Optimization via Separation

Goal: Want to (approximately) solve

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

in blackbox model

$f : \mathbb{R}^n \to \overline{\mathbb{R}} \; (:= \mathbb{R} \cup \{\infty\})$ convex

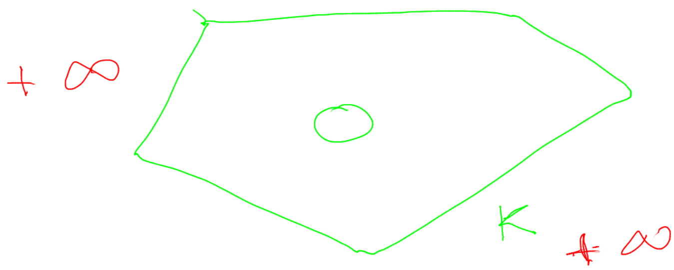
"push" constraints into f

Important concepts:

Domain of f: $\mathrm{Dom}(f) := \{x \in \mathbb{R}^n : f(x) < \infty\}$

convex set (by convexity of f)

Ex:

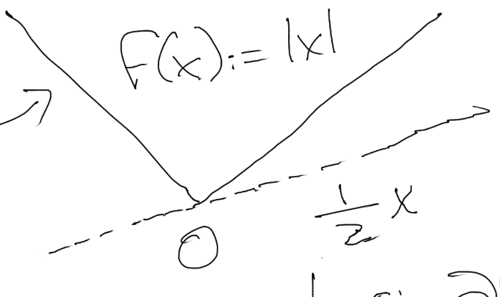


$f(x) = \infty \cdot \mathbb{1}[x \notin K]$

subgradient at x:

$g \in \partial f(x)$ if $\forall y \;\; f(y) \geq f(x) + \langle g, y - x \rangle$

(only for $x \in \mathrm{Dom}(f)$)



if f is differentiable at x then $\partial f(x) = \{\nabla f(x)\}$

$\frac{1}{2} \in \partial f(0)$

Application: Linear Programming

(LP) $Ax \le b$ is feasible

$$(A, b) = \begin{pmatrix} a_1 , & b_1 \\ \vdots & \vdots \\ a_m , & b_m \end{pmatrix} \in \mathbb{Q}^{m \times n+1}$$

$\iff$ (LP')

$$\min_{x \in \mathbb{R}^n} \max_{i \in [m]} \{0, a_i \cdot x - b_i\} \le 0$$

$\uparrow$ f

Exercise: For f describe $\partial f(x)$ for any x.

Khachiyan '79 :

Let $E = \langle A, b \rangle$ bit-encoding length

Then (LP) $Ax \leq b$ is feasible

$\iff$ (LP') has value $\leq \boxed{\frac{1}{2^{O(E)}}}$

Pf sketch: Dual program is

$$\max \; -\sum_{i=1}^{m} \lambda_i b_i$$

$$A^T \lambda = 0$$

$$\lambda \geq 0$$

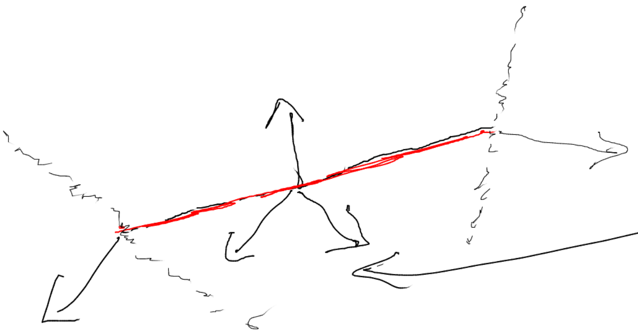
$$\sum \lambda_i \leq 1$$

(i) maximized at vertex λ^* (satisfies m lin. ind. tight constraints)

(ii) Use Cramer's rule to show $-\sum \lambda_i^* b_i = p/q$ where $p, q \in \mathbb{Z}$, $1 \leq q \leq 2^{O(E)}$.

Punchline: Can reduce decisional LP feasibility to approximate convex minimization.

Remark: Lower dimensional LPs are "annoying" :)



these 3 halfspaces reduce dimension from 3 to 1.

Search to Decision?

Yes, iteratively force constraints to be tight while maintaining feasibility until you find a vertex.

(Admittedly, very lame)

LP Optimization?

$$\min \ c \cdot x$$
$$Ax \leq b$$

Exercise: reduce to search version of LP feasibility.

(Hint: combine primal & dual program and set them equal to each other)

