Computing a Nearest Correlation Matrix with Factor Structure

Nick Higham
School of Mathematics
The University of Manchester

higham@ma.man.ac.uk
http://www.ma.man.ac.uk/~higham/

Joint work with
Rüdiger Borsdorf, Marcos Raydan

NAG Quant Event, London, October 2009

Outline

- Properties of structured correlation matrices.
- Nearness problem for factor structured correlation matrices.
- Selection of optimization method.
- Numerical analysis issues.

Correlation Matrix

$n \times n$ symmetric positive semidefinite matrix A with $a_{ii} \equiv 1$.

- symmetric,
- 1s on the diagonal,
- eigenvalues nonnegative *or*
  all principal minors nonnegative.

Properties:

- off-diagonal elements between -1 and 1,
- convex set.

Quiz

Is this a correlation matrix?

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Quiz

Is this a correlation matrix?

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Spectrum: $-0.4142, 1.0000, 2.4142.$

Quiz

Is this a correlation matrix?

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}. \qquad \text{Spectrum: } -0.4142,\ 1.0000,\ 2.4142.$$

For what w is this a correlation matrix?

$$\begin{bmatrix} 1 & w & w \\ w & 1 & w \\ w & w & 1 \end{bmatrix}.$$

Quiz

Is this a correlation matrix?

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}. \qquad \text{Spectrum: } -0.4142,\ 1.0000,\ 2.4142.$$

For what w is this a correlation matrix?

$$\begin{bmatrix} 1 & w & w \\ w & 1 & w \\ w & w & 1 \end{bmatrix}. \qquad \frac{-1}{n-1} \le w \le 1.$$

Structured Correlation Matrices

- Nonnegative:

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & 1 & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & 1 \end{bmatrix}.$$

Structured Correlation Matrices

- Nonnegative:

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & 1 & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & 1 \end{bmatrix}.$$

- Low rank:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Structured Correlation Matrices

- Nonnegative:

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & 1 & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & 1 \end{bmatrix}.$$

- Low rank:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

- Factor structure:

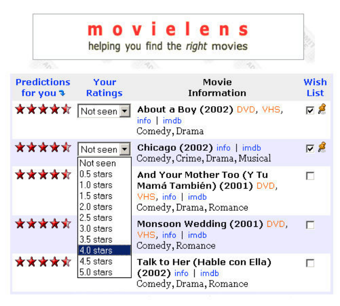
$$\begin{bmatrix} 1 & x_1 x_2 & x_1 x_3 \\ x_1 x_2 & 1 & x_2 x_3 \\ x_1 x_3 & x_2 x_3 & 1 \end{bmatrix}.$$

Approximate Correlation Matrices

Empirical correlation matrices often not true correlation matrices, due to

- asynchronous data
- missing data
- limited precision
- stress testing

http://www.movielens.org



Nearest Correlation Matrix

Find X achieving

$$\min\{ \|A - X\|_F : X \text{ is a correlation matrix} \},$$

where $\|A\|_F^2 = \sum_{i,j} a_{ij}^2$.

- ★ Constraint set is a closed, convex set, so unique minimizer.
- ★ Nonlinear optimization problem.

Optimization Problems—Issues

- Existence and uniqueness of solution.
- Convexity?
- Explicit, closed-form solution?
- Choice of algorithm.
- Availability of derivatives.
- Starting matrix and convergence criterion.
- Practical behaviour.

Quick and “Dirty” Differentiation

Analytic function $f : \mathbb{R} \to \mathbb{R}$, $i = \sqrt{-1}$.

Complex step approximation:

$$\boxed{f'(x) \approx \operatorname{Im} \frac{f(x+ih)}{h}.}$$

E.g., $h = 10^{-100}$.

Quick and "Dirty" Differentiation

Analytic function $f : \mathbb{R} \to \mathbb{R}$, $i = \sqrt{-1}$.

