Local Minima and Convergence in Low-Rank Semidefinite Programming

Sam Burer (University of Iowa)

Renato Monteiro (Georgia Tech)

Workshop on Large Scale

Nonlinear and Semidefinite Programming

University of Waterloo

May 13, 2004

OUTLINE

- Description of problem
- Motivation of algorithm
- Low-rank algorithm #1 (without convergence proof)
- Low-rank algorithm #2 (with convergence proof)
- Computational results
- Conclusions

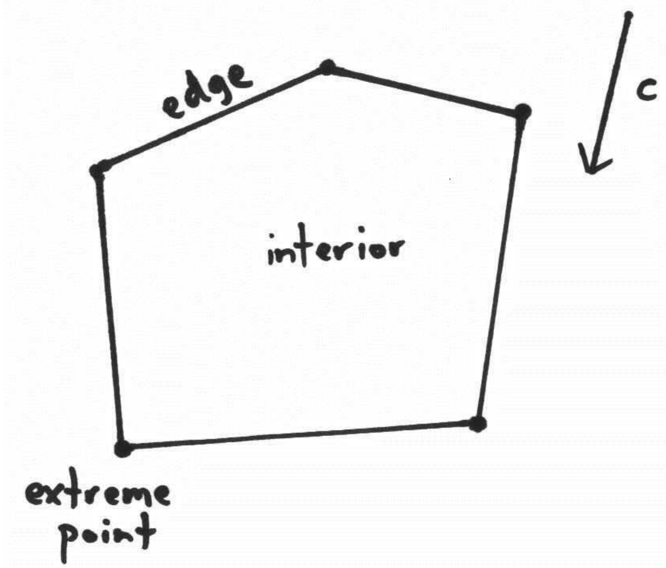
DESCRIPTION OF PROBLEM

$$\text{(SDP)} \qquad \min\{C \bullet X \;:\; A_i \bullet X = b_i, \quad i = 1, \ldots, m, \quad X \succeq 0\}$$

- SDP is a convex optimization problem.
- Although polynomial, interior-point methods can be computationally inefficient when the data is dense, n is large, and / or m is large.
- Do other alternatives exist?
- First-order nonlinear programming methods for SDP have been successful in solving some key classes of large-scale SDPs.
  - Spectral bundle method of Helmberg and Rendl;
    parallel spectral bundle method of Nayakkankuppam and Tymofyeyev;
    bundle method of Fischer, Gruber, Rendl, Sotirov.

Motivation of Algorithm

(LP) $$\min\left\{c^T x \;:\; a_i^T x = b_i, \quad i = 1, \ldots, m, \quad x \geq 0\right\}$$



nz $:=$ # of nonzero entries of x

x extreme point $\Longrightarrow$ nz $\leq m$

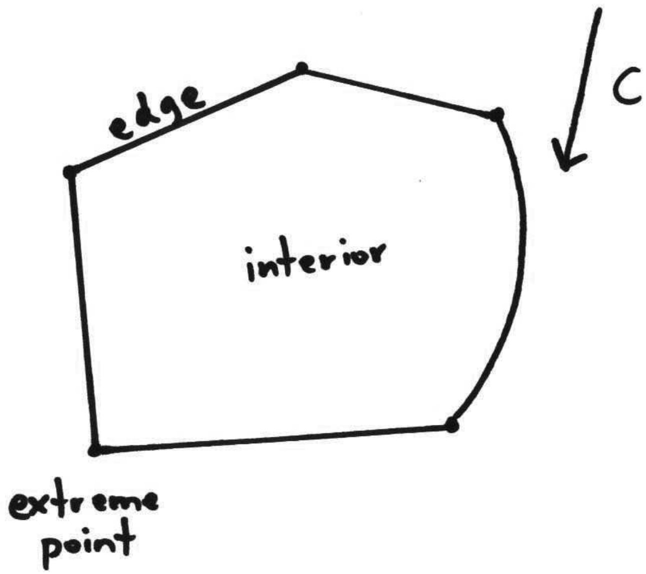
x edge point $\Longrightarrow$ nz $\leq m + 1$

x interior point $\Longrightarrow$ nz $= n$

Interior-point methods for LP are well developed.

