Ellipsoid Method

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

Challenges in cutting-plane methods

- can be difficult to compute appropriate next query point
- localization polyhedron grows in complexity as algorithm progresses

can get around these challenges . . .

ellipsoid method is another approach

- developed in 70s by Shor and Yudin
- used in 1979 by Khachian to give polynomial time algorithm for LP

Ellipsoid algorithm

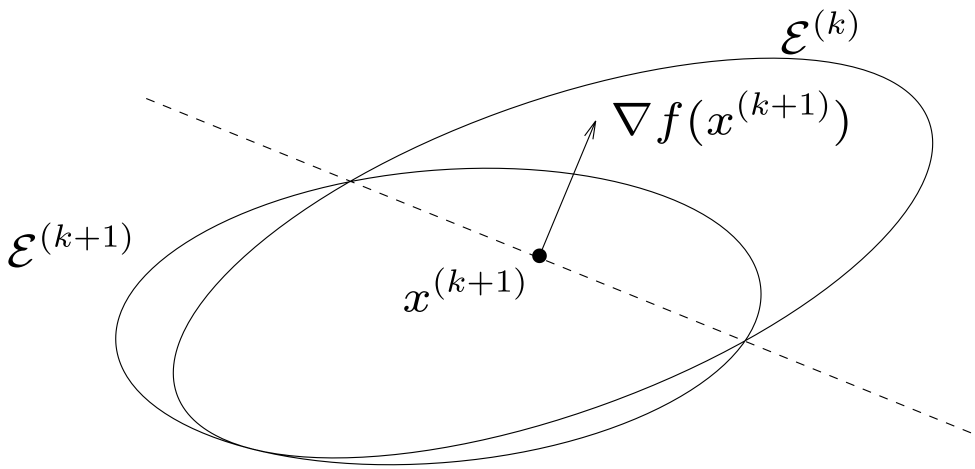
idea: localize $x^\star$ in an **ellipsoid** instead of a **polyhedron**

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

$$x^\star \in \mathcal{E}^{(k)} \cap \{z \mid \nabla f(x^{(k+1)})^T (z - x^{(k+1)}) \leq 0\}$$

   (a half-ellipsoid)

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



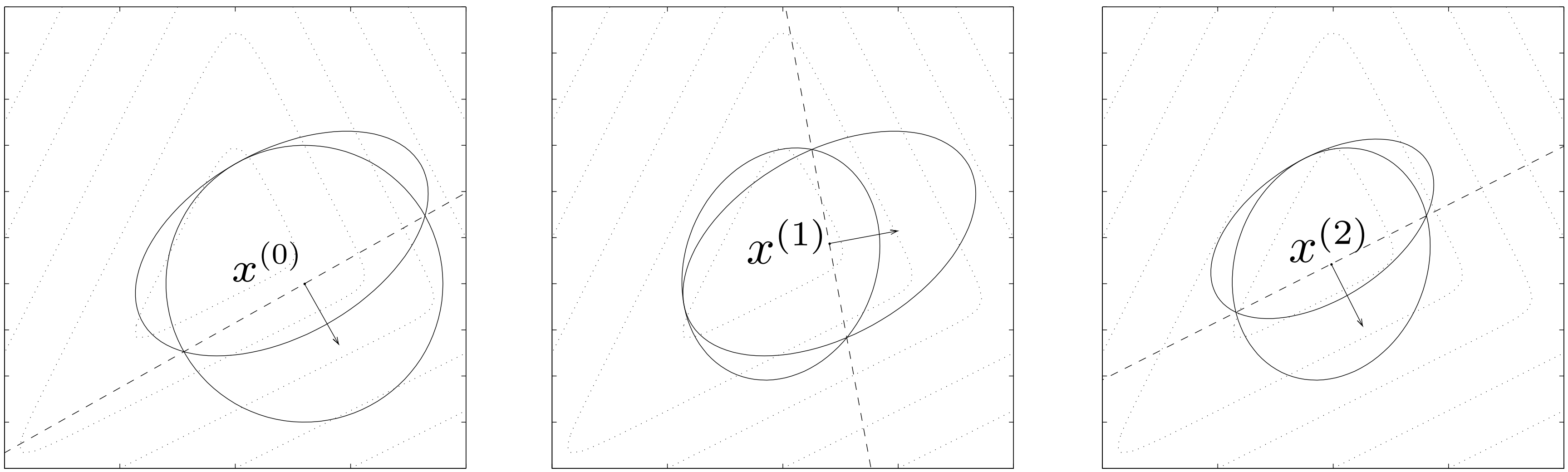
compared to cutting-plane method:

