4-6TE3: Midterm Solution 2011-12

Christopher Anand

TA: Shefali Kulkarni-Thaker

November 6, 2011

---

1. **20 points.**

$$
\begin{aligned}
f(x,y) &= \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & -4 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= \begin{pmatrix} x+y & -4x-y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= x^2 + xy - 4xy - y^2 \\
&= x^2 - 3xy - y^2
\end{aligned}
$$

(a) **2.5+2.5 points.** Gradient and Hessian are given as

$$
\begin{aligned}
\nabla f &= \begin{pmatrix} 2x - 3y \\ -3x - 2y \end{pmatrix} \\
\nabla^2 f &= \begin{pmatrix} 2 & -3 \\ -3 & -2 \end{pmatrix}
\end{aligned}
$$

(b) **2.5 points** Second order Taylor's series at $x_0 = (0,0)^T$ is given as:

$$
f(x) = f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)
$$

$$
f(0,0) = 0 \qquad \nabla f(0,0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad x - x_0 = \begin{pmatrix} x - 0 \\ y - 0 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}
$$

$$
\begin{aligned}
f(x) &= 0 + \begin{pmatrix} 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 2 & -3 \\ -3 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 2 & -3 \\ -3 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 2 & -3 \\ -3 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} 2x - 3y & -3x - 2y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= \frac{1}{2}(2x^2 - 3xy - 3xy - 2y^2) \\
&= \frac{1}{2}(2x^2 - 6xy - 2y^2) = x^2 - 3xy - y^2 = f(x,y)
\end{aligned}
$$

[2.5 points] At $x_0 = (1, -1)^T$, the second order Taylor's series is given as: Here

$$f(1,-1) = 3 \qquad \nabla f(1,-1) = \begin{pmatrix} 5 \\ -1 \end{pmatrix} \qquad (x - x_0) = \begin{pmatrix} x-1 \\ y+1 \end{pmatrix}$$

$$\begin{aligned}
f(x) &= 3 + \begin{pmatrix} 5 & -1 \end{pmatrix} \begin{pmatrix} x-1 \\ y+1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x-1 & y+1 \end{pmatrix} \begin{pmatrix} 2 & -3 \\ -3 & -2 \end{pmatrix} \begin{pmatrix} x-1 \\ y+1 \end{pmatrix} \\
&= 3 + 5x - 5 - y - 1 + \frac{1}{2} \begin{pmatrix} x-1 & y+1 \end{pmatrix} \begin{pmatrix} 2 & -3 \\ -3 & -2 \end{pmatrix} \begin{pmatrix} x-1 \\ y+1 \end{pmatrix} \\
&= 5x - y - 3 + \frac{1}{2} \begin{pmatrix} 2(x-1) - 3(y+1) & -3(x-1) - 2(y+1) \end{pmatrix} \begin{pmatrix} x-1 \\ y+1 \end{pmatrix} \\
&= 5x - y - 3 + \frac{1}{2} \begin{pmatrix} 2x - 3y - 5 \\ -3x - 2y + 1 \end{pmatrix}^T \begin{pmatrix} x-1 \\ y+1 \end{pmatrix} \\
&= 5x - y - 3 + \frac{1}{2}\{(2x - 3y - 5)(x-1) + (-3x - 2y + 1)(y+1)\} \\
&= 5x - y - 3 + \frac{1}{2}\{2x^2 - 10x - 6xy - 2y^2 + 2y + 6\} \\
&= 5x - y - 3 + x^2 - 5x - 3xy - y^2 + y + 3 \\
&= x^2 - 3xy - y^2 = f(x,y)
\end{aligned}$$

(c) **5 points** At point $(0,0)^T$ eigenvalues of the Hessian are:

$$\begin{aligned}
\begin{pmatrix} 2-\lambda & -3 \\ -3 & -2-\lambda \end{pmatrix} &= 0 \\
-(2-\lambda)(2+\lambda) - 9 &= 0 \\
-4 - \lambda^2 - 9 &= 0 \\
\lambda^2 - 13 &= 0 \\
\lambda = \pm\sqrt{13}
\end{aligned}$$

One if the eigenvalues is negative and hence the function is *not convex.*

(d) **5 points** Newton's method is given by: $x_1 = x_0 - \nabla^2 f(x_0) \nabla f(x_0)$. $\nabla f(x_0) = 0$ at point $x_0 = (0,0)^T$. Hence $x_1 = x_0 = (0,0)^T$.

