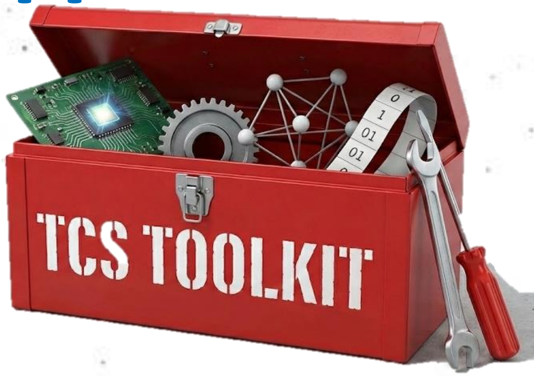
CS 58500 – Theoretical Computer Science Toolkit

Lecture 8/9 (02/12, 02/17)

Convex Geometry and Optimization I



https://ruizhezhang.com/course_spring_2026.html

Today's Lecture

- Why Non-Convex Optimization Is Hard
- Convexity
- Cutting-Plane Method
- Log-concavity
- Brunn-Minkowski Inequality

Why Non-Convex Optimization Is Hard

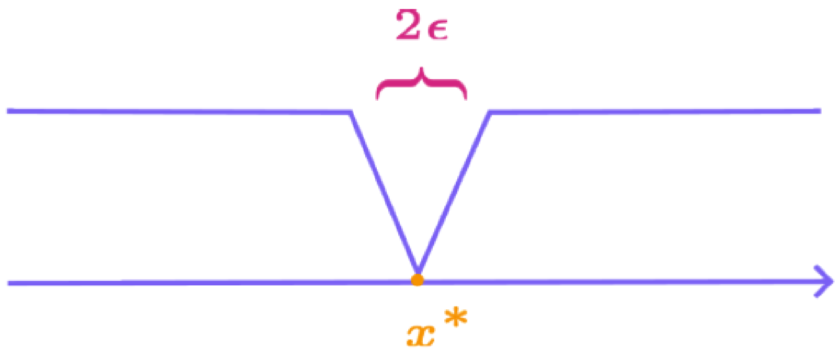
Simple Counterexample:

$$f(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x} = \boldsymbol{x}^\star \\ 1 & \text{othewise} \end{cases}$$

Lipschitz Counterexample:

$$f(\boldsymbol{x}) = \min\{\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2, \epsilon\} \qquad \forall \boldsymbol{x} \in B_2^n$$

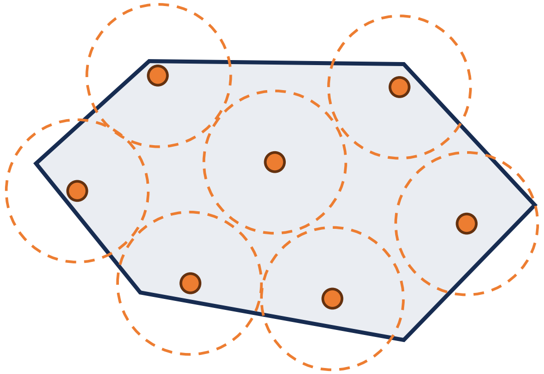
- f is 1-Lipschitz
- To be able to minimize f, we need get access to the region where $f(\boldsymbol{x}) \neq \epsilon$
- $\mathrm{Vol}_n(\{\boldsymbol{x} : f(\boldsymbol{x}) \neq \epsilon\}) = \epsilon^n \cdot \mathrm{Vol}_n(B_2^n)$
- Need $\Omega(1/\epsilon^n)$ queries to f



Detour: Packing and Covering

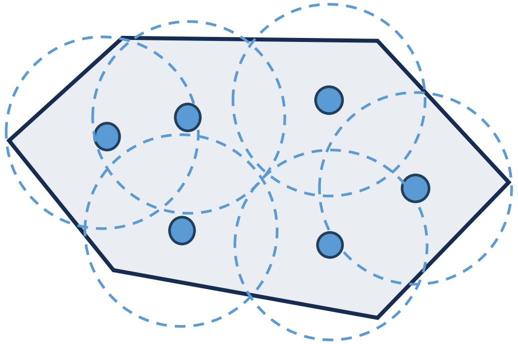
Let $S \subseteq \mathbb{R}^n$

- Packing number $M(S,\epsilon) \coloneqq \max\left\{M : \exists\{\boldsymbol{x}_i\}_{i\in[M]} \subset S \text{ such that } \left\|x_i - x_j\right\|_2 > \epsilon \text{ for all } i \neq j\right\}$
- Covering number $N(S,\epsilon) \coloneqq \min\left\{N : \{\boldsymbol{x}_i\}_{i\in[N]} \subset S \text{ such that } S \subseteq \bigcup_{i\in[N]}(\boldsymbol{x}_i + \epsilon B_2^n)\right\}$



Theorem.

- $M(S,2\epsilon) \leq N(S,\epsilon) \leq M(S,\epsilon)$
- $\left(\frac{1}{\epsilon}\right)^n \frac{\mathrm{Vol}_n(S)}{\mathrm{Vol}_n(B_2^n)} \leq N(S,\epsilon) \leq M(S,\epsilon) \leq \left(\frac{3}{\epsilon}\right)^n \frac{\mathrm{Vol}_n(S)}{\mathrm{Vol}_n(B_2^n)}$



Why Non-Convex Optimization Is Hard

Lemma. No algorithm $\mathcal{A}$ which accesses a target 1-Lipschitz function $f: B_2^n \to \mathbb{R}$ using a value oracle can optimize f to additive error ϵ in $< \left(\frac{1}{2\epsilon}\right)^n - 1$ queries.

Proof.

- Since $M(B_2^n, 2\epsilon) \geq \left(\frac{1}{2\epsilon}\right)^n$, let $\{\boldsymbol{x}_i\}_{i\in[M]}$ be such a maximal packing of B_2^n
- For each $i \in [M]$, define $f_i(\boldsymbol{x}) = \min\{\|\boldsymbol{x} - \boldsymbol{x}_i\|_2, \epsilon\}$
- If #queries $< \left(\frac{1}{2\epsilon}\right)^n - 1$, then $\exists i \neq j \in [M]$ such that no query point lies in $(\boldsymbol{x}_i + \epsilon B_2^n) \cup (\boldsymbol{x}_j + \epsilon B_2^n)$
- Thus, $\mathcal{A}$ cannot distinguish f_i and f_j

$\blacksquare$

Today's Lecture

- Why Non-Convex Optimization Is Hard
- Convexity
- Cutting-Plane Method
- Log-concavity
- Brunn-Minkowski Inequality

Convexity

Definition (Convex set). We say a set $\mathcal{X} \subseteq \mathbb{R}^n$ is convex if for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ the line segment between $\boldsymbol{x}$ and $\boldsymbol{x}'$ lies in $\mathcal{X}$, i.e.,

$$(1-\lambda)\boldsymbol{x} + \lambda\boldsymbol{x}' \in \mathcal{X} \qquad \forall\lambda \in [0,1]$$

Correspondingly, for any $\lambda \in [0,1]$, we call $(1-\lambda)\boldsymbol{x} + \lambda\boldsymbol{x}'$ a convex combination of $\boldsymbol{x}$ and $\boldsymbol{x}'$.

Definition (Convex function). We say a function $f: \mathcal{X} \to \mathbb{R}$ is convex if $\mathcal{X}$ is convex and for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ the linear interpolation of $f(\boldsymbol{x})$ and $f(\boldsymbol{x}')$ overestimates the function on the line segment between $\boldsymbol{x}$ and $\boldsymbol{x}'$, i.e.,

$$f\big((1-\lambda)\boldsymbol{x} + \lambda\boldsymbol{x}'\big) \leq (1-\lambda)f(\boldsymbol{x}) + \lambda f(\boldsymbol{x}') \qquad \forall\lambda \in [0,1]$$

Convexity

Convex set $\implies$ convex function:

- For any convex $S \subseteq \mathbb{R}^n$, define the indicator function of S by

$$\delta_S(\boldsymbol{x}) := \begin{cases} 0 & \text{if } \boldsymbol{x} \in S \\ \infty & \text{if } \boldsymbol{x} \notin S \end{cases} \qquad \forall \boldsymbol{x} \in \mathbb{R}^n$$

- δ_S is a convex function

Convex function $\implies$ convex set:

- The epigraph of $f: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is

$$\text{epi}(f) := \{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R} : t \geq f(\boldsymbol{x})\}$$

- f is convex if and only if $\text{epi}(f)$ is convex
- $\min_{\boldsymbol{x}} f(\boldsymbol{x}) \iff \min_{(\boldsymbol{x},t) \in \text{epi}(f)} t$

Convexity

Convex set $\Longrightarrow$ convex function:

- For any convex $S \subseteq \mathbb{R}^n$, define the indicator function of S by

$$\delta_S(\boldsymbol{x}) := \begin{cases} 0 & \text{if } \boldsymbol{x} \in S \\ \infty & \text{if } \boldsymbol{x} \notin S \end{cases} \qquad \forall \boldsymbol{x} \in \mathbb{R}^n$$

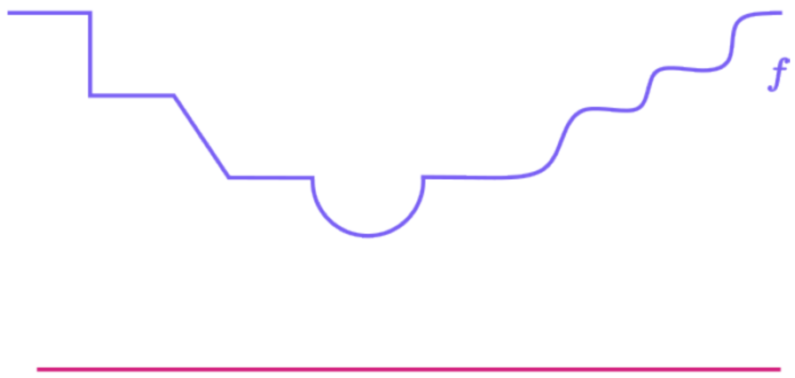
- δ_S is a convex function

Convex function $\Longrightarrow$ convex set:

- For $t \in \mathbb{R}$, the level set of f at t is

$$\{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) \leq t\}$$

- If f is convex then any level set is convex
- The converse does not hold



Quasi-convex

Convexity: Analytic Properties

Lemma. Let $f: \mathcal{X} \to \mathbb{R}$ be differentiable. Then f is convex if and only if

$$f(\boldsymbol{x}') \geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{x}' - \boldsymbol{x} \rangle$$

Proof of "⟹".

