A Proximal Cutting Plane Method Using Chebychev Center for Nonsmooth Convex Optimization

Adam Ouorou

Orange Labs

JFRO in honour of Pierre Huard
Carré des sciences - Paris

Consider the problem

$$(P) \quad \min f(x) \text{ subject to } x \in \mathbb{R}^n$$

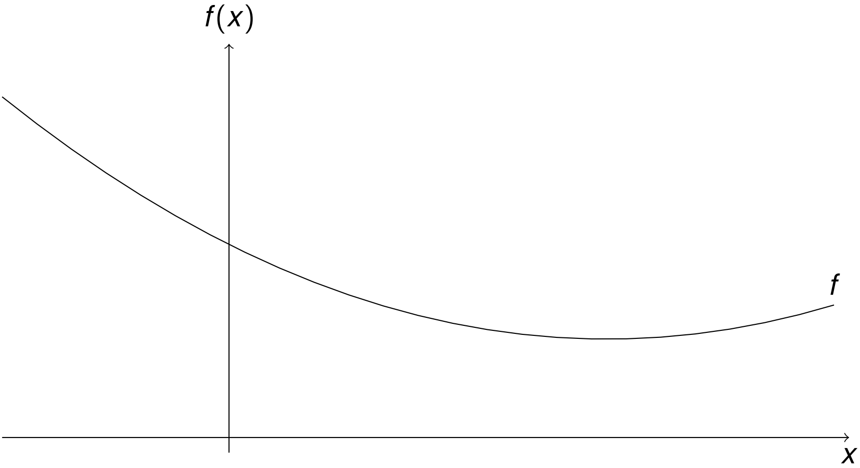
where f is a convex (not necessarily smooth) closed proper function.

Assumption : We are given an ***oracle*** **that compute $f(z)$ and $g \in \partial f(z)$ for any $z \in \mathbb{R}^n$.**

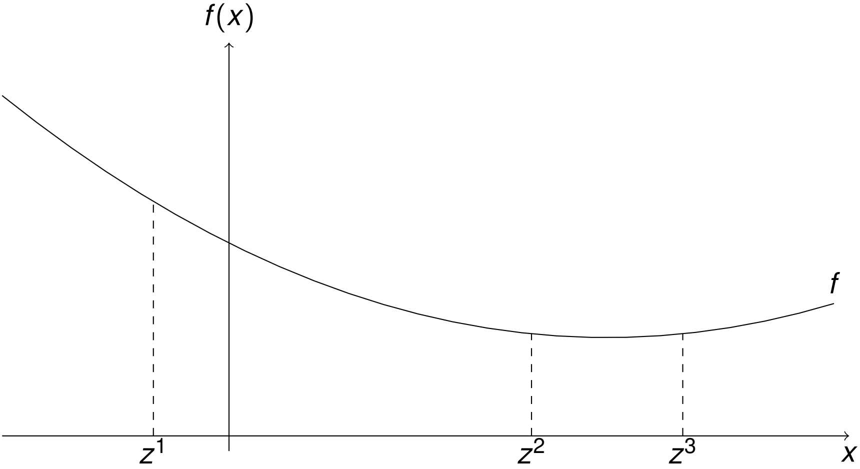
Given some $z^1, z^2, \ldots, z^k \in \mathbb{R}^n$, the function $\check{f}_k$ defined by

$$\check{f}_k(x) = \max\{f^i(x), \;\; i \in I_k\}.$$

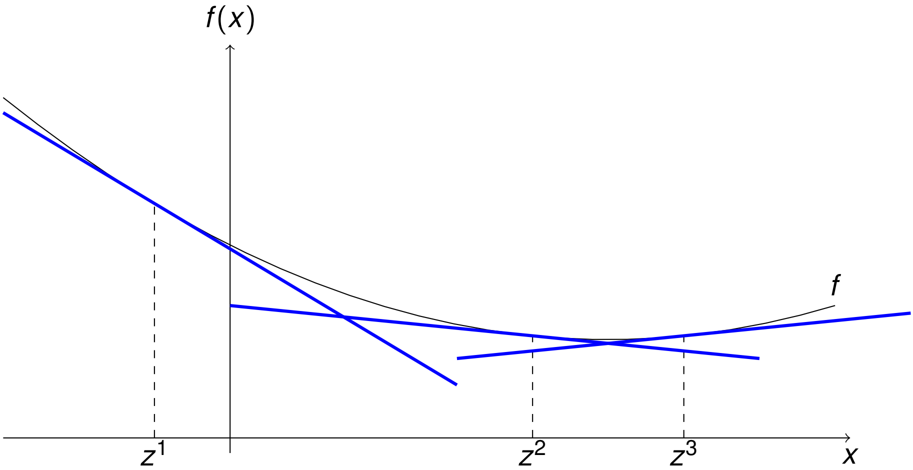
where $f^i(x) = f(z^i) + \langle g^i, \; x - z^i \rangle, \; g^i \in \partial f(z^i), \; I_k = \{1, \ldots, k\}$, is a polyhedral approximation model of f.



