Several open problems about uncertain Linear Matrix Inequalities

ACO Student Seminar, March 26, 2008

Arkadi Nemirovski

♣ For starters:

Open question: Let $a \in \mathbf{R}^n$, $\|a\|_2 \leq 1$, be a deterministic vector and $\xi_1, ..., \xi_n$ be independent random variables taking values ± 1 with probability 0.5.

Prove or disprove:

$$\mathrm{Prob}\left\{ |\sum_{i=1}^{n} a_i \xi_i| \leq 1 \right\} \geq \kappa = \frac{1}{2} \qquad (*)$$

♠ Comments:

• Origin: The question is a byproduct of certain meaningful reasoning aimed at investigating tightness of the *Semidefinite Relaxation* of a specific NP-hard problem. In the context os this question, a whatever positive absolute constant in the role of κ is more or less ok, and I originally proved that one can take $\kappa = 1.\text{e-}7$.

One of my co-authors (Prof. C. Roos from Delft University, The Netherlands) became "obsessed" with this problem; based on numerical experiments, he guessed that one can take κ as large as $1/2$.

• **Current state:** **A.** **Taking** $n = 2$, $a_1 = a_2 = 1/\sqrt{2}$, **we see that** $(*)$ **is, in general, invalid when** $\kappa > 1/2$.

B. P. van der Wal (TU Delft, 2001; see SIAM J. Optim. 13:2 (2002), 535–560**) proved, using a nice "genuine" probabilistic argument, that one can take** $\kappa = 1/3$.

C. The result of van der Wal can be slightly improved, but I am not aware of a proof that $(*)$ **holds true with, say,** $\kappa = 0.4$.

• **Intrinsic difficulty: Potential presence of** $O(1)$ **"large"** a_i**'s.**

When n **is large, "typical" unit vectors** a **have small** $\|\cdot\|_\infty$**-norm. When** $\|a\|_\infty$ **is small, by the Central Limit Theorem the distribution of the random sum**

$$\zeta = \sum_i a_i \xi_i$$

should approach the standard Gaussian distribution $\mathcal{N}(0,1)$**, and for** $\zeta \sim \mathcal{N}(0,1)$ **one has**

$$\mathrm{Prob}\{|\zeta| \leq 1\} \approx 0.6827$$

which is "much larger" than the desired bound $\kappa = 1/2$.

Semidefinite Programming

♣ **Let** $\nu = (n_1, ..., n_m)$, $n = \sum_i n_i$, **and let** $\mathbf{S}^\nu$ **be the space of symmetric** $n \times n$ **block-diagonal matrices with diagonal blocks of sizes** $n_1, ..., n_m$. $\mathbf{S}^\nu$ **is equipped with the Frobenius inner product**

$$\langle A, B \rangle = \mathrm{Tr}(AB) = \sum_{i,j=1}^{n} A_{ij} B_{ij}.$$

♣ **A** *semidefinite program* **on** $\mathbf{S}^\nu$ **is an optimization problem of the form**

$$\min_{x \in \mathbf{R}^N} \left\{ c^T x : \begin{array}{l} \mathcal{A}(x) \equiv \sum_{j=1}^{N} x_j A_j - A_0 \succeq 0 \\ Bx - b = 0 \end{array} \right\} \qquad \textbf{(SDP)}$$

where $P \succeq Q$ **means that** P, Q **are symmetric matrices of the same size such that** $A - B$ **is positive semidefinite. Thus, the** *Linear Matrix Inequality constraint*

$$\mathcal{A}(x) \succeq 0$$

says that the symmetric matrix $\mathcal{A}(x)$ *affinely depending on* x *should belong to the cone* $\mathbf{S}^\nu_+$ *comprised of positive semidefinite matrices from* $\mathbf{S}^\nu$.

$$\min_{x\in\mathbf{R}^N}\left\{c^Tx:\begin{array}{l}\mathcal{A}(x)\equiv\Sigma_{j=1}^N x_jA_j-A_0\succeq 0\\ Bx-b=0\end{array}\right\} \qquad \textbf{(SDP)}$$

Why do we like Semidefinite Programming?

♣ **SDP possesses extremely powerful "expressive abilities: with slight exaggeration,** *every "explicitly given" convex optimization problem can be written down as a SDP or can be approximated by an SDP in a polynomial time fashion to a whatever high accuracy.*

For example, the messy and highly nonlinear problem

	minimize $\sum\limits_{\ell=1}^{n} x_\ell^2$
(a)	$x \geq 0;$
(b)	$a_\ell^T x \leq b_\ell,\ \ell = 1,...,n;$
(c)	$\\|Px - p\\|_2 \leq c^T x + d;$
(d)	$x_\ell^{\frac{\ell+1}{\ell}} \leq e_\ell^T x + f_\ell,\ \ell = 1,...,n;$
(e)	$x_\ell^{\frac{l}{l+3}} x_{l+1}^{\frac{1}{l+3}} \geq g_\ell^T x + h_\ell,\ \ell = 1,...,n-1;$
(f)	$\mathrm{Det}\begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ x_2 & x_1 & x_2 & \cdots & x_{n-1} \\ x_3 & x_2 & x_1 & \cdots & x_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & x_{n-1} & x_{n-2} & \cdots & x_1 \end{pmatrix} \geq 1;$
(g)	$1 \leq \sum\limits_{\ell=1}^{n} x_\ell \cos(\ell\omega) \leq 1 + \sin^2(5\omega)\ \forall \omega \in \left[-\frac{\pi}{7}, 1.3\right]$

can be converted, *in a systematic way*, **into an equivalent SDP.**

$$\min_{x\in\mathbf{R}^N}\left\{c^Tx : \begin{array}{l}\mathcal{A}(x)\equiv\sum_{j=1}^N x_jA_j - A_0 \succeq 0\\ Bx-b=0\end{array}\right\} \qquad \textbf{(SDP)}$$

