Convex Optimization

Prof. Nati Srebro

Lecture 14:

The Ellipsoid Method

Bubeck Section 2.2

Nemirovskii “Information Based Complexity” Sections 2.3, 3.2

Optional further reading: Bubeck Section 2.3

$$\begin{array}{lll} \min\limits_{x\in\mathbb{R}^n} & f_0(x) & \\ s.t. & f_i(x) \le 0 & i = 1..m \end{array}$$

Interior Point (Log Barrier) Method:

- Access to 2nd order oracle for f_0, f_i

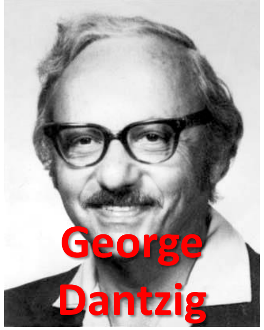
$$x \mapsto f_i(x), \nabla f_i(x), \nabla^2 f_i(x)$$

- If f_i quadratic and f_0 self conc.:
  - Number of Oracle Accesses: $O(\sqrt{m} \log {}^1\!/_\epsilon)$ each f_0, f_i
  - Runtime: $O(\sqrt{m}(n^3 + m\nabla^2) \log {}^1\!/_\epsilon)$
- Inequalities strictly satisfied, converge to x^* from interior
- $(x^{(k)}, \lambda^{(k)})$ satisfy KKT except complementary slackness

Simplex:

- f_0, f_i linear (explicit access, or equiv. 1st order oracle)
- Runtime exponential in worst case; poly-time smoothed analysis
- Work on extreme points, converge to x^* along boundary
- $(x^{(k)}, \lambda^{(k)})$ satisfy KKT except $\lambda \ge 0$

