Localization and Cutting-Plane Methods

- cutting-plane oracle
- finding cutting-planes
- localization algorithms
- specific cutting-plane methods
- epigraph cutting-plane method
- lower bounds and stopping criteria

Localization and cutting-plane methods

- based on idea of 'localizing' desired point in some set, which becomes smaller at each step
- like subgradient methods, require computation of a subgradient of objective or constraint functions at each step
- in particular, directly handle nondifferentiable convex (and quasiconvex) problems
- typically require more memory and computation per step than subgradient methods
- but can be much more efficient (in theory and practice) than subgradient methods

Cutting-plane oracle

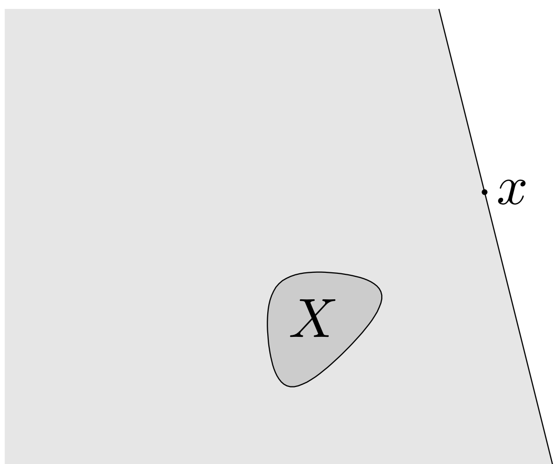
- goal: find a point in convex set $X \subseteq \mathbf{R}^n$, or determine that $X = \emptyset$
- our only access to or description of X is through a *cutting-plane oracle*
- when cutting-plane oracle is *queried* at $x \in \mathbf{R}^n$, it either
  - asserts that $x \in X$, or
  - returns a separating hyperplane between x and X: $a \neq 0$,

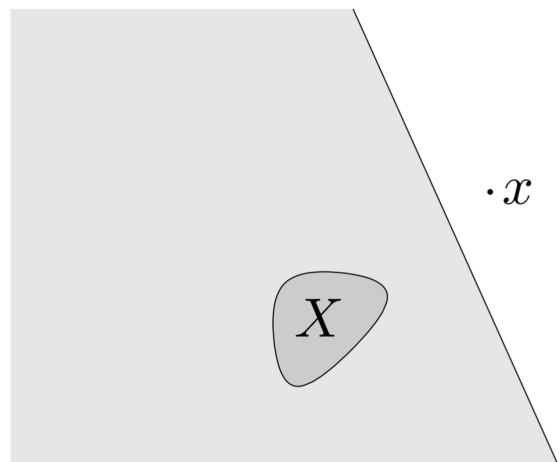
$$a^T z \leq b \text{ for } z \in X, \qquad a^T x \geq b$$

- (a, b) called a *cutting-plane*, or *cut*, since it eliminates the halfspace $\{z \mid a^T z > b\}$ from our search for a point in X

Neutral and deep cuts

- if $a^T x = b$ (x is on boundary of halfspace that is cut) cutting-plane is called *neutral cut*
- if $a^T x > b$ (x lies in interior of halfspace that is cut), cutting-plane is called *deep cut*