♠ **More specifically,**

• **When** $n_i = 1$ **for all** i **(i.e.,** $A_0, A_1, ...$ **are diagonal), (SDP) recovers Linear Programming**

• **When the blocks in** $A_0, A_1, ...$ **are** *arrow matrices*

$$\left[\begin{array}{c|cccc} t & a_1 & a_2 & \ldots & a_k \\ \hline a_1 & t & & & \\ a_2 & & t & & \\ \vdots & & & \ddots & \\ a_k & & & & t \end{array}\right]$$

(SDP) becomes a *conic quadratic program*

$$\min_x\left\{c^Tx : \|P_ix-p_i\|_2 \leq q_i^Tx + r_i, 1\leq i\leq m\right\}$$

Conic Quadratic Programming covers a lot of applications: convex quadratic quadratically constrained problems, convex problems with algebraic constraints like

$$x \geq 0, \quad \sum_\alpha c_\alpha x_1^{\alpha_1}...x_N^{\alpha_N} \geq a^Tx+b$$
$$[\alpha \geq 0 : \text{ rational vectors, } \sum_i \alpha_i \leq 1, c_\alpha \geq 0]$$

and much more.

• **SDP is the natural language for modelling numerous problems in Statistics, Mechanical Engineering, Structural Design, Signal Processing, Control,...**

♣ **Along with its tremendous "expressive abilities", SDP possesses transparent and deep mathematical structure allowing for**

• **Instructive processing of semidefinite programs "on paper", mainly due to the** *Semidefinite Duality*

• **Developing powerful polynomial time solution algorithms for SDP.**

Semidefinite Duality

$$\mathrm{Opt}(P) = \min_x \left\{ c^T x : \begin{array}{l} \sum_{j=1}^N x_j A_j - A_0 \succeq 0 \\ Bx - b = 0 \end{array} \right\} \qquad (P)$$

$$\mathrm{Opt}(D) = \max_{y,Y} \left\{ b^T y + \langle A_0, Y \rangle : \begin{array}{l} Y \succeq 0 \\ \langle A_i, Y \rangle = c_i \\ 1 \leq i \leq N \end{array} \right\} \qquad (D)$$

Semidefinite Duality Theorem:

(i) [Symmetry] *Duality is symmetric: (D) is a semidefinite program, and its semidefinite dual is (equivalent to) (P)*

(ii) [Weak Duality] $\mathrm{Opt}(D) \leq \mathrm{Opt}(P)$

(iii) [Strong Duality] *If one of the problems (P), (D) is strictly feasible and bounded, then the other problems is solvable and*

$$\mathrm{Opt}(P) = \mathrm{Opt}(D). \qquad (!)$$

When both (P) and (D) are strictly feasible, both problems are solvable, and $(!)$ holds true.

♠ Polynomial Time Interior Point methods for SDP

Theorem. *Consider SDP with upper bounds on variables:*

$$\min_{x\in\mathbf{R}^N}\left\{c^Tx:\ \begin{array}{l}\mathcal{A}(x)\equiv\Sigma_{j=1}^N x_jA_j - A_0 \succeq 0\\ \|x\|_2\le R\end{array}\right\} \qquad \text{(SDP)}$$

For every $\epsilon>0$, one can find an ϵ-feasible and ϵ-optimal solution to the problem (or detect correctly that the problem is infeasible) at the cost of solving at most

$$\mathcal{N}=\sqrt{n}\ln\left(\frac{nN(\Theta+1)}{\epsilon}+2\right)$$

$$[\Theta:\ \text{maximal magnitude of data coefficients}]$$

steps, with every step reducing to assembling and solving an $N\times N$ nonsingular system of linear equations. The arithmetic cost of assembling a single system does not exceed

$$O(1)\left[N^2n+N\sum_i n_i^3\right].$$

Randomization and Semidefinite Relaxations of Combinatorial Problems

♣ **Consider a quadratic quadratically constrained problem**

$$\text{Opt} = \max_x \{f_0(x) : f_i(x) \le 0,\ 1 \le i \le m\} \qquad (C)$$
$$\left[f_i(x) = x^T A_i x + 2b_i^T x + c_i\right]$$

Note: We do not assume that f_i are convex and thus can model combinatorial constraints on variables:

$$x_i \in \{0;1\} \Leftrightarrow \left\{\begin{array}{l} x_i - x_i^2 \le 0 \\ -x_i + x_i^2 \le 0 \end{array}\right. ;\ x_i \in \{-1;1\} \Leftrightarrow \left\{\begin{array}{l} 1 - x_i^2 \le 0 \\ -1 + x_i^2 \le 0 \end{array}\right.$$

