svGPFA analysis 30190367 of MC_MAZE_SMALL with LDA latents

Joaquin Rapela

December 15, 2025

Contents

1 Hyper-parameters and data properties **1**

2 Behavioral data **1**

3 Data for LDA **2**

4 Methods **2**

5 Results **3**

5.1 First discriminatory direction separates left from right trials 3

5.2 Second discriminatory direction separates top from bottom trials 3

5.3 Third discriminatory direction separates reaches with positive and negative slope 4

6 LDA for vector of Gaussian Processes **7**

1 Hyper-parameters and data properties

number of latents: 15

number of inducing points: 20

number of trials: 100

number of clusters: 142

2 Behavioral data

Figure 1 plots the behavioral data. In every trial monkeys reach to one of the nine target locations.

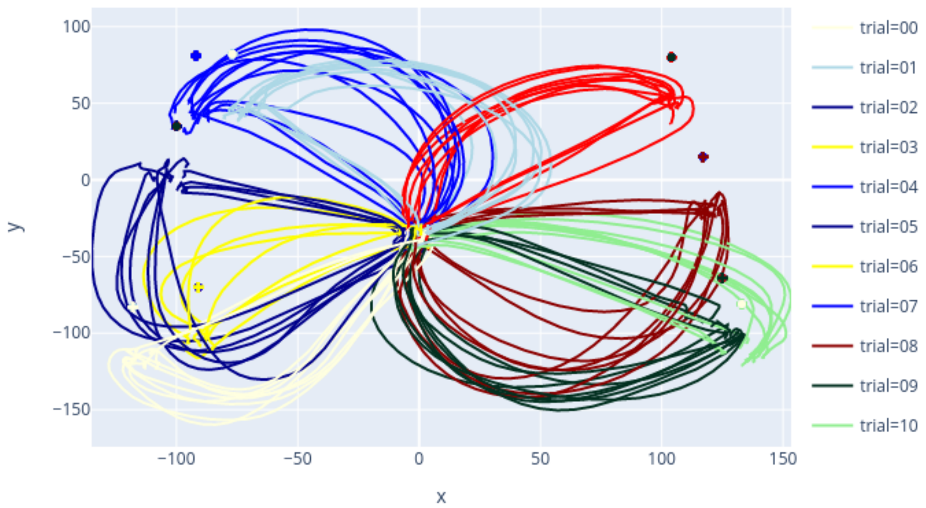

Figure 1: Hand reaching behavior of a monkey. Each trace correspond to a different trial. Trials are colored by target location. There are nine target locations, three to the upper left quadrants, and two to the other quadrants. Click on the figure to get its interactive version.

3 Data for LDA

We provided LDA a matrix of data, $X \in \mathbb{R}^{L\times N}$, and a vector of labels, $\mathbf{y} \in \mathbf{R}^N$. The matrix of data X contains all 15 latents, of all trials, from a 200 ms segment starting at movement onset. A column of X contains the value of the 15 latents corresponding to one sample time point of one trial; thus $L = 15$. The number of columns of X is the number of trials times the number of sample points in the 200 ms segment. The length of $\mathbf{y}$ is the same as the number of columns of X, and the value of $y[i]$ is the target location corresponding to the trial of the ith column of X.

4 Methods

Our implementation of LDA estimated eight (number of target locations minus one) non-orthogonal directions that maximize a criterion of separability, $J_1(A)$, of the projected data Y, where $Y = AX$, $J_1(A) = \text{Trace}(S_b\ S_w^{-1})$, and S_b, S_w and the between and within group scatter matrices of Y. Custom code implementing LDA can be found here. Our LDA implementation produced almost identical results as the scikit-learn one.

The rows of matrix A span the LDA space of dimension number of reach targets minus one. We orthonormalized this space and projected the estimated latents on the orthonormalized space.

