Ellipsoid Method

- ellipsoid method
- convergence proof
- inequality constraints
- feasibility problems

Ellipsoid method

- developed by Shor, Nemirovsky, Yudin in 1970s
- used in 1979 by Khachian to show polynomial solvability of LPs
- each step requires cutting-plane or subgradient evaluation
- modest storage ($O(n^2)$)
- modest computation per step ($O(n^2)$), via analytical formula
- efficient in theory; slow but steady in practice

Motivation

in cutting-plane methods

- serious computation is needed to find next query point (typically $O(n^2m)$, with not small constant)

- localization polyhedron grows in complexity as algorithm progresses (we can, however, prune constraints to keep m proportional to n, $e.g.$, $m = 4n$)

ellipsoid method addresses both issues, but retains theoretical efficiency

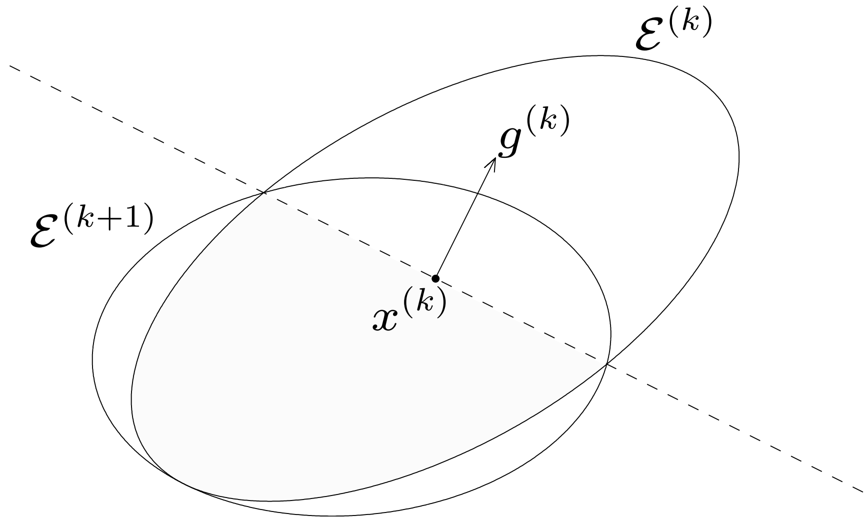
Ellipsoid algorithm for minimizing convex function

idea: localize $x^\star$ in an **ellipsoid** instead of a **polyhedron**

1. at iteration k we know $x^\star \in \mathcal{E}^{(k)}$
2. set $x^{(k)} :=$ center$(\mathcal{E}^{(k)})$; evaluate $g^{(k)} \in \partial f(x^{(k)})$
   ($g^{(k)} = \nabla f(x^{(k)})$ if f is differentiable)
3. hence we know

$$x^\star \in \mathcal{E}^{(k)} \cap \{z \mid g^{(k)T}(z - x^{(k)}) \leq 0\}$$

   (a half-ellipsoid)
4. set $\mathcal{E}^{(k+1)} :=$ minimum volume ellipsoid covering
   $\mathcal{E}^{(k)} \cap \{z \mid g^{(k)T}(z - x^{(k)}) \leq 0\}$