• **Setting** $Q_i = \left[\begin{array}{cc} A_i & b_i \\ b_i^T & c_i \end{array}\right]$, $X(x) = \left[\begin{array}{cc} xx^T & x \\ x^T & 1 \end{array}\right]$, **the problem can be rewritten equivalently as**

$$\text{Opt} = \max_X \left\{\text{Tr}(Q_0 X) : \begin{array}{c} \text{Tr}(Q_i X) \le 0,\ 1 \le i \le m \\ X \in \mathbf{S}_+^{N+1},\ X_{N+1,N+1} = 1 \\ \text{Rank} X = 1 \end{array}\right\}.$$

Removing the constraint $\text{Rank}(X) = 1$**, we end up with** *semidefinite relaxation* **of the problem:**

$$\text{Opt(SDP)} = \max_X \left\{\text{Tr}(Q_0 X) : \begin{array}{c} \text{Tr}(Q_i X) \le 0,\ 1 \le i \le m \\ X \in \mathbf{S}_+^{N+1},\ X_{N+1,N+1} = 1 \end{array}\right\}. \qquad \text{(SDP)}$$

By construction, $\text{Opt} \le \text{Opt(SDP)}$.

$$\begin{array}{c}\mathrm{Opt} = \max_x \left\{ f_0(x) : f_i(x) \leq 0,\, 1 \leq i \leq m \right\} \\ \left[f_i(x) = x^T A_i x + 2b_i^T x + c_i \right]\end{array} \qquad (C)$$

$$\mathrm{Opt(SDP)} = \max_X \left\{ \mathrm{Tr}(Q_0 X) : \begin{array}{l} \mathrm{Tr}(Q_i X) \leq 0,\, 1 \leq i \leq m \\ X \in \mathbf{S}_+^{N+1},\, X_{N+1,N+1} = 1 \end{array} \right\}. \qquad \mathrm{(SDP)}$$

♠ **A useful interpretation of the relaxation is as follows: let us look at** *random* **solutions** $\tilde{x}$ **to (C) which satisfy the constraints** *at average*, **and let us minimize under this restriction the** *expected value* **of the objective. Passing from random solutions** $\tilde{x}$ **to their covariance matrices**

$$X = \mathbf{Cov}\{\tilde{x}\} := \mathbf{E}\left\{ \begin{bmatrix} \tilde{x} \\ 1 \end{bmatrix} \begin{bmatrix} \tilde{x} \\ 1 \end{bmatrix}^T \right\},$$

we arrive at (SDP)**. Vice versa, given a feasible solution** X **to** (SDP)**, we can define (various) random solutions** $\tilde{x}$ **to** (C) **such that** $X = \mathbf{Cov}\{\tilde{x}\}$**. Analyzing these solutions, we, with some luck, can bound the "gap" between** Opt **and** Opt(SDP)**.**

♣ **Example:** **Consider the problem**

$$\mathrm{Opt} = \max_x \left\{ x^T A x + 2b^T x : x^T Q_i x \le q_i,\ 1 \le i \le m \right\}. \quad (C)$$

The SDP relaxation here is

$$\mathrm{Opt(SDP)} = \min_{\lambda,\mu} \left\{ \sum_i q_i \lambda_i + \mu : \left\{ \begin{array}{l} \left[\begin{array}{cc} \sum_i \lambda_i Q_i - A & b \\ b^T & \mu \end{array} \right] \succeq 0 \\ \lambda_i \ge 0, 1 \le i \le m \end{array} \right\} \right\} \quad \mathrm{(SDP)}$$

There are two known cases when one can bound the gap between Opt **and** Opt(SDP)**:**

A. **The matrices** $Q_1, ..., Q_m$ **commute with each other and** $A \succeq 0$**. In this case,**

$$\mathrm{Opt} \le \mathrm{Opt(SDP)} \le \frac{\pi}{2} \mathrm{Opt}$$

(Nesterov, Optim. Methods Softw. 9 (1998), 141-160**)**

B. **The matrices** $Q_1, ..., Q_m$ **are positive semidefinite with positive definite sum. In this case,**

$$\mathrm{Opt} \le \mathrm{Opt(SDP)} \le O(1) \ln(m) \mathrm{Opt}$$

(Ben-Tal, Nem., Roos, SIAM J. Optim. 13:2 (2002), 535–560**).**

♠ **Here is the sketch of the proof for B:**

- **With moderate effort, the situation reduces to the following one: We are given**
- **a vector** a, $\|a\|_2 \leq 1$,
- m **matrices** $R_i \succeq 0$, $1 \leq i \leq m$, **with** $\mathrm{Tr}(R_i) \leq 1$, **and**
- **a diagonal matrix** S.

  We want to find a vector $\bar{y}$ **such that**

$$\bar{y}^T S \bar{y} = \mathrm{Tr}(S); \;\; |a^T \bar{y}| \leq 1; \;\; \bar{y}^T R_i \bar{y} \leq \Omega = O(1)\ln(m). \quad (*)$$

To this end, consider a random vector ξ **with independent entries taking values** ± 1 **with probabilities** $1/2$. **Then**

$$\xi^T S \xi \equiv \mathrm{Tr}(S) \;\&\; \mathbf{E}\{\xi^T R_i \xi\} = \mathrm{Tr}(R_i) = 1, \, 1 \leq i \leq m.$$

From $\mathrm{Tr}(R_i) = 1$ **and** $R_i \succeq 0$ **it follows with some effort that**

$$\forall \Theta > 0 : \mathrm{Prob}\{\xi^T R_i \xi > \Theta\} \leq C \exp\{-\Theta/C\}, \quad C = O(1). \quad (!)$$

