Séminaire de Mathématiques Appliquées du CERMICS

École des Ponts
ParisTech

Matrix completion with deterministic pattern

Alexander Shapiro (GeorgiaTech)

19 mars 2019

Matrix completion with deterministic pattern

A. Shapiro
Joint work with Yao Xie and Rui Zhang

School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, Georgia 30332-0205, USA

École des Ponts ParisTech

Paris, March 2019

Minimum Rank Matrix Completion (MRMC) problem

$$\min_{Y \in \mathbb{R}^{n_1 \times n_2}} \text{rank}(Y) \quad \text{subject to } Y_{ij} = M_{ij}, \ (i,j) \in \Omega, \tag{1}$$

where $\Omega \subset \{1, ..., n_1\} \times \{1, ..., n_2\}$ is an index set of cardinality m and M_{ij}, $(i,j) \in \Omega$, are given values. Let M be the $n_1 \times n_2$ matrix with entries M_{ij} at $(i,j) \in \Omega$, and all other entries equal zero.

Problem (1) can be written as

$$\min_{Y \in \mathbb{V}_{\Omega^c}} \text{rank}(M + Y), \tag{2}$$

where $\Omega^c = \{(i,j) : (i,j) \notin \Omega\}$ is the complement of index set Ω and

$$\mathbb{V}_{\Omega^c} := \left\{ Y \in \mathbb{R}^{n_1 \times n_2} : Y_{ij} = 0, \ (i,j) \in \Omega \right\}.$$

This linear space represents the set of matrices that are filled with zeros at the locations of the observed entries.

The MRMC problem can be formulated in the following equivalent form

$$\min_{X \in \mathbb{W}_{\tau^c}} \text{rank}(\Xi + X) \text{ subject to } \Xi + X \succeq 0, \tag{3}$$

where $\mathbb{S}^p$ denotes the space of $p \times p$ symmetric matrices, $\Xi \in \mathbb{S}^p$, $p = n_1 + n_2$, is of the form $\Xi = \begin{bmatrix} 0 & M \\ M^\top & 0 \end{bmatrix}$, and

$$\mathbb{W}_{\tau^c} := \{X \in \mathbb{S}^p : X_{ij} = 0, \ (i,j) \in \tau\},$$

where τ is (symmetric) index set associated with Ω and τ^c is its symmetric complement. Indeed consider $Y = VW^\top$, where V and W are matrices of the respective order $n_1 \times r$ and $n_2 \times r$ and common rank r and take $X := UU^\top - \Xi$, where

$$U := \begin{bmatrix} V \\ W \end{bmatrix}, \text{ i.e., } X = \begin{bmatrix} VV^\top & Y - M \\ (Y - M)^\top & WW^\top \end{bmatrix}.$$

We can consider minimum rank problems of the form (3) with symmetric index set $\tau \subset \{1, ..., p\} \times \{1, ..., p\}$ which does not include diagonal entries, i.e., $(i, i) \notin \tau$, $i = 1, ..., p$.

Minimum Rank Factor Analysis (MRFA) problem

$$\min_{X \in \mathbb{D}^p} \text{rank}(\Sigma - X) \text{ s.t. } \Sigma - X \succeq 0,$$

where $\Sigma \in \mathbb{S}^p$ is a positive definite matrix (covariance matrix). By $\mathbb{S}^p$ we denote the space of $p \times p$ symmetric matrices and by $\mathbb{D}^p$ the space of $p \times p$ diagonal matrices.

In that case $\tau = \{(i, j) : i \neq j\}$ is the index set of off-diagonal elements and $\mathbb{D}^p = \mathbb{W}_{\tau^c}$.

Generic bounds for the minimal rank

Suppose that specified values of the minimum rank problems are observed with noise. In the MRFA the covariance matrix Σ is estimated by the sample covariance matrix S based on a sample of N observations. In the MRMC the entries M_{ij} are observed with random noise. In such settings there are the following generic lower bounds for the minimal rank.

For the MRFA we have the following lower bound (Shapiro (1982)), which holds for a.e. $\Sigma \in \mathbb{S}^p$ (i.e., for all $\Sigma \in \mathbb{S}^p$ except in a set of Lebesgue measure zero):

$$\text{rank}(\Sigma + X) \geq \frac{2p + 1 - \sqrt{8p + 1}}{2}, \ \ \forall X \in \mathbb{D}^p.$$

This is based on that the set

$$\mathcal{W}_r := \{A \in \mathbb{S}^p : \text{rank}(A) = r\}$$

of matrices of rank r forms C^∞ smooth manifold, in the linear space $\mathbb{S}^p$, of dimension

$$\dim(\mathcal{W}_r) = p(p+1)/2 - (p-r)(p-r+1)/2$$

and the transversality condition. Mapping $\mathcal{A}(X) := \Sigma - X$, from $\mathbb{D}^p$ into $\mathbb{S}^p$, intersects $\mathcal{W}_r$ transverally if for every $X \in \mathbb{D}^p$ either $\mathcal{A}(X) \notin \mathcal{W}_r$ or $\mathcal{A}(X) \in \mathcal{W}_r$ and

$$\mathbb{D}^p + T_{\mathcal{W}_r}(\mathcal{A}(X)) = \mathbb{S}^p.$$

The above generic lower bound means that

$$p + \dim(\mathcal{W}_r) \geq \dim(\mathbb{S}^p),$$

i.e., that $p \geq (p-r)(p-r+1)/2$.

