A Sequential Convex Programming Approach for Rank Constrained Matrix Optimization Problems

Defeng Sun

Department of Mathematics and Risk Management Institute
National University of Singapore

December 16, 2011

Rank Feasibility Problem

Let $\mathcal{A}$ be a linear operator from $\Re^{m\times n}$ (assuming $m \le n$) to $\Re^p$, $b \in \Re^p$ be a vector and $\mathcal{Q} \subseteq \Re^p$ be a closed convex cone.

One may ask the following question:

Does there exist a matrix $X \in \Omega := \{X \in \Re^{m\times n} \,|\, \mathcal{A}(X) \in b + \mathcal{Q}\}$ such that

$$X \in \Omega \cap M_r?$$

Here

$$M_r := \{X \in \Re^{m\times n} \,|\, \mathrm{rank}(X) \le r\}?$$

We use $\sigma_1(X) \geq \sigma_2(X) \geq \ldots \geq \sigma_m(X)$ to denote all (nonnegative) singular values of X. Define

$$s_k(X) := \sum_{j=1}^{k} \sigma_j(X) - \texttt{a convex function.}$$

The function $s_k(\cdot)$ is called Ky Fan's k-norm and the nuclear norm of X is $\|X\|_* := s_m(X)$.

One can solve the following matrix rank optimization problem

$$\begin{array}{ll} \min & \mathrm{rank}(X) \\ \text{s.t.} & X \in \Omega \end{array}$$

to give an answer to the above question.

Since the rank function is even not continuous, the above problem is hard to solve. One is tempted to consider its convex relaxation:

$$\begin{array}{ll} \min & \|X\|_* \\ \text{s.t.} & X \in \Omega. \end{array}$$

The convex relaxation approach is promising (even theoretical justifications do exist for special cases). But what shall we do if it fails to obtain a meaningful solution?

Our view is that for a nonconvex problem, we should not expect too much from a single convex relaxation. One has to go far beyond.

Fact : $\quad \mathrm{rank}(X) \leq r \Longleftrightarrow \sigma_{r+1}(X) = 0\,.$

Then the above Rank Feasibility Problem is completely equivalent to checking if there is a global solution whose optimal value is zero for

$$\begin{array}{ll} \min & p_r(x) := \sigma_{r+1}(X) + \ldots + \sigma_m(X) \qquad [\text{or } \sigma_{r+1}(X)] \\ \text{s.t.} & X \in \Omega; \end{array}$$

or

$$\begin{array}{ll} \min & \frac{1}{2}\|\mathcal{A}(X) - b - \Pi_{\mathcal{Q}}(\mathcal{A}X - b)\|^2 + p_r(X) \\ \text{s.t.} & X \in \Re^{m\times n}. \end{array}$$

Still not convex. What's the point?

Both problems are still nonconvex. But the function $p_r(\cdot)$ is now globally Lipschitz continuous and strongly semismooth.

Similarly, one can consider finding an $X \in \Omega$ of the smallest rank via

- binary search; or
- using surrogate functions, e.g.,

$$\sum_{i=1}^{m} \frac{\sigma_i(X)}{\sigma_i(X)^2 + \delta}, \quad \delta \downarrow 0.$$

Rank Optimization Problem

In applications, we not only seek a feasible point with a prescribed rank, but also want to find a best one, measured by some merit function $f : \Re^{m\times n} \to \Re$. That is, we consider the Rank Optimization Problem

$$\begin{aligned} \min \quad & f(X) \\ \text{s.t.} \quad & \mathcal{A}(X) \in b + \mathcal{Q}\,, \\ & \mathrm{rank}(X) \le r. \end{aligned}$$

Here f can be a smooth function (not necessarily convex) plus a (simple) convex function. For simplicity in discussions, let us assume that f is continuously differentiable.

One situation can be stated as follows: we may have an initial point X^0 which is not feasible. Then we may set $f(X) := \|X - X^0\|^2/2$ in order to project X^0 onto the set $\Omega \cap M_r$.

