Maximizing a Quadratic Form over the Intersection of Ellipsoids with Common Center

A. Nemirovski
Technion, Haifa, Israel

C. Roos, T. Terlaky
Delft University of Technology
e-mail: `C.Roos@its.tudelft.nl`
URL: `http://ssor.twi.tudelft.nl/~roos`

Workshop on

Novel Approaches to Hard Discrete Optimization

University of Waterloo

Waterloo, ON, CANADA

April 26-28, 2001

Outline of the talk

- The problem and its relaxation
- Related results
- A randomized rounding procedure
- Sharpness of the result
- Derandomization
- Extensions
- References

$$\max\left\{x^TAx \ : \ x^TA_ix \le 1,\, i=1,...,m\right\}. \qquad (P)$$

The $n \times n$ matrices A_i $(1 \le i \le m)$, are psd with pd sum.

This problem is NP-hard (take, e.g., $m = n$ and $A_i = e_ie_i^T$, where e_i is the i-th standard basic vector in $\mathbb{R}^n$; then (P) becomes the problem of maximizing a homogeneous quadratic form over the unit cube).

Semidefinite relaxation: $(x^TA_ix = \mathrm{Tr}(A_ixx^T))$

$$\max\left\{\mathrm{Tr}(AX) \ : \ \mathrm{Tr}(A_iX) \le 1,\, i=1,...,m,\, X \succeq 0\right\}. \qquad (SDP)$$

Main result:

$$\mathrm{Opt}(P) \ge \frac{\mathrm{Opt}(SDP)}{2\ln(2m\mu)}, \qquad \mu = \min[m; \max_i \mathrm{rank}\,(A_i)].$$

A feasible solution x to (P) can be generated efficiently such that

$$x^TAx \ge \frac{\mathrm{Opt}(SDP)}{2\ln(2m^2)} \ge \frac{\mathrm{Opt}(P)}{2\ln(2m^2)}.$$

Related results

1. Goemans and Williamson (1995): if $m = n$, $A_i = e_i e_i^T$ and A is psd with nonpositive non-diagonal entries and row sums 0 (which corresponds to the Maximum Cut problem), then

$$\frac{\text{Opt}(P)}{\text{Opt}(SDP)} \geq 0.87856...$$

2. Nesterov (1997): if A_i are as in 1., and A is psd, then

$$\frac{\text{Opt}(P)}{\text{Opt}(SDP)} \geq \frac{2}{\pi} = 0.6366... \tag{1}$$

   Ye (1998) and Bertsimas and Ye (1998) found closely related results.

3. Nesterov (1998):
   3.a. (1) holds if A_i commute with each other and $A \succeq 0$;
   3.b. If $A_i = a_i a_i^T$, then, for certain efficiently solvable convex optimization program (Cnv)(different from (SDP)) one has

$$\text{Opt(Cnv)} \geq \text{Opt(P)} \geq \frac{1}{e \ln m}\text{Opt(Cnv)}.$$

Assumption on the rank of X

We first show that we may assume $\text{rank}\,(X) \leq m$. Indeed, since $X \succeq 0$, we can efficiently represent X as

$$X = \sum_{j=1}^{n} g_j g_j^T \qquad (g_j \in \mathbf{R}^n)\,.$$

Now consider the LO-problem

$$\max \left\{ c^T\lambda = \text{Tr}(A \sum_{j=1}^{n} \lambda_j g_j g_j^T) \;:\; \text{Tr}(A_i \sum_{j=1}^{n} \lambda_j g_j g_j^T) \leq 1 \; (1 \leq i \leq m),\, \lambda \in \mathbf{R}^n_+ \right\}.$$

This problem is feasible (it contains $\lambda^0 = (1, ..., 1)^T$) and, due to $\sum_i A_i \succ 0$, bounded above. Let λ^* be an extreme point solution. Then the number of nonzero weights λ_j^* is at most m, and hence the matrix

$$X' = \sum_{j=1}^{n} \lambda_j^* g_j g_j^T$$

has rank $\leq m$. Moreover $c^T\lambda^* \geq c^T\lambda^0$; in other words, the matrix X' is a feasible solution to (SDP) with the value of the objective at least as at X.

Hence, without loss of generality, we may assume that we have a feasible solution X to (SDP) of rank $\leq m$. Using this, we present below a randomized algorithm $\mathcal{R}$ which, given a feasible solution X to (SDP) with $\text{Tr}AX > 0$, generates random feasible solutions x to (P).