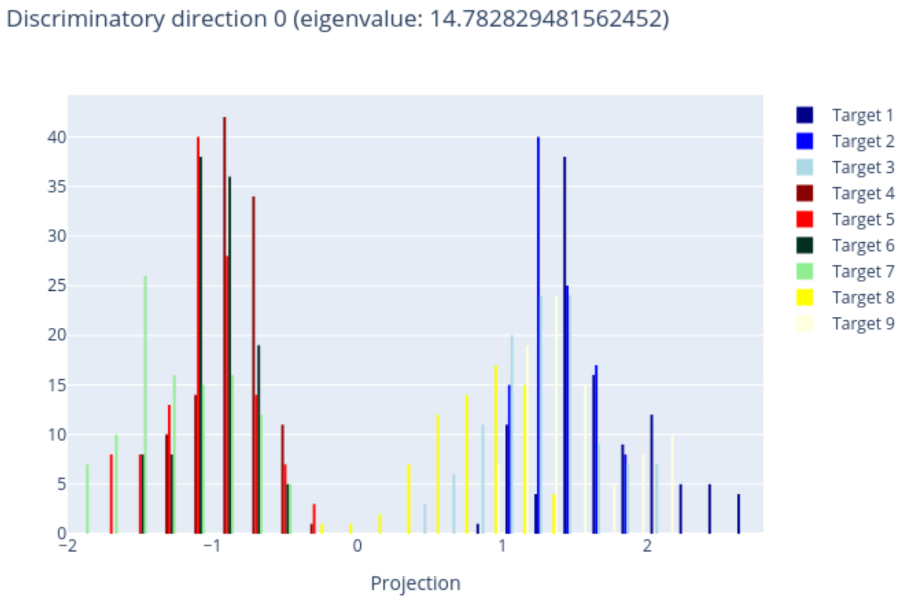

Figure 2: Histograms of projections onto the first LDA direction. Different histograms correspond to the different reach target locations show in Figure 1. This LDA direction separates right (red and green) from left (blue and yellow) trials. Click on the figure to get its interactive version.

5 Results

5.1 First discriminatory direction separates left from right trials

Figure 2 plots histograms of projections of columns of X onto the first discriminatory directions obtained from LDA. There are nine histograms, as many of reaching directions. The histogram for the ith reaching direction contains projections of all columns of X corresponding to trials where the subject reached to the ith direction. The color of the histogram corresponds to the color of the reaching direction in Figure 1. The title of the figure shows the eigenvalue corresponding to this discriminatory direction, which indicates the contribution of this direction to the optimized separability criterion $J_1(A)$ (Section 4). $J_1(A) = \text{Trace}(S_b\ S_w^{-1}) = \text{eigval}_1 + \text{eigval}_2 + \cdots + \text{eigval}_{\text{nTargetLocs}-1}$ (Fukunaga, 1990, Eq. 10.19). This discriminatory direction separates well right (red and green, Figure 1) from left (blue and yellow, Figure 1) trials.

Also, latents corresponding to right and left trials are well separated in the 0.0-0.2 time interval when projected onto the first dimension of the orthonormlised LDA space (see Section 4), as show in Figure 3.

5.2 Second discriminatory direction separates top from bottom trials

Figure 4 plots histograms of projections of columns of X onto the second discriminatory directions obtained from LDA. Note that the eigenvalue corresponding to this discrim-

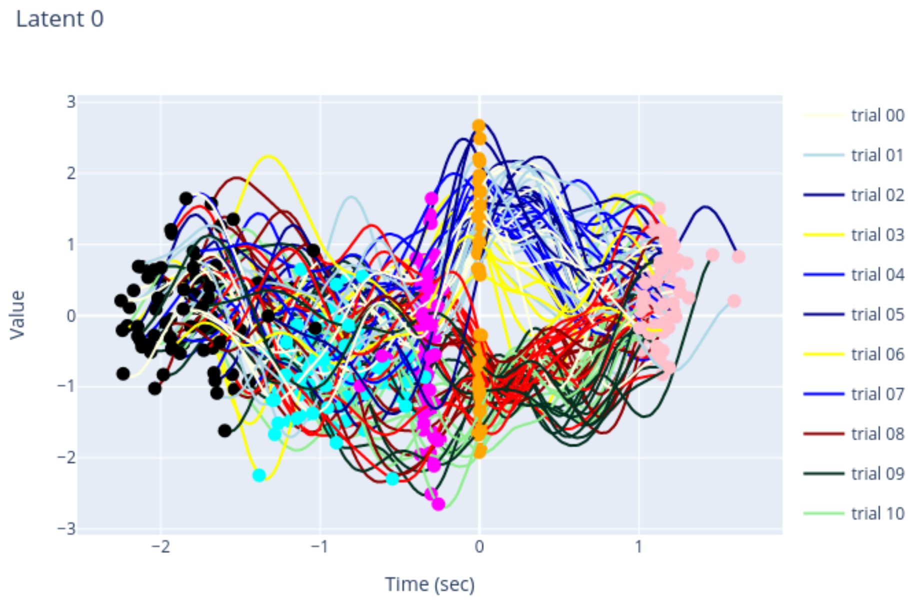

Figure 3: Latents projected onto the first direction of the orthonormalized LDA space. Latents corresponding to trials where the target was on the left (blue and yellow) are well separated in the interval 0.0-0.2 sec from those of trials where the target was on the right (red and green). Click on the figure to get its interactive version.

