Chapter 5

Mathematical Foundations

We saw in Sections 4.2, 4.3, and 4.4 an intuitive presentation of the space of variation for one, two and more dimensions. In each case, we presented also an intuitive definition of distance in the space of variation. For a given shape invariant model, we see in Section 5.1 some necessary conditions for the space of variation to be a manifold. We also show in Sections 5.3, 5.5, and 5.6 that the functions intuitively presented in Sections 4.2, 4.3, and 4.4 are metrics. While we discussed in Subsections 4.2.2, 4.3.2, and 4.4.2 how to estimate the center and spread of variation from the data, we prove consistency of these estimates, for 1-dimensional, 2-dimensional, and d'-dimensional manifolds of variation in Sections 5.3, 5.5, 5.6, 5.7 respectively. The estimates of center and spread in the data converge to their continuous counterparts, the Fréchet mean and the Fréchet variance, under certain conditions.

5.1 Shape invariant models and space of variation

We are interested in this dissertation in decomposing and quantifying the variation in data of common shape along predetermined modes. So, we assume that the data can be represented in a shape invariant model, where the common shape is known and the predetermined modes could be parameterized by a vector θ.

5.1.1 General shape invariant model

All the shape invariant models presented in Chapters 3 and 4 are examples of the general model

$$z_{i,j} = R(\theta_i, t_j) + \epsilon_{i,j}, 1 \leq i \leq n; 1 \leq j \leq d$$

where the response $z_{i,j}$ varies nonlinearly as a function of $t_j \in \mathbb{R}$, where $\theta_i \in \mathbb{R}^{d'}$ is the vector of parameters of variation, and where R is the common regression function.

5.1.2 Space of variation

For data in $\mathbb{R}^d$, we define the **space of variation** V, corresponding to the parameter space $\Theta \subset \mathbb{R}^{d'}$, the subspace of $\mathbb{R}^d$ such that

$$x = (x_1, \ldots, x_d) \in V \Leftrightarrow \exists \theta \in \Theta \text{ such that } x_j = R(\theta, t_j) \forall j; 1 \leq j \leq d.$$

5.1.3 Examples

Let z be the common shape function, we saw in Chapter 4 some one parameter spaces of variation ($d' = 1$ in cases (1)-(3)), two-parameter spaces of variation ($d' = 2$ in cases (4)-(6)) and a three-parameter space of variation ($d' = 3$ in case (7)).

1. When $\theta_i = h_i$ and $R(\theta_i, t) = z(t) + h_i$ then the space of variation is the linear vertical shift.

2. When $\theta_i = m_i$, and $R(\theta_i, t) = z(t - m_i)$ then the space of variation is the nonlinear horizontal shift, this variation is called registration in the FDA literature.

3. When $\theta_i = w_i$ and $R_{\theta_i} = w_i z(w_i(t - m))$, then the space of variation is the nonlinear generalist-specialist or change in width.

4. When $\theta_\mathbf{i} = (m_i, h_i)$ and $R(\theta_\mathbf{i}, t) = z(t - m_i) + h_i$, then the space of variation is the nonlinear horizontal shift and the linear vertical shift.

5. When $\theta_\mathbf{i} = (w_i, h_i)$ and $R(\theta_\mathbf{i}, t) = w_i * z(w_i(t - m)) + h_i$ then the space of variation is the nonlinear generalist-specialist and the linear horizontal shift.

6. When $\theta_\mathbf{i} = (w_i, m_i)$ and $R(\theta_\mathbf{i}, t) = w_i * z(w_i(t - m_i))$ then the space of variation is the nonlinear generalist-specialist and the nonlinear horizontal shift.

7. Finally, when we have three simultaneous modes of variation then $\theta_\mathbf{i} = (w_i, m_i, h_i)$ and $R(\theta_\mathbf{i}, t) = w_i z(w_i(t - m_i)) + h_i$.

The amplitude problem, i.e

$$R(\theta_i, t) = z(\theta_i t),$$

is another example of shape invariant models in FDA, see [Ramsay and Silverman (1997)]. We saw in Chapter 4 that in cases (1)-(3), the space of variation V is a curve or a line. We can also show that for the amplitude problem, the space of variation V is a curve. Lines and curves are one dimensional *manifolds* under certain conditions on the parametrization. In cases (4)-(6), we saw also that the space of variation V is a plane or a surface which are, under certain conditions, two-dimensional manifolds. Manifolds are subspaces defined in differential geometry, which include all linear subspaces. We see in Theorem 1 a condition on the common shape and the parametrization under which the space of variation V is a manifold of the same dimension as the parameter space. The mathematical definition of manifold are given in Subsection 5.1.4. We see in Subsection 5.1.5, that the conditions of Theorem 1 are satisfied for all cases (1)-(7) and for any polynomial template shape. This makes V a manifold of dimension 1 in cases (1)-(3), of dimension 2 in cases (4)-(6) and of dimension 3 in case (7).

5.1.4 Manifold

This subsection gives some background in differential geometry. The differential geometry definitions in this subsection follow the development of [Aubin (2000)]. Note that the curves of variation and the surfaces of variation we saw in Sections 4.2, and 4.3 did not cross and did not have any discontinuities. As we see in this subsection, a manifold of dimension d' is a type of topological space. It is Hausdorff which is a condition on the separability of the space (defined in Definition 1). It is also locally homeomorphic to $\mathbb{R}^{d'}$ which guarantees some smoothness and identifies the dimensionality d' of a manifold (defined in Definition 2).

Definition 1. Hausdorff space. *A topological space is Hausdorff if any two distinct points have disjoint neighborhood.*

Definition 2. Manifold. *A manifold M of dimension d' is a Hausdorff topological space such that each point P of M has a neighborhood Ω homeomorphic to $\mathbb{R}^{d'}$. We call this homeomorphism ϕ_P.*

Although in this most general definition of manifold, the homeomorphism ϕ_P is local, i.e depends on the point P, we focus in this Chapter on the particular case of the homeomorphism ϕ being the same for all points in the manifold.

From the definition of manifold, we have the following sufficient condition for V to be a manifold of the same dimension as the parameter space $\Theta \subset \mathbb{R}^{d'}$.

