Semidefinite Programming with Applications in Geometry and Machine Learning

Alexandre d’Aspremont, CNRS & Ecole Polytechnique.

Support from NSF, ERC and Google.
A linear program (LP) is written

\[ \begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b \\
& \quad x \geq 0
\end{align*} \]

where \( x \geq 0 \) means that the coefficients of the vector \( x \) are nonnegative.

- Starts with Dantzig’s simplex algorithm in the late 40s.
- First efficient algorithm with polynomial complexity derived by Karmarkar [1984], using interior point methods.
A semidefinite program (SDP) is written

\[
\begin{align*}
\text{minimize} & \quad \text{Tr}(CX) \\
\text{subject to} & \quad \text{Tr}(A_i X) = b_i, \quad i = 1, \ldots, m \\
& \quad X \succeq 0
\end{align*}
\]

where $X \succeq 0$ means that the matrix variable $X \in S_n$ is positive semidefinite.

- Nesterov and Nemirovskii [1994] showed that the interior point algorithms used for linear programs could be extended to semidefinite programs.
- Key result: self-concordance analysis of Newton’s method (affine invariant smoothness bounds on the Hessian).
Introduction

- **Modeling**
  - Linear programming started as a toy problem in the 40s, many applications followed.
  - Semidefinite programming has much stronger expressive power, many new applications being investigated today (cf. this talk).
  - Similar conic duality theory.

- **Algorithms**
  - Robust solvers for solving large-scale linear programs are available today (e.g. MOSEK, CPLEX, GLPK).
  - Not (yet) true for semidefinite programs. Very active work now on first-order methods, motivated by applications in statistical learning (matrix completion, NETFLIX, structured MLE, . . . ).
Outline

- **Introduction**

- **Semidefinite programming**
  - Conic duality
  - A few words on algorithms

- **Recent applications**
  - Combinatorial relaxations
  - Ellipsoidal approximations
  - Distortion, embedding
  - Mixing rates for Markov chains & maximum variance unfolding
  - Moment problems & positive polynomials
  - Gordon-Slepian and the maximum of Gaussian processes

- **Dictionary metrics**
  - Sparse recovery conditions
  - Tractable performance bounds
Semidefinite Programming
Semidefinite programming: conic duality

Direct extension of LP duality results. Start from a semidefinite program

\[
\begin{align*}
\text{minimize} & \quad \text{Tr}(CX) \\
\text{subject to} & \quad \text{Tr}(A_i X) = b_i, \quad i = 1, \ldots, m \\
& \quad X \succeq 0
\end{align*}
\]

which is a convex minimization problem in \( X \in S_n \). The cone of positive semidefinite matrices is \textbf{self-dual}, i.e.

\[
Z \succeq 0 \iff \text{Tr}(ZX) \geq 0, \quad \text{for all } X \succeq 0,
\]

so we can form the \textbf{Lagrangian}

\[
L(X, y, Z) = \text{Tr}(CX) + \sum_{i=1}^{m} y_i (b_i - \text{Tr}(A_i X)) - \text{Tr}(ZX)
\]

with \textbf{Lagrange multipliers} \( y \in \mathbb{R}^m \) and \( Z \in S_n \) with \( Z \succeq 0 \).
Rearranging terms, we get

$$L(X, y, Z) = \text{Tr} \left( X \left( C - \sum_{i=1}^{m} y_i A_i - Z \right) \right) + b^T y$$

hence, after minimizing this affine function in $X \in S_n$, the dual can be written

maximize $b^T y$
subject to $Z = C - \sum_{i=1}^{m} y_i A_i$
$Z \succeq 0$,

which is another semidefinite program in the variables $y, Z$. Of course, the last two constraints can be simplified to

$$C - \sum_{i=1}^{m} y_i A_i \succeq 0.$$
Semidefinite programming: conic duality

- Primal dual pair

\[
\begin{align*}
\text{minimize} & \quad \text{Tr}(C X) \\
\text{subject to} & \quad \text{Tr}(A_i X) = b_i \\
& \quad X \succeq 0,
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad b^T y \\
\text{subject to} & \quad C - \sum_{i=1}^{m} y_i A_i \succeq 0.
\end{align*}
\]

- Simple constraint qualification conditions guarantee \text{strong duality}.

