Lecture 5. SDP Relaxations: Robust PCA, Sparse PCA and Graph Realization as MDS with Uncertainty

Yuan Yao

Hong Kong University of Science and Technology

Outline

PCA

- Let $X \in \mathbb{R}^{p \times n}$ be a data matrix. Classical PCA looks for a matrix decomposition

$$X = L + E$$

where

  - L is of low-rank (*e.g.*at most rank k),
  - error matrix E has a small Frobenius norm, which is usually the case for Gaussian noise

PCA

- Classical PCA solves

$$\begin{aligned} \min \quad & \|X - L\| \\ \text{subject to} \quad & \mathbf{rank}(L) \le k \end{aligned} \tag{1}$$

where the norm here is any unitary invariant matrix norms, *e.g.*

  – Schatten's p-norm $\|M\|_p = (\sum_i \sigma_i(M)^p)^{1/p}$ $(p \ge 1)$ when M admits the Singular Value Decomposition (SVD) $M = USV^T$ with $S = \mathbf{diag}(\sigma_1, \ldots, \sigma_k, \ldots)$ $(p = 2$ is the Frobenius norm, $p = 1$ is the nuclear norm, and $p = \infty$ gives the spectral norm).

  – SVD provides a solution with $L = \sum_{i \le k} \sigma_i u_i v_i^T$ where $X = \sum_i \sigma_i u_i v_i^T$ $(\sigma_1 \ge \sigma_2 \ge \ldots)$.

PCA is sensitive to outliers

- However, if some outliers exists, *i.e.* there are a small amount of sample points which are largely deviated from the main population of samples, the classical PCA is well-known very sensitive to such outliers.

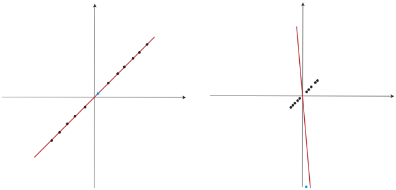


Figure: Classical PCA is sensitive to outliers

Outline

Robust PCA

- To address this issue, Robust PCA looks for the following decomposition instead

$$X = L + S$$

  where
  - L is a low rank matrix;
  - S is a sparse matrix.

Example: rank-1 spike model

Example (Spike model)

- In the spike signal model,

$$X = \alpha u + \sigma_\epsilon \epsilon, \qquad \alpha \sim \mathcal{N}(0, \sigma_u^2) \text{ and } \epsilon \sim \mathcal{N}(0, I_p).$$

  - X is thus subject to the following normal distribution $\mathcal{N}(0, \Sigma)$ where

$$\Sigma = \sigma_u^2 u u^T + \sigma_\epsilon^2 I.$$

  - So $\Sigma = L + S$ has such a rank-sparsity structure with

$$L = \sigma_u^2 u u^T, \qquad S = \sigma_\epsilon^2 I.$$

Example: Surveillance video

Example (Surveillance Video Decomposition)

Figure: Surveillance video as a rank-sparse model: Left = low-rank (middle) + sparse (right)

Example: Gaussian Graphical Model

Example (Gaussian Graphical Model)

Let $X = [X_1, \ldots, X_p]^T \sim \mathcal{N}(0, \Sigma)$ be multivariate Gaussian random variables.

- The following characterization holds

$$X_i \text{ and } X_j \text{ are conditionally independent given other variables}$$

$$\Leftrightarrow (\Sigma^{-1})_{ij} = 0$$

  We denote it by $X_i \perp X_j | X_{-i,-j}$.

- Let $G = (V, E)$ be a undirected graph where V represent p random variables and

$$(i, j) \in E \Leftrightarrow x_i \perp x_j | x_k (k \notin \{i, j\}).$$

  G is called a (Gaussian) graphical model of X.

Example: Gaussian Graphical Model (continued)

- Divide the random variables into observed and hidden (a few) variables $X = (X_o, X_h)^T$ (in semi-supervised learning, labeled vs. unlabeled, respectively) and

$$\Sigma = \begin{bmatrix} \Sigma_{oo} & \Sigma_{oh} \\ \Sigma_{ho} & \Sigma_{hh} \end{bmatrix} \quad \text{and} \quad Q = \Sigma^{-1} = \begin{bmatrix} Q_{oo} & Q_{oh} \\ Q_{ho} & Q_{hh} \end{bmatrix}$$

- The following Schur Complement equation holds for covariance matrix of observed variables

$$\Sigma_{oo}^{-1} = Q_{oo} + Q_{oh} Q_{hh}^{-1} Q_{ho}.$$

  Note that
  - Observable variables are often conditional independent given hidden variables, so Q_{oo} is expected to be *sparse*;
  - Hidden variables are of small number, so $Q_{oh} Q_{hh}^{-1} Q_{ho}$ is of *low-rank*.

