CMO - Tutorial

Sheet 3 — E0230

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 $x_{k+1} = Ay_{k+1}$.

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

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

   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 $||\nabla f(x_k)||_2 \geq \eta$, then $f(x_{k+1}) - f(x_k) \leq -\alpha\beta\eta^2 \frac{m}{M^2}$.

3. *Approximate Hessian inverse*

   Consider the problem of minimizing the convex quadratic function $f(x) = \frac{1}{2}x^T Qx - b^T x$, 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?

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 $\alpha_k > 0$.

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 $i \in 1, \ldots, k$.

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?

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.