Asymptotic Consistency of a Fréchet Mean Estimator for the Latent Deformation Model

Cody Carroll

Nov. 2026

1 Setup and Motivation

Carroll and Müller (2023) propose the latent deformation model (LDM)

$$X_{ij}(t) = A_{ij}\,(\lambda \circ \Psi_j \circ H_i)(t), \qquad i = 1, \dots, n, \quad j = 1, \dots, p, \tag{1}$$

where λ is the latent curve with $\|\lambda\|_\infty = 1$, $A_{ij} > 0$ are random amplitude factors, Ψ_j are fixed component-level deformations, and H_i are random subject-level deformations. Without loss of generality we take $\mathcal{T} = [0, 1]$. The deformations satisfy the standardization conditions

$$\frac{1}{p}\sum_{j=1}^{p} \Psi_j^{-1}(t) = t \quad \text{for all } t \in \mathcal{T}, \tag{2}$$

$$E\, H_i^{-1}(t) = t \quad \text{for all } t \in \mathcal{T}. \tag{3}$$

The jth component tempo is $\gamma_j = \lambda \circ \Psi_j$, so that the normalized curves are $X_{ij}/\|X_{ij}\|_\infty = \gamma_j \circ H_i$.

Carroll and Müller (2023) estimate the component tempos as follows. For each component j, the normalized curves $\{X_{ij}/\|X_{ij}\|_\infty\}_{i=1}^n$ are aligned by pairwise warping (Tang and Müller, 2008) with penalty parameter η_1, which yields estimates $\tilde{H}_i^{(j)}$ of the subject-level deformations. These are averaged over components and used to align each component:

$$\hat{H}_i = \frac{1}{p}\sum_{j=1}^{p} \tilde{H}_i^{(j)}, \qquad \hat{\gamma}_j = \frac{1}{n}\sum_{i=1}^{n} \frac{X_{ij} \circ \hat{H}_i^{-1}}{\|X_{ij}\|_\infty}, \qquad j = 1, \dots, p. \tag{4}$$

When the curves are observed with noise on a grid of m time points, kernel-smoothed curves with bandwidth b are substituted for X_{ij}. The latent curve is estimated by a global alignment estimator: one component curve is selected at random for each subject, these n normalized curves are aligned by pairwise warping, and the aligned curves are averaged.

The convergence results of Carroll and Müller (2023) are obtained under the following assumptions on (L) the components of the LDM and (S) the smoothing step.

(L1) The latent curve $\lambda \in C^2(\mathcal{T})$ is a bounded function. For any nondegenerate interval $\mathcal{T}_0 \subset \mathcal{T}$,

$0 < \int_{\mathcal{T}_0} \lambda'(t)^2\, dt < \infty$.

(L2) For $j = 1, \ldots, p$, $\sup_{1 \le i \le n} A_{ij} = \mathcal{O}_P(1)$ and $\sup_{1 \le i \le n} A_{ij}^{-1} = \mathcal{O}_P(1)$.

(S0) The time points $t_1, \ldots, t_m$ depend on the sample size, $m = m(n)$, and constitute a dense regular design generated by a smooth design density f with $\inf_{t \in \mathcal{T}} f(t) > 0$ and bounded second derivative, according to $t_s = F^{-1}\big(\frac{s-1}{m-1}\big)$, $s = 1, \ldots, m$, where F^{-1} is the quantile function associated with f.

(S1) The kernel function K is a probability density function with support $[-1, 1]$, symmetric around zero, and uniformly continuous on its support, with $\int_{-1}^{1} K^2(u)\, du < \infty$.

(S2) The sequences $m = m(n)$ and $b = b(n)$ satisfy $0 < b < \infty$, and $m \to \infty$, $b \to 0$, and $mb^2(\log b)^{-1} \to \infty$ as $n \to \infty$.

