Explicit Solutions to Correlation Matrix Completion Problems for Risk and Insurance

Prof. Gareth W. Peters (FIOR, YAS-RSE)
Chair of Statistics for Risk and Insurance,
Department of Actuarial Mathematics and Statistics,
Heriot-Watt, Edinburgh, UK

Georgescu D. I., Higham N.J. and Peters G.W. (2018)
"**Explicit Solutions to Correlation Matrix Completion Problems, with an Application to Risk Management and Insurance**"
Royal Society Open Science.

May 28, 2019

Table of contents

Table of contents

Toy Example Motivation

- Correlation coefficients are typically fully specified in the business unit with the BU-specific risk.
- Correlations are also specified between similar risk families in different business units.

Toy Example Motivation

- Correlation coefficients are typically fully specified in the business unit with the BU-specific risk.
- Correlations are also specified between similar risk families in different business units.

Toy example: just two business units BU_1 and BU_2:

- Both are exposed to risks x and y, but only BU_1 is exposed to risk z.
- Correlations are specified between risk z, x, and y in BU_1, but not between x and y in BU_2, and z in BU_1.

correlations		x	y	z	x	y
BU_1	x	1	0.7	0.85	0.85	0.75
	y	0.7	1	0.6	0. 5	0.85
	z	0.85	0.6	1	*	*
BU_2	x	0.85	0.5	*	1	0.75
	y	0.75	0.85	*	0.75	1

Many possible approaches to address this problem in practice - we consider two.

Insurance Application - Toy Motivation

Many possible approaches to address this problem in practice - we consider two.

- Approach 1: (most common approach in practice)
    - STEP 1: Obtain an approximate correlation matrix (at least symmetric BUT COMPLETE) via formal statistical methods or heuristic completions....
    - STEP 2: find the nearest correlation matrix under some projection or norm minimization objective.

Insurance Application - Toy Motivation

Many possible approaches to address this problem in practice - we consider two.

- Approach 1: (most common approach in practice)
  - STEP 1: Obtain an approximate correlation matrix (at least symmetric BUT COMPLETE) via formal statistical methods or heuristic completions....
  - STEP 2: find the nearest correlation matrix under some projection or norm minimization objective.
- Approach 2: Find an onerous completion of the matrix explicitly straight away!

Why do we care about correlation in this problem?

- Consider a collection of risks $X_1, \ldots, X_n$ with aggregate risk measure $\varrho[\cdot]$ and individual risk capital denoted by $\varrho_i = \varrho[X_i]$.

Why do we care about correlation in this problem?

- Consider a collection of risks $X_1, \dots, X_n$ with aggregate risk measure $\varrho[\cdot]$ and individual risk capital denoted by $\varrho_i = \varrho[X_i]$.
- If these risks are combined into one business, then the total capital (coherent risk measures) for the business satisfies

$$\varrho[X_1 + \cdots + X_n] \leq \varrho_1 + \cdots + \varrho_n$$

Why do we care about correlation in this problem?

- Consider a collection of risks $X_1, \ldots, X_n$ with aggregate risk measure $\varrho[\cdot]$ and individual risk capital denoted by $\varrho_i = \varrho[X_i]$.
- If these risks are combined into one business, then the total capital (coherent risk measures) for the business satisfies
$$\varrho[X_1 + \cdots + X_n] \leq \varrho_1 + \cdots + \varrho_n$$
- **Dependence between loss processes can cause increases or decreases in aggregate capital!**
  - *Correlation completion methods can strongly affect the outcome*

HERIOT WATT UNIVERSITY

Why do we care about correlation in this problem?

- Consider a collection of risks $X_1, \ldots, X_n$ with aggregate risk measure $\varrho[\cdot]$ and individual risk capital denoted by $\varrho_i = \varrho[X_i]$.
- If these risks are combined into one business, then the total capital (coherent risk measures) for the business satisfies

$$\varrho[X_1 + \cdots + X_n] \leq \varrho_1 + \cdots + \varrho_n$$

- **Dependence between loss processes can cause increases or decreases in aggregate capital!**
  - *Correlation completion methods can strongly affect the outcome*

$\Rightarrow$ regulators and industry require guidance on mathematical best practice to avoid moral hazard in artificial capital reduction!

Brief Examples: Banking and Insurance
Business Lines and Risk Types

Application Context

- Correlations likely to be missing in areas of risk management and insurance where data and loss event history is scarce $\Rightarrow$ **large gaps in the data records**:
    - in operational risk,
    - reinsurance,
    - catastrophe insurance,
    - life insurance, and
    - cyber risk.

Application Context

- Correlations likely to be missing in areas of risk management and insurance where data and loss event history is scarce $\Rightarrow$ **large gaps in the data records**:
  - in operational risk,
  - reinsurance,
  - catastrophe insurance,
  - life insurance, and
  - cyber risk.
- The estimation of missing correlations is also important in banking capital calculations
  - Example in the internal model-based approach to market risk and the advanced measurement approach (AMA) and (SMA) for operational risk.

Application Context

- Correlations likely to be missing in areas of risk management and insurance where data and loss event history is scarce $\Rightarrow$ **large gaps in the data records**:
  - in operational risk,
  - reinsurance,
  - catastrophe insurance,
  - life insurance, and
  - cyber risk.
- The estimation of missing correlations is also important in banking capital calculations
  - Example in the internal model-based approach to market risk and the advanced measurement approach (AMA) and (SMA) for operational risk.
- Other important applications include correlation effects in **stress testing** and **scenario analysis**

Banking Motivations from Operational Risk

Banking Risk: Business Lines and Risk Types

Banking has three main risk classes:

- Market,
- Credit and
- Operational Risk
  - *Operational Risk is evolving, by loss events incurred, to be the leading risk type in banking out of the three core risks!*

Banking has three main risk classes:

- Market,
- Credit and
- Operational Risk
    - *Operational Risk is evolving, by loss events incurred, to be the leading risk type in banking out of the three core risks!*

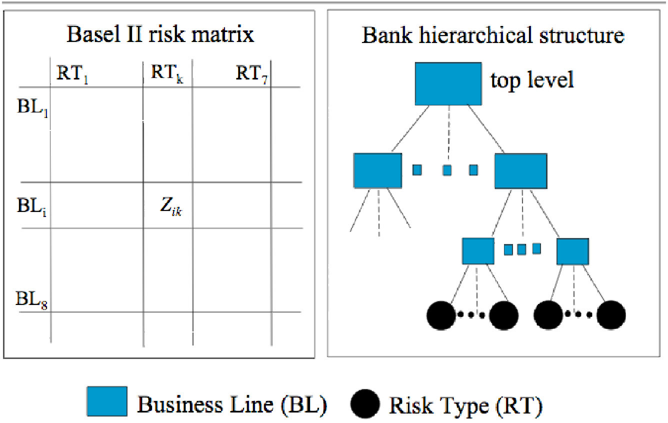
In Operational Risk

- At level 1: Basel II/III requires 56 business unit/risk type loss processes.
- At level 2 and greater granularity: this can reach 100's to 1000's of BuRT cells in practice.

