An Analytic Center Cutting Plane Method in Conic Programming

by **Vasile L. Basescu**
Department of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, NY 12180 USA
vbasescu@cox.net

John E. Mitchell
Department of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, NY 12180 USA
mitchj@rpi.edu

Outline

- Introduce the problem
- Analytic Centers
- Adding cuts
- Recovery of feasibility
- Recovery of the Analytic Center
- Complexity Analysis
- Numerical results
- Future work

What do we want to do?

Solve the **feasibility problem**:

"Given an m-dimensional Hilbert space $(Y, \langle \cdot, \cdot \rangle_Y)$, find an interior point y in the convex bounded set $\Gamma \subset Y$."

- The problem is treated in the context of Conic Optimization

- The method: Analytic Center Cutting Plane Method

Assumptions about the problem

- Γ contains a small ball of radius ε

$$\exists y \in Y : B_Y(y, \varepsilon) \in \Gamma$$

  (Γ is not too flat)

- Γ is contained in a bounded set Ω_0

$$\Gamma \in \Omega_0 := \{y \in Y : c_1 \preceq y \preceq c_2\}$$

  (Γ is bounded)

Theoretical Settings

- $f : D_f \to \mathbb{R}$ is a C^2, strictly convex functional
- D_f - open, convex; $D_f \subseteq (X, \langle \cdot, \cdot \rangle_X)$

For any $x \in D_f$ the **local inner product** (at x):

$$\langle u, v \rangle_x := \langle u, v \rangle_{H(x)} = \langle u, H(x)v \rangle_X.$$

We will use:

- f is **intrinsically self-conjugate**
- K is **self-scaled** cone

Local inner products

- For each cone K fix an arbitrary element $e \in \text{int}(K)$

- Starting now, the inner product will be the one induced by e:

  $\langle u, v \rangle = \langle u, H(e)v \rangle_X$ where $\langle \cdot, \cdot \rangle_X$ is the original inner-product on X

- Use $(K^*, g, H, A^*, \ldots)$ for $(K_e^*, g_e, H_e, A_e^*, \ldots)$

- The vector e has some immediate and useful properties:

$$\|e\| = \sqrt{\theta_f}, g(e) = -e, H(e) = I.$$

Analytic Centers

- $(X, \langle \cdot, \cdot \rangle_X)$ and $(Y, \langle \cdot, \cdot \rangle_Y)$ be two Hilbert spaces
- K self-scaled cone in X
- $f : X \to I\!R$ the intrinsically self-conjugate barrier functional
- $A : X \to Y$ a surjective linear operator

The Analytic Center of:

- $\mathcal{F}_P := \{x \in K : Ax = 0\}$ with respect to $f(x) + \langle c, x \rangle_X$:

$$\begin{array}{rl} \text{min} & f(x) + \langle c, x \rangle_X \\ \text{subject to} & Ax = 0 \\ & x \in K \end{array} \qquad (P_1)$$

- $\mathcal{F}_D := \{s \in K : A^* y + s = c\}$ with respect to $f_e^*(s)$:

$$\begin{array}{rl} \text{min} & f_e^*(s) \\ \text{subject to} & A^* y + s = c \\ & s \in K \end{array} \qquad (D_1)$$

More on Analytic Centers

The KKT equalities:

$$\begin{array}{rcl} g(x)+s &=& 0, \\ g(s)+x &=& 0, \\ Ax &=& 0, \\ A^*y+s &=& c, \\ x,s &\in& K. \end{array}$$

(x, y, s) is a θ - Analytic Center for $\mathcal{F}_P$, $\mathcal{F}_D$ iff

$$\begin{array}{rcl} & x \in \mathcal{F}_P, s \in \mathcal{F}_D & \\ \|I - H(x)^{-\frac{1}{2}} H(s)^{-\frac{1}{2}}\| &\leq& \frac{\theta}{\sqrt{\theta_f}} \end{array}$$