Complex step approximation:

$$\boxed{f'(x) \approx \operatorname{Im} \frac{f(x+ih)}{h}.}$$

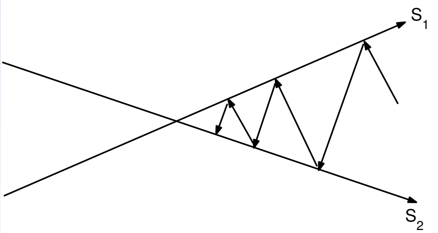
E.g., $h = 10^{-100}$.

$$\begin{aligned} f'(x) &= \operatorname{Im} \frac{f(x+ih)}{h} + O(h^2), \\ f(x) &= \operatorname{Re} f(x+ih) + O(h^2). \end{aligned}$$

See **MIMS EPrint 2009.31, April 2009**.

Alternating Projections Method

H (2002): repeatedly **project** onto the positive semidefinite matrices then the unit diagonal matrices.



- Easy to implement.
- Guaranteed convergence, at a linear rate.
- Can add further constraints/projections, e.g., fixed elements (Lucas, 2001).

Newton Method

Qi & Sun (2006): **Newton method** based on theory of strongly semismooth matrix functions.

- Applies Newton to **dual** (unconstrained) of $\min \frac{1}{2}\|A - X\|_F^2$ problem.
- **Globally** and **quadratically** convergent.
- H & Borsdorf (2007) improve efficiency and reliability by
  - using appropriate iterative method and eigensolver,
  - preconditioning the Newton direction solve.

The basis of **G02AAF** (nearest correlation matrix) in NAG Library Mark 22.

Factor Model

$$\xi = \underbrace{X}_{n\times k}\underbrace{\eta}_{k\times 1} + \underbrace{\mathrm{diag}(f_i)}_{n\times n}\underbrace{\varepsilon}_{n\times 1}, \qquad \eta, \varepsilon \in N(0,1).$$

Since $E(\xi) = 0$,

$$\mathrm{cov}(\xi) = E(\xi\xi^T) = XX^T + F^2.$$

Assume $\mathrm{var}(\xi_i) \equiv 1$. Then $\sum_{j=1}^k x_{ij}^2 + f_{ii}^2 = 1$, so

$$\sum_{j=1}^k x_{ij}^2 \le 1, \qquad i = 1\colon n.$$

- Collateralized debt obligations (CDOs),
- multivariate time series.

Structured Correlation Matrix

Yields correlation matrix of form

$$C(X) = D + XX^T = D + \sum_{j=1}^{k} x_j x_j^T,$$

$$D = \mathrm{diag}(I - XX^T), \qquad X = [x_1, \dots, x_k].$$

$C(X)$ has k **factor correlation matrix structure**.

Structured Correlation Matrix

Yields correlation matrix of form

$$C(X) = D + XX^T = D + \sum_{j=1}^{k} x_j x_j^T,$$

$$D = \mathrm{diag}(I - XX^T), \qquad X = [x_1, \dots, x_k].$$

$C(X)$ has k **factor correlation matrix structure**.

$$C(X) = \begin{bmatrix} 1 & y_1^T y_2 & \dots & y_1^T y_n \\ y_1^T y_2 & 1 & \dots & \vdots \\ \vdots & & \ddots & y_{n-1}^T y_n \\ y_1^T y_n & \dots & y_{n-1}^T y_n & 1 \end{bmatrix}, \quad y_i \in \mathbb{R}^k.$$

Aims

For k factor correlation matrices, investigate

- mathematical properties,
- nearness problem.

1-Parameter Correlation Matrix

$$X(w) = \begin{bmatrix} 1 & w & w \\ w & 1 & w \\ w & w & 1 \end{bmatrix}, \qquad w \in \mathbb{R}.$$

Theorem

$\min\{\, \|A - X(w)\|_F : X(w)$ *a corr. matrix* $\}$ *has unique solution the projection of*

$$w = \frac{e^T A e - \mathrm{trace}(A)}{n^2 - n},$$

onto $[-1/(n-1), 1]$.