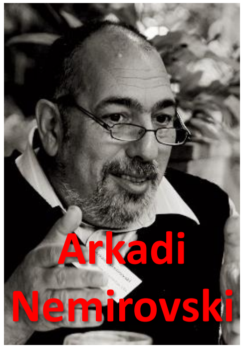
Formulation of LP



Simplex

A Levin
D Newman

Center of Mass



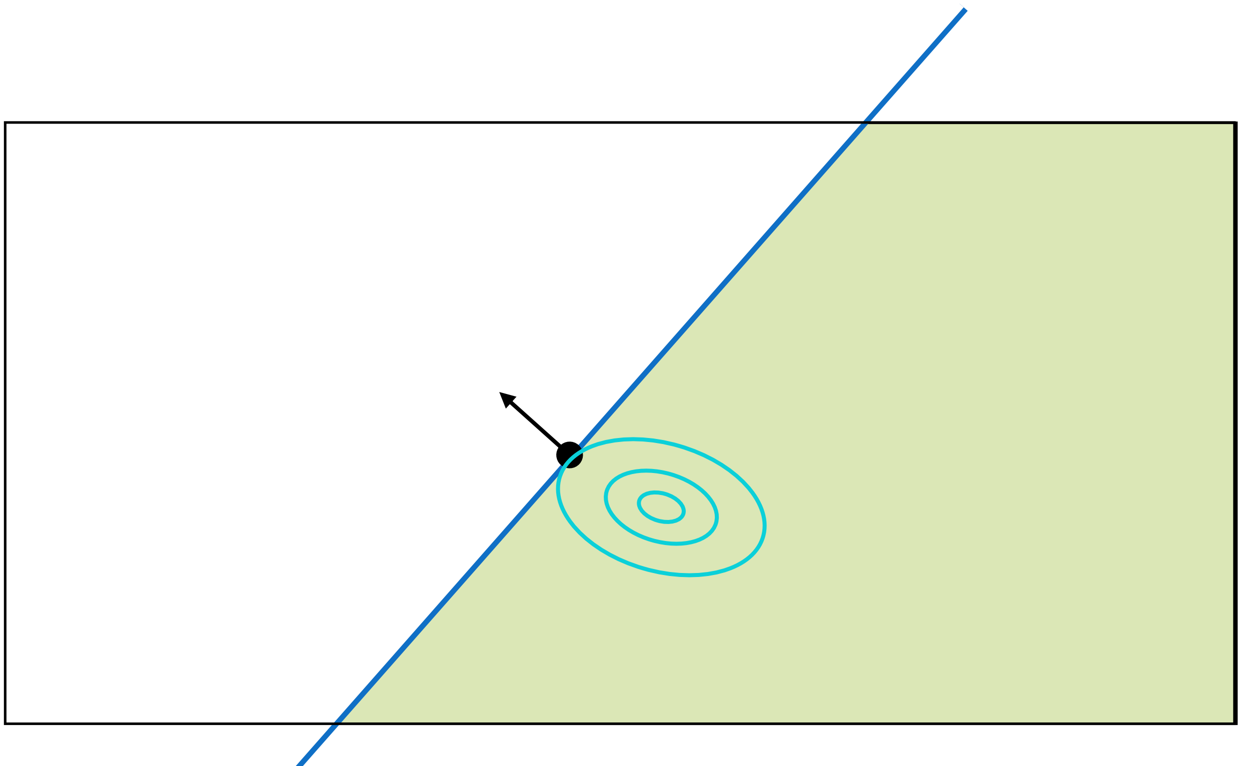
Ellipsoid Method

1939 1947 1965 1972

Center-of-Mass Method

$$\boxed{\min f(x)}$$

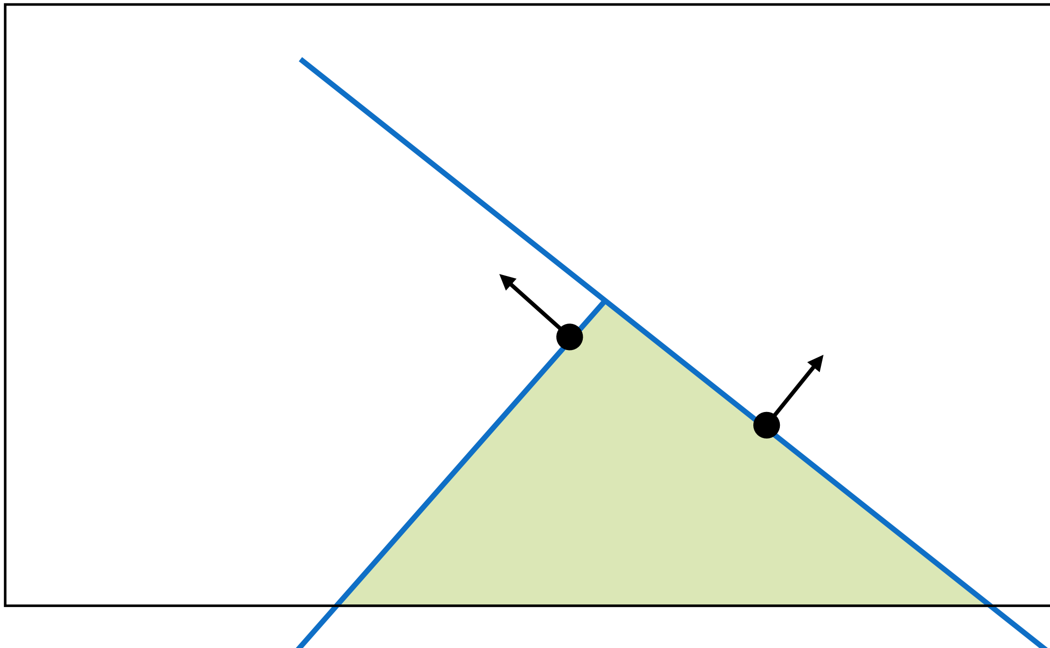
Init	convex $G^{(0)}$ s.t. $x^* \in G^{(0)}$
Iterate	$x^{(k)} =$ center of mass of $G^{(k)}$
	$g^{(k)} = \nabla f_0\left(x^{(k)}\right)$, which implies $\langle g^{(k)}, x^* - x^{(k)} \rangle < 0$
	$G^{(k+1)} \leftarrow G^{(k)} \cap \left\{x \middle| \langle g^{(k)}, x - x^{(k)} \rangle\right\}$



Center-of-Mass Method

$$\boxed{\min f(x)}$$

Init	convex $G^{(0)}$ s.t. $x^* \in G^{(0)}$
Iterate	$x^{(k)}$ = center of mass of $G^{(k)}$
	$g^{(k)} = \nabla f_0\left(x^{(k)}\right)$, which implies $\langle g^{(k)}, x^* - x^{(k)} \rangle < 0$
	$G^{(k+1)} \leftarrow G^{(k)} \cap \left\{x \middle| \langle g^{(k)}, x - x^{(k)} \rangle\right\}$



Center-of-Mass Method

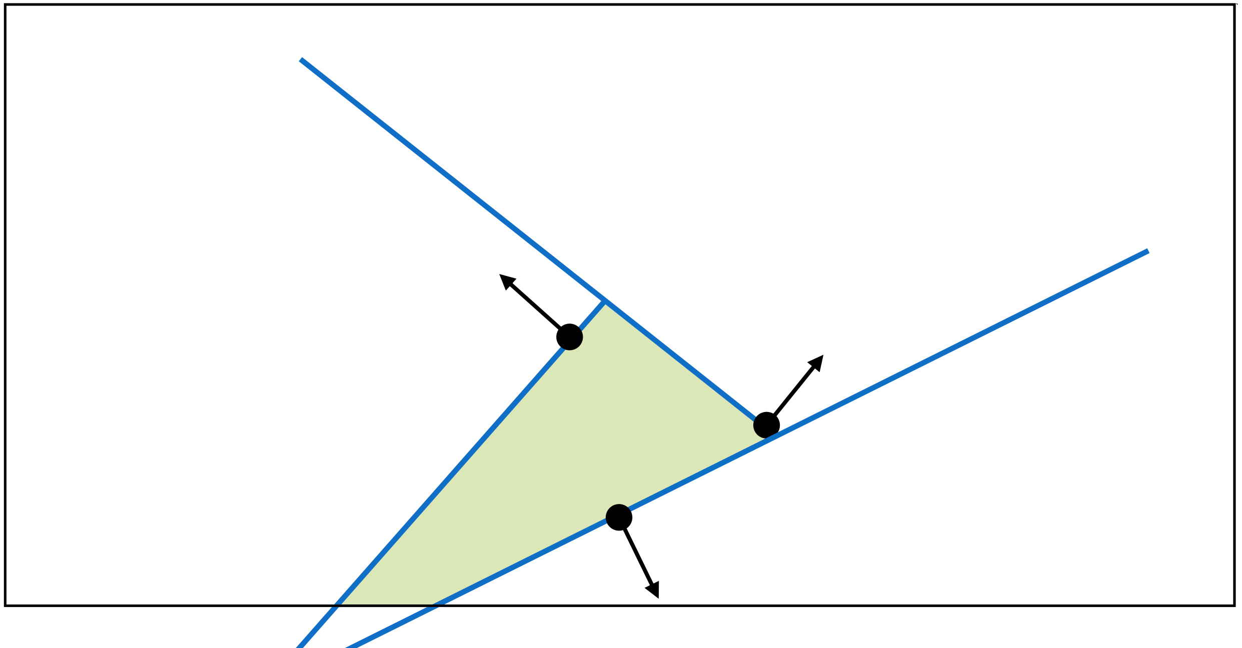
$$\boxed{\min f(x)}$$

Init convex $G^{(0)}$ s.t. $x^* \in G^{(0)}$

Iterate $x^{(k)} =$ center of mass of $G^{(k)}$

$g^{(k)} = \nabla f_0\left(x^{(k)}\right)$, which implies $\langle g^{(k)}, x^* - x^{(k)}\rangle < 0$

