EE363 homework 4

1. *Estimating an unknown constant from repeated measurements.* We wish to estimate $x \sim \mathcal{N}(0,1)$ from measurements $y_i = x + v_i$, $i = 1, \ldots, N$, where v_i are IID $\mathcal{N}(0, \sigma^2)$, uncorrelated with x. Find an explicit expression for the MMSE estimator $\hat{x}$, and the MMSE error.

2. *Estimator error variance and correlation coefficient.* Suppose $(x, y) \in \mathbf{R}^2$ is Gaussian, and let $\hat{x}$ denote the MMSE estimate of x given y, and $\bar{x}$ denote the expected value of x. We define the relative mean square estimator error as $\eta = \mathbf{E}(\hat{x} - x)^2 / \mathbf{E}(\bar{x} - x)^2$. Show that η can be expressed as a function of ρ, the correlation coefficient of x and y. Does your answer make sense?

3. *MMSE predictor and interpolator.* A scalar time series $y(0), y(1), \ldots$ is modeled as

$$y(t) = a_0 w(t) + a_1 w(t-1) + \cdots + a_N w(t-N),$$

   where $w(-N), w(-N+1), \ldots$ are IID $\mathcal{N}(0,1)$. The coefficients $a_0, \ldots, a_N$ are known.

   (a) *Predicting next value from current value.* Find the MMSE predictor of $y(t+1)$ based on $y(t)$. (Note: we really mean based on just $y(t)$, and not based on $y(t), y(t-1), \ldots$) Your answer should be as explicit as possible.

   (b) *MMSE interpolator.* Find the MMSE predictor of $y(t)$ (for $t > 1$) based (only) on $y(t-1)$ and $y(t+1)$ (for $t \geq 1$). Your answer should be as explicit as possible.

4. *Estimating initial subpopulations from total growth observations.* A sample that contains three types of bacteria (called A, B, and C) is cultured, and the total bacteria population is measured every hour. The bacteria populations grow, independently of each other, exponentially with different growth rates: A grows 2% per hour, B grows 5% per hour, and C grows 10% per hour. The goal is to estimate the initial bacteria populations based on the measurements of total population.

   Let $x_A(t)$ denote the population of bacteria A after t hours (say, measured in grams), for $t = 0, 1, \ldots$, and similarly for $x_B(t)$ and $x_C(t)$, so that

$$x_A(t+1) = 1.02 x_A(t), \quad x_B(t+1) = 1.05 x_B(t), \quad x_C(t+1) = 1.10 x_C(t).$$

   The total population measurements are $y(t) = x_A(t) + x_B(t) + x_C(t) + v(t)$, where $v(t)$ are IID, $\mathcal{N}(0, 0.25)$. (Thus the total population is measured with a standard deviation of 0.5).

   The prior information is that $x_A(0)$, $x_B(0)$, $x_C(0)$ (which are what we want to estimate) are IID $\mathcal{N}(5, 2)$. (Obviously the Gaussian model is not completely accurate since it

allows the initial populations to be negative with some small probability, but we'll ignore that.)

How long will it be (in hours) before we can estimate $x_A(0)$ with a mean square error less than 0.01? How long for $x_B(0)$? How long for $x_C(0)$?

5. *Sensor selection.* Suppose 5 scalar measurements $y_1, \ldots, y_5$ are related to 3 scalar variables $x_1,\ x_2,\ x_3$ by

$$y = Ax + v, \quad A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 1 & 3 \\ 1 & -2 & 0 \\ -1 & 0 & -3 \\ 1 & 2 & -3 \end{bmatrix},$$

where v_i are IID $\mathcal{N}(0, 0.01)$. Thus the measurement errors have standard deviation 0.1. The variable $x \in \mathbf{R}^3$ is deterministic but unknown.

You are to find an estimator $\hat{x} = By$ that satisfies the following specifications:

- $BA = I$. This means that $\hat{x} = x$ when $v = 0$, and also that the estimator is unbiased, *i.e.*, $\mathbf{E}\,\hat{x} = x$.
- $\mathbf{E}\,\|\hat{x} - x\|^2 \leq 0.005$.

