1 Theory

1. Consider the two eigenvector problems
$$C_{\mathbf{x}}\mathbf{u} = \lambda_x \mathbf{u}$$
and
$$C_{\mathbf{s}}\mathbf{v} = \lambda_s \mathbf{v}$$
where the matrices are related by $C_{\mathbf{x}} = C_{\mathbf{s}} + \alpha I$, where α is a real number and I is the usual identity matrix. Show that if $\mathbf{u}$ is an eigenvector of $C_{\mathbf{x}}$, then it is also an eigenvector of $C_{\mathbf{s}}$ associated with eigenvalue $\lambda_s = \lambda_x - \alpha$.

2. Let A be a real $m \times n$. Show that the matrix M defined as
$$M = \alpha^2 I + AA^T, \alpha \neq 0$$
is nonsingular, where $I = I_m$ and α is a nonzero real number.

3. Show that the between-class scatter matrix, S_B, in the multi-class *Fisher Discriminant Analysis* is given by
$$S_B = \sum_{i=1}^{M} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T,$$
where M is the total number of distinct classes, n_i is the number of data points in class i, $\mathbf{m}_i$ is the class mean of the i^{th} class, and $\mathbf{m}$ is the mean across all n data points. You may use the facts that
$$S_T = S_B + S_W, \quad S_W = \sum_{i=1}^{M} \sum_{x \in D_i} (x - \mathbf{m}_i)(x - \mathbf{m}_i)^T, \quad \text{and} \quad S_T = \sum_{i=1}^{n} (x_i - \mathbf{m})(x_i - \mathbf{m})^T.$$

2 Computing

1. This project concerns the application of the KL procedure for incomplete data [3]. Let the complete data set be translation- invariant:
$$f(x_m, t_\mu) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{k} \sin[k(x_m - t_\mu)],$$
where $m = 1, \ldots, M$, with M dimension of the ambient space (size of the spatial grid), and $\mu = 1, \ldots, P$, with P the number of points in the ensemble. Let $x_m = \dfrac{(m-1)2\pi}{M}$ and $t_\mu = \dfrac{(\mu-1)2\pi}{P}$. Select an ensemble of masks $\left\{\mathbf{m}^{(\mu)}\right\}$, $\mu = 1, \ldots, P$, where 10% of the indices are selected to be zero for each mask. Each pattern in the incomplete ensemble may be written as
$$\tilde{\mathbf{x}}^{(\mu)} = \mathbf{m}^{(\mu)}.\mathbf{f}^{(\mu)},$$
where $\left(\mathbf{f}^{(\mu)}\right)_m = \dfrac{1}{N} \displaystyle\sum_{k=1}^{N} \frac{1}{k} \sin[k(x_m - t_\mu)]$. Let $P = M = 64$ and $N = 3$.

   (a) Compute the eigenvectors of this ensemble using the gappy algorithm [3].

   (b) Plot the eigenvalues as a function of the iteration, and continue until they converge.

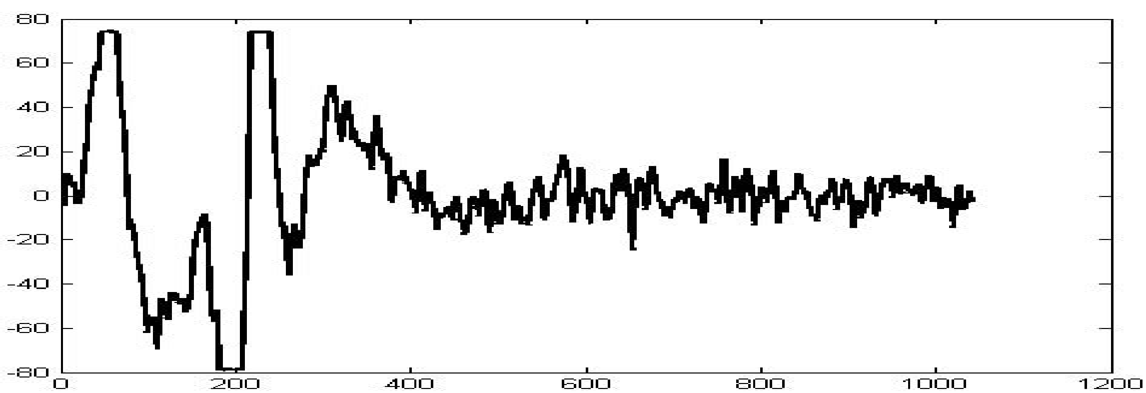
(c) Plot your final eigenfunctions corresponding to the 10 largest eigenvalues.