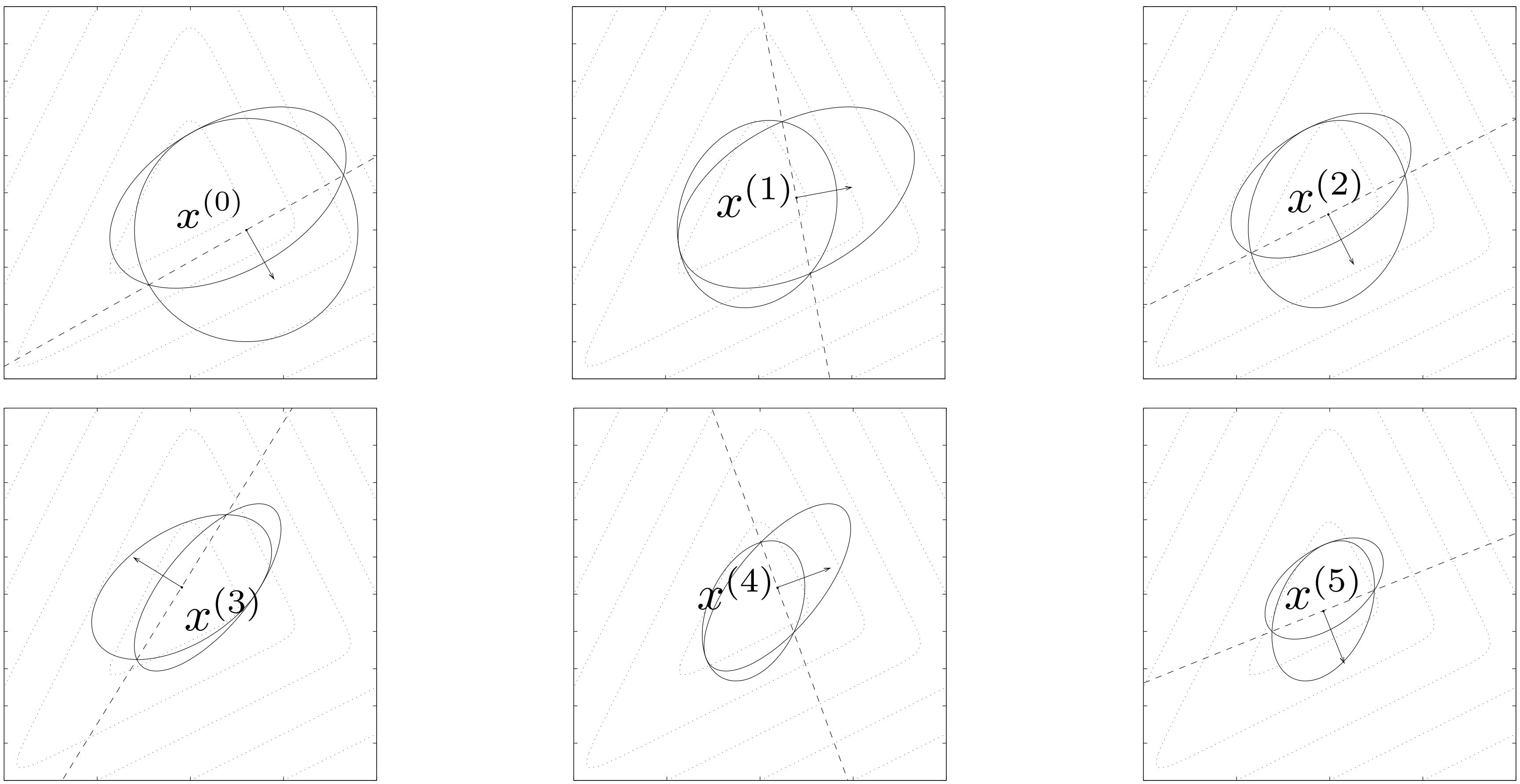
compared to cutting-plane methods:

- localization set doesn't grow more complicated
- easy to compute query point
- but, we add unnecessary points in step 4

Properties of ellipsoid method

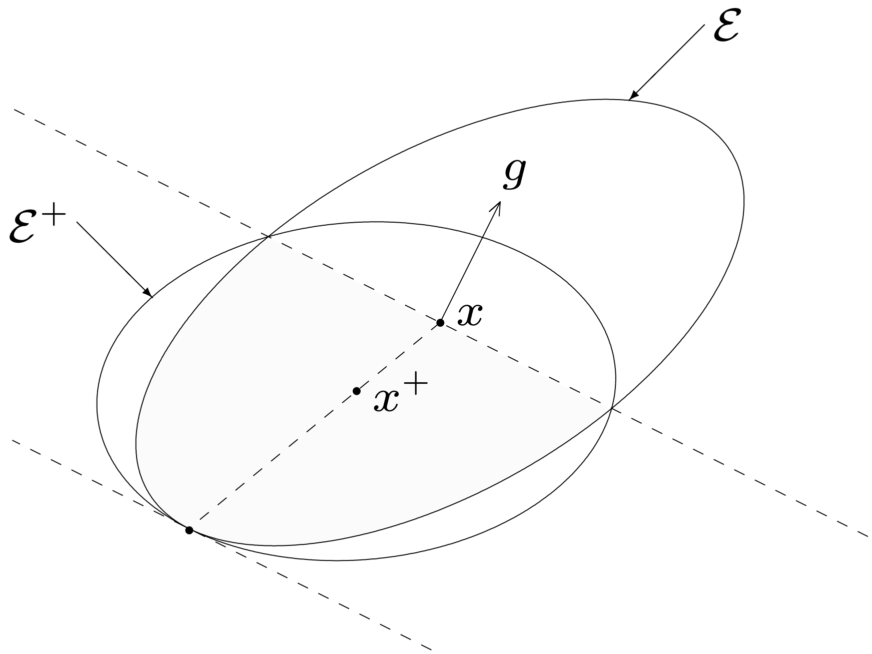
- reduces to bisection for $n = 1$
- simple formula for $\mathcal{E}^{(k+1)}$ given $\mathcal{E}^{(k)}$, $g^{(k)}$
- $\mathcal{E}^{(k+1)}$ can be larger than $\mathcal{E}^{(k)}$ in diameter (max semi-axis length), but is always smaller in volume
- $\mathbf{vol}(\mathcal{E}^{(k+1)}) < e^{-\frac{1}{2n}}\,\mathbf{vol}(\mathcal{E}^{(k)})$
  (volume reduction factor degrades rapidly with n, compared to CG or MVE cutting-plane methods)
- $\log \mathbf{vol}\,\mathcal{E}^{(k+1)} \leq \log \mathbf{vol}\,\mathcal{E}^{(k)} - 1/(2n)$
  (uncertainty in location of $x^\star$ decreases by a fixed number of bits each iteration)

