Nonlinear Optimization	**Homework 4 (Solutions)**
MATH 408 Spring 2019	

Exercises: 4.3, 4.7, 5.2, 7.1, 7.3, 7.4, 7.29, 8.1 in "Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB"

Problem 4.3 (5 pts)
The three methods (implemented as written out in the text) should take 3301, 3732, 1271 iterations, with exact line search being the fastest method.

Problem 4.7 (4 pts) Let $f(x) = x^T Ax + 2b^T x + c$ where $A \in \mathbb{S}^n$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$. Show that the smallest Lipschitz constant of ∇f is $2\|A\|$
Recall the definition of the operator norm (also called the 2-norm) of a matrix:

$$\|A\| := \max_{z \neq 0} \frac{\|Az\|}{\|z\|}$$

As shown in the chapter, for all x, y

$$\|\nabla f(x) - \nabla f(y)\| = 2\|Ax - Ay\| = 2\|A(x - y)\| \leq 2\|A\|\|x - y\|$$

so $2\|A\|$ is a global Lipschitz constant for f. To show it is the smallest possible one, it suffices to find x, y such that $\|\nabla f(x) - \nabla f(y)\| \geq 2\|A\|\|x - y\|$. For a symmetric matrix A, it can be shown that $\|A\| = \max_{i=1,\ldots,n} |\lambda_i(A)|$. Suppose $|\lambda_1(A)| = \max i = 1, \ldots, n|\lambda_i(A)|$. Taking x to be an eigenvector of A corresponding to $\lambda_1(A)$ and $y = 0$, we have

$$\|\nabla f(x) - \nabla f(0)\| = 2\|A(x - 0)\| = 2|\lambda_1(A)|\|x\| = 2\|A\|\|x\|$$

proving the result.

Problem 5.2 (8 pts)
Should find stationary pts for part a to be $(5, 4)$, $(\frac{53+4\sqrt{22}}{3}, \frac{2+\sqrt{22}}{3})$, $(\frac{53-4\sqrt{22}}{3}, \frac{2-\sqrt{22}}{3})$ which are global min, saddle pt, and local min, respectively. The number of iterations for gradient method with backtracking should be ≈2000 iterations to converge for each initial vector compared with 8, 8, 16, 13 for the hybrid Newton method for the 4 initial vectors respectively. For the last initial vector, you should find that the gradient method convegres to the global minimizer while the Newton hybrid converges to the local minimizer found in a. You should notice that Newton's method greatly speeds up convergence, in particular when the initial pt is close to a stationary point.

Problem 7.1 (2pts each) Determine if the following are convex.

i $C_1 = \{x \in \mathbb{R}^n : \|x\|^2 = 1\}$

In $\mathbb{R}^2$, this is the unit circle, which is certainly not convex.

ii $C_2 = \{x \in \mathbb{R}^n : \max_{i=1,\ldots,n} x_i \leq 1\}$

Observe that the function $f(x_1, \ldots, x_n) = \max_{i=1,\ldots,n} x_i$ is convex by Theorem 7.25 and C_2 is the 1 sublevel set of f, hence is convex.

iii $C_3 = \{x \in \mathbb{R}^n : \min_{i=1,\ldots,n} x_i \leq 1\}$

In $\mathbb{R}^2$, $x = (2,1), y = (1,2) \in C_3$, but $x/2 + y/2 = (3/2, 3/2) \notin C_3$ so C_3 is not convex.

iv $C_4 = \{x \in \mathbb{R}^n_{++} : \prod x_i \geq 1\}$

Observe that $x \in C_4$ iff

$$\sum \log(x_i) = \log(\prod x_i) \geq \log(1) = 0$$

or equivalently if $-\sum \log(x_i) \leq 0$. Since $-\log(x_i)$ is convex for each i, the function $f(x) = -\sum \log(x_i)$ is convex. Then C_4 is convex as it is the 0 sublevel set of f.

Problem 7.3 (4 pts) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex as well as concave function. Show that f is an affine function, that is, there exist $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that $f(x) = a^T x + b$.