For a θ - Analytic Center:

$$\|x + g(s)\|_{-g(s)} \leq \sqrt{\theta_f} \|I - H(s)^{-\frac{1}{2}} H(x)^{-\frac{1}{2}}\|$$

So:

$$\|x + g(s)\|_{-g(s)} \leq \theta$$

$$\|s + g(x)\|_{-g(x)} \leq \theta$$

Oracle

The existence of an oracle is assumed:

Given a point $\hat{y}$, the oracle:

- recognizes that $\hat{y} \in \Gamma$

or

- returns:
  - $(X, \langle\cdot,\cdot\rangle_X)$ a p-dimensional Hilbert space
  - $A : X \to Y$ an injective linear operator such that:

    $$\Gamma \subseteq \{y \in Y : A^*(\hat{y} - y) \in K\}$$

    K is a self-scaled cone in $(X, \langle\cdot,\cdot\rangle_X)$.
    A defines p central cuts.

- Example (LP). If $A^T\hat{y} \geq c$, then add central cut:

$$\{y : A^T y \leq A^T \hat{y}\}$$

The Idea

- start with Ω_0; take AC_0 - the θ - Analytic Center; i=0;
- Start loop:
- call oracle at AC_i
- if $AC_i \in \Gamma$ STOP with solution
- else
  - add p_i cuts through AC_i
  - generate Ω_{i+1} with AC_{i+1} the θ - AC
  - i=i+1
- Return

What do we analyze?

- The process of adding cuts

- The number of Newton steps required to move from a θ - AC to the next one.

- The total number of cuts added:

  $$O^*(\frac{mP^3\Theta^3}{\varepsilon^2\Lambda^2})$$

  Compare to:

  - Linear Programming: $O^*(\frac{m^2P^2}{\varepsilon^2})$
  - Semidefinite Programming, adding a single cut of size at most P: $O(\frac{m^3P}{\varepsilon^2\mu^2})$
  - Second Order Cone Programming, adding a single cut: $0(\frac{m}{\varepsilon^2\mu})$.

The recovery of feasibility

Before adding the cuts:

- $(X_1, \langle \cdot, \cdot \rangle_1)$, K_1, f_1, $A_1 : X_1 \to Y$

$$\Omega_1 = \{y \in Y : A_1^* y + s = c_1, s \in K_1\},$$

- Let (x_1, y_1, s_1) be a θ - Analytic Center for $\mathcal{F}_P$, $\mathcal{F}_D$:

$$\begin{array}{l} A_1 x_1 = 0, \\ A_1^* y_1 + s_1 = c_1, \\ x_1, s_1 \in K_1 \text{ and } y_1 \in Y \end{array}$$

Add p central cuts: $(X_2, \langle \cdot, \cdot \rangle_2)$, K_2, f_2, $A_2 : X_2 \to Y$

Changes:

- $\Omega_2 := \Omega_1 \cap \{y \in Y : A_2^* y + s = c_2, s \in K_2\}$.

- $\mathcal{F}_P := \{(x, \beta) : A_1 x + A_2 \beta = 0 \text{ with } x \in K_1, \beta \in K_2\}$

- $\mathcal{F}_D := \{(s, \gamma) : A_1^* y + s = c_1, A_2^* y + \gamma = c_2, s \in K_1, \gamma \in K_2, y \in Y\}$.

After adding the cuts (x_1, y_1, s_1) becomes (x_2, y_2, s_2):

$$x_2 = (x_1, \beta), y_2 = y_1, s_2 = (s_1, \gamma),$$

with y_2 on the boundary of Ω_2, $\beta = 0$ and $\gamma = 0$

The recovery of feasibility cont.

