Rank Constrained Matrix Optimization Problems

Defeng Sun

Department of Mathematics and Risk Management Institute
National University of Singapore

This talk is based on a joint work with Yan Gao at NUS

May 20, 2010

To use a low rank matrix to approximate a given matrix dates back to E. Schmidt [Math. Ann. 63 (1907), pp. 433–476] and C. Eckart and G. Young [Psychometrika 1 (1936), pp. 211–218]:

$$\begin{array}{ll} \min & \frac{1}{2}\|X - Z\|_F^2 \\ \text{s.t.} & \text{rank}(X) \leq r \end{array} \tag{1}$$

admits an analytic solution for a given $Z \in \Re^{m \times n}$ ($m \leq n$ without loss of generality):

$$X^* = \sum_{i=1}^{r} \sigma_i(Z) u_i v_i^T,$$

where Z has the following singular value decomposition (SVD):

$$Z = U[\text{diag}(\sigma(Y)) \; 0]V^T .$$

The matrix completion example:

$$\min \Big\{ \text{rank}(X) \ : \ X_{ij} \approx M_{ij} \ \ \forall \ (i,j) \in \Omega \Big\},$$

where

$$\Omega \in \{1, \ldots, p\} \times \{1, \ldots, q\} :$$

$$\begin{bmatrix} * & & & * & \\ & * & & & * \\ * & & & * & \\ & & * & * & \end{bmatrix}$$

A relaxed convex problem:

$$\min \Big\{ \|X\|_* \ : \ X_{ij} \approx M_{ij} \ \ \forall \ (i,j) \in \Omega \Big\},$$

$$\|X\|_* = \sum_{i=1}^{k} \sigma_i(X)$$

and $\sigma_i(X)$ are the singular values of X.
Further relaxation:

$$\min \Big\{ \frac{1}{2} \sum_{(i,j)\in\Omega} (X_{ij} - M_{ij})^2 + \rho \|X\|_* \Big\}.$$

The Netflix Prize problem: the convex relaxation is pretty good.
http://www.netflixprize.com/index

It works very well in practice and has a theoretical guarantee [refer to as the "l_1"-revolution – Donoho, Tao, Candes]

In many applications such as in image processing, we not only seek a low rank matrix, but also want the matrix to have certain desirable properties:

- $X \geq 0$ -component-wisely
- X is in special class of matrices (Hankel, Toeplitz, tri-diagonal, for examples)
- The rank of X may not be small, but be less than a given number
- Many others.

The theory breaks down ...

Let us look at an example from finance (**rank constrained covariance matrix problem**):

$$
\begin{aligned}
\min \quad & \|H \circ (X - G)\|_F^2 \\
\text{s.t.} \quad & X_{ii} = 1, \ i = 1, \ldots, n \\
& X_{ij} = e_{ij}, \ (i,j) \in \mathcal{B}_e \,, \\
& X_{ij} \geq l_{ij}, \ (i,j) \in \mathcal{B}_l \,, \\
& X_{ij} \leq u_{ij}, \ (i,j) \in \mathcal{B}_u \,, \\
& X \in \mathcal{S}_+^n \,, \\
& \mathrm{rank}(X) \leq r \,,
\end{aligned}
\tag{2}
$$

where $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u$ are three index subsets of $\{(i,j) \,|\, 1 \leq i < j \leq n\}$ satisfying $\mathcal{B}_e \cap \mathcal{B}_l = \emptyset$, $\mathcal{B}_e \cap \mathcal{B}_u = \emptyset$, and $l_{ij} < u_{ij}$ for any $(i,j) \in \mathcal{B}_l \cap \mathcal{B}_u$.

continued

Here $\mathcal{S}^n$ and $\mathcal{S}^n_+$ are, respectively, the space of $n \times n$ symmetric matrices and the cone of positive semidefinite matrices in $\mathcal{S}^n$.

$\|\cdot\|_F$ is the Frobenius norm defined in $\mathcal{S}^n$ and " $\circ$ " is the Hardamard product [component-wise multiplication of two matrices].

$H \geq 0$ is a weight matrix.

- H_{ij} is larger if G_{ij} is better estimated.
- $H_{ij} = 0$ if G_{ij} is missing.

A matrix $X \in \mathcal{S}^n$ is called a correlation matrix if $X \succeq 0$ (i.e., $X \in \mathcal{S}^n_+$) and $X_{ii} = 1$, $i = 1, \ldots, n$.

The bad news is that for a correlation matrix $X \in \mathcal{S}_+^n$:

$$||X||_* = \text{trace}(X) = n.$$

So any convex relaxation of using the nuclear norm directly is doomed as one will simply add a **CONSTANT TERM** if one does so.

Worse than that: the rank r cannot be satisfied even if it may work in some cases.

A cure for these problems?

On January 15, 2010, I received the following email:
From: XXX@grupobbva.com
Sent: Friday, January 15, 2010 5:14 PM
To: Sun Defeng
Cc: XXX XXX
Subject: Nearest Correlation Matrix: Faster code request

Dear Mr. Sun,
Please let me introduce myself. My name is XXX and I work in one of Spain's major banks, BBVA. The position that I hold is Quantitative Analyst.

We have been looking for quite a while for "nearest correlation matrix problem" algorithms until we found your paper "An augmented Lagrangian dual approach for the H-weighted nearest correlation matrix problem" ...,

which shows not only a feasible approach, but also robust and fast results. I was also happy to check and test the MATLAB code that you provide in your web page ..., with outstanding results. We are planning to apply your algorithm to large scale problems (around 2000x2000 correlation matrixes) through a C++ implementation using LAPACK library routines; this is why we are particularly interested in performance. Could you please provide us with any faster code (MATLAB or other) for this matter?
Thank you in advance and sorry for any inconvenience this may cause you.
Regards,
XXX

On November 18, 2009, I received the following email:
From: XXXXX@fortis.com
Sent: Wednesday, November 18, 2009 5:11 PM
To: Sun Defeng
Subject: nearest correlation matrix

Dear Professor Sun,

For R&D purpose, I am currently using your algorithms CorNewton and CorNewton3_Wnorm, which I downloaded from your webpage.

The results look very satisfactory. I was wondering whether you would have another version of the algorithm available in C or C++.

