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

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

Our algorithm does this optimization sequentially, with one step optimizing for the best polynomial coefficients over all curves and the next step optimizing for the best parameters (w_i, m_i, h_i) in each curve,i.e

1. Initialization: $k = 0$, choose an initial set of coefficients of the common polynomial, call it P_0, and initial parameters $(w_{i,0}, m_{i,0}, h_{i,0})$ for each curve i.

2. algorithm step $k \geq 1$.

   - Find the set of optimal coefficients of the common polynomial at the kth iteration, call it P_k (optimization with initial value P_{k-1}) such that

   $$P_k = \text{Argmin}_P n \sum_{i=1}^{n} \sum_{j=1}^{d} \frac{n_i}{\sum_{k=1}^{n} n_k} ||z_i(t_j) - w_{i,k-1} P(t_j - m_{i,k-1}) - h_{i,k-1}||^2$$

   - Find the parameters of variation $(w_{i,k}, m_{i,k}, h_{i,k})$ for each curve i (optimization with initial value $(w_{i,k-1},, m_{i,k-1}, h_{i,k-1})$ for each curve i), such that

   $$\forall i, (w_{i,k}, m_{i,k}, h_{i,k}) = \text{Argmin}_{(w,m,h)\in\Theta} \sum_{j=1}^{d} \frac{n_i}{\sum_{k=1}^{n} n_k} ||z_i(t_j) - wP_k(t_j - m) - h||^2$$

   - Let $SSE_k = n \sum_{i=1}^{n} \sum_{j=1}^{d} \frac{n_i}{\sum_{k=1}^{n} n_k} ||z_i(t_j) - w_{i,k} P_k(t_j - m_{i,k}) - h_{i,k}||^2$

3. If $|SSE_k - SSE_{k-1}|$ less than a chosen small threshold, then let $K = k$ and go to step 4. Otherwise, go to step 2.

4. The fitted values are $w_{i,K}$ for width parameter, $m_{i,K}$ for location, and $h_{i,K+1}$ for height. Where $h_{i,K+1}$ satisfies the identifiability condition $\sum_{i=1}^{K} h_{i,K+1} = 0$. To find $h_{i,K+1}$, let $\alpha = -\frac{\sum_{i=1}^{n} h_{i,K}}{\sum_{i=1}^{n} w_{i,K}}$, then the fitted $h_{i,K+1}$'s are

$$h_{i,K+1} = h_{i,K} + \alpha * w_{i,K}$$

See Subsection 6.1.3 for details of why this gives identifiability.

6.1.2 Algorithm for estimating distances along the manifold

To estimate all metrics we discussed in Chapter 5, it is sufficient to estimate arc-distances along each mode. Estimation of arc-distances is done by linear approximation, i.e for some high number N and a parameter of variation θ,

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

6.1.3 Identifiability of the parameters

We will prove that finding optimal coefficients of a polynomial of degree 3 or higher and coefficients (w, m, h) is an identifiable problem assuming that $\sum_{i=1}^{n} h_i = 0$. Let P_1 and P_2 be the two polynomials of degree $k \geq 3$, and dimension $d \geq k+1$ such that,

$$P_m(t) = \sum_{l=0}^{k} \alpha_{m,l} t^l, m = 1, 2$$

We have to show that if

$$w_i P_1(w_i(t_j - m_i)) + h_i = w_i' P_2(w_i'(t_j - m_i')) + h_i', \forall i \leq n; j \leq d$$

and

$$\sum_{i=1}^{n} h_i = \sum_{i=1}^{n} h_i'$$

then

$$P_1 = P_2, (w_i = w_i'), (m_i = m_i'), \text{ and } (h_i = h_i')$$

From the equality of polynomials, we have

$$\alpha_{1,k}w_i^{k+1} = \alpha_{2,k}w_i'^{k+1} \tag{6.1}$$

$$\alpha_{1,k-1}w_i^{k} - k\alpha_{1,k}w_i^{k+1}m_i = \alpha_{2,k-1}w_i'^{k} - k\alpha_{2,k}w_i'^{k+1}m_i' \tag{6.2}$$

$$\text{for all } s \geq 2, \sum_{l=0}^{s} C_{s-l}^{k-l}\alpha_{1,k-l}w_i^{k-l+1}(-m_i)^l = \sum_{l=0}^{s} C_{s-l}^{k-l}\alpha_{2,k-l}w_i'^{k-l+1}(-m_i')^l \tag{6.3}$$

$$\sum_{l=0}^{k} C_{k-l}^{k-l}\alpha_{1,k-l}w_i^{k-l+1}(-m_i)^l + h_i = \sum_{l=0}^{k} C_{k-l}^{k-l}\alpha_{2,k-l}(w_i')^{k-l+1}(-m_i')^l + h_i' \tag{6.4}$$

From equality (6.1) we have that

$$\exists\lambda \text{ such that } w_i' = \lambda w_i, \forall i$$

and

$$\alpha_{2,k} = \alpha_{1,k}(\lambda)^{-(k+1)}$$

Using this result, we have from equality (6.2)

$$\begin{aligned} m_i' &= -\frac{(\alpha_{1,k-1} - \lambda^k\alpha_{2,k-1})}{k\lambda^{k+1}\alpha_{2,k}w_i} + \frac{\alpha_{1,k}}{\alpha_{2,k}\lambda^{k+1}}m_i' \\ &= m_i - \frac{\alpha_{1,k-1} - \lambda^k\alpha_{2,k-1}}{k\alpha_{1,k}}\frac{1}{w_i} \end{aligned}$$

If we use those two identities in equations (3), we find that m_i could be written as a function of w_i for all i and this function differs from equation to equation. By substitution we show that

$$\lambda = 1, \text{so} w_i = w_i', m_i = m_i' \text{ and } \alpha_{1,s} = \alpha_{2,s}, \forall 1 \leq s \leq k. \tag{6.5}$$

Finally using this result, in (6.5), in the equation 6.4 we have

$$h_i' = h_i + (\alpha_{1,0} - \alpha_{2,0})w_i$$

or

$$h_i' = h_i + (\alpha_{1,0} - \alpha_{2,0})w_i'$$

If $\sum_{i=1}^n h_i = 0$ and $\sum_{i=1}^n h_i' = 0$, then this last equation implies that

$$(\alpha_{1,0} - \alpha_{2,0})\bar{w} = 0$$

Since $\bar{w} \neq 0$, we have that

$$\alpha_{1,0} = \alpha_{2,0} \text{ and } h_i' = h_i.$$

□

6.2 Toy Example

In section 6.2.2 we will identify the linear spaces of variations of the vertical shift (or faster-slower) mode as defined in Chapter 3. Those results hold for any family of curves with a common template shape. In section 6.2.3, we will identify the horizontal shift (or hotter-colder) variation for the parabola template shape toy example used in Chapter 4. We will show that in the point cloud representation, the data with this variation lie in a parabola curve in an affine space. The characterization of the affine space is given in Proposition 15, and the equation of the parabola curve is derived in Proposition 16. Finally, the affine space and the equation of the parabola are derived for the special case of equally spaced temperature measurements.

6.2.1 Notations, and assumptions

We will use the following notation in this section of the appendix

- d is the dimension of the problem, it is the number of distinct temperature measurements. i.e in the caterpillar example $d = 6$.

- n is the number of families, i.e in the caterpillar example $n = 29$.

- T is the temperature domain (interval), and t is the individual or the vector of temperature measurements.

We will suppose that we have a family of curves f_i with a common template shape f_0. The vector ft is the value of the function f at the temperatures $(t_1, \ldots, t_d)$. i.e. $ft = (f(t_1), \ldots, f(t_d))$.

6.2.2 Vertical shift

The proposition 13 will state that for any family of curve with a non constant template shape f_0 and a vertical shift variation, the data will fall in the line generated by the d-vector $\mathbb{I}_d$, $\mathbb{I}_d = (1, 1, \ldots, 1)$ in the point cloud representation.