Under (L1), (L2), and (S0–S2), Theorem 1 of Carroll and Müller (2023) gives

$$\sup_{t \in \mathcal{T}} |\hat{\gamma}_j(t) - \gamma_j(t)| = \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(\tau_m^{1/2}) + \mathcal{O}(\eta_1^{1/2}), \qquad j = 1, \ldots, p, \tag{5}$$

and the same rate for the global alignment estimator of λ. Here $\tau_m = m^{-(1-\delta)/3}$ for an arbitrarily small $\delta > 0$, m is the number of observations per curve, and η_1 is the penalty parameter of the pairwise warping step.

The global alignment estimator uses one of the p observed curves per subject and discards the rest. A natural alternative is to estimate λ from the p estimated tempos $\hat{\gamma}_1, \ldots, \hat{\gamma}_p$, each of which already pools all n subjects. Since the tempos are warped versions of one curve, λ is their mean in a space where curves are compared up to time warping, that is, a Fréchet (or Karcher) mean. The standardization (2) then selects which warped version of that mean is λ. Consistency of an estimator of this type was not established in Carroll and Müller (2023). In this note we show that a Fréchet mean of the estimated tempos under the uniform metric, centered according to (2), attains the rate in (5).

The organization of this note is as follows. Section 2 defines the estimator. Section 3 states the two additional assumptions and the results, with a plain-language overview of the argument. Section 4 contains the proofs.

2 The Fréchet Mean Estimator

Let $\mathcal{W}$ denote the set of all strictly increasing continuous bijections of $\mathcal{T}$ onto itself. This set contains the deformation space of Carroll and Müller (2023) and is closed under composition, inversion, and convex combinations. For continuous functions f, g on $\mathcal{T}$, define the warping-invariant distance

$$d(f, g) = \inf_{\varphi \in \mathcal{W}} \|f \circ \varphi - g\|_\infty. \tag{6}$$

Since $\|f \circ \varphi - g \circ \varphi\|_\infty = \|f - g\|_\infty$ for every $\varphi \in \mathcal{W}$, the function d is symmetric, satisfies the triangle inequality, and satisfies $d(f, g) \le \|f - g\|_\infty$ and $d(f \circ \varphi, g) = d(f, g)$. It is a pseudometric on $C(\mathcal{T})$ under which each tempo is at distance zero from the latent curve: $d(\gamma_j, \lambda) = 0$, with the infimum attained at $\varphi = \Psi_j^{-1}$.

We assume the estimated tempos $\hat{\gamma}_j$ in (4) are continuous. The sample Fréchet functional is

$$\hat{V}(\mu) = \frac{1}{p} \sum_{j=1}^{p} d^2(\hat{\gamma}_j, \mu), \qquad \mu \in C(\mathcal{T}).$$

A minimizer of $\hat{V}$ is determined at most up to time warping, and we do not assume that the infimum is attained. We therefore allow a tolerance $\rho \ge 0$ in the two optimization steps. The Fréchet mean estimator (FME) of λ is obtained as follows.

Step 1. (Template) Choose $\hat{\mu} \in C(\mathcal{T})$ such that $\hat{V}(\hat{\mu}) \le \inf_{\mu \in C(\mathcal{T})} \hat{V}(\mu) + \rho^2$.

Step 2. (Alignment) For $j = 1, \ldots, p$, choose $\hat{\varphi}_j \in \mathcal{W}$ such that $\|\hat{\gamma}_j \circ \hat{\varphi}_j - \hat{\mu}\|_\infty \le d(\hat{\gamma}_j, \hat{\mu}) + \rho$.

Step 3. (Centering) Set $\bar{\varphi} = p^{-1} \sum_{j=1}^{p} \hat{\varphi}_j$ and

$$\hat{\lambda}_F = \hat{\mu} \circ \bar{\varphi}^{-1}. \tag{7}$$

For every $\rho > 0$, Steps 1 and 2 can be carried out by the definition of the infimum. The centering step is the sample version of (2): the warping functions $\hat{\varphi}_j \circ \bar{\varphi}^{-1}$ that carry the estimated tempos to $\hat{\lambda}_F$ average to the identity, just as the functions Ψ_j^{-1} that carry the tempos to λ do.