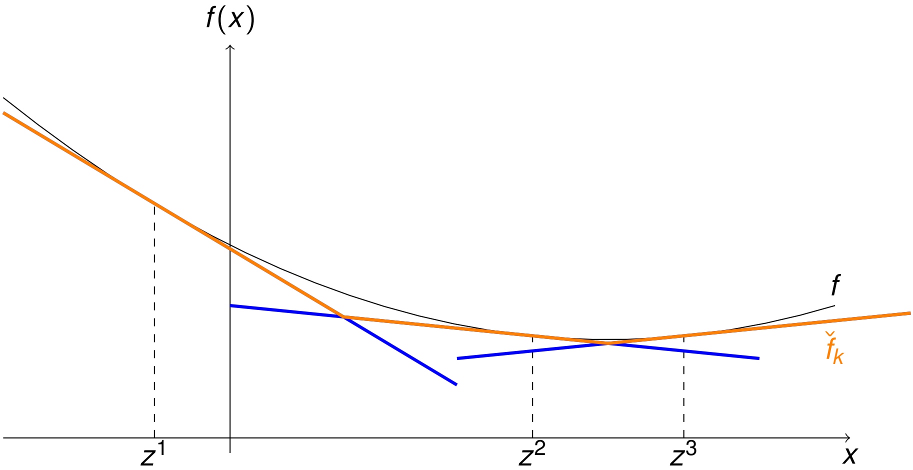
$$\check{f}_k(x) \leq f(x)\ \forall x \in \mathbb{R}^n,\ \check{f}_k(z^i) = f(z^i) \text{ and } g^i \in \partial \check{f}_k(z^i),\ \forall i \in I_k.$$



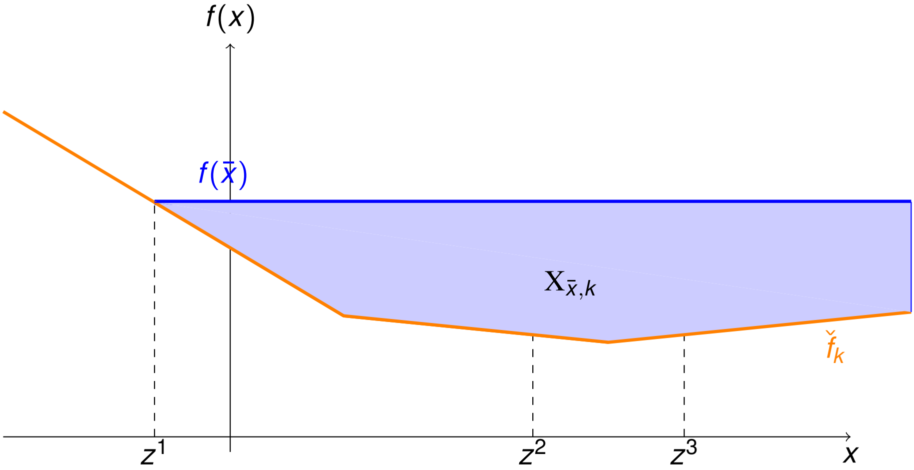
$\check{f}_k(x) \leq f(x)\ \forall x \in \mathbb{R}^n,\ \check{f}_k(z^i) = f(z^i)$ and $g^i \in \partial \check{f}_k(z^i),\ \forall i \in I_k.$



$$\check{f}_k(x) \leq f(x)\ \forall x \in \mathbb{R}^n,\ \check{f}_k(z^i) = f(z^i) \text{ and } g^i \in \partial \check{f}_k(z^i),\ \forall i \in I_k.$$

