Less is More:
Sparsity in Principal Component Analysis and in Linear Systems

Laurent El Ghaoui

EECS dept., UC Berkeley

Goals

- Examine *linear algebra problems with cardinality constraints*
- Develop new formulations, and corresponding convex relaxations
- New formulations may offer insights into problem
- Ultimate objective is to derive estimates of the quality of the convex relaxations

Outline

- *Principal component analysis*
- The sparse PCA problem
- New formulation and SDP relaxation
- Quality estimate
- Sparsity in linear systems

Principal component analysis

PCA is a classic tool in multivariate data analysis

- Input: a $n \times n$ covariance matrix Σ
- Output: a sequence of *factors* ranked by *variance*
- Each factor is a *linear* combination of the problem variables

Typical use: *reduce the number of dimensions* of a model while *maximizing the information* (variance) contained in the simplified model

Solving the PCA problem

- The PCA problem can be solved via the *eigenvalue decomposition* of the covariance matrix:

$$\Sigma = \sum_{i=1}^{n} \lambda_i x_i x_i^T$$

- $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$ are the eigenvalues of Σ

- The corresponding eigenvectors x_i are called the *principal components*, or factors.

PCA and rank-one approximation

- The first principal component, x_1, can be obtained via the solution to the *rank-one approximation problem:*

$$\min_z \ \|\Sigma - zz^T\|_F,$$

the solution of which is $z = \lambda_1 x_1 x_1^T$.

(Here, $\|A\|_F^2 = \mathbf{Tr}\, A^T A$ denotes the Frobenius norm of a matrix A.)

- Above problem can be reduced to the *variational problem:*

$$\max_x \ x^T \Sigma x \ : \ \|x\|_2 = 1,$$

the solution of which is $x = x_1$.

Outline

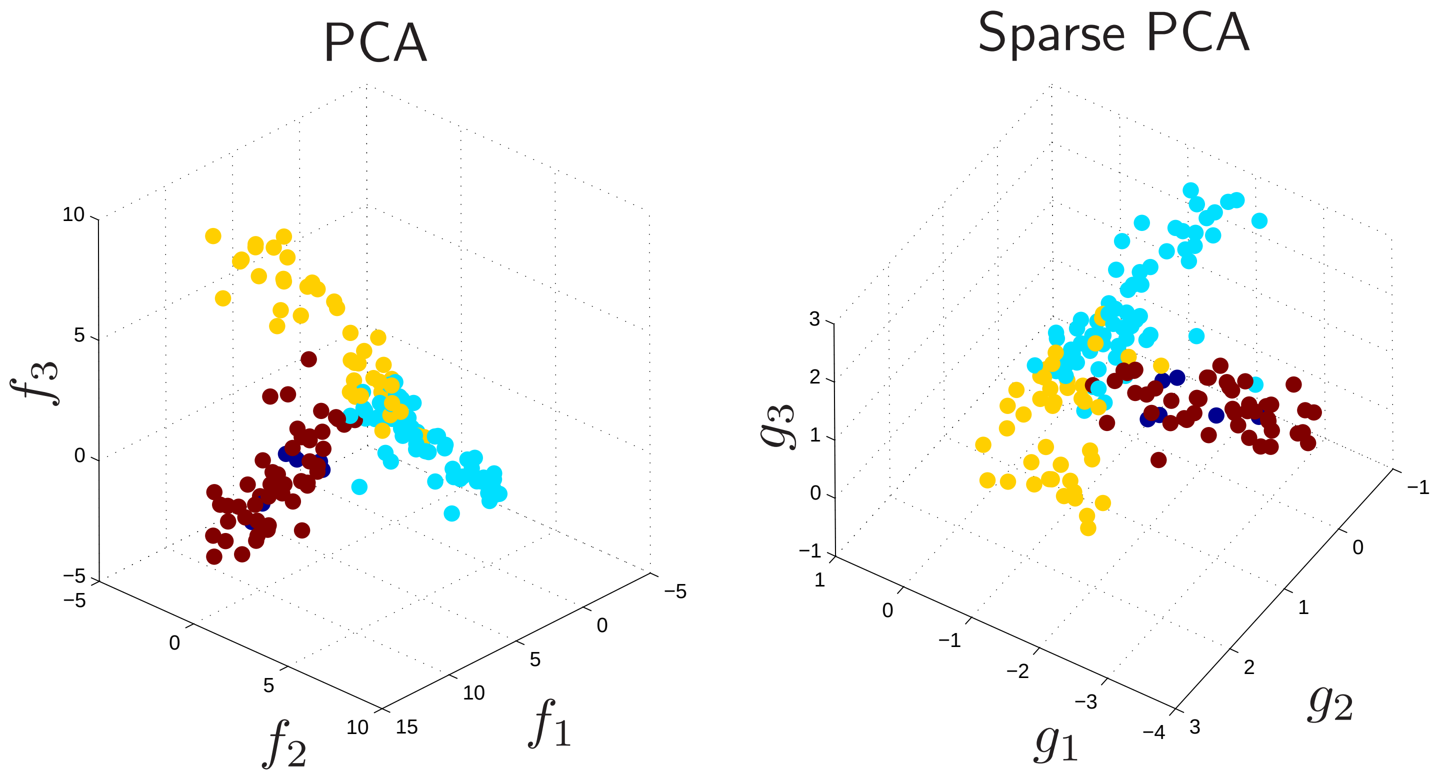
- Principal component analysis
- *The sparse PCA problem*
- New formulation and SDP relaxation
- Quality estimate
- Sparsity in linear systems