Among all matrices B that satisfy these specifications, you are to find one that minimizes the number of nonzero columns.

To understand the practical meaning of this, note that if the jth column of B is zero, then the estimate $\hat{x}$ does not use the jth measurement, so the jth sensor is not needed. Thus we are seeking the smallest sensor configuration (in terms of number of sensors required) that can be used to estimate x with mean square error not exceeding 0.005.

Make sure to describe exactly how you are going to solve the problem, as well as giving your explicit solution (*i.e.*, B). If no B satisfies the specifications, say so, and show why.

6. *MMSE estimation example.* We wish to estimate $x \in \mathbf{R}^n$, given a set of measurements $y \in \mathbf{R}^m$, where

$$y = Ax + v, \qquad x \sim \mathcal{N}(\bar{x}, \Sigma_x), \qquad v \sim \mathcal{N}(0, \Sigma_v),$$

with x and v independent. We'll consider the specific case with $n = 4$, $m = 6$, and measurement matrix

$$A = \begin{bmatrix} 2 & 3 & -1 & 4 \\ 1 & 0 & 0 & -2 \\ 2 & 1 & 1 & 0 \\ -3 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 \end{bmatrix}.$$

The prior distribution of x is given by $\bar{x} = (2, -1, 0.5, -3)$,

$$\sigma_1^2 = 2, \quad \sigma_2^2 = 5, \quad \sigma_3^2 = 1, \quad \sigma_4^2 = 2,$$

and

$$\rho_{12} = -0.1, \quad \rho_{13} = 0.025, \quad \rho_{23} = -0.01.$$

Here σ_i is the standard deviation of x_i, and ρ_{ij} is the correlation coefficient between x_i and x_j; correlation coefficients not given are zero. The noise statistics are characterized by standard deviations

$$\tilde{\sigma}_1^2 = 2, \quad \tilde{\sigma}_2^2 = 1, \quad \tilde{\sigma}_3^2 = 2, \qquad \tilde{\sigma}_4^2 = 3, \qquad \tilde{\sigma}_5^2 = 2, \qquad \tilde{\sigma}_6^2 = 1,$$

and correlation coefficients

$$\tilde{\rho}_{13} = -0.25, \quad \tilde{\rho}_{24} = 0.5, \quad \tilde{\rho}_{35} = 0.3, \quad \tilde{\rho}_{46} = -0.04,$$

with other correlation coefficients zero.

In this problem you will compare the performance of two estimators. The first is the simple pseudo-inverse estimator, $\hat{x}_{\text{pinv}} = A^\dagger y$, (as studied in EE263). The second is the MMSE estimator, denoted $\hat{x}_{\text{mmse}}$. You will first make some predictions about the performance of the two estimators, based on some linear algebra calculations; then, you will carry out simulations of the two to verify your calculations.

(a) Find $\mathbf{E}\,\|\hat{x}_{\text{pinv}} - x\|^2$ and $\mathbf{E}\,\|\hat{x}_{\text{mmse}} - x\|^2$, explaining your method. Briefly comment on the results.

(b) Let $\mathcal{E} = \{\hat{x} - x \mid (\hat{x} - x)^T \Sigma^{-1} (\hat{x} - x) \leq \alpha\}$ be the 90% confidence ellipsoid for the MMSE estimation error $\hat{x}_{\text{mmse}} - x$. Find Σ and α.

(c) Generate 1000 samples of (x,v) pairs from the distributions described above. For each sample, find $\hat{x}_{\text{mmse}} - x$ and $\hat{x}_{\text{pinv}} - x$. Plot the empirical distributions (histograms) of $\|\hat{x}_{\text{mmse}} - x\|^2$ and $\|\hat{x}_{\text{pinv}} - x\|^2$. (Plot the histograms on the same horizontal axis, so they are easier to compare.) Verify that the empirical averages of these are close to values predicted in part (a).

