Wilks Memorial Seminar, Princeton University

Regression graphs and sparsity-inducing reparametrisations

Heather Battey
Department of Mathematics, Imperial College London

April 3, 2025

STARTING POINT

Q^*: For a given covariance matrix, not obviously sparse in any domain, can a non-trivial sparsity-inducing reparametrisation be deduced?

A^*: ...

TRIVIAL VS NON-TRIVIAL REPARAMETRISATION

Q*: For a given covariance matrix, not obviously sparse in any domain, can a **non-trivial** sparsity-inducing reparametrisation be deduced?

A*: ...

By **non-trivial** we mean that we are able to discriminate more effectively on the new scale between elements that are large and elements that are small.

Example of a trivial reparametrisation: $\Sigma \mapsto c\Sigma$ for $c > 0$ close to zero.

TWO DISTINCT TYPES OF MOTIVATION

1. The reparametrised covariance may be the interest parameter by virtue of the interpretation ascribed to its zeros.
2. If the covariance matrix or its inverse is a nuisance parameter, a sparsity assumption allows construction of estimators that are consistent in relevant matrix norms when dimension exceeds sample size.

Positive definiteness enforces additional constraints on how sparsity can legitimately manifest.

PROOF OF CONCEPT FOR Q^*

Possibility of increasing sparsity through reparametrisation.

One proof of concept is immediately available: covariance matrices associated with Gaussian graphical models have no zeros as long as the underlying conditional independence graph is connected, while the inverse covariance matrix may have many zeros. Other examples?

Q*: For a given covariance matrix can a sparsity-inducing reparametrisation be deduced?

Sparsity-inducing reparametrisations for covariance matrices

Battey, H. S. (2017). Eigen structure of a new class of structured covariance and inverse covariance matrices. *Bernoulli*, 23, 3166–3177.

Rybak, J. and Battey, H. S. (2021). Sparsity induced by covariance transformation: some deterministic and probabilistic results. *Proc. Roy. Soc. Lond. A*, 477.

Rybak, J., Battey, H. S., Bharath, K. (2025). Regression graphs and sparsity-inducing reparametrisation, *arXiv:2402.05708*

NON-STANDARD PARAMETRISATION: FIRST EXAMPLE

The matrix logarithm L of a covariance matrix Σ is defined as

$$\Sigma = \exp(L) = \sum_{k=0}^{\infty} \frac{1}{k!} L^k.$$

Spectral decomposition:

$$\begin{aligned} \Sigma &= \Gamma \Lambda \Gamma^T, \quad \Lambda \triangleq \text{diag}\{\lambda_1, \ldots, \lambda_p\} \\ L &= \Gamma \Delta \Gamma^T, \quad \Delta \triangleq \text{diag}\{\log(\lambda_1), \ldots, \log(\lambda_p)\}. \end{aligned}$$

The inverse satisfies $\Sigma^{-1} = \exp(-L)$.

WHAT STRUCTURE IS INDUCED ON Σ THROUGH SPARSITY OF L?

$$\Sigma,\ \Sigma^{-1} \in \mathcal{V}_p^+(\mathbb{R}) \triangleq \{S \in \mathcal{M}_p(\mathbb{R}) : S = S^T,\ S \succ 0\} \quad \text{(open cone)}$$
$$L \in \mathcal{V}_p(\mathbb{R}) \triangleq \{S \in \mathcal{M}_p(\mathbb{R}) : S = S^T\} \quad \text{(vector space)}.$$

Natural symmetrised basis for $\mathcal{V}_p(\mathbb{R})$ of the form $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$:

$$\mathcal{B}_1 = \{B : B = e_j e_j^T,\ j \in [p]\}$$
$$\mathcal{B}_2 = \{B : B = e_j e_k^T + e_k e_j^T,\ j, k \in [p],\ j \neq k\}.$$

By contrast, $\mathcal{V}_p^+(\mathbb{R})$ does not possess a basis.

$$L = \textstyle\sum_{m=1}^{|B|} \alpha_m B_m \ \text{ where } B_1, \ldots, B_{|\mathcal{B}|} \in \mathcal{B}.$$

WHAT STRUCTURE IS INDUCED ON Σ THROUGH SPARSITY OF L?

Impose sparsity on

$$L = \sum_{m=1}^{|\mathcal{B}|} \alpha_m B_m \quad \text{where } B_1, \ldots, B_{|\mathcal{B}|} \in \mathcal{B}.$$

through the basis coefficients. Specifically:

$$\alpha = (\alpha_1, \ldots, \alpha_{|\mathcal{B}|}) \text{ satisfies } \|\alpha\|_0 = s^* < p.$$

The eigenvectors and eigenvalues of Σ inherit substantial structure.

