Lecture 4

Previously

LPs
- 4 cases: unbounded, infeasible, unique, non-unique
- simplex algorithm

NOTE: hyperplane: \( a^Tx = b \)  
\[ a \in \mathbb{R} \quad b \in \mathbb{R} \]

necessary conditions for optimality:
- unconstrained optimization
- quadratic + linear  \( x^TAx + c^Tx \)
- quadratic s.t. linear equality constraints
- quadratic s.t. linear inequality constraints

convex analysis:
- line segment between 2 points
- convex sets
- arbitrary intersection of convex sets is convex

examples of convex sets
- convex hull of A:  \( \bigcap_{c \in C} c \subset \mathbb{R} \)

convex combinations
- convex sets contain all convex combinations
- equivalent definition of convex hull

operations that preserve convexity of sets
- graph & epigraph of functions
- convex functions
\( f: A \to \mathbb{R} \quad A \subseteq \mathbb{R}^n \)

\[ \text{epi} \, f := \{ (x,y) : x \in A, \ y \in \mathbb{R}, \ y \geq f(x) \} \subseteq \mathbb{R}^{n+1} \]

What is the normal vector to the epigraph at a point \((x, f(x))\)?

\[
\begin{bmatrix}
\nabla f(x) \\
-1
\end{bmatrix}
\]

What directions can we go in and stay on the graph?

\((x, f(x)) \mapsto (x + v, f(x) + z)\)

to be on the graph (to first order)

\[ f(x) + z = f(x + v) \]

\[ z \approx \nabla f(x)^T (v - x) \]

\[ z \approx \nabla f(x)^T v - \nabla f(x)^T x \]

\[ \underbrace{\nabla f(x)}_{\\text{constant}} \begin{bmatrix} v \\ z \end{bmatrix} + c \approx 0 \]

**f: A \to \mathbb{R} \quad A \subseteq \mathbb{R}^n**

f is a **convex function** if \( \text{epi} \, f \) is a convex set.

f is a **closed convex function** if \( \text{epi} \, f \) is a closed convex set.
EX: \( f : (-1,1) \rightarrow \mathbb{R} \)

\[
f(x) = \begin{cases} 
0 & x \in (-1,1) \\
+\infty & \text{o.w.}
\end{cases}
\]

not closed

\[ f : \mathbb{R}_{>0} \rightarrow \mathbb{R} \]

\[
f(x) = \begin{cases} 
1 & x > 0 \\
+\infty & \text{o.w.}
\end{cases}
\]
domain open

epi \( f \) closed

\[ f : [-1,1] \rightarrow \mathbb{R} \]

\[
f(x) = \begin{cases} 
0 & x \in (-1,1) \\
1 & x = 1 \\
+\infty & \text{o.w.}
\end{cases}
\]

weird boundary stuff

example

\[
f(x) = \min \{ y : (x,y) \in \text{epi} f \}
\]
Proposition: \( f: A \rightarrow \mathbb{R} \) is a convex function (as defined by epi \( f \) being convex)

then:
- \( A \) (domain of \( f \)) is a convex set
- \( \forall \ x_1, x_2 \in A \)
- \( \forall \ \theta \in [0,1] \)

\[
f(\theta x_1 + (1-\theta) x_2) \leq \theta f(x_1) + (1-\theta) f(x_2)
\]

the graph lies below the secant lines

\( f \) is a concave function if \( -f \) is convex

\( f \) is concave \(\Rightarrow\) domain of \( f \) is convex

**Ex** convex functions (co-domain is always \( \mathbb{R} \))

- affine: \( x \mapsto a^T x + b \)
  \( \mathbb{R}^n \rightarrow \mathbb{R} \)
- exponential: \( x \mapsto \exp(ax) \) \( a \in \mathbb{R} \)
  \( \mathbb{R} \rightarrow \mathbb{R} \)
- powers: \( x \mapsto x^\alpha \) \( \alpha \geq 1 \)
  \( \mathbb{R}_{>0} \rightarrow \mathbb{R} \) \( \alpha \leq 0 \)
- powers of abs. value: \( x \mapsto |x|^p \) \( p \geq 1 \)
  \( \mathbb{R} \rightarrow \mathbb{R} \)
\[ \text{relu: } x \mapsto \max(0, x) \]

concave
- affine: \( x \mapsto a^T x + b \)
- power: \( x \mapsto x^\alpha \) \( \alpha \in [0, 1] \)
  \( \mathbb{R}_{\geq 0} \to \mathbb{R} \)
- log: \( x \mapsto \log x \)
  \( \mathbb{R}_{>0} \to \mathbb{R} \)
- entropy: \( x \mapsto -x \log x \)
  \( \mathbb{R}_{>0} \to \mathbb{R} \)

\[ \text{convex functions:} \]
- norms: \( x \mapsto \|x\| \)
  \( \mathbb{R}^n \to \mathbb{R} \)
- sum of squares: \( x \mapsto \|x\|_2^2 = \sum_{i=1}^n x_i^2 \)
  \( \mathbb{R}^n \to \mathbb{R} \)
- max: \( x \mapsto \max_i x_i \)

\[ \text{f convex iff restrictions to lines are all convex} \]
\[ \text{f: } A \to \mathbb{R} \text{ convex iff} \]
\[ t \mapsto f(x+tv) \text{ is convex for all } x \in A \quad v \in \mathbb{R}^n \]
\[ \text{domain of this: } \{ t: (x+tv) \in A \} \]
Exercise 1: Let \( S^n \subseteq \mathbb{R}^{n \times n} \) be symmetric.

