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

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

$$\mathsf{AC}(\mathcal{P}) = \operatorname*{argmin}_z - \sum_{i=1}^{m} \log(b_i - a_i^T z)$$

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

ACCPM algorithm

given an initial polyhedron $\mathcal{P}_0$ known to contain X.
$k := 0$.
repeat
  Compute $x^{(k+1)} = \mathrm{AC}(\mathcal{P}_k)$.
  Query cutting-plane oracle at $x^{(k+1)}$.
  If $x^{(k+1)} \in X$, quit.
  Else, add returned cutting-plane inequality to $\mathcal{P}$.
    $\mathcal{P}_{k+1} := \mathcal{P}_k \cap \{z \mid a^T z \leq b\}$
  If $\mathcal{P}_{k+1} = \emptyset$, quit.
  $k := k + 1$.

Constructing cutting-planes

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

$f_0, \ldots, f_m : \mathbf{R}^n \to \mathbf{R}$ 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 (deep) *objective cut*

$$g_0^T(z - x) + f_0(x) - f_{\rm best}^{(k)} \leq 0, \qquad g_0 \in \partial f_0(x)$$

Computing the analytic center

we must solve the problem

$$\text{minimize} \quad \Phi(x) = -\sum_{i=1}^m \log(b_i - a_i^T x)$$

where $\mathbf{dom}\,\Phi = \{x \mid a_i^T x < b_i,\ i = 1, \ldots, m\}$

- **challenge**: we are not given a point in $\mathbf{dom}\,\Phi$
- some options:
  - use phase I method to find a point in $\mathbf{dom}\,\Phi$ (or determine that $\mathbf{dom}\,\Phi = \emptyset$); then use standard Newton method to compute AC
  - use infeasible start Newton method starting from a point outside $\mathbf{dom}\,\Phi$

Infeasible start Newton method

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^m \log y_i \\ \text{subject to} & y = b - Ax \end{array}$$

with variables x and y

- can be started from *any* x and *any* $y \succ 0$
- $e.g.$: take initial x as previous point x_{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

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

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

- Newton step at (x, y, ν) 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 $H = \mathbf{diag}(1/y_i^2)$ is Hessian of the objective

- solve this system by block elimination

$$\begin{array}{rcl} \Delta x & = & -(A^T H A)^{-1}(A^T g - A^T H r_p) \\ \Delta y & = & -A\Delta x - r_p \\ \Delta \nu & = & -H\Delta y - g - \nu \end{array}$$

- options for computing Δx:
  - form $A^T H A$, then use dense or sparse Cholesky factorization
  - solve (diagonally scaled) least-squares problem

$$\Delta x = \text{argmin}_z \left\| H^{1/2} A z - H^{1/2} r_p + H^{-1/2} g \right\|_2$$

  - use iterative method such as conjugate gradients to (approximately) solve for Δx

Infeasible start Newton method algorithm

given starting point x, $y \succ 0$, 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 $y + t\Delta y \not\succ 0$, $\qquad t := \beta t$.
   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 $\|r(x, y, \nu)\|_2 \leq \epsilon$.

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 $\mathbf{dom}\,\Phi \neq \emptyset$, $t = 1$ occurs in finite number of steps
- if $\mathbf{dom}\,\Phi = \emptyset$, 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 $\eta_i = \dfrac{b_i - a_i^T x^*}{\sqrt{a_i^T H^{*-1} a_i}}$

- η_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

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0, \\ & Cx \preceq d \end{array}$$

  (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

$$\begin{array}{ll} \text{minimize} & \hat{f}_0(x) \\ \text{subject to} & \hat{f}_1(x) \leq 0, \\ & Cx \preceq d \end{array}$$

(can be solved via LP)

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

- 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

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

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

$$\sum_{i=1}^{q} \frac{g_0^{(i)}}{f_{\rm best}^{(i)} - f_0(x^{(i)}) - g_0^{(i)T}(x^{(k+1)} - x^{(i)})} \quad +$$

$$\sum_{i=q+1}^{k} \frac{g_1^{(i)}}{-f_1(x^{(i)}) - g_1^{(i)T}(x^{(k+1)} - x^{(i)})} \quad + \quad \sum_{i=1}^{m} \frac{c_i}{d_i - c_i^T x^{(k+1)}} = 0.$$

- take $\tau_i = 1/(f_{\rm best}^{(i)} - f_0(x^{(i)}) - g_0^{(i)T}(x^{(k+1)} - x^{(i)}))$ for $i = 1, \ldots, q$.