(d) Plot the element $\tilde{\mathbf{x}}^{(1)}$ and the vector $\tilde{\mathbf{x}}_D$ repaired according to Equation

$$\tilde{\mathbf{x}} \approx \tilde{\mathbf{x}}_D = \sum_{n=1}^{D} \tilde{a}_n \phi^{(n)}. \tag{1}$$

Determine the value of D that provides the best approximation to the original non-gappy pattern vector.

2. This project allows you to apply the two-class *Linear Discriminant Analysis (LDA)* on a simple EEG data. You will download the zipped file EEG_for_LDA from the course website. Once you unzip the archive, you will find 20 files whose file names follow the format "class-C_seq-T", where C stands for the task number (C = 2 and C = 3) and T stands for the trial number which ranges from 0 to 9. The participants were asked to count in task 2 and to perform visual rotation in task 3. The EEG data were collected in 19 channels with sampling at 256 Hz over 10 trials for each task. Upon loading the files, the variable "class_C_seq_T" is a 19-by-1040 matrix, where each row represents a reading from one of the 19 channels (electrodes on the skull) and each column represents a reading at a single time stamp. A sample reading for task 2, trial 0 is shown below.



(a) Write a MATLAB routine to produce an optimal projection direction, w, using the two-class LDA criterion

$$w = \arg\max_{w} J(w) = \arg\max_{w} \frac{w^T S_B w}{w^T S_W w},$$

where

$$S_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T \quad \text{and} \quad S_W = \sum_{i=1}^{2} \sum_{x \in D_i} (x - \mathbf{m}_i)(x - \mathbf{m}_i)^T$$

are the between-class scatter matrix and the within-class scatter matrix, respectively. That is, your code should take in a set of data points with a clear indication which points belong to class one and which points belong to class 2, and output a single vector w that is the solution of the generalized eigenvalue problem $S_B w = \lambda S_W w$. (If you are interested in the implementation of multi-class LDA, see [1] for more details on how to deal with the singularity of S_W.)

(b) Now, use your subroutine in part (a) to project the EEG data onto a real line. Particularly, we can form a data point in $\mathbb{R}^{1040\times 19}$ by concatenating the columns for each trial, therefore having 10 data points for task 2 and 10 data points for task 3. You would then project these 20 points onto the real line with the w found with part (a). Plot the projected data on the real line and distinguish the classes with different symbols. Do you see a clear separation? Analyze your results.

3. Construct a $n \times 10$ matrix (choose $n \geq 250$) to serve as a ground truth data set so that each column is a n-dimensional time series. Next, add **correlated noise** to each column to create a noisy data set, X. The goal of this problem is to implement the Maximum Noise Fraction method to recover the ground truth as closely as possible from the noisy data. Suppose the source of the noise is unknown, you may estimate the noise covariance, $N^T N$, using the difference matrix as $N^T N = \frac{1}{2} dX^T dX$, where if

$$X = \begin{pmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_p(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_p(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_n) & x_2(t_n) & \cdots & x_p(t_n) \end{pmatrix}$$

then

$$dX = \begin{pmatrix} x_1(t_2) - x_1(t_1) & x_2(t_2) - x_2(t_1) & \cdots & x_p(t_2) - x_p(t_1) \\ x_1(t_3) - x_1(t_2) & x_2(t_3) - x_2(t_2) & \cdots & x_p(t_3) - x_p(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_n) - x_1(t_{n-1}) & x_2(t_n) - x_2(t_{n-1}) & \cdots & x_p(t_n) - x_p(t_{n-1}) \end{pmatrix}.$$

Notice that $X \in \mathbb{R}^{n\times p}$ and $dX \in \mathbb{R}^{(n-1)\times p}$. Interested readers may see [2] for a description of the conditions where this method works. In your report, examine and elaborate on the effect of a D-mode reconstruction on a single noisy signal for various values of D (i.e., choose a single column to filter). In a single graph, visually display the result of the original signal, noisy signal, and filtered (de-noised) data (with your best choice of D) to compare. Use the graph legend to distinguish each.

References

[1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[2] F. Emdad. *Signal fraction analysis for subspace processing of high dimensional data*. PhD thesis, Colorado State University, Fort Collins, CO, USA, December 2007.

[3] R. Everson and L. Sirovich. The karhunen-loéve transform for incomplete data. *J. Opt. Soc. Am., A*, 12(8):1657–1664, 1995.