Best Regards,

Dr. XXX XXX
BNP Paribas Equity Derivatives Quantitative Research

On October 27, 2009, I received this from Universiteit van Tilburg:

My thesis is about correlations in a pension fund pooling. It is important for economic capital calculations. For some risks such as operational risk, I dont have data and hence I need to consult for an expert opinion. Then I might end up with not PSD matrices. Therefore, I need to calculate nearest correlation matrix.

In my given correlation matrix, I want to fix the correlations, which are data driven and I want the rest of the correlations not smaller than 0.1 from original matrix.

Your code is very convenient for my study. However, ...

On November 3, 2009:

Thank you for your valuable time, comments and helping me about solving my problem.

I gave no chance that my fixed constraints could be non-PSD before. Your advice solves the problem. I will modify my study in the light of it.

A simple correlation matrix model

$$
\begin{aligned}
\min \quad & \|H \circ (X - G)\|_F^2 \\
\text{s.t.} \quad & X_{ii} = 1\,, \;\; i = 1, \ldots, n \\
& X \succeq 0\,, \\
& \operatorname{rank}(X) \leq r\,.
\end{aligned}
\tag{3}
$$

The simplest corr. matrix model

$$
\begin{aligned}
\min \quad & \|(X - G)\|_F^2 \\
\text{s.t.} \quad & X_{ii} = 1\,, \;\; i = 1, \ldots, n \\
& X \succeq 0\,, \\
& \operatorname{rank}(X) \leq r\,.
\end{aligned}
\tag{4}
$$

In finance and statistics, correlation matrices are in many situations found to be inconsistent, i.e., $X \ngeq 0$.

These include, but are not limited to,

- Structured statistical estimations; data come from different time frequencies
- Stress testing regulated by Basel II;
- Expert opinions in reinsurance, and etc.

One correlation matrix

Partial market data[1]

$$G = \begin{bmatrix} 1.0000 & 0.9872 & 0.9485 & 0.9216 & -0.0485 & -0.0424 \\ 0.9872 & 1.0000 & 0.9551 & 0.9272 & -0.0754 & -0.0612 \\ 0.9485 & 0.9551 & 1.0000 & 0.9583 & -0.0688 & -0.0536 \\ 0.9216 & 0.9272 & 0.9583 & 1.0000 & -0.1354 & -0.1229 \\ -0.0485 & -0.0754 & -0.0688 & -0.1354 & 1.0000 & 0.9869 \\ -0.0424 & -0.0612 & -0.0536 & -0.1229 & 0.9869 & 1.0000 \end{bmatrix}$$

The eigenvalues of G are: $0.0087, 0.0162, 0.0347, 0.1000$, 1.9669, and 3.8736.

[1] RiskMetrics (`www.riskmetrics.com/stddownload_edu.html`)

Stress tested

Let's change G to

[change $G(1,6) = G(6,1)$ from -0.0424 to -0.1000]

$$\begin{bmatrix} 1.0000 & 0.9872 & 0.9485 & 0.9216 & -0.0485 & \mathbf{-0.1000} \\ 0.9872 & 1.0000 & 0.9551 & 0.9272 & -0.0754 & -0.0612 \\ 0.9485 & 0.9551 & 1.0000 & 0.9583 & -0.0688 & -0.0536 \\ 0.9216 & 0.9272 & 0.9583 & 1.0000 & -0.1354 & -0.1229 \\ -0.0485 & -0.0754 & -0.0688 & -0.1354 & 1.0000 & 0.9869 \\ \mathbf{-0.1000} & -0.0612 & -0.0536 & -0.1229 & 0.9869 & 1.0000 \end{bmatrix}$$

The eigenvalues of G are: -0.0216, 0.0305, 0.0441, 0.1078, 1.9609, and 3.8783.

Missing data

On the other hand, some correlations may not be reliable or even missing:

$$G = \begin{bmatrix} 1.0000 & 0.9872 & 0.9485 & 0.9216 & -0.0485 & --- \\ 0.9872 & 1.0000 & 0.9551 & 0.9272 & -0.0754 & -0.0612 \\ 0.9485 & 0.9551 & 1.0000 & 0.9583 & -0.0688 & -0.0536 \\ 0.9216 & 0.9272 & 0.9583 & 1.0000 & -0.1354 & -0.1229 \\ -0.0485 & -0.0754 & -0.0688 & -0.1354 & 1.0000 & 0.9869 \\ --- & -0.0612 & -0.0536 & -0.1229 & 0.9869 & 1.0000 \end{bmatrix}$$

Drop the rank constraint

Let us first consider the problem without the rank constraint:

$$\begin{array}{ll} \min & \frac{1}{2}\|H \circ (X - G)\|_F^2 \\ \text{s.t.} & X_{ii} = 1\,,\ i = 1, \ldots, n \\ & X \succeq 0\,. \end{array} \tag{5}$$

When $H = E$, the matrix of ones, we get

$$\begin{array}{ll} \min & \frac{1}{2}\|X - G\|_F^2 \\ \text{s.t.} & X_{ii} = 1\,,\ i = 1, \ldots, n \\ & X \succeq 0\,. \end{array} \tag{6}$$

which is known as the nearest correlation matrix (NCM) problem, a terminology coined by Nick Higham (2002).

The story starts

The NCM problem is a special case of the best approximation problem

$$\begin{aligned} \min \quad & \frac{1}{2}\|x-c\|^2 \\ \text{s.t.} \quad & \mathcal{A}x \in b+Q\,, \\ & x \in K\,, \end{aligned}$$

where $\mathcal{X}$ is a real Hilbert space equipped with a scalar product $\langle\cdot,\cdot\rangle$ and its induced norm $\|\cdot\|$, $\mathcal{A}:\mathcal{X}\to\Re^m$ is a bounded linear operator, $Q=\{0\}^p\times\Re^q_+$ is a polyhedral convex cone, $1\le p\le m$, $q=m-p$, and K is a closed convex cone in $\mathcal{X}$.

The KKT conditions

The Karush-Kuhn-Tucker conditions are

$$\begin{cases} (x - z) - c - \mathcal{A}^* y = 0 \\ \\ Q^* \ni y \perp \mathcal{A}x - b \in Q \\ \\ K^* \ni z \perp x \in K\,, \end{cases},$$

where "$\perp$" means the orthogonality. Q^* is the dual cone of Q and K^* is the dual cone of K.