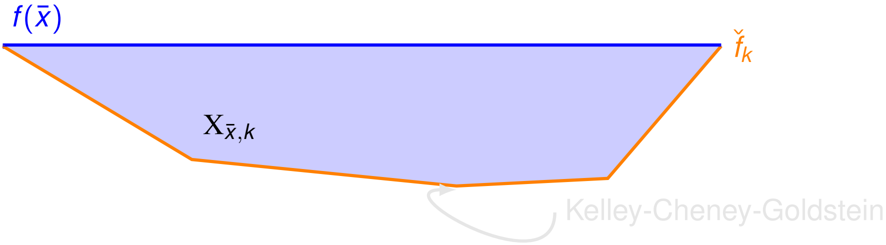


$$\check{f}_k(x) \leq f(x)\ \forall x \in \mathbb{R}^n,\ \check{f}_k(z^i) = f(z^i) \text{ and } g^i \in \partial \check{f}_k(z^i),\ \forall i \in I_k.$$



The set $\mathrm{X}_{\bar{x},k} = \left\{(r,x) \in \mathbb{R} \times \mathbb{R}^n : \; r \leq f(\bar{x}), \; f^i(x) \leq r, \; i \in I_k\right\}$ contains the optimal set of the problem.

Assuming $X_{\bar{x},k}$ *bounded*.

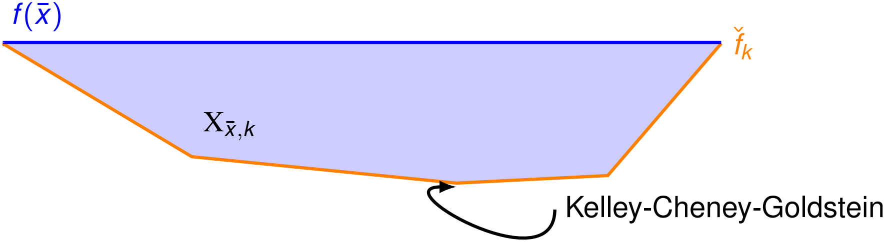


$$(x^{k+1}, r^{k+1}) = \text{analytic center}$$
$$= \text{min product of slacks}$$

$$(x^{k+1}, r^{k+1}) = \text{center of the largest sphere inside } X_{\bar{x},k}$$
$$\sigma = \text{radius of the sphere}$$

Bundle methods overcome the drawback of KCG method (Hiriart-Urruty, Lemaréchal, *Springer-Verlarg, 1993*).

Assuming $X_{\bar{x},k}$ *bounded*.

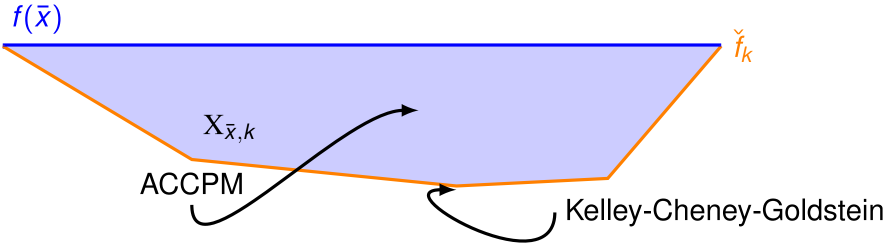


$$\begin{aligned}(x^{k+1}, r^{k+1}) &= \text{analytic center}\\ &= \text{min product of slacks}\end{aligned}$$

$$\begin{aligned}(x^{k+1}, r^{k+1}) &= \text{center of the largest sphere inside } X_{\bar{x},k}\\ \sigma &= \text{radius of the sphere}\end{aligned}$$

Bundle methods overcome the drawback of KCG method (Hiriart-Urruty, Lemaréchal, *Springer-Verlarg, 1993*).

Assuming $X_{\bar{x},k}$ *bounded*.