- construct a dual feasible point by taking

$$\begin{array}{rcl}\lambda_i & = & \left\{\begin{array}{ll} \tau_i/\mathbf{1}^T\tau & \mbox{for } i = 1,\ldots,q \\ 1/(-f_1(x^{(i)}) - g_1^{(i)T}(x^{(k+1)} - x^{(i)}))(\mathbf{1}^T\tau) & \mbox{for } i = q+1,\ldots,k, \end{array}\right. \\ \mu_i & = & 1/(d_i - c_i^T x^{(k+1)})(\mathbf{1}^T\tau) \qquad i = 1,\ldots,m. \end{array}$$

- using these values of λ and μ, we conclude that

$$p^\star \geq l^{(k+1)},$$

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

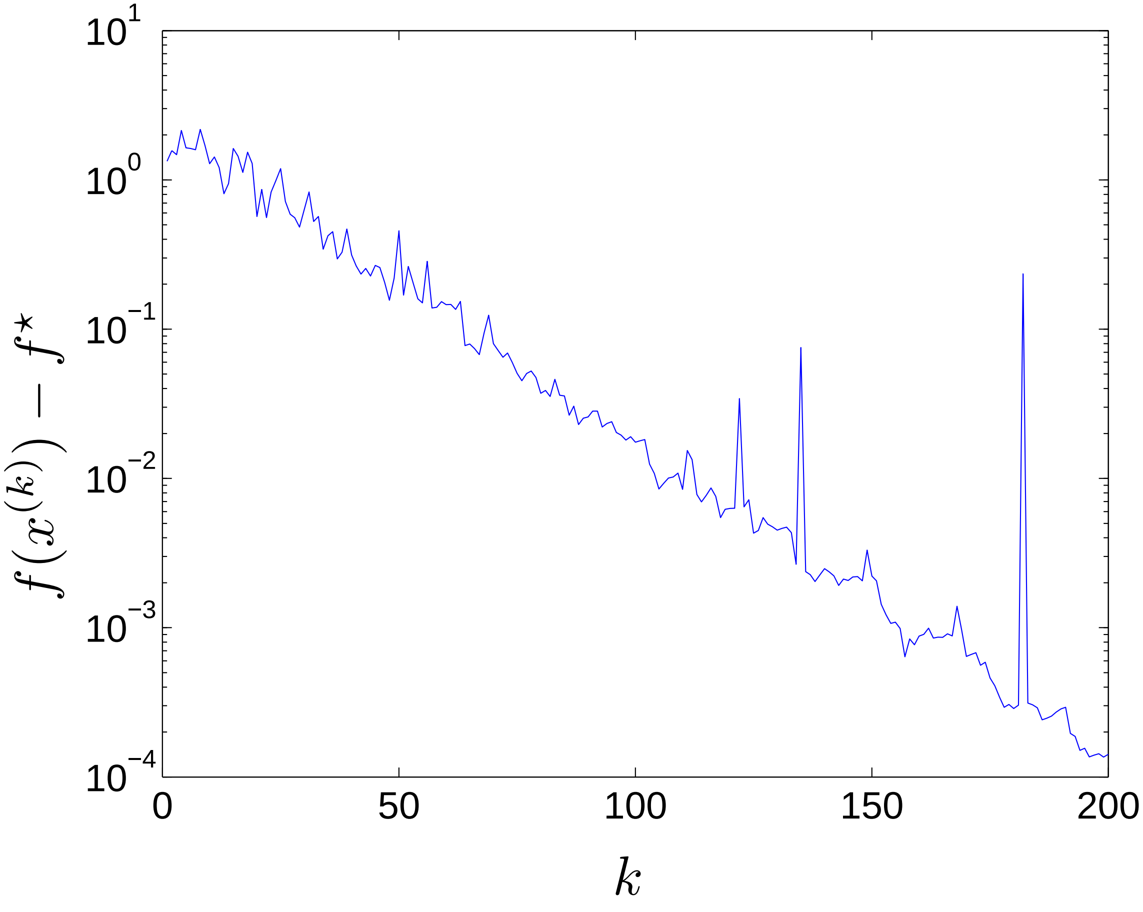
Stopping criterion

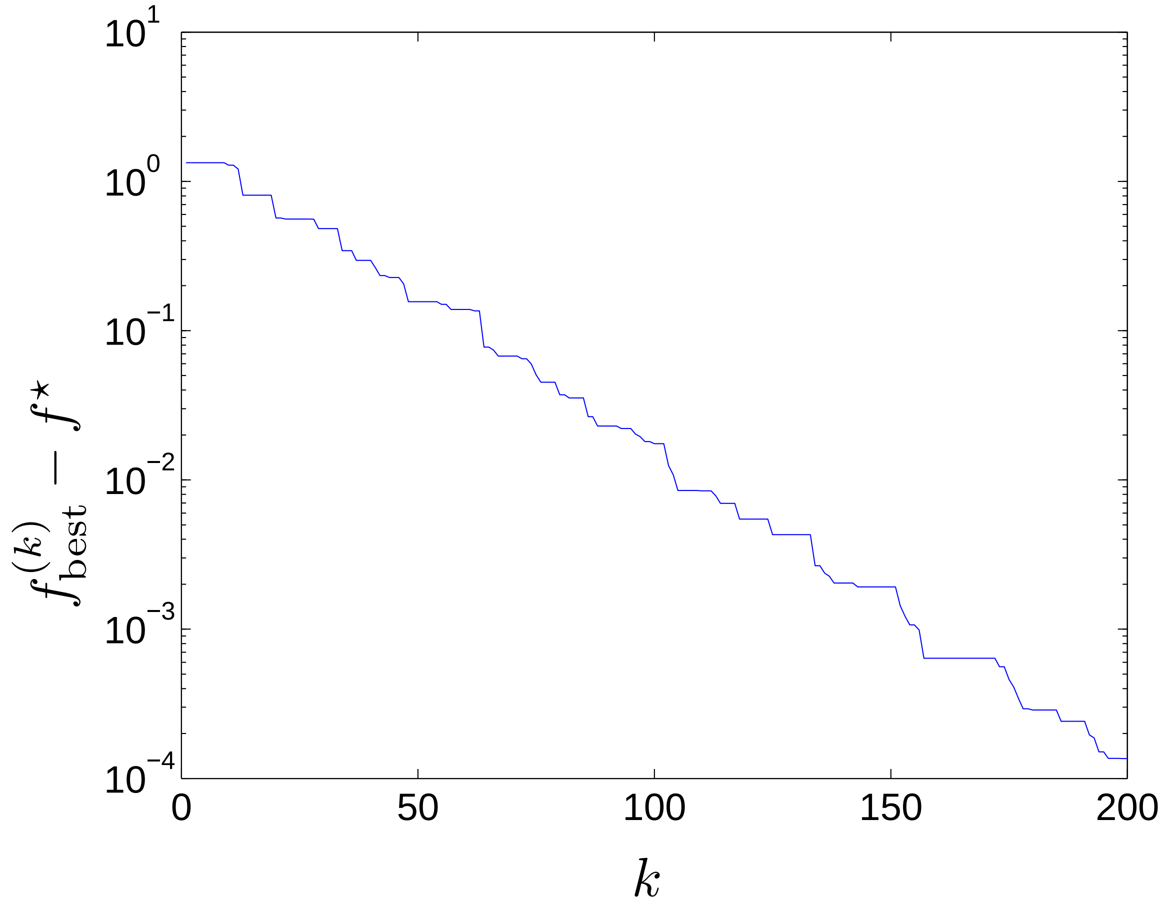
since ACCPM isn't a descent method, we keep track of best point found, and best lower bound

- best function value so far: $f_{\rm best}^{(k)} = \min_{i=1,\ldots,k} f_0(x^{(k)})$
- best lower bound so far: $l_{\rm best}^{(k)} = \max_{i=1,\ldots,k} l(x^{(k)})$
- can stop when $f_{\rm best}^{(k)} - l_{\rm best}^{(k)} \leq \epsilon$
- guaranteed to be ϵ-suboptimal

