C&O367: Nonlinear Optimization
(Winter 2013)
Assignment 3
H. Wolkowicz

Posted Sat, Feb. 2

Due: Tuesday, Feb. 12 10:00AM (before class),

1 Convex Sets and Functions

1.1 Operations that Preserve Convexity

1. Suppose that $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ is an affine mapping[1] and that $D \subseteq \mathbb{R}^n$ is a convex set. Show that $\mathcal{A}(D)$ is a convex set in $\mathbb{R}^m$. (I.e., affine mappings preserve convexity.)

   Solution: Assume that $x, y \in \mathcal{A}(D)$. This means there exists $\bar{x}, \bar{y} \in \mathbb{R}^n$ such that $x = A\bar{x} + b$ and $y = A\bar{y} + b$. Then for any $\lambda \in [0, 1]$, we have

   $$\lambda x + (1-\lambda)y = A(\lambda\bar{x} + (1-\lambda)\bar{y}) + \lambda b + (1-\lambda)b = A(\lambda\bar{x} + (1-\lambda\bar{y})) + b$$

   so $\lambda x + (1-\lambda)y = \mathcal{A}(\lambda\bar{x} + (1-\lambda\bar{y}))$; $\lambda x + (1-\lambda)y \in \mathcal{A}(D)$ as we wanted.

2. Suppose that for each $i = 1, \ldots, k$, the functions $f_i : D \to \mathbb{R}$ are convex on a convex set $D \subseteq \mathbb{R}^n$, and suppose that $w_i > 0$ are positive weights. Show that $f(x) := \sum_{i=1}^k w_i f_i(x)$ is convex on D.

   Solution: For any $x, y \in D$ and any $\lambda \in [0, 1]$ we have

   $$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \sum_{i=1}^k w_i f_i(\lambda x + (1-\lambda)y) \\ &\leq \sum_{i=1}^k w_i \lambda f_i(x) + w_i(1-\lambda)f_i(y) \\ &= \lambda \sum_{i=1}^k w_i f_i(x) + (1-\lambda)\sum_{i=1}^k w_i f_i(y) \\ &= \lambda f(x) + (1-\lambda)f(y), \end{aligned}$$

   where in the inequality I used the convexity of f_i's and $w_i > 0$.

[1] $\mathcal{A}(x) = Ax + b$ for appropriate matrix A and vector b.

3. Let f_1, f_2 be convex functions defined on a convex set $D \subseteq \mathbb{R}^n$. Show that the pointwise maximum function

$$f(x) := \max\{f_1(x), f_2(x)\}$$

is a convex function on D. (HINT: Use the epigraph characterization of convex functions.) Extend this result to the supremum of an infinite number (a set) of functions.

Solution: The epigraph of $f(x)$, say epi (f), is all the points (y, x) that we have $y \geq f(x)$. For $f(x) = \max\{f_1(x), f_2(x)\}$, $y \geq f(x)$ if and only if $y \geq f_1(x)$ and $y \geq f_2(x)$. This means that $(y, x) \in$ epi (f) if and only if $(y, x) \in$ epi (f_1) and $(y, x) \in$ epi (f_2), so we have:

$$\text{epi}\,(f) = \text{epi}\,(f_1) \cap \text{epi}\,(f_2).$$

$f_1(x)$ and $f_2(x)$ are convex functions, so epi (f_1) and epi (f_2) are convex sets. By question 1.1.1 of assignment 2, the intersection of two convex sets is convex. Hence epi (f) is a convex set and $f(x)$ is a convex function. The intersection of any collection of convex sets is convex. So if $\mathcal{F}$ is a set of convex functions (can be infinite), then

$$g(x) := \sup_{f \in \mathcal{F}}\{f(x)\}$$

is also a convex function.

4. Let $\emptyset \neq C \subseteq \mathbb{R}^n$. Let $\|\cdot\|$ denote a norm on $\mathbb{R}^n$. Show that the distance to the farthest point of C

$$f(x) := \sup_{y \in C} \|x - y\|$$

is a convex function.

