1 Overview

In this lecture, we continue to discuss (twice) differentiable convex function over convex set, including first/second necessary and sufficient conditions for optimality.

2 Continue on differentiable convex function

2.1 Conditions for convex function

Theorem 1 If \( f : c \subseteq \mathbb{R}^n \rightarrow \mathbb{R} \) is twice differentiable on \( c \), where \( c \) is a convex set with non-empty interior; then \( f \) is convex iff the Hessian \( \nabla^2 f \) is positive semidefinite.

Proof:

- Using second order Taylor theorem, we know that for \( x, y \in c \), \( \exists 0 \leq \alpha \leq 1 \),
  \[
  f(y) = f(x) + \nabla f(x)(y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(\alpha y + (1 - \alpha) x)(y - x)
  \]
  so that if \( \nabla^2 f(z) \) is positive semidefinite for all \( z \in c \), then choosing \( x, y \in c \) arbitrarily, \( f(y) \geq f(x) + \nabla f(x)(y - x) \), since the second item is non-negative, concluding that \( f \) is convex on \( c \).

- if \( f \) is convex on \( c \), choose \( x \in c \); if \( \nabla^2 f(x) \) is non-positive semidefinite at \( x \), there is \( d \in \mathbb{R}^n \), \( d^T \nabla^2 f(x)d < 0 \). So there is \( \alpha > 0 \), define \( y = x + \alpha d \),
  \[
  (y - x)^T \nabla^2 f(x)(y - x) < 0;
  \]
  By continuity of \( \nabla^2 f(x) \), at \( x \), \( \exists \alpha \), such that,
  \[
  (x + \alpha d)^T \nabla^2 f(x)(x + \alpha d) < 0,
  \]
for every $0 < \alpha < \overline{\alpha}$,

\[
f(y) = f(x) + \nabla f(x)(y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(\alpha y + (1 - \alpha)x)(y - x),
\]

since the last item < 0,

\[
f(y) < f(x) + \nabla f(x)(y - x)
\]

contradicting that $f$ is convex.

\[\blacksquare\]

**Theorem 2** Let $f : c \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ be convex over convex set $c$, and $f$ has continuous first order derivative. If $x^*$ is a point where $\nabla f(x^*)(y - x^*) \geq 0$ for every $y \in c$, then $x^*$ is a global minimum of $f$.

**Proof:** We know $f(x) \geq f(x^*) + \nabla f(x^*)(y - x^*) \geq 0$ by convexity of $f$, since the second item $\geq 0$, we have $f(x) \geq f(x^*)$.

\[\blacksquare\]

**2.2 Definition: Feasible direction**

Let $f : c \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous function over $c$. A vector $d$ is a feasible direction at a point $x \in c$ if there exists some $\overline{\alpha} \geq 0$ such that $x + \alpha d \in c$ for every $0 \leq \alpha \leq \overline{\alpha}$.
3 Conditions for optimality

3.1 First order necessary condition for optimality

**Theorem 3** Let \( f : c \subseteq \mathbb{R}^n \rightarrow \mathbb{R} \) be a \( c^1 \) function. If \( x^* \) is a local minimum of \( f \), then for every feasible direction of \( d \), \( \nabla f(x^*)d \geq 0 \). In particular, if \( x^* \in \text{int}(c) \), then \( \nabla f(x^*) = 0 \).

**Proof:**

- Let \( d \) be a feasible direction, and define \( g : [0, \alpha] \rightarrow \mathbb{R} \) where \( g(\alpha) = f(x^* + \alpha d) \),

\[
g(\alpha) - g(0) = g'(0)\alpha + o(\alpha).
\]

If \( g'(0) < 0 \), then \( \exists \alpha > 0 \) such that the right hand side becomes negative, which means \( g(\alpha) < g(0) \). Which in turn implies that \( f(x^* + \alpha d) < f(x^*) \), contradicting the \( x^* \) being a local minimum.