To get an optimal low rank 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]:

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

admits an analytic solution for a given $Z \in \Re^{m \times n}$:

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

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

$$Z = U[\operatorname{diag}(\sigma(Z)) \; 0]V^T, \qquad \sigma_1(Z) \geq \sigma_2(Z) \geq \ldots \geq \sigma_m(Z) \geq 0 \,.$$

A model from quantitative finance

For an example, let us focus on the following rank constrained correlation matrix problem

$$
\begin{array}{lll}
\min & \frac{1}{2}\|H \circ (X - G)\|_F^2 & \\
\text{s.t.} & X_{ii} = 1,\ i = 1, \ldots, n\,, & \texttt{(all diagonals are ones)} \\
& X_{ij} = e_{ij},\ (i,j) \in \mathcal{B}_e\,, & \texttt{(fixed off diagonals)} \\
& X_{ij} \geq l_{ij},\ (i,j) \in \mathcal{B}_l\,, & \texttt{(lower bounds)} \\
& X_{ij} \leq u_{ij},\ (i,j) \in \mathcal{B}_u\,, & \texttt{(upper bounds)} \\
& X \in \mathcal{S}_+^n\,, & \texttt{(covariance matrix)} \\
& \mathrm{rank}(X) \leq r\,, & \texttt{(rank)}
\end{array}
\tag{1}
$$

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 $(i,j) \in \mathcal{B}_l \cap \mathcal{B}_u$.

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 $\mathcal{S}^n$.

$H \geq 0$ 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 $X_{ii} = 1$, $i = 1, \ldots, n$.

Simple ones

One simpler model:

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

The simplest model:

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

In finance and statistics, correlation matrices are in many situations found to be inconsistent, i.e., $X \not\succeq 0$.

These include, but are not limited to,

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

Drop the rank constraint

Let us first consider the problem without the rank constraint:

$$\begin{array}{ll} \min & \frac{1}{2}\|H \circ (X - G)\|_F^2 \\ \text{s.t.} & X_{ii} = 1\,,\ i = 1, \ldots, n \\ & X \succeq 0\,. \end{array} \tag{4}$$

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

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

which is known as the nearest correlation matrix (NCM) problem, a terminology coined by Nick Higham (02).

The best approximation problem

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 finite dimensional real Hilbert space equipped with a scalar product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$, c is a given vector in $\mathcal{X}$, $\mathcal{A} : \mathcal{X} \to \Re^m$ is a 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 dual formulation

The dual problem is

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

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

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

Then $\nabla\theta(y) = \mathcal{A}(\Pi_K(c+\mathcal{A}^*y)) - b$ and the first order optimality condition for the dual is:

$$F(y) := y - \Pi_{Q^*}(y - \nabla\theta(y)) = 0\,.$$

Now, we only need to solve a system of nonlinear equations

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

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

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

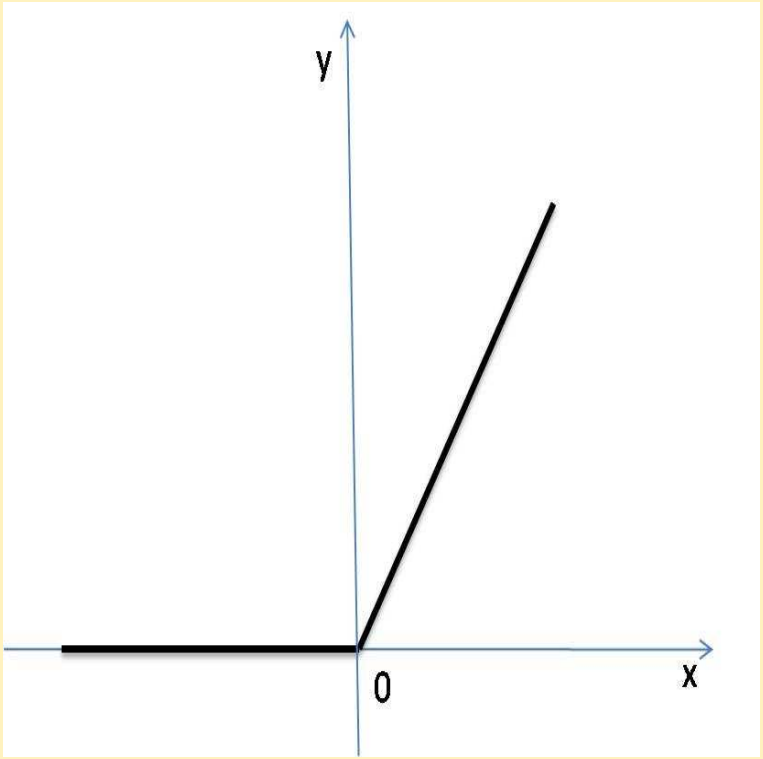
Note that

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

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

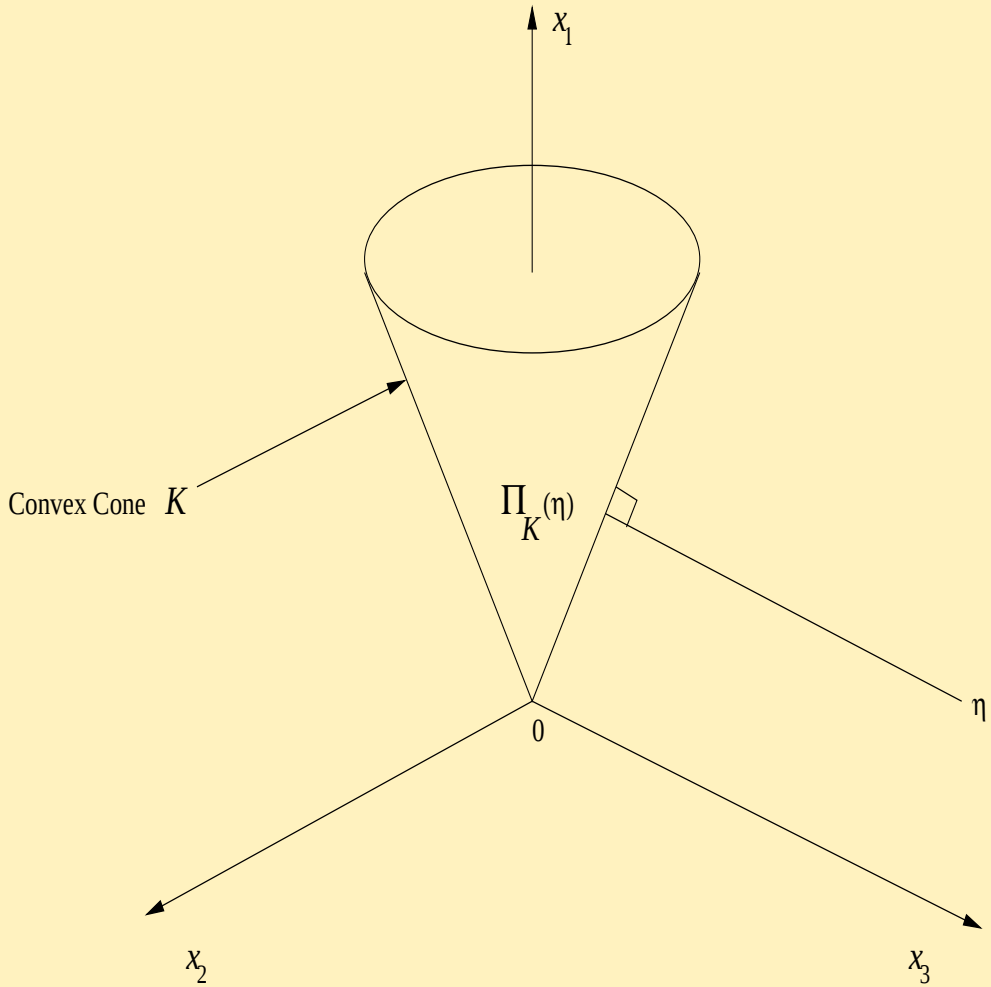
but is not twice continuously differentiable.

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[1]

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

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

Then

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

[1] Use the divide and conquer algorithm, which is much faster than the shifted QR decomposition based algorithm.]

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 ($n > 1$), but strongly semismooth [Sun[2], 02].