Looking for sparse factors

Gene expression data analysis: "explaining data with a few genes"

- PCA is used for clustering and visualizing data (gene responses vs. drugs)
- principal axes represent a combination of genes that are important in explaining data
- the *sparser* the axes, the less genes are involved
- ultimately, a short list of genes that explain data could yield a *universal diagnostic chip*

PCA vs. sparse PCA: example



Clustering of gene expression data in the PCA versus sparse PCA basis with 500 genes. The factors f on the left are dense and each use all 500 genes while the sparse factors g_1, g_2 and g_3 on the right involve *6, 4 and 4 genes* respectively. *(Data source: Iconix Pharmaceuticals, Inc.)*

Some previous work

- Vines (2000): restrict the factors' coefficients in a small set of integers, such as 0, 1, and -1
- Cadima and Jolliffe (1995): simple threshold approach
- Jolliffe and Udin (2003): SCoTLASS
- Zou, Hastie and Tibshirani (2004): write PCA as a regression problem, and add a l_1-norm penalty to it
- d'Aspremont, El Ghaoui, Jordan, Lanckriet (2004): Direct sparse PCA

Direct Sparse PCA

- *Cardinality-penalized variational problem:*

$$\max_x \; x^T \Sigma x - \rho \|x\|_0 \;\; : \;\; \|x\|_2 = 1$$

  where $\rho > 0$, and $\|x\|_0$ denotes the number of non-zero elements in x

- Let $X = xx^T$, and approximate problem by

$$\max_X \; \mathbf{Tr}\, \Sigma X - \rho \|X\|_1 \;\; : \;\; X \succeq 0, \;\; \mathbf{Tr}\, X = 1, \;\; \mathbf{Rank}(X) = 1$$

  ($\|\cdot\|_1$ denotes sum of absolute values)

- Dropping the rank constraint leads to an SDP

Solving direct sparse PCA

- The direct sparse PCA problem

$$\max_X \ \mathbf{Tr}\, \Sigma X - \rho \|X\|_1 \ : \ X \succeq 0, \ \ \mathbf{Tr}\, X = 1$$

can be solved as an SDP, via general-purpose interior-point methods

Complexity: $O(n^6 \log(1/\epsilon))$

- For large-scale problems, first-order methods (Nesterov, 2005) can be used

Complexity: $O(n^4 \sqrt{\log n}/\epsilon)$

Problems with direct sparse PCA

- Direct sparse PCA relies on *two* relaxation steps:
  - *Lower bound on $\|\cdot\|_0$-norm:* via Cauchy-Schwartz inequality,

$$\forall\, x,\ \|x\|_2 = 1\ :\ \|x\|_0 \geq \|x\|_1^2$$

  - *Rank relaxation:* lift $xx^T \to X$, and drop rank constraint on X

- Analysis of the *quality* of the approximation seems to be difficult

Outline

- Principal component analysis
- The sparse PCA problem
- *New formulation and SDP relaxation*
- Quality estimate
- Sparsity in linear systems