(d) For how many of your 1000 samples from part (c) does $\hat{x}_{\text{mmse}} - x$ fall inside the 90% confidence ellipsoid $\mathcal{E}$ found in part (b)? (Obviously this number should be near 900. If the number turns out to exactly 900, change your seed and run the simulations again.)

Some Matlab hints.

- `sqrtm(A)` finds the (matrix) square root of matrix `A`.
- `randn(n,1)` generates a column vector of dimension `n` drawn from a normal distribution $\mathcal{N}(0, I)$. (Do not confuse `randn` with `rand`, which draws the entries of the vector from a uniform distribution on $[0, 1]$.)

- `hist(y,M)` bins the elements of `y` into `M` equally spaced bins and plots the results. (For 1000 samples, `M` around 50 gives a reasonable plot.)
- `chi2inv(p,n)` returns the inverse of the chi-square cumulative distribution with `n` degrees of freedom, at the value `p`.

7. *Cholesky decomposition and estimation.* Every symmetric positive definite matrix P can be decomposed as $P = LL^T$, with L a lower triangular matrix with diagonal entries $L_{ii} > 0$. This decomposition is unique and is called the *Cholesky decomposition* of P. The Cholesky decomposition comes up in many contexts, such as solving linear equations with symmetric positive definite coefficient matrix, and least-squares and least-norm problems. In this problem we explore the connection between the Cholesky factorization and estimation. Let $x \sim \mathcal{N}(0, \Sigma)$ and let $\Sigma = LL^T$ be the Cholesky decomposition of Σ.

   Notation: We use $k:l$ to denote a range of integers, $k, k+1, \ldots, l$. If $z \in \mathbf{R}^n$ is a vector, we let $z_{k:l}$ denote the subvector with the corresponding entries, *i.e.*, $(z_k, \ldots, z_l)$. (We assume here that $1 \leq k \leq l \leq n$.) For a matrix A, we use $A_{k:l,p:q}$ to denote the submatrix of A consisting of rows from k to l and the columns from p to q. For example, $A_{1:i,1:i}$ denotes the leading (top left) $i \times i$ submatrix of A.

   (a) Show that $L_{ii}^2 = \mathbf{E}\left(x_i - \mathbf{E}(x_i|x_{1:i-1})\right)^2$. Thus, L_{ii} is the standard deviation of the error in estimating x_i, given $x_1, \ldots, x_{i-1}$.

   (b) Let $y = x_{i:n} - \mathbf{E}(x_{i:n}|x_{1:i-1})$ be the estimation error in predicting the last $n-i+1$ components of x, given the first $i-1$ components. Show that $\mathbf{E}\,\|y\|^2 = \|L_{i:n,i:n}\|_F^2$. (Here $\|\cdot\|_F$ denotes the Frobenius norm of a matrix, defined as the squareroot of the sum of the squares of its components.)

   Hint: The Cholesky factorization of $\Sigma_{1:i,1:i}$ is $L_{1:i,1:i}L_{1:i,1:i}^T$.

8. *Hadamard product.* The Hadamard product (also called the elementwise product) of two matrices $F \in \mathbf{R}^{p\times q}$, $G \in \mathbf{R}^{p\times q}$, denoted $H = F \circ G$, is defined by $H_{ij} = F_{ij}G_{ij}$. (The Hadamard product of two vectors is usually called the elementwise product.)

   Suppose x and y are independent random vectors in $\mathbf{R}^n$, with means $\mathbf{E}\,x = \bar{x}$, $\mathbf{E}\,y = \bar{y}$, and second moments $\mathbf{E}\,xx^T = X$, $\mathbf{E}\,yy^T = Y$, respectively. Define $z = x \circ y$. Show that $\mathbf{E}\,z = \bar{x} \circ \bar{y}$, and $\mathbf{E}\,zz^T = X \circ Y$.

   Remark. This implies that the Hadamard product of two positive semidefinite matrices is positive semidefinite, which is not obvious.