- We can write a conic version of the KKT optimality conditions

\[
\begin{cases}
C - \sum_{i=1}^{m} y_i A_i = Z, \\
\text{Tr}(A_i X) = b_i, \quad i = 1, \ldots, m, \\
\text{Tr}(XZ) = 0, \\
X, Z \succeq 0.
\end{cases}
\]
So what?

- Weak duality produces simple bounds on e.g. combinatorial problems.
- Consider the MAXCUT relaxation

\[
\begin{align*}
\text{max.} \quad & x^T C x \\
\text{s.t.} \quad & x_i^2 = 1 \\
\end{align*}
\]

is bounded by

\[
\begin{align*}
\text{max.} \quad & \text{Tr}(X C') \\
\text{s.t.} \quad & \text{diag}(X) = 1 \\
& X \succeq 0,
\end{align*}
\]

in the variables \(x \in \mathbb{R}^n\) and \(X \in \mathbb{S}_n\) (more later on these relaxations).

- The dual of the SDP on the right is written

\[
\min_y n \lambda_{\text{max}}(C - \text{diag}(y)) + 1^T y
\]

in the variable \(y \in \mathbb{R}^n\).

- By **weak duality**, plugging **any** value \(y\) in this problem will produce an upper bound on the optimal value of the combinatorial problem above.
Semidefinite programming: algorithms

**Algorithms** for semidefinite programming

- Following [Nesterov and Nemirovskii, 1994], most of the attention was focused on interior point methods.
- Newton’s method, with efficient linear algebra solving for the search direction.
- Fast, and robust on small problems \((n \sim 500)\).
- Computing the Hessian is too hard on larger problems.

**Solvers**

- Open source solvers: SDPT3, SEDUMI, SDPA, CSDP, . . .
- Very powerful modeling systems: CVX
Solving the maxcut relaxation

\[ \begin{align*}
\text{max.} & \quad \text{Tr}(XC) \\
\text{s.t.} & \quad \text{diag}(X) = 1 \\
& \quad X \succeq 0,
\end{align*} \]

is written as follows in CVX/MATLAB

```matlab
cvx_begin
    variable X(n,n) symmetric
    maximize trace(C*X)
    subject to
        diag(X)==1
        X==semidefinite(n)
cvx_end
```
Solving large-scale problems is a bit more problematic. . .

- No universal algorithm known yet. No CVX like modeling system.
- Performance and algorithmic choices heavily depends on problem structure.
- Very basic codes only require computing one leading eigenvalue per iteration, with complexity $O(n^2 \log n)$ using e.g. Lanczos.
- Each iteration requires about 300 matrix vector products, but making progress may require many iterations. Typically $O(1/\epsilon^2)$ or $O(1/\epsilon)$ in some cases.
- In general, most optimization algorithms are purely sequential, so only the linear algebra subproblems benefit from the multiplication of CPU cores.
Outline

- Introduction
- Semidefinite programming
  - Conic duality
  - A few words on algorithms
- Recent applications
  - Combinatorial relaxations
  - Distortion, embedding
  - Ellipsoidal approximations
  - Mixing rates for Markov chains & maximum variance unfolding
  - Moment problems & positive polynomials
  - Gordon-Slepian and the maximum of Gaussian processes
- Dictionary metrics
  - Sparse recovery conditions
  - Tractable performance bounds
Applications

- Many classical problems can be cast as or approximated by semidefinite programs.
- Recognizing this is not always obvious.
- At reasonable scales, numerical solutions often significantly improve on classical closed-form bounds.
- A few examples follow...
Combinatorial relaxations
Combinatorial relaxations


Semidefinite programs with constant trace often arise in \textit{convex relaxations} of combinatorial problems. Use MAXCUT as an example here.

The problem is written

\[
\max \quad x^\top C x \\
\text{s.t.} \quad x \in \{-1,1\}^n
\]

in the binary variables \( x \in \{-1,1\}^n \), with parameter \( C \in \mathcal{S}_n \) (usually \( C \succeq 0 \)). This problem is known to be \textbf{NP-Hard}. Using

\[\begin{align*}
x \in \{-1,1\}^n & \iff x_i^2 = 1, \quad i = 1, \ldots, n
\end{align*}\]

we get

\[
\max \quad x^\top C x \\
\text{s.t.} \quad x_i^2 = 1, \quad i = 1, \ldots, n
\]

which is a nonconvex quadratic program in the variable \( x \in \mathbb{R}^n \).
Combinatorial relaxations

We now do a simple change of variables, setting $X = xx^T$, with

$$X = xx^T \iff X \in S_n, \quad X \succeq 0, \quad \text{Rank}(X) = 1$$

and we also get

$$\text{Tr}(CX) = x^T Cx$$
$$\text{diag}(X) = 1 \iff x_i^2 = 1, \quad i = 1, \ldots, n$$

so the original combinatorial problem is equivalent to

$$\max. \quad \text{Tr}(CX)$$
$$\text{s.t.} \quad \text{diag}(X) = 1$$
$$X \succeq 0, \quad \text{Rank}(X) = 1$$

which is now a nonconvex problem in $X \in S_n$. 