For the MRMC problem we can proceed in a similar way. The set

$$\mathcal{M}_r := \{A \in \mathbb{R}^{n_1 \times n_2} : \text{rank}(A) = r\}$$

of matrices of rank $r \leq \min\{n_1, n_2\}$ forms a smooth manifold, in the linear space $\mathbb{R}^{n_1 \times n_2}$, of dimension

$$\dim(\mathcal{M}_r) = r(n_1 + n_2 - r).$$

Consider mapping $\mathcal{A}_M(Y) := M + Y$, $Y \in \mathbb{V}_{\Omega^c}$. Note that $M \in \mathbb{V}_\Omega$ and the image $\mathcal{A}_M(\mathbb{V}_{\Omega^c})$ defines the space of feasible points of the MRMC problem. The transversality condition: either $\mathcal{A}_M(Y) \notin \mathcal{M}_r$ or $\mathcal{A}_M(Y) \in \mathcal{M}_r$ and

$$\mathbb{V}_{\Omega^c} + T_{\mathcal{M}_r}(\mathcal{A}_M(Y)) = \mathbb{R}^{n_1 \times n_2}.$$

It follows that generically (i.e., for a.e. values M_{ij}) the rank r of matrix $Y \in \mathbb{R}^{n_1 \times n_2}$ such that $Y_{ij} = M_{ij}$, $(i,j) \in \Omega$, should satisfy

$$r(n_1 + n_2 - r) \geq \mathsf{m}.$$

(Recall that $\mathsf{m} = |\Omega|$.) Equivalently it holds generically (almost surely) that the minimal rank

$$r \geq \mathfrak{R}(n_1, n_2, \mathsf{m}),$$

where

$$\mathfrak{R}(n_1, n_2, \mathsf{m}) := (n_1 + n_2)/2 - \sqrt{(n_1 + n_2)^2/4 - \mathsf{m}}.$$

In particular, for $n_1 = n_2 = n$, $\mathfrak{R}(n, n, \mathsf{m}) = n - \sqrt{n^2 - \mathsf{m}}$.

It also follows that unless $\mathfrak{R}(n_1, n_2, \mathsf{m})$ is an integer, almost surely the set of optimal solutions of the MRMC problem is not a singleton.

As a heuristic it was suggested in Fazel (2002) to approximate the MRMC problem by the following problem

$$\min_{Y \in \mathbb{V}_{\Omega^c}} \|Y + M\|_*, \tag{4}$$

where $\|\cdot\|_*$ is the nuclear norm $\|Y\|_* = \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(Y)$, with $\sigma_i(Y)$ denotes the i-th largest singular value of Y.

For the nuclear norm its dual norm is given by $\sigma_1(\cdot)$ and the respective unit ball of the dual norm can be written as $\mathcal{B} = \left\{Q \in \mathbb{R}^{n_1 \times n_2} : \lambda_1(Q^\top Q) \le 1\right\}$, where $\lambda_1(Q^\top Q)$ is the largest eigenvalue of matrix $Q^\top Q$. It follows that $\|Y\|_*$ is equal to the optimal value of the following SDP problem

$$\max_{Q \in \mathbb{R}^{n_1 \times n_2}} \operatorname{tr}(Q^\top Y) \text{ s.t. } \begin{bmatrix} I_{n_1} & Q \\ Q^\top & I_{n_2} \end{bmatrix} \succeq 0.$$

The Lagrangian dual of this problem is

$$\min_{\Lambda_1 \in \mathbb{S}^{n_1}, \Lambda_2 \in \mathbb{S}^{n_2}} \operatorname{tr}(\Lambda_1 + \Lambda_2) \ \text{ s.t. } \begin{bmatrix} \Lambda_1 & -\frac{1}{2}Y \\ -\frac{1}{2}Y^\top & \Lambda_2 \end{bmatrix} \succeq 0.$$

It follows that problem (4) can be written as the SDP problem

$$\min_{X \in \mathbb{W}_{\tau^c}} \operatorname{tr}(X) \text{ subject to } \Xi + X \succeq 0. \qquad (5)$$

with $p = n_1 + n_2$, the index set τ being the symmetric index set corresponding to the index set Ω and $\Xi := \begin{bmatrix} 0 & M \\ M^\top & 0 \end{bmatrix}$ (the coefficient $-1/2$ can be absorbed into X).