inatory direction is almost 80% smaller than that for the first discriminatory direction, indicating that this direction contributes almost 80% less to the discriminatory criterion $J_1(A)$. This discriminatory direction separates top (red and blue, Figure 1) from bottom (green and yellow, Figure 1) trials. However, this separation is weaker than that with the first discriminative direction (Figure 2).

Also, latents corresponding to top and bottom trials are well separated in the 0.0-0.2 time interval when projected onto the second dimension of the orthonormlised LDA space (see Section 4), as show in Figure 5. However, this separation is weaker than that in Figure 3.

5.3 Third discriminatory direction separates reaches with positive and negative slope

Figure 6 plots histograms of projections of columns of X onto the third discriminatory directions obtained from LDA. The eigenvalue corresponding to this discriminatory direction is 90% smaller than that for the first discriminatory direction, indicating that this direction contributes 90% less to the discriminatory criterion $J_1(A)$. This discriminatory direction separates reaches with positive slopw (red and blue, Figure 1) from those with negative slope (green and yellow, Figure 1).

Also, latents corresponding to top and bottom trials are well separated in the 0.0-0.2 time interval when projected onto the second dimension of the orthonormlised LDA space (see Section 4), as show in Figure 7. However, this separation is weaker than that in Figure 3.

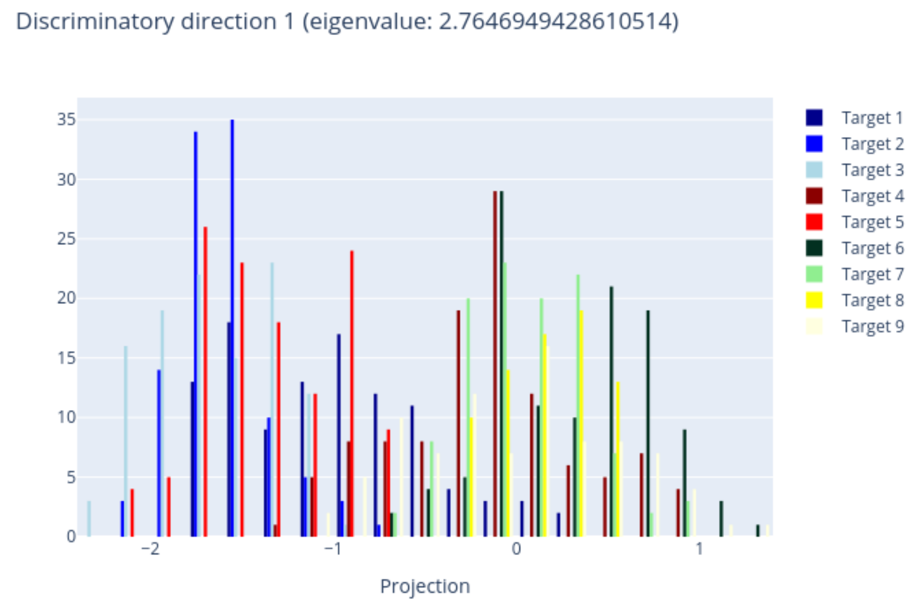

Figure 4: Histograms of projections onto the second LDA direction. Same format as in Figure 2. This LDA direction separates top (red and blue) from bottom (green and yellow) trials. Click on the figure to get its interactive version.