Proposition 13. *Let ft_0 be a non constant d-vector.i.e $ft_0 \neq \bar{f}t_0 * \mathbb{I}_d$. Let $a = (a_i)_{i\in\{1,\ldots,n\}}$ and $h = (h_i)_{i\in\{1,\ldots,n\}}$ be two n vectors. Then, given a family of curves $a_i ft_0 + h_i\mathbb{I}_d$, for $i \in \{1, \ldots, n\}$. The Var-Cov matrix of this family is at most two dimensional. In particular*

- *For the Faster-Slower variation $a = \bar{a} * \mathbb{I}_d$, and $h \neq \bar{h} * \mathbb{I}_d$, then the Var-Cov matrix is one dimensional and $\mathbb{I}_d$ is the eigenvector.*

Proof

Let denote the data by the $d * n$ matrix X

$$X = ft_0 * a' + \mathbb{I}_d * h'$$

and use the following notations

$$\begin{aligned}
\bar{a} &= \frac{1}{n}\sum_{i=1}^{n} a_i, \text{ and} \\
\mathrm{S}^2(a) &= \frac{1}{n}(a - \bar{a} * \mathbb{I}_n)'(a - \bar{a} * \mathbb{I}_n), \text{ likewise} \\
\bar{h} &= \frac{1}{n}\sum_{i=1}^{n} h_i, \text{ and} \\
\mathrm{S}^2(h) &= \frac{1}{n}(h - \bar{h} * \mathbb{I}_n)'(h - \bar{h} * \mathbb{I}_n), \text{ and} \\
\mathrm{r}^2(a,h) &= \frac{1}{n}(h - \bar{h} * \mathbb{I}_n)'(a - \bar{a} * \mathbb{I}_n). \text{ Moreover} \\
ft_c &= ft_0 - \bar{ft}_0\mathbb{I}_d, ft_c \perp \mathbb{I}_n.
\end{aligned}$$

Then

$$\begin{aligned}
\bar{X} &= \frac{1}{n}X * \mathbb{I}_n \\
&= \bar{a}ft_0 + \bar{h} * \mathbb{I}_d
\end{aligned}$$

And let X_c be the centered data

$$\begin{aligned}
X_c &= X - \bar{X} * \mathbb{I}_n' \\
&= ft_0 * (a - \bar{a} * \mathbb{I}_n)' + \mathbb{I}_d * (h - \bar{h} * \mathbb{I}_n)' \\
&= ft_c * (a - \bar{a} * \mathbb{I}_n)' + \mathbb{I}_d * \big(h - \bar{h} * \mathbb{I}_n) + \bar{ft}_0(a - \bar{a} * \mathbb{I}_n)\big)' \\
\mathrm{S}^2(X) &= X_c * X_c' \\
&= n\mathrm{S}^2(a)ft_c * ft_c' + n\left(\bar{ft}_0{}^2\mathrm{S}^2(a) + \mathrm{S}^2(h) + 2\bar{ft}_0\mathrm{r}^2(a,h)\right)\mathbb{I}_d * \mathbb{I}_d'.
\end{aligned}$$

Since the two vectors ft_c and $\mathbb{I}_d$ are orthogonal to each other, the results of proposition 13 follow from this last decomposition.

6.2.3 Horizontal shift in the parabolic case

When $d = 3$, the growth rate ft was measured at three distinct temperatures t_1, t_2, and t_3. Let f_m be a downward open parabola with maximum at m, the location parameter. Let a and h be the scaling parameters of the parabola. For the Hotter-Colder variation, a and h are common to all families and m varies for each family.

$$
\begin{aligned}
t &= \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}, \\
t^2 &= \begin{pmatrix} t_1^2 \\ t_2^2 \\ t_3^2 \end{pmatrix}, \text{and} \\
ft_m &= \begin{pmatrix} f_m(t_1) \\ f_m(t_2) \\ f_m(t_3) \end{pmatrix} = -a(m^2 * (1,1,1)' - 2m * t + t^2) + h * (1,1,1)'.
\end{aligned}
$$

A specific orthogonal basis (u_1, u_2, u_3) in this three dimensional space is defined by:

$$
\begin{aligned}
u_1 &= \begin{pmatrix} -2t_1 + t_2 + t_3 \\ t_1 - 2t_2 + t_3 \\ t_1 + t_2 - 2t_3 \end{pmatrix} = H_1 * t \\
u_2 &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\
u_3 &= \begin{pmatrix} t_2 - t_3 \\ t_3 - t_1 \\ t_1 - t_2 \end{pmatrix} = H_3 * t.
\end{aligned}
$$

Where

$$
\begin{aligned}
H_1 &= \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}. \\
H_3 &= \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}.
\end{aligned}
$$

The vector u_2 is independent of the data whereas the vectors u_1 and u_3 depend on the temperature measurements. However, u_1 and u_3 depend only on the spacing between the temperature measurements, so their expression is simplified in the equally spaced case (see section 6.2.3).

Proposition 14. (u_1, u_2, u_3) *is an orthogonal basis of* $\mathbb{R}^3$.

Proof

$$
\begin{aligned}
H_3' * H_1 &= 0, \\
H_1' * u_2 &= 0, \text{and} \\
H_3' * u_3 &= 0.
\end{aligned}
\tag{6.6}
$$

Then the three vectors are orthogonal to each other. So (u_1, u_2, u_3) is a basis of the three dimensional space. □

Some properties of H_1 and H_3

I will use the following properties in the following derivations of Proposition 16 and Proposition 15 and Corollary 3.

- H_1 is symmetric, and its H_1^2 is a multiple of H_1.

$$
\begin{aligned}
H_1' &= H_1, \\
H_1^2 &= -3H_1.
\end{aligned}
$$

- Some properties of H_3 are

$$
\begin{aligned}
H_3' &= -H_3, \\
u' * H_3 * u &= 0 \text{ for all } u \text{ in } \mathbb{R}^3.
\end{aligned}
$$

- H_1 and H_3 are orthogonal to each other,

$$\begin{aligned} H_3 * H_1 &= 0, \text{ and related by the equation} \\ H_3^2 &= H_1. \end{aligned}$$

- From the four preceding equations, we have that

$$\begin{aligned} \|u_1\|^2 &= 3\|u_3\|^2, \\ \|u_2\|^2 &= 3, \\ \|u_3\|^2 &= -t' * H_1 * t. \end{aligned}$$

Definition 6. *The hotter-colder basis is defined by* $(hc_1, hc_2, hc_3) = \left(\frac{u_1}{\|u_1\|}, \frac{u_2}{\|u_2\|}, \frac{u_3}{\|u_3\|}\right)$. *It is an orthonormal basis of* $\mathbb{R}^3$.

Proposition 15. *The vector* ft_m *is in the affine space perpendicular to* hc_3 *at the point* $(0, 0, k)_{(hc_1, hc_2, hc_3)}$, *where*

$$k = -a(hc_3' * t^2)$$

We call this affine space the hotter-colder affine space.

Proof

$$\begin{aligned}
u_3' * ft_m &= t' * H_3 * \big(-a * (m^2(1,1,1)' - 2mt + t^2) + h * (1,1,1)\big) \\
&\quad \text{since } H_3 * (1,1,1)' = 0, \\
u_3' * ft_m &= 2a(t' * H_3 * t)m - a(t' * H_3 * t^2) \\
&\quad \text{but } t' * H_3 * t = 0, \text{ so} \\
u_3' * ft_m &= -a(u_3' * t^2). \\
&\quad \text{Finally} \\
hc_3' * ft_m &= -a(hc_3' * t^2), \text{ which is a constant independent of } m.
\end{aligned}$$

□

Proposition 16. *ft_m lie in a parabola of equation $Y = A*(X-M)^2 + H$ in the hotter-colder affine space. X (respectively Y) is the projection on hc_1 (respectively hc_2), and the parameters A, M, and H are*

$$\begin{aligned}
A &= -\frac{9\sqrt{3}}{4a\|u_1\|^2}, \\
M &= -a\left(\frac{2}{3}\|u_1\|\bar{t} + (hc_1' * t^2)\right), \textit{ and} \\
H &= \sqrt{3}\left(h - a\left(\bar{t^2} - \bar{t}^2\right)\right).
\end{aligned}$$