Besides this, we know from the "From starters" problem that

$$\mathrm{Prob}\{|a^T \xi| \leq 1\} \geq \kappa = O(1).$$

It follows that for every $\Theta > 0$ **one has**

$$\begin{array}{c}\mathrm{Prob}\left\{|a^T \xi| > 1 \textbf{ or } \exists i : |\xi^T R_i \xi| > \Theta\right\} \\ \leq (1-\kappa) + mC \exp\{-C\Theta\}.\end{array}$$

Choosing Θ **from the condition** $mC \exp\{-C\Theta\} = \frac{\kappa}{2}$, **we get** $\Theta = O(1)\ln(m)$ **and the probability in** $(!)$ **is** $\leq 1 - \kappa/2$. **It follows that a realization** ξ **with probability** $\kappa/2 > O(1)$ **satisfies** $(*)$ **with the announced value of** Ω.

Question: *How can we "derandomize" the construction? Specifically, given m positive semidefinite matrices $R_1, ..., R_m$, we need to find efficiently, in a deterministic fashion, a vector ξ with entries ± 1 such that*

$$\xi^T R_i \xi \leq O(1) \ln(m) \mathrm{Tr}(R_i), \; 1 \leq i \leq m$$

Current state: I know how to derandomize efficiently, but I believe a much simpler algorithm can be found. What I know, is based on Math. Progr. 86 (1999), 463–473.

♣ **It was mentioned that**

"From $\mathrm{Tr}(R) = 1$ *and* $R \succeq 0$ *it follows with some effort that*

$$\forall \Theta > 0 : \mathrm{Prob}\{\xi^T R_i \xi > \Theta\} \le C \exp\{-\Theta/C\}, \quad C = O(1). \qquad (!)$$

where $\xi \sim \mathrm{Uniform}\left(\{-1;1\}^N\right)$*."*

What is actually said is as follows. Since $R \succeq 0$**, we have** $R = D^T D$ **with certain** D**, and**

$$\xi^T R \xi = \xi^T D^T D \xi = \|D\xi\|_2^2.$$

On the other hand, $\mathrm{Tr}(R) = \sum_{i,j} D_{ij}^2 = 1$**. Denoting** D_i i**-th column of** D**, the above statement reads as follows:**

If $D_i \in \mathbf{R}^n$, $1 \le i \le N$, *are such that* $\sum_i \|D_i\|_2^2 \le 1$, *then with* $C = O(1)$ *one has*

$$\forall \Theta > 0 : \mathrm{Prob}\left\{\| \sum_{i=1}^N \xi_i D_i\|_2 > \Theta\right\} \le C \exp\{-\Theta^2/C\}.$$

or, by homogeneity:

(!) *If* $\|\cdot\| = \|\cdot\|_2$, $D_i \in \mathbf{R}^n$, $1 \le i \le N$, *and* $\xi_i \sim \mathrm{Uniform}\left(\{-1;1\}^N\right)$, *then with* $C = O(1)$ *one has*

$$\forall \Theta \ge 0 : \mathrm{Prob}\left\{\| \sum_i \xi_i D_i\| > \Theta \sqrt{\sum_i \|D_i\|^2}\right\} \le C \exp\{-\Theta^2/C\}.$$

The fact that ξ **is uniformly distributed over the vertices of the unit cube is of no importance here. What matters is:**

$$\xi_i \text{ are independent }, \mathbf{E}\{\xi_i\} = 0, \mathbf{E}\{\exp\{\xi_i^2\}\} \le \exp\{1\} \quad (*)$$

On the other hand, this natural "large deviation" result is sensitive to the type of the norm: *a "typical" $\|\cdot\|_1$-norm of a random sum $\sum_i \xi_i D_i$ with $D_i \in \mathbf{R}^n$ can be as large as* $\sqrt{n}\sqrt{\sum_{i=1}^N \|D_i\|_1^2}$, **at least when $N \leq n$.**

♣ **The question about large deviations of** *random vector sums* **in $\mathbf{R}^n$ can be posed as follows:**

Let $\sigma_i > 0$ be reals, $\zeta_i \in \mathbf{R}^n$ be independent random vectors with $\mathbf{E}\{\zeta_i\} = 0$ and with "typical norms of order of σ_i":

$$\mathbf{E}\{\exp\{\|\zeta_i\|^2/\sigma_i^2\}\} \leq \exp\{1\},$$

where $\|\cdot\|$ is certain norm. What should we require from the norm to ensure that

$$\forall \Theta \geq 0 : \mathrm{Prob}\left\{\|\sum_{i=1}^N \zeta_i\| \geq \Theta\sqrt{\sum_{i=1}^n \sigma_i^2}\right\} \leq C\exp\{-\frac{\Theta^2}{\kappa C}\} \tag{!}$$

with $C = O(1)$ and a "moderate" κ depending solely on the norm $\|\cdot\|$?

