Fixing A Broken Correlation Matrix

Brian Spector

July 15th 2014

Actuarial Research Conference
University of California,
Santa Barbara

Agenda

- NAG Introduction
- The Nearest Correlation Matrix Problem
- Numerical computation – why bother
  - Problems in numerical computation
- Computational problems in Actuarial Science

Numerical Algorithms Group - What We Do

- NAG provides mathematical and statistical algorithm libraries widely used in industry and academia
- Established in 1970 with offices in Oxford, Manchester, Chicago, Taipei, Tokyo
- Not-for-profit organization committed to research & development
- Library code written and contributed by some of the world's most renowned mathematicians and computer scientists
- NAG's numerical code is embedded within many vendor libraries such as AMD and Intel
- Many collaborative projects – e.g. CSE Support to the UK's largest supercomputer, HECToR

NAG Library Contents

- C05: Root Finding
- C06: Summation of Series
- D01: Quadrature
- D02: ODEs
- D03: PDEs
- D04: Numerical Differentiation
- D05: Integral Equations
- E01: Interpolation
- E02: Curve and Surface Fitting
- E04: Local Optimization
- E05: Global Optimization
- F: Linear Algebra
- G01: Statistical Functions
- G02: Correlation / Regression
- G03: Multivariate Methods
- G05: RNGs
- G07: Univariate Estimation
- G08: Nonparametric Statistics
- G10: Smoothing in Statistics
- G12: Survival Analysis
- G13: Time Series Analysis
- H: Operations Research
- S: Special Functions

NAG Portfolio

- Numerical Libraries
  - Highly flexible for use in many computing languages, programming environments, hardware platforms and for high performance computing methods
- Connector Products for Excel, MATLAB, .NET, R, and Java
  - Giving users of the spreadsheets and mathematical software packages access to NAG's library of highly optimized and often superior numerical routines
- Consultancy services

Correlation Matrix

- Mathematically, a correlation matrix $C \in \mathbb{R}^{n \times n}$ is ...
  1. Square, Symmetric Matrix with ones on diagonal
  2. Positive semi-definite: $x^T C x \geq 0$ for all $x \in \mathbb{R}^n$
- Often estimated from "real world" – which is messy
- ☐ Ensuring (1) is trivial often (2) is tricky

Correlation Matrix

- Mathematically, a correlation matrix $C \in \mathbb{R}^{n \times n}$ is ...
  1. Square, Symmetric Matrix with ones on diagonal
  2. Positive semi-definite: $x^T C x \geq 0$ for all $x \in \mathbb{R}^n$
- Often estimated from "real world" – which is messy
- Ensuring (1) is trivial often (2) is tricky
- Is C = $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ a correlation matrix?

Correlation Matrix

- Mathematically, a correlation matrix $C \in \mathbb{R}^{n \times n}$ is ...
  1. Square, Symmetric Matrix with ones on diagonal
  2. Positive semi-definite: $x^T C x \geq 0$ for all $x \in \mathbb{R}^n$
- Often estimated from "real world" – which is messy
  - Ensuring (1) is trivial often (2) is tricky
-
- Is C $= \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ a correlation matrix?
- No ... Eigenvalues = {−0.4142, 1.0000, 2.4142}

Correlation Matrix

1	2	3	4	5	6	7	8	9	10
0.30609		-0.31969	-0.23833	-0.55802		-0.17671	-0.19964	0.12005	0.1637
0.43764		-0.23543	0.023406	-0.21202		-0.0294	0.088448	0.32388	0.26884
0.55743		-5.9688	0.32931	-5.6395		1.4488		6.8003	11.156
0.77245		-11.205	1.0352	-10.17		0.19865		10.2	11.35
0.65135		0.056145	-0.66712	-0.61098	1.3086	1.5162		-0.12741	1.4649
0.91497		3.178	-1.3864	1.7916	2.106	-0.17897		-3.2569	-3.7605
0.26635		0.21764	-0.34687	-0.12923	0.63255	0.10337		-0.25066	-1.0276
0.56408		-0.13557	-0.59245	-0.72802	0.62265	0.4615		0.57055	0.44818
0.95547		-0.48116	-0.58159	-1.0627	0.32174	0.16247		-0.74853	0.36455
0.23232		0.72734	0.56775	1.2951	-1.0064	-0.23799		-1.0875	-1.573
0.49441		-1.9103	-0.77478	-2.685	1.1146	2.671		2.5405	2.0808
0.34322		1.422	-0.68669	0.73529	1.3519	0.93479		-1.1706	-2.6415
0.36899		0.013091	-0.01921	-0.00612	-0.4148	-0.17068		-0.04897	0.70969
0.68344		0.001903		-0.87466	1.302	-0.20656		-0.62145	-1.4875
0.82884		-0.12102		-1.0508	1.0127	1.1156		-0.00505	0.14418
0.90667	-0.2333	0.90418		-0.12908	1.14	0.051782		-0.8835	-0.95283
0.080234	1.0532	0.8562		1.7494	-0.973	0.4542		-0.79168	-1.2658
0.6408	-0.4982	-0.8781		-1.6635	1.139	0.13208	-1.6678	-0.78969	0.73345