Motivation of Algorithm (cont'd)

(SDP) $\quad \min \{C \bullet X \;:\; A_i \bullet X = b_i, \quad i = 1, \ldots, m, \quad X \succeq 0\}$



$r :=$ rank of X

X extreme point $\Longrightarrow r(r+1)/2 \leq m$ (Pataki)

X edge point $\Longrightarrow r(r+1)/2 \leq m+1$ (Pataki)

X interior point $\Longrightarrow r = n$

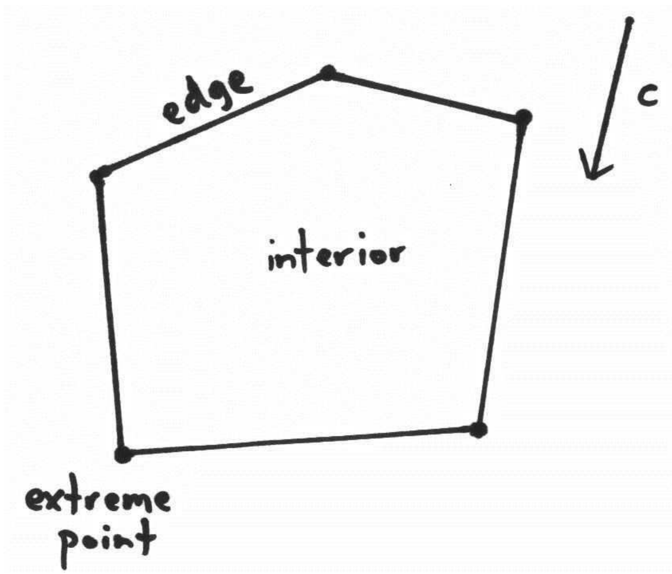
Interior-point methods for SDP are well developed.

Motivation of Algorithm (cont'd)

Simplex method for LP. From current extreme point, move along an improving edge to a new extreme point. If no improving edge exists, current extreme point is optimal.

Another perspective: optimizes over low-dimensional faces. . .

$$\min \left\{ c^T x \; : \; a_i^T x = b_i, \quad i = 1, \ldots, m, \quad x \geq 0, \quad \text{nz} \leq m + 1 \right\}$$



Motivation of Algorithm (cont'd)

$$\text{(LP}'\text{)} \qquad \min\left\{c^T x \,:\, a_i^T x = b_i, \quad i = 1, \ldots, m, \quad x \geq 0, \quad \text{nz} \leq m + 1\right\}$$

Theorem (Simplex Method). Let x be a local min of (LP$'$). If x is an extreme point, then x is optimal. Otherwise, x is in the relative interior of a face which is "flat" with respect to the objective function.

As an algorithm for (LP$'$), the simplex method:

1. keeps all iterates feasible;
2. avoids getting trapped in flat faces.

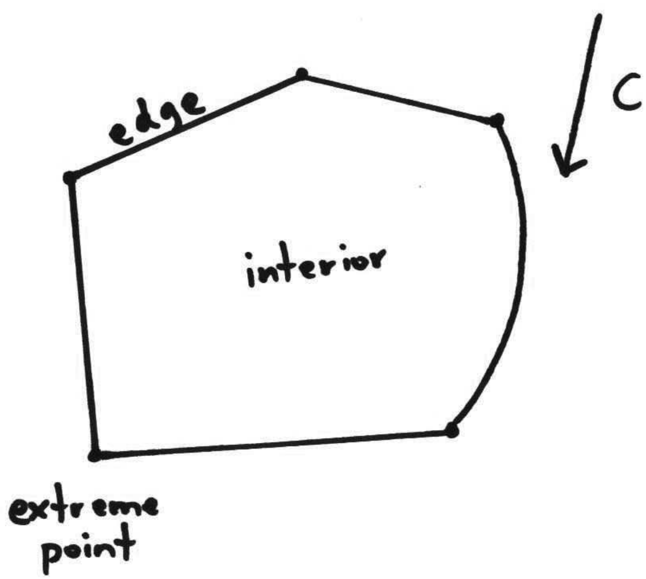
Motivation of Algorithm (cont'd)

Simplex method for SDP? Research ongoing... Krishnan-Mitchell, Goldfarb, Krishnan-Pataki-Zhang

Even still, can we optimize over low-dimensional faces?