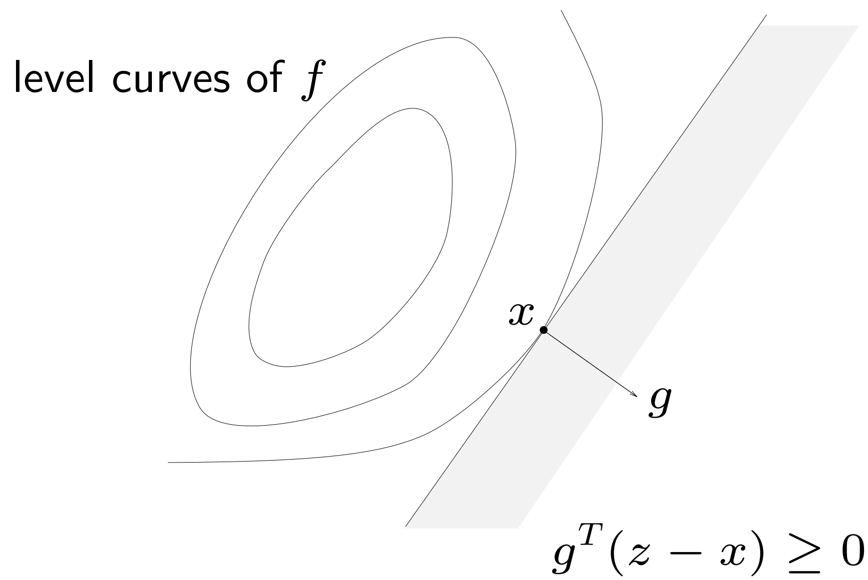
Unconstrained minimization

- minimize convex $f : \mathbf{R}^n \to \mathbf{R}$
- X is set of optimal points (minimizers)
- given x, find $g \in \partial f(x)$
- from $f(z) \geq f(x) + g^T(z - x)$ we conclude

$$g^T(z - x) > 0 \quad \Longrightarrow \quad f(z) > f(x)$$

  i.e., all points in halfspace $g^T(z - x) \geq 0$ are **worse** than x, and in particular not optimal

- so $g^T(z - x) \leq 0$ is (neutral) cutting-plane at x ($a = g$, $b = g^T x$)



- by evaluating $g \in \partial f(x)$ we rule out a halfspace in our search for $x^\star$
- **idea:** get one bit of info (on location of $x^\star$) by evaluating g

Deep cut for unconstrained minimization

- suppose we know a number $\bar{f}$ with $f(x) > \bar{f} \geq f^\star$ ($e.g.$, the smallest value of f found so far in an algorithm)
- from $f(z) \geq f(x) + g^T(z-x)$, we have

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

so we have deep cut

$$g^T(z-x) + f(x) - \bar{f} \leq 0$$

Feasibility problem

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

$f_1, \ldots, f_m$ convex; X is set of feasible points

- if x not feasible, find j with $f_j(x) > 0$, and evaluate $g_j \in \partial f_j(x)$
- since $f_j(z) \geq f_j(x) + g_j^T(z - x)$,

$$f_j(x) + g_j^T(z - x) > 0 \quad \Longrightarrow \quad f_j(z) > 0 \quad \Longrightarrow \quad z \notin X$$

  i.e., any feasible z satisfies the inequality $f_j(x) + g_j^T(z - x) \leq 0$

- this gives a deep cut

Inequality constrained problem

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

$f_0, \ldots, f_m : \mathbf{R}^n \to \mathbf{R}$ convex; X is set of optimal points; $p^\star$ is optimal value