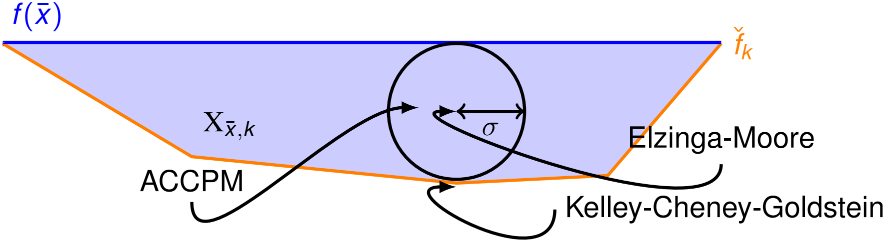


$$\begin{aligned}(x^{k+1}, r^{k+1}) &= \text{analytic center} \\ &= \text{min product of slacks}\end{aligned}$$

$$\begin{aligned}(x^{k+1}, r^{k+1}) &= \text{center of the largest sphere inside } X_{\bar{x},k} \\ \sigma &= \text{radius of the sphere}\end{aligned}$$

Bundle methods overcome the drawback of KCG method (Hiriart-Urruty, Lemaréchal, *Springer-Verlarg, 1993*).

Assuming $X_{\bar{x},k}$ *bounded*.

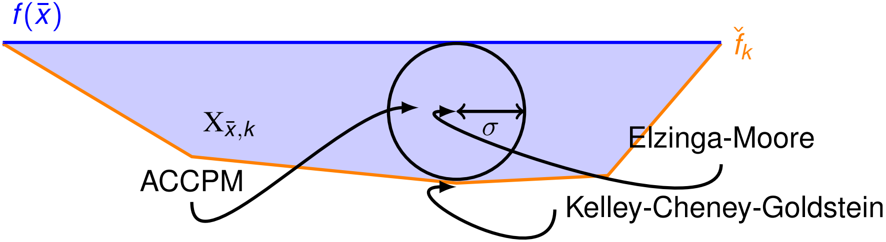


$$
\begin{aligned}
(x^{k+1}, r^{k+1}) &= \text{analytic center} \\
&= \text{min product of slacks}
\end{aligned}
$$

$$
\begin{aligned}
(x^{k+1}, r^{k+1}) &= \text{center of the largest sphere inside } X_{\bar{x},k} \\
\sigma &= \text{radius of the sphere}
\end{aligned}
$$

Bundle methods overcome the drawback of KCG method (Hiriart-Urruty, Lemaréchal, *Springer-Verlarg, 1993*).

Assuming $X_{\bar{x},k}$ *bounded*.



$$(x^{k+1}, r^{k+1}) = \text{analytic center} = \text{min product of slacks}$$

$$(x^{k+1}, r^{k+1}) = \text{center of the largest sphere inside } X_{\bar{x},k}$$

$$\sigma = \text{radius of the sphere}$$

Bundle methods overcome the drawback of KCG method (Hiriart-Urruty, Lemaréchal, *Springer-Verlarg, 1993*).

Finding the centre of the largest sphere inside $\mathrm{X}_{\bar{x},k}$, amounts to solving the linear program

$$
\begin{aligned}
\max \quad & \sigma \\
\text{s.t.} \quad & r + \sigma \le f(\bar{x}), \\
& f^i(x) - r + \sigma\sqrt{1 + \|g^i\|^2} \le 0, \; i \in I_k, \\
& \sigma, \; r \in \mathbb{R}, \quad x \in \mathbb{R}^n.
\end{aligned}
$$

Elzinga-Moore's cutting plane algorithm uses the solution in x as next test point.

When the solution in σ equals 0, the centre is an optimal solution.

Robust deviation (Kouvelis & Yu)

(m, n, K)	# iter		CPU time	
	EM	KCG	EM	KCG
(12, 25, 50)	89	156	17.22	47.47
(17, 26,50)	101	141	14.74	28.74
(19, 34, 50)	235	343	204.09	374.55
(26, 30, 100)	168	504	167.31	477.89
(16, 49, 89)	274	579	1376.06	4086.33
(26, 53, 100)	431	1000*	3517.65	10318.3

* maximum number of iterations (1000) reached.

The approximation is unlikely to be reliable far away from the search points $z^i, \; i \in I_k$. We may consider the following subproblem

$$\begin{aligned} \max \quad & \sigma - \frac{\mu}{2}\|x - x^k\|^2 \\ \text{s.t.} \quad & r + \sigma \leq f(\bar{x}), \\ & f^i(x) - r + \sigma\sqrt{1 + \|g^i\|^2} \leq 0, \quad i \in I_k, \\ & \sigma, \; r \in \mathbb{R}, \quad x \in \mathbb{R}^n, \end{aligned}$$

where $\mu > 0$ is a penalty parameter.

