Cutting-plane methods

- cutting planes
- localization methods

Cutting-plane oracle

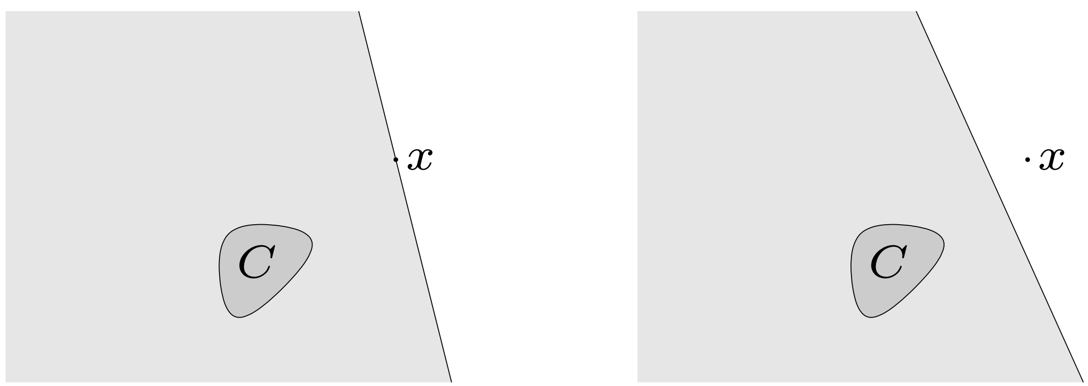
provides a black-box description of a convex set C

- when queried at x, oracle either asserts $x \in C$ or returns $a \neq 0$, b with

$$a^T x \geq b, \qquad a^T z \leq b \quad \forall z \in C$$

  $a^T z = b$ defines a **cutting plane**, separating x and C

- cut is **neutral** if $a^T x = b$: query point is on boundary of halfspace
- cut is **deep** if $a^T x > b$: query point in interior of halfspace that is cut



Localization method

goal: find a point in convex set C described by cutting-plane oracle

algorithm: choose bounded set $\mathcal{P}_0$ containing C; repeat for $k \geq 1$:

- choose a point $x^{(k)}$ in $\mathcal{P}_{k-1}$ and query the cutting-plane oracle at $x^{(k)}$
- if $x^{(k)} \in C$, return $x^{(k)}$; else, add cutting plane $a_k^T z \leq b_k$ to $\mathcal{P}_{k-1}$:

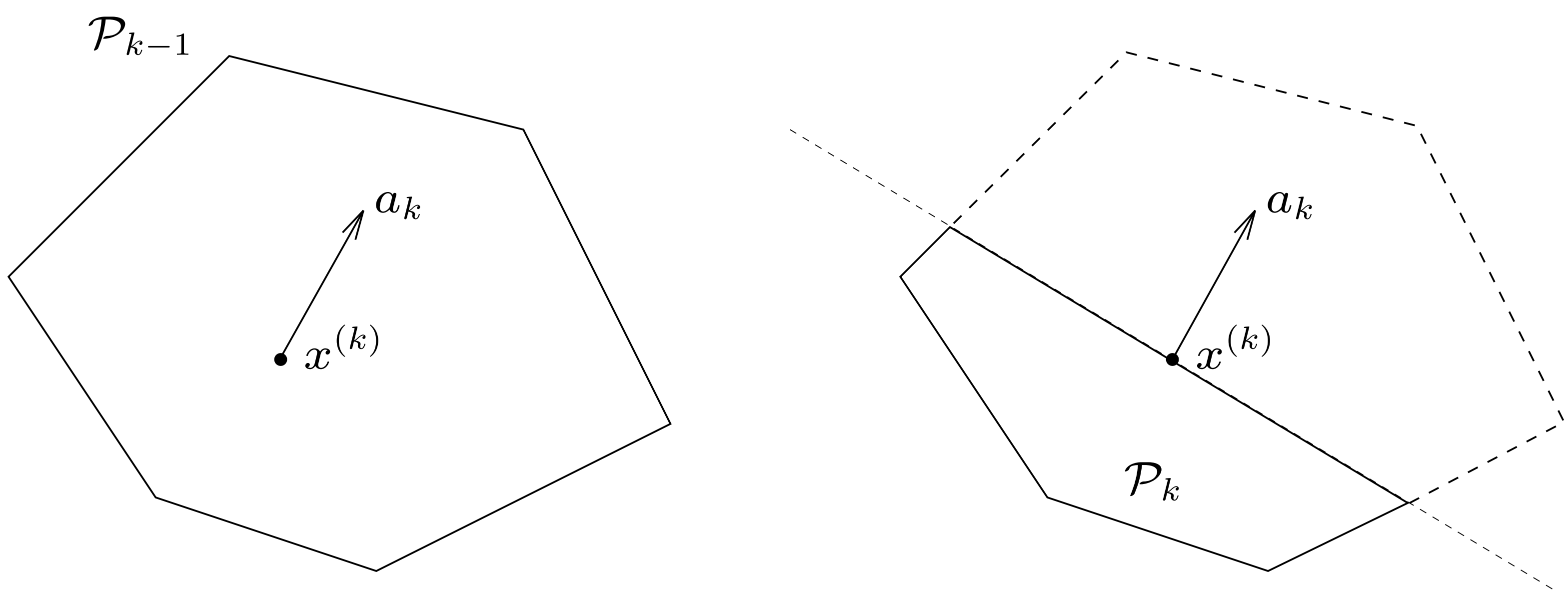
$$\mathcal{P}_k = \mathcal{P}_{k-1} \cap \{z \mid a_k^T z \leq b_k\}$$

