A solution approach for linear optimization with completely positive matrices

Franz Rendl

http://www.math.uni-klu.ac.at

Alpen-Adria-Universität Klagenfurt

Austria

joint work with M. Bomze (Wien) and F. Jarre (Düsseldorf)

Overview

- COP and CP
- Inner approximations with COP matrices
- A Random COP generator
- Computations

Completely Positive Matrices

Let $A = (a_1, \ldots, a_k)$ be a nonnegative $n \times k$ matrix, then

$$X = a_1 a_1^T + \ldots + a_k a_k^T = AA^T$$

is called completely positive.

$$COP = \{X : X \text{ completely positive}\}$$

COP is closed, convex cone

- $X \in COP,\ t \geq 0 \Rightarrow tX \in COP$
- $X = AA^T,\ Y = BB^T \in COP \Rightarrow$
  $\Rightarrow X + Y = [A\ B][A\ B]^T \in COP$

For basics, see the book: A. Berman and N. Shaked - Monderer: Completely Positive Matrices, World Scientific 2003

Copositive Matrices

Dual cone COP^* of COP in S_n (sym. matrices):

$$Y \in COP^* \iff \text{tr}XY \geq 0 \ \ \forall X \in COP$$

$$\iff \text{tr}A^T Y A \geq 0 \ \ \forall A \geq 0$$

$$\iff a^T Y a \geq 0 \ \ \forall \text{ vectors } a \geq 0.$$

By definition, this means Y is copositive, $Y \in CP$.
CP is dual cone to COP!
Bad News: $X \notin CP$ is NP-complete decision problem.

Positive semidefinite matrices PSD:
$Y \in PSD \iff a^T Y a \geq 0$ $\forall$ vectors a.
Well known facts: • $PSD^* = PSD$ (PSD cone is selfdual.)
• $COP \subset PSD \subset CP$

Semidefinite and Copositive Programs

Problems of the form

$$\max \text{tr}CX \text{ s.t. } A(X) = b,\ X \in PSD$$

are called Semidefinite Programs.

Problems of the form

$$\max \text{tr}CX \text{ s.t. } A(X) = b,\ X \in CP$$

or

$$\max \text{tr}CX \text{ s.t. } A(X) = b,\ X \in COP$$

are called Copositive Programs, because the primal or the dual involves copositive matrices.

Why Copositive Programs ?

Copositive Programs can be used to solve combinatorial optimization problems.

• Stable Set Problem:
Let A be adjacency matrix of graph, J be all ones matrix.
DeKlerk and Pasechnik (SIOPT 2002) show the following:

$$\alpha(G) = \max \text{tr} JX \text{ s.t. } \text{tr}(A+I)X = 1, \quad X \in COP$$

$$= \min y \text{ s.t. } y(A+I) - J \in CP.$$

This is a copositive program with only one equation (in the primal problem).
This is a simple consequence of the Motzkin-Strauss Theorem.

A general copositive modeling theorem

Burer (2007) shows the following general result for the power of copositive programming:
The optimal values of P and C are equal: opt(P) = opt(C)

$$(P) \qquad \min x^T Q x + c^T x$$

$$a_i^T x = b_i,\ x \geq 0,\ x_i \in \{0,1\} \forall i \in B.$$

Here $x \in \mathbb{R}^n$ and $B \subseteq \{1, \ldots, n\}$.

$$(C) \qquad \min \text{tr}(QX) + c^T x,\ \text{s.t. } a_i^T x = b_i,$$

$$a_i^T X a_i = b_i^2,\ X_{ii} = x_i\ \forall i \in B,\ \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \in COP$$

Approximating COP

We have now seen the power of copositive programming.

Since optimizing over CP is NP-Hard, it makes sense to get approximations of CP or COP.

To get relaxations, we need supersets of COP, or inner conic approximations of CP (and work on the dual cone). The Parrilo hierarchy uses Sum of Squares and provides such an outer approximation of COP (dual viewpiont!).

Now we consider inner approximations of COP.
This can be viewed as a method to generate feasible solutions of combinatorial optimization problems (primal heuristic!).

Inner approximations of COP

We consider

$$\min \text{tr} CX \text{ s.t. } A(X) = b, \; X \in COP$$

Remember: COP= $\{X : X = VV^T, \; V \geq 0\}$.