STRUCTURE INDUCED ON THE EIGENVECTORS AND EIGENVALUES OF Σ THROUGH SPARSITY OF L

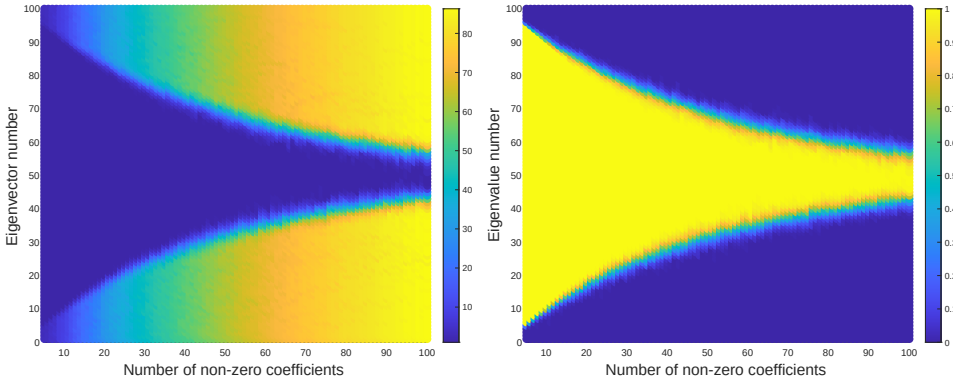


Figure: Simulation average of $\|\gamma_j\|_0$ (left) and $\mathbb{I}\{\lambda_j = 1\}$ (right) for 100 random logarithmically s^*-sparse covariance matrices, plotted against index j of ordered eigenvalues (y-axis) and $s^* \in \{1, \ldots, p\}$ (x-axis) for $p = 100$.

WHAT STRUCTURE IS INDUCED ON Σ THROUGH SPARSITY OF L?

There is a deterministic answer. A random matrix perspective aids interpretation.

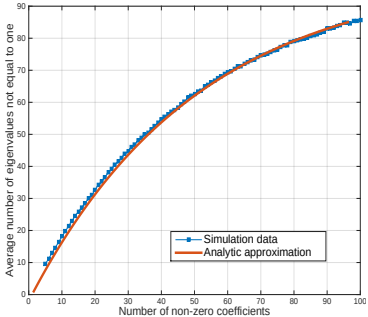
Suppose the support of α is a simple random sample of size s^* from the index set $\{1, \ldots, p(p+1)/2\}$.

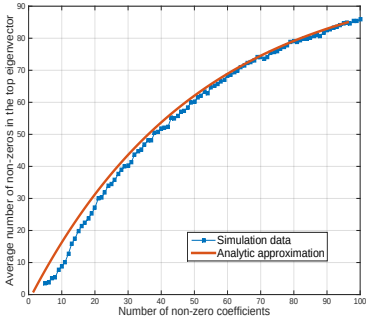
- The expected number of non-unit eigenvalues of $\Sigma = \Sigma(\alpha)$ is approximately $d^* < p$, where

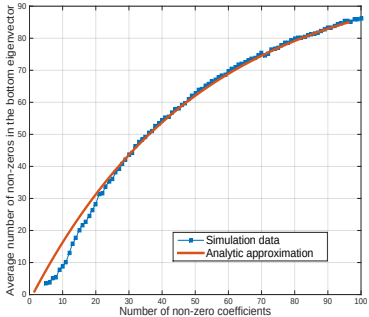
$$d^* = \text{root}\left\{ \frac{4p + p(p-1)}{2(p+1)} \left[\log\left(\frac{p}{p-d} \right) - \frac{d}{2p(p-d)} \right] - s^* \right\}.$$

- The corresponding eigenvectors have d^* non-zeros in expectation.
- The other eigenvectors are of the form e_j.

APPROXIMATION ERROR

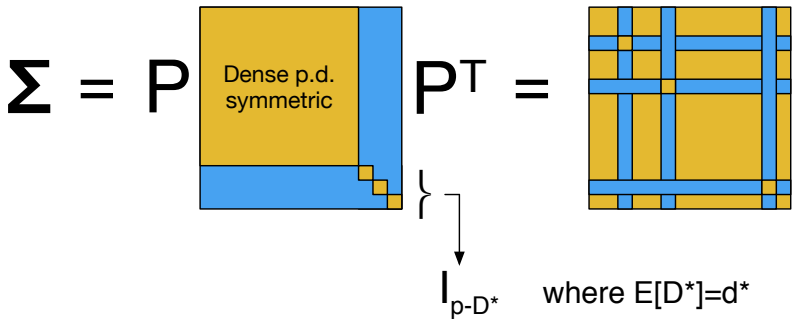






WHAT STRUCTURE IS INDUCED ON Σ THROUGH SPARSITY OF L?