Block Structured Correlation Matrix

$$\left[\begin{array}{cc|cc} 1 & \gamma_{11} & \gamma_{12} & \gamma_{12} \\ \gamma_{11} & 1 & \gamma_{12} & \gamma_{12} \\ \hline \gamma_{12} & \gamma_{12} & 1 & \gamma_{22} \\ \gamma_{12} & \gamma_{12} & \gamma_{22} & 1 \end{array}\right], \qquad C_{ij} = \begin{cases} C(\gamma_{ii}) \in \mathbb{R}^{n_i \times n_i}, & i = j, \\ \gamma_{ij} ee^T \in \mathbb{R}^{n_i \times n_j}, & i \neq j. \end{cases}$$

Objective function:

$$f(\Gamma) = \|A - C(\Gamma)\|_F^2 = \sum_{i=1}^{m} \|A_{ii} - C(\gamma_{ii})\|_F^2 + \sum_{i \neq j} \|A_{ij} - \gamma_{ij} ee^T\|_F^2.$$

- Convex constraint set $\Rightarrow$ unique minimizer.
- Alternating projections converges.

1-Factor Correlation Matrix

$$C(x) = \operatorname{diag}(1 - x_i^2) + xx^T, \qquad x \in \mathbb{R}^n$$

i.e., $c_{ij} = x_i x_j$, $i \neq j$.

Lemma

$$\det(C(x)) = \prod_{i=1}^{n}(1 - x_i^2) + \sum_{i=1}^{n} x_i^2 \prod_{\substack{j=1 \\ j \neq i}}^{n}(1 - x_j^2).$$

Corollary

If $|x| \leq e$ with $x_i = 1$ for at most one i then $C(x)$ is nonsingular. $C(x)$ is singular if $x_i = x_j = 1$ for some $i \neq j$.

Rank Result

$$C(x) = \mathrm{diag}(1 - x_i^2) + xx^T,$$

Theorem

Let $x \in \mathbb{R}^n$ with $|x| \le e$. Then $\mathrm{rank}(C(x)) = \min(p + 1, n)$*, where p is the number of x_i for which $|x_i| < 1$.*

$$x = [1\ 1\ 1\ x_4\ x_5] \quad \Rightarrow \quad C(x) = \begin{bmatrix} 1 & 1 & 1 & x_4 & x_5 \\ 1 & 1 & 1 & x_4 & x_5 \\ 1 & 1 & 1 & x_4 & x_5 \\ x_4 & x_4 & x_4 & 1 & x_4 x_5 \\ x_5 & x_5 & x_5 & x_4 x_5 & 1 \end{bmatrix}.$$

One-Factor Problem

$$\min_{x \in \mathbb{R}^n} f(x) := \|A - C(x)\|_F^2$$
$$\text{subject to} \quad -e \le x \le e.$$

- Objective function is nonconvex.
- The constraint implies $C(x)$ is a correlation matrix.

One-Factor Problem: Derivatives

- **Objective**:

$$f(x) = \langle A - I, A - I \rangle_F - 2x^T(A - I)x + (x^Tx)^2 - \sum_{i=1}^n x_i^4.$$

- **Gradient**:

$$\nabla f(x) = 4((x^Tx)x - (A - I)x - \mathrm{diag}(x_i^2)x).$$

- **Hessian**:

$$\nabla^2 f(x) = 4(2xx^T + (x^Tx + 1)I - A - 3\mathrm{diag}(x_i^2)).$$

- $\nabla f(x)$, $\nabla^2 f(x)$ cheap.
- $f(x)$ has a saddle point at $x = 0$.

