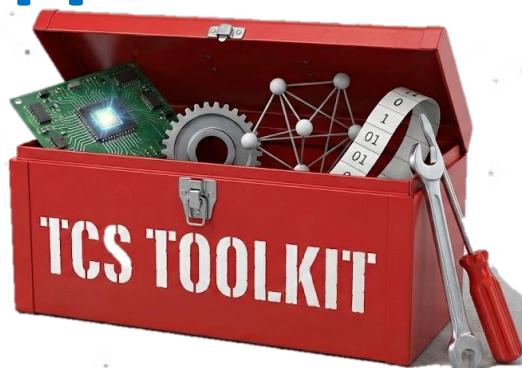


CS 58500 – Theoretical Computer Science Toolkit

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

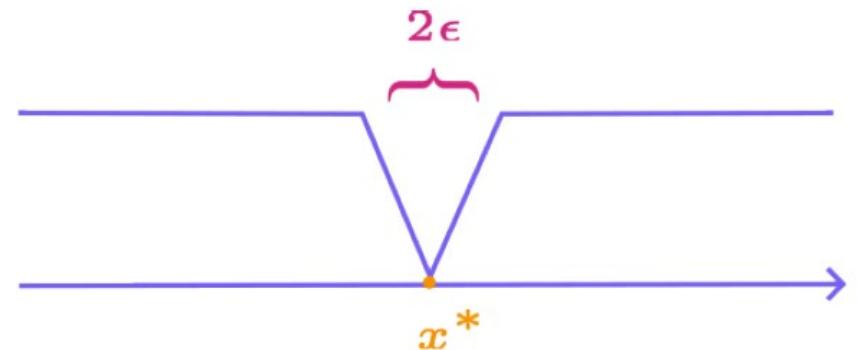
Simple Counterexample:

$$f(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} = \mathbf{x}^* \\ 1 & \text{otherwise} \end{cases}$$

Lipschitz Counterexample:

$$f(\mathbf{x}) = \min\{\|\mathbf{x} - \mathbf{x}^*\|_2, \epsilon\} \quad \forall \mathbf{x} \in B_2^n$$

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



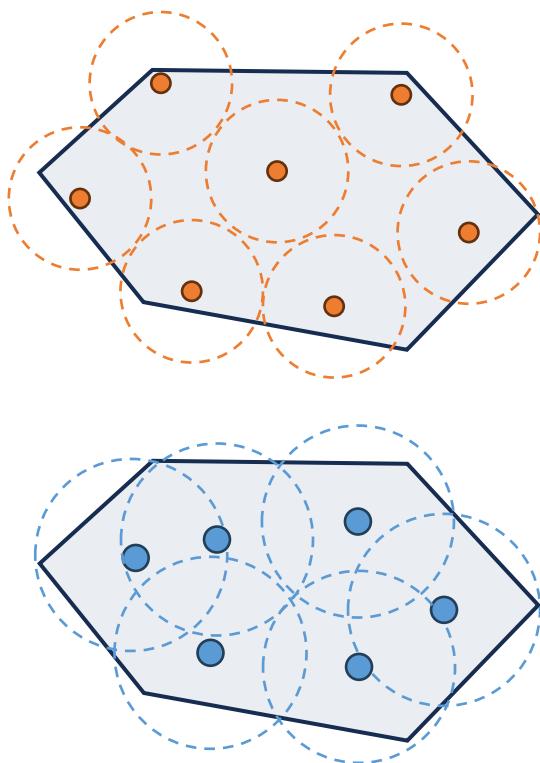
Detour: Packing and Covering

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

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

Theorem.

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



Why Non-Convex Optimization Is Hard

Lemma. No algorithm \mathcal{A} which accesses a target 1-Lipschitz function $f: B_2^n \rightarrow \mathbb{R}$ 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 $\{\mathbf{x}_i\}_{i \in [M]}$ be such a maximal packing of B_2^n
- For each $i \in [M]$, define $f_i(\mathbf{x}) = \min\{\|\mathbf{x} - \mathbf{x}_i\|_2, \epsilon\}$
- If #queries $< \left(\frac{1}{2\epsilon}\right)^n - 1$, then $\exists i \neq j \in [M]$ such that no query point lies in $(\mathbf{x}_i + \epsilon B_2^n) \cup (\mathbf{x}_j + \epsilon B_2^n)$
- Thus, \mathcal{A} cannot distinguish f_i and f_j



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 $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$ the line segment between \mathbf{x} and \mathbf{x}' lies in \mathcal{X} , i.e.,

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

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

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

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

Convexity

Convex set \Rightarrow convex function:

