Semidefinite Programming with Applications in Geometry and Machine Learning

Alexandre d'Aspremont, *CNRS & Ecole Polytechnique*.

Support from NSF, ERC and Google.

Introduction

A **linear program** (LP) is written

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

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 proofs of polynomial complexity by Nemirovskii and Yudin [1979] and Khachiyan [1979] using the ellipsoid method.
- First efficient algorithm with polynomial complexity derived by Karmarkar [1984], using interior point methods.

Introduction

A **semidefinite program** (SDP) is written

$$\begin{array}{ll}\text{minimize} & \mathbf{Tr}(CX)\\ \text{subject to} & \mathbf{Tr}(A_iX) = b_i, \quad i = 1, \ldots, m\\ & X \succeq 0\end{array}$$

where $X \succeq 0$ means that the matrix variable $X \in \mathbf{S}_n$ 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
- Dicitionary metrics
  - Sparse recovery conditions
  - Tractable performance bounds

Semidefinite Programming

Semidefinite programming: conic duality

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

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(CX) \\ \text{subject to} & \mathbf{Tr}(A_i X) = b_i, \quad i = 1, \ldots, m \\ & X \succeq 0 \end{array}$$

which is a convex minimization problem in $X \in \mathbf{S}_n$. The cone of positive semidefinite matrices is **self-dual**, i.e.

$$Z \succeq 0 \quad \Longleftrightarrow \quad \mathbf{Tr}(ZX) \geq 0, \text{ for all } X \succeq 0,$$

so we can form the **Lagrangian**

$$L(X, y, Z) = \mathbf{Tr}(CX) + \sum_{i=1}^{m} y_i \left(b_i - \mathbf{Tr}(A_i X)\right) - \mathbf{Tr}(ZX)$$

with **Lagrange multipliers** $y \in \mathbb{R}^m$ and $Z \in \mathbf{S}_n$ with $Z \succeq 0$.

Semidefinite programming: conic duality

Rearranging terms, we get

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

hence, after minimizing this affine function in $X \in \mathbf{S}_n$, the **dual** can be written

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

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

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(CX) \\ \text{subject to} & \mathbf{Tr}(A_iX) = b_i \\ & X \succeq 0, \end{array} \qquad \begin{array}{ll} \text{maximize} & b^Ty \\ \text{subject to} & C - \sum_{i=1}^m y_iA_i \succeq 0. \end{array}$$

- Simple constraint qualification conditions guarantee **strong duality**.
- We can write a conic version of the KKT optimality conditions

$$\left\{ \begin{array}{rcll} C - \sum_{i=1}^m y_iA_i & = & Z, & \\ \mathbf{Tr}(A_iX) & = & b_i, & i = 1, \ldots, m, \\ \mathbf{Tr}(XZ) & = & 0, & \\ X, Z & \succeq & 0. & \end{array} \right.$$

Semidefinite programming: conic duality

So what?

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

$$\begin{array}{ll} \text{max.} & x^T C x \\ \text{s.t.} & x_i^2 = 1 \end{array} \quad \textbf{is bounded by} \quad \begin{array}{ll} \text{max.} & \mathbf{Tr}(XC) \\ \text{s.t.} & \mathbf{diag}(X) = \mathbf{1} \\ & X \succeq 0, \end{array}$$

  in the variables $x \in \mathbb{R}^n$ and $X \in \mathbf{S}_n$ (more later on these relaxations).
- The dual of the SDP on the right is written

$$\min_y \; n\lambda_{\max}(C - \mathbf{diag}(y)) + \mathbf{1}^T y$$

  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

Semidefinite programming: CVX

Solving the maxcut relaxation

$$\begin{array}{ll} \text{max.} & \mathbf{Tr}(XC) \\ \text{s.t.} & \mathbf{diag}(X) = \mathbf{1} \\ & X \succeq 0, \end{array}$$

is written as follows in CVX/MATLAB

```
cvx_begin
.  variable X(n,n) symmetric
.  maximize trace(C*X)
.  subject to
.    diag(X)==1
.    X==semidefinite(n)
cvx_end
```

Semidefinite programming: large-scale

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
- Dicitionary 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

[Goemans and Williamson, 1995, Nesterov, 1998]

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

The problem is written

$$\begin{array}{ll} \text{max.} & x^T C x \\ \text{s.t.} & x \in \{-1, 1\}^n \end{array}$$

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

$$x \in \{-1, 1\}^n \quad \Longleftrightarrow \quad x_i^2 = 1, \quad i = 1, \ldots, n$$

we get

$$\begin{array}{ll} \text{max.} & x^T C x \\ \text{s.t.} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

