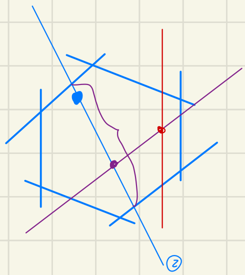
# Lec 6: Cutting plane method

Epigraph Cutting plane method

• A method to calculate center of gravity of a cvx set



① have a point,

② generate a random direction

( uniform distribution over a unit sphere
random gaussian vector, which is
circular symmetric, then normalize it )
$\mathcal{N}(0, I)$

③ generate a uniform distribution on the interval, back to ②

that's a markov chain propagating in the polyhedron } → uniform distribution over P
then averge these points to get CG

Polyhedron : $Ax \le b$
Starting point and a random direction: $x_0 + tv$ } max. / min. $t$
s.t. $A(x_0 + tv) \le b$

• Maximum volume ellipsoid method

$x^{(k+1)}$ is the center of maximum volume ellipsoid in $P^k$ (convex problem)
affine invariant
can show $\mathrm{vol}(P_{k+1}) \le (1 - 1/n)\,\mathrm{vol}(P_k)$
number of steps required : $k \le \dfrac{n\log(R/r)}{-\log(1-1/n)} \approx n^2 \log(R/r)$

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

• Chebyshev center method
not affine-invariant, sensitive to scaling

• Analytic center cutting-plane method

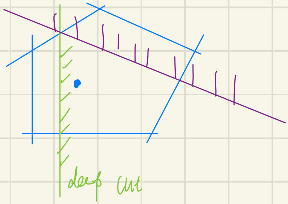
$x^{k+1}$ is the analytic center $P_k = \{ z \mid a_i^T z \le b_i \}$
$x^{(k+1)} = \operatorname{argmin} - \sum_i \log(b_i - a_i^T x)$
( analytic center of a set of inequalities ≠ analytic center of a polyhedron
redundant inequalities can shift the center in set )

- Extensions

multiple cuts:

oracle returns set of linear inequalities instead of one

(all violated inequalities / all inequalities including shallow cuts / multiple deep cuts)



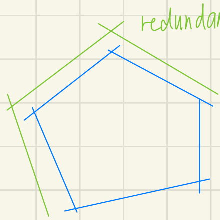

nonlinear cuts

use nonlinear convex inequalities instead of linear ones (eg. ellipsoid)

- Dropping constraints

number of linear inequalities defining $P_k$ increases at each program

(1) drop redundant inequalities




$\rightarrow \begin{bmatrix} a_1^T x \le b_1 \\ \vdots \\ a_m^T x \le b_m \end{bmatrix}$

to check whether 1st inequality

$\left.\begin{array}{ll} \max. & a_1^T x \\ \text{s.t.} & a_i^T x \le b_i \quad i = 2, 3 \cdots \end{array}\right\}$ LP

if $a_1^T x^* \le b_1$, the 1st inequality is redundant

(2). keep only a fixed number $N$ of (the most relevant) inequality constraints (typically $N \approx 5n$)

(can cause localization polyhedron to increase)

· Epigraph cutting-plane method

epigraph form: $\min.\ t$ variable $x, t$

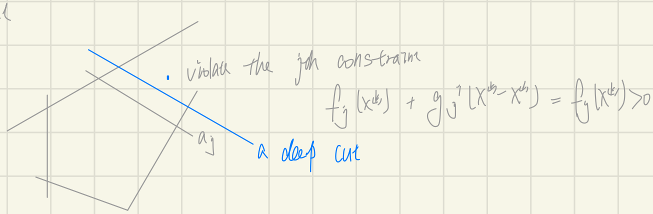
s.t. $f_0(x) \le t$

$f_i(x) \le 0$

at each $(x,t)$, need cutting-plane oracle that separate $(x,t)$ from $(x^*, p^*)$

if $x^{(k)}$ is infeasible for original problem and violates $j$th constraints
add the cutting-plane

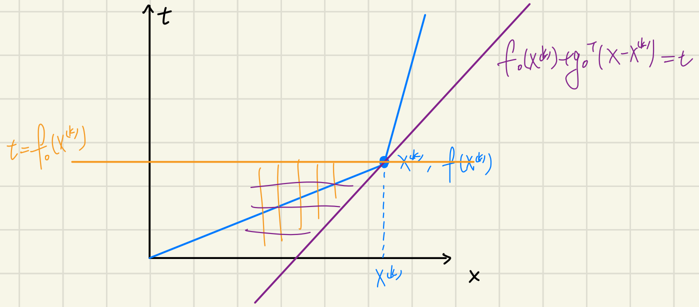
$$\underbrace{f_j(x^{(k)}) + g_j^T (x - x^{(k)}) \le 0}_{\text{a deep cut}} \qquad g_j \in \partial f_j(x^{(k)}) \qquad f_j(x) \ge f_j(x^{(k)}) + g_j^T(x - x^{(k)})$$