Example: Gaussian Graphical Model (continued)

- In semi-supervised learning, X_o is labeled data and X_h is unlabeled. The labeled points are of small number, and the unlabeled points should be as much conditionally independent as possible to each other given labeled points. This implies that the labels should be placed on those most "influential" points.

Robust PCA

- In Robust PCA, the purpose is to solve

$$\begin{aligned} \min \quad & \|X - L\|_0 \\ s.t. \quad & \mathbf{rank}(L) \le k \end{aligned} \tag{2}$$

  where $\|A\|_0 = \#\{A_{ij} \neq 0\}$.

- However both the objective function and the constraint are non-convex, whence it is NP-hard to solve in general.

- In practice, one often uses alternative optimization.

- Here we introduce convex relaxation.

Convex Relaxation

- The simplest convex relaxations:

$$\|S\|_0 := \#\{S_{ij} \neq 0\} \Rightarrow \|S\|_1 \tag{3}$$

$$\mathbf{rank}(L) := \#\{\sigma_i(L) \neq 0\} \Rightarrow \|L\|_* = \sum_i \sigma_i(L), \tag{4}$$

where $\|L\|_*$ is called the *nuclear norm* of L, which has a semi-definite representation

$$\begin{aligned} \|L\|_* = \quad & \min \quad \frac{1}{2}(\mathrm{tr}(W_1) + \mathrm{tr}(W_2)) \\ & s.t. \quad \begin{bmatrix} W_1 & L \\ L^T & W_2 \end{bmatrix} \succeq 0. \end{aligned}$$

Robust PCA via SDP

- The relaxed Robust PCA problem can be solved by the following Semi-Definite Programming (SDP).

$$
\begin{aligned}
\min \quad & \frac{1}{2}(\mathrm{tr}(W_1) + \mathrm{tr}(W_2)) + \lambda \|S\|_1 \\
s.t. \quad & L_{ij} + S_{ij} = X_{ij}, \quad (i,j) \in E \\
& \begin{bmatrix} W_1 & L \\ L^T & W_2 \end{bmatrix} \succeq 0
\end{aligned} \tag{5}
$$

Matlab codes

- The Matlab codes (`testRPCA.m`) realized the SDP algorithm above by CVX (`http://cvxr.com/cvx`).
- Typically CVX only solves SDP problem of small sizes (say matrices of size less than 100). Specific matlab tools have been developed to solve large scale RPCA, which can be found at `http://perception.csl.uiuc.edu/matrix-rank/home.html`.
- Stephen Boyd's website contains ADMM algorithm compared with CVX: `http://web.stanford.edu/~boyd/papers/prox_algs/matrix_decomp.html`

ADMM

- For SDP problem

$$\begin{aligned} \min \quad & \|E\|_F^2 + \gamma_2 \|S\|_1 + \gamma_3 \|L\|_* \\ s.t. \quad & L + S + E = A, \end{aligned} \tag{6}$$

- Augmented Lagrangian:

$$\begin{aligned} & \mathcal{L}(E, L, S; B) \\ = \quad & \|E\|_F^2 + \gamma_2 \|S\|_1 + \gamma_3 \|L\|_* + \ldots \\ & -\langle B, A - L - S - E \rangle + \frac{\rho}{2} \|A - L - S - E\|_F^2 \end{aligned} \tag{7}$$

- ADMM in Stephen Boyd's version: `http://web.stanford.edu/~boyd/papers/prox_algs/matrix_decomp.html`

ADMM

- initialization: $\lambda = 1, \lambda = 1/\rho$, $X_1^0 = X_2^0 = X_3^0 = B^0 = 0^{m\times n}$

- for $k = 0, 1, 2, \ldots$

$$B^{k+1} = B^k + \frac{1}{3}(X_1^k + X_2^k + X_3^k - A), \tag{8a}$$

$$X_1^{k+1} = \frac{1}{1+\lambda}(X_1^k - B^k), \tag{8b}$$

$$X_2^{k+1} = \mathbf{prox}_{\|x\|_1}(X_2^k - B^k, \lambda\gamma_2), \tag{8c}$$

$$X_3^{k+1} = \mathbf{prox}_{\|M\|_*}(X_3^k - B^k, \lambda\gamma_3), \tag{8d}$$

  where $\mathbf{prox}_h(z, c) = \min_x \frac{1}{2}\|x - z\|_F^2 + ch(x)$.