- y_2 on the boundary of the new domain Ω_2
- take step to move in Ω_2:

$$(\Delta x, \beta),\ \Delta y \text{ and } (\Delta s, \gamma)$$

- for feasibility:

$$\begin{aligned} A_1 \Delta x + A_2 \beta &= 0, \\ A_1^* \Delta y + \Delta s &= 0, \\ A_2^* \Delta y + \gamma &= 0. \end{aligned}$$

  with $x_1, x_1 + \Delta x, s_1, s_1 + \Delta s \in K_1$ and $\beta, \gamma \in K_2$.

- $x_1 + \Delta x$, $s_1 + \Delta s$ feasible if $\|\Delta x\|_{x_1} \leq 1$, $\|\Delta s\|_{s_1} \leq 1$
- choose β, γ to minimize f_2 and f_2^*:

$$\min_{\beta \in K_2} \{ f_2(\beta) : A_1 \Delta x + A_2 \beta = 0, \|\Delta x\|_{H_1(s_1)^{-1}} \leq 1 \}$$
$$\min_{\gamma \in K_2} \{ f_2^*(\gamma) : A_2^* \Delta y + \gamma = 0, \|\Delta s\|_{s_1} \leq 1 \}$$

Recovery of the θ - Analytic Center

- use primal-dual potential: $\Phi(x,s) := \langle x, s\rangle + f(x) + f^*(s)$
- w - scaling point for the ordered pair (x,s): $H(w)x = s$
- w^* - scaling point for the ordered pair (s,x): $H(w^*)s = x$
- use two types of steps:
  - scaled Nesterov-Todd directions :

$$\begin{aligned} d_x &:= -P_{L,w}(x + g_w(x)), \\ d_s &:= -P_{L^\perp,w^*}(s + g_{w^*}(s)). \end{aligned}$$

  - scaled Newton steps

$$\begin{aligned} D_x &:= 2P_{L,w}(w - x), \\ D_s &:= 2P_{L^\perp,w^*}(w^* - s) \end{aligned}$$

Recovery of the θ - Analytic Center cont.

- **Theorem** If $\|x - w\|_w \geq \frac{1}{4}$ then:

$$\Phi(x + td_x, s + td_s) \leq \Phi(x, s) - \frac{1}{250}.$$

- **Theorem** If at the current point (x, y, s):

$$\|x - w\|_w < \alpha < \frac{1}{4}$$

then at the new point $(x_+, s_+) := (x + D_x, s + D_s)$:

$$\|s_+ + g(x_+)\|_{-g(x_+)} < (1 + \alpha)\frac{\alpha^2}{1 - \alpha} < \frac{1}{5}.$$

If w_+ is the scaling point for the ordered pair (x_+, s_+), then:

$$\|x_+ - w_+\|_{w_+} < \frac{5\alpha^2(1 + \alpha)}{4(1 - \alpha)} < 3\alpha^2 < \frac{1}{5}.$$

- After k iterations:

$$\|x_k + g(s_k)\|_{-g(s_k)} < 5 \cdot 3^{2^{k+1}-3} \cdot \alpha^{2^{k+1}}.$$

The idea behind the analysis

- generate a sequence of sets Ω_i
- use the exact Analytic Centers s_i^c of the sets Ω_i
- There is an upper bound UB_i for $f_i^*(s_i^c)$, for any i
- Compare two consecutive f_i^* at the corresponding AC s_i^c
- $f_{i+1}^*(s_{i+1}^c) \geq f_i^*(s_i^c) + LB_i$
- after k steps :

$$UB_k \geq f_0^*(s_0^c) + \sum_{i=0}^{k-1} LB_i$$

- $UB_k \rightarrow \infty$ slower than $\sum_{i=0}^{k-1} LB_i$ does
- the algorithm returns a solution before

$$UB_k < f_0^*(s_0^c) + \sum_{i=0}^{k-1} LB_i$$

Assumptions and notations about the problem

- The operators $A_i : X_i \to Y$, $i \geq 1$ - injective
- $\|A_i\| = 1$.
- $\|H_i(e_i)^{-1}\| = 1$ for $i \geq 0$ (in the original norm)
- $\sigma_i := \sqrt{\frac{p_i}{\theta_i}} e_i$. The length of this vector, measured in the local inner product induced by e_i is $\|\sigma_i\| = \sqrt{p_i}$.
- Wlog assume that $f_i(\sigma_i) = 0$.