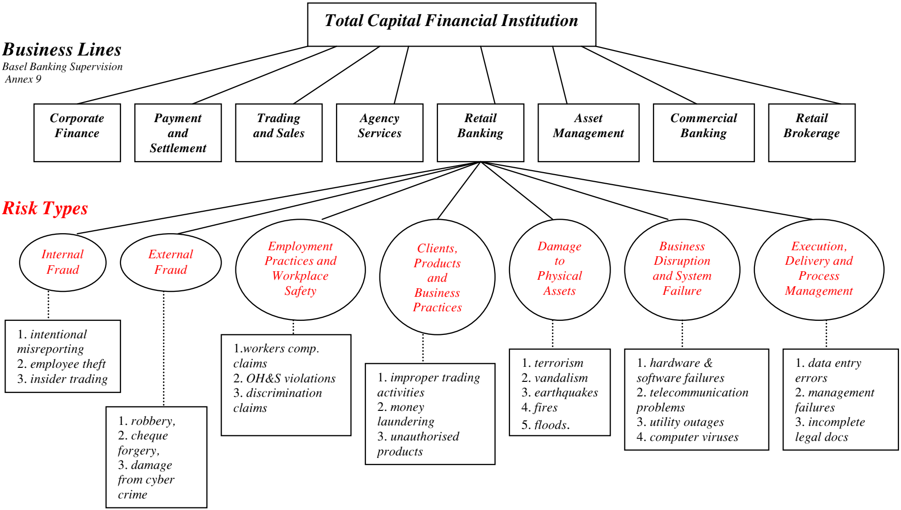
Many unknown/missing correlations present!

HERIOT WATT UNIVERSITY

- Advanced Measurement Approaches (AMA): Internal model for 56 risk cells (7 event types × 8 business lines).



Banking Risk: Business Lines and Risk Types



Insurance Motivations from Solvency II Examples

Lines of Insurance: *one can generally classify insurance companies by the type of insurance policies they write.*

Lines of Insurance: *one can generally classify insurance companies by the type of insurance policies they write.*

- Insurance coverages are often broken down via lines of insurance.

Lines of Insurance: *one can generally classify insurance companies by the type of insurance policies they write.*

- Insurance coverages are often broken down via lines of insurance.
- Information about premiums and losses is frequently analyzed by line of insurance at the company level.

Lines of Insurance: *one can generally classify insurance companies by the type of insurance policies they write.*

- Insurance coverages are often broken down via lines of insurance.
- Information about premiums and losses is frequently analyzed by line of insurance at the company level.

Four Major Lines of Insurance:

- **Property**;
- **Casualty**;
- **Life**;
- **Health and Disability**;

Many large companies write all lines of insurance.

Insurance Risk: Business Lines and Risk Types

Each of the four major categories of insurance can be further subdivided into:

- **Personal**
  - *Personal lines are property-casualty coverages that protect an individual or family.*

Insurance Risk: Business Lines and Risk Types

Each of the four major categories of insurance can be further subdivided into:

- **Personal**
  - *Personal lines are property-casualty coverages that protect an individual or family.*
- **Commercial**
  - *Commercial lines are coverages designed for businesses.*

Insurance Risk: Business Lines and Risk Types

Each of the four major categories of insurance can be further subdivided into:

- **Personal**
  - *Personal lines are property-casualty coverages that protect an individual or family.*
- **Commercial**
  - *Commercial lines are coverages designed for businesses.*

e.g. of commercial lines of business

- professional indemnity;
- product liability;
- political risk;
- financial institutions;
- commercial auto insurance;
- workers compensation insurance;
- federal flood insurance;
- aircraft insurance;
- ocean marine insurance;

Why might there be missing or uncertain correlations between these commercial lines of business and the risk types they are exposed to ?

Insurance Risk: Business Lines and Risk Types

Why might there be missing or uncertain correlations between these commercial lines of business and the risk types they are exposed to ?

- Reason 1: Contract Writing Specificities!
- Reason 2: Specialty & Nature of Insured Risk!

Insurance Risk: Business Lines and Risk Types

Why might there be missing or uncertain correlations between these commercial lines of business and the risk types they are exposed to ?

- Reason 1: Contract Writing Specificities!
- Reason 2: Specialty & Nature of Insured Risk!

Example: Reason 1 Most if not all commercial lines share certain similarities, however it is not unusual that each policy will be tailored for the type of business being covered and the clients unique needs!

Why might there be missing or uncertain correlations between these commercial lines of business and the risk types they are exposed to ?

- Reason 1: Contract Writing Specificities!
- Reason 2: Specialty & Nature of Insured Risk!

Example: Reason 1 Most if not all commercial lines share certain similarities, however it is not unusual that each policy will be tailored for the type of business being covered and the clients unique needs!

- e.g. *structural engineering firm takes professional liability insurance to protect against claims of:*
  - *negligence in creating a buildings plans, performing inspections, and supervising construction, (project specific risks)*
  - *failure to render professional services.*
  - *specific additional coverage for each project, plus coverage for punitive damages can be added on a general cover.*

Example: Reason 2 *Specialty Types of Commercial Lines Insurance*:

- Debris Removal Insurance: *removing debris after a catastrophic events e.g. fires.*
- Builder's risk insurance: *insures buildings while they are being constructed.*
- Glass Insurance: *covers broken windows in a commercial establishment.*
- Business Interruption Insurance: *lost income and expenses resulting from property damage or loss. e.g. fire forces closure for few months, this insurance covers salaries, taxes, rents, and net profits that would have been earned.*

Very challenging to assess / estimate correlations between loss processes in such specialty risk classes!

Application Context

A more general Insurance example arising in the context of Solvency II according to the PRA:

> *The PRA expects firms using an approved internal or partial internal model to calculate their SCR to report the internal model outputs via XBRL using the relevant templates and technical architecture documentation provided in the Appendices. The templates capture selected percentiles of the probability distributions for specified variables (eg risk drivers and lines of business) as well as some information* ***(eg correlation factors)*** *relevant for the PRA to monitor internal models. ...*

[Supervisory Statement — SS25/15 Solvency II: Regulatory reporting, internal model outputs October 2018 (Updating July 2018)]

Application Summary

Example: an insurer needs to complete a correlation matrix to integrate different businesses and risks in the context of Solvency II, where a firm has a hybrid partial IM (Internal Model) and SF (standard formula) composed of

- an IM module,
- some complete SF modules, and
- an incomplete SF module (market risk) where one or more of the submodules have been modelled internally.

The correlations between the SF elements (grey cells) are specified by regulations, and the firm has calculated some coefficients (white cells) but needs to complete the green entries according to a prescribed integration techniques.

Table 2. Example of partial internal model Integration Technique 2, where one of the constituents of the standard formula (SF) market risk module (currency risk) has been included in the IM, so correlations are required between the SF market risk submodules and the other SF modules (that is, the green starred cells).

module	*Submodule*										
IM		1	0.25	0.6	0.55	0.65	0	0.4	0.6	0.2	0.3
SF market risk	*Interest rate*	0.25	1	0	0	0	0	*	*	*	*
	Equity	0.6	0	1	0.75	0.75	0	*	*	*	*
	Property	0.55	0	0.75	1	0.5	0	*	*	*	*
	Spread	0.65	0	0.75	0.5	1	0	*	*	*	*
	Concentration	0	0	0	0	0	1	*	*	*	*
SF default		0.4	*	*	*	*	*	1	0.25	0.25	0.5
SF life		0.6	*	*	*	*	*	0.25	1	0.25	0
SF health		0.2	*	*	*	*	*	0.25	0.25	1	0
SF non-life		0.3	*	*	*	*	*	0.5	0	0	1

One of the prescribed integration techniques for completing the missing entries requires two steps:

- first, determining appropriate upper and lower bounds (based on the firm's risk profile) for the missing correlations and
- second, an optimization step to find the completion such that no other set of correlation coefficients results in a higher SCR, while keeping the matrix positive semi-definite

[Solvency II Delegated Regulation ((EU) 2015/35) Annex XVIII(C) [23]] ... also known as Integration Technique 2, IT2.