Correlation Matrix

- Why do I need a ‘Fixed’ Correlation Matrix?
  - Portfolio Optimization
    - MV Portfolio Sensitive to Estimates
  - Modeling Default Rates
  - Risk Calculations
  - Generating Missing Data
  - ???

Finding a Good Correlation Matrix

1. Compute Row Correlations

1	2	3	4	5	6	7	8	9	10
0.30609		-0.31969	-0.23833	-0.55802		-0.17671	-0.19964	0.12005	0.1637
0.43764		-0.23543	0.023406	-0.21202		-0.0294	0.088448	0.32388	0.26884
0.55743		-5.9688	0.32931	-5.6395		1.4488		6.8003	11.156
0.77245		-11.205	1.0352	-10.17		0.19865		10.2	11.35
0.65135		0.056145	-0.66712	-0.61098	1.3086	1.5162		-0.12741	1.4649
0.91497		3.178	-1.3864	1.7916	2.106	-0.17897		-3.2569	-3.7605
0.26635		0.21764	-0.34687	-0.12923	0.63255	0.10337		-0.25066	-1.0276
0.56408		-0.13557	-0.59245	-0.72802	0.62265	0.4615		0.57055	0.44818
0.95547		-0.48116	-0.58159	-1.0627	0.32174	0.16247		-0.74853	0.36455
0.23232		0.72734	0.56775	1.2951	-1.0064	-0.23799		-1.0875	-1.573
0.49441		-1.9103	-0.77478	-2.685	1.1146	2.671		2.5405	2.0808
0.34322		1.422	-0.68669	0.73529	1.3519	0.93479		-1.1706	-2.6415
0.36899		0.013091	-0.01921	-0.00612	-0.4148	-0.17068		-0.04897	0.70969
0.68344		0.001903		-0.87466	1.302	-0.20656		-0.62145	-1.4875
0.82884		-0.12102		-1.0508	1.0127	1.1156		-0.00505	0.14418
0.90667	-0.2333	0.90418		-0.12908	1.14	0.051782		-0.8835	-0.95283
0.080234	1.0532	0.8562		1.7494	-0.973	0.4542		-0.79168	-1.2658
0.6408	-0.4982	-0.8781		-1.6635	1.139	0.13208	-1.6678	-0.78969	0.73345

Finding a Good Correlation Matrix

1. Compute Row Correlations
2. $C = Q\Lambda Q^{-1}$

- where $\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$

Finding a Good Correlation Matrix

1. Compute Row Correlations
2. $C = Q\Lambda Q^{-1}$

- where $\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$

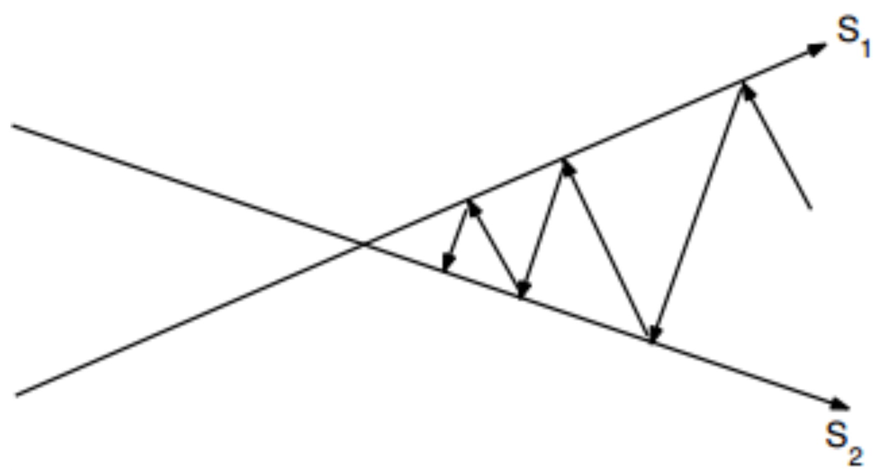
3. Use NAG routines to return “nearest” correlation matrix

$$\min_X ||X - G|| \; s.t. \; X \text{ is a correlation matrix}$$

Solution – Higham 2002

■ $\min_X ||X - G|| \ s.t. \ X$ is a correlation matrix

- Constraint Set is Closed/Convex!
- "closest approximation" to input (non-semidefinite) matrix
- Linear convergence