Minimum Trace Factor Analysis (MTFA) problem (Bentler, 1972)

$$\min_{X \in \mathbb{D}^p} \operatorname{tr}(\Sigma - X) \text{ s.t. } \Sigma - X \succeq 0. \tag{6}$$

For any $\Sigma \in \mathbb{S}^p$ the MTFA problem (6) has unique optimal solution.

Suppose that $\bar{X} \in \mathbb{D}^p$ and $\Sigma - \bar{X} = \gamma\gamma^\top$ for some nonzero vector $\gamma = (\gamma_1, ..., \gamma_p)^\top$, i.e., $\Sigma - \bar{X}$ has rank one. Then $\bar{X}$ is an optimal solution of the MTFA problem (6) iff the following condition holds (Shapiro, 1982)

$$|\gamma_j| \le \tfrac{1}{2} \sum_{i=1}^{p} |\gamma_i|, \; j = 1, ..., p. \tag{7}$$

For a general symmetric set τ not containing diagonal, a point $\bar{X} \in \mathbb{W}_{\tau^c}$ is an optimal solution of problem

$$\min_{X \in \mathbb{W}_{\tau^c}} \mathrm{tr}(X) \text{ subject to } \Xi + X \succeq 0$$

iff $\Xi + \bar{X} \succeq 0$ and there exists $\Lambda \in \mathbb{S}^p_+$ such that $P_{\tau^c}(\Lambda) = I_p$ and the complementary condition holds

$$(\Xi + \bar{X})\Lambda = 0.$$

When the minimal rank is less then the generic lower bound, the minimum trace (minimum nuclear norm) approach "often" recovers the exact minimum rank solution.

Low-Rank Matrix Approximation

Consider the following model of the observed values

$$M_{ij} = Y^*_{ij} + N^{-1/2}\Delta_{ij} + \varepsilon_{ij},\ (i,j) \in \Omega,$$

where Y^* is $n_1 \times n_2$ matrix of rank r (i.e., $Y^* \in \mathcal{M}_r$), Δ_{ij} are some (deterministic) numbers and ε_{ij} are mutually independent random variables such that $N^{1/2}\varepsilon_{ij}$ converge in distribution to normal with mean zero and variance σ^2_{ij}, $(i,j) \in \Omega$. The additional terms $N^{-1/2}\Delta_{ij}$ represent a possible deviation of population values from the "true" model and are often referred to as the population drift or a sequence of local alternatives. It is assumed that $r \le \mathfrak{R}(n_1, n_2, \mathsf{m})$.

It is said that the model is (globally) **identified** (at Y^*) if Y^* is the unique solution of the respective matrix completion problem.

Recall that

$$\mathbb{V}_{\Omega^c} = \left\{ Y \in \mathbb{R}^{n_1 \times n_2} : Y_{ij} = 0,\ (i,j) \in \Omega \right\}$$