- Convexity implies that for any $\lambda \in (0,1)$,

$$\lambda f(\boldsymbol{x}') + (1-\lambda) f(\boldsymbol{x}) \geq f\big((1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{x}'\big) \iff f(\boldsymbol{x}') \geq \frac{f\big((1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{x}'\big) - (1-\lambda) f(\boldsymbol{x})}{\lambda}$$

$$\begin{aligned} f(\boldsymbol{x}') &\geq \lim_{\lambda \to 0} \frac{f\big((1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{x}'\big) - (1-\lambda) f(\boldsymbol{x})}{\lambda} \\ &= f(\boldsymbol{x}) + \lim_{\lambda \to 0} \frac{f(\boldsymbol{x} + \lambda(\boldsymbol{x}' - \boldsymbol{x})) - f(\boldsymbol{x})}{\lambda} \\ &= f(\boldsymbol{x}) + Df(\boldsymbol{x})[\boldsymbol{x}' - \boldsymbol{x}] = f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{x}' - \boldsymbol{x} \rangle \end{aligned}$$

Directional derivative

Convexity: Analytic Properties

Lemma. Let $f: \mathcal{X} \to \mathbb{R}$ be differentiable. Then f is convex if and only if

$$f(\boldsymbol{x}') \geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{x}' - \boldsymbol{x} \rangle$$

Proof of "⟸".

- For any $\lambda \in (0,1)$, let $\boldsymbol{x}_\lambda \coloneqq \lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}$
- $f(\boldsymbol{x}') \geq f(\boldsymbol{x}_\lambda) + \langle \nabla f(\boldsymbol{x}_\lambda), \boldsymbol{x}' - \boldsymbol{x}_\lambda \rangle$
- $f(\boldsymbol{x}) \geq f(\boldsymbol{x}_\lambda) + \langle \nabla f(\boldsymbol{x}_\lambda), \boldsymbol{x} - \boldsymbol{x}_\lambda \rangle$
- Thus, we have

$$\begin{aligned} \lambda f(\boldsymbol{x}') + (1-\lambda) f(\boldsymbol{x}) &\geq f(\boldsymbol{x}_\lambda) + \langle \nabla f(\boldsymbol{x}_\lambda), \lambda(\boldsymbol{x}' - \boldsymbol{x}_\lambda) + (1-\lambda)(\boldsymbol{x} - \boldsymbol{x}_\lambda) \rangle \\ &= f(\boldsymbol{x}_\lambda) + 0 = f(\lambda \boldsymbol{x}' + (1-\lambda)\boldsymbol{x}) \end{aligned}$$

■

Convexity: Analytic Properties

Lemma (First-order optimality). Let $f: \mathcal{X} \to \mathbb{R}$ be differentiable and convex. Then

$$\boldsymbol{x}^\star \in \arg\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x}) \iff \langle \nabla f(\boldsymbol{x}^\star), \boldsymbol{x}^\star - \boldsymbol{x} \rangle \leq 0 \;\; \forall \boldsymbol{x} \in \mathcal{X}$$

Proof.

- If $\langle \nabla f(\boldsymbol{x}^\star), \boldsymbol{x}^\star - \boldsymbol{x} \rangle \leq 0$, then by the previous lemma, for any $\boldsymbol{x} \in \mathcal{X}$, we have
$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^\star) + \langle \nabla f(\boldsymbol{x}^\star), \boldsymbol{x}^\star - \boldsymbol{x} \rangle \geq f(\boldsymbol{x}^\star)$$
- If $\boldsymbol{x}^\star \in \mathcal{X}$ is a minimizer, suppose $\langle \nabla f(\boldsymbol{x}^\star), \boldsymbol{x}^\star - \boldsymbol{x} \rangle > 0$ for some $\boldsymbol{x} \in \mathcal{X}$
- Consider $\boldsymbol{x}_\lambda \coloneqq (1-\lambda)\boldsymbol{x}^\star + \lambda \boldsymbol{x} \in \mathcal{X}$ for $\lambda \in [0,1]$ and $\phi(\lambda) \coloneqq f(\boldsymbol{x}_\lambda)$
- $\phi'(0) = Df(\boldsymbol{x}^\star)[\boldsymbol{x} - \boldsymbol{x}^\star] = \langle \nabla f(\boldsymbol{x}^\star), \boldsymbol{x} - \boldsymbol{x}^\star \rangle < 0$
- Thus, there exists a sufficiently small $\lambda > 0$ such that $\phi(\lambda) < \phi(0)$, i.e., $f(\boldsymbol{x}_\lambda) < f(\boldsymbol{x}^\star)$, contradiction

Constructive!

■

Convexity: Geometric Properties

Lemma (Characterization of minima and maxima). Let $f : \mathcal{X} \to \mathbb{R}$ be convex. Let f have minimizer set $\mathcal{X}_{\min} := \arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$ and maximizer set $\mathcal{X}_{\max} := \arg\max_{\boldsymbol{x}} f(\boldsymbol{x})$. Then,

- $\mathcal{X}_{\min}$ is convex
- If $\mathcal{X}_{\max} \neq \emptyset$, it either contains a boundary point of $\mathcal{X}$ or $\mathcal{X}$ has no boundary points

Moreover, if f is strictly convex, $\mathcal{X}_{\min} = \{\boldsymbol{x}^\star\}$ and $\mathcal{X}_{\max}$ only contains boundary points if it is nonempty.

- f is strictly convex if $f\big((1-\lambda)\boldsymbol{x} + \lambda\boldsymbol{x}'\big) < (1-\lambda)f(\boldsymbol{x}) + \lambda f(\boldsymbol{x}')\ \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}, \lambda \in (0,1)$
- $\boldsymbol{w} \in \mathcal{X}$ is a boundary point $\nexists \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}, \lambda \in (0,1)$ such that $w = (1-\lambda)\boldsymbol{x} + \lambda\boldsymbol{x}'$

Convexity: Geometric Properties

Theorem (Separating Hyperplane Theorem). Let $S \subset \mathbb{R}^n$ be compact and convex and suppose $\boldsymbol{x}_0 \notin S$. There is a separating hyperplane $\boldsymbol{g} \in \mathbb{R}^n$ such that

$$\langle \boldsymbol{g}, \boldsymbol{x}_0 \rangle > \langle \boldsymbol{g}, \boldsymbol{x} \rangle \quad \forall \boldsymbol{x} \in S$$

Moreover, if $\boldsymbol{x}_0 \in S$ is a boundary point, there is a supporting hyperplane $\boldsymbol{g} \in \mathbb{R}^n, \boldsymbol{g} \neq \boldsymbol{0}$ such that

$$\langle \boldsymbol{g}, \boldsymbol{x}_0 \rangle \geq \langle \boldsymbol{g}, \boldsymbol{x} \rangle \quad \forall \boldsymbol{x} \in S$$

Proof.

- Define a potential function $f(\boldsymbol{x}) \coloneqq \|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2$ for $\boldsymbol{x} \in S$
- f is convex, and even strictly convex
- $x^\star \coloneqq \arg\min_{\boldsymbol{x} \in S} f(\boldsymbol{x})$ exists (by the compactness of S) and is unique (by the previous lemma)
- First-order optimality implies that $\langle \nabla f(\boldsymbol{x}^\star), \boldsymbol{x}^\star - \boldsymbol{x} \rangle = 2\langle \boldsymbol{x} - \boldsymbol{x}_0, \boldsymbol{x}^\star - \boldsymbol{x} \rangle \leq 0$ for any $\boldsymbol{x} \in S$

Convexity: Geometric Properties

Theorem (Separating Hyperplane Theorem)**.** Let $S \subset \mathbb{R}^n$ be compact and convex and suppose $\boldsymbol{x}_0 \notin S$. There is a separating hyperplane $\boldsymbol{g} \in \mathbb{R}^n$ such that

$$\langle \boldsymbol{g}, \boldsymbol{x}_0 \rangle > \langle \boldsymbol{g}, \boldsymbol{x} \rangle \quad \forall \boldsymbol{x} \in S$$

Moreover, if $\boldsymbol{x}_0 \in S$ is a boundary point, there is a supporting hyperplane $\boldsymbol{g} \in \mathbb{R}^n, \boldsymbol{g} \neq \boldsymbol{0}$ such that

$$\langle \boldsymbol{g}, \boldsymbol{x}_0 \rangle \geq \langle \boldsymbol{g}, \boldsymbol{x} \rangle \quad \forall \boldsymbol{x} \in S$$

Proof.

- First-order optimality implies that $\langle \nabla f(\boldsymbol{x}^\star), \boldsymbol{x}^\star - \boldsymbol{x} \rangle = 2\langle \boldsymbol{x}^\star - \boldsymbol{x}_0, \boldsymbol{x}^\star - \boldsymbol{x} \rangle \leq 0$ for any $\boldsymbol{x} \in S$
- Define $\boldsymbol{g} \coloneqq \boldsymbol{x}_0 - \boldsymbol{x}^\star \neq \boldsymbol{0}$.
- Then we have

$$\langle \boldsymbol{g}, \boldsymbol{x} \rangle \leq \langle \boldsymbol{g}, \boldsymbol{x}^\star \rangle = \langle \boldsymbol{g}, \boldsymbol{x}_0 \rangle + \langle \boldsymbol{g}, \boldsymbol{x}^\star - \boldsymbol{x}_0 \rangle = \langle \boldsymbol{g}, \boldsymbol{x}_0 \rangle - \|\boldsymbol{g}\|_2^2 < \langle \boldsymbol{g}, \boldsymbol{x}_0 \rangle$$

■

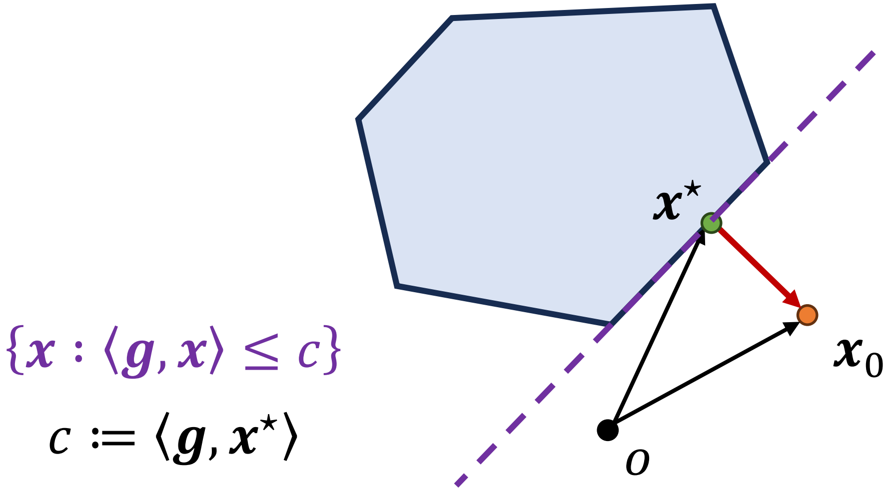
Convexity: Geometric Properties

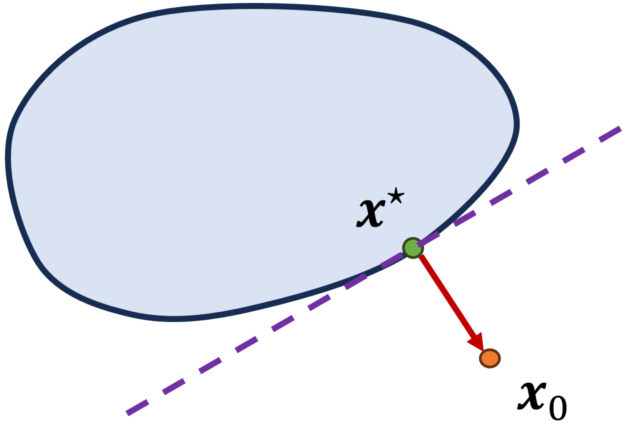
Theorem (Separating Hyperplane Theorem). Let $S \subset \mathbb{R}^n$ be compact and convex and suppose $\boldsymbol{x}_0 \notin S$. There is a separating hyperplane $\boldsymbol{g} \in \mathbb{R}^n$ such that

$$\langle \boldsymbol{g}, \boldsymbol{x}_0 \rangle > \langle \boldsymbol{g}, \boldsymbol{x} \rangle \quad \forall \boldsymbol{x} \in S$$

