COMP SCI and SFWR ENG 4-6TE3 /CES723: MidTerm Solutions

Christopher Anand, Oleksandr Romanko

October 22, 2008

Solutions:

1. Consider the functions

$$\begin{aligned} f(x_1,x_2) &= (2x_1-4x_2)^4 + e^{x_1^2-2x_2} \\ h(x_1,x_2,x_3) &= \min\{(x_1-1)^2,\ (x_1-2)^2+(x_2-1)^2\} + |x_3| \end{aligned}$$

   (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 $f(x)$ is strictly convex.

   (c) Demonstrate that the function $h(x)$ is not convex.

   (d) Examine if the point $(0,0,0)^T$ is a local/global minimum of $h(x)$. **8p**

Solution

(a)

$$\nabla f(x_1,x_2) = \begin{bmatrix} 8\,(2x_1-4x_2)^3 + 2\,x_1 e^{x_1^2-2x_2} \\ -16\,(2x_1-4x_2)^3 - 2\,e^{x_1^2-2x_2} \end{bmatrix}$$

$$H(x_1,x_2) = \begin{bmatrix} 48\,(2x_1-4x_2)^2 + 4\,x_1^2 e^{x_1^2-2x_2} + 2\,e^{x_1^2-2x_2} & -96\,(2x_1-4x_2)^2 - 4\,x_1 e^{x_1^2-2x_2} \\ -96\,(2x_1-4x_2)^2 - 4\,x_1 e^{x_1^2-2x_2} & 192\,(2x_1-4x_2)^2 + 4\,e^{x_1^2-2x_2} \end{bmatrix}$$

Second-order Talor series expansion:

$$\begin{aligned} f(x_1,x_2) &= \left(16+e^{-1}\right) + \left(-64+2e^{-1},\ -128+2e^{-1}\right)\begin{pmatrix} x_1-1 \\ x_2-1 \end{pmatrix} + \\ & (x_1-1,x_2-1)\cdot\begin{bmatrix} 192+6e^{-1} & -384-4e-1 \\ -384-4e-1 & 768+4e^{-1} \end{bmatrix}\begin{pmatrix} x_1-1 \\ x_2-1 \end{pmatrix} \end{aligned}$$

(b) You can prove the convexity by considering the function $f(x)$ as a sum of strictly convex functions. Alternatively, you can show that the Hessian is positive definite. Lets use the first approach. First, as $x_1^2-2x_2$ is a convex function and e^y is a convex non-decreasing function, by the composition rules it follows that $e^{x_1^2-2x_2}$ is convex function. Second, as $2x_1-4x_2$ is linear function and, consequently, convex and concave; y^4 is convex non-decreasing for $y\geq 0$ and convex non-increasing for $y<0$, by the composition rules $(2x_1-4x_2)^4$ is convex function. As the result, sum of two strictly convex functions is strictly convex.

(c) The function $h(x)=\min\{h_1(x),h_2(x)\}$ is not a convex function. To find a counterexample, consider $x_1=(1,0,0)^T$ and $x_2=(2,0,0)^T$. Those two points violate the definition of convexity as $(f(1,0,0)+f(2,0,0))/2 = 0 < f(1.5,0,0) = 0.25$, therefore the function h is not convex.

(d) Since $(x_1-1)^2\geq 0$, $(x_1-2)^2+(x_2-1)^2\geq 0$ and $|x_3|\geq 0$ the minimum of the function $h(x)$ is zero. At the point $(0,0,0)^T$ the value of $h(0,0,0)=1$, so it is not the global minimum of the function. To show that $(0,0,0)^T$ is not a local minimum lets consider the ϵ-neighborhood of the point $(0,0,0)^T$. For the point $(0+\epsilon,0+\epsilon,0+\epsilon)^T$, where $\epsilon>0$, the function value is $h(0+\epsilon,0+\epsilon,0+\epsilon)=(\epsilon-1)^2+|\epsilon| = \epsilon^2-2\epsilon+1+\epsilon = 1+\epsilon^2-\epsilon < 1$. As $h(0,0,0) > h(0+\epsilon,0+\epsilon,0+\epsilon)$, the point $(0,0,0)^T$ is not a local minimum.

2. Consider the bisection line search algorithm.

   (a) Explain how you can compute the order of convergence for the bisection line search algorithm.

   (b) Prove its order and rate of convergence. **4p**

Solution

(a) At each iteration of the bisection line search the distance between the current value x^k and the solution x is divided by two, so $|x^{k+1} - x| = \frac{|x^k - x|}{2}$. So we get the algorithm that converges as the sequence $\alpha_k = (\frac{1}{2})^k$.

(b) The order of convergence of the sequence $\alpha_k = (\frac{1}{2})^k$ is $p^* = 1$. This is the linear convergence with $\beta = 1/2$. To show that, consider the sequence $\alpha_k = (1/2)^k$. Here, $\beta = \frac{0.5^{k+1}}{(0.5^k)^1} = 1/2 \to p = 1$. In addition, $\lim_{k\to\infty} \frac{0.5^{k+1}}{(0.5^k)^{1+\epsilon}} = \frac{0.5}{0.5^{k\epsilon}} = \infty \ \forall \epsilon > 0$, so $p^* = 1$.

3. Consider the function

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

   (a) Let $x^0 = (1, 0)^T$. Apply a full Newton step and give x^1.

   (b) Let $x^0 = (1, 0)^T$. Calculate the Trust-Region search direction with the initial value $\alpha = 4$. Let choose $\mu = 0.2$, $\eta = 0.8$, $\gamma_1 = 0.5$, $\gamma_2 = 2.5$. Would you accept this step in the Trust Region Algorithm or α should be changed? If it needs to be changed, should α be increased or decreased?