3 Results

In addition to (L1), (L2), and (S0–S2), we require the following.

(L3) The derivative λ' has at most finitely many zeros in $\mathcal{T}$, and $\lambda''(z) \neq 0$ at each zero z.

(D) For $j = 1, \ldots, p$, the inverse deformation Ψ_j^{-1} is Lipschitz continuous on $\mathcal{T}$ with constant L.

Assumption (L3) strengthens the requirement in (L1) that λ has no flat stretches: the latent curve has finitely many peaks and valleys, and each has nonzero curvature. Assumption (D) holds whenever the derivatives Ψ_j' are bounded away from zero.

Main idea. The argument has three parts. First, since λ has Fréchet variance at most $\max_j \|\hat{\gamma}_j - \gamma_j\|_\infty^2$, the triangle inequality for d forces any approximate minimizer $\hat{\mu}$ to be close to a warped version of λ. No curvature condition on the Fréchet functional is needed, because the population Fréchet variance is zero. Second, we need to control the warping functions. If $\lambda \circ g$ is uniformly close to λ, then g is close to the identity, but only at a square-root rate near the peaks and valleys

of λ, where the curve is flat to first order. Third, it emerges that this loss does not reach the estimator: the warping error is large only where λ' is small, so its effect on the values of λ is still of first order. This is the content of Lemma 1, which bounds the product $|\lambda'(t)|\,|g(t)-t|$ rather than $|g(t)-t|$ alone.

Throughout, let Z be the set of zeros of λ', K its cardinality, $N \le K+1$ the number of intervals into which $Z \cup \{0,1\}$ partitions $\mathcal{T}$, and $B = \|\lambda''\|_\infty$. On each of the N intervals λ is strictly monotone.

Lemma 1 (Stability of warping)**.** *Suppose $\lambda \in C^2(\mathcal{T})$ satisfies (L3). There exists a constant $c_0 > 0$ depending only on λ such that*

$$\int_s^t |\lambda'(u)|\,du \ge c_0\,(t-s)^2 \qquad \text{for all } 0 \le s \le t \le 1. \tag{8}$$

Moreover, if $g \in \mathcal{W}$ and $\delta \ge 0$ satisfy $\|\lambda \circ g - \lambda\|_\infty \le \delta$, then for all $t \in \mathcal{T}$,

$$(g(t)-t)^2 \le C_\lambda\,\delta \qquad \text{and} \qquad |\lambda'(t)|\,|g(t)-t| \le C_\lambda\,\delta, \tag{9}$$

where $C_\lambda = 4N \max\{1, (1+B)/c_0\}$.

In particular, $\lambda \circ g = \lambda$ with $g \in \mathcal{W}$ implies $g = \mathrm{id}$.

Theorem 1 (Deterministic bound)**.** *Suppose $\lambda \in C^2(\mathcal{T})$ satisfies (L3), the deformations satisfy* (2) *and (D), and $\hat\gamma_1, \dots, \hat\gamma_p \in C(\mathcal{T})$. Let $\varepsilon = \max_{1\le j\le p} \|\hat\gamma_j - \gamma_j\|_\infty$, and let $\hat\lambda_F$ be obtained from Steps 1–3 with tolerance $\rho \ge 0$. Then*

$$\|\hat\lambda_F - \lambda\|_\infty \le C\,(\varepsilon + \rho), \qquad C = (1+\sqrt{p})\,\big\{1 + 2\,C_\lambda L\,(1 + BL/2)\big\}. \tag{10}$$

The bound holds for every choice of $\hat\mu$ and $\hat\varphi_j$ that meets the tolerance, so the estimator does not need to be unique. The case $\varepsilon = \rho = 0$ gives identification at the population level.

