Lecture 6: Cutting-Plane Methods

TTIC 31070 / CMSC 35470 / BUSF 36903 / STAT 31015

Convex Optimization

Prof. Zhiyuan Li

Spring 2026

From Duality Certificates to Algorithms

Lectures 1–5: **what** makes a point optimal (KKT, Slater, certificates).

Today: **how** to find such a point.

Key idea. Recast optimization as a **search problem**. Maintain a shrinking outer approximation S_t of a target set, query a point $x_t \in S_t$, and use a subgradient or separation oracle to **cut** S_t.

Two instantiations:

- **Center-of-mass method** — sharp but expensive
- **Ellipsoid method** — tractable with an extra n factor

Problem Setup and δ-Optimality

Problem: $\min_{x \in X} f(x)$, $X \subseteq \mathbb{R}^n$ convex body, $f : X \to \mathbb{R}$ convex.

Assume $f^* := \min_X f$ is known, and we have a **first-order oracle**: given x, return $f(x)$ and some $g \in \partial f(x)$.

Definition 6.1 (δ-optimal point). For $\delta \geq 0$, a point $x \in X$ is *δ-optimal* if

$$f(x) \leq \inf_{y \in X} f(y) + \delta.$$

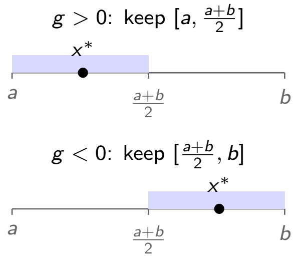
Also define the diameter of the objective values: $\Delta := \sup_{x \in X} f(x) - \inf_{x \in X} f(x)$.

Guarantee form: after T iterations, *one* of the first T query points is δ-optimal.

One-Dimensional Example: Bisection

Lemma 6.1 (One-dimensional subgradient cut). Let $f : [a, b] \to \mathbb{R}$ be convex, $c = (a+b)/2$, $g \in \partial f(c)$.

- $g > 0 \Rightarrow$ minimizer in $[a, c]$
- $g < 0 \Rightarrow$ minimizer in $[c, b]$
- $g = 0 \Rightarrow c$ is a minimizer



The sign of the subgradient tells us which half to keep. Query midpoint $\to$ halve the interval.

One-Dimensional Complexity

Theorem 6.2 (One-dimensional cutting-plane complexity). Apply the bisection procedure to $f : [a_0, b_0] \to \mathbb{R}$ convex. Let I_T be the interval after T steps. Then

$$|I_T| = 2^{-T}|I_0|, \qquad |x - x^*| \leq 2^{-T}|I_0| \; \forall\, x \in I_T.$$

Proof. Each step halves the interval length. Since I_T contains a minimizer, any $x \in I_T$ is at distance $\leq |I_T| = 2^{-T}|I_0|$ from it. □

1D takeaway: $T = O(\log(1/\varepsilon))$ queries to get ε-close. The next question: can we do the same in higher dimensions?

Separation Oracle and Feasibility

Definition 6.2 (Separation oracle for $K \subseteq S_0$). On input $x \in S_0$, the oracle returns either

- a certificate that $x \in K$, or
- an open halfspace H with $K \subseteq H$ and $x \notin H$ (**valid cut**)

Convex feasibility problem. Given $S_0 \supseteq K$ and a separation oracle for K, find a point $x \in K$.

Strategy: maintain a shrinking outer approximation $S_t \supseteq K$. Each oracle call either terminates or produces a cut that refines S_t.

Generic Cutting-Plane Algorithm

Require: $S_0 \subseteq \mathbb{R}^n$ and a separation oracle for $K \subseteq S_0$.

For $t = 0, 1, 2, \dots$:

1. Choose a query point $x_t \in S_t$ and query the separation oracle at x_t.
2. **If** the oracle certifies $x_t \in K$: output x_t and stop.
3. **Else** receive a valid cut H_t for K and construct a new outer approximation S_{t+1} satisfying
$$S_t \cap H_t \subseteq S_{t+1}.$$

Two design choices at each iteration:

- **How to pick** x_t given S_t
- **How to represent** S_{t+1} given the cut H_t

We will see two concrete instantiations — *center-of-mass* and *ellipsoid* — that make opposite choices on this tradeoff.