Since f is concave and convex, it follows immediately that $f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda) f(y)$ for all $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$. It is straightforward (though a bit tedious) to use this to check that the function $g(x) = f(x) - f(0)$ is linear, hence $g(x) = a^T x$, so letting $f(0) = b$, $f(x) = a^T x + b$ is affine

---

Problem 7.4 (5 pts) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable convex function. Show that for any $\varepsilon > 0$, the function

$$g_\varepsilon(x) = f(x) + \varepsilon \|x\|^2$$

is coercive.

Many people copied a overly complicated answer from Stack Exchange almost word for word. A simplified argument in the same vein is below.

By convexity of f

$$f(0) + \nabla f(0)^T x \leq f(x)$$

Then by Cauchy Schwarz, observe that $-\|\nabla f(0)\| \|x\| \leq \nabla f(0)^T x$, so

$$\begin{aligned} f(0) - \|\nabla f(0)\| \|x\| + \varepsilon \|x\|^2 &\leq f(0) + \nabla f(0)^T x + \varepsilon \|x\|^2 \\ &\leq f(x) + \varepsilon \|x\|^2 = g(x) \end{aligned}$$

The quadratic term in $\|x\|$ in $f(0) - \|\nabla f(0)\| \|x\| + \varepsilon \|x\|^2$ dominates the linear term, hence $f(0) - \|\nabla f(0)\| \|x\| + \varepsilon \|x\|^2 \to \infty$ whenever $\|x\| \nearrow \infty$, $g \to \infty$ whenever $\|x\| \nearrow \infty$, so g is coercive

Problem 7.29 (5 pts) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be an extended real-valued function. Show that f is convex iff $\text{epi}(f)$ is convex.

Suppose f is convex. Then for any $(x, s), (y, t) \in \text{epi}(f)$ and $\lambda \in [0, 1]$,

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) \le \lambda s + (1-\lambda)t$$

so $\lambda(x, s) + (1-\lambda)(y, t) = (\lambda x + (1-\lambda)y, \lambda s + (1-\lambda)t) \in \text{epi}(f)$, meaning $\text{epi}(f)$ is convex.

Suppose $\text{epi}(f)$ is convex. Let $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$. If either $f(x) = \infty$ or $f(y) = \infty$, then it is obvious that $f(\lambda x+(1-\lambda)y) \le \lambda f(x)+(1-\lambda)f(y) = \infty$, so suppose $x, y \in \text{dom}(f)$. Then letting $s = f(x)$ and $t = f(y)$, it is clear that $(x, s), (y, t) \in \text{epi}(f)$, so by convexity of $\text{epi}(f)$,

$$f(\lambda x + (1-\lambda)y) \le \lambda s + (1-\lambda)t = \lambda f(x) + (1-\lambda)f(y)$$

so f is convex

Problem 8.1 (6 pts) Consider the problem

$$\begin{array}{lll} \text{(P)} \qquad & \min & f(x) \\ & s.t. & g(x) \le 0 \\ & & x \in X \end{array}$$

where f, g are convex an $X \subseteq \mathbb{R}^n$ is convex. Suppose x^* is an optimal solution of (P) that satisfies $g(x^*) < 0$. Show that x^* is also an optimal solution of the problem

$$\begin{array}{ll} \min & f(x) \\ s.t. & x \in X \end{array}$$

Suppose for sake of contradiction there exists $y \in X \cap \{x : g(x) > 0\}$ such that $f(y) < f(x^*)$. The line segment $[x^*, y]$ lies in X because X is convex. Furthermore, by continuity of g, the intermediate value theorem implies there exists a $z \in [x^*, y]$ such that $g(z) = 0$, i.e., z is feasible for the problem (P) and there exists some $\lambda \in [0, 1]$ such that $z = x^* + \lambda(y - x^*)$. Observe that by convexity of f:

$$f(z) = f(x^* + \lambda(y - x^*)) \le f(x^*) + \lambda \underbrace{(f(y) - f(x^*))}_{<0 \, by \, assumption} < f(x^*)$$

which contradicts x^* being optimal for (P).