CO 466/666: Continuous Optimization
Problem Set 2

H. Wolkowicz

Winter 2012.
Handed out: 2012-Jan-21.
Due: 2012-Jan-30, at the start of class.

Contents

1 Line Search Methods

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function. Then f is called *strongly convex* if there exists $M > 0$ such that

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq M\|x - y\|^2, \forall x, y. \tag{1}$$

The Goldstein conditions to ensure sufficient decrease and sufficiently large steplengths (similar to the Wolfe conditions) are

$$f(x_k) + (1 - c)\alpha_k \nabla f(x_k)^T p_k \leq f(x_k + \alpha_k p_k) \leq f(x_k) + c\alpha_k \nabla f(x_k)^T p_k, \tag{2}$$

where $0 < c < \frac{1}{2}$, x_k is the current iterate, p_k is the current search direction, and α_k is the steplength.

1. Show that an equivalent condition to (1) is

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) - \frac{1}{2}mt(1 - t)\|x - y\|^2, \forall t \in [0, 1].$$

2. Suppose that the current point x_k and the search direction p_k are given and α_k is found using an exact line search. Moreover, suppose that f is a strongly convex <u>quadratic</u> function. Show that the conditions in (2) are satisfied.

2 Trust Region Method

Suppose that the trust region method is applied to a quadratic function with a positive definite Hessian. Show that the optimum is obtained after a finite number of iterations.

3 Derivatives/Taylor Series Approximations

Let $h : \mathbb{S}^n \to \mathbb{R}$ be defined as $h(X) = \det(X)$, where $\mathbb{S}^n$ is the vector space of $n \times n$ symmetric matrices, equipped with the trace inner-product $\langle X, Y \rangle := \operatorname{trace} XY$. Let $g : \mathbb{R}_{++} \to \mathbb{R}$ be the log function on the positive real line, $g(x) = \log(x)$. Let $f : \mathbb{S}^n_{++} \to \mathbb{R}$ be the composite function

$$f(X) = (g \circ h)(X) = \log \det(X),$$

where $\mathbb{S}^n_+$ is the cone of symmetric positive semidefinite matrices and $\mathbb{S}^n_{++}$ is the cone of symmetric positive definite matrices, both of size $n \times n$. (In the following show the details in the derivations.)

1. Derive the gradient of h. (Hint: Recall the definition of the adjugate matrix in terms of the minors and cofactors, and compare this to the cofactor expansion for the determinant.)

2. Use Part 1 above and the chain rule to derive the gradient of f. (Hint: Recall Cramer's rule for the inverse using the determinant and the adjugate.) Then, write down the first order Taylor series approximation of f at a given positive definite matrix $\bar{X}$ acting on the perturbation ΔX, i.e., the first order approximation of $f(X + \Delta X)$. Denote this by $f_1(X + \Delta X)$.

3. Do the same as in Part 2 for the second order approximation of $f(X+\Delta X)$. Denote this by $f_2(X + \Delta X)$.
   (Hint: To derive the second derivative, apply the product differentiation rule to both sides of the equation $X^{-1}X = I$.

4. Let $X := \begin{bmatrix} 3 & 1 & 2 \\ 1 & 2 & -1 \\ 2 & -1 & 4 \end{bmatrix}$ and $\Delta X := \begin{bmatrix} 1 & -1 & -2 \\ -1 & 1 & 1 \\ -2 & 1 & 1 \end{bmatrix}$. Note that $X \succ 0$, is positive definite.

   (a) Using MATLAB find the maximum steplength $\bar{\alpha} > 0$ such that $X + \bar{\alpha}\Delta X \succeq 0$.
   (Hint: Let $S = sqrtm(X)$ be the positive definite square root of X, where sqrtm denotes the MATLAB command. Sylvester's Lemma of inertia yields that a congruence maintains positive definiteness, i.e. $X \succ 0 \implies S\backslash X/S \succ 0$, where $S\backslash X/S = S^{-1}XS^{-1}$.)

   (b) Use the MATLAB command *linspace* to create 100 steps for plotting: steps=linspace$(0, \bar{\alpha} - .1)$, where the $-.1$ guarantees that you stay positive definite. On one figure, plot the three functions f, f_1, f_2 at each of the 100 values in steps, e.g., using the evaluations

$$f(X + steps(i) * \Delta X), f_1(X + steps(i) * \Delta X), f_2(X + steps(i) * \Delta X).$$

Observe and comment on the quality of the approximations.

(c) Redo Part 4b above with the smaller interval, steps=linspace$(0, 5)$. Observe and comment on the quality of the two approximations, i.e., the first and second order approximations.