Case 1: $f(x_*) = 0$

If $f(x_*) = 0$ then $\nabla^2 f(x_*)$ has the form

$$H_n(x) = (x^T x)I + xx^T - 2D^2, \qquad x \in \mathbb{R}^n.$$

For example, for $n = 4$:

$$\begin{bmatrix} x_2^2 + x_3^2 + x_4^2 & x_1 x_2 & x_1 x_3 & x_1 x_4 \\ x_2 x_1 & x_1^2 + x_3^2 + x_4^2 & x_2 x_3 & x_2 x_4 \\ x_3 x_1 & x_3 x_2 & x_1^2 + x_2^2 + x_4^2 & x_3 x_4 \\ x_4 x_1 & x_4 x_2 & x_4 x_3 & x_1^2 + x_2^2 + x_3^2 \end{bmatrix}.$$

Case 1: $f(x_*) = 0$

If $f(x_*) = 0$ then $\nabla^2 f(x_*)$ has the form

$$H_n(x) = (x^T x)I + xx^T - 2D^2, \qquad x \in \mathbb{R}^n.$$

For example, for $n = 4$:

$$\begin{bmatrix} x_2^2 + x_3^2 + x_4^2 & x_1 x_2 & x_1 x_3 & x_1 x_4 \\ x_2 x_1 & x_1^2 + x_3^2 + x_4^2 & x_2 x_3 & x_2 x_4 \\ x_3 x_1 & x_3 x_2 & x_1^2 + x_2^2 + x_4^2 & x_3 x_4 \\ x_4 x_1 & x_4 x_2 & x_4 x_3 & x_1^2 + x_2^2 + x_3^2 \end{bmatrix}.$$

Theorem

H_n is positive semidefinite. Moreover, H_n is nonsingular iff at least three of $x_1, x_2, \ldots, x_n$ are nonzero.

Case 2: $f(x_*) > 0$

Can write

$$\tfrac{1}{4}\nabla^2 f(x) \quad = H_n(x) + E_n(x)$$

where E_n has the form

$$E_4 = \begin{bmatrix} 0 & x_1x_2 - a_{12} & x_1x_3 - a_{13} & x_1x_4 - a_{14} \\ x_2x_1 - a_{21} & 0 & x_2x_3 - a_{23} & x_2x_4 - a_{24} \\ x_3x_1 - a_{31} & x_3x_2 - a_{32} & 0 & x_3x_4 - a_{34} \\ x_4x_1 - a_{41} & x_4x_2 - a_{42} & x_4x_3 - a_{43} & 0 \end{bmatrix}.$$

Hence, if $|x_ix_j - a_{ij}|$ is sufficiently small and H_n positive definite a stationary point will be a local minimizer.

k Factor Problem

$$C(X) := I - \mathrm{diag}(XX^T) + XX^T \quad \text{with } X \in \mathbb{R}^{n\times k}.$$

Representation not unique!

$$\sum_{j=1}^{k} x_{ij}^2 \le 1 \quad \Longrightarrow \quad C(X) \text{ is a correlation matrix.}$$

The k factor problem is

$$\min_{X\in\mathbb{R}^{n\times k}} f(x) := \|A - C(X)\|_F^2 \quad \text{subject to} \quad \sum_{j=1}^{k} x_{ij}^2 \le 1.$$

k Factor Problem: Derivatives

- Gradient

  $$\nabla f(X) = 4(X(X^TX) - AX + X - \mathrm{diag}(XX^T)X)$$

- Hessian given implicitly, can be viewed as a matrix representation of the Fréchet derivative of $\nabla f(X)$.

Choice of Optimization Method

- Derivatives available.
- Ignore the constraints?
- Starting matrix, convergence test?

Choice of Optimization Method

- Derivatives available.
- Ignore the constraints?
- Starting matrix, convergence test?

- Rich set of solvers in NAG Library, Mark 22:
    - E04 - Minimizing or Maximizing a Function
    - E05 - Global Optimization of a Function
- MATLAB Optimization toolbox.

Alternating Directions

$$f(x_{ij}) = \text{const.} + 2\sum_{q\neq i}\left(a_{iq} - \sum_{s=1}^{k} x_{is}x_{qs}\right)^2.$$

Hence $f'(x_{ij}) = 0$ if

$$x_{ij} = \frac{\sum_{q\neq i} x_{qj}\left(a_{iq} - \sum_{s\neq j} x_{is}x_{qs}\right)}{\sum_{q\neq i} x_{qj}^2}.$$

Project x_{ij} onto $[-1, 1]$.