Theorem 1. *For a fixed sampling vector $t \in \mathbb{R}^d$, let the function $R(., \mathbf{t})$ be such that*

$$\begin{aligned} R(., \mathbf{t}) : \mathbb{R}^{d'} &\rightarrow V \\ \theta &\mapsto R(\theta, \mathbf{t}) \end{aligned}$$

If $R(., \mathbf{t})$ is a homeomorphism from $\mathbb{R}^{d'}$ to V, then V is a manifold of dimension d'.

5.1.5 Special case: polynomial template shape

In Chapter 3, we fit the three parameter Shape Invariant Model with a polynomial template shape. In the following Proposition and Corollary, we show that such a parametrization and any polynomial template shape satisfies the conditions of Theorem 1.

Proposition 1. *If two polynomials of degree k are equal in at least $k+1$ distinct points then their coefficients are equal. If a sequence of polynomials of degree k converge in at least $k+1$ distinct points then their coefficients also converge.*

See proof in Appendix Section 6.3.

Corrolary 1. *When z is a polynomial of degree k and $2 \leq k \leq d-1$ (i.e dimension of the data d limits the degree of the polynomial). Then*

- *In cases (1),(2), and (3) (one mode of variation) of Section 5.1, V is a manifold of degree 1.*
- *In cases (4), (5), and (6) (two modes of variation) of Section 5.1, V is a manifold of degree 2.*
- *In case (7) (three simultaneous modes of variation), V is a manifold of degree 3.*

See proof in Appendix Section 6.3.

5.1.6 Fréchet mean and Fréchet variance in a manifold

To quantify the variation along a space of variation V, we need to define an appropriate metric d_V. In Section 5.3, we use the arc-distance as the metric in the manifold of variation V of dimension 1. We define in Section 5.5 and Section 5.6 the appropriate metric in a manifold of variation of dimension 2. In Sections 4.2 and 4.3, we saw how to estimate the center and spread of variation from the data in a given manifold for a given distance. In this subsection, we see a formal definition of mean and variance in a metric manifold as defined by Fréchet

[Fréchet (1948)]. In a metric manifold (M, d) with measure μ, let the Fréchet function be,

$$F(x) = \int_M d^2(x, y) d\mu(y)$$

This function is well defined if $\int_M d^2(x, y) d\mu(y) < \infty, \forall x \in M$

Let X be a random variable with measure μ on M,

Definition 3. Fréchet mean set. *In a metric manifold (M, d), the Fréchet mean set $E_F(X)$ of a random variable X, in the manifold M, with probability measure μ, is the set of points on the manifold which minimize the function F(x). i.e*

$$\tilde{X}_F \in E_F(X) \textit{ iff } F(\tilde{X}_F) = min_{x \in M} F(x).$$

Note that although we can have more than one Fréchet mean, we can define a unique Fréchet variance

Definition 4. Fréchet variance. *The Fréchet variance is the value $F(\tilde{X}_F)$ for any $\tilde{X}_F$ in the Fréchet mean set $E_F(X)$.*

Note also that the Fréchet variance is a scalar, and not a matrix, even for manifolds of dimension $d' \geq 1$.

Estimation

The measure of center and spread that we presented and discussed in Sections 4.2 and 4.3, respectively $\tilde{R}$ and $\widetilde{SSM}$, were estimates of the Fréchet mean and Fréchet variance from the toy example data. We can now define those estimates in a more general case, for given data $R_1, \ldots, R_n$ where $R_i = R(\theta_\mathbf{i}, \mathbf{t})$ in the metric manifold (V, d_V), the estimate of the Fréchet function is

$$F_n(R) = \sum_{i=1}^{n} d_V^2(R_i, R) \tag{5.1}$$

We can derive then the estimates of the Fréchet mean set $E_{F,n}(R)$ and Fréchet variance $\widetilde{SSM}$

$$\begin{aligned}\tilde{R} \in E_{F,n}(R) \quad &\Leftrightarrow \quad F_n(\tilde{R}) = \min_{R \in V} F_n(R) \\ \widetilde{SSM} \quad &= \quad \sum_{i=1}^{n} d_V^2(R_i, \tilde{R})\end{aligned}$$

Note that d_V is the metric in the manifold that we intuitively constructed in Sections 4.2 and 4.3 and that we formally define in Sections 5.3, 5.5 and 5.6, to allow for the decomposition of the variation in the data. Note also that in the particular toy example in Chapter 4, the estimated Fréchet mean set had only one element (the estimate of the Fréchet mean). We show in Sections 5.3 and 5.5 that for one dimensional and some 2-dimensional manifolds of variation, there is only one Fréchet mean and one estimate of the Fréchet mean.

5.2 Linearizing the manifold

Linearizing is the tool which allows us in Sections 5.3 and 5.5 to show existence and uniqueness of the estimates of Fréchet mean and variance and their consistency for our chosen distance d_V. Linearization also allows us to construct an appropriate distance for 2-dimensional manifolds in Section 5.6. In Section 5.3, we see that the appropriate distance in the one-dimensional manifold is the arc-distance. The arc-distance is the distance of the arc between two points in a curve. We call linearization a transformation of the manifold of variation V of dimension d' into $\mathbb{R}^{d'}$ which preserves arc-distances along each mode in V. The linearization gives a new look on variation, as we show in Sections 5.3, 5.5 and 5.6, the metric d_V that we define is equivalent to the Euclidean metric in the linearized space. In particular, if we linearize the one dimensional manifold into $\mathbb{R}$ by preserving the arc-distances, then the Euclidean distance in the linearized space is equivalent to the arc-distance in the manifold. Similarly, when the 2-dimensional manifold satisfies an equality of path condition defined in Section 5.5, the distance d_V that we define is equivalent to the Euclidean distance in $\mathbb{R}^2$. For the more general cases in 2-dimensional manifolds, in Section 5.6, and d'-dimensional manifolds, in Section 5.7, we refer to a general result in [Bhattacharya and Patrangenaru (2003)] for consistency of the estimated Fréchet mean set to the Fréchet mean set. However, linearization is our tool for defining an appropriate distance in this general case.