Example



Updating the ellipsoid

$$\mathcal{E}(x, P) = \left\{ z \mid (z-x)^T P^{-1}(z-x) \le 1 \right\}$$



(for $n > 1$) minimum volume ellipsoid containing half-ellipsoid

$$\mathcal{E} \ \cap \ \left\{z \mid g^T(z - x) \leq 0\right\}$$

is given by

$$\begin{aligned} x^+ &= x - \frac{1}{n+1} P\tilde{g} \\ P^+ &= \frac{n^2}{n^2 - 1}\left(P - \frac{2}{n+1} P\tilde{g}\tilde{g}^T P\right) \end{aligned}$$

where $\tilde{g} = (1/\sqrt{g^T P g})g$

$P\tilde{g}$ is step from x to boundary of $\mathcal{E}$

Ellipsoid update — "Hessian" form

propagate $H = P^{-1}$ instead of P

$$
\begin{aligned}
x^+ &= x - \frac{1}{n+1} H^{-1}\tilde{g} \\
H^+ &= \left(1 - \frac{1}{n^2}\right)\left(H + \frac{2}{n-1}\tilde{g}\tilde{g}^T\right)
\end{aligned}
$$

where $\tilde{g} = (1/\sqrt{g^T H^{-1} g})g$

$H^{-1}\tilde{g}$ is step from x to boundary of $\mathcal{E}$

Simple stopping criterion

$$\begin{aligned} f(x^\star) &\geq f(x^{(k)}) + g^{(k)T}(x^\star - x^{(k)}) \\ &\geq f(x^{(k)}) + \inf_{z \in \mathcal{E}^{(k)}} g^{(k)T}(z - x^{(k)}) \\ &= f(x^{(k)}) - \sqrt{g^{(k)T} P^{(k)} g^{(k)}} \end{aligned}$$

second inequality holds since $x^\star \in \mathcal{E}_k$

simple stopping criterion:

$$\sqrt{g^{(k)T} P^{(k)} g^{(k)}} \leq \epsilon \quad \Longrightarrow \quad f(x^{(k)}) - f(x^\star) \leq \epsilon$$

Basic ellipsoid algorithm

ellipsoid described as $\mathcal{E}(x, P) = \{z \mid (z-x)^T P^{-1} (z-x) \leq 1\}$

given ellipsoid $\mathcal{E}(x, P)$ containing $x^\star$, accuracy $\epsilon > 0$