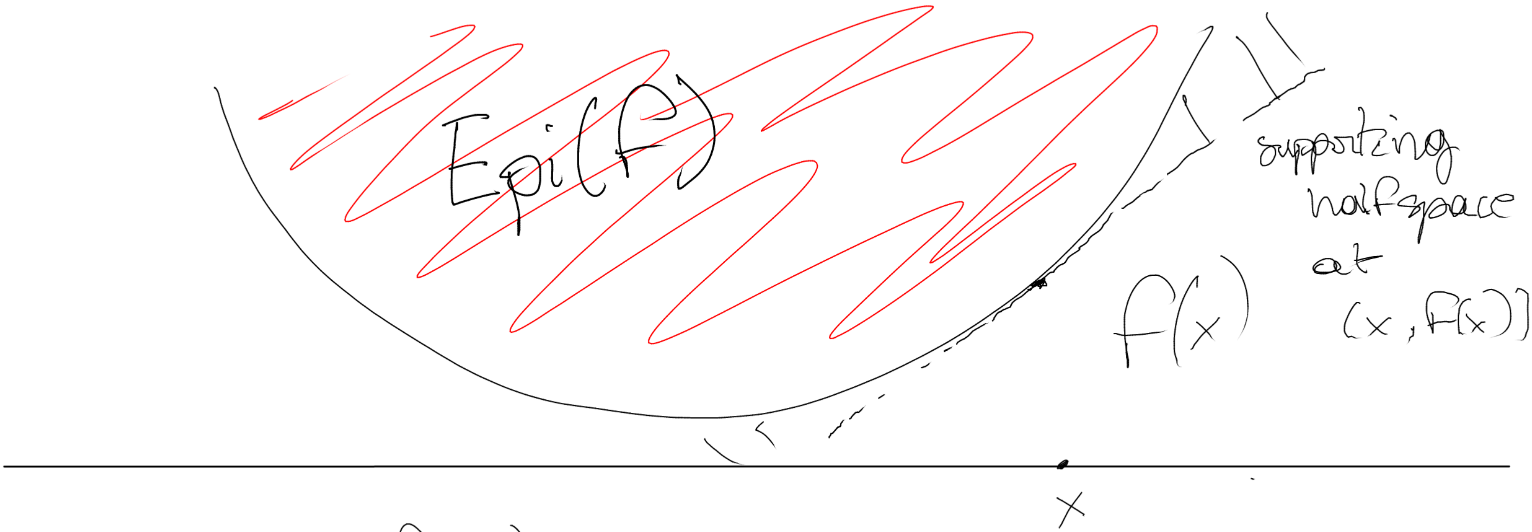
Minimizing convex function

$\Updownarrow$ equivalent to minimizing linear functions on convex sets:

Epigraph of f:

$Epi(f) = \{(x,t) : f(x) \le t,\ t \in \mathbb{R}\}$

convex



$$\min_{x \in \mathbb{R}^n} f(x) \iff \min_{(x,t) \in Epi(f)} 1 \cdot t + 0 \cdot x$$

Remarks:

- $\Pi_x(Epi(f)) = Dom(f)$
- $g \in \partial f(x_0) \iff t \ge f(x_0) + \langle g, x - x_0 \rangle$ for $x \in Dom(f)$

  supporting halfspace for $epi(f)$ at $(x_0, f(x_0))$

(non-empty as long as Epi(f) is closed

)

Moral: To minimize f should only need separation oracle for $\mathrm{Epi}(f)$.

Will use slightly more convenient assumption that we have sep. oracle for <u>level sets</u> of f.

$\boxed{\text{Assumptions on } f:}$

(for simplicity)

$\boxed{1.}$ Have <u>strong</u> separation oracle for level sets of f.

$$\mathrm{lev}_\alpha^{(\leq)}(f) := \{x \in \mathbb{R}^n : f(x) \overset{(\leq)}{<} \alpha\}, \alpha \in \overline{\mathbb{R}}$$

If $x \notin \mathrm{lev}_\alpha(f)$ get g s.t.

$$\sup_{z \in \mathrm{lev}_\alpha(f)} \langle g, z \rangle \leq \langle g, x \rangle$$

Remarks:

- $x \notin \mathrm{lev}_{f(x)}(f)$
- $\mathrm{lev}_\infty(f) = \mathrm{Dom}(f)$

When can we implement this?

Easy if we have

(a) Separation oracle for Dom(f).

(b) Have subgradient oracle for f.

Given f, α, x with $f(x) \geq \alpha$

If $x \notin \mathrm{Dom}(f) \supseteq \mathrm{lev}_\alpha(f)$ apply (a)

Else return $g \in \partial f(x)$ using (b)

correctness

$$\langle g, y\rangle \geq \langle g, x\rangle$$

$$\Longrightarrow f(y) \geq f(x) + \langle g, y-x\rangle$$

$$\geq f(x) \geq \alpha$$

Therefore $y \in \mathrm{lev}_\alpha(f)$

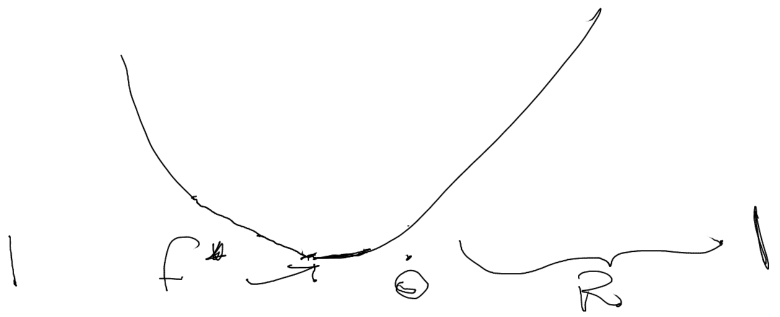
$$\Longrightarrow \langle g, y\rangle < \langle g, x\rangle \quad \square$$

Assumption 2

$f^* := \inf_{x \in \mathbb{R}^n} f(x)$ is attained

and know radius $R > 0$ s.t.

$\mathrm{lev}^{\leq}_{f^*}(f) \cap RB_2^n = \emptyset$ ← ball of radius R



As we will only work within RB_2^n, we can modify f to be ∞ outside RB_2^n.

In particular, we henceforth assume

$$\boxed{\emptyset \neq \mathrm{lev}^{\leq}_{f^*}(f) \subseteq \mathrm{Dom}(f) \subseteq RB_2^n}$$

Exercise: For $f(x) = \max_{i \in [m]} \max\{0, a_i \cdot x - b_i\}$ as in (LP'), show that $R = 2^{O(E)}$ suffices.

Theorem 1: Let f satisfy $\boxed{1} + \boxed{2}$.

For $\varepsilon > 0$, define $V_\varepsilon := \dfrac{\mathrm{vol}_n(\mathrm{lev}_{f^*+\varepsilon}(f))}{\mathrm{vol}_n(RB_2^n)}$.