Corollary 1 (Population identification)**.** *Under the assumptions of Theorem 1, suppose $\hat\gamma_j = \gamma_j$ for all j. Then $\mu = \lambda$ and $\varphi_j = \Psi_j^{-1}$ satisfy Steps 1 and 2 with $\rho = 0$, and every output of Steps 1–3 with $\rho = 0$ is equal to λ.*

Theorem 2 (Rate of convergence)**.** *Suppose assumptions (L1), (L2), and (S0–S2) of Carroll and Müller (2023) hold together with (L3) and (D), and let $\tau_m = m^{-(1-\delta)/3}$ for an arbitrarily small $\delta > 0$. If the FME is computed from the estimated tempos $\hat\gamma_j$ with tolerance ρ_n, then*

$$\sup_{t\in\mathcal{T}} |\hat\lambda_F(t) - \lambda(t)| = \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(\tau_m^{1/2}) + \mathcal{O}(\eta_1^{1/2}) + \mathcal{O}(\rho_n).$$

Corollary 2. *Under the assumptions of Theorem 2, if additionally $\eta_1 = \mathcal{O}(n^{-1})$, $\rho_n = \mathcal{O}(n^{-1/2})$, and m is bounded below by a multiple of $n^{\Delta(1-\delta)^{-1}}$ for some $\Delta \ge 3$, then*

$$\sup_{t\in\mathcal{T}} |\hat\lambda_F(t) - \lambda(t)| = \mathcal{O}_P(n^{-1/2}).$$

The FME therefore attains the same rate as the global alignment estimator, including the parametric rate on dense enough designs, while using all p components of every subject.

4 Proofs

Proof of Lemma 1. We first show (8). If Z is empty, then $c_1 := \min_{u\in\mathcal{T}} |\lambda'(u)| > 0$ and $\int_s^t |\lambda'| \geq c_1(t-s) \geq c_1(t-s)^2$. Otherwise, write $\operatorname{dist}(u, Z)$ for the distance from u to the nearest zero of λ' and consider the ratio $r(u) = |\lambda'(u)|/\operatorname{dist}(u, Z)$ on $\mathcal{T} \setminus Z$. The ratio is continuous and positive, and since $\lambda'(u) = \lambda''(z)(u-z) + o(|u-z|)$ as $u \to z$, it converges to $|\lambda''(z)| > 0$ at each $z \in Z$. Hence r extends to a positive continuous function on the compact set $\mathcal{T}$, and $c_1 := \inf_u r(u) > 0$, that is, $|\lambda'(u)| \geq c_1 \operatorname{dist}(u, Z)$ on $\mathcal{T}$. Now fix $s < t$ and let $\ell = t - s$. The set of $u \in [s, t]$ with $\operatorname{dist}(u, Z) < \ell/(4K)$ has Lebesgue measure at most $2K \cdot \ell/(4K) = \ell/2$, so on a subset of $[s, t]$ of measure at least $\ell/2$ we have $|\lambda'(u)| \geq c_1\ell/(4K)$. Therefore $\int_s^t |\lambda'| \geq c_1\ell^2/(8K)$, and (8) holds in both cases with $c_0 = c_1/(8\max\{K, 1\})$.

Next, let $V(t) = \int_0^t |\lambda'(u)|\, du$, the total variation of λ on $[0, t]$, and let $h = \lambda \circ g$. We claim that

$$|V(g(t)) - V(t)| \leq 2N\delta \qquad \text{for all } t \in \mathcal{T}. \tag{11}$$

Let $0 = a_0 < a_1 < \cdots < a_N = 1$ be the points of $Z \cup \{0, 1\}$, so that λ is monotone on each $[a_{i-1}, a_i]$, and let $q_i = g^{-1}(a_i)$, so that h is monotone on each $[q_{i-1}, q_i]$. Since g maps $[0, t]$ increasingly onto $[0, g(t)]$, the total variation of h on $[0, t]$ is $V(g(t))$. For a function that is monotone on each interval of a partition, the total variation on $[0, t]$ is the sum of the absolute increments over the partition truncated at t, and for any function the total variation is at least that sum. Applying this to h with the partition $\{q_i\}$, and using $\|h - \lambda\|_\infty \leq \delta$ on each of the at most N nonzero increments,

$$V(g(t)) = \sum_{i=1}^{N} |h(q_i \wedge t) - h(q_{i-1} \wedge t)| \leq \sum_{i=1}^{N} |\lambda(q_i \wedge t) - \lambda(q_{i-1} \wedge t)| + 2N\delta \leq V(t) + 2N\delta.$$