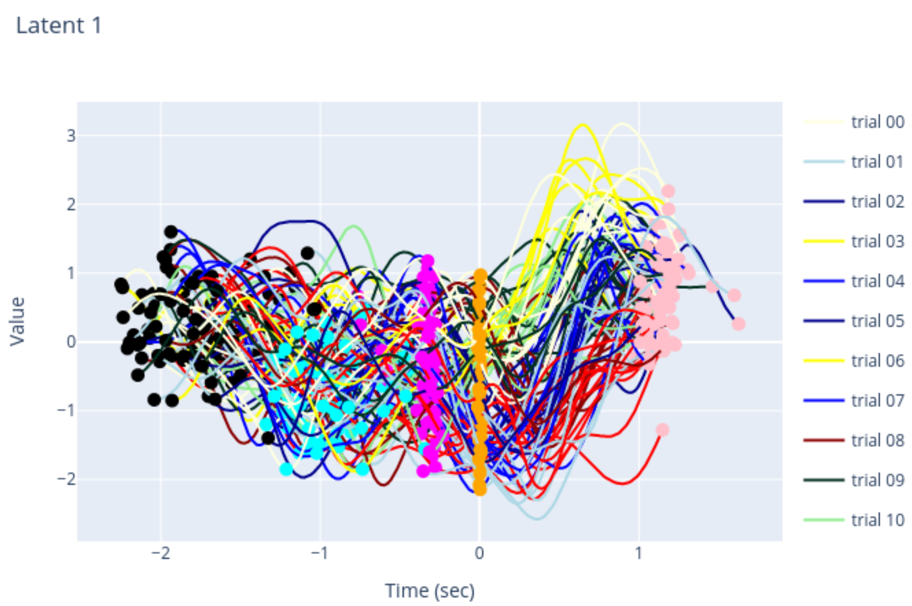

Figure 5: Latents projected onto the second direction of the orthonormalized LDA space. In the interval 0.0-0.2 sec, latents corresponding to top trials (red and blue) are separated from those corresponding to bottom trials (yellow and green), but to a lower degree than in Figure 7. Click on the figure to get its interactive version.

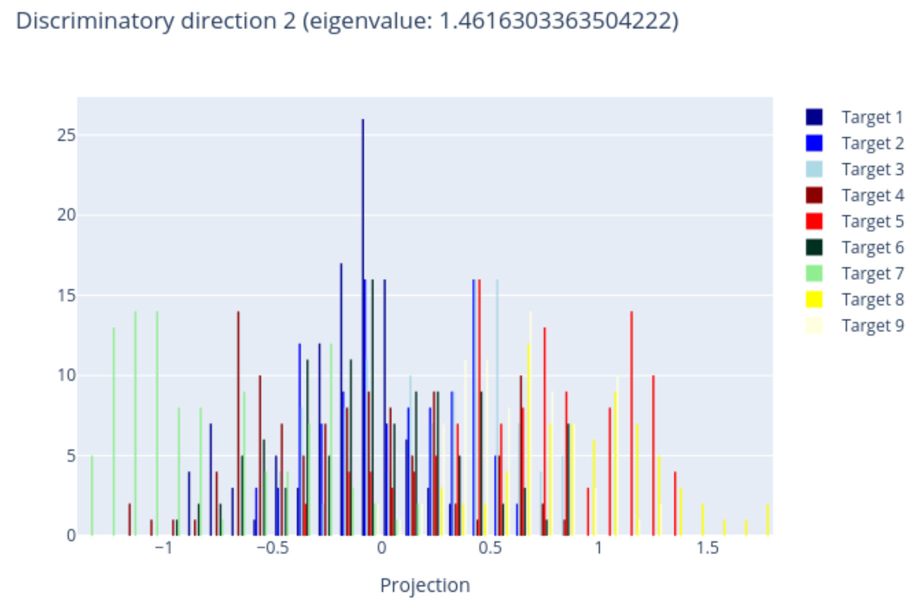

Figure 6: Histograms of projections onto the second LDA direction. Same format as in Figure 2. This LDA direction separates reaches with positive (red and yellow) and negative (blue and green) slopes. Click on the figure to get its interactive version.