Application Summary

In banking and insurance applications there are many business units with many BU-specific risks as well as different numbers of risk families

- Correlation matrices for each BuRT can have hundreds of columns!
- Many of the correlations between diverse BuRT's are completely unknown!

We want to complete the partial correlation matrix $\bar{\Sigma}$ to a fully specified correlation matrix; that is, since the diagonal is fully specified as ones, to a positive definite matrix.

- Many completions are possible $\Rightarrow$ *introduces uncertainty around the range of potential capital outcomes*!
- Completion of most interest is usually a best-estimate completion in some sense.
- *A good candidate is that completion which has maximum determinant*!

Table of contents

- Correlation matrices are real, symmetric positive semidefinite matrices with ones on the diagonal....

- Often in practice one encounters ...
  a matrix that is supposed to be a correlation matrix but for a variety of possible reasons is not...

  e.g. missing data in some of the records...

CLASSIC PROBLEM: *So if we have somehow obtained a complete approximate correlation matrix that does not satisfy symmetry or positive semidefiniteness how do we obtain a correlation matrix... ?*

Nearest Correlation Matrix Problems

One solution is to compute the
"NEAREST CORRELATION MATRIX":

Nearest Correlation Matrix Problems

One solution is to compute the
"NEAREST CORRELATION MATRIX":

i.e. find the nearest correlation matrix X to a given symmetric matrix $A \in \mathbb{R}^{n \times n}$ that is the solution to

$$\min \left\{ \frac{1}{2} ||A - X||_F^2 : X = X^T, \; X \geq 0, \; Diag(X) = e \right\}$$

where for symmetric matrices X and Y:

- $X \geq Y$ here denotes that $X - Y$ is positive semidefinite,
- $Diag(X)$ is the vector of diagonal elements of X,
- e is the vector of ones, and
- $||X||_F = trace(X^T X)^{1/2}$ is the Frobenius norm.

NOTE: since $\mathbb{R}^{n \times n}$ is a Hilbert space with inner product $\langle X, Y \rangle = trace(X^T Y)$ and the constraints are closed convex sets $\Rightarrow$ this optimization program admits a unique solution, see e.g. [Deutsch, 2001]

Nearest Correlation Matrix Problems

NOTE: since $\mathbb{R}^{n \times n}$ is a Hilbert space with inner product $\langle X, Y \rangle = trace(X^T Y)$ and the constraints are closed convex sets $\Rightarrow$ this optimization program admits a unique solution, see e.g. [Deutsch, 2001]

We still need to find a solution...NOT CLOSED FORM

Nearest Correlation Matrix Problems

NOTE: since $\mathbb{R}^{n \times n}$ is a Hilbert space with inner product $\langle X, Y \rangle = trace(X^T Y)$ and the constraints are closed convex sets $\Rightarrow$ this optimization program admits a unique solution, see e.g. [Deutsch, 2001]

We still need to find a solution...NOT CLOSED FORM

- Introducing auxiliary variables one could reformulate this problem as a semidefinite program or second order cone program.

NOTE: since $\mathbb{R}^{n \times n}$ is a Hilbert space with inner product $\langle X, Y \rangle = trace(X^T Y)$ and the constraints are closed convex sets $\Rightarrow$ this optimization program admits a unique solution, see e.g. [Deutsch, 2001]

We still need to find a solution...NOT CLOSED FORM

- Introducing auxiliary variables one could reformulate this problem as a semidefinite program or second order cone program.

$\Rightarrow$ interior point algorithm solutions can be attempted!

Nearest Correlation Matrix Problems

NOTE: since $\mathbb{R}^{n\times n}$ is a Hilbert space with inner product $\langle X, Y\rangle = trace(X^T Y)$ and the constraints are closed convex sets $\Rightarrow$ this optimization program admits a unique solution, see e.g. [Deutsch, 2001]

We still need to find a solution...NOT CLOSED FORM

- Introducing auxiliary variables one could reformulate this problem as a semidefinite program or second order cone program.

  $\Rightarrow$ interior point algorithm solutions can be attempted!

  However: if n is large direct interior point methods struggle!

Numerous works propose numerical solutions via projections:

> *A lot of the literature is concerned with ad hoc methods that are not guaranteed to solve the problem!*

Numerous works propose numerical solutions via projections:

> *A lot of the literature is concerned with ad hoc methods that are not guaranteed to solve the problem!*

Can try alternating projection approaches of [Dykstra, 1983] & [Higham, 2002] such as:

- [Knol and ten Berge, 1989] write $X = Y^T Y$ and iteratively minimize the objective function over each unit 2-norm column of Y .
- [Lurie and Goldberg, 1998] adopt a GaussNewton method to minimize $||A - R^T R||_F^2$, where matrix R is upper triangular with columns of unit 2-norm.

Nearest Correlation Matrix Problems

HERIOT WATT UNIVERSITY

Numerous works propose numerical solutions via projections:

> *A lot of the literature is concerned with ad hoc methods that are not guaranteed to solve the problem!*

Can try alternating projection approaches of [Dykstra, 1983] & [Higham, 2002] such as:

- [Knol and ten Berge, 1989] write $X = Y^T Y$ and iteratively minimize the objective function over each unit 2-norm column of Y .
- [Lurie and Goldberg, 1998] adopt a GaussNewton method to minimize $||A - R^T R||_F^2$, where matrix R is upper triangular with columns of unit 2-norm.

If they converge: these type of alternating projections converge at best linearly!

[Qi and Sun,2006] recognized this problem as a special case of the following convex optimization problem

$$\min \frac{1}{2}||x^0 - x||^2$$
$$s.t. \ \mathcal{A}x = b$$
$$x \in K$$

where

- $K \subseteq \mathcal{X}$ is a closed convex subset in a Hilbert space $\mathcal{X}$
- Space $\mathcal{X}$ is endowed with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\|\cdot\|$
- $\mathcal{A} : \mathcal{X} \mapsto \mathbb{R}^n$ is a bounded linear operator,
- $b \in \mathbb{R}^n$ and
- $x^0 \in \mathcal{X}$ are given data.

Nearest Correlation Matrix Problems

[Qi and Sun,2006] recognized this problem as a special case of the following convex optimization problem

$$\min \frac{1}{2}||x^0 - x||^2$$
$$s.t.\ \ \mathcal{A}x = b$$
$$x \in K$$

where

- $K \subseteq \mathcal{X}$ is a closed convex subset in a Hilbert space $\mathcal{X}$
- Space $\mathcal{X}$ is endowed with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\|\cdot\|$
- $\mathcal{A} : \mathcal{X} \mapsto \mathbb{R}^n$ is a bounded linear operator,
- $b \in \mathbb{R}^n$ and
- $x^0 \in \mathcal{X}$ are given data.

This problem is also known as the best approximation from a closed convex set in a Hilbert space.

Nearest Correlation Matrix Problems

Under this general formulation of [Qi and Sun,2006] one has the connection to the original problem by setting:

- $\mathcal{X} = \mathcal{S}^n$ the space of symmetric $n \times n$ matrices with real values.;
- $K = \mathcal{S}^n_+$ the cone of $n \times n$ positive semi-definite matrices with real values in $\mathcal{S}^n$.;
- $b = e$ the vector of ones.;
- $x^0 = A$; and
- $\mathcal{A}X = diag(X)$ vector of all diagonal elements of $X \in \mathcal{S}^n$.

