COMP SCI and SFWR ENG 4-6TE3, and CES 723
Midterm Exam Solution (Fall 2006)

Oleksandr Romanko

October 27, 2006

1. Consider the function:

$$f(x_1, x_2) \quad = \quad (x_1 - 2x_2)^2 e^{x_1+x_2}$$

(a) (i) Give the gradient and the Hessian of $f(x_1, x_2)$.
(ii) Give the second-order Taylor series expansion of $f(x_1, x_2)$ at the point $x^0 = (1, -1)^T$.

(b) Prove that the function $f(x)$ is not convex.

(c) Make a step with Newton's method from the point $x^0 = (1, -1)^T$ when you minimize the function $f(x)$.

(d) Examine if the point $x = (-2, 1)^T$ is a local/global minimum of $f(x)$ by using its gradient/Hessian information.

(e) Give the directional derivative of the function $f(x)$ at the point $x^0 = (1, -1)^T$ in the direction $s = (2, 1)^T$. Examine if this direction is a descent or an ascent direction.

10p

Solution

a. The gradient and the Hessian of the function $f(x_1, x_2)$ are:

$$\nabla f(x_1, x_2) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2(x_1 - 2x_2)e^{x_1+x_2} + (x_1 - 2x_2)^2 e^{x_1+x_2} \\ \\ -4(x_1 - 2x_2)e^{x_1+x_2} + (x_1 - 2x_2)^2 e^{x_1+x_2} \end{pmatrix} =$$

$$\begin{pmatrix} (2(x_1 - 2x_2) + (x_1 - 2x_2)^2)e^{x_1+x_2} \\ \\ (-4(x_1 - 2x_2) + (x_1 - 2x_2)^2)e^{x_1+x_2} \end{pmatrix}$$

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} =$$

$$\begin{pmatrix} 2 + 4(x_1 - 2x_2) + (x_1 - 2x_2)^2 & -4 - 2(x_1 - 2x_2) + (x_1 - 2x_2)^2 \\ \\ -4 - 2(x_1 - 2x_2) + (x_1 - 2x_2)^2 & 8 - 8(x_1 - 2x_2) + (x_1 - 2x_2)^2 \end{pmatrix} e^{x_1+x_2}$$

The second order Taylor series expansion of the function $f(x_1, x_2)$ at the point x^0 is:

$$f(x) = f(x^0) + \nabla f(x^0)^T (x - x^0) + \frac{1}{2}(x - x^0)^T \nabla^2 f(x^0)(x - x^0).$$

In our case, for $x^0 = (x_1^0, x_2^0)^T = (1, -1)^T$ we get:

$$\begin{aligned} f(x_1, x_2) &= f(x_1^0, x_2^0) + \nabla f(x_1^0, x_2^0)^T \begin{pmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \end{pmatrix} \\ &\quad + \tfrac{1}{2} \begin{pmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \end{pmatrix}^T \nabla^2 f(x_1^0, x_2^0) \begin{pmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \end{pmatrix} \\ &= 9 + \begin{pmatrix} 15 \\ -3 \end{pmatrix}^T \begin{pmatrix} x_1 - 1 \\ x_2 + 1 \end{pmatrix} + \tfrac{1}{2} \begin{pmatrix} x_1 - 1 \\ x_2 + 1 \end{pmatrix}^T \begin{pmatrix} 23 & -1 \\ -1 & -7 \end{pmatrix} \begin{pmatrix} x_1 - 1 \\ x_2 + 1 \end{pmatrix} \\ &= -9x_1 - 10x_2 + \tfrac{23}{2}x_1^2 - x_1x_2 - \tfrac{7}{2}x_2^2. \end{aligned}$$

b. At the point $x^0 = (2, 0)^T$, the principal minor

$$\frac{\partial^2 f}{\partial x_2^2} = (8 - 8(x_1 - 2x_2) + (x_1 - 2x_2)^2)e^{x_1+x_2} = (8 - 8 \cdot 2 + 2^2)e^2 = -4e^2 < 0$$

which means that the Hessian $\nabla^2 f(x_1, x_2)$ of the function $f(x_1, x_2)$ is not positive-semidefinite at that point. So, $f(x_1, x_2)$ is not a convex function.

c. Apply a full Newton step for $f(x_1, x_2) = (x_1 - 2x_2)^2 e^{x_1+x_2}$ with $x^0 = (1, -1)^T$.