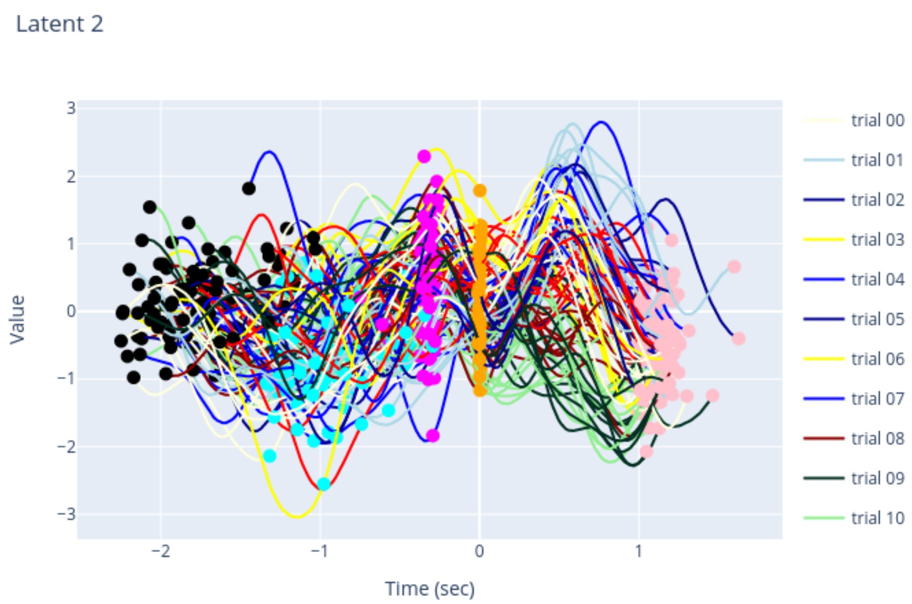

Figure 7: Latents projected onto the third direction of the orthonormalized LDA space. In the 0.0-0.2 sec interval, latents corresponding to reaches with positive slope (red and yellow) are separated from those corresponding to reaches with negative slope (blue and green). Click on the figure to get its interactive version.

6 LDA for vector of Gaussian Processes

Can you tell me how to extend Fisher LDA from vectors of numbers to vectors of Gaussian processes?

Let me clarify. In Fisher LDA I have a set of vectors and for each vector I have a corresponding class. I am seeking a matrix A that transforms X to Y, i.e., $Y = AX$, such that the class separability of Y is maximized.

If X is now a vector of independent mean zero Gaussian processes; i.e., $X_i \sim \mathrm{GP}(0, \kappa_i(\cdot,\cdot))$, and each sample of X is associated with a class, how can I choose linear transformation A ($Y = AX$) such that the samples of Y are maximally separated?

Great question! There are (at least) two clean ways to extend Fisher's LDA from **vectors** to **vectors of Gaussian processes** (i.e., each coordinate is a function). They are equivalent in spirit:

1. **Functional LDA in a Hilbert space** (choose linear functionals of the processes), and

2. **Kernel Fisher Discriminant Analysis (KFDA)** with kernels induced by the GP covariances.

Both end up with the same generalized-eigenvalue problem as classical LDA, just written with **operators** (functional LDA) or **Gram matrices** (KFDA).

Set-up

You observe samples ($\{X^{(n)}, y^{(n)}\}_{n=1}^N$) with class labels ($y^{(n)} \in \{1, \dots, C\}$). Each sample is a **vector of functions**:

$$X^{(n)}(t) = \big(X_1^{(n)}(t), \dots, X_p^{(n)}(t)\big)^\top, \qquad X_i^{(n)} \sim \mathrm{GP}\big(0, \kappa_i(\cdot,\cdot)\big),$$

and (you assume) **independence across coordinates** (i).

A **linear transformation** (A) that maps functions to ($\mathbb{R}^r$) can be written as

$$Y^{(n)} = AX^{(n)}, \qquad Y_j^{(n)} = \sum_{i=1}^p \big\langle w_{j,i}, X_i^{(n)} \big\rangle,$$

where each ($w_{j,i}$) is a weight function and ($\langle\cdot,\cdot\rangle$) is an inner product (choice discussed below). Stack $w_j = (w_{j,1}, \dots, w_{j,p})^\top$; then A is the collection ($\{w_j\}_{j=1}^r$).

Goal: choose (A) so that (Y) has **maximal class separability** (Fisher ratio).

Path A — Functional LDA with Covariance Operators

Let ($\mathcal{H}_i$) be a Hilbert space for coordinate (i). Two common choices:

- $\mathcal{H}_i = L^2(\mathcal{T})$ with the usual $\int f(t)g(t)\,dt$ inner product; or

- $\mathcal{H}_i$ the **RKHS** of (κ_i) (very natural for GPs).