- Convergence guaranteed.
- Limit may not be feasible for $k > 1$.

Principal Factors Method

Anderson, Sidenius & Basu (2003): with
$F(X) = I - \mathrm{diag}(XX^T)$,

$$X_i = \mathrm{argmin}_{X \in \mathbb{R}^{n\times k}} \|A - F(X_{i-1}) - XX^T\|_F.$$

Min obtained by eigendecomposition of $A - F(X_{i-1})$.
Equivalent to **alternating projections method** for

$$\mathcal{U} := \{W \in \mathbb{R}^{n\times n} : w_{ij} = a_{ij} \text{ for } i \neq j\} \quad \textbf{convex,}$$

$$\mathcal{S} := \{W \in \mathbb{R}^{n\times n} : W = XX^T \text{ for } X \in \mathbb{R}^{n\times k}\} \quad \textbf{nonconvex!}$$

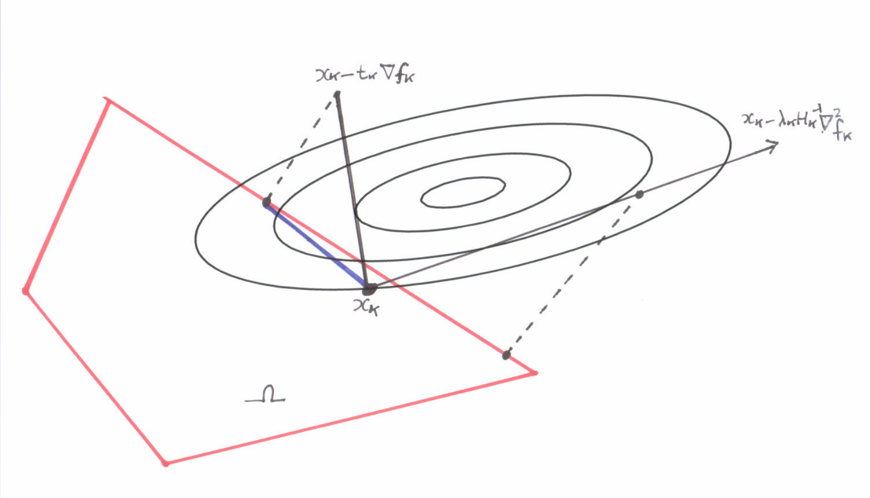
- Alt proj theory says no guarantee of convergence!
- Constraints ignored, so project final iterate onto them.

Spectral Projected Gradient Method

Birgin, Martínez & Raydan (2000).
To minimize $f : \mathbb{R}^n \to \mathbb{R}$ over convex set Ω:

$$x_{k+1} = x_k + \alpha_k d_k.$$

- $d_k = P_\Omega(x_k - t_k \nabla f(x_k)) - x_k$ is descent direction,
- $\alpha_k \in [-1, 1]$ chosen through **nonmonotone line search** strategy.
- Promising since $P_\Omega(\cdot)$ is cheap to compute.



- Code in ACM TOMS (Alg 813).

Test Examples

- **corr**: `gallery('randcorr',n)`
- **nrand**: $\frac{1}{2}(B + B^T) + \mathrm{diag}(I - B)$ with $B \in [-1, 1]^{n\times n}$ such that $\lambda_{\min}(B) < 0$.

Results averaged over 10 instances.

- **AD**: alternating directions.
- **PFM**: principal factors method.
- **Nwt**: `e04lb` of NAG Toolbox for MATLAB (modified Newton), bound constraints.
- **SPGM**: spectral projected gradient method.

