MANCHESTER 1824

The University of Manchester

M\cr NA
Manchester Numerical Analysis

Recent Progress on the Nearest Correlation Matrix Problem

Nick Higham
School of Mathematics
The University of Manchester

http://www.maths.manchester.ac.uk/~higham/talks

Joint work with
Nataša Strabić and Vedran Šego

New Directions in Numerical Computation,
August 25-28, 2015, Oxford.
In celebration of Nick Trefethen's 60th birthday

Invalid Correlation Matrices

Correlation matrix: symm pos semidef with unit diagonal.

- Sample correlation matrix from empirical data.
- May lack definiteness due to
  - missing observations,
  - asynchronous observations,
  - left-censored data,
  - stress testing,
  - expert judgement,
  - separate blocks joined together (aggregation).

How do we make the matrix (semi)definite?

Correlation Matrix

A **correlation matrix** has

- ones on the diagonal,
- all eigenvalues nonnegative,
- all elements between -1 and 1.

Correlation Matrix

A **correlation matrix** has

- ones on the diagonal,
- all eigenvalues nonnegative,
- all elements between -1 and 1.

Is this a correlation matrix?

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

Correlation Matrix

A **correlation matrix** has

- ones on the diagonal,
- all eigenvalues nonnegative,
- all elements between -1 and 1.

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.

- Vary a_{13}: it must be 1 for a correlation matrix.

London Finance Company Question (2000)

"Given a real symmetric matrix *A* which is almost a correlation matrix what is the best approximating (in Frobenius norm?) correlation matrix?"

- **Massage** the original data, e.g., plug gaps.
- Make **ad hoc modifications** to matrix: e.g., shift negative e'vals up to zero then diagonally scale.
- Find a **nearest** correlation matrix. $\surd$

Literature search: very little found.

Spherical Parametrization

Pinheiro & Bates (1996), Rebonato & Jäckel (2000).

.

Correlation matrix $A = R^T R$ with ($n = 3$):

$$R = \begin{bmatrix} 1 & \cos\theta_{12} & \cos\theta_{13} \\ 0 & \sin\theta_{12} & \sin\theta_{13}\cos\theta_{23} \\ 0 & 0 & \sin\theta_{13}\sin\theta_{23} \end{bmatrix}.$$

Hence $\min\{\, \|A - R^T R\|_F^2 : \theta_{ij} \in [0,\pi],\ 1 \le i < j \le n \,\}$.

Spherical Parametrization

Pinheiro & Bates (1996), Rebonato & Jäckel (2000).

.

Correlation matrix $A = R^T R$ with ($n = 3$):

$$R = \begin{bmatrix} 1 & \cos\theta_{12} & \cos\theta_{13} \\ 0 & \sin\theta_{12} & \sin\theta_{13}\cos\theta_{23} \\ 0 & 0 & \sin\theta_{13}\sin\theta_{23} \end{bmatrix}.$$

Hence $\min\{\, \|A - R^T R\|_F^2 : \theta_{ij} \in [0, \pi],\ 1 \le i < j \le n \,\}$.

Nonlinear, lots of local minima!

Problem

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

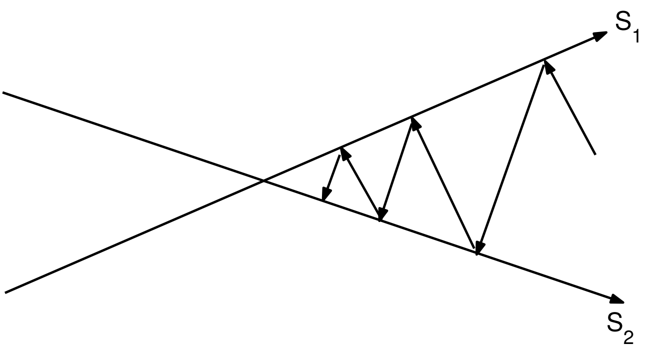
- $X \in S_n \cap \mathcal{U}_n$, where

$$S_n = \{\, X \in \mathbb{R}^{n\times n} : X \text{ is symm pos semidef} \,\},$$
$$\mathcal{U}_n = \{\, X = X^T \in \mathbb{R}^{n\times n} \colon x_{ii} = 1,\ i = 1 : n \,\}.$$

- Constraint a closed, convex set, so unique minimizer.
- H (2002):
  - Characterization of solution using normal cones of convex sets.
  - Alternating projections algorithm.

Alternating Projections

von Neumann (1933), for subspaces.



Dykstra (1983) incorporated corrections for closed convex sets.