The randomized rounding procedure

Decompose X as $X = \Delta^T \Delta$ (with $\text{rank}\,(\Delta) \le m$) and define

$$B_i := \Delta A_i \Delta^T, \qquad B := \Delta A \Delta^T.$$

Since X is feasible for (SDP), we have

$$\begin{array}{lrcl}
(a) & B_i & \succeq & 0,\ i = 1, ..., m; \\
(b) & \text{Tr}(B_i) & = & \text{Tr}(A_i X) \le 1,\ i = 1, ..., m; \\
(c) & \text{rank}\,(B_i) & \le & \mu,\ i = 1, ..., m; \\
(d) & \text{Tr}(B) & = & \text{Tr}(AX).
\end{array}$$

Let U be an orthogonal matrix such that $\widehat{B} = U^T B U$ is diagonal and let

$$\widehat{B}_i := U^T B_i U,\ 1 \le i \le m.$$

Let ξ be a ± 1 -vector, whose entries are independently generated with equal probabilities. ξ gives a feasible solution x of (P) as follows:

$$x = x(\xi) = \frac{\Delta^T U \xi}{\sqrt{\max_i \xi^T \widehat{B}_i \xi}},$$

because

$$x^T A_i x = \frac{\xi^T U^T \Delta A_i \Delta^T U \xi}{\max_i \xi^T \widehat{B}_i \xi} = \frac{\xi^T U^T B_i U \xi}{\max_i \xi^T \widehat{B}_i \xi} = \frac{\xi^T \widehat{B}_i \xi}{\max_i \xi^T \widehat{B}_i \xi} \le 1.$$

Main result

The objective value of the vector x obtained in this way is given by

$$\begin{aligned}
x^T(\xi)Ax(\xi) &= \tfrac{1}{\max_i \xi^T \widehat{B}_i \xi}\mathrm{Tr}(\xi^T \widehat{B}\xi) && \left(\widehat{B} \text{ diagonal }, \xi_i \in \{-1,1\}\right) \\
&= \tfrac{1}{\max_i \xi^T \widehat{B}_i \xi}\mathrm{Tr}(\widehat{B}) && \left(\widehat{B} = U^T B U\right) \\
&= \tfrac{1}{\max_i \xi^T \widehat{B}_i \xi}\mathrm{Tr}(B) && \left(B = \Delta A \Delta^T,\, X = \Delta^T \Delta\right) \\
&= \tfrac{1}{\max_i \xi^T \widehat{B}_i \xi}\mathrm{Tr}(AX).
\end{aligned}$$

Proposition 1 *If X is a feasible solution of (SDP) with $Tr(AX) > 0$, then the above rounding procedure is well-defined. The generated x is feasible for (P). Moreover, for every $\alpha > 0$ one has*

$$\Pr\left\{\frac{x^T Ax}{Tr(AX)} \geq \frac{1}{\alpha}\right\} > 1 - 2m\mu\, e^{-\frac{\alpha}{2}},$$

where $\mu = \min[m, \max_i \mathrm{rank}\,(A)_i]$.

Corollary 1 *With positive probability the generated x satisfies*

$$x^T Ax \geq \frac{\mathrm{Opt}(SDP)}{2\ln(2m\mu)} \left(\geq \frac{\mathrm{Opt}(P)}{2\ln(2m\mu)}\right).$$

Proof: Apply $\mathcal{R}$ to an optimal solution X^* of (SDP), and use Proposition 1 with $\alpha^* = 2\ln 2m\mu$. It follows that with positive probability $\mathcal{R}$ generates a feasible solution x to (P) with the value of the objective $\geq \frac{1}{\alpha_*}\mathrm{Tr}(AX_*) = \frac{1}{\alpha_*}\mathrm{Opt}(SDP)$. Hence there exists a feasible x with $x^T Ax \geq \frac{1}{\alpha_*}\mathrm{Opt}(SDP)$. •

Proof of Proposition 1

Step 1 First we prove that $x(\xi)$ always is well-defined, i.e., that $\max_i \xi^T \widehat{B}_i \xi > 0$ for every vector ξ with coordinates ± 1.