Suppose the support of α is a simple random sample of size s^* from the index set $\{1, \ldots, p(p+1)/2\}$. The resulting Σ is of the form



where P is a permutation matrix. The same structure holds for deterministic logarithmically sparse covariance matrices but the dimension of the identity block is less explicit.

WHAT STRUCTURE IS INDUCED ON Σ THROUGH SPARSITY OF L?

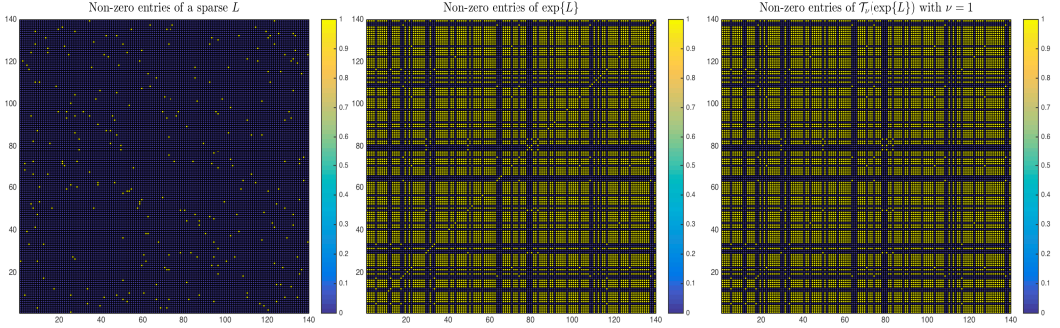
Indicator of non-zero entries for:

Left: one realisation of a random sparse L;

Centre: the corresponding matrix exponential $\Sigma = \exp(L)$

Right: the thresholded version $\mathcal{T}(\Sigma) = \{\Sigma_{ij}\mathbb{I}(|\Sigma_{ij}| \geq 1)\}$.

Yellow entries represent non-zeros. Blue entries represent zeros.



A MORE NUANCED SUMMARY

A sparse L with $s^* = \|\alpha\|_0 < p$ necessarily has more exact zeros than Σ and Σ^{-1}.

Neither Σ nor Σ^{-1} with the specified structure can have more exact zeros than L.

For a randomly generated Σ with the specified structure, however, $L = \log(\Sigma)$ will contain the same number of exact zeros as Σ with probability 1. Any practical advantages are thus more likely to arise in the form of approximate zeros.

QUESTIONS

- Other sparsity-inducing reparametrisations.
- Interpretation of a zero (exact or approximate) in the logarithmic domain.
- Connections to graphical structure and zeros in precision matrix.
- Propagation of estimation errors between scales.

Regression graphs and sparsity-inducing reparametrisations

arXiv:2402.09112

Joint work with Jakub Rybak and Karthik Bharath.

FOUR REPARAMETRISATION MAPS

With $D(d) = \mathrm{diag}(d_1, \ldots, d_p)$, we consider the four maps

$$
\begin{aligned}
\alpha \mapsto \Sigma_{pd}(\alpha) \quad &:= e^{L(\alpha)}, && L(\alpha) \in \mathrm{Sym}(p), && \alpha \in \mathbb{R}^{p(p+1)/2};\\
(\alpha, d) \mapsto \Sigma_{o}(\alpha, d) \quad &:= e^{L(\alpha)} e^{D(d)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \mathrm{Sk}(p), && \alpha \in \mathbb{R}^{p(p-1)/2}, \ d \in \mathbb{R}^{p};\\
\alpha \mapsto \Sigma_{lt}(\alpha) \quad &:= e^{L(\alpha)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \mathrm{LT}(p), && \alpha \in \mathbb{R}^{p(p+1)/2};\\
(\alpha, d) \mapsto \Sigma_{ltu}(\alpha, d) \quad &:= e^{L(\alpha)} e^{D(d)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \mathrm{LT}_{\mathrm{s}}(p), && \alpha \in \mathbb{R}^{p(p-1)/2}, \ d \in \mathbb{R}^{p}.
\end{aligned}
$$

In each case, L belongs to a different vector space in which sparsity can conveniently be studied:

- $\mathrm{Sym}(p)$: the symmetric matrices;
- $\mathrm{Sk}(p)$: the skew-symmetric matrices;
- $\mathrm{LT}(p)$: the lower triangular matrices;
- $\mathrm{LT}_{\mathrm{s}}(p)$: the strictly lower triangular matrices.

