1 Wiener Filter

Let x and y be jointly WSS stochastic processes, and define estimation error process $e = Hy - x$, where H is an LTI filter. By differentiating with respect to each filter coefficient h_n, show that the mean squared error (MSE)

$$\mathcal{E} = \mathbb{E}\left[|x_n - \hat{x}_n|^2\right] \tag{1}$$

is minimized when $h * R_Y Y = R_{XY}$ is satisfied. This verifies a solution covered in course without the orthogonality principal.

Your Solution Here!

2 Dimensionality Reduction

In many data-analysis applications, we may be interested in doing a process of dimensionality reduction. One way of thinking about the problem (which leads to principal component analysis) is to find an orthogonal projection P onto a lower dimensional space such that we minimize

$$\underset{P\in\mathbb{C}^{N\times N}}{\operatorname{argmin}} \sum_i \|x_i - Px_i\|_2^2, \tag{2}$$

where $\{x_i \in \mathbb{C}^N\}$ is a collection of i.i.d. random vectors. Describing our orthogonal projection as the composition of an orthonormal matrix and its adjoint, we may alternatively want to minimize

$$\underset{U\in\mathbb{C}^{N\times M}}{\operatorname{argmin}} \sum_i \|x_i - UU^\top x_i\|_2^2, \tag{3}$$

, where U is a set of orthonormal column vectors.

(a) Show that the above is equivalent to solving

$$\underset{U\in\mathbb{C}^{N\times M}:U^\top U=I}{\operatorname{argmax}} \operatorname{Tr}\left(U^\top \sum_i x_i x_i^\top U\right), \tag{4}$$

where Tr is the trace of a matrix.

Hint: A trace is a linear operator that is invariant to cyclic permutations, i.e. Tr$(ABC) =$ Tr$(BCA) =$ Tr(CBA)

Your Solution Here!

(b) What is the optimal choice of U?

Hint: The trace is also the sum of the eigenvalues, can you see why? Consider combining the spectral decomposition with a cyclic permutation.

Your Solution Here!

(c) If we have strong *a priori* information about our data, we can often do better in an estimation task. In some cases, we may already know the preferred coordinate system. Suppose each x_i is a Gaussian circularly WSS random vector. Here we are defining circularly WSS in the same manner as circular convolution, the periodically extended vector is WSS.

Define U_k as a random matrix generated by solving the above optimization problem (PCA) for k random vectors $\{x_1, ..., x_k\}$. Roughly, what is the limiting behavior of U_k as $k \to \infty$? Justify your claim.

Your Solution Here!

3 Linear MMSE Estimation

Let $X = \left[X_1, ..., X_N\right]^\top$ and $Y = \left[Y_1, ..., Y_N\right]^\top$ be jointly distributed random vectors with mean vectors

$$\mu_x := \mathbb{E}\left[X\right] \tag{5}$$
$$\mu_y := \mathbb{E}\left[Y\right] \tag{6}$$

and the covariance matrices

$$\Sigma_x := \mathbb{E}\left[(X-\mu_x)(X-\mu_x)^*\right], \tag{7}$$
$$\Sigma_y := \mathbb{E}\left[(Y-\mu_y)(Y-\mu_y)^*\right], \tag{8}$$
$$\Sigma_{xy} := \mathbb{E}\left[(X-\mu_x)(Y-\mu_y)^*\right]. \tag{9}$$
$$\tag{10}$$

Let V be the space of all random vectors of length N and define the subset

$$V_Y := \left\{AY + B \middle| A \in \mathbb{C}^{N\times M}, \quad B \in \mathbb{C}^N\right\} \subset V \tag{11}$$

(a) Show that V_Y is a subspace of V, and find its dimension (note that Y is fixed), assuming Σ_Y is invertible.
Hint: note that $V_y = rng(g)$ for $g : \mathbb{C}^{N\times M} \times \mathbb{C}^N \to \mathbb{C}^N$ defined by

$$g(A, B) = AY + B. \tag{12}$$

Show that g is linear in (A, B), and has a trivial nullspace, and the dimension should follow.

Your Solution Here!

(b) The expression $\langle X_1, X_2\rangle = \mathbb{E}\left[X_2^* X_1\right]$ is an inner product on V. Use the orthogonality principal to show that the vector $\hat{X} \in V_Y$ defined as

$$\hat{X} = \hat{A}Y + \hat{B} \qquad \text{where} \qquad \hat{A} = \Sigma_{xy}\Sigma_y^{-1}, \qquad \hat{B} = \mu_x - \Sigma_{xy}\Sigma_y^{-1}\mu_y \tag{13}$$

is the Linear Minimum Mean Square Error (LMMSE) estimator of X as a function of Y, that is, the optimal estimator of X in V_Y. *Hint: Show that $\langle X - \hat{X}, AY + B\rangle = 0$ for all A, B*

Your Solution Here!