What is known:

- **(!) is always true when $\kappa = \sqrt{n}$ (this is the best possible κ when $\|\cdot\| = \|\cdot\|_1$);**
- **When $\|\cdot\| = \|\cdot\|_p$ with $2 \leq p \leq \infty$, (!) holds true with $\kappa = O(1)\min[p, \ln(n+1)]$.**

Let $\sigma_i > 0$ be reals, $\zeta_i \in \mathbf{R}^n$ be independent random vectors with $\mathbf{E}\{\zeta_i\} = 0$ and with "typical norms of order of σ_i":

$$\mathbf{E}\{\exp\{\|\zeta_i\|^2/\sigma_i^2\}\} \leq \exp\{1\},$$

where $\|\cdot\|$ is certain norm. What should we require from the norm to ensure that

$$\forall \Theta \geq 0 : \mathrm{Prob}\left\{\|\sum_{i=1}^{N} \zeta_i\| \geq \Theta\sqrt{\sum_{i=1}^{n} \sigma_i^2}\right\} \leq C\exp\{-\frac{\Theta^2}{\kappa C}\} \qquad (!)$$

with $C = O(1)$ and a "moderate" κ depending solely on the norm $\|\cdot\|$?

♠ **Now let $\mathbf{R}^n = \mathbf{R}^{m\times k}$ be the space of $m \times k$ matrices. It can be equipped with the norms**

$$|A|_p = \|\sigma(A)\|_p, \qquad [1 \leq p \leq \infty]$$

where $\sigma(A)$ is the vector of singular values of A. In particular, $\|A\|_\infty$ is the usual matrix norm $\|A\| = \max_x\{\|Ax\|_2 : \|x\|_2 \leq 1\}$.

- **When $\|\cdot\|$ is the norm $|\cdot|_p$, $2 \leq p \leq \infty$, on $\mathbf{R}^{m\times k}$, (!) holds true with $\kappa = O(1)\min[p, \ln(1+\min[m,k])]$**

(see http://www2.isye.gatech.edu/∼nemirovs/LargeDev2004.pdf)

In particular,

In the case when $\|\cdot\|$ is the matrix norm on $\mathbf{R}^{m\times k}$, (!) holds true with $\kappa = O(1)\ln(1+\min[m,k])$.

♣ **<u>Potential application:</u> Consider a** *randomly perturbed LMI*

$$A_0[x] - \sum_{i=1}^{N} \xi_i A_i[x] \succeq 0,$$

where

• $A_0[x], ..., A_N[x]$ **are matrices from $\mathbf{S}^\nu$ affinely depending on the decision vector** x

• ξ_i **are independent random perturbations satisfying**

$$\xi_i \text{ are independent }, \mathbf{E}\{\xi_i\} = 0, \mathbf{E}\{\exp\{\xi_i^2\}\} \le \exp\{1\} \quad (*)$$

♠ **A natural way to process** $(*)$ **is to pass to its** *chance constrained* **form:**

$$\mathrm{Prob}\left\{\sum_{i=1}^{N} \xi_i A_i[x] \preceq A_0[x]\right\} \ge 1 - \epsilon. \qquad \text{(ChC)}$$

This resulting deterministic constraint, however, is difficult to verify; besides, it typically defines a non-convex set, which makes it difficult to optimize under such constraint(s).

♣ **What we are interested in, is a** *safe tractable approximation of a given chance constrained LMI*, **that is,** *a computationally tractable convex set W_ϵ which is contained in the feasible set of* (ChC).

$$\text{Prob}\left\{\sum_{i=1}^{N}\xi_i A_i[x] \preceq A_0[x]\right\} \geq 1-\epsilon. \qquad \text{(ChC)}$$

- **Aside of pathological cases, at a feasible solution to** (ChC) **one should have** $A_0[x] \succ 0$**. Assuming that it is the case, we can rewrite** (ChC) **equivalently as**

$$\text{Prob}\left\{\sum_{i=1}^{N}\xi_i B_i[x] \preceq I\right\} \geq 1-\epsilon, \; B_i[x] = A_i^{-1/2}[x]A_i[x]A_i^{-1/2}[x].$$

Assuming the distributions of ξ_i **symmetric w.r.t. 0, this is essentially the same as to require that**

$$\text{Prob}\left\{-I \preceq \sum_{i=1}^{N}\xi_i B_i[x] \preceq I\right\} \geq 1-\epsilon$$
$$\Updownarrow$$
$$\boxed{\text{Prob}\left\{\|\sum_{i=1}^{N}\xi_i B_i[x]\| \leq 1\right\} \geq 1-\epsilon.}$$

♣ **We arrive at the question:**

Given deterministic matrices $B_1, ..., B_N \in \mathbf{S}^\nu$ and random variables ξ_i satisfying

$$\xi_i \text{ are independent }, \mathbf{E}\{\xi_i\} = 0, \mathbf{E}\{\exp\{\xi_i^2\}\} \leq \exp\{1\}$$

what are verifiable (and not too conservative) sufficient conditions for the relation

$$\text{Prob}\left\{\|\sum_{i=1}^N \xi_i B_i\| \leq 1\right\} \geq 1-\epsilon \qquad (!)$$

♠ **First attempt: Invoking our large deviation results to random matrices $\zeta_i = \xi_i B_i$, we arrive at a sufficient condition as follows:**

Assume that the matrices B_i satisfy the condition

$$\sum_{i=1}^N \|B_i\|^2 \leq \Upsilon \equiv \frac{c}{\sqrt{\ln(n)\ln(1/\epsilon)}}$$

with appropriately chosen $c = O(1)$. Then (!) is satisfied.