Algorithm (H, 2002)

Given $A = A^T \in \mathbb{R}^{n\times n}$, compute nearest correlation matrix.

$\Delta S_0 = 0,\ Y_0 = A$
for $k = 1, 2, \ldots$
  $R_k = Y_{k-1} - \Delta S_{k-1}$
  $X_k = \mathcal{P}_{S_n}(R_k)$ % Project onto S_n.
  $\Delta S_k = X_k - R_k$ % Dykstra's correction.
  $Y_k = \mathcal{P}_{\mathcal{U}_n}(X_k)$ % Project onto $\mathcal{U}_n$.
end
Return Y_k.

- X_k and Y_k both converge to solution.
- Linear convergence, at best.
- Can add further constraints/projections ...

Unexpected Applications

Some recent papers (all use alternating projections):

- **Simulating wireless links in vehicular networks** (2014)
- **Analysing carbon dioxide storage resources** (2013)
- **Applying stochastic small-scale damage functions to German winter storms** (2012)
- **Predicting breeding values for eventing disciplines and grades in sport horses** (2012)
- **Characterisation of tool marks on cartridge cases by combining multiple images** (2012)
- **Experiments in reconstructing twentieth-century sea levels** (2011)

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.
- Dual problem is ctsly differentiable, but *not twice differentiable* $\Rightarrow$ use generalized Jacobian of gradient.
- **Globally** and **quadratically** convergent.

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.
- Dual problem is ctsly differentiable, but *not twice differentiable* $\Rightarrow$ use generalized Jacobian of gradient.
- **Globally** and **quadratically** convergent.
- Practical improvements: Borsdorf & H (2010).
- NAG code `g02aaf`: order of magnitude faster than alt proj. See NCM blog post (2013)
- Cannot incorporate fixed elements!

Accelerating a Fixed-Point Iteration

We are looking for x_* such that $g(x_*) = x_*$ for $g\colon \mathbb{R}^n \to \mathbb{R}^n$.

Fixed-point iteration

$$x_{k+1} = g(x_k), \quad k \geq 1, \qquad x_0 \in \mathbb{R}^n \text{ given.}$$

Accelerating a Fixed-Point Iteration

We are looking for x_* such that $g(x_*) = x_*$ for $g\colon \mathbb{R}^n \to \mathbb{R}^n$.

Fixed-point iteration

$$x_{k+1} = g(x_k), \quad k \geq 1, \qquad x_0 \in \mathbb{R}^n \text{ given.}$$

Acceleration

The iteration history for some chosen $m \geq 0$ is

$$\begin{array}{cccccc} \dots & x_{k-m} & x_{k-(m-1)} & \dots & x_{k-1} & x_k \\ \dots & g(x_{k-m}) & g(x_{k-(m-1)}) & \dots & g(x_{k-1}) & g(x_k). \end{array}$$

- Define x_{k+1} using all of this information.

Anderson Acceleration

Given history length m.

$x_1 = g(x_0)$
for $k = 1, 2, \ldots$ until convergence
    $m_k = \min(m, k)$
    Determine $\theta_1, \ldots, \theta_{m_k}$ to minimize $\|u_k - v_k\|_2^2$, where

$$u_k = x_k + \sum_{j=1}^{m_k} \theta_j (x_{k-j} - x_k),$$

$$v_k = g(x_k) + \sum_{j=1}^{m_k} \theta_j \big(g(x_{k-j}) - g(x_k)\big).$$

    $x_{k+1} = v_k$
end

- If g linear, objective function is $\|u_k - g(u_k)\|_2^2$.

History of Anderson Acceleration

- Originates with Anderson (1965): integral equations.
- In quantum chemistry known as *Pulay mixing* or *direct inversion in the iterative subspace* (DIIS) (Pulay, 1980).
- Recent papers by numerical analysts, e.g., Walker & Ni (2011), Toth & Kelley (2015).

Theory

No general guarantees of convergence!

- Does not require the iteration to be linearly convergent.
- Related to multisecant quasi-Newton methods; equivalent to "bad" Broyden.
- For $Ax = b$ with $m_k = k$, essentially GMRES.
- Some analysis on the convergence for contractive mappings of Anderson acceleration with *fixed m*.

Practicalities

- AA is implemented using differences of function values and iterates.
- A **linear least squares** problem is solved at each step (the main cost of AA).
- Can write alternating projections for NCM in fixed-point, vector form and apply AA.