- return $E = X_1^k$, $S = X_2^k$, $L = X_3^k$.

Question

How does SDP work?

Outline

Exact Recovery Theory

- A fundamental question about Robust PCA is: given $X = L_0 + S_0$ with low-rank L and sparse S, under what conditions that one can recover X by solving SDP in (6)?

Exact Recovery Theory

- It is necessary to assume that
  - the low-rank matrix L_0 can not be sparse;
  - the sparse matrix S_0 can not be of low-rank.

Exact Recovery Theory

- Such an assumption can be characterized using the following algebraic language. Define

$$T(L_0) = \{UA^T + BV^T : \forall A, B \in \mathbb{R}^{n \times p}, L_0 = USV^T\}$$

which is the tangent space at L_0 varying in the same column and row spaces of L_0, and

$$\Omega(S_0) = \{S : \mathbf{supp}(S) \subseteq \mathbf{supp}(S_0)\},$$

which is the tangent space at S_0 varying within the same support of S_0. The assumptions above are equivalent to say that tangent spaces $T(L_0)$ and $\Omega(S_0)$ are transversal with only intersection at 0,

$$\text{Transversality:} \quad T(L_0) \bigcap \Omega(S_0) = \{0\}.$$

Exact Recovery Theory

- The following two incoherence constants measure the "diffusive behaviours" of sparse (low-rank) matrices onto low-rank (sparse) opponents.

$$\mu(S_0) = \max_{S \in \Omega(S_0), \|S\|_\infty \leq 1} \|S\|_2$$

$$\xi(L_0) = \max_{L \in T(L_0), \|L\|_2 \leq 1} \|L\|_\infty$$

Exact Recovery Theory

- V. Chandrasekaran, S. Sanghavi, P.A. Parrilo, and A. Willsky (2011) showed the following uncertainty principle, for any matrix M, $\mu(M) \cdot \xi(M) \geq 1$. Therefore a sufficient condition holds,

  $$\mu(S_0) \cdot \xi(L_0) < 1 \Rightarrow T(L_0) \bigcap \Omega(S_0) = \{0\}.$$

  Moreover, the following deterministic recovery conditions is shown for SDP

  $$\mu(S_0) \cdot \xi(L_0) < 1/6 \Rightarrow \text{SDP recovers } L_0 \text{ and } S_0.$$

Incoherence Condition

- Probabilistic recovery conditions are given by Candes and Recht (2009). Assume that $L_0 \in \mathbb{R}^{n\times n} = U\Sigma V^T$ and $r = \mathbf{rank}(L_0)$. **Incoherence condition** (Candes-Recht (2009)): there exists a $\mu \geq 1$ such that for all $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$,

$$\|U^T e_i\|^2 \leq \frac{\mu r}{n}, \quad \|V^T e_i\|^2 \leq \frac{\mu r}{n},$$

  and

$$|UV^T|_{ij}^2 \leq \frac{\mu r}{n^2}.$$

- These conditions, roughly speaking, ensure that the singular vectors are not sparse, i.e. well-spread over all coordinates and won't concentrate on some coordinates.
  - The incoherence condition holds if $|U_{ij}|^2 \vee |V_{ij}|^2 \leq \mu/n$. In fact, if U represent random projections to r-dimensional subspaces with $r \geq \log n$, we have $\max_i \|U^T e_i\|^2 \asymp r/n$.
  - To meet the second condition, we simply assume that the sparsity pattern of S_0 is uniformly random.

Probabilistic Recovery Theorem

Theorem (Candes-Recht (2009))

Assume the following holds,

1. L_0 is n-by-n with $\mathbf{rank}(L_0) \leq \rho_r n \mu^{-1} (\log n)^{-2}$,
2. S_0 is uniformly sparse of cardinality $m \leq \rho_s n^2$.

