Analytic Center Cutting-Plane Method

- analytic center cutting-plane method
- computing the analytic center
- pruning constraints
- lower bound and stopping criterion
Analytic center cutting-plane method

The analytic center of polyhedron $\mathcal{P} = \{ z \mid a_i^T z \preceq b_i, \ i = 1, \ldots, m \}$ is

$$\text{AC}(\mathcal{P}) = \arg \min_z - \sum_{i=1}^{m} \log(b_i - a_i^T z)$$

ACCPM is localization method with next query point $x^{(k+1)} = \text{AC}(\mathcal{P}_k)$ (found by Newton’s method)
ACCPM algorithm

given an initial polyhedron $\mathcal{P}_0$ known to contain $X$.

$k := 0$.

repeat

Compute $x^{(k+1)} = AC(\mathcal{P}_k)$.

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

If $x^{(k+1)} \in X$, quit.

Else, add returned cutting-plane inequality to $\mathcal{P}$.

\[ \mathcal{P}_{k+1} := \mathcal{P}_k \cap \{z \mid a^Tz \leq b\} \]

If $\mathcal{P}_{k+1} = \emptyset$, quit.

$k := k + 1$.\]
Constructing cutting-planes

minimize \( f_0(x) \)
subject to \( f_i(x) \leq 0, \quad i = 1, \ldots, m \)

\( f_0, \ldots, f_m : \mathbb{R}^n \rightarrow \mathbb{R} \) convex; \( X \) is set of optimal points; \( p^* \) is optimal value

• if \( x \) is not feasible, say \( f_j(x) > 0 \), we have (deep) feasibility cut

\[
f_j(x) + g_j^T(z - x) \leq 0, \quad g_j \in \partial f_j(x)
\]

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

\[
g_0^T(z - x) + f_0(x) - f_{\text{best}}^{(k)} \leq 0, \quad g_0 \in \partial f_0(x)
\]
Computing the analytic center

we must solve the problem

\[
\text{minimize } \Phi(x) = -\sum_{i=1}^{m} \log(b_i - a^T_i x)
\]

where \( \text{dom } \Phi = \{x | a^T_i x < b_i, \ i = 1, \ldots, m\} \)

• **challenge**: we are not given a point in \( \text{dom } \Phi \)

• some options:
  
  – use phase I method to find a point in \( \text{dom } \Phi \) (or determine that \( \text{dom } \Phi = \emptyset \)); then use standard Newton method to compute AC
  
  – use infeasible start Newton method starting from a point outside \( \text{dom } \Phi \)
Infeasible start Newton method

\[
\text{minimize} \quad -\sum_{i=1}^{m} \log y_i \\
\text{subject to} \quad y = b - Ax
\]

with variables \( x \) and \( y \)

- can be started from \textit{any} \( x \) and \textit{any} \( y \succ 0 \)
- \textit{e.g.}: take initial \( x \) as previous point \( x_{\text{prev}} \), and choose \( y \) as

\[
y_i = \begin{cases} 
  b_i - a_i^T x & b_i - a_i^T x > 0 \\
  1 & \text{otherwise}
\end{cases}
\]
• define primal and dual residuals as

\[ r_p = y + Ax - b, \quad r_d = \begin{bmatrix} A^T \nu \\ g + \nu \end{bmatrix} \]

where \( g = -\text{diag}(1/y_i)1 \) is gradient of objective and \( r = (r_d, r_p) \)

• Newton step at \((x, y, \nu)\) is defined by

\[
\begin{bmatrix}
0 & 0 & A^T \\
0 & H & I \\
A & I & 0
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta y \\
\Delta \nu
\end{bmatrix} = -\begin{bmatrix}
 r_d \\
 r_p
\end{bmatrix},
\]

where \( H = \text{diag}(1/y_i^2) \) is Hessian of the objective
• solve this system by block elimination

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

\[
\Delta y = -A \Delta x - r_p
\]

\[
\Delta \nu = -H \Delta y - g - \nu
\]