2. **20 points**

It is given that computing $x = A^{-1}b$ required N operations. Let us calculate the total number of operations required to compute $y = (A + uv^T)^{-1}c$ using Sherman-Morrison formula.

$$\begin{aligned}
y &= \Big(A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1} u}\Big)c \\
&= A^{-1}c - \frac{A^{-1}uv^T A^{-1}c}{1 + v^T A^{-1} u}
\end{aligned}$$

(a) Computing $A^{-1}u = w_1 \implies N$ operations. $w_1 \in R^n$

(b) Computing $A^{-1}c = w_2 \implies N$ operations. $w_2 \in R^n$. Thus, $y = w_2 - \frac{w_1 v^T w_2}{1 + v^T w_1}$

(c) Computing $v^T w_2 = t_2 \implies n$ operations (ignoring additions) because $v, w_2 \in R^n$.

(d) Computing $v^T w_1 = t_1 \implies n$ operations (ignoring additions) because $v, w_1 \in R^n$. Hence, $y = w_2 - w_1 \frac{t_2}{1+t_1}$

(e) $\alpha = \frac{t_2}{1+t_1}$ is a scalar division, because t_1, t_2 are scalars, requiring constant time. Thus, $y = w_2 - \alpha w_1$

(f) $q_1 = \alpha w_1$ requires n operations as $w_1 \in R^n$ and $\alpha \in R$

Thus, ignoring subtractions in $y = w_2 - q_1$, we get $2N + 3n$ operations to compute y in $y = (A + uv^T)^{-1}c$ using Sherman-Morrison formula.

3. **20 points**

   Newton's method will converge in one iteration for a convex quadratic function. (**10 points**) Depending on the starting point chosen, Newton's method may not converge in one step. Figure 1 shows Newton's taking more than one iteration for function $f(x) = \dfrac{x}{1+x^2}$, with starting point $x_0 = 0.5$ (**10 points**)

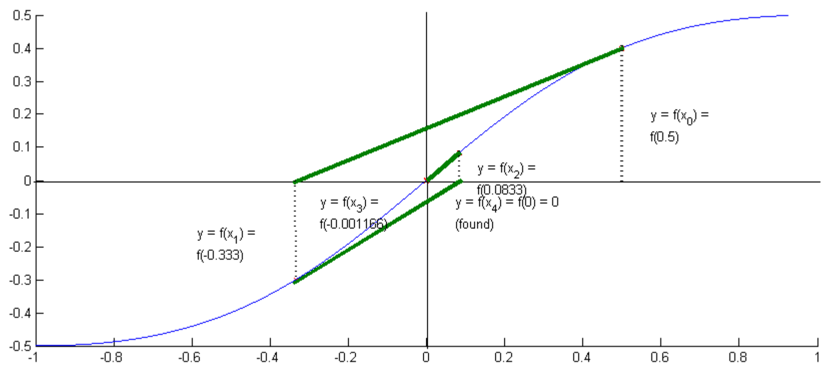

Figure 1: Newton's method converging with $x_0 = 0.5$

4. **20 points**

   In exact line search algorithms, like secant method, am exact minimizer α_k is found in the given search direction. i.e. $\min_{\alpha_k > 0} f(x^k + \alpha_k s^k)$. The minimum of this optimization problem is found by solving the equation, $\nabla f(x^k + \alpha_k s^k) = 0$. In most iterations, the direction s^k may not necessary point to the true minimum. Moreover, it is not always easy to calculate gradient of the function. Thus, exact line search method can turn out to be computationally expensive operation.

   In contrast to exact line search methods, inexact line search methods will try to find an approximate value of $\alpha_k > 0$ so that there is sufficient decrease in the function value at the new point. In this case the step-length can take range of values where the function is approximately minimized. This will save a lot of computational effort that is required by exact line search methods. The examples of inexact line searches are Goldstein-Armijo and Wolfe conditions.

5. **20 points**

   G is the left inverse of H, if $GH = I$, where I is an identity matrix. Thus, we need to prove: $\left(A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1+v^TA{-1}u}\right)(A+uv^T) = I$

$$\begin{aligned}
&\left(A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1+v^TA^{-1}u}\right)(A+uv^T) \\
&= A^{-1}A + A^{-1}uv^T - \frac{A^{-1}uv^TA^{-1}A + A^{-1}uv^TA^{-1}uv^T}{1+v^TA^{-1}u} \\
&= I + A^{-1}uv^T - \frac{A^{-1}uv^T + A^{-1}uv^TA^{-1}uv^T}{1+v^TA^{-1}u} \\
&= I + A^{-1}uv^T - A^{-1}\frac{(1+v^TA^{-1}u)}{1+v^TA^{-1}u}uv^T \\
&= I + A^{-1}uv^T - A^{-1}uv^T \\
&= I \qquad \blacksquare
\end{aligned}$$