Nearest Correlation Matrix Problems

[Deutsch, Li and Ward,1997] demonstrated that the unique solution x^* of this generalised problem statement has representation:

$$x^* = \Pi_K(x^0 + \mathcal{A}^* y^*)$$

with

- $\Pi_K(\cdot)$ the metric projection operator onto K under inner product $\langle \cdot, \cdot \rangle$;
- $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$; and
- y^* is the solution of equation

$$\mathcal{A}\Pi_K(x^0 + \mathcal{A}^* y) = b,$$

This representation holds: iff the STRONG Canonical Hull Intersection Property (CHIP) holds for the set $\{K, \mathcal{A}^{-1}(b)\}$ ... see [Deutsch, Li and Ward,1997]

- REMARK: it can be challenging to verify in general!

Alternatively a sufficient condition for this solution form can be more readily verified and is given by the condition

$$b \in ri\,(\mathcal{A}(K))$$

where

- $\mathcal{A}(K)$ is the data cone when K is a cone in $\mathcal{X}$; and
- $ri(\cdot)$ is the relative interior.

Recall: the relative interior of a set S is defined as its interior within the affine hull of S.

$$ri(S) := \{x \in S : \exists \epsilon > 0, N_\epsilon(x) \cap aff(S) \subseteq S\},$$

where

- $aff(S)$ is the affine hull of S, and
- $N_\epsilon(x)$ is a ball of radius ϵ centered on x.
    - Any metric can be used for the construction of the ball; all metrics define the same set as the relative interior.

Nearest Correlation Matrix Problems

HERIOT WATT UNIVERSITY

Applying this condition a provably quadratically convergent solution was developed in [Qi and Sun,2006]

- To achieve this they dualized the linear constraints in the nearest correlation matrix problem... to produce UNCONSTRAINED convex optimization problem:

Applying this condition a provably quadratically convergent solution was developed in [Qi and Sun,2006]

- To achieve this they dualized the linear constraints in the nearest correlation matrix problem... to produce UNCONSTRAINED convex optimization problem:

$$\min_{y \in \mathbb{R}^n} f(y) := \frac{1}{2}||(A + diag(y))_+||_F^2 - e^T y$$

where

- $diag(y)$ for $y \in \mathbb{R}^n$ denotes the diagonal matrix whose elements are those of the vector y
- $diag(A)$ for $A \in \mathbb{R}^{n \times n}$ denotes $diag([a_{11}, a_{12}, \ldots, a_{nn}])$
- $(\cdot)_+$ is an operator projecting onto positive semidefinite matrices: for symmetric $C \in \mathbb{R}^{n \times n}$ with spectral decomposition $C = Q \Lambda Q^T$ ($Q^T Q = I$, $\Lambda = diag(\lambda_i)$)... then

$$(C)_+ = Q \; diag(max(\lambda_i, 0)) \; Q^T$$

is nearest positive semidefinite matrix to C in Frobenius norm.

Nearest Correlation Matrix Problems

[Malick, 2004] proved the following Lemma which then allowed [Qi and Sun, 2006] to show that a gradient based solution for the dual problems results and it achieves **quadratic convergence**...

Nearest Correlation Matrix Problems

[Malick, 2004] proved the following Lemma which then allowed [Qi and Sun, 2006] to show that a gradient based solution for the dual problems results and it achieves **quadratic convergence**...

Lemma

The dual problem has the properties:

- *f is convex continuously differentiable and has a unique minimizer...*
- *the gradient ∇f is given by*

$$\nabla f(y) = Diag\,((A + diag(y))_+) - e$$

  and is Lipschitz continuous with Lipschitz constant 1.
- *the solutions y_* of the dual problem and X_* of the primal problem are related by*

$$X_* = (A + diag(y_*))_+$$

Nearest Correlation Matrix Problems

[Malick, 2004] proved the following Lemma which then allowed [Qi and Sun, 2006] to show that a gradient based solution for the dual problems results and it achieves **quadratic convergence**...

Lemma

The dual problem has the properties:

- *f is convex continuously differentiable and has a unique minimizer...*
- *the gradient ∇f is given by*

$$\nabla f(y) = Diag\left((A + diag(y))_+\right) - e$$

  and is Lipschitz continuous with Lipschitz constant 1.

- *the solutions y_* of the dual problem and X_* of the primal problem are related by*

$$X_* = (A + diag(y_*))_+$$

- One can interpret this result as showing that the original constrained problem with $(n^2 - n)/2$ variables is equivalent to an unconstrained problem with just n variables

Nearest Correlation Matrix Problems

[Malick, 2004] proved the following Lemma which then allowed [Qi and Sun, 2006] to show that a gradient based solution for the dual problems results and it achieves **quadratic convergence**...

Lemma

The dual problem has the properties:

- *f is convex continuously differentiable and has a unique minimizer...*
- *the gradient ∇f is given by*

$$\nabla f(y) = Diag\left((A + diag(y))_+\right) - e$$

  and is Lipschitz continuous with Lipschitz constant 1.
- *the solutions y_* of the dual problem and X_* of the primal problem are related by*

$$X_* = (A + diag(y_*))_+$$

- One can interpret this result as showing that the original constrained problem with $(n^2 - n)/2$ variables is equivalent to an unconstrained problem with just n variables

A quadratically convergent solution is developed....

Overview

Approach 2: In this work we consider an alternative related problem....
not exactly falling in category of problems explained above...

Approach 2: In this work we consider an alternative related problem....
not exactly falling in category of problems explained above...

We are concerned with problems in which the **missing values are in the correlation matrix itself.**

- Some of the matrix entries are known, having been:
  - *estimated*;
  - *prescribed by regulations*; or
  - *assigned by expert judgement*,

  however, the **other entries are unknown!**

Approach 2: In this work we consider an alternative related problem....
not exactly falling in category of problems explained above...

We are concerned with problems in which the **missing values are in the correlation matrix itself.**

- Some of the matrix entries are known, having been:
  - *estimated*;
  - *prescribed by regulations*; or
  - *assigned by expert judgement*,

  however, the **other entries are unknown!**

Nearest correlation matrix solutions will preclude this important case as the projections required distort all elements!

Overview

- The aim is to fill in the missing entries in order to produce a correlation matrix
- **Of course there are, in general, many possible completions!!!**

For example, the partially specified matrix

$$A = \begin{bmatrix} 1 & a_{12} \\ a_{12} & 1 \end{bmatrix}$$

is a correlation matrix for any a_{12} such that $|a_{12}| \leq 1$.

HERIOT WATT UNIVERSITY

- The aim is to fill in the missing entries in order to produce a correlation matrix
- **Of course there are, in general, many possible completions!!!**

For example, the partially specified matrix

$$A = \begin{bmatrix} 1 & a_{12} \\ a_{12} & 1 \end{bmatrix}$$

is a correlation matrix for any a_{12} such that $|a_{12}| \leq 1$.

- **Our focus is on the completion with maximal determinant.**
  - given by $a_{12} = 0$ in this example.

It is always unique when completions exist!

Table of contents

Correlation Completion via MaxDet

MaxDet has several useful theoretical properties:

1. ***Existence and uniqueness***: if positive semidefinite completions exist then there is exactly **one MaxDet completion** [Grone et al., 84].

Correlation Completion via MaxDet

MaxDet has several useful theoretical properties:

1. ***Existence and uniqueness***: if positive semidefinite completions exist then there is exactly **one MaxDet completion** [Grone et al., 84].
2. ***Maximum entropy model***: **MaxDet is the maximum entropy completion for the multivariate normal model**, where maximum entropy is a principle of favouring the simplest explanations [Good,63].
   Aligned with Solvency II Standard Formula

Correlation Completion via MaxDet

HERIOT WATT UNIVERSITY

MaxDet has several useful theoretical properties:

1. ***Existence and uniqueness***: if positive semidefinite completions exist then there is exactly **one MaxDet completion** [Grone et al., 84].
2. ***Maximum entropy model***: **MaxDet is the maximum entropy completion for the multivariate normal model**, where maximum entropy is a principle of favouring the simplest explanations [Good,63].
   Aligned with Solvency II Standard Formula
3. ***Maximum likelihood estimation***: **MaxDet is the maximum likelihood estimate of the correlation matrix** of the unknown underlying multivariate normal model.