- \( g'(0) \geq 0 \), then the directional derivative of \( f \) at \( x^* \) in the direction of \( d \) is non-negative, thus: \( \nabla f(x^*)d \geq 0 \). For the second part, since all directions are feasible, we have \( \nabla f(x^*)d \geq 0 \), \( \nabla f(x^*)(-d) \geq 0 \), so \( \nabla f(x^*) = 0 \).

\[\square\]

3.2 Second order necessary condition for optimality

**Theorem 4** Let \( f : c \subseteq \mathbb{R}^n \rightarrow \mathbb{R} \) be a \( c^2 \) function. If \( x^* \) is a local minimum of \( f \), then for every feasible direction of \( d \) at \( x^* \), we have:

- \( \nabla f(x^*)d \geq 0 \).
- If \( \nabla f(x^*) = 0 \), then \( d^T f^2(x^*)d \geq 0 \).
Proof: Let $\nabla f(x^*)d = 0$, define $g(\alpha) = f(x^* + \alpha d)$. In this case $g'(\alpha) = 0$, then $g(\alpha) - g(0) = \frac{1}{2}g''(0)\alpha^2 + o(\alpha^2)$. If $g''(0) < 0$, then for sufficiently small $\alpha$ the right hand side becomes negative and then $g(\alpha) < g(0)$, which implies $f(x^* + \alpha d) < f(x^*)$, contradicting local minimality of $x^*$. $g''(0) = d^T\nabla^2 f(x^*)d \geq 0$.

3.3 Second order sufficient condition for local optimality

**Theorem 5** For function $f : c \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ to be $c^2$ function, the interior point $x^*$ is a local minimum if $\nabla f(x^*) = 0$ and the Hessian $\nabla^2 f(x^*)$ is positive definitive.

**Proof:** Since $\nabla^2 f(x^*)$ is positive definitive, there $\exists \alpha > 0$ such that $d^T\nabla^2 f(x^*)d > \alpha \|d\|^2$. Then for any $d$, $f(x^* + d) - f(x^*) = \frac{1}{2}d^T\nabla^2 f(x^*)d + o(\|d\|^2) \geq \frac{1}{2}\alpha \|d\|^2 + o(\|d\|^2) > 0$

A is a square matrix which is invertible and symmetric $A \geq 0$.

$$f(x) = X^TAX + C^TX + d = \sum_{i,j} a_{i,j}x_i x_j + \sum_i c_i x_i + d$$

so that, $$\frac{\partial f}{\partial x_i} = \sum_j a_{i,j}x_j + c_i = A_{i.,x} + c_i$$

and, $$\nabla f(x) = (..., \frac{\partial f}{\partial x_i}, ...) = (..., A_{i.,x} + c_i, ...) + C^T = \{AX\}^T + C^T = X^T A^T + C^T$$

thus, $$\nabla f(x) = 0 \Rightarrow X^T A = -C^T \Rightarrow AX = -C \Rightarrow x^* = -A^{-1} C$$

for the second order, $$\frac{\partial f}{\partial x_i \partial x_j} \left( \sum_{i,j} a_{i,j} x_i x_j \right) = A_{i,j} + A_{j,i} = A + A^T = 2A \geq 0$$

since $A$ is symmetric. Concluding that $A$ is positive semidef.

3.4 Example: Linear Multiple Regression

$$\tilde{x}_1 \rightarrow y_1$$

$$\tilde{x}_2 \rightarrow y_2$$

$$\vdots$$

$$\tilde{x}_m \rightarrow y_m$$
and,
\[ y = a_0 + a_1 x_1 + ... a_n x_n \]

Our objective is to,
\[
\begin{align*}
\min & \quad \|Y - XA\|^2 \\
\Rightarrow & \quad \min \|Y - XA\|^T \|Y - XA\| \\
= & \quad A^T X^T X A - 2Y^T X A + Y^T Y \\
\end{align*}
\]

Using the result from Theorem 5, \( x^* = -A^{-1} C \), we get,
\[
A^* = 2(X^T X)^{-1} X^T Y
\]