Then with probability $1 - O(n^{-10})$, (6) with $\lambda = 1/\sqrt{n}$ is exact, *i.e.* its solution $\hat{L} = L_0$ and $\hat{S} = S_0$.

Remark

- Note that if L_0 is a rectangular matrix of $n_1 \times n_2$, the same holds with $\lambda = 1/\sqrt{(\max n_1, n_2)}$.
- The result can be generalized to $1 - O(n^{-\beta})$ for $\beta > 0$.
- Extensions and improvements of these results to incomplete measurements can be found in (Candes-Tao (2010); Gross (2011)), which solves the following SDP problem.

$$\begin{aligned} \min \quad & \|L\|_* + \lambda \|S\|_1 \\ s.t. \quad & L_{ij} + S_{ij} = X_{ij}, \quad (i,j) \in \Omega_{obs}. \end{aligned} \tag{9}$$

Probabilistic Recovery with Missing Values

Theorem

Assume the following holds,

1. L_0 is n-by-n with $\mathbf{rank}(L_0) \leq \rho_r n \mu^{-1} (\log n)^{-2}$,
2. Ω_{obs} is a uniform random set of size $m = 0.1n^2$,
3. each observed entry is corrupted with probability $\tau \leq \tau_s$.

Then with probability $1 - O(n^{-10})$, (6) with $\lambda = 1/\sqrt{0.1n}$ is exact, *i.e.* its solution $\hat{L} = L_0$. The same conclusion holds for rectangular matrices with $\lambda = 1/\sqrt{\max dim}$. All these results hold irrespective to the magnitudes of L_0 and S_0.

Matrix Completion

- When there are no sparse perturbation in optimization problem (9), the problem becomes the classical Matrix Completion problem with uniformly random sampling:

$$\begin{aligned} \min \quad & \|L\|_* \\ s.t. \quad & L_{ij} = L^0_{ij}, \quad (i,j) \in \Omega_{obs}. \end{aligned} \tag{10}$$

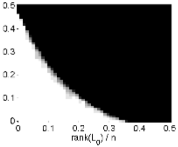
- Assumed the same condition as before, Candes and Tao (2010) gives the following result: solution to SDP (10) is exact with probability at least $1 - n^{-10}$ if $m \geq \mu n r \log^a n$ where $a \leq 6$, which can be improved by Gross (2011) to be near-optimal

$$m \geq \mu n r \log^2 n.$$

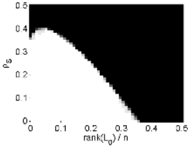
Phase Transitions or Random Matrix Completion

- Take $L_0 = UV^T$ as a product of $n \times r$ i.i.d. $\mathcal{N}(0, 1)$ random matrices. There is a phase transition of successful recovery probability over sparsity ratio $\rho_s = m/n^2$ and low rank ratio r/n.

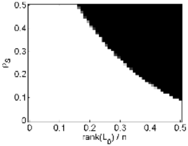
Phase Transitions of Random Matrix Completion



(a) PCP, Random Signs



(b) PCP, Coherent Signs



(c) Matrix Completion

Figure: Phase Transitions in Probability of Successful Recovery

Phase Transitions

- White color indicates the probability equals to 1 and black color corresponds to the probability being 0. A sharp phase transition curve can be seen in the pictures. (a) and (b) respectively use random signs and coherent signs in sparse perturbation, where (c) is purely matrix completion with no perturbation. Increasing successful recovery can be seen from (a) to (c).

Outline

Sparse PCA

- Recall that classical PCA is to solve

$$\begin{aligned} \max \quad & x^T \Sigma x \\ s.t. \quad & \|x\|_2 = 1 \end{aligned}$$

which gives the maximal variation direction of covariance matrix Σ.

- What if only a few coordinates in x are nonzeros in PCA? For example, in human genomics, only a few genes influence a certain disease.

A Convex Relaxation of PCA by SDP

- Note that $x^T \Sigma x = \mathrm{tr}(\Sigma(xx^T))$. Classical PCA can thus be relaxed as follows after dropping the rank-1 constraint,

$$\begin{aligned} \max \quad & \mathrm{tr}(\Sigma X) \\ s.t. \quad & \mathrm{tr}(X) = 1 \\ & X \succeq 0 \end{aligned}$$

  The optimal solution gives a rank-1 X along the first principal component.

