

MANIFOLD-ADAPTED MODELS FOR TIME SERIES OF
COVARIANCE AND CORRELATION MATRICES WITH
APPLICATIONS TO ELECTROENCEPHALOGRAPHY DATA

TAO DING

Thesis submitted for the degree of
Doctor of Philosophy



*School of Mathematics, Statistics & Physics
Newcastle University
Newcastle upon Tyne
United Kingdom*

December 2023

To my beloved mother, may you find eternal peace in heaven.

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor, Dr. Tom Nye, for guiding me through the main period of my Ph.D. journey. His meticulous approach to scientific research and unique thinking perspective have been invaluable to me. Without his continual encouragement and support, the completion of this thesis would not have been possible. I would also like to extend my thanks to Professor JianQing Shi from the Southern University of Science and Technology in China. Despite his departure from Newcastle University during my first year of Ph.D., his influence in instilling good research habits has played a crucial role in my academic development. Simultaneously, I am deeply grateful for the opportunity he provided me to pursue a Ph.D. in the UK. Furthermore, I am very thankful to Dr. Shirley Coleman for offering me the chance to collaborate on a short-term project. Her guidance has been enlightening, showing me how to apply the gained knowledge in practical life.

I am also grateful for the financial support from the China Scholarship Council, which sponsored my living expenses throughout the four years of my studies abroad. This support has made it possible for me, hailing from a rural background, to fulfill my educational aspirations overseas. Heartfelt thanks to the School of Mathematics, Statistics, and Physics at Newcastle University for generously waiving my tuition fees and providing invaluable services. My sincere appreciation extends to the dedicated lecturers and staff who have played a pivotal role in shaping my academic journey.

During my four years in Newcastle upon Tyne, I had the privilege of meeting many wonderful people, and the joyful moments shared with them are unforgettable. I cannot imagine navigating the long and tedious research journey without the companionship of colleagues Aida, Ashlin, and Francesca. Our laughter resonated from the office to the badminton courts, pubs, restaurants, and even the mountains and lakes of the UK. Special thanks to my friends Haiyan and Yang for their advice and assistance in my research, and I will always cherish the times we spent hiking together during leisure moments. To my friend Yan, thank you for your companionship, particularly during the challenging times of the COVID-19 pandemic, which made my life in the UK less lonely and more manageable. Gratitude to all the people I have had the privilege of meeting.

Lastly, I want to express my deepest gratitude to my partner Mengmeng and my family. Thank you for your unwavering understanding and support as I pursued my dreams. Your boundless and selfless love gives meaning to all my efforts and sacrifices. I will cherish and love you forever.

Abstract

EEG (electroencephalography) is a method for recording the brain’s electrical activity in real-time by placing electrodes on the subject’s scalp or, in some circumstances, surgically within the cranium, allowing for the measurement of neural signals. Analysing dynamic brain patterns with EEG is crucial for diagnosing and treating epilepsy. Statistical analysis of EEG data commonly relies on $p \times p$ covariance (or correlation) matrices derived from pre-processed signals, where p represents the number of electrodes. However, the space of covariance (or correlation) matrices is not a vector space with the usual additive structure, and so analysis of samples or time series of covariance (or correlation) matrices must make some geometrical assumptions about the underlying space. Fortunately, both the set of $p \times p$ non-singular covariance and correlation matrices form Riemannian manifolds. Riemannian geometry provides a systematic mathematical framework that allows conventional linear statistical methods to be adapted to the non-linear geometrical setting.

The number of electrodes p can be large, necessitating dimensional reduction to make analysis more computationally tractable. Moreover, electromagnetic artefacts and high correlations between sets of electrodes can result in rank deficiencies in the observed matrix-valued time series. These singularities make analysis more difficult and represent redundancies in the data. To address this, we employ dimensional reduction techniques by identifying linear combinations of channels and selecting channel subsets. These approaches ensure that the reduced time series of covariance (or correlation) matrices remain strictly positive definite, and the data thereby lie on certain smooth manifolds with the natural Riemannian structure.

Our focus is on modelling time series $\{S_i : i = 1, \dots, n\}$ of full-rank covariance (or correlation) matrices. This data can be examined within one of three spaces: $\mathcal{C}^+(p) \subset \mathcal{S}^+(p) \subset \text{Sym}(p)$. In this context, $\text{Sym}(p)$ comprises symmetric matrices and is equipped with the Frobenius norm for Euclidean geometry. $\mathcal{S}^+(p)$ represents the space of symmetric positive definite matrices, equipped with an affine invariant metric that preserves invariance under affine transformations, especially for high-magnitude covariance matrices. By factoring out variances from covariance matrices, the set of full-rank correlation matrices is represented in the quotient geometry, denoted as $\mathcal{C}^+(p)$. Thus, we intrinsically analyse EEG matrix-valued time series data within these three spaces.

Although manifold-valued data have gained substantial attention and applications in various fields recently, the literature on manifold-valued time series remains limited. This research aims to address two main objectives. First, we aim to develop manifold-adapted models for time series of matrix-valued EEG data with interpretable parameters for different possible modes of EEG dynamics. The model specifies a distribution for the tangent

direction vector at any time point, combining an autoregressive term, a mean-reverting term, and a form of Gaussian noise. This model effectively captures a wide range of potential dynamics governing the evolution of EEG data, from a smooth progression along geodesics to a noisy mean-reverting random walk within the underlying manifold. Secondly, we aim to explore the extent to which modelling results are affected by the choice of the manifold and its associated geometry. Manifold-adapted models are implemented in different tangent spaces of $\text{Sym}(p)$, $\mathcal{S}^+(p)$, and $\mathcal{C}^+(p)$. This enables modelling time series of covariance matrices in $\text{Sym}(p)$ and $\mathcal{S}^+(p)$, and time series of correlation matrices in $\text{Sym}(p)$, $\mathcal{S}^+(p)$, and $\mathcal{C}^+(p)$. Note that $\text{Sym}(p)$ can be specified as a Riemannian manifold and is convenient to present it in that way for comparison with geometries on $\mathcal{S}^+(p)$ and $\mathcal{C}^+(p)$. The comparison of these geometries sheds light on their relative advantages.

To handle the potentially large number of parameters, we simplify the general manifold-adapted model to two simpler models with fewer parameters. These simplified coefficients reveal the relative coefficients of each dynamics mode at each time point for each pair of electrodes. Parameter inference is carried out through maximum likelihood estimation. The Mahalanobis distance serves as a metric to gauge the dissimilarity of seizures based on estimated coefficients and their asymptotic covariance matrices. The results effectively discriminate between epileptic ictal (during a seizure) and interictal (between seizures) periods in patients and quantify the dissimilarity among seizures. The affine invariant geometry and quotient geometry also provide a better fit for time series of covariance matrices and correlation matrices, respectively.

In this research, we primarily construct manifold-adapted models for time series of covariance and correlation matrices derived from EEG data for epilepsy patients. We also contribute to the research on correlation matrix space by introducing a quotient metric inspired by the affine invariant metric in the covariance matrix space. The geometric concepts within the Riemannian structure of three spaces open avenues for future work related to non-Euclidean statistical models using manifold-valued data.

Contents

1	Introduction	1
1.1	Background of EEG data analysis	1
1.2	Matrix-valued data analysis	2
1.3	Aims of the research and contributions	4
1.4	Thesis outline	5
1.5	R package: <code>geomTS</code>	7
2	Riemannian geometry and statistics for manifold-valued data	8
2.1	Introduction	8
2.1.1	Basic notions on smooth manifolds	8
2.1.2	Statistics on Riemannian manifolds	10
2.1.3	Matrix computations	11
2.2	The Euclidean geometry on $(\text{Sym}(p), g^{\text{euc}})$	12
2.3	The affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$	13
2.3.1	The affine invariant metric	14
2.3.2	Exponential and logarithm maps	15
2.3.3	Parallel transport and orthonormal coordinate system	15
2.3.4	Fréchet sample mean and variance	17
2.4	The quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$	19
2.4.1	Riemannian submersion and quotient manifold	20
2.4.2	The quotient metric	26
2.4.3	Optimizing along the fibre	28
2.4.4	Exponential and logarithm maps	32
2.4.5	Parallel transport	34
2.4.6	Orthonormal coordinate system	37
2.4.7	Fréchet sample mean and variance	41
3	Data reduction and preliminary analysis	44
3.1	Data acquisition	44

3.2	Dimensional reduction on covariance and correlation matrices	47
3.2.1	Methods	47
3.2.2	Results	49
3.3	Data exploration	52
3.3.1	Covariance matrices in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$	52
3.3.2	Correlation matrices in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$	53
4	Manifold-adapted models for matrix-valued time series data	57
4.1	Rationale	57
4.2	Model specification	59
4.2.1	General model	59
4.2.2	Vectorization of tangent vectors	60
4.2.3	Models with reduced parameters	62
4.3	Model inference and selection	63
4.3.1	Model inference	63
4.3.2	Asymptotic covariance matrix of estimates	66
4.3.3	Model selection	71
4.4	Model diagnostics	72
4.4.1	Residual analysis	73
4.4.2	Goodness of fit	73
4.5	Simulation studies	74
4.5.1	Simulation study on the scalar coefficient model	74
4.5.2	Simulation study on the diagonal matrix coefficient model	79
4.5.3	Summary	81
5	Applications to EEG data	83
5.1	Brain dynamics of EEG data under manifold-adapted models	84
5.1.1	Model interpretation in scalar coefficient and diagonal coefficient models	84
5.1.2	Goodness of modelling fit	89
5.1.3	Model selection of lags	90
5.1.4	Residual analysis after modelling fit with fixed maximum lag $L = 4$	91
5.1.5	Summary	95
5.2	Comparing model results on different geometries	96
5.2.1	Covariance matrix-valued time series in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$	96
5.2.2	Correlation matrix-valued time series in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$	98
5.2.3	Summary	100
5.3	Seizure and patient variability	100

5.3.1	Seizure variability within patients	101
5.3.2	Seizure variability between patients	102
5.3.3	Patient variability	104
5.3.4	Summary	108
6	Discussion	109
6.1	Summary	109
6.2	Limitations of the study	111
6.3	Future Research	112
A	Model results for Patient ID6, ID7, and ID13	114
A.1	Patient ID6	115
A.1.1	Covariance matrices in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$	115
A.1.2	Correlation matrices in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$	117
A.2	Patient ID7	119
A.2.1	Covariance matrices in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$	119
A.2.2	Correlation matrices in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$	121
A.3	Patient ID13	123
A.3.1	Covariance matrices in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$	123
A.3.2	Correlation matrices in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$	125

List of Figures

- 1.1 Diagram representing a time series $\{S_i : i = 1, \dots, n\}$ of full rank covariance (or correlation) matrices. The ambient space is the space of $p \times p$ symmetric matrices $\text{Sym}(p)$, while the shaped surface represents the manifold of symmetric positive definite matrices $\mathcal{S}^+(p)$ embedded within $\text{Sym}(p)$. A perturbation of S_i by an arbitrary estimated ‘velocity’ $V_i \in \text{Sym}(p)$ does not necessarily lie within the manifold $\mathcal{S}^+(p)$ 5
- 2.1 Diagram of geodesic, exponential and logarithm maps. $\gamma(t)$ (bold line) and $\gamma^*(t)$ (dashed line) are two paths in $\mathcal{S}^+(p)$, and $\gamma(t), t \in [0, 1]$ is the geodesic, which is the unique and shortest path connecting S_1 and S_2 . The exponential map projects V_1 in the tangent space $T_{S_1}\mathcal{S}^+(p)$, coloured grey, onto the point $S_2 \in \mathcal{S}^+(p)$. The inverse map, the logarithm map, computes the tangent vector $V_1 \in T_{S_1}\mathcal{S}^+(p)$ of S_2 at S_1 . Note that the tangent vector V_1 does not necessarily lie within the manifold $\mathcal{S}^+(p)$ but must be in $\text{Sym}(p)$. 16
- 2.2 Values of Fréchet function for the gradient descent algorithm with various step sizes, ranging from $dt = 0.1$ to 1.5 and Sturm’s algorithm. The x -axis represents iterative steps, and the y -axis represents the values of Fréchet function $\mathcal{F}$. The ticks on the x -axis indicate the iterative steps at which the algorithm converged. 19
- 2.3 Illustration of Riemannian submersion and the decomposition of the tangent space $T_S\mathcal{S}^+(p)$. The submersion π maps any point $S \in \mathcal{S}^+(p)$ to $C = \pi(S) \in \mathcal{C}^+(p)$. $d\pi_S$ is the derivative of π at S , which projects the tangent vector $V \in T_S\mathcal{S}^+(p)$ to $X \in T_C\mathcal{C}^+(p)$. The grey colored horizontal subspace $\mathcal{H}_S$ is isomorphic to the grey tangent space $T_C\mathcal{C}^+(p)$, and S is an arbitrary element in the fiber $\pi^{-1}(C)$. The tangent vector $V \in T_S\mathcal{S}^+(p)$ can be decomposed into the horizontal component $U \in \mathcal{H}_S$ and the vertical component $W \in \mathcal{V}_S$ 22

2.4	An illustration of the horizontal lift. Given a point $C \in \mathcal{C}^+(p)$ and its tangent vector $X \in T_C \mathcal{C}^+(p)$, we can obtain an element S on the fibre $\pi^{-1}(C) \subset \mathcal{S}^+(p)$. The horizontal lift $X^\sharp$ is computed by pulling X back into the horizontal subspace $\mathcal{H}_S$. Note that $c(t), t \in [0, 1]$ is the geodesic in $\mathcal{C}^+(p)$	25
2.5	Convergence of the gradient descent algorithm with various step sizes, ranging from 0.01 to 0.2. The x -axis represents iterative steps, and the y -axis represents the values of the objective function $f(D)$. The ticks on the x -axis indicate the iterative steps at which the algorithm converged.	31
2.6	Optimal positions along fibres. C_1 and S_1 are both on the fibre $\pi^{-1}(C_1)$, C_2^* and S_2^* are both on the fibre $\pi^{-1}(C_2)$, and $C_1 \& C_2^*$ and $S_1 \& S_2^*$ are in optimal positions. Geodesic $\gamma_{C_1, C_2^*}^{\mathcal{H}}$ ($\gamma_{S_1, S_2^*}^{\mathcal{H}}$) connecting two optimal positions C_1 and C_2^* (S_1 and S_2^*) is horizontal. $\text{Log}_{C_1}(C_2^*)$ and $\text{Log}_{S_1}(S_2^*)$ are both horizontal vectors lying in the horizontal subspaces $\mathcal{H}_{C_1}$ and $\mathcal{H}_{S_1}$, respectively.	34
2.7	The diagram of the parallel transport in $\mathcal{C}^+(p)$ for any given $C_1, C_2 \in \mathcal{C}^+(p)$ and $X_1 \in T_{C_1} \mathcal{C}^+(p)$. The general steps are: (i) find the optimal position $C_2^* = DC_2 D \in \mathcal{S}^+(p)$ in the fibre $\pi^{-1}(C_2)$ with respect to $C_1 \in \mathcal{S}^+(p)$; (ii) calculate the horizontal lift $X^\sharp$ of horizontal vector $X \in T_{C_1} \mathcal{C}^+(p)$ onto $\mathcal{H}_{C_1}$; (iii) parallel transport horizontal lift $X^\sharp$ from horizontal subspace $\mathcal{H}_{C_1}$ to horizontal subspace $\mathcal{H}_{C_2^*}$ along horizontal geodesic $\gamma^{\mathcal{H}}$ connecting from C_1 to C_2^* ; (iv) project the transported vector $\mathcal{P}_{C_1 \rightarrow C_2^*}(X^\sharp)$ onto the tangent space $T_{C_2} \mathcal{C}^+(p)$ by the derivative of the submersion π at C_2^* , i.e. $X_{2,1} = d\pi_{C_2^*}(\mathcal{P}_{C_1 \rightarrow C_2^*}(X^\sharp))$	36
2.8	Illustration of $\text{Vec}(X)$ in $\mathcal{C}^+(p)$. (i): project horizontal basis in $\mathcal{H}_{I_p}$ into tangent space $T_{I_p} \mathcal{C}^+(p)$; (ii): do parallel transport of basis F_r from $T_{I_p} \mathcal{C}^+(p)$ to $T_C \mathcal{C}^+(p)$. (iii): pull back X and translated basis in F_r^C to the horizontal subspace $\mathcal{H}_S$ by the horizontal lift and S is any base point in $\pi^{-1}(C)$. (iv): vectorize the tangent vector X by calculating Riemannian metric $g_S^{\text{aff}} \langle X^\sharp, (F_r^C)^\sharp \rangle$	39
2.9	Coordinate vectors by the horizontal lift. $C^\dagger = \pi^{-1}(C) = D^\dagger C D^\dagger \subset \mathcal{S}^+(p)$ is in optimal positions with respect to $I \in \mathcal{S}^+(p)$. $C^{\dagger 1/2} \text{hor}(E_r) C^{\dagger 1/2}$ is the orthonormal basis in the horizontal subspace $\mathcal{H}_{C^\dagger}$, $r = 1, \dots, \frac{1}{2}p(p-1)$	40
2.10	Values of the Fréchet function for the same data set in the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$ (red), using Sturm's algorithm on $(\mathcal{C}^+(p), g^{\text{quo}})$ (green), and using the submersion of the Fréchet sample mean in $\mathcal{S}^+(p)$ (blue).	42

3.1	Proportion of variance retained in seizures after reduction to $p = 10, 15, 20$ dimensions. Each boxplot shows the distribution across 112 seizures. Method I and II are coloured by red and green, respectively.	50
3.2	$\log_{10}(\omega(\mathfrak{C}))$ values from Method I (coloured by red) and Method II (coloured by green) with respect to the reduced data set for correlation matrices across 112 seizures choosing $p = 10, 15$, and 20. The value of $\log_{10}(\omega(\mathfrak{C}))$ for original full $q \times q$ correlation matrices is coloured by black.	51
3.3	The Fréchet sample variance for each seizure time series and the corresponding interictal time series, broken down by patient ID, using the affine invariant metric. Dimensional reduction was performed independently for the interictal series. The median points of the interictal and seizure distributions for each patient are linked by a line to highlight the increase.	53
3.4	MDS of seizure time series for Patient ID18 using the Euclidean and affine invariant metrics to construct matrices of distance $d_{g^{\text{euc}}}$ in (3.4a) and $d_{g^{\text{aff}}}$ in (3.4b). Patient ID18 has 5 seizures (first 5 panels). Panel 6 shows the result for the interictal period corresponding to Seizure 5. Plots are coloured to show the development of time series over time. Each panel title in brackets shows the stress majorization of 2-dimensional MDS.	54
3.5	The Fréchet sample variance for each seizure time series and the corresponding interictal time series, broken down by patient ID on quotient geometry. Interictal series keep the same selected channels by dimensional reduction to seizure series. The median points of the interictal and seizure distributions for each patient are linked by a line to highlight the increase.	55
3.6	MDS of seizure time series for Patient ID18 using the Euclidean, affine invariant, and quotient metrics to construct matrices of distance $d_{g^{\text{euc}}}$ in (3.6a), $d_{g^{\text{aff}}}$ in (3.6b), and $d_{g^{\text{quo}}}$ in (3.6c). Patient ID18 has 5 seizures (first 5 panels). Panel 6 shows the result for the interictal period corresponding to Seizure 5. Plots are coloured to show the development of time series over time. Each panel title in brackets shows the stress majorization of 2-dimensional MDS plots.	56

4.1	Top Left: Successive distances via the geodesic distance $d_{g\text{aff}}$. We omit the burn-in period and fit the model to the last 100 data points, highlighted in red. Top Right: Connected scatter plot. Utilizing the affine invariant metric, the plot exhibits a gradient of colours from red to green, culminating in blue. The percentage value in brackets represents the stress of the metric MDS plot. Bottom Left: AIC values. AIC values for lags 1 to 10 are depicted. The red point is selected by minimizing AIC values. Bottom Right: Estimated parameters and confidence intervals. The plot displays estimated parameters (red points) along with 95% confidence intervals (vertical intervals). True parameters are marked in blue.	75
4.2	Top Left: Successive distance via the geodesic distance $d_{g\text{quo}}$. We omit the burn-in period and fit the model to the last 100 data points, highlighted in red. Top Right: Connected scatter plot. Utilizing the quotient metric, the plot exhibits a gradient of colours from red to green, culminating in blue. The percentage value in brackets represents the stress of the metric MDS plot. Bottom Left: AIC values. AIC values for lags 1 to 10 are depicted. The red point is selected by minimizing AIC values. Bottom Right: Estimated parameters and confidence intervals. The plot displays estimated parameters (red points) along with 95% confidence intervals (vertical intervals). True parameters are marked in blue.	78
4.3	Top left: Successive distance via the geodesic distance $d_{g\text{aff}}$. After discarding the burn-in period, the plot showcases the last 100 data points, highlighted in red. Top right: Connected scatter plot. Utilizing the affine invariant metric, this plot exhibits a gradient of colors from red to green, culminating in blue. The percentage value in brackets represents the stress of the metric MDS plot. Bottom left: AIC values. AIC values for lags 1 to 10 are displayed, with the red point selected by minimizing AIC values. Bottom right: Boxplots of parameters. This plot presents boxplots comparing setting parameters (in red) with estimated parameters (in green).	80
4.4	Top left: Successive distance via the geodesic distance $d_{g\text{quo}}$. After discarding the burn-in period, the plot showcases the last 100 data points, highlighted in red. Top right: Connected scatter plot. Utilizing the quotient metric, this plot exhibits a gradient of colours from red to green, culminating in blue. The percentage value in brackets represents the stress of the metric MDS plot. Bottom left: AIC values. AIC values for lags 1 to 10 are displayed, with the red point selected by minimizing AIC values. Bottom right: Boxplots of parameters. This plot presents boxplots comparing setting parameters (in red) with estimated parameters (in green).	81

5.1	The estimated coefficients in the scalar and diagonal coefficient model (4.2) with fixed lags $L = 8$ for five seizures and an interictal period within Patient ID18 in geometries on $(\mathcal{S}^+(p), g^{\text{aff}})$ (top panel) and $(\mathcal{C}^+(p), g^{\text{quo}})$ (bottom panel). In the top panel, the red points represent the estimated scalar coefficients in the model with constant parameters. The boxes depict the distribution of estimated diagonals in the model with diagonal coefficients. The x -axis displays $\hat{A}_1$ to $\hat{A}_8$, as well as $\hat{B}$, and $\hat{\Sigma}$	85
5.2	Histograms of normalized inner products between tangent vectors V_k and $V_{k,1}$ for 5 seizures and an interictal series of Patient 18.	86
5.3	The heatmap of the distributions of diagonal matrix coefficients $\hat{A}_l$ and $\hat{B}$ for Patient ID18, where $l = 1, \dots, L$. The colours represent values ranging from the minimum in black, 0 in white, to the maximum in red.	88
5.4	PACF for 5 seizures and an interictal data within Patient ID18 in the scalar coefficient model. The maximum lags are 8. Since all partial correlations are negative, only lower bounds of confidence intervals with blue dashed lines are shown in the plot.	92
5.5	PACF for 5 seizures and an interictal data within Patient ID18 in the diagonal model. The maximum lags are 8. The black cells indicate the estimated $\hat{a}_{lr}$ are outside the 95% confidence intervals.	93
5.6	Histograms of the value of L selected for the scalar coefficient model fitted to seizure and interictal series in Euclidean and affine invariant geometries.	94
5.7	Scalar coefficients for Patient ID 18: comparison of estimated values for five seizures and corresponding interictal series, using Euclidean and affine invariant geometries (Top and bottom row, respectively). The values of model coefficients $\alpha_{11}, \dots, \alpha_4$, β , and σ are shown with 95% confidence intervals marked.	96
5.8	Scalar coefficients for Patient ID 18: comparison of estimated values for five seizures and corresponding interictal series, using Euclidean, affine invariant, and quotient geometries (Top, middle, and bottom row, respectively). The value of model coefficients $\alpha_1, \dots, \alpha_4, \beta$ and σ are shown with 95% confidence intervals marked.	99

5.9	Metric MDS of Mahalanobis distances between seizures for Patient ID6, ID7, ID13 and ID18. Each epileptic patient, with a different number of seizures, is represented by a unique colour and shape. Points representing seizures for each patient are linked with a line representing the time ordering of seizure occurrence. The stress of the metric MDS is represented in the proportion shown in the subfigure titles. The left panel illustrates seizure variability in the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$ using covariance matrices, while the right panel displays seizure variability in the quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$	103
5.10	Seizure variability for all seizures among 18 patients. Each epileptic patient, with a different number of seizures, is represented by a unique colour and shape. More details could be found in caption of Figure 5.9.	103
5.11	Estimated parameters (left panel) and patient dissimilarity (right panel) for 18 patients. Each epileptic patient is represented by a unique colour.	106
5.12	Metric MDS of Mahalanobis distances between patients in different geometries using covariance and correlation matrices, separately. Each epileptic patient, with a different number of seizures, is represented by a unique colour and shape. The percentages in brackets represent stress values from metric MDS plots.	107
A.1	MDS plots for Patient ID6 using the Euclidean and affine invariant metrics.	115
A.2	Estimated parameters for Patient ID6 in $(\text{Sym}(p), g^{\text{euc}})$ (left panel) and $(\mathcal{S}^+(p), g^{\text{aff}})$ (right panel)	116
A.3	MDS plots for Patient ID6 using the Euclidean, affine invariant and quotient metrics.	117
A.4	Estimated parameters for Patient ID6 in $(\text{Sym}(p), g^{\text{euc}})$ (left panel), $(\mathcal{S}^+(p), g^{\text{aff}})$ (middle panel), and $(\mathcal{C}^+(p), g^{\text{quo}})$ (right panel).	118
A.5	MDS plots for Patient ID7 using the Euclidean and affine invariant metrics.	119
A.6	Estimated parameters for Patient ID7 in $(\text{Sym}(p), g^{\text{euc}})$ (left panel) and $(\mathcal{S}^+(p), g^{\text{aff}})$ (right panel).	120
A.7	MDS plots for Patient ID7 using the Euclidean, affine invariant and quotient metrics.	121
A.8	Estimated parameters for Patient ID7 in $(\text{Sym}(p), g^{\text{euc}})$ (left panel), $(\mathcal{S}^+(p), g^{\text{aff}})$ (middle panel), and $(\mathcal{C}^+(p), g^{\text{quo}})$ (right panel).	122
A.9	MDS plots for Patient ID13 using the Euclidean and affine invariant metrics.	123
A.10	Estimated parameters for Patient ID13 in $(\text{Sym}(p), g^{\text{euc}})$ (left panel) and $(\mathcal{S}^+(p), g^{\text{aff}})$ (right panel).	124
A.11	MDS plots for Patient ID13 using the Euclidean, affine invariant and quotient metrics.	125

A.12 Estimated parameters for Patient ID13 in $(\text{Sym}(p), g^{\text{euc}})$ (left panel), $(\mathcal{S}^+(p), g^{\text{aff}})$ (middle panel), and $(\mathcal{C}^+(p), g^{\text{quo}})$ (right panel).	126
---	-----

List of Tables

3.1	Patient information.	46
4.1	True and estimated parameters with 95% confidence intervals.	76
4.2	Two sample Kolmogorov-Smirnov test for two sets of parameters.	80
5.1	The goodness of modelling fit by R^2 for Patient ID18 with different types of reduced parameters and lags using covariance matrices in $(\mathcal{S}^+(p), g^{\text{aff}})$. . .	89
5.2	The goodness of modelling fit by R^2 for Patient ID18 with different types of reduced parameters and lags using correlation matrices in $(\mathcal{C}^+(p), g^{\text{quo}})$. . .	89
5.3	Counts of dimensions satisfying assumptions for covariance matrices in $(\mathcal{S}^+(p), g^{\text{aff}})$ within Patient ID18.	94
5.4	Counts of dimensions satisfying assumptions for correlation matrices in $(\mathcal{C}^+(p), g^{\text{quo}})$ within Patient ID18.	95
5.5	The goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$ for Patient ID18	97
5.6	The goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$ for Patient ID18.	99
5.7	Seizure dissimilarity in $(\mathcal{S}^+(p), g^{\text{aff}})$ using covariance matrices for Patient ID18	101
5.8	Seizure dissimilarity in $(\mathcal{C}^+(p), g^{\text{quo}})$ using correlation matrices for Patient ID18	102
A.1	Goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$ for Patient ID6	116
A.2	Goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$ for Patient ID6.	118
A.3	Goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$ for Patient ID7	120
A.4	Goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$ for Patient ID7.	122

A.5	Goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$ for Patient ID13.	124
A.6	Goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$ for Patient ID13.	126

Chapter 1

Introduction

1.1 Background of EEG data analysis

Electroencephalography (EEG) is a method of measuring electrical fields in the brain. In EEG, electrodes are placed either on the surface of the scalp or on the exposed surface of the brain to record the electrical activity generated by currents in and around neurons. Since EEG is a direct real-time measurement of neural activity, it can characterise the integrity of specific neurophysiological pathways, states of consciousness/sleep, and the precise temporal dynamics of brain functions (Kiloh *et al.*, 2013; Biasiucci *et al.*, 2019). Therefore, analysing brain dynamic patterns from EEG data can help specialists study disorders related to the electrical activity of the brain, such as epilepsy, brain tumours, memory problems, etc. (Klimesch, 1997; Noachtar & Rémi, 2009; Preuß *et al.*, 2015).

Focal epilepsy is characterised by spontaneous recurrent seizures arising from localised cortical sites (Rosenow & Lüders, 2001). An open question is how much seizures themselves vary among individual patients. Past studies have shown that seizures in individual patients share common features and evolve through similar sequences or characteristic pathways in spatiotemporal neural dynamics (Truccolo *et al.*, 2011; Schindler *et al.*, 2011; Burns *et al.*, 2014; Karoly *et al.*, 2018). Meanwhile, there is evidence that seizures may vary in the same patient. Clinically, some patients have multiple seizure onset sites, each of which produces its own unique seizure dynamics (Spencer *et al.*, 1981). Additionally, long-term EEG recordings suggest that a subset of patients have multiple seizure evolutions, and seizure patterns may vary within the same patient (Cook *et al.*, 2016; Freestone *et al.*, 2017; Naftulin *et al.*, 2018). This variability likely stems from fluctuations in underlying brain states, suggesting that background neural activity influences seizure evolution and affects seizure severity (Badawy *et al.*, 2009; Karoly *et al.*, 2017). Therefore, seizures are not a completely deterministic sequence of abnormal brain activity patterns. A given treatment may only address a subset of the patient’s seizures; for example, a single neu-

rostimulation regimen may not control all seizures, and a single prediction algorithm may not predict all seizures (Freestone *et al.*, 2017).

To design optimal and comprehensive treatments, we need to understand the prevalence and characteristics of the variability of patients' and within-patient seizures. Hypothesising that fluctuations in certain features captured in continuously recorded EEG may serve as biomarkers of epilepsy regulatory processes (Höller & Nardone, 2021), several studies have focused on quantifying and characterising seizure variability in the evolution of epileptic networks (Kuhnert *et al.*, 2010; Burns *et al.*, 2014).

An EEG procedure using q electrodes records q real-valued time series, each measuring the strength of electrical signals received from each electrode throughout a time period (referred to as *channels*). Covariance and correlation matrices are then computed to delineate relationships between the activity recorded from channels situated in different regions of the brain. Analysing these relationships can yield more detailed information about brain dynamics than examining the raw, real-valued signal data (Keogh *et al.*, 2019). Currently, a relevant and straightforward method involves evaluating functional network dynamics (also known as functional connectivity patterns) to uncover the evolutions of seizure networks (Schindler *et al.*, 2008; Burns *et al.*, 2014; Schroeder *et al.*, 2020).

Nowadays, the utilisation of Riemannian geometries for matrix-valued EEG data is gaining increasing attention due to accumulating evidence of its simplicity, accuracy, robustness, and transfer learning capabilities (Yger *et al.*, 2016; Congedo *et al.*, 2017). While most geometrical models are centred around image processing in Diffusion Tensor Imaging (DTI), seizure detection, clustering/classification, etc. (Congedo *et al.*, 2017; Guan *et al.*, 2019; Yamamoto *et al.*, 2020), the analysis of brain dynamics from EEG time series data based on Riemannian geometry remains underexplored, particularly in measuring seizure variability.

1.2 Matrix-valued data analysis

Covariance and correlation matrices are symmetric positive semi-definite, but we further restrict datasets to strictly positive definite matrices to avoid redundancies in experimental data. Then, the set of $p \times p$ symmetric positive definite matrices forms a differentiable manifold with a natural Riemannian structure (Bhatia, 2009). In this research, we argue that it is important not only to take this restriction into account but also to consider the curvature of the space when viewed as a Riemannian manifold. Analysis of matrix-valued data is attracting increasing research attention in statistics and machine learning due to the availability of increasing volumes of such data generated by a range of different experimental techniques in data science. An excellent overview of the area and related topics is given by Marron & Alonso (2014). Directional data (Jupp & Mardia, 2009) and some

shape data (Dryden & Mardia, 2016) lie on Riemannian manifolds, and manifold-based statistics play an important role in medical image analysis (Pennec *et al.*, 2019). Analysis of datasets of symmetric positive definite matrices has received particular attention in this context (Pennec *et al.*, 2019, Chapter 5), due to the tractable nature of the geometry and the range of applications, especially in different forms of medical imaging such as DTI (Lenglet *et al.*, 2006). Datasets of covariance matrices arise in a number of different experimental contexts, and Dryden *et al.* (2009) describe several alternative geometries for the space of $p \times p$ covariance matrices. Some of these geometries take rank deficiency into account, but this can make mathematical analysis more involved since data then lie on a manifold-stratified space (Marron & Alonso, 2014, Section 6). Since we assume the data consist of full-rank matrices, it follows that they lie on a manifold equipped with different Riemannian metrics (called *manifold-valued data*), for example, the Log-Euclidean metric (Arsigny *et al.*, 2006), the affine invariant metric (Moakher & Zéraï, 2011; Lenglet *et al.*, 2006), and the Bures-Wasserstein metric (Malagò *et al.*, 2018; Han *et al.*, 2021).

A variety of statistical methods have been developed for manifold-valued data. The definition of the sample mean in Euclidean space does not generally apply on a manifold, so an intrinsic mean was developed (Fréchet, 1948), together with associated theory for asymptotic properties (Bhattacharya & Patrangenaru, 2003, 2005). Analogs to principal component analysis (Fletcher *et al.*, 2004; Huckemann & Ziezold, 2006) and linear regression (Fletcher, 2013) have been proposed, replacing straight lines in Euclidean space with geodesics in Riemannian manifolds, and other higher dimensional subspaces to replace linear Euclidean subspaces (Jung *et al.*, 2012). Manifold-valued Gaussian processes have been developed and applied to functional data on manifolds (Mallasto & Feragen, 2018). Furthermore, there is an existing method for learning trajectories from manifold-valued longitudinal data (Schiratti *et al.*, 2015). Recently, continuous-time manifold-valued Markov processes, which are analogs of the Euclidean Ornstein-Uhlenbeck process, have been developed (Bui *et al.*, 2023). However, the literature on manifold-valued time series appears to be more limited.

To the best of our knowledge, the affine invariant geometry has been widely acknowledged as a comprehensive and popular Riemannian framework, and used in a variety of applications involving covariance matrices of EEG signal data (Congedo *et al.*, 2017; Yger *et al.*, 2016). While correlation matrices can be computed from covariance matrices, sharing the same rank, there are limited tools available for the intrinsic computation with correlation matrices from a geometric standpoint. Correlation matrices are often treated as symmetric positive definite matrices within the affine invariant geometry (Varoquaux *et al.*, 2010). Interestingly, recent discussions have explored full-rank correlation matrices, which constitute a strict sub-manifold of the cone of symmetric positive definite matrix space, referred to as the quotient manifold (Thanwerdas & Pennec, 2021; David, 2019).

However, these discussions on characteristics in the quotient manifold are limited and its applications are still evolving, with ongoing research in this area.

1.3 Aims of the research and contributions

Let $\text{Sym}(p)$ denote the set of (real) symmetric $p \times p$ matrices and let $\mathcal{S}^+(p) \subset \text{Sym}(p)$ denote the subset of symmetric positive definite matrices. When modelling a time series of $p \times p$ covariance (or correlation) matrices $\{S_i \in \mathcal{S}^+(p) : i = 1, \dots, n\}$, it is natural to consider representing the $(i + 1)$ th matrix as

$$S_{i+1} = S_i + V_i + \epsilon_i, \quad (1.1)$$

where $V_i \in \text{Sym}(p)$ is a deterministic perturbation which might depend on preceding matrices in the time series or some covariates, and where ϵ_i represents the noise. For example, a very simple model is defined by $V_i = S_{i+1} - S_i$, so that V_i is an estimate of the ‘direction of travel’ of the time series at the i th time point, and by assuming that elements of ϵ_i are distributed according to some appropriate multivariate normal distribution. However, without additional constraints on V_i and ϵ_i , the right-hand side of Equation (1.1) will not generally be positive definite, and so the model is ill-defined in the sense that it gives probability mass to regions of the sample space that cannot contain data. The additive model in Equation (1.1) therefore needs modifying in order to take into account the non-Euclidean nature of the space of $p \times p$ positive definite matrices (see Figure 1.1). This can be achieved canonically by regarding $\mathcal{S}^+(p)$ as a Riemannian manifold, and working intrinsically within the Riemannian geometry. The notion of perturbing a matrix $S_i \in \mathcal{S}^+(p)$ in some direction $V_i \in \text{Sym}(p)$, expressed in Equation (1.1) by the simple addition, is instead achieved by the Riemannian exponential map which operates intrinsically within $\mathcal{S}^+(p)$.

As a key part of this study, we analyse time series $\{S_i : i = 1, \dots, n\}$ with the standard Euclidean geometry on $\text{Sym}(p)$, and compare the results to those obtained when the model is adapted to work intrinsically within Riemannian manifolds equipped with the affine invariant geometry on $\mathcal{S}^+(p)$ and the quotient geometry on $\mathcal{C}^+(p)$, representing the set of $p \times p$ full-rank correlation matrices, where $\mathcal{C}^+(p) \subset \mathcal{S}^+(p) \subset \text{Sym}(p)$. To achieve this, we construct closed formulae for the various operations we require in order to specify our model, for example Riemannian metrics, logarithm and exponential maps, parallel transports, orthonormal coordinate systems, and Fréchet sample means in these geometries.

Moreover, electromagnetic artefacts and high correlations between sets of electrodes can result in rank deficiencies in the observed matrix-valued time series, and so we propose two distinctive dimensional reduction methods to ensure that the reduced time series of

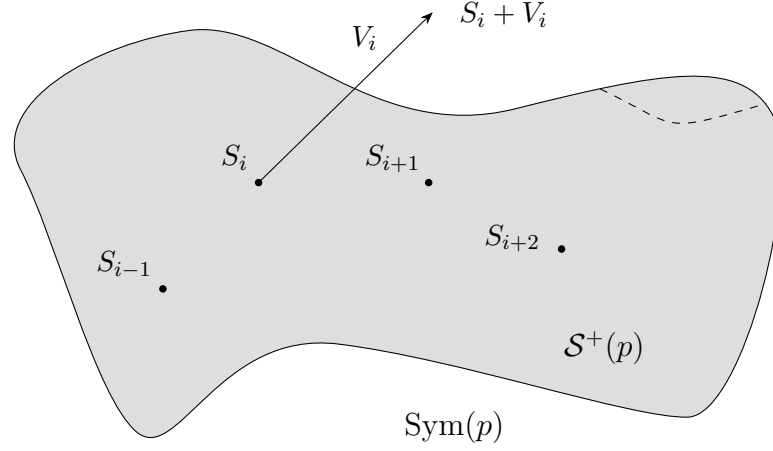


Figure 1.1: Diagram representing a time series $\{S_i : i = 1, \dots, n\}$ of full rank covariance (or correlation) matrices. The ambient space is the space of $p \times p$ symmetric matrices $\text{Sym}(p)$, while the shaped surface represents the manifold of symmetric positive definite matrices $\mathcal{S}^+(p)$ embedded within $\text{Sym}(p)$. A perturbation of S_i by an arbitrary estimated ‘velocity’ $V_i \in \text{Sym}(p)$ does not necessarily lie within the manifold $\mathcal{S}^+(p)$.

covariance (or correlation) matrices remain strictly symmetric positive definite. As a result, the data thereby lie on certain smooth manifolds with the natural Riemannian structure.

In addition to assessing the effect of underlying geometries, we propose a model which is capable of capturing different data dynamics within the time series $\{S_i : i = 1, \dots, n\}$, with the aim of providing an interpretable set of parameters for each data set. Specifically, the tangent (or direction) vector V_i is modelled as the sum of three terms: an autoregressive term, a mean-reverting term, and a noise term. By combining these terms, the model is able to capture various possible dynamics for the evolution of EEG data, from smooth flow along geodesics to a noisy mean-reverting random walk on the underlying manifold, for example. Furthermore, leveraging the model results allows for an easy measurement of seizure dissimilarity within epileptic patients and variations among patients. It overcomes the difficulties of differences in electrode numbers and placements between patients. Although we apply the model to EEG data in $\text{Sym}(p)$, $\mathcal{S}^+(p)$, and $\mathcal{C}^+(p)$, it is formulated in a general way and could be applied to data on other Riemannian manifolds.

1.4 Thesis outline

In this chapter, we introduce the basic background of EEG data analysis and applications of manifold-valued data. We also outline the main aims and contributions and provide an overview of the general structure of this thesis.

Chapter 2 will describe Riemannian geometries on different manifolds. Without loss

of generality, Section 2.1 sets out the basic notions on a smooth manifold and discusses geometrical statistics. Since our matrix-valued time series EEG data are covariance and correlation matrices, and based on their properties, we introduce three different Riemannian manifolds: the symmetric matrix space $(\text{Sym}(p), g^{\text{euc}})$ equipped with the Frobenius norm in Section 2.2, the symmetric positive definite matrix space $(\mathcal{S}^+(p), g^{\text{aff}})$ equipped with an affine invariant metric in Section 2.3. We also consider a newly developed Riemannian manifold of $p \times p$ full-rank positive definite correlation matrices, named the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$ in Section 2.4. Existing literature about this manifold is limited, and formulae for certain aspects of the geometry, such as parallel transport and orthonormal coordinate systems, have not previously been established. Therefore, Section 2.4 develops these formulae required in our proposed model and extends them to a more general tool for geometrical statistics.

After acquiring original EEG signals from an open source, we perform dimensional reduction for preprocessing covariance and correlation matrices to ensure that the reduced matrix-valued time series remain strictly positive definite. The data thereby lie on certain smooth manifolds with a natural Riemannian structure, as described in Chapter 2. Moreover, preliminary analysis on reduced covariance and correlation matrices has been done by comparing the Fréchet sample variance of seizure series and corresponding *inter-ictal* (i.e., normal EEG recorded between seizures) series and performing multidimensional scaling plots using different Riemannian metrics at the end of Chapter 3.

Following this preliminary analysis, we next construct manifold-adapted models for the manifold-valued time series data in Chapter 4. This model specifies a distribution for the tangent direction vector at any time point, combining an autoregressive term, a mean-reverting term, and a form of wrapped Gaussian noise, as described in Sections 4.1 and 4.2. The model effectively captures a wide range of potential dynamics governing the evolution of EEG data. Model inference and fitness are discussed, followed by model specification in Sections 4.3 and 4.4. To check the validity and efficiency of manifold-adapted models and model inference, Section 4.5 simulates the manifold-valued data with fixed parameters and performs model inference to verify the initial setting.

Chapter 5 discusses the model results from different perspectives, including parameter structures and data set types (covariance and correlation matrices) in Section 5.1. This section also interprets estimated parameters in the model with respect to brain dynamics in seizures. As one of the key parts of this research, Section 5.2 compares the results when the model is adapted to work intrinsically within $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$ and explores the extent to which modelling results are affected by the choice of manifold and its associated geometry. Lastly, we propose the Mahalanobis distance to measure the seizure dissimilarity between and within epileptic patients in Section 5.3.

Finally, Chapter 6 discusses the limitations and future work of this research.

1.5 R package: geomTS

In this research, we have introduced various operations in different Riemannian manifolds, including the Euclidean symmetric geometry on $(\text{Sym}(p), g^{\text{euc}})$, the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$, and the quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$. While the R package `Riemann` provides some metric functions and statistical models, several essential operations necessitating this research remain undefined. Notably, a newly developed quotient manifold is still in progress, and there are currently no related open packages for this area.

Therefore, a significant contribution of this research is the development of a comprehensive R package, named `geomTS`, to assist researchers working with the aforementioned manifolds. This package addresses the gaps left by existing tools and provides essential operations such as parallel transport and orthonormal coordinate systems in different geometries. Furthermore, the package includes dimensional reduction methods and manifold-adapted models discussed in this thesis. Interested readers can install the package from GitHub:

```
# install.packages("devtools")
devtools::install_github("TaoDing2/geomTS")
library(geomTS)
```

Chapter 2

Riemannian geometry and statistics for manifold-valued data

In this chapter, we provide a concise overview of essential concepts of Riemannian geometry relevant to our analysis. While we refrain from much of the detail of differential geometry and Lie groups, comprehensive resources such as textbooks by Do Carmo & Flaherty Francis (1992), Lang (2012), and Gallier & Quaintance (2020) offer in-depth explanations. We present this material in the context of a general Riemannian manifold and general geometrical statistics in Section 2.1. Subsequently, we proceed to furnish explicit formulae for specific instances of Riemannian manifolds encompassing the symmetric matrix space with the Euclidean metric $(\text{Sym}(p), g^{\text{euc}})$ in Section 2.2, the symmetric positive definite matrix space with the affine invariant metric $(\mathcal{S}^+(p), g^{\text{aff}})$ in Section 2.3, and the full-rank correlation matrix space with the quotient metric $(\mathcal{C}^+(p), g^{\text{quo}})$ in Section 2.4.

2.1 Introduction

This section commences with an introduction to the fundamental concepts of a general Riemannian manifold. Following this, we delve into geometrical statistics, introducing ideas such as the wrapped Gaussian distribution and the Fréchet sample mean, which are indispensable for analysing manifold-valued data within statistical models. Finally, as the EEG data we obtained consists of covariance and correlation matrices, we conclude by presenting basic operations for matrix computations.

2.1.1 Basic notions on smooth manifolds

Smooth manifolds are mathematical structures that exhibit local similarities to $\mathbb{R}^m$, where m represents a fixed dimension. However, unlike vector spaces, smooth manifolds are more general in nature and can incorporate curvature. A *smooth manifold*, denoted as $\mathcal{M}$, is

accompanied by charts, which are maps $\varphi : U \subset \mathcal{M} \rightarrow \varphi(U) \subset \mathbb{R}^m$. These charts collectively cover the entire manifold $\mathcal{M}$ and satisfy a compatibility condition. For any pair of charts (φ, U) and (ψ, V) , the transition map $\psi \circ \varphi^{-1} : \varphi(U \cap V) \rightarrow \psi(U \cap V)$ is a smooth, bijective function with a smooth inverse.

A path $\gamma : [a, b] \rightarrow \mathcal{M}$ is considered smooth if, when composed with every chart, $\varphi \circ \gamma : [a, b] \rightarrow \mathbb{R}^m$, it becomes a smooth map. The *tangent vector* to the curve γ at a given point t , denoted as $\gamma'(t)$, represents the derivative of γ at t . Thereby, the *tangent space* $T_x\mathcal{M}$ at a point $x \in \mathcal{M}$ can be understood as the vector space consisting of direction vectors associated with smooth paths passing through x .

A *Riemannian metric*, denoted as g , defined on a smooth manifold $\mathcal{M}$ is a mapping that assigns an inner product g_x to each tangent space $T_x\mathcal{M}$ in a smooth manner, where x varies over $\mathcal{M}$. Consequently, a smooth manifold equipped with a Riemannian metric g is referred to as a *Riemannian manifold*, denoted as $(\mathcal{M}, g)$. For a given smooth path $\gamma : [a, b] \rightarrow \mathcal{M}$, the Riemannian norm of its velocity at a point $t \in [a, b]$ is denoted as $g_{\gamma(t)}(\gamma'(t), \gamma'(t))^{1/2}$. This norm represents the square root of the inner product of the velocity vector with itself. By integrating this, the Riemannian norm along the path yields the length of the path γ . Specifically, let x and y be any two points in the manifold $\mathcal{M}$. The quantity $d_g(x, y)$ denotes the infimum of the lengths of all smooth paths connecting x to y , where the lengths are calculated using the Riemannian norm. It is worth noting that d_g satisfies the properties of a metric on $\mathcal{M}$. A path between x and y that achieves this infimum is referred to as a *geodesic*, which corresponds to the shortest path between x and y in terms of Riemannian length.

The Riemannian *exponential map*, denoted as $\text{Exp}_x : T_x\mathcal{M} \rightarrow \mathcal{M}$, is defined for tangent vectors of sufficiently small norm at a point x . It can be visualized as the result of following a geodesic from x in the direction of a specific tangent vector. For example, if we have a tangent vector $v \in T_x\mathcal{M}$, then the path $\text{Exp}_x(tv)$ defined for $t \in [0, 1]$ represents a geodesic from x to $\text{Exp}_x(v)$. The inverse of the exponential map at x , known as the Riemannian *logarithm map*, is denoted as $\text{Log}_x : \mathcal{M} \rightarrow T_x\mathcal{M}$. Typically, it is defined within a neighbourhood of x . For two points x and y in $\mathcal{M}$, the logarithm map $\text{Log}_x(y)$ can be understood as the tangent vector at x corresponding to the geodesic that connects x and y . In the case of the specific class of manifolds being considered, the exponential map is defined for the entire tangent space $T_x\mathcal{M}$ at each point x in $\mathcal{M}$. Likewise, the logarithm map $\text{Log}_x(y)$ is defined for all pairs of points x and y belonging to $\mathcal{M}$.

When the manifold $\mathcal{M}$ is the Euclidean space $\mathbb{R}^m$, the tangent spaces $T_{\gamma(t)}\mathbb{R}^m$ along any path $\gamma(t)$ are naturally isomorphic to $\mathbb{R}^m$, $t \in [0, 1]$. This means that at any point $\gamma(t)$, the tangent vector v also belongs to $T_{\gamma(t)}\mathbb{R}^m$. However, for a curve γ connecting two points x and y on a general manifold, Riemannian geometry allows for a natural isomorphism between $T_x\mathcal{M}$ and $T_y\mathcal{M}$ through the process of *parallel transport* along γ ,

denoted as $\mathcal{P}_{x \rightarrow y} : T_x \mathcal{M} \rightarrow T_y \mathcal{M}$. This can be achieved using the concept of the covariant derivative of a vector field with respect to Levi-Civita connection. Specifically, a *Levi-Civita connection* on $\mathcal{M}$, compatible with Riemannian metric and being torsion free, is a $\mathbb{R}$ -bilinear map $\nabla : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \rightarrow \mathfrak{X}(\mathcal{M})$, where $\nabla_{\gamma'} V \in \mathfrak{X}(\mathcal{M})$ is called the *covariant derivative* of a vector field $V \in \mathfrak{X}(\mathcal{M})$ with respect to a vector field $\gamma' \in \mathfrak{X}(\mathcal{M})$. A vector field V along γ is parallel iff $\nabla_{\gamma'} V = 0$.

Importantly, the parallel transport map $\mathcal{P}_{x \rightarrow y}$ preserves the Riemannian inner product g and, therefore, is an *isometry*. Consequently, if we have an *orthonormal basis* $\{w_i : i = 1, \dots, m\}$ for $T_x \mathcal{M}$, then for any point $y \in \mathcal{M}$, the vectors $v_i = \mathcal{P}_{x \rightarrow y}(w_i)$ form an orthonormal basis for $T_y \mathcal{M}$. A Riemannian manifold that possesses a globally defined smooth parallel orthonormal basis of vector fields is referred to as a *parallelizable manifold*. This implies that there exists a set of vector fields defined on the entire manifold such that, at every point, these vector fields form an orthonormal basis. It is crucial to recognize that the property of being parallelizable manifold does not necessarily imply that the manifold is diffeomorphic to Euclidean space. In other words, the existence of a parallel orthonormal basis does not guarantee that the manifold shares the same geometric properties as Euclidean space: examples of parallelizable manifolds include certain spheres, such as S^1, S^3, S^7 , and $S^+(p)$ with the affine invariant geometry.

2.1.2 Statistics on Riemannian manifolds

Gaussian distributions on manifolds

There are several different ways to define Gaussian distributions for statistics on Riemannian manifolds. We will adopt an approach first developed in the context of directional data (Jupp & Mardia, 2009) and use *wrapped* Gaussian distributions. Given a Riemannian manifold $\mathcal{M}$, let $\{w_i : i = 1, \dots, m\}$ be an orthonormal basis in $T_x \mathcal{M}$, and let Σ be an $m \times m$ covariance matrix. A random point $X \in \mathcal{M}$ has the *wrapped Gaussian distribution* centered at $x \in \mathcal{M}$ with covariance Σ in the basis $\{w_i : i = 1, \dots, m\}$ if

$$\begin{aligned} Y &= (Y_1, \dots, Y_m)^T \sim N(\mathbf{0}, \Sigma), \\ U &= \sum_i^m Y_i w_i, \quad \text{and} \\ X &= \text{Exp}_x(U) \end{aligned} \tag{2.1}$$

Here $N(\mathbf{0}, \Sigma)$ is the zero-mean multivariate Gaussian distribution on $\mathbb{R}^m$ with covariance Σ . When Σ is a diagonal matrix, the distribution is defined independently of any basis.

The wrapped Gaussian distribution represents a bump of density around the point x , with the degree of spread and shape of the density contours determined by the matrix Σ . In fact, we will work with a related set of distributions obtained by adding a non-zero

mean $\mathbf{u} \in \mathbb{R}^m$ to the Gaussian distribution in Equation (2.1). In this case, we write

$$U \sim N_{T_x\mathcal{M}}(\mathbf{u}, \Sigma) \quad (2.2)$$

to denote the distribution of the random tangent vector U . The resulting distribution of the random point $X = \text{Exp}_x(U)$ closely approximates the wrapped Gaussian with zero mean centered at the point $\text{Exp}_x(\sum_i w_i u_i)$ when the vector u_i is sufficiently small (Mallasto & Feragen, 2018), and w_i is the i th orthonormal basis in $T_x\mathcal{M}$, $i = 1, \dots, m$.

The Fréchet mean

Since the addition operation is not defined for points on a general manifold, the definition of a sample mean needs to be adapted. Given a set of data points $\{x_i : i = 1, \dots, n\} \in \mathcal{M}$, define the *Fréchet function* $\mathcal{F}$ by

$$\mathcal{F}(y; \{x_i\}) = \frac{1}{n} \sum_{i=1}^n d_g(y, x_i)^2 \quad (2.3)$$

where d_g is the geodesic distance in $\mathcal{M}$. A point $\tilde{x} \in \arg \min_y \mathcal{F}(y; \{x_i\})$ is called a *Fréchet sample mean* of the data $\{x_i : i = 1, \dots, n\}$. Considering the Fréchet function $\mathcal{F}$ as an objective function, the gradient descent algorithm is commonly used to find its local minimum. Additionally, the value of $\mathcal{F}$ at the Fréchet sample mean is called the *Fréchet sample variance*. However, Fréchet sample means do not necessarily exist for all data sets on a general manifold, nor are they generally unique. Of course, on $\mathbb{R}^m$ with the Euclidean metric, the Fréchet sample mean and standard sample mean coincide.

2.1.3 Matrix computations

Let $\mathcal{M}(p)$ denote the space of $p \times p$ real matrices. The subset of $p \times p$ symmetric matrices in $\mathcal{M}(p)$ is denoted by

$$\text{Sym}(p) = \{A \in \mathcal{M}(p), A^T = A\}.$$

A symmetric matrix A is positive definite, iff $\mathbf{x}^T A \mathbf{x} > 0$ for any non-zero vector $\mathbf{x}$, or equivalently, if all eigenvalues of matrix A are greater than 0. The open subset of $p \times p$ symmetric positive definite matrices in $\mathcal{S}^+(p)$ is denoted by

$$\mathcal{S}^+(p) = \{A \in \text{Sym}(p), \mathbf{x}^T A \mathbf{x} > 0 \text{ for all } \mathbf{x} \in \mathbb{R}^p \setminus \{0\}\}.$$

The exponential of a matrix $A \in \mathcal{M}(p)$ is given by the convergent series

$$\exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}, \quad k \in \mathbb{Z}^+.$$

The logarithm of any $B \in \text{Sym}(p)$ is defined to be solutions of the matrix equation $\exp(A) = B$. If all the eigenvalues of B are small enough, logarithm of a matrix $B \in \text{Sym}(p)$ is given by a convergent series

$$\log(B) = - \sum_{k=1}^{\infty} \frac{(I_p - B)^k}{k}, \quad k \in \mathbb{Z}^+,$$

where I_p is the $p \times p$ identity matrix.

The square root of the symmetric matrix A can be defined as:

$$A^{1/2} = \{B \in \text{Sym}(p), B^2 = A\},$$

with the property that

$$A^{1/2} = \exp\left(\frac{1}{2} \log(A)\right).$$

Suppose $A \in \mathcal{S}^+(p)$, and A could be written in diagonal form as $A = U\Lambda U^T$, $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_p)$. Then, we have

- (1) $A^{1/2} = U\text{Diag}(\lambda_1^{1/2}, \dots, \lambda_p^{1/2})U^T$
- (2) $\exp(A) = U\text{Diag}(e^{\lambda_1}, \dots, e^{\lambda_p})U^T$
- (3) $\log(A) = U\text{Diag}(\log(\lambda_1), \dots, \log(\lambda_p))U^T$

The above matrix definitions and computations can be found in books Curtis (2012) and Golub & Van Loan (2013).