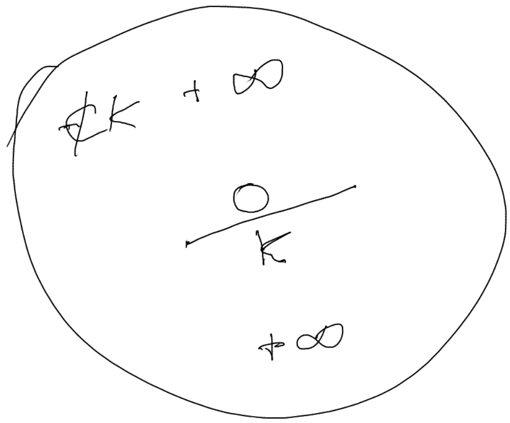
Then the Center of Gravity (CoG) makes

$\boxed{O(\log 1/V_\varepsilon)}$ level set separation queries

and computes x satisfying $f(x) \leq f^* + \varepsilon$.

Remark: V_ε could be zero!

E.g.



$f(x) = \infty \cdot \mathbb{1}[x \notin K]$

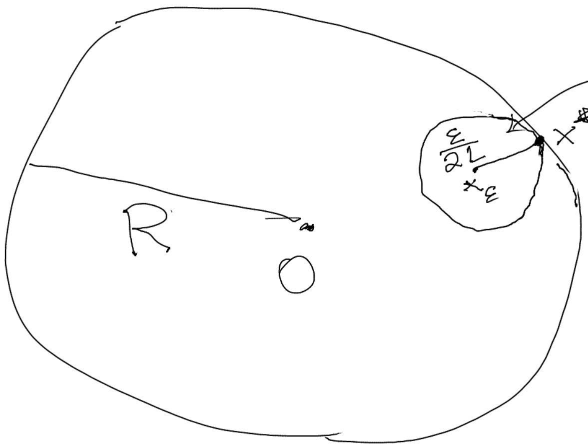
K convex &
lower dimensional

When is V_ε "reasonably" lower bounded?

Lemma: If f is L-Lipschitz

then $V_\varepsilon \geq \left(\frac{\varepsilon}{2LR}\right)^n \Rightarrow \log\frac{1}{V_\varepsilon} \leq \boxed{n \log \frac{2LR}{\varepsilon}}$

Pf:



$$f(x) \leq f(x^*) + L\|x - x^*\|$$
$$\leq f(x^*) + \varepsilon$$

$$V_\varepsilon \geq \frac{\mathrm{vol}_n\left(\frac{\varepsilon}{2L} B_2^n\right)}{\mathrm{vol}_n\left(R B_2^n\right)} = \left(\frac{\varepsilon}{2LR}\right)^n$$

Corollary: For $f : \mathbb{R}^n \to \mathbb{R}$ L-Lipschitz satisfying $\boxed{1} + \boxed{2}$, CoG makes $O\left(n \log \frac{LR}{\varepsilon}\right)$ separation queries to compute an ε-optimal point x $\left(f(x) < f^* + \varepsilon\right)$.

Ex: Show that (LP') objective is $L = \max_{i \in [m]} \|a_i\|_2 = 2^{O(L)}$ - Lipschitz.

Proof Theorem 1: Geometric "Binary Search"

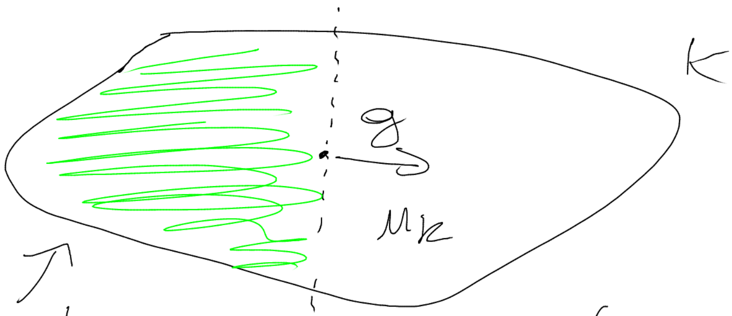
Will use Center of gravity method

For convex body $K \subseteq \mathbb{R}^n$
(compact convex with non-empty interior)

define $\mu_K := \mathbb{E}[X]$, $X \sim \text{uniform}(K)$

Grünbaum's Theorem: $\forall g \in \mathbb{R}^n / \{0\}$
60's

$$\text{vol}_n\left(K \cap \{x : \langle g, x \rangle \leq \langle g, \mu_K \rangle\}\right) \leq \left(1 - \frac{1}{e}\right) \text{vol}_n(K)$$



at most $1 - \frac{1}{e}$
fraction of volume.

(reduces to 1D inequality
for logconcave distributions)

CoG Method(f, R, # iter N)

