ASE 381P Problem Set #2

Posting Date: Sept. 13, 2019

You need not hand in anything. Instead, be prepared to answer any of these problems on an upcoming take-home exam. You may discuss your solutions with classmates up until the time that the exam becomes available, but do not swap work.

1. Suppose that you are given the scalar function of a vector, $J(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T P \boldsymbol{x} + \boldsymbol{g}^T\boldsymbol{x}$, where $\boldsymbol{x}$ is the $n \times 1$ vector argument of the function, P is a given symmetric $n \times n$ matrix, and $\boldsymbol{g}$ is a given $n \times 1$ vector. If $\partial J/\partial \boldsymbol{x}$ is defined to be the $n \times 1$ column vector $[\partial J/\partial x_1, \partial J/\partial x_2, \ldots, \partial J/\partial x_n]^T$, then prove that $\partial J/\partial \boldsymbol{x} = P\boldsymbol{x} + \boldsymbol{g}$. Also, if $\partial^2 J/\partial \boldsymbol{x}^2$ is defined to be the symmetric $n \times n$ matrix whose ijth element is $\partial^2 J/\partial x_j \partial x_i$, then prove that $\partial^2 J/\partial \boldsymbol{x}^2 = P$.

   Hints: Use summation notation to represent the value of $J(\boldsymbol{x})$ in terms of the elements of $\boldsymbol{x}$, $\boldsymbol{g}$, and P. Differentiate this expression with respect to the elements of $\boldsymbol{x}$ in order to get summation formulas for the required derivatives. Show that these summation formulas are equivalent to the matrix-vector formulas that are given above for the two required derivative expressions.

2. Suppose you are given the scalar quadratic form

$$C(\boldsymbol{x}|\boldsymbol{z}) = \frac{1}{2}\left[(\boldsymbol{x}-\bar{\boldsymbol{x}})^T, (\boldsymbol{z}-\bar{\boldsymbol{z}})^T\right] \begin{bmatrix} V_{xx} & V_{xz} \\ V_{xz}^T & V_{zz} \end{bmatrix} \begin{bmatrix} (\boldsymbol{x}-\bar{\boldsymbol{x}}) \\ (\boldsymbol{z}-\bar{\boldsymbol{z}}) \end{bmatrix}$$

   where the vectors $\bar{\boldsymbol{x}}, \boldsymbol{z}$, and $\bar{\boldsymbol{z}}$ and the matrices V_{xx}, V_{xz}, and V_{zz} are known quantities. Prove that $\partial C/\partial \boldsymbol{x} = V_{xx}(\boldsymbol{x}-\bar{\boldsymbol{x}}) + V_{xz}(\boldsymbol{z}-\bar{\boldsymbol{z}})$ and that $\partial^2 C/\partial \boldsymbol{x}^2 = V_{xx}$.

   Hints: You should be able to use your results from Problem 1. First you might want to re-express $C(\boldsymbol{x}|\boldsymbol{z})$ as a sum of scalar terms by carrying out the matrix-vector and matrix-matrix multiplications that correspond to the above block quadratic form. You should not have to resort to writing $C(\boldsymbol{x}|\boldsymbol{z})$ in terms of elemental summation formulas.

3. Given the matrices

$$A = \begin{bmatrix} 1 & 1 \\ 1 & (1+10^{-8}) \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 \\ 1 & (1+10^{-3}) \end{bmatrix}$$

   and the vectors

$$\boldsymbol{y}_A = \begin{bmatrix} 1 \\ (1-3\times 10^{-8}) \end{bmatrix} \quad \text{and} \quad \boldsymbol{y}_B = \begin{bmatrix} 1 \\ (1-3\times 10^{-3}) \end{bmatrix}$$

   solve the equations

$$A\boldsymbol{x}_A = \boldsymbol{y}_A$$

$$B\boldsymbol{x}_B = \boldsymbol{y}_B$$

   for the 2×1 vectors $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$ in the following four ways:

(a) Analytically

(b) By using the Matlab left division operator, $\boldsymbol{x}_A = A\backslash\boldsymbol{y}_A$, and $\boldsymbol{x}_B = B\backslash\boldsymbol{y}_B$.