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

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

- if x is feasible, we have (neutral) *objective cut*

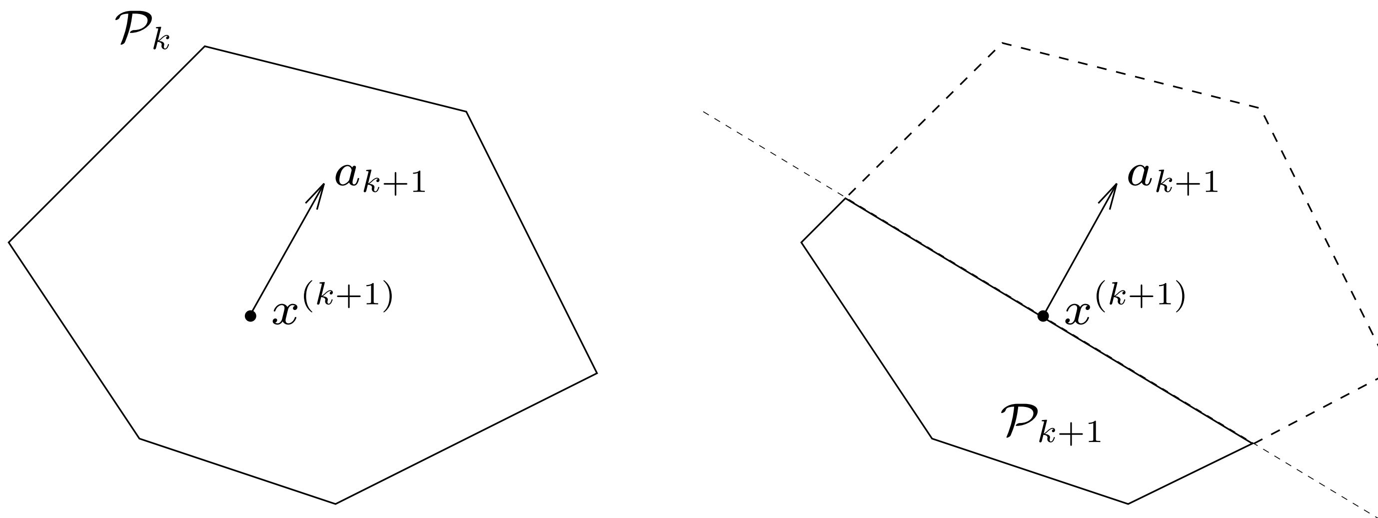
$$
g_0^T(z - x) \leq 0, \qquad g_0 \in \partial f_0(x)
$$

(or, deep cut $g_0^T(z - x) + f_0(x) - \bar{f} \leq 0$ if $\bar{f} \in [p^\star, f_0(x))$ is known)

Localization algorithm

basic (conceptual) localization (or cutting-plane) algorithm:

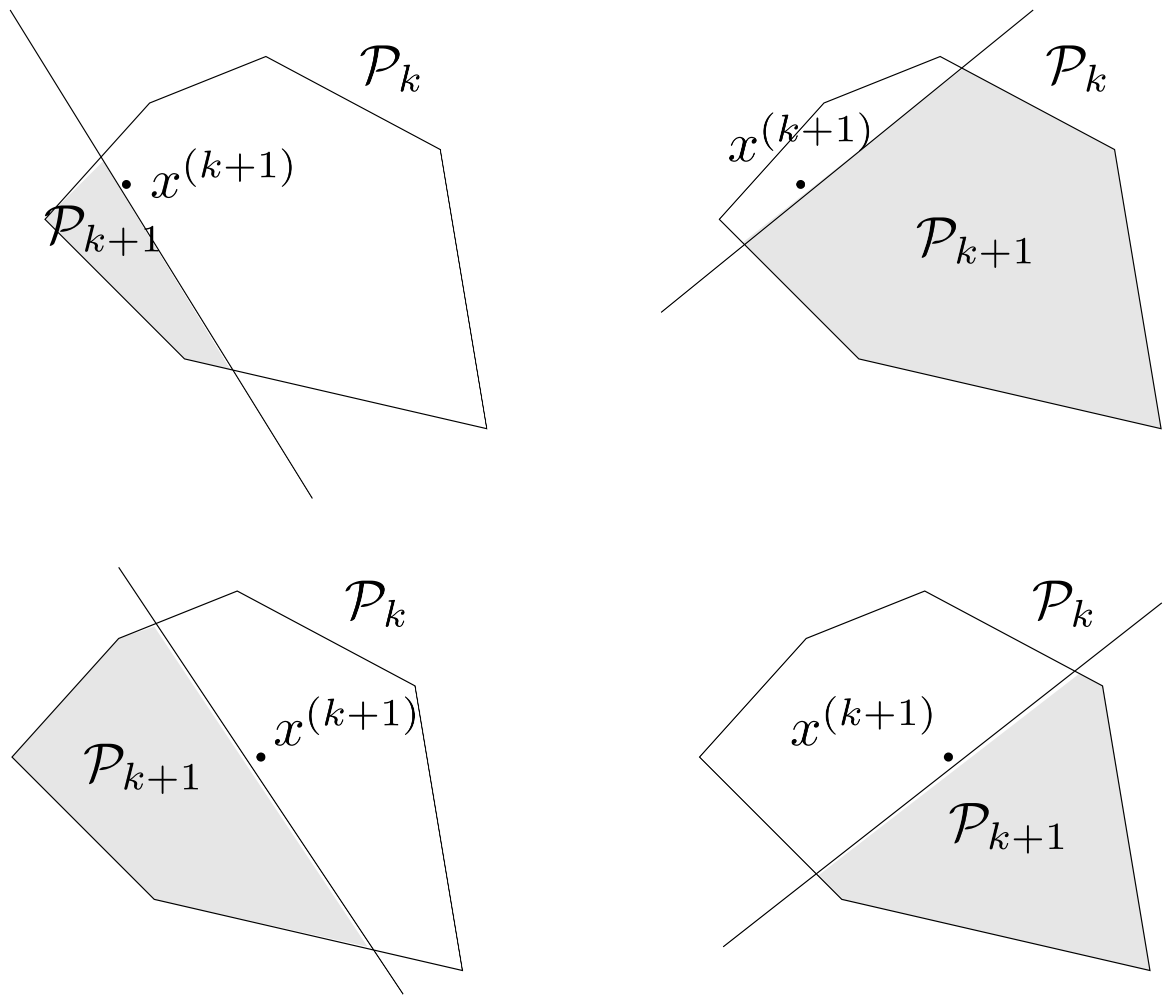
given initial polyhedron $\mathcal{P}_0 = \{z \mid Cz \preceq d\}$ known to contain X
$k := 0$
repeat
  Choose a point $x^{(k+1)}$ in $\mathcal{P}_k$
  Query the cutting-plane oracle at $x^{(k+1)}$
  If $x^{(k+1)} \in X$, quit
  Else, add new cutting-plane $a_{k+1}^T z \leq b_{k+1}$:
    $\mathcal{P}_{k+1} := \mathcal{P}_k \cap \{z \mid a_{k+1}^T z \leq b_{k+1}\}$
  If $\mathcal{P}_{k+1} = \emptyset$, quit
  $k := k + 1$



- $\mathcal{P}_k$ gives our uncertainty of $x^\star$ at iteration k

- want to pick $x^{(k+1)}$ so that $\mathcal{P}_{k+1}$ is as small as possible, no matter what cut is made

- want $x^{(k+1)}$ near center of $\mathcal{P}^{(k)}$


$\mathcal{P}_k$
$x^{(k+1)}$
$\mathcal{P}_{k+1}$
$\mathcal{P}_k$
$x^{(k+1)}$
$\mathcal{P}_{k+1}$
$\mathcal{P}_k$
$x^{(k+1)}$
$\mathcal{P}_{k+1}$
$\mathcal{P}_k$
$x^{(k+1)}$
$\mathcal{P}_{k+1}$