Proof

$$
\begin{aligned}
u_1' * ft_h &= t' * H_1 * ft_h \\
&= t' * H_1 * (-a * (t - m(1,1,1)')^2 + h(1,1,1)') \\
&= -t' * H_1 * a * (t - m(1,1,1)')^2, \\
&\quad \text{because } H_1 * (1,1,1)' = 0, \\
&= -a * t' * H_1 * (m^2(1,1,1)' - 2m * t + t^2) \\
&= 2a(t' * H_1 * t)m - a(t' * H_1 * t^2) \\
&= 2a\left(-\frac{\|u_1\|^2}{3}\right) m - a(t' * H_1 * t^2) \\
&= -\left(\frac{2}{3}a\|u_1\|^2\right) m - a(u_1 * t^2).
\end{aligned}
$$

So,

$$
\begin{aligned}
X &= hc_1' * ft_h = -(\frac{2}{3}a\|u_1\|)m - aC_1, \text{ and} \\
X^2 &= \left(\frac{2}{3}a\|u_1\|\right)^2 m^2 + \left(\frac{4}{3}a^2\|u_1\|C_1\right) m + a^2C_1^2.
\end{aligned}
$$

Where $C_1 = hc_1' * t^2$. So, for all A, M, and H

$$
\begin{aligned}
A(X-M)^2 + H &= AX^2 - 2AMX + AM^2 + H \\
&= A\left(\frac{4}{9}a^2\|u_1\|^2\right) m^2 + \left(\frac{4}{3}A\|u_1\|a\right)(aC_1 + M)m \\
&\quad + Aa^2C_1^2 + 2aAMC_1 + AM^2 + H. \qquad (6.7)
\end{aligned}
$$

On the other hand,

$$\begin{aligned} Y = e_2' * ft_h &= \frac{1}{\sqrt{3}}\left((1,1,1) * (-a * (t - m(1,1,1)')^2 + h(1,1,1)')\right) \\ &= -\frac{1}{\sqrt{3}}\left(a * (1,1,1) * (m^2(1,1,1)' - 2m * t + t^2) + 3h\right) \\ &= -\frac{1}{\sqrt{3}}\left((3a)m^2 + (6a\bar{t})m - 3a * \bar{t^2} + 3h\right). \end{aligned}$$

So,

$$Y = hc_2' * ft_h = -(a\sqrt{3}) * m^2 + (2a\bar{t}\sqrt{3}) * m + C_2. \tag{6.8}$$

Where $C_2 = \sqrt{3}(h - a\bar{t^2})$.

We are looking for A, M, and H such that $hc_2' * ft_h = A(X - M)^2 + H$ for all m in $\mathbb{R}$. Then, by identification of the two equations 6.7 and 6.8

$$\begin{aligned} A\left(\tfrac{4}{9}a^2\|u_1\|^2\right) &= -(a\sqrt{3}) \\ \left(\tfrac{4}{3}A\|u_1\|a\right)(aC_1 + M) &= (2a\bar{t}\sqrt{3}) \\ Aa^2C_1^2 + 2aAMC_1 + AM^2 + H &= C_2. \end{aligned} \tag{6.9}$$

So,

$$\begin{aligned} A &= -\frac{\sqrt{3}a}{\left(\frac{2}{3}a * \|u_1\|\right)^2} = -\frac{9\sqrt{3}}{4a\|u_1\|^2}. \\ M &= \frac{2a\bar{t}\sqrt{3}}{\frac{4}{3}A\|u_1\|a} - aC_1 = -a\left(\frac{2}{3}\|u_1\|\bar{t} + (hc_1' * t^2)\right). \end{aligned}$$

And finally,

$$
\begin{aligned}
H &= C_2 - Aa^2C_1^2 - 2aAMC_1 - AM^2 \\
&= \sqrt{3}(h - a\bar{t^2}) + \frac{9a\sqrt{3}}{4\|u_1\|^2}(hc_1' * t^2)^2 - \frac{9a\sqrt{3}}{2\|u_1\|^2}\left(\frac{2}{3}\|u_1\|\bar{t} + (hc_1' * t^2)\right)(hc_1' * t^2) \\
&\quad + \frac{9a\sqrt{3}}{4\|u_1\|^2}\left(\frac{2}{3}\|u_1\|\bar{t} + (hc_1' * t^2)\right)^2 \\
&= \sqrt{3}\left(h - a\left(\bar{t^2} - \bar{t}^2\right)\right).
\end{aligned}
$$

□

Particular case: equally spaced measurements

Let c be the spacing between temperatures. Then, the expression of (hc_1, hc_2, hc_3) is simplified.

$$
\begin{aligned}
t &= \begin{pmatrix} t_1 \\ t_1 + c \\ t_1 + 2c \end{pmatrix} \\
t^2 &= \begin{pmatrix} t_1^2 \\ t_1^2 + c^2 + 2t_1c \\ t_1^2 + 4c^2 + 4t_1c \end{pmatrix} \\
hc_1 &= \frac{\sqrt{2}}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \\
hc_2 &= \frac{\sqrt{3}}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\
hc_3 &= \frac{\sqrt{6}}{6} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \\
\|u_1\| &= 3c\sqrt{2}.
\end{aligned}
$$

Corrolary 3. *When the temperatures are equally spaced*

$$
\begin{aligned}
k &= \frac{ac^2\sqrt{6}}{3} \\
A &= -\frac{\sqrt{3}}{8c^2a} \\
M &= 0 \\
H &= \sqrt{3}\left(h - a\left(\overline{t^2} - \bar{t}^2\right)\right).
\end{aligned}
$$

Proof

From the previous two propositions for the general case, we have that.

$$
\begin{aligned}
k &= -a(hc_3' * t^2), \\
A &= -\frac{9\sqrt{3}}{4a\|u_1\|^2}, \\
H &= \sqrt{3}\left(h - a\left(\overline{t^2} - \bar{t}^2\right)\right).
\end{aligned}
$$

So by in the equally spaced particular case,

$$
\begin{aligned}
k &= -\frac{a\sqrt{6}}{6}((-1,2,-1) * t^2) = \frac{c^2a\sqrt{6}}{3}, \\
A &= -\frac{9\sqrt{3}}{4a(3c\sqrt{2})^2} = \frac{\sqrt{3}}{8c^2a}, \\
H &= \sqrt{3}\left(h - a\left(\overline{t^2} - \bar{t}^2\right)\right).
\end{aligned}
$$

Also from the previous proposition:

$$
M = -a\left(\frac{2}{3}\|u_1\|\bar{t} + (hc_1' * t^2)\right).
$$

Using that

$$hc_1' * t^2 \quad = \quad -2\sqrt{2}c(c+t_1).$$

And

$$\|u_1\|\bar{t} \quad = \quad (3c\sqrt{2})(t_1+c).$$

We finally have, $M = 0.\square$

6.3 Mathematical Foundation

We restate in this section the theorems, propositions and lemmas from Chapter 5 before we show the proof

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'.

Proof V is a subspace of $\mathbb{R}^{d'}$, it is a topological space, its topology is induced by the topology in $(\mathbb{R}^d, ||.||)$. Since $\mathbb{R}^{d'}$ is separable, and $R(.,\mathbf{t})$ is an homeomorphism then V is also separable. Since $R(.,\mathbf{t})$ is a homeomorphism from $(\mathbb{R}^{d'}, ||.||)$ to $(V, ||.||)$, then all neighborhood Ω in V is homeomorphic to a neighborhood in $\mathbb{R}^{d'}$. $\square$

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.

