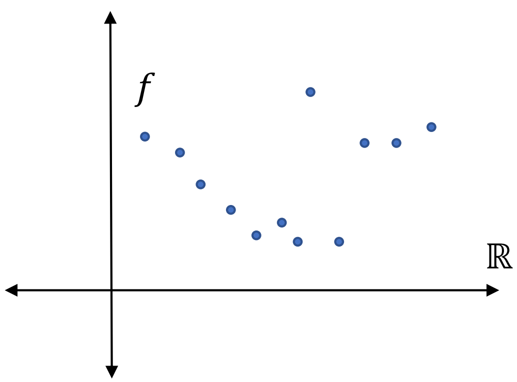
Introduction to Optimization Theory

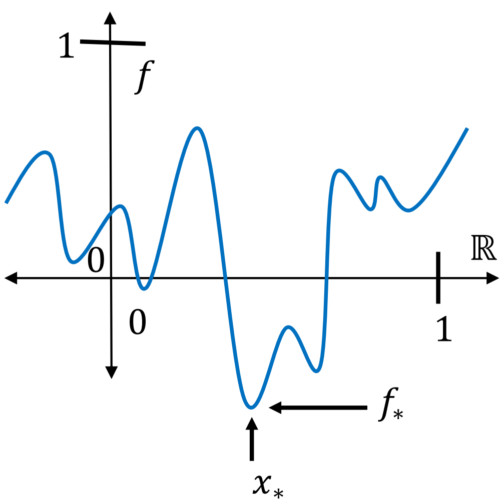
Lecture #18 - 11/16/20

MS&E 213 / CS 2690



Aaron Sidford

sidford@stanford.edu



Plan for Today

Recap
- Rates, subgradients, FTRL, and mirror descent

Feasibility Problem
- An abstraction for separation oracles

Cutting Plane Methods
- Methods for solving feasibility problem

Interior Point Methods

How to obtain improved rates when more structure.

Recap

Problem

$$\min_{x\in\mathbb{R}^n} f(x)$$

Regularity	Oracle	Goal	Algorithm	Iterations
$n=1$, $f(x)\in[0,1]$, $x_*\in[0,1]$	value	½-optimal	anything	∞
$n=1$, $x_*\in[0,1]$, L-Lipschitz	value	ϵ-optimal	ϵ-net	$\Theta(L/\epsilon)$
$x_*\in[0,1]$, L-Lipschitz in $\|\cdot\|_\infty$	value	ϵ-optimal	ϵ-net	$(\Theta(L/\epsilon))^n$
L-smooth and bounded	value, gradient	ϵ-optimal	ϵ-net	exponential
L-smooth	gradient	ϵ-critical	gradient descent	$O(L(f(x_0)-f_*)\epsilon^{-2})$
L-smooth μ-strongly convex	gradient	ϵ-optimal	gradient descent	$O((L/\mu)\log([f(x_0)-f_*]/\epsilon))$
L-smooth convex	gradient	ϵ-optimal	gradient descent	$O(L\|x_0-x_*\|_2^2/\epsilon)$
L-smooth μ-strongly convex	gradient	ϵ-optimal	gradient descent	$O(\sqrt{L/\mu}\log([f(x_0)-f_*]/\epsilon))$
L-smooth μ-strongly convex	gradient	ϵ-optimal	gradient descent	$O\left(\sqrt{L\|x_0-x_*\|_2^2/\epsilon}\right)$
L-Lipschitz, convex	subgradient	ϵ-optimal	Mirror descent, FTRL	$O(L^2\|x_0-x_*\|_2^2/\epsilon^2)$

Convex Function Oracle

Approach

- Reduce to feasibility problem
- Solve feasibility problem

query
$x \in \mathbb{R}^n$ → **separation oracle** →

output
$g \in \mathbb{R}^n$ such that
$\mathrm{level}_{\leq}\big(f, f(x)\big) \subseteq H_{\geq}\,(g, g^\top x)$
With $g = 0$ only if $f_* = f(x)$

Cutting Plane Methods

query
$x \in \mathbb{R}^n$ → **subgradient oracle** ✓ →

output
$g \in \partial f(x)$

Subgradient descent

Subgradient: *g is subgradient of f at x (i.e. $g \in \partial f(x)$) if and only if $f(y) \geq f(x) + g^\top (y - x)$ for all $y \in \mathbb{R}^n$*

Approach

- Reduce to online linear optimization
- Solve online linear optimization

Plan for Today

Recap ✓

- Rates, subgradients, FTRL, and mirror descent

Feasibility Problem

- An abstraction for separation oracles

Cutting Plane Methods

- Methods for solving feasibility problem

Interior Point Methods

How to obtain improved rates when more structure.