   (c) Is it more advantageous to use the Trust Region method or the Newton's method when minimizing the given function $f(x_1, x_2)$. Explain why. **6p**

Solution

(a) Newton step:

$$x^1 = x^0 + s^0$$

$$s^0 = -[\nabla^2 f(1, 0)]^{-1} \nabla f(1, 0)$$

$$\nabla f(1, 0) = (1 + 2/x_1, -1 - e^{x_2})^T = (3, -2)^T$$

$$\nabla^2 f(1, 0) = \begin{bmatrix} -2/x_1^2 & 0 \\ 0 & -e^{x_2} \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}$$

$$s^0 = -[\nabla^2 f(1, 0)]^{-1} \nabla f(1, 0) = (1.5, -2)^T$$

$$x^1 = (1, 0)^T + (1.5, -2)^T = (2.5, -2)^T$$

(b) The trust-region step:

$$H = (\nabla^2 f + \alpha I)$$

$$H(1, 0) = \begin{bmatrix} -2 + 4 & 0 \\ 0 & -1 + 4 \end{bmatrix}$$

$$s^0 = -H(1, 0)^{-1} \nabla f(1, 0) = (-3/2, 2/3)^T$$

$$x^1 = x^0 + s^0 = (-1/2, 2/3)^T$$

To determine if the step is good enough, let's check the ratio

$$\rho_0 = \frac{f(x^0) - f(x^1)}{f(x^0) - q(x^1)}$$

$$\text{actual decrease} = f(x^0) - f(x^1) = f(1,0) - f(-1/2, 2/3) = 0 - (-3.72) = 3.72$$

$$\text{predicted decrease} = f(x^0) - q(x^1) = -\nabla f(x^0)^T(x^1 - x^0) - 1/2(x^1 - x^0)^T \nabla^2 f(x^0)(x^1 - x^0) = 8.32$$

$\rho_0 =$ actual decrease / predicted decrease $= 0.45$ and as $\mu \leq \rho_0 \leq \eta$ the step is good and we accept it. The $\alpha^1 = \alpha^0 = 4$ remains unchanged.

(c) As the objective function is not convex, the Newton's method may not converge while Trust Region is globally convergent.

4. Consider the application of Newton's method to minimize the function $f(x) = (x-1)^4 + (x-1)^2$:

   (a) Give conditions under which Newton's method will always converge, and give stronger conditions under which Newton's method will converge in one step.

   (b) Will Newton's method for the function $f(x)$ above converge starting from the point $x = 4$? Will it converge in one step?

   (c) Give conditions under which Newton's method would fail to converge. Alternatively, set up an example function, and starting point and show that Newton's method when applied to this point will not converge. You can do this using equations, or by using a diagram, labelling the points x_i, and describing the properties of the slopes you need to make your example work.

6p

Solution

(a) Netwon's method will converge for strictly convex functions. Newton's method approximates the function by the quadratic model around the given point and after that takes step toward minimum of that quadratic function. So, it will converge in one step for quadratic functions.

(b) The function above will converge as the function $f(x)$ is strictly convex. It will not converge in one step as $x^1 = x^0 - f'(x^0)/f''(x^0) = 4 - 114/110 = 2.96 \neq 1$ and we know that the minimum of $f(x)$ is reached for $x = 1$.

(c) For non-convex functions Newton's method may fail to converge.

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

   (a) The epigraph of the function $f(x) = |x|$ is not convex as the function $f(x)$ is not differentiable at the point $x = 0$.

   (b) Two orthogonal vectors s^1 and s^2 are also conjugate with respect to the identity matrix I.

   (c) Steepest descent (gradient) algorithm has local quadratic convergence.

   (d) The function $f(x) = \frac{|x|}{x}$ is a convex function.

   (e) Set $\mathcal{C} = \{(x_1, x_2) \,|\, x_1 \geq 0,\ x_2 \geq 0\}$ is a convex pointed cone.

   (f) If any two people in any two positions of a valley can always see each other, the height function of the valley is convex.

   (g) Sum of one convex and one non-convex functions is always non-convex.

   (h) In constrained optimization (minimizie $f(x)$ subject to $g(x) = 0$), the gradient, ∇f, will be zero at the minimum. **8p**

Solution

a. False. The function $f(x) = |x|$ is convex (as $f(x) = -x$ and $f(x) = x$ are convex for $x < 0$ and $x \geq 0$ respectively) and its epigraph is a convex set.

b. True. For orthogonal vectors $(x^1)^T s^2 = 0$ and for the vectors conjugate with respect to I we have $(x^1)^T I s^2 = (x^1)^T s^2 = 0$.

c. False. Steepest descent algorithm converges linearly.

d. False. The function $f(x) = \frac{|x|}{x}$ is discontinuous at $x = 0$ and so it cannot be convex.

e. True. It is convex cone as the set is convex and $\lambda x \in \mathcal{C},\ \lambda \geq 0$; $\mathcal{C}$ is pointed as it does not contain a line.

f. True. For the epigraph of the height function of the valley, the line between any two points is in the set (epigraph), so the height function of the valley is convex.

g. False. The sum of convex function $f_1(x) = 2x^2$ and non-convex function $f_2(x) = -x^2$ is equal to $f(x) = f_1(x) + f_2(x) = x^2$ which is convex.

h. False. For the functions $f(x) = x^3$ and $g(x) = x - 2 = 0$, the gradient at the point of minimum is equal to $\nabla f(x) = 3x^2 = 12$.