Proof Let t be a d-vector in $\mathbb{R}^d$, we can write the equation $P_k(t) = q$ where P_k is a polynomial of degree k as a linear equation $T_k\alpha = q$, where T_k is a $d*(k+1)$ matrix such as $(T_k)_{i,j} = t_i^{j-1}$ for i in 1 to d and j from 1 to $k+1$, i.e

$$T_k = \begin{bmatrix} 1 & t_1 & \dots & t_1^k \\ 1 & t_2 & \dots & t_2^k \\ \dots & \dots & \dots & \dots \\ 1 & t_d & \dots & t_d^k \end{bmatrix}$$

α is the $(k+1)*1$ vector of coefficients of the polynomial P_k. If the coordinates of t are different from each other, than the matrix T_k is of dimension $k+1$ (could show that by induction). So if the two polynomials of degree k: P_k and Q_k agree on d points $(t_1, \dots, t_d)$ then

$$T_k\alpha_P = T_k\alpha_Q \tag{6.10}$$

Since T_k is of dimension $k+1$ then Equation 6.10 results in

$$\alpha_P = \alpha_Q.$$

Similarly if $T_k\alpha_n$ converges to $T_k\alpha$ then α_n converges to α. □

Corollary 1 *When z is a polynomial of degree k and $2 \le k \le d-1$ (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.*

Proof V is a subspace of $\mathbb{R}^d$. For the one dimensional cases (1), (2) and (3), i.e when θ is one parameter. V is a line when V is the vertical shift space of variation, and a curve when V is the horizontal shift space of variation or when V is the generalist-specialist space of variation. To prove that V is a one dimensional manifold by Theorem 1, we only need to prove that

$$\theta \mapsto R_\theta$$

is a homeomorphism from $\mathbb{R}$ to V, i.e R_θ is continuous, invertible and its inverse is continuous.
Invertibility: We need to show that for any $t \in \mathbb{R}^d$,

$$R_{\theta_1}(t) = R_{\theta_2}(t) \Leftrightarrow \theta_1 = \theta_2. \tag{6.11}$$

Since z is a polynomial of degree k, $R_{\theta_1}(t)$ and $R_{\theta_2}(t)$ are also polynomials of degree k on t in all situations (1)-(7) defined in Section 5.1. Because the two polynomials agree at d points, i.e $t \in \mathbb{R}^d$, and d is greater than $k+1$, the two polynomials have the same coefficients. So, in particular by looking at the coefficients of the three highest degrees we have

1. $\theta_i = h_i$. If $R_{h_1}(t) = R_{h_2}(t)$ then $z(t) + h_1 = z(t) + h_2$, so $h_1 = h_2$ for any t and any k.

2. $\theta_i = m_i$. Then by Equation 6.11 we have that $z(t - m_1) = z(t - m_2)$, with highest coefficients (the one for power $k-1$):

$$c\alpha_k * m_1 + \alpha_{k-1} = c\alpha_k * m_2 + \alpha_{k-1}$$

where c in a nonzero constant. So $m_1 = m_2$. This only requires $k \geq 1$, and the equality of the second highest coefficients which is guaranteed if $d > k$.

3. $\theta_i = w_i$. From Equation 6.11, we have that $w_1 z(w_1 t) = w_2 z(w_2 t)$. From the equality of the highest coefficients it follows that $w_1^{k+1} = w_2^{k+1}$. Since w is positive, this implies $w_1 = w_2$.

4. $\theta_i = (m_i, h_i)$ then,

$$z(t - m_1) + h_1 = z(t - m_2) + h_2. \tag{6.12}$$

By equality of the $(k-1)$th coefficient, we have that

$$c\alpha_k * m_1 + \alpha_{k-1} = c\alpha_k * m_2 + \alpha_{k-1}$$

so $m_1 = m_2$. When we substitute this last equality in Equation 6.12, we have $h_1 = h_2$.

5. $\theta_i = (w_i, h_i)$ and

$$w_1 z(w_1 t) + h_1 = w_2 z(w_2 t) + h_2 \tag{6.13}$$

then similarly to situation 3 (i.e $\theta = w$), we find by equality of highest coefficients that $w_1 = w_2$. We substitute this in Equation 6.13 and we have $h_1 = h_2$.

6. $\theta_i = (m_i, w_i)$, we find by equality of the two highest coefficients that $w_1 = w_2$ and $m_1 = m_2$.

7. $\theta_i = (w_i, m_i, h_i)$. Then

$$w_1 z(w_1(t - m_1)) + h_1 = w_2 z(w_2(t - m_2)) + h_2$$

For $k \geq 2$ and equality of the two highest coefficients and the constant term we have that $(h_1, m_1, w_1) = (h_2, m_2, w_2)$.

Continuity: Let d' be the dimension of the vector θ, i.e d' is 1, 2, or 3.

Showing the continuity of $\theta \mapsto R_\theta(t)$ for any t is equivalent to showing that as η converges to 0 in $(\mathbb{R}^{d'}, ||.||)$, $R_{\theta+\eta}$ converges to R_θ in $(V, ||.||)$. Since the function z is continuous in all cases (1)-(7), then the continuity of R_θ follows immediately.

Continuity of the inverse: is equivalent to show that if $R_{\theta+h}$ converges to R_θ, then h converges to 0. We have this last property by Proposition 1. □

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

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

Proof

(i) Continuity of L_{θ_0}: If $\mathrm{Arcd}(x_n, x)$ converges to 0, then by isometry $|L_{\theta_0}(x_n) - L(x)|$ converges to 0. So, L_{θ_0} is continuous.

(ii) Invertibility of L_{θ_0}: We need to show that $L_{\theta_0}(x_1) = L_{\theta_0}(x_2) \Rightarrow x_1 = x_2$. By isometry the left hand side implies that $\mathrm{Arcd}(x_1, x_2) = 0$, which implies that $x_1 = x_2$ (because Arcd is a metric in V).

(ii) Continuity of the inverse of L_{θ_0}: We need to show that

$$|L_{\theta_0}(x_n) - L_{\theta_0}(x)| \to 0 \Rightarrow \mathrm{Arcd}(x_n, x) \to 0$$

. This property follows directly from the isometry.

□

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

$$\begin{aligned}\tilde{R} &= L_{\theta_0}^{-1}(\overline{L_{\theta_0}}) \text{ for all } \theta_0 \\ \text{where} & \\ \overline{L_{\theta_0}} &= \frac{1}{n}\sum_{i=1}^{n} L_{\theta_0}(R_i)\end{aligned}$$

Proof The estimate of the Fréchet function $F_n(R)$ is

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

By isometry,

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

We have that

$$\mathrm{Argmin}_{x\in\mathbb{R}} \sum_{i=1}^{n}(L_{\theta_0}(R_i) - x)^2 = \overline{L_{\theta_0}}$$

where,

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

Let $R_0 = L_{\theta_0}^{-1}(\overline{L_{\theta_0}})$, we can show that R_0 is an estimate of the Fréchet mean set,

$$\begin{aligned}\forall R, F_n(R) &\geq \sum_{i=1}^{n}(L_{\theta_0}(R_i) - \overline{L_{\theta_0}}), \text{ and by isometry,} \\ &\geq \sum_{i=1}^{n} \mathrm{Arcd}^2(R_i, R_0)\end{aligned}$$

So, R_0 is a minimizer (then it is a Fréchet mean estimate). On the other hand, let $\tilde{R}$ be a Fréchet mean estimate. If $\tilde{R} \neq R_0$, then $L_{\theta_0}(\tilde{R}) \neq \overline{L_{\theta_0}}$, so

$$F_n(\tilde{R}) > F_n(R_0)$$

which is a contradiction (since both $\tilde{R}$ and R_0 are Fréchet mean estimates, $F_n(\tilde{R}) = F_n(R_0)$). So, the Fréchet mean estimate is unique. □.