Convex Function Oracle

Approach
- Reduce to feasibility problem
- Solve feasibility problem

query
$x \in \mathbb{R}^n$ → **separation oracle** →

output
$g \in \mathbb{R}^n$ such that
$\text{level}_{\leq}\big(f, f(x)\big) \subseteq H_{\geq}\,(g, g^{\top}x)$
With $g = 0$ only if $f_* = f(x)$

Cutting Plane Methods

query
$x \in \mathbb{R}^n$ → **subgradient oracle** →

output
$g \in \partial f(x)$

Subgradient descent

Subgradient*: g is subgradient of f at x (i.e. $g \in \partial f(x)$) if and only if $f(y) \geq f(x) + g^{\top}(y - x)$ for all $y \in \mathbb{R}^n$*

Approach
- Reduce to online linear optimization
- Solve online linear optimization

Picture

Geometric problem abstraction?

query

$x \in \mathbb{R}^n$

separation oracle

output

$g \in \mathbb{R}^n$ such that

$\text{level}_{\leq}\big(f, f(x)\big) \subseteq H_{\geq}\,(g, g^\top x)$

With $g = 0$ only if $f_* = f(x)$

Analogous role to online linear optimization problem

Box of radius R: $\{x \in \mathbb{R}^n \mid \|x - c\|_\infty \leq R\}$

(R, r, n)-Feasibility Problem

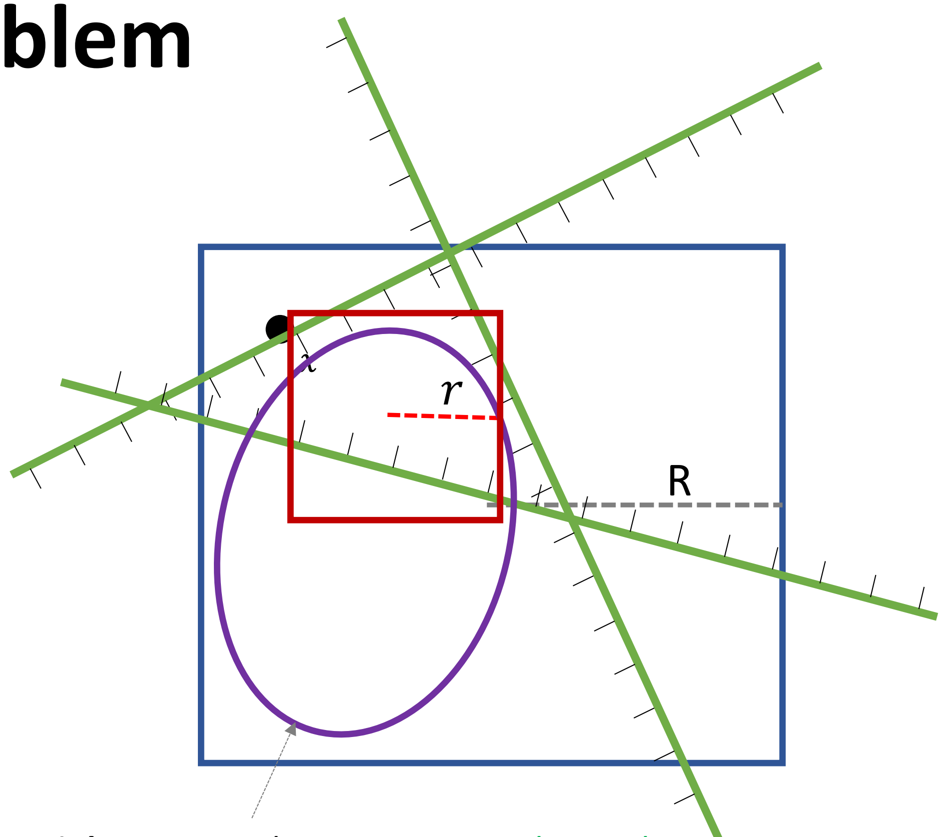
Given:

- n-dimensional box B of radius R:
- **Separation oracle**: when evaluated at point $x \in B$ in time T either outputs "success" or a halfspace with x on boundary.

Goal:

- Get "success" or
- Prove that the intersection of halfspaces and initial box does not contain a box of radius r

Note: Want $O(poly(n, T, \log(R/r))$



***Special Case**: we obtain separating hyperplanes for some convex set the is contain in a Box of radius R and contains a ball of radius r.*

Analogous role to online linear optimization problem

Box of radius R: $\{x \in \mathbb{R}^n \mid \|x - c\|_\infty \leq R\}$
$B_\infty(R) = \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq R\}$

$(\boldsymbol{R}, \boldsymbol{r}, \boldsymbol{n})$-Feasibility Problem

Given:

- n-dimensional box B of radius R:
- **Separation oracle**: when evaluated at point $x \in B$ in time T either outputs "success" or a halfspace with x on boundary.

Goal:

- Get "success" or
- Prove that the intersection of halfspaces and initial box does not contain a box of radius r

Solution Concept:

- An algorithm is a (T_q, T_t)-solution if it solves the problem with $O(T_q)$ queries and $O(T_t)$ time.

Note

- Choice of norms somewhat arbitrary
- Consider solutions which depend polylogarithmically on nR/r and therefore, norm choice.

Why?

query $x \in \mathbb{R}^n$ → **separation oracle** → output $g_x \in \mathbb{R}^n$ with $\text{level}_{\leq}(f, f(x)) \subseteq H_{\geq}(g_x, g_x^\top x)$

With $g_x = \vec{0}$ only if $f_* = f(x)$

Lemma: Suppose $f: \mathbb{R}^n \to \mathbb{R}$ obtains its minimum value at $x_* \in B_\infty(R)$ and there is a box of radius r where every point in it is ϵ-optimal, then given a (T_q, T_t)-solution to the (R, r, n)-feasibility problem, can compute an ϵ-optimal point with $O(T_q)$ queries to a separation oracle and value oracle for f and $O(T_t)$-time.

Proof: apply solution with separation oracle as oracle and output point queried which has minimum value.

- If query x and $g_x = \vec{0}$ then output minimizer!
- Otherwise, since intersection doesn't contain r-box, have for some x it is the case that $H_{\geq}(g_x, g_x^\top x)$ doesn't contain all ϵ-optimal points. Since $\text{level}_{\leq}(f, f(x)) \subseteq H_{\geq}(g_x, g_x^\top x) \Rightarrow x$ is ϵ-optimal

query
$x \in \mathbb{R}^n$ → **separation oracle** → output
$g_x \in \mathbb{R}^n$ with $\text{level}_{\leq}\big(f, f(x)\big) \subseteq H_{\geq}\,(g_x, g_x^\top x)$
With $g_x = \vec{0}$ only if $f_* = f(x)$

Constraints

Note

- Can apply a similar approach to constrained optimization if have separation oracle for constraint set S, i.e. an oracle which when queried outputs $\vec{0}$ if $x \in S$ and $g \neq \vec{0}$ with $S \subseteq H_{\geq}(g, g^\top x)$ when $x \notin S$.
- **Idea**: apply feasibility solution and output separation oracle for S when queried point $\notin S$ and separation oracle for f otherwise.

Ball of radius R in $\|\cdot\|$: $B_{\|\cdot\|}(R,c)=\{x\in\mathbb{R}^n \mid \|x-c\|\le R\}$

Size of set of ϵ-opt points?