Since the matrices $\widehat{B}_i = U^T B_i U = U^T \Delta A_i \Delta^T U$ are positive semidefinite, otherwise $\max_i \xi^T \widehat{B}_i \xi = 0$ for some ± 1-vector ξ. This implies that there exists a ± 1-vector ξ such that

$$\xi^T \widehat{B}_i \xi = 0, \quad 1 \le i \le m.$$

On the other hand, since $\widehat{B}$ is diagonal,

$$\xi^T \widehat{B} \xi = \mathrm{Tr}(\widehat{B}) = \mathrm{Tr}(B) = \mathrm{Tr}(AX) > 0.$$

With

$$x = x(\xi) = \Delta^T U \xi,$$

we now have

$$\begin{aligned} x^T A_i x &= \xi^T U^T \Delta A_i \Delta^T U \xi = \xi^T U^T B_i U \xi = \xi^T \widehat{B}_i \xi = 0, \quad 1 \le i \le m, \\ x^T A x &= \xi^T U^T \Delta A \Delta^T U \xi = \xi^T U^T B U \xi = \xi^T \widehat{B} \xi > 0. \end{aligned}$$

This contradicts the assumption that $\sum_i A_i \succ 0$.

Proof of Proposition 1 (cont.)

We have shown that $x = x(\xi)$ is always well-defined and feasible for (P). Recall that the objective value of $x = x(\xi)$ in (P) satisfies

$$x^T Ax = \frac{\mathrm{Tr}(\xi^T \widehat{B}\xi)}{\max_i \xi^T \widehat{B}_i \xi} = \frac{\mathrm{Tr}(\widehat{B})}{\max_i \xi^T \widehat{B}_i \xi} = \frac{\mathrm{Tr}(B)}{\max_i \xi^T \widehat{B}_i \xi} = \frac{\mathrm{Tr}(AX)}{\max_i \xi^T \widehat{B}_i \xi}.$$

Step 2 We need to show that

$$\Pr\left\{\frac{x^T Ax}{\mathrm{Tr}\,(AX)} \geq \frac{1}{\alpha}\right\} > 1 - 2m\mu\, e^{-\frac{\alpha}{2}}, \quad \forall \alpha > 0$$

or, equivalently,

$$\Pr\left\{\max_i \xi^T \widehat{B}_i \xi > \alpha\right\} < 2m\mu e^{-\frac{\alpha}{2}}, \quad \forall \alpha > 0.$$

Since

$$\Pr\left\{\max_i \xi^T \widehat{B}_i \xi > \alpha\right\} \leq \sum_{i=1}^{m} \Pr\left\{\xi^T \widehat{B}_i \xi > \alpha\right\},$$

this certainly holds if

$$\Pr\left\{\xi^T \widehat{B}_i \xi > \alpha\right\} < 2\mu e^{-\frac{\alpha}{2}}, \quad \forall i, \forall \alpha > 0.$$

Since $\mathrm{Tr}\,\widehat{B}_i \leq 1$, this inequality is a consequence of the next lemma, and hence the proof of the proposition will be complete after this lemma.

Proof of Proposition 1 (cont.)

Lemma 1 *Let* $\operatorname{rank}(B) = k$, $B \succeq 0$ *and* x *a random* ± 1*-vector such that* $\Pr(x_i = \pm 1) = \frac{1}{2}$. *Then*

$$\Pr\left\{x^T B x \geq \alpha \operatorname{Tr} B\right\} \leq 2k e^{\frac{-\alpha}{2}}, \qquad \forall \alpha > 0.$$

Proof: Writing

$$B = \sum_{i=1}^{k} b_i b_i^T$$

the statement in the lemma can be rewritten as

$$\Pr\left\{\sum_{i=1}^{k} \left(b_i^T x\right)^2 \geq \alpha \sum_{i=1}^{k} \|b_i\|^2\right\} \leq 2k e^{\frac{-\alpha}{2}}, \qquad \forall \alpha > 0.$$

We have

$$\begin{aligned}
p_k \;&:=\; \Pr\left\{\sum_{i=1}^{k} \left(b_i^T x\right)^2 \geq \alpha \sum_{i=1}^{k} \|b_i\|^2\right\} = \Pr\left\{\sum_{i=1}^{k} \left(\left(b_i^T x\right)^2 - \alpha \|b_i\|^2\right) \geq 0\right\} \\
&\leq\; \Pr\left\{\max_i \left(\left(b_i^T x\right)^2 - \alpha \|b_i\|^2\right) \geq 0\right\} \leq \sum_{i=1}^{k} \Pr\left\{\left(b_i^T x\right)^2 \geq \alpha \|b_i\|^2\right\} \\
&=\; \sum_{i=1}^{k} \Pr\left\{\left|b_i^T x\right| \geq \sqrt{\alpha}\, \|b_i\|\right\} = 2 \sum_{i=1}^{k} \Pr\left\{b_i^T x \geq \sqrt{\alpha}\, \|b_i\|\right\}.
\end{aligned}$$

Proof of Proposition 1 (cont.)