$K_0 \leftarrow R B_2^n$, $x_0 \leftarrow 0$

For $t = 1$ to N

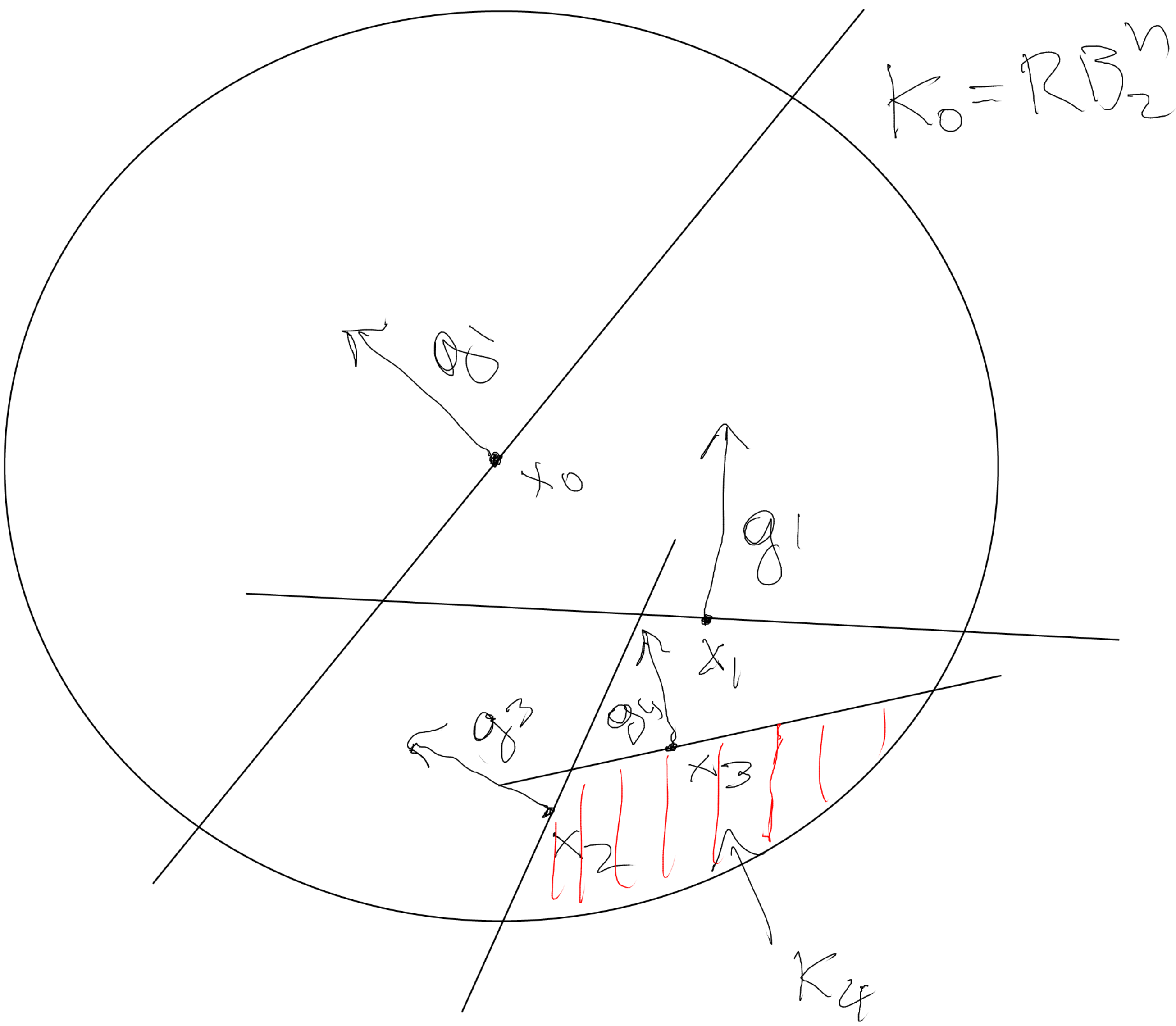
$g_t \leftarrow$ separator for x_{t-1} wrt $\mathrm{lev}_{f(x_{t-1})}(f)$

If $g_t = 0$ return x_{t-1}

$K_t \leftarrow K_{t-1} \cap \{x : \langle g_t, x \rangle \leq \langle g_t, x_{t-1} \rangle\}$

$x_t \leftarrow \mu_{K_t}$ center of gravity of K_t

Return best solution from $x_0, x_1, \ldots, x_N$



Claim: For $N > \lceil \log 1/v_\varepsilon / \log \frac{e}{e-1} \rceil$,

$$v_N := \min_{i \in [N]} f(x_i) < f^* + \varepsilon.$$

Pf: We first show by induction that $\mathrm{lev}_{v_N}(f) \subseteq K_N$. For $K_0 = RB_2^n$ by assumption $\mathrm{lev}_{v_N}(f) \subseteq \mathrm{Dom}(f) \subseteq K_0$.

For K_t, $1 \le t \le N$, we have

$$K_t = K_{t-1} \cap \{x : \langle g_t, x \rangle \le x_{t-1}\}$$

$$\supseteq \underbrace{\mathrm{lev}_{v_N}(f)}_{\text{induction hypothesis}} \cap \underbrace{\mathrm{lev}_{f(x_{t-1})}(f)}_{\text{separation guarantee}}$$

$$= \mathrm{lev}_{v_N}(f).$$

Now if $v_N \ge f^* + \varepsilon$, we have that

$$\mathrm{lev}_{f^*+\varepsilon}(f) \subseteq \mathrm{lev}_{v_N}(f) \subseteq K_N$$

$$\Longrightarrow \mathrm{vol}_n(\mathrm{lev}_{f^*+\varepsilon}(f)) \le \mathrm{vol}_n(K_N).$$

But by Grünbaum's theorem

$$\begin{aligned}\mathrm{vol}_n(K_N) &\le \left(1-\frac{1}{e}\right)^N \mathrm{vol}_n(RB_2^n)\\ &< V_\varepsilon \,\mathrm{vol}_n(RB_2^n)\\ &= \mathrm{vol}_n(\mathrm{lev}_{f^*+\varepsilon}(F)),\end{aligned}$$

a clear contradiction.

Therefore, we must have $v_N < f^* + \varepsilon$.

