ECE 174 Homework # 5 Solutions Spring 2009

FINAL EXAM is Tuesday, June 9, 3-6pm, in our regular classroom.

The final exam is closed notes and book. Bring only a pen (no pencils). I can ask vocabulary questions on the final exam and expect you to know technical terms used in the problem statements.¹ The best way to study for the exam is to understand concepts and study all of the homework solutions, lecture notes, projects, and lecture supplements.

Solutions

1. (a) We first determine that $m = 3$, $n = 1$, $r = 1$, $\nu = 0$, and $\mu = 2$. Once the dimensions of the various subspaces are known, we can systematically work the matrix $A$ into the appropriate factorization,²

\[
A = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} (\sqrt{2})(1) = U_1 S V_1^T
\]

\[
= \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} (\sqrt{2})(1) = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \begin{pmatrix} S \\ 0 \end{pmatrix} (V_1^T) = U \Sigma V^T.
\]

All the columns of $U$ provide an orthonormal basis for $\gamma = \mathbb{R}^3$. The column of $U_1$ spans $\mathcal{R}(A)$ and the two columns of $U_2$ provide an orthonormal basis for $\mathcal{N}(A^T)$. The single element of $V^T = V_1^T$ spans $\mathcal{R}(A^T) = \mathbb{R} = \mathcal{X}$. The nullspace of $A$ is trivial. With the SVD in hand, it is readily determined using the formulas given earlier that $\sigma_1 = \sqrt{2} > 0$ and

\[
A^+ = \frac{1}{2} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix};
\]

\[
P_{\mathcal{R}(A)} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix};
\]

\[
P_{\mathcal{N}(A^T)} = \frac{1}{2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix};
\]

\[
P_{\mathcal{R}(A^T)} = 1;
\]

\[
P_{\mathcal{N}(A)} = 0.
\]

¹Thus during the exam I will not answer questions about questions asking for clarifications to to terms or concepts that you are expected to already know.
²Note that these matrices have been hand crafted to be amenable to this approach. Generally numerical techniques must be used to obtain the SVD.
Note that because $A$ has full column rank we can also compute $A^+$ directly as $A^+ = (A^T A)^{-1} A^T$.

(b) We first determine that $m = 3$, $n = 2$, $r = 2$, $\nu = 0$, and $\mu = 1$. Again, once the dimensions of the various subspaces are known, we can systematically work the matrix $A$ into the appropriate factorization,

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 1 \end{pmatrix} (1 \ 0) = U_1 S V_1^T$$

$$= \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 1 \end{pmatrix} (1 \ 0) = (U_1 \ U_2) \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U \Sigma V^T.$$ 

All the columns of $U$ provide an orthonormal basis for $Y = \mathbb{R}^3$. The two columns of $U_1$ gives an orthonormal basis for $\mathcal{R}(A)$ and the column of $U_2$ spans $\mathcal{N}(A^T)$. Both rows of $V^T = V_1^T$ provide an orthonormal basis for $\mathcal{R}(A^T) = \mathbb{R}^2 = X$. The nullspace of $A$ is trivial. It now can be readily determined that $\sigma_1 = \frac{\sqrt{2}}{2} > \sigma_2 = 1 > 0$ and

$$A^+ = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix};$$

$$P_{\mathcal{R}(A)} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix};$$

$$P_{\mathcal{N}(A^T)} = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix};$$

$$P_{\mathcal{R}(A)} = I, \quad P_{\mathcal{N}(A)} = 0.$$ 

Note that because $A$ has full column rank we can also compute $A^+$ directly as $A^+ = (A^T A)^{-1} A^T$.

(c) Here $m = n = 3$ while $r = 1$. For this case, the matrix $A$ is rank–deficient (i.e., is neither one–to–one nor onto) and hence cannot be constructed using an analytic formula as can be done for the full–rank cases. We also have $\nu = \mu = 1$. The
matrix $A$ factors as

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 & \sqrt{2} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \\ 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} \end{pmatrix} = U_1 SV_1^T$$

$$= \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \\ 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

$$= (U_1 \ U_2) \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U \Sigma V^T.$$  

All the columns of $U$ provide an orthonormal basis for $\mathcal{Y} = \mathbb{R}^3$. The two columns of $U_1$ gives an orthonormal basis for $\mathcal{R}(A)$ and the column of $U_2$ spans $\mathcal{N}(A^T)$. All the rows of $V^T$ provide an orthonormal basis for $\mathcal{X} = \mathbb{R}^3$. The two rows of $V_1^T$ provide an orthonormal basis for $\mathcal{R}(A^T)$. The nullspace of $A$ is spanned by the row of $V_2^T$. It can now be readily determined that $\sigma_1 = 2 > \sigma_2 = 1 > 0$ and 

$$P_{\mathcal{R}(A)} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

$$P_{\mathcal{N}(A^T)} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

$$P_{\mathcal{R}(A^T)} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

$$P_{\mathcal{N}(A)} = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$  

Note that because $A$ is symmetric it just happens to be the case that $P_{\mathcal{R}(A)} = P_{\mathcal{R}(A^T)}$ and $P_{\mathcal{N}(A)} = P_{\mathcal{N}(A^T)}$; however this is not true for a general nonsymmetric matrix $A$.