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

$$X = xx^T \quad \Longleftrightarrow \quad X \in \mathbf{S}_n,\ X \succeq 0,\ \mathbf{Rank}(X) = 1$$

and we also get

$$\begin{aligned} &\mathbf{Tr}(CX) = x^T C x \\ &\mathbf{diag}(X) = \mathbf{1} \quad \Longleftrightarrow \quad x_i^2 = 1, \quad i = 1, \ldots, n \end{aligned}$$

so the original combinatorial problem is equivalent to

$$\begin{array}{ll} \text{max.} & \mathbf{Tr}(CX) \\ \text{s.t.} & \mathbf{diag}(X) = 1 \\ & X \succeq 0,\ \mathbf{Rank}(X) = 1 \end{array}$$

which is now a nonconvex problem in $X \in \mathbf{S}_n$.

Combinatorial relaxations

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

$$\begin{array}{ll}\text{max.} & x^TCx\\ \text{s.t.} & x\in\{-1,1\}^n\end{array} \qquad \textbf{is bounded by} \qquad \begin{array}{ll}\text{max.} & \mathbf{Tr}(CX)\\ \text{s.t.} & \mathbf{diag}(X)=1\\ & X\succeq 0,\end{array}$$

  which is a semidefinite program in $X \in \mathbf{S}_n$.

- **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

$$\frac{2}{\pi}SDP \leq OPT \leq SDP$$

  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, . . .

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

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

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

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

Proof. Randomization, concentration results on Gaussian quadratic forms.

See [Barvinok, 2002, Ben-Tal, El Ghaoui, and Nemirovski, 2009] for more details.

Distortion, embedding problems, . . .

A particular case: Given N vectors $v_i \in \mathbb{R}^d$, construct their Gram matrix $X \in \mathbf{S}_N$, 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 $D_{ij} \in \mathbf{S}_n$ such that

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

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

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

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 $\mathcal{S}$ 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.

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

  in the variables $D \in \mathbf{S}_n$ 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. . .

Distortion, embedding problems, . . .

[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

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

in the variables $X \in \mathbf{S}_n$ 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 $\mathcal{E}$ s.t. $C \subseteq \mathcal{E}$ (Löwner-John ellipsoid).

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

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

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**:

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

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

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

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 λ_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

$$\begin{array}{ll} \mbox{minimize} & \log\det\left(\sum_{i=1}^p \lambda_i v_i v_i^T\right)^{-1} \\ \mbox{subject to} & \mathbf{1}^T\lambda = 1,\ \lambda \geq 0 \end{array}$$

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

$$\begin{array}{ll} \mbox{minimize} & \log\det(W)^{-1} \\ \mbox{subject to} & v_i^T W v_i \leq 1 \end{array}$$

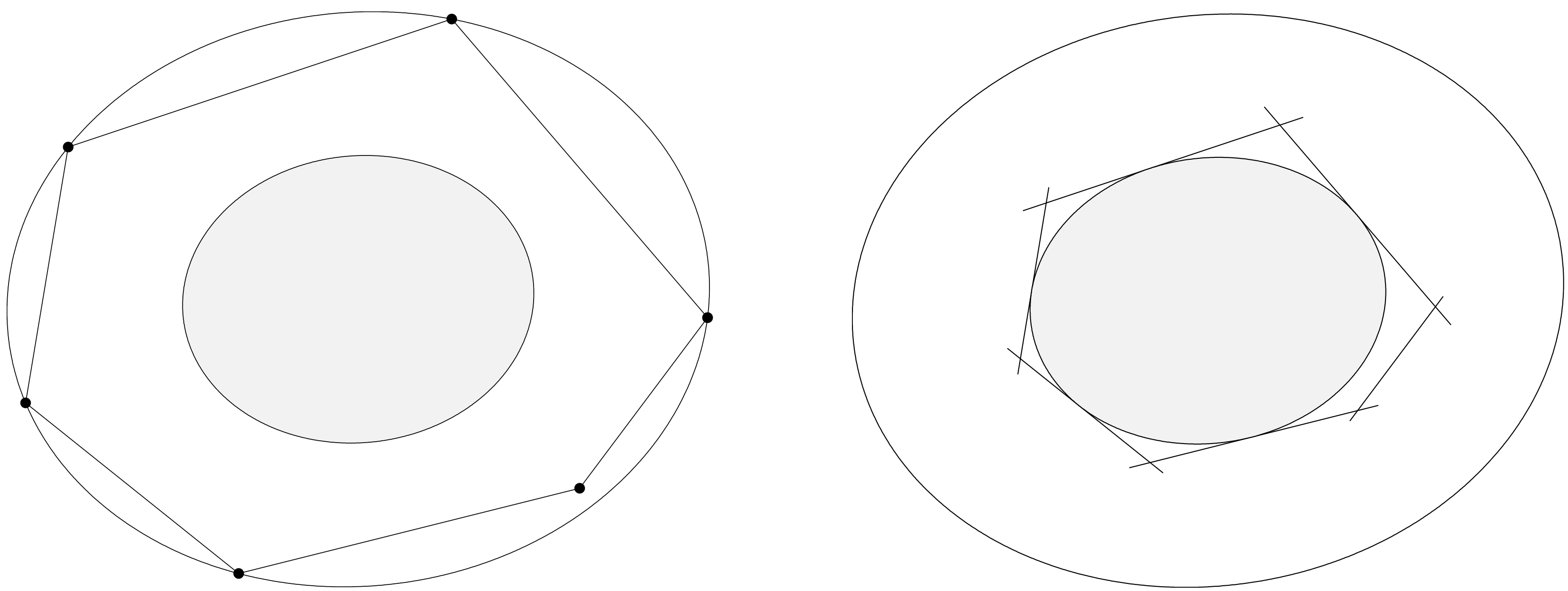
which is a minimum volume ellipsoid problem in the variable $W \in \mathbf{S}_n$.