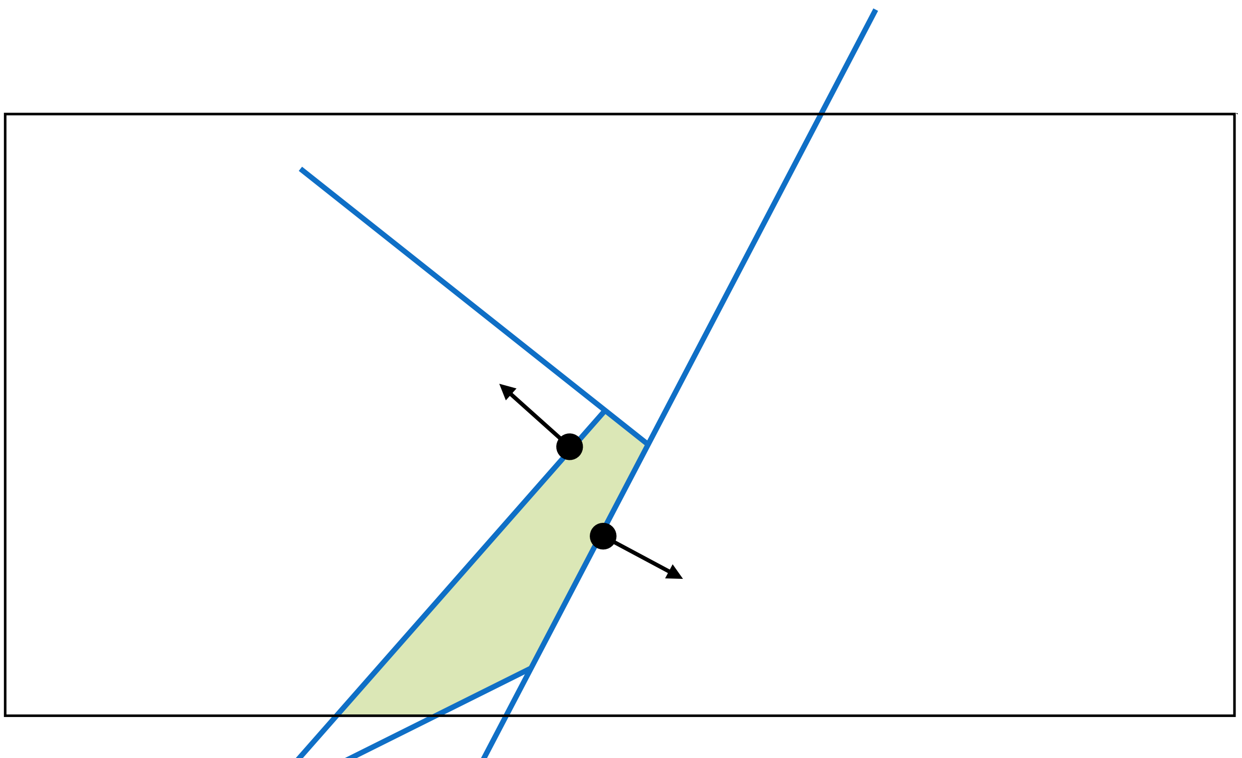
$G^{(k+1)} \leftarrow G^{(k)} \cap \left\{x \middle| \langle g^{(k)}, x - x^{(k)}\rangle\right\}$



Center-of-Mass Method

$$\boxed{\min f(x)}$$

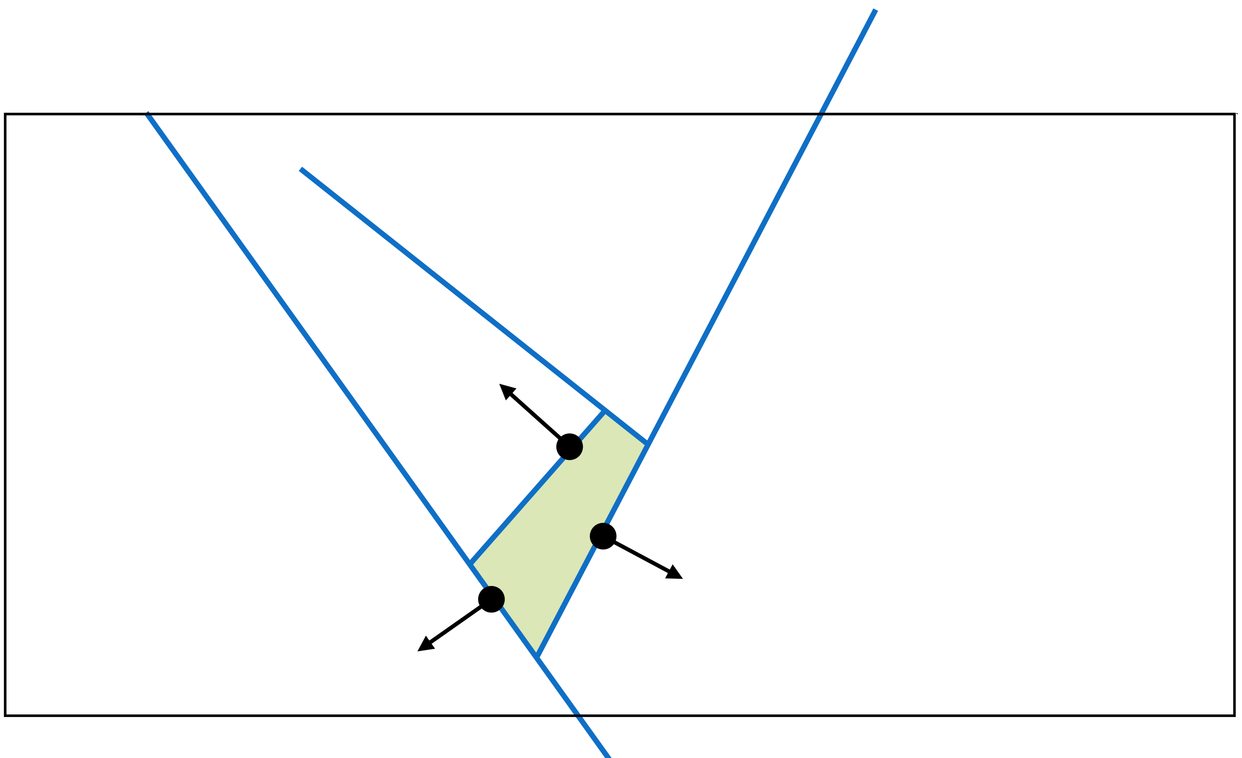
Init	convex $G^{(0)}$ s.t. $x^* \in G^{(0)}$
Iterate	$x^{(k)} =$ center of mass of $G^{(k)}$
	$g^{(k)} = \nabla f_0\left(x^{(k)}\right)$, which implies $\langle g^{(k)}, x^* - x^{(k)}\rangle < 0$
	$G^{(k+1)} \leftarrow G^{(k)} \cap \left\{x \middle\| \langle g^{(k)}, x - x^{(k)}\rangle\right\}$