Equivalently,

$$\begin{cases} (x - z) - c - \mathcal{A}^* y = 0 \\ Q^* \ni y \perp \mathcal{A}x - b \in Q \\ x - \Pi_K(x - z) = 0 \end{cases},$$

where $\Pi_K(x)$ is the unique optimal solution to

$$\begin{array}{ll} \min & \frac{1}{2}\|u - x\|^2 \\ \text{s.t.} & u \in K\,. \end{array}$$

Consequently, by first eliminating $(x - z)$ and then x, we get

$$Q^* \ni y \perp \mathcal{A}\Pi_K(c + \mathcal{A}^* y) - b \in Q\,,$$

which is equivalent to

$$F(y) := y - \Pi_{Q^*}[y - (\mathcal{A}\Pi_K(c + \mathcal{A}^* y) - b)] = 0, \quad y \in \Re^m\,.$$

The dual formulation

The above is nothing but the first order optimality condition to the convex dual problem

$$\begin{aligned} \max \quad & -\theta(y) := -\left[\frac{1}{2}\|\Pi_K(c + \mathcal{A}^* y)\|^2 - \langle b, y\rangle - \frac{1}{2}\|c\|^2\right] \\ \text{s.t.} \quad & y \in Q^* \,. \end{aligned}$$

Then F can be written as

$$F(y) = y - \Pi_{Q^*}(y - \nabla\theta(y)) \,.$$

Now, we only need to solve

$$F(y) = 0, \quad y \in \Re^m .$$

However, the difficulties are:

- F is not differentiable at y;
- F involves two metric projection operators;
- Even if F is differentiable at y, it is too costly to compute $F'(y)$.

The NCM problem

For the nearest correlation matrix problem,

- $\mathcal{A}(X) = \mathrm{diag}(X)$, a vector consisting of all diagonal entries of X.

.

- $\mathcal{A}^*(y) = \mathrm{diag}(y)$, the diagonal matrix.
- $b = e$, the vector of all ones in $\Re^n$ and $K = \mathcal{S}^n_+$.

Consequently, F can be written as

$$F(y) = \mathcal{A}\Pi_{\mathcal{S}^n_+}(G + \mathcal{A}^* y) - b.$$

The projector

For $n = 1$, we have

$$x_+ := \Pi_{\mathcal{S}^1_+}(x) = \max(0, x).$$

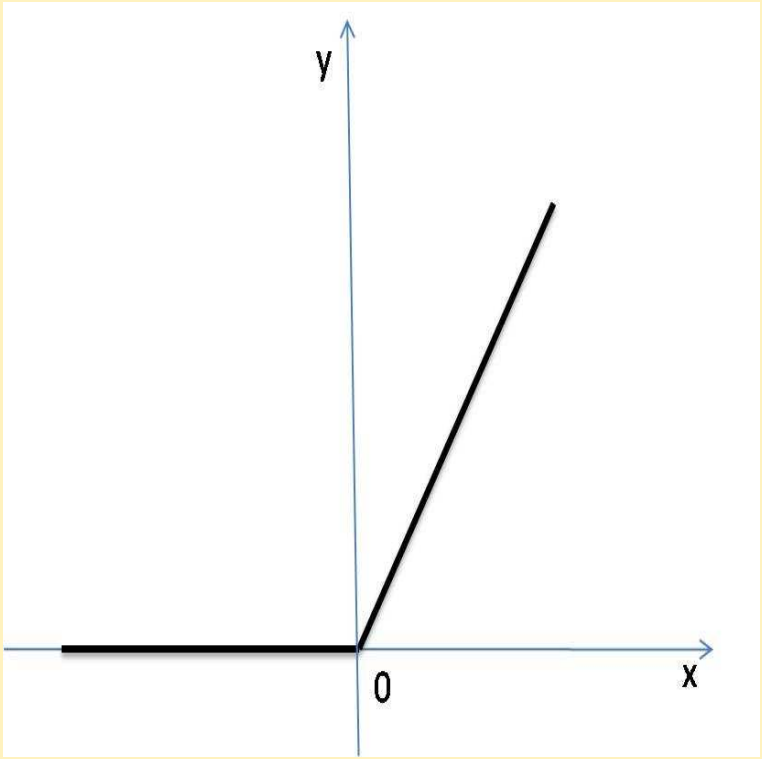
Note that

- x_+ is only piecewise linear, but not smooth.
- $(x_+)^2$ is continuously differentiable with

$$\nabla\left\{\frac{1}{2}(x_+)^2\right\} = x_+,$$

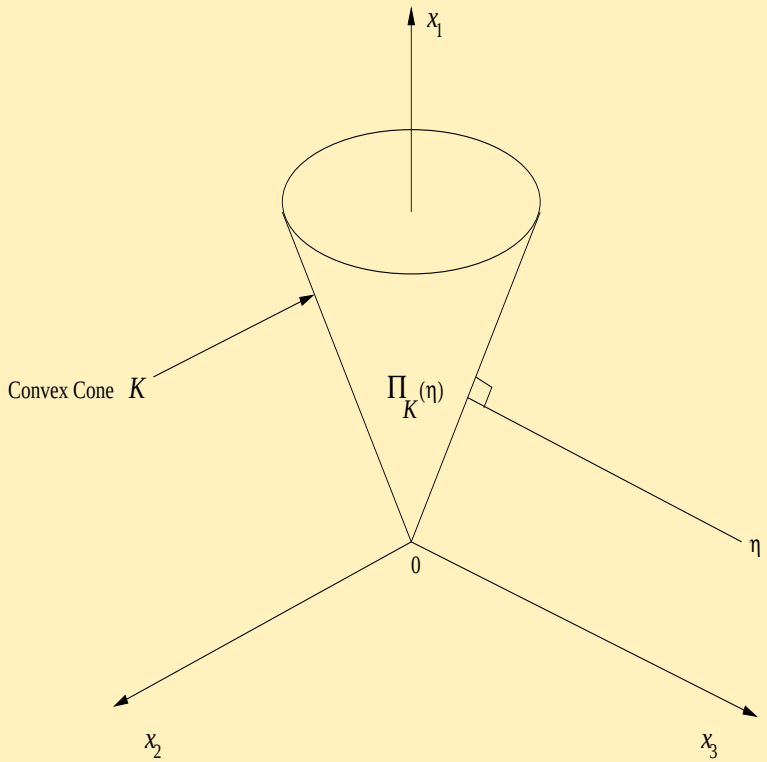
but is not twice continuously differentiable.