Equality vs. inequality model

Sparse PCA problem:

$$\phi := \max_x \; x^T \Sigma x - \rho \|x\|_0 \;\;:\;\; \|x\|_2 = 1$$

We will develop SDP bounds for the related quantity:

$$\tilde{\phi} := \max_x \; x^T \Sigma x - \rho \|x\|_0 \;\;:\;\; \|x\|_2 \leq 1$$

Fact: (assume WLOG $\Sigma_{11} \geq \ldots \geq \Sigma_{nn}$)

- If $\rho \geq \Sigma_{11}$, then $\tilde{\phi} = 0$, $\phi = \Sigma_{11} - \rho$ (with optimizer $x^* = e_1$)
- If $\rho < \Sigma_{11}$, then $\tilde{\phi} = \phi > 0$

In the sequel, assume $\rho < \Sigma_{11}$

Towards a new formulation

Our problem:

$$\phi := \max_x \; x^T \Sigma x - \rho \|x\|_0 \;\;:\;\; \|x\|_2 \le 1 \tag{1}$$

We have

$$\phi = \max_{u \in \{0,1\}^n} \; \max_{y^T y \le 1} \; y^T D(u) \Sigma D(u) y - \rho \cdot \mathbf{1}^T u, \tag{2}$$

where $D(u) := \mathbf{diag}(u)$

- The boolean vector u represents the *sparsity pattern* of an optimal solution
- Optimal (y, u) in (2) related to optimal x in (1) by

$$x = D(u) y$$

Towards a new formulation (cont'd)

Eliminating y in (2), obtain

$$
\begin{aligned}
\phi &= \max_x \; x^T \Sigma x - \rho \|x\|_0 \; : \; \|x\|_2 \leq 1 \\
&= \max_{u \in \{0,1\}^n} \; \max_{y^T y \leq 1} \; y^T D(u) \Sigma D(u) y - \rho \cdot \mathbf{1}^T u \\
&= \max_{u \in \{0,1\}^n} \; \lambda_{\max}(D(u) \Sigma D(u)) - \rho \cdot \mathbf{1}^T u
\end{aligned}
$$

- Optimal y is an eigenvector corresponding to $\lambda_{\max}$ above
- Optimal x is $x = D(u)y$

Towards a new formulation (cont'd)

- *Cholesky decomposition:* Let $\Sigma = A^T A$, where $A = [a_1 \ldots a_n]$, with $a_i \in \mathbb{R}^m$, $i = 1, \ldots, n$, and $m = \mathbf{Rank}(\Sigma)$
- Our previous formulation leads to a formulation based on *eigenvalue maximization:*

$$\phi = \max_{u \in \{0,1\}^n} \lambda_{\max}(D(u) A^T A D(u)) - \rho \cdot \mathbf{1}^T u$$

Towards a new formulation (cont'd)

- *Cholesky decomposition:* Let $\Sigma = A^T A$, where $A = [a_1 \dots a_n]$, with $a_i \in \mathbb{R}^m$, $i = 1, \dots, n$, and $m = \mathbf{Rank}(\Sigma)$

- Our previous formulation leads to a formulation based on *eigenvalue maximization:*

$$\begin{aligned} \phi &= \max_{u \in \{0,1\}^n} \lambda_{\max}(D(u) A^T A D(u)) - \rho \cdot \mathbf{1}^T u \\ &= \max_{u \in \{0,1\}^n} \lambda_{\max}(A D(u)^2 A^T) - \rho \cdot \mathbf{1}^T u \end{aligned}$$

