Selective Gram-Schmidt orthonormalization for conic cutting surface algorithms

John E. Mitchell   Vasile L. Basescu

Department of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, NY 12180 USA

ICCOPT II / MOPTA 07
August 13-16, 2007
McMaster University
Hamilton, Ontario, Canada

Summary

- Want to solve a convex feasibility problem.
- Cutting surface algorithm adds conic cuts such as SDP cuts or SOCP cuts, or linear cuts.
- With SDP or SOCP cuts, need to solve an NLP to find a new feasible interior point.
- With SDP and SOCP, complexity to find a solution to the feasibility problem depends on a condition number, which is a property of the added constraints.

Benefits of selective orthonormalization

Selective orthonormalization is a method for modifying the cuts that

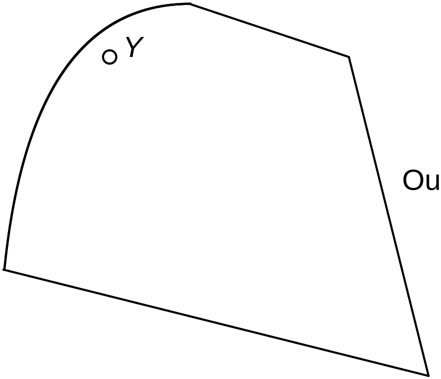
1. Makes it easy to restart
2. Removes the need for a condition number in the global convergence analysis.

Convex feasibility problem

Given a set Y, find a point $y \in Y$ or determine that Y is empty.
Assumptions:

- Y is a convex, bounded set contained in $\mathbb{R}^m$, containing a ball $B(., \epsilon)$ of radius ϵ.
- If $\bar{y} \notin Y$, a separation oracle returns a conic inequality $\mathcal{G}^* y + s = h$, $s \in K_0$ satisfied by all $y \in Y$ and violated by $\bar{y}$. K_0 is a full-dimensional self-scaled cone in R^p. $\mathcal{G}^*$ is surjective.
- The cone K_0 has a self-concordant barrier function $f_0(K_0)$.

Convex feasibility problem



Outer approximation of Y

Conic relaxation

The feasibility problem is approximated by a conic program:

$$\begin{array}{llcccll} \max & 0 & & & & & \\ \text{subject to} & \mathcal{A}^* y & + & s & = & c & \qquad (CD) \\ & & & s & \in & K & \end{array}$$

where K is a full-dimensional self-scaled cone.

Note that K may be a product of smaller cones.
Eg, $K = R^n_+$, $K =$ SDP cone, $K =$ product of SOCP cones, ...

Duality

Call (CD) the dual problem. The corresponding primal problem is:

$$\begin{array}{lrcl} \min & & < c, x > & \\ \text{subject to} & \mathcal{A}x & = & 0 \\ & x & \in & K \end{array} \qquad (CP)$$

Note that we assume that K is self-dual.

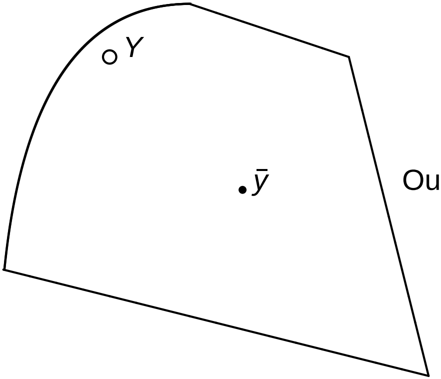
Cutting planes

- We find an approximate analytic center $(\bar{x}, \bar{y}, \bar{s})$ for (CP) and (CD).
- If $\bar{y} \in Y$, DONE.
- Otherwise, the oracle returns a cut violated by $\bar{y}$ and satisfied by all $y \in Y$:

$$\begin{array}{rcrcl} \mathcal{G}^* y & + & s & = & h \\ & & s & \in & K_0 \end{array}$$

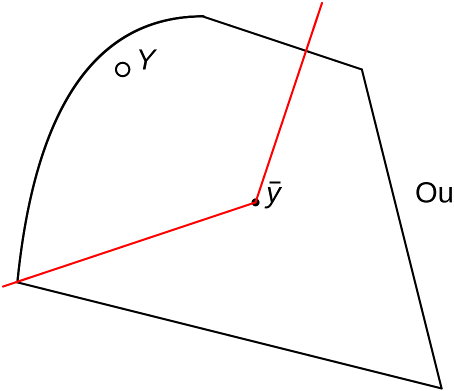
- We assume the cut is shifted to be central, so $h = \mathcal{G}^* \bar{y}$.