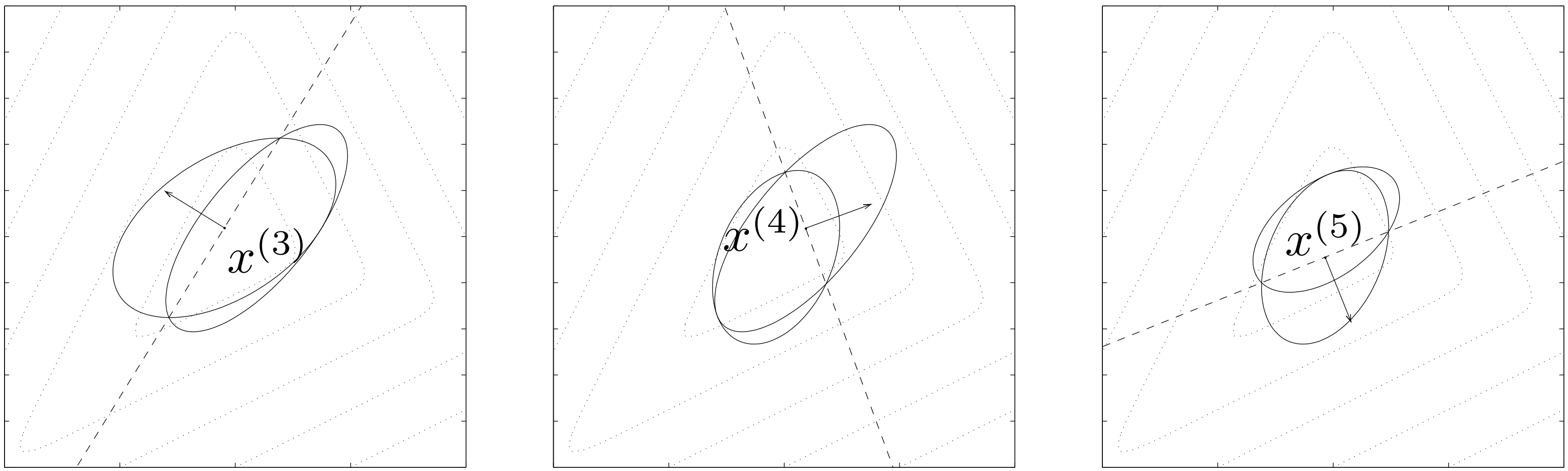
- localization set doesn't grow more complicated
- easy to compute query point
- but, we add unnecessary points in step 4

Properties of ellipsoid method

- reduces to bisection for $n = 1$
- simple formula for $\mathcal{E}^{(k+1)}$ given $\mathcal{E}^{(k)}$, $\nabla f(x^{(k+1)})$
- $\mathcal{E}^{(k+1)}$ can be larger than $\mathcal{E}^{(k)}$ in diameter (max semi-axis length), but is always smaller in volume
- $\mathbf{vol}(\mathcal{E}^{(k+1)}) < e^{-\frac{1}{2n}}\mathbf{vol}(\mathcal{E}^{(k)})$
(note that volume reduction factor depends on n)