Towards a new formulation (cont'd)

- *Cholesky decomposition:* Let $\Sigma = A^T A$, where $A = [a_1 \dots a_n]$, with $a_i \in \mathbb{R}^m$, $i = 1, \dots, n$, and $m = \mathbf{Rank}(\Sigma)$

- Our previous formulation leads to a formulation based on *eigenvalue maximization:*

$$
\begin{aligned}
\phi &= \max_{u \in \{0,1\}^n} \lambda_{\max}(D(u)A^T A D(u)) - \rho \cdot \mathbf{1}^T u \\
&= \max_{u \in \{0,1\}^n} \lambda_{\max}(A D(u)^2 A^T) - \rho \cdot \mathbf{1}^T u \\
&= \max_{u \in \{0,1\}^n} \lambda_{\max}(A D(u) A^T) - \rho \cdot \mathbf{1}^T u
\end{aligned}
$$

Eigenvalue maximization problem

- Using the convexity of the largest eigenvalue function, we obtain the representation

$$\phi = \max_{u \in [0,1]^n} \lambda_{\max}\left(\sum_{i=1}^{n} u_i a_i a_i^T\right) - \rho \cdot \mathbf{1}^T u.$$

- Set $B_i := a_i a_i^T - \rho \cdot I_m$, $i = 1, \ldots, n$, and express ϕ as

$$\phi = \max_{u \in [0,1]^n} \lambda_{\max}\left(\sum_{i=1}^{n} u_i B_i\right),$$

- The computation of ϕ can be interpreted as a *eigenvalue maximization problem*, where the sparsity pattern u is the decision variable

An alternate expression

We have

$$
\begin{aligned}
\phi &= \max_{u\in[0,1]^n} \lambda_{\max}\left(\sum_{i=1}^n u_i B_i\right) \\
&= \max_{u\in[0,1]^n} \max_{\xi^T\xi\leq 1} \xi^T\left(\sum_{i=1}^n u_i a_i a_i^T\right)\xi - \rho\cdot\mathbf{1}^T u \\
&= \max_{\xi^T\xi\leq 1} \sum_{i=1}^n ((a_i^T\xi)^2 - \rho\xi^T\xi)_+ \\
&= \max_{\xi^T\xi=1} \sum_{i=1}^n ((a_i^T\xi)^2 - \rho)_+
\end{aligned}
$$

Try rank relaxation?

$$\begin{aligned}
\phi &= \max_{\xi^T\xi=1} \sum_{i=1}^{n} ((a_i^T\xi)^2 - \rho)_+ \\
&= \max_{X} \sum_{i=1}^{n} \left(a_i^T X a_i - \rho\right)_+ \ : \ X \succeq 0, \ \ \mathbf{Tr}\, X = 1, \ \ \mathbf{Rank}(X) = 1 \\
&\leq \max_{X} \sum_{i=1}^{n} \left(a_i^T X a_i - \rho\right)_+ \ : \ X \succeq 0, \ \ \mathbf{Tr}\, X = 1
\end{aligned}$$

- Rank relaxation is actually exact ($\leq$ is an equality) . . .
- . . . Unfortunately, it is useless as the rank-relaxed problem is still not convex!

Recovering the sparsity pattern

We have obtained

$$\phi = \max_{\xi^T \xi = 1} \sum_{i=1}^{n} ((a_i^T \xi)^2 - \rho)_+$$

- An optimal sparsity pattern u is obtained from an optimal solution ξ to the above problem by setting

$$u_i = \begin{cases} 1 & \text{if } (a_i^T \xi)^2 > \rho, \\ 0 & \text{otherwise} \end{cases}$$

  Thus, for every i such that $\rho \geq a_i^T a_i$, we can always assume that the optimal sparsity pattern satisfies $u_i = 0$ *(ignore a_i)*