5.3 One mode of variation

As we intuitively justified in Section 4.2, when V is a space of variation of dimension one, i.e either a line or a curve, the arc-distance or distance along the curve is the appropriate metric to measure the spread. Let R_1 and R_2 be two points in the manifold, then there exist two parameters θ_1 and θ_2 such that $R_i = R(\theta_i, \mathbf{t})$ for $i = 1, 2$. Without loss of generality, suppose $\theta_1 \leq \theta_2$, then the arc-distance defined as the refinement limit of linear interpolations, is

$$\mathrm{Arcd}(R_1, R_2) = \lim_{n\to\infty} \sum_{i=0}^{n} \left|\left| R\left(\theta_1 + \frac{(i+1)*(\theta_2-\theta_1)}{n}, \mathbf{t}\right) - R\left(\theta_1 + \frac{i*(\theta_2-\theta_1)}{n}, \mathbf{t}\right)\right|\right|$$

If the function $R(\theta, \mathbf{t})$ is differentiable with respect to θ, the arc-distance is equivalent to

$$\mathrm{Arcd}(R_1, R_2) = \int_{\theta_1}^{\theta_2} \left|\left| \frac{\partial}{\partial\theta} R(\theta, \mathbf{t}) \right|\right| d\theta$$

We define linearizing functions, in this 1-dimensional case, in Subsection 5.3.1. Linearization is a useful tool to show existence, uniqueness and consistency of the estimates of the Fréchet mean and the Fréchet variance in Subsection 5.3.2. Although the space of variation is defined through parametrization, we note in Subsection 5.3.3 that the metric as well as the estimates are *parameter-free.*

5.3.1 Linearizing one mode

The arc-distance is a metric in the one dimensional space of variation. If we linearize the one dimensional manifold into a line by keeping the arc-distances between the points fixed, then the arc-distance between two points in the manifold corresponds to the Euclidean distance in the line. Let $\mathcal{L}$ be the class of linearizing functions, $\mathcal{L} = \{L_\theta, \theta \in \Theta\}$ such that for all θ_0 in Θ

$$\begin{aligned} L_{\theta_0} : V &\rightarrow R \\ x = R(\theta, \mathbf{t}) &\mapsto L_{\theta_0}(x) \\ \text{such that } L_{\theta_0}(x) &= \text{sign}(\theta - \theta_0)\text{Arcd}(R(\theta, \mathbf{t}), R(\theta_0, \mathbf{t})) \end{aligned}$$

Then for a given parameter θ_0, we can consider $R(\theta_0, \mathbf{t})$ as the origin in the manifold V by the transformation L_{θ_0} because by definition $L_{\theta_0}(R(\theta_0, \mathbf{t})) = 0$. For all θ_0, L_{θ_0} satisfies the isometry property, i.e

$$|L_{\theta_0}(R_1) - L_{\theta_0}(R_2)| = \text{Arcd}(R_1, R_2),$$

and the distance between two points $L_{\theta_0}(R_1)$ and $L_{\theta_0}(R_2)$ does not depend on the origin. Note that the existence of an ordering ($\leq$) in the parameter space Θ gives a way of ordering in the one dimensional manifold V. By definition of the space of variation and by using the isometry property, the following proposition follows

Proposition 2. *For all θ_0, the linearizing function L_{θ_0} is a homeomorphism from $(V, Arcd)$ to $\mathbb{R}$. Moreover,*

$$L_{\theta_1}(x) = L_{\theta_0}(x) + sign(\theta_1 - \theta_0)L_{\theta_1}(\theta_0)$$

See proof in Section 6.3 of Appendix.

5.3.2 Mean and variance

The linearizing tool allows us to prove the existence, uniqueness and consistency of estimates of Fréchet mean and Fréchet variance for a one-dimensional manifold of variation V. In the one dimensional case, for given data $R_1, \ldots, R_n$ of measure μ in the space of variation V, such that $R_i = R(\theta_i, \mathbf{t})$, the estimate of the Fréchet function is

$$F_n(R) = \sum_{i=1}^{n} \mathrm{Arcd}^2(R_i, R)$$

And by isometry of L_{θ_0}, we have

$$F_n(R) = \sum_{i=1}^{n} (L_{\theta_0}(R_i) - L_{\theta_0}(R))^2 \tag{5.2}$$

Because the $L_{\theta_0}(R_i)$'s are real values random variables, and $L_{\theta_0}(R) \in \mathbb{R}$, the minimization of $F_n(R)$ is straightforward. In particular, we can show the following proposition of existence and uniqueness of estimates of the Fréchet mean in a manifold of variation of dimension 1 satisfying Theorem 1

Proposition 3. *For a one-dimensional differentiable space of variation $(V, Arcd)$ satisfying conditions of Theorem 1, we have that*

$$\tilde{R} = L_{\theta_0}^{-1}(\overline{L_{\theta_0}}) \text{ for all } \theta_0$$

where

$$\overline{L_{\theta_0}} = \frac{1}{n} \sum_{i=1}^{n} L_{\theta_0}(R_i)$$

Using the linearizing function and the law of large numbers for i.i.d random variables in $\mathbb{R}$ we can also show that $\tilde{R}$ and $\frac{1}{n}\widetilde{SSM}$ are consistent estimators of the Fréchet mean and the

Fréchet variance respectively.

Theorem 2. *For i.i.d random variables $R_1, \ldots, R_n$ of measure μ in $(V, Arcd)$ of finite Fréchet mean and variance, and V satisfying the conditions in Proposition 3. We have that*

$$\begin{aligned} d_V(\tilde{R}, \tilde{R}_F) &\rightarrow 0 \ (a.s) \\ \left| \frac{1}{n}\widetilde{SSM} - Var_F \right| &\rightarrow 0 \ (a.s) \end{aligned}$$

Where $\tilde{R}_F$ is **the** *Fréchet mean and Var_F is the Fréchet variance.*

See detailed proof in Appendix Section 6.3

5.3.3 Equivalence of Parameterizations

Several parameterizations might lead to the same space of variation V. For example, in the generalist-specialist case, the space V is defined as

$$V = \{x \in \mathbb{R}^d, \exists w > 0 \text{ such that } x = w * z(w\mathbf{t}), \mathbf{t} \in \mathbb{R}^d\}$$

We can equivalently parameterize the same space V as

$$V = \{x \in \mathbb{R}^d, \exists \mu \in \mathbb{R} \text{ such that } x = \exp \mu * z(\exp \mu\mathbf{t}), \mathbf{t} \in \mathbb{R}^d\}$$

A question of interest is would a different parametrization of the same space V lead to a different center and spread? Because the arc-distance between two points does not depend on the parametrization, the measure of center and spread do not depend on a particular parametrization of V. So, although the space of variation is defined through a particular parametrization, the center and spread in the manifold stay the same modulo an equivalent parametrization which makes them in some sense *parametrization-free*.