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

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

- δ_S is a convex function

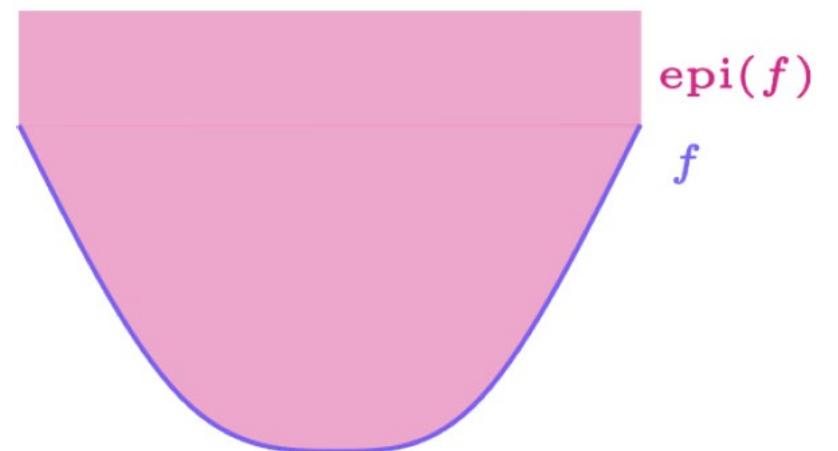
Convex function \Rightarrow convex set:

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

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

- f is convex if and only if $\text{epi}(f)$ is convex

- $\min_x f(x) \Leftrightarrow \min_{(x,t) \in \text{epi}(f)} t$



Convexity

Convex set \Rightarrow convex function:

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

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

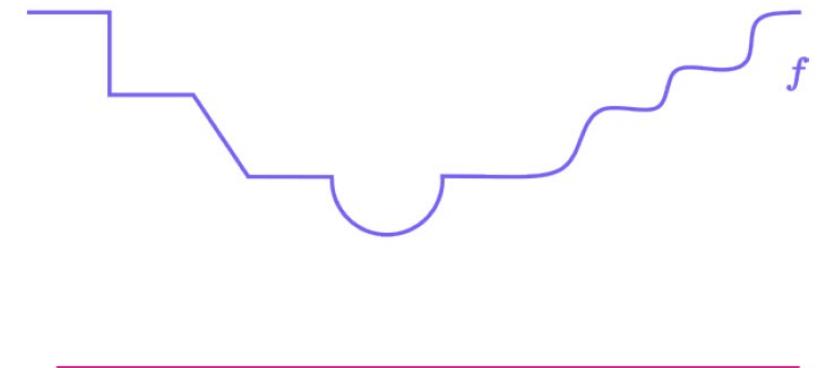
- δ_S is a convex function

Convex function \Rightarrow convex set:

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

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

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



Quasi-convex

Convexity: Analytic Properties

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

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

Proof of “ \Rightarrow ”.

- Convexity implies that for any $\lambda \in [0,1]$,

$$\lambda f(\mathbf{x}') + (1 - \lambda)f(\mathbf{x}) \geq f((1 - \lambda)\mathbf{x} + \lambda\mathbf{x}') \Leftrightarrow f(\mathbf{x}') \geq \frac{f((1 - \lambda)\mathbf{x} + \lambda\mathbf{x}') - (1 - \lambda)f(\mathbf{x})}{\lambda}$$

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

Directional derivative

Convexity: Analytic Properties

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

$$\mathbf{x}^* \in \arg \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \iff \langle \nabla f(\mathbf{x}^*), \mathbf{x}^* - \mathbf{x} \rangle \leq 0 \quad \forall \mathbf{x} \in \mathcal{X}$$

Proof.

- If $\langle \nabla f(\mathbf{x}^*), \mathbf{x}^* - \mathbf{x} \rangle \leq 0$, then by the previous lemma, for any $\mathbf{x} \in \mathcal{X}$, we have

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \langle \nabla f(\mathbf{x}^*), \mathbf{x}^* - \mathbf{x} \rangle \geq f(\mathbf{x}^*)$$

- If $\mathbf{x}^* \in \mathcal{X}$ is a minimizer, suppose $\langle \nabla f(\mathbf{x}^*), \mathbf{x}^* - \mathbf{x} \rangle > 0$ for some $\mathbf{x} \in \mathcal{X}$

- Consider $\mathbf{x}_\lambda := (1 - \lambda)\mathbf{x}^* + \lambda\mathbf{x} \in \mathcal{X}$ for $\lambda \in [0,1]$ and $\phi(\lambda) := f(\mathbf{x}_\lambda)$

- $\phi'(0) = Df(\mathbf{x}^*)[\mathbf{x} - \mathbf{x}^*] = \langle \nabla f(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle < 0$

- Thus, there exists a sufficiently small $\lambda > 0$ such that $\phi(\lambda) < \phi(0)$, i.e., $f(\mathbf{x}_\lambda) < f(\mathbf{x}^*)$, contradiction

Constructive!

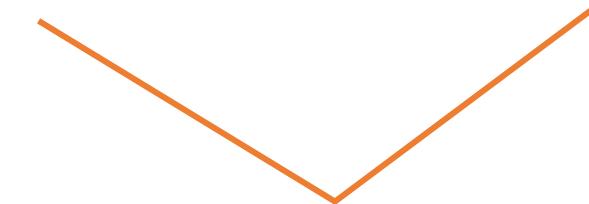
Convexity: Geometric Properties

Lemma (Characterization of minima and maxima). Let $f: \mathcal{X} \rightarrow \mathbb{R}$ be convex. Let f have minimizer set $\mathcal{X}_{\min} := \arg \min_x f(x)$ and maximizer set $\mathcal{X}_{\max} := \arg \max_x f(x)$. 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**, $\mathcal{X}_{\min} = \{\mathbf{x}^*\}$ and \mathcal{X}_{\max} only contains boundary points if it is nonempty.