Example: Bisection on R

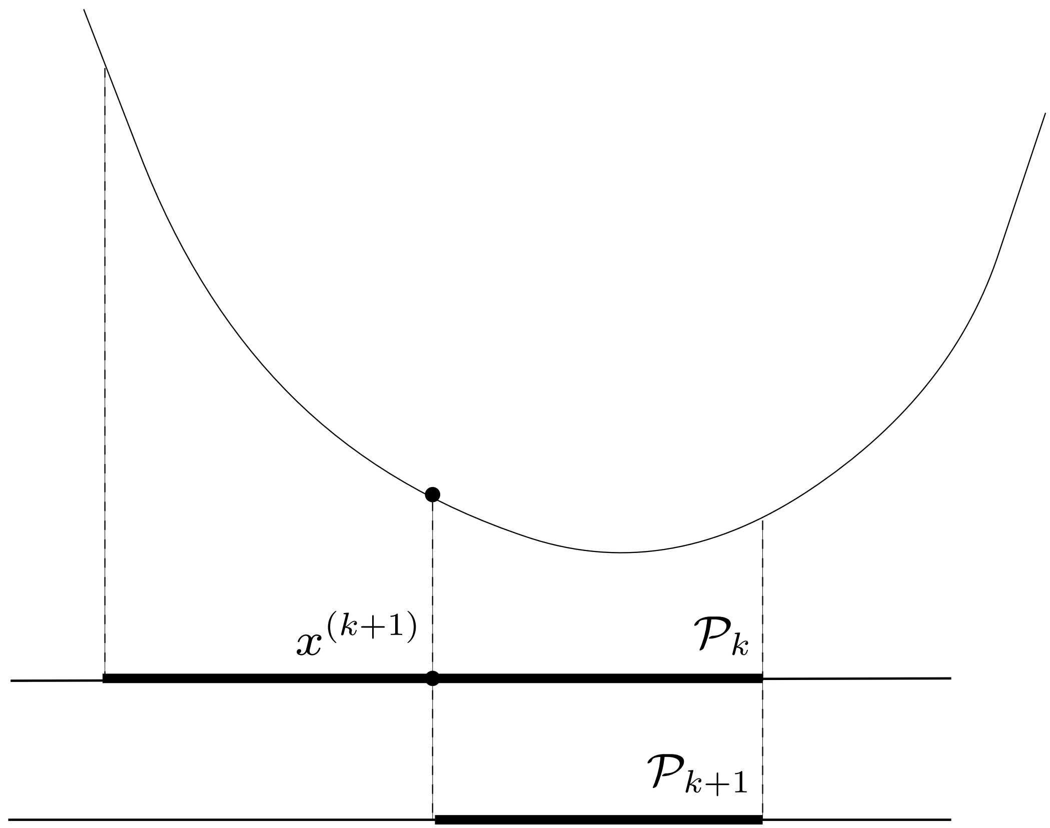
- minimize convex $f : \mathbf{R} \to \mathbf{R}$
- $\mathcal{P}_k$ is interval
- obvious choice for query point: $x^{(k+1)} :=$ midpoint$(\mathcal{P}_k)$

bisection algorithm

given interval $\mathcal{P}_0 = [l, u]$ containing $x^\star$

repeat

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


$x^{(k+1)}$
$\mathcal{P}_k$
$\mathcal{P}_{k+1}$

$$\text{length}(\mathcal{P}_{k+1}) = u_{k+1} - l_{k+1} = \frac{u_k - l_k}{2} = (1/2)\text{length}(\mathcal{P}_k)$$

and so $\text{length}(\mathcal{P}_k) = 2^{-k}\text{length}(\mathcal{P}_0)$

interpretation:

- $\text{length}(\mathcal{P}_k)$ measures our uncertainty in $x^\star$
- uncertainty is halved at each iteration; get exactly one bit of info about $x^\star$ per iteration
- $\#$ steps required for uncertainty (in $x^\star$) $\leq r$:

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

Specific cutting-plane methods

methods vary in choice of query point

- *center of gravity (CG) algorithm*:
  $x^{(k+1)}$ is center of gravity of $\mathcal{P}_k$
- *maximum volume ellipsoid (MVE) cutting-plane method*:
  $x^{(k+1)}$ is center of maximum volume ellipsoid contained in $\mathcal{P}_k$
- *Chebyshev center cutting-plane method*:
  $x^{(k+1)}$ is Chebyshev center of $\mathcal{P}_k$
- *analytic center cutting-plane method (ACCPM)*:
  $x^{(k+1)}$ is analytic center of (inequalities defining) $\mathcal{P}_k$

