OR 631: Mathematical Programming II. Spring 2014.
Homework Set 4. Due: Tuesday April 29.

1. Suppose we wish to solve the problem

$$\min f(x), \qquad h_i(x) \le 0,\; i = 1, \ldots, m, \quad x \in B(0, R),$$

and suppose that f is convex and has range at most 1 on $B(0, R)$, that each h_i is convex and is at most 1 on $B(0, R)$, and that the problem has an optimal solution x_*.

Consider the following algorithm, given $\epsilon > 0$. Start with $x_0 = 0$. At iteration k, if $\|x_k\| > R$, choose $v_k = x_k$. If for some i, $h_i(x_k) > \epsilon$, choose any such i and set v_k to be a subgradient of h_i at x_k. If $h_i(x_k) \le \epsilon$ for all i, set v_k to be a subgradient of f at x_k. If $v_k = 0$, stop; otherwise set $x_{k+1} = x_k - \frac{\epsilon R}{\|v_k\|} v_k$.

Show that within ϵ^{-2} iterations, the method will give an x_k with $h_i(x_k) \le \epsilon$ for all i and $f(x_k) \le f(x_*) + \epsilon$.

The proof follows that of Theorem 1 in Lecture 21. We consider the small ball $B(\bar{x}, \epsilon R)$, which is also $x_* + \epsilon B(0, R)$. Here $\bar{x} := (1 - \epsilon)x_*$. All points in this ball are in $B(0, R)$, have objective value within ϵ of that of x_*, and have all h_i values at most ϵ, by convexity of the h_i's and the fact that each h_i is at most 1 on $B(0, R)$. So for all cases of v_k, we find that

$$v_k^T(x_k - (\bar{x} + \frac{\epsilon R v_k}{\|v_k\|})) \ge 0,$$

as long as $x_k \notin B(0, R)$, or $h_i(x_k) > \epsilon$ for some i, or $f(x_k) > f(x_*) + \epsilon$. The proof then continues exactly as before, and we conclude that a suitable solution is found within ϵ^{-2} iterations.

2. Assume that f is a twice continuously differentiable convex function on $\Re^n$ and that for every x, the eigenvalues of $\nabla^2 f(x)$ are bounded between $\ell > 0$ and $L < \infty$. Show that, for every $x, y \in \Re^n$,

$$f(x) + \nabla f(x)^T(y - x) + \frac{\ell}{2}\|y - x\|^2 \le f(y) \le f(x) + \nabla f(x)^T(y - x) + \frac{L}{2}\|y - x\|^2.$$

Let $\phi(\lambda) := f(x + \lambda(y - x))$ so that $\phi'(\lambda) = \nabla f(x + \lambda(y - x))^T(y - x)$ and $\phi''(\lambda) = (y - x)^T \nabla^2 f(x + \lambda(y - x))(y - x)$. Thus $l\|y - x\|^2 \le \phi''(\lambda) \le L\|y - x\|^2$.

Now

$$\begin{aligned}
\phi(1) &= \phi(0) + \int_0^1 \phi'(\lambda) d\lambda \\
&= \phi(0) + \phi'(0) + \int_0^1 (\phi'(\lambda) - \phi'(0)) d\lambda \\
&= \phi(0) + \phi'(0) + \int_0^1 (\int_0^\lambda \phi''(\mu) d\mu) d\lambda
\end{aligned}$$

Inserting the given bounds on ϕ'' gives the desired result.

[Incidentally, if we are dealing with this function class, better complexity bounds are possible: basically, we can assure linear convergence, so the bound is $O(\ln(1/\epsilon))$, not $O(\sqrt{1/\epsilon})$.]

3. a) Suppose that g is a convex function on $\Re^n$. Show that, for every $z \in \Re^n$ and $L > 0$, the problem

$$(P(z)) \qquad \min g(x) + \frac{L}{2}\|x - z\|^2$$

has an optimal solution.

b) Assume that, for any $z \in \Re^n$, you can solve $(P(z))$ efficiently. Show how you can solve the problem

$$\min f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{L}{2}\|x - x_k\|^2 + g(x)$$

efficiently. What is the solution if $g \equiv 0$?

c) Find the solution to $(P(z))$ explicitly if $g(x) := \|x\|_1$.

a) We are assuming that g is a real-valued function on $\Re^n$, so that the interior of its domain is $\Re^n$. Thus it is continuous, and has a subgradient at every point. Let h be a subgradient at z. Then for all $x \in \Re^n$, $g(x)+(L/2)\|x-z\|^2 \geq g(z)+h^T(x-z)+(L/2)\|x-z\|^2$. This is greater than $g(z)$ for $\|x - z\| > (2/L)\|h\|$. Thus minimizing this function over $\Re^n$ is equivalent to minimizing it over the ball centered at z with radius $(2/L)\|h\|$. Since this set is compact and the function is continuous, it has a minimizer on this set, and hence on $\Re^n$.

b) Note that $(L/2)\|x - (x_k - (1/L)\nabla f(x_k))\|^2 = (L/2)\|x - x_k\|^2 + \nabla f(x_k)^T(x - x_k) + \|\nabla f(x_k)\|^2/(2L)$. Hence the given function is a constant plus $g(x) + (L/2)\|x - z\|^2$ for $z := x_k - (1/L)\nabla f(x_k)$, so minimizing it is equivalent to $(P(z))$ for this z. If $g \equiv 0$, then $(P(z))$ is solved by $x = z$, so the solution then is $x_k - (1/L)\nabla f(x_k)$.

c) If $g(x) = \|x\|_1$, then $g(x) + (L/2)\|x - z\|^2 = \sum_j[|x^{(j)}| + (L/2)(x^{(j)} - z^{(j)})^2]$, so we can minimize it by minimizing each term. But each term is a strictly convex function, so it has a unique minimizer at a point where 0 is a subgradient. By considering the three cases for the sign of $x^{(j)}$, we find the optimal $x^{(j)}$ is $z^{(j)} - 1/L$ if this is positive, $z^{(j)} + 1/L$ if this is negative, and 0 otherwise. (This can also be written $\text{mid}(z^{(j)} - 1/L, 0, z^{(j)} + 1/L)$, where mid gives the median of its three arguments, a very useful function.)