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



Convexity: Geometric Properties

Theorem (Separating Hyperplane Theorem). Let $S \subset \mathbb{R}^n$ be compact and convex and suppose $x_0 \notin S$. There is a **separating hyperplane** $g \in \mathbb{R}^n$ such that

$$\langle g, x_0 \rangle > \langle g, x \rangle \quad \forall x \in S$$

Moreover, if $x_0 \in S$ is a boundary point, there is a **supporting hyperplane** $g \in \mathbb{R}^n, g \neq \mathbf{0}$ such that

$$\langle g, x_0 \rangle \geq \langle g, x \rangle \quad \forall x \in S$$

Proof.

- Define a potential function $f(x) := \|x - x_0\|_2^2$ for $x \in S$
- f is convex, and even strictly convex
- $x^* := \arg \min_{x \in S} f(x)$ **exists** (by the compactness of S) and is **unique** (by the previous lemma)
- **First-order optimality** implies that $\langle \nabla f(x^*), x^* - x \rangle = \langle x - x_0, x^* - x \rangle \leq 0$ for any $x \in S$

Convexity: Geometric Properties

Theorem (Separating Hyperplane Theorem). Let $S \subset \mathbb{R}^n$ be compact and convex and suppose $x_0 \notin S$. There is a **separating hyperplane** $\mathbf{g} \in \mathbb{R}^n$ such that

$$\langle \mathbf{g}, x_0 \rangle > \langle \mathbf{g}, x \rangle \quad \forall x \in S$$

Moreover, if $x_0 \in S$ is a boundary point, there is a **supporting hyperplane** $\mathbf{g} \in \mathbb{R}^n, \mathbf{g} \neq \mathbf{0}$ such that

$$\langle \mathbf{g}, x_0 \rangle \geq \langle \mathbf{g}, x \rangle \quad \forall x \in S$$

Proof.

- **First-order optimality** implies that $\langle \nabla f(x^*), x^* - x \rangle = \langle x^* - x_0, x^* - x \rangle \leq 0$ for any $x \in S$
- Define $\mathbf{g} := x_0 - x^* \neq \mathbf{0}$.
- Then we have

$$\langle \mathbf{g}, x \rangle \leq \langle \mathbf{g}, x^* \rangle = \langle \mathbf{g}, x_0 \rangle + \langle \mathbf{g}, x^* - x_0 \rangle = \langle \mathbf{g}, x_0 \rangle - \|\mathbf{g}\|_2^2 < \langle \mathbf{g}, x_0 \rangle$$



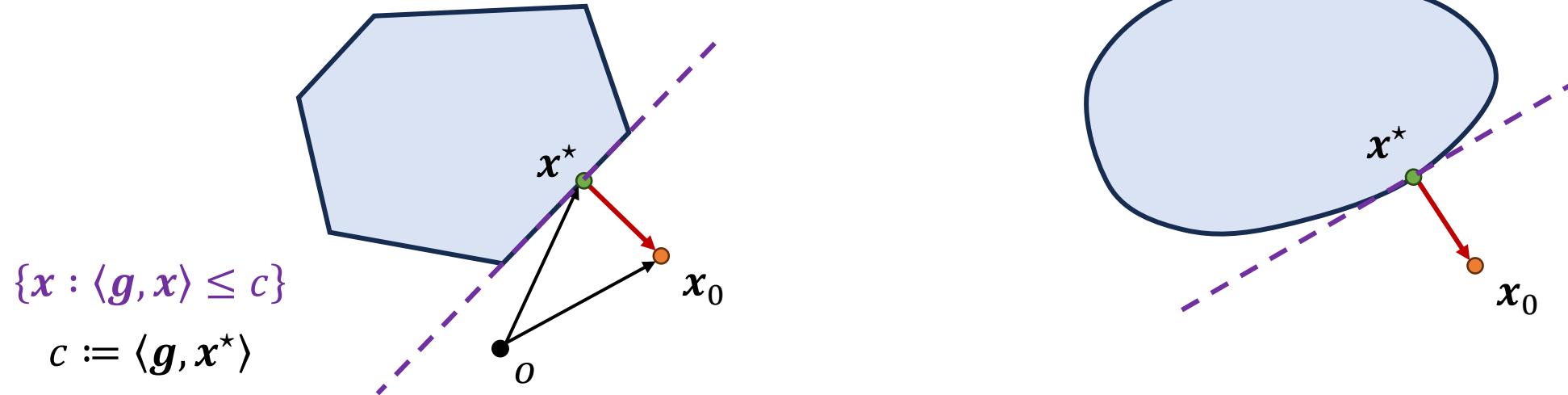
Convexity: Geometric Properties

Theorem (Separating Hyperplane Theorem). Let $S \subset \mathbb{R}^n$ be compact and convex and suppose $x_0 \notin S$. There is a **separating hyperplane** $g \in \mathbb{R}^n$ such that

$$\langle g, x_0 \rangle > \langle g, x \rangle \quad \forall x \in S$$

Moreover, if $x_0 \in S$ is a boundary point, there is a **supporting hyperplane** $g \in \mathbb{R}^n, g \neq 0$ such that