Let $\bar{r}$ be the smallest integer such that $\bar{r}(\bar{r}+1)/2 \geq m+1$...

$$\min \{ C \bullet X \ : \ A_i \bullet X = b_i, \quad i = 1, \ldots, m, \quad X \succeq 0, \quad r \leq \bar{r} \}$$



Motivation of Algorithm (cont'd)

(SDP$'$) $\min\{C \bullet X \;:\; A_i \bullet X = b_i, \;\; i = 1, \ldots, m, \;\; X \succeq 0, \;\; r \leq \bar{r}\}$

Theorem (BM). Let X be a local min of (SDP$'$). If X is an extreme point, then X is optimal. Otherwise, X is in the relative interior of a face which is "flat" with respect to the objective function.

We propose an algorithm for (SDP$'$) that:

1. is based on nonlinear programming (in particular, is an infeasible method);
2. avoids getting trapped in flat faces;
3. has certain computational advantages over interior-point methods.

Low-Rank Algorithm #1

$$\text{(SDP}'\text{)} \quad \min\{C \bullet X \ : \ A_i \bullet X = b_i, \ \ i = 1, \ldots, m, \ \ X \succeq 0, \ \ r \leq \bar{r}\}$$

$$\longrightarrow \quad X \succeq 0, \ r \leq \bar{r} \quad \Longleftrightarrow \quad X = RR^T \text{ for some } R \in \Re^{n \times \bar{r}} \quad \longleftarrow$$

In 2001, we proposed the following nonlinear reformulation of (SDP′):

$$\text{(nSDP}'\text{)} \quad \min\left\{C \bullet RR^T \ : \ A_i \bullet RR^T = b_i, \ \ i = 1, \ldots, m, \ \ R \in \Re^{n \times \bar{r}}\right\}$$

Obs. Because $\bar{r} \approx \sqrt{2m}$, (nSDP′) has fewer variables than (SDP′), especially when m is small.

Low-Rank Algorithm #1 (cont'd)

$$(\text{nSDP}') \quad \min \left\{ C \bullet RR^T \ : \ A_i \bullet RR^T = b_i, \quad i = 1, \ldots, m, \quad R \in \Re^{n \times \bar{r}} \right\}$$

We used a first-order augmented Lagrangian method to "solve" (nSDP$'$)...

Introducing Lagrange multipliers y_i and a penalty parameter σ, R^k is a stationary point of

$$\mathcal{L}_k(R) := C \bullet RR^T + \sum_{i=1}^{m} y_i^k (b_i - A_i \bullet RR^T) + \frac{\sigma_k}{2} \sum_{i=1}^{m} (A_i \bullet RR^T - b_i)^2.$$

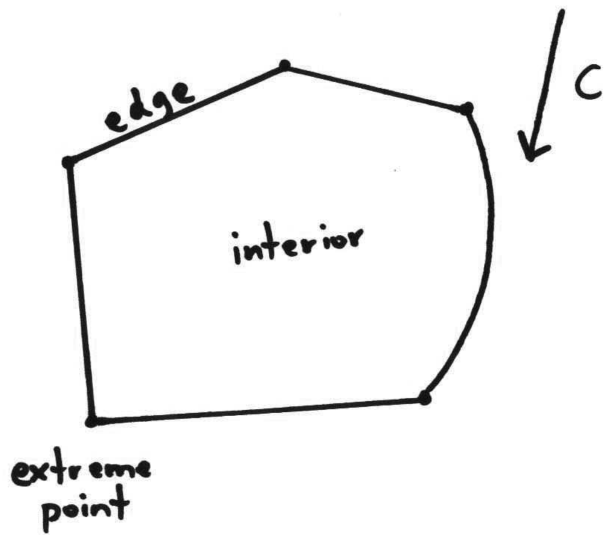
- In theory, $\{R^k\}$ converges to a stationary point of (nSDP$'$).
- But we always observed convergence to a global minimum.
- Moreover, the speed was very competitive on certain classes of problems.
- We wondered: could convergence be proved?