FOUR REPARAMETRISATION MAPS

With $D(d) = \mathrm{diag}(d_1, \ldots, d_p)$, we consider the four maps

$$
\begin{aligned}
\alpha \mapsto \Sigma_{pd}(\alpha) \quad &:= e^{L(\alpha)}, && L(\alpha) \in \mathrm{Sym}(p), && \alpha \in \mathbb{R}^{p(p+1)/2};\\
(\alpha, d) \mapsto \Sigma_{o}(\alpha, d) \quad &:= e^{L(\alpha)} e^{D(d)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \mathrm{Sk}(p), && \alpha \in \mathbb{R}^{p(p-1)/2}, \ d \in \mathbb{R}^{p};\\
\alpha \mapsto \Sigma_{lt}(\alpha) \quad &:= e^{L(\alpha)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \mathrm{LT}(p), && \alpha \in \mathbb{R}^{p(p+1)/2};\\
(\alpha, d) \mapsto \Sigma_{ltu}(\alpha, d) \quad &:= e^{L(\alpha)} e^{D(d)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \mathrm{LT}_{\mathrm{s}}(p), && \alpha \in \mathbb{R}^{p(p-1)/2}, \ d \in \mathbb{R}^{p}.
\end{aligned}
$$

The subscripts on Σ indicate which of the matrix sets are represented as the image of the exponential map: $\mathrm{PD}(p)$ (positive definite), $\mathrm{SO}(p)$ (special orthogonal), $\mathrm{LT}_{+}(p)$ (lower triangular, w/ positive diagonal) and $\mathrm{LT}_{\mathrm{u}}(p)$ (lower triangular w/ unit diagonal).

FOUR REPARAMETRISATION MAPS

With $D(d) = \text{diag}(d_1, \ldots, d_p)$, we consider the four maps

$$
\begin{aligned}
\alpha \mapsto \Sigma_{pd}(\alpha) \quad &:= e^{L(\alpha)}, && L(\alpha) \in \text{Sym}(p), && \alpha \in \mathbb{R}^{p(p+1)/2}; \\
(\alpha, d) \mapsto \Sigma_{o}(\alpha, d) \quad &:= e^{L(\alpha)} e^{D(d)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \text{Sk}(p), && \alpha \in \mathbb{R}^{p(p-1)/2}, \ \ d \in \mathbb{R}^{p}; \\
\alpha \mapsto \Sigma_{lt}(\alpha) \quad &:= e^{L(\alpha)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \text{LT}(p), && \alpha \in \mathbb{R}^{p(p+1)/2}; \\
(\alpha, d) \mapsto \Sigma_{ltu}(\alpha, d) \quad &:= e^{L(\alpha)} e^{D(d)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \text{LT}_{\text{s}}(p), && \alpha \in \mathbb{R}^{p(p-1)/2}, \ \ d \in \mathbb{R}^{p}.
\end{aligned}
$$

- Sparsity of α in the map $\alpha \mapsto \Sigma_{pd}(\alpha)$ was presented a few slides earlier (Battey, 2017).
- The map $(\alpha, d) \mapsto \Sigma_{o}(\alpha, d)$ was studied by Rybak and Battey (2021).
- The maps $\alpha \mapsto \Sigma_{lt}(\alpha)$ and $(\alpha, d) \mapsto \Sigma_{ltu}(\alpha, d)$ are new and have a graphical interpretation.
- There is an encompassing formulation.