$$x^1 = x^0 - (\nabla^2 f(x^0))^{-1} \nabla f(x^0)$$

$$\nabla f(x) = \begin{pmatrix} (2(x_1 - 2x_2) + (x_1 - 2x_2)^2)e^{x_1+x_2} \\ \\ (-4(x_1 - 2x_2) + (x_1 - 2x_2)^2)e^{x_1+x_2} \end{pmatrix}$$

$$\nabla^2 f(x) = \left(\begin{pmatrix} 2 & -4 \\ -4 & 8 \end{pmatrix} + (x_1 - 2x_2) \begin{pmatrix} 4 & -2 \\ -2 & -8 \end{pmatrix} + (x_1 - 2x_2)^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right) e^{x_1+x_2}$$

$$\nabla f(x^0) = \begin{pmatrix} 15 \\ -3 \end{pmatrix} \qquad \nabla f^2(x^0) = \begin{pmatrix} 2 & -4 \\ -4 & 8 \end{pmatrix} + 3 \begin{pmatrix} 4 & -2 \\ -2 & -8 \end{pmatrix} + 9 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 23 & -1 \\ -1 & -7 \end{pmatrix}$$

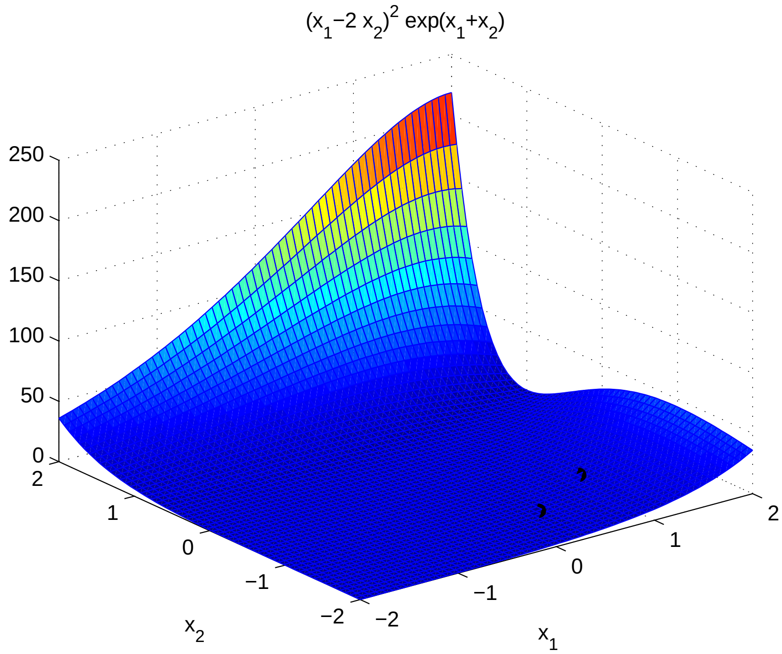
Instead of inverting matrix $\nabla^2 f(x^0)$, which is costly, we can solve the system of equations. Note that if we want to compute $y = A^{-1}b$, we can solve the system of equations $Ay = b$ to find y. In our case, we can solve the system of equations $\nabla^2 f(x^0)y = \nabla f(x^0)$ to get $y = (\nabla^2 f(x^0))^{-1} \nabla f(x^0)$:

$$\begin{pmatrix} 23 & -1 \\ -1 & -7 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 15 \\ -3 \end{pmatrix} \implies y_1 = \frac{2}{3},\ y_2 = \frac{1}{3}.$$

Finally,

$$x^1 = x^0 - (\nabla^2 f(x^0))^{-1} \nabla f(x^0) = \begin{pmatrix} 1 \\ \\ -1 \end{pmatrix} - \begin{pmatrix} \frac{2}{3} \\ \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \\ -\frac{4}{3} \end{pmatrix}$$

$$f(x_1) = 3.3109 < f(x_0) = 9$$



d. The point $x = (-2, 1)^T$ is not a local/global minimum of $f(x)$ because the gradient at that point is not equal to zero:

$$\nabla f(x) = \begin{pmatrix} (2(x_1 - 2x_2) + (x_1 - 2x_2)^2)e^{x_1+x_2} \\ \\ (-4(x_1 - 2x_2) + (x_1 - 2x_2)^2)e^{x_1+x_2} \end{pmatrix} = \begin{pmatrix} 8e^{-1} \\ \\ 32e^{-1} \end{pmatrix}.$$

e. Directional derivative of the function $f(x)$ at the point $(1, -1)^T$ in the direction $s = (2, 1)^T$. The directional derivative is:

$$\delta f(x, s) = \nabla f(x)^T s = \begin{pmatrix} (2(x_1 - 2x_2) + (x_1 - 2x_2)^2)e^{x_1+x_2} \\ \\ (-4(x_1 - 2x_2) + (x_1 - 2x_2)^2)e^{x_1+x_2} \end{pmatrix}^T s = (15, -3) \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 27 > 0$$

This is an ascent direction.