Moreover, if $\boldsymbol{x}_0 \in S$ is a boundary point, there is a supporting hyperplane $\boldsymbol{g} \in \mathbb{R}^n, \boldsymbol{g} \neq \boldsymbol{0}$ such that

$$\langle \boldsymbol{g}, \boldsymbol{x}_0 \rangle \geq \langle \boldsymbol{g}, \boldsymbol{x} \rangle \quad \forall \boldsymbol{x} \in S$$





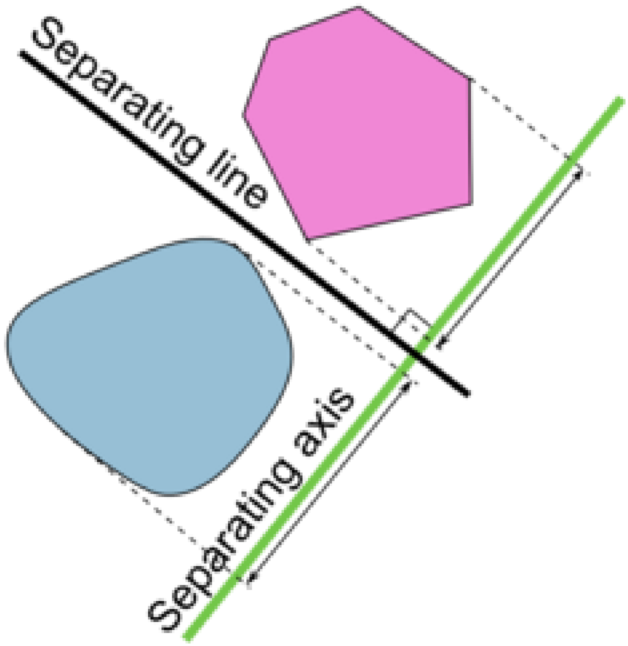
Convexity: Geometric Properties

Theorem (Separating Hyperplane Theorem, General form)**.** Let $A, B \subset \mathbb{R}^n$ be two disjoint convex subsets of $\mathbb{R}^n$. Then, there exist $\boldsymbol{g} \neq \boldsymbol{0} \in \mathbb{R}^n$ and $c \in \mathbb{R}$ such that

$$\langle \boldsymbol{g}, \boldsymbol{x} \rangle \geq c \geq \langle \boldsymbol{g}, \boldsymbol{y} \rangle \quad \forall \boldsymbol{x} \in A, \boldsymbol{y} \in B$$

Moreover, both A and B are closed, and at least one of them is compact, then there exist $c_1, c_2 \in \mathbb{R}$ such that

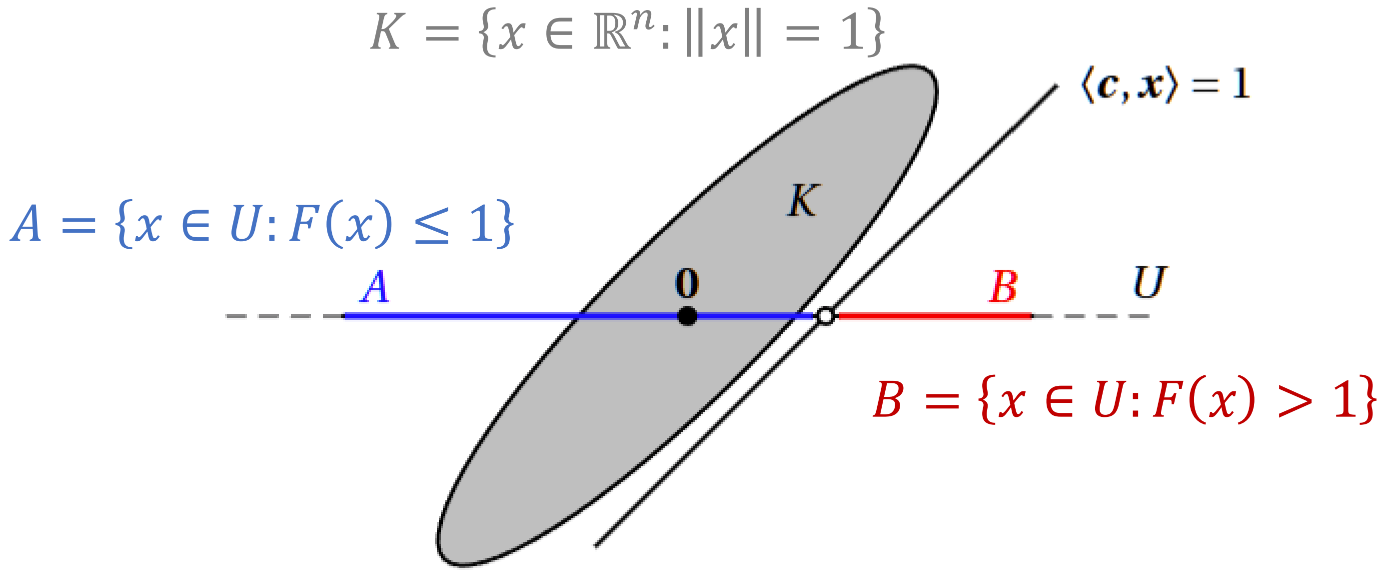
$$\langle \boldsymbol{g}, \boldsymbol{x} \rangle > c_1 > c_2 > \langle \boldsymbol{g}, \boldsymbol{y} \rangle \quad \forall \boldsymbol{x} \in A, \boldsymbol{y} \in B$$



Convexity: Geometric Properties

Theorem (Hahn-Banach Theorem). Let $(\mathbb{R}^n, \|\cdot\|)$ be a Banach space and let $U \subseteq \mathbb{R}^n$ be a subspace. Suppose $F: U \to \mathbb{R}$ is a linear function with $F(\boldsymbol{x}) \leq \|\boldsymbol{x}\|$ for all $\boldsymbol{x} \in U$. Then there exists a linear function $\tilde{F}: \mathbb{R}^n \to \mathbb{R}$ so that

- $\tilde{F}(\boldsymbol{x}) = F(\boldsymbol{x}) \ \forall \boldsymbol{x} \in U$
- $\tilde{F}(\boldsymbol{x}) \leq \|\boldsymbol{x}\| \ \forall \boldsymbol{x} \in \mathbb{R}^n$



Application of Separating Hyperplane Theorem

Primal LP

$$\max \boldsymbol{c}^\top \boldsymbol{x}$$
$$\text{s.t. } \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}$$
$$\boldsymbol{x} \in \mathbb{R}^n_{\ge 0}$$

Dual LP

$$\min \boldsymbol{b}^\top \boldsymbol{y}$$
$$\text{s.t. } \boldsymbol{A}^\top \boldsymbol{y} \ge \boldsymbol{c}$$
$$\boldsymbol{y} \in \mathbb{R}^m_{\ge 0}$$

Theorem (Duality Theorem for LPs). If P and D are a primal-dual pair of LPs, then one of the four cases happens:

1. Both are infeasible
2. P is unbounded and D is infeasible
3. D is unbounded and P is infeasible
4. Both are feasible and there exist optimal solutions x, y to P and D such that $c^\top x = b^\top y$

Application of Separating Hyperplane Theorem

Lemma (Farkas). Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. Then exactly one of the following two cases holds:

1. There exists $\boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ and $\boldsymbol{x} \geq \boldsymbol{0}$
2. There exists $\boldsymbol{y} \in \mathbb{R}^m$ such that $\boldsymbol{A}^\top \boldsymbol{y} \geq \boldsymbol{0}$ and $\boldsymbol{b}^\top \boldsymbol{y} < 0$

Proof.

- Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be the columns of $\boldsymbol{A}$
- Define the convex cone:

$$C \coloneqq \{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \geq \boldsymbol{0}\} = \left\{\sum_{i=1}^{n} x_i \boldsymbol{a}_i : x_i \geq 0\right\}$$

- Case 1 is equivalent to "$\boldsymbol{b} \in C$?"
- Suppose Case 1 is false, i.e., $\boldsymbol{b} \notin C$, we'll prove that Case 2 must hold

Application of Separating Hyperplane Theorem

Lemma (Farkas). Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. Then exactly one of the following two cases holds:

1. There exists $\boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{Ax} = \boldsymbol{b}$ and $\boldsymbol{x} \geq \mathbf{0}$
2. There exists $\boldsymbol{y} \in \mathbb{R}^m$ such that $\boldsymbol{A}^\top \boldsymbol{y} \geq \mathbf{0}$ and $\boldsymbol{b}^\top \boldsymbol{y} < 0$

Proof.