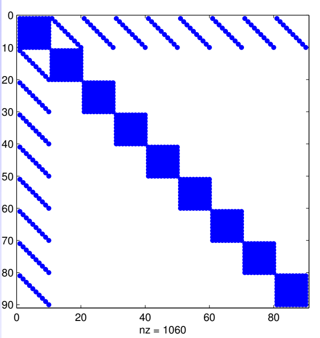
Experiment 1

- Five invalid correlation matrices from the literature.
- Iterations: `nearcorr` vs. `nearcorr_AA`.

n	`it`	`itAA` $m=1$	$m=2$	$m=3$	$m=4$	$m=5$
4	39	15	10	9	9	9
5	27	17	14	12	11	10
6	801	305	212	117	126	40
7	33	15	10	10	10	9

Experiment 2: Fixed Elements (1)

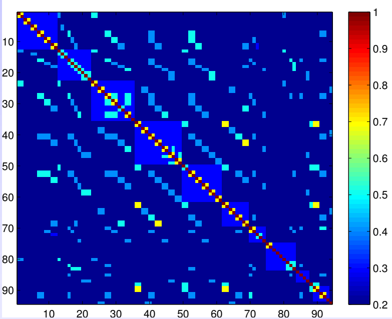
george: 90×90



Keep fixed:

- (1,1) block.
- Main diagonal (not unit).
- "Small" diagonals.

madalyn: 94×94



Keep fixed:

- All diagonal blocks (respective sizes 12, 5, 1, 14, 12, 1, 10, 4, 5, 9, 13 and 8).

Experiment 2: Fixed Elements (2)

- `nearcorr` vs. `nearcorr_fe` vs. `nearcorr_fe_AA`
- $n = 90$: finance.
- $n = 94$: carbon dioxide storage.

n	it	it_fe	itAA_fe				
			$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$
90	29	169	93	70	55	45	39
94	18	40	15	14	12	12	12

H & Strabić (2015), **Anderson acceleration of the alternating projections method for computing the nearest correlation matrix**, MIMS EPrint 2015.39. Codes **available on Github**.

higham / **anderson-accel-ncm**

Unwatch 1 Star 1 Fork 0

Anderson acceleration for the alternating projections method for the nearest correlation matrix problem — Edit

12 commits **1** branch **0** releases **1** contributor

Code
Issues 0
Pull requests 0
Wiki
Pulse
Graphs
Settings

Branch: **master** anderson-accel-ncm / +

Change filename to lower case in nearcorr_aa.m.

higham authored 22 minutes ago latest commit 029ac6078c

license.txt	Punctuation corrected in license.	19 days ago
nearcorr_aa.m	Change filename to lower case in nearcorr_aa.m.	22 minutes ago
nearcorr_new.m	Typo in comment corrected.	28 minutes ago
readme.md	Minor change to wording of readme.	19 days ago
test_anderson.m	Changes to comments, remove unused parameter beta from nearcorr_aa.m.	18 hours ago

HTTPS clone URL

`https://github.com/h`

You can clone with HTTPS, SSH, or Subversion.

Clone in Desktop

Download ZIP

readme.md

`anderson-accel-ncm` - MATLAB Codes for Anderson acceleration of the alternating projections method for the nearest correlation matrix

Shrinking

- **Invalid** correlation matrix C.
- **Target** correlation matrix T.
- Replace C by $S(\alpha) = (1-\alpha)C + \alpha T$, where $\alpha \in [0,1]$.

Large literature on shrinking in which

- C and T are *both* cov/correl matrices.
- α is chosen subject to *statistical considerations*.

Our optimal shrinking parameter:

$$\alpha_* = \min\{\alpha \in [0,1] : S(\alpha) \text{ is pos semidef}\}.$$

Additional Requirement: Fixed Block

Data: N random variables, K observations $\rightsquigarrow X \in \mathbb{R}^{K \times N}$.
Arrange so first m columns have no missing data:

$$X = [\, x_1 \quad \dots \quad x_m \quad x_{m+1} \quad \dots \quad x_{m+n} \,].$$

$$C = \begin{matrix} & m & n \\ m & A & Y \\ n & Y^T & B \end{matrix}$$

- Symmetric. ✓
- Unit diagonal. ✓
- Elements in $[-1, 1]$. ✓
- Positive semidefinite. ✗

Transform C into

- a valid correlation matrix,
- while keeping a positive semidefinite block A.