- In the sequel, we assume WLOG $a_i^T a_i > \rho$ for every i

SDP relaxation

Our new formulation is (having set $B_i = a_i a_i^T - \rho \cdot I_m$):

$$\phi = \max_{u \in [0,1]^n} \lambda_{\max}\left(\sum_{i=1}^n u_i B_i\right)$$

SDP relaxation:

$$\phi \leq \psi := \min_{(Y_i)_{i=1}^n} \lambda_{\max}\left(\sum_{i=1}^n Y_i\right) \ : \ Y_i \succeq B_i, \ \ Y_i \succeq 0, \ \ i = 1, \ldots, n$$

SDP relaxation

Our new formulation is (having set $B_i = a_i a_i^T - \rho \cdot I_m$):

$$\phi = \max_{u \in [0,1]^n} \lambda_{\max} \left(\sum_{i=1}^n u_i B_i \right)$$

SDP relaxation:

$$\phi \leq \psi := \min_{(Y_i)_{i=1}^n} \lambda_{\max} \left(\sum_{i=1}^n Y_i \right) \ : \ Y_i \succeq B_i, \ \ Y_i \succeq 0, \ \ i = 1, \dots, n$$

Proof: if $(Y_i)_{i=1}^n$ is feasible for the above SDP, then for every $\xi \in \mathbb{R}^m$, $\xi^T \xi \leq 1$, and $u \in [0,1]^n$, we have

$$\xi^T \left(\sum_{i=1}^n u_i B_i \right) \xi \leq \sum_{i=1}^n (\xi^T B_i \xi)_+ \leq \xi^T \left(\sum_{i=1}^n Y_i \right) \xi \leq \psi$$

Dual of SDP relaxation

Dual problem is

$$\psi = \max_{X,(P_i)_{i=1}^n} \sum_{i=1}^n \mathbf{Tr}\, P_i B_i \;:\; X \succeq P_i \succeq 0, \;\; i = 1, \ldots, n, \;\; \mathbf{Tr}\, X = 1$$

$$= \max_X \sum_{i=1}^n \mathbf{Tr} \left(X^{1/2} a_i a_i^T X^{1/2} - \rho X \right)_+ \;:\; X \succeq 0, \;\; \mathbf{Tr}\, X = 1,$$

where $\mathbf{Tr}\, B_+ =$ sum of non-negative eigenvalues of symmetric matrix B

Dual of SDP relaxation (cont'd)

The bound $\phi \le \psi$ can also be inferred directly from the dual:

$$
\begin{aligned}
\phi &= \max_{\xi^T\xi=1} \sum_{i=1}^{n} ((a_i^T\xi)^2-\rho)_+ \\
&= \max_X \sum_{i=1}^{n} \left(a_i^T X a_i - \rho\right)_+ \;:\; X \succeq 0,\; \mathbf{Tr}\, X = 1,\; \mathbf{Rank}(X)=1 \\
&= \max_X \sum_{i=1}^{n} \mathbf{Tr}\left(X^{1/2} a_i a_i^T X^{1/2} - \rho X\right)_+ \;:\; X \succeq 0,\; \mathbf{Tr}\, X = 1,\; \mathbf{Rank}(X) = \\
&\le \max_X \left\{ \sum_{i=1}^{n} \mathbf{Tr}\left(X^{1/2} a_i a_i^T X^{1/2} - \rho X\right)_+ \;:\; X \succeq 0,\; \mathbf{Tr}\, X = 1 \right\} = \psi
\end{aligned}
$$

If $\mathbf{Rank}(X)=1$ at the optimum of the dual problem, then $\le$ becomes an equality, and $\phi=\psi$

Outline

- Principal component analysis
- The sparse PCA problem
- New formulation and SDP relaxation
- *Quality estimate*
- Sparsity in linear systems

Quality of SDP relaxation (1)

(Inspired by Ben-Tal & Nemirovski, 2002)

Upper bound: $\phi \leq \psi = \max_{X \succeq 0, \, \mathbf{Tr} X = 1} \sum_{i=1}^{n} \mathbf{Tr} \left(X^{1/2} a_i a_i^T X^{1/2} - \rho X \right)_+$