The one dimensional case



The multi-dimensional case

The projector for $K = \mathcal{S}_+^n$:



Let $X \in \mathcal{S}^n$ have the following spectral decomposition

$$X = P\Lambda P^T,$$

where Λ is the diagonal matrix of eigenvalues of X and P is a corresponding orthogonal matrix of orthonormal eigenvectors.

Then

$$X_+ := \Pi_{\mathcal{S}^n_+}(X) = P\Lambda_+ P^T.$$

We have

- $\|X_+\|^2$ is continuously differentiable with

$$\nabla\left(\frac{1}{2}\|X_+\|^2\right) = X_+,$$

but is not twice continuously differentiable.

- X_+ is not piecewise smooth, but strongly semismooth[2].

[2] D.F. Sun and J. Sun. Semismooth matrix valued functions. *Mathematics of Operations Research* 27 (2002) 150–169.

A quadratically convergent Newton's method is then designed by Qi and Sun[3]. The written code is called CorNewton.m.

"This piece of research work is simply great and practical. I enjoyed reading your paper." – March 20, 2007, a home loan financial institution based in McLean, VA.

"It's very impressive work and I've also run the Matlab code found in Defeng's home page. It works very well."– August 31, 2007, a major investment bank based in New York city.

[3] H.D. Qi and D.F. Sun, A quadratically convergent Newton method for computing the nearest correlation matrix. *SIAM Journal on Matrix Analysis and Applications* 28 (2006), pp. 360–385.

Inequality constraints

If we have lower and upper bounds on X, F takes the form

$$F(y) = y - \Pi_{Q^*}[y - (\mathcal{A}\Pi_{\mathcal{S}^n_+}(G + \mathcal{A}^* y) - b)]\,,$$

which involves double layered projections over convex cones.

A quadratically convergent inexact smoothing Newton-BICGStab method is designed by Gao and Sun[4].

Again, highly efficient.

[4] Y. Gao and D.F. Sun, Calibrating least squares covariance matrix problems with equality and inequality constraints, SIAM Journal on Matrix Analysis and Applications 31 (2009), pp. 1432–1457.

Back to the rank constraint

$$\begin{aligned} \min \quad & \frac{1}{2}\|H \circ (X - G)\|_F^2 \\ \text{s.t.} \quad & \mathcal{A}X \in b + Q\,, \\ & X \in \mathcal{S}_+^n\,, \\ & \operatorname{rank}(X) \leq k, \end{aligned}$$

equivalently,

$$\begin{aligned} \min \quad & \frac{1}{2}\|H \circ (X - G)\|_F^2 \\ \text{s.t.} \quad & \mathcal{A}X \in b + Q\,, \\ & X \in \mathcal{S}_+^n\,, \\ & \lambda_i(X) = 0,\ i = k+1, \ldots, n. \end{aligned}$$

The penalty approach

Given $c > 0$, we consider a penalized version

$$\begin{aligned} \min \quad & \frac{1}{2}\|H \circ (X - G)\|_F^2 + c \sum_{i=k+1}^{n} \lambda_i(X) \\ \text{s.t.} \quad & \mathcal{A}X \in b + Q\,, \\ & X \in \mathcal{S}_+^n\,, \end{aligned}$$

or equivalently

$$\begin{aligned} \min \quad & f_c(X) := \frac{1}{2}\|H \circ (X - G)\|_F^2 + c\langle I, X\rangle - c\sum_{i=1}^{k} \lambda_i(X) \\ \text{s.t.} \quad & \mathcal{A}X \in b + Q\,, \\ & X \in \mathcal{S}_+^n\,. \end{aligned}$$

Majorization functions

Let $h(X) := \sum_{i=1}^{k} \lambda_i(X) - \langle I, X \rangle$. Since h is a convex function, for given X^k, we have

$$h(X) \geq h^k(X) := h(X^k) + \langle V^k, X - X^k \rangle,$$

where $V^k \in \partial h(X^k)$. Thus, $-h$ is majorized by $-h^k$.

Let $d \in \Re^n$ be a positive vector such that

$$H \circ H \leq dd^T.$$

For example, $d = \max(H_{ij})e$. Let $D^{1/2} = \mathrm{diag}(d_1^{0.5}, \ldots, d_n^{0.5})$.

Let

$$g(X) := \frac{1}{2}\|H \circ (X - G)\|_F^2.$$

Then g is majorized by

$$g^k(X) := g(X^k) + \langle H \circ H(X^k - G), X - X^k \rangle + \frac{1}{2}\|D^{1/2}(X - X^k)D^{1/2}\|_F^2.$$

Thus, at X^k, f_c is majorized by

$$f_c(X) \leq f^k(X) := g^k(X) - ch^k(X)$$

and $f_c(X^k) = f^k(X^k)$.

The idea of majorization

Instead of solving the penalized problem, the idea of the majorization is to solve, for given X^k, the following problem

$$\begin{aligned} \min \quad & f_c^k(X) = g^k(X) - ch^k(X) \\ \text{s.t.} \quad & \mathcal{A}X \in b + Q\,, \\ & X \in \mathcal{S}_+^n\,, \end{aligned}$$

which is a diagonal weighted least squares correlation matrix problem

$$\begin{aligned} \min \quad & \frac{1}{2}\|D^{1/2}(X - X^k)D^{1/2}\|_F^2 \\ \text{s.t.} \quad & \mathcal{A}X \in b + Q\,, \\ & X \in \mathcal{S}_+^n\,. \end{aligned}$$

Now, we can use the two Newton methods introduced earlier for the majorized subproblems!

$$f_c(X^{k+1}) < f_c(X^k) < \cdots < f_c(X^1).$$

Where is the rank condition?

Looks good? But how can one guarantee that we can get a final X^* such that its rank is less or equal to k?

The answer is: increase c. That is, to have a sequence of $\{c_k\}$ with $c_{k+1} \geq c_k$.

Will it work? Numerical stability? Does not need a large c_k in numerical computations.

There are no known methods that can solve the general rank constrained problem. For the H-normed correlation matrix problems (without constraints on the off diagonal entries), the major.m of R. Pietersz and J.F. Groenen (2004) is the most efficient one so far [write $X = YY^T$ for $Y \in \Re^{n \times k}$ and apply component-by-component majorization.].