  terminate if $\mathcal{P}_k = \emptyset$

variation: to keep $\mathcal{P}_k$ simple, choose $\mathcal{P}_k \supseteq \mathcal{P}_{k-1} \cap \{z \mid a_k^T z \leq b_k\}$

we'll discuss specific algorithms later

geometry



$\mathcal{P}_k$ gives uncertainty of C after iteration k

Unconstrained minimization

C is optimal set for convex f

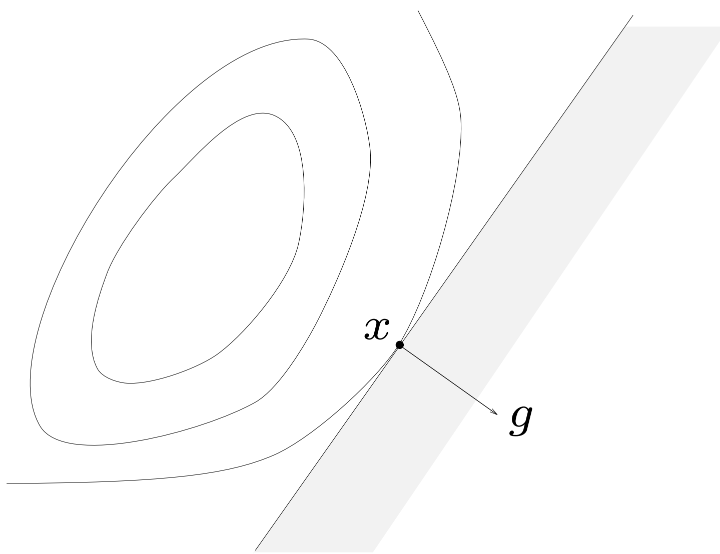
neutral cut: if $f(x) > f^\star$ and $g \in \partial f(x)$, then a neutral cut at x is

$$g^T z \leq g^T x$$

proof: $g^T z > g^T x$ implies $f(z) \geq f(x) + g^T(z - x) > f(x) > f^\star$

interpretation: by evaluating $g \in \partial f(x)$

- we rule out halfspace in search for $x \in C$
- we get one 'bit' of info on C



Deep cut for unconstrained minimization

suppose we know a number $\bar{f}$ with

$$f(x) > \bar{f} \geq f^\star$$

for example, $\bar{f}$ is the smallest value of f found so far in an algorithm

deep cut: if $f(x) > f^\star$ and $g \in \partial f(x)$, then a deep cut at x is given by