\( S^n_{>0} \subseteq S^n \)

P.D. matrices

\[ f: S^n_{>0} \rightarrow \mathbb{R} \]

\[ f(X) = \log \det X \]

using the above fact:

pick any \( X \in S^n_{>0} \), \( V \in S^n \)

\[ g(t) = \log \det (X + tV) \]

\[ \mathbb{R} \rightarrow \mathbb{R} \]

\[ \det(AB) = \det(A) \det(B) \]

\[ X + tV = X^{\frac{1}{2}} [I + tX^{-\frac{1}{2}} VX^{-\frac{1}{2}}] X^{\frac{1}{2}} \]

\[ \det(X + tV) = \det(X^{\frac{1}{2}}) \det(I + tX^{-\frac{1}{2}} VX^{-\frac{1}{2}}) \]

\[ = \det(X) \det(I + tX^{-\frac{1}{2}} VX^{-\frac{1}{2}}) \]

\[ \lambda_i \text{ be eigenvalues of } X^{\frac{1}{2}} VX^{\frac{1}{2}} \]

\[ \det(I + tX^{\frac{1}{2}} VX^{\frac{1}{2}}) \]

\[ = \prod_{i=1}^{n} (1 + t\lambda_i) \]

we can that \( \lambda_i \) are constants!
\( \log \det (X + tv) = \log \det (X) + \sum_{i=1}^{n} \log (1 + t \lambda_i) \)

\( t \mapsto \sum_{i=1}^{n} \log (1 + t \lambda_i) \) concave

\( \forall X \in \mathbb{S}_{>0}^n, \forall V \in \mathbb{S}^n \),

\( \Rightarrow f(X) = \log \det X \) concave

---

**First-order condition**

\( f : \mathbb{A} \to \mathbb{R}^n \)

\( f \) is differentiable

\( f \) convex iff

\( f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0) \)

\( \forall x, x_0 \in \mathbb{A} \)

Graph of \( f \) lies above the tangent hyperplane

intuition: tangent is the limit of secants
\( \implies \) Suppose \( f \) is convex.

\[
\text{WTS: } \forall x, x_0 \quad f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0)
\]

by convexity: \( \forall \theta \in [0, 1] \quad \theta x + (1 - \theta) x_0 \)

\[
\theta f(x) + (1 - \theta) f(x_0) \geq f(x + \theta(x-x_0))
\]

\[
f(x) - f(x_0) \geq \frac{f(x_0 + \theta(x-x_0)) - f(x_0)}{\theta}
\]

\[
lim_{\theta \to 0} f(x) - f(x_0) \geq \nabla f(x_0)^T (x - x_0)
\]

\( \Leftarrow \) we will show later (supporting hyperplane theorem)

\[
\text{supponting hyperplane of } C \text{ at } x_0 \in C
\]

\[
a^T x \leq a^T x_0 \quad \forall x \in C
\]

\[
C : \text{ boundary of } C
\]

\[
\text{supporting hyperplane of } C \text{ at } x_0
\]

\[
\partial C \subset \text{ boundary of } C
\]

\[
\text{supporting hyperplane of } C \text{ at } x_0
\]

\[
(\partial C \cap C \setminus \text{int } C)
\]

every point on the boundary has a supporting hyperplane \( \implies \) convexity

---

Subgradients

\( f : A \to \mathbb{R} \)

A \( \subset \mathbb{R}^n \)

\( f \) is not necessarily differentiable

\( g \in \mathbb{R}^n \) is a subgradient of \( f \) at \( x_0 \)

if \( f(x) \geq f(x_0) + g^T (x-x_0) \) \( \forall x \)

linear global underapproximation of \( f \)
if $f$ is differentiable at $x_0$, then the only subgradient is $g = \nabla f(x)$

more generally:

$$f(x) = |x|$$

$\mathbb{R} \to \mathbb{R}$

$g \in \mathbb{R} \cup \mathbb{R}$ is a subgradient at $x = 0$

function: linear global
underapproximations
set: supporting hyperplanes

$x_{opt}$ is minimum of $f$ if $0$ is a subgradient of $f$
at $x_{opt}$

$$f(x) \geq f(x_{opt}) + g^T(x - x_{opt})$$

subdifferentials

$f: A \to \mathbb{R}$

$A \subset \mathbb{R}^n$

$\partial f: A \to 2^{\mathbb{R}^n}$

'multifunction': $\partial f: A \to 2^{\mathbb{R}^n}$
\( \partial f(x) = \text{set of all subgradients of } f \text{ at } x \quad x \in A \)

\[ \frac{d \partial f(x)}{dx} = \partial f(x) \exists \forall f(x) \]

Sometimes \( \partial f(x) = \emptyset \)

**EX** \( f(x) = |x| \)

![Graph of |x| function](attachment:graph1.png)

\( f(x) = \frac{1}{2} \cdot \text{b}^2 \)

![Graph of \( \frac{1}{2} \cdot \text{b}^2 \) function](attachment:graph2.png)

**Property:** for any \( f \) (not nec. convex) \( \partial f(x) \) is convex & closed for any \( x \)