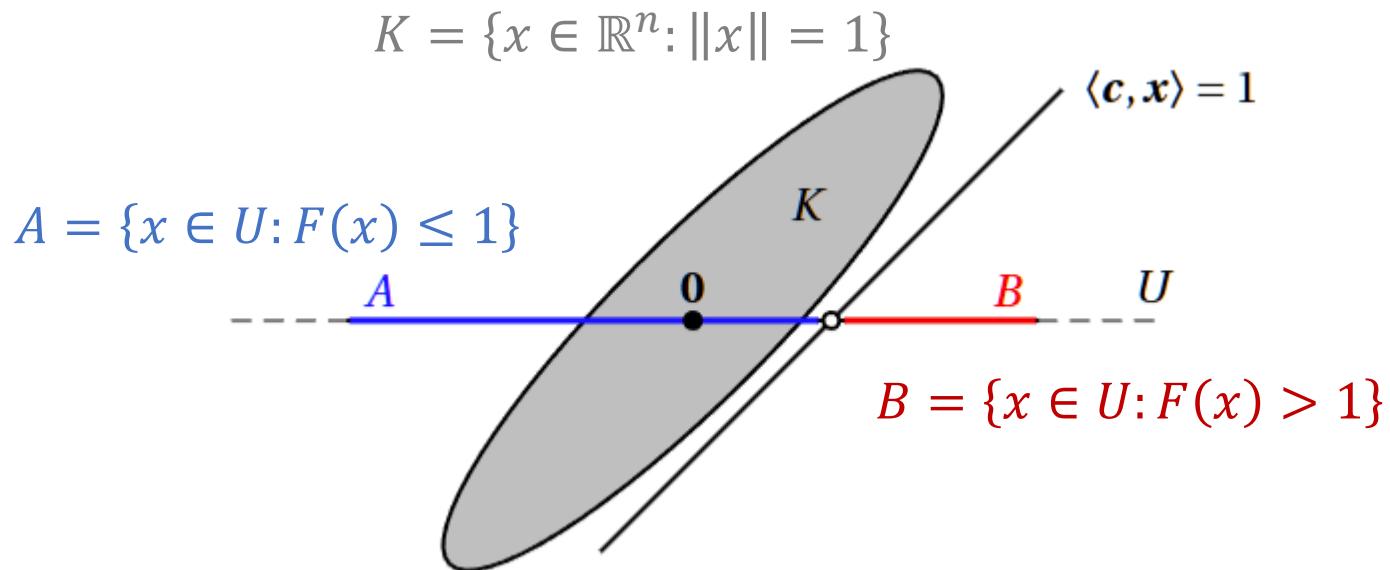
$$\langle g, x_0 \rangle \geq \langle g, x \rangle \quad \forall x \in S$$



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 $F: U \rightarrow \mathbb{R}$ is a linear function with $F(x) \leq \|x\|$ for all $x \in U$. Then there exists a linear function $\tilde{F}: \mathbb{R}^n \rightarrow \mathbb{R}$ so that

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



Application of Separating Hyperplane Theorem

Primal LP

$$\max \mathbf{c}^T \mathbf{x}$$

$$\text{s. t. } A\mathbf{x} \leq \mathbf{b}$$

$$\mathbf{x} \in \mathbb{R}_{\geq 0}^n$$

Dual LP

$$\min \mathbf{b}^T \mathbf{y}$$

$$\text{s. t. } A^T \mathbf{y} \geq \mathbf{c}$$

$$\mathbf{y} \in \mathbb{R}_{\geq 0}^m$$

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 $\mathbf{c}^T \mathbf{x} = \mathbf{b}^T \mathbf{y}$

Convexity: Subgradient

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

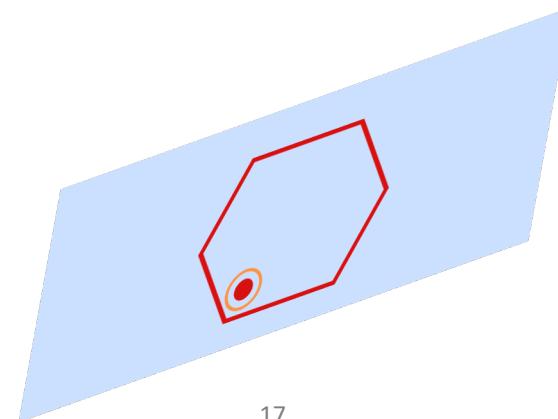
$$f(x') \geq f(x) + \langle \mathbf{g}, x' - x \rangle \quad \forall x' \in \mathcal{X}$$

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

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

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

- 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 $f: \mathcal{X} \rightarrow \mathbb{R}$. We say \mathbf{g} is a **subgradient** of f at $x \in \mathcal{X}$ if

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

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

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

Proof.