- By the separating hyperplane theorem, there exists $\boldsymbol{y} \neq \mathbf{0}$ and $\alpha \in \mathbb{R}$ such that
$$\langle \boldsymbol{y}, \boldsymbol{z} \rangle \geq \alpha > \langle \boldsymbol{y}, \boldsymbol{b} \rangle \quad \forall \boldsymbol{z} \in C$$
- Since $\mathbf{0} \in C$, we have $\alpha \leq 0$ and $\boldsymbol{b}^\top \boldsymbol{y} < 0$
- Since $\lambda \boldsymbol{a}_i \in C$ for any $\lambda > 0$ and $i \in [n]$, we have $\lambda \boldsymbol{a}_i^\top \boldsymbol{y} \geq \alpha$
- Thus, $\boldsymbol{a}_i^\top \boldsymbol{y} \geq 0$. That is, $\boldsymbol{A}^\top \boldsymbol{y} \geq \mathbf{0}$
- It remains to show that Cases 1 & 2 cannot be true at the same time (exercise)

Convexity: Subgradient

Let $f: \mathcal{X} \to \mathbb{R}$. We say $\boldsymbol{g}$ is a subgradient of f at $\boldsymbol{x} \in \mathcal{X}$ if

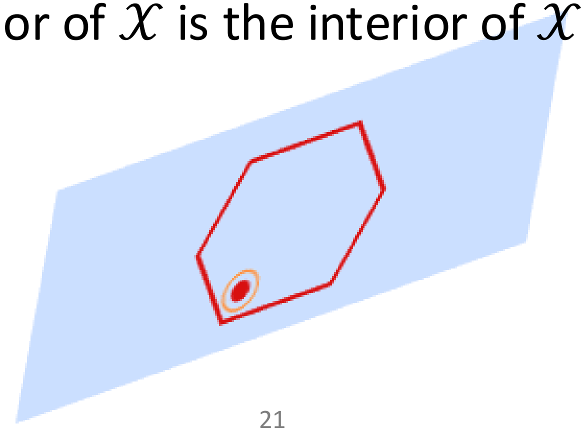
$$f(\boldsymbol{x}') \geq f(\boldsymbol{x}) + \langle \boldsymbol{g}, \boldsymbol{x}' - \boldsymbol{x} \rangle \quad \forall \boldsymbol{x}' \in \mathcal{X}$$

We denote the set of subgradients of f at x by $\partial f(\boldsymbol{x})$

- If f is convex, then ∇f is a subgradient everywhere

Lemma. Let $f: \mathcal{X} \to \mathbb{R}$ be convex and suppose $\mathcal{X} \subseteq \mathbb{R}^n$. For all $\boldsymbol{x} \in \mathrm{relint}(\mathcal{X})$, the relative interior of $\mathcal{X}$, $\partial f(\boldsymbol{x}) \neq \emptyset$

- $\boldsymbol{x}$ is in the interior of $\mathcal{X}$ if there is an open neighborhood of x contained in $\mathcal{X}$. When $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{X}$ is not full-dimensional, the relative interior of $\mathcal{X}$ is the interior of $\mathcal{X}$ within the smallest subspace containing it



Convexity: Subgradient

Let $f: \mathcal{X} \to \mathbb{R}$. We say $\boldsymbol{g}$ is a subgradient of f at $\boldsymbol{x} \in \mathcal{X}$ if

$$f(\boldsymbol{x}') \geq f(\boldsymbol{x}) + \langle \boldsymbol{g}, \boldsymbol{x}' - \boldsymbol{x} \rangle \quad \forall \boldsymbol{x}' \in \mathcal{X}$$

We denote the set of subgradients of f at x by $\partial f(\boldsymbol{x})$

Lemma. Let $f: \mathcal{X} \to \mathbb{R}$ be convex and suppose $\mathcal{X} \subseteq \mathbb{R}^n$. For all $\boldsymbol{x} \in \mathrm{relint}(\mathcal{X})$, the relative interior of $\mathcal{X}$, $\partial f(\boldsymbol{x}) \neq \emptyset$

Proof.

- Since $(\boldsymbol{x}, f(\boldsymbol{x}))$ lies on the boundary of $\mathrm{epi}(f)$, by the hyperplane separation theorem, there exists $(\boldsymbol{a}, b) \neq \boldsymbol{0} \in \mathbb{R}^n \times \mathbb{R}$ such that $\langle \boldsymbol{a}, \boldsymbol{x} \rangle + bf(\boldsymbol{x}) \geq \langle \boldsymbol{a}, \boldsymbol{y} \rangle + bt$ for any $(\boldsymbol{y}, t) \in \mathrm{epi}(f)$
- Since t can be arbitrarily large, $b \leq 0$. We claim that $b \neq 0$
- If $\boldsymbol{a} = \boldsymbol{0}$, then $b \neq 0$ as otherwise $(\boldsymbol{a}, b) = \boldsymbol{0}$
- If $\boldsymbol{a} \neq 0$, we can take $\boldsymbol{y} := \boldsymbol{x} + \epsilon \boldsymbol{a} \in \mathcal{X}$ for a sufficiently small $\epsilon > 0$ (since $\boldsymbol{x} \in \mathrm{relint}(\mathcal{X})$)
- $\langle \boldsymbol{a}, \boldsymbol{x} \rangle + bf(\boldsymbol{x}) \geq \langle \boldsymbol{a}, \boldsymbol{y} \rangle + bt = \langle \boldsymbol{a}, \boldsymbol{x} \rangle + \epsilon \|\boldsymbol{a}\|_2^2 + bt \Longrightarrow bf(\boldsymbol{x}) > bt \Longrightarrow b < 0$

Convexity: Subgradient

Let $f: \mathcal{X} \to \mathbb{R}$. We say $\boldsymbol{g}$ is a subgradient of f at $\boldsymbol{x} \in \mathcal{X}$ if

$$f(\boldsymbol{x}') \geq f(\boldsymbol{x}) + \langle \boldsymbol{g}, \boldsymbol{x}' - \boldsymbol{x} \rangle \quad \forall \boldsymbol{x}' \in \mathcal{X}$$

We denote the set of subgradients of f at x by $\partial f(\boldsymbol{x})$

Lemma. Let $f: \mathcal{X} \to \mathbb{R}$ be convex and suppose $\mathcal{X} \subseteq \mathbb{R}^n$. For all $\boldsymbol{x} \in \mathrm{relint}(\mathcal{X})$, the relative interior of $\mathcal{X}$, $\partial f(\boldsymbol{x}) \neq \emptyset$

Proof.

- $\langle \boldsymbol{a}, \boldsymbol{x} \rangle + b f(\boldsymbol{x}) \geq \langle \boldsymbol{a}, \boldsymbol{y} \rangle + bt$ for any $(\boldsymbol{y}, t) \in \mathrm{epi}(f)$ and $b < 0$
- Since $\left(\boldsymbol{x}', f(\boldsymbol{x}')\right) \in \mathrm{epi}(f)$,

$$\langle \boldsymbol{a}, \boldsymbol{x} \rangle + b f(\boldsymbol{x}) \geq \langle \boldsymbol{a}, \boldsymbol{x}' \rangle + b f(\boldsymbol{x}') \quad \Longrightarrow \quad f(\boldsymbol{x}') \geq f(\boldsymbol{x}) + \left\langle -\frac{\boldsymbol{a}}{b}, \boldsymbol{x}' - \boldsymbol{x} \right\rangle$$

- Thus, $\boldsymbol{g} := -\boldsymbol{a}/b \in \partial f(\boldsymbol{x})$

$\blacksquare$

Convexity: Convex Hull

For $A \subseteq \mathbb{R}^n$, the convex hull of A is the set of all convex combinations of points from A:

$$\operatorname{conv}(A) \coloneqq \left\{ \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i : k \geq 1, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in A, \lambda_i \geq 0, \sum_{i=1}^{k} \lambda_i = 1 \right\}$$

Theorem (Carathéodory)**.** Let $A \subseteq \mathbb{R}^n$ and let $x \in \operatorname{conv}(A)$. Then x is the convex combination of at most $n + 1$ points in A

Convexity: Convex Hull

Theorem (Carathéodory). Let $A \subseteq \mathbb{R}^n$ and let $x \in \operatorname{conv}(A)$. Then x is the convex combination of at most $n+1$ points in A

Proof.

- Attach an additional coordinate to each vector, and we have

$$\begin{pmatrix} \boldsymbol{x} \\ 1 \end{pmatrix} = \sum_{i=1}^{k} \lambda_i \begin{pmatrix} \boldsymbol{a}_i \\ 1 \end{pmatrix}, \qquad \boldsymbol{a}_i \in A, \lambda_i > 0, \sum_{i=1}^{k} \lambda_i = 1$$

- Let k be the smallest possible for which the above equation holds, and $k > n+1$
- Then, the k vectors in $\mathbb{R}^{n+1}$ are linearly-dependent:

$$\exists \mu_i \in \mathbb{R} \quad \text{s.t.} \quad \sum_{i=1}^{k} \mu_i \begin{pmatrix} \boldsymbol{a}_i \\ 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ 0 \end{pmatrix}$$

Convexity: Convex Hull

Theorem (Carathéodory). Let $A \subseteq \mathbb{R}^n$ and let $x \in \mathrm{conv}(A)$. Then x is the convex combination of at most $n+1$ points in A

Proof.

- Thus, for any $t \in \mathbb{R}$,

$$\begin{pmatrix} \boldsymbol{x} \\ 1 \end{pmatrix} = \sum_{i=1}^{k} (\lambda_i + t\mu_i) \begin{pmatrix} \boldsymbol{a}_i \\ 1 \end{pmatrix}, \qquad \boldsymbol{a}_i \in A, \lambda_i > 0, \sum_{i=1}^{k} \lambda_i = 1, \mu_i \in \mathbb{R}$$

- Since all $\lambda_i > 0$, by choosing t sufficiently small we can make at least one coefficient $\lambda_j = 0$ while keeping all others positive
- This contradicts the minimality of k

∎

Today’s Lecture

- Why Non-Convex Optimization Is Hard
- Convexity
- Cutting-Plane Method
- Log-concavity
- Brunn-Minkowski Inequality

Cutting-plane methods

Theorem (Polynomial-time convex optimization). Let $f: \mathcal{X} \to \mathbb{R}$ be convex for $\mathcal{X} \subset \mathbb{R}^n$ and assume the range $\sup f - \inf f$ bounded by $\mathrm{poly}(n)$. There is an algorithm which uses $\mathcal{O}\left(n \log\left(\frac{n}{\epsilon}\right)\right)$ queries to an evaluation and subgradient oracle for f, and $\mathrm{poly}\left(n, \log\frac{1}{\epsilon}\right)$ additional time, such that with high probability, the algorithm returns $\boldsymbol{x}$ satisfying

$$f(\boldsymbol{x}) \leq \min_{\boldsymbol{x}^\star \in \mathcal{X}} f(\boldsymbol{x}^\star) + \epsilon$$

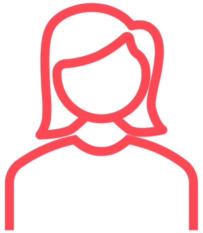
Cutting-plane methods

Game theory perspective of cutting-plane methods

Turn t (starting from $t = 0$):

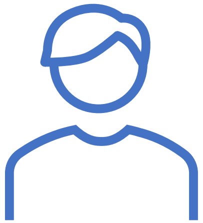
If $\mathrm{Vol}_n(S_t) < \mathrm{V}_{\min}$, the game ends

Alice



Alice choose $\boldsymbol{x}_t$ and send to Bob

Bob



If $x_t \in S^\star$, the game ends

Bob choose $\boldsymbol{g}_t \in \mathbb{R}^n$ s.t. $\langle \boldsymbol{g}_t, \boldsymbol{x}_t \rangle > \langle \boldsymbol{g}_t, \boldsymbol{x} \rangle \ \forall \boldsymbol{x} \in S^\star$ and send to Alice

$S_0 \supseteq S^\star$

Alice update $S_{t+1} \supseteq S_t \cap H_t$ where
$H_t \coloneqq \{x \in \mathbb{R}^d : \langle \boldsymbol{g}_t, \boldsymbol{x} \rangle < \langle \boldsymbol{g}_t, \boldsymbol{x}_t \rangle\}$