Center-of-Mass Method

$$\boxed{\min f(x)}$$

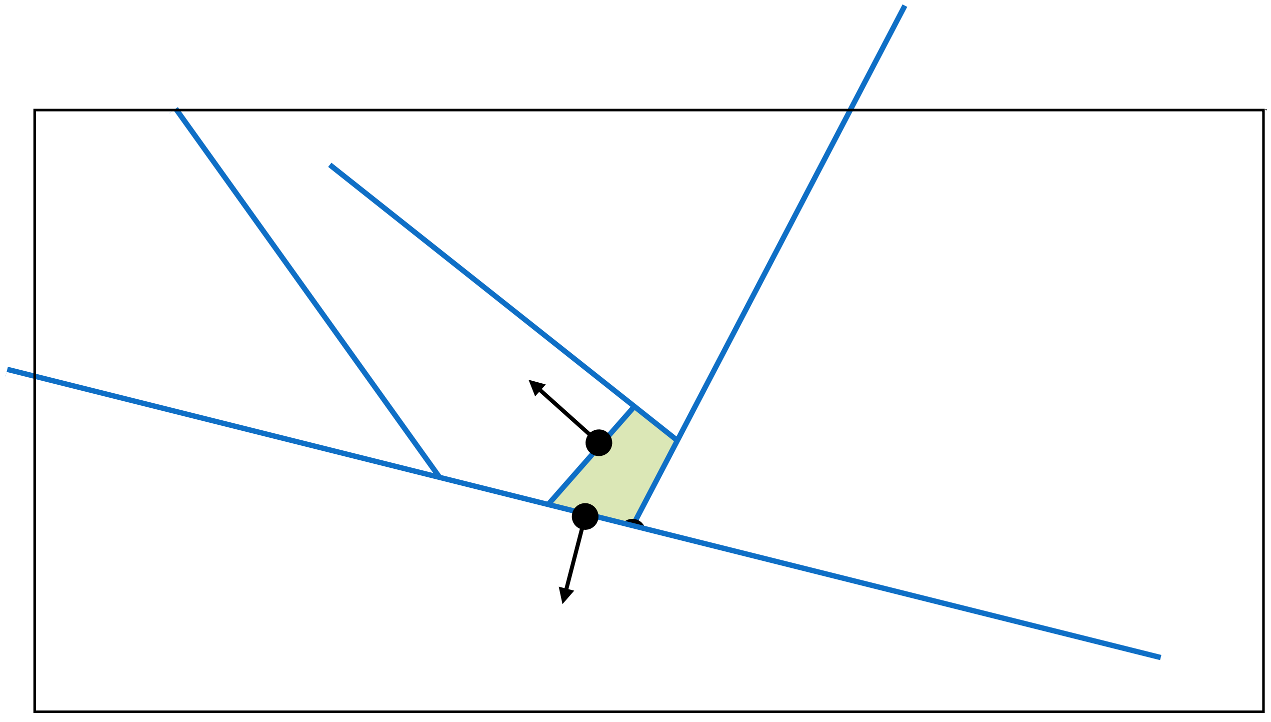
Init	convex $G^{(0)}$ s.t. $x^* \in G^{(0)}$
Iterate	$x^{(k)} =$ center of mass of $G^{(k)}$
	$g^{(k)} = \nabla f_0\left(x^{(k)}\right)$, which implies $\langle g^{(k)}, x^* - x^{(k)}\rangle < 0$
	$G^{(k+1)} \leftarrow G^{(k)} \cap \left\{x \middle\| \langle g^{(k)}, x - x^{(k)}\rangle\right\}$



Center-of-Mass Method

$$\boxed{\min f(x)}$$

Init	convex $G^{(0)}$ s.t. $x^* \in G^{(0)}$
Iterate	$x^{(k)}$ = center of mass of $G^{(k)}$
	$g^{(k)} = \nabla f_0\left(x^{(k)}\right)$, which implies $\langle g^{(k)}, x^* - x^{(k)} \rangle < 0$
	$G^{(k+1)} \leftarrow G^{(k)} \cap \left\{x \middle| \langle g^{(k)}, x - x^{(k)} \rangle\right\}$