The Shrinking Method

$$S(\alpha) := \alpha \underbrace{\begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}}_{\text{target}} + (1-\alpha) \begin{bmatrix} A & Y \\ Y^T & B \end{bmatrix}, \quad \alpha \in [0,1].$$

$$S(\alpha) = \begin{bmatrix} A & (1-\alpha)Y \\ (1-\alpha)Y^T & \alpha I + (1-\alpha)B \end{bmatrix}$$

- Symmetric. ✓
- Unit diagonal. ✓
- Upper-left block A. ✓

The Shrinking Method

$$S(\alpha) := \alpha \underbrace{\begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}}_{\text{target}} + (1-\alpha)\begin{bmatrix} A & Y \\ Y^T & B \end{bmatrix}, \quad \alpha \in [0,1].$$

$$S(\alpha) = \begin{bmatrix} A & (1-\alpha)Y \\ (1-\alpha)Y^T & \alpha I + (1-\alpha)B \end{bmatrix}$$

- Symmetric. ✓
- Unit diagonal. ✓
- Upper-left block A. ✓

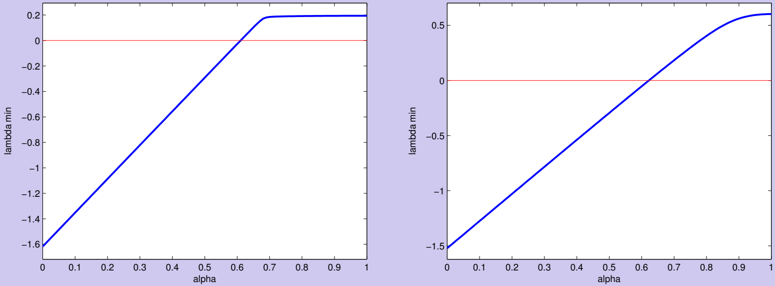
Find

$$\begin{aligned} \alpha_* &= \min\{\, \alpha \in [0,1] : S(\alpha) \text{ pos semidef} \} \\ &= \min\{\, \alpha \in [0,1] : f(\alpha) := \lambda_{\min}(S(\alpha)) \geq 0 \,\}. \end{aligned}$$

Minimal uniform rel change to each unfixed element.

Properties of $f(\alpha) = \lambda_{\min}(S(\alpha))$

A positive definite



- f is concave.
- $f(0) < 0$, $f(1) > 0$.
- α_* is the unique zero of f in $[0, 1]$.

Bisection with Cholesky

$$C := \begin{array}{c c} & \begin{array}{cc} m & n \end{array} \\ \begin{array}{c} m \\ n \end{array} & \begin{bmatrix} A & Y \\ Y^T & B \end{bmatrix} \end{array}$$ invalid correlation matrix, A pos def.

Algorithm

- $A = R_{11}^T R_{11}$ (Cholesky decomposition).
- $R_{11}^T X = Y, \; Z = X^T X$
- Interval $[\alpha_\ell, \alpha_r] \equiv [0, 1]$.

1. $\alpha_m = (\alpha_\ell + \alpha_r)/2$
2. $T = \alpha_m I + (1 - \alpha_m)B - (1 - \alpha_m)^2 Z$
3. Attempt Cholesky of T and set $\alpha_m = \alpha_\ell$ or $\alpha_m = \alpha_r$ accordingly.

Bisection with Cholesky

$$C := \begin{array}{c c} & \begin{array}{cc} m & n \end{array} \\ \begin{array}{c} m \\ n \end{array} & \begin{bmatrix} A & Y \\ Y^T & B \end{bmatrix} \end{array}$$ invalid correlation matrix, A pos def.

Algorithm

- $A = R_{11}^T R_{11}$ (Cholesky decomposition).
- $R_{11}^T X = Y, \; Z = X^T X$
- Interval $[\alpha_\ell, \alpha_r] \equiv [0, 1]$.

1. $\alpha_m = (\alpha_\ell + \alpha_r)/2$
2. $T = \alpha_m I + (1 - \alpha_m)B - (1 - \alpha_m)^2 Z$
3. Attempt Cholesky of T and set $\alpha_m = \alpha_\ell$ or $\alpha_m = \alpha_r$ accordingly.

- Guarantee psd matrix by taking $\alpha_* \leftarrow \alpha_r$ in bisection.

Generalized Eigenvalue Approach

Looking for

$$\alpha_* = \min\{\alpha \in [0,1]\colon S(\alpha) \text{ is positive semidefinite}\},$$

where

$$S(\alpha) = \alpha T + (1-\alpha)C =: E - \alpha F.$$

E and F are both symm indef.

