CO 367 Fall 2018: Homework 2

Due: November 12th, 1:30pm

Instructions For every nontrivial step you perform, you must justify why the step is valid and what assumptions it exploits. In other words, you do not need to justify basic algebraic operations (rearranging or distributing terms, multiplying both sides of an equation by a constant, etc.), but you *do* need to explain all steps that exploit hypotheses and assumptions (positive semidefiniteness of a matrix, continuity or convexity of a function, taking a limit that must exist, etc.). If you exploit a result seen in class, or an elementary theorem, clearly state which one.

Question 1 [8 marks] Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1-smooth function over $\mathbb{R}^n$. Let us fix a point $x^k \in \mathbb{R}^n$ and a descent direction $p^k \in \mathbb{R}^n$. We define a function $\psi : \mathbb{R} \to \mathbb{R}$ for the line search problem, i.e. $\psi(\alpha) = f(x^k + \alpha p^k)$.
(i) Prove that if ∇f is Lipschitz continuous with constant L, then $\frac{d}{d\alpha}\psi$ is also Lipschitz continous. Give a Lipschitz constant K for $\frac{d}{d\alpha}\psi$.
(ii) Assuming that K is known, prove that setting $\alpha = \frac{2}{3K}\left(-\frac{d}{d\alpha}\psi(0)\right)$ will satisfy the sufficient decrease condition for line search with $\sigma = \frac{1}{3}$.
(iii) If f is strongly convex, give a constant γ such that the curvature condition is satisfied for all $\alpha > \gamma$. The expression of γ may involve $\psi(0)$ and p^k.

Solution: (i) By Lipschitz continuity of ∇f, we have

$$||\nabla f(y) - \nabla f(x)||_2 \leq L||y - x||_2 \qquad \text{for all } x, y \in \mathbb{R}^n.$$

By the definition of ψ, and denoting $\psi'(\alpha) := \frac{d}{d\alpha}$ for conciseness, we get

$$\begin{aligned}
|\psi'(\beta) - \psi'(\alpha)| &= |\nabla f(x + \beta p)^T p - \nabla f(x + \alpha p)^T p| \\
&= |(\nabla f(x + \beta p) - \nabla f(x + \alpha p))^T p| \\
&\leq ||(\nabla f(x + \beta p) - \nabla f(x + \alpha p))|| \cdot ||p|| \\
&\leq L||(\beta - \alpha)p|| \cdot ||p|| \\
&= L|\beta - \alpha|||p|| \cdot ||p|| = L|\beta - \alpha|||p||^2 = K|\beta - \alpha|,
\end{aligned}$$

where, successively, we used the definition of directional derivatives, Cauchy-Schwartz inequality, Lipschitz continuity of ∇f, and let $K := L||p||^2$.

(ii) We want to show

$$\psi(\alpha) \leq \psi(0) + \frac{1}{3}\alpha\psi'(0).$$

By the mean value theorem, there exists $0 \leq \beta \leq \alpha$ such that

$$\psi(\alpha) = \psi(0) + \alpha\psi'(\beta).$$

On the other hand, by the Lipschitz continuity of ψ', we have

$$\begin{aligned} |\psi'(\beta) - \psi'(0)| &\leq K|\beta - 0| \\ \psi'(\beta) - \psi'(0) &\leq K\beta \\ \psi'(\beta) &\leq K\beta + \psi'(0). \end{aligned}$$

Using both, we get

$$\begin{aligned} \psi(\alpha) &\leq \psi(0) + \alpha(K\beta + \psi'(0)) \\ &\leq \psi(0) + \alpha(K\alpha + \psi'(0)) \\ &= \psi(0) + \alpha(K(-\frac{2}{3K}\psi'(0)) + \psi'(0)) \\ &= \psi(0) + \alpha\psi'(0)(1 - \frac{2K}{3K}) \\ &= \psi(0) + \alpha\frac{1}{3}\psi'(0). \end{aligned}$$