2.2 The Euclidean geometry on $(\text{Sym}(p), g^{\text{euc}})$

When $\mathcal{M}(p) = \text{Sym}(p) \subset \mathbb{R}^{p \times p}$ with the dimension of $m = p(p+1)/2$, which denotes the set of symmetric matrices, it can be shown that every tangent space is a copy of $\text{Sym}(p)$. A Riemannian metric is then defined at $S \in \text{Sym}(p)$ for any two tangent vectors $V, W \in T_S \text{Sym}(p) = \text{Sym}(p)$ as follows:

$$g_S^{\text{euc}} \langle V, W \rangle = \text{Tr}(V^T W). \quad (2.4)$$

This implies the inner product is in fact independent of the base point S . The corresponding Riemannian geometry is exactly that of the symmetric matrices equipped with the standard Euclidean (or Frobenius) distance. Specifically, it is straightforward to show that for $S_1, S_2 \in \text{Sym}(p)$:

$$d_{g^{\text{euc}}}(S_1, S_2) = \|S_1 - S_2\|_F \quad (2.5)$$

where $\|\cdot\|_F$ is the Frobenius norm and it is computed by $\|A\|_F = \sqrt{\sum_i \sum_j a_{ij}^2}$, where a_{ij} is an element in matrix A , $i, j = 1, \dots, p$. Furthermore, given a tangent vector $V \in \text{Sym}(p)$ at the base point $S \in \text{Sym}(p)$, the exponential map is defined as:

$$\text{Exp}_S(V) = S + V. \quad (2.6)$$

Substituting tV for the tangent vector in the exponential map for any $t \in [0, 1]$ gives the geodesic from S in the direction V as $S + tV$, i.e. the straight line segment in the direction V . The inverse map of the exponential map, known as the logarithm map, is given by:

$$\text{Log}_{S_1}(S_2) = S_2 - S_1. \quad (2.7)$$

Since parallel transport of V does not change along the geodesic, we have

$$\mathcal{P}_{S_1 \rightarrow S_2}(V) = V. \quad (2.8)$$

Let e_i denote the i th standard basis vector in $\mathbb{R}^p$, where $e_i = (0, \dots, 1, \dots, 0)_p^T$. For any $1 \leq i \leq j \leq p$, we can construct an orthonormal basis in $T_S \text{Sym}(p)$ for all $S \in \text{Sym}(p)$ as

$$E_{ij} = \begin{cases} e_i e_j^T & \text{if } i = j, \text{ and} \\ \frac{\sqrt{2}}{2} (e_i e_j^T + e_j e_i^T) & \text{if } 1 \leq i < j \leq p, \end{cases} \quad (2.9)$$

To simplify the notation, let E_{ij} denote as E_r where $r = 1, \dots, m = p(p+1)/2$. It represents the r th basis vector in any tangent space $T_S \text{Sym}(p)$ at the point $S \in \text{Sym}(p)$.

The Fréchet sample mean of the set of matrices $\{S_i : i = 1, \dots, n\} \in \text{Sym}(p)$ with respect to $d_{g^{\text{euc}}}$ is the standard Euclidean mean

$$\tilde{S} = \frac{1}{n} \sum_{i=1}^n S_i. \quad (2.10)$$

Since positive definiteness is a convex property, if $S_i \in \mathcal{S}^+(p)$ for all i it follows that the Euclidean mean is also positive definite.

2.3 The affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$

Let $\mathcal{M}_p = \mathcal{S}^+(p)$ denote the set of $p \times p$ symmetric positive definite matrices, i.e. $\mathcal{S}^+(p) = \{A \in \mathcal{M}_p, A^T = A; \mathbf{x}^T A \mathbf{x} > 0 \text{ for all } \mathbf{x} \in \mathbb{R}^p \setminus \{0\}\}$. This set yields a differentiable manifold equipped with a natural Riemannian structure (Moakher & Zéraï, 2011). Various Riemannian metrics in $\mathcal{S}^+(p)$ have been introduced in literatures, for example, the Log-Euclidean metric (Huang *et al.*, 2015), the Bures-Wasserstein metric (Bhatia *et al.*, 2019;

Han *et al.*, 2021), and the affine invariant metric (Moakher, 2005; Pennec *et al.*, 2006).

In this section, we explain the reasons why we choose the affine invariant metric on covariance matrix space of $(\mathcal{S}^+(p), g^{\text{aff}})$ and give explicit formulae for this metric, exponential and logarithm maps, parallel transport, orthonormal coordinate system, and Fréchet sample mean.

2.3.1 The affine invariant metric

The affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$ arises naturally from the information geometry of multivariate normal distributions with zero mean (Skovgaard, 1984; Lenglet *et al.*, 2006). The metric satisfies the property

$$d_{g^{\text{aff}}}(AS_1A^T, AS_2A^T) = d_{g^{\text{aff}}}(S_1, S_2) \quad (2.11)$$

for all $S_1, S_2 \in \mathcal{S}^+(p)$ and A is an element in the linear group $GL(p)$, so in other words, it is invariant under changes of basis in $\mathbb{R}^p$ when S_1 and S_2 are symmetric positive definite matrices. This is particularly relevant when analysing covariance matrices from EEG recordings: the overall scale of measurements is affected by instrument sensitivity and positioning which between electrodes and so is largely artefactual. The metric is invariant under re-scaling of the signal from each electrode, which corresponds to the action of a diagonal matrix A with positive entries in Equation (2.11). The affine invariant geometry offers additional advantages over the Euclidean geometry, for example, in that the determinant of matrices along geodesics is better behaved. (In the Euclidean geometry the determinant can become inflated relative to its value at the endpoints of the geodesic.)

Suppose $S \in \mathcal{S}^+(p)$, it is straightforward to show that $T_S\mathcal{S}^+(p) = \text{Sym}(p)$ for all S . The Riemannian inner product on $T_S\mathcal{S}^+(p)$ is defined by

$$g_S^{\text{aff}}\langle V, W \rangle = \text{Tr} \left(S^{-1/2} V S^{-1/2} S^{-1/2} W S^{-1/2} \right) = \text{Tr} \left(S^{-1} V S^{-1} W \right), \quad (2.12)$$

where $V, W \in T_S\mathcal{S}^+(p)$. It follows that

$$d_{g^{\text{aff}}}(S_1, S_2) = \left\| \log \left(S_1^{-1/2} S_2 S_1^{-1/2} \right) \right\|_F = \left(\sum_{r=1}^p \log^2 \lambda_r \right)^{1/2}, \quad (2.13)$$

where $\lambda_r, r = 1, \dots, p$ are the eigenvalues of the matrix $S_1^{-1/2} S_2 S_1^{-1/2}$.

2.3.2 Exponential and logarithm maps

The Riemannian exponential map, which projects a tangent vector $V_1 \in T_{S_1}\mathcal{S}^+(p)$ back to the manifold $\mathcal{S}^+(p)$, is given by

$$\text{Exp}_{S_1}(V_1) = S_1^{1/2} \exp(S_1^{-1/2} V_1 S_1^{-1/2}) S_1^{1/2}. \quad (2.14)$$

Inversely, we can uniquely map a matrix $S_2 \in \mathcal{S}^+(p)$ back to the tangent space $T_{S_1}\mathcal{S}^+(p)$ at $S_1 \in \mathcal{S}^+(p)$ via a logarithm map:

$$\text{Log}_{S_1}(S_2) = S_1^{1/2} \log(S_1^{-1/2} S_2 S_1^{-1/2}) S_1^{1/2}. \quad (2.15)$$

The general geodesic curve $\gamma(t) \in \mathcal{S}^+(p)$ emanating from $S_1 = \gamma(0) \in \mathcal{S}^+(p)$ in the direction $V_1 = \gamma'(0) \in T_{S_1}\mathcal{S}^+(p)$ has the form of

$$\gamma(t) = \text{Exp}_{S_1}(tV_1), \quad (2.16)$$

where $t \in [0, 1]$. Therefore, the unique geodesic joining S_1 and S_2 has the shortest path length and is given by

$$\begin{aligned} \gamma(t) &= \text{Exp}_{S_1}(t \text{Log}_{S_1}(S_2)) \\ &= S_1^{1/2} \exp(t \log(S_1^{-1/2} S_2 S_1^{-1/2})) S_1^{1/2} \\ &= S_1^{1/2} \left(S_1^{-1/2} S_2 S_1^{-1/2} \right)^t S_1^{1/2}. \end{aligned} \quad (2.17)$$

In Figure 2.1, there are paths $\gamma(t)$ and $\gamma^*(t)$ both connecting S_1 to S_2 . However, the shortest path of $\gamma(t), t \in [0, 1]$ (bold line) is called a geodesic on the surface of $\mathcal{S}^+(p)$. The tangent vector $V_1 = \text{Log}_{S_1}(S_2)$ lies in the grey tangent space $T_{S_1}\mathcal{S}^+(p)$. Its exponential map $\text{Exp}_{S_1}(\cdot)$ gives us a collection of one-to-one and complete maps of the manifold, centred at point S_1 . Pennec *et al.* (2006); Moakher & Zéraï (2011) provide detailed descriptions of these formulae, and matrix computations such as square root, exponential, and logarithm map can be found in Section 2.1.3.

2.3.3 Parallel transport and orthonormal coordinate system

In Euclidean space, for any vector $V \in T_{\gamma(0)}\text{Sym}(p)$, it can be straightforwardly moved along the curve $\gamma(t)$, where $t \in [0, 1]$. This means that at any point $\gamma(t)$, the tangent vector V remains in $T_{\gamma(t)}\text{Sym}(p)$. However, when dealing with a curved manifold, the presence of curvature influences the behaviour of the vector during translation, making the definition of parallel transport more intricate.

The papers by Sra & Hosseini (2015) and Yair *et al.* (2019) provide a comprehensive

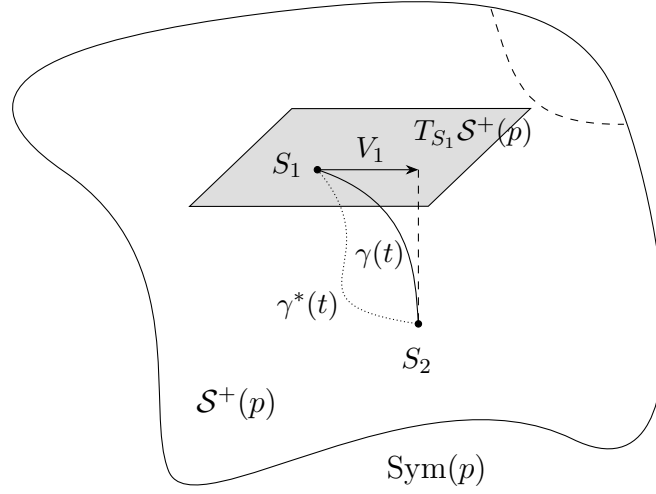


Figure 2.1: Diagram of geodesic, exponential and logarithm maps. $\gamma(t)$ (bold line) and $\gamma^*(t)$ (dashed line) are two paths in $\mathcal{S}^+(p)$, and $\gamma(t), t \in [0, 1]$ is the geodesic, which is the unique and shortest path connecting S_1 and S_2 . The exponential map projects V_1 in the tangent space $T_{S_1}\mathcal{S}^+(p)$, coloured grey, onto the point $S_2 \in \mathcal{S}^+(p)$. The inverse map, the logarithm map, computes the tangent vector $V_1 \in T_{S_1}\mathcal{S}^+(p)$ of S_2 at S_1 . Note that the tangent vector V_1 does not necessarily lie within the manifold $\mathcal{S}^+(p)$ but must be in $\text{Sym}(p)$.

explanation of the definition of parallel transport in $\mathcal{S}^+(p)$. Briefly, parallel transport involves the translation of a tangent vector V_1 from the tangent space $T_{S_1}\mathcal{S}^+(p)$ at point $S_1 \in \mathcal{S}^+(p)$ to the tangent space $T_{S_2}\mathcal{S}^+(p)$ at point $S_2 \in \mathcal{S}^+(p)$, and its mathematical formula is given by:

$$\mathcal{P}_{S_1 \rightarrow S_2}(V_1) = W V_1 W^T. \quad (2.18)$$

Here, $W = S_1^{1/2} \exp\left(\frac{1}{2} S_1^{-1/2} \text{Log}_{S_2}(S_1) S_1^{-1/2}\right) S_1^{-1/2}$, and it can be simplified to $W = (S_2 S_1^{-1})^{1/2}$ (Yair *et al.*, 2019).

An orthonormal basis in $T_S\mathcal{S}^+(p)$ for each $S \in \mathcal{S}^+(p)$ can be defined via parallel transport of a choice of orthonormal basis in the tangent space at the identity matrix $I_p \in T_{I_p}\mathcal{S}^+(p)$. The matrices $\{E_r : r = 1, \dots, m = p(p+1)/2\}$, defined in Equation (2.9), determine an orthonormal basis in the tangent space at I_p with respect to the Riemannian inner product in Equation (2.12). Therefore, given any point $S \in \mathcal{S}^+(p)$, we can obtain an orthonormal basis in $T_S\mathcal{S}^+(p)$ using the Equation (2.18) as

$$E_r^S = \mathcal{P}_{I_p \rightarrow S}(E_r) = S^{1/2} E_r S^{1/2}. \quad (2.19)$$

This basis can be thought of as a global smooth section of the frame bundle of the Riemannian manifold $(\mathcal{S}^+(p), g^{\text{aff}})$. By definition, it follows that $(\mathcal{S}^+(p), g^{\text{aff}})$ is parallelizable.

For statistical operations, we can define the classical operator Vec that maps the sym-

metric tangent vectors into $m = p(p+1)/2$ dimensional coordinate vectors which realizes an explicit isomorphism between $T_S\mathcal{S}^+(p)$ and $\mathbb{R}^m$ with the canonical metric (Pennec *et al.*, 2006). Thereby, we have the following definition of vectorizing projection of Vec .

Definition 2.1. *Given any tangent vector $V \in T_S\mathcal{S}^+(p)$ at the base of $S \in \mathcal{S}^+(p)$, we can calculate its orthonormal coordinate vector $\mathbf{v} = (v_1, \dots, v_m)^T$ in $\mathbb{R}^m$ by the vectorization operator $Vec(V)$ as*

$$\begin{aligned} v_r &= g_S^{\text{aff}} \langle V, E_r^S \rangle \\ &= \langle V, S^{1/2} E_r S^{1/2} \rangle_S \\ &= \text{Tr}(S^{-1/2} V S^{-1/2} E_r), \end{aligned} \tag{2.20}$$

where E_r is the orthonormal basis in $T_{I_p}\mathcal{S}^+(p)$ as shown in Equation (2.9) and $r = 1, \dots, m$.

2.3.4 Fréchet sample mean and variance

To obtain the intrinsic mean in a Riemannian manifold, we can find it by minimizing the Fréchet function $\mathcal{F}$ as defined in Equation (2.3), giving $\tilde{x} \in \arg \min_y \mathcal{F}(y, \{x_i\})$. It has been proved that the Fréchet mean exists and is unique when the smooth manifold has non-positive sectional curvature (Karcher, 1977), which is the case for the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$ (Skovgaard, 1984; Kendall, 1990). Therefore, given a sample of points $\{S_i : i = 1, \dots, n\} \in \mathcal{S}^+(p)$, the Fréchet sample mean in the affine invariant manifold could be obtained by minimizing the Fréchet function using the geodesic distance $d_{g^{\text{aff}}}$ as

$$\tilde{S} = \arg \min_{\tilde{S}} \frac{1}{n} \sum_{i=1}^n d_{g^{\text{aff}}}(S_i, \tilde{S})^2. \tag{2.21}$$

Generally, this can be derived by the numerical gradient descent algorithm with the step operator $\tilde{S}_{k+1} = \tilde{S}_k - \delta \nabla \mathcal{F}$, where δ is the step size and $\nabla \mathcal{F}$ is the derivative of the Fréchet function $\mathcal{F}$. More detailed processes of this can be found in Lenglet *et al.* (2006). We show the final step operator for computing the Fréchet sample mean in $\mathcal{S}^+(p)$ as

$$\begin{aligned} \tilde{S}_{k+1} &= \text{Exp}_{\tilde{S}_k} \left(-\frac{\delta}{n} \sum_{i=1}^n \text{Log}_{\tilde{S}_k}(S_i) \right) \\ &= \tilde{S}_k^{1/2} \exp \left(\frac{\delta}{n} \sum_{i=1}^n \log \left(\tilde{S}_k^{-1/2} S_i \tilde{S}_k^{-1/2} \right) \right) \tilde{S}_k^{1/2} \end{aligned} \tag{2.22}$$

whose associated flow converges toward the Fréchet sample mean for initial guess $\tilde{S}_0 \in \mathcal{S}^+(p)$, and $\delta \in [0, 1]$ is the step size, which should be small. The geometrical interpretation

of the gradient approach for the computation of the Fréchet sample mean can be referred to in Kume & Le (2003).

Another distinctive approach to compute the Fréchet sample mean involves utilizing Sturm's algorithm (Sturm, 2003). This method iteratively computes the sample mean along the geodesics γ , connecting any point in the data set $\{S_i : i = 1, \dots, n\} \in \mathcal{S}^+(p)$ to the iterated point $\tilde{S}_k$, and the iterated point $\tilde{S}_k$ eventually converges to the Fréchet sample mean (See Algorithm 1).

Algorithm 1 Sturm's algorithm for Fréchet sample mean

Require: Data $\{S_i : i = 1, \dots, n\} \in \mathcal{S}^+(p)$, initial guess $\tilde{S}_0 \in \mathcal{S}^+(p)$, and maximum iterations *max.ites*.

$k = 1$

while $k > \text{max.ites}$ or $\left| \frac{1}{n} \sum_{i=1}^n d_{g^{\text{aff}}}(S_i, \tilde{S}_k)^2 - \frac{1}{n} \sum_{i=1}^n d_{g^{\text{aff}}}(S_i, \tilde{S}_{k+1})^2 \right| < 10^{-4}$ **do**

 Select a random point $S_0 \in \mathcal{S}^+(p)$ from the data set $\{S_i : i = 1, \dots, n\}$.

 Update the Fréchet sample mean $\tilde{S}_k = \gamma(t_k)$, where $\gamma(t_k)$ is the geodesic connecting the iterated mean $\tilde{S}_k$ and S_0 at t_k , where $t_k = 1/(1+k)$.

 Update $k = k + 1$.

end while

We implement the gradient descent algorithm using the step operator in Equation (2.22) and Sturm's algorithm (Algorithm 1) on a given data set $\{S_i : i = 1, \dots, n\} \in \mathcal{S}^+(p)$. As an example, we use the reduced seizure data of 15×15 full-rank covariance matrices with the time length of 116 seconds in $\mathcal{S}^+(15)$ from Seizure 1 within Patient ID18 (Data information will be provided in Chapter 3). The step sizes in the gradient descent algorithm are set as $dt = 0.1, 0.25, 0.5, 0.75, 1, 1.25$, and 1.5 . Iterations continue until either the iteration count reaches the maximum iteration steps (set to *max.ites* = 200) or the difference between consecutive values of the Fréchet function is small enough (set to 10^{-4}), represented as $\left| \frac{1}{n} \sum_{i=1}^n d_{g^{\text{aff}}}(S_i, \tilde{S}_k)^2 - \frac{1}{n} \sum_{i=1}^n d_{g^{\text{aff}}}(S_i, \tilde{S}_{k+1})^2 \right| < 10^{-4}$. Figure 2.2 displays iterations and their corresponding values of the Fréchet function $\mathcal{F}$ under different parameter settings. Sturm's algorithm fails to converge within the maximum iteration limit. The step size reaches its breakdown point at 1.5 within the given range of dt . The remaining curves of $\mathcal{F}$ achieve convergence at different iteration steps. The x -axis records the final iterative steps when the convergence criterion is met. For instance, the first (red) line with a step size of $dt = 0.1$ converges after the 54th iteration.

To identify an appropriate algorithm and optimal step size, we considered the convergence time. After implementing the aforementioned algorithms, we observed the following iteration time lengths: 31.03s, 12.97s, 6.43s, 3.57s, 5.25s, 24.56s, 108.98s for the different step size dt , and 34.74s for Sturm's algorithm, respectively. To mitigate the risk of local minimization of the objective function and to save computing time, in this research, we prefer to select a step size of $dt = 0.75$ for the gradient descent algorithm to compute the

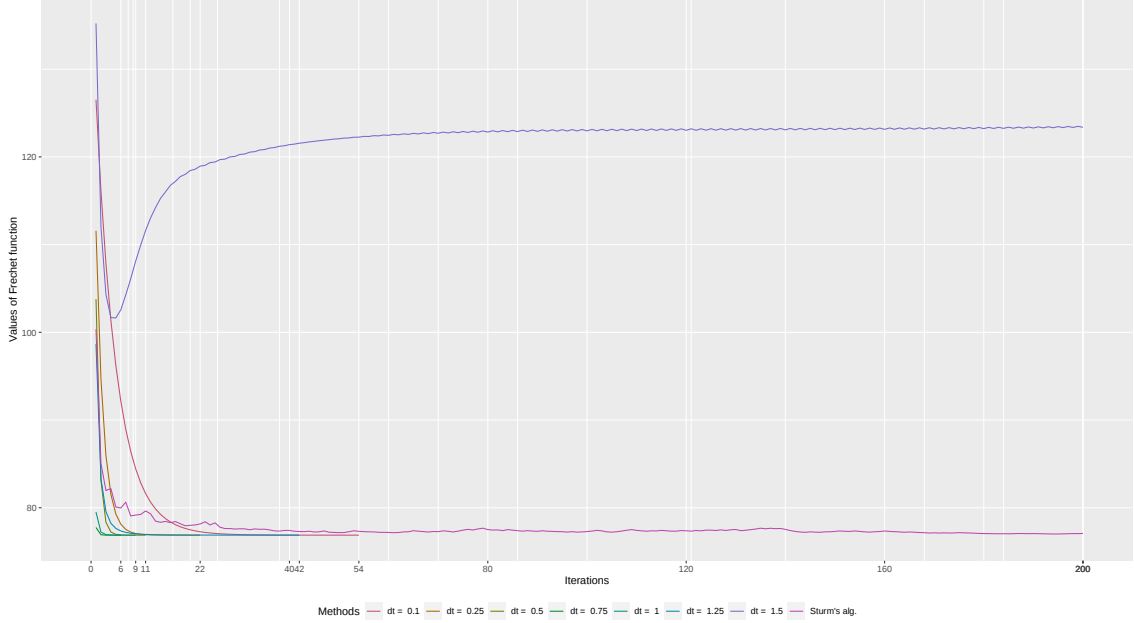


Figure 2.2: Values of Fréchet function for the gradient descent algorithm with various step sizes, ranging from $dt = 0.1$ to 1.5 and Sturm's algorithm. The x -axis represents iterative steps, and the y -axis represents the values of Fréchet function $\mathcal{F}$. The ticks on the x -axis indicate the iterative steps at which the algorithm converged.

Fréchet sample mean in the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$ across all seizure data sets. Readers may choose other parameters within the range of 0.1 to 1.25 depending on their real data.

2.4 The quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$

In this section, we will describe a newly developed Riemannian manifold consisting of the set of full-rank $p \times p$ correlation matrices, named as the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$. This includes basic concepts of the Riemannian metric g^{quo} , exponential and logarithm maps, parallel transport, orthonormal coordinate system, and geometrical statistics within this space, analogous to the covariance matrix space of $(\mathcal{S}^+(p), g^{\text{aff}})$. To aid readers in understanding the fundamental concepts of the quotient manifold, we briefly elaborate on the concepts of Riemannian submersion, a mapping that takes the manifold $(\mathcal{S}^+(p), g^{\text{aff}})$ to the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$ along with its corresponding properties. Additional detailed information can be found in Section 18.3 and Section 23.3 of the book by Gallier & Quaintance (2020) and in a PhD thesis by (Lueg, 2023, Chapter 3). It is assumed that readers possess basic knowledge of differential geometry, Lie groups, and group actions. Alternatively, relevant background information can be found in the aforementioned book and thesis.

2.4.1 Riemannian submersion and quotient manifold

Riemannian submersion

Here, we briefly recall the definition of Riemannian submersion.

Definition 2.2. Let $\phi : \mathcal{M} \rightarrow \mathcal{N}$ be a smooth map of Riemannian manifolds from $(\mathcal{M}, g)$ to $(\mathcal{N}, h)$.

- (1) The smooth map ϕ is a submersion of $\mathcal{M}$ into $\mathcal{N}$ iff the derivative $d\phi_x$ is surjective for all $x \in \mathcal{M}$.
- (2) The smooth map ϕ is a local isometry between tangent spaces $(T_x\mathcal{M}, g_x)$ and $(T_x\mathcal{N}, h_{\phi(x)})$ such that every $x \in \mathcal{M}$ and every $u, v \in T_x\mathcal{M}$,

$$\langle u, v \rangle_x = \langle d\phi_x(u), d\phi_x(v) \rangle_{\phi(x)},$$

And, ϕ is an isometry iff it is a local isometry and a diffeomorphism.

Then, ϕ is a Riemannian submersion from Riemannian manifold $(\mathcal{M}, g)$ to Riemannian manifold $(\mathcal{N}, h)$ if the above properties hold.

Quotient manifold on $(\mathcal{C}^+(p), g^{\text{quo}})$

Now, we will see that Riemannian submersion arises when $\mathcal{N}$ is obtained from a free and proper action of a Lie group acting by isometries on $\mathcal{M}$. The definitions related to Lie groups and group actions can be found in the textbook by Gallier & Quaintance (2020).

Theorem 2.1. Given a Riemannian manifold $(\mathcal{M}, g)$, let $\cdot : G \times \mathcal{M} \rightarrow \mathcal{M}$ be a smooth, free, and proper action, with G a Lie group acting by isometries of $\mathcal{M}$. Then there is a unique Riemannian metric on $\mathcal{N} = \mathcal{M}/G$ such that $\pi : \mathcal{M} \rightarrow \mathcal{N}$ is a Riemannian submersion.

The proof can be found in (Gallier & Quaintance, 2020, Theorem 23.14). As a case of a Riemannian submersion, the *quotient manifold* $\mathcal{N}$ is obtained from a Lie group G acting on a Riemannian manifold $\mathcal{M}$ by isometries. Considering the covariance matrices equipped with the affine invariant manifold $(\mathcal{S}^+(p), g^{\text{aff}})$, the set of correlation matrices generates a quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$ through the Riemannian submersion $\pi : \mathcal{S}^+(p) \rightarrow \mathcal{C}^+(p)$. We have the following definition.

Definition 2.3. The Lie group $\mathcal{D}^+(p)$ of diagonal matrices with positive entries acts smoothly, properly, and freely on the affine invariant manifold $(\mathcal{S}^+(p), g^{\text{aff}})$ via the action $\phi : \mathcal{D}^+(p) \times \mathcal{S}^+(p) \rightarrow \mathcal{S}^+(p)$ at every point $S \in \mathcal{S}^+(p)$:

$$\phi_D(S) = DSD,$$

where $D \in \mathcal{D}^+(p)$. Subsequently, the quotient manifold $\mathcal{C}^+(p) = \mathcal{S}^+(p)/\mathcal{D}^+(p)$, resulting from the Riemannian submersion $\pi : \mathcal{S}^+(p) \rightarrow \mathcal{C}^+(p) = \mathcal{S}^+(p)/\mathcal{D}^+(p)$, is a smooth manifold with a unique metric.

For example, a full-rank correlation matrix is calculated from a covariance matrix by

$$\pi : S \in \mathcal{S}^+(p) \rightarrow C = \Delta_S^{-1} S \Delta_S^{-1} \in \mathcal{C}^+(p),$$

where $\Delta_S = \sqrt{\text{Diag}(S)}$ and $\text{Diag}(S) \in \mathcal{D}^+(p)$ is the diagonal matrix of S (David, 2019, Theorem 2.1). For every $C \in \mathcal{C}^+(p)$, $\pi^{-1}(C) \subset \mathcal{S}^+(p)$ is the *fibre/orbit* of S , i.e., the point S in the fibre is computed by $S \in \pi^{-1}(C) = \{DCD \mid D \in \mathcal{D}^+(p)\}$. Notably, the map π as defined above is a surjective submersion (Lueg, 2023, Theorem 3.5.1).

Decomposition of the tangent space $T_S \mathcal{S}^+(p)$

Generally, given a tangent space $T_x \mathcal{M}$ at $x \in \mathcal{M}$, it can be decomposed into two subspaces:

$$T_x \mathcal{M} = \mathcal{V}_x \oplus \mathcal{H}_x,$$

where $\mathcal{H}_x$ and $\mathcal{V}_x$ are *horizontal* and *vertical* subspaces of $T_x \mathcal{M}$, respectively. They are orthogonal, i.e., $\mathcal{H}_x = (\mathcal{V}_x)^\perp$. This means any tangent vector $v \in T_x \mathcal{M}$ can be decomposed into the vertical component w in the vertical subspace $\mathcal{V}_x$ and the horizontal component u in the horizontal subspace $\mathcal{H}_x$, i.e., $v = w + u$.

A tangent vector $v \in T_x \mathcal{M}$ is said to be *horizontal* if and only if $v \in \mathcal{H}_x$ (equivalently, if $w = 0$).

Because π is a Riemannian submersion, $d\pi_x$ gives a linear isomorphism between $\mathcal{H}_x$ and $T_{\pi(x)} \mathcal{N}$, and the differential geometry of $\mathcal{N}$ can be studied by ‘lifting’ from $\mathcal{N}$ to $\mathcal{M}$ (Gallier & Quaintance, 2020, Section 18.3). The following commutative diagram shows the relations between different manifolds and tangent spaces, taking an example of $\pi : \mathcal{S}^+(p) \rightarrow \mathcal{C}^+(p)$:

$$\begin{array}{ccc} \mathcal{S}^+(p) & \xrightarrow{\pi} & \mathcal{C}^+(p) \\ \downarrow \text{Log.} & & \downarrow \\ T_S \mathcal{S}^+(p) = \mathcal{V}_S \oplus \mathcal{H}_S & \xrightarrow{d\pi_S} & T_C \mathcal{C}^+(p) \end{array}$$

Specifically, given any point $S \in \mathcal{S}^+(p)$ and Riemannian submersion π , we can find the quotient manifold $\mathcal{C}^+(p)$ and corresponding point $C = \pi(S) \in \mathcal{C}^+(p)$ to S . After performing the logarithm map at the base S , a tangent vector $V \in T_S \mathcal{S}^+(p)$ splits into a vertical component W in the vertical subspace $\mathcal{V}_S$ and a horizontal component U in the horizontal subspace $\mathcal{H}_S$ (the orthogonal complement of $\mathcal{V}_S$, i.e., $V = W + U$). Figure 2.3 gives an illustration of Riemannian submersion from the affine invariant manifold $(\mathcal{S}^+(p), g^{\text{aff}})$ to the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$ and the decomposition of tangent vectors $V \in T_S \mathcal{S}^+(p)$.

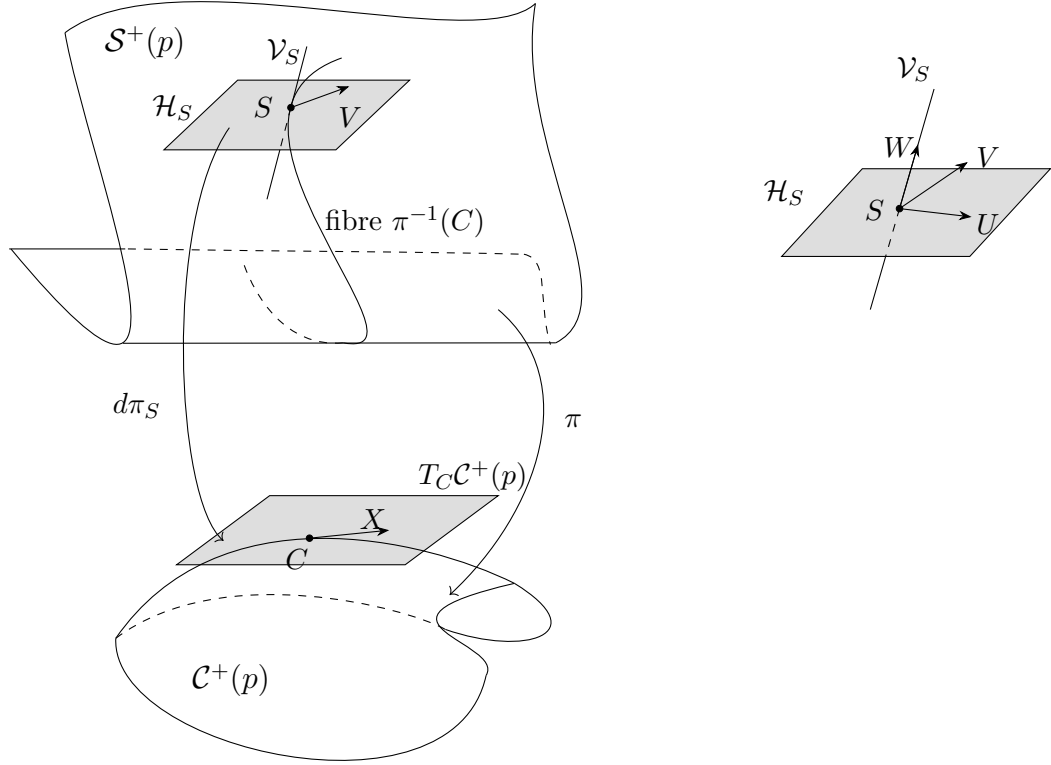


Figure 2.3: Illustration of Riemannian submersion and the decomposition of the tangent space $T_S \mathcal{S}^+(p)$. The submersion π maps any point $S \in \mathcal{S}^+(p)$ to $C = \pi(S) \in \mathcal{C}^+(p)$. $d\pi_S$ is the derivative of π at S , which projects the tangent vector $V \in T_S \mathcal{S}^+(p)$ to $X \in T_C \mathcal{C}^+(p)$. The grey colored horizontal subspace $\mathcal{H}_S$ is isomorphic to the grey tangent space $T_C \mathcal{C}^+(p)$, and S is an arbitrary element in the fiber $\pi^{-1}(C)$. The tangent vector $V \in T_S \mathcal{S}^+(p)$ can be decomposed into the horizontal component $U \in \mathcal{H}_S$ and the vertical component $W \in \mathcal{V}_S$.

Lemma 2.1. *Given any tangent vector $V \in T_S \mathcal{S}^+(p)$, the vertical and horizontal components of V are computed as follows:*

- (1) *The vertical component $W \in \mathcal{V}_S$:*

$$W = D_p S + S D_p$$

where $D_p = \text{Diag}((S^{-1} \circ S + I_p)^{-1} \text{diag}(S^{-1} V))$. Here, $\text{diag}(\cdot)$ is the diagonal vector of matrix $(\cdot)$, $\text{Diag}(\cdot)$ is the diagonal matrix with diagonals $(\cdot)$, and operation $\circ$ is the Hadamard product on matrices defined by $[A \circ B]_{ij} = A_{ij} B_{ij}$.

- (2) *The horizontal component $U \in \mathcal{H}_S$ is the complement of W :*

$$U = V - W$$

Proof. (2) is obvious as it follows the definition of the decomposition of $V \in T_S \mathcal{S}^+(p)$. Now, we prove the (1).

Let $\gamma(t)$ be a geodesic in the affine invariant manifold $(\mathcal{S}^+(p), g^{\text{aff}})$, $t \in [0, 1]$ and $\gamma(0) = S$, $V = \gamma'(0) \in T_S \mathcal{S}^+(p)$ be the velocity of the geodesic at the point S . π is a Riemannian submersion between $\mathcal{S}^+(p)$ and $\mathcal{C}^+(p)$. We can compute the derivative of π at S with respect to V as:

$$\begin{aligned}
 d\pi_S(V) &= \left. \frac{\partial \pi(\gamma(t))}{\partial t} \right|_{t=0} \\
 &= \left. \frac{\partial \sqrt{\text{Diag}(\gamma(t))}^{-1} \gamma(t) \sqrt{\text{Diag}(\gamma(t))}^{-1}}{\partial t} \right|_{t=0} \\
 &= \left(\frac{\partial \gamma_{ii}^{-\frac{1}{2}}(t) \gamma_{ij}(t) \gamma_{jj}^{-\frac{1}{2}}(t)}{\partial t} \right)_{i,j=1}^p \Big|_{t=0} \\
 &= \left(\frac{\partial \left(-\frac{1}{2} \gamma_{ii}^{-\frac{3}{2}}(t) \gamma'_{ii}(t) \gamma_{ij}(t) \gamma_{jj}^{-\frac{1}{2}}(t) \right)}{\partial t} \right)_{i,j=1}^p \Big|_{t=0} + \\
 &\quad \left(\frac{\partial \gamma_{ii}^{-\frac{1}{2}}(t) \gamma'_{ij}(t) \gamma_{jj}^{-\frac{1}{2}}(t)}{\partial t} \right)_{i,j=1}^p \Big|_{t=0} + \left(\frac{\partial \left(-\frac{1}{2} \gamma_{ii}^{-\frac{1}{2}}(t) \gamma_{ij}(t) \gamma_{jj}^{-\frac{3}{2}}(t) \gamma'_{jj}(t) \right)}{\partial t} \right)_{i,j=1}^p \Big|_{t=0} \\
 &= -\frac{1}{2} \Delta_S^{-3} \text{Diag}(V) S \Delta_S^{-1} + \Delta_S^{-1} V \Delta_S^{-1} - \frac{1}{2} \Delta_S^{-1} S \Delta_S^{-3} \text{Diag}(V) \\
 &= \Delta_S^{-1} \left[V - \frac{1}{2} (\Delta_S^{-2} \text{Diag}(V) S + S \text{Diag}(V) \Delta_S^{-2}) \right] \Delta_S^{-1}, \tag{2.23}
 \end{aligned}$$

where $\Delta_S = \sqrt{\text{Diag}(S)}$.

Since $d\pi_S$ gives a linear isomorphism between $\mathcal{H}_S$ and $T_S \mathcal{S}^+(p)$ by Riemannian submersion, we have

$$\begin{aligned}
 d\pi_S(V) &= d\pi_S(W + U) = d\pi_S(U), \text{ and} \\
 d\pi_S(W) &= 0,
 \end{aligned}$$

where U is the horizontal component in the horizontal subspace $\mathcal{H}_S$ and W is the vertical component in the vertical subspace $\mathcal{V}_S$. Thereby, for any point $S \in \mathcal{S}^+(p)$, a characteri-

zation of the vertical subspace $W \in \mathcal{V}_S$ is given by

$$\begin{aligned}
 W \in \mathcal{V}_S &\iff d\pi_S(W) = 0 \\
 &\iff \Delta_S^{-1}[W - \frac{1}{2}(\Delta_S^{-2}\text{Diag}(W)S + S\text{Diag}(W)\Delta_S^{-2})]\Delta_S^{-1} = 0 \\
 &\iff W = \frac{1}{2}(\text{Diag}(S)^{-1}\text{Diag}(W)S + S\text{Diag}(W)\text{Diag}(S)^{-1}) \\
 &\iff W_{ij} = \frac{1}{2}S_{ij} \left(\frac{W_{ii}}{S_{ii}} + \frac{W_{jj}}{S_{jj}} \right), \text{ for all } i, j = 1, \dots, p \\
 &\iff W \in \{D_p S + S D_p \mid D_p \in \mathcal{D}(p)\},
 \end{aligned}$$

where D_p is a diagonal matrix in *diagonal matrix space* $\mathcal{D}(p)$, which is the set of diagonal matrices.

Furthermore, a characterization of the horizontal component $U \in \mathcal{H}_S$ is given by

$$\begin{aligned}
 U \in \mathcal{H}_S &\iff \text{Tr}(S^{-1}W S^{-1}U) = 0 \\
 &\iff \text{Tr}(S^{-1}(D_p S + S D_p)S^{-1}U) = 0 \text{ for all } D_p \in \mathcal{D}(p), \\
 &\iff \text{Tr}(D_p U S^{-1} + D_p S^{-1}U) = 0 \\
 &\iff U S^{-1} + S^{-1}U \in \text{Hol}(p) \Leftrightarrow S^{-1}U \in \text{Hol}(p) \Leftrightarrow U S^{-1} \in \text{Hol}(p)
 \end{aligned}$$

where $\text{Hol}(p)$ is the vector space of symmetric matrices with vanishing diagonals (*symmetric hollow matrices*), i.e., $\text{Hol}(p) = \{H = (h_{ij})_{i,j=1}^p \in \text{Sym}(p) : h_{ii} = 0, i = 1, \dots, p\}$. Hence, $U \in \mathcal{H}_S$ iff it is the unique solution to Sylvester's equation for some $H \in \text{Hol}(p)$ (Sylvester, 1884), such that, $S^{-1}U + U S^{-1} = H$.

Since $V = W + U$, it also can be written as $V = (D_p S + S D_p) + U$. We left multiply matrix S^{-1} to both sides of this equation and consider the diagonal vector $\text{diag}(\cdot) \in \mathbb{R}^p$ as

$$\begin{aligned}
 V &= (D_p S + S D_p) + U \\
 &\iff \text{diag}(S^{-1}V) = \text{diag}(S^{-1}(D_p S + S D_p) + S^{-1}U) \\
 &\iff \text{diag}(S^{-1}V) = \text{diag}(S^{-1}D_p S + D_p) \quad (\text{as } S^{-1}U \in \text{Hol}(p)) \\
 &\iff \text{diag}(S^{-1}V) = (S^{-1} \circ S + I_p)\text{diag}(D_p) \\
 &\iff \text{diag}(D_p) = (S^{-1} \circ S + I_p)^{-1}\text{diag}(S^{-1}V),
 \end{aligned}$$

where operation $\circ$ is the Hadamard product on matrices. Therefore, $D_p = \text{Diag}((S^{-1} \circ S + I_p)^{-1}\text{diag}(S^{-1}V))$ and $W = D_p S + S D_p$ as proved. $\square$

Simply, we also represent the horizontal component of V as $U = \text{hor}(V) \in \mathcal{H}_S$ and the vertical component of V as $W = \text{ver}(V) \in \mathcal{V}_S$ such that $V = \text{hor}(V) + \text{ver}(V)$ for every $V \in T_S \mathcal{S}^+(p)$. When V is a horizontal vector, $V = \text{hor}(V)$, i.e. $\text{ver}(V) = 0$.

Horizontal lift

Because π is a submersion, $d\pi_S$ gives a linear isomorphism between $\mathcal{H}_S$ and $T_{\pi(S)}\mathcal{C}^+(p)$ for every $S \in \mathcal{S}^+(p)$ (Gallier & Quaintance, 2020, Definition of 18.3). This implies we can study geometry on $\mathcal{C}^+(p)$ by ‘lifting’ from $\mathcal{C}^+(p)$ to $\mathcal{S}^+(p)$ (see Figure 2.4).

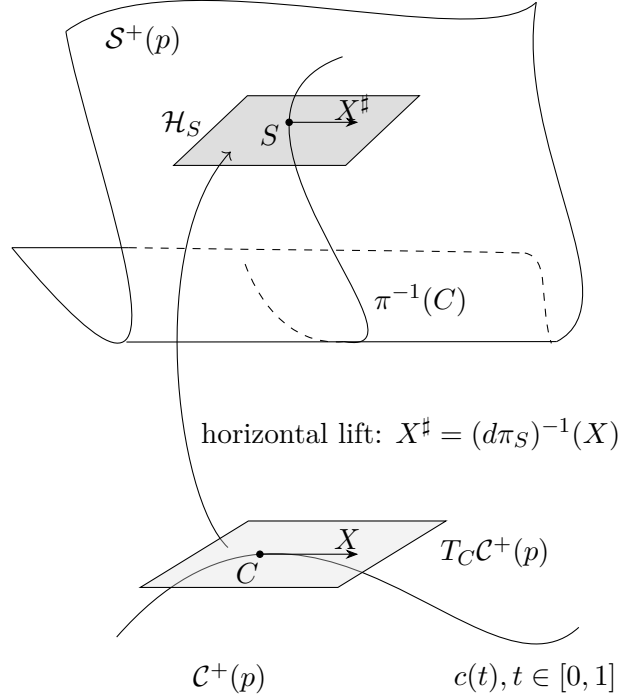


Figure 2.4: An illustration of the horizontal lift. Given a point $C \in \mathcal{C}^+(p)$ and its tangent vector $X \in T_C \mathcal{C}^+(p)$, we can obtain an element S on the fibre $\pi^{-1}(C) \subset \mathcal{S}^+(p)$. The horizontal lift $X^\#$ is computed by pulling X back into the horizontal subspace $\mathcal{H}_S$. Note that $c(t), t \in [0, 1]$ is the geodesic in $\mathcal{C}^+(p)$.

Definition 2.4. Given a Riemannian submersion $\pi : \mathcal{S}^+(p) \rightarrow \mathcal{C}^+(p)$ and an isomorphism $d\pi_S$ between $\mathcal{H}_S$ and $T_{\pi(S)}\mathcal{C}^+(p)$, for every $C \in \mathcal{C}^+(p)$, $S \in \pi^{-1}(C) \subset \mathcal{S}^+(p)$, and $X \in T_C \mathcal{C}^+(p)$, the unique horizontal lift $X^\#$ in horizontal subspace $\mathcal{H}_S$ at $S \in \mathcal{S}^+(p)$ is defined as

$$X^\# = \text{hor}(\Delta_S X \Delta_S) \in \mathcal{H}_S. \quad (2.24)$$

Recalling the first derivative of submersion π at S as shown in (2.23), we have :

$$\begin{aligned}
 d\pi_S(X^\sharp) &= d\pi_S(\text{hor}(\Delta_S X \Delta_S)) \\
 &\stackrel{d\pi_S(\text{ver}(\Delta_S X \Delta_S))=0}{=} d\pi_S(\Delta_S X \Delta_S) \\
 &= \Delta_S^{-1} [\Delta_S X \Delta_S - \frac{1}{2} (\Delta_S^{-2} \text{Diag}(\Delta_S X \Delta_S) S + S \text{Diag}(\Delta_S X \Delta_S) \Delta_S^{-2})] \Delta_S^{-1} \\
 &\stackrel{\text{Diag}(X)=0}{=} X - \frac{1}{2} \Delta_S^{-1} (\underbrace{\Delta_S^{-2} \text{Diag}(\Delta_S X \Delta_S)}_{=0} S + S \underbrace{\text{Diag}(\Delta_S X \Delta_S) \Delta_S^{-2}}_{=0}) \Delta_S^{-1} \\
 &= X
 \end{aligned}$$

as shown in Definition 2.4. Notably, in the forth line, we use the condition that $X \in \text{Hol}(p)$. This could be verified as following:

$$\begin{aligned}
 \text{Diag}(X) &= \text{Diag}(d\pi_S(V)) \\
 &= \text{Diag} \left(\left. \frac{\partial \pi(\gamma(t))}{\partial t} \right|_{t=0} \right) \\
 &= \left(\left. \frac{\partial \gamma_{ii}^{-\frac{1}{2}}(t) \gamma_{ii}(t) \gamma_{ii}^{-\frac{1}{2}}(t)}{\partial t} \right|_{t=0} \right)^p_{i=1} = 0.
 \end{aligned}$$

This illustrates that each tangent space in the quotient manifold can be seen as a copy of the space $\text{Hol}(p)$ consisting of symmetric hollow matrices.

In particular, the horizontal lift at $C \in \mathcal{S}^+(p)$ is $X^\sharp = \text{hor}(X)$.

2.4.2 The quotient metric

For any $C \in \mathcal{C}^+(p)$, $X, Y \in T_C \mathcal{C}^+(p)$, and $S \in \pi^{-1}(C)$ in $\mathcal{S}^+(p)$, there exists a unique horizontal lift $X^\sharp$ and $Y^\sharp$, both in the horizontal subspace $\mathcal{H}_S$, such that $d\pi_S(X^\sharp) = X$ and $d\pi_S(Y^\sharp) = Y$. Since $d\pi_S$ is a linear isomorphism between the horizontal subspace $\mathcal{H}_S$ and tangent space $T_S \mathcal{C}^+(p)$, we have the following formula for the quotient metric g^{quo} :

Definition 2.5. *The unique quotient metric in the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$ is defined as*

$$g_C^{\text{quo}} \langle X, Y \rangle = g_S^{\text{aff}} \langle X^\sharp, Y^\sharp \rangle, \quad (2.25)$$

which makes $T_C \mathcal{C}^+(p)$ isometric to $\mathcal{H}_S$.

Theorem 2.2. *The quotient metric g_C^{quo} does not depend on the choice of S in the fibre $\pi^{-1}(C) \subset \mathcal{S}^+(p)$.*

Proof. Firstly, we recall the following properties of Lie group action and Riemannian submersion:

- (1) Recalling the group action $\phi : \mathcal{S}^+(p) \times \mathcal{D}^+(p) \rightarrow \mathcal{S}^+(p)$ acts smoothly and transitively on $\mathcal{S}^+(p)$, for any $S_1, S_2 \in \pi^{-1}(C)$, there exists $D \in \mathcal{D}^+(p)$ such that $\phi_D(S_1) = DS_1D = S_2$, and thus $(d\phi_D)_{S_1}(V) = DVD$, where V is a tangent vector in $T_{S_1}\mathcal{S}^+(p)$. Hence, for given a point $S_1 \in \mathcal{S}^+(p)$ and two tangent vectors $V, W \in T_{S_1}\mathcal{S}^+(p)$, we have

$$\begin{aligned} g_{\phi_D(S_1)}^{\text{aff}} \langle (d\phi_D)_{S_1}(V), (d\phi_D)_{S_1}(W) \rangle &= \text{Tr}[(DS_1D)^{-1}DVD(DS_1D)^{-1}DWD] \\ &= \text{Tr}[S_1^{-1}VS_1^{-1}W] \\ &= g_{S_1}^{\text{aff}} \langle V, W \rangle. \end{aligned}$$

This also illustrates the diffeomorphism $\phi_D : \mathcal{S}^+(p) \rightarrow \mathcal{S}^+(p)$ is an isometry $((d\phi_D)_{S_1} : T_{S_1}\mathcal{S}^+(p) \rightarrow T_{\phi_D(S_1)}\mathcal{S}^+(p))$ for all $S_1 \in \mathcal{S}^+(p)$.

- (2) The chain rule gives $(d\pi)_{\phi_D(S_1)} \circ (d\phi_D)_{S_1} = (d\pi)_{S_1}$ (Lueg, 2023).

Therefore, given any $C \in \mathcal{C}^+(p)$ and tangent vectors $X, Y \in T_C\mathcal{C}^+(p)$, we have

$$\begin{aligned} g_{S_1}^{\text{aff}} \langle X^\sharp, Y^\sharp \rangle &= g_{\phi_D(S_1)} \langle (d\phi_D)_{S_1}(X^\sharp), (d\phi_D)_{S_1}(Y^\sharp) \rangle \\ &= g_{\phi_D(S_1)} \langle (d\phi_D)_{S_1}((d\pi_{S_1})^{-1}(X)), (d\phi_D)_{S_1}((d\pi_{S_1})^{-1}(Y)) \rangle \\ &= g_{\phi_D(S_1)} \langle ((d\pi)_{\phi_D(S_1)})^{-1}(X), ((d\pi)_{\phi_D(S_1)})^{-1}(Y) \rangle \\ &= g_{\phi_D(S_1)} \langle X^\sharp, Y^\sharp \rangle, \end{aligned}$$

where $^\sharp$ is the horizontal lift from $\mathcal{C}^+(p)$ to $\mathcal{S}^+(p)$, and two points S_1 and $\phi_D(S_1)$ both lie in the fibre $\pi^{-1}(C)$. It follows that the quotient invariant metric $(\mathcal{C}^+(p), g^{\text{quo}})$ in Equation (2.25) does not depend on the chosen $S \in \pi^{-1}(C)$. $\square$

Subsequently, given any $C \in \mathcal{C}^+(p)$ and tangent vectors $X, Y \in T_C\mathcal{C}^+(p)$, the quotient metric is defined by pull-backing through the horizontal lift:

$$\begin{aligned} g_C^{\text{quo}} \langle X, Y \rangle &= g_S^{\text{aff}} \langle X^\sharp, Y^\sharp \rangle \\ &= g_S^{\text{aff}} \langle \text{hor}(\Delta_S X \Delta_S), \text{hor}(\Delta_S Y \Delta_S) \rangle \\ &= g_S^{\text{aff}} \langle \Delta_S X \Delta_S - \text{ver}(\Delta_S X \Delta_S), \Delta_S Y \Delta_S - \text{ver}(\Delta_S Y \Delta_S) \rangle \\ &= g_S^{\text{aff}} \langle \Delta_S X \Delta_S, \Delta_S Y \Delta_S \rangle - g_S^{\text{aff}} \langle \Delta_S X \Delta_S, \text{ver}(\Delta_S Y \Delta_S) \rangle \\ &\quad - g_S^{\text{aff}} \langle \Delta_S Y \Delta_S, \text{ver}(\Delta_S X \Delta_S) \rangle + g_S^{\text{aff}} \langle \text{ver}(\Delta_S X \Delta_S), \text{ver}(\Delta_S Y \Delta_S) \rangle \\ &= g_S^{\text{aff}} \langle \Delta_S X \Delta_S, \Delta_S Y \Delta_S \rangle - g_S^{\text{aff}} \langle \text{ver}(\Delta_S X \Delta_S), \text{ver}(\Delta_S Y \Delta_S) \rangle. \end{aligned}$$

Here, S is the any point in the fibre $\pi^{-1}(C) \subset \mathcal{S}^+(p)$.

Geodesic distance

Lemma 2.2. *If γ is a geodesic in $\mathcal{S}^+(p)$ such that $\gamma'(0)$ is a horizontal vector, then γ is horizontal geodesic in $\mathcal{S}^+(p)$ (which means that $\gamma'(t)$ is a horizontal vector for all t), and $c(t) = \pi(\gamma(t))$ is a geodesic in $\mathcal{C}^+(p)$ of the same length with $\gamma(t), t \in [0, 1]$.*

The proof can be seen in (Huckemann *et al.*, 2010). The specific computation equation will be introduced in Equation (2.32).

In case of an isometric action ϕ_D , we have that any horizontal geodesic segment γ joining S_1 and S_2 in $\mathcal{S}^+(p)$ has the same length as the geodesic segment $c = \pi(\gamma)$ joining $\pi(S_1)$ and $\pi(S_2)$. Hence the geodesic distance in $\mathcal{C}^+(p)$ can be computed by

$$d_{g^{\text{quo}}}(\pi(S_1), \pi(S_2)) = d_{g^{\text{aff}}}(S_1, S_2)$$

for any $S_1, S_2 \in \mathcal{S}^+(p)$, and thus

$$d_{g^{\text{quo}}}(\pi(S_1), \pi(S_2)) = \inf_{D \in \mathcal{D}^+(p)} d_{g^{\text{aff}}}(S_1, \phi_D(S_2)).$$

where $\phi_D(S_2)$ is said to be in the *optimal position* to S_1 .

Definition 2.6. *Given any two points $C_1, C_2 \in \mathcal{C}^+(p)$ and chosen $C_1 = I_p C_1 I_p \in \pi^{-1}(C_1)$, the geodesic distance in $\mathcal{C}^+(p)$ is computed by*

$$d_{g^{\text{quo}}}(C_1, C_2) = \inf_{D \in \mathcal{D}^+(p)} d_{g^{\text{aff}}}(C_1, DC_2 D), \quad (2.26)$$

where $D \in \mathcal{D}^+(p)$. C_1 and $DC_2 D \in \pi^{-1}(C_2)$ are *optimally positioned*.

In general, optimally positioned points will not be uniquely determined (Huckemann *et al.*, 2010). Moreover, the relation being in optimal positions may not be transitive (Ziezold, 1977, p.602).

2.4.3 Optimizing along the fibre

Recalling the Riemannian metric between $C_1, C_2 \in \mathcal{C}^+(p)$ in Equation (2.26), it is calculated via minimizing the affine invariant distance between data point $C_1 \in \mathcal{C}^+(p)$ (also in $\mathcal{S}^+(p)$) and the point in the fibre $\pi^{-1}(C_2) \subset \mathcal{S}^+(p)$ over $C_2 \in \mathcal{C}^+(p)$ as

$$\begin{aligned} d_{g^{\text{quo}}}(C_1, C_2) &= \inf_{D \in \mathcal{D}^+(p)} d_{g^{\text{aff}}}(C_1, DC_2 D) \\ &= \inf_{D \in \mathcal{D}^+(p)} \left\| \log \left(C_1^{-1/2} DC_2 DC_1^{-1/2} \right) \right\|_F \\ &= \inf_{D \in \mathcal{D}^+(p)} \text{Tr} \left[\log^2 (C_1^{-1/2} DC_2 DC_1^{-1/2}) \right]^{1/2}. \end{aligned} \quad (2.27)$$

Although there is no mathematical solution for $D \in \mathcal{D}^+(p)$ in the above equation, fortunately, David (2019) proposed a gradient descent algorithm which seeks the optimal position along the fibre numerically and checked its convergence. Inspired by this, we employ a simple gradient descent algorithm to the objective function in Equation (2.27), rewritten as

$$\begin{aligned} f(D) &= \frac{1}{2} d_{g^{\text{aff}}}(C_1, DC_2D)^2 \\ &= \frac{1}{2} \text{Tr} \left(\log^2(C_1^{-1/2} DC_2 DC_1^{-1/2}) \right). \end{aligned} \quad (2.28)$$

with respect to $D \in \mathcal{D}^+(p)$ such that C_1 and $C_2^* = DC_2D \in \mathcal{S}^+(p)$ are in optimal positions.