♠ **Oops... When substituting into our sufficient condition $B_i = B_i[x] = A_0^{-1/2}[x]A_i[x]A_0^{-1/2}[x]$, the condition becomes**

$$\sum_{i=1}^N \|A_0^{-1/2}[x]A_i[x]A_0^{-1}[x]A_i[x]A_0^{-1/2}\| \leq \Upsilon,$$

and the corresponding feasible set is, in general, *nonconvex*, **aside of the case when $A_0[x] \succ 0$ is independent of x.**

Besides, the condition can be too conservative. E.g., when $N = n$, $B_i = e_i e_i^T$ and $\|\xi\|_\infty \leq 1$, we have

$$\text{Prob}\left\{\| \sum_{i=1}^{N} \xi_i B_i \| \leq 1\right\} = 1,$$

while $\Sigma_{i=1}^{n} \|B_i\|^2 = n$, i.e., the sufficient condition is very far from being satisfied.

Given deterministic matrices $B_1, ..., B_N \in \mathbf{S}^\nu$ *and random variables* ξ_i *satisfying*

$$\xi_i \text{ are independent }, \mathbf{E}\{\xi_i\} = 0, \mathbf{E}\{\exp\{\xi_i^2\}\} \leq \exp\{1\}$$

what are verifiable (and not too conservative) sufficient conditions for the relation

$$\mathrm{Prob}\left\{\|\sum_{i=1}^N \xi_i B_i\| \leq 1\right\} \geq 1 - \epsilon \qquad (!)$$

♣ **Second attempt: Let us "start with the opposite end". Let S be the random matrix $\sum_{i=1}^N \xi_i B_i$. Assume that**

$$\xi \sim \mathrm{Uniform}\left(\{-1;1\}^N\right)$$

and (!) holds true for certain $\epsilon \ll 1$. Then it is natural to guess (and is indeed true) that

$$\mathbf{E}\{\underbrace{\|S\|^2}_{=\|S^2\|}\} \leq O(1),$$

whence $\sum_{i=1}^N B_i^2 = \mathbf{E}\left\{S^2\right\} \leq O(1)I$. **Thus, a** *necessary* **condition for (!) is**

$$\sum_{i=1}^N B_i^2 \preceq O(1)I, \qquad (!!)$$

and intuition says that this *necessary* **condition should not be very far from a** *sufficient* **one. Indeed, under this condition the expectation of the** *positive semidefinite* **random matrix S^2 is $O(1)I$, and "a light-tail** *nonnegative* **random variable should be typically of order of its expectation" – the rule which is true for random** *reals* **(i.e., for random 1×1 symmetric matrices).**

♣ **We arrive at**

Conjecture: *Let $B_1, ..., B_N$ be deterministic matrices from $\mathbf{S}^\nu$ such that*

$$\sum_{i=1}^N B_i^2 \preceq \Upsilon I$$

with appropriately chosen and "not too close to 0" $\Upsilon = \Upsilon(n, \epsilon)$. Then

$$\mathrm{Prob}\left\{\| \sum_{i=1}^N \xi_i B_i\| \leq 1\right\} \geq 1 - \epsilon,$$

provided $\xi \sim \mathrm{Uniform}\left(\{-1; 1\}^N\right)$.

♠ **On a detailed inspection, "not too close to 0" Υ should satisfy the upper bound**

$$\Upsilon(n, \epsilon) \leq \frac{O(1)}{\ln(n) + \ln(1/\epsilon)},$$

so that a scientific version of the Conjecture reads:

Conjecture: *Let $B_1, ..., B_N$ be deterministic matrices from $\mathbf{S}^\nu$ such that*

$$\sum_{i=1}^N B_i^2 \preceq \Upsilon I, \; \Upsilon = \frac{c}{\ln(n) + \ln(1/\epsilon)}$$

with appropriately chosen $c = O(1)$. Then

$$\mathrm{Prob}\left\{\| \sum_{i=1}^N \xi_i B_i\| \leq 1\right\} \geq 1 - \epsilon,$$

provided $\xi \sim \xi \sim \mathrm{Uniform}\left(\{-1; 1\}^N\right)$.

Conjecture: *Let $B_1, ..., B_N$ be deterministic matrices from $\mathbf{S}^\nu$ such that*

$$\sum_{i=1}^{N} B_i^2 \preceq \Upsilon I,\ \Upsilon = \frac{c}{\ln(n) + \ln(1/\epsilon)}$$

with appropriately chosen $c = O(1)$. Then

$$\mathrm{Prob}\left\{\|\sum_{i=1}^{N} \xi_i B_i\| \leq 1\right\} \geq 1 - \epsilon,$$

provided $\xi \sim \xi \sim \mathrm{Uniform}\left(\{-1;1\}^N\right)$.