repeat

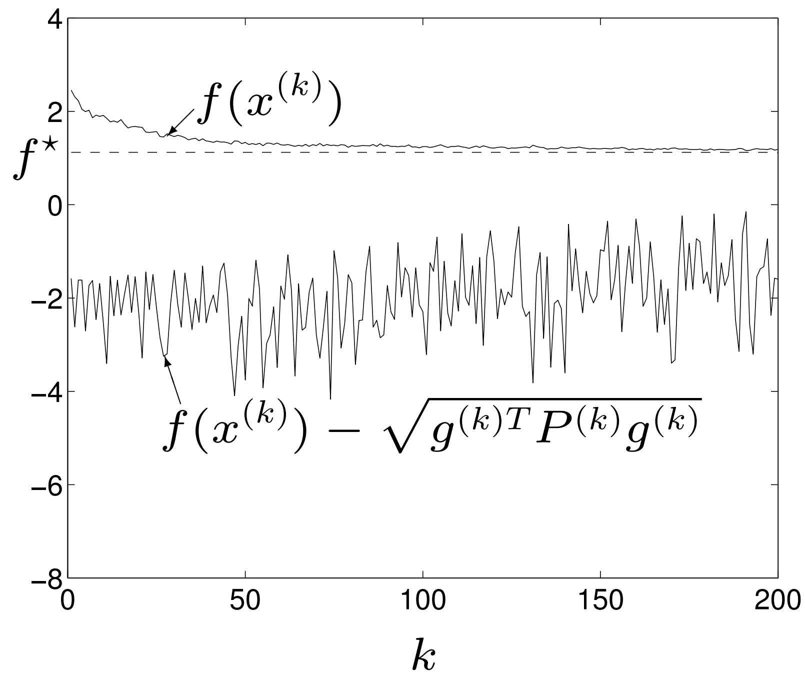
1. evaluate $g \in \partial f(x)$
2. if $\sqrt{g^T P g} \leq \epsilon$, return($x$)
3. update ellipsoid
   - 3a. $\tilde{g} := \frac{1}{\sqrt{g^T P g}} g$
   - 3b. $x := x - \frac{1}{n+1} P \tilde{g}$
   - 3c. $P := \frac{n^2}{n^2-1} \left(P - \frac{2}{n+1} P \tilde{g} \tilde{g}^T P \right)$

