Nonlinear Optimization Lecture 6

Garrick Aden-Buie

Thursday, January 28, 2016

Continuing from last time...

5

Let $f\colon S \to \mathbb{R}$ and $S \subset \mathbb{R}$: open, convex, non-empty.

Consider: $f(x) = ax^2$. This is convex depending on $a > 0$. If $a < 0$ it is not convex.

Consider. $f(x) = x^n$. If n is even $\Rightarrow$ convex. If $n = 1$? Look at $f'(x) = nx^{n-1}$ and $f''(x)n(n-1)x^{n-2}$. $f'' \geq 0, \forall x$. If $n = 3$, $f'' = 6x$, if $n = 4$, then $f'' = 12x^2 \geq 0$.

Returning to this item, if $f \in C^2(S)$, then f is convex on S if and only if $\mathrm{H}(x)$ is *positive semi-definite* (PSD) for all $x \in S$.

Def. A square matrix (n x n) A is **positive semi-definite** if and only if $\nu^T A \nu \geq 0$, $\forall \nu \in \mathbb{R}^n$.

$$
\begin{aligned}
f(x) &= x^n \\
f'(x) &= nx^{n-1} \\
f''(x) - n(n-1)x^{n-2} &= \mathrm{H}(x) \geq 0 \forall x \in S
\end{aligned}
$$

This definition is for *positive semi-definite*, but a matrix A is **positive definite** if and only if $\nu^T A \nu > 0$,

$$
\begin{aligned}
\text{Positive Definite} \Leftrightarrow \quad \nu^T A \nu &> 0 \quad \forall \nu \in \mathbb{R}^n \\
\text{Negative Semi-Definite} \qquad\qquad &\leq 0 \\
\text{Negative Definite} \qquad\qquad &< 0
\end{aligned}
$$

Proof ($\Rightarrow$).

Let $x, \bar{x} \in S$.

f is convex $\Rightarrow f(x) \geq f(\bar{x}) + [\nabla f(\bar{x})]^T (x - \bar{x})$.

Taylor Expansion $\Rightarrow f(x) = f(\bar{x}) + [\nabla f(\bar{x})]^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T H(\bar{x})(x - \bar{x}) + \|x - \bar{x}\|^2 \alpha(\bar{x}; x - \bar{x})$. Note the part starting with the fracion thereafteer is always ≥ 0.

Let $x = \bar{x} + \lambda d$, $d \in \mathbb{R}$ $\lambda\, 0$.

$$
\begin{aligned}
\frac{\lambda^2}{2} d^T H(\bar{x}) d + \lambda^2 \|d\|^2 \alpha(\bar{x}; \lambda d) \geq 0 \\
\Rightarrow d^T H(\bar{x}) d + 2\|d\|^2 \alpha(\bar{x}; \lambda d) \geq 0 \\
\text{Let } \lambda \to 0^+ \\
d^T H(\bar{x}) d \geq 0 \ \forall d \in \mathbb{R}^n \\
\Rightarrow H(\bar{x}) \text{ is PSD.}
\end{aligned}
$$

Proof (⇐).

Let $x, \bar{x} \in S$. From the SOMVT, $\exists \hat{x} = \lambda x + (1-\lambda)\bar{x},\ \forall x \in (0,1)$ such that

$$f(x) = f(\bar{x}) + (x-\bar{x})(x-\bar{x})^T \nabla f(\bar{x}) + \frac{1}{2}(x-\bar{x})^T H(\hat{x})(x-\bar{x})$$

S is convex $\Leftrightarrow \hat{x} \in S \Rightarrow H(\hat{x})$ is PSD. (Note again that the portion starting with $frac12$ is ≥ 0.)

$$\Rightarrow f(x) {\geq} f(\bar{x}) + (x-\bar{x})^T \nabla f(\bar{x}) \Leftrightarrow f \text{ is convex}$$

6

Consider a function $p\colon \mathbb{R}^n \to \mathbb{R}$ that satisfies:

$$
\begin{aligned}
p(x) &\geq 0 \\
p(\alpha x) &= |\alpha| p(x) \\
p(x+y) &\leq p(x) + p(y) \\
p(x) &= 0 & \Leftrightarrow x = 0
\end{aligned}
$$

This is a **norm**. The most common is the *Euclidean norm* ($\|x\|$).

Any vector norm is a convex function:

For any $\lambda \in [0,1]$:

$$\|\lambda x + (1-\lambda) y\| \leq \|\lambda x\| + \|(1-\lambda) y\| = \lambda \|x\| + (1-\lambda)\|y\|$$

7

Let $f(x) = x^T Q x$ and Q is a symmetric matrix.

f is convex $\Leftrightarrow Q$ is PSD.

f is strictly convex $\Leftrightarrow Q$ is PD.

8

In general, for $f(x)$, $H(x)$ is PD $\forall x \Rightarrow f$ is strictly convex.

But, if f is *strictly convex* $\Rightarrow H(x)$ is PSD $\forall x$.

