Lecture 12
Ellipsoid method

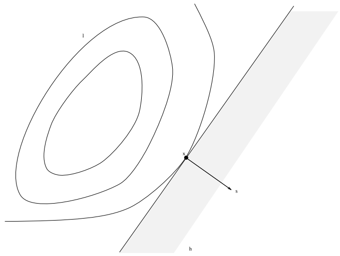
- idea of localization methods
- bisection on **R**
- center of gravity algorithm
- ellipsoid method

Localization

- $f : \mathbf{R}^n \to \mathbf{R}$ convex (and for now, differentiable)
- **problem:** minimize f
- **oracle model:** for any x we can evaluate f and $\nabla f(x)$ (at some cost)

recall: $f(x) \geq f(x_0) + \nabla f(x_0)^T(x - x_0)$

hence $\nabla f(x_0)^T(x - x_0) \geq 0 \Rightarrow f(x) \geq f(x_0)$



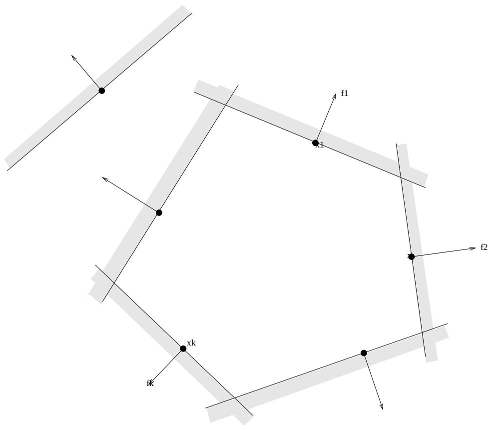
by evaluating ∇f we rule out a halfspace in our search for $x^\star$:

$$x^\star \in \{x \mid \nabla f(x_0)^T(x - x_0) \leq 0\}$$

idea: get one bit of info by evaluating ∇f

suppose we have evaluated $\nabla f(x_1), \ldots, \nabla f(x_k)$

then we know $x^\star \in \{x \mid \nabla f(x_i)^T(x - x_i) \leq 0\}$



on the basis of $\nabla f(x_1), \ldots, \nabla f(x_k)$, we have **localized** $x^\star$ to a polyhedron

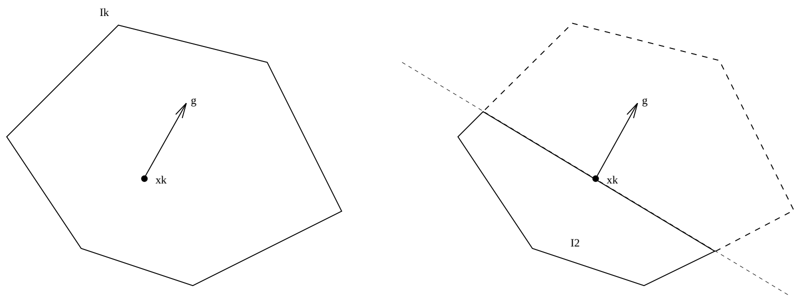
What is a 'good' point x_{k+1} at which to evaluate ∇f?

localization algorithm (idea)

1. after iteration $k-1$ we know $x^\star \in C^{(k-1)}$:

   $$C^{(k-1)} = \{x \mid \nabla f(x^{(i)})^T(x - x^{(i)}) \leq 0,\ i = 1, \ldots, k-1\}$$

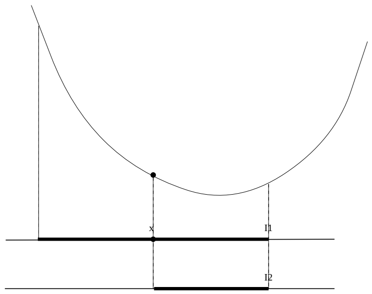
2. evaluate $\nabla f(x^k)$ for some $x^{(k)} \in C^{(k-1)}$
3. $C^{(k)} := C^{(k-1)} \cap \{x \mid \nabla f(x^{(k)})^T(x - x^{(k)}) \leq 0\}$