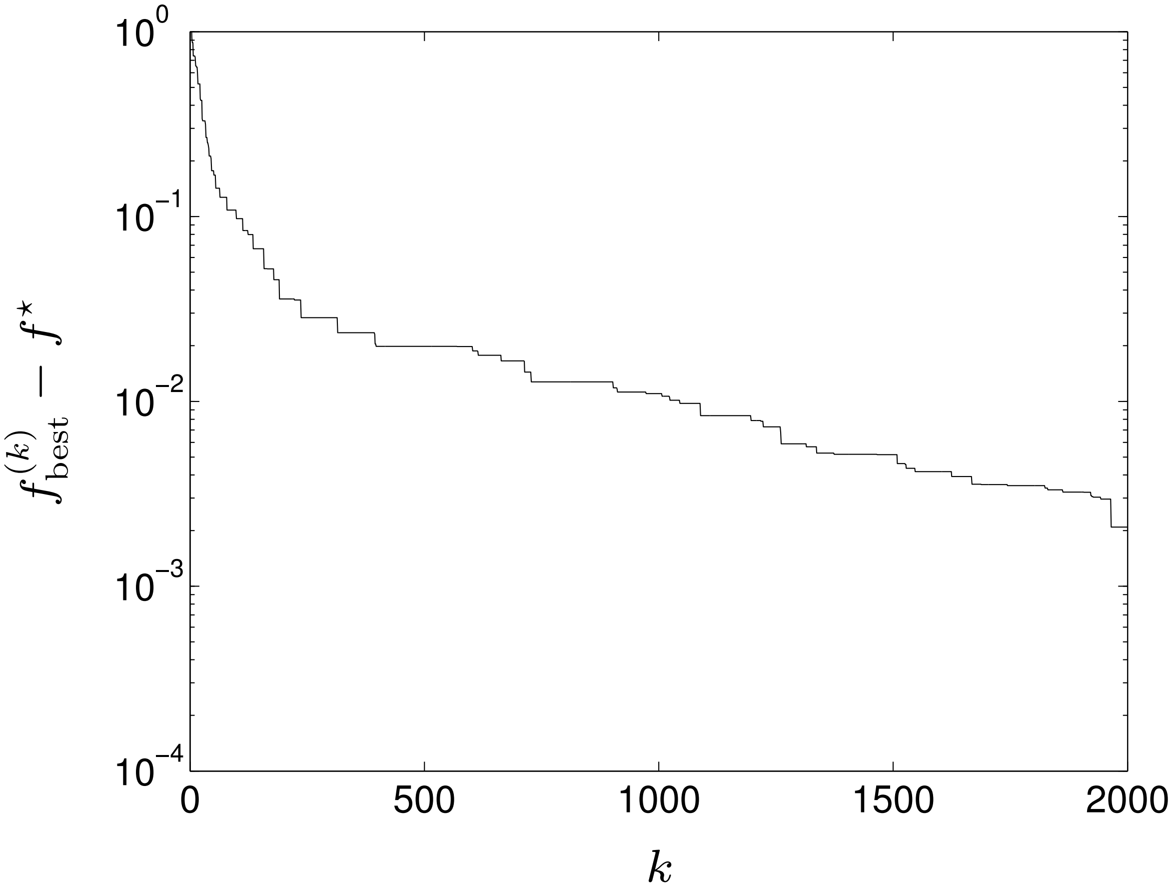
Interpretation

- change coordinates so uncertainty is isotropic (same in all directions), $i.e.$, $\mathcal{E}$ is unit ball
- take subgradient step with fixed length $1/(n+1)$
- Shor calls ellipsoid method 'gradient method with space dilation in direction of gradient' (which, strangely enough, didn't catch on)

Example

PWL function $f(x) = \max_{i=1}^{m}(a_i^T x + b_i)$, with $n = 20$, $m = 100$




10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
$f_{\rm best}^{(k)} - f^\star$
0
500
1000
1500
2000
k