Correlation Completion via MaxDet

MaxDet has several useful theoretical properties:

1. ***Existence and uniqueness***: if positive semidefinite completions exist then there is exactly **one MaxDet completion** [Grone et al., 84].
2. ***Maximum entropy model***: **MaxDet is the maximum entropy completion for the multivariate normal model**, where maximum entropy is a principle of favouring the simplest explanations [Good,63].
   Aligned with Solvency II Standard Formula
3. ***Maximum likelihood estimation***: **MaxDet is the maximum likelihood estimate of the correlation matrix** of the unknown underlying multivariate normal model.
4. ***Analytic center***: **MaxDet is the analytic centre of the feasible region described by the positive semidefiniteness constraints**, where the analytic centre is defined as the point that maximizes the product of distances to the defining hyperplanes [Vandenberghe et al, 98].

Correlation Completion via MaxDet

When considering MaxDet solutions for a correlation matrix:

- we have a simple upper bound; and
- sufficient conditions for existance and uniqueness

Correlation Completion via MaxDet

When considering MaxDet solutions for a correlation matrix:

- we have a simple upper bound; and
- sufficient conditions for existance and uniqueness

Upper Bound:

- The determinant of a correlation matrix is at most 1 via Hadamard's inequality.

Let matrix $A = [a_{ij}]$ be an $n \times n$ positive definite matrix. Then:

$$\det A \leq a_{11} \cdots a_{nn}$$

with equality iff A is diagonal.

Correlation Completion via MaxDet

When considering MaxDet solutions for a correlation matrix:

- we have a simple upper bound; and
- sufficient conditions for existance and uniqueness

Upper Bound:

- The determinant of a correlation matrix is at most 1 via Hadamard's inequality.

Let matrix $A = [a_{ij}]$ be an $n \times n$ positive definite matrix. Then:

$$\det A \leq a_{11} \cdots a_{nn}$$

with equality iff A is diagonal.

Conditions for Existence: [Grone et al., 84] showed a partially specified Hermitian matrix with specified positive diagonal entries and positive principal minors (where specified) can be completed to a positive definite matrix regardless of the values of the entries ...

Correlation Completion via MaxDet

When considering MaxDet solutions for a correlation matrix:

- we have a simple upper bound; and
- sufficient conditions for existance and uniqueness

Upper Bound:

- The determinant of a correlation matrix is at most 1 via Hadamard's inequality.

Let matrix $A = [a_{ij}]$ be an $n \times n$ positive definite matrix. Then:

$$\det A \leq a_{11} \cdots a_{nn}$$

with equality iff A is diagonal.

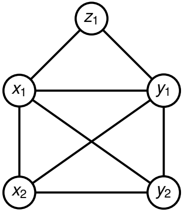
Conditions for Existence: [Grone et al., 84] showed a partially specified Hermitian matrix with specified positive diagonal entries and positive principal minors (where specified) can be completed to a positive definite matrix regardless of the values of the entries ...

- *iff the undirected graph of the specified entries (ignoring the leading diagonal) is chordal.*

Correlation Completion via MaxDet

- *A graph is chordal if every cycle of length ≥ 4 has a chord, which is an edge that is not part of the cycle but connects two vertices of the cycle.*
- If the graph is not chordal, then whether a positive symmetric definite completion exists depends on the specified entries.
- All the sparsity patterns considered in this work are chordal $\Rightarrow$ a positive definite symmetric completion is possible!!!

Adjacency graph for the case in previous 2 BU example below:



[Grone et al. 84] showed that if a positive definite completion exists then there is a unique matrix in the class of all positive definite completions whose determinant is maximal.

Dealing with large matrices with block patterns of specified and unspecified entries, it is convenient to introduce the definition of a "block chordal" graph.

Block Chordal Graphs:

- A block is a subgraph which is complete in terms of edges (a clique).
- Two blocks are connected by an edge if every vertex has an edge to every other vertex, so the two blocks considered together also form a clique.
- A graph is block chordal if every cycle of blocks of length ≥ 4 has a chord.
- Finally, a block chordal graph is also chordal since every block is either fully specified or fully unspecified, so collapsing each block into one node means that we do not lose any information in the graph.

Table of contents

[Dempster, 72] proposed a related problem known as **covariance selection**

MaxDet and the Dual Covariance Selection

[Dempster, 72] proposed a related problem known as **covariance selection**

Covariance selection: aims to simplify the covariance structure of a multivariate normal population by ***setting elements of the inverse of the covariance matrix to zero***.

- The statistical interpretation is that ***certain variables are set to be pairwise conditionally independent.***

MaxDet and the Dual Covariance Selection

[Dempster, 72] proposed a related problem known as **covariance selection**

Covariance selection: aims to simplify the covariance structure of a multivariate normal population by ***setting elements of the inverse of the covariance matrix to zero***.

- The statistical interpretation is that ***certain variables are set to be pairwise conditionally independent.***

[Dahl et al, 08] and [Vandenberghe et al, 98] show that **MaxDet completion and covariance selection are duals!**

MaxDet and the Dual Formulation

Another way to see that a determinant-maximizing completion must have zeros in the inverse corresponding to the free elements of Σ is by a perturbation argument.

We need the following lemma [Chan, 84].

Lemma

For $v, w, x, y \in \mathbb{R}^n$,

$$\det(I + vx^T + wy^T) = (1 + v^T x)(1 + w^T y) - (v^T y)(w^T x).$$

Using the lemma, we consider how the determinant of a symmetric positive definite matrix $A \in \mathbb{R}^{n \times n}$ changes when we perturb a_{ij} (and a_{ji}, by symmetry).

Perturbed Matrix: let

$$A(\epsilon) = A + \epsilon(e_i e_j^T + e_j e_i^T),$$

where e_i is the ith column of the identity matrix.

MaxDet and the Dual Formulation

Let $B = A^{-1}$ and partition $B = [b_1, \ldots, b_n]$. Apply the lemma:

$$
\begin{aligned}
\det A(\epsilon) &= \det(A(I + \epsilon(b_i e_j^T + b_j e_i^T))) \\
&= \det(A) \det(I + \epsilon(b_i e_j^T + b_j e_i^T)) \\
&= \det(A) \big[(1 + \epsilon b_i^T e_j)(1 + \epsilon b_j^T e_i) - \epsilon^2 (b_i^T e_i)(b_j^T e_j)\big] \\
&= \det(A) \big[(1 + \epsilon b_{ji})(1 + \epsilon b_{ij}) - \epsilon^2 b_{ii} b_{jj}\big] \\
&= \det(A) \big(1 + 2\epsilon b_{ij} + \epsilon^2 (b_{ij}^2 - b_{ii} b_{jj})\big).
\end{aligned}
$$

We want to know when $\det A(0)$ is maximal. Since

$$\frac{d}{d\epsilon} \det A(\epsilon)|_{\epsilon=0} = 2 \det(A) b_{ij},$$

we need $b_{ij} = 0$ for a stationary point at $\epsilon = 0$, and from

$$\frac{d^2}{d\epsilon^2} \det A(\epsilon)|_{\epsilon=0} = 2 \det(A)(b_{ij}^2 - b_{ii} b_{jj}) < 0$$

(since B is positive definite), we see that when $b_{ij} = 0$, the quadratic function $\det A()$ has a maximum at $= 0$.