- Let $X \succeq 0$, $\mathbf{Tr}\, X = 1$, be optimal for ψ, so that

$$\psi = \sum_{i=1}^{n} \alpha_i,$$

where

$$B_i(X) := X^{1/2} B_i X^{1/2} = X^{1/2} (a_i a_i^T - \rho I) X^{1/2}, \quad \alpha_i := \mathbf{Tr}(B_i(X)_+)$$

- Let $k := \mathbf{Rank}(X)$, assume $k > 1$

Quality of SDP relaxation (2)

(Fix $i \in \{1, \ldots, n\}$, drop subscript on $\alpha_i, B_i(X) = X^{1/2}(a_i a_i^T - \rho I)X^{1/2}$)

- In view of our assumption $\min_i a_i^T a_i > \rho$, $B(X)$ has exactly *one positive eigenvalue*, equal to $\alpha = \mathbf{Tr}\, B_+$

- Denote by $-\beta_j$ ($\beta_j > 0$) the remaining non-zero eigenvalues; one can show that
$$\sum_{j=1}^{k-1} \beta_j \leq \rho.$$

- Assume $\xi \sim \mathcal{N}(0, I_m)$; by rotational invariance of the normal distribution:

$$\mathbf{E}(\xi^T B(X)\xi)_+ = \mathbf{E}\left(\alpha\xi_1^2 - \sum_{j=1}^{k-1} \beta_j \xi_{j-1}^2\right)_+$$

Quality of SDP relaxation (3)

Thus

$$\begin{aligned}
\mathbf{E}(\xi^T B(X)\xi)_+ &\geq \min_{\beta\geq 0,\, \sum_j \beta_j \leq \rho} \mathbf{E}\left(\alpha\xi_1^2 - \sum_{j=1}^{k-1} \beta_j \xi_{j+1}^2\right)_+ \\
&= \mathbf{E}\left(\alpha\xi_1^2 - \frac{\rho}{k-1}\sum_{j=1}^{k-1} \xi_{j+1}^2\right)_+ \\
&\geq \left(\alpha - \rho + \frac{2}{\pi}\sqrt{\alpha^2 + \frac{\rho^2}{k-1}}\right)_+
\end{aligned}$$

Here we have used a result in Ben-Tal & Nemirovski (2002):

$$\forall\, \gamma \in \mathbb{R}^d \;:\; \mathbf{E}\left|\sum_{i=1}^{d} \gamma_i \xi_i^2\right| \geq \frac{2}{\pi}\|\gamma\|_2$$

Quality of SDP relaxation (4)

Summing over i, and with $\alpha_i := \mathbf{Tr}(B_i(X)_+)$, $\psi = \sum_{i=1}^n \alpha_i$, we get:

$$\begin{aligned}
\mathbf{E}\sum_{i=1}^n (\xi^T B_i(X)\xi)_+ &\geq \sum_{i=1}^n \left(\alpha_i - \rho + \frac{2}{\pi}\sqrt{\alpha_i^2 + \frac{\rho^2}{k-1}}\right)_+ \\
&\geq \frac{1}{2}\left(\psi - n\rho + \frac{2}{\pi}\sqrt{\psi^2 + \frac{n^2\rho^2}{k-1}}\right)_+ \\
&\geq \frac{1}{\pi}\psi \quad \left(= \frac{1}{\pi}\psi\,\mathbf{E}(\xi^T X\xi)\right),
\end{aligned}$$

provided $\psi \geq n\rho$.