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

Mixing rates for Markov chains & unfolding

[Sun, Boyd, Xiao, and Diaconis, 2006]

- 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 $L \in \mathbf{S}_n$ 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
$$\lambda_2(L) \geq t \quad \Longleftrightarrow \quad L(w) \succeq t(\mathbf{I} - (1/n)\mathbf{1}\mathbf{1}^T),$$
- Maximizing the mixing rate of the Markov chain means solving
$$\begin{array}{ll} \text{maximize} & t \\ \text{subject to} & L(w) \succeq t(\mathbf{I} - (1/n)\mathbf{1}\mathbf{1}^T) \\ & \sum_{(i,j)\in V} d_{ij}^2 w_{ij} \leq 1 \\ & w \geq 0 \end{array}$$
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**.

Mixing rates for Markov chains & unfolding

- We can also form the **dual** of the maximum MC mixing rate problem.
- The dual means solving

$$\begin{array}{ll} \mbox{maximize} & \mathbf{Tr}(X(\mathbf{I}-(1/n)\mathbf{1}\mathbf{1}^T)) \\ \mbox{subject to} & X_{ii}-2X_{ij}+X_{jj} \leq d_{ij}^2, \quad (i,j)\in V \\ & X \succeq 0, \end{array}$$

  in the variable $X \in \mathbf{S}_n$.
- 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

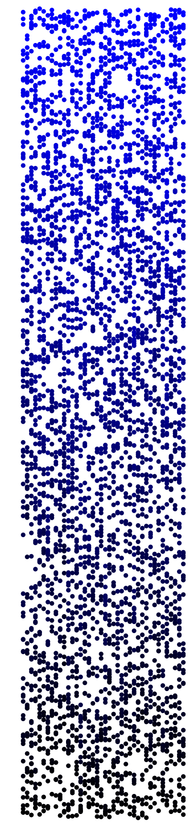
$$\mathbf{Tr}(X(\mathbf{I}-(1/n)\mathbf{1}\mathbf{1}^T)) = \textstyle\sum_i \|v_i\|_2^2 - \|\sum_i v_i\|_2^2$$

  while the constraints bound pairwise distances

$$X_{ii}-2X_{ij}+X_{jj} \leq d_{ij}^2 \quad \Longleftrightarrow \quad \|v_i - v_j\|_2^2 \leq d_{ij}^2$$

- This is a **maximum variance unfolding problem** [Weinberger and Saul, 2006, Sun et al., 2006].

Mixing rates for Markov chains & unfolding



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

Moment problems & positive polynomials

[Nesterov, 2000]. Hilbert's 17^{th} problem has a positive answer for univariate polynomials: a polynomial is nonnegative iff it is a **sum of squares**

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

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

$$\sum_i \lambda_i u_i u_i^T = M \succeq 0 \quad \Longleftrightarrow \quad p(x) = v(x)^T M v(x) = \sum_i \lambda_i (u_i^T v(x))^2$$

where (λ_i, u_i) are the eigenpairs of M.