Exchanging the roles of λ and h and using the partition $\{a_i\}$ gives $V(t) \leq V(g(t)) + 2N\delta$, which proves (11).

Fix t and let $\ell = |g(t) - t|$. By (11) and (8),

$$c_0\, \ell^2 \leq \Big| \int_t^{g(t)} |\lambda'(u)|\, du \Big| = |V(g(t)) - V(t)| \leq 2N\delta,$$

so $\ell^2 \leq 2N\delta/c_0 \leq C_\lambda\delta$, which is the first bound in (9). For the second, if $2B\ell \leq |\lambda'(t)|$, then $|\lambda'(u)| \geq |\lambda'(t)| - B\ell \geq |\lambda'(t)|/2$ for all u between t and $g(t)$, and (11) gives $|\lambda'(t)|\, \ell/2 \leq 2N\delta$, that is, $|\lambda'(t)|\, \ell \leq 4N\delta$. If instead $|\lambda'(t)| < 2B\ell$, then $|\lambda'(t)|\, \ell < 2B\ell^2 \leq 4BN\delta/c_0$. In both cases $|\lambda'(t)|\, \ell \leq C_\lambda\delta$. □

Proof of Theorem 1. Let $\delta_0 = (1 + \sqrt{p})(\varepsilon + \rho)$.

Step (a): the template is close to the tempos. Since $d(\gamma_j, \lambda) = 0$ and $d(\hat{\gamma}_j, \gamma_j) \le \|\hat{\gamma}_j - \gamma_j\|_\infty \le \varepsilon$, the triangle inequality gives $d(\hat{\gamma}_j, \lambda) \le \varepsilon$ for all j, hence $\hat{V}(\lambda) \le \varepsilon^2$ and, by Step 1 of the estimator, $\hat{V}(\hat{\mu}) \le \varepsilon^2 + \rho^2$. Each summand of $\hat{V}(\hat{\mu})$ is then at most $p(\varepsilon^2 + \rho^2)$, so $d(\hat{\gamma}_j, \hat{\mu}) \le \sqrt{p}\,(\varepsilon + \rho)$. By Step 2 of the estimator and the invariance of the uniform norm under warping,

$$\|\gamma_j \circ \hat{\varphi}_j - \hat{\mu}\|_\infty \le \|\gamma_j - \hat{\gamma}_j\|_\infty + \|\hat{\gamma}_j \circ \hat{\varphi}_j - \hat{\mu}\|_\infty \le \varepsilon + \sqrt{p}\,(\varepsilon + \rho) + \rho = \delta_0, \qquad j = 1, \dots, p. \tag{12}$$

Step (b): the alignment warps are stable. Let $\hat{\rho} = \Psi_1 \circ \hat{\varphi}_1 \in \mathcal{W}$ and $g_j = \Psi_j \circ \hat{\varphi}_j \circ \hat{\rho}^{-1} \in \mathcal{W}$, so that $g_1 = \mathrm{id}$. Since $\gamma_j \circ \hat{\varphi}_j = \lambda \circ \Psi_j \circ \hat{\varphi}_j$, applying (12) for j and for 1 gives $\|\lambda \circ \Psi_j \circ \hat{\varphi}_j - \lambda \circ \hat{\rho}\|_\infty \le 2\delta_0$, and composing with $\hat{\rho}^{-1}$,

$$\|\lambda \circ g_j - \lambda\|_\infty \le 2\delta_0, \qquad j = 1, \dots, p.$$