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*

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

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

Proof We will first show that the Fréchet mean is unique. We have that

$$E(L_{\theta_0}(R)) = \text{Argmin}_{x \in \mathbb{R}} \int (L_{\theta_0}(R) - x)^2 d\mu$$

Let $R_0 = L_{\theta_0}^{-1}(E(L_{\theta_0}(R_i)))$, we can show that R_0 is the Fréchet mean. Let $F(R)$ be the Fréchet function, then

$$\begin{aligned} \forall R \neq R_0, F(R) &> \int (L_{\theta_0}(R_i) - E(L_{\theta_0}(R)))^2 d\mu, \text{ and by isometry,} \\ &> \int \text{Arcd}^2(R_i, R_0) d\mu \end{aligned}$$

So, R_O is the only element in the Fréchet mean set. Since by the law of large numbers,

$$\overline{L_{\theta_0}} \rightarrow E(L_{\theta_0}(R))(a.s) \tag{6.14}$$

then by isometry of L_{θ_0}, we have that

$$d_V(\tilde{R}, \tilde{R}_F) \to 0 \text{ (a.s)}$$

Similarly, since $(L_{\theta_0}(R_i) - E(L_{\theta_0}(R_i)))^2$ are i.i.d and of finite mean Var_F, then by law of large numbers

$$\left| \frac{1}{n} \sum (L_{\theta_0}(R_i) - E(L_{\theta_0}(R_i)))^2 - \text{Var}_F \right| \to 0 \text{ (a.s)} \tag{6.15}$$

On the other hand,

$$\frac{1}{n}\widetilde{SSM} - \text{Var}_F = \left(\frac{1}{n} \sum (L_{\theta_0}(R_i) - E(L_{\theta_0}(R_i)))^2 - \text{Var}_F \right) + (\bar{L_{\theta_0}} - E(L_{\theta_0}(R)))^2 \tag{6.16}$$

From equality (6.16), and using the two convergence results in (6.15) and (6.14), we have that

$$\left| \frac{1}{n}\widetilde{SSM} - \text{Var}_F \right| \to 0 \text{ (a.s)}$$

□

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

Proof Without loss of generality, we will show the equality of path condition for case (5) because the proof for case (6) is the same.

(i) For all $h \in \mathbb{R}$,

$$\frac{\partial}{\partial m}(z(t-m)+h) = -z'(t-m)$$

Without loss of generality, let $m_1 \leq m_2$. Then, by definition,

$$\text{Arcd}((z(t_j - m_1) + h)_j, (z(t_j - m_2) + h)_j) = \int_{m_1}^{m_2} \frac{\partial}{\partial m} ||(z(t_j - m) + h)_j|| dm$$

By combining the two previous equality, we conclude that the value of $\text{Arcd}((z(t_j - m_1))_j + h * (1, \ldots, 1), (z(t_j - m_2))_j + h * (1, \ldots, 1))$ depends only on the value of m_1 and m_2 and not on the value of h.

(ii) For all $m, h_1, h_2 \in \mathbb{R}$,

$$||(z(t_j - m) + h_1)_j - (z(t_j - m) + h_2)_j|| = d|h_1 - h_2|$$

The right hand side depends only on the values of h_1 and h_2 and not on the value of m.

From the results in (i) and (ii), we have that case (5) satisfies the property of equality of paths. $\square$.

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*

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

is a distance in V

Proof By the equality of path property, we have that

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

(i) By definition, d_V is a non-negative function.

(ii) d_V is symmetric because C_{α_1,α_2} and C_{β_1,β_2} are symmetric.

(iii) If $d_V\left(R((\alpha_1,\beta_1),\mathbf{t}),R((\alpha_2,\beta_2),\mathbf{t})\right)=0$, then

$$C_{\alpha_1,\alpha_2}=C_{\beta_1,\beta_2}=0.$$

So,

$$\begin{aligned}\text{Arcd}\left(R((\alpha_1,\beta),\mathbf{t}),R((\alpha_2,\beta),\mathbf{t})\right) &= 0\forall\beta\\ \text{Arcd}\left(R((\alpha,\beta_1),\mathbf{t}),R((\alpha,\beta_2),\mathbf{t})\right) &= 0\forall\alpha\end{aligned}$$

Since Arcd is a distance, this implies that $R((\alpha_1,\beta),\mathbf{t})=R((\alpha_2,\beta),\mathbf{t})$ for all β, and $R((\alpha,\beta_1),\mathbf{t})=R((\alpha,\beta_2),\mathbf{t})$ for all α. Since R is a homeomorphism, these two equalities imply that $\alpha_1=\alpha 2$, and $\beta_1=\beta_2$. So, $R((\alpha_1,\beta_1),\mathbf{t})=R((\alpha_2,\beta_2),\mathbf{t})$

(iv) Triangular inequality. Let $R_i=R((\alpha_i,\beta_i),\mathbf{t})$, let's show that

$$d_V(R_1,R_3)\leq d_V(R_1,R_2)+d_V(R_2,R_3)$$

this is equivalent to

$$d_V^2(R_1,R_3)\leq d_V^2(R_1,R_2)+d_V^2(R_2,R_3)+2*d_V(R_1,R_2)d_V(R_2,R_3). \tag{6.17}$$

We have that,

$$\begin{aligned}
\text{L.H.S of 6.17} &= C^2_{\alpha_1,\alpha_3} + C^2_{\beta_1,\beta_3} \text{ (by definition)} \\
&\leq C^2_{\alpha_1,\alpha_2} + C^2_{\alpha_2,\alpha_3} + 2 * C_{\alpha_1,\alpha_2} C_{\alpha_2,\alpha_3} + \\
&\quad + C^2_{\beta_1,\beta_2} + C^2_{\beta_2,\beta_3} + 2 * C_{\beta_1,\beta_2} C_{\beta_2,\beta_3} \text{ (b/c Arcd is a metric)} \\
&\leq d_V^2(R_1, R_2) + d_V^2(R_2, R_3) + 2 * C_{\alpha_1,\alpha_2} C_{\alpha_2,\alpha_3} + 2 * C_{\beta_1,\beta_2} C_{\beta_2,\beta_3}
\end{aligned}$$

Then, to show 6.17, it is sufficient to show that

$$C_{\alpha_1,\alpha_2} C_{\alpha_2,\alpha_3} + C_{\beta_1,\beta_2} C_{\beta_2,\beta_3} \leq d_V(R_1, R_2) d_V(R_2, R_3)$$

which is equivalent to

$$C^2_{\alpha_1,\alpha_2} C^2_{\alpha_2,\alpha_3} + C^2_{\beta_1,\beta_2} C^2_{\beta_2,\beta_3} + 2 * C_{\alpha_1,\alpha_2} C_{\alpha_2,\alpha_3} C_{\beta_1,\beta_2} C_{\beta_2,\beta_3} \leq d_V^2(R_1, R_2) d_V^2(R_2, R_3)$$

which is equivalent to

$$2 * C_{\alpha_1,\alpha_2} C_{\alpha_2,\alpha_3} C_{\beta_1,\beta_2} C_{\beta_2,\beta_3} \leq C^2_{\alpha_1,\alpha_2} C\beta_2, \beta_3{}^2 + C^2_{\alpha_2,\alpha_3} C\beta_1, \beta_2{}^2$$

which is equivalent to

$$0 \leq (C_{\alpha_1,\alpha_2} C\beta_2, \beta_3 - C_{\alpha_2,\alpha_3} C\beta_1, \beta_2)^2$$

Since the last inequality is always true, d_V satisfies the triangular inequality.

From (i), (ii), (iii), and (iv), we have that d_V is a metric in V.□

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

Proof This proof is equivalent to the proof of Proposition 2. Since d_V is a metric and L_{α_0,β_0} satisfies the isometry property, then we can show that L_{α_0,β_0} is continuous, invertible and the inverse is continuous. □.

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

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

Proof This proof is equivalent to the proof of Proposition 3 in the one dimensional case. □.

Proposition 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*

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

Proof The proof of this theorem is equivalent to the proof of Theorem 2 in the one dimensional case.□

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

Proof Let $R_{i,j} = R((\alpha_i, \beta_j), \mathbf{t})$.

(i) $d_{1,O}$ is positive by definition.

(ii) $d_{1,O}(R_{1,1}, R_{2,2}) = ||L_{1,O}(R_{1,1}) - L_{1,O}(R_{2,2})||$. So, $d_{1,O}$ is symmetric.