To find the gradient descent of $f(D)$, we give the following facts which are essential in the development of our analysis.

- (1) Given a smooth matrix-valued function $X(t)$ of a real variable $t \in [0, 1]$ such that $X^{-1}(t)$ exists for all t , we have the lemma given in Moakher (2005) that

$$\frac{d}{dt} \frac{1}{2} \text{Tr} [\log^2(X(t))] = \text{Tr} [\log(X(t)) X^{-1}(t) X'(t)]. \quad (2.29)$$

- (2) $\text{Tr}(AB) = \text{Tr}(BA)$.

- (3) $A \log(B) A^{-1} = \log(ABA^{-1})$.

Now, we take

$$X(t) = C_1^{-1/2} s(t) C_2 s(t) C_1^{-1/2}$$

where $s(t)$ is a geodesic in the Riemannian manifold $(\mathcal{D}^+(p), g^{\text{aff}})$ which preserves the same affine invariant metric as defined in $(\mathcal{S}^+(p), g^{\text{aff}})$ such that

$$s(t) = \text{Exp}_D(t\Delta) = D^{1/2} \exp(tD^{-1/2} \Delta D^{-1/2}) D^{1/2}$$

starting from $s(0) = D$ with the direction $s'(0) = \Delta$. Subsequently, we have $X(0) = C_1^{-1/2} DC_2 DC_1^{-1/2}$, and $X'(0) = C_1^{-1/2} (\Delta C_2 D + DC_2 \Delta) C_1^{-1/2}$. Therefore, the gradient of the objective function $f(D)$ at $t = 0$ using the above facts is computed by

$$\begin{aligned} & \left. \frac{d}{dt} f(D) \right|_{t=0} \\ &= \left. \frac{d}{dt} \left(\frac{1}{2} \text{Tr} \left(\log^2 \left(C_1^{-1/2} s(t) C_2 s(t) C_1^{-1/2} \right) \right) \right) \right|_{t=0} \\ &\stackrel{\text{Facts(1)}}{=} \text{Tr} \left(\log(C_1^{-1/2} DC_2 DC_1^{-1/2}) (C_1^{-1/2} DC_2 DC_1^{-1/2})^{-1} C_1^{-1/2} (\Delta C_2 D + DC_2 \Delta) C_1^{-1/2} \right) \\ &= \text{Tr} \left(\log(C_1^{-1/2} DC_2 DC_1^{-1/2}) C_1^{1/2} D^{-1} C_2^{-1} D^{-1} \Delta C_2 DC_1^{-1/2} \right) \end{aligned}$$

$$\begin{aligned}
 & + \text{Tr} \left(\log(C_1^{-1/2} D C_2 D C_1^{-1/2}) C_1^{1/2} D^{-1} \Delta C_1^{-1/2} \right) \\
 & \stackrel{\text{Facts}(2)}{=} \text{Tr} \left((C_2 D C_1^{-1/2}) \log(C_1^{-1/2} D C_2 D C_1^{-1/2}) (C_2 D C_1^{-1/2})^{-1} D^{-1} \Delta \right) \\
 & \quad \text{Tr} \left(C_1^{-1/2} \log(C_1^{-1/2} D C_2 D C_1^{-1/2}) C_1^{1/2} D^{-1} \Delta \right) \\
 & \stackrel{\text{Facts}(3)}{=} \text{Tr} \left(\log(C_2 D C_1^{-1} D) D^{-1} \Delta \right) + \text{Tr} \left(\log(C_1^{-1} D C_2 D) D^{-1} \Delta \right).
 \end{aligned}$$

Let the initial direction of geodesic $s(t)$ be $s'(0) = \Delta = I_p$. Then, we take the Hadamard product with respect to I_p to guarantee this quantity restricted to diagonal matrices in $\mathcal{D}^+(p)$ such that

$$\begin{aligned}
 \nabla f(D) &= I_p \circ [D^{-1} \log(C_2 D C_1^{-1} D) + D^{-1} D \log(C_1^{-1} D C_2 D) D^{-1}] \\
 &\stackrel{\text{Facts}(2)}{=} I_p \circ [D^{-1} \log(C_2 D C_1^{-1} D) + (D^{-1} \log(C_2 D C_1^{-1} D))^T] \\
 &= I_p \circ 2\text{Sym} [D^{-1} \log(C_2 D C_1^{-1} D)],
 \end{aligned}$$

where symmetrization operator $\text{Sym}(A) = \frac{1}{2}(A + A^T)$.

Therefore, we can minimize the objective function $f(D)$ via following iterative steps with the step size $\delta > 0$ at the k th iteration:

$$\begin{aligned}
 \Delta_k &= I_p \circ 2\text{Sym} [D_k^{-1} \log(C_2 D_k C_1^{-1} D_k)]; \\
 D_{k+1} &= D_k - \delta \Delta_k \\
 &= \text{Exp}_{D_k}(-\delta \Delta_k) \\
 &= D_k^{1/2} \exp(-\delta D_k^{-1/2} \Delta_k D_k^{-1/2}) D_k^{1/2}
 \end{aligned} \tag{2.30}$$

until the iteration reaches the maximum iteration or the difference of two successive values from objective function is less than 10^{-4} . The detailed algorithm can be seen in Algorithm 2. Once we find an optimal Lie group element $D^* \in \mathcal{D}^+(p)$, as a result of the minimum of the objective function $f(D)$, we write our optimal element in fibre over C_2 as $C_2^* = D^* C_2 D^* \in \pi^{-1}(C_2)$.

Now, we test the above algorithm via the following simulation. We draw any two points $S_1, S_2 \in \mathcal{S}^+(15)$ from a wrapped Gaussian distribution as described in Equation (2.1) with zero mean and covariance $\Sigma = 0.2I_{120}$, resulting in two correlation matrices $C_1, C_2 \in \mathcal{C}^+(15)$. By implementing the gradient descent algorithm with various step sizes $\delta = 0.01, 0.05, 0.1, 0.15$, and 0.2 , along with a maximum iteration limit of $\text{max.ite} = 200$, we compute the values of the objective function $f(D)$ that minimizes the geodesic distance between C_1 and C_2^* , as shown in Figure 2.5.

Indeed, it is important to recognize that larger step sizes in the algorithm tend to converge faster. However, caution must be exercised to ensure that the step size remains

Algorithm 2 Optimization along the fibre via the gradient descent algorithm

Require: Observations $C_1, C_2 \in \mathcal{C}^+(p)$, initial point $D_0 \in \mathcal{D}^+(p)$ and $\Delta = I_p$, step size $\delta > 0$
 $k = 0$
while $k > \text{max.ite}$ or $|f(D_{k+1}) - f(D_k)| < 10^{-4}$ **do**
 $\Delta_k = I_p \circ 2\text{Sym} [D_k^{-1} \log(C_2 D_k C_1^{-1} D_k)]$
 $D_{k+1} = D_k^{1/2} \exp(-\delta D_k^{-1/2} \Delta_k D_k^{-1/2}) D_k^{1/2}$
 $k = k + 1$
end while
 $D^* = D_k$
 $C_2^* = D^* C_2 D^* \in \pi^{-1}(C_2)$

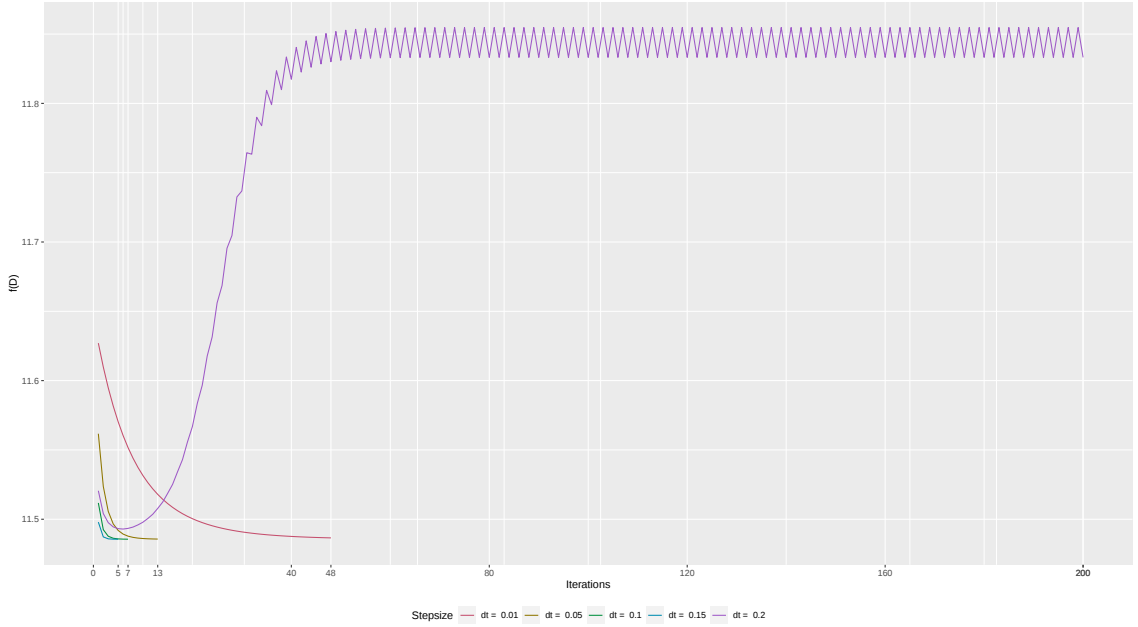


Figure 2.5: Convergence of the gradient descent algorithm with various step sizes, ranging from 0.01 to 0.2. The x -axis represents iterative steps, and the y -axis represents the values of the objective function $f(D)$. The ticks on the x -axis indicate the iterative steps at which the algorithm converged.

within an appropriate range. If the step size becomes too large, the algorithm may overshoot the minimum and fail to converge properly. For instance, using a step size of $\delta = 0.2$ or larger can result in the breakdown of the algorithm.

To strike a balance between convergence and stability, it is crucial to carefully select a step size that allows the algorithm to progress efficiently while avoiding overshooting and maintaining convergence to the desired solution. In this thesis, we choose the default $\delta = 0.05$ in our following research. Also, readers can choose other values depending on their data.

Moreover, David (2019) demonstrated the convergence of the proposed method and compared it with other approaches. However, the theoretical proof for the uniqueness and even existence of the minimizer D^* in his thesis remains elusive. Although Lueg (2023) established the coercivity of the objective function to be minimized, implying the existence of a solution to Equation (2.27), the uniqueness of the optimally positioned point has not been explicated. We will not explore this further in this thesis. Instead, we utilize the numerical minimizer D^* to obtain points in their optimal positions directly and discuss that the uniqueness of optimal positions does not impact our statistical analysis (see Theorem 2.4 and Proposition 2.1).

2.4.4 Exponential and logarithm maps

Theorem 2.3. *The geodesic $c(t)$ in the quotient manifold is a projection of the horizontal geodesic $\gamma^{\mathcal{H}}(t)$ in the manifold $\mathcal{S}^+(p)$. The geodesic connecting any two points in optimal positions is horizontal.*

The proof can be seen in (Huckemann *et al.*, 2010). Generally, it is not true that two points on a horizontal geodesic segment are in optimal positions.

Based on Theorem 2.3, we can define exponential and logarithm maps in the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$.

Definition 2.7. *For any point $C \in \mathcal{C}^+(p)$ and tangent vector $X \in T_C\mathcal{C}^+(p)$ at the base point C , we can ‘lift’ X from $T_C\mathcal{C}^+(p)$ to the horizontal subspace $\mathcal{H}_C$ and apply the exponential map in $\mathcal{S}^+(p)$ which is defined in (2.14). Finally, π is used to project this point back into the quotient manifold so that the exponential map in $\mathcal{C}^+(p)$ is defined as*

$$\text{Exp}_C(X) = \pi(\text{Exp}_C(X^\sharp)) = \pi(C^{1/2} \exp(C^{-1/2} \text{hor}(X) C^{-1/2}) C^{1/2}), \quad (2.31)$$

where $\text{Exp}_C(X^\sharp)$ is the exponential map as defined in the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$.

Subsequently, supposing that $c(0) = C$, $c'(0) = X$, the geodesic $c(t)$ in the direction $c'(0) = X \in T_C\mathcal{C}^+(p)$ is

$$c(t) = \pi(C^{1/2} \exp(t C^{-1/2} \text{hor}(X) C^{-1/2}) C^{1/2}), \quad t \in [0, 1]. \quad (2.32)$$

According to the preceding subsection, given any $C_1, C_2 \in \mathcal{C}^+(p)$, we can find the optimal position in the fiber $\pi^{-1}(C_2) \subset \mathcal{S}^+(p)$ with respect to $C_1 \in \mathcal{S}^+(p)$ by optimizing the objective function, i.e., obtaining $C_2^* = DC_2D \in \pi^{-1}(C_2) \subset \mathcal{S}^+(p)$, so that $\text{Log}_{C_1}(C_2^*) \in T_{C_1}\mathcal{S}^+(p)$ is a tangent vector, also a horizontal vector, i.e., $\text{Log}_{C_1}(C_2^*) = \text{hor}(\text{Log}_{C_1}(C_2^*))$ according to Theorem 2.3 and $\text{Log}_{C_1}(C_2^*)$ is the logarithm map defined in the affine in-

variant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$. Consequently, we can define the logarithm map in the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$.

Definition 2.8. *Given $C_1, C_2 \in \mathcal{C}^+(p)$, $C_2^* \in \pi^{-1}(C_2)$ and C_1 are in optimal positions in $\mathcal{S}^+(p)$. The logarithm map in $\mathcal{C}^+(p)$ is defined as*

$$\text{Log}_{C_1}(C_2) = d\pi_{C_1}(\text{hor}(\text{Log}_{C_1}(C_2^*))) = d\pi_{C_1}(\text{Log}_{C_1}(C_2^*)) \in T_{C_1}\mathcal{C}^+(p), \quad (2.33)$$

where $\text{Log}_C(C_2^*)$ is the logarithm map as defined in the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$.

Before introducing the theorem of the uniqueness of the tangent vector $X \in T_C\mathcal{C}^+(p)$, we give the following lemma.

Lemma 2.3. *Given a Riemannian submersion $\pi : \mathcal{S}^+(p) \rightarrow \mathcal{C}^+(p)$, $d\pi_S$ gives a linear isomorphism between $\mathcal{H}_S$ and $T_{\pi(S)}\mathcal{C}^+(p)$ and it is surjective, i.e., the derivative of horizontal vectors along the fibre is uniquely determined in $T_C\mathcal{C}^+(p)$.*

Proof. A proof for the first part (linear isomorphism) can be found in the book by (Gallier & Quaintance, 2020, Section 18.3). Since π is a surjective submersion (Lueg, 2023), $d\pi_S$ will be a surjective map from $\mathcal{H}_S$ to $T_{\pi(S)}\mathcal{C}^+(p)$ as well. Specifically, for every $C \in \mathcal{C}^+(p)$ and the tangent vector $X \in T_C\mathcal{C}^+(p)$, the horizontal lift from $T_C\mathcal{C}^+(p)$ to $\mathcal{H}_S$ is defined as $X^\sharp = (d\pi_S)^{-1}(X)$, for any $S \in \pi^{-1}(C)$. This implies that for any S_1, S_2 both in the fibre $\pi^{-1}(C)$, i.e., $\pi(S_1) = \pi(S_2) = C \in \mathcal{C}^+(p)$, the tangent vector field X could be uniquely horizontally lifted to the horizontal subspace $\mathcal{H}_{S_1}$ as $(X^\sharp)_{S_1}$ or the horizontal subspace $\mathcal{H}_{S_2}$ as $(X^\sharp)_{S_2}$. Since $d\pi_{S_1}$ (or $d\pi_{S_2}$) is an isomorphism between $\mathcal{H}_{S_1}$ (or $\mathcal{H}_{S_2}$) and $T_C\mathcal{C}^+(p)$, we have $d\pi_{S_1}(X^\sharp) = d\pi_{S_2}(X^\sharp) = X$. Therefore, the derivative of horizontal vectors at any base point in the fibre $\pi^{-1}(C)$, is unique in $T_C\mathcal{C}^+(p)$. $\square$

Theorem 2.4. *Given any two points $C_1, C_2 \in \mathcal{C}^+(p)$, the logarithm map $X = \text{Log}_{C_1}(C_2)$ as defined in Equation (2.33) is unique.*

Proof. Given any two points $C_1, C_2 \in \mathcal{C}^+(p)$, the logarithm map can be written as $X = \text{Log}_{C_1}(C_2) \in T_{C_1}\mathcal{C}^+(p)$, which is the tangent vector along the geodesic $c(t), t \in [0, 1]$ connecting C_1 to C_2 . Suppose C_1, S_1 are both in the fiber $\pi^{-1}(C_1)$, C_2^*, S_2^* are both in the fibre $\pi^{-1}(C_2)$, and C_1, C_2^* and S_1, S_2^* are in optimal positions (see Figure 2.6). Tangent vectors $\text{Log}_{C_1}(C_2^*)$ and $\text{Log}_{S_1}(S_2^*)$ are both horizontal, and their connection is a horizontal geodesic according to Theorem 2.3. Therefore, according to Lemma 2.3, we have

$$d\pi_{C_1}(\text{Log}_{C_1}(C_2^*)) = d\pi_{S_1}(\text{Log}_{S_1}(S_2^*)).$$

Additionally, if it is true that the optimal positions may not be uniquely determined, there exists another optimally positioned C_2' with respect to C_1 such that $V^* = \text{Log}_{C_1}(C_2^*)$

and $V' = \text{Log}_{C_1}(C_2')$ are both horizontal in the horizontal subspace $\mathcal{H}_{C_1}$. According to Theorem 2.3, $d\pi_{C_1}(V^*) = d\pi_{C_1}(V')$ still holds true, which means the logarithm map $X = \text{Log}_{C_1}(C_2)$ does not depend on the optimal positions C_2^* along the fibre $\pi^{-1}(C_2)$.

In summary, given any two points $C_1, C_2 \in \mathcal{C}^+(p)$, the definition of the logarithm map $X = d\pi_{C_1}(\text{Log}_{C_1}(C_2^*))$ is uniquely determined even if the optimal position $C_2^* \in \pi^{-1}(C_2)$ with respect to C_1 is not unique. Similarly, the logarithm map in $(\mathcal{C}^+(p), g^{\text{quo}})$ does not depend on the chosen point in the fibre $\pi^{-1}(C_1)$. $\square$

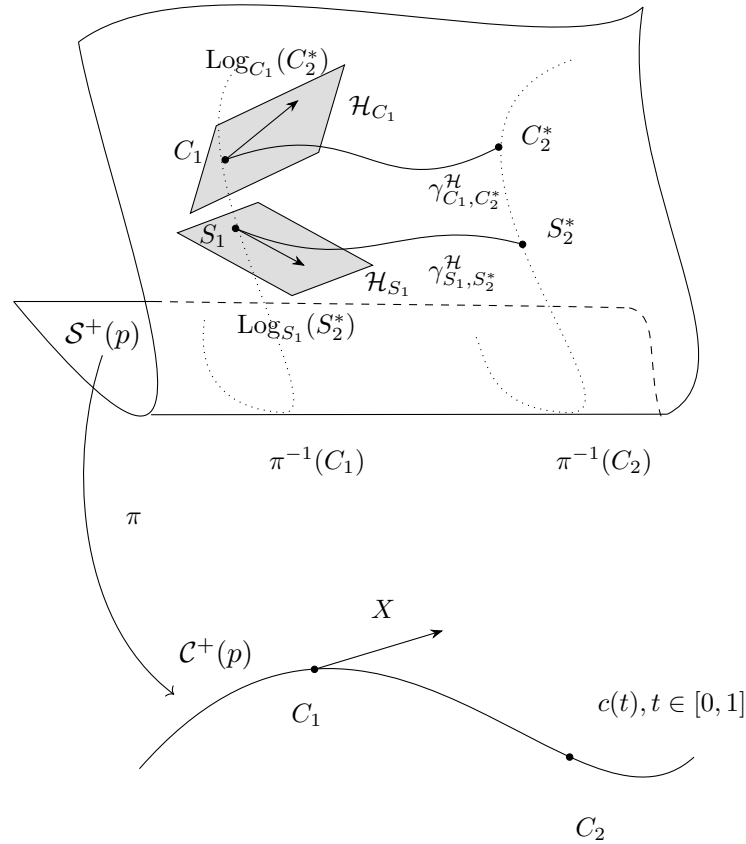


Figure 2.6: Optimal positions along fibres. C_1 and S_1 are both on the fibre $\pi^{-1}(C_1)$, C_2^* and S_2^* are both on the fibre $\pi^{-1}(C_2)$, and C_1 & C_2^* and S_1 & S_2^* are in optimal positions. Geodesic $\gamma_{C_1, C_2^*}^{\mathcal{H}}$ ($\gamma_{S_1, S_2^*}^{\mathcal{H}}$) connecting two optimal positions C_1 and C_2^* (S_1 and S_2^*) is horizontal. $\text{Log}_{C_1}(C_2^*)$ and $\text{Log}_{S_1}(S_2^*)$ are both horizontal vectors lying in the horizontal subspaces $\mathcal{H}_{C_1}$ and $\mathcal{H}_{S_1}$, respectively.

2.4.5 Parallel transport

Le (2003) characterizes the parallel transport on the shape space in terms of its behaviour on the pre-shape space via Riemannian submersion. This follows from an important

theorem presented below.

Theorem 2.5. *Given a Riemannian submersion $\pi : \mathcal{M} \rightarrow \mathcal{N}$ for two Riemannian manifolds, $\mathcal{M}$ and $\mathcal{N}$, each equipped with its respective Levi-Civita connections, denoted by ∇ and $\dot{\nabla}$, respectively:*

- (1) *For any given tangent vectors u_1, u_2 both in the tangent space $T_y \mathcal{N}$ at y , and their corresponding horizontal lift to the horizontal subspace $\mathcal{H}_x$ as $u_1^\sharp, u_2^\sharp$, where $x \in \pi^{-1}(y) \subset \mathcal{M}$, the relationship $d\pi(\nabla_{u_1^\sharp} u_2^\sharp) = \dot{\nabla}_{u_1} u_2$ holds true.*
- (2) *Let $v^\mathcal{H} \in T_x \mathcal{M}$ be a tangent vector field along the horizontal geodesic $\gamma^\mathcal{H}(t), t \in [0, 1]$ with an initial velocity $\gamma^{\mathcal{H}'}(0) = v_1^\mathcal{H}$, which is horizontal. The projection $d\pi(v^\mathcal{H})$ is the parallel transport of $d\pi(v_1^\mathcal{H})$ along $\pi(\gamma^\mathcal{H}) \in \mathcal{N}$ if and only if*

$$d\pi(\nabla_{v_1^\mathcal{H}} v^\mathcal{H}) = 0.$$

The proof follows from Lemma 1 and Theorem 1 of Le (2003). Part (1) of Theorem 2.5 gives the relationship of Levi-Civita connections for both manifolds $\mathcal{M}$ and $\mathcal{N}$. Part (2) explains the condition of parallel transport in $\mathcal{C}^+(p)$. When the tangent vector $v^\mathcal{H} \in T_x \mathcal{M}$ is horizontal, i.e., its vertical component is 0, $v^\mathcal{V} = 0$, we say that $v^\mathcal{H} \in T_{\gamma^\mathcal{H}} \mathcal{M}$ is parallel along $\gamma^\mathcal{H}$ if and only if $\nabla_{\gamma^{\mathcal{H}'}} v^\mathcal{H} = 0$ (Gallier & Quaintance, 2020, Definition 15.6). This follows the definition of parallel transport in the quotient manifold. Specifically, the horizontal lift $u_1^\sharp \in \mathcal{H}_S$ could be considered as an initial velocity of the horizontal geodesic $\gamma^{\mathcal{H}'}(0) \in \mathcal{S}^+(p)$. Horizontal lift $u^\sharp$ is a horizontal vector field along the horizontal geodesic $\gamma^\mathcal{H}$. Therefore, $\nabla_{u_1^\sharp} u^\sharp = 0$ such that $d\pi(\nabla_{u_1^\sharp} u^\sharp) = d\pi(\nabla_{v_1^\mathcal{H}} v^\mathcal{H}) = 0 = \dot{\nabla}_{u_1} u$ holds true.

Now, we can give the definition of parallel transport in the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$.

Definition 2.9. *Given any two points $C_1, C_2 \in \mathcal{C}^+(p)$ and a tangent vector $X \in T_{C_1} \mathcal{C}^+(p)$, we can find an optimal position with respect to $C_1 \in \mathcal{S}^+(p)$ in the fibre $\pi^{-1}(C_2)$ as $C_2^* = DC_2 D \in \pi^{-1}(C_2)$, which guarantees the geodesic $\gamma(t), t \in [0, 1]$ in $\mathcal{S}^+(p)$ joining $\gamma(0) = C_1 \in \mathcal{S}^+(p)$ and $\gamma(1) = C_2^* \in \mathcal{S}^+(p)$ is horizontal. Henceforth, the parallel transport of X from tangent space $T_{C_1} \mathcal{C}^+(p)$ to tangent space $T_{C_2} \mathcal{C}^+(p)$ is defined as*

$$\begin{aligned} \mathcal{P}_{C_1 \rightarrow C_2}(X) &= d\pi_{C_2^*}(\mathcal{P}_{C_1 \rightarrow C_2^*}(\text{Log}_{C_1}(C_2^*))) \\ &= d\pi_{C_2^*}\left((C_2^* C_1^{-1})^{1/2} \text{Log}_{C_1}(C_2^*) \left((C_2^* C_1^{-1})^{1/2}\right)^T\right) \end{aligned} \quad (2.34)$$

where $C_2^* \in \mathcal{S}^+(p)$ is the optimal position in $\pi^{-1}(C_2)$ with respect to $C_1 \in \mathcal{S}^+(p)$ and $\text{Log}_{C_1}(C_2^*)$ is the logarithm map in the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$ for given optimal positions $C_1, C_2^* \in \mathcal{S}^+(p)$.

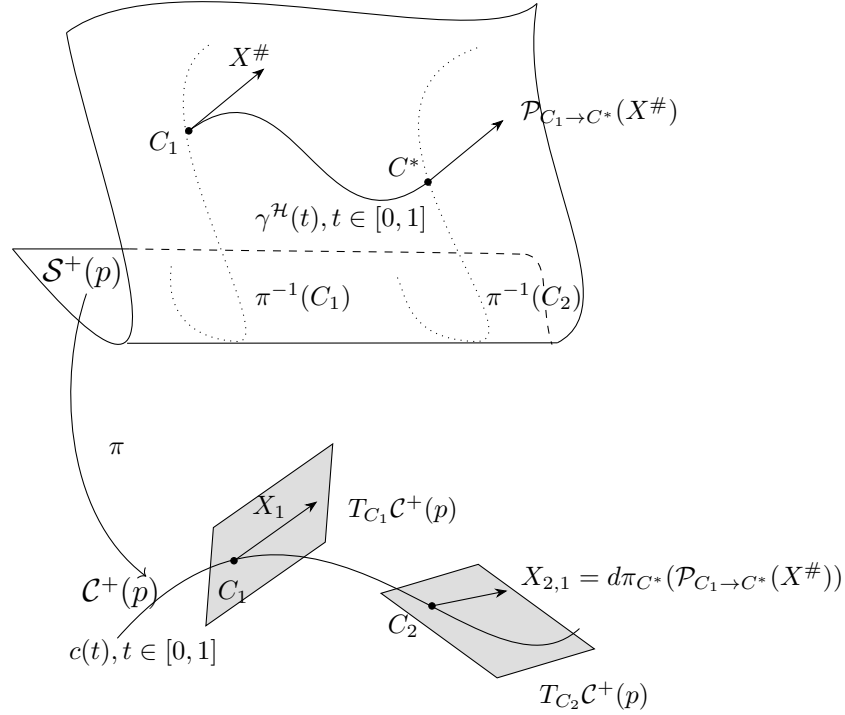


Figure 2.7: The diagram of the parallel transport in $\mathcal{C}^+(p)$ for any given $C_1, C_2 \in \mathcal{C}^+(p)$ and $X_1 \in T_{C_1}\mathcal{C}^+(p)$. The general steps are: (i) find the optimal position $C_2^* = DC_2D \in \mathcal{S}^+(p)$ in the fibre $\pi^{-1}(C_2)$ with respect to $C_1 \in \mathcal{S}^+(p)$; (ii) calculate the horizontal lift $X^\#$ of horizontal vector $X \in T_{C_1}\mathcal{C}^+(p)$ onto $\mathcal{H}_{C_1}$; (iii) parallel transport horizontal lift $X^\#$ from horizontal subspace $\mathcal{H}_{C_1}$ to horizontal subspace $\mathcal{H}_{C_2^*}$ along horizontal geodesic $\gamma^{\mathcal{H}}$ connecting from C_1 to C_2^* ; (iv) project the transported vector $\mathcal{P}_{C_1 \rightarrow C_2^*}(X^\#)$ onto the tangent space $T_{C_2}\mathcal{C}^+(p)$ by the derivative of the submersion π at C_2^* , i.e. $X_{2,1} = d\pi_{C_2^*}(\mathcal{P}_{C_1 \rightarrow C_2^*}(X^\#))$.

The detailed process of parallel transport in the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$ is illustrated in Figure 2.7.

Analogous to the uniqueness of logarithm map in $\mathcal{C}^+(p)$, we have the following theorem on the parallel transport in $\mathcal{C}^+(p)$.

Proposition 2.1. *Given any $C_1, C_2 \in \mathcal{C}^+(p)$ and tangent vector $X \in T_{C_1}\mathcal{C}^+(p)$, parallel transport $\mathcal{P}_{C_1 \rightarrow C_2}(X)$ as defined in Equation (2.34) is unique.*

Proof. The geodesic connecting two optimal positions C_1, C_2^* is horizontal such that the parallel transport of horizontal vectors along such geodesic is horizontal everywhere. According to the proof of Theorem 2.4, we know that $d\pi_{\pi^{-1}(C_2)}$ is a surjective map; and it is uniquely determined without considering the uniqueness of optimal positions and the choice of base point in the fibre $\pi^{-1}(C_1)$, i.e. $d\pi_{S_2}(\mathcal{P}_{S_1 \rightarrow S_2}(X^\#))$ is uniquely determined iff any $S_1 \in \pi^{-1}(C_1)$ and $S_2 \in \pi^{-1}(C_2)$ are optimally positioned. $\square$

Alternatively, we can compute parallel transport directly by the definition of Levi-

Civita connection $\dot{\nabla}$ in the quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$ as

$$\begin{aligned} \left(\dot{\nabla}_X(Y) \right)_{|C} = d_C Y(X) + \text{sym} \left[\text{Diag}(X^\#)Y^\# + \text{Diag}(Y^\#)X^\# + \text{Diag}(X^\#C^{-1}Y^\#)C \right. \\ \left. - X^\#C^{-1}Y^\# - \frac{1}{2}\text{Diag}(X^\#)C\text{Diag}(Y^\#) - \frac{3}{2}\text{Diag}(X^\#)\text{Diag}(Y^\#)C \right], \end{aligned}$$

where $d_C Y(X)$ is the derivative of X with respect to Y at the base C for every $C \in \mathcal{C}^+(p)$ and $X, Y \in T_C \mathcal{C}^+(p)$ (Thanwerdas & Pennec, 2021). If $\left(\dot{\nabla}_X(Y) \right)_{|C} = 0$, then we can solve the parallel translated vector Y along the geodesic $c(t) \in \mathcal{C}^+(p), t \in [0, 1]$ with the initial velocity $X = c'(0)$ such that parallel transport in quotient manifold is well-defined. However, the computation is a huge challenge and it is not clear how to solve $\left(\dot{\nabla}_X(Y) \right)_{|C}$.

Therefore, in this research, we compute the parallel transport in $\mathcal{C}^+(p)$ using Equation (2.34) with the numerical computation.

2.4.6 Orthonormal coordinate system

Recalling the orthonormal basis in the tangent space $T_{I_p} \mathcal{S}^+(p)$ at the identity point $I_p \in \mathcal{S}^+(p)$ and decomposing it into horizontal and vertical components as

$$E_{ij} = \begin{cases} \text{hor}(E_{ij}) = 0; & \text{ver}(E_{ij}) = e_i e_j^T & \text{if } i = j \\ \text{hor}(E_{ij}) = \frac{\sqrt{2}}{2}(e_i e_j^T + e_j e_i^T); & \text{ver}(E_{ij}) = 0 & \text{if } 1 \leq i < j \leq p \end{cases} \quad (2.35)$$

Since the horizontal subspace $\mathcal{H}_S$ is diffeomorphic to $T_{\pi(S)} \mathcal{C}^+(p)$ under $d\pi_S$ for every $S \in \mathcal{S}^+(p)$, we can use the push-forward to define an orthonormal basis in the tangent space $T_{I_p} \mathcal{C}^+(p)$ at the identity point $I_p \in \mathcal{C}^+(p)$ as

$$E_{ij}^{I_p} = d\pi_{I_p}(\text{hor}(E_{ij})) = E_{ij},$$

where $1 \leq i < j \leq p$. Obviously, E_{ij} is the basis both in the horizontal subspace $\mathcal{H}_{I_p}$ and the tangent space $T_{I_p} \mathcal{C}^+(p)$. (It is easy to check that the horizontal lift $(E_{ij})_{I_p}^\# = E_{ij}$).

To distinguish the orthonormal basis in $T_{I_p} \mathcal{S}^+(p)$ and $T_{I_p} \mathcal{C}^+(p)$ with different dimensions, we introduce a new notation $F_{ij} = E_{ij}$ as the orthonormal basis in $\mathcal{H}_{I_p}$ and $T_{I_p} \mathcal{C}^+(p)$ for $i \neq j$ and simplify it as $F_r = F_{ij}, r = 1, \dots, m = p(p-1)/2$.

According to the definition of parallel transport in $\mathcal{C}^+(p)$ as Equation (2.34), we can define the orthonormal basis in any tangent space $T_C \mathcal{C}^+(p)$.

Definition 2.10. *The orthonormal basis in tangent space $T_C \mathcal{C}^+(p)$ at any base point*

$C \in \mathcal{C}^+(p)$ is defined as

$$\begin{aligned} F_r^C &= \mathcal{P}_{I_p \rightarrow C}(F_r) = d\pi_{C^\dagger} \left[(C^\dagger I_p^{-1})^{1/2} (F_r)^\sharp ((C^\dagger I_p^{-1})^{1/2})^T \right] \\ &= d\pi_{C^\dagger} \left[C^{\dagger 1/2} F_r C^{\dagger 1/2} \right] \end{aligned} \quad (2.36)$$

where $C^\dagger = DCD \in \pi^{-1}(C)$ is the optimal position with respect to I_p , and $r = 1, \dots, m = p(p-1)/2$.

Theorem 2.6. *The tangent space in the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$ is parallelizable.*

Proof. At every point $S \in \pi^{-1}(C) \subset \mathcal{S}^+(p)$, the geodesic connecting two optimal positions S and I_p is horizontal such that the tangent vectors $\{F_r^S : r = 1, \dots, m\}$ provide a set of basis for the horizontal subspace $\mathcal{H}_S$ as $F_r^S = S^{1/2} F_r S^{1/2}$ defined in Equation (2.19). It follows that the differentiable horizontal subspace $\mathcal{H}_S$ is parallelizable.

In accordance with Lemma 2.3 and Proposition 2.1, the orthonormal basis $E_r^C = d\pi_S (S^{1/2} F_r S^{1/2})$, $r = 1, \dots, m$, defined in (2.36), in the quotient manifold is uniquely determined by the diffeomorphism $d\pi_S$. Thus, the tangent space in $\mathcal{C}^+(p)$ possesses a frame of global sections, i.e., the tangent space $T_C \mathcal{C}^+(p)$ at $C \in \mathcal{C}^+(p)$ is trivial. Moreover, a manifold whose tangent space is trivial is parallelizable, as proven by (Gallier & Quaintance, 2020). $\square$

Subsequently, we can define the vectorization of any tangent vector $X \in T_C \mathcal{C}^+(p)$ obtaining a coordinate vector $\text{Vec}(X) \in \mathbb{R}^m$.

Definition 2.11. *Given any tangent vector $X \in T_C \mathcal{C}^+(p)$ at the base point $C \in \mathcal{C}^+(p)$, the vectorization of tangent vector $X \in T_C \mathcal{C}^+(p)$, written as $\text{Vec}(X) = \mathbf{u} \in \mathbb{R}^m$, is defined as*

$$\begin{aligned} u_r &= g_C^{\text{quo}}(X, F_r^C) \\ &= g_S^{\text{aff}}(X^\sharp, (F_r^C)^\sharp) \\ &= \left\langle X^\sharp, (F_r^C)^\sharp \right\rangle_S, \text{ (where } S \in \pi^{-1}(C)) \\ &= \left\langle \text{hor}(\Delta_S X \Delta_S), \text{hor}(\Delta_S F_r^C \Delta_S) \right\rangle_S \\ &= \left\langle \text{hor}(\Delta_S X \Delta_S), \text{hor} \left(\Delta_S d\pi_{C^\dagger} (C^{\dagger 1/2} F_r C^{\dagger 1/2}) \Delta_S \right) \right\rangle_S \end{aligned} \quad (2.37)$$

$$\begin{aligned} &= \left\langle \text{hor}(\Delta_{C^\dagger} X \Delta_{C^\dagger}), C^{\dagger 1/2} F_r C^{\dagger 1/2} \right\rangle_{C^\dagger} \text{ (when } S = C^\dagger) \\ &= \text{Tr} \left(C^{\dagger -1/2} \text{hor}(\Delta_{C^\dagger} X \Delta_{C^\dagger}) C^{\dagger -1/2} F_r \right) \end{aligned} \quad (2.38)$$

where $S \in \pi^{-1}(C)$ is any point in the fibre $\pi^{-1}(C)$ and $\mathbf{u} = (u_1, \dots, u_r)^T$, $r = 1, \dots, m = p(p-1)/2$. $C^\dagger$ is the optimal position with respect to I_p .

In general, Equation (2.37) is the vectorization solution $Vec(X)$ for any point $S \in \pi^{-1}(C)$ as the quotient metric does not depend on the chosen $S \in \pi^{-1}(C)$. Note that Equation (2.38) relies on finding optimally positioned $S = C^\dagger$ in the fibre $\pi^{-1}(C)$. The diagram in Figure 2.8 shows the whole processes of the vectorization of $X \in T_C \mathcal{C}^+(p)$ to the coordinate vector $\mathbf{u} \in \mathbb{R}^m$ in a general way.

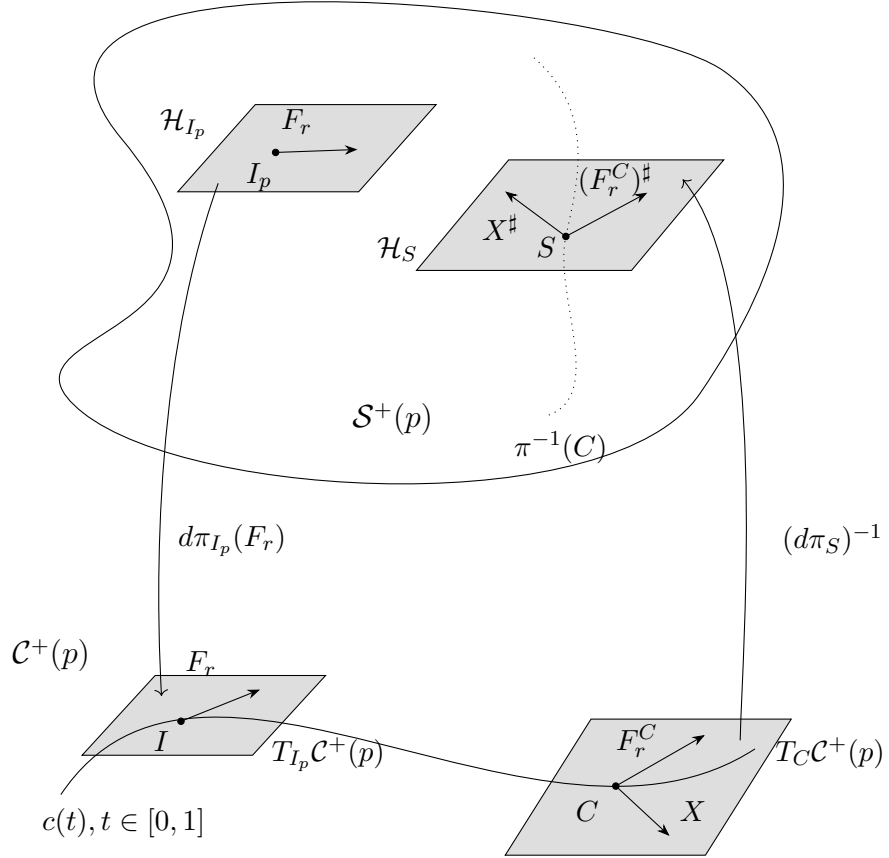


Figure 2.8: Illustration of $Vec(X)$ in $\mathcal{C}^+(p)$. (i): project horizontal basis in $\mathcal{H}_{I_p}$ into tangent space $T_{I_p} \mathcal{C}^+(p)$; (ii): do parallel transport of basis F_r from $T_{I_p} \mathcal{C}^+(p)$ to $T_C \mathcal{C}^+(p)$. (iii): pull back X and translated basis in F_r^C to the horizontal subspace $\mathcal{H}_S$ by the horizontal lift and S is any base point in $\pi^{-1}(C)$. (iv): vectorize the tangent vector X by calculating Riemannian metric $g_S^{\text{aff}} \langle X^\sharp, (F_r^C)^\sharp \rangle$.

The above push-forward method for computing the vectorization of tangent vector $X \in T_C \mathcal{C}^+(p)$ can be simplified without considering the orthonormal basis in $T_C \mathcal{C}^+(p)$.

Lemma 2.4. *Let $I_p \in \mathcal{S}^+(p)$ and $C^\dagger \in \pi^{-1}(C)$ be in optimal positions. The horizontal orthonormal basis in $\mathcal{H}_{C^\dagger}$ is $C^{\dagger 1/2} F_r C^{\dagger 1/2}$. Lifting X from smooth vector field $T_C \mathcal{C}^+(p)$ to horizontal subspace $\mathcal{H}_{C^\dagger}$, we can compute the affine invariant metric between $(X^\sharp)_{C^\dagger}$ and orthonormal basis $C^{\dagger 1/2} F_r C^{\dagger 1/2}$, and then the vectorization of any tangent vector*

$X \in T_C \mathcal{C}^+(p)$ also could be defined as

$$\begin{aligned}
 u_r &= g_{C^\dagger}^{\text{aff}}(X^\sharp, C^{\dagger 1/2} F_r C^{\dagger 1/2}) \\
 &= \langle X^\sharp, C^{\dagger 1/2} F_r C^{\dagger 1/2} \rangle_{C^\dagger} \\
 &= \text{Tr} \left(C^{\dagger -1/2} \text{hor}(\Delta_{C^\dagger} X \Delta_{C^\dagger}) C^{\dagger -1/2} F_r \right) \\
 &= \text{Tr} \left(C^{\dagger -1/2} \text{hor}(\Delta_{C^\dagger} X \Delta_{C^\dagger}) C^{\dagger -1/2} F_r \right)
 \end{aligned} \tag{2.39}$$

Apparently, the above result of Equation(2.39) is exactly the same as from Equation (2.38) when S is chosen by the optimal position $C^\dagger$ with respect to I_p . Figure 2.9 shows the construction of the vectorization $\text{Vec}(X)$ ignoring the definition of orthonormal basis in $T_C \mathcal{C}^+(p)$.

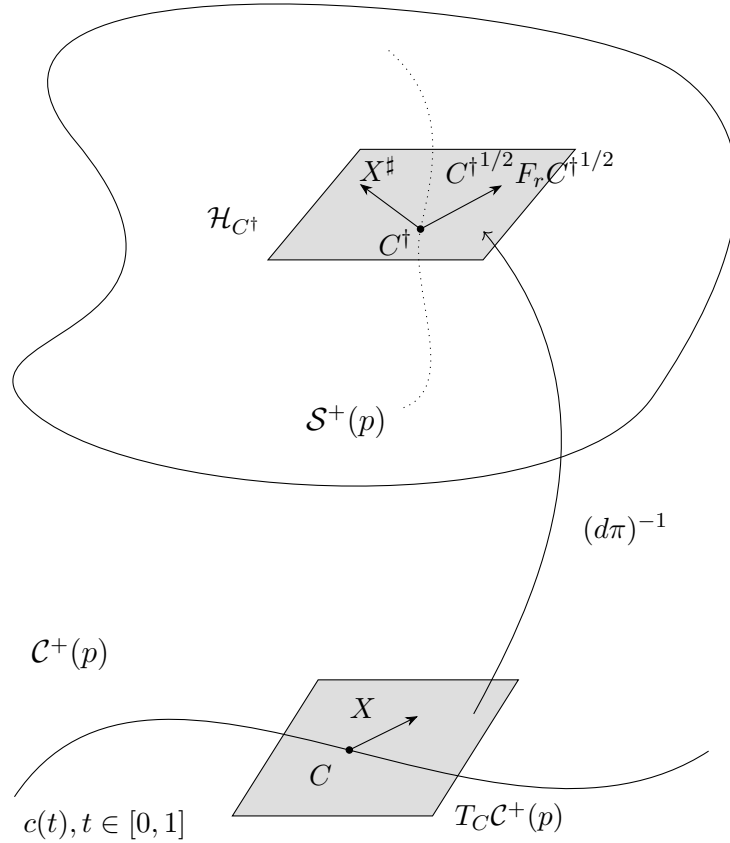


Figure 2.9: Coordinate vectors by the horizontal lift. $C^\dagger = \pi^{-1}(C) = D^\dagger C D^\dagger \subset \mathcal{S}^+(p)$ is in optimal positions with respect to $I \in \mathcal{S}^+(p)$. $C^{\dagger 1/2} \text{hor}(E_r) C^{\dagger 1/2}$ is the orthonormal basis in the horizontal subspace $\mathcal{H}_{C^\dagger}$, $r = 1, \dots, \frac{1}{2}p(p-1)$.

Remark 1. In Lemma 2.4, we cannot lift X into an arbitrary horizontal subspace $\mathcal{H}_S$, where S is any point in the fiber $\pi^{-1}(C)$. Only optimal positions with respect to $I_p \in \mathcal{S}^+(p)$ are eligible for Equation (2.39). Alternatively, we can horizontally lift the orthonormal basis as defined in Equation (2.36) and the tangent vector $X \in T_C \mathcal{C}^+(p)$ together onto any horizontal subspace $\mathcal{H}_S$. In this case, S could be selected at any point in the fibre $\pi^{-1}(C)$.

The basis $\mathbf{E}^S$ in the tangent space $T_S \mathcal{S}^+(p)$ can be decomposed into its horizontal component $\mathbf{E}^{\mathcal{H}_S} = \text{hor}(\mathbf{E}^S)$ and vertical component $\mathbf{E}^{\mathcal{V}_S} = \text{ver}(\mathbf{E}^S)$, represented as $\mathbf{E}^S = \mathbf{E}^{\mathcal{H}_S} \oplus \mathbf{E}^{\mathcal{V}_S}$. The condition $\mathbf{F}^S = \mathbf{E}^{\mathcal{H}_S}$ holds if and only if I_p and S are in optimal positions, where $\mathbf{F}^S = S^{1/2} \mathbf{F} S^{1/2}$ denotes the orthonormal basis in the horizontal subspace $\mathcal{H}_S$. This condition arises from the fact that the geodesic connecting I_p and S is horizontal, ensuring that the vectors along this geodesic are horizontal everywhere. However, in general, there is no guarantee that any S in the fibre $\pi^{-1}(C)$ possesses an orthonormal basis $\mathbf{F}^S$ that is horizontal. Furthermore, the orthonormal basis in the horizontal subspace $\mathbf{E}^{\mathcal{H}_S}$ cannot simply be considered as the horizontal components of $\mathbf{F}^S$, i.e., $\mathbf{E}^{\mathcal{H}_S} \neq \text{hor}(\mathbf{F}^S)$ if I_p and S are not optimally positioned. This is why we lift $X \in T_C \mathcal{C}^+(p)$ to the horizontal subspace $\mathcal{H}_{C^\dagger}$ in Figure 2.9, as I_p and $C^\dagger$ are in optimal positions.

In practice, ensuring that $\mathbf{F}^{C^\dagger} \in \mathcal{H}_{C^\dagger}$ is entirely horizontal is rarely feasible. This is because $C^\dagger$ is typically computed numerically using the gradient descent algorithm, as described in Section 2.4.3. Improving the precision of optimal positions is a topic for further research and will not be addressed in this thesis.

2.4.7 Fréchet sample mean and variance

Karcher (1977) has proved the existence and uniqueness of the Riemannian Fréchet mean for manifolds with non-positive sectional curvature. However, the determination of the sectional curvature's sign for the quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$ remains elusive (Thanwerdas & Pennec, 2021). In this thesis, we will not attempt to calculate the sign of the sectional curvature either.

We will assume that Fréchet sample mean in $(\mathcal{C}^+(p), g^{\text{quo}})$ with the quotient geometry exists and is unique. The definition of the Fréchet sample mean in $(\mathcal{C}^+(p), g^{\text{quo}})$ is analogous to the definition in $(\mathcal{S}^+(p), g^{\text{aff}})$ and can be defined as follows:

Definition 2.12. The Fréchet sample mean in the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$ is computed by minimizing the Fréchet function, which finds a point $\tilde{C}$ such that

$$\tilde{C} \in \arg \inf_{\tilde{C}} \sum_{i=1}^n \frac{1}{n} d_{g^{\text{quo}}}(C_i, \tilde{C})^2, \quad (2.40)$$

where the distance between C_i and $\tilde{C}$ is given by

$$d_{g^{\text{quo}}}(C_i, \tilde{C})^2 = \inf_{D_i \in \mathcal{D}^+(p)} d_{g^{\text{aff}}}(D_i C_i D_i, \tilde{C})^2,$$

and D_i is solved by optimizing the objective function $f(D)$, defined in Equation (2.28), for all $\{C_i : i = 1, \dots, n\}$.

Although Riquelme (2021) does not prove the existence and uniqueness of the Fréchet mean in $\mathcal{C}^+(p)$, he directly computes optimal positions for all observations with respect to an iterated mean estimate $\tilde{C}_k$ and then obtains the Fréchet mean of these optimal positions in $\mathcal{S}^+(p)$ as $\tilde{S}_k$. Finally, the intrinsic mean in $\mathcal{C}^+(p)$ is the submersion of $\tilde{S}_k$, i.e., $\tilde{C}_{k+1} = \pi(\tilde{S}_k) \in \mathcal{C}^+(p)$. Based on the assumption that the Fréchet sample mean exists and is unique, we can compute it directly using the Sturm algorithm in $\mathcal{C}^+(p)$. Specifically, we update the iterated mean $\tilde{C}_k$ along the geodesic $c(t_k)$, $t_k \in [0, 1]$, i.e., $\tilde{C}_{k+1} = c(t_k)$ at $t_k = 1/(k+1)$ along the geodesic $c(t)$ ($t \in [0, 1]$), which connects the iterated mean $\tilde{C}_k$ to any random data point C_i , $i = 1, \dots, n$ at the k th iteration. The detailed process of computing the Fréchet sample mean using Sturm's algorithm could be extended from Algorithm 1.

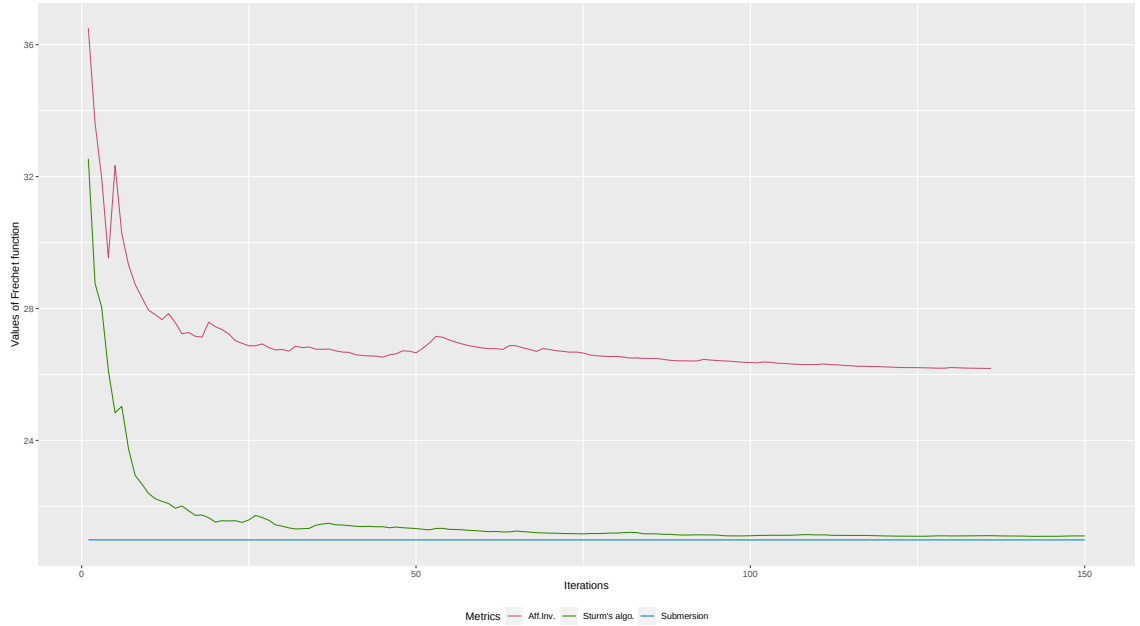


Figure 2.10: Values of the Fréchet function for the same data set in the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$ (red), using Sturm's algorithm on $(\mathcal{C}^+(p), g^{\text{quo}})$ (green), and using the submersion of the Fréchet sample mean in $\mathcal{S}^+(p)$ (blue).

We consider a time series of 116 correlation matrices with $p = 15$ for Seizure 1 within Patient ID18 (data information will be provided in Chapter 3), and compute Fréchet

sample means using the affine invariant and quotient geometries separately. Figure 2.10 shows the Fréchet function values using different Riemannian metrics for the same correlation matrices (this is applicable as the correlation matrices also lie on $(\mathcal{S}^+(p), g^{\text{aff}})$, i.e., $\mathcal{C}^+(p) \subset \mathcal{S}^+(p)$). We find that the Fréchet function values using the quotient metric are lower than those computed using the affine invariant metric. This is because the quotient metric aims to find the two optimal positions in $\mathcal{S}^+(p)$ by minimizing the affine invariant distance. Furthermore, Sturm's algorithm converged in the quotient geometry for this example. There are minor variations in the values of the Fréchet functions at each estimated iterated mean $\tilde{C}_k$ using Riquelme's method.

Chapter 3

Data reduction and preliminary analysis

Instead of modelling the raw time series of EEG data, researchers often analyse the covariance or correlation between channels to compress the information from all recording channels into a meaningful measure. In this chapter, we will introduce the basic information about an open-source EEG dataset from patients with epilepsy. Following this introduction, we will construct covariance and correlation matrix-valued time series data, as discussed in Section 3.1. The high dimensionality of raw covariance or correlation matrices presents a computational challenge for analysis. Additionally, electromagnetic artefacts and high correlations between channel sets can lead to rank deficiencies in the observed matrix-valued time series. These issues complicate the analysis further by introducing singularities. To address these challenges, Section 3.2 proposes two distinct dimensional reduction approaches. In Section 3.3, we continue our analysis by examining covariance and correlation matrix-valued time series EEG data that lie on different manifolds discussed in Chapter 2. We utilise multidimensional scaling plots for this purpose. Furthermore, we compare the Fréchet sample variances between interictal (between seizures) and ictal (seizure) periods to provide insights into the data.

3.1 Data acquisition

In this study, we analyse an open-source dataset of EEG recordings from patients with drug-resistant focal epilepsy, available from <http://ieeg-swez.ethz.ch/>. This data was collected at the Sleep-Wake-Epilepsy-Center (SWEC) of the University Department of Neurology at the Inselspital Bern (Burrello *et al.*, 2019). The full dataset includes 116 seizures captured during 2656 hours of long-term intracranial electroencephalography (iEEG) signals from 18 patients. The number of electrodes varies from 24 to 128 between

patients, and the length of recording varies from 2 to 12 days. Samples were recorded at a rate of either 512 or 1024 Hz, depending on the experimental subject. The stored iEEG data were provided in a preprocessed form. Specifically, signals were median-referenced and band-pass filtered with a broad bandwidth (0.5-120 Hz) using a 3rd order Butterworth filter (forward and backward). To ensure the reliability and validity of the iEEG recordings, an experienced epileptologist visually inspected all iEEG recordings to identify seizure onset and termination, and to exclude channels with persistent artefacts. These steps were conducted independently of this study and resulted in the publicly available data and annotations. All subjects formally consented to their iEEG data being used for research purposes. Without loss of generality, we will use ‘EEG’ rather than ‘iEEG’ in the following contexts as they both refer to electrical signals, with electrodes placed either on the scalp or on the exposed surface of the brain.