$$g^T z \leq g^T x - f(x) + \bar{f}$$

proof: $g^T z > g^T x - f(x) + \bar{f}$ implies

$$f(z) \geq f(x) + g^T (z - x) > \bar{f} \geq f^\star$$

Feasibility problem

C is solution set of convex inequalities

$$f_i(x) \leq 0, \quad i = 1, \ldots, m$$

deep cut: if $x \notin C$, find j with $f_j(x) > 0$ and evaluate $g_j \in \partial f_j(x)$;

$$g_j^T z \leq g_j^T x - f_j(x)$$

is a deep cut at x

proof: $g_j^T z > g_j^T x - f_j(x)$ implies $z \notin C$ because

$$f_j(z) \geq f_j(x) + g_j^T (z - x) > 0$$

Inequality constrained problem

C is optimal set of convex problem

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

feasibility cut: if x is not feasible, say $f_j(x) > 0$, we have a deep cut

$$g_j^T z \leq g_j^T x - f_j(x) \quad \text{where } g_j \in \partial f_j(x)$$

objective cut: if x is feasible, but $f_0(x) > p^\star + \epsilon$, we have a neutral cut

$$g_0^T z \leq g_0^T x \quad \text{where } g_0 \in \partial f_0(x)$$

moreover, if $\bar{f}$ with $f_0(x) > \bar{f} \geq p^\star$ is known, we have a deep cut

$$g_0^T z \leq g_0^T x - f_0(x) + \bar{f}$$

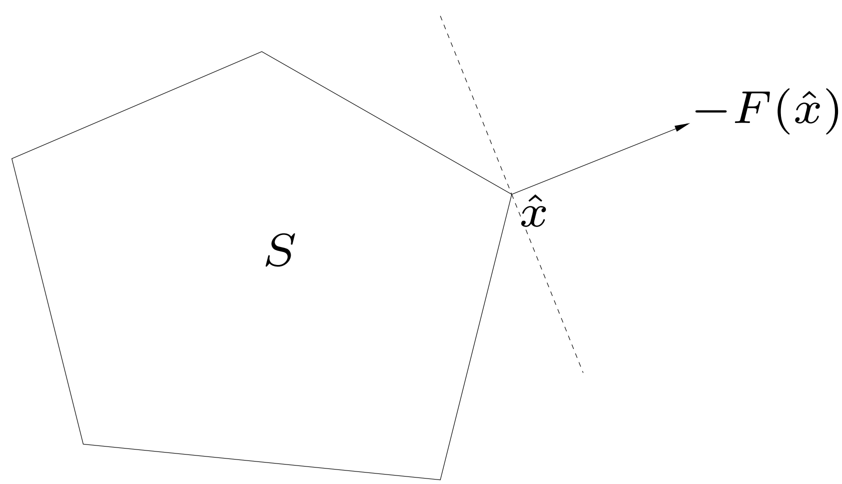
Variational inequality

monotone mapping: a mapping $F : \mathbf{R}^n \to \mathbf{R}^n$ is monotone if

$$(F(x) - F(y))^T (x - y) \geq 0 \quad \forall x, y$$

monotone variational inequality: given closed convex S, find $\hat{x} \in S$ with

$$F(\hat{x})^T (x - \hat{x}) \geq 0 \quad \forall x \in S$$



equivalently, $\hat{x} = P_S(\hat{x} - F(\hat{x}))$ where P_S is projection on S

Convex optimization problem as variational inequality

variational inequality with $F(x) = \nabla f(x)$:

$$\hat{x} \in S, \qquad \nabla f(\hat{x})^T (x - \hat{x}) \geq 0 \quad \forall x \in S$$

- F is monotone if f is convex (see p. 1-9)
- variational inequality is optimality condition for convex problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in S \end{array}$$