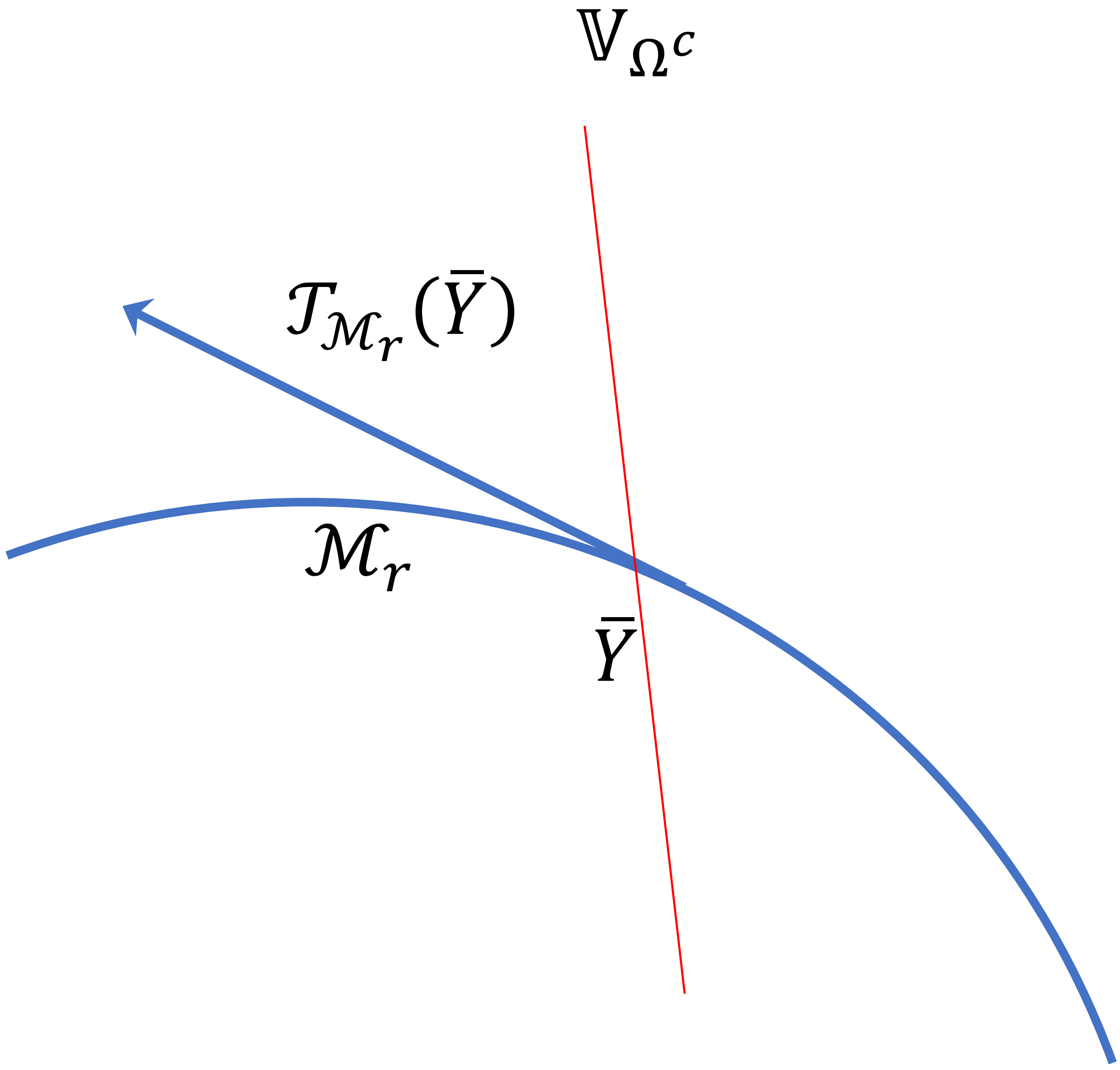
and

$$\mathbb{V}_{\Omega} = \left\{ Y \in \mathbb{R}^{n_1 \times n_2} : Y_{ij} = 0,\ (i,j) \in \Omega^c \right\}.$$

By P_Ω we denote the projection onto the space $\mathbb{V}_\Omega$, i.e., $P_\Omega(Y) = Y$ for $Y \in \mathbb{V}_\Omega$, and $P_\Omega(Y) = 0$ for $Y \in \mathbb{V}_{\Omega^c}$.

Definition 1 (Well-posedness condition) *We say that a matrix* $\bar{Y} \in \mathcal{M}_r$ *is* well-posed, *for the MRMC problem, if* $P_\Omega(\bar{Y}) = M$ *and the following condition holds*

$$\mathbb{V}_{\Omega^c} \cap T_{\mathcal{M}_r}(\bar{Y}) = \{0\}.$$


$\mathbb{V}_{\Omega^c}$
$\mathcal{T}_{\mathcal{M}_r}(\bar{Y})$
$\mathcal{M}_r$
$\bar{Y}$

Theorem 1 (Sufficient conditions for local uniqueness) *If $\bar{Y} \in \mathcal{M}_r$ is well-posed, then matrix $\bar{Y} \in \mathcal{M}_r$ is a locally unique solution of the MRMC problem, i.e., there is a neighborhood $\mathcal{V}$ of $\bar{Y}$ such that* $\text{rank}(Y) > \text{rank}(\bar{Y})$ *for any $Y \in \mathcal{V}$, $Y \neq \bar{Y}$.*

Algebraic condition for well-posedness. The tangent space to $\mathcal{M}_r$ at $Y \in \mathcal{M}_r$ can be written as

$$T_{\mathcal{M}_r}(Y) = \left\{ H \in \mathbb{R}^{n_1 \times n_2} : FHG = 0 \right\},$$

where F is an $(n_1 - r) \times n_1$ matrix of rank $n_1 - r$ such that $FY = 0$ (referred to as a *left side complement* of Y) and G is an $n_2 \times (n_2 - r)$ matrix of rank $n_2 - r$ such that $YG = 0$ (referred to as a *right side complement* of Y.

Matrix $\bar{Y} \in \mathcal{M}_r$ is well-posed if and only if for any left side complement F and right side complement G of $\bar{Y}$, the column vectors $g_j^\top \otimes f_i$, $(i,j) \in \Omega^c$, are linearly independent.