Volume Threshold Principle

Proposition 6.3 (Volume threshold implies success). Assume all returned halfspaces are valid cuts for $K \subseteq S_0$. If after T iterations the algorithm has not yet queried a point of K, then $K \subseteq S_T$. Consequently, if

$$\mathrm{vol}(S_T) < \mathrm{vol}(K),$$

then one of the first T query points must already lie in K.

Proof. Every cut H_t is valid for K, so $K \subseteq S_t \cap H_t \subseteq S_{t+1}$. Hence $K \subseteq S_T$. If $\mathrm{vol}(S_T) < \mathrm{vol}(K) \leq \mathrm{vol}(S_T)$ then a contradiction — so the hypothesis must have failed. □

Plan: bound the shrinkage rate of $\mathrm{vol}(S_t)$ and compare with a known lower bound on $\mathrm{vol}(K)$.

Reducing Optimization to Feasibility

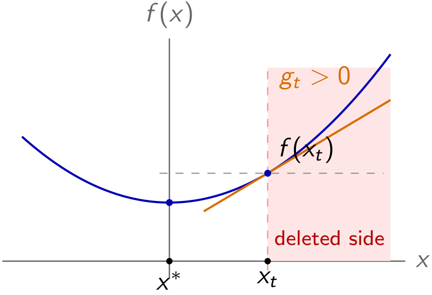
Target set: $K_\delta := \{x \in X : f(x) \le f^* + \delta\}$ (δ-approximate optimal set).

Proposition 6.4 (Sublevel oracle). A first-order oracle for f and a separation oracle for X together give a separation oracle for K_δ: at x_t,

- if $x_t \notin X$: use separation oracle for X
- if $f(x_t) \le f^* + \delta$: certify $x_t \in K_\delta$
- else: use $g_t \in \partial f(x_t)$ to cut with $H_t = \{x : \langle g_t, x - x_t \rangle < 0\}$

The subgradient inequality $f(x) \ge f(x_t) + \langle g_t, x - x_t \rangle$ means every $x \in K_\delta$ satisfies $\langle g_t, x - x_t \rangle < 0$.

The Subgradient Cut in Pictures



If $g_t > 0$ then $f(x) > f(x_t)$ for all $x > x_t$, so the side $x > x_t$ is **safely deleted**. The kept side still contains points worse than $f(x_t)$, but it also contains any minimizer.

A Lower Bound on vol(K_δ)

Proposition 6.5 (Witness-set lower bound). For $\alpha \in (0, 1]$, let $W_\alpha := (1 - \alpha) \operatorname{argmin}_{x \in X} f(x) + \alpha X$. Then

$$W_\alpha \subseteq K_{\alpha\Delta} \quad \text{and} \quad \operatorname{vol}(K_{\alpha\Delta}) \geq \alpha^n \operatorname{vol}(X).$$

Proof sketch. For $y = (1 - \alpha)z + \alpha x$ with $z \in \operatorname{argmin} f$, $x \in X$:

$$f(y) \leq (1 - \alpha)f^* + \alpha f(x) \leq f^* + \alpha\Delta,$$

so $W_\alpha \subseteq K_{\alpha\Delta}$. Fix any minimizer x^* and apply $T_\alpha(x) := (1 - \alpha)x^* + \alpha x$: its linear part is αI, so $\operatorname{vol}(T_\alpha(X)) = \alpha^n \operatorname{vol}(X)$. Since $T_\alpha(X) \subseteq W_\alpha$, done. □

With $\alpha = \delta/\Delta$: $\operatorname{vol}(K_\delta) \geq (\delta/\Delta)^n \operatorname{vol}(X)$.

Plug into volume threshold principle: T iterations suffice once $\operatorname{vol}(S_T) < (\delta/\Delta)^n \operatorname{vol}(X)$.

Grünbaum's Theorem

Definition 6.3 (Center of mass). For $S \subseteq \mathbb{R}^n$ measurable, bounded, positive volume: $\mathrm{center}(S) := \frac{1}{\mathrm{vol}(S)} \int_S x\, dx$.

