Robust Nonparametric Inference on Manifold Spaces

by

Ahmad Mozaffari

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2025

© Ahmad Mozaffari 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Professor Matias Salibian Barrera
Department of Statistics
The University of British Columbia

Supervisor: Professor Shojaeddin Chenouri
Department of Statistics and Actuarial Science
University of Waterloo

Internal Member: Professor Paul Marriott
Department of Statistics and Actuarial Science
University of Waterloo

Internal Member: Professor Pengfei Li
Department of Statistics and Actuarial Science
University of Waterloo

Internal-External Member: Professor Ehsan Toyserkani
Department of Mechanical and Mechatronics Engineering
University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

We propose rank-based procedures for robust and nonparametric statistical inference on manifold spaces. Particularly, we focus on the problems of multi-sample hypothesis testing, multiple change point analysis, and statistical process monitoring when data lie on a Riemannian manifold. These methodologies provide a unified framework to deal with various types of data structures such as matrices, curves, surfaces, networks, to name a few. These types of datasets frequently appear in a broad set of applications such as communication networks, manufacturing, computer vision, autonomous systems and robotics. We evaluate the proposed methods considering various types of object data such as matrices, curves, text mining data, networks, shape data and landmarks.

In Chapter 2, we develop robust and nonparametric methods for hypothesis testing when data lie on a manifold. We demonstrate that ranks generated from data depth can be used for two-sample and multiple sample hypothesis testing of change in location and scale parameters. Several important properties of these tests such as asymptotic convergence, size and power, robustness with respect to qualitative-robustness and breakdown point are developed under mild nonparametric assumptions. These tests have several advantages, they have a simple distribution under null, they are computationally cheap, and they enjoy invariance properties. We demonstrate the efficacy of these methods with a numerical simulation and a data analysis. We show that these tests are robust when data are heavy tailed or skewed, and have higher power compared to their competitors in some situations, while still maintaining a reasonable size.

In Chapter 3, we propose robust and nonparametric single and multiple change point detection methods for stochastic processes defined on manifolds. These methods consider a variant of CUSUM statistic which is based on the rank of data depth. We demonstrate that changes in the rank of depth values can be used to detect change in the distribution of data lie on manifolds. To detect more than one change point, we consider binary segmentation and wild binary segmentation algorithms along with the proposed data depth rank CUSUM statistic. We demonstrate that both of these algorithms are consistent estimators of the number of change point(s) and the location of change point(s). In addition to asymptotic results, we develop nonasymptotic sharp bounds for single and multiple change point estimators. These test statistics can be applied to both intrinsic and extrinsic manifold analysis frameworks. In simulation, we compare our methods against several methods from the literature, and demonstrate that the proposed methods outperform their competitors in some situations where dataset is contaminated with outliers. We also present the application of our methods to vehicle health monitoring, traffic monitoring on highways, and mall pedestrian surveillance.

In Chapter 4, we extend these methods to the setting of statistical process monitoring. We investigate statistical process monitoring scheme on general metric spaces, and propose exponentially weighted moving average, CUSUM, and Mann-Whitney moving average Shewhart control charts using rank of data depth. These methods are nonparametric and robust to outliers through the use of data depth ranks. We show that when sample size is large, our methods have simple behaviour under the null hypothesis. Since our methods are based on data depth ranks, we do not need the estimate of covariance operator which makes our method computationally cheap. Such advantages make these methods a favorable choice for online process monitoring. We demonstrate the robustness of these methods theoretically and numerically. We extract several nonparametric control charts from the literature for comparative study. Simulation results indicated that the proposed methods outperform their competitors in many situations in terms of out-of-control average run length, while keeping the in-control average run length at a reasonable level. We present the application of our methods to laser power-bed fusion additive manufacturing process.

In Chapter 5, we present some possible directions for future research related to dynamic network and longitudinal data analysis on Riemannian manifolds. It is anticipated that the contributions achieved in this thesis will be applicable to a wide range of interdisciplinary research problems.

Acknowledgements

I would like to acknowledge my supervisor for providing guidance and support throughout preparing my thesis. I would also like to thank the members of my committee for taking the time to read the thesis and provide feedback. In addition, I would like to thank my entire family for supporting me along the way! Special thanks to all of my friends for their support in writing this thesis.

Dedication

To my friends and family, who offered words of encouragement and shared in the journey of finishing this study.

Table of Contents

List of Figures xi

List of Tables xix

1 Introduction 1

1.1 Outline of Research . . . 1
1.2 Some Areas of Application . . . 3
1.3 Potential Contributions . . . 5
1.4 Statistics on Manifolds . . . 7
1.4.1 Analysis on General Manifold Spaces . . . 7
1.4.2 Some Manifold Spaces . . . 9
1.5 Robust Statistics on Manifolds . . . 19
1.5.1 Depth Functions on Manifolds . . . 19
1.5.2 Ranking Based on Depths . . . 35
1.6 Literature Review . . . 41

2 Hypothesis Testing 51

2.1 Hypothesis Testing for Manifold Data . . . 51
2.2 Nonparametric Hypothesis Testing . . . 55
2.2.1 Two-sample Hypothesis Testing Of Independence . . . 56
2.2.2 Hypothesis Testing for Quality Monitoring . . . 62

2.2.3 Multiple Sample Hypothesis Testing Of Location and Scale Change 64
2.2.4 Convergence of Empirical Rank-Depth Tests and Power Study . . . 66
2.2.5 Rank-Depth Tests with Estimators of Geodesic Distance 67
2.2.6 Fisher's Permutation Rank-Depth Tests 73
2.2.7 Robustness of Rank-Depth Tests 74
2.3 Simulation Study 88
2.3.1 Two-sample Hypothesis Testing of Independence 89
2.3.2 Hypothesis Testing for Quality Monitoring 103
2.3.3 Multiple Sample Hypothesis Testing of Location and Scale Change 112
2.3.4 Robustness Analysis 116
2.4 Discussion and Future Work 126

3 Change Point Detection **128**
3.1 Change Point Detection for Manifold Data 128
3.2 Change Point Analysis on General Metric Spaces 134
3.3 Rank-CUSUM on Manifold Spaces 139
3.4 Multiple Change Point Analysis 141
3.5 Robustness of Rank-Depth Change Point Detection Methods 145
3.6 Simulation Study 149
3.6.1 Single Change Point Detection 149
3.6.2 Multiple Change Point Detection 168
3.6.3 Robustness Analysis 186
3.7 Discussion and Future Work 187

4 Process Monitoring **189**
4.1 Process Monitoring for Manifold Data 189
4.2 Control Charts 194
4.2.1 Rank-Depth EWMA Control Chart 195

4.2.2 Rank-Depth CUSUM Control Chart . . . 196
4.2.3 Rank-Depth MWMASHC Control Chart . . . 198
4.2.4 Robustness of Rank-Depth Control Charts . . . 199
4.2.5 Average Run Length . . . 202
4.3 Numerical Study . . . 206
4.3.1 Numerical Simulation . . . 206
4.3.2 Robustness Analysis . . . 223
4.4 Additive Manufacturing . . . 227
4.5 Discussion and Future Work . . . 251

5 Future Work **259**
5.1 Longitudinal Data Analysis . . . 260
5.2 Dynamical Networks . . . 261

References **263**

APPENDICES **288**

A appendix **289**
A.1 Proofs from Chapter 2 . . . 289
A.2 Proofs from Chapter 3 . . . 314
A.2.1 Change point on General Metric Spaces . . . 314
A.2.2 Single Change point Detection . . . 329
A.2.3 Multiple Change point Detection . . . 338
A.3 Proofs from Chapter 4 . . . 354

List of Figures

1.1 Schematic illustration of manifold, tangent space, normal space, exponential map, and a path on manifold. 8

1.2 Schematic illustration of 3-spider. 18

1.3 Visualization of SPD(2) matrices using ellipses. A red ellipse is the data point with highest depth and a blue ellipse is the data point with the lowest depth. Depths are calculated using affine invariant metric considering manifold geometry. 24

1.4 Pair-wise comparison of depth values for SPD(2) data using geodesic depth, Zonoid depth, spatial depth, and half-space depth. 25

1.5 Visualization of SPD(2) matrices using ellipses. A red ellipse is the point with the highest depth and a blue ellipse is the point with the lowest depth. Depths are calculated using Euclidean metric. 26

1.6 Geodesic path between two 3-spiders. 28

1.7 Visualization of 3-spider data. Depths are calculated using intrinsic embedding. Each axis represents different tree topology and the location of data on axis represents the length of the interior edge. Red circles are the 3-spider with highest depth and blue circles are the 3-spiders with lowest depth. . . 29

1.8 Pair-wise comparison of depth values for 3-spider data using geodesic depth, Zonoid depth, spatial depth, and half-space depth. 30

1.9 The complete set of 60 3-spider tree structures with the deepest structure shown inside the red box. 31

1.10 Visualization of 3-spider data. Depths are calculated using extrinsic embedding. Each axis represents different tree topology and the location of data on axis represents the length of the interior edge. Red circles are the 3-spider with highest depth and blue circles are the 3-spiders with lowest depth. 32

1.11 Depth color map for circular data for geodesic depth and spatial depth with Euclidean metric and ISOMAP, (a) samples from the distribution together with the density color map, (b) color map contour plot of GD, (c) color map contour plot of SD, (d) weighted undirected graph $\mathcal{G}$ on the sample points using k smallest distances, (e) color map contour plot of GD based on ISOMAP, (f) color map contour plot of SD based on ISOMAP. 34

1.12 Circular normal distribution centered at angular frequency ω. 37

1.13 (a) Distribution of points sampled from F_1 and F_2, (b) SD values of samples drawn from $\mathrm{N}\big(0, 2\big)$, and $\mathrm{N}\big(0, 4.25\big)$. 38

1.14 Ellipse with highest depth (red color) and ellipses with lowest depth (blue and purple color). Density colormap for the eigenvalues of the SPD(2) dataset. The points with highest and lowest depths are shown inside circles in the colormap. 40

1.15 Rank of SPD(2) data, data point with highest rank is shown with red circle and data points with lowest ranks are shown with blue and purple circles. . 40

2.1 Schematic illustration of the graph distance estimator of geodesic distance in the constructed neighbourhood graph. The estimated distance equals the length of the blue edges which form the shortest path from x to y on the graph. 68

2.2 Schematic illustration of Midsagittal landmarks from CT scans of 24 adult human (16 females, 8 males) with a total of 87 landmarks comprises of 32 anatomical landmarks (white landmarks) and 55 semilandmarks (black dots). 71

2.3 (a) Three possible paths from node 5 to node 21, the shortest path "5-8-15-16-21" is shown in bold red color, (b) the correlation between GGD based on graph distance of k-ad shape data and GD based on distance induced by the embedded Veronese-Whitney space for 24 sample data, (c) graph of 100 k-ad samples obtained from Mardia-Dryden distribution, (d) illustration of the 100 samples, (e) correlation between GGD and GD for 100 samples, (f) illustration of the 1000 samples, (g) correlation between GGD and GD for 1000 samples. 72

2.4 Power plots for data from Wishart distribution, (a) Power curves for Spearman correlation coefficient test with HD, GD, and SD depths, (b) Power curves for Quadrant test with HD, GD, and SD depths, (c) Power curves for Spearman correlation coefficient test with GD, Quadrant with GD, and ball covariance test. 91

2.5 Schematic illustration of torus with data generated from a von Mises distribution. 92

2.6 Power plots for data from a von Mises distribution, (a) Power curves for Spearman correlation coefficient test with HD, GD, and SD depths, (b) Power curves for Quadrant test with HD, GD, and SD depths, (c) Power curves for Spearman correlation coefficient test with GD, Quadrant with GD, and ball covariance test. 94

2.7 Power plots for von Mises Fisher distribution data, (a) Power curves for Spearman correlation coefficient test with HD, GD, and SD depths, (b) Power curves for Quadrant test with HD, GD, and SD depths, (c) Power curves for Spearman correlation coefficient test with GD, Quadrant with GD, and ball covariance test. 96

2.8 Schematic illustration of communication networks for the first two months. 101

2.9 Schematic illustration of communication networks for the last two months. 101

2.10 Power plots for community network data, (a) Power curves for Spearman correlation coefficient test with HD, GD, and SD depths, (b) Power curves for Quadrant test with HD, GD, and SD depths, (c) Power curves for Spearman correlation coefficient test with GD, Quadrant with GD, and ball covariance test. 102

2.11 Schematic illustration of a healthy product. 104

2.12 Schematic illustration of a faulty product. 104

2.13 Schematic illustration of the contour line of healthy products. 105

2.14 Schematic illustration of the contour line of faulty products. 105

2.15 Schematic illustration of DNA data, top panels indicate 3D view of a sample of real world healthy DNA molecule structure data with 2D projections, bottom panels indicate 3D view of damaged DNA molecule with deformed structures as well as their 2D projections generated from Mardia-Dryden distribution. 106

2.16 (a) Geodesic-based Rank-Depth for reference sample, (b) 3D view of healthy DNA molecule structure, (c) 3D view of damaged DNA molecule structures with lowest geodesic depth. 107

2.17 Power plot for DNA molecule data with power curves for Wilcoxon test with GD, energy test, GET, and geodesic T^2 test. 108

2.18 The deepest phylogenetic tree with respect to geodesic depth. 109

2.19 Four individual phylogenetic trees with the least geodesic depth value. . . . 111

2.20 Power plot for phylogenetic tree data with power curves for Wilcoxon test with GD, energy test, GET, and geodesic T^2 test. 112

2.21 Power plots for covariance matrices data considering Kruskal-Wallis test with GD, energy test, and ball divergence test, (a) case I, (b) case II. . . . 115

2.22 Power plot for covariance matrices data with power curves for Kruskal-Wallis test with GD, energy test, and ball divergence test. 117

2.23 (a) level breakdown point histogram for Spearman test with $n_y = 50$, (d) power breakdown point histogram for Spearman test with $n_y = 50$, (e) Effect of contamination level on Spearman test statistic value under null hypothesis, (f) Effect of contamination level on Spearman test statistic value under alternative hypothesis. 118

2.24 (a) Distribution of data with density contour map, (b) grid points on sample space domain, (c) level breakdown point histogram for Wilcoxon test with $n_y = 50$, (d) power breakdown point histogram for Wilcoxon test with $n_y = 50$, (e) Effect of contamination level on Wilcoxon test statistic value under null hypothesis, (f) Effect of contamination level on Wilcoxon test statistic value under alternative hypothesis. 121

2.25 A sample of $K = 3$ groups of data generated from VMF distribution on the space $\mathcal{S}^2$. 123

2.26 (a) Distribution of data with density contour map (top-front view), (b) grid points on sample space domain, (c) level breakdown point histogram for Kruskal-Wallis test with $n_y = 50$, (d) power breakdown point histogram for Kruskal-Wallis test with $n_y = 50$, (e) Effect of contamination level on Kruskal-Wallis test statistic value under null hypothesis, (f) Effect of contamination level on Kruskal-Wallis test statistic value under alternative hypothesis. 124

3.1 Statistical summaries for the estimated locations of CUSUM and r-CUSUM for Wishart distribution with breaking fractions $\theta = \{0.25, 0.50, 0.75\}$ and sample sizes $n = \{50, 100, 200\}$. 151

3.2 Probability density function data objects for no change, change in location and change in scale scenarios. 152

3.3 Statistical summaries for the estimated locations of CUSUM and r-CUSUM for pdf curve data under change in location scenario with breaking fractions $\theta = \{0.25, 0.50, 0.75\}$ and sample sizes $n = \{50, 100, 200\}$. 154

3.4 Statistical summaries for the estimated locations of CUSUM and r-CUSUM for pdf curve data under change in scale scenario with breaking fractions $\theta = \{0.25, 0.50, 0.75\}$ and sample sizes $n = \{50, 100, 200\}$. 155

3.5 Yearly maternal age probability density functions for the age interval $[12, 55]$ for the period of 1939 to 2021. 156

3.6 Absolute value of Z_n test statistic for r-CUSUM and d-CUSUM for Finland fertility data. 157

3.7 Cluster analysis and MDS plot for Dickens and Austen novels. 160

3.8 Networks indicating top 100 pairs of words for Dickens and Austen. 160

3.9 Data sampled from von Mises distribution on 2-torus $\mathbb{T}^2$. Left panel indicates two groups of data sampled at different locations $(\mu_1 = \pi/2, \nu_1 = \pi/2)$ and $(\mu_2 = 3\pi/2, \nu_2 = \pi/2)$ with the same scale parameters $(\kappa_1 = 5, \kappa_2 = 5, \kappa_3 = 2)$. Right panel indicates the corresponding depth values of these two groups. 164

3.10 The performance of SS-CUSUM chart under null and alternative with effect size $\delta = 0.7$. 165

3.11 Empirical power for 2-torus data, (a) performance of SS-CUSUM with HD, GD, and SD depth functions for location change, (b) performance of r-CUSUM with HD, GD, and SD depth functions for scale change. 166

3.12 Data sampled from von Mises distribution on 2-torus $\mathbb{T}^2$. Left panel indicates two groups of data sampled at same location $(\mu_1 = \pi/2, \nu_1 = \pi/2)$ and different concentration parameters $(\kappa_1 = 3, \kappa_2 = 50, \kappa_3 = 0)$ and $(\kappa_1 = 50, \kappa_2 = 3, \kappa_3 = 0)$. Right panel indicates the corresponding depth values of these two groups. 167

3.13 Landmark representation of pupfishes are shown in left panel. Digitized representation of the average shape is given in right panel. 177

3.14 Principal component plot of shape variation in the population of pupfishes. 178

3.15 Shape difference between marsh and sinkhole pupfish specimens displayed as (a) thin-plate spline deformation grid for two extremes, (b) change direction vector of two extremes with respect to mean shape. 178

3.16 Change points obtained by r-BS. Three panels shown the location of changes in three stages of binary segmentation corresponding to locations $15, 29, 44$. 179

3.17 Geodesic depth values for flag object data. 182

3.18 Detected changes in the intensity of vehicles moving on the road. 182

3.19 Detected changes in the intensity of passengers for mall pedestrian surveillance data. 184

3.20 Test vehicles: (a) A Cadillac ATS, with electronic limited slip differential on the rear axles; (b) fully electrified vehicle with independent-drive configurations 184

3.21 Detected changes for the correlation matrices of the considered vehicle platforms, (a) Chevrolet Equinox segments (i) $[1, 63]$, (ii) $[64, 119]$, (iii) $[120, 158]$, (iv) $[159, 170]$; (b) Cadillac ATS segments (i) $[1, 122]$, (ii) $[123, 142]$, (iii) $[143, 158]$, (iv) $[159, 190]$; (c) Cadillac CTS segments (i) $[1, 117]$, (ii) $[118, 170]$ 185

3.22 Gross-error sensitivities for r-CUSUM, d-CUSUM, AMOC, and ECP for Wishart distribution under null and alternative hypotheses with different values of ε 186

4.1 Average run length for IC and OC conditions. 202

4.2 Flowchart of Phase-I and Phase-II stages of process monitoring. 205

4.3 MC error for OC-ARL with r-EWMA and replication numbers $(1, 10000)$ for Wishart distribution with dimensions $p = (3, 6, 10)$. 211

4.4 OC-ARL values for the process monitoring control charts for Wishart and contaminated Wishart with shift sizes $(0, 10)$ and dimensions $(3, 11, 21)$. . . 212

4.5 OC-ARL values for the process monitoring control charts for VMF and contaminated VMF with shift sizes of range $(0, 10)$ and dimensions $(3, 10, 20)$. 215

4.6 Schematic illustration of Steifel manifold data corresponding to 2-frames in $\mathbb{R}^3$. 216

4.7 OC-ARL values for the process monitoring control charts for Gaussian Steifel and contaminated Gaussian Steifel with shift sizes of range $(0, 0.1)$ and dimensions $(3, 11, 20)$. 219

4.8 Computation time for the considered control charts. 221
4.9 Process monitoring with r-EWMA, r-CUSUM, r-MWMASHC, IPD, DDMA, and RE control charts and their control limits for Wishart distribution. . . 222
4.10 ARL boxplots for the process monitoring control charts r-EWMA, r-CUSUM, and r-MWMASHC under contamination noises with fractions $(0\%, 10\%, 30\%, 50\%)$. 223
4.11 Gross-error sensitivities for r-EWMA, r-CUSUM, and r-MWMASHC for IC and OC conditions with different values of ε. 224
4.12 Gross-error sensitivities for r-EWMA, r-CUSUM, and r-MWMASHC for IC and OC conditions with different values of ε. 225
4.13 Gross-error sensitivities for r-EWMA, r-CUSUM, and r-MWMASHC for IC and OC conditions with different values of ε. 226
4.14 Shape point clouds representing a part of turbine propeller fan, and illustration of a layer-wise optical tomography image data object. 230
4.15 Two different off-axial camera mounting strategies (Taherkhani, 2022). . . 232
4.16 Process monitoring algorithm for image data taken from each layer of product. 232
4.17 Illustration of healthy product along with a number of faulty products obtained in a photodiode additive manufacturing process. Right panel shows how a control chart monitors the quality of images taken from photodiode AM process in Phase-II. 233
4.18 First panel indicates the change in the distribution of depths for image data. Next panels indicate the performance of r-EWMA, r-CUSUM, and r-MWMASHC control charts in Phase-II. 234
4.19 Different views of spiral bevel gear and coin loading tube geometries. . . . 235
4.20 Fitted Fourier series to the contours of spiral bevel gear and coin loading tube. 237
4.21 Shape point clouds representing spiral bevel gear and coin loading tube. . . 240
4.22 Schematic illustration of a number of detected products with defective contours, (a) spiral bevel gear, (b) coin loading tube. 241
4.23 The performance of r-EWMA, r-CUSUM, and r-MWMASHC control charts in Phase-II. 243
4.24 Illustration of healthy and faulty products observed in noise. The left panel indicates the contour of a healthy product. The right panel indicates the contour of a deformed product generated by Mardia-Dryden distribution. . 244

4.25 Sample images taken from set R2 location 8 (Taherkhani, 2022). 245

4.26 Top panel indicates a printed product. Bottom left panel indicates an image generated by software from a layer of the print. Bottom right panel indicates 49 regions of interest around each defect at the zoomed location (Taherkhani, 2022). 246

4.27 The performance of r-EWMA, r-CUSUM, and r-MWMASHC control charts in Phase-II. 247

4.28 A sample of healthy and faulty images from layers of printed product. The faulty images include void defects in the layers of printed product (Taherkhani, 2022). 250

4.29 The performance of r-EWMA, r-CUSUM, and r-MWMASHC control charts in Phase-II. 251

4.30 Gross-error sensitivities for r-EWMA, r-CUSUM, and r-MWMASHC for IC and OC conditions with different values of ε. 252

List of Tables

2.1 Independence test simulation results for Wishart distribution. Reported values are the power of the test statistics. 90

2.2 Independence test simulation results for von Mises distribution. Reported values are the power of the test statistics. 93

2.3 Independence test simulation results for von Mises-Fisher distribution. Reported values are the power of the test statistics. 95

2.4 Simulation results for community network distribution, Case I. Reported values are the size of the test statistics. 100

2.5 Simulation results for community network distribution, Case II. Reported values are the power of the test statistics. 100

2.6 Simulation results under null hypothesis for multiple groups of covariance matrices. 114

2.7 Simulation results under alternative hypothesis for multiple groups of covariance matrices. 114

2.8 Simulation results for social network data. 116

2.9 Simulation results for Spearman correlation coefficient test. The results are reported for level breakdown point (ε_{SA}), and power breakdown point (ε_{SR}). 119

2.10 Comparison of ε_{SA} and ε_{SR} values for Spearman correlation coefficient test, Quadrant test, and Ball covariance test. 119

2.11 Simulation results for Wilcoxon test. The results are reported for level breakdown point (ε_{SA}), and power breakdown point (ε_{SR}). 120

2.12 Comparison of ε_{SA} and ε_{SR} values for Wilcoxon test, geodesic T^2 test, energy test, and GET test. 122

2.13 Simulation results for Kruskal-Wallis test. The results are reported for level breakdown point (ε_{SA}), and power breakdown point (ε_{SR}). 125

2.14 Comparison of ε_{SA} and ε_{SR} values for Kruskal-Wallis test, BallDiv test, and energy test. 125

3.1 Empirical size under the null for Wishart distribution for 5000 simulation runs. 150

3.2 Empirical power under the alternative for Wishart distribution for 5000 simulation runs. 150

3.3 Empirical size under the null for pdf curve data for 5000 simulation runs. . 153

3.4 Empirical power under the change in location alternative for pdf curve data for 5000 simulation runs. 153

3.5 Empirical power under the change in scale alternative for pdf curve data for 5000 simulation runs. 153

3.6 List of Jane Austen and Charles Dickens novels. 159

3.7 Summary statistics for change point location estimates using Bootstrapping with $B = 1000$. 161

3.8 Empirical power of change point detection methods. 168

3.9 Summary statistics for the empirical distribution of $\widehat{\ell}$, and MSE, MAD, and relative efficiency (w.r.t. WBS) of $\widehat{\ell} - \ell$ for pdf curve data over 5000 simulation runs. 169

3.10 Estimated change points of the rival methods for pdf curve data over 5000 simulation runs. 170

3.11 Summary statistics for the empirical distribution of $\widehat{\ell}$, and MSE, MAD, and relative efficiency (w.r.t. WBS) of $\widehat{\ell} - \ell$ for Wishart distribution over 5000 simulation runs. 171

3.12 Estimated change points of the rival methods for Wishart distribution. . . 173

3.13 Summary statistics for the empirical distribution of $\widehat{\ell}$, and MSE, MAD, and relative efficiency (w.r.t. WBS) of $\widehat{\ell} - \ell$ for Wishart distribution contaminated with noise over 5000 simulation runs. 174

3.14 Estimated change points of the rival methods for Wishart distribution contaminated with noise. 175

3.15 Estimated change points of the rival methods for pupfish data. 180

4.1 Control limits coefficient γ_0 for r-EWMA, r-CUSUM, and r-MWMASHC control charts for $ARL_0 = 200$ and $ARL_0 = 370$ for different manifold intrinsic dimensions. 207

4.2 IC-ARL values for r-EWMA, r-CUSUM, and r-MWMASHC for Wishart distribution for 5000 simulation runs. The values in the parentheses are standard deviations. 209

4.3 OC-ARL values for r-EWMA, r-CUSUM, and r-MWMASHC for Wishart distribution for 5000 simulation runs. The values in the parentheses are standard deviations. 210

4.4 IC-ARL values for r-EWMA, r-CUSUM, and r-MWMASHC for von Mises-Fisher distribution for 5000 simulation runs. The values in the parentheses are standard deviations. 213

4.5 OC-ARL values for r-EWMA, r-CUSUM, and r-MWMASHC for von Mises-Fisher distribution for 5000 simulation runs. The values in the parentheses are standard deviations. 214

4.6 IC-ARL values for r-EWMA, r-CUSUM, and r-MWMASHC for Uniform distribution on Steifel manifold for 5000 simulation runs. The values in the parentheses are standard deviations. 217

4.7 OC-ARL values for r-EWMA, r-CUSUM, and r-MWMASHC for Uniform and Gaussian distributions on Steifel manifold for 5000 simulation runs. The values in the parentheses are standard deviations. 218

4.8 Percentage of times a method outperformed the other rival methods. 220

4.9 IC-ARL and OC-ARL values for IC and OC data obtained over 5000 simulation. 233

4.10 Computed UCL values. Numbers in parentheses are the iteration until convergence. 236

4.11 IC-ARL values for in-control simulation scenario. Numbers in parentheses are standard deviations over 5000 simulation runs. 238

4.12 OC-ARL values for out-of-control simulation scenario. Numbers in parentheses are standard deviations over 5000 simulation runs. 238

4.13 IC-ARL values for in-control simulation scenario. Numbers in parentheses are standard deviations over 5000 simulation runs. 239

4.14 OC-ARL values for out-of-control simulation scenario. Numbers in parentheses are standard deviations over 5000 simulation runs. 239

4.15 IC-ARL and OC-ARL values for IC and OC data obtained over 5000 simulation. 244
4.16 IC-ARL and OC-ARL values for IC and OC data obtained over 5000 simulation. 247
4.17 IC-ARL and OC-ARL values for IC and OC data obtained over 5000 simulation. 249

Chapter 1

Introduction

In this chapter, the outline of the carried out research will be discussed. The importance of the statistical inference on manifold spaces will be demonstrated considering several important areas of applications which require developing tools appropriate for non-Euclidean spaces, i.e. space of symmetric positive definite (SPD) matrices, networks, graphs, curves, shapes, etc. Thereafter, the mathematical tools required for implementing statistical methods will be provided. These include depth functions on manifold spaces, ranking methods, and several examples of manifold spaces which will be used for simulation studies throughout this thesis. Finally, a literature review will be given which discusses some important research areas in the field as well as advances in statistical inference on manifold spaces.

1.1 Outline of Research

In recent decades, with the emergence of high-speed computational systems and advanced sampling tools, the field of statistical data analysis has witnessed a significant revolution. In particular, it is now possible to collect high-dimensional datasets with the goal of identifying various aspects and key features of real-world processes. Statistics on high-dimensional datasets is not an easy task, and careful analysis is required to properly extract representative geometrical information, which can provide insights into the key properties of the corresponding processes. In line with such an interest, several research areas have been introduced, which include regularized regression on high-dimensional spaces , variable selection methods, functional data analysis, principal component analysis, see the nice reviews in Nosedal-Sanchez et al. (2012), O'Hara and Sillanpää (2009), Wang, Chiou and Müller (2016), Abdi and Williams (2010), respectively. One difficulty associated with

such approaches is that for many real-world datasets, these seemingly high-dimensional data actually lie in a lower dimensional subspace which is nonlinear. Therefore, traditional multivariate statistical analysis tools cannot afford desirable outcomes mainly due to the fact that the nonlinearity of such subspaces has not been taken into account.

Statistics on manifold data has originally been proposed to learn appropriate low-dimensional subspaces which can properly represent data lying on high-dimensional and infinite-dimensional spaces, see for example (Murray and Rice, 1993). It has been proven that after finding the intrinsic dimension of high-dimensional datasets, coordinate charts can be developed to make statistical analysis possible considering the local geometry of the manifold. Such a strategy is known as intrinsic manifold analysis, which is based on differential geometry and finding a collection of charts to represent the whole manifold space. The other possible approach for manifold data analysis is extrinsic analysis which is based on applying a mapping which transfers data from a nonlinear manifold space to a high dimensional Euclidean space such that the distance of any two points is preserved (Qiu et al., 2008). It has been demonstrated that statistical analysis on manifolds yields promising outcomes (Izenman, 2012). In particular, inference on the low dimensional space can afford more realistic conclusions and insights, the computational power will be saved by working on a lower dimensional dataset, which preserves the information stored in the seemingly high-dimensional datasets, and asymptotic analysis can be performed as a function of intrinsic dimension which makes the convergence rates unaffected from the dimension of the seemingly high-dimensional datasets (Cheng and Xie, 2021*a*).

Although research on statistics on manifolds has gained remarkable attention and several research domains such as analysis of longitudinal and functional data (Muralidharan and Fletcher, 2012), hypothesis testing (Zhang, 2021), change point detection Duan et al. (2019), dimensionality reduction (Ham et al., 2004), classification, clustering (Lee and Wilkinson, 2019), shape analysis (Hanik et al., 2020), and tensor computing (Dryden et al., 2009) on manifold spaces have been introduced in the literature, there is a clear gap related to robust statistics on manifold spaces which needs further attention. Robust statistics has been an active research field over the past several decades. It has been proven to show improved results when the collected data follow heavy-tail and asymmetric distribution or are contaminated with outliers. The success of robust statistics has been proven in domains such as regression, exploratory data analysis, hypothesis testing, change point detection and analysis of variance (Maronna et al., 2019; Huber, 2011).

Given the promising reports in the literature regarding the use of robust statistics for multivariate and functional data analysis, in the current research, the author intends to investigate and assess the potential of robust statistics on manifold spaces. In particular, the goal of the research is to implement a robust ranking mechanism based on the

depth of data sampled from general metric spaces and curved manifold spaces and use it for developing statistics for hypothesis testing, change point detection, and statistical process monitoring. Comparative study will be carried out considering various synthetic and real-world datasets lying on various manifolds, which are contaminated with various types of heavy-tailed distributions and outliers in order to demonstrate the efficacy of robust analysis on manifolds. Simulation results will be supported by theoretical analysis concerning the convergence of the robust estimators as well as the power of ranking-based tests implemented for hypothesis testing, change point detection, and process monitoring.

The rest of the chapter is organized as follows. In the second section, some important areas of application for statistics on manifold spaces are given to clarify the importance of the topic of the current research. The third section is devoted to discussing the contributions of the research, where we try to address some research areas with open ground for theoretical and numerical analysis. The fourth section is devoted to mathematical definitions and the required background for implementing manifolds of interest with several real-world examples. In the fifth section, we discuss robust statistics on manifolds. Several variants of depth functions and the concept of ranking based on depth functions will be introduced. Examples are presented which elaborate on the importance of developing depth functions taking advantage of affine invariant geometry when working with data which have non-Euclidean nature. The last section is devoted to a literature review explaining contributions related to statistics on data sampled from various manifold spaces together with their theoretical and numerical findings.

1.2 Some Areas of Application

Statistics on manifold spaces has been applied to a wide range of complicated problems for which standard Euclidean-based statistical analysis is not suitable. A generalized version of manifold data is known as object data which includes datasets defined on Riemannian manifolds as well as datasets which belong to non-Riemannian spaces, i.e. spaces for which differential geometry and a proper generalized Riemannian metric cannot be defined (Gromov et al., 1999). Techniques and methodologies required for applying the manifold analysis include Frèchet mean and variance calculation, curve registration, visualization of a graph structure of the manifolds, smoothing, dimension reduction and intrinsic dimension, data visualization via heat map of matrix data, curve view of matrices and modes of variation, scatter plots, scree plot, etc. (Marron and Dryden, 2021). Such methods have successfully been applied to various important real-world applications. Examples include medicine and human cognition, e.g. skull analysis, gene expression analysis, clinical data

analysis, human behaviour analysis, and investigating brain connectivity, industrial applications, e.g. process monitoring of shape of products and surface curves and heat map matrices, text mining and inference about handwritten data, etc.

In the listed applications, the theories are usually developed in an infinite dimensional continuous space, and then for data analysis purpose, the results are simplified via discretization, e.g. graphs, and matrices, in order to make the numerical analysis and statistical inference possible. Manifold analysis also enables us to directly apply the theoretical findings to spaces such as graphs and matrices, which are valuable because many applications in the real-world are defined on spaces such as matrix spaces and graphs. Below, some important real-world examples are presented, and the potential of manifold analysis for statistical inferences are discussed.

- **Medical Imaging:** Combining change point analysis method with techniques available in localized manifold learning, such as using Laplacian operator, and distances such as those used in graphs to calculate the degree of centrality for nodes can be used to obtain information about the importance of any pixel in images of brain and other organs. Doing so enables us to obtain a statistic (functional on space of manifolds) and a density on manifold with respect to the statistic (using the history of collected images) and then monitor sudden significant changes in the structure of images (a localized analysis of pixels) (Brosch et al., 2013). In Rué-Queralt et al. (2021), the authors proposed a manifold learning framework to estimate the intrinsic dimension of brain dynamics to reveal the nonlinear differences of brain activity in wakefulness and sleeping stages.

- **Social Networking and Community Detection:** Manifold analysis can be used to come up with statistics to test hypotheses related to important questions arising in the field of social networking (You et al., 2016). This can be done by taking a localized analysis approach on manifolds, and techniques such as kernels and Laplacian can be advantageous (Deng et al., 2021). Similar to analysis in stochastic block model (SBM), it is possible to develop graph models to extend the analysis to general manifolds and explore ideas which enable us to detect hidden structural properties that lie within the collected set of manifold data.

- **Shape Analysis and Process Monitoring:** An important application of manifold learning is in monitoring the quality of products based on the structural properties of their shape functions. We can use localized and intrinsic manifold learning approaches to verify whether the products are in control by quantifying the distribution of healthy products based on obtaining a distribution on the domain in which the

shapes are defined (Liu et al., 2021). Standard charts such as Shewhart, CUSUM and EWMA can be extended to manifolds for process monitoring (Zhao and Del Castillo, 2021).

- **Set with Constraints for Hypothesis Testing and Smooth Tests:** Set with constraints can be defined on manifolds. For example, conditional distributions given constraints, e.g. sufficient statistics, can be viewed as a set of points forming a manifold. Hypothesis tests can be implemented with respect to such manifolds, e.g. testing if a random sample is independently and identically distributed from a Gamma distribution. Similarly, the Neyman's smooth test can be defined on such spaces, which can be used to test if the CDF from which a sample is collected is smooth (Diaconis et al., 2013).

- **Identifying Hidden Pattern in a High Dimensional Dataset:** Manifold learning (Izenman, 2012) can be used to unveil the hidden low-dimensional manifold that lies within a seemingly high-dimensional space. To do so, tools available to obtain the intrinsic dimension can be used to represent the data on a manifold in lower dimensional spaces (Raducanu and Dornaika, 2012). Pattern recognition in such a frame can be applied to problems working with point clouds and can help to understand images and extract geometrical information from complex datasets. Applications include problems arising in security, such as facial expression.

From the abovementioned research areas, it is apparent that manifold learning plays a key role in facilitating the modeling and the statistical inference analysis. Specially, many of these complex problems are represented with high-dimensional data for which it is possible to estimate a low dimensional nonlinear manifold which preserves the important geometrical information and, therefore, provides a useful insight into the true structure of the high-dimensional dataset.

1.3 Potential Contributions

In this section, the main motivating goals for the research are remarked upon, which will be scrutinized in the next chapters for developing methodologies for statistical manifold analysis. The main goals are as follows:

- To do a robust analysis on manifold spaces, it is important to investigate and demonstrate the efficacy of depth functions defined on manifold spaces. Depth functions

on manifold spaces are congruence invariant (equivalent to affine invariance in vector spaces) and, therefore can be properly used with a rank function to develop robust statistic for manifold analysis. Also, theoretical results will be provided to investigate the rates of convergence for the estimators based on rank of depth functions.

- Hypothesis testing on manifold spaces is an important field of research which has attracted significant attention. However, to the author's best knowledge, research on robust hypothesis testing for manifold data still needs further attention. In line with this, robust versions of two-sample and multiple-sample hypothesis tests for manifold data will be proposed, and their theoretical properties will be investigated. A comparative study, together with real-world examples, will be considered to evaluate the performance of the tests.

- Change point analysis is an important research topic which has recently received a considerable attention in multivariate and functional data. Formulating a change-point detection algorithm for manifold data and performing theoretical analysis is of paramount importance due to the fact that modern statistical applications is more related to the study of complex phenomena with random objects sampled from general metric spaces. In line with such an interest, the author proposes a robust change point detection on manifold spaces. The efficacy of the proposed framework is demonstrated on a number of important real-world problems. Theoretical result are also provided to support the numerical findings and give insights into the asymptotic properties of the proposed robust change point statistic.

- Process monitoring and quality control on Euclidean spaces are well-established research fields. One of the other goals of the current research is to probe the potential of robust manifold learning for developing control charts for analyzing object data sampled from manifold spaces. Theoretical results will be provided to explore the asymptotic properties of the proposed methodologies. The finding will be endorsed by numerical simulation.

In the next sections, the required background including the concept of depth function and ranking as well as robust statistics are provided to lay the foundation for the development of the methodologies. Various types of manifolds are given, which will be used for numerical simulations. Furthermore, a literature review on the topic of statistics on manifolds is presented to give the readers an insight into what has been done and what are the potentials for further research, clarifying the importance of the development of robust statistics on manifold spaces.

1.4 Statistics on Manifolds

In this section, details related to performing statistical analysis on data sampled from manifold spaces are given. Firstly, the tools and concepts required for statistical analysis on general manifold spaces are provided, and thereafter, the analysis is extended to a number of manifold spaces (Lee et al., 2009) considered in the current study.

1.4.1 Analysis on General Manifold Spaces

A subset $\mathcal{M}$ of ambient space $\mathbb{R}^{m+k}$ has intrinsic dimension m if for any $x \in \mathcal{M}$, there exists a neighbourhood $U \subset \mathbb{R}^{m+k}$ and an everywhere surjective differential map $f : U \to \mathbb{R}^k$ such that $U \cap \mathcal{M} = f^{-1}(0)$. The set $\mathcal{M}$ is called an m-dimensional submanifold of $\mathbb{R}^{m+k}$. The tangent space $T_x\mathcal{M}$ at $x \in \mathcal{M}$ is an m-dimensional subspace with tangent vectors having the form $\alpha'_\epsilon(0)$ with $\alpha : (-\epsilon, \epsilon) \to \mathcal{M}$ being curves on manifold such that $\alpha_\epsilon(0) = x$ and $\epsilon > 0$. The tangent space $T_x\mathcal{M}$ can be represented by m vectors e_i's forming an orthonormal basis for the space. The tangent bundle TM is the union of all tangent spaces at each point of the manifold, i.e.

$$TM = \bigcup_{x \in \mathcal{M}} T_x\mathcal{M} \ .$$

A Riemannian manifold is endowed with the Riemannian inner product $\langle \cdot, \cdot \rangle_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}$, which is a bilinear, symmetric and positive definite inner product at each point $x \in T_x\mathcal{M}$ varying smoothly through points on manifold, whose value is equal to inner product in the ambient space $\mathbb{R}^{m+k}$.

The length of a piecewise differentiable path $\alpha : [0, 1] \to \mathcal{M}$ is defined as

$$\ell(\alpha) = \int_0^1 \sqrt{\langle \frac{d\alpha}{dt}, \frac{d\alpha}{dt} \rangle_x} dt \ .$$

The Riemannian distance between two points $x, y \in \mathcal{M}$ is defined as

$$\delta(x, y) = \inf \left\{ \ell(\alpha); \ \alpha : [0, 1] \to \mathcal{M}, \alpha(0) = x, \alpha(1) = y \right\} \ .$$

If $\mathcal{M}$ is a completely connected Riemannian manifold, then by Hopf-Rinow theorem, $\mathcal{M}$ is a geodesic space, implying that there exists a unique geodesic segment joining any two points $x, y \in \mathcal{M}$.

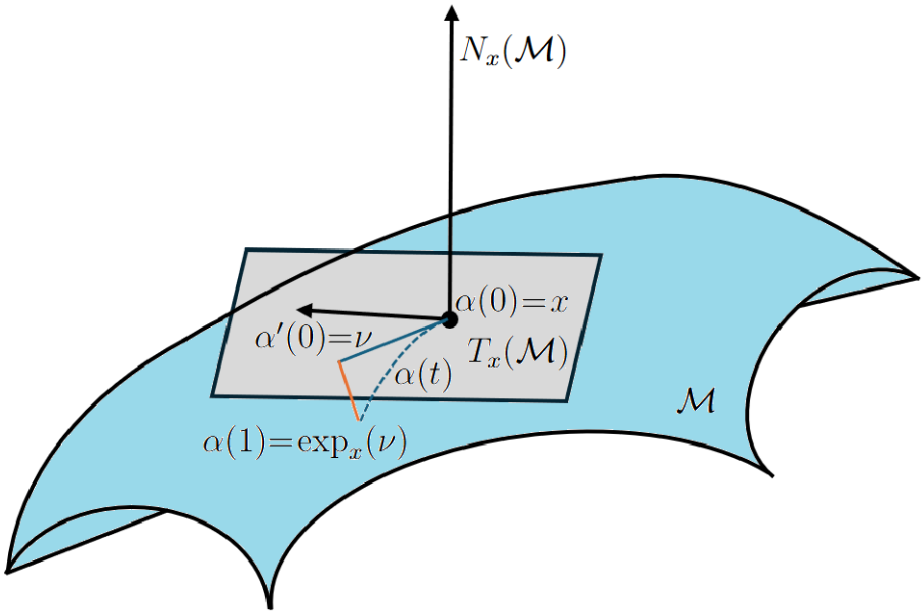


Figure 1.1: Schematic illustration of manifold, tangent space, normal space, exponential map, and a path on manifold.

The Riemannian exponential map at $x \in \mathcal{M}$ is a map $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ such that $\exp_x(\nu)$ is the endpoint $\alpha(1)$ of a unit-speed geodesic $\alpha : [0,1] \to \mathcal{M}$ with $\alpha(0) = x$, $\alpha'(0) = \nu$ and $\|\alpha'(t)\| = 1$ for all t. The inverse of an exponential map is called logarithmic (inverse exponential) map denoted by $\log_x : \mathcal{M} \to T_x\mathcal{M}$. Figure 1.1 illustrates these analysis tools and ideas. For a smooth function $f : \mathcal{M} \to \mathbb{R}$, the Riemannian gradient $\nabla_x f$ in the direction of vector $\nu \in T_x\mathcal{M}$ determines the rate of change in the direction of vector ν, which implies that for any curve $\alpha : [0,1] \to \mathcal{M}$ with $\alpha(0) = x$, $\alpha'(0) = \nu$, it holds that

$$\langle \nu, \nabla_x f \rangle_x = \frac{d}{dt} f(\alpha(t)) \mid_{t=0} \quad . \tag{1.1}$$

The above tools and concepts enable us to extend statistical analysis from Euclidean spaces to nonlinear manifold spaces. The general steps of data analysis on manifolds can be summarized as follows:

- Recovering data topology.
- Determining intrinsic dimension m of underlying manifold.

- Developing a neighbourhood system.
- Representing intrinsic coordinates via calculating local parametrization of the neighbourhood system.

1.4.2 Some Manifold Spaces

In this section, the geometrical properties of a number of manifolds are studied. The manifolds listed in this section are important in the sense that they entail several emerging real-world problems. In each of the subsections, the possible applications of the considered manifolds are also discussed to demonstrate their practical importance.

Example 1 (**Matrix Lie Groups**)**.** A group G is called a Lie group if it has a differentiable structure, i.e. for $x, y \in G$ the operators

$$\jmath : G \times G \to G, \quad (x, y) \to x \cdot y \ ,$$

$$\imath : G \to G, \quad x \to x^{-1} \ ,$$

are differentiable (Baker, 2003). A Lie group whose elements are matrices is called a matrix Lie group. The tangent space of Lie group at identity is called Lie algebra $\mathfrak{g}$, which is a vector space endowed with an operator called Lie bracket, satisfying the Jacobi identity. Therefore, the elements of G and $\mathfrak{g}$ are related through exponential and logarithmic maps. The exponential map is homomorphic mapping from a neighbourhood of identity $\mathbf{0} \in \mathfrak{g}$ to a neighbourhood of identity $I \in G$.

The product operator in logarithmic coordinate is a map $\nu : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ satisfying the following property

$$\exp(u)\exp(v) = \exp(\nu(u, v)) \ .$$

Taylor series of $\nu(U_1, U_2)$ is represented by Baker-Campbell-Hausdorff formula around point $(\mathbf{0}, \mathbf{0})$

$$\nu(u, v) = u + v + [u, v] + O(\mid (u, v) \mid^3) \ ,$$

where $[u, v]$ denotes the Lie bracket on the Lie algebra and is matrix commutator

$$[u, v] = uv - vu \ , \forall u, v \in \mathfrak{g} \ .$$

All matrix Lie groups are subsets of linear general groups, denoted by $\mathrm{GL}(n, \mathbb{R})$, which is a differentiable curved space comprising of invertible $n \times n$ real matrices, with group multiplication operator defined as matrix multiplication, the inverse operator defined as

matrix inverse, and the identity element being the identity matrix I. The Lie algebra of $\mathrm{GL}(n, \mathbb{R})$ is denoted by $\mathfrak{gl}(n, \mathbb{R})$, which is an open subset of the set of real $n \times n$ matrices, with Lie bracket defined as matrix commutator.

Let us denote the matrix Lie group and matrix Lie algebra by S and $\mathfrak{s}$. Given an element of matrix Lie algebra $U \in \mathfrak{s}$, the exponential map is defined as:

$$\exp(U) = \sum_{i=0}^{\infty} \frac{U^i}{i!} \ ,$$

which is the matrix exponential operator. The inverse exponential map is defined as

$$\log(X) = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{(X-I)^i}{i} \ ,$$

for X in a neighbourhood of identity $I \in S$.

For $X, Y \in S$, and $U \in T_X S$, the geometric structure of the Riemannian manifold can be realized using the following operators

$$L_X Y = XY, \quad (L_X)_* U = XU, \quad R_X Y = YX^{-1}, \quad (R_{X^{-1}})_* U = UX,$$

where L and R denote the left translation and right translation, and $(L_X)_*$ and $(R_{X^{-1}})_*$ represent the tangent mapping of L_X and $R_{X^{-1}}$, respectively. For $X \in S$ and $U, V \in T_X S$, the left invariant metric on S is defined as

$$\langle U, V \rangle_X = \langle (L_{X^{-1}})_* U, (L_{X^{-1}})_* V \rangle_I = \langle X^{-1}U, X^{-1}V \rangle_I = \mathrm{tr}[(X^{-1}U)^\top X^{-1} V] \ .$$

where tr represents the trace of matrix.

Example 2 (**Special Euclidean Space**). Special Euclidean Space $SE(n)$, also called Special Euclidean group, is a subgroup of $\mathrm{GL}(n+1, \mathbb{R})$, which is the result of the product of special orthogonal group $SO(n)$ with $\mathbb{R}^n$ (Duan et al., 2013). Here, $SO(n)$ is the group of rotations in $\mathbb{R}^n$, mathematically expressed as

$$SO(n) = \{ X \in \mathrm{GL}(n, \mathbb{R}) \mid X^\top X = I \text{ and } \det X = 1 \} \ .$$

Notice that $SE(n)$ can be expressed as

$$SE(n) = \left\{ \begin{pmatrix} X & b \\ 0 & 1 \end{pmatrix} \mid X \in SO(n) \ , \ b \in \mathbb{R}^n \right\} \ .$$

Therefore, $SE(n)$ is a curved space whose elements represent a displacement, where X and b stand for the orientation and translation of the rigid body in $\mathbb{R}^n$. The corresponding Lie algebra is represented by $\mathfrak{se}(n)$, which can be expressed as

$$\mathfrak{se}(n) = \left\{ \begin{pmatrix} \Omega & \nu \\ 0 & 0 \end{pmatrix} \mid \Omega^\top = -\Omega \ , \ \nu \in \mathbb{R}^n \right\} .$$

When $n = 3$, $SE(3)$ represents the rigid body motion, consisting of rotations and translations in $\mathbb{R}^3$. For this case, the matrix Ω is expressed as

$$\Omega = \begin{pmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{pmatrix} .$$

Let $\omega = (\omega_x, \omega_y, \omega_z) \in \mathbb{R}^3$ represent the angular velocity of rigid body. The amount of rotation along ω can be obtained by $\|\omega\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm. The exponential map $\mathfrak{se}(3) \to SE(3)$ for $u \in \mathfrak{se}(3)$ is

$$\exp(u) = I_4 + u + \frac{1 - \cos(\theta)}{\theta^2} u^2 + \frac{\theta - \sin(\theta)}{\theta^3} u^3 ,$$

where $\theta^2 = \omega_x^2 + \omega_y^2 + \omega_z^2$. Also, the logarithmic map $SE(3) \to \mathfrak{se}(3)$ for $U \in \mathfrak{se}(3)$ can be given as

$$\log(U) = r_1(r_2 I_4 - r_3 U + r_4 U^2 - r_5 U^3) ,$$

where

$$\begin{aligned} r_1 &= \frac{1}{8} \csc^3\left(\frac{\theta}{2}\right) \sec^3\left(\frac{\theta}{2}\right) , \\ r_2 &= \theta \cos(2\theta) - \sin(\theta) , \\ r_3 &= \theta \cos(\theta) + 2\theta \cos(2\theta) - \sin(\theta) - \sin(2\theta) , \\ r_4 &= 2\theta \cos(\theta) + \theta \cos(2\theta) - \sin(\theta) - \sin(2\theta) , \\ r_5 &= \theta \cos(\theta) - \sin(\theta) , \end{aligned}$$

and $\mathrm{tr}(U) = 2 + 2\cos(\theta)$ for $-\pi < \theta < \pi$.

Let us denote $U, V \in SE(n)$ by

$$U = \begin{pmatrix} X_1 & b_1 \\ 0 & 1 \end{pmatrix} \ , \ V = \begin{pmatrix} X_2 & b_2 \\ 0 & 1 \end{pmatrix} ,$$

then, the geodesic $\delta_{U,V}$ between U and V on $SE(n)$ is given by

$$\delta_{U,V}(t) = \begin{pmatrix} X_1 (X_1^\top X_2)^t & b_1 + (b_2 - b_1)t \\ 0 & 1 \end{pmatrix} .$$

Example 3 (**Symmetric Positive Definite Matrix Space**). The space of symmetric positive definite matrices of dimension $k \times k$

$$SPD(k) = \{A \in \mathbb{R}^{k\times k} \mid A = A^\top, \text{ and } x^\top Ax > 0\} \ ,$$

is another well-known curved space (Arsigny et al., 2007), with manifold dimension $\frac{k(k+1)}{2}$, which plays an important role in applied statistical analysis and inference. Correlation matrix and covariance matrix spaces are examples of SPD spaces. The space is endowed with an affine invariant Riemannian metric. The geodesic distance between $U, V \in SPD(k)$ can be obtained as

$$\delta(U, V) = \| \log(U^{-1/2} V U^{-1/2}) \|_F \ .$$

Let $\mathcal{M}$ denote the open subset of symmetric positive definite matrices, which is a smooth manifold.

The tangent space $T_X\mathcal{M}$ at point $X \in \mathcal{M}$ is the space of tangent vectors u at point $X \in \mathcal{M}$, which can be identified by $k \times k$ symmetric matrices V, as follows

$$u = \{X^{1/2} V X^{1/2} \mid V = V^\top\} \ ,$$

and it admits the following affine invariant Riemannian inner product on manifold $\mathcal{M}$

$$\langle U, V \rangle_X = \text{tr}(X^{-1} V X^{-1} U), \quad V, U \in T_X\mathcal{M} \ .$$

Let E_{ij} be a matrix whose elements are 0 except for $(i, j)_{th}$ entry which equals 1. An orthonormal basis for tangent space $T_X\mathcal{M}$ at point $X \in SPD(k)$ can be given as

$$B_{ij} = \begin{cases} \frac{(E_{ij}+E_{ji})}{\sqrt{2}}, & i < j \\ E_{ii}, & i = j \end{cases} \ ,$$

for $1 \leq i \leq j \leq k$.

The logarithmic map $\mathcal{M} \to T_X\mathcal{M}$ and exponential map $T_X\mathcal{M} \to \mathcal{M}$ at point $X \in SPD(k)$ can be represented as

$$\log(U) = X^{1/2} \log(X^{-1/2} U X^{-1/2}) X^{1/2} \ ,$$

$$\exp(u) = X^{1/2} \exp(X^{-1/2} u X^{-1/2}) X^{1/2} \ .$$

The statistical inference on $SPD(k)$ enables us to approach real-world problems where collected data are in the form of covariance matrices, correlation matrices, etc. A related application is on health monitoring of vehicle sensors, which can be done by detecting possible outliers and / or change-points given a hypothetical distribution for the correlation of sensor readings mounted on a vehicle.

Example 4 (**Stiefel Manifold Space**). The Stiefel manifold $St(k,p)$ is the space of k-frames in $\mathbb{R}^p$, which is a space with elements being k orthonormal vectors in $\mathbb{R}^p$ (James, 1976). This set of $p \times k$ orthonormal matrices is defined as

$$St(k,p) = \{X \in \mathbb{R}^{p\times k} \mid X^\top X = I_k\} \ .$$

The intrinsic dimension of Stiefel manifold is $pk - \frac{1}{2}k(k+1)$. Stiefel manifold is an embeded submanifold of $\mathbb{R}^{p\times k}$. It can be shown that is a compact manifold. To see this, let $\mathcal{S}_{sym}(k)$ represent the set of all symmetric $k\times k$ matrices. Consider the function $F : \mathbb{R}^{p\times k} \to \mathcal{S}_{sym}(k)$ where maps X to $X^\top X - I_k$. It can be seen that $St(k,p) = F^{-1}(0_k)$. The Stiefel manifold is closed because it is the inverse of the the closed set $\{0_k\}$ under continuous mapping F. It is also bounded because the Frobenius norm of $X \in St(k,p)$ equals $\sqrt{k}$. Therefore, it is a compact manifold.

When $k = 1$, Stiefel manifold becomes the manifold of unit sphere $\mathcal{S}^{p-1}$ which is an embeded manifold in $\mathbb{R}^p$. When $k = p$, Stiefel manifold becomes the space of $p \times p$ orthogonal matrices, i.e. $SO(p)$, with dimension $\frac{1}{2}p(p-1)$.

The tangent space of any given element $X_0 \in St(k,p)$ can be obtained as follows. Let $\gamma(t)$ denote a curve in $St(k,p)$, such that $\gamma(0) = X_0$ and $\gamma(t) \in St(k,p)$ for all t. Differentiating both sides of $\gamma^\top(t)\gamma(t) = I_k$ with respect to t and evaluating at $t = 0$ gives

$$\dot\gamma^\top(0)\gamma(0) + \gamma^\top(0)\dot\gamma(0) = 0 \ .$$

Let $Z = \dot\gamma^\top(0)$, we have

$$Z^\top X_0 + X_0^\top Z = 0 \ ,$$

indicating that Z belongs to the tangent space $T_{X_0}St(k,p)$. Also, the normal space at X_0 is

$$N_{X_0}St(k,p) = \{N \in St(k,p) \mid tr(N^\top X_0) = 0\} \ .$$

It can be inferred that for any X_0, any element $Z \in St(k,p)$ satisfies

$$\pi_N(Z) := X_0\Big(\frac{X_0^\top Z + (X_0^\top Z)^\top}{2}\Big) \in N_{X_0}St(k,p) \ .$$

Now given the decomposition

$$Z = \pi_N(Z) + \pi_T(Z) \ ,$$

it can be seen that

$$\pi_T(Z) = X_0\Big(\frac{X_0^\top Z - (X_0^\top Z)^\top}{2}\Big) + (I_p - X_0X_0^\top)Z \ .$$

For the case that $k = 1$, the tangent space of unit sphere $\mathcal{S}^{p-1}$ can be obtained in a similar fashion. It holds that

$$T_{X_0}\mathcal{S}^{p-1} = \{Z \in \mathbb{R}^p \mid X_0^\top Z = 0\} \ .$$

The Riemannian exponential map at $X_0 \in \mathcal{S}^{p-1}$, i.e. $\exp_{X_0} : T_{X_0}\mathcal{M} \to \mathcal{S}^{p-1}$, is

$$\exp(u) = \cos(\|u\|)X_0 + \sin(\|u\|)\frac{u}{\|u\|} \ .$$

The logarithmic map denoted by $\log_{X_0} : \mathcal{S}^{p-1} \to T_{X_0}$ is

$$\log(U) = \frac{\theta}{\sin(\theta)}(U - \cos(\theta)X_0) \ , \quad \theta = \cos^{-1}\left(\langle U, X_0\rangle\right) \ .$$

An important application of Stiefel manifolds is related to statistical inference for computer vision tasks such as estimating the parameters of spatio-temporal dynamical models (Turaga et al., 2008). Such models can be used for understanding the underlying dynamics of physical phenomena observed in nature, such as dynamic textures, and human joint angle trajectories. Models such as auto regressive moving average (ARMA) can be used for such tasks, where its parameters lie on Stiefel manifold (Chakraborty and Vemuri, 2019). Therefore, using a set of instance data, we can estimate the conditional probability density function over the parameter space, which is Stiefel manifold.

Example 5 (**Grassmannian Manifold Space**)**.** The Grassmannian manifold $G(k, p)$ is a manifold space whose elements are k dimensional hyperplanes (subspaces) in $\mathbb{R}^p$ (Kozlov, 1997). Each k dimensional hyperplane can be viewed as a unique $p \times p$ orthogonal projection matrix P with rank $k \leq p$. The intrinsic dimension of $G(k, p)$ is $k(p - k)$.

Let $\mathbb{R}_*^{p\times k}$ denote the space of Steifel manifold, which is the space of $p \times k$ matrices whose columns are independent. It can be shown that $G(k, p) = \mathbb{R}_*^{p\times k}/\mathrm{GL}(k, \mathbb{R})$, where $\mathbb{R}_*^{p\times k}/\mathrm{GL}(k, \mathbb{R})$ is a qoutient set endowed with a qoutient manifold structure. As a result, $G(k, p)$ has a Riemannian manifold structure when $\mathbb{R}_*^{p\times k}$ is endowed with Riemannian metric

$$g_{X_0}(Z_1, Z_2) = \mathrm{tr}\Big((X_0^\top X_0)^{-1} Z_1^\top Z_2\Big) \ ,$$

Also, the elements of tangent space at X_0 satisfy the following

$$T_{X_0}G(k, p) = \{Z \in \mathbb{R}^{p\times k} \mid Z^\top X_0 = 0_k\} \ .$$

For $Z \in T_{X_0}G(k, p)$, consider SVD decomposition of $Z = USV$. Also, let

$$A = X_0 V \cos(S)V^\top + U \sin(S)V^\top \ .$$

The output of exponential map is

$$\exp(Z) = Q \ ,$$

where Q can be obtained using QR decomposition $A = QR$. Also, the logarithmic map can be obtained as follows. Consider the SVD decomposition

$$USV = (Y^\top X_0)^{-1}(Y^\top - Y^\top X X^\top) \ .$$

It can be shown that

$$\log Y = V \arctan(S)\ U^\top \ .$$

Applications of Grassmannian manifolds arises in fields such as shape analysis where projective spaces play an important role. Therefore, estimating probability distribution on shape spaces requires statistical inference on Grassmannian manifolds (Yataka et al., 2023). The other important application is related to dynamic environments and robotic systems, where the navigation in unknown spaces requires algorithms for tracking appropriate subspaces because the system operation cannot be represented by a static model (Bernstein, 2017; Saad-Falcon et al., 2024). This problem is related to Grassmannian manifolds, where the appropriate PCA subspace is identified as a point in Grassmannian manifold.

Example 6 (**Directional and Axis (Projective) Spaces**)**.** A direction in $\mathbb{R}^p$ is a unit vector e, which can be identified as a point on unit sphere $\mathcal{S}^{p-1} = \{e \in \mathbb{R}^p \mid \|e\| = 1\}$ (Mardia and Patrangenaru, 2005). An axes in $\mathbb{R}^p$ can be represented as a pair of directions $\pm e$, which corresponds to a pair of diametrically opposed points on $\mathcal{S}^{p-1}$. Another way for representing an axis is to consider it as an orthogonal projection in $\mathbb{R}^p$ whose rank is 1. Indeed, when $k = 1$, the Grassmannian manifold reduces to real projective space, which is the set of all orthogonal projections of $\mathbb{R}^p$ into one dimensional subspaces. Therefore, given $\pm e$, $e^\top e$ has one eigenvector with rank one (representing the axis $\pm e$), and other eigenvectors equal 0.

Consider two given directions e_1 and e_2 in $\mathbb{R}^2$ and let $\theta \in [0, 2\pi]$ represent the counterclockwise rotation that takes e_1 to e_2, i.e. $\rho_2(\theta) = \big(\cos(\theta), \sin(\theta)\big)^\top$, which can be equivalently written as

$$r_2(e_1, e_2) = \big(e_1^\top e_2, \det(e_1, e_2)\big)^\top \ .$$

The corresponding rotation matrix becomes

$$\begin{pmatrix} e_1^\top e_2 & -\det(e_1, e_2) \\ \det(e_1, e_2) & e_1^\top e_2 \end{pmatrix} \ .$$

For two axes $\pm e_1$ and $\pm e_2$ in $\mathbb{R}^2$, there are two counterclockwise rotations which can take $\pm e_1$ to $\pm e_2$. The rotation angle θ takes e_1 into e_2 and $-e_1$ into $-e_2$. The rotation angle $\theta + \pi$ [mod 2π] takes e_1 into $-e_2$ and $-e_1$ into e_2. Given that $\rho_2(\theta + \pi) = -\rho_2(\theta)$, we can represent the pair of rotation that brings axis $\pm e_1$ to $\pm e_2$ by $\pm\rho_2(\theta)$, or equivalently by the pair $\{\theta, \theta + \pi\}$[mod 2π].

Consider e_1 and e_2 as two direction vectors in $\mathbb{R}^3$. Let us represent the normal vector to the plane determined by e_1 and e_2 by ν. Applying right hand rule rotation takes e_1 into e_2 with corresponding rotation angle $\theta \in [0, 2\pi]$. This rotation can be represented by the angle pair

$$\rho_3(\theta, \nu) = \left(\cos(\theta), \sin(\theta)\nu^\top\right)^\top ,$$

which can be equivalently written as

$$r_3(e_1, e_2) = \left(e_1.e_2, (e_1 \times e_2)^\top\right)^\top ,$$

where $\times$ represents the cross-product operator, which can be represented as an orthogonal matrix

$$d_1 d_1' + (\nu \times d_1)(\nu \times d_2)^\top + \nu\nu^\top , \qquad \nu = \frac{d_1 \times d_2}{\mid d_1 \times d_2 \mid} .$$

For two axes $\pm e_1$ and $\pm e_2$ in $\mathbb{R}^3$, the rotation angle (θ, ν) takes e_1 into e_2 and $-e_1$ into $-e_2$. The rotation angle $(\pi - \theta, -\nu)$ takes e_1 into $-e_2$ and $-e_1$ into e_2. Given that $\rho_3(\pi - \theta, -\nu) = -\rho_3(\theta, \nu)$, the pair of rotations can be represented as $\pm\rho_3(\theta, \nu)$, or equivalently by $\{(\theta, \nu), (\pi - \theta, -\nu)\}$.

Example 7 (**Tree Space**)**.** Tree spaces are among the most important variants of manifold spaces with applications in fields such as evolutionary biology, where the ancestral relationships among a collection of species can be represented by rooted phylogenetic trees (Billera et al., 2001; Barden et al., 2013). Such spaces form non-Riemannian geodesic manifold spaces due to singularity at origin, i.e. the resulting start-shape tree at origin has degenerate interior edges. Data sampled from such spaces are called data objects. Statistical inference in such spaces can become difficult, and combinatorial analysis is needed to find a proper representation for doing statistical inference in such spaces.

A rooted phylogenetic tree is comprised of leaf set $\mathcal{X} = \{0, 1, 2, ..., N\}$, where root vertex has label 0, and other leaf vertices have label $1, 2, ..., N$. Rooted phylogenetic tree is an undirected acyclic graph with $N + 1$ degree 1 vertices (pendant edges) and no vertices with degree 2. The internal edges can vary, and the maximum number of internal edges is $N - 2$. This maximum number of edges is reached when tree has a binary structure, where all internal vertices have degree 3, and the resulting tree is called fully resolved. Any other

tree with fewer edges is called unresolved which will have at-least one vertex with degree 4. A strictly positive weight is assigned to each edge in the tree, which also indicates the edge length.

Let us denote the set of all possible trees with leaf set $\mathcal{X}$, known as tree space, by $\mathcal{T}_N$. Given a tree $x \in \mathcal{T}_N$, the set of edges of x is denoted by $\mathcal{E}_x$, and the weights assigned to edges $e \in \mathcal{E}_x$ is denoted by $\mid e \mid_x$. The weight 0 is assigned to an edge e which is not a member of $\mathcal{E}_x$. We can embed tree space $\mathcal{T}_N$ into $\mathbb{R}^M$, where $M = 2^N - 1$. For a given tree $x \in \mathcal{T}_N$, the tree topology can be identified by the set of possible splits. If we cut any given edge e of tree x, then we end up with two disjoint sub-trees x_e and x_e^c, where x_e is the sub-tree containing the root vertex. As a result, there are $2^N - 1$ ways to split the tree x, and the collection of splits $x_e \mid x_e^c$ identifies the topology of x. The notation $\mathcal{E}_x$ is used to represent the set of splits given that the splits and edges are equivalent. Therefore, we can represent each tree $x \in \mathcal{T}_N$ as vector in $\mathbb{R}^M$ which has at-most $2N - 2$ positive edge weights.

The local structure of $\mathcal{T}_N$ can be revealed by embedding the tree space to a proper Euclidean space. For a given tree x with m internal edges and topology τ, the set of trees with topology τ forms a subset $\mathcal{O}_\tau$ which is isomorphic to $\mathbb{R}^m$, which corresponds to a local Euclidean structure. The boundary of the subset corresponds to trees formed by shrinking a single or multiple edges from a tree in the subset of trees with topology τ. The k-codimension of a fully resolved tree topology τ can be obtained by removing $N - k - 2$ internal edges, for $k = 1, ..., N - 2$.

It can be shown that there are $(2N-3)!!$ possible fully resolved rooted tree topologies in $\mathcal{T}_N$, and each of these topologies correspond to a positive orthant of $\mathbb{R}^{N-2}$ and boundaries of these orthants obtained by shrinking the edges of these $(2N - 3)!!$ fully resolved tree topologies. A tree with no internal edges is called a star tree, and is included in the boundary of every positive orthant $\mathcal{O}_\tau$.

For any two members $x, y \in \mathcal{T}_N$, there exists a continuous path $\Gamma(x, y)$, which is the geodesic path. The length of the path is denoted by $d(x, y)$, which defines a metric on tree space $\mathcal{T}_N$. If x and y belong to the same orthant, the distance will simply be the length of the straight line segment. Otherwise, the distance will be the length of a series of straight line segments traversing orthants within the geodesic path.

Two splits $x_e \mid x_e^c$ and $x_f \mid x_f^c$ are called compatible if one of the four sets $x_e \cap x_f$, $x_e^c \cap x_f$, $x_e \cap x_f^c$, and $x_e^c \cap x_f^c$ is an empty set. Given $x, y \in \mathcal{T}_N$, let the set $\mathcal{C}(x, y)$ contains the set of splits in $\mathcal{E}_x \cup \mathcal{E}_y$ compatible with every split in $\mathcal{E}_x$ and $\mathcal{E}_y$. For given trees $x, y \in \mathcal{T}_N$, the disjoint set of splits

$$A^{(1)}, ..., A^{(i_{xy})} \subset \mathcal{E}_x \quad , \quad B^{(1)}, ..., B^{(i_{xy})} \subset \mathcal{E}_y \ .$$

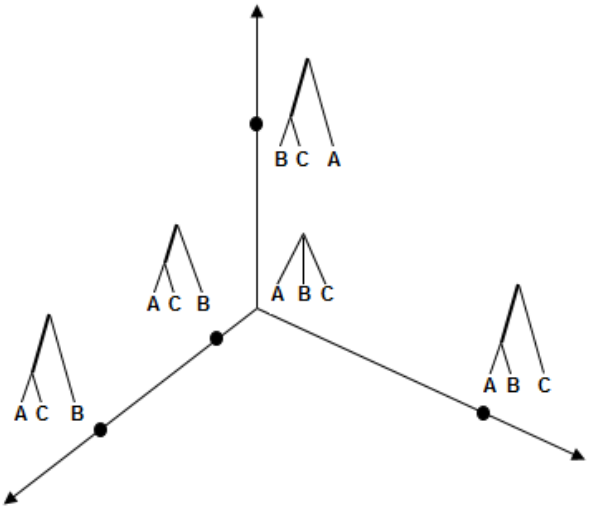


Figure 1.2: Schematic illustration of 3-spider.

characterize the geodesic path $\Gamma(x, y)$, where i_{xy} is an integer which depends on the topologies of x and y, and the split

$$B^{(1)} \cup ... \cup B^{(j)} \cup A^{(j+1)} \cup ... \cup A^{(i_{xy})} \cup \mathcal{C}(x, y) \quad , \quad j = 0, ..., i_{xy} \ ,$$

forms the j^{th} topology visited.

For any subset of edges $E \subset \mathcal{E}_x$, define the norm

$$\| E \|_x = \Big(\sum_{e \in E} \mid e \mid_x^2 \Big)^{\frac{1}{2}} \ .$$

Also, let A_{xy} and B_{xy} be i_{xy}-dimensional vectors with j^{th} elements $\| A_{xy}^{(j)} \|_x$ and $\| B_{xy}^{(j)} \|_x$, respectively, and C_{xy} and D_{xy} be $\mid \mathcal{C}(x, y) \mid$-dimensional vectors containing edges e_X and e_y in the set $\mathcal{C}(x, y)$, respectively. Then the distance of x and y can be obtained as

$$\delta(x, y) = \sqrt{\| A_{xy} + B_{xy} \|^2 + \| C_{xy} - D_{xy} \|^2} \ .$$

3-spider is a special rooted tree with a root vertex and three leaf vertices. Possible representations of 3-spider is shown in Figure 1.2.

1.5 Robust Statistics on Manifolds

In many real-world applications, the collected data can be contaminated with heavy tailed noises, and outliers, and the distribution can be asymmetrical. This will undermine the performance of standard statistical procedures which are usually developed under the assumption that the underlying distribution is normal, symmetric, etc. Robust statistical tools make it possible to reduce or even remove the effect of such undesirable properties, and, therefore, make the resulting statistical inference more accurate. The difficulties of dealing with the mentioned undesirable properties can become even more critical when the collected data lie on manifolds because the physical understanding and interpretation of the basic concepts, such as distance between two data points on such non-Euclidean spaces may not be that clear. Therefore, to develop effective robust statistical tools, it is important to understand the basic properties of the underlying manifold space from which the data is collected.

In this section, we focus on defining important tools and concepts which enable us to do robust statistics on manifolds. The considered concepts include depth functions, developed for statistical analysis on manifolds, and appropriate rank based statistics. These tools will be used to develop robust statistics on manifolds for hypothesis testing, change-point detection, and process monitoring, in the subsequent chapters.

1.5.1 Depth Functions on Manifolds

Depth functions are very powerful tools for exploratory data analysis and can be used to extract important features of the underlying distributions, such as identifying the centrality or outlyingness of data points. Depth functions are used for statistical inference tasks such as developing rank based nonparametric statistics for hypothesis testing, and classification (Chau et al., 2019), among others. A depth function is a nonincreasing function which assigns lower values to outskirt observations and larger values to observations near the center so that the most central observations take the largest value and the depth values decrease as we move from center outwards.

Depth functions were originally proposed to rank multivariate data based on their centrality with respect to the underlying distribution and allowed generalization of the concepts such as order statistics, centrality regions (Zuo and Serfling, 2000), and robust classification and inference (Feige et al., 2018). Tukey's depth (Tukey, 1975), Oja depth (Oja, 1983), and Liu depth (Liu, 1990) are among the most well-known depth functions that have been used for multivariate statistical inference, see for example (Zhang, 2002;

Agostinelli and Romanazzi, 2011; Mosler, 2013). It has been proven that the median estimate based on depth functions is a robust measure of centrality, which makes depth functions a favorable tool for robust nonparametric inference and developing graphical summarization tools.

These depth functions are originally proposed for Euclidean spaces and may not be effective when data lies on manifolds. Given the fact that many of the recent advancements are related to the data belonging to manifold spaces, there is a need for developing depth functions which are appropriate for manifold data analysis. In recent years, developing depth functions for manifold data has gained interest. Examples are depth functions designed for circles and spheres (Liu and Singh, 1992), torus (Fraiman et al., 2019), positive definite matrices (Chau et al., 2019), networks (Zhang et al., 2021), tree structures (Dai et al., 2022), scatter matrices of Euclidean data (Paindaveine and Van Bever, 2018), etc.

Let $\mathcal{M}$ represent the manifold space from which samples are observed, and $P(\mathcal{M})$ represent a family of probability measures on manifold $\mathcal{M}$. Depth of $x \in \mathcal{M}$ is denoted by $D(P, x)$ where $P \in P(\mathcal{M})$ is a distribution on $\mathcal{M}$. A proper depth function is a non-negative function satisfying the following four properties:

- Congruence invariance: A depth function is invariant to congruent transformation $x \to a * x$ where $a \in \mathrm{GL}(n, \mathbb{R})$ is linear general group, the space of invertible $n \times n$ real matrices. This implies that

$$D(P, x) = D(P_a, a * x), \quad \forall x \in \mathcal{M},$$

  where P_a is the distribution of random variable $a * x$, with $x \sim P$. In the Euclidean setup, this is equivalent to the condition of affine invariance of depth functions $D(P, x) = D(P_{A,b}, Ax + b)$ for $x \in \mathbb{R}^n$ and $P_{A,b}$ the distribution of random vector $Ax + b$ where $A \in \mathrm{GL}(n, \mathbb{R})$ and $b \in \mathbb{R}^n$.

- Maximality at center: Let $\mu \in \mathcal{M}$ be the unique central point of distribution P, then, the depth function attains its maximum value at μ, i.e.

$$D(P, \mu) = \sup_{x \in \mathcal{M}} D(P, x).$$

  This is also known as the deepest point with respect to P.

- Monotonicity relative to center: This property requires that as $x \in \mathcal{M}$ moves away from the deepest point μ along the geodesic curve starting from μ, the depth function should behave monotonically non-increasing. Let $\exp_\mu(th)$, $t \geq 0$ be the geodesic from μ with unit tangent vector h, then,

$$D(P, \exp_\mu(t_1 h)) \geq D(P, \exp_\mu(t_2 h)), \quad \forall\ 0 \leq t_1 \leq t_2\ .$$

- Vanishing at infinity: The depth of a point $x \in \mathcal{M}$ should approach 0 as the point x moves away from the center μ.

  For example, consider the case where X is the space of Hermitian positive definite matrices, then the depth function should approach 0 as the point X converges to a singular matrix, i.e. a matrix with an eigenvalue 0 or infinity, i.e.

$$\lim_{M \to \infty} \sup_{\|\log(X)\|_F \geq M} D(P, X) = 0.$$

For our study, we consider several important depth functions, including Zonoid depth, spatial depth, geodesic depth, and half-space depth for nonparametric statistical inference on manifolds, including hypothesis testing, change point detection, and process monitoring.

Zonoid Depth Function

When geodesic convex hulls are well-defined on manifold spaces, we can generalize the definition of the simplicial depth (Liu, 1990). However, when working on manifold spaces, the difficulty is that the computation of simplicial depth can become very challenging, as we possibly need to compute many convex hulls which can become complex as the dimension of manifold increases. Zonoid depth reduces the computational complexity by trimming the convex depth regions (Chau et al., 2019). A subset $\mathcal{K} \subseteq \mathcal{M}$ is geodesically convex if for each pair of points $p_1, p_2 \in \mathcal{K}$, the geodesic segment $[p_1, p_2]$ is contained entirely in $\mathcal{K}$. For $\mathcal{S} \subseteq \mathcal{M}$ the convex hull of $\mathcal{S}$, denoted by $\text{conv}(\mathcal{S})$, is the smallest convex set containing $\mathcal{S}$. The convex set is defined as

$$\text{conv}(\mathcal{S}) = \left\{ p \in \mathcal{S} : p = \exp_p \Big(\int_{\mathcal{S}} \log_p(x) w(x) P(dx) \Big), w : \mathcal{S} \to [0,1], \int_{\mathcal{S}} w(x) P(dx) = 1 \right\}.$$

where P is a Lebesgue measure on metric space $(\mathcal{M}, \delta)$ and w is a measurable function. The Zonoid α-trimmed region is defined as

$$D_\alpha^{\mathcal{M}}(P) = \Big\{ p \in \mathcal{M} \mid p = \exp_p \Big(\int_{\mathcal{M}} \log_p(x) w(x) P(dx) \Big), \\ w : \mathcal{M} \to [0, 1/\alpha], \int_{\mathcal{M}} w(x) P(dx) = 1 \Big\}.$$

The Zoinoid depth function with respect to distribution P can be defined as

$$ZD(P, x) = \sup\{\alpha : x \in D_\alpha^{\mathcal{M}}(P)\}.$$

Spatial Depth Function

General metric spatial depth is the extension of classical spatial depth to data defined on complete and separable metric spaces (Virta, 2023). Let $\mathcal{M}$ be a manifold space equipped with metric δ. Also, let $P \in P(\mathcal{M})$, then the spatial depth of $x \in \mathcal{M}$ with respect to the distribution P is defined as

$$SD(P, x) = 1 - \frac{1}{2} \int_{\mathcal{M}} \int_{\mathcal{M}} \left(\frac{\delta^2(y_1, x) + \delta^2(y_2, x) - \delta^2(y_1, y_2)}{\delta(y_1, x)\delta(y_2, x)} \right) P(dy_1)P(dy_2) \ .$$

It can be seen that SD captures the geometrical structure of $\mathcal{M}$ in a natural way. For $x_1, x_2, x_3 \in \mathcal{M}$, Let $L(x_1, x_2, x_3)$ denote the case where $\delta(x_1, x_3) = \delta(x_1, x_2) + \delta(x_2, x_3)$, which implies that the three points are on a line so that x_2 is in the middle of x_1 and x_3. It can be shown that the maximum value of depth $SD(P, x)$ occurs when $\Pr(L(x_1, x, x_3)) = 1$ for $x_1, x_2 \sim P$, namely the point x is in the middle of other points x_1, x_3 drawn from P. Its minimum value is achieved when $\Pr(L(x_1, x_2, x) \cup L(x_2, x_1, x)) = 1$, i.e. the point x is extreme compared to other points x_1, x_2 drawn from P. Therefore, SD provides a measure of centrality of points, as for a point x to have the maximum depth value, it should almost surely reside between any other two points drawn from P.

Geodesic Depth Function

Geodesic depth is an extension of the arc distance depth (Liu and Singh, 1992) for data defined on spheres. Geodesic depth can be computed for geodesically complete manifolds. Let $P \in P(\mathcal{M})$, then the geodesic depth of $x \in \mathcal{M}$ with respect to the distribution P is defined as

$$GD(P, x) = \exp\Big(- \int_{\mathcal{M}} \delta(x, y) P(dy) \Big) \ ,$$

where δ is the Riemannian distance defined on manifold $\mathcal{M}$. The unique point with maximal depth coincides with the median of the distribution defined as

$$m = \underset{x \in \text{supp}(P)}{\text{argmin}} \int_{\mathcal{M}} \delta(x, y) P(dy) \ .$$

Half-Space Depth Function

To define half-space depth on a manifold, we need to define the notion of half-spaces. For any two points p_1, p_2 in the manifold $\mathcal{M}$, the half-space (Chau et al., 2019) is defined as

$$H_{p_1, p_2} = \Big\{ y \in \mathcal{M} \mid \delta(y, p_1) \leq \delta(y, p_2) \Big\} \ .$$

which is anchored at (p_1, p_2). In words, half-space H_{p_1,p_2} is the set of all points that are not farther from p_1 than from p_2 with respect to Riemannian distance δ. Let us denote the collection of all half-spaces by $\mathcal{H} = \{H_{p_1,p_2} \mid p_1 \neq p_2 \in \mathcal{M}\}$ and the set of half-spaces containing x by $\mathcal{H}_x = \{H_{p_1,p_2} \in \mathcal{H} \mid x \in \mathcal{H}_{p_1,p_2}\}$. The half-space depth at $x \in \mathcal{M}$ with respect to distribution P is defined as

$$\mathrm{HD}(P, x) = \inf_{\substack{p_1,p_2 \in \mathcal{M} \\ \delta(p_1,x) \leq \delta(p_2,x)}} \int_{\mathcal{M}} \mathbb{I}[\delta(p_1, y) \leq \delta(p_2, y)] d(P(y)) \ .$$

For n i.i.d. observations $y_1, ..., y_n$ with empirical distribution P_n, the empirical version of HD is defined as

$$\mathrm{HD}(P_n, x) = \inf_{\substack{p_1,p_2 \in \mathcal{M} \\ \delta(p_1,x) \leq \delta(p_2,x)}} n^{-1} \sum_{i=1}^{n} \mathbb{I}[\delta(p_1, y_i) \leq \delta(p_2, y_i)] \ .$$

An important advantage of distance-based depths such as GD and HD is that the depth function incorporates the geometry of $\mathcal{M}$ via the metric δ, and therefore there is no need for parametrization and linearization onto the tangent space. Moreover, if we have access to the discrete graph of pairwise geodesic distances, the approximated depth function can be easily used. This makes such distance-based functions applicable to many real-world scenarios in which geodesic distances are estimated from point clouds using graph-based methods.

Example 8 (**Depth for SPD(2) Data**)**.** In this example, we evaluate the depth functions using matrix data sampled from manifold $\mathcal{M} = SPD(2)$. We considered the affine invariant Riemannian metric $\delta(x, y) = \| \log(U^{-1/2} V U^{-1/2}) \|_F$. To show the importance of working with a manifold depth function and affine invariant Riemannian metric when working with non-Euclidean data, we also considered Euclidean distance based on vectorized lower triangular elements. All of the considered depth functions, i.e. geodesic depth, Zonoid depth, half-space depth, and spatial depth, are implemented using affine invariant geometry and Euclidean geometry. To generate $SPD(2)$ samples, we considered sample size $n = 75$ and i.i.d. observations from $\mathcal{M} = SPD(2)$ where $X_i = \exp_\mu V_i$ and $V_i = \begin{pmatrix} v_{i,1} & v_{i,2} \\ v_{i,2} & v_{i,3} \end{pmatrix}$, with $v_{i,j}$ independently sampled from $\mathrm{N}(0, \sigma^2)$, $\sigma_1 = 1$, $\sigma_2 = 0.25$, $\sigma_2 = 1.25$, for $i = 1, \ldots, n$. We considered ellipsoids for visualization of $SPD(2)$ matrices. Semi-major and semi-minor axes of each ellipsoid indicate the eigenvectors and the lengths of these vectors equal the square root of their corresponding eigenvalue. Figure 1.3 indicates the $SPD(2)$ matrices visualized using ellipses. For the considered depth functions, the data points with highest

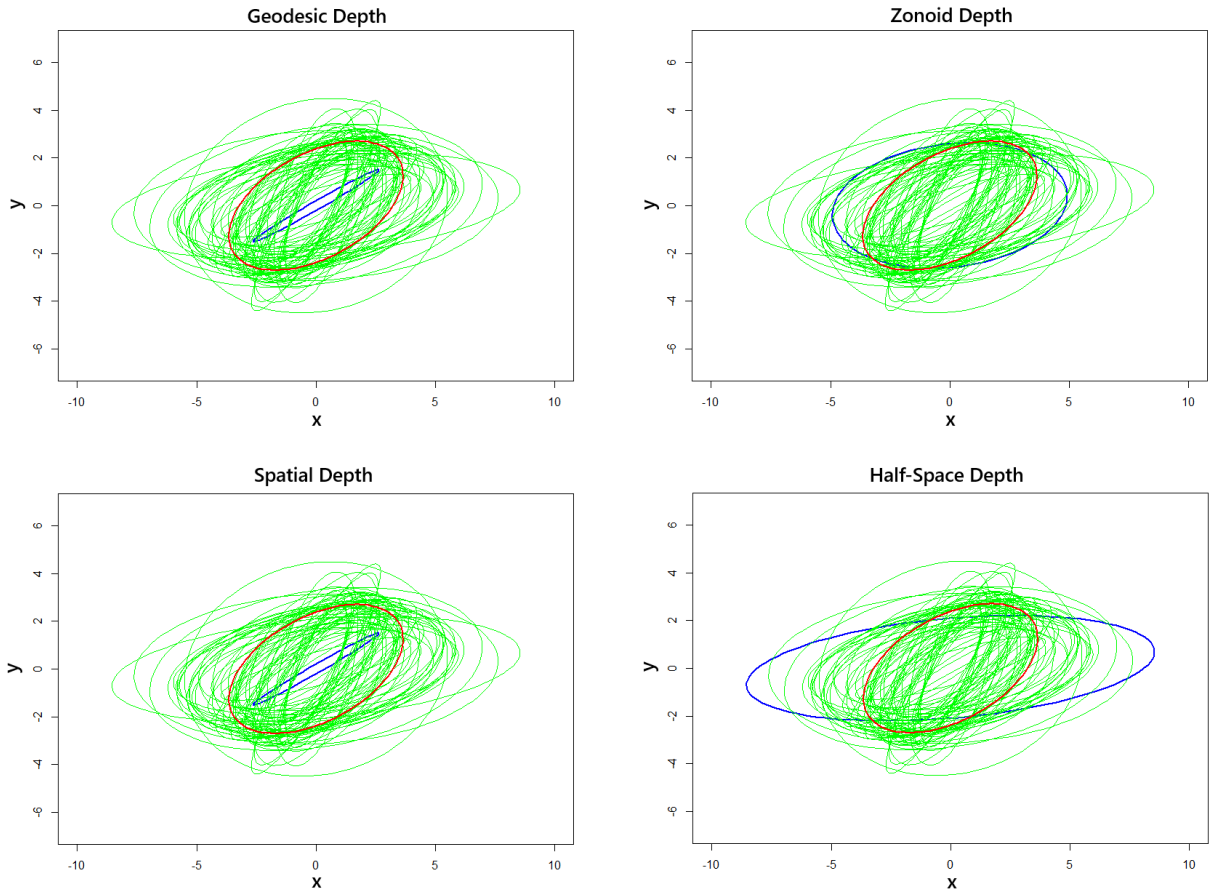


Figure 1.3: Visualization of SPD(2) matrices using ellipses. A red ellipse is the data point with highest depth and a blue ellipse is the data point with the lowest depth. Depths are calculated using affine invariant metric considering manifold geometry.

and lowest depths are shown with red and blue colors. Clearly, all of the depth functions with affine invariant metric find a similar data point with maximum depth, which is the median data point. For the data points with the lowest depth, all of these depth functions find extreme ellipses. For geodesic depth and spatial depth, the blue ellipses are extreme in the sense that their eigenvalues are much smaller than most of the data points. For half-space depth and Zonoid depth, the ellipses are those with larger eigenvalues compared to most of the data points. By comparing the observed depth values, it can be verified that all these ellipses have low depth values and are on the tails of the sample distribution.

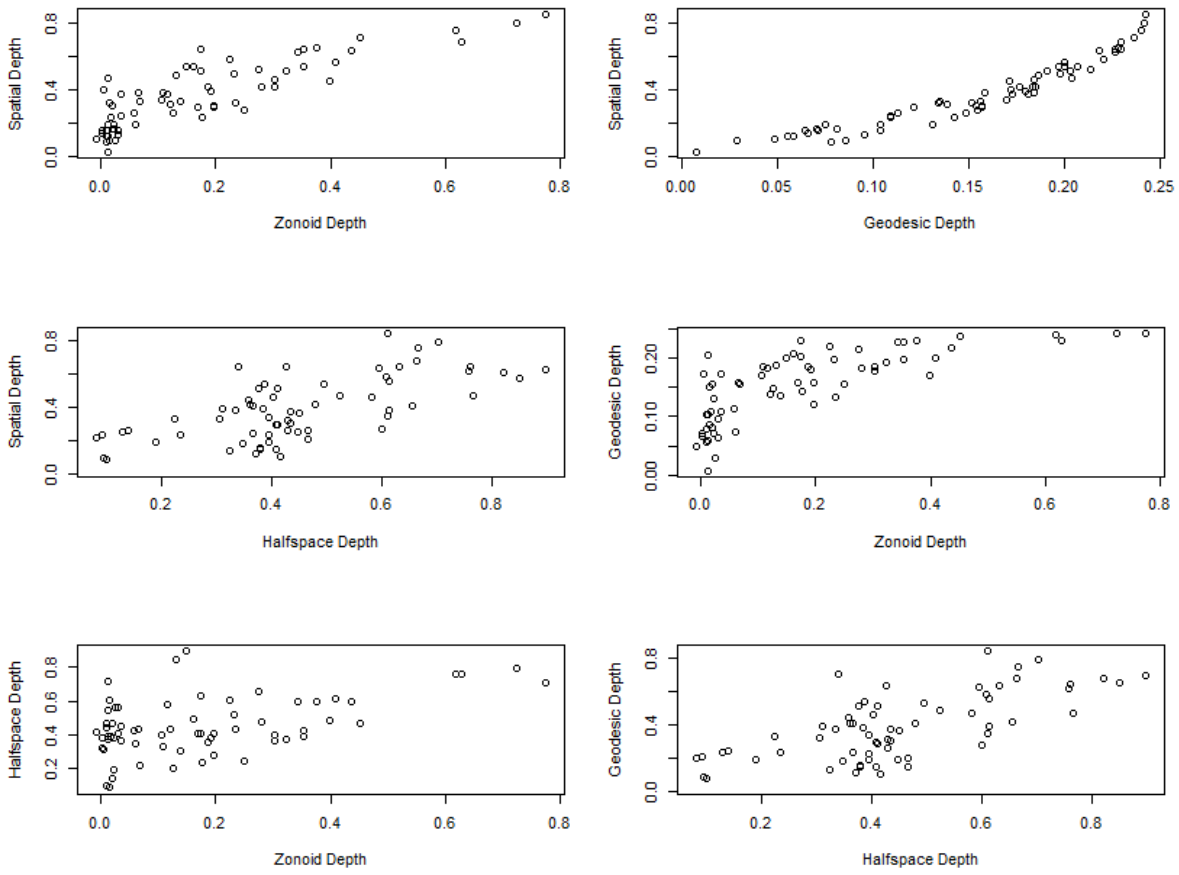


Figure 1.4: Pair-wise comparison of depth values for SPD(2) data using geodesic depth, Zonoid depth, spatial depth, and half-space depth.

Figure 1.4 compares the data depths obtained by geodesic depth, Zonoid depth, spatial depth, and half-space depth. As can be inferred, the depth values are positively correlated and follow a similar pattern. This can also be realized from Figure 1.3 where all these depth functions find the same median point. To illustrate the effectiveness of the depth functions implemented by affine invariant geometry, we also depict the corresponding depths for the same data using Euclidean depth functions. Figure 1.5 depicts the $SPD(2)$ matrices visualized using ellipses with highest and lowest data points shown with red and blue colors.

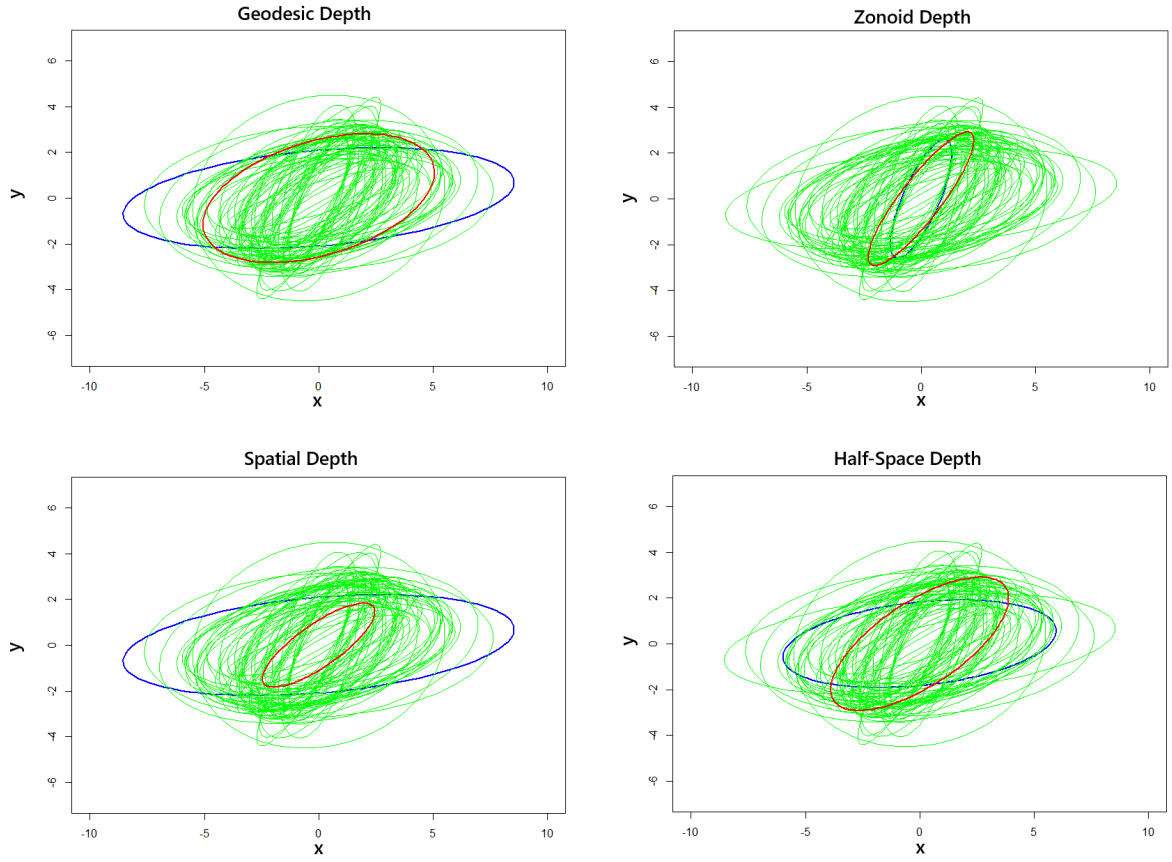


Figure 1.5: Visualization of SPD(2) matrices using ellipses. A red ellipse is the point with the highest depth and a blue ellipse is the point with the lowest depth. Depths are calculated using Euclidean metric.

It can be realized that there is no compromise between these deepest points. The Zonoid and half-space depths cannot yield appropriate data, as the points with lowest depth value do not fall on the tail of the sample distribution. The main reason behind this deficiency is due to the swelling effect, namely the determinant of Euclidean average can become strictly greater than those of the individual matrices (Arsigny et al., 2007). This undesirable effect cannot be observed when calculating the average, i.e. Fréchet mean, using affine invariant metric. Similar behaviour can be observed for data obtained from other types of manifold spaces. This makes the affine invariant geometry a preferable choice for calculating depth functions when working with data which have non-Euclidean nature.

Example 9 (**Depth for $\mathcal{T}_3$ Data**). In this example, the goal is to analyze the performance of the depth functions for object data sampled from the space of trees with three leaves $\mathcal{T}_3$, also known as 3-spider space, which includes 1-dimensional Euclidean rays glued together at the origin. We consider both intrinsic and extrinsic analysis. For intrinsic analysis, the geodesic metric is used to calculate the distances between two trees, which is the length of the geodesic path from one tree to another tree in the manifold. Figure 1.6 indicates an illustrative example of a geodesic path from one 3-spider tree to another 3-spider tree. We can represent a 3-spider tree with $[0, \infty) \times \{1, 2, 3/\sim\}$. Note that $(0, j)$, $j = 1, 2, 3$ is the same point identified through the equivalent relationship $\sim$. For two points $x_1 = (r_{x_1}, j_{x_1})$ and $x_2 = (r_{x_2}, j_{x_2})$ on $\mathcal{T}_3$, the distance can be calculated as follows

$$\delta(x_1, x_2) = \begin{cases} \mid r_x - r_y \mid & j_{x_1} = j_{x_2} \\ r_x + r_y, & j_{x_1} \neq j_{x_2} \end{cases},$$

which is the length of geodesic path to traverse along x_1 to x_2. To generate samples from $\mathcal{T}_3$, we considered a sample size of $n = 75$, with random variables $Y_i \sim \mathrm{N}(0.5, 1)$ for the locations and $Z_i \sim$ Bernoulli(0.5) for branches, which gives the following i.i.d. observations

$$X_i = \begin{cases} (Y_i, 2), & Y_i \geq 0 \\ (-Y_i, 1), & Y_i < 0 \text{ and } Z_i = 0 \\ (-Y_i, 3), & Y_i < 0 \text{ and } Z_i = 1 \end{cases}.$$

Figure 1.7 illustrates the 3-spider tree structures with each axis representing different tree topologies and location on axis representing the length of the interior edge. The deepest point is shown with a red circle and the point with the lowest depth is shown with a blue circle. In this representation $j = 1, 2, 3$ indicate trees on z, x, and y axis, respectively.

As can be inferred, the deepest tree calculated by all of the depth functions are close to each other and are located on $j = 2$. The point with lowest depth are the same for geodesic and spatial depth functions and outperform those calculated by Zonoid and half-space depth functions.

In general, it can be concluded that all of these depth functions show a comparable performance. Apparently, the region which include the deepest point is a denser region with a higher number of circles, suggesting that it is close to the center of the distribution from which 3-spiders are sampled. Moreover, the points with lowest depth are located on tail of the distribution and are extreme compared to other points, i.e. green circles, shown in the plot.

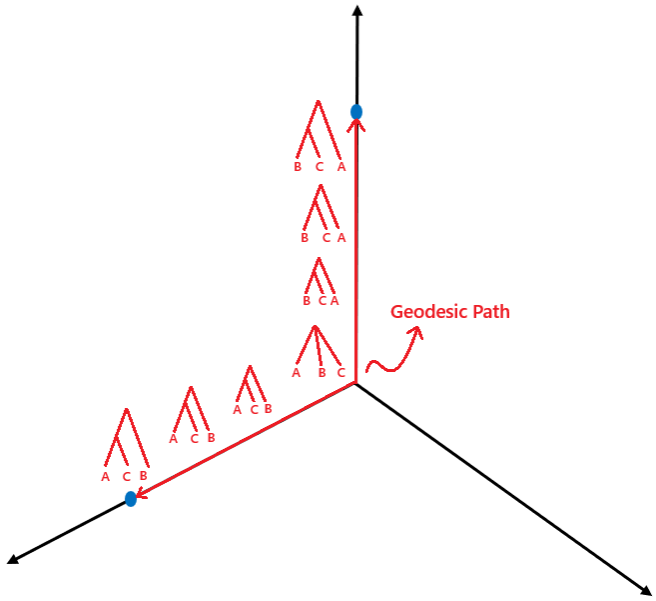


Figure 1.6: Geodesic path between two 3-spiders.

It is apparent that geodesic and spatial depth outperform Zonoid and half-space depth with respect to the calculated extreme point, as the point with lowest depth suggested by these methods are far away from the center of the distribution with respect to both axis and length of the interior edge on this axis.
Figure 1.8 indicates the correlation between each pair of depth functions. All of these plots suggest that there is a relatively strong positive correlation between the depths calculated by the depth functions.
Figure 1.9 presents a complete set of 56 3-spider structures along with the deepest 3-spider shown inside the red box. As it can be seen, the calculated median 3-spider is in the central region of the distribution with respect to both axis and the length of the interior edge on the axis, demonstrating the efficacy of the intrinsic depth functions for analyzing object data sampled from $\mathcal{T}_3$.
For the extrinsic analysis, we embed the 3-spider $[0, \infty) \times \{1, 2, 3/\sim\}$ into $\mathbb{R}^2$ using the map $h(r, j) = (r\cos(2(j-1)/3), r\sin(2(j-1)/3)$. Depth functions are then applied to the embedded data in $\mathbb{R}^2$.
Figure 1.10 depicts the depths calculated by extrinsic depth functions. It can be observed that the calculated medians are comparable with those obtained using intrinsic analysis.

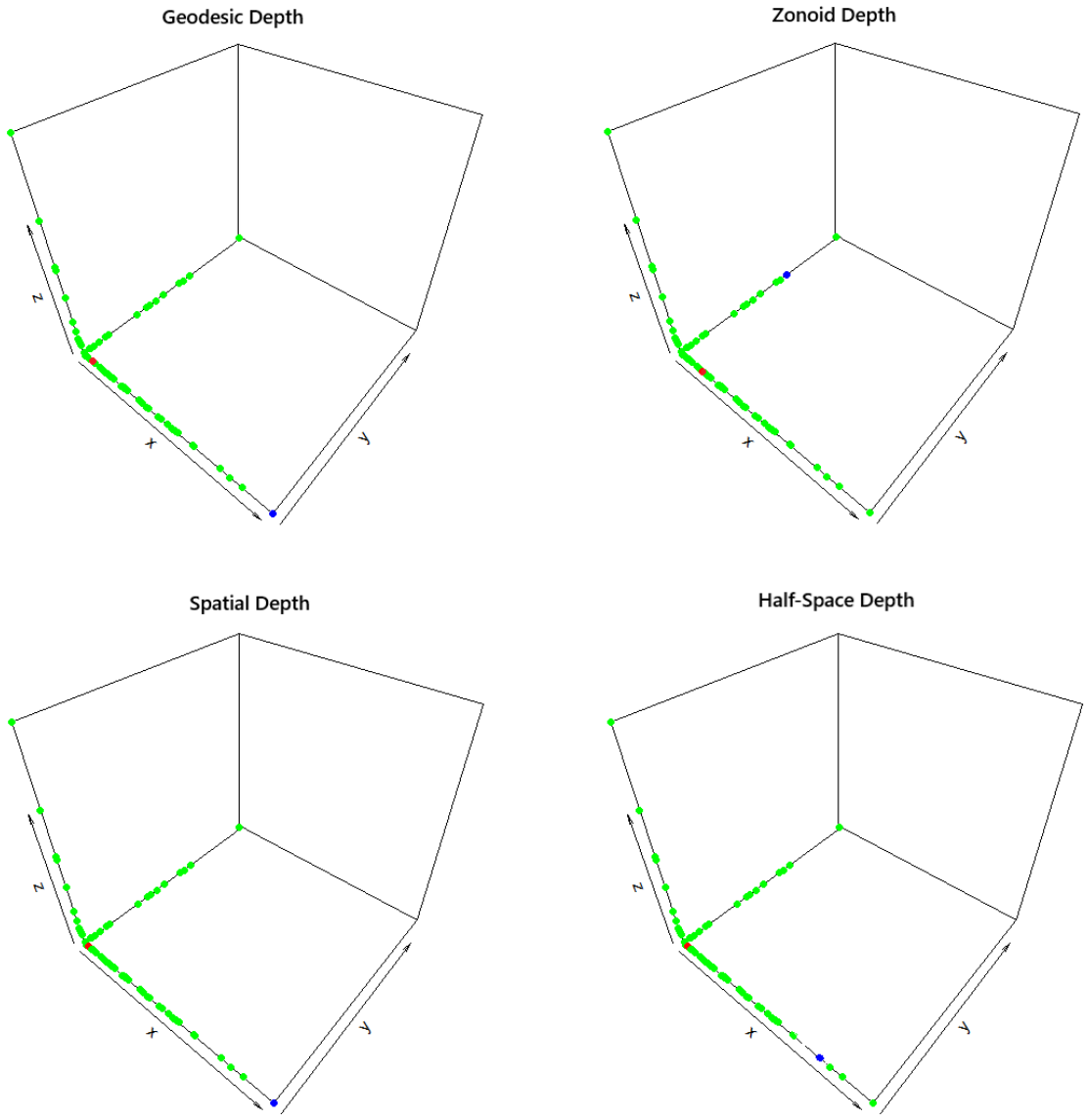


Figure 1.7: Visualization of 3-spider data. Depths are calculated using intrinsic embedding. Each axis represents different tree topology and the location of data on axis represents the length of the interior edge. Red circles are the 3-spider with highest depth and blue circles are the 3-spiders with lowest depth.

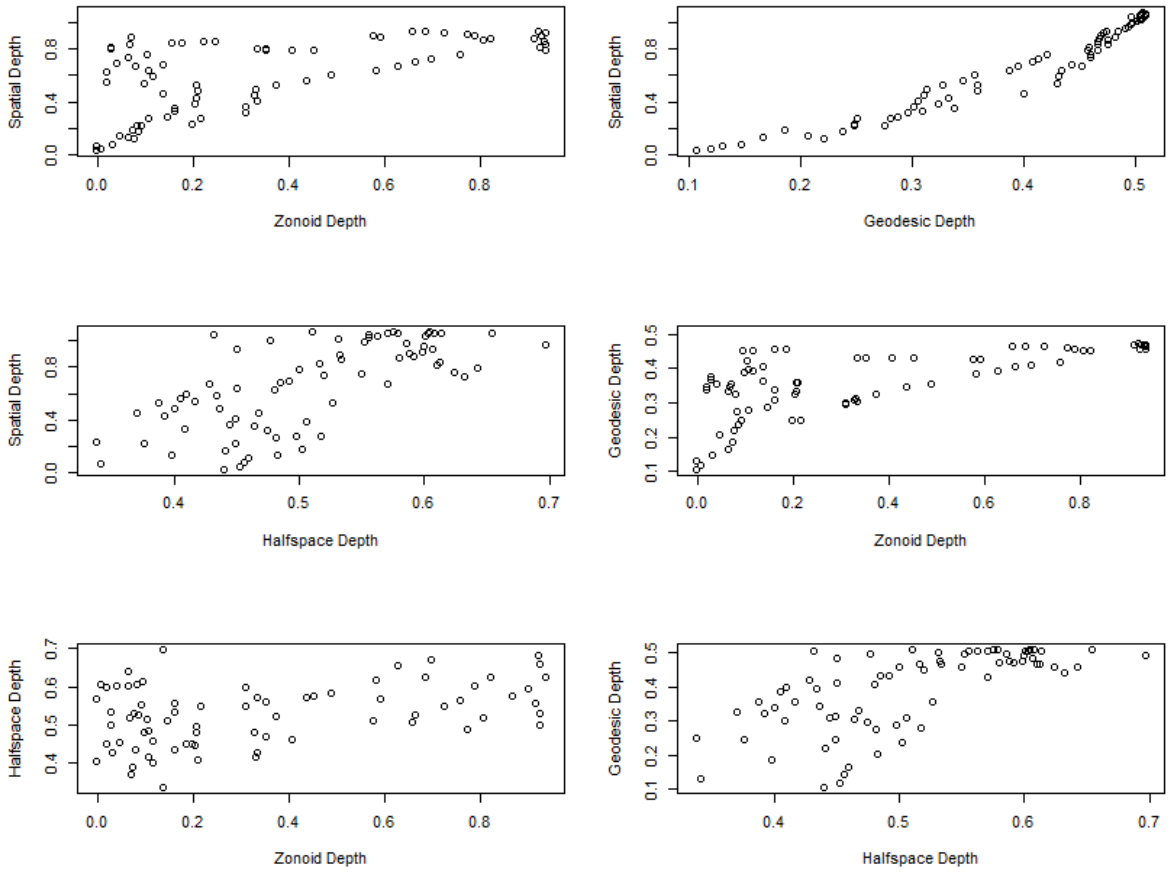


Figure 1.8: Pair-wise comparison of depth values for 3-spider data using geodesic depth, Zonoid depth, spatial depth, and half-space depth.

Comparing Figure 1.7 to Figure 1.10, it is observed that for most of these cases, the deepest points calculated using intrinsic geometry outperform those obtained using extrinsic embedding method. This suggests that intrinsic analysis can provide a better insight due to the fact that it uses properties inherent to the manifold itself, i.e. geodesic, curvature, and distances calculated using the manifold's structure. Note that however the intrinsic analysis can become computationally challenging and less intuitive, especially when dealing with complex manifolds. Therefore, the optimal choice of the manifold analysis approach depends on the geometrical properties as well as the complexity of the manifold itself.



Figure 1.9: The complete set of 60 3-spider tree structures with the deepest structure shown inside the red box.

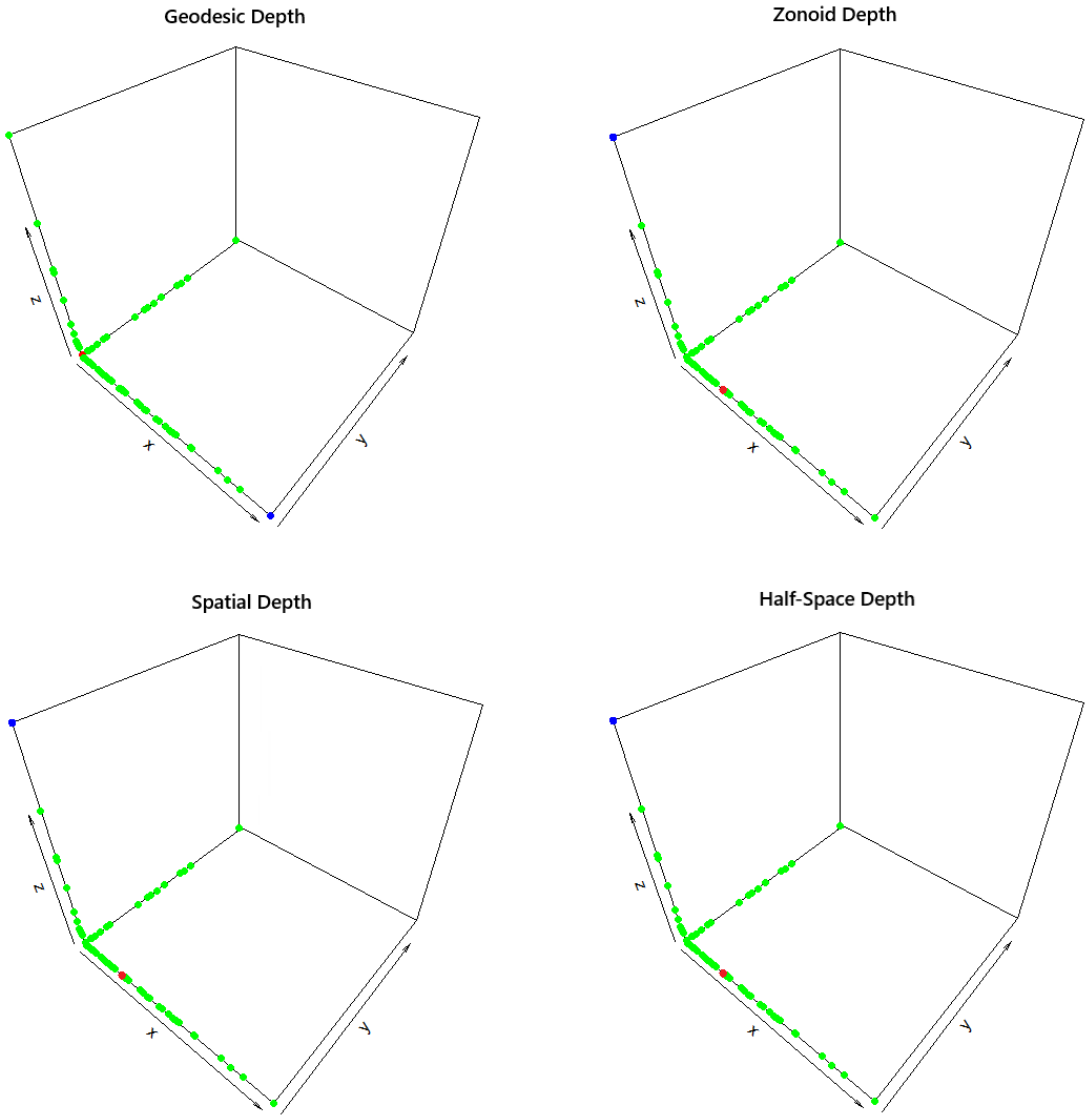


Figure 1.10: Visualization of 3-spider data. Depths are calculated using extrinsic embedding. Each axis represents different tree topology and the location of data on axis represents the length of the interior edge. Red circles are the 3-spider with highest depth and blue circles are the 3-spiders with lowest depth.

Example 10 (**Depth for Non-convex Distributions**). Depth functions in Euclidean space $\mathbb{R}^p$ produce convex depth regions, which is undesirable for data sampled from spaces with non-convex shapes. The goal of this example is to show how the proposed depth functions can be used to provide a meaningful measures of depth in such scenarios. For this simulation, we consider a bivariate sample of size $n = 1000$ generated as $x_i = 2\left(\cos(\theta_i), \sin(\theta_i)\right) + 0.1^{1/2}\epsilon_i$, where θ_i is sampled i.i.d. from Uniform$[0, 2\pi]$ and ϵ_i are sampled i.i.d. from bivariate standard normal distribution.
Figure 1.11(a) indicates the samples from the distribution together with the density color map. It can be realized that the samples resemble a circle around the origin in the plane. We examine the depth functions with standard Euclidean metric for this non-convex dataset. The plane is divided into 1600 equal size grids and the depth of these grid points are computed using GD and SD with Euclidean metric.
Figure 1.11(b) and Figure 1.11(c) indicate the color map contour plot of GD and SD depths, respectively. It can be seen that both depths fail to capture the natural shape of the sample, assigning highest depth to grid points around the origin where no sample points lie. For GD, the grid points x in central region will take small values of $\mathbb{E}\left[\|x - z\|\right]$ with respect to measure induced by the Euclidean structure, and therefore their depth values become large. For SD, it can be verified that the length of the average of unit length vectors pointing towards the sample points become small for grid points in the central region, resulting in high depth values. As grid points get farther away from the center, both GD and SD become smaller. To solve this deficiency, we combine GD and SD depths with a well-known manifold learning method known as ISOMAP (Tenenbaum et al., 2000). ISOMAP uses local Euclidean distance information and discards the longer Euclidean distances to capture the curve structure in the data. To do so, we only retain k smallest distances from a given point, where k is a tuning parameter. The k smallest distances are used to construct a weighted undirected graph $\mathcal{G}$ on the sample points, as shown in Figure 1.11(d). A grid point x can be included in the graph using its k nearest neighbour points (kNN), which are used to calculate the distance between x and sample points on the graph. For small values of k, this distance approximates the geodesic distance with respect to the underlying manifold structure. We use $k = 10$ for the analysis.
Figure 1.11(e) and Figure 1.11(f) indicate the color map contour plot of GD and SD depths based on ISOMAP, respectively. An important observation is that the performance of both depth methods is significantly improved, and they approximate the underlying distribution of the non-convex sample data. For SD with kNN, grid points x in the central region are equally likely to have any point as their neighbour, as the the origin is roughly equally far away from all sample points. This makes kNN incapable of identifying the local structure of the manifold data for grid points around the origin, so that the depth is calculated with respect to the euclidean structure, making the depth large for central region, with

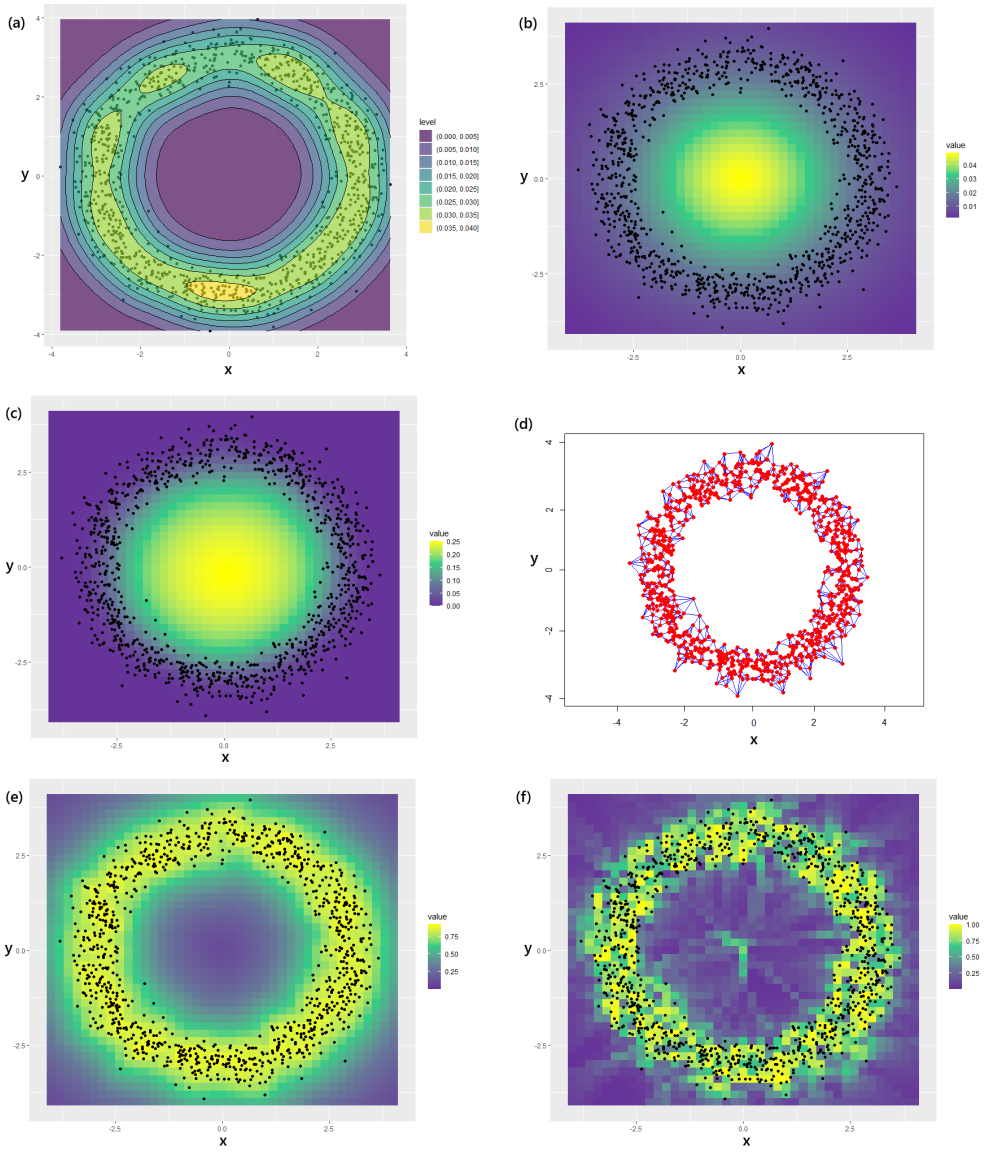


Figure 1.11: Depth color map for circular data for geodesic depth and spatial depth with Euclidean metric and ISOMAP, (a) samples from the distribution together with the density color map, (b) color map contour plot of GD, (c) color map contour plot of SD, (d) weighted undirected graph $\mathcal{G}$ on the sample points using k smallest distances, (e) color map contour plot of GD based on ISOMAP, (f) color map contour plot of SD based on ISOMAP.

respect to Euclidean geodesic path. In this scenario, for SD, $\Pr(L(x_1, x, x_3))$ becomes large for $x_1, x_2 \sim P$ and as a result, the depth become larger. As grids move towards an orthant of the circle, it is more likely that the k nearest neighbours are chosen from the sample points on the part of the circle on the corresponding orthant, so that the point in the middle lie on the tail with respect to the structure induced by the selected points. Hence, $\Pr(L(x_1, x_2, x) \cup L(x_2, x_1, x)) = 1$ becomes larger, which implies smaller depth. As the points on the grid get closer to the sample points, $\Pr(L(x_1, x, x_3))$ becomes large with respect to the manifold structure induced by the k nearest neighbours and therefore, depth gets larger values. It can be inferred from the plot that GD is less sensitive to this phenomenon for an optimal choice of k nearest neighbours, and better approximates the underlying distribution of the data, as well as the structure of the manifold. In general, it can be concluded that using kNN distance results in a much better performance compared to Euclidean metric for both SD and GD methods.

1.5.2 Ranking Based on Depths

Ranking based on data depth functions defined in the previous section can be used to develop robust nonparametric inference on manifold spaces. Ranking of depth measures can be done by calculating the depth values of data points based on distribution $P \in P(\mathcal{M})$, e.g. (Chenouri et al., 2020). Consider the data sample $x_1, ..., x_n$ from manifold space $\mathcal{M}$. We consider a center outward ranking of data points based on depth functions, i.e. for a data point x_i, its rank is

$$R_i = \#\big\{x_j \mid D(P, x_j) \leq D(P, x_i), j = 1, ..., n\big\}.$$

Notice that by definition, data points closer to the center with respect to distribution P have higher ranks, and points farther away from the center have lower ranks. The advantages of using rank-based test statistics are as follows

- Test statistics based on rank of the depth functions are distribution free compared to those developed based on depth of the data. Such a salient feature makes rank-based statistics more desirable for inference on manifold spaces as data cleaning and outlier removal can become quite intricate when data belong to manifold spaces, and consequently, distribution dependent, or non-robust statistics can lose their power.
- Test statistics based on ranks of depth functions are inherently robust to heavy-tailed and skewed distributions.

- As long as the geometrical properties of manifold space are properly identified, the empirical significance level based on rank-based test statistics will not depend on the dimensionality of the dataset. Furthermore, the finite sample null distribution for test statistics can be derived using permutation approaches, which can then be used for small sample sizes.

We will discuss the details for the development of permutation-based and asymptotic distributions for rank-based statistics for various inference problems, including hypothesis testing, change-point analysis, and process monitoring. In the later chapters, we will establish results connecting the geometrical properties of manifolds such as intrinsic dimensionality to the convergence rates.

Example 11 (**von Mises-Fisher Distribution on Circle**)**.** Consider the von Mises-Fisher distribution on circle $\mathcal{S}^1$ (also known as circular normal distribution). The probability distribution for 2-dimensional unit vector $\mathbf{x}$ is

$$f(\mathbf{x}; \boldsymbol{\mu}, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\left(\kappa \boldsymbol{\mu}^\top \mathbf{x}\right) \ ,$$

where $\kappa \geq 0$, $\|\boldsymbol{\mu}\| = 1$ and I_0 is the modified Bessel function of first kind at order 0. The support of the distribution is $\{\mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x}\| = 1\}$. Figure 1.12 depicts the density of this distribution in angular coordinate on $\mathcal{S}^1$. Suppose that we use the spatial depth (SD) on $\mathcal{S}^1$ to calculate the depth of data sampled from von Mises-Fisher distribution. The depth function can be formulated as

$$SD(P, x) = 1 - \frac{1}{2}\mathbb{E}_P \left[\frac{\delta^2(z_1, x) + \delta^2(z_2, x) - \delta^2(z_1, z_2)}{\delta(z_1, x)\delta(z_2, x)} \right] \ .$$

where $x \in \mathcal{S}^1$, $z_1, z_2 \sim P$, and δ is an appropriate distance on $\mathcal{S}^1$.
We can map the data points from the manifold $\mathcal{S}^1$ to tangent space at center ω, which is a line, as shown in Figure 1.12. We can use the Euclidean version of SD for the analysis on this tangent space. It can be verified that the data on the tangent space follow Normal distribution centered at ω. We translate the points on this tangent space such that the center of the data becomes 0.
Consider the sample points shown in Figure 1.13(a), sampled from distributions F_1 and F_2. As seen, the points sampled from distribution F_2 are more dispersed compared to those sampled from F_1, and both of the distributions have the same median μ. In this example, the performance of SD for ranking these points is analyzed, and it is shown that the outcomes of SD follow a pattern similar to that of Ansari-Bradly ranking procedure

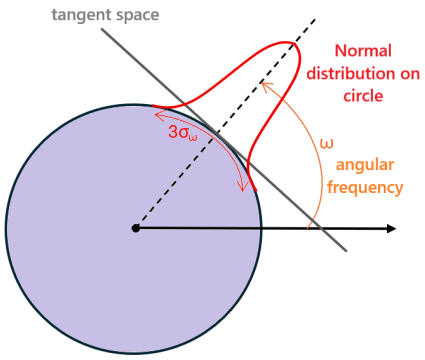


Figure 1.12: Circular normal distribution centered at angular frequency ω.

(Ansari and Bradley, 1960). The SD depth for a point $x \in \mathbb{R}$, with respect to random variable z sampled from combined distribution $F_* = \theta F_1 + (1-\theta)F_2$ is given by

$$SD(F_*, x) = \frac{1}{1 + E_{F_*}|z - x|} \ .$$

Then, the point corresponding to the maximum SD can be obtained by solving the optimization problem $\arg\min_x E_{F_*}|z-x|$, which gives μ, the median of F_*. For any two points $x_1 < x_2 < \mu$, we can write

$$E_{F_*}|z - x_i| = \int_{-\infty}^{x_i} F_*(z)dz + \int_{x_i}^{\mu} \big(1 - F_*(z)\big)dz + \int_{\mu}^{\infty} \big(1 - F_*(z)\big)dz, \quad i = 1, 2 \ .$$

It follows that

$$\begin{aligned} E_{F_*}|z - x_1| - E_{F_*}|z - x_2| &= \int_{-\infty}^{x_1} F_*(z)dz + \int_{x_1}^{\mu} \big(1 - F_*(z)\big)dz \\ &\quad - \int_{-\infty}^{x_2} F_*(z)dz - \int_{x_2}^{\mu} \big(1 - F_*(z)\big)dz \\ &= \int_{x_1}^{x_2} \big(1 - F_*(z)\big)dz - \int_{x_1}^{x_2} F_*(z)dz = \int_{x_1}^{x_2} \big(1 - 2F_*(z)\big)dz \ . \end{aligned}$$

Due to the fact that for $x_1 < x_2 < \mu$, $F_*(z) < 0.5$, it can be shown that

$$E_{F_*}|z - x_1| - E_{F_*}|z - x_2| = \int_{x_1}^{x_2} \big(1 - 2F_*(z)\big)dz > 0 \ ,$$

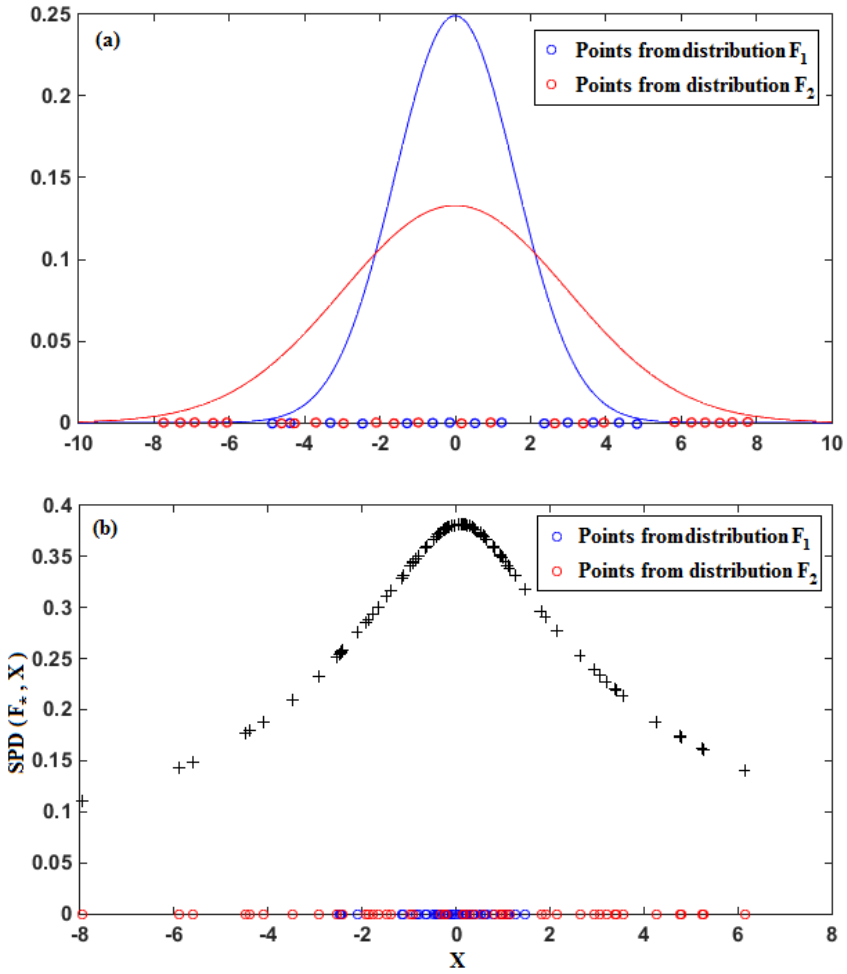


Figure 1.13: (a) Distribution of points sampled from F_1 and F_2, (b) SD values of samples drawn from $\mathrm{N}(0, 2)$, and $\mathrm{N}(0, 4.25)$.

which means $SD(F_*, x_1) < SD(F_*, x_2)$. On the other hand, for $\mu < x_3 < x_4$, it can be shown that

$$E_{F_*}|z - x_i| = \int_{-\infty}^{\mu} F_*(z)dz + \int_{\mu}^{x_i} F_*(z)dz + \int_{x_i}^{\infty} \big(1 - F_*(z)\big)dz, \quad i = 3, 4 \ .$$

This results in

$$E_{F_*}|z - x_4| - E_{F_*}|z - x_3| = \int_{\mu}^{x_4} F_*(z)dz + \int_{x_4}^{\infty} \big(1 - F_*(z)\big)dz$$

$$
\begin{aligned}
&- \int_{\mu}^{x_3} F_*(z)dz + \int_{x_3}^{\infty} \big(1 - F_*(z)\big)dz \\
&= \int_{x_3}^{x_4} F_*(z)dz - \int_{x_3}^{x_4} \big(1 - F_*(z)\big)dz = \int_{x_1}^{x_2} \big(2F_*(z) - 1\big)dz \ .
\end{aligned}
$$

When $\mu < x_3 < x_4$, $F_*(z) > 0.5$, and by following a similar argument, we get $SD(F_*, x_4) < SD(F_*, x_3)$. The above results show that SPD is maximum in the median, and the more we approach to the tails, SD becomes smaller. Therefore, for an ordered set of points $x_1, x_2, ..., x_{\frac{k+n}{2}}, ..., x_{k+n-1}, x_{k+n}$ (where $k+n$ is even), it holds that

$$
\begin{cases}
SD(F_*, x_1) \leq SD(F_*, x_2) \leq \cdots \leq SD(F_*, x_{\frac{k+n}{2}}) \\
\\
SD(F_*, x_{\frac{k+n}{2}}) \geq \cdots \geq SD(F_*, x_{k+n-1}) \geq SD(F_*, x_{k+n})
\end{cases} ,
$$

which follows a pattern similar to the ranking process of Ansari and Bradly. Note that the same holds when $k+n$ is odd, and the only difference is that the index of median which becomes $\frac{k+n+1}{2}$. An example of the shape of SD profile for ordered sample points is given in Figure 1.13(b). For this simulation, $k = n = 50$, $N = n + k = 100$, $F_1 = \mathrm{N}\big(0, 2\big)$, and $F_2 = \mathrm{N}\big(0, 4.25\big)$.

In the next chapters, we will see that this idea can be used for developing nonparametric rank-based hypothesis tests, change point detection statistics, and process monitoring control charts.

Example 12. Recall Example 8. For the SPD(2) data, the ellipses with highest depth, as well as two ellipses with lowest depth are shown in Figure 1.14. Also, in the plot, the eigenvalues of the samples is shown where the eigenvalues corresponding to the selected ellipses are depicted with the same color. It is apparent that the extreme ellipses fall on the opposite tails of the distribution. Figure 1.15 shows the corresponding ranks of the data. It can be verified that the points on the two tails of the distribution have small rank and the median has the largest rank. This observation is in agreement with the results obtained in Example 11.

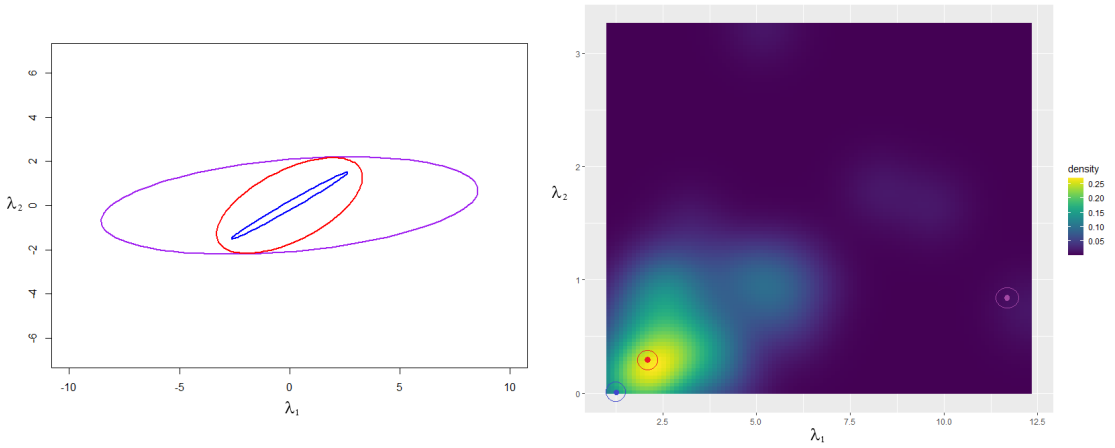


Figure 1.14: Ellipse with highest depth (red color) and ellipses with lowest depth (blue and purple color). Density colormap for the eigenvalues of the SPD(2) dataset. The points with highest and lowest depths are shown inside circles in the colormap.

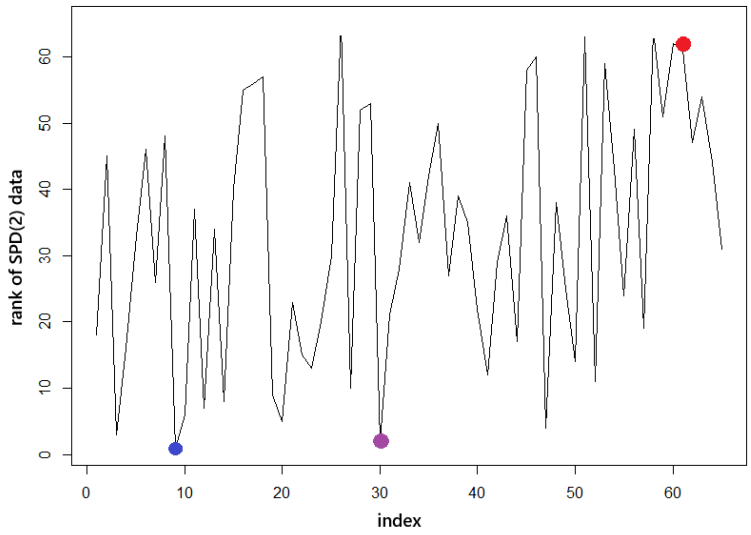


Figure 1.15: Rank of SPD(2) data, data point with highest rank is shown with red circle and data points with lowest ranks are shown with blue and purple circles.

1.6 Literature Review

In this section we review the research literature related to statistical analysis on manifold spaces. Statistical analysis on manifold spaces has gained significant interest in recent years from both theoretical and practical view points, thanks to the emergence of high speed computational systems and the methodological and theoretical developments in related research fields such as functional spaces and general metric spaces. The current literature review encompasses a wide range of research domains demonstrating the potential of manifold statistical analysis methodologies for tackling complicated real-world problems.

Statistical analysis on manifold spaces has gained interest in statistical machine learning over the past decades with main research objectives being dimensionality reduction, local embedding, finding the intrinsic dimension of a manifold, etc. In Ham et al. (2004), the authors proposed a kernel-based dimensionality reduction algorithm which can be viewed as a kernel principal component analysis on a Gram matrix. It was demonstrated how Isomap, graph Laplacian eigenmap, and locally linear embedding can be translated into this KPCA frame and their differences and similarities were investigated. In Coifman and Lafon (2006), the authors investigated the applicability of diffusion processes for extracting useful geometrical features from a dataset. To do so, the eigenfunctions of the Markov matrix is used to construct diffusion coordinates called diffusion maps which represent the geometrical structure of the dataset. It was proved that the methodology is useful for dimensionality reduction and finding the intrinsic dimension, spectral graph theory and data parametrization. In particular, the authors indicated that when the dataset represents a manifold, it is possible to unveil the geometry of the manifold by approximating the Laplace-Beltrami operator. In Miranda et al. (2017), the authors proposed a nonlinear dimensionality reduction algorithm for efficient information extraction from manifold data. The strategy proposed relies on local embedding on nonlinear spaces to linear spaces for understanding the geometrical properties of images. The methodology is applied to facial image feature extraction, clustering in Riemannian manifold spaces, and understanding the relationship of manifold learning with deep learning methods. In Pennec (2006), the authors investigated invariance properties on manifold spaces and proposed a Riemannian metric to develop statistical tools for transformation groups. The authors proposed a notion of Normal law based on maximization of an entropy using known mean and covariance of the distribution. It was demonstrated that the proposed law is applicable to general metric spaces ranging from vector space to sphere space to get normal density. Also, the resulting family of pdfs comprises the whole range of distributions including uniform distribution to point mass. A test for outlier detection based on χ^2 law for general metric spaces is also proposed. In Jayasumana et al. (2015), the authors proposed a Gaussian

radial basis function to embed points on a nonlinear manifold to a higher dimensional linear space. It was demonstrated that methodologies originally developed for linear spaces, such as support vector machine, discriminant analysis and principal component analysis can be generalized to manifold spaces using the proposed embedding. The applicability of the strategy is investigated for computer vision applications on manifolds such as positive definite matrices and Grassmannian manifold, e.g. pedestrian detection, visual object categorization, texture recognition, segmentation, action recognition, pose categorization, etc. In Weinberger et al. (2004), the authors developed an algorithm to learn a kernel matrix for high dimensional data that lies on or near a low dimensional manifold. The proposed map unfold the manifold from which the data points are sampled. It was demonstrated that the kernel learning method outperforms Gaussian and polynomial kernels for manifold learning. In Zhou and Sun (2016), a supervised dimensionality reduction algorithm is proposed which is called manifold partition discrimination analysis. The method is based on a local embedding along the direction of the local variation of the data while ensuring that the nearby data from different classes are well-separated. The proposed method has the advantage of capturing the pairwise and higher order interactions among data points improving the measures of within-class similarities, and consequently a higher quality of dimensionality reduction. The efficacy of the proposed method is proposed using various datasets related to face recognition, text recognition and vehicle systems. In Zhang and Zha (2004), the authors proposed a manifold learning algorithms for nonlinear dimensionality reduction which learns the local geometry of the manifold from data points sampled under noise from a parametrized manifold. The tangent spaces are then aligned to develop a global coordinate system for the underlying manifolds. The performance of the algorithm is demonstrated considering curves and surfaces embedded in high dimensional Euclidean spaces.

In recent years, statistical analysis on manifold spaces has been extended from research on dimensionality reduction to other areas of statistics such as hypothesis testing (Chu and Dai, 2024), change point detection (Wang et al., 2025), longitudinal data analysis (Dai et al., 2021) and shape analysis (Yataka et al., 2023) and several theoretical results related to the consistency of the estimators and asymptotic distribution of statistics have been reported (Chakraborty and Vemuri, 2019; Huckemann and Eltzner, 2021).