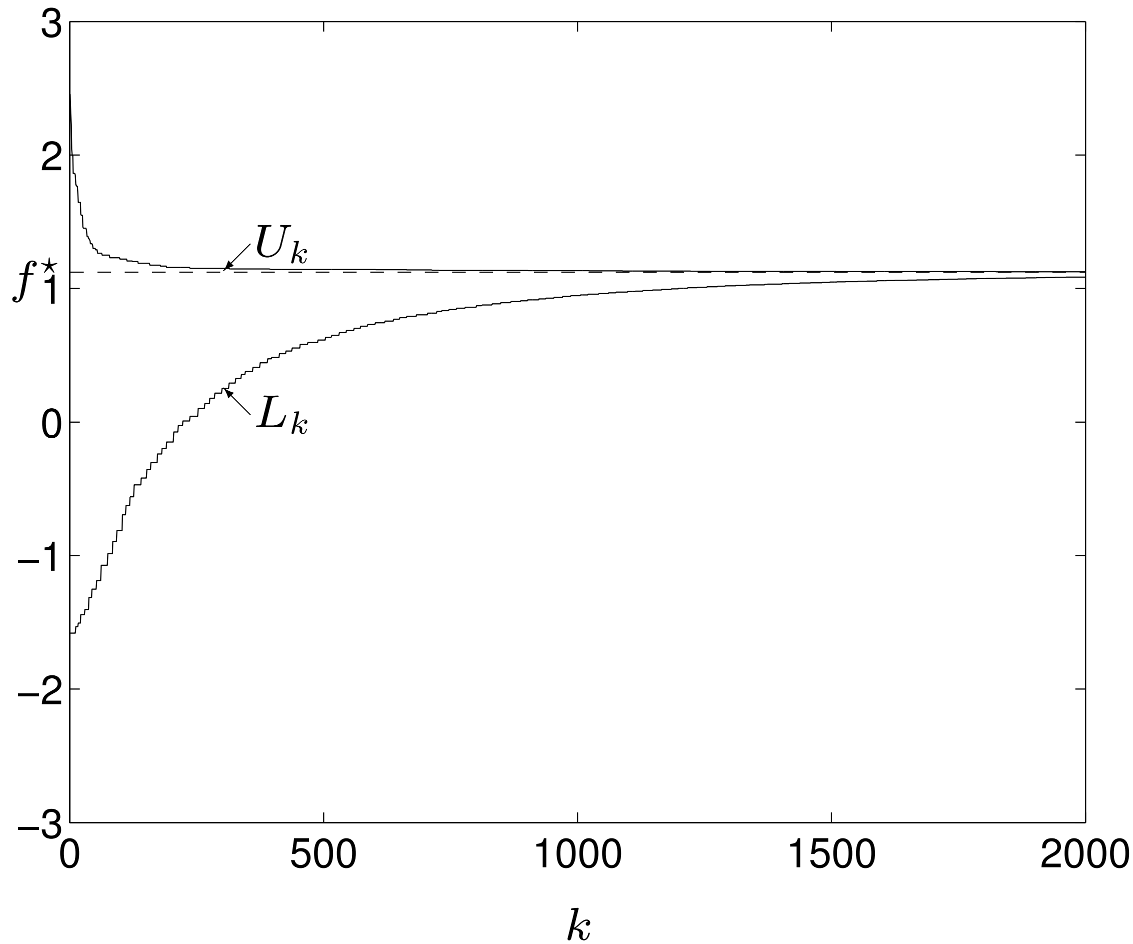
Improvements

- keep track of best upper and lower bounds:

$$u_k = \min_{i=1,\ldots,k} f(x^{(i)}), \qquad l_k = \max_{i=1,\ldots,k} \left(f(x^{(i)}) - \sqrt{g^{(i)T} P^{(i)} g^{(i)}} \right)$$

  stop when $u_k - l_k \leq \epsilon$

- can propagate Cholesky factor of P
  (avoids problem of $P \not\succ 0$ due to numerical roundoff)


3
2
$f^\star$
1
0
–1
–2
–3
0
500
1000
1500
2000
U_k
L_k
k

Proof of convergence

assumptions:

- f is Lipschitz: $|f(y) - f(x)| \leq G\|y - x\|$
- $\mathcal{E}^{(0)}$ is ball with radius R

suppose $f(x^{(i)}) > f^\star + \epsilon$, $i = 0, \ldots, k$

then

$$f(x) \leq f^\star + \epsilon \Longrightarrow x \in \mathcal{E}^{(k)}$$

since at iteration i we only discard points with $f \geq f(x^{(i)})$

from Lipschitz condition,

$$\|x - x^\star\| \leq \epsilon/G \Longrightarrow f(x) \leq f^\star + \epsilon \Longrightarrow x \in \mathcal{E}^{(k)}$$

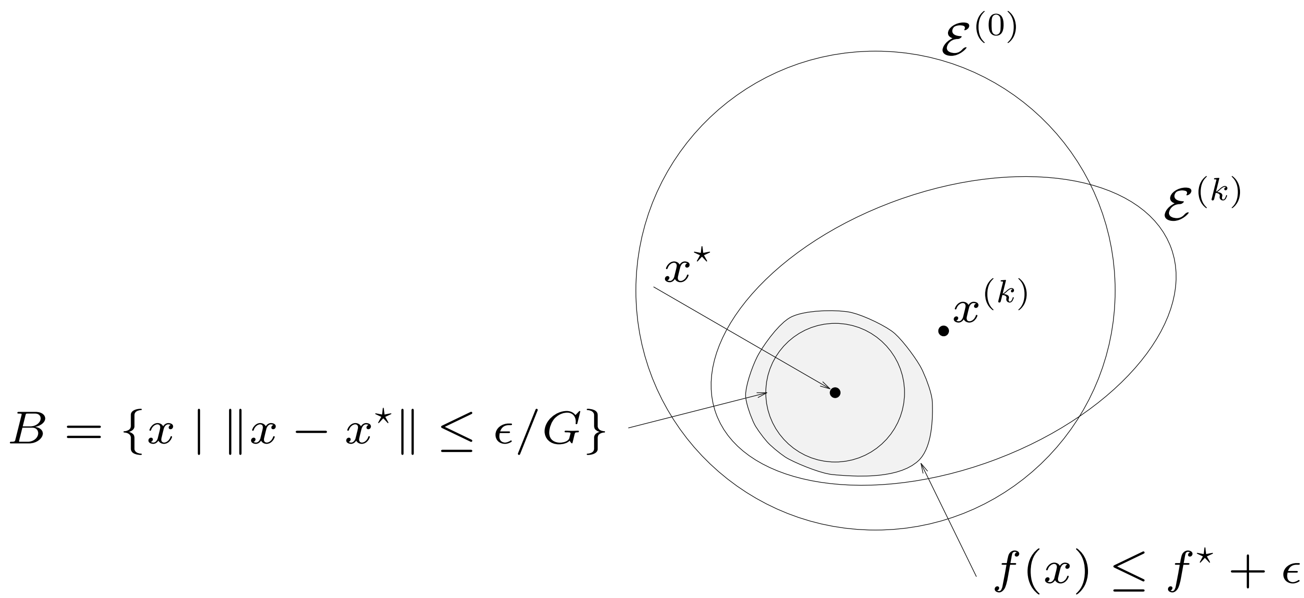
so $B = \{x \mid \|x - x^\star\| \leq \epsilon/G\} \subseteq \mathcal{E}^{(k)}$

hence $\mathbf{vol}(B) \leq \mathbf{vol}(\mathcal{E}^{(k)})$, so

$$\alpha_n(\epsilon/G)^n \leq e^{-k/2n}\,\mathbf{vol}(\mathcal{E}^{(0)}) = e^{-k/2n}\alpha_n R^n$$