FOUR REPARAMETRISATION MAPS

With $D(d) = \text{diag}(d_1, \ldots, d_p)$, we consider the four maps

$$
\begin{aligned}
\alpha \mapsto \Sigma_{pd}(\alpha) \quad &:= e^{L(\alpha)}, && L(\alpha) \in \text{Sym}(p), && \alpha \in \mathbb{R}^{p(p+1)/2};\\
(\alpha, d) \mapsto \Sigma_{o}(\alpha, d) \quad &:= e^{L(\alpha)} e^{D(d)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \text{Sk}(p), && \alpha \in \mathbb{R}^{p(p-1)/2}, \ \ d \in \mathbb{R}^{p};\\
\alpha \mapsto \Sigma_{lt}(\alpha) \quad &:= e^{L(\alpha)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \text{LT}(p), && \alpha \in \mathbb{R}^{p(p+1)/2};\\
(\alpha, d) \mapsto \Sigma_{ltu}(\alpha, d) \quad &:= e^{L(\alpha)} e^{D(d)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \text{LT}_{\text{s}}(p), && \alpha \in \mathbb{R}^{p(p-1)/2}, \ \ d \in \mathbb{R}^{p}.
\end{aligned}
$$

- Sparsity of α in the map $\alpha \mapsto \Sigma_{pd}(\alpha)$ was presented a few slides ago (Battey, 2017).
- The map $(\alpha, d) \mapsto \Sigma_{o}(\alpha, d)$ was studied by Rybak and Battey (2021).

- The maps $\alpha \mapsto \Sigma_{lt}(\alpha)$ and $(\alpha, d) \mapsto \Sigma_{ltu}(\alpha, d)$ are new and have a graphical interpretation.
- There is an encompassing formulation.

FOUR REPARAMETRISATION MAPS

With $D(d) = \text{diag}(d_1, \ldots, d_p)$, we consider the four maps

$$
\begin{aligned}
\alpha \mapsto \Sigma_{pd}(\alpha) \quad &:= e^{L(\alpha)}, && L(\alpha) \in \text{Sym}(p), && \alpha \in \mathbb{R}^{p(p+1)/2};\\
(\alpha, d) \mapsto \Sigma_{o}(\alpha, d) \quad &:= e^{L(\alpha)} e^{D(d)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \text{Sk}(p), && \alpha \in \mathbb{R}^{p(p-1)/2}, \ \ d \in \mathbb{R}^{p};\\
\alpha \mapsto \Sigma_{lt}(\alpha) \quad &:= e^{L(\alpha)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \text{LT}(p), && \alpha \in \mathbb{R}^{p(p+1)/2};\\
(\alpha, d) \mapsto \Sigma_{ltu}(\alpha, d) \quad &:= e^{L(\alpha)} e^{D(d)} (e^{L(\alpha)})^{\mathrm{T}}, && L(\alpha) \in \text{LT}_{\text{s}}(p), && \alpha \in \mathbb{R}^{p(p-1)/2}, \ \ d \in \mathbb{R}^{p}.
\end{aligned}
$$

- Sparsity of α in the map $\alpha \mapsto \Sigma_{pd}(\alpha)$ was presented a few slides ago (Battey, 2017).
- The map $(\alpha, d) \mapsto \Sigma_{o}(\alpha, d)$ was studied by Rybak and Battey (2021).

- The maps $\alpha \mapsto \Sigma_{lt}(\alpha)$ and $(\alpha, d) \mapsto \Sigma_{ltu}(\alpha, d)$ are new and have a graphical interpretation.
- There is an encompassing formulation.

OBJECTIVES

To provide insight into how sparsity interacts with these parametrisations.

Not to advocate any particular sparsity scale, but to clarify, to the extent feasible, the implications of choosing it.

MATRIX DECOMPOSITIONS

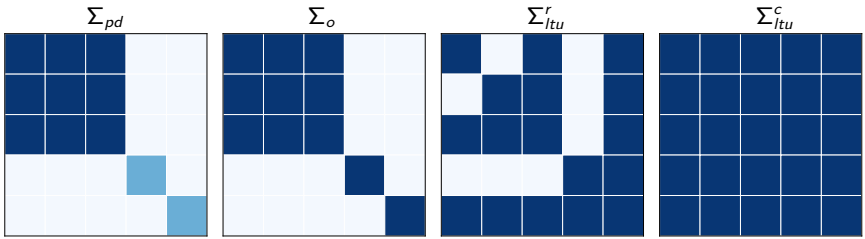
The parametrisations correspond to matrix-logarithmic transformation following different matrix decompositions.

Σ_{pd}: no preliminary decomposition.

Σ_{o}: spectral decomposition.

Σ_{lt}: Cholesky decomposition.

Σ_{ltu}: LDL decomposition.



Structure of $\Sigma(\alpha)$ induced by sparsity of α. Zero entries are light blue, unit entries are medium blue, unrestricted entries are dark blue.