(see EE236B page 4–9)

note: in the general variational inequality, F is not necessarily a gradient

Saddle-point problem

suppose $f(u, v)$ is convex in u, concave in v, and U, V are convex sets

saddle point: $(\hat{u}, \hat{v}) \in U \times V$ is a saddle point if

$$f(\hat{u}, v) \leq f(\hat{u}, \hat{v}) \leq f(u, \hat{v}) \qquad \forall u \in U, v \in V$$

variational inequality formulation (for differentiable f):

$$\left[\begin{array}{c} \nabla_u f(\hat{u}, \hat{v}) \\ -\nabla_v f(\hat{u}, \hat{v}) \end{array} \right]^T \left[\begin{array}{c} u - \hat{u} \\ v - \hat{v} \end{array} \right] \geq 0 \qquad \forall (u, v) \in U \times V$$

- $\hat{u}$ minimizes $f(u, \hat{v})$ over $u \in U$; $\hat{v}$ minimizes $-f(\hat{u}, v)$ over $v \in V$
- a variational inequality with $F(u, v) = (\nabla f_u(u, v), -\nabla f_v(u, v))$

monotonicity of $F(u,v) = (\nabla f_u(u,v), -\nabla f_v(u,v))$

if f is convex-concave, then for all $w = (u,v)$, $\hat{w} = (\hat{u},\hat{v})$

$$
\begin{array}{l}
(F(w) - F(\hat{w}))^T(w - \hat{w}) \\
\quad \begin{array}{rl}
= & (\nabla_u f(w) - \nabla_u f(\hat{w}))^T(u - \hat{u}) - (\nabla_v f(w) - \nabla_v f(\hat{w}))^T(v - \hat{v}) \\
\geq & -f(\hat{u},v) + f(u,v) - f(u,\hat{v}) + f(\hat{u},\hat{v}) + f(u,\hat{v}) - f(u,v) \\
& \quad + f(\hat{u},v) - f(\hat{u},\hat{v}) \\
= & 0
\end{array}
\end{array}
$$

Cutting planes for variational inequality

to generate cutting plane at x

- if $x \notin S$, use feasibility cut (cutting plane that separates x from S)
- if $x \in S$ and not a solution, use the cutting plane

$$F(x)^T z \leq F(x)^T x$$

proof: if $F(x)^T z > F(x)^T x$ then, by monotonicity,

$$F(z)^T (z - x) \geq F(x)^T (z - x) > 0$$

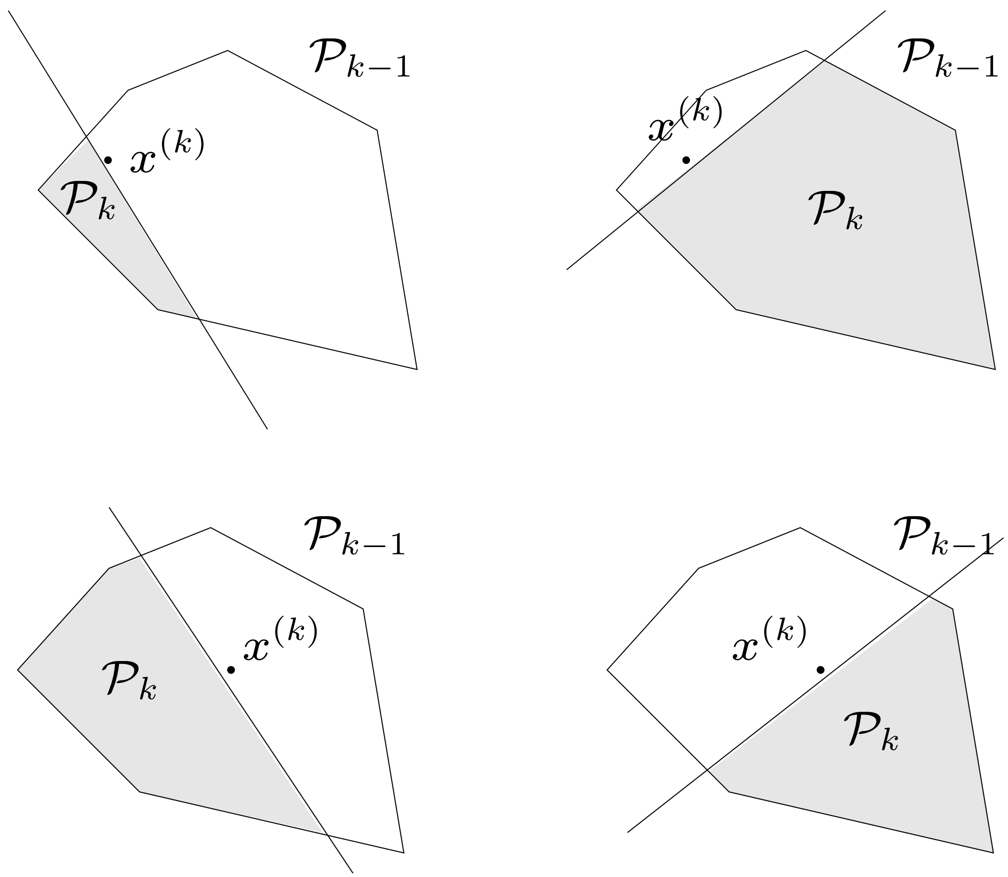
therefore z is not a solution of the variational inequality

