Fixing A Broken Correlation Matrix

Brian Spector

July 15th 2014

Actuarial Research Conference
University of California,
Santa Barbara

Experts in numerical algorithms and HPC services
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

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.30609</td>
<td>-0.31969</td>
<td>-0.23833</td>
<td>-0.55802</td>
<td>-0.17671</td>
<td>-0.19964</td>
<td>0.12005</td>
<td>0.1637</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.43764</td>
<td>-0.23543</td>
<td>0.023406</td>
<td>-0.21202</td>
<td>-0.0294</td>
<td>0.088448</td>
<td>0.32388</td>
<td>0.26884</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.55743</td>
<td>-5.9688</td>
<td>0.32931</td>
<td>-5.6395</td>
<td>1.4488</td>
<td>-6.8003</td>
<td>11.156</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.77245</td>
<td>-11.205</td>
<td>1.0352</td>
<td>-10.17</td>
<td>0.19865</td>
<td>-10.2</td>
<td>11.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.65135</td>
<td>0.056145</td>
<td>-0.66712</td>
<td>-0.61098</td>
<td>1.3086</td>
<td>1.5162</td>
<td>-0.12741</td>
<td>1.4649</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.91497</td>
<td>3.178</td>
<td>-1.3864</td>
<td>1.7916</td>
<td>2.106</td>
<td>-0.17897</td>
<td>-3.2569</td>
<td>-3.7605</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.26635</td>
<td>0.21764</td>
<td>-0.34687</td>
<td>-0.12923</td>
<td>0.63255</td>
<td>0.10337</td>
<td>-0.25066</td>
<td>-1.0276</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.56408</td>
<td>-0.13557</td>
<td>-0.59245</td>
<td>-0.72802</td>
<td>0.62265</td>
<td>0.4615</td>
<td>0.57555</td>
<td>0.44818</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.95547</td>
<td>-0.48116</td>
<td>-0.58159</td>
<td>-1.0627</td>
<td>0.32174</td>
<td>0.16247</td>
<td>-0.74853</td>
<td>0.36455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.23232</td>
<td>0.72734</td>
<td>0.56775</td>
<td>1.2951</td>
<td>-1.0064</td>
<td>-0.23799</td>
<td>-1.0875</td>
<td>-1.573</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.49441</td>
<td>-1.9103</td>
<td>-0.77478</td>
<td>-2.685</td>
<td>1.1146</td>
<td>2.671</td>
<td>2.5405</td>
<td>2.0808</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.34322</td>
<td>1.422</td>
<td>-0.68669</td>
<td>0.73529</td>
<td>1.3519</td>
<td>0.93479</td>
<td>-1.1706</td>
<td>-2.6415</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.36899</td>
<td>0.013091</td>
<td>-0.01921</td>
<td>-0.00612</td>
<td>-0.4148</td>
<td>-0.17068</td>
<td>-0.04897</td>
<td>0.70969</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.68344</td>
<td>0.001903</td>
<td>-0.87466</td>
<td>1.302</td>
<td>-0.20656</td>
<td>-0.62145</td>
<td>-1.4875</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.82884</td>
<td>-0.12102</td>
<td>-1.0508</td>
<td>1.0127</td>
<td>1.1156</td>
<td>-0.00505</td>
<td>0.14418</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90667</td>
<td>-0.2333</td>
<td>0.90418</td>
<td>-0.12908</td>
<td>1.14</td>
<td>0.051782</td>
<td>-0.8835</td>
<td>-0.95283</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.080234</td>
<td>1.0532</td>
<td>0.8562</td>
<td>1.7494</td>
<td>-0.973</td>
<td>0.4542</td>
<td>-0.79168</td>
<td>-1.2658</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.6408</td>
<td>-0.4982</td>
<td>-0.8781</td>
<td>-1.6635</td>
<td>1.139</td>
<td>0.13208</td>
<td>-1.6678</td>
<td>-0.78969</td>
<td>0.73345</td>
<td></td>
</tr>
</tbody>
</table>
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

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

1. Compute Row Correlations

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

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$$
Finding a Good Correlation Matrix

1. Compute Row Correlations

2. \( C = Q \Lambda Q^{-1} \)

\[
\Lambda = \begin{bmatrix}
\lambda_1 & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & \lambda_n
\end{bmatrix}
\]

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

\[
\min_X ||X - G|| \quad s.t. \ X \text{ is a correlation matrix}
\]
Solution – Higham 2002

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

- Constraint Set is Closed/Convex!
- “closest approximation” to input (non-semidefinite) matrix
- Linear convergence
Qi and Sun (2006)

- Instead of:
  \[ \min_X \frac{1}{2} \|X - A\| \quad s.t. \quad X \text{ is a correlation matrix} \]

- Work on dual:
  \[ \min_{y \in \mathbb{R}^n} \frac{1}{2} \|A + \text{diag}(y)_+\| - e^T y \]

- With gradient
  \[ \nabla = \text{diag}(A + \text{diag}(y)_+) - e \]

- Quadratic Convergence!
Nearest Correlation Matrices performance

Improvements to the NCM Algorithm

<table>
<thead>
<tr>
<th>N</th>
<th>CL09</th>
<th>CL23</th>
<th>FS23</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>4.5</td>
<td>4.2</td>
<td>6.7%</td>
</tr>
<tr>
<td>2,000</td>
<td>34.5</td>
<td>28.5</td>
<td>17.4%</td>
</tr>
<tr>
<td>3,000</td>
<td>110</td>
<td>101</td>
<td>8.2%</td>
</tr>
<tr>
<td>4,000</td>
<td>288</td>
<td>273</td>
<td>5.2%</td>
</tr>
<tr>
<td>5,000</td>
<td>572</td>
<td>533</td>
<td>6.8%</td>
</tr>
<tr>
<td>10,000</td>
<td>4,237</td>
<td>3,952</td>
<td>6.7%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>n-1</td>
<td></td>
<td>1E+16</td>
<td></td>
<td></td>
<td>Average</td>
<td></td>
<td>1E+16</td>
</tr>
<tr>
<td>5</td>
<td>n</td>
<td>1E+16</td>
<td></td>
<td></td>
<td>StDev</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>n+1</td>
<td>1E+16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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**


**NAGNews**

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

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