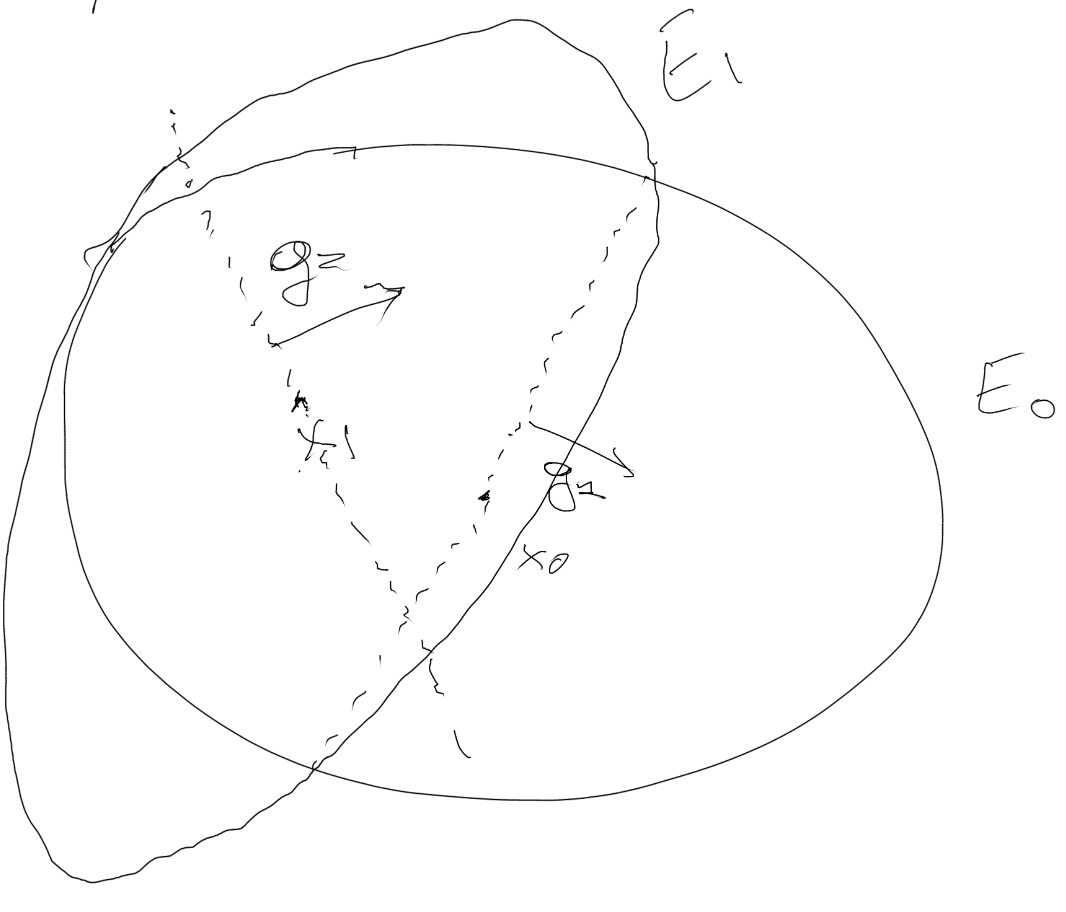
$\square$

Ellipsoid Method:

Center of gravity method has great convergence, but is not easy to implement, i.e. how to compute $\mathbb{E}_{X\sim K_t}[X]$?

(Can in fact be approximated using random walk methods, e.g. Dyer, Frieze, Kannan '91)

Will discuss deterministic method based on approximating optimality set by sequence of ellipsoids.



Will compute "shrinking" sequence ellipsoids
(in volume)

$$E_0, E_1, \ldots, E_N \supseteq \mathrm{lev}_{v_N}(f)$$

$x_0 \quad x_1 \quad x_N$

$\min f(x_i)$

centers

in analoguous way to CoG..

Outline of iteration t:

$g_t \leftarrow$ separator for x_{t-1} wrt $\mathrm{lev}_{f(x_{t-1})}(f)$

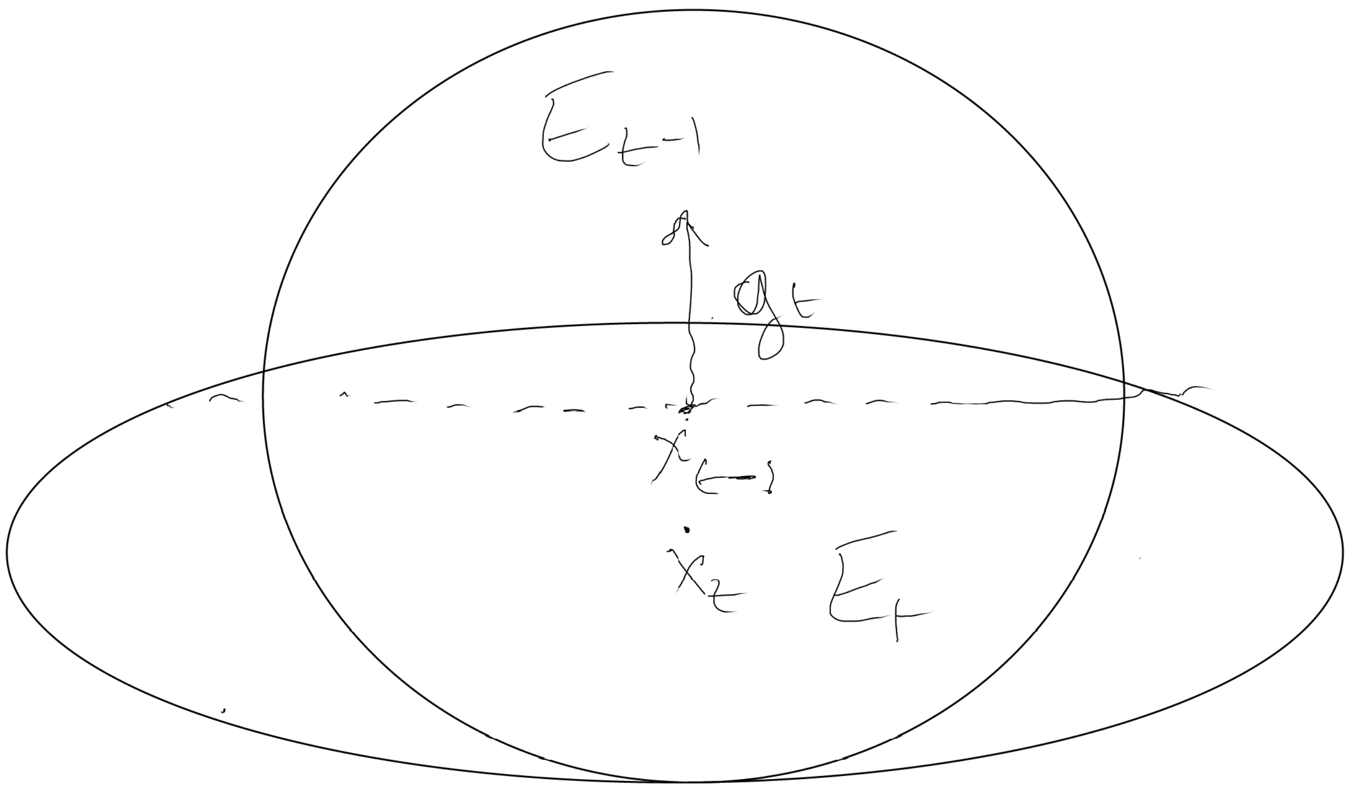
$E_t \leftarrow$ minimum volume ellipsoid
containing

$$E_{t-1} \cap \{x : \langle x, g_t \rangle \le \langle x_{t-1}, g_t \rangle\}$$

$x_t \leftarrow$ center of E_t

Central Problems:

(1) How to compute E_t?