A locally Lipschitz function $G : \Re^n \to \Re^m$ is said to be strongly semismooth at x if the directional derivative of G at x along h, $G'(x;h)$, exists; and

$$G(x+h) - G(x) - \partial G(x+h)h = O(\|h\|^2)\,.$$

Choice of algorithms

- Interior point methods (for $n \leq 100$ only);
- First order methods are too costly (too many iterations);
- Quasi-Newton methods (okay, but can be slow as F is not smooth);
- A quadratically convergent semismooth Newton-CG method [Qi & S., 06].
  - It needs 1 to 10 steps for both random and market data;
  - For a typical example with $n = 2000$, one needs less than 50 secs on my PC running MatLab to get very high accuracy.

Inequality constraints

People from the financial institutions started to use it immediately.
NAG `http://www.nag.com/` includes it in their financial products now.

Two key points: 1) F is strongly semismooth; and
2) $\partial F(X^*)$ is positive definite that makes CG method possible [roughly: non-smoothness allows non-singularity!] Smoothness = Singularity.

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 highly efficient quadratically convergent inexact smoothing Newton-BiCGStab method is available [Gao & S., 09]

Back to the rank constraint

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

equivalently,

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

The penalty approach

Given $c > 0$, we consider a penalized version (trade-off between the objective function and the rank constraint)

$$\begin{array}{ll} \min & \frac{1}{2}\|H \circ (X - G)\|_F^2 + c p_r(X) \qquad \left[p_r(X) = \sum_{j=r+1}^{n} \sigma_j(X)\right] \\ \text{s.t.} & \mathcal{A}X \in b + Q\,, \\ & X \in \mathcal{S}_+^n\,, \end{array}$$

or equivalently

$$\begin{array}{ll} \min & f_c(X) := \frac{1}{2}\|H \circ (X - G)\|_F^2 + c(\|X\|_* - s_r(X)) \\ \text{s.t.} & \mathcal{A}X \in b + Q\,, \\ & X \in \mathcal{S}_+^n\,. \end{array}$$

Ky Fan's r-norm function

Recall that Ky Fan's r-norm function $s_r(Y) = \sigma_1(Y) + \ldots, \sigma_r(Y)$ is convex. Then for any given X and $J \in \partial s_r(X)$, we have

$$s_r(Y) \geq l_X(Y) := s_r(X) + \langle J, Y - X \rangle \quad \forall\, Y \in \Re^{m \times n}.$$

Suppose that X has the SVD decomposition

$$X = U[\Sigma(X)\ 0]V^T,$$

where U and V are orthogonal matrices. The set $\partial s_r(X)$ is completely characterized [Overton & Womersley, 93]. In particular,

$$J := \sum_{j=1}^{r} u_j v_j^T \in \partial s_r(X).$$

Majorized functions

For given $X^0 = 0$, we wish to use the largest possible support function $l_{X^0}(Y)$ by choosing a good $J \in \partial S_r(X^0)$. Since

$$l_{X^0}(Y) = \langle J, Y \rangle \leq s_r(Y) \quad \forall\, Y \in \Re^{m \times n},$$

the "neutral" choice is $J = 0 \in \partial_r S(X^0)$ and $l_{X^0}(Y) \equiv 0$. Consequently, $p_r(X)$ is majorized at X^0 by

$$s_m(X) = \|X\|_* \geq p_r(X) \quad \forall\, X$$

and $f(X) + cp_r(X)$ is majorized by

$$f(X) + c\|X\|_* \geq f(X) + cp_r(X) \quad \forall\, X.$$

Then our majorized penalty approach (suppose that f is simple first) will try to solve the following nuclear norm regularized optimization problem for X^1:

$$\begin{aligned} \min \quad & f(X) + c\|X\|_* \\ \text{s.t.} \quad & \mathcal{A}(X) \in b + Q\,. \end{aligned}$$

A good starting point X^0 may be available in practice. Then we only need to solve the following problem for X^1:

$$\begin{aligned} \min \quad & f(X) + c(\|X\|_* - \langle J_0, X\rangle) \\ \text{s.t.} \quad & \mathcal{A}(X) \in b + Q\,. \end{aligned} \tag{6}$$

where $J_0 = \sum_{j=1}^{r} u_j v_j^T \in \partial s_r(X_0)$.