Center-of-Mass Method

$$\boxed{\min f(x)}$$

Init	convex $G^{(0)}$ s.t. $x^* \in G^{(0)}$
Iterate	$x^{(k)} =$ center of mass of $G^{(k)}$
	$g^{(k)} = \nabla f_0\left(x^{(k)}\right)$, which implies $\langle g^{(k)}, x^* - x^{(k)}\rangle < 0$
	$G^{(k+1)} \leftarrow G^{(k)} \cap \left\{x \middle| \langle g^{(k)}, x - x^{(k)}\rangle\right\}$
Return	$\tilde{x} = \arg\min_{i=0..k} f\left(x^{(k)}\right)$

- Granbaum: for any half-plane H through center of G

$$Vol_n(G \cap H) \leq \left(1 - \frac{1}{e}\right) Vol_n(G) \leq 0.64 Vol_n(G)$$

➔ $Vol_n\left(G^{(k)}\right) \leq 0.64^k\, Vol\left(G^{(0)}\right)$

- Claim: If $x^* \in A \subseteq G, \forall_{x\in G} |f(x)| \leq B$ then $\min_{x\in(G-A)} f(x) \leq f(x^*) + 2B \sqrt[n]{\frac{Vol_n(A)}{Vol_n(G)}}$
- Claim: $\tilde{x}^{(k)} \leq \inf_{x\in\left(G^{(0)} - G^{(k+1)}\right)} f(x)$

Conclusion: If we start with $G^{(0)}$ s.t. $x^* \in G^{(0)}$ and $\sup_{x\in G^{(0)}} f(x) \leq B$ then :

$$\min_{j=0..k} f\left(x^{(j)}\right) \leq f(x^*) + 2B(0.64)^{k/n} \;➔\; k = 2.2\, n \cdot \log B/\epsilon \text{ iterations}$$

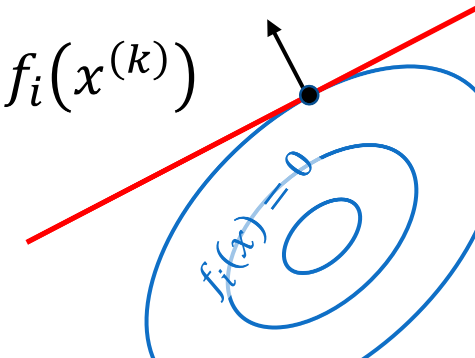
Cutting Planes with Constraints

$$\min_{x\in\mathbb{R}^n} \quad f_0(x)$$
$$s.t. \quad f_i(x) \le 0 \qquad i = 1..m$$

- If $x^{(k)}$ is not an optimum, we want $H = \{\langle g, x - x^{(k)}\rangle < 0\}$ s.t. $x^* \in H$

- If $x^{(k)}$ is feasible:
$$f_0\left(x^{(k)}\right) + \langle \nabla f_0\left(x^{(k)}\right), x^* - x^{(k)}\rangle \le f_0\ (x^*) < f_0\left(x^{(k)}\right)$$
➔ use $g = \nabla f_0\left(x^{(k)}\right)$

- If $x^{(k)}$ is not feasible, $f_i\left(x^{(k)}\right) > 0$ and so
$$f_i\left(x^{(k)}\right) + \langle \nabla f_i\left(x^{(k)}\right), x - x^{(k)}\rangle \le f_i(x^*) \le 0 < f_i\left(x^{(k)}\right)$$
➔ use $g = \nabla f_i\left(x^{(k)}\right)$



Center of Mass with Constraints

$$\min_{x \in \mathbb{R}^n} \quad f_0(x)$$
$$s.t. \quad f_i(x) \le 0 \qquad i = 1..m$$

Init $\quad G^{(0)}$

Iterate $\quad x^{(k)} =$ center of mass of $G^{(k)}$

If $\exists_i \; f_i\left(x^{(k)}\right) \ge \epsilon$, then $g^{(k)} = \nabla f_i\left(x^{(k)}\right)$

Else, $g^{(k)} = \nabla f_0\left(x^{(k)}\right)$

$G^{(k+1)} \leftarrow G^{(k)} \cap \left\{x \middle| \langle g^{(k)}, x - x^{(k)} \rangle\right\}$

Return $\quad \tilde{x} = \arg \min_{\forall_i f\left(x^{(k)}\right) < \epsilon} f\left(x^{(k)}\right)$

If $x^* \in G^{(0)}$, $\sup_{x \in G^{(0)}} |f_i(x)| \le B$ for $i = 0..m$, then after

$k = 2.2\, n \log B/\epsilon$ iterations:

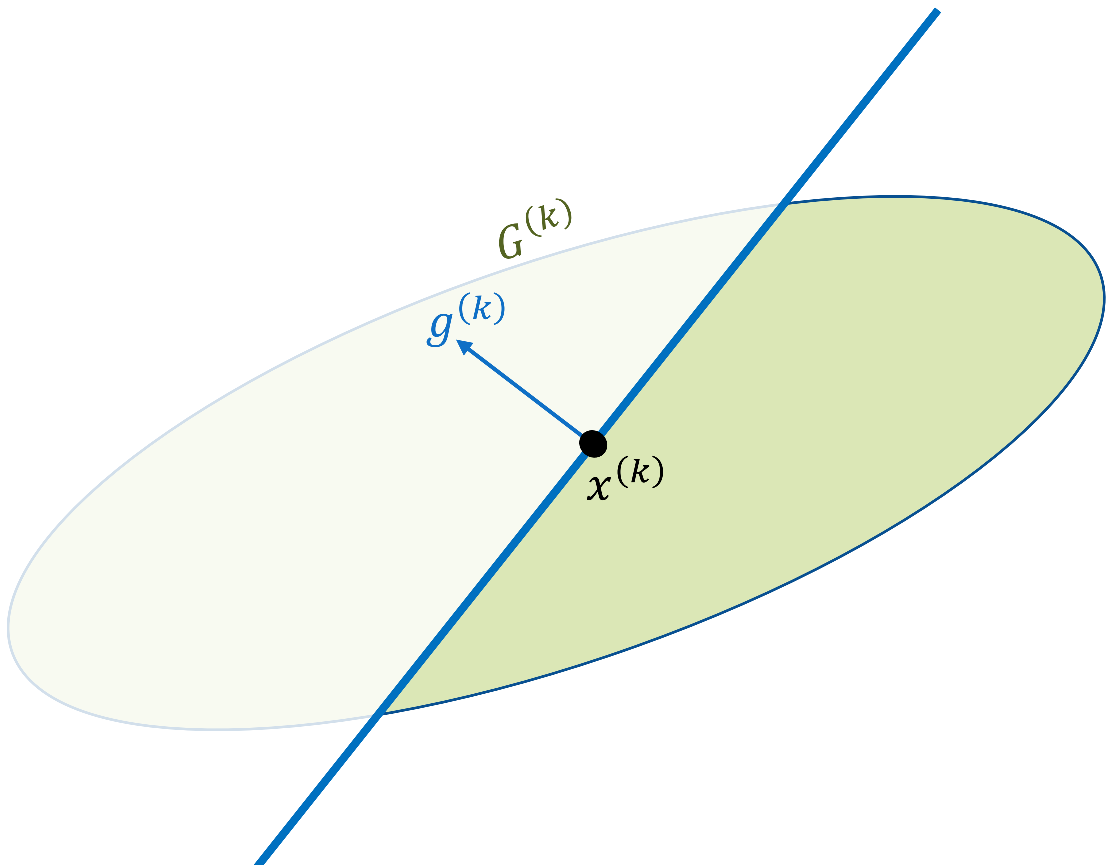
$$f_0(\tilde{x}) \le f_0(x^*) + \epsilon \qquad f_i(\tilde{x}) < \epsilon$$

Runtime of each iteration?

From Center-of-Mass to Ellipsoid

- Instead of maintaining a polytope $G^{(k)} \ni x^*$, maintain Ellipsoid
$$G^{(k)} = \left\{x = B^{(k)}u + x^{(k)} \middle| \|u\| \leq 1\right\} \ni x^*$$

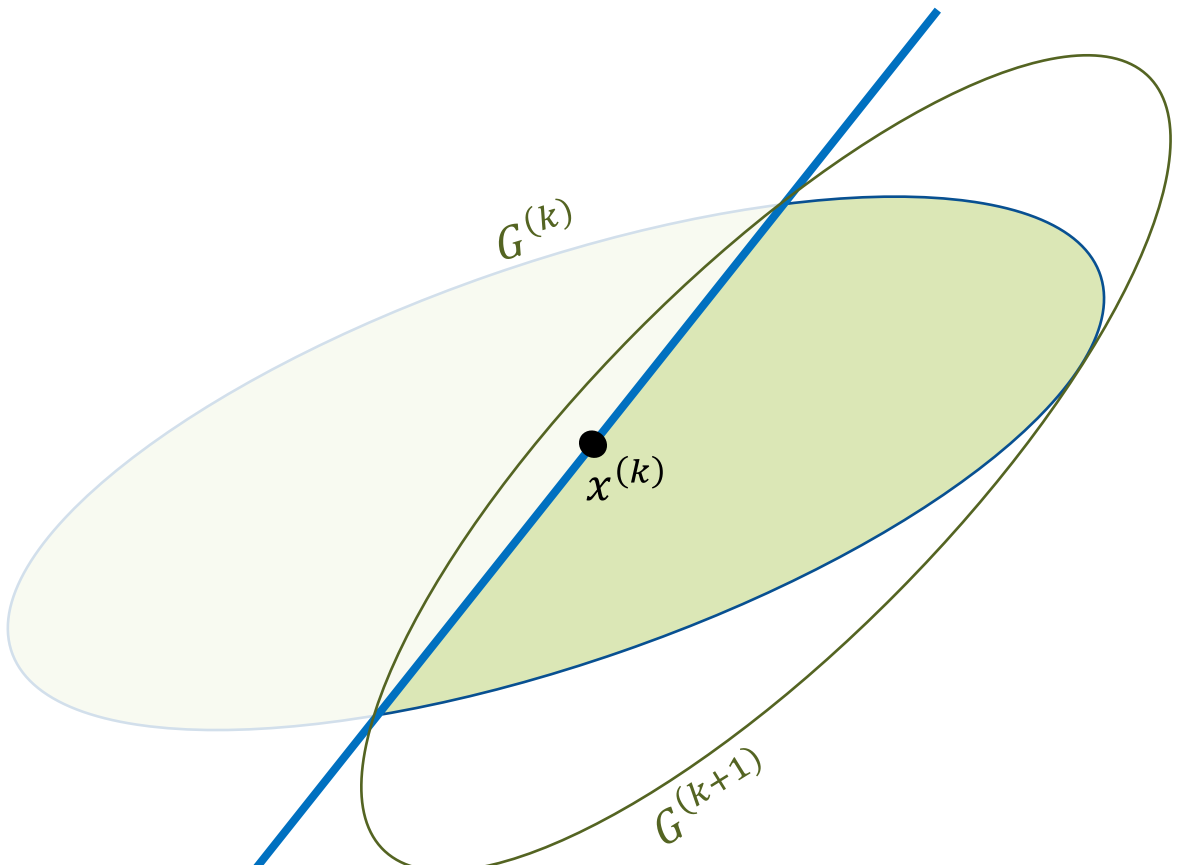
- At each iteration, need to find $G^{(k+1)} \supseteq G^{(k)} \cap \left\{\langle g^{(k)}, x - x^{(k)} \rangle < 0\right\}$