SOME REMARKS ON THE PREVIOUS FIGURE

- The structure depicted is that of $P\Sigma P^{\mathrm{T}}$ for some permutation matrix P.
- P reflects the arbitrary ordering of variables.
- Σ^{r}_{ltu} and Σ^{c}_{ltu} are induced by two configurations of 6 non-zero entries in $\alpha \in \mathbb{R}^{10}$.
- Σ^{c}_{ltu} is completely dense on the original scale, but sparse after reparametrisation. Similarly for Σ_{lt} (not depicted).

Some formal statements

A GENERAL RESULT

Consider any p-dimensional matrix M of the form $M = e^L$, where L belongs to a vector space (e.g. any of the four defined earlier). Let d_r^* and d_c^* be the number of non-zero rows and columns of L respectively. Then:

- M has $p - d_r^*$ rows of the form e_j^{T} for some $j \in [p]$, all distinct, and $p - d_c^*$ columns of the form e_j.
- Of these, $p - d^*$ coincide after transposition.
- If M is normal, i.e. $M^{\mathrm{T}}M = MM^{\mathrm{T}}$, then $d_r^* = d_c^* = d^*$.

Zero rows and columns of L are likely to arise when $s^* = \|\alpha\|_0 \ll p$. Probabilistic statements can be made when the positions of non-zero entries are picked totally at random.

EXAMPLE STRUCTURE OF $M = e^L$

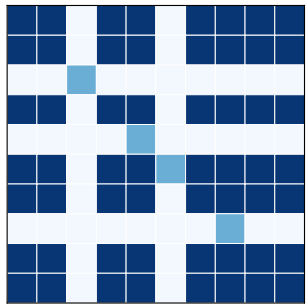


Figure: Example of a structure of M as as described on the last slide with $p = 10$, $d_r^* = 7$, $d_c^* = 8$ and $d^* = 9$. Zero, unit and unconstrained entries are light, medium and dark blue respectively.

The specific vector spaces of interest impose additional constraints.

THE MAP $\alpha \mapsto \Sigma_{pd}(\alpha)$

Corollary

The image of the map $\alpha \mapsto \Sigma_{pd}(\alpha) = e^{L(\alpha)}$ is logarithmically sparse in the sense that $\|\alpha\|_0 = s^*$ in the basis representation for $L(\alpha)$ if and only if Σ is of the form $\Sigma = P\Sigma^{(0)}P^{\mathrm{T}}$, where $P \in \mathrm{P}(p)$ is a permutation matrix and $\Sigma^{(0)} = \Sigma_1^{(0)} \oplus I_{p-d^*}$ with $\Sigma_1^{(0)} \in \mathrm{PD}(d^*)$ of maximal dimension, in the sense that it is not possible to find another permutation $P \in \mathrm{P}(p)$ such that the dimension of the identity block is larger than $p - d^*$.

THE MAP $\alpha \mapsto \Sigma_o(\alpha)$

Corollary

The image of the map $\alpha \mapsto \Sigma_o(\alpha) = e^{L(\alpha)} e^{D} (e^{L(\alpha)})^{\mathrm{T}}$ is logarithmically sparse in the sense that $\|\alpha\|_0 = s^*$ in the basis representation for $L(\alpha)$ if and only if Σ is of the form $\Sigma = P \Sigma^{(0)} P^{\mathrm{T}}$, where $P \in \mathrm{P}(p)$ is a permutation matrix and $\Sigma^{(0)} = \Sigma_1^{(0)} \oplus D_{p-d^*}$, where $D_{p-d^*} \in \mathrm{D}(p - d^*)$ and $\Sigma_1^{(0)} \in \mathrm{PD}(d^*)$ is of maximal dimension, in the sense that it is not possible to find another permutation $P \in \mathrm{P}(p)$ such that the dimension of the diagonal block is larger than $p - d^*$.

THE MAP $\alpha \mapsto \Sigma_{lt}(\alpha)$

Corollary

The image of the map $\alpha \mapsto \Sigma_{lt}(\alpha) = e^{L(\alpha)}(e^{L(\alpha)})^{\mathrm{T}}$ is logarithmically sparse in the sense that $\|\alpha\|_0 = s^*$ in the basis representation for $L(\alpha)$ if and only if Σ is of the form $\Sigma = VV^{\mathrm{T}}$, where $V = I_p + \Theta$ and $\Theta \in \mathrm{LT}_+(p)$ has $p - d_r^*$ zero rows and $p - d_c^*$ zero columns, of which $p - d^*$ coincide.