Among the above research areas, hypothesis testing has attracted the most interest from researchers. In Chen and Friedman (2017) a two-sample test for non-Euclidean and object data was proposed based on similarity graph obtained from pooled samples using a closeness notion such as minimum spanning tree. It was demonstrated that the test statistic has power for both change in location and scale of the two samples. The asymptotic distribution under the null hypothesis is derived and it was demonstrated that the

test statistic is powerful considering real world problems including the assessment of the covariance balance in a matched observational study, and the comparison of network data. In Chen et al. (2018), a modified two-sample test is proposed for the test of location and scale change of object data using similarity measure graph. It is a modified version of the test proposed in Chen and Friedman (2017) which considers a weighting to balance the effect of sample size when the size of the two samples are different. The within-sample edges are assigned different weight according to which sample they are from. The test statistic is known as a weighted edge count test. It was proven that the test statistic has an asymptotic Normal distribution. The performance of the test statistic is attested using mobile network data with the objective of finding out if the phone call pattern on the weekdays is different from the phone call pattern on the weekends. In Zhan and Kang (2022), the authors proposed an automatic framework based on deep neural network to approximate uniformly most powerful test for composite hypothesis testing while controlling type-I error of the test statistic. The applicability of the proposed general framework is demonstrated using the adaptive COVID-19 treatment trial and the multiple Sclerosis and extract of Cannabis trial. It was proven that deep neural network can approximate test statistic and critical value to enhance the power under the alternative. In Dubey and Müller (2019), the authors proposed the use of Fréchet mean and variance for random objects and data sampled from spaces without geometrical structure, and accordingly proposed Fréchet analysis of variance to compare k populations of metric space-valued data objects. The finite sample behaviour of the proposed test statistic is investigated using simulation studies including probability distribution curve data, graph Laplacians, network data, mortality profile of several countries, and resting-state functional state magnetic resonance imaging data. Theoretical results for the consistency of sample Fréchet mean and variance are developed and accordingly the asymptotic distribution of test statistic under null is derived which is a χ^2 distribution with $k-1$ degrees of freedom where k is the number of groups. Under contiguous alternative, optimal rates for the convergence of power function to 1 was derived for the considered test statistic. In Hanik et al. (2020), the authors proposed a variant of Hotelling's T^2 statistic and Bhattacharayya distance for data taking values in Lie groups. The method is validated using manifold data indicating differences in hippocampal shape between individuals with mild impairment and normal controls. In Ginestet et al. (2017), the authors proposed a multiple sample hypothesis test for testing the difference of two samples of network data using the Laplacian matrices and geometrical properties of Laplacian space. The test statistic is obtained by half-vectorization mapping of the Laplace matrices to vectors. It was demonstrated that under the null distribution, the proposed test statistic converges to χ^2 distribution with $\binom{d}{2}$ degrees of freedom where d denotes the number of vertices on a graph. The efficacy of the method was proven using 1000 Functional Connectomes Project (FCP) network dataset which is a commonly used

neuroimaging dataset. In Cheng and Xie (2021*a*), the authors proposed a kernel based two-sample test on manifold spaces which uses reproducing kernel Hilbert space (RKHS) kernel maximum mean discrepancy statistic. The asymptotic rates of convergence and power analysis were carried out as a function of sample size, kernel bandwidth and intrinsic dimension of the manifold space. The benefit of presenting the consistency result as a function of intrinsic dimension is that the rates will not be affected by the curse of dimensionality when the manifold lies in a low dimensional space. The efficacy of the proposed method was proven using examples including clean and noise corrupted image data. In Hendriks and Landsman (1998), the authors proposed a two-sample hypothesis test for the equality of mean for data sampled from a Steifel manifold and special orthogonal group. Asymptotic normality of the hypothesis test was proven and the efficacy of the method was demonstrated through a MCMC simulation. In Zhang et al. (2018), the authors proposed a test statistic for multiple sample test of the equality of high dimensional covariance matrices. Consistency results were provided and it was proved that the test statistic asymptotically converges in law to standard Normal distribution. The performance of the test statistic was evaluated through a comparative study and a Monte Carlo simulation. In Cheng and Xie (2021*b*), the authors proposed a neural network framework based on neural tangent kernel and maximum mean discrepancy. The neural kernel theory states that the optimization problem can be represented using RKHS formulation, which enabled the authors to connect the neural optimization problem to neural kernel theory and formulate a two-sample test for data in general metric spaces. The theoretical properties of the proposed neural tangent kernel based statistic, e.g. type-I error and power of the test statistic, was derived using well-known kernel maximum mean discrepancy results. Numerical simulation on problems such as human activity change point detection and MNIST distribution abundance change indicated that the test statistic has power under the alternative. In Beran and Fisher (1998), a nonparametric test statistic is proposed for pairwise and multiple comparison of the mean of direction and axis datasets. The test statistic is distribution free with no shape assumption on the distribution of the sample from which the data is collected. Confidence intervals for α significance levels were derived for the test statistics and critical values were given. The asymptotic coverage probability rates were presented for the formulated confidence intervals. In Bhattacharya (2008), the author proposed a nonparametric statistics for hypothesis testing on manifold spaces with application to shape analysis. Hypothesis tests for the equality of sample variation and sample mean were proposed and their asymptotic behaviour in terms of convergence in distribution were studied. The proposed tests statistics were applied to shape analysis problems under 2-D shape space of k-ads setup. The simulation results demonstrated the potential of the method for hypothesis testing of data sampled from manifold shape spaces. In Osborne et al. (2013), the authors proposed a two-sample procedure for the equality of

generalized Frobenius means of two independent populations on the space of symmetric positive matrices. It was proven that the test statistic converges in distribution to χ^2 with m degrees of freedom, where m denotes the dimensionality of the manifold space. The proposed hypothesis test were applied to diffusion tensor imaging data. In Zheng et al. (2017), the authors proposed Riemannian variance filtering for statistical tests on manifold-valued data. The proposed procedure has two stages, where in the first stage some voxels are filtered based on a Riemannian variance filter and in the second stage voxels that passed the variance filter are used for the hypothesis testing. The test is then applied to brain imaged data.

From the above literature, it is apparent that hypothesis testing on manifold spaces of various nature, e.g. Grassmannian, positive definite matrices, shape spaces, projective spaces, space of direction and axis, has been actively followed by statistical community and is gaining an increasing interest. In spite of significant contributions in the field and theoretical results, there are still lots of potential to advance the research field. In particular, research on robust and nonparametric hypothesis testing on manifold spaces of various types needs further attention with convincing real-world examples properly demonstrating the benefit of working with such methods. One can refer to Fletcher et al. (2008); Lin et al. (2020); Asta (2024) for some nice reports on robust and nonparametric analysis for manifold data.

Research on statistical analysis on manifold spaces has been extended to the area of change point detection problem. In Duan et al. (2019), the authors proposed a matrix information geometric method for detecting change points of a rigid body motion. It was indicated that the set of rigid body motions is the set of special Euclidean group $SE(3)$, and consequently, the change point detection problem was formulated based on the Riemannian mean of the Lie group structure of $SE(3)$. It was indicated that the proposed change point detection algorithm can be extended to the space of general matrix Lie groups. The performance of the method was evaluated considering the motion of a cylinder problem and the motion control of a robot manipulator problem. In Zambon et al. (2018), the authors proposed a change point detection algorithm in graph streams based on constant-curvature manifold embeddings. The methodology consists of embedding graphs into a geometric space and performing change point detection on the resulting space using standard methods for numerical streams. The change point detection method was evaluated considering embedding graph streams on Euclidean, spherical and hyperbolic manifolds. In Xu et al. (2022), the authors proposed change point detection of high-dimensional streaming via dynamic sparse subspace learning. The methodology works based on the fact that high-dimensional streaming data usually lie in multiple low-dimensional subspaces, and therefore, manifold structures may change abruptly due several reasons, e.g. occurrence of

anomalies, pattern shift, etc. A subspace learning algorithm was proposed based on LASSO optimization problem for the change point detection of high-dimensional streaming data. The asymptotic properties of the change point detection algorithm was investigated and it was demonstrated that the estimated subspace is a consistent estimator of the true subspace. The method was applied to human motion tracking problem and its effectiveness was demonstrated. In Grattarola et al. (2019), the authors proposed a change point detection algorithm to detect change in stationarity of graph distribution based on neural network and constant-curvature Riemannian manifold. Autoencoder neural network estimates a constant-curvature embedding space for graphs and then change point detection algorithm is formulated based on geodesic distance in the corresponding manifold space and it was demonstrated that the statistics converges in distribution to χ^2 distribution. The performance of the algorithm was evaluated using synthetic graphs of type Delaunay triangulation of classes 0-4, functional connectivity network, and hostile behaviour for which the data is considered as human skeletal structure. In Vellekoop and Clark (2006), the authors proposed a nonlinear filtering approach for change point detection on the manifold space of Gaussian densities. The proposed method works based on the differential geometric theory of statistical information to project data from the space of infinite-dimensional unnormalized probability densities into a finite dimensional space. It was indicated that a three-dimensional nonlinear filter can yield global and unique solutions for data defined on the manifold space of Uniform densities and Normal densities.

Statistics in tree manifold spaces has attracted attention in recent years due to the emergent of complex real-world problems which requires considering a sampling space endued with the tree geometry. For example, it is well-known that evolutionary relationships can be represented by trees, known as Phylogenetic trees. Therefore, we need a sample of these trees to do statistical inference. In Nye et al. (2017), the authors proposed a variant of principal component analysis based on Geodesic paths (principal Geodesic) on the space of tree spaces to estimate a low-dimensional estimation that preserves much of the information from the high-dimensional sample space. Theoretical results were given indicating how to perform principal Geodesic analysis with respect to the topology of the space. Simulation studies considering two examples containing Apicomplexa and African coelacanth demonstrated the applicability of the proposed method to complex real-world problems. In Weyenberg et al. (2014), the authors proposed a nonparametric approach called KDETREES to estimate the distribution of Phylogenetic trees to identify outlier trees in a collected sample. The method was then used to identify gene expressions which were outliers in a history of collected genes. The simulation results considering a dataset of Epichloë, fungi symbiotic demonstrated the potential of the method in identifying outliers. In Weyenberg et al. (2016), a modified variant of KDETREES was proposed where

the kernel constant was estimated using a novel holonomic gradient method. The kernel was used to estimate the distribution of samples from a Billera-Holmes-Vogtman treespace with the goal of identifying outliers. In Billera et al. (2001), the authors studied the geometry of the space of Phylogenetic trees with aim of facilitating the statistical analysis of samples from this space. The efficacy of the developed metric was elaborated considering several mitochondrial DNA sequence data. In Barden et al. (2013), the authors investigated central limit theorem for Fréchet mean estimator in the space of Phylogenetic trees using a generalized version of delta method. Several results related to the convergence in distribution of the estimator to Normal distribution was presented. In Basrak (2010), the authors investigated the asymptotic properties of a location estimator for data sampled from a binary tree space. The estimator is recursively constructed using weighted interpolation between the current mean and the new data. The analysis was performed using tripod tree topology which is a 3-spider and binary tree topology. Theoretical findings demonstrated the usefulness of the hypothesis tests for the equality of location parameter of two samples. In particular, strong consistency results for the estimators of barycenter were established using a proper distance function on the space of trees. In Akhøj et al. (2022), the authors proposed tangent Phylogenetic PCA for dimension reduction on the dataset of shapes considering the nonlinear structure and Phylogenetic covariance structure. The method operates on data which are leaf nodes of Phylogenetic trees where it is assumed that the observations are not independent. A Brownian motion on Phylogenetic trees is proposed to model the evolution of covariance matrix. Theoretical results related to the joint Normality of the leaf-nodes are established. The methodology was then applied to shape dataset of mammal jaw and the principal components of the covariance matrix were estimated which capture the variability in the dataset. In Polly et al. (2013), the authors investigated the propoerties of Phylogenetic PCA and demonstrated that it can produce a shape space which preserves the Procrustes distance. A comparative analysis was conducted using ordinary PCA and it was indicated that phylogenetic grouping is more evident in Phylogenetic component of variation compared to PCA components of variation.

From the above literature, it can be realized that research on manifold analysis of tree structures has gained a significant attention. However, due to the complexity of the underlying manifold geometry, theoretical analysis reported in the literature still needs more work. In particular, to the author's best knowledge, results related to the rates of convergence given the complexity of the geometrical space is not established and needs careful analysis. Moreover, k-sample hypothesis testing and change-point analysis in tree spaces seems to be a promising research area given the importance of modeling stochastic processes representing the evolution of natural phenomena such as gene expressions.

Statistics on manifold spaces has been applied to real-world shape analysis problems. In Mardia and Patrangenaru (2005), the authors proposed a projective shape space for projective shape analysis with application to machine vision. Parametric and nonparametric methods were developed for multivariate directional and multivariate axial analysis. Frèchet extrinsic mean was investigated for axial and directional mean. Asymptotic distribution for the Hotelling's T-test statistics on tangent space was developed for testing the equality of means for two samples. Extrinsic mean as a location parameter for a distribution on shape space was developed and it was indicated that it is consistent estimator of location parameter. A Hotelling's T-test for the equality of mean for two samples was developed and it was indicated that the statistic asymptotically converges in distribution to χ^2_M, where $M \times M$ is the dimension of the covariance matrix of the corresponding distribution in the embedded Euclidean space. Applying the method to image processing problems demonstrated the potential of the developed hypothesis test for shape analysis. In Ambellan et al. (2019), the authors proposed a nonlinear statistical shape analysis model based on the differential coordinates viewed as the group of orientation preserving linear transformations $\mathrm{GL}^+(3)$. The geometric properties of the manifold space was used to derive the exponential and inverse exponential map and the principal components were calculated in the tangent space. The efficacy of the developed method was evaluated considering two anatomical datasets namely knee Osteoarthritis data and skeletal human hand data. In Patrangenaru (2001), the authors investigated the asymptotic properties of intrinsic mean associated with Riemannian metrics which are locally flat in the geodesically convex neighbourhood around the support of the probability measure defined on the shape space. Central limit theorem for the pivotal statistic on the shape space was derived with asymptotic Normal distribution. For finite sample, the approximation of the distribution of the pivotal statistic based bootstrap was proposed and empirical confidence interval was developed.

It is known that tensors are common sources of geometrical information which makes the development of statistical tools for analysis on such spaces very desirable. In line with this objective, statistical analysis on manifolds has also successfully been extended to tensor computing. In Pennec et al. (2006), the authors endowed the tensor space with an affine invariant Riemannian metric and demonstrated that their framework leads to strong theoretical properties. Based on the proposed framework various statistics and data mining algorithms were extended to tensor metric space, e.g. interpolation, Gaussian filtering and restoration of missing data. The developed methods were applied to diffusion tensor imaging (DTI) slice data around the corpus Callosum. It was observed that the use of Riemannian distance inheritted from the chosen space results in efficient interpretation. In Dryden et al. (2009), the authors proposed non-Euclidean statistics for covariance ma-

trices with applications to diffusion tensor in medical image analysis. The objective of the study was to estimate the mean covariance matrix, to perform principal component analysis of diffusion tensors, and to use Procrustes size and shape space. The simulation demonstrated that the proposed methods are much more efficient than usual Euclidean distance based methods especially when the datasets are of high-dimensional matrix type and near deficient rank. In Lenglet et al. (2006), the authors proposed statistics on the manifold of multivariate normal distributions with application to diffusion tensor MRI processing. Several important results such as geodesic distance between normal distributions, i.e. elements of the space, and intrinsic mean were established. All of these results were then used to characterize a generalized Normal law on the space of zero-mean multivariate Normal distributions. The efficacy of the method was demonstrated through MCMC simulation and real-world dataset for three DTI processing applications including generation of random elements in the space, robust gradient descent algorithm for the estimation of diffusion tensors, and for level-set segmentation of DTI dataset. In Fletcher et al. (2008), the authors proposed robust statistics on Riemannian manifold using geometric median. The median was estimated as the minimizer of the geodesic distances of the data points. Existence and uniqueness of the geometric median on manifolds with non-positive sectional curvature was proven and sufficient conditions for the uniqueness on positively curved data were given. Furthermore, an algorithm based on the generalization of Weiszfeld procedure for computing the geometric median on general manifold spaces was proposed. It was indicated that the algorithm converges to a unique solution when such a solution exists. Therefore, the algorithm gives a robust central point for data sampled from a manifold. The robustness properties of the estimates were investigated considering 3D rotation group, tensor manifold, and shape space data contaminated with outliers and by comparing the results with Frèchet mean. In Fletcher and Joshi (2004, 2007), the authors proposed principal geodesic analysis on symmetric spaces to perform statistical analysis on the space of diffusion tensors. It was indicated that the space of diffusion tensors is a Riemannian manifold and methods such as averages and modes of variation were developed. Principal geodesic analysis for the symmetric spaces was developed and applied to computing the variability of diffusion tensor data. It was proven that the method is suitable for the registration of diffusion tensor images, production of statistical atlases from diffusion tensor data, and for the quantification of anatomical variability incurred from diseases.

Functional data analysis and longitudinal data analysis has also been extended to the manifold spaces. In Dai et al. (2021), the authors proposed an intrinsic functional principle component analysis for analyzing sparse longitudinal data on Riemannian manifold. The method is based on calculating the Frèchet mean with local regression analysis and smoothing the covariance structure of the data linearized on tangent space around the

mean. Asymptotic convergence rates for uniform consistency for mean and covariance estimators were obtained. The method is used for studying the development of brain connectivity in the longitudinal cohort of Alzheimer's disease and normal participants. It was shown that the method has merits in terms of yielding interpretable eigenfunctions and principal component scores. In Dai and Müller (2018), the authors developed a Riemannian principal component analysis for Riemannian manifold-valued functional data. Riemannian functional data analysis is performed by mapping the manifold-valued data to the tangent space around the Frèchet mean function via logarithmic map, and then carrying out standard functional principal component analysis in the tangent space. The obtained components are then transferred back to the manifold space using exponential map around the Frèchet mean function. Central limit theorem for the estimated mean was derived and also probabilistic convergence rates for other estimated components were given. The effectiveness of the proposed method was then proven considering fruitfly behaviour longitudinal data and flight trajectory on unit sphere data. In Muralidharan and Fletcher (2012), the authors proposed a method for analyzing longitudinal data in Riemannian manifolds. A generative hierarchical model is developed in which each individual is modeled by a geodesic trend. It was indicated that each geodesic in the model can be uniquely parameterized by a starting point and a velocity, which is a point in the tangent bundle. Sasaki metric is used as a natural distance on the tangent bundle for comparing these parameters. Hotelling's T^2 test statistic was used for developing a two sample hypothesis test on the manifold space. The method was then utilized for characterizing anatomical shape changes and for discriminating trends in anatomy that are healthy from trends that are due to disease. In Chen and Müller (2012), the authors developed a representation framework for analyzing functional data lying on a nonlinear low-dimensional space. The framework is based on manifold learning for estimating manifold modes of functional variation and functional manifold components. Asymptotic consistency results and probabilistic convergence rates for the estimates were derived. Simulation results using Berkeley growth data, Yeast cell cycle gene expression data and human mortality across countries data demonstrated the superiority of the proposed method over traditional cross-sectional mean and functional principal components.

Chapter 2

Hypothesis Testing

In this chapter, nonparametric multiple sample hypothesis tests are proposed based on both depth function and rank of depth functions for data points sampled from a manifold space $\mathcal{M}$. The rest of the chapter is organized as follows. In Section 1, a review is carried out on research for developing hypothesis tests for manifold data. In Section 2, nonparametric hypothesis tests based on rank-depth are proposed for the hypothesis testing of independence, two-sample hypothesis testing for quality monitoring, and multiple sample hypothesis testing for change in location and scale. Several important properties of these tests, i.e. asymptotic convergence, size and power, robustness with respect to qualitative-robustness (q-robustness) and level and power breakdown point, are studied. Section 3 is devoted to simulation study and discussion where we investigate the performance of these test statistics on several numerical and real-world examples. In section 4, we discuss the findings of the research study and outline some areas for future research.

2.1 Hypothesis Testing for Manifold Data

Hypothesis testing for object data and observations from manifold spaces is an emerging research area which has gained significant interest from statisticians in recent years. In this section, we discuss in detail some relevant important research work related to hypothesis testing on manifolds.

In Chen and Friedman (2017), the authors proposed a novel nonparametric test, called generalized edge-count test, based on minimum spanning tree, which is a variant of graph-based testing. The test can be used once a proper distance can be defined on the space from

which the samples are collected, and therefore is applicable to object data and manifold data. It was indicated that when we have location alternative and the dimension is low, the standard edge-count test can outperform the proposed test statistic; however, when the change is in scale and the dimension is moderate to high, the standard edge-count test quickly loses power while the proposed test has acceptable power. The exact null distribution of statistic as well as its asymptotic distribution (χ^2_2) are derived, and based on simulation study considering covariates appraisal (a study for comparing ultimate educational attainment for students who start college at two-year versus four-year college) it was indicated that the asymptotic distribution works properly for finite sample scenarios.

In Chen et al. (2018), the authors proposed a modified version of standard edge-count test called weighted edge-count test for two-sample hypothesis testing scenarios, which is best suited for cases where the number of observations in the two-samples are unbalanced. The authors obtained the exact permutation distribution under null as well as its asymptotic distribution under mild condition on similarity graph (normal distribution), and indicated that it is an accurate approximation for finite sample cases, by comparing the asymptotic p-value and exact permutation p-value. Considering MCMC simulation as well as a real data obtained from MIT Media Laboratory to test the difference between graph topology of phone calls during weekdays and weekends, it was observed that the proposed test statistic outperforms the generalized edge-count test for cases where the number of observations in the two-samples are unbalanced. It was also indicated that the generalized edge-count test is applicable to wider range of alternative scenarios compared to weighted edge-count test; however, under location alternative the power of weighted edge-count test is higher. Therefore, it was suggested that both tests can be used in a complementary way.

In Dubey and Müller (2019), the authors proposed a test statistic to compare k populations of metric spaced value objects in terms of the equality of variance and mean. The central limit theorem for the Frechet variance was derived, under mild conditions based on empirical process theory, and a consistent estimator of asymptotic variance was obtained. The asymptotic distribution for the test statistic was derived (χ^2_{k-1} where k is the number of different groups), and it was indicated that the statistic is consistent under contiguous alternatives that capture departures from null distribution of equality of Frechet means and Frechet variances. The performance of the proposed test statistic was studied considering wide range of object data including probability distributions (testing populations of curves) and graph Laplacians (testing population of networks). It was proven that the test statistic has a very good finite sample performance for simulations on groups of Normal probability distribution function data (which are curves defined on manifolds) using L^2-Wasserstein distance, and brain network of Alzheimer's patients data using graph Laplacians.

In Hanik et al. (2020), the authors proposed a generalization of Hotelling T^2 statistic for Riemannian manifolds as well as a general version of Mahalanobis and Bhattacharayya distance. The proposed modifications make the standard statistic and distance suitable for the cases where the data belong to Lie group and can be used for shape analysis. Both modified Mahalanobis and Bhattacharayya distances are bi-invariant, and therefore will be unaffected by the right or left translation of data points. Therefore, we can translate the mean to identity without changing the distances. The invariance property is useful since it enables us to centralize covariance matrices and then define a pooled covariance matrix. The p-value of the test statistic was computed using permutation setup yielding a bi-invariant distribution-free two-sample test. The test statistic was applied to group tests on malformations of right hippocampi due to mild cognitive impairment.

In Ginestet et al. (2017), the authors proposed one sample, two-sample and k sample hypothesis tests for the equality of Frechet means considering geometric characterization of the space of graph Laplacian to extend the classical tests to the cases of network data. A central limit theorem was derived for the half-vectorization map of Laplacian matrices, and based on that a test statistic for one sample test was obtained which asymptotically converges to χ^2_k distribution. Also, the variant of the statistic for multiple sample comparison was derived. The proposed statistics were applied on neuroimaging data on human subjects from the 1000 functional connectomes project. The power of the test statistics was found to be empirically well-behaved for all of the considered topologies (dimension of network) and sample sizes. In particular, it was observed that for a fixed sample size, the type-I error increases beyond nominal rate as dimension d increases. Also, it was indicated that the proposed test statistics outperform the mass-univariate approach, where a single hypothesis is run for each voxel (local inference), which is a standard method in the field of network hypothesis testing.

In Cheng and Xie (2021*a*), the authors proposed a Kernel two-sample tests which works based on the idea of maximum mean discrepancy on a manifold setup. Theoretical analysis was conducted to investigate the relation of the power of test with the kernel bandwidth, the number of samples and intrinsic dimension. It was proven that as long as data lie on or near a lower dimensional manifold, the power of the test statistic will not be affected by increasing the dimension of ambient space, therefore, demonstrating the importance of having a theoretical analysis which considers the intrinsic dimension for obtaining sharp bounds. Also, optimal rate for sample size as a function of distance between pair of distributions on manifolds were given, as long as the pdfs are Hölder type. The optimal rate of bandwidth as a function of intrinsic dimension and sample size was also given. Conditions for extension of the results to the case of bounded manifolds and near-manifold data were also derived. The efficiency of the test was studied considering several

examples, i.e. image data lying in manifolds under two rotation angle distributions, which is an example of a manifold with boundary, and image data contaminated with pixel-wise noises, and the case where the kernel used is not PSD. The simulation results indicated that the test has power as long as the variability of contamination noise remains at a small level, and loses its power as the level of contamination noise increases.

In Cheng and Xie (2021*b*), the authors proposed a hypothesis testing method which combines the idea of neural tangent kernel and maximum mean discrepancy. The test statistic is called neural network maximum mean discrepancy. Neural tangent kernel (NTK) characterizes the neural network training dynamics in early stages. NTK theory shown that network optimization can be approximated by RKHS formulation. The main idea is that the short-time training of a network is equivalent to computing the witness function of a kernel maximum mean discrepancy statistic with neural tangent kernel at time zero. In this way, the proposed NTKMMD approximates MMD with NTK. It was shown that the proposed statistic which is a function of network training objective function can be used for two-sample testing and detects difference in distribution if the statistic passes a threshold. Theoretical studies were performed to characterize the statistical properties of the test including type-I error and power, and conditions to guarantee the power of the test were given.

In Bhattacharya (2008), The authors conducted a throughout statistical analysis which includes obtaining asymptotic distribution of both intrinsic and extrinsic mean based on central limit theorem on metric spaces, and consistency of Frechet variance and its asymptotic distribution. Accordingly, two-sample test statistics on shape spaces was implemented for the equality of both instrinsic and extrinsic means, which asymptotically converge to χ^2 distribution. The test statistics are applied to two-sample testing using manifold data including schizophrenic children and Gorilla skulls. Based on a comparative study using well-known two-sample test statistics, it was indicated that both intrinsic and extrinsic test statistics are powerful and can detect change in the two-sample sets.

In Osborne et al. (2013), the authors proposed a two-sample procedure for testing the equality of means of two independent populations based on data collected from the space of symmetric positive matrices. The focus of the study was on random objects defined on Hadamard manifolds, and an invariant Riemannian metric was defined which is called Frobenius metric. The asymptotic normality of the proposed two-sample test statistic was proven on Hadamard manifold space. Also a nonparametric bootstrap approximation was presented which is suitable for small sample sizes. Numerical analysis was carried out on pairwise spatially registered DT images between two groups of children where one group consisted of children with normal reading ability and the other group consisted of children with dyslexia. The results were compared to a parametric test and it was shown that the

proposed nonparametric test has a reasonable performance, especially when the sample size is small.

In Zheng et al. (2017), the authors used independent filtering for data defined on manifolds, e.g. diffusion tensor imaging, Cauchy deformation tensors, which have versatile applications in neuroimaging studies. Compared to standard large-scale hypothesis testing which needs multiple testing adjustment for controlling the level of the test, it was shown that performing statistical tests at voxel locations that pass the filter increases the power of hypothesis tests for manifold valued data. The proposed group difference hypothesis test statistics are mean-based permutation test and Cramer's test, which are nonparametric and their p-values can be obtained using permutation null distribution.

From the above literature, it can be realized that research on developing robust and nonparametric tests based on rank and depth function needs further attention given the fact that they are free from assumptions regarding the distribution and are less sensitive to outliers. These are important observations which should be taken into account for analyzing complex, non-Euclidean data. In particular, we investigate the following:

- Motivated by the mentioned shortcoming, we develop rank-depth test statistics applicable to a wide range of manifold data with various geometrical properties. We demonstrate the robustness of these test statistics against outliers, and show that these tests have high power, considering a wide range of manifold spaces.
- We investigate various properties of the rank-depth tests and propose a graph-based approach which can be used when the manifold from which data are sampled is unknown.
- We develop asymptotic results for the proposed test statistics. The performance of the proposed hypothesis testing methods are evaluated via finite-sample Monte-Carlo simulation, based on which we demonstrate the validity these convergence bounds.

2.2 Nonparametric Hypothesis Testing

In this section, we present the proposed nonparametric hypothesis testing methods for data from manifold spaces. The proposed hypothesis tests are hypothesis testing of independence, two-sample hypothesis testing for quality monitoring, and multiple sample hypothesis testing for location and scale change.

2.2.1 Two-sample Hypothesis Testing Of Independence

Suppose that $X : \Omega \to \mathcal{M}$ is a random element with density f_X with respect to measure $P \in P(\mathcal{M})$, and $Y : \Omega \to \mathcal{M}$ is random element with density f_Y with respect to measure $Q \in P(\mathcal{M})$. The interest is to test the null hypothesis

$$H_0 : X \perp Y \quad v.s. \quad H_A : X \not\perp Y \ ,$$

where $X \perp Y$ denotes the independence of X and Y.

Suppose that $Z : \Omega \to \mathcal{M}$ is a random element with density f_Z with respect to measure $G \in P(\mathcal{M})$, where G is unknown. Consider independent random elements X^* and Y^* with known densities with respect to measures P^* and Q^*, respectively. Motivated by Sidak et al. (1999) in Euclidean setting, we can test the null hypothesis using the random variables

$$X = \exp_\mu \left((1-\delta) \log_\mu(X^*) + \delta \log_\mu(Z) \right) \ ,$$
$$Y = \exp_\mu \left((1-\delta) \log_\mu(Y^*) + \delta \log_\mu(Z) \right) \ ,$$

where δ is a constant within the range $[0, 1]$. Note that the summation can be done by calculating the median $\mu \in \mathcal{M}$ of the dataset $\{X, Y, Z\}$ and mapping the variables to the tangent space using logarithmic map $\log_\mu$. Then the resulting random variables are mapped back to the manifold $\mathcal{M}$ using the corresponding exponential map $\exp_\mu$. The random elements X and Y are independent if $\delta = 0$. We use this setup for our numerical simulation study, where we investigate the power of tests with respect to effect size δ. The proposed test statistics are applicable to other simulation scenarios and real-world problems. We examine our test statistics of independence considering problems in the area of community network.

A well-known approach for testing the independence of random variables is rank correlation coefficient (Kendall, 1949; Kruskal, 1958; Sidak et al., 1999). In this investigation, we extend the idea to the case of testing independence on manifold spaces using the metric-based notion of depth function and the rank of depth values, based on the fact that the considered depth functions are Borel measurable.

Suppose that we collected N pairs of samples $(X_1, Y_1), (X_2, Y_2), ..., (X_N, Y_N)$. We can calculate the paired depth of the sample, $\big(D(P, X_1), D(Q, Y_1)\big), \big(D(P, X_2), D(Q, Y_2)\big), \ldots, \big(D(P, X_N), D(Q, Y_N)\big)$, and rank the observed samples using the rank of depths, i.e. $(R_1^x, R_2^x, \ldots, R_N^x)$ and $(R_1^y, R_2^y, ..., R_N^y)$. We can use rank statistics which are invariant under permutation applied to R_i^x's and R_i^y's. Let d_i denote the antiranks of X_i's such that

$R_{d_i} = i$. Arrange all N pairs of observed depths as an ascending function of the ranks of $D(P, X_i)$'s, i.e. R_i's as below

$$\big(D(P, X_{d_1}), D(Q, Y_{d_1})\big), \big(D(P, X_{d_2}), D(Q, Y_{d_2})\big), ..., \big(D(P, X_{d_N}), D(Q, Y_{d_N})\big) \ .$$

Also, let R_i^0 be the rank of $D(Q, Y_{d_i})$. With the above properties, we can formulate depth based statistics which are functions of R_i^0's.

We can use rank-depth function to study the independence between random objects X and Y on manifold space $\mathcal{M}$. We consider geodesic depth for the analysis. We want to show that if our interest is to test the hypothesis of the independence of $X \sim P$ and $Y \sim Q$ we can use depth values $D(P, X)$ and $D(Q, Y)$, given that depth functions are Borel measurable. To do so, we use moment generating function (MGF) and probability scale residuals (PSR) (Shepherd et al., 2016) of depth values. Assume that X and Y are independent. Then, their joint distribution satisfies $\Gamma(P, Q) = P \cdot Q$. Consider the probability density function with the following form

$$f(x, y; \mu_x, \mu_y, \kappa) = \frac{1}{C(\mu_x, \mu_y, \kappa)} \exp\left(\frac{-\delta\big((x, y), (\mu_x, \mu_y)\big)}{\kappa}\right) \ ,$$

where $\mu_x, \mu_y \in \mathcal{M}$ and $\kappa > 0$ are the parameters of the distribution, $C(\mu_x, \mu_y, \kappa)$ is a normalization constant, and δ is a distance on $\mathcal{M}$, see for example von Mises distribution (Forbes et al., 2011) , Wishart distribution (Nydick, 2012), and Kent distribution (Kasarapu, 2015). Also, we assume isometric mapping such that $\delta(x, y) = \| \log_\mu x - \log_\mu y \|$.

For the submanifolds $\mathcal{M}_\mathcal{X}, \mathcal{M}_\mathcal{Y} \subseteq \mathcal{M}$, we have

$$\begin{aligned}
P(X \in \mathcal{M}_\mathcal{X}, Y \in \mathcal{M}_\mathcal{Y}) &= \frac{1}{C(\mu_x, \mu_y, \kappa)} \int_{\mathcal{M}_\mathcal{X}} \int_{\mathcal{M}_\mathcal{Y}} \exp\left(\frac{-\delta\big((x, y), (\mu_x, \mu_y)\big)}{\kappa}\right) d\Gamma(P, Q) \\
&= \frac{1}{C(\mu_x, \mu_y, \kappa)} \int_{\mathcal{M}_\mathcal{X}} \int_{\mathcal{M}_\mathcal{Y}} \exp\left(\frac{-\big(\|(\log_\mu x, \log_\mu y) - (\log_\mu \mu_x, \log_\mu \mu_y)\|\big)}{\kappa}\right) d\Gamma(P, Q) \\
&= \frac{1}{C(\mu_x, \mu_y, \kappa)} \int_{\mathcal{M}_\mathcal{X}} \int_{\mathcal{M}_\mathcal{Y}} \exp\left(\frac{-\big(\| \log_\mu x - \log_\mu \mu_x \| + \| \log_\mu y - \log_\mu \mu_y)\|\big)}{\kappa}\right) dP dQ \\
&= \left(\frac{1}{C'(\mu_x, \kappa)} \int_{\mathcal{M}_\mathcal{X}} \exp\left(\frac{-\big(\| \log_\mu x - \log_\mu \mu_x \|\big)}{\kappa}\right) dP\right) \\
&\qquad \times \left(\frac{1}{C'(\mu_y, \kappa)} \int_{\mathcal{M}_\mathcal{Y}} \exp\left(\frac{-\big(\| \log_\mu y - \log_\mu \mu_y \|\big)}{\kappa}\right) dQ\right) \\
&= P(X \in \mathcal{M}_\mathcal{X}) P(Y \in \mathcal{M}_\mathcal{Y}) \ .
\end{aligned}$$

For the MGF of depth values, we can write

$$
\begin{aligned}
M_{GD(\Gamma(P,Q),(X,Y))}(s,t) &= \mathbb{E}_{\Gamma(P,Q)}\left[\exp\left(sGD(P,X)+tGD(Q,Y)\right)\right] \\
&= \frac{1}{C(\mu_x,\mu_y,\kappa)}\int_{\mathcal{M}_{\mathcal{X}}}\int_{\mathcal{M}_{\mathcal{Y}}}\exp\left(sGD(P,x)+tGD(Q,y)\right)\exp\left(\frac{-\delta\big((x,y),(\mu_x,\mu_y)\big)}{\kappa}\right)d\Gamma(P,Q) \\
&= \left(\frac{1}{C'(\mu_x,\kappa)}\int_{\mathcal{M}_{\S}}\exp\left(sGD(P,x)\right)\exp\left(\frac{-\left(\|\log_\mu x-\log_\mu \mu_x\|\right)}{\kappa}\right)dP\right) \\
&\qquad\qquad \times\left(\frac{1}{C'(\mu_y,\kappa)}\int_{\mathcal{M}_{\mathcal{Y}}}\exp\left(tGD(Q,y)\right)\exp\left(\frac{-\left(\|\log_\mu y-\log_\mu \mu_y\|\right)}{\kappa}\right)dQ\right) \\
&= \mathbb{E}_P\left[\exp\left(sGD(P,X)\right)\right]\mathbb{E}_Q\left[\exp\left(tGD(Q,Y)\right)\right] = M_{GD(P,X)}(s)M_{GD(Q,Y)}(t)\ ,
\end{aligned}
$$

for all $t,s\in\mathbb{R}$, implying that $GD(P,X)$ and $GD(Q,Y)$ are independent.

Now, we assume that the depth functions $D(P,X)$ and $D(Q,Y)$ are independent, i.e.

$$
M_{GD(\Gamma(P,Q),(X,Y))}(s,t) = M_{GD(P,X)}(s)M_{GD(Q,Y)}(t)\ .
$$

Using the Taylor approximation $e^{-u}\approx 1-u$, we get

$$
\begin{aligned}
M_{GD(\Gamma(P,Q),(X,Y))}(s,t) &= \mathbb{E}_{\Gamma(P,Q)}\left[\exp\left(sGD(P,X)+tGD(Q,Y)\right)\right] \\
&= \left(\frac{1}{C(\mu_x,\mu_y,\kappa)}\int_{\mathcal{M}_{\mathcal{X}}}\int_{\mathcal{M}_{\mathcal{Y}}}\exp\left(sGD(P,x)\right)\exp\left(\frac{-\delta\big((x,y),(\mu_x,\mu_y)\big)}{\kappa}\right)d\Gamma(P,Q)\right) \\
&\quad\times\left(\frac{1}{C(\mu_x,\mu_y,\kappa)}\int_{\mathcal{M}_{\mathcal{X}}}\int_{\mathcal{M}_{\mathcal{Y}}}\exp\left(tGD(Q,y)\right)\exp\left(\frac{-\delta\big((x,y),(\mu_x,\mu_y)\big)}{\kappa}\right)d\Gamma(P,Q)\right) \\
&= \left(\frac{e^s}{C(\mu_x,\mu_y,\kappa)}\int_{\mathcal{M}_{\mathcal{X}}}\int_{\mathcal{M}_{\mathcal{Y}}}\exp\left(-s\mathbb{E}_{Z\sim P}\left[\delta(x,Z)\right]\right)\exp\left(\frac{-\delta\big((x,y),(\mu_x,\mu_y)\big)}{\kappa}\right)d\Gamma(P,Q)\right) \\
&\quad\times\left(\frac{e^t}{C(\mu_x,\mu_y,\kappa)}\int_{\mathcal{M}_{\mathcal{X}}}\int_{\mathcal{M}_{\mathcal{Y}}}\exp\left(-t\mathbb{E}_{Z'\sim Q}\left[\delta(y,Z')\right]\right)\exp\left(\frac{-\delta\big((x,y),(\mu_x,\mu_y)\big)}{\kappa}\right)d\Gamma(P,Q)\right) \\
&= \left(e^{t+s}\right)\mathbb{E}_{\Gamma(P,Q)}\left[\exp\left(s\mathbb{E}_{Z\sim P}\left[\delta(X,Z)\right]+t\mathbb{E}_{Z'\sim Q}\left[\delta(Y,Z')\right]\right)\right]\ ,
\end{aligned}
$$

for $t, s \in \mathbb{R}$. Moreover, we have

$$\begin{aligned}
M_{GD(P,X)}(s) M_{GD(Q,Y)}(t) &= \left(\frac{1}{C'(\mu_x, \kappa)} \int_{\mathcal{M}_\mathcal{X}} \exp\left(sGD(P,x)\right) \exp\left(\frac{-\left(\| \log_\mu x - \log_\mu \mu_x \|\right)}{\kappa} \right) dP \right) \\
&\quad \times \left(\frac{1}{C'(\mu_y, \kappa)} \int_{\mathcal{M}_\mathcal{Y}} \exp\left(tGD(Q,y)\right) \exp\left(\frac{-\left(\| \log_\mu y - \log_\mu \mu_y \|\right)}{\kappa} \right) dQ \right) \\
&= \left(\frac{1}{C'(\mu_x, \kappa)} \int_{\mathcal{M}_\mathcal{X}} \exp\left(-s\mathbb{E}_{Z\sim P}\left[\delta(x,Z)\right]\right) \exp\left(\frac{-\left(\| \log_\mu x - \log_\mu \mu_x \|\right)}{\kappa} \right) dP \right) \\
&\quad \times \left(\frac{1}{C'(\mu_y, \kappa)} \int_{\mathcal{M}_\mathcal{Y}} \exp\left(-t\mathbb{E}_{Z'\sim Q}\left[\delta(y,Z')\right]\right) \exp\left(\frac{-\left(\| \log_\mu y - \log_\mu \mu_y \|\right)}{\kappa} \right) dQ \right) \\
&= \left(e^{t+s}\right) \mathbb{E}_P\left[\exp\left(s\mathbb{E}_{Z\sim P}\left[\delta(X,Z)\right]\right)\right] \mathbb{E}_Q\left[\exp\left(t\mathbb{E}_{Z'\sim Q}\left[\delta(Y,Z')\right]\right)\right] \ ,
\end{aligned}$$

for $t, s \in \mathbb{R}$. Hence

$$\mathbb{E}_{\Gamma(P,Q)}\left[\exp\left(s\mathbb{E}_{Z\sim P}\left[\delta(X,Z)\right] + t\mathbb{E}_{Z'\sim Q}\left[\delta(Y,Z')\right]\right)\right] = \mathbb{E}_P\left[\exp\left(s\mathbb{E}_{Z\sim P}\left[\delta(X,Z)\right]\right)\right] \mathbb{E}_Q\left[\exp\left(t\mathbb{E}_{Z'\sim Q}\left[\delta(Y,Z')\right]\right)\right] \ ,$$

for all $t, s \in \mathbb{R}$, implying that $\delta(X, z)$ and $\delta(Y, z)$ are independent for all $z \in \mathcal{M}$. As long as distance function δ applied to X and Y is Borel measurable, we have

$$\mathbb{E}_{\Gamma(P,Q)}\left[\mathbf{s}^\top \log_\mathcal{M} X + \mathbf{t}^\top \log_\mathcal{M} Y\right] = \mathbb{E}_P\left[\mathbf{s}^\top \log_\mathcal{M} X\right] \mathbb{E}_P\left[\mathbf{t}^\top \log_\mathcal{M} Y\right] \ ,$$

for all $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$, where d is the intrinsic dimension of manifold $\mathcal{M}$. Therefore, X and Y are independent (Nikolaos et al., 2025). Similar result holds for the other considered depth functions.

To use Spearman correlation coefficient as a measure of independence of $D(P, X)$ and $D(Q, Y)$, we can use probability scale residuals (PSR) (Shepherd et al., 2016) as well as the difference between the probability of concordance and discordance, which is a measure of association (Liu et al., 2018; Kruskal, 1958). In the univariate setup, let us consider the Spearman correlation between depth values formulated as scaled difference between the probability of concordance and the probability of discordance between (X, Y) and (X_0, Y_0), where X_0 and Y_0 have the same marginal distribution as X and Y; however, X_0 is independent of Y_0 and (X, Y) are independent of (X_0, Y_0) (Kruskal, 1958). This can be formulated as follows

$$\gamma_{X,Y} = c(P_C - P_D) \ ,$$

where c is a constant chosen such that $-1 \leq c(P_C - P_D) \leq 1$, and

$$P_C = \Pr\left(\text{sign}\left(X, X_0\right) \text{sign}\left(Y, Y_0\right) = 1\right) \ ,$$

$$P_D = \Pr\left(\text{sign}\left(X, X_0\right)\text{sign}\left(Y, Y_0\right) = -1\right) \ ,$$

where $\text{sign}(a, b) = \{-1, 0, 1\}$ if $a < b$, $a = b$, and $a > b$, respectively. Without loss of generality, assume that both random variables are 0 centered, i.e. $\mu_x = \mu_y = 0$. Note that X, Y, X_0 and Y_0 can be replaced by $(X - 0)$, $(Y - 0)$, $(X_0 - 0)$ and $(Y_0 - 0)$, which are measure of centrality of random variables.

Now, let us get back to to manifold spaces. In the setting of object data on the manifold space $\mathcal{M}$, we can replace these with depth values. The Spearman correlation between depth values can be formulated as scaled difference between the probability of concordance and the probability of discordance between $(D(P, X), D(Q, Y))$ and $(D(P, X_0), D(Q, Y_0))$, where X_0 and Y_0 have the same marginal distribution as X and Y; however, X_0 is independent of Y_0 and (X, Y) are independent of (X_0, Y_0). This equivalence relation can be modeled as follows

$$\gamma_{X,Y} = c'(P_C^* - P_D^*) \ ,$$

where c' is chosen such that $-1 \leq c'(P_C^* - P_D^*) \leq 1$, and

$$P_C^* = \Pr\left(\text{sign}\left(D(P, X), D(P, X_0)\right)\text{sign}\left(D(Q, Y), D(Q, Y_0)\right) = 1\right) \ ,$$

$$P_D^* = \Pr\left(\text{sign}\left(D(P, X), D(P, X_0)\right)\text{sign}\left(D(Q, Y), D(Q, Y_0)\right) = -1\right) \ ,$$

If there is no association between these depth values, $\mid c'(P_C^* - P_D^*) \mid$ gets closer to 0, and $\text{Cov}\left(\text{res}\left(D(P, X), P\right), \text{res}\left(D(Q, Y), Q\right)\right)$ gets closer to 0, where PSR $\text{res}\left(D(P, x), P\right)$ and $\text{res}\left(D(Q, y), Q\right)$ are defined as follows

$$\text{res}\left(D(P, x), P\right) = \mathbb{E}\big[\text{sign}\left(D(P, x), D(P, X_0)\right)\big] \ ,$$

$$\text{res}\left(D(Q, y), Q\right) = \mathbb{E}\big[\text{sign}\left(D(Q, y), D(Q, Y_0)\right)\big] \ .$$

gets closer to 0. As a result, the value of Spearman correlation coefficient test statistic defined as

$$S = \sum_{i=1}^{N} R_i^x R_i^y \ ,$$

becomes small. Similarly, it can be shown that the value of Quadrant test statistic becomes small. If X and Y are dependent, the value of $\mid c'(P_C^* - P_D^*) \mid$ gets closer to 1, and $\text{Cov}\left(\text{res}\left(D(P, X), P\right), \text{res}\left(D(Q, X), Q\right)\right)$ gets larger in magnitude. As a result the value of Spearman correlation coefficient test statistic becomes large. The same is true for Quadrant test statistic.

In what follows, we formulate Spearman correlation coefficient test and Quadrant test to test the independence of random elements X and Y using rank of depths. In the

next section, the performance of these tests will be studied considering data sampled from manifold spaces of various geometries.

The Spearman correlation coefficient test can be formulated for testing the independence of random elements X and Y

$$S = \sum_{i=1}^{N} R_i^x R_i^y = \sum_{i=1}^{N} i R_i^0 \ ,$$

which is equivalent to testing the well-known rank correlation coefficient test

$$\rho = \frac{12}{N^3 - N} \sum_{i=1}^{N} \big(R_i^x - \frac{N+1}{2}\big)\big(R_i^y - \frac{N+1}{2}\big) = \frac{12}{N^3 - N} \sum_{i=1}^{N} \big(i - \frac{N+1}{2}\big)\big(R_i^0 - \frac{N+1}{2}\big) \ .$$

The expectation and variance of the test statistics S and ρ satisfy the following conditions

$$\mathbb{E}[S] = \frac{1}{4} N(N+1)^2 \ ,$$

$$\mathrm{Var}[S] = \frac{1}{4} \frac{N^2(N+1)^2(N-1)}{144} \ ,$$

or equivalently $\mathbb{E}[\rho] = 0$ and $\mathrm{Var}[\rho] = \frac{1}{N-1}$.

Theorem 1. *Under the null hypothesis, each of the statistics* $\dfrac{S - \mathbb{E}[S]}{\sqrt{Var[S]}}$ *and* $\dfrac{\rho - \mathbb{E}[\rho]}{\sqrt{Var[\rho]}}$ *are asymptotically Normally distributed.*

Another test statistic for testing the independence of random elements X and Y can be given by Quadrant test as follows. Define

$$\begin{aligned} S^* &= \sum_{i=1}^{N} \mathrm{sign}\big(R_i^x - \frac{1}{2}(N+1)\big)\mathrm{sign}\big(R_i^y - \frac{1}{2}(N+1)\big) \\ &= \sum_{i=1}^{N} \mathrm{sign}\big(i - \frac{1}{2}(N+1)\big)\mathrm{sign}\big(R_i^0 - \frac{1}{2}(N+1)\big) \ . \end{aligned}$$

Let us denote the median of $\big(D(P, X_1), D(P, X_2), ..., D(P, X_N)\big)$ by m_X and the median of $\big(D(Q, Y_1), D(Q, Y_2), ..., D(Q, Y_N)\big)$ by m_Y.

A test statistic equivalent to the above test, known as Quadrant test, is

$$S = S_1 + S_2 + S_3 + S_4 \ ,$$

where S_1 is the number of pairs $\big(D(P, X_i), D(Q, Y_i)\big)$ for which $D(P, X_i) > m_X$ and $D(Q, Y_i) > m_Y$, $S_2 = \frac{1}{2}$ if there exist some i for which $D(P, X_i) = m_X$ and $D(Q, Y_i) > m_Y$, $S_3 = \frac{1}{2}$ if there exist some i for which $D(Q, Y_i) = m_Y$ and $D(P, X_i) > m_X$, $S_4 = \frac{1}{4}$ if $D(P, X_i) = m_X$ and $D(Q, Y_i) = m_Y$ for some i.

Under the null distribution, we have

$$\mathbb{E}[S] = \frac{1}{4}N \ ,$$
$$\mathrm{Var}[S] = \begin{cases} \frac{1}{16}\Big(\frac{N^2}{N-1}\Big) & \text{for } N \text{ even} \\ \frac{1}{16}(N-1) & \text{for } N \text{ odd} \end{cases} .$$

Theorem 2. *Under the null hypothesis, the statistic* $\dfrac{S - \mathbb{E}[S]}{\sqrt{Var[S]}}$ *is asymptotically Normally distributed.*

2.2.2 Hypothesis Testing for Quality Monitoring

In this section, we propose a variant of Wilcoxon test statistic to test the outlyingness of one distribution defined on a manifold $\mathcal{M}$ with respect to the other distribution defined on a same manifold $\mathcal{M}$, using the rank of data depth functions and an index denoted by $\mathcal{Q}$. The proposed two-sample test statistic is in particular important for phase-II quality monitoring, where the interest is to detect the loss of precision and the accuracy of a manufacturing process. The test statistic can detect both change in location and scale simultaneously.

This type of test is useful for data lying on manifold spaces such as shape spaces, graphs, etc., where the interest lies in detecting possible deflections on the products with respect to the distribution of healthy products. The efficacy of the quality monitoring test will be indicated considering a 3D printing problem where the quality of manufactured products is monitored with respect to the healthy distribution of in control products obtained at phase-I monitoring step.

Consider random element $Y : \Omega \to \mathcal{M}$ with measure $Q \in P(\mathcal{M})$, and random element $X : \Omega \to \mathcal{M}$ with measure $P \in P(\mathcal{M})$. For $y \in \mathcal{M}$ let us define the function $\Lambda(P, y) = \Pr_P(D(P, X) \leq D(P, y))$. $\Lambda(P, y)$ is the fraction of the population of X with respect to the measure $P \in P(\mathcal{M})$ that is less central than element $y \in \mathcal{M}$.

In Liu and Singh (1993), for multivariate data, the authors shown that under the null hypothesis

$$H_0 : X \text{ and } Y \text{ have the same distribution,}$$

$\Lambda(P, y)$ follows Uniform$[0, 1]$ distribution. We extend these results to the case where data are sampled from non-Euclidean spaces. We can define a quality index $\mathcal{Q}$

$$\mathcal{Q}_{P,Q} = \mathbb{E}_Q\Big[\Pr_P\Big(D(P, X) \leq D(P, Y)\Big)\Big] \ .$$

The index $\mathcal{Q}_{P,Q} = \mathbb{E}_Q[\Lambda(P, Y)]$ is the average of fractions $\Lambda(P, y)$ over all y's distributed with respect to measure $Q \in P(\mathcal{M})$.

Given that the considered depth functions are invariant to congruence transformation $x \to a * x$, it is easy to verify that the function Λ and index $\mathcal{Q}$ are also congruent invariant

$$\Lambda(P, y) = \Lambda(P_a, a * y) \ , \forall y \in \mathcal{M} \ ,$$

where P_a is the distribution of the transformed random element $a * X$ with X distributed with respect to measure P. Similarly

$$\mathcal{Q}_{P_a,Q_a} = \mathcal{Q}_{P,Q} \ .$$

If $\mathcal{Q} = \frac{1}{2}$, then the distribution of random elements X and Y are identical. If the index $\mathcal{Q} < \frac{1}{2}$, then on average more than 50% of the population of the random element X is deeper than any given y distributed according to distribution with measure Q, implying inconsistency between the distributions of X and Y.

We can use a variant of Wilcoxon rank sum statistic to test the hypothesis

$$H_0 : \mathcal{Q} = \frac{1}{2} \quad v.s. \quad H_A : \mathcal{Q} < \frac{1}{2}$$

which is equivalent to testing for the shift in the location and scale of the two distributions. Suppose we draw random samples $X = (X_1, X_2, ..., X_n)$ and $Y = (Y_1, Y_2, ..., Y_m)$ from distributions with measures P and Q, respectively. Consider also random sample $Z = (Z_1, Z_2..., Z_{n_0})$ from the distribution with respect to measure P. In practice, the empirical measure $\widetilde{P}_{n_0} = \frac{1}{n_0} \sum_{i=1}^{n_0} \delta_{Z_i}$ can be used as the reference population. Define $\Lambda(\widetilde{P}_{n_0}, X_i)$, the

proportion of Z samples having lower depth than X_i, i.e. the proportion of Z_k's satisfying $D(\widetilde{P}_{n_0}, Z_k) \leq D(\widetilde{P}_{n_0}, X_i)$. Similarly, we can define $\Lambda(\widetilde{P}_{n_0}, Y_i)$. Sort the $m+n$ calculated values of $\Lambda(\widetilde{P}_{n_0}, X_i)$ and $\Lambda(\widetilde{P}_{n_0}, Y_j)$ for all i and j in the ascending order and assign ranks $1, ..., m+n$ to them accordingly. The Wilcoxon statistic is defined as

$$W = \sum_{j=1}^{m} R(\Lambda(\widetilde{P}_{n_0}, Y_j)) \ ,$$

which can detect shift in location and scale simultaneously.

Theorem 3. *Under the null hypothesis, the test statistic W is normally distributed with*

$$\mathbb{E}[W] = \frac{m(n+m+1)}{2} \ ,$$

$$\mathrm{Var}[W] = \frac{mn(n+m+1)}{12} \ .$$

We investigate the consistency of the test statistic when we use empirical measures to calculate depth and rank-depth. The performance of the proposed rank-depth test statistic will be evaluated considering several quality control examples with object data sampled from manifold spaces of various geometries. In the next section, we extend the concept using rank of depths considering the underlying geometry of manifold spaces.

2.2.3 Multiple Sample Hypothesis Testing Of Location and Scale Change

In this section, we investigate multiple sample hypothesis testing using depth of rank test for comparing two or more distributions. Consider the sample $\mathbf{X}_j = \{X_{1,j}, ..., X_{n_j,j}\}$ generated from a population with distribution P_j for $j = 1, 2, \ldots, t$. We assume the samples $\mathbf{X}_1, \ldots, \mathbf{X}_t$ are independent.

The combined measure for t samples becomes $P = n^{-1} \sum_{j=1}^{t} n_j P_j$, where $n = \sum_{j=1}^{t} n_j$. Let $\mathcal{X} = \cup_{j=1}^{t} \mathbf{X}_j$ denote the combined sample. Let $D_{ij} = D(X_{i,j}, P)$ be the depth of $X_{i,j}$ for $i = 1, ..., n_j$ and $j = 1, ..., t$ with respect to P. Also, let $D_{ij}(k) = D(X_{i,j}, P_k)$ for $k = 1, 2, \ldots, t$. Let R_{ij} denote the rank of $X_{i,j}$ with respect to the depth function $D(\cdot, P)$.

Under the null hypothesis, we assume that all these samples are generated from a common distribution P, i.e.

$$H_0 : P_1 = P_2 = \cdots = P_t = P \ .$$

If we assume no ties in ranks, under the null hypothesis we have

$$\Pr_{H_0}(R_{ij} = r) = \begin{cases} n^{-1} & \text{if } r = 1, ..., n \\ 0 & \text{otherwise} \end{cases}$$

and for $(i, j) \neq (i', j')$, the joint probability of R_{ij} and $R_{i'j'}$ is

$$\Pr_{H_0}(R_{ij} = r, R_{i'j'} = r') = \begin{cases} \dfrac{1}{n(n-1)} & \text{if } r \neq r' \\ 0 & \text{otherwise} \end{cases}$$

Let

$$R_{\cdot j} = \sum_{i=1}^{n_j} R_{ij} \quad \text{and} \quad \bar{R}_{\cdot j} = \frac{R_{\cdot j}}{n_j} .$$

Under H_0, it holds that

$$\mathbb{E}_{H_0}(R_{\cdot j}) = \frac{n_j(n+1)}{2} ,$$

$$\mathrm{Var}_{H_0}(R_{\cdot j}) = \frac{n_j(n-n_j)(n+1)}{12} ,$$

$$\mathrm{Cov}_{H_0}(R_{\cdot j}, R_{\cdot j'}) = \frac{-n_j n_{j'}(n+1)}{12} , \quad j \neq j' .$$

In Chenouri and Small (2012), the authors proposed a variant of Kruskal-Wallis test statistic which has power against location and scale change alternatives, when depth function is applied to multivariate data. We extend the Kruskal-Wallis test using the proposed rank-depth for manifold data.

Let $R_{ij}(k)$ denote the rank of $D_{ij}(k)$. The advantage of working with $R_{ij}(k)$ instead of R_{ij} is that $R_{ij}(k)$ can discriminate samples in group k from those in group j when $j \neq k$. Let us define

$$H_j(k) = \frac{12 n_j}{(n-n_j)(n+1)} \Big(\bar{R}_{\cdot j}(k) - \frac{n+1}{2} \Big)^2 .$$

Let the weighted average of $H_j(k)$'s be defined as

$$\begin{aligned} H(k) &= \sum_{j=1}^{t} (1 - \frac{n_j}{n}) H_j(k) \\ &= \frac{12}{n(n+1)} \sum_{j=1}^{t} n_j \Big(\bar{R}_{\cdot j}(k) - \frac{n+1}{2} \Big)^2 . \end{aligned}$$

To make the test statistic independent of the choice of k, the modified kruskal-Wallis test statistic is defined as

$$\begin{aligned} H &= \frac{1}{t}\sum_{k=1}^{t} H(k) \\ &= \frac{12}{n(n+1)t}\sum_{k=1}^{t}\sum_{j=1}^{t} n_j\Big(\bar{R}_{\cdot j}(k) - \frac{n+1}{2}\Big)^2 \\ &= \frac{12}{n(n+1)t}\sum_{k=1}^{t}\sum_{j=1}^{t} \frac{R^2_{\cdot j}(k)}{n_j} - 3(n+1)\ . \end{aligned} \tag{2.1}$$

Theorem 4. *Under the null hypothesis, the Kruskal-Wallis test statistic H is asymptotically chi-squared distributed with $t-1$ degrees of freedom.*

2.2.4 Convergence of Empirical Rank-Depth Tests and Power Study

The above asymptotic results are based on the assumption that the test statistics are implemented using theoretical rank-depths. However, in practice, we have a sample from which we need to estimate the ranks based on empirical depth values. The following theorem establishes results on the consistency of the test statistics based on empirical ranks.

The theoretical results are derived based on the convergence of probability measures by means of integral probability metric (IPM) (Müller, 1997; Han et al., 2023) and the assumption of Lipschitz continuity and boundedness of depth functions with respect to probability measures. For any two probability measures $P, Q \in \mathcal{P}(\mathcal{M})$, IPM metric associated with class of functions $\mathcal{F}$ is defined as

$$D_{\mathcal{F}}(P,Q) = \sup_{f\in\mathcal{F}}\left|\int_{\mathcal{M}} f d(P-Q)\right|\ .$$

Assumption 1. *The class of functions $f \in \mathcal{F}$ is bounded $\sup_{f\in\mathcal{F}}|f| < \infty$, and Lipschitz continuous.*

Assumption 2. *The class of functions $\mathcal{F} \subset L^2(\mathcal{M})$ is a Vapnik-Chervonenkis (VC) class (Dudley, 1978) with finite VC-dimension which is controlled by the complexity of the function class $\mathcal{F}$ as well as the intrinsic dimension d of the manifold $\mathcal{M}$.*

Assumption 3. *Depth functions are Lipschitz continuous with respect to probability measure P and Q, that is*

$$\sup_{x\in\mathcal{M}} \mid D(x,P) - D(x,Q) \mid \leq C\delta(P,Q) \ ,$$

where $C > 0$ is a constant, and δ is a distance between probability measures.

Theorem 5. *Under Assumptions 1-3, the empirical rank-depth test statistics are consistent estimators of statistics based on theoretical rank-depths.*

$$T(P_n) - T(P) = O_p\left(n^{-\alpha}(d\log d)^{\alpha}\right) \ ,$$

where d is the intrinsic dimension of the manifold, $\alpha = \frac{1}{2}$ for the Spearman correlation coefficient test, Quadrant test, and Wilcoxon test, and $\alpha = 1$ for the Kruskal-Wallis test.

In the next theorem, we investigate the consistency of the rank-depth test statistics under general alternatives. It will be demonstrated that the proposed test statistics can control the size under null hypothesis and are consistent under general alternatives, i.e. their power approaches 1, as $n \to \infty$.

Theorem 6. *Under Assumptions 1-3, the empirical rank-depth test statistics satisfy the following:*

(a) Test Size: We assume that the α-level threshold c_0 of the test statistic T is well-defined. Under H_0, we have

$$P(T(P_n) \geq c_0 \mid H_0) = \alpha \ .$$

(b) Test Power: Suppose that $X \sim P$ and $Y \sim Q$. Under H_A, the test statistic T is consistent, namely

$$\lim_{n\to\infty} P(T(P_n) \geq c_0 \mid H_A) \to 1 \ .$$

2.2.5 Rank-Depth Tests with Estimators of Geodesic Distance

For calculating the depth functions and rank-depth test statistics, we considered geodesic distance on manifold $\mathcal{M}$ with the metric δ, for any two points $x, y \in \mathcal{M}$, as follows

$$\delta(x,y) = \inf_{\gamma} \int_0^1 \|\gamma'(t)\| dt \ ,$$

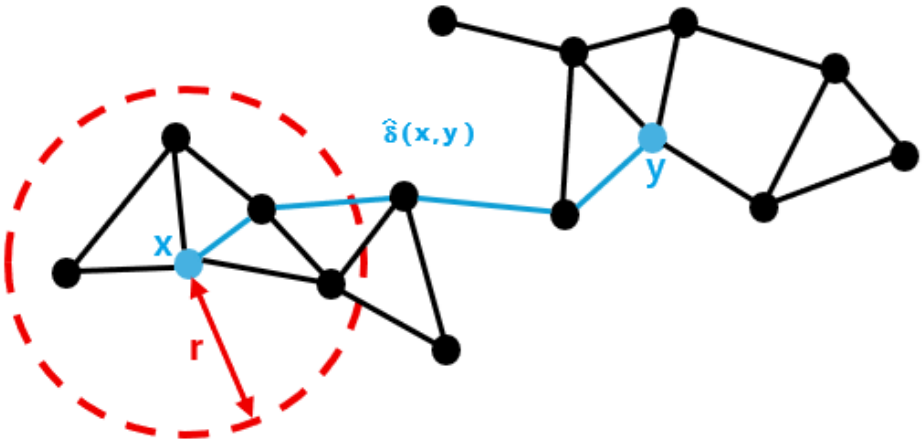


Figure 2.1: Schematic illustration of the graph distance estimator of geodesic distance in the constructed neighbourhood graph. The estimated distance equals the length of the blue edges which form the shortest path from x to y on the graph.

where the infimum is calculated over all smooth paths $\gamma : [0,1] \to \mathcal{M}$ from x to y, and $\|\cdot\|$ denotes the ambient Euclidean norm. In practice, we need to estimate the geodesic distance since the manifold topology is unknown. The following algorithm was proposed in Tenenbaum et al. (2000) and Chu and Dai (2024) for estimating geodesic distance based on the graph distance on a neighbourhood graph:

- Step 1: Input observations $\mathcal{Z}$, a pair of points (x, y) from observed data, and neighbourhood maximum radius size r.
- Step 2: Construct a graph by connecting all the pairs of observations $z_1, z_2 \in \mathcal{Z}$ for which the edges are such that $\|z_1 - z_2\| \leq r$.
- Step 3: Estimate the geodesic distance $\widehat{\delta}(x, y)$ as the length of the shortest path in the graph connecting x to y.
- Step 4: Output the estimated geodesic distance $\widehat{\delta}(x, y)$.

Figure 2.1 presents an schematic illustration of the graph distance estimator of geodesic distance. We can estimate the manifold depth function using $\widehat{\delta}(x, y)$. We propose the following graph-based estimators of depth functions.

Definition (Graph Spatial Depth Function). Let $\widehat{\delta}$ denote the graph-based estimation of geodesic metric δ from an unknown manifold $\mathcal{M}$. Also, let P_n denote the empirical probability measure, then the spatial depth of x with respect to P_n is defined as

$$\widehat{GSD}(P_n, x) = 1 - \frac{1}{2\binom{n}{2}} \sum_{i=1}^{n} \sum_{j=i}^{n} \frac{\widehat{\delta}^2(y_i, x) + \widehat{\delta}^2(y_j, x) - \widehat{\delta}^2(y_i, y_j)}{\widehat{\delta}(y_i, x)\widehat{\delta}(y_j, x)} .$$

Definition (Graph Geodesic Depth Function). Graph-based geodesic depth of point x with respect to the empirical measure P_n can be computed as

$$\widehat{GGD}(P_n, x) = \exp\Big(-\frac{1}{n} \sum_{i=1}^{n} \widehat{\delta}(x, y_i) \Big) .$$

Definition (Graph Half-Space Depth Function). Graph-based half-space depth at point x with respect to the empirical measure P_n is defined as

$$\widehat{GHD}(P_n, x) = \inf_{\substack{p_1, p_2 \in \mathcal{A} \\ \widehat{\delta}(p_1, x) \leq \widehat{\delta}(p_2, x)}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left[\widehat{\delta}(p_1, y_i) \leq \widehat{\delta}(p_2, y_i) \right] ,$$

where $\mathcal{A}$ is an anchor set.

We next provide convergence rate for depth functions estimated by graph distance $\widehat{\delta}$ based on the convergence result from (Aaron and Bodart, 2018) which depend on the intrinsic dimension d of the manifold $\mathcal{M}$. The following assumptions are needed to derive the asymptotic convergence rate.

Assumption 4. *The sampled data lie exactly on the manifold $\mathcal{M}$, which is a d-dimensional compact manifold, isometrically and smoothly embedded in the ambient space $\mathbb{R}^k$.*

Assumption 5. *The neighbourhood graph is constructed from pooled sample $\mathcal{Z}$ with maximum radius satisfying $r_n = c\left(\max_{x \in \mathcal{Z}} \left(\min_{y \neq x; y \in \mathcal{Z}} \|x - y\|\right)\right)^{2/3}$, where $c > 0$ is a constant.*

Assumption 6. *For two-sample test, $\mathcal{Z}$ is the pooled sample set with size $n = n_1 + n_2$, satisfying $n_1, n_2 \to \infty$, $n_1/(n_1 + n_2) \to \lambda \in (0, 1)$. For multiple sample test, $\mathcal{Z}$ is the pooled sample with size $n = n_1 + n_2 + \cdots + n_K$, satisfying $n_1, n_2, ..., n_K \to \infty$ with $\lambda_1, ..., \lambda_K \in (0, 1)$ and $\sum_{i=1}^{K} \lambda_i = 1$.*

Theorem 7. *Under Assumptions 4-6, the following convergence rate can be achieved for the general metric graph-based depth function estimators*

$$\sup_{x\in\mathcal{M}} |\widehat{GD}(\nu_n, x) - D(\nu, x)| = O\left(\left(\frac{\log n}{n}\right)^{2/(3d)}\right) ,$$

almost surely, where d is the intrinsic dimension of the manifold $\mathcal{M}$.

The following theorem proves the consistency of the rank-depth test statistics based on graph-based depth estimators.

Theorem 8. *Under Assumptions 4-6, the rank-depth test statistics T based on graph depth are consistent*

$$|\widehat{T}_{\mathcal{G}}(P_n) - T(P)| = O\left(n^{\alpha}\left(\frac{\log n}{n}\right)^{2/(3d)}\right) ,$$

almost surely, where $\alpha = -\frac{1}{2}$ for Spearman correlation coefficient test and Wilcoxon test, $\alpha = -1$ for Kruskal-Wallis test, and d is the intrinsic dimension of the manifold $\mathcal{M}$.

Example 1 (Graph-based Depth for Human Crania Data): The goal of this example is to investigate the performance the proposed graph-based depth estimator on a real-world manifold data related to quantitative and evolutionary biology (Mitteroecker et al., 2020). The dataset consisted of a sample of 24 adult human crania. Analyzing the size and shape variation of human crania is an important problem in evolutionary biology. A proper analysis requires representations which properly model functional and spatial formation of human crania. Manifold analysis can serve as a promising tool in such applications.

We use k landmark points in plane to represent the shape of products, i.e. the well-known Kendall's two-dimensional shape space Σ_2^k (Bhattacharya, 2008). The shape of each product is represented using k-ad, the set of ordered k points on a two-dimensional image of an object

$$\mathbf{z} = \{(x_j, y_j), 1 \le j \le k\} ,$$

where $k > 2$. The shape space Σ_2^k includes equivalent classes of all such k points under translation, rotation and scaling. The translation is removed using the transformation $\mathbf{z} - \langle \mathbf{z} \rangle$ where $\langle \mathbf{z} \rangle = \frac{1}{k}\sum_{j=1}^{k}(x_j, y_j)$. The scaling effect is removed using $\mathbf{w} = \frac{\mathbf{z}-\langle \mathbf{z} \rangle}{\|\mathbf{z}-\langle \mathbf{z} \rangle\|_2}$. The transformed $\mathbf{w}$ is called the preshape of k-ad $\mathbf{z}$. We use an extrinsic embedding called Veronese-Whitney embedding ϕ_E of Σ_2^k into the space $\mathcal{H}(k, \mathbb{C})$ of $k \times k$ complex Hermitian

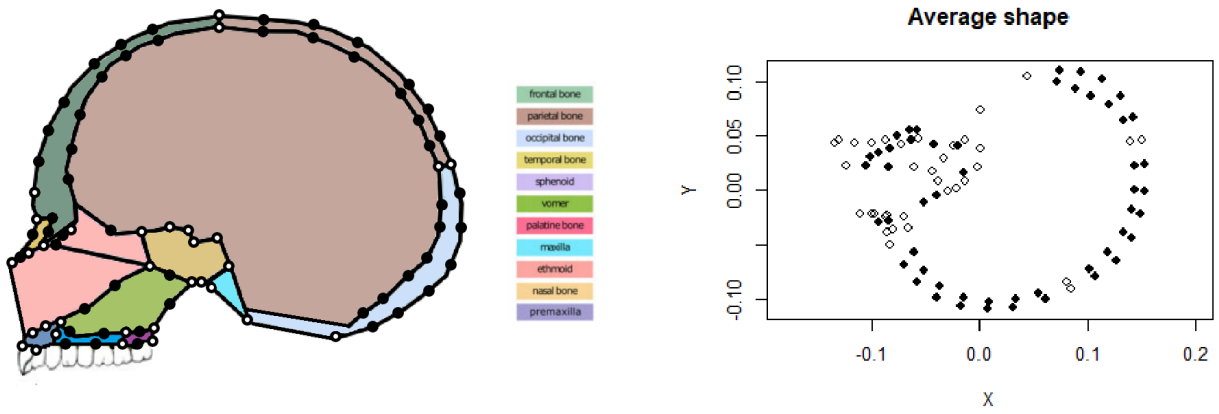


Figure 2.2: Schematic illustration of Midsagittal landmarks from CT scans of 24 adult human (16 females, 8 males) with a total of 87 landmarks comprises of 32 anatomical landmarks (white landmarks) and 55 semilandmarks (black dots).

matrices $\phi_{_E}([\mathbf{z}]) = \mathbf{w}\mathbf{w}^*$, where $\mathbf{w}^*$ is the transpose of the complex conjugate of $\mathbf{w}$, and $[\mathbf{z}]$ is a preshape (Bhattacharya, 2008). The norm and distance are defined as

$$\|A\|_{\mathcal{H}}^2 = \mathrm{tr}(AA^*) = \mathrm{tr}(A^2) \ ,$$

$$d(A,B) = \|A-B\|_{\mathcal{H}}^2 = \mathrm{tr}((A-B)^2) \ ,$$

where $d(\phi_{_E}([\mathbf{z}_1]), \phi_{_E}([\mathbf{z}_2])) = \rho_{_E}([\mathbf{z}_1],[\mathbf{z}_2])$ is the induced distance between preshapes $[\mathbf{z}_1]$ and $[\mathbf{z}_2]$, known as Procrustes distance (Bhattacharya, 2008).

Figure 2.2 illustrates the Midsagittal landmarks from CT scans of 24 adult human. There are a total of 87 landmarks from which 32 landmarks are anatomical and the rest of 55 landmarks are semilandmarks. In this analysis, we compare the graph-based GD estimator with GD obtained from an extrinsic embedding of the manifold space Σ_2^k. The second panel of the plot provides an schematic illustration of these coordinates for the average shape. Figure 2.3(a) illustrates the formed neighbouring graph with $r = 12.5$. Three possible paths are shown from vertex 5 to vertex 21, including the longest path, a possible path, and the shortest path, where the shortest path "5-8-15-16-21" is shown in bold red color. We use this formation of the graph and the corresponding distances to estimate the geodesic depths GGD. Figure 2.3(b) compares the calculated GGD and GD for 24 samples. It can be observed that there is a positive correlation between GGD and GD. To investigate the effect of sample size on the accuracy of estimated depth functions, we use the mean shape to generate a fabricated sample of size 100 of human crania using

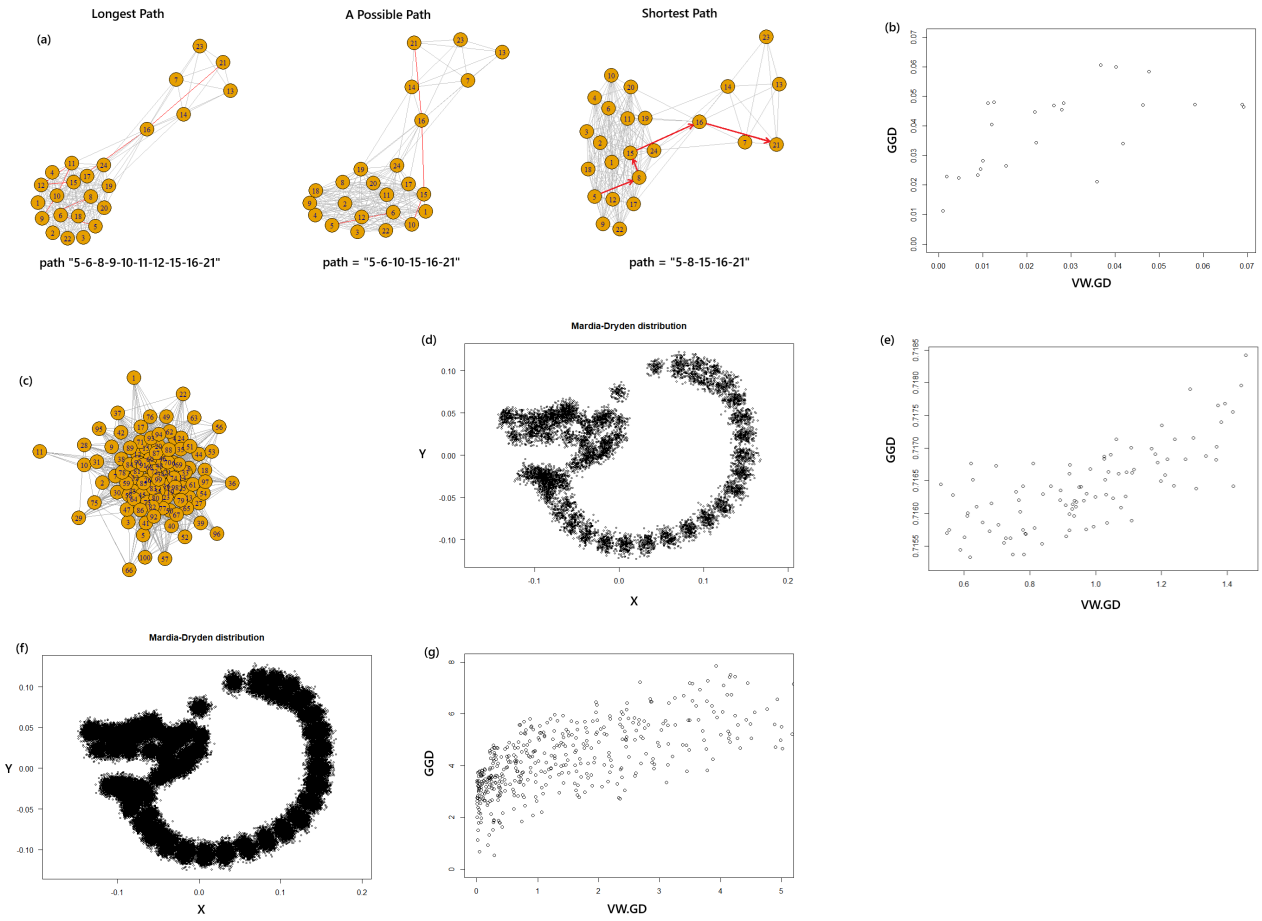


Figure 2.3: (a) Three possible paths from node 5 to node 21, the shortest path "5-8-15-16-21" is shown in bold red color, (b) the correlation between GGD based on graph distance of k-ad shape data and GD based on distance induced by the embedded Veronese-Whitney space for 24 sample data, (c) graph of 100 k-ad samples obtained from Mardia-Dryden distribution, (d) illustration of the 100 samples, (e) correlation between GGD and GD for 100 samples, (f) illustration of the 1000 samples, (g) correlation between GGD and GD for 1000 samples.

the well-known Mardia-Dryden distribution (Dryden and Mardia, 2016). This distribution is frequently used in statistical shape analysis for analyzing the shape of triangles as well as other landmark configurations. The idea is to displace landmark points using independent Gaussian perturbations. Figure 2.3(c) depicts the neighbouring graph of these 100 samples. Figure 2.3(d) illustrates these samples in the shape space Σ_2^k. Comparison of GGD and GD are shown in Figure 2.3(e). It can be realized that the correlation between GGD

and GD is improved by approximately 20% for $n = 100$ compared to $n = 24$. In a final experiment, we simulate 1000 samples from mean shape using Mardia-Dryden distribution. Figure 2.3(f) depicts these samples in the shape space Σ_2^k. Figure 2.3(g) compares the corresponding GGD and GD values. As can be seen, the correlation between GGD and GD is significantly improved, around 50% improvement for $n = 1000$ compared to $n = 24$, indicating the approximation power of GGD for moderately large sample sizes.

2.2.6 Fisher's Permutation Rank-Depth Tests

Permutation tests based on ranks can be used when the sample sizes are small to effectively approximate the critical values of test statistics. The permutation strategy is effective when the parametric assumptions are suspected, and therefore there is a need for a nonparametric approach. The idea is to estimate the null distribution using permutation distribution. Let $\mathcal{Z} = \{Z_1, \ldots, Z_{n+m}\} = \{X_1, \ldots, X_n, Y_1, \ldots, Y_m\}$ denote the pooled sample. Then, the permutation null distribution is the distribution of statistic $T_{\boldsymbol{\pi}} = T(Z_{\pi_1}, \ldots, Z_{\pi_{n+m}})$ where $(\pi_1, \ldots, \pi_{n+m})$ is a random permutation uniformly distributed on the set of all possible $(n+m)!$ permutations of $(1, 2, \ldots, n+m)$. For large sample sizes, the calculation of all possible permutations can become computationally complex. Therefore, in practice, the permutation null distribution is approximated by the empirical distribution of $T_{\boldsymbol{\pi}_1}, \ldots, T_{\boldsymbol{\pi}_B}$ for a large number B of permutations. In this section, we express a strategy called permutation test of Fisher (Tibshirani and Efron, 1993), which helps us to approximate the p-value for two/multiple sample hypothesis testing when the sample size is small. The steps are given below:

- Step 1: Compute the test statistic T based on the collected data $Z_1, ..., Z_N$ and call it T_N, where N is the total number of samples.
- Step 2: Randomly permute the combined sample a total of B times where B is a sufficiently large number. We denote these permutation vectors by $\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_B$.
- Step 3: For all of the permutations, group the data given the number of samples in a group, i.e. assign n_j samples to j-th group for $j = 1, ..., K$.
- Step 4: Compute the value of test statistic $T_{\boldsymbol{\pi}_b}$ for $b = 1, ..., B$. The null distribution can be approximated using the empirical distribution of $T_{\boldsymbol{\pi}_b}$.
- Step 5: The p-value can be approximated using the following

$$\widehat{\mathrm{P}}_{H_0}(T \geq T_N) = \frac{1}{B} \sum_{b=1}^{B} I(T_{\boldsymbol{\pi}_b} > T_N) \ .$$

Let us represent the joint distribution of two/multiple populations by G. The power functional can be formulated as

$$\Pi(G) = \mathrm{P}_G\Big(\widehat{\mathrm{P}}_{H_0}(T \geq T_N) \leq \alpha\Big) \ .$$

Under null hypothesis, $\Pi(G)$ approximates the type-I error and under a given alternative hypothesis, $\Pi(G)$ approximates the power of the test at G. We can use standard Monte Carlo methods to estimate $\Pi(G)$.

We can prove that the Fisher's permutation rank-depth test effectively controls the size under the null H_0, and is consistent under the alternative H_A.

Theorem 9. *Let $c_{n,m} = c_{n,m}(Z_1, \ldots, Z_{n+m})$ denote the α-upper quantile of permutation test $T_{\boldsymbol{\pi}}$ given sample $\mathcal{Z}$.*

(a) Test Size: We assume that for large enough value of $N = n + m$ ($N = n_1 + \ldots, n_K$ for multiple sample), the exact upper α-quantile $c_{n,m}$ of $T_{\boldsymbol{\pi}}$ given $\mathcal{Z}$ is well-defined. Under H_0, we have

$$P(T_{n+m} \geq c_{n,m} \mid H_0) = \alpha \ .$$

(b) Test Power: Supper that $X \sim P$ and $Y \sim Q$. Under H_A, the test statistic is consistent, namely

$$\lim_{n,m\to\infty} P(T_{n+m} \geq c_{n,m} \mid H_A) \to 1 \ .$$

2.2.7 Robustness of Rank-Depth Tests

Rank-based test statistics can be superior to their parametric competitors in terms of robustness. In particular, it can be realized that the impact of successive outlier data on rank-scores in the worst case is strictly decreasing as compared to the maximum influence of each outlier on the parametric test statistic competitors. Moreover, the values of the absolute deviation of outlier data can be arbitrarily increased without changing the rank of the data, and the value of the corresponding rank statistic. Thus, it is important to analyze the robustness of the proposed rank-depth test statistics. In this section, we analyze the robustness of the proposed hypothesis tests using the concepts of qualitative robustness (Mizera, 2010; Huber, 2011) and breakdown point (He et al., 1990).

Qualitative Robustness of Tests

We analyze the robustness of a sequence of rank-depth test statistics by using the idea of equicontinuity of error probabilities with respect to total variation metric and bounded

Lipschitz metric as well as the geometrical complexity of the underlying manifold space from which the object data are sampled. This idea is known as the qualitative definition of robust estimators, originally defined in Hampel (1971). After the introduction of qualitative robustness by Hampel (1971), it has been applied to frameworks such as hypothesis testing (Lambert, 1982; Liu and Briol, 2024), time series (Papantoni-Kazakos, 1984; Boldin, 2013), etc.

Consider a random element X on manifold $\mathcal{M}$ which is a measurable map on a probability space $(\Omega, \mathcal{F}, P)$ to $(\mathcal{M}, \mathcal{S})$, where $\mathcal{S}$ denotes the Borel sigma-algebra of $\mathcal{M}$. For two given probability measures P and Q defined on $\mathcal{S}$, the total variation distance (Rieder, 1982) is defined as

$$\delta_{TV}(P, Q) = \sup_{A \in \mathcal{S}} \mid P(A) - Q(A) \mid \ .$$

Let t_n denote the sequence of test statistics. For $X \sim P$, the distribution of t_n is denoted by $\mathcal{G}_P(t_n)$. A test statistic t_n is qualitatively robust (q-robust) at P if for any $\epsilon > 0$ there exists $\delta > 0$ such that

$$\delta_{TV}(P, Q) \leq \delta \quad \text{implies} \quad \delta_{TV}(\mathcal{G}_P(t_n), \mathcal{G}_Q(t_n)) \leq \epsilon \ ,$$

for all sufficiently large n.

Consider a class of probability measures $\mathcal{P}(\mathcal{M})$ on manifold $\mathcal{M}$. Let T be a functional (mapping) defined for all $P \in \mathcal{P}(\mathcal{M})$ such that t_n is the restriction of T to empirical measure P_n, i.e. $t_n = T \mid P_n$ (Huber, 2011). In Hampel (1971), it was demonstrated that for functionals T defined on $\mathcal{P}(\mathcal{M})$, the idea of q-robustness can be connected to weak continuity (Huber, 2011), i.e. the continuity implied by the weak convergence of probability measures.

Therefore, for the test statistics t_n written in the following format

$$t_n(x_1, \ldots, x_n) = T(P_n) \ ,$$

where P_n is the empirical probability supported by points $x_1, \ldots, x_n$, we have q-robustness as a result of weak continuity.

Assumption 7. *We assume that the representation of t_n by functional T is regular in the sense that the following properties are satisfied (i) it is consistent for any P in the domain of functional T, i.e. $t_n \xrightarrow{p} T(P)$, and (ii) for any $\tau \in T(P)$ there is a sequence of empirical probability measures P_n converging to P such that $T(P_n)$ converges to τ.*

The proposed rank-depth test statistics t_n can be re-written in the form of an equivalent functional form $T \mid P_n$, using square integrable score function φ (Sidak et al., 1999). Let φ be a real-valued function defined on $[0,1]$, i.e. $\varphi : [0,1] \to \mathbb{R}$, such that

$$\int_0^1 [\varphi(u) - \overline{\varphi}]^2 du < \infty \ ,$$

with $\overline{\varphi} = \int_0^1 \varphi(u) du$.

For the Spearman correlation coefficient test, we have

$$S_n = \sum_{i=1}^{n} R_i^x R_i^y = \sum_{i=1}^{n} i R_i^0 = \sum_{i=1}^{n} a(i) b(i) \ ,$$

where $a(i) = i$ for $i = 1, 2, ..., n$ and $b(i) = R_i^0 \in \{1, ..., n\}$ are score functions.

Note that R_i^0 is the rank of $D(Q, Y_{d_i})$, therefore, we can equivalently work with permutation function $\pi(i) = h(d_i)$, where d_i is the antirank of X_i's.

The function φ is chosen such that

$$a(i) = \mathbb{E}[\varphi(U_{(i)})] = \frac{n!}{(i-1)!(n-i)!} \int_0^1 \varphi(x) x^{i-1} (1-x)^{n-i} dx \ ,$$

where $U_{(i)}$ is the i^{th} order statistic from a Uniform$[0,1]$ random sample of size n. Let P_n and Q_n denote the probability measure associated with random samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ respectively.

The functional $T \mid P_n, Q_n$ has the following form

$$\begin{aligned} t_n &= \sum_{i=1}^{n} a(i) b(i) = \sum_{i=1}^{n} \mathbb{E}[\varphi_1(U_{(i)})] \mathbb{E}[\varphi_2(U_{(\pi(i))})] \\ &= \sum_{i=1}^{n} \left(\frac{n!}{(i-1)!(n-i)!} \int_0^1 \varphi_1(x) x^{i-1} (1-x)^{n-i} dx \right) \left(\frac{n!}{(\pi(i)-1)!(n-\pi(i))!} \int_0^1 \varphi_2(x) x^{\pi(i)-1} (1-x)^{n-\pi(i)} dx \right) \\ &= T \mid P_n, Q_n \ \ , \end{aligned}$$

where $\pi(i)$ is an appropriate permutation which is a function of observed samples $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$.

For Quadrant test, we can similarly show

$$t_n = S^* = \sum_{i=1}^{n} \text{sign} \left(i - \frac{n+1}{2} \right) \text{sign} \left(R_i^0 - \frac{n+1}{2} \right) = \sum_{i=1}^{n} a(i) b(i) \ ,$$

where $a(i) = \text{sign}\left(i - \frac{n+1}{2}\right)$ and $b(i) = \text{sign}\left(R_i^0 - \frac{n+1}{2}\right)$. The functional $T \mid P_n, Q_n$ has a similar form as in the Spearman correlation coefficient test.

For Wilcoxon test, the functional $T \mid P_n, Q_m$ has the following form

$$\begin{aligned} t_{m+n} = W &= \sum_{i=1}^{m} R(\Lambda(\tilde{P}_{n_0}, Y_i)) \\ &= \sum_{i=1}^{m} R_i^y = \sum_{i=1}^{m} \mathbb{E}[\varphi(U_{\pi(i)})] \\ &= \sum_{i=1}^{m} \frac{(n+m)!}{(\pi(i)-1)!(n+m-\pi(i))!} \int_0^1 \phi(x) x^{\pi(i)-1}(1-x)^{n+m-\pi(i)} dx \\ &= T \mid P_n, Q_m \ , \end{aligned}$$

where $\pi(i)$ is an appropriate permutation which is a function of observed samples $X_1, \ldots, X_n$, $Y_1, \ldots, Y_m$.

For Kruskal-Wallis test, we have

$$\begin{aligned} t_n = H &= \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{1}{n_j} \left(\sum_{i \in S_j} R_i \right)^2 - 3(n+1) \\ &= \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{1}{n_j} \left(\sum_{i \in S_j} a(i) \right)^2 - 3(n+1) \\ &= \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{1}{n_j} \left(\sum_{i \in S_j} \mathbb{E}[\varphi(U_{(i)})] \right)^2 - 3(n+1) \\ &= \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{1}{n_j} \left(\sum_{i \in S_j} \frac{n!}{(i-1)!(n-i)!} \int_0^1 \varphi(x) x^{i-1}(1-x)^{n-i} dx \right)^2 - 3(n+1) \\ &= T \mid P_n \ . \end{aligned}$$

For the theoretical format of the above functionals, i.e. $T \mid P$, we use the fact that $T \mid P_n$ is written as a function of order statistics

$$T(P_n) = \sum_{i=1}^{n} h(U_{(i)}) = \int_0^1 h\left(F_{n,U}^{-1}(s)\right) ds \ .$$

for some function h, and F_U the cdf of Uniform$[0,1]$ distribution. Then, it holds that

$$T(P) = \int_0^1 h\left(F_U^{-1}(s)\right) ds \ .$$

Given the derived $T \mid P_n$ and $T \mid P$, it is easy to see that the regularity conditions of Assumption 7 are satisfied for the derived functional representations of the test statistics.

Theorem 10. *Under the conditions of Assumption 7, the proposed rank-depth test statistics t_n are q-robust.*

Breakdown Point of Tests

The breakdown point is defined as the critical radius, or fraction of gross error, with respect to a contamination ball, i.e. $0 \leq \varepsilon \leq 1$, $\mathcal{B}_\varepsilon(P) = \{(1-\varepsilon)P + \varepsilon Q \mid Q \in \mathcal{P}(\mathcal{M})\}$, beyond which the test statistic is unable to correctly decide for either acceptance or rejection i.e. 0 or 1 (He et al., 1990; Zhang, 1996; Roy et al., 2023). Let $X_1, \ldots, X_n$ be a random sample of size n. Denote the test function by $\phi(X)$. Also, let $X^{(k)} = (X_1, \ldots, X_k)$ be the first k components of X. For $X^{(k)} = (X_1, \ldots, X_k)$ and $Y^{(\ell)} = (Y_1, \ldots, Y_\ell)$, denote $X^{(k)} \cup Y^{(\ell)} = (X_1, \ldots, X_k, Y_1, \ldots, Y_\ell)$. For $\mathcal{N} = \{1, \ldots, n\}$, $\mathcal{I}_k = \{i_1, \ldots, i_k\} \subset \mathcal{N}$, $1 \leq k \leq n$, let

$$\Pi_n\left(\mathcal{I}_k\right) = \{(Z_1, \ldots, Z_n) \mid Z_i = X_i,\ i \in \mathcal{I}_k,\ Z_j \in \mathcal{M},\ j \notin \mathcal{I}_k\} \ .$$

We can evaluate the robustness of a test statistic t_n by the effect of a fraction of outliers on its corresponding decision function ϕ (Loh, 2024). The replacement sample breakdown points of the acceptance and rejection decisions of ϕ is defined as

$$\varepsilon_A = \frac{1}{n}\min\left\{0 \leq m \leq n \,\middle|\, \max_{\mathcal{I}_{n-m} \subset \mathcal{N}} \sup_{Z \in \Pi_n(\mathcal{I}_{n-m})} \phi(Z) = 1\right\} \ ,$$

$$\varepsilon_R = \frac{1}{n}\min\left\{0 \leq m \leq n \,\middle|\, \min_{\mathcal{I}_{n-m} \subset \mathcal{N}} \inf_{Z \in \Pi_n(\mathcal{I}_{n-m})} \phi(Z) = 0\right\} \ .$$

We can measure the breakdown robustness of $\phi(X)$ as follows

$$P\left(\varepsilon_A(X) \geq v \mid \phi(X) = 0\right) \ ,$$
$$P\left(\varepsilon_R(X) \geq v \mid \phi(X) = 1\right) \ ,$$

where $v \geq 0$.

Computation of ε_A and ε_R can become difficult in practice because we need to take into account all possible $\binom{n}{k}$ ways of selecting k samples from the set of samples of size n. A simplified version is proposed by (Ylvisaker, 1977) which is based on replacing the components of sample set from right to left. The breakdown point based on simplified replacement is defined as

$$\varepsilon_{_{SA}} = \frac{1}{n}\min\left\{0 \le m \le n \,\middle|\, \sup_{Y^{(m)}\in\mathcal{M}} \phi\left(X^{(n-m)} \cup Y^{(m)}\right) = 1\right\} , \tag{2.2}$$

$$\varepsilon_{_{SR}} = \frac{1}{n}\min\left\{0 \le m \le n \,\middle|\, \inf_{Y^{(m)}\in\mathcal{M}} \phi\left(X^{(n-m)} \cup Y^{(m)}\right) = 0\right\} . \tag{2.3}$$

For the two-sample tests, consider another set of sample $W_1, \ldots, W_u$, and let the corresponding decision function be denoted by $\phi(X \cup W)$. Then $\varepsilon_A(X \cup W)$ and $\varepsilon_R(X \cup W)$ are the breakdown points of ϕ at $X \cup W$. The extension to multiple sample tests can be done similarly.

It can be verified that the proposed test statistics can be re-written using a measurable function ψ such that $T(x) = \sum_{i=1}^{n} \psi(x_i)/n$, with $\inf_x \psi(x) \le c_n \le \sup_x \psi(x)$. The following generalized formula for breakdown point (Zhang, 1996; Huber, 2011) can be used

$$\varepsilon_A = \frac{1}{n}\min\left\{0 \le m \le n-1 \,\middle|\, \min_{\mathcal{I}_{n-m}\subset\mathcal{N}} \frac{-T(X(\mathcal{I}_{n-m})) + c_n}{-T(X(\mathcal{I}_{n-m})) + \sup_x \psi(x)} \le \frac{m}{n}\right\} ,$$

$$\varepsilon_R = \frac{1}{n}\min\left\{0 \le m \le n-1 \,\middle|\, \min_{\mathcal{I}_{n-m}\subset\mathcal{N}} \frac{T(X(\mathcal{I}_{n-m})) - c_n}{T(X(\mathcal{I}_{n-m})) - \inf_x \psi(x)} \le \frac{m}{n}\right\} .$$

Suppose the test function $\phi(X)$ has the form

$$\phi(X) = \begin{cases} 1, & T(X) > c_n \\ 0, & \text{otherwise} \end{cases}$$

where c_n is the critical value for the test statistic.

It is well-known that the sample breakdown point of acceptance (rejection) is non-decreasing (non-increasing) in c_n, and the type-I and type-II error are non-increasing (non-decreasing). Hence, lower type-II (type-I) error are in conflict with that of higher sample breakdown point of acceptance (rejection) decisions.

The formulation of breakdown point for the proposed test statistics, i.e. Spearman correlation test, Quadrant test, Wilcoxon test and Kruskal-Wallis test as well as their asymptotic strong breakdown point limits are given below.

Spearman Correlation Test: For the Spearman correlation coefficient test, we can write

$$S = S(X,Y) = \sqrt{\frac{576}{n^2(n+1)^2(n-1)}} \sum_{i=1}^{n} \left(R_i^x R_i^y - \frac{1}{4}(n+1)^2 \right) ,$$

with test function

$$\phi(X \cup Y) = \begin{cases} 1, & S(X,Y) > c_n \\ 0, & \text{otherwise} \end{cases}$$

For $\mathcal{N} = \{1, \ldots, n\}$, $\mathcal{I}_k = \{i_1, \ldots, i_k\} \subset \mathcal{N}$, $1 \leq k \leq n$, let

$$\Pi_n(\mathcal{I}_k) = \{(Z_1, \ldots, Z_n) \mid Z_i = Y_i,\ i \in \mathcal{I}_k,\ Z_j \in \mathcal{M},\ j \notin \mathcal{I}_k\} .$$

The replacement sample breakdown points of the acceptance and rejection decisions for Spearman correlation test can be calculated as follows

$$\begin{aligned}
\varepsilon_A &= \frac{1}{n} \min \left\{ 0 \leq m \leq n \,\middle|\, \max_{\mathcal{I}_{n-m} \subset \mathcal{N}} \sup_{Z \in \Pi_n(\mathcal{I}_{n-m})} \phi(X \cup Z) = 1 \right\} \\
&= \frac{1}{n} \min \left\{ 0 \leq m \leq n \,\middle|\, \max_{\mathcal{I}_{n-m} \subset \mathcal{N}} \sup_{Z \in \Pi_n(\mathcal{I}_{n-m})} S(X,Z) > c_n \right\} , \\
\varepsilon_R &= \frac{1}{n} \min \left\{ 0 \leq m \leq n \,\middle|\, \min_{\mathcal{I}_{n-m} \subset \mathcal{N}} \inf_{Z \in \Pi_n(\mathcal{I}_{n-m})} \phi(X \cup Z) = 0 \right\} \\
&= \frac{1}{n} \min \left\{ 0 \leq m \leq n \,\middle|\, \max_{\mathcal{I}_{n-m} \subset \mathcal{N}} \inf_{Z \in \Pi_n(\mathcal{I}_{n-m})} S(X,Z) \leq c_n \right\} .
\end{aligned}$$

Suppose that we contaminate the sample Y with outliers $V^{(m)}$. For the strong breakdown point limit of ε_{SA} and ε_{SR}, we have

$$\begin{aligned}
\sup_{V^{(m)} \in \mathcal{M}} S(X,Z) &= \sup_{V^{(m)} \in \mathcal{M}} S\left(X, Y^{(n-m)} \cup V^{(m)}\right) \\
&= \sup_{V^{(m)} \in \mathcal{M}} \left(\frac{24}{\sqrt{n^2(n+1)^2(n-1)}} \right) \left(\sum_{i=1}^{n-m} R_i^x R_i^{y,Y^{(n-m)}} - \left(\frac{n-m}{n} \right) \frac{n(n+1)^2}{4} \right. \\
&\qquad \left. + \sum_{i=1}^{m} R_{n-m+i}^x R_{n-m+i}^{y,V^{(m)}} - \left(\frac{m}{n} \right) \frac{n(n+1)^2}{4} \right) ,
\end{aligned}$$

where

$$\sup_{V^{(m)}\in\mathcal{M}}\sum_{i=1}^{m}R^{x}_{n-m+i}R^{y,V^{(m)}}_{n-m+i}=\sup_{V^{(m)}\in\mathcal{M}}\sum_{i=1}^{m}(n-m+i)R^{0,V^{(m)}}_{n-m+i}=\sum_{i=1}^{m}(n+i-m)^2=\frac{6n^2m+7m^3+3m^2+2m-6nm^2}{6}.$$

Therefore, as $n\to\infty$, we have that

$$\left(\frac{24}{\sqrt{n^2(n+1)^2(n-1)}}\right)\sup_{V^{(m)}\in\mathcal{M}}\left\{\sum_{i=1}^{n-m}R^{x}_{i}R^{y,Y^{(n-m)}}_{i}-\left(\frac{n-m}{n}\right)\frac{n(n+1)^2}{4}\right\}$$

$$=O\left(\left(1+\frac{m}{n}\right)\sqrt{n}\right)+o(1)\ ,$$

$$\left(\frac{24}{\sqrt{n^2(n+1)^2(n-1)}}\right)\left(\sup_{V^{(m)}\in\mathcal{M}}\sum_{i=1}^{m}R^{x}_{n-m+i}R^{y,V^{(m)}}_{n-m+i}-\left(\frac{m}{n}\right)\frac{n(n+1)^2}{4}\right)$$

$$=O\left(\left(\frac{m}{n}\right)^3\sqrt{n}\right)+O\left(\left(\frac{m}{n}\right)^2\sqrt{n}\right)+O\left(\left(\frac{m}{n}\right)\sqrt{n}\right)+o(1)\ ,$$

almost surely. If we take $c_n=o(\sqrt{n})$, it holds that $\varepsilon_{_{SA}}>0$, implying that $\varepsilon_{_{SA}}$ does not converge to 0 almost surely as $n\to\infty$.

For $\varepsilon_{_{SR}}$, we have that

$$\inf_{V^{(m)}\in\mathcal{M}}\sum_{i=1}^{m}R^{x}_{n-m+i}R^{y,V^{(m)}}_{n-m+i}=\inf_{V^{(m)}\in\mathcal{M}}\sum_{i=1}^{m}(n-m+i)R^{0,V^{(m)}}_{n-m+i}=\sum_{i=1}^{m}(n-m+i)(m-i+1)=\frac{nm(m+1)}{2}.$$

Therefore

$$\begin{aligned}\inf_{V^{(m)}\in\mathcal{M}}S(X,Z)&=\left(\frac{24}{\sqrt{n^2(n+1)^2(n-1)}}\right)\inf_{V^{(m)}\in\mathcal{M}}\left\{\sum_{i=1}^{n-m}R^{x}_{i}R^{y,Y^{(n-m)}}_{i}-\left(\frac{n-m}{n}\right)\frac{n(n+1)^2}{4}\right\}\\&\quad+\left(\frac{24}{\sqrt{n^2(n+1)^2(n-1)}}\right)\left(\inf_{V^{(m)}\in\mathcal{M}}\sum_{i=1}^{m}R^{x}_{n-m+i}R^{y,V^{(m)}}_{n-m+i}-\left(\frac{m}{n}\right)\frac{n(n+1)^2}{4}\right)\\&=O\left(\left(\frac{m}{n}\right)\sqrt{n}\right)+o(1)+O\left(\left(\frac{m}{n}\right)^2\sqrt{n}\right)+O\left(\left(\frac{m}{n}\right)\sqrt{n}\right)+o(1)\ ,\end{aligned}$$

almost surely, which gives $\varepsilon_{SR} > 0$, implying that ε_{SR} does not converge to 0 almost surely as $n \to \infty$. The same asymptotic convergence rate can be achieved for level breakdown and power breakdown of the Spearman correlation coefficient test statistic. This indicates that the rank-depth test statistics are robust to outliers, as they can handle a substantial fraction of contaminated data without becoming unreliable. This shows the superiority of rank-depth test compared to parametric tests which can have asymptotic level and power breakdown of $o(1)$ (Zhang, 1996).

Quadrant test: For the Quadrant test, we have

$$S^* = S^*(X,Y) = \sqrt{\frac{16(n-1)}{n^2}} \left(\sum_{i=1}^{n} \text{sign}\left(R_i^x - \frac{n+1}{2} \right) \text{sign}\left(R_i^y - \frac{n+1}{2} \right) - \frac{1}{4} \right) ,$$

which has the following decision function

$$\phi(X \cup Y) = \begin{cases} 1, & S^*(X,Y) > c_n \\ 0, & \text{otherwise} \end{cases}$$

For $\mathcal{N} = \{1, \ldots, n\}$, $\mathcal{I}_k = \{i_1, \ldots, i_k\} \subset \mathcal{N}$, $1 \le k \le n$, let

$$\Pi_n(\mathcal{I}_k) = \{(Z_1, \ldots, Z_n) \mid Z_i = Y_i,\ i \in \mathcal{I}_k,\ Z_j \in \mathcal{M},\ j \notin \mathcal{I}_k\} .$$

The replacement sample breakdown points of the acceptance and rejection decisions for Quadrant test can be calculated as follows

$$\begin{aligned}
\varepsilon_A &= \frac{1}{n} \min \left\{ 0 \le m \le n \,\middle|\, \max_{\mathcal{I}_{n-m} \subset \mathcal{N}} \sup_{Z \in \Pi_n(\mathcal{I}_{n-m})} \phi(X \cup Z) = 1 \right\} \\
&= \frac{1}{n} \min \left\{ 0 \le m \le n \,\middle|\, \max_{\mathcal{I}_{n-m} \subset \mathcal{N}} \sup_{Z \in \Pi_n(\mathcal{I}_{n-m})} S^*(X,Z) > c_n \right\} , \\
\varepsilon_R &= \frac{1}{n} \min \left\{ 0 \le m \le n \,\middle|\, \min_{\mathcal{I}_{n-m} \subset \mathcal{N}} \inf_{Z \in \Pi_n(\mathcal{I}_{n-m})} \phi(X \cup Z) = 0 \right\} \\
&= \frac{1}{n} \min \left\{ 0 \le m \le n \,\middle|\, \max_{\mathcal{I}_{n-m} \subset \mathcal{N}} \inf_{Z \in \Pi_n(\mathcal{I}_{n-m})} S^*(X,Z) \le c_n \right\} .
\end{aligned}$$

For the strong breakdown point limit of ε_{SA} and ε_{SR}, we contaminate the sample Y

with outliers $V^{(m)}$. Then we have

$$\begin{aligned}\sup_{V^{(m)}\in\mathcal{M}} S^*(X,Z) &= \sup_{V^{(m)}\in\mathcal{M}} S^*\left(X, Y^{(n-m)}\cup V^{(m)}\right)\\ &= \sup_{V^{(m)}\in\mathcal{M}}\left(\sqrt{\frac{16(n-1)}{n^2}}\right)\left(\sum_{i=1}^{n-m}\operatorname{sign}\left(R_i^x-\frac{n+1}{2}\right)\operatorname{sign}\left(R_i^{y,Y^{(n-m)}}-\frac{n+1}{2}\right)-\left(\frac{n-m}{n}\right)\frac{n}{4}\right.\\ &\qquad\left.+\sum_{i=1}^{m}\operatorname{sign}\left(R_{n-m+i}^x-\frac{n+1}{2}\right)\operatorname{sign}\left(R_{n-m+i}^{y,V^{(m)}}-\frac{n+1}{2}\right)-\left(\frac{m}{n}\right)\frac{n}{4}\right).\end{aligned}$$

It can be shown that

$$\begin{aligned}&\sup_{V^{(m)}\in\mathcal{M}}\sum_{i=1}^{m}\operatorname{sign}\left(R_{n-m+i}^x-\frac{n+1}{2}\right)\operatorname{sign}\left(R_{n-m+i}^{y,V^{(m)}}-\frac{n+1}{2}\right)\\ &\quad=\sup_{V^{(m)}\in\mathcal{M}}\sum_{i=1}^{m}\operatorname{sign}\left((n-m+i)-\frac{n+1}{2}\right)\operatorname{sign}\left(R_{n-m+i}^{0,V^{(m)}}-\frac{n+1}{2}\right)\\ &\quad=\begin{cases}\sum_{i=1}^{m}\operatorname{sign}\left((n-m+i)-\frac{n+1}{2}\right) & ;m\le\lfloor n+1/2\rfloor\\ \sum_{i=1}^{m}\operatorname{sign}\left((n-m+i)-\frac{n+1}{2}\right)-\sum_{i=m+1}^{\lfloor n+1/2\rfloor}\operatorname{sign}\left((n-m+i)-\frac{n+1}{2}\right) & ;m>\lfloor n+1/2\rfloor\end{cases}.\end{aligned}$$

Hence, as $n\to\infty$, we have that

$$\begin{aligned}&\left(\sqrt{\frac{16(n-1)}{n^2}}\right)\sup_{V^{(m)}\in\mathcal{M}}\left\{\sum_{i=1}^{n-m}\operatorname{sign}\left(R_i^x-\frac{n+1}{2}\right)\operatorname{sign}\left(R_i^{y,Y^{(n-m)}}-\frac{n+1}{2}\right)-\left(\frac{n-m}{n}\right)\frac{n}{4}\right\}\\ &\qquad=O\left(\left(1+\frac{m}{n}\right)\sqrt{n}\right)+o(1)\ ,\\ &\left(\sqrt{\frac{16(n-1)}{n^2}}\right)\left(\sup_{V^{(m)}\in\mathcal{M}}\sum_{i=1}^{m}\operatorname{sign}\left(R_{n-m+i}^x-\frac{n+1}{2}\right)\operatorname{sign}\left(R_{n-m+i}^{y,V^{(m)}}-\frac{n+1}{2}\right)-\left(\frac{m}{n}\right)\frac{n}{4}\right)\\ &\qquad=\begin{cases}O\left(\left(\frac{m}{n}\right)\sqrt{n}\right)+o(1) & ;m\le\lfloor n+1/2\rfloor\\ O\left(\left(1+\frac{m}{n}\right)\sqrt{n}\right)+o(1) & ;m>\lfloor n+1/2\rfloor\end{cases},\end{aligned}$$

almost surely. If we take $c_n=o(\sqrt{n})$, it holds that $\varepsilon_{_{SA}}>0$, implying that $\varepsilon_{_{SA}}$ does not converge to 0 almost surely as $n\to\infty$

For ε_{SR}, it holds that

$$\begin{aligned}
&\inf_{V^{(m)}\in\mathcal{M}} \sum_{i=1}^{m} \operatorname{sign}\left(R^{x}_{n-m+i} - \frac{n+1}{2}\right) \operatorname{sign}\left(R^{y,V^{(m)}}_{n-m+i} - \frac{n+1}{2}\right) \\
&\quad = \inf_{V^{(m)}\in\mathcal{M}} \sum_{i=1}^{m} \operatorname{sign}\left((n-m+i) - \frac{n+1}{2}\right) \operatorname{sign}\left(R^{0,V^{(m)}}_{n-m+i} - \frac{n+1}{2}\right) \\
&\quad = \begin{cases} -\sum_{i=1}^{m} \operatorname{sign}\left((n-m+i) - \frac{n+1}{2}\right) & ; m \le \lfloor n+1/2 \rfloor \\ \pm \sum_{i=1}^{m} \operatorname{sign}\left((n-m+i) - \frac{n+1}{2}\right) \mp \sum_{i=m+1}^{\lfloor n+1/2 \rfloor} \operatorname{sign}\left((n-m+i) - \frac{n+1}{2}\right) & ; m > \lfloor n+1/2 \rfloor \end{cases} .
\end{aligned}$$

Therefore

$$\begin{aligned}
\inf_{V^{(m)}\in\mathcal{M}} S^{*}(X,Z) &= \left(\sqrt{\frac{16(n-1)}{n^2}}\right) \sup_{V^{(m)}\in\mathcal{M}} \left\{ \sum_{i=1}^{n-m} \operatorname{sign}\left(R^{x}_{i} - \frac{n+1}{2}\right) \operatorname{sign}\left(R^{y,Y^{(n-m)}}_{i} - \frac{n+1}{2}\right) - \left(\frac{n-m}{n}\right)\frac{n}{4} \right\} \\
&\quad + \left(\sqrt{\frac{16(n-1)}{n^2}}\right) \left(\sup_{V^{(m)}\in\mathcal{M}} \sum_{i=1}^{m} \operatorname{sign}\left(R^{x}_{n-m+i} - \frac{n+1}{2}\right) \operatorname{sign}\left(R^{y,V^{(m)}}_{n-m+i} - \frac{n+1}{2}\right) - \left(\frac{m}{n}\right)\frac{n}{4} \right) \\
&= O\left(\left(\frac{m}{n}\right)\sqrt{n}\right) + o(1) ,
\end{aligned}$$

almost surely, which gives which gives $\varepsilon_{SR} > 0$, implying that ε_{SR} does not converge to 0 almost surely as $n \to \infty$. The same asymptotic convergence rate can be achieved for level breakdown and power breakdown of the Quadrant test statistic.

Wilcoxon test: Consider the Wilcoxon test

$$W = W(X,Y) = \frac{1}{\sqrt{n_1 n_2 (n_1+n_2+1)/12}} \sum_{i=1}^{n_2} \left(R^{y}_{i} - \frac{(n_1+n_2+1)}{2} \right) ,$$

with test function

$$\phi(X \cup Y) = \begin{cases} 1, & W(X,Y) > c_n \\ 0, & \text{otherwise} \end{cases}$$

For $\mathcal{N} = \{1,\ldots,n_2\}$, $\mathcal{I}_k = \{i_1,\ldots,i_k\} \subset \mathcal{N}$, $1 \le k \le n_2$, let

$$\Pi_{n_2}(\mathcal{I}_k) = \{(Z_1,\ldots,Z_{n_2}) \mid Z_i = Y_i,\ i \in \mathcal{I}_k,\ Z_j \in \mathcal{M},\ j \notin \mathcal{I}_k\} .$$

The replacement sample breakdown points of the acceptance and rejection decisions for Wilcoxon test can be calculated as follows

$$\begin{aligned}
\varepsilon_A &= \frac{1}{n_2}\min\left\{0 \le m \le n_2 \,\middle|\, \max_{\mathcal{I}_{n_2-m}\subset\mathcal{N}} \sup_{Z\in\Pi_{n_2}(\mathcal{I}_{n_2-m})} \phi(X\cup Z) = 1\right\} \\
&= \frac{1}{n_2}\min\left\{0 \le m \le n_2 \,\middle|\, \max_{\mathcal{I}_{n_2-m}\subset\mathcal{N}} \sup_{Z\in\Pi_{n_2}(\mathcal{I}_{n_2-m})} W(X,Z) > c_n\right\} \ , \\
\varepsilon_R &= \frac{1}{n_2}\min\left\{0 \le m \le n_2 \,\middle|\, \min_{\mathcal{I}_{n_2-m}\subset\mathcal{N}} \inf_{Z\in\Pi_{n_2}(\mathcal{I}_{n_2-m})} \phi(X\cup Z) = 0\right\} \\
&= \frac{1}{n_2}\min\left\{0 \le m \le n_2 \,\middle|\, \max_{\mathcal{I}_{n_2-m}\subset\mathcal{N}} \inf_{Z\in\Pi_{n_2}(\mathcal{I}_{n_2-m})} W(X,Z) \le c_n\right\} \ .
\end{aligned}$$

The strong breakdown point limit for ε_{SA} and ε_{SR} can be obtained as follows. Let $n = n_1+n_1$ and $n_2 = \alpha n$ where $\alpha = (\frac{n_1}{n_2}+1)^{-1}$. We denote the outliers by $V^{(m)} = (V_1,\ldots,V_m)$, where $m \le \alpha n$. We have

$$\begin{aligned}
\sup_{V^{(m)}\in\mathcal{M}} W(X,Z) &= \sup_{V^{(m)}\in\mathcal{M}} W\left(X, Y^{(\alpha n-m)}\cup V^{(m)}\right) \\
&= \sup_{V^{(m)}\in\mathcal{M}} \left(\frac{1}{\sqrt{\alpha(1-\alpha)n^2(n+1)/12}}\right)\left(\sum_{i=1}^{\alpha n-m} R_i^{Y^{(\alpha n-m)}} - \left(\frac{\alpha n-m}{\alpha n}\right)\frac{\alpha n(n+1)}{2}\right. \\
&\qquad\qquad \left. + \sum_{i=1}^{m} R_i^{V^{(m)}} - \left(\frac{m}{\alpha n}\right)\frac{\alpha n(n+1)}{2}\right) \ ,
\end{aligned}$$

where

$$\sup_{V^{(m)}\in\mathcal{M}} \sum_{i=1}^{m} R_i^{V^{(m)}} = \sum_{i=1}^{m}(n-i+1) = n(m+1) - \frac{m+1}{2}(m+2) \ .$$

Therefore, as $n\to\infty$, we have that

$$\begin{aligned}
&\left(\frac{1}{\sqrt{\alpha(1-\alpha)n^2(n+1)/12}}\right) \sup_{V^{(m)}\in\mathcal{M}} \left\{\sum_{i=1}^{\alpha n-m} R_i^{Y^{(\alpha n-m)}} - \left(\frac{\alpha n-m}{\alpha n}\right)\frac{\alpha n(n+1)}{2}\right\} \\
&\qquad\qquad = O\left(\left(1+\frac{m}{n}\right)\sqrt{n}\right) + o(1) \ ,
\end{aligned}$$

$$\left(\frac{1}{\sqrt{\alpha(1-\alpha)n^2(n+1)/12}}\right)\left(\sup_{V^{(m)}\in\mathcal{M}}\sum_{i=1}^{m}R_i^{V^{(m)}}-\left(\frac{m}{\alpha n}\right)\frac{\alpha n(n+1)}{2}\right)$$
$$=O\left(\left(\frac{m}{n}\right)^2\sqrt{n}\right)+O\left(\left(\frac{m}{n}\right)\sqrt{n}\right)+o(1)\ ,$$

almost surely. If we take $c_n = o(\sqrt{n})$, it can be verified that $\varepsilon_{SA} > 0$, implying that ε_{SR} does not converge to 0 almost surely as $n \to \infty$.

For ε_{SR}, note that

$$\inf_{V^{(m)}\in\mathcal{M}}\sum_{i=1}^{m}R_i^{V^{(m)}}=\sum_{i=1}^{m}i=\frac{m(m+1)}{2}\ .$$

Therefore

$$\inf_{V^{(m)}\in\mathcal{M}}W(X,Z)=\left(\frac{1}{\sqrt{\alpha(1-\alpha)n^2(n+1)/12}}\right)\inf_{V^{(m)}\in\mathcal{M}}\left\{\sum_{i=1}^{\alpha n-m}R_i^{Y^{(\alpha n-m)}}-\left(\frac{\alpha n-m}{\alpha n}\right)\frac{\alpha n(n+1)}{2}\right\}$$
$$+\left(\frac{1}{\sqrt{\alpha(1-\alpha)n^2(n+1)/12}}\right)\left(\inf_{V^{(m)}\in\mathcal{M}}\sum_{i=1}^{m}R_i^{V^{(m)}}-\left(\frac{m}{\alpha n}\right)\frac{\alpha n(n+1)}{2}\right)$$
$$=O\left(\left(\frac{m}{n}\right)\sqrt{n}\right)+o(1)+O\left(\left(\frac{m}{n}\right)^2\sqrt{n}\right)+O\left(\left(\frac{m}{n}\right)\sqrt{n}\right)+o(1)\ ,$$

almost surely, which gives which gives $\varepsilon_{SR} > 0$, implying that ε_{SR} does not converge to 0 almost surely as $n \to \infty$. The same asymptotic convergence rate can be achieved for level breakdown and power breakdown of the Wilcoxon test statistic.

Kruskal-Wallis test: For the Kruskal-Wallis test, we have

$$H = H\left(\mathbf{X}_1,\ldots,\mathbf{X}_k\right)=\frac{12}{n(n+1)}\sum_{j=1}^{k}\frac{1}{n_j}\left(\sum_{i\in S_j}R_i\right)^2-3(n+1)\ ,$$

with test function

$$\phi\left(\bigcup_{i=1}^{k}\mathbf{X}_i\right)=\begin{cases}1, & H\left(\mathbf{X}_1,\ldots,\mathbf{X}_k\right)>c_n\\ 0, & \text{otherwise}\end{cases}$$

Define $\mathbf{X} = \left(\mathbf{X}_1^{(n_1)}, \ldots, \mathbf{X}_k^{(n_k)}\right) = (X_1, \ldots, X_n)$. For $\mathcal{N} = \{1, \ldots, n\}$, and $\mathcal{I}_\ell = \{i_1, \ldots, i_\ell\} \subset \mathcal{N}$, $1 \leq \ell \leq n$, let

$$\Pi_n(\mathcal{I}_\ell) = \{(Z_1, \ldots, Z_n) \mid Z_i = X_i,\ i \in \mathcal{I}_\ell,\ Z_j \in \mathcal{M},\ j \notin \mathcal{I}_\ell\} \ .$$

The replacement sample breakdown points of the acceptance and rejection decisions for Spearman correlation coefficient test can be calculated as follows

$$\begin{aligned}
\varepsilon_A &= \frac{1}{n}\min\left\{0 \leq m \leq n \,\middle|\, \max_{\mathcal{I}_{n-m} \subset \mathcal{N}} \sup_{Z \in \Pi_n(\mathcal{I}_{n-m})} \phi\left(\bigcup_{i=1}^{k} \mathbf{Z}_i\right) = 1\right\} \\
&= \frac{1}{n}\min\left\{0 \leq m \leq n \,\middle|\, \max_{\mathcal{I}_{n-m} \subset \mathcal{N}} \sup_{Z \in \Pi_n(\mathcal{I}_{n-m})} H(\mathbf{Z}_1, \ldots, \mathbf{Z}_k) > c_n\right\} \ , \\
\varepsilon_R &= \frac{1}{n}\min\left\{0 \leq m \leq n \,\middle|\, \min_{\mathcal{I}_{n-m} \subset \mathcal{N}} \inf_{Z \in \Pi_n(\mathcal{I}_{n-m})} \phi\left(\bigcup_{i=1}^{k} \mathbf{Z}_i\right) = 0\right\} \\
&= \frac{1}{n}\min\left\{0 \leq m \leq n \,\middle|\, \max_{\mathcal{I}_{n-m} \subset \mathcal{N}} \inf_{Z \in \Pi_n(\mathcal{I}_{n-m})} H(\mathbf{Z}_1, \ldots, \mathbf{Z}_k) \leq c_n\right\} \ .
\end{aligned}$$

Suppose that we contaminate the sample $X_k^{(n_k)}$ with outliers $V_k^{(m_k)}$, $\sum_{i=1}^{k} m_k = m$. For the strong breakdown point limit of ε_{SA} and ε_{SR}, we have

$$\begin{aligned}
\sup_{V^{(m)} \in \mathcal{M}} H(\mathbf{Z}_1, \ldots, \mathbf{Z}_k) &= \sup_{V^{(m)} \in \mathcal{M}} H\left(X_1^{(n_1-m_1)} \cup V_1^{(m_1)}, \ldots, X_k^{(n_k-m_k)} \cup V_k^{(m_k)}\right) \\
&= \frac{12}{n(n+1)} \sup_{V^{(m)} \in \mathcal{M}} \sum_{j=1}^{k} \frac{1}{n_j} \left(\sum_{i=1}^{n_j-m_j} R_i^{X_j^{n_j-m_j}} + \sum_{i=1}^{m_j} R_{n-m_j+i}^{V_j^{(m_j)}}\right)^2 - 3(n+1) \\
&= O\left(\left(1 + \frac{m}{n}\right)\sqrt{n}\right) + o(1) \ ,
\end{aligned}$$

almost surely. If we take $c_n = o(\sqrt{n})$, it holds that $\varepsilon_{SA} > 0$, implying that ε_{SA} does not converge to 0 almost surely as $n \to \infty$.

For ε_{SR}, we have that

$$\begin{aligned}
\inf_{V^{(m)} \in \mathcal{M}} H(\mathbf{Z}_1, \ldots, \mathbf{Z}_k) &= \inf_{V^{(m)} \in \mathcal{M}} H\left(X_1^{(n_1-m_1)} \cup V_1^{(m_1)}, \ldots, X_k^{(n_k-m_k)} \cup V_k^{(m_k)}\right) \\
&= \frac{12}{n(n+1)} \inf_{V^{(m)} \in \mathcal{M}} \sum_{j=1}^{k} \frac{1}{n_j} \left(\sum_{i=1}^{n_j-m_j} R_i^{X_j^{n_j-m_j}} + \sum_{i=1}^{m_j} R_{n-m_j+i}^{V_j^{(m_j)}}\right)^2 - 3(n+1) \\
&= O\left(\left(\frac{m}{n}\right)\sqrt{n}\right) + o(1) \ ,
\end{aligned}$$

almost surely, which gives which gives $\varepsilon_{_{SR}} > 0$, implying that $\varepsilon_{_{SR}}$ does not converge to 0 almost surely as $n \to \infty$. The same asymptotic convergence rate can be achieved for level breakdown and power breakdown of the Kruskal-Wallis test statistic.

All the above results are stated under the assumption that we have knowledge of the true distribution for $X_1, \ldots, X_n$, that is the statistic functional can be written as a function of true distribution, i.e. $T(P)$. However, this is not the case in practice where we only have a sample of observations making the use of empirical distribution P_n possible. In the following theorem, we will prove that the asymptotic breakdown point of $T(P_n)$ is at-least as large as that of $T(P)$. For the proof, we need the following assumption.

Assumption 8. *The estimated probability density function $\widehat{f}_n$ converges to the theoretical one f in L^p on $\mathcal{M}$, i.e. $\int |\widehat{f}_n - f|^p \to 0$, with probability 1.*

For kernel density estimate of $\widehat{f}_n$, given below

$$\widehat{f}_n(x_0) = \frac{1}{h_n} \int_{\mathcal{M}} K\left(\frac{\delta(x_0, x)}{h_n}\right) dP_n(x) \ , \ x_0 \in \mathcal{M} \ ,$$

where K is a kernel function, h_n is the bandwidth of the kernel, and δ is a distance metric on $\mathcal{M}$, Roy et al. (2023) proved the following convergence result

$$\int_{\mathcal{M}} |\widehat{f}_n(x) - f(x)|^p dx \xrightarrow{a.s.} 0 \ ,$$

where $h_n = o(n^{-2/9})$. For data sampled from SPD matrices space, one choice of kernel is $K(x_0, x) = \exp\left(\frac{-\delta^2(x_0,x)}{2h_n}\right)$, with $\delta(x_0, x) = \|\log(x) - \log(x_0)\|_F$ (Jayasumana et al., 2013).

Theorem 11. *Suppose that the condition of Assumption 8 is satisfied. Let ε_1^* denote the asymptotic breakdown point for functional $T(P)$, and ε_2^* denote the asymptotic breakdown point for functional $T(P_n)$. It can be shown that $\varepsilon_2^* \geq \varepsilon_1^*$.*

2.3 Simulation Study

In this section, we evaluate the performance of the proposed nonparametric test statistics using Monte Carlo simulations, as well as real-world examples from various research disciplines including additive manufacturing, community and social networks, and phylogenetic trees.

2.3.1 Two-sample Hypothesis Testing of Independence

Example 2 (Wishart Distributions): Consider the manifold space of positive definite symmetric matrices

$$SPD(k) = \{A \in \mathbb{R}^{k\times k} \mid A = A^\top, \text{ and } x^\top Ax > 0\} \ ,$$

and the space of the general linear group $\mathrm{GL}(k,\mathbb{R})$. For $G \in \mathrm{GL}(k,\mathbb{R})$ and $A \in SPD(k)$, we know that the group $\mathrm{GL}(k,\mathbb{R})$ acts on $SPD(k)$ by the action

$$\mathrm{GL}(k,\mathbb{R}) \times SPD(k) \to SPD(k) \ , \quad (G,A) \to G^\top AG \ .$$

Let A denote a $k \times k$ symmetric matrix of random variables that is positive definite, and V denote a fixed $k \times k$ symmetric positive definite matrix.

For $N \geq k$, A has a Wishart distribution with N degrees of freedom and probability density function

$$f(A) = \frac{1}{2^{Nk/2} \mid V \mid^{N/2} \Gamma_k(N/2)} \mid A \mid^{N-k-1/2} \exp\big(-\frac{1}{2}\mathrm{tr}(V^{-1}A)\big) \ ,$$

where $\mid A \mid$ is the determinant of A, and $\Gamma_k(N/2)$ is the multivariate Gamma function, defined as

$$\Gamma_k(N/2) = \pi^{k(k-1)/4} \prod_{j=1}^{k} \Gamma\left(\frac{N}{2} - \frac{j-1}{2}\right) \ .$$

The proposed rank-depth Spearman correlation test and Quadrant test statistics are used to test for independence. Consider

$$X_i^* \sim W_{20}(2I_2), \quad Y_i^* \sim W_{20}(6I_2), \quad Z_i \sim W_{15}(I_2) \ ,$$

and

$$\begin{aligned} X_i &= \exp_\mu\big((1-\delta)\log_\mu(X_i^*) + \delta \log_\mu(Z_i)\big) \ , \\ Y_i &= \exp_\mu\big((1-\delta)\log_\mu(Y_i^*) + \delta \log_\mu(Z_i)\big) \ . \end{aligned}$$

Our interest is to test the null hypothesis

$$H_0 : X \perp Y \quad v.s. \quad H_A : X \not\perp Y \ .$$

Table 2.1: Independence test simulation results for Wishart distribution. Reported values are the power of the test statistics.

		Spearman Test				Quadrant Test			
		$\delta = 0$	$\delta = 0.1$	$\delta = 0.8$	$\delta = 1$	$\delta = 0$	$\delta = 0.1$	$\delta = 0.8$	$\delta = 1$
$n = 20$	HD	0.06	0.03	1	0.99	0.05	0.13	1	0.98
$n = 50$	HD	0.06	0.06	0.97	1	0.04	0.18	1	1
$n = 100$	HD	0.07	0.06	1	1	0.08	0.18	1	1
$n = 500$	HD	0.04	0.07	1	1	0.03	0.13	1	1
$n = 20$	GD	0.04	0.03	0.99	1	0.03	0.15	0.98	1
$n = 50$	GD	0.05	0.06	1	1	0.08	0.11	0.99	1
$n = 100$	GD	0.06	0.07	1	1	0.07	0.16	1	1
$n = 500$	GD	0.04	0.06	1	1	0.08	0.13	1	1
$n = 20$	SD	0.03	0.04	0.99	1	0.06	0.16	0.97	1
$n = 50$	SD	0.04	0.06	1	1	0.08	0.17	0.99	1
$n = 100$	SD	0.06	0.07	1	1	0.08	0.09	1	1
$n = 500$	SD	0.04	0.08	1	1	0.02	0.14	1	1

We can use Fréchet approach to calculate the median of paired samples (X_1, Y_1), $(X_2, Y_2), \ldots, (X_n, Y_n)$ and calculate the HD, GD, and SD depth values accordingly. Then, the Spearman correlation coefficient and Quadrant tests can be applied for testing the independence of the pairs of observations.

The simulation results for different values of δ for 5000 simulation runs are reported in Table 2.1. The performed experiments revealed that the MC error becomes stable for replications greater than 2000 and is not affected by the dimensionality of the manifold space. As can be seen, Spearman test shows a good performance under null and alternative hypothesis tests at $\alpha = 0.05$ significance level. The interesting observation is that both Spearman correlation coefficient test and Quadrant test keep their acceptable empirical size for small sample size of $n = 20$ and also has a good power regardless of the sample size. It is also observed that the Quadrant test is more sensitive to small effect size values. In general both methods can effectively detect dependency in the data; however, Spearman correlation coefficient test has smaller type-I error compared to Quadrant test.

For the comparative power studies, we controlled the size of the test statistics at $\alpha = 0.05$ significance level. Figure 2.4(a) compares the performance of Spearman test with HD, GD, and SD depth functions. As it can be seen, all of the considered depth functions show comparable performance. Figure 2.4(b) compares HD, GD, and SD applied to Quadrant test. As it can be seen, the performance of the Quadrant test is also acceptable with all of the considered depth functions. Figure 2.4(c) compares the power curves for Spearman correlation coefficient test, Quadrant test, and ball covariance (BallCov) (Zhu et al., 2021)

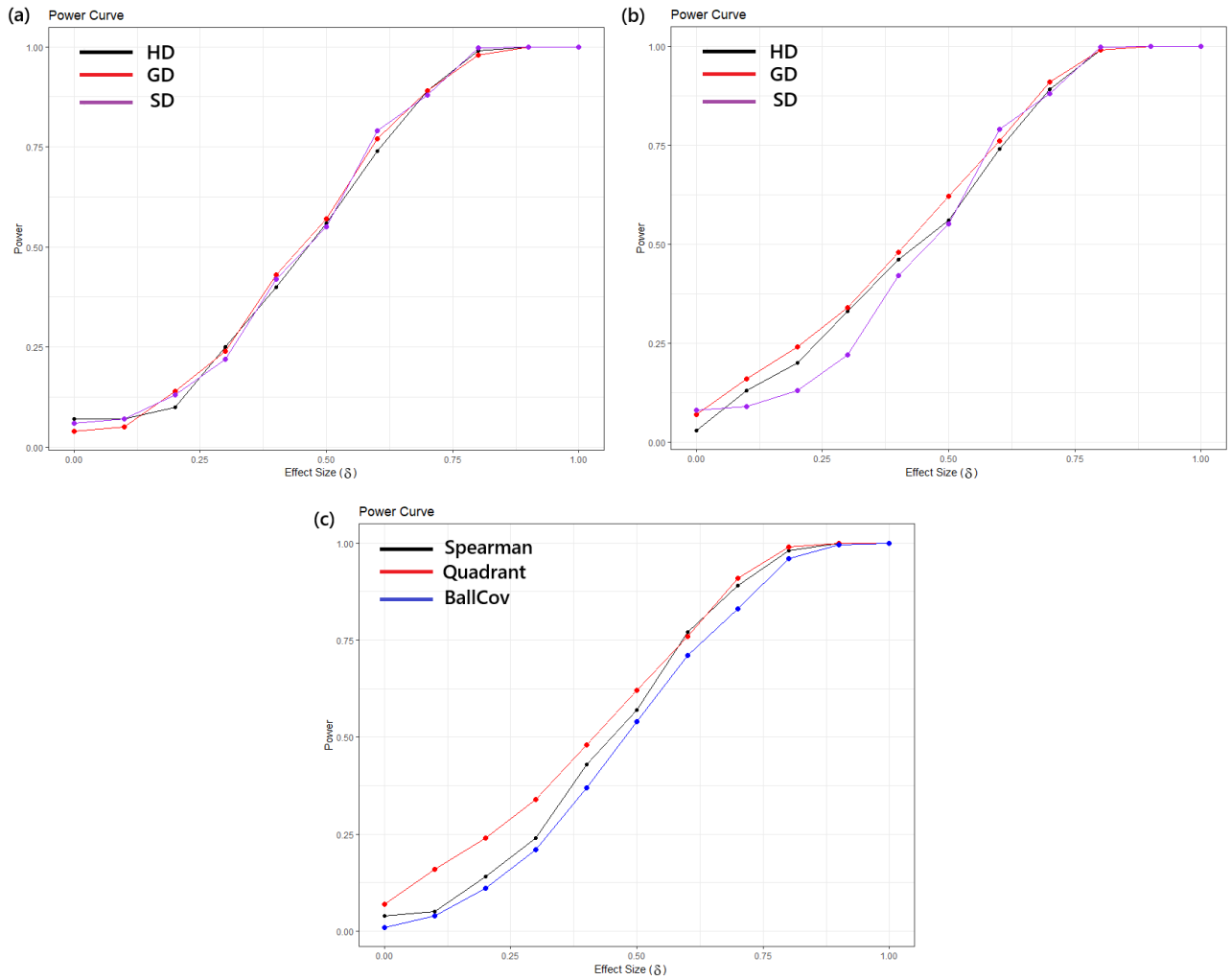


Figure 2.4: Power plots for data from Wishart distribution, (a) Power curves for Spearman correlation coefficient test with HD, GD, and SD depths, (b) Power curves for Quadrant test with HD, GD, and SD depths, (c) Power curves for Spearman correlation coefficient test with GD, Quadrant with GD, and ball covariance test.

methods. It is obvious that Quadrant test is more sensitive to small effect sizes, and also has larger type-I error. Spearman correlation coefficient test outperforms the other approaches as it can control type-I error and is powerful as the effect size increases.

Example 3 (von Mises Distributions): Consider sampling data on a torus manifold

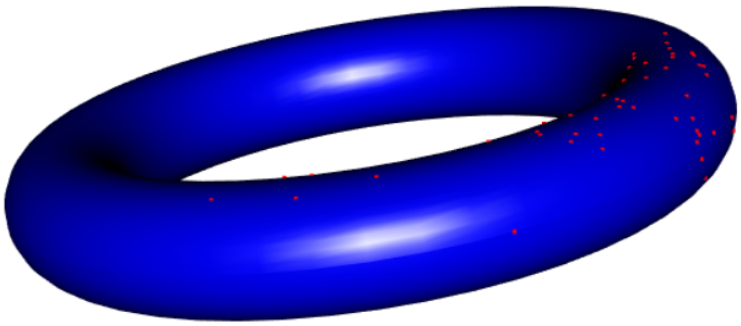


Figure 2.5: Schematic illustration of torus with data generated from a von Mises distribution.

space $\mathbb{T}^2 = \mathcal{S}^{2-1} \times \mathcal{S}^{2-1}$. We can generate data from $\mathbb{T}^2$ using multivariate von Mises distribution $MVM(\mu, \kappa, \delta)$ with probability density function for all $\theta \in \mathbb{T}^2$ defined as

$$f(\theta; \mu, \kappa, \Delta) = \frac{1}{Z(\kappa, \Delta)} \exp\{\kappa^\top c(\theta) + s(\theta)\Delta s(\theta)/2\} \ ,$$

where $\mu \in \mathbb{T}^2$ is the mean parameter, $\kappa \in \mathbb{R}^k$ is the concentration parameter and $\Delta = (\lambda_{i,j})$ is a symmetric matrix in $\mathbb{R}^{k \times k}$ with zero diagonal entries, $\lambda_{i,i} = 0$ for $i = 1, ..., k$, $Z(\kappa, \Delta)$ is the normalization constant, $c_i(\theta) = \cos(\theta_i - \mu_i)$, and $s_i(\theta) = \sin(\theta_i - \mu_i)$ for all $i = 1, 2, ..., k$. Consider

$$X_i^* \sim MVM\left(\begin{pmatrix} \pi/2 \\ 0 \end{pmatrix}, \begin{pmatrix} 20 \\ 20 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right) \ ,$$

$$Y_i^* \sim MVM\left(\begin{pmatrix} 7\pi/4 \\ 0 \end{pmatrix}, \begin{pmatrix} 100 \\ 100 \end{pmatrix}, \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix} \right) \ ,$$

$$Z_i \sim MVM\left(\begin{pmatrix} \pi/2 \\ 0 \end{pmatrix}, \begin{pmatrix} 50 \\ 50 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right) \ ,$$

and

$$X_i = \exp_\mu \left((1 - \delta) \log_\mu(X_i^*) + \delta \log_\mu(Z_i) \right) \ ,$$
$$Y_i = \exp_\mu \left((1 - \delta) \log_\mu(Y_i^*) + \delta \log_\mu(Z_i) \right) \ .$$

Table 2.2: Independence test simulation results for von Mises distribution. Reported values are the power of the test statistics.

		Spearman Test				Quadrant Test			
		$\delta = 0$	$\delta = 0.1$	$\delta = 0.8$	$\delta = 1$	$\delta = 0$	$\delta = 0.1$	$\delta = 0.8$	$\delta = 1$
$n = 20$	HD	0.01	0.06	0.82	0.99	0.03	0.14	0.86	1
$n = 50$	HD	0.03	0.05	0.94	0.98	0.04	0.12	0.75	1
$n = 100$	HD	0.04	0.07	0.99	1	0.04	0.12	0.97	1
$n = 500$	HD	0.04	0.06	1	1	0.03	0.13	1	1
$n = 20$	GD	0.05	0.06	0.89	1	0.07	0.14	0.83	1
$n = 50$	GD	0.03	0.07	0.85	1	0.08	0.15	0.82	1
$n = 100$	GD	0.05	0.06	0.98	1	0.09	0.08	1	1
$n = 500$	GD	0.04	0.08	1	1	0.03	0.12	1	1
$n = 20$	SD	0.05	0.03	0.76	0.89	0.07	0.13	0.77	0.90
$n = 50$	SD	0.05	0.06	0.82	1	0.09	0.18	0.78	1
$n = 100$	SD	0.03	0.05	1	1	0.02	0.14	1	1
$n = 500$	SD	0.02	0.07	1	1	0.04	0.16	1	1

Our interest is to test the null hypothesis

$$H_0 : X \perp Y \quad v.s. \quad H_A : X \not\perp Y \ .$$

Schematic illustration of torus with points generated according to von Mises distribution is shown in Figure 2.5. We can use Fréchet approach for calculating the median of the paired data and calculate the depths of pair observations accordingly. Thereafter, the Spearman correlation coefficient test and Quadrant test can be applied to test for the independence of X and Y. The simulation results for different values of δ over 5000 simulation runs are reported in Table 2.2. As can be seen, Spearman test with HD, GD, ans SD shows promising performance under both null and alternative. Quadrant test also effectively controls the test size under null hypothesis; however, for some cases it has a large type-I error, and its empirical size is significantly different from $\alpha = 0.05$. Also, it can be observed that the power for both Spearman correlation coefficient test and Quadrant test improve as the sample size gets larger. In summary, both Spearman correlation coefficient test and Quadrant test show promising performance but Spearman test effectively retains its empirical size below significance level while Quadrant test has a large false rejection rate for some cases under the null hypothesis. Figure 2.6(a) compares the power curves for Spearman correlation coefficient test with HD, GD, and SD depth functions. Figure 2.6(b) compares the power curves for Quadrant test with HD, GD, and SD depth functions. It can be seen that both methods show promising performance with the considered depth functions. In Figure 2.6(c) , we compare the power curves of Spearman correlation coefficient test, Quadrant test, and BallCov test (Zhu et al., 2021).