Solution 1: Let $x, z \in \mathbb{R}^n$ and $\lambda \in [0, 1]$. We have:

$$\begin{aligned}
f(\lambda x + (1-\lambda)z) &= \sup_{y \in C} \|\lambda x + (1-\lambda)z - y\| = \sup_{y \in C} \|\lambda(x - y) + (1-\lambda)(z - y)\| \\
&\leq^{(a)} \sup_{y \in C} \lambda\|(x - y)\| + (1-\lambda)\|(z - y)\| \\
&\leq^{(b)} \lambda \sup_{y \in C} \|(x - y)\| + (1-\lambda) \sup_{y \in C} \|(z - y)\| \\
&= \lambda f(x) + (1-\lambda) f(y),
\end{aligned}$$

Inequality (a) is by using the triangle inequality and $\lambda, (1-\lambda) \geq 0$. For inequality (b), this is true for any two functions g and h;

$$\sup_{x \in C}(g(x) + h(x)) \leq \sup_{x \in C} g(x) + \sup_{x \in C} h(x).$$

To prove this, let $a := \sup_{x \in C}(g(x) + h(x))$. Then there is a sequence $\{x_n\}$ in C such that $g(x_n) + h(x_n) \to a$. We have $g(x_n) \leq \sup_{x \in C} g(x)$ and $h(x_n) \leq \sup_{x \in C} h(x)$ for all n. Hence, $a = \lim g(x_n) + h(x_n) \leq \sup_{x \in C} g(x) + \sup_{x \in C} h(x)$ as we want.

Solution 2: By using triangle inequality, the function $f_y(x) = \|x - y\|$ is a convex function. Hence, by Question 1.1.3 above, $f(x) = \sup_{y \in C} f_y(x) = \sup_{y \in C} \|x - y\|$ is a convex function.

2 Differentiability

2.1 Gradient

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and strictly convex. Show that $\nabla f(x) = \nabla f(y)$ if, and only if, $x = y$.

Solution: It is clear that if $x = y$ then $\nabla f(x) = \nabla f(y)$. For that other side, f is a strictly convex function, so for any two points $x \neq y$ we have:

$$f(y) > f(x) + \nabla f(x)^T (y - x)$$
$$f(x) > f(y) + \nabla f(y)^T (x - y)$$

By adding these two inequality and subtracting the common terms from both sides, we have $0 > \nabla(f(x) - f(y))^T (y - x)$. This means that $f(x) - f(y) \neq 0$ if $x \neq y$ and we are done.

2.2 Hessian

Let $f : S \to \mathbb{R}$ be defined by

$$f(x) = \frac{(\alpha^T x)^2}{(\beta^T x)},$$

where S is a convex subset of $\mathbb{R}^n$, $\alpha, \beta \in \mathbb{R}^n$, and where $\beta^T x > 0, \forall x \in S$. Derive an expression for the Hessian of f and, hence, verify that f is convex over S.

Solution: By symple calculations we have:

$$(\nabla f(x))_i = \frac{2\alpha_i (\alpha^T x)}{(\beta^T x)} - \frac{\beta_i (\alpha^T x)^2}{(\beta^T x)^2},$$
$$(\nabla^2 f(x))_{ij} = \frac{2\alpha_i \alpha_j}{(\beta^T x)} - 2\frac{\alpha_i \beta_j (\alpha^T x)}{(\beta^T x)^2} - 2\frac{\alpha_j \beta_i (\alpha^T x)}{(\beta^T x)^2} + 2\frac{\beta_i \beta_j (\alpha^T x)^2}{(\beta^T x)^3}.$$

This is equivalent to:

$$\nabla^2 f(x) = \frac{2\alpha\alpha^T}{(\beta^T x)} - 2\frac{\alpha\beta^T (\alpha^T x)}{(\beta^T x)^2} - 2\frac{\beta\alpha^T (\alpha^T x)}{(\beta^T x)^2} + 2\frac{\beta\beta^T (\alpha^T x)^2}{(\beta^T x)^3}.$$

To show that f is convex on S, we have to show that $\nabla^2 f(x)$ is positive semidefinite for all $x \in S$; we have to show that $y^T \nabla^2 f(x) y \geq 0$ for all $y \in \mathbb{R}^n$. We can write:

$$\begin{aligned} y^T \nabla^2 f(x) y &= \frac{2}{(\beta^T x)} \left[(\alpha^T y)^2 - 2\frac{(\alpha^T y)(\beta^T y)(\alpha^T x)}{(\beta^T x)} + \frac{(\beta^T y)^2 (\alpha^T x)^2}{(\beta^T x)^2} \right] \\ &= \frac{2}{(\beta^T x)} \left[\alpha^T y - \frac{(\beta^T y)(\alpha^T x)}{(\beta^T x)} \right]^2 \geq 0. \end{aligned}$$

The last inequality is because $(\beta^T x) > 0$ for $x \in S$.