Theorem 6.6 (Grünbaum, 1960). Let $S \subseteq \mathbb{R}^n$ be convex, bounded, positive volume, and let $c = \mathrm{center}(S)$. For any closed halfspace H whose boundary passes through c,

$$\mathrm{vol}(S \cap H) \geq \left(\frac{n}{n+1}\right)^n \mathrm{vol}(S).$$

The sequence $(n/(n+1))^n$ is strictly decreasing in n and tends to $1/e$, so

$$\left(\frac{n}{n+1}\right)^n \geq \frac{1}{e} \qquad \forall\, n \geq 1.$$

Dimension-free corollary: cutting through the center of mass always keeps at least a $1/e \approx 0.37$ fraction of the volume, regardless of n.

Center-of-Mass Search Complexity

Theorem 6.7. Run the generic cutting-plane method with $x_t = \text{center}(S_t)$ and $S_{t+1} = S_t \cap H_t$, assuming each cut passes through x_t. Then

$$\text{vol}(S_T) \le (1 - e^{-1})^T \, \text{vol}(S_0).$$

Consequently, if $\text{vol}(K) \ge \eta$, then $\lfloor 3 \log(\text{vol}(S_0)/\eta) \rfloor + 1$ queries suffice.

Proof idea. Grünbaum $\Rightarrow$ each kept half has volume at most $(1 - e^{-1}) \, \text{vol}(S_t)$. Iterate T times, then apply the volume threshold principle. □

Corollary 6.8 (Optimization complexity). With target $K = K_\delta$ and $\alpha = \delta/\Delta$:

$$T = \lfloor 3n \log(\Delta/\delta) \rfloor + 1 \text{ iterations suffice.}$$

Dimension factor n comes from the α^n in Proposition 6.5.

From Center-of-Mass to Ellipsoid

Problem with center-of-mass method. After a few cuts, $S_t = S_0 \cap H_0 \cap \cdots \cap H_{t-1}$ becomes an irregular polytope. Computing center(S_t) is expensive in high dimension.

Idea. Replace S_t by a tractable **outer ellipsoid** $E_t \supseteq K$. At each step, cover $E_t \cap H_t$ by a new ellipsoid E_{t+1}.

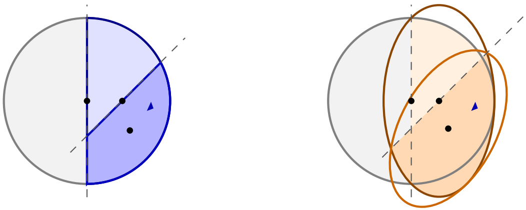
Definition 6.4 (Ellipsoid). For $Q \in \mathbb{R}^{n \times n}$ symmetric positive definite, $c \in \mathbb{R}^n$:

$$E(c, Q) := \{x \in \mathbb{R}^n : (x - c)^\top Q^{-1}(x - c) \leq 1\}.$$

Tradeoff: weaker shrinkage factor $e^{-1/(4n)}$ instead of $(1 - e^{-1})$, but an explicit update formula.

Center-of-Mass vs. Ellipsoid

(a) Two center-of-mass cuts (b) Two ellipsoid updates



(a) Exact localization $K_0 \supset K_1 \supset K_2$ becomes irregular after two cuts. (b) Same cut directions, but each E_t is replaced by an explicit covering ellipsoid.

The Ellipsoid Update

Theorem 6.9 ($n \geq 2$). Let $E(c, Q)$ be an ellipsoid and let $H = \{x : \langle a, x \rangle \leq \langle a, c \rangle\}$ be a central halfspace. Define

$$c^+ = c - \frac{1}{n+1} \frac{Qa}{\sqrt{a^\top Qa}}, \qquad Q^+ = \frac{n^2}{n^2 - 1}\left(Q - \frac{2}{n+1} \frac{Qaa^\top Q}{a^\top Qa}\right).$$

Then

$$E(c, Q) \cap H \subseteq E(c^+, Q^+) \qquad \text{and} \qquad \operatorname{vol}(E(c^+, Q^+)) \leq e^{-1/(4n)} \operatorname{vol}(E(c, Q)).$$

Reading:

- c^+ shifts the center by a rank-1 update in direction Qa
- Q^+ is a rank-1 surgery on Q, rescaled by $n^2/(n^2 - 1)$
- All operations are $O(n^2)$ per iteration
- Shrinkage $e^{-1/(4n)} \approx 1 - 1/(4n)$: weaker than Grünbaum's $(1 - 1/e)$, but tractable

Where the Ellipsoid Formulas Come From

Step 1: Normalize. $x = c + Q^{1/2}z$ sends $E(c, Q)$ to the unit ball B; after a rotation, the cut direction is $\nu = e_1$.