The implications of the initial assumptions

Let (x^k, y^k, s^k) - AC of Ω_k (the outer-approximation set of Γ after k iterations)

- **Theorem** At any instance k of the algorithm described by the space X, the cone K and the functional f, if s_{AC} is the exact AC:

  $$f^*(s_{AC}) \leq \sum_{i=0}^{k} \theta_{f_i} \ln \frac{\theta_{f_i}}{\varepsilon_i}$$

  where $\varepsilon_i = \varepsilon\sqrt{\lambda_{\min}(A_i^* A_i)}$ for $i \geq 1$ and $\varepsilon_0 = \varepsilon\sqrt{2}$.

- Γ contains a ball of radius ε.

  $$\mathcal{M} := \{y \in Y : y \in \Omega_k, B_Y(y, \varepsilon) \subset \Omega_k\} \neq \emptyset.$$

  Then, (x^k, y^k, s^k) - AC

  $$f^*(s^k) \leq f^*(s), \forall s \in \mathcal{M}_s := \{s : s = c - A^* y \text{ with } y \in \mathcal{M}\}.$$

  Lemma Let s be an arbitrary point in the set $\mathcal{M}_s$, with $s_i \in K_i$, the corresponding components.

  $$d(s_i, \partial K_i) \geq \varepsilon\sqrt{\lambda_{\min}(A_i^* A_i)}.$$

The result

Theorem 1 *The algorithm stops with a solution as soon as:*

$$(\sum_{l=1}^{k} \theta_{f_l}) \ln H \frac{(2m \ln \frac{1}{2m}(tr(B_0)+\sum_{i=1}^{k} p_i)-\ln(\det B_0))}{\sum_{i=1}^{k} p_i} \leq 2f_0^*(s_0) - 2\theta_{f_0} \ln \frac{\theta_{f_0}}{\varepsilon\sqrt{2}}.$$

with $H = Const.\frac{\Theta^3 P^2}{\varepsilon^2 \Lambda^2}$, $\Lambda := \min_{i=1,\ldots,k} \sqrt{\lambda_{\min}(A_i^* A_i)}$.

The number of cuts added is at most $O^*(\frac{mP^3\Theta^3}{\varepsilon^2\Lambda^2})$ *(here* O^* *means that terms of low order are ignored).*

Compare to:

- Complexity of the analytic center cutting plane method with multiple cuts for LP was done by Ye. The solution to the feasibility problem can be obtained in no more than $O^*(\frac{m^2P^2}{\varepsilon^2})$ iterations.

- The semidefinite programming case is treated by by Oskoorouchi and Goffin. The bound on the total size of the relaxation before obtaining the solution is $O(\frac{m^3P^2}{\varepsilon^2\mu^2})$ when a single cut of size at most P is added at each iteration.

- The second order cone case has also been analyzed by Oskoorouchi and Goffin. The total number of cuts required to obtain the solution is $O(\frac{m}{\varepsilon^2\mu})$ when a single cut of size at most P is added at each iteration.

Numerical results - feasibility/optimization

Problems from

- "The Seventh DIMACS Implementation Challenge Problems"
- subproblems from sparse "Partial Least Squares"