Qi and Sun (2006)

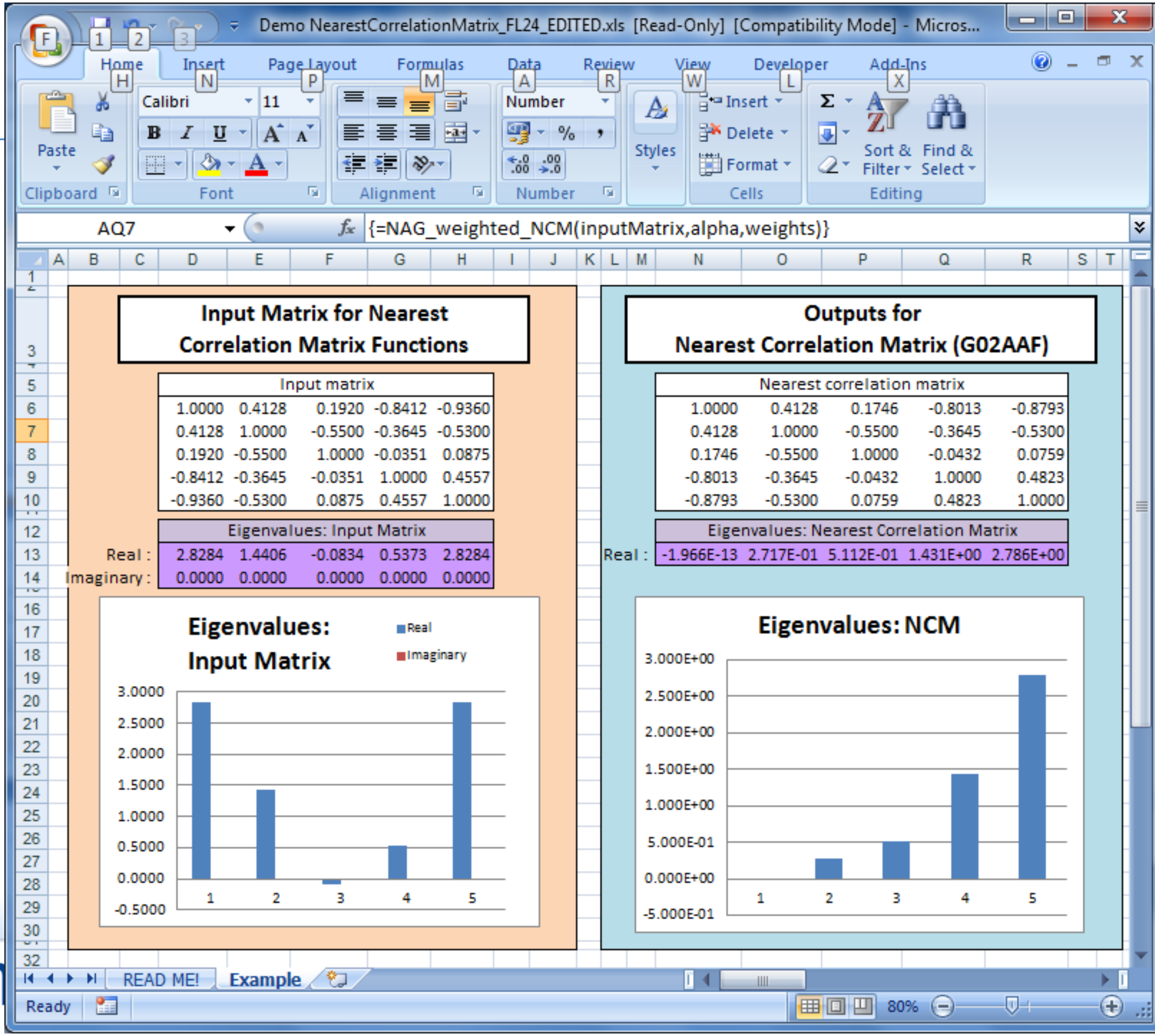
- Instead of:
  - $\min_{X} \frac{1}{2} ||X - A|| \; s.t. \; X$ is a correlation matrix
- Work on dual:
-
  - $\min_{y \in R^n} \frac{1}{2} ||\mathrm{A} + diag(y)_+|| - e^T y$
- With gradient
  - ∇ = diag$(\mathrm{A} + diag(y)_+) - e$
- Quadratic Convergence!