(iii) Triangular inequality is satisfied. i.e

$$\begin{aligned} d_{1,O}(R_{1,1}, R_{3,3}) &\leq ||L_{1,O}(R_{1,1}) - L_{1,O}(R_{2,2})|| + ||L_{1,O}(R_{2,2}) - L_{1,O}(R_{3,3})|| \\ &\leq d_{1,O}(R_{1,1}, R_{2,2}) + d_{1,O}(R_{2,2}, R_{3,3}) \end{aligned}$$

(iv) If $d_{1,O}(R_{1,1}, R_{2,2}) = 0$, then $\mathrm{Arcd}^2(R_{1,0}, R_{2,0}) = 0$, and $\mathrm{Arcd}^2(R_{0,1}, R_{0,2}) = 0$ which imply $\alpha_1 = \alpha_2$, and $\beta_1 = \beta_2$. So, $R_{1,1} = R_{2,2}$.

From (i), (ii), (iii), and (iv), we have that $d_{1,O}$ is a distance.□

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.

Proof (i), (ii), and (iii) are true by definition. If $L_{2,O}$ is one-to-one, then $d_{2,O}(R_1, R_2) = 0$ implies $||L_{2,O}(R_1) - L_{2,O}(R_2)|| = 0$, which implies $R_1 = R_2$. When conditions (i)-(iv) are verified, $d_{2,O}$ is a metric.□

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.*

Proof When the equality of paths is satisfied, then for all origins O

$$\begin{aligned} d_{1,O}(R_{1,1}, R_{2,2}) &= \sqrt{C^2_{\alpha_1,\alpha_2} + C^2_{\beta_1,\beta_2}} \text{ and} \\ d_{2,O}(R_{1,1}, R_{2,2}) &= \sqrt{C^2_{\alpha_1,\alpha_2} + C^2_{\beta_1,\beta_2}} \end{aligned}$$

So, $d_{1,O} = d_{2,O} = d_V$.□

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 $[0, 1]$.

Proof

(i) d is positive by definition.

(ii) d is symmetric since d_1 and d_2 are symmetric.

(iii) If $d(x, y) = 0$, then $d_1(x, y) = 0$ and $d_2(x, y) = 0$ which implies that $x = y$ since d_1 and d_2 are distances.

(iv) Triangular inequality, we need to prove that

$$d(x, y) \leq d(x, z) + d(z, y)$$

or equivalently that

$$d^2(x, y) \leq d^2(x, z) + d^2(z, y) + 2d(x, z)d(z, y) \tag{6.18}$$

Let's consider the left hand side of the inequality

$$\begin{aligned}
d^2(x, y) &= \gamma d_1^2(x, y) + (1-\gamma) d_2^2(x, y) \\
&\leq \gamma d_1^2(x, z) + \gamma d_1^2(z, y) + \\
&\quad +2\gamma d_1(x, z) d_1(z, y) + (1-\gamma) d_2^2(x, z) + \\
&\quad +(1-\gamma) d_2^2(z, y) + 2(1-\gamma) d_2(x, z) d_2(z, y) \\
d^2(x, y) &\leq d^2(x, z) + d^2(z, y) + \\
&\quad +2\gamma d_1(x, z) d_1(z, y) + 2(1-\gamma) d_2(x, z) d_2(z, y)
\end{aligned}$$

To prove the inequality 6.18, it is sufficient to prove that

$$\gamma d_1(x, z) d_1(z, y) + (1-\gamma) d_2(x, z) d_2(z, y) \leq d(x, z) d(z, y) \tag{6.19}$$

We will reason by equivalence, inequality (6.18) is equivalent to

$$(\gamma d_1(x, z) d_1(z, y) + (1-\gamma) d_2(x, z) d_2(z, y))^2 \leq d^2(x, z) d^2(z, y) \tag{6.20}$$

On one hand,

$$\text{L.H.S of inequality (6.20)} = \gamma^2 d_1^2(x,z)d_1^2(z,y) + (1-\gamma)^2 d_2^2(x,z)d_2^2(z,y) + \\ +2\gamma(1-\gamma)d_1(x,z)d_1(z,y)d_2(x,z)d_2(z,y)$$

On the other hand,

$$\begin{aligned}\text{R.H.S of inequality (6.20)} &= \left(\gamma d_1^2(x,z) + (1-\gamma)d_2^2(x,z)\right) * \\ &\quad * \left(\gamma d_1^2(z,y) + (1-\gamma)d_2^2(z,y)\right) \\ &= \gamma^2 d_1^2(x,z)d_1^2(z,y) + (1-\gamma)^2 d_2^2(x,z)d_2^2(z,y) + \\ &\quad +\gamma(1-\gamma)\left(d_1^2(x,z)d_2^2(z,y) + d_1^2(z,y)d_2^2(x,z)\right)\end{aligned}$$

After cancelling the common terms from the left and the right hand side of inequality (6.20), we have the the inequality

$$0 \leq \gamma(1-\gamma)\left(d_1(x,z)d_2(z,y) + d_1(x,z)d_2(z,y)\right)^2$$

Since this last inequality is always true, then the triangular inequality is always true.

Since the function d satisfies (i), (ii), (iii), and (iv) then it is a metric in M.□

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

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

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

Proof This proposition derives from Lemma 1 and the properties of $d_{1,O}$, and $d_{2,O}$.□

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

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

Proof Let $F_n(x)$ be the estimate of the Fréchet function from the sample, i.e

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} d^2_V(R_i, R)$$

Let x_n be a minimizer of F_n, i.e

$$F_n(x_n) = \min_{x\in V} F_n(x)$$

To use the results by [Bhattacharya and Patrangenaru (2003)], we need to show that (V, d_V) is a complete metric space and all bounded closed set is compact.

1. Completeness: Let x_n, and x_m two points in V, then by definition of d_V

$$d_V(x_n, x_m) = \sqrt{\gamma d_1^2(x_n, x_m) + (1-\gamma) d_2^2(x_n, x_m)}$$

So, if (x_n) is a cauchy sequence in (V, d_V), it is cauchy in (V, d_1) and (V, d_2). We know by construction of d_1 and d_2, (V, d_1) and (V, d_2) are both homeomorphic to $(\mathbb{R}^2, ||.||)$ (where $||.||$ denotes the norm derived from the Euclidean metric). By this homeomorphism, a cauchy sequence (x_n) in (V, d_1) converges to a_1 in V. Similarly, a cauchy sequence (x_n) in (V, d_2) converges to a_2 in V. Let's finally show that $a_1 = a_2$. Since V is a subspace of $\mathbb{R}^{d'}$, we have that

$$||x_n - a_1|| \leq d_1(x_n, a_1) \tag{6.21}$$

(b/c the Euclidean distance is the shortest distance), so (x_n) converges to a_1 in $\mathbb{R}^{d'}$. Similarly, we have that

$$||x_n - a_2|| \leq d_2(x_n, a_2) \tag{6.22}$$

(b/c the Euclidean distance is the shortest distance), so (x_n) converges to a_2 in $\mathbb{R}^{d'}$. By equations (6.21) and (6.22), and using the triangular inequality of the Euclidean distance, we have that

$$||a_1 - a_2|| = 0$$

So, $a_1 = a_2$. Then, $d_V(x_n, a_1)$ converges, so (x_n) converges.

2. Compact sets in (V, d_V). Similarly, using the homeomorphism, a closed and bounded set A in (V, d_V) is closed and bounded in (V, d_1) and (V, d_2). By the homeomorphisms to $(\mathbb{R}^2, ||.||)$, the space A is compact in (V, d_1) and (V, d_2), so it is compact in (V, d_V).