Table of contents

MaxDet and the Dual Covariance Selection

In general, solving the MaxDet completion problem (or, equivalently, *the covariance selection problem*) requires solving a convex optimization problem on the set of positive definite matrices [Dahl et al, 08].

- **We develop explicit, easily implementable solutions for some practically important cases arising in the insurance application.**

Let Σ denote the solution of the MaxDet completion problem for the partially-specified correlation matrix $\bar{\Sigma}$.

We give a result for an L-shaped pattern of unspecified entries in $\bar{\Sigma}$.

> *Note that we do not require a unit diagonal, so it applies more generally than just to correlation matrices.*

MaxDet and the Dual Covariance Selection

HERIOT WATT UNIVERSITY

Theorem

Consider the symmetric partially specified matrix

$$\bar{\Sigma} = \begin{array}{c} \\ n_1 \\ n_2 \\ n_3 \\ n_4 \end{array} \begin{array}{c} \begin{array}{cccc} n_1 & n_2 & n_3 & n_4 \end{array} \\ \left[\begin{array}{cccc} A_{11} & B & C & D \\ B^T & A_{22} & E & F \\ C^T & E^T & A_{33} & G \\ D^T & F^T & G^T & A_{44} \end{array} \right] \end{array} \in \mathbb{R}^{(n_1+n_2+n_3+n_4)\times(n_1+n_2+n_3+n_4)},$$

where C, E, and F are unspecified, the diagonal blocks A_{ii}, $i = 1:4$ are all positive definite, and all specified principal minors are positive.

The maximal determinant completion is

$$C = DA_{44}^{-1}G^T, \quad F = B^TA_{11}^{-1}D, \quad E = FA_{44}^{-1}G^T.$$

Brief comments on Proof:

- The result can be derived by permuting $\bar{\Sigma}$ so that the unspecified matrices appear in the block (1,3), (1,4), and (2,4) positions and then applying the results of [Dym et al, 81] on completion of block banded matrices.
- The result can also be obtained from [Johnson and Lundquist, 93], in which the unspecified elements of the MaxDet completion are given elementwise in terms of the clique paths in the graph of the specified elements.

Alternatively, we develop an elementary proof based on Gaussian elimination, using the property that Σ^{-1} will contain zeros in the positions of the unspecified entries in $\bar{\Sigma}$.

- It is easy to check that the graph of the specified entries is block chordal, and therefore a unique determinant maximizing positive definite completion exists!

To find it, we need to solve the linear system

$$\begin{bmatrix} A_{11} & B & C & D \\ B^T & A_{22} & E & F \\ C^T & E^T & A_{33} & G \\ D^T & F^T & G^T & A_{44} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \Gamma_3 \\ \Gamma_4 \end{bmatrix},$$

that is,

$$A_{11}X_1 + BX_2 + CX_3 + DX_4 = \Gamma_1, \quad (1)$$

$$B^TX_1 + A_{22}X_2 + EX_3 + FX_4 = \Gamma_2, \quad (2)$$

$$C^TX_1 + E^TX_2 + A_{33}X_3 + GX_4 = \Gamma_3, \quad (3)$$

$$D^TX_1 + F^TX_2 + G^TX_3 + A_{44}X_4 = \Gamma_4, \quad (4)$$

by Gaussian elimination in order to identify the inverse of the matrix $\bar{\Sigma}$. *...steps of the solution in our paper...*

Computation Efficiency and Accuracy

Accuracy and Efficiency of Computation: should be evaluated as follows, avoiding explicit computation of matrix inverses.

Compute Cholesky factorizations:
(decomposed into lower triangular matrix R)

- $A_{11} = R_{11}^T R_{11}$ and
- $A_{44} = R_{44}^T R_{44}$,

then evaluate unspecified matrix components according to:

$$C = (DR_{44}^{-1})(R_{44}^{-T} G^T), \quad F = (B^T R_{11}^{-1})(R_{11}^{-T} D), \quad E = (FR_{44}^{-1})(R_{44}^{-T} G^T).$$

Each of the terms in parentheses should be evaluated as the solution of a triangular linear system with multiple right handsides.

- The term $R_{44}^{-T} G^T$ can be calculated once and reused.

We identify two useful special cases.

Corollary

Consider the symmetric matrix

$$\begin{array}{c} \\ n_1 \\ n_2 \\ n_3 \end{array} \begin{array}{c} \begin{array}{ccc} n_1 & n_2 & n_3 \end{array} \\ \left[\begin{array}{ccc} A_{11} & B & C \\ B^T & A_{22} & E \\ C^T & E^T & A_{33} \end{array} \right] \end{array} \in \mathbb{R}^{(n_1+n_2+n_3)\times(n_1+n_2+n_3)},$$

where E is unspecified, all the diagonal blocks are positive definite, and all specified principal minors are positive.

The maximal determinant completion is $E = B^T A_{11}^{-1} C$.

HERIOT WATT UNIVERSITY

Corollary

Consider the symmetric matrix

$$\begin{array}{c c} & \begin{array}{ccc} n_1 & n_2 & n_3 \end{array} \\ \begin{array}{c} n_1 \\ n_2 \\ n_3 \end{array} & \left[\begin{array}{ccc} A_{11} & B & C \\ B^T & A_{22} & E \\ C^T & E^T & A_{33} \end{array}\right] \end{array} \in \mathbb{R}^{(n_1+n_2+n_3)\times(n_1+n_2+n_3)},$$

where C is unspecified, all the diagonal blocks are positive definite, and all specified principal minors are positive.

The maximal determinant completion is $C = BA_{22}^{-1}E$.

MaxDet and the Dual Covariance Selection

Now we consider a pattern of unspecified elements that arises when (for example) an insurance company has four business units where correlations between BU-specific risks are known

- described by the specified blocks A_{11}, A_{22}, A_{33} and A_{44} and
- all the correlations are known for the first group of risks (for example, risk drivers such as interest rates or exchange rates).

Now we consider a pattern of unspecified elements that arises when (for example) an insurance company has four business units where correlations between BU-specific risks are known

- described by the specified blocks A_{11}, A_{22}, A_{33} and A_{44} and
- all the correlations are known for the first group of risks (for example, risk drivers such as interest rates or exchange rates).

So here we have a complete first block row and column
$\Rightarrow$ ***this case cannot be obtained by permuting rows and columns in the previous Theorem***.

MaxDet and the Dual Covariance Selection

HERIOT WATT UNIVERSITY

Theorem

Consider the symmetric matrix

$$\begin{array}{c} \\ n_1 \\ n_2 \\ n_3 \\ n_4 \end{array} \begin{array}{c} \begin{array}{cccc} n_1 & n_2 & n_3 & n_4 \end{array} \\ \left[\begin{array}{cccc} \boldsymbol{A_{11}} & \boldsymbol{B} & \boldsymbol{C} & \boldsymbol{D} \\ \boldsymbol{B^T} & A_{22} & E & F \\ \boldsymbol{C^T} & E^T & A_{33} & G \\ \boldsymbol{D^T} & F^T & G^T & A_{44} \end{array}\right] \end{array} \in \mathbb{R}^{(n_1+n_2+n_3+n_4)\times(n_1+n_2+n_3+n_4)},$$

where E, F, and G are unspecified, all the diagonal blocks are positive definite, and all specified principal minors are positive.

The maximal determinant completion of the matrix is

$$E = B^T A_{11}^{-1} C, \quad F = B^T A_{11}^{-1} D, \quad G = C^T A_{11}^{-1} D.$$

Finally, we consider the case where C, E and F are unspecified, and B and G are partly specified.

