ECE 174 - Homework # 5 Solutions

1. (a) The MLE, $\widehat{x}_{\text{MLE}}$, is determined as

$$\widehat{x}_{\text{MLE}} = \arg\min_x \{-\ln p(y\,|x)\} = \arg\min_x \sum_{i=1}^{2} \frac{1}{\sigma_i^2}(y_i - x)^2 = \arg\min_x \|y - Ax\|_W^2 \,,$$

where

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \,, \quad A = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \,, \quad W = \begin{pmatrix} w_1 & 0 \\ 0 & w_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{pmatrix} \,.$$

We have that the adjoint operator is given by

$$A^* = A^T W = (w_1 \ , \ w_2) = \left(\frac{1}{\sigma_1^2} \ , \ \frac{1}{\sigma_2^2} \right) \,,$$

so that,

$$A^* A = w_1 + w_2 = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \,,$$

which is invertible (as we already knew since A has full column rank). Continuing, we obtain,

$$(A^* A)^{-1} = \frac{1}{w_1 + w_2} = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \,,$$

and

$$A^+ = (A^* A)^{-1} A^* = \left(\frac{w_1}{w_1 + w_2} \ , \ \frac{w_2}{w_1 + w_2} \right) = \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \ , \ \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \right) \,.$$

Thus,

$$\widehat{x}_{\text{MLE}} = A^+ y = \alpha_1 \, y_1 + \alpha_2 \ y_2 \,,$$

where

$$0 \le \alpha_i = \frac{w_i}{w_1 + w_2} = \frac{\sigma_i^2}{\sigma_1^2 + \sigma_2^2} \le 1 \,, \quad \text{for} \quad i = 1, 2 \,,$$

and $\alpha_1 + \alpha_2 = 1$. This shows that the MLE is a so–called *convex combination* (i.e., a normalized weighted average) of the measurements y_1 and y_2. Recalling that $w_i = \frac{1}{\sigma_i^2}$ gives the *precision* of the measurement y_i, we see that the weight α_i is a measure of the normalized (or relative) precision.

(b) In the limit that $\sigma_1^2 \to 0$ ($w_1 \to \infty$), we see from the general solution derived above that $\widehat{x}_{\text{MLE}} \to y_1$. This makes sense as $\sigma_1^2 \to 0$ ($w_1 \to \infty$) corresponds to y_1 becoming a perfect, non-noisy (infinitely precise) measurement of the unknown quantity x. In the limit that $\sigma_1^2 \to \infty$, ($w_1 \to 0$)the general solution shows that $\widehat{x}_{\text{MLE}} \to y_2$. This makes sense as $\sigma_1^2 \to \infty$ ($w_1 \to 0$)corresponds to the first measurement becoming so noisy (so imprecise) that it is worthless relative to the second measurement.

(c) In the case that $\sigma_1^2 = \sigma_2^2 \equiv \sigma^2$ ($w_1 = w_2 \equiv w$) we have,

$$\widehat{x}_{\text{MLE}} = \arg\min_x \|y - Ax\|_W^2 = \arg\min_x \frac{1}{\sigma^2} \|y - Ax\|^2 = \arg\min_x \|y - Ax\|^2 ,$$

the last expression being an unweighted least–squares problem. Using the condition $\sigma_1^2 = \sigma_2^2$ ($w_1 = w_2$), the general solution derived earlier becomes,

$$\widehat{x}_{\text{MLE}} = \frac{1}{2}(y_1 + y_2) .$$

This solution makes sense because the condition $\sigma_1^2 = \sigma_2^2$ means that the two measurements are equally precise, $w_1 = w_2$, so that we have *no* rational reason to prefer one measurement over the other. This is because the errors in the two measurements are additive, independent, zero mean, and identically symmetrically-distributed measurement errors. This symmetry condition yields a linear estimator which is the symmetric sample–mean solution.[1]

2. Note that $\mathrm{E}\{y_i\} = \alpha t_i$.

(a) The least–squares problem to be solved is

$$\min_\alpha \|y - A\alpha\|^2 ,$$

where

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} t_1 \\ \vdots \\ t_m \end{pmatrix} .$$

The least–squares solution is

$$\widehat{\alpha}(m) = A^+ y = (A^T A)^{-1} A^T y = \frac{\sum_{i=1}^m t_i y_i}{\sum_{i=1}^m t_i^2} .$$

(b)

$$\mathrm{E}\{\widehat{\alpha}(m)\} = \frac{\sum_{i=1}^m t_i \mathrm{E}\{y_i\}}{\sum_{i=1}^m t_i^2} = \alpha \frac{\sum_{i=1}^m t_i^2}{\sum_{i=1}^m t_i^2} = \alpha .$$

(c)

$$\mathrm{Var}\{\widetilde{\alpha}(m)\} = \sigma^2 (A^T A)^{-1} = \frac{\sigma^2}{\sum_{i=1}^m t_i^2} \to 0 \quad \text{as} \quad m \to \infty .$$

[1] In advanced courses we would say that "the solution to the optimal linear estimator is obvious from symmetry."

3. **Vocabulary.** The relevant definitions can be found in your class lecture notes and/or in the lecture supplement handouts.