So, using the results from [Bhattacharya and Patrangenaru (2003)], we have that

$$(x_n) \rightarrow R_F(a.s)$$

where R_F is a Fréchet mean. On the other hand, we have that

$$F_n(R_F) \rightarrow Var_F$$

Since

$$|F_n(x_n) - Var_F| \leq |F_n(x_n) - F_n(R_F)| + |F_n(R_F) - Var_F|$$

We have that

$$F_n(x_n) \rightarrow R_F(a.s)$$

□

Lemma 2 *If a set of real functions (F_n) is converges uniformly to a function F. For all n, $\tilde{\theta}_n$ is a minimizer of F_n, and the sequence $(\tilde{\theta_n})$ is convergent, then it converges to a minimizer of F.*

Proof The following lemma is useful for the proof of Lemma 2

Lemma 3. *If a set of real functions (F_n) converges uniformly to a function F, and θ_n is a minimizer of F_n for all n then the real sequence $(F_n(\theta_n))$ is Cauchy.i.e $|F_n(\theta_n) - F_m(\theta_m)|$ converges to 0 as n and m converge to ∞.*

Proof of Lemma 3 Since the sequence of function (F_n) uniformly converges to F, it is uniformly Cauchy. We would like to show that

$$|F_n(\theta_n) - F_m(\theta_m)| \leq ||F_n - F_m||_\infty, \forall n, m \tag{6.23}$$

This follows from

$$\begin{aligned} |F_n(\theta_n) - F_m(\theta_m)| &= |\mathrm{Inf}_{\theta \in A}(F_n(\theta)) - \mathrm{Inf}_{\theta \in A}(F_m(\theta))| \\ &= |||-F_n||_\infty - ||-F_m||_\infty| \\ &\leq ||-F_n - (-F_m)||_\infty = ||F_n - F_m||_\infty \end{aligned}$$

The last inequality is true because for all real functions f and g, $|(||f||_\infty - ||g||_\infty)| \leq ||f-g||_\infty$

□

Proof of Lemma 2. Let the limit of the sequence (θ_n) be θ'. We will prove that θ' is a minimizer of F, i.e $\forall \epsilon > 0, \exists \eta_\epsilon \in A$ such that

$$F(\eta_\epsilon) - \epsilon < F(\theta') \leq F(\eta_\epsilon).$$

Let $\epsilon > 0$, then since θ_n minimizes F_n, we have that $\exists \eta_{\epsilon,n}$ such that

$$F_n(\eta_{\epsilon,n}) - \epsilon \leq F_n(\theta_n) < F_n(\eta_{\epsilon,n}) \tag{6.24}$$

Note that $\eta_{\epsilon,n}$ depends on ϵ and also on n, so before taking this inequality to the limit, we need to find a lower bound that is independent of n. From Lemma 3, since $(F_n(\tilde{\theta}_n))$ is Cauchy, for all n, m greater than $n_{1,\epsilon}$:

$$F_m(\tilde{\theta}_m) - \frac{\epsilon}{3} < F_n(\tilde{\theta}_n) \tag{6.25}$$

On the other hand, since $\tilde{\theta}_m$ is a minimizer of F_m then there exists $\eta_{\epsilon,m}$ such that

$$F_m(\eta_{\epsilon,m}) - \frac{\epsilon}{3} < F_m(\tilde{\theta}_m) \tag{6.26}$$

Finally, since the sequence (F_n) is uniformly Cauchy, for all n, m greater than $n_{2,\epsilon}$

$$F_m(x) \quad > \quad F_n(x) - \frac{\epsilon}{3}; \forall x, \text{ and in particular} \tag{6.27}$$

$$F_m(\eta_{\epsilon,m}) \quad > \quad F_n(\eta_{\epsilon,m}) - \frac{\epsilon}{3} \tag{6.28}$$

So, let n_0 be greater than $n_{1,\epsilon}$ and $n_{2,\epsilon}$ than from the three inequalities 6.25, 6.26 and 6.28 we have for all n greater than n_0 the three inequalities (for simplicity we will denote η_{ϵ,n_0} by η_0)

$$\begin{aligned} F_n(\tilde{\theta}_n) &> F_{n_0}(\tilde{\theta}_{n_0}) - \frac{\epsilon}{3} \\ F_{n_0}(\tilde{\theta}_{n_0}) &> F_{n_0}(\eta_0) - \frac{\epsilon}{3} \\ F_{n_0}(\eta_0) &> F_n(\eta_0) - \frac{\epsilon}{3} \end{aligned}$$

So, finally for all $\epsilon > 0$ and all n greater than n_0, there exists η_0 (depends on ϵ but not on n) such as

$$F_n(\eta_0) - \epsilon < F_n(\tilde{\theta}_n) \leq F_n(\eta_0) \tag{6.29}$$

Since we have uniform convergence of F_n to F, it follows that $\forall a > 0$, $\exists n_a$ s.a $\forall n > n_a$

$$-a < \quad F(\theta) - F_n(\theta) \quad < a, \forall \theta \text{ and in particular} \tag{6.30}$$

$$F_n(\tilde{\theta}_n) - a < \quad F(\tilde{\theta}_n) \quad < F_n(\tilde{\theta}_n) + a \text{ and} \tag{6.31}$$

$$F_n(\eta_0) - a < \quad F(\eta_0) \quad < F_n(\eta_0) + a \tag{6.32}$$

From Inequalities 6.29 and 6.32, we have that for all $\epsilon > 0$ and all $a > 0$ and n greater than $\max(n_{1,\epsilon}, n_{2,\epsilon}, n_a)$

$$\begin{aligned} F(\eta_0) - \epsilon - 2a &< F_n(\eta_0) - \epsilon - a < F_n(\tilde{\theta}_n) - a \leq F(\tilde{\theta}_n), \text{ and} \\ F(\tilde{\theta}_n) &< F_n(\tilde{\theta}_n) + a \leq F_n(\eta_0) + a \leq F(\eta_0) + a, \text{ So} \\ F(\eta_0) - \epsilon - 2a < F(\tilde{\theta}_n) &\leq F(\eta_0) + a \end{aligned}$$

Since $\tilde{\theta}_n$ converges to θ' and F is continuous, then $F(\tilde{\theta}_n)$ converges to $F(\theta')$, and we have for all $\epsilon > 0$ and all $a > 0$, that there exists η_0 (depending only on the choice of ϵ) such that

$$F(\eta_0) - \epsilon - 2a < F(\theta') \leq F(\eta_0) + a$$

So

$$\forall \epsilon > 0, F(\eta_0) - \epsilon < F(\theta') \leq F(\eta_0)$$

which proves that the θ' minimizes F. □

Proposition 11 *The sequence of minimizers (O_n) converges to a minimizer O_B*

Proof We have that $f_n(O)$ converges to $F(O)$ for all O in a compact set K, which implies uniform convergence of f_n to F. Moreover, for a sequence (O_n) in K, there is a convergent subsequence. Then, by lemma 2 the sequence (O_n) has a convergent subsequence to a minimizer O.□

BIBLIOGRAPHY

[Anderson (1971)] Anderson, T. W. (1971) *An introduction to multivariate statistical analysis.* Wiley: Wiley series in probability and mathematical statistics.

[Aubin (2000)] Aubin, T. (2000) *A course in differential geometry.* Graduate studies in mathematics, ISS 1065-7339 ; v.27.

[Bhattacharya and Patrangenaru (2003)] Bhattacharya, R. Patrangenaru, V. (2003), Large sample theory of intrinsic and extrinsic sample means on manifolds *Annals of Statistics*, **31**, 1-29.

[Bookstein (1991)] Bookstein, F. L. (1991) *Morphometric Tools for Landmark Data: Geometry and Biology.* Cambridge: Cambridge University Press.

[Capra and Muller (1997)] Capra, W.B., Muller, H.G. (1997), An accelerated-time model for response curves, *Journal of American Statististical Association*, **92**, 72-83.

[Castro et al. (1986)] Castro, P. E., Lawton, W. H. and Sylvestre, E. A. (1986), Principal modes of variation for processes with continuous sample curves, *Technometrics*, **28**, 329-337.

[Cheng and Gasser (1999)] Cheng, Ming-Yen , Gasser, T. , and Hall, P. (1999), Nonparametric density estimation under unimodality and monotonicity constraints, *Journal of Computational and Graphical Statistics*, **8**, 1-21.

[Chiou, Muller and Wang (2003)] Chiou, J.M., Muller, H.G., Wang, J.L. (2003), Functional quasi-likelihood regression models with smooth random effects, *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **65**, 405-423, Part 2.

[Delicado (2001)] Delicado P. (2001). Another look at principal curves and surfaces. *Journal of Multivariate Analysis*, **77**, 84-116.