Instead of adding the nuclear norm term $c\|X\|_*$, we add a weighted nuclear norm term in (6). Applying von Neumann's trace inequality to

$$\langle J_0, \pm X\rangle \leq \sigma(J_0)^T\sigma(X) = s_r(X),$$

we know that for any $X \in \Re^{m\times n}$,

$$\|X\|_* + s_r(X) \geq g_{J_0}(X) := \|X\|_* - \langle J_0, X\rangle \geq p_r(X) \geq 0.$$

The added term $g_{J_0}(X)$ is actually a semi norm, but not a norm.

Thus, instead of solving a nuclear norm regularized optimization problem, we are solving a semi norm regularized optimization problem.

The symmetric case: things are slightly different.
Let $X^0 = 0$. We wish to use the largest possible support function $l_{X^0}(Y)$ by choosing a good $J \in \partial S_r(X^0)$. Since

$$l_{X^0}(Y) = \langle J, Y \rangle \leq s_r(Y) \quad \forall\, Y \in \mathcal{S}^n_+,$$

the worst choice is $J = 0 \in \partial_r S(X^0)$ and $l_{X^0}(Y) \equiv 0$. Then our majorized penalty approach will try to solve the following trace regularized optimization problem for X^1:

$$\begin{aligned} \min \quad & f(X) + c\langle I, X \rangle \\ \text{s.t.} \quad & \mathcal{A}(X) \in b + Q\,, \\ & X \in \mathcal{S}^n_+\,. \end{aligned}$$

The fact that the above choice is the worst among all possible choices can be seen clearly from the following:

$$\langle I, X \rangle = n \qquad \texttt{if X is a correlation matrix}.$$

Since X is positive semi-definite, the best choice is $J_0 = \sum_{j=1}^{r} u_i u_i^T \in \partial S_r(X^0)$ and $l_{X^0}(Y) = s_r(X_0) + \langle J_0, X - X^0 \rangle$. Consequently, $p_r(X)$ is majorized at X^0 by

$$g_{J_0}(X) := \|X\|_* - \langle J_0, X \rangle = \langle I - J_0, X \rangle \geq p_r(X) \quad \forall\, X \in \mathcal{S}_+^n\,.$$

By noting that $I - J_0$ is a positive semi-definite and its rank is $n - r$, we know that $g_{J_0}(X)$ is a singular weighted trace.

Consequently, we will solve the following weighted trace regularized optimization problem for X^1:

$$\begin{aligned} \min \quad & f(X) + c\langle W_0, X\rangle \\ \text{s.t.} \quad & \mathcal{A}(X) \in b + Q\,, \\ & X \in \mathcal{S}^n_+\,, \end{aligned} \tag{7}$$

where

$$I \succeq W_0 := I - J_0 \succeq 0\,.$$

For $X^0 \neq 0$, we can discuss it similarly. The problem (7) may still be difficult to solve. We will use a simple majorized function to approximate it.

The idea of majorization

Let us consider

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

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 $D^{1/2} = \mathrm{diag}(d_1^{0.5}, \ldots, d_n^{0.5})$.
Then f is majorized by

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

Thus, at X^k, f_c is majorized by

$$f_c(X) \leq f_c^k(X) := f^k(X) + cg_{J_k}(X)$$

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

If ∇f is Lipschitz continuous, we can always find a majorized function for f, thus for f_c.

If ∇f is not known to be Lipschitz continuous, we can combine the idea of line search to select "locally" majorized functions for f at each step.

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

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

which is a diagonal weighted least squares correlation matrix problem

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

Now, we can use the smoothing Newton BiCGStab method introduced earlier or any other method for solving convex least squares semi-definite programming for the majorized subproblems!

We have the following property:

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

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.

How do I know the quality of the obtained solution?