Quality of SDP relaxation (5)

Assuming $\psi \geq n\rho$:

- The previous bound implies that there exist $\xi \in \mathbb{R}^m$ such that

$$\sum_{i=1}^{n} (\xi^T B_i(X)\xi)_+ \geq \frac{\psi}{\pi}(\xi^T X \xi).$$

- Thus, with $u_i = 1$ if $\xi^T B_i(X)\xi > 0$, $u_i = 0$ otherwise, we obtain that there exist $\xi \in \mathbb{R}^m$ and $u \in [0,1]^n$ such that

$$\xi^T \left(\sum_{i=1}^{n} u_i B_i(X) \right) \xi \geq \frac{\psi}{\pi}(\xi^T X \xi).$$

Quality of SDP relaxation (5)

- With $z = X^{1/2}\xi$:

$$z^T \left(\sum_{i=1}^{n} u_i B_i \right) z \geq \frac{\psi}{\pi} \cdot (z^T z).$$

- We conclude that there exist $u \in [0,1]^n$ such that

$$(\psi \geq \phi \geq)\ \lambda_{\max} \left(\sum_{i=1}^{n} u_i B_i \right) \geq \frac{1}{\pi} \psi$$

Quality of SDP relaxation (6)

When is condition $\psi \geq n\rho$ met?

Find a lower bound on ψ:

$$(\psi \geq)\phi = \max_X \sum_{i=1}^n \left(a_i^T X a_i - \rho\right)_+ \; : \; X \succeq 0, \;\; \mathbf{Tr}\, X = 1$$

$$\geq \max_i a_i^T a_i - \rho \quad \text{(choose } X = a_j a_j^T/(a_j^T a_j)\text{, where } j := \arg\max_i a_i^T a_i\text{)}$$

Thus, condition $\psi \geq n\rho$ is met when $\rho \leq \frac{1}{n+1} \max_i a_i^T a_i$ . . .

. . . Don't forget we assumed *$\rho < a_i^T a_i$ for every i* . . .

Quality of SDP relaxation: summary

Theorem: Assume

$$\rho < \min\left(\min_{1\leq i\leq n} \Sigma_{ii}, \frac{1}{n+1} \max_{1\leq i\leq n} \Sigma_{ii}\right).$$

Then,

$$\frac{1}{\pi}\psi \leq \phi \leq \psi. \tag{3}$$

Quality of SDP relaxation: summary

Theorem: Assume

$$\rho < \min\left(\min_{1\le i\le n} \Sigma_{ii}, \frac{1}{n+1} \max_{1\le i\le n} \Sigma_{ii}\right).$$

Then,

$$\frac{1}{\pi}\psi \le \phi \le \psi. \tag{4}$$

Corollary: Assume (WLOG) $\Sigma_{11} \ge \ldots \ge \Sigma_{nn}$. If Σ satisfies

$$\forall\, p \in \{2, \ldots, n\} \ : \ \Sigma_{pp} < \frac{1}{p} \max_i \Sigma_{ii},$$

Then (4) holds for every $\rho < \Sigma_{22}$.

Outline

- Principal component analysis
- The sparse PCA problem
- New formulation and SDP relaxation
- Quality estimate
- *Sparsity in linear systems*

Sparse solutions of linear equations

Minimum cardinality problem:

$$\phi := \min \ \|x\|_0 \ : \ Ax = b$$

where

- $m \leq n$, $A = [a_1, \ldots, a_n] \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$
- $\|x\|_0$ denotes the number of non-zero elements of x

The minimum cardinality problem

- Problem arises in a number of fields (compression, signal processing, etc)
- Problem is NP-hard
- A vast body of literature is attached to it

A classical approach: obtain a suboptimal solution by solving the LP

$$\min \|x\|_1 \ : \ Ax = b$$

In signal processing, approach is called *"basis pursuit"*

Some previous approaches

- *Convex approximation methods:*
  - Chen, Donoho (1994): basis pursuit
  - Tropp (2004-5): analyze l_1-norm approximation using QP duality
- *Bayesian methods:* Lewicki & Sejnowski (2000), Miller (2002)
- *Greedy methods:* *e.g.* Orthogonal Matching Pursuit, see Miller (2002)
- *Global optimization:* see Miller (2002)
- *Nonlinear optimization:* Rao, Kreutz-Delgado (1999)

A modified problem

We consider a slightly modified problem:

$$\phi(\sigma) := \min_x \; \|x\|_0 \;\; : \;\; Ax = b, \;\; \|x\|_2 \leq \sigma,$$

where $\sigma > 0$ is given.