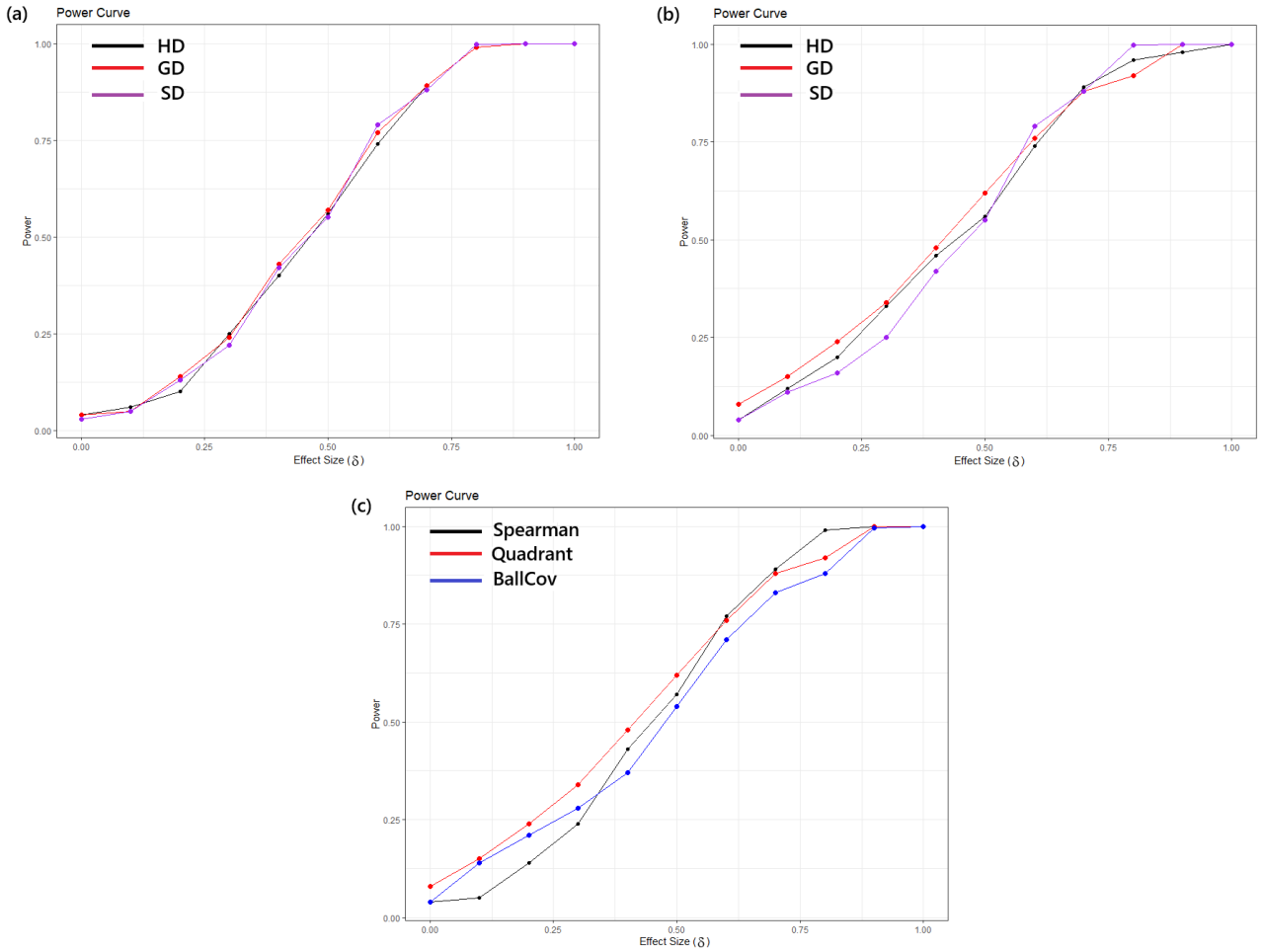


Figure 2.6: Power plots for data from a von Mises distribution, (a) Power curves for Spearman correlation coefficient test with HD, GD, and SD depths, (b) Power curves for Quadrant test with HD, GD, and SD depths, (c) Power curves for Spearman correlation coefficient test with GD, Quadrant with GD, and ball covariance test.

It can be seen that the best performance belongs to Spearman correlation coefficient test, as it has a good balance for small and large effect sizes.

Example 4 (von Mises-Fisher Distributions): Consider the manifold space of unit sphere $\mathcal{S}^{3-1} = \{X \in \mathbb{R}^3 \mid \|X\| = 1\}$ following von Mises-Fisher distribution $VMF(\mu, \kappa)$ with density

$$f(x, \mu, \kappa) = C(x) \exp \kappa \mu^\top x \mathbb{I}_{\mathcal{S}^{3-1}}(x) \ ,$$

Table 2.3: Independence test simulation results for von Mises-Fisher distribution. Reported values are the power of the test statistics.

		Spearman Test				Quadrant Test			
		$\delta = 0$	$\delta = 0.1$	$\delta = 0.8$	$\delta = 1$	$\delta = 0$	$\delta = 0.1$	$\delta = 0.8$	$\delta = 1$
$n = 20$	HD	0.04	0.07	0.62	0.98	0.06	0.18	0.86	1
$n = 50$	HD	0.02	0.06	0.98	1	0.03	0.18	0.96	1
$n = 100$	HD	0.06	0.08	0.97	0.99	0.04	0.13	1	1
$n = 500$	HD	0.03	0.07	1	1	0.02	0.06	1	1
$n = 20$	GD	0.01	0.06	0.89	1	0.02	0.12	0.87	1
$n = 50$	GD	0.05	0.04	1	1	0.07	0.08	1	1
$n = 100$	GD	0.06	0.07	1	1	0.05	0.13	1	1
$n = 500$	GD	0.04	0.08	1	1	0.02	0.15	1	1
$n = 20$	SD	0.02	0.05	0.89	1	0.03	0.06	0.76	1
$n = 50$	SD	0.01	0.06	0.87	1	0.08	0.19	0.85	1
$n = 100$	SD	0.02	0.07	1	1	0.01	0.10	1	1
$n = 500$	SD	0.04	0.06	1	1	0.04	0.03	1	1

where $\kappa \geq 0$ and $\mu \in \mathcal{S}^{3-1}$ are known as concentration and directional median parameters, and $C(x)$ is a normalization constant. Consider

$$X_i^* \sim VMF\big((1,0,0),10\big), \quad Y_i^* \sim VMF\big((0,0,1),50\big), \quad Z_i \sim VMF\big((0,1,0),20\big) \ ,$$

and

$$\begin{aligned} X_i &= \exp_\mu \left((1-\delta)\log_\mu(X_i^*) + \delta \log_\mu(Z_i)\right) \ , \\ Y_i &= \exp_\mu \left((1-\delta)\log_\mu(Y_i^*) + \delta \log_\mu(Z_i)\right) \ . \end{aligned}$$

Our interest is to test the null hypothesis

$$H_0 : X \perp Y \quad v.s. \quad H_A : X \not\perp Y \ .$$

We can use HD, GD, and SD together with the Spearman correlation coefficient test and Quadrant test to test the independence of X and Y. The simulation results for different values of δ over 5000 simulation runs are reported in Table 2.3. As can be seen, Spearman test shows promising performance for both null and alternative scenarios and is not affected by the sample size. Quadrant test is also effective method for test of independence. Both methods improve power as δ approaches 1. It should be also noted for small sample sizes Quadrant test can have large type-I error; however, it keeps the empirical size below the significance level $\alpha = 0.05$ as the sample size increases.

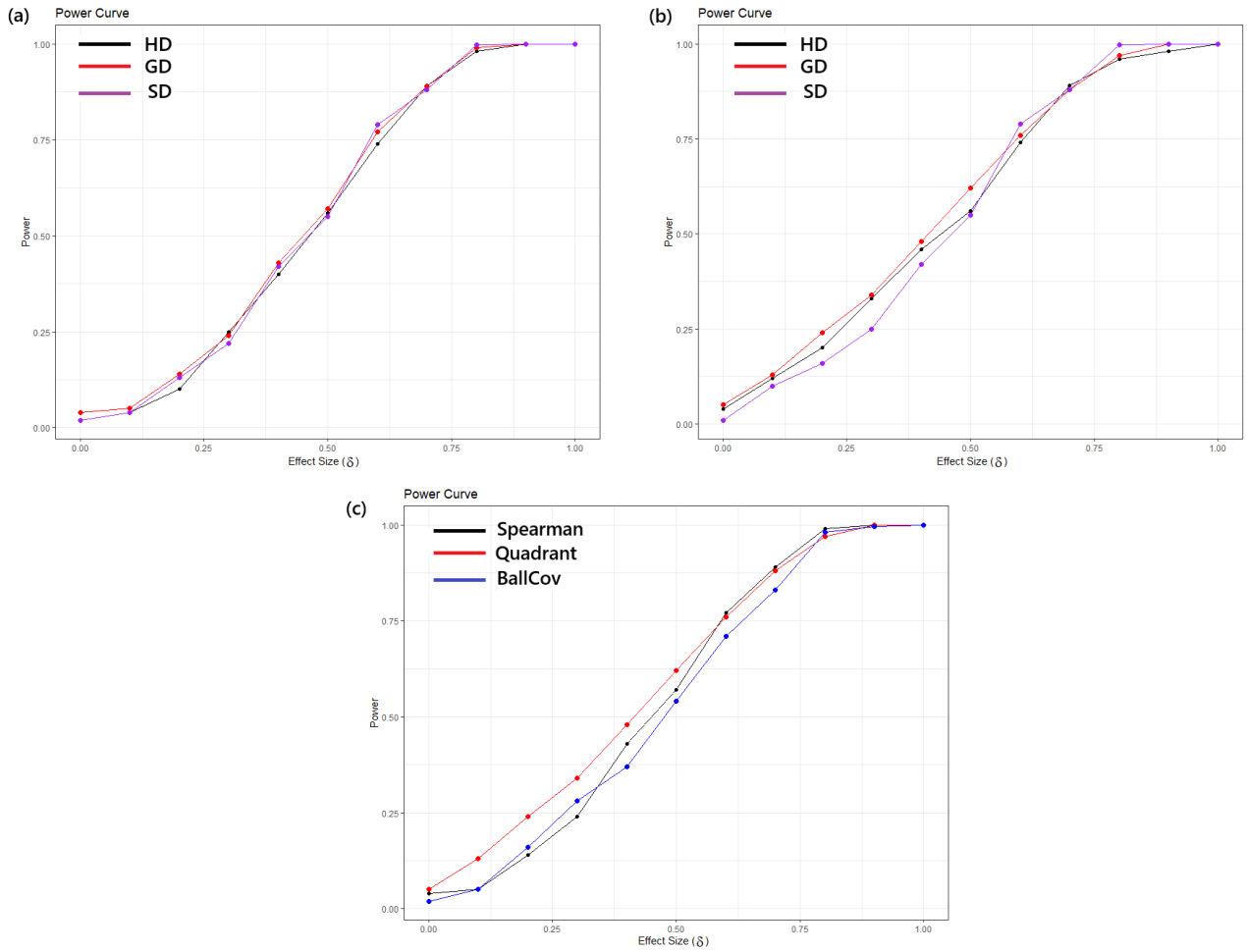


Figure 2.7: Power plots for von Mises Fisher distribution data, (a) Power curves for Spearman correlation coefficient test with HD, GD, and SD depths, (b) Power curves for Quadrant test with HD, GD, and SD depths, (c) Power curves for Spearman correlation coefficient test with GD, Quadrant with GD, and ball covariance test.

Figure 2.7(a) compares the power curves for Spearman correlation coefficient test with HD, GD, and SD depth functions. It can be seen that the Spearman test performs well with the considered depth functions. Figure 2.7(b) compares the power curves for Quadrant test with HD, GD, and SD depth functions. It can be seen that Quadrant test show promising performance with the considered depth functions. In Figure 2.7(c), we evaluate the power curves of Spearman correlation coefficient test, Quadrant test, and BallCov test.

It can be seen that Quadrant test is sensitive to small effect sizes and cannot effectively control empirical size under null hypothesis. Both Spearman correlation coefficient test and BallCov tests can control the empirical size under the null hypothesis; however, the power of Spearman test approaches the nominal value 1 faster than BallCov test. It can be concluded that using depth functions result in a more effective nonparametric test compared to BallCov which is based on the use of distance function.

Example 5 (Community Network Distributions): Stochastic block models (SBMs) can be used for modeling heterogeneous communication networks with latent connection structure (Lee and Wilkinson, 2019). In this example, the interest is to find out whether there is a dependence between the distributions generating communication networks for two groups of societies. Suppose that our interest is to determine whether there is a difference in the facebook communication pattern among students at a college in the first two months of the academic year. We choose students from three different classes for the experiment. We can use SBM networks to simulate these communication patterns. Let $d \in \mathbb{Z}$ denote the number of vertices (the number of students chosen for the experiment) and $k \in \mathbb{Z}$ denote the number of communities (the number of classes considered for the analysis) and $\mathbf{p} = (p_1, p_2, ..., p_k)$ be a probability vector on $[k] = \{1, 2, ..., k\}$ and W be a $k \times k$ symmetric matrix known as connectivity probability. We can generate (X, G) under SBM$(d, \mathbf{p}, W)$ where X is a d-dimensional random vector identifying the label of community group according to probability model $\mathbf{p}$, and G is a d-vertex simple graph with vertices i and j connected with probability W_{X_i, X_j}. The community sets are defined by $\Omega_i = \Omega_i(X) = \{\nu \in [d] \mid X_\nu = i\}$, for $i \in [k]$.

Suppose we generate n_1 communication networks according to the above model representing the facebook communication of the d chosen students from k classes for the first two months of academic year. We consider $\mathbf{p} = (0.3, 0.5, 0.2)$ and

$$W = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.3 & 0.5 & 0.2 \\ 0.6 & 0.2 & 0.2 \end{bmatrix}.$$

For the second set of samples which corresponds to n_2 communication network representing facebook communication at M days of the last two months of the academic year, we generate independently considering $\mathbf{p} = (0.1, 0.3, 0.6)$ and

$$W = \begin{bmatrix} 0.1 & 0.7 & 0.2 \\ 0.7 & 0.2 & 0.1 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}.$$

We choose $d = \{20, 50, 100\}$ students and $n_1 = n_2 = \{20, 50, 100\}$ communication network graphs to test the null hypothesis of independence of the distributions of communication network distributions for the first and last two months.

If the distributions are dependent, then it implies that students have inclination towards keeping the friendship pattern they developed during the early weeks of the academic year for the rest of the academic year. We can simulate such patterns considering layered multiplex SBM networks. For the second scenario, we generate dependent graphs using multiplex SBM with two layers considering $\mathbf{p} = (0.1, 0.3, 0.6)$ and the connectivity probabilities

$$P_{00} = \begin{bmatrix} 0.1 & 0.1 & 0.8 \\ 0.1 & 0.8 & 0.1 \\ 0.8 & 0.1 & 0.1 \end{bmatrix} .$$

$$P_{10} = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.3 & 0.5 & 0.2 \\ 0.6 & 0.2 & 0.2 \end{bmatrix} .$$

$$P_{01} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

$$P_{11} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

It is well-known that undirected graphs $G(\mathcal{V}, \mathcal{E})$ can be realized by their Laplacian, which is defined as a matrix $L = D - W$ with W denoting the $d \times d$ adjacency matrix of graph G with $\mid V \mid = d$, and D is a diagonal matrix with the vertex degrees along the diagonal. Apparently, the degree matrix

$$D = \text{diag}\big(\sum_{j=1}^{d} w_{1,j}, \sum_{j=1}^{d} w_{2,j}, ..., \sum_{j=1}^{d} w_{d,j}\big) ,$$

contains information about the importance of each node, which is a partial information about the network. Also, the adjacency matrix contains all the information about the

network; however, information about the importance of each node is indirectly available. The Laplacian matrix L contains both of these information

$$L = D - W = \begin{cases} -w_{ij} & \text{if } i \neq j \\ \sum_{k \neq i} w_{ik} & \text{if } i = j \end{cases}$$

for $1 \leq i, j \leq d$.

Therefore, the degree information is directly available on diagonal elements and adjacency information as negative off-diagonal elements. Moreover, spectral analysis of L can give important information about the main characteristics of the network, such as spectral gap, the smallest non-zero eigenvalue of L, and the sparsest cut of a graph, which minimizes the ratio of the number of edges across the cut and the number of vertices in the smaller part of the graph. Also, when there is no self-loops and no multi-edges, the correspondence between G and L are one-to-one. Therefore, the analysis can be done by focusing on the space of Laplacians

$$\mathcal{L}_d = \{L = (l_{ij}) : L = L^\top;\ l_{ij} \leq 0,\ \forall i \neq j;\ L\mathbf{1}_{d,1} = \mathbf{0}_{d,1}\}\ .$$

Therefore, we can work with the space of graph Laplacians $\mathcal{L}_d$, which is a submanifold of Euclidean space $\mathbb{R}^{d^2}$ with intrinsic dimension $\frac{d(d+1)}{2}$ (Ginestet et al., 2017). Laplacian matrices are PSD, and the corresponding metric space is $(\mathcal{L}_d, \delta_F)$ with δ_F being the Frobenius norm

$$\delta_F = \|X - Y\|_F^2 = \sum_{i=1}^{d} \sum_{j=1}^{d} (x_{ij} - y_{ij})^2\ ,$$

for any pair of matrices $X, Y \in \mathcal{L}_d$. Therefore, we can use all properties of the manifold space

$$SPD(d) = \{A \in \mathbb{R}^{d \times d} \mid A = A^\top,\ \text{and } x^\top A x > 0\}\ .$$

We consider HD, GD, and SD together with the Spearman correlation coefficient test and Quadrant test to test the independence of two random objects X and Y representing the distribution of facebook communication networks in the first and last two months of the academic year. We choose $d = \{20, 50, 100\}$ students and $n_1 = n_2 = \{20, 50, 100\}$ communication network graphs to test the null hypothesis.

Figure 2.8 and Figure 2.9 indicate connectivity matrices for a sample of three generated SBM networks for the first and last two months of the year. It is apparent that the connectivity structure for the two sets of SBM networks is entirely different.

Table 2.4: Simulation results for community network distribution, Case I. Reported values are the size of the test statistics.

		Spearman Test			Quadrant Test		
		$d = 20$	$d = 50$	$d = 100$	$d = 20$	$d = 50$	$d = 100$
$n = 20$	HD	0.02	0.03	0.06	0.08	0.03	0.04
$n = 50$	HD	0.02	0.01	0.05	0.01	0.04	0.03
$n = 100$	HD	0.01	0.04	0.02	0.03	0.03	0.06
$n = 500$	HD	0.04	0.02	0.05	0.02	0.01	0.03
$n = 20$	GD	0.05	0.04	0.06	0.07	0.05	0.08
$n = 50$	GD	0.04	0.03	0.04	0.08	0.03	0.06
$n = 100$	GD	0.04	0.03	0.05	0.02	0.04	0.06
$n = 500$	GD	0.03	0.05	0.04	0.05	0.04	0.05
$n = 20$	SD	0.05	0.04	0.08	0.03	0.04	0.07
$n = 50$	SD	0.06	0.04	0.07	0.06	0.04	0.06
$n = 100$	SD	0.05	0.05	0.04	0.02	0.05	0.03
$n = 500$	SD	0.04	0.05	0.03	0.04	0.02	0.05

Table 2.5: Simulation results for community network distribution, Case II. Reported values are the power of the test statistics.

		Spearman Test			Quadrant Test		
		$d = 20$	$d = 50$	$d = 100$	$d = 20$	$d = 50$	$d = 100$
$n = 20$	HD	0.32	0.72	0.76	0.56	0.81	0.76
$n = 50$	HD	0.62	0.90	0.92	0.64	0.95	0.90
$n = 100$	HD	0.90	1	1	0.83	1	1
$n = 500$	HD	0.95	1	1	0.89	1	1
$n = 20$	GD	0.44	0.79	0.66	0.63	0.84	0.89
$n = 50$	GD	0.56	0.86	0.88	0.77	0.94	0.95
$n = 100$	GD	0.87	1	1	1	1	1
$n = 500$	GD	0.90	1	1	1	1	1
$n = 20$	SD	0.77	0.85	0.86	0.79	0.86	0.95
$n = 50$	SD	0.92	0.95	0.94	0.75	0.87	0.95
$n = 100$	SD	1	1	1	1	1	1
$n = 500$	SD	1	1	1	1	1	1



Figure 2.8: Schematic illustration of communication networks for the first two months.



Figure 2.9: Schematic illustration of communication networks for the last two months.

The simulation results for the first scenario are reported in Table 2.4. The Spearman correlation coefficient test and Quadrant test retain their acceptable empirical size regardless of the dimensionality of the manifold and the sample size. The false rejection rates is less than or equal 0.05 for the $\alpha = 0.05$ significant level.

The obtained results endorse the capability of the formulated test statistic given the fact that the method shows promising results when the manifold has a high dimension $D = \frac{100(101)}{2} = 5050$ and sample size is small $n = 20$. It is inferred that both Spearman correlation coefficient test and Quadrant test yield promising empirical size at the $\alpha = 0.05$ significance level, and therefore, are practical choices for high dimensional manifolds.

For the second scenario, the simulation results are reported in Table 2.5. It can be seen that the performance of Spearman correlation coefficient test improves as the sample size gets larger. The power of the test statistic is not acceptable for small sample sizes, e.g. $n = 20$, especially when dimension is lower. However, it is improved for larger sample sizes, e.g. $n = 100$. The same behaviour can be observed for Quadrant test, however, it cannot improve its performance as fast as Spearman test for smaller dimensions.

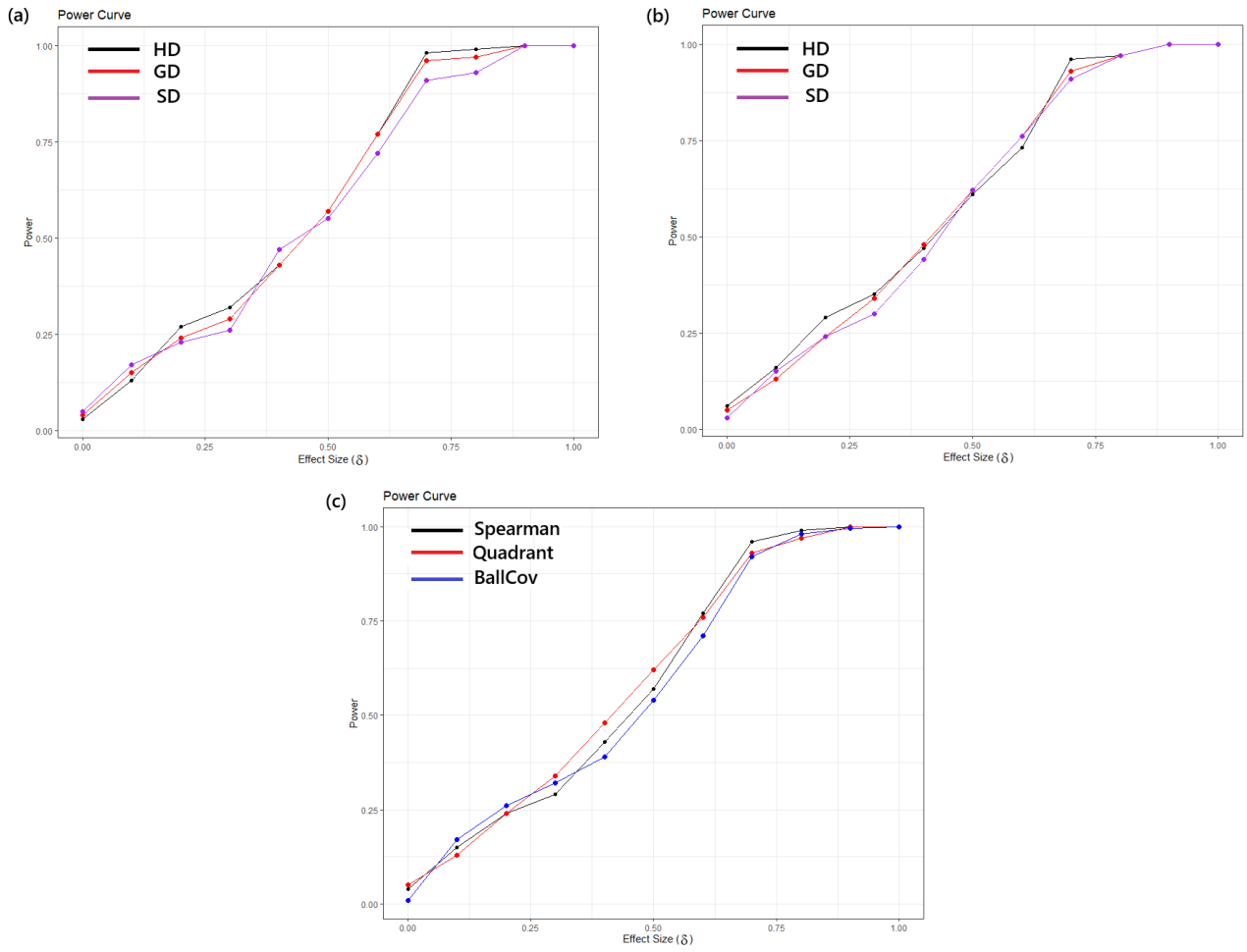


Figure 2.10: Power plots for community network data, (a) Power curves for Spearman correlation coefficient test with HD, GD, and SD depths, (b) Power curves for Quadrant test with HD, GD, and SD depths, (c) Power curves for Spearman correlation coefficient test with GD, Quadrant with GD, and ball covariance test.

Figure 2.10(a) compares the power curves for Spearman correlation coefficient test with HD, GD, and SD depth functions. It can be realized that the Spearman test performs effectively with all of the considered depth functions. Figure 2.10(b) compares the power curves for Quadrant test with HD, GD, and SD depth functions. It can be seen that Quadrant test also show promising performance with the considered depth functions. In Figure 2.10(c), we compare the power curves of Spearman correlation coefficient test,

Quadrant test, and BallCov test. All of the considered methods, i.e. Spearman correlation coefficient test, Quadrant test, and BallCov tests, can control the empirical size under the null hypothesis; however, the power of Spearman correlation coefficient test and Quadrant test approaches the nominal value 1 faster than BallCov test, endorsing the efficacy of using depth functions for developing rank-based nonparametric methods.

2.3.2 Hypothesis Testing for Quality Monitoring

Example 6 (Phase-II Process Monitoring): Consider the process monitoring problem where the interest is to monitor the quality of the shape of products. We can use k landmark points in plane to represent the shape of products, i.e. the well-known Kendall's two-dimensional shape space Σ_2^k with the induced Procrustes distance for the extrinsic analysis (Bhattacharya, 2008).

We collect m of these products from a batch and their distributions are compared to the distributions of n healthy sample products collected in the phase-I monitoring stage. Suppose we collected $(X_1, ..., X_n)$ healthy samples from the manifold space $\mathcal{M}$ representing the landmark sets of product shapes. In the production process $(Y_1, ..., Y_m)$ samples are collected from a batch of quality monitoring. This can be formulated using the Wilcoxon statistic based on the rank of the manifold data. The n_0 samples $(Z_1, ..., Z_{n_0})$ are drawn from the pool of healthy samples available so far. The empirical measure $\widetilde{P}_{n_0}$ is used as reference sample.

The Wilcoxon test statistic $W = \sum_{j=1}^{m} R(\Lambda(\widetilde{P}_{n_0}, Y_j))$ is calculated and the null hypothesis for the equality of the distributions of high quality data X and in process collected data Y is tested using asymptotic distribution $\mathrm{N}(\frac{m(n+m+1)}{2}, \frac{mn(n+m+1)}{12})$ with $\alpha = 0.01$, which is a commonly used significance level for process monitoring.

The schematic illustration of healthy and faulty products along with the contour line of a set of products are shown in Figure 2.11, Figure 2.12, Figure 2.13, and Figure 2.14.

Fisher's permutation is used to calculate the p-value of the test statistic, which is 0.001 and therefore we reject the null hypothesis of the equality of the distribution of the healthy and inline produced products at $\alpha = 0.01$ significant level.

It is apparent that the test statistic correctly identified the produced batch of products as outlier with respect to healthy Phase-I collected products, as the products have significant deviations as well as deformation faults on their surfaces, compared to healthy heart surfaces.

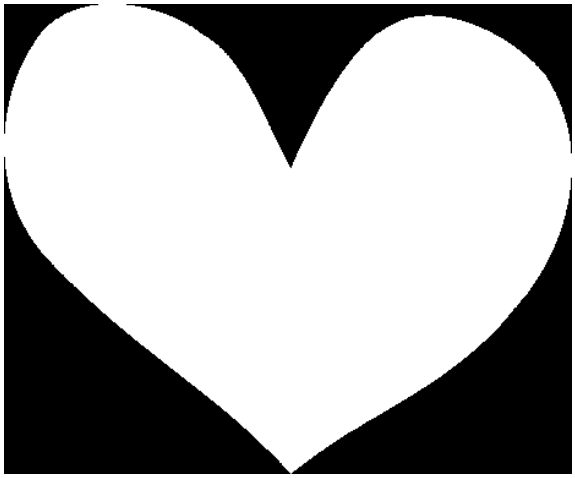


Figure 2.11: Schematic illustration of a healthy product.

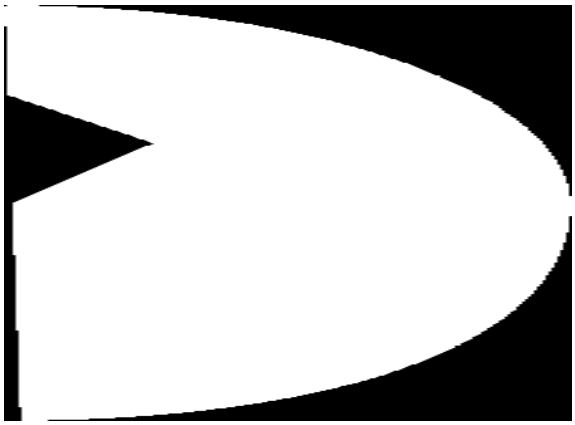


Figure 2.12: Schematic illustration of a faulty product.

Example 7 (DNA Molecule Data): In this example, we evaluate the depth-rank Wilcoxon test for studying DNA damage data (Dryden et al., 2017; Marron and Dryden, 2021). The dataset is a TFC molecule which is a type of damaged DNA molecule, and the observations are temporally thinned to 50^{th} observation (Marron and Dryden, 2021). The dataset has $n = 50$ samples with $k = 22$ landmarks in $d = 3$ dimensions. The goal of the analysis is to detect significant difference in the mean and variability of the size and shape of point sets evolving over time. We model the manifold data using k-ads shape space in 3 dimensions Σ_3^k, represented by the set of ordered k points on a 3 dimensional image $\{(x_j, y_j, z_j), j = 1, \ldots, k\}$ with $k = 22$. After rescaling the data set to

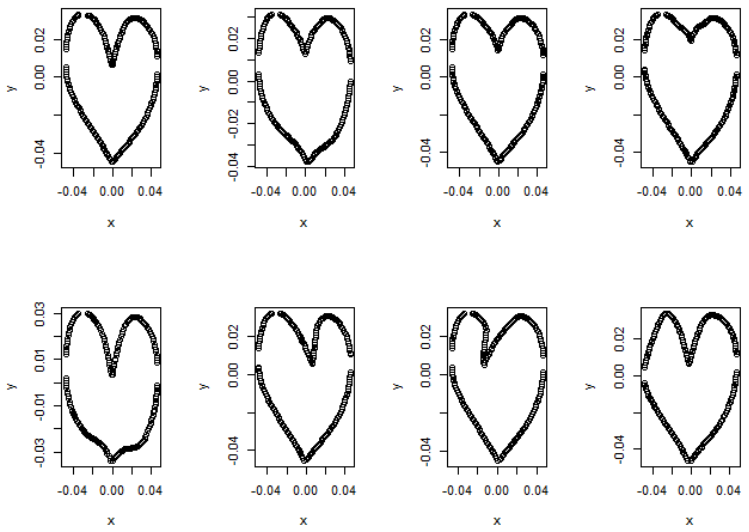


Figure 2.13: Schematic illustration of the contour line of healthy products.

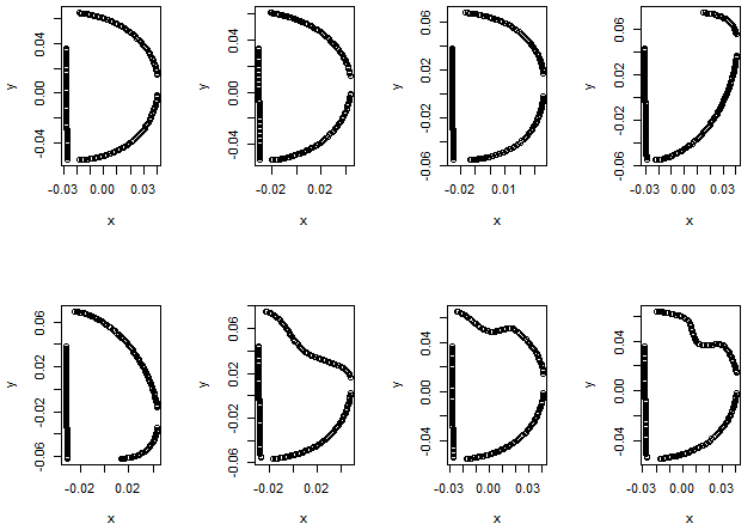


Figure 2.14: Schematic illustration of the contour line of faulty products.

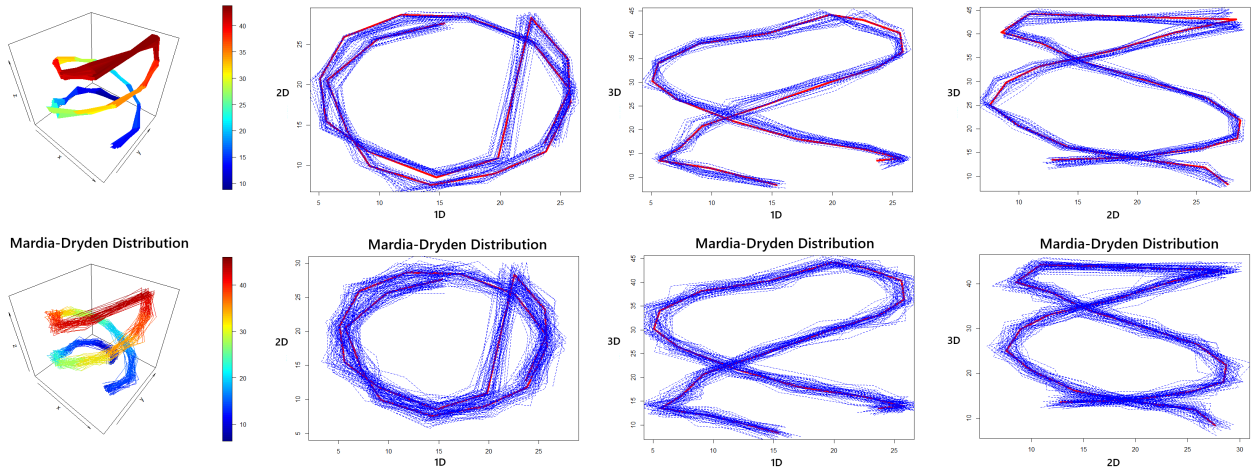


Figure 2.15: Schematic illustration of DNA data, top panels indicate 3D view of a sample of real world healthy DNA molecule structure data with 2D projections, bottom panels indicate 3D view of damaged DNA molecule with deformed structures as well as their 2D projections generated from Mardia-Dryden distribution.

have Euclidean norm 1, the dataset can be represented on $\mathcal{S}^{3k-4}$ in $\mathbb{R}^{3k-3}$, for which we can perform intrinsic analysis using the underlying geometry of $\mathcal{S}^{3k-4}$.

Figure 2.15 top panel indicates the schematic illustration of DNA molecule structure data in 3D as well as their 2D projections. The Fréchet mean is shown in red color in corresponding 2D projections. We generate damaged DNA molecule structure data of size $n = 50$ using Mardia-Dryden distribution (Dryden and Mardia, 2016) with variance $\sigma^2 = 0.1$ and expected shape being the Frećhet mean from the first sample set.

Figure 2.15 bottom panel indicates the generated data in 3D as well as their corresponding 2D projectoins. As can be seen, the dataset include deformed DNA molecule structures. The proposed rank-depth Wilcoxon test is used to detect instability in the DNA molecule structures using the reference sample extracted from (Marron and Dryden, 2021).

Figure 2.16(a) indicates the geodesic depth ranks of the reference sample as well as the phase-II sample. As can be seen, The difference in the distribution of reference and Dryden-Mardia generated DNA samples are detected by GD. Figure 2.16(b) indicates a 3D view of a healthy DNA molecule structure. Rank-depth Wilcoxon test successfully detected the change in the distribution. Figure 2.16(c) indicates the set of 3 damaged DNA molecule structures with lowest geodesic depth.

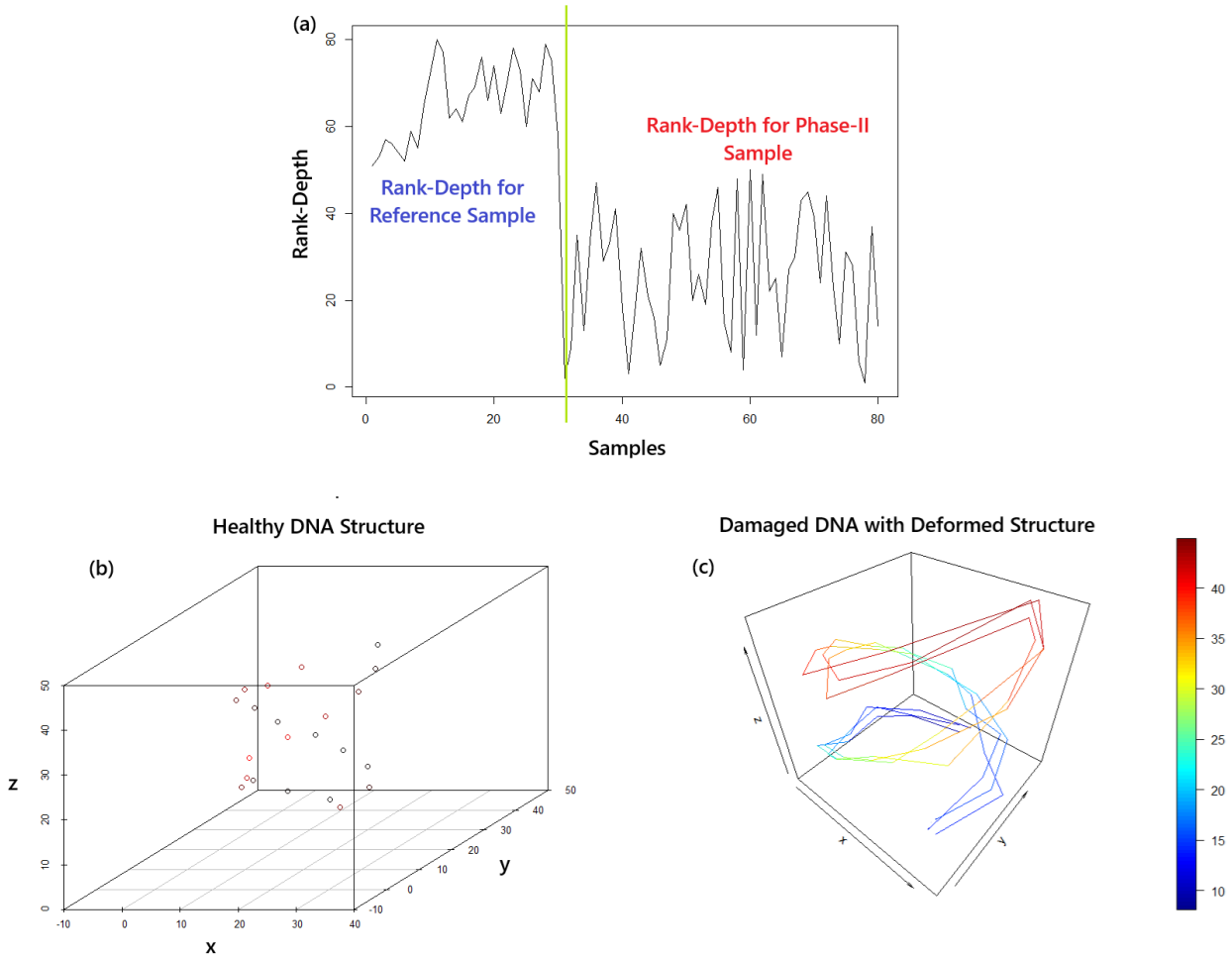


Figure 2.16: (a) Geodesic-based Rank-Depth for reference sample, (b) 3D view of healthy DNA molecule structure, (c) 3D view of damaged DNA molecule structures with lowest geodesic depth.

Figure 2.17 compares the power plot for DNA molecule data with power curves for Wilcoxon test with GD, energy test (Chu and Dai, 2024), GET (Chen and Friedman, 2017), and geodesic T^2 test (Alekseyenko, 2016). As it can be inferred, the distance-based tests show large type-I error and are sensitive to small effect sizes. Edge count test GET and geodesic rank-depth Wilcoxon show superior performance and their power grows to 1 as the effect size increases.

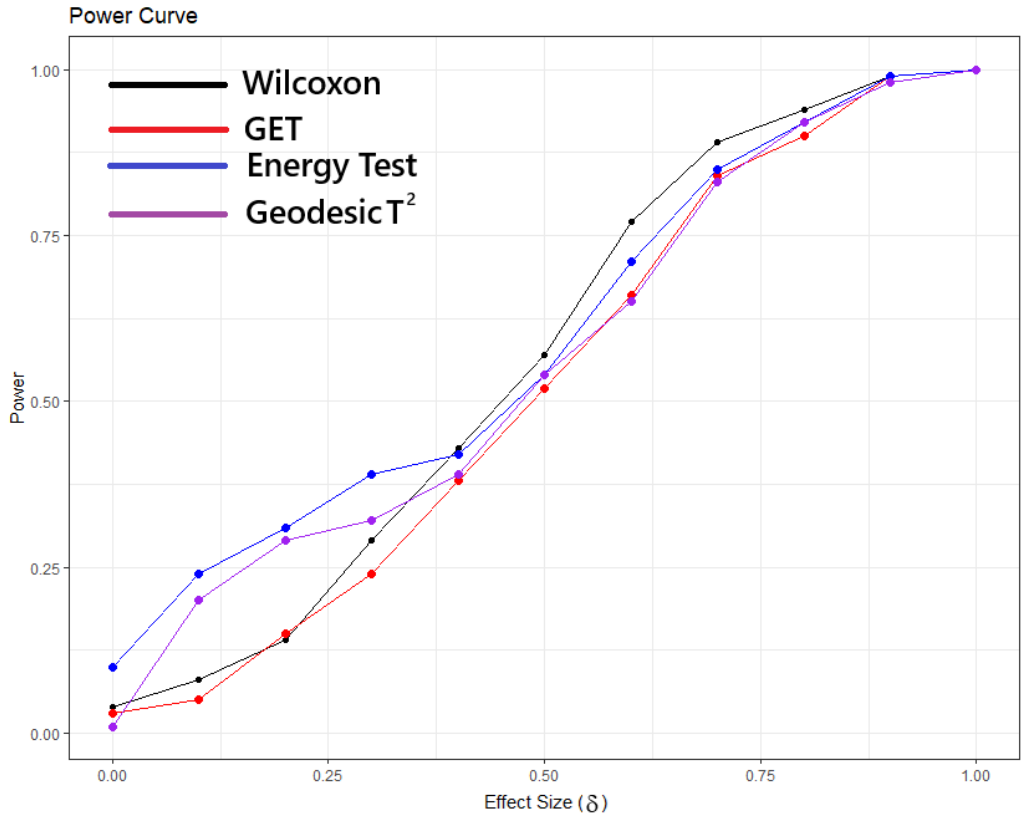


Figure 2.17: Power plot for DNA molecule data with power curves for Wilcoxon test with GD, energy test, GET, and geodesic T^2 test.

Among the rival methods, geodesic rank-depth Wilcoxon test outperforms the other methods as it has higher power for moderate to large effect sizes and its power grows faster to 1 as the effect size increases.

Example 8 (Apicomplexa Phylogenetic Tree Data): In evolutionary biology, tree structure is used to model the ancestral relationship among a fixed collection of species (Kuo et al., 2008). Such tree structures are known as phylogenetic trees. Each leaf represents a species, each interior node a speciation event, each edge transition from an ancestral to a descendant, and the length of each edge indicates the evolutionary divergence. For statistical analysis, a sample of phylogenetic trees are considered, where each individual tree is constructed from the sequence of a specific gene presented in the species in question. Biological variation and sequence misalignment are the main reasons for the source of variation in such samples of phylogenetic trees. Tree space geometry aware methods have

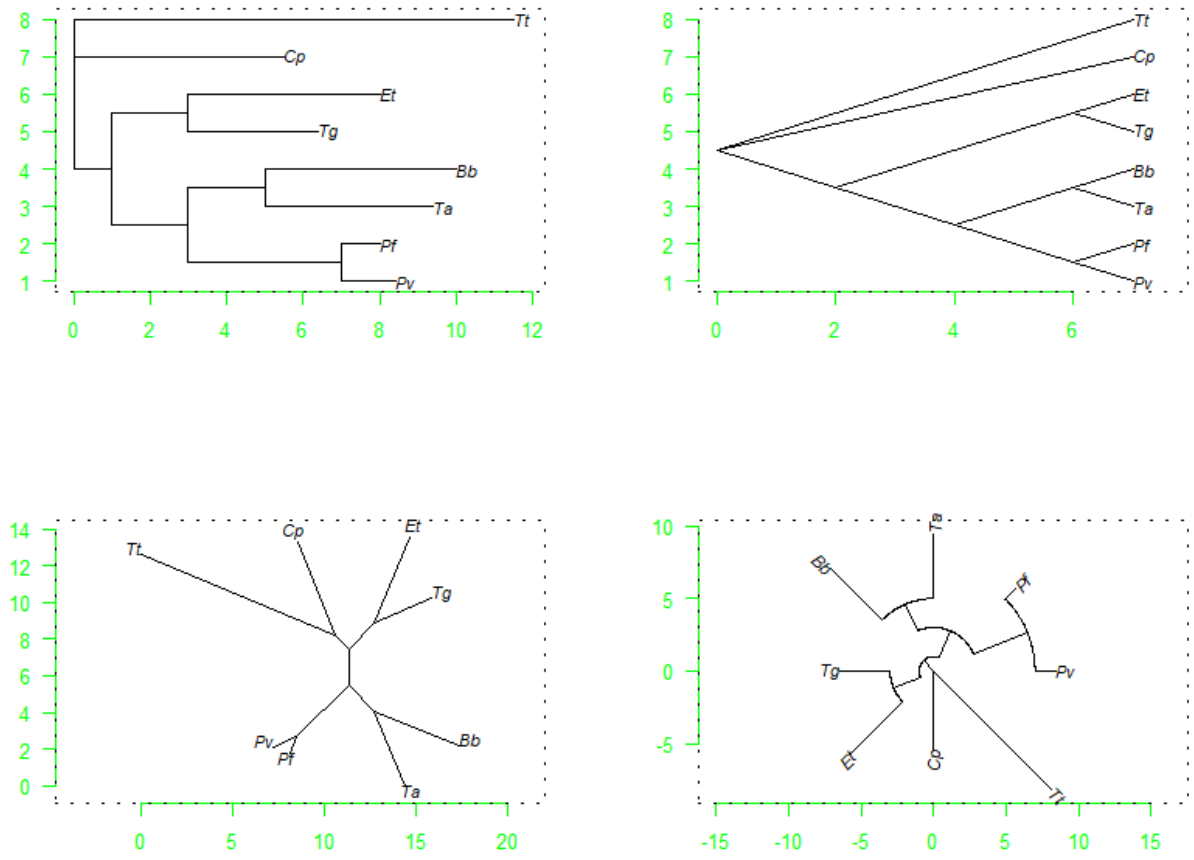


Figure 2.18: The deepest phylogenetic tree with respect to geodesic depth.

been proven to be effective tools for statistical analysis of phylogenetic data (Weyenberg et al., 2014; Dai et al., 2022). Tree data are non-Riemannian due to singularity at origin, i.e. start-shape tree structure at origin has degenerate interior edges.

In this example, we further investigate the potential of statistical methods by implementing rank-depth Wilcoxon test with respect to the underlying geometry of tree structure for monitoring the stability of a sample of phylogenetic data considering a reference sample. The phylogenetic trees contain 7 pathogenic Apicomplexan species related to a sample of

species of 268 individual gene tree structures originally introduced by (Kuo et al., 2008). The Apicomplexa phylum contains various types of detrimental pathogenic parasites. In (Kuo et al., 2008), the Apicomplexan species entail malaria pathogens Plasmodium falciparum (Pf), Plasmodium vivax (Pv), tick-borne haemopathogens Babesia bovis (Bb), Theilera annulata (Ta), coccidian parasites Eimeria tenella (Et), Toxoplasma gondii (Tg), and Cryptosporidium parvum (Cp), which infect intestines. The outgroup Tetrahymena thermophila (Tt) is a remotely related model species which is considered to root the phylogeny. In this model, the root is the point joining the outgroup with the Apicomplexan species.

We use the product space $\mathcal{T}_8 \times \mathbb{R}^8$ to model the evolutionary divergence between species and their ancestors, where the gene tree structure is modeled by $\mathcal{T}_8$ and the pendant edge lengths are modeled by $\mathbb{R}^8$. To extract the phylogenetic tree that best represents the sample, we apply geodesic depth function with generalized Robinson-Foulds distance metric (Robinson and Foulds, 1981).

Figure 2.18 depicts the schematic illustration of the deepest phylogenetic structure calculated from dataset (Kuo et al., 2008). The suggested structure is in agreement with the findings reported in (Dai et al., 2022; Kuo et al., 2008; Nye et al., 2017). It can be seen that Cp is the deepest rooted species. Moreover, tick parasites Bb, Ta, Pf, and Pv share the same immediate ancestors, i.e. they are monophyletic, with Coccidian parasites Et and Tg forming their sister group. For the simulation, we collect $n = 50$ reference samples and generate $m = 50$ samples from the collection of gene tree structures with an outlying distribution with respect to that of the reference sample. Therefore, we expect that rank-depth Wilcoxon test detect this change of distribution, i.e. instability in the gene tree structure of the generated sample.

Figure 2.19 depicts four individual phylogenetic trees with the least geodesic depth value generated from phase-II distribution. As seen, these gene tree structures are outliers comparing to the deepest tree, and differ from the findings reported in the literature (Dai et al., 2022; Kuo et al., 2008). The simulation results demonstrate the effectiveness of the rank-depth Wilcoxon test for monitoring the outlyingness/instability of gene tree structures and for detecting phylogenetic trees with uncommon features.

Figure 2.20 indicates the power plot for phylogenetic tree data with power curves of Wilcoxon test with GD, energy test, GET, and geodesic T^2 test. The effect size δ represents the amount of shift/increase in variability we have when sampling the length of pendant edges. Under the null, the edges are distributed around the deepest tree with variability $\sigma^2 = 0.1$ and under the alternative, we have $\delta + \sigma^2$. As it can be inferred, the distance-based tests, i.e. energy test and geodesic T^2 have large type-I error and are affected by

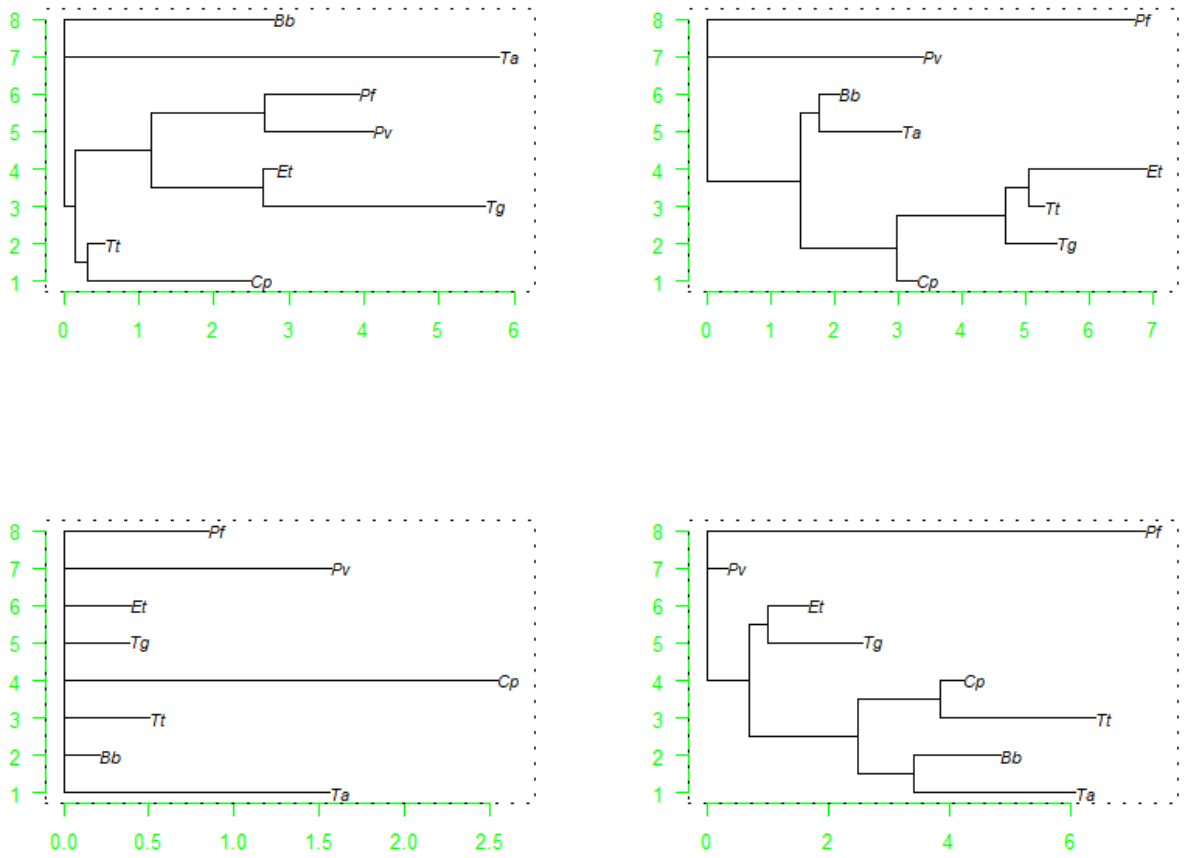


Figure 2.19: Four individual phylogenetic trees with the least geodesic depth value.

small effect sizes. The proposed rank-depth Wilcoxon test shows superior performance with respect to both small and large effect sizes, and its power grows faster to the nominal value 1 as the effect size δ increases.

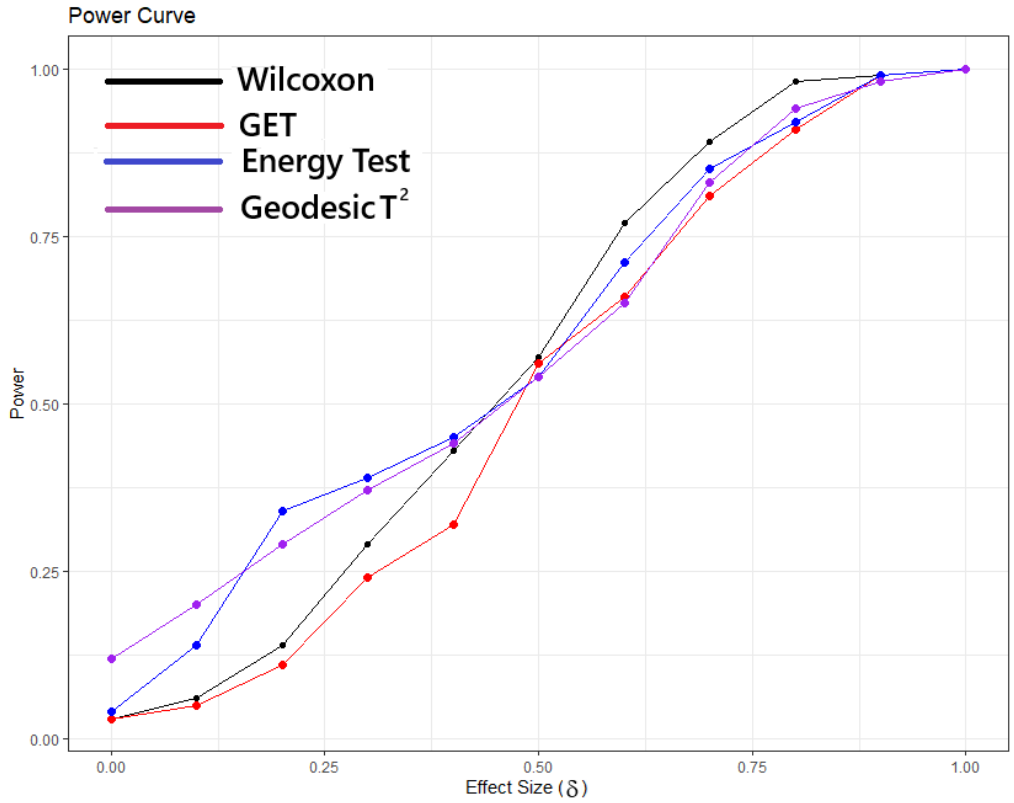


Figure 2.20: Power plot for phylogenetic tree data with power curves for Wilcoxon test with GD, energy test, GET, and geodesic T^2 test.

2.3.3 Multiple Sample Hypothesis Testing of Location and Scale Change

Example 9 (Multiple Group of Covariance Matrices): In this simulation study, we consider several covariance matrix structures to test the equality of the distributions of covariance matrices across groups.

The first hypothesis testing problem is

$$\text{(I) } H_0 : F_{\Sigma_1} = F_{\Sigma_2} = F_{\Sigma_3} \ ,$$

where $\Sigma_i = \Omega\Delta_{i-1}\Omega$, for $i = 1, 2, 3$. $\Omega = \text{diag}(\omega_1, \omega_2, ..., \omega_p)$ with $w \overset{\text{iid}}{\sim} U(1, 5)$, and Δ_i is a $p \times p$ matrix with (a, b)-th element being $(-1)^{a+b}\left(\frac{i+2}{5}\right)^{|a-b|^{1/10}}$.

The second hypothesis testing problem is

$$\text{(II) } H_0 : F_{\Sigma_1} = F_{\Sigma_2} = F_{\Sigma_3} \ ,$$

where $\Sigma_i = U_i \Lambda U_i'$, for $i = 1, 2, 3$, and $\Lambda = \text{diag}(\gamma_1, ..., \gamma_p)$, where $\gamma_1, ..., \gamma_p \overset{\text{iid}}{\sim} \Gamma(4, 0.5)$, $U_i = (Z_i' Z_i)^{-1/2} Z_i'$, and Z_i is a $p \times p$ matrix whose entries are i.i.d. $\text{N}(10, i)$.

We use $n_1 = n_2 = n_3 = \{20, 50, 100\}$ and $p = \{3, 5, 10\}$ to generate the covariance matrix samples for each group. The proposed Kruskal-Wallis test statistic H together with HD, GD, and SD depth functions are used to test the two null hypotheses.

Simulation result for the two scenarios under the null hypothesis is given in Table 2.6. It can be seen that for both cases, the test statistic controls the empirical size to be less than or close to 0.05 at $\alpha = 0.05$ significance level indicating its appropriateness for testing change in distribution on manifold spaces. It is also observed that the empirical size does not get affected as we increase the dimensionality of the covariance matrices. Table 2.7 tabulates the simulation results under the alternative hypothesis. It is obvious that the test statistic has a large empirical power for the considered sample sizes and dimensionalities.

In general, it can be concluded that the test statistic with HD, GD, and SD depth functions can control the type-I error and has a small type-II error, achieving power 1 as the sample size increases. The finite sample Monte Carlo simulation verifies the asymptotic results we obtained for size and power of the Kruskal-Wallis test statistic, as the test statistic performs quite well for sample sizes as large as $n = 100$.

Figure 2.21 compares the power curves of Kruskal-Wallis test, energy test (Chu and Dai, 2024) and ball divergence (BallDiv) (Zhu et al., 2021). For case I, the effect size δ changes the value of Δ_i with $i = 1, 1 + \delta, 1 + 2\delta$. For case II, the effect size δ changes the entries of Z_i with $\text{N}(10, i)$, $\text{N}(10, i + \delta)$, and $\text{N}(10, i + 2\delta)$. It can be seen that the proposed rank-depth Kruskal-Wallis test statistic outperforms energy test and BallDiv test as its power is larger for moderate effect sizes and grows to 1 faster than its rival methods. All in all, it can be concluded that all the methods can hold a good balance between size and power and are effective for multiple sample test of change in distribution.

Example 10 (Difference Among Groups of Social Networks): In this example, our interest is to find out whether there is a difference among the distributions of social networks of students friendship across different cities. The social networks consist of connections via facebook for students at schools in different cities, e.g. Kitchener, Waterloo, and Cambridge. Erdős-Réyni network can be used to model the facebook friendship connections, assuming homogeneous connection structure.

Table 2.6: Simulation results under null hypothesis for multiple groups of covariance matrices.

		Case I			Case II		
		$p = 3$	$p = 5$	$p = 10$	$p = 3$	$p = 5$	$p = 10$
$n = 20$	HD	0.02	0.01	0.01	0.02	0.06	0.08
$n = 50$	HD	0.02	0.03	0.01	0.06	0.02	0.01
$n = 100$	HD	0.01	0.04	0.01	0.01	0.05	0.03
$n = 500$	HD	0.04	0.03	0.05	0.04	0.03	0.02
$n = 20$	GD	0.02	0.04	0.01	0.03	0.04	0.06
$n = 50$	GD	0.03	0.03	0.03	0.01	0.03	0.04
$n = 100$	GD	0.06	0.01	0.04	0.02	0.03	0.05
$n = 500$	GD	0.03	0.02	0.02	0.04	0.02	0.05
$n = 20$	SD	0.01	0.04	0.02	0.04	0.02	0.03
$n = 50$	SD	0.03	0.02	0.05	0.04	0.01	0.02
$n = 100$	SD	0.05	0.01	0.01	0.03	0.02	0.03
$n = 500$	SD	0.03	0.02	0.02	0.01	0.05	0.04

Table 2.7: Simulation results under alternative hypothesis for multiple groups of covariance matrices.

		Case I			Case II		
		$p = 3$	$p = 5$	$p = 10$	$p = 3$	$p = 5$	$p = 10$
$n = 20$	HD	0.97	0.99	1	0.87	0.98	1
$n = 50$	HD	0.98	1	1	1	1	1
$n = 100$	HD	1	1	1	1	1	1
$n = 500$	HD	1	1	1	1	1	1
$n = 20$	GD	1	1	1	0.79	0.95	1
$n = 50$	GD	1	1	1	1	1	1
$n = 100$	GD	1	1	1	1	1	1
$n = 500$	GD	1	1	1	1	1	1
$n = 20$	SD	0.98	1	1	0.76	1	1
$n = 50$	SD	1	1	1	1	1	1
$n = 100$	SD	1	1	1	1	1	1
$n = 500$	SD	1	1	1	1	1	1

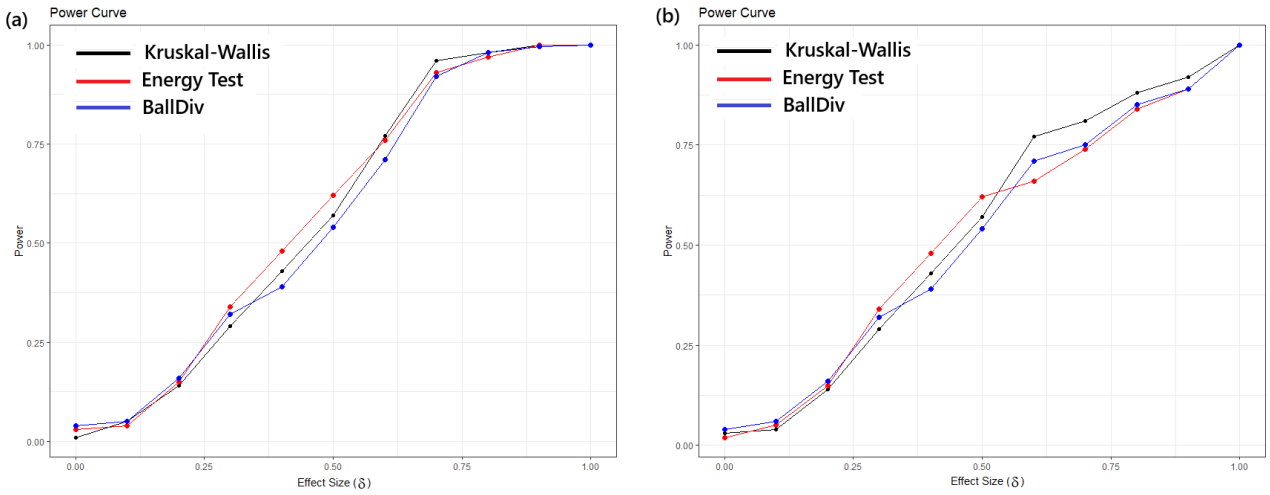


Figure 2.21: Power plots for covariance matrices data considering Kruskal-Wallis test with GD, energy test, and ball divergence test, (a) case I, (b) case II.

For the first scenario, suppose that the friendship network ERG(n, p) for these three cities have the following degree distributions

$$\text{(I):} \begin{cases} \text{Kitchener:} & P[(i,j) \in \mathcal{E}_1] = p_1 = 0.2 \quad \text{for} \quad 1 \leq i < j \leq d \\ \text{Waterloo:} & P[(i,j) \in \mathcal{E}_2] = p_2 = 0.6 \quad \text{for} \quad 1 \leq i < j \leq d \\ \text{Cambridge:} & P[(i,j) \in \mathcal{E}_3] = p_3 = 0.9 \quad \text{for} \quad 1 \leq i < j \leq d \end{cases}$$

where $\mathcal{E}_1$, $\mathcal{E}_2$, and $\mathcal{E}_3$ are the vertices set for Kitchener, Waterloo, and Cambridge, respectively. Apparently, the connection of any edges follows Bernoulli distribution. Suppose that we observe n_1, n_2, and n_3 of such networks from schools located at Kitchener, Waterloo and Cambridge, respectively, and our interest is to test the hypothesis

$$H_0 : F_{G_1} = F_{G_2} = F_{G_3} \ .$$

Given that we have undirected graphs $G(\mathcal{V}, \mathcal{E})$, we can use their Laplacians, $L = D - W$ for the analysis. Using the properties of $SPD(d)$ manifold on which L lies, and considering Frobenius distance, we can use HD, GD, and SD depth functions together with the Kruskal-Wallis test statistic given in Eq. 2.1.

For the second scenario, we consider three models with the same degree distribution

$$\text{(II):} \ P[(i,j) \in \mathcal{E}_k] = p_k = 0.5 \ \text{ for } 1 \leq i < j \leq d \ ,$$

Table 2.8: Simulation results for social network data.

		Case I			Case II		
		$d = 3$	$d = 5$	$d = 10$	$d = 3$	$d = 5$	$d = 10$
$n = 20$	HD	0.85	0.98	1	0.04	0.06	0.08
$n = 50$	HD	0.99	0.98	1	0.07	0.09	0.07
$n = 100$	HD	1	1	1	0.06	0.04	0.07
$n = 500$	HD	1	1	1	0.03	0.05	0.04
$n = 20$	GD	0.92	1	1	0.10	0.13	0.11
$n = 50$	GD	0.96	1	1	0.08	0.11	0.04
$n = 100$	GD	1	1	1	0.06	0.03	0.07
$n = 500$	GD	1	1	1	0.01	0.04	0.01
$n = 20$	SD	0.97	1	1	0.06	0.03	0.04
$n = 50$	SD	0.98	1	1	0.03	0.05	0.05
$n = 100$	SD	1	1	1	0.05	0.01	0.03
$n = 500$	SD	1	1	1	0.04	0.05	0.03

for $k = 1, 2, 3$.

For both scenarios, we collect $n_1 = n_2 = n_3 = \{20, 50, 100, 500\}$ data for each test and let the number of vertices $\mid V \mid = d = \{3, 5, 10\}$ to evaluate the performance of test statistic H for different network complexities.

Table 2.8 indicates the simulation results under the null and alternative scenarios. It can be observed that the test statistic has power under the alternative hypothesis with empirical power close to 1. Also, under the null hypothesis, the empirical size remains close to significance level $\alpha = 0.05$. It is observed that the test statistic is not affected by the dimensionality of the manifold. For sample sizes larger than $n = 100$, the method holds a good balance between type-I error and type-II error.

Figure 2.22 compares the power curves for Kruskal-Wallis test, energy test (Chu and Dai, 2024), and BallDiv test (Zhu et al., 2021). It can be seen that the performance of the methods are similar and their power approaches the nominal value 1 as effect size increases.

2.3.4 Robustness Analysis

In this section, we will evaluate the robustness of the proposed rank-depth test statistics using examples from various manifold spaces with contaminated data. The goal is to investigate the robustness of these statistics in terms of their breakdown point.

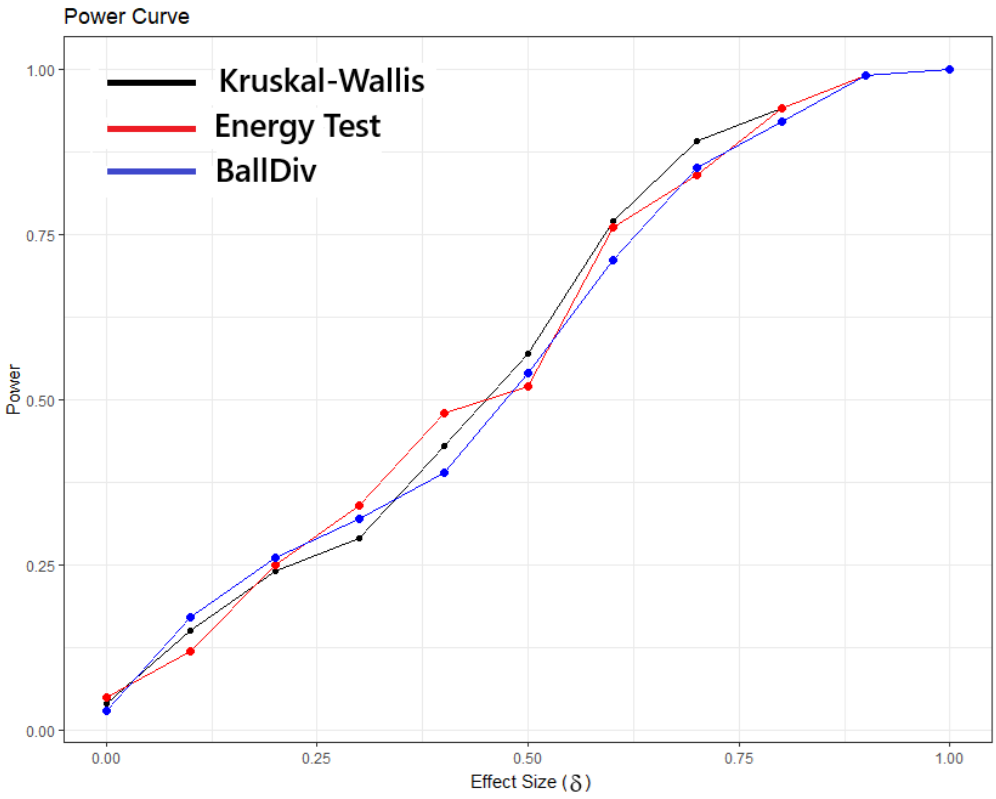


Figure 2.22: Power plot for covariance matrices data with power curves for Kruskal-Wallis test with GD, energy test, and ball divergence test.

Example 11 (Spearman Correlation Test for Wishart Distribution on SPD(3)): In this example, we carry out Monte Carlo simulation considering Wishart distribution on $SPD(3)$ to investigate the robustness of Spearman correlation coefficient test with respect to level breakdown point and power breakdown point. We consider multiple sample test with 3 groups ($K = 3$). For this simulation, paired samples of size $n = \{100, 200, 400, 1000\}$ are used where $n_x = n_y = n/2$.

For comparative study, we adopted ball-covariance test (Zhu et al., 2021) which is designed for testing mutual independence for data on general metric spaces. The data points are generated from Wishart distribution with degree of freedom 10 and identity concentration matrix of size 3×3. We consider the significance level of the test to be 0.05, and the number of replications equals 5000. For each of the considered sample sizes n, we generate 5000 samples of size n and calculate ε_{SA} and ε_{SR}. Then for each $1 \leq j \leq n$, we

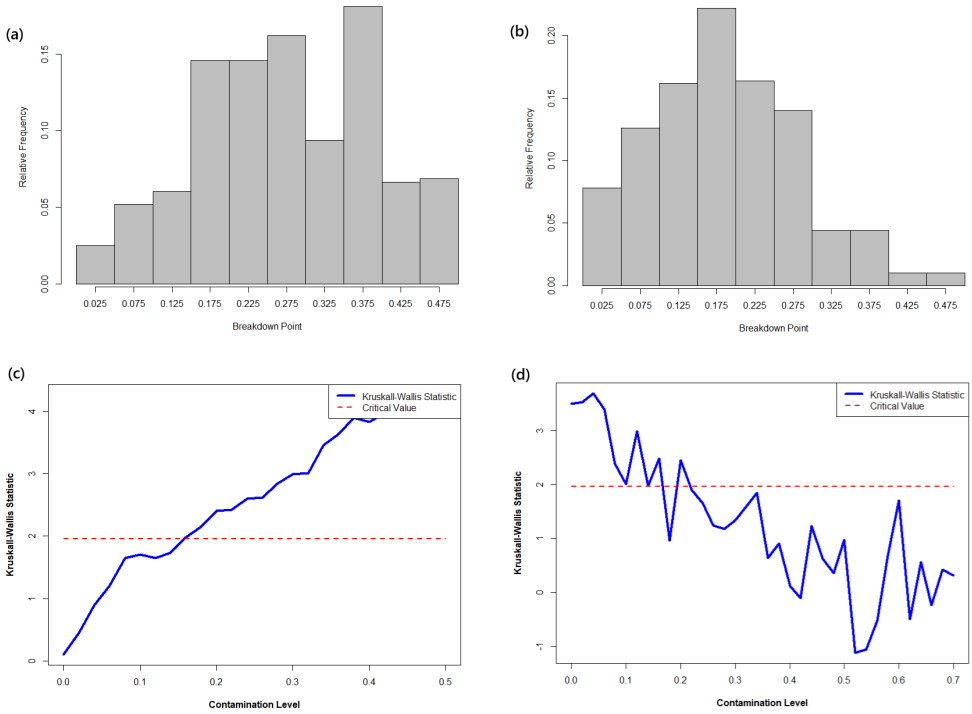


Figure 2.23: (a) level breakdown point histogram for Spearman test with $n_y = 50$, (d) power breakdown point histogram for Spearman test with $n_y = 50$, (e) Effect of contamination level on Spearman test statistic value under null hypothesis, (f) Effect of contamination level on Spearman test statistic value under alternative hypothesis.

compute the frequency $\#\{1 \leq i \leq 5000 \mid \varepsilon_{SA} = j/n_y\}/(5000 \times 0.95)$ for level breakdown point and frequency $\#\{1 \leq i \leq 5000 \mid \varepsilon_{SR} = j/n_y\}/(5000 \times 0.95)$ for power breakdown point (Zhang, 1996), which is the Monte Carlo approximation of theoretical probabilities

$$P\left(\varepsilon_{SA}(X) = j/n_y \mid \phi(X) = 0\right)/0.95 \ ,$$

$$P\left(\varepsilon_{SR}(X) = j/n_y \mid \phi(X) = 1\right)/0.95 \ .$$

Figure 2.23(a) illustrates the level breakdown point histogram for Spearman correlation coefficient test with $n_y = 50$, and Figure 2.23(b) depicts the power breakdown point histogram for Spearman correlation coefficient test with $n_y = 50$. In Figure 2.23(c) and Figure 2.23(d) , we evaluate the effect of contamination level on the performance of Spearman correlation coefficient test statistic under null and alternative hypotheses. It can be

Table 2.9: Simulation results for Spearman correlation coefficient test. The results are reported for level breakdown point (ε_{SA}), and power breakdown point (ε_{SR}).

	ε_{SA}				ε_{SR}			
	$n_y = 50$	$n_y = 100$	$n_y = 200$	$n_y = 500$	$n_y = 50$	$n_y = 100$	$n_y = 200$	$n_y = 500$
$j = 1$	0.021	0.055	0.083	0.001	0.065	0.001	0.001	0.001
$j = 5$	0.059	0.085	0.074	0.051	0.151	0.009	0.003	0.004
$j = 10$	0.153	0.117	0.083	0.060	0.203	0.026	0.024	0.041
$j = 15$	0.134	0.192	0.151	0.074	0.096	0.086	0.037	0.052
$j = 20$	0.081	0.206	0.224	0.145	0.028	0.151	0.236	0.014
$j = 25$	0.064	0.182	0.254	0.145	0.010	0.231	0.212	0.140
$j = 30$	0.000	0.136	0.106	0.163	0.000	0.238	0.358	0.192
$j = 40$	0.000	0.029	0.074	0.180	0.000	0.098	0.118	0.225
$j = 50$	0.000	0.002	0.004	0.166	0.000	0.042	0.026	0.166
$j = 70$	0.000	0.000	0.000	0.067	0.000	0.018	0.000	0.111

Table 2.10: Comparison of ε_{SA} and ε_{SR} values for Spearman correlation coefficient test, Quadrant test, and Ball covariance test.

	Spearman Test	**Quadrant Test**	**Ball Covariance Test**
ε_{SA}	0.40	0.40	0.24
ε_{SR}	0.40	0.40	0.24

realized that the Spearman correlation coefficient test becomes ineffective under both null and alternative hypotheses, as we increase the contamination level.

In Table 2.9, simulation results for computing the level and power breakdown points are reported. It can be observed that with probability around 0.6 the level breakdown point lies within the range $15 \leq n\varepsilon_{SA} \leq 30$, and the power breakdown point lies within the range $25 \leq n\varepsilon_{SR} \leq 40$.

Table 2.10 compares the level and power breakdown points for Spearman correlation coefficient test, Quadrant test, and BallCov test. It can be seen that the proposed rank-depth methods are superior in terms of robustness compared to BallCov test statistic.

Example 12 (Wilcoxon Test for Non-convex Distribution): In Example 10, Chapter 1, we studied the performance of the manifold based depth functions for distributions which are non-convex. In this example, we use the same distribution to investigate the robustness of Wilcoxon test for ε_{SA} and ε_{RA} using Monte Carlo simulation. For the simulation, bivariate samples of sizes $n = \{100, 200, 400, 1000\}$ are generated as $x_i = 2\,(\cos(\theta_i), \sin(\theta_i)) + 0.1^{1/2}\epsilon_i$, where θ_i is sampled i.i.d. from Uniform$[0, 2\pi]$ and ϵ_i are sampled i.i.d. from bivariate standard normal distribution. The reference and test

Table 2.11: Simulation results for Wilcoxon test. The results are reported for level breakdown point (ε_{SA}), and power breakdown point (ε_{SR}).

	ε_{SA}				ε_{SR}			
	$n_y = 50$	$n_y = 100$	$n_y = 200$	$n_y = 500$	$n_y = 50$	$n_y = 100$	$n_y = 200$	$n_y = 500$
$j = 1$	0.007	0.014	0.018	0.003	0.109	0.071	0.072	0.066
$j = 5$	0.016	0.029	0.035	0.019	0.042	0.096	0.075	0.060
$j = 10$	0.218	0.068	0.066	0.089	0.042	0.028	0.049	0.064
$j = 15$	0.223	0.129	0.107	0.073	0.070	0.024	0.021	0.068
$j = 20$	0.021	0.121	0.157	0.121	0.078	0.036	0.019	0.063
$j = 25$	0.000	0.070	0129	0.168	0.076	0.052	0.017	0.060
$j = 30$	0.000	0.001	0.168	0.215	0.089	0.056	0.035	0.051
$j = 40$	0.000	0.000	0.041	0.268	0.050	0.064	0.043	0.036
$j = 50$	0.000	0.000	0.001	0.178	0.000	0.147	0.130	0.361
$j = 70$	0.000	0.000	0.000	0.042	0.000	0.066	0.261	0.385

sample sizes are chosen to be $n_x = n_y = n/2$. For the comparison, we also consider two-sample t-test which is an extension of the work of (Osborne et al., 2013; Bhattacharya, 2008) to general metric spaces. To calculate the level and power breakdown points, we partitioned the space to equal-sized grids and the optimization problem is solved using these grid points. The significance level of the test is chosen to be 0.05, and the number of replications equals 5000. For each sample size n, 5000 samples of size n are generated and ε_{SA} and ε_{RA} are computed. Then for each $1 \leq j \leq n$, we compute the frequency $\#\{1 \leq i \leq 5000 \mid \varepsilon_{SA} = j/n_y\}/(5000 \times 0.95)$ for level breakdown point and frequency $\#\{1 \leq i \leq 5000 \mid \varepsilon_{SR} = j/n_y\}/(5000 \times 0.95)$ for power breakdown point.

Figure 2.24(a) depicts the distribution of the data points, and Figure 2.24(b) illustrates the equal-spaced grid points on the space. Figure 2.24(c) indicates the level breakdown point histogram for Wilcoxon test with $n_y = 50$, and Figure 2.24(d) indicates the power breakdown point histogram for Wilcoxon test with $n_y = 50$. In Figure 2.24(e) and Figure 2.24(f), we study the effect of contamination level on the performance of Wilcoxon test statistic under null and alternative hypotheses. It can be inferred that increasing the contamination level undermines the performance of Wilcoxon test under both null and alternative hypotheses.

The simulation results reported in Table 2.11 can be used to calculate the level and power breakdown points. It can be inferred that with probability around 0.6 the level breakdown point lies within the range $20 \leq n\varepsilon_{SA} \leq 40$, and the power breakdown point lies within the range $40 \leq n\varepsilon_{SR} \leq 60$.

Table 2.12 compares the level and power breakdown points for Spearman correlation coefficient test, geodesic T^2 test, energy test, and GET test. It can be seen that the

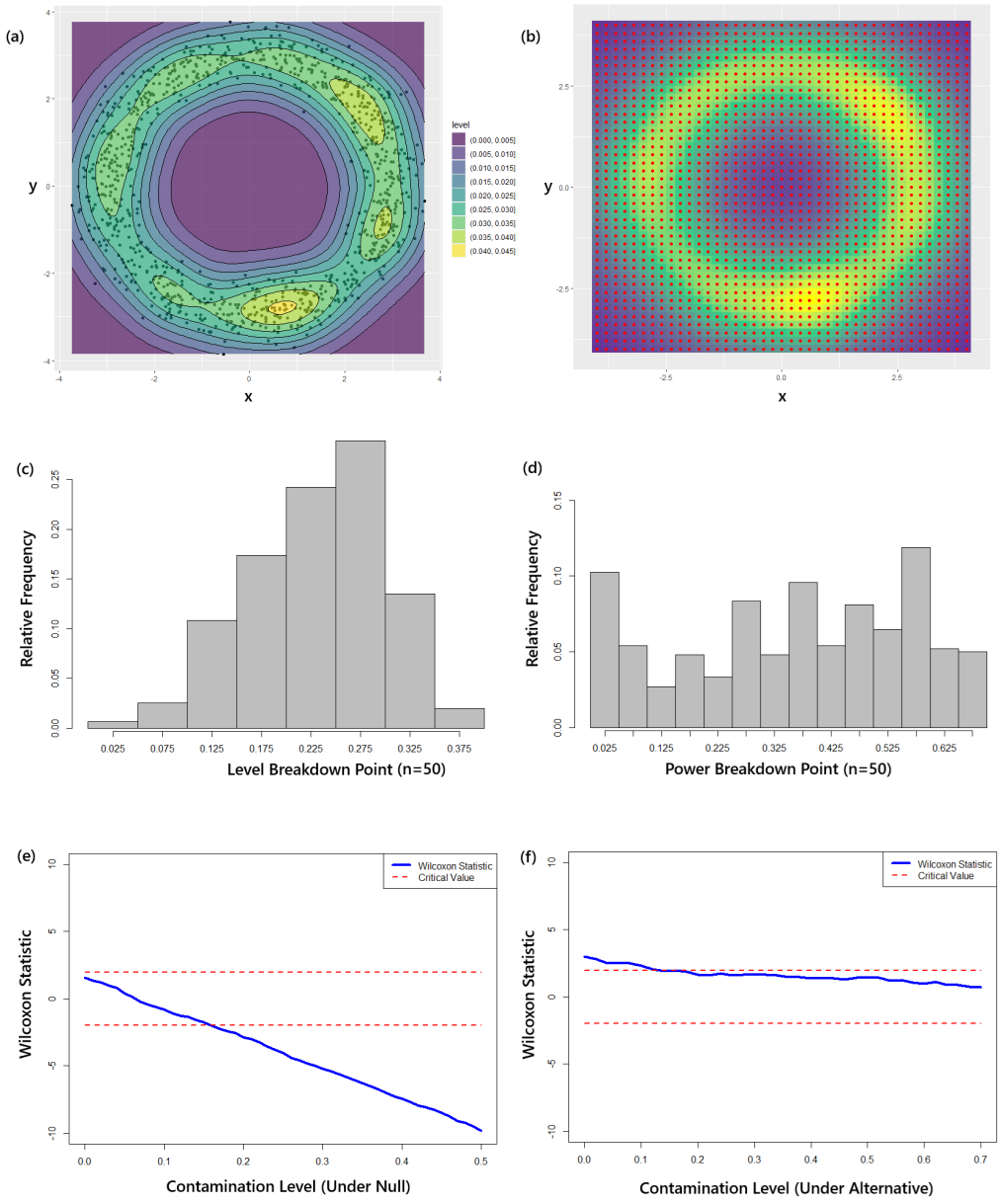


Figure 2.24: (a) Distribution of data with density contour map, (b) grid points on sample space domain, (c) level breakdown point histogram for Wilcoxon test with $n_y = 50$, (d) power breakdown point histogram for Wilcoxon test with $n_y = 50$, (e) Effect of contamination level on Wilcoxon test statistic value under null hypothesis, (f) Effect of contamination level on Wilcoxon test statistic value under alternative hypothesis.

Table 2.12: Comparison of ε_{SA} and ε_{SR} values for Wilcoxon test, geodesic T^2 test, energy test, and GET test.

	Wilcoxon Test	**Geodesic T^2 Test**	**Energy Test**	**GET Test**
ε_{SA}	0.50	0.20	0.39	0.20
ε_{SR}	0.50	0.20	0.39	0.20

proposed rank-depth Wilcoxon test statistic outperforms the rival methods in terms of robustness, indicating the efficacy of using depth function for developing rank-based test statistics compared to distance-based ranking for object data.

Example 13 (Kruskal-Wallis Test for VMF Distribution on $\mathcal{S}^2$:) In this example, we use von Mises-Fisher distribution on $\mathcal{S}^2$ to probe the robustness of Kruskal-Wallis test in terms of level breakdown point and power breakdown point using Monte Carlo simulation. We consider multiple sample test with 3 groups ($K = 3$). For the simulation, i.i.d. samples of size $n = \{150, 300, 600, 1500\}$ are considered where $n_1 = n_2 = n_3 = n/3$. For the comparison, we adopted a multiple sample test called ball-divergence statistic (Zhu et al., 2021) and an extension of Kolmogorov-Smirnov test to general metric spaces (Sidak et al., 1999). Figure 2.25 depicts a sample of 3 groups of data generated from VMF distribution on the space of $\mathcal{S}^2$. To calculate the level and power breakdown points, we partitioned the sphere space $\mathcal{S}^2$ into equally spaced grids, and solve the optimization problem on these grid points. For each of these sample sizes, the simulation is repeated for 5000 times and for each $1 \leq j \leq n$, the frequency $\#\{1 \leq i \leq 5000 \mid \varepsilon_{SA} = j/(n/K)\}/(5000 \times 0.95)$ for level breakdown point and frequency $\#\{1 \leq i \leq 5000 \mid \varepsilon_{SR} = j/(n/K)\}/(5000 \times 0.95)$ for power breakdown point are calculated, which are approximations of theoretical probabilities

$$P\left(\varepsilon_{SA}(X) = j/(n/K) \mid \phi(X) = 0\right)/0.95 \ ,$$

$$P\left(\varepsilon_{SR}(X) = j/(n/K) \mid \phi(X) = 1\right)/0.95 \ .$$

Figure 2.26(a) illustrates the distribution of the data points on sphere $\mathcal{S}^2$, and Figure 2.26(b) shows the equal-spaced grid points on $\mathcal{S}^2$. Figure 2.26(c) depicts the level breakdown point histogram for Kruskal-Wallis test with $n_1 = n_2 = n_3 = 50$, and Figure 2.26(d) indicates the power breakdown point histogram for Kruskal-Wallis test with $n_1 = n_2 = n_3 = 50$. Figure 2.26(e) and Figure 2.26(f) , we study the effect of contamination level on the performance of Kruskal-Wallis test statistic under null and alternative hypotheses. It can be realized that by increasing the contamination level, the performance of Kruskal-Wallis test statistic is deteriorated under both null and alternative hypotheses.

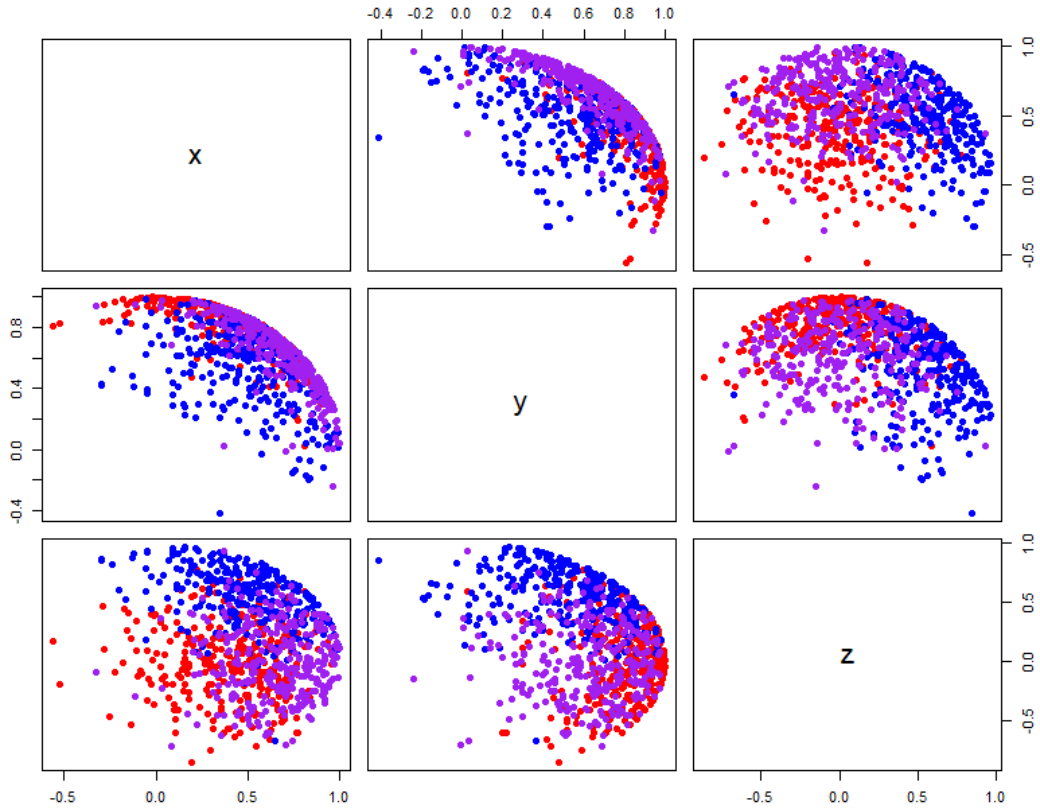


Figure 2.25: A sample of $K = 3$ groups of data generated from VMF distribution on the space $\mathcal{S}^2$.

In Table 2.13, the simulation results for computing the level and power breakdown points are reported. It can be seen that with probability around 0.6 the level breakdown point lies within the range $20 \leq n\varepsilon_{SA} \leq 40$, and the power breakdown point lies within the range $40 \leq n\varepsilon_{SR} \leq 60$. Table 2.14 compares the level and power breakdown point of Kruskal-Wallis test, BallDiv test, and energy test. As it can be seen, the proposed test statistic outperforms the rival methods in terms of robustness.

Example 14 (Final Breakdown Point for Rank-Depth Tests): The concept of final breakdown point of rank statistics (Rieder, 1982) can be used to analytically evaluate the robustness of rank statistics. The point of final breakdown of a sequence of statistics is defined as the critical amount of contamination beyond which a test statistic is unable to

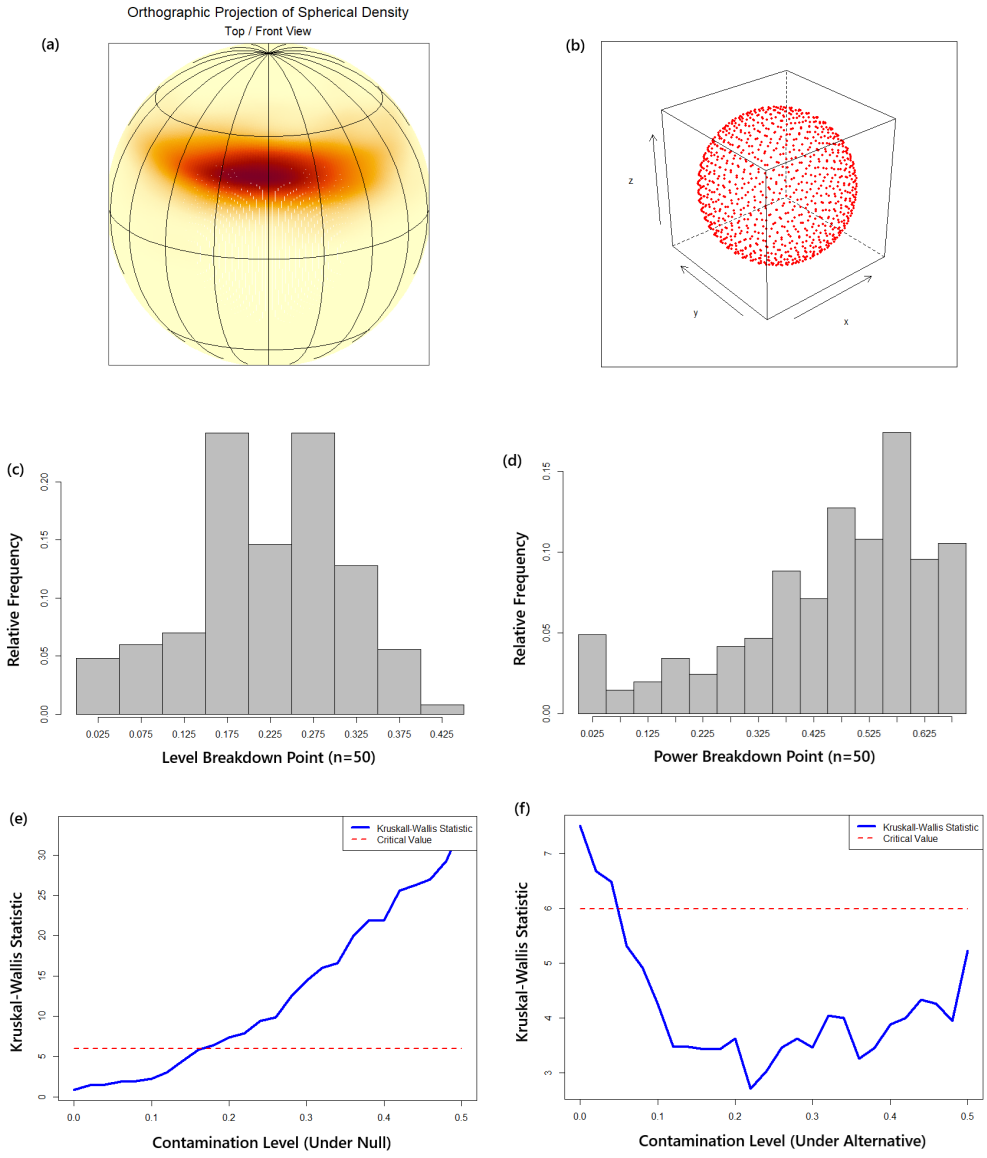


Figure 2.26: (a) Distribution of data with density contour map (top-front view), (b) grid points on sample space domain, (c) level breakdown point histogram for Kruskal-Wallis test with $n_y = 50$, (d) power breakdown point histogram for Kruskal-Wallis test with $n_y = 50$, (e) Effect of contamination level on Kruskal-Wallis test statistic value under null hypothesis, (f) Effect of contamination level on Kruskal-Wallis test statistic value under alternative hypothesis.

Table 2.13: Simulation results for Kruskal-Wallis test. The results are reported for level breakdown point (ε_{SA}), and power breakdown point (ε_{SR}).

	ε_{SA}				ε_{SR}			
	$n_1 = 50$	$n_1 = 100$	$n_1 = 200$	$n_1 = 500$	$n_1 = 50$	$n_1 = 100$	$n_1 = 200$	$n_1 = 500$
$j = 1$	0.040	0.048	0.080	0.001	0.041	0.004	0.001	0.000
$j = 5$	0.068	0.100	0.096	0.006	0.018	0.007	0.005	0.008
$j = 10$	0.204	0.183	0.052	0.080	0.030	0.004	0.008	0.005
$j = 15$	0.194	0.286	0.113	0.124	0.046	0.002	0.008	0.008
$j = 20$	0.033	0.264	0.155	0.102	0.083	0.007	0.011	0.016
$j = 25$	0.008	0.136	0.174	0.179	0.123	0.010	0.019	0.047
$j = 30$	0.000	0.054	0.128	0.224	0.141	0.016	0.016	0.018
$j = 40$	0.000	0.006	0.034	0.199	0.105	0.227	0.222	0.227
$j = 50$	0.000	0.000	0.004	0.192	0.000	0.176	0.288	0.168
$j = 70$	0.000	0.000	0.000	0.053	0.000	0.288	0.168	0.288

Table 2.14: Comparison of ε_{SA} and ε_{SR} values for Kruskal-Wallis test, BallDiv test, and energy test.

	Kruskal-Wallis Test	**Ball Divergence Test**	**Energy Test**
ε_{SA}	0.39	0.30	0.31
ε_{SR}	0.39	0.30	0.31

asymptotically distinguish any two probailities. It can be shown that the points of final breakdown of rank statistics coincide with the breakdown points of the corresponding rank estimators (Hampel, 1971). Let us define the ranks of the form

$$r_i^* = \sum_{i=1}^{n} \mathbb{I}\left[\|x_j\| \leq \|x_i\|\right] \ ,$$

$i = 1, \ldots, n$. We consider the following rank statistics

$$T(P_n) = \frac{1}{n}\sum_{i=1}^{n} \text{sign}(h(x_i)) a_n(r_i^*) = 2\int_{(0,\infty)} a_n\left(n(F_n(t) - F_n(-t))\right) F_n(dt) - \frac{1}{n}\sum_{i=1}^{n} a_n(i) \ ,$$

where $h : \mathcal{M} \to \mathbb{R}$, the scores are increasing, i.e. $a_n(1) \leq a_n(2) \leq \cdots \leq a_n(n)$, and F_n is c.d.f. corresponding to the empirical measure P_n. There exists a function $a(\cdot)$ and Lebesgue measure λ such that

$$\lim_{n\to\infty} \sum_{i=1}^{n} \left| \int_{[(i-1)/n, i/n]} (a_n(i) - a(s))\, \lambda(ds) \right| = 0 \ .$$

Then we can write

$$T(P) = 2\int_{(0,\infty)} a\left(F(t) - F(-t)\right) F(dt) - \int a d\lambda \ .$$

In Rieder (1982), it was shown that $T(P_n) \to T(P)$ almost surely, as $n \to \infty$. Also, it was proved that the point of final breakdown is the maximum solution ε^* of the following equation

$$\int_{(0,1-\varepsilon^*)} a d\lambda = \int_{(1-\varepsilon^*,1)} a d\lambda \ .$$

It can be shown that for sign tests we get $\varepsilon^* \approx 0.5$, for test statistics with normal score function we have $\varepsilon^* \approx 0.239$, for Spearman correlation coefficient test and Quadrant test we get $\varepsilon^* \approx 0.21$, for Wilcoxon test we get $\varepsilon^* \approx 0.293$, and for Kruskal-Wallis test we have $\varepsilon^* \approx 0.21$. These values can be verified from the plots given in the previous examples studying the effect of contamination level on the behaviour of the empirical rank-depth test statistics under both null and alternative hypotheses as well as the tables reporting the simulation results for the level and power breakdown point of the empirical rank-depth test statistics.

2.4 Discussion and Future Work

In this chapter, we implemented several hypothesis testing methods for object data belonging to various manifold spaces. The performance of the considered test statistics were evaluated using Monte Carlo simulation and real-world datasets. The efficacy of the hypothesis testing methods were confirmed through numerical experiment and theoretical investigation. In particular, it was observed that these test statistics are robust and can be applied to object data belonging to various types of manifold curved spaces, e.g. SPD, Kendall's shape space, Torus space, sphere space, and network data including ERG and SBM. An interesting observation was that these rank-based test statistics were powerful regardless of the dimensionality of the manifold space and kept their power when the sample size was as small as $n = 20$. Finite sample Monte Carlo power studies endorsed the quality of the asymptotic results derived for the considered test statistics. Furthermore, it was realized that the proposed methods are effective for object data with large dimensions. Several rival methods are adopted from the literature to evaluate the performance of the nonparametric tests. It can be concluded that rank-depth nonparametric tests can control type-I error with respect to the specified level and also have low type-II error. It was observed that the power of the proposed test statistics grows faster to 1 as the effect size

increases. This general observation is valid for object data sampled from various types of manifold spaces with specific geometrical features. Also, it was observed that the proposed methods can successfully be applied to both intrinsic and extrinsic manifold analysis frameworks. Robustness analysis was carried out from both theoretical and numerical aspects. From theoretical viewpoint, we performed q-robustness analysis and derived asymptotic results for level and power breakdown point of the proposed test statistics. Comparative study substantiated by theoretical findings demonstrated that the rank-depth methods can outperform other methods available in the literature, and are less sensitive to the contaminating outlier data.

Future research direction involves applying the proposed nonparametric methods to real-world emerging problems from various research disciplines ranging from automotive and robotics to process monitoring. Besides, it is important to further evaluate the asymptotic results for object data with large dimensions and various geometrical features. It is believed that the rank-depth methods can play a very important role given their generalization power and the fact that they free the analysis from any assumption regarding the distribution of the data.

Chapter 3

Change Point Detection

In this chapter, a class of nonparametric change point detection for stochastic processes defined on manifold spaces $\mathcal{M}$ is proposed. The rest of the chapter is organized as follows. In Section 1, we introduce the change point detection problem on manifold spaces and discuss in detail some important research works on general change point detection related to the topic of this chapter. In Section 2, a formulation of a change point detection scheme on general metric spaces is presented. In Section 3, a ranking based version of CUSUM statistic based on depth function defined on manifold space is formulated. In Section 4, the formulation of the change point detection is extended to multiple change point detection. Section 5 is devoted to robustness analysis of change point detection methods. Simulation study is carried out in Section 6. In Section 7, the findings of the research as well as future research directions are outlined.

3.1 Change Point Detection for Manifold Data

Change point detection methods are powerful statistical tools particularly developed for estimating the number, location and magnitude of changes in the underlying stochastic structure of an evolvable process. In doing so, such stochastic processes are divided into a number of homogeneous sub-processes with specific structures, such as constant mean and/or constant variance, constant correlation structure, and other distributional properties such as symmetry, frequency, tailedness, etc. Given the effectiveness of such statistical tools, they have successfully been applied to identifying change in the underlying stochastic structure of various real-world processes, including social, financial, environmental, and biological processes (Niu et al., 2016). A remarkable number of research papers exist

on both single and multiple change point detection methods. See for example Xie and Siegmund (2013); Aminikhanghahi and Cook (2017); Hušková (2013); Horváth and Rice (2014); Truong et al. (2020). Multiple change point detection methods are essential in the sense that for many real-world processes, there is a need for accurate estimation and isolation of underlying sub-processes with homogeneous stochastic structures. For example, one may use multiple change point detection for identifying tumour cells in which the interest is to detect regions in the chromosomes with copious deletion or amplification of DNA (Pierre-Jean et al., 2015).

The first research paper on change point analysis was published in 1955 (Page, 1955), where the author explained the difference between a change point detection problem and a two-sample test problem. For change point detection problem the point where the distributional properties of the data changes is unknown as opposed to the two sample comparison test for which the point is known before hand and does not need to be estimated. After this seminal publication which clarifies the importance of change point detection, an extensive research has been carried out for proposing and analyzing single and multiple change point detection methods for various types of data. A comprehensive review of the state-of-the-art can be found in Aminikhanghahi and Cook (2017). From the available reports in the literature, it can be inferred that multiple change point detection methods are in particular important because they are more general in the sense that they can detect and estimate one or more change points in the data sequence. Well-known multiple change point detection methods include standard binary segmentation (Vostrikova, 1981; Venkatraman, 1993), least-square estimation (Yao and Au, 1989), maximum likelihood estimation based on robust metrics (Arunajadai, 2009), Schwarz-like penalties (Ciuperca, 2011), wild binary segmentation (Fryzlewicz, 2014), CUSUM based on the correlation matrix of a sequence of random variables (Galeano and Wied, 2017), tail-greedy unbalance Haar (TGUH) (Fryzlewicz, 2018), narrowest over threshold (NOT) (Baranowski et al., 2019), and isolate detection (ID) (Anastasiou and Fryzlewicz, 2019). In spite of the fact that nonparametric single change point analysis has extensively been explored in the literature (Brodsky and Darkhovsky, 2013), there are not much research on nonparametric statistics for multiple change point detection.

Earliest works on developing rank-based statistics for nonparametric two sample and multiple sample hypothesis tests comprises of celebrated tests such as Wilcoxon signed-rank and sum rank tests (Wilcoxon, 1945), Mann-Whitney test (Mann and Whitney, 1947), Fisher-Yates-Terry-Hoeffding test (Hoeffding, 1951; Terry, 1952), van der Waerden test (Van der Waerden, 1952), Kruskal-Wallis test (Kruskal and Wallis, 1952), Kolmogorov-Smirnov and Cramer-von Mises tests (Kiefer, 1959), and scale test (Ansari and Bradley, 1960). These tests have mainly been proposed for comparing two or more population

in terms of one or more statistical characteristics of these populations, such as location, scale, correlation, symmetry, etc. It has been proven that such test statistics can become asymptotically as powerful as most powerful tests for testing change in the location or scale of two or more parametric distributions (Siegel and Tukey, 1960). Moreover, the other advantage of working with rank-based statistics is that such tests automatically introduce robustness to the testing procedures, which potentially makes it possible to effectively apply such methods to datasets contaminated with heavy tailed noises and outliers.

Following the promising reports in the literature regarding the application of rank-based statistics for developing nonparametric statistics for two sample and multiple sample hypothesis tests, researchers have been persuaded to apply rank-based statistics for developing nonparametric change point detection methods. Most of the rank-based change point detection methods are developed for single change point problems.

Lombard (1987) can be enumerated as one of the earliest research works on using rank-based statistics and studying their asymptotic properties for change point detection. The proposed test statistic is based on the square of the normalized scores of ranks for the detection both single and multiple abrupt and smooth changes. Katsevich and Ramm (1993) proposed a variant of rank-based Durbin-Watson statistic to test two classes of alternatives, namely stochastic monotonocity and arbitrary continuous trend, in a sequence of univariate/multivariate observations. Sugiura and Ogden (1994) proposed several variants of rank tests including a weighted version of Lombard's test statistic, a rank-based likelihood ratio test, and rank-based Bayes tests to detect abrupt/smooth changes in a sequence of independent observations. Koziol (1996) developed several versions of signed rank tests to detect abrupt and smooth changes in the location parameter of the underlying distribution of a sequence of random variables. Venter and Steel (1996) proposed the use of maximum likelihood estimator and ranking via standard score functions namely Wilcoxon and van der Waerden scores for detecting multiple abrupt changes in a sequence of nearly normal random variables. Aly and BuHamra (1996) developed Kruskal-Wallis-type tests to evaluate the null hypothesis of no change against two change points alternative in a sequence of independent observations.

Antoch and Hušková (2001) developed a variant of rank test generated by kernel functions to detect location shifts. Asymptotic distribution under the null hypothesis of no change point has been derived for the proposed test statistic. Besides, consistency of the test was proved under the alternative. Due to the slow convergence of the asymptotic distribution under null, Monte Carlo simulations were carried out using various types of kernels under different sources of contaminating noises and sample-based critical values were also reported. Antoch and Hušková (2009) performed theoretical studies on using permutation tests to detect changes in location and investigated the relation between permutation

tests and rank-based tests. There are also some important contributions in the literature on extending the area of application of rank-based statistics to testing for change in the dependence structure of random variables (Hušková, 2003), testing for change when there are censored data in the sample (Hušková and Neuhaus, 2004), regression rank statistics (Hušková, 2004), and empirical characteristic function of ranks (Hušková and Meintanis, 2006). The outcomes of these research studies demonstrated that rank-based change point detection methods are as efficient as most of the well-known standard tests for single and multiple change detection scenarios. Also, replacing the original observations with their corresponding ranks makes standard statistics robust against different types of outliers and contamination noises.

It is worth pointing out that there exists research in the literature where the authors used parametric models to formulate change point detection for other types of noises. In Wu and Fitzgerald (1995), the authors proposed an analytical framework using Bayesian inference for the identification of time series discontinuity. The algorithm is formulated in the form of change point detection in impulsive Laplacian noise. Exact expressions for posterior density of change point positions are derived and its performance was compared with Monte-Carlo method based on Gaussian assumption for the noise. Given the fact that it is not always possible to derive parametric models for formulating change point detection with non-Gaussian noises, the rank-based statistics can serve as a general tool for implementing nonparametric change detection algorithm when the noise is non-Gaussian.

Most of the above literature are focused on cases with single change point in the data sequence. In recent years, research on change point detection has been extended to multiple change point detection scenarios. Binary segmentation (BS) is among one of the most well-established strategies for sequential detection of multiple change points. BS is based on splitting the whole sample into two sub-samples at the location of a change point. After that the two sub-samples are searched for possible change points. This detection and segmentation strategy is continued until no further change point can be detected in the sample. BS has advantageous properties including low computational complexity and being easy to implement; however, its performance can become deficient for certain configuration of multiple change points. In particular, when the spacing between consecutive change points is less than a threshold, binary segmentation becomes inconsistent.

To cope with this shortcoming, Fryzlewicz (2014) proposed wild binary segmentation (WBS) strategy which uses a localized CUSUM statistic based on random sampling of sub-segments. It has been demonstrated that this strategy significantly reduces the required space between neighbouring change points. There also exists some modified versions of WBS such as narrowest over threshold (NOT) (Baranowski et al., 2019) and isolate detection (ID) (Anastasiou and Fryzlewicz, 2019) which are appropriate for detecting changes

in more general scenarios such as detection of changes in trend of piecewise linear models with possible discontinuities in the signal. The performed simulation studies indicated that these methods can surpass the standard BS method. However, these methods are best suited for signals contaminated with Gaussian noises. Extension of multiple change point detection algorithms to robust change point detection methods has recently been carried out. In Wang et al. (2020), the authors proposed a method called RANK which is different from BS in the sense that it is based on a Kruskal-Wallis type test and a variant of dynamic programming (DP) optimization for change point detection and estimation. It has been demonstrated that RANK can show robust performance; however, when the sample size grows the computational cost can become significant.

In recent years, the area of application of change point detection has been extended to the cases with data sampled from various types of manifold spaces (Wang et al., 2025). Such change point detection methods are usually implemented based on distances defined on general metric spaces. In Duan et al. (2019), a version of change point detection method for special Euclidean group $SE(3)$ has been proposed which is based on approximation of Riemannian mean using Baker-Campbell-Hausdorff formula and gradient descent algorithm. The efficacy of the algorithm was attested considering numerical examples and robot manipulator experiment. In Zambon et al. (2018), a change point detection algorithm was proposed based on mapping complex data into suitable lower dimensional manifold, i.e. embedding graphs into constant-curvature manifolds, e.g. hyper-spherical and hyperbolic manifolds. Simulation studies indicated that the geometry of complex network can have a hyperbolic geometry. It has been demonstrated that the performance of the proposed method is comparable to the groundtruth methods applied directly to the complex space of graph data.

In Huang et al. (2020), the authors proposed a change point detection algorithm for anomaly detection in temporal graphs based on short-term / long-term modeling via the spectrum of the Laplacian matrix. The algorithm is applied to real-world dynamical networks UCI message network, US senate co-sponsorship network and Canadian bill voting network. It was observed that the algorithm can detect anomalous time points according to significant real events. In Xie et al. (2012), the authors proposed a sub-manifold learning framework for change point detection of high-dimensional data streams with missing components. The idea is based on estimating the intrinsic dimension of the high-dimensional space and detecting abrupt or unexpected changes in the underlying manifold subspace. Simulation studies on real-world data, e.g. solar flare detection in a video stream from solar data observatory, and asymptotic analysis demonstrated the efficacy of the proposed method for detecting the abrupt change in the otherwise slowly varying low-dimensional manifold subspace. In Huang et al. (2021), the authors proposed a deep learning sub-

manifold representation framework for change point detection and applied it to fMRI data to detect in cognition and behaviour. The method is formulated in the manifold space of SPD matrix data and the correlation matrices are used to model the functional connectivity and interaction between distance brain region. The change point detection algorithm detects dynamical shifts along the states of brain network topology. Simulation results demonstrated that the method has an accurate performance compared to other change point detection methods.

In Zhang (2021), the authors proposed a change point detection method for networks of point processes where scanning score for network data is used to model the sequential events. Theoretical study is carried out considering the false alarm rate and average run length of the process under the null hypothesis. The method is applied to memetracker data and stock data. In Grattarola et al. (2018), the authors proposed a graph learning mechanism to embed the graph data into a constant curvature manifold subspace. The method is used to formulate a change point detection method for estimating changes in the stationarity of graph generating process. Simulation studies were carried out on graph stream data including functional networks extracted from iEEG data (Bayram and Arserim, 2021) to detect uncontrolled burst of electrical activity in brain. In Yger et al. (2012), the authors formulated a canonical correlation analysis for the space of matrix manifolds and applied it for change point detection in the correlation matrix of two time series. The proposed method was applied to BCI dataset (Blankertz et al., 2004).

In Auret and Aldrich (2010), the authors proposed a change point detection for dynamic systems and time series data where random forest is used to embed the time series data to nonlinear manifold space and extracting features from the resulting lagged trajectory matrix. In Huang et al. (2021), the authors proposed a method for detecting brain state change using manifold mean shifting. The method is based on defining a set of landmarks on the space of Riemannian manifold and considering mean shift to detect cognition changes by identifying representative networks of the distribution of functional networks. Simulation studies considering neuroimaging data obtained from human Connectome project demonstrated the accuracy and consistency of the method in detecting and estimating cognition changes. In Deng et al. (2024), the authors proposed Riemannian geometric method for change detection of correlation matrices data based on CUSUM statistic and geodesic distance between current observation and Fréchet of previous observations. Simulation results indicated the effectiveness of the method in terms of detection accuracy, average detection delay, and efficiency.

From the above research studies, it can be perceived that standard change point detection can be extended to cases where data lie in manifold spaces. The formulation of these change point detection algorithms depends on the geometrical properties of the man-

ifold space and heavily relies on concepts such as the calculation of Fréchet mean and Fréchet variance (Wang et al., 2024). Also, effectiveness of the interpretations is related to formulation of an algorithm to properly identify the intrinsic dimension of the seemingly high-dimensional data. In this chapter, the author intends to investigate the potential of robust statistical methods for implementation of change point detection methods on various types of manifold spaces. The idea we seek here is based on adapting change point detection statistics to general metric spaces based on applying rank of depth function appropriate for manifold data.

The proposed method is a variant of rank-CUSUM developed for manifold data and is implemented for detecting both single change point and multiple change point detection. Theoretical study is carried out to investigate the consistency properties of the test statistic as a function of sample size and intrinsic dimension. The power of the test under the general alternative of change is investigated and asymptotic distribution for the change point test statistic is derived. Several manifold dataset of various types are considered to endorse the veracity of the obtained theoretical results regarding the performance of the change point detection algorithm.

3.2 Change Point Analysis on General Metric Spaces

In this section, a generalized version of single change point detection for change in location and scale is implemented based on the concept of Fréchet mean and Fréchet variance. Consider the random object $X \sim P$ defined on metric space (Ω, δ). Then, the Fréchet mean and Fréchet variance are defined as

$$\mu = \arg\min_{\omega \in \Omega} \mathbb{E}[\delta^2(X, \omega)] \ ,$$
$$\nu = \min_{\omega \in \Omega} \mathbb{E}[\delta^2(X, \omega)] \ .$$

Note that for multivariate distribution $X \sim P$ on $(\mathbb{R}^p, \|\cdot\|)$ with $\mathbb{E}[\|X\|^2] < \infty$, we have

$$\mu = \mathbb{E}[X] \ ,$$
$$\nu = \text{trace}(\text{Cov}(X)) = \sum_{i=1}^{p} \sigma_i^2 \ .$$

For $X_1, X_2, ..., X_n$, the sample version of Fréchet mean and Fréchet variance can be defined as

$$\widehat{\mu}_n = \arg\min_{\omega \in \Omega} \frac{1}{n} \sum_{i=1}^{n} \delta^2(X_i, \omega) \ ,$$
$$\widehat{\nu}_n = \frac{1}{n} \sum_{i=1}^{n} \delta^2(X_i, \widehat{\mu}_n) \ .$$

In Dubey and Müller (2019), it was shown that for bounded metric space (Ω, δ), we have the following consistency results

$$\begin{aligned} &\delta(\mu, \widehat{\mu}_n) = o_p(1), \quad n \to \infty \ , \\ &|P_n - P| \leq \left| \frac{1}{n} \sum_{i=1}^{n} [\delta^2(X_i, \widehat{\mu}_n) - \delta^2(X_i, \mu)] \right| + \left| \frac{1}{n} \sum_{i=1}^{n} [\delta^2(X_i, \mu) - P] \right| \\ &\qquad \leq 2 \ \mathrm{diam}(\Omega) \ \delta(\widehat{\mu}_n, \mu) + \left| \frac{1}{n} \sum_{i=1}^{n} [\delta^2(X_i, \mu) - P] \right| = o_p(1) \ . \end{aligned}$$

Now, consider a sequence of random objects $X_1, X_2, ..., X_n$ taking values in a bounded metric space (Ω, δ). Our interest is to test the null hypothesis

$$H_0 : X_1, X_2, ..., X_n \sim P_1 \ ,$$

against the alternative hypothesis

$$H_A : \text{there exists } 0 < \theta < 1 \text{ such that } \begin{cases} X_1, X_2, ..., X_{[n\theta]} \sim P_1 \\ X_{[n\theta]+1}, X_{[n\theta]+2}, ..., X_n \sim P_2 \end{cases}$$

where $\theta \in [0, 1]$ is called breaking fraction.

Suppose u is a possible change point location under the alternative. This divides the data sequence into two segments, observations located before $[nu]$ and observations located after $[nu]$. The estimated Fréchet mean and Fréchet variance for observations in the first segment can be obtained as

$$\widehat{\mu}_{[0,u]} = \arg\min_{\omega \in \Omega} \frac{1}{[nu]} \sum_{i=1}^{[nu]} \delta^2(X_i, \omega) \ ,$$

$$\widehat{V}_{[0,u]} = \frac{1}{[nu]} \sum_{i=1}^{[nu]} \delta^2(X_i, \widehat{\mu}_{[0,u]}) \ ,$$

and for observations in the second segment, the estimates are

$$\widehat{\mu}_{[u,1]} = \arg\min_{\omega \in \Omega} \frac{1}{n - [nu]} \sum_{i=[nu]+1}^{n} \delta^2(X_i, \omega) \ ,$$

$$\widehat{V}_{[u,1]} = \frac{1}{n - [nu]} \sum_{i=[nu]+1}^{n} \delta^2(X_i, \widehat{\mu}_{[u,1]}) \ .$$

Under the null hypothesis of no change point $\widehat{\mu}_{[0,u]}$ is close to $\widehat{\mu}_{[u,1]}$ and consequently $\widehat{V}_{[u,1]} - \widehat{V}_{[0,u]}$ will be small.

In Dubey and Müller (2020), the authors proposed using contaminated version of the above estimates to develop a unified test statistic for detecting change in location and scale. The contaminated version of the estimates can be obtained by replacing the Fréchet mean from the corresponding segment with the Fréchet mean from the other segment. The complement variance is defined as

$$\widehat{V}^C_{[0,u]} = \frac{1}{[nu]} \sum_{i=1}^{[nu]} \delta^2(X_i, \widehat{\mu}_{[u,1]}) \ ,$$

$$\widehat{V}^C_{[u,1]} = \frac{1}{n - [nu]} \sum_{i=[nu]+1}^{n} \delta^2(X_i, \widehat{\mu}_{[0,u]}) \ .$$

It was indicated that for $\widehat{\mu}_{[0,u]} \neq \widehat{\mu}_{[u,1]}$, the difference can be detected using the statistic $(\widehat{V}^C_{[0,u]} - \widehat{V}_{[0,u]} + \widehat{V}^C_{[u,1]} - \widehat{V}_{[u,1]})^2$ and for $\widehat{V}_{[0,u]} \neq \widehat{V}_{[u,1]}$, the difference can be detected using $(\widehat{V}_{[0,u]} - \widehat{V}_{[u,1]})^2$. Therefore, the following statistics was proposed for change in both location and scale

$$Z_n(u) = \frac{u(1-u)}{\widehat{\sigma}^2} \Big\{ (\widehat{V}_{[0,u]} - \widehat{V}_{[u,1]})^2 + (\widehat{V}^C_{[0,u]} - \widehat{V}_{[0,u]} + \widehat{V}^C_{[u,1]} - \widehat{V}_{[u,1]})^2 \Big\}$$

where $0 < u < 1$ and

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \delta^4(X_i, \widehat{\mu}) - \widehat{V}^2 \ .$$

It was shown that under the null hypothesis, the test statistic $\{nZ_n(u) : 0 < u < 1\}$ converges in distribution to the square of the stochastic process

$$\left\{\mathcal{G}(u) = \frac{\mathcal{B}(u)}{\sqrt{u(1-u)}} : 0 < u < 1\right\} ,$$

where $\mathcal{B}(u)$ is a Brownian Bridge on the interval $[0, 1]$, which is a process with mean 0 and covariance kernel $K(s, t) = \min(s, t) - st$. Therefore, the test statistic for detecting a change point is defined as

$$T_n = \sup_{0<u<1} nZ_n(u) ,$$

which converges in distribution to

$$\sup_{0<u<1} \mathcal{G}^2(u) .$$

Let us denote the significance level for the process by $q_{1-\alpha}$, then the rejection region becomes

$$R_{n,\alpha} = \{T_n > q_{1-\alpha}\} .$$

The formulated change point detection method is not robust in the sense that it assumes that the metric space (Ω, δ) is totally bounded. In addition, it is parametric because it uses Fréchet mean and Fréchet variance of the underlying probability measure. Rank of depth functions can be used to develop nonparametric change point detection method which is robust against outliers.

For a metric space (Ω, δ) with the probability measure $X \sim P$, and any $r > 0$, the following depth function can be defined

$$D(P, \omega) = (1 + \mathbb{E}_P[\delta^r(X, \omega)])^{-1}$$

to measure how central the point $\omega \in \Omega$ is w.r.t. probability measure P.

For $X_1, X_2, ..., X_n$, the empirical version of the depth function is defined as

$$D(P_n, \omega) = \left(1 + \frac{1}{n}\sum_{i=1}^{n} \delta^r(X_i, \omega)\right)^{-1}$$

For the above depth function, we can achieve the following strong uniform consistency and a nonasymptotic sharp bound using Bernstein-type McDiarmid inequality. The convergence results are based on density estimation on manifold $\mathcal{M}$ by constructing a basis of wavelet functions which can adapt to the local geometry of the manifold making the

localized and efficient representation of probability density function possible. The benefit of such multi-resolution basis representation is to cope with the challenges such as complex curvature, high dimension, etc., by capturing features at different scales (Kerkyacharian et al., 2012).

Theorem 12. *Suppose that probability measure P on manifold $\mathcal{M}$ with $\dim(\mathcal{M}) = d$ has bounded probability density function f. Then we have*

$$\limsup_{n\to\infty} \sqrt{\frac{n}{2^{j_n d/2} j_n^2}} \sup_{y\in\mathcal{M}} \mid D(P_n, y) - D(P, y) \mid \leq K \ ,$$

where K is a constant which depends on manifold $\mathcal{M}$ and $\|f\|_\infty$, and j_n determines the number of basis in wavelet-type estimator of f satisfying

$$\frac{2^{j_n d/2} j_n^2}{n} \to \infty \ .$$

Theorem 13. *Suppose that probability measure P on manifold $\mathcal{M}$ with $\dim(\mathcal{M}) = d$ has bounded probability density function f. Let us define the the deviation and variance term as follows*

$$B = \max_{i\in[n]} \sup_{\substack{x_1,x_2,...,x_{i-1},\\ x_{i+1},...,x_n}} \left| \widehat{f}_n(x_1, x_2, ..., x_{i-1}, X_i, x_{i+1}, ..., x_n) - \mathbb{E}_{X_i}\left[\widehat{f}_n(x_1, x_2, ..., x_{i-1}, X_i, x_{i+1}, ..., x_n)\right] \right| ,$$

$$V_i = \sup_{\substack{x_1,x_2,...,x_{i-1},\\ x_{i+1},...,x_n}} \mathbb{E}_{X_i}\left[\widehat{f}_n(x_1, x_2, ..., x_{i-1}, X_i, x_{i+1}, ..., x_n) - \mathbb{E}_{X_i}\left[\widehat{f}_n(x_1, x_2, ..., x_{i-1}, X_i, x_{i+1}, ..., x_n)\right]\right]^2 ,$$

$$\sigma^2 = \sum_{i=1}^{n} V_i \ .$$

Then, for any $t > 0$, we have the following Bernstein form of McDiarmid's nonasymptotic sharp bound

$$P\Big(\sup_{y\in\mathcal{M}} \mid D(P_n, y) - D(P, y) \mid \geq t \Big) \leq \exp\Big(- \frac{t^2}{2\sigma^2(1 + \frac{Bt}{3\sigma^2})} \Big) \ .$$

In the next sections, the concept of rank-depth is used to develop nonparametric and robust change point detection method when there are single or multiple change points in data.

3.3 Rank-CUSUM on Manifold Spaces

It is well-known that using depth function can improve the robustness of standard CUSUM change point detection methods, and is also an effective strategy for developing change point methods for multivariate data (Chenouri et al., 2020). The same strategy can be used for developing change point detection methods for a sequence of data lie on the manifold space.

Consider a sequence of random objects $X_1, X_2, ..., X_n$ taking values in a manifold space $\mathcal{M}$. Let us calculate the depth of the random objects with respect to the measure P, i.e. $D(P, X_1), D(P, X_2), ..., D(P, X_n)$.

The goal is to test the null hypothesis that

$$H_0 : X_1, X_2, ..., X_n \text{ are identically distributed}$$

against the alternative hypothesis

$$H_A : \begin{cases} X_1, X_2, ..., X_{[n\theta]} \sim P_1 \\ X_{[n\theta]+1}, X_{[n\theta]+2}, ..., X_n \sim P_2 \end{cases}$$

where $\theta \in [0, 1]$.

Let us denote the depth ranks by $R_1, R_2, ..., R_n$. Suppose that there is a change point at breaking fraction θ. Given that the depth of random objects $X_1, X_2, ..., X_{[n\theta]}$ are on average smaller/larger than the depths of $X_{[n\theta]+1}, X_{[n\theta]+2}, ..., X_n$ and thus the rank of depths $R_1, R_2, ..., R_{[n\theta]}$ will be larger/smaller than $R_{[n\theta]+1}, R_{[n\theta]+2}, ..., R_n$. Furthermore, under the null hypothesis of no changes, the ranks R_i are uniformly distributed on the integers between 1 and n, regardless of the distribution of random objects X_i.

We have

$$\mathbb{E}[R_i] = \frac{n+1}{2} \quad , \quad \text{Var}[R_i] = \frac{n^2-1}{12} \, .$$

The rank-CUSUM test statistic can be formulated as

$$Z_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nu]} \frac{R_i - (n+1)/2}{\sqrt{(n^2-1)/12}} \, ,$$

where $u \in [0, 1]$.

The test statistic for change point detection can be formulated as

$$T_n = \sup_{0<u<1} \mid Z_n(u) \mid \, ,$$

which is the maximally selected two sample Wilcoxon test statistic over all possible breaking fractions applied to data depths.

Theorem 14. *Under H_0, the test statistic T_n converges in distribution to the maximum of the absolute value of Brownian Bridge*

$$T_n \xrightarrow{D} \sup_{0<u<1} \mid \mathcal{B}(u) \mid \ .$$

The rejection region is

$$R_{n,\alpha} = \{T_n > q_{1-\alpha}\} \ ,$$

where $q_{1-\alpha}$ denotes the significance level for the process $\sup_{0<u<1} \mid \mathcal{B}(u) \mid$.

The next step is to investigate the asymptotic consistency of the test statistic under at-most one change point alternative.

Under H_A and under the independence of observations, it can be seen that the empirical measure P_n based on the whole sample converges in sup-norm by Glivenko-Cantelli theorem to

$$P_* = \alpha P_1 + (1-\alpha) P_2 \ .$$

Given the fact that the test statistic T_n is based on the depth-ranks, the condition for consistency naturally depends on the distribution of depths themselves. We define $H_1(\epsilon) = \Pr(D(P_*, X) \leq \epsilon)$ and $H_2(\epsilon) = \Pr(D(P_*, Y) \leq \epsilon)$, where $X \sim P_1$ and $Y \sim P_2$. Assume the following continuity condition holds.

Assumption 9. *$H_1(\epsilon)$ and $H_2(\epsilon)$ are each Lipschitz with constant C, implying that for all $\epsilon_1, \epsilon_2 \in \mathbb{R}$ we have*

$$\mid H_i(\epsilon_1) - H_i(\epsilon_2) \mid \leq C \mid \epsilon_1 - \epsilon_2 \mid \quad \textit{for} \ \ i = 1, 2 \ .$$

Theorem 15. *Suppose that H_A, and Assumption 9 hold. If*

$$\int_{-\infty}^{\infty} \{\alpha H_1(u) + (1-\alpha) H_2(u)\} \, dH_1(u) \neq \frac{1}{2}, \tag{3.1}$$

then $T_n \xrightarrow{P} \infty$.

Under H_A, it is often of interest to estimate the change point k^* or equivalently the corresponding breaking fraction θ. Natural estimators of k^* and θ are

$$\widehat{k^*} = \min \left\{ k \ : \mid Z_n(\lfloor k/n \rfloor) \mid = \sup_{0 \leq t \leq 1} \mid Z_n(t) \mid \right\}, \ \text{ and } \ \widehat{\theta} = \frac{\widehat{k^*}}{n},$$

The following theorem proves that the above estimators are consistent.

Theorem 16. *Under the condition of Theorem 15,* $\mid \widehat{\theta} - \theta \mid = o_p(1)$.

Theorem 17. *Consider the same assumptions made in Theorem 15, then, under* H_A*, the following consistency result holds*

$$a_n^{-1}\Big(Z_n(\widehat{\theta}) - Z_n(\theta)\Big) = o_p(1) \ .$$

where $a_n \to \infty$ *and* $\dfrac{a_n}{n} \to 0$*, as* $n \to \infty$.

It is also important to derive nonasymptotic sharp bounds for the estimator. This result is given in the following theorem.

Theorem 18. *Suppose that the condition of Theorem 15 is satisfied. Also, assume that* $p = \Pr(D(P_1, X_1) < D(P_n, X_n)) \neq \frac{1}{2}$*, for* n *i.i.d. sample from manifold* $\mathcal{M}$*. Then, the following nonasymptotic bound for* $\widehat{k}^*$ *can be achieved*

$$\Pr\left(\left|\widehat{k}^* - k^*\right| > tn\right) < c_1 \exp\left(-c_2 n t^2 (p - 1/2)^2\right) \ ,$$

where $c_1 > 0$ *and* $c_2 > 0$ *are constants.*

3.4 Multiple Change Point Analysis

In this section, we will extend the formulation of the single change point detection problem to a multiple change point detection problem. Consider $X_1, X_2, ..., X_n$ taking values in a manifold space $\mathcal{M}$ with null hypothesis

$$H_0 : X_1, X_2, ..., X_n \text{ are identically distributed}$$

against the alternative hypothesis

$$H_A : X_{\lfloor n\tau_i \rfloor}, ..., X_{\lfloor n\tau_{i+1} \rfloor} \text{ for } i = 0, ..., \ell \text{ are change point locations} \ ,$$

where ℓ is an integer representing the number of change points, and $\tau_i \in [0, 1]$ values are breaking fractions such that $\tau_0 = 0$ and $\tau_{\ell+1} = 1$.

We can calculate the depth of these random objects with respect to the measure P, i.e. $D(P, X_1), D(P, X_2), ..., D(P, X_n)$. We can accordingly calculate the rank of the depth values $R_1, R_2, ..., R_n$. We can use binary segmentation (BS) strategy to detect multiple

change points. At the first stage, the rank-CUSUM test statistic is used to search for the first change point

$$Z_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nu]} \frac{R_i - (n+1)/2}{\sqrt{(n^2-1)/12}} \,,$$

where $u \in [0,1]$ and

$$\mathbb{E}[R_i] = \frac{n+1}{2} \quad , \quad \text{Var}[R_i] = \frac{n^2-1}{12} \,.$$

The test statistic for detecting the possible change point can be formulated as the maximally selected two sample Wilcoxon sum rank test over all possible break fractions

$$T_n = \max_{0<t<1} \mid Z_n(t) \mid \quad .$$

It can be easily shown that under the hypothesis of no change the above statistic converges in distribution to the maximum absolute value of the Brownian Bridge process, which can be used to obtain the thresholds for the detection and estimation of change points. See for example Chenouri et al. (2020).

The estimated change point location can be then obtained by

$$\widehat{k}^* = \min \left\{ k \,; \;\; \left| Z_n \left(\frac{k}{n} \right) \right| > c_\alpha \right\} \,, \tag{3.2}$$

where c_α is the critical value at the $1-\alpha$ level of significance. This procedure can be easily extended to multiple change point detection problems by using the binary segmentation strategy. Also, to update the critical value for multiple change point scenario, Bonferroni correction is taken into account. In this way, after the i^{th} change point is detected, α_i is obtained by $\alpha_i = 1 - (1-\alpha)^{\frac{1}{i+1}}$. This controls the significance level to be kept at α thorough the entire process.

Once the first change point X_{τ_1} is detected, the sample sequence is divided into two segments, i.e. $X_1, ..., X_{\lfloor n\tau_1 \rfloor}$ and $X_{\lfloor n\tau_1 \rfloor + 1}, ..., X_n$ and the two segments are explored for change points. The processes of binary segmentation and change point detection is repeated until no more change point can be detected.

Theorem 19. *Let $\widehat{\ell}$ denote the number of change points estimated by rank binary segmentation and*

$$\mathcal{B}_n := \{ \widehat{\ell} = \ell, \max_{i=1,...,\ell} \mid \widehat{k}_i^* - k_i^* \mid \leq C n^\alpha \} \,,$$

where $\dfrac{1}{2} < \alpha < 1$, *and C is a positive constant. Then* $\Pr(\mathcal{B}_n) \to 1$ *as* $n \to \infty$.

Theorem 20. *Suppose that the conditions of Theorem 18 are satisfied. Let ℓ denote the number of change points. The following nonasymptotic bound for $\widehat{k}_i^*$, $i = 1, \ldots, \ell$ can be achieved*

$$\Pr\left(\max_{i=1,\ldots,\ell}\left|\widehat{k}_i^* - k_i^*\right| > tn\right) < c_1 \exp\left(-c_2 n t^2 (p - 1/2)^2\right) \ ,$$

where $c_1 > 0$ and $c_2 > 0$ are constants.

The pseudo-code of rank-CUSUM binary segmentation is given in Algorithm 1.

The second proposed method is based on the combination of rank-CUSUM statistic and the wild binary segmentation strategy. Proposed by Fryzlewicz (2014), wild binary segmentation is a modified version of binary segmentation which is based on a random selection of M_N arbitrary intervals in each segment to find a potential change point. The main advantage of such a strategy lies in its capability to estimate change points that are not necessarily dominant and work against each other. See Figure 1 of Fryzlewicz (2014) and the respective discussion therein. In such cases, narrower intervals are required to make change points detectable. Another advantage of wild binary segmentation over binary segmentation is the use of optimization, which gives the location of the change point $\widehat{k}^*$ with the maximum violation of threshold instead of simply using Eq.(3.2). It is also worth mentioning that the computational time required for solving the optimization problem in wild binary segmentation is not considerable compared to methods using a dynamic programming algorithm (Fryzlewicz, 2014).

Let $\mathbb{M}_{s,e}$ be the set of M_n randomly drawn intervals $[s_m, e_m]$, $m = 1, \ldots, M_n$ from the segment $[s, e]$, which is constructed from sample points, where the set of M_n starting and ending points are drawn independently with replacement from discrete Uniform distribution, and satisfy the constraints $s_m \in \{s, \ldots, e-1\}$ and $e_m \in \{s_m + 1, \ldots, e\}$, for $m \in \{1, \ldots, M_n\}$. Then, the algorithmic implementation of rank wild binary segmentation can be given as follows.

Start by forming the set $\mathbb{M}_{s,e}$ which contains M_n intervals from the $\{X_s, \ldots, X_e\}$. Calculate the value of the statistic for each of the M_n intervals, and define

$$(m^*, \widehat{k}_{m^*}^*) = \underset{\substack{k_m \in [s_m, e_m] \\ m \in \{1, \ldots, M_n\}}}{\operatorname{argmax}} \left\{\left|Z_n\left(\frac{k_m}{n}\right)\right|\right\} \ .$$

Then, if $|Z_n(\widehat{k}_{m^*}^*/n)| > c_\alpha$, $\widehat{k}_1^* = k_m$, and the sample is divided into two sub-sections $\{X_1, \ldots, X_{\widehat{k}_1^* - 1}\}$ and $\{X_{\widehat{k}_1^*}, \ldots, X_n\}$. Again, the set $\mathbb{M}_{s,e}$ containing M_n intervals is formed

Algorithm 1: Pseudo-code of rank-CUSUM binary segmentation algorithm

$i \leftarrow 1$;
$\mathcal{K} = \varnothing,\ \mathcal{K}' = \varnothing$;
$\mathcal{S}_i = \{X_1, \ldots, X_n\},\ \mathcal{S}' \leftarrow \mathcal{S}_i$;
Apply Wilcoxon sum rank test over all break fractions of $\mathcal{S}_i$;
Calculate $\widehat{k}_i^*$ using Eq. 3.2;
$\mathcal{K} \leftarrow \mathcal{K} \cup \widehat{k}_i^*$;
$\mathcal{K}' \leftarrow \widehat{k}_i^*$;
while $\mathcal{K}' \neq \varnothing$ **do**
    $\mathcal{K}'' \leftarrow \mathcal{K}',\ \mathcal{S}'' \leftarrow \mathcal{S}'$;
    $\mathcal{K}' = \varnothing, \mathcal{S}' = \varnothing$;
    for *j = 1 : size($\mathcal{K}''$)* **do**
        $S_j \leftarrow \mathcal{S}''[j]$;
        $s \leftarrow$ index $\mathcal{S}_j(1)$;
        $e \leftarrow$ index $\mathcal{S}_j(end)$;
        $\mathcal{S}_{2j} \leftarrow \mathcal{S}_j(s : \widehat{k}_j^*)$;
        $\mathcal{S}_{2j+1} \leftarrow \mathcal{S}_j(\widehat{k}_j^* + 1 : e)$;
        Apply Wilcoxon sum rank test over all break fractions of $\mathcal{S}_{2j}$ and $\mathcal{S}_{2j+1}$;
        Calculate $\widehat{k}_{2j}^*$ and $\widehat{k}_{2j+1}^*$ using Eq. 3.2;
        $\mathcal{K}' \leftarrow \{\widehat{k}_{2j}^*, \widehat{k}_{2j+1}^*\}$;
        $\mathcal{K} \leftarrow \mathcal{K} \cup \{\widehat{k}_{2j}^*, \widehat{k}_{2j+1}^*\}$;
        if $\widehat{k}_{2j}^* \neq \varnothing$ **then**
            $\mathcal{S}' \leftarrow \mathcal{S}' \cup \mathcal{S}_{2j}$;
        end
        if $\widehat{k}_{2j+1}^* \neq \varnothing$ **then**
            $\mathcal{S}' \leftarrow \mathcal{S}' \cup \mathcal{S}_{2j+1}$;
        end
    end
end
Sort the members of $\mathcal{K}$ and report the ordered breaking fractions;

for each of these segments, and calculations are repeated. This iterative search is continued until no more change point can be detected. The locations of these change points are sorted, and the ordered breaking fractions $\widehat{\theta}_i = \widehat{k}_i^*/n$, for $i = 1, \ldots, \widehat{\ell}$ are reported.

Theorem 21. *Suppose the assumptions of Theorem 19 hold. Let $M_n = O(\log n)$, where M_n is the number of sub-intervals randomly drawn from a uniform distribution in each of the segments. Denote the number of change points estimated via rank-wild binary segmentation by $\widehat{\ell}$ and*

$$\mathcal{D}_n := \bigcap_{i=1}^{\ell} \bigcup_{m_i=1}^{M_n} \{[s_{m_i}, e_{m_i}] \in \mathcal{I}_i\},$$

where $\mathcal{I}_i$ is the set of all sub-intervals that only include i^{th} change point. Then, we have $\Pr(\mathcal{B}_n \mid \mathcal{D}_n) \to 1$, as $n \to \infty$, where $\mathcal{B}_n$ is defined in Theorem 19.

Theorem 22. *Assume that the conditions of Theorem 18 hold. Let $M_n = O(\log n)$ and ℓ denote the number of change points. The following nonasymptotic bound for $\widehat{k}_i^*$, $i = 1, \ldots, \ell$ can be achieved*

$$\Pr\left(\max_{i=1,\ldots,\ell} \left|\widehat{k}_i^* - k_i^*\right| > tn\right) < c_1 \exp\left(-c_2 n t^2 (p - 1/2)^2\right) \ ,$$

where $c_1 > 0$ and $c_2 > 0$ are constants.

The pseudo-code of rank-wild binary segmentation is given in Algorithm 2.

3.5 Robustness of Rank-Depth Change Point Detection Methods

In this section, our goal is to study the robustness of the proposed rank-depth change point detection methods using q-robustness (Huber, 2011) and breakdown point for sequential processes (He et al., 1990; Zhang et al., 2022; Zhang, 2025).

Qualitative Robustness of Change Point Detection Methods

We probe the robustness of a sequence of rank-depth change point detection statistics, denoted by T_n, via equicontinuity of error probabilities using the total variation metric, bounded Lipschitz metric and the geometrical features of manifold spaces.