Outline

- cutting planes
- **cutting-plane methods**

Choice of query point

should be near center of $\mathcal{P}_{k-1}$

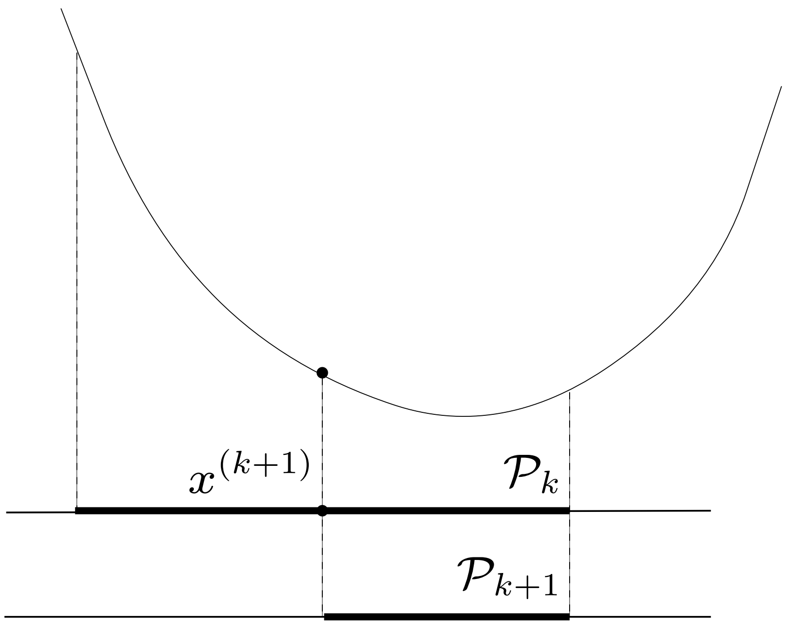


want to pick $x^{(k)}$ so that $\mathcal{P}_k$ is as small as possible, for any cut

Example: bisection in R

for minimizing differentiable convex $f : \mathbf{R} \to \mathbf{R}$

given: interval $\mathcal{P}_0 = [l, u]$ containing $x^\star$
repeat:
  $x := (l + u)/2$;
  if $f'(x) < 0$, $l := x$; else $u := x$



iteration complexity

$$\text{length}(\mathcal{P}_k) = \frac{\text{length}(\mathcal{P}_{k-1})}{2} = \frac{\text{length}(\mathcal{P}_0)}{2^k}$$

- $\text{length}(\mathcal{P}_k)$ measures uncertainty in $x^\star$
- uncertainty is halved at each iteration (exactly one bit of info)

#steps required to reduce uncertainty (in $x^\star$) to below r:

$$k = \log_2 \frac{\text{length}(\mathcal{P}_0)}{r} = \log_2 \frac{\text{initial uncertainty}}{\text{final uncertainty}}$$