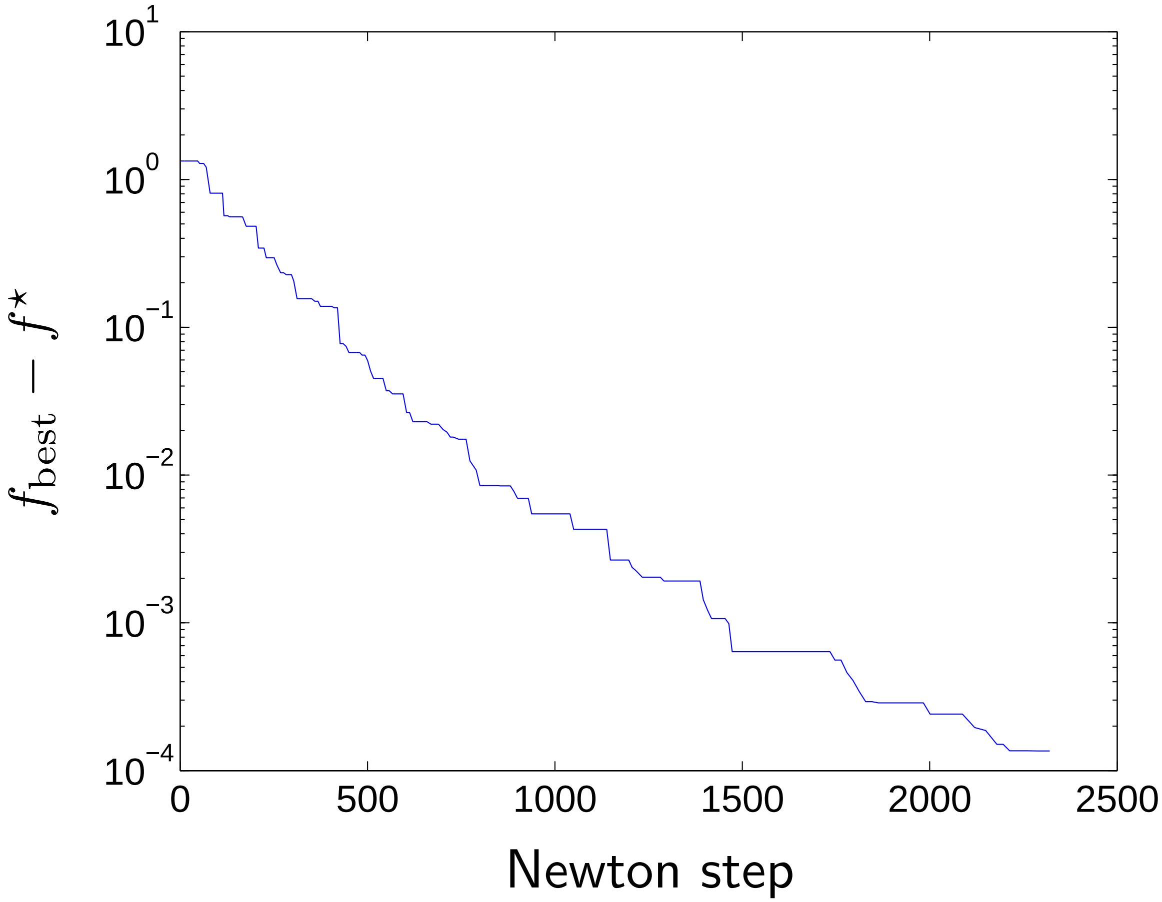
Example: Piecewise linear minimization

problem instance with $n = 20$ variables, $m = 100$ terms, $f^\star \approx 1.1$

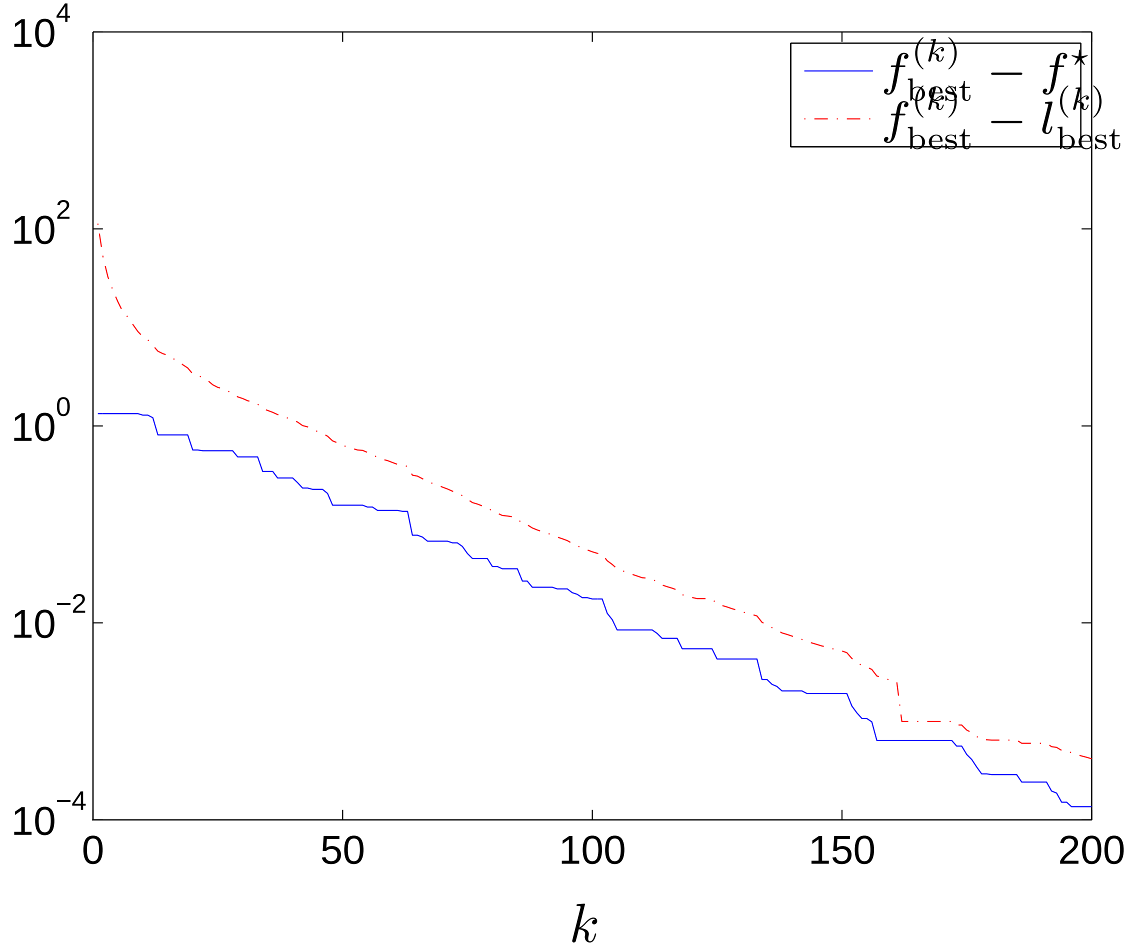




10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0