Use the product space $\mathcal{H} = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_p$ with inner product $\langle w, x \rangle_{\mathcal{H}} = \sum_i \langle w_i, x_i \rangle_{\mathcal{H}_i}$. Define **class mean functions** and (centered) **within-class covariance operator**:

$$\mu_c = \mathbb{E}[X \mid y = c] \in \mathcal{H}, \qquad \Sigma_W = \sum_{c=1}^{C} \pi_c \, \mathbb{E}\big[(X - \mu_c) \otimes (X - \mu_c) \mid y = c\big],$$

and the **between-class covariance operator**:

$$\Sigma_B = \sum_{c=1}^{C} \pi_c \, (\mu_c - \mu) \otimes (\mu_c - \mu), \quad \mu = \sum_c \pi_c \mu_c.$$

For any direction ($w \in \mathcal{H}$), the Fisher ratio is

$$J(w) = \frac{\langle w, \, \Sigma_B w \rangle_{\mathcal{H}}}{\langle w, \, \Sigma_W w \rangle_{\mathcal{H}}}.$$

The **functional Fisher problem** is

$\Sigma_B w = \lambda \, (\Sigma_W + \gamma I) \, w$ (generalized eigenproblem with small regularizer ($\gamma > 0$) for invertibility).

Take the top (r) eigenfunctions ($w_1, \ldots, w_r$); then the linear features are ($Y_j = \langle w_j, X \rangle_{\mathcal{H}}$).

Notes

- If your classes differ mainly by their **mean functions** (μ_c), this is exactly the functional analogue of LDA.
- If classes differ only by **covariances** (means equal), $\Sigma_B = 0$ and Fisher LDA cannot separate them—you'd need quadratic/second-order features or QDA-style functionals.

Practical estimation

- Replace expectations by sample estimates (sample means ($\hat{\mu}_c$), empirical covariance operators from training data).
- The operators are compact $\rightarrow$ add ridge (γI) (Tikhonov).
- You can also **truncate** in a basis (e.g., KL/eigenfunctions of (κ_i)) to get a finite-dimensional LDA.

Path B — Kernel Fisher Discriminant Analysis (KFDA) with GP Kernels

This is often the easiest to implement and scales well to irregular sampling.

Feature map viewpoint

Let ($\phi_i : \mathcal{X}_i \to \mathcal{H}_i$) be the RKHS feature map of (κ_i). For the full vector, use the **sum kernel**

$$K\big(X^{(n)}, X^{(m)}\big) = \sum_{i=1}^{p} \kappa_i\big(X_i^{(n)}, X_i^{(m)}\big),$$

which corresponds to the direct-sum feature space $\mathcal{H} = \oplus_i \mathcal{H}_i$. (Independence across coordinates makes the sum kernel a natural choice; you can also weight each (κ_i)).

In KFDA, you look for directions ($w \in \mathcal{H}$) but, by the representer theorem, (w) lies in the **span of training features**:

$$w = \sum_{n=1}^{N} \alpha_n \, \phi(X^{(n)}).$$

Let $K \in \mathbb{R}^{N\times N}$ be the Gram matrix ($K_{nm} = K(X^{(n)}, X^{(m)})$). Let $H_c = I_{n_c} - \frac{1}{n_c}\mathbf{1}\mathbf{1}^\top$ be the centering matrix within class (c), embedded in ($N \times N$). Define (one common formulation; two-class shown for clarity):

- **Between-class term** (M): if $m = \frac{1}{n_1}\mathbf{1}_{C_1} - \frac{1}{n_2}\mathbf{1}_{C_2}$ then $KMK = (Km)(Km)^\top$.
- **Within-class term** (N): $KNK = \sum_c K_c H_c K_c$ (block sums over classes).

Then KFDA solves the generalized eigenproblem

$$(KMK)\,\alpha = \lambda\,\big(KNK + \lambda_{\text{reg}}\,K\big)\,\alpha,$$

and the discriminant score for a new sample (X) is

$$f(X) = \sum_{n=1}^{N} \alpha_n \, K\big(X^{(n)}, X\big).$$

For ($p > 2$) classes, use the multiclass generalization (multiple eigenvectors).

How to build the GP kernels in practice

- If each sample (X_i) is observed on a grid ($t_{i,1}, \ldots$), define the inner product in ($\mathcal{H}_i$) via (κ_i) (RKHS) and compute (K) from data directly (no need to reconstruct on a common grid).
- If sampling times differ per sample, GP kernels handle it naturally; you can, e.g., take $\kappa_i\big(X_i^{(n)}, X_i^{(m)}\big) = \mathbf{z}_n^\top C_{nm} \mathbf{z}_m$ for coefficients from a GP regression representation, or simply use empirical features (KL truncation) and then a linear kernel.

