CMO - Tutorial

Sheet 3 Solutions — E0230

For Tutors Only — Not For Distribution

1. *Newton method invariance.*

   Show that Newton's method is invariant under a linear change of coordinates, i.e, given function $f(x)$ and function $\bar{f}(y) = f(Ay)$, where A is a nonsingular matrix, if $x_k = Ay_k$ and x and y are updated using Newton's method, then $x_{k+1} = Ay_{k+1}$.

   Solution: We have $\nabla \bar{f}(y) = A^T \nabla f(x)$ and $\nabla^2 \bar{f}(y) = A^T \nabla^2 f(x) A$, where $x = Ay$. So the Newton step for $\bar{f}$ becomes

$$\begin{aligned} y_{k+1} &= y_k - (A^T \nabla^2 f(x) A)^{-1} (A^T \nabla f(x)) \\ &= y_k - A^{-1} \nabla^2 f(x)^{-1} \nabla f(x) \\ &= y_k + A^{-1}(x_{k+1} - x_k) \\ &= y_k + A^{-1} x_{k+1} - y_k. \end{aligned}$$

2. *Inexact line search*

   Consider an iteration of Newton's method

$$x_{k+1} = x_k - t(\nabla^2 f(x_k))^{-1} \nabla f(x_k),$$

   in which the step size t is chosen by backtracking, i.e, starting at initial value $t = 1$, t is repeatedly updated as $t \leftarrow \beta t$ until it satisfies

$$f(x + tu_k) \le f(x) + \alpha t \nabla f(x_k)^T u_k$$

   where $u_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$ is the descent direction and $\alpha \in (0, \frac{1}{2})$, $\beta \in (0, 1)$ are fixed parameters. Assume $mI \preceq \nabla^2 f(x) \preceq MI$.

   Show that if $||\nabla f(x_k)||_2 \ge \eta$, then $f(x_{k+1}) - f(x_k) \le -\alpha \beta \eta^2 \frac{m}{M^2}$.

   Solution: Let $\lambda(x) = (\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x))^{1/2}$. This means $\lambda(x)^2 = u_k^T \nabla^2 f(x) u_k \ge m||u_k||^2$. Using Taylor expansion, we have

$$\begin{aligned} f(x^{(k+1)}) &\le f(x^{(k)}) + t \nabla f(x^{(k)})^T u_k + \frac{M||u_k||^2}{2} t^2 \\ &\le f(x^{(k)}) - t\lambda(x^{(k)})^2 + \frac{M}{2m} t^2 \lambda(x^{(k)})^2. \end{aligned}$$

   The step size $\hat{t} = \frac{m}{M}$ satisfies the exit condition of the line search, since

$$f(x^{(k)} + \hat{t} u_k) \le f(x^{(k)}) - \frac{m}{2M} \lambda(x^{(k)})^2 \le f(x^{(k)}) - \alpha \hat{t} \lambda(x^{(k)})^2.$$

So, the line search returns a step size $t \geq \frac{\beta m}{M}$, so the change in the objective function becomes

$$\begin{aligned} f(x^{(k+1)}) - f(x^{(k)}) &\leq -\alpha t \lambda(x^{(k)})^2 \\ &\leq -\alpha\beta \frac{m}{M} \lambda(x^{(k)})^2 \\ &\leq -\alpha\beta \frac{m}{M^2} ||\nabla f(x^{(k)})||^2 \\ &\leq -\alpha\beta\eta^2 \frac{m}{M^2}, \end{aligned}$$

where the third inequality is because $\lambda(x)^2 = \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) \leq \frac{1}{M} ||\nabla f(x)||^2$.

3. *Approximate Hessian inverse*

   Consider the problem of minimizing the convex quadratic function $f(x) = \frac{1}{2} x^T Q x - b^T x$, where $Q = I + V$ and V is symmetric with eigenvalues bounded by $e < 1$ in magnitude. For applying Newton's method, consider the approximation of Q^{-1} as $I - V$, so that $x_{k+1} = x_k - \alpha_k (I - V) g_k$ at each time step and α_k is chosen using exact line search.

   What is the rate of convergence of this method?