Let

$$\mathcal{S}_+^n(r) := \{X \in \mathcal{S}_+^n \,|\, \mathrm{rank}(X) \le r\}.$$

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 $\Pi_{\mathcal{S}_+^n(r)}(Y)$, which can be completely characterized. Furthermore, we have

$$\partial\Psi_r(Y) = Y - \mathrm{conv}(\Pi_{\mathcal{S}_+^n(r)}(Y)).$$

The Lagrangian dual

The Lagrangian dual for

$$\begin{array}{ll} \min & \frac{1}{2}\|X-G\|_F^2 \\ \text{s.t.} & \mathcal{A}(X) \in b+Q\,, \\ & X \in \mathcal{S}_+^n(r) \end{array} \tag{9}$$

takes the form:

$$\begin{array}{ll} \max & v(y) := -\left[\frac{1}{2}\|\Pi_{\mathcal{S}_+^n(r)}(G+\mathcal{A}^*y)\|^2 - \langle b, y\rangle - \frac{1}{2}\|G\|^2\right] \\ \text{s.t.} & y \in Q^*\,. \end{array}$$

Let $\bar{y} \in Q^*$ be an optimal solution to the dual problem [the existence can be guaranteed under Slater's condition for the rank-free problem]. Then

$$\texttt{optimal value for (9)} \geq v(\bar{y}).$$

That is, $v(\bar{y})$ provides a valid lower bound. If we can find a feasible X^* such that

$$\frac{1}{2}\|X^* - G\|_F^2 = v(\bar{y}),$$

then X^* must be a global optimal solution.

Note that $v(\cdot)$ is a non-smooth concave function. We may use any first known method to solve the (convex) dual problem. Our choice is Overton's BFGS code as it works well for the equality constrained problems.

Global optimality checking

Theorem. Let $\bar{y} \in Q^*$ be an optimal solution to the dual problem. If there exists a matrix

$$\overline{X} \in \Pi_{\mathcal{S}^n_+(r)}\big(G + \mathcal{A}^*\bar{y}\big)$$

such that

$$b - \mathcal{A}\overline{X} \in \mathcal{N}_{Q^*}(\bar{y}),$$

then $\overline{X}$ and $\bar{y}$ 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. In particular, if

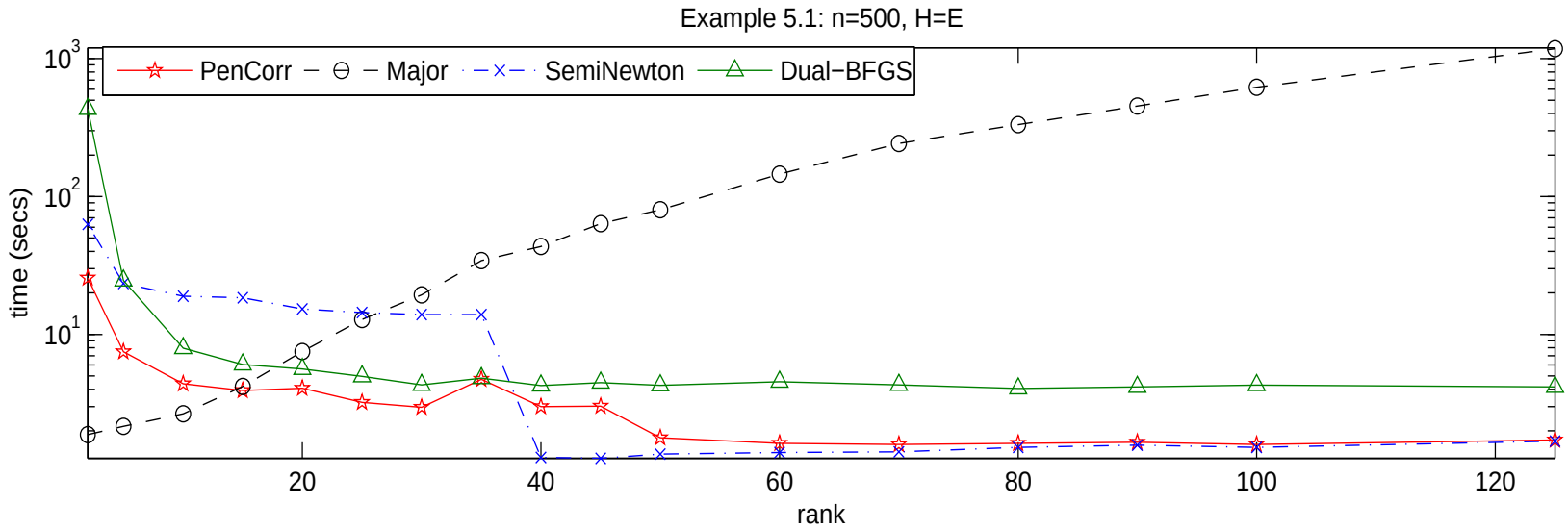
$$\lambda_r\big(G + \mathcal{A}^*\bar{y}\big) > \lambda_{r+1}\big(G + \mathcal{A}^*\bar{y}\big) > 0 \quad \text{or} \quad \lambda_{r+1}\big(G + \mathcal{A}^*\bar{y}\big) < 0\,,$$