Some previous work by:

- Bomze, DeKlerk, Nesterov, Pasechnik, others:

Get stable sets by approximating COP formulation of the stable set problem using optimization of quadratic over standard simplex, or other local methods.

Incremental version

A general feasible descent approach:

Let $X = VV^T$ with $V \geq 0$ be feasible. Consider the regularized, and convex descent step problem:

$$\min \epsilon \langle C, \Delta X \rangle + (1-\epsilon) \|\Delta X\|^2,$$

$$\text{such that } A(\Delta X) = 0, \quad X + \Delta X \in COP.$$

Incremental version

A general feasible descent approach:
Let $X = VV^T$ with $V \geq 0$ be feasible. Consider the regularized, and convex descent step problem:

$$\min \epsilon\langle C, \Delta X\rangle + (1-\epsilon)\|\Delta X\|^2,$$

such that $A(\Delta X) = 0, \quad X + \Delta X \in COP.$

For small $\epsilon > 0$ we approach the true optimal solution, because we follow the continuous steepest descent path, projected onto COP.

Unfortunately, this problem is still not tractable. We approximate it by working in the V-space instead of the X-space.

Incremental version: V-space

$$X^+ = (V + \Delta V)(V + \Delta V)^T \text{ hence },$$

$$X = VV^T$$

$$\Delta X = \Delta X(\Delta V) = V\Delta V^T + \Delta V V^T + (\Delta V)(\Delta V^T).$$

Now linearize and make sure ΔV is small.

Incremental version: V-space

$$X^+ = (V + \Delta V)(V + \Delta V)^T \text{ hence },$$

$$X = VV^T$$

$$\Delta X = \Delta X(\Delta V) = V\Delta V^T + \Delta V V^T + (\Delta V)(\Delta V^T).$$

Now linearize and make sure ΔV is small. We get

$$\min \epsilon \langle 2CV, \Delta V\rangle + (1-\epsilon)\|\Delta V\|^2 \text{ such that}$$

$$\langle 2A_i, \Delta V\rangle = b_i - \langle A_i V, V\rangle \ \ \forall i,$$

$$V + \Delta V \geq 0$$

This is convex approximation of nonconvex version in $\Delta X(\Delta V)$.

Correcting the linearization

The linearization obtained by replacing $\Delta X(\Delta V)$ with $V\Delta V^T + \Delta V V^T$ introduces an error both in the cost function and in the constraints $A(X) = b$.

We therefore include corrector iterations of the form

$$\Delta V = \Delta V_{old} + \Delta V_{corr}$$

before actually updating $V \leftarrow V + \Delta V$.

Convex quadratic subproblem

The convex subproblem is of the following form, after appropriate redefinition of data and variables $x = vec(V + \Delta V), \ldots$.

$$\min c^T x + \rho x^T x \text{ such that } Rx = r, \; x \geq 0.$$

Convex quadratic subproblem

The convex subproblem is of the following form, after appropriate redefinition of data and variables $x = vec(V + \Delta V), \ldots$.

$$\min c^T x + \rho x^T x \text{ such that } Rx = r, \; x \geq 0.$$

If V is $n \times k$ (k columns generating V), x if of dimension nk and there are m equations.

Convex quadratic subproblem

The convex subproblem is of the following form, after appropriate redefinition of data and variables $x = vec(V + \Delta V), \ldots$.

$$\min c^T x + \rho x^T x \text{ such that } Rx = r,\ x \geq 0.$$

If V is $n \times k$ (k columns generating V), x if of dimension nk and there are m equations.
Since Hessian of cost function is identity matrix, this problem can be solved efficiently using interior-point methods. (convex quadratic with sign constraints and linear equations)

Main effort is solving a linear system of order m, essentially independent of n and k.

Test data sets

COP problems coming from formulations of NP-hard problems are too difficult (Stable Set, Coloring, Quadratic Assignment) to test new algorithms.

Would like to have:

Data (A, b, C) and (X, y, Z) such that
• (X, y, Z) is optimal for primal and dual (no duality gap).
• COP is nontrivial (optimum not given by optimizing over semidefiniteness plus nonnegativity)
• generate instances of varying size both in n and m.

Hard part: Z provably copositive !!