(2) How small is

$$\frac{\mathrm{vol}(E_t)}{\mathrm{vol}(E_{t-1})}\ ?$$

Answers:

(1) Can give explicit formula:

(2) $$\frac{\mathrm{vol}(E_t)}{\mathrm{vol}(E_{t-1})} \leq \boxed{e^{-\frac{1}{2(n+1)}}}$$

Factor $\underline{n}$ slower than CoG.

Theorem 1 with ellipsoid requires

$$O\left(n \log \frac{1}{v_\varepsilon}\right) = O\left(n^2 \log \frac{R}{\varepsilon L}\right)$$

iterations.

Ellipsoids Formally:

Will consider only full dimensional and compact ellipsoids in $\mathbb{R}^n$.

Many equivalent defs:

1) Invertible affine transform

$$\underbrace{T}_{\text{invertible}} B_2^n + \underbrace{c}_{\text{center}} \quad \text{of ball}$$

2) $(x-c)^T Q (x-c) \le 1$

$\underbrace{Q \succ 0}_{\text{positive definite}}$ (set $Q = T^{-T} T^{-1}$,

note $\det(Q)^{-1} = \det(T)^2$)

2') $q(x) \geq 0$

q strictly concave quadratic polynomial satisfying

$$q^* := \max_{x \in \mathbb{R}^n} q(x) > 0.$$

center is implicit here:

$$c_q = \operatorname{argmax}_{x \in \mathbb{R}^n} q(x)$$

Hessian is constant

$$q(x) = q^* + \frac{1}{2}(x - c_q)^T \nabla^2 q(0)(x - c_q)$$

$-\nabla^2 q(0) \succ 0$ by strict concavity.

Ellipsoid: $(x - c_q)^T \dfrac{-\nabla^2 q(0)}{2q^*}(x - c_q) \leq 1$

Exercise: Show

- $c_q = -\nabla^2 q(0)^{-1} \nabla q(0)$
- $q^* = q(0) - \frac{1}{2} \nabla q(0)^T \nabla^2 q(0)^{-1} \nabla q(0)$

can replace 0 with any $x \in \mathbb{R}^n$

Volume Formula:

1) $\mathrm{vol}_n(T B_2^n + c) = |\det(T)| \, \mathrm{vol}_n(B_2^n)$

"reverse engineer"

2) $\mathrm{vol}_n(\{x : q(x) \geq 0\})$

$$= \frac{(q^*)^{n/2}}{\det(-\nabla^2 q(0)/2)^{1/2}} \mathrm{vol}_n(B_2^n)$$

Optimizing over ellipsoids:

Lemma: For $R \succ 0$, $c \in \mathbb{R}^n$, $g \in \mathbb{R}^n$

$$\min\{\langle g, x\rangle : (x-c)^T R (x-c) \leq 1\} = \langle g, c\rangle - \sqrt{g^T R^{-1} g}$$

and minimizer is $c - \dfrac{R^{-1} g}{\sqrt{g^T R^{-1} g}}$.

Pf: Assume $(x-c)^T R (x-c) \leq 1$

then $\langle g, x\rangle = \langle g, c\rangle + \langle g, x-c\rangle$

$$= \langle g, c\rangle + \langle R^{-1/2} g, R^{1/2}(x-c)\rangle \quad \text{(Cauchy-Schwarz)}$$

$$\geq \langle g, c\rangle - \|R^{-1/2} g\|_2 \|R^{1/2}(x-c)\|_2$$

$$= \langle g, c\rangle - \sqrt{g^T R^{-1} g}\, \underbrace{\sqrt{(x-c)^T R (x-c)}}_{\leq 1}$$

Take $x = c - \dfrac{R^{-1} g}{\sqrt{g^T R^{-1} g}}$.

Then $(x-c)^T R (x-c) = \dfrac{g^T R^{-1} R R^{-1} g}{g^T R^{-1} g} = 1$

and $\langle g, x\rangle = \langle g, c\rangle - \dfrac{g^T R^{-1} g}{\sqrt{g^T R^{-1} g}}$

$$= \langle g, c\rangle - \sqrt{g^T R^{-1} g}.$$

Why reprensentation (2')?

Allows us to reprensent min volume ellipsoid as <u>convex program</u>:

Löwner ellipsoid for convex (min volume containing) body K:

$$\max \ \ln \det\left(-\nabla^2 q(0)/2\right)$$

normalization $q^* \leq 1$
$\hookrightarrow$ $q(x) \leq 1 \quad \forall x \in \mathbb{R}^n$

