Problem 1: Consider the 2×2 matrix

$$\boldsymbol{A} = \frac{1}{15}\begin{bmatrix} 0 & 5 \\ -3 & -4 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 \\ 0 & \frac{1}{3\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{-3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix}.$$

(a) Compute the pseudo-inverse $\boldsymbol{A}^\dagger$ of $\boldsymbol{A}$. (In this case, the pseudo-inverse also happens to be the actual inverse.)

$\boldsymbol{A}^\dagger =$

(b) Suppose we observe

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{e}, \quad \text{where} \quad \boldsymbol{x}_0 = \begin{bmatrix} 0 \\ 3 \end{bmatrix},$$

and then take

$$\hat{\boldsymbol{x}} = \boldsymbol{A}^\dagger \boldsymbol{y}.$$

Fill in the blanks below

$$\_\_\_\_ \, \|\boldsymbol{e}\|_2^2 \;\le\; \|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2^2 \;\le\; \_\_\_\_ \, \|\boldsymbol{e}\|_2^2$$

Problem 4: Suppose that X is a Gaussian random vector in $\mathbb{R}^2$ with

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \text{Normal}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} \right)$$

(a) What is the variance $\mathrm{E}[(X_2 - \mathrm{E}[X_2])^2]$ of X_2?

Answer:

(b) Suppose we observe $X_1 = -2$. Find the solution to the following optimization program:

$$\underset{g}{\text{minimize}}\, \mathrm{E}\left[(X_2 - g)^2 | X_1 = -2\right]$$

Answer:

(c) Let $\hat{g}$ be your answer to part (b). What is $\mathrm{E}[(X_2 - \hat{g})^2 | X_1 = -2]$?

Answer:

Problem 5: Suppose that we observe samples of a known function $g(t) = t^3$ with unknown amplitude θ at (known) arbitrary locations $t_1, \ldots, t_N$, and these samples are corrupted by Gaussian noise. That is, we observe the sequence of random variables

$$X_n = \theta t_n^3 + Z_n, \quad n = 1, \ldots, N,$$

where the Z_n are independent and $Z_n \sim \text{Normal}(0, \sigma^2)$ [1].

(a) Given $X_1 = x_1, \ldots, X_N = x_N$, compute the log likelihood function

$$\ell(\theta; x_1, \ldots, x_N) = \log f_{X_1,\ldots,X_N}(x_1, \ldots, x_N; \theta) = \log\left(f_{X_1}(x_1;\theta) f_{X_2}(x_2;\theta) \cdots f_{X_N}(x_N;\theta)\right).$$

Note that the X_n are independent (as the last equality is suggesting) but not identically distributed (they have different means).

$\ell(\theta; x_1, \ldots, x_N) =$

(b) Compute the MLE for θ.

$\hat{\theta}_{\text{MLE}} =$

Problem 1: For all parts of this problem, take

$$\boldsymbol{A} = \begin{bmatrix} 1 & -3 \\ -1 & 3 \\ 1 & -3 \\ -1 & 3 \end{bmatrix}.$$

(a) Find the SVD of $\boldsymbol{A}$, $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$. (Notice that $\boldsymbol{A}$ has a rank of 1.)

$\boldsymbol{U} =$	$\boldsymbol{\Sigma} =$	$\boldsymbol{V} =$

(b) Let $\boldsymbol{y} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1 \end{bmatrix}$. Let $\mathcal{X}$ be the set of solutions[1] to

$$\underset{\boldsymbol{x} \in \mathbb{R}^2}{\text{minimize}} \; \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$$

Find the element of $\mathcal{X}$ with smallest Euclidean norm.

$\hat{\boldsymbol{x}} =$

Problem 3: Circle the appropriate responses below (showing your work is not necessary).

(a) If $\hat{\Theta}_N$ is an estimator of the parameter θ formed from independent and identically distributed random variables $X_1, \ldots, X_N$ with pdf $f_X(\boldsymbol{x}; \theta_0)$, $\theta_0 \in \mathcal{T}$ that obeys $\mathrm{E}[\hat{\Theta}_N] = \theta_0$, then for any fixed $\epsilon > 0$, $\mathrm{P}\left(\left|\hat{\Theta}_N - \theta_0\right| > \epsilon\right) \to 0$ as $N \to \infty$.

Always true Not always true

(b) If $\hat{\Theta}_N$ is an estimator of the parameter θ formed from independent and identically distributed random variables $X_1, \ldots, X_N$ with pdf $f_X(\boldsymbol{x}; \theta_0)$, $\theta_0 \in \mathcal{T}$ that obeys, for any fixed $\epsilon > 0$, $\mathrm{P}\left(\left|\hat{\Theta}_N - \theta_0\right| > \epsilon\right) \to 0$ as $N \to \infty$, then $\mathrm{E}[\hat{\Theta}_N] \to \theta_0$ as $N \to \infty$.

Always true Not always true

(c) Let $X_1, \ldots, X_N$ be independent and identically distributed random variables with $\mathrm{E}[X_n] = \theta_0$ and $\mathrm{E}[(X_n - \theta_0)^2] < \infty$. Set $\hat{\Theta}_N = \frac{1}{N}\sum_{n=1}^{N} X_n$. Then for any fixed $\epsilon > 0$, $\mathrm{P}\left(\left|\hat{\Theta}_N - \theta_0\right| > \epsilon\right) \to 0$ as $N \to \infty$.

Always true Not always true

(d) Let $X_1, \ldots, X_N$ be independent and identically distributed random variables with known mean μ and unknown finite variance θ_0. Set $\hat{\Theta}_N = \frac{1}{N}\sum_{n=1}^{N}(X_n - \mu)^2$. Circle all the conditions that are sufficient for $\mathrm{P}\left(\left|\hat{\Theta}_N - \theta_0\right| > \epsilon\right) \to 0$ as $N \to \infty$ for any $\epsilon > 0$.

(i) The X_n are discrete

(ii) The X_n are Gaussian

(iii) The X_n are zero mean ($\mu = 0$)

(iv) The X_n have finite fourth moment ($\mathrm{E}[X_n^4] < \infty$)

(e) Let $\boldsymbol{H}$ be an $N \times N$ symmetric matrix with eigenvalues $\lambda_1 = \lambda_2 = \cdots = \lambda_N = 1$. Then $\boldsymbol{H} = \mathbf{I}$.

Always true Not always true

(f) The method of conjugate gradients (CG) for solving the optimization program

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \ \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x},$$

where $\boldsymbol{H}$ is an $N \times N$ symmetric positive definite matrix, converges in at most N iterations.

Always true Not always true

(g) Consider as above the optimization program

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \ \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x},$$

where $\boldsymbol{H}$ is an $N \times N$ symmetric positive definite matrix. If $\boldsymbol{b} = \mathbf{0}$ and the initial point $\boldsymbol{x}_0$ is a linear multiple of an eigenvector of $\boldsymbol{H}$, then both gradient descent (with optimal step sizes) and conjugate gradients converge in exactly one iteration.

Always true Not always true

(h) Suppose that $\boldsymbol{H}$ is symmetric but not positive semi-definite; that is, it has at least one negative eigenvalue. What is the smallest value $\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x}$ can take for $\boldsymbol{x} \in \mathbb{R}^N$?