Let $Y \in \mathcal{S}^n$ be arbitrarily chosen. Suppose that Y has the spectral decomposition

$$Y = U\Sigma(Y)U^T, \tag{7}$$

where $U \in \mathcal{O}_n$ is a corresponding orthogonal matrix of orthonormal eigenvectors of Y and $\Sigma(Y) := \mathrm{diag}(\sigma(Y))$ where $\sigma(Y) = (\sigma_1(Y), \ldots, \sigma_n(Y))^T$ is the column vector containing all the eigenvalues of Y being arranged in the non-increasing order in terms of their absolute values, i.e.,

$$|\sigma_1(Y)| \geq \cdots \geq |\sigma_n(Y)|,$$

and whenever the equality holds, the larger one comes first, i.e.,

$$\text{if } |\sigma_i(Y)| = |\sigma_j(Y)| \text{ and } \sigma_i(Y) > \sigma_j(Y), \text{ then } i < j\,.$$

Define

$$\bar{\alpha} := \{i \mid |\sigma_i(Y)| > |\sigma_r(Y)|\}, \quad \bar{\beta} := \{i \mid |\sigma_i(Y)| = |\sigma_r(Y)|\},$$

$$\bar{\gamma} := \{i \mid |\sigma_i(Y)| < |\sigma_r(Y)|\},$$

and $\bar{\beta}^+ := \{i \mid \sigma_i(Y) = |\sigma_r(Y)|\}, \quad \bar{\beta}^- := \{i \mid \sigma_i(Y) = -|\sigma_r(Y)|\}.$

Denote

$$\begin{array}{rll} \Psi_r(Y) := & \min & \frac{1}{2}\|Z - Y\|^2 \\ & \text{s.t.} & Z \in \mathcal{S}^n(r). \end{array} \tag{8}$$

Denote the set of optimal solutions to (8) by $\Pi_{\mathcal{S}^n(r)}(Y)$.

Projection onto $\mathcal{S}^n(r)$

Lemma 1. *Let $Y \in \mathcal{S}^n$ have the spectral decomposition as in (7). Then the solution set $\Pi_{\mathcal{S}^n(r)}(Y)$ to problem (8) can be characterized as follows*

$$\Pi_{\mathcal{S}^n(r)}(Y) = \left\{ [\,U_{\bar\alpha}\; U_{\bar\beta}Q_{\bar\beta}\; U_{\bar\gamma}\,]\mathsf{diag}(v)\,[\,U_{\bar\alpha}\; U_{\bar\beta}Q_{\bar\beta}\; U_{\bar\gamma}\,]^T \Big| \right.$$
$$\left. v \in \mathcal{V}\,, Q_{\bar\beta} = \begin{bmatrix} Q_{\bar\beta+} & 0 \\ 0 & Q_{\bar\beta-} \end{bmatrix}, Q_{\bar\beta+} \in \mathcal{O}_{|\bar\beta+|}, Q_{\bar\beta-} \in \mathcal{O}_{|\bar\beta-|} \right\} \tag{9}$$

where

$$\mathcal{V} := \left\{ v \in \Re^n \,\middle|\, v_i = \sigma_i(Y) \text{ for } i \in \bar\alpha \cup \bar\beta_1,\; v_i = 0 \text{ for } i \in (\bar\beta \setminus \bar\beta_1) \cup \bar\gamma\,, \right.$$
$$\left. \text{where } \bar\beta_1 \subseteq \bar\beta \text{ and } |\bar\beta_1| = r - |\bar\alpha| \right\}. \tag{10}$$

Global Optimality Checking

Theorem 1. [5] *The optimal solution* $(\bar{y}, \overline{Y}) \in \mathcal{Q}^* \times \mathcal{S}^n$ *to the the dual problem satisfies*

$$b - \mathcal{A}\Pi_{\mathcal{S}^n_+}\left(C + \mathcal{A}^*\bar{y} + \overline{Y}\right) \in \mathcal{N}_{\mathcal{Q}^*}(\bar{y}) \tag{11}$$

and

$$\Pi_{\mathcal{S}^n_+}\left(C + \mathcal{A}^*\bar{y} + \overline{Y}\right) \in \operatorname{conv}\left\{\Pi_{\mathcal{S}^n(r)}(C - \overline{Y})\right\}, \tag{12}$$

where $\Pi_{\mathcal{S}^n(r)}(\cdot)$ *is defined as in Lemma 1. Furthermore, if there exists a matrix* $\overline{X} \in \Pi_{\mathcal{S}^n(r)}(C - \overline{Y})$ *such that* $\overline{X} = \Pi_{\mathcal{S}^n_+}\left(C + \mathcal{A}^*\bar{y} + \overline{Y}\right)$*, then* $\overline{X}$ *and* $(\bar{y}, \overline{Y})$ *globally solve the primal problem with* $H = E$ *and the corresponding dual problem, respectively and there is no duality gap between the primal and dual problems.*

---

[5] Y. Gao and D.F. Sun, A majorized penalty approach for calibrating rank constrained correlation matrix problems, manuscript, March 2010.

Testing Examples

The testing examples to be reported are given below.

Example 1. *Let* $n = 500$ *and the weight matrix* $H = E$*. For* $i, j = 1, \ldots, n$*,* $C_{ij} = 0.5 + 0.5e^{-0.05|i-j|}$*. The index sets are* $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$*.*