Lemma: If $f:\mathbb{R}^n\to\mathbb{R}$ convex and $\epsilon>0$ and the set of ϵ-optimal points contains a ball of radius R in $\|\cdot\|$ then for all $\alpha\in(0,1)$ the set of $\alpha\epsilon$-points contains a ball of radius αR in that norm.

Proof

- By assumption, there is x_0 with $B_{\|\cdot\|}(R,x_0)\subseteq \mathrm{level}_{\le}(f,f_*+\epsilon)$
- Let $S=B_{\|\cdot\|}(\alpha R,\alpha x_0+(1-\alpha)x_*)$
- If $x\in S$ then $\|x-\alpha x_0-(1-\alpha)x_*\|\le\alpha R$ and $x=\alpha y+(1-\alpha)x_*$ where $\|\alpha(y-x_0)\|\le\alpha R$, i.e. $y\in B_{\|\cdot\|}(R,x_0)$
- $\Rightarrow f(x)\le\alpha f(y)+(1-\alpha)f(x_*)\le\alpha\cdot(f_*+\epsilon)+(1-\alpha)f_*=f_*+\alpha\epsilon$

Can apply to smooth, strongly convex functions!

Minimizing Bounded Convex Functions

Lemma: Suppose $f: \mathbb{R}^n \to \mathbb{R}$ is differentiable, convex, minimized in $B_\infty(R)$, and has $|f(x)| \leq M$ for al $x \in B_\infty(R)$. Then given a first order oracle for f and a (T_q, T_t)-solution to the $(R, \frac{\epsilon R}{2M}, n)$-feasibility problem can compute an ϵ-optimal point with $O(T_q)$-queries and $O(T_t)$-time.

Proof: run solution with -gradient oracle and output queried point of min value

- -gradient oracle is separation oracle
- Every point in $B_\infty(R)$ is $2M$-optimal
- The set of ϵ-optimal points contain a box of radius $\frac{\epsilon R}{2M}$ by preceding lemma

Plan for Today

- **Recap** ✓ — Rates, subgradients, FTRL, and mirror descent
- **Feasibility Problem** ✓ — An abstraction for separation oracles
- **Cutting Plane Methods** — Methods for solving feasibility problem

Interior Point Methods

How to obtain improved rates when more structure.

Analogous role to online linear optimization problem

Box of radius R: $\{x \in \mathbb{R}^n \mid \|x - c\|_\infty \leq R\}$
$B_\infty(R) = \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq R\}$

$(\boldsymbol{R}, \boldsymbol{r}, \boldsymbol{n})$-Feasibility Problem

Given:

- n-dimensional box B of radius R:
- **Separation oracle**: when evaluated at point $x \in B$ in time T either outputs "success" or a halfspace with x on boundary.

Goal:

- Get "success" or
- Prove that the intersection of halfspaces and initial box does not contain a box of radius r

Solution Concept:

- An algorithm is a (T_q, T_t)-solution if it solves the problem with $O(T_q)$ queries and $O(T_t)$ time.

Algorithm Idea?

- **Warmup:** 1-dimension?

One Dimensional Problem

Theorem: Can solve the $(R, r, 1)$-problem using $O\left(\log\left(\frac{R}{r}\right)\right)$ queries.

n-Dimensional Problem

Grunbaum's Theorem: Intersecting convex body with halfspace through center of gravity decreases volume by at least $(1 - 1/e)$

Algorithm: repeatedly query center of gravity of intersection of box and all halfspaces

Analysis: volume starts at $(2R)^n$ and if volume $\leq (2r)^n$ algorithm ends

Theorem: Can solve the (R, r, n)-problem using $O\left(n \cdot \log\left(\frac{R}{r}\right)\right)$ queries

Algorithm Framework

- Start with "set" containing the box

What set?

- Get separating hyperplane through "center"

What center?