The initial dataset used in this study consisted of 18 patients (designated ID1 to ID18) who were part of the epilepsy surgery programme. The number of seizures for each patient varied from 2 to 23, while the total duration of interictal recordings ranged from 41 to 293 hours. To better understand differences in seizure evolution, for each seizure we identified a comparator interictal period exactly 2 hours before the seizure, lasting for the same duration as the seizure (as suggested by the epileptologist). This comparator period was unavailable for three seizures out of the total 116 seizures from 18 patients, and so these were removed from consideration (Seizure 1 and 20 of Patient ID9, Seizure 1 of Patient ID10). Additionally, one seizure with a duration of less than 10 seconds was also removed (Seizure 11 of Patient ID4), giving a remaining set of 112 seizures and associated interictal periods. Table 3.1 provides an overview of the number of electrodes, total number of seizures, overall recording duration, and seizure duration in seconds for each patient.

The complete dataset is extensive, with data recorded hundreds of times per second for thousands of hours across tens of electrodes in each patient. After the signal pre-processing mentioned above, an EEG recording for a single seizure consists of a time series $\{z_i \in \mathbb{R}^q\}$, where $i = 1, \dots, m$, and i represents the index across time. The coordinates $j = 1, \dots, q$ are referred to as channels, with each channel recording activity at a specific electrode. Each time series encompasses seizures along with interictal periods.

Rather than modelling the raw time series z_i of EEG signals, researchers often analyse the covariance (or correlation) between channels, as the primary scientific interest lies in dependencies between channels rather than the absolute measurement at each channel (Prabhu *et al.*, 2020). To calculate a sample covariance or correlation matrix, measurements z_i within a sliding window are used. While the use of a sliding window to derive time series of covariance or correlation matrices is common in EEG data analysis, it presents challenges. Specifically, the length of the window chosen can impact the dynamics revealed

Table 3.1: Patient information.

Subject	Duration (Hours)	No. of electrodes	No. of seizures	Seizure duration (Seconds)		
				Minimum	Maximum	Mean
ID01	293	88	2	589	613	601
ID02	235	66	2	86	89	88
ID03	158	64	4	60	68	64
ID04	41	32	13	31	68	44
ID05	110	128	4	15	17	16
ID06	146	32	8	29	126	45
ID07	69	75	4	14	98	69
ID08	144	61	4	17	413	189
ID09	41	48	21	22	136	40
ID10	42	32	16	61	106	70
ID11	212	32	2	83	99	91
ID12	191	56	9	106	194	146
ID13	104	64	7	40	188	103
ID14	161	24	2	46	60	53
ID15	196	98	2	69	119	94
ID16	177	34	5	120	245	190
ID17	130	60	2	97	98	98
ID18	205	42	5	71	300	199
Total	2656		112			
Average		58		86	169	122

by the data. To detect faster oscillations, which are known to play a crucial role in seizure initiation (Schindler *et al.*, 2011), we employ a sliding window of 1 second in this research, and the window contains 512 or 1024 samples, denoted as $f = 512/1024$, depending on subjects. This approach is also recommended by the dataset reference (Burrello *et al.*, 2019).

We assume that there are f EEG measurements per second and that we use non-overlapping sliding windows of exactly 1 second duration to compute covariance (or correlation) matrices, so that $m = f \times n$ where m is the length of the time series z_i and n is the length of the matrix-valued time series $\{S'_i \text{ (or } C'_i)\}$, for $i = 1, \dots, n$. Thereby, we obtain the full $q \times q$ sample covariance (or correlation) matrices

$$S'_i = \text{Cov}(\{z_{(i-1)f+1}, \dots, z_{if}\}),$$

$$C'_i = \text{Corr}(\{z_{(i-1)f+1}, \dots, z_{if}\}).$$

3.2 Dimensional reduction on covariance and correlation matrices

Following data acquisition and preprocessing, we derive the raw covariance matrix S'_i and correlation matrix C'_i , where $i = 1, \dots, n$. However, the large number of electrodes q may be overwhelming, requiring dimensional reduction to achieve computational tractability in the analysis. Moreover, electromagnetic artefacts and high correlations among sets of channels can lead to rank deficiencies in the observed matrix-valued time series. These singularities pose challenges to the analysis and signify redundancies in the data.

3.2.1 Methods

Some form of dimensional reduction is often performed over $\mathbb{R}^q$. This serves two purposes. First, the computation time of analysis typically increases polynomially in the number of channels q , but q can be large so dimensional reduction is required for computational feasibility. Secondly, certain linear combinations of channels can have very low variance. These represent a redundancy in the data set, and removing these speeds up analysis, as well as ensuring the covariance (or correlation) matrices obtained are strictly positive definite. The end result of these steps is a time series $\{S_i \text{ (or } C_i) : i = 1, \dots, n\}$ of $p \times p$ symmetric positive definite matrices where $p \leq q$, and i now indices the sequence of windows.

We propose two methods for the dimensional reduction. The first aims at selecting linear combinations of channels that maximize the variation in the raw time series z_i . The second aims to remove redundancies and identifies a set of channels to retain in the dimensionally-reduced data, as opposed to a set of linear combinations of channels.

Method I: Maximizing variation. Computing the full $q \times q$ sample covariance matrices $\{S'_i : i = 1, \dots, n\}$, we seek linear combinations of channels $u \in \mathbb{R}^q$ which maximize $\frac{1}{n} \sum_i u^T S'_i u$, where $(\cdot)^T$ denotes the transpose. In other words, we find the eigenvectors $u_1, \dots, u_p$ of $\frac{1}{n} \sum_i S'_i$ with the largest p eigenvalues. The time series of covariance (or correlation) matrices $\{S_i\}$ (or $\{C_i\}$) are then defined as the sample covariance (or correlation) of these linear combinations:

$$S_i = \text{Cov}(\{U z_{(i-1)f+1}, \dots, U z_{if}\}) = U S'_i U^T$$

$$\text{or } C_i = \text{Corr}(\{U z_{(i-1)f+1}, \dots, U z_{if}\}),$$

where the $p \times q$ matrix U has rows $u_1^T, \dots, u_p^T$.

For the interictal time series, dimensional reduction was performed (i) independently from the corresponding seizure time series and (ii) with the same U matrix obtained from dimensional reduction of the seizure data, thereby giving two comparators of interictal

time series. The first set of interictal series retains a higher proportion of variance, but parameters estimated from the data, such as the Fréchet mean of covariance matrices, cannot be directly compared to estimates obtained from the seizure data since they are in different coordinates. The second set of interictal series retains a lower proportion of variance, but direct comparisons can be made with the corresponding seizure data. We will elucidate the methods of dimensional reduction for interictal time series used according to the different scenarios in the following studies.

Method II: Minimizing redundancy. If a covariance (or correlation) matrix has a zero eigenvalue, then there is a linear combination of channels with zero variance. This implies redundancy in the data, as the value of one or more channels can be derived from the other channels, at least over the time interval represented by the covariance (or correlation) matrix. Our aim is to select channels to avoid such redundancies as follows. Suppose we have a set of channels $\mathfrak{C}$, and let $S'_i(\mathfrak{C})$ (or $C'_i(\mathfrak{C})$) denote the restriction of the full covariance (or correlation) matrix S'_i (or C'_i) to $\mathfrak{C}$. Also, define

$$\omega(\mathfrak{C}) = \frac{1}{n} \sum_i \min \sigma[S'_i(\mathfrak{C})]$$

$$\text{or } \omega(\mathfrak{C}) = \frac{1}{n} \sum_i \min \sigma[C'_i(\mathfrak{C})]$$

where $\sigma(A)$ denotes the set of eigenvalues of $A \in \text{Sym}(p)$, and the minimum is taken over this set for each i . A greedy algorithm was used to construct sets $\mathfrak{C}$ that maximize $\omega(\mathfrak{C})$ for a fixed value of p . Specifically, we first consider every set $\mathfrak{C}$ consisting of a pair of channels to find the optimal set $\mathfrak{C}_2$. Then, we consider every set $\mathfrak{C}$ containing $\mathfrak{C}_2$ and one additional channel to find the optimal set $\mathfrak{C}_3$, and so on, until the last channel has been added to $\mathfrak{C}$.

Similarly, two comparators are considered for the reduced interictal time series by Method II: (i) the same set of channels $\mathfrak{C}$ with the seizure data and (ii) the set of channels $\mathfrak{C}$ independently from the seizure. The advantages and disadvantages will be the opposite cases of these comparators for the reduced interictal time series by Method I. When analysing the linear relationship between pairs of channels, our primary interest lies in comparing the same set of channels with the interictal series. Therefore, we opt to use the same $\mathfrak{C}$ for seizures and corresponding interictal series for the series comparison. Additionally, the independent set of channels $\mathfrak{C}$ of the interictal period will be used to analyse the data dynamics of interictal series. We will expound upon the methods used in the subsequent studies according to the research purposes.

Discussion

Generally, the dimensional reduction of Method I decided above involves performing principal component analysis (PCA) on the average covariance matrix of EEG time series. The top p eigenvectors associated with the largest p eigenvalues of the average covariance matrix could be considered as a filter, such that the reduced covariance matrix of signals retains most of the variation of the raw data. Although this method has the great advantage of trying to capture the greatest variation in the data, the artefact suppression still exists, which makes the reduced covariance (or correlation) matrices potentially have small eigenvalues, ensuing matrices are not strictly positive definite. Furthermore, it loses interpretability since it works with linear combinations of channels, not individual channels.

The channel selection of Method II aims to identify the combination of individual channels, which bounds the eigenvalues of the matrices away from zero. This technique substitutes highly correlated covariance (or correlation) of channels with a single channel that exhibits a close evolutionary pattern of signals, resulting in small eigenvalues in covariance (or correlation) matrices. As a result, the reduced data strictly adheres to being strictly positive definite, lying within the Riemannian manifold $(\mathcal{S}^+(p), g^{\text{aff}})$ (or $(\mathcal{C}^+(p), g^{\text{quo}})$). However, a significant drawback of this approach is that it loses the information of channels which are removed.

3.2.2 Results

Since Method I and Method II have different properties, two statistics are used to evaluate the efficiency of these methods separately: (i) the proportion of variance retained in the reduced dataset, defined as

$$\frac{\sum_{i=1}^p \lambda_i}{\sum_{j=1}^q \lambda_j}, \quad (3.1)$$

where $\lambda_1 \geq \dots \geq \lambda_q \geq 0$ are the eigenvalues of the matrix $\frac{1}{n} \sum_{i=1}^n S'_i$, and (ii) observed $\omega(\mathfrak{C})$ in the reduced dataset with different chosen values of p .

Literally, the reduced covariance and correlation matrices from Method I are both obtained from the filtered signals U , i.e., $S_i = \text{Cov}(\{Uz_{(i-1)f+1}, \dots, Uz_{if}\})$ and $C_i = \text{Corr}(\{Uz_{(i-1)f+1}, \dots, Uz_{if}\})$, where the filter U is the top p eigenvectors associated with the largest p eigenvalues of the mean sample covariance matrices, $i = 1, \dots, n$ indices the sequence of windows, and f is EEG measurements per second. Henceforth, we will only show the proportion of explained variance for the covariance matrices as both covariance and correlation matrices have the same proportion of the explained variance of the reduced data. Figure 3.1 illustrates the proportion of explained variance for the reduced covariance matrix, as defined in Equation (3.1), across 112 seizures. This analysis was conducted

using both methods across three different dimensions: $p = 10, 15, 20$ in this research. It is evident that Method II of dimensional reduction, which identifies sets of channels, yields inferior results in terms of the proportion of variance retained during seizures.

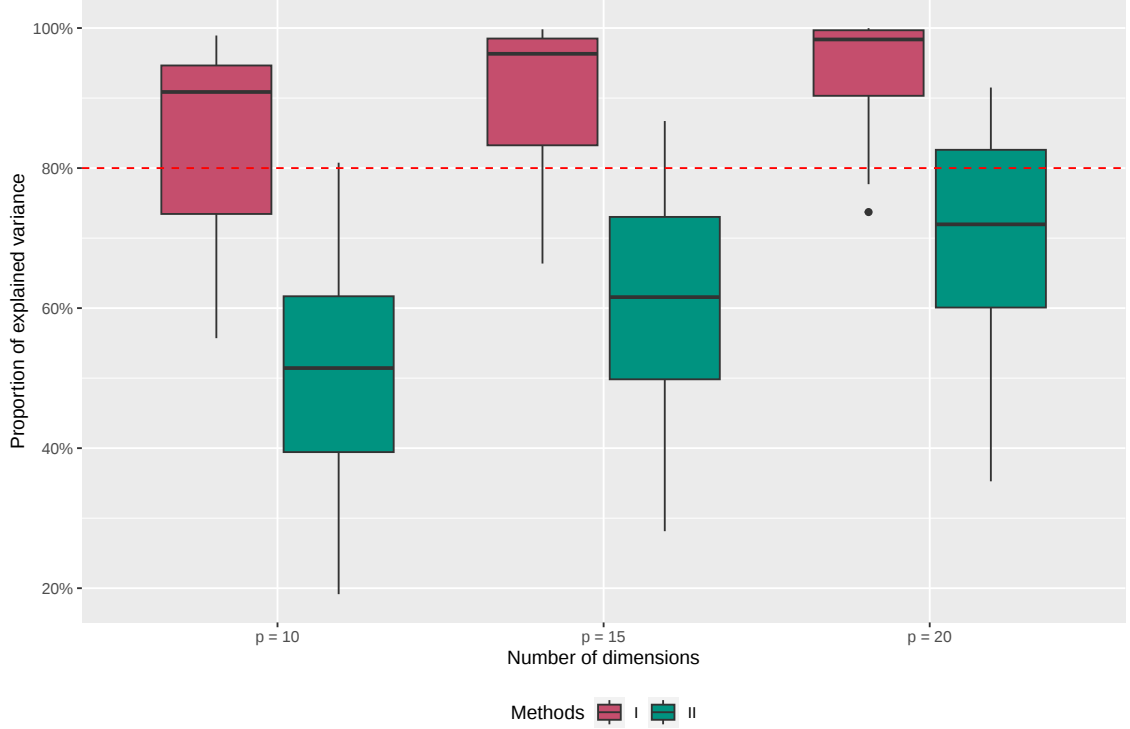


Figure 3.1: Proportion of variance retained in seizures after reduction to $p = 10, 15, 20$ dimensions. Each boxplot shows the distribution across 112 seizures. Method I and II are coloured by red and green, respectively.

In experimental EEG data, electrodes are placed by epileptologists to monitor the real-time signal dynamics of brain regions ensuring no two electrodes yield the exact same covariance between channels. Moreover, due to the nature of neural activities, the covariance magnitudes between channels are typically large, leading to very large values of $\omega(\mathfrak{C})$. Therefore, the reduced covariance matrices obtained from both methods are strictly positive definite. In contrast, channel correlations represent scaled data that factor out the channel variances, ranging from -1 to 1. This scaling can result in some eigenvalues of the correlation matrix approaching zero, leading to considerably smaller values of $\omega(\mathfrak{C})$. Therefore, Figure 3.2 only illustrates $\log_{10}(\omega(\mathfrak{C}))$ of the reduced correlation matrices across 112 seizures, following dimensional reduction to $p = 10, 15$, and 20, as well as the full q for the comparison. Opting for a smaller p in Method II allows for the elimination of highly correlated channels, replacing them with a single channel that closely mirrors the signal's evolutionary pattern. However, as the dimension p increases, the advantages of Method II over Method I begin to diminish. This is because additional channels can explain more

variance, but they also introduce more highly correlated channel pairs.

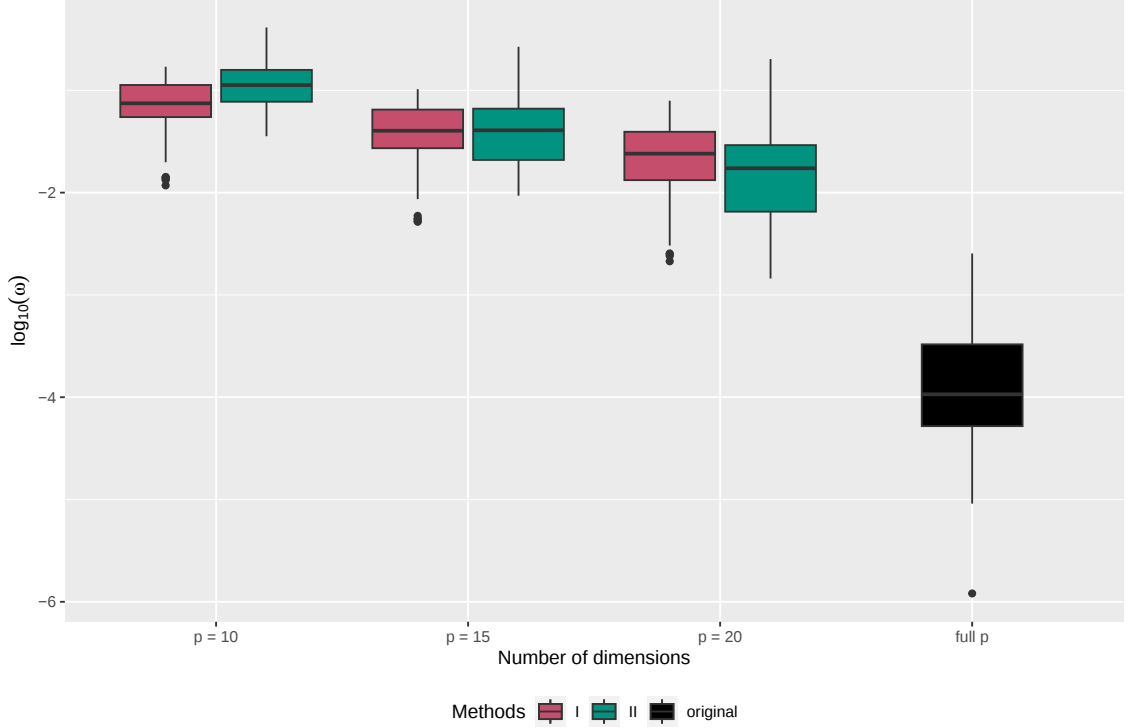


Figure 3.2: $\log_{10}(\omega(\mathfrak{C}))$ values from Method I (coloured by red) and Method II (coloured by green) with respect to the reduced data set for correlation matrices across 112 seizures choosing $p = 10, 15$, and 20. The value of $\log_{10}(\omega(\mathfrak{C}))$ for original full $q \times q$ correlation matrices is coloured by black.

Based on the properties of Methods I and II, along with the insights provided by Figures 3.1 and 3.2, we have chosen to employ Method I for reducing the dimensionality of covariance matrices, setting $p = 15$. This choice enables us to capture over 80% of the explained variance from the original data for more than 80% of seizure time series. Thereby, we opt to use $p = 15$ for all patients and seizures with a reduced covariance matrix. Conversely, for correlation matrices, we opt for Method II with the objective of minimizing channel redundancies. While Method II does not significantly outperform Method I in terms of the $\log_{10}(\omega(\mathfrak{C}))$ values for the reduced correlation matrices, it offers more interpretable results for modelling functional brain networks. This is because Method II preserves more channel information in the reduced data, making the functional network modelling more explainable. To maintain consistency and ensure that all $\omega(\mathfrak{C})$ values are larger than 10^{-2} , we finalize our choice of $p = 15$ for the reduced correlation data.

Figures 3.1 and 3.2 only list results of dimensional reduction with three dimensions $p = 10, 15$ and 20 and we finally choose $p = 15$ for both the reduced covariance and correlation matrices in this thesis based on the above discussions. Notably, these may not be the best and unique choice of p , readers can set different reduced dimensions p

according to different data sets and discover the corresponding results.

3.3 Data exploration

In this section, an initial analysis of seizure and interictal matrix-valued time series data is carried out using the Euclidean, the affine invariant, and the quotient geometries. Notably, the interictal time series was prepared with dimensional reduction independent from the seizure time series in this section. This analysis involves: (i) calculating sample means and variances, and (ii) performing multidimensional scaling (MDS) (Mead & A., 1992). Due to the substantial quantity of the data, we do not present a comprehensive set of results; instead, we provide results for a representative selection of patients. Patient ID18's results will be emphasized throughout the thesis, and results for Patient ID6, ID7, and ID13 can be found in Appendix A. The selection of these patients is based on the number and duration of seizures, aiming for representativeness. Specifically, Patient ID6, ID7, ID13, and ID18 were chosen as they have more than 2 seizures but less than 8, with the occurrence and duration of seizures varying as much as possible within each patient. This information for these patients is highlighted in bold in Table 3.1.

3.3.1 Covariance matrices in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$

Figure 3.3 displays the Fréchet sample variances for each seizure time series and the corresponding interictal time series, categorized by patient ID, using the affine invariant metric. The plot illustrates a significant increase of the variance in seizure series compared to interictal series for all patients. We do not show the sample variance for time series using the Euclidean metric as the sample variance of seizures is considerably larger than that of corresponding interictal time series by orders of magnitude.

Figure 3.4 shows the results of MDS plots of the seizure time series and one interictal series for Patient ID18, using Euclidean and the affine invariant metrics respectively. Note that metric MDS is used throughout this thesis. It minimizes the loss function which is the residual sum of squares:

$$\text{Stress} = \sqrt{\sum_{i \neq j=1, \dots, n} (d_{ij} - \|x_i - x_j\|)^2}, \quad (3.2)$$

where d_{ij} is the Riemannian manifold metrics, e.g. $d_{g^{\text{euc}}}, d_{g^{\text{aff}}}$; x_i, x_j are points in $\mathbb{R}^2$ after MDS. Optimization is performed by stress majorization. Lower stress values indicate a better fit. The stress values with percentages are shown behind the seizure titles within brackets in each MDS plot in Figure 3.4. The interictal time series were prepared with dimensional reduction independent from the seizure time series. MDS plots with the affine invariant metric for seizures 1, 2, 4, and 5 are quite similar: the plots show evolution in

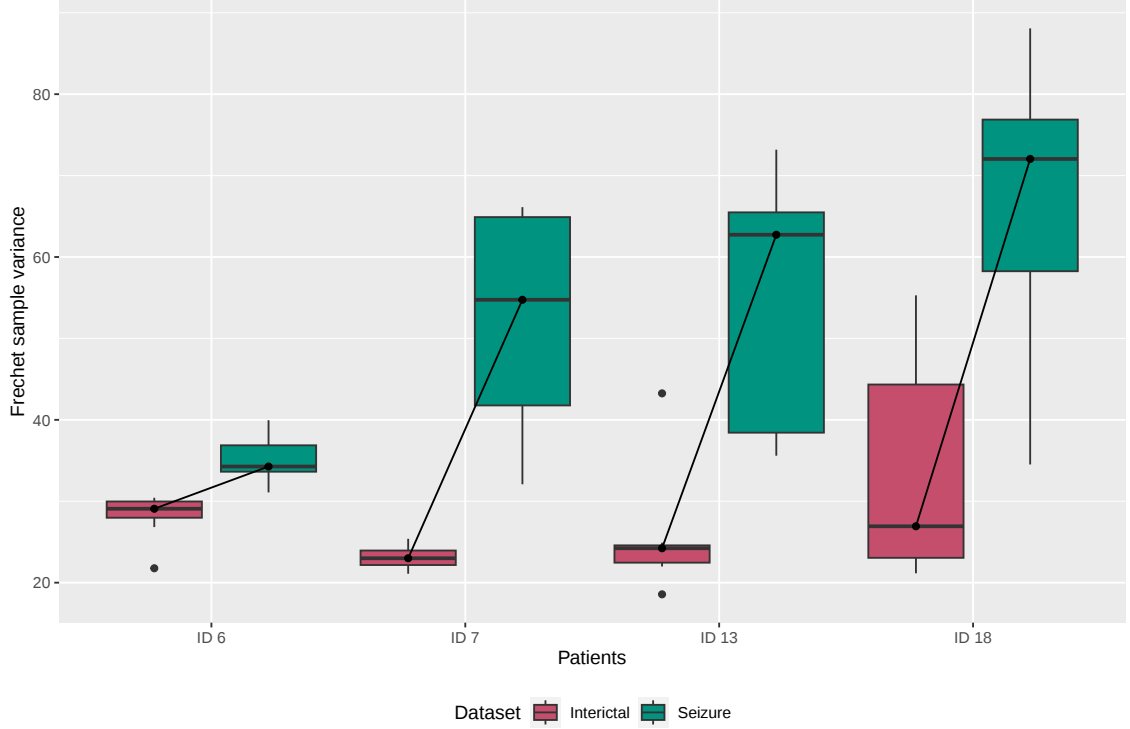


Figure 3.3: The Fréchet sample variance for each seizure time series and the corresponding interictal time series, broken down by patient ID, using the affine invariant metric. Dimensional reduction was performed independently for the interictal series. The median points of the interictal and seizure distributions for each patient are linked by a line to highlight the increase.

each case along a curving trajectory but with a degree of local noise. Seizure 3 is of shorter duration and a curved trajectory is less apparent. For the interictal period, there is no apparent trajectory over time, and the plot largely resembles noise. This was also the case for the interictal periods for the other seizures.

3.3.2 Correlation matrices in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$

Similarly, we conduct a preliminary analysis on seizure and interictal time series data for full-rank reduced correlation matrices. Assuming the uniqueness of the Fréchet sample mean in quotient geometry $(\mathcal{C}^+(p), g^{\text{quo}})$, as discussed in Section 2.4.7, Figure 3.5 illustrates the Fréchet sample variances for all seizures and their corresponding interictal time series. In general, more variation is observed in seizures compared to their corresponding interictal time series. However, in some seizures, larger variation is exhibited in interictal time series than in seizures. This may be explained by the loss of information from dimensional reduction, leading to the higher Fréchet sample variance in interictal time series.

To visualize trajectories of seizure time series for Patient ID18 under different geometries, we compute different distance matrices for the reduced correlation matrices using the

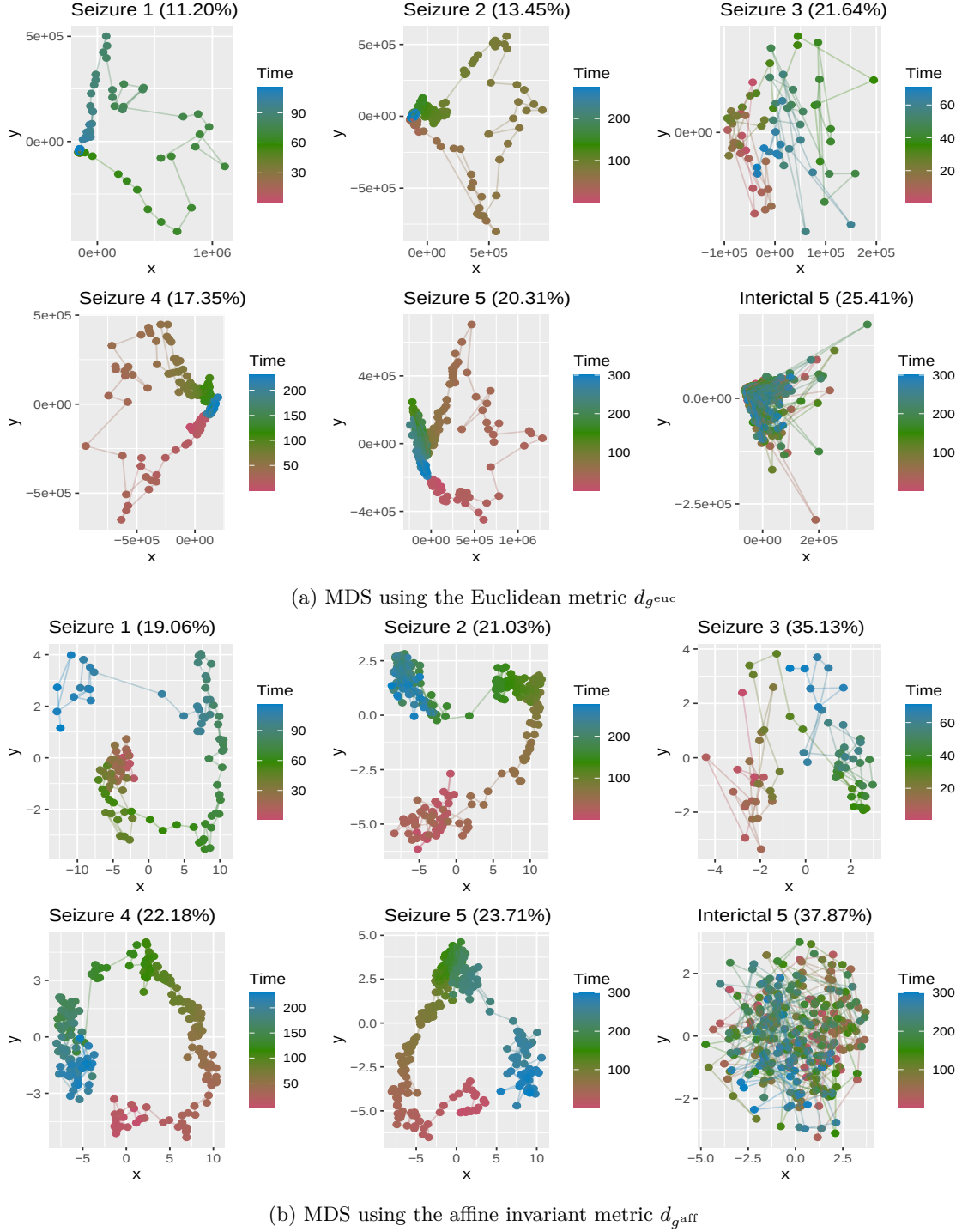


Figure 3.4: MDS of seizure time series for Patient ID18 using the Euclidean and affine invariant metrics to construct matrices of distance d_g^{euc} in (3.4a) and d_g^{aff} in (3.4b). Patient ID18 has 5 seizures (first 5 panels). Panel 6 shows the result for the interictal period corresponding to Seizure 5. Plots are coloured to show the development of time series over time. Each panel title in brackets shows the stress majorization of 2-dimensional MDS.

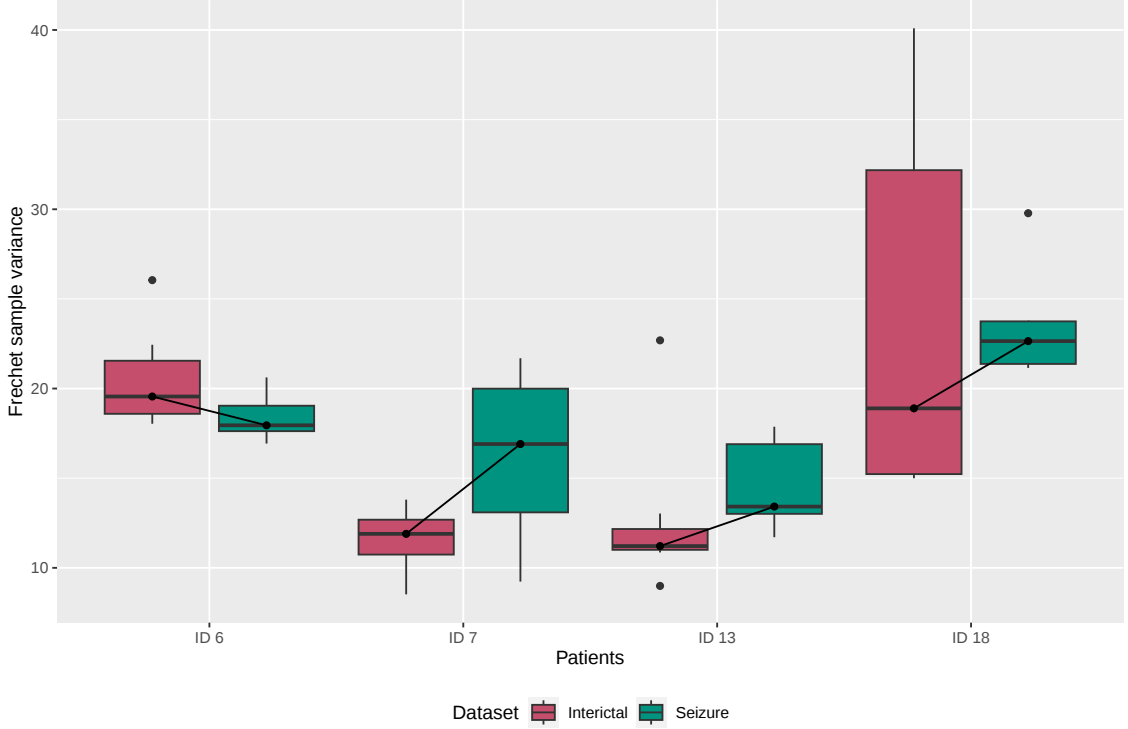


Figure 3.5: The Fréchet sample variance for each seizure time series and the corresponding interictal time series, broken down by patient ID on quotient geometry. Interictal series keep the same selected channels by dimensional reduction to seizure series. The median points of the interictal and seizure distributions for each patient are linked by a line to highlight the increase.

Euclidean metric on $(\text{Sym}(p), g^{\text{euc}})$, the affine invariant metric on $(\mathcal{S}^+(p), g^{\text{aff}})$, and the quotient metric on $(\mathcal{C}^+(p), g^{\text{quo}})$ (See Figure 3.6). This is due to the fact that a correlation matrix is a symmetric positive definite matrix with 1s on the diagonal, i.e., existing as nested manifolds $\mathcal{C}^+(p) \subset \mathcal{S}^+(p) \subset \text{Sym}(p)$.

In Figure 3.4, there are distinguishing differences of MDS plots between the Euclidean and affine invariant metrics with respect to covariance matrices, and the shape of trajectories in the affine invariant metric is easily discernible. However, we can rarely find any particular patterns of trajectories, neither with the Euclidean metric, the affine invariant metric, nor the quotient metric for correlation matrix-valued time series data in Figure 3.6. Compared to covariance matrices with a high magnitude of variance/covariance, a correlation matrix is the scaled data explaining the linear relationship between pairs of channels. This explains why the ranges of the x and y axes in the three sub-figures are close to each other. MDS plots using the quotient metric have similar percentages of stress to each other. Although the MDS results under the quotient geometry do not show advantages over other geometries, these are more apparent when we perform statistical modelling in subsequent chapters.

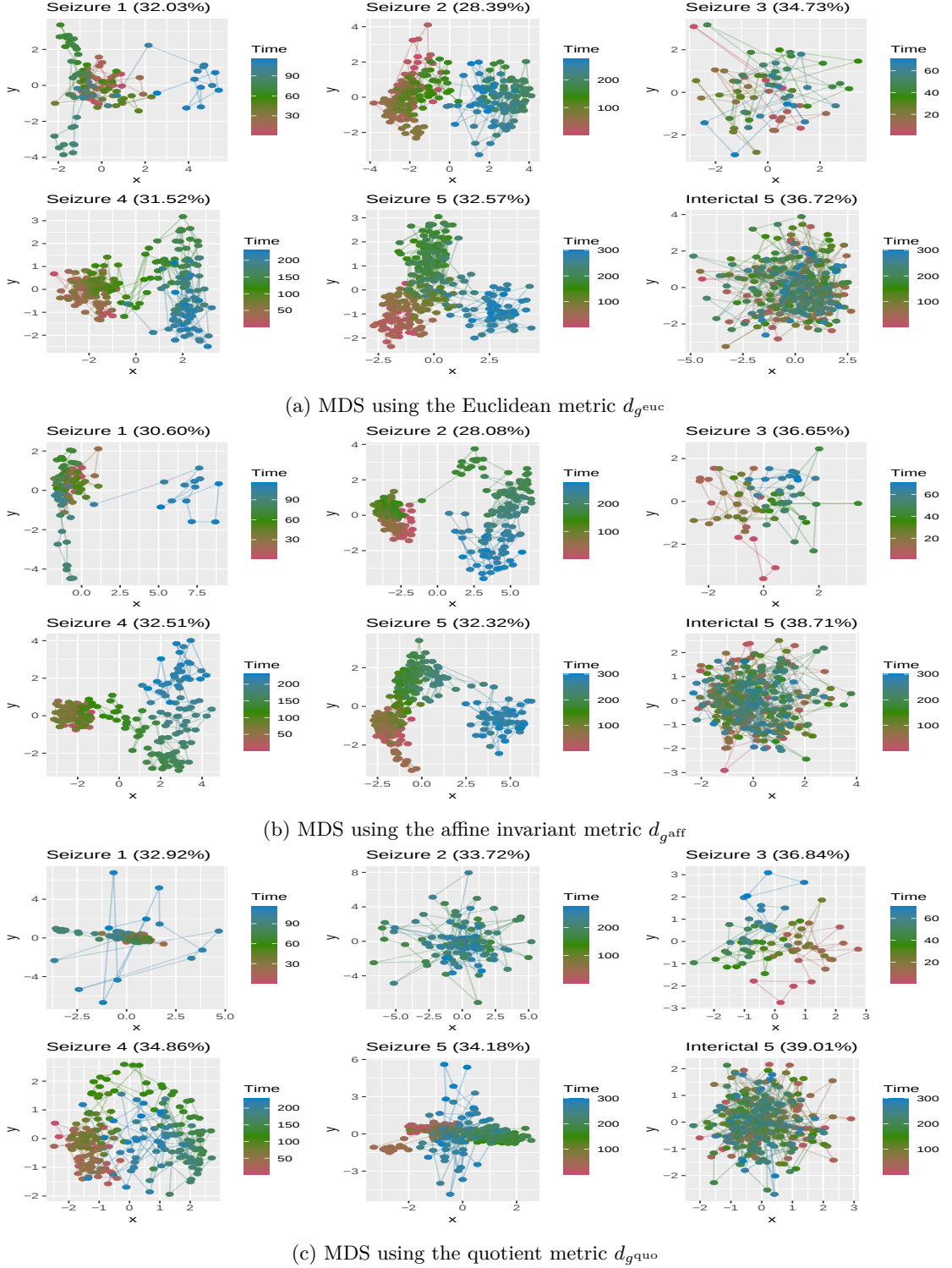


Figure 3.6: MDS of seizure time series for Patient ID18 using the Euclidean, affine invariant, and quotient metrics to construct matrices of distance $d_{g^{euc}}$ in (3.6a), $d_{g^{aff}}$ in (3.6b), and $d_{g^{quo}}$ in (3.6c). Patient ID18 has 5 seizures (first 5 panels). Panel 6 shows the result for the interictal period corresponding to Seizure 5. Plots are coloured to show the development of time series over time. Each panel title in brackets shows the stress majorization of 2-dimensional MDS plots.

Chapter 4

Manifold-adapted models for matrix-valued time series data

In Chapter 2, we discussed different Riemannian geometries for time series of covariance and correlation matrices. We also conducted preliminary analysis on the EEG time series dataset in Chapter 3. To capture and delineate diverse modes of evolution within time series of seizures, comprising covariance and correlation matrices, this chapter constructs manifold-adapted time series models and corresponding inference through maximum likelihood estimation (MLE).

Section 4.1 outlines the rationale behind the manifold-adapted model, which includes three terms: an autoregressive term, a mean-reverting term, and a noise term, each elucidating a distinct brain dynamics pattern in Section 4.2. Subject to the specific model assumptions discussed in Section 4.4, we employ MLE to find the coefficients in manifold-adapted models and discuss their asymptotic distribution in Section 4.3. Lastly, Section 4.5 simulates covariance and correlation matrices under the framework of manifold-adapted models in the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$ and the quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$. Simulation results underscore the validity and efficiency of the model specification and inference, importantly affirming the robust performance of our code in the R package: `geomTS`.

4.1 Rationale

When modelling a time series of $p \times p$ covariance (or correlation) matrices $\{S_i : i = 1, \dots, n\}$, a natural approach is to express the $(i + 1)$ th matrix as follows:

$$S_{i+1} = S_i + V_i + \epsilon_i \tag{4.1}$$

Here, $V_i \in \text{Sym}(p)$ serves as a deterministic perturbation, potentially dependent on preceding matrices in the time series or some covariates, and ϵ_i represents noise. Section 1.3 and Figure 1.1 demonstrate the necessary modification of the additive model in Equation (4.1) to account for the non-Euclidean nature of the space of $p \times p$ positive definite matrices. This research will work the data set $\{S_i : i = 1, \dots, n\}$ intrinsically within Riemannian geometries. The notion of perturbing a matrix S_i in some direction V_i , expressed in Equation (4.1) by simple addition, is instead achieved by the Riemannian exponential map, which operates intrinsically.

There are two types of matrix-valued time series data: covariance matrices (referred to as S_i) and correlation matrices (referred to as C_i). In light of the definitions pertaining to manifold-valued data and their corresponding properties, our initial analysis centres on the time series of covariance matrices $\{S_i : i = 1, \dots, n\}$. We approach this analysis by considering the data as residing within the Riemannian geometries on $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$. Concurrently, we consider the time series of correlation matrices $\{C_i : i = 1, \dots, n\}$, which are derived from the covariance matrices S_i and also lie on $(\mathcal{S}^+(p), g^{\text{aff}})$. In this regard, we analyse the time series $\{C_i\}$ by regarding the data as situated within Riemannian geometries on $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$, in turn. The primary objective here is to discern dissimilarities in the outcomes of our modelling efforts and to observe how these disparities vary based on the underlying geometries we have imposed.

While our central focus revolves around processed EEG data represented as covariance (or correlation) matrices, it is important to note that we adopt a cohesive approach within a single overarching model. This model is adept at accommodating the diverse dynamics intrinsic to the time series $\{S_i\}$ (or $\{C_i\}$), aiming to yield a set of interpretable parameters for each data set. Specifically, the tangent vector V_i (or X_i) is modelled as the sum of three terms:

- (1) An autoregressive term which combines direction vectors at previous lagged time points. For certain parameter values, this term can, for example, give rise to approximate movement along a geodesic in the manifold if this term dominates the other two terms. In the Euclidean setting, this would correspond to drift in an approximately constant direction.
- (2) A mean reverting term. It is plausible that the brain has some ‘mean’ state about which it varies, and that observed EEG data should also have a tendency to return to a certain point or region of the state space. In the absence of the autoregressive term, the mean reverting and noise term could be considered as a discrete-time analogue of an Ornstein-Uhlenbeck process (Maller *et al.*, 2009).
- (3) A noise term. This is a multivariate normal distribution on the tangent space, which

gives rise to the so-called wrapped Gaussian distribution on the manifold as described in Section 2.1.2. If this term dominates, then the time series will resemble a type of random walk over the manifold.

By combining these terms, the model is able to capture various possible dynamics for the evolution of EEG data, from smooth flow along geodesics to a noisy mean reverting random walk on the underlying manifold, for example.

4.2 Model specification

In this section, we propose a model for a time series $\{S_i : i = 1, \dots, n\} \in \mathcal{M}$ for a parallelizable Riemannian manifold $(\mathcal{M}, g)$. A simpler version is also specified for manifolds that are not parallelizable. The general model is formulated in terms of distributional assumptions for S_{i+1} based on the preceding values $S_i, \dots, S_1$. Specifically, we develop a linear model in the tangent space $T_{S_i}\mathcal{M}$ for the direction vector $V_i = \text{Log}_{S_i}(S_{i+1})$ of S_{i+1} at the base point S_i . As described in Section 4.1, the expression for V_i has three terms, each corresponding to a different possible type of dynamics for the time series.

4.2.1 General model

Here, we set up notation and specify a general version of our model. Let $V_i = \text{Log}_{S_i}(S_{i+1}) \in T_{S_i}\mathcal{M}$, for $i = 1, \dots, n-1$, be the direction vector at time i and define:

$$V_{il} = \mathcal{P}_{S_{i-l} \rightarrow S_i}(V_{i-l})$$

for $l = 1, \dots, L$ and $i > L$. Here, L is some fixed maximum lag, and V_{il} is the parallel translation of the direction vector V_{i-l} at S_{i-l} to the tangent space $T_{S_i}\mathcal{M}$ at time i . The model for S_{i+1} is specified by:

$$\begin{aligned} S_{i+1} &= \text{Exp}_{S_i}(V_i) \\ V_i &= \sum_{l=1}^L A_l(S_i)V_{il} + B\text{Log}_{S_i}(S^*) + \epsilon_i, \\ \epsilon_i &\sim N_{T_{S_i}\mathcal{M}}(0, \Sigma_{S_i}) \end{aligned} \tag{4.2}$$

The terms in the model are defined as follows:

1. **Autoregressive Term:** The sum $\sum_l A_l(S_i)V_{il}$ is a manifold-adapted version of a conventional vector autoregressive (VAR) model. For each l , $A_l(S)$ is an endomorphism of $T_S\mathcal{M}$, which varies smoothly with S . Each term in the sum results from a linear map acting on the lagged direction vector V_{i-l} parallel transported to $T_{S_i}\mathcal{M}$.

Given a parallel orthonormal basis $\{\omega_r(S) : r = 1, \dots, m\}$ of $T_S\mathcal{M}$, one way to determine the bundle endomorphisms $A_l(S)$ is via a constant matrix representation in this basis. Each matrix A_l is of dimension $m \times m$, where m is the dimension of $\mathcal{M}$ (for example, $m = \frac{1}{2}p(p+1)$ when $\mathcal{M} = \mathcal{S}^+(p)$, and $m = \frac{1}{2}p(p-1)$ when $\mathcal{M} = \mathcal{C}^+(p)$). Defined in this general way, the model has a very large number of parameters, and we will discuss lower dimensional parametrizations below.

2. **Mean Reversion Term:** Like the terms $A_l(S)$, the term $B(S)$ is an endomorphism of $T_S\mathcal{M}$ which varies smoothly with S . The vector $\text{Log}_{S_i}(S^*)$ is the direction from the current point towards a fixed ‘attractor’ point $S^* \in \mathcal{M}$. The attractor point $S^* \in \mathcal{M}$ can either be fixed a priori or estimated in some way from the data. We do not call S^* a mean point; it is not necessarily the Fréchet mean of a set of data generated according to the model, for example. In the same way as for the autoregressive terms if $\mathcal{M}$ is parallelizable, we can use a fixed $m \times m$ matrix B and the parallel basis $\{\omega_r(S)\}$ to define $B(S)$ globally. The term $B\text{Log}_{S_k}(S^*)$ models a tendency to move towards the attractor point S^* , and we refer to this as a mean reversion term.
3. **Noise Term:** The distribution of the random tangent vector $\epsilon_i \in T_{S_i}\mathcal{M}$ is determined by a covariance matrix $\Sigma_{S_i} \in \text{Sym}(m)$, and this definition implicitly depends on the existence of some orthonormal basis of $T_{S_i}\mathcal{M}$. Using the parallel orthonormal basis $\{\omega_r(S)\}$, we can take Σ_{S_i} to be a fixed matrix in this basis, denoted as Σ .

We model each seizure independently within and between patients. In order to better understand differences in seizure evolution, we identify a comparator interictal period exactly 2 hours before the seizure, and lasting for the same duration as the seizure. Note that each paired seizure and interictal series are prepared using the same dimensional reduction to ensure that the Fréchet sample mean of the interictal series is in the same basis as the seizure series.

4.2.2 Vectorization of tangent vectors

In order to conduct statistical analysis of the geometric data, it is essential to establish an orthonormal coordinate system that isomorphically maps the tangent space $T_S\mathcal{M}$ onto $\mathbb{R}^m$, which is equipped with a canonical metric (Pennec *et al.*, 2006). This can be accomplished by calculating the Riemannian metric between the tangent vector $V \in T_S\mathcal{M}$ and the orthonormal basis $\{\omega_r(S) : r = 1, \dots, m\}$ in the tangent space $T_S\mathcal{M}$ as:

$$v_r = g_S\langle V, \omega_r(S) \rangle, \quad r = 1, \dots, m,$$

where g_S represents the Riemannian metric on $(\mathcal{M}, g_S)$. In general, any tangent vector $V \in T_S \mathcal{M}$ can then be vectorized as a coordinate vector $\mathbf{v} = (v_1, \dots, v_m)^T \in \mathbb{R}^m$ by vectorization $Vec(V)$. In our proposed general model (4.2), we need to vectorize three different types of tangent vectors: the response vector $V_i = \text{Log}_{S_i}(S_{i+1})$, the parallel translated lagged vector $V_{il} = \mathcal{P}_{S_{i-l} \rightarrow S_i}(V_{i-l})$, and the mean-reverting vector $V_{i0} = \text{Log}_{S_i}(S^*)$.

Here, we recall the vectorization $Vec(V)$ in different Riemannian manifolds: $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$, all defined in Chapter 2. The vectorization $Vec(V)$ of the tangent vector $V \in T_S \text{Sym}(p)$ in the Euclidean geometry on $(\text{Sym}(p), g^{\text{euc}})$ is:

$$v_r = g_S^{\text{euc}} \langle V, E_r \rangle = \text{Tr}(V E_r), \quad r = 1, \dots, m,$$

where the orthonormal basis $\{E_r\}$ is defined in Equation (2.9) and the inner product is independent of the base $S \in \text{Sym}(p)$. m is the dimension of the Euclidean geometry. It follows that $\mathbf{v} = (V_{11}, \sqrt{2}V_{12}, \dots, \sqrt{2}V_{1p}, V_{22}, \sqrt{2}V_{23}, \dots, \sqrt{2}V_{2p}, \dots, V_{pp})^T \in \mathbb{R}^{p(p+1)/2}$ if S_i is a covariance matrix, i.e., the vector $\mathbf{v}$ consists of the diagonals of matrix V and its upper triangular elements multiplied by $\sqrt{2}$. Analogously, when we consider correlation matrices $\{C_i : i = 1, \dots, n\}$, the dimension of $\mathcal{M}$ will be $(p-1)p/2$, henceforth, $\mathbf{v} = (V_{12}, \dots, V_{1p}, V_{23}, \dots, V_{2p}, \dots, V_{(p-1)p})^T \in \mathbb{R}^{(p-1)p/2}$, which are the elements from the strict upper triangular matrix of matrix V .

In $(\mathcal{S}^+(p), g^{\text{aff}})$, the vectorization $Vec(V)$ is defined in Equation (2.20) as:

$$v_r = g_S^{\text{aff}} \langle V, S^{1/2} E_r S^{1/2} \rangle = \text{Tr}(S^{-1/2} V S^{-1/2} E_r), \quad r = 1, \dots, m.$$

Subsequently, the coordinate vector $\mathbf{v}$ by the vectorization $Vec(V)$ is isometric with the tangent space $T_S \mathcal{M}$ and it is computed as $\mathbf{v} = (W_{11}, \sqrt{2}W_{12}, \dots, \sqrt{2}W_{1p}, W_{22}, \dots, W_{pp})^T \in \mathbb{R}^m$, where $W = S^{-1/2} V S^{-1/2}$ and $m = p(p+1)/2$.

In $(\mathcal{C}^+(p), g^{\text{quo}})$, recalling the vectorization $Vec(X)$ in the quotient manifold as defined in Equation (2.39) for every $X \in T_C \mathcal{C}^+(p)$, we have

$$\mathbf{v} = g_C^{\text{quo}} \langle X, \mathbf{F}^C \rangle = \text{Tr} \left(C^{\dagger -1/2} \text{hor}(\Delta_{C^\dagger} X \Delta_{C^\dagger}) C^{\dagger -1/2} F_r \right),$$

where $C^\dagger \in \mathcal{S}^+(p)$ is the optimal position with respect to the identity matrix I_p , and $\{F_r : r = 1, \dots, m\}$ is the orthonormal basis in $T_{I_p} \mathcal{C}^+(p)$. Since X is a hollow matrix with vanishing diagonals, the dimension of $\mathbf{v} \in \mathbb{R}^m$ is $m = (p-1)p/2$.

Subsequently, the final model with reduced parameters from the general model (4.2) can be represented as

$$\mathbf{v}_i = \sum_{l=1}^L A_l \mathbf{v}_{il} + B \mathbf{v}_{i0} + \boldsymbol{\epsilon}_i, \quad (4.3)$$

where $\mathbf{v}_i, \mathbf{v}_{il}$, and $\mathbf{v}_{i0}$ are all mapped vectors in $\mathbb{R}^m$, representing the tangent vector V_i ,

lagged tangent vectors V_{il} , and mean-reverting vector V_{i0} , separately, and $\epsilon_i \sim N(0, \Sigma)$. $A_1, \dots, A_L, B, \Sigma, S^*$ and the lag L are unknown parameters in the model as defined above.

4.2.3 Models with reduced parameters

In a conventional VAR model, the model coefficients A_l and B should be $m \times m$ matrices. Therefore, there are $(L+2)m^2$ unknown autoregressive parameters (including mean reversion and noise terms). However, considering the potentially large number of parameters required by the general model (4.3) is unnecessary for the following reasons:

- (1) In the practical analysis of EEG data, a comprehensive evaluation of interaction effects between pairs of channels is infrequent. Typically, the analysis focuses solely on the evolution pattern of data for individual channel pairs. This implies components in the vectorized vector $\mathbf{v} \in \mathbb{R}^m$ are independent. In this case, the effective structure of regression coefficients A_l can be represented as a diagonal matrix, denoted by $A_l = \text{Diag}(a_{l1}, \dots, a_{lm})$ for $l = 1, \dots, L$.
- (2) The reduced model, characterized by $(L+2)m$ parameters as described above, remains challenging for the general interpretation of patterns in brain dynamics. It is assumed that the regression coefficients $a_{lr}, r = 1, \dots, m$ for the r th pair of coordinates are independently and normally distributed with a mean of $\alpha_l \in \mathbb{R}$. As a result, we can simplify A_l to scalar coefficients, denoted as $A_l = \text{Diag}(a_{l1}, \dots, a_{lm}) = \alpha_l I_m$, where I_m represents an identity matrix. This simplification facilitates a more straightforward quantification of the strength of directional influence from V_{i-l} to V_i .
- (3) Particularly, when the Riemannian manifold is not parallelizable, the time series model (4.3) with scalar coefficients is well-defined. This is due to the unavailability of a basis with a dimension of m that is orthogonal. After applying the vectorization $\text{Vec}(\cdot)$, the model with scalar coefficients disregards the linear relationship among bases and consolidates all components with the same regression coefficient for each lagged vector.

In summary, there are two types of lower reduced parameters in the manifold-adapted models which we fit to data in practice.

Diagonal model. We assume that the matrices A_ℓ , B , and Σ are all diagonal:

$$\begin{aligned} A_l &= \text{Diag}(a_{l1}, \dots, a_{lm}), \\ B &= \text{Diag}(b_1, \dots, b_m), \quad \text{and} \\ \Sigma &= \text{Diag}(\sigma_1^2, \dots, \sigma_m^2). \end{aligned}$$

It should be noted that these models are diagonal in the m -dimensional tangent space $T_S \mathcal{M}$. There is one parameter for every pair of coordinates. For $p = 15$, we have $m = 120$,

so for example, the reversion term and the noise term each involve 120 scalar parameters.

Scalar coefficient model. We assume that the matrices A_ℓ , B , and Σ are all scalar multiples of the identity on $T_S\mathcal{M}$:

$$A_\ell = \alpha_\ell I_m, \quad B = \beta I_m, \quad \Sigma = \sigma^2 I_m.$$

Generally, with this set of assumptions, the model is defined independently of the basis $\{\omega_r(S)\}$ and so it can be applied to a manifold which is not parallelizable.

4.3 Model inference and selection

For a fixed L and attract point S^* , we introduce model inference through MLE for the remaining parameters. Estimating these parameters reduces to a set of linear regressions within each tangent space, as discussed in Subsection 4.3.1. Subsequently, Subsection 4.3.2 computes the asymptotic covariance matrix of these estimates. Subsection 4.3.3 discusses methods of model selection to determine the optimal maximum value for L .

Differentiating the likelihood function with respect to S^* is not straightforward, making joint MLE of S^* and the other parameters challenging. As a result, when analysing a seizure time series $\{S_i : i = 1, \dots, n\}$, we fix S^* to be the Fréchet sample mean of the corresponding interictal series, assuming the brain has a tendency to return to the ‘rest’ state. Furthermore, to ensure that the Fréchet sample mean of the interictal series is in the same basis as the seizure series, this research applies the same dimensional reduction to paired seizure and interictal series.

4.3.1 Model inference

MLE is a statistical approach that allows for the estimation of probability distribution parameters. To perform MLE, a probability distribution is initially selected to represent the observed data. Subsequently, using the observed data, the likelihood of the data is calculated for different values of the distribution’s parameters. In order to facilitate mathematical calculations, it is often convenient to take the natural logarithm of the likelihood function and then differentiate it with respect to each parameter. By setting the derivatives equal to zero, the resulting equations can be solved to obtain the parameter values that maximize the likelihood function. This approach allows for the determination of the parameter values that are most likely to have generated the observed data.

Assuming ϵ is normally distributed, the likelihood function of our manifold-adapted

model (4.3) is

$$\begin{aligned} \mathcal{L}(A_1, \dots, A_L, B, \Sigma | \mathbf{v}) &= \prod_{i=L+1}^n (2\pi)^{-\frac{m}{2}} |\Sigma|^{-\frac{1}{2}} \\ &\times \exp \left\{ -\frac{1}{2} \left(\mathbf{v}_i - \sum_{l=1}^L A_l \mathbf{v}_{il} - B \mathbf{v}_{i0} \right)^T \Sigma^{-1} \left(\mathbf{v}_i - \sum_{l=1}^L A_l \mathbf{v}_{il} - B \mathbf{v}_{i0} \right) \right\}. \end{aligned} \quad (4.4)$$

The following two subsections give the detailed description of MLE, considering two different types of reduced parameters in the model involving scalar coefficients and diagonal matrix coefficients. We also discuss the asymptotic covariance of the estimated parameters in both reduced parameter models but in the next subsection.