- A recursive application of the algorithm may lead to top k principal components. That is, one first to find a rank-1 approximation of Σ and extract it from $\Sigma_0 = \Sigma$ to get $\Sigma_1 = \Sigma - X$, then pursue the rank-1 approximation of Σ_1, and so on.

Sparse PCA

- Now we are looking for sparse principal components, i.e. $\#\{X_{ij} \neq 0\}$ are small. Using 1-norm convex relaxation, we have the following SDP formulation by dAspremont, El Ghaoui, Jordan, Lanckriet (2007) for Sparse PCA

$$\begin{aligned} \max \quad & \mathrm{tr}(\Sigma X) - \lambda \|X\|_1 \\ s.t. \quad & \mathrm{tr}(X) = 1 \\ & X \succeq 0 \end{aligned}$$

Matlab Codes for Sparse PCA

- The Matlab codes (`testSPCA.tex`) realized the SDP algorithm above by CVX (`http://cvxr.com/cvx`).
- Python package `scikit-learn` includes: `http://scikit-learn.org/stable/modules/generated/sklearn.decomposition.SparsePCA.html`

Other Approaches to Sparse PCA

- There are many other algorithms for sparse PCA, e.g. regression with LASSO (Hui Zou; Trevor Hastie; Robert Tibshirani (2006)), alternative nonconvex optimization etc.

- A recent survey: Hui Zou; Lingzhou Xue (2018). "A Selective Overview of Sparse Principal Component Analysis". Proceedings of the IEEE. 106 (8): 13111320.

Outline

Linear Programming: Primal Problem

- LP (Linear Programming): for $x \in \mathbb{R}^n$ and $c \in \mathbb{R}^n$,

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{11}$$

This is the primal linear programming problem.

Linear Programming: Primal Problem

- SDP (Semi-definite Programming): for $X, C \in \mathbb{R}^{n\times n}$

$$\begin{aligned} \min \quad & C \bullet X = \sum_{i,j} c_{ij} X_{ij} \\ s.t. \quad & A_i \bullet X = b_i, \quad \text{for } i = 1, \cdots, m \\ & X \succeq 0 \end{aligned} \tag{12}$$

- In SDP, nonnegative variables x is replaced by positive semi-definite matrices X.
- In SDP, the inner product between vectors $c^T x$ in LP will change to Hadamard inner product (denoted by $\bullet$) between matrices.

From Primal to Dual

- Linear programming has a dual problem via the Lagrangian.

- The Lagrangian of the primal problem is

$$\max_{\mu \geq 0, y} \min_{x} L_{x;y,\mu} = c^T x + y^T (b - Ax) - \mu^T x$$

which implies that

$$\frac{\partial L}{\partial x} = c - A^T y - \mu = 0$$

$$\Longleftrightarrow c - A^T y = \mu \geq 0$$

$$\Longrightarrow \max_{\mu \geq 0, y} L = \max_{\mu \geq 0, y} y^T b$$

which leads to the following dual problem.

Linear Programming: Dual Problem

- LD (Dual Linear Programming):

$$\begin{aligned} \max \quad & b^T y \\ s.t. \quad & \mu = c - A^T y \geq 0 \end{aligned} \tag{13}$$

Semi-Definite Programming: Dual Problem

- SDD (Dual Semi-definite Programming):

$$\begin{aligned} \max \quad & b^T y \\ s.t. \quad & S = C - \sum_{i=1}^{m} A_i y_i \succeq 0 =: C - \langle A, y \rangle \end{aligned} \tag{14}$$

  where

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

Weak Duality

- Define the feasible set of primal and dual problems are $\mathbb{F}_p = \{X \succeq 0; A_i \bullet X = b_i\}$ and $\mathbb{F}_d = \{(y, S) : S = C - \sum_i y_i A_i \succeq 0\}$, respectively.

Theorem (Weak Duality of SDP)

If $\mathbb{F}_p \neq \emptyset, \mathbb{F}_d \neq \emptyset$, then

$$C \bullet X \geq b^T y,$$

for $\forall X \in \mathbb{F}_p$ and $\forall (y, S) \in \mathbb{F}_d$.