Solution: Since exact line search is used for minimizing the convex quadratic function, we have

$$\begin{aligned} \frac{E(x^{(k+1)})}{E(x^{(k)})} &= 1 - \frac{(g_k^T u_k)^2}{u_k^T Q u_k g_k^T Q^{-1} g_k} \\ &= 1 - \frac{(g_k^T (I - V) g_k)^2}{u_k^T (I + V) u_k g_k^T Q^{-1} g_k}, \end{aligned}$$

where $g_k = \nabla f(x^{(k)})$ and $u_k = -(I - V) g_k$ is the descent direction. The eigenvalues of $Q = I + V$ lie between $1 - e$ and $1 + e$, those of $I - V$ also lie between $1 - e$ and $1 + e$, and those of Q^{-1} lie between $\frac{1}{1+e}$ and $\frac{1}{1-e}$. So, we have

$$\begin{aligned} g_k^T (I - V) g_k &\geq (1 - e) ||g_k||^2, \\ u_k^T (I + V) u_k &\leq (1 + e) ||u_k||^2 \\ &= (1 + e) ||(I - V) g_k||^2 \\ &\leq (1 + e)^3 ||g_k||^2, \\ g_k^T Q^{-1} g_k &\leq \frac{1}{1 - e} ||g_k||^2. \end{aligned}$$

Therefore,

$$\frac{E(x^{(k+1)})}{E(x^{(k)})} \leq 1 - \frac{(1 - e)^3}{(1 + e)^3}.$$

4. *Descent condition*

   Suppose f is twice continuously differentiable. The convergence of Newton's method requires the Hessian of f to satisfy some conditions and if not, the update equation for x_k has to be modified in some way. Consider the alternative update equation:

   $$x_{k+1} = x_k - \alpha_k(\nabla^2 f(x_k) + \mu_k I)^{-1}\nabla f(x_k)$$

   Determine the range of values of μ_k for which $f(x_{k+1}) < f(x_k)$ for some $\alpha_k > 0$.

   Solution: The update direction should be a descent direction, so we need
   $(-(\nabla^2 f(x_k) + \mu_k I)^{-1}\nabla f(x_k))^T\nabla f(x_k) < 0$, or
   $\nabla f(x_k)^T(\nabla^2 f(x_k) + \mu_k I)^{-1}\nabla f(x_k) > 0$. So, the eigenvalues of $\nabla^2 f(x_k) + \mu_k I$ should be positive, i.e, $\mu_k > -\lambda_1(\nabla^2 f(x_k))$, the minimum eigenvalue of the Hessian.

5. *Conjugate Gram-Schmidt*

   (a) Given linearly independent vectors $v_1, v_2, \ldots, v_k \in \mathbb{R}^d$, construct vectors $d_1, d_2, \ldots d_k \in \mathbb{R}^d$ such that the linear span of $\{v_1, \ldots, v_i\}$ is the same as the linear span of $\{d_1, \ldots, d_i\}$ for each $i \in 1, \ldots, k$ and all d_i are orthogonal to each other. (Hint: Write d_i as a linear combination of v_i and $d_1, \ldots d_{i-1}$ and determine the required coefficients. The resultant method is called Gram-Schmidt orthogonalization.)

   (b) Given linearly independent vectors $v_1, v_2, \ldots, v_k \in \mathbb{R}^d$ and positive definite matrix $Q \in \mathbb{R}^{d\times d}$, construct Q-conjugate vectors $d_1, d_2, \ldots d_k \in \mathbb{R}^d$ such that the linear span of $\{v_1, \ldots, v_i\}$ is the same as the linear span of $\{d_1, \ldots, d_i\}$ for each $i \in 1, \ldots, k$.

   Solution:

   (a)

   $$d_i = v_i + \sum_{j=1}^{i-1} c_{i,j}d_j$$
   $$\implies d_j^T d_i = d_j^T v_i + c_{i,j}d_j^T d_j = 0,$$

   so, $c_{i,j} = -\frac{d_j^T v_i}{d_j^T d_j}$ and

   $$d_i = v_i - \sum_{j=1}^{i-1} \frac{d_j^T v_i}{d_j^T d_j} d_j.$$

(b) Similar to (a),

$$d_i = v_i - \sum_{j=1}^{i-1} \frac{d_j^T Q v_i}{d_j^T Q d_j} d_j.$$

6. Consider a function $f : \Omega\ (\subseteq \mathbb{R}^d) \to \mathbb{R}$. Suppose x^* is a local minimizer of f over $\Omega' \subset \Omega$. Is x^* a local minimizer of f on Ω? If yes, prove it. If no, what are the conditions under which it is true?

   Solution: If x^* is an interior point of Ω', then it is a local minimizer of Ω, else there may be a new descent direction possible in Ω which was not possible in Ω'.

7. Let $c \in \mathbb{R}^d, c \neq 0$ and consider the problem of minimizing the function $f(x) = c^T x$ over a constraint set $\Omega \subset \mathbb{R}$. Show that we cannot have a solution lying in the interior of Ω.

   Solution: For an interior point x to be a local minimum, a necessary condition is $\nabla f(x) = 0$, i.e, $c = 0$, which does not hold true.