MLE for the model with scalar coefficients

Assume all coefficients are scalar values and denote $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_L, \beta)^T$ and $\Sigma = \sigma^2 I_m$. Taking the logarithm of likelihood function $\mathcal{L}$ calculated in Equation (4.4) for the scalar coefficient model, we have

$$\ell = -\frac{m(n-L)}{2} \log(2\pi) - m(n-L) \log(\sigma) - \frac{1}{2\sigma^2} \sum_{i=L+1}^n \left\| \mathbf{v}_i - \sum_{l=1}^L \alpha_l \mathbf{v}_{il} - \beta \mathbf{v}_{i0} \right\|^2, \quad (4.5)$$

where $\|\cdot\|^2$ is the Euclidean norm. For example, $\|\mathbf{x}\|^2 = \sum_{i=1}^n x_i^2$ for any $\mathbf{x} \in \mathbb{R}^n$.

Subsequently, we can calculate its first derivatives with respect to $\alpha_l (l = 1, \dots, L)$ and β as

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha_l} &= \frac{1}{\sigma^2} \sum_{i=L+1}^n \mathbf{v}_{il}^T (\mathbf{v}_i - \sum_{l=1}^L \alpha_l \mathbf{v}_{il} - \beta \mathbf{v}_{i0}) = 0 \\ \frac{\partial \ell}{\partial \beta} &= \frac{1}{\sigma^2} \sum_{i=L+1}^n \mathbf{v}_{i0}^T (\mathbf{v}_i - \sum_{l=1}^L \alpha_l \mathbf{v}_{il} - \beta \mathbf{v}_{i0}) = 0. \end{aligned}$$

Therefore, parameter vector $\hat{\boldsymbol{\theta}} = (\hat{\alpha}_1, \dots, \hat{\alpha}_L, \hat{\beta})^T$ is solving by

$$\sum_{i=L+1}^n \begin{bmatrix} \|\mathbf{v}_{i1}\|^2 & \mathbf{v}_{i1}^T \mathbf{v}_{i2} & \cdots & \mathbf{v}_{i1}^T \mathbf{v}_{iL} & \mathbf{v}_{i1}^T \mathbf{v}_{i0} \\ \mathbf{v}_{i2}^T \mathbf{v}_{i1} & \|\mathbf{v}_{i2}\|^2 & \cdots & \mathbf{v}_{i2}^T \mathbf{v}_{iL} & \mathbf{v}_{i2}^T \mathbf{v}_{i0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{v}_{iL}^T \mathbf{v}_{i1} & \mathbf{v}_{iL}^T \mathbf{v}_{i2} & \cdots & \|\mathbf{v}_{iL}\|^2 & \mathbf{v}_{iL}^T \mathbf{v}_{i0} \\ \mathbf{v}_{i0}^T \mathbf{v}_{i1} & \mathbf{v}_{i0}^T \mathbf{v}_{i2} & \cdots & \mathbf{v}_{i0}^T \mathbf{v}_{iL} & \|\mathbf{v}_{i0}\|^2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_L \\ \beta \end{bmatrix} = \sum_{i=L+1}^n \begin{bmatrix} \mathbf{v}_{i1}^T \mathbf{v}_i \\ \mathbf{v}_{i2}^T \mathbf{v}_i \\ \vdots \\ \mathbf{v}_{iL}^T \mathbf{v}_i \\ \mathbf{v}_{i0}^T \mathbf{v}_i \end{bmatrix},$$

such that $\hat{\boldsymbol{\theta}} = (Z^T Z)^{-1} Z^T Y$, where $Z = \sum_{i=L+1}^n \begin{bmatrix} \|\mathbf{v}_{i1}\|^2 & \cdots & \mathbf{v}_{i1}^T \mathbf{v}_{i0} \\ \vdots & \ddots & \vdots \\ \mathbf{v}_{i0}^T \mathbf{v}_{i1} & \cdots & \|\mathbf{v}_{i0}\|^2 \end{bmatrix}$ and $Y = \sum_{i=L+1}^n \begin{bmatrix} \mathbf{v}_{i1}^T \mathbf{v}_i \\ \vdots \\ \mathbf{v}_{i0}^T \mathbf{v}_i \end{bmatrix}$ for $i = L+1, \dots, n$.

Additionally, $\hat{\sigma}$ is solved by computing the first derivative of log-likelihood function ℓ with respect to σ and letting it be 0, so that

$$\begin{aligned} \frac{\partial \ell}{\partial \sigma} &= -\frac{m(n-L)}{\sigma} + \frac{1}{\sigma^3} \sum_{i=L+1}^n \|\mathbf{v}_i - \hat{\mathbf{v}}_i\|^2 \\ \hat{\sigma} &= \sqrt{\frac{1}{m(n-L)} \sum_{i=L+1}^n \|\mathbf{v}_i - \hat{\mathbf{v}}_i\|^2}, \end{aligned}$$

where $\hat{\mathbf{v}}_i = \sum_{l=1}^L \hat{\alpha}_l \mathbf{v}_{il} - \hat{\beta} \mathbf{v}_{i0}$ is the estimated values.

MLE for the model with diagonal coefficients

When $A_l = \text{Diag}(a_{l1}, \dots, a_{lm})$, $B = \text{Diag}(b_1, \dots, b_m)$, and $\Sigma = \text{Diag}(\sigma_1^2, \dots, \sigma_m^2)$ are diagonal matrices, we can represent model (4.3) as

$$\begin{aligned} \mathbf{v}_i &= \sum_{l=1}^L \begin{bmatrix} v_{il,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & v_{il,m} \end{bmatrix} \begin{bmatrix} a_{l1} \\ \vdots \\ a_{lm} \end{bmatrix} + \begin{bmatrix} v_{i0,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & v_{i0,m} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{im} \end{bmatrix} \\ &= \sum_{l=1}^L \mathbf{u}_{il} \mathbf{a}_l + \mathbf{u}_{i0} \mathbf{b} + \boldsymbol{\epsilon}_i, \end{aligned}$$

where $\mathbf{u}$ are diagonal matrices of $\mathbf{v}$, i.e. $\mathbf{u} = \text{Diag}(v_1, \dots, v_m)$; $\mathbf{a}_l = (a_{l1}, \dots, a_{lm})^T$, $\mathbf{b} = (b_1, \dots, b_m)^T$ are vectors from diagonal matrix $A_l, B, l = 1, \dots, L$. Furthermore, $\boldsymbol{\epsilon}_i \sim N(0, \Sigma)$ and $\Sigma = \text{Diag}(\sigma_1^2, \dots, \sigma_m^2)$.

The log-likelihood function is then written as

$$\begin{aligned} \ell &= -\frac{m(n-L)}{2} \log(2\pi) - \frac{n-L}{2} \log(\sigma_1^2 \cdots \sigma_m^2) \\ &\quad - \frac{1}{2} \sum_{i=L+1}^n \left(\mathbf{v}_i - \sum_{l=1}^L \mathbf{u}_{il} \mathbf{a}_l - \mathbf{u}_{i0} \mathbf{b} \right)^T \Sigma^{-1} \left(\mathbf{v}_i - \sum_{l=1}^L \mathbf{u}_{il} \mathbf{a}_l - \mathbf{u}_{i0} \mathbf{b} \right). \end{aligned} \quad (4.6)$$

According to the process of MLE, the first derivative of log-likelihood with respect to coefficients A_l and B are calculated and set to 0. Then, the estimated $\hat{\Theta} = (\hat{\mathbf{a}}_1, \dots, \hat{\mathbf{a}}_L, \hat{\mathbf{b}})^T$

is solved by

$$\hat{\Theta} = (Z^T Z)^{-1} Z^T Y,$$

where $Z = \sum_{i=L+1}^n \begin{bmatrix} \mathbf{u}_{i1}\mathbf{u}_{i1} & \cdots & \mathbf{u}_{i1}\mathbf{u}_{iL} & \mathbf{u}_{i1}\mathbf{u}_{i0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{u}_{iL}\mathbf{u}_{i1} & \cdots & \mathbf{u}_{iL}\mathbf{u}_{iL} & \mathbf{u}_{iL}\mathbf{u}_{i0} \\ \mathbf{u}_{i0}\mathbf{u}_{i1} & \cdots & \mathbf{u}_{i0}\mathbf{u}_{iL} & \mathbf{u}_{i0}\mathbf{u}_{i0} \end{bmatrix}$ and $Y = \sum_{i=L+1}^n \begin{bmatrix} \mathbf{u}_{i1}\mathbf{v}_i \\ \vdots \\ \mathbf{u}_{iL}\mathbf{v}_i \\ \mathbf{u}_{i0}\mathbf{v}_i \end{bmatrix}$.

Similarity, $\hat{\sigma}_r$ is solved by its the first derivative of log-likelihood ℓ as

$$\begin{aligned} \frac{\partial \ell}{\partial \sigma_r} &= -\frac{n-L}{\sigma_r} + \frac{1}{\sigma_r^3} \sum_{i=L+1}^n (\mathbf{v}_i - \hat{\mathbf{v}}_i)_r^2 = 0 \\ \hat{\sigma}_r &= \sqrt{\frac{\sum_{i=L+1}^n (\mathbf{v}_i - \hat{\mathbf{v}}_i)_r^2}{n-L}}, \end{aligned}$$

where $(\mathbf{v}_i - \hat{\mathbf{v}}_i)_r$ is the r th component of residual vector $\sum_{i=L+1}^n (\mathbf{v}_i - \hat{\mathbf{v}}_i)^2$, $r = 1, \dots, m$. It follows the estimated $\hat{\Sigma}$ as $\hat{\Sigma} = \text{Diag}(\hat{\sigma}_1, \dots, \hat{\sigma}_m)$.

4.3.2 Asymptotic covariance matrix of estimates

The asymptotic covariance matrix of maximum likelihood estimates of $\Phi = (\alpha_1, \dots, \alpha_L, \beta, \sigma)^T$ in the scalar coefficient model (or $\Phi = (\mathbf{a}_1, \dots, \mathbf{a}_L, \mathbf{b}, \sigma)^T$ in the diagonal coefficient model) is used to construct confidence intervals and provides a measurement of the uncertainty associated with the estimated parameters.

Fisher information matrix of estimates in the scalar coefficient model

In model (4.3) with scalar coefficients, we firstly compute the *Hessian matrix*, which is the second derivative of log-likelihood function defined as

$$H(\ell) = \begin{bmatrix} \frac{\partial^2 \ell}{\partial \alpha_1^2} & \cdots & \frac{\partial^2 \ell}{\partial \alpha_1 \partial \alpha_L} & \frac{\partial^2 \ell}{\partial \alpha_1 \partial \beta} & \frac{\partial^2 \ell}{\partial \alpha_1 \partial \sigma} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\partial^2 \ell}{\partial \alpha_1 \partial \alpha_L} & \cdots & \frac{\partial^2 \ell}{\partial \alpha_L^2} & \frac{\partial^2 \ell}{\partial \alpha_L \partial \beta} & \frac{\partial^2 \ell}{\partial \alpha_L \partial \sigma} \\ \frac{\partial^2 \ell}{\partial \alpha_1 \partial \beta} & \cdots & \frac{\partial^2 \ell}{\partial \alpha_L \partial \beta} & \frac{\partial^2 \ell}{\partial \beta^2} & \frac{\partial^2 \ell}{\partial \beta \partial \sigma} \\ \frac{\partial^2 \ell}{\partial \alpha_1 \partial \sigma} & \cdots & \frac{\partial^2 \ell}{\partial \alpha_L \partial \sigma} & \frac{\partial^2 \ell}{\partial \beta \partial \sigma} & \frac{\partial^2 \ell}{\partial \sigma^2} \end{bmatrix}$$

where the second cross-partial derivative of ℓ is

$$\frac{\partial^2 \ell}{\partial \alpha_l \partial \alpha_o} = -\frac{1}{\sigma^2} \sum_{i=L+1}^n \mathbf{v}_{il}^T \mathbf{v}_{io}$$

$$\begin{aligned}
 \frac{\partial^2 \ell}{\partial \alpha_l \partial \beta} &= -\frac{1}{\sigma^2} \sum_{i=L+1}^n \mathbf{v}_{il}^T \mathbf{v}_{i0} \\
 \frac{\partial^2 \ell}{\partial \sigma^2} &= \frac{m(n-L)}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=L+1}^n \left\| \mathbf{v}_i - \sum_{l=1}^L \hat{\alpha}_l \mathbf{v}_{il} - \hat{\beta} \mathbf{v}_{i0} \right\|^2 \\
 \frac{\partial^2 \ell}{\partial \alpha_l \partial \sigma} &= \frac{\partial}{\partial \alpha_l} \left(-\frac{m(n-L)}{\sigma} + \frac{1}{\sigma^3} \sum_{i=L+1}^n \left\| \mathbf{v}_i - \sum_{l=1}^L \hat{\alpha}_l \mathbf{v}_{il} - \hat{\beta} \mathbf{v}_{i0} \right\|^2 \right) \\
 &= \frac{\partial}{\partial \sigma} \left(\frac{1}{\sigma^2} \sum_{i=L+1}^n \mathbf{v}_{il}^T \left(\mathbf{v}_i - \sum_{l=1}^L \hat{\alpha}_l \mathbf{v}_{il} - \hat{\beta} \mathbf{v}_{i0} \right) \right) \\
 &= -\frac{2}{\sigma^3} \sum_{i=L+1}^n \mathbf{v}_{il}^T \left(\mathbf{v}_i - \sum_{l=1}^L \hat{\alpha}_l \mathbf{v}_{il} - \hat{\beta} \mathbf{v}_{i0} \right)
 \end{aligned}$$

where $l, o = 1, \dots, L$.

Recall the model $\mathbf{v}_i = \sum_{l=1}^L \alpha_l \mathbf{v}_{il} + \beta \mathbf{v}_{i0} + \boldsymbol{\epsilon}_i$. We have

$$\begin{aligned}
 \mathbb{E} \left[\mathbf{v}_i - \sum_{l=1}^L \alpha_l \mathbf{v}_{il} - \beta \mathbf{v}_{i0} \right] &= 0 \\
 \mathbb{E} \left[\left\| \mathbf{v}_i - \sum_{l=1}^L \alpha_l \mathbf{v}_{il} - \beta \mathbf{v}_{i0} \right\|^2 \right] &= m\sigma^2.
 \end{aligned}$$

Consequently, the negative expectation of the second partial derivative of ℓ are

$$\begin{aligned}
 -\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \alpha_l \partial \alpha_o} \right] &= \frac{1}{\sigma^2} \sum_{i=L+1}^n \mathbf{v}_{il}^T \mathbf{v}_{io} \\
 -\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \alpha_l \partial \beta} \right] &= \frac{1}{\sigma^2} \sum_{i=L+1}^n \mathbf{v}_{il}^T \mathbf{v}_{i0} \\
 -\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \sigma^2} \right] &= -\frac{m(n-L)}{\sigma^2} + \frac{3}{\sigma^4} \sum_{i=L+1}^n \mathbb{E} \left[\left\| \mathbf{v}_i - \sum_{l=1}^L \alpha_l \mathbf{v}_{il} - \beta \mathbf{v}_{i0} \right\|^2 \right] \\
 &= -\frac{m(n-L)}{\sigma^2} + \frac{3}{\sigma^4} \sum_{i=L+1}^n m\sigma^2 \\
 &= -\frac{m(n-L)}{\sigma^2} + \frac{3m(n-L)\sigma^2}{\sigma^4} \\
 &= \frac{2m(n-L)}{\sigma^2} \\
 -\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \alpha_l \partial \sigma} \right] &= -\frac{2}{\sigma^3} \sum_{i=L+1}^n \left(\mathbb{E}[\mathbf{v}_{il}^T] \mathbb{E} \left[\mathbf{v}_i - \sum_{l=1}^L \alpha_l \mathbf{v}_{il} - \beta \mathbf{v}_{i0} \right] \right) = 0
 \end{aligned}$$

Therefore, the *Fisher information matrix* is

$$\begin{aligned} I(\Phi) &= -\mathbb{E}[H(\ell)] \\ &= \frac{1}{\sigma^2} \sum_{i=L+1}^n \begin{bmatrix} \mathbf{v}_{i1}^T \mathbf{v}_{i1} & \cdots & \mathbf{v}_{i1}^T \mathbf{v}_{iL} & \mathbf{v}_{i1}^T \mathbf{v}_{i0} & 0 \\ \vdots & \ddots & \vdots & \vdots & 0 \\ \mathbf{v}_{i0}^T \mathbf{v}_{i1} & \cdots & \mathbf{v}_{i0}^T \mathbf{v}_{iL} & \mathbf{v}_{i0}^T \mathbf{v}_{i0} & 0 \\ 0 & \cdots & 0 & 0 & 2m \end{bmatrix} \end{aligned}$$

Fisher information matrix of estimates in the diagonal model

Analogously, we can obtain the Fisher information matrix of maximum likelihood estimates: $\Phi = (a_{11}, \dots, a_{1m}, a_{21}, \dots, a_{Lm}, b_1, \dots, b_m, \sigma_1, \dots, \sigma_m)^T$ from the Hessian matrix. Recalling the log-likelihood function of the manifold-adapted model with diagonal matrix coefficients and rewriting them as vectors, the log-likelihood function in (4.6) could be written as

$$\begin{aligned} \ell &= -\frac{m(n-L)}{2} \log(2\pi) - \frac{n-L}{2} \log(\sigma_1^2 \cdots \sigma_m^2) \\ &\quad - \frac{1}{2} \sum_{i=L+1}^n \left(\sum_{r=1}^m \frac{1}{\sigma_r^2} (v_{ir} - \sum_{l=1}^L a_{lr} v_{il,r} - v_{i0,r} b_r)^2 \right) \end{aligned} \quad (4.7)$$

where $r = 1, \dots, m; l = 1, \dots, L; i = L+1, \dots, n$. Compute the first derivatives of log-likelihood function with respect to all parameters as

$$\begin{aligned} \frac{\partial \ell}{\partial a_{lr}} &= \frac{1}{\sigma_r^2} \sum_{i=L+1}^n v_{il,r} (v_{i,r} - \sum_{l=1}^L a_{lr} v_{il,r} - b_r v_{i0,r}) \\ \frac{\partial \ell}{\partial b_r} &= \frac{1}{\sigma_r^2} \sum_{i=L+1}^n v_{i0,r} (v_{i,r} - \sum_{l=1}^L a_{lr} v_{il,r} - b_r v_{i0,r}) \\ \frac{\partial \ell}{\partial \sigma_r} &= -\frac{n-L}{\sigma_r} + \frac{1}{\sigma_r^3} \sum_{i=L+1}^n \left(v_{i,r} - \sum_{l=1}^L a_{lr} v_{il,r} - b_r v_{i0,r} \right)^2, \end{aligned}$$

and their second derivatives are

$$\begin{aligned} \frac{\partial^2 \ell}{\partial a_{lr} \partial a_{ko}} &= -\frac{1}{\sigma_r^2} \sum_{i=L+1}^n v_{il,r} v_{ik,o} \\ \frac{\partial^2 \ell}{\partial a_{ko} \partial a_{lr}} &= -\frac{1}{\sigma_o^2} \sum_{i=L+1}^n v_{il,r} v_{ik,o} \\ \frac{\partial^2 \ell}{\partial a_{lr} \partial b_r} &= -\frac{1}{\sigma_r^2} \sum_{i=L+1}^n v_{il,r} v_{i0,r} \end{aligned}$$

$$\begin{aligned}
 \frac{\partial^2 \ell}{\partial a_{lr} \partial \sigma_r} &= -\frac{2}{\sigma_r^3} \sum_{i=L+1}^n v_{il,r} (v_{i,r} - \sum_{l=1}^L a_{lr} v_{il,r} - b_r v_{i0,r}) \\
 \frac{\partial^2 \ell}{\partial b_r \partial a_{lr}} &= -\frac{1}{\sigma_r^2} \sum_{i=L+1}^n v_{i0,r} v_{il,r} \\
 \frac{\partial^2 \ell}{\partial b_r^2} &= -\frac{1}{\sigma_r^2} \sum_{i=L+1}^n v_{i0,r}^2 \\
 \frac{\partial^2 \ell}{\partial b_o \partial a_{lr}} &= -\frac{1}{\sigma_o^2} \sum_{i=L+1}^n v_{i0,o} v_{il,r} \\
 \frac{\partial^2 \ell}{\partial b_r \partial \sigma_r} &= -\frac{2}{\sigma_r^3} \sum_{i=L+1}^n v_{i0,r} (v_{i,r} - \sum_{l=1}^L a_{lr} v_{il,r} - b_r v_{i0,r}) \\
 \frac{\partial^2 \ell}{\partial \sigma_r \partial a_{lr}} &= -\frac{2}{\sigma_r^3} \sum_{i=L+1}^n v_{il,r} (v_{i,r} - \sum_{l=1}^L a_{lr} v_{il,r} - b_r v_{i0,r}) \\
 \frac{\partial^2 \ell}{\partial \sigma_r \partial b_r} &= -\frac{2}{\sigma_r^3} \sum_{i=L+1}^n v_{i0,r} (v_{i,r} - \sum_{l=1}^L a_{lr} v_{il,r} - b_r v_{i0,r}) \\
 \frac{\partial^2 \ell}{\partial \sigma_r^2} &= \frac{n-L}{\sigma_r^2} - \frac{3}{\sigma_r^4} \sum_{i=L+1}^n \left(v_{i,r} - \sum_{l=1}^L a_{lr} v_{il,r} - b_r v_{i0,r} \right)^2 \\
 \frac{\partial^2 \ell}{\partial \sigma_o \partial a_{lr}} &= \frac{\partial \ell^2}{\partial \sigma_o \partial b_r} = \frac{\partial \ell^2}{\partial \sigma_o \partial \sigma_r} = 0,
 \end{aligned}$$

where $l, k = 1, \dots, L; r, o = 1, \dots, m$.

Therefore, we can obtain a large Hessian matrix of dimension $((L+2)m \times (L+2)m)$ as

$$H(\ell) = \begin{bmatrix}
 \frac{\partial^2 \ell}{\partial a_{11}^2} & \cdots & \frac{\partial^2 \ell}{\partial a_{11} \partial a_{1m}} & \cdots & \frac{\partial^2 \ell}{\partial a_{11} \partial a_{Lm}} & \frac{\partial^2 \ell}{\partial a_{11} \partial b_1} & \cdots & \frac{\partial^2 \ell}{\partial a_{11} \partial \sigma_1} & \cdots & \frac{\partial^2 \ell}{\partial a_{11} \partial \sigma_m} \\
 & \ddots & & \ddots & & & \ddots & & \ddots & \\
 \frac{\partial^2 \ell}{\partial a_{1m} \partial a_{11}} & \cdots & \frac{\partial^2 \ell}{\partial a_{1m}^2} & \cdots & \frac{\partial^2 \ell}{\partial a_{1m} \partial a_{Lm}} & \frac{\partial^2 \ell}{\partial a_{1m} \partial b_1} & \cdots & \frac{\partial^2 \ell}{\partial a_{1m} \partial \sigma_1} & \cdots & \frac{\partial^2 \ell}{\partial a_{1m} \partial \sigma_m} \\
 & \ddots & & \ddots & & & \ddots & & \ddots & \\
 \frac{\partial^2 \ell}{\partial a_{Lm} \partial a_{11}} & \cdots & \frac{\partial^2 \ell}{\partial a_{Lm}^2} & \cdots & \frac{\partial^2 \ell}{\partial a_{Lm} \partial a_{Lm}} & \frac{\partial^2 \ell}{\partial a_{Lm} \partial b_1} & \cdots & \frac{\partial^2 \ell}{\partial a_{Lm} \partial \sigma_1} & \cdots & \frac{\partial^2 \ell}{\partial a_{Lm} \partial \sigma_m} \\
 \frac{\partial^2 \ell}{\partial b_1 \partial a_{11}} & \cdots & \frac{\partial^2 \ell}{\partial b_1 \partial a_{Lm}} & \cdots & \frac{\partial^2 \ell}{\partial b_1 \partial a_{Lm}} & \frac{\partial^2 \ell}{\partial b_1^2} & \cdots & \frac{\partial^2 \ell}{\partial b_1 \partial \sigma_1} & \cdots & \frac{\partial^2 \ell}{\partial b_1 \partial \sigma_m} \\
 & \ddots & & \ddots & & & \ddots & & \ddots & \\
 \frac{\partial^2 \ell}{\partial \sigma_1 \partial a_{11}} & \cdots & \frac{\partial^2 \ell}{\partial \sigma_1 \partial a_{1m}} & \cdots & \frac{\partial^2 \ell}{\partial \sigma_1 \partial a_{Lm}} & \frac{\partial^2 \ell}{\partial \sigma_1 \partial b_1} & \cdots & \frac{\partial^2 \ell}{\partial \sigma_1^2} & \cdots & \frac{\partial^2 \ell}{\partial \sigma_1 \partial \sigma_m} \\
 & \ddots & & \ddots & & & \ddots & & \ddots & \\
 \frac{\partial^2 \ell}{\partial \sigma_m \partial a_{11}} & \cdots & \frac{\partial^2 \ell}{\partial \sigma_m \partial a_{1m}} & \cdots & \frac{\partial^2 \ell}{\partial \sigma_m \partial a_{Lm}} & \frac{\partial^2 \ell}{\partial \sigma_m \partial b_1} & \cdots & \frac{\partial^2 \ell}{\partial \sigma_m \partial \sigma_1} & \cdots & \frac{\partial^2 \ell}{\partial \sigma_m^2}
 \end{bmatrix}$$

Now, we compute the expectation of each derivative in the above Fisher information

matrix which is defined as the negative expectation of Hessian matrix $H(\ell)$, i.e., $I(\Phi) = -\mathbb{E}[H(\ell)]$. According to the model, it is known that

$$\begin{aligned}\mathbb{E}\left[v_r - \sum_l a_{lr}v_{l,r} - b_r v_{0,r}\right] &= 0 \\ \mathbb{E}\left[\left(v_r - \sum_l a_{lr}v_{l,r} - b_r v_{0,r}\right)^2\right] &= \sigma_r^2.\end{aligned}$$

It follows that

$$\begin{aligned}-\mathbb{E}\left[\frac{\partial^2 \ell}{\partial a_{lr} \partial \sigma_r}\right] &= -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial b_r \partial \sigma_r}\right] = \mathbb{E}\left[\frac{\partial^2 \ell}{\partial \sigma_r \partial a_{lr}}\right] = \mathbb{E}\left[\frac{\partial^2 \ell}{\partial \sigma_r \partial b_r}\right] = 0 \\ -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \sigma_r^2}\right] &= -\frac{n-L}{\sigma_r^2} + \frac{3(n-L)}{\sigma_r^2} = \frac{2(n-L)}{\sigma_r^2}.\end{aligned}$$

The Fisher information matrix is then

$$I(\Phi) = \sum_{i=L+1}^n \begin{bmatrix} \frac{v_{i1,1}^2}{\sigma_1^2} & \cdots & \frac{v_{i1,1}v_{i1,m}}{\sigma_1^2} & \cdots & \frac{v_{i1,1}v_{iL,m}}{\sigma_1^2} & \frac{v_{i1,1}v_{i0,1}}{\sigma_1^2} & \cdots & 0 & \cdots & 0 \\ & \ddots & & \ddots & & & \ddots & & \ddots & \\ \frac{v_{i1,m}v_{i1,1}}{\sigma_m^2} & \cdots & \frac{v_{i1,m}^2}{\sigma_m^2} & \cdots & \frac{v_{i1,m}v_{iL,m}}{\sigma_m^2} & \frac{v_{i1,m}v_{i0,1}}{\sigma_m^2} & \cdots & 0 & \cdots & 0 \\ & \ddots & & \ddots & & & \ddots & & \ddots & \\ \frac{v_{iL,m}v_{i1,1}}{\sigma_m^2} & \cdots & \frac{v_{iL,m}v_{i1,m}}{\sigma_m^2} & \cdots & \frac{v_{iL,m}^2}{\sigma_m^2} & \frac{v_{iL,m}v_{i0,1}}{\sigma_m^2} & \cdots & 0 & \cdots & 0 \\ \frac{v_{i0,1}v_{i1,1}}{\sigma_1^2} & \cdots & \frac{v_{i0,1}v_{i1,m}}{\sigma_1^2} & \cdots & \frac{v_{i0,1}v_{iL,m}}{\sigma_1^2} & \frac{v_{i0,1}^2}{\sigma_1^2} & \cdots & 0 & \cdots & 0 \\ & \ddots & & \ddots & & & \ddots & & \ddots & \\ 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & \frac{2}{\sigma_1^2} & \cdots & 0 \\ & \ddots & & \ddots & & & \ddots & & \ddots & \\ 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \frac{2}{\sigma_m^2} \end{bmatrix}.$$

Asymptotic covariance matrices of maximum likelihood estimates

Given $\Phi = (\alpha_1, \dots, \alpha_L, \beta, \sigma)^T$ in the scalar coefficient model or $\Phi = (\mathbf{a}_1, \dots, \mathbf{a}_L, \mathbf{b}, \sigma)^T$ in the diagonal model, the maximum likelihood estimator is asymptotically normal for sufficiently large n : $\sqrt{n}(\hat{\Phi} - \Phi)$ converges to a multivariate normal distribution with zero mean and covariance matrix $I(\Phi)^{-1}$ (Newey & McFadden, 1994, Theorem 3.3), i.e.,

$$\sqrt{n}(\hat{\Phi} - \Phi) \rightarrow N(0, I(\Phi)^{-1}).$$

More precisely, for ‘large’ n , the distribution of the vector $\hat{\Phi}$ can be approximated by a multivariate normal distribution with mean Φ and covariance matrix

$$\frac{1}{n}I(\Phi)^{-1}.$$

Hence, we can establish 95% confidence intervals for the estimated parameters of Φ :

$$\hat{\Phi}_j \pm 1.96I(\Phi)^{-1/2}_{jj}$$

where $j = 1, \dots, J$ and J is the total number of estimated parameters, e.g., $J = L + 2$ in the scalar coefficient model and $J = (L + 2)m$ in the diagonal matrix model.

4.3.3 Model selection

There are many methods for model selection with respect to time series models. We review two typical approaches in this research: the Akaike Information Criterion (AIC) (Burnham & Anderson, 2004) and the partial autocorrelation function (PACF) (Box *et al.*, 2015).

Akaike Information Criterion

In the process of assessing which model best fits the data, it is crucial to acknowledge that the objective function, the log-likelihood ℓ , always improves when extra parameters are introduced. As a result, the strategy is to counterbalance increasing complexity by implementing a penalty, as represented by the AIC value, given by:

$$\text{AIC}(L) = -2\ell + 2K,$$

where ℓ is the log-likelihood at the MLE and K denotes the total number of parameters. The penalty grows with the increase in K . Finally, the optimal model is selected by minimizing $\text{AIC}(L)$.

Partial autocorrelation function

The AIC takes into account the goodness of fit of a model while penalizing for the number of parameters used. A high lag suggested by AIC might indicate that the model is overfitting the data by including unnecessary parameters. Additionally, a high lag could be capturing random fluctuations or noise in the data, which may not have any meaningful relationship with the underlying process. To overcome this issue, we need to measure the direct relationship between the current observation and its lagged values while controlling for the effects of the intermediate lags. That is to say, the partial autocorrelation at lag

l is the autocorrelation between $\mathbf{v}_i$ and $\mathbf{v}_{i-l}$ that is not accounted for by lags 1 to $l - 1$ ($l = 1, \dots, L$).

In practice, there are algorithms for computing the partial autocorrelation based on the sample autocorrelations (Box *et al.*, 2015; Brockwell & Davis, 1991). Without delving into the mathematical details of the partial autocorrelation function, there is a simple method for obtaining partial autocorrelation yielding the same results (Hyndman & Athanasopoulos, 2018). Specifically, the first partial autocorrelation is identical to the first autocorrelation, and each subsequent partial autocorrelation can be estimated as the last coefficient in an autoregressive model. For example, the PACF value at lag L can be simply considered as α_L from the autoregressive model $\text{AR}(L)$. Note that the PACF value of $L - 1$ should be estimated by the model $\text{AR}(L - 1)$ (the last regression coefficient in the model $\text{AR}(L - 1)$) rather than α_{L-1} in $\text{AR}(L)$.

Subsequently, we will check for significant PACF values at particular lags, suggesting that those lags are important in explaining the current value of the series. Typically, PACF values beyond a given lag, denoted as L , exhibit independence and behave as approximately independent random variables following a $N(0, 1/n)$ distribution. This implies that around 95% of sample PACF values beyond lag L are expected to lie within the range of $\pm 1.96/\sqrt{n}$ (Brockwell & Davis, 2002).

4.4 Model diagnostics

Considering the intricacies inherent in our model, defined on Riemannian manifolds, traditional stationary tests designed for Euclidean spaces are deemed inappropriate. This is due to the dynamic nature of the endogenous variables, which undergo continuous changes owing to the parallel translation between tangent spaces on the manifold. Departing from conventional practices in time series analysis, we choose not to assess stationarity directly within the manifold-valued time series. Instead, this thesis pivots towards scrutinizing the stationarity of model residuals, which assume a pivotal role as an indicator of the model's effectiveness in capturing the latent structures embedded in the data. In evaluating our model, we place emphasis on the desired properties of residuals generated by a robust regression model. Ideally, these residuals should exhibit characteristics such as independence, homoscedasticity, normality, and stationarity, representing essential qualities that validate the adequacy of our manifold-adapted time series model.

Therefore, this section will analyse the residuals $\epsilon = \mathbf{v} - \hat{\mathbf{v}}$ by checking the assumptions mentioned above, and additionally, we will assess the goodness of modelling fit.

4.4.1 Residual analysis

Due to the independence among components of ϵ , tests for m -dimensional residuals should be assessed independently through:

- (1) **Normality:** Each component should independently and identically follow a normal distribution, specifically $N(0, \sigma^2 I_m)$ or $N(0, \Sigma_{rr})$ where $\Sigma = \text{Diag}(\sigma_1^2, \dots, \sigma_m^2)$ and $r = 1, \dots, m$. To satisfy this condition, it is imperative to subject each dimension to a normality test, ensuring that residuals exhibit both a mean of zero and constant variance. This requirement is particularly crucial when employing the maximum likelihood estimation in our model.

Kolmogorov–Smirnov test is a nonparametric test with the null hypothesis that the sample is drawn from the reference distribution $N(0, \hat{\sigma}^2)$ or $N(0, \hat{\Sigma})$ (Berger & Zhou, 2014).

- (2) **Stationarity:** In the context of time series analysis, the stationarity of residuals is a critical consideration to prevent spurious regressions. This stationarity implies that both the mean and variance of residuals remain constant over time. When the individual components of residuals exhibit independent stationarity, it indicates that the model has effectively captured all the temporal characteristics of the time series data, and consequently, no systematic patterns can be discerned in the residuals over time.

Augmented Dickey-Fuller (ADF) test: The ADF test is a commonly used test for stationarity (Mushtaq, 2011). It tests the null hypothesis that a unit root is present in time series, which would indicate non-stationarity.

- (3) **Independence:** The errors should not exhibit temporal correlation, indicating that they should resemble white noise, devoid of any discernible patterns over time.

Ljung-Box test is a statistical test that tests whether the residuals of the time series model are independently distributed (Box & Pierce, 1970). It is a common method to check for the presence of autocorrelation in time series residuals, particularly at different lags.

4.4.2 Goodness of fit

Like a general regression model, the R-squared (R^2) is used to assess the goodness of modelling fit. We directly compute the R^2 values for our manifold-adapted models as follows:

$$R^2 = 1 - \frac{\text{Unexplained variation}}{\text{Total variation}} = 1 - \frac{\sum_{i=1}^n \|\mathbf{v}_i - \hat{\mathbf{v}}_i\|^2}{\sum_{i=1}^n \|\mathbf{v}_i - \bar{\mathbf{v}}\|^2}. \quad (4.8)$$

Here, $\|\cdot\|^2$ represents the Euclidean norm of vectors. For example, $\|\mathbf{v}_i\|^2 = \sum_r v_{ir}^2$. $\bar{\mathbf{v}}$ is the mean vector of $\mathbf{v}$, i.e. $\bar{\mathbf{v}} = 1/n \sum_{i=1}^n \mathbf{v}_i$.

4.5 Simulation studies

In this section, simulation studies will be conducted to test our models and model inference. The underlying approach involves generating initial points from a wrapped Gaussian distribution. Subsequently, with fixed lags and initial parameters $\Phi = (A_1, \dots, A_L, B, \Sigma)^T$, the matrices S_{i+1} under the general manifold-valued model (4.2) are generated. After discarding a burn-in period, the simulated manifold-valued data are utilized to perform MLE and model selection using AIC, obtaining estimated parameters. A comparison is then conducted between these estimated parameters and the initial parameters.

4.5.1 Simulation study on the scalar coefficient model

Covariance matrices on the affine invariant geometry $(\mathcal{S}^+(p), g^{\text{aff}})$

We start the simulation study using the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$ and the scalar coefficient model described in (4.2). Specifically, given initial set of points $S_1, \dots, S_{L+1}$ within $\mathcal{S}^+(p)$, we proceed to generate the subsequent points through the following steps:

Step 1 Autoregressive terms:

$$\begin{aligned} V_{i-l} &= \text{Log}_{S_{i-l}}(S_{i-l+1}), \\ V_{il} &= \mathcal{P}_{S_{i-l} \rightarrow S_i}(V_{i-l}), \\ v_{il,r} &= g_{S_i}^{\text{aff}} \langle V_{il}, E_r^{S_i} \rangle, \end{aligned}$$

where $l = 1, \dots, L$ and $\{E_r^{S_i} : r = 1, \dots, m\}$ is the orthonormal basis in the tangent space $T_{S_i}\mathcal{S}^+(p)$ such that $\mathbf{v}_{il} = (v_{il,1}, \dots, v_{il,m})^T \in \mathbb{R}^m$. Refer to Section 2.3 for detailed operations in the affine invariant geometry.

Step 2 Mean reversion term:

$$\begin{aligned} V_{i0} &= \text{Log}_{S_i}(S^*), \\ v_{i0,r} &= g_{S_i}^{\text{aff}} \langle V_{i0}, E_r^{S_i} \rangle, \end{aligned}$$

where S^* is the attract point, and $r = 1, \dots, m$.

Step 3 Wrapped Gaussian distribution:

$$\begin{aligned} \mathbf{v}_i &\sim N\left(\sum_{l=1}^L \alpha_l \mathbf{v}_{il} + \beta \mathbf{v}_{i0}, \sigma^2 I_m\right), \\ V_i &= \sum_{r=1}^m v_{ir} E_r^{S_i}, \\ S_{i+1} &= \text{Exp}_{S_i}(V_i). \end{aligned}$$

Here, we provide an example of a manifold-adapted model (4.2) with scalar coefficients in $\mathcal{S}^+(10)$. We set the maximum lag to $L = 3$ and initialize the parameters $(\alpha_1, \alpha_2, \alpha_3) \sim N(0, 0.4^2)$, $\beta \sim \Gamma(5, 25)$, and $\sigma \sim \Gamma(5, 30)$. We simulate a set of 10×10 covariance matrix-valued time series with a length of $n = 300$ and an attract point S^* , all drawn from a wrapped Gaussian distribution as defined in Equation (2.2) with zero mean and covariance matrix $\Sigma = 0.2I_{10}$ at the identity matrix I_{10} . For the sake of illustrating simulation results and model inference, we choose fixed values for the initial parameters: $(\alpha_1, \alpha_2, \alpha_3)^T = (-0.167, -0.282, 0.413)^T$, $\beta = 0.252$, and $\sigma = 0.212$.

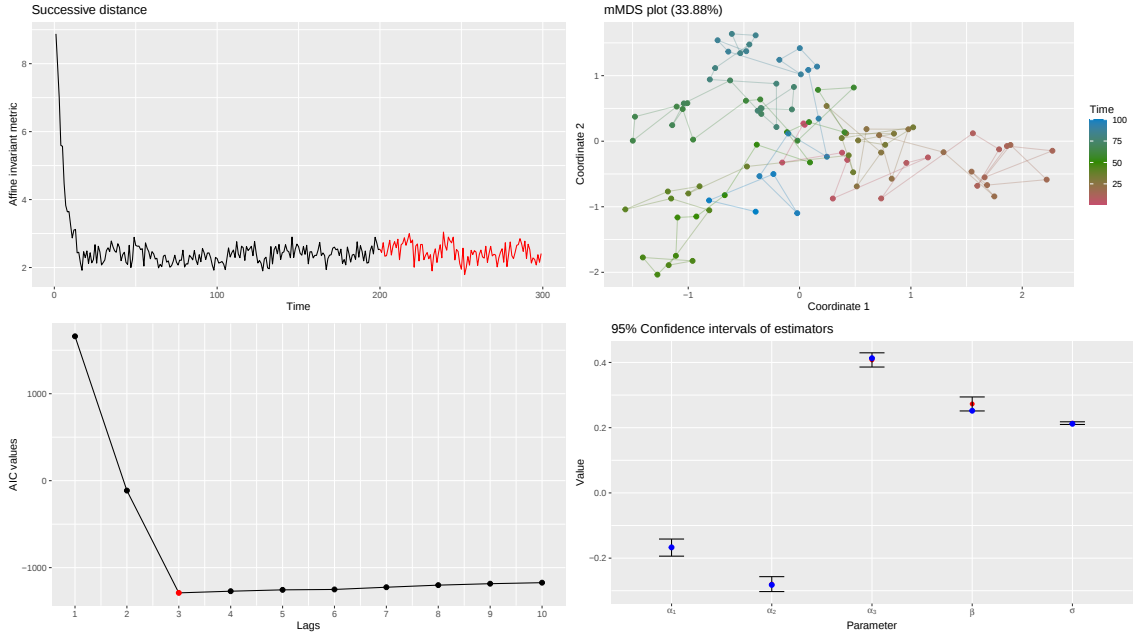


Figure 4.1: Top Left: Successive distances via the geodesic distance $d_{g^{\text{aff}}}$. We omit the burn-in period and fit the model to the last 100 data points, highlighted in red. Top Right: Connected scatter plot. Utilizing the affine invariant metric, the plot exhibits a gradient of colours from red to green, culminating in blue. The percentage value in brackets represents the stress of the metric MDS plot. Bottom Left: AIC values. AIC values for lags 1 to 10 are depicted. The red point is selected by minimizing AIC values. Bottom Right: Estimated parameters and confidence intervals. The plot displays estimated parameters (red points) along with 95% confidence intervals (vertical intervals). True parameters are marked in blue.

Following a burn-in period in the simulated data, we extract the last 100 data points (depicted by the red line) in the successive distance plot shown in Figure 4.1. The attract point is computed using the Fréchet sample mean of the extracted data. Subsequently, our model is applied to fit this simulated data set. The resulting connected scatter plot on the right side of Figure 4.1 employs a gradient colour scheme ranging from red to blue. Notably, through model selection, we infer that a lag value of $L = 3$ represents the optimal choice for the simulated data set, aligning with the initial configuration. Consequently, MLE is employed to estimate the parameters. These estimated parameters are listed in Table 4.1, and the bottom-right panel shows 95% confidence intervals, which include both the estimated parameters (depicted by red dots) and the initially set parameters (depicted by blue dots).

Table 4.1: True and estimated parameters with 95% confidence intervals.

	True parameters	Estimates	95% CI
α_1	-0.167	-0.168	(-0.194 , -0.141)
α_2	-0.282	-0.280	(-0.303, -0.257)
α_3	0.413	0.408	(0.386, 0.430)
β	0.252	0.273	(0.251 , 0.294)
σ	0.212	0.214	(0.210, 0.218)

Moreover, the goodness of the model fit is assessed using R^2 , defined in Equation (4.8), yielding a value of 58.03%. This implies that 58.03% of the variation in the response variables is explained by the autoregressive and mean-reverting terms.

Correlation matrices on the quotient geometry $(\mathcal{C}^+(p), g^{\text{quo}})$

Similarly, we extend our simulation to data generated under the model with scalar coefficients in the quotient geometry $(\mathcal{C}^+(p), g^{\text{quo}})$. The simulation process in $(\mathcal{C}^+(p), g^{\text{quo}})$ is inherently more intricate compared to that in $(\mathcal{S}^+(p), g^{\text{aff}})$. The following steps encapsulate the simulation procedure:

Step 1 Optimal positions within the fibre $\pi^{-1}(C_i)$ with respect to I_p are denoted as S_i , ensuring that both I_p and S_i are optimally positioned within $\mathcal{S}^+(p)$. Following this, we define optimal positions within the fibres $\pi^{-1}(C_{i-1}), \dots, \pi^{-1}(C_{i-L}), \pi^{-1}(C^*)$, represented as $S_{i-1}, \dots, S_{i-L}, S^* \in \mathcal{S}^+(p)$ with respect to I_p .

Step 2 Autoregressive terms:

$$\begin{aligned}
 X_{i-l} &= \text{Log}_{C_{i-l}}(C_{i-l+1}), \\
 X_{il}^\sharp &= (\mathcal{P}_{C_{i-l} \rightarrow C_i}(X_{i-l}))^\sharp, \\
 u_{il,r} &= g_{S_i}^{\text{aff}} \langle X_{il}^\sharp, F_r^{S_i} \rangle,
 \end{aligned}$$

where $l = 1, \dots, L$. Here, $X_{il}^\sharp$ represents the horizontal lift of X_{il} , computed by parallel transporting X_{i-l} from the tangent space $T_{C_{i-l}}\mathcal{C}^+(p)$ to $T_{C_i}\mathcal{C}^+(p)$. We lift the tangent vector $X_{il} \in T_{C_i}\mathcal{C}^+(p)$ to the horizontal subspace $\mathcal{H}_{S_i}$, where $S_i \in \pi^{-1}(C_i)$ and I_p are optimally positioned within $\mathcal{S}^+(p)$. The set $\{F_r^{S_i} : r = 1, \dots, m\}$ forms an orthonormal basis in the horizontal subspace $\mathcal{H}_{S_i}$, such that $\mathbf{u}_{il} = (u_{il,1}, \dots, u_{il,m})^T \in \mathbb{R}^m$. For further details on these operations in quotient geometry, please refer to Section 2.4.

Step 3 Mean reversion term:

$$\begin{aligned} X_{i0} &= \text{Log}_{C_i}(C^*), \\ u_{i0,r} &= g_{S_i}^{\text{aff}} \langle X_{i0}^\sharp, F_r^{S_i} \rangle, \end{aligned}$$

where C^* is the attract point and $r = 1, \dots, m$.

Step 4 Wrapped Gaussian distribution:

$$\begin{aligned} \mathbf{u}_i &\sim N\left(\sum_{l=1}^L \alpha_l \mathbf{u}_{il} + \beta \mathbf{u}_{i0}, \sigma^2 I_m\right), \\ V_i &= \sum_{r=1}^m u_{ir} F_r^{S_i}; X_i = d\pi_{S_i}(V_i), \\ C_{i+1} &= \text{Exp}_{C_i}(X_i). \end{aligned}$$

The simulated results are presented in Figure 4.2. In this case, we generated 300 covariance matrices ($n = 300$) and an attract point S^* from a wrapped Gaussian distribution with zero mean and covariance matrix $\Sigma = 0.2I_{10}$ at the identity matrix I_{10} and then computed their correlation matrices. Obtaining the data set of correlation matrices, we selected the data from the last 100 data points, which is in the post-burn-in period. The goodness of modelling fit with R^2 value was 43.77%. To avoid redundant descriptions of predictable simulation results, we have omitted the display of estimated parameters and 95% confidence intervals in a table. Instead, the bottom-right panel illustrates $\hat{\Phi}$ with red points and confidence intervals depicted by vertical intervals. The true (initial set) parameters are denoted in blue.

It is notable that there are variations in the estimated parameters $\hat{\alpha}_3$, β , and $\hat{\sigma}$ compared to the real setting parameters. This discrepancy arises from the fact that the estimated Fréchet sample mean $\tilde{C}$ computed from the simulated data is not close to the real attractor point C^* , leading to a higher $\hat{\beta}$ and slightly lower $\hat{\alpha}_3$ after the model inference. We also attempted to use the original setting point C^* as the attractor, but yielded the similar results. It also may be due to the incremental variation in optimization for

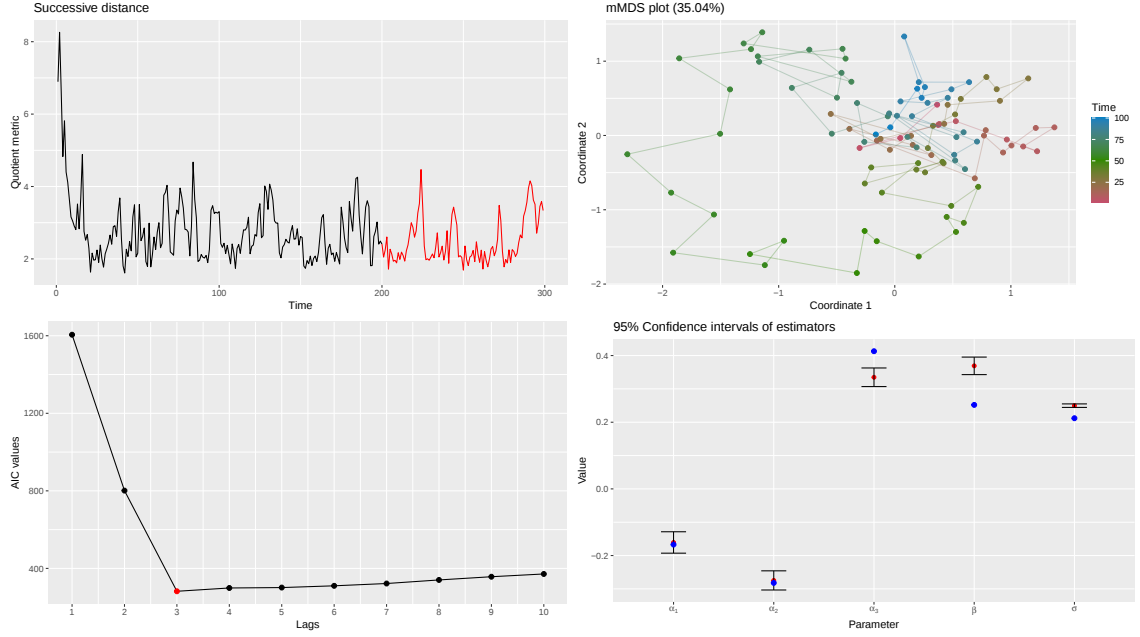


Figure 4.2: Top Left: Successive distance via the geodesic distance $d_{g^{\text{quo}}}$. We omit the burn-in period and fit the model to the last 100 data points, highlighted in red. Top Right: Connected scatter plot. Utilizing the quotient metric, the plot exhibits a gradient of colours from red to green, culminating in blue. The percentage value in brackets represents the stress of the metric MDS plot. Bottom Left: AIC values. AIC values for lags 1 to 10 are depicted. The red point is selected by minimizing AIC values. Bottom Right: Estimated parameters and confidence intervals. The plot displays estimated parameters (red points) along with 95% confidence intervals (vertical intervals). True parameters are marked in blue.

seeking the optimal positions (described in Section 2.4.3) leads to this outcome. Specifically, for every simulated C_i , we consistently find an optimal position with respect to I_p as $S_i \in \pi^{-1}(C_i) \subset \mathcal{S}^+(p)$, ensuring that any tangent vector along the geodesic from I_p to S_i is horizontal. However, in practice, the numerical experiment results of S_i cannot guarantee that the horizontal basis in $\mathcal{H}_{I_p}$ to $\mathcal{H}_{S_i}$ remains horizontal. This results in the Riemannian submersion of the basis in $T_{C_i}\mathcal{C}^+(p)$ being not completely orthonormal when implemented as the computer code. In practice, the inner product of the basis $g_{C_i}^{\text{quo}}\langle F_r^{C_i}, F_r^{C_i} \rangle$ is close but not equal to 1, and $g_{C_i}^{\text{quo}}\langle F_r^{C_i}, F_o^{C_i} \rangle$ is close but not equal to 0, where $r, o = 1, \dots, m$, and $r \neq o$.

In the future research, improving the precision of optimization for seeking an optimal position and proving the uniqueness of the Fréchet sample mean in the quotient geometry may be the key to overcome the above limitations.

4.5.2 Simulation study on the diagonal matrix coefficient model

Covariance matrices on the affine invariant geometry $(\mathcal{S}^+(p), g^{\text{aff}})$

Similar to the simulation study under the framework of the model with the scalar coefficients in $(\mathcal{S}^+(p), g^{\text{aff}})$, the only necessary adjustment is to modify Step 3, incorporating the wrapped Gaussian distribution of $\mathbf{v}_i$ from the previous subsection as follows:

$$\mathbf{v}_i \sim N\left(\sum_{l=1}^L A_l \mathbf{v}_{il} + B \mathbf{v}_{i0}, \Sigma\right),$$

$$V_i = \sum_{r=1}^m v_{ir} E_r^{S_i}.$$

For a more comprehensive comparison with the simulation study on the model with scalar coefficients, we generate the diagonals of A_1, A_2 and A_3 following normal distributions with the means of α_1, α_2 and α_3 and variance of 0.1^2 . B and Γ follows the same distributions with β and σ . Specifically, we establish the following parameter values based on the setting parameters in Section 4.5.1:

$$p = 10, L = 3,$$

$$a_{1r} \sim N(-0.167, 0.1^2), a_{2r} \sim N(-0.282, 0.1^2), a_{3r} \sim N(0.413, 0.1^2),$$

$$b_r \sim \Gamma(5, 25),$$

$$\sigma_r \sim \Gamma(5, 30),$$

where $A_l = \text{Diag}(a_{l1}, \dots, a_{lm})$, $B = \text{Diag}(b_1, \dots, b_m)$, and $\Sigma = \text{Diag}(\sigma_1^2, \dots, \sigma_m^2)$ are diagonal matrices and $l = 1, 2, 3; r = 1, \dots, m$.

In practical implementation, we set the variance of distributions of diagonals in A_l and B being 0.1^2 , as a larger standard deviation was found to lead to simulation instability. Following the implementation of the simulation process with these parameters, we extract simulated data after the burn-in period. Subsequently, model inference and selection are performed to obtain the estimated parameters, as depicted in Figure 4.3.

The goodness of the modeling fit, measured by R^2 value, is 70.47%. To assess parameter distributions, we employ the two-sample Kolmogorov-Smirnov test, comparing each pair of parameters. The null hypothesis is that, for example, $\text{Diag}(A_l)$ and $\text{Diag}(\hat{A}_l)$ are drawn from the same probability distribution (See Table 4.2). Importantly, we accept all hypotheses indicating that our manifold-adapted model with diagonal coefficients, along with the model inference, effectively captures data dynamics and performs exceptionally well.

We can also examine the confidence intervals for each element a_{lr}, b_r, σ_r ($l = 1, 2, 3; r =$

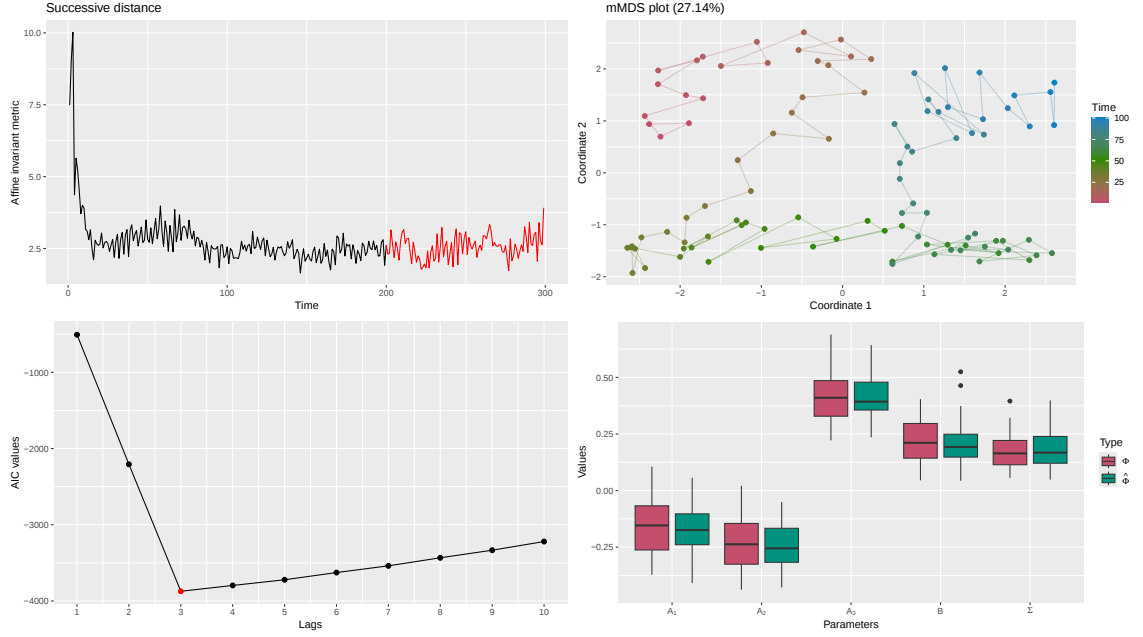


Figure 4.3: Top left: Successive distance via the geodesic distance $d_{g^{\text{aff}}}$. After discarding the burn-in period, the plot showcases the last 100 data points, highlighted in red. Top right: Connected scatter plot. Utilizing the affine invariant metric, this plot exhibits a gradient of colors from red to green, culminating in blue. The percentage value in brackets represents the stress of the metric MDS plot. Bottom left: AIC values. AIC values for lags 1 to 10 are displayed, with the red point selected by minimizing AIC values. Bottom right: Boxplots of parameters. This plot presents boxplots comparing setting parameters (in red) with estimated parameters (in green).