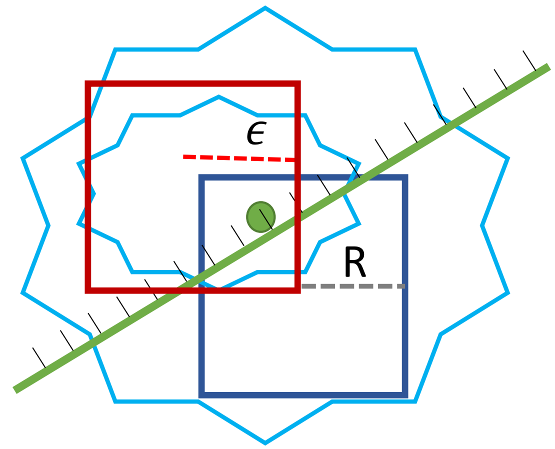
- Update set (must contain intersection of halfspaces)

How update?

- Repeat until "size" is small

How measure size?

- **Given**: n-dimension box of radius R
- **Given**: Separating oracle takes time T
- **Goal**: Prove intersection doesn't contain box radius r
- **Notation**: $\kappa \stackrel{\text{def}}{=} nR/r$



Goal:

- Minimize total running time:
  - Number of times call oracle (oracle complexity)
  - Additional running time (overhead)

Ellipsoid Method

- Given: n-dimension box of radius R
- Given: Separating oracle takes time T
- Goal: Prove intersection doesn't contain box radius r
- Notation: $\kappa \stackrel{\text{def}}{=} nR/r$

- Start with "set" containing the box

Ellipsoid

- Get separating hyperplane through "center"

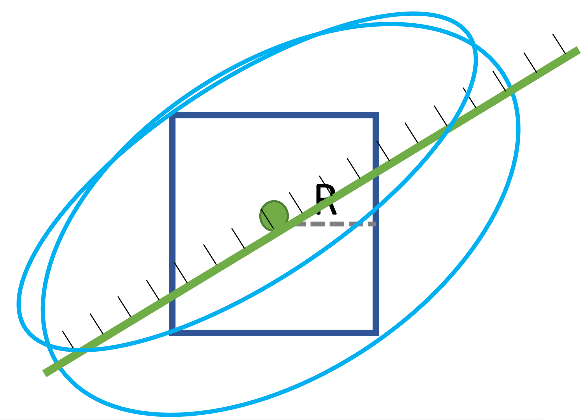
Center of ellipsoid

- Update set (must contain intersection of halfspaces)

$O(n^2)$

- Repeat until "size" is small

Volume of ellipsoid decreases by $(1 - 1/n)$ each iteration



Result

- Time $O(n^2 T \log \kappa + n^4 \log \kappa)$
- Great cost per iteration
- Bad oracle complexity

Center of Gravity

- Given: n-dimension box of radius R
- Given: Separating oracle takes time T
- Goal: Prove intersection doesn't contain box radius r
- Notation: $\kappa \stackrel{\text{def}}{=} nR/r$

- Start with "set" containing the box

Intersection of all halfspaces

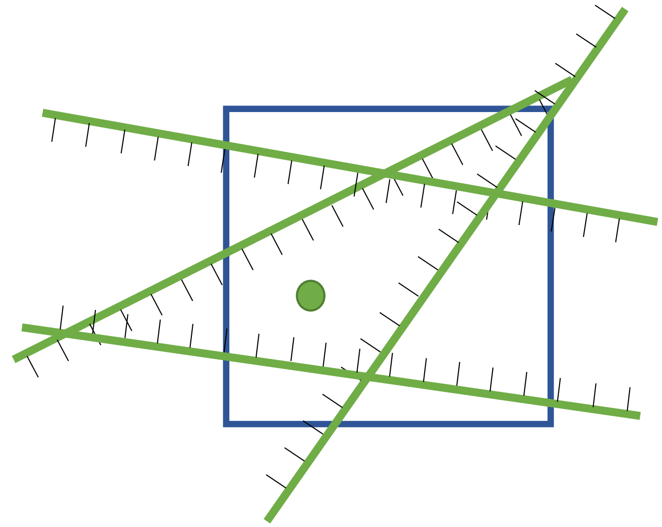
- Get separating hyperplane through "center"

Center of gravity

- Update set (must contain intersection of halfspaces)

Can approximate center in poly time

- Repeat until "size" is small



Grunbaum's Theorem

Intersecting convex body with halfspace through center of gravity decreases volume by at least $(1 - 1/e)$.