---

2. Give a sequence that has sub-linear converges to -1, and prove its convergence rate.

3p

Solution Sequence that converges to -1 sub-linearly. Consider the sequence $\alpha_k = \frac{1}{k} - 1$. This sequence converges to -1 when $k \to \infty$. Let us take $p = 1$, then $\lim_{k\to\infty} \frac{1/(k+1)}{(1/k)^1} = \lim_{k\to\infty} \frac{k}{k+1} = 1$. In addition, $\lim_{k\to\infty} \frac{1/(k+1)}{(1/k)^{1+\epsilon}} = \lim_{k\to\infty} \frac{kk^\epsilon}{k+1} = \lim_{k\to\infty} k^\epsilon = \infty \; \forall\, \epsilon > 0$. So, the order of convergence is $p^* = 1$. Moreover, $\beta = \lim_{k\to\infty} \frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|} = \lim_{k\to\infty} \frac{1/(k+1)}{1/k} = 1$ which implies sub-linear convergence.

---

3. Consider the function $f(x_1, x_2) = (x_1 - 2x_2)^2 e^{x_1+x_2}$.

(a) Let $x^0 = (2, 0)^T$, $x^1 = (1, -1)^T$, $x^2 = (-1, -1)^T$. Apply one cycle of the Nelder-Mead simplex algorithm to minimize the function when $\alpha = 0.5$, $\beta = 2/3$ and $\gamma = 2$.

(b) Make a line-search step by using the Goldstein-Armijo method from the point $x^0 = (-1, 1)^T$ in the direction $s = (1, -1)^T$, when $\mu_1 = 0.2,\ \ \mu_2 = 0.9$ and $\alpha_0 = 0.8$. Let the multiplication factor be 2, if the step-length has to be increased; 0.5 if it has to be decreased; and when the step-length interval is identified, use bi-section to find the Goldstein-Armijo step-length.

8p

Solution

a. Sort the points in the order of ascending function value

$$f(x^0) = f(-1,-1) = e^{-2} = 0.1353 \leq f(x^1) = f(1,-1) = 9 \leq f(x^2) = f(2,0) = 4e^2 = 29.5562;$$

$\bar{x} = \frac{x^0+x^1}{2} = (0,-1)^T$; $x^r = \bar{x} + (\bar{x} - x^2) = (-2,-2)^T$;
$f(x^r) = 4e^{-4} = 0.0733 < f(x^0)$;
$x^e = \bar{x} + \gamma(\bar{x} - x^2) = (-4,-3)^T$; $f(x^e) = 4e^{-7} = 0.0036 < f(x^r)$;
Drop $x^2 = (2,0)^T$ and add $x^e = (-4,-3)^T$.

b. Goldstein-Armijo line-search step from the point $x^0 = (-1,1)^T$ in the direction $s^0 = (1,-1)^T$ when $\mu_1 = 0.2, \mu_2 = 0.9, \alpha_0 = 0.8$:

$f(x^0) = f(-1,1) = 9,$
$f(x^1) = f(x^0 + \alpha_0 s^0) = f(x^0 + 0.8 s^0) = f(-0.2, 0.2) = 0.36.$

$$\nabla f(x) = \begin{pmatrix} (2(x_1 - 2x_2) + (x_1 - 2x_2)^2)e^{x_1+x_2} \\ \\ (-4(x_1 - 2x_2) + (x_1 - 2x_2)^2)e^{x_1+x_2} \end{pmatrix}.$$

Check if $-\alpha_0\mu_1\nabla f(x^0)^T s^0 \leq f(x^0) - f(x^1) \leq -\alpha_0\mu_2\nabla f(x^0)^T s^0$:
$-\alpha_0\mu_1\nabla f(x^0)^T s^0 = -0.8 \cdot 0.2 \cdot (3,21)(1,-1)^T = 2.88$
$-\alpha_0\mu_2\nabla f(x^0)^T s^0 = -0.8 \cdot 0.9 \cdot (3,21)(1,-1)^T = 12.96$

So, $2.88 \leq f(x^0) - f(x^1) = 8.64 \leq 12.96$. The Goldstein-Armijo conditions are satisfied $\Longrightarrow$ the step-length α_0 should be accepted.