• options for computing \( \Delta x \):
  – form \( A^T H A \), then use dense or sparse Cholesky factorization
  – solve (diagonally scaled) least-squares problem

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

  – use iterative method such as conjugate gradients to (approximately) solve for \( \Delta x \)
Infeasible start Newton method algorithm

given starting point $x, y > 0$, tolerance $\epsilon > 0$, $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$.
$
\nu := 0.
$
repeat

1. Compute Newton step $(\Delta x, \Delta y, \Delta \nu)$ by block elimination.
2. Backtracking line search on $\|r\|_2$.
   
   \[ t := 1. \]
   
   while $y + t\Delta y \neq 0$, $t := \beta t$.
   
   while $\|r(x + t\Delta x, y + t\Delta y, \nu + t\Delta \nu)\|_2 > (1 - \alpha t)\|r(x, y, \nu)\|_2$, $t := \beta t$.

3. Update. $x := x + t\Delta x$, $y := y + t\Delta y$, $\nu := \nu + t\Delta \nu$.

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

EE364b, Stanford University
Properties

• once any equality constraint is satisfied, it remains satisfied for all future iterates

• once a step size $t = 1$ is taken, all equality constraints are satisfied

• if $\text{dom} \Phi \neq \emptyset$, $t = 1$ occurs in finite number of steps

• if $\text{dom} \Phi = \emptyset$, algorithm never converges
Pruning constraints

- let $x^*$ be analytic center of $\mathcal{P} = \{ z \mid a_i^T z \preceq b_i, \; i = 1, \ldots, m \}$

- let $H^*$ be Hessian of barrier at $x^*$,

$$H^* = -\nabla^2 \sum_{i=1}^{m} \log(b_i - a_i^T z) \bigg|_{z=x^*} = \sum_{i=1}^{m} \frac{a_i a_i^T}{(b_i - a_i^T x^*)^2}$$

- then, $\mathcal{P} \subseteq \mathcal{E} = \{ z \mid (z - x^*)^T H^* (z - x^*) \leq m^2 \}$

define (ir)relevance measure $\eta_i = \frac{b_i - a_i^T x^*}{\sqrt{a_i^T H^* - 1} a_i}$
• $\eta_i/m$ is normalized distance from hyperplane $a_i^T x = b_i$ to outer ellipsoid

• if $\eta_i \geq m$, then constraint $a_i^T x \leq b_i$ is redundant

common ACCPM constraint dropping schemes:

• drop all constraints with $\eta_i \geq m$ (guaranteed to not change $\mathcal{P}$)

• drop constraints in order of irrelevance, keeping constant number, usually $3n - 5n$
PWL lower bound on convex function

• suppose $f$ is convex, and $g^{(i)} \in \partial f(x^{(i)}), i = 1, \ldots, m$

• then we have

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

• $\hat{f}$ is PWL lower bound on $f$
Lower bound in ACCPM

• in solving convex problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_1(x) \leq 0, \\
& \quad Cx \preceq d
\end{align*}
\]

(by taking max of constraint functions we can assume there is only one)

• we have evaluated \(f_0\) and subgradient \(g_0\) at \(x^{(1)}, \ldots, x^{(q)}\)

• we have evaluated \(f_1\) and subgradient \(g_1\) at \(x^{(q+1)}, \ldots, x^{(k)}\)

• form piecewise-linear approximations \(\hat{f}_0, \hat{f}_1\)
• form PWL relaxed problem

\[
\begin{align*}
\text{minimize} & \quad \hat{f}_0(x) \\
\text{subject to} & \quad \hat{f}_1(x) \leq 0, \\
& \quad Cx \preceq d
\end{align*}
\]

(can be solved via LP)

• optimal value is a lower bound on \( p^* \)

• can easily construct a lower bound on the PWL relaxed problem, as a by-product of the analytic centering computation