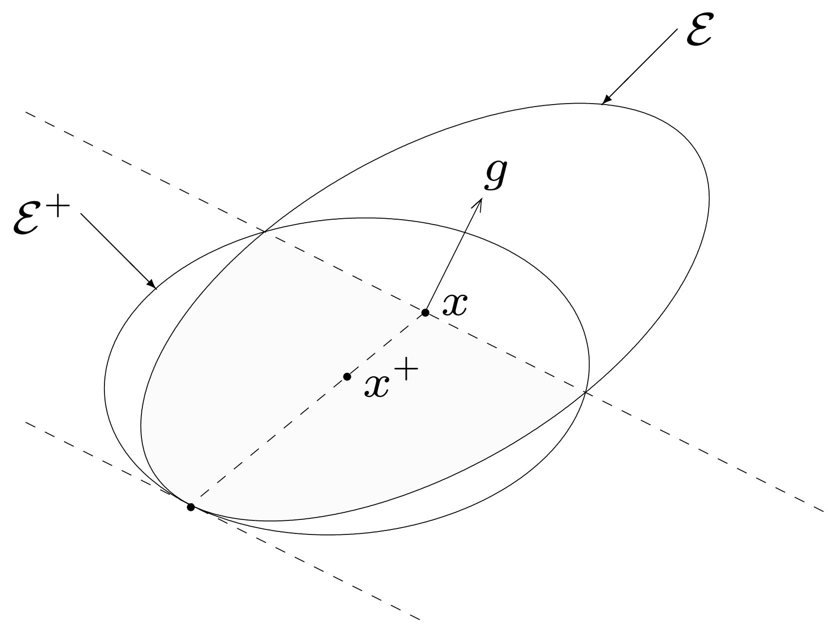
Example




$x^{(3)}$
$x^{(4)}$
$x^{(5)}$

Updating the ellipsoid

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



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

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

is given by

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

where $\tilde{g} \stackrel{\Delta}{=} g \Big/ \sqrt{g^T A g}$