The assumption on $\mathrm{X}_{\bar{x},k}$ *to be bounded is now eliminated since the quadratic term guarantees compactness.*

The convex function f can be written as the envelope of its supporting hyperplanes, i.e.

$$f(x) = \sup \{ f(z) + \langle g,\ x - z \rangle :\ z \in \mathbb{R}^n,\ g \in \partial f(z) \} .$$

Given some *level point* $\bar{x} \in \mathbb{R}^n$, $f(\bar{x})$ is an upper bound to our problem. The part of the epigraph that is below $f(\bar{x})$ defines a (level) set

$$\begin{aligned} \mathrm{X}_{\bar{x}} \ &= \ \Big\{ (x, r) \in \mathbb{R}^n \times \mathbb{R} :\ r \leq f(\bar{x}),\ f(z) + \langle g,\ x - z \rangle \leq r, \\ &\quad\ z \in \mathbb{R}^n,\ g \in \partial f(z) \Big\} \end{aligned}$$

which contains the optimal set.

Remark : $\mathrm{X}_{\bar{x},k}$ *is an outer approximation of* $\mathrm{X}_{\bar{x}}$*.*

Let $\nu = -\sigma$, computing the the Chebychev center of the set $\mathrm{X}_{\bar{x}}$ i.e. solve the semi-infinite linear program

$$\begin{array}{ll} \min & \nu \\ \text{s.t.} & r - \nu \leq f(\bar{x}), \\ & f(z) + \langle g,\ x - z\rangle - r - \nu\sqrt{1 + \|g\|^2} \leq 0,\ z \in \mathbb{R}^n,\ g \in \partial f(z), \\ & \nu,\ r \in \mathbb{R},\quad x \in \mathbb{R}^n. \end{array}$$

At optimality, the constraint $r - \nu \leq f(\bar{x})$ is tight : we can eliminate r and the problem becomes

$$\begin{array}{ll} \min & \nu \\ \text{s.t.} & \dfrac{\langle g,\ x - z\rangle + f(z) - f(\bar{x})}{1 + \sqrt{1 + \|g\|^2}} \leq \nu,\ z \in \mathbb{R}^n,\ g \in \partial f(z), \\ & \nu \in \mathbb{R},\quad x \in \mathbb{R}^n. \end{array}$$

Define the function $\psi_{\bar{x}}$ by

$$\psi_{\bar{x}}(x) = \sup \left\{ \frac{\langle g,\ x - z\rangle + f(z) - f(\bar{x})}{1 + \sqrt{1 + \|g\|^2}} :\ z \in \mathbb{R}^n,\ g \in \partial f(z) \right\}.$$

$\psi_{\bar{x}}$ is a convex function. It can be viewed as a function representation of $\mathrm{X}_{\bar{x}}$

Computing the the Chebychev center of $\mathrm{X}_{\bar{x}}$ is equivalent to

$$\min_{x \in \mathbb{R}^n} \psi_{\bar{x}}(x).$$

Properties of $\psi_{\bar{x}}$

Define the *translated* function at $\bar{x}$ as $f_{\bar{x}}(x) = f(x) - f(\bar{x})$. Then,

- If $f_{\bar{x}}(x) > 0$, then $f_{\bar{x}}(x) \geq 2\psi_{\bar{x}}(x) > 0$.
- If $f_{\bar{x}}(x) \leq 0$, then $f_{\bar{x}}(x) \leq 2\psi_{\bar{x}}(x) \leq 0$.
- If $\psi_{\bar{x}}(x_{\bar{x}}) = 0$ then $\bar{x}$ and $x_{\bar{x}}$ are optimal, otherwise

$$f(x_{\bar{x}}) \leq f(\bar{x}) + 2\psi_{\bar{x}}(x_{\bar{x}}) < f(\bar{x}).$$

Abstract algorithm (AA)

1. Choose some $\bar{x} \in \mathbb{R}^n$.
2. Solve
$$(P_{X_{\bar{x}}}) \quad \min_{x \in \mathbb{R}^n} \psi_{\bar{x}}(x),$$
and let $x_{\bar{x}}$ be an optimal solution.
3. If $\psi_{\bar{x}}(x_{\bar{x}}) = 0$ stop : $\bar{x}$ and $x_{\bar{x}}$ solve the original problem.
4. Set $\bar{x} = x_{\bar{x}}$ and loop to 2.