Center of gravity algorithm

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

$$\mathsf{CG}(\mathcal{P}_k) = \int_{\mathcal{P}_k} x \, dx \Bigg/ \int_{\mathcal{P}_k} dx$$

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

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

(independent of dimension n)

hence in CG algorithm, $\mathbf{vol}(\mathcal{P}_k) \leq 0.63^k\,\mathbf{vol}(\mathcal{P}_0)$

Convergence of CG cutting-plane method

- suppose $\mathcal{P}_0$ lies in ball of radius R, X includes ball of radius r (can take X as set of ϵ-suboptimal points)

- suppose $x^{(1)}, \dots, x^{(k)} \notin X$, so $\mathcal{P}_k \supseteq X$

- we have

$$\alpha_n r^n \leq \mathbf{vol}(\mathcal{P}_k) \leq (0.63)^k \, \mathbf{vol}(\mathcal{P}_0) \leq (0.63)^k \alpha_n R^n$$

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

- so $k \leq 1.51 n \log_2(R/r)$ (cf. bisection on $\mathbf{R}$)

advantages of CG-method

- guaranteed convergence
- affine-invariance
- number of steps proportional to dimension n, log of uncertainty reduction

disadvantages

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

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

Maximum volume ellipsoid method

- $x^{(k+1)}$ is center of maximum volume ellipsoid in $\mathcal{P}_k$
  (can compute as convex problem)

- affine-invariant

- 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

Chebyshev center method

- $x^{(k+1)}$ is center of largest Euclidean ball in $\mathcal{P}_k$
  (can compute via LP)

- not affine invariant; sensitive to scaling

Analytic center cutting-plane method

- $x^{(k+1)}$ is analytic center of $\mathcal{P}_k = \{z \mid a_i^T z \le b_i,\ i = 1, \ldots, q\}$

$$x^{(k+1)} = \operatorname*{argmin}_x - \sum_{i=1}^{q} \log(b_i - a_i^T x)$$

- $x^{(k+1)}$ can be computed using infeasible start Newton method

- works quite well in practice (more on this next lecture)

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
- the solution: drop or prune constraints
  - drop redundant constraints
  - keep only a fixed number N of (the most relevant) constraints (can cause localization polyhedron to increase!)

Epigraph cutting-plane method

apply cutting-plane method to epigraph form problem

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

with variables $x \in \mathbf{R}^n$ and t

at each (x, t), need cutting-plane oracle that separates (x, t) from $(x^\star, p^\star)$

- if $x^{(k)}$ is infeasible for original problem and violates jth constraint, add the cutting-plane

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

- if $x^{(k)}$ is feasible for original problem, add *two* cutting-planes

$$f_0(x^{(k)}) + g_0^T(x - x^{(k)}) \leq t, \qquad t \leq f_0(x^{(k)})$$

where $g_0 \in \partial f_0(x^{(k)})$

PWL lower bound on convex function

- suppose we have evaluated f and a subgradient of f at $x^{(1)}, \ldots, x^{(q)}$
- for all z,

$$f(z) \geq f(x^{(i)}) + g^{(i)T}(z - x^{(i)}), \quad i = 1, \ldots, q$$

  and so

$$f(z) \geq \hat{f}(z) = \max_{i=1,\ldots,q} \left(f(x^{(i)}) + g^{(i)T}(z - x^{(i)}) \right).$$

- $\hat{f}$ is a convex piecewise-linear global underestimator of f

Lower bound

- in solving convex problem

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

  we have evaluated some of the f_i and subgradients at $x^{(1)}, \ldots, x^{(k)}$

- form piecewise-linear approximations $\hat{f}_0, \ldots, \hat{f}_m$

- form PWL relaxed problem

$$\begin{array}{ll} \text{minimize} & \hat{f}_0(x) \\ \text{subject to} & \hat{f}_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Cx \preceq d \end{array}$$

(can be solved via LP)

- optimal value is a lower bound on $p^\star$