Generalized Eigenvalue Approach

Looking for

$$\alpha_* = \min\{\alpha \in [0,1] : S(\alpha) \text{ is positive semidefinite}\},$$

where

$$S(\alpha) = \alpha T + (1-\alpha)C =: E - \alpha F.$$

E and F are both symm indef. Rewrite

$$S(\alpha) = (1-\alpha)\left(\frac{\alpha}{1-\alpha}T + C\right) =: (1-\alpha)\left(-\mu T + C\right).$$

α_* from **smallest generalized eigenvalue** of the **definite pencil** $C - \mu T$. Solve by

- $T = R^T R$
- $G = R^{-T} C R^{-1}$
- Find smallest e'val μ_* of G.

Generalized Eigenvalue Approach

Looking for

$$\alpha_* = \min\{\alpha \in [0,1] \colon S(\alpha) \text{ is positive semidefinite}\},$$

where

$$S(\alpha) = \alpha T + (1-\alpha) C =: E - \alpha F.$$

E and F are both symm indef. Rewrite

$$S(\alpha) = (1-\alpha)\left(\frac{\alpha}{1-\alpha} T + C\right) =: (1-\alpha)\left(-\mu T + C\right).$$

α_* from **smallest generalized eigenvalue** of the **definite pencil** $C - \mu T$. Solve by

- $T = R^T R$
- $G = R^{-T} C R^{-1}$
- Find smallest e'val μ_* of G.

Easily modified to exploit fixed (1,1) block.

Experiment: Shrinking versus NCM

- $n = 1399, 3120$: matrices from finance industry.
- `g02aa` is NAG Newton NCM code.
- Shrinking is done by bisection.
- Times (secs).

n	shrinking		`g02aa`		Distance $\|\cdot\|_F$	
	1e-3	1e-6	1e-3	1e-6	shrinking	NCM
1399	0.2	0.3	3.7	4.4	321.0	21.0
3120	1.0	2.2	28.1	34.3	178.7	5.4
2798	0.7	1.6	44.2	50.9	1221.2	1089.5
4519	2.3	5.0	220.9	234.7	1761.5	1631.5
6240	7.1	17.5	447.3	449.9	2578.1	2446.8

Shrinking Summary

- Attractive way to restore definiteness.
- Order of magnitude **faster** than computing NCM.
- Can easily incorporate **weighting**.
- Codes **available on Github**.
- Code `g02anf` in NAG Library Mark 25.
  Weighting is being added.

H, Strabić & Šego (2014), **Restoring definiteness via shrinking, with an application to correlation matrices with a fixed block**, MIMS EPrint 2014.54; to appear in SIAM Review.

Conclusions

- Invalid correlation matrices are **ubiquitous**.
  Frequently replaced by *nearest correlation matrix*.
- Practitioners use **alt proj** because easily available (MATLAB, R).
- For NCM with fixed element constraints, recommend **alternating projections + Anderson acceleration**.
- Anderson acceleration reduces # iterations by
  - **at least a half** for standard NCM and
  - **at least a third** for the (harder) variants.
- Shrinking is an attractive alternative: order of magnitude faster than NCM.

References I

D. G. Anderson.
Iterative procedures for nonlinear integral equations.
J. Assoc. Comput. Mach., 12(4):547–560, Oct. 1965.

R. Borsdorf and N. J. Higham.
A preconditioned Newton algorithm for the nearest correlation matrix.
IMA J. Numer. Anal., 30(1):94–107, 2010.

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

References II

N. J. Higham and N. Strabić.
Anderson acceleration of the alternating projections method for computing the nearest correlation matrix.
MIMS EPrint 2015.39, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, Aug. 2015.
22 pp.

References III

N. J. Higham, N. Strabić, and V. Šego.
Restoring definiteness via shrinking, with an application to correlation matrices with a fixed block.
MIMS EPrint 2014.54, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, Nov. 2014.
19 pp.
Revised March 2015. To appear in SIAM Rev.

J. Pinheiro and D. M. Bates.
Unconstrained parametrizations for variance-covariance matrices.
Statistics and Computing, 6(3):289–296, 1996.

References IV

P. Pulay.
Convergence acceleration of iterative sequences. The case of SCF iteration.
Chem. Phys. Lett., 73(2):393–398, 1980.

H. 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.

R. Rebonato and P. Jäckel.
The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes.
Journal of Risk, 2(2):17–27, 2000.

References V

A. Toth and C. T. Kelley.
Convergence analysis for Anderson Acceleration.
SIAM J. Numer. Anal., 53(2):805–819, 2015.

H. F. Walker and P. Ni.
Anderson acceleration for fixed-point iterations.
SIAM J. Numer. Anal., 49(4):1715–1735, 2011.