Specific cutting-plane algorithms

methods vary in choice of query point

center of gravity (CG) algorithm

$x^{(k)}$ is center of gravity of $\mathcal{P}_{k-1}$

maximum volume ellipsoid (MVE) cutting-plane method

$x^{(k)}$ is center of maximum volume ellipsoid contained in $\mathcal{P}_{k-1}$

Chebyshev center cutting-plane method

$x^{(k)}$ is center of largest ball contained in $\mathcal{P}_{k-1}$

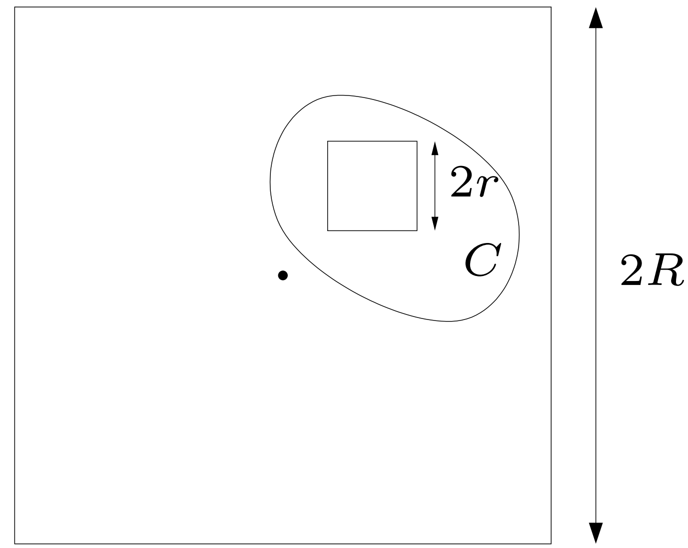
analytic center cutting-plane method (ACCPM) (next lecture)

$x^{(k)}$ is analytic center of (inequalities defining) $\mathcal{P}_{k-1}$

Lower bound on complexity

problem class: find $x \in C \subseteq \mathbf{R}^n$

- C is convex
- C is contained in $\{x \mid \|x\|_\infty \leq R\}$
- C contains an ℓ_∞-norm ball of radius r
- C is described by a cutting-plane oracle



bound on complexity

no localization algorithm can guarantee a complexity lower than

$$n \log_2 \left(\frac{R}{2r} \right)$$

iterations (queries to oracle)

proof: suppose we run a localization algorithm for

$$k < n \log_2(R/(2r)) \text{ iterations}$$

we will construct a 'resisting oracle' for a hyperrectangle

$$C = \{x \mid c - d \preceq x \preceq c + d\}$$

that does not contain any of the k query points and satisfies

$$\max_i (|c_i| + d_i) \leq R, \qquad \min_i d_i \geq \frac{R}{2^{\lceil k/n \rceil}} \geq r$$

therefore, the algorithm failed to find a point in C in k steps even though

$$\{x \mid \|x - c\|_\infty \leq r\} \subseteq C \subseteq \{x \mid \|x\|_\infty \leq R\}$$

the oracle and c, d are constructed as follows: initially, $c = 0$, $d = R$

at iteration j,

- define $i = j - n\lfloor (j-1)/n \rfloor$ (*i.e.*, cycle through the n coordinates)
- if x is the query point at iteration j, then
  - if $x_i \geq c_i$, update c, d as

$$c_i := c_i - d_i/2, \qquad d_i := d_i/2$$

    and return the cut $e_i^T(z - x) \leq 0$

  - if $x_i < c_i$, update c, d as

$$c_i := c_i + d_i/2, \qquad d_i := d_i/2$$

    and return the cut $-e_i^T(z - x) \leq 0$

Center of gravity algorithm

choose as $x^{(k)}$ the center of gravity of $\mathcal{P}_{k-1}$ (denoted $\mathrm{CG}(\mathcal{P}_{k-1})$)

$$x^{(k)} = \mathrm{CG}(\mathcal{P}_{k-1}) = \frac{\displaystyle\int_{\mathcal{P}_{k-1}} x \, dx}{\displaystyle\int_{\mathcal{P}_{k-1}} dx}$$

theorem: if $S \subseteq \mathbf{R}^n$ convex, $x_{\mathrm{cg}} = \mathrm{CG}(S)$, $g \neq 0$,

$$\mathbf{vol}\left(S \cap \left\{x \mid g^T(x - x_{\mathsf{cg}}) \leq 0\right\}\right) \leq (1 - 1/e)\,\mathbf{vol}(S)$$