Comparison: $k = 1, n = 2000$

	tol$= 10^{-3}$			tol$= 10^{-6}$		
	t(sec.)	**#its**	$\sqrt{f(X^*)}$	**t(sec.)**	**#its**	$\sqrt{f(X^*)}$
corr, $f(X_0) = 26.0$						
AD	3.3	5.2	26.0	3938	7282	26.0
PFM	68	1.1	26.0	827	18	26.0
Nwt	23	1.8	26.0	36	5.0	26.0
SPGM	9.8	5.2	26.0	638	760	26.0
nrand $f(X_0) = 825.13$						
AD	3.8	7.2	815.79	3.4	10.0	815.79
PFM	22	3.0	815.81	19.0	4.0	815.81
Nwt	4167	1222	815.79	4312	1229	815.79
SPGM	9.4	7.2	815.79	11	9.6	815.79

Comparison: $k = 6$, $n = 1000$

	tol$= 10^{-3}$			tol$= 10^{-6}$		
	t(sec.)	**#its**	$\sqrt{f(X^*)}$	**t(sec.)**	**#its**	$\sqrt{f(X^*)}$
corr, $f(X_0) = 18.5$						
AD	704	836	18.38	5060	5955	18.38
PFM	10	4.1	18.38	95	28.1	18.38
Nwt	167	52	18.38	280	68.2	18.38
SPGM	24	235	18.38	108	892	18.38
nrand , $f(X_0) = 415$						
AD	8694	9816	421	1.13e4	1.28e4	414
PFM	10.1	6.0	421	9.8	10	420
Nwt	146	40.8	421	109	56	420
SPGM	122	1263	407	276	2925	407

Conclusions

- Performance of methods depends on
  - the problem type,
  - the required tolerance,
  - the problem size.
- **Alternating directions** good for $k = 1$, low accuracy.
- **Principal factors method** generally fast, but may not converge to feasible point.
- **Spectral projected gradient** method wins overall.
- Incorporating the constraints need not hurt performance.
- Exploit convexity when present.

References I

L. Anderson, J. Sidenius, and S. Basu.
All your hedges in one basket.
Risk, pages 67–72, Nov. 2003.
|www.risk.net|.

J. Barzilai and J. M. Borwein.
Two-point step size gradient methods.
IMA J. Numer. Anal., 8:141–148, 1988.

E. G. Birgin, J. M. Martínez, and M. Raydan.
Nonmonotone spectral projected gradient methods on convex sets.
SIAM J. Optim., 10(4):1196–1211, 2000.

References II

- E. G. Birgin, J. M. Martínez, and M. Raydan.
  Algorithm 813: SPG—Software for convex-constrained optimization.
  ACM Trans. Math. Software, 27(3):340–349, 2001.
- E. G. Birgin, J. M. Martínez, and M. Raydan.
  Spectral projected gradient methods.
  In C. A. Floudas and P. M. Pardalos, editors, *Encyclopedia of Optimization*, pages 3652–3659. Springer-Verlag, Berlin, second edition, 2009.
- P. Glasserman and S. Suchintabandid.
  Correlation expansions for CDO pricing.
  Journal of Banking & Finance, 31:1375–1398, 2007.

References III

N. J. Higham.
Computing the nearest correlation matrix—A problem from finance.
IMA J. Numer. Anal., 22(3):329–343, 2002.

C. Lucas.
Computing nearest covariance and correlation matrices.

M.Sc. Thesis, University of Manchester, Manchester, England, Oct. 2001.

References IV

H.-D. Qi and D. Sun.
A quadratically convergent Newton method for computing the nearest correlation matrix.
SIAM J. Matrix Anal. Appl., 28(2):360–385, 2006.

P. Sonneveld, J. J. I. M. van Kan, X. Huang, and C. W. Oosterlee.
Nonnegative matrix factorization of a correlation matrix.
Linear Algebra Appl., 431:334–349, 2009.

References V

A. Vandendorpe, N.-D. Ho, S. Vanduffel, and P. Van Dooren.
On the parameterization of the CreditRisk$^+$ model for estimating credit portfolio risk.
Insurance: Mathematics and Economics, 42(2): 736–745, 2008.