A. d’Aspremont
Combinatorial relaxations

- If we simply drop the rank constraint, we get the following relaxation:

\[
\begin{align*}
\text{max.} & \quad x^T C x \\
\text{s.t.} & \quad x \in \{-1, 1\}^n
\end{align*}
\]

is bounded by

\[
\begin{align*}
\text{max.} & \quad \text{Tr}(CX) \\
\text{s.t.} & \quad \text{diag}(X) = 1 \\
& \quad X \succeq 0,
\end{align*}
\]

which is a semidefinite program in \( X \in \mathbb{S}_n \).

- **Rank constraints** in semidefinite programs are usually hard. All semi-algebraic optimization problems can be formulated as rank constrained SDPs.

- Randomization techniques produce bounds on the approximation ratio. When \( C \succeq 0 \) for example, we have

\[
\frac{2}{\pi} \text{SDP} \leq \text{OPT} \leq \text{SDP}
\]

for the MAXCUT relaxation (more details in [Ben-Tal and Nemirovski, 2001]).

- Applications in graph, matrix approximations (CUT-Norm, \( \| \cdot \|_{2 \to 1} \)) [Frieze and Kannan, 1999, Alon and Naor, 2004, Nemirovski, 2005]
Distortion, embedding problems, . . .
We cannot hope to always get low rank solutions, unless we are willing to admit some distortion. The following result from [Ben-Tal, Nemirovski, and Roos, 2003] gives some guarantees.

**Theorem**

**Approximate S-lemma.** Let $A_1, \ldots, A_N \in S_n$, $\alpha_1, \ldots, \alpha_N \in \mathbb{R}$ and a matrix $X \in S_n$ such that

$$A_i, X \succeq 0, \quad \text{Tr}(A_i X) = \alpha_i, \quad i = 1, \ldots, N$$

Let $\epsilon > 0$, there exists a matrix $X_0$ such that

$$\alpha_i (1 - \epsilon) \leq \text{Tr}(A_i X_0) \leq \alpha_i (1 + \epsilon) \quad \text{and} \quad \text{Rank}(X_0) \leq 8 \frac{\log 4N}{\epsilon^2}$$

**Proof.** Randomization, concentration results on Gaussian quadratic forms.

A particular case: Given $N$ vectors $v_i \in \mathbb{R}^d$, construct their Gram matrix $X \in S_N$, with

$$X \succeq 0, \quad X_{ii} - 2X_{ij} + X_{jj} = \|v_i - v_j\|_2^2, \quad i, j = 1, \ldots, N.$$  

The matrices $D_{ij} \in S_n$ such that

$$\text{Tr}(D_{ij} X) = X_{ii} - 2X_{ij} + X_{jj}, \quad i, j = 1, \ldots, N$$

satisfy $D_{ij} \succeq 0$. Let $\epsilon > 0$, there exists a matrix $X_0$ with

$$m = \text{Rank}(X_0) \leq 16 \frac{\log 2N}{\epsilon^2},$$

from which we can extract vectors $u_i \in \mathbb{R}^m$ such that

$$\|v_i - v_j\|_2^2 (1 - \epsilon) \leq \|u_i - u_j\|_2^2 \leq \|v_i - v_j\|_2^2 (1 + \epsilon).$$

In this setting, the **Johnson-Lindenstrauss** lemma is a particular case of the approximate $S$ lemma...
Distortion, embedding problems, . . .

The problem of reconstructing an $N$-point Euclidean metric, given partial information on pairwise distances between points $v_i$, $i = 1, \ldots, N$ can also be cast as an SDP, known as and Euclidean Distance Matrix Completion problem.

$$\begin{align*}
\text{find} & \quad D \\
\text{subject to} & \quad 1v^T + v1^T - D \succeq 0 \\
& \quad D_{ij} = \|v_i - v_j\|_2^2, \quad (i, j) \in S \\
& \quad v \geq 0
\end{align*}$$

in the variables $D \in S_n$ and $v \in \mathbb{R}^n$, on a subset $S \subset [1, N]^2$.