(α_n is volume of unit ball in $\mathbf{R}^n$)

therefore $k \leq 2n^2 \log(RG/\epsilon)$



conclusion: for $k > 2n^2 \log(RG/\epsilon)$,

$$\min_{i=0,\ldots,k} f(x^{(i)}) \leq f^\star + \epsilon$$

Interpretation of complexity

since $x^\star \in \mathcal{E}_0 = \{x \mid \|x - x^{(0)}\| \leq R\}$, our prior knowledge of $f^\star$ is

$$f^\star \in [f(x^{(0)}) - GR, f(x^{(0)})]$$

our prior uncertainty in $f^\star$ is GR

after k iterations our knowledge of $f^\star$ is

$$f^\star \in \left[\min_{i=0,\ldots,k} f(x^{(i)}) - \epsilon, \min_{i=0,\ldots,k} f(x^{(i)}) \right]$$

posterior uncertainty in $f^\star$ is $\leq \epsilon$

iterations required:

$$2n^2 \log \frac{RG}{\epsilon} = 2n^2 \log \frac{\text{prior uncertainty}}{\text{posterior uncertainty}}$$

efficiency: $0.72/n^2$ bits per subgradient evaluation

Deep cut ellipsoid method

minimum volume ellipsoid containing ellipsoid intersected with halfspace

$$\mathcal{E} \ \cap \ \left\{ z \mid g^T(z-x) + h \le 0 \right\}$$

with $h \ge 0$, is given by

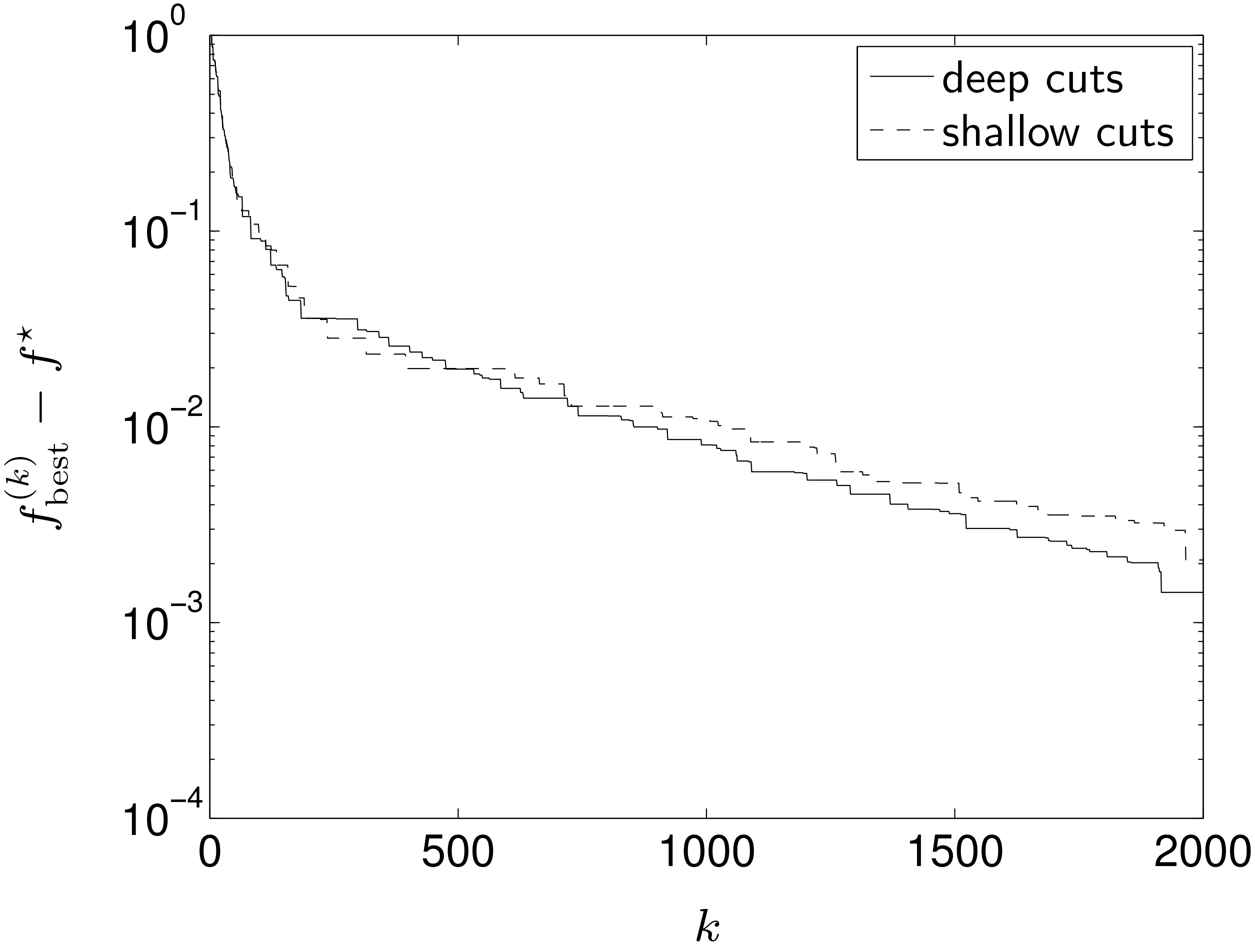
$$\begin{aligned} x^+ &= x - \frac{1+\alpha n}{n+1} P\tilde{g} \\ P^+ &= \frac{n^2(1-\alpha^2)}{n^2-1} \left(P - \frac{2(1+\alpha n)}{(n+1)(1+\alpha)} P\tilde{g}\tilde{g}^T P \right) \end{aligned}$$