For any random variable y, with distribution F, we have for any $\rho \geq 0$,

$$\mathcal{E}\left(e^{\rho y}\right) = \int_{-\infty}^{0} e^{\rho y} dF(y) + \int_{0}^{\infty} e^{\rho y} dF(y) \geq 0 + \int_{0}^{\infty} dF(y) = \Pr\left\{y \geq 0\right\}.$$

Hence,

$$\Pr\left\{b_i^T x \geq \sqrt{\alpha}\, \|b_i\|\right\} \leq \mathcal{E}\left(e^{\rho\left(b_i^T x - \sqrt{\alpha}\|b_i\|\right)}\right) = \mathcal{E}\left(e^{\rho\,(b_i^T x)}\right)\; e^{-\rho\sqrt{\alpha}\,\|b_i\|}.$$

Furthermore, using the inequality $\cosh t \leq e^{\frac{1}{2}t^2}$, we have

$$\mathcal{E}\left(e^{\rho(b_i^T x)}\right) = \prod_{j=1}^{n} \mathcal{E}\left(e^{\rho b_{ij} x_j}\right) = \prod_{j=1}^{n} \cosh\left(\rho b_{ij}\right) \leq \prod_{j=1}^{n} e^{\frac{1}{2}\rho^2 b_{ij}^2} = e^{\frac{1}{2}\rho^2 \|b_i\|^2}.$$

Substitution gives

$$\Pr\left\{b_i^T x \geq \sqrt{\alpha}\, \|b_i\|\right\} \leq e^{\frac{1}{2}\rho^2\|b_i\|^2 - \rho\sqrt{\alpha}\|b_i\|}, \qquad \rho \geq 0.$$

The right hand side is minimal if $\rho = \sqrt{\alpha}/\|b_i\|$. Thus we obtain

$$\Pr\left\{b_i^T x \geq \sqrt{\alpha}\, \|b_i\|\right\} \leq e^{\frac{-\alpha}{2}}.$$

From this we derive the inequality

$$p_k \leq 2\sum_{i=1}^{k} \Pr\left\{b_i^T x \geq \sqrt{\alpha}\, \|b_i\|\right\} \leq 2\sum_{i=1}^{k} e^{\frac{-\alpha}{2}} = 2k e^{\frac{-\alpha}{2}},$$

which completes the proof of the lemma. •

Proposition 2 *For every positive integer* $m \geq 3$, *there exists a problem* (P) *with positive definite* A *and* $n = O(\ln m)$ *such that*

$$\frac{\mathrm{Opt}(SDP)}{\mathrm{Opt}(P)} \geq \kappa \ln m,$$

where κ *is a positive absolute constant (for large* m, *one can take* $\kappa = 0.55$ *).*

Proof: Fix $\phi \in (0, \pi/2)$, and let, for positive integer n, Γ_ϕ be a maximal set of unit vectors from $\mathbb{R}^n$ such that the angle between every two distinct vectors from the set is $> \phi$. Let $M(n, \phi)$ denote the cardinality of Γ_ϕ. Then

$$\ln(M(n,\phi)) \leq \left[n \ln \left(\frac{1}{\sin(\phi/2)} \right) \right] (1 + o_\phi(1)),$$

where $o_\phi(1) \to 0$ as $n \to \infty$. Now consider the following instance of (P), with $m = M(n, \phi)$:

$$\max \left\{ x^T I x \ : \ x^T [vv^T] x \leq 1 \quad \forall v \in \Gamma_\phi \right\}. \quad (P_{n,\phi})$$

The matrix $X = I$ clearly is a feasible solution for the associated (SDP), and for this solution $\mathrm{Tr}(AX) = \mathrm{Tr}(I^2) = n$, so that $\mathrm{Opt}(SDP) \geq n$. On the other hand, one can show that

$$\mathrm{Opt}(P_{n,\phi}) \leq \frac{1}{\cos^2(\phi)}.$$

Sharpness of the result (cont.)

We found

$$\mathrm{Opt}(SDP) \geq n, \qquad \mathrm{Opt}(P_{n,\phi}) \leq \frac{1}{\cos^2(\phi)}.$$

Hence

$$\frac{\mathrm{Opt}(SDP)}{\mathrm{Opt}(P_{n,\phi})} \geq n\cos^2(\phi) \geq \frac{\cos^2(\phi)}{\ln \sin^{-1}(\phi/2)}(1+o(1)) \ln M(n,\phi).$$

Specifying $\phi = \frac{3}{16}\pi$, we get for all large enough values of n

$$\frac{\mathrm{Opt}(SDP)}{\mathrm{Opt}(P_{n,\phi})} \geq 0.55 \ln M(n,\phi) = 0.55 \ln m. \qquad \square$$

Derandomization

Given a feasible solution X to (SDP), let $\gamma := \mathrm{Tr}\,(AX) > 0$. For any real $\alpha \geq 2\ln(2m\mu)$, one can find, in an efficient *deterministic* fashion, a feasible solution x to (P) with $x^T Ax \geq \frac{\gamma}{\alpha}$.