Low-Rank Algorithm #2

Problem. (SDP$'$) may have many local minima corresponding to flat faces.

How can we guarantee that $X^k := R^k(R^k)^T$ does not converge to a flat face?

Key Idea. For points $X = RR^T$ in an edge, $\det(R^T R)$ has no local minima except at the extreme points, for which $\det(R^T R) = 0$.



Algorithm. Choose positive $\mu_k \to 0$, and compute R^k as a local minimum of

$$\mathcal{L}_k(R) + \mu_k \det(R^T R).$$

Low-Rank Algorithm #2 (cont'd)

Theorem (BM). Suppose that each R^k is a local minimum of the k-th *perturbed* augmented Lagrangian subproblem. Suppose moreover that $\{R^k\}$ attains feasibility in the limit and that $\{R^k\}$ is bounded. Then

(a) every accumulation point of $\{R^k(R^k)^T\}$ is an optimal solution of SDP;

(b) the sequence $\{(y_1^k, \ldots, y_m^k)\}$ is bounded and any of its accumulation points is an optimal solution of the dual.

Remark. Feasibility can be guaranteed by taking σ_k sufficiently large.

Remark. In many problems, feasibility implies boundedness of $\{R^k\}$.

Computational Results

Note. We actually implement algorithm #1.

- Algorithm works well for sparse, large-scale SDPs, such as relaxations of combinatorial optimization problems (COPs).
- Since it is a “primal” algorithm, can it be used to bound COPs?
- Lagrange multipliers are optimal in the limit but infeasible along the way.
- However, for the special case of binary quadratic COPs, $(y_1^k, \ldots, y_m^k)$ can be easily shifted to dual feasibility, which yields a bound.

COMPUTATIONAL RESULTS (CONT'D)

- We bound the quadratic assignment problem via the “gangster” SDP relaxation (Zhao-Karisch-Rendl-Wolkowicz).
- Test problems are a representative sample from QAPLIB of all problems having between 30 and 40 facilities and locations.
- To our knowledge, only problems up to size 32 have been reported in the SDP literature (Lin-Saigal, Zhao-Karisch-Rendl-Wolkowicz, Rendl-Sotirov).
- All experiments done on a Pentium 2.4 GHz.

Conclusion 1. In comparison with other SDP algorithms, we receive comparable bound quality in much less time.

Conclusion 2. We are able to solve much larger instances.

Computational Results (cont'd)

problem	feasible val	n	m	lower bd	time (s)
esc32a	130	960	30721	-144	480
esc32h	438	960	30721	225	527
kra30a	*88900	840	25201	78255	58359
kra30b	*91420	840	25201	79165	48846
kra32	*88700	960	30721	76669	58103
lipa30a	*13178	840	25201	12934	2294
lipa30b	*151426	840	25201	151357	14862
lipa40a	*31538	1520	60801	30560	8753
lipa40b	*476581	1520	60801	476417	93621
nug30	*6124	840	25201	5629	2161
ste36a	*9526	1224	44065	7156	25703
ste36b	*15852	1224	44065	10350	552860
tai30a	1818146	840	25201	1577013	72911
tai35a	2422002	1155	40426	2029376	155143
tai40a	3139370	1520	60801	2592756	421348
tho30	*149936	840	25201	135535	81454
tho40	240516	1520	60801	214593	219336

CONCLUSIONS

- We have given a convergence proof for the low-rank SDP algorithm.
- The proof uses the geometry of SDP and the idea of a “reverse barrier.”
- Previous computational results have been extended:
  - dual bounds for combinatorial optimization problems;
  - application to large quadratic assignment problems.
- Extension to SDPs having lots of inequality constraints?
- More generally: extension to block SDPs?