Compact, convex
$S^\star \subset \mathbb{R}^n$

Cutting-plane methods

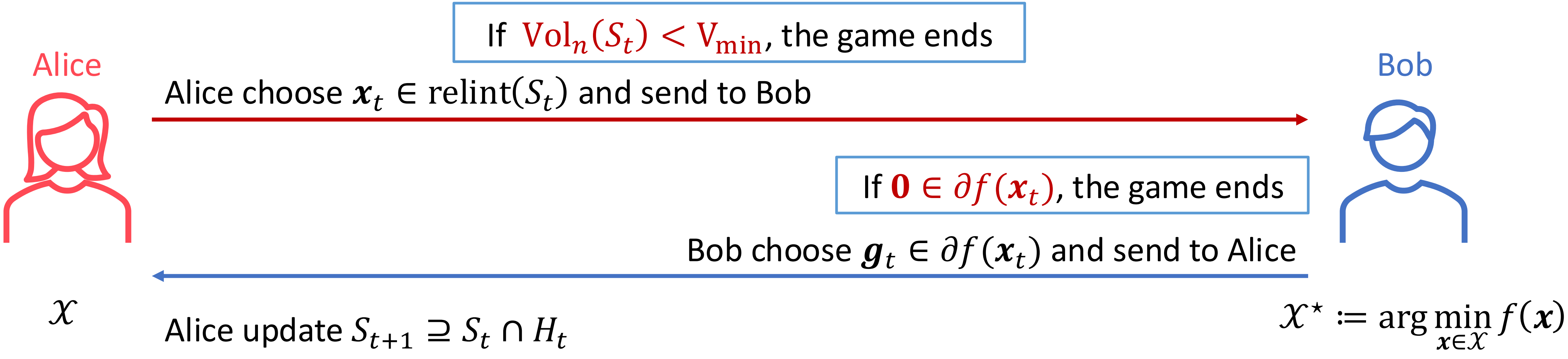
Lemma. Let $f: \mathcal{X} \to \mathbb{R}$ be convex for $\mathcal{X} \subset \mathbb{R}^n$, and suppose f has minimizer set $\mathcal{X}^\star \coloneqq \arg\min_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x})$.

Suppose we play the cutting-plane game initialized from $S_0 \leftarrow \mathcal{X}$, and let $\alpha \coloneqq \left(\frac{\mathrm{V}_{\min}}{\mathrm{Vol}_n(S_0)}\right)^{\frac{1}{n}}$.

Further, suppose Alice always chooses $\boldsymbol{x}_t \in \mathrm{relint}(S_t)$, and Bob (who holds $S^\star \leftarrow \mathcal{X}^\star$) plays by ending the game if $\mathbf{0} \in \partial f(\boldsymbol{x}_t)$, and returning $\boldsymbol{g}_t \in \partial f(\boldsymbol{x}_t)$ otherwise.

If the game terminates in T iterations, letting $f^\star \coloneqq \min f(\boldsymbol{x})$, we have

$$\min_{t \in [T]} f(\boldsymbol{x}_t) \leq f^\star + \alpha \left(\max_{\boldsymbol{z} \in \mathcal{X}} f(\boldsymbol{z}) - f^\star \right)$$



Cutting-plane methods

Lemma. Let $f: \mathcal{X} \to \mathbb{R}$ be convex for $\mathcal{X} \subset \mathbb{R}^n$, and suppose f has minimizer set $\mathcal{X}^\star \coloneqq \arg\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$.

Suppose we play the cutting-plane game initialized from $S_0 \leftarrow \mathcal{X}$, and let $\alpha \coloneqq \left(\frac{\mathrm{V}_{\min}}{\mathrm{Vol}_n(S_0)}\right)^{\frac{1}{n}}$.

Further, suppose Alice always chooses $\boldsymbol{x}_t \in \mathrm{relint}(S_t)$, and Bob (who holds $S^\star \leftarrow \mathcal{X}^\star$) plays by ending the game if $\boldsymbol{0} \in \partial f(\boldsymbol{x}_t)$, and returning $\boldsymbol{g}_t \in \partial f(\boldsymbol{x}_t)$ otherwise.

If the game terminates in T iterations, letting $f^\star \coloneqq \min f(\boldsymbol{x})$, we have

$$\min_{t\in[T]} f(\boldsymbol{x}_t) \leq f^\star + \alpha\left(\max_{\boldsymbol{z}\in\mathcal{X}} f(\boldsymbol{z}) - f^\star\right)$$

Proof.

- $\boldsymbol{x}_t \in \mathcal{X}^\star$ is equivalent to:
  $$f(\boldsymbol{x}_t) \leq f(\boldsymbol{x}) \;\; \forall \boldsymbol{x} \in \mathcal{X} \iff f(\boldsymbol{x}_t) + \langle \boldsymbol{0}, \boldsymbol{x} - \boldsymbol{x}_t\rangle \leq f(\boldsymbol{x}) \;\; \forall \boldsymbol{x} \in \mathcal{X} \iff \boldsymbol{0} \in \partial f(\boldsymbol{x}_t)$$
- If $\boldsymbol{x}_t \notin \mathcal{X}^\star$, then for any $\boldsymbol{x}^\star \in \mathcal{X}^\star$, we have
  $$0 > f(\boldsymbol{x}^\star) - f(\boldsymbol{x}_t) \geq \langle \boldsymbol{g}_t, \boldsymbol{x}^\star - \boldsymbol{x}_t\rangle \iff \langle \boldsymbol{g}_t, \boldsymbol{x}_t\rangle > \langle \boldsymbol{g}_t, \boldsymbol{x}^\star\rangle$$
  Thus, Bob's choice is valid

Cutting-plane methods

Lemma. Let $f: \mathcal{X} \to \mathbb{R}$ be convex for $\mathcal{X} \subset \mathbb{R}^n$, and suppose f has minimizer set $\mathcal{X}^\star \coloneqq \arg\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$.

Suppose we play the cutting-plane game initialized from $S_0 \leftarrow \mathcal{X}$, and let $\alpha \coloneqq \left(\frac{\mathrm{V}_{\min}}{\mathrm{Vol}_n(S_0)}\right)^{\frac{1}{n}}$.

Further, suppose Alice always chooses $\boldsymbol{x}_t \in \mathrm{relint}(S_t)$, and Bob (who holds $S^\star \leftarrow \mathcal{X}^\star$) plays by ending the game if $\boldsymbol{0} \in \partial f(\boldsymbol{x}_t)$, and returning $\boldsymbol{g}_t \in \partial f(\boldsymbol{x}_t)$ otherwise.

If the game terminates in T iterations, letting $f^\star \coloneqq \min f(\boldsymbol{x})$, we have

$$\min_{t\in[T]} f(\boldsymbol{x}_t) \le f^\star + \alpha\left(\max_{z\in\mathcal{X}} f(\boldsymbol{z}) - f^\star\right)$$

Proof.

- For any $x^\star \in \mathcal{X}^\star$, consider the set $S_\alpha \coloneqq \{\boldsymbol{x} = (1-\alpha)\boldsymbol{x}^\star + \alpha\boldsymbol{z} : \boldsymbol{z} \in \mathcal{X}\} = (1-\alpha)\boldsymbol{x}^\star + \alpha\mathcal{X}$
- We have $\mathrm{Vol}_n(S_\alpha) = \alpha^n \mathrm{Vol}_n(\mathcal{X}) = \alpha^n \mathrm{Vol}_n(S_0) = \mathrm{V}_{\min}$
- If $\mathrm{Vol}_n(S_T) < \mathrm{V}_{\min}$, then $\exists \boldsymbol{x} \in S_\alpha \backslash S_T$ such that $\boldsymbol{x} = (1-\alpha)\boldsymbol{x}^\star + \alpha\boldsymbol{z}$
- Then, $\exists t \in [T]$ such that $\boldsymbol{x} \notin H_t$

Cutting-plane methods

Lemma. Let $f: \mathcal{X} \to \mathbb{R}$ be convex for $\mathcal{X} \subset \mathbb{R}^n$, and suppose f has minimizer set $\mathcal{X}^\star \coloneqq \arg\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$.

Suppose we play the cutting-plane game initialized from $S_0 \leftarrow \mathcal{X}$, and let $\alpha \coloneqq \left(\frac{\mathrm{V_{min}}}{\mathrm{Vol}_n(S_0)}\right)^{\frac{1}{n}}$.

Further, suppose Alice always chooses $\boldsymbol{x}_t \in \mathrm{relint}(S_t)$, and Bob (who holds $S^\star \leftarrow \mathcal{X}^\star$) plays by ending the game if $\mathbf{0} \in \partial f(\boldsymbol{x}_t)$, and returning $\boldsymbol{g}_t \in \partial f(\boldsymbol{x}_t)$ otherwise.

If the game terminates in T iterations, letting $f^\star \coloneqq \min f(\boldsymbol{x})$, we have

$$\min_{t\in[T]} f(\boldsymbol{x}_t) \le f^\star + \alpha\left(\max_{\boldsymbol{z}\in\mathcal{X}} f(\boldsymbol{z}) - f^\star\right)$$

Proof.

- Then, $\exists t \in [T]$ such that $\boldsymbol{x} \notin H_t$, which implies that

$$f(\boldsymbol{x}) \ge f(\boldsymbol{x}_t) + \langle \boldsymbol{g}_t, \boldsymbol{x} - \boldsymbol{x}_t\rangle \ge f(\boldsymbol{x}_t)$$

- Thus,

$$f(\boldsymbol{x}_t) \le f(\boldsymbol{x}) \le (1-\alpha) f(\boldsymbol{x}^\star) + \alpha f(\boldsymbol{z}) \le f^\star + \alpha\left(\max_{\boldsymbol{z}\in\mathcal{X}} f(\boldsymbol{z}) - f^\star\right)$$

■

Cutting-plane methods

Lemma. Let $f: \mathcal{X} \to \mathbb{R}$ be convex for $\mathcal{X} \subset \mathbb{R}^n$, and suppose f has minimizer set $\mathcal{X}^\star \coloneqq \arg\min_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x})$.

Suppose we play the cutting-plane game initialized from $S_0 \leftarrow \mathcal{X}$, and let $\alpha \coloneqq \left(\frac{\mathrm{V_{min}}}{\mathrm{Vol}_n(S_0)}\right)^{\frac{1}{n}}$.

Further, suppose Alice always chooses $\boldsymbol{x}_t \in \mathrm{relint}(S_t)$, and Bob (who holds $S^\star \leftarrow \mathcal{X}^\star$) plays by ending the game if $\mathbf{0} \in \partial f(\boldsymbol{x}_t)$, and returning $\boldsymbol{g}_t \in \partial f(\boldsymbol{x}_t)$ otherwise.