Problem	optimality (sedumi)	feasibility (ACCPM)	Size
[100 x 3]	1.17	0.56	107 x 500
[100 x 5]	0.89	1.23	111 x 700
[100 x 25]	0.58	0.61	151 x 2700
[500 x 3]	1.38	1.31	507 x 2500
[500 x 5]	1.55	0.78	511 x 3500
[500 x 25]	3.63	0.88	551 x 13500
[1000 x 3]	2.05	1.56	1007 x 5000
[1000 x 5]	2.48	1.29	1011 x7000
[1000 x 25]	6.73	1.29	1051 x 27000
[5000 x 3]	7.73	3.52	5007 x 25000
[5000 x 5]	11.75	3.69	5011 x 35000
[5000 x 25]	58.02	3.52	5051 x 135000
[10000 x 3]	18.95	5.89	10007 x 50000
[10000 x 5]	26.92	5.84	10011 x 70000
[100,000 x 3]	889.48	151.34	100007 x 500000
[300,000 x 3]	8278.10	269.37	300007 x 1,500,000
nql30	1.688	3.2	3680 x 6302
nql60	7.03	14.46	14,560 x 25,202
nql180	246.87	243.48	130,080 x 226,802
nb	4.516	1.57	123 x 2,383
nb-L1	5.015	34.95	915 x 3,176

Comments

Factors that influence performance

- the oracle
- the size of initial set Ω_0 versus the thickness of the set
- number of cuts added
- use of weights

Other comments:

- feasibility problems: $5\% - 100\%$ cuts used
- optimality problems: $37\% - 91\%$ cuts used
- second type of Newton steps - never used
- frequency for the used cuts : $1 - 66$(feasibility) $8 - 57$ (optimality)
- Feasibility algorithm performs better when problems are very large

An algorithm for optimization problems

Input
  setup the initial set Ω_0
  initialize the point (x, y, s)
  get the θ - analytic center for Ω_0
  compute the duality gap
  get the total violation
while $|\text{duality gap}| \geq \tau_{dg}$ or total violation $> \tau_{feas}$
  call the oracle at (x, y, s)
  if the point is feasible in Γ
    decrease barrier parameter μ $(\mu := (1 - \Theta)\mu)$
    get the new θ - analytic center
    compute the duality gap and total violation
  else the point is outside Γ
    add the cuts, generate a new outer-approximation set for Γ
    get the new θ - analytic center
    compute the duality gap and total violation
  return
STOP

Numerical results - optimization problems

Name	Ω_0	AC's	Newton	Cuts(Freq.)	%	Time
nql30	10	326	858	2791 (20)	0.61	494.81
nql30	5000	149	275	3975 (20)	0.88	262.05
nql60	10	424	889	16830 (57)	0.93	3150.48
nql60	5000	318	654	16454 (22)	0.91	2904.20
nb	10	561	646	252 (11)	0.31	248.07
nb	500	604	1063	258 (12)	0.32	376.86
nb-L1	10	217	271	794 (20)	0.49	1331.42
nb-L1	500	322	639	1557 (51)	0.97	2562.84

Size	AC's	Newton	Cuts (Freq.)	%	Time
$[100 \times 3]$	132	293	171 (8)	0.85	29.43
$[100 \times 5]$	165	563	166 (12)	0.83	51.31
$[100 \times 25]$	183	504	158 (10)	0.79	185.17
$[500 \times 3]$	190	587	799 (13)	0.79	165.92
$[500 \times 5]$	185	438	871 (11)	0.87	227.71
$[500 \times 25]$	195	615	746 (9)	0.74	1396.12
$[1000 \times 3]$	159	395	1296 (9)	0.64	218.26
$[1000 \times 5]$	170	442	1381 (9)	0.69	352.26
$[1000 \times 25]$	226	746	1489 (13)	0.74	4039.45
$[5000 \times 3]$	176	462	6576 (9)	0.65	948.70
$[5000 \times 5]$	195	582	6936 (9)	0.69	1379.97
$[10000 \times 3]$	203	462	13005 (9)	0.65	2653.94
$[10000 \times 5]$	200	582	13746 (9)	0.68	3807.11

Future work

- extend the optimization algorithm to the general case
- analyze particular cases SOCP and SDP
- implement the algorithm for the LP + SOCP + SDP
- find better ways to estimate the size of the Ω_0
- find a strategy for decreasing μ
- how to deal with non-fully dimensional cases