where

$$\tilde{g} = \frac{g}{\sqrt{g^T P g}}, \qquad \alpha = \frac{h}{\sqrt{g^T P g}}$$

(if $\alpha > 1$, intersection is empty)

Ellipsoid method with deep objective cuts



Inequality constrained problems

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \end{array}$$

- if $x^{(k)}$ feasible, update ellipsoid with objective cut

$$g_0^T(z - x^{(k)}) + f_0(x^{(k)}) - f_{\mathrm{best}}^{(k)} \leq 0, \qquad g_0 \in \partial f_0(x^{(k)})$$

$f_{\mathrm{best}}^{(k)}$ is best objective value of feasible iterates so far

- if $x^{(k)}$ infeasible, update ellipsoid with feasibility cut

$$g_j^T(z - x^{(k)}) + f_j(x^{(k)}) \leq 0, \qquad g_j \in \partial f_j(x^{(k)})$$

assuming $f_j(x^{(k)}) > 0$

Stopping criterion

if $x^{(k)}$ is feasible, we have lower bound on $p^\star$ as before:

$$p^\star \geq f_0(x^{(k)}) - \sqrt{g_0^{(k)T} P^{(k)} g_0^{(k)}}$$

if $x^{(k)}$ is infeasible, we have for all $x \in \mathcal{E}^{(k)}$

$$\begin{aligned} f_j(x) &\geq f_j(x^{(k)}) + g_j^{(k)T}(x - x^{(k)}) \\ &\geq f_j(x^{(k)}) + \inf_{z \in \mathcal{E}^{(k)}} g^{(k)T}(z - x^{(k)}) \\ &= f_j(x^{(k)}) - \sqrt{g_j^{(k)T} P^{(k)} g_j^{(k)}} \end{aligned}$$

hence, problem is infeasible if for some j,

$$f_j(x^{(k)}) - \sqrt{g_j^{(k)T} P^{(k)} g_j^{(k)}} > 0$$

stopping criteria:

- if $x^{(k)}$ is feasible and $\sqrt{g_0^{(k)T} P^{(k)} g_0^{(k)}} \leq \epsilon$ ($x^{(k)}$ is ϵ-suboptimal)
- if $f_j(x^{(k)}) - \sqrt{g_j^{(k)T} P^{(k)} g_j^{(k)}} > 0$ (problem is infeasible)