If the game terminates in T iterations, letting $f^\star \coloneqq \min f(\boldsymbol{x})$, we have

$$\min_{t \in [T]} f(\boldsymbol{x}_t) \leq f^\star + \alpha \left(\max_{\boldsymbol{z} \in \mathcal{X}} f(\boldsymbol{z}) - f^\star \right)$$

To implement the cutting-plane game / cutting-plane method, we need:

- Evaluation oracle: $O_f(\boldsymbol{x}) = f(\boldsymbol{x})$ if $\boldsymbol{x} \in \mathcal{X}$ and ∞ otherwise
- Subgradient oracle: $O_{\partial f}(\boldsymbol{x})$ returns an element in $\partial f(\boldsymbol{x})$ (if $\mathbf{0} \in \partial f(\boldsymbol{x})$ returns $\mathbf{0}$ by default)
- To obtain $\min_{t \in [T]} f(x_t) \leq f^\star + \epsilon$, let R be the range of f and we can take $\alpha \coloneqq \frac{\epsilon}{R}$

Cutting-plane methods

- To obtain $\min_{t\in[T]} f(x_t) \le f^\star + \epsilon$, let R be the range of f and we can take $\alpha := \frac{\epsilon}{R}$
- By assumption, $R = \mathrm{poly}(n)$, and we need

$$\frac{\mathrm{Vol}_n(S_T)}{\mathrm{Vol}_n(S_0)} \le \alpha^n = \frac{\epsilon^n}{n^{\Omega(n)}}$$

Theorem (Grünbaum). Let $S \subseteq \mathbb{R}^n$ be convex and let $\overline{\boldsymbol{x}}_S$ be the barycenter of S. Then any halfspace

$$H = \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{g}, \boldsymbol{x}\rangle \le \langle \boldsymbol{g}, \overline{\boldsymbol{x}}_S\rangle\}$$

whose defining halfplane passes through $\overline{\boldsymbol{x}}_S$ satisfies $\mathrm{Vol}_n(S \cap H) \ge \frac{1}{e}\mathrm{Vol}_n(S)$

- If Alice choose $\boldsymbol{x}_t = \overline{\boldsymbol{x}}_S$ and $S_{t+1} = S_t \cap H_t$, then $\mathrm{Vol}_n(S_{t+1}) \le \left(1 - \frac{1}{e}\right)\mathrm{Vol}_n(S_t)$
- Thus, it suffices to take $T = \mathcal{O}\left(n \log \frac{n}{\epsilon}\right)$

Cutting-plane methods

Theorem (Polynomial-time convex optimization). Let $f: \mathcal{X} \to \mathbb{R}$ be convex for $\mathcal{X} \subset \mathbb{R}^n$ and assume the range $\sup f - \inf f$ bounded by $\mathrm{poly}(n)$. There is an algorithm which uses $\mathcal{O}\left(n \log\left(\frac{n}{\epsilon}\right)\right)$ queries to an evaluation and subgradient oracle for f, and $\mathrm{poly}\left(n, \log\frac{1}{\epsilon}\right)$ additional time, such that with high probability, the algorithm returns $\boldsymbol{x}$ satisfying

$$f(\boldsymbol{x}) \le \min_{\boldsymbol{x}^\star \in \mathcal{X}} f(\boldsymbol{x}^\star) + \epsilon$$

- CPM can be viewed as a "binary search" in high dimensions
- We only proved the query complexity. For polynomial-time implementation, we need to estimate the barycenter by uniformly sampling points inside S_t (a polytope)
- The ellipsoid method is much easier to implement in each iteration, but it needs $\tilde{\mathcal{O}}(n^2)$ iterations
- The state-of-the-art CPM implementation is by (Jiang-Lee-Song-Wong '20)

Today's Lecture

- Why Non-Convex Optimization Is Hard
- Convexity
- Cutting-Plane Method
- Log-concavity
- Brunn-Minkowski Inequality

Log-concavity

A function $F: \mathbb{R}^n \to \mathbb{R}$ is called log-concave if one of the following equivalent conditions hold:

- $F(\boldsymbol{x}) = \exp(-G(\boldsymbol{x}))$ for some convex function $G: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$
- $\ln(F(\boldsymbol{x}))$ is concave
- $F(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \geq F(\boldsymbol{x})^{\lambda} \cdot F(\boldsymbol{y})^{1-\lambda}$ for all $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$

- Log-concave functions frequently appear in sampling where the probability density functions are log-concave.
- Intuitively, just as convex optimization being easy since convexity can rule out disjoint sets of minimizers, log-concave sampling is easy since log-concavity can rule out disjoint modes
- We often say $\mu(\boldsymbol{x}) \propto F(\boldsymbol{x})$ is a log-concave distribution, which means the normalizing constant is hidden:

$$\mu(\boldsymbol{x}) = \frac{F(\boldsymbol{x})}{Z}, \qquad Z := \int F(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

Log-concavity

A function $F: \mathbb{R}^n \to \mathbb{R}$ is called log-concave if one of the following equivalent conditions hold:

- $F(\boldsymbol{x}) = \exp\big(-G(\boldsymbol{x})\big)$ for some convex function $G: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$
- $\ln\big(F(\boldsymbol{x})\big)$ is concave
- $F(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \geq F(\boldsymbol{x})^{\lambda} \cdot F(\boldsymbol{y})^{1-\lambda}$ for all $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$

Examples:

- $\mu(\boldsymbol{x}) \propto \exp(-\|\boldsymbol{x}\|_2^2)$, i.e., multivariate Gaussian distribution
- $\mu(\boldsymbol{x}) \propto \mathbf{1}[\boldsymbol{x} \in S]$ for a convex set S, i.e., uniform distribution over a convex body
- $F: \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ concave, then F is also log-concave
- (Nontrivial) Let K be a convex set, $\theta \in \mathbb{R}^n$, define $G(t) \coloneqq \mathrm{Vol}_n(K \cap \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle \leq t\})$. Then $G(t)^{1/n}$ is concave and $G(t)$ is log-concave

Log-concavity

Lemma (1d log-concavity). Let $G: \mathbb{R} \to \mathbb{R}_{>0}$ be log concave. Then for $t^\star \in \mathbb{R}$ we have

$$G(t) \leq G(t^\star) \exp\left((t - t^\star) \cdot \frac{G'(t^\star)}{G(t^\star)} \right) \qquad \forall t \in \mathbb{R}$$

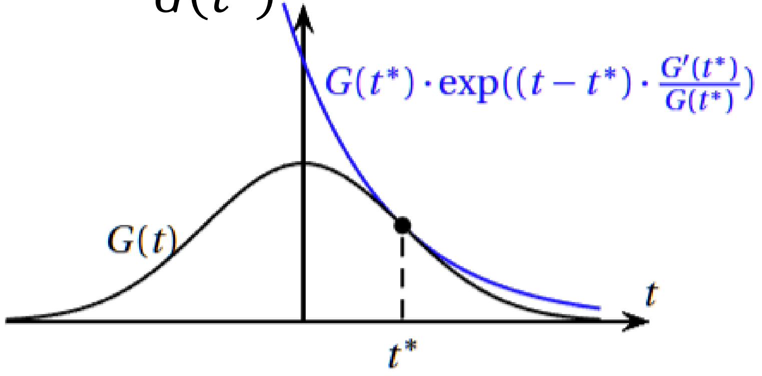
➢ If the derivative of a log-concave function < 0 at any point, then it must decay at an exponential rate

Proof.

- Since $\ln G$ is concave, we have

$$\ln G(t) \leq \ln G(t^\star) + (t - t^\star) \cdot (\ln G(t^\star))' = \ln G(t^\star) + (t - t^\star) \cdot \frac{G'(t^\star)}{G(t^\star)}$$

- Exponentiating both sides proves the lemma



Log-concavity

Theorem (Grünbaum). Let $K \subseteq \mathbb{R}^n$ be convex with $\mathrm{Vol}_n(K) = 1$ and let $\mathbf{0}$ be the barycenter of K. Then any halfspace $H = \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{g}, \boldsymbol{x} \rangle \leq 0\}$ satisfies $\mathrm{Vol}_n(K \cap H) \geq 1/e$

Proof.

- We can scale $\boldsymbol{g}$ so that for any $x \in K$, $\langle \boldsymbol{g}, \boldsymbol{x} \rangle \in [-1,1]$
- Define $G(t) \coloneqq \mathrm{Vol}_n(K \cap \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{g}, \boldsymbol{x} \rangle \leq t\})$. Then G is log-concave (see the nontrivial example)
- $G(-1) = 0$, $G(1) = 1$, and $G(0) = \mathrm{Vol}_n(K \cap H)$
- $G'(t) = \mathrm{Vol}_{n-1}(K \cap \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{g}, \boldsymbol{x} \rangle = t\})$ (i.e. the area of the hyperplane)
- **Fact:** $\mathbf{0}$ is the barycenter of K iff $\int_{\mathbb{R}} t \cdot \mathrm{Vol}_{n-1}(\{\boldsymbol{x} \in K : \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle = t\})\mathrm{d}t = 0$ for any $\boldsymbol{\theta} \in \mathbb{R}^{n-1}$
- Thus,

$$0 = \int_{-1}^{1} tG'(t)\mathrm{d}t = tG(t)\Big|_{-1}^{1} - \int_{-1}^{1} G(t)\mathrm{d}t \quad \Longrightarrow \quad \int_{-1}^{1} G(t)\mathrm{d}t = 1$$

Log-concavity

Theorem (Grünbaum). Let $K \subseteq \mathbb{R}^n$ be convex with $\mathrm{Vol}_n(K) = 1$ and let $\mathbf{0}$ be the barycenter of K. Then any halfspace $H = \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{g}, \boldsymbol{x} \rangle \leq 0\}$ satisfies $\mathrm{Vol}_n(K \cap H) \geq 1/e$

Proof.