[Diggle et al. (1994)] Diggle, P. J., Liang, K.-Y. and Zeger, S.L.(1994) *Analysis of longitudinal Data.* New York: Oxford

[Dong and McAvoy (1996)] Dong D. and McAvoy T.J. (1996), Nonlinear Principal Component Analysis-Based on Principal Curves and Neural Networks, *Computers and Chemical Engineeering*

[Dryden and Mardia (1998)] . Dryden, I. L., and Mardia, K. V. (1998) Statistical Shape Analysis. *W*iley Series in Probability and Statistics. New York John Wiley & Sons, Ltd. (UK).

[Fréchet (1948)] Fréchet M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. (French) Ann. Inst. H. Poincaré 10, (1948), 215–310.

[Gasser, and Kneip 1995] Gasser, T., Kneip, A. (1995). Searching for structure in curve samples, *Journal of the American Statistical Association*, **90**, 1179-1188.

[Gasser et al. (1994)] Gasser, T., Kneip, A., Ziegler, P., Molinari, L., Prader, A. and Largo, R. (1994) Development and outcome of indices of obesity in normal children, *Annals of Human Biology*, **21**, 275-286.

[Gasser et al. (1984)] Gasser, T., Muller, H.-G., Kohler, W., Molinari, L. and Prader, A. (1984), Nonparamteric regression analysis of growth curves, *Annals of Statistics*, **12**, 210-229.

[Gilchrist (1996)] Gilchrist G. W. (1996), A quantitative genetic analysis of thermal sensitivity in the locomotor performance curve of *APHIDUS ERVI*, *Evolution*, **50**, 1560-1572.

[Girard and Iovleff (2002)] Girard S., Iovleff S. (2002), Auto-Associative Modes and Generalized Principal Component Analysis, *Rapport de recherche de l'Institut National de Recherche en Informatique et en Automatique.*

[Gomulkiewicz (1992)] Gomulkiewicz, R., and Kirkpatrick M. (1992), Quantitative genetics and the evolution of reaction norms, *Evolution*, **46**, 390-411.

[Hardle and Marron (1990)] Hardle, W. and Marron, J.S. (1990), Semiparametric comparison of regression curves, *Annals of statistics*, **18**, 63-90.

[Hastie and Stuetzle (1989)] Hastie, T. , Stuetzle, W. (1989), Principal curves, *Journal of the american statistical association*, **84**, 502-516.

[He, Muller, and Wang (2003)] He, G.Z., Muller, H.G., Wang, J.L. (2003), Functional canonical analysis for square integrable stochastic processes, *Journal of Multivariate Analysis*, **85 (1)**, 54-77.

[Izem, Marron, and Kingsolver (2004)] Izem R., Marron J.S , Kingsolver J.G (2004) Analyzing nonlinear variation of thermal performance curves.*Proceedings of the American Statistical Association.*

[Izem and Marron (2003)] Izem R., Marron J.S (2003), Quantifying nonlinear modes of variation of curves. *Proceeding of the International Statistical Institute 54th session, Berlin, Germany.*

[Kingsolver et al. (2004)] Kingsolver J.G, Ragland G.J, Schlichta J.G (2004), Quantitative Genetics of Continuous Reaction Norms: Thermal sensitivity of caterpillar growth rates, In press in *Evolution.*

[Kingsolver et al. (2001)] Kingsolver J.G., Gomulkiewicz R. , Carter P.A. (2001), Variation, selection and evolution of function-valued traits, *Genetica*, **112-113**, 87-104.

[Kirkpatrick, Hill, and Thompson (1994)] Kirkpatrick, M., Hill, W.G. , and Thompson, R.(1994), Estimating the covariance structure of traits during growth and aging, illustrated with lactatin in dairy cattle, *Genetical Research*,**64**, 57-69.

[Kirkpatrick and Lofsvold (1992)] Kirkpatrick, M., and D. Lofsvold. (1992), Measuring Selection and Constraint in the Evolution of Growth, *Evolution*, **46**, 954-971.

[Kirkpatrick et al. (1990)] Kirkpatrick, M., Lofsvold D. , and Bulmer. M. (1990) Analysis of the inheritance, selection, and evolution of growth trajectories. *Genetics*, **124**, 979-993.

[Kirkpatrick and Heckman (1989)] Kirkpatrick, M. and Heckman, N. (1989), A quantitative genetic model for growth, shape, reaction norms and other infinite-dimensional characters, *Journal of Mathematical Biology*, **27**, 429-450.

[Kirkpatrick and Lofsvold (1989)] Kirkpatrick, M., and D. Lofsvold. (1989), The evolution of growth trajectories and other complex quantitative characters, *Genome*, **31**, 778-783.

[Kneip et al. (2000)] Kneip, A. , Li, X. , MacGibbon, K. B. , and Ramsay, J. O. (2000), Curve registration by local regression, *The Canadian Journal of Statistics*, **28**, 19-29

[Kneip (1994)] Kneip, A. (1994), Nonparametric estimation of common regressors for similar curve data, *Annals of Statistics*, **22**, 1386-1427.

[Kneip and Engel (1995)] Kneip, A. and Engel, J. (1995), Model estimation in nonlinear regression under shape invariance, *Annals of Statistics*, **23**, 551-570.

[Kneip and Gasser (1988)] Kneip, A. and Gasser, T. (1988), Convergence and consistency results for self-modeling nonlinear regression, *Annals of Statistics*, **16**, 82-112.

[Kneip and Gasser (1992)] Kneip, A. and Gasser, T. (1992), Statisticals tools to analyze data representing a sample of curves, *Annals of Statistics*, **20**, 1266-1305.

[Kohonen (1995)] Kohonen, T., (1995), *Self Organizing Maps*, Berlin, Heidelberg: Springer, 1995.

[Chang and Ghosh (2001)] Chang, K. and Ghosh J. (2001), A unified model for probabilistic surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23:1**, 22-41.

[Lawton et al. (1972)] Lawton, W.H, Sylvestre, E.A and Maggo, M.S (1972), Self modeling regression, *Technometrics*, **14**, 513-532.

[Le and Kume (2000)] Le, H.L. Kume, A. (2000), Estimating Fréchet means in Bookstein's shape space, *A*dvances In Applied Probability, **32**, 101-113.

[Le (2001)] Le, H.L.(2001), Locating Fréchet means with application to shape spaces, Advances in Applied Probability, **33**, 324 - 338.

[Lynch and Walsh (1998)] Lynch, M. and Walsh B. (1998), *Genetics and Analysis of quantitative traits*, Sinauer Associates Inc.

[Meyer (1991)] Meyer, K. (1991), Estimating variances and covariances for multivariate animal models by Restricted Maximum Likelihood, *Genetics, Selection, and Evolution*, **23**, 67-83.

[Meyer (1998)] Meyer, K. (1998), Estimating covariance function for longitudinal data using a random regression model, Genetics, Selection, and Evolution, **30**, 221-240.

[Meyer (2001)] Meyer, K. (2001) Estimates of direct and maternal covariance functions for growth of Australian beef calves from birth to weaning, *Genetics Selection Evolution*, **33**, 487-514.

[Ramsay and Silverman (1997)] Ramsay J. O. , and Silverman B. W. (1997), *Functional data analysis*, Springer: Springer Series in Statistics.

[Ramsay and Silverman (2002)] Ramsay, J.O. and Silverman, B.W. (2002), *Functional data analysis : methods and case studies*, Springer: Springer Series in Statistics.

[Silverman (1995)] Silverman, B. W. (1995), Incorporating parametric effects into functional principal components, *Journal of the Royal Statistical Society*, **57**, 673-689.

[Small (1996)] Small C. G (1996) *The Statistical Theory of Shape*, Springer: Springer Series in Statistics.

[Tarpey and Flury (1996)] Tarpey T. and Flury B. (1996) Self-Consistency: A fundamental Concept in Statistics, *Statistical Science*, **11:3**, 229-243.

[Wang and Gasser (1997)] Wang, K. , and Gasser, T. (1997), Alignment of curves by dynamic time warping, *The Annals of Statistics*, **25** , 1251-1276.

[Wang and Gasser (1999)] Wang, K. , and Gasser, T. (1999), Synchronizing sample curves nonparametrically, *The Annals of Statistics*, **27**, 439-460.