Algorithm 2: Pseudo-code of rank-wild binary segmentation algorithm

initialize M_n;
$i \leftarrow 1$;
$\mathcal{K} = \varnothing,\ \mathcal{K}' = \varnothing$;
$\mathcal{S}_i = \{X_1, \ldots, X_n\},\ \mathcal{S}' \leftarrow \mathcal{S}_i$;
Apply rank-wild binary segmentation on $\mathcal{S}_i$;
Calculate $\widehat{k}_i^*$;
$\mathcal{K} \leftarrow \mathcal{K} \cup \widehat{k}_i^*$;
$\mathcal{K}' \leftarrow \widehat{k}_i^*$;
while $\mathcal{K}' \neq \varnothing$ **do**
  $\mathcal{K}'' \leftarrow \mathcal{K}',\ \mathcal{S}'' \leftarrow \mathcal{S}'$;
  $\mathcal{K}' = \varnothing,\ \mathcal{S}' = \varnothing$;
  for *j = 1 : size($\mathcal{K}''$)* **do**
    $S_j \leftarrow \mathcal{S}''[j]$;
    $s \leftarrow$ index $\mathcal{S}_j(1)$;
    $e \leftarrow$ index $\mathcal{S}_j(end)$;
    $\mathcal{S}_{2j} \leftarrow \mathcal{S}_j(s : \widehat{k}_j^*)$;
    $\mathcal{S}_{2j+1} \leftarrow \mathcal{S}_j(\widehat{k}_j^* + 1 : e)$;
    Apply rank-wild binary segmentation on $\mathcal{S}_{2j}$ and $\mathcal{S}_{2j+1}$;
    Calculate $\widehat{k}_{2j}^*$ and $\widehat{k}_{2j+1}^*$;
    $\mathcal{K}' \leftarrow \{\widehat{k}_{2j}^*, \widehat{k}_{2j+1}^*\}$;
    $\mathcal{K} \leftarrow \mathcal{K} \cup \{\widehat{k}_{2j}^*, \widehat{k}_{2j+1}^*\}$;
    if $\widehat{k}_{2j}^* \neq \varnothing$ **then**
      $\mathcal{S}' \leftarrow \mathcal{S}' \cup \mathcal{S}_{2j}$;
    end
    if $\widehat{k}_{2j+1}^* \neq \varnothing$ **then**
      $\mathcal{S}' \leftarrow \mathcal{S}' \cup \mathcal{S}_{2j+1}$;
    end
  end
end
Sort the members of $\mathcal{K}$ and report the ordered breaking fractions;

Consider a random element X on manifold $\mathcal{M}$ as a measurable map on a probability space $(\Omega, \mathcal{F}, P)$ to $(\mathcal{M}, \mathcal{S})$, where $\mathcal{S}$ denotes the Borel sigma-algebra of $\mathcal{M}$. For $X \sim P$, the distribution of T_n is represented by $\mathcal{G}_P(T_n)$. For two given probability measures P and Q defined on $\mathcal{S}$, a sequence of the maximal CUSUM statistic T_n is q-robust at P if for any $\epsilon > 0$ there is $\delta > 0$ such that

$$\delta_{TV}(P, Q) \leq \delta \quad \text{implies} \quad \delta_{TV}(\mathcal{G}_P(T_n), \mathcal{G}_Q(T_n)) \leq \epsilon \ ,$$

for all sufficiently large n. Let T be a functional (mapping) for all $P \in \mathcal{P}(\mathcal{M})$ so that i.e. $T_n = T \mid P_n$. For functionals T defined on $\mathcal{P}(\mathcal{M})$, the idea of q-robustness can be related to weak continuity (Huber, 2011).

Let P_n denote the empirical probability supported by points $x_1, \ldots, x_n$. For the change point detection method

$$T_n(x_1, \ldots, x_n) = T(P_n) \ ,$$

we have q-robustness as a result of weak continuity.

Assumption 10. *The representation of T_n by functional T is regular such that (i) it is consistent for any P in the domain of functional T, i.e. $T_n \xrightarrow{p} T(P)$, and (ii) for any $\tau \in T(P)$ there exists a sequence of empirical probability measures P_n converging to P so that $T(P_n)$ converges to τ.*

Similar to Chapter 2, we can show that the empirical and theoretical versions of the proposed change point detection methods can be written in the form of functionals $T \mid P_n$ and $T \mid P$. Hence, it can be verified that the regularity conditions given in Assumption 10 hold true.

Theorem 23. *Under the conditions of Assumption 10, the proposed rank-depth change point detection methods T_n are q-robust.*

Breakdown Point of Change Point Detection Statistic

We can formulate the change point detection scheme with contaminated data using Huber's gross error (Huber, 2011). Let P and Q two probability measures and define the following ϵ-contaminated probability measure classes

$$\mathcal{H}_{0,\epsilon} = \{(1-\epsilon)P + \epsilon G_0 \mid G_0 \in \mathcal{P}(\mathcal{M})\} \ ,$$
$$\mathcal{H}_{1,\epsilon} = \{(1-\epsilon)Q + \epsilon G_1 \mid G_1 \in \mathcal{P}(\mathcal{M})\} \ .$$

We assume data points are generated from $\mathcal{H}_{0,\epsilon}$, and at some time point t_{k^*}, the distribution of the data changes to $\mathcal{H}_{1,\epsilon}$. We can represent our sequential change point detection problem as a statistical procedure with stopping time T_{k^*}, which is the time we trigger an alarm and declare that a change has occurred. T_{k^*} is a random variable and decision $\{T_{k^*} = t_{k^*}\}$ depends only on the observations collected in the first t_{k^*} time steps. Let $P_{H_0}^{(\infty)}$ and $\mathbb{E}_{H_0}^{(\infty)}$ denote the probability and expectation under no change. Also, let $P_{H_0,H_1}^{(k^*)}$ and $\mathbb{E}_{H_0,H_1}^{(k^*)}$ denote the probability and expectation when change occurred at time t_{k^*} and data have distribution $\mathcal{H}_{1,\epsilon}$.

In (Zhang et al., 2022; Zhang, 2025), the authors extended the concept of level breakdown point and power breakdown point in offline hypothesis testing to sequential change point detection. Here, we adopt these definitions for analyzing the robustness of the proposed rank-depth change point detection methods.

The false-alarm breakdown point ε_F is defined as

$$\varepsilon_F = \min\left\{0 \le \epsilon \le 1 \Big| \lim_{\mathbb{E}_P^{(\infty)} \to \infty} \inf_{H_0 \in \mathcal{H}_{0,\epsilon}} \frac{\mathbb{E}_{H_0}^{(\infty)}(T_{k^*})}{\mathbb{E}_P^{(\infty)}(T_{k^*})} = 0\right\} .$$

The worst-case detection delay is defined as

$$\mathrm{WADD}_{H_0,H_1} = \sup_{t_{k^*} \ge 1} \sup \mathbb{E}_{H_0,H_1}^{(k^*)}\left((T_{k^*} - t_{k^*} + 1)^+ \mid X_1, \ldots, X_{t_{k^*}-1}\right) ,$$

where $H_0 \in \mathcal{H}_{0,\epsilon}$ and $H_1 \in \mathcal{H}_{1,\epsilon}$. The detection-delay breakdown point ε_D is defined as

$$\varepsilon_D = \min\left\{0 \le \epsilon \le 1 \Big| \lim_{\mathbb{E}_P^{(\infty)} \to \infty} \sup_{H_1 \in \mathcal{H}_{1,\epsilon}} \frac{\mathrm{WADD}_{P,H_1}(T_{k^*})}{\mathrm{WADD}_{P,Q}(T_{k^*})} = \infty\right\} .$$

The above results are derived for the case that we have knowledge of the true distribution P for $X_1, \ldots, X_n$, i.e. the statistics can be written as a function of true distribution $T(P)$. For the case that we use empirical distribution P_n, the asymptotic breakdown point for $T(P_n)$ is at-least as large as that of $T(P)$.

Theorem 24. *Suppose that the condition of Assumption 8 is satisfied. Let ε_1^* denote the asymptotic breakdown point for functional $T(P)$, and ε_2^* denote the asymptotic breakdown point for functional $T(P_n)$. It can be shown that $\varepsilon_2^* \ge \varepsilon_1^*$.*

3.6 Simulation Study

In this section, we perform Monte Carlo simulation to evaluate the performance of the proposed rank-depth change point detection methods considering various types of manifolds. For single change point scenario, we consider at-most-one change point (AMOC) method from package 'changepoint', nonparametric ECP method (James and Matteson, 2015), and d-CUSUM which is a Riemannian change point detection method. For multiple change point scenario, we consider pruned exact linear time (PELT) method from R package 'changepoint' and ECP method. Also, standard BS and WBS are considered to evaluate the performance of rank-depth tests r-BS and r-WBS. Implementation of BS, WBS and d-CUSUM are based on robust centroid estimation proposed in (Wang et al., 2025).

3.6.1 Single Change Point Detection

Example 1 (Wishart Distribution): Consider the manifold space of positive definite symmetric matrices

$$SPD(k) = \{A \in \mathbb{R}^{k\times k} \mid A = A^\top, \text{ and } x^\top Ax > 0\} \ .$$

The probability density function of standard Wishart distribution with N degrees of freedom is defined as

$$f(A) = \frac{1}{2^{Nk/2} \mid V \mid^{N/2} \Gamma_k(N/2)} \mid A \mid^{N-k-1/2} \exp\big(-\frac{1}{2}\text{tr}(V^{-1}A)\big) \ ,$$

where $\mid A \mid$ is the determinant of $A \in SPD(k)$, and $\Gamma_k(N/2)$ is the multivariate Gamma function, and V denote a fixed $k \times k$ symmetric positive definite matrix.

Geodesic depth GD is used for calculating the depth of the object data. For this simulation, we consider rank-based CUSUM (r-CUSUM) and standard CUSUM based on the depth of the data (d-CUSUM). To evaluate the performance of the change point detection statistics, we consider sample sizes $n = \{50, 100, 200\}$ and breaking fractions $\theta = \{0.25, 0.50, 0.75\}$.

The simulation is performed under both null and alternative scenarios to evaluate the power of the CUSUM statistics. For the alternative scenario, the distributions for the two fractions are

$$X_i \sim W_3(20I_2), \quad Y_i \sim W_{20}(6I_2) \ .$$

Table 3.1: Empirical size under the null for Wishart distribution for 5000 simulation runs.

		d-CUSUM	r-CUSUM	AMOC	ECP
n = 50	**fraction = 0.25**	0.03	0.03	0.04	0.03
	fraction = 0.50	0.02	0.03	0.01	0.03
	fraction = 0.75	0.05	0.04	0.04	0.03
n = 100	**fraction = 0.25**	0.06	0.05	0.03	0.05
	fraction = 0.50	0.05	0.05	0.01	0.03
	fraction = 0.75	0.04	0.05	0.02	0.04
n = 200	**fraction = 0.25**	0.04	0.04	0.03	0.02
	fraction = 0.50	0.05	0.02	0.01	0.03
	fraction = 0.75	0.05	0.05	0.02	0.01

Table 3.2: Empirical power under the alternative for Wishart distribution for 5000 simulation runs.

		d-CUSUM	r-CUSUM	AMOC	ECP
n = 50	**fraction = 0.25**	1.00	1.00	0.92	0.87
	fraction = 0.50	1.00	1.00	0.98	0.83
	fraction = 0.75	0.93	0.91	0.95	0.91
n = 100	**fraction = 0.25**	1.00	1.00	0.91	0.96
	fraction = 0.50	1.00	1.00	0.90	0.94
	fraction = 0.75	0.95	0.92	0.92	0.93
n = 200	**fraction = 0.25**	1.00	1.00	1.00	1.00
	fraction = 0.50	1.00	1.00	1.00	1.00
	fraction = 0.75	1.00	1.00	1.00	1.00

Simulation results for the null scenario are given in Table 3.1. It can be seen that the empirical size is kept at the significant level $\alpha = 0.05$ for the considered test statistics. All of these statistics show very close performance regardless of the sample size and breaking fraction. Table 3.2 reports the results for the alternative scenario. It can be seen that rank-depth methods show very good performance for all sample sizes and breaking fractions and can outperform the rival methods. Clearly, increasing sample size can improve the performance of these methods. Figure 3.1 presents the summary statistics for the estimated locations for the proposed rank-depth methods. Obviously, both methods show very close performance and most of the times estimated the correct location of change regardless of sample size and breaking fraction. It can be concluded that these methods can be reliably applied to change point detection problems for object data belonging to SPD manifold. Furthermore, it is observed that the methods show very good power for small sample sizes indicating that asymptotic results can be reliably used for finite sample simulations. Our simulations indicated that the rank-depth methods can estimate the location of changes with smaller error compared to AMOC and ECP.

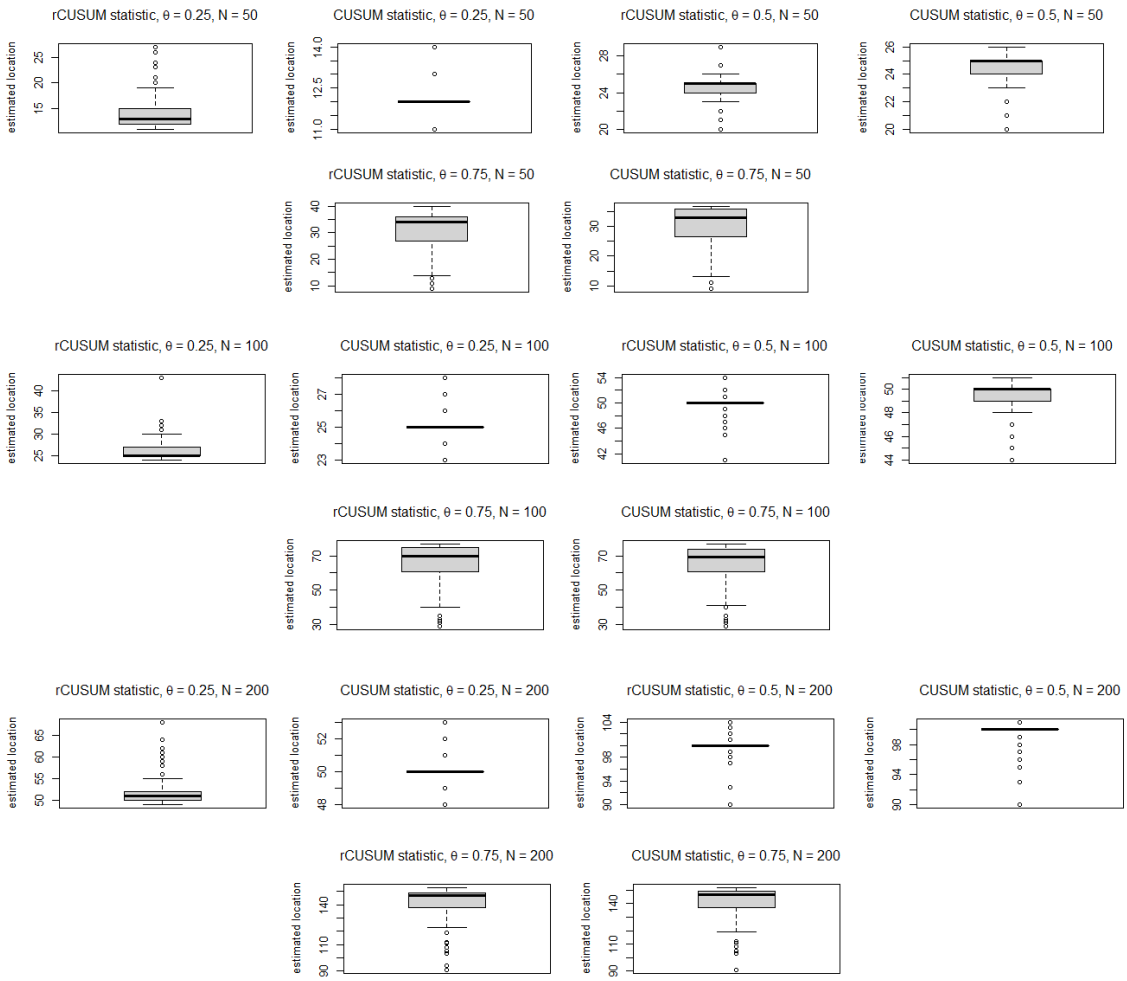


Figure 3.1: Statistical summaries for the estimated locations of CUSUM and r-CUSUM for Wishart distribution with breaking fractions $\theta = \{0.25, 0.50, 0.75\}$ and sample sizes $n = \{50, 100, 200\}$.

Example 2 (pdf Cruve Space with Wasserstein Distance): In this example, we consider probability density data objects to evaluate the performance of the considered change point detection methods. Probability density functions are commonly used in the literature as object data for evaluating the performance of the statistical methods developed for manifold spaces (Dryden et al., 2009).

The considered dataset are pdf's of univariate normal distribution $\mathrm{N}(\mu, 1)$, where each

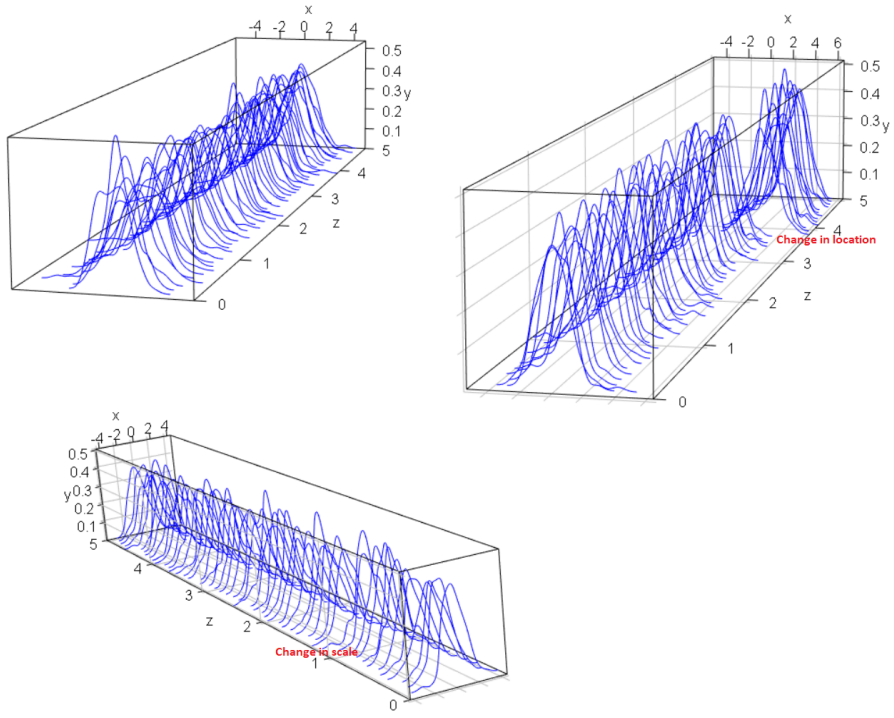


Figure 3.2: Probability density function data objects for no change, change in location and change in scale scenarios.

datum is $\mathrm{N}(\mu, 1)$. The simulation is performed under no change, change in location, and change in scale scenarios. GD is used to calculate the depth of the data objects. To calculate the distance of objects in the space of pdfs, we can use 2-Wasserstein distance between measures ν_1 and ν_2

$$W_2(\nu_1, \nu_2) = \Big(\int_0^1 \big(F_1^{-1}(q) - F_2^{-1}(q)\big)^2 dq \Big)^2 ,$$

where F_1^{-1} and F_2^{-1} are quantile functions of ν_1 and ν_2, respectively.

To investigate change in location, we generate the first data segment using $\mathrm{N}(0, 0.5^2)$ and the second data segment using $\mathrm{N}(2, 0.5^2)$. For the change in scale scenario, we generate the first segment from $\mathrm{N}(0, 1.5^2)$ and the second segment from $\mathrm{N}(0, 0.75^2)$. The considered sample sizes and breaking fractions are $n = \{50, 100, 200\}$ and $\theta = \{0.25, 0.50, 0.75\}$, respectively. Figure 3.2 indicates the profiles together with the change point locations. The depicted panels include the case of no change, change in location and change in scale scenarios.

Table 3.3: Empirical size under the null for pdf curve data for 5000 simulation runs.

		d-CUSUM	r-CUSUM	AMOC	ECP
n = 50	**fraction = 0.25**	0.04	0.04	0.03	0.04
	fraction = 0.50	0.06	0.05	0.03	0.05
	fraction = 0.75	0.05	0.06	0.03	0.04
n = 100	**fraction = 0.25**	0.06	0.06	0.02	0.03
	fraction = 0.50	0.02	0.03	0.03	0.04
	fraction = 0.75	0.04	0.03	0.03	0.04
n = 200	**fraction = 0.25**	0.06	0.04	0.05	0.03
	fraction = 0.50	0.02	0.06	0.02	0.03
	fraction = 0.75	0.04	0.05	0.02	0.06

Table 3.4: Empirical power under the change in location alternative for pdf curve data for 5000 simulation runs.

		d-CUSUM	r-CUSUM	AMOC	ECP
n = 50	**fraction = 0.25**	1.00	1.00	0.85	0.91
	fraction = 0.50	0.97	0.99	0.93	0.89
	fraction = 0.75	1.00	1.00	0.92	0.89
n = 100	**fraction = 0.25**	1.00	1.00	0.88	0.92
	fraction = 0.50	1.00	1.00	0.89	0.92
	fraction = 0.75	0.99	1.00	0.93	0.85
n = 200	**fraction = 0.25**	1.00	1.00	0.91	0.93
	fraction = 0.50	1.00	1.00	0.89	0.91
	fraction = 0.75	1.00	1.00	0.88	0.91

Table 3.5: Empirical power under the change in scale alternative for pdf curve data for 5000 simulation runs.

		d-CUSUM	r-CUSUM	AMOC	ECP
n = 50	**fraction = 0.25**	0.95	0.94	0.92	0.88
	fraction = 0.50	0.86	0.88	0.92	0.89
	fraction = 0.75	0.95	0.95	0.92	0.88
n = 100	**fraction = 0.25**	0.93	0.94	0.87	0.89
	fraction = 0.50	0.88	0.90	0.88	0.91
	fraction = 0.75	0.97	0.95	0.86	0.92
n = 200	**fraction = 0.25**	0.94	0.92	0.81	0.88
	fraction = 0.50	0.92	0.90	0.81	0.82
	fraction = 0.75	0.96	0.92	0.88	0.85

Table 3.3 tabulates the results for the case of no change for 100 simulation runs. It can be seen that the empirical size is kep at 0.05 level for all sample sizes and breaking fractions. The performance of the considered statistics are comparable. Table 3.4 tabulates

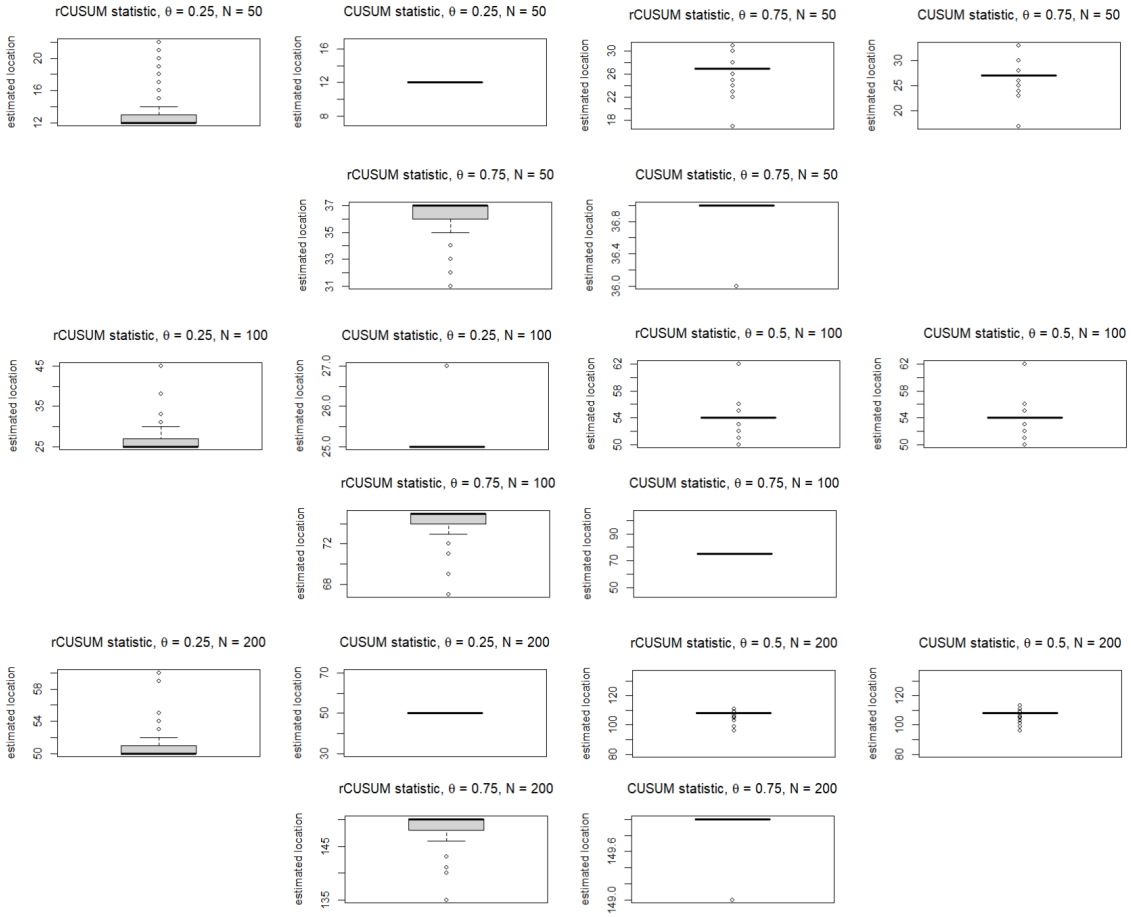


Figure 3.3: Statistical summaries for the estimated locations of CUSUM and r-CUSUM for pdf curve data under change in location scenario with breaking fractions $\theta = \{0.25, 0.50, 0.75\}$ and sample sizes $n = \{50, 100, 200\}$.

the results for the case of change in location. As seen, the empirical power of the proposed test statistics are close or equal to 1 for all of the cases; however, AMOC and ECP show inferior performance in terms of power. Table 3.5 tabulates the simulation results for the case of change in scale. It can be seen that the performance is not as good as the case of change in location. This is in agreement with our observation from Figure 3.2 where it is observed that that pdf curves for the case of change in scale do not differ much in shape compared to the case of change in location. However, this fact does not significantly undermine the performance of the proposed statistics as the empirical powers are close

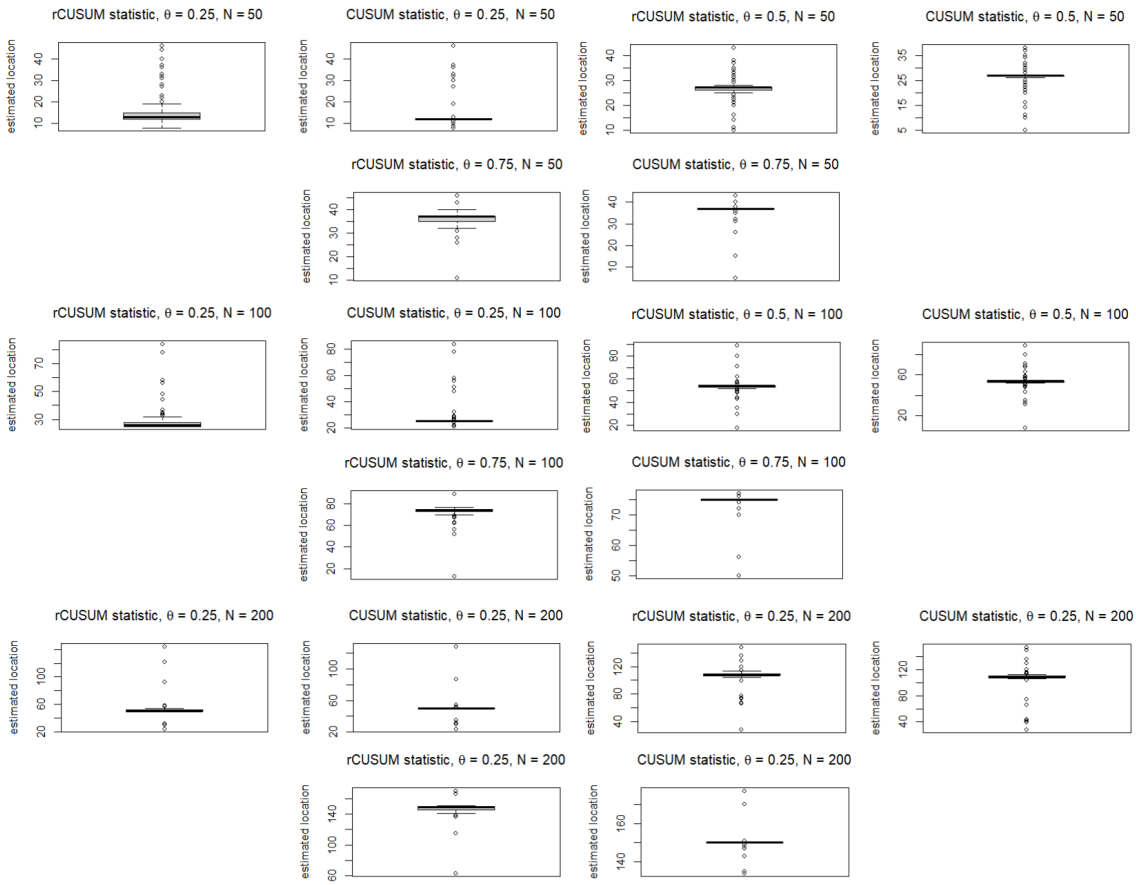


Figure 3.4: Statistical summaries for the estimated locations of CUSUM and r-CUSUM for pdf curve data under change in scale scenario with breaking fractions $\theta = \{0.25, 0.50, 0.75\}$ and sample sizes $n = \{50, 100, 200\}$.

to 0.95 for most of the cases. The above results demonstrate the capability of the test statistic to be applied to both change in location and scale scenarios. Figures 3.3 and 3.4 indicate the boxplots for the estimated locations. It is obvious that for most of the cases, the median and mode of the estimated locations is the same as the true location point. Simulation results indicated that the estimation error of AMOC and ECP was larger than d-CUSUM and r-CUSUM for the change location estimates.

Example 3 (Finland Fertility Data): Finland fertility database provides a cohort fertility data which is commonly used in the literature for investigating the evolution of fertility in Finland for the period of 1939 to 2021. The dataset consists of age-based total

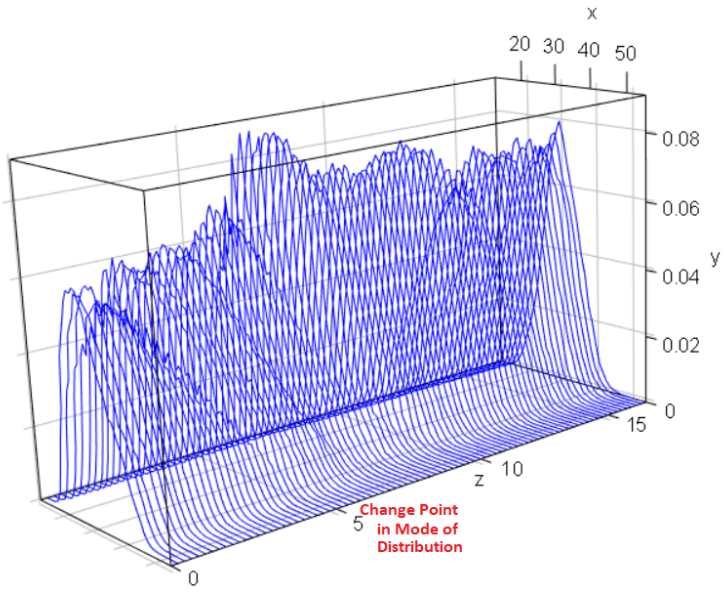


Figure 3.5: Yearly maternal age probability density functions for the age interval $[12, 55]$ for the period of 1939 to 2021.

birth counts aggregated in a yearly fashion for the age interval $[12, 55]$. The preprocessing is carried out by collecting these data in the form of a histogram and smoothing them in the form of probability density functions.

These pdfs are treated as object data and are used for detecting possible changes in the distribution of yearly age-based birth rate pdf curves. Figure 3.5 indicate the pdf object data. It is obvious that there is a change in the mode of the distribution of age-based birth rates at a point around 1982. We use r-CUSUM and d-CUSUM with 2-Wasserstein distance to detect changes in the distribution of the pdf curve data.

Figure 3.6 indicates the absolute value of Z-statistic curves for the considered change point detection algorithms. It can be verified that the maximum of the Z-statistic curves for r-CUSUM and d-CUSUM statistics are located at years 1981 and 1983, respectively. The estimated change points are in a very good agreement with the true change point location, i.e. 1982, observed from the evolution of curves in Figure 3.5.

It can be concluded that the considered methods are powerful for detecting change in the distributional characteristics such as the mode of the distribution. Given the results obtained in the previous example where the efficacy of the methods were verified for change in location and scale, it can be concluded that the proposed test statistics are reliable

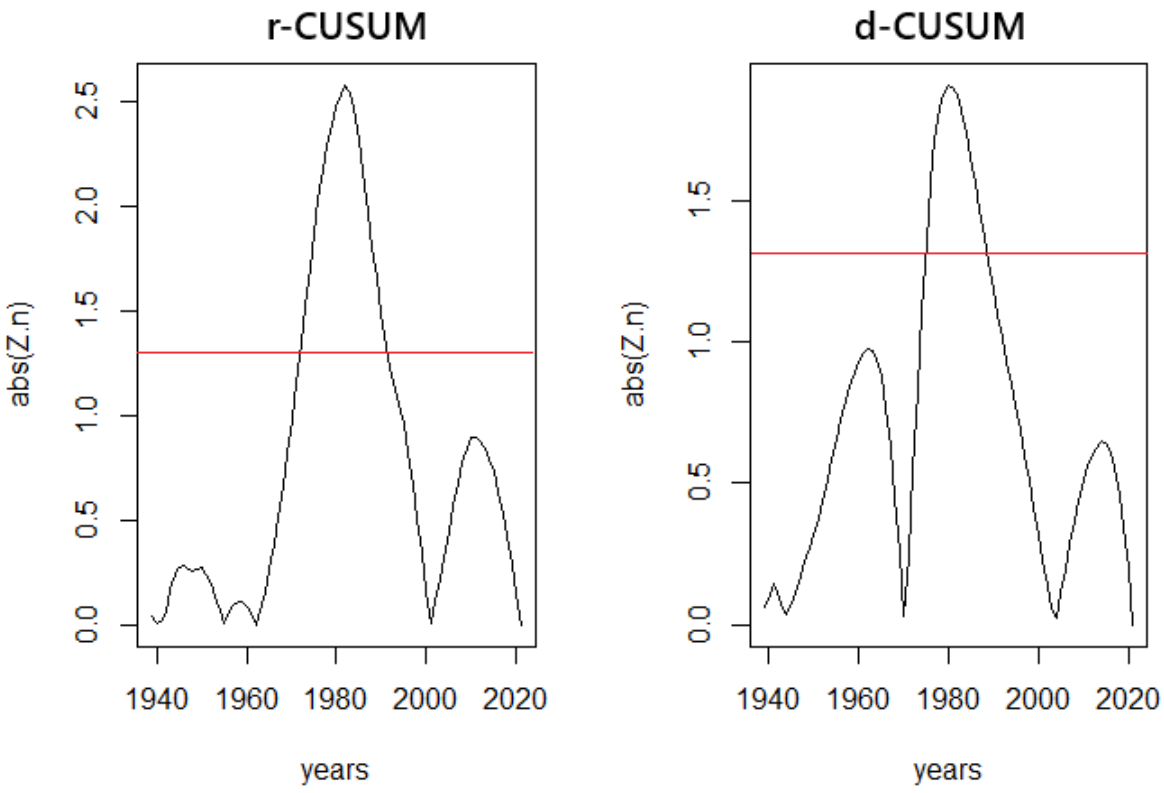


Figure 3.6: Absolute value of Z_n test statistic for r-CUSUM and d-CUSUM for Finland fertility data.

for detecting change in various distributional characteristics of object data lying in the manifold space of pdf curves.

Example 4 (Natural Language Processing): In this example, we focus on a network data which is related to natural language processing (NLP) (Dryden et al., 2009). The problem is known as corpus linguistics where the networks are used to model documents comprising of text corpus. Each node is a word and each edge indicates words that are coocurred within a five words span. This modeling is in particular useful to characterize typing style of novel writers. For the current analysis, we consider novels written by Jane Austen and Charles Dickens, as listed in Table 3.6. There are 7 novels written by Austen and 16 novels written by Dickens. Our goal for this analysis is to find out whether

there is a significant statistical difference in the distribution of network data for the two authors, which helps us realize whether these two writers have significantly distinct writing styles. To get an initial idea, a pre-analysis is carried out using classical multidimensional scaling and hierarchical cluster analysis. The results for the first two multidimensional scaling variables and cluster trees can be found in Figure 3.7. It is obvious that there is a separation in the MDS space between Austen's and Dicken's novels. Also, the novels are clustered in distinct groups. Therefore, it can be concluded that there is a statistically significant difference in the writing style of the two authors.

For further evaluation, networks indicating top 100 pairs of words for Dickens and Austen are illustrated in Figure 3.8. Exploratory graphical analysis indicates that there is a clear difference in the writing style of the authors. For Austen, there are common pairings of words with terms like *her*, *herself*, and *she*. Also, it is observed that Austen tends to use these words with emotional words such as *feelings*, *felt*, *feel*, *kindness*, *pleasure*, *happiness*. For Dickens, there are more frequent use of abbreviations, e.g. *it's*, *I'll*, and *that's*. It is also observed that Dickens pairs words related to body parts, e.g. *arms*, *shoulders*, *eyes*, *feet*, *hand* with words like *his*, *the*. Moreover, Dickens uses a more pronounced pairing of words such as *dark*, *black*, and *dead*. These findings further endorse the fact that there is clear difference in the writing style followed by the two authors.

To carry out the change point analysis, we calculate the adjacency matrix, which are undirected graphs $G(\mathcal{V}, \mathcal{E})$, for the top $d = 1000$ words used in the novels and calculate the Laplacian of the adjacency matrices which is defined as a matrix $L = D - W$ with W denoting the $d \times d$ adjacency matrix of graph G with $\mid V \mid = d$, and D is a diagonal matrix with the vertex degrees along the diagonal. The degree information is directly available on diagonal elements and adjacency information as negative off-diagonal elements. Moreover, spectral analysis of L can give important information about the main characteristics of the network, such as spectral gap, the smallest non-zero eigenvalue of L, and the sparsest cut of a graph, which minimizes the ratio of the number of edges across the cut and the number of vertices in the smaller part of the graph. The analysis can be done by focusing on the space of Laplacians

$$\mathcal{L}_d = \{L = (l_{ij}) : L = L^\top;\ l_{ij} \leq 0,\ \forall i \neq j;\ L\mathbf{1}_{d,1} = \mathbf{0}_{d,1}\}\ .$$

Therefore, we can work with the space of graph Laplacians $\mathcal{L}_d$, which is a submanifold of Euclidean space $\mathbb{R}^{d^2}$ with intrinsic dimension $\frac{d(d+1)}{2}$ (Ginestet et al., 2017). Laplacian matrices are PSD, and the corresponding metric space is $(\mathcal{L}_d, \rho_F)$ with ρ_F being the Frobenius norm

$$\rho_F = \|X - Y\|_F^2 = \sum_{i=1}^{d}\sum_{j=1}^{d}(x_{ij} - y_{ij})^2\ ,$$

Table 3.6: List of Jane Austen and Charles Dickens novels.

Author	Novel	Abbreviation	Year Published
Austen	Lady Susan	LS	1794
Austen	Sense and Sensibility	SE	1795
Austen	Pride and Prejudice	PR	1796
Austen	Northanger Abbey	NO	1798
Austen	Mansfield Park	MA	1811
Austen	Emma	EM	1814
Austen	Persuation	PE	1815
Dickens	The Pickwick Papers	PP	1836
Dickens	Oliver Twist	OT	1837
Dickens	Nicholas Nickleby	NN	1838
Dickens	The Old Curiosity Shop	OCS	1840
Dickens	Barnaby Rudge	BR	1841
Dickens	Martin Chuzzlewit	MC	1843
Dickens	A Christmas Carol	C	1843
Dickens	Dombey and Son	DS	1846
Dickens	David Copperfield	DC	1849
Dickens	Bleak House	BH	1852
Dickens	Hard Times	HT	1854
Dickens	Little Dorrit	LD	1855
Dickens	A Tale of Two Cities	TTC	1859
Dickens	Great Expectations	GE	1860
Dickens	Our Mutual Friend	OMF	1864
Dickens	The Mystery of Edwin Drood	ED	1870

for any pair of matrices $X, Y \in \mathcal{L}_d$. Therefore, we can use all properties of the manifold space

$$SPD(d) = \{A \in \mathbb{R}^{d\times d} \mid A = A^\top, \text{ and } x^\top Ax > 0\} \ .$$

We consider HD depth for the analysis. Given that there are a very small sample, 7 for Austen and 16 for Dickens, we can evaluate and verify the power and capacity of change point detection method for finite sample cases. Bootstrapping is carried out with $B = 1000$ where samples are selected with replacement from the two groups and the r-CUSUM and d-CUSUM statistics are calculated. Table 3.7 lists the summary statistics. It can be seen that the mode is located at 6 which is very close to the true change point, 7. Also, 99% confidence interval based on normality assumption includes the true change point 7, indicating the change point detection works properly for high-dimensional problem with manifold dimensionality $\frac{1000\times1001}{2} = 500500$ and sample size as small as 25. This is

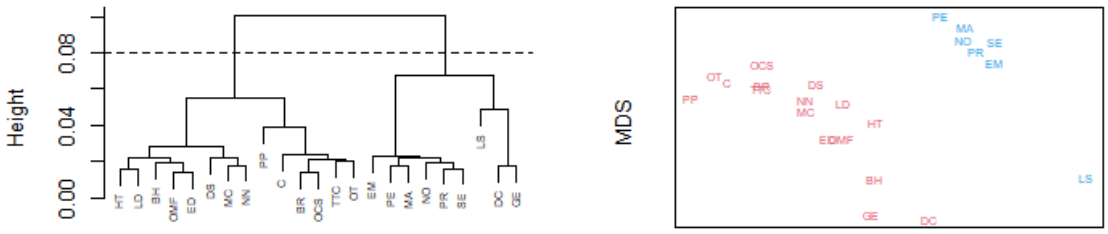


Figure 3.7: Cluster analysis and MDS plot for Dickens and Austen novels.

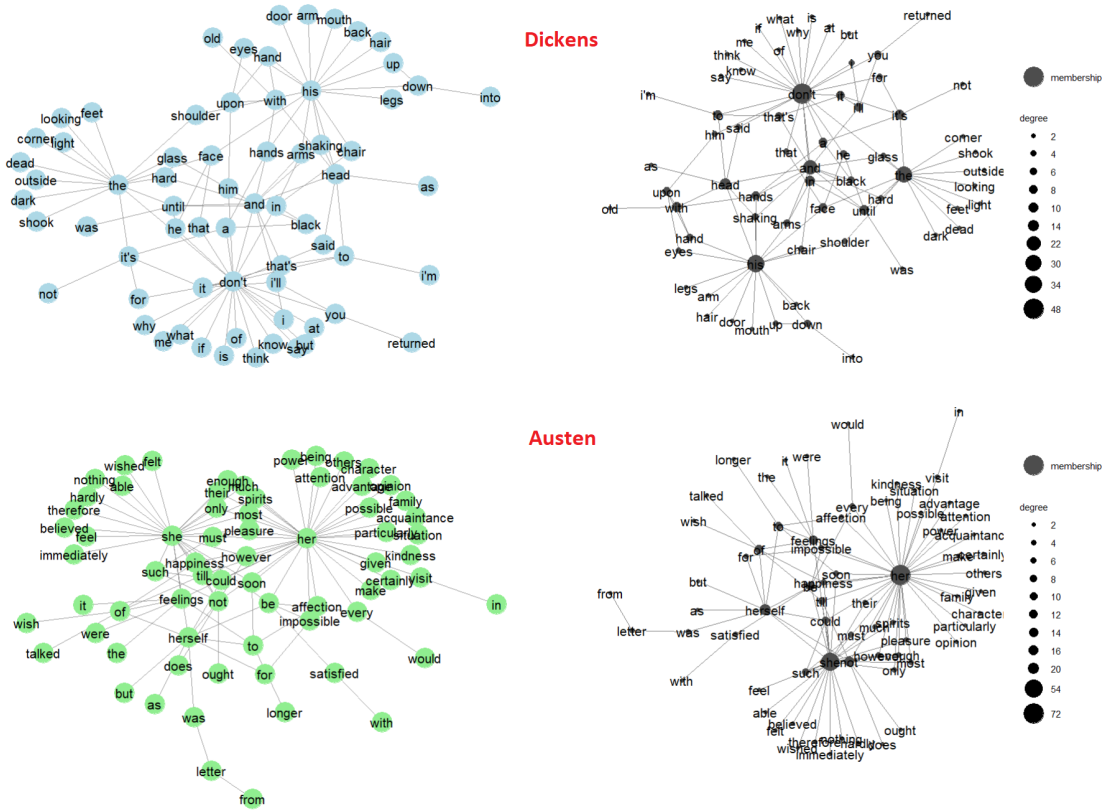


Figure 3.8: Networks indicating top 100 pairs of words for Dickens and Austen.

a promising observation demonstrating the high capability of the considered change point detection methods for change point detection problems arising in the field of linguistic corpus analysis and NLP. It is observed that r-CUSUM and d-CUSUM outperform AMOC and ECP with more accurate CI estimate for the change point location.

Table 3.7: Summary statistics for change point location estimates using Bootstrapping with $B = 1000$.

Methods	min	median	mode	max	99% CI
r-CUSUM	4	6	6	16	[5,7]
d-CUSUM	4	6	6	16	[5,7]
AMOC	4	6	8	18	[5,9]
ECP	4	6	8	18	[5,9]

Example 5 (Singular Case for von Mises-Fisher Distribution): Consider the von Mises-Fisher distribution on sphere $\mathcal{S}^d$ with probability density function for $(d+1)$-dimensional unit vector $\mathbf{x}$ defined as

$$f(\mathbf{x}; \boldsymbol{\mu}, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\left(\kappa \boldsymbol{\mu}^\top \mathbf{x}\right) \ ,$$

where $\|\boldsymbol{\mu}\| = 1$ is the location parameter, $\kappa \geq 0$ is the scale parameter, and I_0 is the modified Bessel function of first kind at order 0. The support of the distribution is $\{\mathbf{x} \in \mathbb{R}^{d+1} \mid \|\mathbf{x}\| = 1\}$.

Suppose we are at a stage of change detection in which the proposed method is searching a segment with exactly one change point located at the breaking fraction θ resulting in the combined distribution

$$H(\mathbf{x} \mid \boldsymbol{\mu}, \kappa) = \theta\, Q_1(\mathbf{x} \mid \boldsymbol{\mu}_{_1}, \kappa_1) + (1-\theta)\, Q_2(\mathbf{x} \mid \boldsymbol{\mu}_{_2}, \kappa_2)\,.$$

Assume that $Q_1(\mathbf{x} \mid \boldsymbol{\mu}_{_1}, \kappa_1) = \text{VMF}(\boldsymbol{\mu}_{_1}, \kappa_1)$ and $Q_2(\mathbf{x} \mid \boldsymbol{\mu}_{_2}, \kappa_2) = \text{VMF}(\boldsymbol{\mu}_{_2}, \kappa_2)$, for brevity shown by $Q_1(\mathbf{x})$ and $Q_2(\mathbf{x})$ hereafter. It can be observed that

$$\begin{aligned}
\mathrm{E}\left[Z_{_n}(\theta)\right] &= \sqrt{\frac{12}{n(n^2-1)}}\,\mathrm{E}\left[\sum_{j=1}^{\lfloor n\theta \rfloor} \widehat{R}_j - \mathrm{E}[\widehat{R}_j]\right] \\
&= \sqrt{\frac{12n^2}{n(n^2-1)}}\left(\mathrm{E}\left[\sum_{j=1}^{\lfloor n\theta \rfloor} \frac{\widehat{R}_j}{n}\right] - \frac{\lfloor n\theta \rfloor}{n}\mathrm{E}[\widehat{R}_j]\right) \\
&= \sqrt{\frac{12n}{(n^2-1)}}\,\lfloor n\theta \rfloor \left(\int_{S^d} \left[\,\theta\, Q_1(\mathbf{x}) + (1-\theta)\, Q_2(\mathbf{x})\,\right] dQ_1(\mathbf{x}) - \frac{1}{2}\right) \\
&\approx \sqrt{12n}\,\theta \left(\int_{S^d} \left[\,\theta\, Q_1(\mathbf{x}) + (1-\theta)\, Q_2(\mathbf{x})\,\right] dQ_1(\mathbf{x}) - \frac{1}{2}\right) .
\end{aligned}$$

We can rewrite the probability density function of von Mises-Fisher distribution in the form of multivariate normal distribution with isotropic covariance κI and mean $\boldsymbol{\mu}$ such that $\|\boldsymbol{\mu}\| = 1$, as follows

$$f(\mathbf{x}; \boldsymbol{\mu}, \kappa) = \left(\sqrt{\frac{\kappa}{2\pi}}\right)^{p} \exp\left(-\kappa \frac{(\mathbf{x}-\boldsymbol{\mu})^{\top}(\mathbf{x}-\boldsymbol{\mu})}{2}\right) ,$$

where $\|\mathbf{x}\| = 1$.

To have the singularity condition, the constant term in the parenthesis must become zero, which means $\int_{-\infty}^{\infty} Q_2(\mathbf{x})\, dQ_1(\mathbf{x}) = \frac{1}{2}$. Since this is the expectation of the cdf of $\text{VMF}(\mu_{_2}, \kappa_2)$ with respect to $\text{VMF}(\mu_{_1}, \kappa_1)$, it can be represented as

$$\begin{aligned}
\int_{S^d} Q_2(x)\, dQ_1(x) &= \int_{S^d} \left(\int_{\mathbf{x}} \left(\sqrt{\frac{\kappa_1}{2\pi}}\right)^{p} \exp\left(-\kappa_2 \frac{(\mathbf{y}-\boldsymbol{\mu}_2)^{\top}(\mathbf{y}-\boldsymbol{\mu}_2)}{2}\right) dy \right) \\
&\qquad \left(\sqrt{\frac{\kappa_1}{2\pi}}\right)^{p} \exp\left(-\kappa_1 \frac{(\mathbf{x}-\boldsymbol{\mu}_1)^{\top}(\mathbf{x}-\boldsymbol{\mu}_1)}{2}\right) dx \\
&= \int_{S^d} \Phi(\kappa_2(v/\kappa_1 - \boldsymbol{\mu}_2 + \boldsymbol{\mu}_1))\, d\Phi(v) ,
\end{aligned}$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution, and the last equation is obtained by the change of variables $\mathbf{u} = \sqrt{\kappa_2}(\mathbf{y} - \boldsymbol{\mu}_2)$ and $\mathbf{v} = \sqrt{\kappa_1}(\mathbf{x} - \boldsymbol{\mu}_1)$. Now, for any other variable $\mathbf{z} \sim \text{MVN}(\mathbf{0}, \mathbf{I})$ independent of $\mathbf{v}$, we can write

$$\begin{aligned}
\text{E}_{\mathbf{v}}\left[\Phi\left(\kappa_2(\mathbf{v}/\kappa_1 - \boldsymbol{\mu}_2 + \boldsymbol{\mu}_1)\right)\right] &= \text{E}_{\mathbf{v}}\left[\mathbb{I}\left(\mathbf{z} \le \kappa_2(\mathbf{v}/\kappa_1 - \boldsymbol{\mu}_2 + \boldsymbol{\mu}_1)\right)\right] \\
&= \Pr\left(\sqrt{\kappa_1 + \kappa_2}(\mathbf{z}/\kappa_2 - \mathbf{v}/\kappa_1) \le \sqrt{\kappa_1 + \kappa_2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right) \\
&= \Phi\left(\sqrt{\kappa_1 + \kappa_2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right) .
\end{aligned}$$

This means that whatever the scale parameters κ_1 and κ_2 are, the singular condition happens for VMF distributed data when $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ (no change in location). For the case of multiple undetected change points (in location), there exists at-least one dominant change point. Our simulation study shows that this singular situation arises when there is a single change in scale and rank-depth change point detection might not have a large power for this case study.

In general, for object data sampled from manifold spaces, it is not easy to categorize the change in form of location change or scale change, e.g. shape spaces, matrix spaces, flag manifolds, etc. Our simulation indicated that the proposed rank-depth change detection methods have good power in detecting change in distribution for object data sampled from various types of manifolds.

Example 6 (Change in Location and Scale of Distribution): Consider the von Mises distribution on torus $\mathbb{T}^2 = S^1 \times S^1$ with probability density function defined as

$$f(\phi, \psi) = Z_c(\kappa_1, \kappa_2, \kappa_3) \exp\{\kappa_1 \cos(\phi - \mu) + \kappa_2 \cos(\psi - \nu) - \kappa_3 \cos(\phi - \mu - \psi + \nu)\} \ ,$$

where μ and ν are the means, κ_1 and κ_2 are the concentration parameters, and κ_3 is related to their correlation. For this manifold space, our goal is to study the effect of change in location/scale parameters of the distribution. Consider two groups of data of size 500 sampled from torus with different location parameters $\mu_1 = \pi/2, \nu_1 = \pi/2$ and $\mu_2 = 3\pi/2, \nu_2 = \pi/2$, and the same scale parameters $\kappa_1 = 5, \kappa_2 = 5, \kappa_3 = 2$. Figure 3.9 depicts these two groups of data on 2-torus $\mathbb{T}^2$. It is clear that these data are sampled at two different locations and their variability is the same. In such a scenario, the difference between these two groups may not be well identified by depth functions, as the depth functions measure the centrality of the points and therefore reflect possible changes in scale. This is clear in the right panel plot in Figure 3.9, where the distribution of the depth values for these two groups remains the same. Therefore, we need to modify the proposed depth-based change point detection methods to make sure the change detection will have power when the change is only in location.

We propose a combined rank-based change point detection method based on rank-depth and rank of observed data on the corresponding tangent space using the standard sign rank. Given metric space $(\mathcal{M}, d)$, consider an isometric embedding map $J : \mathcal{M} \to \mathbb{R}^d$, such that $\|J(x) - J(y)\| = d(x, y)$ for $x, y \in \mathcal{M}$. Let $U(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$ denote the spatial sign function for $\mathbf{x} \in \mathbb{R}^d$. We can use the spatial sign function to modify the proposed r-CUSUM change point detection method. The idea is to embed the data on the manifold space to the ambient Euclidean space. This suggests a setup for extrinsic analysis so that the spatial sign in the ambient space $\mathbb{R}^d$ can be calculated. The proposed statistic is the spatial sign CUSUM (SS-CUSUM) which is an extension of the work of (Li et al., 2013; Crosier, 1988) to manifold spaces and for change point detection applications. The

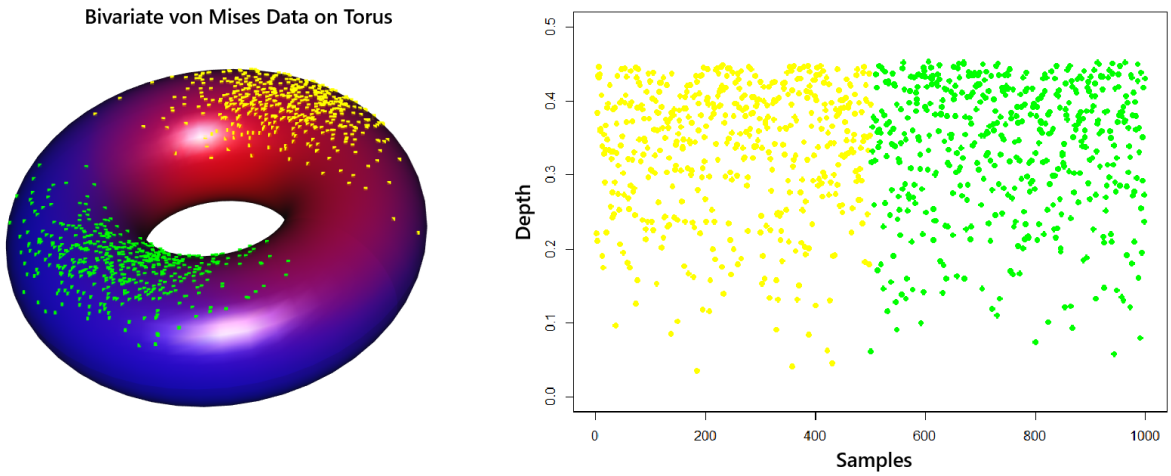


Figure 3.9: Data sampled from von Mises distribution on 2-torus $\mathbb{T}^2$. Left panel indicates two groups of data sampled at different locations ($\mu_1 = \pi/2, \nu_1 = \pi/2$) and ($\mu_2 = 3\pi/2, \nu_2 = \pi/2$) with the same scale parameters ($\kappa_1 = 5, \kappa_2 = 5, \kappa_3 = 2$). Right panel indicates the corresponding depth values of these two groups.

proposed nonparametric SS-CUSUM is formulated as follows

$$
\begin{aligned}
C_n &= \sqrt{(S_{n-1} + U(\mathbf{x}_n) - \boldsymbol{\mu})^\top \Sigma_0^{-1} (S_{n-1} + U(\mathbf{x}_n) - \boldsymbol{\mu})}\ , \\
S_n &= \begin{cases} 0 & \text{if } C_n \le k \\ (S_{n-1} + U(\mathbf{x}_n) - \boldsymbol{\mu}) \left(1 - \dfrac{k}{C_n}\right) & \text{if } C_n > k \end{cases}\ , \\
T_n &= \sup_{0<t<1} \left| \left(S_n(t)^\top \Sigma_0^{-1} S_n(t)\right)^{1/2} \right|\ ,
\end{aligned}
$$

where $\boldsymbol{\mu}$ and $\Sigma_0 = \mathbf{I}_d/d$ are sample mean and sample variance.

A change point occurs if $T_n > h$, where h is a threshold determined by k and Brownian bridge. We can use both r-CUSUM and SS-CUSUM to identify change in location and scale of a distribution. To calculate the spatial sign values, we need to embed the data

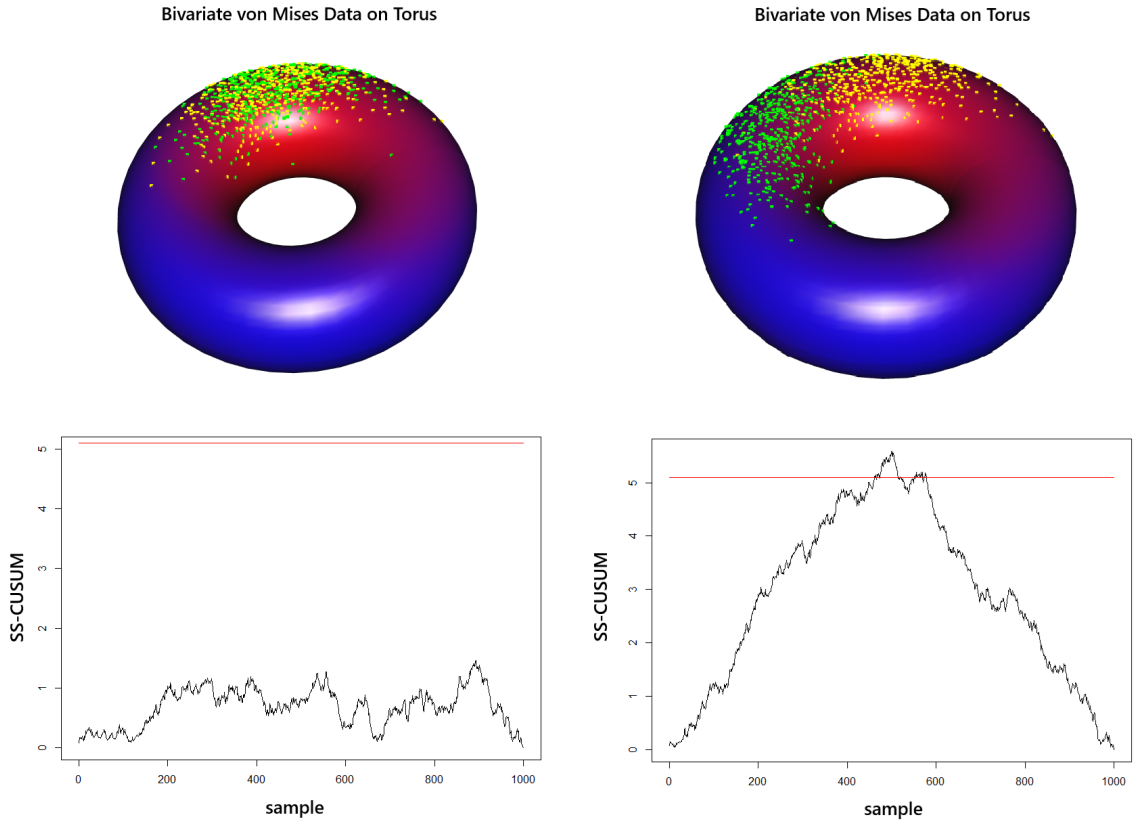


Figure 3.10: The performance of SS-CUSUM chart under null and alternative with effect size $\delta = 0.7$.

from manifold space to ambient Euclidean space $\mathbb{R}^3$, using the following mapping

$$
\begin{aligned}
x &= (r + R\cos(\psi))\cos(\phi) \ , \\
y &= (r + R\cos(\psi))\sin(\phi) \ , \\
z &= r\sin(\psi) \ ,
\end{aligned}
$$

where r is the minor radius and R is the major radius of the torus. The above map is a local isometric embedding of $\mathbb{T}^2$ to $\mathbb{R}^3$ (Lin, 2008). The proposed method has an empirical power of 1 with estimated location 500, indicating the efficacy of the proposed method in detecting change in location. Figure 3.10 indicates the performance of SS-CUSUM under null hypothesis of no location change and alternative hypothesis of location shift

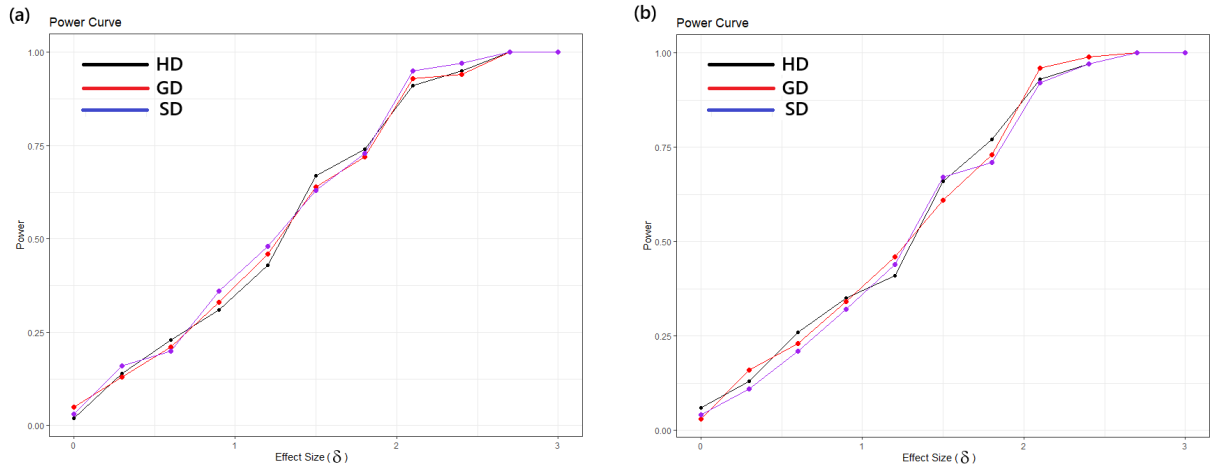


Figure 3.11: Empirical power for 2-torus data, (a) performance of SS-CUSUM with HD, GD, and SD depth functions for location change, (b) performance of r-CUSUM with HD, GD, and SD depth functions for scale change.

with $(\mu_2 = 1.7\pi/2, \nu_2 = \pi/2)$. Clearly, SS-CUSUM successfully detected the changes in the location parameter. Figure 3.11 compares the power plots of SS-CUSUM for location change and r-CUSUM for scale change with effect sizes δ in the range $(0, 3)$ based on 5000 independent simulation. As seen, both SS-CUSUM and r-CUSUM methods are successful for detecting change in location and scale of the distribution, respectively. This suggests that we can use both of these nonparametric methods in tandem to detect changes in scale and location.

SS-CUSUM test statistic is applicable to any type of manifold data as long as tangent space can be calculated or the manifold data can be embedded in higher dimensional Euclidean space. Without loss of generality, consider the manifold to be $\text{Sym}^+(d)$ of $d \times d$ psd matrices. half-vectorization mapping can be used for embedding the data on manifold $\text{Sym}^+(d)$ to ambient Euclidean space $\mathbb{R}^{\frac{d(d+1)}{2}}$ from which we can calculate the sign rank. This test statistic work best for manifold spaces in which we can categorize the change in the form of change in location/scale, e.g. von Mises-Fisher distribution on d-Sphere, von Mises distribution on d-torus.

In general, our simulation studies demonstrated that rank-depth change detection methods have a high power for detecting change in the distribution of data sampled from various types of manifolds, e.g. shape spaces, complex projective spaces, matrix spaces, phylogenetic trees, etc.

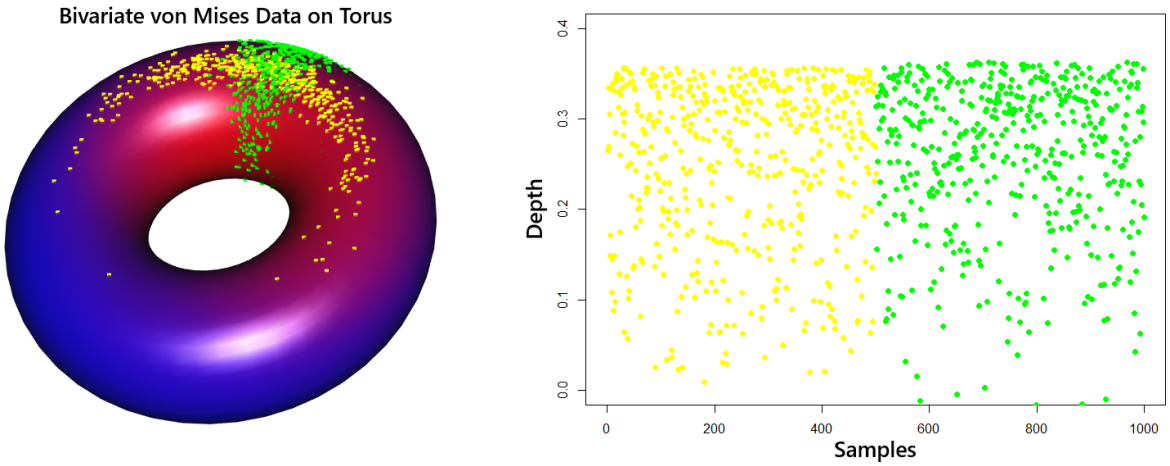


Figure 3.12: Data sampled from von Mises distribution on 2-torus $\mathbb{T}^2$. Left panel indicates two groups of data sampled at same location ($\mu_1 = \pi/2, \nu_1 = \pi/2$) and different concentration parameters ($\kappa_1 = 3, \kappa_2 = 50, \kappa_3 = 0$) and ($\kappa_1 = 50, \kappa_2 = 3, \kappa_3 = 0$). Right panel indicates the corresponding depth values of these two groups.

Example 7 (Change in Direction of Variability): In this example, our goal is to indicate a weakness of change point detection methods for a relatively difficult case of change in distribution. For most of the cases, change in the scale of the distribution appears in the form of expansion/contraction. However, if the nature of the change of the variability of the distribution is that the scale remains the same but the direction changes, then in general we expect the depths before and after the change to be similar (Chenouri et al., 2020). In such cases, rank-depth methods may not have power. To investigate this issue, we consider the von Mises distribution on 2-torus $\mathbb{T}^2$ with concentration parameters $\kappa_1 = 3, \kappa_2 = 50, \kappa_3 = 0$ and $\kappa_1 = 50, \kappa_2 = 3, \kappa_3 = 0$ and the same location parameter $\mu = \pi/2, \nu = \pi/2$.

Figure 3.12 depicts the data sampled from von Mises distribution on 2-torus $\mathbb{T}^2$. Left panel indicates two groups of data sampled at same location ($\mu_1 = \pi/2, \nu_1 = \pi/2$) and different concentration parameters ($\kappa_1 = 3, \kappa_2 = 50, \kappa_3 = 0$) and ($\kappa_1 = 50, \kappa_2 = 3, \kappa_3 = 0$). Right panel indicates the corresponding depth values of these two groups. Clearly, the depth values are not altered by change in the direction of variability. Table 3.8 reports the results obtained from applying the nonparametric change point detection methods under the considered simulation setup for 5000 simulation runs. As expected, in this case, the depth-based methods do not have power; however, as sample size N increases, r-CUSUM

Table 3.8: Empirical power of change point detection methods.

Methods	n=10	n=50	n=100	n=500
r-CUSUM-GD	0.13	0.43	0.72	1.00
r-CUSUM-SD	0.03	0.06	0.14	0.42
SS-CUSUM-GD	0.04	0.03	0.07	0.06
SS-CUSUM-SD	0.05	0.07	0.03	0.04
AMOC	0.02	0.01	0.13	0.21
ECP	0.06	0.02	0.13	0.16

with GD depth can detect the change in the direction of variability to a good extend.

3.6.2 Multiple Change Point Detection

Example 8 (pdf Curve Space with Wasserstein Distance): In this example, we consider a sequence of pdf curve data objects (Dryden et al., 2009). The considered pdf's are univariate normal distribution $\mathrm{N}(\mu, 1)$ where each datum is $\mathrm{N}(\mu, 1)$. The considered change points are in scale and location. In particular, the first segment is distributed according to $\mathrm{N}(\mu_1, 1)$ where $\mu_1 \sim \mathrm{N}(0, 1.5^2)$, and the second segment follows $\mathrm{N}(\mu_2, 1)$ where $\mu_2 \sim \mathrm{N}(0, 0.75^2)$ and the last segment follows $\mathrm{N}(\mu_3, 1)$ where $\mu_3 \sim \mathrm{N}(2, 0.5^2)$. This scenario makes it possible to evaluate the performance of the method under both change in location and scale of the distribution. GD is used to calculate the depth of the data objects. To calculate the distance of objects in the space of pdfs, we use 2-Wasserstein distance between measures ν_1 and ν_2

$$W_2(\nu_1, \nu_2) = \Big(\int_0^1 (F_1^{-1}(q) - F_2^{-1}(q))^2 dq \Big)^2,$$

where F_1^{-1} and F_2^{-1} are quantile functions of ν_1 and ν_2, respectively.

To investigate the effect of change point location on the performance of the methods, we consider breaking fractions $\{\theta_1, \theta_2\} = \{0.16, 0.83\}$ and $\{\theta_1, \theta_2\} = \{0.33, 0.66\}$ with sample sizes $n = \{60, 150, 300\}$.

Table 3.9 gives the simulation result for r-BS, r-WBS, BS and WBS, PELT and ECP. It can be inferred from the simulation results that the rank based versions can dominate the competitor multiple change detection methods for almost all of the scenarios with respect to the considered metrics. Moreover, it can be realized from the summary statistics for $\widehat{\ell}$ that the performance of r-BS and r-WBS remains more stable compared to other techniques

Table 3.9: Summary statistics for the empirical distribution of $\widehat{\ell}$, and MSE, MAD, and relative efficiency (w.r.t. WBS) of $\widehat{\ell} - \ell$ for pdf curve data over 5000 simulation runs.

			$\widehat{\ell}$			$\widehat{\ell} - \ell$	
	Methods	**Min.**	**Median**	**Mode**	**Max.**	**MSE**	**MAD**
				n = 60			
$k = (20, 40),\ \ell = 2$	**r-BS**	1	2	2	3	0.30	0.00
	r-WBS	1	2	2	3	0.28	0.00
	BS	1	2	2	5	0.50	0.00
	WBS	1	4.5	5	14	15.06	2.22
	PELT	1	4	4	7	11.83	1.98
	ECP	1	3	3	4	0.13	0.00
$k = (10, 50),\ \ell = 2$	**r-BS**	1	2	2	3	0.30	0.00
	r-WBS	1	2	2	4	0.98	1.48
	BS	1	2	2	5	0.54	0.00
	WBS	1	3	3	10	6.12	1.48
	PELT	1	4	4	7	10.13	1.82
	ECP	1	2	2	3	0.22	0.00
				n = 150			
$k = (50, 100),\ \ell = 2$	**r-BS**	1	2	2	4	0.18	0.00
	r-WBS	2	2	2	6	0.96	0.00
	BS	1	2	2	8	1.74	0.00
	WBS	1	6	4	24	35.18	2.96
	PELT	1	2	2	7	1.13	0.00
	ECP	1	2	2	5	0.98	0.00
$k = (25, 125),\ \ell = 2$	**r-BS**	1	2	2	4	0.44	0.00
	r-WBS	1	3	3	5	1.18	1.48
	BS	1	2	2	7	1.40	0.00
	WBS	1	5	4	15	19.86	2.96
	PELT	1	4	5	7	12.71	2.22
	ECP	1	2	2	3	0.41	0.00
				n = 300			
$k = (100, 200),\ \ell = 2$	**r-BS**	1	2	2	4	0.36	0.00
	r-WBS	1	2	2	6	1.44	0.00
	BS	2	2	2	5	0.56	0.00
	WBS	4	9	11	19	79.32	4.44
	PELT	1	2	2	4	0.24	0.00 0
	ECP	1	3	3	5	2.16	1.23
$k = (50, 250),\ \ell = 2$	**r-BS**	1	2	2	5	0.64	0.00
	r-WBS	2	3	4	5	2.68	1.48
	BS	1	2	2	9	2.28	0.00
	WBS	2	5	2	15	30.80	4.44
	PELT	1	3	3	5	2.11	1.17
	ECP	1	2	2	4	0.41	0.00

over the considered simulation scenarios. In particular, median and max of $\widehat{\ell}$ has more variation as the sample size n varies. It can be also realized that WBS shows inferior performance as the sample size increases. The main reason is the non-robustness of the method for data lying on manifold spaces, which results in the detection of extra change points during the sub-segment random sampling. Obviously, this is not an issue when using r-WBS where the number of detected change points $\widehat{\ell}$ remains close to ℓ regardless

Table 3.10: Estimated change points of the rival methods for pdf curve data over 5000 simulation runs.

	Methods	$\widehat{k}$	$\widehat{\ell}$
n = 60			
$k = (20, 40),\ \ell = 2$	**r-BS**	$(21, 41)$	2
	r-WBS	$(21, 41)$	2
	BS	$(20, 40)$	2
	WBS	$(24, 30, 31, 34, 35, 40)$	6
	PELT	$(20, 29, 31, 40)$	4
	ECP	$(13, 21, 40)$	3
$k = (10, 50),\ \ell = 2$	**r-BS**	$(11, 51)$	2
	r-WBS	$(11, 51)$	2
	BS	$(10, 50)$	2
	WBS	$(8, 10, 50)$	3
	PELT	$(3, 10, 50, 58)$	4
	ECP	$(11, 51)$	2
n = 150			
$k = (50, 100),\ \ell = 2$	**r-BS**	$(51, 101)$	2
	r-WBS	$(51, 101)$	2
	BS	$(50, 100)$	2
	WBS	$(5, 17, 21, 44, 50, 100)$	6
	PELT	$(50, 100)$	2
	ECP	$51, 101$	2
$k = (25, 125),\ \ell = 2$	**r-BS**	$(29, 126)$	2
	r-WBS	$(26, 57, 125)$	3
	BS	$(25, 125)$	2
	WBS	$(9, 25, 74, 125)$	4
	PELT	$(25, 31, 56, 125)$	4
	ECP	$(26, 126)$	2
n = 300			
$k = (100, 200),\ \ell = 2$	**r-BS**	$(101, 201)$	2
	r-WBS	$(101, 201)$	2
	BS	$(100, 200)$	2
	WBS	$(9, 104, 115, 136, 150, 167, 174, 186, 200)$	9
	PELT	$(100, 200)$	2
	ECP	$(10, 100, 200)$	3
$k = (50, 250),\ \ell = 2$	**r-BS**	$(51, 251)$	2
	r-WBS	$(53, 67, 115, 251)$	4
	BS	$(50, 250)$	2
	WBS	$(50, 250)$	2
	PELT	$(50, 216, 250)$	3
	ECP	$(51, 251)$	2

Table 3.11: Summary statistics for the empirical distribution of $\widehat{\ell}$, and MSE, MAD, and relative efficiency (w.r.t. WBS) of $\widehat{\ell}-\ell$ for Wishart distribution over 5000 simulation runs.

		$\widehat{\ell}$				$\widehat{\ell}-\ell$	
	Methods	**Min.**	**Median**	**Mode**	**Max.**	**MSE**	**MAD**
				n = 60			
$k=(20,40),\ \ell=2$	**r-BS**	2	2	2	3	0.01	0.00
	r-WBS	2	2	2	4	0.13	0.00
	BS	2	2	2	9	2.20	0.00
	WBS	1	4.5	4	10	11.07	2.22
	PELT	2	4	4	5	5.73	2.12
	ECP	2	2	2	3	0.09	0.00
$k=(10,50),\ \ell=2$	**r-BS**	1	2	2	3	0.69	0.00
	r-WBS	1	3	3	5	1.09	0.00
	BS	2	2	2	6	0.73	0.00
	WBS	2	4	3	10	5.41	1.48
	PELT	2	3	3	5	1.52	0.00
	ECP	2	2	2	3	0.41	0.00
				n = 150			
$k=(50,100),\ \ell=2$	**r-BS**	2	2	2	2	0.00	0.00
	r-WBS	2	2	2	5	0.55	0.00
	BS	2	2	2	7	1.91	0.00
	WBS	2	7	6	18	41.08	2.96
	PELT	2	2	2	4	1.33	0.00
	ECP	2	2	2	4	1.21	0.00
$k=(25,125),\ \ell=2$	**r-BS**	2	2	2	4	0.22	0.00
	r-WBS	2	3	3	5	1.81	0.00
	BS	2	2	2	6	1.04	0.00
	WBS	2	5	4	13	16.34	2.96
	PELT	2	3	3	5	1.42	0.00
	ECP	2	2	2	4	0.31	0.00
				n = 300			
$k=(100,200),\ \ell=2$	**r-BS**	2	2	2	3	0.04	0.00
	r-WBS	2	2	2	7	1.47	0.00
	BS	2	2	2	8	2.33	0.00
	WBS	2	10	8	17	75.54	2.96
	PELT	2	2	2	3	0.32	0.00
	ECP	2	2	2	3	0.01	0.00
$k=(50,250),\ \ell=2$	**r-BS**	2	2	2	4	0.18	0.00
	r-WBS	2	3	3	8	3.99	1.48
	BS	2	2	2	7	1.38	0.00
	WBS	2	7	8	13	31.23	1.48
	PELT	2	2	2	3	0.14	0.00
	ECP	1	2	2	4	1.12	0.00

of sample size n. This fact is also reflected in MSE of estimated $\widehat{\ell}$ where the MSE for WBS gets larger when sample size increases. Table 3.10 reports the estimated locations for the considered methods. It can be seen that the estimated locations for r-BS and r-WBS are very close to true locations for all of the scenarios.

Example 9 (Wishart Distribution): In this example, we consider change point detection for a sequence of data distributed according to Wishart distribution. We consider

two change point locations in the data sequence, where the first segment is distributed according to $W_3(20I_2)$, the second segment is distributed according to $W_{20}(6I_2)$ and the last segment follows $W_{30}(10I_2)$. Similar to the previous example, the breaking fractions are set as $\{\theta_1, \theta_2\} = \{0.16, 0.83\}$ and $\{\theta_1, \theta_2\} = \{0.33, 0.66\}$ with sample sizes $n = \{60, 150, 300\}$. Table 3.11 presents the simulation result for r-BS, r-WBS, BS and WBS, PELT and ECP. It can be realized that rank-depth statistics show promising results regardless of the location of the change points and sample sizes.

Summary statistics also reveal that BS and WBS, PELT and ECP have a more unstable performance with respect to the number of detected change point as the location of the change points and the number of samples varies. r-BS outperforms the other methods with respect to MSE of estimated $\widehat{\ell}$. Table 3.12 presents the estimated change points for the considered methods. It can be observed that the estimated change points are close to true change point locations for the most of the considered scenarios.

Example 10 (Wishart Distribution with Contamination Noise): In this example, we will investigate the robustness of the rank based change point detection methods considering data sequence distributed according to Wishart distribution contaminated with noise. We will consider mixture of Wisharts to simulate the noise contaminated scenario. For $A, V \in SPD(k)$, let us denote the probability density function of Wishart distribution with covariance parameter V by $f(V^{-1}A)$, which is a function of $\text{tr}(V^{-1/2}AV^{-1/2}) = \text{tr}(V^{-1}A)$.

Consider a random matrix V with continuous mixing density w. The continuous mixing Wishart density is given by

$$g(A) = \int_{\mathcal{M}} w(V) f(V^{-1}A) dV \ ,$$

where $\mathcal{M} = SPD(k)$.

For the considered simulation, the distribution of mixtures for the first segment is

$$\alpha W_3(20I_2) + (1-\alpha) W_3(2I_2) \ ,$$

the distribution of the second segment is

$$\alpha W_{20}(6I_2) + (1-\alpha) W_3(2I_2) \ ,$$

and the distribution of the last segment is

$$\alpha W_{30}(10I_2) + (1-\alpha) W_3(2I_2) \ ,$$

Table 3.12: Estimated change points of the rival methods for Wishart distribution.

	Methods	$\widehat{k}$	$\widehat{\ell}$
n = 60			
$k = (20, 40),\ \ell = 2$	**r-BS**	$(21, 41)$	2
	r-WBS	$(21, 41)$	2
	BS	$(20, 39)$	2
	WBS	$(20, 39)$	2
	PELT	$(7, 11, 20, 40)$	4
	ECP	$(21, 41)$	2
$k = (10, 50),\ \ell = 2$	**r-BS**	$(11, 51)$	2
	r-WBS	$(11, 25, 50)$	3
	BS	$(10, 50)$	2
	WBS	$(8, 11, 46)$	3
	PELT	$(10, 45, 50)$	3
	ECP	$(11, 51)$	2
n = 150			
$k = (50, 100),\ \ell = 2$	**r-BS**	$(51, 101)$	2
	r-WBS	$(51, 101)$	2
	BS	$(50, 100)$	2
	WBS	$(10, 16, 17, 19, 50, 101)$	6
	PELT	$(50, 100)$	2
	ECP	$(51, 101)$	2
$k = (25, 125),\ \ell = 2$	**r-BS**	$(26, 126)$	2
	r-WBS	$(26, 91, 127)$	3
	BS	$(25, 125)$	2
	WBS	$(16, 17, 25, 125)$	4
	PELT	$(25, 103, 125)$	3
	ECP	$(26, 126)$	2
n = 300			
$k = (100, 200),\ \ell = 2$	**r-BS**	$(101, 201)$	2
	r-WBS	$(101, 201)$	2
	BS	$(100, 199)$	2
	WBS	$(9, 20, 53, 60, 63, 76, 100, 200)$	8
	PELT	$(100, 200)$	2
	ECP	$(101, 201)$	2
$k = (50, 250),\ \ell = 2$	**r-BS**	$(51, 251)$	2
	r-WBS	$(51, 154, 251)$	3
	BS	$(50, 251)$	2
	WBS	$(2, 25, 33, 34, 41, 45, 50, 250)$	8
	PELT	$(50, 250)$	2
	ECP	$(56, 251)$	2

Table 3.13: Summary statistics for the empirical distribution of $\widehat{\ell}$, and MSE, MAD, and relative efficiency (w.r.t. WBS) of $\widehat{\ell} - \ell$ for Wishart distribution contaminated with noise over 5000 simulation runs.

			$\widehat{\ell}$			$\widehat{\ell} - \ell$	
	Methods	**Min.**	**Median**	**Mode**	**Max.**	**MSE**	**MAD**
				n = 60			
$k = (20, 40),\ \ell = 2$	**r-BS**	1	2	2	2	0.88	0.00
	r-WBS	1	2	2	4	0.55	0.00
	BS	1	2	2	16	4.85	1.48
	WBS	2	7	6	13	32.78	3.96
	PELT	2	3	3	5	1.35	0.00
	ECP	1	4	4	5	2.42	1.15
$k = (10, 50),\ \ell = 2$	**r-BS**	1	2	2	3	1.24	0.00
	r-WBS	1	2	3	5	1.92	1.48
	BS	1	2	2	14	7.44	2.00
	WBS	2	10	10	16	62.77	3.96
	PELT	2	8	10	13	32.43	3.25
	ECP	2	5	5	7	6.32	2.14
				n = 150			
$k = (50, 100),\ \ell = 2$	**r-BS**	2	2	2	3	0.04	0.00
	r-WBS	2	3	2	5	1.90	1.48
	BS	2	3	2	10	4.18	1.48
	WBS	3	14	16	24	163.95	4.44
	PELT	5	10	13	17	75.21	4.10
	ECP	6	12	16	20	84.52	4.31
$k = (25, 125),\ \ell = 2$	**r-BS**	2	2	2	3	0.12	0.00
	r-WBS	2	3	3	6	1.91	0.00
	BS	2	2	2	13	5.43	0.00
	WBS	10	20	18	32	339.48	4.44
	PELT	4	6	6	10	27.32	2.58
	ECP	6	10	14	20	83.41	3.74
				n = 300			
$k = (100, 200),\ \ell = 2$	**r-BS**	2	2	2	3	0.04	0.00
	r-WBS	2	3	2	6	2.03	1.48
	BS	2	2.5	2	8	2.86	0.74
	WBS	10	21	18	39	432.13	5.93
	PELT	4	9	9	16	126.43	2.31
	ECP	4	10	10	17	133.51	2.14
$k = (50, 250),\ \ell = 2$	**r-BS**	2	2	2	3	0.17	0.00
	r-WBS	2	3	3	9	2.39	0.00
	BS	2	2.5	2	8	5.17	0.74
	WBS	19	33.5	32	45	1025.22	5.18
	PELT	6	14	14	23	614.31	4.72
	ECP	8	20	20	33	732.13	5.06

Table 3.14: Estimated change points of the rival methods for Wishart distribution contaminated with noise.

	Methods	$\widehat{k}$	$\widehat{\ell}$
n = 60			
$k = (20, 40),\ \ell = 2$	**r-BS**	$(23, 41)$	2
	r-WBS	$(18, 42)$	2
	BS	$(20, 36)$	2
	WBS	-	16
	PELT	$(20, 38, 40)$	3
	ECP	$(4, 21, 37, 40)$	4
$k = (10, 50),\ \ell = 2$	**r-BS**	$(11, 51)$	2
	r-WBS	$(11, 23, 46)$	3
	BS	$(9, 50)$	2
	WBS	-	18
	PELT	-	8
	ECP	$(11, 25, 30, 43, 50)$	5
n = 150			
$k = (50, 100),\ \ell = 2$	**r-BS**	$(51, 99)$	2
	r-WBS	$(58, 100)$	2
	BS	$(44, 99)$	2
	WBS	-	18
	PELT	-	13
	ECP	-	16
$k = (25, 125),\ \ell = 2$	**r-BS**	$(26, 122)$	2
	r-WBS	$(26, 87, 126)$	3
	BS	$(25, 124)$	2
	WBS	-	32
	PELT	$(25, 37, 45, 55, 117, 125)$	6
	ECP	-	14
n = 300			
$k = (100, 200),\ \ell = 2$	**r-BS**	$(101, 201)$	2
	r-WBS	$(103, 203)$	2
	BS	$(94, 200)$	2
	WBS	-	10
	PELT	-	9
	ECP	-	10
$k = (50, 250),\ \ell = 2$	**r-BS**	$(51, 250)$	2
	r-WBS	$(48, 235, 249)$	3
	BS	$(50, 250)$	2
	WBS	-	32
	PELT	-	14
	ECP	-	20

where $\alpha \sim \text{Bern}(0.9)$. As can be seen, for all the three segments, the contamination noise is distributed according to $W_3(2I_2)$.

Table 3.13 presents the simulation results for the breaking fractions $\{\theta_1, \theta_2\} = \{0.16, 0.83\}$ and $\{\theta_1, \theta_2\} = \{0.33, 0.66\}$ with sample sizes $n = \{60, 150, 300\}$.

The summary statistics for $\widehat{\ell}$ demonstrates the efficacy of the rank-based statistics for multiple change point detection when data is contaminated with outliers. The number of detected change points can become very large for BS, WBS, PELT and ECP when there are actually two change points in the sequence. This is even more apparent as the sample size gets larger and the number of outliers in the data sequence increases. For such scenarios, BS, WBS, PELT and ECP became unstable and in some cases detected around 16, 45, 23, and 33 change points. r-BS and r-WBS do not suffer from this deficiency and can keep the number of detected change points close to the true ℓ regardless of the sample size.

Also, significantly lower MSE of r-BS and r-WBS for $\widehat{\ell}$ verifies the capability of rank-based statistics when data are contaminated with noise. For most of the simulation scenarios, MSE is around 50% smaller for the rank-based method compared to other competitor methods.

Table 3.14 presents the estimated change points for the considered scenarios. By comparing the rank-based methods with the standard methods, it can be realized that r-BS can outperform the other methods in terms of detecting the correct location of change points.

Example 11 (Phenotypic Change in Pupfish Population): In this example, our goal is to investigate phynotypic change in important biological disciplines including phenotypic plasticity, morphological and physiological characteristics, and ecological evolution in populations of pupfishes. In Collyer et al. (2015), the authors proposed nonparametric high-dimensional statistics to approach this problem, where anatomic landmarks are considered as geomorphic features to describe organism shape.

Our purpose is to extend the application of nonparametric manifold-based statistical analysis to study dimorphisms in body shape for different populations of pupfishes. Landmark data are collected from 54 museum specimens of Pecos pupfishes which are inhabitants of Pecus river in Eastern New Mexico, USA. These 54 specimens are collected from a large marsh system (29 fishes, 16 females and 13 males) and small sinkhole (25 fishes, 12 females and 13 males). Experiments revealed that body shape traits such as phenotypic plasticity in body shape can vary according to sex and environmental factors such as salinity, as well as the presence of predators. The body shapes are represented using 12 fixed anatomical landmarks and 44 sliding semi-landmarks (56 landmarks in total).

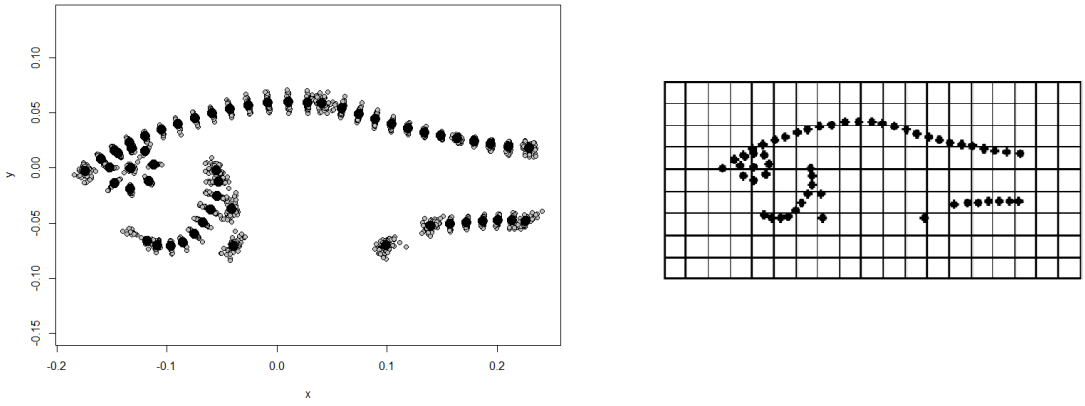


Figure 3.13: Landmark representation of pupfishes are shown in left panel. Digitized representation of the average shape is given in right panel.

Figure 3.13 indicates the landmark representation of the sample pupfishes along with the digitized representation of the average shape. As it can be seen, there is an apparent variation between landmark characteristics of the collected pupfishes indicating that there may exist shift points in the distribution of the collected samples. Our goal is to investigate this using the proposed rank-depth change point analysis method.

To study the patterns of shape variation in the shape space Σ_2^k, we carry out principal component analysis. Figure 3.14 depicts the principal component plot of shape variation in the population of pupfishes. It is apparent, that the first two dimensions explain 67% of shape variation in the collected dataset, and also reveal distinct cluster of specimens, indicating the presence of shape difference in the collected data. The first principal coordinate mostly explains the change in sex of specimens and the second principal coordinate explains the change in the environmental factors of specimens. Bottom left shape represent the landmarks of a female marsh pupfish, and bottom right shape represents the landmarks of a male marsh pupfish. The top left and right shapes represent the landmarks for female and male pupfishes which are inhabitants of sinkhole.

The difference in shape with respect to population mean is shown for the two extremes of each principal component using thin-plate spline deformation grid. As seen, these results indicate that there may exist two different species in these two distinct environments, i.e. marsh and sinkhole. Also, it is inferred that sex can play an important role in the body shape of pupfishes. These observations justify the use of change point analysis methods to investigate for significant shape difference between species.

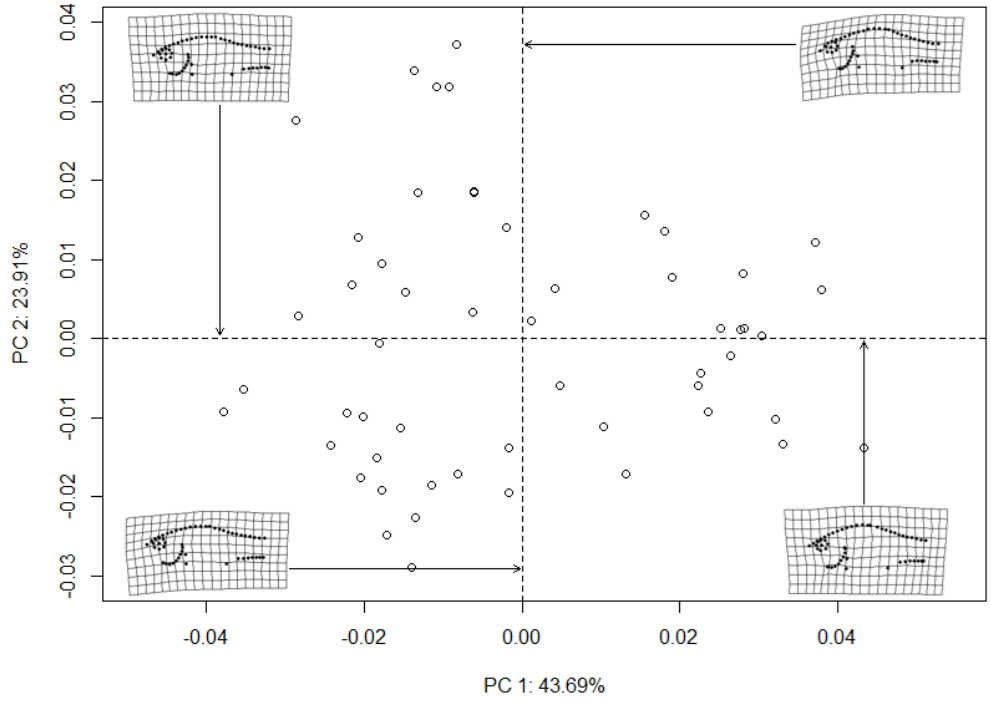


Figure 3.14: Principal component plot of shape variation in the population of pupfishes.

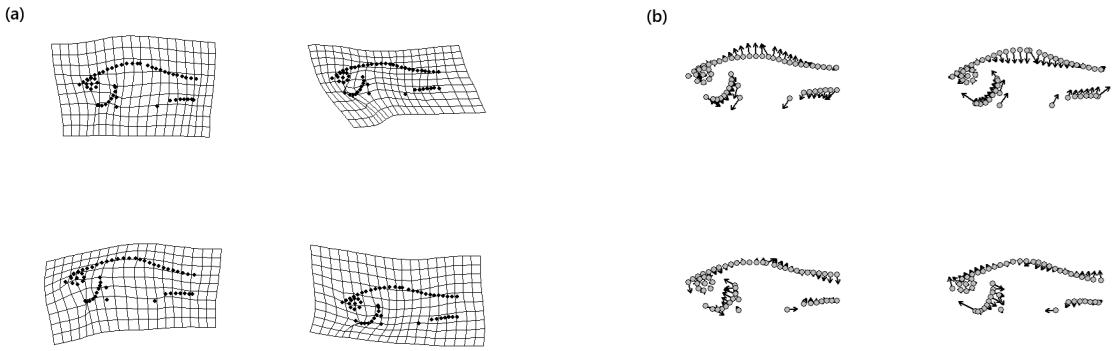


Figure 3.15: Shape difference between marsh and sinkhole pupfish specimens displayed as (a) thin-plate spline deformation grid for two extremes, (b) change direction vector of two extremes with respect to mean shape.

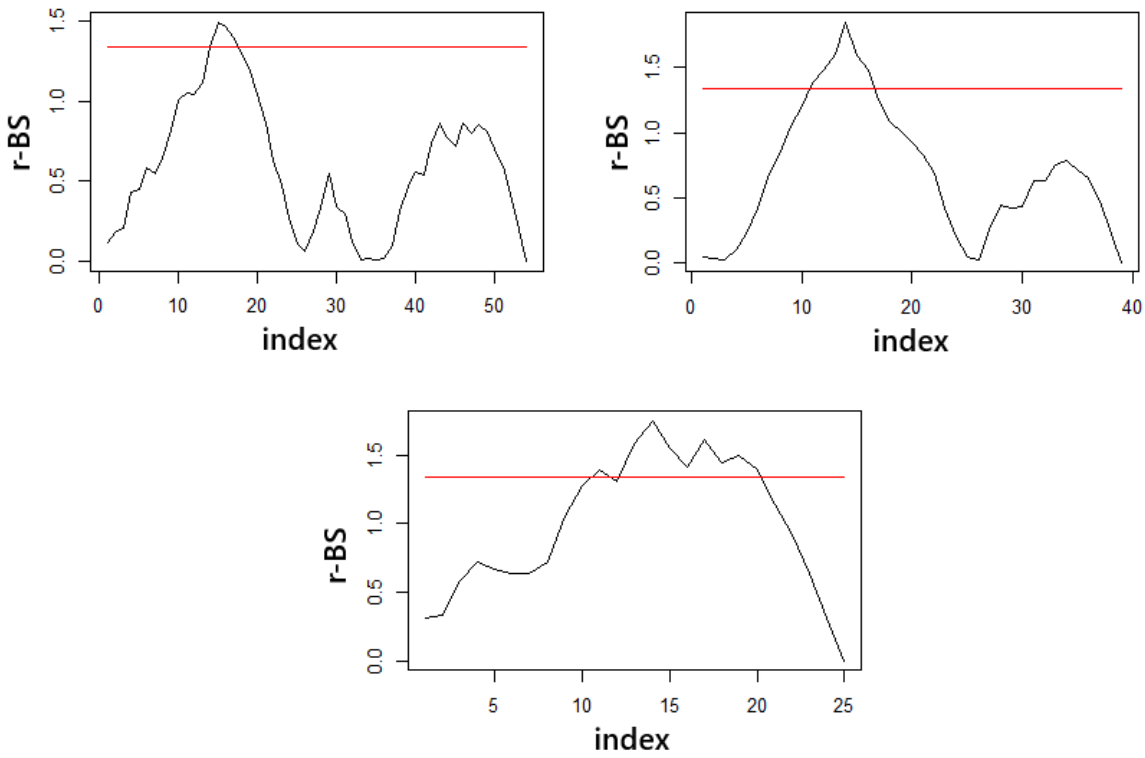


Figure 3.16: Change points obtained by r-BS. Three panels shown the location of changes in three stages of binary segmentation corresponding to locations $15, 29, 44$.

To further investigate the shape difference among groups of pupfishes, we plot the shape difference between the extreme shapes at each of these two principal components. Figure 3.15 indicates the differences in thin-plate spline deformation grid for two extremes as well as in directional change vector for the two extremes with respect to mean shape. We amplified the shape differences by a factor of 2 to aid in the description of shape differences and to facilitate biological interpretation. From the principal component analysis of shape variation, it is observed that there is a significant deformation in the morphological characteristics of the samples selected from extreme parts of principal coordinates. This also justifies that the body shape features of inhabitants large marsh system and small sinkhole are affected by environmental factors such as salinity and the presence of predators and other competitor species.

Figure 3.16 indicates the binary segmentation procedure and the location of change

Table 3.15: Estimated change points of the rival methods for pupfish data.

	Methods					
	r-BS	**r-WBS**	**BS**	**WBS**	**PELT**	**ECP**
$\widehat{k}$	$(15, 29, 44)$	$(15, 29, 43)$	$(15, 29, 49)$	$(16, 29, 49)$	$(16, 29, 35, 49)$	$(16, 30, 44, 52)$
$\widehat{\ell}$	3	3	3	3	4	4

points detected by r-BS. There are 3 changes in the distribution of the data at locations $15, 29, 44$. Table 3.15 compares the performance of the considered methods for pupfish data. The obtained results indicate that r-BS and r-WBS outperform the other methods as the locations and number of detected change points coincide with the locations where the sex of specimens and their living environment changes. This is inline with our findings in the PCA analysis of shapes where it was indicated that sex and living environment of specimens contribute to a total of 67% of variability in the data.

Example 12 (Video Processing): In this example, we consider two real-world problems related to video processing, which are traffic monitoring on highway and mall pedestrian surveillance, to elaborate on the performance of the proposed change point detection method. To do the simulation, we need to formulate the corresponding problems using the idea of flag manifolds (Ma et al., 2020).

Consider an increasing sequence $0 < v_1 < v_2 < \cdots < v_k < d$ of k positive integers, a flag f is defined as a nested sequence of sub-spaces

$$\varnothing \subseteq \mathbb{V}_1 \subseteq \mathbb{V}_2 \subseteq \cdots \subseteq \mathbb{V}_k \subseteq \mathbb{R}^d \ , \ \dim \mathbb{V}_i = v_i \ , \ i = 1, ..., k \ .$$

Flag manifold $\mathrm{Fl}(v_1, ..., v_k; d)$ is the space of single flags f. When $k = 1$, we recover the Grassmannian manifold $\mathrm{Gr}(v; d)$ which is the space of all v-dimensional sub-spaces of $\mathbb{R}^d$. Flag manifold $\mathrm{Fl}(v_1, ..., v_k; d)$ is diffeomorphic to the space

$$SO(d)/S(O(v_1) \times O(v_2 - v_1) \times O(v_3 - v_2) \times \cdots \times O(v_k - v_{k-1}) \times O(d - v_k))$$

where the stabilizer $S(O(v_1) \times O(v_2 - v_1) \times O(v_3 - v_2) \times \cdots \times O(v_k - v_{k-1}) \times O(d - v_k))$ can be mathematically expressed as

$$\begin{bmatrix} Q_1 & 0 & \dots & 0 \\ 0 & Q_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Q_{k+1} \end{bmatrix} \in O(d), \ Q_i \in O(v_i - v_{i-1}) \ , \ i = 1, ..., k+1, \ \prod_{i=1}^{k+1} \det(Q_i) = 1$$

which is the space of unit-determinant block diagonal matrices with orthogonal blocks.

Let $so(v)$ denote the vector space of $v \times v$ skew-symmetric matrices. The tangent space V is the lie algebra of the stabilizer

$$\begin{aligned}
\mathrm{v} &= \{Q : \exp(Q) \in V\} \\
&= \left\{ Q : \exp\left\{ \begin{bmatrix} Q_1 & 0 & \dots & 0 \\ 0 & Q_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Q_{k+1} \end{bmatrix} \right\}^\top \exp\left\{ \begin{bmatrix} Q_1 & 0 & \dots & 0 \\ 0 & Q_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Q_{k+1} \end{bmatrix} \right\}^\top = I \right\} \\
&= \{Q : Q_i + Q_i^\top = 0 \ , \ i = 1, ..., k+1 \in V\} \\
&= \{Q : Q_i \in so(v_i - v_{i-1}) \ , \ i = 1, ..., k+1 \in V\} \\
&= so(v_1) \times so(v_2 - v_1) \times \cdots \times so(v_k - v_{k-1}) \times so(d - v_k) \ .
\end{aligned}$$

The orthonormal complement of v, denoted by h, is represented as

$$\mathrm{h} = \left\{ \begin{bmatrix} 0 & B_{1,2} & \dots & B_{1,k+1} \\ -B_{1,2}^\top & 0 & \dots & B_{2,k+1} \\ \vdots & \vdots & \ddots & \vdots \\ -B_{1,k+1}^\top & -B_{2,k+1}^\top & \dots & 0 \end{bmatrix} \in so(d) : B_{i,j} = \mathbb{R}^{(v_i - v_{i-1}) \times (v_j - v_{j-1})}, 1 \le i < j < k+1 \right\}$$

The tangent space at flag f in the manifold space $\mathrm{Fl}(v_1, ..., v_k; d)$ is denoted by

$$T_f(\mathrm{Fl}(v_1, ..., v_k; d)) = \{fB \in \mathbb{R}^{d \times d} : B \in \mathrm{h}\} \ .$$

Consider a sample of pixel matrices of size $m \times \ell$ obtained from image data, namely $X_1, \cdots, X_n$. Each of these pixel matrices can be fattened in a row giving a vector of size $m\ell$. If we concatenate all these rows, we will obtain a matrix of size $n \times m\ell$, for which we can calculate the principal components and corresponding eigenvalues. We can truncate the first $p << d$ components corresponding to the first p eigenvalues representing the 95% variability in the data. These p ordered eigenbasis can then be represented as a flag, in the form of a sequence of nested subspaces, belonging to the manifold space $\mathrm{Fl}(1, 2, ..., p; d)$.

For the change point analysis, we can concatenate the observed flattened pixel matrices sequentially and calculate these $p << d$ eigenvalues, which gives our flag data for the analysis. The distance between any two flags f_i and f_j can be computed using an approximated Grassmannian projection. Consider the Grassmannian manifold $\mathrm{Gr}(1, d)$ induced by the first principal component. Define the Grassmannian projector

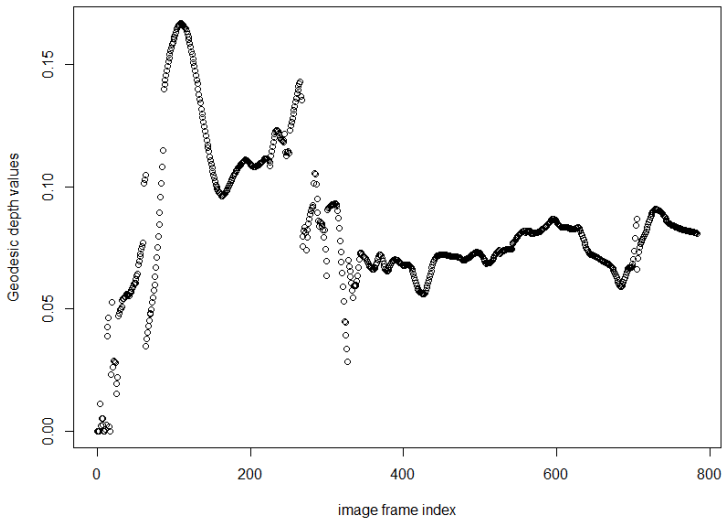


Figure 3.17: Geodesic depth values for flag object data.



Figure 3.18: Detected changes in the intensity of vehicles moving on the road.

$\phi : \mathrm{Gr}(1,d) \to \mathrm{Sym}(d) : [\mathbf{x}] \to \mathbf{x}\mathbf{x}^\top$, which maps the basis to the space of symmetric matrices which is itself a Hilbert space equipped with induced distance

$$d^2_{gr}([\mathbf{x}],[\mathbf{y}]) = \|\phi([\mathbf{x}]) - \phi([\mathbf{y}])\|^2_{\mathcal{H}} = 2 - 2(\mathbf{x}^\top\mathbf{y})^2$$

The projection can be extended for data of the form $f \in \mathrm{Fl}(1,2,...,p;d)$ using flag projector $\psi : \mathrm{Fl}(1,2,...,p;d) \to \mathrm{Sym}(pd)$ which maps flag data to the space of symmetric matrices with induced distance

$$d^2_{fl}([\mathbf{X}],[\mathbf{Y}]) = \|\psi([\mathbf{X}]) - \psi([\mathbf{Y}])\|^2_{\mathcal{H}} = 2\sum_{i=1}^{p}(1 - (\mathbf{x}_i^\top\mathbf{y}_i)^2)$$

GD depth can be calculated using the induced distance d^2_{fl} from which the ranks can be calculated to implement r-BS and r-WBS.

To elaborate on the performance of the methods, we used highway video data, which comprises of 1700 frames collected in a day, where each frame is represented using a grayscale image of size 240×320. Applying the strategy mentioned before, we can concatenate the flattened vectors of size 76800, as new frames are added to the database. Empirical studies indicated that 95% of the variability can be explained by the first 10 principal components, hence leading to a dataset with objects belonging to $\mathrm{Fl}(1, ..., 10; 76800)$, obtained by the first 10 ordered eigenbasis. Figure 3.17 indicates the calculated GD depth values for the flag data objects. From the calculated depth it can be inferred that the stochastic structure of the object data changes over time. r-BS and r-WBS detect 7 and 9 change points for the upcoming highway data. Comparison indicates that the detected change points are close and important changes in the intensity of vehicles passing through the road was identified accurately by both methods. Figure 3.18 indicates the frames detected by r-BS. By comparing the frames, it can be realized that significant changes occur in the intensity of vehicles on the road. For comparison, we also applied WBS to the same dataset. The method detects a large number of change points, which makes the analysis and the verification of the results difficult. This also endorses the efficacy of robust rank-based nonparametric method as the complexity of the problem and uncertainties associated with the nature of such dataset does not affect the analysis. The proposed change point detection method can be used for traffic monitoring on highways with application to important tasks such as traffic scheduling. In this way, for hours in a day with more vehicles on the road, appropriate traffic control strategies can be devised.

The next dataset contains mall pedestrians surveillance images with 2000 frames collected in a day, where each image is represented using grayscale pixel matrices of size 649×480. The pixel matrices data are flattened in the form of vectors of size 311520. Based on an empirical analysis, it was observed that considering the first 10 principal components can capture 95% of the variability in the dataset, thus suggesting object data belonging to $\mathrm{Fl}(1, ..., 10; 311520)$. Both r-BS and r-WBS detected 3 change points in the collected sequence of images. Figure 3.19 indicates the detected frames from which we can realize a significant change in the intensity of passengers in the mall passage. WBS and BS are also applied to the same dataset, yielding a large number of change points, thus cannot be used for a practical analysis. The change point detection method can be used for mall security service and passage control, which is an important real-world task. In this way, for busier places in the mall as well as for the time segments of a day with more passengers in the mall, appropriate security and monitoring plans can be deployed to guarantee safety and shoppers' comfort.

For both of these datasets, one can realize that the type of change is change in the location, as the corresponding images belong to different locations on the space of flag



Figure 3.19: Detected changes in the intensity of passengers for mall pedestrian surveillance data.

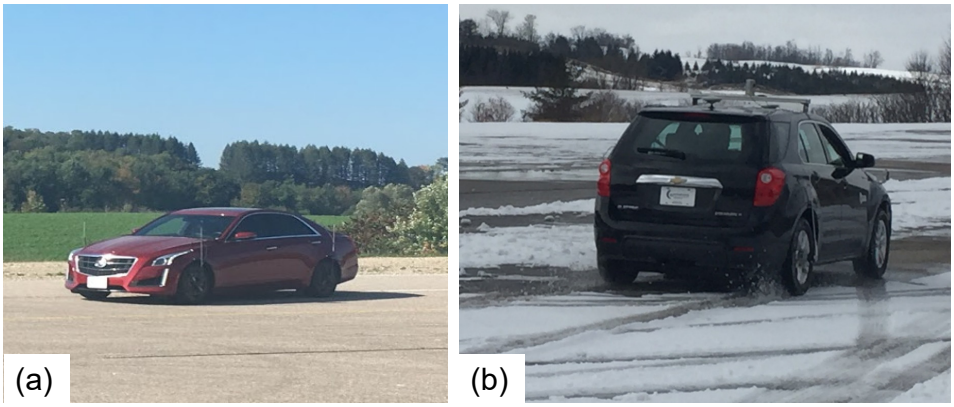


Figure 3.20: Test vehicles: (a) A Cadillac ATS, with electronic limited slip differential on the rear axles; (b) fully electrified vehicle with independent-drive configurations

manifold. Variability of the data in the corresponding segments does not have a significant change. This can be verified by the fact that for each segment, the distance of images from the median image of the corresponding segment falls in a ball of a comparable radius.

Example 13 (Vehicle Health Monitoring): In this example, we carry out experiments to evaluate the performance of change point detection method for vehicle health monitoring (Sun et al., 2023). The objective is to monitor for possible faulty readings in the sensors mounted in the vehicle. The analysis is done using a set of experimental data for estimating the vehicle lateral speed on roads. This dataset is provided by mechatronic vehicle systems (MVS) lab at the University of Waterloo.

The input signals are the rotational speed of vehicle four wheels (in rad/s), longitudinal/lateral accelerations (in m/s^2), yaw rate (in rad/s), and steering wheel angle (in rad), and the lateral speed. The dataset is represented by $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$. The test vehicle platforms used for evaluation of the proposed method are Cadillac ATS and

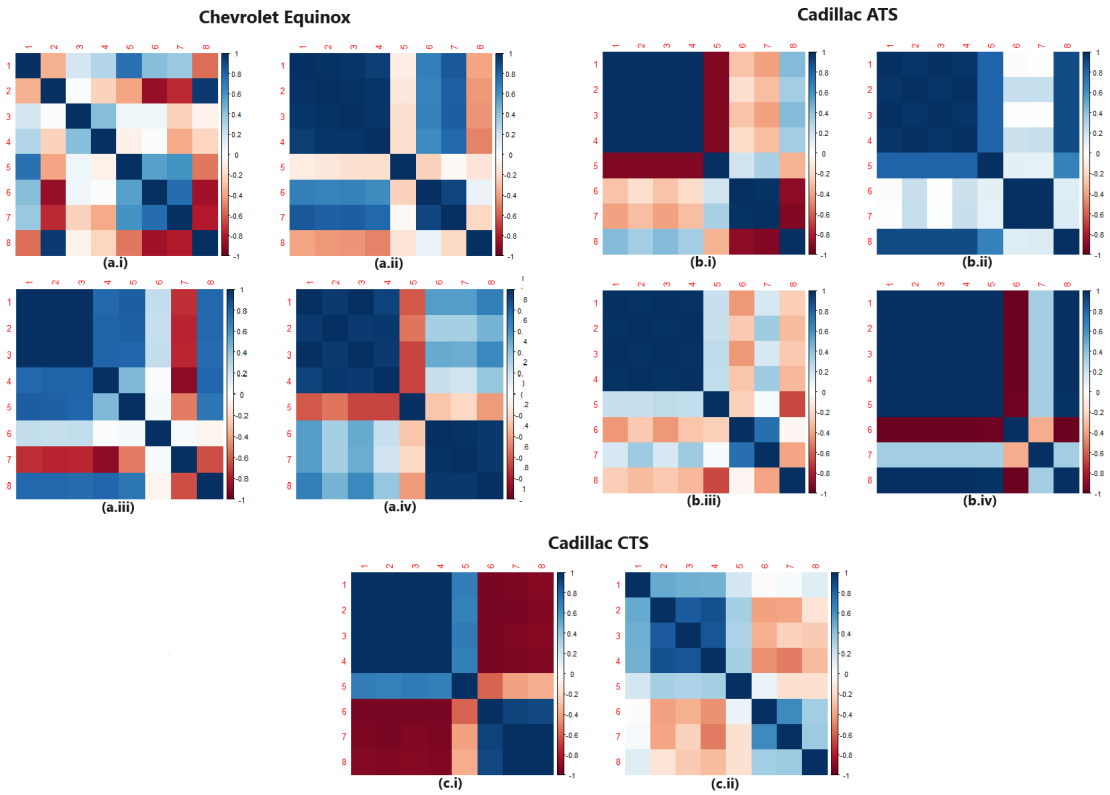


Figure 3.21: Detected changes for the correlation matrices of the considered vehicle platforms, (a) Chevrolet Equinox segments (i) $[1, 63]$, (ii) $[64, 119]$, (iii) $[120, 158]$, (iv) $[159, 170]$; (b) Cadillac ATS segments (i) $[1, 122]$, (ii) $[123, 142]$, (iii) $[143, 158]$, (iv) $[159, 190]$; (c) Cadillac CTS segments (i) $[1, 117]$, (ii) $[118, 170]$

Cadillac CTS, with electronic limited slip differential on the rear axles, and a fully electrified sport utility vehicle (Chevrolet Equinox) with independent-drive configurations as shown in Figure 3.20. The authors performed a pre-processing and discarded data points with missing arrays. The algorithm is tested on an *Intel Core* 2, CPU 6600, 2.40 GHz, and 2 GB RAM using the sensory measurements (i.e., wheel speeds, active front steering, accelerations, and yaw rate) from the three test vehicle platforms.

At each time point the 8 sensors read the relevant signals and then a batch of data of size 100 is collected from which we calculate a correlation matrix as an object data. The matrix data are then considered as Hermitian positive definite manifold data and the depth and rank of these object data are calculated. In case that any fault is introduced into the

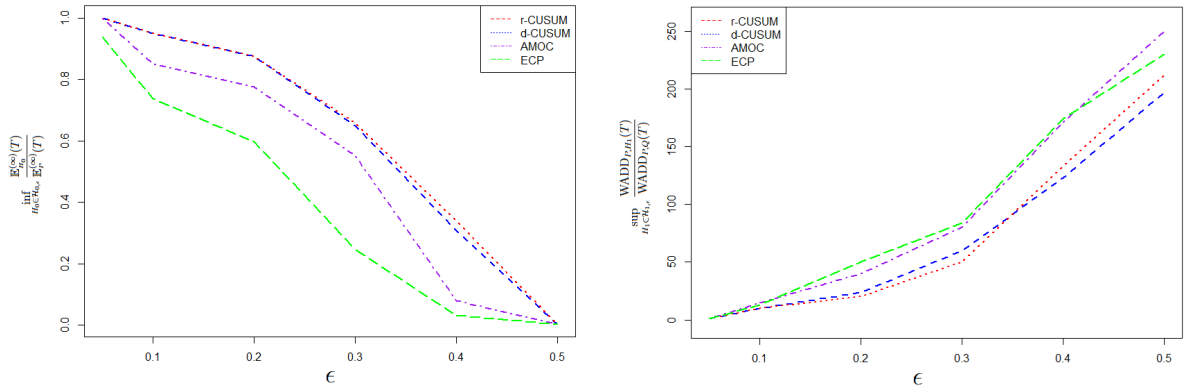


Figure 3.22: Gross-error sensitivities for r-CUSUM, d-CUSUM, AMOC, and ECP for Wishart distribution under null and alternative hypotheses with different values of ε

system, this faulty signal reading will emerge as an abrupt change or significant deviation which can result in object data with outlying behaviour. Such changes will also affect the statistical linear dependence among sensor readings, calculated by correlation matrices, and therefore the depth and rank of the data will have a statistically significant change. If such a change is detected, then it is easy to monitor the corresponding individual readings. Figure 3.21 indicates the change points detected by r-BS for the correlation matrix data. It can be seen that there is a significant change in the correlation matrix readings at each of the segments, which endorses the validity of the change point detection. By checking the corresponding sensor readings, it is also observed that some of the sensors have outlier readings (significant deviation from immediate neighbour data), which could be because of environmental factors such as wind, traffic, turning point, etc. Note that the same experiment was performed by r-WBS which yielded similar result to r-BS.

3.6.3 Robustness Analysis

Example 14 (Breakdown Point Analysis for Wishart Distribution on SPD(3)): In this example, we investigate the robustness of r-CUSUM, d-CUSUM, AMOC, and ECP using breakdown point for Wishart distribution. Under the the null hypothesis condition, we calculate the ratio $\inf_{H_0 \in \mathcal{H}_{0,\epsilon}} \frac{\mathbb{E}_{H_0}^{(\infty)}(T)}{\mathbb{E}_P^{(\infty)}(T)}$ as a function of contamination level ϵ. Under the alternative hypothesis, we calculate the ratio $\sup_{H_1 \in \mathcal{H}_{1,\epsilon}} \frac{\text{WADD}_{P,H_1}(T)}{\text{WADD}_{P,Q}(T)}$ as a function of contamination level ϵ.

Figure 3.22 depicts the gross-error sensitivities for r-CUSUM, d-CUSUM, AMOC, and ECP under the null and alternative hypotheses for different values of ϵ. The left panel evaluates the robustness of the proposed methods under the null hypothesis. Clearly, increasing the contamination level decreases the ratio $\inf_{H_0 \in \mathcal{H}_{0,\epsilon}} \frac{\mathbb{E}_{H_0}^{(\infty)}(T)}{\mathbb{E}_{P}^{(\infty)}(T)}$ so that it approaches 0. The obtained results suggest false-alarm breakdown point $\varepsilon_F = 0.4$ for r-CUSUM and d-CUSUM and $\varepsilon_F = 0.3$ for AMOC and ECP. The right panel compares the robustness of the considered methods under the alternative hypothesis. It can be seen that increasing the contamination level increases the ratio $\sup_{H_1 \in \mathcal{H}_{1,\epsilon}} \frac{\text{WADD}_{P,H_1}(T)}{\text{WADD}_{P,Q}(T)}$ so that it becomes very large. The results suggest detection-delay breakdown point $\varepsilon_D = 0.4$ for r-CUSUM and d-CUSUM and $\varepsilon_D = 0.3$ for AMOC and ECP. It can be concluded that rank-depth change point detection method is robust given that the obtained values for ε_F and ε_D are reasonably large. Moreover, it can be inferred that the use of depth makes the results more robust compared to centroid estimation based Riemannian change point detection strategy, as the value of ε_F and ε_D for r-CUSUM is larger than those of AMOC and ECP.

3.7 Discussion and Future Work

In this chapter, we formulated rank-depth change point detection methods for single and multiple change point analysis where data are sampled from manifold spaces. The effectiveness of the proposed methods were evaluated considering Monte Carlo simulation and real-world analysis. It was demonstrated that rank-depth change point detection methods are robust and can effectively cope with object data from various types of manifold spaces, e.g. sphere, torus, SPD, flag manifold, shape space, networks and graphs.

Simulation results proved that these change point detection methods can keep the empirical size at a nominal level while maximizing the empirical power under alternative hypothesis. Besides, they could effectively tackle problems with small sample sizes and large dimensions. Finite sample Monte Carlo studies endorsed the validity of the asymptotic results derived for these nonparametric change point detection methods. Several well-known methods including AMOC, PELT, ECP, BS, and WBS were extracted from the literature and used to evaluate these rank-depth change point detection methods.

The comparative simulation results demonstrated that the proposed methods could outperform these competitors for most of the cases. Also, it was observed that the proposed methods can be successfully applied to both intrinsic and extrinsic manifold analysis frameworks. Robustness of the methods were studied from both theoretical and numerical viewpoints. Theoretical analysis involved q-robustness analysis as well as asymptotic

breakdown point analysis. It was observed that these methods can show superior performance when data are contaminated with outliers.

Future research direction involves applying the proposed nonparametric change point detection methods to problems from various research disciplines including dynamic networks, longitudinal data, robotics and computer vision. It is believed that rank-depth change point detection methods are effective in such applications given their generalization power and the fact that they require no assumption regarding the distribution of object data.

Chapter 4

Process Monitoring

In this chapter, nonparameteric control chart based on rank of depths are proposed for monitoring processes defined on manifold spaces $\mathcal{M}$. The rest of the chapter is organized as follows. In Section 1, an introduction along with a discussion of important methodologies of process monitoring is presented. In Section 2, a formulation of process monitoring scheme on general metric spaces as well as ranking based version of control charts such as exponentially weighted moving average (EWMA), CUSUM, and Mann-Whitney moving average Shewhart (MWMASH) are presented. In Section 3, a comparative simulation study is carried out to investigate the performance of the proposed methods. In Section 4, the proposed control charts are considered for monitoring additive manufacturing (AM) processes. Real-world examples are given to demonstrate the efficacy of the proposed methods. In Section 5, the findings of the study as well as future research directions are outlined.

4.1 Process Monitoring for Manifold Data

Development of advanced industrial tools and manufacturing systems have motivated researchers from various disciplines, e.g. control and mechatronics engineering and statistical society, to propose algorithms for monitoring the quality of products. In such applications, known as statistical process monitoring (SPM), the main objective is to ensure high-quality products as well as high-quality services in order to satisfy customer needs in the modern business world. There has been a significant amount of research publications since the inception of this idea, covering various aspects of process monitoring and quality control (Ge et al., 2013; Tapia and Elwany, 2014). Usually, computers and data acquisition equipments

are used online to collect some control variables/measurements. These measurements represent several quality characteristics of the underlying process.

Traditionally, the development of quality control charts for real-time surveillance of the quality characteristics is done in two phases. In Phase I of SPM, the focus is on understanding process variability, outlier detection, parameter estimation, data cleaning, and selection of appropriate control variates based on which statistical control charts and in-control (IC) operating conditions of the process are developed. In Phase II of SPM, the developed control chart and reference sample, i.e., in-control distribution, are used online to monitor the process and detect possible out-of-control (OC) performance. Fault can occur (i.e., out-of-control operating conditions) when the data-generating process shifts in location and/or scale. Once an out-of-control condition is detected, the process is stopped in order to investigate for the special causes of variation (Montgomery, 2009). Control charts are usually designed to monitor small persistent shifts, e.g. exponentially weighted moving average control chart (EWMAC), or abrupt big shifts, e.g. Shewhart control chart (SHC) (Montgomery, 2009), in the in-control distribution. For a comprehensive literature review on this topic, readers can refer to seminal review papers (Yin et al., 2014; Qin, 2012; Joe Qin, 2003).

A recent trend has emerged in the literature towards extending the classical univariate SPM methods. This is indeed necessary due to the availability of sensors and internet technology, which enabled the collection of hundreds of measurements related to production. As a result of this progress, multivariate SPM methods have been introduced in the literature, which makes it possible to control several correlated quality characteristics simultaneously (Ge and Song, 2012; Kresta et al., 1991). These methods are usually based on the assumption of normality of control variables based on which parametric control charts such as Hotelling T^2 can be implemented. The outcome of these research studies demonstrated the superiority of multivariate SPM (MSPM) compared to using individual univariate SPM for each control variable separately. The reason is that MSPM takes the correlation among these control variables into account. As a result, there is no loss of information when monitoring for out-of-control conditions and studying causes of variability.

In line with such fruitful observations and with the invention of high-rate data acquisition sensors, research has been extended to the areas where the outcome of the process has a profile nature, i.e. functional quality characteristics (Colosimo and Pacella, 2010). Real-time monitoring of functional data, also known as profile monitoring, requires techniques such as registration, dimensionality reduction, phase and amplitude decomposition, and time-warping methods for aligning phase and amplitude components of functional signals (Centofanti et al., 2022). Hotelling T^2 control chart can be implemented based on the score and variance of principal components for real-time monitoring of such processes. In

Centofanti et al. (2022), it was demonstrated that these types of control charts can successfully distinguish between normal sources of variation (common causes) from assignable sources of variation (special causes).

Despite the remarkable progress in developing MSPM and profile monitoring methods, the literature lacks research publications on processes where data points belong to general non-Euclidean manifold spaces. In del Castillo and Zhao (2014), it was indicated that in many modern applications, the seemingly high-dimensional data lie on lower-dimensional curved subspaces or manifolds. It was demonstrated that multivariate control charts can be developed based on spectrum analysis of the Laplace-Beltrami operator. There exist some publications in the literature which address the use of local-global feature extraction methods and traditional manifold learning algorithms for process monitoring of manifold data (Zhan et al., 2019; del Castillo and Zhao, 2014; Zhang et al., 2023; Liu et al., 2021). One drawback associated with such manifold learning methods is that these methods can become computationally intensive, especially for process monitoring tasks where real-time speed is of the essence. In line with this observation, in this chapter, a nonparametric control chart method is developed that only requires rank-of-depth functions to monitor the in-control operating condition of manifold data processes. It will be demonstrated that this general framework can be applied to a wide range of processes with object data sampled from various manifold spaces. The developed method is nonparametric in the sense that it does not require any assumption on the distribution of the manifold data, which makes it favorable for process monitoring applications. The method is robust against outlying and object data collected from heavy-tail, asymmetric distributions.

Some literature on robust and nonparametric MSPM follows the use of rank and depth function for developing control charts (Pandolfo et al., 2021; Haanchumpol et al., 2020; Qiu and Hawkins, 2001). This is crucial because any unnecessary assumption on the distributional characteristics of data can limit the application of the SPM method. Also, selecting a wrong distribution for the data can undermine the performance of the control chart method and result in a large false alarm rate. For the reason of such deficiencies, several rank-based SPM methods have been proposed in the literature.

In Chen et al. (2016), the authors proposed a nonparametric method based on the Wilcoxon rank-sum test for monitoring multivariate quality variables and data streams. The method is developed for changes in location parameters when only a small reference sample is available. It was indicated that the proposed charting statistic can achieve a specified false alarm rate. The advantage associated with the monitoring scheme is the use of data-dependent control limits, which are calculated online. Theoretical results indicated that the control chart attains the desired in-control run length regardless of the distribution and dimension of data. In particular, the method effectively detects multivariate shifts for

heavy-tailed and skewed data. In Qiu and Hawkins (2001), the authors proposed a CUSUM control chart for multivariate data based on cross-sectional anti-ranks of the measurements. It was indicated that when the joint distribution of multivariate data satisfies certain regularity conditions, the CUSUM statistic based on anti-ranks follows χ^2 distribution. When there is a shift in the mean vector of the process, the distribution of anti-ranks changes for out-of-control operating conditions. The method is nonparametric in the sense that all of the properties of the control chart can be specified based on the distribution of anti-ranks only, rather than the distribution of data. The performance of the method was verified considering a multivariate SPM task where the normality assumption was inapplicable. In Megahed et al. (2011), the authors reviewed parametric and nonparametric SPM methods applied to machine vision systems and medical devices, which are based on image-based analysis. It was discussed that charting statistics such as spatial control charts, e.g. based on a spatial design using a moving mask (a window) across an image, can be used for such applications and can afford promising outcomes. In Li et al. (2017), the authors developed a self-starting control chart which combines a multivariate spatial rank test with an EWMA charting scheme based on forward variable selection for sparse mean shift detection. Theoretical and numerical simulation investigations indicated that the test statistic is robust to non-normal data and efficient in sparse mean shift detection, especially when data is heavy-tailed and skewed. The efficacy of the method was proven by considering real-world examples such as the white wine production process.

In Zou et al. (2011), the authors formulated SPM within the framework of a two-sample variable selection problem based on the Bayesian information criterion (BIC). Adaptive LASSO variable selection is combined with BIC to perform SPM with a computational cost similar to least square regression. The effectiveness of the method was demonstrated using several SPM examples, e.g. multistage process control and profile monitoring. In Shu and Fan (2018), the authors proposed a novel distribution-free control chart based on interpoint distances for monitoring mean vectors in high-dimensional settings. It was indicated that the method is generalizable in the sense that standard Euclidean distance can be replaced with general distances such as Minkowski distance. The method is distribution-free and only depends on distances. It was demonstrated that the control chart effectively detects the mean shift in high-dimensional data for both symmetric and skewed distributions. In Li et al. (2013), the authors proposed two novel nonparametric multivariate CUSUM control charts based on the spatial sign and data depth for detecting location and scale change, which do not assume the normality of data. It was indicated that the two test statistics are affine invariant and distribution-free and can be effectively applied to a broad family of distributions. In Liu et al. (2004), the authors proposed a control chart based on data depth-based moving averages for monitoring multivariate data. The proposed control chart

is distribution-free and can be used for both changes in location and scale of the process.

In Merlo et al. (2022), the authors proposed a nonparametric Shewhart-type control chart based on the combination of Mahalanobis distance with Mann-Whitney statistic for multivariate process monitoring with guaranteed IC performance for shift in location. The method is based on ranking of the Mahalanobis distance and does not require any assumption regarding the mean vector or covariance matrix of the underlying multivariate distribution. It was demonstrated that the approach enables a quick determination of control limits which guarantee a minimum performance given a Phase-I sample.

In Mukherjee and Marozzi (2021), the authors proposed a control chart for high-dimensional setup where the number of quality variables can become much larger than the sample size. The proposed method is based on Shewhart-type nonparametric control chart using distance metrics for monitoring high-dimensional processes. The control chart can be used for simultaneous monitoring of both location and scale parameters.

In Zhang et al. (2020), the authors proposed a nonparametric control chart based on spatial rank for applications where joint distribution of process variable is usually unknown and not normal. The method is applicable to high dimensional processes where the number of dimension is much larger than the number of sample. The method is implemented based on random projection where the original process is randomly projected into several subprocesses using ensemble random projection for dimensionality reduction. The control chart based on spatial rank is then applied to each of these subprocesses and the results are then fused for a final decision making. The authors also implemented a diagnosis framework based on square-root LASSO algorithm.

In Wang et al. (2023), the authors proposed a rank-based EWMA control chart for monitoring high-dimensional processes with heteroscedasticity. The method is implemented for detecting shift in the location of the high-dimensional processes. Bootstrapping is then used to determine the control limits by attaining a prespecified false alarm rate. The method has the advantage that it is robust to heteroscedaisticity and is sensitive to small and moderate mean shifts.

In Chakraborty and Finkelstein (2024), the authors proposed a distribution-free multivariate process monitoring scheme based on rank-energy statistic-based approach for high-dimensional processes. The rank-energy statistic is based on multivariate ranking based on measure transportation theory. Multivariate ranks can be viewed as solution to an optimization problem in a multivariate setting in the form of measure transportation problem. It can be shown that the method favors a comparable degree of robustness compared to methods such as spatial sign and depth-based ranking.

In Ding et al. (2016), the authors proposed a rank-based control chart for mixed-type

data where the underlying processes contain both continuous and categorical control variates. Such processes are becoming more important in modern manufacturing applications. The method is implemented based on the assumption that we can find a latent continuous variable that determines the attribute level of categorical variables. The proposed control chart is a variant of EWMA based on ranks. Numerical simulations indicated that the method can detect location shifts and is robust to various distributional properties.

From the abovementioned literature review, it can be realized that nonparametric and rank-based control charts are effective methods for robust statistical process monitoring of processes with various distributional features. The current study extends the above literature by applying the idea of depth-based ranking to data belonging to non-Euclidean manifold spaces. Moreover, theoretical analysis is carried out to investigate the finite-sample properties of the developed control chart for data belonging to a wide range of manifold spaces with various geometrical features.

4.2 Control Charts

The proposed control charts are nonparametric and are based on using the rank of the depth of the object data collected from corresponding manifold spaces. To design the control charts, two sample sets are used. The first sample set is called Phase-I observations which is assumed to be in-control with Fréchet mean μ_x and Fréchet variance ν_x. We denote this set by $(X_1, X_2, ..., X_{n_1})$. The second sample set $(Y_1, Y_2, ..., Y_{n_2})$ is called Phase-II observation which is usually collected in batches. We denote the Fréchet mean and Fréchet variance of this set by μ_y and ν_y. We assume these sample sets take values in a bounded metric space (Ω, δ) with probability measures $X \sim P$ and $Y \sim Q$.

Due to the non-Euclidean nature of data collected from additive manufacturing (AM) processes, there is not much flexibility in the selection of charts and working with distance measures, as it is commonly available in the development of univariate and multivariate charts.

In the previous section, it was mentioned that a proper way for mathematical modeling and statistical analysis for AM monitoring is to work with general metric spaces and notions such as Frechet mean and Frechet variance w.r.t. distance on the corresponding space. Once a proper distance measure is used for the analysis, data depths can be calculated which enable us to use nonparametric control charts based on the ranks of data depths, e.g. rank-based CUSUM, rank-based EWMA, rank-based MWMASHC etc.

The Fréchet mean and Fréchet variance are defined as

$$\mu_x = \arg\min_{\omega\in\Omega} \mathbb{E}[\delta^2(X,\omega)] \ , \quad \mu_y = \arg\min_{\omega\in\Omega} \mathbb{E}[\delta^2(Y,\omega)] \ ,$$

$$\nu_x = \min_{\omega\in\Omega} \mathbb{E}[\delta^2(X,\omega)] \ , \quad \nu_y = \min_{\omega\in\Omega} \mathbb{E}[\delta^2(Y,\omega)] \ .$$

Once these datasets are collected, we can use empirical version of Fréchet mean and Fréchet variance defined as

$$\widehat{\mu}_x = \arg\min_{\omega\in\Omega} \frac{1}{n_1}\sum_{i=1}^{n_1} \delta^2(X_i,\omega) \ , \quad \widehat{\mu}_y = \arg\min_{\omega\in\Omega} \frac{1}{n_2}\sum_{i=1}^{n_2} \delta^2(Y_i,\omega) \ ,$$

$$\widehat{\nu}_x = \frac{1}{n_1}\sum_{i=1}^{n_1} \delta^2(X_i,\widehat{\mu}_x) \ , \quad \widehat{\nu}_y = \frac{1}{n_2}\sum_{i=1}^{n_2} \delta^2(Y_i,\widehat{\mu}_y) \ .$$

In Dubey and Müller (2019), it was shown that for bounded metric space (Ω,δ), the sample mean and variance are consistent estimators of theoretical mean and variance, i.e. $\delta(\mu,\widehat{\mu}) = o_p(1)$ and $|\widehat{\nu}-\nu| = o_p(1)$.

For a metric spaces (Ω,δ), with probability measures $X \sim P$ and $Y \sim Q$, and any $r > 0$, the following depth function can be defined

$$D(P,\omega) = (1+\mathbb{E}_P[\delta^r(X,\omega)])^{-1} \ , \quad D(Q,\omega) = (1+\mathbb{E}_Q[\delta^r(Y,\omega)])^{-1} \ ,$$

to measure how central the point $\omega \in \Omega$ is w.r.t. a given probability measure.

For $(X_1, X_2, ..., X_{n_1})$ and $(Y_1, Y_2, ..., Y_{n_2})$, the empirical version of the depth functions are defined as

$$D(P_{n_1},\omega) = \Big(1+\frac{1}{n_1}\sum_{i=1}^{n_1}\delta^r(X_i,\omega)\Big)^{-1} \ , \quad D(Q_{n_2},\omega) = \Big(1+\frac{1}{n_2}\sum_{i=1}^{n_2}\delta^r(Y_i,\omega)\Big)^{-1} \ .$$

Once the depths are calculated, we can calculate the depth ranks $R_1, R_2, ..., R_{n_1+n_2}$. The proposed control charts are developed using the rank of depths and serve as nonparametric process monitoring tools.

4.2.1 Rank-Depth EWMA Control Chart

The first control chart is r-EWMA which is derived by implementing EWMA statistics using rank of data depths. Let R_t denote the rank of the depth value of t^{th} observation. The r-EWMA control chart can be implemented as follows

$$\mathbf{S}_t^{EWMA} = (1-\lambda)\mathbf{S}_{t-1}^{EWMA} + \lambda R_t \ , \quad t = 1, 2, ..., n_1+n_2 \ ,$$

where $\mathbf{S}_0^{EWMA} = 0$ and $\lambda \in (0, 1]$.

As can be inferred, r-EWMA uses a weighted average of the current observation and the chart value obtained at the previous time to monitor the process. The value λ determines the importance of the historical data for monitoring the current products, and also can be set in a way that the chart detects small but persistent shifts from IC condition. The statistic $\mathbf{S}_t^{EWMA}$ gives an upward shift signal if $\mathbf{S}_t^{EWMA} > UCL$ and a downward shift signal if $\mathbf{S}_t^{EWMA} < LCL$, where LCL and UCL are set such that ARL_0 is kept at a specified value.

Assumption 11. *It is assumed that random objects X_i's are independent and identically distributed from probability measure P on manifold space $\mathcal{M}$.*

Assumption 12. *It is assumed that X's are sampled from distribution P on manifold space $\mathcal{M}$, and* $\text{var}(X) < \infty$.

Theorem 25. *Suppose that assumptions 11-12 are satisfied. Under H_0, it holds that the two-sided control chart $\mathbf{S}_t^{EWMA}$ converges in distribution to*

$$\sqrt{\frac{2-\lambda}{\lambda}}\left(\frac{\mathbf{S}_t^{EWMA} - \zeta}{\sigma}\right) \xrightarrow{\mathcal{D}} \text{N}(0, 1) \ ,$$

and the one-sided control chart $\mathbf{\Lambda}_t^{EWMA}$ converges in distribution to

$$\mathbf{\Lambda}_t^{EWMA} = \frac{2-\lambda}{\lambda\sigma^2}\left(\mathbf{S}_t^{EWMA} - \zeta\right)^2 \xrightarrow{\mathcal{D}} \chi_1^2 \ ,$$

where $\sigma^2 = Var[R_t]$, and $\zeta = \mathbb{E}[R_t]$.

4.2.2 Rank-Depth CUSUM Control Chart

The next nonparametric control chart is r-CUSUM derived by implementing CUSUM chart using rank of data depths. Consider a sample of size n from distribution P, r-CUSUM uses the normalized depth rank function

$$R(x) = \frac{\#\{X_i \mid D(P, X_i) \leq D(P, x), i = 1, ..., n\}}{n} \ .$$

If a sample is closer to the center of the empirical distribution P_n, there will be many observations with smaller depth, thus, $R(x)$ will be closer to 1. If an observation is on

a tail of the distribution, $R(x)$ will be closer to 0. Therefore, $1 - R(x)$ can be used to quantify the relative distance between x and the center of the reference sample. The r-CUSUM control chart can be implemented as follows

$$\mathbf{S}_t^{CUSUM_1} = \max\{0, \mathbf{S}_{t-1}^{CUSUM_1} + (1 - R_t) - k\} \ , \quad t = 1, 2, ..., n_1 + n_2 \ ,$$

$$\mathbf{S}_t^{CUSUM_2} = \min\{0, \mathbf{S}_{t-1}^{CUSUM_2} + (1 - R_t) + k\} \ , \quad t = 1, 2, ..., n_1 + n_2 \ ,$$

where $\mathbf{S}_0^{CUSUM_1} = \mathbf{S}_0^{CUSUM_2} = 0$ and $k > 0$.

As can be seen, r-CUSUM is developed based on the cumulative summation of the relative distance of the observations from the center of the sample distribution. The control parameter k can be optimally tuned so that the chart detects small but persistent shifts from IC condition.

The statistic $\mathbf{S}_t^{CUSUM_1}$ gives an upward shift signal if $\mathbf{S}_t^{CUSUM_1} > UCL$ and the statistic $\mathbf{S}_t^{CUSUM_2}$ gives a downward shift signal if $\mathbf{S}_t^{CUSUM_2} < LCL$, where LCL and UCL are set such that IC average run length (ARL_0) is kept at a specified value, where ARL_0 is defined as

$$ARL_0 := \mathbb{E}_P\big[\min\{t \mid \mathbf{S}_t > \text{UCL or } \mathbf{S}_t < \text{LCL}\}\big] \ ,$$

where $\mathbb{E}_P$ is the expectation operator with respect to the IC distribution P.