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

$$f_0(x^{(k)}) + g_0^T (x - x^{(k)}) \le t \qquad g_0 \in \partial f_0(x^{(k)})$$

$$g^T x + [f_0(x^{(k)}) - g^T x^{(k)}] \le t$$



- **Piece-wise-linear lower bound on convex function**

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

for all $z$

$$f(z) \geq f(x^{(i)}) + g^{(i)T}(z - x^{(i)}) \qquad \text{for any } x^{(i)}$$

So

$$f(z) \geq \hat{f}(z) = \max_i \left\{ f(x^{(i)}) + g^{(i)T}(z - x^{(i)}) \right\}$$

a single subgradient gives an affine lower bound.

a set of subgradients gives a piece-wise-linear lower bound

- **Lower bound**

in solving convex problem

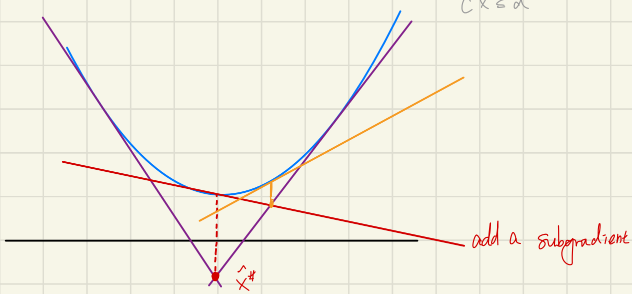
$$\begin{aligned} \min.\ & f_0(x) \\ \text{s.t.}\ & f_i(x) \leq 0 \\ & Cx \leq d \end{aligned}$$

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

form piecewise-linear relaxed problem

$$\begin{aligned} \min.\ & \max_k \left\{ f_0(x^{(k)}) + g_i^{(k)T}(x - x^{(k)}) \right\} \\ \text{s.t.}\ & \max_k \left\{ f_i(x^{(k)}) + g_i^{(k)T}(x - x^{(k)}) \right\} \leq 0 \\ & Cx \leq d \end{aligned} \qquad \text{LP} \implies \qquad \begin{aligned} \min.\ & t \\ \text{s.t.}\ & f_0(x^{(k)}) + g_0^{(k)T}(x - x^{(k)}) \leq t \\ & u \leq 0 \\ & f_i(x^{(k)}) + g_i^{(k)}(x - x^{(k)}) \leq u \\ & Cx \leq d \end{aligned}$$

the optimal value of PWL relaxed problem is a lower bound of the original problem



- Analytic center cutting plane method

analytic center of polyhedron $P = \{z \mid Az \le b\}$ is

$$AC(P) = \underset{x}{\arg\min} - \sum_i \log(b_i - a_i^T x)$$

ACCPM is localization method with next query point $x^{(k+1)} = AC(P_k)$

ACCPM algorithm

given an initial polyhedron $P_0$ known to contain $X$

$k := 0$

repeat:

compute $x^{(k+1)} = AC(P_k)$

Query cutting-plane oracle at $x^{(k+1)}$

if $x^{(k+1)} \in X$

return $x^{(k+1)}$

else:

add returned inequality to $P_k$ $\quad P_{k+1} = P_k \cap \{z \mid a^T z \le b\}$

if $P_{k+1} = \emptyset$ quit

$k := k+1$

- Construction Cutting Planes (inequality-constrained)

$$\min.\ f_0(x)$$
$$\text{s.t. } f_i(x) \le 0$$

if $x$ is not feasible, chose any violate constraint $f_j(x) > 0$, we have (deep) feasibility cut

$$\{z \mid f_j(x) + g_j^T(z-x) \le 0\} \quad g_j \in \partial f_j(x)$$

$\left\{\begin{array}{l}\text{for any } z \text{ s.t. } f(x) + g^T(z-x) > 0 \\ f(z) \ge 0\end{array}\right.$

if $x$ is feasible, we have (deep) objective cut

$$\{z \mid g_0^T(z-x) + f_0(x) - f_{best}^{(k)} \le 0\} \quad g_0 \in \partial f_0(x)$$

for any $z$ s.t. $f(x) + g^T(z-x) > f_{best}^{(k)}$

$f(z) \ge f_{best}^{(k)}$

- Computing the analytic center

$$\min \; -\sum_i \log(b_i - a_i^T x) \qquad \mathbf{dom}\,\phi = \{x \mid a_i^T x < b\}$$

① use phase I method to find a point in dom $\phi$ (or determine dom $\phi = \emptyset$) then use standard newton method