The problem $(P_{X_{\bar{x}}})$ in Step 2 has no reason to be easy, computing $\psi_{\bar{x}}$-values is already a difficult issue.

First attempt to perform Step 2

Recall that

$$\psi_{\bar{x}}(x) = \sup \left\{ \frac{\langle g,\ x - z\rangle + f(z) - f(\bar{x})}{1 + \sqrt{1 + \|g\|^2}} :\ z \in \mathbb{R}^n,\ g \in \partial f(z) \right\}.$$

With the set of sample points $z^i,\ i \in I^k$, we can build the following lower approximation of $\psi_{\bar{x}}$

$$\check{\psi}_{\bar{x},k}(x) = \max_{i \in I^k} \left\{ \frac{\langle g^i,\ x - z^i\rangle + f(z^i) - f(\bar{x})}{1 + \sqrt{1 + \|g^i\|^2}} \right\},$$

The problem

$$\min_{x \in \mathbb{R}^n} \check{\psi}_{\bar{x},k}(x)$$

is then an approximation of $(P_{X_{\bar{x}}})$.

Elzinga-Moore algorithm

1. Choose some $\bar{x} \in \mathbb{R}^n$. Set $z^1 = \bar{x}$, compute $f(z^1)$, $g^1 \in \partial f(z^1)$. Let $k = 1$.
2. Compute $z^{k+1} = \arg\min_{x \in \mathbb{R}^n} \check{\psi}_{\bar{x},k}(x)$.
3. If $\check{\psi}_{\bar{x},k}(z^{k+1}) = 0$ stop : $\bar{x}$ is an optimal solution.
4. Compute $f(z^{k+1})$ and $g^{k+1} \in \partial f(z^{k+1})$. If $f(z^{k+1}) < f(\bar{x})$ set $\bar{x} = z^{k+1}$.
5. Set $k = k + 1$ and go to Step 2.

Solving $\min_{x \in \mathbb{R}^n} \check{\psi}_{\bar{x},k}(x)$ is equivalent to computing the Chebychev center of $\check{\mathrm{X}}_{\bar{x},k}$.

& If $f(\bar{x}') \leq f(\bar{x})$, then $\psi_{\bar{x}}(x_{\bar{x}}) \leq \psi_{\bar{x}'}(x_{\bar{x}'})$.

Second attempt to perform Step 2 of AA

Define the *Moreau-Yosida* regularization of $\psi_{\bar{x}}$ by

$$\phi_{\bar{x}}(x) = \min_{z \in \mathbb{R}^n} \left\{ \psi_{\bar{x}}(z) + \frac{\mu}{2} \|z - x\|^2 \right\},$$

which is a differentiable convex function, and $\min_{x \in \mathbb{R}^n} \psi_{\bar{x}}(x)$ is equivalent to $\min_{x \in \mathbb{R}^n} \phi_{\bar{x}}(x)$. This can be done with the following algorithm :

1. Set $y^1 = \bar{x}$, $j = 1$.
2. Compute

$$p_{\psi_{\bar{x}}}(y^j) = \arg\min_{z \in \mathbb{R}^n} \left\{ \psi_{\bar{x}}(z) + \frac{\mu}{2} \|z - y^j\|^2 \right\}$$

3. If $y^j = p_{\psi_{\bar{x}}}(y^j)$, stop : $x_{\bar{x}} = y^j$. Otherwise set $y^{j+1} = p_{\psi_{\bar{x}}}(y^j)$.
4. Increase j by 1 and loop to Step 2.

Bundle scheme to compute $p_{\psi_{\bar{x}}}(y^j)$

1. Set $z^1 = y^j$ and $k = 1$.
2. Compute $f(z^k),\ g^k \in \partial f(z^k)$ and update $\check{\psi}_{\bar{x},k}$.

$$z^{k+1} = \arg\min_{z \in \mathbb{R}^n} \left\{ \check{\psi}_{\bar{x},k}(z) + \frac{\mu}{2}\|z - y^j\|^2 \right\}.$$

3. If $\psi_{\bar{x}}(z^{k+1}) = \check{\psi}_{\bar{x},k}(z^{k+1})$, set $p_{\psi_{\bar{x}}}(y^j) = z^{k+1}$ and stop.
4. Increase k by 1 and loop to Step 2.