(iii) We want

$$\psi(2\alpha) > \psi(0) + \frac{1}{3}2\alpha\psi'(0).$$

By the definition of strong convexity, using the points x and $x + \beta p$ there exists $\ell > 0$ such that

$$\begin{aligned} \left(\nabla f(x + \beta p) - \nabla f(x)\right)^T \beta p &\geq \ell||\beta p||^2 \\ \beta\left(\nabla f(x + \beta p) - \nabla f(x)\right)^T p &\geq \ell\beta^2||p||^2 \\ \nabla f(x + \beta p)^T p - \nabla f(x)^T p &\geq \ell\beta||p||^2 \\ \psi'(\beta) - \psi'(0) &\geq \ell\beta||p||^2. \end{aligned}$$

Now let us write

$$\begin{aligned} \psi(2\alpha) &= \psi(0) + \int_0^{2\alpha} \psi'(\beta)d\beta \\ &\geq \psi(0) + \int_0^{2\alpha} \psi'(0) + \ell\beta||p||^2 d\beta \\ &= \psi(0) + 2\alpha\psi'(0) + \frac{(2\alpha)^2}{2}\ell||p||^2 \\ &= \psi(0) + 2\alpha\psi'(0) + 2\alpha^2\ell||p||^2. \end{aligned}$$

In order to satisfy the curvature condition, we need

$$\begin{aligned} 2\alpha\psi'(0) + 2\alpha^2\ell||p||^2 &> \frac{1}{3}2\alpha\psi'(0) \\ \psi'(0) + \alpha\ell||p||^2 &> \frac{1}{3}\psi'(0) \\ \alpha\ell||p||^2 &> -\frac{2}{3}\psi'(0). \end{aligned}$$

Thus we get that the curvature condition is satisfied (at least) for all α strictly larger than $\gamma = \frac{2}{3\ell||p||^2}(-\psi'(0))$.

Question 2 [4 marks] In class, we did not use line search for Newton's method: we simply set $x^{k+1} = x^k + p^k$, where $p^k \neq 0$ is given by

$$p^k = -(\nabla^2 f(x^k))^{-1} \nabla f(x^k).$$

Instead, consider the possibility of doing line search along the direction p^k, i.e. we set $x^{k+1} = x^k + \alpha p^k$, for some $\alpha > 0$. We consider the special case where f is a strictly convex quadratic function.
(i) Prove that $\psi(\alpha) = f(x^k + \alpha p^k)$ is also a strictly convex quadratic function.
(ii) Briefly explain why the unique global minimizer of ψ is $\alpha^* = 1$.
(iii) Prove that setting $\alpha = 1$ satisfies the sufficient decrease condition and the curvature condition for $\sigma \leq \frac{1}{2}$.

Solution: (i) Since f is a strictly convex quadratic function, we know that it takes the form $f(x) = x^T Ax + b^T x + c$ for some $A \in \mathbb{R}^{n\times n}$ positive definite, and $b \in \mathbb{R}^n$. Letting $x = x^k$ and $p = p^k$ for conciseness, we get

$$\begin{aligned}
\psi(\alpha) &= f(x + \alpha p) \\
&= (x + \alpha p)^T A(x + \alpha p) + b^T(x + \alpha p) + c \\
&= x^T Ax + \alpha^2 p^T Ap + 2\alpha x^T Ap + b^T x + \alpha b^T p + c \\
&= (p^T Ap)\alpha^2 + (b^T p + 2x^T Ap)\alpha + (x^T Ax + b^T x + c) \\
&= q\alpha^2 + r\alpha + s,
\end{aligned}$$

where $q = p^T Ap$, $r = b^T p + 2x^T Ap$ and $s = x^T Ax + b^T x + c$. The second derivative of ψ is a constant function $\psi''(\alpha) = q$. We have $q = p^T Ap > 0$ since A is positive definite and $p \neq 0$. The Hessian of ψ is the constant 1×1 matrix $\nabla^2\psi(\alpha) = [q]$ which is positive definite, so ψ is strictly convex.