It is known that $1 - R_t$ follows Uniform$[0, 1]$ distribution, when process is IC. Given that the mean of Uniform$[0, 1]$ is 0.5, we can modify the above control chart as follows

$$\begin{aligned}\mathbf{S}_t^{CUSUM_1} &= \max\{0, \mathbf{S}_{t-1}^{CUSUM_1} + ((1 - R_t) - 0.5) - k\} \\ &= \max\{0, \mathbf{S}_{t-1}^{CUSUM_1} + (0.5 - R_t) - k\} \ ,\end{aligned}$$

$$\begin{aligned}\mathbf{S}_t^{CUSUM_2} &= \min\{0, \mathbf{S}_{t-1}^{CUSUM_2} + ((1 - R_t) - 0.5) + k\} \\ &= \min\{0, \mathbf{S}_{t-1}^{CUSUM_2} + (0.5 - R_t) + k\} \ .\end{aligned}$$

It was observed that the modified version of r-CUSUM has a more robust performance compared to the standard r-CUSUM version. Therefore, we will use the modified version for our simulation study.

Theorem 26. *Suppose that assumptions 11-12 are satisfied. Under H_0, the two-sided control chart $\mathbf{S}_t^{CUSUM}$ converges in distribution to*

$$\frac{\mathbf{S}_t^{CUSUM} - \zeta}{\sigma} \xrightarrow{\mathcal{D}} \text{N}(0, 1) \ ,$$

and the one-sided control chart Λ_t^{CUSUM} *converges in distribution to*

$$\mathbf{\Lambda}_t^{CUSUM} = \frac{1}{\sigma^2}\Big(\mathbf{S}_t^{CUSUM} - \zeta\Big)^2 \xrightarrow{\mathcal{D}} \chi_1^2 \ ,$$

where $\sigma^2 = Var[R_t]$*, and* $\zeta = \mathbb{E}[R_t]$*.*

4.2.3 Rank-Depth MWMASHC Control Chart

The last control chart considered in this investigation is a variant of Shewhart chart (SHC). The standard implementation of SHC requires batch data $j = 1, 2, ..., m$ for each sample product $i = 1, 2, ..., n_2$. For conditions, such as the problems considered in the current case study, in which only a single product is produced at each time instance t, the idea of moving window of size m can be applied. This is done by grouping the n_2 data points into $n_2 - m + 1$ groups by defining the collection of object data $\mathbf{g}_t = (Y_{t-m+1}, ..., Y_t)$ for $t = m, m+1, ..., n_2$, where Y_i's are the original sample points obtained sequentially in Phase-II.

The depth ranks of these datasets can be calculated using reference sample $(X_1, X_2, ..., X_{n_1})$. We denote the grouped rank at time t by $\mathbf{R}_{g_t} = (R_{t-m+1}, ..., R_t)$ for $t = m, m+1, ..., n_2$. We use Mann-Whitney test statistic to design the moving average based Shewhart control chart (MWMASHC), formulated as follows

$$\mathbf{S}_{0,t}^{MWMASHC} = \sum_{i=1}^{n_1}\sum_{j=1}^{m}\mathbb{I}\Big(\mathbf{R}_{g_t}(j) > R_i\Big) \ , \quad t = m, m+1, ..., n_2 \ ,$$

where m is the batch size / horizon length of MWMASHC and can be tuned optimally. As the main goal of SHC is to detect large transient shifts, it is logical to select smaller values of m.

We use the normalized version of the above control chart for monitoring, defined as follows

$$\mathbf{S}_t^{MWMASHC} = \frac{\mathbf{S}_{0,t}^{MWMASHC} - \zeta}{\sigma} \ ,$$

where

$$\zeta = \frac{mn_1}{2} \ ,$$
$$\sigma^2 = \frac{mn_1(m + n_1 + 1)}{12} \ .$$

The statistic $\mathbf{S}_t^{MWMASHC}$ gives an upward shift from IC condition at instance t if $\mathbf{S}_t^{MWMASHC} > UCL$ and a downward shift from IC condition at instance t if $\mathbf{S}_t^{MWMASHC} < LCL$.

Theorem 27. *Suppose that assumptions 11-12 are satisfied. Under H_0, the two-sided control chart $\mathbf{S}_t^{MWMASHC}$ converges in distribution to*

$$\frac{\mathbf{S}_t^{MWMASHC} - \zeta}{\sigma} \xrightarrow{\mathcal{D}} \mathrm{N}(0,1) \ ,$$

and the one-sided control chart Λ_t converges in distribution to

$$\mathbf{\Lambda}_t = \frac{1}{\sigma^2}\left(\mathbf{S}_{0,t}^{MWMASHC} - \zeta\right)^2 \xrightarrow{\mathcal{D}} \chi_1^2 \ .$$

In practice, we need to work with the sample version of the above control charts, which are functions of empirical ranks. The following asymptotic strong consistency result holds for the sample version of the above control charts.

Theorem 28. *Assume that random objects X_i's are independent and identically distributed from probability measure P on manifold space $\mathcal{M}$. Under Assumptions 1-3, the empirical rank-depth control charts $\widehat{\mathbf{S}}_t$ and $\widehat{\mathbf{\Lambda}}_t$ based on empirical rank-depths $\widehat{R}_t$ are consistent estimators of control charts based on theoretical rank-depths R_t, i.e.*

$$\widehat{\mathbf{S}}_t(P_n) - \mathbf{S}_t(P) = O_p\left(n^{-1/2}(d\log d)^{1/2}\right) \ ,$$
$$\widehat{\mathbf{\Lambda}}_t(P_n) - \mathbf{\Lambda}_t(P) = O_p\left(n^{-1}(d\log d)\right) \ ,$$

where d is the intrinsic dimension of the manifold $\mathcal{M}$.

In practice, $\mathbf{S}_t^{EWMA}$ and $\mathbf{S}_t^{CUSUM}$ are used when there is a small persistent shift in the distribution of the Phase-II data, and $\mathbf{S}_t^{MWMASHC}$ is used when there is a large transient shift in the distribution of the Phase-II data. The above control charts are robust and nonparametric, as they do not assume any distribution for the collected object data.

Theorem 29. *The control charts r-EWMA, r-CUSUM, and r-MWMASHC are asymptotically distribution-free for any continuous distribution on a given manifold space $\mathcal{M}$.*

4.2.4 Robustness of Rank-Depth Control Charts

In this section, we investigate the robustness of the proposed rank-depth control charts using the concepts of q-robustness (Huber, 2011) and breakdown point for sequential process (He et al., 1990; Zhang, 2025).

Qualitative Robustness of Control Charts

We investigate the robustness properties of a sequence of rank-depth control chart statistics, denoted by $\mathbf{S}_n$, using the idea of equicontinuity of error probabilities, the total variation metric, bounded Lipschitz metric as well as the geometrical properties of the underlying manifold spaces.

Consider a random variable X on manifold $\mathcal{M}$ as a measurable map on a probability space $(\Omega, \mathcal{F}, P)$ to $(\mathcal{M}, \mathcal{S})$, where $\mathcal{S}$ denotes the Borel sigma-algebra of $\mathcal{M}$. For $X \sim P$, the distribution of $\mathbf{S}_n$ is denoted by $\mathcal{G}_P(\mathbf{S}_n)$. For two given probability measures P and Q defined on $\mathcal{S}$, a sequence of control chart $\mathbf{S}_n$ is q-robust at P if for any $\epsilon > 0$ there is $\delta > 0$ such that

$$\delta_{TV}(P, Q) \leq \delta \quad \text{implies} \quad \delta_{TV}(\mathcal{G}_P(\mathbf{S}_n), \mathcal{G}_Q(\mathbf{S}_n)) \leq \epsilon \ ,$$

for all sufficiently large n.

Let S be a functional (mapping) defined for all $P \in \mathcal{P}(\mathcal{M})$ such that $\widehat{\mathbf{S}}_n$ is the restriction of S to empirical measure P_n, i.e. $\widehat{\mathbf{S}}_n = S \mid P_n$. For functionals S defined on $\mathcal{P}(\mathcal{M})$, the idea of q-robustness can be connected to weak continuity (Huber, 2011), i.e. the continuity implied by the weak convergence of probability measures. Let P_n denote the empirical probability supported by points $x_1, \ldots, x_n$.

For the control chart $\widehat{\mathbf{S}}_n$ below

$$\widehat{\mathbf{S}}_n(x_1, \ldots, x_n) = S(P_n) \ ,$$

we have q-robustness as a result of weak continuity.

Assumption 13. *The representation of $\mathbf{S}_n$ by functional S is regular such that the following properties are satisfied (i) it is consistent for any P in the domain of functional S, i.e. $\widehat{\mathbf{S}}_n \xrightarrow{p} S(P)$, and (ii) for any $\tau \in S(P)$ there exists a sequence of empirical probability measures P_n converging to P such that $S(P_n)$ converges to τ.*

Following steps similar to those in Chapter 2, it is straightforward to show that the empirical and theoretical versions of the proposed control charts can be written in the form of functionals $S \mid P_n$ and $S \mid P$. As a result, it is easy to see that the regularity conditions given in Assumption 13 hold true.

Theorem 30. *Under the conditions of Assumption 13, the proposed rank-depth control charts $\mathbf{S}_n$ are q-robust.*

Breakdown Point of Control Charts

In this section, we extend the breakdown point analysis to the process monitoring application by using a definition which is appropriate for sequential/online analysis (Zhang, 2025). We formulate the process monitoring scheme for data contaminated by outliers modeled by Huber's gross error (Huber, 2011). Let P and Q represent IC and OC probability measures and define the following ϵ-contaminated probability measure classes

$$\mathcal{H}_{0,\epsilon} = \{(1-\epsilon)P + \epsilon G_0 \mid G_0 \in \mathcal{P}(\mathcal{M})\} \ ,$$
$$\mathcal{H}_{1,\epsilon} = \{(1-\epsilon)Q + \epsilon G_1 \mid G_1 \in \mathcal{P}(\mathcal{M})\} \ .$$

We assume that the IC data are generated from $\mathcal{H}_{0,\epsilon}$, and at some time point t_*, the distribution of the data changes to $\mathcal{H}_{1,\epsilon}$ due to some special causes of variation. An online process monitoring procedure can be represented by stopping time T_*, which is the time we trigger an alarm due to fault. Clearly, T_* is a random variable and decision $\{T_* = t_*\}$ depends only on the observations collected in the first t_* time steps.

Denote by $P^{(\infty)}_{H_0}$ and $\mathbb{E}^{(\infty)}_{H_0}$ the probability and expectation under no change and by $P^{(*)}_{H_0,H_1}$ and $\mathbb{E}^{(*)}_{H_0,H_1}$ the probability and expectation when change occurred at time t_* and data have distribution $\mathcal{H}_{1,\epsilon}$.

In (Zhang, 2025), the authors extended the concept of level breakdown point and power breakdown point in offline hypothesis testing to sequential process monitoring scheme. Here, we use similar definitions for analyzing the robustness of the proposed rank-depth control charts.

The false-alarm breakdown point ε_F is defined as

$$\varepsilon_F = \min\left\{0 \le \epsilon \le 1 \Big| \lim_{\mathbb{E}^{(\infty)}_P \to \infty} \inf_{H_0 \in \mathcal{H}_{0,\epsilon}} \frac{\mathbb{E}^{(\infty)}_{H_0}(T_*)}{\mathbb{E}^{(\infty)}_P(T_*)} = 0\right\} \ .$$

Define the worst-case detection delay

$$\mathrm{WADD}_{H_0,H_1} = \sup_{t_* \ge 1} \sup \mathbb{E}^{(*)}_{H_0,H_1}\left((T_* - t_* + 1)^+ \mid X_1, \ldots, X_{t_*-1}\right) \ ,$$

where $H_0 \in \mathcal{H}_{0,\epsilon}$ and $H_1 \in \mathcal{H}_{1,\epsilon}$.

The detection-delay breakdown point ε_D is defined as

$$\varepsilon_D = \min\left\{0 \le \epsilon \le 1 \Big| \lim_{\mathbb{E}^{(\infty)}_P \to \infty} \sup_{H_1 \in \mathcal{H}_{1,\epsilon}} \frac{\mathrm{WADD}_{P,H_1}(T_*)}{\mathrm{WADD}_{P,Q}(T_*)} = \infty\right\} \ .$$

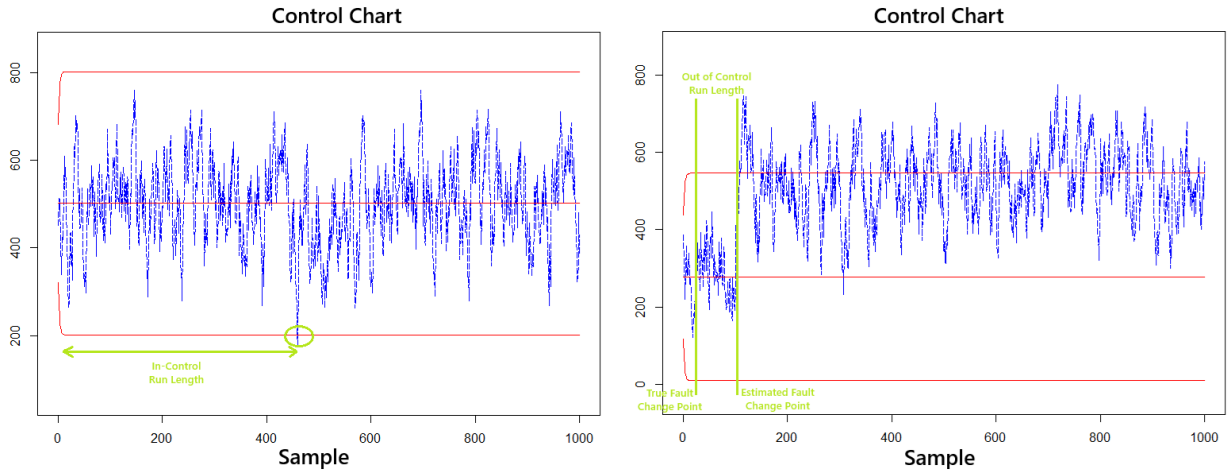


Figure 4.1: Average run length for IC and OC conditions.

For deriving the above results, we assumed that we have knowledge of the true distribution for $X_1, \ldots, X_n$, i.e. the statistics can be written as a function of true distribution $S(P)$. For the case that we use empirical distribution P_n, it can be shown that the asymptotic breakdown point for $S(P_n)$ is at-least as large as that of $S(P)$.

Theorem 31. *Suppose that the condition of Assumption 8 is satisfied. Let ε_1^* denote the asymptotic breakdown point for functional $S(P)$, and ε_2^* denote the asymptotic breakdown point for functional $S(P_n)$. It can be shown that $\varepsilon_2^* \geq \varepsilon_1^*$.*

4.2.5 Average Run Length

In this section, we introduce and discuss the concept of average run length (ARL) which we need for using the control charts and also for evaluating their performance. We also conduct theoretical analysis to investigate various properties of ARL for the proposed control charts. Once a proper lower control limit (LCL) and upper control limit (UCL) are obtained, the proposed control charts can be used for finite sample simulation.

Figure 4.1 depicts the average run lengths for in-control (IC) and out-of-control (OC) conditions, i.e. IC-ARL and OC-ARL. The control limit are selected such that the IC-ARL is kept at a nominal level, commonly used values are 200 and 370, while minimizing OC-ARL. Asymptotic results can also be considered for choosing control limits; however, it is possible that they do not achieve desirable performance in practice due to the fact

that usually a large sample size is needed to guarantee that the asymptotic control limits yield small OC-ARL (ARL_1) values while keeping IC-ARL (ARL_0) at a nominal level.

Definition ($\mathcal{K}$-homogeneous space). Consider the space $(\mathcal{M}, \delta)$, a metric space with manifold structure. For manifold sample spaces $(\mathcal{M}, \delta)$ equipped with homogeneity property, that is any two points (x, y) on the manifold $\mathcal{M}$ can be brought into coincidence by an isometry, we can define the transitivity of the left action of the isometry group. Define the group $(\mathcal{K}, \cdot)$ with a map $\alpha : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$, left action of $\mathcal{K}$ on $\mathcal{M}$. For all $k \in \mathcal{K}$, let $\alpha_k : \mathcal{M} \to \mathcal{M}$. The defined left action satisfies the following properties

$$\alpha_{k \cdot h} = \alpha_k \circ \alpha_h \ , \quad \forall h, k \in \mathcal{K} \ ,$$

$$\alpha_e = \text{identity map} \quad \mathcal{M} \to \mathcal{M} \ .$$

The action α_k is called transitive if for all $x_1, x_2 \in \mathcal{M}$, there exists a left action $k \in \mathcal{K}$ so that $\alpha_k(x_1) = x_2$. The resulting space is a $\mathcal{K}$-homogeneous space $(\mathcal{M}, \mathcal{K}, \alpha)$ (Osborne et al., 2013). If this space is equipped with metric δ, we have a homogeneous metric space.

Theorem 32. *For $\mathcal{K}$-homogeneous space $(\mathcal{M}, \mathcal{K}, \alpha)$, the control charts r-EWMA, r-CUSUM, and r-MWMASHC are congruence-invariant. As a result, the IC-ARL and OC-ARL are congruence-invariant. That is*

$$\begin{aligned} &\mathbb{E}_{F_X}\big[\min\{t \mid \mathbf{S}_t(X) > UCL \text{ or } \mathbf{S}_t(X) < LCL\}\big] \\ &\qquad = \mathbb{E}_{F_Y}\big[\min\{t \mid \mathbf{S}_t(Y) > UCL \text{ or } \mathbf{S}_t(Y) < LCL\}\big] \ , \end{aligned}$$

under the action $y = \alpha_k(x)$.

Theorem 33. *The IC-ARL of control charts r-EWMA, r-CUSUM, and r-MWMASHC follows geometric distribution with $ARL_0 = 1/\alpha$, where α is the false alarm rate. Therefore, for any distribution of X on a given manifold space $\mathcal{M}$, for random variable run length (RL), we have $P\{RL = n\} = \alpha(1-\alpha)^{n-1}$.*

In the current investigation, an algorithmic procedure is implemented which considers both Phase-I and Phase-II monitoring tasks. In the Phase-I monitoring, a control chart is defined and a number of IC data are used to estimate the median and the variance of the distribution, and to find lower and upper control limit (LCL and UCL) for monitoring task. Also, it is a common practice to do data cleaning by removing possible outliers from the dataset.

At the Phase-II monitoring stage, LCL and UCL values are used together with the considered chart for the online/sequential detection of OC products. For the current analysis,

r-EWMA, r-CUSUM and r-MWMASHC are considered for process monitoring. From the definition of ARL_0, it is apparent that $ARL_0 = 1/\alpha$, where false alarm rate α is equivalent to Type I error in hypothesis testing.

The algorithmic procedure used for the Phase-I monitoring is given below:

- Step 1: Clean the dataset by removing the extreme data points. Also, remove their corresponding depth values.
- Step 2: Use the IC samples $(X_1, X_2, ..., X_{n_1})$ from Step 1 for the computation of LCL and UCL values for each of the considered charts.
- Step 3: Record the data including their depth values $D(P_{n_1}, X_i)$ for $i = 1, 2, ..., n_1$, the computed median and variance, and the corresponding LCL and UCL values for each of the considered charts.

At the Phase-II stage, the IC distribution obtained from Phase-I analysis is used for the online monitoring of production.

This is done using the following procedure:

- Step 1: Extract the required geometrical information (median, variance, etc.) of the products produced at Phase-II.
- Step 2: Calculate the corresponding depth and rank of the Phase-II data with respect to the reference sample.
- Step 3: If the calculated value of control chart exceeds the corresponding LCL or UCL values, send an OC signal and stop the production process; otherwise, add the collected data to IC databse.
- Step 4: If an adaptive version of the statistics is used, the new incoming IC data can be used to update the LCL and UCL values.
- Step 5: Return to Step 1 and repeat the algorithm for another produced product.

Figure 4.2 depicts the flowchart of the steps we take at phase-I and phase-II stages of process monitoring, along with root cause diagnosis analysis.

In the next section, the details for the calculation of LCL and UCL values for the implemented charts are given, and simulation study is carried out to evaluate the performance of the proposed control charts.

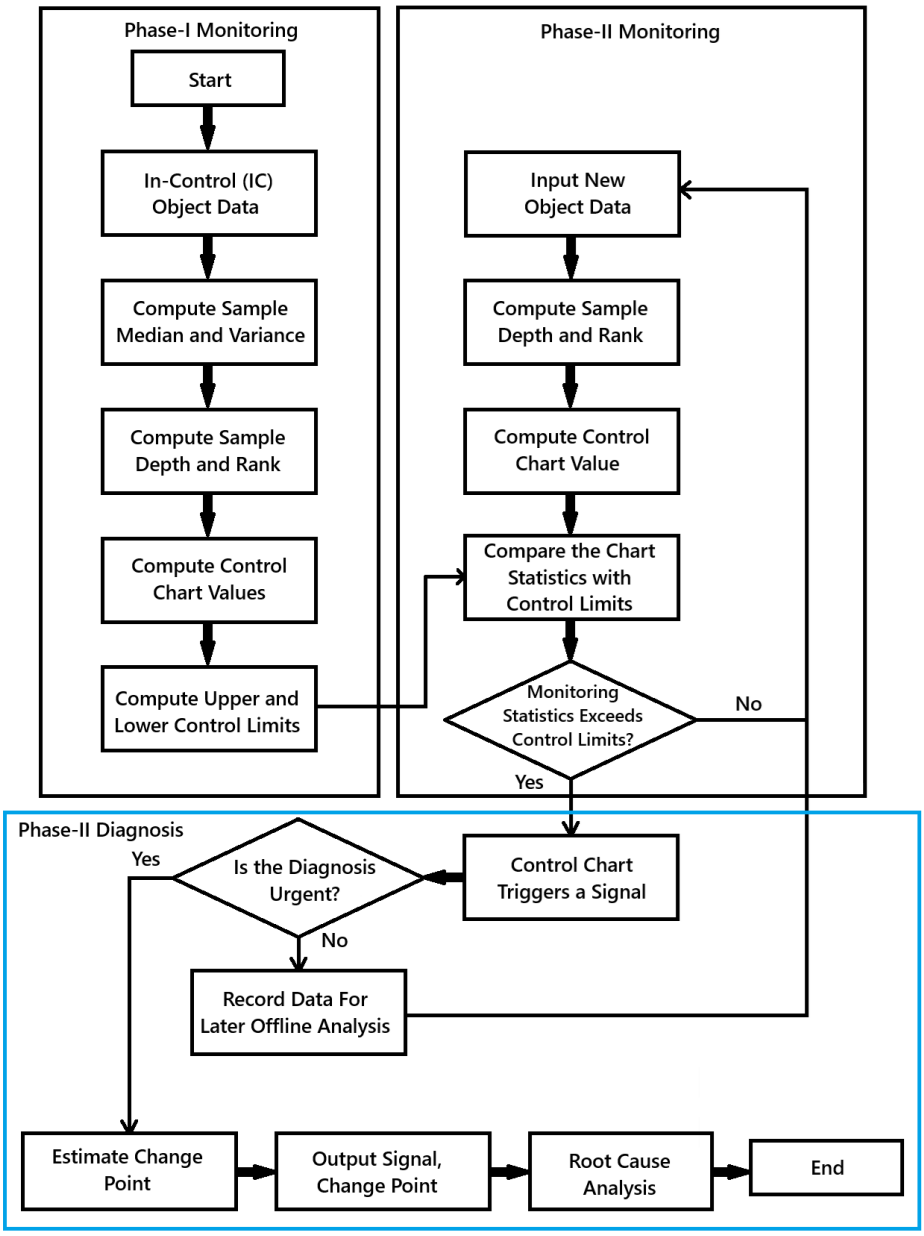


Figure 4.2: Flowchart of Phase-I and Phase-II stages of process monitoring.

4.3 Numerical Study

In this section, we carry out Monte Carlo simulation to evaluate the performance of the proposed rank-depth process monitoring methods considering various types of manifold spaces.

4.3.1 Numerical Simulation

In the first stage, we perform a numerical experiment to calculate the LCL and UCL for the proposed control charts. Given the fact that ranks are asymptotically uniformly distributed, we can generate large samples from uniform distribution to calculate these control limits (Li et al., 2013; Wang et al., 2023).

The control limits coefficients γ for given nominal IC-ARL values (ARL_0) are obtained using 5000 in-control samples, which is sufficiently large so that our asymptotic results approximately hold. All the the simulation results are based on 5000 replicates. The control coefficients for different control charts are obtained using bisectioning optimization method. The algorithmic steps are as follows:

- Step 1: On a grid of values for γ_0, calculate the corresponding IC-ARL by simulating 5000 in-control sample paths and averaging out the run lengths over 5000 replicates. Based on this, we extract the control coefficients whose IC-ARL values satisfy $ARL_0^{\gamma_1} < ARL_0$ and $ARL_0^{\gamma_2} > ARL_0$.
- Step 2: Find γ_3, the midpoint of γ_1 and γ_2, and calculate its corresponding $ARL_0^{\gamma_3}$.
- Step 3: If $ARL_0^{\gamma_3} < ARL_0$ assign $h_1 = h_3$. If $ARL_0^{\gamma_3} > ARL_0$ assign $h_2 = h_3$.
- Step 4: Repeat Steps 2 and 3 until $ARL_0^{\gamma_3}$ is sufficiently close to ARL_0.
- Step 5: Use γ_3 as the control limit.

Table 4.1 reports the control limit coefficients for r-EWMA, r-CUSUM, and r-SHC control charts for $ARL_0 = 200$ and $ARL_0 = 370$ for manifold intrinsic dimensions $2, 5, 10, 15, 20$ and 30. The reported control limit coefficients can be used as LCL and UCL for normalized values of control charts. For the general analysis, we can use these coefficient values as well as the average and std. of the control chart values calculated in Phase-I to calculate the updated LCL and UCL values, i.e. $f(\gamma_0, \mu, \sigma)$. From the reported results, it can be seen that the value of control limit coefficients are approximately insensitive to the

Table 4.1: Control limits coefficient γ_0 for r-EWMA, r-CUSUM, and r-MWMASHC control charts for $ARL_0 = 200$ and $ARL_0 = 370$ for different manifold intrinsic dimensions.

$ARL_0 = 200$						
	p = 2	**p = 5**	**p = 10**	**p = 15**	**p = 20**	**p = 30**
r-EWMA($\lambda = 0.05$)	2.223	2.254	2.243	2.271	2.292	2.304
r-EWMA($\lambda = 0.1$)	2.493	2.478	2.575	2.589	2.649	2.674
r-EWMA($\lambda = 0.2$)	2.607	2.643	2.662	2.697	2.718	2.734
r-CUSUM($k = 0.1$)	5.910	5.990	5.995	6.107	6.291	6.351
r-CUSUM($k = 0.2$)	2.807	2.884	3.047	3.075	3.153	3.223
r-CUSUM($k = 0.3$)	1.329	1.343	1.439	1.467	1.476	1.582
r-MWMASHC($m = 5$)	2.151	2.535	2.631	2.650	2.866	3.015
r-MWMASHC($m = 10$)	1.853	2.464	2.579	2.623	2.686	2.856
r-MWMASHC($m = 15$)	1.863	2.437	2.516	2.602	2.642	2.774
$ARL_0 = 370$						
	p = 2	**p = 5**	**p = 10**	**p = 15**	**p = 20**	**p = 30**
r-EWMA($\lambda = 0.05$)	2.528	2.546	2.646	2.676	2.688	2.701
r-EWMA($\lambda = 0.1$)	2.705	2.819	2.875	2.882	2.884	2.903
r-EWMA($\lambda = 0.2$)	2.757	2.868	2.886	2.921	2.922	2.931
r-CUSUM($k = 0.1$)	6.895	6.901	6.962	7.086	7.149	7.183
r-CUSUM($k = 0.2$)	3.419	3.649	3.986	4.027	4.125	4.217
r-CUSUM($k = 0.3$)	1.629	1.734	1.829	1.860	1.880	2.136
r-MWMASHC($m = 5$)	3.087	3.105	3.229	3.296	3.352	3.364
r-MWMASHC($m = 10$)	2.842	3.042	3.101	3.163	3.264	3.331
r-MWMASHC($m = 15$)	2.808	3.029	3.039	3.063	3.187	3.236

intrinsic dimension of the manifold space, and vary as a function of the parameter values of the control charts, i.e. λ, k, and m as well as the considered nominal IC-ARL values. Through an extensive simulation study, it was demonstrated that the reported values have a good generalization power and can be effectively applied to various types of manifolds with specific geometrical properties.

For comparison we adopt well-known nonparametric process monitoring methods from the literature which are proposed for non-Euclidean data. The interpoint distance test (IPD) (Shu and Fan, 2018) is a nonparametric control chart based on Minkowski distance which can be applied to manifold data as long as a proper distance for the underlying space is defined. In Saha et al. (2017), a quadratic functional was proposed which is applied to the distance to calculate the value of IPD control chart. In Shu and Fan (2018), three variants

of functionals were used which increase slowly as distance increases. These functionals are

$$\begin{aligned}\phi_1(x) &= \log(1+x)\ ,\\ \phi_2(x) &= x^2/(1+x^2)\ ,\\ \phi_2(x) &= 1-e^{-x}\ .\end{aligned}$$

It was observed that using these functional makes the performance of IPD robust against outliers. Following the findings of (Shu and Fan, 2018), we use the above functionals to calculate the values of IPD test statistic. In Liu et al. (2004), the authors proposed a data depth based moving average (DDMA) control chart, which was proven to be powerful for the detection of changes in location and scale. Following the promising report, we extend the DDMA to manifold data with geodesic depth functions. In Chakraborty and Finkelstein (2024), the authors proposed rank-energy (RE) control chart which is based on measure transportation. It was demonstrated that RE is robust and can outperform other nonparametric methods which are not based on rank. We also consider interpoint distance based high-dimensional Shewhart control chart based on distance from origin (HDSOR) and distance from reference data point (HDSIP) (Mukherjee and Marozzi, 2021).

For the numerical simulation, we consider Wishart distribution, von Mises-Fisher distribution as well as uniform and Gaussian distributions on Steifel manifold.

Example 1 (Wishart Distributions): In this example, we perform a comparative study to evaluate the performance of the perposed rank-depth process monitoring algorithm on manifold data generated from Wishart distributions with intrinsic dimensions $3, 10, 21$, corresponding to matrices of size 2×2, 4×4 and 6×6. The evaluate the robustness of the rank-depth control charts, we also contaminate these object data with outliers, with contamination level of 50%.

Table 4.2 reports the results of simulation study for in-control scenario. It can be seen that the proposed rank-depth control charts outperform the other nonparametric methods for most of the case studies. Besides, their superiority becomes more apparent as the intrinsic dimension of the manifold space is increased. For most of the cases, for the both choices of ARL_0, 200 and 370, the obtained IC-ARL stays above the nominal value. It is also observed that the performance of these methods are not affected by the contamination of the original data. Table 4.3 reports the results of simulation study for out-of-control scenario. For the OC performance, we considered effect sizes $\delta = 2, 5, 10$. It is clear that rank-depth control charts surpass the other nonparametric methods for most of the simulation scenarios and have a small OC-ARL values. This indicates the effectiveness of using rank-depth for monitoring manifold data given the fact that the IC-ARL is kept at the nominal value and OC-ARL is small. It is also observed that contaminating the original

Table 4.2: IC-ARL values for r-EWMA, r-CUSUM, and r-MWMASHC for Wishart distribution for 5000 simulation runs. The values in the parentheses are standard deviations.

	Standard Wishart			**Contaminated Wishart**		
			$ARL_0 = 200$			
	p = 3	**p = 10**	**p = 21**	**p = 3**	**p = 10**	**p = 21**
r-EWMA($\lambda = 0.05$)	206.5 (187.7)	174.2 (167.1)	203.1 (186.7)	208.4 (192.4)	204.5 (186.5)	209.5 (203.9)
r-EWMA($\lambda = 0.1$)	204.9 (189.3)	212.7 (195.7)	213.2 (196.6)	229.2 (204.5)	192.3 (201.5)	241.5 (240.6)
r-EWMA($\lambda = 0.2$)	208.1 (190.1)	215.8 (195.7)	211.4 (190.4)	196.6 (208.3)	234.3 (222.3)	228.4 (208.3)
r-CUSUM($k = 0.1$)	267.4 (208.1)	210.9 (209.8)	228.6 (227.7)	240.1 (191.7)	214.8 (192.1)	237.1 (216.4)
r-CUSUM($k = 0.2$)	237.1 (187.5)	212.3 (227.4)	195.4 (248.1)	196.1 (208.7)	209.1 (220.0)	223.7 (219.1)
r-CUSUM($k = 0.3$)	251.1 (212.7)	253.1 (225.2)	211.9 (232.2)	235.9 (166.6)	230.4 (203.9)	215.1 (239.0)
r-MWMASHC($m = 5$)	165.2 (52.2)	240.6 (196.1)	258.4 (242.8)	213.2 (109.4)	238.0 (139.8)	**273.9 (263.1)**
r-MWMASHC($m = 10$)	227.4 (114.7)	210.1 (143.4)	222.3 (237.8)	204.9 (117.3)	**259.9 (213.8)**	231.0 (217.6)
r-MWMASHC($m = 15$)	204.5 (212.4)	**267.4 (218.9)**	**275.2 (226.4)**	**241.4 (155.1)**	255.3 (207.2)	193.3 (229.8)
IPD (ϕ_1)	218.9 (254.8)	193.1 (249.7)	244.2 (288.5)	128.3 (116.2)	118.7 (107.5)	129.4 (105.3)
IPD (ϕ_2)	206.2 (185.5)	191.3 (179.1)	236.9 (215.2)	84.4 (87.2)	130.8 (127.9)	142.7 (127.8)
IPD (ϕ_3)	187.1 (156.4)	193.9 (201.6)	213.3 (184.2)	164.3 (181.2)	91.4 (99.3)	89.6 (94.3)
DDMA ($m = 5$)	188.7 (173.2)	196.3 (201.1)	203.8 (180.8)	157.5 (163.3)	179.1 (182.7)	184.5 (174.7)
DDMA ($m = 10$)	189.6 (206.4)	204.5 (199.7)	188.1 (161.7)	159.1 (159.5)	209.7 (207.3)	179.4 (155.4)
DDMA ($m = 15$)	207.6 (201.7)	208.9 (176.6)	225.2 (232.1)	164.5 (162.4)	212.6 (203.2)	172.9 (163.7)
RE	187.6 (177.8)	150.1 (90.1)	215.3 (211.4)	123.6 (178.0)	229.8 (164.1)	153.5 (214.1)
HDSOR	223.8 (225.1)	259.4 (227.9)	229.9 (198.1)	174.5 (176.7)	129.1 (123.2)	126.9 (125.7)
HDSIP	**285.7 (259.1)**	206.9 (241.5)	176.3 (198.9)	134.1 (122.7)	170.7 (227.1)	156.1 (116.1)
			$ARL_0 = 370$			
	p = 3	**p = 10**	**p = 21**	**p = 3**	**p = 10**	**p = 21**
r-EWMA($\lambda = 0.05$)	331.7 (220.1)	377.7 (229.5)	366.2 (244.1)	316.2 (200.6)	376.4 (230.1)	386.1 (244.3)
r-EWMA($\lambda = 0.1$)	342.1 (237.6)	404.7 (269.4)	403.2 (255.3)	342.1 (244.8)	392.1 (262.1)	380.5 (242.3)
r-EWMA($\lambda = 0.2$)	359.1 (239.4)	364.3 (249.7)	391.1 (260.4)	331.6 (225.8)	397.3 (283.4)	414.3 (276.1)
r-CUSUM($k = 0.1$)	386.1 (238.2)	356.7 (242.7)	355.3 (230.9)	369.6 (220.4)	389.1 (223.6)	325.4 (203.7)
r-CUSUM($k = 0.2$)	**447.7 (285.2)**	**476.7 (268.4)**	**443.4 (268.1)**	377.0 (220.3)	393.4 (290.6)	364.0 (288.5)
r-CUSUM($k = 0.3$)	385.3 (254.3)	379.7 (254.7)	424.6 (259.9)	384.9 (280.6)	418.1 (277.3)	349.8 (279.9)
r-MWMASHC($m = 5$)	414.4 (229.7)	447.1 (247.5)	383.5 (265.9)	406.6 (254.6)	**452.2 (258.2)**	397.5 (99.3)
r-MWMASHC($m = 10$)	404.7 (247.6)	391.6 (216.2)	394.8 (274.7)	368.5 (259.9)	375.7 (234.4)	400.9 (244.0)
r-MWMASHC($m = 15$)	424.7 (257.6)	421.9 (207.6)	411.9 (256.5)	**418.5 (218.6)**	451.7 (242.9)	**434.1 (295.1)**
IPD (ϕ_1)	349.6 (252.1)	370.1 (290.1)	360.4 (254.7)	246.8 (206.1)	241.1 (233.4)	203.7 (189.8)
IPD (ϕ_2)	411.1 (277.2)	344.6 (250.1)	424.3 (167.8)	205.5 (143.3)	362.7 (239.7)	327.5 (263.3)
IPD (ϕ_3)	428.5 (330.7)	382.3 (257.7)	349.1 (233.4)	373.9 (251.2)	377.2 (263.2)	212.2 (216.0)
DDMA ($m = 5$)	349.1 (251.7)	355.9 (265.1)	341.8 (253.6)	380.8 (279.1)	333.9 (295.7)	300.5 (201.1)
DDMA ($m = 10$)	410.7 (288.3)	412.1 (214.3)	392.6 (297.7)	313.2 (290.2)	332.9 (218.2)	377.6 (220.1)
DDMA ($m = 15$)	401.2 (287.7)	385.2 (297.3)	381.1 (292.7)	327.8 (203.2)	316.7 (206.5)	346.2 (219.1)
RE	312.8 (302.3)	317.7 (91.9)	396.6 (210.9)	265.1 (181.6)	317.7 (192.9)	304.3 (131.3)
HDSOR	336.9 (255.9)	372.1 (262.6)	348.1 (245.0)	298.3 (252.0)	260.8 (238.6)	230.4 (230.9)
HDSIP	326.8 (208.1)	337.7 (215.9)	353.9 (268.4)	277.9 (287.9)	307.5 (159.6)	306.1 (216.1)

Table 4.3: OC-ARL values for r-EWMA, r-CUSUM, and r-MWMASHC for Wishart distribution for 5000 simulation runs. The values in the parentheses are standard deviations.

		Standard Wishart			Contaminated Wishart		
		p = 3	**p = 10**	**p = 21**	**p = 3**	**p = 10**	**p = 21**
r-EWMA($\lambda = 0.05$)	$\delta = 2$	31.76 (86.54)	20.56 (65.31)	8.64 (8.47)	59.15 (122.23)	32.06 (100.54)	15.42 (37.08)
	$\delta = 5$	**1.61 (5.18)**	1.97 (3.91)	1.65 (1.15)	4.72 (4.53)	2.72 (5.10)	**1.25 (6.28)**
	$\delta = 10$	**1.04 (3.61)**	1.76 (2.28)	1.75 (2.28)	**1.08 (2.09)**	1.24 (4.13)	1.60 (6.39)
r-EWMA($\lambda = 0.1$)	$\delta = 2$	17.58 (73.12)	30.32 (77.06)	9.89 (8.44)	61.77 (134.33)	31.90 (76.70)	15.31 (18.36)
	$\delta = 5$	5.81 (6.27)	1.75 (1.14)	1.73 (3.43)	3.28 (5.88)	2.55 (4.15)	2.23 (4.22)
	$\delta = 10$	1.67 (2.54)	**1.03 (0.17)**	**1.02 (0.14)**	1.41 (3.64)	**1.08 (4.31)**	1.55 (3.33)
r-EWMA($\lambda = 0.2$)	$\delta = 2$	62.31 (109.15)	26.67 (56.83)	10.75 (13.39)	83.46 (164.91)	44.62 (102.47)	16.02 (20.033)
	$\delta = 5$	2.18 (2.80)	**1.31 (4.69)**	**1.35 (4.26)**	**1.58 (5.37)**	**1.49 (4.48)**	2.00 (3.23)
	$\delta = 10$	1.78 (1.92)	1.23 (1.43)	1.03 (0.17)	1.47 (3.32)	1.33 (3.44)	1.16 (4.14)
r-CUSUM($k = 0.1$)	$\delta = 2$	24.69 (56.38)	11.12 (30.11)	**2.27 (4.16)**	24.32 (40.64)	11.77 (9.67)	4.36 (6.66)
	$\delta = 5$	5.61 (12.35)	5.44 (4.15)	5.27 (8.89)	6.71 (13.06)	5.80 (12.84)	5.71 (12.41)
	$\delta = 10$	4.16 (11.56)	4.57 (7.62)	5.46 (5.67)	4.73 (3.20)	4.09 (3.09)	5.80 (9.90)
r-CUSUM($k = 0.2$)	$\delta = 2$	26.75 (52.96)	22.56 (46.73)	5.40 (10.08)	25.61 (57.21)	13.92 (31.20)	**1.28 (4.94)**
	$\delta = 5$	4.25 (5.50)	4.90 (5.50)	4.52 (5.31)	4.37 (8.54)	4.01 (5.50)	4.14 (4.35)
	$\delta = 10$	3.02 (8.85)	3.26 (10.20)	3.43 (8.88)	3.56 (4.49)	3.55 (8.27)	4.34 (11.56)
r-CUSUM($k = 0.3$)	$\delta = 2$	23.24 (45.90)	22.19 (21.93)	13.01 (17.54)	70.61 (131.75)	35.07 (58.92)	7.17 (13.44)
	$\delta = 5$	3.49 (5.95)	3.47 (11.51)	3.92 (4.07)	3.10 (1.00)	3.84 (10.88)	3.30 (11.57)
	$\delta = 10$	2.65 (10.15)	3.40 (7.06)	3.05 (5.61)	2.27 (9.10)	3.13 (9.73)	3.67 (3.07)
r-MWMASHC($m = 5$)	$\delta = 2$	50.60 (86.85)	21.27 (24.12)	9.73 (16.47)	32.51 (57.26)	27.20 (26.02)	16.18 (15.63)
	$\delta = 10$	4.59 (1.40)	4.05 (1.02)	4.04 (0.82)	4.98 (2.56)	4.49 (0.89)	3.94 (0.96)
	$\delta = 15$	3.73 (0.82)	3.81 (0.80)	4.00 (0.82)	3.88 (0.96)	3.73 (0.89)	3.84 (0.77)
r-MWMASHC($m = 10$)	$\delta = 2$	34.75 (54.22)	18.88 (16.21)	11.61 (5.96)	23.90 (27.17)	16.26 (24.57)	10.42 (5.45)
	$\delta = 5$	6.22 (1.63)	5.68 (1.32)	5.83 (1.27)	6.55 (1.63)	6.06 (1.48)	5.74 (1.33)
	$\delta = 10$	5.53 (1.23)	5.17 (1.28)	5.69 (1.38)	5.32 (1.36)	5.35 (1.39)	5.54 (1.42)
r-MWMASHC($m = 15$)	$\delta = 2$	29.03 (37.04)	16.06 (16.16)	6.89 (1.92)	27.75 (37.83)	19.83 (14.05)	15.37 (9.46)
	$\delta = 5$	7.95 (2.27)	7.16 (1.75)	7.16 (1.65)	8.69 (2.07)	7.64 (1.83)	7.39 (1.71)
	$\delta = 10$	6.32 (1.55)	6.39 (1.65)	6.71 (1.62)	6.78 (1.72)	6.89 (1.69)	7.93 (1.68)
IPD (ϕ_1)	$\delta = 2$	**8.50 (4.03)**	8.41 (4.13)	4.95 (2.47)	8.16 (4.80)	7.45 (3.13)	5.18 (2.37)
	$\delta = 5$	2.67 (1.47)	2.66 (1.09)	1.66 (0.86)	2.91 (1.40)	2.63 (1.33)	1.63 (0.84)
	$\delta = 10$	1.36 (0.55)	1.06 (0.23)	1.05 (0.66)	1.47 (0.74)	1.16 (0.39)	**1.12 (0.38)**
IPD (ϕ_2)	$\delta = 2$	32.73 (58.16)	39.66 (51.16)	30.77 (28.14)	59.24 (84.83)	53.52 (58.51)	32.44 (32.12)
	$\delta = 5$	46.23 (54.23)	36.23 (24.62)	13.03 (3.74)	25.27 (15.30)	24.78 (15.15)	13.22 (3.78)
	$\delta = 10$	23.05 (22.72)	20.21 (9.93)	10.23 (0.50)	16.81 (7.12)	15.45 (6.01)	11.42 (2.57)
IPD (ϕ_3)	$\delta = 2$	72.14 (55.08)	60.78 (50.72)	26.95 (20.98)	79.87 (78.94)	64.31 (54.93)	32.79 (23.78)
	$\delta = 5$	54.81 (59.64)	34.56 (23.85)	12.36 (3.49)	51.69 (57.20)	23.35 (14.01)	11.96 (3.35)
	$\delta = 10$	42.35 (47.29)	19.09 (8.77)	10.04 (0.56)	36.90 (37.85)	15.35 (6.27)	10.26 (1.83)
DDMA ($m = 5$)	$\delta = 2$	91.58 (103.06)	59.75 (115.88)	16.42 (21.07)	83.34 (86.36)	**2.63 (1.70)**	12.71 (15.11)
	$\delta = 5$	5.12 (3.78)	3.45 (1.68)	2.90 (1.02)	5.93 (4.85)	3.81 (2.79)	2.92 (1.06)
	$\delta = 10$	2.53 (1.68)	2.55 (1.67)	2.49 (1.70)	2.74 (1.79)	2.77 (1.66)	2.46 (1.73)
DDMA ($m = 10$)	$\delta = 2$	73.22 (73.89)	38.20 (59.26)	13.20 (15.77)	72.87 (77.23)	28.58 (33.92)	10.97 (10.55)
	$\delta = 5$	5.66 (2.62)	4.41 (1.74)	4.01 (1.39)	6.30 (3.63)	4.91 (1.75)	4.10 (1.48)
	$\delta = 10$	3.65 (1.25)	3.78 (1.10)	3.91 (1.09)	3.95 (1.41)	4.26 (1.13)	3.83 (1.23)
DDMA ($m = 15$)	$\delta = 2$	86.55 (94.14)	27.33 (40.06)	12.97 (9.60)	64.17 (76.35)	32.03 (33.63)	15.04 (12.26)
	$\delta = 5$	6.38 (3.29)	5.87 (2.04)	5.38 (2.12)	7.35 (4.03)	6.41 (2.48)	5.43 (1.97)
	$\delta = 10$	4.97 (2.01)	5.44 (1.85)	5.24 (1.91)	5.28 (1.92)	5.02 (1.99)	5.55 (2.05)
RE	$\delta = 2$	46.50 (34.67)	65.79 (91,41)	37.72 (39.42)	**7.34 (7.07)**	16.73 (5.16)	5.46 (8.89)
	$\delta = 5$	1.94 (1.52)	1.34 (1.82)	1.89 (1.92)	1.45 (1.32)	1.52 (1.02)	1.31 (2.23)
	$\delta = 10$	1.14 (1.25)	1.74 (1.72)	1.84 (1.63)	1.53 (1.35)	1.73 (1.13)	1.24 (2.53)
HDSOR	$\delta = 2$	13.50 (9.55)	**7.70 (2.80)**	6.91 (2.11)	11.18 (6.28)	7.57 (2.40)	6.78 (1.71)
	$\delta = 5$	5.18 (1.15)	4.96 (0.96)	5.04 (0.99)	5.18 (1.32)	5.21 (1.05)	4.90 (1.21)
	$\delta = 10$	4.92 (0.99)	4.87 (0.98)	5.01 (0.86)	5.16 (1.20)	5.08 (1.03)	5.01 (0.93)
HDSIP	$\delta = 2$	38.07 (30.33)	57.20 (25.75)	38.89 (51.41)	34.87 (46.96)	31.38 (14.18)	24.30 (20.98)
	$\delta = 5$	9.26 (9.01)	7.96 (11.25)	4.22 (5.90)	29.07 (29.61)	15.86 (18.17)	7.10 (3.97)
	$\delta = 10$	7.61 (7.94)	2.30 (2.49)	2.23 (2.47)	8.25 (5.92)	10.07 (10.97)	6.90 (11.08)

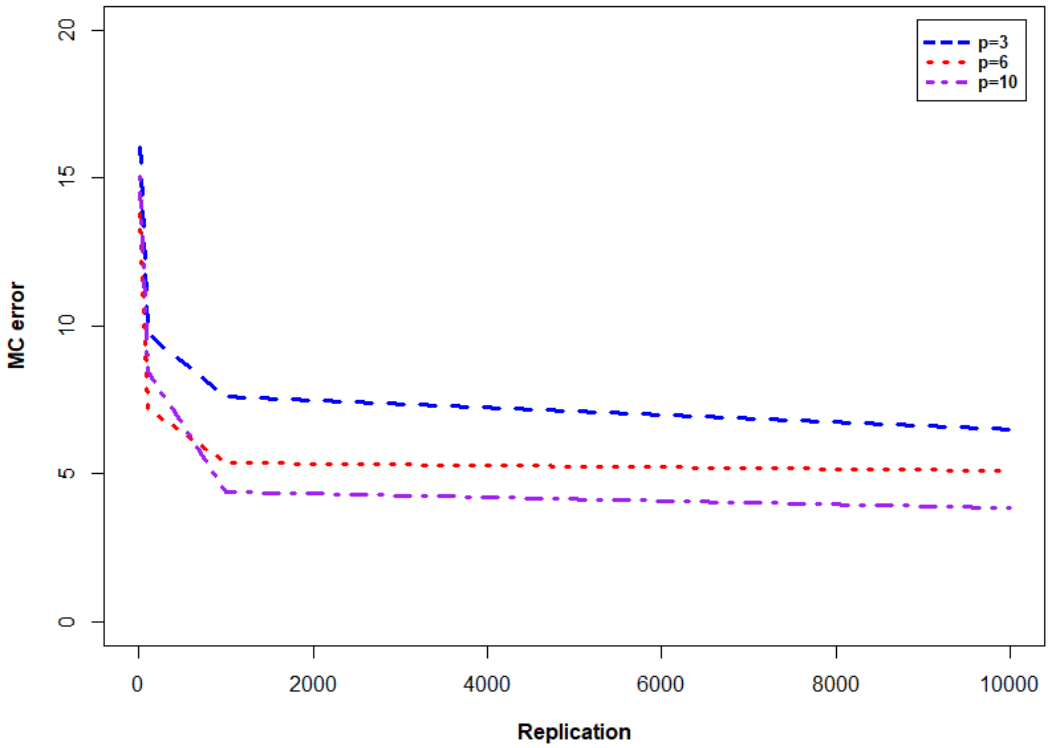


Figure 4.3: MC error for OC-ARL with r-EWMA and replication numbers $(1, 10000)$ for Wishart distribution with dimensions $p = (3, 6, 10)$.

data does not affect the performance of the proposed rank-depth control charts. It is also observed that the performance of IPD_{ϕ_1}, DDMA, HDSOR and RE are comparable to the rank-depth control charts. However, they are inferior for small effect sizes, e.g. $\delta = 2$, especially when dimension is large, i.e. $p = 21$. For some of the simulation scenarios, their performances are significantly affected by the contamination of the original data.

Figure 4.3 indicates the MC error of r-EWMA for OC-ARL scenario using Wishart distributionn with dimensions $p = 3, 6, 10$. It can be seen that that MC error decreases exponentially fast as the number of replications gets larger. The carried out experiments indicated that the MC error becomes stable for replications greater than 2000 regardless of the dimensionality of the manifold space. Figure 4.4 indicates the performance of the considered nonparametric methods for the effect sizes in the range $(0, 10)$. Top panels are the log(ARL) values for original data with dimensions $3, 10, 21$ and bottom panels are the log(ARL) values for contaminated data with dimensions $3, 10, 21$. It is clear that the rank-

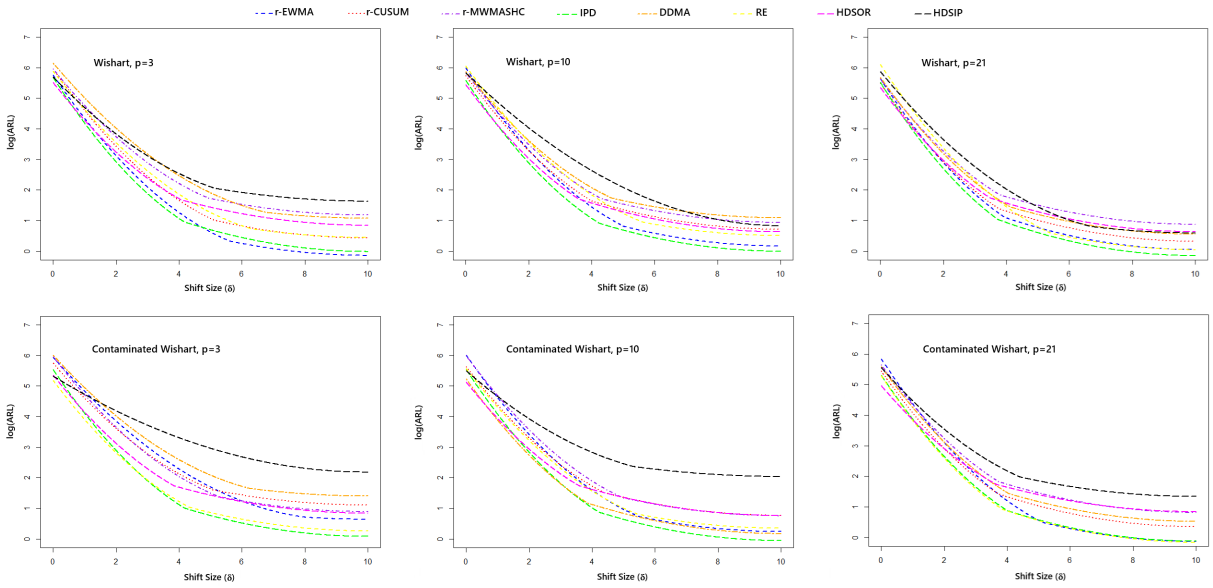


Figure 4.4: OC-ARL values for the process monitoring control charts for Wishart and contaminated Wishart with shift sizes $(0, 10)$ and dimensions $(3, 11, 21)$.

depth methods keep a very good balance at various effect size values and are superior or comparable to the other nonparametric control charts. It is observed that the performance of IPD is also very close to the rank-depth methods while HDSIP has a large OC-ARL value for most of the cases. It can also be realized that the log(ARL) value for all of these methods decreases exponentially fast as the effect size increases. For most of the cases the fastest decrease belongs to either r-EWMA or IPD while the performance of r-CUSUM and r-MWMASHC remains comparable to these two control charts. These results demonstrate the effectiveness of the rank-depth methods as they keep a good balance between IC-ARL and OC-ARL regardless of the contamination of the original data and the dimensionality of the manifold.

Example 2 (von Mises-Fisher Distributions): In this example, we evaluate the performance of the rank-depth process monitoring methods on data generated from von Mises-Fisher distribution with intrinsic dimensions $3, 10, 20$. The robustness of the methods are also evaluated by contaminating the data with outliers, with contamination level 50%.

Table 4.4 reports the obtained results for IC scenario. As it can be seen, r-EWMA and r-CUSUM outperform the rival methods for most of the cases. Having said that the performance of IPD is also comparable. For the both choices of ARL_0, 200 and 370, the

Table 4.4: IC-ARL values for r-EWMA, r-CUSUM, and r-MWMASHC for von Mises-Fisher distribution for 5000 simulation runs. The values in the parentheses are standard deviations.

	Standard VMF			**Contaminated VMF**		
			$ARL_0 = 200$			
	p = 3	**p = 10**	**p = 20**	**p = 3**	**p = 10**	**p = 20**
r-EWMA($\lambda = 0.05$)	215.6 (187.4)	183.8 (76.9)	191.1 (96.7)	208.9 (179.5)	**237.3 (183.6)**	223.2 (177.6)
r-EWMA($\lambda = 0.1$)	206.6 (176.1)	174.8 (74.2)	183.5 (103.6)	249.1 (211.1)	178.7 (80.1)	206.6 (103.2)
r-EWMA($\lambda = 0.2$)	236.8 (237.1)	196.2 (91.3)	210.5 (125.7)	**259.1 (216.6)**	171.1 (73.6)	182.1 (103.3)
r-CUSUM($k = 0.1$)	226.2 (177.9)	243.9 (223.1)	209.2 (163.3)	220.1 (164.7)	204.5 (231.1)	233.4 (195.1)
r-CUSUM($k = 0.2$)	229.2 (175.4)	**261.8 (243.1)**	**253.1 (247.9)**	199.2 (163.3)	208.1 (158.0)	233.9 (202.2)
r-CUSUM($k = 0.3$)	209.4 (168.4)	211.5 (194.7)	217.5 (218.2)	221.6 (162.6)	210.1 (203.1)	216.5 (168.6)
r-MWMASHC($m = 5$)	191.4 (162.9)	153.3 (95.6)	187.7 (141.7)	187.6 (171.2)	180.6 (152.8)	174.4 (105.2)
r-MWMASHC($m = 10$)	223.5 (215.4)	180.7 (148.3)	248.3 (228.2)	232.4 (242.1)	145.8 (81.4)	205.8 (129.9)
r-MWMASHC($m = 15$)	178.1 (168.7)	197.1 (185.5)	190.3 (189.2)	255.7 (261.8)	228.7 (212.6)	198.9 (170.1)
IPD (ϕ_1)	198.5 (145.0)	213.3 (198.3)	196.5 (167.7)	175.8 (170.61)	199.2 (203.6)	210.7 (184.1)
IPD (ϕ_2)	216.8 (213.5)	194.2 (258.5)	209.1 (221.1)	205.1 (252.5)	156.3 (168.7)	124.1 (133.8)
IPD (ϕ_3)	**244.5 (284.4)**	193.1 (233.2)	235.6 (236.6)	239.5 (241.3)	179.5 (199.4)	180.6 (177.6)
DDMA ($m = 5$)	206.7 (201.1)	240.4 (180.7)	216.8 (232.7)	213.9 (199.9)	162.8 (156.9)	226.2 (169.5)
DDMA ($m = 10$)	195.8 (190.3)	231.4 (170.1)	226.3 (185.5)	165.1 (193.5)	177.1 (173.1)	174.6 (228.1)
DDMA ($m = 15$)	207.3 (218.1)	192.5 (157.2)	212.5 (201.9)	179.5 (152.9)	176.4 (185.4)	160.5 (171.3)
RE	206.8 (153.9)	245.1 (183.1)	170.1 (71.1)	184.4 (173.7)	227.3 (149.1)	**236.6 (185.4)**
HDSOR	201.3 (207.2)	229.3 (198.3)	220.1 (243.8)	180.3 (206.9)	132.3 (114.3)	185.9 (121.7)
HDSIP	178.7 (201.4)	195.1 (177.3)	194.1 (171.4)	132.1 (109.5)	146.1 (113.1)	138.4 (101.4)
			$ARL_0 = 370$			
	p = 3	**p = 10**	**p = 20**	**p = 3**	**p = 10**	**p = 20**
r-EWMA($\lambda = 0.05$)	378.2 (237.3)	340.4 (243.4)	354.4 (144.8)	361.8 (225.1)	301.3 (98.3)	354.1 (153.2)
r-EWMA($\lambda = 0.1$)	348.5 (242.1)	314.1 (298.3)	334.6 (159.8)	**430.7 (284.8)**	409.6 (248.4)	343.6 (161.8)
r-EWMA($\lambda = 0.2$)	394.5 (249.5)	330.2 (193.5)	355.6 (154.5)	397.1 (266.2)	312.1 (92.1)	362.8 (171.78)
r-CUSUM($k = 0.1$)	368.5 (234.3)	334.6 (216.0)	340.2 (222.8)	324.7 (249.8)	330.5 (236.4)	395.3 (237.4)
r-CUSUM($k = 0.2$)	350.4 (244.7)	390.9 (268.7)	374.7 (264.9)	361.5 (248.0)	405.4 (243.9)	351.5 (254.4)
r-CUSUM($k = 0.3$)	**416.6 (257.2)**	401.1 (264.9)	372.3 (282.1)	376.1 (256.9)	415.6 (264.6)	**416.0 (268.4)**
r-MWMASHCC($m = 5$)	401.9 (288.5)	233.4 (175.3)	336.1 (215.1)	406.7 (272.4)	270.1 (216.1)	298.6 (175.4)
r-MWMASHC($m = 10$)	354.6 (257.8)	322.8 (231.2)	368.7 (279.4)	351.6 (224.1)	357.9 (277.1)	371.5 (265.1)
r-MWMASHC($m = 15$)	368.3 (273.3)	340.4 (225.5)	343.7 (239.7)	386.1 (269.4)	360.4 (245.1)	415.6 (273.5)
IPD (ϕ_1)	392.6 (271.6)	413.9 (268.7)	370.8 (263.8)	422.9 (315.1)	323.6 (266.6)	351.7 (244.4)
IPD (ϕ_2)	373.7 (290.3)	368.5 (280.4)	**404.0 (312.6)**	385.7 (300.3)	320.3 (264.4)	331.9 (319.2)
IPD (ϕ_3)	415.2 (266.1)	**425.6 (217.2)**	342.5 (284.1)	321.3 (294.2)	337.6 (229.3)	331.1 (253.1)
DDMA ($m = 5$)	391.1 (184.3)	416.7 (281.2)	370.8 (246.7)	337.7 (191.6)	331.5 (203.1)	268.2 (175.4)
DDMA ($m = 10$)	360.3 (221.2)	347.4 (138.9)	347.3 (243.2)	278.5 (206.8)	335.6 (201.6)	349.7 (201.8)
DDMA ($m = 15$)	403.3 (233.5)	357.8 (185.1)	394.5 (201.2)	322.2 (195.4)	343.2 (215.1)	311.9 (189.0)
RE	380.6 (245.9)	349.3 (111.8)	329.5 (203.8)	301.2 (191.3)	**416.6 (236.3)**	284.3 (107.6)
HDSOR	376.4 (231.0)	333.7 (266.8)	322.6 (241.2)	309.8 (236.1)	373.8 (261.3)	254.5 (250.4)
HDSIP	379.1 (209.3)	371.7 (250.8)	361.1 (234.1)	371.1 (237.1)	304.0 (158.4)	347.6 (214.3)

IC-ARL values remain stably above the specified threshold, and are not affected by the dimension of the data. It is also observed that contaminating the original data undermines the nonparametric methods which are based on distances, i.e. IPD, RE, HDSIP. However, r-EWMA, r-CUSUM, and r-MWMASHC are not affected by contaminating the dataset.

Table 4.5 reports the simulation results for OC scenario. The considered effect sizes for OC performance are $\delta = 2, 5, 10$. As it can be seen, for most of the considered simulation case studies, the performance of rank-depth method r-CUSUM is significantly better. However, the OC-ARL for r-EWMA and r-MWMASHC remains large for small effect size values, i.e. $\delta = 2$. It is also obvious that RE outperforms many of the considered methods for some of the cases and its results is comparable with r-CUSUM. It is observed that contaminating the original data undermines the performance of RE unlike r-CUSUM which shows a stable performance and is not affected by contamination of data with outliers. Increasing the dimension of these object data has a negative impact on the performance

Table 4.5: OC-ARL values for r-EWMA, r-CUSUM, and r-MWMASHC for von Mises-Fisher distribution for 5000 simulation runs. The values in the parentheses are standard deviations.

		Standard VMF			**Contaminated VMF**		
		p = 3	**p = 10**	**p = 20**	**p = 3**	**p = 10**	**p = 20**
r-EWMA($\lambda = 0.05$)	$\delta = 2$	74.43 (72.79)	92.51 (89.78)	84.08 (85.05)	118.13 (108.26)	132.22 (124.07)	123.71 (100.08)
	$\delta = 5$	36.74 (43.45)	70.52 (95.12)	36.93 (30.93)	38.33 (42.36)	49.72 (47.27)	38.25 (31.20)
	$\delta = 10$	17.42 (12.95)	45.51 (60.94)	27.41 (31.57)	14.62 (15.32)	48.04 (48.83)	48.98 (66.24)
r-EWMA($\lambda = 0.1$)	$\delta = 2$	61.96 (66.53)	82.76 (74.80)	105.39 (106.50)	48.16 (66.82)	102.58 (68.77)	119.28 (117.40)
	$\delta = 5$	25.60 (33.01)	58.19 (78.14)	43.52 (42.10)	20.87 (26.38)	81.89 (62.64)	50.26 (44.10)
	$\delta = 10$	11.02 (13.15)	23.06 (34.93)	35.57 (37.52)	13.83 (15.29)	39.97 (42.81)	43.31 (44.05)
r-EWMA($\lambda = 0.2$)	$\delta = 2$	72.56 (90.98)	135.07 (98.34)	145.98 (120.91)	34.49 (56.16)	130.32 (154.65)	115.21 (132.31)
	$\delta = 5$	22.69 (10.58)	65.74 (54.50)	60.44 (58.57)	30.30 (41.44)	89.74 (96.18)	43.79 (64.08)
	$\delta = 10$	8.49 (3.34)	48.65 (34.73)	36.39 (38.21)	8.99 (7.78)	45.72 (47.80)	56.45 (60.47)
r-CUSUM(k = 0.1)	$\delta = 2$	14.23 (35.73)	**14.64 (34.68)**	31.75 (54.64)	**8.88 (11.88)**	22.49 (54.23)	63.18 (108.35)
	$\delta = 5$	7.79 (12.44)	7.31 (6.44)	**6.75 (6.84)**	9.57 (9.64)	8.48 (10.23)	9.17 (11.97)
	$\delta = 10$	5.51 (5.74)	6.51 (9.78)	6.88 (3.43)	6.38 (6.46)	5.69 (9.57)	3.96 (5.98)
r-CUSUM(k = 0.2)	$\delta = 2$	**6.95 (4.49)**	19.39 (23.79)	**29.37 (13.07)**	26.66 (52.26)	46.16 (57.76)	52.02 (54.78)
	$\delta = 5$	5.59 (2.43)	**6.72 (8.01)**	8.51 (3.61)	5.05 (8.74)	**4.04 (11.01)**	**8.31 (11.65)**
	$\delta = 10$	4.65 (9.91)	3.12 (2.95)	4.86 (5.87)	5.17 (13.93)	6.06 (14.06)	4.85 (6.20)
r-CUSUM(k = 0.3)	$\delta = 2$	29.21 (65.78)	46.72 (113.38)	46.50 (55.45)	25.62 (53.19)	50.84 (97.18)	34.19 (40.69)
	$\delta = 5$	10.83 (30.69)	8.57 (31.11)	27.85 (54.11)	5.43 (14.16)	19.77 (55.80)	17.66 (23.29)
	$\delta = 10$	8.49 (13.48)	4.25 (19.85)	8.24 (13.58)	4.27 (9.58)	5.65 (20.66)	6.39 (8.34)
r-MWMASHC(m = 5)	$\delta = 2$	71.01 (116.08)	79.97 (89.96)	113.13 (103.95)	162.80 (161.91)	92.87 (84.04)	93.78 (118.64)
	$\delta = 5$	54.97 (97.43)	68.43 (76.44)	40.04 (43.86)	100.32 (122.15)	78.64 (67.70)	39.69 (50.86)
	$\delta = 10$	13.57 (37.82)	49.87 (28.66)	83.59 (101.76)	91.48 (162.61)	61.69 (87.18)	52.81 (87.64)
r-MWMASHC(m = 10)	$\delta = 2$	58.84 (84.33)	61.33 (78.20)	138.18 (177.46)	83.91 (126.21)	134.45 (151.79)	121.63 (122.72)
	$\delta = 5$	23.61 (49.05)	50.78 (55.71)	26.40 (36.54)	44.91 (56.10)	48.18 (48.28)	40.15 (38.70)
	$\delta = 10$	7.93 (2.90)	36.02 (52.76)	27.25 (60.14)	23.38 (71.93)	29.93 (24.72)	68.56 (96.99)
r-MWMASHC(m = 15)	$\delta = 2$	62.43 (73.92)	47.52 (75.20)	120.06 (126.28)	71.11 (74.65)	63.73 (67.56)	150.54 (152.90)
	$\delta = 5$	24.48 (37.41)	46.03 (51.78)	29.95 (29.78)	42.44 (58.56)	29.66 (32.01)	33.18 (31.46)
	$\delta = 10$	10.23 (3.68)	28.08 (44.63)	27.59 (49.72)	17.44 (13.50)	23.77 (19.38)	55.22 (72.26)
IPD (ϕ_1)	$\delta = 2$	139.98 (90.35)	149.39 (164.43)	142.49 (147.57)	154.22 (113.31)	104.07 (109.46)	157.32 (113.62)
	$\delta = 5$	77.41 (104.79)	53.72 (74.36)	91.03 (106.77)	121.92 (123.39)	55.95 (65.35)	109.93 (105.68)
	$\delta = 10$	58.15 (77.89)	49.45 (23.13)	42.32 (40.02)	85.72 (60.74)	43.48 (25.94)	42.08 (14.03)
IPD (ϕ_2)	$\delta = 2$	151.28 (164.17)	133.18 (105.76)	160.69 (143.20)	161.47 (173.26)	132.77 (129.89)	120.08 (138.49)
	$\delta = 5$	127.12 (126.12)	92.20 (90.65)	101.69 (123.49)	151.06 (158.81)	70.15 (68.78)	127.42 (148.73)
	$\delta = 10$	108.41 (107.54)	22.18 (15.75)	14.66 (23.04)	133.64 (127.91)	34.46 (35.63)	17.77 (26.15)
IPD (ϕ_3)	$\delta = 2$	142.39 (135.09)	143.26 (151.44)	60.51 (73.10)	158.27 (130.21)	149.57 (118.40)	116.70 (123.14)
	$\delta = 5$	102.82 (105.05)	126.62 (125.26)	26.93 (24.05)	104.77 (106.97)	117.39 (119.02)	104.09 (109.12)
	$\delta = 10$	108.15 (108.42)	83.36 (90.87)	32.56 (42.06)	70.33 (88.63)	108.51 (102.82)	102.58 (107.34)
DDMA (m = 5)	$\delta = 2$	48.70 (57.68)	18.47 (20.41)	72.02 (72.33)	36.14 (51.41)	24.07 (22.20)	**28.81 (41.50)**
	$\delta = 5$	27.40 (39.27)	14.08 (25.42)	52.80 (62.71)	13.45 (14.83)	11.45 (‘8.46)	13.45 (19.11)
	$\delta = 10$	7.15 (6.02)	5.37 (4.55)	23.45 (20.23)	10.11 (11.95)	5.54 (5.61)	6.11 (10.73)
DDMA (m = 10)	$\delta = 2$	31.06 (39.09)	29.60 (34.04)	45.65 (58.93)	52.68 (55.61)	48.87 (53.32)	41.48 (59.07)
	$\delta = 5$	9.66 (8.44)	9.67 (10.54)	15.88 (46.80)	9.32 (5.16)	16.42 (19.82)	12.40 (9.06)
	$\delta = 10$	5.62 (4.14)	4.94 (2.63)	4.69 (3.12)	5.87 (3.12)	4.92 (2.85)	5.70 (2.98)
DDMA (m = 15)	$\delta = 2$	43.76 (53.45)	28.87 (52.35)	32.07 (40.19)	45.39 (58.93)	**21.35 (21.08)**	44.68 (40.95)
	$\delta = 5$	11.90 (11.83)	9.96 (8.88)	10.05 (6.77)	20.69 (50.60)	9.81 (8.31)	15.01 (11.75)
	$\delta = 10$	7.14 (5.09)	5.96 (3.38)	5.86 (2.85)	7.97 (4.48)	6.15 (3.50)	6.10 (3.13)
RE	$\delta = 2$	44.08 (63.01)	73.11 (72.16)	30.64 (22.63)	62.23 (60.61)	84.44 (112.92)	29.33 (10.72)
	$\delta = 5$	**1.46 (1.53)**	10.73 (1.34)	11.10 (2.82)	**1.33 (2.08)**	6.01 (3.45)	20.50 (21.30)
	$\delta = 10$	**2.64 (3.15)**	**1.89 (1.61)**	**1.36 (2.41)**	**1.57 (1.33)**	**1.64 (1.42)**	**1.73 (1.53)**
HDSOR	$\delta = 2$	31.16 (51.31)	42.80 (49.26)	89.81 (101.24)	37.91 (33.07)	49.79 (49.46)	71.94 (89.06)
	$\delta = 5$	14.15 (11.79)	9.45 (4.40)	11.68 (6.51)	11.51 (9.46)	10.35 (5.36)	12.96 (7.89)
	$\delta = 10$	7.62 (2.58)	5.49 (1.38)	6.15 (2.07)	7.73 (3.36)	5.73 (1.54)	6.11 (1.85)
HDSIP	$\delta = 2$	98.69 (44.17)	93.59 (45.33)	113.48 (53.13)	91.50 (12.19)	105.62 (29.94)	79.18 (50.80)
	$\delta = 5$	84.29 (18.78)	80.30 (15.63)	86.77 (20.88)	76.51 (12.55)	82.46 (14.45)	46.05 (15.34)
	$\delta = 10$	80.06 (11.41)	73.72 (16.06)	26.07 (38.87)	64.66 (8.53)	17.88 (16.98)	26.52 (35.22)

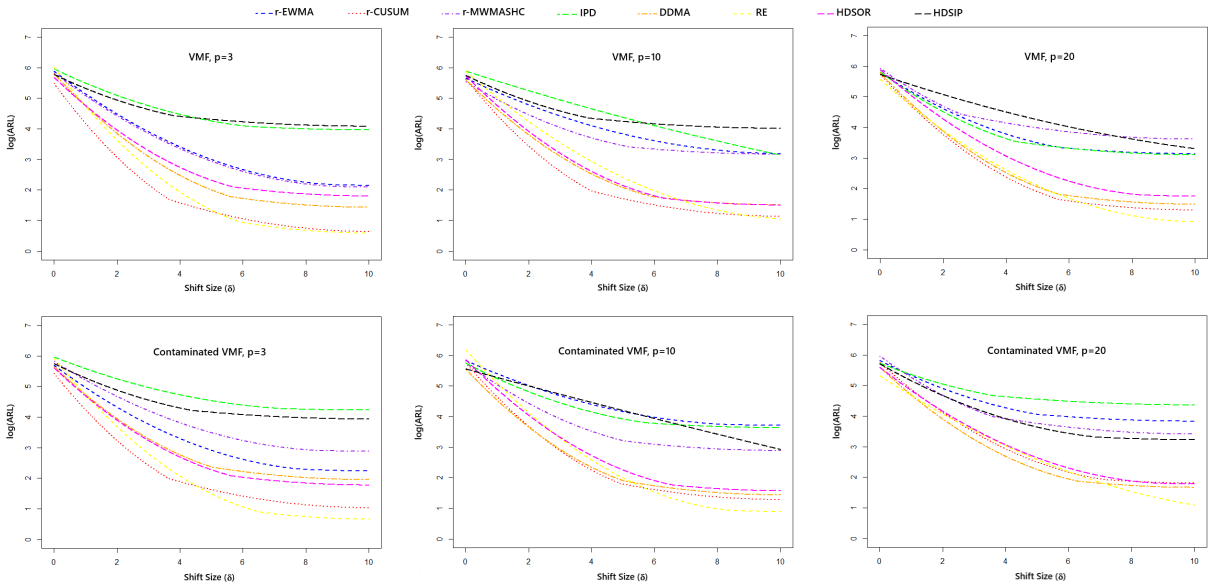


Figure 4.5: OC-ARL values for the process monitoring control charts for VMF and contaminated VMF with shift sizes of range $(0, 10)$ and dimensions $(3, 10, 20)$.

of the distance-based methods such as HDSOR, HDSIP, and IPD. This is not the case for rank-depth methods and their results are stable as the dimension is increased.

Figure 4.5 depicts the logarithm of ARL values for the effect sizes in the range $(0, 10)$. Top panels depict the log(ARL) values for original data with dimensions $3, 10, 20$ and bottom panels are the log(ARL) values for contaminated data with dimensions $3, 10, 20$. Clearly, r-CUSUM and RE show a very good balance for the considered range of effect sizes and are superior or comparable to the other nonparametric methods. Also, it can be realized that DDMA shows promising results for some of the cases, while r-EWMA and r-MWMASHC show stably acceptable performance for the considered simulation scenarios. It is observed that IPD and HDSIP show inferior performance for most of the considered cases, which augurs the fact that their performance can become unstable with respect to the considered manifold spaces, given that the IPD performed superior for Wishart distributed object data. For most of the considered cases, the fastest decrease belongs to either r-CUSUM or RE while r-MWMASHC and DDMA remain comparable to these two control charts. All in all, the obtained results endorse the effectiveness of the proposed rank-depth methods given the fact that they successfully keep a good balance between IC-ARL and OC-ARL for both original and contaminated data over the considered dimensions of the manifold.

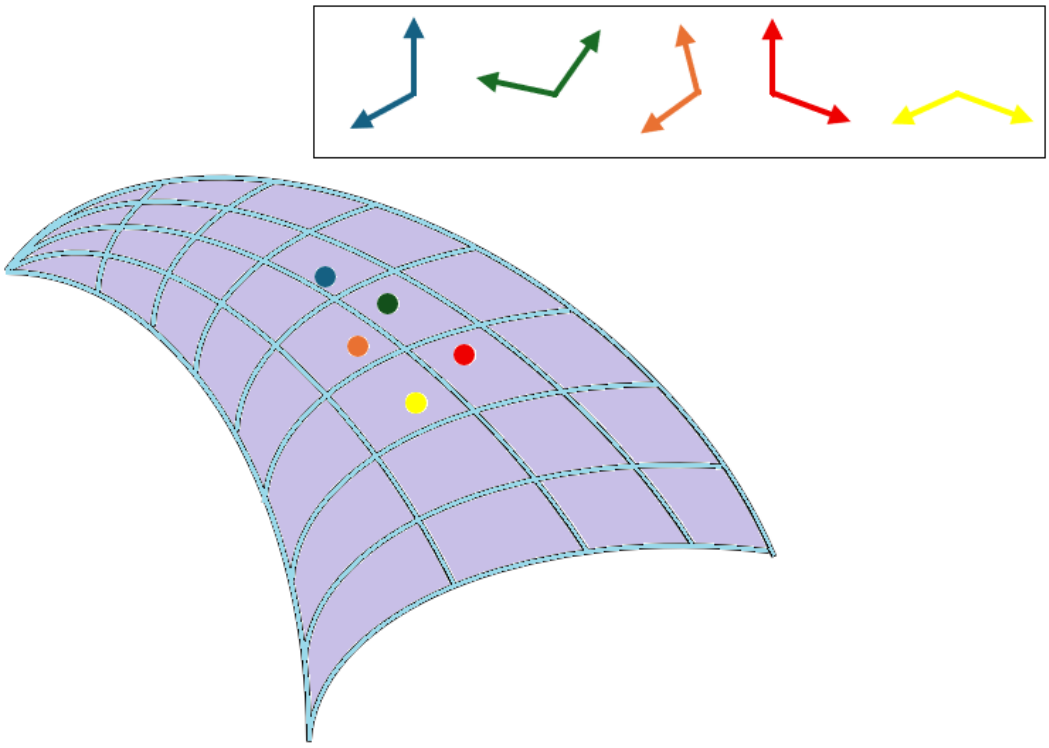


Figure 4.6: Schematic illustration of Steifel manifold data corresponding to 2-frames in $\mathbb{R}^3$.

Example 3 (Uniform and Gaussian Distributions on Steifel Manifold): In this simulation, we compare the rank-depth process monitoring methods with adopted non-parametric methods on data generated from distributions on Steifel manifold.

Figure 4.6 indicates the schematic illustration of object data on Steifel manifold corresponding to 2-frames in $\mathbb{R}^3$. For the IC condition, we draw samples using Uniform distribution on Steifel manifold. For OC condition, we generate from Gaussian distribution on Steifel manifold. The simulation is carried out for dimensions $3, 11, 21$ which correspond to 1-frame in $\mathbb{R}^3$, 2-frame in $\mathbb{R}^6$, and 3-frame in $\mathbb{R}^8$. Our experiments revealed that this problem is more challenging than the other two examples, as distinguishing data sampled from uniform and Gaussian on Steifel manifold, object data are k-frames in $\mathbb{R}^d$, can become difficult. To evaluate the robustness of the methods, we consider contaminating the original data with outliers at 50% level. For IC condition, the contaminating data are sampled from Gaussian distribution and for OC condition the contaminating data are sampled from uniform distribution.

Table 4.6: IC-ARL values for r-EWMA, r-CUSUM, and r-MWMASHC for Uniform distribution on Steifel manifold for 5000 simulation runs. The values in the parentheses are standard deviations.

	Steifel Unif			**Contaminated Steifel Unif**		
			$ARL_0 = 200$			
	p = 3	**p = 11**	**p = 21**	**p = 3**	**p = 11**	**p = 21**
r-EWMA($\lambda = 0.05$)	220.5 (232.3)	192.3 (218.2)	161.1 (143.1)	246.1 (211.6)	175.2 (145.8)	153.1 (150.3)
r-EWMA($\lambda = 0.1$)	229.4 (233.8)	208.52 (230.6)	168.1 (193.1)	**261.9 (275.2)**	185.9 (179.4)	177.6 (170.9)
r-EWMA($\lambda = 0.2$)	240.2 (247.07)	233.8 (232.9)	177.7 (205.8)	254.1 (264.5)	208.4 (215.7)	187.6 (201.3)
r-CUSUM($k = 0.1$)	246.3 (223.8)	206.4 (221.3)	234.9 (242.4)	235.7 (200.4)	217.6 (253.1)	**243.9 (230.4)**
r-CUSUM($k = 0.2$)	**250.7 (243.0)**	222.1 (213.2)	220.6 (214.1)	224.3 (236.9)	184.8 (205.6)	197.7 (208.5)
r-CUSUM($k = 0.3$)	184.6 (204.1)	202.4 (217.46)	**250.4 (247.3)**	185.5 (209.3)	**241.5 (207.3)**	197.9 (207.9)
r-MWMASHC($m = 5$)	190.1 (200.4)	212.6 (211.1)	188.6 (173.9)	173.9 (189.2)	224.3 (215.7)	227.2 (232.6)
r-MWMASHC($m = 10$)	244.7 (245.5)	193.9 (249.4)	243.2 (232.9)	216.1 (203.2)	207.7 (196.3)	238.8 (225.9)
r-MWMASHC($m = 15$)	204.5 (208.1)	214.4 (185.2)	205.6 (198.3)	206.2 (204.6)	198.1 (196.6)	222.4 (231.2)
IPD (ϕ_1)	227.5 (192.2)	225.5 (209.8)	229.7 (208.4)	182.2 (209.6)	151.1 (149.2)	201.5 (227.1)
IPD (ϕ_2)	205.4 (180.5)	**234.5 (207.2)**	210.5 (180.9)	134.5 (129.3)	118.5 (116.9)	178.1 (190.3)
IPD (ϕ_3)	229.5 (198.5)	198.7 (173.2)	208.3 (173.9)	135.2 (106.7)	122.2 (151.3)	139.1 (150.3)
DDMA ($m = 5$)	165.2 (89.1)	187.2 (96.8)	199.1 (131.9)	166.1 (98.1)	171.2 (100.2)	173.9 (114.4)
DDMA ($m = 10$)	168.7 (116.1)	177.5 (86.1)	192.5 (117.6)	172.8 (102.3)	147.0 (166.5)	157.9 (167.7)
DDMA ($m = 15$)	177.1 (157.5)	181.2 (78.2)	179.7 (107.2)	154.4 (79.8)	166.0 (120.7)	172.3 (98.1)
RE	113.8 (61.1)	118.5 (133.2)	134.5 (137.9)	86.4 (126.9)	168.7 (120.5)	93.2 (73.2)
HDSOR	234.6 (243.9)	216.6 (185.2)	236.3 (226.1)	164.5 (228.6)	211.1 (200.4)	176.3 (207.7)
HDSIP	180.3 (139.6)	155.2 (123.2)	156.3 (101.1)	177.2 (131.8)	165.1 (127.3)	170.6 (122.1)
			$ARL_0 = 370$			
	p = 3	**p = 11**	**p = 21**	**p = 3**	**p = 11**	**p = 21**
r-EWMA($\lambda = 0.05$)	**422.4 (269.2)**	348.8 (265.9)	395.1 (244.1)	310.3 (284.5)	390.4 (259.4)	377.1 (289.1)
r-EWMA($\lambda = 0.1$)	401.6 (262.8)	367.1 (258.5)	375.2 (268.7)	357.1 (283.1)	310.4 (267.2)	347.4 (274.8)
r-EWMA($\lambda = 0.2$)	385.2 (265.8)	386.5 (267.3)	385.2 (279.4)	355.3 (283.8)	321.3 (281.9)	310.5 (269.7)
r-CUSUM($k = 0.1$)	387.7 (268.9)	410.5 (269.2)	346.5 (271.4)	395.1 (287.6)	391.2 (256.4)	333.88 (273.1)
r-CUSUM($k = 0.2$)	359.2 (289.2)	**438.3 (295.5)**	374.7 (297.4)	371.6 (291.3)	362.1 (281.8)	**405.9 (337.3)**
r-CUSUM($k = 0.3$)	364.9 (261.7)	375.7 (303.9)	398.1 (304.5)	352.1 (256.6)	**420.6 (274.7)**	376.5 (295.3)
r-MWMASHC($m = 5$)	385.5 (282.6)	431.1 (317.3)	391.6 (280.7)	365.7 (232.8)	395.9 (270.6)	385.0 (271.4)
r-MWMASHC($m = 10$)	395.9 (301.8)	436.5 (265.6)	382.7 (236.5)	398.9 (295.1)	388.2 (267.3)	358.9 (230.5)
r-MWMASHC($m = 15$)	369.4 (290.4)	417.6 (248.6)	401.1 (280.2)	419.6 (256.1)	352.8 (298.4)	365.7 (283.9)
IPD (ϕ_1)	313.3 (269.3)	412.1 (282.6)	397.3 (134.2)	315.6 (256.6)	396.1 (264.1)	305.1 (224.7)
IPD (ϕ_2)	395.8 (270.6)	412.0 (224.1)	**438.6 (286.3)**	319.1 (248.9)	319.3 (232.7)	392.9 (240.6)
IPD (ϕ_3)	414.9 (270.9)	412.3 (269.6)	377.4 (246.2)	315.5 (221.3)	363.8 (210.8)	315.7 (248.5)
DDMA ($m = 5$)	319.4 (125.5)	315.9 (170.7)	338.2 (158.6)	292.7 (121.8)	282.6 (188.1)	308.7 (146.1)
DDMA ($m = 10$)	317.4 (141.5)	326.7 (113.1)	337.4 (107.3)	296.4 (115.1)	284.4 (191.2)	296.1 (122.0)
DDMA ($m = 15$)	334.1 (198.2)	319.5 (110.3)	315.5 (197.4)	332.4 (140.9)	320.6 (149.6)	310.0 (168.8)
RE	323.4 (229.6)	285.2 (184.6)	262.6 (182.2)	220.6 (106.9)	323.2 (259.3)	297.1 (154.3)
HDSOR	382.6 (254.0)	339.8 (247.5)	361.0 (261.2)	**419.7 (237.9)**	346.7 (260.2)	337.9 (248.9)
HDSIP	331.5 (157.9)	287.3 (143.7)	288.1 (115.6)	305.8 (144.8)	297.6 (146.4)	308.4 (137.8)

Table 4.7: OC-ARL values for r-EWMA, r-CUSUM, and r-MWMASHC for Uniform and Gaussian distributions on Steifel manifold for 5000 simulation runs. The values in the parentheses are standard deviations.

		Steifel Gaussian			**Contaminated Steifel Gaussian**		
		p = 3	**p = 11**	**p = 21**	**p = 3**	**p = 11**	**p = 21**
r-EWMA($\lambda = 0.05$)	$\sigma^2 = 10$	27.85 (90.48)	23.90 (65.81)	**24.13 (20.07)**	29.13 (50.06)	39.51 (46.09)	23.60 (29.38)
	$\sigma^2 = 30$	**38.20 (81.59)**	**32.50 (58.77)**	34.83 (38.99)	38.98 (58.66)	31.50 (47.95)	23.99 (48.46)
	$\sigma^2 = 50$	46.24 (90.18)	43.50 (60.40)	43.40 (32.01)	52.36 (85.04)	41.53 (78.77)	22.50 (53.40)
r-EWMA($\lambda = 0.1$)	$\sigma^2 = 10$	38.29 (88.13)	35.06 (71.89)	29.48 (45.38)	35.61 (52.06)	33.33 (70.13)	21.95 (19.99)
	$\sigma^2 = 30$	41.59 (57.43)	40.83 (39.17)	**31.50 (30.59)**	32.29 (43.45)	**30.15 (57.98)**	23.30 (45.93)
	$\sigma^2 = 50$	44.77 (94.90)	**35.89 (26.82)**	38.20 (24.79)	34.50 (59.19)	**40.50 (95.43)**	**20.34 (56.94)**
r-EWMA($\lambda = 0.2$)	$\sigma^2 = 10$	**24.60 (67.78)**	34.09 (69.66)	24.85 (63.42)	**21.62 (32.69)**	41.48 (54.98)	26.91 (23.38)
	$\sigma^2 = 30$	28.50 (72.41)	53.05 (83.07)	35.89 (92.18)	**22.17 (59.82)**	45.89 (24.57)	31.97 (68.35)
	$\sigma^2 = 50$	**31.82 (49.99)**	56.48 (99.04)	51.95 (34.82)	**30.59 (59.48)**	45.58 (54.91)	32.05 (52.18)
r-CUSUM($k = 0.1$)	$\sigma^2 = 10$	75.95 (94.36)	110.40 (106.82)	41.48 (66.12)	98.64 (91.26)	110.65 (103.83)	66.86 (94.34)
	$\sigma^2 = 30$	77.46 (78.54)	83.77 (136.65)	116.56 (130.68)	130.54 (97.09)	92.56 (104.15)	76.30 (62.69)
	$\sigma^2 = 50$	113.11 (130.31)	100.29 (107.98)	126.22 (114.66)	98.50 (104.61)	106.91 (102.84)	109.23 (122.71)
r-CUSUM($k = 0.2$)	$\sigma^2 = 10$	62.78 (103.52)	**10.77 (24.43)**	50.13 (84.04)	91.02 (85.67)	62.24 (63.22)	81.71 (84.38)
	$\sigma^2 = 30$	79.54 (114.86)	67.15 (98.73)	93.22 (79.26)	76.34 (107.10)	44.15 (59.66)	45.38 (95.27)
	$\sigma^2 = 50$	59.24 (33.58)	71.17 (51.39)	85.92 (103.77)	111.31 (75.47)	91.78 (112.81)	44.03 (52.48)
r-CUSUM($k = 0.3$)	$\sigma^2 = 10$	52.34 (92.91)	57.57 (97.09)	28.24 (57.53)	69.96 (87.52)	**25.45 (55.01)**	16.01 (27.74)
	$\sigma^2 = 30$	62.87 (63.56)	62.26 (115.22)	44.57 (58.77)	62.31 (77.32)	86.25 (76.94)	53.67 (50.71)
	$\sigma^2 = 50$	33.81 (66.89)	55.35 (62.55)	81.38 (88.11)	63.21 (64.02)	83.87 (103.62)	31.94 (214.46)
r-MWMASHC($m = 5$)	$\sigma^2 = 10$	42.54 (77.15)	43.67 (52.52)	55.84 (54.82)	40.58 (62.75)	58.93 (55.52)	**15.18 (55.77)**
	$\sigma^2 = 30$	41.93 (70.52)	71.51 (68.89)	37.30 (57.36)	53.83 (43.17)	65.49 (57.12)	**23.20 (49.34)**
	$\sigma^2 = 50$	55.39 (58.14)	54.59 (51.13)	**38.12 (54.57)**	48.94 (50.29)	67.62 (73.55)	48.14 (41.43)
r-MWMASHC($m = 10$)	$\sigma^2 = 10$	60.61 (72.62)	50.13 (76.65)	52.36 (81.80)	45.87 (62.12)	59.11 (90.49)	69.63 (81.46)
	$\sigma^2 = 30$	68.65 (88.21)	79.85 (80.13)	57.52 (67.79)	53.55 (72.86)	88.77 (91.13)	61.52 (88.49)
	$\sigma^2 = 50$	72.27 (100.53)	105.52 (108.87)	55.82 (68.69)	76.79 (70.76)	67.54 (73.12)	79.32 (98.88)
r-MWMASHC($m = 15$)	$\sigma^2 = 10$	70.63 (97.17)	80.68 (100.89)	71.53 (118.64)	83.42 (100.31)	64.66 (70.17)	64.74 (92.85)
	$\sigma^2 = 30$	75.51 (90.94)	91.12 (113.79)	76.53 (123.46)	93.73 (137.89)	97.23 (114.39)	57.53 (98.14)
	$\sigma^2 = 50$	69.55 (77.27)	107.65 (97.63)	82.27 (127.72)	79.21 (100.98)	86.53 (118.16)	72.51 (79.73)
IPD (ϕ_1)	$\sigma^2 = 10$	128.23 (131.33)	103.34 (111.95)	149.71 (152.26)	129.53 (132.17)	134.58 (147.34)	159.93 (162.56)
	$\sigma^2 = 30$	160.14 (176.36)	131.53 (155.67)	160.68 (114.89)	128.02 (139.21)	136.56 (140.18)	176.53 (180.24)
	$\sigma^2 = 50$	173.21 (171.95)	141.86 (147.97)	179.65 (114.88)	147.12 (149.93)	131.68 (134.63)	197.41 (194.27)
IPD (ϕ_2)	$\sigma^2 = 10$	132.13 (103.88)	115.18 (109.07)	150.51 (149.65)	100.04 (89.31)	123.86 (165.69)	152.88 (107.96)
	$\sigma^2 = 30$	132.88 (132.95)	121.32 (109.48)	159.46 (106.61)	143.96 (112.17)	139.50 (144.02)	175.71 (101.58)
	$\sigma^2 = 50$	156.21 (117.72)	137.08 (149.02)	164.83 (126.22)	153.63 (152.74)	158.38 (149.80)	149.97 (105.83)
IPD (ϕ_3)	$\sigma^2 = 10$	129.56 (119.41)	101.41 (88.78)	145.51 (120.50)	122.68 (109.07)	136.38 (145.91)	97.48 (100.49)
	$\sigma^2 = 30$	152.25 (161.23)	123.73 (124.84)	148.38 (155.99)	144.54 (157.64)	141.35 (148.05)	151.11 (154.98)
	$\sigma^2 = 50$	162.10 (152.27)	146.77 (151.52)	154.96 (143.87)	150.95 (142.49)	145.64 (151.80)	190.31 (200.29)
DDMA ($m = 5$)	$\sigma^2 = 10$	178.88 (47.75)	128.89 (71.60)	131.58 (90.09)	131.82 (75.35)	127.50 (119.30)	58.73 (76.79)
	$\sigma^2 = 30$	186.66 (85.27)	166.87 (116.15)	143.68 (93.22)	164.66 (79.44)	130.31 (134.69)	127.02 (67.94)
	$\sigma^2 = 50$	194.85 (72.12)	168.72 (69.90)	151.25 (117.29)	166.78 (89.40)	159.30 (96.97)	131.30 (125.98)
DDMA ($m = 10$)	$\sigma^2 = 10$	162.36 (109.74)	79.23 (126.18)	138.59 (85.35)	148.30 (65.68)	173.11 (110.84)	131.39 (103.85)
	$\sigma^2 = 30$	169.49 (101.73)	143.53 (100.43)	149.29 (52.77)	158.70 (74.73)	164.50 (93.86)	160.98 (87.52)
	$\sigma^2 = 50$	183.34 (98.23)	158.98 (90.85)	186.90 (138.91)	167.17 (99.54)	162.60 (123.26)	180.35 (123.16)
DDMA ($m = 15$)	$\sigma^2 = 10$	122.84 (109.46)	50.37 (97.39)	128.71 (94.99)	148.56 (61.65)	128.36 (132.02)	135.73 (124.24)
	$\sigma^2 = 30$	145.34 (70.83)	73.60 (101.89)	154.24 (97.28)	159.66 (90.20)	155.10 (82.93)	158.61 (81.97)
	$\sigma^2 = 50$	172.34 (110.55)	137.89 (115.48)	166.75 (96.94)	175.39 (108.25)	161.81 (97.06)	163.31 (88.67)
RE	$\sigma^2 = 10$	90.57 (87.75)	72.80 (115.83)	74.96 (111.47)	64.04 (54.14)	93.62 (176.68)	105.72 (112.42)
	$\sigma^2 = 30$	114.06 (126.39)	129.52 (146.62)	96.73 (74.01)	54.61 (59.39)	105.28 (154.73)	100.18 (70.52)
	$\sigma^2 = 50$	116.52 (142.12)	100.07 (128.08)	120.66 (89.16)	103.47 (150.73)	102.22 (145.21)	110.60 (160.29)
HDSOR	$\sigma^2 = 10$	119.83 (119.01)	114.68 (100.77)	121.23 (110.80)	136.18 (91.21)	150.77 (103.50)	139.05 (169.12)
	$\sigma^2 = 30$	130.90 (116.53)	126.23 (151.73)	138.02 (123.30)	157.69 (111.68)	157.06 (124.57)	148.28 (145.82)
	$\sigma^2 = 50$	145.86 (155.70)	153.47 (118.68)	143.43 (115.33)	180.92 (134.49)	162.50 (148.40)	153.87 (201.28)
HDSIP	$\sigma^2 = 10$	65.48 (82.76)	78.21 (71.22)	52.81 (53.62)	101.13 (107.20)	79.12 (57.23)	57.67 (46.54)
	$\sigma^2 = 30$	106.20 (101.06)	98.08 (128.24)	85.53 (70.63)	108.74 (108.75)	94.04 (100.05)	87.23 (68.19)
	$\sigma^2 = 50$	137.40 (150.61)	118.92 (123.81)	107.18 (102.74)	136.55 (135.95)	124.14 (70.24)	90.21 (46.59)

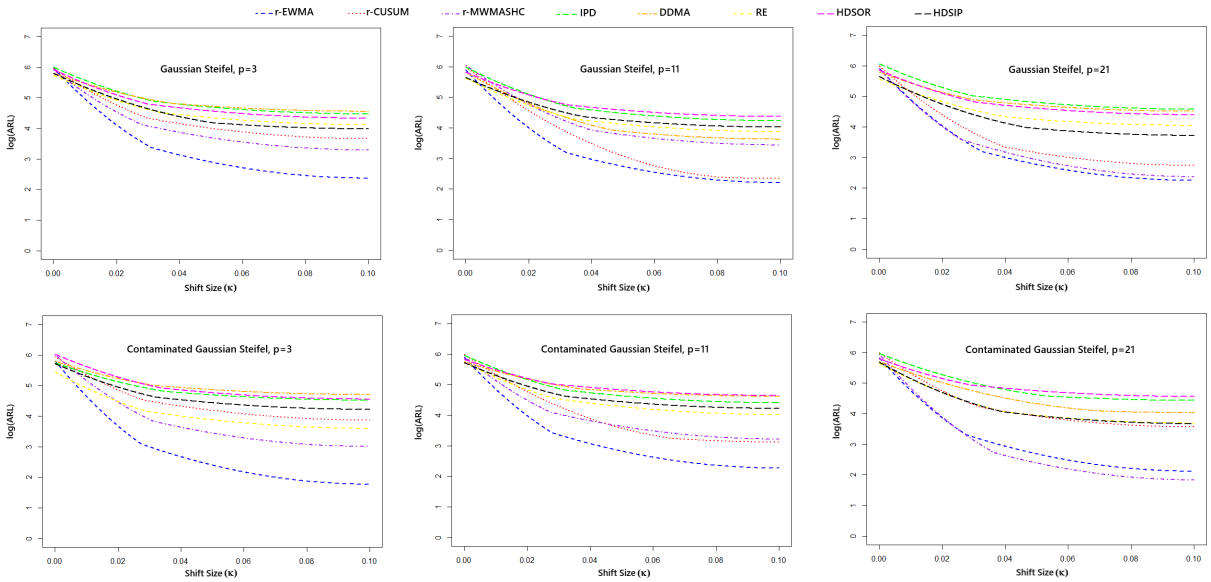


Figure 4.7: OC-ARL values for the process monitoring control charts for Gaussian Steifel and contaminated Gaussian Steifel with shift sizes of range $(0, 0.1)$ and dimensions $(3, 11, 20)$.

Table 4.6 compares the results for IC condition. It is clear that rank-depth methods perform significantly better than the adopted nonparametric methods. For most of the cases, r-EWMA, r-CUSUM, and r-MWMASHC have IC-ARL above the specified value. This is not the case for nonparametric methods as they cannot control the IC-ARL at the specified value. The performance of rank-depth methods is stable after contaminating data with outliers. For the other nonparametric methods, the performance deteriorates as the original data is contaminated with outliers. From the results of the performed experiments, it can be inferred that rank-depth methods enjoy generalization power as these methods are stably superior to the adopted nonparametric methods regardless of the geometric properties of manifold. This is more apparent for the Steifel manifold data which is more challenging compared to Wishart and VMF case studies.

Table 4.7 reports the results for OC condition for effect sizes $\kappa = 1/\sigma^2$ where $\sigma^2 = 10, 30, 50$ is the variance of normal distribution on Steifel manifold, and κ is the concentration parameter of the distribution. Both uniform and Gaussian distributions are sampled around origin, upper $k \times k$ elements of $d \times k$ matrix are 1 and others are 0 (Chakraborty and Vemuri, 2019), so the effect size increases as the variance of Gaussian distribution decreases and tails become shorter. It can be seen that the obtained OC-ARL values for r-EWMA is

Table 4.8: Percentage of times a method outperformed the other rival methods.

Proposed			**Literature**				
r-EWMA	**r-CUSUM**	**r-MWMASHC**	**IPD**	**DDMA**	**RE**	**HDSOR**	**HDSIP**
32.2%	28.9%	12.2%	7.8%	3.4%	12.2%	2.2%	1.1%
	Total				**Total**		
	73.3%				26.7%		

smaller than those of the other rival methods. For most of the considered cases, the rank-depth methods show a significantly better performance than the adopted nonparametric methods. It can also be observed that contaminating the original data with outliers does not undermine the performance of these methods. For the other nonparametric methods, introducing contamination data can affect their performance.

Figure 4.7 compares the OC-ARL values for the process control charts on Gaussian steifel and contaminated Gaussian Steifel with shift sizes in the range $(0, 0.1)$ and dimensions $3, 11, 20$. As it can be seen, r-EWMA, r-CUSUM, and r-MWMASHC show much better performance compared to the adopted nonparametric control charts. The logarithm of ARL values for these control charts decreases exponentially faster than the adopted techniques as the effect size increases. This behaviour remains unchanged as the original data are contaminated with outliers, demonstrating the efficacy of using depth functions for developing nonparametric robust control charts.

Example 4 (Computational Time and General Discussion): In this example, we discuss our general findings from the carried out experiments in Examples 1-3. Table 4.8 reports the percentage of times a method outperformed other rival approaches considering the whole simulation case studies reported in the above tables.

It can be seen that in total for 73.3% of simulation experiments, the proposed rank-depth process monitoring methods outperformed the rival approaches, while for 26.7% of simulation experiments, the adopted nonparametric methods outperformed the proposed approaches. The table also reports the percentage of times any single one of these methods was superior. It can be seen that r-EWMA has the best performance with 32.2% of times being superior, while r-CUSUM stands in the second place with 28.9% of times being superior. Both r-MWMASHC and RE stand in the third place with 12.2% of times being superior to others. IPD stands in the fourth place with being superior to other methods for 7.8% of times, followed by DDMA, HDSOR, and HDSIP being superior for 3.4%, 2.2%, and 1.1% of times.

It can be concluded that robustness of the proposed methods has a significant impact on their success given the fact that for half of these simulations, the dataset are contami-

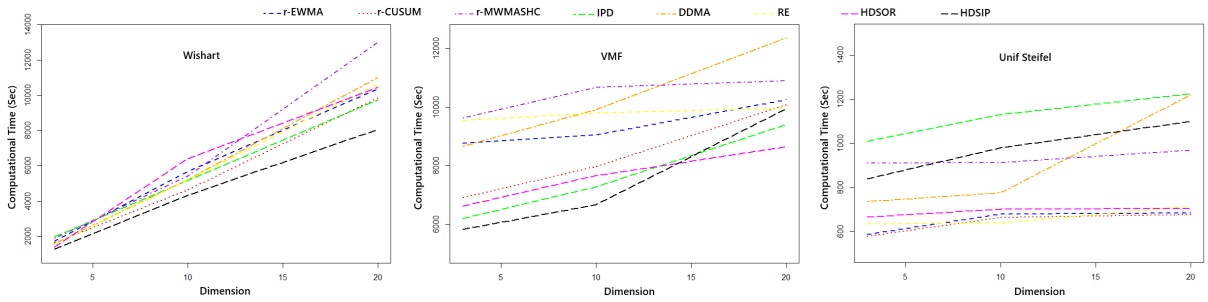


Figure 4.8: Computation time for the considered control charts.

nated with outliers. For many of these simulation cases, it was observed that introducing contaminated data weakens the performance of adopted nonparametric methods which are based on distances between data points rather than depth values. It is also worth noting that our simulations indicated that RE can be computationally more expensive than rank-depth methods which further justifies the use of the proposed methods.

Figure 4.8 compares the average computational time for the considered control charts for Wishart, VMF, and Uniform Steifel cases as a function of intrinsic dimension of the corresponding manifold spaces. It can be seen that the computation time increases linearly as the dimension of the underlying manifold spaces increases. Distance-based approaches yield less stable results as calculating the inter-distances can increase the computational time depending on the geometrical features of the underlying manifold spaces. This behaviour is apparent in HDSIP and HDSOR. The computational time of the rank-depth methods remain reasonably stable and for most of these cases are less than or comparable to most of the adopted nonparametric methods. This superiority is clear for the case of r-CUSUM which not only stands as the most successful method among the rival techniques, but also consumes a very low computational time for all of the cases.

Figure 4.9 depicts the phase-II nonparametric process monitoring results with r-EWMA, r-CUSUM, r-MWMASHC, IPD, DDMA, and RE control charts along with their control limits for Wishart distribution on OC condition. The first 100 data are for phase-I for which we have IC condition and the chart values stay within the control limits.

It is observed that when we get into the OC condition, the chart values violate the control limits and trigger alarm indicating faulty performance. This pattern can be observed in all of the considered nonparametric control charts. It is also clear that r-EWMA, r-CUSUM, r-MWMASHC, and DDMA are two-sided control charts while IPD and RE are one-sided control charts.

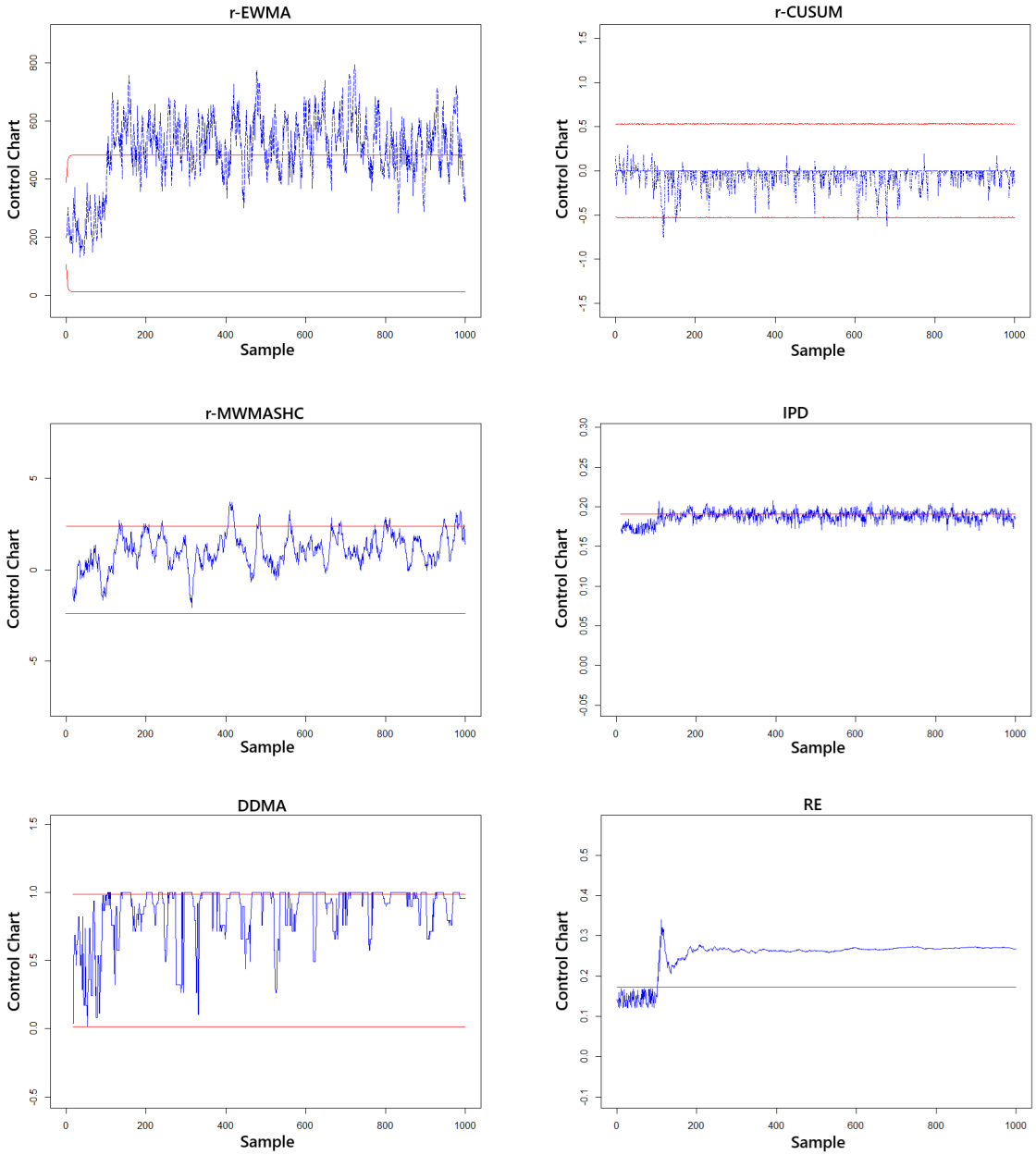


Figure 4.9: Process monitoring with r-EWMA, r-CUSUM, r-MWMASHC, IPD, DDMA, and RE control charts and their control limits for Wishart distribution.

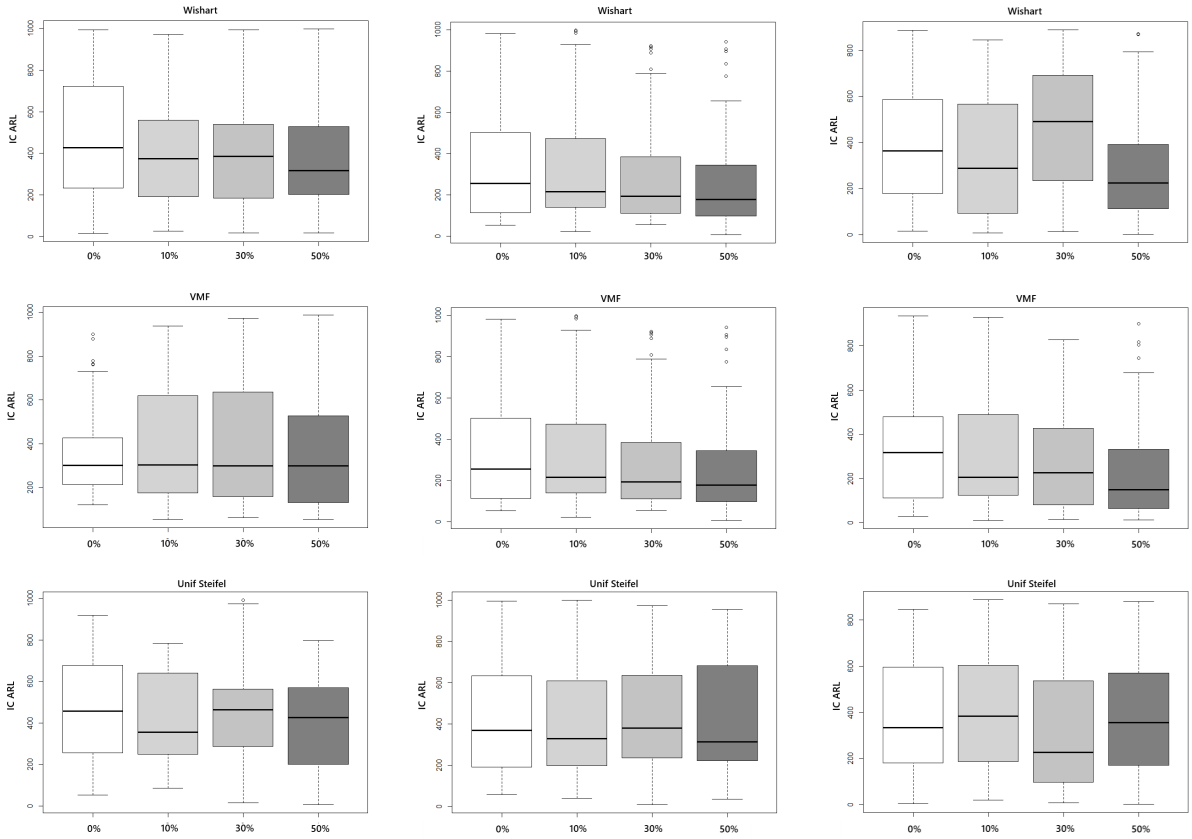


Figure 4.10: ARL boxplots for the process monitoring control charts r-EWMA, r-CUSUM, and r-MWMASHC under contamination noises with fractions $(0\%, 10\%, 30\%, 50\%)$.

4.3.2 Robustness Analysis

In this section, we investigate the robustness of the proposed rank-depth process monitoring methods using MCMC numerical simulation and theoretical breakdown point analysis considering Wishart distribution on $SPD(3)$, VMF on $\mathcal{S}^2$, and uniform distribution on $St(2,3)$.

Example 5 (Contaminating Wishart, VMF, and Uniform Steifel Data): In this example, we investigate the robustness of the rank-depth control charts by contaminating the original data with outliers at levels $0\%, 10\%, 30\%, 50\%$. The calculations are done using GD depth function. Figure 4.10 depicts the ARL boxplots for the process monitoring control charts r-EWMA, r-CUSUM, and r-MWMASHC under contamination noises with

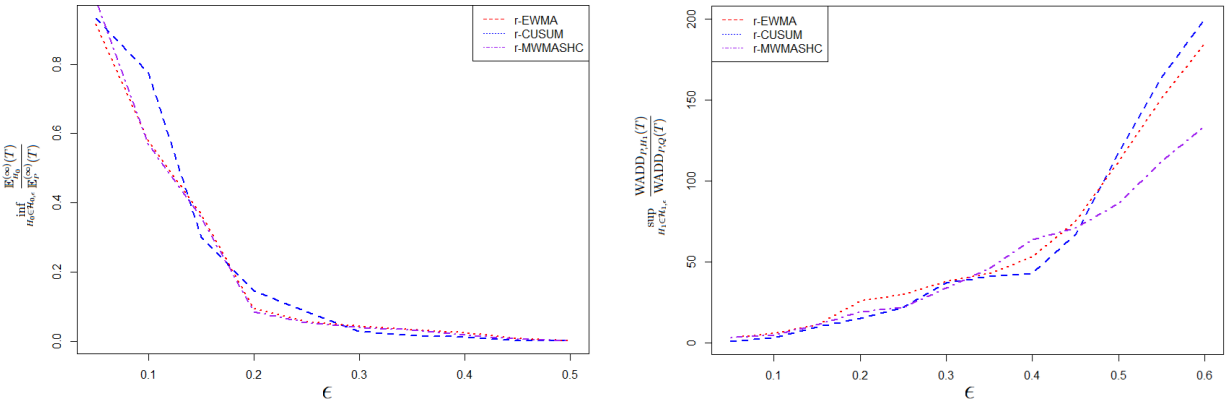


Figure 4.11: Gross-error sensitivities for r-EWMA, r-CUSUM, and r-MWMASHC for IC and OC conditions with different values of ε.

fractions $0\%, 10\%, 30\%, 50\%$. The top three panles compare the performance of r-EWMA, r-CUSUM, and r-MWMASHC for Wishart distribution, respectively.

As it can be seen, the ARL values remains almost stable as we increase the level of contamination noise. The interesting observation is that as the level of contamination increases a slight decrease in both median ARL and third quartile ARL can be observed while first quartile is almost unchanged. This is a reasonable observation as introducing outliers should affect the best-case scenario, i.e. the maximum ARL value achieved under IC condition, however the worst-case scenario, the minimum ARL value achieved, remains unchanged. A similar behaviour is observed for VMF and Uniform Steifel manifold data, demonstrating the robustness of the proposed methods against contaminating outliers, thanks to the use of depth function and ranking for calculating ARL values.

Example 6 (Breakdown Point Analysis for Wishart Distribution on SPD(3)): In this example, we evaluate the robustness of r-EWMA, r-CUSUM, and r-MWMASHC using breakdown point for Wishart distribution. For IC condition, we calculate the ratio $\inf_{H_0 \in \mathcal{H}_{0,\epsilon}} \frac{\mathbb{E}_{H_0}^{(\infty)}(T)}{\mathbb{E}_{P}^{(\infty)}(T)}$ as a function of contamination level ϵ. For OC condition, we calculate the ratio $\sup_{H_1 \in \mathcal{H}_{1,\epsilon}} \frac{\mathrm{WADD}_{P,H_1}(T)}{\mathrm{WADD}_{P,Q}(T)}$ as a function of contamination level ϵ.

Figure 4.11 illustrates the gross-error sensitivities for r-EWMA, r-CUSUM, and r-MWMASHC for IC and OC conditions with different values of ϵ. The left panel evaluates the robustness of the proposed methods for IC condition. It is observed that increasing the contamination level decreases IC-ARL so that the ratio $\inf_{H_0 \in \mathcal{H}_{0,\epsilon}} \frac{\mathbb{E}_{H_0}^{(\infty)}(T)}{\mathbb{E}_{P}^{(\infty)}(T)}$ approaches 0.

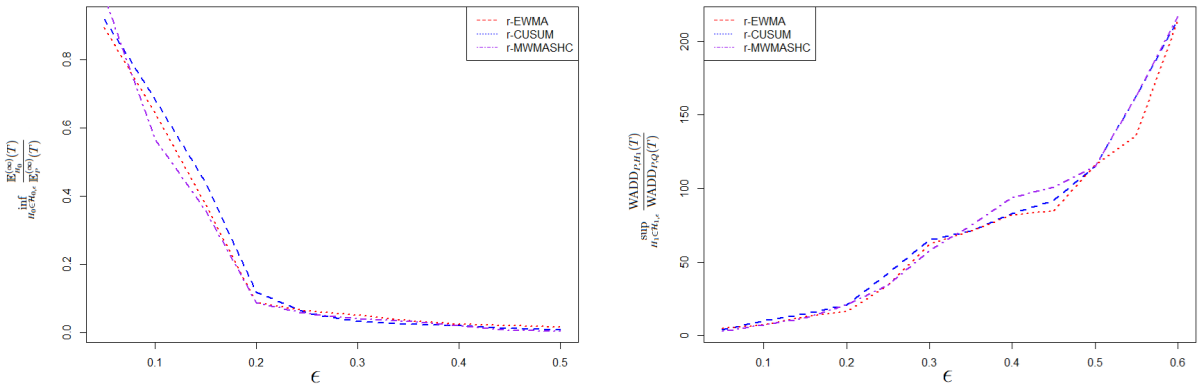


Figure 4.12: Gross-error sensitivities for r-EWMA, r-CUSUM, and r-MWMASHC for IC and OC conditions with different values of ε.

The performance of the proposed methods is comparable and suggests false-alarm breakdown point $\varepsilon_F = 0.3$. The right panel compares the robustness of these methods for OC condition. It can be inferred that increasing the contamination level increases the OC-ARL so that the ratio $\sup_{H_1 \in \mathcal{H}_{1,\epsilon}} \frac{\text{WADD}_{P,H_1}(T)}{\text{WADD}_{P,Q}(T)}$ becomes very large. The results indicate that the performance of these methods are comparable with detection-delay breakdown point $\varepsilon_D = 0.4$. Overall, the results demonstrate the robustness of the proposed methods given that the obtained values for ε_F and ε_D are reasonably large.

Example 7 (Breakdown Point Analysis for VMF Distribution on $\mathcal{S}^2$): In this example, we evaluate the breakdown point robustness of r-EWMA, r-CUSUM, and r-MWMASHC using VMF distribution. Figure 4.12 indicates the gross-error sensitivities for r-EWMA, r-CUSUM, and r-MWMASHC for IC and OC conditions with respect to contamination level ϵ. The left panel evaluates the robustness of the proposed methods for IC condition. As it can be seen, increasing the contamination level decreases IC-ARL so that the ratio $\inf_{H_0 \in \mathcal{H}_{0,\epsilon}} \frac{\mathbb{E}_{H_0}^{(\infty)}(T)}{\mathbb{E}_{P}^{(\infty)}(T)}$ approaches 0. From the plot it can be inferred that the proposed methods have a comparable robustness with false-alarm breakdown point $\varepsilon_F = 0.3$.

The right panel compares the robustness of these methods for OC condition. It is observed that increasing the contamination level increases the OC-ARL so that the ratio $\sup_{H_1 \in \mathcal{H}_{1,\epsilon}} \frac{\text{WADD}_{P,H_1}(T)}{\text{WADD}_{P,Q}(T)}$ becomes very large. The results suggest detection-delay breakdown point of $\varepsilon_D = 0.4$ for the proposed process monitoring methods.The obtained values for ε_F and ε_D endorse the robustness of the proposed methods.

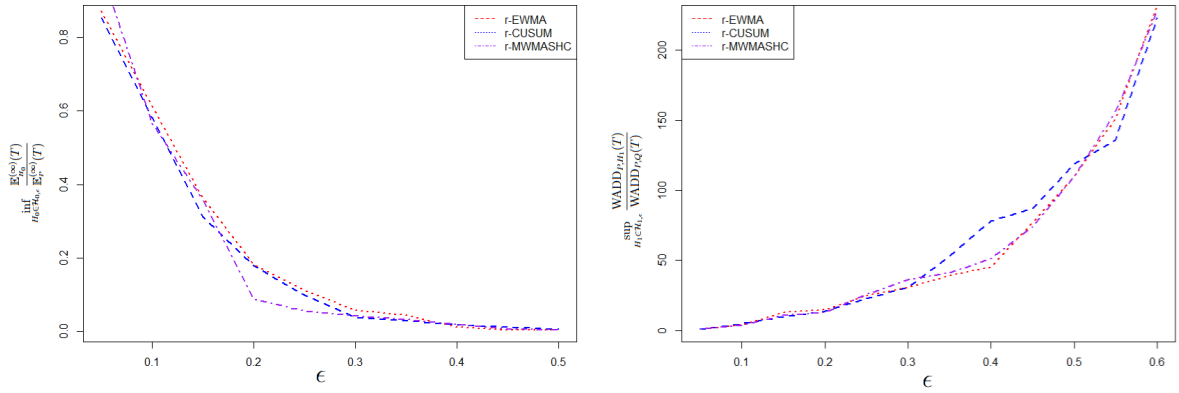


Figure 4.13: Gross-error sensitivities for r-EWMA, r-CUSUM, and r-MWMASHC for IC and OC conditions with different values of ε.

Example 8 (Breakdown Point Analysis for Uniform Distribution on St$(2,3)$): In this example, we evaluate the breakdown point robustness of r-EWMA, r-CUSUM, and r-MWMASHC considering uniform Steifel distribution. Figure 4.13 indicates the gross-error sensitivities for r-EWMA, r-CUSUM, and r-MWMASHC for IC and OC conditions. The left panel evaluates the robustness of the proposed methods for IC condition and the right panel evaluates the robustness of these methods for OC condition.

It is clear that increasing the contamination level ϵ decreases IC-ARL so that the ratio $\inf_{H_0\in\mathcal{H}_{0,\epsilon}} \frac{\mathbb{E}_{H_0}^{(\infty)}(T)}{\mathbb{E}_{P}^{(\infty)}(T)}$ approaches 0. It can be realized that the proposed methods are comparable in terms of robustness with false-alarm breakdown point $\varepsilon_F = 0.4$.

It is also observed that increasing the contamination level increases the OC-ARL so that the ratio $\sup_{H_1\in\mathcal{H}_{1,\epsilon}} \frac{\text{WADD}_{P,H_1}(T)}{\text{WADD}_{P,Q}(T)}$ becomes very large. The results suggest detection-delay breakdown point $\varepsilon_D = 0.4$ for the proposed methods. The obtained values for ε_F and ε_D demonstrate the robustness of r-EWMA, r-CUSUM, and r-MWMASHC.

From the above experiment, we conclude that the rank-depth process monitoring methods are reliable and and can be considered as a robust method for a wide range of manifold spaces with various geometrical properties. Also, our experiments revealed that ε_F and ε_D are effective measures for evaluating the robustness of process monitoring methods under IC and OC conditions.

4.4 Additive Manufacturing

Research on Statistical Process Monitoring for Additive Manufacturing

Additive manufacturing (AM), also called 3D printing, is the process of fabricating 3D products, layer-by-layer under the control of computers. This is done using a collection of digital data (or, if available, a mathematical model) representing the geometry of interest which enables computer aided design (CAD). Unlike conventional manufacturing technologies that involve material removal in their process, AM produces objects layer-by-layer in an additive manner, and thus, there is no waste of material (Tapia and Elwany, 2014). This can benefit industrialists economically. Also, it can prevent life hazards in cases that the material are poisonous, because of no waste during the production (Qian, 2017). Such a feature enables AM to directly produce products with complicated 3D geometries. This makes AM a game changing technology in the manufacturing industry and over the past few decades, this technology has gained a tremendous interest with a wide range of applications, including architecture design (Yin et al., 2018), biomedicine (Melchels et al., 2010), aeronautics and astronautics (Schiller, 2015), and *etc.* Comprehensive reviews on the applications of AM to different key industries and its societal impact can be found in (Guo and Leu, 2013; Huang et al., 2013).

However, in spite of its obvious advantages, today's AM technology suffers from two important flaws, namely unacceptable quality and repeatability of products. This incurs some technical difficulties (especially for applications like aerospace in which product quality is of high importance), and has enforced researchers to seek for methods and techniques to overcome the mentioned problems.

Fortunately, AM is performed under full control of computers, and this enables the use of process controlling and monitoring computer programs to guarantee the quality and repeatability of desired products. In line with such interests, in recent years, the use of statistical monitoring and outlier detection concepts has been investigated by researchers and is continued to be considered as one of the most important research goals of additive manufacturers (Tapia and Elwany, 2014; Colosimo et al., 2018).

The purpose of the current investigation is to probe the potential of some recent statistical tools and concepts to be used for quality assurance in material extrusion AM method.

In particular, control charts are implemented for detecting products which deviate significantly from desired geometry. In general, process monitoring in AM is a complicated task, since data presenting geometries of interest usually have non-euclidean (object data) and functional nature (in the form of planar or spatial). Therefore, more care should be

exerted to make sure the considered process monitoring methods are capable of identifying the sources of variability as well as significant outliers in a set of collected functional / object data (Srivastava and Klassen, 2016).

For the simulation study in this paper, the primary goal is to evaluate the performance of the considered method for detecting significantly deformed 3D geometries produced during a 3D printing process. To do so, the geometries of interest are mathematically modelled using manifold analysis tools and statistical shape analysis methods.

The other important issue which can be considered as the final goal for the problem at hand is the design of efficient process monitoring / controlling algorithms. Due to its importance, different researchers have done remarkable efforts to design effective methods for improving the quality of products in AM. In Cohen and Lipson (2010), a feedback control algorithm was proposed for monitoring the quality of geometrical parts in solid freeform fabrication (SFF) process. Conducted simulation indicated that the proposed method could compensate a broader range of geometrical errors compared to other methods, and also was computationally inexpensive which made it proper for real-time monitoring application.

In Berumen et al. (2010), different aspects and parameters affecting the quality control of products in AM were discussed. In Tapia and Elwany (2014), a review on the advances in process monitoring and control for metal-based AM was performed, which can be viewed as the first comprehensive review on the subject. Based on a throughout analysis, some remarks for future researches were pointed out. In Huang et al. (2014), a statistical predictive modeling was proposed to compensate the geometrical deviations of three-dimensional printed products. Simulation on a number of polyhedron shapes demonstrated the capability of the method for predicting and compensating errors for a wide class of products.

In Lu et al. (2014), a layer-to-layer predictive modelling and feedback control was proposed to control ink-jet 3D printing process. Also, an extension of the algorithm was implemented for two material printing process to enable the printing of complex 3D geometries. Simulation results indicated that the method could reduce the shrinking of parts, and made top layer surface smoother. In Huang et al. (2015), an analytical framework was proposed for the offline optimal compensation of shape shrinkage for 3D printing processes. Experiments indicated that the proposed method can reduce the geometrical error of cylindrical products by an order of magnitude. In Ridwan (2015), the potential of the process monitoring of AM with the goal of improving the quality of products was investigated.

In Jin et al. (2016), an out-of-plane predictive control algorithm was proposed for the offline compensation of geometric errors in AM. Simulation results indicated the accuracy of the proposed control approach. In Everton et al. (2016), a review on the in-situ process

monitoring and metrological methods for metal AM was done. In this review, the suitability of inspection methodologies compatible with AM for identifying typical material discontinuities and failure modes were investigated, and based on that, some remarks for future researches were given.

In Xu et al. (2017), a reverse compensation framework for controlling of shape deformation in AM was proposed. Simulations indicated improvements in the compensation of geometrical errors. In Chua et al. (2017), a review on process monitoring systems for metal AM was performed, and a real-time inspection method together with a closed-loop control algorithm was proposed for quality control in AM.

In Colosimo (2018), a review on modeling and monitoring methods for spatial and image data with application to AM was carried out, and some of the main challenges were discussed. Also solutions to these problems together with some open issues for future research were outlined. In Grasso et al. (2018), a process monitoring method that consists of a feature extraction technique and a data-driven statistical rule was proposed for monitoring powder bed fusion processes. Comparative analysis demonstrated the effectiveness of the proposed method.

In Zang and Qiu (2018*a*), a phase I control chart was proposed for monitoring of spatial surface data from 3D printing. A comparative study was conducted to prove the efficacy of the proposed control chart. In Zang and Qiu (2018*b*), a nonparametric control chart was proposed for phase II monitoring of spatial surfaces produced in 3D printing. The veracity of the proposed method was ascertained by numerical simulation.

As can be inferred, researchers are becoming more and more interested in using standard statistical quality control methods for AM, and it is expected that in near future, more specific investigations will be conducted to fill the gap in the current literature.

Statistics on Manifold for Additive Manufacturing Process

In the current investigation, we considered two types of manifold spaces to represent 3D products. These are the Kendall's shape space Σ_2^k which comprises of a set of points, k-ads, as well as flag manifold $\mathrm{Fl}(v_1, ..., v_k; d)$ representing the evolution of image data in sequential analysis.

Figure 4.14 indicates the considered approaches for representing the 3D products. In the left panel, the volume point cloud representing a part of turbine propeller fan is depicted. The right panel indicates a layer-wise optical tomography image data object. For the point data, we can utilize k-ads shape space Σ_2^k, and for the image data, we use flag manifold

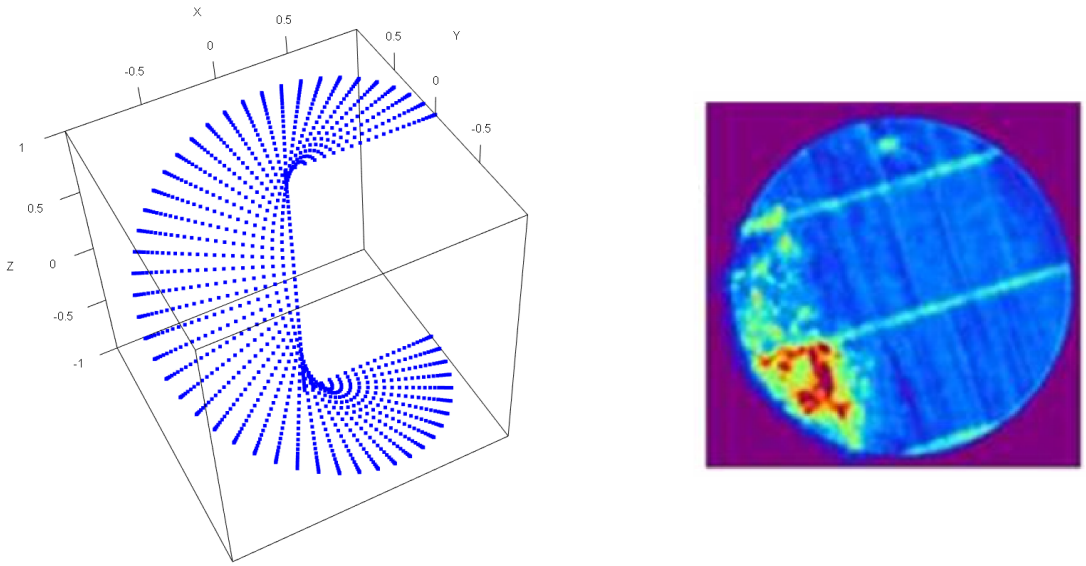


Figure 4.14: Shape point clouds representing a part of turbine propeller fan, and illustration of a layer-wise optical tomography image data object.

$\mathrm{Fl}(v_1, ..., v_k; d)$ which models the evolution of k nested subspaces of d principal component coordinates of the image data. We use these manifold spaces for the simulation studies considered in this section. The depth value can be calculated using the standard distances on these two manifold spaces $\mathcal{M}$, from which the ranks are calculated and the control charts are used for monitoring the quality of products.

For AM applications, given that products are produced layer-by-layer in an additively fashion, it is important to consider dependency between these produced layers. This is because any deformation, deflection, or environmental factor which affects the quality of a layer of the product can also affect the quality of other layers. To model this phenomenon, we can view the production process as a stochastic process $\mathcal{G}(t)$ so that at each time-point t, a layer is produced according to distribution P.

Suppose that the intrinsic dimension of the manifold $\mathcal{M}$ is d. The $\mathbb{R}^d$-valued logarithm process is defined as $L_{\mathcal{G}}(t) = \log_{\mu(t)} \mathcal{G}(t)$ for all $t \in \mathcal{T}$. It holds that

$$\mathbb{E}L_{\mathcal{G}}(\cdot) = 0 \quad , \quad \mathbb{E}\|L_{\mathcal{G}}(t)\|_2^2 = \mathbb{E}d(\mu(t), \mathcal{G}(t)) < \infty \ ,$$

for every $t \in \mathcal{T}$. The covariance function is defined as

$$\Gamma(s,t) = \mathbb{E}\left[L(s)L(t)^\top\right] \ , \quad s,t \in \mathcal{T} \ .$$

Assumption 14. *The stochastic process $\mathcal{G}(t)$ is second-order and the Fréchet mean function exists.*

Assumption 15. *The domain $\mathcal{T}$ is a compact interval, and the sample path $\mathcal{G}(t)$ is twice continuously differentiable for $t \in \mathcal{T}$ almost surely.*

Theorem 34. *Under Assumptions 14 and 15, it holds that*

$$\sup_{s,t\in\mathcal{T}} \|\widehat{\Gamma}(s,t) - \Gamma(s,t)\|_F^2 = o_p(1) \ .$$

Theorem 35. *Assume that random objects are sampled from stochastic process $\mathcal{G}(t)$ with covariance function $\Gamma(s,t)$ on compact time interval $\mathcal{T}$ defined on manifold $\mathcal{M}$. Under Assumptions 1-3, 14 and 15, the control charts $\widehat{\boldsymbol{S}}_t$ and $\widehat{\boldsymbol{\Lambda}}_t$ based on empirical ranks-depths $\widehat{R}_t$ are consistent estimators of control charts based on theoretical rank-depths R_t, i.e.*

$$\widehat{\mathbf{S}}_t - \mathbf{S}_t = O_p\left(n^{-1/2}(d\log d)^{1/2}\right) \ ,$$
$$\widehat{\boldsymbol{\Lambda}}_t - \boldsymbol{\Lambda}_t = O_p\left(n^{-1}(d\log d)\right) \ ,$$

where d is the intrinsic dimension of the manifold.

Simulation Case Studies for Additive Manufacturing Process

Example 9 (Photodiode AM Image Dataset): In this example, we consider photodiode light intensity and optical tomography dataset which comprises of images taken from a fabricated cylinder-shape object. The images are taken in real-time from each layer of product using a camera mounted outside 3D printing closure. Figure 4.15 indicates two different off-axial camera mounting strategies (Taherkhani, 2022).

Figure 4.16 depicts the process monitoring algorithm for image data taken from each layer of product. Products are built layer-by-layer inside 3D printing chamber while off-axial camera takes images in real-time from each layer. These images are considered as object data generated from an unknown distribution on Hermitian matrix manifold space for which standard tools such as tangent space, logarithmic and exponential map are available. Thereafter, we use geodesic depth to calculate the depth of each data and rank them accordingly. Control chart values are calculated in real-time using the proposed rank-depth process monitoring methods, and if these values exceed the control limit, a fault alarm is triggered.

Figure 4.17 depicts the image taken from a layer of healthy product along with a number of artificial faulty products. The fault appears in the form of seams, holes, and

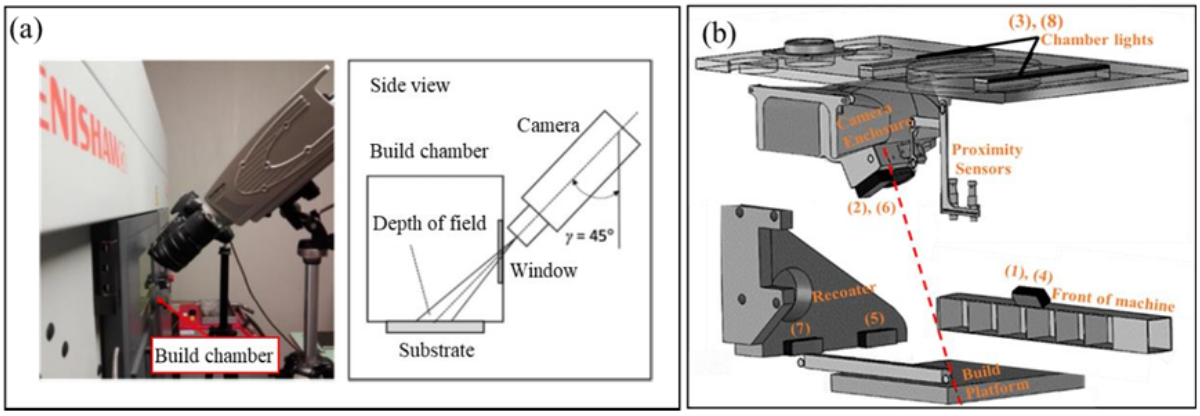


Figure 4.15: Two different off-axial camera mounting strategies (Taherkhani, 2022).

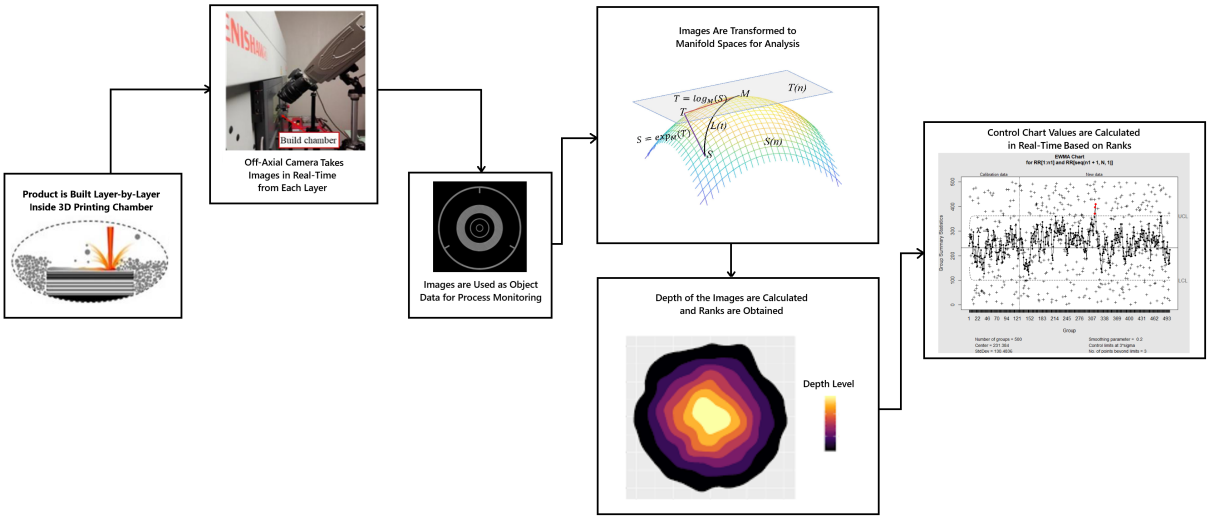


Figure 4.16: Process monitoring algorithm for image data taken from each layer of product.

cuts. The plot also illustrates how a control chart monitors the quality of images taken from photodiode AM process in Phase-II. Once a faulty product is produced, the values of these control charts violate the control limits and the process control method detects the occurred fault.

Figure 4.18 depicts the results of simulation for 100 images of size 1369×1320 pixels where the last 50 images are taken from layers of faulty products. We generate faulty images of various defect types in Phase-II using uniform distribution. Also, to evaluate

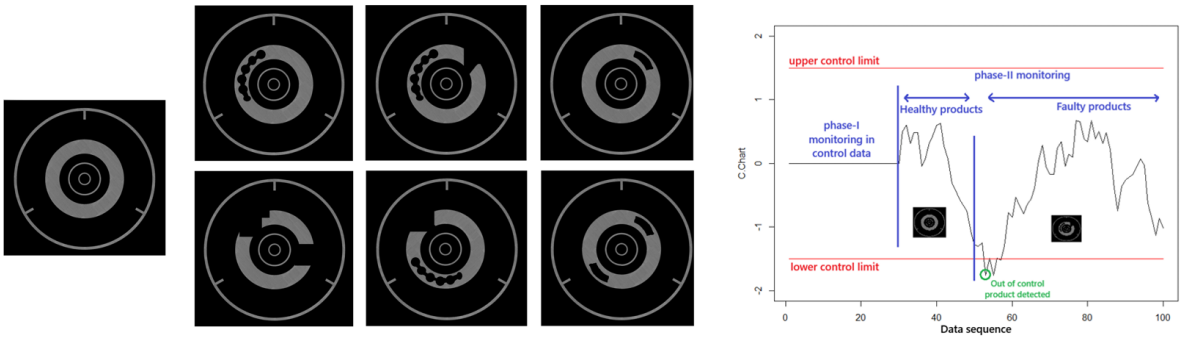


Figure 4.17: Illustration of healthy product along with a number of faulty products obtained in a photodiode additive manufacturing process. Right panel shows how a control chart monitors the quality of images taken from photodiode AM process in Phase-II.