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

Convexity: Subgradient

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

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

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

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

Proof.

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

$$\langle \mathbf{a}, x \rangle + bf(x) \geq \langle \mathbf{a}, x' \rangle + bf(x') \implies f(x') \geq f(x) + \left\langle -\frac{\mathbf{a}}{b}, x' - x \right\rangle$$

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



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 :

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

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

Proof.

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

$$\begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} = \sum_{i=1}^k \lambda_i \begin{pmatrix} \mathbf{a}_i \\ 1 \end{pmatrix}, \quad \mathbf{a}_i \in A, \lambda_i \geq 0, \sum_{i=1}^k \lambda_i = 1$$

Convexity: Convex Hull

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

Proof.

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

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

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

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

Convexity: Convex Hull

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

Proof.

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

$$\binom{x}{1} = \sum_{i=1}^k (\lambda_i + t\mu_i) \binom{a_i}{1}, \quad a_i \in A, \lambda_i > 0, \sum_{i=1}^k \lambda_i = 1, \mu_i \in \mathbb{R}$$

- 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 $f: \mathcal{X} \rightarrow \mathbb{R}$ be convex for $\mathcal{X} \subset \mathbb{R}^n$ and assume the range $\sup f - \inf f$ bounded by $\text{poly}(n)$. 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 $\text{poly}\left(n, \log\frac{1}{\epsilon}\right)$ additional time, such that with high probability, the algorithm returns x satisfying

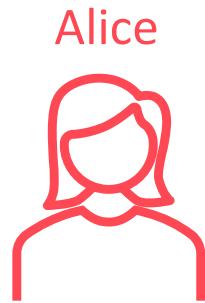
$$f(x) \leq \min_{x^* \in \mathcal{X}} f(x^*) + \epsilon$$

Cutting-plane methods

Game theory perspective of cutting-plane methods

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

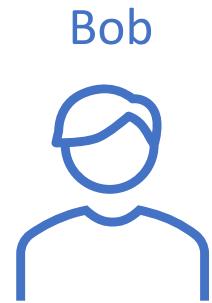
If $\text{Vol}_n(S_t) < V_{\min}$, the game ends



$$S_0 \supseteq S^*$$

Alice choose x_t and to Bob

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



Bob choose $g_t \in \mathbb{R}^n$ s.t. $\langle g_t, x_t \rangle > \langle g_t, x \rangle \quad \forall x \in S^*$ and send to Alice

Alice update $S_{t+1} \supseteq S_t \cap H_t$ where
 $H_t := \{x \in \mathbb{R}^d : \langle g_t, x \rangle < \langle g_t, x_t \rangle\}$

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

Cutting-plane methods

Lemma. Let $f: \mathcal{X} \rightarrow \mathbb{R}$ be convex for $\mathcal{X} \subset \mathbb{R}^n$, and suppose f has minimizer set $\mathcal{X}^* := \arg \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$.

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

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

If the game terminates in T iterations, letting $f^* := \min f(\mathbf{x})$, we have

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

Proof.

- $\mathbf{x}_t \in \mathcal{X}^*$ is equivalent to:

$$f(\mathbf{x}_t) \leq f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X} \iff f(\mathbf{x}_t) + \langle \mathbf{0}, \mathbf{x} - \mathbf{x}_t \rangle \leq f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X} \iff \mathbf{0} \in \partial f(\mathbf{x}_t)$$

- If $\mathbf{x}_t \notin \mathcal{X}^*$, then for any $\mathbf{x}^* \in \mathcal{X}^*$, we have

$$0 > f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq \langle \mathbf{g}_t, \mathbf{x}^* - \mathbf{x}_t \rangle$$

Thus, Bob's choice is valid

Cutting-plane methods

Lemma. Let $f: \mathcal{X} \rightarrow \mathbb{R}$ be convex for $\mathcal{X} \subset \mathbb{R}^n$, and suppose f has minimizer set $\mathcal{X}^* := \arg \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$.

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

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

If the game terminates in T iterations, letting $f^* := \min f(\mathbf{x})$, we have

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

Proof.

- For any $\mathbf{x}^* \in \mathcal{X}^*$, consider the set $S_\alpha := \{\mathbf{x} = (1 - \alpha)\mathbf{x}^* + \alpha\mathbf{z} : \mathbf{z} \in \mathcal{X}\} = (1 - \alpha)\mathbf{x}^* + \alpha\mathcal{X}$
- We have $\text{Vol}_n(S_\alpha) = \alpha^n \text{Vol}_n(\mathcal{X}) = \alpha^n \text{Vol}_n(S_0) = V_{\min}$
- If $\text{Vol}_n(S_T) < V_{\min}$, then $\exists \mathbf{x} \in S_\alpha \setminus S_T$ such that $\mathbf{x} = (1 - \alpha)\mathbf{x}^* + \alpha\mathbf{z}$
- Then, $\exists t \in [T]$ such that $\mathbf{x} \notin H_t$

Cutting-plane methods

Lemma. Let $f: \mathcal{X} \rightarrow \mathbb{R}$ be convex for $\mathcal{X} \subset \mathbb{R}^n$, and suppose f has minimizer set $\mathcal{X}^* := \arg \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$.