In practice

- Approximate proximal points :

$$\psi_{\bar{x}}(z^{k+1}) \leq \psi_{\bar{x}}(y^j) - \varrho \left[(\psi_{\bar{x}}(y^j) - \check{\psi}_{\bar{x},k}(z^{k+1}) \right]$$

  for some $0 < \varrho < 1$, which indicates that z^{k+1} approximates $p_{\psi_{\bar{x}}}(y^j)$.

- Difficulty to check the approximate proximal point condition : $\psi_{\bar{x}}(y^j)$ and $\psi_{\bar{x}}(z^{k+1})$ are out of reach.

- We should find a way to identify approximate proximal points through the oracle for f :

$$f(z^{k+1}) \leq f(\bar{x}) + 2\kappa\check{\psi}_{\bar{x},k}(z^{k+1}), \quad 0 < \kappa < 1.$$

1. Select the stopping tolerance ε, the parameter $0 < \kappa < 1$, an initial point $z^1 \in \mathbb{R}^n$. Compute $f(z^1),\ g^1 \in \partial f(z^1)$. Set $x^1 = z^1,\ k = 1$.
2. If $g^k = 0$, terminate.
3. Compute

$$z^{k+1} = \arg\min_{x \in \mathbb{R}^n} \left\{ \check{\psi}_{x^k,k}(x) + \frac{\mu}{2}\|x - x^k\|^2 \right\}$$

   and set $\sigma^k = -\check{\psi}_{x^k,k}(z^{k+1})$. If $\sigma^k \leq \varepsilon$, terminate.
4. Compute $f(z^{k+1})$ and $g^{k+1} \in \partial f(z^{k+1})$.
5. If $f(z^{k+1}) \leq f(x^k) - 2\kappa\sigma^k$, set $x^{k+1} = z^{k+1}$ (descent step). Otherwise $x^{k+1} = x^k$ (null step).
6. Increase k by 1 and loop to 2.

- We use a convergence analysis with identical ideas and techniques as in bundle methods
- Choice of the penalty parameter μ_k
  - Adapatation of Kiwiel's proximity control (Math. Prog 46(1), 1990)
- Size of the QP subproblem can be reduced using aggregation techniques
  - Specialized codes for this type of problem exist (Kiwiel, Frangioni)

Problem	Name	n	f^*
1	CB2	2	1.952224
2	CB3	2	2
3	DEM	2	-3
4	QL	2	7.2
5	LQ	2	$-\sqrt{2}$
6	Mifflin1	2	-1
7	Mifflin2	2	-1
8	Rosen-Suzuki	4	-44
9	Shor	5	22.600162
10	Maxquad	10	-0.841408
11	Maxq	20	0
12	Maxl	20	0

Problem	Name	n	f^*
13	TSP29	29	-9015
14	Badguy	30	-2048
15	TR48	48	-638565
16	Goffin	50	0
17	MxHilb	50	0
18	L1Hilb	50	0
19	Ury100	100	
20	TSP120	120	-1606.3125
21	TSP442	442	-50499.5
22	TSP1173	1173	-56351
23	TSP3038	3038	-136587.5

- Cplex 9.0 for all the LP and QP problems
- Oracle fortran routines from Lukšan and Lemaréchal.
- Implementation in C language
- No constraints delation rules in the algorithms
- All the runs in double precision.

Pb	em			kelley		
	$\#fg$	$\lvert\mathcal{K}\rvert$	f	$\#fg$	$\lvert\mathcal{K}\rvert$	f
1	23	8	1.952226	26	16	1.952225
2	24	10	2.000001	16	8	2
3	21	14	-2.999996	9	5	-2.999999
4	22	17	7.200004	25	13	7.2
5	21	11	-1.414212	30	14	-1.414213
6	25	7	-0.999998	20	9	-0.999999
7	24	10	-0.999999	29	9	-0.999999
8	73	18	-43.999993	79	26	-43.999999
9	82	25	22.600167	76	34	22.600162
10	483	119	-0.841402	995	72	-0.841407
11	17	8	1.846905E-07	17	7	3.662894E-07
12	3	1	0	3	2	0

Pb	em			kelley		
	$\#fg$	$\lvert\mathcal{K}\rvert$	f	$\#fg$	$\lvert\mathcal{K}\rvert$	f
13	353	32	-9015	612	37	-9015
14	270	9	-2047.999877	270	11	-2047.999877
15	455	67	-638564.999801	919	60	-638564.999999
16	51	1	3.390244E-13	51	2	0
17	3	1	0	3	1	0
18	3	1	0	3	1	0
19	448	72	1209.869668	$\star$	80	1209.869928
20	$\star$	23	-1563.691114	$\star$	0	-1402
21	$\star$	0	-46858	$\star$	0	-46858
22	$\star$	0	-51477	$\star$	0	-51477
23	$\star$	0	-127387	$\star$	0	-127387
-	6401	454	-	8183	407	-

$\star$ maximum number of oracle calls (1000) reached.