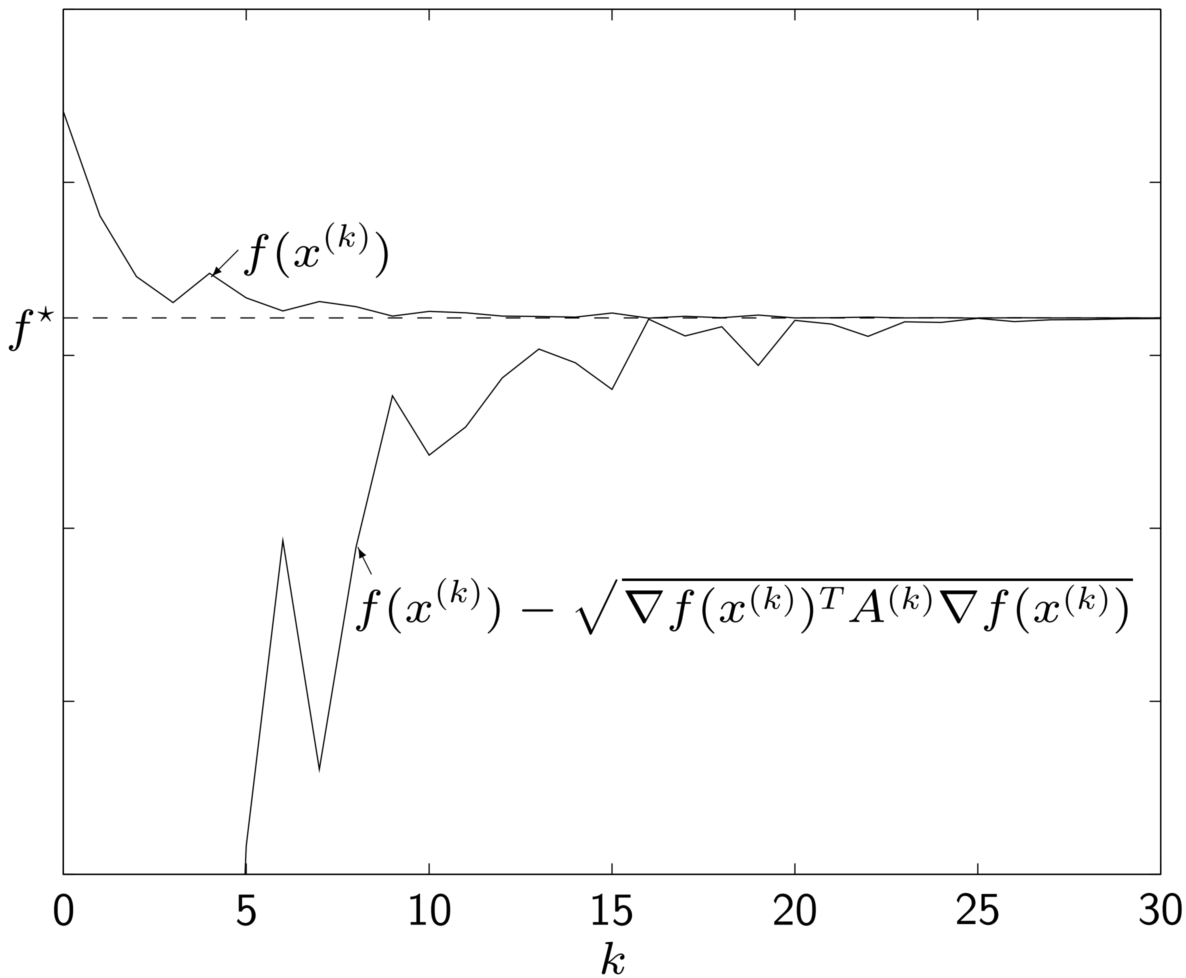
Stopping criterion

$x^\star \in \mathcal{E}_k$, so

$$\begin{array}{rcl} f(x^\star) & \geq & f(x^{(k)}) + \nabla f(x^{(k)})^T(x^\star - x^{(k)}) \\ & \geq & f(x^{(k)}) + \inf_{x \in \mathcal{E}^{(k)}} \nabla f(x^{(k)})^T(x - x^{(k)}) \\ & = & f(x^{(k)}) - \sqrt{\nabla f(x^{(k)})^T A^{(k)} \nabla f(x^{(k)})} \end{array}$$

simple stopping criterion:

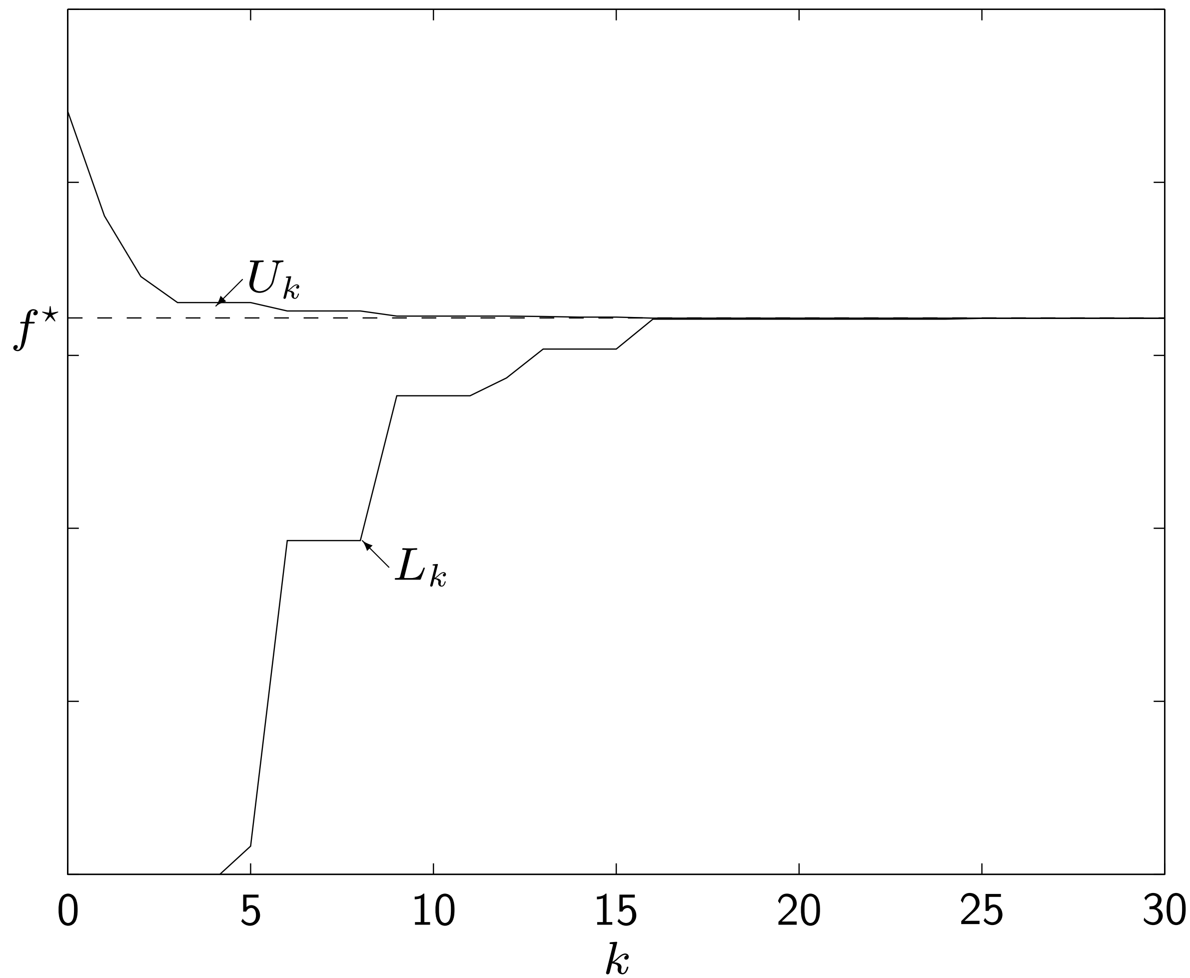
$$\sqrt{\nabla f(x^{(k)})^T A^{(k)} \nabla f(x^{(k)})} \leq \epsilon$$



$f(x^{(k)})$
$f^\star$
$f(x^{(k)}) - \sqrt{\nabla f(x^{(k)})^T A^{(k)} \nabla f(x^{(k)})}$
0
5
10
15
20
25
30
k

more sophisticated stopping criterion: $U_k - L_k \leq \epsilon$, where

$$\begin{aligned} U_k &= \min_{i \leq k} f(x^{(i)}) \\ L_k &= \max_{i \leq k} \left(f(x^{(i)}) - \sqrt{\nabla f(x^{(i)})^T A^{(i)} \nabla f(x^{(i)})} \right) \end{aligned}$$



U_k
$f^\star$
L_k
0
5
10
15
20
25
30
k

Basic ellipsoid algorithm

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

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

repeat

1. evaluate $\nabla f(x)$ (or $g \in \partial f(x)$)
2. if $\sqrt{\nabla f(x)^T A \nabla f(x)} \leq \epsilon$, return($x$)
3. update ellipsoid

   3a. $\tilde{g} := \nabla f(x) \Big/ \sqrt{\nabla f(x)^T A \nabla f(x)}$

   3b. $x := x - \frac{1}{n+1} A\tilde{g}$

   3c. $A := \frac{n^2}{n^2-1}\left(A - \frac{2}{n+1} A\tilde{g}\tilde{g}^T A\right)$

Interpretation

- change coordinates so uncertainty ($\mathcal{E}$) is unit ball
- take gradient (or subgradient) step with fixed length $1/(n+1)$

properties:

- can propagate Cholesky factor of A; get $O(n^2)$ update
- **not** a descent method
- often slow but robust in practice