We can add further constraints to this problem given additional structural info on the configuration.

Applications in sensor networks, molecular conformation reconstruction etc. . .
[Dattorro, 2005] 3D map of the USA reconstructed from pairwise distances on 5000 points. Distances reconstructed from Latitude/Longitude data.
Distortion, embedding problems, . . .

**Theorem**

**Embedding.** [Bourgain, 1985] Every $n$-point metric space $(X, d)$ can be embedded in an $O(\log n)$-dimensional Euclidean space with an $O(\log n)$ distortion.

Let $(X = \{x_1, \ldots, x_n\}, d)$ be a finite metric space, we can find the **minimum distortion embedding** by solving

\[
\begin{align*}
\text{minimize} & \quad z \\
\text{subject to} & \quad X_{ii} + X_{jj} - 2X_{ij} \geq d(x_i, x_j)^2 \\
& \quad X_{ii} + X_{jj} - 2X_{ij} \leq zd(x_i, x_j)^2, \quad i, j = 1, \ldots, n \\
& \quad X \succeq 0
\end{align*}
\]

in the variables $X \in S_n$ and $z \in \mathbb{R}$.

The result above shows $\sqrt{z^*}$ is $O(\log n)$ in the worst case.
Ellipsoidal approximations
Ellipsoidal approximations

Minimum volume ellipsoid $E$ s.t. $C \subseteq E$ (Löwner-John ellipsoid).

- parametrize $E$ as $E = \{v \mid \|Av + b\|_2 \leq 1\}$ with $A \succ 0$.
- $\text{vol } E$ is proportional to $\det A^{-1}$; to compute minimum volume ellipsoid,

$$\begin{align*}
\text{minimize (over } A, b \text{)} & \quad \log \det A^{-1} \\
\text{subject to } & \quad \sup_{v \in C} \|Av + b\|_2 \leq 1
\end{align*}$$

convex, but the constraint can be hard (for general sets $C$).

Finite set $C = \{x_1, \ldots, x_m\}$, or polytope with polynomial number of vertices:

$$\begin{align*}
\text{minimize (over } A, b \text{)} & \quad \log \det A^{-1} \\
\text{subject to } & \quad \|Ax_i + b\|_2 \leq 1, \quad i = 1, \ldots, m
\end{align*}$$

also gives Löwner-John ellipsoid for polyhedron $C \circ \{x_1, \ldots, x_m\}$.

Similar result for the maximum volume inscribed ellipsoid when $C$ is a polyhedron given by its facets $\{x \mid a_i^T x \leq b_i, \quad i = 1, \ldots, m\}$. 
**Ellipsoidal approximations**

**D-Optimal Experiment Design.**

Given experiment vectors \( v_i \in \mathbb{R}^n \), we minimize the MLE estimation error in

\[
y_j = v_j^T x + w_j
\]

for Gaussian noise \( w \). Assuming we run \( \lambda_i \) times experiment \( v_i \), the covariance matrix of the estimation error \( x - \hat{x} \) is given by \( \sum_{i=1}^{p} \lambda_i v_i v_i^T \) and we solve

\[
\begin{align*}
\text{minimize} & \quad \log \det \left( \sum_{i=1}^{p} \lambda_i v_i v_i^T \right)^{-1} \\
\text{subject to} & \quad 1^T \lambda = 1, \lambda \geq 0
\end{align*}
\]

in the variable \( \lambda \in \mathbb{R}^p \). The dual of this last problem is written

\[
\begin{align*}
\text{minimize} & \quad \log \det(W)^{-1} \\
\text{subject to} & \quad v_i^T W v_i \leq 1
\end{align*}
\]

which is a minimum volume ellipsoid problem in the variable \( W \in \mathbb{S}_n \).
Ellipsoidal approximations

$C \subseteq \mathbb{R}^n$ convex, bounded, with nonempty interior

- Löwner-John ellipsoid, shrunk by a factor $n$, lies inside $C$
- maximum volume inscribed ellipsoid, expanded by a factor $n$, covers $C$

**example** (for two polyhedra in $\mathbb{R}^2$)

factor $n$ can be improved to $\sqrt{n}$ if $C$ is symmetric. See [Boyd and Vandenberghe, 2004] for further examples.
Mixing rates for Markov chains & maximum variance unfolding
Let $G = (V, E)$ be an **undirected graph** with $n$ vertices and $m$ edges.