Lemma 1 then yields, for all $t \in \mathcal{T}$ and all j,

$$(g_j(t) - t)^2 \le 2C_\lambda \delta_0 \qquad \text{and} \qquad |\lambda'(t)|\,|g_j(t) - t| \le 2C_\lambda \delta_0. \tag{13}$$

Step (c): centering. By construction $\hat{\varphi}_j = \Psi_j^{-1} \circ g_j \circ \hat{\rho}$, so $\bar{\varphi} = \kappa \circ \hat{\rho}$ with $\kappa = p^{-1} \sum_{j=1}^p \Psi_j^{-1} \circ g_j \in \mathcal{W}$. The standardization (2) and (D) give

$$|\kappa(t) - t| = \Big|\frac{1}{p} \sum_{j=1}^{p} \big\{\Psi_j^{-1}(g_j(t)) - \Psi_j^{-1}(t)\big\}\Big| \le L \max_{1 \le j \le p} |g_j(t) - t|. \tag{14}$$

Since $\bar{\varphi}^{-1} = \hat{\rho}^{-1} \circ \kappa^{-1}$, the definition (7) and (12) with $j = 1$ give

$$\begin{aligned} \|\hat{\lambda}_F - \lambda\|_\infty &\le \|\hat{\mu} \circ \hat{\rho}^{-1} \circ \kappa^{-1} - \lambda \circ \kappa^{-1}\|_\infty + \|\lambda \circ \kappa^{-1} - \lambda\|_\infty \\ &= \|\hat{\mu} - \lambda \circ \hat{\rho}\|_\infty + \|\lambda - \lambda \circ \kappa\|_\infty \le \delta_0 + \|\lambda \circ \kappa - \lambda\|_\infty. \end{aligned} \tag{15}$$

Step (d): first-order effect of the centering error. A second-order Taylor expansion of λ at t, together with (14) and (13), gives

$$\begin{aligned} |\lambda(\kappa(t)) - \lambda(t)| &\le |\lambda'(t)|\,|\kappa(t) - t| + \tfrac{1}{2} B\,(\kappa(t) - t)^2 \\ &\le L \max_j |\lambda'(t)|\,|g_j(t) - t| + \tfrac{1}{2} B L^2 \max_j (g_j(t) - t)^2 \\ &\le 2C_\lambda L\,(1 + BL/2)\,\delta_0 \end{aligned}$$

for all $t \in \mathcal{T}$. Combining this with (15) proves (10). □

Proof of Corollary 1. With $\hat{\gamma}_j = \gamma_j$ we have $\hat{V}(\lambda) = 0$, so $\mu = \lambda$ satisfies Step 1 with $\rho = 0$, and $\gamma_j \circ \Psi_j^{-1} = \lambda$ shows that $\varphi_j = \Psi_j^{-1}$ satisfies Step 2 with $\rho = 0$. Theorem 1 with $\varepsilon = \rho = 0$ gives $\|\hat{\lambda}_F - \lambda\|_\infty = 0$ for every output of Steps 1–3. □

Proof of Theorem 2. The constant C in (10) depends only on λ, L, and p, and $\varepsilon = \max_j \|\hat{\gamma}_j - \gamma_j\|_\infty$ has the rate in (5) by Theorem 1b of Carroll and Müller (2023), since p is fixed. The result follows

from Theorem 1. □

Proof of Corollary 2. If $m \geq c\, n^{\Delta(1-\delta)^{-1}}$ with $\Delta \geq 3$, then $\tau_m^{1/2} = m^{-(1-\delta)/6} = \mathcal{O}(n^{-\Delta/6}) = \mathcal{O}(n^{-1/2})$. The remaining terms are $\mathcal{O}(n^{-1/2})$ by assumption. □

References

Carroll, C. & Müller, H.-G. (2023). Latent deformation models for multivariate functional data and time-warping separability. *Biometrics*, 79(4), 3345–3358.

Tang, R. & Müller, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika*, 95(4), 875–889.