- $\phi = \lim_{\sigma \to +\infty} \phi(\sigma)$
- Assume that A is full row rank, and that the above problem is feasible,*i.e.*

$$b^T (AA^T)^{-1} b \leq \sigma^2$$

- Norm constraint often makes sense from a practical point of view

A boolean SDP formulation

The problem can be formulated as

$$\phi(\sigma) = \min_{u,x} \mathbf{1}^T u \ : \ AD(u)y = b, \ \ \|y\|_2 \le \sigma, \ \ u \in \{0,1\}^n,$$

where $D(u) := \mathbf{diag}(u)$, and $x = D(u)y$

Lemma: $\exists\, y \in \mathbb{R}^n, \ \|y\|_2 \le 1, \ By = b \Longleftrightarrow BB^T \succeq bb^T$

Thus

$$\begin{aligned} \phi(\sigma) &= \min_{u} \mathbf{1}^T u \ : \ \sigma^2 AD(u)^2 A^T \succeq bb^T, \ \ u \in \{0,1\}^n \\ &= \min_{u} \mathbf{1}^T u \ : \ \sigma^2 \sum_{i=1}^n u_i a_i a_i^T \succeq bb^T, \ \ u \in \{0,1\}^n \end{aligned}$$

SDP bound

Relax the boolean constraint and obtain the lower bound

$$\phi(\sigma) \geq \psi(\sigma) := \min_u \mathbf{1}^T u \ : \ \sigma^2 \sum_{i=1}^n u_i a_i a_i^T \succeq bb^T, \ \ u \in [0,1]^n$$

This an SDP, with *dual:*

$$\psi(\sigma) = \max_{X \succeq 0} (b^T X b)/\sigma^2 - \sum_{i=1}^n (a_i^T X a_i - 1)_+$$

An SOCP representation of the bound

The SDP bound can be expressed as

$$\psi(\sigma) = \psi(\sigma) := \min_u \mathbf{1}^T u \ : \ b^T \left(\sum_{i=1}^n u_i a_i a_i^T \right)^{-1} b \leq \sigma^2, \ \ u \in [0,1]^n$$

Usign QCQP duality, we obtain the equivalent representation

$$\psi(\sigma) = \max_{z, \, \mu \geq 0} 2b^T z - \mu\sigma^2 - \sum_{i=1}^n ((a_i^T z)^2/\mu - 1)_+$$

- Above problem can be expressed as a *(rotated cone) SOCP*
- As such, can be efficiently solved

Link with the l_1-norm approximation

We can also express the previous SOCP as the (non-convex) QCQP

$$\psi(\sigma) = \max_{\xi} \, (b^T \xi)^2 / \sigma^2 - \sum_{i=1}^{n} ((a_i^T \xi)^2 - 1)_+$$

For $\sigma \to \infty$, the solution set to above problem converges to that of the LP

$$\psi = \max_{y} \, b^T \xi \ : \ |a_i^T \xi| \leq 1, \ \ i = 1, \ldots, n,$$

which is the (dual of) the classical LP relaxation

Extensions

New formulations and bounds can be extended to other problems:

- sparse solutions to linear *inequalities:*

$$\phi := \min \ \|x\|_0 \ : \ Ax \leq b$$

(Hint: previous formulation is convex in b . . .)

- *penalized* versions, such as

$$\phi := \min_x \ \|Ax - b\|_2^2 + \rho \|x\|_0$$

Challenges

- Evaluate the quality of the SOCP bound
- Investigate the results for $\sigma \to \infty$

Wrap-up

- We investigated problems involving sparsity and linear systems
- We devised new formulations and corresponding SDP relaxations
- For the sparse PCA problem we obtained a quality estimate valid for small penalty ρ
- Refined results in

L. El Ghaoui, Eigenvalue Maximization in Sparse PCA,
`http://arxiv.org/abs/math.OC/0601448`