– $C^{(k)}$ gives our 'uncertainty' of $x^\star$ at iteration k
– pick $x^{(k)}$ so that $C^{(k+1)}$ is as 'small' as possible
– clearly want $x^{(k)}$ near center of $C^{(k)}$

Example: bisection on R

- $f : \mathbf{R} \to \mathbf{R}$
- $C^{(k)}$ is interval
- obvious choice: $x^{(k+1)} :=$ midpoint$(C^{(k)})$



bisection algorithm

given interval $C = [l, u]$ containing $x^\star$

repeat

1. $x := (l+u)/2$
2. evaluate $f'(x)$
3. if $f'(x) < 0$, $l := x$; else $u := x$

we have length$(C^{(k+1)}) =$ length$(C^{(k)})/2$,

so length$(C^{(k)}) = 2^{-k}$length$(C^{(0)})$

interpretation:

- length$(C^{(k)})$ measures our uncertainty in $x^\star$
- uncertainty is halved at each iteration (get exactly one bit info about $x^\star$ per iteration)
- $\#$ steps required for uncertainty $\leq \epsilon$:

$$\log_2 \frac{\text{length}(C^{(0)})}{\epsilon} = \log_2 \frac{\text{initial uncertainty}}{\text{final uncertainty}}$$

question:

- can bisection be extended to $\mathbf{R}^n$?
- or is it special since $\mathbf{R}$ is linear ordering?

Center of gravity algorithm

take $x^{(k+1)} = \mathsf{CG}(C^{(k)})$ (center of gravity)

$$\mathsf{CG}(C^{(k)}) = \int_{C^{(k)}} x \, dx \Big/ \int_{C^{(k)}} dx$$

theorem. if $C \subseteq \mathbf{R}^n$ convex, $x_{\mathsf{cg}} = \mathsf{CG}(C)$, $g \neq 0$,

$$\begin{aligned} \mathbf{vol}\left(C \cap \{x \mid g^T(x - x_{\mathsf{cg}}) \leq 0\}\right) &\leq (1 - 1/e)\mathbf{vol}(C) \\ &\approx 0.63 \, \mathbf{vol}(C) \end{aligned}$$

(independent of n)

hence $\mathbf{vol}(C^{(k)}) \leq 0.63^k \mathbf{vol}(C^{(0)})$

- $\mathbf{vol}(C^{(k)})$ measures uncertainty at iteration k
- uncertainty reduced by 0.63 or more per iteration
- max. # steps required for uncertainty $\leq \epsilon$:

$$1.85 \log \frac{\mathbf{vol}(C^{(0)})}{\epsilon} = 1.85 \log \frac{\text{initial uncertainty}}{\text{final uncertainty}}$$

- from this can prove $f(x^{(k)}) \to f(x^\star)$ (later)

advantages of CG-method

- guaranteed convergence
- number of steps independent of dimension n

disadvantages

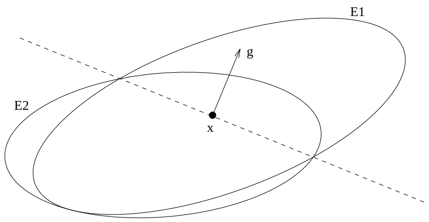
- finding $x^{(k+1)} = \mathsf{CG}(C^{(k)})$ is harder than original problem
- $C^{(k)}$ becomes more complex as k increases (removing redundant constraints is harder than solving original problem)

(but, can modify CG-method to work)

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 $\nabla f(x^{(k+1)})$
3. hence we know
$$x^\star \in \mathcal{E}^{(k)} \cap \{z \mid \nabla f(x^{(k)})^T(z - x^{(k)}) \le 0\}$$
(a half-ellipsoid)
4. set $\mathcal{E}^{(k+1)} :=$ minimum volume ellipsoid covering $\mathcal{E}^{(k)} \cap \{z \mid \nabla f(x^{(k+1)})^T(z - x^{(k+1)}) \le 0\}$



compared to CG method:

– localization set doesn't grow more complicated,

– but, we add unnecessary points in step 4

properties of ellipsoid method

– reduces to bisection for $n = 1$

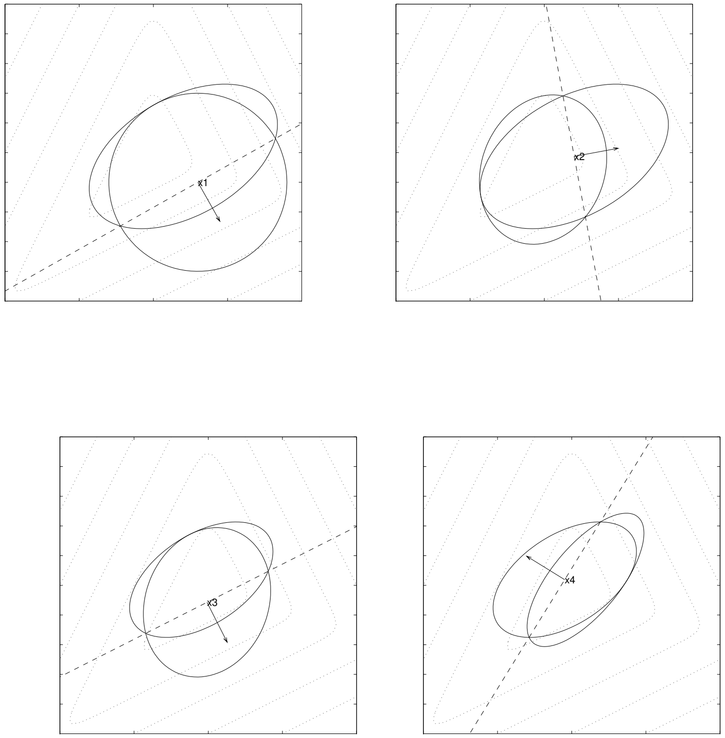
– simple formula for $\mathcal{E}^{(k+1)}$ given $\mathcal{E}^{(k)}$, $\nabla f(x^{(k+1)})$

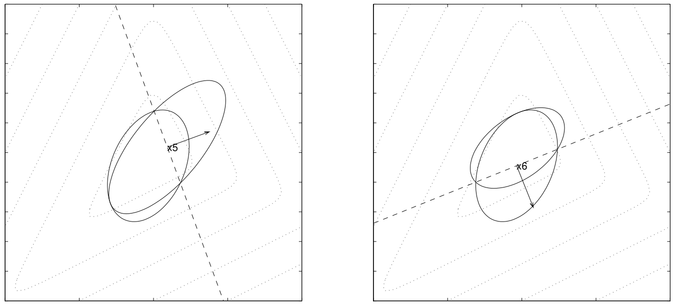
– $\mathcal{E}^{(k+1)}$ can be larger than $\mathcal{E}^{(k)}$ in diameter (max semi-axis length), but is always smaller in volume

– **vol**$(\mathcal{E}^{(k+1)}) < e^{-\frac{1}{2n}}$**vol**$(\mathcal{E}^{(k)})$
(note that volume reduction factor depends on n)

– extends to nondifferentiable, constrained, quasiconvex (more later)

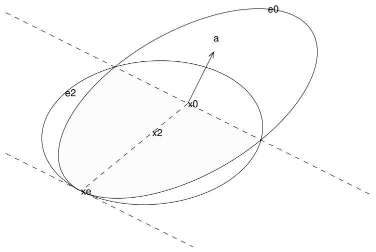
Example





Updating the ellipsoid

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



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

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

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 $\tilde{g} \triangleq g \Big/ \sqrt{g^T A g}$

Proof of convergence

assumptions:

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

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

then $f(x) \le f^\star + \epsilon \Longrightarrow x \in \mathcal{E}^{(k)}$
since at iter i we only discard points with $f \ge f(x^{(i)})$

from Lipschitz condition,

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

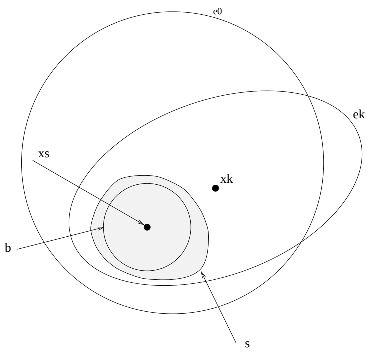
i.e., $B = \{x \mid \|x - x^\star\| \le \epsilon/G\} \subseteq \mathcal{E}^{(k)}$