- Define $G(t) \coloneqq \mathrm{Vol}_n(K \cap \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{g}, \boldsymbol{x} \rangle \leq t\})$. Then G is log-concave (see the nontrivial example)
- $G(-1) = 0$, $G(1) = 1$, and $G(0) = \mathrm{Vol}_n(K \cap H)$

$$1 = \int_{-1}^{1} G(t)\mathrm{d}t \leq \int_{-1}^{1} \min\left\{G(0)\exp\left(t\frac{G'(0)}{G(0)}\right), 1\right\}\mathrm{d}t = \int_{-1}^{\alpha} G(0)\exp\left(t\frac{G'(0)}{G(0)}\right)\mathrm{d}t + (1-\alpha)$$

$$\leq G(0)\frac{G(0)}{G'(0)}\exp\left(t\frac{G'(0)}{G(0)}\right)\Bigg|_{-\infty}^{\alpha} + (1-\alpha) = \frac{G(0)^2}{G'(0)}\exp\left(\alpha\frac{G'(0)}{G(0)}\right) + (1-\alpha)$$

$$\alpha = -\frac{G(0)}{G'(0)}\ln G(0)$$

$$= \frac{G(0)}{G'(0)} + 1 + \frac{G(0)}{G'(0)}\ln G(0) \quad \Rightarrow \quad \frac{G(0)}{G'(0)}(1 + \ln G(0)) \geq 0 \quad \Rightarrow \quad G(0) \geq \frac{1}{e}$$

■

Today's Lecture

- Why Non-Convex Optimization Is Hard
- Convexity
- Cutting-Plane Method
- Log-concavity
- Brunn-Minkowski Inequality

Brunn-Minkowski Inequality

Theorem (Brunn-Minkowski, v1). Let $A, B \subseteq \mathbb{R}^n$ be non-empty compact sets. Then

$$\mathrm{Vol}_n(\lambda A + (1-\lambda)B)^{\frac{1}{n}} \geq \lambda \cdot \mathrm{Vol}_n(A)^{\frac{1}{n}} + (1-\lambda) \cdot \mathrm{Vol}_n(B)^{\frac{1}{n}} \quad \forall \lambda \in [0,1]$$

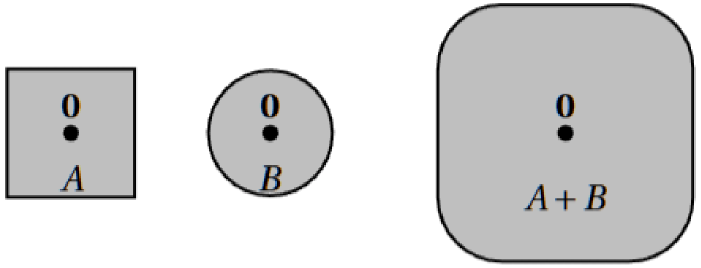
Theorem (Brunn-Minkowski, v2). Let $A, B \subseteq \mathbb{R}^n$ be non-empty compact sets. Then

$$\mathrm{Vol}_n(A+B)^{\frac{1}{n}} \geq \mathrm{Vol}_n(A)^{\frac{1}{n}} + \mathrm{Vol}_n(B)^{\frac{1}{n}}$$

Theorem (Brunn-Minkowski, v3). Let $A, B \subseteq \mathbb{R}^n$ be non-empty compact sets. Then

$$\mathrm{Vol}_n(\lambda A + (1-\lambda)B) \geq \mathrm{Vol}_n(A)^{\lambda} \cdot \mathrm{Vol}_n(B)^{1-\lambda} \quad \forall \lambda \in [0,1]$$

- $A + B \coloneqq \{\boldsymbol{a} + \boldsymbol{b} : \boldsymbol{a} \in A, \boldsymbol{b} \in B\}$
- Very powerful results since they still hold even if A and B are non-convex



Brunn-Minkowski Inequality

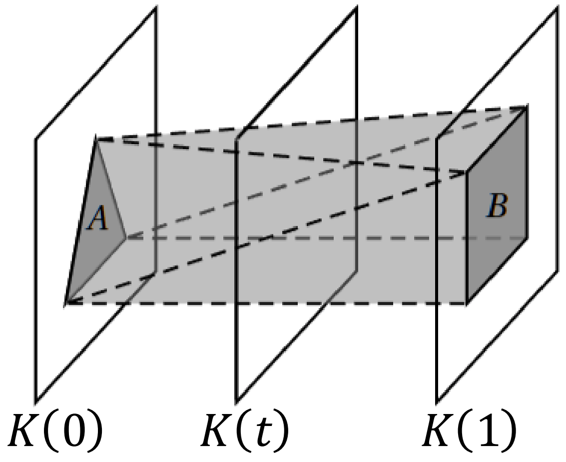
Theorem (Brunn-Minkowski, v0). Let $A, B \subseteq \mathbb{R}^n$ be non-empty compact convex sets. Then

$$\mathrm{Vol}_n(\lambda A + (1-\lambda)B)^{\frac{1}{n}} \geq \lambda \cdot \mathrm{Vol}_n(A)^{\frac{1}{n}} + (1-\lambda) \cdot \mathrm{Vol}_n(B)^{\frac{1}{n}} \quad \forall \lambda \in [0,1]$$

Proof.

- We increase the dimension from n to $n+1$, and embed A to the hyperplane $x_{n+1} = 0$, B to $x_{n+1} = 1$
- Let $K \coloneqq \mathrm{conv}\big((A \times \{0\}) \cup (B \times \{1\})\big)$, and $K(t) \coloneqq \{\boldsymbol{x} \in \mathbb{R}^n : (\boldsymbol{x}, t) \in K\}$ for $t \in [0,1]$
- $K(t) = (1-t) \cdot A + t \cdot B$
- If we can show that $t \mapsto \mathrm{Vol}_n\big(K(t)\big)$ is log-concave, then

$$\underbrace{\mathrm{Vol}_n\big(K(\lambda \cdot 0 + (1-\lambda) \cdot 1)\big)}_{\lambda A + (1-\lambda)B} \geq \mathrm{Vol}_n\big(\underbrace{K(0)}_{A}\big)^{\lambda} \cdot \mathrm{Vol}_n\big(\underbrace{K(1)}_{B}\big)^{1-\lambda}$$



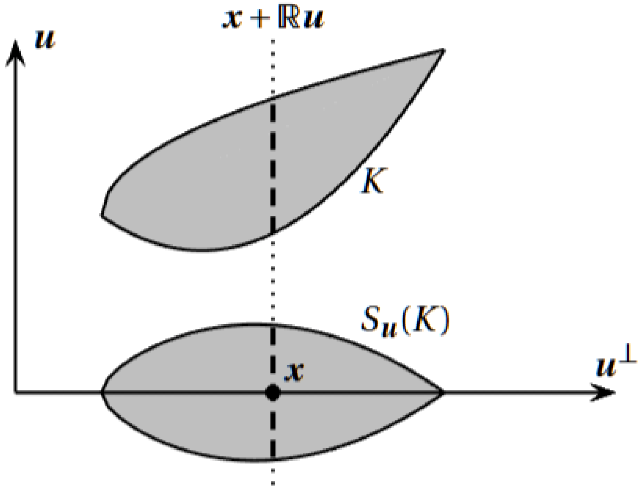
- For the additive form, take $A' \coloneqq \frac{A}{\mathrm{Vol}_n(A)^{\frac{1}{n}}}$, $B' \coloneqq \frac{B}{\mathrm{Vol}_n(B)^{\frac{1}{n}}}$, and $\lambda \coloneqq \frac{\mathrm{Vol}_n(A)^{\frac{1}{n}}}{\mathrm{Vol}_n(A)^{\frac{1}{n}} + \mathrm{Vol}_n(B)^{\frac{1}{n}}}$

Detour: Steiner Symmetrization

Let $K \subseteq \mathbb{R}^n$ be a convex body and $\boldsymbol{u} \in \mathbb{S}^{n-1}$ a unit direction. The Steiner symmetral $S_{\boldsymbol{u}}(K) \subseteq \mathbb{R}^n$ of K in direction $\boldsymbol{u}$ is defined so that for every $\boldsymbol{x} \in \boldsymbol{u}^\perp$,

$$\mathrm{Vol}_1\big((\boldsymbol{x} + \mathbb{R}\boldsymbol{u}) \cap K\big) = \mathrm{Vol}_1\big((\boldsymbol{x} + \mathbb{R}\boldsymbol{u}) \cap S_{\boldsymbol{u}}(K)\big)$$

where $(\boldsymbol{x} + \mathbb{R}\boldsymbol{u}) \cap S_{\boldsymbol{u}}(K)$ is an interval centered at $\boldsymbol{x}$



Properties:

- $S_{\boldsymbol{u}}(K)$ is symmetric w.r.t. the hyperplane $\boldsymbol{u}^\perp$
- K convex $\Longrightarrow S_{\boldsymbol{u}}(K)$ convex
- $\mathrm{Vol}_n\big(S_{\boldsymbol{u}}(K)\big) = \mathrm{Vol}_n(K)$
- Suppose $\mathrm{Vol}_n(K) = \mathrm{Vol}_n(rB_2^n)$. Then there exists a sequence of vectors $\boldsymbol{u}_i \in \mathbb{S}^{n-1}$ so that the sequence $K_i \coloneqq S_{\boldsymbol{u}_i}(K_{i-1})$ with $K_0 \coloneqq K$ converges to rB_2^n

Detour: Steiner Symmetrization

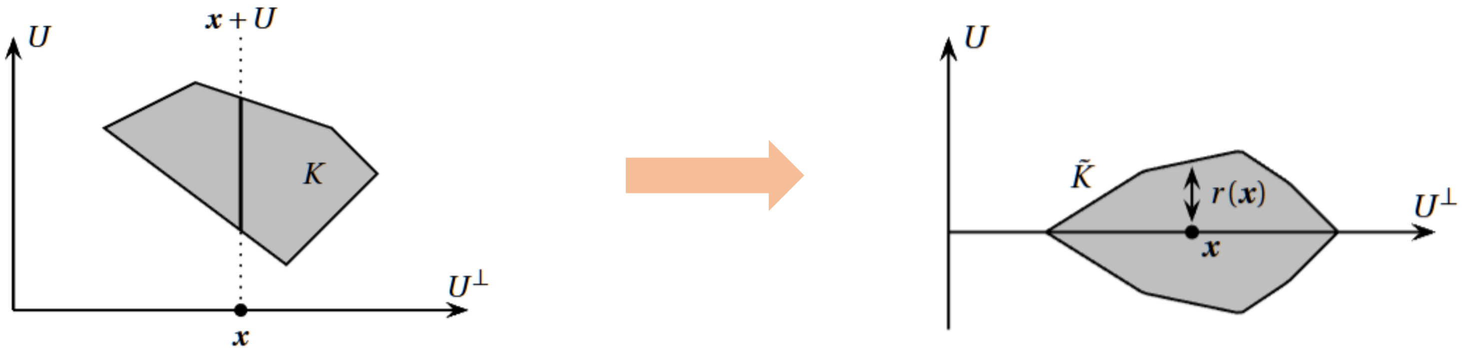
Theorem (Brunn's Concavity Principle)**.** Let $K \subseteq \mathbb{R}^n$ be a convex body and let $U \subseteq \mathbb{R}^n$ be a k-dimensional subspace. Then the function $F\colon U^\perp \to \mathbb{R}$ defined by $F(\boldsymbol{x}) \coloneqq \mathrm{Vol}_k\big(K \cap (U + \boldsymbol{x})\big)^{1/k}$ is concave on its support.

Proof.