(independent of dimension n)

Convergence of CG cutting-plane method

assumptions

- $\mathcal{P}_0 \subseteq \{x \mid \|x\|_\infty \leq R\}$
- C contains an ℓ_∞-ball of radius r

iteration complexity

if $x^{(1)}, \ldots, x^{(k)} \notin C$, then $C \subseteq \mathcal{P}_k$ (no part of C is cut) and

$$(2r)^n \leq \mathbf{vol}(\mathcal{P}_k) \leq \left(1 - \frac{1}{e}\right)^k \mathbf{vol}(\mathcal{P}_0) \leq \left(1 - \frac{1}{e}\right)^k (2R)^n$$

therefore

$$k \leq \frac{n \log(R/r)}{-\log(1 - 1/e)} = 1.51\, n \log_2 \left(\frac{R}{r}\right)$$

advantages of CG-method

- guaranteed convergence
- affine-invariance
- iteration complexity is near optimal (see page 18)

disadvantage

finding $x^{(k)} = \mathrm{CG}(\mathcal{P}_{k-1})$ is **much harder** than original problem

(but, can modify CG-method to work with approximate CG computation)

Maximum volume ellipsoid method

$x^{(k)}$ is center of maximum volume ellipsoid in $\mathcal{P}_{k-1}$

- can be computed via convex optimization
- affine-invariant

complexity

- can show $\mathbf{vol}(\mathcal{P}_{k+1}) \leq (1 - 1/n)\,\mathbf{vol}(\mathcal{P}_k)$
- hence can bound number of steps:

$$k \leq \frac{n \log(R/r)}{-\log(1 - 1/n)} \approx n^2 \log(R/r)$$

if cutting-plane oracle cost is not small, MVE is a good practical method

Extensions

multiple cuts

- oracle returns set of linear inequalities instead of just one, $e.g.$,
  - all violated inequalities
  - all inequalities (including *shallow cuts*)
  - multiple deep cuts
- at each iteration, append (set of) new inequalities to those defining $\mathcal{P}_k$

nonlinear cuts

- use nonlinear convex inequalities instead of linear ones
- localization set no longer a polyhedron
- some methods ($e.g.$, ACCPM) still work

Dropping constraints

the problem

- number of linear inequalities defining $\mathcal{P}_k$ increases at each iteration
- hence, computational effort to compute $x^{(k+1)}$ increases

solutions

- drop redundant constraints
- keep only a fixed number of (the most relevant) constraints
- at each iteration, replace localization set by upper bound

first two solutions discussed in lecture ; third solution in lecture

Epigraph cutting-plane method

cutting-plane method applied to epigraph form problem

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

cutting-plane oracle (queried at x)

- if x is infeasible for original problem (say, $f_j(x) > 0$), add cutting-plane

$$\left[\begin{array}{c} g_j \\ 0 \end{array}\right]^T \left[\begin{array}{c} z \\ t \end{array}\right] \leq g_j^T x - f_j(x) \quad (g_j \in \partial f_j(x))$$

- if x is feasible for original problem, add *two* cutting-planes

$$\left[\begin{array}{c} 0 \\ 1 \end{array}\right]^T \left[\begin{array}{c} z \\ t \end{array}\right] \leq f_0(x), \qquad \left[\begin{array}{c} g_0 \\ -1 \end{array}\right]^T \left[\begin{array}{c} z \\ t \end{array}\right] \leq g_0^T x - f_0(x) \quad (g_0 \in \partial f_0(x))$$

References

- Yu. Nesterov, *Introductory Lectures on Convex Optimization. A Basic Course* (2004) (§3.2.5 and §3.2.6)

- S. Boyd, course notes for EE364b, Convex Optimization II