50
100
150
200
k
$f_{\rm best}^{(k)} - f^\star$



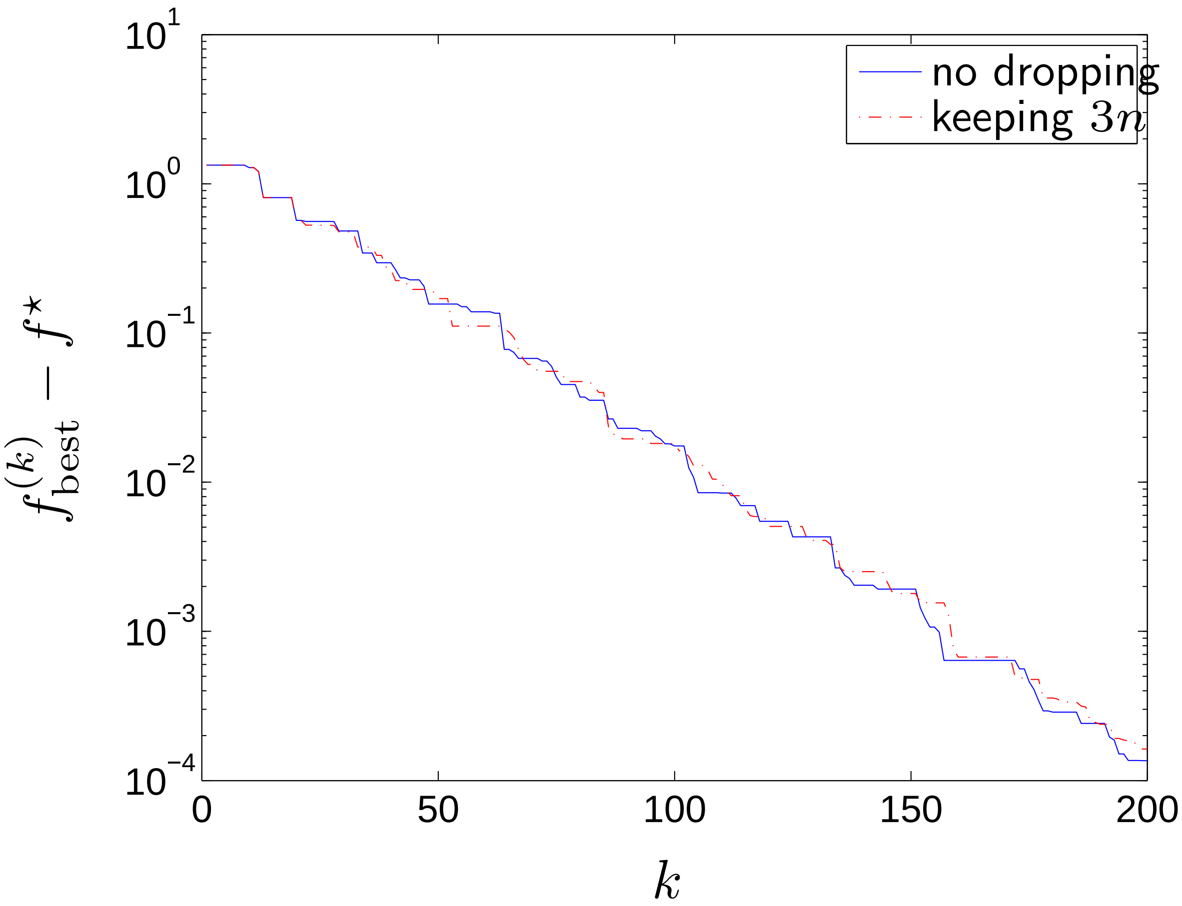
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0
500
1000
1500
2000
2500
$f_{\mathrm{best}} - f^{\star}$
Newton step