In a certain sense the well-posedness condition is generic. Denote by $\mathcal{F}_r \subset \mathbb{R}^{n_1 \times r}$ and $\mathcal{X}_r \subset \mathbb{R}^{n_2 \times r}$ the respective sets of matrices of rank r. Consider the set $\Theta := \mathcal{F}_r \times \mathcal{X}_r \times \mathbb{V}_{\Omega^c}$ viewed as a subset of $\mathbb{R}^{n_1 r + n_2 r + n_1 n_2 - \mathsf{m}}$, and mapping $\mathfrak{F} : \Theta \to \mathbb{R}^{n_1 \times n_2}$ defined as

$$\mathfrak{F}(\theta) := VW^\top + X, \;\; \theta = (V, W, X) \in \Theta.$$

Let $\Delta(\theta)$ be the Jacobian matrix of mapping $\mathfrak{F}$. We refer to

$$\varrho := \max_{\theta \in \Theta} \Big\{ \mathrm{rank}\Big(\Delta(\theta)\Big) \Big\} \tag{8}$$

as the *characteristic rank* of mapping $\mathfrak{F}$ and say that $\theta \in \Theta$ is a *regular point* of $\mathfrak{F}$ if $\mathrm{rank}\Big(\Delta(\theta)\Big) = \varrho$. We say that $(V, W) \in \mathcal{F}_r \times \mathcal{X}_r$ is regular if $\theta = (V, W, X)$ is regular for some $X \in \mathbb{V}_{\Omega^c}$.

Consider

$$\mathfrak{f}(r, m) := \dim(\mathcal{M}_r) + \dim(\mathbb{V}_{\Omega^c}) = r(n_1 + n_2 - r) + n_1 n_2 - m.$$

We have that $\varrho \leq \mathfrak{f}(r, m)$.

Theorem 2 *The following holds. (i) Almost every point* $(V, W) \in \mathcal{F}_r \times \mathcal{X}_r$ *is regular. (ii) The set of regular points forms an open subset of* $\mathcal{F}_r \times \mathcal{X}_r$*. (iii) For any regular point* $(V, W) \in \mathcal{F}_r \times \mathcal{X}_r$*, the corresponding matrix* $Y = VW^\top$ *satisfies the well-posedness condition if and only if the characteristic rank* ϱ *is equal to* $\mathfrak{f}(r, m)$*. (iv) If* $\varrho < \mathfrak{f}(r, m)$ *and a point* $(\bar{V}, \bar{W}) \in \mathcal{F}_r \times \mathcal{X}_r$ *is regular, then for any* $Y \in \mathcal{M}_r$ *in a neighborhood of* $\bar{Y} = \bar{V}\bar{W}^\top$ *there exists* $X \in \mathbb{V}_{\Omega^c}$ *such that* $Y = \bar{Y} + X$*.*

Consider the weighted least squares problem

$$\min_{Y \in \mathcal{M}_r} \sum_{(i,j) \in \Omega} w_{ij} \left(M_{ij} - Y_{ij} \right)^2,$$

where $w_{ij} := 1/\hat{\sigma}^2_{ij}$ with $\hat{\sigma}^2_{ij}$ being consistent estimates of σ^2_{ij} (i.e., $\hat{\sigma}^2_{ij}$ converge in probability to σ^2_{ij} as $N \to \infty$).

The model

$$M_{ij} = Y^*_{ij} + N^{-1/2} \Delta_{ij} + \varepsilon_{ij},\ (i, j) \in \Omega,$$

is tested by the following statistic

$$T_N(r) := N \min_{Y \in \mathcal{M}_r} \sum_{(i,j) \in \Omega} w_{ij} \left(M_{ij} - Y_{ij} \right)^2.$$

A sufficient condition for a feasible $\bar{Y}$ to be a strict local solution of the MRMC problem is

$$\mathrm{tr}\left[P_\Omega(H)^\top P_\Omega(H)\right] > 0, \ \ \forall H \in T_{\mathcal{M}_r}(\bar{Y}) \setminus \{0\}.$$

This condition is equivalent to well posedness of $\bar{Y}$.

Theorem 3 (Asymptotic properties of test statistic) *Suppose that the model is globally identified at $Y^* \in \mathcal{M}_r$ and Y^* is well-posed. Then the test statistic $T_N(r)$ converges in distribution to noncentral chi square with degrees of freedom* $\mathrm{df}_r = \mathrm{m} - r(n_1 + n_2 - r)$ *and the noncentrality parameter*

$$\delta_r = \min_{H \in T_{\mathcal{M}_r}(Y^*)} \sum_{(i,j)\in\Omega} \sigma_{ij}^{-2} \left(\Delta_{ij} - H_{ij}\right)^2.$$

In Factor Analysis this type of results is going back to Steiger, Shapiro, and Browne (1985).

The noncentrality parameter can be approximated as

$$\delta_r \approx N \min_{Y \in \mathcal{M}_r} \sum_{(i,j)\in\Omega} w_{ij} \left(Y_{ij}^* + N^{-1/2}\Delta_{ij} - Y_{ij} \right)^2 .$$

That is, the noncentrality parameter is approximately equal to N times the fit to the "true" model of the alternative population values $Y_{ij}^* + N^{-1/2}\Delta_{ij}$ under small perturbations of order $O(N^{-1/2})$.

