Optimisation Combinatoire et Convexe.

Semidefinite programming

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
  - Eigenvalue problems
  - Combinatorial relaxations
  - Ellipsoidal approximations
  - Distortion, embedding
  - Mixing rates for Markov chains & maximum variance unfolding
  - Moment problems & positive polynomials

Semidefinite Programming

Semidefinite programming: conic duality

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

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

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**

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

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

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

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**
  - Eigenvalue problems, 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
  - Collaborative prediction

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. . .

Eigenvalue problems

Eigenvalue problems

Start from a semidefinite program **with constant trace**

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

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

$$\max_{\substack{\mathbf{Tr}(X)=1,\\ X \succeq 0}} \mathbf{Tr}(CX) = \lambda_{\max}(C),$$

the dual semidefinite program is written

$$\min_y \ \lambda_{\max}\left(C - \sum_{i=1}^m y_i\right) + b^T y$$

in the variable $y \in \mathbb{R}^m$.

Maximum eigenvalue minimization problems are usually easier to solve using first-order methods.

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

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

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, $\|\cdot\|_{1\to 2}$) [Frieze and Kannan, 1999, Alon and Naor, 2004, Nemirovski, 2005]

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,

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

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**:

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

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

Ellipsoidal approximations

Maximum volume ellipsoid $\mathcal{E}$ inside a convex set $C \subseteq \mathbb{R}^n$

- parametrize $\mathcal{E}$ as $\mathcal{E} = \{Bu + d \mid \|u\|_2 \leq 1\}$ with $B \succ 0$.
- $\mathbf{vol}\,\mathcal{E}$ is proportional to $\det B$, we can compute $\mathcal{E}$ by solving

$$\begin{array}{ll} \text{maximize} & \log\det B \\ \text{subject to} & \sup_{\|u\|_2 \leq 1} I_C(Bu + d) \leq 0 \end{array}$$

(where $I_C(x) = 0$ for $x \in C$ and $I_C(x) = \infty$ for $x \notin C$) again, this is a **convex problem**, but evaluating the constraint can be hard (for general C)

Polyhedron given by its **facets** $\{x \mid a_i^T x \leq b_i,\ i = 1, \ldots, m\}$:

$$\begin{array}{ll} \text{maximize} & \log\det B \\ \text{subject to} & \|Ba_i\|_2 + a_i^T d \leq b_i, \quad i = 1, \ldots, m \end{array}$$

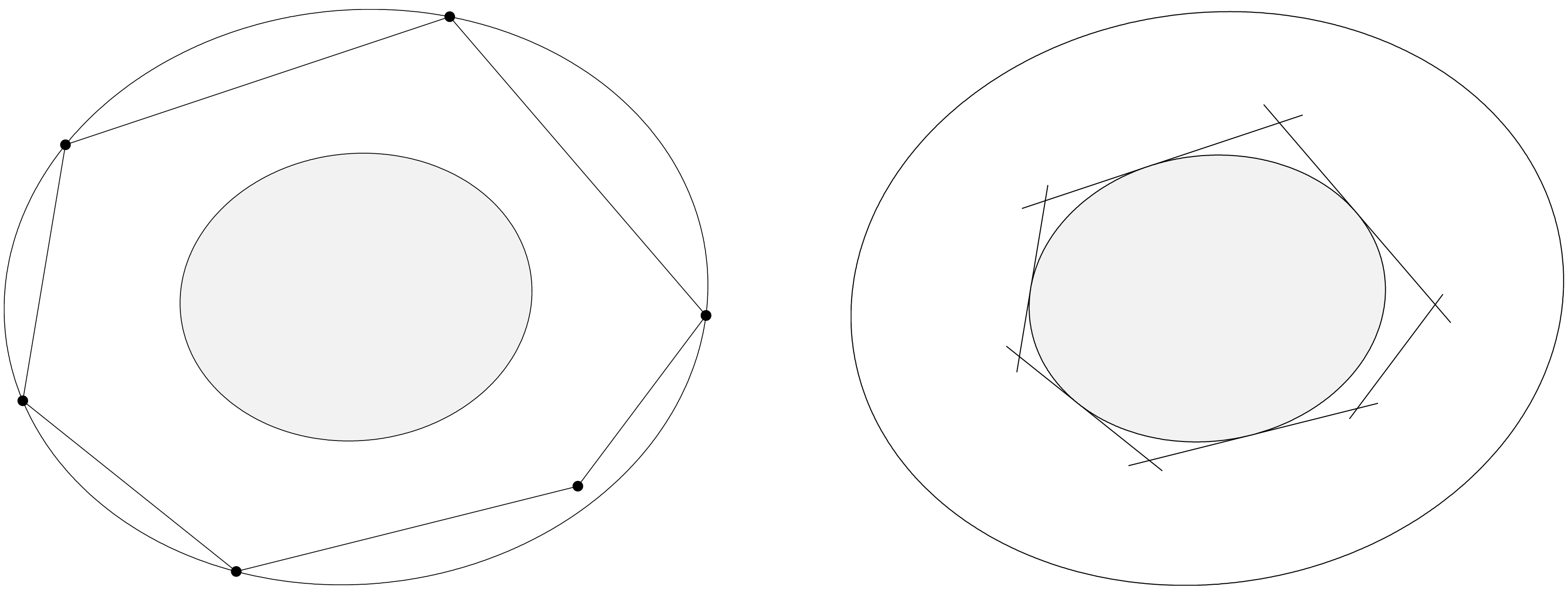
(constraint follows from $\sup_{\|u\|_2 \leq 1} a_i^T(Bu + d) = \|Ba_i\|_2 + a_i^T d$)

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.