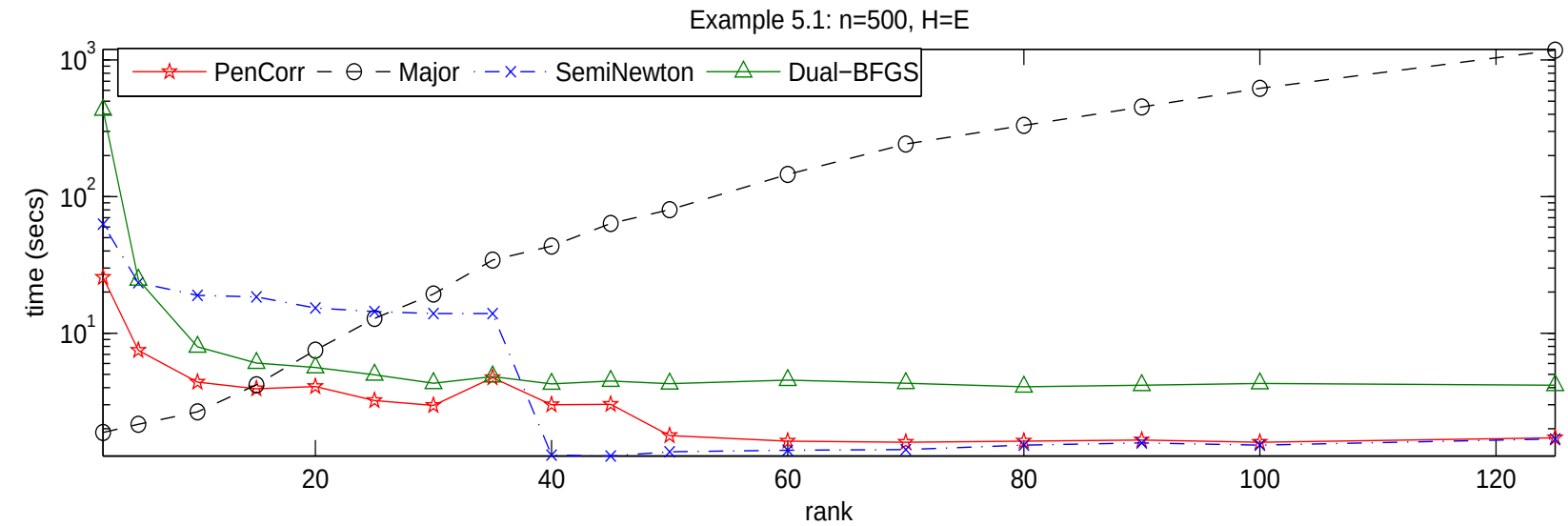
Example 2. *Let* $n = 500$ *and the weight matrix* $H = E$*. The matrix* C *is extracted from the correlation matrix which is based on a* $10,000$ *gene micro-array data set obtained from* 256 *drugs treated rat livers. The index sets are* $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$*.*

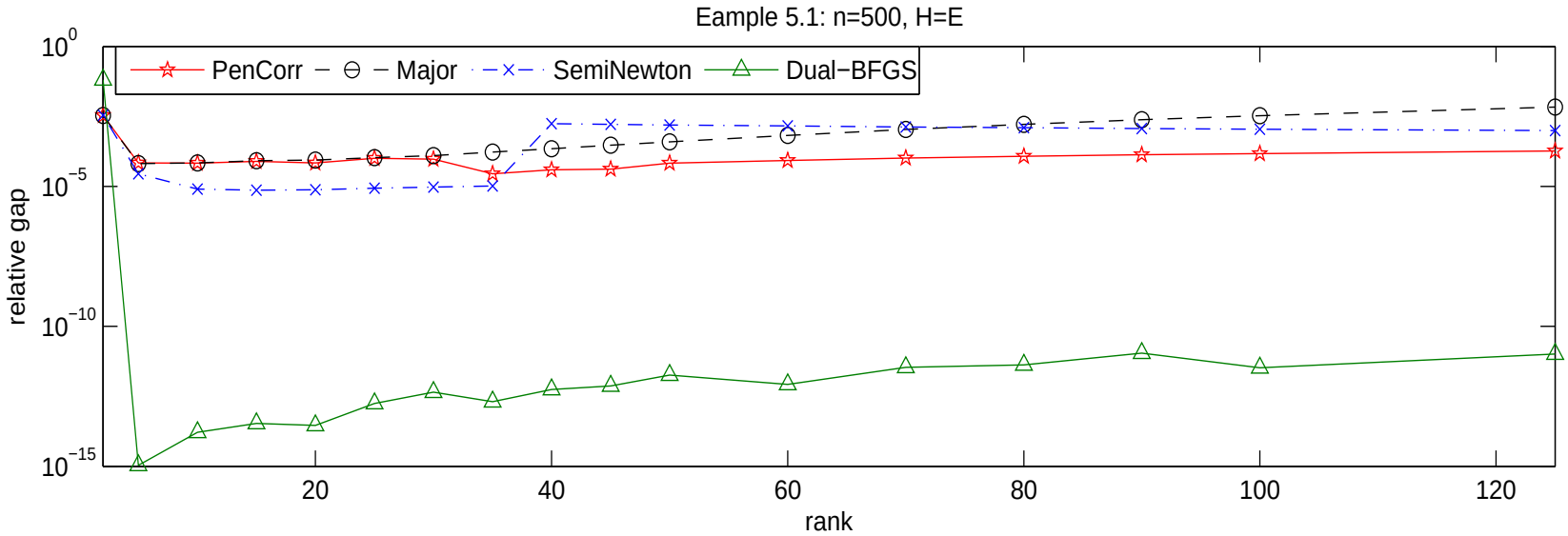
Example 3. *Let* $n = 500$*. The matrix* C *is the same as in Example 1, i.e.,* $C = 0.5 + 0.5e^{-0.05|i-j|}$ *for* $i, j = 1, \ldots, n$*. The index sets are* $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$*. The weight matrix* H *is generated in the way such that all its entries are uniformly distributed in* $[0.1, 10]$ *except for* 2×100 *entries in* $[0.01, 100]$*.*

Example 4. *Let $n = 500$. The matrix C is the same as in Example 2. The index sets are $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$. The weight matrix H is generated in the same way as in Example 3.*

Example 5. *The matrix C is obtained from the gene data sets with dimension $n = 1,000$ as in Example 2. The weight matrix H is the same as in Example 3. The index sets $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u \subset \{(i,j) \,|\, 1 \le i < j \le n\}$ consist of the indices of $\min(\hat{n}_r, n-i)$ randomly generated elements at the ith row of X, $i = 1, \ldots, n$ with $\hat{n}_r = 5$ for $\mathcal{B}_e$ and $\hat{n}_r = 10$ for $\mathcal{B}_l$ and $\mathcal{B}_u$. We take $e_{ij} = 0$ for $(i,j) \in \mathcal{B}_e$, $l_{ij} = -0.1$ for $(i,j) \in \mathcal{B}_l$ and $u_{ij} = 0.1$ for $(i,j) \in \mathcal{B}_u$.*