(ii) By construction with Newton's step, $x^{k+1} = x^k + p^k$ is a minimizer for a quadratic approximation of f at x^k. Since f is quadratic, the approximation is f itself and since f is strictly convex, x^{k+1} is its unique global minimizer. Since $\psi(1) = f(x^{k+1})$ and $\psi(\alpha) = f(y)$ for some $y \neq x^{k+1}$ when $\alpha \neq 1$, we know that $\psi(1) < \psi(\alpha)$ for all $\alpha \neq 1$.

(iii) Recall that

$$\begin{aligned}
\psi(\alpha) &= q\alpha^2 + r\alpha + s \\
\psi'(\alpha) &= 2q\alpha + r.
\end{aligned}$$

and that $\alpha^* = 1$ is the global minimizer, thus a stationary point, and $\psi'(1) = 0$ yields $r = -2q$. We get

$$\begin{aligned}
\psi(\alpha) &= q\alpha^2 - 2q\alpha + s \\
\psi'(\alpha) &= 2q\alpha - 2q,
\end{aligned}$$

for some $q = p^T A p > 0$. We can now verify the sufficient decrease condition:

$$\begin{aligned} \psi(\alpha^*) &= \psi(1) \;=\; q - 2q + s \;=\; s - q \\ &= s + \frac{1}{2}1(-2q) \\ &\le \psi(0) + \sigma\alpha^*\psi'(0) \end{aligned}$$

and the curvature condition:

$$\begin{aligned} \psi(2\alpha^*) &= \psi(2) \;=\; 4q - 4q + s \;=\; s \\ &> s - 4\sigma q \;=\; s + \sigma 2(-2q) \;=\; \psi(0) + \sigma 2\alpha^*\psi'(0). \end{aligned}$$

Question 3 [3 marks] Assume that we can solve the trust region subproblem. In other words, we are given a function $s(A, b) = \operatorname{argmin}\{x^T Ax + b^T x \;:\; ||x||_2 \le 1\}$ that is defined for any $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$. Consider the problem

$$\min\{x^T Cx + d^T x \;:\; (x - z)^T R(x - z) \le 1\}, \tag{E}$$

where we are given constants $C \in \mathbb{R}^{n\times n}$, $d \in \mathbb{R}^n, z \in \mathbb{R}^n$, $R \in \mathbb{R}^{n\times n}$, and where R is symmetric and positive definite. Give an expression for an optimal solution x^* to (E). This expression may use the function s.

Solution: We need to express $(x - z)^T R(x - z)$ as a Euclidean norm constraint. Since R is symmetric and positive semidefinite, we know that it can be expressed as $R = GG^T$. As we have seen in class, this construction is obtained by letting $R = QDQ^T$ where $Q^{-1} = Q^T$ and $D_{ii} > 0$ for all i, then building $G = QD^{1/2}$. Since R is positive definite, $D^{1/2}$ is invertible, and so is G.

If we let $y := G^T(x - z)$, we get that

$$\begin{aligned} (x - z)^T R(x - z) &= (x - z)^T GG^T(x - z) \\ &= y^T y \;=\; ||y||_2^2. \end{aligned}$$

Therefore, $(x - z)^T R(x - z) \le 1$ is equivalent to $||y||_2^2 \le 1$ which is in turn equivalent to $||y||_2 \le 1$, as desired. We now have to reformulate the objective function in terms of y. We have $x = G^{-T}y + z$ and

$$\begin{aligned} x^T Cx + d^T x &= (G^{-T}y + z)^T C(G^{-T}y + z) + d^T(G^{-T}y + z) \\ &= y^T G^{-1}CG^{-T}y + 2z^T CG^{-T}y + z^T Cz + d^T G^{-T}y + d^T z \\ &= y^T(G^{-1}CG^{-T})y + (2z^T CG^{-T} + d^T G^{-T})y + z^T Cz + d^T z, \end{aligned}$$

thus an optimal solution to (E) is can be found using

$$x^* = s(G^{-1}CG^{-T}, 2z^T CG^{-T} + d^T G^{-T}).$$