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.

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.

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

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

  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

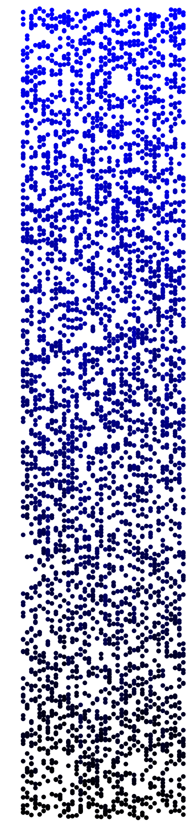
$$\mathbf{Tr}(X(\mathbf{I}-(1/n)\mathbf{1}\mathbf{1}^T)) = \textstyle\sum_i \|v_i\|_2^2 - \|\sum_i v_i\|_2^2$$

  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

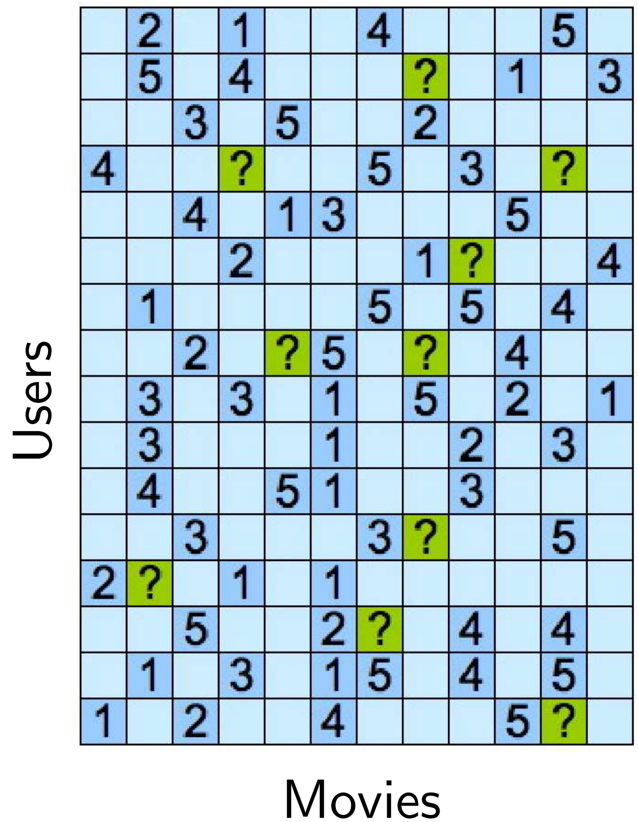
$$\mathbf{E}_\mu[x^i] = q_i,\ i = 0, \ldots, d \quad \Longleftrightarrow \quad \begin{pmatrix} q_0 & q_1 & \cdots & q_d \\ q_1 & q_2 & & q_{d+1} \\ \vdots & & \ddots & \vdots \\ q_d & q_{d+1} & \cdots & q_{2d} \end{pmatrix} \succeq 0$$

- [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.

Collaborative prediction

Collaborative prediction

- Users assign **ratings** to a certain number of movies:



- Objective: make recommendations for other movies. . .

Collaborative prediction

- Infer **user preferences** and **movie features** from user ratings.
- We use a linear prediction model:

$$rating_{ij} = u_i^T v_j$$

  where u_i represents user characteristics and v_j movie features.
- This makes collaborative prediction a **matrix factorization** problem
- Overcomplete representation. . .

Collaborative prediction

- **Inputs**: a matrix of ratings $M_{ij} = \{-1, +1\}$ for $(i, j) \in S$, where S is a subset of all possible user/movies combinations.
- We look for a linear model by factorizing $M \in \mathbb{R}^{n \times m}$ as:

$$M = U^T V$$

where $U \in \mathbb{R}^{n \times k}$ represents user characteristics and $V \in \mathbb{R}^{k \times m}$ movie features.

- **Parsimony**... We want k to be as small as possible.
- **Output**: a matrix $X \in \mathbb{R}^{n \times m}$ which is a low-rank approximation of the ratings matrix M.

Least-Squares

- Choose Means Squared Error as measure of discrepancy.
- Suppose S is the full set, our problem becomes:

$$\min_{\{X:\, \mathbf{Rank}(X)=k\}} \|X - M\|^2$$

- This is just a **singular value decomposition** (SVD)...

Problem: Not true when S is not the full set (partial observations). Also, MSE not a good measure of prediction performance...