- The week duality says that the primal value is always an upper bound of dual value. The gap, $\gamma = C \bullet X - b^T y > 0$, is called the duality gap.

Strong Duality

Theorem (Strong Duality SDP)

Assume the following hold,

1. $\mathbb{F}_p \neq \emptyset, \mathbb{F}_d \neq \emptyset$;
2. At least one feasible set has an interior.

Then X^* is optimal iff

1. $X^* \in \mathbb{F}_p$
2. $\exists (y^*, S^*) \in \mathbb{F}_d$

s.t. $C \bullet X^* = b^T y^*$ or $X^* S^* = 0$ (note: in matrix product)

Remark

- The strong duality says that the existence of an interior point ensures the vanishing duality gap between primal value and dual value, as well as the complementary conditions hold. In this case, to check the optimality of a primal variable, it suffices to find a dual variable which meets the complementary condition with the primal. This is often called the *witness* method.

- The existence of an interior solution implies the complementary condition of optimal solutions. Under the complementary condition, we have

$$\mathbf{rank}(X^*) + \mathbf{rank}(S^*) \leq n$$

  for every optimal primal X^* and dual S^*.

Outline

Recall: MDS

- Recall that in classical MDS, given all pairwise distances $d_{ij} = \|x_i - x_j\|^2$ among a set of points $x_i \in \mathbb{R}^p$ ($i = 1, 2, \cdots, n$) whose coordinates are unknown, our purpose is to find $y_i \in \mathbb{R}^k (k \leq p)$ such that

$$\min \sum_{i,j=1}^{n} \left(\|y_i - y_j\|^2 - d_{ij} \right)^2 . \tag{15}$$

MDS with Incomplete and Uncertain Information

- What about the following scenarios?
  - Noisy perturbations: $d_{ij} \to \widetilde{d_{ij}} = d_{ij} + \epsilon_{ij}$
  - Incomplete measurments: only partial pairwise distance measurements are available on an edge set of graph, *i.e.* $G = (V, E)$ and d_{ij} is given when $(i, j) \in E$ (*e.g.* x_i and x_j in a neighborhood).
  - Anchors: sometimes we may fixed the locations of some points called *anchors*, *e.g.* in sensor network localization (SNL) problem.

- In other words, we are looking for MDS on graphs with partial and noisy information, often called **Graph Realization**.

Semi-Definite Relaxation of MDS

Lemma

The quadratic constraint

$$\|y_i - y_j\|^2 = d_{ij}^2, \quad (i,j) \in E$$

has a semi-definite relaxation:

$$\begin{cases} Z_{1:k,1:k} = I \\ (0; e_i - e_j)(0; e_i - e_j)^T \bullet Z = d_{ij}^2, \quad (i,j) \in E \\ Z = \begin{bmatrix} I_k & Y \\ Y^T & X \end{bmatrix} \succeq 0. \end{cases}$$

where $\bullet$ denotes the Hadamard inner product, *i.e.*
$A \bullet B := \sum_{i,j=1}^n A_{ij} B_{ij}$.

Proof of Lemma

- Denote $Y = [y_1, \cdots, y_n]^{k \times n}$ where $y_i \in \mathbb{R}^k$, and

$$e_i = (0, 0, \cdots, 1, 0, \cdots, 0) \in \mathbb{R}^n.$$

- Then we have

$$\|y_i - y_j\|^2 = (y_i - y_j)^T (y_i - y_j) = (e_i - e_j)^T Y^T Y (e_i - e_j)$$

Set $X = Y^T Y$, which is symmetric and positive semi-definite. Then

$$\|Y_i - Y_j\|^2 = (e_i - e_j)(e_i - e_j)^T \bullet X.$$

So

$$\|Y_i - Y_j\|^2 = d_{ij}^2 \Leftrightarrow (e_i - e_j)(e_i - e_j)^T \bullet X = d_{ij}^2$$

which is linear with respect to X.