- Apply the Steiner Symmetrization for all directions $\boldsymbol{u} \in U$ and let $\widetilde{K}$ be the limiting convex body
- $\widetilde{K} \cap (U + \boldsymbol{x})$ is a k-dimensional ball of radius $r(\boldsymbol{x})$ such that

$$r(\boldsymbol{x})^k \cdot \mathrm{Vol}_k\big(B_2^k\big) = \mathrm{Vol}_k\left(\widetilde{K} \cap (U + \boldsymbol{x})\right) = \mathrm{Vol}_k\big(K \cap (U + \boldsymbol{x})\big) = F(\boldsymbol{x}) \quad \forall x \in U^\perp$$

- K convex $\Longrightarrow \widetilde{K}$ convex $\Longrightarrow r(\boldsymbol{x})$ concave $\Longrightarrow F(x)$ concave

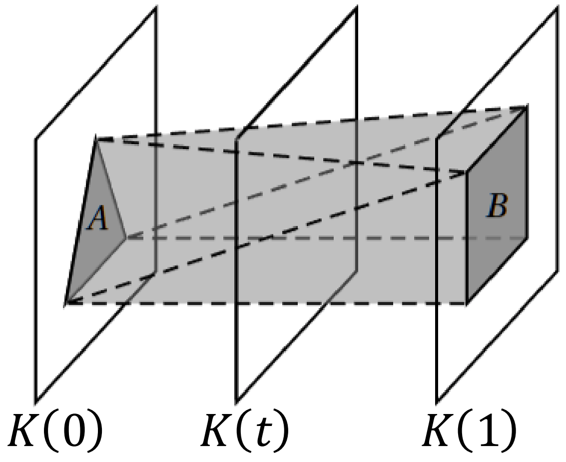


■

Brunn-Minkowski Inequality

Theorem (Brunn's Concavity Principle). Let $K \subseteq \mathbb{R}^n$ be a convex body and let $U \subseteq \mathbb{R}^n$ be a k-dimensional subspace. Then the function $F: U^\perp \to \mathbb{R}$ defined by $F(\boldsymbol{x}) \coloneqq \mathrm{Vol}_k\big(K \cap (U + \boldsymbol{x})\big)^{1/k}$ is concave on its support.

- We take $U = \{\boldsymbol{x} \in \mathbb{R}^{n+1} : x_{n+1} = 0\}$, and $k = n$
- Then, $t \mapsto \mathrm{Vol}_n\left(K \cap \left(U + (\mathbf{0}, t)\right)\right) = \mathrm{Vol}_n\big(K(t)\big)^{1/n}$ is concave
- By the property of log-concavity, $\mathrm{Vol}_n\big(K(t)\big)^{1/n}$ is also log-concave
- Thus, $\frac{1}{n} \ln \mathrm{Vol}_n\big(K(t)\big)$ concave $\Longrightarrow \mathrm{Vol}_n\big(K(t)\big)$ log-concave
- Therefore, the B-M inequality holds for convex A, B



■

Brunn-Minkowski Inequality

Theorem (Prékopa-Leindler Inequality). For $0 < \lambda < 1$, let $f, g, h \colon \mathbb{R}^n \to \mathbb{R}$ be measurable functions so that

$$h(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \geq f(\boldsymbol{x})^{\lambda} g(\boldsymbol{y})^{1-\lambda} \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

Then

$$\int_{\mathbb{R}^n} h(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \geq \left(\int_{\mathbb{R}^n} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right)^{\lambda} \cdot \left(\int_{\mathbb{R}^n} g(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right)^{1-\lambda}$$

Corollary. Let $\mu, \mu' \colon \mathbb{R}^n \to \mathbb{R}$ be log-concave.

1. Let $S \subseteq [n]$ and let $\mu_S(\boldsymbol{x}_S) := \int_{\mathbb{R}[d] \setminus S} \mu(\boldsymbol{x}_S, \boldsymbol{x}_{-S}) \mathrm{d}\boldsymbol{x}_{-S}$ be the marginal on S. Then μ_S is log-concave
2. Let $(\mu * \mu')(\boldsymbol{x}) = \int \mu(\boldsymbol{x} - \boldsymbol{y}) \mu'(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$ be the convolution of μ and μ'. Then $\mu * \mu'$ is log-concave

Brunn-Minkowski Inequality

Corollary. Let $\mu, \mu' : \mathbb{R}^n \to \mathbb{R}$ be log-concave.

1. Let $S \subseteq [n]$ and let $\mu_S(\boldsymbol{x}_S) := \int_{\mathbb{R}[d]\setminus S} \mu(\boldsymbol{x}_S, \boldsymbol{x}_{-S}) \mathrm{d}\boldsymbol{x}_{-S}$ be the marginal on S. Then μ_S is log-concave
2. Let $(\mu * \mu')(\boldsymbol{x}) = \int \mu(\boldsymbol{x} - \boldsymbol{y})\mu'(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$ be the convolution of μ and μ'. Then $\mu * \mu'$ is log-concave

Proof of 1.

- For $\boldsymbol{x}_S, \boldsymbol{x}'_S \in \mathbb{R}^S$ and $\lambda \in (0,1)$, let $f(\boldsymbol{x}_{-S}) := \mu(\boldsymbol{x}_S, \boldsymbol{x}_{-S})$, $g(\boldsymbol{x}_{-S}) := \mu(\boldsymbol{x}'_S, \boldsymbol{x}_{-S})$, $h(\boldsymbol{x}_{-S}) := \mu(\lambda \boldsymbol{x}_S + (1-\lambda)\boldsymbol{x}'_S, \boldsymbol{x}_{-S})$
- $h(\lambda \boldsymbol{x}_{-S} + (1-\lambda)\boldsymbol{x}'_{-S}) = \mu(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{x}') \geq \mu(\boldsymbol{x})^\lambda \mu(\boldsymbol{x}')^{1-\lambda} = f(\boldsymbol{x}_{-S})^\lambda g(\boldsymbol{x}'_{-S})^{1-\lambda}$
- By Prékopa-Leindler Inequality, we have

$$\underbrace{\int_{\mathbb{R}^n} \mu(\lambda \boldsymbol{x}_S + (1-\lambda)\boldsymbol{x}'_S, \boldsymbol{x}_{-S}) \mathrm{d}\boldsymbol{x}_{-S}}_{\mu_S(\lambda \boldsymbol{x}_S + (1-\lambda)\boldsymbol{x}'_S)} \geq \Bigg(\underbrace{\int_{\mathbb{R}^n} \mu(\boldsymbol{x}_S, \boldsymbol{x}_{-S}) \mathrm{d}\boldsymbol{x}_{-S}}_{\mu_S(\boldsymbol{x}_S)}\Bigg)^\lambda \cdot \Bigg(\underbrace{\int_{\mathbb{R}^n} \mu(\boldsymbol{x}'_S, \boldsymbol{x}_{-S}) \mathrm{d}\boldsymbol{x}_{-S}}_{\mu_S(\boldsymbol{x}'_S)}\Bigg)^{1-\lambda}$$

Brunn-Minkowski Inequality

Corollary. Let $\mu, \mu': \mathbb{R}^n \to \mathbb{R}$ be log-concave.

1. Let $S \subseteq [n]$ and let $\mu_S(\boldsymbol{x}_S) := \int_{\mathbb{R}^{[d]\setminus S}} \mu(\boldsymbol{x}_S, \boldsymbol{x}_{-S}) \mathrm{d}\boldsymbol{x}_{-S}$ be the marginal on S. Then μ_S is log-concave
2. Let $(\mu * \mu')(\boldsymbol{x}) = \int \mu(\boldsymbol{x} - \boldsymbol{y})\mu'(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$ be the convolution of μ and μ'. Then $\mu * \mu'$ is log-concave

Proof of 2.

- For $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^n$ and $\lambda \in (0,1)$, we want to show

$$\int_{\mathbb{R}^n} \underbrace{\mu(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{x}' - \boldsymbol{y})\mu'(\boldsymbol{y})}_{=:\, h(\boldsymbol{y})}\mathrm{d}\boldsymbol{y} \geq \left(\int_{\mathbb{R}^n} \underbrace{\mu(\boldsymbol{x} - \boldsymbol{y})\mu'(\boldsymbol{y})}_{=:\, f(\boldsymbol{y})}\mathrm{d}\boldsymbol{y}\right)^{\lambda} \cdot \left(\int_{\mathbb{R}^n} \underbrace{\mu(\boldsymbol{x}' - \boldsymbol{y})\mu'(\boldsymbol{y})}_{=:\, g(\boldsymbol{y})}\mathrm{d}\boldsymbol{y}\right)^{1-\lambda}$$

- We need to check the condition in Prékopa-Leindler Inequality: $\forall \boldsymbol{y}, \boldsymbol{y}' \in \mathbb{R}^n$,

$$\begin{aligned} h(\lambda\boldsymbol{y} + (1-\lambda)\boldsymbol{y}') &= \mu\big(\lambda(\boldsymbol{x} - \boldsymbol{y}) + (1-\lambda)(\boldsymbol{x}' - \boldsymbol{y}')\big)\mu'(\lambda\boldsymbol{y} + (1-\lambda)\boldsymbol{y}') \\ &\geq \mu(\boldsymbol{x} - \boldsymbol{y})^{\lambda}\mu(\boldsymbol{x}' - \boldsymbol{y}')^{1-\lambda}\mu'(\boldsymbol{y})^{\lambda}\mu'(\boldsymbol{y}')^{1-\lambda} = f(\boldsymbol{y})^{\lambda} g(\boldsymbol{y})^{1-\lambda} \end{aligned}$$

Brunn-Minkowski Inequality

We show one more application of Prékopa-Leindler Inequality

- Suppose we want to prove that for a convex set $S \subseteq \mathbb{R}^n$ that is symmetric w.r.t. $\mathbf{0}$,
$$\arg\sup_{\boldsymbol{u}\in\mathbb{R}^n} \Pr_{\boldsymbol{x}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_n)}[\boldsymbol{x} \in (S+\boldsymbol{u})] = \mathbf{0}$$
That is, a random Gaussian vector falls in $S + \boldsymbol{u}$ with the highest probability when $\boldsymbol{u} = \mathbf{0}$
- Such kind of result is useful in anti-concentration or small ball probability

- Consider $\mu(\boldsymbol{x}) \propto \exp(-\|\boldsymbol{x}\|_2^2/2)$ and $\mu'(\boldsymbol{x}) = \exp(-\delta_S(\boldsymbol{x}))$ where δ_S is the indicator function for S
- Corollary (2) gives that $\mu' * \mu$ is log-concave
$$(\mu' * \mu)(\boldsymbol{x}) = \int_{\mathbb{R}^n} \mu'(\boldsymbol{x}-\boldsymbol{y})\mu(\boldsymbol{y})\mathrm{d}\boldsymbol{y} = \int_{\mathbb{R}^n} \mu(\boldsymbol{y})\mathbf{1}[\boldsymbol{x}-\boldsymbol{y} \in S]\mathrm{d}\boldsymbol{y} = \int_{\boldsymbol{x}+S} \mu(\boldsymbol{y})\mathrm{d}\boldsymbol{y}$$
- By log-concavity and the symmetry, $\arg\sup_{u\in\mathbb{R}^n} (\mu' * \mu)(\boldsymbol{x}) = \mathbf{0}$