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

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

(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:

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

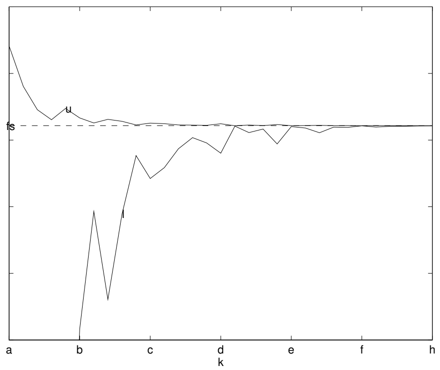
efficiency: $0.72/n^2$ bits per gradient evaluation
(note: degrades with n)

Stopping criterion

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

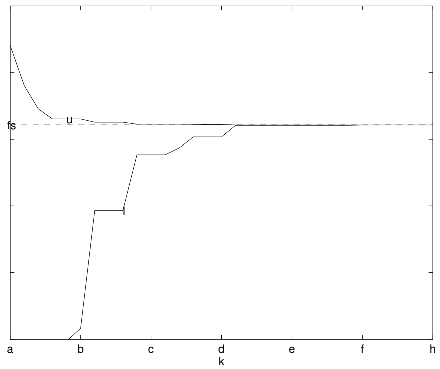
simple stopping criterion:

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



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

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



Basic ellipsoid algorithm

given ellipsoid $\mathcal{E}(x, A)$ containing $x^\star$
repeat
1. evaluate $\nabla f(x)$
2. if $\sqrt{\nabla f(x)^T A \nabla f(x)} \le \epsilon$, 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 coords so uncertainty ($\mathcal{E}$) is unit ball
- take gradient step with length $1/(n+1)$

properties:

- *not* a descent method
- like quasi-Newton method with fixed step length
- much slower convergence than BFGS, etc.
- but, extends to nondifferentiable f

Ellipsoid method for standard problem

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

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

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

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

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, $f_j(x^{(k)}) > 0$;

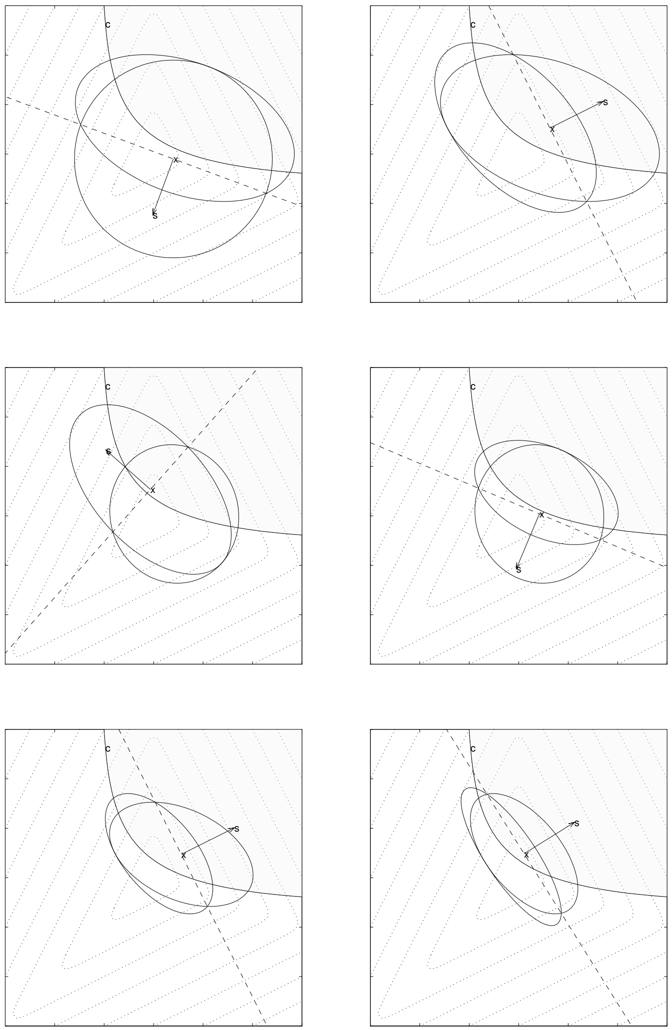
then

$\nabla f_j(x^{(k)})^T (x - x^{(k)}) \ge 0 \Longrightarrow f_j(x) > 0 \Longrightarrow x$ infeasible