Note that division by a matrix in Matlab is accomplished, effectively, by computing the matrix's inverse and then multiplying by that inverse. Therefore, the above solution procedures are equivalent to $\boldsymbol{x}_A = A^{-1}\boldsymbol{y}_A$, and $\boldsymbol{x}_B = B^{-1}\boldsymbol{y}_B$ if A and B are square nonsingular matrices, which they are. Note that Matlab uses QR factorization to compute the effective inverse of the matrix when an expression involving matrix division is used.

(c) By using the Matlab left division operator as part of a least-squares solution procedure, $\boldsymbol{x}_A = (A^TA)\backslash(A^T\boldsymbol{y}_A)$ and $\boldsymbol{x}_B = (B^TB)\backslash(B^T\boldsymbol{y}_B)$. Least-squares techniques are not needed for these problems because they are not over-determined, but least-squares still should give the correct answers, at least in theory. Differences may arise, however, due to the build-up of numerical round-off errors in the computer.

(d) By using the Matlab matrix inversion function, `inv()`, as part of a least-squares solution procedure $\boldsymbol{x}_A = \texttt{inv}(A^TA)(A^T\boldsymbol{y}_A)$ and $\boldsymbol{x}_B = \texttt{inv}(B^TB)(B^T\boldsymbol{y}_B)$.

Compare the four different values of $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$ that you get from these four different techniques. Be sure to use the "format long" command in Matlab so that you can see all of the digits of your numerical answers. What can you say about the various different procedures for solving the problem? What is the effect of the squaring of the matrix that occurs in the ordinary least squares solution procedure? Before you answer these questions, compute the condition numbers of A and B by using the Matlab function `cond`(). Also, compute the condition numbers of A^TA and B^TB.

In discussing your answers, keep in mind that the condition number of a matrix gives an idea of whether one can invert the matrix on a finite-precision computer and get sensible results when using the computed inverse to solve a system of linear equations. If a matrix has a high condition number, then numerical round-off errors will build up a lot during the inversion process (even if the inversion algorithm is very stable numerically), and the solution of the associated system of linear equations will not be very accurate. If one is working with a computer that is executing double-precision arithmetic, i.e., that retains 16 significant digits in its real numbers, then the number of valid significant digits in the solution to a system of linear equations of the form $A\boldsymbol{x} = \boldsymbol{y}$ should be approximately $16 - \log_{10}(\mathrm{cond}(A))$. Thus, if $\mathrm{cond}(A) = 10^{16}$, then the computed answer $\boldsymbol{x}$ will be garbage, but if $\mathrm{cond}(A) = 10^{10}$, then one would expect the first 6 significant digits of the computed $\boldsymbol{x}$ to be valid. Note that a condition number of infinity corresponds to a singular matrix, and a large condition number means that the matrix is close to being singular.

4. Consider the following binary hypothesis testing problem, which arises in the context of spoofing detection in a GPS receiver. The quantity z is a scalar measurement. You may

assume that $\sigma_1^2 > \sigma_0^2$ but don't assume any particular relationship between μ_0 and μ_1.

$$\begin{aligned} H_0 : &\quad z \sim \mathcal{N}(\mu_0, \sigma_0^2) \\ H_1 : &\quad z \sim \mathcal{N}(\mu_1, \sigma_1^2) \end{aligned}$$

(a) Find an expression for the likelihood ratio

$$\Lambda(z) = \frac{p(z|H_1)}{p(z|H_0)}$$

(b) The optimal form of the hypothesis test is as follows:

$$\Lambda(z) \underset{H_0}{\overset{H_1}{\gtrless}} \nu$$

where ν is the detection threshold. In most cases, we can simplify the test by performing equivalent operations on $\Lambda(z)$ and on ν. For example, we can often simplify by taking the log of both $\Lambda(z)$ and ν:

$$\Lambda'(z) = \log \Lambda(z) \underset{H_0}{\overset{H_1}{\gtrless}} \log \nu = \nu'$$

Further manipulation might result in an even simpler test, which we could denote as

$$\Lambda''(z) \underset{H_0}{\overset{H_1}{\gtrless}} \nu''$$

Let $\Lambda^*(z)$ denote whatever expression we get on the left-hand side of the comparison after repeated simplification, and let ν^* denote the corresponding threshold. We call $\Lambda^*(z)$ the *detection statistic*.