♠ **Comments:**

• **The assumption $\xi \sim \mathrm{Uniform}\left(\{-1;1\}^N\right)$ plays no significant role: if the Conjecture were true in this case, it would be true for every collection of (independent) ξ_i with $\mathbf{E}\{\xi_i\} = 0$ and $\mathbf{E}\{\exp\{\xi_i^2\}\} \leq \exp\{1\}$.**

• **All we need to prove the Conjecture is a "rough" fact like this:**

Let $B_1, ..., B_N$ be deterministic matrices from $\mathbf{S}^\nu$ such that $\Sigma_{i=1}^N B_i^2 \preceq I$. Then with $\xi \sim \mathrm{Uniform}\left(\{-1;1\}^N\right)$ and an appropriate $c_1, c_2 = O(1)$ one has

$$\mathrm{Prob}\left\{\|\sum_{i=1}^{N} \xi_i B_i\| > c_1\sqrt{\ln n}\right\} \leq 1 - c_2.$$

• **The Conjecture is trivially true when B_i are diagonal matrices. It is also true (even with $\Upsilon = \frac{O(1)}{\ln(1/\epsilon)}$) when**

$$B_i = A_0^{-1/2} A_i A_0^{-1/2}$$

and $A_0 \succ 0, A_1, ..., A_N$ are arrow matrices.

• **For general-type symmetric matrices** B_i**, the best known provably true approximation of the Conjecture is the implication**

$$\begin{array}{c}\Sigma_{i=1}^N B_i^2 \preceq \Upsilon I,\ \Upsilon = \frac{c}{n^{1/3}+\ln(1/\epsilon)} \\ \Downarrow \\ \mathrm{Prob}\left\{\| \Sigma_{i=1}^N \xi_i B_i\| \leq 1\right\} \geq 1-\epsilon,\end{array}$$

where $c = O(1)$ **and** $\xi \sim \xi \sim \mathrm{Uniform}\left(\{-1;1\}^N\right)$**.** (Math. Progr. 109:2-3 (2007), 283-317)**.**

• **Assume we know an** $\Upsilon = \Upsilon(n,\epsilon)$ **such that the implication**

$$\Sigma_{i=1}^N \left[A_0^{-1/2}[x]A_i[x]A_0^{-1/2}[x]\right]^2 \preceq \Upsilon I \qquad (*)$$

$$\Downarrow$$

$$\mathrm{Prob}\left\{\| \Sigma_{i=1}^N \xi_i A_0^{-1/2}[x]A_i[x]A_0^{-1/2}[x]\| \leq 1\right\} \geq 1-\epsilon. \qquad (!)$$

holds true for all x**. Then** $(*)$ **is a safe approximation of the chance constrained LMI**

$$\mathrm{Prob}\left\{\sum_{i=1}^N \xi_i A_i[x] \preceq A_0[x]\right\} \geq 1-\epsilon,$$

and this approximation is given by the *arrow LMI*

$$\left[\begin{array}{c|cccc} \Upsilon A_0[x] & A_1[x] & \ldots & A_N[x] \\ \hline A_1[x] & \Upsilon A_0[x] & & \\ \vdots & & \ddots & \\ A_N[x] & & & \Upsilon A_0[x]\end{array}\right] \succeq 0$$

and is therefore a not only safe, but convex and computationally tractable.

Open question: *Prove or disprove the Conjecture.*

♣ **A nice open problem (a byproduct of one of many attempts to prove the Conjecture):**

> **Prove or disprove:** *Let A, B be positive semidefinite. Consider a product P of p copies of A and q copies of B, taken in an arbitrary order, like $P = AB^2AB$ $(p = 2, q = 3)$. Then* $\mathrm{Tr}(P) \leq \mathrm{Tr}(A^pB^q)$.

♠ **Comment: The fact is definitely true when $p+q = 4$, and there is certain numerical evidence that it is true in general. At the same time, to the best of my memory, the validity status of the guess is unknown already when $p + q = 5$, and it definitely is unknown in the general case.**

♣ **Application: Nonconvex quadratic optimization under norm constraints.**

♠ **The problem: Let**

$$x = \{x_i\}_{i=1}^N \in \mathbf{R}^{p_1\times q_1} \times ... \times \mathbf{R}^{p_N\times q_N} =: \mathbf{R}^M$$

be a collection of N variable matrices of prescribed sizes, and let $\mathcal{A}$, $\mathcal{Q}_1, ...\mathcal{Q}_m$ be symmetric $M\times M$ matrices such that $\mathcal{Q}_1, ..., \mathcal{Q}_m$ are positive semidefinite with positive definite sum. Consider the optimization problem

$$\mathrm{Opt} = \max_x \left\{\langle x, \mathcal{A}x\rangle : \begin{array}{l}\langle x, \mathcal{Q}_\ell x\rangle \le 1,\ \ell = 1, ..., m\\ \|x_i\| \le 1,\ i = 1, ..., N\end{array}\right\}.$$

This problem typically is NP-hard. It, however, admits a semidefinite relaxation as follows: Let $\mathcal{X}$ be the symmetric $M\times M$ matrix with entries which are pairwise products of entries in x. Then

- $\langle x, \mathcal{A}x\rangle = \mathrm{Tr}(\mathcal{A}\mathcal{X})$, $\langle x, \mathcal{Q}_\ell x\rangle = \mathrm{Tr}(\mathcal{Q}_\ell\mathcal{X})$
- $x_i x_i^T$ **and** $x_i^T x_i$ **are obtained from** $\mathcal{X}$ **by linear transformation:**

$$x_i x_i^T = \mathcal{R}_i(\mathcal{X}),\ x_i^T x_i = \mathcal{S}_i(\mathcal{X}).$$