② use infeasible start newton method starting from a point outside dom $\phi$

③ solve the dual

- Infeasible start newton method

$$\min \; -\sum_i \log y_i$$
$$\text{s.t.} \quad y = b - Ax$$

Can start from any $x$ and any $y > 0$ eg. $y_i = \begin{cases} b_i - a_i^T x & b_i - a_i^T x > 0 \\ 1 & \text{otherwise} \end{cases}$

If an equality constraint is satisfied. it remains satisfied in later iterations

$$\min f(x) \qquad \mathcal{L}(x,v) = f(x) + v^T(Ax - b) \qquad \nabla_x \mathcal{L}(x+\Delta x) \approx \nabla f(x) + \nabla^2 f(x)\,\Delta x + A^T v$$
$$\text{s.t. } Ax = b \qquad \nabla_x \mathcal{L} = \nabla f(x) + A^T v$$

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ v \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ b - Ax \end{bmatrix} \qquad \begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta v \end{bmatrix} = -\begin{bmatrix} \nabla f(x) + A^T v \\ Ax - b \end{bmatrix}$$

$$\begin{aligned} &\min \; -\sum_i \log y_i \\ &\text{s.t.} \; y = b - Ax \end{aligned} \;\Rightarrow\; \begin{aligned} &\min \; -[0 \;\; 1] \cdot \log \begin{bmatrix} x \\ y \end{bmatrix} \\ &\text{s.t.} \; [A \;\; I] \begin{bmatrix} x \\ y \end{bmatrix} = b \end{aligned}$$

$$\mathcal{L}\left(\begin{bmatrix} x \\ y \end{bmatrix}, v\right) = -[0, 1] \cdot \log \begin{bmatrix} x \\ y \end{bmatrix} + v^T\left([A \;\; I]\begin{bmatrix} x \\ y \end{bmatrix} - b\right)$$

$$\begin{cases} \text{dual residual} = \nabla \mathcal{L}\left(\begin{bmatrix} x+\Delta x \\ y+\Delta y \end{bmatrix}, v^+\right) \approx -\begin{bmatrix} 0 \\ 1/y \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & H \end{bmatrix}\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} + [A \;\; I]^T v^+ := 0 \\ \text{primal residual} = b - [A \;\; I]\begin{bmatrix} x+\Delta x \\ y+\Delta y \end{bmatrix} := 0 \end{cases}$$

$$\underbrace{\begin{bmatrix} 0 & 0 & A^T \\ 0 & H & I \\ A & I & 0 \end{bmatrix}}_{\text{arrow-shape}} \begin{bmatrix} \Delta x \\ \Delta y \\ v^+ \end{bmatrix} = \begin{bmatrix} 0 \\ 1/y \\ b - Ax - y \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & 0 & A^T \\ 0 & H & I \\ A & I & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta v \end{bmatrix} = \begin{bmatrix} -A^T v \\ 1/y - v \\ b - Ax - y \end{bmatrix} = \begin{bmatrix} -r_d \\ \\ -r_p \end{bmatrix}$$

$$= \begin{bmatrix} -A^T v \\ -g - v \\ b - Ax - y \end{bmatrix}$$

$$H = \mathrm{diag}(1/y^2) \qquad g = -1/y$$

$$r_d = \begin{bmatrix} A^T v \\ g + v \end{bmatrix} \qquad r_p = [y + Ax - b]$$

$r_3 - H^{-1} r_2$: $\begin{bmatrix} 0 & 0 & A^T \\ 0 & H & I \\ A & 0 & -H^{-1} \end{bmatrix}$

$r_1 + A^T H (r_3 - H^{-1} r_2) = -A^T v - A^T H r_p - A^T(-g - v)$
$= A^T g - A^T H r_p$

$r_1 + A^T H (r_3 - H^{-1} r_2)$:

$$\begin{bmatrix} A^T H A & 0 & 0 \\ 0 & H & I \\ A & I & 0 \end{bmatrix} \cdot \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta v \end{bmatrix} = \begin{bmatrix} A^T g - A^T H r_p \\ -g - v \\ -r_p \end{bmatrix}$$

$\Delta x = (A^T H A)^{-1} (A^T g - A^T H r_p)$

$r_2 + H (r_3 - H^{-1} r_2)$:

$r_2 + H (r_3 - H^{-1} r_2) = -H r_p$

$$\begin{bmatrix} 0 & 0 & A^T \\ HA & H & 0 \\ A & 0 & -H^{-1} \end{bmatrix} \cdot \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta v \end{bmatrix} = \begin{bmatrix} \\ -H r_p \\ \\ \end{bmatrix}$$

$HA \Delta x + H \Delta y = -H r_p$

$\Delta y = -A \Delta x - r_p$

$$\begin{bmatrix} 0 & 0 & A^T \\ 0 & H & I \\ A & I & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta v \end{bmatrix} = \begin{bmatrix} -A^T v \\ -g - v \\ -r_p \end{bmatrix}$$