4. **Scalar Generalized Gradient Descent Algorithms.**

   Note that in the scalar case the loss function is

$$\ell(x) = \frac{1}{2}\left(y - h(x)\right)^2 .$$

   Also the gradient of $\ell(x)$ is just the derivative of $\ell(x)$ with respect to x,

$$\ell'(x) = \nabla\ell(x) = \frac{d}{dx}\ell(x) = -h'(x)\left(y - h(x)\right) ,$$

   where $h'(x) = \frac{d}{dx}h(x)$. Note also that the second derivative of the loss function is

$$\ell''(x) = h'^2(x) - h''(x)\left(y - h(x)\right) .$$

   (a) **Gradient Descent Method.**

$$\widehat{x}_{k+1} = \widehat{x}_k - \alpha_k \ell'(\widehat{x}_k) = \widehat{x}_k + \alpha_k h'(\widehat{x}_k)\left(y - h(\widehat{x}_k)\right) ,$$

   where the step size $\alpha_k > 0$ is used for convergence control. It is evident that $Q_k = 1$ for all k.

   (b) **Gauss' Method (Gauss–Newton Method).**

   To find a correction to the current estimate $\widehat{x}$, we linearize the nonlinear inverse problem $y = h(x)$ about the point $\widehat{x}$,

$$y \approx h(\widehat{x}) + h'(\widehat{x})(x - \widehat{x}) = h(\widehat{x}) + h'(\widehat{x})\Delta x ,$$

   where $\Delta x = x - \widehat{x}$, and find a solution which is optimal in the least–squares sense for this linearized problem. Note that the loss function for this linearized problem is

$$\ell_{\text{gauss}}(x) = \frac{1}{2}\left(y - [h(\widehat{x}) + h'(\widehat{x})\left(x - \widehat{x}\right)]\right)^2 = \frac{1}{2}\left(\Delta y - h'(\widehat{x})\Delta x\right)^2 = \ell_{\text{gauss}}(\Delta x) ,$$

   where $\Delta y = y - h(\widehat{x})$. Also note that because

$$x = \widehat{x} + \Delta x ,$$

   where $\widehat{x}$ is given and fixed, it is evident that optimizing over x is equivalent to optimizing over Δ. An equivalent problem, then, is to find the correction Δx which is optimal in the least–squares sense by minimizing $\ell_{\text{gauss}}(\Delta x)$ with respect to Δx.

The solution is

$$\Delta x = \frac{1}{h'(\widehat{x})}\ \Delta y = \frac{1}{h'(\widehat{x})}\ \ (y - h(\widehat{x}))\ ,$$

assuming that $h'(\widehat{x}) \neq 0$. Incorporating a step size $\alpha > 0$ for convergence control, yields

$$\Delta x = \alpha(\widehat{x})\,\frac{1}{h'(\widehat{x})}\ \ (y - h(\widehat{x})) = \alpha(\widehat{x})\,Q(\widehat{x})h'(\widehat{x})\,(y - h(\widehat{x}))\ ,$$

where

$$Q(\widehat{x}) = \frac{1}{h'^2(\widehat{x})}\,.$$

This yields the iterative algorithm,

$$\widehat{x}_{k+1} = \widehat{x}_k + \alpha_k \frac{1}{h'(\widehat{x}_k)}\,(y - h(\widehat{x}_k)) = \widehat{x}_k + \alpha_k Q_k h'(\widehat{x}_k)\,(y - h(\widehat{x}_k))$$

where

$$Q_k = \frac{1}{h'^2(\widehat{x}_k)}\,.$$

(c) **Newton's Method.**

To find a correction to the current estimate $\widehat{x}$ we expand the loss function $\ell(x)$ about $\widehat{x}$ to second order and then minimize this quadratic approximation to $\ell(x)$. With $\Delta x = x - \widehat{x}$, the quadratic approximation is

$$\ell(x) \approx \ell_{\text{quad}}(x) = \ell(\widehat{x}) + \ell'(\widehat{x})\Delta x + \frac{1}{2}\ell''(\widehat{x})(\Delta x)^2 = \ell_{\text{quad}}(\Delta x)\,.$$

Note that $\ell_{\text{quad}}(x) = \ell_{\text{quad}}(\Delta x)$ can be equivalent minimized either with respect to x or with respect to Δx. If we take the derivative of $\ell_{\text{quad}}(\Delta x)$ with respect Δx and set it equal to zero we get

$$\Delta x = -\frac{\ell'(\widehat{x})}{\ell''(\widehat{x})} = \frac{h'(\widehat{x})\,(y - h(\widehat{x}))}{h'^2(\widehat{x}) - h''(\widehat{x})\,(y - h(\widehat{x}))}\,. \tag{1}$$

If we multiply the correction Δx by a step size $\alpha > 0$ in order to control convergence we obtain the iterative algorithm

$$\widehat{x}_{k+1} = \widehat{x}_k + \alpha_k Q_k h'(\widehat{x}_k)\,(y - h(\widehat{x}_k))\ ,$$

where

$$Q_k = \frac{1}{h'^2(\widehat{x}) - h''(\widehat{x})\,(y - h(\widehat{x}))}\,.$$

Comparing the values of Q_k for the Gauss method and the Newton method, ***we see that when*** $y - h(\widehat{x}_k) \approx 0$ ***the two methods become essentially***

equivalent.[2] Also note that when $h'' \equiv 0$ the two methods become exactly equivalent.

5. **Simple Nonlinear Inverse Problem Example.**

Let y denote the known number for which you wish to determine the cube root. Let x denote the unknown cube root of y. We need a relationship $y = h(x)$, which is *obviously* given by $h(x) = x^3$. This relationship defines an inverse problem that we wish to solve. (I.e., given the relationship $y = h(x) = x^3$ and a value for y determine a corresponding value for x.) The relevant derivatives needed to implement the descent algorithms are

$$h'(x) = 3x^2 \quad \text{and} \quad h''(x) = 6x\,.$$

[2]This fact also holds for the more general vector case. For example, because this latter situation holds for the GPS computer assignment, the Gauss-Newton method used in that assignment is essentially equivalent to the Newton method and therefore has the very fast convergence rate associated with the Newton method.