Lecture 4
Scribe?

Last time

- 2nd-order optimality cond.
- Basic convexity

More on convexity

**Agenda**
- More on convexity
- Characterization smooth convex.
- Subgradients.

More on convexity

**Lemma:** Assume that $C_1, C_2 \subseteq \mathbb{R}^n$ convex sets.

Then, the following are convex

1. (Scaling) $\lambda C_1 = \{ \lambda x \mid \lambda \geq 0 \text{ and } x \in C_1 \}$
2. (Sums) $C_1 + C_2 = \{ x_1 + x_2 \mid x_1 \in C_1, x_2 \in C_2 \}$
3. (Intersections) $C_1 \cap C_2$
4. (Linear images and preimages)

Let $A : \mathbb{R}^d \to \mathbb{R}^n$ is linear,

$A C_1$ and $A^{-1} C_3$ are convex.

**Intuition**
Proof: Exercise

Equivalence of operations

<table>
<thead>
<tr>
<th>Function</th>
<th>Epigraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda f(x/a) )</td>
<td>( \lambda \text{epi } f )</td>
</tr>
<tr>
<td>( \max_i f_i )</td>
<td>( \cap \text{epi } f )</td>
</tr>
<tr>
<td>( f(AX) )</td>
<td>( [AXI]^{-1} \text{epi } f )</td>
</tr>
</tbody>
</table>

**Lemma** (First-order characterization of convexity)

Suppose that \( f : \mathbb{R}^d \to \mathbb{R} \) is differentiable. Then, the following are equivalent:

1. \( f \) is convex.
2. \( \forall x,y \in \mathbb{R}^d \) \( f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle \)
3. \( \forall x, y \) \( \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0 \)

Intuition for (2): \( f(x) + \langle \nabla f(x), y-x \rangle \) supports the epigraph.
Intuition for (3): In 1D the function is monotone.

**Proof:** (1) $\Rightarrow$ (2) Let $x, y \in \mathbb{R}^d$ and $t \in (0, 1)$. Convexity ensures that

$$f(t \cdot x + \lambda(y - x)) \leq (1-t)f(x) + \lambda f(y)$$

$$\Rightarrow$$

$$f(x + \lambda(y - x)) - f(x) \leq f(y) - f(x)$$

Taking $\lambda \to 0 \Rightarrow \langle f(x), x-y \rangle + f(x) \leq f(y)$.

(2) $\Leftarrow$ (1) Let $x, y \in \mathbb{R}, \lambda \in [0,1]$ and $t \in (0,1)$.

$$z_t = (1-t)x + ty$$

$$\Rightarrow$$

$$f(x) \geq f(z_t) + \langle \nabla f(z_t), x - z_t \rangle \quad (\text{1})$$

$$f(y) \geq f(z_t) + \langle \nabla f(z_t), y - z_t \rangle \quad (\text{2})$$

$$\Rightarrow$$

$$(1-t) \cdot (\text{1}) + t \cdot (\text{2}) \text{ gives}$$

$$(1-t)f(x) + tf(y) \geq f(z_t) + \langle \nabla f(z_t), (1-t)x + ty - z_t \rangle$$

$$= f(z_t).$$
(2) ⇒ (3)  \[ f(x) = f(y) + \nabla f(y)^T (x - y) \]
\[ + f(y) = f(x) + \nabla f(x)^T (y - x) \]
\[ 0 \geq (\nabla f(x) - \nabla f(y))^T (y - x) \]

(3) ⇒ (2)  Define \( \Psi(t) = f(x + t(y - x)) \)

Then \( f(y) = \Psi(1) = \Psi(0) + \int_0^1 \Psi'(t) \, dt \)
\[ = \Psi(0) + \Psi'(0) + \int_0^1 \left[ \Psi'(t) - \Psi'(0) \right] \, dt \]
\[ = f(x) + \nabla f(x)(y - x) \]
\[ = f(x) + \nabla f(x)(y - x) + \int_0^1 \nabla f(x + t(y - x))^T (y - x) \, dt \]
\[ \geq f(x) + \nabla f(x)(y - x) \]

\[ \square \]

Lemma 2nd-order characterization
Assume \( f \) twice differentiable. Then,
\( f \) is convex \( \iff \) \( \nabla^2 f(x) \geq 0 \ \forall x. \)
Intuition

Second order model never curves down!

\[ y = f(x) + \langle v; y-x \rangle \]

Proof \hspace{1cm} \Box

Question: How can we assess optimality for general convex functions?

Idea: Consider a convex function \( f: \mathbb{R}^d \rightarrow \mathbb{R} \). The subdifferential of \( f \) at \( x \) is

\[ \partial f(x) = \{ v \mid \forall y \in \mathbb{R}^d \quad f(y) \geq f(x) + \langle v; y-x \rangle \} \]
Examples

1. \( f(x) = |x| \)

2. \( f(x) = \|x\| \)

3. \( f(x) = \max \{0, x\} \)  
   \[ \Theta f(x) = \begin{cases} x & x \geq 0 \\ 0 & -1 \leq x < 0 \end{cases} \]
What do we just gained? general

**Theorem:** Optimality cond for convex func.
Suppose \( f: \mathbb{R}^d \to \mathbb{R} \) is convex. Then \( x^* \) is a minimizer iff \( 0 \in \partial f(x^*). \)

Intuition

\[ \text{Nothing goes under.} \]

**Proof:** Assume \( x^* \) is a minimizer.
\[
f(x^*) + \langle 0, y - x \rangle \leq f(y) \quad \forall y.
\]
Assume that \( 0 \in \partial f(x^*). \)

**Proposition:** Subdifferential calculus
Suppose that \( f, h: \mathbb{R}^d \to \mathbb{R} \) are convex functions. Then the following holds

\[
\left( \sum_{\varepsilon} \partial (f + h)(x) = \partial f(x) + \partial h(x). \right)
\]