$H \Delta y + \Delta v = -g - v$

$\Delta v = -H \Delta y - g - v$

options for computing $\Delta x$ $\quad (A^T H A)^{-1} (A^T g - A^T H r_p)$

(1). form $A^T H A$ and solve with cholesky factorization

(2) solve least-square problem

$$\Delta x = \arg\min_z \| H^{1/2} A z + H^{1/2} r_p - H^{-1/2} g \|_2$$

(3). use iterative method such as conjugate gradients to solve for $\Delta x$

**given** starting point $x$, $y \succ 0$, tolerance $\epsilon > 0$, $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$.
$\nu := 0$.
Compute residuals from (2).
**repeat**
1. Compute Newton step $(\Delta x, \Delta y, \Delta \nu)$ using (2).
2. *Backtracking line search on* $\|r\|_2$.
   $t := 1$.
   **while** $y + t\Delta y \not\succ 0$, $\quad t := \beta t$.
   **while** $\|r(x + t\Delta x, y + t\Delta y, \nu + t\Delta\nu)\|_2 > (1 - \alpha t)\|r(x, y, \nu)\|_2$, $\quad t := \beta t$.
3. *Update.* $x := x + t\Delta x$, $y := y + t\Delta y$, $\nu := \nu + t\Delta\nu$.

**until** $y = b - Ax$ and $\|r(x, y, \nu)\|_2 \le \epsilon$.

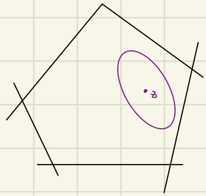
- Pruning constraints

let $x^*$ be the analytic center of $P=\{z \mid Az \le b\}$

let $H^*$ be the hessian of barrier at $x^*$

$$H^* = \nabla^2 - \sum_i \log(b_i - a_i^T z)\Big|_{z=x^*} = \sum_i \frac{a_i a_i^T}{(b_i - a_i^T x^*)^2}$$

} that's $A^T H A$ when calculating the analytic center, calculated at the last step of newton




$\{x \mid (x-z)^T H(z)(x-z) \le 1\}$

short-axis → large eigenvalue of Hessian → "more convex"

this ellipsoid is guaranteed to be in this polyhedron (whether or not $z$ is the analytic center)

---

$P = \{x \mid Ax \le b\}$

$\mathcal{E}_{inner} = \{x \mid (x-x_{ac})^T H (x-x_{ac}) \le 1\}$

$\mathcal{E}_{out} = \{x \mid (x-x_{ac})^T H (x-x_{ac}) \le m(m-1)\}$

($x \in \mathcal{E}_{out} \Rightarrow x \in P$ : CVX book p422)

suppose $x \in \mathcal{E}_{inner}$

$$(x-x_{ac})^T H (x-x_{ac}) \le 1$$

$$= (x-x_{ac})^T \sum_i \frac{a_i a_i^T}{(b_i - a_i^T x_{ac})^2} (x-x_{ac})$$

$$= \sum_i \left(\frac{a_i^T (x-x_{ac})}{b_i - a_i^T x_{ac}}\right)^2 \le 1$$

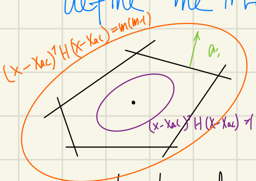
$\therefore$ $a_i^T (x-x_{ac}) \le b_i - a_i^T x_{ac}$ for all $i$s

$\therefore$ $a_i^T x \le b_i$

$\therefore$ if $x \in \mathcal{E}_{inner}$, $x \in P$

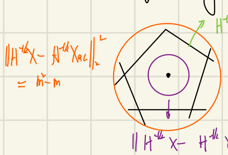
---

define the irrelevance measure: $\eta_i = \dfrac{b_i - a_i^T x}{\sqrt{a_i^T H^{*-1} a_i}} = \dfrac{b_i - a_i^T x^*}{\| H^{-1/2} a_i \|_2}$




polyhedron, calculate the analytic center

↓ change of coordinate $\bar{x} = H^{1/2} x$




the Hessian of log barrier $-\sum \log(b_i - a_i^T x)$ is $I$

a ball of radius 1 fits inside the polyhedron

a ball of radius $m$ is outside the polyhedron

$\eta_i$ is the length in the transformed coordinates

$\eta_i / m$ is the normalized distance from $a_i^T x - b$ to outer ellipsoid

if $\eta_i \ge m$ then constraint $a_i^T x \le b$ is redundant } → safe drop

drop all constraints with $\eta_i > m$ (does not change $P$)

drop constraints in order of irrelevance, keeping constant number, usually $3n \sim 5n$