3 AGM

Use AGM to solve the following:

$$\max\{xy^2z^3 : x^3 + y^2 + z = 39, \text{and } x, y, z > 0\}$$

Solution: Let's define a and b such that $z = ay^2$ and $x^3 = by^2$, then we have $y^2(b + 1 + a) = 39$. We can write:

$$xy^2z^3 = b^{\frac{1}{3}} a^3 \left(\frac{39}{b+1+a}\right)^{\frac{13}{3}} = \left(\frac{39}{b^{\frac{12}{13}} a^{\frac{-9}{13}} + b^{\frac{-1}{13}} a^{\frac{-9}{13}} + b^{\frac{-1}{13}} a^{\frac{4}{13}}}\right)^{\frac{13}{3}}.$$

We use the AGM to find a lower bound on the dinaminator of the last expression as:

$$\begin{aligned} b^{\frac{12}{13}} a^{\frac{-9}{13}} + b^{\frac{-1}{13}} a^{\frac{-9}{13}} + b^{\frac{-1}{13}} a^{\frac{4}{13}} &= b^{\frac{12}{13}} a^{\frac{-9}{13}} + 3\left(\frac{b^{\frac{-1}{13}} a^{\frac{-9}{13}}}{3}\right) + 9\left(\frac{b^{\frac{-1}{13}} a^{\frac{4}{13}}}{9}\right) \\ &\geq 13\left(\left(\frac{1}{3}\right)^3 \left(\frac{1}{9}\right)^9\right)^{\frac{1}{13}} \end{aligned}$$

with equality iff $b^{\frac{12}{13}} a^{\frac{-9}{13}} = \frac{b^{\frac{-1}{13}} a^{\frac{-9}{13}}}{3} = \frac{b^{\frac{-1}{13}} a^{\frac{4}{13}}}{9} \quad \Rightarrow \quad a = 3, \quad , b = \frac{1}{3}$ Hence $xy^2z^3 \leq 3^{\frac{34}{3}}$ with equality iff $x = 3^{\frac{1}{3}}$, $y = 3$, and $z = 27$.

4 Iterative Methods

4.1 Example in $\mathbb{R}$

Consider the function

$$f(x) = \begin{cases} 4x^3 - 3x^4 & \text{if } \lambda \geq 0 \\ 4x^3 + 3x^4 & \text{if } \lambda < 0 \end{cases}$$

1. Show that f is twice continuously differentiable on $\mathbb{R}$.

Solution: It is clear that

$$f'(x) = \begin{cases} 12x^2 - 12x^3 & \text{if } \lambda > 0 \\ 12x^2 + 12x^3 & \text{if } \lambda < 0 \end{cases}$$

$$f''(x) = \begin{cases} 24x - 36x^2 & \text{if } \lambda > 0 \\ 24x + 36x^2 & \text{if } \lambda < 0 \end{cases}$$

It is easy to check that $f''(x)$ is continuous at $x = 0$.

2. Apply Newton's method on f for minimization from the starting point $x_0 = .40$. Show the results for the first $k = 6$ iterations. What is x_k converging to?

   Solution: The Newton's step is $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$ for $f''(x_k) \neq 0$. By starting from $x_0 = .4$ we have

   $$x_1 = 0.1,\ x_2 = 0.0471,\ x_3 = 0.0229,\ x_4 = 0.0113,\ x_5 = 0.0056,\ x_6 = 0.0028,$$

   x_k is converging to 0.

3. Now start with the initial iteration $x_0 = .60$. What can you conclude about the convergence now?

   Solution: By starting from $x_0 = .6$ we have

   $$x_1 = -0.6,\quad x_2 = 0.6,\quad x_3 = -0.6,\quad x_4 = 0.6,\quad x_5 = -0.6,\quad x_6 = 0.6,$$

   x_k is swapping between 0.6 and -0.6 and is not convergent.

4.2 MATLAB Example

Use MATLAB to find the minimum of the following functions, i.e., use the routine fminunc in MATLAB. And, solve three ways: (i) with no derivatives; (ii) with first derivatives; (iii) with first and second derivatives. (Provide the input you used and the output obtained.)

1. $$f(x) = (x_1^2 + x_2^2)^2 - x_1^2 - x_2 + x_3^2$$

   with starting point $(1, 1, 1)$.

2. $$f(x) = \pi \exp^{x_2}(5x_1^2 + 3x_1^2 + 2.3x_1x_2 + 2x_2 + 7)$$

   (State the starting point you used.)