Table 4.2: Two sample Kolmogorov-Smirnov test for two sets of parameters.

	$A_1 \& \hat{A}_1$	$A_2 \& \hat{A}_2$	$A_3 \& \hat{A}_3$	$B \& \hat{B}$	$\Sigma \& \hat{\Sigma}$
Statistic	0.16	0.15	0.11	0.16	0.13
p value	0.46	0.61	0.90	0.46	0.77

$1, \dots, m)$, resulting in a total of $(L + 2) \times m = 275$ intervals. However, to conserve space, we will not present the comparison of the initial and estimated parameters along with their 95% confidence intervals in this thesis. Interested readers can access the code and computed results through our R package: `geomTS`.

Correlation matrices on the quotient geometry $(\mathcal{C}^+(p), g^{\text{quo}})$

Within our R package, another simulation study is available for the model with diagonal matrix coefficients in the quotient geometry $(\mathcal{C}^+(p), g^{\text{quo}})$ generating time series of covariance matrices. To prevent redundant representation of results similar to simulations in $(\mathcal{S}^+(p), g^{\text{aff}})$, we exclusively present Figure 4.4 from simulation studies on $(\mathcal{C}^+(p), g^{\text{quo}})$. This presentation includes successive distance, MDS, AIC values, and distributions of initial and estimated parameters. The goodness-of-fit for the model is 66.88% in this

simulation study.

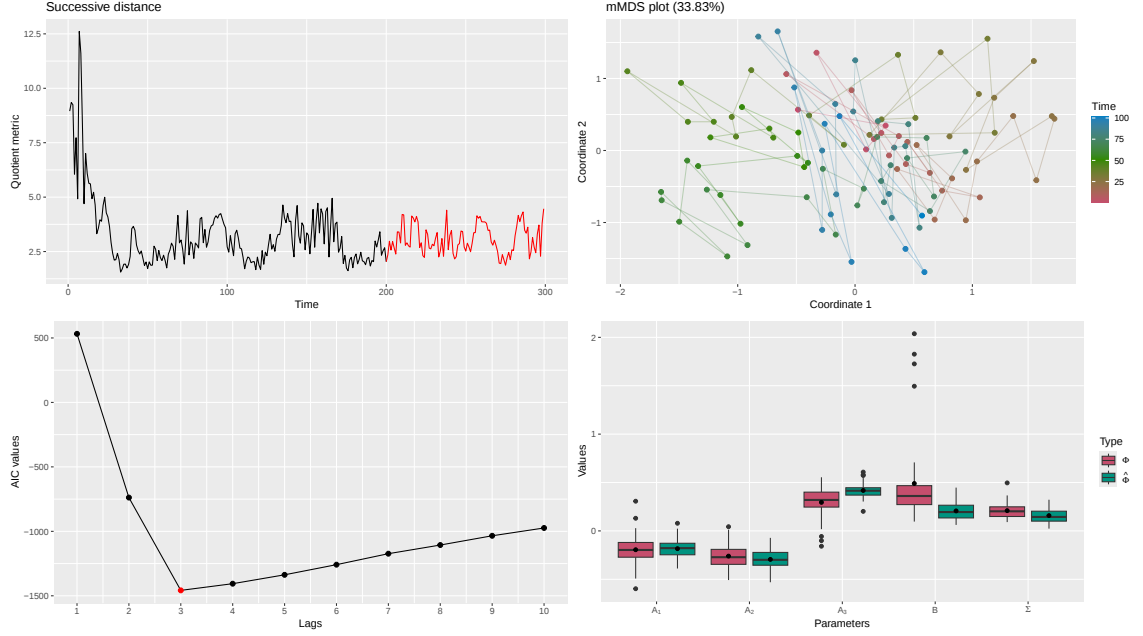


Figure 4.4: Top left: Successive distance via the geodesic distance $d_{g^{\text{quo}}}$. After discarding the burn-in period, the plot showcases the last 100 data points, highlighted in red. Top right: Connected scatter plot. Utilizing the quotient metric, this plot exhibits a gradient of colours from red to green, culminating in blue. The percentage value in brackets represents the stress of the metric MDS plot. Bottom left: AIC values. AIC values for lags 1 to 10 are displayed, with the red point selected by minimizing AIC values. Bottom right: Boxplots of parameters. This plot presents boxplots comparing setting parameters (in red) with estimated parameters (in green).

A noteworthy observation is that $\hat{B}$ exhibit different distributions compared to B . Potential reasons for this discrepancy have been discussed in the previous subsection, primarily stemming from numerically experimental results for optimal positions and the uniqueness of the Fréchet sample mean on the quotient geometry. Despite the slight bias in the estimated parameters, we believe the manifold-adapted model with diagonal matrix coefficients in the quotient geometry's has an ability to effectively capture data dynamics.

4.5.3 Summary

Based on the simulation studies conducted on our manifold-adapted models with scalar and diagonal matrix coefficients for covariance and correlation matrices under different geometries on $(\mathcal{S}^+(p), g^{\text{aff}})$ and $(\mathcal{C}^+(p), g^{\text{quo}})$, we assessed the effectiveness of our models, model inference, and the code in the R package `geomTS`. The estimated parameters obtained from the simulated data fitted into the models with scalar and diagonal matrix coefficients are statistically identical to the initial settings, despite a slight bias in the results on the quotient manifold due to numerical optimization issues. The questions regarding the

presence of numerical rounding errors and the uniqueness of the Fréchet sample mean in $(\mathcal{C}^+(p), g^{\text{quo}})$ still pose an open mathematical challenge.

Chapter 5

Applications to EEG data

The general manifold-adapted model in (4.2) comprises three components: the autoregression term, the mean reversion term, and the noise term. The first component assesses the influence of historical data on the current data, with regression coefficients $A_l, l = 1, \dots, L$ representing the extent of this influence. It is expected that the influence decreases as the lag increases, meaning that data further in the past has a diminishing effect on the present state. The second component quantifies the strength of the mean reversion, referring to the tendency of the data to revert to the attractor point. The parameter B determines the intensity of this attraction. In practical applications, the Fréchet sample mean of the interictal series corresponding to the seizure series is used as the attractor point S^* or C^* . The third component accounts for the presence of noise in the model. Noise represents random fluctuations or unpredictable elements in the data that cannot be attributed to any specific influence or pattern. It adds variability to the model and captures the unexplained portion of the observed data.

In this chapter, we present model results for an exemplar patient, Patient ID18, demonstrating the application of covariance matrix-valued time series data in the Euclidean geometry on $(\text{Sym}(p), g^{\text{euc}})$ and the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$. The goal is to discover the evolution of the covariance of pairs of channels during the seizure period. Additionally, after factoring out the variance amplitudes for pairs of channels, correlation matrix-valued time series data can explicitly depict the data dynamics of linear relationships, and we will discuss model results in geometries on $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and the quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$. The comparison of covariance (or correlation) matrix data in different geometries and the influence of choices of geometry on the model will be discussed in Section 5.2. Before that, we first introduce the results of the manifold-adapted model in $(\mathcal{S}^+(p), g^{\text{aff}})$ and $(\mathcal{C}^+(p), g^{\text{quo}})$ with respect to the covariance matrix and correlation matrix in Section 5.1 as an example showing how to interpret model results under the manifold-adapted models and diagnose model effectiveness. Models with

different reduced parameters are discussed in Section 5.1 as well. Lastly, Section 5.3 will measure the seizure variability within the individual patient and across patients.

5.1 Brain dynamics of EEG data under manifold-adapted models

In the preceding sections, the primary objective of the general manifold-adapted model in (4.2) has been to understand the intricate dynamics inherent in brain activity. In this section, we apply the manifold-adapted model to EEG data from Patient ID18, aiming to elucidate how the model interprets brain dynamics across five seizures and one interictal series for Patient ID18. Notably, the comparative analysis of the interictal series is selected by the last interictal time series corresponding to the final seizure time series (refer to more examples of Patients ID6, ID7, and ID13 in Appendix A).

The Riemannian geometries explored in this section involve time series of covariance matrices on $(\mathcal{S}^+(p), g^{\text{aff}})$ and correlation matrices on $(\mathcal{C}^+(p), g^{\text{quo}})$. By presenting the estimated parameters separately for models with scalar and diagonal coefficients, we seek to clarify the evolution of EEG signals between pairs of channels during seizure and interictal periods in subsection 5.1.1. By checking R^2 values for the goodness of model fit with different lags L in subsection 5.1.2 and discussing model selection in subsection 5.1.3, we conduct model diagnostics to assess adherence to model assumptions with the fixed L in subsection 5.1.4. Finally, subsection 5.1.5 summarizes the model results of EEG time series data.

5.1.1 Model interpretation in scalar coefficient and diagonal coefficient models

To mitigate the computational burden and align our approach with real-life brain dynamics, we restrict the temporal influence to the preceding eight seconds exclusively, configuring a fixed default lag (denoted as $L = 8$) within the general model (4.3). Following the model inference process applied to the EEG time series data of all seizures, AIC always suggests a maximum lag $L = 8$ for both scalar coefficient and diagonal models. As a result, we compile the estimated parameters from both models in Figure 5.1. The box plot illustrates the distributions of diagonal elements in matrices $\hat{A}_l, \hat{B}, \hat{\Sigma}$, for $l = 1, \dots, 8$, in each panel (seizure), with the red points representing the estimated scalar coefficients $\hat{\alpha}_l, \hat{\beta}, \hat{\sigma}$, for $l = 1, \dots, 8$, situated within the boxes.

It is evident that the estimated values of $\hat{\alpha}_l$ and $\hat{A}_l$, where $l = 1, \dots, 8$, are often negative in both geometries on $(\mathcal{S}^+(p), g^{\text{aff}})$ and $(\mathcal{C}^+(p), g^{\text{quo}})$, according to Figure 5.1. Negative values suggest oscillatory behaviours. To investigate this, we first used specific

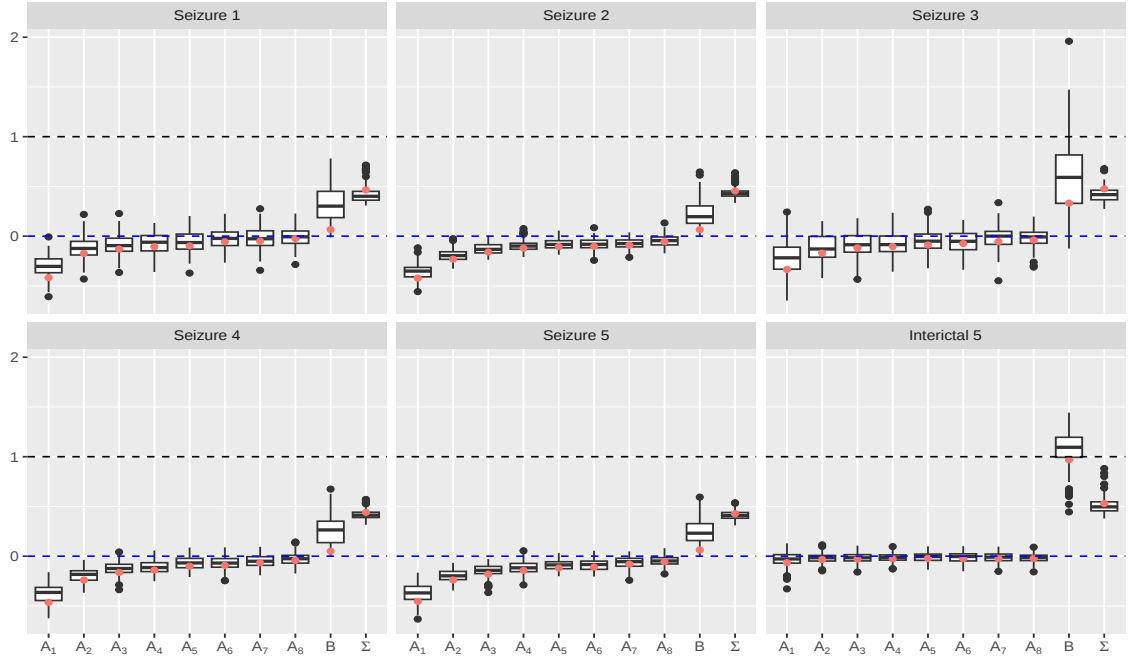
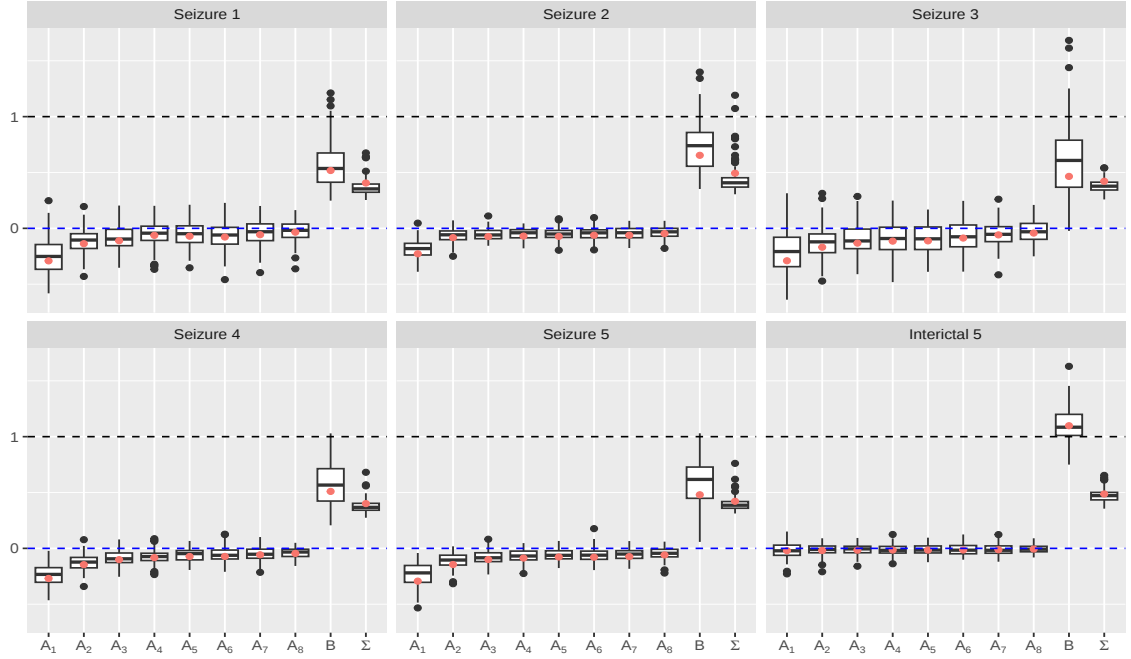

 (a) Covariance matrices in $(\mathcal{S}^+(p), g^{\text{aff}})$

 (b) Correlation matrices in $(\mathcal{C}^+(p), g^{\text{quo}})$

Figure 5.1: The estimated coefficients in the scalar and diagonal coefficient model (4.2) with fixed lags $L = 8$ for five seizures and an interictal period within Patient ID18 in geometries on $(\mathcal{S}^+(p), g^{\text{aff}})$ (top panel) and $(\mathcal{C}^+(p), g^{\text{quo}})$ (bottom panel). In the top panel, the red points represent the estimated scalar coefficients in the model with constant parameters. The boxes depict the distribution of estimated diagonals in the model with diagonal coefficients. The x -axis displays $\hat{A}_1$ to $\hat{A}_8$, as well as $\hat{B}$, and $\hat{\Sigma}$.

window sizes (non-overlapping windows of size 0.5 seconds, 2 seconds, and 3 seconds) to obtain the time series of covariance matrices. The negative inferred values persisted across these repeated analyses of data reduction and model fitting for the scalar coefficient model, suggesting oscillatory behaviour at different time scales.

Next, we calculated the quantities $g_{S_i}^{\text{aff}}(V_i, V_{i,1})/\|V_i\|\|V_{i,1}\|$ for $i = 2, \dots, n$, for each seizure for Patient 18 under the affine invariant geometry as an example. The values were negative, confirming the tendency for tangent vectors to reverse direction. Therefore, our conclusion is that the oscillatory behaviour indicated by the negative autoregressive coefficients is genuine at these time scales. However, the autoregressive terms might not be capturing the long-term evolution of the seizure trajectories observed in the MDS plots depicted in Figures 3.4 and 3.6, which illustrate a locally oscillatory pattern in the data.

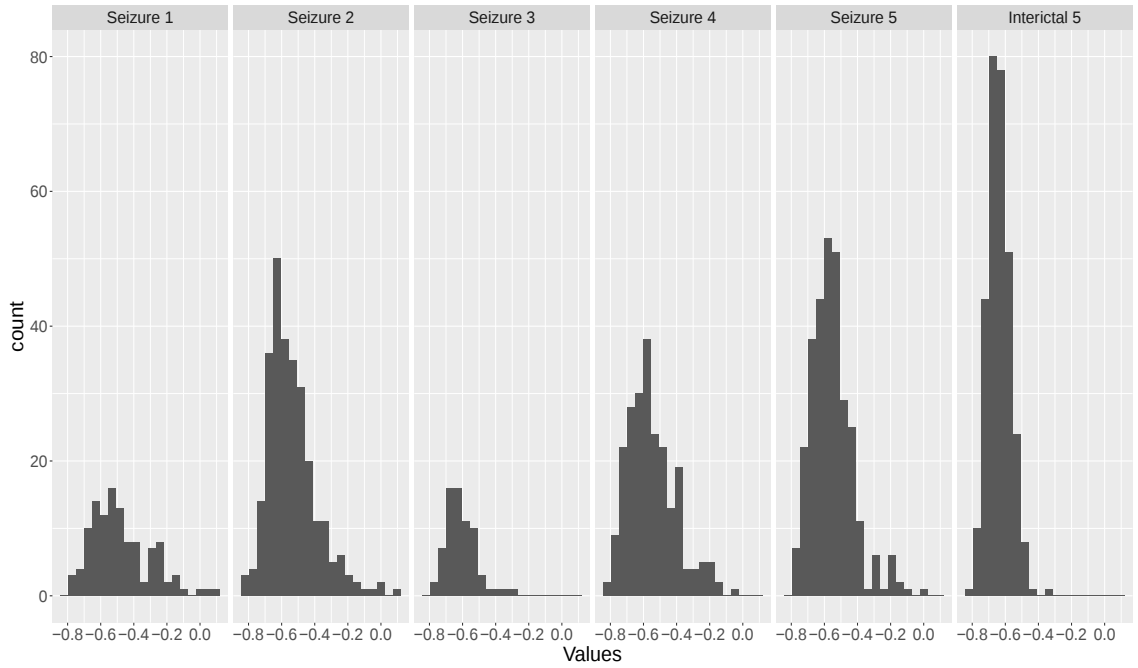


Figure 5.2: Histograms of normalized inner products between tangent vectors V_k and $V_{k,1}$ for 5 seizures and an interictal series of Patient 18.

Notably, the absolute magnitude of $\hat{\alpha}_l$ and $\hat{A}_l$ diminishes as the lag increases, signifying that data with higher lags exert a reduced impact on the current state. Additionally, the estimated values of $\hat{\beta}$ and $\hat{B}$ remain positive. This can be explained by the fact that brain states oscillate around an attractor point, and the velocity of the pathway returns to this attractor point at each time point. Overall, parameters with larger absolute values exert a more pronounced influence on the current data.

For a comprehensive representation of the distributions of $\hat{A}_l$, $\hat{B}$, and $\hat{\Sigma}$, please refer to the heatmap in Figure 5.3. Here, we present the diagonal matrix coefficients as $p \times p$ matrices, where each cell denotes the estimated values $\hat{a}_{lr}$ or b_r , with $l = 1, \dots, L$ and

$r = 1, \dots, m$. These values signify the strength of the influence of historical lagged vectors of l to the r -th pair of channels. It is noteworthy that the model in $\mathcal{S}^+(p)$ has more p dimensions than $\mathcal{C}^+(p)$, as it considers not only the covariance of each pair of channels but also the variance of channels. Visually, this heatmap offers a direct illustration of the evolution of brain dynamics for each pair of channels with different lags. For instance, Seizure 3 and the remaining seizures exhibit varying colour gradients in the heatmap, highlighting their differences. (The same colour gradient scheme is applied in each Riemannian geometry, with darker colours indicating larger influence values on the current brain states.)

Comparing the scalar and diagonal matrix coefficients between Figures 5.1a and 5.3a, we observe that scalar coefficients $\hat{\alpha}_l$ approximate the median of diagonals and primarily fall within the interquartile range of diagonal elements of $\hat{A}_l$. However, $\hat{\beta}$ appears notably distant from the median of $\hat{B}$. This suggests that constant parameters in the scalar coefficient model are not as sensitive in capturing the real pattern of the mean reversion term as they are in autoregressive terms. Furthermore, the diagonals in Figure 5.1a with respect to the attractor point term $\hat{B}$ are notably lighter than others. These are attributed to the Fréchet sample mean computed from the interictal time series, which has a higher magnitude of channel variance compared to the covariance of pairs of channels. To balance the value of the maximum log-likelihood function, the diagonals in the heatmap of $\hat{B}$ are close to 0, and $\hat{\beta}$ approaches 0 as well. Fortunately, this limitation is overcome by the property of the quotient manifold consisting of correlation matrices (see Figure 5.3b), where we only consider the linear relationship among pairs of channels. This is also reflected in Figure 5.1b where all estimated scalar coefficients in this context lie within the 75% quantile of estimated diagonal elements when examining the linear relationship for pairs of channels. Further comparisons of modelling results underlying different geometries will be discussed in Section 5.2.

In summary, although confirming that α_l, β, σ represent the mean or median of the distributions of diagonal elements in A_l, B, Σ poses a challenge, the results suggest that the scalar coefficients closely approximate the diagonal median and primarily fall within the interquartile range of diagonal matrices. This insight indicates that scalar parameters elucidate the overall influence of lagged ‘brain behaviours’ on the current ‘brain state’ within the scalar coefficient model. However, when investigating the lagged influence of each pair of channels, the model with diagonal coefficients would be more appropriate.

Furthermore, by selecting an interictal time series collected within the 2-hour window preceding Seizure 5, the final panel in Figures 5.1 and 5.3 illustrates that the interictal data is primarily influenced by the attractor point. This difference between seizure data and interictal data is evident as $\hat{\alpha}_l, \hat{A}_l \approx 0$ for $l = 1, \dots, L$, while $\hat{\beta}, \hat{B} \approx 1$ in Figure 5.1, and lighter colours indicating autoregression terms and darker red colours indicating mean

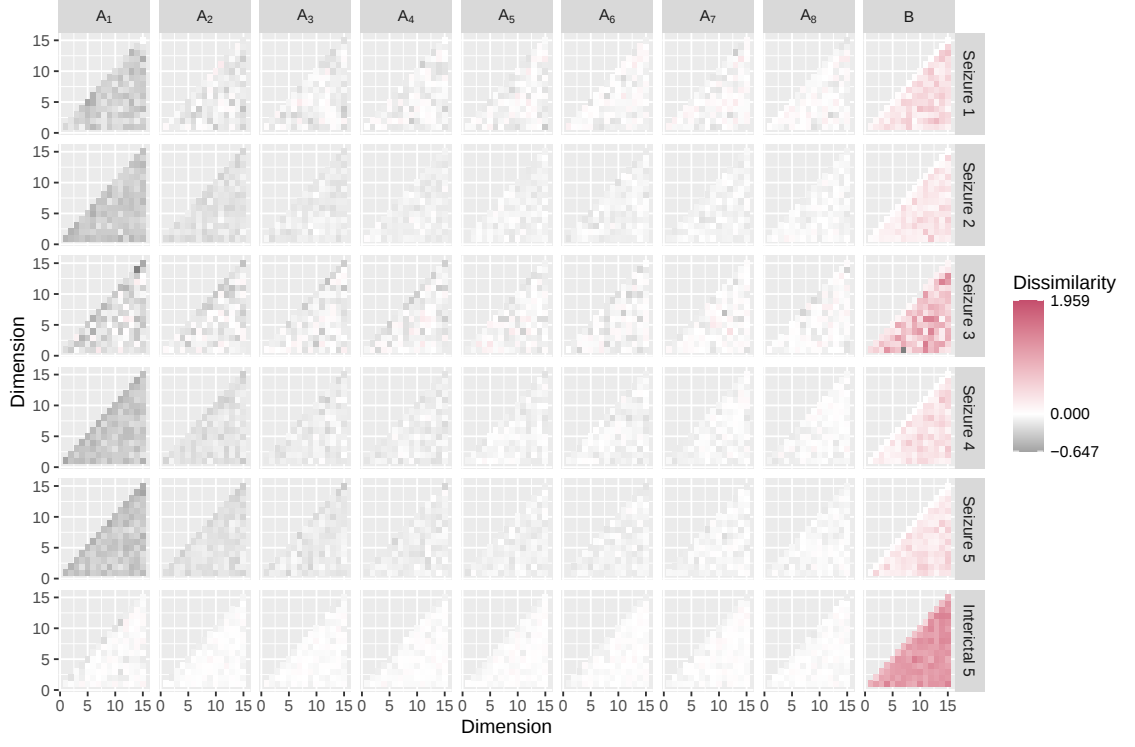
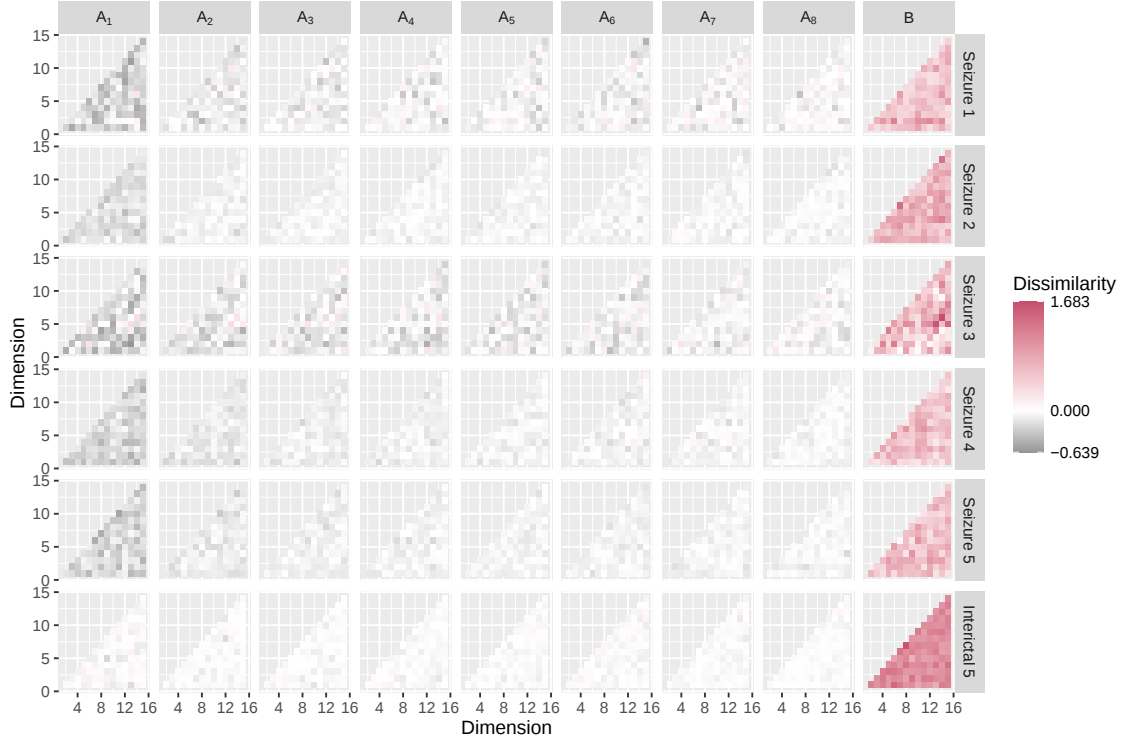

 (a) Covariance matrices in $\mathcal{S}^+(p)$

 (b) Correlation matrices in $\mathcal{C}^+(p)$

Figure 5.3: The heatmap of the distributions of diagonal matrix coefficients $\hat{A}_l$ and $\hat{B}$ for Patient ID18, where $l = 1, \dots, L$. The colours represent values ranging from the minimum in black, 0 in white, to the maximum in red.

reversion terms in Figure 5.3.

Notably, we observe that the values of $\hat{\alpha}_l$ and $\hat{A}_l$, where $l = 1, \dots, L$, for high lags are close to zero for both models under different geometries. This observation suggests that the influence of data with very high lags is not substantial. Section 5.1.3 will explain the reasons behind this and explore the selection of an appropriate maximum lag L .

5.1.2 Goodness of modelling fit

Like a traditional regression model, we implemented R^2 values to assess the goodness of modelling fit, defined in Equation (4.8). Table 5.1 and 5.2 display R^2 values for models featuring various types of coefficients (Sca. and Diag. represent the model with scalar and diagonal matrix coefficients, separately) and lags from 1 to 8. It is important to note that while R^2 values around 0.5 may not be deemed high in a low-dimensional context, they can still be considered reasonable and acceptable in high-dimensional settings where data complexity and noise levels are more pronounced.

Table 5.1: The goodness of modelling fit by R^2 for Patient ID18 with different types of reduced parameters and lags using covariance matrices in $(\mathcal{S}^+(p), g^{\text{aff}})$.

	Seizure 1		Seizure 2		Seizure 3		Seizure 4		Seizure 5		Interictal 5	
	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.
lag=1	0.290	0.371	0.348	0.404	0.482	0.537	0.363	0.433	0.358	0.418	0.615	0.623
lag=2	0.313	0.386	0.396	0.436	0.499	0.555	0.417	0.464	0.405	0.446	0.616	0.624
lag=3	0.323	0.395	0.414	0.449	0.506	0.565	0.435	0.476	0.427	0.462	0.616	0.625
lag=4	0.328	0.403	0.423	0.457	0.509	0.576	0.447	0.487	0.437	0.471	0.616	0.627
lag=5	0.331	0.411	0.427	0.461	0.514	0.587	0.452	0.492	0.443	0.476	0.616	0.628
lag=6	0.332	0.416	0.432	0.467	0.516	0.596	0.457	0.499	0.449	0.482	0.616	0.630
lag=7	0.332	0.422	0.437	0.472	0.513	0.599	0.459	0.503	0.451	0.486	0.617	0.632
lag=8	0.331	0.428	0.439	0.476	0.514	0.607	0.461	0.507	0.454	0.490	0.617	0.634

Table 5.2: The goodness of modelling fit by R^2 for Patient ID18 with different types of reduced parameters and lags using correlation matrices in $(\mathcal{C}^+(p), g^{\text{quo}})$.

	Seizure 1		Seizure 2		Seizure 3		Seizure 4		Seizure 5		Interictal 5	
	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.
lag=1	0.517	0.549	0.550	0.592	0.501	0.531	0.521	0.550	0.533	0.575	0.623	0.629
lag=2	0.531	0.564	0.562	0.600	0.509	0.545	0.540	0.565	0.557	0.591	0.623	0.630
lag=3	0.538	0.574	0.569	0.606	0.510	0.551	0.549	0.573	0.567	0.598	0.623	0.631
lag=4	0.540	0.580	0.575	0.611	0.515	0.563	0.553	0.579	0.574	0.603	0.623	0.633
lag=5	0.542	0.588	0.579	0.615	0.519	0.574	0.557	0.584	0.578	0.607	0.623	0.634
lag=6	0.543	0.595	0.582	0.619	0.523	0.586	0.560	0.589	0.581	0.611	0.623	0.636
lag=7	0.543	0.602	0.585	0.623	0.522	0.592	0.562	0.593	0.584	0.615	0.623	0.637
lag=8	0.544	0.607	0.587	0.626	0.526	0.604	0.564	0.596	0.587	0.618	0.623	0.638

Upon closer examination of the tables, it becomes evident that models with higher lags consistently exhibit slightly larger R^2 values. This trend primarily arises from their inclu-

sion of more autoregression terms to capture dynamic patterns. This observation holds true for the model with diagonal matrix coefficients as well. Furthermore, the diagonal model consistently demonstrates a better fit than the scalar coefficient model due to its incorporation of more explanatory parameters, which account for the covariance/correlation of every pair of channels. However, it should be noted that the improvement in R^2 is not substantial, and the inclusion of more explanatory parameters in the model also leads to increased computational burden.

More importantly, if the manifold is not parallelizable and the goal is to obtain a single index for assessing the influence of historical data on current brain dynamics, it is recommended to opt for the manifold-adapted model (4.3) with scalar coefficients. Conversely, if the aim is to evaluate brain dynamics for each pair of channels, and the geometry space allows for parallelization, the model (4.3) with diagonal matrix coefficients can more accurately describe the systemic pattern of brain dynamics.

5.1.3 Model selection of lags

The model selection criteria employed in simulation studies, as described in Section 4.5, effectively demonstrate their capability to identify suitable lag settings for both scalar and diagonal coefficient models. However, when utilizing the AIC criteria and estimating parameters in both models as shown in Figure 5.1 and 5.3, it becomes evident that the values of $\hat{\alpha}$, $\hat{A}_l, l = 1, \dots, L$ for high lags are close to zero. On the other hand, it has been demonstrated that the higher lagged terms in autoregressive terms do not improve the modelling fit with desirable levels as discussed in Section 5.1.2. These results suggest that the influence of data with very high lags is not substantial. The reason behind this phenomenon lies in the nature of the AIC, which balances model goodness of fit with a penalty for the number of parameters utilized. A high lag recommended by AIC might imply the ill-specified model, including unnecessary parameters, or the capture of random fluctuations or noise that lacks a meaningful relationship with the underlying process.

To address this challenge, it is essential to directly assess the relationship between the current observation and its lagged values while accounting for the effects of intermediate lags. The PACF is employed to determine the order of autoregressive components in an autoregression model as described in Section 4.3.3. A significant PACF value at a specific lag indicates that this lag is crucial in explaining the current value of the time series. Therefore, PACF helps identify the genuine lagged dependencies for autoregressive terms. It is generally assumed that the partial autocorrelations of lags are approximately independent and normally distributed with a mean of 0 (Quenouille, 1949). Confidence intervals can be constructed by dividing a chosen z -score by $\sqrt{n}$. Lags with partial autocorrelations falling outside the confidence interval are considered significant. Figure 5.4 visually represents the partial autocorrelation function for seizures within Patient ID18,

including blue lines indicating the confidence intervals.

In Figure 5.4, only the lower confidence intervals of PACF values are presented. This choice stems from the observation of negative relationships between predictors and response variables in the autoregressive term, as discussed in the previous Section 5.1. Consequently, based on the PACF analysis, it is advisable to employ the model with reduced scalar coefficients while utilizing lower lags. Regarding the model with diagonal matrix coefficients, it is impractical to generate PACF plots for each pair of channels with varying lags. Therefore, in Figure 5.5, we represent 0-1 heatmaps. These heatmaps highlight significant cells of $\hat{a}_{lr}$ or $\hat{b}_r$, with values marked in black if they fall outside the confidence interval, and in white if within.

To achieve a balance between model complexity and capturing direct dependencies, thereby avoiding overfitting and enhancing the efficiency of comparing seizure variability using our models, we examined PACF values across all 112 seizures for covariance matrix-valued time series using the scalar coefficient model, as an example. Specifically, for each value of L , a 95% confidence interval was constructed for α_L to test the hypothesis that $\alpha_L = 0$. The value of L for each time series was determined as the largest for which the null hypothesis was not rejected. This procedure for selecting the maximum lag L resulted in a value of $L = 0$ for the vast majority of interictal series. For seizures, the majority had $L = 1$ or 2 . These results are consistent across both geometries ($\text{Sym}(p), g^{\text{euc}}$) and $(\mathcal{S}^+(p), g^{\text{aff}})$, and are depicted in Figure 5.6. Similar results can be obtained for the diagonal model and the dataset of correlation matrices, as observed in the example of Patient ID18 (see Figure 5.4 and 5.5).

Consequently, we chose to fit all models from this point onwards with $L = 4$ for the following reasons. First, it was essential to use the same value for all time series to facilitate the comparison of seizures between patients. Secondly, selecting $L = 4$ ensured consistency across most series, as it was greater than or equal to the value observed in the vast majority of cases. However, it is noteworthy that opting for $L = 4$ resulted in estimated parameter values close to zero for larger values of l in many time series. Nonetheless, to maintain consistency, we decided on a fixed lag of $L = 4$ for both models across all datasets. Following this decision, we further investigated the validity of the models by examining the residuals after fitting.

5.1.4 Residual analysis after modelling fit with fixed maximum lag $L = 4$

The maximum likelihood estimator assumes that residuals are normally distributed. Since the dimensions of coordinate vectors $\mathbf{v} \in \mathbb{R}^m$ are mutually independent (i.e., Riemannian manifold is parallelizable), we implement the Kolmogorov–Smirnov test to check if the residuals are drawn from a normal distribution $N(0, \hat{\Sigma})$. Furthermore, the Ljung-Box test is used to check the independence of residuals for each dimension. Ideally, there should be

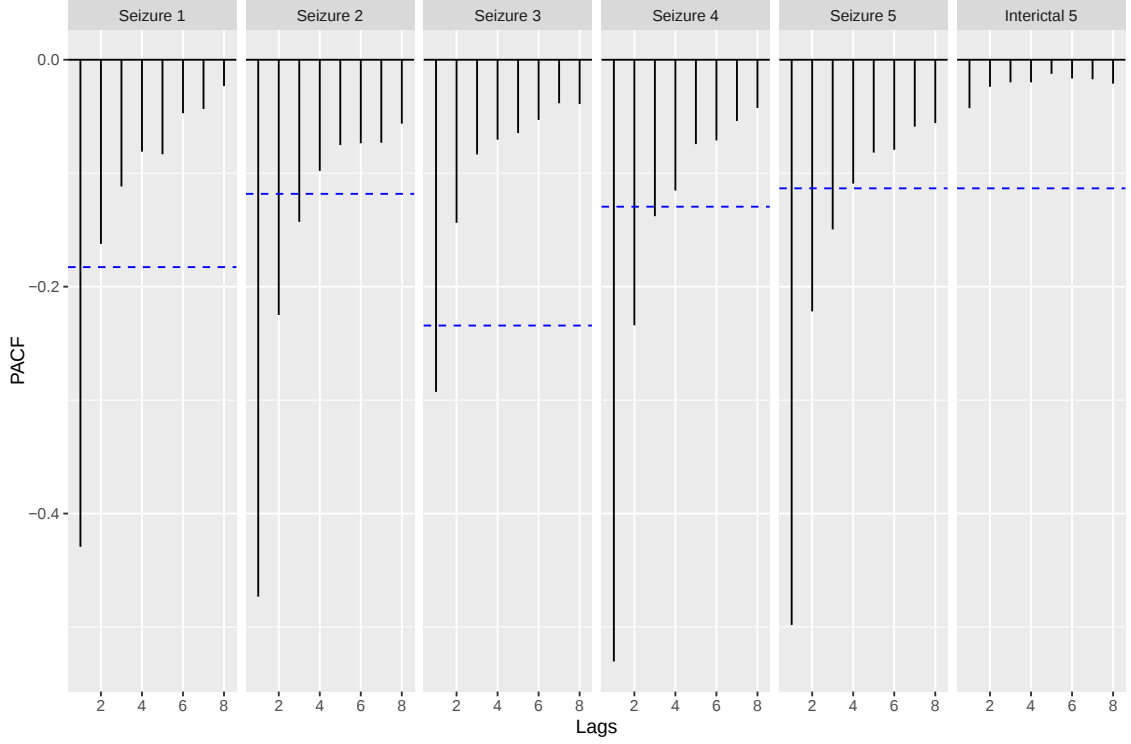
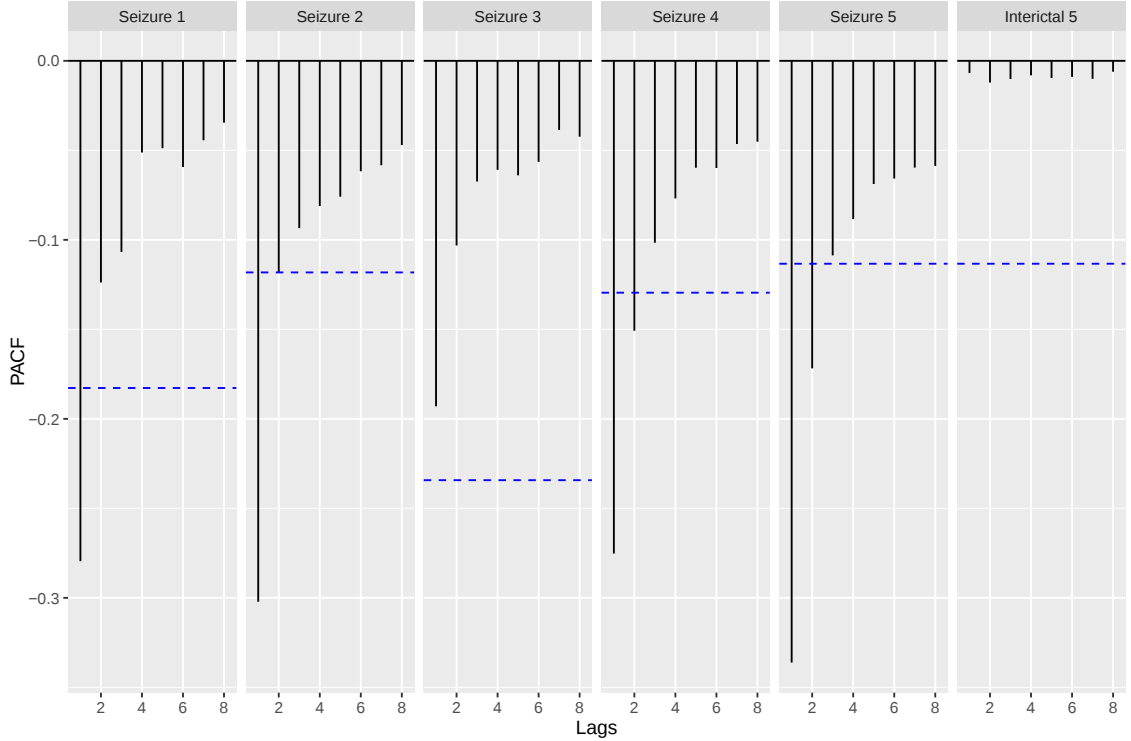
(a) Covariance matrices in $\mathcal{S}^+(p)$ (b) Correlation matrices in $\mathcal{C}^+(p)$

Figure 5.4: PACF for 5 seizures and an interictal data within Patient ID18 in the scalar coefficient model. The maximum lags are 8. Since all partial correlations are negative, only lower bounds of confidence intervals with blue dashed lines are shown in the plot.

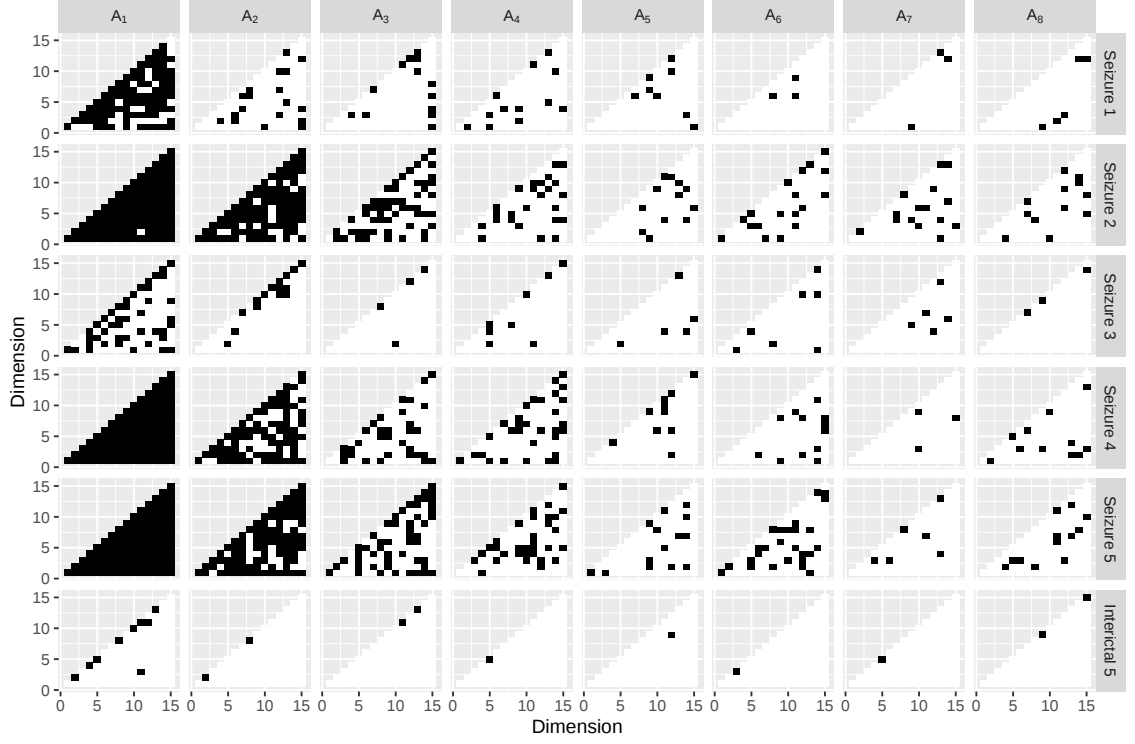
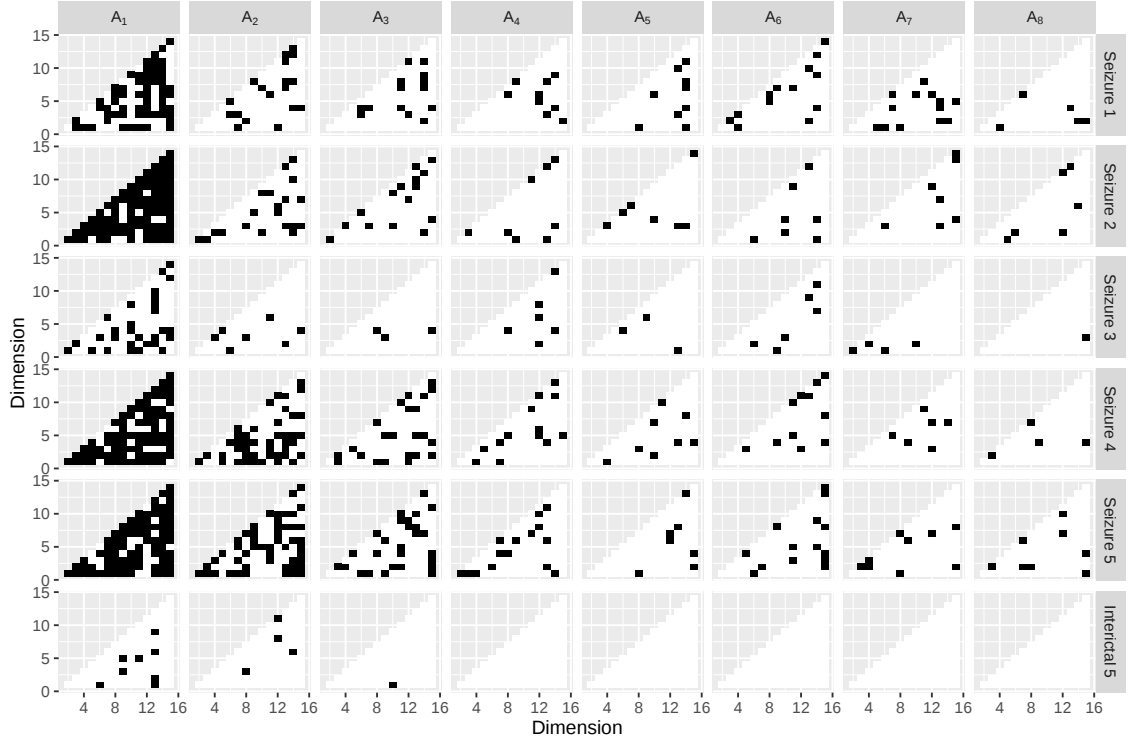
(a) Covariance matrices in $\mathcal{S}^+(p)$ (b) Correlation matrices in $\mathcal{C}^+(p)$

Figure 5.5: PACF for 5 seizures and an interictal data within Patient ID18 in the diagonal model. The maximum lags are 8. The black cells indicate the estimated $\hat{a}_{lr}$ are outside the 95% confidence intervals.

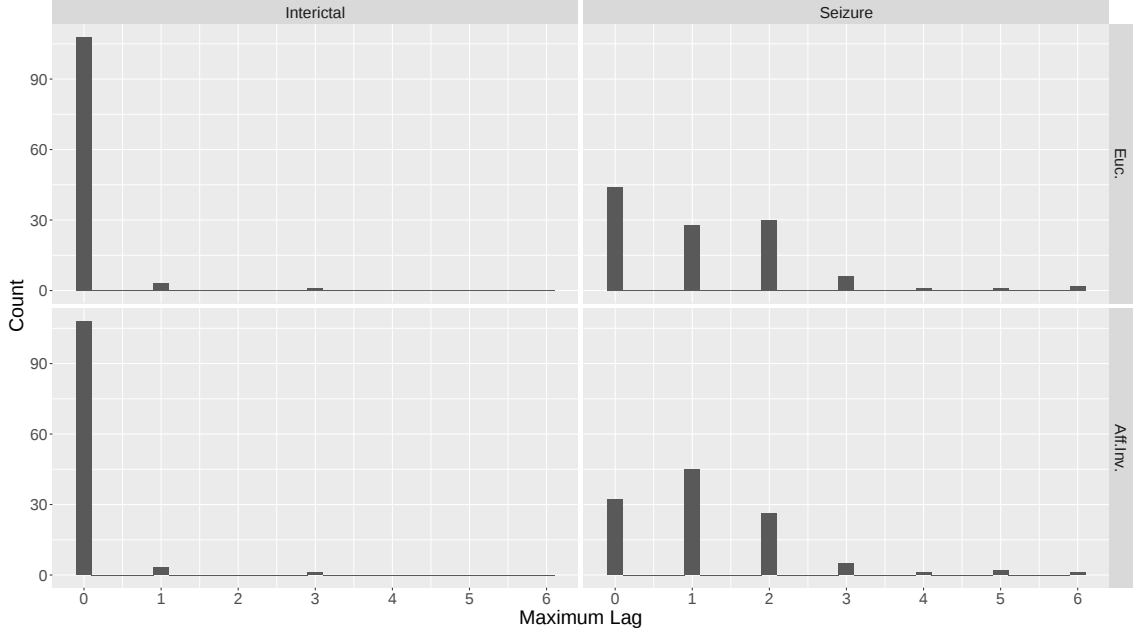


Figure 5.6: Histograms of the value of L selected for the scalar coefficient model fitted to seizure and interictal series in Euclidean and affine invariant geometries.

no autocorrelation among residual variables with any lags. More importantly, stationary residuals indicate that the model has adequately accounted for the autoregressive dynamics. This is often checked by the Augmented Dickey–Fuller test, which computes the test for the null hypothesis that the residual has a unit root.

Implementing the tests to check assumptions for each dimension, Tables 5.3 and 5.4 present the counts of dimensions satisfying the specified assumptions in $(\mathcal{S}^+(p), g^{\text{aff}})$ and $(\mathcal{C}^+(p), g^{\text{quo}})$ for covariance and correlation matrices, respectively. Note dimensions of tangent vectors in $T_S \mathcal{S}^+(p)$ and $T_C \mathcal{C}^+(p)$ are $p(p+1)/2 = 120$ and $(p-1)p/2 = 105$ when $p = 15$. In general, the model with diagonal matrix coefficients leverages brain dynamics for each pair of channels, resulting in a better modelling fit. Consequently, the estimated parameters for binary channels are unbiased for both geometries.

Table 5.3: Counts of dimensions satisfying assumptions for covariance matrices in $(\mathcal{S}^+(p), g^{\text{aff}})$ within Patient ID18.

	Seizure 1		Seizure 2		Seizure 3		Seizure 4		Seizure 5		Interictal 5	
	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.
Normality	107	112	97	100	101	105	100	107	98	101	103	119
Independence	94	118	76	118	107	120	101	119	81	116	100	120
Stationarity	112	117	120	120	103	84	120	120	120	120	120	120

In $(\mathcal{S}^+(p), g^{\text{aff}})$, the majority of dimensions of residuals satisfy the three assumptions mentioned earlier, and we consider the remaining dimensions that deviate from these assumptions as outliers. When examining the dimensions that deviate from the assumptions,

Table 5.4: Counts of dimensions satisfying assumptions for correlation matrices in $(\mathcal{C}^+(p), g^{\text{quo}})$ within Patient ID18.

	Seizure 1		Seizure 2		Seizure 3		Seizure 4		Seizure 5		Interictal 5	
	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.
Normality	77	89	14	47	74	90	44	63	38	50	76	91
Independence	83	100	78	96	93	105	82	96	72	96	96	105
Stationarity	104	104	105	105	88	81	105	105	104	105	105	105

it is interesting to note that the variance of channels tends to exhibit more complicated patterns compared to the covariance of pairs of channels. That is, dimensions which are against the normality assumptions are mainly covered by the variance of channels, i.e., the diagonal indices in $A_l, B, l = 1, \dots, L = 4$.

Unfortunately, this is not the case for the quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$, consisting of the set of correlation matrices. When focusing solely on the linear relationships for pairs of channels, a high number of observations of residuals are not normally distributed with zero mean and variance $\hat{\Sigma}$, even with the model employing diagonal matrix coefficients. We have investigated all dimensions of residuals obtained from the model in $(\mathcal{C}^+(p), g^{\text{quo}})$, and most of them are ‘close’ to following the normal distribution $N(0, \hat{\Sigma})$, either with their mean approaching 0 or their variance being the same as $\hat{\sigma}^2$ by histogram plots of residuals. The solution to solve this problem may involve a new, more precise dynamical model. In this research, we temporarily assume these dimensions are outliers and do not influence our final measurement of seizure dissimilarity.

5.1.5 Summary

We have verified that the diagonal model consistently fits the data better than the scalar model for covariance and correlation matrices. However, it is important to note that the results from the diagonal model fitted by the reduced covariance matrices are challenging to interpret. This is due to the reduced covariance matrices losing their electrode information, making it difficult to explain the estimated parameters for pairs of channels. Conversely, the scalar coefficient model is more convenient for demonstrating how brain dynamics evolve during seizures in a general way. Additionally, when examining the correlation matrices, it is crucial to acknowledge that the dimension of the affine invariant geometry is different from the Euclidean and quotient geometries, which have more p dimensions. In this case, the results from the diagonal model are also challenging to interpret.

To further EEG data analysis for seizure variability, the scalar coefficient model is suggested for our research for the following reasons: (i) The comparison of model results underlying different geometries demonstrates that the scalar coefficient model can effectively capture brain dynamics. This is evident in terms of the goodness of modelling fit and the satisfaction of model assumptions. (ii) There are no significant differences in

results between the scalar coefficient and diagonal models. (iii) The parameters from the scalar coefficient model provide a more straightforward demonstration of general brain dynamics, avoiding the intractable interpretation of the results from the diagonal model.

5.2 Comparing model results on different geometries

5.2.1 Covariance matrix-valued time series in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$

For the covariance matrix-valued time series of EEG data after dimensional reduction by Method I described in Section 3.2.1, it lies on two manifolds: the Euclidean geometry on $(\text{Sym}(p), g^{\text{euc}})$ and the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$. We implement our manifold-adapted model with scalar coefficients based on both spaces. Figure 5.7 presents the estimated parameters $\hat{\Phi} = (\hat{\alpha}_1, \dots, \hat{\alpha}_4, \hat{\theta}, \hat{\sigma})^T$ for seizures and interictal periods for Patient ID18.