Table 4.9: IC-ARL and OC-ARL values for IC and OC data obtained over 5000 simulation.

		IC-ARL		**OC-ARL**	
Methods		**avg.**	**std.**	**avg.**	**std.**
r-EWMA	$\lambda = 0.1$	374.3	231.8	17.13	10.36
r-CUSUM	$k = 0.2$	372.4	252.6	17.16	11.31
r-MWMASHC	$m = 10$	375.7	245.7	21.04	13.63

the robustness of the proposed methods for distinguishing special causes of variation from common causes of variation, a stochastic perturbation process is considered which adds uniform noises to the grayscale values of original pixel images for both healthy and faulty data. Top left panel indicates the depth values for the object data. As it can be seen, the depth distribution changes for the last 50 images. Clearly, introducing faults make the images far from the median of the distribution as the minimum achieved depth values are decreased compared to the first 50 depth values. This is also reflected in the obtained control chart values, as the values exceed the control limits for r-EWMA, r-CUSUM, and r-MWMASHC. For developing these control charts, we used flag manifold space $\mathrm{Fl}(v_1, ..., v_k; d)$ which models the nested subspaces of the coordinate components of these image data in sequential analysis. Table 4.9 reports the IC-ARL and OC-ARL values for r-EWMA, r-CUSUM, and r-MWMASHC over 5000 simulation runs. As it can be seen, for the IC condition, all of these methods can control the ARL around the nominal r-EWMA, r-CUSUM, and r-MWMASHC ARL_0 value 370. For the OC condition, the obtained OC-ARL values are relatively small and comparable for r-EWMA, r-CUSUM, and r-MWMASHC. Based on the results of simulation, the best performance belongs to r-EWMA and r-CUSUM which have

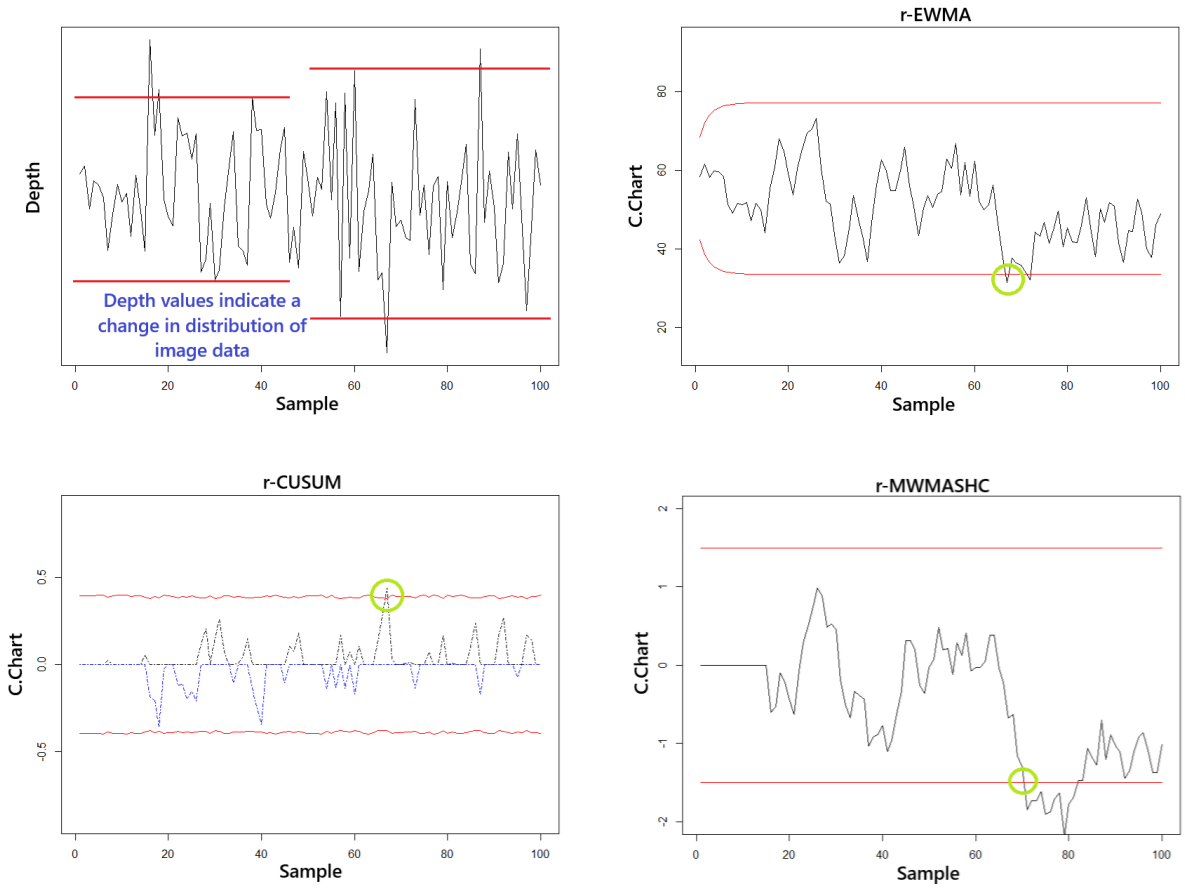


Figure 4.18: First panel indicates the change in the distribution of depths for image data. Next panels indicate the performance of r-EWMA, r-CUSUM, and r-MWMASHC control charts in Phase-II.

smaller OC-ARL compared to r-MWMASHC, while their IC-ARL values are comparable.

Example 10 (Thingi10K Dataset): The purpose of this simulation is to evaluate the efficacy of the proposed control charts for monitoring AM products. Thingi10K (Zhou and Jacobson, 2016) is one of the largest open access databases including AM products with a wide range of features. For the current simulation, two geometries known as spiral bevel gear (model 34785), and coin loading tube (model 36090) are extracted for analysis. Figure 4.19 depicts the considered geometries. Sampling frequency for spiral bevel gear and coin loading tube cases are 168 and 58, respectively.

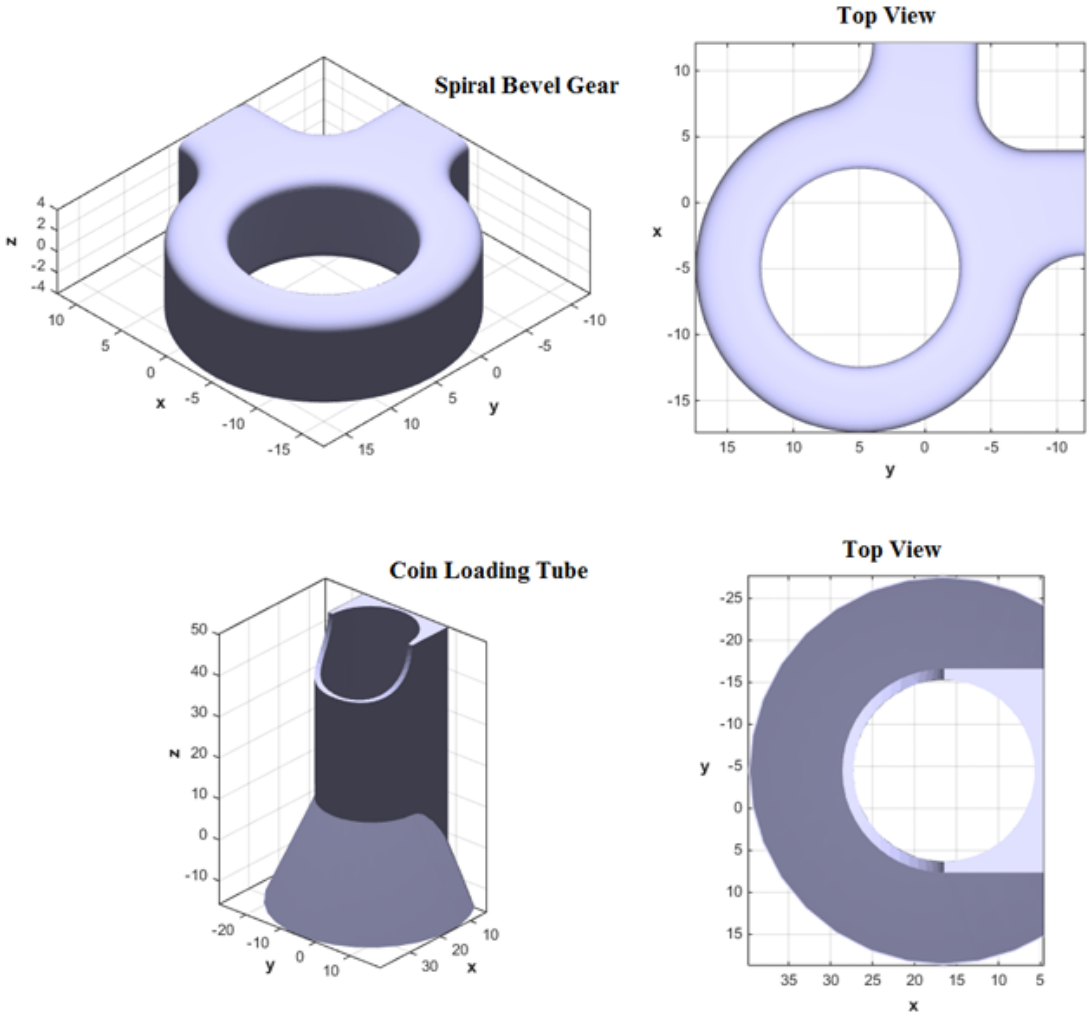


Figure 4.19: Different views of spiral bevel gear and coin loading tube geometries.

Parametric Control Charts: We propose parametric SHC and EWMA control charts which use functional representation of contour shapes and works based on the decomposition of amplitude and phase components. The details of derivation of the proposed control charts is given in Appendix A in the end of chapter. Fourier model together with coordinate functions $x(t)$ and $y(t)$ ($t \in [0, 1]$ is the shape parameter) are used for approximating contour functions. Figure 4.20 depicts the fitted functionals in the $x(t)$ and $y(t)$ coordinates. By modifying the coefficients of trained Fourier model, the following gener-

Table 4.10: Computed UCL values. Numbers in parentheses are the iteration until convergence.

		Spiral bevel gear			Coin loading tube		
	Methods	**N = 100**	**N = 200**	**N = 500**	**N = 100**	**N = 200**	**N = 500**
Amplitude	**SHC** ($m = 2$)	0.410 (45)	0.336 (185)	0.390 (-)	0.365 (76)	0.439 (169)	0.446 (-)
	SHC ($m = 5$)	0.366 (187)	0.318 (140)	0.342 (-)	0.311 (44)	0.374 (152)	0.414 (30)
	SHC ($m = 10$)	0.340 (19)	0.292 (24)	0.314 (53)	0.286 (19)	0.353 (208)	0.394 (-)
	SHC ($m = 20$)	0.311 (89)	0.273 (26)	0.297 (-)	0.265 (37)	0.329 (285)	0.372 (14)
	EWMAC ($\lambda = 0.25$)	0.341 (11)	0.302 (153)	0.329 (-)	0.306 (49)	0.365 (51)	0.401 (101)
	EWMAC ($\lambda = 0.50$)	0.386 (212)	0.320 (47)	0.360 (-)	0.341 (117)	0.415 (49)	0.430 (-)
	EWMAC ($\lambda = 0.75$)	0.399 (190)	0.330 (268)	0.385 (229)	0.361 (30)	0.437 (243)	0.441 (260)
	EWMAC ($\lambda = 0.90$)	0.408 (37)	0.335 (40)	0.394 (179)	0.368 (252)	0.458 (56)	0.449 (191)
Phase	**SHC** ($m = 2$)	0.064 (91)	0.062 (34)	0.060 (46)	0.219 (46)	0.244 (-)	0.220 (-)
	SHC ($m = 5$)	0.056 (30)	0.053 (-)	0.053 (112)	0.168 (32)	0.222 (89)	0.203 (60)
	SHC ($m = 10$)	0.051 (114)	0.048 (178)	0.049 (-)	0.149 (32)	0.199 (43)	0.183 (20)
	SHC ($m = 20$)	0.047 (61)	0.045 (104)	0.045 (187)	0.132 (16)	0.183 (212)	0.167 (-)
	EWMAC ($\lambda = 0.25$)	0.054 (208)	0.051 (17)	0.049 (-)	0.157 (54)	0.201 (155)	0.191 (-)
	EWMAC ($\lambda = 0.50$)	0.060 (80)	0.057 (95)	0.056 (244)	0.192 (112)	0.224 (176)	0.210 (139)
	EWMAC ($\lambda = 0.75$)	0.064 (24)	0.061 (18)	0.059 (82)	0.209 (35)	0.241 (37)	0.218 (-)
	EWMAC ($\lambda = 0.90$)	0.065 (29)	0.062 (58)	0.060 (-)	0.219 (93)	0.243 (191)	0.220 (115)

ating mechanism is developed to collect an IC dataset including geometries with random amplitude components

$$\begin{cases} x^{(i)}(t) = x(t) \\ y^{(i)}(t) = A_0 + \sum_{k=1}^{K} (A_k + u_1)cos(kt) + (B_k + u_1)sin(kt) \end{cases},$$

for $i = 1, 2, ..., n$, where A_0, A_k and B_k are the coefficients of fitted Fourier series, and $\{u_1, u_2\}$ are uniformly distributed random variables. The number of basis K for approximating spiral bevel gear, and coin loading tube are 40, and 25, respectively. Also, for spiral bevel gear and coin loading tube cases, $\{u_1, u_2\} \sim Unif(0, 0.05)$, and $\{u_1, u_2\} \sim Unif(0, 0.01)$, respectively. Phase variation is also incorporated into the above Fourier model using warping function $\gamma_i(t) = t + u_3 t(t-1)$ for $i = 1, 2, ..., N$, where $u_3 \sim Unif(-0.05, 0.05)$, and setting $y^{(i)}(t) = y^{(i)}(\gamma_i(t))$.

For AM data, the first 100 functions (Phase-I) are used to calculate the medians and UCL, and then the next incoming N IC data are used to adaptively calculate the bootstrap UCL. The numerical experiments show that for AM data, due to the mentioned complexity of the monitoring problem, the performance of both of these charts degenerate if the information from newly available data during sequential process is not used. Table 4.10 reports the calculated UCL values for the parametric SHC and EWMAC control charts. Bold values indicate the best achieved IC-ARL values for SHC and EWMAC.

Table 4.11 reports the IC-ARL values for SHC and EWMAC for in-control simulation scenario. The values are reported for different values of Phase-II observations, i.e. $n = \{100, 200, 500\}$. Note that for Phase-I analysis of all of these simulation scenarios, we

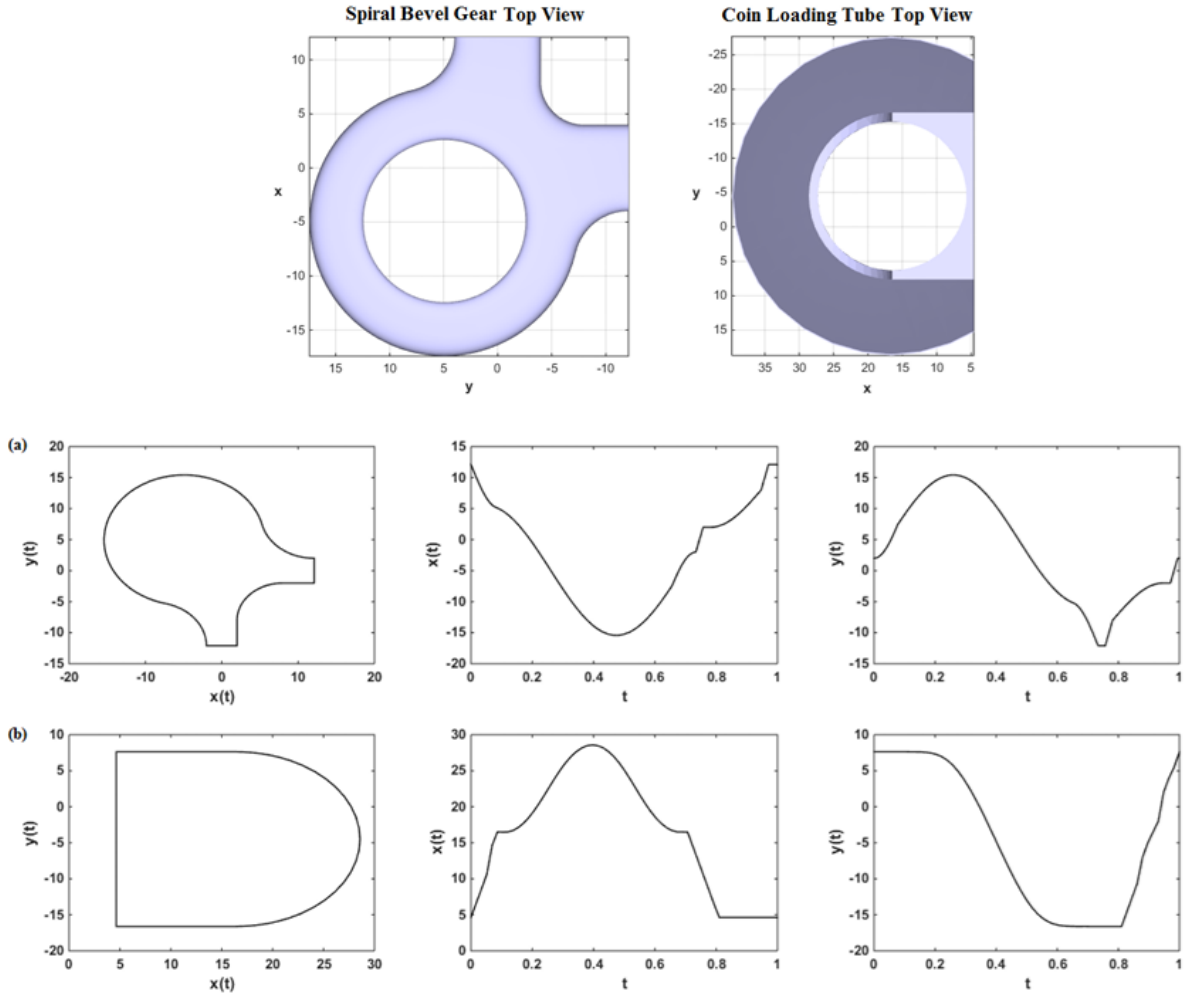


Figure 4.20: Fitted Fourier series to the contours of spiral bevel gear and coin loading tube.

considered 100 in-control data. It can be seen that the performance of both parametric control charts are comparable, and under the considered setting, they achieve the nominal ARL_0 values of 200 and 370 in the sense that the obtained IC-ARL values are either kept at this level or is larger. This behaviour is valid for both spiral bevel gear and coin loading tube objects, demonstrating the generalization capability of these parametric methods.

Table 4.12 reports the OC-ARL values for SHC and EWMAC for out-of-control simulation scenario. Clearly, both of these methods perform a good job and yield small OC-ARL

Table 4.11: IC-ARL values for in-control simulation scenario. Numbers in parentheses are standard deviations over 5000 simulation runs.

		Spiral bevel gear			Coin loading tube		
	Methods	**N = 100**	**N = 200**	**N = 500**	**N = 100**	**N = 200**	**N = 500**
Amplitude	**SHC** ($m = 2$)	100 (0)	173.51 (55.06)	500 (0)	**96.27 (17.29)**	199.64 (3.60)	493.51 (51.96)
	SHC ($m = 5$)	**92.09 (26.99)**	**190.07 (43.50)**	**480.10 (92.38)**	93.08 (25.35)	**189.44 (42.66)**	**475.05 (109.30)**
	SHC ($m = 10$)	82.20 (38.18)	162.30 (78.23)	422.41 (179.95)	84.20 (36.38)	170.17 (71.36)	445.11 (156.91)
	SHC ($m = 20$)	68.36 (46.35)	142.34 (90.67)	360.45 (224.90)	77.98 (41.02)	132.51 (94.50)	385.31 (210.90)
	EWMAC ($\lambda = 0.25$)	**99.13 (6.45)**	200 (0)	500 (0)	100 (0)	199.85 (1.50)	500 (0)
	EWMAC ($\lambda = 0.50$)	100 (0)	**187.90 (37.63)**	500 (0)	100 (0)	199.85 (1.50)	500 (0)
	EWMAC ($\lambda = 0.75$)	100 (0)	137.92 (68.77)	500 (0)	98.05 (10.07)	**196.31 (20.83)**	500 (0)
	EWMAC ($\lambda = 0.90$)	100 (0)	89.49 (67.75)	500 (0)	**95.28 (17.18)**	198.39 (14.66)	500 (0)
Phase	**SHC** ($m = 2$)	100 (0)	200 (0)	499.96 (0.40)	**90.55 (26.82)**	200 (0)	**475.88 (93.79)**
	SHC ($m = 5$)	**94.07 (23.59)**	**176.82 (62.46)**	**475.05 (109.30)**	87.49 (31.85)	200 (0)	470.07 (119.06)
	SHC ($m = 10$)	91.10 (28.44)	140.72 (89.79)	445.87 (154.91)	90.10 (29.84)	155.97 (82.73)	427.08 (175.45)
	SHC ($m = 20$)	81.26 (38.89)	125.15 (95.84)	355.41 (227.38)	69.43 (45.83)	140.38 (91.53)	407.15 (193.58)
	EWMAC ($\lambda = 0.25$)	100 (0)	**196.03 (23.47)**	**493.30 (47.02)**	97.83 (11.86)	199.73 (2.70)	496.90 (31.00)
	EWMAC ($\lambda = 0.50$)	100 (0)	198.78 (12.20)	500 (0)	**95.95 (16.54)**	**196.83 (22.30)**	500 (0)
	EWMAC ($\lambda = 0.75$)	100 (0)	200 (0)	496.51 (34.90)	90.97 (23.28)	200 (0)	495.16 (35.35)
	EWMAC ($\lambda = 0.90$)	100 (0)	199.26 (7.40)	500 (0)	90.91 (23.59)	200 (0)	**491.59 (49.89)**
Joint	**SHC** ($m = 2$)	99.13 (6.45)	173.51 (55.06)	500 (0)	96.27 (17.29)	199.49 (3.88)	493.51 (51.96)
	SHC ($m = 5$)	92.08 (26.99)	178.53 (55.30)	480.10 (92.38)	93.08 (25.35)	189.43 (42.66)	475.05 (109.30)
	SHC ($m = 10$)	82.20 (38.18)	**117.92 (82.06)**	422.41 (179.95)	82.25 (36.92)	167.76 (72.28)	445.11 (156.91)
	SHC ($m = 20$)	**68.36 (46.35)**	66.70 (71.28)	**360.45 (224.90)**	**75.12 (41.64)**	**130.90 (94.48)**	**385.31 (210.90)**
	EWMAC ($\lambda = 0.25$)	100 (0)	196.03 (23.47)	493.26 (47.02)	89.01 (28.18)	199.73 (2.70)	472.78 (98.02)
	EWMAC ($\lambda = 0.50$)	94.07 (23.59)	176.82 (62.46)	475.05 (109.30)	86.65 (32.61)	196.83 (22.30)	470.07 (119.06)
	EWMAC ($\lambda = 0.75$)	91.10 (28.44)	140.72 (89.79)	445.87 (154.91)	82.65 (34.31)	155.97 (82.73)	422.24 (176.97)
	EWMAC ($\lambda = 0.90$)	**81.26 (38.89)**	**125.15 (95.84)**	**355.41 (227.38)**	**63.28 (45.73)**	**140.38 (91.53)**	**398.74 (195.93)**

Table 4.12: OC-ARL values for out-of-control simulation scenario. Numbers in parentheses are standard deviations over 5000 simulation runs.

		Spiral bevel gear			Coin loading tube		
	Methods	**n = 100**	**n = 200**	**n = 500**	**n = 100**	**n = 200**	**n = 500**
Amplitude	**SHC** ($m = 2$)	**1.09 (0.32)**	**2.02 (9.89)**	**1.04 (0.19)**	**1.75 (4.90)**	**4.38 (16.91)**	**6.35 (34.98)**
	SHC ($m = 5$)	2.08 (6.88)	8.99 (26.97)	8.54 (42.67)	3.16 (8.31)	7.70 (23.45)	11.55 (48.92)
	SHC ($m = 10$)	6.98 (15.97)	10.01 (28.44)	23.49 (71.59)	8.61 (16.80)	18.95 (36.88)	31.92 (80.94)
	SHC ($m = 20$)	15.80 (22.50)	27.84 (44.10)	70.78 (112.32)	12.88 (19.83)	31.15 (44.23)	52.06 (99.47)
	EWMAC ($\lambda = 0.25$)	1.10 (0.33)	1.04 (0.19)	1.05 (0.21)	1.32 (0.56)	5.40 (19.41)	**1.41 (0.58)**
	EWMAC ($\lambda = 0.50$)	1.09 (0.32)	**1.02 (0.14)**	**1.03 (0.17)**	1.19 (0.56)	**3.32 (13.89)**	3.73 (24.88)
	EWMAC ($\lambda = 0.75$)	1.09 (0.32)	**1.02 (0.14)**	3.52 (24.89)	1.14 (0.44)	7.17 (23.57)	21.04 (67.85)
	EWMAC ($\lambda = 0.90$)	**1.05 (0.26)**	**1.02 (0.14)**	15.96 (59.42)	**1.09 (0.28)**	15.03 (34.45)	38.43 (89.32)
Phase	**SHC** ($m = 2$)	**2.01 (6.89)**	**2.00 (9.90)**	**5.99 (35.03)**	**6.65 (10.38)**	**5.49 (10.11)**	**5.25 (3.85)**
	SHC ($m = 5$)	3.00 (9.64)	3.00 (13.92)	8.50 (42.68)	7.57 (12.60)	22.15 (37.90)	9.76 (34.51)
	SHC ($m = 10$)	6.49 (15.37)	13.89 (33.45)	15.98 (59.42)	9.74 (13.53)	32.28 (43.51)	30.74 (73.46)
	SHC ($m = 20$)	14.78 (22.07)	29.76 (45.11)	55.82 (103.64)	19.69 (20.00)	51.98 (47.32)	38.44 (78.53)
	EWMAC ($\lambda = 0.25$)	1.53 (4.89)	2.99 (13.92)	1.03 (0.17)	3.66 (1.19)	6.30 (16.59)	5.17 (1.49)
	EWMAC ($\lambda = 0.50$)	**1.02 (0.14)**	**1.99 (9.90)**	3.50 (24.89)	3.56 (4.89)	3.97 (2.01)	5.49 (2.83)
	EWMAC ($\lambda = 0.75$)	**1.02 (0.14)**	4.95 (19.50)	3.50 (24.89)	3.68 (2.23)	3.48 (2.16)	4.38 (2.63)
	EWMAC ($\lambda = 0.90$)	1.51 (4.90)	9.90 (28.47)	**1.01 (0.10)**	**3.05 (2.11)**	**2.54 (1.97)**	**3.14 (2.30)**
Joint	**SHC** ($m = 2$)	**1.03 (0.17)**	**1.01 (0.17)**	**1.02 (0.14)**	**1.28 (0.56)**	**1.47 (0.73)**	**1.41 (0.65)**
	SHC ($m = 5$)	1.04 (0.19)	1.02 (0.14)	1.04 (0.19)	2.70 (6.83)	3.86 (13.84)	1.74 (0.90)
	SHC ($m = 10$)	1.59 (4.90)	2.01 (9.89)	1.04 (0.19)	5.50 (12.33)	8.49 (23.28)	12.45 (48.75)
	SHC ($m = 20$)	6.95 (15.98)	11.94 (31.11)	20.97 (67.87)	7.70 (14.28)	26.45 (41.56)	20.77 (63.25)
	EWMAC ($\lambda = 0.25$)	1.04 (0.19)	1.01 (0.10)	1.03 (0.17)	1.32 (0.56)	2.50 (9.87)	1.41 (0.58)
	EWMAC ($\lambda = 0.50$)	**1.02 (0.14)**	1.00 (0)	**1.01 (0.10)**	1.19 (0.50)	1.40 (0.72)	**1.28 (0.62)**
	EWMAC ($\lambda = 0.75$)	**1.02 (0.14)**	**0.99 (0.10)**	**1.01 (0.10)**	1.14 (0.44)	1.32 (0.61)	1.47 (1.64)
	EWMAC ($\lambda = 0.90$)	**1.02 (0.14)**	1.00 (0.14)	**1.01 (0.10)**	**1.09 (0.28)**	**1.30 (0.88)**	1.48 (1.56)

values. In most of these simulation cases, the OC-ARL value remains less than 10 regardless of the considered sample size for Phase-II analysis. This promising observation together with the results obtained in IC simulation demonstrate the efficacy of the proposed parametric control charts for this simulation study.

It is worth mentioning that these parametric control charts work for cases where the

Table 4.13: IC-ARL values for in-control simulation scenario. Numbers in parentheses are standard deviations over 5000 simulation runs.

		Spiral bevel gear			Coin loading tube		
Methods		**n = 100**	**n = 200**	**n = 500**	**n = 100**	**n = 200**	**n = 500**
r-EWMA	$\lambda = 0.05$	100 (0)	175.45 (97.21)	340.89 (148.63)	93.50 (56.29)	190.14 (123.65)	360.80 (142.66)
r-EWMA	$\lambda = 0.1$	**98.83 (3.39)**	187.30 (131.01)	373.45 (163.40)	93.55 (47.56)	194.90 (153.9)	399.35 (104.39)
r-EWMA	$\lambda = 0.2$	100 (0)	165.85 (94.78)	**456.65 (186.79)**	**96.92 (13.75)**	189.55 (119.04)	462.58 (183.88)
r-CUSUM	$k = 0.1$	100 (0)	174.74 (74.52)	353.83 (182.42)	100 (0)	**195.24 (101.44)**	432.15 (221.83)
r-CUSUM	$k = 0.2$	100 (0)	184.92 (91.17)	363.82 (191.25)	100 (0)	172.13 (112.93)	442.92 (253.31)
r-CUSUM	$k = 0.3$	100 (0)	**192.81 (88.34)**	342.81 (173.82)	100 (0)	174.83 (121.29)	**463.34 (227.21)**
r-MWMASHC	$m = 5$	96.34 (10.32)	181.83 (82.32)	363.83 (170.93)	91.24 (8.91)	186.84 (124.21)	386.93 (121.53)
r-MWMASHC	$m = 10$	92.19 (11.42)	186.81 (121.25)	353.18 (191.32)	93.42 (14.73)	175.83 (120.72)	412.83 (194.82)
r-MWMASHC	$m = 15$	94.75 (9.73)	175.72 (102.31)	362.82 (179.42)	91.74 (8.42)	183.82 (112.82)	401.41 (189.43)

Table 4.14: OC-ARL values for out-of-control simulation scenario. Numbers in parentheses are standard deviations over 5000 simulation runs.

		Spiral bevel gear			Coin loading tube		
Methods		**n = 100**	**n = 200**	**n = 500**	**n = 100**	**n = 200**	**n = 500**
r-EWMA	$\lambda = 0.05$	3.75 (0.91)	2.55 (0.82)	1.85 (0.58)	1.90 (2.13)	**1.20 (2.90)**	1.40 (2.81)
r-EWMA	$\lambda = 0.1$	2.80 (0.65)	2.40 (3.87)	1.45 (0.82)	1.75 (1.93)	1.65 (1.88)	1.70 (2.10)
r-EWMA	$\lambda = 0.2$	2.20 (0.83)	1.50 (1.90)	**1.40 (0.68)**	2.15 (2.22)	1.75 (1.50)	1.80 (1.18)
r-CUSUM	$k = 0.1$	2.43 (1.92)	3.72 (2.21)	2.14 (1.63)	1.53 (0.82)	1.45 (0.74)	1.63 (2.84)
r-CUSUM	$k = 0.2$	1.42 (1.04)	2.13 (2.13)	2.84 (2.12)	**1.42 (1.32)**	1.84 (0.82)	2.14 (1.83)
r-CUSUM	$k = 0.3$	1.62 (2.42)	1.83 (1.93)	2.23 (2.73)	2.42 (1.42)	2.13 (1.42)	1.85 (0.86)
r-MWMASHC	$m = 5$	**1.03 (0.52)**	2.42 (1.32)	1.84 (2.14)	1.53 (0.56)	1.63 (0.74)	3.12 (2.22)
r-MWMASHC	$m = 10$	1.84 (2.13)	**1.43 (0.87)**	2.83 (1.84)	1.92 (2.16)	1.92 (1.31)	1.42 (2.10)
r-MWMASHC	$m = 15$	2.13 (2.63)	3.19 (2.63)	1.80 (1.21)	1.73 (1.83)	1.88 (1.29)	**1.12 (0.26)**

obtained eigenvalues from spectral analysis of phase and amplitude functional components follow normal distribution or they are approximately normally distributed.

Therefore, it is also important to consider nonparametric control charts for such analysis given the fact that nonparametric control charts do not consider any assumption for the distribution from which these object data are sampled.

Nonparametric Control Charts: To investigate the effectiveness of nonparametric methods, we consider r-EWMA, r-CUSUM, and r-MWMASHC control charts and compare their results with the parametric ones. Figure 4.21 depicts the k-ads Σ_2^k shape space representation of spiral bevel gear and coin loading tube.

Table 4.13 reports the IC-ARL values for in-control simulation scenario. Clearly, the results are comparable with those of parametric control charts, indicating the competence of nonparametric method, as they do not assume distribution for data. Moreover, it can be seen that the performance of all three nonparametric methods are comparable, with r-CUSUM taking the first place, followed by r-EWMA and r-MWMASHC.

Table 4.14 reports the OC-ARL values for out-of-control condition. The OC-ARL values are small and comparable with those of parametric control charts. Overall, nonparametric methods perform very well as they control the IC-ARL values at the nominal level and yield small OC-ARL values.

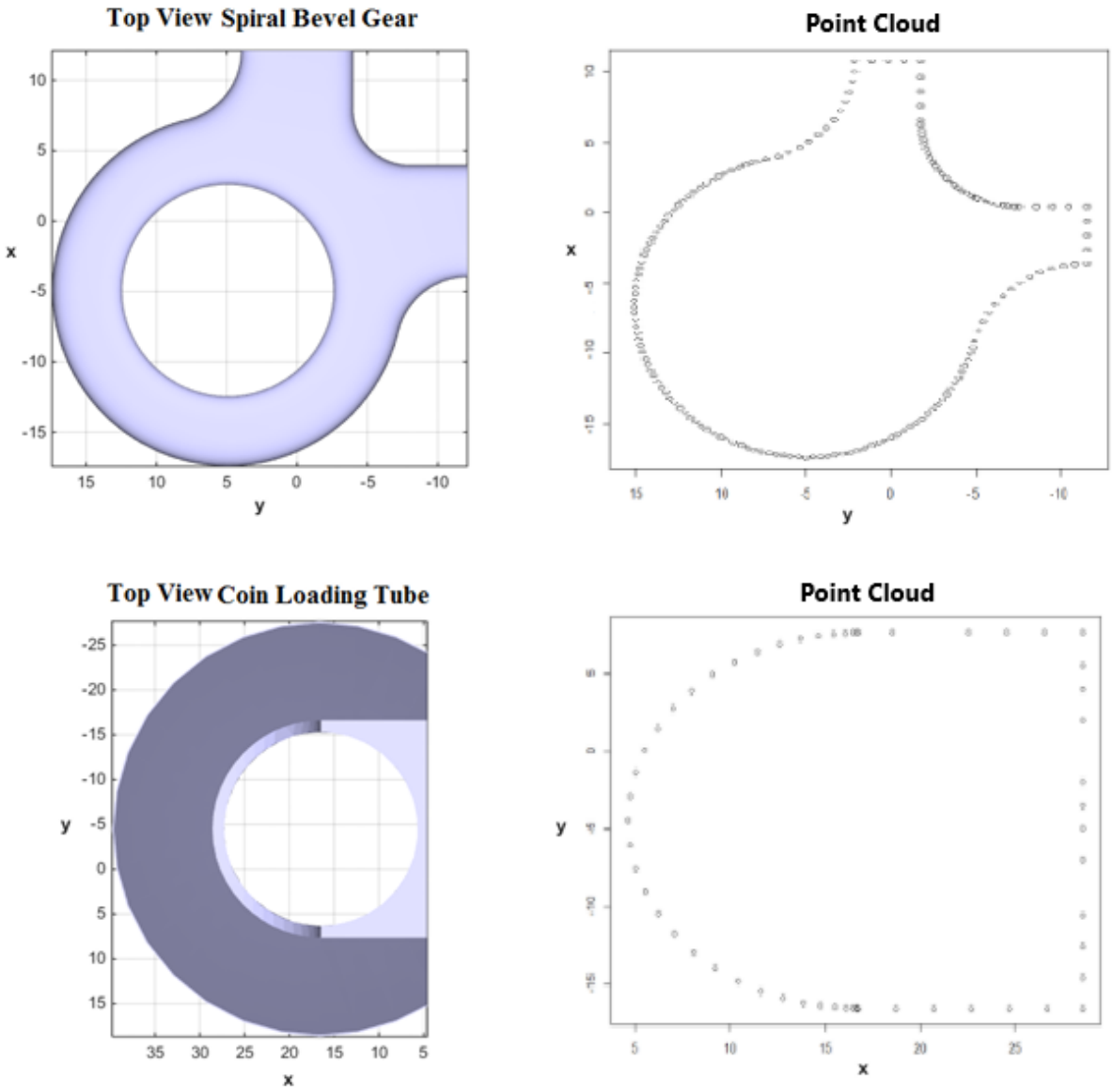


Figure 4.21: Shape point clouds representing spiral bevel gear and coin loading tube.

Figure 4.22 depicts a number of detected disformed products with contour roughness defect for spiral bevel gear and coin loading tube. It is inferred that the considered process monitoring methods are successful in detecting outlier products, and therefore, have poten-

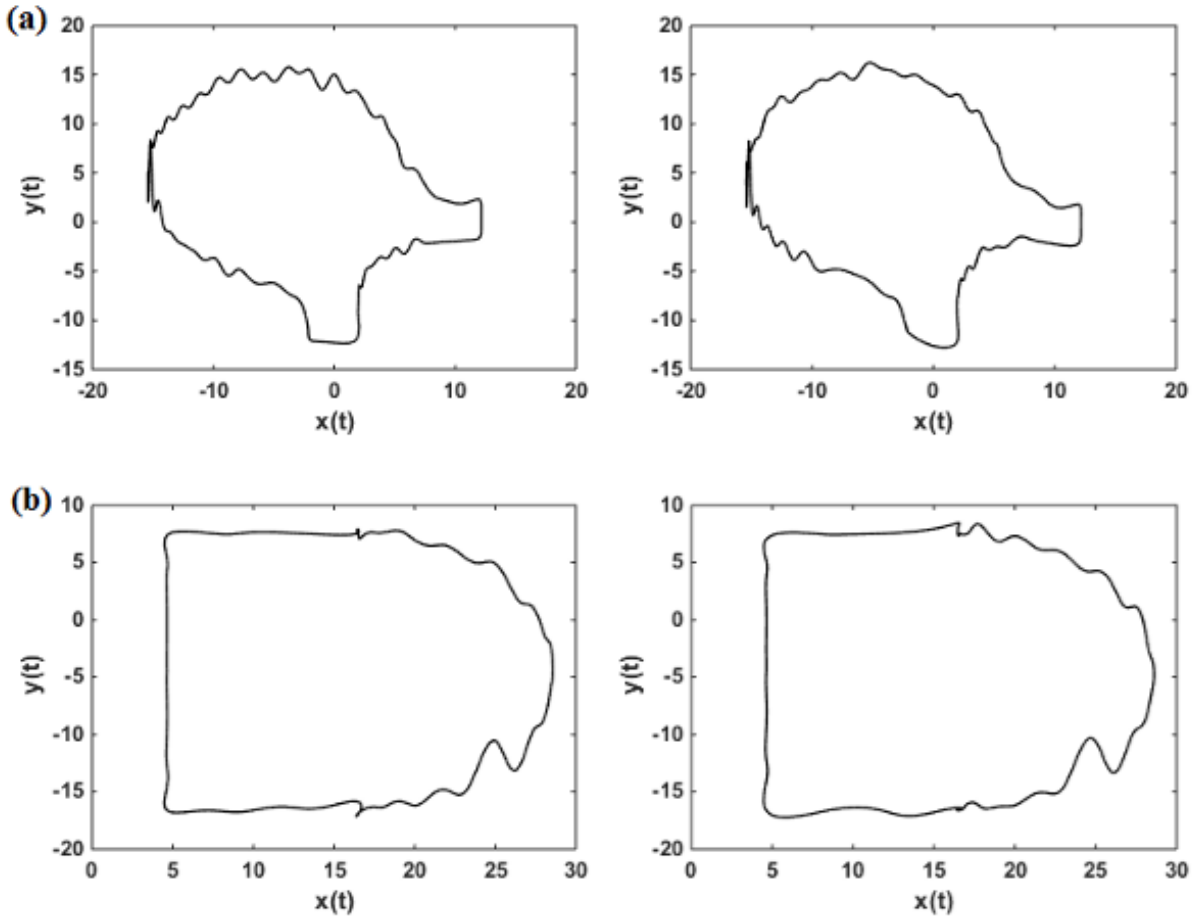


Figure 4.22: Schematic illustration of a number of detected products with defective contours, (a) spiral bevel gear, (b) coin loading tube.

tial to be used for monitoring the quality of the surface of shapes with complex geometrical features.

It can be concluded that rank-depth control charts have the advantage in that they do not require any assumption for the data which makes the analysis much easier, since finding the appropriate distribution for manifold data can be a formidable task given the complexity of the underlying manifold spaces $\mathcal{M}$.

Example 11 (Shape Space $\mathbf{\Sigma}_2^k$ Embedded in Complex Projective Space): In this example, we present an embedding framework for shape objects on $\mathbf{\Sigma}_2^k$, which we will then use for analysis. Consider $k > 2$ points on the complex plane represented by complex numbers $z_j = x_j + iy_j, 1 \leq j \leq k$. The shape of the object can be viewed as equivalent class under removing translation, rotation and scale components. We remove the translation component using $\mathbf{z} - \langle \mathbf{z} \rangle$ where $\langle \mathbf{z} \rangle = \frac{1}{k} \sum_{i=1}^{k} (x_j, y_j)$. The rotation by angle θ and scale by factor r are removed by multiplying $\mathbf{z} - \langle \mathbf{z} \rangle$ by $\gamma = r \exp(i\theta)$. Clearly, the shape can be expressed as lines passing through $\mathbf{z} - \langle \mathbf{z} \rangle$, i.e. $\{\gamma (\mathbf{z} - \langle \mathbf{z} \rangle) \mid \gamma \in \mathbb{C}/\{0\}\}$.

Therefore, $\mathbf{\Sigma}_2^k$ can be represented by complex projective space $\mathbb{C}P^{k-2}$, which is the space of all complex line passing through origin on $\mathbb{C}^{k-1}$. Given the fact that scaling can be removed by $r = 1/\|z - \langle \mathbf{z} \rangle\|$, we can represent the shapes as equivalent class $[z] = \left\{ e^{i\theta}(z - \langle \mathbf{z} \rangle)/\|z - \langle \mathbf{z} \rangle\| \mid \theta \in [0, 2\pi] \right\}$, which lies on unit sphere in $\mathbb{C}^{k-1}$. The distance on $\mathbb{C}P^{k-2}$ is defined as

$$d(u, v) = \arccos\left(\mathrm{Re}(u^* v)\right) \ ,$$

where u^* is the conjugate transpose of u. This metric is used to calculate distances between shape objects $[\mathbf{z}_1]$ and $[\mathbf{z}_2]$, i.e. $d([\mathbf{z}_1], [\mathbf{z}_2])$.

Let P and Q denote two probability measures on $\mathbf{\Sigma}_2^k$, and $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ represent samples from P and Q respectively. Denote by $[\mathbf{x}_1], \ldots, [\mathbf{x}_n]$ and $[\mathbf{y}_1], \ldots, [\mathbf{y}_m]$ the image of samples into $\mathbb{C}^{k-1}$. Let us represent the combined sequential sample by $[\mathbf{z}_1], \ldots, [\mathbf{z}_t]$ for $t = 1, \ldots, n+m$. The geodesic depth of point $[\mathbf{z}_t]$ with respect to empirical measure ν_t can be computed as

$$GD(\nu_t, [\mathbf{z}_t]) = \exp\Big(- \frac{1}{t-1} \sum_{i=1}^{t-1} \arccos\left(\mathrm{Re}\left([\mathbf{z}_t]^* [\mathbf{z}_i]\right)\right) \Big) \ .$$

Denote the mapping from k-ads complex plane space $\mathbf{\Sigma}_2^k$ to $\mathbb{C}^{k-1}$ by ϕ. Then the ranks of shape objects can be computed as follows

$$R(Z_t) = R(\phi(Z_t)) = R(GD(\nu_t, [\mathbf{z}_t])) \ .$$

These ranks can then be used to calculate the values of control charts r-EWMA, r-CUSUM, and r-MWMASHC. Suppose we have collected 20 spiral bevel gear objects in Phase-I and 10 OC spiral bevel gear objects in Phase-II which are generated from Mardia-Dryden distribution with $\sigma^2 = 1.5$. We also add additive noise to the data to evaluate the performance of methods for distinguishing deformation fault from noise. We calculate the

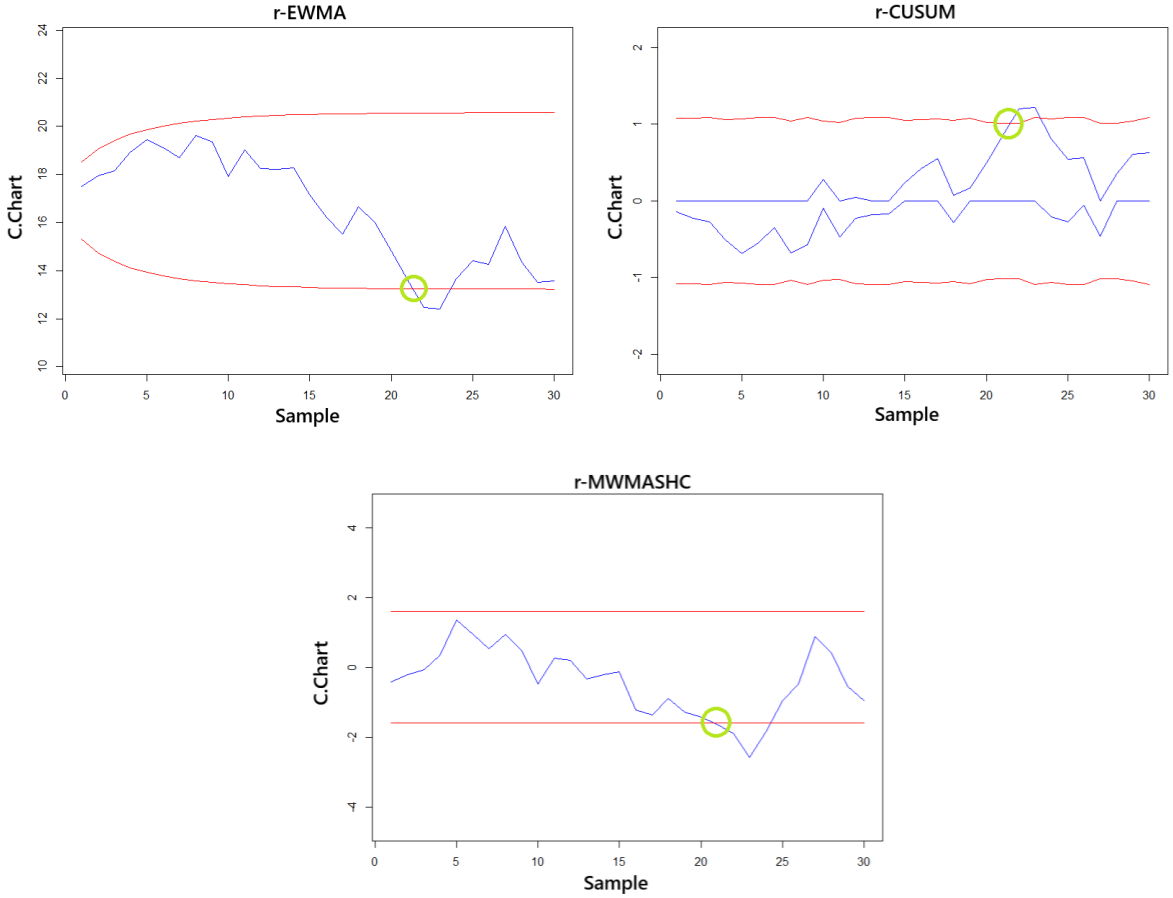


Figure 4.23: The performance of r-EWMA, r-CUSUM, and r-MWMASHC control charts in Phase-II.

geodesic depth value of OC object at time t with respect to collected data $i = 1, \ldots, t-1$. The following geodesic values are obtained

$$GD(\nu, [z_{21}]) = 0.2070717\ ,\ GD(\nu, [z_{22}]) = 0.2068551\ ,\ GD(\nu, [z_{23}]) = 0.2079436$$
$$GD(\nu, [z_{24}]) = 0.2089082\ ,\ GD(\nu, [z_{25}]) = 0.2082980$$
$$GD(\nu, [z_{26}]) = 0.2079714\ ,\ GD(\nu, [z_{27}]) = 0.2103337\ ,\ GD(\nu, [z_{28}]) = 0.2067006$$
$$GD(\nu, [z_{29}]) = 0.2073892\ ,\ GD(\nu, [z_{30}]) = 0.2079773\ .$$

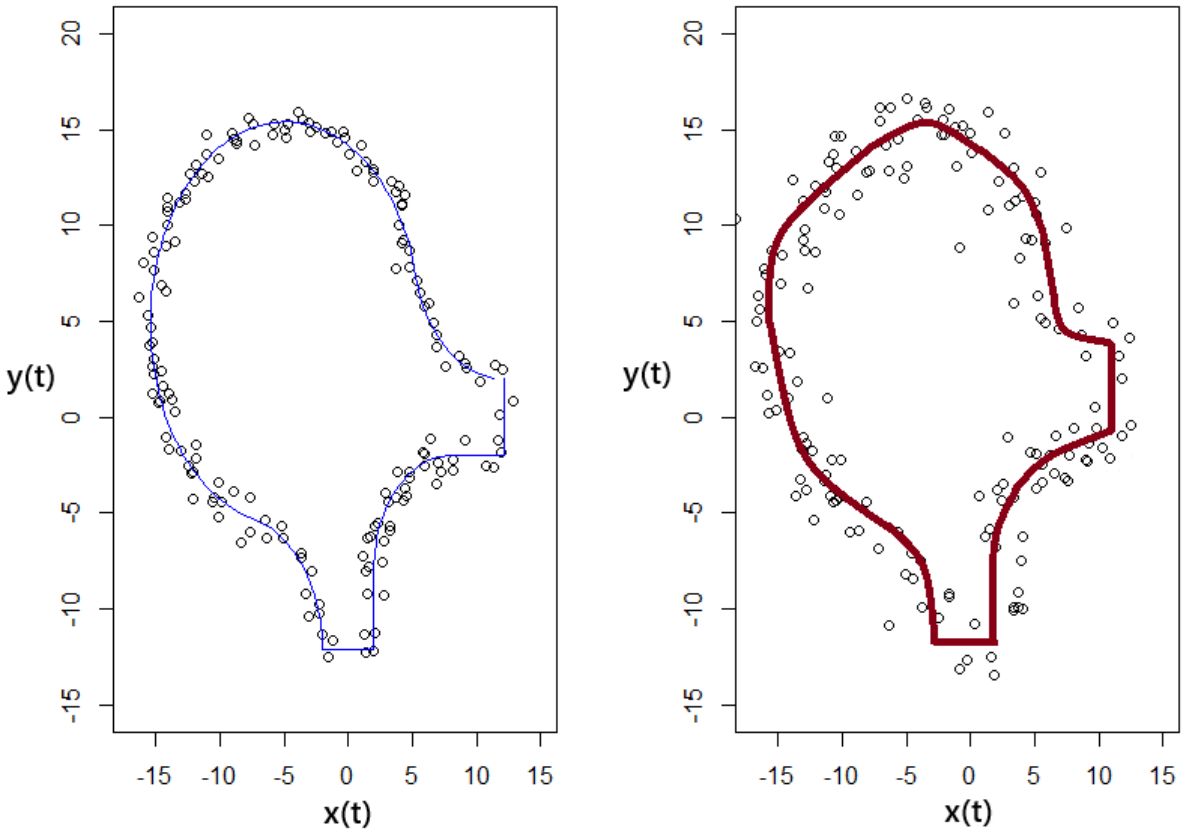


Figure 4.24: Illustration of healthy and faulty products observed in noise. The left panel indicates the contour of a healthy product. The right panel indicates the contour of a deformed product generated by Mardia-Dryden distribution.

Table 4.15: IC-ARL and OC-ARL values for IC and OC data obtained over 5000 simulation.

		IC-ARL		**OC-ARL**	
Methods		**avg.**	**std.**	**avg.**	**std.**
r-EWMA	$\lambda = 0.1$	371.1	253.7	2.52	0.24
r-CUSUM	$k = 0.2$	373.3	231.5	2.46	0.74
r-MWMASHC	$m = 10$	372.5	226.3	1.42	0.61

The corresponding ranks are

$$R_{21} = 3\ ,\ R_{22} = 2\ ,\ R_{23} = 12\ ,\ R_{24} = 25\ ,\ R_{25} = 21$$
$$R_{26} = 13\ ,\ R_{27} = 26\ ,\ R_{28} = 1\ ,\ R_{29} = 6\ ,\ R_{30} = 14\ .$$

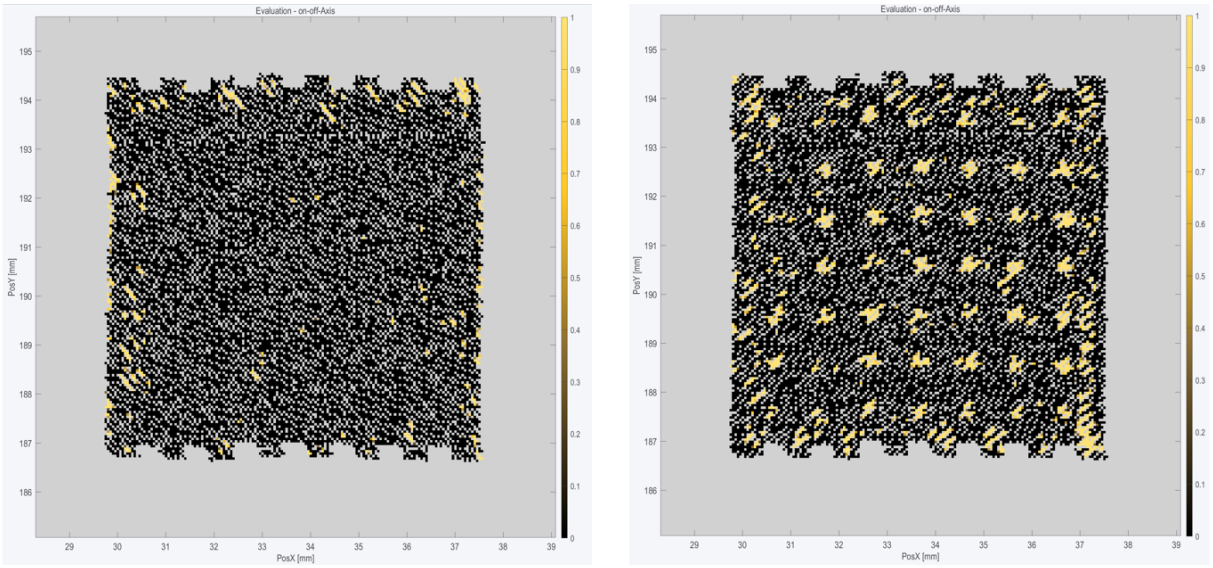


Figure 4.25: Sample images taken from set R2 location 8 (Taherkhani, 2022).

Figure 4.23 compares the performance of r-EWMA, r-CUSUM, and r-MWMASHC control chart. It can be seen that all of these nonparametric control charts correctly detected the faulty products and triggered false alarm rate. Table 4.15 reports the IC-ARL and OC-ARL values for IC and OC conditions. The performance of the methods are comparable for IC and OC conditions with the best performance belongs to r-MWMASHC followed by r-CUSUM and r-EWMA. Figure 4.24 depicts healthy and faulty products observed in noise. Apparently, the control charts are successful in identifying shape deformation in these noisy observations.

Example 12 (Laser Powder-Bed Fusion Process): In this example, we apply the proposed nonparametric process monitoring methods to detect defects in parts fabricated by a commercial laser powder-bed fusion (LPBF) system (Keshavarzkermani et al., 2019). This dataset is provided by the multi-scale additive manufacturing (MSAM) lab at the University of Waterloo.

Light intensity emitted from the melt pool is measured using photodiode monitoring systems to detect pores created from the lack of fusion phenomenon. The experiments are conducted considering two types of defects, which include parts with embedded micro-voids modeling the lack of fusion in parts of product Hastelloy-X and parts with stochastic pores resulting from process disturbance (Taherkhani, 2022). The faults are introduced at different operating parameters including low laser powers of 175W, 150W, 125W, and 100W, high hatching distances of 110 μm, 130 μm, and 150 μm, and process speeds of

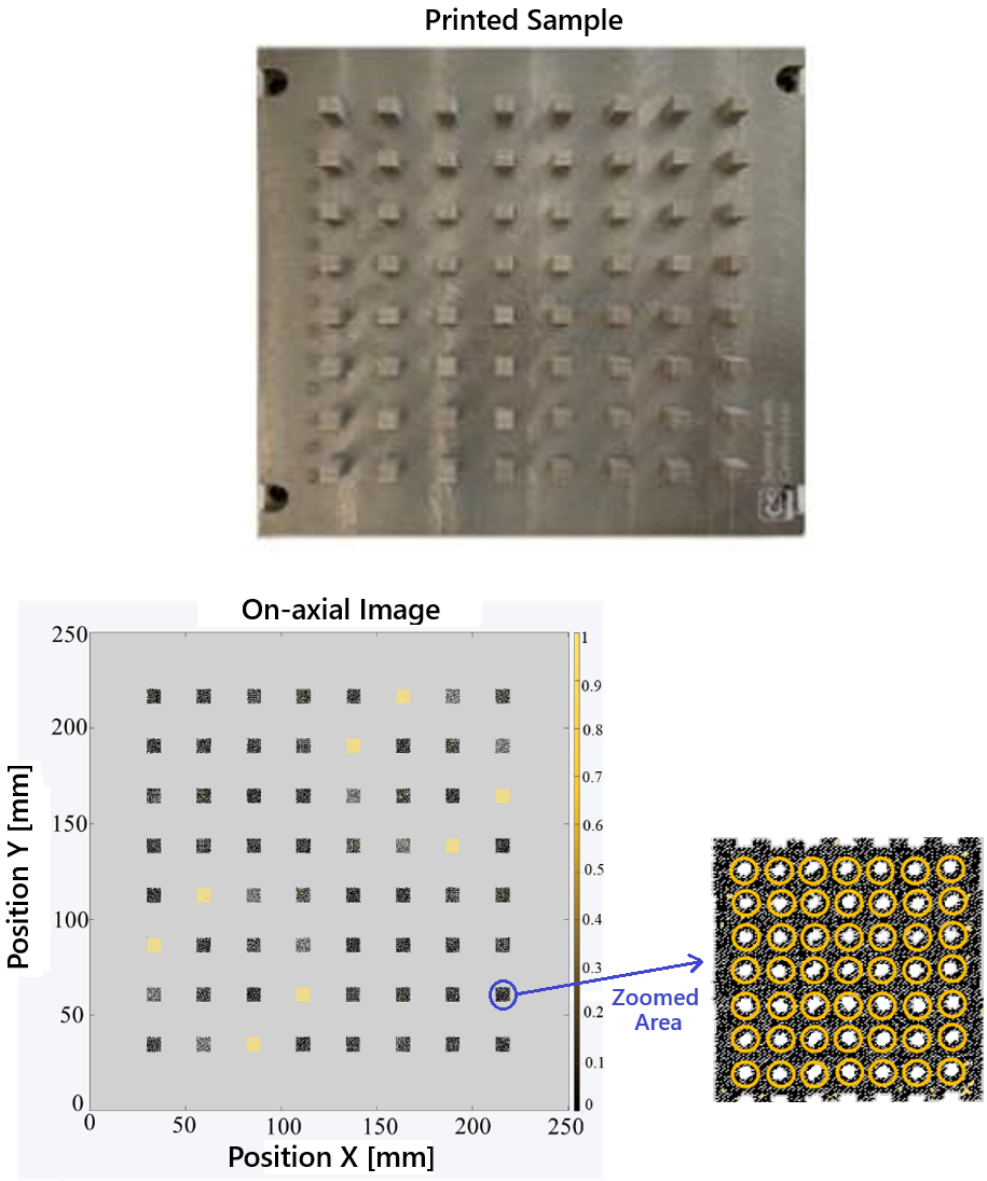


Figure 4.26: Top panel indicates a printed product. Bottom left panel indicates an image generated by software from a layer of the print. Bottom right panel indicates 49 regions of interest around each defect at the zoomed location (Taherkhani, 2022).

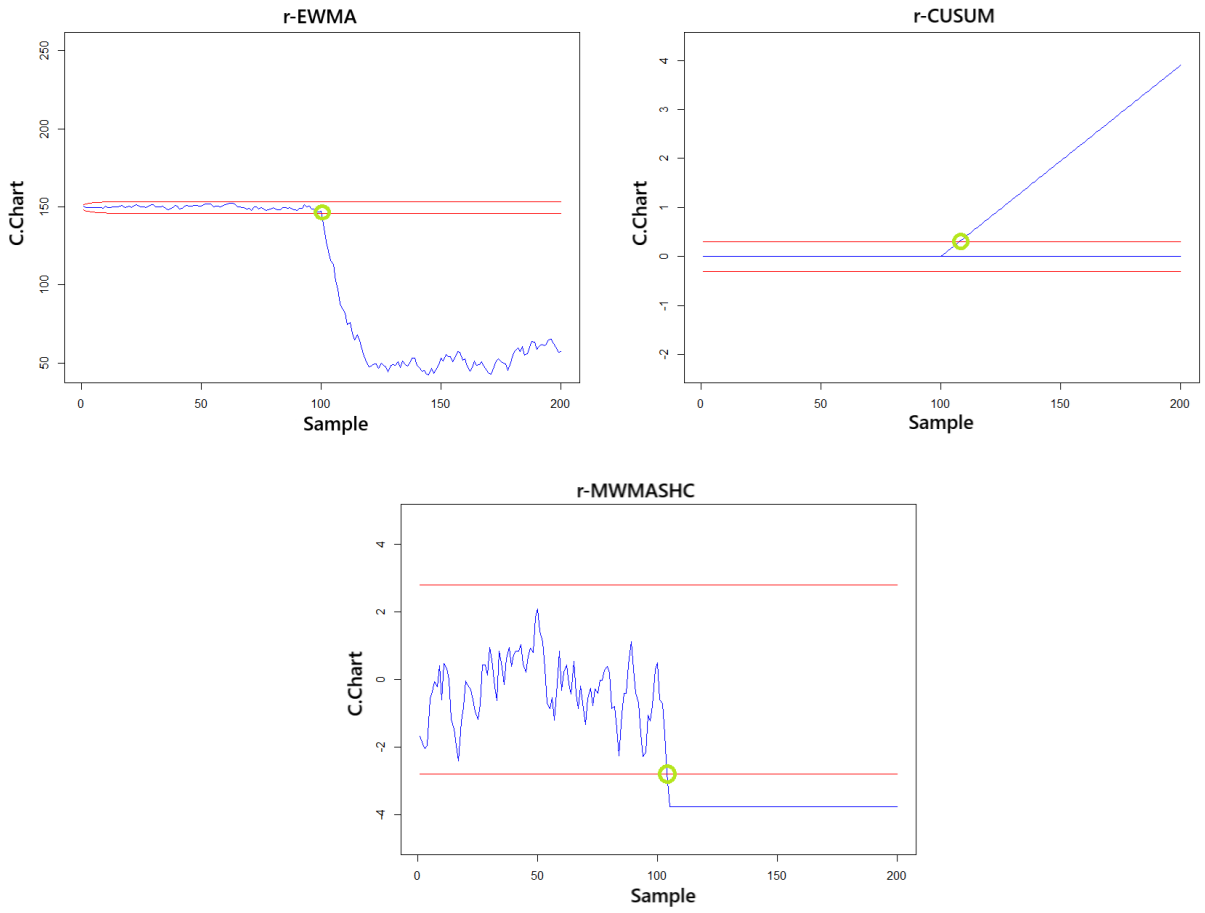


Figure 4.27: The performance of r-EWMA, r-CUSUM, and r-MWMASHC control charts in Phase-II.

Table 4.16: IC-ARL and OC-ARL values for IC and OC data obtained over 5000 simulation.

		IC-ARL		OC-ARL	
Methods		**avg.**	**std.**	**avg.**	**std.**
r-EWMA	$\lambda = 0.1$	373.3	246.4	1.12	0.53
r-CUSUM	$k = 0.2$	372.1	243.2	2.46	0.56
r-MWMASHC	$m = 10$	374.6	227.1	1.32	0.53

1100 mm/sec, 1300 mm/sec, and 1500 mm/sec.

The samples are cubical with a size of $8 \times 8 \times 10$ (W $\times$ L $\times$ H) mm, labeled as R-series.

In these samples, the effect of lack of fusion was simulated by embedding artificial voids in the samples. The considered void geometries are cylindrical (R2, R3, and R5) samples and spherical shapes (R4) with different sizes fabricated using the EOS-M290 LPBF machine. One control sample R1 without artificial void is also printed. Samples R6, R7, and R8 have similar design to R2 but printed with different process parameters.

The EOSTATE software is used to analyze the photodiode signal to identify the location of fluctuations. The result of in-situ analysis is exported as a grayscale TIFF image per layer which shows the nominal points as grayscale pixels and abnormal points as yellow pixels. In these TIFF images, each pixel represents the average value of intensity signal in the area of 5.55×5.55 ($\mu m \times \mu m$). Figure 4.25 indicates two sample images taken from set R2 location 8. We analyze these images using flag manifold $\mathrm{Fl}(v_1, ..., v_k; d)$ which models the evolution of features in the form of k nested subspaces of d principal component coordinates of the image data. The GD depth value are calculated using the standard distance on flag manifold spaces $\mathcal{M}$, from which the ranks are calculated and the control charts are used for monitoring the quality of the printed products. In this example, we considered these preprocessed image data. Our method is also applicable to raw images, without a need for the mentioned preprocessing steps to identify micro-voids in the products.

Figure 4.26 indicates a printed product along with an image generated by software from a layer of the print and a zoomed area indicating 49 regions of interest around each defect. The center of each circular area coincides with the center of each defect. The defects have cylindrical and spherical shapes. For samples with cylindrical defects, i.e. R2, R3, R5, R6, R7, and R8, the radius of area of interest is the same as the radius of artificial defect and for samples with spherical defect, i.e. R4, the radius of area of interest equals the radius of the spherical defect.

For the simulation, we considered 100 IC images from R1 set and 100 OC images generated uniformly from a complete set of 50 images from objects with voids. Figure 4.27 depicts the performance of r-EWMA, r-CUSUM, and r-MWMASHC control chart. Clearly, all of these nonparametric control charts are successful in detecting the faulty products. Table 4.16 compares the IC-ARL and OC-ARL values for IC and OC conditions. The performance of the methods are comparable for IC and OC conditions with the best performance belongs to r-EWMA followed by r-MWMASHC and r-CUSUM.

Example 13 (Robustness under Contamination for LPBF Process): In this example, our goal is to evaluate the robustness of the proposed nonparametric control charts when the images are contaminated with background noise. The simulation is carried out considering high noise to signal ratio. To do so, the images are converted to grayscale matrices and additive noise is added to the pixel values with high noise to signal ratio. Figure

Table 4.17: IC-ARL and OC-ARL values for IC and OC data obtained over 5000 simulation.

		IC-ARL		OC-ARL	
Methods		**avg.**	**std.**	**avg.**	**std.**
r-EWMA	$\lambda = 0.1$	375.4	247.7	9.51	5.13
r-CUSUM	$k = 0.2$	377.5	249.7	10.21	4.22
r-MWMASHC	$m = 10$	373.8	221.6	11.77	7.55

4.28 indicate a sample of healthy and faulty images from layers of the printed product. We contaminate these images with additive noise.

Figure 4.29 compares the performance of the proposed nonparametric control charts for OC condition. As it can be seen, all three methods can successfully detect the OC condition and triggered false alarm. It can be inferred that the use of rank-depth is effective so that the high noise to signal ratio does not affect the outcome of these control charts. Table 4.17 reports the results of simulation for IC and OC conditions. Clearly, the OC-ARL values are larger for the high noise to signal contamination scenario compared to the results reported in Table 4.16. However, these values are still comparable indicating the robustness of the methods in detecting micro void defects from noise contaminated images. This endorses the applicability of these methods for real-time quality monitoring for LPBF process, as the quality of printed products can be reliably monitored without being affected by high noise to signal ratio as well as other nuisance parameters.

Example 14 (Break-down Point Analysis for LPBF Data): In this example, we evaluate the breakdown point robustness of the nonparametric control charts using image data from LPBF process.

Figure 4.30 compares the gross-error sensitivities for r-EWMA, r-CUSUM, and r-MWMASHC for IC and OC conditions. The left panel evaluates the robustness of the proposed methods for IC condition. Clearly, increasing the contamination level decreases IC-ARL so that the ratio $\inf_{H_0 \in \mathcal{H}_{0,\epsilon}} \frac{\mathbb{E}_{H_0}^{(\infty)}(T)}{\mathbb{E}_{P}^{(\infty)}(T)}$ approaches 0. From the plot it can be seen that these control charts have a comparable robustness with false-alarm breakdown point $\varepsilon_F = 0.4$. The right panel evaluates the robustness of the methods for OC condition. As seen, increasing contamination level increases OC-ARL such that the ratio $\sup_{H_1 \in \mathcal{H}_{1,\epsilon}} \frac{\text{WADD}_{P,H_1}(T)}{\text{WADD}_{P,Q}(T)}$ becomes very large. The plot suggests $\varepsilon_D = 0.45$ for rank-depth control charts. It can be concluded that rank-depth control charts enjoy robustness with relatively large breakdown points, and can be considered as a promising quality monitoring approach for real-world AM applications such as LPBF process.

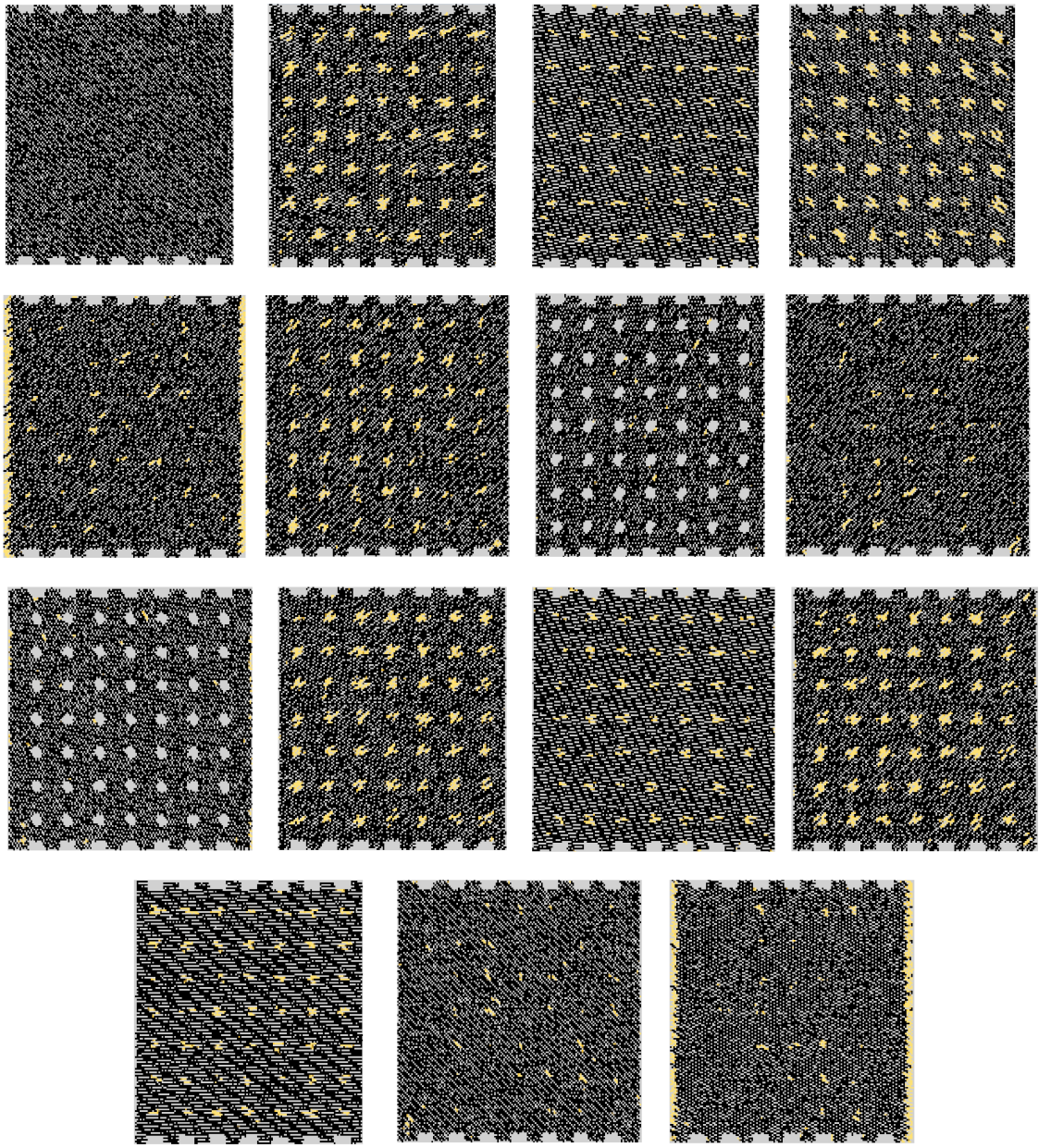


Figure 4.28: A sample of healthy and faulty images from layers of printed product. The faulty images include void defects in the layers of printed product (Taherkhani, 2022).

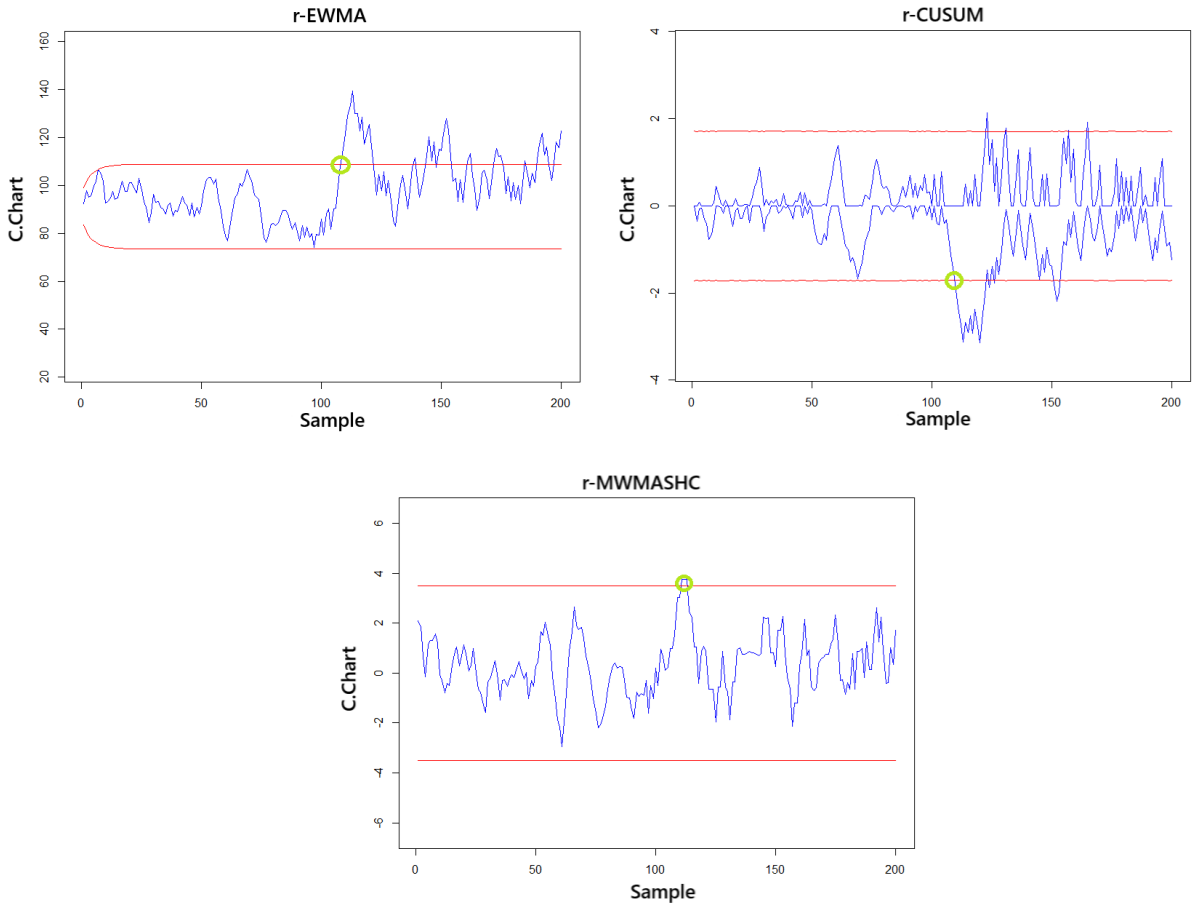


Figure 4.29: The performance of r-EWMA, r-CUSUM, and r-MWMASHC control charts in Phase-II.

4.5 Discussion and Future Work

In this chapter, we implemented rank-depth control charts for monitoring processes which require manifold data analysis. The performance of the proposed methods were evaluated considering Monte Carlo simulation and real-world datasets. The simulation results demonstrated the effectiveness of rank-depth control charts. It was observed that rank-depth control charts enjoy robustness and can effectively handle monitoring task for processes with object data from various types of manifold spaces, e.g. Steifel, SPD, sphere, complex projective space, and flag manifold. Rank-depth control charts were effective in the sense that they could control the IC-ARL at the nominal level and minimize the OC-

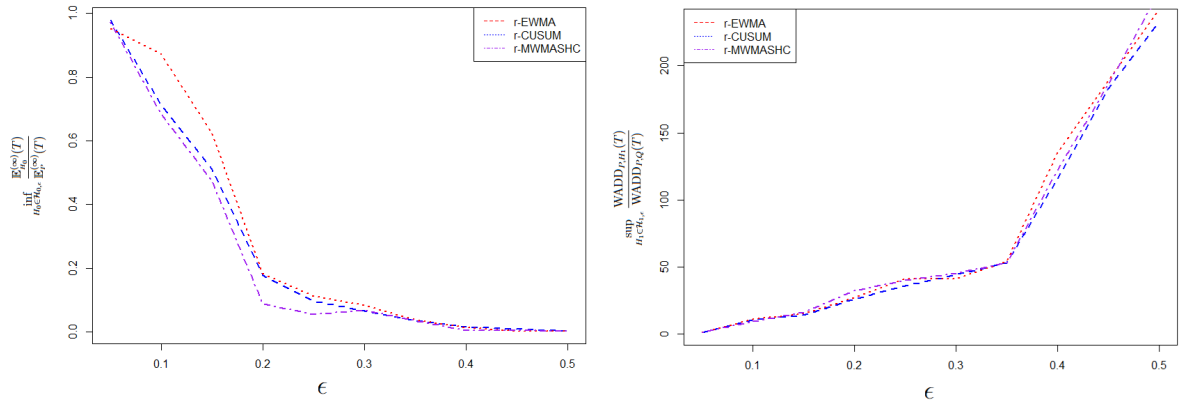


Figure 4.30: Gross-error sensitivities for r-EWMA, r-CUSUM, and r-MWMASHC for IC and OC conditions with different values of ε.

ARL regardless the dimensionality of the underlying space and the types of manifold space. Moreover, these methods could show an acceptable performance for sample sizes as small as $n = 20$.

Finite sample Monte Carlo simulation studies demonstrated the validity of the asymptotic results derived for the nonparametric control charts. Several well-known nonparametric control charts were extracted from the literature and used to evaluate the performance of rank-depth control charts. It was observed that the proposed control charts could show promising results and for most of the cases outperform the rival methods when the dataset was contaminated with outliers. Simulation results proved that rank-depth control charts can be successfully applied to both intrinsic and extrinsic manifold analysis frameworks.

Robustness analysis was carried out from both theoretical and numerical viewpoints. The theoretical analysis involved q-robustness analysis as well as asymptotic breakdown point analysis for both IC and OC conditions. Comparative numerical breakdown point study demonstrated that rank-depth control charts can show superior performance to the considered rival methods, and are less sensitive to the contaminating outlier data.

Future research direction involves applying the proposed nonparametric control charts to important emerging problems from various research disciplines ranging from additive manufacturing to robotics and automotive engineering. It is believed that rank-depth control chart can be very effective in such applications given the generalization power and the fact that they are free from any assumption for the distribution of object data.

Appendix A: Implementation of Parametric Control Chart

We propose a parametric control chart based on an analytical modeling for decomposing warping and phase components of functional data. In the current analysis, the domain of the functions are normalized to lie in the interval $[0, 1]$, and the space of functions is defined as $\mathcal{F} = \{f : [0, 1] \to \mathbb{R} | f \text{ is absolutely continuous}\}$. The absolute continuity condition implies $f(t) = f(0) + \int_0^t \dot{f}(s)ds$, where $\dot{f}$ represents the first derivative of function f. Also, the current analysis uses orientation-preserving diffeomorphism maps γ (called warping functions) which belong to the space $\Gamma = \{\gamma : [0, 1] \to [0, 1] | \dot{\gamma} > 0,\ \gamma(0) = 0,\ \gamma(1) = 1\}$, by defining the warping of f using γ as the decomposition operator $f \circ \gamma = f(\gamma(t))$. This operator enables the decomposition of functional data to phase and amplitude components which serves as a mean to analyze the variability of collected data more efficiently.

Recall that $||f_1 - f_2||_2^2 = \int_0^1 \big(f_1(t) - f_2(t)\big)^2 dt$, for $f_1, f_2 \in \mathcal{F}$. One difficulty with directly applying the above operator on functions f for decomposition is the lack of isometry under L_2. Mathematically speaking, $\forall f_1, f_2 \in \mathcal{F}$, $||f_1 - f_2||_2 \neq ||f_1 \circ \gamma - f_2 \circ \gamma||_2)$, as long as $\dot{\gamma}(\gamma^{-1}(t)) \neq 1$ for $\exists t \in [0, 1]$. To tackle this problem, a bijection mapping known as square root slope function (SRSF), $\mathcal{Q} : \mathcal{F} \to L_2\big([0, 1], \mathbb{R}\big)$ can be taken into account. This map can be mathematically defined as $Q(f(t)) = q(t) := sgn(\dot{f}(t))\sqrt{|\dot{f}(t)|}$, which is invariant under translation, and acts under L_2 by isometry $||q_1 - q_2||_2 = ||q_1 \circ \gamma - q_2 \circ \gamma||_2$. The operator $q \circ \gamma = q(\gamma(t))\sqrt{\dot{\gamma}(t)}$ can be used for the decomposition of phase and amplitude components. The important feature of this mapping lies in the fact that the bjection relation between $\mathcal{F}$ and SRSF space ensures that for each q there exists a unique function $f \in \mathcal{F}$. Let $[q]$ be the equivalent class $[q] = \{q \circ \gamma | \gamma \in \Gamma\}$. From the above description, it is obvious that the decomposition process maps the elements of SRSF space $L_2\big([0, 1], \mathbb{R}\big)$ to the amplitude space $L_2/\Gamma = \{[q] | q \in L_2\}$, and the resulting phase space Γ. The decomposition process can be performed using the following steps:

1. Map the functional data $\{f_1, f_2, \cdots, f_N\}$ to SRSF space to get $\{q_1, q_2, \cdots, q_N\}$.
2. Define the distance $D_a([q_1], [q_2]) := \inf_{\gamma \in \Gamma} ||q_1 - (q_2, \gamma)||_2^2$, and compute the median by solving the optimization problem below

$$[\bar{q}] = \arg \min_{[q] \in L_2/\Gamma} \sum_{i=1}^{N} D_a([q_i], [q])\ . \tag{4.1}$$

The solution to the above problem can be found using gradient descent algorithm.

3. Select an element from the median equivalent class $\bar{q}$ using orbit centering method (Srivastava and Klassen, 2016).

4. Use dynamic programming to allign the SRSFs to the median (the details of the method can be found in page 100 of Srivastava and Klassen (2016)), and record the set of aligned SRSFs $\{\tilde{q}_1, \tilde{q}_2, ..., \tilde{q}_N\}$ (and corresponding $\{\tilde{f}_1, \tilde{f}_2, ..., \tilde{f}_N\}$), amplitude distances $\{d_a^1, d_a^2, ..., d_a^N\}$, and optimal phase functions $\{\gamma_1, \gamma_2, ..., \gamma_N\}$.

The information obtained from the above decomposition can be used to design control charts for process monitoring. In particular, the obtained amplitude distances is used for the development of control charts to monitor for possible out-of-control products with respect to amplitude components. A similar concept is followed to design a chart for monitoring the phase components.

The intrinsic difficulty associated with this task is that the phase space Γ has a nonlinear geometry, which makes the direct use of L_2 impossible. To handle this problem, one approach is using a bijection map called square root transformation (SRT) defined as $\psi(t) := \sqrt{\dot{\gamma}(t)}$. Obviously, any element of such a space satisfies $||\psi||_2 = 1$ and $\psi(t) > 0$ for $t \in [0, 1]$, meaning that it belongs to the posistive orthant of the unit Hillbert sphere space Ψ, for which explicit formulas are available. Note that $\gamma(t)$ can be recovered by the inverse of SRT map $\gamma(t) = \int_0^t \psi^2(s)ds$. The geometry of space Ψ suggests the calculation of distance between γ_1 and γ_2 using the arc-length of their corresponding SRTs, i.e. $D_p(\gamma_1, \gamma_2) = cos^{-1}(\langle \psi_1, \psi_2 \rangle)$. Also, the median $\bar{\psi}$ can be computed by solving the optimization problem

$$\bar{\psi} = \arg\min_{\psi \in \Psi} \sum_{i=1}^{N} D_p(\gamma_i, \gamma) \ , \tag{4.2}$$

from which the tangent space at point $\bar{\psi}$ defined by $T_{\bar{\psi}}(\Psi) := \{\nu | \langle \nu, \bar{\psi} \rangle = 0, \ \nu : [0, 1] \to \mathbb{R}\}$ can be computed. The linearity of $T_{\bar{\psi}}(\Psi)$ enables the use of L_2 for calulating the distances. It is worth mentioning that the distances between phase functions could be obtained using $D_p(\gamma_i, \bar{\gamma})$; however, the authors choose to work with the tangent space because of two reasons: (1) the behaviour of L_2 is well-known and can facilitate the designing of control charts, and (2) obtaining phase distances using L_2 results in a dataset which is in agreement with the obtained amplitude distances, and makes it possible to design a joint monitoring chart for both amplitude and phase components. Obtaining data points for phase analysis is done using the steps below:

1. Map the phase functions $\{\gamma_1, \gamma_2, ..., \gamma_N\}$ obtained from decomposition process to $T_{\bar{\psi}}(\Psi)$ to get the shooting vectors $\{\nu_1, \nu_2, ..., \nu_N\}$, using the inverse exponential map below

$$\nu_i = \exp_{\bar{\psi}}^{-1}(\psi_i) = \frac{\theta}{sin(\theta)}(\psi_i - cos(\theta)\bar{\psi}), \quad \theta = cos^{-1}(\langle \psi_i, \bar{\psi} \rangle) \ , \quad i = 1, ..., n \ .$$

2. Calculate the distances using $D_a(\nu_i, \bar{\nu})$ in Eq.4.1, and record $\{d_p^1, d_p^2, ..., d_p^N\}$.

The phase distance for any new observed phase function during the monitoring process can be calculated following the above steps.

Due to the functional nature of the AM products, and the need for working in a space equipped with a norm which automatically satisfies isometry under the action of warping functions, there is not much flexibility in the selection of charts and working with distance measures, as it is commonly available in the development of univariate and multivariate charts.

We implement and investigate the properties of parametric SHC and EWMAC, suited for the condition inwhich the observed random functions follow Gaussian distribution. Let us represent the observed amplitude functions $q_i \in L_2/\Gamma$ and their corresponding shooting vectors $\nu_i \in T_{\bar{\psi}}(\Psi)$ by $q_i(t) = \bar{q}(t) + \sum_{k=1}^{\infty} c_{i,k}^q g_k(t)$ and $\nu_i(t) = \bar{\nu}(t) + \sum_{k=1}^{\infty} c_{i,k}^{\nu} h_k(t)$, respectively, where g_k and h_k are orthonormal eigenfunctions, and $c_{i,k}^q$ and $c_{i,k}^{\nu}$ are coefficients satisfying $c_{i,k}^q \to 0$ and $c_{i,k}^{\nu} \to 0$, as $k \to \infty$. One way to generate functions following Gaussian distribution is to take the orthonormal eigenfunctions g_k and h_k as fixed and let the coefficients follow Gaussian distribution, i.e. $c_{i,k}^q \sim \mathrm{N}(0, \sigma_{1,k}^2)$ and $c_{i,k}^{\nu} \sim \mathrm{N}(0, \sigma_{2,k}^2)$. Also, let the initial boundary value $f(0) \sim N(\mu_{f_0}, \sigma_3^2)$. It is easy to recover the original functions using the initial value and the composition $f_i \circ \gamma_i$.

As mentioned, the control charts are developed based on the L_2 distance of q and ν functions from their respective median. Using the above setup, it can be shown that $q_i \sim \mathrm{N}(\bar{q}, \sum_{k=1}^{\infty} \sigma_{1,k}^2 g_k^2(t))$ and $\nu_i \sim \mathrm{N}(\bar{\nu}, \sum_{k=1}^{\infty} \sigma_{2,k}^2 h_k^2(t))$. Let's assume that there are exactly K^q and K^{ν} nonzero coefficients in the above representation models, then

$$\begin{aligned} ||q_i - \bar{q}||_2^2 &= \int_0^1 \Big(\sum_{k=1}^{\infty} c_{i,k}^q g_k(t) \Big)^2 dt = \int_0^1 \Big(\sum_{k=1}^{K^q} \sum_{\ell=1}^{K^q} c_{i,k}^q c_{i,\ell}^q g_k(t) g_\ell(t) \Big) dt \\ &= \sum_{k=1}^{K^q} \sum_{\ell=1}^{K^q} c_{i,k}^q c_{i,\ell}^q \Big(\int_0^1 g_k(t) g_\ell(t) dt \Big) = \sum_{k=1}^{K^q} (c_{i,k}^q)^2 \ , \end{aligned}$$

$$||\nu_i - \bar{\nu}||_2^2 = \sum_{k=1}^{K^\nu} (c_{i,k}^\nu)^2 \ , \qquad i = \{1, 2, ..., N\} \ .$$

Note that in the derivation of the above result, the fact that the eigenfunctions are orthonormal is taken into account.

Now, given the fact that $\dfrac{(c_{i,k}^q)^2}{\sigma_{1,k}^2} \sim \chi_1^2 = Gam(\frac{1}{2}, 2)$, each term in the above summations follows a distribution $(c_{i,k}^q)^2 \sim Gam(\frac{1}{2}, 2\sigma_{1,k}^2)$, where $\frac{1}{2}$ and $2\sigma_{1,k}^2$ are the shape and scale parameters, respectively. Similarly, $(c_{i,k}^\nu)^2 \sim Gam(\frac{1}{2}, 2\sigma_{2,k}^2)$. For amplitude SHC, the statistic can be expanded as

$$\begin{aligned} Z_i^{SH} = \frac{1}{M} \sum_{j=1}^{M} \mathbf{g}_{lj} &= \frac{1}{M} \big(||q_i - \bar{q}||_2^2 + ||q_{i+1} - \bar{q}||_2^2 + ... + ||q_{M+i-1} - \bar{q}||_2^2 \big) \\ &= \frac{1}{M} \sum_{k=1}^{K^q} \big((c_{i,k}^q)^2 + (c_{i+1,k}^q)^2 + ... + (c_{M+i-1,k}^q)^2 \big) \ , \\ &\sim \sum_{k=1}^{K^q} Gam\big(\frac{M}{2}, \frac{2\sigma_{1,k}^2}{M} \big) \ , \end{aligned}$$

for $i = 1, 2, ..., N - M + 1$, which is the summation of K^q independent Gamma distributions with the same shape parameter $\dfrac{M}{2}$, and distinct scale parameters $\dfrac{2\sigma_{1,k}^2}{M}$, for $k = 1, 2, ..., K^q$. The derivation of exact distribution for random variable Z_i^{SH} can be carried out by means of moment generating functions, which is a single gamma-series, as below

$$Pr(Z_i^{SH} \leq \omega) = \Bigg(\prod_{k=1}^{K^q} \Big(\frac{\sigma^*}{\sigma_{1,k}^2} \Big)^{\frac{M}{2}} \Bigg) \Bigg(\sum_{j=0}^{\infty} \delta_j \int_0^{\omega} \frac{z^{\frac{MK^q}{2}+j-1} e^{\frac{-Mz}{2\sigma^*}}}{\Gamma_*(\frac{MK^q}{2} + j)(\frac{2\sigma^*}{M})^{\frac{MK^q}{2}+j}} dz \Bigg) \ , \tag{4.3}$$

where $\sigma^* := \min\limits_{k \in \{1,2,...,K^q\}} \{\sigma_{1,k}^2\}$, $\Gamma_*(\cdot)$ is the standard gamma function, and δ_j can be calculated recursively as

$$\delta_j = \frac{1}{j} \sum_{r=1}^{j} r \kappa_r \delta_{j-r} \ , \tag{4.4}$$

with $\delta_0 = 1$, and $\kappa_r = \sum\limits_{k=1}^{K^q} \frac{M(1 - \sigma^*/\sigma_{1,k}^2)^r}{2r}$. The derivation of exact distribution for phase SHC follows the same steps, with K^q and $\sigma_{1,k}^2$ replaced by K^ν and $\sigma_{2,k}^2$.

By expanding the terms in Z_i^{EW}, the distribution of EWMAC can be obtained as below

$$\begin{aligned} Z_i^{EW} = \lambda X_i + (1-\lambda) Z_{i-1}^{EW} &= \lambda ||q_i - \bar{q}||_2^2 + (1-\lambda)\lambda ||q_{i-1} - \bar{q}||_2^2 + \\ &+ (1-\lambda)^2 \lambda ||q_{i-2} - \bar{q}||_2^2 + ... + (1-\lambda)^{i-1} \lambda ||q_1 - \bar{q}||_2^2 \\ &= \sum_{k=1}^{K^q} \left(\lambda (c_{i,k}^q)^2 + (1-\lambda)\lambda (c_{i-1,k}^q)^2 + ... + (1-\lambda)^{i-1} \lambda (c_{1,k}^q)^2 \right) \\ &\sim \sum_{k=1}^{K^q} \sum_{j=1}^{i} Gam\big(\frac{1}{2}, 2\lambda(1-\lambda)^{i-j} \sigma_{1,k}^2\big) \ . \end{aligned}$$

As seen, the distribution of Z_i^{EW} can be determined as the summation of iK^q Gamma distributions with the same shape but distinct scale parameters. Let $\beta_1, ..., \beta_{iK^q}$ present the scale parameters in the summation above. Then, by following steps similar to those of SHC, the following distribution can be obtained

$$Pr(Z_i^{EW} \leq \omega) = \left(\prod_{\ell=1}^{iK^q} \left(\frac{\beta^*}{\beta_\ell} \right)^{\frac{1}{2}} \right) \left(\sum_{j=0}^{\infty} \delta_j \int_0^{\omega} \frac{z^{\frac{iK^q}{2}+j-1} e^{\frac{-z}{\beta^*}}}{\Gamma_*(\frac{iK^q}{2}+j) {\beta^*}^{\frac{iK^q}{2}+j}} dz \right) , \tag{4.5}$$

where $\beta^* := \min_{\ell \in \{1,2,...,iK^q\}} \{\beta_\ell\}$, δ_j is given in Eq.(4.4), with $\kappa_r = \sum_{\ell=1}^{iK^q} \frac{(1-\beta^*/\beta_\ell)^r}{2r}$. The exact distribution for phase EWMAC can be obtained by following the same procedure and replacing K^q and $\sigma_{1,k}^2$ by K^ν and $\sigma_{2,k}^2$.

The authors' numerical experiments indicated that approximating Z_i^{EW} from Eq.(4.5) can become numerically unstable when iK^q (or equivalently iK^ν) becomes large. To avoid this difficulty, an expression resulting from the convergence property of geometric series can be considered (Moschopoulos, 1985; Covo and Elalouf, 2014). This is done as below

$$\begin{aligned} \sum_{j=0}^{i-1} \sigma_{1,k}^2 \lambda (1-\lambda)^j \frac{(c_{i-j,k}^q)^2}{\sigma_{1,k}^2} &\sim \sum_{j=0}^{i-1} \sigma_{1,k}^2 \lambda (1-\lambda)^j \chi_1^2 = \sigma_{1,k}^2 \lambda \frac{1-(1-\lambda)^i}{1-(1-\lambda)} \chi_1^2 \\ &= Gam\big(\frac{1}{2}, 2(1-(1-\lambda)^i)\sigma_{1,k}^2\big) \ . \end{aligned}$$

Also, if required, asymptotic distribution can be used to make the chart independent from i, as follows

$$Z_i^{EW} = \sum_{k=1}^{K^q} Gam\big(\frac{1}{2}, 2(1-(1-\lambda)^i)\sigma_{1,k}^2\big) \xrightarrow[i\to\infty]{\mathcal{D}} Z^{EW} \sim \sum_{k=1}^{K^q} Gam\big(\frac{1}{2}, 2\sigma_{1,k}^2\big) \ .$$

Note that depending on the application, asymptotic probability for EWMAC can reduce the accuracy, and thus, should be used with caution. The distribution for EWMAC can be expressed as

$$Pr(Z_i^{EW} \leq \omega) = \left(\prod_{k=1}^{K^q} \left(\frac{\sigma^*}{\sigma_{1,k}^2} \right)^{\frac{1}{2}} \right) \left(\sum_{j=0}^{\infty} \delta_j \int_0^{\omega} \frac{z^{\frac{K^q}{2}+j-1} e^{\frac{-z}{2(1-(1-\lambda)^i)\sigma^*}}}{\Gamma_*(\frac{K^q}{2}+j)(2(1-(1-\lambda)^i)\sigma^*)^{\frac{K^q}{2}+j}} dz \right) , \tag{4.6}$$

where the definitions of the parameters are the same as those of Eq.(4.3), with $\kappa_r = \sum_{k=1}^{K^q} \frac{(1-\sigma^*/\sigma_{1,k}^2)^r}{2r}$. Again, the same procedure is valid for phase EWMAC.

In a proposal by Covo and Elalouf (2014), a numerical method is derived to approximate the above exact distributions (the theoretical details and results regarding the numerical accuracy of the method can be found in (Covo and Elalouf, 2014)). The main idea is that each of the distributions above can be approximated using a single Gamma distribution, as their levy density is similar to that of the Gamma distribution. Consider the condition that the interest is to approximate the summation of K^q Gamma distributions with shape parameter ϱ_k and scale parameter ξ_k, where $k = 1, 2, ..., K^q$. The parameters of the approximated $Gam(\varrho^*, \xi^*)$ can be obtained, by first solving the equation below for ξ^*

$$\sum_{k=1}^{K^q} \frac{\varrho_k \xi_k}{2} - 2 \sum_{k=1}^{K^q} \frac{\varrho_k \xi_k^3}{(\xi_k + \xi)^2} = 0 ,$$

and then obtaining $\varrho^* = \sum_{k=1}^{K^q} \frac{\varrho_k \xi_k}{\xi^*}$. Solution to the equation above for ξ^* can be found using bisection search. Consequently, ARL_0 with significance level α can be determined from $Gam(\varrho^*, \xi^*)$.

Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be the events that the amplitude and phase charts send false OC signal, respectively. Then, for joint statistic, the probability of sending false signal can be kep at a desired level α using the following approximation

$$Pr(\mathcal{A}_1 \cup \mathcal{A}_2) = Pr(\mathcal{A}_2) + Pr(\mathcal{A}_2) - Pr(\mathcal{A}_1 \cap \mathcal{A}_2) \approx Pr(\mathcal{A}_2) + Pr(\mathcal{A}_2) ,$$

from which the joint ARL_0^* can be computed.

Chapter 5

Future Work

Research related to statistics on manifold spaces is growing rapidly encompassing both theoretical findings and practical aspects. In recent years, the society has witnessed a major breakthrough with respect to theoretical findings whose goal was to explore convergence rate and conditions needed to satisfy consistency of the estimators. In doing so, advanced theoretical results concerning sharp bounds based on the geometrical complexity of such spaces have been developed and in particular it has been proven that the convergence rates can be derived as a function of intrinsic dimension of the manifold, complexity of the geometrical space, and the number of sample sizes (Cheng and Xie, 2021*a*). Moreover, it was demonstrated that the theoretical investigation can be followed under both intrinsic and extrinsic manifold analysis setups (Bhattacharya and Patrangenaru, 2003, 2005). From the practical point of view, the research stream has been followed by applying the developed statistical analysis tools to complex research problems dealing with high-dimension correlation/covariance matrices, high-dimensional landmark sets representing various shapes, curved data sampled from manifold spaces, e.g. probability density functions, and images represented as matrix data (Marron and Dryden, 2021). The areas of applications encompass robotics, process monitoring, computer vision, healthcare and clinical data analysis, biology and genetics, social network analysis, etc. Following such interest and the promising research outcomes, the author is motivated to extend the research towards these directions and explore the potential of robust statistical analysis on manifold spaces. In particular the main interest is to extend the conducted research by developing proper statistics for robust analysis on dynamical environments, such as evolution of graphs, and also longitudinal data analysis on manifold spaces.

The rest of the chapter is organized as follows. In the first section, the author discusses longitudinal data analysis on manifold spaces as a potential research area for future work.

In the second section, the author discusses research on dynamical networks related to the evolution of graphs over time. The author also discusses potential research directions for extending the theoretical findings in these research areas.

5.1 Longitudinal Data Analysis

With the advancement in measurement system and sensor technology, it is now common to collect data in a longitudinal repeated measurements. Longitudinal data analysis is a well-established research field and has attracted a significant attention from the community (Fitzmaurice et al., 2008). However, most of the conducted research focus on dataset sampled from Euclidean space. Recently, interesting research articles are published which demonstrate the potential for extending the classical longitudinal data analysis to the cases where dataset are sampled from manifold spaces (Dai et al., 2021). It has been proven that classical principal component analysis can be extended to the principal manifold component analysis with the goal of extracting the information lies within a sample of longitudinal manifold data. Theoretical investigations endorsed the validity of the developed method and indicated that longitudinal analysis on manifold data can be followed by appropriate modification of classical tools. Following the reported findings, the author intends to explore the potentials of robust estimation for extracting information from a collection of longitudinal manifold data. Research on longitudinal manifold data analysis can be carried out with the following goals:

- It is very common that the longitudinal data is contaminated with outliers which needs to be taken care of with appropriate estimators. To demonstrate the effectiveness of the developed robust longitudinal data analysis tool, the authors will evaluate the statistical properties of the method considering various contamination noise and outliers and also study the effect of sample size and the proportion of outlying data.

- It is very interesting to explore the theoretical properties of the developed robust manifold data analysis tools and compare the convergence rates and efficiency with existing methods. The goal of the analysis is to investigate the efficiency of the robust estimator as a function of sample size and dimensionality of the dataset (or the intrinsic dimension of the underlying manifold space). Breakdown point analysis will be carried out to demonstrate the robustness of the developed methods.

- Another important research direction is to develop robust longitudinal data analysis for data sampled from general metric spaces. The importance of the analysis is to

make sure the developed method is appropriate for a various classes of manifold spaces. To do so, the robust statistic can be implemented as a function of general metrics and then it can be generalized to various manifold spaces.

- It is important to probe the potential of both intrinsic and extrinsic manifold data analysis for longitudinal data. The reason is that for some manifold spaces the best strategy is to work with extrinsic setup, e.g. the landmark data, while for some cases, we can effectively work with local embeddings and intrinsic analysis. These will be explored in detail with the goal of clarifying the effectiveness of both approaches for longitudinal data analysis on manifold spaces.

- To the author's best knowledge, there are rare reports in the literature concerning the development of convergence bounds and asymptotic distribution for the robust statistics on manifold spaces for longitudinal data applications. Therefore, following these research directions should afford interesting outcomes clarifying the efficiency and consistency of robust test statistics for longitudinal data analysis on manifolds.

5.2 Dynamical Networks

Statistical analysis of dynamical networks has been a major research area over the past decade with successful applications in the fields such as social psychology, sociology, physical science, computer science and community detection. Statistical models for analyzing network data comprises of graphical representation schemes and probability models which explains the stochastic structure of the networks. Common probability models include exponential random graph models, Erdős-Rényi model, and stochastic block model which is proper for networks with latent network structure (Fienberg, 2012). An important statistical problem related to analyzing dynamical networks is to investigate changes in the stochastic structure of networks, e.g. change in the latent structure, change in the degree distribution of graph, and anomaly detection in dynamical networks, etc. (Clegg, 2004; Ranshous et al., 2015). A common approach for monitoring change in the stochastic structure and temporal dependency of network data is to calculate graph Laplacian and modeling short-term and long-term dependencies by applying sliding windows (Huang et al., 2020).

Given the fact that graph Laplacian can be viewed as manifold space, the author intends to investigate the potential of change point detection on manifold space $SPD(d)$ for detecting changes in dynamic network. Statistical analysis on dynamic network can be carried out with the following goals:

- Theoretical properties of the test statistics developed for the space of graph Laplacian will be investigated. Robust statistics will be implemented by ranking the data with respect to their depth functions. Robustness analysis will be carried out using breakdown point.

- The generality of the proposed scheme will be evaluated considering various types of networks with different stochastic structures, e.g. homogeneous networks and heterogeneous networks with unknown latent structure.

- For graph evolution data, the goal is to develop a robust estimator which can identify the modes of variation and principal manifold components when the data is contaminated with outlier observations. In particular, the goal is to explore the convergence rate for such estimators with respect to the complexity of the underlying space and sample size. The validity of the theoretical findings, e.g. the efficiency of the robust estimator, will be checked versus methods available in the literature.

References

Aaron, C. and Bodart, O. (2018), 'Convergence rates for estimators of geodesic distances and fréchet expectations', *Journal of Applied Probability* **55**(4), 1001–1013.

Abdi, H. and Williams, L. J. (2010), 'Principal component analysis', *Wiley interdisciplinary reviews: computational statistics* **2**(4), 433–459.

Agostinelli, C. and Romanazzi, M. (2011), 'Local depth', *Journal of Statistical Planning and Inference* **141**(2), 817–830.

Akhøj, M., Pennec, X. and Sommer, S. (2022), 'Tangent phylogenetic pca', *arXiv:2208.12730* .

Alekseyenko, A. V. (2016), 'Multivariate welch t-test on distances', *Bioinformatics* **32**(23), 3552–3558.

Aly, E.-E. A. and BuHamra, S. S. (1996), 'Rank tests for two change points', *Computational statistics & data analysis* **22**(4), 363–372.

Ambellan, F., Zachow, S. and von Tycowicz, C. (2019), An as-invariant-as-possible-based statistical shape model, *in* 'Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy: 4th International Workshop, MBIA 2019, and 7th International Workshop, MFCA 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Proceedings', Springer, pp. 219–228.

Aminikhanghahi, S. and Cook, D. J. (2017), 'A survey of methods for time series change point detection', *Knowledge and information systems* **51**(2), 339–367.

Anastasiou, A. and Fryzlewicz, P. (2019), 'Detecting multiple generalized change-points by isolating single ones', *arXiv:1901.10852* .

Ansari, A. R. and Bradley, R. A. (1960), 'Rank-sum tests for dispersions', *The annals of mathematical statistics* pp. 1174–1189.

Antoch, J. and Hušková, M. (2001), 'Permutation tests in change point analysis', *Statistics & probability letters* **53**(1), 37–46.

Antoch, J. and Hušková, M. (2009), 'Data driven rank statistics in change point analysis', *Mathematica Slovaca* **59**(5), 545–564.

Antoch, J., Hušková, M. and Veraverbeke, N. (1995), 'Change-point problem and bootstrap', *Journaltitle of Nonparametric Statistics* **5**(2), 123–144.

Arsigny, V., Fillard, P., Pennec, X. and Ayache, N. (2007), 'Geometric means in a novel vector space structure on symmetric positive-definite matrices', *SIAM journal on matrix analysis and applications* **29**(1), 328–347.

Arunajadai, S. G. (2009), 'A point process driven multiple change point model: A robust resistant approach', *Mathematical biosciences* **220**(1), 57–71.

Asta, D. M. (2024), 'Non-parametric manifold learning', *Electronic Journal of Statistics* **18**(2), 3903–3930.

Auret, L. and Aldrich, C. (2010), 'Change point detection in time series data with random forests', *Control Engineering Practice* **18**(8), 990–1002.

Baker, A. (2003), *Matrix groups: An introduction to Lie group theory*, Springer Science & Business Media.

Baranowski, R., Chen, Y. and Fryzlewicz, P. (2019), 'Narrowest-over-threshold detection of multiple change points and change-point-like features', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **81**(3), 649–672.

Barden, D., Le, H. and Owen, M. (2013), 'Central limit theorems for fréchet means in the space of phylogenetic trees', *Electron. J. Probab* **18**(25), 1–25.

Baringhaus, L. (1991), 'Testing for spherical symmetry of a multivariate distribution', *The Annals of Statistics* pp. 899–917.

Basrak, B. (2010), 'Limit theorems for the inductive mean on metric trees', *Journal of applied probability* **47**(4), 1136–1149.

Bayram, M. and Arserim, M. A. (2021), 'Analysis of epileptic ieeg data by applying convolutional neural networks to low-frequency scalograms', *IEEE Access* **9**, 162520–162529.

Beran, R. and Fisher, N. I. (1998), 'Nonparametric comparison of mean directions or mean axes', *The Annals of Statistics* **26**(2), 472–493.

Bernstein, A. (2017), Manifold learning in machine vision and robotics, *in* '2016 International Conference on Robotics and Machine Vision', Vol. 10253, SPIE, pp. 81–86.

Berumen, S., Bechmann, F., Lindner, S., Kruth, J.-P. and Craeghs, T. (2010), 'Quality control of laser-and powder bed-based additive manufacturing (am) technologies', *Physics procedia* **5**, 617–622.

Bhattacharya, A. (2008), *Nonparametric statistics on manifolds with applications to shape spaces*, The University of Arizona.

Bhattacharya, R. and Patrangenaru, V. (2003), 'Large sample theory of intrinsic and extrinsic sample means on manifolds', *The Annals of Statistics* **31**(1), 1–29.

Bhattacharya, R. and Patrangenaru, V. (2005), 'Large sample theory of intrinsic and extrinsic sample means on manifolds: Ii', *Annals of Statistics* pp. 1225–1259.

Billera, L. J., Holmes, S. P. and Vogtmann, K. (2001), 'Geometry of the space of phylogenetic trees', *Advances in Applied Mathematics* **27**(4), 733–767.

Billingsley, P. (1968), *Convergence of Probability Measures*, John Wiley & Sons.

Blankertz, B., Muller, K.-R., Curio, G., Vaughan, T. M., Schalk, G., Wolpaw, J. R., Schlogl, A., Neuper, C., Pfurtscheller, G., Hinterberger, T. et al. (2004), 'The bci competition 2003: progress and perspectives in detection and discrimination of eeg single trials', *IEEE Transactions on Biomedical Engineering* **51**(6), 1044–1051.

Boldin, M. (2013), 'Robustness of sign tests for testing hypotheses about order of autoregression', *Theory of Probability & Its Applications* **57**(4), 663–670.

Boucheron, S., Lugosi, G. and Bousquet, O. (2003), Concentration inequalities, *in* 'Summer school on machine learning', Springer, pp. 208–240.

Brodsky, E. and Darkhovsky, B. S. (2013), *Nonparametric methods in change point problems*, Vol. 243, Springer Science & Business Media.

Brosch, T., Tam, R. and Initiative, A. D. N. (2013), Manifold learning of brain mris by deep learning, *in* 'Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II 16', Springer, pp. 633–640.

Canzani, Y. (2013), 'Analysis on manifolds via the laplacian', *Lecture Notes available at: http://www. math. harvard. edu/canzani/docs/Laplacian. pdf* pp. 41–44.

Centofanti, F., Lepore, A., Kulahci, M. and Spooner, M. P. (2022), 'Real-time monitoring of functional data', *arXiv:2205.06256* .

Chakraborty, B., Chaudhuri, P. and Oja, H. (1998), 'Operating transformation retransformation on spatial median and angle test', *Statistica Sinica* pp. 767–784.

Chakraborty, N. and Finkelstein, M. (2024), 'Distribution-free multivariate process monitoring: A rank-energy statistic-based approach', *Quality and Reliability Engineering International* **40**(7), 4068–4087.

Chakraborty, R. and Vemuri, B. C. (2019), 'Statistics on the stiefel manifold: Theory and applications', *The Annals of Statistics* **47**(1).

Chau, J., Ombao, H. and von Sachs, R. (2019), 'Intrinsic data depth for hermitian positive definite matrices', *Journal of Computational and Graphical Statistics* **28**(2), 427–439.

Chen, D. and Müller, H.-G. (2012), 'Nonlinear manifold representations for functional data', *The Annals of Statistics* **40**(1), 1–29.

Chen, H., Chen, X. and Su, Y. (2018), 'A weighted edge-count two-sample test for multivariate and object data', *Journal of the American Statistical Association* **113**(523), 1146–1155.

Chen, H. and Friedman, J. H. (2017), 'A new graph-based two-sample test for multivariate and object data', *Journal of the American statistical association* **112**(517), 397–409.

Chen, N., Zi, X. and Zou, C. (2016), 'A distribution-free multivariate control chart', *Technometrics* **58**(4), 448–459.

Cheng, L., Tsung, F. and Wang, A. (2017), 'A statistical transfer learning perspective for modeling shape deviations in additive manufacturing', *IEEE Robotics and Automation Letters* **2**(4), 1988–1993.

Cheng, X. and Xie, Y. (2021*a*), 'Kernel two-sample tests for manifold data', *arXiv:2105.03425* .

Cheng, X. and Xie, Y. (2021*b*), 'Neural tangent kernel maximum mean discrepancy', *Advances in Neural Information Processing Systems* **34**, 6658–6670.

Chenouri, S., Mozaffari, A. and Rice, G. (2020), 'Robust multivariate change point analysis based on data depth', *Canadian Journal of Statistics* **48**(3), 417–446.

Chenouri, S. and Small, C. G. (2012), 'A nonparametric multivariate multisample test based on data depth', *Electronic Journal of Statistics* **6**, 760–782.

Chernoff, H. and Savage, I. R. (1958), 'Asymptotic normality and efficiency of certain nonparametric test statistics', *The Annals of Mathematical Statistics* pp. 972–994.

Chernoff, H., Savage, I. R. and CA, S. U. (1957), Asymptotic normality and efficiency of certain nonparametric test statistics, *in* 'Annals of Mathematical Statistics', Vol. 28, INST MATHEMATICAL STATISTICS IMS BUSINESS OFFICE-SUITE 7, 3401 INVESTMENT . . . , pp. 1068–1068.

Chu, L. and Dai, X. (2024), 'Manifold energy two-sample test', *Electronic Journal of Statistics* **18**(1), 145–166.

Chua, Z. Y., Ahn, I. H. and Moon, S. K. (2017), 'Process monitoring and inspection systems in metal additive manufacturing: Status and applications', *International Journal of Precision Engineering and Manufacturing-Green Technology* **4**(2), 235–245.

Ciuperca, G. (2011), 'A general criterion to determine the number of change-points', *Statistics & probability letters* **81**(8), 1267–1275.

Clegg, R. G. (2004), The statistics of dynamic networks, PhD thesis, University of York.

Cohen, D. L. and Lipson, H. (2010), 'Geometric feedback control of discrete-deposition sff systems', *Rapid Prototyping Journal* **16**(5), 377–393.

Coifman, R. R. and Lafon, S. (2006), 'Diffusion maps', *Applied and computational harmonic analysis* **21**(1), 5–30.

Collyer, M. L., Sekora, D. J. and Adams, D. C. (2015), 'A method for analysis of phenotypic change for phenotypes described by high-dimensional data', *Heredity* **115**(4), 357–365.

Colosimo, B. M. (2018), 'Modeling and monitoring methods for spatial and image data', *Quality Engineering* **30**(1), 94–111.

Colosimo, B. M., Huang, Q., Dasgupta, T. and Tsung, F. (2018), 'Opportunities and challenges of quality engineering for additive manufacturing', *Journal of Quality Technology* **50**(3), 233–252.

Colosimo, B. M. and Pacella, M. (2010), 'A comparison study of control charts for statistical monitoring of functional data', *International Journal of Production Research* **48**(6), 1575–1601.

Covo, S. and Elalouf, A. (2014), 'A novel single-gamma approximation to the sum of independent gamma variables, and a generalization to infinitely divisible distributions', *Electronic Journal of Statistics* **8**(1), 894–926.

Crosier, R. B. (1988), 'Multivariate generalizations of cumulative sum quality-control schemes', *Technometrics* **30**(3), 291–303.

Dai, X., Lin, Z. and Müller, H.-G. (2021), 'Modeling sparse longitudinal data on riemannian manifolds', *Biometrics* **77**(4), 1328–1341.

Dai, X., Lopez-Pintado, S. and Initiative, A. D. N. (2022), 'Tukey's depth for object data', *Journal of the American Statistical Association* pp. 1–13.

Dai, X. and Müller, H.-G. (2018), 'Principal component analysis for functional data on riemannian manifolds and spheres', *The Annals of Statistics* **46**(6B), 3334–3361.

Dehghan, S. and Faridrohani, M. R. (2020), 'Depth-based signed-rank tests for bivariate central symmetry', *REVSTAT-Statistical Journal* **18**(1), 109–129.

del Castillo, E. and Zhao, X. (2014), 'Statistical process monitoring for manifold data', *Wiley StatsRef: Statistics Reference Online* pp. 1–8.

Deng, C., Chen, Z., Zhao, X., Wang, H., Wang, J., Chen, H. and Gao, J. (2024), 'Rio-cpd: A riemannian geometric method for correlation-aware online change point detection', *arXiv:2407.09698* .

Deng, S., Ling, S. and Strohmer, T. (2021), 'Strong consistency, graph laplacians, and the stochastic block model', *The Journal of Machine Learning Research* **22**(1), 5210–5253.

Devroye, L., Györfi, L., Lugosi, G., Devroye, L., Györfi, L. and Lugosi, G. (1996), 'Vapnik-chervonenkis theory', *A probabilistic theory of pattern recognition* pp. 187–213.

Diaconis, P., Holmes, S. and Shahshahani, M. (2013), Sampling from a manifold, *in* 'Advances in modern statistical theory and applications: a Festschrift in honor of Morris L. Eaton', Vol. 10, Institute of Mathematical Statistics, pp. 102–126.

Ding, D., Tsung, F. and Li, J. (2016), 'Rank-based process control for mixed-type data', *IIE Transactions* **48**(7), 673–683.

Dryden, I., Hill, B., Wang, H. and Laughton, C. (2017), 'Covariance analysis for temporal data, with applications to dna modelling', *Stat* **6**(1), 218–230.

Dryden, I. L., Koloydenko, A. and Zhou, D. (2009), 'Non-euclidean statistics for covariance matrices, with applications to diffusion tensor imaging', *The Annals of Applied Statistics* **3**(3), 1102–1123.

Dryden, I. L. and Mardia, K. V. (2016), *Statistical shape analysis: with applications in R*, John Wiley & Sons.

Duan, X., Sun, H. and Peng, L. (2013), 'Riemannian means on special euclidean group and unipotent matrices group', *The Scientific World Journal* **2013**.

Duan, X., Sun, H. and Zhao, X. (2019), 'A matrix information-geometric method for change-point detection of rigid body motion', *Entropy* **21**(5), 531.

Dubey, P. and Müller, H.-G. (2019), 'Fréchet analysis of variance for random objects', *Biometrika* **106**(4), 803–821.

Dubey, P. and Müller, H.-G. (2020), 'Fréchet change-point detection', *The Annals of Statistics* **48**(6), 3312–3335.

Dubhashi, D. P. and Panconesi, A. (2009), *Concentration of measure for the analysis of randomized algorithms*, Cambridge University Press.

Dudley, R. M. (1978), 'Central limit theorems for empirical measures', *The Annals of Probability* pp. 899–929.

Dudley, R. M. (2010), Universal donsker classes and metric entropy, *in* 'Selected Works of RM Dudley', Springer, pp. 345–365.

Dudley, R. M., Giné, E. and Zinn, J. (1991), 'Uniform and universal glivenko-cantelli classes', *Journal of Theoretical Probability* **4**(3), 485–510.

Dyckerhoff, R., Ley, C. and Paindaveine, D. (2015), 'Depth-based runs tests for bivariate central symmetry', *Annals of the Institute of Statistical Mathematics* **67**, 917–941.

Everton, S. K., Hirsch, M., Stravroulakis, P., Leach, R. K. and Clare, A. T. (2016), 'Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing', *Materials & Design* **95**, 431–445.

Feige, U., Mansour, Y. and Schapire, R. E. (2018), Robust inference for multiclass classification, *in* 'Algorithmic Learning Theory', PMLR, pp. 368–386.

Ferri, M. and Frosini, P. (2008), 'Vc-dimension on manifolds: a first approach', *Mathematical methods in the applied sciences* **31**(5), 589–605.

Fienberg, S. E. (2012), 'A brief history of statistical models for network analysis and open challenges', *Journal of Computational and Graphical Statistics* **21**(4), 825–839.

Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2008), *Longitudinal data analysis*, CRC press.

Fletcher, P. T. and Joshi, S. (2004), Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors, *in* 'Computer vision and mathematical methods in medical and biomedical image analysis', Springer, pp. 87–98.

Fletcher, P. T. and Joshi, S. (2007), 'Riemannian geometry for the statistical analysis of diffusion tensor data', *Signal Processing* **87**(2), 250–262.

Fletcher, P. T., Venkatasubramanian, S. and Joshi, S. (2008), Robust statistics on riemannian manifolds via the geometric median, *in* '2008 IEEE Conference on Computer Vision and Pattern Recognition', IEEE, pp. 1–8.

Forbes, C., Evans, M., Hastings, N. and Peacock, B. (2011), *Statistical distributions*, John Wiley & Sons.

Fraiman, R., Gamboa, F. and Moreno, L. (2019), 'Connecting pairwise geodesic spheres by depth: Dcops', *Journal of Multivariate Analysis* **169**, 81–94.

Fryzlewicz, P. (2014), 'Wild binary segmentation for multiple change-point detection', *The Annals of Statistics* **42**(6), 2243–2281.

Fryzlewicz, P. (2018), 'Tail-greedy bottom-up data decompositions and fast multiple change-point detection', *The Annals of Statistics* **46**(6B), 3390–3421.

Galeano, P. and Wied, D. (2017), 'Dating multiple change points in the correlation matrix', *Test* **26**(2), 331–352.

Ge, Z. and Song, Z. (2012), *Multivariate statistical process control: Process monitoring methods and applications*, Springer Science & Business Media.

Ge, Z., Song, Z. and Gao, F. (2013), 'Review of recent research on data-based process monitoring', *Industrial & Engineering Chemistry Research* **52**(10), 3543–3562.

Geller, D. and Mayeli, A. (2009), 'Continuous wavelets on compact manifolds', *Mathematische Zeitschrift* **262**(4), 895–927.

Geller, D. and Pesenson, I. Z. (2011), 'Band-limited localized parseval frames and besov spaces on compact homogeneous manifolds', *Journal of Geometric Analysis* **21**(2), 334–371.

Ghosh, S. and Ruymgaart, F. H. (1992), 'Applications of empirical characteristic functions in some multivariate problems', *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* pp. 429–440.

Gibbs, A. L. and Su, F. E. (2002), 'On choosing and bounding probability metrics', *International statistical review* **70**(3), 419–435.

Giné, E. and Nickl, R. (2009), 'Uniform limit theorems for wavelet density estimators', *The Annals of Probability* **37**(4), 1605–1646.

Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S. and Kolaczyk, E. D. (2017), 'Hypothesis testing for network data in functional neuroimaging', *The Annals of Applied Statistics* pp. 725–750.

Gombay, E. and Hušková, M. (1998), 'Rank based estimators of the change-point', *Journal of statistical planning and inference* **67**(1), 137–154.

Grasso, M., Demir, A., Previtali, B. and Colosimo, B. (2018), 'In situ monitoring of selective laser melting of zinc powder via infrared imaging of the process plume', *Robotics and Computer-Integrated Manufacturing* **49**, 229–239.

Grattarola, D., Zambon, D., Alippi, C. and Livi, L. (2018), 'Learning graph embeddings on constant-curvature manifolds for change detection in graph streams', *stat* **1050**, 16.

Grattarola, D., Zambon, D., Livi, L. and Alippi, C. (2019), 'Change detection in graph streams by learning graph embeddings on constant-curvature manifolds', *IEEE Transactions on neural networks and learning systems* **31**(6), 1856–1869.

Groeneboom, P. and Wellner, J. A. (2001), 'Computing chernoff's distribution', *Journal of Computational and Graphical Statistics* **10**(2), 388–400.

Gromov, M., Katz, M., Pansu, P. and Semmes, S. (1999), *Metric structures for Riemannian and non-Riemannian spaces*, Vol. 152, Springer.

Guo, N. and Leu, M. C. (2013), 'Additive manufacturing: technology, applications and research needs', *Frontiers of Mechanical Engineering* **8**(3), 215–243.

Haanchumpol, T., Sudasna-na Ayudthya, P. and Singhtaun, C. (2020), 'Modern multivariate control chart using spatial signed rank for non-normal process', *Engineering Science and Technology, an International Journal* **23**(4), 859–869.

Habib, M. (1998), *Probabilistic methods for algorithmic discrete mathematics*, Vol. 16, Springer Science & Business Media.

Ham, J., Lee, D. D., Mika, S. and Schölkopf, B. (2004), A kernel view of the dimensionality reduction of manifolds, *in* 'Proceedings of the twenty-first international conference on Machine learning', p. 47.

Hampel, F. R. (1971), 'A general qualitative definition of robustness', *The annals of mathematical statistics* **42**(6), 1887–1896.

Han, J., Hu, R. and Long, J. (2023), 'A class of dimension-free metrics for the convergence of empirical measures', *Stochastic Processes and their Applications* **164**, 242–287.

Hanik, M., Hege, H.-C. and Tycowicz, C. v. (2020), Bi-invariant two-sample tests in lie groups for shape analysis, *in* 'International Workshop on Shape in Medical Imaging', Springer, pp. 44–54.

Hauberg, S. (2018), Directional statistics with the spherical normal distribution, *in* '2018 21st International Conference on Information Fusion (FUSION)', IEEE, pp. 704–711.

He, X., Simpson, D. G. and Portnoy, S. L. (1990), 'Breakdown robustness of tests', *Journal of the American Statistical Association* **85**(410), 446–452.

Hendriks, H. and Landsman, Z. (1998), 'Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions', *Journal of Multivariate Analysis* **67**(2), 227–243.

Hendriks, H. and Landsman, Z. (2007), 'Asymptotic data analysis on manifolds', *The Annals of Statistics* **35**(1), 109–131.

Hettmansperger, T. P. and Randles, R. H. (2002), 'A practical affine equivariant multivariate median', *Biometrika* **89**(4), 851–860.

Hoeffding, W. (1951), " optimum" nonparametric tests, *in* 'Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability', The Regents of the University of California.

Hoeffding, W. (1963), 'Probability inequalities for sums of bounded random variables', *Journal of the American statistical association* **58**(301), 13–30.

Horváth, L. and Rice, G. (2014), 'Extensions of some classical methods in change point analysis', *Test* **23**(2), 219–255.

Huang, Q. (2016), 'An analytical foundation for optimal compensation of three-dimensional shape deformation in additive manufacturing', *Journal of Manufacturing Science and Engineering* **138**(6), 061010.

Huang, Q., Nouri, H., Xu, K., Chen, Y., Sosina, S. and Dasgupta, T. (2014), 'Statistical predictive modeling and compensation of geometric deviations of three-dimensional printed products', *Journal of Manufacturing Science and Engineering* **136**(6), 061008.

Huang, Q., Zhang, J., Sabbaghi, A. and Dasgupta, T. (2015), 'Optimal offline compensation of shape shrinkage for three-dimensional printing processes', *Iie transactions* **47**(5), 431–441.

Huang, S. H., Liu, P., Mokasdar, A. and Hou, L. (2013), 'Additive manufacturing and its societal impact: a literature review', *The International Journal of Advanced Manufacturing Technology* **67**(5), 1191–1203.

Huang, S., Hitti, Y., Rabusseau, G. and Rabbany, R. (2020), Laplacian change point detection for dynamic graphs, *in* 'Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining', pp. 349–358.

Huang, Z., Cai, H., Dan, T., Lin, Y., Laurienti, P. and Wu, G. (2021), Detecting brain state changes by geometric deep learning of functional dynamics on riemannian manifold, *in* 'Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII 24', Springer, pp. 543–552.

Huber, P. J. (2011), Robust statistics, *in* 'International encyclopedia of statistical science', Springer, pp. 1248–1251.

Huckemann, S. F. and Eltzner, B. (2021), 'Data analysis on nonstandard spaces', *Wiley Interdisciplinary Reviews: Computational Statistics* **13**(3), e1526.

Hušková, M. (2003), 'Serial rank statistics for detection of changes', *Statistics & Probability Letters* **61**(2), 199–213.

Hušková, M. (2004), 'Weak invariance principles for regression rank statistics', *Sequential Analysis* **23**(1), 121–140.

Hušková, M. (2013), Robust change point analysis, *in* 'Robustness and Complex Data Structures', Springer, pp. 171–190.

Hušková, M. and Meintanis, S. G. (2006), 'Change-point analysis based on empirical characteristic functions of ranks', *Sequential Analysis* **25**(4), 421–436.

Hušková, M. and Neuhaus, G. (2004), 'Change point analysis for censored data', *Journal of statistical planning and inference* **126**(1), 207–223.

Izenman, A. J. (2012), 'Introduction to manifold learning', *Wiley Interdisciplinary Reviews: Computational Statistics* **4**(5), 439–446.

James, I. M. (1976), *The topology of Stiefel manifolds*, Vol. 24, Cambridge University Press.

James, N. A. and Matteson, D. S. (2015), 'ecp: An r package for nonparametric multiple change point analysis of multivariate data', *Journal of Statistical Software* **62**, 1–25.

Jayasumana, S., Hartley, R., Salzmann, M., Li, H. and Harandi, M. (2013), Kernel methods on the riemannian manifold of symmetric positive definite matrices, *in* 'proceedings of the IEEE Conference on Computer Vision and Pattern Recognition', pp. 73–80.

Jayasumana, S., Hartley, R., Salzmann, M., Li, H. and Harandi, M. (2015), 'Kernel methods on riemannian manifolds with gaussian rbf kernels', *IEEE transactions on pattern analysis and machine intelligence* **37**(12), 2464–2477.

Jin, Y., Joe Qin, S. and Huang, Q. (2016), 'Offline predictive control of out-of-plane shape deformation for additive manufacturing', *Journal of Manufacturing Science and Engineering* **138**(12), 121005.

Joe Qin, S. (2003), 'Statistical process monitoring: basics and beyond', *Journal of Chemometrics: A Journal of the Chemometrics Society* **17**(8-9), 480–502.

Kasarapu, P. (2015), 'Modelling of directional data using kent distributions', *arXiv preprint arXiv:1506.08105* .

Katsevich, A. and Ramm, A. (1993), 'Consistency of rank tests against some general alternatives', *Mathematical and computer modelling* **18**(12), 49–55.

Kendall, M. (1949), 'Rank correlation methods. by maurice g. kendall, ma [pp. vii+ 160. london: Charles griffin and co. ltd., 42 drury lane, 1948. 18 s]', *J. Inst. Actuar* **75**, 140–141.

Kerkyacharian, G., Nickl, R. and Picard, D. (2012), 'Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds', *Probability Theory and Related Fields* **153**(1), 363–404.

Keshavarzkermani, A., Marzbanrad, E., Esmaeilizadeh, R., Mahmoodkhani, Y., Ali, U., Enrique, P. D., Zhou, N. Y., Bonakdar, A. and Toyserkani, E. (2019), 'An investigation into the effect of process parameters on melt pool geometry, cell spacing, and grain refinement during laser powder bed fusion', *Optics & Laser Technology* **116**, 83–91.

Kiefer, J. (1959), 'K-sample analogues of the kolmogorov-smirnov and cramér-v. mises tests', *The Annals of Mathematical Statistics* pp. 420–447.

Koul, H. L. (1970), 'Some convergence theorems for ranks and weighted empirical cumulatives', *The Annals of Mathematical Statistics* pp. 1768–1773.

Koziol, J. A. (1996), 'A note on signed rank tests for the changepoint problem', *Statistics* **27**(3-4), 325–338.

Kozlov, S. E. (1997), 'Geometry of the real grassmannian manifolds. parts i, ii', *Zapiski Nauchnykh Seminarov POMI* **246**, 84–107.

Kresta, J. V., Macgregor, J. F. and Marlin, T. E. (1991), 'Multivariate statistical monitoring of process operating performance', *The Canadian journal of chemical engineering* **69**(1), 35–47.

Kruskal, W. H. (1958), 'Ordinal measures of association', *Journal of the American Statistical Association* **53**(284), 814–861.

Kruskal, W. H. and Wallis, W. A. (1952), 'Use of ranks in one-criterion variance analysis', *Journal of the American statistical Association* **47**(260), 583–621.

Kuo, C.-H., Wares, J. P. and Kissinger, J. C. (2008), 'The apicomplexan whole-genome phylogeny: an analysis of incongruence among gene trees', *Molecular biology and evolution* **25**(12), 2689–2698.

Lambert, D. (1982), 'Qualitative robustness of tests', *Journal of the American Statistical Association* **77**(378), 352–357.

Lee, C. and Wilkinson, D. J. (2019), 'A review of stochastic block models and extensions for graph clustering', *Applied Network Science* **4**(1), 1–50.

Lee, J. M., Chow, B., Chu, S.-C., Glickenstein, D., Guenther, C., Isenberg, J., Ivey, T., Knopf, D., Lu, P., Luo, F. et al. (2009), 'Manifolds and differential geometry', *Topology* **643**, 658.

Lee, T.-S. (2010), 'Change-point problems: bibliography and review', *Journal of Statistical Theory and Practice* **4**(4), 643–662.

Lenglet, C., Rousson, M., Deriche, R. and Faugeras, O. (2006), 'Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor mri processing', *Journal of Mathematical Imaging and Vision* **25**(3), 423–444.

Li, J., Zhang, X. and Jeske, D. R. (2013), 'Nonparametric multivariate cusum control charts for location and scale changes', *Journal of Nonparametric Statistics* **25**(1), 1–20.

Li, W., Pu, X., Tsung, F. and Xiang, D. (2017), 'A robust self-starting spatial rank multivariate ewma chart based on forward variable selection', *Computers & Industrial Engineering* **103**, 116–130.

Lin, F.-H. (2008), 'On the isometric embedding of torus in r3', *Methods and Applications of Analysis* **15**(2), 197–204.

Lin, L., Lazar, D., Sarpabayeva, B. and Dunson, D. B. (2020), 'Robust optimization and inference on manifolds', *arXiv:2006.06843* .

Liu, C., Kong, Z., Babu, S., Joslin, C. and Ferguson, J. (2021), 'An integrated manifold learning approach for high-dimensional data feature extractions and its applications to online process monitoring of additive manufacturing', *IISE Transactions* **53**(11), 1215–1230.

Liu, L., Yu, L., Kubatko, L., Pearl, D. K. and Edwards, S. V. (2009), 'Coalescent methods for estimating phylogenetic trees', *Molecular phylogenetics and evolution* **53**(1), 320–328.

Liu, Q., Li, C., Wanga, V. and Shepherd, B. E. (2018), 'Covariate-adjusted spearman's rank correlation with probability-scale residuals', *Biometrics* **74**(2), 595–605.

Liu, R. Y. (1990), 'On a notion of data depth based on random simplices', *The Annals of Statistics* pp. 405–414.

Liu, R. Y. and Singh, K. (1992), 'Ordering directional data: concepts of data depth on circles and spheres', *The Annals of Statistics* **20**(3), 1468–1484.

Liu, R. Y. and Singh, K. (1993), 'A quality index based on data depth and multivariate rank tests', *Journal of the American Statistical Association* **88**(421), 252–260.

Liu, R. Y., Singh, K. and Teng*, J. H. (2004), 'Ddma-charts: nonparametric multivariate moving average control charts based on data depth', *Allgemeines Statistisches Archiv* **88**(2), 235–258.

Liu, X. and Briol, F.-X. (2024), 'On the robustness of kernel goodness-of-fit tests', *arXiv:2408.05854* .

Loh, P.-L. (2024), 'A theoretical review of modern robust statistics', *Annual Review of Statistics and Its Application* **12**, 477–496.

Lombard, F. (1987), 'Rank tests for changepoint problems', *Biometrika* **74**(3), 615–624.

Lu, L., Zheng, J. and Mishra, S. (2014), 'A layer-to-layer model and feedback control of ink-jet 3-d printing', *IEEE/ASME Transactions on Mechatronics* **20**(3), 1056–1068.

Luan, H. and Huang, Q. (2016), 'Prescriptive modeling and compensation of in-plane shape deformation for 3-d printed freeform products', *IEEE Transactions on Automation Science and Engineering* **14**(1), 73–82.

Ma, X., Kirby, M. and Peterson, C. (2020), 'The flag manifold as a tool for analyzing and comparing data sets', *arXiv preprint arXiv:2006.14086* .

Mann, H. B. and Whitney, D. R. (1947), 'On a test of whether one of two random variables is stochastically larger than the other', *The annals of mathematical statistics* pp. 50–60.

Mardia, K. V. and Patrangenaru, V. (2005), 'Directions and projective shapes1', *The Annals of Statistics* **33**(4), 1666–1699.

Maronna, R. A., Martin, R. D., Yohai, V. J. and Salibián-Barrera, M. (2019), *Robust statistics: theory and methods (with R)*, John Wiley & Sons.

Marron, J. S. and Dryden, I. L. (2021), *Object oriented data analysis*, Chapman and Hall/CRC.

Massart, P. (2007), *Concentration inequalities and model selection: Ecole d'Eté de Probabilités de Saint-Flour XXXIII-2003*, Springer.

Maurer, A. (2019), 'A bernstein-type inequality for functions of bounded interaction', *Bernoulli* **25**(2), 1451–1471.

Megahed, F. M., Woodall, W. H. and Camelio, J. A. (2011), 'A review and perspective on control charting with image data', *Journal of quality technology* **43**(2), 83–98.

Melchels, F. P., Feijen, J. and Grijpma, D. W. (2010), 'A review on stereolithography and its applications in biomedical engineering', *Biomaterials* **31**(24), 6121–6130.

Merlo, J., Cordero-Franco, A. E. and Tercero-Gómez, V. G. (2022), 'Nonparametric multivariate processes monitoring with guaranteed in-control performance for changes in location', *Computers & Industrial Engineering* **166**, 107940.

Miranda, G. F., Thomaz, C. E. and Giraldi, G. A. (2017), Geometric data analysis based on manifold learning with applications for image understanding, *in* '2017 30th SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T)', IEEE, pp. 42–62.

Mitteroecker, P., Bartsch, S., Erkinger, C., Grunstra, N. D., Le Maître, A. and Bookstein, F. L. (2020), 'Morphometric variation at different spatial scales: coordination and compensation in the emergence of organismal form', *Systematic biology* **69**(5), 913–926.

Mizera, I. (2010), Qualitative robustness and weak continuity: the extreme unction?, *in* 'Nonparametrics and robustness in modern statistical inference and time series analysis: a Festschrift in honor of Professor Jana Jurečková', Vol. 7, Institute of Mathematical Statistics, pp. 169–182.

Montgomery, D. C. (2009), *Statistical quality control*, Vol. 7, Wiley New York.

Moschopoulos, P. G. (1985), 'The distribution of the sum of independent gamma random variables', *Annals of the Institute of Statistical Mathematics* **37**(1), 541–544.

Mosler, K. (2013), 'Depth statistics', *Robustness and complex data structures: Festschrift in Honour of Ursula Gather* pp. 17–34.

Mukherjee, A. and Marozzi, M. (2021), 'Nonparametric phase-ii control charts for monitoring high-dimensional processes with unknown parameters', *Journal of Quality Technology* **54**(1), 44–64.

Müller, A. (1997), 'Integral probability metrics and their generating classes of functions', *Advances in applied probability* **29**(2), 429–443.

Muralidharan, P. and Fletcher, P. T. (2012), Sasaki metrics for analysis of longitudinal data on manifolds, *in* '2012 IEEE conference on computer vision and pattern recognition', IEEE, pp. 1027–1034.

Murray, M. K. and Rice, J. W. (1993), *Differential geometry and statistics*, Vol. 48, CRC Press.

Nikolaos, K., Rafail, V. and Michail, T. (2025), 'On independence testing using the (partial) distance correlation', *arXiv:2506.15659* .

Niu, Y. S., Hao, N. and Zhang, H. (2016), 'Multiple change-point detection: A selective overview', *Statistical Science* pp. 611–623.

Niven, R. K. (2009), 'Combinatorial entropies and statistics', *The European Physical Journal B* **70**(1), 49–63.

Nosedal-Sanchez, A., Storlie, C. B., Lee, T. C. and Christensen, R. (2012), 'Reproducing kernel hilbert spaces for penalized regression: A tutorial', *The American Statistician* **66**(1), 50–60.

Nydick, S. W. (2012), 'The wishart and inverse wishart distributions', *Electronic Journal of Statistics* **6**(1-19).

Nye, T. M., Tang, X., Weyenberg, G. and Yoshida, R. (2017), 'Principal component analysis and the locus of the fréchet mean in the space of phylogenetic trees', *Biometrika* **104**(4), 901–922.