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,

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

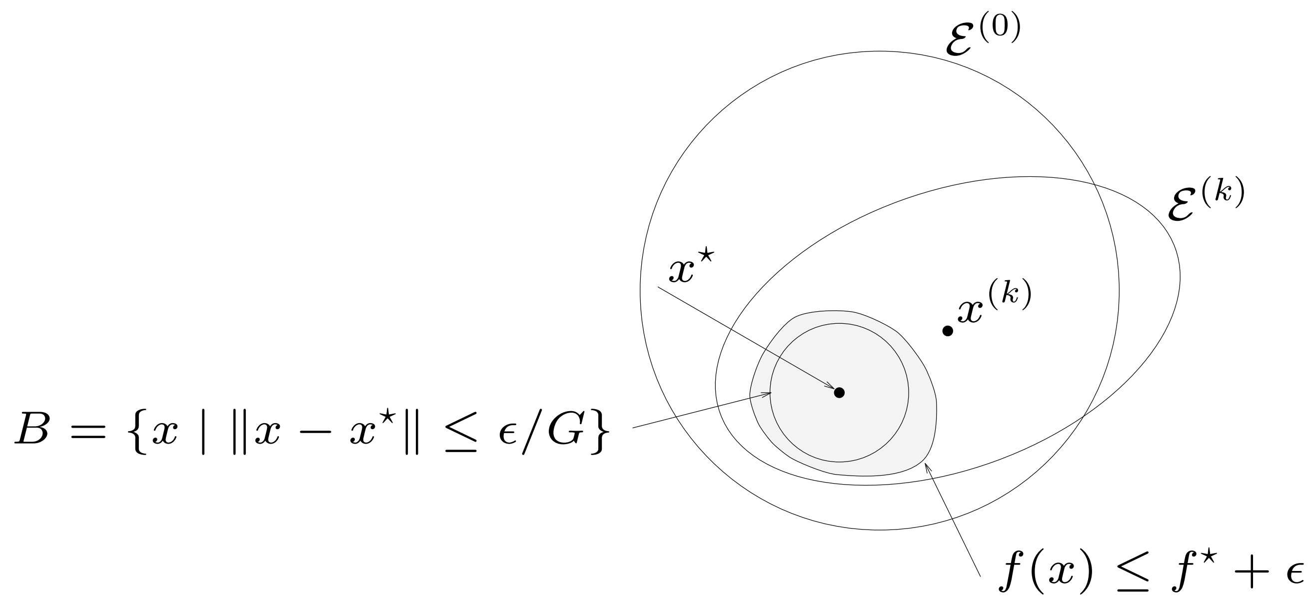
so $B = \{x \mid \|x - x^\star\| \le \epsilon/G\} \subseteq \mathcal{E}^{(k)}$

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

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

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

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



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 gradient evaluation (degrades with n)

