Midterm Preparation Handout

CS/SWE 4/6TE3 October 17, 2010

Trust-Region Method

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

(a) Let $x^0 = (1,1)^T$. Calculate the Trust-Region search direction with the initial value $\alpha = 1$. 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?

$$x^1 = x^0 + s^0(\alpha),\ s^0(\alpha) = -(\nabla^2 f(x^0) + \alpha I)^{-1}\nabla f(x^0)$$

$$\alpha = 1,\ \nabla^2 f(x^0) + \alpha I = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & -1 \\ -1 & 4 \end{pmatrix}$$

Similarly to the computation of the Newton method, we can solve the system of equations $-(\nabla^2 f(x^0) + \alpha I)s^0 = \nabla f(x^0)$ to get $s^0(\alpha) = -(\nabla^2 f(x^0) + \alpha I)^{-1}\nabla f(x^0)$:

$$\begin{pmatrix} -4 & 1 \\ 1 & -4 \end{pmatrix}\begin{pmatrix} s_1^0 \\ s_2^0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \Longrightarrow \quad s^0 = \begin{pmatrix} -\frac{1}{3} \\ -\frac{1}{3} \end{pmatrix}$$

So,

$$x^1 = x^0 + s^0(\alpha) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -\frac{1}{3} \\ -\frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \end{pmatrix}$$

$$f(x^1) = f(x^0 + s^0(\alpha)) = f(2/3, 2/3) = 0.5134 < f(x^0) = f(1,1) = 1$$

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

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

$$-(1\ \ 1)\begin{pmatrix} -\frac{1}{3} \\ -\frac{1}{3} \end{pmatrix} - \frac{1}{2}(-1/3 \quad -1/3)\begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}\begin{pmatrix} -\frac{1}{3} \\ -\frac{1}{3} \end{pmatrix} = \frac{4}{9}$$

$$\rho_k = \frac{f(x^k) - f(x^k + s^k(\alpha))}{f(x^k) - q(x^k + s^k(\alpha))} = \frac{\text{actual decrease}}{\text{predicted decrease}},$$

$$\rho_1 = \frac{f(x^0) - f(x^0 + s^0(\alpha))}{f(x^0) - q(x^0 + s^0(\alpha))} = \frac{f(x^0) - f(x^1)}{f(x^0) - q(x^1)} = \frac{1 - 0.5134}{4/9} = 1.0949\,.$$

As $\rho_1 > \eta$ the step is very good and we accept it, $x^1 = x^0 + s^0(\alpha) = (2/3, 2/3)^T$, α is updated: $\alpha = \gamma_1\alpha = 0.5$.

(b) Compare the Steepest Descent, Newtons and the Trust Region methods w.r.t. applicability, computational cost and convergence.

Cost of one iteration:

Steepest Descent: $O(n)$ + line search

Newton Method: $O(n^3)$

Trust-Region Method: $O(n^3)$

Convergence:

Steepest Descent is globally convergent (with exact line search), rate of convergence is linear, but convergence can be very slow (zig-zagging)

Newton Method: local quadratic convergence, not necessarily globally convergent (Hessian is not necessarily positive definite)

Trust Region: globally convergent

Trust region is a compromise between Steepest Descent and Newton.

Solutions to selected problems from the previous Midterm Exams

1. **Spring 2004:** Consider the functions

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

(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 $x = (-10, 5, 5)^T$ is a local/global minimum of $h(x)$ without using its gradient/Hessian information.

Solution

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

$$\nabla f(x_1, x_2) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} -2 + 16x_1^3 + 2x_1 e^{(x_1^2)} - e^{-x_1-2x_2} \\ -2e^{-x_1-2x_2} \end{pmatrix}$$

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

$$= \begin{pmatrix} 48x_1^2 + 2e^{(x_1^2)} + 4x_1^2 e^{(x_1^2)} + e^{-x_1-2x_2} & 2e^{-x_1-2x_2} \\ 2e^{-x_1-2x_2} & 4e^{-x_1-2x_2} \end{pmatrix}$$

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:

$$\begin{aligned} f(x_1, x_2) &= f(x_1^0, x_2^0) + \nabla f(x_1^0, x_2^0)^T \begin{pmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \end{pmatrix} \\ &\quad + \tfrac{1}{2} \begin{pmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \end{pmatrix}^T \nabla^2 f(x_1^0, x_2^0) \begin{pmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \end{pmatrix} \\ &= 9.0862 + \begin{pmatrix} -23.8044 \\ -0.7358 \end{pmatrix}^T \begin{pmatrix} x_1 + 1 \\ x_2 - 1 \end{pmatrix} \\ &\quad + \tfrac{1}{2} \begin{pmatrix} x_1 + 1 \\ x_2 - 1 \end{pmatrix}^T \begin{pmatrix} 64.6776 & 0.7358 \\ 0.7358 & 1.4715 \end{pmatrix} \begin{pmatrix} x_1 + 1 \\ x_2 - 1 \end{pmatrix} \\ &= 18.3563 + 40.1374x_1 - 1.47152x_2 + 32.3388x_1^2 + .735759x_2x_1 + .735759x_2^2 \end{aligned}$$