The situation can be reduced to:

Given $m\mu$ vectors $g_\ell \in \mathbb{R}^n$, find a ± 1 -vector ξ such that

$$\left| g_\ell^T \xi \right| \leq \sqrt{\alpha}\, \|g_\ell\|_2 \;\; \ell = 1, ..., m\mu.$$

To build such a ξ, a standard derandomization technique works, and $x = x(\xi)$ is the desired vector.

Combined problems

Consider the case when (P) is replaced with the more general problem

$$\max \left\{ x^T A x \;:\; \begin{array}{l} x^T A_i x \leq t_i,\ 1 \leq i \leq m; \\ t \in T \subseteq \mathbb{R}^m_+ \end{array} \right\} \qquad (P^+)$$

where T is a closed and bounded convex set. The natural convex relaxation of (P^+) is the problem

$$\max \left\{ \mathrm{Tr}(AX) \;:\; \begin{array}{l} \mathrm{Tr}(A_i X) \leq t_i,\ 1 \leq i \leq m; \\ t \in T \subseteq \mathbb{R}^m_+ \end{array} \right\} \qquad (C)$$

Applying the same technique as before, we immediately conclude that

$$\mathrm{Opt}(P^+) \geq \frac{\mathrm{Opt}(C)}{2\ln(2m\mu)}$$

and that every feasible solution of (C) with positive value γ of the objective can be efficiently converted into a feasible solution of (P^+) with the value of the objective $\geq \frac{\gamma}{\alpha}$, where $\alpha \geq 2\ln(2m\mu)$.

Inhomogeneous case

Consider the case when (P) is replaced with the problem

$$\max\left\{x^TAx + 2b^Tx \ : \ x^TA_ix \leq 1, \ i = 1, ..., m\right\}. \qquad (P')$$

In that case one can show that

$$\text{Opt}(P') \geq \frac{\text{Opt}(SDP')}{2\ln(2(m+1)\mu)},$$

where $\mu = \min\left[m+1, \max_i \text{rank}\,(A_i)\right]$.

A feasible solution X to (SDP') with a positive value of the objective γ, can be efficiently converted, in a deterministic fashion, into a feasible solution to (P') with the value of the objective $\geq \frac{\gamma}{\alpha}$, where $\alpha \geq 2\ln(2(m+1)\mu)$.

Conjecture 1 *Let* $b \in \mathbb{R}^n$, $\|b\| = 1$ *be given. If* x *is a random* ± 1*-vector such that* $\Pr(x_i = \pm 1) = \frac{1}{2}$, *then*

$$\Pr\left(\left|x^T b\right| \leq 1\right) \geq \frac{1}{2}.$$

Defining

$$p_0 = \Pr\left(\left|x^T b\right| \leq 1\right)$$

the following has been proven:

- $p_0 > 4.10^{-7}$ (Ben-Tal)
- $p_0 > 0.24$ (Nemirovski)
- $p_0 \geq \frac{5}{18} \approx 0.28$ (Van der Wal)
- $p_0 \geq \frac{1}{3}$ (Nemirovski)

References

1. D. Bertsimas, Y. Ye (1997). Semidefinite relaxations, multivariate normal distributions, and order statistics. Working Paper, Department of Management Science, The University of Iowa.

2. M.X. Goemans, D.P. Williamson (1995). Improved approximation algorithms for Maximum Cut and Satisfiability problems using semidefinite programming. *Journal of ACM* **42**, 1115-1145.

3. A. Nemirovski, C. Roos, T. Terlaky (1998). Maximizing a Quadratic form over the Intersection of Ellipsoids with Common Center. *Math. Progr.* **86**, 463–473, 1999.

4. Yu. Nesterov (1997). Semidefinite relaxation and non-convex quadratic optimization. *Optimization Methods and Software* **12** (1997), 1-20.

5. Yu. Nesterov (1998). Global Quadratic Optimization via Conic Relaxation. Working Paper, CORE, Catholic University of Louvaine, 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium.

6. Yinyu Ye. Approximating global quadratic optimization with convex quadratic constraints. *J. Global Optim.*, 15(1):1–17, 1999.