Proof of Lemma (continued)

- Now we relax the constrain $X = Y^T Y$ to

$$X \succeq Y^T Y \Longleftrightarrow X - Y^T Y \succeq 0.$$

Through Schur Complement Lemma we know

$$X - Y^T Y \succeq 0 \Longleftrightarrow \begin{bmatrix} I & Y \\ Y^T & X \end{bmatrix} \succeq 0$$

- We may define a new variable

$$Z \in S^{k+n}, Z = \begin{bmatrix} I_k & Y \\ Y^T & X \end{bmatrix}$$

which gives the result. □

SD Relaxations of MDS

- Given anchors a_k $(k = 1, \ldots, s)$ with known coordinates, find x_i such that
  - $\|x_i - x_j\|^2 = d_{ij}^2$ where $(i, j) \in E_x$ and x_i are unknown locations
  - $\|a_k - x_j\|^2 = \widehat{d_{kj}}^2$ where $(k, j) \in E_a$ and a_k are known locations

- We can exploit the following SD relaxation:
  - $(0; e_i - e_j)(0; e_i - e_j)^T \bullet Z = d_{ij}$ for $(i, j) \in E_x$,
  - $(a_i; e_j)(a_i; e_j)^T \bullet Z = \widehat{d_{ij}}$ for $(i, j) \in E_a$,

  both of which are linear with respect to Z.

- The constraints with equalities of d_{ij}^2 can be replaced by inequalities such as $\leq d_{ij}^2(1 + \epsilon)$ (or $\geq d_{ij}^2(1 - \epsilon)$). This is a system of linear matrix inequalities with positive semidefinite variable Z.

Dual Problem

- The SDD associated with the primal problem above is

$$\min \quad I \bullet V + \sum_{i,j \in E_x} w_{ij} d_{ij} + \sum_{i,j \in E_a} \widehat{w}_{ij} \widehat{d_{ij}} \tag{16}$$

s.t.

$$S = \begin{pmatrix} V & 0 \\ 0 & 0 \end{pmatrix} + \sum_{i,j \in E_x} w_{ij} A_{ij} + \sum_{i,j \in E_a} \widehat{w}_{ij} \widehat{A_{ij}} \succeq 0$$

where

$$A_{ij} = (0; e_i - e_j)(0; e_i - e_j)^T$$

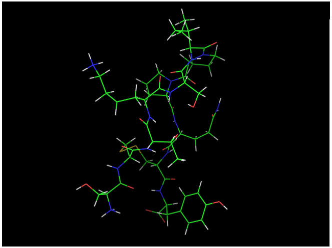
$$\widehat{A_{ij}} = (a_i; e_j)(a_i; e_j)^T.$$

Remark on Dual Problem

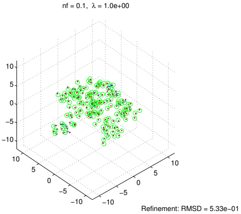
- The variables w_{ij} is the stress matrix on edge between unknown points i and j and $\widehat{w}_{ij}$ is the stress matrix on edge between anchor i and unknown point j.
- The dual is always feasible, as $V = 0$, $y_{ij} = 0$ for all $(i, j) \in E_x$ and $w_{ij} = 0$ for all $(i, j) \in E_a$ is a feasible solution.

Example: Protein 3D Structure Reconstruction

- Here we show an example of using SDP to find 3-D coordinates of a protein molecule based on noisy pairwise distances for atoms in ϵ-neighbors. We use matlab package SNLSDP by Kim-Chuan Toh, Pratik Biswas, and Yinyu Ye, downladable at `http://www.math.nus.edu.sg/~mattohkc/SNLSDP.html`.



(a)



(b)

- Matlab: `testSNL.m`

Question

- A crucial theoretical question is to ask, when $X = Y^T Y$ holds such that SDP embedding Y gives the same answer as the classical MDS?

Question

- Such SDP has the following rank properties:
  A. maximal rank solutions X^* or S^* exist;
  B. minimal rank solutions X^* or S^* exist;
  C. if complementary condition $X^*S^* = 0$ holds, then $\mathbf{rank}(X^*) + \mathbf{rank}(S^*) \leq n$ with equality holds iff strictly complementary condition holds, whence $\mathbf{rank}(S^*) \geq n - k \Rightarrow \mathbf{rank}(X^*) \leq k$.