(c) Let $Y_h = \left[Y_1, ..., Y_M, 1\right]^\top$. Find the LMMSE of X as a function of Y_h in terms of Σ_x, Σ_y, and Σ_{xy}. Show that your result coincides with the LMMSE estimator using correlation matrices, $\hat{X} = A_{LMMSE}Y_h$, where $A_{LMMSE} = \mathbb{E}\left[XY_h^*\right]\mathbb{E}\left[Y_hY_h^*\right]^{-1}$.

Hint: There are two options to solve this problem:
Option 1: Use the Schur complement inversion formula

$$\begin{bmatrix} R & \mu \\ \mu^* & 1 \end{bmatrix}^{-1} = \begin{bmatrix} (R-\mu\mu^*)^{-1} & -R^{-1}\mu q \\ -q\mu^*R^{-1} & q \end{bmatrix} \qquad \textit{where } q = (1-\mu^*R^{-1}\mu)^{-1} \tag{14}$$

and the Woodbury Inversion formula

$$R^{-1} = (\Sigma + \mu\mu^*)^{-1} = \Sigma^{-1} - \Sigma^{-1}\mu(1 + \mu^*\Sigma^{-1}\mu)^{-1}\mu^*\Sigma^{-1}. \tag{15}$$

Option 2: Define the subspace of linear estimators

$$W_Y = \{QY_h | Q \in \mathbb{C}^{N\times(M+1)}\} \subset V \tag{16}$$

and show that $V_Y = W_Y$*. Then show that* A_{LLMSE} *satisfies the orthogonality principle, hence must coincide with* $\hat{A}, \hat{B}$*.*

Your Solution Here!

4 Inverse Covariance Structure

A real-valued finite-dimensional autonomous discrete-time state-space model of a system is described by the following pair of equations

$$x_{n+1} = Fx_n + u_n \tag{17}$$
$$y_n = Hx_n + v_n, \tag{18}$$

where $x_n, u_n \in \mathbb{R}^N$, $y_n, v_n \in \mathbb{R}^M$, and the linear operators F and H can be represented by appropriately sized matrices. y_n represents an observable at time n, and x_n represents the state of the system. We call F the state transition operator, and H the measurement operator.

Consider letting $u_n \sim \mathcal{N}(0, \Sigma_u)$ and $v_n \sim (0, \Sigma_v)$ be i.i.d. (between timesteps) Gaussian random noise driving the system.

Let the system begin from rest, i.e. $x_0 = u_0$.

First, we will ignore the observations y and assume that $y_n = x_n$ for every timestep.

(a) What is the covariance matrix for a vector $\bar{x} = \begin{bmatrix} x_0^\top & x_1^\top & x_2^\top \end{bmatrix}^\top$?

Your Solution Here!

(b) What are the following conditional distributions?

$$f(x_2|x_1, x_0), f(x_1|x_0)$$

Your Solution Here!

(c) What are the following marginal distributions?

$$f(x_2), f(x_1), f(x_0)$$

Your Solution Here!

(d) Using earlier parts of the problem, what is the inverse covariance matrix?

Your Solution Here!

(e) What can you say about the inverse covariance matrix of

$$\bar{z} = \begin{bmatrix} x_0^\top & y_0^\top & x_1^\top & y_1^\top & ... \end{bmatrix}^\top$$

Your Solution Here!

(f) Now, without giving any actual derivations, just some loose justifications, consider a process that branches out in a tree structure. The system is initialized with a single scalar random variable $x_0 \sim \mathcal{N}(0, \sigma^2)$. To construct the next timestep, we generate two independent random variables $u_1 \sim \mathcal{N}(0, \sigma_u^2)$ and $v_1 \sim \mathcal{N}(0, \sigma_v^2)$. The state for that timestep is a pair of random variables $x_{1,u} = x_0 + u_1$ and $x_{1,v} = x_0 + v_1$.

At each timestep, we follow an analogous process, doubling the number of state variables each time. What can you say about the inverse covariance matrix of the collection of random variables? (similar to part e) How does this relate to a potentially more compuationally efficient representation of the probability density?

Your Solution Here!

The notion behind this property is very important to a wider collection of distributions called Markov Random Fields (MRF), on which conditional and marginal distributions have a nice structure. Such random variables are generally represented as a graphical model and allows generalizations of many common algorithms. MRFs encode the notion that the joint probability of entire system of random variables inherently comes form local interactions.

5 Computational Problem

Implement a Wiener filter to denoise and sharpen the included image. The image has been blurred through convolution with a 11×11 square with i.i.d. Gaussian noise added. The image files are named with a pair of numbers, which indicates that the standard deviation of the noise is $0.XX$, where XX is the number in the filename. There are also a couple of without the blurring.

The blurring was done through a circular convolution.

Below is the original image and an example of a blurred image.


Your Solution Here!