Moment problems & positive polynomials

- The dual to the cone of Sum-of-Squares polynomials is the cone of moment matrices

$$\mathbf{E}_\mu[x^i] = q_i,\ i = 0,\ldots,d \quad \Longleftrightarrow \quad \begin{pmatrix} q_0 & q_1 & \cdots & q_d \\ q_1 & q_2 & & q_{d+1} \\ \vdots & & \ddots & \vdots \\ q_d & q_{d+1} & \cdots & q_{2d} \end{pmatrix} \succeq 0$$

- [Putinar, 1993, Lasserre, 2001, Parrilo, 2000] These results can be extended to multivariate polynomial optimization problems over compact semi-algebraic sets.
- This forms exponentially large, ill-conditioned semidefinite programs however.

Gordon-Slepian and the maximum of Gaussian processes

Gordon-Slepian & the max. of Gaussian processes

[Massart, 2007]

- Let $x \sim \mathcal{N}(0, X)$ and $y \sim \mathcal{N}(0, Y)$ be two Gaussian processes such that

$$\begin{cases} X_{ij} \leq Y_{ij} \\ X_{ii} = Y_{ii}, \quad i,j = 1,\ldots,n, \end{cases}$$

  then

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

  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

$$\mathbf{E}\left[\sup_{i=1,\ldots,N} y_i\right] \leq \mathbf{E}\left[\sup_{i=1,\ldots,N} x_i\right]$$

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 $Y \in \mathbf{S}_n$. 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 $X \in \mathbf{S}_n$, then

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

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

Gordon-Slepian & the max. of Gaussian processes

Proposition

SDP bounds on Gaussian processes. *Let* $y \sim \mathcal{N}(0, Y)$ *be a Gaussian vector with covariance* $Y \in \mathbf{S}_n$. *From a matrix* $X \in \mathbf{S}_n$ *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 $X \in \mathbf{S}_n$, *we can construct a low rank matrix* X^r *such that*

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

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

Proof. Combine previous lemma with approximate $\mathcal{S}$-lemma. Explicit randomized procedure to find X^r.

Gordon-Slepian & the max. of Gaussian processes

- The previous result shows that the SDP bound always does as well as the classical bound

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

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

N. Alon and A. Naor. Approximating the cut-norm via grothendieck's inequality. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 72–80. ACM, 2004.

A. Barvinok. *A course in convexity*. American Mathematical Society, 2002.

A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization : analysis, algorithms, and engineering applications*. MPS-SIAM series on optimization. Society for Industrial and Applied Mathematics : Mathematical Programming Society, Philadelphia, PA, 2001.

A. Ben-Tal, A. Nemirovski, and C. Roos. Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM Journal on Optimization*, 13(2):535–560, 2003. ISSN 1052-6234.

A. Ben-Tal, L. El Ghaoui, and A.S. Nemirovski. *Robust optimization*. Princeton University Press, 2009.

J. Bourgain. On lipschitz embedding of finite metric spaces in hilbert space. *Israel Journal of Mathematics*, 52(1):46–52, 1985.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

J. Dattorro. *Convex optimization & Euclidean distance geometry*. Meboo Publishing USA, 2005.

A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

Karin Gatermann and P. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. Technical Report arXiv math.AC/0211450, ETH Zurich, 2002.

M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42:1115–1145, 1995.

N. K. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

L. G. Khachiyan. A polynomial algorithm in linear programming (in Russian). *Doklady Akademiia Nauk SSSR*, 224:1093–1096, 1979.

J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

P. Massart. Concentration inequalities and model selection. *Ecole d'Eté de Probabilités de Saint-Flour XXXIII*, 2007.

A.S. Nemirovski. *Computation of matrix norms with applications to Robust Optimization*. PhD thesis, Technion, 2005.

A. Nemirovskii and D. Yudin. Problem complexity and method efficiency in optimization. *Nauka (published in English by John Wiley, Chichester, 1983)*, 1979.

Y. Nesterov. *Global quadratic optimization via conic relaxation*. Number 9860. CORE Discussion Paper, 1998.

Y. Nesterov. Squared functional systems and optimization problems. Technical Report 1472, CORE reprints, 2000.

Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*. Society for Industrial and Applied Mathematics, Philadelphia, 1994.

P. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, 2000.

M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993.

J. Sun, S. Boyd, L. Xiao, and P. Diaconis. The fastest mixing Markov process on a graph and a connection to a maximum variance unfolding problem. *SIAM Review*, 48(4):681–699, 2006.

F. Vallentin. Symmetry in semidefinite programs. *Linear Algebra and Its Applications*, 430(1):360–369, 2009.

K.Q. Weinberger and L.K. Saul. Unsupervised Learning of Image Manifolds by Semidefinite Programming. *International Journal of Computer Vision*, 70(1):77–90, 2006.