A COP generator

Basic idea: Let G be a graph, with $Q_G = A + I$, and (X_G, y_G, Z_G) be optimal solution of

$$\omega(G) = \max\langle J, X\rangle \text{ such that } \langle Q_G, X\rangle = 1,\ X \in COP$$

$$= \min y \text{ such that } yQ_G - J \in CP.$$

We further assume:

- $G = H * K$ (strong graph product of H and K).
- K is perfect and $H = C_5$ (or some other graph with known clique number and $\omega(G) < \vartheta'(G)$).

This implies:

- $\omega(G) = \omega(H)\omega(K)$ (same for ϑ')
- Max cliques in G through Kronecker products from H, K.

COP generator (2)

Then we have:

- $y_G := \omega(H)\omega(K)$ implies $Z_G = y_G Q_G - J$ is CP.
- Let X_H and X_K convex combinations of outer products of characteristic vectors of max cliques in H and K. This implies that $X_G = X_H \otimes X_K$ is primal optimal.
- Select m and Matrix A of size $m \times n^2$ and set $b = A(X_G)$.
- Set $C := Z_G - A^T(y_G)$.

This implies that (X_G, y_G, Z_G) is optimal for the data set A, b, C and that there is no duality gap, and the problem is nontrivial, because of $\omega(G) < \vartheta'(G)$.

Computational results

A sample instance with $n = 60, \ m = 100$.

$z_{sdp} = -9600, 82, \ \ z_{sdp+nonneg} = -172.19, z_{cop} = -69.75$

it	secs	|b-A(X)|	f(x)
1	5.17	0.166918	-54.5084
10	33.45	0.000028	-69.4370
20	64.94	0.000003	-69.5439
30	96.49	0.000002	-69.6109

The number of inner iterations was set to 3, column 1 shows the outer iteration count.

Computational results

A sample instance with $n = 60,\ m = 100$.

$z_{sdp} = -9600, 82,\ \ z_{sdp+nonneg} = -172.19,$ $z_{cop} = -69.75$

it	secs	|b-A(X)|	f(x)
1	5.17	0.166918	-54.5084
10	33.45	0.000028	-69.4370
20	64.94	0.000003	-69.5439
30	96.49	0.000002	-69.6109

The number of inner iterations was set to 3, column 1 shows the outer iteration count.

But starting point: $V0 = .9$ Vopt $+ .1$ rand

Computational results (2)

Example (continued). recall $n = 60,\ m = 100$.

$z_{sdp} = -9600, 82, \quad z_{sdp+nonneg} = -172.19, z_{cop} = -69.75$

start	iter	|b-A(X)|	f(x)
(a)	30	0.000000	-69.67
(b)	30	0.000002	-69.61
(c)	30	0.000100	-69.03

Different starting points:

(a) V = ,95 * V_{opt} + .05 * rand

(b) V = .90 * V_{opt} + .10 * rand

(c) V = rand(n, 2n)

Similar results for other instances using the COP generator.

More results

n	m	opt	found	$\|b - A(X)\|$
50	100	314.48	315.18	4 10^{-5}
60	120	-266.99	-266.31	4 10^{-5}
70	140	-158.74	-155.78	3 10^{-5}

Starting point in all cases: rand(n,2n)
Inner iterations: 5
Outer iterations: 25

Some experiments with Stable Set

$$\max\langle J, X\rangle \text{ such that } \mathrm{tr}(X) = 1,\ \mathrm{tr}(A_G X) = 0,\ X \in COP$$

Only two equations but many local optima. We consider a selection of graphs from the DIMACS collection. Computation times in the order of a few minutes.

name	n	ω	clique found
keller4	171	11	9
brock200-4	200	17	14
c-fat200-1	200	12	12
c-fat200-5	200	58	58
brock400-1	400	27	24
p-hat500-1	500	9	8

Last Slide

Unfortunately, the subproblem may have local solutions, which are not local minima for the original descent step problem.

The number of columns of V does not need to be larger than $\binom{n+1}{2}$, but for practical purposes, this is too large.

Also, there is dependence on the starting point V.

Further technical details in a forthcoming paper by I. Bomze, F. Jarre and F. R.: Quadratic factorization heuristics for copositive programming, technical report, (2008).