5.4 Two modes of variation, notation and comments

In Section 4.3, we saw three cases of two-dimensional manifolds corresponding to cases (4)-(6) in Section 5.1. In all these cases, the space of variation V is a manifold of dimension 2 which satisfies Theorem 1 and is parameterized by

$$x \in V \Leftrightarrow \exists(\alpha, \beta) \text{ such that } x = R\left((\alpha, \beta), \mathbf{t}\right)$$

There are several possible metrics in a two-dimensional manifold. For example, the distance of the shortest path along the manifold between two points is a metric (geodesic distance). However, this distance does not allow for the decomposition of the variation into the modes of interest which do not always follow geodesics. In Sections 5.5 and 5.6 we define metrics which do allow for the decomposition. In cases (4) and (5), the manifold of variation corresponds to the particular case where two same distance paths go from one point to the next following the curves of variation, as we saw with a toy example in Section 4.3. This condition is mathematically formalized in Section 5.5. Case (6) corresponds to the more general case of inequality of paths. In the first situation, we mathematically define the equality of path condition and show that the distance is a metric in Section 5.5. In Section 5.6, we consider the second situation, the two possible linearizations and the choice of the metric.

5.5 Two modes: Particular case, equality of path between two points

Suppose we have two modes of variation of interest, parameterized respectively by α and β. If we fix a parameter and let one parameter vary and find that the distance between two points depends only on the values of the varying parameter, then we have equality of paths. We mathematically formalize this property is Subsection 5.5.1. We saw in Section 4.3 an intuitive justification of the metric we chose to use in a two-dimensional manifold, we formally define this function and show that it is a metric in Subsection 5.5.2. As in the one dimensional case, the linearization presented in Subsection 5.5.3 is a useful tool to show existence, uniqueness and consistency of the estimates of the Fréchet mean and Fréchet variance in Subsection 5.5.4.

5.5.1 Condition of equality of paths

Figure 4.8 shows intuitively the *equality of path* property for a toy example data set with two modes of variation. This property is formally stated in Definition 5.

Definition 5. *For a two-dimensional manifold of variation V parameterized by (α, β), we say that we have equality of paths if*

$$\begin{aligned} \forall(\alpha,\beta,\alpha_1,\alpha_2,\beta_1,\beta_2) &\in \Theta \\ Arcd(R((\alpha_1,\beta),\mathbf{t}),R((\alpha_2,\beta),\mathbf{t})) &= C_{\alpha_1,\alpha_2}, \ \textit{and} \\ Arcd(R((\alpha,\beta_1),\mathbf{t}),R((\alpha,\beta_2),\mathbf{t})) &= C_{\beta_1,\beta_2} \end{aligned}$$

Where C_{α_1,α_2} (resp. C_{β_1,β_2}) is a function depending only on the values of α_1 and α_2 (resp. β_1 and β_2) and not on the value of β (resp. α).

It is implied here that the arc-distances along each of the two modes are well defined for all fixed values of the parameter of the other mode which requires a certain smoothness of the manifold. Differentiability of the regression function R with respect to the two parameters α (resp. β) for all values of β (resp. α) are two sufficient conditions. We saw in Section 4.3 that the condition of equal paths is satisfied for cases (5) and (6) of the toy example. Proposition 4 states that this property holds for those two cases and for all differentiable template shapes z.

Proposition 4. *For any differentiable template shape z, the space of variation V satisfies the equality of path condition in cases (5) and (6).*

See proof in Appendix Section 6.3

5.5.2 Distance along the manifold

When the equality of path is satisfied, we can simply define a metric in V which allows for the decomposition of the variation into one term which only depends on one parameter and a second term which only depends on the other parameter

Theorem 3. *For two-dimensional differentiable manifold of variation V for which the equality of path condition is satisfied the non-negative function defined as*

$$\begin{array}{rcl} d_V : V * V & \rightarrow & \mathbb{R}^+ \\ (R((\alpha_1, \beta_1), \mathbf{t}), R((\alpha_2, \beta_2), \mathbf{t})) & \mapsto & \sqrt{C^2_{\alpha_1,\alpha_2} + C^2_{\beta_1,\beta_2}} \end{array}$$

is a distance in V

See proof in Appendix Section 6.3.

5.5.3 Another look at distance through the linearizing function

For the one dimensional case in Section 5.3, the distance d_V we defined in Theorem 3 is equivalent to the Euclidean distance in the linearized space of variation. The linearization allows us to prove existence, uniqueness and consistency of the estimates of the Fréchet mean and the Fréchet variance. Consider the following linearizing functions $L_{\alpha_0,\beta_0} \in \mathcal{L}$, associated with origins parameterized by (α_0, β_0).

$$\begin{aligned} L_{\alpha_0,\beta_0} : V &\rightarrow \mathbb{R}^2 \\ x = R((\alpha,\beta),\mathbf{t}) &\mapsto L_{\alpha_0,\beta_0}(x) \\ L_{\alpha_0,\beta_0}(x) &= (\text{sign}(\alpha-\alpha_0) * C_{\alpha,\alpha_0}, \text{sign}(\beta-\beta_0) * C_{\beta,\beta_0}) \end{aligned}$$

Note that (α_0, β_0) parameterize an origin in V by the function L_{α_0,β_0} because by definition

$$L_{\alpha_0,\beta_0}\left(R((\alpha_0,\beta_0),\mathbf{t})\right) = 0$$

As in the one dimensional case, the ordering in the parameter space Θ gives also an ordering in the space of variation V. For all possible (α_0, β_0) in Θ, L_{α_0,β_0} satisfies the isometry property, i.e

$$||L_{\alpha_0,\beta_0}(R_1) - L_{\alpha_0,\beta_0}(R_2)|| = d_V(R_1, R_2), \forall R_1, R_2 \in V$$

By using the isometry property, and for a manifold of variation V of dimension 2 which satisfies Theorem 1 and the equality of path property we can show the following proposition

Proposition 5. *For all* (α_0, β_0)*,* L_{α_0,β_0} *is a homeomorphism from* (V, d_V) *to* $\mathbb{R}^2$.