THE MAP $\alpha \mapsto \Sigma_{ltu}(\alpha)$

Corollary

The image of the map $\alpha \mapsto \Sigma_{ltu}(\alpha) = e^{L(\alpha)} e^{D} (e^{L(\alpha)})^{\mathrm{T}}$ is logarithmically sparse in the sense that $\|\alpha\|_0 = s^*$ in the basis representation for $L(\alpha)$ if and only if Σ is of the form $\Sigma = U \Psi U^{\mathrm{T}}$, where $\Psi = e^{D} \in \mathrm{D}_+(p)$, $U = I_p + \Theta$ and $\Theta \in \mathrm{LT}_{\mathrm{s}}(p)$ has $p - d_r^*$ zero rows and $p - d_c^*$ zero columns, of which $p - d^*$ coincide.

QUESTIONS OF INTERPRETATION

- Interpretation of α and zeros in α.
- Exact zeros vs approximate zeros.
- Interpretation of the structure in Σ induced by/inducing a sparse α in $\alpha \mapsto \Sigma(\alpha)$.

Background to Σ_{ltu} interpretation: causal ordering

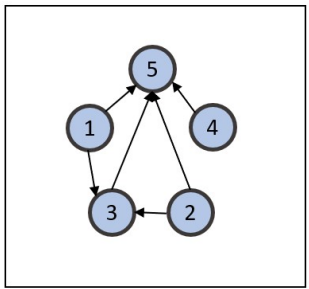
CONTRAST: DIRECTED/UNDIRECTED GRAPHS

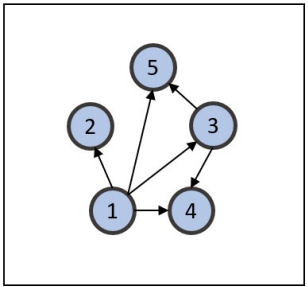
- Multiple causal models compatible with the same structure of zeros in Σ^{-1}.
- An undirected graph whose associated Gaussian model has a sparse Σ^{-1} could be appreciably less sparse in Σ^{-1} when the undirected edges are replaced by directed ones.
- The key factor determining this is whether there are common response variables occurring later in the causal ordering.
- Other parametrisations are more appropriate.

CONTRAST: DIRECTED/UNDIRECTED GRAPHS

- Multiple causal models compatible with the same structure of zeros in Σ^{-1}.
- An undirected graph whose associated Gaussian model has a sparse Σ^{-1} could be appreciably less sparse in Σ^{-1} when the undirected edges are replaced by directed ones.
- The key factor determining this is whether there are common response variables occurring later in the causal ordering.
- Other parametrisations are more appropriate.

CONTRAST: COMMON RESPONSE/COMMON SOURCE VARIABLES





More later. . .

Background to Σ_{ltu} interpretation: Iwasawa coordinates

BLOCK DIAGONALISATION

With $[p] = \{1, \ldots, p\}$, let $a \subset [p]$ and $b = [p] \backslash a$ be disjoint subsets of variable indices. As a consequence of a block-diagonalisation identity for symmetric matrices (Cox and Wermuth, 1993, 2004),

$$L\Sigma L^{\mathrm{T}} = \begin{pmatrix} I_{aa} & 0 \\ -\Sigma_{ba}\Sigma_{aa}^{-1} & I_{bb} \end{pmatrix} \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \begin{pmatrix} I_{aa} & -\Sigma_{aa}^{-1}\Sigma_{ab} \\ 0 & I_{bb} \end{pmatrix} = \begin{pmatrix} \Sigma_{aa} & 0 \\ 0 & \Sigma_{bb.a} \end{pmatrix},$$

so that Σ can be written in terms of $\Pi_{b|a} := \Sigma_{ba}\Sigma_{aa}^{-1} \in \mathbb{R}^{|b| \times |a|}$, $\Sigma_{aa} \in \mathrm{PD}(|a|)$,

$$\Sigma_{bb.a} := \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \in \mathrm{PD}(|b|).$$

These are known in some quarters as the partial Iwasawa coordinates for $\mathrm{PD}(p)$ based on a two-component partition $|a| + |b| = p$ of $[p]$.

This holds independently of any distributional assumptions on the underlying RVs.

INTERPRETATION OF BLOCKS

Let $Y = (Y_a^{\mathrm{T}}, Y_b^{\mathrm{T}})^{\mathrm{T}}$ be a mean-zero random vector with covariance matrix Σ, $\Pi_{b|a}$ is the matrix of regression coefficients on Y_a in a linear regression of Y_b on Y_a and $\Sigma_{bb.a}$ is the error covariance matrix, i.e. $Y_b = \Pi_{b|a} Y_a + \varepsilon_b$ and $\Sigma_{bb.a} = \mathrm{var}(\varepsilon_b)$.