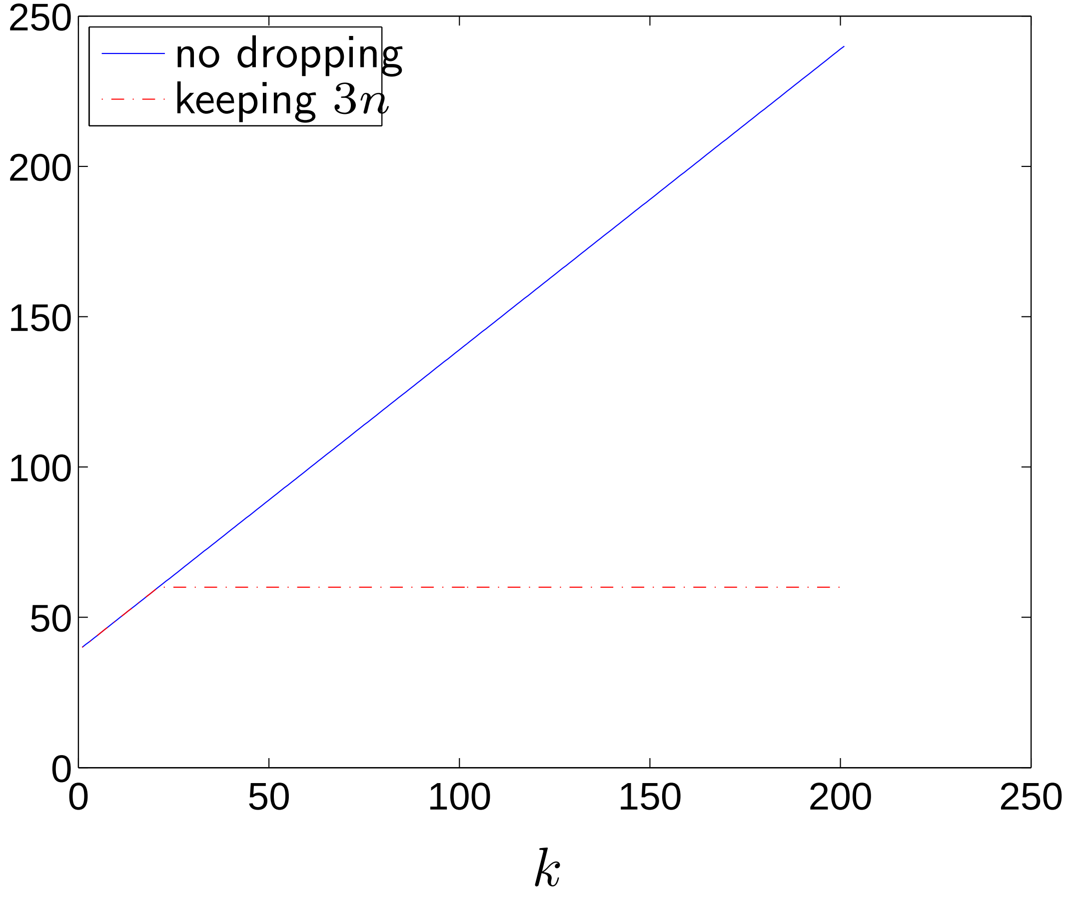
10^4
10^2
10^0
10^{-2}
10^{-4}
0
50
100
150
200
k
$f_{\rm best}^{(k)} - f^\star$
$f_{\rm best}^{(k)} - l_{\rm best}^{(k)}$