Inequality constrained problems

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

same idea: maintain ellipsoids $\mathcal{E}^{(k)}$ that

- contain $x^\star$
- decrease in volume to zero

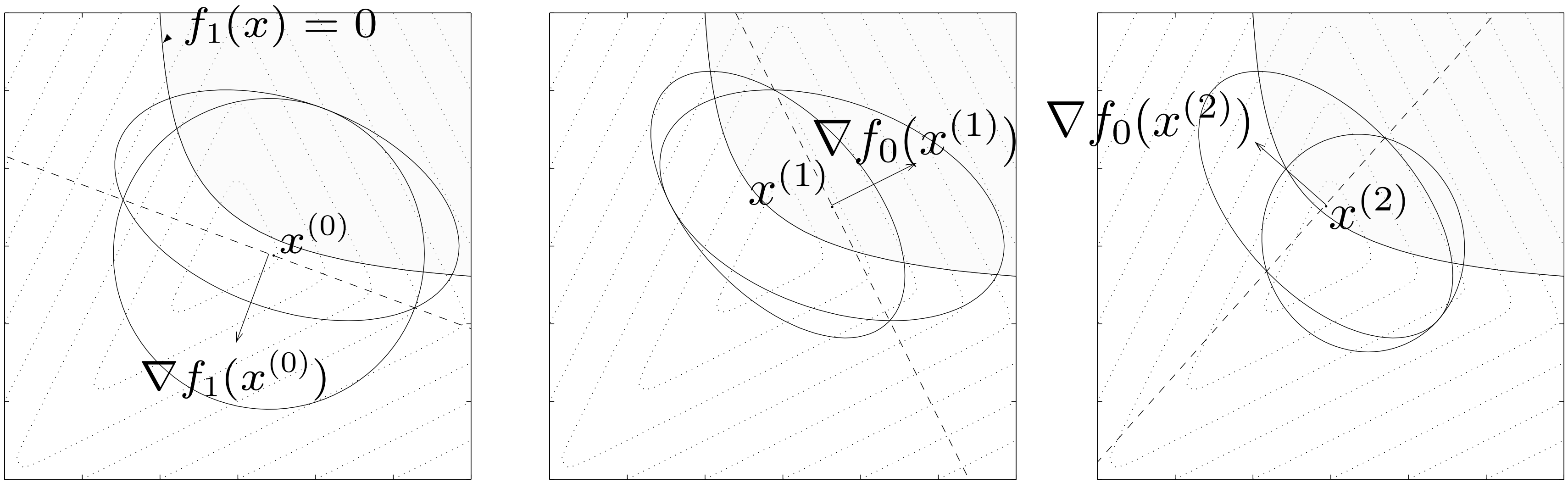
case 1: $x^{(k)}$ feasible, *i.e.*, $f_i(x^{(k)}) \leq 0$, $i = 1, \ldots, m$

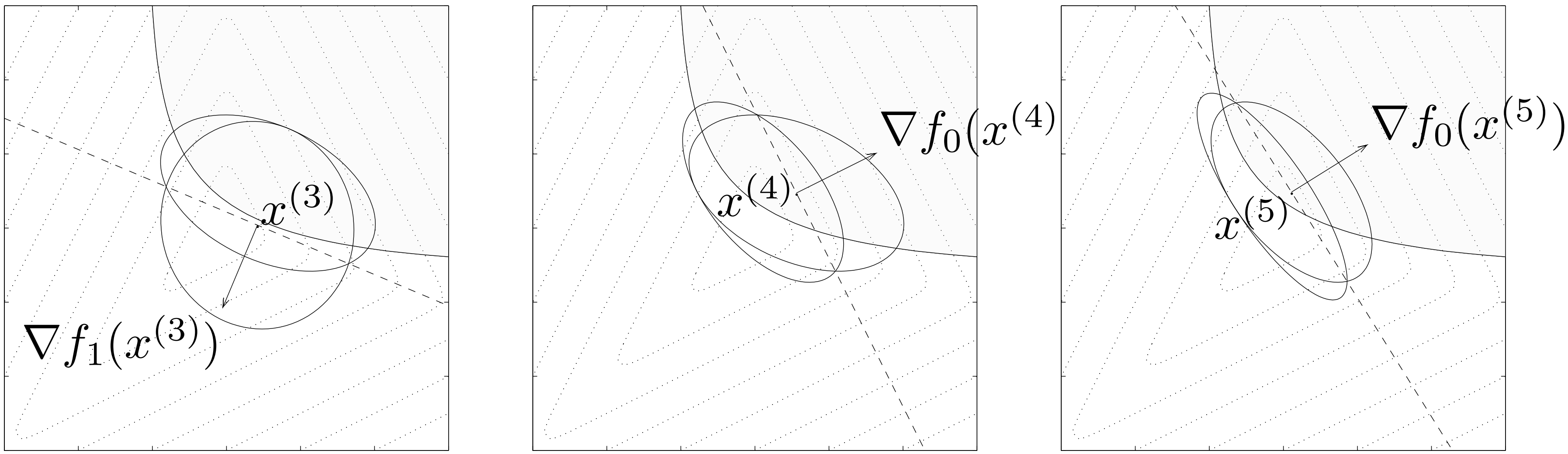
- then do usual update of $\mathcal{E}^{(k)}$ based on $\nabla f_0(x^{(k)})$
- rules out halfspace of points with larger function value than current point

case 2: $x^{(k)}$ infeasible, say, $f_j(x^{(k)}) > 0$;