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

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

If the game terminates in T iterations, letting $f^* := \min f(\mathbf{x})$, we have

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

Proof.

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

$$f(\mathbf{x}) \geq f(\mathbf{x}_t) + \langle \mathbf{g}_t, \mathbf{x} - \mathbf{x}_t \rangle \geq f(\mathbf{x}_t)$$

- Thus,

$$f(\mathbf{x}_t) \leq f(\mathbf{x}) \leq (1 - \alpha)f(\mathbf{x}^*) + \alpha f(\mathbf{z}) \leq f^* + \alpha \left(\max_{\mathbf{z} \in \mathcal{X}} f(\mathbf{z}) - f^* \right)$$



Cutting-plane methods

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

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

Further, suppose Alice always chooses $x_t \in \text{relint}(S_t)$, and Bob (who holds $S^* \leftarrow \mathcal{X}^*$) plays by ending the game if $\mathbf{0} \in \partial f(x_t)$, and returning $g_t \in \partial f(x_t)$ otherwise.

If the game terminates in T iterations, letting $f^* := \min f(x)$, we have

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

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

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

Cutting-plane methods

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

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

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

$$H = \{x \in \mathbb{R}^n : \langle g, x \rangle \leq \langle g, \bar{x}_S \rangle\}$$

whose defining halfplane passes through \bar{x}_S satisfies $\text{Vol}_n(S \cap H) \geq \frac{1}{e} \text{Vol}_n(S)$

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

Cutting-plane methods

Theorem (Polynomial-time convex optimization). Let $f: \mathcal{X} \rightarrow \mathbb{R}$ be convex for $\mathcal{X} \subset \mathbb{R}^n$ and assume the range $\sup f - \inf f$ bounded by $\text{poly}(n)$. 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 $\text{poly}\left(n, \log\frac{1}{\epsilon}\right)$ additional time, such that with high probability, the algorithm returns x satisfying

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

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

- $F(\mathbf{x}) = \exp(-G(\mathbf{x}))$ for some convex function $G: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$
 - $\ln(F(\mathbf{x}))$ is concave
 - $F(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \geq F(\mathbf{x})^\lambda \cdot F(\mathbf{y})^{1-\lambda}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ 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 $\mu(\mathbf{x}) \propto F(\mathbf{x})$ is a log-concave distribution, which means the normalizing constant is hidden:

$$\mu(\mathbf{x}) = \frac{F(\mathbf{x})}{Z}, \quad Z := \int F(\mathbf{x}) d\mathbf{x}$$

Log-concavity

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

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

Examples:

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

Log-concavity

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

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

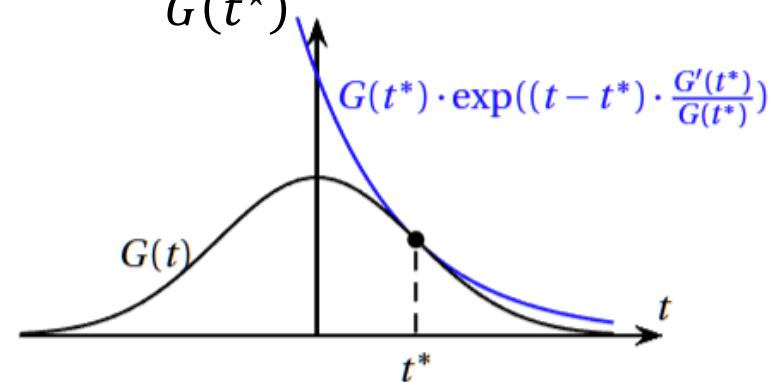
- If the derivative of a log-concave function < 0 at any point, then it must decay at an exponential rate

Proof.

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

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

- Exponentiating both sides proves the lemma



Log-concavity

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

Proof.

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

$$0 = \int_{-1}^1 tG'(t) dt = tG(t) \Big|_{-1}^1 - \int_{-1}^1 G(t) dt \implies \int_{-1}^1 G(t) dt = 1$$

Log-concavity

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

Proof.

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

$$\begin{aligned}
 1 &= \int_{-1}^1 G(t) dt \leq \int_{-1}^1 \min \left\{ G(0) \exp \left(t \frac{G'(0)}{G(0)} \right), 1 \right\} dt = \int_{-1}^{\alpha} G(0) \exp \left(t \frac{G'(0)}{G(0)} \right) dt + (1 - \alpha) \\
 &\leq G(0) \frac{G(0)}{G'(0)} \exp \left(t \frac{G'(0)}{G(0)} \right) \Big|_{-\infty}^{\alpha} + (1 - \alpha) = \frac{G(0)^2}{G'(0)} \exp \left(\alpha \frac{G'(0)}{G(0)} \right) + (1 - \alpha) \quad \boxed{\alpha = -\frac{G(0)}{G'(0)} \ln G(0)} \\
 &= \frac{G(0)}{G'(0)} + 1 + \frac{G(0)}{G'(0)} \ln G(0) \quad \Rightarrow \quad \frac{G(0)}{G'(0)} (1 + \ln G(0)) \geq 0 \quad \Rightarrow \quad G(0) \geq \frac{1}{e}
 \end{aligned}$$