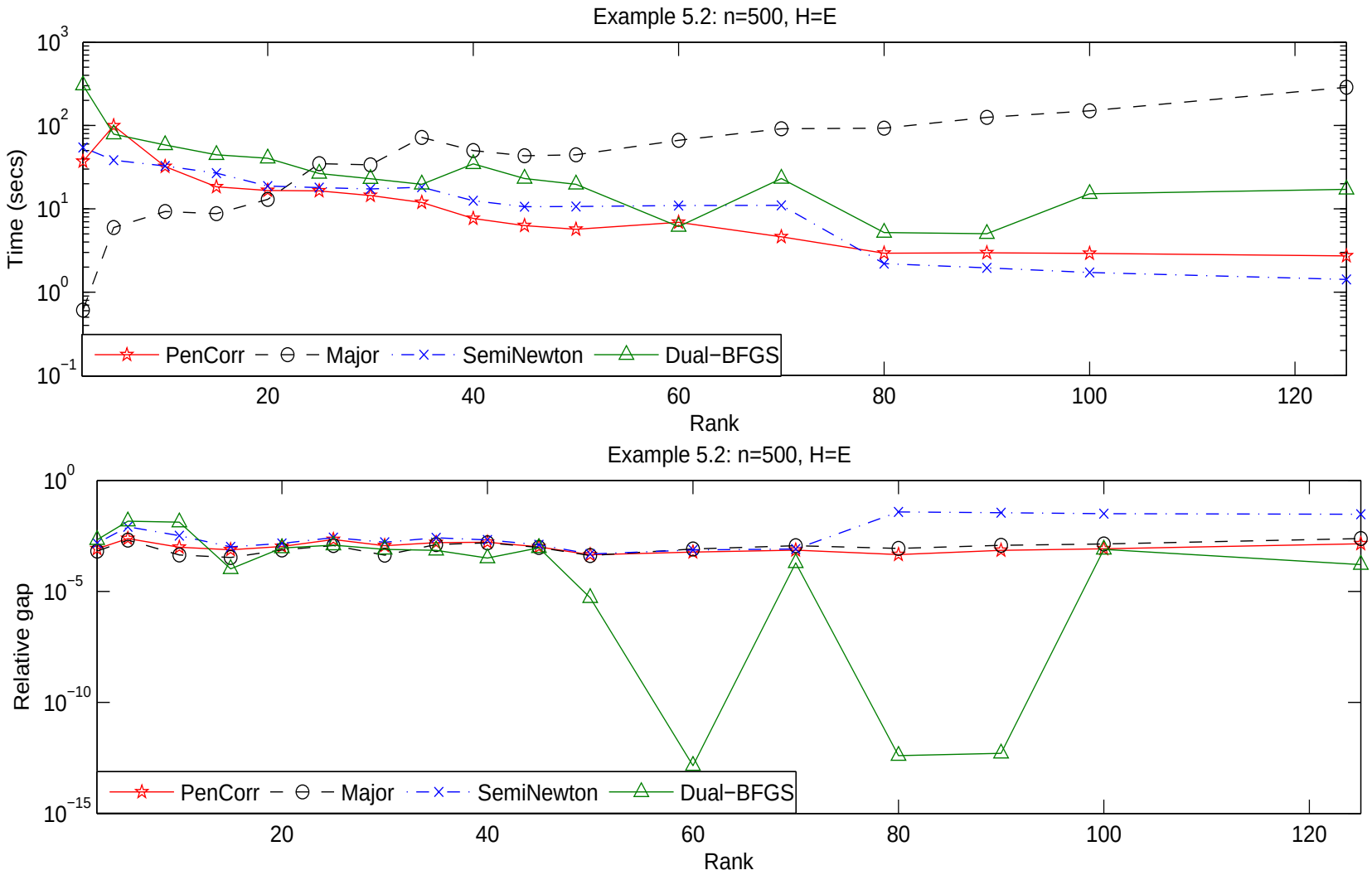
Numerical Results





Eg1	Major			SemiNewton			Dual-BFGS			PenCorr		
rank	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*
2	1.9	1.564e2	3.4e-3	63.0	1.564e2	3.5e-3	432.0	1.660e2	6.5e-2	25.7	1.564e2	3.4e-3
5	2.2	7.883e1	6.5e-5	23.5	7.883e1	2.8e-5	24.6	7.883e1	1.1e-15	7.5	7.883e1	7.0e-5
10	2.7	3.869e1	6.9e-5	19.0	3.868e1	8.0e-6	8.0	3.868e1	1.7e-14	4.4	3.869e1	6.7e-5
15	4.2	2.325e1	8.3e-5	18.5	2.324e1	7.3e-6	6.0	2.324e1	3.4e-14	3.9	2.325e1	7.9e-5
20	7.5	1.571e1	8.8e-5	15.3	1.571e1	7.6e-6	5.6	1.571e1	2.9e-14	4.1	1.571e1	6.9e-5
25	12.8	1.145e1	1.1e-4	14.4	1.145e1	8.6e-6	5.0	1.145e1	1.8e-13	3.2	1.145e1	1.0e-4
30	19.4	8.797e0	1.3e-4	14.0	8.796e0	9.5e-6	4.3	8.795e0	4.4e-13	3.0	8.796e0	9.4e-5
35	34.4	7.020e0	1.7e-4	14.0	7.019e0	1.0e-5	4.8	7.019e0	2.0e-13	4.7	7.019e0	2.8e-5
40	43.4	5.766e0	2.2e-4	1.3	5.774e0	1.7e-3	4.3	5.764e0	5.6e-13	3.0	5.765e0	3.9e-5
45	63.6	4.843e0	3.0e-4	1.3	4.849e0	1.6e-3	4.5	4.841e0	7.4e-13	3.0	4.841e0	4.2e-5
50	80.1	4.141e0	4.0e-4	1.4	4.146e0	1.6e-3	4.3	4.139e0	1.8e-12	1.8	4.139e0	6.8e-5
60	145.0	3.156e0	6.7e-4	1.4	3.158e0	1.4e-3	4.5	3.153e0	8.4e-13	1.6	3.154e0	8.4e-5
70	243.0	2.507e0	1.1e-3	1.4	2.507e0	1.3e-3	4.3	2.504e0	3.4e-12	1.6	2.504e0	1.0e-4
80	333.0	2.053e0	1.6e-3	1.5	2.052e0	1.2e-3	4.1	2.050e0	4.2e-12	1.6	2.050e0	1.2e-4
90	452.0	1.722e0	2.4e-3	1.6	1.720e0	1.2e-3	4.2	1.718e0	1.1e-11	1.7	1.718e0	1.4e-4
100	620.0	1.471e0	3.3e-3	1.5	1.468e0	1.1e-3	4.3	1.467e0	3.3e-12	1.6	1.467e0	1.5e-4
125	1180.0	1.055e0	6.8e-3	1.7	1.049e0	9.9e-4	4.2	1.048e0	1.0e-11	1.7	1.048e0	1.8e-4

