8. Localization and cutting-plane methods

- cutting-plane oracle
- finding cutting-planes
- localization algorithms
- specific cutting-plane methods
- epigraph cutting-plane method

Localization methods

- based on idea of 'localizing' desired point in some set, which becomes smaller at each step
- like subgradient methods, require one subgradient of objective or constraint functions at each step
- 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)

Cutting-plane oracle

goal: find a point in convex set $X \subseteq \mathbf{R}^n$, or determine that X is empty

cutting-plane oracle

- provides black-box description of X
- when 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: $a \neq 0$,

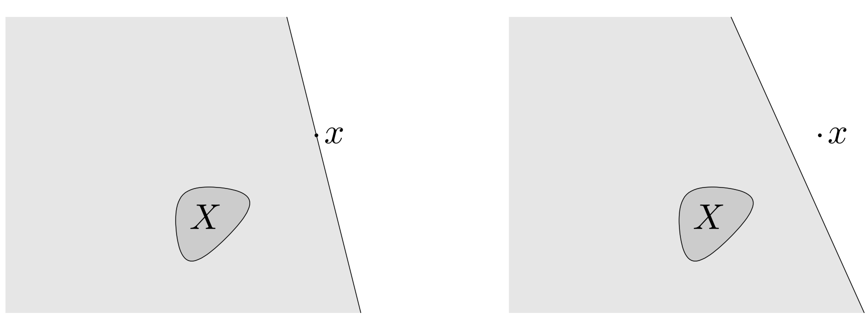
$$a^T z \leq b \text{ for } z \in X, \qquad a^T x \geq b$$

- (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

neutral cut: $a^T x = b$ (x is on boundary of halfspace that is cut)

deep cut: $a^T x > b$ (x lies in interior of halfspace that is cut)



Unconstrained minimization

take set of minimizers of f as X

cutting-plane oracle (for convex f): if $0 \neq g \in \partial f(x)$, then

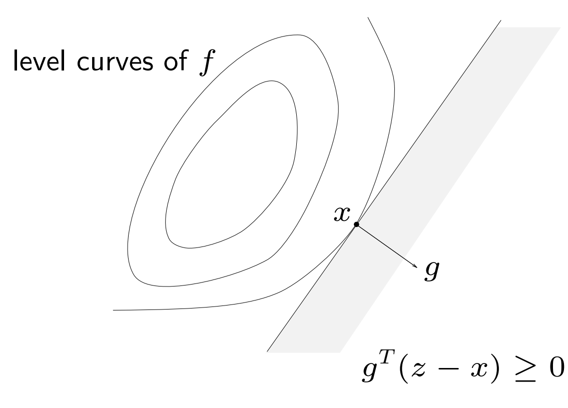
$$g^T(z - x) \leq 0$$

defines a (neutral) cut $(a, b) = (g, g^T x)$ at x

proof: if $g^T(z - x) > 0$, then $z \notin X$

$$f(z) \geq f(x) + g^T(z - x) > f(x)$$

interpretation



- by evaluating $g \in \partial f(x)$ we rule out halfspace in search for optimum
- 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)

deep cut: if $g \in \partial f(x)$, then a deep cut is given by

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

proof: if $f(x) + g^T(z - x) - \bar{f} > 0$, then $z \notin X$

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

Feasibility problem

X is solution set of convex inequalities

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

cutting-plane oracle

if x not feasible, find j with $f_j(x) > 0$, and evaluate $g_j \in \partial f_j(x)$;

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

is a deep cut

proof: if $f_j(x) + g_j^T(z - x) > 0$, then $z \notin X$

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

Inequality constrained problem

X is the set of optimal points of convex problem

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

cutting-plane oracle

- 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)

(Conceptual) localization algorithm

given initial polyhedron $\mathcal{P}_0 = \{z \mid Cz \preceq d\}$ known to contain X

repeat for $k = 1, 2, \ldots$:

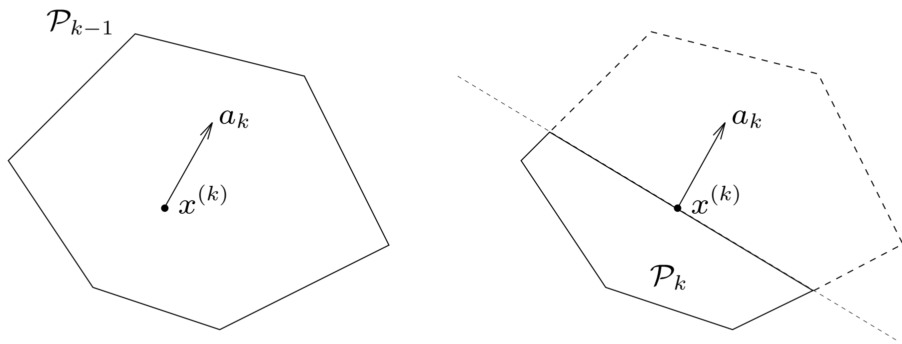
choose a point $x^{(k)}$ in $\mathcal{P}_{k-1}$ and query the cutting-plane oracle at $x^{(k)}$

- if $x^{(k)} \in X$, return $x^{(k)}$
- else, add new cutting-plane $a_k^T z \leq b_k$:

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

  if $\mathcal{P}_k = \emptyset$, quit

geometry



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

Lower bound on complexity

problem class

find $x \in X \subseteq \mathbf{R}^n$, where the following is known about X:

- X is convex
- X is contained in $\{x \mid \|x\|_\infty \leq R\}$
- X contains a ball $\{x \mid \|x - x_\mathrm{c}\|_\infty \leq r\}$
- a cutting-plane oracle for X

bound on complexity

no localization algorithm can guarantee a complexity lower than

$$n \log_2(\frac{R}{2r}) \text{ iterations}$$

proof

suppose we run a localization algorithm for $k < n\log_2(R/(2r))$ iterations

we define an oracle, and construct X as follows

- initialize a rectangle $B = [c-d, c+d]$ with $c = 0$, $d = R\mathbf{1}$
- at iteration j, if x is the query point:
  - update B as

$$d_i := d_i/2, \qquad c_i := \begin{cases} c_i - d_i/2 & \text{if } x_i \geq c_i \\ c_i + d_i/2 & \text{if } x_i < c_i \end{cases}$$

  - let the oracle return the cut

$$e_i^T(z-x) \leq 0 \quad (\text{if } x_i \geq c_i), \qquad -e_i^T(z-x) \leq 0 \quad (\text{if } x_i < c_i)$$

  where $i = j - n\lfloor (j-1)/n \rfloor$ (*i.e.*, cycle through the n coordinates)

after k iterations, take $X = B$

properties of $X = B$:

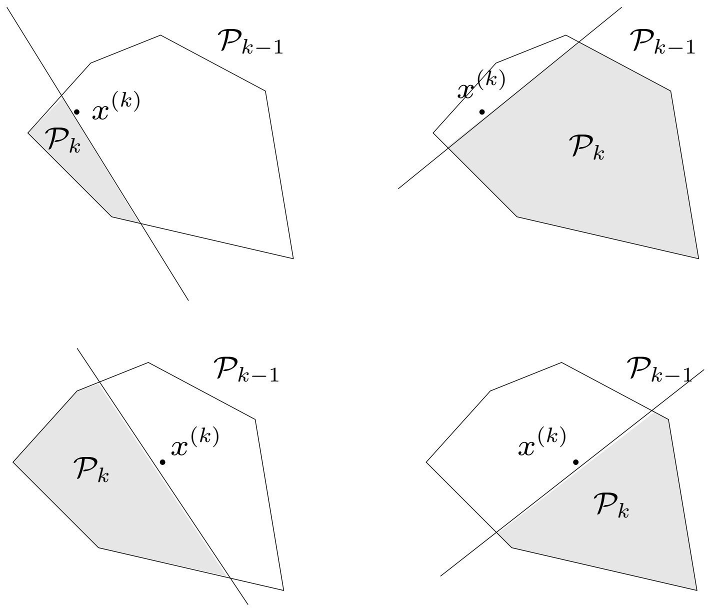
- X is convex and enclosed in $\{x \mid \|x\|_\infty \leq R\}$
- $d \succeq (R/2^{\lceil k/n \rceil})\mathbf{1}$, so X contains an ∞-norm ball with radius

$$R/2^{\lceil k/n \rceil} > r$$

- X does not contain any of the k query points

Choice of query point

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

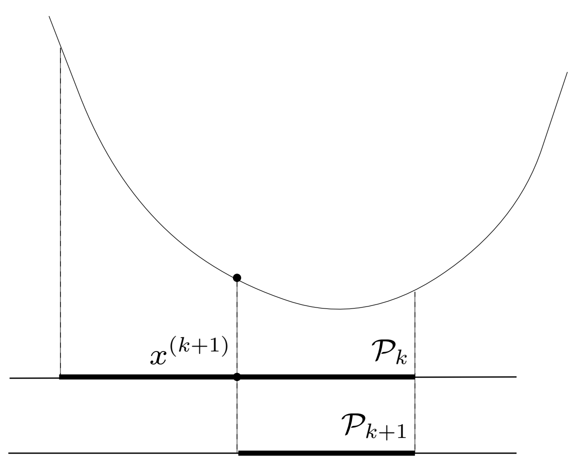


want to pick $x^{(k)}$ so that $\mathcal{P}_k$ is as small as possible, for any cut

Example: bisection on R

for minimizing 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

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

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

#steps required for uncertainty (in $x^\star$) $\leq r$:

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

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

Specific cutting-plane methods

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 Chebyshev center of $\mathcal{P}_{k-1}$

analytic center cutting-plane method (ACCPM)

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

Center of gravity algorithm

take $x^{(k)} = \mathrm{CG}(\mathcal{P}_{k-1})$ (center of gravity)

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

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

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

(independent of dimension n)

Convergence of CG cutting-plane method

assumptions

- $\mathcal{P}_0 \subseteq \{x \mid \|x\|_\infty \leq R\}$
- X contains a ball $\{x \mid \|x - x_{\mathrm{c}}\|_\infty \leq r\}$

iteration complexity

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

$$(2r)^n \leq \mathbf{vol}(\mathcal{P}_k) \leq (0.63)^k\,\mathbf{vol}(\mathcal{P}_0) \leq (0.63)^k (2R)^n$$

therefore

$$k \leq 2.18 n \log(R/r) = 1.51 n \log_2(R/r)$$

advantages of CG-method

- guaranteed convergence
- affine-invariance
- iteration complexity is essentially optimal (see page 8–12)

disadvantage

finding $x^{(k)} = \mathsf{CG}(\mathcal{P}_{k-1})$ 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 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

Chebyshev center method

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

- can be computed via linear programming
- not affine-invariant; sensitive to scaling

Analytic center cutting-plane method

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

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

- $x^{(k)}$ can be computed using Newton method
- works quite well in practice (more later)

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

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

- if $x^{(k)}$ is infeasible for original problem ($f_j(x) > 0$), add 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 $g_0 \in \partial f_0(x^{(k)})$

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!)

Piecewise linear lower bound on convex function

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

then for all z,

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

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

Piecewise linear relaxation

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 $x^{(1)}, \ldots, x^{(k)}$

PWL relaxation

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

$\hat{f}_0, \ldots, \hat{f}_m$ are PWL lower bounds of f_0,. . . , f_m (can be solved as LP)

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

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