Find approximate analytic center $\bar{y}$



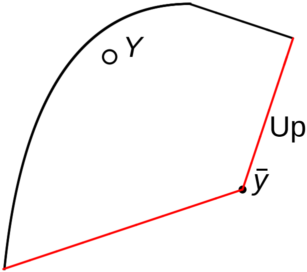
Outer approximation of Y

Add a conic cut at $\bar{y}$



Outer approximation of Y

Update outer approximation



Updated outer approximation of Y

Modifying (CP) and (CD)

$$
\begin{array}{llcllcll}
\max & & 0 & & & & & \\
\text{subject to} & \mathcal{A}^* y & + & s & = & c & & (\overline{CD}) \\
& \mathcal{G}^* y & + & s_0 & = & h & & \\
& s \in K, & & s_0 & \in & K_0 & &
\end{array}
$$

$$
\begin{array}{llcllcll}
\min & < c, x > & + & < h, x_0 > & & & & \\
\text{subject to} & \mathcal{A} x & + & \mathcal{G} x_0 & = & 0 & & (\overline{CP}) \\
& x \in K, & & x_0 & \in & K_0 & &
\end{array}
$$

Finding a new approximate analytic center

- With $x_0 = 0$ and $s_0 = 0$, the current point $(\bar{x}, \bar{y}, \bar{s})$ is on the boundary of the feasible regions of $(\overline{CD})$ and $(\overline{CP})$.
- Can solve an NLP in the new variables in order to find a restart direction.

Potential functions

- The primal-dual potential function is

$$\Phi_{PD} \;=\; < x, s > + f(x) + f^*(s).$$

  It is minimized with value 0 at the analytic center.

- Let ϑ_f denote the complexity value of f, in the terminology of Renegar's text.

  LP: $f(x) = -\sum_{i=1}^n \ln(x_i)$, $\vartheta_f = n$.
  SDP: $f(X) = -\ln\det(X)$, $\vartheta_f = n$.
  SOCP: $f(\xi, x) = -0.5\ln(\xi^2 - \sum_{i=1}^n x_i^2)$, $\vartheta_f = 2$.

Convergence

- Local convergence:
  Get convergence to new approximate analytic center in $O(\vartheta_{f_0} \ln \vartheta_{f_0})$ steps.
- Global convergence:
  Polynomial in m, the required tolerance ϵ,
  and a condition number based on the added constraints.
  Proof uses upper and lower bounds on the dual potential function.

The need for a condition number

- Upper bound obtained from assumption of an ϵ-ball:
  If $c - \mathcal{A}^*(y + \epsilon u) \succeq_K 0$ for any unit vector u then
  $f_K^*(c - \mathcal{A}^* y) \leq f_K^*(\epsilon \mathcal{A}^* u) = f_K^*(\mathcal{A}^* u) - \vartheta_{f_K} \ln \epsilon$.
- Want $f_K^*(\mathcal{A}^* u)$ small in order to get a good upper bound.
- Define condition number μ_K so that
  $\ln \mu_K := \inf\{f_K^*(\mathcal{A}^* u) : \mathcal{A}^* u \succ_K 0,\ ||u|| = 1\}$.

An SDP constraint with a bad condition number

- Look at $\{u \in \mathbb{R}^4 : \mathcal{A}^*(u) := \sum_{i=1}^4 A_i u_i \succeq 0\}$ for

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & \epsilon \end{bmatrix} A_2 = \begin{bmatrix} 0 & 1 \\ 1 & \epsilon \end{bmatrix} A_3 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} A_4 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

- If $||u|| = 1$ then $\det(\mathcal{A}^*(u))$ is at most $O(\epsilon)$.
- Thus, for any choice of direction, need to move quite far to get a reasonable potential function value.

Selective Orthonormalization

Selective Orthonormalization is a method for modifying the added cuts so that

- It is easy to restart
- The complexity does not depend on a condition number
- Drawback: It may weaken the cuts.

Modifying a single cut

Given a central constraint $c - \mathcal{A}^* y \succeq_K 0$:

- Want a direction d such that $\mathcal{A}^* d \prec_K 0$.
- Let e be in interior of K. Can assume $\mathcal{A} e \neq 0$ (else no such d exists).
- Let $\bar{\mathcal{A}} := \mathcal{A} + \lambda \mathcal{A} e e^*$ with $\lambda \geq 0$.
- Then $\bar{\mathcal{A}}^* d \preceq_K \mathcal{A}^* d$ for all d satisfying $\mathcal{A}^* d \preceq_K 0$.

  So $c - \bar{\mathcal{A}}^* y \succeq_K 0$ is a valid constraint.

Example

- Use earlier SDP example with $\mathcal{A}$ defined by

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & \epsilon \end{bmatrix} A_2 = \begin{bmatrix} 0 & 1 \\ 1 & \epsilon \end{bmatrix} A_3 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} A_4 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