Table 1: Numerical results for Example 1


Example 5.2: n=500, H=E
Time (secs)
Rank
PenCorr
Major
SemiNewton
Dual−BFGS
Example 5.2: n=500, H=E
Relative gap
Rank
PenCorr
Major
SemiNewton
Dual−BFGS

Eg2	Major			SemiNewton			Dual-BFGS			PenCorr		
rank	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*
2	0.6	2.858e2	6.5e-4	54.4	2.860e2	1.5e-3	304.5	2.862e2	2.1e-3	37.2	2.859e2	8.2e-4
5	6.0	1.350e2	2.0e-3	38.2	1.358e2	8.1e-3	78.8	1.367e2	1.5e-2	99.2	1.351e2	2.4e-3
10	9.3	6.716e1	4.4e-4	32.7	6.735e1	3.2e-3	58.3	6.802e1	1.3e-2	32.1	6.719e1	9.7e-4
15	8.8	4.097e1	3.4e-4	26.8	4.100e1	1.0e-3	44.6	4.096e1	1.0e-4	18.4	4.099e1	7.5e-4
20	13.0	2.842e1	7.3e-4	18.8	2.844e1	1.4e-3	40.4	2.842e1	8.9e-4	16.6	2.843e1	1.1e-3
25	34.9	2.149e1	1.2e-3	18.0	2.152e1	2.6e-3	26.6	2.149e1	1.2e-3	16.4	2.151e1	2.2e-3
30	33.7	1.693e1	4.3e-4	17.3	1.695e1	1.7e-3	23.0	1.694e1	7.8e-4	14.5	1.694e1	1.2e-3
35	71.8	1.379e1	1.3e-3	18.1	1.381e1	2.6e-3	19.7	1.378e1	7.1e-4	11.9	1.379e1	1.6e-3
40	50.0	1.151e1	1.5e-3	12.5	1.152e1	2.1e-3	34.7	1.145e1	3.2e-4	7.7	1.151e1	1.6e-3
45	43.3	9.733e0	9.6e-4	10.6	9.736e0	1.3e-3	23.1	9.733e0	9.2e-4	6.3	9.733e0	1.0e-3
50	44.5	8.318e0	4.1e-4	10.7	8.319e0	4.8e-4	19.7	8.315e0	5.1e-6	5.7	8.318e0	4.5e-4
60	66.5	6.214e0	8.1e-4	10.9	6.214e0	7.4e-4	6.1	6.209e0	1.4e-13	6.9	6.213e0	5.9e-4
70	91.2	4.733e0	1.1e-3	11.0	4.731e0	8.2e-4	23.1	4.728e0	1.9e-4	4.6	4.731e0	7.2e-4
80	93.0	3.663e0	8.7e-4	2.2	3.800e0	3.8e-2	5.2	3.660e0	4.0e-13	2.9	3.662e0	4.5e-4
90	125.0	2.865e0	1.2e-3	2.0	2.962e0	3.5e-2	5.0	2.862e0	5.1e-13	3.0	2.864e0	7.0e-4
100	150.0	2.255e0	1.4e-3	1.7	2.323e0	3.2e-2	15.1	2.254e0	7.8e-4	2.9	2.254e0	8.3e-4
125	288.6	1.269e0	2.4e-3	1.4	1.304e0	3.0e-2	17.1	1.266e0	1.6e-4	2.7	1.268e0	1.4e-3

Table 2: Numerical results for Example 2

	Example 3				Example 4			
	Majorw		PenCorr		Majorw		PenCorr	
rank	*time*	*residue*	*time*	*residue*	*time*	*residue*	*time*	*residue*
2	8.8	1.805e2	81.2	1.804e2	2.9	3.274e2	141.6	3.277e2
5	27.0	8.984e1	70.0	8.986e1	34.4	1.523e2	245.0	1.522e2
10	38.7	4.382e1	48.7	4.383e1	48.5	7.423e1	98.7	7.428e1
15	55.5	2.616e1	43.7	2.618e1	70.5	4.442e1	79.9	4.446e1
20	84.4	1.751e1	39.1	1.753e1	101.4	2.985e1	67.0	2.987e1
25	117.0	1.265e1	38.2	1.266e1	289.6	2.197e1	69.8	2.204e1
30	171.8	9.657e0	36.5	9.657e0	335.6	1.694e1	65.8	1.699e1
35	250.6	7.639e0	39.8	7.632e0	436.7	1.345e1	71.0	1.343e1
40	324.7	6.213e0	38.8	6.203e0	470.7	1.098e1	50.5	1.098e1
45	408.4	5.169e0	38.4	5.148e0	498.7	9.104e0	47.7	9.094e0
50	502.2	4.391e0	37.5	4.355e0	639.5	7.625e0	48.0	7.623e0
60	654.1	3.290e0	35.6	3.219e0	837.6	5.552e0	44.0	5.523e0
70	972.5	2.579e0	38.2	2.481e0	987.5	4.135e0	44.9	4.084e0
80	1274.9	2.090e0	42.6	1.959e0	1212.0	3.127e0	38.0	3.082e0
90	1526.9	1.740e0	44.0	1.588e0	1417.0	2.393e0	35.6	2.345e0
100	1713.7	1.478e0	40.9	1.310e0	1612.0	1.865e0	32.7	1.814e0
125	2438.1	1.052e0	44.6	8.591e-1	1873.0	1.030e0	27.7	9.748e-1

Table 3: Numerical results for Example 3 and 4

A general example

Example 5	PenCorr	
rank	*time*	*residue*
20	11640.0	1.872e2
50	1570.0	1.011e2
100	899.0	8.068e1
250	318.3	7.574e1
500	326.3	7.574e1

Table 4: Numerical results for Example 5

Final remarks

• A code named PenCorr.m can efficiently solve all sorts of rank constrained correlation matrix problems. Faster when rank is larger.

• The techniques may be used to solve other problems, e.g., low rank matrix problems with sparsity.

• The limitation is that it cannot solve problems for matrices exceeding the dimension $4,000$ by $4,000$ on a PC due to memory constraints.

• The techniques are applicable to general rank constrained matrix (including nonsymmetric matrices) optimization problems.

End of talk

Thank you! :)