The asymptotics of the test statistic $T_N(r)$ depends on r and also on the cardinality m of the index set Ω. Suppose now that more observations become available at additional entries of the matrix. That is we are testing now the model with a larger index set Ω', of cardinality m', such that $\Omega \subset \Omega'$. In order to emphasize that the test statistic also depends on the corresponding index set we add the index set in the respective notations.

Theorem 4 *Consider index sets* $\Omega \subset \Omega'$ *of cardinality* $\mathsf{m} = |\Omega|$ *and* $\mathsf{m}' = |\Omega'|$. *Suppose that the model is globally identified at* $Y^* \in \mathcal{M}_r$ *and the well posedness condition holds at* Y^* *for the smaller model (and hence for both models). Then the statistic* $T_N(r, \Omega') - T_N(r, \Omega)$ *converges in distribution to noncentral chi-square with* $\mathsf{df}_{r,\Omega'} - \mathsf{df}_{r,\Omega} = \mathsf{m}' - \mathsf{m}$ *degrees of freedom and the noncentrality parameter* $\delta_{r,\Omega'} - \delta_{r,\Omega}$, *and* $T_N(r, \Omega') - T_N(r, \Omega)$ *is asymptotically independent of* $T_N(r, \Omega)$.

The (weighted) least squares problem is not convex and it may happen that an optimization algorithm converges to a stationary point which is not a globally optimal solution. Suppose that we run a numerical procedure which identifies a matrix $\bar{Y} \in \mathcal{M}_r$ satisfying the (necessary) first order optimality conditions, i.e., the algorithm converged to a stationary point. Then if $P_\Omega(\bar{Y})$ is sufficiently close to M (i.e., the fit $\sum_{(i,j)\in\Omega} \left(Y_{ij} - M_{ij}\right)^2$ is sufficiently small) and the well posedness condition holds at $\bar{Y}$, then $\bar{Y}$ solves the least squares problem at least locally. Unfortunately it is not clear how to quantify the "sufficiently close" condition.

Uniqueness of the minimal rank solutions

Consider the Minimum Rank Factor Analysis problem. Suppose that it has solution

$$\Sigma = \Lambda\Lambda^\top + X$$

where Λ is $p \times r$ matrix of rank r and $X \in \mathbb{D}^p$ is diagonal matrix. This solution is unique if for any $i \in \{1, ..., p\}$ the matrix Λ has two $r \times r$ disjoint nonsingular submatrices not containing i-th row (this result is going back at least to Anderson and Rubin, 1956). That if $r < p/2$ and every $r \times r$ submatrix of Λ is nonsingular, then the solution is unique.

Example 1 (Wilson and Worcester. 1939) Consider

$$M = \begin{pmatrix} 0 & 0.56 & 0.16 & 0.48 & 0.24 & 0.64 \\ 0.56 & 0 & 0.20 & 0.66 & 0.51 & 0.86 \\ 0.16 & 0.20 & 0 & 0.18 & 0.07 & 0.23 \\ 0.48 & 0.66 & 0.18 & 0 & 0.3 & 0.72 \\ 0.24 & 0.51 & 0.07 & 0.30 & 0 & 0.41 \\ 0.64 & 0.86 & 0.23 & 0.72 & 0.41 & 0 \end{pmatrix}$$

The two rank 3 solutions are given by diagonal matrices:

$$D_1 = \text{Diag}(0.64, 0.85, 0.06, 0.56, 0.50, 0.93),$$

$$D_2 = \text{Diag}(0.425616, 0.902308, 0.063469, 0.546923, 0.386667, 0.998).$$

For the MRFA the corresponding generic lower bound for $p = 6$ is $r \geq 3$, and for the MRMC problem the corresponding generic lower bound for $n_1 = n_2 = 6$ is $r \geq 4$.

Uniqueness of the minimum rank solution is invariant with respect to permutations of rows and columns of matrix M. This motivates to introduce the following definition.