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

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

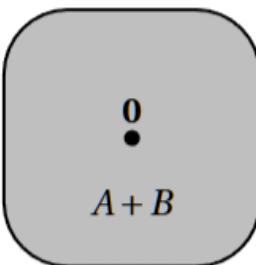
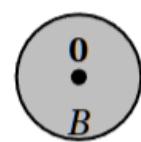
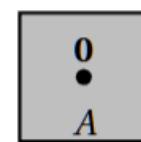
Theorem (Brunn-Minkowski, v2). Let $A, B \subseteq \mathbb{R}^n$ be non-empty compact sets. Then

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

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

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

- $A + B := \{a + b : a \in A, b \in B\}$
- 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

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

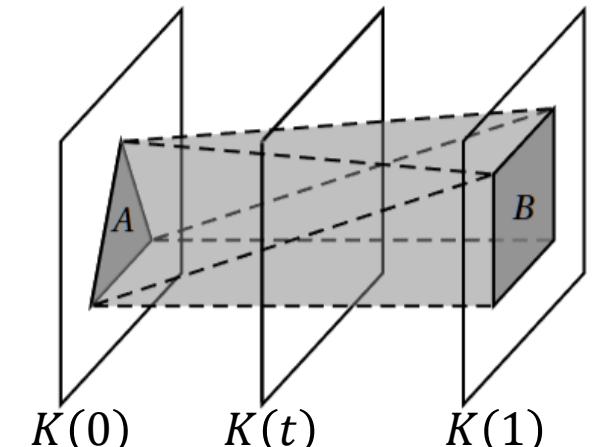
Proof.

- 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 $K := \text{conv}((A \times \{0\}) \cup (B \times \{1\}))$, and $K(t) := \{x \in \mathbb{R}^n : (x, t) \in K\}$ for $t \in [0,1]$
- $K(t) = (1 - t) \cdot A + t \cdot B$
- If we can show that $t \mapsto \text{Vol}_n(K(t))$ is **log-concave**, then

$$\text{Vol}_n(K(\lambda \cdot 0 + (1 - \lambda) \cdot 1)) \geq \text{Vol}_n(K(0))^{\lambda} \cdot \text{Vol}_n(K(1))^{1-\lambda}$$

$\lambda A + (1 - \lambda)B$
 A
 B

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



Detour: Steiner Symmetrization

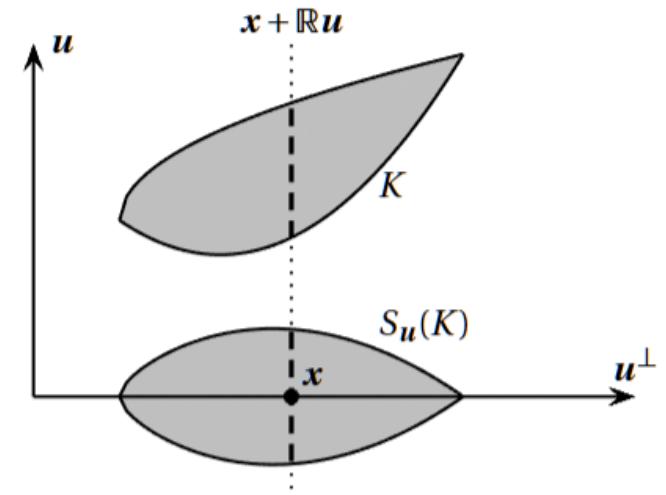
Let $K \subseteq \mathbb{R}^n$ be a convex body and $\mathbf{u} \in \mathbb{S}^{n-1}$ a unit direction. The **Steiner symmetral** $S_{\mathbf{u}}(K) \subseteq \mathbb{R}^n$ of K in direction \mathbf{u} is defined so that for every $x \in \mathbf{u}^\perp$,

$$\text{Vol}_1((x + \mathbb{R}\mathbf{u}) \cap K) = \text{Vol}_1((x + \mathbb{R}\mathbf{u}) \cap S_{\mathbf{u}}(K))$$

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

Properties:

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

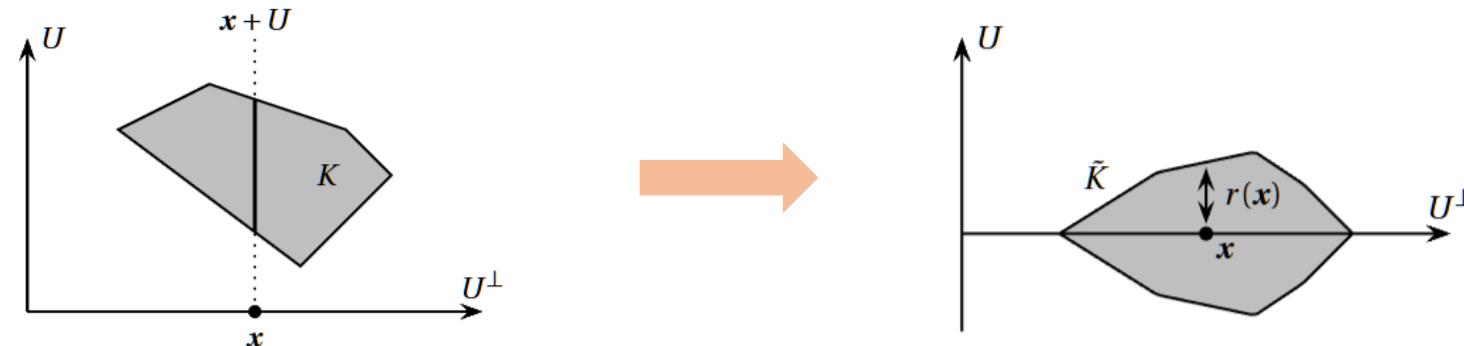


Detour: Steiner Symmetrization

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 $F: U^\perp \rightarrow \mathbb{R}$ defined by $F(x) := \text{Vol}_k(K \cap (U + x))^{1/k}$ is concave on its support.

Proof.

- Apply the Steiner Symmetrization for all directions $\mathbf{u} \in U$ and let \tilde{K} be the limiting convex body
- $\tilde{K} \cap (U + x)$ is a k -dimensional ball of radius $r(x)$ such that
$$r(x)^k \cdot \text{Vol}_k(B_2^k) = \text{Vol}_k(\tilde{K} \cap (U + x)) = \text{Vol}_k(K \cap (U + x)) = F(x) \quad \forall x \in U^\perp$$
- K convex $\Rightarrow \tilde{K}$ convex $\Rightarrow r(x)$ concave $\Rightarrow F(x)$ concave

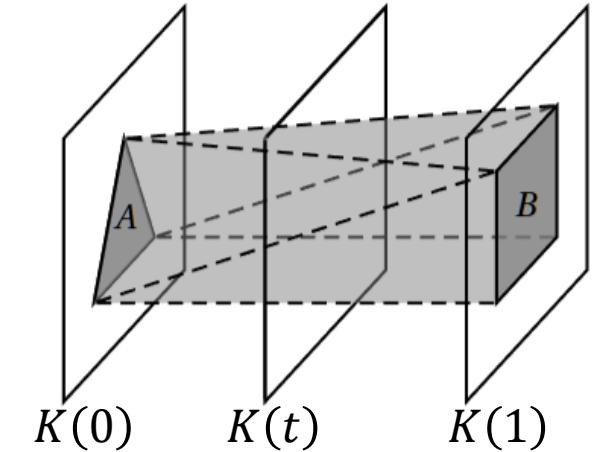


Brunn-Minkowski Inequality

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 $F: U^\perp \rightarrow \mathbb{R}$ defined by $F(x) := \text{Vol}_k(K \cap (U + x))^{1/k}$ is concave on its support.

- We take $U = \{x \in \mathbb{R}^{n+1} : x_{n+1} = 0\}$, and $k = n$
- Then, $t \mapsto \text{Vol}_n(K \cap (U + (\mathbf{0}, t))) = \text{Vol}_n(K(t))^{1/n}$ is concave
- By the property of log-concavity, $\text{Vol}_n(K(t))^{1/n}$ is also log-concave
- Thus, $\frac{1}{n} \ln \text{Vol}_n(K(t))$ concave $\Rightarrow \text{Vol}_n(K(t))$ log-concave

■



Brunn-Minkowski Inequality

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

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

Then

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

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

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

Brunn-Minkowski Inequality

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

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

Proof of 1.

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

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

Brunn-Minkowski Inequality

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

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

Proof of 2.

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

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

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

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

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. $\mathbf{0}$,

$$\arg \sup_{\mathbf{u} \in \mathbb{R}^n} \Pr_{x \sim \mathcal{N}(\mathbf{0}, I_n)} [x \in (S + \mathbf{u})] = \mathbf{0}$$

That is, a random Gaussian vector falls in $S + \mathbf{u}$ with the highest probability when $\mathbf{u} = \mathbf{0}$

- Such kind of result is useful in anti-concentration or small ball probability
- Consider $\mu(x) \propto \exp(-\|x\|_2^2/2)$ and $\mu'(x) = \exp(-\delta_S(x))$ where δ_S is the indicator function for S
- [Corollary \(2\)](#) gives that $\mu' * \mu$ is log-concave

$$(\mu' * \mu)(x) = \int_{\mathbb{R}^n} \mu'(x - y) \mu(y) dy = \int_{\mathbb{R}^n} \mu(y) \mathbf{1}[x - y \in S] dy = \int_{x+S} \mu(y) dy$$

- By log-concavity and the symmetry, $\arg \sup_{\mathbf{u} \in \mathbb{R}^n} (\mu' * \mu)(x) = \mathbf{0}$