(b) Consider the Hessian of $f(x_1, x_2)$:

$$\begin{aligned} \nabla^2 f(x_1, x_2) &= \begin{pmatrix} 48x_1^2 + 2e^{(x_1^2)} + 4x_1^2 e^{(x_1^2)} + e^{-x_1-2x_2} & 2e^{-x_1-2x_2} \\ 2e^{-x_1-2x_2} & 4e^{-x_1-2x_2} \end{pmatrix} \\ &= \begin{pmatrix} (48x_1^2 + (2 + 4x_1^2)e^{(x_1^2)})e^{x_1+2x_2} + 1 & 2 \\ 2 & 4 \end{pmatrix} e^{-x_1-2x_2} \end{aligned}$$

Note that $e^{-x_1-2x_2} > 0$, $e^{x_1+2x_2} > 0$ and $(48x_1^2 + (2 + 4x_1^2)e^{(x_1^2)}) > 0$. This implies that $(48x_1^2 + (2 + 4x_1^2)e^{(x_1^2)})e^{x_1+2x_2} + 1 > 1$ and $\det(\nabla^2 f(x_1, x_2)) > 0$. As all leading principal minors of the matrix are positive, the Hessian is positive definite, and, consequently, $f(x)$ is strictly convex.

(c) Note that $\log(z)$ is non-convex increasing function. Let's consider $x^1 = (0, 0, 0)^T$, $x^2 = (2, 0, 0)^T$. So, $0.5h(x^1) + 0.5h(x^2) = 0.5h(0, 0, 0) + 0.5h(2, 0, 0) = 0.5\log(1) + 0.5\log(3) = 0.5493 < h(0.5x^1 + 0.5x^2) = h(0, 0, 1) = \log(2) = 0.6931$, which violates the definition of convexity.

(d) Note that $\log(1+z), z \geq 0 \in \Re^3$ is monotonically increasing non-convex function defined on the interval $[1, +\infty)$. This function reaches its global minimum on the boundary of its domain (in the point $\log(1)$, when $z = 0$). So, the point $x = (-10, 5, 5)^T$, where $z = |x_1+x_2+x_3| = 0$, is a global minimum of the function $h(x)$. Consequently, this point is a local minimum as well.

---

2. **Spring 2004:** Convergence:

(b) Determine the limit point and the rate of convergence of the following sequence:
$x^k = (\frac{1}{k^4}, \frac{4}{k+4})^T, \qquad k = 1, 2, ...$

Solution

(b) Consider the sequences $\frac{1}{k^4}$ and $\frac{4}{k+4}$ separately.
The sequence $\alpha_k = \frac{1}{k^4}$ converges to 0 when $k \to \infty$ as $\lim_{k\to\infty} \frac{1}{k^4} = 0$. Lets take $p = 1$, then $\lim_{k\to\infty} \frac{\frac{1}{(k+1)^4}}{\frac{1}{k^4}} = \lim_{k\to\infty} (\frac{k}{k+1})^4 = 1$. Also, $\lim_{k\to\infty} \frac{\frac{1}{(k+1)^4}}{\frac{1}{k^{4+\epsilon}}} = \lim_{k\to\infty} \frac{k^4 k^\epsilon}{(k+1)^4} = \lim_{k\to\infty} (\frac{k}{k+1})^4 k^\epsilon = \infty \quad \forall \epsilon > 0$. So, the order of convergence is $p^* = 1$. Moreover, $\beta = \lim_{k\to\infty} (\frac{k}{k+1})^4 = 1$, so the sequence converges *sub-linearly.*
Now, consider the sequence $\alpha_k = \frac{4}{k+4}$ that converges to 0 when $k \to \infty$ as $\lim_{k\to\infty} \frac{4}{k+4} = 0$. Lets take $p = 1$, then $\beta = \frac{\frac{4}{(k+1)+4}}{(\frac{4}{k+4})^1} = \lim_{k\to\infty} \frac{k+4}{k+5} = 1$. In addition, $\lim_{k\to\infty} \frac{\frac{4}{k+5}}{(\frac{4}{k+4})^{1+\epsilon}} = \lim_{k\to\infty} \frac{k+4}{k+5}(\frac{k}{4}+1)^\epsilon = \infty \ \forall \epsilon > 0$. So, this sequence converges *sub-linearly.*
As a result, the sequence $x^k = (\frac{1}{k^4}, \frac{4}{k+4})$ converges to $(0, 0)$ *sub-linearly.*

---

3. **Spring 2004:** Consider the function $f(x_1, x_2) = (x_1 + x_2 - 2)^2 - \log(1 + x_1 - x_2)^2$.