$^3$The special properties $P_{\mathcal{R}(A)} = P_{\mathcal{R}(A^*)}$ and $P_{\mathcal{N}(A)} = P_{\mathcal{N}(A^*)}$ hold for any self-adjoint matrix $A$, $A = A^*$.
(d) Here \( m = 2, n = 3, \) and \( r = 1. \) Thus \( A \) is rank–deficient and non–square (and definitely nonsymmetric).

\[
A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{5}}{2} \\ \frac{\sqrt{5}}{2} \end{pmatrix} \begin{pmatrix} \sqrt{5} & \sqrt{5} & \sqrt{5} \end{pmatrix}
\]

\[
A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{5}}{2} \\ \frac{\sqrt{5}}{2} \end{pmatrix} (1 & 1 & 1)
\]

\[
A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{5}}{2} \sqrt{15} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} = U_1 SV^T
\]

\[
A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{5}}{2} \sqrt{15} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}
\]

\[
A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{5}}{2} \sqrt{15} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{5}}{2} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}
\]

Both columns of \( U \) provide an orthonormal basis for \( \mathcal{Y} = \mathbb{R}^2. \) The column of \( U_1 \) spans \( \mathcal{R}(A) \) and the column of \( U_2 \) spans \( \mathcal{N}(A^T). \) All the rows of \( V^T \) provide an orthonormal basis for \( \mathcal{X} = \mathbb{R}^3. \) The row of \( V_1^T \) spans \( \mathcal{R}(A^T). \) The nullspace of \( A \) is spanned by the two rows of \( V_2^T. \) It can now be readily determined that \( \sigma_1 = \sqrt{15} > 0 \) and

\[
A^+ = \frac{1}{15} \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{pmatrix};
\]

\[
P_\mathcal{R}(A) = \frac{1}{5} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix};
\]

\[
P_\mathcal{N}(A^T) = \frac{1}{5} \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix};
\]

\[
P_\mathcal{R}(A^T) = \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix};
\]

\[
P_\mathcal{N}(A) = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.
\]

2. The MLE, \( \hat{x}_{MLE} \), is determined as

\[
\hat{x}_{MLE} = \arg \min_x \{ -\ln p(y|x) \} = \arg \min_x \sum_{i=1}^{2} \frac{1}{\sigma_i^2} (y_i - x)^2 = \arg \min_x \| y - Ax \|^2_W,
\]
where
\[ y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad A = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad W = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{pmatrix}. \]

We have that the adjoint operator is given by
\[ A^* = A^T W = \begin{pmatrix} \frac{1}{\sigma_1^2} \\ \frac{1}{\sigma_2^2} \end{pmatrix}, \]
so that,
\[ A^* A = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}, \]
which is invertible (as we already knew since \( A \) has full column rank). Continuing, we obtain,
\[ (A^* A)^{-1} = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \]
and
\[ A^+ = (A^* A)^{-1} A^* = \left( \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right). \]
Thus,
\[ \hat{x}_{MLE} = A^+ y = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} y_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} y_2 = \alpha_1 y_1 + \alpha_2 y_2, \]
where
\[ 0 \leq \alpha_i = \frac{\sigma_i^2}{\sigma_1^2 + \sigma_2^2} \leq 1, \quad \text{for} \quad i = 1, 2, \]
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 \( y_2 \).

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

(b) In the case that \( \sigma_1 = \sigma_2 = \sigma \) we have,
\[ \hat{x}_{MLE} = \arg \min_x \| y - Ax \|_W^2 = \arg \min_x \frac{1}{\sigma^2} \| y - Ax \|^2 = \arg \min_x \| y - Ax \|^2, \]
the last expression being an unweighted least-squares problem. Using the condition \( \sigma_1 = \sigma_2 = \sigma \), the general solution derived above becomes,
\[ \hat{x}_{MLE} = \frac{1}{2} (y_1 + y_2). \]
This solution makes sense because the condition $\sigma_1 = \sigma_2 = \sigma$ means that we have no rational reason to prefer one measurement over the other. The errors in both sets of measurements are additive, independent, zero mean, identically symmetrically-distributed measurement errors, and this symmetric condition yields a linear estimator which is the symmetric sample–mean solution. \(^4\)

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

(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

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

(b) 

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

(c) 