So, the distance d_V in the space of variation V is equivalent to the Euclidean distance in the linearized space $L_{\alpha_0,\beta_0}(V)$.

5.5.4 Mean, variance and decomposition

For data $R_1, \ldots, R_n$ in V, the estimate of the Fréchet function is by definition

$$F_n(R) = \sum_{i=1}^{n} d_V^2(R_i, R)$$

By using the isometry property we have this equivalent characterization of $F_n(R)$,

$$F_n(R) = \sum_{i=1}^{n} ||L_{\alpha_0,\beta_0}(R_i) - L_{\alpha_0,\beta_0}(R)||^2 \tag{5.3}$$

This second expression of F_n is easier to minimize because we know that the sums of square $\sum_{i=1}^{n} ||X_i - X||^2$ is minimized by the average $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. So, we have the following proposition for existence and unicity of the estimate of the Fréchet mean.

Proposition 6. *For a two-dimensional differentiable space of variation* (V, d_V) *satisfying equality of path condition, we have that*

$$\begin{aligned} \tilde{R} &= L^{-1}_{\alpha_0,\beta_0}(\overline{L_{\alpha_0,\beta_0}}) \text{ , for all } (\alpha_0, \beta_0) \\ \textit{where} & \\ \overline{L_{\alpha_0,\beta_0}} &= \frac{1}{n}\sum_{i=1}^{n} L_{\alpha_0,\beta_0}(R_i) \end{aligned}$$

See proof Section 6.3 of Appendix.

The spread in the data is quantified by $\widetilde{SSM}$ such that

$$\widetilde{SSM} = \sum_{i=1}^{n} d_V(R_i, \tilde{R})^2 \tag{5.4}$$

By construction, the distance d_V allows us to decompose the variation into two terms, one which depends on the variation along one mode (parameterized by α) and the other which depends on the variation along the second mode (parameterized by β). As we have from the characterization of d_V in Theorem 3,

$$\begin{aligned} d_V(R_i, \tilde{R})^2 &= C^2_{\alpha_i,\alpha} + C^2_{\beta_i,\beta} \\ \text{So,} & \\ \widetilde{SSM} &= \widetilde{SSM}_\alpha + \widetilde{SSM}_\beta \\ \text{Where,} & \\ \widetilde{SSM}_\alpha &= \sum_{i=1}^{n} C^2_{\alpha_i,\alpha} \text{ and} \\ \widetilde{SSM}_\beta &= \sum_{i=1}^{n} C^2_{\beta_i,\beta} \end{aligned}$$

This decomposition lies at the heart of the sum of square data analysis done in Section 3.5. As we have seen in the one-dimensional case in Section 5.3, linearizing is useful to show consistency of the estimate of the Fréchet mean and consistency of the estimate of the Fréchet variance.

Theorem 4. *For i.i.d random variables $R_1, \ldots, R_n$ of measure μ in $(V, Arcd)$ of finite Fréchet mean and Fréchet variance, and V satisfying the conditions in Theorem 3. Let $\tilde{R}_F$ be the Fréchet mean and Var_F the Fréchet variance. Then*

$$\begin{aligned} d_V(\tilde{R}, \tilde{R}_F) &\rightarrow 0 \text{ (a.s)} \\ \left| \frac{1}{n}\widetilde{SSM} - Var_F \right| &\rightarrow 0 \text{ (a.s)} \end{aligned}$$

5.6 Two modes: General case

In the two modes of variation general case, the two paths going from one point to the other point in the manifold are not equal. We define a distance d_V that takes into account those two possible paths and which allows for the decomposition of the variance. When the equality of path is satisfied, this distance is the same as the one we defined in Section 5.5. We illustrated in Section 4.3, and Figures 4.10, 4.11 and 4.12 the two possible linearizations of the two-dimensional space for a toy example. We mathematically define those two linearizations in Subsection 5.6.1, and 5.6.2. For each linearization, we formally define a possible metric, and we combine these two metrics in Subsection 5.6.3. This metric allow us to decompose the variation in Subsection 5.6.4. Using the general consistency result from [Bhattacharya and Patrangenaru (2003)], we show consistency of the estimate of the Fréchet variance in Subsection 5.6.5. As was the case in Sections 5.3 and 5.5, the linearizing functions is defined with respect to an origin $R_0 = O$ corresponding to the parametrization (α_0, β_0). However, the distance between two points in the general case varies with the origin. We discuss the choice of an origin in Subsection 5.6.6.

5.6.1 First linearization

The first linearization maps V into $\mathbb{R}^2$ by keeping the arc-distances along the curves, which cross the origin $O = R((\alpha_0, \beta_0), \mathbf{t})$, constant (see Figure 4.10 and 4.11 for a graphic illustration of this linearization). Let $R_{i,k} = R((\alpha_i, \beta_k), \mathbf{t})$ for $i, j = 0, 1, 2$

$$
\begin{aligned}
L_{1,O} : V &\rightarrow \mathbb{R}^2 \\
R_{1,1} &\mapsto (\eta_1, \eta_2) \text{ such that} \\
\eta_1 &= \text{sign}(\alpha_1 - \alpha_0)\text{Arcd}(R_{1,0}, R_{0,0}) \\
\eta_2 &= \text{sign}(\beta_1 - \beta_0)\text{Arcd}(R_{0,1}, R_{0,0})
\end{aligned}
$$

After linearization, we define the metric $d_{1,O}$ between two points in the manifold as equivalent to the Euclidean distance in the linearized space such that

$$\begin{aligned} d_{1,O} : V * V &\rightarrow \mathbb{R}^+ \\ (R_{1,1}, R_{2,2}) &\mapsto d_{1,O}(R_{1,1}, R_{2,2}) \\ d_{1,O}(R_{1,1}, R_{2,2}) &= ||L_{1,O}(R_{1,1}) - L_{1,O}(R_{2,2})|| \end{aligned}$$

The following Proposition states that this function defines a metric,

Proposition 7. *For all origin O, and for a differentiable manifold V, the function $d_{1,O}$ defines a metric in V.*