contains K
$\hookrightarrow$ $q(x) \geq 0 \quad \forall x \in K$

q quadratic strictly concave polynomial

Interested in special case where K is "half" ellipsoid

$$K = \{x : p(x) \ge 0,\ \langle g, x\rangle \le \langle g, c_p\rangle\}$$

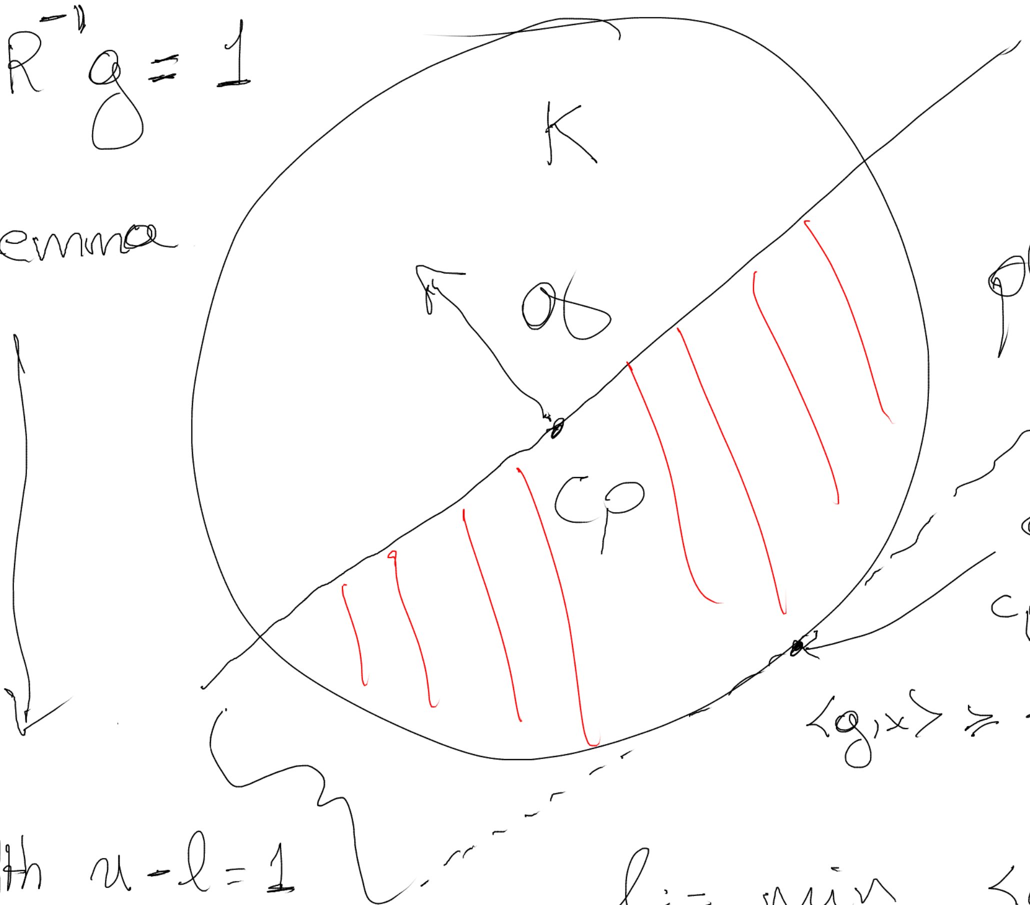
$$p(x) := \underset{p^*}{1} - (x - \underset{\text{center}}{c_p})^T \underset{-\frac{1}{2}\nabla^2 p(0) \succ 0}{R} (x - c_p)$$

Normalization for g

$\langle g, x\rangle \le \langle g, c_p\rangle =: u$

$g^T R^{-1} g = 1$

by Lemma



$p(x) \ge 0$

g minimizer: $c_p - R^{-1} g$

$\langle g, x\rangle \ge \langle g, c_p\rangle - 1$

width $u - l = 1$

$$l := \min_{p(x) \ge 0} \langle g, x\rangle = u - 1$$

Main Idea: [Burrell - Todd '85
"The Ellipsoid Method Generates Dual Variables"]

$$p(x) \geq 0, \quad \underbrace{\ell \leq \langle g, x \rangle \leq u}$$

$\uparrow$ redundant ineq.

$$\Longleftrightarrow \quad p(x) \geq 0 \quad |$$

$$\underbrace{\Delta p(x) := (u - \langle g, x \rangle)(\langle g, x \rangle - \ell)}_{} \geq 0$$

Choose new ellipsoid $q_\lambda(x) \geq 0$

with $q_\lambda = (1-\lambda) p + \lambda \Delta p, \quad \lambda \in [0,1]$

Properties: 1. $q_\lambda(x) \geq 0 \quad \forall x \in K$.

2. $c_p - R_p^{-1} g$ minimizes $\langle g, \cdot \rangle$ on $p(x) \geq 0$ & $\Delta p(x) \geq 0$

$\Longrightarrow$ minimizer on $q_\lambda(x) \geq 0$.

(Min volume ellipsoid will indeed have form $q_\lambda(x) \geq 0$)

Will compute (squared) volume minimizer:

$$\min_{\lambda\in[0,1]} \frac{\mathrm{vol}_n(\{x : q_\lambda(x) \geq 0\})^2}{\mathrm{vol}_n(\{x : p(x)\geq 0\})^2} = \qquad \boxed{\text{MIN-VOL}}$$

$$\min_{\lambda\in[0,1]} \frac{\det(R)\,(q_\lambda^*)^n}{\det(-\frac{1}{2}\nabla q_\lambda(0))} \quad \swarrow \text{ squared volume ratio.}$$

<u>Properties of q_λ:</u>

a) Form:
$$q_\lambda(x) = (1-\lambda)(1-(x-c_p)^T R(x-c_p)) + \lambda\,(\underbrace{u-\langle g,x\rangle}_{(-\langle g,\, x-c_p\rangle)})(\underbrace{\langle g,x\rangle - \ell}_{\cdot(\langle g,\,x-c_p\rangle+1)})$$

$$= (1-\lambda) - \lambda\langle g, x-c_p\rangle - (x-c_p)^T((1-\lambda)R+\lambda gg^T)(x-c_p)$$

b) Hessian: $-\frac{1}{2}\nabla^2 q_\lambda(0) = (1-\lambda)R + \lambda gg^T$

c) Inverse: $((1-\lambda)R+\lambda gg^T)^{-1}$

by normalization $g^T R^{-1} g = 1$ $\leadsto$ $= \frac{R^{-1}}{1-\lambda} - \frac{\lambda}{1-\lambda} R^{-1} gg^T R^{-1}$