• this, in turn, gives a lower bound on the original problem
- form dual of PWL relaxed problem

\[
\text{maximize} \quad \sum_{i=1}^{q} \lambda_i (f_0(x^{(i)}) - g_0^{(i)}T x^{(i)}) + \sum_{i=q+1}^{k} \lambda_i (f_1(x^{(i)}) - g_1^{(i)}T x^{(i)}) - d^T \mu \\
\text{subject to} \quad \sum_{i=1}^{q} \lambda_i g_0^{(i)} + \sum_{i=q+1}^{k} \lambda_i g_1^{(i)} + C^T \mu = 0 \\
\mu \succeq 0, \quad \lambda \succeq 0, \quad \sum_{i=1}^{q} \lambda_i = 1,
\]

- optimality condition for \( x^{(k+1)} \)

\[
\sum_{i=1}^{q} \frac{g_0^{(i)}}{f_{\text{best}} - f_0(x^{(i)}) - g_0^{(i)}T (x^{(k+1)} - x^{(i)})} + \sum_{i=q+1}^{k} \frac{g_1^{(i)}}{f_1(x^{(i)}) - g_1^{(i)}T (x^{(k+1)} - x^{(i)})} + \sum_{i=1}^{m} \frac{c_i}{d_i - c_i^T x^{(k+1)}} = 0.
\]
• take \( \tau_i = 1/(f_{\text{best}}(i) - f_0(x^{(i)}) - g_0^{(i)}T(x^{(k+1)} - x^{(i)})) \) for \( i = 1, \ldots, q \).

• construct a dual feasible point by taking

\[
\lambda_i = \begin{cases} 
\frac{\tau_i}{\mathbf{1}^T \tau} & \text{for } i = 1, \ldots, q \\
\frac{1}{-f_1(x^{(i)}) - g_1^{(i)}T(x^{(k+1)} - x^{(i)})} & \text{for } i = q + 1, \ldots, k,
\end{cases}
\]

\[
\mu_i = \frac{1}{(d_i - c_i^T x^{(k+1)})} \mathbf{1}^T \tau \quad i = 1, \ldots, m.
\]

• using these values of \( \lambda \) and \( \mu \), we conclude that

\[
p^* \geq l^{(k+1)},
\]

where \( l^{(k+1)} = \sum_{i=1}^{q} \lambda_i (f_0(x^{(i)}) - g_0^{(i)}T x^{(i)}) + \sum_{i=q+1}^{k} \lambda_i (f_1(x^{(i)}) - g_1^{(i)}T x^{(i)}) - d^T \mu.\)
Stopping criterion

since ACCPM isn’t a descent method, we keep track of best point found, and best lower bound

- best function value so far: \( f_{\text{best}}^{(k)} = \min_{i=1,\ldots,k} f_0(x^{(k)}) \)

- best lower bound so far: \( l_{\text{best}}^{(k)} = \max_{i=1,\ldots,k} l(x^{(k)}) \)

- can stop when \( f_{\text{best}}^{(k)} - l_{\text{best}}^{(k)} \leq \epsilon \)

- guaranteed to be \( \epsilon \)-suboptimal
Example: Piecewise linear minimization

problem instance with $n = 20$ variables, $m = 100$ terms, $f^* \approx 1.1$
$f_{\text{best}}(k) - f^*$
Newton step

$f_{\text{best}} - f^*$
$f_{\text{best}}^{(k)} - f_{\text{best}}^\star - l_{\text{best}}^{(k)}$
ACCPM with constraint dropping

PWL objective, $n = 20$ variables, $m = 100$ terms
number of inequalities in $\mathcal{P}$:
accuracy versus approximate cumulative flop count

\[ f_{\text{best}}(k) - f^* \]

- no dropping
- keeping \( 3n \)

Megaflops
Epigraph ACCPM

PWL objective, $n = 20$ variables, $m = 100$ terms
$f_{\text{best}}(k) - f^*$ vs. Newton iterations