Nearest Correlation Matrices performance

Improvements to the NCM Algorithm

N	CL09	CL23		FS23	
1,000	4.5	4.2	6.7%	3.3	28.6%
2,000	34.5	28.5	17.4%	22	43.9%
3,000	110	101	8.2%	76	33.7%
4,000	288	273	5.2%	185	37.7%
5,000	572	533	6.8%	421	28.3%
10,000	4,237	3,952	6.7%	3,142	27.7%

run on an AMD quad processor machine (2.6 GHz, 16 cores in total), 64 bit Windows


Demo NearestCorrelationMatrix_FL24_EDITED.xls [Read-Only] [Compatibility Mode] - Micros...
AQ7
{=NAG_weighted_NCM(inputMatrix,alpha,weights)}
Input Matrix for Nearest Correlation Matrix Functions
Input matrix
1.0000 0.4128 0.1920 -0.8412 -0.9360
0.4128 1.0000 -0.5500 -0.3645 -0.5300
0.1920 -0.5500 1.0000 -0.0351 0.0875
-0.8412 -0.3645 -0.0351 1.0000 0.4557
-0.9360 -0.5300 0.0875 0.4557 1.0000
Eigenvalues: Input Matrix
Real : 2.8284 1.4406 -0.0834 0.5373 2.8284
Imaginary : 0.0000 0.0000 0.0000 0.0000 0.0000
Eigenvalues: Input Matrix
Real
Imaginary
Outputs for Nearest Correlation Matrix (G02AAF)
Nearest correlation matrix
1.0000 0.4128 0.1746 -0.8013 -0.8793
0.4128 1.0000 -0.5500 -0.3645 -0.5300
0.1746 -0.5500 1.0000 -0.0432 0.0759
-0.8013 -0.3645 -0.0432 1.0000 0.4823
-0.8793 -0.5300 0.0759 0.4823 1.0000
Eigenvalues: Nearest Correlation Matrix
Real : -1.966E-13 2.717E-01 5.112E-01 1.431E+00 2.786E+00
Eigenvalues: NCM
READ ME!
Example
Ready
80%

Nearest Correlation Matrix

- Additions to the NCM Algorithm
  - Bounds on eigenvalues
  - K-factor structure (can reduce dimensionality)
  - Weights
- NAG is keen to collaborate
  - Collaborative Projects:
    - HPCFinance.eu - http://www.hpcfinance.eu/
  - Academia:
    - Professor Nick Higham (University of Manchester)
  - Industry:
    - ISV: Supporting the porting of applications onto new platforms

Numerical Computations - Why bother?

- Numerical computation is difficult to do accurately
- Problems of
  - Overflow / underflow / rounding
    - How does the computation behave for large / small numbers?
  - Condition
    - How is it affected by small changes in the input?
  - Stability
    - How sensitive is the computation to rounding errors?
- Importance of
    - error analysis
    - information about error bounds on solution

Numerical Computing Problems

- Take 3 numbers: {-1, 0, 1}
  - Mean: 0
  - Standard Deviation: 1

Numerical Computing Problems

- Take 3 numbers: {-1, 0, 1}
  - Mean: 0
  - Standard Deviation: 1
- { (1e16) - 1, (1e16), (1e16) + 1}

	A	B	C	D	E	F	G	H
1								
2								
3								
4			n-1	1E+16		Average	1E+16	
5			n	1E+16		StDev	0	
6			n+1	1E+16				
7								
8								
9								

NAG and Actuarial Science - Summary

- NAG is keen on further collaborations in building actuarial models and risk engines
  - We want to make sure we provide what you need
- Risk engines likely to involve a LOT of computation
  - NAG has *significant* experience in HPC services, consulting and training
  - We know how to do large scale computations efficiently
  - *This is non-trivial*! Our expertise has been sought out and exploited by organisations such as (HECToR, Microsoft, Oracle, major Aerospace and Oil & Gas companies …….)

Keep in touch

brian.spector@nag.com

Industry Articles

http://www.nag.com/IndustryArticles/fixing-a-broken-correlation-matrix.pdf

NAGNews

http://www.nag.co.uk/NAGNews/Index.asp

Blog: http://blog.nag.com/