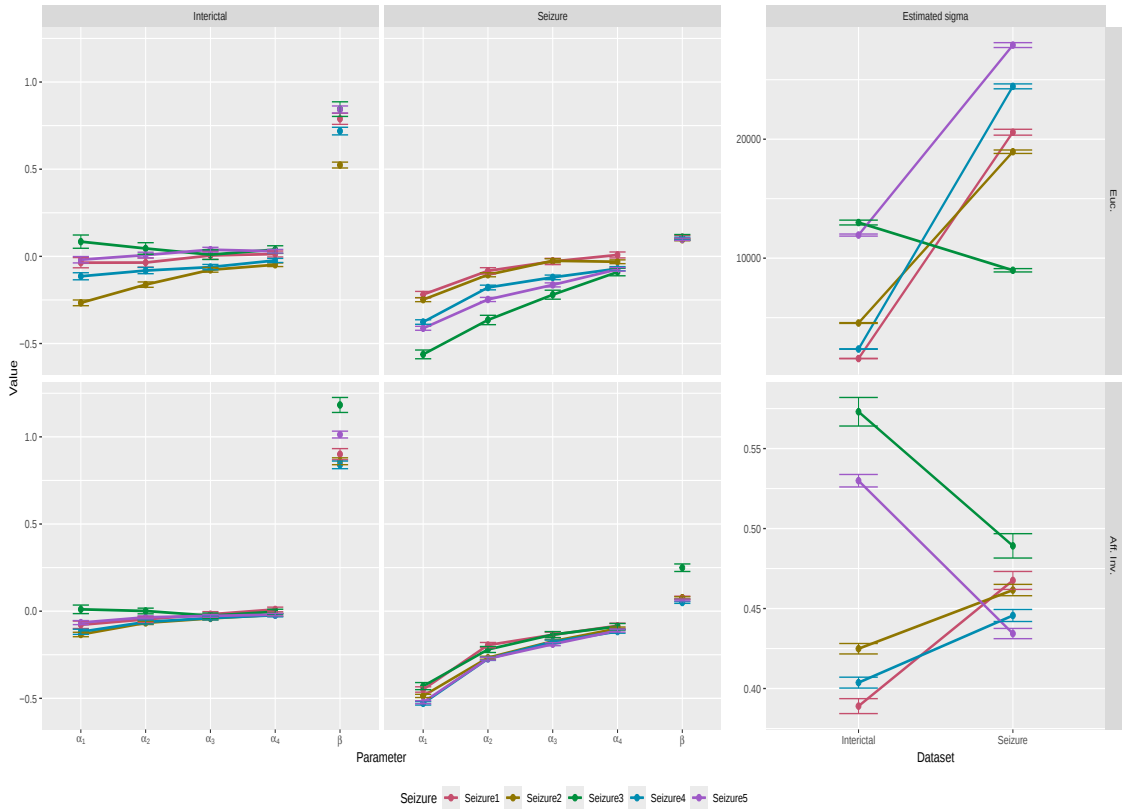


Figure 5.7: Scalar coefficients for Patient ID 18: comparison of estimated values for five seizures and corresponding interictal series, using Euclidean and affine invariant geometries (Top and bottom row, respectively). The values of model coefficients $\alpha_1, \dots, \alpha_4, \beta$, and σ are shown with 95% confidence intervals marked.

Figure 5.7 shows the estimated parameters $\alpha_1, \dots, \alpha_4, \beta$ and σ in the Euclidean and

affine invariant geometries for seizures and interictal periods for Patient 18. We first consider the results for the affine invariant geometry before comparing the geometries below. The results under the affine invariant geometry (bottom row) show that for the interictal series, the autoregressive terms α_l are close to zero, while the mean reversion coefficient β is close to 1. In contrast, for the seizure data, the α_l are negative, with 95% confidence intervals which do not include zero, and with decreasing absolute value as l increases. The mean reversion coefficient β for the seizures is close to zero apart from the value for seizure 3. The inferred values are quite homogeneous between seizures with the exception of seizure 3, which had a noticeably different trajectory in the previous MDS analysis (see Figure 3.4b). These results suggest that during interictal periods, the brain dynamics follow a mean-reverting random walk in $(\mathcal{S}^+(p), g^{\text{aff}})$, while during seizures, the mean reversion behaviour is weaker with a significant autoregressive component which gives rise to the trajectories observed in the MDS plots. The procedure for selecting the maximum lag L (see Figure 5.6) showed that across the entire data set, a number of seizures had no significant autoregressive terms ($L = 0$). These mainly consisted of shorter time series for which the confidence region for α_1 was large.

Next, we compare the results for the affine invariant geometry with those under the Euclidean geometry (top row in Figure 5.7). The values of α_l and β are similar for the two geometries and show qualitatively similar dynamics for seizures and interictal series as the affine invariant model. However, there is greater heterogeneity between series for the estimated parameters under the Euclidean geometry. Greater differences between geometries are apparent in the estimated values of σ shown in the right-hand panel of Figure 5.7. There are very substantial increases in σ between interictal periods and seizures under the Euclidean geometry, while the estimated values under the affine invariant geometry are broadly similar between interictal and seizure series. Thus under the Euclidean model, much of the variation in seizure data is modelled as noise.

Furthermore, to provide a more comprehensive evaluation of the model's performance across different geometries, we have computed the R^2 values to assess the goodness of the modelling fit. The results are summarized in Table 5.5. It shows substantially better model fit for the affine invariant model across all seizures and interictal periods, and this was also the case for other patients. This is a result of the intrinsic nature of the model in the affine invariant geometry. For both geometries, the R^2 values are higher for interictal series than the corresponding seizures.

Table 5.5: The goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$ for Patient ID18

	Seizure 1		Seizure 2		Seizure 3		Seizure 4		Seizure 5		Interictal 5	
	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.
$\text{Sym}(p)$	0.12	0.26	0.12	0.22	0.34	0.43	0.20	0.27	0.22	0.29	0.43	0.44
$\mathcal{S}^+(p)$	0.33	0.40	0.42	0.46	0.51	0.58	0.45	0.49	0.44	0.47	0.62	0.63

Residual analysis has been discussed in Section 5.1.4 for Patient ID18. Although there are a few dimensions of residuals that deviate from the assumptions, we consider them as outliers but still believe the scalar coefficient model is suitable for capturing the brain dynamics by the covariance of pairs of channels based on the boxplots (in Figure 5.1a) and R^2 values (in Table 5.1).

Additional examples of model results for covariance matrices can be found in Appendix A, specifically for Patient ID6, ID7, and ID13.

5.2.2 Correlation matrix-valued time series in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$

For full-rank correlation matrix-valued time series of EEG data after dimensional reduction by Method II described in Section 3.2.1, it lies on three different Riemannian geometries: the Euclidean manifold $(\text{Sym}(p), g^{\text{euc}})$, the affine invariant manifold $(\mathcal{S}^+(p), g^{\text{aff}})$, and the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$. It is well-known that $\mathcal{C}^+(p) \subset \mathcal{S}^+(p) \subset \text{Sym}(p)$. This means correlation matrix-valued time series can be modelled in any of these three nested spaces and then fitted to the model with scalar coefficients. The estimated parameters underlying these nested spaces are shown in Figure 5.8.

Analogous to the result analysis in the previous subsection for covariance time series in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$, the estimated $\hat{\alpha}_l$ and $\hat{\beta}$ show a similar trend pattern for brain dynamics considering the correlation matrix-valued time series. It is coincided that the interictal time series is a random walk centred at an attract point in these three spaces. Observing estimated parameters in Figure 5.8, Seizure 3 has more variation compared to other seizures in $(\text{Sym}(p), g^{\text{euc}})$, and there is less dissimilarity among seizures underlying the geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$ and $(\mathcal{C}^+(p), g^{\text{quo}})$, excluding the high strength of $\hat{\beta}$ in the mean-reverting term in $(\mathcal{C}^+(p), g^{\text{quo}})$.

As the correlation matrix is the scaled covariance matrix factoring out the variance of channels, the influence of data magnitude on model results is rarely found. This implies that the estimated coefficients under different geometries show a similar pattern. One of the main benefits of analysing correlation matrix-valued time series in the quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$ is that it considers the curvature of parallel-transported tangent vectors, similar to what the affine invariant geometry does in $\mathcal{S}^+(p)$ with covariance matrices. While it is true that the correlation matrix is symmetric positive definite and the data set could be analysed in $(\mathcal{S}^+(p), g^{\text{aff}})$, the dimension of $(\mathcal{S}^+(p), g^{\text{aff}})$ is $m = p(p+1)/2$, whereas the dimensions of $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{C}^+(p), g^{\text{quo}})$ are $m = (p-1)p/2$. Therefore, it is hard to interpret the estimated parameters in diagonal indices for the model with diagonal matrix coefficients.

On the other hand, under the model assumptions as described in Section 4.4, R^2 values are computed for different models with reduced parameters and in different geometries

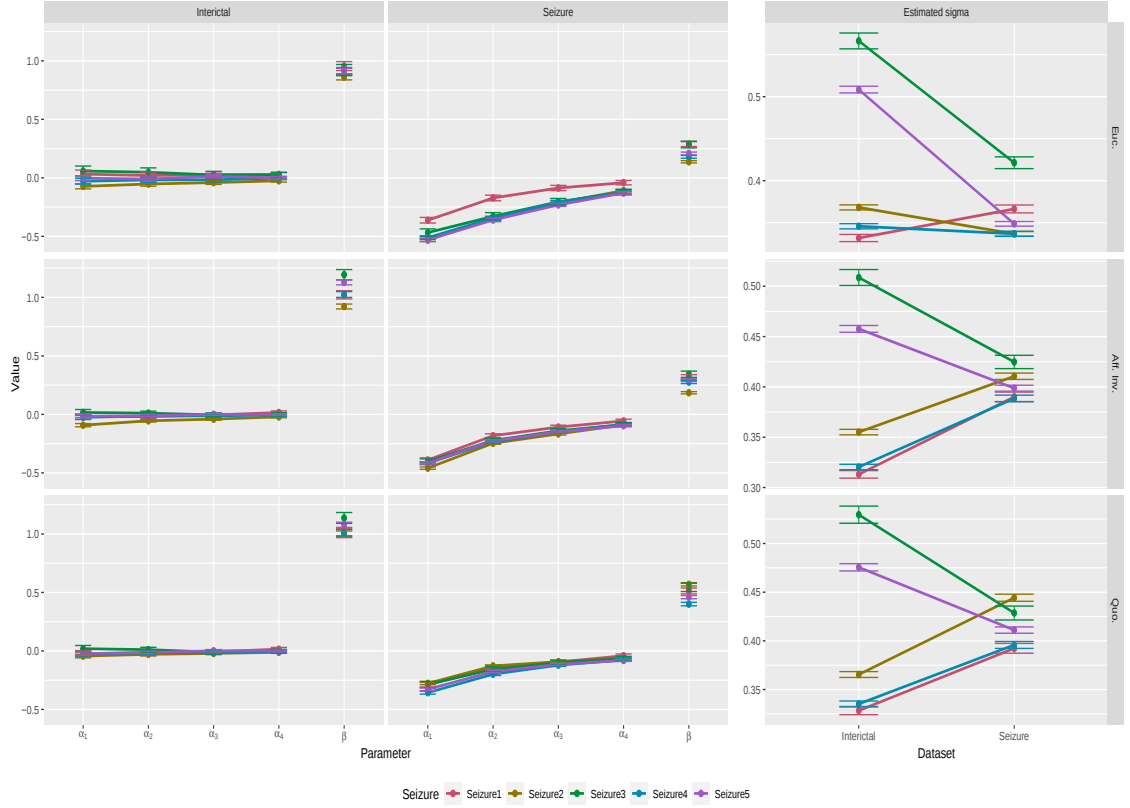


Figure 5.8: Scalar coefficients for Patient ID 18: comparison of estimated values for five seizures and corresponding interictal series, using Euclidean, affine invariant, and quotient geometries (Top, middle, and bottom row, respectively). The value of model coefficients $\alpha_1, \dots, \alpha_4, \beta$ and σ are shown with 95% confidence intervals marked.

(See Table 5.6). Apparently, the quotient manifold-valued model with a diagonal matrix coefficient has better performance than others, but there is an insignificant improvement in the model with scalar coefficients. This is explained by the pairs of channels being independent, and the means of diagonals are close to scalar coefficients.

Table 5.6: The goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$ for Patient ID18.

	Seizure 1		Seizure 2		Seizure 3		Seizure 4		Seizure 5		Interictal 5	
	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.
$\text{Sym}(p)$	0.298	0.339	0.314	0.344	0.367	0.428	0.328	0.357	0.357	0.379	0.456	0.465
$\mathcal{S}^+(p)$	0.437	0.495	0.462	0.498	0.500	0.566	0.469	0.501	0.489	0.520	0.628	0.635
$\mathcal{C}^+(p)$	0.540	0.580	0.575	0.611	0.515	0.563	0.553	0.579	0.574	0.603	0.623	0.633

Recalling the residual analysis from Section 5.1.4, the boxplots in Figure 5.1b and R^2 values in Table 5.2 collectively affirm the reliability of the manifold-adapted model with scalar coefficients. Despite the non-normality of residuals for the majority of dimensions,

as indicated in Table 5.4, we believe that our model, specifically designed for the quotient geometry $(\mathcal{C}^+(p), g^{\text{quo}})$, provides a better fit for correlation matrix-valued time series than other geometries.

5.2.3 Summary

When comparing the model results across various geometric settings, it becomes apparent that the performance of the affine invariant geometry surpasses that of Euclidean space when examining covariance matrices. The goodness of modelling fit significantly improves in $(\mathcal{S}^+(p), g^{\text{aff}})$, benefiting from the affine invariant transformation for high magnitudes. However, when scrutinizing correlation matrices, both Euclidean and affine invariant geometries produce similar results to quotient geometries. This is because the correlation matrices are the scaled data fitted in the models, such that the affine invariant metric loses its advantages, and there is less variation in statistical operations between Euclidean geometry and Riemannian geometry. For example, vectorization and parallel transport in $(\mathcal{S}^+(p), g^{\text{aff}})$ and $(\mathcal{C}^+(p), g^{\text{quo}})$. Nevertheless, the quotient geometry still demonstrates a superior goodness of fit than others.

In summary of the findings from the previous subsections, the affine invariant manifold-valued model proves to be sensitive in capturing the real brain dynamics, particularly considering the covariance of pairs of channels. When it comes to the data dynamics of the linear relationship of pairs of channels, the quotient manifold-valued model demonstrates superior performance compared to other manifolds. The use of Euclidean geometry for comparison purposes reveals that Riemannian properties can aid in analysing matrix-valued data and are superior to traditional Euclidean space.

More broadly, we have constructed a Riemannian framework for geometric statistical analysis of symmetric positive definite matrix-valued data. Beyond time series models, this research has the potential to inspire the development of more non-Euclidean statistical models, incorporating a geometric perspective into statistical ideas.

5.3 Seizure and patient variability

Conventional brain dynamic analysis, such as the model of functional network dynamics, heavily relies on channel information and electrode locations (Schindler *et al.*, 2008; Schroeder *et al.*, 2020). These methods assess the evolution of the recording signal for specific channels and compare the sectional functions of the brain based on electrode placements. However, the dependence on channel information poses limitations. To overcome these limitations and evaluate overall brain dynamics, our manifold-adapted models generate a set of parameters that succinctly describe general brain evolution patterns, providing a simpler way to measure seizure dissimilarity.

Using the estimated parameters, as illustrated in Figures 5.7 and 5.8 for Patient ID18, this section will delve into quantifying the dissimilarity among seizures between and within patients with epilepsy. This becomes particularly relevant when seizures do not exhibit significant visual differences from each other.

5.3.1 Seizure variability within patients

As the estimated parameters are obtained through MLE, we can leverage the asymptotic covariance matrices (computed in Section 4.3) to establish a quantitative measurement of seizure dissimilarity.

Let $\Phi = (\hat{A}_1, \dots, \hat{A}_4, B, \Sigma)^T$ denote the set of all estimated parameters. The seizure dissimilarity between any two seizures, indexed by i and j , is defined using the Mahalanobis distance as follows:

$$d_{i,j} = \frac{1}{2} \left(\sqrt{(\Phi_i - \Phi_j)^T V_j^{-1} (\Phi_i - \Phi_j)} + \sqrt{(\Phi_j - \Phi_i)^T V_i^{-1} (\Phi_j - \Phi_i)} \right)$$

Here, V represents the asymptotic covariance matrix of estimates. As demonstrated in Section 4.3.2, the distribution of $\hat{\Phi}$ is asymptotically normal with the mean of Φ and covariance matrix $\frac{1}{n}I(\Phi)^{-1}$. It is important to note that the division by n is necessary because the asymptotic covariance matrix represents the covariance matrix of $\sqrt{n}(\hat{\Phi} - \Phi)$, whereas we are interested in the covariance of $\hat{\Phi}$ (Newey & McFadden, 1994). Therefore, to approximate the asymptotic covariance matrix of MLE, we utilize the Hessian estimator, computing it as $V = I(\Phi)^{-1}$.

By applying the above procedure, we can construct the dissimilarity matrix for seizures within Patient ID18 by Mahalanobis distance using estimated parameters in the scalar coefficient model. Table 5.7 shows the seizure dissimilarity for covariance matrix data in $(\mathcal{S}^+(p), g^{\text{aff}})$. A smaller distance value represents less variation in seizure evolution. We can observe that Seizure 4 and Seizure 5 have highly similar evolution, while Seizure 3 is remarkably different from the others. This is also validated by the appearance of MDS plots in Figure 3.4b and estimated parameter plot in Figure 5.7.

Table 5.7: Seizure dissimilarity in $(\mathcal{S}^+(p), g^{\text{aff}})$ using covariance matrices for Patient ID18

	Seizure 1	Seizure 2	Seizure 3	Seizure 4
Seizure 2	16.329			
Seizure 3	31.750	39.380		
Seizure 4	22.195	14.883	50.572	
Seizure 5	26.518	18.603	50.498	9.529

Table 5.8 provides the seizure dissimilarity in the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$ using correlation matrices. Overall, Table 5.8 elucidates a different pattern of seizure dissimi-

larity for brain dynamics involving the relationship of pairs of channels compared to the covariance of channels in Table 5.7. Notably, Seizure 2 and Seizure 3 exhibit the closest similarity of brain dynamics with the lowest Mahalanobis distance.

Table 5.8: Seizure dissimilarity in $(\mathcal{C}^+(p), g^{\text{quo}})$ using correlation matrices for Patient ID18

	Seizure 1	Seizure 2	Seizure 3	Seizure 4
Seizure 2	25.952			
Seizure 3	13.180	7.993		
Seizure 4	12.958	35.761	19.284	
Seizure 5	13.250	25.001	11.292	12.471

5.3.2 Seizure variability between patients

Table 5.7 and Table 5.8 both provide insights into seizure variability, capturing the dynamics of covariance and linear relationships for pairs of channels, respectively. To assess the efficacy of measuring seizure variability, considering Patient ID6, ID7, ID13, and ID18 as examples to illustrate the variation among the total of 24 seizures, we compute a dissimilarity matrix for all these seizures based on estimated parameters from the manifold-adapted model using Mahalanobis distance. The results are depicted as scatter plots in Figure 5.9. These plots demonstrate that seizures within individual patients consistently cluster closely under both geometries using covariance and correlation matrices, despite the presence of some outliers. This observation aligns with findings from Burnham & Anderson (2004); Truccolo *et al.* (2011); Schindler *et al.* (2011); Karoly *et al.* (2018), suggesting that seizures in individual patients tend to share common features and evolve with a similar pattern of neural dynamics. It is notable that the arrangement of seizures for each fixed patient within the plots follows some time-ordering, suggesting that differences between seizures depend on the order in which they occur. Furthermore, lower percentages of stress value indicate a better fit to the original distances after MDS, defined in Equation 3.2. This suggests that epileptic seizure variability has been sensitively detected among patients.

To further illustrate seizure variability among patients, scatter plots of seizure dissimilarity for all patients have been measured in the affine invariant and quotient geometries separately (see Figure 5.10). In total, there are 112 seizures considered, taking into account the time series of covariance matrices in the affine invariant geometry on $\mathcal{S}^+(p)$ (shown in the left panel) and the time series of correlation matrices in the quotient geometry on $\mathcal{C}^+(p)$ (shown in the right panel).

The objective of this subsection is to illustrate the variability in seizures both within and across patients. Our manifold-adapted models reveal distinct groups clustered by patients, confirming clinical evidence that seizures within patients exhibit variability while

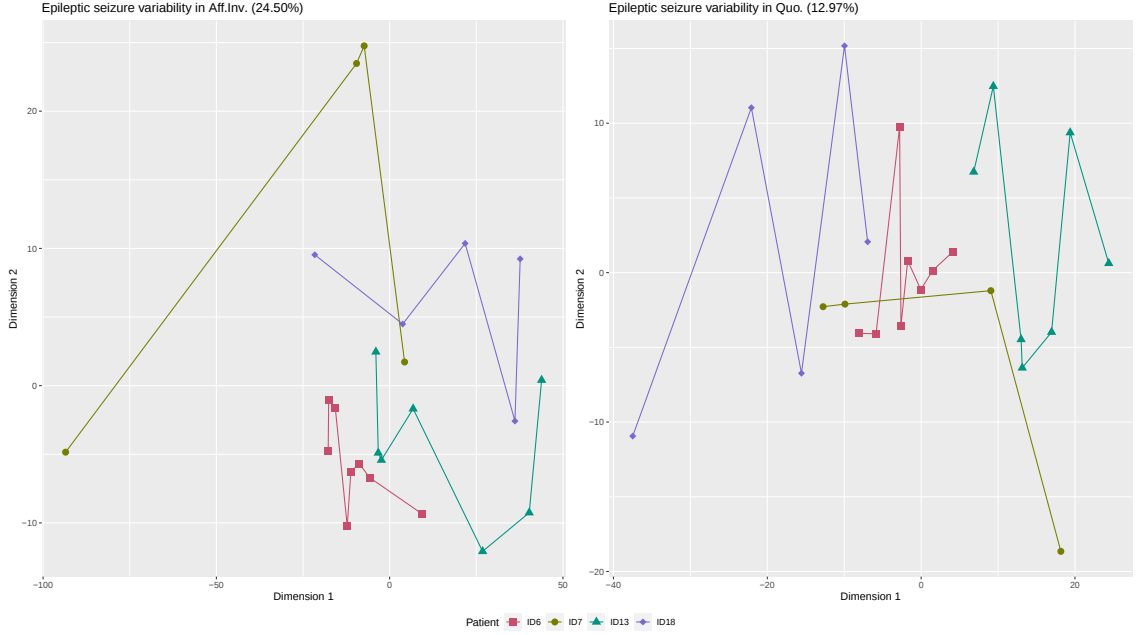


Figure 5.9: Metric MDS of Mahalanobis distances between seizures for Patient ID6, ID7, ID13 and ID18. Each epileptic patient, with a different number of seizures, is represented by a unique colour and shape. Points representing seizures for each patient are linked with a line representing the time ordering of seizure occurrence. The stress of the metric MDS is represented in the proportion shown in the subfigure titles. The left panel illustrates seizure variability in the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$ using covariance matrices, while the right panel displays seizure variability in the quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$.

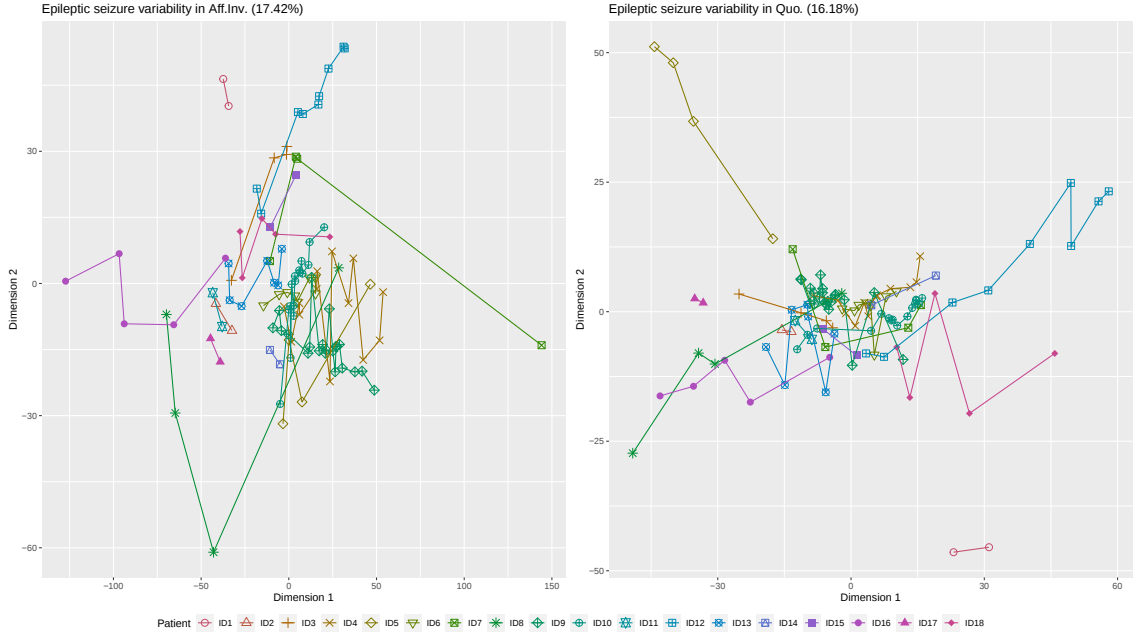


Figure 5.10: Seizure variability for all seizures among 18 patients. Each epileptic patient, with a different number of seizures, is represented by a unique colour and shape. More details could be found in caption of Figure 5.9.

also sharing common features or characteristics of brain dynamics. While it is possible to compute seizure dissimilarity using Mahalanobis distance with maximum likelihood estimates for two seizures across patients, such results may appear inconclusive in epileptic research, given the substantial variation observed in single seizures across different patients. Consequently, the subsequent subsection will focus on measuring patient variability, considering information from all seizures for each patient with epilepsy.

5.3.3 Patient variability

The placement of electrodes and their numbers vary across patients in EEG datasets, making it challenging to measure patient variability based on a single seizure. To address this issue, we propose a method called weighted maximum log-likelihood estimation (wMLE), which evaluates a single parameter vector for all seizures within a patient.

Suppose there are J seizures for a patient, with time lengths of $n_1, \dots, n_J$, separately. We define the weighted maximum log-likelihood function of the manifold-adapted model (4.3) for all seizures within this patient as:

$$\ell = \sum_{j=1}^J w_j \ell_j$$

where w_j denotes the weight assigned to the j th seizure, $j = 1, \dots, J$, which is calculated as the ratio of its time length n_j to the total length of all seizures, i.e., $w_j = n_j / (n_1 + \dots + n_J)$ and $\sum_{j=1}^J w_j = 1$ (Each patient has a limited number of seizures, as well as a limited duration for each seizure.). ℓ_j is the log-likelihood function for Seizure j . Subsequently, wMLE is implemented to estimate unknown parameters of α, β , and σ . The process of model inference is highly similar to the model inference of the scalar coefficient model in Section 4.3. Briefly, we can compute the first derivative of ℓ to obtain $\hat{\theta} = (\hat{\alpha}_1, \dots, \hat{\alpha}_L, \hat{\beta})^T$ and $\hat{\sigma}$ and let them be 0 as:

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha_l} &= \sum_{j=1}^J \omega_j \sum_{i=L+1}^{n_j} \mathbf{v}_{il}^{jT} (\mathbf{v}_i^j - \sum_{l=1}^L \alpha_l \mathbf{v}_{il}^j - \beta \mathbf{v}_{i0}^j) = 0 \\ \frac{\partial \ell}{\partial \beta} &= \sum_{j=1}^J \omega_j \sum_{i=L+1}^{n_j} \mathbf{v}_{i0}^{jT} (\mathbf{v}_i^{sej} - \sum_{l=1}^L \alpha_l \mathbf{v}_{il}^j - \beta \mathbf{v}_{i0}^j) = 0 \\ \frac{\partial \ell}{\partial \sigma} &= \sum_{j=1}^J \omega_j \left(\frac{1}{\sigma^2} \sum_{i=L+1}^{n_j} \left\| \mathbf{v}_i^j - \hat{\mathbf{v}}_i^j \right\|^2 - m(n_j - L) \right) = 0, \end{aligned}$$

where $\mathbf{v}_{il}^j$ and $\mathbf{v}_{i0}^j$ are coordinate vectors of autoregressive and mean-reverting terms as defined in Section 4.2 for the j th Seizure. Note that we do not fix an attract point for all seizures. That is, the attract points for each seizure still are the Fréchet sample mean of

their corresponding interictal data, and we compute the coordinate vectors $\mathbf{v}_0^j$ for the j th seizure. Subsequently, we can construct the matrix formulation with unknown parameter vectors $\theta = (\alpha_1, \dots, \alpha_L, \beta, \sigma)^T$ as:

$$\sum_{j=1}^J \omega_j \sum_{i=L+1}^{n_j} \begin{bmatrix} \|\mathbf{v}_{i1}\|^2 & \cdots & \mathbf{v}_{i1}^T \mathbf{v}_{iL} & \mathbf{v}_{i1}^T \mathbf{v}_{i0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{v}_{iL}^T \mathbf{v}_{i1} & \cdots & \|\mathbf{v}_{iL}\|^2 & \mathbf{v}_{iL}^T \mathbf{v}_i^* \\ \mathbf{v}_{i0}^T \mathbf{v}_{i1} & \cdots & \mathbf{v}_{i0}^T \mathbf{v}_{iL} & \|\mathbf{v}_{i0}\|^2 \end{bmatrix}^j \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_L \\ \beta \end{bmatrix} = \sum_{j=1}^J \omega_j \sum_{i=L+1}^{n_j} \begin{bmatrix} \mathbf{v}_{i1}^T \mathbf{v}_i \\ \vdots \\ \mathbf{v}_{iL}^T \mathbf{v}_i \\ \mathbf{v}_{i0}^T \mathbf{v}_i \end{bmatrix}^j$$

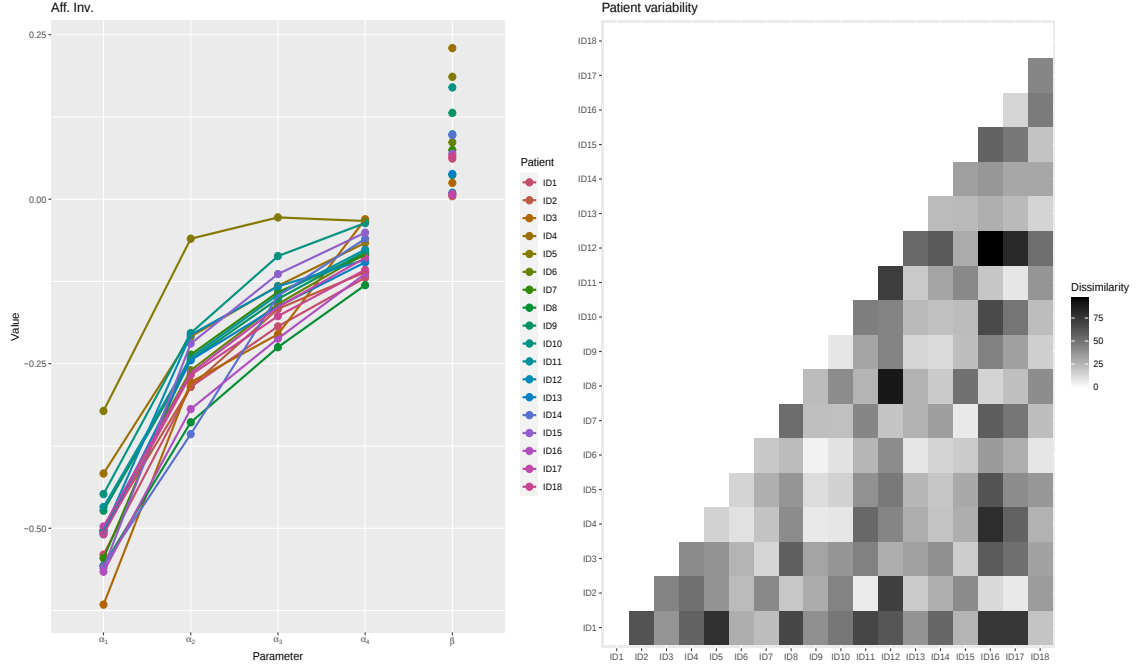
where $[\cdot]^j$ represents the data matrix of j th seizure, $j = 1, \dots, J$.

Utilizing weighted maximum log-likelihood estimation for all seizures within a patient, we compute dissimilarities of brain dynamics among patients. The left panel of Figure 5.11 displays the estimated parameters, while the right panel presents the heatmap of dissimilarities among patients. These results are demonstrated in two geometries, $(\mathcal{S}^+(p), g^{\text{aff}})$ and $(\mathcal{C}^+(p), g^{\text{quo}})$, positioned on the top and bottom, respectively.

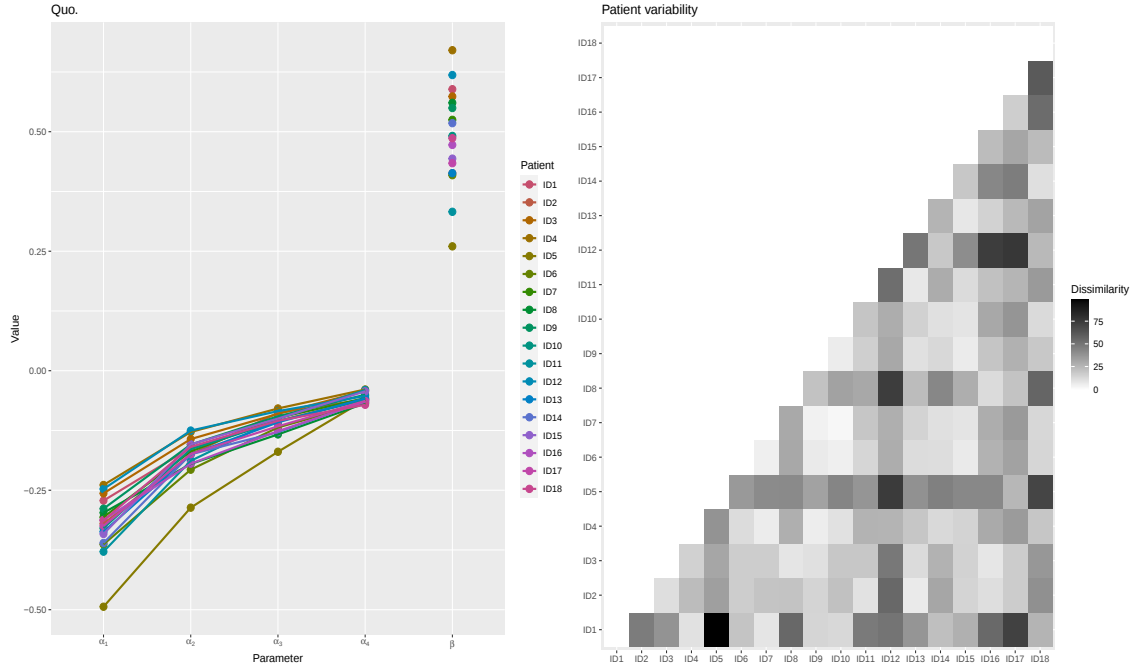
The analysis of Patient ID5 reveals noticeable differences from other patients in both geometries, $(\mathcal{S}^+(p), g^{\text{aff}})$ and $(\mathcal{C}^+(p), g^{\text{quo}})$. These distinctions are visually apparent, and upon further investigation, it is observed that all seizures of this patient have a very short time length, with a mean of 15 seconds for 4 seizures (see Table 3.1). Additionally, the estimated coefficient $\hat{\beta}$ in $(\mathcal{S}^+(p), g^{\text{aff}})$ is not as high as in $(\mathcal{C}^+(p), g^{\text{quo}})$. This difference is attributed to the high magnitude of variance of channels for covariance time series data, as further elaborated in the first subsection of Section 5.2.

However, despite these characteristics, the model still demonstrates efficacy, with the variation of $\hat{\beta}$ across patients ranging from 0 to 0.3. This pattern is particularly evident in the quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$, indicating a higher correlation for the current brain state with the interictal state concerning pairs of channels. Consequently, the influence of autoregressive terms has less impact on the current state compared to the covariance matrix data in the affine invariant geometry.

The heatmaps presented showcase patient variability, with the gradient colour from white to black. Larger and darker values indicate higher dissimilarity among patients. To provide a better visualization of patient variability, metric MDS plots for the dissimilarity distance matrix under both geometries are presented in Figure 5.12. Two scatter plots depict the relative distances among patients, reflecting the relative variability of seizures. The percentages in brackets indicate the stress values of the cost function for metric MDS, where a lower value indicates a better fit.

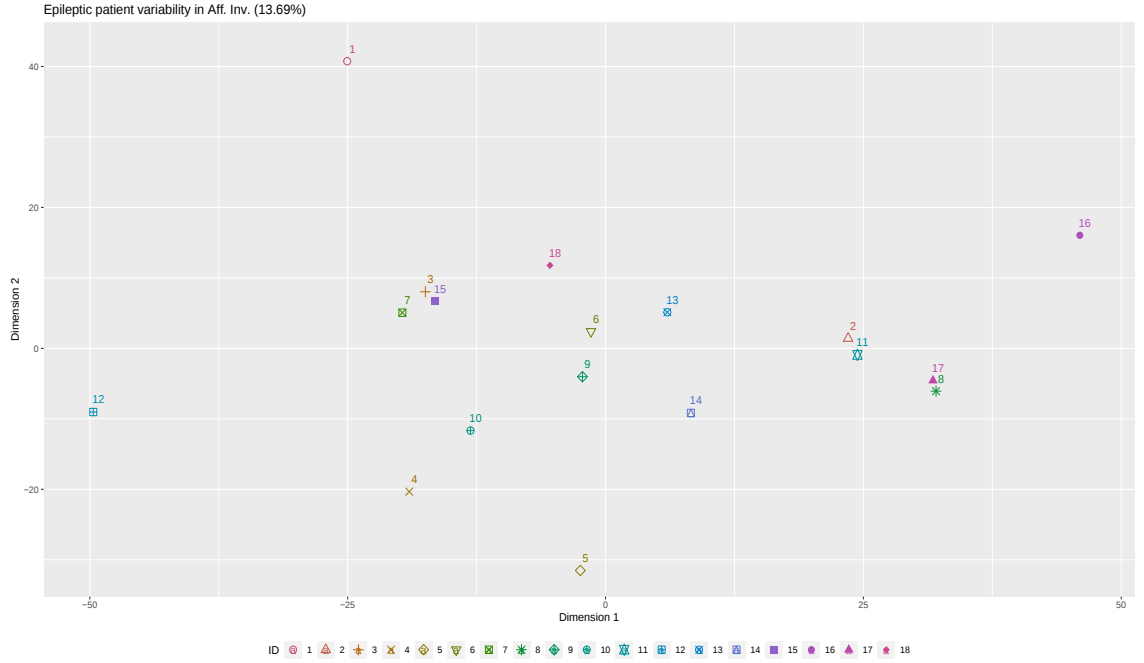


(a) Estimated parameters and patient dissimilarity in $\mathcal{S}^+(p)$ for covariance matrices in $(\mathcal{S}^+(p), g^{\text{aff}})$.

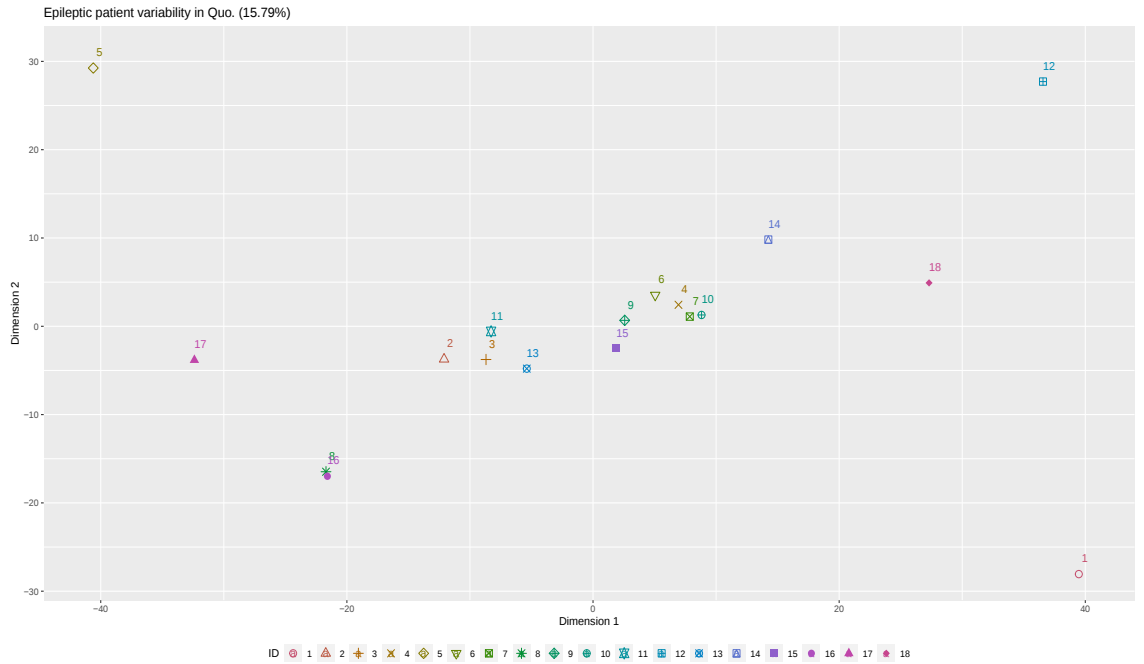


(b) Estimated parameters and patient dissimilarity in $\mathcal{C}^+(p)$ for covariance matrices in $(\mathcal{C}^+(p), g^{\text{quo}})$.

Figure 5.11: Estimated parameters (left panel) and patient dissimilarity (right panel) for 18 patients. Each epileptic patient is represented by a unique colour.



(a) Metric MDS of Mahalanobis distances between patients with respect to covariance matrices in $(\mathcal{S}^+(p), g^{\text{aff}})$.



(b) Metric MDS of Mahalanobis distances between patients with respect to correlation matrices in $(\mathcal{C}^+(p), g^{\text{quo}})$.

Figure 5.12: Metric MDS of Mahalanobis distances between patients in different geometries using covariance and correlation matrices, separately. Each epileptic patient, with a different number of seizures, is represented by a unique colour and shape. The percentages in brackets represent stress values from metric MDS plots.

5.3.4 Summary

It is important to emphasize to readers that the process of generating reduced covariance matrices involves using PCA methods. This results in reduced data that captures the most significant proportion of the original covariance matrix. On the other hand, when working with reduced correlation matrices, the selection of channel combinations aims to maximally reduce the redundancy of the linear relationship between channels. Regardless of the specific dimensions chosen in our model, variations in results are inevitable, potentially influencing the values of seizure dissimilarity. Therefore, the seizure variability presented in our analysis serves as a recommendation for neuroscientists, providing insights into the potential impact of different manifolds with different reduced data representations on the results.

Chapter 6

Discussion

6.1 Summary

In this research, we conducted a comprehensive analysis of manifold-valued time series data, exploring different manifolds, including the symmetric Euclidean manifold $(\text{Sym}(p), g^{\text{euc}})$, the affine invariant manifold $(\mathcal{S}^+(p), g^{\text{aff}})$, and the quotient manifold $(\mathcal{C}^+(p), g^{\text{quo}})$. Our approach involved introducing manifold-adapted models to analyse matrix-valued time series data with the aim of revealing the brain dynamics associated with seizures in epileptic patients. We delved into the examination of seizure variability within patients and epileptic patient variability by fitting the manifold-adapted models to covariance and correlation matrices in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$. Importantly, the general nature of our model formulation allows for its application to data on other Riemannian manifolds. It also provides the basic tools for further geometrical statistics with applications to other fields.

Given that electromagnetic artefacts and high correlations between sets of channels can result in rank deficiencies in the observed matrix-valued time series, we proposed two different approaches to dimensional reduction: (i) capturing the maximum variance of the reduced data set and (ii) substituting highly correlated channels with a single channel. These approaches ensure that the reduced covariance and correlation matrices are strictly positive definite, lying on certain smooth manifolds with a natural Riemannian structure.

Inspired by the widespread use of non-Euclidean statistical tools, this research introduced geometrical operations for statistical analysis in the affine invariant geometry on $(\mathcal{S}^+(p), g^{\text{aff}})$ and the quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$, which is a newly developed manifold (David, 2019; Thanwerdas & Pennec, 2021). Therefore, our research constructed closed formulas for the various operations we require to specify our model, including Riemannian metrics, logarithm and exponential maps, parallel transports, orthonormal coordinate systems, and Fréchet sample means.

To capture different dynamics for EEG matrix-valued data intrinsically, we modelled the direction vector as the sum of three terms: an autoregressive term, a mean-reverting mean, and a noise term. This manifold-adapted additive model was able to capture various possible dynamics for the evolution of EEG data, from smooth flow along geodesics to a noisy mean-reverting random walk on the underlying manifolds.

Considering the potentially large number of parameters required by our model, we simplified the manifold-adapted model to two simpler models with fewer parameters: the scalar coefficient model and the diagonal model. Both models could uncover the brain dynamics effectively according to the model diagnostics. Comparing the scalar coefficient and diagonal models with respect to covariance and correlation matrices separately suggested that the diagonal model had slightly better performance than the scalar coefficient model. However, the scalar coefficient model is more flexible and straightforward to interpret the brain dynamics with a desirable goodness of model fit. Specifically, when the manifold is not parallelizable, the diagonal model is ill-defined. The scalar coefficient model generates general strength values that describe the tendency of lagged vectors and mean reversion term to the current brain state. When we investigated the detailed signal evolution of pairs of channels, the diagonal model is the best choice.

Comparing the model results on different geometries, it became evident that the affine invariant geometry exhibited superior performance compared to Euclidean space when examining covariance matrices. The goodness of modelling fit significantly improved in $(\mathcal{S}^+(p), g^{\text{aff}})$, benefiting from the affine invariant transformation for high-magnitude covariance matrices. However, upon analysing correlation matrices, the Euclidean and affine invariant geometries yielded similar results to the quotient geometries, as they all relied on scaled data. Notably, a higher goodness of fit was still observed in the quotient geometry. Furthermore, the dimensions of the affine invariant geometry differed from other geometries, as they still accounted for the values of the diagonal elements in the correlation matrix.

Maximum likelihood estimators and their asymptotic distributions were used to gauge seizure and patient variability. By mapping the dissimilarity matrices of seizures across patients, we verified that there are variations among seizures within the patients and the seizures within patient share some common features of brain dynamics (Burnham & Anderson, 2004; Truccolo *et al.*, 2011; Schindler *et al.*, 2011; Karoly *et al.*, 2018).

Another significant contribution of this work is that we developed an R Package: `geomTS`. The package includes all operations in Riemannian manifolds, $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$, manifold-adapted models, and model inference. It will be continuously updated.

In summary, our research delved into the analysis of matrix-valued time series of EEG data from epileptic patients, utilizing manifold-adapted models to describe the brain dy-

namical patterns for each pair of channels. The estimated model coefficients shed light on the influence of historical brain states on current brain states and the proximity of the seizure period to the interictal period. By quantifying these strengths, our research effectively characterized seizure and patient variability in the evolution of epileptic networks.

It is important to emphasize to readers that the process of generating reduced covariance matrices involves using PCA methods. This results in reduced data that captures the most significant proportion of the original covariance matrix. On the other hand, when working with reduced correlation matrices, the selection of channel combinations aims to maximally reduce the redundancy of the linear relationship between channels. Regardless of the specific dimensions chosen in our model, variations in results are inevitable, potentially influencing the values of seizure dissimilarity. Therefore, the seizure variability presented in our analysis serves as a recommendation for neuroscientists, providing insights into the potential impact of different manifolds with different reduced data representations on the results.

6.2 Limitations of the study

The manifold-adapted model, though flexible and applicable to various types of manifold-valued time series data, possesses inherent limitations that should be acknowledged, as they may influence the final results.

Upon downloading the EEG dataset from open sources, determining the optimal sliding window size becomes crucial. Currently, there are no specific criteria for choosing the best window size and overlaps, and the various combinations of these factors may impact the model results (Zhuang *et al.*, 2020). For detecting faster oscillations and avoiding large size of data, a sliding window of 1 second without overlap was chosen in this research. Literally, we also employed different window sizes (i.e., 0.5s and 2s) to fit the manifold-adapted model and evaluate estimator variability. However, the main conclusions drawn from the model results remained consistent across these variations.

Two approaches to dimensional reduction, based on maximizing variation and minimizing redundancy, were explored. The first method involves performing PCA on the average covariance of signals over time, resulting in the dimensions of the reduced covariance matrix losing the positions of electrodes. Consequently, for the diagonal coefficients model, the estimated parameters for each pair of channels are interpretable. Conversely, the second method of dimensional reduction avoids this problem by identifying the most effective and exclusive combinations of channels. However, this approach loses some information for discarded channels, potentially leading to large variation in model results. Importantly, the final choice of dimensions remains an open question, requiring careful consideration to balance model efficiency and computation resources.

Currently, there is no mathematical solution to find the optimal position along the fiber concerning any point in $(\mathcal{S}^+(p), g^{\text{aff}})$. Although David (2019) proposed a gradient descent algorithm to find it numerically, variations in the optimization process may impact the precision of the operations in the quotient manifold, including logarithm map, parallel transport, and orthonormal coordinate system. Specifically, optimal positions for each of these operations in the quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$ need to be determined, and variations in these increments may lead the model to insensitively capture the data evolution, as evident in simulation studies on the quotient geometry.

The uniqueness of the Fréchet sample mean in the quotient manifold remains an unsolved problem. Lueg (2023) was the first to discuss this issue, but a definitive solution is still elusive. This research assumed that the Fréchet sample mean in the quotient manifold is unique and computed by submerging the Fréchet sample mean in $\mathcal{S}^+(p)$, determined by all optimal positions concerning the iterated mean $\tilde{S}_k$ in $(\mathcal{S}^+(p), g^{\text{aff}})$. However, the uniqueness of the Fréchet sample mean is not mandatory, as the attract point in the model does not have to be the Fréchet sample mean. Since the interictal time series is a random walk at the centre point, any Fréchet sample mean of correlation matrices of interictal data could serve as the attract point. Intuitively, we believe that the Fréchet means will be close to each other.

Some dimensions of residuals from the quotient manifold-adapted model against normality with zero mean and estimated variance may be attributed to the numerical optimization of optimal positions in $(\mathcal{S}^+(p), g^{\text{aff}})$. Upon checking these residuals, it is evident that they are nearly normally distributed, with means approaching 0 and variance approaching $\hat{\sigma}^2$. Unfortunately, in this research, the proposed quotient manifold-adapted model was not optimized to enhance model precision, which would have better aligned the model's assumptions with the requirement of normality.

6.3 Future Research

The aforementioned limitations offer valuable insights for future research. Although widely used to analyse EEG data, the use of a sliding window to derive time series of covariance or correlation matrices is problematic in that the choice of the length of the window affects the dynamics which can be revealed by such data. A more consistent approach to modelling might be to adopt a continuous time model for the underlying covariance process, like the mean reverting diffusion developed by Bui *et al.* (2023), with the raw EEG signals z_i arising as discrete-time observations from a Gaussian with the underlying covariance process. Nonetheless, incorporating autoregressive or other components suitable for modelling seizures into such a continuous time model could be challenging.

Given the absence of a general expectation for the uniqueness of the intrinsic mean

within $(\mathcal{C}^+(p), g^{\text{quo}})$, Huckemann introduced the concept of the *principal component mean* in Huckemann & Ziezold (2006) and Huckemann *et al.* (2010). Drawing inspiration from PCA on shape spaces, this concept utilizes generalized geodesics. Extending this methodology to the quotient manifold, we can compute the principal component mean as the attract point. Furthermore, refining the gradient descent algorithm for optimizing optimal positions could enhance the precision of Riemannian operations in the quotient geometry on $(\mathcal{C}^+(p), g^{\text{quo}})$.

To explore the EEG dynamical network for epileptic patients more comprehensively, improvements to our manifold-adapted model are necessary for better data fitting, especially concerning correlation matrix-valued time series in the quotient manifold.

Lastly, we have developed geometric statistical tools for analysing Riemannian manifold time series data. Future research could delve into more applications of such data on different geometries, exploring various perspectives of geometrical ideas. Non-Euclidean statistics will be a primary focus in the near future.

Appendix A

Model results for Patient ID6, ID7, and ID13

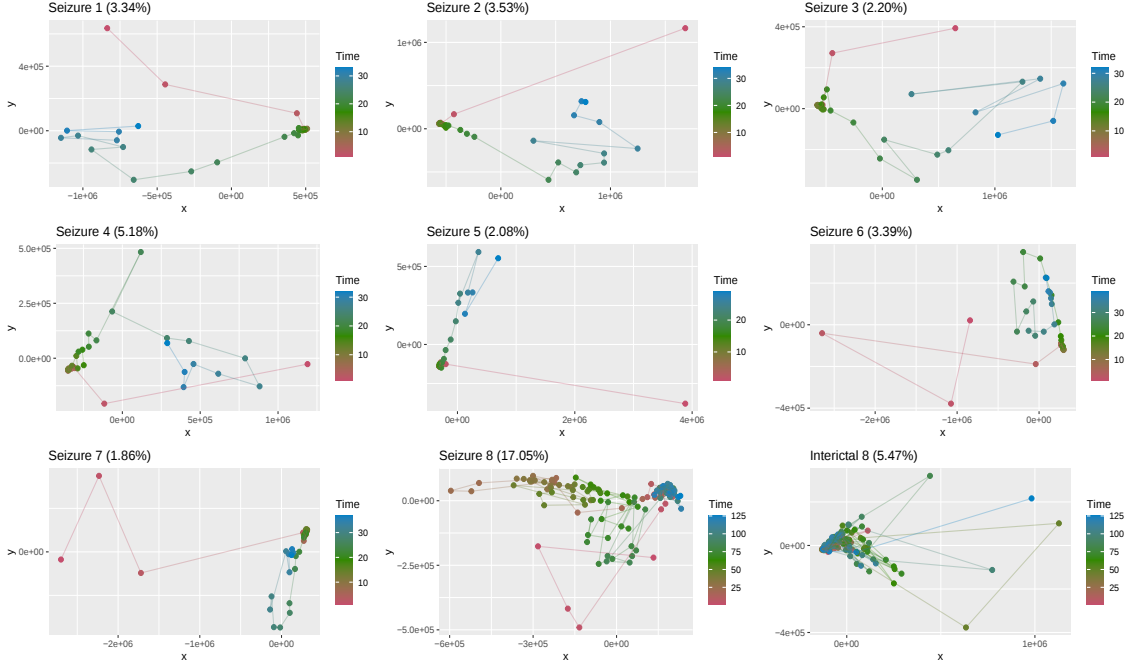
In this appendix, we present MDS plots and scalar coefficient model results for Patient ID6, ID7, and ID13 separately. We utilize the time series of covariance matrices in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$, as well as the time series of correlation matrices in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$.

For each patient, we represent the MDS plots for covariance matrices in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$, followed by the estimated parameters in both geometries. Lastly, the goodness of modelling fit by R^2 is shown in the table.

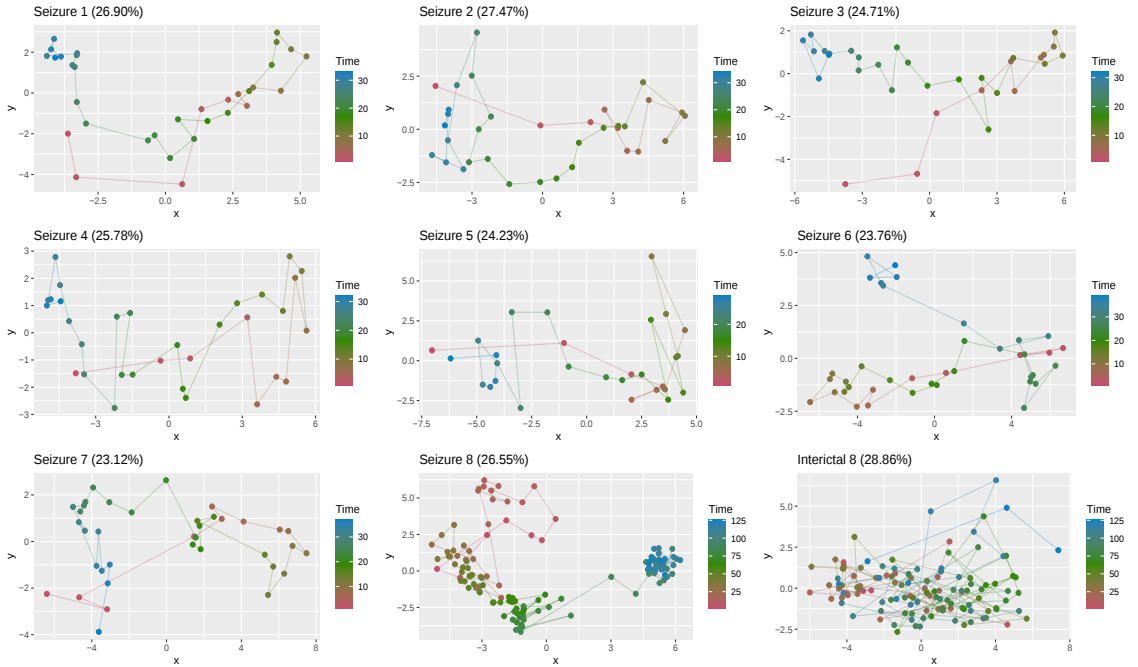
Analogously, MDS plots, estimated parameters, and R^2 values are presented for correlation matrices in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$.

A.1 Patient ID6

A.1.1 Covariance matrices in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$



(a) MDS using the Euclidean metric $d_{g^{\text{euc}}}$.



(b) MDS using the affine invariant metric $d_{g^{\text{aff}}}$.

Figure A.1: MDS plots for Patient ID6 using the Euclidean and affine invariant metrics.