Step 2: Symmetry. Symmetry around e_1 forces the covering ellipsoid to have the form

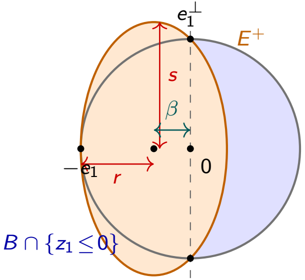
$$\frac{(z_1+\beta)^2}{r^2} + \frac{\|z_\perp\|_2^2}{s^2} \le 1.$$

Step 3: Touch constraints. Tight at $-e_1$ and on the equator: $r = 1 - \beta$, $s^2 = \frac{(1-\beta)^2}{1-2\beta}$.

Step 4: Minimize. Minimize vol $\propto r\, s^{n-1} = (1-\beta)^n(1-2\beta)^{-(n-1)/2}$:

$$\beta = \frac{1}{n+1}, \quad r = \frac{n}{n+1}, \quad s = \frac{n}{\sqrt{n^2-1}}.$$

Unnormalize via $x = c + Q^{1/2}z \Rightarrow$ Theorem 6.9 formulas.



The covering ellipsoid has one axis of length r along e_1 and the $n-1$ remaining axes of length s along $e_1^\perp$, shifted by β from the origin.

Ellipsoid Search and Optimization Complexity

Theorem 6.10. Starting from $E_0 \supseteq K$ and applying Theorem 6.9 at each step:

$$\mathrm{vol}(E_T) \le e^{-T/(4n)} \,\mathrm{vol}(E_0).$$

If $\mathrm{vol}(K) \ge \eta$, then $\lfloor 4n \log(\mathrm{vol}(E_0)/\eta) \rfloor + 1$ queries suffice.

Corollary (Optimization). With target K_δ:

$$T = \lfloor 4n \log(\mathrm{vol}(E_0)/\,\mathrm{vol}(X)) + 4n^2 \log(\Delta/\delta) \rfloor + 1.$$

Comparison with center-of-mass: $O(n \log(\Delta/\delta))$ vs. $O(n^2 \log(\Delta/\delta))$. Extra factor of n from the weaker shrinkage.

Summary: Two Methods, Same Template

	Center-of-mass	**Ellipsoid**
Query point	center(S_t)	center of E_t
Update S_{t+1}	$S_t \cap H_t$	ellipsoid of Thm 6.9
Shrinkage per step	$1 - e^{-1}$	$e^{-1/(4n)}$
Representation	*general* convex body	$O(n^2)$ parameters
Tractable?	*expensive*	**yes**
Opt. iterations	$O(n \log(\Delta/\delta))$	$O(n^2 \log(\Delta/\delta))$

Tradeoff: Ellipsoid gives up one factor of n in the iteration count to gain tractable representation and an explicit update formula. This is the first practical polynomial-time cutting-plane method in high dimension.

Historical Remark

Center-of-mass method. Introduced independently by **Levin (1965)** in the Soviet literature and **Newman (1965)** in the West.

Ellipsoid method.

- **Yudin & Nemirovski (1976)**: "Informational complexity and effective methods for the solution of convex extremal problems"
- **Shor (1977)**: "Cut-off method with space extension in convex programming problems"
- **Khachiyan (1979)**: first polynomial-time algorithm for linear programming
- **Grötschel, Lovász, Schrijver (1988)**: *Geometric Algorithms and Combinatorial Optimization*

Summary & What's Next

Today:

- Cutting-plane framework: query $\to$ cut $\to$ shrink
- Volume threshold principle: match shrinkage with $\text{vol}(K)$ lower bound
- Grünbaum: center-of-mass cut keeps $\geq e^{-1}$ volume fraction
- Center-of-mass method: $O(n \log(\Delta/\delta))$ iterations
- Ellipsoid method: $O(n^2 \log(\Delta/\delta))$, but tractable
- Sublevel-set reduction: optimization $\to$ feasibility for K_δ

Next lecture:

- First-order methods: (sub)gradient descent
- From oracle complexity to per-iteration cost