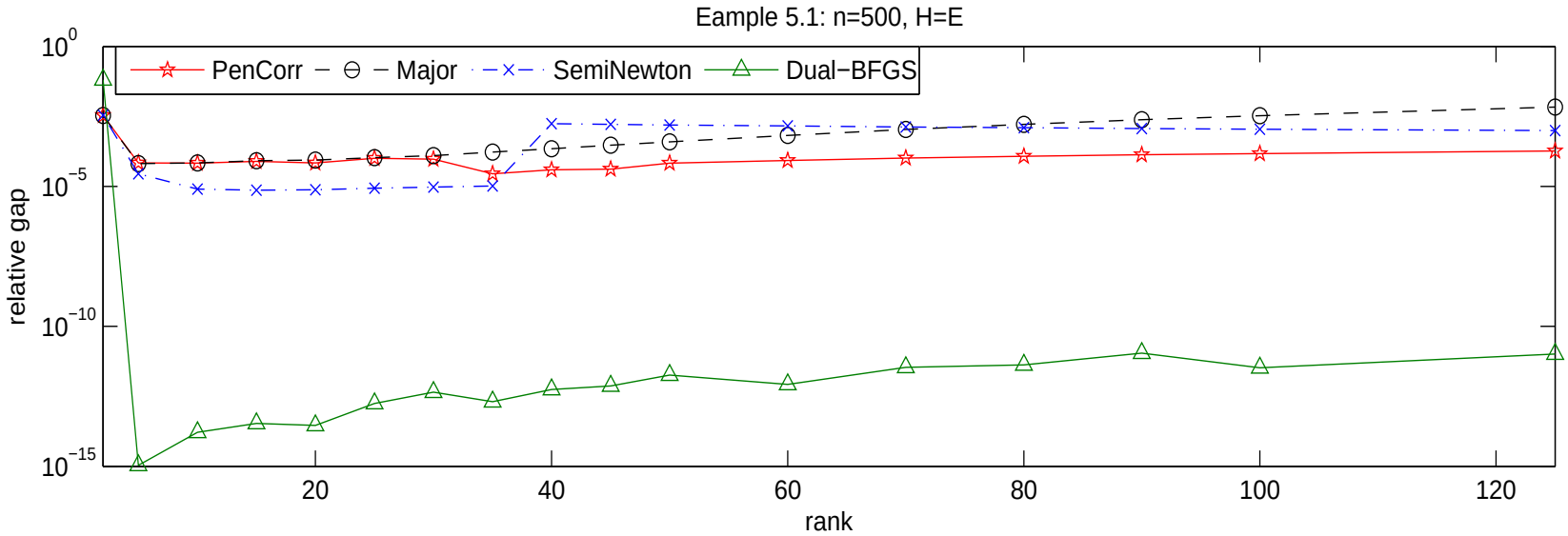
then the above statements hold automatically.

