The final exam is closed notes and book. Bring paper, pencil, and a nonprogrammable calculator. I can ask vocabulary questions on the final exam, or expect you to know technical terms used in the problem statements.\footnote{Thus during the final exam I will not answer questions about terms or concepts that you are expected to know.}

The best way to study for the final exam is to understand concepts and study all of the homework and midterm solutions, lecture notes, sample midterm, computer projects, and lecture supplements. Create a unified geometric and practical understanding of the material presented in the class.

There will be office hours from noon-2 pm Monday, June 11 in room EBU-1 5605.

Solutions

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


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

\[ \ell''(x) = h''(x) (y - h(x)). \]

(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

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

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.\(^2\)


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

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

4. Lagrange Multiplier Example.

We wish to minimize

\[
\ell(x) = 2x_1^2 + 27x_2^3
\]

subject to the constraint

\[
g(x) = x_1 - 9x_2 - 9 = 0. \tag{2}
\]

The lagrangian is given by

\[
\mathcal{L}(x, \lambda) = \ell(x) + \lambda g(x) = 2x_1^2 + 27x_2^3 + \lambda (x_1 - 9x_2 - 9) .
\]

Setting the partial derivative of \( \mathcal{L}(x, \lambda) \) wrt \( \lambda \) equal to 0 retrieves the constraint condition (2) which we can rewrite as

\[
x_1 = 9 + 9x_2. \tag{3}
\]

\(^2\)Because this latter situation holds for the GPS computer assignment, the Gauss method used in that assignment has the very fast convergence associated with the Newton method.
Setting the partial derivative of $L(x, \lambda)$ wrt $x_1$ equal to zero yields
\[ 4x_1 + \lambda = 0. \]  
(4)

Setting the partial derivative of $L(x, \lambda)$ wrt $x_2$ equal to zero yields
\[ 9x_2^2 - \lambda = 0. \]  
(5)

(4)+(5) $\Rightarrow$
\[ 9x_2^2 + 4x_1 = 0. \]  
(6)

(3)+(6) $\Rightarrow$
\[ x_2^2 + 4x_2 + 4 = (x_2 + 2)^2 = 0, \]
or $x_2 = -2$. Substituting this value into (3) then yields $x_1 = -9$.

5. **Root finding and Minimization.**

The equivalence obviously follows from equation (1) (ignore the far right-hand-side of (1) which applies only to the least-squares problem) once we make the identification $f(x) = \ell'(x)$.