Soft Margin

$$\text{minimize} \quad \mathbf{Rank}(X) + c \sum_{(i,j)\in S} \max(0, 1 - X_{ij}M_{ij})$$

non-convex and numerically hard. . .

- Relaxation result in Fazel et al. [2001]: replace $\mathbf{Rank}(X)$ by its convex envelope on the spectahedron to solve:

$$\text{minimize} \quad \|X\|_* + c \sum_{(i,j)\in S} \max(0, 1 - X_{ij}M_{ij})$$

  where $\|X\|_*$ is the **nuclear norm**, $i.e.$ sum of the singular values of X.

- Srebro [2004]: This relaxation also corresponds to multiple large margin SVM classifications.

Soft Margin

- The dual of this program:

$$\begin{array}{ll} \text{maximize} & \sum_{ij} Y_{ij} \\ \text{subject to} & \|Y \odot M\|_2 \leq 1 \\ & 0 \leq Y_{ij} \leq c \end{array}$$

  in the variable $Y \in \mathbb{R}^{n \times m}$, where $Y \odot M$ is the Schur (componentwise) product of Y and M and $\|Y\|_2$ the largest singular value of Y.

- This problem is **sparse**: $Y^*_{ij} = c$ for $(i, j) \in S^c$

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.

Stephen R Becker, Emmanuel J Candès, and Michael C Grant. Templates for convex cone problems with applications to sparse signal recovery. *Mathematical Programming Computation*, 3(3):165–218, 2011.

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.

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

O. Bunk, A. Diaz, F. Pfeiffer, C. David, B. Schmitt, D.K. Satapathy, and JF Veen. Diffractive imaging for periodic samples: retrieving one-dimensional concentration profiles across microfluidic channels. *Acta Crystallographica Section A: Foundations of Crystallography*, 63 (4):306–314, 2007.

E. J. Candes, T. Strohmer, and V. Voroninski. Phaselift : exact and stable signal recovery from magnitude measurements via convex programming. *To appear in Communications in Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

E.J. Candes and B. Recht. Exact matrix completion via convex optimization. *preprint*, 2008.

E.J. Candes and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *Information Theory, IEEE Transactions on*, 56(5): 2053–2080, 2010.

E.J. Candes, Y. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *Arxiv preprint arXiv:1109.0573*, 2011.

A. Chai, M. Moscoso, and G. Papanicolaou. Array imaging using intensity-only measurements. *Inverse Problems*, 27:015005, 2011.

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

L. Demanet and P. Hand. Stable optimizationless recovery from phaseless linear measurements. *Arxiv preprint arXiv:1208.1803*, 2012.

M. Fazel, H. Hindi, and S. Boyd. A rank minimization heuristic with application to minimum order system approximation. *Proceedings American Control Conference*, 6:4734–4739, 2001.

J.R. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15):2758–2769, 1982.

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.

R. Gerchberg and W. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35: 237–246, 1972.

M.X. Goemans and D. Williamson. Approximation algorithms for max-3-cut and other problems via complex semidefinite programming. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 443–452. ACM, 2001.

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.

D. Griffin and J. Lim. Signal estimation from modified short-time fourier transform. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 32(2):236–243, 1984.

R.W. Harrison. Phase problem in crystallography. *JOSA A*, 10(5):1046–1055, 1993.

C. Helmberg, F. Rendl, R. J. Vanderbei, and H. Wolkowicz. An interior–point method for semidefinite programming. *SIAM Journal on Optimization*, 6:342–361, 1996.

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.

M. Kisialiou and Z.Q. Luo. Probabilistic analysis of semidefinite relaxation for binary quadratic minimization. *SIAM Journal on Optimization*, 20:1906, 2010.

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

L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991.

Z.Q. Luo, X. Luo, and M. Kisialiou. An efficient quasi-maximum likelihood decoder for psk signals. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 6, pages VI–561. IEEE, 2003.

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

J. Miao, T. Ishikawa, Q. Shen, and T. Earnest. Extending x-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes. *Annu. Rev. Phys. Chem.*, 59:387–410, 2008.

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. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Mathematics Doklady*, 27(2): 372–376, 1983.

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.

B. Recht, M. Fazel, and P.A. Parrilo. Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *Arxiv preprint arXiv:0706.4138*, 2007.

H. Sahinoglou and S.D. Cabrera. On phase retrieval of finite-length sequences using the initial time sample. *Circuits and Systems, IEEE Transactions on*, 38(8):954–958, 1991.

N.Z. Shor. Quadratic optimization problems. *Soviet Journal of Computer and Systems Sciences*, 25:1–11, 1987.

A.M.C. So. Non–asymptotic performance analysis of the semidefinite relaxation detector in digital communications. *Working Paper*, 2010.

N. Srebro. *Learning with Matrix Factorization*. PhD thesis, Massachusetts Institute of Technology, 2004.

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.

Z. Wen, D. Goldfarb, S. Ma, and K. Scheinberg. Row by row methods for semidefinite programming. Technical report, Technical report, Department of IEOR, Columbia University, 2009.

S. Zhang and Y. Huang. Complex quadratic optimization and semidefinite programming. *SIAM Journal on Optimization*, 16(3):871–890, 2006.