d) Maximum values:

recall Hessian is constant

$$q_\lambda^* = \underbrace{q_\lambda(c_p)}_{} - \frac{1}{2}\underbrace{\nabla q_\lambda(c_p)^T}_{} \underbrace{\nabla^2 q_\lambda(0)^{-1}}_{} \nabla q(c_p) \quad \text{(by exercise)}$$

$$= 1-\lambda - \frac{1}{2}(-\lambda g)^T\left(-2\left((1-\lambda)R + \lambda g g^T\right)\right)^{-1}(-\lambda g) \quad \text{(a) + (b)}$$

$$= (1-\lambda) + \frac{1}{4}\lambda^2 g^T\left((1-\lambda)R + \lambda g g^T\right)^{-1} g$$

$$= (1-\lambda) + \frac{1}{2}\lambda^2 g^T\left(\frac{R^{-1}}{1-\lambda} - \frac{\lambda}{1-\lambda}R^{-1}gg^TR^{-1}\right) g \quad \text{(c)}$$

$$= (1-\lambda) + \frac{1}{2}\lambda^2\left(\frac{1}{1-\lambda} - \frac{\lambda}{1-\lambda}\right) \quad \text{normalization}$$

$$= 1 - \lambda + \frac{1}{2}\lambda^2 = \boxed{\left(1-\frac{\lambda}{2}\right)^2}$$

e) Center of ellipsoid $q_\lambda(x) \geq 0$:

Center is $\operatorname{argmax}_{x \in \mathbb{R}^n} q_\lambda(x)$.

By exercise (as above)

$$c_{q_\lambda} = c_p - \nabla^2 q_\lambda(0)^{-1} \nabla q_\lambda(c_p)$$

$$= c_p - \frac{1}{2}\left((1-\lambda)R + \lambda g g^T\right)^{-1} \lambda g$$

$$= c_p - \frac{1}{2}\lambda R^{-1} g \quad \text{(same as in part (d))}$$

Linearly interpolates between center of $p(x) \geq 0$ and middle of band $l \leq \langle g, x\rangle \leq u$

F) Squared volume ratio (MIN-VOL) for q_λ:

$$\frac{\det(R)\, q_\lambda^{*\,n}}{\det\left(-\frac{1}{2} D^2 q_\lambda(0)\right)} \xrightarrow{(10)+(d)} = \frac{\det(R)\,(1-\lambda/2)^{2n}}{\det\left((1-\lambda)R + \lambda g g^T\right)}$$

$[\det(A\cdot B) = \det(A)\cdot\det(B)]$

$$= \frac{(1-\lambda/2)^{2n}}{(1-\lambda)^n \det\left(I + \frac{\lambda}{1-\lambda} R^{-1} g g^T\right)}$$

$\left[\det(I + a b^T) = 1 + \langle a, b\rangle\right]$

$$= \frac{(1-\lambda/2)^{2n}}{(1-\lambda)^n \left(1 + \frac{\lambda}{1-\lambda} g^T R^{-1} g\right)}$$

$$= \frac{(1-\lambda/2)^{2n}}{(1-\lambda)^n \left(1 + \frac{\lambda}{1-\lambda}\right)} = \frac{(1-\lambda/2)^{2n}}{(1-\lambda)^{n-1}}$$

Comparison between $q_\lambda(x) \geq 0$ & $p(x) \geq 0$:

$$\left[\frac{(1-\lambda/2)^2}{(1-\lambda)}\right]^{n-1} (1-\lambda/2)^2$$

$n-1$ axes get longer

1 axis gets shorter (get "squeezed" in dir. of g)

Note for small $\lambda > 0$

$$\frac{(1-\lambda/2)^{2n}}{(1-\lambda)^{n-1}} \approx \frac{(1-\lambda)^n}{(1-\lambda)^{n-1}} = 1-\lambda$$

so expect some decrease.

Lemma: Squared volume decrease satisfies

$$\min_{\lambda \in [0,1]} \frac{(1-\frac{\lambda}{2})^{2n}}{(1-\lambda)^{n-1}} = \frac{(1-\frac{1}{n+1})^{2n}}{(1-\frac{2}{n+1})^{n-1}} \leq \frac{n}{n+1} \leq e^{-\frac{1}{n+1}}.$$

minimizer $\lambda = \frac{2}{n+1}$.

Pf: Note for $n=1$, denominator is constant, so minimizing $\lambda = 1 = \frac{2}{n+1}$, i.e. we replace interval K by interval $\Delta p(x) \geq 0$ of half size, exactly as in binary search. For $n \geq 2$,

$$\min_{\lambda \in [0,1]} \frac{(1-\lambda/2)^{2n}}{(1-\lambda)^{n-1}} \iff \min_{\lambda \in [0,1]} 2n \ln(1-\lambda/2) - (n-1)\ln(1-\lambda)$$

Easy to solve (*) by setting derivative to 0.

$$\frac{-n}{1-\lambda/2} + \frac{n-1}{1-\lambda} = 0 \implies \boxed{\lambda = \frac{2}{n+1}}$$

MIN-VOL: $\frac{(1-\frac{1}{n+1})^{2n}}{(1-\frac{2}{n+1})^{n-1}} = \left(\frac{n}{n+1}\right)^{n+1}\left(\frac{n}{n-1}\right)^{n-1}$.

$$= \left(\frac{n}{n+1}\right) \cdot \left(\frac{n}{n+1}\right)\left(\frac{n^2}{n^2-1}\right)^{n-1} = \left(\frac{n}{n+1}\right)\left(1-\frac{1}{n+1}\right)\left(1+\frac{1}{n^2-1}\right)^{n-1}$$

$$\leq \left(\frac{n}{n+1}\right) \cdot e^{-\frac{1}{n+1}} \cdot e^{\frac{n-1}{n^2-1}} = \frac{n}{n+1} \quad \square$$

$(1+x \leq e^x)$