Pb	em			pc^{3}pa		
	$\#fg$	$\lvert\mathcal{K}\rvert$	f	$\#fg$	$\lvert\mathcal{K}\rvert$	f
1	23	8	1.952226	19	11	1.952224
2	24	10	2.000001	12	8	2
3	21	14	-2.999996	19	18	-2.999999
4	22	17	7.200004	21	14	7.2
5	21	11	-1.414212	8	7	-1.414212
6	25	7	-0.999998	22	11	-0.999999
7	24	10	-0.999999	17	10	-0.999999
8	73	18	-43.999993	33	17	-43.999999
9	82	25	22.600167	31	18	22.600162
10	483	119	-0.841402	81	49	-0.841407
11	17	8	1.846905E-07	124	32	4.026423E-07
12	3	1	0	42	20	3.271027E-12

Pb	em			pc^{3}pa		
	$\#fg$	$\|\mathcal{K}\|$	f	$\#fg$	$\|\mathcal{K}\|$	f
13	353	32	-9015	52	21	-9014.999999
14	270	9	-2047.999877	247	14	-2047.999877
15	455	67	-638564.999801	162	90	-638564.999761
16	51	1	3.390244E-13	52	35	1.601813E-07
17	3	1	0	20	17	3.749418E-07
18	3	1	0	86	71	2.494463E-06
19	448	72	1209.869668	655	185	1209.869752
20	$\star$	23	-1563.691114	166	44	-1606.312499
21	$\star$	0	-46858	685	84	-50499.499999
22	$\star$	0	-51477	690	75	-56350.999997
23	$\star$	0	-127387	$\star$	86	-136579.740648
-	6401	454	-	4243	937	-

$\star$ maximum number of oracle calls (1000) reached.

Pb	bm			pc^{3}pa		
	$\#fg$	$\|\mathcal{K}\|$	f	$\#fg$	$\|\mathcal{K}\|$	f
1	14	12	1.952224	19	11	1.952224
2	16	12	2	12	8	2
3	13	7	-2.999999	19	18	-2.999999
4	17	13	7.2	21	14	7.2
5	6	5	-1.414213	8	7	-1.414212
6	103	86	-0.999995	22	11	-0.999999
7	17	10	-0.999999	18	12	-0.999999
8	40	17	-43.999999	33	16	-43.999999
9	29	19	22.600162	31	18	22.600162
10	43	24	-0.841407	71	44	-0.841407
11	172	90	6.633164E-07	124	32	4.026423E-07
12	42	20	3.091276E-14	42	20	3.271027E-12

Pb	bm			pc³pa		
	#fg	$\lvert\mathcal{K}\rvert$	*f*	#fg	$\lvert\mathcal{K}\rvert$	*f*
13	47	25	-9014.999999	52	21	-9014.999999
14	281	21	-2047.999877	247	14	-2047.999877
15	212	78	-638564.999860	162	90	-638564.999761
16	52	34	1.080161E-10	52	35	1.601813E-07
17	13	10	2.810345E-07	20	17	3.749418E-07
18	16	12	3.814789E-07	86	71	2.494463E-06
19	⋆	165	1209.890546	655	185	1209.869752
20	156	49	-1606.312499	166	44	-1606.312499
21	887	88	-50499.499993	685	84	-50499.499999
22	571	69	-56350.999989	690	75	-56350.999999
23	⋆	90	-136559.059612	⋆	86	-136579.740648
-	4747	956	-	4243	937	-

⋆ maximum number of oracle calls (1000) reached.

- The proposed approach improves on Elzinga-Moore's method
  - remove the compactness assumption
  - more efficient
- An interesting alternative to proximal bundle method
- New class of methods for nonsmooth optimization
- Extension to separable case (through nonlinear multicommodity flow problems)