---

4. Which of the following statements is true/false. Give a one-sentence justification of your answer.

(a) Every convex function is monotone non-decreasing.
(b) The minimum of two convex functions is convex.
(c) The cost of a step of the steepest descent algorithm is $O(n)$ arithmetic operations.
(d) The Trust Region algorithms requires the Hessian to be positive definite.
(e) The Hessian of a convex function has only real eigenvalues.
(f) The intersection of convex sets is convex.
(g) The bi-section method (for zero finding in 1-dimension) converge quadratically to find the root of the function x^3.
(h) The function $e^{(x^2)}$ is a convex function.
(i) The relative interior of a nonempty convex set is closed.
(j) The superlinear convergence is faster than linear convergence. **10p**

Solution

a. False. A function f is monotonically non-deacreasing if, whenever $x \leq y$, then $f(x) \leq f(y)$. For example, $f(x) = x^2$ is not monotonically non-deacreasing on $(-\infty, 0)$, so, a convex function is not necessarily monotonically non-deacreasing.

b. False. Epigraph of the minimum of convex functions is a union of the functions epigraphs, but union of convex sets (functions epigraphs) is not necessarily convex.

c. False. The cost of a step of the steepest descent algorithm is $O(n)$ plus line search as at each step we need to compute a search direction $s^k = -\nabla f(x^k)$ which is the gradient of dimension n and perform the line search.

d. False. The Hessian is not required to be positive definite as by adding αI to it we can make the resulting matrix positive definite.

e. True. The Hessian of a function is symmetric and, so, has only real eigenvalues.

f. True. If we take two points inside the intersection, the line connecting them still belongs to the intersection (as the line belongs to each set).

g. False. Bisection method converges linearly with the convergence rate $\beta = \frac{1}{2}$ for every function.

h. True. As exponential function is convex nondecreasing, by using composition rules we can say that a function $e^{g(x)}$ is convex if $g(x)$ is convex. The quadratic function $g(x) = x^2$ is convex and, so, $e^{(x^2)}$ is convex as well.

i. False. The relative interior of a non-empty circle is open.

j. True. Though they have the same order of convergence, the superlinear convergence is faster than all linear as it has lower rate of convergence.

---

5. Consider the function $f(x_1, x_2) = 18x_1 + x_2^2 + (x_1 - 2x_2)^2$. Apply the Fletcher-Reeves conjugate gradient method to find the minimum of the function $f(x_1, x_2)$ starting from the point $x^0 = (1, 0)^T$.

4p

Solution

a. Fletcher-Reeves conjugate gradient algorithm with $x^0 = (1, 0)^T$. We have quadratic function of the form $f(x) = \frac{1}{2}x^T Ax - b^T x$, where

$$A = \begin{pmatrix} 2 & -4 \\ -4 & 10 \end{pmatrix}, \quad b = \begin{pmatrix} -18 \\ 0 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\nabla f(x) = \begin{pmatrix} 18 + 2(x_1 - 2x_2) \\ 2x_2 - 4(x_1 - 2x_2) \end{pmatrix} = \begin{pmatrix} 18 + 2x_1 - 4x_2 \\ -4x_1 + 10x_2 \end{pmatrix}.$$

Iteration 1:

$$\text{Initial point: } x^0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$\nabla f(x^0) = \begin{pmatrix} 20 \\ -4 \end{pmatrix},$$

$$s^0 = -\nabla f(x^0) = \begin{pmatrix} -20 \\ 4 \end{pmatrix}.$$

We use exact line search to find $x^1 = \arg\min(x^0 + \lambda_0 s^0)$:

$$\lambda_0 = -\frac{(s^0)^T \nabla f(x^0)}{(s^0)^T A s^0} = 0.26.$$

So,

$$x^1 = x^0 + \lambda_0 s^0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 0.26 \begin{pmatrix} -20 \\ 4 \end{pmatrix} = \begin{pmatrix} -4.2 \\ 1.04 \end{pmatrix}.$$

$$f(x^1) = -35.08 < f(x^0) = 19.$$

Iteration 2:

$$\nabla f(x^1) = \begin{pmatrix} 5.44 \\ 27.2 \end{pmatrix},$$

$$s^1 = -\nabla f(x^1) + \frac{||\nabla f(x^1)||^2}{||\nabla f(x^0)||^2} s^0 = \begin{pmatrix} -5.44 \\ -27.2 \end{pmatrix} + \frac{769.4336}{416} \begin{pmatrix} -20 \\ 4 \end{pmatrix} = \begin{pmatrix} -42.4320 \\ -19.8016 \end{pmatrix}.$$

We can check that s^0 and s^1 are conjugate, i.e., $(s^0)^T A s^1 = 0$.
Using exact line search to find $x^2 = \arg\min(x^1 + \lambda_1 s^1)$, we get:

$$\lambda_1 = -\frac{(s^1)^T \nabla f(x^1)}{(s^1)^T A s^1} = 0.9615.$$

$$x^2 = x^1 + \lambda_1 s^1 = \begin{pmatrix} -4.2 \\ 1.04 \end{pmatrix} + 0.9615 \begin{pmatrix} -42.4320 \\ -19.8016 \end{pmatrix} = \begin{pmatrix} -45 \\ -18 \end{pmatrix}.$$

We can check that x^2 is the optimal solution as $\nabla f(x^2) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

THE END