HERIOT WATT UNIVERSITY

Theorem

Consider the symmetric matrix

$$\bar{\Sigma} = \begin{array}{c} \\ n_1 \\ n_2 \\ n_3 \\ n_4 \end{array} \begin{array}{c} \begin{array}{cccc} n_1 & n_2 & n_3 & n_4 \end{array} \\ \begin{bmatrix} A_{11} & B & C & D \\ B^T & A_{22} & E & F \\ C^T & E^T & A_{33} & G \\ D^T & F^T & G^T & A_{44} \end{bmatrix} \end{array} \in \mathbb{R}^{(n_1+n_2+n_3+n_4)\times(n_1+n_2+n_3+n_4)},$$

where C, E, and F are unspecified, B and G are partly specified (possibly fully unspecified), all the diagonal blocks are positive definite, all specified principal minors are positive, and the graph of the specified entries is block chordal.

If B and G are fully unspecified then the maximal determinant completion of the matrix is

$$\Sigma = \begin{array}{c} \\ n_1 \\ n_2 \\ n_3 \\ n_4 \end{array} \begin{array}{c} \begin{array}{cccc} n_1 & n_2 & n_3 & n_4 \end{array} \\ \begin{bmatrix} A_{11} & 0 & 0 & D \\ 0 & A_{22} & 0 & 0 \\ 0 & 0 & A_{33} & 0 \\ D^T & 0 & 0 & A_{44} \end{bmatrix} \end{array}. \tag{5}$$

Otherwise, the maximal determinant completion of B and G is independent of the entries in D.

Conclusions

- We have derived explicit solutions for completions with maximal determinant of a wide class of partially specified correlation matrices that arise in the context of insurers calculating economic capital requirements.
  - Further results are contained in paper for other patterns...

Conclusions

- We have derived explicit solutions for completions with maximal determinant of a wide class of partially specified correlation matrices that arise in the context of insurers calculating economic capital requirements.
  - Further results are contained in paper for other patterns...
- The patterns supported are block diagonal, with either cross-shaped or (inverted) L-shaped gaps on the off-diagonal.

Conclusions

- We have derived explicit solutions for completions with maximal determinant of a wide class of partially specified correlation matrices that arise in the context of insurers calculating economic capital requirements.
  - Further results are contained in paper for other patterns...
- The patterns supported are block diagonal, with either cross-shaped or (inverted) L-shaped gaps on the off-diagonal.
- The solutions are easy to evaluate, being expressed in terms of products and inverses of known matrices.

Conclusions

- We have derived explicit solutions for completions with maximal determinant of a wide class of partially specified correlation matrices that arise in the context of insurers calculating economic capital requirements.
  - Further results are contained in paper for other patterns...
- The patterns supported are block diagonal, with either cross-shaped or (inverted) L-shaped gaps on the off-diagonal.
- The solutions are easy to evaluate, being expressed in terms of products and inverses of known matrices.
- Possible directions for future work include developing explicit solutions for more general patterns of unspecified entries and allowing semidefinite diagonal blocks and zero principal minors.

Conclusions

- We have derived explicit solutions for completions with maximal determinant of a wide class of partially specified correlation matrices that arise in the context of insurers calculating economic capital requirements.
  - Further results are contained in paper for other patterns...
- The patterns supported are block diagonal, with either cross-shaped or (inverted) L-shaped gaps on the off-diagonal.
- The solutions are easy to evaluate, being expressed in terms of products and inverses of known matrices.
- Possible directions for future work include developing explicit solutions for more general patterns of unspecified entries and allowing semidefinite diagonal blocks and zero principal minors.

Please see further details in:

Georgescu D. I., Higham N.J. and Peters G.W. (2018)
*"**Explicit Solutions to Correlation Matrix Completion Problems, with an Application to Risk Management and Insurance**"*
Royal Society Open Science. (to appear)

Scottish Financial Risk Academy

Collection of Heriot-Watt University and Edinburgh University in partnership with Scottish Financial Enterprise.

- `https://www.sfrascottishfinancialriskacademy.com/`
- `http://www.sfe.org.uk/`

Table of contents

Table of contents

MaxDet and the Dual Covariance Selection

Brief comments on Proof:

- The result can be derived by permuting $\bar{\Sigma}$ so that the unspecified matrices appear in the block (1,3), (1,4), and (2,4) positions and then applying the results of [Dym et al, 81] on completion of block banded matrices.
- The result can also be obtained from [Johnson and Lundquist, 93], in which the unspecified elements of the MaxDet completion are given elementwise in terms of the clique paths in the graph of the specified elements.

Alternatively, we develop an elementary proof based on Gaussian elimination, using the property that Σ^{-1} will contain zeros in the positions of the unspecified entries in $\bar{\Sigma}$.

- It is easy to check that the graph of the specified entries is block chordal, and therefore a unique determinant maximizing positive definite completion exists!

To find it, we need to solve the linear system

$$\begin{bmatrix} A_{11} & B & C & D \\ B^T & A_{22} & E & F \\ C^T & E^T & A_{33} & G \\ D^T & F^T & G^T & A_{44} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \Gamma_3 \\ \Gamma_4 \end{bmatrix},$$

that is,

$$A_{11}X_1 + BX_2 + CX_3 + DX_4 = \Gamma_1, \quad (6)$$

$$B^TX_1 + A_{22}X_2 + EX_3 + FX_4 = \Gamma_2, \quad (7)$$

$$C^TX_1 + E^TX_2 + A_{33}X_3 + GX_4 = \Gamma_3, \quad (8)$$

$$D^TX_1 + F^TX_2 + G^TX_3 + A_{44}X_4 = \Gamma_4, \quad (9)$$

by Gaussian elimination in order to identify the inverse of the matrix $\bar{\Sigma}$.

Basic Proof Steps

In this system we can think of C, E, and F as representing any positive definite completions, so that the coefficient matrix is positive definite.

- **We find the determinant maximizing completions by enforcing zeros in relevant blocks of the inverse.**

The following patterns arise frequently in the working below so we assign them variable names to condense the formulae:

$$\mathcal{B} = B - DA_{44}^{-1}F^T, \qquad \mathcal{K} = \mathcal{E} - \mathcal{B}^T\Delta\mathcal{C},$$
$$\mathcal{C} = C - DA_{44}^{-1}G^T, \qquad \mathcal{M} = A_{44}^{-1} + A_{44}^{-1}D^T\Delta DA_{44}^{-1},$$
$$\mathcal{E} = E - FA_{44}^{-1}G^T, \qquad \Delta = (A_{11} - DA_{44}^{-1}D^T)^{-1},$$
$$\mathcal{F} = F - \mathcal{B}^T\Delta D, \qquad \Phi = (A_{22} - FA_{44}^{-1}F^T - \mathcal{B}^T\Delta\mathcal{B})^{-1},$$
$$\mathcal{G} = G - \mathcal{C}^T\Delta D, \qquad \Xi = (A_{33} - GA_{44}^{-1}G^T - \mathcal{C}^T\Delta\mathcal{C} - \mathcal{K}^T\Phi\mathcal{K})^{-1}.$$

- *Inverses in definitions of Δ, Φ, and Ξ exist since matrices being inverted are Schur complements arising in block elimination of the positive definite matrix $\bar{\Sigma}$, so are themselves positive definite.*

Basic Proof Steps

We first solve for X_4 in (9), to obtain

$$X_4 = A_{44}^{-1}(\Gamma_4 - D^T X_1 - F^T X_2 - G^T X_3),$$

and substitute this expression into (6) to obtain

$$A_{11}X_1 + BX_2 + CX_3 + DA_{44}^{-1}(\Gamma_4 - D^T X_1 - F^T X_2 - G^T X_3) = \Gamma_1.$$