Result

- Time $O(nT\ log\ \kappa\ + (n\ log\ \kappa)^{O(1)})$
- Great oracle complexity
- Bad cost per iteration

John Ellipse

- Start with “set” containing the box

Intersection of all halfspaces

- Get separating hyperplane through “center”

Center of maximum volume ellipse in set

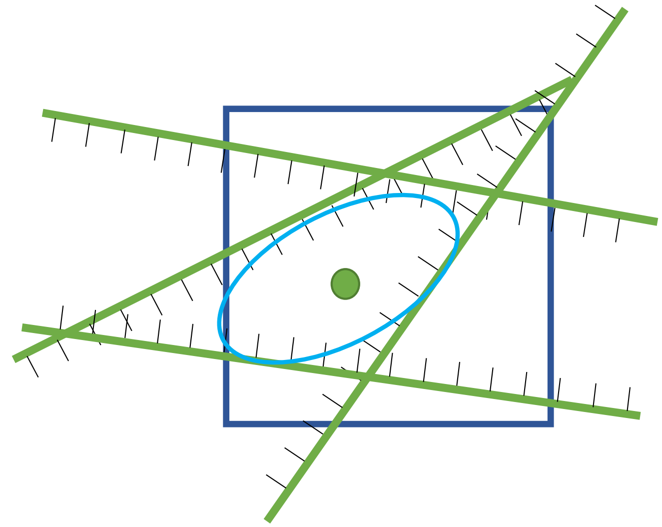
- Update set (must contain intersection of halfspaces)

Polynomial time

- Repeat until “size” is small

Volume of maximum volume ellipse decreases by constant.

- **Given**: n-dimension box of radius R
- **Given**: Separating oracle takes time T
- **Goal**: Prove intersection doesn’t contain box radius r
- **Notation**: $\kappa \stackrel{\text{def}}{=} nR/r$



Result

- Time $O(nT \log \kappa + (n \log \kappa)^{O(1)})$
- Great oracle complexity
- Better but still high iteration cost

Practice : Analytic Center

- Start with "set" containing the box

Intersection of some halfspace

- Get separating hyperplane through "center"

Barrier – analytic center

- Update set (must contain intersection of halfspaces)

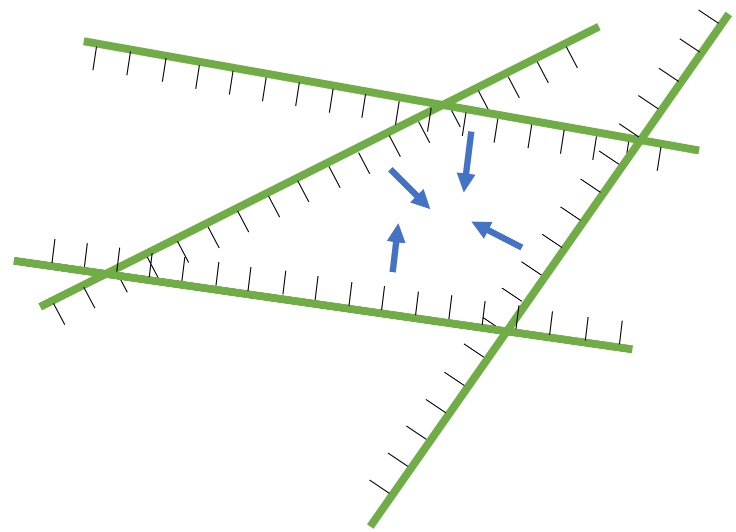
Fast - comparable to interior point

- Repeat until "size" is small

Sometimes good in practice, various theoretical guarantees, hard to analyze

Barrier

A nice function from interior of set to the reals that goes to infinity as you approach boundary of set



- For polytope $P = \{x : \boldsymbol{A}x \geq b\}$
- Let $s_x = \boldsymbol{A}x - b$
- Log barrier $\mathrm{b_l}(x) = -\sum_i \log s_x(i)$
- Analytic center $\underset{x}{\operatorname{argmin}}\, b_l\,(x)$

Volumetric center

- Start with "set" containing the box

Intersection of some halfspace

- Get separating hyperplane through "center"

Volumetric center

- Update set (must contain intersection of halfspaces)

Matrix multiplication time

- Repeat until "size" is small

Dropping constraint decreases $b_v(x)$ by small constant but adding constant increases $b_v(x)$ by larger constant.