We define a **Markov chain** on this graph, and let $w_{ij} \geq 0$ be the transition rate for edge $(i, j) \in V$.

Let $\pi(t)$ be the state distribution at time $t$, its evolution is governed by the heat equation

$$d\pi(t) = -L\pi(t)dt$$

with

$$L_{ij} = \begin{cases} 
-w_{ij} & \text{if } i \neq j, (i, j) \in V \\
0 & \text{if } (i, j) \notin V \\
\sum_{(i,k) \in V} w_{ik} & \text{if } i = j
\end{cases}$$

the **graph Laplacian** matrix, which means

$$\pi(t) = e^{-Lt}\pi(0).$$

The matrix $L \in S_n$ satisfies $L \succeq 0$ and its smallest eigenvalue is zero (associated with the uniform distribution).
Mixing rates for Markov chains & unfolding

- With

\[ \pi(t) = e^{-Lt}\pi(0) \]

the **mixing rate** is controlled by the second smallest eigenvalue \( \lambda_2(L) \).

- Since the smallest eigenvalue of \( L \) is zero, with eigenvector \( \mathbf{1} \), we have

\[ \lambda_2(L) \geq t \iff L(w) \succeq t(\mathbf{1} - (1/n)\mathbf{1}\mathbf{1}^T), \]

- Maximizing the mixing rate of the Markov chain means solving

\[
\begin{align*}
\text{maximize} & \quad t \\
\text{subject to} & \quad L(w) \succeq t(\mathbf{1} - (1/n)\mathbf{1}\mathbf{1}^T) \\
& \quad \sum_{(i,j) \in V} d_{ij}^2 w_{ij} \leq 1 \\
& \quad w \geq 0
\end{align*}
\]

in the variable \( w \in \mathbb{R}^m \), with (normalization) parameters \( d_{ij}^2 \geq 0 \).

- Since \( L(w) \) is an affine function of the variable \( w \in \mathbb{R}^m \), this is a semidefinite program in \( w \in \mathbb{R}^m \).

- Numerical solution usually performs better than **Metropolis-Hastings**.
We can also form the dual of the maximum MC mixing rate problem. The dual means solving

\[
\begin{align*}
\text{maximize} \quad & \quad \text{Tr}(X(I - (1/n)11^T)) \\
\text{subject to} \quad & \quad X_{ii} - 2X_{ij} + X_{jj} \leq d_{ij}^2, \quad (i, j) \in V \\
& \quad X \succeq 0,
\end{align*}
\]

in the variable \( X \in S_n \).

Here too, we can interpret \( X \) as the gram matrix of a set of \( n \) vectors \( v_i \in \mathbb{R}^d \). The program above maximizes the variance of the vectors \( v_i \)

\[
\text{Tr}(X(I - (1/n)11^T)) = \sum_i \|v_i\|_2^2 - \|\sum_i v_i\|_2^2
\]

while the constraints bound pairwise distances

\[
X_{ii} - 2X_{ij} + X_{jj} \leq d_{ij}^2 \iff \|v_i - v_j\|_2^2 \leq d_{ij}^2
\]

This is a maximum variance unfolding problem [Weinberger and Saul, 2006, Sun et al., 2006].
From [Sun et al., 2006]: we are given pairwise 3D distances for $k$-nearest neighbors in the point set on the right. We plot the maximum variance point set satisfying these pairwise distance bounds on the right.
Moment problems & positive polynomials
[Nesterov, 2000]. Hilbert’s 17th problem has a positive answer for univariate polynomials: a polynomial is nonnegative iff it is a **sum of squares**

\[ p(x) = x^{2d} + \alpha_{2d-1}x^{2d-1} + \ldots + \alpha_0 \geq 0, \text{ for all } x \iff p(x) = \sum_{i=1}^{N} q_i(x)^2 \]

We can formulate this as a linear matrix inequality, let \( v(x) \) be the moment vector

\[ v(x) = (1, x, \ldots, x^d)^T \]

we have

\[ \sum_{i} \lambda_i u_i u_i^T = M \succeq 0 \iff p(x) = v(x)^T M v(x) = \sum_{i} \lambda_i (u_i^T v(x))^2 \]

where \((\lambda_i, u_i)\) are the eigenpairs of \(M\).
The dual to the cone of Sum-of-Squares polynomials is the cone of moment matrices

\[ E_\mu[x^i] = q_i, \ i = 0, \ldots, d \iff \begin{pmatrix} q_0 & q_1 & \cdots & q_d \\ q_1 & q_2 & \cdots & q_{d+1} \\ \vdots & \vdots & \ddots & \vdots \\ q_d & q_{d+1} & \cdots & q_{2d} \end{pmatrix} \succeq 0 \]