Oja, H. (1983), 'Descriptive statistics for multivariate distributions', *Statistics & Probability Letters* **1**(6), 327–332.

Osborne, D., Patrangenaru, V., Ellingson, L., Groisser, D. and Schwartzman, A. (2013), 'Nonparametric two-sample tests on homogeneous riemannian manifolds, cholesky decompositions and diffusion tensor image analysis', *Journal of Multivariate Analysis* **119**, 163–175.

O'Hara, R. and Sillanpää, M. (2009), 'A review of bayesian variable selection methods: What, how and which', *Bayesian Analysis* **4**(1), 85–118.

Page, E. (1955), 'A test for a change in a parameter occurring at an unknown point', *Biometrika* **42**(3/4), 523–527.

Paindaveine, D. and Van Bever, G. (2018), 'Halfspace depths for scatter, concentration and shape matrices', *The Annals of Statistics* **46**(6B), 3276–3307.

Pandolfo, G., Iorio, C., Staiano, M., Aria, M. and Siciliano, R. (2021), 'Multivariate process control charts based on the l p depth', *Applied Stochastic Models in Business and Industry* **37**(2), 229–250.

Papantoni-Kazakos, P. (1984), Some aspects of qualitative robustness in time series, *in* 'Robust and Nonlinear Time Series Analysis: Proceedings of a Workshop Organized by the Sonderforschungsbereich 123 "Stochastische Mathematische Modelle", Heidelberg 1983', Springer, pp. 218–230.

Patrangenaru, V. (2001), 'New large sample and bootstrap methods on shape spaces in high level analysis of natural images', *Communications in Statistics-Theory and Methods* **30**(8-9), 1675–1693.

Pennec, X. (2006), 'Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements', *Journal of Mathematical Imaging and Vision* **25**(1), 127–154.

Pennec, X., Fillard, P. and Ayache, N. (2006), 'A riemannian framework for tensor computing', *International Journal of computer vision* **66**(1), 41–66.

Pierre-Jean, M., Rigaill, G. and Neuvial, P. (2015), 'Performance evaluation of dna copy number segmentation methods', *Briefings in bioinformatics* **16**(4), 600–615.

Pitcan, Y. (2017), 'A note on concentration inequalities for u-statistics', *arXiv:1712.06160* .

Polly, P. D., Lawing, A. M., Fabre, A.-C. and Goswami, A. (2013), 'Phylogenetic principal components analysis and geometric morphometrics', *Hystrix* **24**(1), 33.

Qian, C. (2017), 'Statistical issues in 3d printing', *MSc. Thesis, University of Waterloo* .

Qin, S. J. (2012), 'Survey on data-driven industrial process monitoring and diagnosis', *Annual reviews in control* **36**(2), 220–234.

Qiu, A., Younes, L. and Miller, M. I. (2008), 'Intrinsic and extrinsic analysis in computational anatomy', *Neuroimage* **39**(4), 1803–1814.

Qiu, P. and Hawkins, D. (2001), 'A rank-based multivariate cusum procedure', *Technometrics* **43**(2), 120–132.

Raducanu, B. and Dornaika, F. (2012), 'A supervised non-linear dimensionality reduction approach for manifold learning', *Pattern Recognition* **45**(6), 2432–2444.

Ramsay, K. (2022), Data Depth Inference for Difficult Data, PhD thesis, University of Waterloo.

Ranshous, S., Shen, S., Koutra, D., Harenberg, S., Faloutsos, C. and Samatova, N. F. (2015), 'Anomaly detection in dynamic networks: a survey', *Wiley Interdisciplinary Reviews: Computational Statistics* **7**(3), 223–247.

Ridwan, S. (2015), *Process monitoring in additive manufacturing aimed toward part qualification*, The University of Texas at El Paso.

Rieder, H. (1982), 'Qualitative robustness of rank tests', *The Annals of Statistics* **10**(1), 205–211.

Robinson, D. F. and Foulds, L. R. (1981), 'Comparison of phylogenetic trees', *Mathematical biosciences* **53**(1-2), 131–147.

Roy, S., Sarkar, A., Ghosh, A. and Basu, A. (2023), 'Asymptotic breakdown point analysis for a general class of minimum divergence estimators', *arXiv:2304.07466* .

Rudnick, Z. and Wigman, I. (2008), On the volume of nodal sets for eigenfunctions of the laplacian on the torus, *in* 'Annales Henri Poincare', Vol. 9, Springer, pp. 109–130.

Rué-Queralt, J., Stevner, A., Tagliazucchi, E., Laufs, H., Kringelbach, M. L., Deco, G. and Atasoy, S. (2021), 'Decoding brain states on the intrinsic manifold of human brain dynamics across wakefulness and sleep', *Communications Biology* **4**(1), 854.

Saad-Falcon, A., Ancelin, B. and Romberg, J. (2024), Subspace tracking with dynamical models on the grassmannian, *in* '2024 IEEE 13rd Sensor Array and Multichannel Signal Processing Workshop (SAM)', IEEE, pp. 1–5.

Sabbaghi, A., Dasgupta, T., Huang, Q. and Zhang, J. (2014), 'Inference for deformation and interference in 3d printing', *The Annals of Applied Statistics* pp. 1395–1415.

Saha, E., Sarkar, S. and Ghosh, A. K. (2017), 'Some high-dimensional one-sample tests based on functions of interpoint distances', *Journal of Multivariate Analysis* **161**, 83–95.

Schiller, G. (2015), Additive manufacturing for aerospace, *in* '2015 IEEE Aerospace Conference', IEEE, pp. 1–8.

Sen, B. (2018), 'A gentle introduction to empirical process theory and applications', *Lecture Notes, Columbia University* **11**, 28–29.

Shepherd, B. E., Li, C. and Liu, Q. (2016), 'Probability-scale residuals for continuous, discrete, and censored data', *Canadian Journal of Statistics* **44**(4), 463–479.

Shu, L. and Fan, J. (2018), 'A distribution-free control chart for monitoring high-dimensional processes based on interpoint distances', *Naval Research Logistics (NRL)* **65**(4), 317–330.

Sidak, Z., Sen, P. K. and Hajek, J. (1999), *Theory of rank tests*, Elsevier.

Siegel, S. and Tukey, J. W. (1960), 'A nonparametric sum of ranks procedure for relative spread in unpaired samples', *Journal of the American Statistical Association* **55**(291), 429–445.

Srivastava, A. and Klassen, E. P. (2016), *Functional and shape data analysis*, Vol. 1, Springer.

Sugiura, N. and Ogden, R. (1994), 'Testing change-points with linear trend', *Communications in Statistics-Simulation and Computation* **23**(2), 287–322.

Sun, C., Zhang, R., Lu, Y., Cui, Y., Deng, Z., Cao, D. and Khajepour, A. (2023), 'Toward ensuring safety for autonomous driving perception: Standardization progress, research advances, and perspectives', *IEEE Transactions on Intelligent Transportation Systems* **25**(5), 3286–3304.

Sun, Y. and Genton, M. G. (2011), 'Functional boxplots', *Journal of Computational and Graphical Statistics* **20**(2), 316–334.

Taherkhani, K. (2022), In-Situ Monitoring and Quality Assurance Algorithms for Laser Powder-Bed Fusion, PhD thesis, University of Waterloo.

Talagrand, M. (1994), 'Sharper bounds for gaussian and empirical processes', *The Annals of Probability* pp. 28–76.

Tapia, G. and Elwany, A. (2014), 'A review on process monitoring and control in metal-based additive manufacturing', *Journal of Manufacturing Science and Engineering* **136**(6), 060801.

Tenenbaum, J. B., Silva, V. d. and Langford, J. C. (2000), 'A global geometric framework for nonlinear dimensionality reduction', *science* **290**(5500), 2319–2323.

Terry, M. E. (1952), 'Some rank order tests which are most powerful against specific parametric alternatives', *The Annals of Mathematical Statistics* pp. 346–366.

Tibshirani, R. J. and Efron, B. (1993), 'An introduction to the bootstrap', *Monographs on statistics and applied probability* **57**(1).

Truong, C., Oudre, L. and Vayatis, N. (2020), 'Selective review of offline change point detection methods', *Signal Processing* **167**, 107299.

Tukey, J. W. (1975), Mathematics and the picturing of data, *in* 'Proceedings of the international congress of mathematicians', Vol. 2, Vancouver, pp. 523–531.

Turaga, P., Veeraraghavan, A. and Chellappa, R. (2008), Statistical analysis on stiefel and grassmann manifolds with applications in computer vision, *in* '2008 IEEE conference on computer vision and pattern recognition', IEEE, pp. 1–8.

Van der Vaart, A. W. (2000), *Asymptotic statistics*, Vol. 3, Cambridge university press.

Van der Waerden, B. (1952), Order tests for the two-sample problem and their power, *in* 'Indagationes Mathematicae (Proceedings)', Vol. 55, Elsevier, pp. 453–458.

Vellekoop, M. H. and Clark, J. M. C. (2006), 'A nonlinear filtering approach to change-point detection problems: Direct and differential-geometric methods', *SIAM review* **48**(2), 329–356.

Venkatraman, E. S. (1993), 'Consistency results in multiple change-point problems', *Technical Report, Stanford University* **24**.

Venter, J. and Steel, S. (1996), 'Finding multiple abrupt change points', *Computational statistics & data analysis* **22**(5), 481–504.

Virta, J. (2023), 'Spatial depth for data in metric spaces', *arXiv:2306.09740* .

Vostrikova, L. Y. (1981), Detecting "disorder" in multidimensional random processes, *in* 'Doklady Akademii Nauk', Vol. 259, Russian Academy of Sciences, pp. 270–274.

Wang, A., Song, S., Huang, Q. and Tsung, F. (2016), 'In-plane shape-deviation modeling and compensation for fused deposition modeling processes', *IEEE Transactions on Automation Science and Engineering* **14**(2), 968–976.

Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016), 'Functional data analysis', *Annual Review of Statistics and its application* **3**, 257–295.

Wang, X., Borsoi, R. A. and Richard, C. (2024), Non-parametric online change point detection on riemannian manifolds, *in* '41st International Conference on Machine Learning, ICML 2024'.

Wang, X., Borsoi, R., Breloy, A. and Richard, C. (2025), 'Riemannian change point detection on manifolds with robust centroid estimation', *arXiv:2508.18045* .

Wang, Y., Wang, Z. and Zi, X. (2020), 'Rank-based multiple change-point detection', *Communications in Statistics-Theory and Methods* **49**(14), 3438–3454.

Wang, Z., Goedhart, R. and Zwetsloot, I. M. (2023), 'Monitoring high-dimensional heteroscedastic processes using rank-based ewma methods', *Computers & Industrial Engineering* **184**, 109544.

Weinberger, K. Q., Sha, F. and Saul, L. K. (2004), Learning a kernel matrix for nonlinear dimensionality reduction, *in* 'Proceedings of the twenty-first international conference on Machine learning', p. 106.

Weyenberg, G., Huggins, P. M., Schardl, C. L., Howe, D. K. and Yoshida, R. (2014), 'Kdetrees: non-parametric estimation of phylogenetic tree distributions', *Bioinformatics* **30**(16), 2280–2287.

Weyenberg, G., Yoshida, R. and Howe, D. (2016), 'Normalizing kernels in the billera-holmes-vogtmann treespace', *IEEE/ACM transactions on computational biology and bioinformatics* **14**(6), 1359–1365.

Wilcoxon, F. (1945), 'Individual comparisons by ranking methods', *Biometrics* **1**(6), 80–83.

Wu, M. and Fitzgerald, W. (1995), 'Analytical approach to changepoint detection in laplacian noise', *IEE Proceedings-Vision, Image and Signal Processing* **142**(3), 174–180.

Xie, W., Kurtek, S., Bharath, K. and Sun, Y. (2017), 'A geometric approach to visualization of variability in functional data', *Journal of the American Statistical Association* **112**(519), 979–993.

Xie, Y., Huang, J. and Willett, R. (2012), 'Change-point detection for high-dimensional time series with missing data', *IEEE Journal of Selected Topics in Signal Processing* **7**(1), 12–27.

Xie, Y. and Siegmund, D. (2013), Sequential multi-sensor change-point detection, *in* '2013 Information Theory and Applications Workshop (ITA)', IEEE, pp. 1–20.

Xu, K., Kwok, T.-H., Zhao, Z. and Chen, Y. (2017), 'A reverse compensation framework for shape deformation control in additive manufacturing', *Journal of Computing and Information Science in Engineering* **17**(2).

Xu, R., Wu, J., Yue, X. and Li, Y. (2022), 'Online structural change-point detection of high-dimensional streaming data via dynamic sparse subspace learning', *Technometrics* pp. 1–14.

Yao, Y.-C. and Au, S.-T. (1989), 'Least-squares estimation of a step function', *Sankhyā: The Indian Journal of Statistics, Series A* pp. 370–381.

Yataka, R., Hirashima, K. and Shiraishi, M. (2023), 'Grassmann manifold flows for stable shape generation', *Advances in Neural Information Processing Systems* **36**, 72377–72411.

Yger, F., Berar, M., Gasso, G. and Rakotomamonjy, A. (2012), 'Adaptive canonical correlation analysis based on matrix manifolds', *arXiv:1206.6453* .

Yin, H., Qu, M., Zhang, H. and Lim, Y. (2018), '3d printing and buildings: A technology review and future outlook', *Technology— Architecture+ Design* **2**(1), 94–111.

Yin, S., Ding, S. X., Xie, X. and Luo, H. (2014), 'A review on basic data-driven approaches for industrial process monitoring', *IEEE Transactions on Industrial electronics* **61**(11), 6418–6428.

Ylvisaker, D. (1977), 'Test resistance', *Journal of the American Statistical Association* **72**(359), 551–556.

You, T., Cheng, H.-M., Ning, Y.-Z., Shia, B.-C. and Zhang, Z.-Y. (2016), 'Community detection in complex networks using density-based clustering algorithm and manifold learning', *Physica A: Statistical Mechanics and its Applications* **464**, 221–230.

Zambon, D., Livi, L. and Alippi, C. (2018), Anomaly and change detection in graph streams through constant-curvature manifold embeddings, *in* '2018 International Joint Conference on Neural Networks (IJCNN)', IEEE, pp. 1–7.

Zang, Y. and Qiu, P. (2018*a*), 'Phase i monitoring of spatial surface data from 3d printing', *Technometrics* **60**(2), 169–180.

Zang, Y. and Qiu, P. (2018*b*), 'Phase ii monitoring of free-form surfaces: An application to 3d printing', *Journal of Quality Technology* **50**(4), 379–390.

Zha, W. and Anand, S. (2015), 'Geometric approaches to input file modification for part quality improvement in additive manufacturing', *Journal of Manufacturing Processes* **20**, 465–477.

Zhan, C., Li, S. and Yang, Y. (2019), 'Improved process monitoring based on global–local manifold analysis and statistical local approach for industrial process', *Journal of Process Control* **75**, 107–119.

Zhan, T. and Kang, J. (2022), 'Finite-sample two-group composite hypothesis testing via machine learning', *Journal of Computational and Graphical Statistics* pp. 1–10.

Zhang, C., Bai, Z., Hu, J. and Wang, C. (2018), 'Multi-sample test for high-dimensional covariance matrices', *Communications in Statistics-Theory and Methods* **47**(13), 3161–3177.

Zhang, C., Chen, N. and Wu, J. (2020), 'Spatial rank-based high-dimensional monitoring through random projection', *Journal of Quality Technology* **52**(2), 111–127.

Zhang, F., Zhang, J. and Ma, J. (2023), 'Data-manifold-based monitoring and anomaly diagnosis for manufacturing process', *Journal of Intelligent Manufacturing* **34**(7), 3159–3177.

Zhang, J. (1996), 'The sample breakdown points of tests', *Journal of statistical planning and inference* **52**(2), 161–181.

Zhang, J. (2002), 'Some extensions of tukey's depth function', *Journal of Multivariate Analysis* **82**(1), 134–165.

Zhang, R. (2021), Hypothesis Test for Manifolds and Networks, PhD thesis, Georgia Institute of Technology.

Zhang, R. (2025), 'Robust sequential change detection: The approach based on breakdown points and influence functions', *IEEE Transactions on Information Theory* **71**(7), 5620–5632.

Zhang, R., Mei, Y. and Shi, J. (2022), 'Robust change detection for large-scale data streams', *Sequential Analysis* **41**(1), 1–19.

Zhang, X., Tian, Y., Guan, G. and Gel, Y. R. (2021), 'Depth-based classification for relational data with multiple attributes', *Journal of Multivariate Analysis* **184**, 104732.

Zhang, Z. and Zha, H. (2004), 'Principal manifolds and nonlinear dimensionality reduction via tangent space alignment', *SIAM journal on scientific computing* **26**(1), 313–338.

Zhao, X. and Del Castillo, E. (2021), 'An intrinsic geometrical approach for statistical process control of surface and manifold data', *Technometrics* **63**(3), 295–312.

Zheng, L., Kim, H. J., Adluru, N., Newton, M. A. and Singh, V. (2017), Riemannian variance filtering: An independent filtering scheme for statistical tests on manifold-valued data, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops', pp. 20–29.

Zhou, Q. and Jacobson, A. (2016), 'Thingi10k: A dataset of 10,000 3d-printing models', *arXiv:1605.04797* .

Zhou, Y. and Sun, S. (2016), 'Manifold partition discriminant analysis', *IEEE transactions on cybernetics* **47**(4), 830–840.

Zhu, J., Pan, W., Zheng, W. and Wang, X. (2021), 'Ball: An r package for detecting distribution difference and association in metric spaces', *Journal of Statistical Software* **97**, 1–31.

Zhu, Z., Anwer, N. and Mathieu, L. (2017), 'Deviation modeling and shape transformation in design for additive manufacturing', *procedia CIRP* **60**, 211–216.

Zou, C., Jiang, W. and Tsung, F. (2011), 'A lasso-based diagnostic framework for multivariate statistical process control', *Technometrics* **53**(3), 297–309.

Zuo, Y. and Serfling, R. (2000), 'General notions of statistical depth function', *Annals of statistics* pp. 461–482.

Zuowei, Z., Keimasi, S., Anwer, N., Mathieu, L. and Lihong, Q. (2017), Review of shape deviation modeling for additive manufacturing, *in* 'Advances on Mechanics, Design Engineering and Manufacturing', Springer, pp. 241–250.

APPENDICES

Appendix A

Proofs of Theoretical Results

A.1 Proofs from Chapter 2

Lemma 1. (Sidak et al., 1999) *Let π_n be a permutation of $(1, 2, ..., N)$. Let $a_n(i)$ and $b_n(i)$ be two arbitrary functions such that*

$$S_n = \sum_{i=1}^{n} a_n(i) b_n(\pi_n(i))$$

If

$$n \frac{\max_{1 \le i \le n} [a_n(i) - \bar{a}_n]^2}{\sum_{i=1}^{n} [a_n(i) - \bar{a}_n]^2} \cdot \frac{\max_{1 \le i \le n} [b_n(i) - \bar{b}_n]^2}{\sum_{i=1}^{n} [b_n(i) - \bar{b}_n]^2} = o(1)$$

Then, by Hajek CLT

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{Var(S_n)}} \to N(0, 1)$$

where

$$\mathbb{E}[S_n] = n \bar{a}_n \bar{b}_n \quad , \quad Var[S_n] = \frac{n^2}{n-1} \sigma_a^2 \sigma_b^2$$

Proof of Theorem 1. It can be shown that

$$S_n = \sum_{i=1}^{n} R_i^x R_i^y = \sum_{i=1}^{n} i R_i^0 = \sum_{i=1}^{n} a_n(i) b_n(i)$$

where $a_n(i) = i$ for $i = 1, 2, ..., n$ and $b_n(i) = R_i^0 \in \{1, ..., n\}$. For any given permutation $\pi_n(i)$ of $(1, ..., n)$, we have

$$\mathbb{E}[a_n(\pi_n(i))] = \bar{a}_n = \frac{1}{n}\sum_{i=1}^{n} a_n(i) = \frac{1}{n}\sum_{i=1}^{n} i = \frac{n+1}{2}$$

$$\begin{aligned}\mathrm{Var}[a_n(\pi_n(i))] = \sigma_a^2 &= \frac{1}{n}\sum_{i=1}^{n}[a_n(i) - \bar{a}_n(i)]^2 = \frac{1}{n}\sum_{i=1}^{n}[a_n(i) - \frac{n+1}{2}]^2 \\ &= \frac{1}{n}[\sum_{i=1}^{n} a_n^2(i) + \sum_{i=1}^{n}(\frac{n+1}{2})^2 - 2(\frac{n+1}{2})\sum_{i=1}^{n} a_n(i)] \\ &= \frac{1}{n}(\sum_{i=1}^{n} i^2 + n(\frac{n+1}{2})^2 - (n+1)\sum_{i=1}^{n} i) = (n+1)(\frac{n-1}{12})\end{aligned}$$

$$\mathbb{E}[b_n(\pi_n(i))] = \bar{b}_n = \frac{1}{n}\sum_{i=1}^{n} b_n(i) = \frac{1}{n}\sum_{i=1}^{n} R_i^0 = \frac{n+1}{2}$$

$$\begin{aligned}\mathrm{Var}[b_n(\pi_n(i))] = \sigma_b^2 &= \frac{1}{n}\sum_{i=1}^{n}[b_n(i) - \bar{b}_n(i)]^2 = \frac{1}{n}\sum_{i=1}^{n}[b_n(i) - \frac{n+1}{2}]^2 \\ &= \frac{1}{n}[\sum_{i=1}^{n} b_n^2(i) + \sum_{i=1}^{n}(\frac{n+1}{2})^2 - 2(\frac{n+1}{2})\sum_{i=1}^{n} b_n(i)] \\ &= \frac{1}{n}(\sum_{i=1}^{n}(R_i^0)^2 + n(\frac{n+1}{2})^2 - (n+1)\sum_{i=1}^{n} R_i^0) = (n+1)(\frac{n-1}{12})\end{aligned}$$

Also

$$\max_{1\le i\le n}[a_n(i) - \bar{a}_n]^2 = (n - \frac{n+1}{2})^2 = \frac{(n-1)^2}{4}$$

$$\max_{1\le i\le n}[b_n(i) - \bar{b}_n]^2 = \frac{(n-1)^2}{4}$$

Therefore

$$n\frac{\max_{1\le i\le n}[a_n(i) - \bar{a}_n]^2}{\sum_{i=1}^{n}[a_n(i) - \bar{a}_n]^2} \cdot \frac{\max_{1\le i\le n}[b_n(i) - \bar{b}_n]^2}{\sum_{i=1}^{n}[b_n(i) - \bar{b}_n]^2} = o(1)$$

which satisfies the required conditions for applying Hajek CLT, and thus

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} \to \mathrm{N}(0,1)$$

where

$$\mathbb{E}[S_n] = n\bar{a}_n\bar{b}_n = \frac{1}{4}n(n+1)^2$$
$$\mathrm{Var}[S_n] = \frac{n^2}{n-1}\sigma_a^2\sigma_b^2 = \frac{n^2(n+1)^2(n-1)}{144}$$

which completes the proof. □

Proof of Theorem 2. It can be shown that

$$S_n^* = \sum_{i=1}^{n} \mathrm{sign}(i - \frac{n+1}{2})\mathrm{sign}(R_i^0 - \frac{n+1}{2}) = \sum_{i=1}^{n} a_n(i)b_n(i)$$

where $a_n(i) = \mathrm{sign}(i - \frac{n+1}{2})$ and $b_n(i) = \mathrm{sign}(R_i^0 - \frac{n+1}{2})$. A simpler way for the above statistic is to use the equivalent test statistic where $\mathrm{sign}(i - \frac{n+1}{2})$ is replaced by the number of observation from the sample exceeding the median increased by $\frac{1}{2}$, i.e.

$$S = \sum_{i=1}^{n} \frac{1}{2}[\mathrm{sign}(i - \frac{n+1}{2}) + 1]\frac{1}{2}[\mathrm{sign}(R_i^0 - \frac{n+1}{2}) + 1]$$

where $a_n(i) = \frac{1}{2}[\mathrm{sign}(i - \frac{n+1}{2}) + 1]$ and $b_n(i) = \frac{1}{2}[\mathrm{sign}(R_i^0 - \frac{n+1}{2}) + 1]$.

For any given permutation $\pi_n(i)$ of $(1, ..., n)$, we have

$$\mathbb{E}[a_n(\pi_n(i))] = \bar{a}_n = \frac{1}{n}\sum_{i=1}^{n} a_n(i) = \frac{1}{2n}\sum_{i=1}^{n}[\mathrm{sign}(i - \frac{n+1}{2}) + 1] = \frac{1}{2}$$

$$\begin{aligned}\mathrm{Var}[a_n(\pi_n(i))] = \sigma_a^2 &= \frac{1}{n}\sum_{i=1}^{n}[a_n(i) - \bar{a}_n(i)]^2 = \frac{1}{n}\sum_{i=1}^{n}[a_n(i) - \frac{n+1}{2}]^2 \\ &= \frac{1}{n}[\sum_{i=1}^{n} a_n^2(i) + \sum_{i=1}^{n}(\frac{n+1}{2})^2 - 2(\frac{n+1}{2})\sum_{i=1}^{n} a_n(i)] \\ &= \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{2})^2 = \frac{1}{4}\end{aligned}$$

$$\mathbb{E}[b_n(\pi_n(i))] = \bar{b}_n = \frac{1}{n}\sum_{i=1}^{n} b_n(i) = \frac{1}{2n}\sum_{i=1}^{n}(\text{sign}(R_i^0 - \frac{n+1}{2}) + 1) = \frac{1}{2}$$

$$\text{Var}[b_n(\pi_n(i))] = \sigma_b^2 = \frac{1}{4}$$

Also

$$\max_{1\le i\le n}[a_n(i) - \bar{a}_n]^2 = \big(2(\frac{1}{2}) - \frac{1}{2}\big)^2 = \frac{1}{4}$$

$$\max_{1\le i\le n}[b_n(i) - \bar{b}_n]^2 = \frac{1}{4}$$

Therefore

$$n\frac{\max_{1\le i\le n}[a_n(i) - \bar{a}_n]^2}{\sum_{i=1}^{n}[a_n(i) - \bar{a}_n]^2} \cdot \frac{\max_{1\le i\le n}[b_n(i) - \bar{b}_n]^2}{\sum_{i=1}^{n}[b_n(i) - \bar{b}_n]^2} = o(1)$$

which satisfies the required conditions for applying Hajek CLT, and thus

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}(S_n)}} \to \text{N}(0,1)$$

where

$$\mathbb{E}[S_n] = n\bar{a}_n\bar{b}_n = \frac{n}{4}$$

$$\text{Var}[S_n] = \begin{cases} \frac{1}{16}\Big(\frac{n^2}{n-1}\Big) & \text{for } n \text{ even} \\ \frac{1}{16}\Big(\frac{(n-1)^2}{n-1}\Big) & \text{for } n \text{ odd} \end{cases} = \begin{cases} \frac{1}{16}\Big(\frac{n^2}{n-1}\Big) & \text{for } n \text{ even} \\ \frac{1}{16}(n-1) & \text{for } n \text{ odd} \end{cases}$$

which completes the proof. □

Proof of Theorem 3. For two independent sample of random objects $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_m$, the Wilcoxon rank test is defined as

$$W = \sum_{i=1}^{m} R(\Lambda(\tilde{P}_{n_0}, Y_i))$$

can be rewritten as follows

$$U_i = \begin{cases} X_i & \text{if } 1 \leq i \leq m \\ Y_{i-m} & \text{if } m+1 \leq i \leq m+n \end{cases}$$

If we define

$$a_{m+n}(i) = R(\Lambda(\tilde{P}_{n_0}, Y_i)) \quad , \quad b_{m+n}(i) = \begin{cases} 1 & \text{if } 1 \leq i \leq m \\ 0 & \text{if } m+1 \leq i \leq m+n \end{cases}$$

$$W = \sum_{i=1}^{m} R(\Lambda(\tilde{P}_{n_0}, Y_i)) = \sum_{i=1}^{m+n} a_{m+n}(i) b_{m+n}(\pi_{m+n}(i))$$

Also, it can be shown that

$$\mathbb{E}[a_{m+n}(\pi_{m+n}(i))] = \bar{a}_{m+n} = \frac{1}{m+n} \sum_{i=1}^{m+n} a_{m+n}(i) = \frac{m+n+1}{2}$$

$$\mathrm{Var}[a_{m+n}(\pi_{m+n}(i))] = \sigma_a^2 = \frac{1}{n+m} \sum_{i=1}^{m+n} [a_{m+n}(i) - \bar{a}_{m+n}(i)]^2 = \frac{(m+n)^2 - 1}{12}$$

$$\mathbb{E}[b_{m+n}(\pi_{m+n}(i))] = \bar{b}_{m+n} = \frac{1}{m+n} \sum_{i=1}^{m+n} b_{m+n}(i) = \frac{m}{m+n}$$

$$\mathrm{Var}[b_{m+n}(\pi_{m+n}(i))] = \sigma_b^2 = \frac{1}{n+m} \sum_{i=1}^{m+n} [b_{m+n}(i) - \bar{b}_{m+n}(i)]^2 = \frac{m}{m+n}\big(1 - \frac{m}{m+n}\big)$$

Also

$$\max_{1 \leq i \leq n} [a_n(i) - \bar{a}_n]^2 = O((m+n)^2)$$

$$\max_{1 \leq i \leq n} [b_n(i) - \bar{b}_n]^2 = \max[\big(\frac{m}{m+n}\big)^2, \big(1 - \frac{m}{m+n}\big)^2]$$

Therefore

$$n\frac{\max_{1\le i\le n}[a_n(i)-\bar{a}_n]^2}{\sum_{i=1}^n[a_n(i)-\bar{a}_n]^2}\cdot\frac{\max_{1\le i\le n}[b_n(i)-\bar{b}_n]^2}{\sum_{i=1}^n[b_n(i)-\bar{b}_n]^2}=o(1)$$

which satisfies the required conditions for applying Hajek CLT, and thus

$$\frac{S_n-\mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}\to \mathrm{N}(0,1)$$

where

$$\mathbb{E}[S_n]=(m+n)\bar{a}_{m+n}\bar{b}_{m+n}=\frac{m(m+n+1)}{2}$$

$$\mathrm{Var}[S_n]=\big(\frac{mn}{n+m-1}\big)\left(\frac{(n+m)^2-1}{12}\right)=\frac{mn(n+m+1)}{12}$$

which completes the proof. □

Proof of Theorem 4. Let us define

$$S_{n_j}=\sum_{i\in s_j}a_n(R_{n_j})\quad ,1\le j\le t$$

$$S_{n_1,\dots,n_t}=(n-1)\Big\{\sum_{i=1}^n[a_n(i)-\bar{a}_n]^2\Big\}^{-1}\sum_{j=1}^t\frac{(S_{n_j}-n_j\bar{a}_n)^2}{n_j}$$

The test statistic can be rewritten as

$$H(k)=\frac{12}{n(n+1)}\sum_{j=1}^t n_j(\bar{R}_{.j}(k)-\frac{n+1}{2})^2=\frac{12}{n(n+1)}\sum_{j=1}^t\frac{(\sum_{i=1}^{n_j}R_{ij}(k)-n_j\frac{n_j(n+1)}{2})^2}{n_j}$$

where

$$(\sigma^2)^{-1}=(n-1)\Big\{\sum_{i=1}^n[a_n(i)-\bar{a}_n]^2\Big\}^{-1}=\frac{12}{n(n+1)}$$

$$\sum_{j=1}^t\frac{(S_{n_j}-n_j\bar{a}_n)^2}{n_j}=\sum_{j=1}^t\frac{(\sum_{i=1}^{n_j}R_{ij}(k)-n_j\frac{n_j(n+1)}{2})^2}{n_j}$$

$$S_{n_j}=\sum_{i=1}^{n_j}R_{ij}(k)\ ,\quad \bar{a}_n=\frac{n_j(n+1)}{2}$$

Let $\kappa_j = \frac{n_j}{n}$. By following steps similar to Theorem 1, it can be shown that

$$\zeta_j = \sigma^{-1}(S_{n_j} - n_j \bar{a}_{n_j}) n_j^{-1/2} \xrightarrow{D} Z_j \sim \mathrm{N}(0, 1 - \frac{n_j}{n})$$

where $\mathrm{Cov}(Z_j, Z_g) = -\kappa_j \kappa_g$ for $1 \leq j \neq g \leq t$. Let

$$c_{ij} = \begin{cases} 1 & i \in s_j \\ 0 & o.w. \end{cases}$$

Then, we have

$$\sum_{i=1}^{n} (c_{ij} - \bar{c}_j)(c_{ig} - \bar{c}_g) = -\frac{n_j n_g}{n} \ , \quad 1 \leq j \neq g \leq t$$

For any real vector $(\lambda_1, ..., \lambda_t)$, let

$$\sum_{j=1}^{t} \lambda_j \zeta_j = \sigma^{-1} \sum_{i=1}^{n} c_i [a_n(R_i) - \bar{a}_n]$$

where $c_i \in \lambda_j n_j^{-1/2}$ if $i \in s_j$. It can be shown that

$$\sum_{i=1}^{n} (c_i - \bar{c})^2 \to \sum_{j=1}^{t} \lambda_j^2 - (\sum_{j=1}^{t} \lambda_j \kappa_j)^2 > 0$$

$$\max_{1 \leq i \leq n} (c_i - \bar{c})^2 = \max_{1 \leq i \leq n} (\lambda_j n_j^{-1/2} - n^{-1} \sum_{j=1}^{t} \lambda_j n_j^{1/2})^2 \to 0$$

$$\frac{\max_{1 \leq i \leq n} (c_i - \bar{c})^2}{\sum_{i=1}^{n} (c_i - \bar{c})^2} = o(1)$$

Therefore, $\sum_{j=1}^{t} \lambda_j \zeta_j$ converges in distribution to $\sum_{j=1}^{t} \lambda_j Z_j$ and by Cramer-Wold theorem, we conclude that

$$(\zeta_1, ..., \zeta_t) \xrightarrow{D} (Z_1, ..., Z_t)$$

This implies that

$$H(k) = \sum_{j=1}^{t} \zeta_j^2 \xrightarrow{D} \sum_{j=1}^{t} Z_j^2$$

which is χ^2_{t-1}.

Given that the statistic H is the average of $H(k)$'s calculated by the construction of empirical measure ν_k, it also holds that H is asymptotically χ^2_{t-1} distributed. This completes the proof. □

Proof of Theorem 5. Let $\mathcal{X} = (x_1, \ldots, x_n) \in \mathcal{M}$, we can define the empirical Rademacher complexity (Sen, 2018) as follows

$$\mathrm{Rad}_n(\mathcal{F}, \mathcal{X}) = \frac{1}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}\zeta_i f(x_i)\right|\right] ,$$

where $\zeta_1, \ldots, \zeta_n$ are i.i.d random variables generated from Rademacher distribution with $P(\zeta_i = 1) = P(\zeta_i = -1) = \frac{1}{2}$.

For a function class $\mathcal{F}$ on $\mathcal{M}$ and $\varepsilon > 0$, $S \subseteq \mathcal{F}$ is an ε-covering of $\mathcal{F}$ if for any $f \in \mathcal{F}$, there exists $f_s \in S$, such that $\|f - f_s\|_{L^2(\mathcal{X})} \leq \varepsilon$. The covering number is defined as

$$\mathcal{N}(\mathcal{F}, \varepsilon, L^2(\mathcal{X})) = \min\{|\ S\ |: S \text{ is a } \varepsilon\text{-cover of } \mathcal{F}\} .$$

The Dudley entropy integral (Dudley, 2010; Talagrand, 1994) can be used to bound the Rademacher complexity

$$\mathrm{Rad}_n(\mathcal{F}, \mathcal{X}) \leq \inf_{\varepsilon>0}\left\{4\varepsilon + \frac{12}{\sqrt{n}}\int_{\varepsilon}^{c}\sqrt{\log\mathcal{N}(\mathcal{F}, t, L^2(\mathcal{X}))}dt\right\} ,$$

where $c = \sup_{f\in\mathcal{F}}\sqrt{\sum_{i=1}^{n} f^2(x_i)/n}$. By Assumption 2, the class of functions $\mathcal{F}$ is VC with VC-dimension d_0. The ε-covering of $\mathcal{F}$ can be bounded as follows

$$\mathcal{N}(\mathcal{F}, \varepsilon, L^2(\mathcal{X})) \leq c_0 K^{d_0 \log d_0}\left(\frac{1}{\varepsilon}\right)^{2d_0+2} .$$

The VC-dimension of a class of bounded functions defined on a manifold space depends on the intrinsic dimension of the manifold d as well as the complexity of the function class $\mathcal{F}$, which we denote by $d_0 = h_{\mathcal{F}}(d)$ (Ferri and Frosini, 2008). For example, the VC-dimension of the class of indicator functions on d-dimensional sphere is $d+1$, and for the d-dimension Torus space, the VC-dimension grows with the rate $d\log d$. The VC-dimension of linear functions on d-dimensional manifold space is $d+1$. For simplicity we assume that for the considered class of functions $\mathcal{F}$, this relation is linear, i.e. $d_0 = c'd$, for $c' > 0$. Therefore, it holds that $\log\mathcal{N}(\mathcal{F}, t, L^2(\mathcal{X})) = O(d\log d)$, and thus $\mathrm{Rad}_n(\mathcal{F}, \mathcal{X}) = O_p\left(n^{-1/2}(d\log d)^{1/2}\right)$.

Let us denote the empirical measure by

$$P_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \ .$$

The next step is to find an upper bound for $\mathbb{E}[D_{\mathcal{F}}(P, P_n)]$ and $\mathrm{Var}[D_{\mathcal{F}}(P, P_n)]$. It holds that

$$\begin{aligned}
\mathbb{E}[D_{\mathcal{F}}(P, P_n)] &= \frac{1}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X_i)]\right|\right] \\
&\leq \frac{2}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n} \zeta_i f(X_i)\right|\right] \\
&\leq 2\inf_{\varepsilon>0}\left\{4\varepsilon + \frac{12}{\sqrt{n}}\int_{\varepsilon}^{c}\sqrt{\log \mathcal{N}(\mathcal{F}, t, L^2(\mathcal{X}))}dt\right\} \ .
\end{aligned}$$

By Efron-Stein-Steele inequality (Boucheron et al., 2003), we have that

$$\begin{aligned}
\mathrm{Var}[D_{\mathcal{F}}(P, P_n)] &= \mathrm{Var}\left[\frac{1}{n}\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X_i)]\right|\right] \\
&\leq \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left[\left(\frac{1}{n}\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X_i)]\right|\right.\right. \\
&\qquad \left.\left. - \frac{1}{n}\sup_{f\in\mathcal{F}}\left|\sum_{j=1,\ j\neq i}^{n}[f(X_j) - \mathbb{E}[f(X_j)]] + f(Y_i) - \mathbb{E}[f(X_i)]\right|\right)^2\right] \ ,
\end{aligned}$$

where $Y_1, \ldots, Y_n \sim P$ are independent of $X_1, \ldots, X_n$. Therefore,

$$\begin{aligned}
\mathbb{E}[D_{\mathcal{F}}(P, P_n)] &= O\left(n^{-1/2}(d\log d)^{1/2}\right) \ , \\
\mathrm{Var}[D_{\mathcal{F}}(P, P_n)] &= O\left(n^{-1}(d\log d)^{1}\right) \ .
\end{aligned}$$

By the application of Chebyshev's inequality, we get

$$D_{\mathcal{F}}(P, P_n) = O_p\left(n^{-1/2}(d\log d)^{1/2}\right) \ .$$

By (Han et al., 2023), we know that $D_{\mathcal{F}} = 0$ is equivalent to weak convergence, $P_n \xrightarrow{d} P$, with rate $O\left(n^{-1/2}(d\log d)^{1/2}\right)$. If $P_n \xrightarrow{d} P$, for a bounded and Lipschitz continuous function

g defined on $\mathcal{M}$, we have $g_{P_n} \xrightarrow{p} g_P$ with the same rate. Therefore, Assumption 3 and the boundedness of depth functions imply that

$$\sup_{x \in \mathcal{M}} \mid D(P_n, x) - D(P, x) \mid = O_p\left(n^{-1/2}(d \log d)^{1/2}\right) \ .$$

Let

$$A_{i,\,j} = \{D(P,\, X_j) \leq D(P,\, X_i)\} \cap \{D(P_n,\, X_j) > D(P_n,\, X_i)\}\,,$$
$$B_{i,\,j} = \{D(P,\, X_j) > D(P,\, X_i)\} \cap \{D(P_n,\, X_j) \leq D(P_n,\, X_i)\}\,.$$

Let $\gamma_n = \max\limits_{1 \leq j \leq n} \mid D(P,\, \mathbf{X}_j) - D(P_n,\, \mathbf{X}_j) \mid$, then it follows that

$$A_{i,\,j} \subset \{\mid D(P,\, X_i) - D(P,\, X_j) \mid \leq 2\gamma_n\}\,,$$

$$B_{i,\,j} \subset \{\mid D(P,\, X_i) - D(P,\, X_j) \mid \leq 2\gamma_n\}\,.$$

Furthermore, we have

$$\gamma_n \leq \sup_{\mathbf{x} \in \mathcal{M}} \mid D(P_n,\, \mathbf{x}) - D(P,\, \mathbf{x}) \mid .$$

Therefore

$$\mathbb{I}(A_{i,j}) \leq \mathbb{I}(\{\mid D(P,\, X_i) - D(P,\, X_j) \mid \leq 2 \sup_{\mathbf{x} \in \mathcal{M}} \mid D(P_n,\, x) - D(P,\, x) \mid\}) \ ,$$

$$\mathbb{I}(B_{i,j}) \leq \mathbb{I}(\{\mid D(P,\, X_i) - D(P,\, X_j) \mid \leq 2 \sup_{\mathbf{x} \in \mathcal{M}} \mid D(P_n,\, x) - D(P,\, x) \mid\}) \ .$$

If we use relative rank score function $a(i) = R_i/n$ for rank-depth statistics, it can be shown that

$$\begin{aligned} \mid \widehat{a}(i) - a(i) \mid &\leq \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}(A_{i,j}) + \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}(B_{i,j}) \\ &\leq 4 \sup_{\mathbf{x} \in \mathcal{M}} \mid D(P_n,\, x) - D(P,\, x) \mid\} \ . \end{aligned}$$

Hence, it holds that

$$\widehat{a}(i) - a(i) = O_p\left(n^{-1/2}(d \log d)^{1/2}\right) \ ,$$

We can use the above result to obtain convergence bounds for Spearman correlation coefficient test, Quadrant test, Wilcoxon test and Kruskal-Wallis test statistics.

Spearman Correlation Coefficient Test: Let us define the rank score $a'_i = (i/n)(R^0_i/n)$. Then, we have

$$T(P_n) = \frac{1}{\sigma_n} \sum_{i=1}^{N} \widehat{a}'(i) - \bar{a}'_n \ .$$

where

$$\bar{a}'_n = \frac{1}{n} \sum_{i=1}^{n} a'_i \ ,$$

$$\sigma_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (a'_i - \bar{a}'_n)^2 \ .$$

By taking similar steps to (Koul, 1970), it can be shown that

$$\mid T(P_n) - T(P) \mid = \frac{1}{\sigma_n} \sum_{i=1}^{N} \frac{i}{n} \left(\frac{\widehat{R}^0_i - R^0_i}{n} \right) \leq \max_{i=1,\ldots,n} \mid \widehat{a}(i) - a(i) \mid = O_p\left(n^{-1/2} (d \log d)^{1/2}\right) \ .$$

Similar convergence rate holds for Quadrant test.

Wilcoxon Test: Let us define the rank score $a_i = \left(\frac{1}{n}\right) R_i \mathbb{I}[i > n]$. Then, by following similar steps as above, we have

$$\mid T(P_n) - T(P) \mid = O_p\left(n^{-1/2} (d \log d)^{1/2}\right) \ .$$

Kruskal-Wallis Test: Let us define the rank score $\left(\frac{1}{n}\right) R_i$. Then, we have

$$T(P_n) = \frac{12}{n(n+1)} \sum_{j=1}^{K} \frac{1}{n_j} \left(\sum_{i \in S_j} a_i - \bar{a}_n \right)^2 \ .$$

Following similar argument, it holds that

$$\mid T(P_n) - T(P) \mid = O_p\left(n^{-1} (d \log d)\right) \ .$$

As a result

$$T(P_n) - T(P) = O_p\left(n^{-1/2} (d \log d)^{1/2}\right) \ ,$$

for Spearman correlation coefficient test, Quadrant test, and Wilcoxon test, and

$$T(P_n) - T(P) = O_p\left(n^{-1} (d \log d)\right) \ ,$$

for Kruskal-Wallis test. This completes the proof.

□

Proof of Theorem 6. The proof can be done following the steps given in (Katsevich and Ramm, 1993). The results hold for all of the proposed rank-depth test statistics.

(a) Size of tests: Fix $0 < \alpha < 1$, the probability of rejecting H_0 under null hypothesis, i.e. type-I error, and find the threshold value c_0 by solving the equation

$$P(T(P) \geq c_0 \mid H_0) = \alpha \ .$$

Note that we can use asymptotic results for the proposed test statistics to find the exact value of threshold c_0. For small values of n, we can use Monte Carlo simulation using a sufficiently large number of permutations of $\{1, \ldots, n\}$. By Theorem 5, we have

$$T(P_n) - T(P) = O_p\left(n^{-\alpha}(d \log d)^{\alpha}\right) = o_p(1) \ ,$$

where $\alpha = 1/2$ for Spearman correlation coefficient test, Quadrant test, and Wilcoxon test, and $\alpha = 1$ for Kruskal-Wallis test. Therefore

$$P(T(P_n) \geq c_0 \mid H_0) = \alpha \ ,$$

as $n \to \infty$. This implies that the size of the test statistics are controlled at level α.

(b) Power of tests: To study the power of the proposed test statistics, first note that by the compactness of $\mathcal{M}$ and the theory of U-statistics (Van der Vaart, 2000), it can be shown that $c_0 = O(n^{-1/2})$, as $n \to \infty$. Moreover, under H_A, we have

$$\int_{-\infty}^{\infty} \Big(\epsilon H_1(u) + (1-\epsilon)H_2(u)\Big) dH_1(u) \neq \frac{1}{2}$$

where $H_1(u) = P(D(P_n, X) \leq u)$ and $H_2(u) = P(D(Q_n, Y) \leq u)$. Also, it can be shown that

$$T(P_n) = \sqrt{n}\left(\int_{-\infty}^{\infty} \epsilon H_1(u) + (1-\epsilon)H_2(u) dH_1(u) - \frac{1}{2} \right) = O(\sqrt{n}) \ ,$$

This rate holds for Spearman correlation coefficient test, Quadrant test, and Wilcoxon test. For Kruskal-Wallis we can follow a similar argument to show $T(P_n) = O(n)$. Therefore, it holds that

$$\begin{aligned} P(T(P_n) \leq c_0 \mid H_A) &\leq P(\mid T(P_n) - \mu_n \mid \geq \mu_n - c_0 \mid H_A) \\ &\leq \frac{\mathrm{Var}(T(P_n))}{(c_0 - \mu_n)^2} = O(n^{-2}) \ , \end{aligned}$$

as $n \to \infty$, implying that

$$\lim_{n \to \infty} P(T(P_n) \geq c_0 \mid H_A) \to 1 \ .$$

□

Lemma 2. (Aaron and Bodart, 2018) *Let $\mathcal{M}$ be a d-dimensional compact manifold embedded in $\mathbb{R}^p$ with $p \geq 2$. Let the observations $X_1, \ldots, X_n$ follow probability distribution P with a continuous density with volume measure on $\mathcal{M}$ bounded from below by a positive constant c_0. For any $c > 0$ and maximum radius $r_n = c\left(\max_{x\in\mathcal{Z}}\left(\min_{y\neq x; y\in\mathcal{Z}} \|x-y\|\right)\right)^{2/3}$, the following convergence rate holds for graph-based estimator of geodesic distance*

$$\sup_{x,y\in\mathcal{Z}} |\widehat{\delta}(x,y) - \delta(x,y)| = O\left(\left(\frac{\log n}{n}\right)^{2/(3d)}\right) ,$$

almost surely, where d is the intrinsic dimension of the manifold $\mathcal{M}$.

Proof of Theorem 7. Under the Assumptions 4-6, all of the conditions of Lemma 2 are satisfied. Let $\widehat{\Delta}(x,y) = \widehat{\delta}(x,y) - \delta(x,y)$. We have

$$\sup_{x,y\in\mathcal{Z}} |\widehat{\Delta}(x,y)| = O\left(\left(\frac{\log n}{n}\right)^{2/(3d)}\right) ,$$

almost surely, where d is the intrinsic dimension of the manifold. We prove the asymptotic convergence rate for graph-based estimate of geodesic depth, half-space depth, and spatial depth.

Geodesic Depth: For the geodesic depth, we have

$$\sup_{x\in\mathcal{M}} |\widehat{GGD}(P_n,x) - GD(P,x)| = \sup_{x\in\mathcal{M}} \left|\exp\Big(-\frac{1}{n}\sum_{i=1}^{n}\widehat{\delta}(x,y_i)\Big) - \exp\Big(-\frac{1}{n}\sum_{i=1}^{n}\delta(x,y_i)\Big)\right| .$$

To obtain the convergence rate, first note that

$$\begin{aligned}\sup_{x\in\mathcal{M}}\left|\frac{1}{n}\sum_{i=1}^{n}\widehat{\delta}(x,y_i) - \frac{1}{n}\sum_{i=1}^{n}\delta(x,y_i)\right| &= \sup_{x\in\mathcal{M}}\left|\frac{1}{n}\sum_{i=1}^{n}\widehat{\Delta}(x,y_i)\right| \\ &\leq c\max_{x_0,y_0\in\mathcal{Z}}\left|\widehat{\Delta}(x_0,y_0)\right| = O\left(\left(\frac{\log n}{n}\right)^{2/(3d)}\right) .\end{aligned}$$

For two random variables x_n and x such that $|x_n - x| = O(n^{\alpha})$ with $0 < \alpha < 1$, it can be shown that

$$\left|e^{-x_n} - e^{-x}\right| = \left|e^{-x}\left(e^{-(x_n-x)} - 1\right)\right| =\approx e^{-x}|x_n - x| = O(n^{\alpha}) ,$$

where we used Taylor expansion to approximate $e^{-u} - 1 \approx -u$. Hence

$$\sup_{x \in \mathcal{M}} |\widehat{GGD}(P_n, x) - GD(P, x)| = O\left(\left(\frac{\log n}{n}\right)^{2/(3d)}\right) ,$$

almost surely.

Half-Space Depth: For the half-space depth, we have

$$\sup_{x \in \mathcal{M}} |\widehat{GHD}(P_n, x) - HD(P, x)| = \sup_{x \in \mathcal{M}} \left| \inf_{\substack{p_1, p_2 \in \mathcal{A} \\ \widehat{\delta}(p_1, x) \leq \widehat{\delta}(p_2, x)}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left[\widehat{\delta}(p_1, y_i) \leq \widehat{\delta}(p_2, y_i)\right] \right.$$
$$\left. - \inf_{\substack{p_1, p_2 \in \mathcal{A} \\ \delta(p_1, x) \leq \delta(p_2, x)}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left[\delta(p_1, y_i) \leq \delta(p_2, y_i)\right] \right| .$$

The convergence rate can be achieved given the fact that if x_n converges to x with speed $O(n^\alpha)$, then $\mathbb{I}[x_n \in A]$ converges to $\mathbb{I}[x \in A]$ with the same speed $O(n^\alpha)$. Therefore, we have

$$\left| \inf_{\substack{p_1, p_2 \in \mathcal{A} \\ \widehat{\delta}(p_1, x) \leq \widehat{\delta}(p_2, x)}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left[\widehat{\delta}(p_1, y_i) \leq \widehat{\delta}(p_2, y_i)\right] - \inf_{\substack{p_1, p_2 \in \mathcal{A} \\ \delta(p_1, x) \leq \delta(p_2, x)}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left[\delta(p_1, y_i) \leq \delta(p_2, y_i)\right] \right|$$
$$\leq \sup_{p_1, p_2, y \in \mathcal{M}} \left| \mathbb{I}\left[\widehat{\delta}(p_1, y) \leq \widehat{\delta}(p_2, y)\right] - \mathbb{I}\left[\delta(p_1, y) \leq \delta(p_2, y)\right] \right| = O\left(\left(\frac{\log n}{n}\right)^{2/(3d)}\right) ,$$

almost surely, and therefore

$$\sup_{x \in \mathcal{M}} |\widehat{GHD}(P_n, x) - HD(P, x)| = O\left(\left(\frac{\log n}{n}\right)^{2/(3d)}\right) ,$$

almost surely.

Spatial Depth: For the spatial depth, we have

$$\widehat{GSD}(P_n, x) = 1 - \frac{1}{2\binom{n}{2}} \sum_{i=1}^{n} \sum_{j=i}^{n} \left(\frac{\widehat{\delta}(x, y_i)}{\widehat{\delta}(x, y_j)} + \frac{\widehat{\delta}(x, y_j)}{\widehat{\delta}(x, y_i)} - \frac{\widehat{\delta}(y_i, y_j)}{\widehat{\delta}(x, y_i)\widehat{\delta}(x, y_j)} \right) ,$$

which gives

$$\begin{aligned}\sup_{x\in\mathcal{M}}|\widehat{GSD}(P_n,x)-SD(P,x)| \leq & \frac{1}{2\binom{n}{2}}\sum_{i=1}^{n}\sum_{j=i}^{n}\left|\frac{\widehat{\delta}(x,y_i)}{\widehat{\delta}(x,y_j)}-\frac{\delta(x,y_i)}{\delta(x,y_j)}\right| \\ & + \frac{1}{2\binom{n}{2}}\sum_{i=1}^{n}\sum_{j=i}^{n}\left|\frac{\widehat{\delta}(x,y_j)}{\widehat{\delta}(x,y_i)}-\frac{\delta(x,y_j)}{\delta(x,y_i)}\right| \\ & + \frac{1}{2\binom{n}{2}}\sum_{i=1}^{n}\sum_{j=i}^{n}\left|\frac{\delta(y_i,y_j)}{\widehat{\delta}(x,y_i)\widehat{\delta}(x,y_j)}-\frac{\delta(y_i,y_j)}{\delta(x,y_i)\delta(x,y_j)}\right| .\end{aligned}$$

The convergence can be obtained using the following result

$$\begin{aligned}\left|\frac{\widehat{\delta}(x,y_i)}{\widehat{\delta}(x,y_j)}-\frac{\delta(x,y_i)}{\delta(x,y_j)}\right| &= \left|\frac{\delta(x,y_j)\widehat{\delta}(x,y_i)-\delta(x,y_i)\widehat{\delta}(x,y_j)}{\widehat{\delta}(x,y_j)\delta(x,y_j)}\right| \\ &= \left|\frac{\delta(x,y_i)\left(\widehat{\delta}(x,y_j)-\widehat{\delta}(x,y_j)\right)}{\widehat{\delta}(x,y_j)\delta(x,y_j)}-\frac{\delta(x,y_j)\left(\widehat{\delta}(x,y_i)-\widehat{\delta}(x,y_i)\right)}{\widehat{\delta}(x,y_j)\delta(x,y_j)}\right| \\ &\leq \left|\frac{\delta(x,y_i)\left(\widehat{\delta}(x,y_j)-\widehat{\delta}(x,y_j)\right)}{\widehat{\delta}(x,y_j)\delta(x,y_j)}\right|+\left|\frac{\delta(x,y_j)\left(\widehat{\delta}(x,y_i)-\widehat{\delta}(x,y_i)\right)}{\widehat{\delta}(x,y_j)\delta(x,y_j)}\right|\end{aligned}$$

Given that $\widehat{\delta}(x,y_j)$ converges to $\delta(x,y_j)$, we can continue as follows

$$\begin{aligned}&\frac{\delta(x,y_i)}{\delta^2(x,y_j)}\left|\widehat{\delta}(x,y_j)-\delta(x,y_j)\right|+\frac{1}{\delta(x,y_j)}\left|\widehat{\delta}(x,y_i)-\delta(x,y_i)\right| \\ &\qquad\leq c\max_{x,y\in\mathcal{Z}}\mid\widehat{\Delta}(x,y)\mid=O\left(\left(\frac{\log n}{n}\right)^{2/(3d)}\right) ,\end{aligned}$$

almost surely. Similar convergence rate can be achieved for the other two terms. Therefore

$$\sup_{x\in\mathcal{M}}|\widehat{GSD}(P_n,x)-SD(P,x)|=O\left(\left(\frac{\log n}{n}\right)^{2/(3d)}\right) ,$$

almost surely. This completes the proof. □

Proof of Theorem 8. To do the proof, note that all of the test statistics can be written as a function of relative rank, which has the following forms for theoretical depth function and graph-based depth function

$$R_i = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}\left[D(P, x_j) \leq D(P, x_i)\right] \ ,$$

$$\widehat{R}_i = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}\left[\widehat{GD}(P_n, x_j) \leq \widehat{GD}(P_n, x_i)\right] \ .$$

Therefore, it can be shown that

$$\begin{aligned}
\left|\widehat{R}_i - R_i\right| &= \left|\frac{1}{n}\sum_{j=1}^{n}\left(\mathbb{I}\left[D(P, x_j) \leq D(P, x_i)\right] - \mathbb{I}\left[\widehat{GD}(P_n, x_j) \leq \widehat{GD}(P_n, x_i)\right]\right)\right| \\
&\leq \max_{x,y\in\mathcal{Z}}\left|\mathbb{I}\left[D(P, x) \leq D(P, y)\right] - \mathbb{I}\left[\widehat{GD}(P_n, x) \leq \widehat{GD}(P_n, y)\right]\right| \\
&= O\left(\left(\frac{\log n}{n}\right)^{2/(3d)}\right) \ .
\end{aligned}$$

The rest of the proof gives convergence bounds for graph-based Spearman correlation coefficient test, Wilcoxon test and Kruskal-Wallis test statistics.

Spearman Correlation Coefficient Test: For $\widehat{S}_n = \widehat{T}(P_n)$, we can show the following

$$\begin{aligned}
\sup_{\mathcal{Z}\in\mathcal{M}}\left|\widehat{T}(P_n) - T(P)\right| &= \sup_{\mathcal{Z}\in\mathcal{M}}\left|\widehat{S}_n - S_n\right| \\
&= \left(\frac{24}{\sqrt{n^2(n+1)^2(n-1)}}\right)\sup_{\mathcal{Z}\in\mathcal{M}}\left|\sum_{i=1}^{n}\widehat{R}_i^x\widehat{R}_i^y - \sum_{i=1}^{n}R_i^xR_i^y\right| \\
&= \left(\frac{24}{\sqrt{n^2(n+1)^2(n-1)}}\right)\left|\sum_{i=1}^{n} i\left(\widehat{R}_i^0 - R_i^0\right)\right| \\
&= O\left(n^{-1/2}\left(\frac{\log n}{n}\right)^{2/(3d)}\right) \ ,
\end{aligned}$$

almost surely.

Wilcoxon Test: For the Wilcoxon test $\widehat{W}_n = \widehat{T}(P_n)$, we have

$$\begin{aligned}\sup_{\mathcal{Z}\in\mathcal{M}}\left|\widehat{T}(P_n) - T(P)\right| &= \sup_{\mathcal{Z}\in\mathcal{M}}\left|\widehat{W}_n - W_n\right| \\ &= \left(\frac{1}{\sqrt{n_1 n_2(n_1+n_2+1)/12}}\right)\sup_{\mathcal{Z}\in\mathcal{M}}\left|\sum_{i=1}^{n_2}\widehat{R}_i^y - R_i\right| \\ &= O\left(n^{-1/2}\left(\frac{\log n}{n}\right)^{2/(3d)}\right) ,\end{aligned}$$

almost surely.

Kruskal-Wallis Test: For Kruskal-Wallis test $\widehat{H}_n = \widehat{T}(P_n)$, it holds that

$$\begin{aligned}\sup_{\mathcal{Z}\in\mathcal{M}}\left|\widehat{T}(P_n) - T(P)\right| &= \sup_{\mathcal{Z}\in\mathcal{M}}\left|\widehat{H}_n - H_n\right| \\ &= \left(\frac{12}{n(n+1)}\right)\sup_{\mathcal{Z}\in\mathcal{M}}\left|\sum_{j=1}^{K}\frac{1}{n_j}\left(\sum_{i\in S_j}\sum_{l\in S_j}\left(\widehat{R}_i\widehat{R}_l - R_i R_l\right)\right)\right| .\end{aligned}$$

It can be shown that

$$\begin{aligned}\left|\widehat{R}_i\widehat{R}_l - R_i R_l\right| &= \left|\widehat{R}_i\widehat{R}_l - \widehat{R}_i R_l + \widehat{R}_i R_l - R_i R_l\right| \\ &= \left|\widehat{R}_i\left(\widehat{R}_l - R_l\right) + R_l\left(\widehat{R}_i - R_i\right)\right| = \left|\widehat{R}_i + R_l\right| O\left(\left(\frac{\log n}{n}\right)^{2/(3d)}\right) ,\end{aligned}$$

almost surely. Therefore

$$\sup_{\mathcal{Z}\in\mathcal{M}}\left|\widehat{T}(P_n) - T(P)\right| = O\left(n^{-1}\left(\frac{\log n}{n}\right)^{2/(3d)}\right) ,$$

almost surely. This completes the proof. □

Lemma 3. *Condition on $\mathcal{Z}$, under H_0, as $n, m \to \infty$, we have*

$$T_{n+m}\left(Z_{\pi_1}, \ldots, Z_{\pi_{n+m}}\right) = O_p\left((n+m)^{-1/2}\right) ,$$

for Spearman correlation coefficient and Wilcoxon test, and

$$T_{n+m}\left(Z_{\pi_1}, \ldots, Z_{\pi_{n+m}}\right) = O_p\left((n+m)^{-1}\right) ,$$

for Kruskal-Wallis test.

Proof of Lemma 3. Let $N = n + m$ for Wilcoxon test, $N = n$ for Spearman correlation coefficient test, and $N = n_1+n_2+\ldots+n_K$ for Kruskal-Wallis test. Consider the permutation $(\pi_1, \ldots, \pi_N)$ of indices $(1, \ldots, N)$ and define $T_{\boldsymbol{\pi}} = T(Z_{\pi_1}, \ldots, Z_{\pi_{n+m}})$.

Spearman Correlation Coefficient Test: Let us define the rank score $a_i = (i/N)(R_i^0/N)$. Then, under H_0, we have

$$T_{\boldsymbol{\pi}} = T\left(Z_{\pi_1}, \ldots, Z_{\pi_{n+m}}\right) = \frac{1}{\sigma_N} \sum_{i=1}^{N} a_{\pi_i} \ .$$

Therefore

$$\mathbb{E}[T_{\boldsymbol{\pi}} \mid \mathcal{Z}] = N\sigma_N^{-1}\bar{a}_N \ ,$$

where $\bar{a}_N = N^{-1}\sum_{i=1}^{N} a_i$, and

$$\sigma_N^2 = \frac{1}{N-1}\left(\sum_{i=1}^{N}(a_i - \bar{a}_N)^2\right)$$

$$\mathrm{Var}[T_{\boldsymbol{\pi}} \mid \mathcal{Z}] = \sigma_N^{-2}\sum_{i=1}^{N}\mathrm{Var}[a_{\pi_i}] = N\sigma_N^{-2}\frac{1}{N-1}\left(\sum_{i=1}^{N}(a_i - \bar{a}_N)^2\right) \ .$$

By compactness of $\mathcal{M}$, and U-statistic CLT (Van der Vaart, 2000; Chu and Dai, 2024), it can be shown that

$$\mathbb{E}[T_{\boldsymbol{\pi}} \mid \mathcal{Z}] = O_p(N^{-1/2}) \ ,$$

and

$$\mathrm{Var}[T_{\boldsymbol{\pi}} \mid \mathcal{Z}] = O_p(N^{-1}) \ .$$

Therefore, by Chebychev's inequality, condition on $\mathcal{Z}$, we have

$$T_{\boldsymbol{\pi}} = O_p(N^{-1/2}) \ .$$

Wilcoxon Test: Let us define the rank score $a_i = \left(\frac{1}{n}\right) R_i \mathbb{I}[i > n]$. Then, under H_0, we have

$$T_{\boldsymbol{\pi}} = T\left(Z_{\pi_1}, \ldots, Z_{\pi_{n+m}}\right) = \frac{1}{\sqrt{nm(N+1)/12}} \sum_{i=1}^{N} a_{\pi_i} \ .$$

Similar to Spearman test, we get

$$T_{\boldsymbol{\pi}} = O_p(N^{-1/2}) \ ,$$

condition on $\mathcal{Z}$.

Kruskal-Wallis Test: Let us define the rank score $\left(\frac{1}{n}\right) R_i$. Then, under H_0, we have

$$T_{\boldsymbol{\pi}} = T\left(Z_{\pi_1}, \ldots, Z_{\pi_{n+m}}\right) = \frac{12}{N(N+1)} \sum_{j=1}^{K} \frac{1}{n_j} \left(\sum_{i \in S_j} a_{\pi_i} - \bar{a}_{S_j} \right)^2 .$$

It holds that

$$\mathbb{E}[T_{\boldsymbol{\pi}} \mid \mathcal{Z}] = O_p(N^{-1}) ,$$

and

$$\mathrm{Var}[T_{\boldsymbol{\pi}} \mid \mathcal{Z}] = O_p(N^{-2}) ,$$

which gives

$$T_{\boldsymbol{\pi}} = O_p(N^{-1}) ,$$

condition on $\mathcal{Z}$. This completes the proof. □

Proof of Theorem 9. We provide the proof for the size and power of Spearman correlation coefficient test, Wilcoxon test, and Kruskal-Wallis test.

(a) Size of tests: Under H_0, we have the following result for $T_{\boldsymbol{\pi}}$

$$P\left(T_{\boldsymbol{\pi}} \geq c_{m,n} \mid \mathcal{Z}, H_0\right) = \alpha ,$$

which unconditionally gives

$$P\left(T_{\boldsymbol{\pi}} \geq c_{m,n} \mid H_0\right) = \alpha .$$

Given that T_{n+m} has the same distribution as $T_{\boldsymbol{\pi}}$, we have that

$$P\left(T_{n+m} \geq c_{m,n} \mid H_0\right) = \alpha .$$

It is straightforward to show that the same is true for all of the considered test statistics.

(b) Power of tests: For the power study, we need a propoer distance function related to the proposed test statistics. For Wilcoxon and Kruskal-Wallis tests, we can use the energy distance (Chu and Dai, 2024) between two distributions P and Q, defined as follows

$$D_E(P, Q) = 2\mathbb{E}\left[\delta(X, Y)\right] - \mathbb{E}\left[\delta(X, X')\right] - \mathbb{E}\left[\delta(Y, Y')\right] ,$$

where $X, X' \sim P$ and $Y, Y' \sim Q$, and δ is a proper distance metric on manifold space $\mathcal{M}$. For Wilcoxon and Kruskal-Wallis test statistics, under the alternative, by (Van der Vaart, 2000), it can be shown that

$$T_{n+m} = (\gamma_{n,m})\, D_E(P, Q) + o_p(1) ,$$

as $n, m \to \infty$, and

$$T_{n_1+n_2+\ldots+n_K} = (\gamma_{n_1,n_2,\cdots,n_K}) \sum_{i=1}^{K} \sum_{j=1}^{K} D_E(P_i, P_j) + o_p(1) \ ,$$

as $n_1, n_2, \ldots, n_K \to \infty$. where γ depends on sample size.

For Spearman correlation coefficient test, we can use the probability of concordance (Liu et al., 2018) between paired observations (X, Y) and (X_0, Y_0) where X and Y have the same marginal distribution as X_0 and Y_o; however, X_0 and Y_0 are independent. Then, we can write

$$D_C(P, Q) = \mathbb{E}_{(X,Y)|(X_0,Y_0)}\big[\ | \ c(P_C - P_D) \ | \ \big] \ ,$$

where c is chosen such that $-1 \leq c(P_C - P_D) \leq 1$, and

$$P_C = \Pr\left(\text{sign}\left(D(P, X), D(P, X_0)\right) \text{sign}\left(D(Q, Y), D(Q, Y_0)\right) = 1\right) \ ,$$

$$P_D = \Pr\left(\text{sign}\left(D(P, X), D(P, X_0)\right) \text{sign}\left(D(Q, Y), D(Q, Y_0)\right) = -1\right) \ ,$$

where $\text{sign}(a, b) = \{-1, 0, 1\}$ if $a < b$, $a = b$, and $a > b$, respectively. For Spearman correlation coefficient test, under the alternative, it can be shown that

$$T_n = (\gamma_n) \, D_C(P, Q) + o_p(1) \ ,$$

as $n \to \infty$.

Under the alternative, we have

$$\lim_{n,m\to\infty} D_E(P_n, Q_m) > 0 \ ,$$

$$\lim_{n_1,n_2,\ldots,n_K\to\infty} \sum_{i=1}^{K} \sum_{j=1}^{K} D_E(P_i, P_j) > 0 \ ,$$

$$\lim_{n\to\infty} D_C(P_n, Q_n) > 0 \ ,$$

Also, the we have

$$\lim_{n,m\to\infty} \gamma_{n,m} = O(\sqrt{n+m}) \ ,$$

$$\lim_{n_1,n_2,\ldots,n_K\to\infty} \gamma_{n_1,n_2,\cdots,n_K} = O(\sqrt{n_1+n_2+\ldots+n_K}) \ ,$$

$$\lim_{n\to\infty} \gamma_n = O(\sqrt{n}) \ .$$

Therefore, the corresponding test statistics diverge with rate $O(\sqrt{N})$. By Lemma 3, condition on $\mathcal{Z}$ and under null, we have

$$T_{\boldsymbol{\pi}} = o_p(1) \ ,$$

and thus $c_{n,m} = o_p(1)$, unconditionally. Combining all these results implies that,

$$\lim_{n,m\to\infty} P(T_{n+m} \geq c_{n,m} \mid H_A) \to 1 \ ,$$

under the alternative. This completes the proof.

□

Lemma 4. *Under the conditions of Assumption 7, the functionals $T(P)$ representing the proposed rank-depth test statistics t_n, are weakly continuous at P.*

Proof of Lemma 4. A functional T is weakly continuous at probability measure P if and only if

$$\forall \epsilon > 0, \ \exists \delta > 0, \ \exists n_0, \ \forall n, m \geq n_0 \ ,$$
$$\forall P_n, P_m : \delta_{TV}(P_n, P) < \delta \ \text{ and } \delta_{TV}(P_m, P) < \delta \Rightarrow \mid T(P_n) - T(P_m) \mid < \epsilon \ .$$

For $\delta_1 > 0$ and $\delta_{TV}(P_n, P) < \delta_1$ there exists n_0 such that for all $n > n_0$, the diameter of the closure of the compact set satisfies $\text{diam}\{T(P_n) : \delta_{TV}(P_n, P) < \delta_1, n > n_0\} \leq 2\epsilon_1$. As $n_0 \to \infty$, it can be verified that $\epsilon_1 \to 0$ and the set converges to the center point $T(P)$. Similarly, it can be seen that for $\delta_2 > 0$ and $\delta_{TV}(P_m, P) < \delta_2$ there exists n_0 such that for all $n > n_0$, the diameter of the closure of the compact set satisfies $\text{diam}\{T(P_m) : \delta_{TV}(P_m, P) < \delta_2, m > n_0\} \leq 2\epsilon_2$, such that $\epsilon_2 \to 0$ as $n_0 \to \infty$, and the set converges to the center point $T(P)$. Let $\delta = \max\{\delta_1, \delta_2\}$ such that $\delta_{TV}(P_n, P) < \delta$ and $\delta_{TV}(P_m, P) < \delta$. Also, let $\epsilon = \max\{2\epsilon_1, 2\epsilon_2\}$, then for all $n, m \geq n_0$, we have $\mid T(P_n) - T(P_m) \mid < \epsilon$, satisfying $\epsilon \to 0$ as $n_0 \to \infty$. It is easy to see that the above argument can be extended to the derived functional format of the proposed rank-depth test statistics under $t_n = T \mid P_n, Q_n$ for Spearman correlation coefficient test and quadrant test, and $t_{m+n} = T \mid P_n, Q_m$ for Wilcoxon test, and $T \mid P_n$ for kruskal-Wallis test. Therefore, these test statistics are weakly continuous. This completes the proof. □

Proof of Theorem 10. Consider i.i.d random variables $X_1, X_2, \ldots, X_n$ on $(\mathcal{M}, \mathcal{S})$, where $\mathcal{S}$ denotes the Borel sigma-algebra of manifold $\mathcal{M}$. For the proof, we need bounded Lipschitz metric δ_{BL}, defined as follows

$$\delta_{BL}(P, Q) = \sup_{f \in BL(\mathcal{S})} \left| \int f dP - \int f dQ \right| ,$$

where $BL(\mathcal{S})$ denotes the set of all real functions on $\mathcal{S}$, such that

$$\sup_{u\in\mathcal{S}}|f(u)|+\sup_{u,v\in\mathcal{S}}\frac{|f(u)-f(v)|}{d(u,v)}\leqslant 1\ .$$

A set $A\in\mathcal{S}$ is totally bounded if for any ϵ and a metric ℓ defined on $\mathcal{S}$, we can find a finite collection of ϵ-balls, i.e. balls with radius ϵ in metric ℓ, covering A. We denote the minimal cardinality of such a collection by $N(\epsilon,A,\ell)$, ϵ-covering number of A in ℓ. For a totally bounded subset $A\subseteq\mathcal{S}$, we can show by (Dudley et al., 1991) the following upper bound for ϵ-covering number

$$2N\left(6\epsilon,BL(A),\ell^1_{P_n}\right)e^{-18n\epsilon^2}\leq 2N\left(6\epsilon,BL(A),\ell^\infty_A\right)e^{-18n\epsilon^2}\ .$$

Approximating the functions in $BL(A)$ by stepwise functions (Mizera, 2010) gives the following inequality

$$N\left(6\epsilon,BL(A),\ell^\infty_A\right)\leq\left(\frac{1}{2\epsilon}\right)^{N(2\epsilon,A,d)}\ ,$$

where d is a metric for $\mathcal{S}$. Given that A is totally bounded, we can realize that $N\left(6\epsilon,BL(A),\ell^\infty_A\right)$ is bounded. Therefore, by (Dudley et al., 1991), it can be shown that

$$\sup_{Q\in\mathcal{P}(\mathcal{S})}P_Q\left[\sup_{f\in BL(A)}\left|\int f dQ_n-\int f dQ\right|>48\epsilon\right]\leq 2N\left(6\epsilon,BL(A),\ell^\infty_A\right)e^{-18n\epsilon^2}\ , \qquad (A.1)$$

which tends to 0, as $n\to\infty$, uniformly in $Q\in\mathcal{P}(\mathcal{S})$.

For $A\subseteq S$, and any $\epsilon>0$, assume that $Q(A)\leq\epsilon$, for all $Q\in\mathcal{P}(\mathcal{S})$. For the Bernoulli sequence of events $p=Q_n(A)\leq\epsilon$, we have

$$P_Q\left[Q_n(A)\geq 2\epsilon\right]\leq P_Q\left[|Q_n(A)-p|\geq\epsilon\right]\leq\frac{\mathrm{Var}\left(Q_n\right)}{\epsilon^2}=\frac{p(1-p)}{n\epsilon^2}\leq\frac{1}{n\epsilon}\ , \qquad (A.2)$$

for all $Q\in\mathcal{P}(\mathcal{S})$, which tends to 0. For any $\epsilon>0$, we can choose a subset A such that $P(A)\geq 1-\frac{\epsilon}{2}$. Choose n_0 such that Eq.A.1 is bounded by $\eta/3$, where $\eta>0$ is a fix value. Also, choose n_1 such that

$$\frac{n_1(1-\epsilon)}{2}\geq n_0\ \ ,\ \ \frac{4\epsilon}{n_1(1-\epsilon)}\leq\frac{\eta}{3}\ \ ,\ \ \frac{1}{48^2n_1\epsilon^2}\leq\frac{\eta}{3}\ . \qquad (A.3)$$

For any $Q \in \mathcal{P}(\mathcal{S})$ such that $\delta_{TV}(Q,P) \leq \epsilon/2$, we get $Q(A) \geq 1-\epsilon$. Denote the random number of $X_i \in A$ by N_Q and define $Q_{\cdot|A}$ the conditional probability on A. Let $p = Q(A) \geq 1-\epsilon$.

For $n \geq n_1$, we obtain that

$$\begin{aligned}
&P_Q\left[\sup_{f\in BL(\mathcal{S})}\left|\frac{1}{np}\sum_{X_i\in A} f\left(X_i\right) - \frac{1}{p}\int_A f dQ\right| \geq \frac{96\epsilon}{p}\right] \\
&\leq P_Q\left[\sup_{f\in BL(\mathcal{S})}\left|\frac{1}{np}\sum_{X_i\in A} f\left(X_i\right) - \frac{1}{N_Q}\sum_{X_i\in A} f\left(X_i\right)\right| \geq \frac{48\epsilon}{p}\right] \\
&+ P_Q\left[\sup_{f\in BL(\mathcal{S})}\left|\frac{1}{N_Q}\sum_{X_i\in A} f\left(X_i\right) - \frac{1}{Q(A)}\int f dQ\right| \geq \frac{48\epsilon}{p}\right] \\
&\leq P_Q\left[\left|\frac{N_Q}{n} - p\right| > 48\epsilon\right] + P_Q\left[\sup_{f\in BL(A)}\left|\frac{1}{N_Q}\sum_{X_i\in A} f\left(X_i\right) - \int f dQ_{\cdot|A}\right| \geq 48\epsilon\right] \ .
\end{aligned}$$

For the first term, we can use an argument similar to that of Eq.A.2 to get an upper bound, as follows

$$P_Q\left(\left|\frac{N_Q}{n} - p\right| > 48\epsilon\right) \leq \frac{p(1-p)}{48^2 n\epsilon^2} \leq \frac{\eta}{3} \ ,$$

where we used results from Eq.A.3.

For the second term, we can rewrite it as follows

$$\sum_{m=1}^{\infty} P_Q\left[\sup_{f\in BL(A)}\left|\frac{1}{N_Q}\sum_{X_i\in A} f\left(X_i\right) - \int f dQ_{\cdot|A}\right| \geq 48\epsilon \middle| N_Q = m\right] P_Q\big[N_Q = m\big] \ .$$

For $m \leq n$, it is easy to verify that this term is bounded by $\eta/3$. For $m > n$, we can rewrite the inequality using $Y_1, Y_2, \ldots, Y_m$ sampled i.i.d from $Q_{\cdot|A}$, as below

$$\sum_{m>n}^{\infty} P_{Q_{\cdot|A}}\left[\sup_{f\in BL(A)}\left|\frac{1}{m}\sum_{i=1}^{m} f\left(Z_i\right) - \int f dQ_{\cdot|A}\right| \geq 48\epsilon\right] P_Q\big[N_Q = m\big] \leq \frac{\eta}{3} \ .$$

Given that η is arbitrarily chosen, we conclude that for any $\epsilon > 0$, there exists a totally bounded subset A of $\mathcal{S}$ such that

$$P_Q\left[\sup_{f\in BL(\mathcal{S})}\left|\int_A f dQ_n - \int_A f dQ\right| > 96\epsilon\right] \to 0 \ , \tag{A.4}$$

uniformly in all $Q \in \mathcal{P}(\mathcal{S})$, such that $\delta_{TV}(P,Q) \leq \epsilon/2$.

Let $d_{min}(\mathcal{S}) = \inf_{x,y\in\mathcal{S}} \ell(x,y)$ and $\text{diam}(\mathcal{S}) = \sup_{x,y\in\mathcal{S}} \ell(x,y)$ (Gibbs and Su, 2002). Given $P \in \mathcal{P}(\mathcal{S})$, and α, choose ϵ such that $96\epsilon \leq (\frac{d_{min}(\mathcal{S})}{2(\text{diam}(\mathcal{S})+1)})\alpha$. Using the relation between δ_{TV} and δ_{BL}

$$\left(\frac{d_{min}(\mathcal{S})}{\text{diam}(\mathcal{S})+1}\right)\delta_{T_V}(P,Q) \leq \delta_{BL}(P,Q) \leq \delta_{T_V}(P,Q)\ ,$$

and using the argument given in Eq.A.4, we get

$$\begin{aligned}
P_Q\left(\delta_{TV}\left(Q_n,P\right) \geq \alpha\right) &\leq P_Q\left[\delta_{BL}\left(Q_n,P\right) \geq \left(\frac{d_{min}(\mathcal{S})}{\text{diam}(\mathcal{S})+1}\right)\alpha\right] \\
&= P_Q\left[\sup_{f\in BL(\mathcal{S})}\left|\int f dP - \int f dQ_n\right| \geq \left(\frac{d_{min}(\mathcal{S})}{\text{diam}(\mathcal{S})+1}\right)\alpha\right] \\
&\leq P_Q\left[\sup_{f\in BL(\mathcal{S})}\int_{S/A}|f|dQ_n + \sup_{f\in BL(S)}\int_{S/A}|f|dQ \geq \left(\frac{d_{min}(\mathcal{S})}{2(\text{diam}(\mathcal{S})+1)}\right)\alpha\right] \\
&+ P_Q\left[\sup_{f\in BL(\mathcal{S})}\left|\int_A f dQ_n - \int_A f dQ\right| \geq \left(\frac{d_{min}(\mathcal{S})}{2(\text{diam}(\mathcal{S})+1)}\right)\alpha\right] \\
&\leqslant P_Q\left[Q_n(S/k) > \left(\frac{d_{min}(\mathcal{S})}{2(\text{diam}(\mathcal{S})+1)}\right)\alpha\right] + P_Q\left[\sup_{f\in BL(\mathcal{S})}\left|\int_A f dQ_n - \int_A f dQ\right| > 96\epsilon\right]\ .
\end{aligned}$$

Similar to Eq. A.4, we can show that the second term is bounded by $\eta/3$. Given that $Q(\mathcal{S}/A) \leq \epsilon$ (since $Q(A) \geq 1-\epsilon$), using Eq. A.2, we get the following bound for the first term

$$P_Q\left[Q_n(\mathcal{S}/A) \geq \left(\frac{d_{min}(\mathcal{S})}{2(\text{diam}(\mathcal{S})+1)}\right)\alpha\right] \leq P_Q\left[Q_n(\mathcal{S}/A) \geq 96\epsilon\right] \leq P_Q\left[Q_n(\mathcal{S}/A) \geq 2\epsilon\right] \leq \frac{\eta}{3}\ .$$

Therefore, for any $\alpha, \eta > 0$, there exists $\delta > 0$, such that

$$P_Q\left[\delta_{TV}\left(Q_n,P\right) > \alpha\right] < \frac{\eta}{3}\ , \tag{A.5}$$

when $\delta_{TV}(P,Q) < \delta$.

By Lemma 4, for $\delta_{TV}(P,Q) \leq \delta$, we have for $\tau \in T(Q)$ that $|\tau - T(P)| \leq \epsilon$.

Using Eq. A.5 for any $\tau \in T(Q_n)$, we have

$$P_Q\left[|\tau - T(p)| \geq 2\epsilon\right] < \frac{\eta}{3} .$$

Using an argument similar to (Huber, 2011), we get

$$\delta_{TV}\left(\mathcal{G}_Q\left(t_n\right), \delta_{T(P)}\right) \leq \frac{\eta}{3} ,$$

where $\delta_{T(P)}$ is the Dirac measure concentrated in $T(P)$. By Assumption 7 and Eq. A.5, we get

$$\delta_{TV}\left(\mathcal{G}_Q\left(t_n\right), \mathcal{G}_P\left(t_n\right)\right) \leq \frac{\eta}{3}$$

when $\delta_{TV}(P, Q) \leq \delta$. This completes the proof. □

Proof of Theorem 11. The proof can be done by contradiction. Note that this argument holds for both ε_A and ε_R. Let $\epsilon_2^* < \epsilon_1^*$. Then, for any choice of $\varepsilon_2^* < \varepsilon < \varepsilon_1^*$, the sequence of tests $T(P_n)$ breakdown at ε-level of contamination. Therefore, there exists a sequence of contaminating distributions $\{Q_m\}_{m=1}^{\infty}$ with probability density function $\{\widehat{g}_m\}_{m=1}^{\infty}$ such that for contaminated distribution $P_{n,\varepsilon,m} = (1-\varepsilon)\, P_n + \varepsilon Q_m$, we have $|T(P_{n,\varepsilon,m}) - T(P)| \to \infty$. Since $\varepsilon < \varepsilon_1^*$, the sequence of tests $T(P)$ will not breakdown at ε-level of contamination, i.e. the functional $T(P_{\varepsilon,m})$ will not breakdown. Therefore, we will have a contradiction if we can show $|T(P_{\varepsilon,m}) - T(P)| \to \infty$, as $m \to \infty$.

Let us define $f_{\varepsilon,m} = (1-\varepsilon) f + \varepsilon g_m$ and $\widehat{f}_{n,\varepsilon,m} = (1-\varepsilon)\widehat{f}_n + \varepsilon g_m$, for any fixed m. Given that $\int_{\mathcal{M}} |\widehat{f}_n(x) - f(x)|^p dx \overset{a.s.}{\to} 0$, there exists a subsequence $\{\widehat{f}_{n_j}\}_{j=1}^{\infty}$ for which we have a pointwise converge $\widehat{f}_{n_j}(x) \overset{a.s.}{\to} f(x)$ for any $x \in \mathcal{M}$, as $n_j \to \infty$. Let $\delta_{n_j,\varepsilon,m} = |\widehat{f}_{n_j,\varepsilon,m} - f_{\varepsilon,m}|$. It can be shown that $\int_{\mathcal{M}} \delta_{n,\varepsilon,m}^p dx \overset{a.s.}{\to} 0$, therefore, by the same argument we have a pointwise converge $f_{n_j,\varepsilon,m}(x) \overset{a.s.}{\to} f_{\varepsilon,m}(x)$ for any $x \in \mathcal{M}$, as $n_j \to \infty$. Therefore, for any fixed ε, it holds that

$$\sup_{f:\ P \in \mathcal{P}(\mathcal{M})} \left\{ |T(P_{n_j,\varepsilon,m}) - T(P_{\varepsilon,m})| \right\} \overset{a.s.}{\to} 0 ,$$

which implies

$$\sup_{f:\ P \in \mathcal{P}(\mathcal{M})} \left\{ \left| \max_{\mathcal{I}_{n-m} \subset \mathcal{N}} \sup_{Z \in \Pi_n(\mathcal{I}_{n-m})} \phi_{n_j}(Z) - \max_{\mathcal{I}_{n-m} \subset \mathcal{N}} \sup_{Z \in \Pi_n(\mathcal{I}_{n-m})} \phi(Z) \right| \right\} \overset{a.s.}{\to} 0 ,$$

$$\sup_{f:\ P \in \mathcal{P}(\mathcal{M})} \left\{ \left| \min_{\mathcal{I}_{n-m} \subset \mathcal{N}} \inf_{Z \in \Pi_n(\mathcal{I}_{n-m})} \phi_{n_j}(Z) - \min_{\mathcal{I}_{n-m} \subset \mathcal{N}} \inf_{Z \in \Pi_n(\mathcal{I}_{n-m})} \phi(Z) \right| \right\} \overset{a.s.}{\to} 0 ,$$

where ϕ_n is the empirical decision function based on $T(P_n)$.

Therefore, for any $\varepsilon < \varepsilon_1^*$, and any $\delta > 0$, there exists $m > M$, $n > n'_m$, with $\varepsilon = m/n$, such that

$$\sup_{f:\ P\in\mathcal{P}(\mathcal{M})} \left\{|T(P_{n,\varepsilon,m}) - T(P)|\right\} > \delta \ ,$$

$$\sup_{f:\ P\in\mathcal{P}(\mathcal{M})} \left\{ \left| \max_{\mathcal{I}_{n-m}\subset\mathcal{N}} \sup_{Z\in\Pi_n(\mathcal{I}_{n-m})} \phi_n(Z) - \phi(X) \right| \right\} = 1 \ ,$$

$$\sup_{f:\ P\in\mathcal{P}(\mathcal{M})} \left\{ \left| \max_{\mathcal{I}_{n-m}\subset\mathcal{N}} \inf_{Z\in\Pi_n(\mathcal{I}_{n-m})} \phi_n(Z) - \phi(X) \right| \right\} = 1 \ ,$$

and

$$\sup_{f:\ P\in\mathcal{P}(\mathcal{M})} \left\{|T(P_{\varepsilon,m})(\omega) - T(P)|\right\} > \delta \ ,$$

$$\sup_{f:\ P\in\mathcal{P}(\mathcal{M})} \left\{ \left| \max_{\mathcal{I}_{n-m}\subset\mathcal{N}} \sup_{Z\in\Pi_n(\mathcal{I}_{n-m})} \phi(Z) - \phi(X) \right| \right\} = 1 \ ,$$

$$\sup_{f:\ P\in\mathcal{P}(\mathcal{M})} \left\{ \left| \max_{\mathcal{I}_{n-m}\subset\mathcal{N}} \inf_{Z\in\Pi_n(\mathcal{I}_{n-m})} \phi(Z) - \phi(X) \right| \right\} = 1 \ .$$

Since $\delta > 0$ is arbitrary, the result holds for both ε_A and ε_R. □

A.2 Proofs from Chapter 3

A.2.1 Change point on General Metric Spaces

Proof of Theorem 12. On a compact manifold $\mathcal{M}$ with dimension d, define the lie group G acting on $\mathcal{M}$, $(g, x) \in G \times \mathcal{M} \to g.x \in \mathcal{M}$. Also, define the Lebesgue spaces $L^p(\mathcal{M})$ for $1 \leq p \leq \infty$. For the current analysis, we focus on the class of functions $f \in L^p(\mathcal{M})$. The lie algebra $\text{Lie}(G)$ of G is defined as $X \in \text{Lie}(G)$ such that $e^X \in G$. On this space, there exists a Euclidean structure $\langle \cdot, \cdot \rangle$ such that

$$\forall g \in G,\ \forall x, y \in \text{Lie}(G) \ ,\ \langle gxg^{-1}, gyg^{-1} \rangle = \langle x, y \rangle \ ,\ |x|^2 = \langle x, x \rangle \ .$$

This induces the following metric on $\mathcal{M}$

$$d_{\mathcal{M}}(x, y) = \inf \left\{ \mid X \mid : e^X \cdot x = y \right\} .$$

which is invariant under action $g \in G$, i.e.

$$\forall g \in G,\ x, y \in \mathcal{M},\ d_{\mathcal{M}}(g \cdot x, g \cdot y) = d_{\mathcal{M}}(x, y).$$

Also, define

$$D_X f(x) = \frac{d}{dt} f\left(e^{tX} \cdot x\right)\Big|_{t=0},\ x \in \mathcal{M},\ t \in \mathbb{R},\ f \in L^p(\mathcal{M})\ .$$

Consider an orthonormal basis X_i, $i = 1, ..., \tau$, on Lie(G), we have the following Laplace operator on $L^2(\mathcal{M})$

$$\mathcal{L}_{\mathcal{M}} = D^2_{X_1} + D^2_{X_2} + \ldots + D^2_{X_n}\ .$$

It can be verified that the above operator coincide with Laplace-Beltrami operator on the spaces of $\mathcal{S}^d$ and $\mathbb{T}^d$, i.e. d-dimensional sphere and torus spaces (Canzani, 2013).

By spectral decomposition we get eigenfunctions $\{e_k\}_{k\in\mathbb{N}}$, which form an orthonormal basis on $L^2(\mathcal{M})$. We can also define for $n \in \mathbb{N}$, a finite-dimensional subspace of $\mathcal{M}$, i.e. $\mathcal{E}_{n_0}(\mathcal{M})$ of $L^2(\mathcal{M})$ spanned by the eigenfunctions e_k of $\mathcal{L}_{\mathcal{M}}$

$$\mathcal{E}_{n_0}(\mathcal{M}) = \left\{ x \mapsto \sum_{k:\lambda_k \le n_0} c_k e_k(x) \,|\, c_k \in \mathbb{R}, \lambda_k \text{ are eigenvalues of } e_k \right\}\ .$$

For the proof, we use a wavelet-type representation of density function $f \in L^2(\mathcal{M})$ (Kerkyacharian et al., 2012; Giné and Nickl, 2009). For two functions $h_1, h_2 \in L^2(\mathcal{M})$, let $\langle h_1, h_2 \rangle$ denote their $L^2(\mathcal{M})$-inner product. Also let $\|h\|_\infty = \sup_{y\in\mathcal{M}} |h(y)|$ denote the supremum norm of $h \in L^2(\mathcal{M})$. Consider infinitely-differentiable nonnegative function γ defined on $[0, \infty)$ such that γ equals 1 on $(0, 1/2]$ and compact on $[0, 1]$. The liner operator S_j, $j \in \mathbb{N}$ is defined as

$$S_0 h = \int_{\mathcal{M}} h(x) dx\ ,$$
$$S_j h(x) = \int_{\mathcal{M}} S_j(x, y) f(y) dy,\ j \in \mathbb{N}\ ,$$

where

$$S_j(x, y) = \sum_{k:\lambda_k < 2^{2j}} \left(\frac{\lambda_k}{2^{2j}}\right) e_k(x)\overline{e_k}(x)\ ,$$

satisfying $S_j h = h$

$$\langle S_j h, h\rangle = \sum_{k:\lambda_k<2^{2j}} \gamma\left(\frac{\lambda_k}{2^{2j}}\right) \langle e_k\overline{e_k}h, h\rangle \le \sum_{k:\lambda_k<2^{2j}} (1) \langle e_k\overline{e_k}h, h\rangle = \|h\|^2\ .$$

Also, by the definition of function $\gamma(\cdot)$, it can be verified that, for $h \in E_{2^{j-1}}(\mathcal{M})$

$$\begin{aligned}\langle S_j h, h\rangle &= \sum_{k:\lambda_k<2^{2j}} \gamma\left(\frac{\lambda_k}{2^{2j}}\right)\langle e_k\overline{e_k}h, h\rangle \\ &= \sum_{k:\lambda_k<2^{2j}} (1)\,\langle e_k\overline{e_k}h, h\rangle = \|h\|^2 = \langle h, h\rangle \ . \end{aligned} \tag{A.6}$$

For any $h \in L^2(\mathcal{M})$, it holds that $\lim_{j\to\infty} \|S_j h - h\| = 0$. We use the following representation of the kernel S_j

$$S_j(x,y) = \int_{\mathcal{M}} S_j^{(*)}(x,u) S_j^{(*)}(u,y) du \ ,$$

where

$$S_j^{(*)}(x,y) = \sum_{k:\lambda_k<2^{2j}} \sqrt{\gamma\left(\frac{\lambda_k}{2^{2j}}\right)} e_k(x)\overline{e_k}(y) \ .$$

For the current analysis, we use cubature formula on $E_k(\mathcal{M})$ proposed by (Geller and Pesenson, 2011). For $k \in \mathbb{N}$, consider the subset Ω_k of $\mathcal{M}$ with $|\Omega_k| \leq Ck^{d/2}$ and positive real numbers $\alpha_\eta > 0$, where η's are the elements of Ω_k, such that

$$\forall h \in E_k(\mathcal{M}), \quad \int_{\mathcal{M}} h(x)dx = \sum_{\eta\in\Omega_k} \alpha_\eta h(\eta) \ .$$

Given that $S_j^{(*)}(x,u) \in E_{2^{2j}}(\mathcal{M})$, we infer that $S_j^{(*)}(x,u)S_j^{(*)}(u,y) \in E_{\tau 2^{2j+2}}(\mathcal{M})$ (Geller and Pesenson, 2011). Therefore

$$S_j(x,y) = \int_{\mathcal{M}} S_j^{(*)}(x,u) S_j^{(*)}(u,y) du = \sum_{\eta\in\Omega_{\tau 2^{2j+2}}} \alpha_\eta S_j^{(*)}(x,\eta) S^{(*)}(\eta,y) \ .$$

$$S_j h(x) = \int_{\mathcal{M}} S_j(x,y)h(y)dy = \sum_{\eta\in\Omega_{\tau 2^{2j+2}}} \sqrt{\alpha_\eta} S_j^{(*)}(x,\eta) \int_{\mathcal{M}} \sqrt{\alpha_\eta} S_j^{(*)}(\eta,y)h(y)dy \ .$$

Using the above cubature formula, we define the multi-resolution wavelet-type approximation of probability density function $f \in L^2(\mathcal{M})$, as follows

$$S_j f = S_0 f + \sum_{0\leqslant l\leqslant j-1} \sum_{\eta} \langle f, \psi_{l\eta}\rangle \psi_{l\eta} \ ,$$

which is the projection of $f \in L^2(\mathcal{M})$ to $E_{\tau 2^{2l+2}}$ at resolution levels $0 \leq l \leq j-1$. In this representation, $\psi_{l\eta}$ is wavelet scaling function indexed by cubature points $\eta \in \Omega_{\tau 2^{2l+2}}$ defined as

$$\psi_{l\eta}(x) = \sqrt{\alpha_{\eta l}} \sum_{k:\lambda_k < 2^{2l}} \sqrt{\gamma\left(\frac{\lambda_k}{2^{2l+1}}\right) - \gamma\left(\frac{\lambda_k}{2^{2l}}\right)} e_k(x)\overline{e_k}(\eta) \ .$$

As $j \to \infty$, $\|f - A_j f\| \to 0$. We can achieve upper bonds for $\|\psi_{l\eta}\|_2$ and $\|\psi_{l\eta}\|_\infty$, for $l \in \mathbb{N}$ and $\eta \in \Omega_{\tau 2^{2l+2}}$, as follows

$$x \mapsto \left(\sum_{k:\lambda_k \leq 2^{2l}} e_k(x)\overline{e_k}(\eta)\right)^2 \in E_{\tau 2^{2l+2}}(\mathcal{M}) \ ,$$

$$\begin{aligned}\int_{\mathcal{M}} \left(\sum_{k:\lambda_k \leq 2^{2l}} e_k(x)\overline{e_k}(\eta)\right)^2 dx &= \sum_{\xi \in \Omega_{\tau 2^{2l+2}}} \alpha_\xi \left(\sum_{k:\lambda_k \leq 2^{2l}} e_k(\xi)\overline{e_k}(\eta)\right)^2 \\ &\geq \alpha_\eta \left(\sum_{k:\lambda_k \leq 2^{2l}} e_k(\eta)\overline{e_k}(\eta)\right)^2 \ ,\end{aligned}$$

implying that

$$\alpha_\eta \leq \frac{1}{\sum_{k:\lambda_k \leq 2^{2l}} e_k(\eta)\overline{e_k}(\eta)} \ ,$$

where we used the fact that

$$\int_{\mathcal{M}} \left(\sum_{k:\lambda_k \leq 2^{2l}} e_k(x)\overline{e_k}(\eta)\right)^2 dx = \sum_{k:\lambda_k \leq 2^{2l}} e_k(\eta)\overline{e_k}(\eta) \ .$$

Therefore, for any $\eta \in \Omega_{\tau 2^{2l+2}}$, we have

$$\begin{aligned}\|\psi_{l\eta}\|_2 = \int_{\mathcal{M}} \psi_{l\eta}^2(x)dx &= \alpha_\eta \sum_{k:\lambda_k \leq 2^{2l}} \left(\gamma\left(\frac{\lambda_k}{2^{2l+1}}\right) - \gamma\left(\frac{\lambda_k}{2^{2l}}\right)\right) e_k(\eta)\overline{e_k}(\eta) \\ &\leq \frac{\sum_{k:\lambda_k \leq 2^{2l}} \left(\gamma\left(\frac{\lambda_k}{2^{2l+1}}\right) - \gamma\left(\frac{\lambda_k}{2^{2l}}\right)\right) e_k(\eta)\overline{e_k}(\eta)}{\sum_{k:\lambda_k \leq 2^{2l}} e_k(\eta)\overline{e_k}(\eta)} \\ &\leq \sum_{k:\lambda_k \leq 2^{2l}} (1) e_k(\eta)\overline{e_k}(\eta) \left(\sum_{k:\lambda_k \leq 2^{2l}} e_k(\eta)\overline{e_k}(\eta)\right)^{-1} = 1.\end{aligned}$$

In (Geller and Mayeli, 2009), it was proved that for a function h applied to elliptical operator $h\left(\mathcal{L}/2^l\right)$ cm and any integer $n \geq 0$, we have

$$\sum_{k:\lambda_k \leq 2^{2l}} h\left(\frac{\lambda_k}{2^{2l}}\right) e_k(x)\overline{e_k}(\eta) \leq \frac{c_n(f)2^{ld}}{\left(1+2^{ld}d(\eta,x)\right)^n} \ .$$

Therefore, if we take $h(t) = \sqrt{\gamma(t/2)-\gamma(t)}$, an upper bound on $\|\psi_{l\eta}\|_\infty$ can be achieved, which has the form

$$\|\psi_{l\eta}\|_\infty \leq C_1(\mathcal{M})2^{ld/2} \ , \tag{A.7}$$

where $C_1(\mathcal{M})$ is a constant which depends on the manifold $\mathcal{M}$.

Let $X_1, X_2, \ldots, X_n$ denote i.i.d random variables taking values on $\mathcal{M}$, with probability measure P which has probability density function $f : \mathcal{M} \to (0,\infty)$, i.e. for $\Omega \subseteq M$, we have $P(\Omega) = \int_\Omega f(x)dx$. The wavelet-type estimator of f is defined as

$$\widehat{f}_n(j,y) = \frac{1}{n}\sum_{i=1}^n S_j\left(X_i,y\right) = \int_{\mathcal{M}} S_j(x,y)dP_n(x) = S_j\left(P_n\right)(y), \quad y \in \mathcal{M} \ .$$

It can be shown that conditioning on $\dot{J}_n, 2^{j_n} \approx n^\epsilon, 0<\epsilon<1$ such that $\sqrt{\dfrac{n}{2j_n^{d/2}j_n^2}} \to \infty$, we get the following

$$\limsup_{n\to\infty} \sqrt{\frac{n}{2j_n^{d/2}j_n^2}} \sup_{y\in\mu}\left(\widehat{f}_n(j,y) - \mathbb{E}[\widehat{f}_n(j,y)]\right) = O(1) \ ,$$

almost surely. To show this, note that the supremum of any $\widehat{f}_n \in E_{2^{2j-1}}(\mathcal{M})$ can be bounded using Eq. A.6 and Eq. A.7, as follows

$$\widehat{f}_n(y) = S_j\widehat{f}_n(y) = \sum_{0\leqslant l\leqslant j-1}\sum_{\eta\in\Omega_{\tau 2^{2l+2}}} \langle\widehat{f}_n,\psi_{l\eta}\rangle\psi_{l\eta} \ ,$$

where we assumed $S_0\widehat{f}_n = 0$. It holds that

$$\begin{aligned}\sup_{x\in\mathcal{M}}|\widehat{f}_n(y)| &\leq \max_{\eta\in\Omega_{\tau 2^{2l+2}}}|\langle\widehat{f}_n,\psi_{l\eta}\rangle| \sup_{y\in\mathcal{M}}\sum_{0\leqslant l\leqslant j-1}\sum_{\eta\in\Omega_{\tau 2^{2l+2}}}|\psi_{l\eta}(y)| \\ &\leq j_n c_0(\mathcal{M})2^{j_n d/2}\max_{\eta\in\Omega_{\tau 2^{2l+2}}}|\langle\widehat{f}_n,\psi_{l\eta}\rangle| \ .\end{aligned}$$

Substituting $\widehat{f}_n(j,y) = S_j\left(P_n\right)(y)$, we get

$$\begin{aligned}\sup_{y\in\mathcal{M}}\left\|\widehat{f}_n(j,y) - \mathbb{E}[\widehat{f}_n(j,y)]\right\| &= \sup_{y\in\mathcal{M}}\left|\sum_{0\le l\le j-1}\sum_{\eta\in\Omega_{\tau 2^{2l+2}}}\psi_{l\eta}(y)\left\langle P_n - P, \psi_{l\eta}\right\rangle\right| \\ &\le j_n C_1(\mathcal{M}) 2^{j_n d/2}\max_{\eta\in\Omega_{\tau 2^{2l+2}}}\left|\langle P_n - P, \psi_{l\eta}\rangle\right| .\end{aligned}$$

For a finite empirical process bounded by the sequence of functions $\{\psi_{\eta_k l}\}_{k=1}^{|\Omega_{\tau 2^{2l+2}}|}$ with envelope $U = 2^{j_n d/2} C_1(\mathcal{M})$, the class of functions

$$\mathcal{G} = \left\{\psi_{j\eta_1}/2U, \ldots, \psi_{j\eta_{|\Omega_{\tau 2^{2l+2}|}}}/2U\right\} ,$$

is uniformly bounded by $1/2$, and has the variance $\mathbb{E}[g^2(X)] = \dfrac{\|f\|_\infty}{2^{\frac{ld}{2}+2} C_1^2(\mathcal{M})}$.

Therefore

$$\begin{aligned}\max_{\eta\in\Omega_{\tau 2^{2l+2}}}\left|\langle P_n - P, \psi_{l\eta}\rangle\right| \le \|P_n - P\|_{\mathcal{G}} &= \sup_{g\in\mathcal{G}}\left|\int_{y\in\mathcal{M}} g(y) d\left(P_n - P\right)(y)\right| \\ &= \max_{m=1,\ldots,|\Omega_{\tau 2^{2l+2}}|}\left|\frac{1}{n}\sum_{i=1}^{n}\left(g_m\left(X_i\right) - \mathbb{E}[g_m(X)]\right)\right| \\ &\le \sum_{m=1}^{|\Omega_{\tau 2^{2l+2}}|}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(g_m\left(X_i\right) - \mathbb{E}[g_m(X)]\right)\right|\right) \\ &= 2^{ld+2}\left(\left(l/n\sum_{i=1}^{n}\left(g_m\left(x_i\right) - Eg_m(x)\right)\right)\right) .\end{aligned}$$

The final result can be obtained using Bernstein's inequality for the sum of centered Gaussian random variables bounded in absolute value by 1 (Massart, 2007), as follows

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} Z_i\right| \ge \sqrt{\frac{2t\mathbb{E}[Z_i^2]}{n}} + \frac{t}{3n}\right) = 2e^{-t} ,$$

which gives

$$\begin{aligned}\sup_{y\in\mathcal{M}}\left|\widehat{f}_n(j,y) - \mathbb{E}[\widehat{f}_n(j,y)]\right| &\le \left(j_n c_0(\mathcal{M}) 2^{j_n d/2}\right)\left(\sqrt{\frac{2\|f\|_\infty}{n 2^{\frac{j_n d}{2}+2} C_1^2(\mathcal{M})}}\right) \\ &= \sqrt{\frac{2^{j_n d/2} j_n^2}{n}}\left(\frac{\|f\|_\infty}{2}\right)^{1/2}\left(\frac{c_0\left(\mathcal{M}\right)}{C_1(\mathcal{M})}\right) .\end{aligned}$$

Therefore, we have the following asymptotic convergence result

$$\limsup_{n\to\infty} \sqrt{\frac{n}{2^{j_n d/2} j_n^2}} \sup_{y\in\mathcal{M}} \left| \widehat{f}_n(j,y) - \mathbb{E}[\widehat{f}_n(j,y)] \right| \leq K \ ,$$

where constant K depends on manifold $\mathcal{M}$ and $\|f\|_\infty$.

The approximation bound due to bias, i.e. $\left\| \mathbb{E}[\widehat{f}_n(j,y)] - f(j,y) \right\|_\infty$ can be controlled by $C2^{-j\kappa}$ assuming that f is a κ-regular function, i.e. $f \in C^\kappa(\mathcal{M})$ for $\kappa \in \mathbb{N}$ (Kerkyacharian et al., 2012).

As a result, we get

$$\limsup_{n\to\infty} \sqrt{\frac{n}{2^{j_n d/2} j_n^2}} \sup_{y\in\mathcal{M}} \left| \widehat{f}_n(j,y) - f(j,y) \right| \leq K \ .$$

By using Lebesgue's dominated convergence theorem, we get the following result for the convergence of empirical probability measure P_n to theoretical probability measure P

$$\limsup_{n\to\infty} \sqrt{\frac{n}{2^{j_n d/2} j_n^2}} \sup_{\Omega\subseteq\mathcal{M}} |P_n(\Omega) - P(\Omega)| \leq K \ .$$

Given that the depth function $D(P,y)$ is a continuous function of probability measure P, by continuous mapping theorem and Assumption 3, we get the following result for the asymptotic convergence of empirical depth function

$$\limsup_{n\to\infty} \sqrt{\frac{n}{2^{j_n d/2} j_n^2}} \sup_{y\in\mathcal{M}} |D(P_n,y) - D(P,y)| \leq K \ .$$

To determine the constant K on sphere and torus manifolds, we can use the following result.

Sphere: Let $h(x_1,\ldots,x_d)$ represent the spherical coordinates for $\mathcal{S}^d \subset \mathbb{R}^d$. We can represent any $x \in \mathbb{R}^{d+1}$ as $rh(x_1,\ldots,x_d)$ for $r>0$. It can be verified that the following relation holds for Laplacian operator Δ on $\mathbb{R}^{d+1}$ and $\mathcal{S}^d$

$$\Delta_{\mathbb{R}^{d+1}} = -\frac{1}{r^d}\frac{\partial}{\partial r}\left(r^d \frac{\partial}{\partial r}\right) + \frac{1}{r^2}\Delta_{\mathcal{S}^d} \ .$$

Let

$$\begin{aligned}
\mathcal{P}_k &= \{\text{homogeneous polynomials of degree } k\} \ , \\
\mathcal{H}_k &= \left\{P \in \mathcal{P}_k : \Delta_{R^{d+1}} P = 0\right\} \ , \\
H_k &= \left\{P_{|\mathcal{S}^d} : P \in \mathcal{H}_k\right\} \ .
\end{aligned}$$

where $P_{|\mathcal{S}^d}$ is the sum of spherical harmonics of degree $\leq k$ for $P \in \mathcal{P}_k$ (Canzani, 2013). For $r^2\mathcal{P}_{k-2} = \{r^2 P : P \in \mathcal{P}_{k-2}\}$, we have that

$$\mathcal{P}_k = \mathcal{H}_k \oplus r^2 \mathcal{P}_{k-2} \ .$$

Therefore

$$\mathcal{P}_{2k} = \mathcal{H}_{2k} \oplus r^2\mathcal{H}_{2k-2} \oplus r^4\mathcal{H}_{2k-4} \oplus \ldots \oplus r^{2k}\mathcal{H}_0 \ ,$$

satisfying

$$\begin{aligned}
\dim\left(H_k\right) &= \dim\left(\mathcal{P}_k\right) - \dim\left(\mathcal{P}_{k-2}\right) \\
&= \binom{d+k}{d} - \binom{d+k-2}{d} \\
&= (2k+d-1)\frac{(k+d-2)!}{k!(d-1)!} \ .
\end{aligned}$$

Given that $\mathcal{L}$ coincides with the Laplace-Beltrami operator

$$L^2\left(\mathcal{S}^d\right) = \bigoplus_k \mathcal{H}_k\left(S^d\right) = \ker\left(\Delta - \lambda_k I\right) \ ,$$

with eigenfunctions e_k which are spherical harmonics with eigenvalues $k(k+d-1)$, we have that $\mathcal{E}_{n_0}\left(\mathcal{S}^d\right)$ corresponds to the spaces $\mathcal{P}_{n_0}\left(\mathcal{S}^d\right)$ of spherical polynomials, spanned by mutually orthogonal spaces $\mathcal{H}_k\left(\mathcal{S}^d\right), 0 \leq k \leq n_0$ of spherical harmonics.

For every $x \in \mathcal{S}^d$, we have

$$\sum_i \left|e_i^k(x)\right|^2 = \frac{\dim\left(\mathcal{H}_k\left(S^d\right)\right)}{|S^d|} \ .$$

Therefore

$$\begin{aligned}
\|\psi_{l\eta}\|_\infty &= \sup_{x\in\mathcal{S}^d}\sqrt{\alpha_{\eta l}}\sum_{k:\lambda_k<2^{2l}}\sqrt{\gamma\left(\frac{\lambda_k}{2^{2l+1}}\right)-\gamma\left(\frac{\lambda_k}{2^{2l}}\right)}e_k(x)\overline{e_k}(\eta)\\
&\leq \sqrt{\sum_{k:\lambda_k<2^{2l}}\left|e_i^k(x)\right|^2}=\sqrt{\frac{\dim\left(\mathcal{H}_k\left(S^d\right)\right)}{|\mathcal{S}^d|}}\\
&= \sqrt{\frac{\dim\left(P_{[k]_{\lambda_k\leq 2^{2l}}}\left(\mathcal{S}^d\right)\right)}{|\mathcal{S}^d|}}\\
&= \sqrt{\frac{\begin{pmatrix} d+[k]_{\lambda_k\leq 2^{2l}} \\ d\end{pmatrix}}{|\mathcal{S}^d|}}\\
&= \sqrt{\binom{d+[k]_{\lambda_k\leq 2^{2l}}}{d}\frac{\Gamma\left(\frac{d+1}{2}\right)}{2\pi^{(d+1)/2}}}\\
&\leq \left(2^{ld/2}\right)\sqrt{\frac{\Gamma\left(\frac{d+1}{2}\right)}{2\pi^{(d+1)/2}}}\ ,
\end{aligned}$$

which implies $C_1\left(\mathcal{S}^d\right)=\dfrac{1}{\sqrt{|\mathcal{S}^d|}}=\sqrt{\dfrac{\Gamma\left(\frac{d+1}{2}\right)}{2\pi^{(d+1)/2}}}$.

Torus: Let Γ denote an d-dimensional lattice in $\mathbb{R}^d$, ie. Γ is generated by $\{r_1e_1,\ldots,r_d,e_d\}$ over $\mathbb{Z}$, where $\{e_1,\ldots,e_d\}$ is a basis of $\mathbb{R}^d$, for some $r_{1,},\ldots,r_d$. The d-dimensional Torus becomes

$$\mathbb{T}^d=\mathbb{R}^d/\Gamma\ .$$

For $H:\mathbb{R}^d\to\mathbb{R}^d$ such that $H\left(Z^d\right)=\Gamma$, we have $\mathrm{vol}\left(\mathbb{T}^d\right)=\det(H)$. The eigenfunctions of $\mathbb{T}^d$ are those of $\mathbb{R}^d$, which are invariant under deck transformation group $\{\alpha_\gamma(x)=x+\gamma:\gamma\in\Gamma\}$. Let the dual lattice be defined as

$$\Gamma^*=\left\{x\in\mathbb{R}^d:\langle x,\gamma\rangle\in\mathbb{Z},\forall\gamma\in\Gamma\right\}\ .$$

The family of eigenfunctions of $\mathbb{T}^d$ becomes

$$\varphi_y(x)=e^{2\pi i\langle x,y\rangle}\ ,\ y\in\Gamma^*\ ,$$

with eigenvalues $4\pi^2|y|^2$ which form a dense basis on $\mathbb{T}^d$, i.e. $\mathcal{B} = \text{span}\{\varphi_y : y \in \Gamma^*\}$ is dense in $\mathbb{T}^d$. It can be seen that these eigenfunctions are indeed the eigenfunctions of Laplacians on $\mathbb{T}^d$ (Canzani, 2013; Rudnick and Wigman, 2008).

We have that $\mathcal{E}_{n_0}\left(\mathbb{T}^d\right)$ corresponds to the spaces $\mathcal{B}_{n_0}\left(T^d\right)$, which is the span of eigenfunctions with eigenvalues $4\pi^2|y|^2 \leq n_0$. It holds that

$$\sum_i \left|e_i^k(x)\right|^2 = \frac{\dim\left(\mathcal{B}_k\left(\mathbb{T}^d\right)\right)}{|\mathbb{T}^d|} \ .$$

Therefore

$$\begin{aligned}
\|\psi_{l\eta}\|_\infty &= \sup_{x\in\mathbb{T}^d} \sqrt{\alpha_{\eta l}} \sum_{k:\lambda_k<2^{2l}} \sqrt{\gamma\left(\frac{\lambda_k}{2^{2l+1}}\right) - \gamma\left(\frac{\lambda_k}{2^{2l}}\right)} e_k(x)\overline{e_k}(\eta) \\
&\leq \sqrt{\sum_{k:\lambda_k<2^{2l}} \left|e_i^k(x)\right|^2} = \sqrt{\frac{\dim(\mathcal{B}_k(\mathbb{T}^d))}{|\mathbb{T}^d|}} \\
&\leq \sqrt{\frac{2^{ld}}{|\ \mathbb{T}^d\ |}} = \left(2^{ld/2}\right) \sqrt{\frac{1}{2\pi^d r_0^{d-s} R_0^s}} \ ,
\end{aligned}$$

where R_0 is the radius of circles which trace out the larger structure of torus, and r_0 is the radius of circles that form the cross-sections. Hence, $C_1\left(\mathbb{T}^d\right) = \sqrt{\frac{1}{|\mathbb{T}^d|}} = \sqrt{\frac{1}{2\pi^d r_0^{d-s} R_0^s}}$.

Projective Space: Let $h\left(x_1, \ldots, x_d\right)$ denote the coordinates of d-dimensional prohective space $\mathcal{P}^d$. The Laplace-Beltrami operator of $\mathcal{P}^d$ is defined as

$$\Delta = (1 + x\overline{x}^t) \sum_{i,j=1}^{d} (\delta_{ij} + x_i\overline{x}_j) \frac{\partial^2}{\partial\overline{x}_i \partial x_j} \ .$$

It can be shown that the vector space $\mathcal{H}_k$ of λ_k-eigenfunctions has dimension

$$\dim\left(\mathcal{H}_k\right) = \left(\begin{array}{c} d+k \\ k \end{array}\right)^2 - \left(\begin{array}{c} d+k-1 \\ k-1 \end{array}\right)^2 \ ,$$

with eigenvalues $\lambda_k = 4k(d+k)$. Also, it holds that

$$|\ \mathcal{P}^d\ | = \int_{\mathcal{P}^d} dx = \frac{2\pi^d}{(d+1)!} \ .$$

Therefore

$$\begin{aligned}\|\psi_{l\eta}\|_\infty &= \sup_{x\in\mathcal{P}_d}\sqrt{\alpha_{\eta l}}\sum_{k:\lambda_k<2^{2l}}\sqrt{\gamma\left(\frac{\lambda_k}{2^{2l+1}}\right)-\gamma\left(\frac{\lambda_k}{2^{2l}}\right)}e_k(x)\overline{e_k}(\eta)\\ &\le \sqrt{\sum_{k:\lambda_k<2^{2l}}\left|e_i^k(x)\right|^2}=\sqrt{\frac{\dim\left(\mathcal{H}_k\left(\mathcal{P}^d\right)\right)}{|\mathcal{P}^d|}}\\ &= \sqrt{\frac{\left(\begin{array}{c} d+[k]_{\lambda_k\le 2^{2l}} \\ [k]_{\lambda_k\le 2^{2l}}\end{array}\right)^2-\left(\begin{array}{c} d+[k]_{\lambda_k\le 2^{2l}}-1 \\ [k]_{\lambda_k\le 2^{2l}}-1\end{array}\right)^2}{|\mathcal{P}^d|}}\\ &\le \sqrt{\frac{2^{ld}}{|\mathcal{P}^d|}}=\left(2^{ld/2}\right)\sqrt{\frac{(d+1)!}{2\pi^d}}\,,\end{aligned}$$

which implies $C_1\left(\mathcal{P}^d\right)=\sqrt{\dfrac{1}{|\mathcal{P}^d|}}=\sqrt{\dfrac{(d+1)!}{2\pi^d}}$.

This completes the proof.

□

Proof of Theorem 13. The following nonasymptotic sharp bound can be achieved for empirical density function $\widehat{f}_n$ using Bernstein form of McDiarmid's inequality (Habib, 1998; Dubhashi and Panconesi, 2009).

By applying (Geller and Mayeli, 2009), an upper bound for $S_j(x,y)$ can be achieved, as below

$$\begin{aligned}S_jh(x) &= \int_{\mathcal{M}} S_j^2(x,y)f(y)dy\\ &= \int_{\mathcal{M}}\left(\sum_{k:\lambda_k\le 2^{2l}}\sum_{k':\lambda_{k'}\le 2^{2l}}\gamma\left(\frac{\lambda_k}{2^{2l}}\right)e_k(x)\overline{e_k}(y)\gamma\left(\frac{\lambda_{k'}}{2^{2l}}\right)e_{k'}(x)\overline{e_{k'}}(y)\right)dx,\ j\in\mathbb{N}\,,\\ &= \sum_{k:\lambda_k\le 2^{2l}}\gamma^2\left(\frac{\lambda_k}{2^{2l}}\right)e_k(x)\overline{e_k}(x)\le C_2(\mathcal{M})2^{ld/2}\,. \qquad (A.8)\end{aligned}$$

It can be seen that $\widehat{f}_n$ is a function of $X_1, \ldots, X_n$, which satisfies the bounded difference property i.e. there are constants $c_1, \ldots, c_n$ such that for any $1 \leqslant i \leqslant n$, we have

$$\sup_{x_i' \in \mathcal{M}} \mid \widehat{f}_n \left(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n\right) - \widehat{f}_n \left(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n\right) \mid \leqslant c_i \ .$$

To use the Bernstein-type McDiarmid inequality, we need to define variance term for each function argument, as follows

$$B = \max_{i \in [n]} \sup_{x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n} \left| \widehat{f}_n \left(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n\right) - \mathbb{E}_{X_i}[\widehat{f}_n \left(x_1, \ldots, x_{i-1}, X_i, x_{i+1}, \ldots, x_n\right)] \right| \ ,$$

$$V_i = \sup_{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n} \mathbb{E}_{X_i} \Big(\widehat{f}_n \left(x_1, \cdots, x_{i-1}, X_i, x_{i+1}, \ldots, x_n\right) - \mathbb{E}_{X_i}[\widehat{f}_n \left(x_1, \cdots, x_{i-1}, X_i, x_{i+1}, \ldots, x_n\right)] \Big)^2 \ ,$$

$$\sigma^2 = \sum_{i=1}^{n} V_i \ .$$

From the results given in Eq. A.8, we can derive upper bounds for B and σ^2, as follows

$$\begin{aligned}
B &= \max_{i \in [n]} \sup_{x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n} \Big| \widehat{f}_n \left(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n\right) \\
&\qquad\qquad - \mathbb{E}_{X_i}[\widehat{f}_n \left(x_1, \ldots, x_{i-1}, X_i, x_{i+1}, \ldots, x_n\right)] \Big| \ , \\
&= \sup_{x_1, \ldots, x_{i-1}, x_{i+1} \ldots, x_n} \left| \frac{1}{n} \sum_{\substack{k=1 \\ k \neq i}}^{n} S_j \left(x_k, y\right) + \frac{1}{n} S_j \left(X_i, y\right) - \left(\frac{1}{n} \sum_{\substack{k=1 \\ k \neq i}}^{n} S_j \left(x_k, y\right) + \frac{1}{n} \mathbb{E}[S_j \left(X_i, y\right)] \right) \right| \\
&= \left| \frac{1}{n} \Big(S_j \left(X_i, y\right) - \mathbb{E}[S_j \left(X_i, y\right)] \Big) \right| \leq \sup_{x_i' \in \mathcal{M}} \left| S_j \left(x_i', y\right) - \mathbb{E}[S_j \left(X_i, y\right)] \right| \leq \frac{2}{n} \sup_{x_i' \in \mathcal{M}} \| S_j \left(x_i', y\right) \|_2 \\
&\leq \frac{2}{n} C_2(\mathcal{M}) 2^{ld/2} \ .
\end{aligned}$$

$$
\begin{aligned}
V_i &= \sup_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n} \mathbb{E}_{X_i}\Bigg(\widehat{f}_n\left(x_1,\cdots,x_{i-1},X_i,x_{i+1},\ldots,x_n\right) \\
&\qquad\qquad - \mathbb{E}_{X_i}[\widehat{f}_n\left(x_1,\cdots,x_{i-1},X_i,x_{i+1},\ldots,x_n\right)]\Bigg)^2 \\
&= \left|\mathbb{E}_{X_i}\left(\frac{1}{n}\Big(S_j\left(X_i,y\right)-\mathbb{E}_{X_i}[S_j\left(X_i,y\right)]\Big)\right)^2\right| \leq \frac{2}{n^2}\sup_{x_i'\in\mathcal{M}}\left\|S_j\left(x_i',y\right)\right\|_2 \\
&\leq \frac{2}{n^2}C_2(\mathcal{M})2^{ld}\ , \\
\sigma^2 &= \sum_{i=1}^{n} V_i \leq \frac{2}{n}C_2(\mathcal{M})2^{ld}\ .
\end{aligned}
$$

Define the Doob martingale sequence $Z_i = \mathbb{E}\left[f \mid X_1,\ldots,X_i\right]$, such that $Z_0 = \mathbb{E}\left[f\left(X_1,\ldots,X_n\right)\right]$ and $Z_n = f\left(X_1,\ldots,X_n\right)$ (Dubhashi and Panconesi, 2009). Using Eq. A.8, it can be shown that the following difference is bounded

$$
|Z_i - Z_{i-1}| = |\mathbb{E}\left[f \mid x_1,\ldots,x_i\right] - \mathbb{E}\left[f \mid x_1,\ldots,x_{i-1}\right]| \leq c_i'\ ,
$$

where $c_i' = 2C_2(\mathcal{M})2^{\frac{jd}{2}}$. By applying Markov's inequality

$$
P\left(Z_n \geq t\right) \leq \min_{\lambda>0}\frac{\mathbb{E}\left[e^{\lambda Z_n}\right]}{e^{\lambda t}}\ ,
$$

where we set $\lambda = \frac{t}{2\sigma^2}$. Define the martingale difference sequence $D_i = Z_i - Z_{i-1}$, $i \geq 1$. It can be shown that

$$
\begin{aligned}
&\mathbb{E}\left[e^{\lambda Z_n}\right] = \mathbb{E}\left[e^{\lambda(Z_{n-1}+D_n)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[e^{\lambda(Z_{n-1}+D_n)}\right] \mid X_1,\ldots,X_{n-1}\right] \\
&= \mathbb{E}\left[e^{\lambda Z_{n-1}}\mathbb{E}\left[e^{\lambda D_n} \mid X_1,\ldots,X_{n-1}\right]\right]\ .
\end{aligned}
$$

It be shown that

$$
V_i = \sup \mathrm{Var}\left(D_i \mid X_1,\ldots,X_{i-1}\right).
$$

Using the inequality $e^x \leq 1 + x + x^2$ for $|x| \leq 1$, and $\lambda\left|D_i\right| \leq 1$, it can be shown that

$$
\mathbb{E}\left[e^{\lambda D_i} \mid X_1,\ldots,X_{i-1}\right] \leq e^{\lambda^2 V_i}\ .
$$

Therefore

$$
\mathbb{E}\left[e^{\lambda Z_n}\right] \leq \mathbb{E}\left[e^{\lambda Z_{n-1}}\right]e^{\lambda^2 V_n} \leq e^{\lambda^2\sigma^2}\ .
$$

Therefore

$$P\left(Z_n \geq Z_0\right) \leq \frac{e^{\lambda^2\sigma^2}}{e^{\lambda t}} = e^{\frac{t^2}{2\sigma^2}} \ .$$

Using Bernstein's inequality (Maurer, 2019; Dubhashi and Panconesi, 2009), it can be shown that

$$P\left(Z_n \geq Z_0 + t\right) \leq \exp\left(-\frac{t^2}{2\sigma^2\left(1+\frac{Bt}{3V}\right)}\right) \ .$$

Substituting $Z_0 = \mathbb{E}\left[f\left(X_1, \ldots, X_n\right)\right]$ and $Z_n = f\left(X_1, \ldots, X_n\right)$ gives

$$P\left(\sup_{y\in\mathcal{M}}\left|\widehat{f}_n(j,y) - \mathbb{E}\left[\widehat{f}_n(j,y)\right]\right| \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma^2\left(1+\frac{Bt}{3\sigma^2}\right)}\right).$$

The approximation bound due to bias, i.e. $\left\|\mathbb{E}[\widehat{f}_n(j,y)] - f(j,y)\right\|_\infty$ can be controlled by $C2^{-j\kappa}$ assuming that f is a κ-regular function, i.e. $f \in C^\kappa(\mathcal{M})$ for $\kappa \in \mathbb{N}$ (Kerkyacharian et al., 2012).

As a result, we get

$$P\left(\sup_{y\in\mathcal{M}}\left|\widehat{f}_n(j,y) - f(j,y)\right| \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma^2\left(1+\frac{Bt}{3\sigma^2}\right)}\right).$$

By using Lebesgue's dominated convergence theorem, following result holds for the convergence of empirical probability measure P_n to theoretical probability measure P

$$P\left(\sup_{\Omega\subseteq\mathcal{M}}\left|P_n(\Omega) - P(\Omega)\right| \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma^2\left(1+\frac{Bt}{3\sigma^2}\right)}\right).$$

Given that the depth function $D(P,y)$ is a continuous function of probability measure P, by continuous mapping theorem and Assumption 3, we get the following result for the asymptotic convergence of empirical depth function

$$P\left(\sup_{y\in\mathcal{M}}\left|D(P_n,y) - D(P,y)\right| \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma^2\left(1+\frac{Bt}{3\sigma^2}\right)}\right).$$

To determine the constant $C_2\left(\mathcal{M}\right)$ on sphere and torus manifolds, we can use the following result.

Sphere: Following steps similar to those in Theorem 12, it can be shown that

$$\begin{aligned}
\int_{\mathcal{S}_d} S_j^2(x,y)dx &= \sum_{k:\lambda_k \leq 2^{2l}} \gamma^2\left(\frac{\lambda_k}{2^{2l}}\right) e_k(x)\overline{e_k}(x) \\
&= \sum_{k:\lambda_k \leq 2^{2l}} \gamma^2\left(\frac{\lambda_k}{2^{2l}}\right) \frac{\dim\left(\mathcal{H}_k\left(\mathcal{S}^d\right)\right)}{|\mathcal{S}^d|} \\
&= \frac{\dim\left(P_{[k]_{\lambda_k \leq 2^{2l}}}\left(\mathcal{S}^d\right)\right)}{|\mathcal{S}^d|} \\
&= \frac{\begin{pmatrix} d+[k]_{\lambda_k \leq 2^{2l}} \\ d \end{pmatrix}}{|\mathcal{S}^d|} \\
&= \binom{d+[k]_{\lambda_k \leq 2^{2l}}}{d} \frac{\Gamma\left(\frac{d+1}{2}\right)}{2\pi^{(d+1)/2}} \\
&\leq \left(2^{ld}\right) \frac{\Gamma\left(\frac{d+1}{2}\right)}{2\pi^{(d+1)/2}} \ ,
\end{aligned}$$

which implies $C_2\left(\mathcal{S}^d\right) = \dfrac{1}{|\mathcal{S}^d|} = \dfrac{\Gamma\left(\frac{d+1}{2}\right)}{2\pi^{(d+1)/2}}$.

Torus: By following a similar argument as in Theorem 12, it can be seen that

$$\begin{aligned}
\int S_j^2(x,y)dx &= \sum_{k:\lambda_k \leq 2^{2l}} \gamma^2\left(\frac{\lambda_k}{2^{2l}}\right) e_k(x)\overline{e_k}(x) \\
&= \sum_{k:\lambda_k \leq 2^{2l}} \gamma^2\left(\frac{\lambda_k}{2^{2l}}\right) \frac{\dim(\mathcal{B}_k(\mathbb{T}^d))}{|\ \mathbb{T}^d\ |} \\
&\leq \frac{2^{ld}}{|\ \mathbb{T}^d\ |} = \frac{2^{ld}}{2\pi^d r_0^{d-s} R_0^s} \ ,
\end{aligned}$$

where R_0 is the radius of circles which trace out the larger structure of torus, and r_0 is the radius of circles that form the cross-sections. Therefore $C_2\left(\mathbb{T}^d\right) = \dfrac{1}{|\mathbb{T}^d|} = \dfrac{1}{2\pi^d r_0^{d-s} R_0^s}$.

Projective Space: Following steps similar to those in Theorem 12, it can be shown that

$$\begin{aligned}
\int_{\mathcal{P}_d} S_j^2(x,y)dx &= \sum_{k:\lambda_k \le 2^{2l}} \gamma^2\left(\frac{\lambda_k}{2^{2l}}\right) e_k(x)\overline{e_k}(x) \\
&= \sum_{k:\lambda_k \le 2^{2l}} \gamma^2\left(\frac{\lambda_k}{2^{2l}}\right) \frac{\dim\left(\mathcal{H}_k\left(\mathcal{P}^d\right)\right)}{|\mathcal{P}^d|} \\
&= \frac{\left(\begin{array}{c} d+[k]_{\lambda_k \le 2^{2l}} \\ [k]_{\lambda_k \le 2^{2l}} \end{array}\right)^2 - \left(\begin{array}{c} d+[k]_{\lambda_k \le 2^{2l}}-1 \\ [k]_{\lambda_k \le 2^{2l}}-1 \end{array}\right)^2}{|\mathcal{P}^d|} \\
&\le \frac{2^{ld}}{|\mathcal{P}^d|} = \left(2^{ld}\right)\frac{(d+1)!}{2\pi^d}\,,
\end{aligned}$$

which implies $C_2\left(\mathcal{P}^d\right) = \dfrac{1}{|\mathcal{P}^d|} = \dfrac{(d+1)!}{2\pi^d}$.

This completes the proof.

□

A.2.2 Single Change point Detection

Proof of Theorem 14. It can be seen that under H_0, the process $\{Z_n(u),\ 0 \le u \le 1\}$ converges weakly in $\mathcal{D}[0,1]$ endowed with the Skorokhod topology (see Chapter 3 of Billingsley (1968)) to a standard Brownian bridge

$$\{B(u),\ 0 \le u \le 1\}\,.$$

To see this, define

$$U_i = \frac{\widehat{R}_i - (n+1)/2}{\sqrt{n\,(n^2-1)/12}}, \qquad i = 1, \ldots, n\,.$$

Evidently $Z_n(u) = \sum_{i=1}^{\lfloor nu \rfloor} U_i$. Under H_0, the random variables $U_1, \ldots, U_n$ are exchangeable, and $\sum_{i=1}^{n} U_i = 0$ while also $\sum_{i=1}^{n} U_i^2 = 1$. In addition,

$$\max_{1 \le i \le n} |\,U_i\,| = \frac{n-(n+1)/2}{\sqrt{n\,(n^2-1)/12}} = \frac{n-1}{\sqrt{n(n^2-1)/3}} \to 0\,, \quad \text{as} \quad n \to \infty\,.$$

This shows that the conditions of Theorem 24.1 of Billingsley (Billingsley, 1968) are satisfied for the process $Z_n(u) = \sum_{i=1}^{\lfloor nu \rfloor} U_i$, from which the result follows. As a consequence of the continuous mapping theorem, we get the result:

$$T_n \xrightarrow{\mathcal{D}} \sup_{0 \le u \le 1} |B(u)|.$$

□

Proof of Theorem 15. Throughout the proof we let c_i, $i \geq 0$, denote unimportant numerical constants. Clearly $T_n \geq \mid Z_n(\theta) \mid$, and so the theorem follows if $\mid Z_n(\theta) \mid \xrightarrow{P} \infty$, as $n \to \infty$. For $0 \leq u \leq 1$, let

$$\widetilde{Z}_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nu \rfloor} \frac{R_i - (n+1)/2}{\sqrt{(n^2-1)/12}},$$

where

$$R_i = \#\{\mathbf{X}_j \ ; \ D(P_* \, ; \, \mathbf{X}_j) \leq D(P_* \, ; \, \mathbf{X}_i), \ j = 1, \ldots n\}\,. \tag{A.9}$$

Evidently

$$Z_n(\theta) = G_n + \widetilde{Z}_n(\theta), \tag{A.10}$$

with $G_n = Z_n(\theta) - \widetilde{Z}_n(\theta)$. By the triangle inequality

$$\mid G_n \mid \leq \frac{1}{\sqrt{n}} \left(\frac{n^2-1}{12} \right)^{-1/2} \sum_{i=1}^{\lfloor n\theta \rfloor} \mid \widehat{R}_i - R_i \mid \leq c_0 \, n^{-3/2} \sum_{i=1}^{\lfloor n\theta \rfloor} \mid \widehat{R}_i - R_i \mid .$$

Let

$$\begin{aligned} A_{i,\,j} &= \{D(P_*,\, \mathbf{X}_j) \leq D(P_*,\, \mathbf{X}_i)\} \cap \{D(P_n,\, \mathbf{X}_j) > D(P_n,\, \mathbf{X}_i)\}\,, \\ B_{i,\,j} &= \{D(P_*,\, \mathbf{X}_j) > D(P_*,\, \mathbf{X}_i)\} \cap \{D(P_n,\, \mathbf{X}_j) \leq D(P_n,\, \mathbf{X}_i)\}\,. \end{aligned}$$

Then clearly

$$\mid \widehat{R}_i - R_i \mid \leq \sum_{j=1}^{n} \mathbf{1}(A_{i,j}) + \sum_{j=1}^{n} \mathbf{1}(B_{i,j}), \tag{A.11}$$

where $\mathbf{1}(A)$ denotes the indicator of the event A. It follows according to the definition of $A_{i,j}$ that if $\gamma_n = \max_{1 \le j \le n} |\, D(P_*, \mathbf{X}_j) - D(P_n, \mathbf{X}_j)\,|$, then

$$A_{i,\,j} \subset \{|\, D(P_*, \mathbf{X}_i) - D(P_*, \mathbf{X}_j)\,| \le 2\gamma_n\}\,.$$

Moreover, by the definitions of $D(P_*, \mathbf{X}_i)$ and $D(P_n, \mathbf{X}_i)$, it follows that

$$\gamma_n \le \sup_{\mathbf{x} \in \mathbb{R}^p} |\, D(P_n, \mathbf{x}) - D(P_*, \mathbf{x})\,|\,,$$

and hence

$$\mathbf{1}(A_{i,j}) \le \mathbf{1}(\{|\, D(P_*, \mathbf{X}_i) - D(P_*, \mathbf{X}_j)\,| \le 2 \sup_{\mathbf{x} \in \mathbb{R}^p} |\, D(P_n, \mathbf{x}) - D(P_*, \mathbf{x})\,|\}). \tag{A.12}$$

According to Assumption 9, and since $D(P_*, \mathbf{X}_i)$ and $D(P_*, \mathbf{X}_j)$ are independent for $i \neq j$, the random variable $|\, D(P_*, \mathbf{X}_i) - D(P_*, \mathbf{X}_j)\,|$ has a Lipschitz distribution function for $i \neq j$, from which it follows that for any $1 \le i, j \le n$,

$$\begin{aligned}
&\mathrm{E}[\mathbf{1}(\{|\, D(P_*, \mathbf{X}_i) - D(P_*, \mathbf{X}_j)\,| \le 2 \sup_{\mathbf{x} \in \mathbb{R}^p} |\, D(P_n, \mathbf{x}) - D(P_*, \mathbf{x})\,|\})] \\
&\quad = \mathrm{E}[\mathrm{E}[\mathbf{1}(\{|\, D(P_*, \mathbf{X}_i) - D(P_*, \mathbf{X}_j)\,| \\
&\quad \le 2 \sup_{\mathbf{x} \in \mathbb{R}^p} |\, D(P_n, \mathbf{x}) - D(P_*, \mathbf{x})|\})| \sup_{\mathbf{x} \in \mathbb{R}^p} |\, D(P_n, \mathbf{x}) - D(P_*, \mathbf{x})\,|]] \\
&\quad \le c_1 \mathrm{E}[\sup_{\mathbf{x} \in \mathbb{R}^p} |\, D(P_n, \mathbf{x}) - D(P_*, \mathbf{x})\,|] = O(n^{-1/2}),
\end{aligned}$$

where the last line can be derived from Dvoretzky-Keifer-Wolfowitz inequality.