See proof, Appendix Section 6.3

We can further derive the expression of $d_{1,O}$ as a function of the parameters of variation rather than the linearizing function as,

$$\begin{aligned} d^2_{1,O}(R_{1,1}, R_{2,2}) &= A^2 + B^2, \text{ where} \\ A^2 &= (\text{sign}(\alpha_1 - \alpha_0)\text{Arcd}(R_{1,0}, R_{0,0}) - \text{sign}(\alpha_2 - \alpha_0)\text{Arcd}(R_{2,0}, R_{0,0}))^2 \\ B^2 &= (\text{sign}(\beta_1 - \beta_0)\text{Arcd}(R_{0,1}, R_{0,0}) - \text{sign}(\beta_2 - \beta_0)\text{Arcd}(R_{0,2}, R_{0,0}))^2 \end{aligned}$$

We can simplify A^2 and B^2 further

$$\begin{aligned} A^2 &= (\text{sign}(\alpha_2, \alpha_1)\text{Arcd}(R_{1,0}, R_{2,0}))^2 \\ &= \text{Arcd}^2(R_{1,0}, R_{2,0}) \end{aligned}$$

Similarly, we can show that

$$B^2 = \text{Arcd}^2(R_{0,1}, R_{0,2})$$

Finally,

$$d^2_{1,O}(R_{1,1}, R_{2,2}) = \text{Arcd}^2(R_{1,0}, R_{2,0}) + \text{Arcd}^2(R_{0,1}, R_{0,2})$$

5.6.2 Second linearization

The second linearization also maps V into $\mathbb{R}^2$. However, this linearization keeps the arc-distances along the curves which cross the linearized point constant (See Figure 4.10 and 4.12 for a graphic illustration of this linearization). The second linearization $L_{2,O}$ is defined as follows,

$$\begin{aligned} L_{2,O} : V &\rightarrow \mathbb{R}^2 \\ R_{1,1} &\mapsto (\eta_1, \eta_2) \text{ such that} \\ \eta_1 &= \text{sign}(\alpha_1 - \alpha_0)\text{Arcd}(R_{1,1}, R_{0,1}) \\ \eta_2 &= \text{sign}(\beta_1 - \beta_0)\text{Arcd}(R_{1,1}, R_{1,0}) \end{aligned}$$

With this linearization, we can define a function $d_{2,O}$ which corresponds to the Euclidean distance in the linearized space.

$$\begin{aligned} d_{2,O} : V * V &\rightarrow \mathbb{R}^+ \\ d_{2,O}(R_{1,1}, R_{2,2}) &= ||L_{2,O}(R_{1,1}) - L_{2,O}(R_{2,2})|| \end{aligned}$$

We can rewrite $d_{2,O}$ as a function of the parametrization as follows

$$\begin{aligned} d_{2,O}(R_{1,1}, R_{2,2}) &= \sqrt{A^2 + B^2}, \text{ where} \\ A^2 &= (\text{sign}(\alpha_1 - \alpha_0)\text{Arcd}(R_{1,1}, R_{0,1}) - \text{sign}(\alpha_2 - \alpha_0)\text{Arcd}(R_{2,2}, R_{0,2}))^2 \\ B^2 &= (\text{sign}(\beta_1 - \beta_0)\text{Arcd}(R_{1,1}, R_{1,0}) - \text{sign}(\beta_2 - \beta_0)\text{Arcd}(R_{2,2}, R_{2,0}))^2 \end{aligned}$$

Proposition 8. *The function $d_{2,O}$ satisfies the following conditions*

(i) $d_{2,O}$ *is non-negative.*

(ii) $d_{2,O}$ *is symmetric.*

(iii) $d_{2,O}$ *satisfies the triangular inequality.*

If the mapping $L_{2,O}$ is one-to-one, then $d_{2,O}$ satisfies the condition

$$d_{2,O}(R_{1,1}, R_{2,2}) = 0 \Leftrightarrow R_{1,1} = R_{2,2}$$

which, in addition to the above conditions, makes $d_{2,O}$ a distance.

See proof, Appendix Section 6.3

Although $d_{2,O}$ is not always a distance, it always satisfies conditions *(i), (ii)* and *(iii)*, and we can combine it with $d_{1,O}$ to define a distance in Subsection 5.6.3. We can show that in the particular case when the space of variation V is two dimensional and satisfies the equal path property, the distance d_V we defined in Section 5.5 is the same as the two distances $d_{1,O}$ and $d_{2,O}$ that we have defined in this section.

Proposition 9. *When V satisfies the equality of paths property, and d_V is the distance we defined in Section 5.5, we have that $d_{1,O} = d_{2,O} = d_V$ for all origins O in the manifold.*

5.6.3 Distance in the manifold

For a fixed origin $O = R_{0,0}$, we have defined two distances $d_{1,O}$ and $d_{2,O}$ which correspond to two different linearizations $L_{1,O}$ and $L_{2,O}$ of the space of variation V into $\mathbb{R}^2$. Each one of these two distances takes into account one possible path to go from one point to another in the manifold of variation V. We can combine those two distances using the following lemma

Lemma 1. *If d_1 and d_2 are two distances in a space M, then the function d such that*

$$\begin{aligned} d : M * M &\rightarrow \mathbb{R}^+ \\ (x, y) &\mapsto d(x, y) = \sqrt{\gamma d_1(x, y)^2 + (1 - \gamma) d_2(x, y)^2} \end{aligned}$$

is a distance in M for all γ in $[0, 1]$.

See Proof in Appendix Section 6.3

The parameter γ is a weight of each distance. From the lemma, we can show that

Proposition 10. *The non-negative function $d_{V,O,\gamma}$ defined as*

$$\begin{aligned} d_{V,O,\gamma} : V * V &\rightarrow \mathbb{R}^+ \\ (R_{1,1}, R_{2,2}) &\mapsto d_{V,O,\gamma}(R_{1,1}, R_{2,2}) \\ \textit{such that} & \\ d_{V,O,\gamma}(R_{1,1}, R_{2,2}) &= \sqrt{\gamma d_{1,0}^2(R_{1,1}, R_{2,2}) + (1 - \gamma) d_{2,O}^2(R_{1,1}, R_{2,2})} \end{aligned}$$

is a metric for all $\gamma \in [0, 1)$ in a differentiable manifold V

In the decomposition we presented in Chapter 3, the distance in the manifold generated by the generalist-specialist and horizontal shift was with equal weights ($\gamma = \frac{1}{2}$) for each distance

$$d_{V,O}(R_{1,1}, R_{2,2}) = \sqrt{\frac{1}{2} d_{1,0}^2(R_{1,1}, R_{2,2}) + \frac{1}{2} d_{2,O}^2(R_{1,1}, R_{2,2})}$$

We chose equal weight because there was no a-priori reason why one path should be weighted more than the other path. We repeated the decomposition by changing the distance from full weight on one distance ($\gamma = 0$) to full weight the other distance ($\gamma = 1$) and the results were similar.