This forms exponentially large, ill-conditioned semidefinite programs however.
Gordon-Slepian
and the maximum of Gaussian processes
Let \( x \sim \mathcal{N}(0, X) \) and \( y \sim \mathcal{N}(0, Y) \) be two Gaussian processes such that

\[
\begin{align*}
X_{ij} &\leq Y_{ij} \\
X_{ii} &= Y_{ii}, \quad i, j = 1, \ldots, n,
\end{align*}
\]

then

\[
E \left[ \prod_{i=1}^{N} f(x_i) \right] \leq E \left[ \prod_{i=1}^{N} f(y_i) \right]
\]

for every nonnegative and nonincreasing differentiable \( f \) such that \( f \) and \( f' \) are bounded on \( \mathbb{R} \). This is first order stochastic dominance.

This implies in particular that

\[
E \left[ \sup_{i=1,\ldots,N} y_i \right] \leq E \left[ \sup_{i=1,\ldots,N} x_i \right]
\]
Gordon-Slepian & the max. of Gaussian processes

Lemma

**Simple bound on Gaussian processes.** Let \( y \sim \mathcal{N}(0, Y) \) be a Gaussian vector with covariance \( Y \in S_n \). Suppose \( X \succeq 0 \) satisfies

\[
Y_{ii} - 2Y_{ij} + Y_{jj} \leq X_{ii} - 2X_{ij} + X_{jj}, \quad i, j = 1, \ldots, n
\]

which are convex inequalities in \( X \in S_n \), then

\[
E \left[ \sup_{i=1,\ldots,n} y_i \right] \leq 2 \left( \text{Rank}(X) \max_{i=1,\ldots,n} X_{ii} \right)^{1/2}
\]

**Proof.** Use Gordon-Slepian together with Cauchy inequality.

Write \( x = Vg \) with \( V \in \mathbb{R}^{n \times k} \) such that \( X = VV^T \) and \( k = \text{Rank}(X) \), so

\[
\sup_{i=1,\ldots,n} x_i = \sup_{i=1,\ldots,n} \sum_{j=1}^{k} V_{ij} g_j \leq \left( \sup_{i=1,\ldots,n} \|V_i\|_2 \right) \|g\|_2 \quad \text{where } g \text{ i.i.d. Gaussian, with } E[\|g\|_2] \leq \sqrt{k} \text{ and } \|V_i\|_2 = X_{ii}.\]
Proposition

SDP bounds on Gaussian processes. Let $y \sim \mathcal{N}(0, Y)$ be a Gaussian vector with covariance $Y \in \mathbb{S}_n$. From a matrix $X \in \mathbb{S}_n$ satisfying

$$Y_{ii} - 2Y_{ij} + Y_{jj} \leq (X_{ii} - 2X_{ij} + X_{jj}), \quad i, j = 1, \ldots, n$$

which are convex inequalities in $X \in \mathbb{S}_n$, we can construct a low rank matrix $X^r$ such that

$$E \left[ \sup_{i=1,\ldots,n} y_i \right] \leq c_2 \sqrt{\log n} \left( \max_{i=1,\ldots,n} \sqrt{X_{ii}} \right)$$

where $c_1, c_2 > 0$ are absolute constants.

Proof. Combine previous lemma with approximate $S$-lemma. Explicit randomized procedure to find $X^r$. 
The previous result shows that the SDP bound always does as well as the classical bound

$$
E \left[ \sup_{i=1,\ldots,n} y_i \right] \leq 2\sqrt{\log n} \left( \max_{i=1,\ldots,n} \sqrt{Y_{ii}} \right),
$$

up to a multiplicative constant.

We can exploit symmetries in the matrix $Y$ to block diagonalize the SDP and reduce its complexity. [Gatermann and Parrilo, 2002, Vallentin, 2009]
Conclusion

- Semidefinite programming formulation to some problems in geometry, probability, statistics.

- Improvements over classical closed-form bounds on reasonably large problems.

Next: applications to **compressed sensing**...
References