This combined with (A.12) gives that

$$\mathrm{E}\left(\sum_{j=1}^{n} 1(A_{i,j})\right) = O(n^{1/2}).$$

The same arguments imply that the above bound also holds for the sum of the $B_{i,j}$ terms in (A.11), which gives that

$$\mathrm{E}\,|\,\widehat{R}_i - R_i\,| = O(n^{1/2}). \tag{A.13}$$

This along with (A.11) implies that $\mathrm{E}|G_n| = O(1)$, and hence $G_n = O_P(1)$ by Markov's inequality. The random variables $D(P_*, \mathbf{X}_i)$, $1 \le i \le n$ satisfy the conditions of the

Chernoff and Savage Theorem (Sidak et al., 1999), from which it follows that $\widetilde{Z}_n(\theta)$ is asymptotically Gaussian with bounded variance and asymptotic mean

$$\sqrt{n}\theta\left[\int_{-\infty}^{\infty}\theta H_1(u)+(1-\theta)H_2(u)dH_1(u)-\frac{1}{2}\right],$$

which tends in absolute value to positive infinity under (3.1). This implies that $|\,\widetilde{Z}_n(\theta)\,|\xrightarrow{P}\infty$, which, along with (A.10) and the fact that $G_n=O_P(1)$, proves the result. □

Proof of Theorem 16. Let

$$Z_k=Z_{_n}(k)=\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\frac{\widehat{R}_i-(n+1)/2}{\sqrt{(n^2-1)/12}},$$

$$\widetilde{Z}_k=\widetilde{Z}_{_n}(k)=\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\frac{R_i-(n+1)/2}{\sqrt{(n^2-1)/12}},$$

$$G_k=Z_k-\widetilde{Z}_k=\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\frac{R_i-\widehat{R}_i}{\sqrt{(n^2-1)/12}}.$$

Under these definitions, it is clear that according to the definition of $\widehat{k}^*$,

$$\begin{aligned}\widehat{k}^*&=\min\{k\ :\ |\,Z_k\,|=\max_{1\le i\le n}|\,Z_i\,|\}\\&=\min\{k\ :\ |\,Z_k\,|^2=\max_{1\le i\le n}|\,Z_i\,|^2\}\\&=\min\{k\ :\ |\,Z_k\,|^2-|\widetilde{Z}_{k^*}\,|^2=\max_{1\le i\le n}|\,Z_i\,|^2-|\widetilde{Z}_{k^*}|^2\}.\end{aligned}$$

From the definition of $\widehat{\theta}$, the statement of Theorem 16 is equivalent with the existence of a real valued sequence $a_{_n}$ satisfying $a_{_n}\to\infty$, $a_{_n}/n\to 0$, and

$$\Pr(|\,\widehat{k}^*-k^*\,|>a_{_n})=\Pr(\widehat{k}^*<k^*-a_{_n})+\Pr(\widehat{k}^*>k^*+a_{_n})\to 0,\quad n\to\infty. \tag{A.14}$$

We now show that for the sequence $a_{_n}=n^{\kappa}$, $1/2<\kappa<1$,

$$\Pr(\widehat{k}^*<k^*-a_{_n})\to 0\,,\quad as\quad n\to\infty\,.$$

The second term on the right hand side of (A.14) can be handled in a similar way, and so the details are omitted. Evidently

$$\Pr(\widehat{k}^*<k^*-a_{_n})\le\Pr\left(\max_{1\le i\le k^*-2a_{_n}}|\,Z_i\,|^2-|\widetilde{Z}_{k^*}|^2>\max_{k^*-a_{_n}\le i\le k^*}|\,Z_i\,|^2-|\widetilde{Z}_{k^*}|^2\right). \tag{A.15}$$

Since $Z_i = \widetilde{Z}_i + G_i$, we have by some simple algebra that for $1 \leq i \leq k^*$,

$$| Z_i |^2 - | \widetilde{Z}_{k^*} |^2 = A_i + B_i + C_i + G_i^2 + 2\widetilde{Z}_i G_i,$$

where

$$A_i = \left(\widetilde{Z}_i - \frac{i}{k^*} \widetilde{Z}_{k^*} \right)^2, \quad B_i = 2 \left(\widetilde{Z}_i - \frac{i}{k^*} \widetilde{Z}_{k^*} \right) \frac{i}{k^*} \widetilde{Z}_{k^*},$$

and

$$C_i = \widetilde{Z}_{k^*}^2 \left(\left(\frac{i}{k^*} \right)^2 - 1 \right).$$

It follows under the conditions of Theorem 15 and the Chernoff-Savage Theorem (Chernoff et al., 1957) that there exists a nonzero constants c_0 and $c_1 > 0$ so that $\widetilde{Z}_{k^*}^2 = n\, c_0^2\, (1 + o_P(1))$, and hence for each $i \leq k^*$,

$$C_i = c_0^2\, n \left(\frac{i^2 - (k^*)^2}{(k^*)^2} \right) (1 + o_P(1)) = c_1 \left(\frac{i^2 - (k^*)^2}{k^*} \right) (1 + o_P(1)). \tag{A.16}$$

Moreover, it follows by the definition of C_i that for all n, $C_i < C_j$ for all $i < j$, and by the above property and the fact that $k^* = \lfloor n\theta \rfloor$, if $a_n = n^\kappa$ for $1/2 < \kappa < 1$, we get with a little algebra that $C_{k^* - a_n} - C_{k^* - 2\, a_n} = 2\, a_n (1 + o_P(a_n)) \xrightarrow{P} \infty$, as $n \to \infty$. We now aim to show the following four results:

$$\max_{1 \leq k \leq k^* - 2a_n} \frac{A_k}{| C_k |} = o_P(1), \tag{A.17}$$

$$\max_{1 \leq k \leq k^* - 2a_n} \frac{B_k}{| C_k |} = o_P(1), \tag{A.18}$$

$$\max_{1 \leq k \leq k^* - 2a_n} \frac{G_k^2}{| C_k |} = o_P(1), \tag{A.19}$$

$$\max_{1 \leq k \leq k^* - 2a_n} \frac{\widetilde{Z}_k G_k}{| C_k |} = o_P(1). \tag{A.20}$$

Towards establishing (A.17), we note that for all $k \leq k^*$,

$$\left(\frac{k^*}{(k^*)^2 - k^2} \right) \leq \frac{k^*}{k(k^* - k)}.$$

Since the ranks R_k, $k < k^*$, are the ranks of independent and identically distributed $D(P, \mathbf{X}_1), \ldots, D(P, \mathbf{X}_k)$, and $\widetilde{Z}_k$ is the partial sum of these ranks, it follows from equation

(2.22) in Gombay and Hušková (1998) that for all $\varepsilon > 0$,

$$\begin{aligned}
&\Pr\left(\max_{1\le k\le k^*-2a_n} \frac{k^*}{k(k^*-k)} A_k < \varepsilon \right) \\
&= \Pr\left(\max_{1\le k\le k^*-2a_n} \frac{k^*}{k(k^*-k)} \left(\sum_{i=1}^{k} \frac{R_i-(n+1)/2}{\sqrt{(n^2-1)/12}} - \frac{k}{k^*} \sum_{i=1}^{k^*} \frac{R_i-(n+1)/2}{\sqrt{(n^2-1)/12)}} \right)^2 < \varepsilon n \right) \\
&\quad \longrightarrow 1 \text{ as } n \to \infty,
\end{aligned}$$

proving (A.17). (A.18) can be established using a similar argument as that used to establish (2.18) in Gombay and Hušková (1998), so the details are omitted. In order to show (A.19), we note that according to the definition of G_k and the triangle inequality that

$$\begin{aligned}
\max_{1\le k\le k^*-2a_n} |\, G_k \,| &= \max_{1\le k\le k^*-2a_n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{k} \frac{R_i-\widehat{R}_i}{\sqrt{(n^2-1)/12}} \right| \\
&\le \max_{1\le k\le k^*-2a_n} \frac{1}{\sqrt{n}} \sum_{i=1}^{k} \left| \frac{R_i-\widehat{R}_i}{\sqrt{(n^2-1)/12}} \right| \\
&\le \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| \frac{R_i-\widehat{R}_i}{\sqrt{(n^2-1)/12}} \right| = O_P(1),
\end{aligned} \tag{A.21}$$

where the last line follows from (A.13) and Chebyshev's inequality.

Evidently then $\max_{1\le k\le k^*-2\,a_n} |\, G_k \,|^2 = O_P(1)$, and also for some positive constant c_2 and for $k < k^*$,

$$\left| c_0^2\, n \left(\frac{k^2-(k^*)^2}{(k^*)^2} \right) \right| \ge c_2\, n,$$

from which with (A.16), (A.19) now follows. (A.20) follows similarly. In addition to (A.17)–(A.20), we also have that

$$\max_{k^*-a_n\le k\le k^*} A_k = O_P(1), \tag{A.22}$$

$$\max_{k^*-a_n\le k\le k^*} B_k = O_P(1), \tag{A.23}$$

$$\max_{k^*-a_n\le k\le k^*} G_k^2 = O_P(1), \tag{A.24}$$

$$\max_{k^*-a_n\le k\le k^*} \widetilde{Z}_k\, G_k = o_P(a_n). \tag{A.25}$$

(A.22) and (A.23) follow from Lemma 2.1 in Gombay and Hušková (1998). Equation (A.19) follows from (A.21), and the crude approximation that, with some positive constant c_5,

$$\begin{aligned}
\max_{1\le k\le n} |\,\widetilde{Z}_k\,| &= \max_{1\le k\le n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\frac{R_i-(n+1)/2}{\sqrt{(n^2-1)/12}}\right| \\
&\le \max_{1\le k\le n}\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\left|\frac{R_i-(n+1)/2}{\sqrt{(n^2-1)/12}}\right| \\
&\le \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|\frac{R_i-(n+1)/2}{\sqrt{(n^2-1)/12}}\right| \\
&\le c_5\,\sqrt{n},
\end{aligned}$$

where in the last line we used that $|\,R_i-(n+1)/2\,|\le n$. Now combining the above results, we have for all $\varepsilon>0$ and n sufficiently large,

$$\begin{aligned}
&\Pr\left(\max_{1\le k\le k^*-2\,a_n}|\,Z_k\,|^2-|\,\widetilde{Z}_{k^*}\,|^2>\max_{k^*-a_n\le k\le k^*}|\,Z_k\,|^2-|\,\widetilde{Z}_{k^*}\,|^2\right)\\
&\qquad\le \Pr\left(\max_{1\le k\le k^*-2\,a_n} C_k\left(1+\frac{A_k}{|\,C_k\,|}+\frac{B_k}{|\,C_k\,|}+\frac{G_k^2}{|\,C_k\,|}+\frac{\widetilde{Z}_k\,G_k}{|\,C_k\,|}\right)\right.\\
&\qquad\qquad\left.>\max_{k^*-a_n\le k\le k^*}A_k+B_k+C_k+G_k^2+2\widetilde{Z}_k\,G_k\right)\\
&\qquad\le \Pr(C_{k^*-a_n}\Lambda_n-C_{k^*-2\,a_n}<\max_{k^*-a_n\le k\le k^*}A_k+B_k+G_k^2+2\widetilde{Z}_kG_k)\\
&\qquad\to 0,\quad as\quad n\to\infty,
\end{aligned}$$

since $\Lambda_n=\max_{1\le k\le k^*-2\,a_n}(1+\frac{A_k}{|\,C_k\,|}+\frac{B_k}{|\,C_k\,|}+\frac{G_k^2}{|\,C_k\,|})\xrightarrow{P}1$ as $n\to\infty$ by (A.17)–(A.20), which gives the result. □

Proof of Theorem 17. The proof results from the consistency of the estimated change point, namely $|\widehat{k}-k|=o_p(a_n)$. For $k<\widehat{k}$, the following holds

$$a_n^{-1}\Big(Z_n(\widehat{\theta})-Z_n(\theta)\Big)=\frac{a_n^{-1}}{\sqrt{n}}\Big(\sum_{i=k+1}^{\widehat{k}}e_i\Big)\le a_n^{-1}\frac{|\widehat{k}-k|}{\sqrt{n}}O_p(1)=o_p(1)\ ,$$

The same is also true for $k>\widehat{k}$. This completes the proof. □

Lemma 5. *Consider an i.i.d. sample of size n, $\mathbf{X}_1, \ldots, \mathbf{X}_n$ sampled from manifold $\mathcal{M}$, with change at location k^*. For $1 \leq m \leq k^*$, and $\delta > 0$, the partial sum $S_m = \sum_{i=1}^{m} R_i$ satisfies*

$$Pr\left(|S_m - \mathbb{E}[S_m]| > \delta\right) \leq 4 \exp\left(-2\min(m, k^* - m, n - k^*)\left(\frac{\delta}{2\max(k^* - m, n - k^*)m}\right)^2\right) .$$

Proof of Lemma 5. The proof is similar to the proof of Lemma 8 in (Ramsay, 2022). □

Proof of Theorem 18. let $p = \Pr(D(P_1, X_1) < D(P_2, X_n))$. Using Hoeffding (1963), it can be shown that

$$\begin{aligned}\sum_{i=1}^{m} R_i &= \frac{m(m+1)}{2} + \sum_{j=1}^{m}\sum_{i=m+1}^{k^*} \mathbb{I}\left(D(P_*, X_j) \leq D(P_*, X_i)\right) \\ &\quad + \sum_{j=1}^{m}\sum_{i=k^*+1}^{n} \mathbb{I}\left(D(P_*, X_j) \leq D(P_*, X_i)\right) \\ &= \frac{m(m+1)}{2} + (k^* - m)m\mathrm{U}_{1,n} + m(n - k^*)\mathrm{U}_{2,n} ,\end{aligned}$$

where $\mathrm{U}_{1,n}$ and $\mathrm{U}_{2,n}$ are U-statistics of order 2 with $\mathbb{E}[\mathrm{U}_{1,n}] = \frac{1}{2}$ and $\mathbb{E}[\mathrm{U}_{2,n}] = p$.

For $Z_n(k^*/n)$, we can write

$$\sqrt{n}Z_n(k^*/n) = \frac{k^*(k^*+1)}{2\sigma_n} + \frac{k^*(n-k^*)}{\sigma_n}\mathrm{U}' - \frac{k^*(n+1)}{2\sigma_n} ,$$

where U' is a two-sample U-statistic, and

$$\sqrt{n}\mathbb{E}[Z_n(k^*/n)] = \frac{k^*(k^*+1)}{2\sigma_n} + \frac{k^*(n-k^*)p}{\sigma_n} - \frac{k^*(n+1)}{2\sigma_n} = k^*(n-k^*)\frac{p - 1/2}{\sigma_n} .$$

Using Hoeffding (1963) and Pitcan (2017), we have

$$\begin{aligned}\Pr\left(Z_n(k^*/n) < a_n\right) &= \Pr\left(Z_n(k^*1/n) - \mathbb{E}\left[Z_n(k^*/n)\right] < a_n - k^*(n-k^*)\frac{p-1/2}{\sigma_n}\right) \\ &\leq \exp\left(-2n^2\min(\theta, 1-\theta)\left(p - \frac{1}{2} - \frac{a_n\sigma_n}{n^2\theta(1-\theta)}\right)^2\right) ,\end{aligned}$$

where $0 < \theta < 1$ is the breaking fraction, i.e. $k^* = n\theta$.

For $t > 0$, it holds that

$$\begin{aligned}
\Pr\left(\left|\widehat{k}^* - k^*\right| > tn\right) &= \Pr\left(\widehat{k}^* - k^* > tn\right) + \Pr\left(k^* - \widehat{k}^* > tn\right) \\
&\leq \Pr\left(\max_{c_0 n \leq i \leq k^* - tn} |Z_n(i/n)| - Z(k^*/n) > 0\right) \\
&\leq \Pr\left(\max_{c_0 n \leq i \leq k^* - tn} Z_n(i/n) - Z_n(k^*/n) > 0\right) \\
&\quad + \Pr\left(\max_{c_0 n \leq i \leq k^* - tn} Z_n(i/n) + Z_n(k^*/n) < 0\right) \ .
\end{aligned}$$

We have that

$$\max_{c_0 n \leq i \leq k^* - tn} Z_n(i/n) - Z_n(k^*/n) = \max_{c_0 n \leq i \leq k^* - tn} -\sum_{j=i+1}^{k^*} \frac{R_j - \mu_n}{\sigma_n} = \max_{tn \leq i \leq k^* - c_0 n} -Z_n(i/n) \ ,$$

where the last line follows from the exchangability of $R_1, \ldots, R_{k^*}$.

For $tn < i < k^* - c_0 n$, we have that

$$\sqrt{n}\mathbb{E}[Z_n(i/n)] = \frac{i}{\sigma_n}(n - k^*)(p - 1/2) \geq \frac{tn}{\sigma_n}(n - k^*)(p - 1/2) \ .$$

By Lemma 5, we have

$$\Pr\left(\max_{tn \leq i \leq k^* - c_0 n} -Z_n(i/n) > 0\right) \leq c\theta^2 \exp\left(-c_2 n t^2 (p - 1/2)^2\right) \ .$$

Similar results can be obtained for

$$\Pr\left(\max_{c_0 n \leq i \leq k^* - tn} Z_n(i/n) + Z_n(k^*/n) < 0\right) \ .$$

Therefore

$$\Pr\left(\widehat{k}^* - k^* > tn\right) \leq c\theta^2 \exp\left(-c_2 n t^2 (p - 1/2)^2\right) \ .$$

By following the same argument for $\Pr\left(k^* - \widehat{k}^* > tn\right)$, we get

$$\Pr\left(\left|\widehat{k}^* - k^*\right| > tn\right) \leq c_1 \exp\left(-c_2 n t^2 (p - 1/2)^2\right) \ .$$

This completes the proof.

□

A.2.3 Multiple Change point Detection

To establish theoretical results for the proposed multiple change point detection method, we need to define the following

$$\widehat{R}_k^{s,e} := \#\{y_j \mid y_j \leq y_k, j = s, ..., e\} \ ,$$

and

$$\widehat{Z}_{e,s}(k) := \frac{1}{n_{s,e}} \sum_{i=s}^{e} \frac{\widehat{R}_i^{s,e} - (n_{s,e}+1)/2}{\sqrt{(n_k^{s,e})/12}} \ ,$$

$$Z_{e,s}(k) := \frac{1}{n_{s,e}} \sum_{i=s}^{e} \frac{R_i^{s,e} - (n_{s,e}+1)/2}{\sqrt{(n_k^{s,e})/12}} \ ,$$

where $n_{s,e} = e - s + 1$, s is the starting point of the segment and e is the end point of the segment, $\widehat{R}_k^{s,e}$ is the estimated rank based on observations in segment $[s, e]$ and R is the corresponding theoretical ranks.

Moreover, we need the following assumptions and lemmas to finish the proofs.

Assumption 16. *It is assumed that the number of change points ℓ is fixed, and the locations of the change points are well spread throughout the dataset. More precisely, we assume that the distance between any two adjacent change points satisfies $\delta_n = O(n)$.*

Assumption 17. *Let $k_i^* = k_i^*(n)$ denote the true location of i^{th} change point within the dataset, such that $k_i^*/n \to \theta_i \in (0, 1)$, as $n \to \infty$, for $i = 1, 2, \ldots, \ell$. Also, let $k_0^* = 1$ and $k_{\ell+1}^* = n$.*

Assumption 18. *It is assumed that the critical value $c_{\alpha_i} = o(\sqrt{n})$, for $i = 1, 2, ..., \ell$ and obeys the condition $\lim_{\ell \to \infty} c_{\alpha_i} = \infty$.*

Remark 1. *Assumption 16 is a reasonable assumption for many cases, including scenarios where the number of shifts is fixed and finite. Also, the assumption that $k_i^* - k_{i-1}^* = O(n)$ for all $i = 1, \ldots, \ell$, is reasonable, as for breaking fractions $\theta_1, \theta_2, \ldots, \theta_\ell \in (0, 1)$, and the sample of size n, it is expected to have $\lfloor n\theta_1 \rfloor$ of observations fall within the interval $[0, k_1^*)$, $\lfloor n\theta_i \rfloor - \lfloor n\theta_{i-1} \rfloor$ fall in the interval $[k_{i-1}^*, k_i^*)$ for $i = 2, \ldots, \ell - 1$, and $n - \lfloor n\theta_\ell \rfloor$ fall in the interval $[k^* t_\ell, n]$. Given this, Assumption 17 is also very natural. Assumption 18 is a technical assumption on the divergence speed of critical value due to Bonferroni correction. Under the hypothesis of no change, $c_{\alpha_i} = o(\sqrt{n})$ is a reasonable assumption. For numerical simulation, the critical value is kept fixed after some point (see for example Galeano and Wied (2017)), which automatically satisfies this assumption.*

The results pertinent to the consistency of the proposed multiple change-points detection stem from the classical work of Chernoff and Savage (1958) on two samples rank tests. In Lemma 6, we present a modified version of this two-sample test statistic. In Lemma 7, the relation between the proposed test statistic and this modified two-sample test statistic is given. In Lemma 8, the change point detection power of the proposed test statistic is shown when there is at least one change point in the segment. Lemmas 9 to 12 are devoted to proving the consistent estimation of the multiple change points for the proposed test statistic. The abovementioned lemmas will be used to finish the proofs of Theorem 19 and Theorem 21.

Lemma 6. *Let the Assumptions 16 to 18 hold. For a given true change point location $k = k_i^*$, $i \in \{1, 2, ..., \ell\}$, and fixed s, $e \in \{k_{i'}^* \mid i' \neq i,\ i' = 0,\ 1,\ 2,\ \ldots,\ \ell+1\}$ with $s < k < e$, consider the two-sample test statistic proposed by Chernoff and Savage (1958)*

$$T_{s,\,e}(k) := \frac{1}{k-s} \sum_{j=s}^{n_{s,\,e}} J\left(\frac{j-s+1}{n_{s,\,e}}\right) \zeta_{j,n_{s,\,e}}(k) = \int_{\mathcal{M}} J(\widehat{H}_{s,\,e}(x))\, d\widehat{\nu}_k(x)\,.$$

where $\widehat{\nu}_k$ is the empirical distribution measure of observation in the segment $[s,\, k-1]$, x is an open ball in manifold $\mathcal{M}$, and $\widehat{H}_{s,\,e}$ is the combined empirical distribution measure of observations in the segment $[s,\, e]$, $J(\cdot)$ is a score function satisfying the four assumptions of Theorem 1 of Chernoff and Savage (1958), and $\zeta_{j,n_{s,\,e}}(k)$ is a function of j^{th} smallest sample after ordering the data, defined as

$$\zeta_{j,n_{s,\,e}}(k) := \begin{cases} 1, & \text{if} \quad j \in \{s,\,\ldots,\,k-1\} \\ 0, & \text{if} \quad j \in \{k,\ldots,\,e\} \end{cases},$$

Then $\mu_{n_{s,\,e}}(k) = \mathrm{E}\left[T_{s,\,e}(k) \mid s,\, e\right] = O(1)$.

Lemma 7. *For any fixed $k \in [s,\, e]$, it holds that $Z_{s,\,e}(k) = \sqrt{n_{s,\,e}}\left(\dfrac{k-s}{n_{s,\,e}}\right) T_{s,\,e}(k)$.*

Lemma 8. *Under the assumptions made, and when there is at-least one change point in the segment, for any sequence $a_n = o(\sqrt{n_{s,\,e}})$, it holds that*

$$\lim_{n\to\infty} \Pr\left(\left|\frac{Z_{s,\,e}(k_i^*)}{a_n}\right| > \Delta \,\middle|\, s,\, e\right) \to 1, \qquad \textit{for all } \Delta > 0\,,$$

where $s,\, e \in \{k_{i'}^* \mid i' \neq i,\ i' = 0,\ 1,\ 2,\ \ldots,\ \ell+1\}$.

Lemma 9. *Let the Assumptions 16 to 18 hold. Recall $Z_{s,e}(k)$ and $\widetilde{Z}_{s,e}(k)$ defined as above, and fix s, $e \in \{k^*_{i'} \mid i' \neq i,\ i' = 0,\ 1,\ 2,\ \ldots,\ \ell+1\}$. Define*

$$A_n := \max_{s<k<e} \mid \widetilde{Z}_{s,e}(k) - Z_{s,e}(k) \mid \ .$$

Then $A_n = O_p(1)$.

Lemma 10. *Let the Assumptions 16 to 18 hold, and $a_n = o(\sqrt{n_{s,e}})$. Then*

$$\lim_{n\to\infty} \Pr\left(\left| \frac{\widetilde{Z}_{s,e}(k^*_i)}{a_n} \right| > \Delta \mid s,\ e \right) \to 1, \quad \textit{for all } \Delta > 0\,,$$

where s, $e \in \{k^*_{i'} \mid i' \neq i,\ i' = 0,\ 1,\ 2,\ \ldots,\ \ell+1\}$.

Lemma 11. *Let the Assumptions 16 to 18 hold, and $R_k^{s,e}$ be defined as above. Consider two arbitrary consecutive change points s and e, then*

(a) $\widetilde{Z}_{s,e}(k)$ *is strictly monotonic in k.*

(b) $\underset{s\le k\le e}{\operatorname{argmax}} \mid \widetilde{Z}_{s,e}(k) \mid$ *is reached exactly at one of the change points.*

Lemma 12. *Let the Assumptions 16 to 18 hold. For*

$$b = \underset{s\le k\le e}{\operatorname{argmax}} \mid Z_{s,e}(k) \mid,$$

it follows that

$$\min_{\substack{j\in\{1,\ldots,\ell\} \\ k^*_j\in[s,e]}} \mid b - k^*_j \mid = \gamma_n \ ,$$

where s, $e \in \{k^*_{i'} \mid i' \neq i,\ i' = 0,\ 1,\ 2,\ \ldots,\ \ell+1\}$, and $\gamma_n = O(n^\alpha)$, with $\frac{1}{2} < \alpha < 1$.

Proof of Lemma 6. Without loss of generality, consider the case of the whole segment where $[s,\ e] = [1,\ n]$. Suppose $X_i \sim Q_i$, $i = 1,\ \ldots,\ n$ and there are ℓ change points located at $1 = k^*_0 < k^*_1 < \cdots < k^*_\ell < k^*_{\ell+1} = n$, that is

$$Q_1 = \cdots = Q_{k^*_1-1} \neq Q_{k^*_1} = \cdots = Q_{k^*_2-1} \neq \cdots \neq Q_{k^*_\ell} = \cdots = Q_n\,.$$

Suppose for the segment $[s,\, e]$ the dominant change point is $k = k_i^*$ for some $i \in \{1,\, \ldots,\, \ell\}$. Let F_k, G_k, and $H_{s,\,e}$ be the true distribution of the observations in the segments $[s,\, k-1]$, $[k,\, e]$, and the whole segment $[s,\, e]$, respectively. Then

$$F_k = \frac{n_{s,e}}{(k-1)} \sum_{j=1}^{i-1} \left(\frac{k_j^* - k_{j-1}^*}{n_{s,e}} \right) Q_{k_j^*} \,,$$

$$G_k = \frac{n_{s,e}}{(e-k+1)} \sum_{j=i}^{\ell+1} \left(\frac{k_j^* - k_{j-1}^*}{n_{s,e}} \right) Q_{k_j^*} \,,$$

where F_k and G_k are the mixture of distributions on the left and right hand sides of the change point k, respectively. Note that for the test statistic proposed in this paper, $J(t) = t - \frac{1}{2}$ for all $t \in [0, 1]$, from which we have

$$J(\widehat{H}_{s,\,e}(x)) = J(H_{s,\,e}(x)) + \widehat{H}_{s,\,e}(x) - H_{s,\,e}(x) \,,$$
$$H_{s,\,e}(x) = \left(\frac{k-1}{n_{s,e}} \right) F_k(x) + \left(\frac{e-k+1}{n_{s,e}} \right) G_k(x) \,.$$

The test statistic $T_{s,\,e}(k)$ can be decomposed as

$$T_{s,\,e}(k) = \int_{\mathcal{M}} J(H_{s,\,e}(x))\, dF_k(x) + \int_{\mathcal{M}} J(H_{s,\,e}(x))\, d(\widehat{F}_k(x) - F_k(x)) \tag{A.26}$$
$$+ \int_{\mathcal{M}} (\widehat{H}_{s,\,e}(x) - H_{s,\,e}(x))\, dF_k(x) + \int_{\mathcal{M}} (\widehat{H}_{s,\,e}(x) - H_{s,\,e}(x))\, d(\widehat{F}_k(x) - F_k(x)) \,. \tag{A.27}$$

The first term is the conditional mean of $T_{s,e}(k)$ given s and e, which is a bounded constant satisfying

$$\begin{aligned} \left| \int_{\mathcal{M}} J(H_{s,\,e}(x))\, dF_k(x) \right| &= \left| \int_{\mathcal{M}} \left(H_{s,\,e}(x) - \frac{1}{2} \right)\, dF_k(x) \right| \\ &= \left| \int_{\mathcal{M}} H_{s,\,e}(x)\, dF_k(x) - \frac{1}{2} \right| \leq \left| \int_{\mathcal{M}} H_{s,\,e}(x)\, dF_k(x) \right| + \frac{1}{2} \\ &\leq \frac{1}{\lambda_0} \left| \int_{\mathcal{M}} H_{s,\,e}(x)\, dH_{s,\,e}(x) \right| + \frac{1}{2} = \frac{1}{2} \left(\frac{1}{\lambda_0} + 1 \right) = O_p(1) \,. \end{aligned}$$

Note that the last inequality is obtained by the result from section 7.A.8 of Chernoff and Savage (1958), stating that $dF_k(x) \leq \lambda_0^{-1}\, dH_{s,\,e}(x)$, where

$$0 < \lambda_0 < \frac{k-1}{n_{s,e}} < 1 - \lambda_0 < 1 \,.$$

By using Dvoretzky–Kiefer–Wolfowitz (DKW) inequality, which implies

$$\max_{x\in\mathbb{R}} \mid \widehat{F}_k(x) - F_k(x) \mid = O_p\left(\frac{1}{\sqrt{n_{s,e}}}\right) ,$$

and the fact that $\max_{x\in\mathbb{R}} H_{s,\,e}(x) = 1$, for the second term of A.27 we have

$$\left| \int_{\mathcal{M}} J(H_{s,\,e}(x))\, d(\widehat{F}_k(x) - F_k(x)) \right| \leq \frac{1}{2} \left| \int_{\mathcal{M}} d(\widehat{F}_k(x) - F_k(x)) \right| = O_p\left(\frac{1}{\sqrt{n_{s,e}}}\right) .$$

Also, by using DKW, for the third term of A.27 we obtain

$$\left| \int_{\mathcal{M}} (\widehat{H}_{s,\,e}(x) - H_{s,\,e}(x))\, dF_k(x) \right| \leq \int_{\mathcal{M}} \max_{x\in\mathcal{M}} \mid \widehat{H}_{s,\,e}(x) - H_{s,\,e}(x) \mid \, dF_k(x) = O_p\left(\frac{1}{\sqrt{n_{s,e}}}\right) .$$

Finally, to bound the fourth term of A.27, we use the following decomposition

$$\begin{aligned} \int_{\mathcal{M}} (\widehat{H}_{s,e}(x) - H_{s,e}(x))\, d(\widehat{F}_k(x) - F_k(x)) &= \left(\frac{k-1}{n_{s,e}}\right) \int_{\mathcal{M}} (\widehat{F}_k(x) - F_k(x))\, d(\widehat{F}_k(x) - F_k(x)) \\ &+ \left(\frac{e-k+1}{n_{s,e}}\right) \int_{\mathcal{M}} (\widehat{G}_k(x) - G_k(x))\, d(\widehat{F}_k(x) - F_k(x)) . \end{aligned}$$

By integration by parts and DKW, it follows that

$$\begin{aligned} \int_{\mathcal{M}} (\widehat{F}_k(x) - F_k(x))\, d(\widehat{F}_k(x) - F_k(x)) &= \frac{1}{2}(\widehat{F}_k(x) - F_k(x))^2 \\ &\leq \frac{1}{2} \max_{x\in\mathcal{M}} (\widehat{F}_k(x) - F_k(x))^2 = O_p(n_{s,e}^{-1}) . \end{aligned}$$

To get the bounds on the second part of the decomposition above, we use the result

$$\widehat{G}_k(x) - G_k(x) \sim \mathrm{N}\left(0, \frac{G_k(x)(1 - G_k(x))}{e-k+1}\right) ,$$

and the fact $\{X_s, X_{s+1}, \ldots, X_{k-1}\}$ is independent from $\{X_k, X_{k+1}, \ldots, X_e\}$. Then, we can proceed as follows

$$\mathrm{E}\Big[\int_{\mathcal{M}} (\widehat{G}_k(x) - G_k(x)) d(\widehat{F}_k(x) - F_k(x)) | s, e \Big]$$

$$= \mathrm{E}\Big[\mathrm{E}\Big[\int_{\mathcal{M}} (\widehat{G}_k(x) - G_k(x)) d(\widehat{F}_k(x) - F_k(x)) | X_1, X_2, \ldots, X_{k-1}\Big] \mid s, e\Big]$$
$$\leq C\, \mathrm{E}\Big[\max_{x \in \mathcal{M}} \mid \widehat{G}_k(x) - G_k(x) \mid \Big] = O(n_{s,e}^{-\frac{1}{2}}) \ ,$$

where C is a positive constant. Let

$$V_1, V_2, \ldots, V_{k-1} \sim \mathrm{N}\left(0, \frac{F_k(x)(1 - F_k(x))}{k-1}\right) \ ,$$

it can be seen that

$$\begin{aligned}
&\mathrm{Var}\Big[\int_{\mathcal{M}} (\widehat{G}_k(x) - G_k(x)) d(\widehat{F}_k(x) - F_k(x)) | s, e\Big] \\
&= \mathrm{E}\Big[\Big(\int_{\mathcal{M}} (\widehat{G}_k(x) - G_k(x)) d(\widehat{F}_k(x) - F_k(x))\Big)^2 | s, e\Big] \\
&= \frac{1}{(k-1)^2}\left(\sum_{i=1}^{k-1} \mathrm{E}\Big[(\widehat{G}_k(V_i) - G_k(V_i))^2\Big] + \sum_{i=1}^{k-1}\sum_{\substack{j=1 \\ j \neq i}}^{k-1} \mathrm{E}\Big[(\widehat{G}_k(V_i) - G_k(V_i))(\widehat{G}_k(V_j) - G_k(V_j))\Big]\right) \\
&\approx \frac{1}{(k-1)^2}\left(\sum_{i=1}^{k-1} \frac{G_k(Z_i)(1 - G_k(Z_i))}{e-k+1} + \sum_{i=1}^{k-1}\sum_{\substack{j=1 \\ j \neq i}}^{k-1} \frac{G_k(Z_i)(1 - G_k(Z_j))}{e-k+1}\right) \\
&= \frac{1}{(k-1)(e-k+1)} \int_{\mathcal{M}} G_k(x)(1 - G_k(x)) d(\widehat{F}_k(x) - F_k(x)) \\
&\qquad + \frac{2}{e-k+1} \int_{\mathcal{M}} \int_x G_k(x)(1 - G_k(y)) d(\widehat{F}_k(x) - F_k(x)) d(\widehat{F}_k(y) - F_k(y)) \ ,
\end{aligned}$$

which is $O_p(n_{s,e}^{-\frac{5}{2}}) + O_p(n_{s,e}^{-2})$, given the fact that $G_k(x)(1 - G_k(x))$ and $G_k(x)(1 - G_k(y))$ are non-deterministic and always less than 1. Then by Chebyshev's inequality, we get

$$\int_{\mathcal{M}} (\widehat{G}_k(x) - G_k(x)) d(\widehat{F}_k(x) - F_k(x)) = O_p(n_{s,e}^{-\frac{1}{2}}) \ .$$

This means that the fourth therm in the decomposition becomes $O_p(n_{s,e}^{-\frac{1}{2}})$, and thus

$$T_{s,e}(k) = O_p(1) + O_p(n_{s,e}^{-\frac{1}{2}}) + O_p(n_{s,e}^{-\frac{1}{2}}) + O_p(n_{s,e}^{-\frac{1}{2}}) \ .$$

This completes the proof. □

Proof of Lemma 7. The rank-based statistic $Z_{s,e}(k)$ can be represented as a function of $T_{s,e}(k)$, as below

$$
\begin{aligned}
Z_{s,e}(k) :&= \frac{1}{\sqrt{n_{s,e}}} \sum_{i=s}^{k} \frac{\widehat{R}_i^{s,e} - (n_{s,e}+1)/2}{\sqrt{(n_{s,e}^2-1)/12}} \\
&= \frac{1}{\sqrt{n_{s,e}}} \sum_{i=s}^{n_{s,e}} \frac{n_{s,e}(\widehat{R}_i^{s,e}/n_{s,e}) - (n_{s,e}+1)/2}{\sqrt{(n_{s,e}^2-1)/12}} \, \zeta_{i,n_{s,e}}(k) \\
&= \frac{1}{\sqrt{n_{s,e}}} \sum_{i=s}^{n_{s,e}} \frac{n_{s,e}\,\Gamma(\widehat{R}_i^{s,e}/n_{s,e})}{\sqrt{(n_{s,e}^2-1)/12}} \, \zeta_{i,n_{s,e}}(k) \\
&= \sqrt{\frac{n_{s,e}^3}{(n_{s,e}^2-1)/12}} \left(\frac{k-s}{n_{s,e}}\right) T_{s,e}(k) \;,
\end{aligned}
$$

where we used the fact that $(k-s)\,T_{s,e}(k) = \sum_{i=1}^{n_{s,e}} \Gamma(\widehat{R}_i^{s,e}/n_{s,e})\,\zeta_{i,n_{s,e}}(k)$, as well as the approximation $(n_{s,e}+1)/2 \approx n_{s,e}/2$. The polynomial $\Gamma(\widehat{R}_i^{s,e}/n_{s,e}) = \widehat{R}_i^{s,e}/n_{s,e} - 1/2$ is non-degenerate satisfying the second condition of *Theorem 4* of Chernoff and Savage (1958). In the continuous format, $\Gamma(\widehat{R}_i^{s,e}/n_{s,e})$ is equivalent to $J(H)$. Thus, $Z_{s,e}(k) = \sqrt{n_{s,e}} \left(\frac{k-s}{n_{s,e}}\right) T_{s,e}(k)$, which completes the proof. □

Proof of Lemma 8. From Lemma 6 and Lemma 7, it can be shown that $Z_{s,e}(k) = O_p(\sqrt{n_{s,e}})$. Therefore, it follows that

$$
\left| \frac{Z_{s,e}(k_i^*)}{a_n} \right| = O_p\left(n_{s,e}^{\epsilon}\right) \;, \qquad \text{for all } \epsilon > 0 \,.
$$

Since by Assumption 16, the number of change points ℓ is fixed and thus for any $\Delta > 0$, we have

$$
\lim_{n\to\infty} \Pr\left(\left| \frac{Z_{s,e}(k_i^*)}{a_n} \right| > \Delta \,\middle|\, s,\, e \right) \to 1 \,,
$$

for $s,\, e \in \{k_{i'}^* \mid i' \neq i,\, i' \in \{0,\, 1,\, 2,\, \ldots,\, \ell+1\}\}$. This completes the proof. □

Proof of Lemma 9. By using Lemma 7, and $H_{s,e}(x)$ defined in the proof of Lemma 6, it can be shown that

$$\begin{aligned}
\max_{s<k<e} \mid \widetilde{Z}_{s,e}(k) - Z_{s,e}(k) \mid &= \max_{s\le k\le e} \frac{1}{\sqrt{n_{s,e}}} \left| \sum_{i=s}^{k} \frac{R_i^{s,e} - \widehat{R}_i^{s,e}}{\sqrt{(n_{s,e}^2-1)/12}} \right| \\
&= \max_{s\le k\le e} \left| \sqrt{n_{s,e}} \left(\frac{k-s}{n_{s,e}}\right) \int_{\mathcal{M}} \left(J(H_{s,e}(x)) - J(\widehat{H}_{s,e}(x)) \right) d\widehat{F}_k(x) \right| \\
&= \max_{s\le k\le e} \left| \sqrt{n_{s,e}} \left(\frac{k-s}{n_{s,e}}\right) \int_{\mathcal{M}} \left(H_{s,e}(x) - \widehat{H}_{s,e}(x) \right) d\widehat{F}_k(x) \right|
\end{aligned}$$

where $\widehat{H}_{s,e}(x)$ is the sample version of $H_{s,e}(x)$. It is clear that the integral above is equivalent to the last two terms in the decomposition of $T_{s,e}(k)$ in the proof of Lemma 6. To get a convergence rate, we require an upper bound for $n_{s,e}$. It is clear that for rank-binary segmentation, $n_{s,e} = O(n)$, as long as $s,\, e \in \{k_{i'}^* \mid i' \neq i,\, i' = 0,\, 1,\, 2,\, \ldots,\, \ell+1\}$. However, for rank-wild binary segmentation, due to random sampling performed in each segment, bounding $n_{s,e}$ needs a more careful discussion. First, let us do the proof for rank-binary segmentation. From Lemma 6, it can be easily checked that

$$\max_{s\le k\le e} \left| \sqrt{n_{s,e}} \left(\frac{k-s}{n_{s,e}}\right) \int_{\mathcal{M}} \left(H_{s,e}(x) - \widehat{H}_{s,e}(x) \right) d\widehat{F}_k(x) \right| = \sqrt{n_{s,e}}\, O_p(n_{s,e}^{-\frac{1}{2}}) = O_p(1)\ ,$$

for $s,\, e \in \{k_{i'}^* \mid i' \neq i,\, i' = 0,\, 1,\, 2,\, \ldots,\, \ell+1\}$. This gives the result for the rank-binary segmentation method.

Now, for rank-wild binary segmentation, due to the random selection of sub-intervals in each segment, there is no guarantee that $n_{s,e} = O(n)$ is always satisfied. We do the proofs by considering two different conditions, when the selected sub-interval is large enough, satisfying $n_{s,e} = O(n)$, and when the selected sub-interval is small, i.e. $n_{s,e} < C$, where C is a constant finite positive integer. Let us indicate this two conditions by $\Theta_1 := \{n_{s,e} = O(n)\}$ and $\Theta_2 := \{n_{s,e} < C\}$. For the condition Θ_1, the above results regarding rank-binary segmentation holds. However, for Θ_2, it can be shown that

$$\max_{s\le k\le e} \left| \widetilde{Z}_{s,e}(k) - Z_{s,e}(k) \right| = O_p(1)\,.$$

It is also true that $\widetilde{Z}_{s,e}(k) = O_p(1)$, and thus, there is no convergence. This is an undesirable property, which can potentially degenerate the performance of rank-wild binary segmentation if all sub-intervals satisfy condition Θ_2. However, as the sample size n grows to infinity, probability of such problematic intervals shrink to zero. To see this, without loss

of generality we work with the segment $[1,\, n]$. Let $\mathcal{J}_1$ and $\mathcal{J}_2$ be the set of all sub-intervals satisfying condition Θ_1 and Θ_2, respectively. Then

$$\frac{\#\mathcal{J}_2}{\#\mathcal{J}_1} = \frac{1 + n + \binom{n}{2} + \ldots \binom{n}{C}}{2^n} = \frac{O(n^C)}{O(2^n)} = o(1)\,.$$

This completes the proof. □

Remark 2. *Our simulation studies indicated that even for the choice of $M_n = O(\log n)$, rank-wild binary segmentation efficiently detected the true change points, and the sub-intervals were chosen properly.*

Proof of Lemma 10. The proof is easily followed by the results of Lemma 8 and Lemma 9.

□

Proof of Lemma 11. (a): Let us denote the deterministic step function f_t by $f_t = \sum_{i=1}^{\ell+1} \mu_i \, \mathbb{I}_{\mathcal{A}_i}(t)$, where $\mathcal{A}_i = \{\lfloor n\theta_{i-1} \rfloor, \ldots, \lfloor n\theta_i \rfloor - 1\}$, $\mathbb{I}_{\mathcal{A}_i}(t)$ is the indicator function of set $\mathcal{A}_i$, and $\theta_1 < \theta_2 < \ldots < \theta_\ell$ are the breaking fractions, $\theta_0 = 0$ and $\theta_{\ell+1} = 1$. Then, based on the definition of $R_i^{s,\,e}$, the ranks remain constant between the change points $\lfloor n\theta_{i-1} \rfloor$, $\lfloor n\theta_i \rfloor$. Therefore

$$\widetilde{Z}_{s,\,e}(k) = \frac{1}{\sqrt{n_{s,\,e}}} \sum_{j=s}^{k} \frac{R_j^{s,\,e} - (n_{s,\,e}+1)/2}{\sqrt{(n_{s,\,e}^2 - 1)/12}}\,,$$

becomes strictly monotonic, for $k \in [s,\, e]$, where $s = \lfloor n\theta_{i-1} \rfloor$ and $e = \lfloor n\theta_i \rfloor - 1$.

(b) Due to the fact that $\widetilde{Z}_{s,\,e}(k)$ is strictly monotonic in $k \in [s, e]$, the maximum value is reached exactly at one of the change points.

□

Proof of Lemma 12. Let $b_1 = \underset{s \le k \le e}{\operatorname{argmax}} \mid \widetilde{Z}_{s,\,e}(k) \mid$. By Lemma 9, it holds that

$$\mid \widetilde{Z}_{s,\,e}(b_1) \mid \le \mid Z_{s,\,e}(b_1) \mid + \lambda_n \le \mid Z_{s,\,e}(b) \mid + \lambda_n \le \mid \widetilde{Z}_{s,\,e}(b) \mid + 2\lambda_n\,,$$

where $\lambda_n = O_p(1)$. Let $b \in (k_i^* + \gamma_n, k_{i+1}^* - \gamma_n)$, where $i = \underset{j \in \{1,\,\ldots,\,\ell\}}{\operatorname{argmin}} \mid b - k_j^* \mid$. Without loss of generality, let $\widetilde{Z}_{s,\,e}(b) > 0$. From Lemma 10, we know that $\widetilde{Z}_{s,\,e}(b)$ is strictly

monotonic in $[k_i^*,\, k_{i+1}^*]$. Thus, we have $\widetilde{Z}_{s,\,e}(b) < \max\{\widetilde{Z}_{s,\,e}(k_i^*), \widetilde{Z}_{s,\,e}(k_{i+1}^*)\}$. If $\widetilde{Z}_{s,\,e}(b)$ is locally decreasing at b, then $\widetilde{Z}_{s,\,e}(k_i^*) > \widetilde{Z}_{s,\,e}(b)$. Also, it can be shown that there exists $b' \in (k_i^*,\, k_i^* + \gamma_n]$, such that

$$\begin{aligned}
\widetilde{Z}_{s,\,e}(k_i^*) - \widetilde{Z}_{s,\,e}(b') &= \frac{1}{\sqrt{n_{s,\,e}}} \sum_{j=s}^{k_i^*} \frac{R_j^{s,\,e} - (n_{s,\,e}+1)/2}{\sqrt{(n_{s,\,e}^2-1)/12}} - \frac{1}{\sqrt{n_{s,\,e}}} \sum_{j=s}^{b'} \frac{R_j^{s,\,e} - (n_{s,\,e}+1)/2}{\sqrt{(n_{s,\,e}^2-1)/12}} \\
&= \frac{1}{\sqrt{n_{s,\,e}}} \sum_{j=k_i^*+1}^{b'} \frac{(n_{s,\,e}+1)/2 - R_j^{s,\,e}}{\sqrt{(n_{s,\,e}^2-1)/12}} \\
&= O_p\left(n^{\alpha-\frac{1}{2}}\right) \geq 2\,\lambda_n\ , \qquad \text{for all } \alpha > \frac{1}{2}\,.
\end{aligned}$$

This implies $|\,\widetilde{Z}_{s,\,e}(b_1)\,| \;>\; |\,\widetilde{Z}_{s,\,e}(b)\,| + 2\lambda_n$, which is a contradiction, as such a point cannot exist when $A_n = O_p(1)$, due to Lemma 9 ($|\,\widetilde{Z}_{s,\,e}(b_1)\,| > |\,Z_{s,\,e}(b_1)\,| + \lambda_n$ contradicts the results of Lemma 9). The same argument can be followed when $\widetilde{Z}_{s,\,e}(b)$ is locally increasing at b. Thus, $b \in (k_i^*,\, k_i^* + \gamma_n]$. Note that the upper bound $\alpha < 1$ is automatically satisfied since, due to Assumption 16, $\delta_n = O(n)$.

□

Proof of Theorem 19. The proof is given by investigating convergence properties of the following two events

$$\begin{aligned}
\Psi_1 &:= \{\widehat{\ell} = \ell\}\ , \\
\Psi_2 &:= \{\max_{i=1,\,\ldots,\,\ell} |\,\widehat{k}_i^* - k_i^*\,| \leq C n^{\alpha}\ ,\ \frac{1}{2} < \alpha < 1\}\ .
\end{aligned}$$

First consider the complement of the event Ψ_1, i.e. $\Psi_1^c = \{\widehat{\ell} \neq \ell\} = \{\widehat{\ell} < \ell\} \cup \{\widehat{\ell} > \ell\}$. If the event $\{\widehat{\ell} < \ell\}$ occurs, there exists at-least one segment which includes at-least one undetected change point, say $k_{i'}^* \in (k_i^*,\, k_{i+2}^*)$, where $i' = i+1$. Then, by Lemma 12, we must have $|\,b - k_{i'}^*\,| = O_p(\gamma_n)$, where

$$b = \underset{k \in [k_i^*,\, k_{i+2}^*]}{\operatorname{argmax}} \ |\,Z_{k_i^*,\,k_{i+2}^*}(k)\,|\ .$$

Additionally, there is an important observation pertaining to the fact that the endpoints of segments are estimated and thus random. Therefore, the results derived so far for fixed

segments cannot be directly applied. Denote the estimated change point within the random interval $[\widehat{k}_i^*,\ \widehat{k}_{i+2}^*]$ by $\widehat{k}_{i+1}^*$. It can be shown that

$$Z_{\widehat{k}_i^*,\,\widehat{k}_{i+2}^*}(\widehat{k}_{i+1}^*) = \sqrt{\widehat{k}_{i+2}^* - \widehat{k}_i^*}\,\left(\frac{\widehat{k}_{i+1}^* - \widehat{k}_i^*}{\widehat{k}_{i+2}^* - \widehat{k}_i^*}\right)\int_{\mathcal{M}} J(\widehat{H}_{\widehat{H}_{\widehat{k}_i^*,\,\widehat{k}_{i+2}^*}}(x))\,d\widehat{F}_{\widehat{k}_{i+1}^*}(x)\ ,$$

where

$$\widehat{F}_{\widehat{k}_{i+1}^*}(x) = \frac{1}{\widehat{k}_{i+1}^* - \widehat{k}_i^*}\sum_{j=\widehat{k}_i^*}^{\widehat{k}_{i+1}^*-1}\mathbb{I}(X_j\ x)\,,$$

$$\widehat{H}_{\widehat{k}_i^*,\,\widehat{k}_{i+2}^*}(x) = \frac{1}{\widehat{k}_{i+2}^* - \widehat{k}_i^*}\sum_{j=\widehat{k}_i^*}^{\widehat{k}_{i+2}^*}\mathbb{I}(X_j \in x)\,.$$

By decompositions

$$J(\widehat{H}_{\widehat{k}_i^*,\,\widehat{k}_{i+2}^*}(x)) = J(\widehat{H}_{k_i^*,\,k_{i+2}^*}(x)) + \widehat{H}_{\widehat{k}_i^*,\,\widehat{k}_{i+2}^*}(x) - \widehat{H}_{k_i^*,\,k_{i+2}^*}(x)\,,$$
$$\widehat{F}_{\widehat{k}_{i+1}^*}(x) = \widehat{F}_{k_{i+1}^*}(x) + \widehat{F}_{\widehat{k}_{i+1}^*}(x) - \widehat{F}_{k_{i+1}^*}(x)\,,$$

it follows that

$$\begin{aligned}\int_{\mathcal{M}} J(\widehat{H}_{\widehat{k}_i^*,\,\widehat{k}_{i+2}^*}(x))\,d\widehat{F}_{\widehat{k}_{i+1}^*}(x) = &\int_{\mathcal{M}} J(\widehat{H}_{k_i^*,\,k_{i+2}^*}(x))\,d\widehat{F}_{k_{i+1}^*}(x)\\ &+\int_{\mathcal{M}} J(\widehat{H}_{k_i^*,\,k_{i+2}^*}(x))\,d(\widehat{F}_{\widehat{k}_{i+1}^*}(x) - \widehat{F}_{k_{i+1}^*}(x))\\ &+\int_{\mathcal{M}} (\widehat{H}_{\widehat{k}_i^*,\,\widehat{k}_{i+2}^*}(x) - \widehat{H}_{k_i^*,\,k_{i+2}^*}(x))\,d\widehat{F}_{k_{i+1}^*}(x)\\ &+\int_{\mathcal{M}} (\widehat{H}_{\widehat{k}_i^*,\,\widehat{k}_{i+2}^*}(x) - \widehat{H}_{k_i^*,\,k_{i+2}^*}(x))\,d(\widehat{F}_{\widehat{k}_{i+1}^*}(x) - \widehat{F}_{k_{i+1}^*}(x))\,.\end{aligned}$$

By Lemma 12, for the event $\{\widehat{k}^*_{i+2} > k^*_{i+2},\, \widehat{k}^*_i < k^*_i\}$, we have

$$\begin{aligned}
\max_{x\in\mathbb{R}} \mid \widehat{H}_{\widehat{k}^*_i,\,\widehat{k}^*_{i+2}}(x) - \widehat{H}_{k^*_i,\,k^*_{i+2}}(x) \mid &= \max_{x\in\mathbb{R}} \left| \frac{1}{\widehat{k}^*_{i+2} - \widehat{k}^*_i} \sum_{j=\widehat{k}^*_i}^{\widehat{k}^*_{i+2}} \mathbb{I}(X_j \in x) - \frac{1}{k^*_{i+2} - k^*_i} \sum_{j=k^*_i}^{k^*_{i+2}} \mathbb{I}(X_j \in x) \right| \\
&= \frac{\max\limits_{x\in\mathbb{R}} \left| (k^*_{i+2} - k^*_i) \sum\limits_{j=\widehat{k}^*_i}^{\widehat{k}^*_{i+2}} \mathbb{I}(X_j \in x) - (\widehat{k}^*_{i+2} - \widehat{k}^*_i) \sum\limits_{j=k^*_i}^{k^*_{i+2}} \mathbb{I}(X_j \in x) \right|}{(\widehat{k}^*_{i+2} - \widehat{k}^*_i)(k^*_{i+2} - k^*_i)} \\
&\le \max_{x\in\mathbb{R}} \left| \frac{(k^*_{i+2} - k^*_i) - (\widehat{k}^*_{i+2} - \widehat{k}^*_i)}{(\widehat{k}^*_{i+2} - \widehat{k}^*_i)(k^*_{i+2} - k^*_i)} \sum_{j=k^*_i}^{k^*_{i+2}} \mathbb{I}(X_j \in x) \right| \\
&+ \frac{\max\limits_{x\in\mathbb{R}} \left| \sum\limits_{j=\widehat{k}^*_i}^{k^*_i - 1} \mathbb{I}(X_j \in x) \right|}{(\widehat{k}^*_{i+2} - \widehat{k}^*_i)} + \frac{\max\limits_{x\in\mathbb{R}} \left| \sum\limits_{j=k^*_{i+2}+1}^{\widehat{k}^*_{i+2}} \mathbb{I}(X_j \in x) \right|}{(\widehat{k}^*_{i+2} - \widehat{k}^*_i)} \\
&\le 2 \left| \frac{k^*_i - \widehat{k}^*_i}{\widehat{k}^*_{i+2} - \widehat{k}^*_i} \right| + 2 \left| \frac{\widehat{k}^*_{i+2} - k^*_{i+2}}{\widehat{k}^*_{i+2} - \widehat{k}^*_i} \right| = O_p(n^{\alpha-1}) \,.
\end{aligned}$$

which in turn, since $1/2 < \alpha < 1$, implies probabilistic convergence to zero with the rate $n^{\alpha-1}$, as $n \to \infty$. The same convergence rate is valid for the events $\{\widehat{k}^*_{i+2} < k^*_{i+2},\, \widehat{k}^*_i > k^*_i\}$, $\{\widehat{k}^*_{i+2} > k^*_{i+2},\, \widehat{k}^*_i > k^*_i\}$, and $\{\widehat{k}^*_{i+2} < k^*_{i+2},\, \widehat{k}^*_i < k^*_i\}$. Similarly, it can be shown that

$$\max_{x\in\mathbb{R}} \mid \widehat{F}_{\widehat{k}^*_{i+1}}(x) - \widehat{F}_{k^*_{i+1}}(x) \mid = O_p(n^{\alpha-1}) \,.$$

The above results can be easily extended to the case with more than one undetected change point at each random interval. Hence, by following steps similar to Lemma 6, it can be shown that

$$\begin{aligned}
\int_{\mathcal{M}} J(\widehat{H}_{\widehat{k}^*_i,\,\widehat{k}^*_{i+2}}(x))\, d\widehat{F}_{\widehat{k}^*_{i+1}}(x) &= \int_{\mathcal{M}} J(\widehat{H}_{k^*_i,\,k^*_{i+2}}(x))\, d\widehat{F}_{k^*_{i+1}}(x)(x) + O_p(n^{\alpha-1}) \\
&= O_p(1) + O_p(n^{-\frac{1}{2}}) + O_p(n^{\alpha-1}) \\
&= O_p(n^{\alpha-1}) \,,
\end{aligned}$$

meaning that when there exists at-least one undetected change point k_i, the statistic with random interval satisfies

$$\lim_{n\to\infty} \Pr\left(\left| \frac{Z_{\widehat{k}^*_i,\,\widehat{k}^*_{i+2}}(\widehat{k}^*_{i+1})}{c_{\alpha_i}} \right| > \Delta \,\middle|\, \widehat{k}^*_i,\, \widehat{k}^*_{i+2} \right) \to 1 \,, \qquad \text{for all } \Delta > 0 \,.$$

This implies $\Pr(\{\widehat{\ell} < \ell\}) = o(1)$. By the same argument, it can be inferred that if there are undetected change points remaining, the algorithm proceeds and detects them all.

Now, consider the event $\{\widehat{\ell} > \ell\}$ holds, which implies the test statistic falsely rejected the null hypothesis at certain stages of the binary segmentation process. Thus

$$\begin{aligned}
\Pr\Big(\{\widehat{\ell} > \ell\}\Big) &\leq \Pr\Big(\exists\, [\widehat{k}_i^*,\, \widehat{k}_{i+1}^*] \text{ that the null hypothesis of no change point is rejected } \Big) \\
&\leq \sum_{i=0}^{\ell} \Pr\Big(\text{ the null hypothesis of no change point is rejected for } [\widehat{k}_i^*, \widehat{k}_{i+1}^*]\Big) \\
&= \sum_{i=0}^{\ell} \Pr\left(\max_{k \in [\widehat{k}_i^*,\, \widehat{k}_{i+1}^*]} \mid \widetilde{Z}_{\widehat{k}_i^*,\, \widehat{k}_{i+1}^*}(k) \mid > c_{\alpha_i} \right) \\
&\leq (\ell+1) \max_{i \in \{1,\ldots,\ell\}} \Pr\left(\max_{k \in [\widehat{k}_i^*,\, \widehat{k}_{i+1}^*]} \mid \widetilde{Z}_{\widehat{k}_i^*,\, \widehat{k}_{i+1}^*}(k) \mid > c_{\alpha_i} \right) .
\end{aligned}$$

By *Theorem 3.1* of Chenouri et al. (2020), under the hypothesis of no change, the test statistic $\max_{k \in [s,e]} \mid \widetilde{Z}_{s,e}(k) \mid$ converges in distribution to the distribution of the maximum of the absolute value of the standard Brownian Bridge process, which implies that

$$\max_{i \in \{1,\ldots,\ell\}} \Pr\Big(\max_{k \in [\widehat{k}_i^*, \widehat{k}_{i+1}^*]} \mid \widetilde{Z}_{\widehat{k}_i^*, \widehat{k}_{i+1}^*}(k) \mid > c_{\alpha_i}\Big) = o(1)\,,$$

since $c_{\alpha_i} = o(\sqrt{n})$. Thus, it also holds that $\Pr(\{\widehat{\ell} > \ell\}) = o(1)$, as ℓ is a finite positive integer. Therefore $\Pr(\{\widehat{\ell} < \ell\} \cup \{\widehat{\ell} > \ell\}) = o(1)$, which in turn implies $\Pr(\Psi_1) \to 1$, as $n \to \infty$.

For the event Ψ_2, using Bonferroni inequality and knowing that by Lemma 12, we have

$$\Pr(\mid \widehat{k}_i^* - k_i^* \mid > Cn^{\alpha}) = o(1) \quad \text{for} \quad i = 1, \ldots, \ell\,,$$

we get

$$\begin{aligned}
\Pr(\Psi_2) &= \Pr\left(\max_{i \in \{1,\ldots,\ell\}} \mid \widehat{k}_i^* - k_i^* \mid \leq Cn^{\alpha} \right) \\
&= \Pr\left(\bigcap_{i=1}^{\ell} \mid \widehat{k}_i^* - k_i^* \mid \leq Cn^{\alpha} \right) \\
&\geq 1 - \sum_{i=1}^{\ell} \Pr\Big(\mid \widehat{k}_i^* - k_i^* \mid > Cn^{\alpha} \Big) .
\end{aligned}$$

That is $\Pr(\Psi_2) \to 1$ as $n \to \infty$. Combining the results for the events Ψ_1 and Ψ_2, we have $\Pr(\Psi_1 \cap \Psi_2) \to 1$ as $N \to \infty$. This completes the proof. □

Proof of Theorem 20. The proof can be done following similar steps given in the proof of Theorem 18 and Theorem 19.

□

Proof of Theorem 21. By the sampling strategy of rank-wild binary segmentation, at each stage of segmentation, M_n sub-inervals are uniformly drawn from the initial interval. Suppose there is only one change point in the interval $[s, e]$. Also, note that $s_m \in \{s, \ldots, e-1\}$ and $e_m \in \{s_m + 1, ..., e\}$, for all $m \in \{1, \ldots, M_n\}$. Let $\mathcal{J}$ be the set of all possible sub-intervals that can be drawn from $[s, e]$. By following such a sampling mechanism, the total number of sub-intervals, $\#\mathcal{J}$, can be obtained as

$$\sum_{i=s}^{e-1}(e-i) = \frac{(e-s)(e-s-1)}{2} \ .$$

Now, let us partition $\mathcal{J} = \mathcal{J}_1 \cup \mathcal{J}_1^c$, where $\mathcal{J}_1$ is the set of those sub-intervals which include the change point, and $\mathcal{J}_1^c$ includes the rest of the intervals. Denote the change point location by k^*. Based on Assumption 16, as $n \to \infty$, it holds that $k^*/n \to \theta \in (0, 1)$. It is easy to check that $\#\mathcal{J}_1 = (e - k^*)(k^* - s)$. Then, the probability of the event that none of the M_n sub-intervals contain the change point becomes

$$\begin{aligned}
\Pr\left(\text{no sub-interval include the change point }\right) &= \prod_{m=1}^{M_n} \Pr\left([s_m, e_m] \in \mathcal{J}_1^c\right) \\
&= \left(1 - \frac{\#\mathcal{J}_1}{\#\mathcal{J}}\right)^{M_n} \\
&\approx \left(1 - \theta(1-\theta)\right)^{M_n} \ .
\end{aligned}$$

This implies that, as $n \to \infty$, with probability exponentially converging to 1 at-least one sub-interval includes the change point. This can be extended to the case with more than one change point within the segment, given the stage-wise (one by one) detection of change points in rank-wild binary segmentation, and the fact that the total number of change points is kept fixed while $n \to \infty$. Let $\{k_1^*, k_2^*, \ldots, k_\ell^*\}$ and $\{\theta_1, \theta_2, \ldots, \theta_\ell\}$ be the

locations and breaking fractions of these change points. Then, it can be shown that

$$\begin{aligned}\Pr(\mathcal{D}_n^c) &\leq \sum_{i=1}^{\ell} \prod_{m_i=1}^{M_n} \Pr\left([s_{m_i},\, e_{m_i}] \in \mathcal{J}_{1,i}^c\right) \\ &\leq \ell \max_{i\in\{1,\ldots,\ell\}} \prod_{m_i=1}^{M_n} \Pr\left([s_{m_i},\, e_{m_i}] \in \mathcal{J}_{1,\,i}^c\right) \\ &\approx \ell(1-\theta^*(1-\theta^*))^{M_n}\ ,\end{aligned}$$

which goes to zero since ℓ is a finite positive integer number.

Note that $\theta^* = \min\{\theta_1, \theta_2, \ldots, \theta_\ell\}$. Therefore, $\Pr(\mathcal{D}_n) \gtrapprox 1 - \ell(1-\theta^*(1-\theta^*))^{M_n}$, and as $N \to \infty$, $\Pr(\mathcal{D}_n) \to 1$. Also, $A_n = O_p(1)$ is satisfied due to Lemma 9. Therefore, there exists an interval $[s_{m_i}, e_{m_i}]$ that includes change point k_i^* for which all of the arguments given in Theorem 19 is valid, because rank-wild binary segmentation and rank-binary segmentation use the same statistic. This completes the proof.

□

Proof of Theorem 22. The proof can be done following similar steps given in the proof of Theorem 18 and Theorem 21.

□

Proof of Theorem 23. The proof can be done following similar steps given in the proof of Theorem 10.

□

Proof of Theorem 24. The proof can be done by contradiction. The argument holds for both ε_F and ε_D. Let $\epsilon_2^* < \epsilon_1^*$. For any choice of $\varepsilon_2^* < \varepsilon < \varepsilon_1^*$, the sequence of statistics $T(P_n)$ breakdown at ε-level of contamination. Hence, there exists a sequence of contaminating distributions $\{Q_m\}_{m=1}^{\infty}$ with probability density function $\{\widehat{g}_m\}_{m=1}^{\infty}$ so that for contaminated distribution $P_{n,\varepsilon,m} = (1-\varepsilon)\, P_n + \varepsilon Q_m$, we have $|T(P_{n,\varepsilon,m}) - T(P)| \to \infty$. Since $\varepsilon < \varepsilon_1^*$, the sequence of statistics $T(P)$ will not breakdown at ε-level of contamination, i.e. the functional $T(P_{\varepsilon,m})$ will not breakdown. Therefore, we will have a contradiction if $|T(P_{\varepsilon,m}) - T(P)| \to \infty$, as $m \to \infty$.

Let us define $f_{\varepsilon,m} = (1-\varepsilon)f + \varepsilon g_m$ and $\widehat{f}_{n,\varepsilon,m} = (1-\varepsilon)\widehat{f}_n + \varepsilon g_m$, for any fixed m. Given that $\int_{\mathcal{M}} |\widehat{f}_n(x) - f(x)|^p dx \xrightarrow{a.s.} 0$, there exists a subsequence $\{\widehat{f}_{n_j}\}_{j=1}^{\infty}$ for which we have a

pointwise converge $\widehat{f}_{n_j}(x) \overset{a.s.}{\to} f(x)$ for any $x \in \mathcal{M}$, as $n_j \to \infty$. Let $\delta_{n_j,\varepsilon,m} = |\widehat{f}_{n_j,\varepsilon,m} - f_{\varepsilon,m}|$. It can be shown that $\int_{\mathcal{M}} \delta^p_{n,\varepsilon,m} dx \overset{a.s.}{\to} 0$, therefore, by the same argument we have a pointwise converge $f_{n_j,\varepsilon,m}(x) \overset{a.s.}{\to} f_{\varepsilon,m}(x)$ for any $x \in \mathcal{M}$, as $n_j \to \infty$. Therefore, for any fixed ε, it holds that

$$\sup_{f:\ P \in \mathcal{P}(\mathcal{M})} \left\{ |T(P_{n_j,\varepsilon,m}) - T(P_{\varepsilon,m})| \right\} \overset{a.s.}{\to} 0 \ ,$$

which implies

$$\sup_{f:\ P \in \mathcal{P}(\mathcal{M})} \left\{ \left| \lim_{\mathbb{E}_P^{(\infty)} \to \infty} \inf_{H_0 \in \mathcal{H}_{0,\epsilon}} \frac{\mathbb{E}^{(\infty)}_{H_0}(T_{k^*})}{\mathbb{E}^{(\infty)}_{P_{n_j}}(T_{k^*})} - \lim_{\mathbb{E}_P^{(\infty)} \to \infty} \inf_{H_0 \in \mathcal{H}_{0,\epsilon}} \frac{\mathbb{E}^{(\infty)}_{H_0}(T_{k^*})}{\mathbb{E}^{(\infty)}_{P}(T_{k^*})} \right| \right\} \overset{a.s.}{\to} 0 \ ,$$

$$\sup_{f:\ P \in \mathcal{P}(\mathcal{M})} \left\{ \left| \lim_{\mathbb{E}_P^{(\infty)} \to \infty} \sup_{H_1 \in \mathcal{H}_{1,\epsilon}} \frac{\mathrm{WADD}_{P,H_1}(T_{k^*})}{\mathrm{WADD}_{P_{n_j},Q_{n_j}}(T_{k^*})} - \lim_{\mathbb{E}_P^{(\infty)} \to \infty} \sup_{H_1 \in \mathcal{H}_{1,\epsilon}} \frac{\mathrm{WADD}_{P,H_1}(T_{k^*})}{\mathrm{WADD}_{P,Q}(T_{k^*})} \right| \right\} \overset{a.s.}{\to} 0 \ .$$

Therefore, for any $\varepsilon < \varepsilon_1^*$, and any $\delta > 0$, there exists $m > M$, $n > n'_m$, with $\varepsilon = m/n$, such that

$$\sup_{f:\ P \in \mathcal{P}(\mathcal{M})} \left\{ |T(P_{n,\varepsilon,m}) - T(P)| \right\} > \delta \ ,$$

$$\sup_{f:\ P \in \mathcal{P}(\mathcal{M})} \left\{ \lim_{\mathbb{E}_P^{(\infty)} \to \infty} \inf_{H_0 \in \mathcal{H}_{0,\epsilon}} \frac{\mathbb{E}^{(\infty)}_{H_0}(T_{k^*})}{\mathbb{E}^{(\infty)}_{P_n}(T_{k^*})} \right\} = 0 \ ,$$

$$\sup_{f:\ P \in \mathcal{P}(\mathcal{M})} \left\{ \lim_{\mathbb{E}_P^{(\infty)} \to \infty} \sup_{H_1 \in \mathcal{H}_{1,\epsilon}} \frac{\mathrm{WADD}_{P,H_1}(T_{k^*})}{\mathrm{WADD}_{P_n,Q_n}(T_{k^*})} \right\} = \infty \ ,$$

and

$$\sup_{f:\ P \in \mathcal{P}(\mathcal{M})} \left\{ |T(P_{\varepsilon,m})(\omega) - T(P)| \right\} > \delta \ ,$$

$$\sup_{f:\ P \in \mathcal{P}(\mathcal{M})} \left\{ \lim_{\mathbb{E}_P^{(\infty)} \to \infty} \inf_{H_0 \in \mathcal{H}_{0,\epsilon}} \frac{\mathbb{E}^{(\infty)}_{H_0}(T_{k^*})}{\mathbb{E}^{(\infty)}_{P}(T_{k^*})} \right\} = 0 \ ,$$

$$\sup_{f:\ P \in \mathcal{P}(\mathcal{M})} \left\{ \lim_{\mathbb{E}_P^{(\infty)} \to \infty} \sup_{H_1 \in \mathcal{H}_{1,\epsilon}} \frac{\mathrm{WADD}_{P,H_1}(T_{k^*})}{\mathrm{WADD}_{P,Q}(T_{k^*})} \right\} = \infty \ ,$$

Since $\delta > 0$ is arbitrary, the result holds for both ε_F and ε_D.

□

A.3 Proofs from Chapter 4

Proof of Theorem 25. Let $a_n(i) = \frac{1}{n}R_i$ be the rank score and $b_n(i) = (1-\lambda)^{t-i}\lambda$. It can be verified that

$$n\frac{\max_{1\leq i\leq n}[a_n(i)-\bar{a}_n]^2}{\sum_{i=1}^n[a_n(i)-\bar{a}_n]^2}\cdot\frac{\max_{1\leq i\leq n}[b_n(i)-\bar{b}_n]^2}{\sum_{i=1}^n[b_n(i)-\bar{b}_n]^2} = o(1)\ .$$

Therefore the conditions of Hajek CLT given in Lemma 1 are satisfied, which gives

$$\sqrt{\frac{2-\lambda}{\lambda}}\left(\frac{\mathbf{S}_t^{EWMA}-\zeta}{\sigma}\right)\xrightarrow{\mathcal{D}}\mathrm{N}(0,1)\ ,$$

$$\mathbf{\Lambda}_t^{EWMA} = \frac{2-\lambda}{\lambda\sigma^2}\left(\mathbf{S}_t^{EWMA}-\zeta\right)^2\xrightarrow{\mathcal{D}}\chi_1^2\ ,$$

where $\sigma^2 = \mathrm{Var}[R_t]$, and $\zeta = \mathbb{E}[R_t]$. This completes the proof.

□

Proof of Theorem 26. Let $a_n(i) = \frac{1}{n}R_i$ be the rank score and $b_n(i) = 1$. It can be shown that

$$n\frac{\max_{1\leq i\leq n}[a_n(i)-\bar{a}_n]^2}{\sum_{i=1}^n[a_n(i)-\bar{a}_n]^2}\cdot\frac{\max_{1\leq i\leq n}[b_n(i)-\bar{b}_n]^2}{\sum_{i=1}^n[b_n(i)-\bar{b}_n]^2} = o(1)\ ,$$

satisfying the conditions of Hajek CLT. Therefore

$$\frac{\mathbf{S}_t^{CUSUM}-\zeta}{\sigma}\xrightarrow{\mathcal{D}}\mathrm{N}(0,1)\ ,$$

$$\mathbf{\Lambda}_t^{CUSUM} = \frac{1}{\sigma^2}\left(\mathbf{S}_t^{CUSUM}-\zeta\right)^2\xrightarrow{\mathcal{D}}\chi_1^2\ ,$$

where $\sigma^2 = \mathrm{Var}[R_t]$, and $\zeta = \mathbb{E}[R_t]$. This completes the proof.

□

Proof of Theorem 27. Let $a_n(i,j) = \mathbb{I}\Big(R_j > R_i\Big)$ be the rank score and $b_n(i,j) = 1$. We have

$$n_1 m\frac{\max_{1\leq i\leq n_1,1\leq j\leq m}[a_n(i,j)-\bar{a}_n]^2}{\sum_{i=1}^{n_1}\sum_{j=1}^{m}[a_n(i,j)-\bar{a}_n]^2}\cdot\frac{\max_{1\leq i\leq n_1,1\leq j\leq m}[b_n(i,j)-\bar{b}_n]^2}{\sum_{i=1}^{n_1}\sum_{j=1}^{m}[b_n(i,j)-\bar{b}_n]^2} = o(1)\ ,$$

satisfying the conditions of Hajek CLT. Hence

$$\frac{\mathbf{S}_t^{MWMASHC} - \zeta}{\sigma} \xrightarrow{\mathcal{D}} \mathrm{N}(0,1) \ ,$$

$$\mathbf{\Lambda}_t = \frac{1}{\sigma^2}\left(\mathbf{S}_{0,t}^{MWMASHC} - \zeta\right)^2 \xrightarrow{\mathcal{D}} \chi_1^2 \ .$$

where

$$\zeta = \frac{mn_1}{2} \quad , \quad \sigma^2 = \frac{mn_1(m+n_1+1)}{12} \quad .$$

This completes the proof.

□

Proof of Theorem 28. The proof is based on bounding the rank scores using empirical Rademacher complexity following steps similar to Theorem 5. For $\mathcal{X} = (x_1, \ldots, x_n) \in \mathcal{M}$, define the empirical Rademacher complexity as follows

$$\mathrm{Rad}_n(\mathcal{F}, \mathcal{X}) = \frac{1}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}\zeta_i f(x_i)\right|\right] \ ,$$

where $\zeta_1, \ldots, \zeta_n$ are i.i.d random variables following Rademacher distribution with $P(\zeta_i = 1) = P(\zeta_i = -1) = \frac{1}{2}$, and $\mathcal{F}$ is a function class on $\mathcal{M}$.

The Dudley entropy integral (Dudley, 2010; Talagrand, 1994) can be used to bound the Rademacher complexity

$$\mathrm{Rad}_n(\mathcal{F}, \mathcal{X}) \leq \inf_{\varepsilon>0}\left\{4\varepsilon + \frac{12}{\sqrt{n}}\int_{\varepsilon}^{c}\sqrt{\log \mathcal{N}(\mathcal{F}, t, L^2(\mathcal{X}))}dt\right\} \ ,$$

where $\mathcal{N}(\mathcal{F}, t, L^2(\mathcal{X}))$ is the covering number, and $c = \sup_{f\in\mathcal{F}}\sqrt{\sum_{i=1}^{n} f^2(x_i)/n}$.

By Assumption 2, and following steps similar to Theorem 5, we get the following bound for Rademacher complexity $\mathrm{Rad}_n(\mathcal{F}, \mathcal{X}) = O_p\left(n^{-1/2}(d\log d)^{1/2}\right)$, where the class of functions $\mathcal{F}$ is VC with VC-dimension d. Also, by following similar steps, we get the following convergence bound for IPM

$$D_{\mathcal{F}}(P, P_n) = O_p\left(n^{-1/2}(d\log d)^{1/2}\right) \ .$$

By Han et al. (2023), $D_{\mathcal{F}} = 0$ is equivalent to weak convergence, $P_n \xrightarrow{d} P$, with rate $O\left(n^{-1/2}(d\log d)^{1/2}\right)$. If $P_n \xrightarrow{d} P$, for a bounded and Lipschitz continuous function g defined on $\mathcal{M}$, we have $g_{P_n} \xrightarrow{p} g_P$ with the same rate. Hence, by Assumption 3 and the boundedness of depth functions we get

$$\sup_{x\in\mathcal{M}} \mid D(P_n, x) - D(P, x) \mid = O_p\left(n^{-1/2}(d\log d)^{1/2}\right) \ .$$

r-EWMA: Let $a(i) = \frac{1}{n}R_i$ be the rank score. It holds that

$$\mathbf{S}_t^{EWMA} = (1-\lambda)\mathbf{S}_{t-1}^{EWMA} + \lambda a(t) \ , \quad t = 1, 2, ..., n_1 + n_2 \ ,$$

where $\mathbf{S}_0^{EWMA} = 0$ and $\lambda \in (0, 1]$. Therefore

$$\begin{aligned}\mathbf{S}_t &= (1-\lambda)^{t-1}\lambda a(1) + (1-\lambda)^{t-2}\lambda a(2) + \ldots + \lambda a(t) \\ &= \sum_{i=0}^{t-1}(1-\lambda)^i \lambda a(t-i) \ .\end{aligned}$$

Let $t = n$, we can write

$$\begin{aligned}\mathbb{E}[\mathbf{S}_n^{EWMA}] &= \sum_{i=0}^{n-1}(1-\lambda)^i\lambda\mathbb{E}[a(n-i)] = \lambda\frac{1-(1-\lambda)^n}{1-(1-\lambda)}\mathbb{E}[a(i)] \\ &= (1-(1-\lambda)^n)\bar{a}_n \ , \\ \mathrm{Var}[\mathbf{S}_n^{EWMA}] &= \sum_{i=0}^{n-1}(1-\lambda)^{2i}\lambda^2\mathrm{Var}[a(n-i)] = \lambda^2\frac{1-(1-\lambda)^{2n}}{1-(1-\lambda)^2}\sigma_n^2 \ .\end{aligned}$$

By the argument given in Theorem 5, we have

$$\mid \widehat{a}(i) - a(i) \mid = O_p\left(n^{-1/2}(d\log d)^{1/2}\right) \ .$$

Let $\mathbf{S}_n^* = \dfrac{\mathbf{S}_n - \mathbb{E}[\mathbf{S}_n]}{\sqrt{\mathrm{Var}[\mathbf{S}_n]}}$. By following similar steps to Theorem 5, we get

$$\mid \mathbf{S}_n^* - \widehat{\mathbf{S}}_n^* \mid \leq \max_{i=1,\ldots,n} \mid \widehat{a}(i) - a(i) \mid = O_p\left(n^{-1/2}(d\log d)^{1/2}\right) \ .$$

Similarly, we get

$$\mid \mathbf{\Lambda}_n^* - \widehat{\mathbf{\Lambda}}_n^* \mid = O_p\left(n^{-1}(d\log d)\right) \ .$$

r-CUSUM: Let $a(i) = \frac{1}{n}R_i$ be the rank score. For r-CUSUM control chart, we have

$$\begin{aligned}
\mathbf{S}_n^{CUSUM_1} &= \max\{0, \mathbf{S}_{n-1}^{CUSUM_1} + (0.5 - a(n)) - k\} \\
&= \max\{0, \sum_{i=0}^{n-1}(0.5 - a(i)) - k\} \ , \\
\mathbf{S}_n^{CUSUM_2} &= \min\{0, \mathbf{S}_{n-1}^{CUSUM_2} + (0.5 - a(n)) + k\} \\
&= \min\{0, \sum_{i=0}^{n-1}(0.5 - a(i)) + k\} \ .
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbb{E}[\mathbf{S}_n^{CUSUM_1}] &= \max\{0, n(0.5 - \bar{a}_n - k)\} \ , \\
\mathbb{E}[\mathbf{S}_n^{CUSUM_2}] &= \max\{0, n(0.5 - \bar{a}_n + k)\} \ , \\
\mathrm{Var}[\mathbf{S}_n^{CUSUM_1}] &= \left(\mathbb{I}[\mathbf{S}_n^{CUSUM_1} > 0]\right) n\sigma_n^2 \ , \\
\mathrm{Var}[\mathbf{S}_n^{CUSUM_2}] &= \left(\mathbb{I}[\mathbf{S}_n^{CUSUM_2} < 0]\right) n\sigma_n^2 \ .
\end{aligned}$$

Following similar steps as above, we get

$$\begin{aligned}
\mid \mathbf{S}_n^* - \widehat{\mathbf{S}}_n^* \mid &\leq \max_{i=1,\ldots,n} \mid \widehat{a}(i) - a(i) \mid = O_p\left(n^{-1/2}(d\log d)^{1/2}\right) \ , \\
\mid \mathbf{\Lambda}_n^* - \widehat{\mathbf{\Lambda}}_n^* \mid &= O_p\left(n^{-1}(d\log d)\right) \ .
\end{aligned}$$

r-MWMASHC: Let $a_{ij} = \mathbb{I}\Big(\mathbf{R}_{g_n}(j) > R_i\Big)$ be the rank score. For r-MWMASHC control chart, we have

$$\begin{aligned}
\mathbf{S}_{0,n}^{MWMASHC} &= \sum_{i=1}^{n_1}\sum_{j=1}^{m}\mathbb{I}\Big(\mathbf{R}_{g_n}(j) > R_i\Big) = \sum_{i=1}^{n_1}\sum_{j=n-m+1}^{n}\mathbb{I}\Big(R_j > R_i\Big) \ , \\
\mathbb{E}[\mathbf{S}_{0,n}^{MWMASHC}] &= \sum_{i=1}^{n_1}\sum_{j=1}^{m}P\Big(\mathbf{R}_{g_t}(j) > R_i\Big) = \sum_{i=1}^{n_1}\sum_{j=n-m+1}^{n}\mathbb{E}[a_{i,j}] \ ,
\end{aligned}$$

$$
\begin{aligned}
\text{Var}[\mathbf{S}_{0,n}^{MWMASHC}] &= \sum_{i=1}^{n_1} \sum_{j=n-m+1}^{n} \text{Var}\,(a_{ij}) \\
&+ \sum_{i=1}^{n_1} \sum_{j=n-m+1}^{n} \sum_{\substack{l=1 \\ l\neq i}}^{n_1} \sum_{\substack{h=n-m+1 \\ h\neq j}}^{n} \text{Cov}\,(a_{ij}, a_{lk}) \\
&= \sum_{i=1}^{n_1} \sum_{j=n-m+1}^{m} P\Big(R_j > R_i\Big)\Big(1 - P(R_j > R_i)\Big) \\
&+ \sum_{i=1}^{n_1} \sum_{j=n-m+1}^{n} \sum_{\substack{h=n-m+1 \\ h\neq j}}^{n} P\Big(R_j > R_i, R_h > R_i\Big) - P\Big(R_j > R_i\Big)P\Big(R_h > R_i\Big) \ .
\end{aligned}
$$

By the argument given in Theorem 5, we have

$$
\mid \widehat{a}_{i,j} - a_{i,j} \mid = O_p\left(n^{-1/2}(d\log d)^{1/2}\right) \ .
$$

Following similar steps as above, we get

$$
\begin{aligned}
\mid \mathbf{S}_n^* - \widehat{\mathbf{S}}_n^* \mid &\leq \max_{i=1,\ldots,n} \mid \widehat{a}_{i,j} - a_{i,j} \mid = O_p\left(n^{-1/2}(d\log d)^{1/2}\right) \ , \\
\mid \mathbf{\Lambda}_n^* - \widehat{\mathbf{\Lambda}}_n^* \mid &= O_p\left(n^{-1}(d\log d)\right) \ .
\end{aligned}
$$

This completes the proof.

□

Proof of Theorem 29. The proof is given in two parts. In the first part, the proof is given for intrinsic isometric embedding of Riemannian manifolds. In the second part, the proof is given for extrinsic isometric embedding of Riemannian manifolds.

(I): Intrinsic Embedding. Consider the Phase-I sample set $(X_1, X_2, ..., X_n)$ which is assumed to be in-control with mean μ_x and variance ν_x. Also, consider the Phase-II sample set $(Y_1, Y_2, ..., Y_m)$. We denote the mean and variance of this set by μ_y and ν_y. We assume these sample sets take values in a bounded metric space (Ω, d) with probability measures $X \sim P$ and $Y \sim Q$. For any two points p_1, p_2 in the manifold $\mathcal{M}$, the half-space (Chau et al., 2019) is defined as

$$
H_{p_1,p_2} = \Big\{ y \in \mathcal{M} \mid g_{\mathcal{M}}(y, p_1) \leq g_{\mathcal{M}}(y, p_2) \Big\} \ .
$$

which is anchored at (p_1, p_2). In words, half-space H_{p_1,p_2} is the set of all points that are not farther from p_1 than from p_2 with respect to Riemannian metric g. Let us denote the collection of all half-spaces by $\mathcal{H} = \{H_{p_1,p_2} \mid p_1 \neq p_2 \in \mathcal{M}\}$. The half-space depth at $\omega \in \mathcal{M}$ with respect to theoretical distribution is defined as

$$\mathrm{HD}(P, \omega) = \inf_{\substack{p_1,p_2 \in \mathcal{M} \\ g_{\mathcal{M}}(p_1,\omega) \leq g_{\mathcal{M}}(p_2,\omega)}} \int_{\mathcal{M}} \mathbb{I}[g_{\mathcal{M}}(p_1, x) \leq g_{\mathcal{M}}(p_2, x)] P(d(x)) \ ,$$

$$\mathrm{HD}(Q, \omega) = \inf_{\substack{p_1,p_2 \in \mathcal{M} \\ g_{\mathcal{M}}(p_1,\omega) \leq g_{\mathcal{M}}(p_2,\omega)}} \int_{\mathcal{M}} \mathbb{I}[g_{\mathcal{M}}(p_1, y) \leq g_{\mathcal{M}}(p_2, y)] Q(d(y)) \ ,$$

to measure the centrality of the point ω. Also, we define the empirical version of the HD as follows

$$\mathrm{HD}(P_n, \omega) = \inf_{\substack{p_1,p_2 \in \mathcal{M} \\ g_{\mathcal{M}}(p_1,\omega) \leq g_{\mathcal{M}}(p_2,\omega)}} n^{-1} \sum_{i=1}^{n} \mathbb{I}[g_{\mathcal{M}}(p_1, y_i) \leq g_{\mathcal{M}}(p_2, y_i)] \ ,$$

$$\mathrm{HD}(Q_m, \omega) = \inf_{\substack{p_1,p_2 \in \mathcal{M} \\ g_{\mathcal{M}}(p_1,\omega) \leq g_{\mathcal{M}}(p_2,\omega)}} m^{-1} \sum_{i=1}^{m} \mathbb{I}[g_{\mathcal{M}}(p_1, y_i) \leq g_{\mathcal{M}}(p_2, y_i)] \ .$$

Let $\mathcal{C}$ be a class of sets and $\mathcal{S}$ be an arbitrary set. We define

$$\mathcal{C} \ \cap \ \mathcal{S} = \{c \ \cap \mathcal{S} \mid c \in \mathcal{C}\} \ .$$

The set $\mathcal{S}$ is shattered by $\mathcal{C}$ if $\mathcal{C} \ \cap \ \mathcal{S}$ includes all subsets of $\mathcal{S}$, i.e. $\mid \mathcal{C} \ \cap \ \mathcal{S} \mid = 2^{|\mathcal{S}|}$. Vapnik-Chervonenkis (VC) dimension of a set $\mathcal{C}$ is the smallest number of n such that no set of size n can be shattered by $\mathcal{C}$. This can be defined as

$$\mathrm{VC}(\mathcal{C}) = \inf\{n \mid \max_{x_1,...,x_n} \Delta_n(\mathcal{C}, x_1, ..., x_n) < 2^n\} \ ,$$

where $\Delta_n(\mathcal{C}, x_1, ..., x_n) = \mid \{c \ \cap \ \{x_1, ..., x_n\} \mid c \in \mathcal{C}\} \mid$, i.e. the number of subsets of $\{x_1, ..., x_n\}$ picked out by set $\mathcal{C}$.

Consider the separable and complete Riemannian manifold $(\mathcal{M}, g)$. $\mathcal{M}$ is a connected manifold with Riemannian metric g, equipped with metric space structure. We assume that for any two points p_1 and p_2, there is a unique geodesic equal to the infimum of lengths of curves connecting two points p_1 and p_2 with probability 1 w.r.t. tensorial probability measure $d\zeta(p_1, p_2) = f(p_1)f(p_2)d\nu^*(p_1)d\nu^*(p_2)$. For any pair $p_1, p_2 \in \mathcal{M}$ we denote this unique geodesic by $\overline{p_1 p_2}$. If $\mathcal{I} \in \mathcal{M}$ is compact set, then the family of half-spaces $\mathcal{H}_{p_1,p_2} =$

$\{H_{p_1,p_2} \mid p_1, p_2 \in \mathcal{I}\}$ is a Glivenko-Cantelli class in $\mathcal{I}$, since the class of half-spaces is Vapnik-Chervonenkis (Devroye et al., 1996; Dudley, 1978; Niven, 2009). In fact, for Euclidian space $\mathbb{R}^d$, VC-dimension of half-space is $d+2$. For the manifold version of half-space depth, given that the depth is defined by two points, it is reasonable to expect $\mathrm{VC}(\mathcal{H})$ is finite. For example, for sphere $\mathcal{M} = \mathcal{S}^d$, $\mathrm{VC}(\mathcal{H}) < d+3$, and for the space of phylogenetic trees $\mathcal{M} = \mathcal{T}_3$ with three leaves, $\mathrm{VC}(\mathcal{H}) = 4$ (Dai et al., 2022).

Let us define the set of half-spaces containing $\omega \in \mathcal{M}$ by

$$\mathcal{H}(\omega) = \{H_{p_1,p_2} \in \mathcal{H}_{p_1,p_2} \mid \omega \in H_{p_1,p_2}\} \ .$$

Consider a random object $X : \Omega \to \mathcal{M}$ with density f_X with respect to measure ν on Riemannian manifold $(\mathcal{M}, g)$. Let H_{p_1,p_2} be a half-space such that $P_X(H_{p_1,p_2}) = P_H(p_1,p_2)\mathbb{I}_{H_{p_1,p_2}}(\omega)$. We have

$$\begin{aligned} HD^*(\omega) &= \inf_{p_1,p_2;\omega\in H_{p_1,p_2}} P_H(p_1,p_2) = \inf_{p_1,p_2\in\mathcal{M}} \mathbb{I}_{H_{p_1,p_2}}(\omega) P_H(p_1,p_2) \\ &= \inf_{p_1,p_2\in\mathcal{M}} \int_{\mathcal{M}\times\mathcal{M}} \mathcal{K}^*_{q_1,q_2}(p_1,p_2,\omega) d\zeta(q_1,q_2) \\ &= \inf_{p_1,p_2\in\mathcal{M}} \int_{\mathcal{M}\times\mathcal{M}} \mathcal{K}^*_{q_1,q_2}(p_1,p_2,\omega) f(q_1) f(q_2) d\nu^*(q_1) d\nu^*(q_2) \ , \end{aligned}$$

where $\mathcal{K}^*$ is a kernel function.

For IC random objects $q_1, ..., q_n$, the empirical version of $HD^*(\omega)$ can be defined as

$$\widehat{HD}^*_n(\omega) = \inf_{p_1,p_2\in\mathcal{M}} U^2_n(\omega) = \inf_{p_1,p_2\in\mathcal{M}} \frac{1}{\binom{n}{2}} \sum_{1\le i\le j\le n} \mathcal{K}^*_{q_i,q_j}(p_1,p_2,\omega) \ ,$$

where U^2_n is a U-statistic of order 2.

The manifold $\mathcal{M}$ belongs to VC class if for the family of sets $\mathcal{C} = \{C_{p_1,p_2} \mid p_1, p_2 \in \mathcal{M}\}$, with

$$C_{p_1,p_2} = \{(x,y) \in \mathcal{M}\times\mathcal{M} \mid xy/2 \in H_{p_1,p_2}\} \ ,$$

it holds that $\mathcal{C}$ has a finite VC-dimension. Note that $xy/2$ represents the midpoint of geodesic $\overline{xy}$.

To show that the control charts are distribution free, we first need to show that the empirical depth converges almost surely and uniformly for all points $\omega \in \mathcal{M}$ to theoretical depth, i.e.

$$\sup_{\omega\in\mathcal{M}} \mid \widehat{HD}^*_n(\omega) - HD^*(\omega) \mid \to 0 \ , \quad a.s.$$

In a seminal paper by Fraiman et al. (2019), a framework was proposed for proving strong uniform consistency of empirical depth function defined by connecting pairwise geodesic spheres. We extend these steps to prove the above result for the proposed HD^* half-space depth function. The proof for strong uniform consistency can be accomplished by showing the vanishing at infinity property for HD^*, and the continuity property for HD^*.

$(*)$ Vanishing at infinity: For compact manifolds the proof is straightforward. We can choose δ such that for $g_{\mathcal{M}}(\omega, q) > \delta$,

$$HD^*(\omega) = \inf_{p_1,p_2;\omega\in H_{p_1,p_2}} P_X(H_{p_1,p_2}) < P\big(g_{\mathcal{M}}(q, X) > \delta\big) \to 0 \ ,$$

as $\delta \to \infty$. For noncompact manifold spaces, let p_2 be a fixed point on $\mathcal{M}$ and $p_{1,n}$ be a sequence of points on $\mathcal{M}$ such that $\delta_n = g_{\mathcal{M}}(p_{1,n}, p_2) \to \infty$ as $n \to \infty$. Also, let ω_n be a sequence of points on $\mathcal{M}$ such that $g_{\mathcal{M}}(\omega_n, p_{1,n}) \to 0$, as $n \to \infty$. For any $x \in \mathrm{B}(p_2, \delta_n/2)$, a ball centered at p_2 with radius $\delta_n/2$, it holds that

$$g_{\mathcal{M}}(p_{1,n}, x) \geq g_{\mathcal{M}}(p_{1,n}, p_2) - g_{\mathcal{M}}(p_2, x) > \delta_n - \frac{\delta_n}{2} = \frac{\delta_n}{2} > g_{\mathcal{M}}(p_2, x) \ .$$

This implies that $\mathrm{B}(p_2, \delta_n/2) \in H^C_{p_{1,n},p_2}$. Let $X \sim P$, it can be shown that

$$\begin{aligned} HD(P, \omega_n) = HD(P, p_{1,n}) &= P_X(\mathcal{H}_{p_{1,n},p_2}) = 1 - P_X(\mathcal{H}^C_{p_{1,n},p_2}) \\ &\leq 1 - P_X\Big(\mathrm{B}(p_2, \delta_n/2)\Big) \to 0 \ , \ n \to \infty \ , \end{aligned}$$

given that the probability measure P is tight on complete separable metric space $\mathcal{M}$. Therefore, for fixed $q \in \mathcal{M}$ and $\delta \to \infty$, it holds that

$$\sup_{g_{\mathcal{M}}(\omega,q)>\delta} HD^*(\omega) \to 0 \ .$$

Assume that for any $\delta > 0$ there exists an event $\Omega_1 \subset \Omega$ with nonzero probability, i.e. $P(\Omega_1) > 0$, for which we can find $\gamma > 0$ and $\omega \in \mathcal{M}$ satisfying $\widehat{HD}^*_n(\omega) > \gamma > 0$ for $g_{\mathcal{M}}(\omega, q) > \delta$ for Ω_1. If we choose δ_0 such that $P\big(g_{\mathcal{M}}(q, X) > \delta_0\big) < P(\Omega_1)/2n$. Given that $\Omega_1 \subset \bigcup_{i=1}^n \{g_{\mathcal{M}}(q, X_i) > \delta_0\}$, we have $P(\Omega_1) < nP\big(g_{\mathcal{M}}(q, X) > \delta_0\big) < P(\Omega_1)/2$, which is a contradiction. Similar steps as above can be followed to show the vanishing property for $HD^*(\omega)$ on noncompact manifold $\mathcal{M}$. Therefore, for fixed $q \in \mathcal{M}$ and $\delta \to \infty$, it holds that

$$\lim_{\delta\to\infty} \sup_{g_{\mathcal{M}}(\omega,q)>\delta} \widehat{HD}^*_n(\omega) = 0 \ , \ \text{a.s.}$$

$(*)$ Continuity: Given two points ω, η on $\mathcal{M}$ such that $g_{\mathcal{M}}(\omega, \eta) \to 0$, it can be shown that

$$\begin{aligned}
\Big|HD^*(\omega) - HD^*(\eta)\Big| &= \Big| \inf_{p_1,p_2;\omega\in H_{p_1,p_2}} P_H(p_1,p_2) - \inf_{p_1',p_2';\eta\in H_{p_1',p_2'}} P_H(p_1',p_2') \Big| \\
&= \inf_{\substack{p_1,p_2\in\mathcal{M}\\ p_1',p_2'\in\mathcal{M}}} \left| \int_{\mathcal{M}\times\mathcal{M}} \big(\mathcal{K}^*_{q_1,q_2}(p_1,p_2,\omega) - \mathcal{K}^*_{q_1,q_2}(p_1',p_2',\eta)\big) f(q_1)f(q_2)d\nu^*(q_1)d\nu^*(q_2) \right| \\
&= \inf_{\substack{p_1,p_2\in\mathcal{M}\\ p_1',p_2'\in\mathcal{M}}} \left| \int_{\mathcal{M}\times\mathcal{M}} \big(\mathcal{K}^*_{q_1,q_2}(p_1,p_2,\omega)\big)\big(1 - \mathcal{K}^*_{q_1,q_2}(p_1',p_2',\eta)\big) f(q_1)f(q_2)d\nu^*(q_1)d\nu^*(q_2) \right. \\
&\qquad \left. - \int_{\mathcal{M}\times\mathcal{M}} \big(\mathcal{K}^*_{q_1,q_2}(p_1',p_2',\eta)\big)\big(1 - \mathcal{K}^*_{q_1,q_2}(p_1,p_2,\omega)\big) f(q_1)f(q_2)d\nu^*(q_1)d\nu^*(q_2) \right| \\
&= \Big| \inf_{\substack{p_1,p_2;\omega\in H_{p_1,p_2}\\ p_1',p_2';\eta\in H_{p_1',p_2'}}} P_H\left(H_{p_1,p_2}, H^c_{p_1',p_2'}\right) - \inf_{\substack{p_1,p_2;\omega\in H_{p_1,p_2}\\ p_1',p_2';\eta\in H_{p_1',p_2'}}} P_H\left(H^c_{p_1,p_2}, H_{p_1',p_2'}\right) \Big| \\
&\le \max_{\substack{p_1,p_2;\omega\in H_{p_1,p_2}\\ p_1',p_2';\eta\in H_{p_1',p_2'}}} \left\{ P_H\left(H_{p_1,p_2}, H^c_{p_1',p_2'}\right), P_H\left(H^c_{p_1,p_2}, H_{p_1',p_2'}\right) \right\} \to 0 \ .
\end{aligned}$$

Therefore, it holds that

$$\Big|HD^*(\omega) - HD^*(\eta)\Big| \to 0 \ \ , \ as \ \ g_{\mathcal{M}}(\omega,\eta) \to 0 \ .$$

Let $r^* = \mathrm{diam}(\mathcal{M})$ and denote the symmetric set difference of two sets A and B by $A\Delta B$, it holds that

$$\begin{aligned}
\Big|\widehat{HD}^*_n(\omega) - \widehat{HD}^*_n(\eta)\Big| &= \inf_{\substack{p_1,p_2\in\mathcal{M}\\ p_1',p_2'\in\mathcal{M}}} \left| \frac{1}{\binom{n}{2}} \sum_{1\le i\le j\le n} \mathcal{K}^*_{q_i,q_j}(p_1,p_2,\omega) - \mathcal{K}^*_{q_i,q_j}(p_1',p_2',\eta) \right| \\
&\le \inf_{\substack{p_1,p_2\in\mathcal{M}\\ p_1',p_2'\in\mathcal{M}}} \frac{1}{\binom{n}{2}} \sum_{1\le i\le j\le n} \left| \mathcal{K}^*_{q_i,q_j}(p_1,p_2,\omega) - \mathcal{K}^*_{q_i,q_j}(p_1',p_2',\eta) \right| \\
&\le \frac{1}{\binom{n}{2}} \sum_{1\le i\le j\le n} \left| \mathcal{K}_{H_{\omega,q_i}\Delta H_{\eta,q_j}}\left(\frac{q_iq_j}{2}\right) \right| ,
\end{aligned}$$

where $\mathcal{K}_{\omega,\eta}\big(\frac{xy}{2}\big)$ is the kernel of the U-statistic of order 2 and $\frac{q_iq_j}{2}$ is the midpoint of geodesic

path connecting points q_i and q_j. Let $q^*_{ij} = \frac{p_i p_j}{2}$, It can be shown that

$$\begin{aligned}\left|\widehat{HD}^*_n(\omega) - \widehat{HD}^*_n(\eta)\right| &\leq \frac{1}{\binom{n}{2}} \sum_{1\leq i\leq j\leq n} \left|\mathcal{K}_{H_{\omega,q_i}\Delta H_{\eta,q_j}}\left(q^*_{ij}\right)\right| \\ &= \frac{1}{\binom{n}{2}} \sum_{1\leq i\leq j\leq n} \left|\mathcal{K}_{H_{\omega,q_i}\Delta H_{\eta,q_j}}\left(q^*_{ij}\right) - \mathbb{E}\Big[\mathcal{K}_{H_{\omega,q_i}\Delta H_{\eta,q_j}}\left(q^*_{ij}\right)\Big] + \mathbb{E}\Big[\mathcal{K}_{H_{\omega,q_i}\Delta H_{\eta,q_j}}\left(q^*_{ij}\right)\Big]\right| \\ &\leq \frac{1}{\binom{n}{2}} \sum_{1\leq i\leq j\leq n} \left|\mathcal{K}_{H_{\omega,q_i}\Delta H_{\eta,q_j}}\left(q^*_{ij}\right) - \mathbb{E}\Big[\mathcal{K}_{H_{\omega,q_i}\Delta H_{\eta,q_j}}\left(q^*_{ij}\right)\Big]\right| + \mathbb{E}\Big[\mathcal{K}_{H_{\omega,q_i}\Delta H_{\eta,q_j}}\left(q^*_{ij}\right)\Big] \ .\end{aligned}$$

Given that the family of sets $\mathcal{C}^*_{p_1,p_2} = \{C_{p_1}\Delta C_{p_2}\}_{(p_1,p_2)\in E\times E}$ has a finite VC-dimension, as $n\to\infty$, it holds that

$$L_n = \sup_{(p_1,p_2)\in E\times E} \frac{1}{\binom{n}{2}} \sum_{1\leq i\leq j\leq n} \left|\mathcal{K}_{H_{\omega,q_i}\Delta H_{\eta,q_j}}\left(q^*_{ij}\right) - \mathbb{E}\Big[\mathcal{K}_{H_{\omega,q_i}\Delta H_{\eta,q_j}}\left(q^*_{ij}\right)\Big]\right| \to 0 \quad , \quad \text{a.s.}$$

Let $\gamma(\epsilon)$ denote a deterministic function which converges to 0 as $\epsilon\to 0$. By applying Lebesgue's dominant convergence theorem, as $\epsilon\to 0$, it holds that

$$\mathbb{E}\Big[\mathcal{K}_{H_{\omega,q_i}\Delta H_{\eta,q_j}}\left(q^*_{ij}\right)\Big] \leq 2C\nu\{H_{\omega,r^*}\Delta H_{\eta,r^*}\} = \gamma(\epsilon) \to 0 \ .$$

Therefore, for manifold space $\mathcal{M}$ satisfying VC condition and for $\epsilon > 0$, it can be shown that

$$\sup_{g_{\mathcal{M}}(\omega,\eta)<\epsilon} \left|\widehat{HD}^*_n(\omega) - \widehat{HD}^*_n(\eta)\right| < \gamma(\epsilon) + L_n \ .$$

which proves the continuity condition as $\epsilon\to 0$.

$(*)$ Strong uniform consistency: For $\epsilon > 0$, by the vanishing property of the depth function, for δ and n sufficiently large, it holds that

$$\sup_{g_{\mathcal{M}}(\omega,q)\geq\delta} \left|\widehat{HD}^*_n(\omega) - HD^*(\omega)\right| \leq \sup_{g_{\mathcal{M}}(\omega,q)\geq\delta} \left|\widehat{HD}^*_n(\omega)\right| + \sup_{g_{\mathcal{M}}(\omega,q)\geq\delta} \left|HD^*(\omega)\right| < \epsilon \ , \ \text{a.s.}$$

For the compact set $\mathrm{B}(q,\delta)$ we can choose $\delta_0 > 0$ small enough to use the continuity property of the depth function. For separable $\mathcal{M}$ there exists a finite set $S = \{\omega_1, ..., \omega_m\} \subset \mathrm{B}(q,\delta)$ so that $\mathrm{B}(q,\delta)$ can be covered by $\bigcup_{j=1}^m \mathrm{B}(\omega_j,\delta_0)$, i.e. $\mathrm{B}(q,\delta) \subset \bigcup_{j=1}^m \mathrm{B}(\omega_j,\delta_0)$. Let $\omega\in\mathrm{B}(\omega_j,\delta_0)$, it can be shown that

$$\begin{aligned}\left|\widehat{HD}^*_n(\omega) - HD^*(\omega)\right| &\leq \left|\widehat{HD}^*_n(\omega) - \widehat{HD}^*_n(\omega_0)\right| + \left|\widehat{HD}^*_n(\omega_0) - HD^*(\omega_0)\right| + \left|HD^*(\omega_0) - HD^*(\omega)\right| \\ &\leq 2\epsilon + L_n + \sup_{x\in S} \left|\widehat{HD}^*_n(x) - HD^*(x)\right| \ ,\end{aligned}$$

where we used $\gamma(\epsilon) = \epsilon$. Thus, it holds that

$$\sup_{\omega \in \mathrm{B}(q,\delta)} \left| \widehat{HD}_n^*(\omega) - HD^*(\omega) \right| \leq 2\epsilon + L_n + \sup_{x \in S} \left| \widehat{HD}_n^*(x) - HD^*(x) \right| .$$

Given that the kernel of the U-statistic of order 2, i.e. $\mathcal{K}_{\omega,\eta}(x, y)$, lies between 0 and 1 almost surely, Hoeffding's inequality can be applied as follows

$$P\Big\{ \left| \widehat{HD}_n^*(\omega) - HD^*(\omega) \right| > t \Big\} \leq \exp(-nt^2/2) .$$

where $t > 0$. Therefore, it holds that

$$P\Big\{ \sup_{x \in S} \left| \widehat{HD}_n^*(x) - HD^*(x) \right| > t \Big\} \leq \sum_{i=1}^{m} P\Big\{ \left| \widehat{HD}_n^*(\omega_i) - HD^*(\omega_i) \right| > t \Big\} \leq m \exp(-nt^2/2) ,$$

which implies $\sup_{x \in S} \left| \widehat{HD}_n^*(x) - HD^*(x) \right|$ converges almost surely to 0 by Borel-Cantelli lemma. This implies that

$$\sup_{\omega \in \mathcal{M}} \left| \widehat{HD}_n^*(\omega) - HD^*(\omega) \right| \to 0 \ , \ \text{a.s.} \ ,$$

as $n \to \infty$, which proves the strong uniform consistency of sample depth function. When the process is IC, it is well-known that the relative rank of theoretical depths asymptotically converge in distribution to uniform distribution (Liu and Singh, 1993), i.e.

$$R(HD^*(\omega)) \xrightarrow{\mathcal{D}} \text{Uniform}[0, 1] .$$

Therefore, the control charts are asymptotically distribution-free. The extension of the above results to other types of depth functions defined on manifold $\mathcal{M}$ can be done similarly.

(II): Extrinsic Embedding. Consider a metric space $(\mathcal{M}, d)$ with isometric embedding map $J : \mathcal{M} \to \mathbb{R}^D$, such that $\|J(u) - J(v)\| = d(u, v)$ for $u, v \in \mathcal{M}$. Without loss of generality, let the manifold to be $\mathrm{Sym}^+(p)$ of $p \times p$ psd matrices. We can use half-vectorization mapping to embed the data on manifold $\mathrm{Sym}^+(p)$ to ambient Euclidean space $\mathbb{R}^{\frac{p(p+1)}{2}}$. Let $U(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$ denote the spatial sign function. For this space, we can use the spatial depth

$$SD(F_{\mathbf{X}}, \mathbf{y}) = 1 - \left\| \mathbb{E}_{F_{\mathbf{X}}} \left[\frac{\mathbf{y} - \mathbf{X}}{\|\mathbf{y} - \mathbf{X}\|} \right] \right\| = 1 - \left\| \mathbb{E}_{F_{\mathbf{X}}} \big[U(\mathbf{y} - \mathbf{X}) \big] \right\| .$$

Let A_0 and b_0 denote the solution to the following transformation-retransformation equation

$$\mathbb{E}_{F_0}\left(\frac{A(\mathbf{X}-\mathbf{b})}{\|A(\mathbf{X}-\mathbf{b})\|}\right)=\mathbf{0} \quad , \quad \mathbb{E}_{F_0}\left(\frac{A(\mathbf{X}-\mathbf{b})(\mathbf{X}-\mathbf{b})^\top A^\top}{\|A(\mathbf{X}-\mathbf{b})\|^2}\right)=\frac{2}{p(p+1)}\mathbb{I}_{p(p+1)/2} \ .$$

Also, let $(\widehat{A}_n, \widehat{\mathbf{b}}_n)$ denote the solution to the sample version of transformation-retransformation equation

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{A}_n(\mathbf{X}_i-\widehat{\mathbf{b}}_n)}{\|\widehat{A}_n(\mathbf{X}_i-\widehat{\mathbf{b}}_n)\|}=\mathbf{0} \ , \quad \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{A}_n(\mathbf{X}_i-\widehat{\mathbf{b}}_n)(\mathbf{X}_i-\widehat{\mathbf{b}}_n)^\top \widehat{A}_n^\top}{\|\widehat{A}_n(\mathbf{X}_i-\widehat{\mathbf{b}}_n)\|^2}=\frac{2}{p(p+1)}\mathbb{I}_{p(p+1)/2} \ .$$

It can be shown that $\widehat{A}_n \xrightarrow{a.s.} A_0$ and $\widehat{\mathbf{b}}_n \xrightarrow{a.s.} b_0$ (Hettmansperger and Randles, 2002). Let $\mathbf{X}_1, ..., \mathbf{X}_n$ denote the Phase-I reference sample and $\mathbf{V}_i = A_0(\mathbf{X}_i - \mathbf{b_0})$. We use F_n and G_n for the empirical distribution of $\mathbf{X}_1, ..., \mathbf{X}_n$ and $\mathbf{V}_1, ..., \mathbf{V}_n$, respectively. For any $\mathbf{y} \in \mathbb{R}^{\frac{p(p+1)}{2}}$, it holds that

$$\begin{aligned}
\left|\mathrm{SD}(F_n,\mathbf{y})-\mathrm{SD}(G_n,\mathbf{y})\right| &= \left|\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{A}_n(\mathbf{y}-\mathbf{X}_i)}{\|\widehat{A}_n(\mathbf{y}-\mathbf{X}_i)\|}\right\| - \left\|\frac{1}{n}\sum_{i=1}^{n}\frac{A_0(\mathbf{y}-\mathbf{X}_i)}{\|A_0(\mathbf{y}-\mathbf{X}_i)\|}\right\|\right| \\
&\le \left\|\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{A}_n(\mathbf{y}-\mathbf{X}_i)}{\|\widehat{A}_n(\mathbf{y}-\mathbf{X}_i)\|}-\frac{1}{n}\sum_{i=1}^{n}\frac{A_0(\mathbf{y}-\mathbf{X}_i)}{\|A_0(\mathbf{y}-\mathbf{X}_i)\|}\right\| \\
&\le \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\widehat{A}_n(\mathbf{y}-\mathbf{X}_i)}{\|\widehat{A}_n(\mathbf{y}-\mathbf{X}_i)\|}-\frac{A_0(\mathbf{y}-\mathbf{X}_i)}{\|A_0(\mathbf{y}-\mathbf{X}_i)\|}\right\| \\
&\le \frac{2}{n}\sum_{i=1}^{n}\frac{\|(\widehat{A}_n-A_0)(\mathbf{y}-\mathbf{X}_i)\|}{\|A_0(\mathbf{y}-\mathbf{X}_i)\|} \\
&= \frac{2}{n}\sum_{i=1}^{n}\sqrt{\frac{(\mathbf{y}-\mathbf{X}_i)^\top(\widehat{A}_n-A_0)^\top(\widehat{A}_n-A_0)(\mathbf{y}-\mathbf{X}_i)}{(\mathbf{y}-\mathbf{X}_i)^\top A_0^\top A_0(\mathbf{y}-\mathbf{X}_i)^\top}} \\
&\le 2\sqrt{\lambda^*_{\max}\Big((A_0^\top A_0)^{-1}(\widehat{A}_n-A_0)^\top(\widehat{A}_n-A_0)\Big)} \ ,
\end{aligned}$$

where $\lambda^*_{\max}(A)$ is the largest eigenvalue of matrix A. Given that $\lambda^*_{\max}(A)$ is a bounded continuous function and $\widehat{A}_n \xrightarrow{a.s.} A_0$, the following strong uniform consistency result holds

$$\sup_{\mathbf{y}\in\mathbb{R}^{p(p+1)/2}}\left|\mathrm{SD}(F_n,\mathbf{y})-\mathrm{SD}(G_n,\mathbf{y})\right| \xrightarrow{a.s.} 0 \ .$$

Also, by the uniform convergence of sample spatial depth (Li et al., 2013), we have

$$\sup_{\mathbf{y}\in\mathbb{R}^{p(p+1)/2}} \left|\mathrm{SD}(G_n,\mathbf{y}) - \mathrm{SD}(G,\mathbf{y})\right| \xrightarrow{a.s.} 0 \ .$$

Therefore

$$\sup_{\mathbf{y}\in\mathbb{R}^{p(p+1)/2}} \left|\mathrm{SD}(F_n,\mathbf{y}) - \mathrm{SD}(G,\mathbf{y})\right| \xrightarrow{a.s.} 0 \ .$$

When the process is IC, the relative rank of theoretical depths asymptotically converge in distribution to uniform distribution (Liu and Singh, 1993), i.e.

$$R(HD^*(\omega)) \xrightarrow{\mathcal{D}} \mathrm{Uniform}[0,1] \ .$$

Therefore, the control charts are asymptotically distribution-free. This completes the proof.

□

Proof of Theorem 30. Proof can be finished following similar steps given in the proof of Theorem 10.

□

Proof of Theorem 31. The proof can be done following similar steps given in the proof of Theorem 24.

□

Proof of Theorem 32. The proof of the theorem is given in three parts. In the first part, the congruence-invariance of control charts under the notion of intrinsic isometric embedding of Riemannian manifolds is demonstrated. In the second part, the congruence-invariance of the control charts is proven under the notion of extrinsic isometric embedding of Riemannian manifolds. In the last part, the congruence-invariance of IC-ARL and OC-ARL is shown.

(I): Intrinsic Embedding. For intrinsic analysis, we focus on Riemannian manifold $(\mathcal{M}, g)$ which is a connected manifold with Riemannian metric g, equipped with metric space structure with distance function equal to the infimum of lengths of curves connecting two points x, y. For differentiable function $f : \mathcal{M} \to \mathcal{M}$ and $x \in \mathcal{M}$, the differential of f at point x, i.e. $d_x f : T_x\mathcal{M} \to T_{f(x)}\mathcal{M}$, is defined as

$$d_x f\left(\frac{d\gamma}{dt}(0)\right) = \frac{d(f\circ\gamma)}{dt}(0) \ , \ \gamma(-\epsilon,\epsilon) \to \mathcal{M} \text{ and } \gamma(0) = x \ .$$

This structure admits Riemannian isometry embedding $J : \mathcal{M} \to \mathbb{R}^D$, satisfying $g_x(\eta_1, \eta_2) = d_x J(\eta_1)^\top d_x J(\eta_2)$, for all $x \in \mathcal{M}$ and for all $\eta_1, \eta_2 \in T_x\mathcal{M}$. The $\mathcal{K}$-homogeneous space $(\mathcal{M}, \mathcal{K}, \alpha)$ is $\mathcal{K}$-invariant, i.e. $g_{\alpha_k(x)}\big((d_x\alpha_k)(\eta_1), (d_x\alpha_k)(\eta_2)\big) = g_x(\eta_1, \eta_2)$, for all $x \in \mathcal{M}$ and for all $\eta_1, \eta_2 \in T_x\mathcal{M}$ (Osborne et al., 2013).

Without loss of generality, assume that the manifold space is $\mathrm{Sym}^+(p)$ of $p \times p$ psd matrices. Then, $\mathcal{K}$ is the group of matrices of positive determinant, i.e. $\mathrm{GL}^+(p, \mathbb{R})$ and $\alpha : \mathrm{GL}^+(p, \mathbb{R}) \times \mathrm{Sym}^+(p) \to \mathrm{Sym}^+(p)$ and $\alpha_k(x) = kxk^\top$. Apparently, α is the transitive left action of $\mathrm{GL}^+(p, \mathbb{R})$ on $\mathrm{Sym}^+(p)$, i.e. for $x, y \in \mathrm{Sym}^+(p)$, there exists $\alpha_k(x_1) = x_2$. Also, the identity map is the class of special orthogonal group. For this space, if U is an open set in $\mathbb{R}^D$, the tangent space at point $x \in U$ is $T_xU = \{X\} \times U$ and the tangent bundle of U is $TU = \bigcup_{x \in U} T_xU = U \times \mathbb{R}^D$. Therefore, for $x \in \mathrm{Sym}^+(p)$, $T_x\mathrm{Sym}^+(p) = \{x\} \times \mathrm{Sym}(p)$ and $T\mathrm{Sym}^+(p) = \mathrm{Sym}^+(p) \times \mathrm{Sym}(p)$. This structure admits the Frobenius scalar product on $\mathrm{Sym}(p)$ and the Riemannian inner product g_{I_p} on $T_{I_p}(\mathrm{Sym}^+(p))$ as follows

$$g_{I_p}\Big((I_p, u), (I_p, v)\Big) = \mathrm{tr}(uv) \ ,$$

which is the unique $\mathrm{GL}^+(p, \mathbb{R})$-invariant Riemannian metric on $\mathrm{Sym}^+(p)$, i.e.

$$g_{I_p}\Big(d_{I_p}\alpha_k(I_p, u), d_{I_p}\alpha_k(I_p, v)\Big) = g_{I_p}\Big((I_p, u), (I_p, v)\Big) \\ , \quad \forall k \in \mathrm{GL}^+(p, \mathbb{R}) \ , \forall (I_p, u), (I_p, v) \in T\mathrm{Sym}^+(p) \ ,$$

where $d_x\alpha_k(x, u) = (kxk^\top, kuk^\top)$, for all $(x, u) \in TSym^+(p)$.

Let $(\Omega, \mathcal{F}, P)$ represent the probability space associated with the complete Riemannian manifold $(\mathcal{M}, g)$ and $\mathcal{B}_\mathcal{M}$ denote the Borel σ-field of open subsets of $\mathcal{M}$. Random object $X : \Omega \to \mathcal{M}$ is $(\mathcal{F}, \mathcal{B}_\mathcal{M})$-measurable with probability measure $F_X(B) = P(X^{-1}(B))$ for all $B \in \mathcal{B}_\mathcal{M}$. The intrinsic mean of set of X with respect to the Riemannian structure g can be obtained as follows

$$\begin{aligned} \mu_g &= \arg\min_{y \in \mathcal{M}} \mathbb{E}_{F_X}\Big[g^2(X, y)\Big] = \arg\min_{y \in \mathcal{M}} \int_\mathcal{M} g^2(x, y) F_X(dx) \\ &= \arg\min_{y \in \mathcal{M}} \int_\mathcal{M} g^2_{I_p}\Big((I_p, y), (I_p, x)\Big) F_X(dx) \\ &= \arg\min_{y \in \mathcal{M}} \int_\mathcal{M} \Big(\mathrm{tr}(xy)\Big)^2 F_X(dx) \ . \end{aligned}$$

Let $r = 2$, μ_g maximizes the depth function

$$D(F_X, \mu_g) = (1 + \mathbb{E}_{F_X}[g^r(\mu_g, X)])^{-1} \ .$$

Define the random object $Z = \alpha_k(X)$, with probability measure $F_{\alpha_k(X)} = F_Z$. It can be verified that the intrinsic mean μ'_g satisfies the following $\mathcal{K}$-equivariant property

$$\begin{aligned}
\mu'_g &= \arg\min_{y\in\mathcal{M}} \mathbb{E}_{F_Z}\Big[g^2(Z,y)\Big] \\
&= \arg\min_{y\in\mathcal{M}} \int_{\mathcal{M}} g^2(z,y) F_Z(dz) \ , \\
&= \arg\min_{y\in\mathcal{M}} \int_{\mathcal{M}} g^2_{I_p}\Big(d_{I_p}\alpha_k(I_p,x), d_{I_p}\alpha_k(I_p,y)\Big) F_Z(dz) \ , \\
&= \arg\min_{y\in\mathcal{M}} \int_{\mathcal{M}} g^2_{I_p}\Big((I_p,x),(I_p,y)\Big) F_Z(dz) \ , \quad \text{by invariance property} \\
&= k\Bigg[\arg\min_{y\in\mathcal{M}} \int_{\mathcal{M}} g^2_{I_p}\Big((I_p,x),(I_p,y)\Big) F_X(dx)\Bigg]k^\top \ , \quad \text{by change of variable} \\
&= k\Bigg[\arg\min_{y\in\mathcal{M}} \int_{\mathcal{M}} \Big(\mathrm{tr}(xy)\Big)^2 F_X(dx)\Bigg]k^\top \\
&= k\mu_g k^\top \ .
\end{aligned}$$

Also, for $r = 2$, μ'_g is the maximizer of the depth function

$$D(F_Z, \mu'_g) = (1 + \mathbb{E}_{F_Z}[g^r(\mu'_g, Z)])^{-1} \ .$$

By the congruence-invariant property of the depth functions, it can be easily verified that the depths of $\alpha_k(X_1), ..., \alpha_k(X_{n_1}), \alpha_k(Y_1), ..., \alpha_k(Y_{n_2})$ with respect to $F_{\alpha_k(X)}$ are the same as the depths of $X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2}$ with respect to F_X. As a result, the ranks are also the same, i.e. $R^X_t = R^{\alpha_k(X)}_t$, for $t = 1, 2, ..., n_1+n_2$. This implies that the control charts $\mathbf{S}^{EWMA}_t$, $\mathbf{S}^{CUSUM}_t$, and $\mathbf{S}^{MWMASHC}_t$ are invariant under left action $\alpha_k \in \mathcal{K}$ on random objects $X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2}$ sampled from manifold $\mathcal{M}$, and therefore congruence-invariant. This result can be extended to other classes of manifold spaces, e.g. Steifel manifold, projective spaces groups of rotations, spheres, etc., which are $\mathcal{K}$-homogeneous space.

(II): Extrinsic Embedding. Given metric space $(\mathcal{M}, d)$, consider an isometric embedding map $J : \mathcal{M} \to \mathbb{R}^D$, such that $\|J(u) - J(v)\| = d(u, v)$ for $u, v \in \mathcal{M}$. Without loss of generality, consider the manifold to be $\mathrm{Sym}^+(p)$ of $p \times p$ psd matrices. half-vectorization mapping can be used for embedding the data on manifold $\mathrm{Sym}^+(p)$ to ambient Euclidean space $\mathbb{R}^{\frac{p(p+1)}{2}}$. Let $U(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$ denote the spatial sign function. For this space, we can use the spatial depth

$$SD(F_{\mathbf{X}}, \mathbf{y}) = 1 - \left\|\mathbb{E}_{F_{\mathbf{X}}}\left[\frac{\mathbf{y} - \mathbf{X}}{\|\mathbf{y} - \mathbf{X}\|}\right]\right\| = 1 - \Big\|\mathbb{E}_{F_{\mathbf{X}}}\big[U(\mathbf{y} - \mathbf{X})\big]\Big\| \ .$$

To achieve affine-invariance when using spatial sign function, it is common to consider the transformation-retransformation suggested in Chakraborty et al. (1998), as follows

$$\mathbf{X}' = \widehat{A}_n(\mathbf{X} - \widehat{\mathbf{b}}_n) \ , \ \mathbf{Y}' = \widehat{A}_n(\mathbf{Y} - \widehat{\mathbf{b}}_n) \ ,$$

where $\widehat{A}_n$ is an upper-triangular $\frac{p(p+1)}{2} \times \frac{p(p+1)}{2}$ matrix with positive diagonal entries and $\widehat{\mathbf{b}}_n \in \mathbb{R}^{\frac{p(p+1)}{2}}$. $(\widehat{A}_n, \widehat{\mathbf{b}}_n)$ is the solution to the following system of equations

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{A}_n(\mathbf{X}_i - \widehat{\mathbf{b}}_n)}{\|\widehat{A}_n(\mathbf{X}_i - \widehat{\mathbf{b}}_n)\|} = \mathbf{0} \ , \quad \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{A}_n(\mathbf{X}_i - \widehat{\mathbf{b}}_n)(\mathbf{X}_i - \widehat{\mathbf{b}}_n)^\top \widehat{A}_n^\top}{\|\widehat{A}_n(\mathbf{X}_i - \widehat{\mathbf{b}}_n)\|^2} = \frac{2}{p(p+1)}\mathbb{I}_{p(p+1)/2} \ .$$

Let $\mathbf{X}_1, ..., \mathbf{X}_n$ denote the Phase-I reference sample and $\mathbf{Y}_1, ..., \mathbf{Y}_m$ denote the Phase-II sample. Also let $\mathbf{V}_i = R\mathbf{X}_i + \mathbf{u}$ and $\mathbf{W}_i = R\mathbf{Y}_i + \mathbf{u}$ denote the samples obtained after rotation R and translation u, where R is $\frac{p(p+1)}{2} \times \frac{p(p+1)}{2}$ nonsingular matrix and $\mathbf{u}$ is a $\frac{p(p+1)}{2}$-dimensional vector.

The control charts are affine-invariant if the statistic calculated for $\mathbf{Y}_i$'s using $(\widehat{A}_n^x, \widehat{\mathbf{b}}_n^x)$ are the same as the statistic calculated for $\mathbf{W}_i$'s using $(\widehat{A}_n^v, \widehat{\mathbf{b}}_n^v)$. It can be seen that $\widehat{\mathbf{b}}_n^v = R\widehat{\mathbf{b}}_n^x + \mathbf{u}$. Also, $\widehat{A}_n^v = \widehat{A}_n^x R^{-1}$ is a solution to rescaling problem

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{A}_n(\mathbf{V}_i - \widehat{\mathbf{b}}_n^z)(\mathbf{V}_i - \widehat{\mathbf{b}}_n^z)^\top \widehat{A}_n^\top}{\|\widehat{A}_n(\mathbf{V}_i - \widehat{\mathbf{b}}_n^z)\|^2} = \frac{2}{p(p+1)}\mathbb{I}_{p(p+1)/2} \ .$$

Given the fact that the above equation has a unique solution up to a positive multiplication term, we can define $\widehat{A}_n^{v,*} = \kappa \widehat{A}_n^x R^{-1}$ as the general solution to rescaling problem.

This can be verified with a little algebra as follows

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{A}_n^{v,*}(\mathbf{V}_i - \widehat{\mathbf{b}}_n^z)(\mathbf{V}_i - \widehat{\mathbf{b}}_n^z)^\top (\widehat{A}_n^{v,*})^\top}{\|\widehat{A}_n^{v,*}(\mathbf{V}_i - \widehat{\mathbf{b}}_n^z)\|^2} \\
&\quad = \frac{1}{n}\sum_{i=1}^{n} \frac{\kappa \widehat{A}_n^x R^{-1}(R\mathbf{X}_i + \mathbf{u} - R\widehat{\mathbf{b}}_n^x - \mathbf{u})(R\mathbf{X}_i + \mathbf{u} - R\widehat{\mathbf{b}}_n^x - \mathbf{u})^\top (\kappa \widehat{A}_n^x R^{-1})^\top}{\|\kappa \widehat{A}_n^x R^{-1}(R\mathbf{X}_i + \mathbf{u} - R\widehat{\mathbf{b}}_n^x - \mathbf{u})\|^2} \\
&\quad = \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{A}_n^x(\mathbf{X}_i - \widehat{\mathbf{b}}_n^x)(\mathbf{X}_i - \widehat{\mathbf{b}}_n^x)^\top (\widehat{A}_n^x)^\top}{\|\widehat{A}_n^x(\mathbf{X}_i - \widehat{\mathbf{b}}_n^x)\|^2} = \frac{2}{p(p+1)}\mathbb{I}_{p(p+1)/2} \ .
\end{aligned}$$

Let $\Theta = \kappa^{-1}\widehat{A}_n^v R(\widehat{A}_n^x)^{-1}$, it can be verified that

$$\widehat{A}_n^{v,*}(\mathbf{V}_i - \widehat{\mathbf{b}}_n^z) = \widehat{A}_n^{v,*} R(\mathbf{X}_i - \widehat{\mathbf{b}}_n^x) = \kappa\Theta\widehat{A}_n^x(\mathbf{X}_i - \widehat{\mathbf{b}}_n^x) \ , \ i = 1, ..., n \ ,$$

and thus, we have the following property for the spatial sign function

$$U_{\mathbf{V}_i} = \frac{\widehat{A}_n^{v,*}(\mathbf{V}_i - \widehat{\mathbf{b}}_n^z)}{\|\widehat{A}_n^{v,*}(\mathbf{V}_i - \widehat{\mathbf{b}}_n^z)\|} = \Theta \frac{\widehat{A}_n^{x}(\mathbf{X}_i - \widehat{\mathbf{b}}_n^z)}{\|\widehat{A}_n^{x}(\mathbf{X}_i - \widehat{\mathbf{b}}_n^z)\|} = \Theta U_{\mathbf{X}_i} \ .$$

Given the fact that Θ is orthogonal, any orthogonally invariant statistic computed on spatial sign function will be affine-invariant. Let $F_{\mathbf{X}}^*$ and $G_{\mathbf{V}}^*$ denote the empirical distribution for $\mathbf{X}_1, ..., \mathbf{X}_n$ and $\mathbf{V}_1, ..., \mathbf{V}_n$, respectively. It can be shown that the spatial depth function satisfies the following

$$\begin{aligned} SD(G_{\mathbf{V}}^*, \mathbf{W}_j) &= 1 - \Big\| \mathbb{E}_{G_{\mathbf{V}}^*}\big[U(\mathbf{W}_j - \mathbf{V})\big] \Big\| = 1 - \left\| \frac{1}{n} \sum_{i=1}^n \frac{\widehat{A}_n^{v}(\mathbf{W}_j - \mathbf{V}_i)}{\|\widehat{A}_n^{v}(\mathbf{W}_j - \mathbf{V}_i)\|} \right\| \\ &= 1 - \left\| \frac{1}{n} \sum_{i=1}^n \frac{\Theta \widehat{A}_n^{x}(\mathbf{Y}_j - \mathbf{X}_i)}{\|\widehat{A}_n^{x}(\mathbf{Y}_j - \mathbf{X}_i)\|} \right\| = 1 - \Big\| \mathbb{E}_{F_{\mathbf{X}}^*}\big[U(\mathbf{Y}_j - \mathbf{X})\big] \Big\| = SD(F_{\mathbf{X}}^*, \mathbf{Y}_j) \ . \end{aligned}$$

As a result, the ranks are also the same, i.e. $R_t^X = R_t^v$, for $t = 1, 2, ..., n_1 + n_2$. This implies that the control charts $\mathbf{S}_t^{EWMA}$, $\mathbf{S}_t^{CUSUM}$, and $\mathbf{S}_t^{MWMASHC}$ are affine-invariant under the extrinsic transformation, and thus, the results are congruence-invariant under action $\alpha_k \in \mathcal{K}$ on random objects.

(III): IC-ARL. Given the fact that the control charts are congruence-invariant under action $\alpha_k \in \mathcal{K}$ on random objects, the following holds for IC-ARL and OC-ARL

$$\begin{aligned} &\mathbb{E}_{F_X}\big[\min\{t \mid \mathbf{S}_t(X) > \text{UCL or } \mathbf{S}_t(X) < \text{LCL}\}\big] \\ &\quad = \mathbb{E}_{F_{\alpha_k(X)}}\big[\min\{t \mid \mathbf{S}_t(\alpha_k(X)) > \text{UCL or } \mathbf{S}_t(\alpha_k(X)) < \text{LCL}\}\big] \ . \end{aligned}$$

This implies the congruence-invariant property of IC-ARL and OC-ARL, which completes the proof. □

Proof of Theorem 33. Let us denote the sample version of probability measure on manifold $\mathcal{M}$ by $\widehat{P}_n$. Also, let us denote the control limit by $CL(\alpha)$.

The proof can be done by induction. By the definition, we know $P\big\{\widehat{\mathbf{S}}_1 < CL(\alpha) \mid \widehat{P}_1\big\} = 1 - \alpha$. Let us assume that

$$P\Big\{\widehat{\mathbf{S}}_1 < CL(\alpha), ..., \widehat{\mathbf{S}}_{k-1} < CL(\alpha)\Big|\widehat{P}_1, ..., \widehat{P}_{k-1}\Big\} = (1-\alpha)^{k-1} \ .$$

Given that $\widehat{\mathbf{S}}_i$, $i = 1, ..., k-1$, are independent of R_k, we know that they are independent from $\widehat{P}_k$ condition on $\widehat{P}_i$, $i = 1, ..., k-1$. Also, the distribution of $\widehat{\mathbf{S}}_k$ is independent of $\widehat{P}_i$, $i = 1, ..., k-1$, condition on $\widehat{P}_k$.

We have that

$$
\begin{aligned}
P\Big\{\widehat{\mathbf{S}}_i < CL(\alpha), i = 1, ..., k \Big| \widehat{P}_1, ..., \widehat{P}_k\Big\} &= P\Big\{\widehat{\mathbf{S}}_k < CL(\alpha), \widehat{\mathbf{S}}_i < CL(\alpha), i = 1, ..., k-1 \Big| \widehat{P}_1, ..., \widehat{P}_k\Big\} \\
&= P\Big\{\widehat{\mathbf{S}}_k < CL(\alpha) \Big| \widehat{\mathbf{S}}_i < CL(\alpha), i = 1, ..., k-1, \widehat{P}_k\Big\} \\
&\quad P\Big\{\widehat{\mathbf{S}}_i < CL(\alpha), i = 1, ..., k-1 \Big| \widehat{P}_1, ..., \widehat{P}_{k-1}\Big\} \\
&= P\Big\{\widehat{\mathbf{S}}_k < CL(\alpha) \Big| \widehat{P}_k\Big\} \\
&\quad P\Big\{\widehat{\mathbf{S}}_i < CL(\alpha), i = 1, ..., k-1 \Big| \widehat{P}_1, ..., \widehat{P}_{k-1}\Big\} \\
&= (1-\alpha)(1-\alpha)^{k-1} = (1-\alpha)^k \ ,
\end{aligned}
$$

which completes the induction proof.

Similarly, we can show the following result for the distribution of RL

$$
\begin{aligned}
P\{\mathrm{RL} = n\} &= \mathbb{E}\big[\mathbb{I}_{\mathrm{RL}=n}\big] = \mathbb{E}\Big[\mathbb{E}\Big[\mathbb{I}_{\mathrm{RL}=n}\Big| \widehat{P}_1, ..., \widehat{P}_n\Big]\Big] \\
&= \mathbb{E}\Big[P\Big\{\mathrm{RL} = n\Big| \widehat{P}_1, ..., \widehat{P}_n\Big\}\Big] \\
&= \mathbb{E}\Big[P\Big\{\widehat{\mathbf{S}}_n > CL(\alpha), \widehat{\mathbf{S}}_i < CL(\alpha), i = 1, ..., n-1\Big\}\Big| \widehat{P}_1, ..., \widehat{P}_n\Big] \\
&= \mathbb{E}\Big[P\Big\{\widehat{\mathbf{S}}_n > CL(\alpha)\Big| \widehat{\mathbf{S}}_i < CL(\alpha), i = 1, ..., n-1, \widehat{P}_n\Big\} \\
&\qquad P\Big\{\widehat{\mathbf{S}}_i < CL(\alpha), i = 1, ..., n-1\Big| \widehat{P}_1, ..., \widehat{P}_{n-1}\Big\}\Big] \\
&= \mathbb{E}\Big[P\Big\{\widehat{\mathbf{S}}_n > CL(\alpha)\Big| \widehat{P}_n\Big\} P\Big\{\widehat{\mathbf{S}}_i < CL(\alpha), i = 1, ..., n-1\Big| \widehat{P}_1, ..., \widehat{P}_{n-1}\Big\}\Big] \\
&= \alpha(1-\alpha)^{n-1} \ .
\end{aligned}
$$

This completes the proof.

□

Proof of Theorem 34. Let $K_h(T_{ij} - t)$ denote a symmetric kernel function with bandwidth $h > 0$, evaluated at time t, and T_{ij} are times of sampling from the underlying stochastic process with $i = 1, \ldots, N$ samples and $j = 1, \ldots, m_i$ random time-points. We consider a kernel estimate $\widehat{\Gamma}(s,t) = \widehat{A}_0$ proposed by Dai and Müller (2018) which is the

solution to the following problem

$$\left(\widehat{A}_0, \widehat{A}_1, \widehat{A}_2\right) = \arg\min_{A_0,A_1,A_2} \sum_{i=1}^{N} v_i \sum_{1\le j\ne l\le m_i} K_h(T_{ij}-s)K_h(T_{il}-t) \times \|\widehat{\Gamma}_{ijl} - A_0 - (T_{ij}-s)A_1 - (T_{il}-t)A_2\|_F^2 \ ,$$

where $\widehat{\Gamma}_{ijl} = \log_{\widehat{\mu}(T_{ij})}(X_{ij}) \log_{\widehat{\mu}(T_{il})}(X_{il})^\top$ and v_i's are weights such that $v_i = 1/\sum_{i=1}^{N} m_i(m_i-1)$, where m_i is the number of sampling time-points for i^{th} observation from stochastic process defined on compact time interval $\mathcal{T}$. Dai and Müller (2018) formulated the estimate $\widehat{\Gamma}(s,t)$ in terms of S_{ab} and $\widehat{E}_{ab}$

$$S_{ab} = \sum_{i=1}^{N} v_i \sum_{1\le j\ne l\le m_i} K_h(T_{ij}-s)K_h(T_{il}-t)\left(\frac{T_{ij}-s}{h}\right)^a \left(\frac{T_{il}-t}{h}\right)^b \ ,$$

$$\widehat{E}_{ab} = \sum_{i=1}^{N} v_i \sum_{1\le j\ne l\le m_i} K_h(T_{ij}-s)K_h(T_{il}-t)\left(\frac{T_{ij}-s}{h}\right)^a \left(\frac{T_{il}-t}{h}\right)^b \widehat{\Gamma}_{ijl} \ ,$$

for $a, b = 0, 1, 2$. We have

$$\|E_{ab} - \widehat{E}_{ab}\|_F^2 \le c_0 \sup_{t\in\mathcal{T}} d_{\mathcal{M}}^2(\widehat{\mu}(t) - \mu(t)) = o_p(1) \ .$$

Therefore

$$\sup_{s,t\in\mathcal{T}} \|\widehat{\Gamma}(s,t) - \Gamma(s,t)\|_F^2 = o_p(1) \ .$$

This completes the proof.

□

Proof of Theorem 35. Under assumptions 1-3, by following steps similar to Theorem 5, it can be shown that $\sup_{x\in\mathcal{M}} \mid D(P_n,x) - D(P,x) \mid = O_p\left(n^{-1/2}(d\log d)^{1/2}\right)$.

r-EWMA: Let $a(i) = \frac{1}{n}R_i$ be the rank score. For r-EWMA, we have

$$\mathbf{S}_t^{EWMA} = (1-\lambda)\mathbf{S}_{t-1}^{EWMA} + \lambda a(t) \ , \quad t = 1, 2, ..., n_1+n_2 \ ,$$

where $\mathbf{S}_0^{EWMA} = 0$ and $\lambda \in (0, 1]$, we have

$$\begin{aligned}\mathbf{S}_t &= (1-\lambda)^{t-1}\lambda a(1) + (1-\lambda)^{t-2}\lambda a(2) + \ldots + \lambda a(t) \\ &= \sum_{i=0}^{t-1}(1-\lambda)^i \lambda a(t-i) \ .\end{aligned}$$

Let $t = n$, it holds that

$$\begin{aligned}\mathbb{E}[\mathbf{S}_n^{EWMA}] &= \sum_{i=0}^{n-1}(1-\lambda)^i\lambda\mathbb{E}[a(n-i)] = \lambda\frac{1-(1-\lambda)^n}{1-(1-\lambda)}\mathbb{E}[a(i)] \\ &= (1-(1-\lambda)^n)\bar{a}_n \ .\end{aligned}$$

$$\begin{aligned}\mathrm{Var}[\mathbf{S}_n^{EWMA}] &= \sum_{i=0}^{n-1}(1-\lambda)^{2i}\lambda^2\mathrm{Var}[a(n-i)] + 2\sum_{i=0}^{n-1}\sum_{j=i}^{n-1}(1-\lambda)^i(1-\lambda)^j\lambda^2\mathrm{Cov}(a(i),a(j)) \\ &= \lambda^2\frac{1-(1-\lambda)^{2n}}{1-(1-\lambda)^2}\sigma_n^2 + 2\sum_{i=0}^{n-1}\sum_{j=i}^{n-1}(1-\lambda)^{i+j}\lambda^2\mathbb{E}\left[a(i)a(j) - a(i)\bar{a}_n - a(j)\bar{a}_n + \bar{a}_n^2\right] \ .\end{aligned}$$

From Theorem 34, we have

$$|\mathbb{E}\left[a(i)a(j)\right] - \mathbb{E}\left[\widehat{a}(i)\widehat{a}(j)\right]| = O_p\left(n^{-1}(d\log d)\right) \ .$$

From Theorem 5, we have $|\ \widehat{a}(i) - a(i)\ | = O_p\left(n^{-1/2}(d\log d)^{1/2}\right)$.

Let $\mathbf{S}_n^* = \dfrac{\mathbf{S}_n - \mathbb{E}[\mathbf{S}_n]}{\sqrt{\mathrm{Var}[\mathbf{S}_n]}}$. By following similar steps to Theorem 5, we get

$$|\ \mathbf{S}_n^* - \widehat{\mathbf{S}}_n^*\ | \leq \max_{i=1,\ldots,n}|\ \widehat{a}(i) - a(i)\ | = O_p\left(n^{-1/2}(d\log d)^{1/2}\right) \ .$$

Similarly, we get $|\ \mathbf{\Lambda}_n^* - \widehat{\mathbf{\Lambda}}_n^*\ | = O_p\left(n^{-1}(d\log d)\right)$.

r-CUSUM: Let $a(i) = \frac{1}{n}R_i$ be the rank score. For r-CUSUM, using the representation given in Theorem 28 and following similar steps as above, we get

$$|\ \mathbf{S}_n^* - \widehat{\mathbf{S}}_n^*\ | \leq \max_{i=1,\ldots,n}|\ \widehat{a}(i) - a(i)\ | = O_p\left(n^{-1/2}(d\log d)^{1/2}\right) \ .$$

$$|\ \mathbf{\Lambda}_n^* - \widehat{\mathbf{\Lambda}}_n^*\ | = O_p\left(n^{-1}(d\log d)\right) \ .$$

r-MWMASHC: Let $a_{ij} = \mathbb{I}\Big(\mathbf{R}_{g_t}(j) > R_i\Big)$ be the rank score. For r-MWMASHC, using the representation given in Theorem 28 and following similar steps as above, we get

$$\mid \mathbf{S}_n^* - \widehat{\mathbf{S}}_n^* \mid \leq \max_{i=1,\ldots,n} \mid \widehat{a}(i) - a(i) \mid = O_p\left(n^{-1/2}(d\log d)^{1/2}\right) \ .$$

$$\mid \mathbf{\Lambda}_n^* - \widehat{\mathbf{\Lambda}}_n^* \mid = O_p\left(n^{-1}(d\log d)\right) \ .$$

This completes the proof.

□