(c) Give the directional derivative of the function $f(x)$ at the point $(1, 1)^T$ in the direction $(1, 1)^T$.

Solution

(c) Directional derivative of the function $f(x)$ at the point $(1, 1)^T$ in the direction $s = (1, 1)^T$.

The directional derivative is:

$$\delta f(x, s) = \nabla f(x)^T s = \nabla f(1, 1)^T s = \begin{pmatrix} 2(x_1 + x_2 - 2) - \frac{2}{1+x_1-x_2} \\ 2(x_1 + x_2 - 2) + \frac{2}{1+x_1-x_2} \end{pmatrix}^T s = (-2, 2) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0$$

4. **Spring 2005:** Convex functions:

(a) Prove that if the function $f(x) : R^n \to R$ is convex, then for any set of points x^1, x^2, x^3 the inequality

$$f\left(\sum_{i=1}^{3} \lambda_i x^i\right) \leq \sum_{i=1}^{3} \lambda_i f(x^i),$$

where $\sum_{i=1}^{3} \lambda_i = 1$ and $0 \leq \lambda_i \leq 1, \ \ \forall i$ holds.

(b) Show that $\exp(-3x+4) : R \to R$ is a convex function.

Solution

a. We need to prove that $f(\lambda_1 x^1 + \lambda_2 x^2 + \lambda_3 x^3) \leq \lambda_1 f(x^1) + \lambda_2 f(x^2) + \lambda_3 f(x^3)$ for $\lambda_1 + \lambda_2 + \lambda_3 = 1$ and $0 \leq \lambda_1 \leq 1$, $0 \leq \lambda_2 \leq 1$, $0 \leq \lambda_3 \leq 1$. Let us denote $\mu_2 y = \lambda_2 x^2 + \lambda_3 x^3$, where $\mu_2 = 1 - \lambda_1$. As function f is convex, applying the definition of convex function, we get $f(\lambda_1 x^1 + \mu_2 y) \leq \lambda_1 f(x^1) + \mu_2 f(y)$. In its turn $f(y) = f(\frac{\lambda_2}{\mu_2} x^2 + \frac{\lambda_3}{\mu_2} x^3) \leq \frac{\lambda_2}{\mu_2} f(x^2) + \frac{\lambda_3}{\mu_2} f(x^3)$. Equivalently, $\mu_2 f(y) = \mu_2 f(\frac{\lambda_2}{\mu_2} x^2 + \frac{\lambda_3}{\mu_2} x^3) \leq \lambda_2 f(x^2) + \lambda_3 f(x^3)$ and $\lambda_1 f(x^1) + \mu_2 f(y) \leq \lambda_1 f(x^1) + \lambda_2 f(x^2) + \lambda_3 f(x^3)$. So, we got that $f(\lambda_1 x^1 + \lambda_2 x^2 + \lambda_3 x^3) \leq \lambda_1 f(x^1) + \mu_2 f(y) \leq \lambda_1 f(x^1) + \lambda_2 f(x^2) + \lambda_3 f(x^3)$.

b. 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 linear function $g(x) = -3x + 4$ is convex and, so, e^{-3x+4} is convex as well.

---

5. **Spring 2004:** Which of the following statements is true/false.
Give a one-sentence justification of your answer.

(a) A two times differentiable function is strictly convex if and only if its Hessian is positive definite.

(b) Minimum of convex functions is convex.

(c) The steepest descent algorithm converges quadratically.

(d) An exact line-search is used in trust region algorithms.

(e) The derivative of convex functions is convex as well.

(f) The bisection method converges superlinearly.

(g) A two times differentiable function is convex if and only if its Hessian is symmetric.

(h) If a sequence converges quadratically, then it converges superlinearly too.

Solution

(a) False. While positive definite Hessian implies strictly convex function, the converse is not true: x^4 is strictly convex with its Hessian 0 at $x = 0$.

(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. It converges linearly, only Newton has local quadratic convergence.

(d) False. There is no line search in the trust region algorithms, instead the paramter α is decreased or increased dynamically.

(e) False. For $f(x) = x^4$, its derivative $f'(x) = 4x^3$ is not convex.

(f) False. For the bisection method the sequence is $\alpha_k = (\frac{1}{2})^k$ which converges linearly with the convergence rate $\beta = \frac{1}{2}$.

(g) False. The Hessian of a two times differentiable function is always symmetric and such function is convex if the Hessian is positive semidefinite.

(h) True. Quadratic convergence imply suprelinear convergence as well, if the order of convergence is $p^* = \sup p = 2$, it is implied that for $p < 2$ the limit is finite as well.