5.6.4 Mean, variance and decomposition

For a fixed origin $O = R_{0,0}$. Let $\tilde{R}_O = R_{\tilde{O}_1,\tilde{O}_2}$ be an estimate of a Fréchet mean (i.e in the Fréchet mean set) then we can see how to decompose the variation with the distance $d_{V,O}$

$$\begin{aligned}
\widetilde{SSM}_O &= \sum_{i=1}^{n} d^2_{V,O}(R_{i,i},\tilde{R}_O) \\
d^2_{V,O}(R_{i,i},\tilde{R}_O) &= \frac{1}{2} d^2_{1,0}(R_{i,i},\tilde{R}_O) + \frac{1}{2} d^2_{2,O}(R_{i,i},\tilde{R}_O) \\
d^2_{1,0}(R_{i,i},\tilde{R}_O) &= \text{Arcd}^2(R_{i,0},R_{\tilde{O}_1,0}) + \text{Arcd}^2(R_{0,i},R_{0,\tilde{O}_2}) \\
d^2_{2,0}(R_{i,i},\tilde{R}_O) &= \Big(\text{sign}(\alpha_i-\alpha_0)\text{Arcd}(R_{i,i},R_{0,i}) - \text{sign}(\tilde{O}_1-\alpha_0)\text{Arcd}(\tilde{R}_O,R_{0,\tilde{O}_2})\Big)^2 + \\
&\quad + \Big(\text{sign}(\beta_i-\beta_0)\text{Arcd}(R_{i,i},R_{i,0}) - \text{sign}(\tilde{O}_2-\beta_0)\text{Arcd}(\tilde{R}_O,R_{\tilde{O}_1,0})\Big)^2
\end{aligned}$$

The advantage of defining the distance $d_{V,O}$ by $d_{1,O}$ and $d_{2,O}$ is that we can reorganize the terms of the $\widetilde{SSM}_O$ such that

$$\widetilde{SSM}_O = \widetilde{SSM}_{1,O} + \widetilde{SSM}_{2,O}$$

Such that

$$\begin{aligned}\widetilde{SSM}_{1,O} &= \frac{1}{2}\sum_{i=1}^{n} \mathrm{Arcd}^2(R_{i,0}, R_{\tilde{O}_1,0}) + \\ &+ \frac{1}{2}\sum_{i=1}^{n} \Big(\mathrm{sign}(\alpha_i - \alpha_0)\mathrm{Arcd}(R_{i,i}, R_{0,i}) - \mathrm{sign}(\tilde{O}_1 - \alpha_0)\mathrm{Arcd}(\tilde{R}_O, R_{0,\tilde{O}_2})\Big)^2\end{aligned}$$

and

$$\begin{aligned}\widetilde{SSM}_{2,O} &= \frac{1}{2}\mathrm{Arcd}^2(R_{0,i}, R_{0,\tilde{O}_2}) + \\ &+ \frac{1}{2}\Big(\mathrm{sign}(\beta_i - \beta_0)\mathrm{Arcd}(R_{i,i}, R_{i,0}) - \mathrm{sign}(\tilde{O}_2 - \beta_0)\mathrm{Arcd}(\tilde{R}_O, R_{\tilde{O}_1,0})\Big)^2\end{aligned}$$

For a given origin O, the term $\widetilde{SSM}_{1,O}$ quantifies the variation along the mode parameterized by α and the term $\widetilde{SSM}_{2,O}$ quantifies the variation along the mode parameterized by β.

5.6.5 Consistency

Showing consistency in the general case is not trivial. The following theorem establishes consistency of the Fréchet mean estimates for a complete manifold, this results is shown in [Bhattacharya and Patrangenaru (2003)]

Theorem 5. *Let X be a random variable with measure μ on a metric manifold (M, d). For a finite Fréchet function, and for a complete metric manifold (M, d) in which every bounded closed set is compact, the estimates of the Fréchet mean set are consistent estimators, i.e*

$$\tilde{X}_n(set) \longrightarrow \tilde{X}_F(set)a.s$$

We use this result to show consistency of $\frac{1}{n}\widetilde{SSM}_O$ as an estimate of the Fréchet variance

$\text{Var}_{F,O}$ for a fixed origin O in the metric manifold $(V, d_{V,O})$.

Corrolary 2. *For all fixed origins O in a metric manifold $(V, d_{V,O})$*

$$\frac{1}{n}\widetilde{SSM}_O \longrightarrow \widetilde{Var}_{F,O} a.s$$

5.6.6 Choice of an origin

In the decomposition presented in Chapter 3, we chose an origin $\tilde{O} = R((\alpha_0, \beta_0), \mathbf{t})$ in a compact set K in V to be one which minimizes the Fréchet Variance, i.e

$$\begin{aligned} Var_{F,\tilde{O}} &= \min_{O\in K}\text{Var}_{F,O} \\ &\text{where} \\ Var_{F,O} &= \int d^2_{V,O}(R, \tilde{R}_{F,O})d\mu(R) \end{aligned}$$

where $\tilde{R}_{F,O}$ is an element of the Fréchet mean set. A question of interest is, can this origin be estimated from the data? Let $f_n(O)$ be the estimate of the Fréchet variance from the data, i.e

$$f_n(O) = \frac{1}{n}\widetilde{SSM}_{F,O} = \frac{1}{n}\sum_{i=1}^{n} d^2_O(R_{i,i}, \tilde{R}_{n,O})$$

where $\tilde{R}_{n,O}$ is the estimate of the Fréchet mean associated to the origin O and the data $R_{1,1}, \dots, R_{n,n}$. Let O_n be the estimate of $\tilde{O}$ origin from the data

$$f_n(O_n) = min_{O\in K} f_n(O)$$

We have from Corollary 2 that $f_n(O)$ converges almost surely to $Var_{F,O}$ for every origin O. Then, since O is in a compact set, this implies uniform convergence of (f_n) to $Var_{F,O}$. Since the sequence O_n is bounded, it has a converging subsequence. This lemma help us prove

Proposition 11, which proves convergence of (O_n) to $\tilde{O}$.