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

4. (a) Note that for $A$ one–to–one, $A^+ A = (A^T A)^{-1} A^T A = I$. Let $e$ be the vector of all ones, $e = (1 \cdots 1)^T$. Note that

$$E\{n\} = b \quad \text{where} \quad b = \beta e.$$ 

We have that

$$\tilde{x} = \hat{x} - x = A^+ y - x = A^+(Ax + n) - x = x + A^+n - x = A^+n.$$ 

Thus

$$E\{\tilde{x}\} = A^+ E\{n\} = A^+ b \neq 0.$$ 

\(^4\)In advanced courses we would say that “the solution to the optimal linear estimator is obvious from symmetry.”
(b) We have
\[ y = Ax + n = Ax + b + (n - b) = Ax + b + \nu \]
where \( \nu = n - b = n - \beta e \) has zero mean. Continuing,
\[ y = \begin{pmatrix} A & e \end{pmatrix} \begin{pmatrix} x \\ \beta \end{pmatrix} + \nu = Ax + \nu, \]
where
\[ A = \begin{pmatrix} A & e \end{pmatrix} \quad \text{and} \quad x = \begin{pmatrix} x \\ \beta \end{pmatrix}. \]
Assuming that \( A \) is one-to-one, we have
\[ \hat{x} = A^+ y = (A^T A)^{-1} A^T y \]
and
\[ \tilde{x} = A^+ \nu. \]
Since \( \nu \) has zero mean, \( \tilde{x} \) is unbiased.

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

6. **Scalar Generalized Gradient Descent Algorithms.**

Note that in the scalar case the loss function is
\[ \ell(x) = \frac{1}{2} (y - h(x))^2. \]
Also the gradient of \( \ell(x) \) is just the derivative of \( \ell(x) \) with respect to \( x \),
\[ \ell'(x) = \nabla \ell(x) = \frac{d}{dx} \ell(x) = -h'(x) (y - h(x)) , \]
where \( h'(x) = \frac{d}{dx} h(x) \). Note also that the second derivative of the loss function is
\[ \ell''(x) = h''(x) - h''(x) (y - h(x)) . \]

(a) **Gradient Descent Method.**
\[ \hat{x}_{k+1} = \hat{x}_k - \alpha_k \ell'(\hat{x}_k) = \hat{x}_k + \alpha_k h'(\hat{x}_k) (y - h(\hat{x}_k)) , \]
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 \( \hat{x} \), we linearize the nonlinear inverse problem \( y = h(x) \) about the point \( \hat{x} \),

\[
y \approx h(\hat{x}) + h'(\hat{x})(x - \hat{x}) = h(\hat{x}) + h'(\hat{x})\Delta x,
\]

where \( \Delta x = x - \hat{x} \), 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

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

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

\[
x = \hat{x} + \Delta x,
\]

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

The solution is

\[
\Delta x = \frac{1}{h'(\hat{x})} \Delta y = \frac{1}{h'(\hat{x})} (y - h(\hat{x})) ,
\]

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

\[
\Delta x = \alpha(\hat{x}) \frac{1}{h'(\hat{x})} (y - h(\hat{x})) = \alpha(\hat{x}) Q(\hat{x}) h'(\hat{x}) (y - h(\hat{x})) ,
\]

where

\[
Q(\hat{x}) = \frac{1}{h'^2(\hat{x})} .
\]

This yields the iterative algorithm,

\[
\hat{x}_{k+1} = \hat{x}_k + \alpha_k \frac{1}{h'(\hat{x}_k)} (y - h(\hat{x}_k)) = \hat{x}_k + \alpha_k Q_k h'(\hat{x}_k) (y - h(\hat{x}_k))
\]

where

\[
Q_k = \frac{1}{h'^2(\hat{x}_k)} .
\]

(c) Newton’s Method.

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

\[
\ell(x) \approx \ell_{\text{quad}}(x) = \ell(\hat{x}) + \ell'(\hat{x})\Delta x + \frac{1}{2} \ell''(\hat{x})(\Delta x)^2 = \ell_{\text{quad}}(\Delta x) .
\]
Note that \( \ell_{\text{quad}}(x) = \ell_{\text{quad}}(\Delta x) \) can be equivalent minimized either with respect to \( x \) or with respect to \( \Delta x \). If we take the derivative of \( \ell_{\text{quad}}(\Delta x) \) with respect \( \Delta x \) and set it equal to zero we get

\[
\Delta x = -\left( \frac{\ell'(\hat{x})}{\ell''(\hat{x})} \right) = \frac{h'(\hat{x})(y - h(\hat{x}))}{h'^2(\hat{x}) - h''(\hat{x})(y - h(\hat{x}))} .
\]  

(1)

If we multiply the correction \( \Delta x \) by a step size \( \alpha > 0 \) in order to control convergence we obtain the iterative algorithm

\[
\hat{x}_{k+1} = \hat{x}_k + \alpha_k Q_k h'(\hat{x}_k)(y - h(\hat{x}_k)) ,
\]

where

\[
Q_k = \frac{1}{h'^2(\hat{x}) - h''(\hat{x})(y - h(\hat{x}))} .
\]

Comparing the values of \( Q_k \) for the Gauss method and the Newton method, we see that when \( y - h(\hat{x}_k) \approx 0 \) the two methods become essentially equivalent.\(^5\) Also note that when \( h'' \equiv 0 \) the two methods become exactly equivalent.

7. 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 \( y = h(x) = x^3 \) determine \( x \).) The relevant derivatives needed to implement the descent algorithms are

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

\(^5\)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.