From Center-of-Mass to Ellipsoid

- Instead of maintaining a polytope $G^{(k)} \ni x^*$, maintain Ellipsoid
$$G^{(k)} = \left\{x = B^{(k)}u + x^{(k)} \middle| \|u\| \le 1\right\} \ni x^*$$

- At each iteration, need to find $G^{(k+1)} \supseteq G^{(k)} \cap \left\{\langle g^{(k)}, x - x^{(k)}\rangle < 0\right\}$



From Center-of-Mass to Ellipsoid

- Instead of maintaining a polytope $G^{(k)} \ni x^*$, maintain Ellipsoid
$$G^{(k)} = \left\{x = B^{(k)}u + x^{(k)} \middle| \|u\| \leq 1\right\} \ni x^*$$

- At each iteration, need to find $G^{(k+1)} \supseteq G^{(k)} \cap \left\{\langle g^{(k)}, x - x^{(k)}\rangle < 0\right\}$

$$x^{(k+1)} = x^{(k)} - \frac{1}{n+1} \frac{B^{(k)}B^{(k)}g^{(k)}}{\left\|B^{(k)}g^{(k)}\right\|}$$

$$B^{(k+1)} = \sqrt{\frac{n^2}{n^2-1}} B^{(k)} + \sqrt{\frac{n^2}{n^2-1}} \left(1 - \sqrt{\frac{n-1}{n+1}}\right) \frac{B^{(k)}B^{(k)}g^{(k)}{g^{(k)}}^\top B^{(k)}}{\left\|B^{(k)}g^{(k)}\right\|^2}$$

- Claim:
  - $G^{(k+1)} \supseteq G^{(k)} \cap \left\{\langle g^{(k)}, x - x^{(k)}\rangle < 0\right\}$
  - $Vol\left(G^{(k+1)}\right) \leq e^{-\frac{1}{2(n-1)}} Vol\left(G^{(k)}\right)$

➔ $Vol\left(G^{(k+1)}\right) \leq e^{-\frac{k}{2(n-1)}} Vol\left(G^{(0)}\right)$, need n times as many iterations

Ellipsoid Algorithm

Init	$G^{(0)} = \{x = B^{(0)}u + x^{(0)} \,\big|\, \|u\| \le 1\}$
Iterate	If $\exists_i \; f_i\left(x^{(k)}\right) \ge \epsilon$, then $g^{(k)} = \nabla f_i\left(x^{(k)}\right)$
	Else, $g^{(k)} = \nabla f_0\left(x^{(k)}\right)$
	$x^{(k+1)} = x^{(k)} - \frac{1}{n+1} \frac{B^{(k)} B^{(k)} g^{(k)}}{\left\|B^{(k)} g^{(k)}\right\|}$
	$B^{(k+1)} = \sqrt{\frac{n^2}{n^2-1}} B^{(k)} + \sqrt{\frac{n^2}{n^2-1}} \left(1 - \sqrt{\frac{n-1}{n+1}}\right) \frac{B^{(k)} B^{(k)} g^{(k)} {g^{(k)}}^{\mathsf{T}} B^{(k)}}{\left\|B^{(k)} g^{(k)}\right\|^2}$
Return	$\tilde{x} = \arg \min_{\forall_i f\left(x^{(k)}\right) < \epsilon} f\left(x^{(k)}\right)$

If $x^* \in G^{(0)}$, $\sup_{x \in G^{(0)}} |f_i(x)| \le B$ for $i = 0..m$, then after

$$k = 2n^2 \log B/\epsilon \quad \text{iterations:}$$

$$f_0(\tilde{x}) \le f_0(x^*) + \epsilon \qquad f_i(\tilde{x}) < \epsilon$$

Runtime:

$$O\left(n^4 \log \frac{B}{\epsilon}\right) + O\left(n^2 \log \frac{B}{\epsilon}\right) \text{ access to each first order oracle}$$

Finding Violating Constraints

- To use cutting plane methods (inc. Ellipsoid), we needed at each iteration
  - Decide if x is feasible
  - Or, find violated constraint $f_i(x) > 0$ (or $\geq \epsilon$)

- Straight-forward implementation:
  - At each iteration, enumerate over constraints and check them

- Instead of enumerating over constraints, enough to have efficient method (e.g. oracle) for finding violating constraint $f_i(x) > 0$

$$x \mapsto \text{“feasible” or } i \text{ s.t. } f_i(x) > 0$$

  - OK to have lots of constraints, as long as we can provide such an oracle
  - Runtime doesn't depend on #constraints

$$O\left(n^4 \log \frac{B}{\epsilon}\right) + O\left(n^2 \log \frac{B}{\epsilon}\right) \text{ oracle accesses}$$

Example: Min Cost Arborescence

- Input: directed graph $G(V, E)$ with costs $c(u \to v) \in \mathbb{R}$ on edges, and a root vertex $r \in V$
- Goal: find a minimum cost subset of edges, s.t. there is a path from r to every other edge
- LP relaxation (which is tight, i.e. LP has integer opt):

$$\min_{x(u \to v)} \quad \sum_{u \to v} c(u \to v) x(u \to v)$$

$$\text{s.t.} \quad \sum_{u \in S, v \notin S} x(u \to v) \geq 1 \quad \forall r \in S \subset V$$

$$0 \leq x(u \to v) \leq 1$$

- Exponentially many constraints, but easy to check feasibility and find violating constraint:
  - For each $v \in V$, find min-cut between r and v