Definition 2 *We say that the index set* Ω *is* reducible *if by permutations of rows and columns, matrix* M *can be represented in a block diagonal form, i.e.,* $M = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}$. *Otherwise we say that* Ω *is irreducible.*

Theorem 5 *The following holds.*
(i) *If the index set* Ω *is reducible, then any minimum rank solution* $\bar{Y}$, *such that* $\bar{Y}_{ij} \neq 0$ *for all* $(i,j) \in \Omega^c$, *is not locally (and hence globally) unique.*

(ii) *Suppose that* Ω *is irreducible,* $M_{ij} \neq 0$ *for all* $(i,j) \in \Omega$, *and every row and every column of the matrix* M *have at least one element* M_{ij}. *Then any rank one solution is globally unique.*

Statistical inference of SDP problems

Consider the following Semidefinite Programming (SDP) problem

$$\min_{x\in\mathbb{R}^n} c^\top x \text{ subject to } \Sigma + \mathcal{A}(x) \succeq 0, \tag{9}$$

where $\Sigma \in \mathbb{S}^p$ and $A_i \in \mathbb{S}^p$ are given matrices and $\mathcal{A} : \mathbb{R}^n \to \mathbb{S}^p$ is the linear mapping $\mathcal{A}(x) := \sum_{i=1}^n x_i A_i$.

The (Lagrangian) dual of problem (9) is the problem

$$\max_{\Lambda\in\mathbb{S}^p_+} -\text{tr}(\Lambda\Sigma) \ \text{ s.t. } \text{tr}(\Lambda A_i) = c_i,\ i = 1, ..., n. \tag{10}$$

Let x^* be optimal solutions of the primal problem (9) and Λ be an optimal solution of the dual problem (10). Then

$$(\Sigma + \mathcal{A}(x^*))\Lambda = 0.$$

It is said that the strict complementarity condition holds at Λ if

$$\text{rank}(\Sigma + \mathcal{A}(x^*)) + \text{rank}(\Lambda) = p.$$

Suppose that Σ is estimated (approximated) by a matrix $S \in \mathbb{S}^p$ and hence the SDP problem (9) is approximated by

$$\min_{x \in \mathbb{R}^n} c^\top x \text{ subject to } S + \mathcal{A}(x) \succeq 0. \tag{11}$$

What can be said about statistical properties of the optimal value and optimal solutions of the SDP problem (11) considered as estimates of their counterparts of problem (9).

Consider $p^2 \times 1$ vectors $s := \text{vec}(S)$ and $\sigma := \text{vec}(\Sigma)$ formed from columns of the respective matrices stacked columnwise. We assume that $N^{1/2}(s-\sigma)$ converges in distribution to multivariate normal with mean vector zero and $p^2 \times p^2$ covariance matrix Γ as the sample size N tends to ∞. This can be justified by the Central Limit Theorem. In particular if S is the sample covariance matrix of normally distributed population, then

$$\Gamma_{ij,k\ell} = \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk},$$

where $\sigma_{ij} = \Sigma_{ij}$.

Statistical inference is based on perturbation (sensitivity) analysis of SDP programs and the so-called Delta Theorem.

It is said that $x^* \in \mathbb{R}^n$ is a nondegenerate point of mapping $x \mapsto \Sigma + \mathcal{A}(x)$ if for $\Upsilon := \Sigma + \mathcal{A}(x^*)$ and $r := \text{rank}(\Upsilon)$ it follows that

$$\mathcal{A}(\mathbb{R}^n) + T_{\mathcal{W}_r}(\Upsilon) = \mathbb{S}^p, \tag{12}$$

otherwise point x^* is said to be degenerate. That is, x^* is nondegenerate if mapping $x \mapsto \Sigma + \mathcal{A}(x)$ intersects the smooth manifold $\mathcal{W}_r$ transversally at $\Sigma + \mathcal{A}(x^*) \in \mathcal{W}_r$.

If x^* is nondegenerate, then

$$n \geq (p - r)(p - r + 1)/2.$$

Denote by $\vartheta(\Sigma)$ the optimal value of the the SDP problem (9) considered as a function of matrix Σ, by $\mathsf{Sol}(P)$ the set of optimal solutions of the SDP problem (9) and by $\mathsf{Sol}(D)$ the set of optimal solutions of its dual problem.

Proposition 1 *Suppose that Slater condition holds for the reference problem* (9) *and its optimal value* $\vartheta(\Sigma)$ *is finite. Then the set* $\mathsf{Sol}(D)$ *is nonempty, convex and compact and the optimal value function* $\vartheta(\cdot)$ *is continuous and Fréchet directionally differentiable at* Σ *with the directional derivative*

$$\vartheta'(\Sigma, H) = \sup_{\Lambda \in \mathsf{Sol}(D)} \mathrm{tr}(\Lambda H). \tag{13}$$

Let $\widehat{\vartheta}_N$ be the optimal value of the approximate SDP problem based on a sample of size N.

Theorem 6 *Suppose that the optimal value $\vartheta^* = \vartheta(\Sigma)$ is finite and Slater condition for the true problem holds. Then*

$$N^{1/2}(\widehat{\vartheta}_N - \vartheta^*) \Rightarrow \sup_{\lambda \in \mathrm{Sol}(D)} \lambda^\top Z,$$

where Z is a random vector having multivariate normal distribution $\mathcal{N}(0, \Gamma)$.