so update $\mathcal{E}^{(k)}$ based on $\nabla f_j(x^{(k)})$

rules out halfspace of infeasible points

Example



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 $x \in \mathcal{E}^{(k)}$

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

hence, problem is infeasible if for some j,

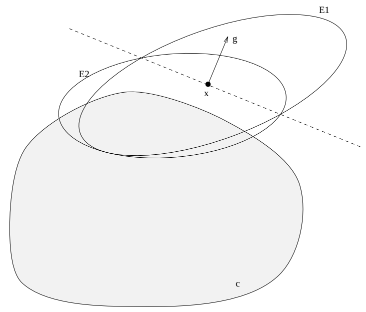
$$f_j(x^{(k)}) - \sqrt{\nabla f_j(x^{(k)})^T A^{(k)} \nabla f_j(x^{(k)})} > 0$$

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

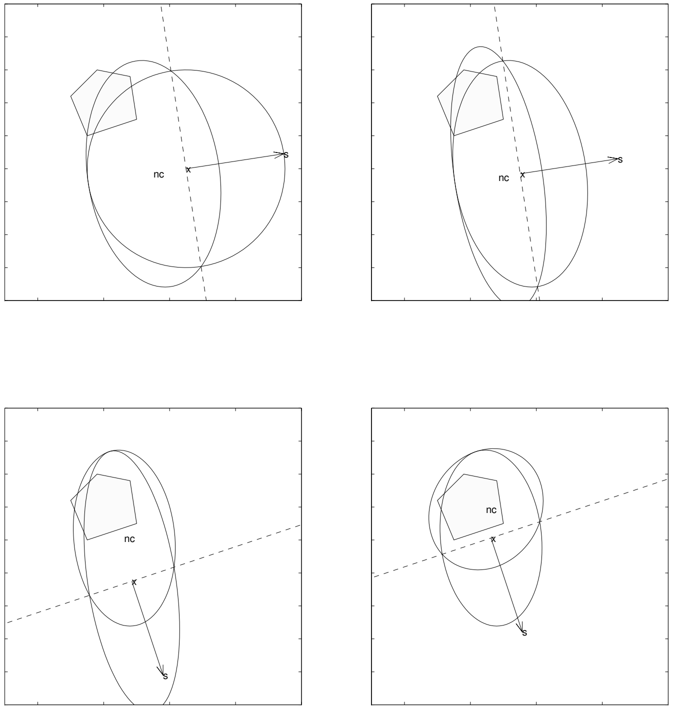
- confirms $x \in C$, or
- returns $g \neq 0$ s.t. $z \in C \Rightarrow g^T(z - x) \leq 0$



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



Ellipsoid algorithm for nondifferentiable probs

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

f_i convex, $\textbf{dom}\, f_i = \mathbf{R}^n$

idea:

- work with subgradients instead of gradients
- if constraints are satisfied at current point, use objective subgradient
- if not, use constraint information

Algorithm

given ellipsoid $\mathcal{E}$ with $x^\star \in \mathcal{E}$ if feasible, $\epsilon > 0$
repeat
  if for some j, $f_j(x) > 0$ (*i.e.*, x infeasible)
    find $g \in \partial f_j(x)$;
    if $f_j(x) + \inf_{z \in \mathcal{E}} g^T(z - x) > 0$, return(infeas);
  else (*i.e.*, x feasible)
    $u := f_0(x)$;
    find $g \in \partial f_0(x)$;
    $l := f_0(x) + \inf_{z \in \mathcal{E}} g^T(z - x)$;
    if $u - l < \epsilon$ return(x);
  $\mathcal{E}$:= min. volume ellipsoid containing
  $\mathcal{E} \cap \{z \mid g^T(z - x) \leq 0\}$;
  x := center($\mathcal{E}$);