We can then express X_1 and X_4 in terms of X_2 and X_3 only:

$$\begin{aligned} X_1 &= (A_{11} - DA_{44}^{-1}D^T)^{-1}\Big(\Gamma_1 - DA_{44}^{-1}\Gamma_4 - (B - DA_{44}^{-1}F^T)X_2 - (C - DA_{44}^{-1}G^T)X_3\Big) \\ &= \Delta(\Gamma_1 - DA_{44}^{-1}\Gamma_4 - \mathcal{B}X_2 - \mathcal{C}X_3) \qquad (10) \end{aligned}$$

and

$$\begin{aligned} X_4 &= A_{44}^{-1}\Big(\Gamma_4 - D^T\Delta(\Gamma_1 - DA_{44}^{-1}\Gamma_4 - \mathcal{B}X_2 - \mathcal{C}X_3) - F^T X_2 - G^T X_3\Big) \\ &= A_{44}^{-1}\Big(- D^T\Delta\Gamma_1 + \Gamma_4 + D^T\Delta DA_{44}^{-1}\Gamma_4 - (F^T - D^T\Delta\mathcal{B})X_2 - (G^T - D^T\Delta\mathcal{C})X_3\Big) \\ &= A_{44}^{-1}\Big(- D^T\Delta\Gamma_1 + \Gamma_4 + D^T\Delta DA_{44}^{-1}\Gamma_4 - \mathcal{F}^T X_2 - \mathcal{G}^T X_3\Big). \qquad (11) \end{aligned}$$

Working with (7) next, and separating the X_2 and X_3 variables, we have:

$$\begin{aligned}
A_{22}X_2 &= \Gamma_2 - B^T X_1 - EX_3 - FX_4 \\
&= \Gamma_2 - B^T\Delta(\Gamma_1 - DA_{44}^{-1}\Gamma_4 - \mathcal{B}X_2 - \mathcal{C}X_3) - EX_3 \\
&\quad - FA_{44}^{-1}(-D^T\Delta\Gamma_1 + \Gamma_4 + D^T\Delta DA_{44}^{-1}\Gamma_4 - \mathcal{F}^T X_2 - \mathcal{G}^T X_3) \\
&= -(B^T - FA_{44}^{-1}D^T)\Delta\Gamma_1 + \Gamma_2 - (F - \mathcal{B}^T\Delta D)A_{44}^{-1}\Gamma_4 \\
&\quad + (B^T\Delta\mathcal{B} + FA_{44}^{-1}\mathcal{F}^T)X_2 - \big(E - FA_{44}^{-1}G^T - (B^T - FA_{44}^{-1}D^T)\Delta\mathcal{C}\big)X_3 \\
&= -\mathcal{B}^T\Delta\Gamma_1 + \Gamma_2 - \mathcal{F}A_{44}^{-1}\Gamma_4 + (B^T\Delta\mathcal{B} + FA_{44}^{-1}\mathcal{F}^T)X_2 - (\mathcal{E} - \mathcal{B}^T\Delta\mathcal{C})X_3.
\end{aligned}$$

Therefore

$$(A_{22} - B^T\Delta\mathcal{B} - FA_{44}^{-1}\mathcal{F}^T)X_2 = -\mathcal{B}^T\Delta\Gamma_1 + \Gamma_2 - \mathcal{F}A_{44}^{-1}\Gamma_4 - \mathcal{K}X_3.$$

Notice that the left-hand side simplifies to one of our inverse equations:

$$(A_{22} - B^T\Delta\mathcal{B} - FA_{44}^{-1}\mathcal{F}^T)X_2 = (A_{22} - FA_{44}^{-1}F^T - \mathcal{B}^T\Delta\mathcal{B})X_2 = \Phi^{-1}X_2,$$

$$X_2 = \Phi\big(-\mathcal{B}^T\Delta\Gamma_1 + \Gamma_2 - \mathcal{F}A_{44}^{-1}\Gamma_4 - \mathcal{K}X_3\big). \qquad (12)$$

Substituting (12) into the expressions (10) and (11) we have

$$X_1 = (\Delta + \Delta\mathcal{B}\Phi\mathcal{B}^T\Delta)\Gamma_1 - \Delta\mathcal{B}\Phi\Gamma_2 - \Delta(D - \mathcal{B}\Phi\mathcal{F})A_{44}^{-1}\Gamma_4 - \Delta(\mathcal{C} - \mathcal{B}\Phi\mathcal{K})X_3,$$

$$X_4 = A_{44}^{-1}(-D^T + \mathcal{F}^T\Phi\mathcal{B}^T)\Delta\Gamma_1 - A_{44}^{-1}\mathcal{F}^T\Phi\Gamma_2 + (\mathcal{M} + A_{44}^{-1}\mathcal{F}^T\Phi\mathcal{F}A_{44}^{-1})\Gamma_4$$
$$+ A_{44}^{-1}(\mathcal{F}^T\Phi\mathcal{K} - \mathcal{G}^T)X_3. \qquad (13)$$

Finally, we substitute these expressions into (8) to obtain X_3 in terms of Γ_1, Γ_2, Γ_3, and Γ_4 which is simplified to

$$\Xi^{-1}X_3 = (-\mathcal{C}^T + \mathcal{K}^T\Phi\mathcal{B}^T)\Delta\Gamma_1 - \mathcal{K}^T\Phi\Gamma_2 + \Gamma_3$$
$$+ (\mathcal{K}^T\Phi\mathcal{F}A_{44}^{-1} - G\mathcal{M} + C^T\Delta DA_{44}^{-1})\Gamma_4. \qquad (14)$$

Basic Proof Steps

The only blocks of interest in the inverse of $\bar{\Sigma}$ are those that we denote $X_3(\Gamma_1)$ and $X_3(\Gamma_2)$, which are defined by

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} A_{11} & B & C & D \\ B^T & A_{22} & E & F \\ C^T & E^T & A_{33} & G \\ D^T & F^T & G^T & A_{44} \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \Gamma_3 \\ \Gamma_4 \end{bmatrix} = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ X_3(\Gamma_1) & X_3(\Gamma_2) & \times & \times \\ \times & \times & \times & \times \end{bmatrix} \tag{15}$$

where "$\times$" denotes a block that is not of interest.

Comparing (14) and (15), we find that

$$X_3(\Gamma_2) = -\Xi \mathcal{K}^T \Phi,$$

and we require this expression to be zero for the maximal determinant completion.

- Since Φ and Ξ are inverses, they cannot be zero, therefore we require $\mathcal{K}^T = 0$.

HERIOT WATT UNIVERSITY

Similarly, we have

$$X_3(\Gamma_1) = \Xi(-\mathcal{C}^{\mathcal{T}} + \mathcal{K}^T \Phi \mathcal{B}^T)\Delta,$$

and since $\mathcal{K}^T = 0$ (and Δ and Ξ are nonsingular) we require that $\mathcal{C} = 0$, which implies that

$$C = DA_{44}^{-1}G^T. \tag{16}$$

The equations $\mathcal{C} = 0$ and $\mathcal{K} = 0$ imply $\mathcal{E} = 0$, and hence

$$E = FA_{44}^{-1}G^T.$$

Denoting by Π the permutation matrix that reverses the order of the blocks in $\bar{\Sigma}$, we have

$$\Pi^T \bar{\Sigma} \Pi = \begin{bmatrix} A_{44} & G^T & F^T & D^T \\ G & A_{33} & E^T & C^T \\ F & E & A_{22} & B^T \\ D & C & B & A_{11} \end{bmatrix}.$$

The block F^T now takes the role of C in the original matrix, so from (16) we obtain, after transposing, $F = B^T A_{11}^{-1} D$.

We have now found the MaxDet completion!