In the Neyman-Pearson formulation, one sets the threshold ν^* to satisfy a desired probability of false alarm P_F. It is also typically desirable to determine the probability of detection P_D associated with the chosen value of ν^*. To find the proper value of ν^* and the associated P_D, it is not enough to have an expression for $\Lambda^*(z)$; we must also know the probability distribution of $\Lambda^*(z)$ under H_0 and under H_1. In other words, we need to find $p(\Lambda^*(z)|H_j)$ for $j = 0, 1$. One can therefore appreciate the need to find a detection statistic $\Lambda^*(z)$ whose distribution is easily calculated.

Find a simple detection statistic $\Lambda^*(z)$ for the current hypothesis test and find $p(\Lambda^*(z)|H_j)$ for $j = 0, 1$. (Hint: You may wish to complete the square.)

(c) Now repeat steps (a) and (b) for a similar problem with $\boldsymbol{z}$ an n-by-1 vector:

$$H_0 : \quad \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu_0}, \sigma_0^2 I) \tag{1}$$
$$H_1 : \quad \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu_1}, \sigma_1^2 I) \tag{2}$$

5. Consider a hypothesis test about the music coming through a loudspeaker. Is the music a Brahms cello sonata (soft and lyrical) or Stravinsky's Rite of Spring second movement (loud and percussive)? For some reason, you can't hear the loudspeaker but you can measure the deflection of its diaphragm. Let the random variable z represent a measurement of the diaphragm deflection. The Brahms-vs-Stravinsky hypothesis can then be stated as

$$\begin{aligned} H_0 : &\quad z \sim \mathcal{N}(0, \sigma_0^2) \\ H_1 : &\quad z \sim \mathcal{N}(0, \sigma_1^2) \end{aligned}$$

with $\sigma_0 < \sigma_1$, where σ_0 corresponds to Brahms and σ_1 to Stravinsky.

(a) Derive a simple detection statistic for this hypothesis test. Justify simplifications as you make them.

(b) Show how to determine the value of the detection threshold for $P_F = 0.03$. Express the threshold in terms of σ_0 or σ_1—whichever is appropriate.

(c) Find the probability of detection P_D associated with this threshold for the following ratios of σ_1/σ_0: 4, 6, 8, 10. Do you think the P_D is satisfactory when $\sigma_1/\sigma_0 = 4$?

(d) Suppose we decide to take more measurements to improve P_D. We space these measurements far enough apart in time so that they are independent of one another. Derive a simple detection statistic that makes use of N measurements $z_1, z_2, ..., z_N$ of the diaphragm deflection. Again, justify simplifications as you make them.

(e) Assuming $\sigma_1/\sigma_0 = 4$, how large must N be to ensure $P_D > 0.99$ for $P_F = 0.001$?

6. You've been given a coin that when tossed shows heads with probability θ and tails with probability $1 - \theta$. You wish to estimate θ by experimentation. You toss the coin N times, yielding N independent observations $y_1, y_2, ..., y_N$, where

$$y_k = \begin{cases} 1 & \text{if } k\text{th toss shows heads} \\ 0 & \text{if } k\text{th toss shows tails} \end{cases}$$

The observations are packaged together in a vector $\boldsymbol{y} = [y_1, y_2, ..., y_N]^T$.

(a) Suppose $N = 3$ and $\boldsymbol{y}^* = [1, 1, 0]^T$ is the outcome of your experiment. Fill out the table below with the probability of the outcome $\boldsymbol{y}^*$ given θ.

θ	$P(\boldsymbol{y} = \boldsymbol{y}^* \vert \theta)$
0.2	
0.4	
0.8	
1.0	

(b) Let $n_1 = \sum_{k=1}^{N} y_k$ be the number of heads and $n_0 = N - n_1$ be the number of tails. Write an expression for $P(\boldsymbol{y}|\theta)$ in terms of θ, n_1, and n_0.

(c) Derive $\hat{\theta}_{\mathrm{ML}}$, the maximum likelihood estimate of θ, from the expression in part (b).

Also do the following problems from Bar Shalom: 1-9, 2-1 (Clarification: in parts 2, 3, and 4, the MSEs and "associated variance" are all conditioned on z), 2-3 (assume that A is a symmetric matrix), 2-5, 2-7 (the given distribution should be $p(x|z)$).