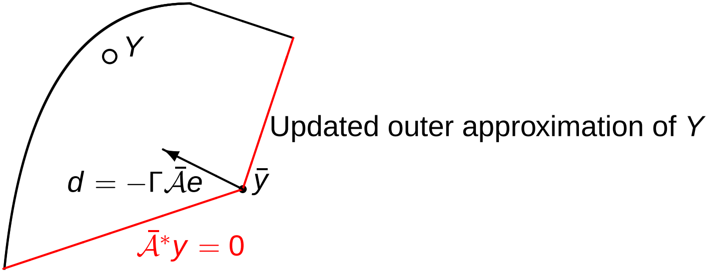
- Take e to be the 2×2 identity matrix.
- Then the modification results in adding terms to the diagonals of the matrices, so (ignoring normalization):

$$A_1 \approx \begin{bmatrix} \frac{7}{6} & 0 \\ 0 & \frac{1}{6} \end{bmatrix} A_2 \approx \begin{bmatrix} 0 & 1 \\ 1 & \epsilon \end{bmatrix} A_3 \approx \begin{bmatrix} \frac{7}{6} & 0 \\ 0 & \frac{1}{6} \end{bmatrix} A_4 \approx \begin{bmatrix} \frac{7}{6} & 1 \\ 1 & \frac{1}{6} \end{bmatrix}$$

- Now there are vectors u with $||u|| = O(1)$ and $\det(\mathcal{A}^*(u)) = O(1)$.

Choosing λ for strict feasibility

Want to move in the direction $d = -\Gamma \bar{\mathcal{A}} e$, where $\Gamma := (AH(\bar{s})A^*)^{-1}$. Need $\bar{\mathcal{A}}^* d \prec_K 0$.



Chossing λ to control the condition number

- To control the condition number, update $\bar{\mathcal{A}}$ so that $\bar{\mathcal{A}}^* \bar{\mathcal{A}} e \succeq_K \omega e$, where ω is chosen to be a scalar between 0 and 1.
- Then $\ln \mu_K \leq -\vartheta_{f_K}(1 + \ln(\omega))$

  and $f_K^*(c - \mathcal{A}^* y) \leq -\vartheta_{f_K}(1 + \ln(\epsilon\omega))$ if $B(y, \epsilon) \subseteq K$.

  (Assume $\mathcal{A}$ and e each have norm 1, and $f_K(e) = 0$, so $f_K^*(e) = -\vartheta_{f_K}$.)

  Thus, we have a reasonable upper bound on the potential function.

Technicalities of picking λ

- If $\bar{\mathcal{A}}^*\bar{\mathcal{A}}e \not\succeq_K 0$, set $\bar{\lambda} := \min\{\lambda : \lambda e + \frac{1}{||\bar{\mathcal{A}}e||}\bar{\mathcal{A}}^*\bar{\mathcal{A}}e \succeq_K 0\}$. Update $\bar{\mathcal{A}} \leftarrow \bar{\mathcal{A}} + \bar{\lambda}\bar{\mathcal{A}}ee^*$ and renormalize so that $||\bar{\mathcal{A}}|| = 1$.
- If $\bar{\mathcal{A}}^*\bar{\mathcal{A}}e \not\succeq_K \omega e$, update $\bar{\mathcal{A}} \leftarrow (1-\sqrt{\omega})\bar{\mathcal{A}} + \frac{\sqrt{\omega}}{||\bar{\mathcal{A}}e||}\bar{\mathcal{A}}ee^*$. Renormalize so that $||\bar{\mathcal{A}}|| = 1$.
- Let $\eta = \sqrt{e^*\bar{\mathcal{A}}^*\Gamma\bar{\mathcal{A}}e}$. If $\bar{\mathcal{A}}^*\Gamma\bar{\mathcal{A}}e \not\succeq_K \nu\eta^2 e$, update $\bar{\mathcal{A}} \leftarrow (1-\nu)\bar{\mathcal{A}} + \nu\bar{\mathcal{A}}ee^*$. Renormalize so that $||\bar{\mathcal{A}}|| = 1$.

Multiple cuts

- Assume multiple central cuts are added simultaneously.
- Can extend Selective Orthonormalization to ensure that can still restart easily.
- Need to modify using terms of the form

$$\bar{\mathcal{A}}_p \leftarrow \bar{\mathcal{A}}_p + \lambda \bar{\mathcal{A}}_q e_q e_p^*$$

  to ensure $\bar{\mathcal{A}}_p^* \Gamma \bar{\mathcal{A}}_q e_q \succeq_K 0$.
- Use direction

$$d = -\sum_{p=1}^{l} \frac{1}{\eta_p} \Gamma \bar{\mathcal{A}}_p e_p$$

Conclusions

- Given a convex feasibility problem with an oracle that returns conic cuts, the cutting surface method can determine feasibility in fully polynomial time.
- If selective orthonormalization is used then the complexity does not depend on a condition number, and the algorithm can be restarted easily.
- Open question: can the condition number be eliminated from the complexity without weakening the cuts?

References

V. L. Basescu and J. E. Mitchell.
An analytic center cutting plane approach for conic programming.
Technical report, Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, June 2005. Revised June 2007.

J. E. Mitchell and V. L. Basescu.
Selective Gram-Schmidt orthonormalization for conic cutting surface algorithms.
November 2006. Accepted for publication in *Mathematical Methods of Operations Research*.

J. E. Mitchell and S. Ramaswamy.
Using selective orthonormalization to update the analytic center after the addition of multiple cuts.
Journal of Optimization Theory and Applications, 125(2):431–451, 2005.