A Concrete Functional Form for (A)

If you prefer an explicit "matrix (A)" analogue, write each output as

$$Y_j = \sum_{i=1}^{p} \int_{\mathcal{T}} w_{j,i}(t)\, X_i(t)\, dt \quad (\text{in } L^2) \quad \text{or} \quad Y_j = \sum_{i=1}^{p} \langle w_{j,i}, X_i \rangle_{\mathcal{H}_i} \ (\text{in RKHS}).$$

Stacking the $(w_{j,i})$ gives your operator $A : \mathcal{H} \to \mathbb{R}^r$. The optimization is exactly the generalized eigenproblem

$$\Sigma_B w_j = \lambda_j (\Sigma_W + \gamma I) w_j, \qquad j = 1, \ldots, r,$$

estimated from data.

Practical Recipe (Works Well)

1. **Choose the space**: RKHS of the GP kernels (most coherent with your (κ_i)).
2. **Build a kernel**: $K = \sum_i \omega_i K_i$ (weights $(\omega_i \geq 0)$ optional).
3. **Run KFDA** with a small ridge (λ_{reg}) (ill-posed otherwise).
4. **Project**: your discriminant features are $f_j(X) = \sum_n \alpha_{j,n} K(X^{(n)}, X)$.
5. **Classifier**: apply linear LDA in that projected space (or just threshold for two classes).

Two Important Caveats

- **Means vs covariances.** Classical Fisher criteria separate classes by **mean differences** relative to within-class variability. If your classes differ **only in covariance structure** (same mean), standard LDA/KFDA may not separate; you'd need **quadratic features** or covariance-aware discriminants.
- **Regularization is essential.** Covariance operators in function space are compact $\to$ non-invertible. Always add ridge (γI in functional LDA; $\lambda_{\text{reg}} K$ in KFDA) or truncate to a finite basis (e.g., top M KL modes per process) before running finite-dimensional LDA.

TL;DR

- View each sample as an element of a Hilbert space (e.g., the RKHS of your GP kernels).
- Define **linear functionals** of the processes and maximize the **functional Fisher ratio**:

$$\max_w \frac{\langle w, \Sigma_B w \rangle}{\langle w, (\Sigma_W + \gamma I) w \rangle}.$$

- Solve the **generalized eigenproblem** (functional LDA), or do it in data space via **KFDA** with kernel $K = \sum_i \kappa_i$.
- The resulting A consists of weight functions $(w_{j,i})$ giving $Y_j = \sum_i \langle w_{j,i}, X_i \rangle$, which maximally separates classes in the Fisher sense.

List of Figures

1 Hand reaching behavior of a monkey. Each trace correspond to a different trial. Trials are colored by target location. There are nine target locations, three to the upper left quadrants, and two to the other quadrants. Click on the figure to get its interactive version. 2
2 Histograms of projections onto the first LDA direction. Different histograms correspond to the different reach target locations show in Figure 1. This LDA direction separates right (red and green) from left (blue and yellow) trials. Click on the figure to get its interactive version. 3
3 Latents projected onto the first direction of the orthonormalized LDA space. Latents corresponding to trials where the target was on the left (blue and yellow) are well separated in the interval 0.0-0.2 sec from those of trials where the target was on the right (red and green). Click on the figure to get its interactive version. 4
4 Histograms of projections onto the second LDA direction. Same format as in Figure 2. This LDA direction separates top (red and blue) from bottom (green and yellow) trials. Click on the figure to get its interactive version. 5
5 Latents projected onto the second direction of the orthonormalized LDA space. In the interval 0.0-0.2 sec, latents corresponding to top trials (red and blue) are separated from those corresponding to bottom trials (yellow and green), but to a lower degree than in Figure 7. Click on the figure to get its interactive version. 5
6 Histograms of projections onto the second LDA direction. Same format as in Figure 2. This LDA direction separates reaches with positive (red and yellow) and negative (blue and green) slopes. Click on the figure to get its interactive version. 6
7 Latents projected onto the third direction of the orthonormalized LDA space. In the 0.0-0.2 sec interval, latents corresponding to reaches with positive slope (red and yellow) are separated from those corresponding to reaches with negative slope (blue and green). Click on the figure to get its interactive version. 6

References

Fukunaga, K. (1990). *Introduction to statistical pattern recognition.* Academic Press.