ACCPM with constraint dropping

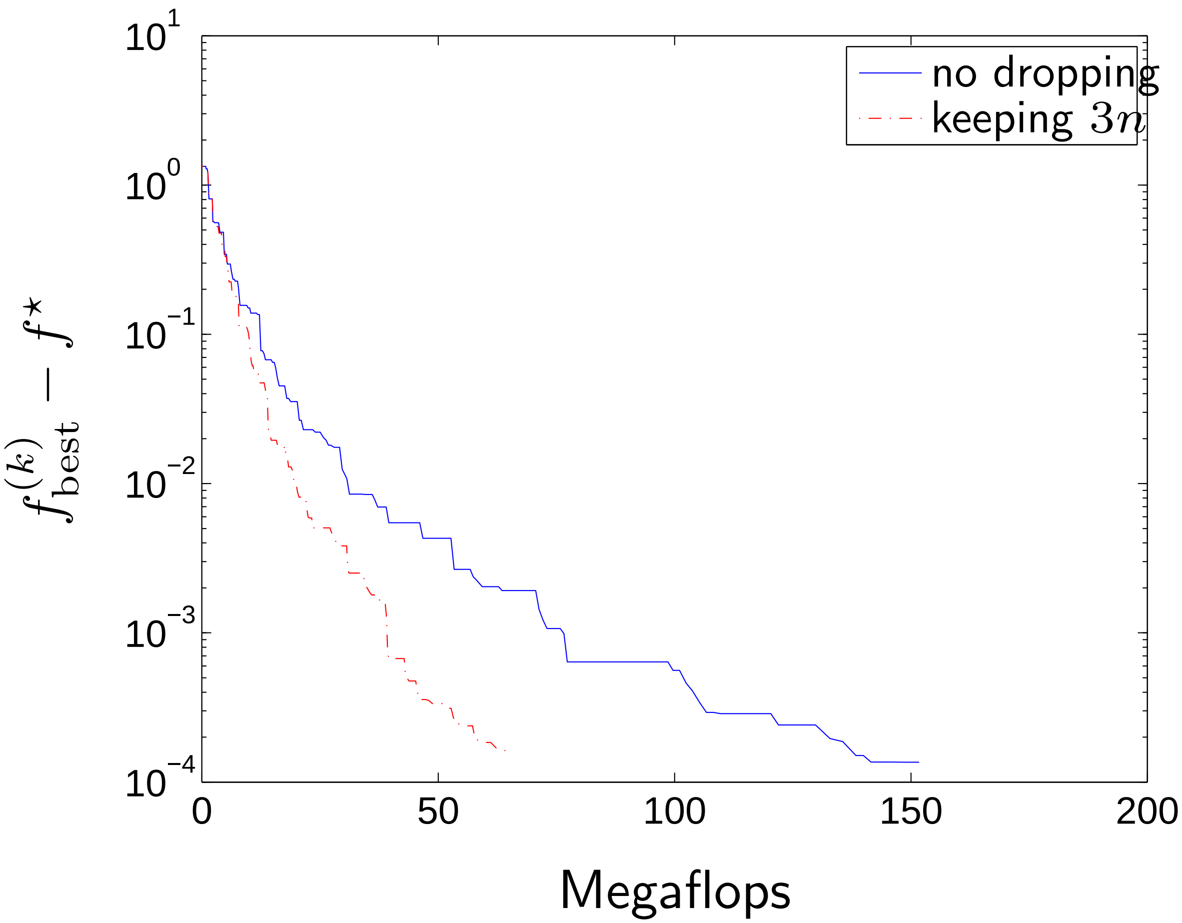
PWL objective, $n = 20$ variables, $m = 100$ terms



number of inequalities in $\mathcal{P}$:

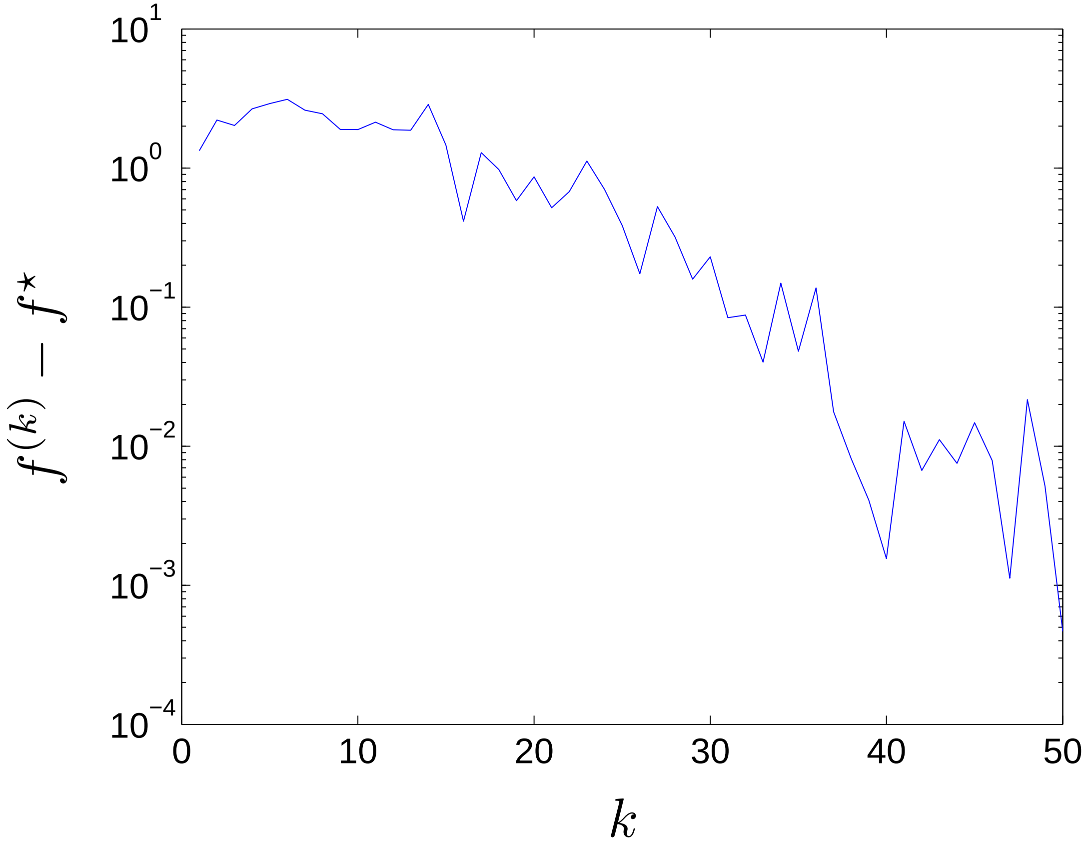


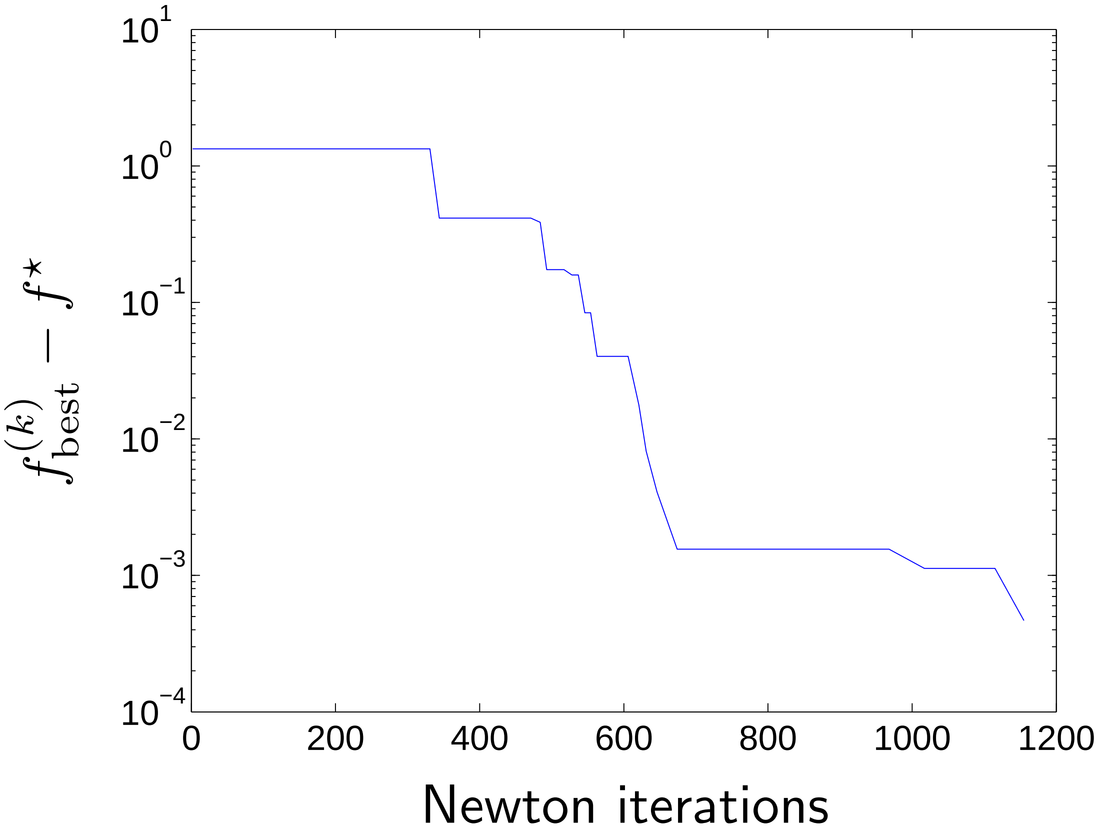
accuracy versus approximate cumulative flop count



Epigraph ACCPM

PWL objective, $n = 20$ variables, $m = 100$ terms





10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0
200
400
600
800
1000
1200
$f_{\text{best}}^{(k)} - f^\star$
Newton iterations