Lemma 2. *For a set of real functions (F_n) converging uniformly to a function F. Let θ_n be a convergent sequence of minimizers of F_n, then this sequence converges to a minimizer of F.*

Proposition 11. *The sequence (O_n) converges to a minimizer of the variance $\tilde{O}$.*

5.7 Multiple modes of variation

Case (7) defined in Section 5.1 is an example of three simultaneous modes of variation. This space is a particular case of a three dimensional manifold, it satisfies an equality of path property that we state in Subsection 5.7.1. We discuss in Subsection 5.7.2 an extension of the distance we defined in Section 5.6 to the most general case of d-dimensional manifolds of variation.

5.7.1 Three modes of variation, particular case

Because in case (7) one of the modes of variation is the vertical shift, the space of variation is created by a shifting nonlinear 2-dimensional manifold generated by the horizontal shift and generalist-specialist modes. Let V be a 3-dimensional manifold of variation parameterized by (α, β, γ) and let $R_{i,j,k} = R((\alpha_i, \beta_j, \gamma_k), \mathbf{t})$ be a point in the manifold. We denote the spaces of variation V_γ as the 2-dimensional subspaces of V such that

$$R \in V_\gamma \Leftrightarrow \exists(\alpha_i, \beta_i) \text{ such that } R = R((\alpha_i, \beta_i, \gamma), \mathbf{t})$$

We can define in V_γ a distance d_{V_γ} as defined in Section 5.5 or Section 5.6. We can use d_{V_γ} to define a distance d_V in V in particular cases such as (7). The following property states the property and the construction of d_V.

Proposition 12. *If V satisfies the following property,*

$$\begin{array}{rcl} & \textit{for all } (\alpha_1,\beta_1,\alpha_2,\beta_2,\gamma_1,\gamma_2) & \\ d_{V_\gamma}(R_{1,1,1},R_{2,2,1}) & = & C_{\alpha_1,\beta_1,\alpha_2,\beta_2} \\ Arcd(R_{1,1,1},R_{1,1,2}) & = & C_{\gamma_1,\gamma_2} \end{array}$$

where $C_{\alpha_1,\beta_1,\alpha_2,\beta_2}$ (resp. C_{γ_1,γ_2}) is a function of $(\alpha_1,\beta_1,\alpha_2,\beta_2)$ (resp. of (γ_1,γ_2)) and does not depend on γ_1 (resp. on $(\alpha_1,\beta_1,\alpha_2,\beta_2)$). Then the function d_V defined as

$$\begin{array}{rcl} d_V : V * V & \rightarrow & \mathbb{R}^+ \\ (R_{1,1,1},R_{2,2,2}) & \mapsto & d_V(R_{1,1,1},R_{2,2,2}) \textit{ where} \\ d_V(R_{1,1,1},R_{2,2,2}) & = & \sqrt{d^2_{V_\gamma}(R_{1,1,1},R_{2,2,1}) + Arcd^2(R_{1,1,1},R_{1,1,2})} \end{array}$$

is a metric in V.

As we saw in Sections 5.6 and 5.5 for the two dimensional case, the distance d_V allows for the decomposition $\widetilde{SSM}$ into three terms, each quantifying one mode. Using also the same theorem as in Section 5.5, we can show consistency of our estimates.

5.7.2 Discussion

For the general case of a 2-dimensional manifold, we found two possible linearizing functions, each preserving one (out of two) possible 2-steps paths between two points along the curves of variation. In the general d'-dimensional manifold, the number of linearizing functions is the number of possible d'-step paths between two points along the curves of variations which is $d'!$. For each possible path i, we can define a linearization $L_{i,O}$, which allows us to define a metric $d_{i,O}$. We chose the metric in the manifold to be d_O such that for any two points R_1 and R_2 in the manifold

$$d^2_O(R_1,R_2) = \frac{1}{d'!}\sum_{i=1}^{d'!} d^2_{i,O}(R_1,R_2)$$

5.8 Data with errors

We have discussed so far how to estimate the variation in a manifold of variation. In practise, the data does not lie in the manifold but is projected onto the manifold, i.e the data $z_i \in \mathbb{R}^d$ is such that

$$z_i = R_i + \epsilon_i$$

where R_i is the projection of z_i onto the manifold and ϵ_i is the additive error. In practise, the projection is the point in the manifold which is closest to the data in $\mathbb{R}^d$, i.e minimizes SSE such that

$$||z_i - R_i|| = \min_{x \in V} ||z_i - x||$$

This projection might not be unique. We saw in Section 3.5 several visual assessments of the fit that would help detect multiplicity of fits. We can also use local information to choose between multiple projections, so that for data which is close together in $\mathbb{R}^d$, their projections onto V should be close together.

Chapter 6

Appendix

Here are the proofs of results cited in Chapter 3, 4, and 5. The Proposition is restated before each proof.

6.1 Optimization

6.1.1 Fitting algorithm

In Chapter 3, we fit the data $z_i(t_j)$ by the polynomial fit $\tilde{z}_i(t_j)$ such that

$$\tilde{z}_i(t_j) = w_i P(w_i(t_j - m_i)) + h_i$$

where P is a polynomial of degree k, and (w_i, m_i, h_i) are respectively the width, the location and the height parameter of the ith individual. We optimize for the coefficients of the polynomial P which are common to all curves and for the three parameters (w_i, m_i, h_i) which are specific to each curve. Let n be the number of families and d be the dimension of the vector $\mathbf{t}$, the criterion that we use for optimization is minimizing the weighted SSE (weighted by the sample size n_i in each family) such that

$$SSE = n * \sum_{i=1}^{n} \sum_{j=1}^{d} \frac{n_i}{\sum_{k=1}^{n} n_k} ||z_i(t_j) - w_i P(t_j - m_i) - h_i||^2$$