The result is that PD is a good test for strict convexity. But PD does not always guarantee strict convexity $\leftarrow$ **look this up!**.

Note. The notation for PD is generally $H \succ 0$ and for PSD it is $H \succeq 0$.

Note. Read the textbook for some notes on test for PD and PSD.

Optimization

Consider the problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t} \quad & x \in S \subset \mathbb{R}^n \end{aligned}$$

(1) $\bar{x}$ is feasible if $\bar{x} \in S$ (admissable).

(2) $\bar{x}$ is a global min if $\bar{x} \in S$ and $f(\bar{x}) \leq f(x)\ \forall x \in S$.

(3) $\bar{x}$ is a local minimum if $\bar{x} \in S$ and $\exists \epsilon > 0\colon f(\bar{x}) \leq f(x)\ \forall x \in S \cap N_\epsilon(\bar{x})$.

(4) $\bar{x}$ is the unique global minimum if $\bar{x} \in S$ and $f(\bar{x}) < f(x)\ \forall x \in S$ and $x \neq \bar{x}$.

Theorem

Theorem. Consider the same problem above, where $S \subset \mathbb{R}^n$ is convex and non-empty. Let $\bar{x}$ be a local minimum.

(1) f is convex on $S \Rightarrow \bar{x}$ is global minimum.

(2) If f is strictly convex on $S \Rightarrow \bar{x}$ is unique global minimum.

Proof (1). Suppose not (proof by contradiction).

- Assume that $\bar{x}$ is not a global minimum (but still a local minimum): then $\exists \hat{x} \in S\colon f(\hat{x}) < f(\bar{x})$.
- Let $x_\lambda = \lambda\hat{x} + (1-\lambda)\bar{x},\ \forall \lambda \in (0,1)$,
- $f(x_\lambda) \leq \lambda f(\hat{x}) + (1-\lambda)f(\bar{x}) < \lambda f(\bar{x}) + (1-\lambda)f(\bar{x}) = f(\bar{x})$.
  - The idea here is that if a better solution exist ($\hat{x}$), then you can find a point between $\bar{x}$ and $\hat{x}$ that is a better solution than $\bar{x}$.
- Then $\bar{x}$ is not a local minimum.

- Contradiction. $\bar{x}$ must be global minimum.

Proof (2). Is done similarly and left as an exercise to the class.

This applies to solvers: find local minima, check Hessian at each point, decide on best solution (no guarantee of global minimum).

Theorem Variational Inequality Problem

Theorem. Convex problem, let $f \in C^1(S)$ and

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t} \quad & x \in S \end{aligned}$$

Then $\bar{x} \in S$ is a global minimum if and only if

$$[\nabla f(\bar{x})]^T (x - \bar{x}) \geq 0 \; \forall x \in S$$

.

Sidenote: Normally we say things like $\bar{x} \in S \Rightarrow$ Condition A is true (necessary condition). Example, for $f(x) = ax^2$, then setting $f'(x) = 2ax = 0$ and finding $\bar{x} = 0$. The necessary condition is that $f' = 0$. But the reverse is not true because a may be negative, so the arrow has to $\Rightarrow$ and $\Leftarrow$ is not necessarily implied by Condition A. These are called: $\Rightarrow A$, condition A is a necessary condition; $\Leftarrow B$, condition B is a sufficient condition; $\Leftrightarrow C$, condition C is called *sufficient and necessary.*

Proof ($\Rightarrow$). Suppose not (proof by contradiction), that the second statement is false.

- Assume that $\exists \hat{x} \in S$ such that $[\nabla f(\bar{x})]^T (x - \bar{x}) < 0$.
- Let $x_\lambda = \lambda \hat{x} + (1 - \lambda)\bar{x}, \; \lambda \in (0, 1)$.

$$\begin{aligned} f(x_\lambda) &= f(\bar{x} + \lambda(\hat{x} - \bar{x})) \\ &= f(\bar{x}) + \lambda [\nabla f(\bar{x})]^T (\hat{x} - \bar{x}) + \lambda[\cdots] \\ &= f(\bar{x}) + \lambda \{[\nabla f(\bar{x})]^T (\hat{x} - \bar{x}) + \lambda[\cdots]\} \\ &= f(\bar{x}) + \{< 0 \text{ for sufficiently small } \lambda\} \end{aligned}$$

- That is, for sufficiently small λ, $f(x_\lambda) < f(\bar{x}) \Rightarrow \bar{x}$ is *not* a global minimum.
- Contradiction.

Proof ($\Leftarrow$). f is convex.

$$
\begin{aligned}
f(x) &\geq f(\bar{x}) + [\nabla f(\bar{x})]^T (x - \bar{x}) \ \forall x \in S \\
&= f(\bar{x}) + [\geq 0] \\
\Rightarrow f(x) &\geq f(\bar{x}) \ \forall x \in S
\end{aligned}
$$

$\Rightarrow \bar{x}$ is a global min.