Moreover, if $\mathrm{Sol}(D) = \{\Lambda\}$ is a singleton, then $N^{1/2}(\widehat{\vartheta}_N - \vartheta^*)$ converges in distribution to normal with zero mean and variance $\sigma^2 = \lambda^\top \Gamma \lambda$, where $\lambda := \mathrm{vec}(\Lambda)$. In particular, if the population is normally distributed, then the asymptotic variance $\sigma^2 = 2\mathrm{tr}(\Sigma\Lambda\Sigma\Lambda)$.

It follows, under mild regularity conditions, that if the set Sol(D) is not a singleton, then the bias

$$\mathbb{E}[\hat{\vartheta}_N - \vartheta^*] = N^{-1/2}\mathbb{E}\left[\sup_{\lambda \in \text{Sol}(D)} \lambda^\top Z\right] + o(N^{-1/2}),$$

with $Z \sim \mathcal{N}(0, \Gamma)$. That is, the bias of $c^\top \hat{x}_N$ considered as an estimator of $c^\top x^*$ is of order $O(N^{-1/2})$.

It is considerably more difficult to derive asymptotic distribution of optimal solutions of the approximate SDP problem. For that we will need considerably more restrictive assumptions.

We need to verify that the following so-called quadratic growth condition holds at an optimal solution $x^* \in \mathrm{Sol}(P)$:
there is $\kappa > 0$ such that

$$c^\top x \geq c^\top x^* + \kappa \|x - x^*\|^2$$

for all feasible points of the problem (9) in a neighborhood of the point x^*.

Proposition 2 *Suppose that* $\mathrm{Sol}(P) = \{x^*\}$ *is a singleton and the strict complementarity condition holds at some* $\Lambda \in \mathrm{Sol}(D)$*. Then the quadratic growth condition follows.*

We discuss now differentiability of an optimal solution $\bar{x}(\sigma)$ of the SDP problem (9) considered as a function of $\sigma = \text{vec}(\Sigma)$.

Suppose that $\text{Sol}(P) = \{x^*\}$ is a singleton and that x^* is a non-degenerate point of $\Sigma + \mathcal{A}(\cdot)$, and hence $\text{Sol}(D) = \{\Lambda\}$ is a singleton. Suppose also that the strict complementarity condition holds. Let $\Upsilon := \Sigma + \mathcal{A}(x^*)$ and $\Lambda = E\Theta E^\top$ be the spectral decomposition of matrix Λ. Recall that $\Upsilon\Lambda = 0$ (complementarity condition), $\Upsilon \succeq 0$, $\Lambda \succeq 0$, and because of the strict complementarity assumption E is $p \times (p - r)$ matrix of rank $p - r$ where $r = \text{rank}(\Upsilon)$.

Consider the following optimization problem, depending on $\Delta \in \mathbb{S}^p$,

$$\begin{array}{ll} \min\limits_{h \in \mathbb{R}^n} & \operatorname{tr}\left[\Lambda(\mathcal{A}(h)+\Delta)\Upsilon^\dagger(\mathcal{A}(h)+\Delta)\right] \\ \text{s.t.} & E^\top \mathcal{A}(h) E + E^\top \Delta E = 0, \end{array} \tag{14}$$

where $\Upsilon^\dagger$ is the Moore- Penrose pseudoinverse of matrix Υ.

This is a problem of minimization of quadratic function subject to linear constraints. Under the above assumptions, problem (14) has a unique optimal solution $\bar{h}(\delta)$, which is a linear function of $\delta = \operatorname{vec}(\Delta)$. That is $\bar{h}(\delta) = J^\top \delta$, where J is the corresponding $p^2 \times n$ matrix.

Theorem 7 *Suppose that* Sol$(P) = \{x^*\}$ *is a singleton, and that* x^* *is a nondegenerate point of* $\Sigma + \mathcal{A}(\cdot)$ *and the strict complementarity condition holds. Then* $\bar{x}(\cdot)$ *is differentiable at* $\sigma = \text{vec}(\Sigma)$ *and*

$$\bar{x}(s) = \bar{x}(\sigma) + J^\top (s - \sigma) + o(\|s - \sigma\|), \qquad (15)$$

where $J^\top \delta$ *is the optimal solution of problem* (14) *with* Λ *being the optimal solution of the dual problem and* E *being a matrix whose columns are orthonormal and generate the null space of the matrix* $\Sigma + \mathcal{A}(x^*)$.

That is, J is the Jacobian matrix of $\bar{x}(\cdot)$ at σ.

Let $\widehat{x}_N$ be an optimal solution of the approximation problem. By the Delta Theorem it follows that under the above assumptions, $N^{1/2}(\widehat{x}_N - x^*)$ converges in distribution to normal $\mathcal{N}(0, J^\top \Gamma J)$, where J is the $p^2 \times n$ matrix such that $J^\top \delta$ is the optimal solution of problem (14).