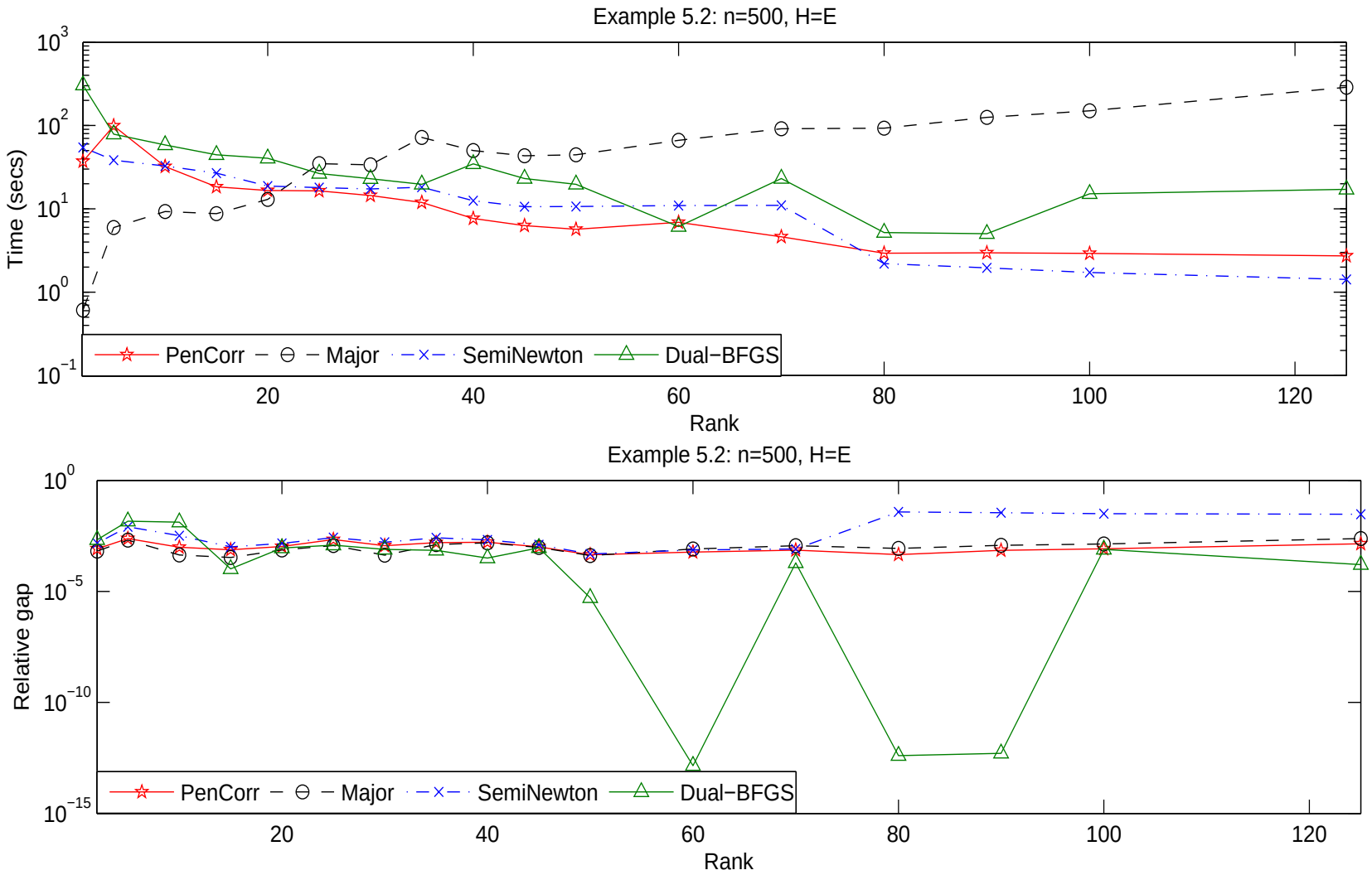
Two examples for error checking

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 5.1: n=500, H=E
PenCorr
Major
SemiNewton
Dual−BFGS
time (secs)
rank


Eample 5.1: n=500, H=E
PenCorr
Major
SemiNewton
Dual−BFGS
relative gap
rank


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

Final remarks

• All results hold in a parallel way for the non-symmetric case.

• Solve a sequence of convex least squares matrix optimization problems with weighted nuclear semi norm regularization by using tools from non-smooth and semi-smooth analysis.

• 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 and hard constraints.

Thank you! :)