Fact

Volume of ellipse is approximately volume of set at center if not too many constraints.

Algorithm [V89]

- If constraint has high *leverage score*

$$[S_x A (A S_x^{-2} A)^{-1} A^\top S_x]_{ii}$$

drop it and compute new center.

- Otherwise, add call oracle at center, add nearby, and re-center.
- Repeat

- F
- L
- V
- Volumetric center $\arg\min_x b_v(x)$

Result

- Time $O(nT \log \kappa + n^{\omega+1} \log \kappa)$

$\omega < 2.373$ [W14]

Cutting Plane Methods

- Maintain a convex set
- Query separation oracle at "center" of convex set
- Update convex set
- Repeat until "size" is sufficiently small

Method	Set	Center	Size Tracked	Time ($\kappa = nR/\epsilon$)
Ellipsoid [YN76,S77,K80]	Ellipse	Ellipse	Volume	$O(n^2 T \log \kappa + n^4 \log \kappa)$
Center of Gravity [L65,BV02]	All half-spaces	Gravity	Volume	$O\left(nT \log \kappa + n^5 \log^{O(1)} \kappa\right)$
Inscribed Ellipse [KTE88,NN92]	All half-spaces	John Ellipse	Volume	$O(nT \log \kappa + n^{\omega+1.5} \log^{O(1)} \kappa)$
Volumetric Center [V89]	Some half-spaces	Volumetric	Volume, width	$O(nT \log \kappa + n^{\omega+1} \log \kappa)$
Analytic Center [AV95]	Some half-spaces	Analytic	Volume, width	$O(nT \log^{O(1)} \kappa + n^{O(1)} \log^{O(1)} \kappa)$

$\omega < 2.373$ [W14]

Cutting Plane Methods

- Maintain a convex set
- Query separation oracle at “center” of convex set
- Update convex set
- Repeat until “size” is sufficiently small

Method	Set	Center	Size Tracked	Time ($\kappa = nR/\epsilon$)
Ellipsoid [YN76,S77,K80]	Ellipse	Ellipse	Volume	$O(n^2 T \log\kappa + n^4 \log\kappa)$
Center of Gravity [L65,BV02]	All half-spaces	Gravity	Volume	$O\left(nT \log\kappa + n^5 \log^{O(1)}\kappa\right)$
Inscribed Ellipse [KTE88,NN92]	All half-spaces	John Ellipse	Volume	$O(nT \log\kappa + n^{\omega+1.5} \log^{O(1)}\kappa)$
Volumetric Center [V89]	Some half-spaces	Volumetric	Volume, width	$O(nT \log\kappa + n^{\omega+1} \log\kappa)$
Analytic Center [AV95]	Some half-spaces	Analytic	Volume, width	$O(nT \log^{O(1)}\kappa + n^{O(1)} \log^{O(1)}\kappa)$
[LSW15]	**Some half-spaces**	**Hybrid**	**Volume, width**	$\boldsymbol{O(nT \log\kappa + n^3 \log^{O(1)}\kappa)}$
[JLSW20]	**Some half-spaces**	**Volumetric**	**Volume, width**	$\boldsymbol{O\left(nT \log\kappa + n^3 \log\kappa\right)}$

- Implication: L-smooth, μ-strongly convex ϵ-opt given ϵ_0-initial error with $O\left(n \log\left(\frac{nL\epsilon_0}{\mu\epsilon}\right)\right)$ gradient evaluations
- Implication: faster matroid intersection, semidefinite programming, submodular optimization

$\omega < 2.373$ [W14]

Plan for Today

Recap

- Rates, subgradients, FTRL, and mirror descent

Feasibility Problem

- An abstraction for separation oracles

Cutting Plane Methods

- Methods for solving feasibility problem

Interior Point Methods

How to obtain improved rates when more structure.