Example: SDP

- Instead of the constraint:

$$X \preccurlyeq 0$$

use the scalar linear constraints:

$$v^\top X v \leq 0 \quad \forall\, v \in \mathbb{R}^n$$

- Infinitely many constraints, but easy to find violating constraint:
  - Find largest eigenpair (v, λ)
  - If $\lambda \leq 0$, return "feasible"
  - If $\lambda > 0$, return $f_v(X) = v^\top X v$

Separation Oracles

$$\boxed{\begin{array}{ll} \min_{x \in \mathbb{R}^n} & f_0(x) \\ s.t. & x \in K \end{array}}$$

- For cutting-plane methods, enough to have:
  - 1st order oracle: $x \mapsto f_0(x), \nabla f_0(x)$
  - Separation oracle: $x \mapsto$ "feasible" or g s.t.
    $$K \subseteq \{x' | \langle g, x' - x \rangle < 0\}$$

- E.g., for $K = \{X \preccurlyeq 0\}$:
  - $X \mapsto$ negative eigenvector if one exists, or "feasible" if not

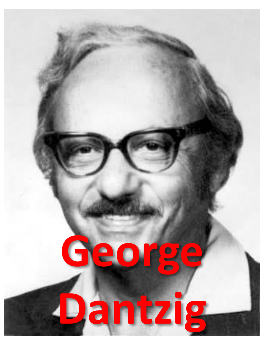
Ellipsoid Method

- Runtime:
  - $O(n^4 \log {}^1/_\epsilon)$
  - $O(n^2 \log {}^1/_\epsilon)$ accesses to 1st order and separation oracles
- Compare with IP methods:
  - $O\left(m^{1/2}(n^3 + m) \log {}^1/_\epsilon\right)$
  - $O\left(m^{1/2} \log {}^1/_\epsilon\right)$ accesses to 2nd order oracle for each $f_0, f_1, \dots, f_m$
- In practice: Ellipsoid really n^4, whereas for IP, closer to n^3 with Newton, faster with quasi-Newton

But:

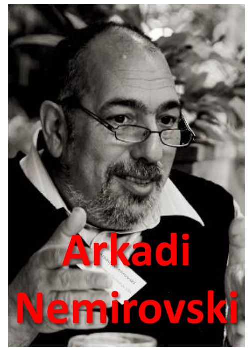
- Historical significance

Formulation of LP



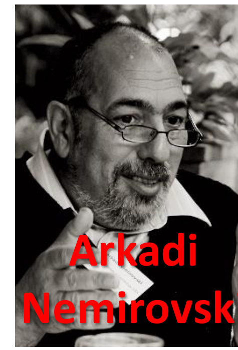
Simplex

A Levin
D Newman
Center of Mass



Ellipsoid Method

Ellipsoid is Poly time for LP



General IP

IP for LP

1939 1947 1965 1972 1979 1984 1989-1994

Ellipsoid Method

- Runtime:
  - $O(n^4 \log {}^1\!/\!_\epsilon)$
  - $O(n^2 \log {}^1\!/\!_\epsilon)$ accesses to 1st order and separation oracles
- Compare with IP methods:
  - $O\left(m^{1/2}(n^3 + m) \log {}^1\!/\!_\epsilon\right)$
  - $O\left(m^{1/2} \log {}^1\!/\!_\epsilon\right)$ accesses to 2nd order oracle for each $f_0, f_1, \dots, f_m$
- In practice: Ellipsoid really n^4, whereas for IP, closer to n^3 with Newton, faster with quasi-Newton

But:

- Historical significance:
  - First poly-time method for LP
  - First, and for a long time only, poly-time method for SDP
- Useful for combinatorial algorithm (at least in theory), since it can handle infinitely many constraints

Cutting Plane Methods

- Reduce problem with infinitely many constraints, or only separation oracle, to LP

Init	small set of linear const $L^{(0)}$
	(e.g. only box constraints on each variable)
Iterate	Solve LP: $x^{(k)} \to \min f_0(x) \;\; s.t. L^{(k)}$
	Query separation oracle with $x^{(k)}$
	If not feasible, and oracle returns $g^{(k)}$,
	$L^{(k+1)} \leftarrow L^{(k)} \cup \{ \langle g^{(k)}, x - x^{(k)} \rangle \leq 0 \}$

Can we do better?

- **Ellipsoid**: $O(n^2 \log {}^1/_\epsilon)$ iterations, $O(n^4 \log {}^1/_\epsilon)$ runtime
- **Center of Mass**: $O(n \log {}^1/_\epsilon)$ iterations
  - Exact computation (likely) requires exponential time (#P-complete)
  - Using **random-walk sampling** to aprox center-of-mass [Bertsimas Vempala 2004]: $\tilde{O}(n^6)$ per iteration ➔ $\tilde{O}(n^7 \log {}^1/_\epsilon)$ total, but only $O(n \log {}^1/_\epsilon)$ oracle accesses
- Faster cutting plane method?
  - need to keep track of $O(n \log {}^1/_\epsilon)$ hyperplanes, each of dim n, i.e. $\Omega(n^2)$ nums
  - also with ellipsoids: $n \times n$ matrix representing ellipsoid has $\Omega(n^2)$ numbers
  - seems like at least $\Omega(n^2)$ per iteration ➔ $\Omega(n^3 \log^2 {}^1/_\epsilon)$ overall
- **Vaidya's cutting plane algorithm** [Vaidya 1989][Lee Sidford Wang 2015]
  - Keep track of polytope, adding *and removing* halfspaces
  - Use minimum of "volumetric barrier" instead of center of mass
  - $\tilde{O}(n \log {}^1/_\epsilon)$ oracle access, $\tilde{O}(n^3 \log {}^1/_\epsilon)$ total runtime
- Can we optimize with $< \omega(n)$ iterations (1st order/separation oracle accesses)?
  - E.g. for IP methods, #iterations Newton independent of n