Question

- Strong duality of SDP tells us that an interior point feasible solution in primal or dual problem will ensure the complementary condition and the zero duality gap. Now we assume that $d_{ij} = \|x_i - x_j\|$ precisely for some unknown $x_i \in \mathbb{R}^k$. Then the primal problem is feasible with $Z = (I_k; Y)^T (I_k; Y)$. Therefore the complementary condition holds and the duality gap is zero. In this case, assume that Z^* is a primal feasible solution of SDP embedding and S^* is an optimal dual solution, then
    1. $\mathbf{rank}(Z^*) + \mathbf{rank}(S^*) \leq k + n$ and $\mathbf{rank}(Z^*) \geq k$ $\Rightarrow \mathbf{rank}(S^*) \leq n$;
    2. $\mathbf{rank}(Z^*) = k \Longleftrightarrow X = Y^T Y$.
- It follows that if an optimal dual S^* has rank n, then every primal solution Z^* has rank k, which ensures $X = Y^T Y$. Therefore it suffices to find a maximal rank dual solution S^* whose rank is n.

Universal Rigidity

Definition (Universal Rigidity (UR) or Unique Localization (UL))

$\exists ! y_i \in \mathbb{R}^k \hookrightarrow \mathbb{R}^l$ where $l \geq k$ s.t. $d_{ij}^2 = \|y_i - y_j\|^2, \widehat{d_{ij}}^2 = \|a_k - y_j\|^2$.

- It simply says that there is no nontrivial extension of $y_i \in \mathbb{R}^k$ in $\mathbb{R}^l$ satisfying $d_{ij}^2 = \|y_i - y_j\|^2$ and $\widehat{d_{ij}}^2 = \|(a_k; 0) - y_j\|^2$.

Graph Realization with Universal Rigidity

(A) (Schoenberg 1938) G is complete $\Longrightarrow$ UR

(B) (So-Ye 2007) G is incomplete: UR $\Longleftrightarrow$ SDP has a maximal rank solution $\mathbf{rank}(Z^*) = k$.

Graph Realization Theorem

Theorem (So-Ye (2007))

The following statements are equivalent.

1. The graph is universally rigid or has a unique localization in $\mathbb{R}^k$.
2. The max-rank feasible solution of the SDP relaxation has rank k;
3. The solution matrix has $X = Y^T Y$ or $\mathrm{tr}(X - Y^T Y) = 0$.

Moreover, the localization of a UR instance can be computed approximately in a time polynomial in n, k, and the accuracy $\log(1/\epsilon)$.

Noisy Graph Realization

- In practice, we often meet problems with noisy measurements $\alpha d_{ij}^2 \geq \tilde{d}_{ij}^2 \leq \beta d_{ij}^2$.

- If we relax the constraint $\|y_i - y_j\|^2 = d_{ij}^2$ or equivalently $A_i \bullet X = b_i$ to inequalities, we can achieve **arbitrarily small rank** solution.

- To see this, assume that for $i = 1, \ldots, m$, we replace

$$A_i X = b_i \quad \mapsto \quad \alpha b_i \leq A_i X \leq \beta b_i,$$

where $\beta \geq 1 > \alpha > 0$.

Noisy Graph Realization Theorem

Theorem (So, Ye, and Zhang (2008))

For every $d \geq 1$, there is a SDP solution $\widehat{X} \succeq 0$ with rank $\mathbf{rank}(\widehat{X}) \leq d$, if the following holds,

$$\beta = \begin{cases} 1 + \dfrac{18 \ln 2m}{d} & 1 \leq d \leq 18 \ln 2m \\ 1 + \dfrac{\sqrt{18 \ln 2m}}{d} & d \geq 18 \ln 2m \end{cases}$$

$$\alpha = \begin{cases} \dfrac{1}{e(2m)^{2/d}} & 1 \leq d \leq 4 \ln 2m \\ \max \left\{ \dfrac{1}{e(2m)^{2/d}}, 1 - \sqrt{\dfrac{4 \ln 2m}{d}} \right\} & d \geq 4 \ln 2m \end{cases}$$

Remark

- Note that α, β are independent to n.

- Arbitrary dimension $d \geq 1$ embedding is achievable as long as distortion levels β and α are properly chosen.

Summary

- We have introduced semi-definite programming (relaxations) to the following problems
  - Robust PCA
  - Sparse PCA
  - Graph Realization as MDS with Uncertainy
- Many spectral methods allow SDP relaxations with powerful theoretical guarantees.