It follows that the problem of interest is exactly equivalent to the problem

$$\mathrm{Opt} = \max_{\mathcal{X}} \left\{\mathrm{Tr}(\mathcal{A}\mathcal{X}) : \begin{array}{l}\mathrm{Tr}(\mathcal{Q}_\ell\mathcal{X}) \le 1\,\forall \ell\\ \mathcal{X} \succeq 0\\ \mathcal{R}_i(\mathcal{X}) \preceq I_{p_i}, \mathcal{S}_i(\mathcal{X}) \preceq I_{q_i}\,\forall i\\ \mathrm{Rank}(\mathcal{X}) = 1\end{array}\right\}.$$

Eliminating the rank constraint, we arrive at the SDP relaxation

$$\mathrm{Opt(SDP)} = \max_{\mathcal{X}} \left\{\mathrm{Tr}(\mathcal{A}\mathcal{X}) : \begin{array}{l}\mathrm{Tr}(\mathcal{Q}_\ell\mathcal{X}) \le 1\,\forall \ell\\ \mathcal{X} \succeq 0\\ \mathcal{R}_i(\mathcal{X}) \preceq I_{p_i}, \mathcal{S}_i(\mathcal{X}) \preceq I_{q_i}\,\forall i\end{array}\right\}.$$

$$\mathrm{Opt} = \max_{\mathcal{X}} \left\{ \mathrm{Tr}(\mathcal{A}\mathcal{X}) : \begin{array}{l} \mathrm{Tr}(\mathcal{Q}_\ell \mathcal{X}) \le 1 \,\forall \ell \\ \mathcal{X} \succeq 0 \\ \mathcal{R}_i(\mathcal{X}) \preceq I_{p_i}, \mathcal{S}_i(\mathcal{X}) \preceq I_{q_i} \,\forall i \\ \mathrm{Rank}(\mathcal{X}) = 1 \end{array} \right\}.$$

$$\mathrm{Opt}(\mathrm{SDP}) = \max_{\mathcal{X}} \left\{ \mathrm{Tr}(\mathcal{A}\mathcal{X}) : \begin{array}{l} \mathrm{Tr}(\mathcal{Q}_\ell \mathcal{X}) \le 1 \,\forall \ell \\ \mathcal{X} \succeq 0 \\ \mathcal{R}_i(\mathcal{X}) \preceq I_{p_i}, \mathcal{S}_i(\mathcal{X}) \preceq I_{q_i} \,\forall i \end{array} \right\}.$$

♠ **Applying the same approach as in the case when no norm constraints are present and invoking what was called "the best known provably true approximation of the Conjecture", we end up with the following result** (Math. Progr. 109:2-3 (2007), 283-317)**:**

Theorem: *One has*

$$\mathrm{Opt} \le \mathrm{Opt}(\mathrm{SDP}) \le \Omega \mathrm{Opt},$$
$$\Omega = O(1) \max\left[\max_i \min[p_i, q_i, (p_i + q_i)^{1/3}]; \ln(N); \ln(m)\right]$$

♣ **Example:** *Procrustes problem.* **Assume we are given** N **matrices** A_i **of common size** $p \times r$**. Interpreting the columns of** A_i **as points in** $\mathbf{R}^q$**,** A_i **becomes "a pattern" – an ordered collection of** r **points in** $\mathbf{R}^q$**. The problem is to rotate every pattern in order to bring them as close to each other as possible:**

Procrustes problem: *Given N matrices $A_i \in \mathbf{R}^{q\times r}$, find orthogonal $q\times q$ matrices x_i which minimize the quadratic objective*

$$F(x) = \sum_{i\neq j} \mathrm{Tr}\left([x_jA_j - x_iA_i]^T[x_jA_j - x_iA_i]\right)$$

♠ **When** x_i **are orthogonal, we have**

$$\begin{array}{rcl} F(x) & = & \sum_{i\neq j} \mathrm{Tr}\left([x_jA_j - x_iA_i]^T[x_jA_j - x_iA_i]\right) \\ & = & \sum_{i\neq j}[\mathrm{Tr}(A_j^TA_j) + \mathrm{Tr}(A_i^TA_i)] - G(x), \\ G(x) & = & 2\sum_{i\neq j} \mathrm{Tr}\left(A_j^Tx_j^Tx_iA_i\right) \end{array}$$

We see that the problem reduces to the form

$$\mathrm{Opt} = \max_{x=\{x_1,\dots,x_N\}} \left\{\langle x, \mathcal{G}x\rangle : x_i \in \mathbf{R}^{q\times q} \text{ are orthogonal}\right\}.$$

where $\mathcal{G}$ **is a symmetric matrix** *such that $\langle x, \mathcal{G}x\rangle$ is linear w.r.t. every x_i.* **This observation combines with the fact that the orthogonal matrices are exactly the extreme points of the matrix norm ball** $\{u \in \mathbf{R}^{q\times q} : \|u\| \leq 1\}$ **to imply that**

$$\mathrm{Opt} = \max_{x=\{x_1,\dots,x_N\}} \left\{\langle x, \mathcal{G}x\rangle : x_i \in \mathbf{R}^{q\times q}, \|x_i\| \leq 1, 1 \leq i \leq N\right\}.$$

Applying SDP relaxation, we arrive at a tractable problem (SDP) **such that**

$$\mathrm{Opt} \leq \mathrm{Opt(SDP)} \leq O(1)\max[q^{1/3}, \ln(N)]\mathrm{Opt}.$$

♠ **Numerical experiments reported in** (Math. Progr. 109:2-3 (2007), 283-317) **demonstrate surprisingly good practical performance of the relaxation.**