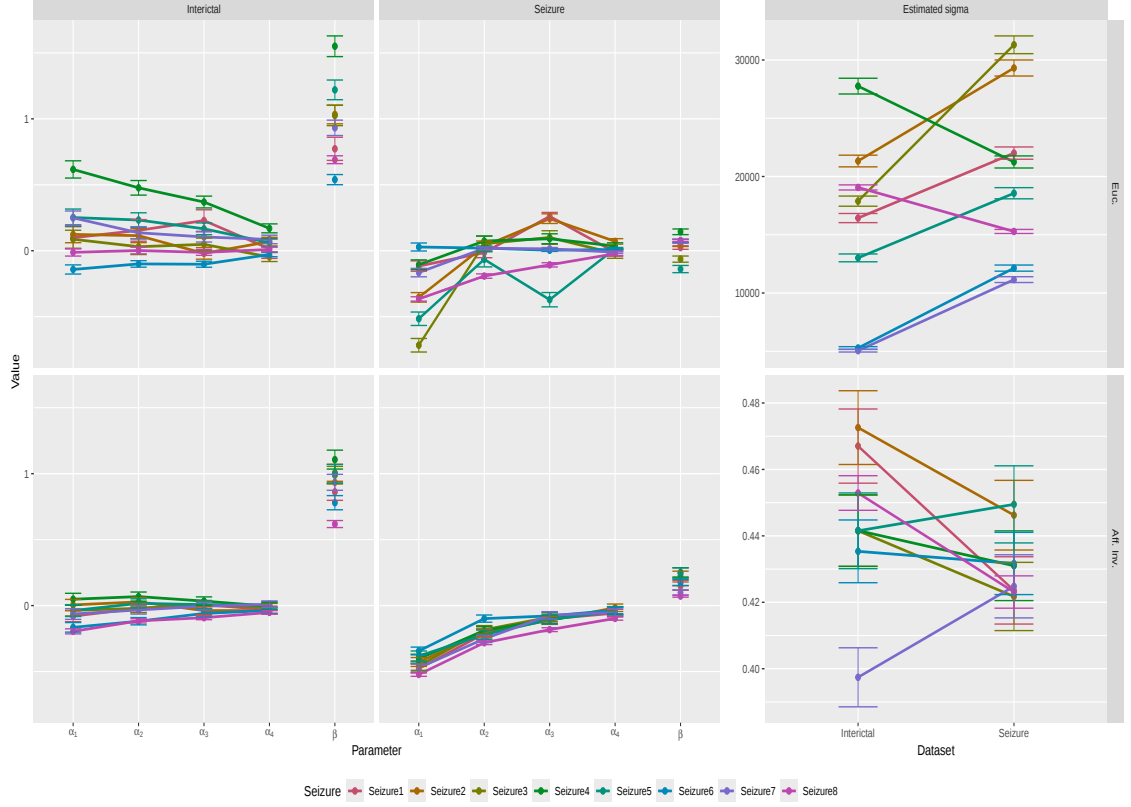
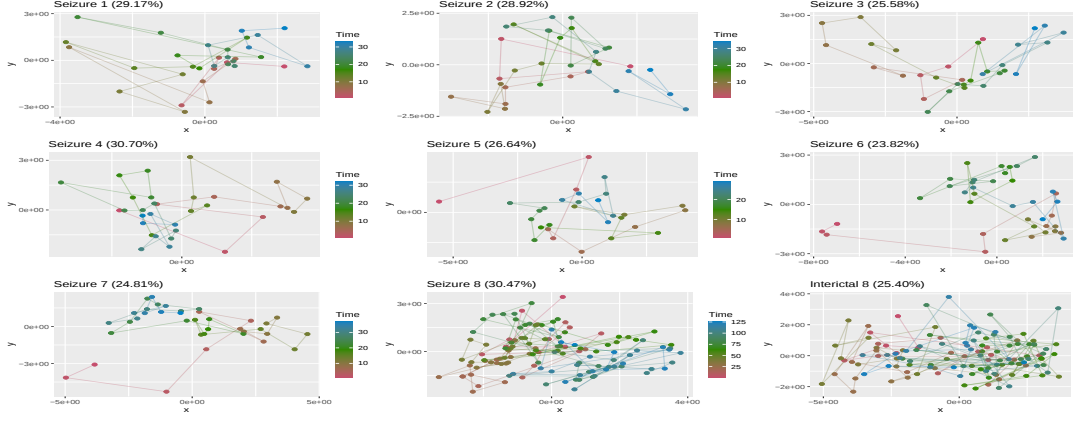


Figure A.2: Estimated parameters for Patient ID6 in $(\text{Sym}(p), g^{\text{euc}})$ (left panel) and $(\mathcal{S}^+(p), g^{\text{aff}})$ (right panel)

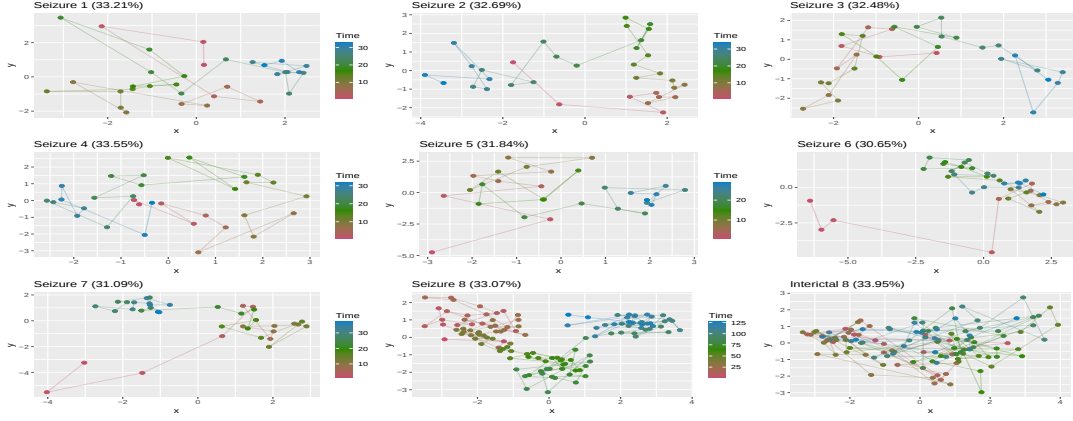
Table A.1: Goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$ for Patient ID6

	Seizure 1		Seizure 2		Seizure 3		Seizure 4		Seizure 5		Seizure 6		Seizure 7		Seizure 8		Interictal 8	
	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.
$\text{Sym}(p)$	0.081	0.293	0.170	0.287	0.390	0.472	0.079	0.339	0.187	0.445	0.025	0.154	0.066	0.249	0.179	0.266	0.301	0.320
$\mathcal{S}^+(p)$	0.356	0.506	0.389	0.511	0.376	0.511	0.388	0.523	0.355	0.504	0.271	0.417	0.323	0.462	0.391	0.443	0.447	0.489

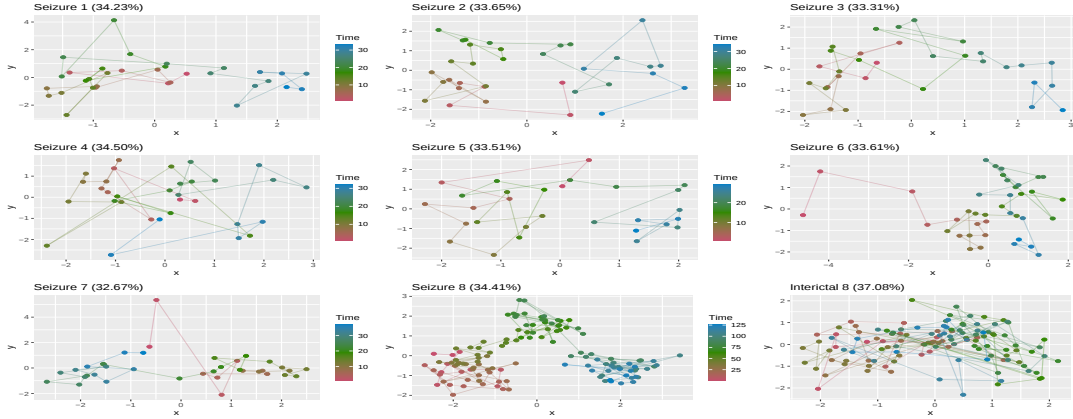
A.1.2 Correlation matrices in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$



(a) MDS using the Euclidean metric $d_{g^{\text{euc}}}$.



(b) MDS using the affine invariant metric $d_{g^{\text{aff}}}$.



(c) MDS using the quotient metric $d_{g^{\text{quo}}}$.

Figure A.3: MDS plots for Patient ID6 using the Euclidean, affine invariant and quotient metrics.

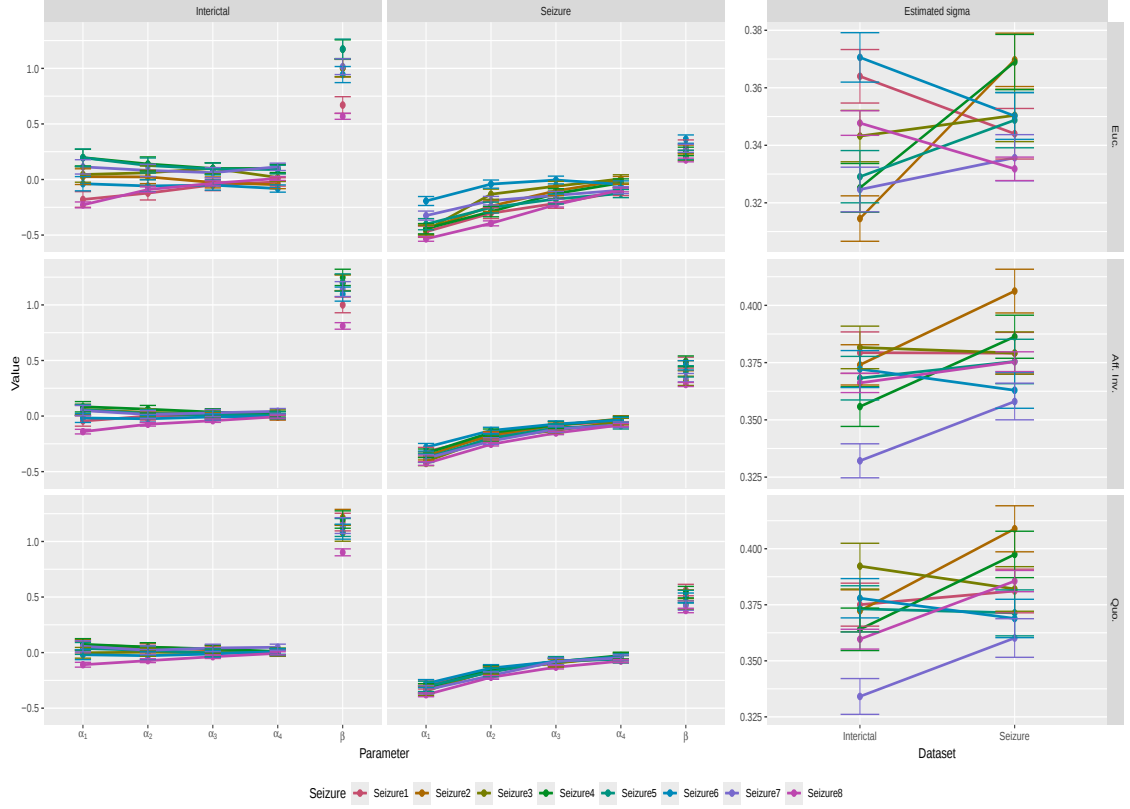


Figure A.4: Estimated parameters for Patient ID6 in $(\text{Sym}(p), g^{\text{euc}})$ (left panel), $(\mathcal{S}^+(p), g^{\text{aff}})$ (middle panel), and $(\mathcal{C}^+(p), g^{\text{quo}})$ (right panel).

Table A.2: Goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$ for Patient ID6.

	Seizure 1		Seizure 2		Seizure 3		Seizure 4		Seizure 5		Seizure 6		Seizure 7		Seizure 8		Interictal 8	
	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.
$\text{Sym}(p)$	0.381	0.474	0.337	0.456	0.311	0.422	0.329	0.445	0.291	0.432	0.255	0.395	0.286	0.419	0.341	0.386	0.391	0.414
$\mathcal{S}^+(p)$	0.472	0.575	0.480	0.590	0.431	0.542	0.488	0.594	0.467	0.596	0.415	0.530	0.432	0.547	0.459	0.503	0.518	0.547
$\mathcal{C}^+(p)$	0.484	0.582	0.488	0.594	0.441	0.546	0.482	0.606	0.458	0.592	0.423	0.544	0.434	0.541	0.466	0.512	0.539	0.561

A.2 Patient ID7

A.2.1 Covariance matrices in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$



Figure A.5: MDS plots for Patient ID7 using the Euclidean and affine invariant metrics.

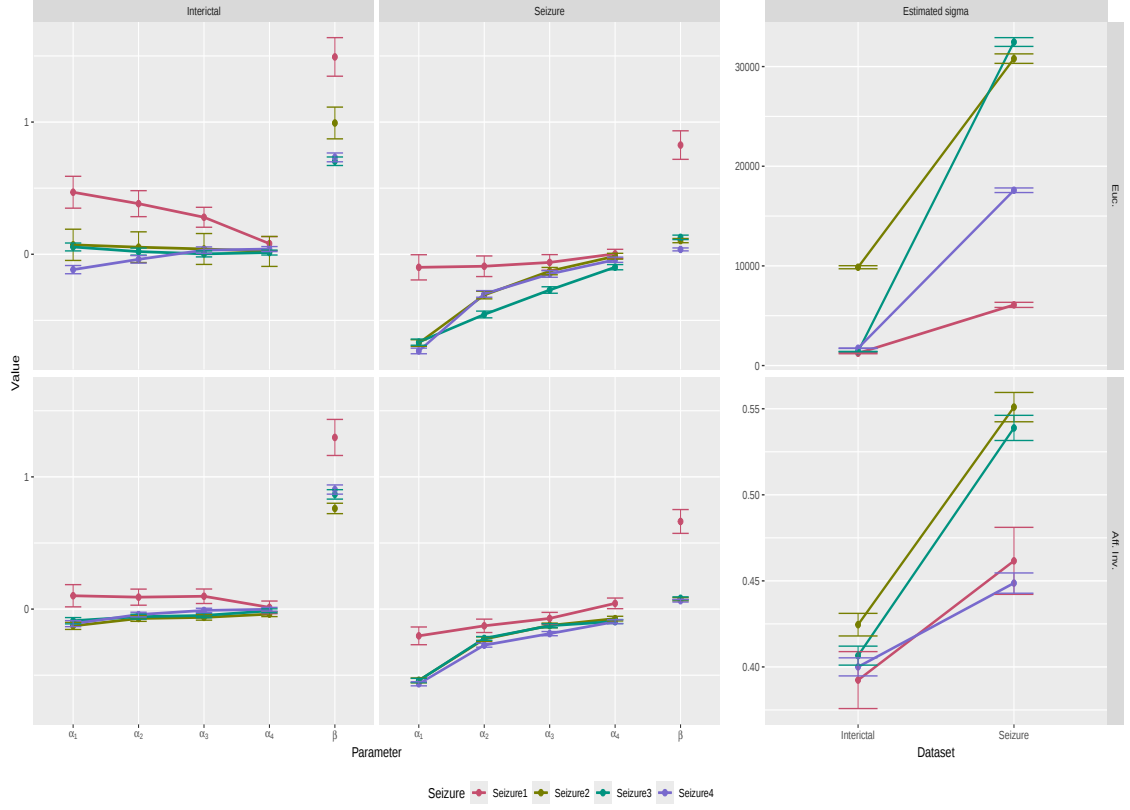
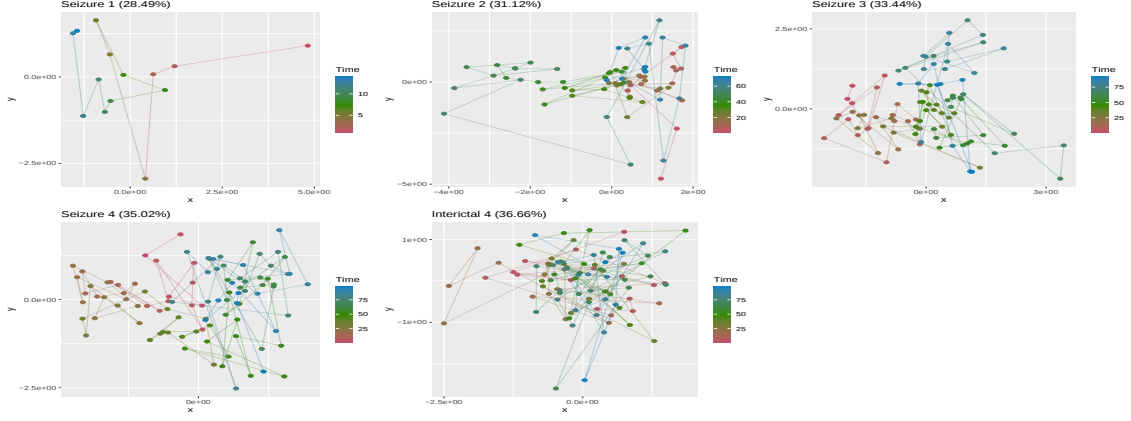


Figure A.6: Estimated parameters for Patient ID7 in $(\text{Sym}(p), g^{\text{euc}})$ (left panel) and $(\mathcal{S}^+(p), g^{\text{aff}})$ (right panel).

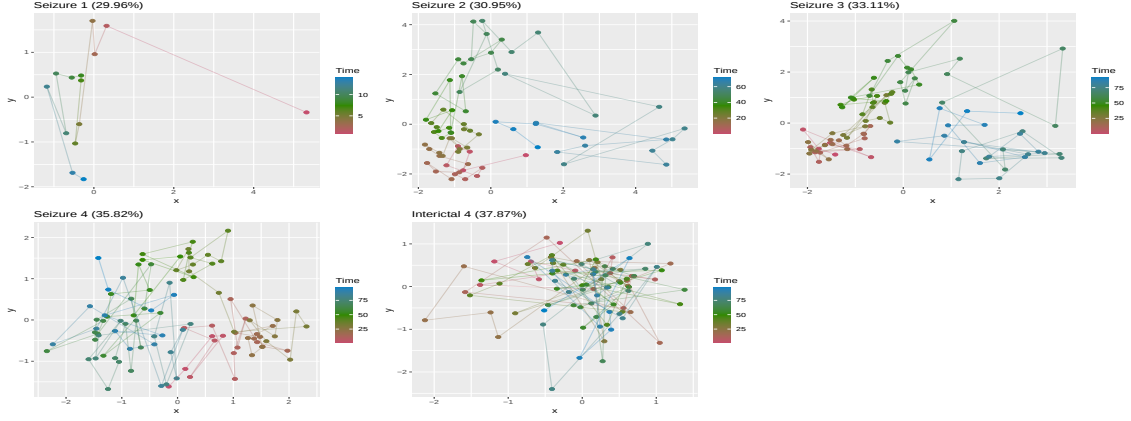
Table A.3: Goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$ for Patient ID7

	Seizure 1		Seizure 2		Seizure 3		Seizure 4		Interictal 4	
	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.
$\text{Sym}(p)$	0.467	0.642	0.389	0.453	0.385	0.464	0.387	0.472	0.420	0.469
$\mathcal{S}^+(p)$	0.492	0.762	0.466	0.545	0.493	0.571	0.453	0.522	0.551	0.583

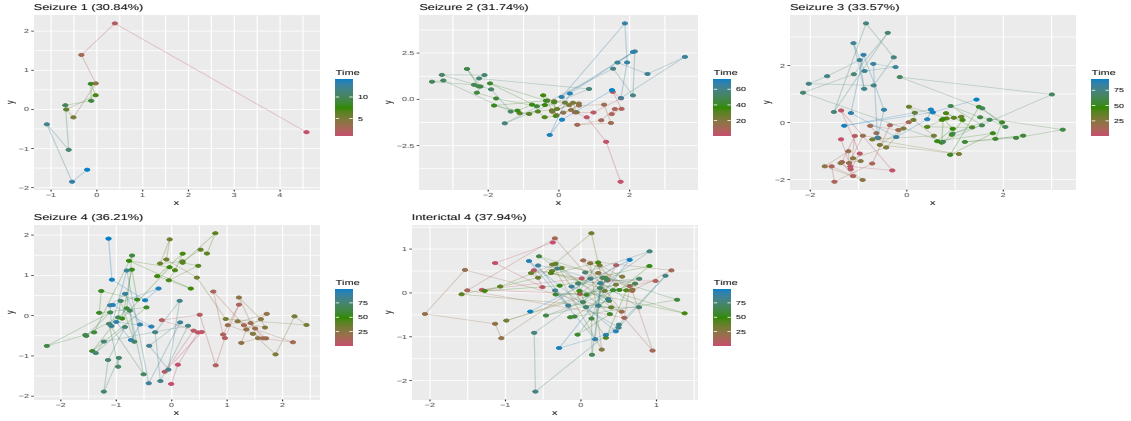
A.2.2 Correlation matrices in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$



(a) MDS using the Euclidean metric $d_{g^{\text{euc}}}$.



(b) MDS using the affine invariant metric $d_{g^{\text{aff}}}$.



(c) MDS using the quotient metric $d_{g^{\text{quo}}}$.

Figure A.7: MDS plots for Patient ID7 using the Euclidean, affine invariant and quotient metrics.

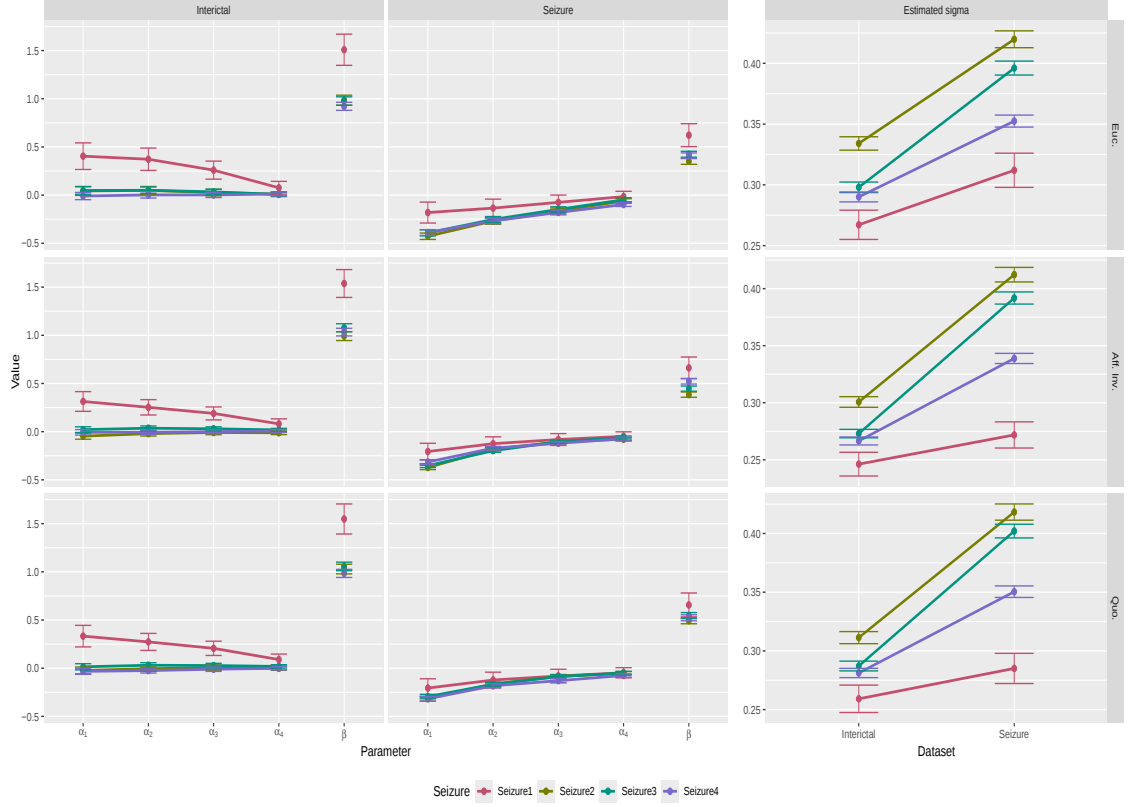


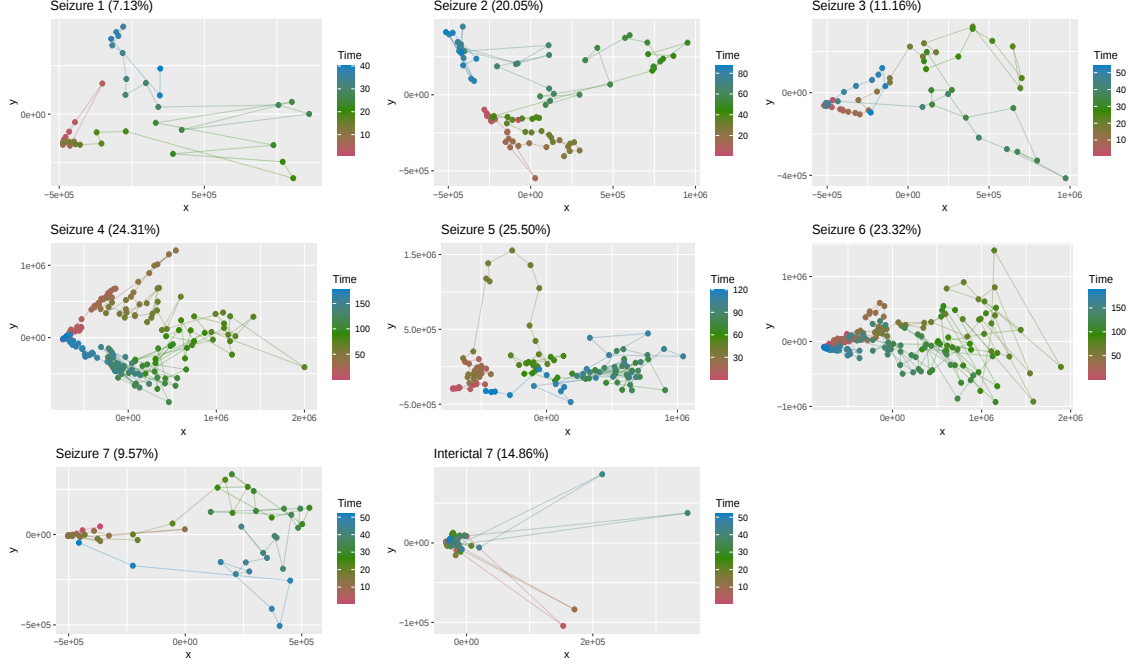
Figure A.8: Estimated parameters for Patient ID7 in $(\text{Sym}(p), g^{\text{euc}})$ (left panel), $(\mathcal{S}^+(p), g^{\text{aff}})$ (middle panel), and $(\mathcal{C}^+(p), g^{\text{quo}})$ (right panel).

Table A.4: Goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$ for Patient ID7.

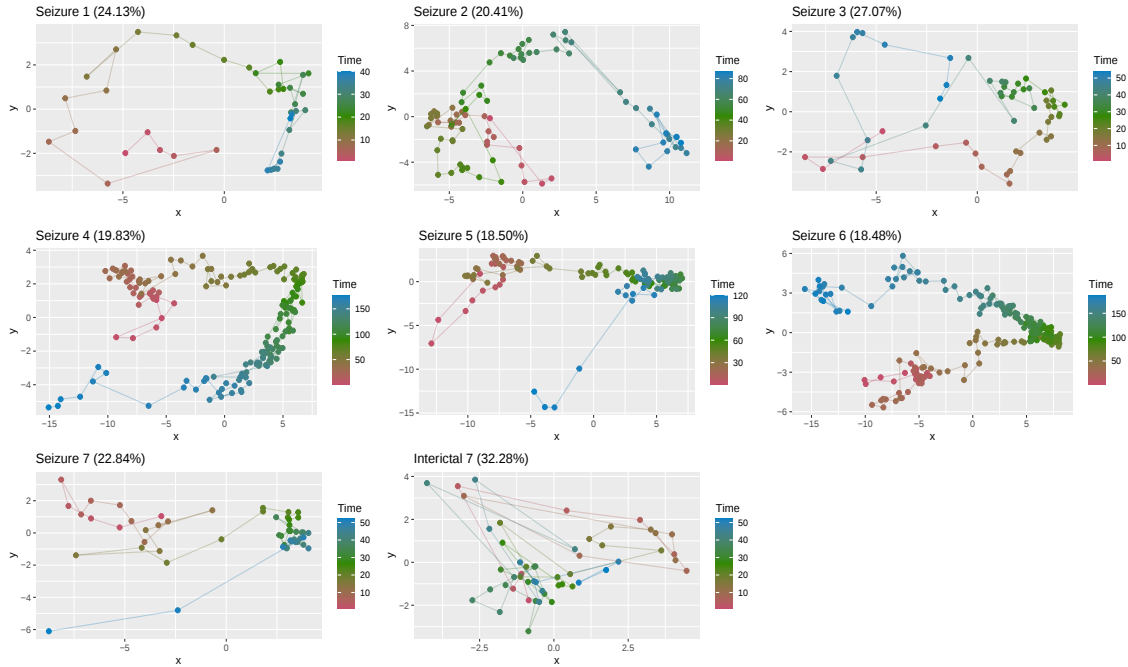
	Seizure 1		Seizure 2		Seizure 3		Seizure 4		Interictal 4	
	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.
$\text{Sym}(p)$	0.380	0.669	0.375	0.448	0.396	0.456	0.389	0.442	0.464	0.492
$\mathcal{S}^+(p)$	0.454	0.720	0.501	0.577	0.519	0.575	0.500	0.544	0.540	0.563
$\mathcal{C}^+(p)$	0.445	0.724	0.508	0.578	0.510	0.561	0.496	0.540	0.530	0.554

A.3 Patient ID13

A.3.1 Covariance matrices in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$



(a) MDS using Euclidean metric $d_{g^{\text{euc}}}$.



(b) MDS using affine invariant metric $d_{g^{\text{aff}}}$.

Figure A.9: MDS plots for Patient ID13 using the Euclidean and affine invariant metrics.

Appendix A. Model results for Patient ID6, ID7, and ID13

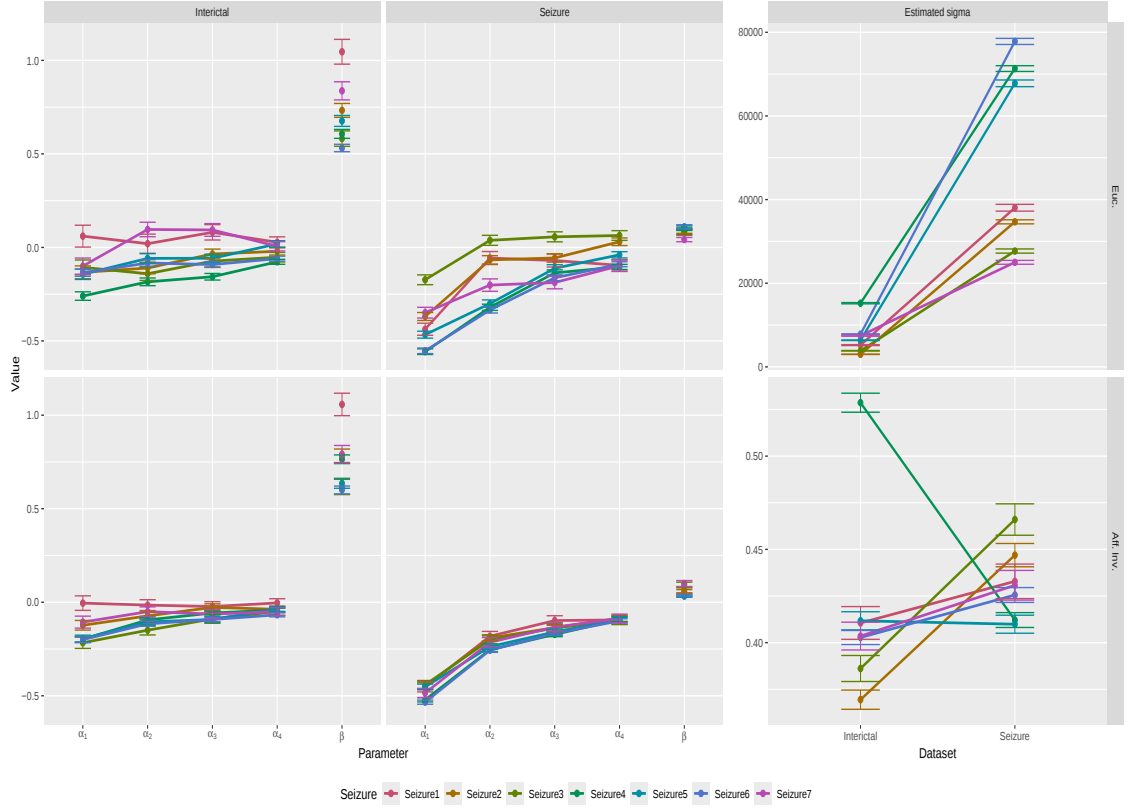
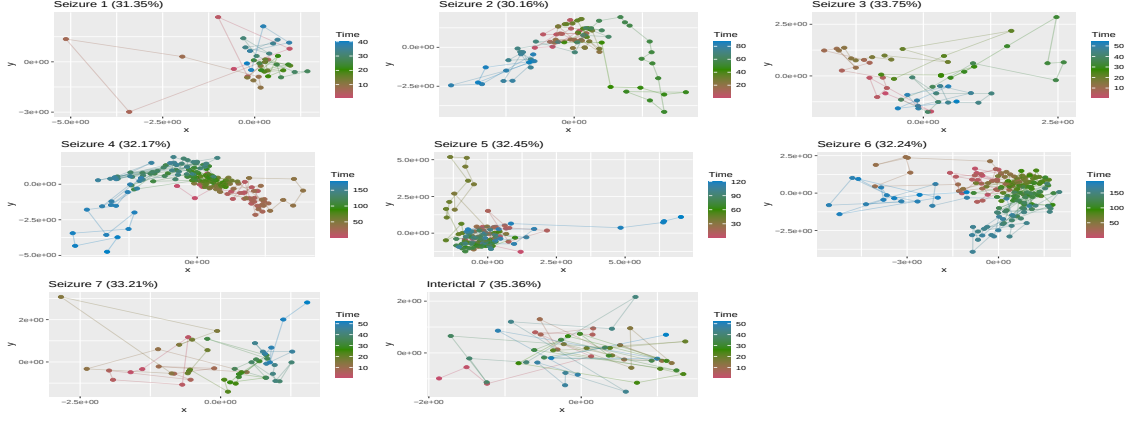


Figure A.10: Estimated parameters for Patient ID13 in $(\text{Sym}(p), g^{\text{euc}})$ (left panel) and $(\mathcal{S}^+(p), g^{\text{aff}})$ (right panel).

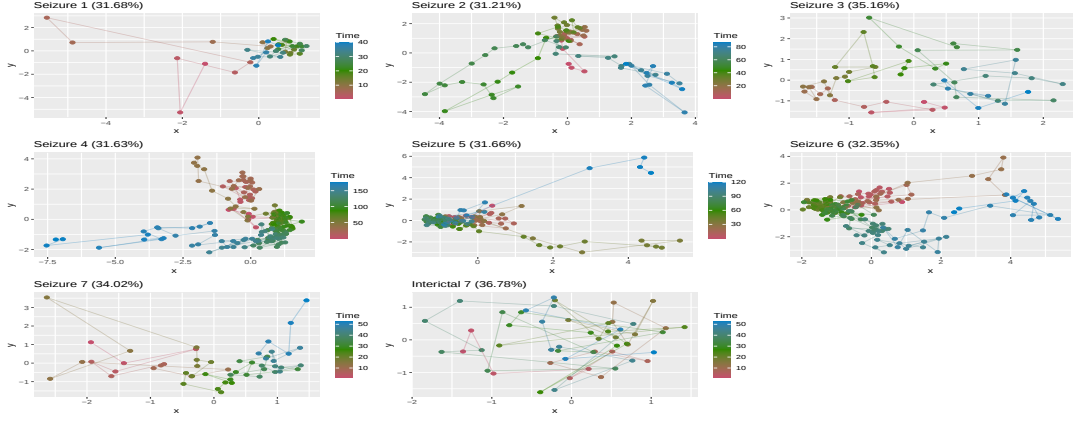
Table A.5: Goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$ and $(\mathcal{S}^+(p), g^{\text{aff}})$ for Patient ID13.

	Seizure 1		Seizure 2		Seizure 3		Seizure 4		Seizure 5		Seizure 6		Seizure 7		Interictal 7	
	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.
$\text{Sym}(p)$	0.244	0.292	0.187	0.282	0.080	0.333	0.310	0.395	0.257	0.352	0.308	0.377	0.136	0.235	0.486	0.520
$\mathcal{S}^+(p)$	0.316	0.447	0.312	0.372	0.351	0.482	0.367	0.420	0.282	0.367	0.392	0.443	0.335	0.439	0.503	0.562

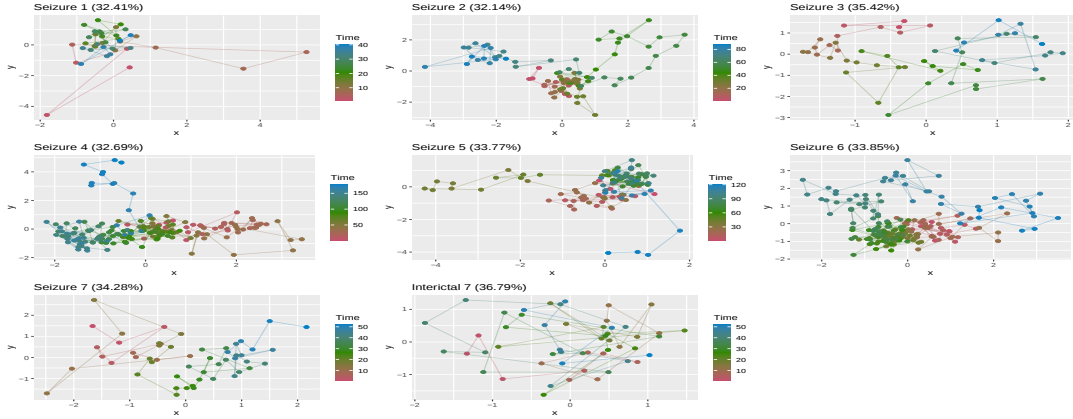
A.3.2 Correlation matrices in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$



(a) MDS using the Euclidean metric $d_{g^{\text{euc}}}$.



(b) MDS using the affine invariant metric $d_{g^{\text{aff}}}$.



(c) MDS using the quotient metric $d_{g^{\text{quo}}}$.

Figure A.11: MDS plots for Patient ID13 using the Euclidean, affine invariant and quotient metrics.

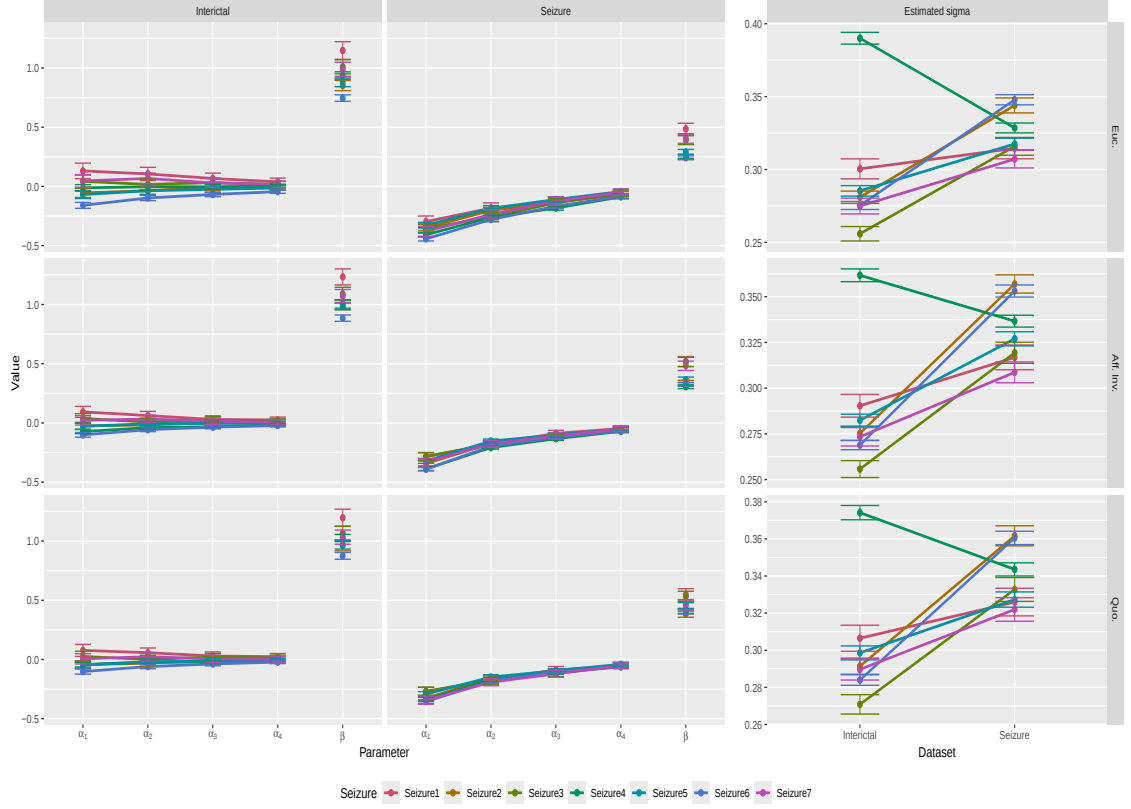


Figure A.12: Estimated parameters for Patient ID13 in $(\text{Sym}(p), g^{\text{euc}})$ (left panel), $(\mathcal{S}^+(p), g^{\text{aff}})$ (middle panel), and $(\mathcal{C}^+(p), g^{\text{quo}})$ (right panel).

Table A.6: Goodness of modelling fit by R^2 in $(\text{Sym}(p), g^{\text{euc}})$, $(\mathcal{S}^+(p), g^{\text{aff}})$, and $(\mathcal{C}^+(p), g^{\text{quo}})$ for Patient ID13.

	Seizure 1		Seizure 2		Seizure 3		Seizure 4		Seizure 5		Seizure 6		Seizure 7		Seizure 7	
	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.	Sca.	Diag.
$\text{Sym}(p)$	0.390	0.491	0.273	0.327	0.361	0.433	0.311	0.345	0.281	0.329	0.322	0.363	0.379	0.460	0.473	0.525
$\mathcal{S}^+(p)$	0.480	0.591	0.409	0.465	0.463	0.523	0.423	0.465	0.386	0.445	0.449	0.495	0.464	0.537	0.541	0.585
$\mathcal{C}^+(p)$	0.482	0.591	0.417	0.468	0.457	0.515	0.439	0.473	0.424	0.466	0.460	0.497	0.460	0.534	0.531	0.574

Bibliography

- ARSIGNY, V., FILLARD, P., PENNEC, X. & AYACHE, N. 2006 Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **56** (2), 411–421.
- BADAWY, R., MACDONELL, R., JACKSON, G. & BERKOVIC, S. 2009 The peri-ictal state: cortical excitability changes within 24 h of a seizure. *Brain* **132** (4), 1013–1021.
- BERGER, V. W. & ZHOU, Y. 2014 Kolmogorov–smirnov test: Overview. *Wiley statsref: Statistics reference online* .
- BHATIA, R. 2009 *Positive definite matrices*. Princeton university press.
- BHATIA, R., JAIN, T. & LIM, Y. 2019 On the bures–wasserstein distance between positive definite matrices. *Expositiones Mathematicae* **37** (2), 165–191.
- BHATTACHARYA, R. & PATRANGENARU, V. 2003 Large sample theory of intrinsic and extrinsic sample means on manifolds. *The Annals of Statistics* **31** (1), 1–29.
- BHATTACHARYA, R. & PATRANGENARU, V. 2005 Large sample theory of intrinsic and extrinsic sample means on manifolds II. *Ann. Stat.* **33**, 1225–1259.
- BIASIUCCI, A., FRANCESCHIELLO, B. & MURRAY, M. M. 2019 Electroencephalography. *Current Biology* **29** (3), R80–R85.
- BOX, G. E., JENKINS, G. M., REINSEL, G. C. & LJUNG, G. M. 2015 *Time series analysis: forecasting and control*. John Wiley & Sons.
- BOX, G. E. & PIERCE, D. A. 1970 Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association* **65** (332), 1509–1526.
- BROCKWELL, P. J. & DAVIS, R. A. 1991 *Time series: theory and methods*. Springer science & business media.

- BROCKWELL, P. J. & DAVIS, R. A. 2002 *Introduction to time series and forecasting*. Springer.
- BUI, M. N., POKERN, Y. & DELLAPORTAS, P. 2023 Inference for partially observed Riemannian Ornstein-Uhlenbeck diffusions of covariance matrices. *Bernoulli* **29** (4), 2961–2986.
- BURNHAM, K. P. & ANDERSON, D. R. 2004 Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research* **33** (2), 261–304.
- BURNS, S. P., SANTANIELLO, S., YAFFE, R. B., JOUNY, C. C., CRONE, N. E., BERGEY, G. K., ANDERSON, W. S. & SARMA, S. V. 2014 Network dynamics of the brain and influence of the epileptic seizure onset zone. *Proceedings of the National Academy of Sciences* **111** (49), E5321–E5330.
- BURRELLO, A., CAVIGELLI, L., SCHINDLER, K., BENINI, L. & RAHIMI, A. 2019 Laelaps: An energy-efficient seizure detection algorithm from long-term human iieeg recordings without false alarms. In *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 752–757. IEEE.
- CONGEDO, M., BARACHANT, A. & BHATIA, R. 2017 Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces* **4** (3), 155–174.
- COOK, M. J., KAROLY, P. J., FREESTONE, D. R., HIMES, D., LEYDE, K., BERKOVIC, S., O'BRIEN, T., GRAYDEN, D. B. & BOSTON, R. 2016 Human focal seizures are characterized by populations of fixed duration and interval. *Epilepsia* **57** (3), 359–368.
- CURTIS, M. L. 2012 *Matrix groups*. Springer Science & Business Media.
- DAVID, P. 2019 A riemannian quotient structure for correlation matrices with applications to data science. PhD thesis, The Claremont Graduate University.
- DO CARMO, M. P. & FLAHERTY FRANCIS, J. 1992 *Riemannian geometry*, , vol. 6. Springer.
- DRYDEN, I. L., KOLOYDENKO, A. & ZHOU, D. 2009 Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3** (3), 1102–1123.
- DRYDEN, I. L. & MARDIA, K. V. 2016 *Statistical shape analysis: with applications in R*, , vol. 995. John Wiley & Sons.

- FLETCHER, P. T., LU, C., PIZER, S. M. & JOSHI, S. 2004 Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE T. Med. Imaging* **23** (8), 995–1005.
- FLETCHER, T. P. 2013 Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vision* **105**, 171–185.
- FRÉCHET, M. 1948 Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut Henri Poincaré*, , vol. 10, pp. 215–310.
- FREESTONE, D. R., KAROLY, P. J. & COOK, M. J. 2017 A forward-looking review of seizure prediction. *Current opinion in neurology* **30** (2), 167–173.
- GALLIER, J. & QUAINANCE, J. 2020 *Differential geometry and Lie groups: a computational perspective*, , vol. 12. Springer Nature.
- GOLUB, G. H. & VAN LOAN, C. F. 2013 *Matrix computations*, , vol. 3. JHU press.
- GUAN, S., ZHAO, K. & YANG, S. 2019 Motor imagery eeg classification based on decision tree framework and riemannian geometry. *Computational intelligence and neuroscience* **2019**.
- HAN, A., MISHRA, B., JAWANPURIA, P. & GAO, J. 2021 Generalized bures-wasserstein geometry for positive definite matrices. *arXiv preprint arXiv:2110.10464* .
- HÖLLER, Y. & NARDONE, R. 2021 Quantitative eeg biomarkers for epilepsy and their relation to chemical biomarkers. *Advances in clinical chemistry* **102**, 271–336.
- HUANG, Z., WANG, R., SHAN, S., LI, X. & CHEN, X. 2015 Log-euclidean metric learning on symmetric positive definite manifold with application to image set classification. In *International conference on machine learning*, pp. 720–729. PMLR.
- HUCKEMANN, S., HOTZ, T. & MUNK, A. 2010 Intrinsic shape analysis: Geodesic pca for riemannian manifolds modulo isometric lie group actions. *Statistica Sinica* pp. 1–58.
- HUCKEMANN, S. & ZIEZOLD, H. 2006 Principal component analysis for riemannian manifolds, with an application to triangular shape spaces. *Advances in Applied Probability* **38** (2), 299–319.
- HYNDMAN, R. J. & ATHANASOPOULOS, G. 2018 *Forecasting: principles and practice*. OTexts.
- JUNG, S., DRYDEN, I. L. & MARRON, J. S. 2012 Analysis of principal nested spheres. *Biometrika* **99** (3), 551–568.
- JUPP, P. E. & MARDIA, K. V. 2009 *Directional statistics*. John Wiley & Sons.

- KARCHER, H. 1977 Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics* **30** (5), 509–541.
- KAROLY, P. J., KUHLMANN, L., SOUDRY, D., GRAYDEN, D. B., COOK, M. J. & FREESTONE, D. R. 2018 Seizure pathways: A model-based investigation. *PLoS computational biology* **14** (10), e1006403.
- KAROLY, P. J., UNG, H., GRAYDEN, D. B., KUHLMANN, L., LEYDE, K., COOK, M. J. & FREESTONE, D. R. 2017 The circadian profile of epilepsy improves seizure forecasting. *Brain* **140** (8), 2169–2182.
- KENDALL, W. S. 1990 Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence. *Proceedings of the London Mathematical Society* **3** (2), 371–406.
- KEOGH, C., PINI, G., GEMO, I. & TROPEA, D. 2019 Statistical modelling of cortical connectivity using non-invasive electroencephalograms. *JoVE (Journal of Visualized Experiments)* (153), e60249.
- KILOH, L. G., MCCOMAS, A. J. & OSSELTON, J. W. 2013 *Clinical electroencephalography*. Butterworth-Heinemann.
- KLIMESCH, W. 1997 Eeg-alpha rhythms and memory processes. *International Journal of psychophysiology* **26** (1-3), 319–340.
- KUHNERT, M.-T., ELGER, C. E. & LEHNERTZ, K. 2010 Long-term variability of global statistical properties of epileptic brain networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **20** (4).
- KUME, A. & LE, H. 2003 On fréchet means in simplex shape spaces. *Advances in Applied Probability* **35** (4), 885–897.
- LANG, S. 2012 *Fundamentals of differential geometry*, , vol. 191. Springer Science & Business Media.
- LE, H. 2003 Unrolling shape curves. *Journal of the London Mathematical Society* **68** (2), 511–526.
- LENGLET, C., ROUSSON, M., DERICHE, R. & 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.
- LUEG, J. 2023 Phylogenetic forest spaces .

- MALAGÒ, L., MONTRUCCHIO, L. & PISTONE, G. 2018 Wasserstein riemannian geometry of gaussian densities. *Information Geometry* **1** (2), 137–179.
- MALLASTO, A. & FERAGEN, A. 2018 Wrapped gaussian process regression on riemannian manifolds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5580–5588.
- MALLER, R. A., MÜLLER, G. & SZIMAYER, A. 2009 Ornstein–uhlenbeck processes and extensions. *Handbook of financial time series* pp. 421–437.
- MARRON, J. S. & ALONSO, A. M. 2014 Overview of object oriented data analysis. *Biometrical J.* **56** (5), 732–753.
- MEAD & A. 1992 Review of the development of multidimensional scaling methods. *Journal of the Royal Statistical Society* **41** (1), 27.
- MOAKHER, M. 2005 A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications* **26** (3), 735–747.
- MOAKHER, M. & ZÉRAÏ, M. 2011 The riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data. *Journal of Mathematical Imaging and Vision* **40** (2), 171–187.
- MUSHTAQ, R. 2011 Augmented dickey fuller test .
- NAFTULIN, J. S., AHMED, O. J., PIAONTI, G., EICHENLAUB, J.-B., MARTINET, L.-E., KRAMER, M. A. & CASH, S. S. 2018 Ictal and preictal power changes outside of the seizure focus correlate with seizure generalization. *Epilepsia* **59** (7), 1398–1409.
- NEWAY, W. K. & MCFADDEN, D. 1994 Large sample estimation and hypothesis testing. *Handbook of econometrics* **4**, 2111–2245.
- NOACHTAR, S. & RÉMI, J. 2009 The role of eeg in epilepsy: a critical review. *Epilepsy & Behavior* **15** (1), 22–33.
- PENNEC, X., FILLARD, P. & AYACHE, N. 2006 A riemannian framework for tensor computing. *International Journal of computer vision* **66** (1), 41–66.
- PENNEC, X., SOMMER, S. & FLETCHER, T. 2019 *Riemannian geometric statistics in medical image analysis*. Academic Press.
- PRABHU, S. S., SINHA, N. ET AL. 2020 Sleep eeg analysis utilizing inter-channel covariance matrices. *Biocybernetics and Biomedical Engineering* **40** (1), 527–545.

- PREUSS, M., PREISS, S., SYRBE, S., NESTLER, U., FISCHER, L., MERKENSCHLAGER, A., BERTSCHE, A., CHRISTIANSEN, H. & BERNHARD, M. K. 2015 Signs and symptoms of pediatric brain tumors and diagnostic value of preoperative eeg. *Child's Nervous System* **31** (11), 2051–2054.
- QUENOUILLE, M. H. 1949 Approximate tests of correlation in time-series 3. In *Mathematical Proceedings of the Cambridge Philosophical Society*, , vol. 45, pp. 483–484. Cambridge University Press.
- RIQUELME, A. 2021 A riemannian approach to multivariate geostatistical modeling. *arXiv preprint arXiv:2109.14550* .
- ROSENOW, F. & LÜDERS, H. 2001 Presurgical evaluation of epilepsy. *Brain* **124** (9), 1683–1700.
- SCHINDLER, K., GAST, H., STIEGLITZ, L., STIBAL, A., HAUF, M., WIEST, R., MARIANI, L. & RUMMEL, C. 2011 Forbidden ordinal patterns of periictal intracranial eeg indicate deterministic dynamics in human epileptic seizures. *Epilepsia* **52** (10), 1771–1780.
- SCHINDLER, K. A., BIALONSKI, S., HORSTMANN, M.-T., ELGER, C. E. & LEHNERTZ, K. 2008 Evolving functional network properties and synchronizability during human epileptic seizures. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **18** (3), 033119.
- SCHIRATTI, J.-B., ALLASSONNIERE, S., COLLIOT, O. & DURRLEMAN, S. 2015 Learning spatiotemporal trajectories from manifold-valued longitudinal data. In *Advances in Neural Information Processing Systems*, , vol. 28, pp. 2404–2412.
- SCHROEDER, G. M., DIEHL, B., CHOWDHURY, F. A., DUNCAN, J. S., DE TISI, J., TREVELYAN, A. J., FORSYTH, R., JACKSON, A., TAYLOR, P. N. & WANG, Y. 2020 Seizure pathways change on circadian and slower timescales in individual patients with focal epilepsy. *Proceedings of the National Academy of Sciences* **117** (20), 11048–11058.
- SKOVGAARD, L. T. 1984 A riemannian geometry of the multivariate normal model. *Scandinavian journal of statistics* pp. 211–223.
- SPENCER, S. S., SPENCER, D. D., WILLIAMSON, P. D. & MATTSON, R. H. 1981 Ictal effects of anticonvulsant medication withdrawal in epileptic patients. *Epilepsia* **22** (3), 297–307.
- SRA, S. & HOSSEINI, R. 2015 Conic geometric optimization on the manifold of positive definite matrices. *SIAM Journal on Optimization* **25** (1), 713–739.

- STURM, K.-T. 2003 Probability measures on metric spaces of nonpositive. *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces: Lecture Notes from a Quarter Program on Heat Kernels, Random Walks, and Analysis on Manifolds and Graphs: April 16-July 13, 2002, Emile Borel Centre of the Henri Poincaré Institute, Paris, France* **338**, 357.
- SYLVESTER, J. J. 1884 Sur l'équation en matrices $px = xq$. *CR Acad. Sci. Paris* **99** (2), 67–71.
- THANWERDAS, Y. & PENNEC, X. 2021 Geodesic of the quotient-affine metrics on full-rank correlation matrices. *arXiv preprint arXiv:2103.04621*.
- TRUCCOLO, W., DONOGHUE, J. A., HOCHBERG, L. R., ESKANDAR, E. N., MADSEN, J. R., ANDERSON, W. S., BROWN, E. N., HALGREN, E. & CASH, S. S. 2011 Single-neuron dynamics in human focal epilepsy. *Nature neuroscience* **14** (5), 635–641.
- VAROQUAUX, G., BARONNET, F., KLEINSCHMIDT, A., FILLARD, P. & THIRION, B. 2010 Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 200–208. Springer.
- YAIR, O., BEN-CHEN, M. & TALMON, R. 2019 Parallel transport on the cone manifold of spd matrices for domain adaptation. *IEEE Transactions on Signal Processing* **67** (7), 1797–1811.
- YAMAMOTO, M. S., SADATNEJAD, K., TANAKA, T., ISLAM, R., TANAKA, Y. & LOTTE, F. 2020 Detecting eeg outliers for bci on the riemannian manifold using spectral clustering. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 438–441. IEEE.
- YGER, F., BERAR, M. & LOTTE, F. 2016 Riemannian approaches in brain-computer interfaces: a review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25** (10), 1753–1762.
- ZHUANG, X., YANG, Z., MISHRA, V., SREENIVASAN, K., BERNICK, C. & CORDES, D. 2020 Single-scale time-dependent window-sizes in sliding-window dynamic functional connectivity analysis: a validation study. *NeuroImage* **220**, 117111.
- ZIEZOLD, H. 1977 On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the 1974 European Meeting of Statisticians*, pp. 591–602. Springer.