- then $\nabla f_j(x^{(k)})^T(x - x^{(k)}) \geq 0 \Longrightarrow f_j(x) > 0 \Longrightarrow x$ infeasible so update $\mathcal{E}^{(k)}$ based on $\nabla f_j(x^{(k)})$
- rules out halfspace of infeasible points

Example





$x^{(3)}$
$\nabla f_1(x^{(3)})$
$x^{(4)}$
$\nabla f_0(x^{(4)}$
$x^{(5)}$
$\nabla f_0(x^{(5)})$

Stopping criterion

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

$$f^\star \geq f(x^{(k)}) - \sqrt{\nabla f(x^{(k)})^T A^{(k)} \nabla f(x^{(k)})}$$

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

$$\begin{array}{rcl} f_j(x) & \geq & f_j(x^{(k)}) + \nabla f_j(x^{(k)})^T (x - x^{(k)}) \\ & \geq & f_j(x^{(k)}) + \inf_{x \in \mathcal{E}^{(k)}} \nabla f_j(x^{(k)})^T (x - x^{(k)}) \\ & = & f_j(x^{(k)}) - \sqrt{\nabla f_j(x^{(k)})^T A^{(k)} \nabla f_j(x^{(k)})} \end{array}$$

hence, problem is infeasible if for some j,

$$f_j(x^{(k)}) - \sqrt{\nabla f_j(x^{(k)})^T A^{(k)} \nabla f_j(x^{(k)})} > 0$$

stopping criteria:

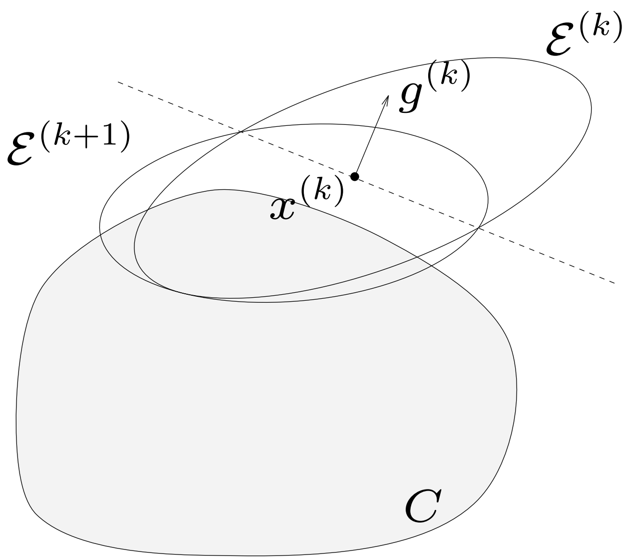
- if $x^{(k)}$ is feasible and $\sqrt{\nabla f_0(x^{(k)})^T A^{(k)} \nabla f_0(x^{(k)})} \leq \epsilon$ ($x^{(k)}$ is ϵ-suboptimal)
- if $f_j(x^{(k)}) - \sqrt{\nabla f_j(x^{(k)})^T A^{(k)} \nabla f_j(x^{(k)})} > 0$ (problem is infeasible)

Ellipsoid method for feasibility

abstract feasibility problem: find $x \in C \subset \mathbf{R}^n$ or determine $C = \emptyset$

separating hyperplane oracle: for any x, oracle either

- confirms $x \in C$, or
- returns $g \neq 0$ s.t. $z \in C \Rightarrow g^T(z - x) \leq 0$



start with $\mathcal{E}^{(0)}$ which intersects C

1. If $x^{(k)} :=$ center$(\mathcal{E}^{(k)}) \in C$, quit. Else, compute $g \neq 0$, s.t. $x \in C \Rightarrow g^T(x - x^{(k)}) \leq 0$
2. $\mathcal{E}^{(k+1)} :=$ minimum volume ellipsoid covering

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

Example