The entries of $\Pi_{b|a}$ encapsulate dependencies between each variable in b and those of a, conditional on other variables in a, but marginalising over the remaining variables in b.

Σ_{ltu} interpretation

Σ_{ltu} FROM RECURSIVE BLOCK-DIAGONALISATION

With $|b| = 1$, recursively apply the identity

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} = \begin{pmatrix} I_{aa} & 0 \\ \Sigma_{ba}\Sigma_{aa}^{-1} & I_{bb} \end{pmatrix} \begin{pmatrix} \Sigma_{aa} & 0 \\ 0 & \Sigma_{bb.a} \end{pmatrix} \begin{pmatrix} I_{aa} & \Sigma_{aa}^{-1}\Sigma_{ab} \\ 0 & I_{bb} \end{pmatrix}.$$

This leads to the representation $\Sigma = U\Psi U^{\mathrm{T}}$ based on p blocks of size 1×1 where the general form of $U = e^L$ ignoring sparsity is

$$U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \beta_{2.1} & 1 & 0 & 0 \\ \beta_{3.1} & \beta_{3.21} & 1 & 0 \\ \beta_{4.1} & \beta_{4.21} & \beta_{4.3[2]} & 1 \end{pmatrix}. \tag{1}$$

Notation: e.g. $\beta_{4.21}$ is the coefficient on Y_2 in a linear regression of Y_4 on Y_1 and Y_2.
This is not new: it is implicit in Cox and Wermuth (1993, 2004).

INTERPRETATION OF ENTRIES OF U IN $\Sigma_{ltu} = U\Psi U^{\mathrm{T}}$

Three variables (Y_1, Y_2, Y_3). The total effect of Y_1 on Y_3 is related to the conditional effects through Cochran's formula:

$$\beta_{3.1} = \beta_{3.12} + \beta_{3.21}\beta_{2.1}.$$

Notation.

- $\beta_{3.1}$: regression coefficient on Y_1 in a regression of Y_3 on Y_1 only, having marginalised over Y_2.
- $\beta_{3.12}$: coeffient on Y_1 in a regression of Y_3 on Y_1 and Y_2.

POPULATION-LEVEL DEFINITION OF COEFFICIENTS

The coefficient $\beta_{3.1}$ is the total derivative of

$$f(y_1, \bar{y}_2) := \mathbb{E}(Y_3 \mid Y_1 = y_1, Y_2 = \bar{y}_2),$$

treating y_1 and $\bar{y}_2 = \bar{y}_2(y_1) = \mathbb{E}(Y_2 \mid Y_1 = y_1)$ as free variables, i.e.

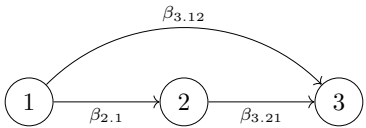
$$\beta_{3.1} = \frac{Df(y_1, \bar{y}_2)}{Dy_1} = \underbrace{\frac{\partial f(y_1, \bar{y}_2)}{\partial y_1}}_{\beta_{3.12}} + \underbrace{\frac{\partial f(y_1, \bar{y}_2)}{\partial \bar{y}_2}}_{\beta_{3.21}} \underbrace{\frac{d\bar{y}_2(y_1)}{dy_1}}_{\beta_{2.1}}.$$

INTERPRETATION OF ENTRIES OF U IN $\Sigma_{ltu} = U\Psi U^{\mathrm{T}}$

The right hand side of Cochran's formula:

$$\beta_{3.1} = \beta_{3.12} + \beta_{3.21}\beta_{2.1}.$$

corresponds to tracing the effects of Y_3 on Y_1 along two paths connecting the nodes in a system of random variables (Y_1, Y_2, Y_3), with edge weights given by the corresponding regression coefficients.



MARGINALISATION AND CONDITIONING

Marginalisation, indicated by $\not\oslash$, induces an edge between i and j if the marginalised variable is a transition node or a source node.

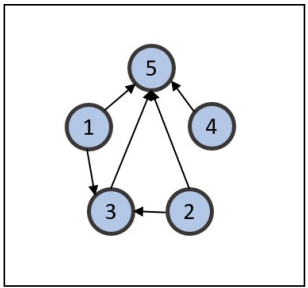
$$i \longleftarrow \not\oslash \longrightarrow j, \qquad i \longleftarrow \not\oslash \longleftarrow j,$$
$$i \text{ ---- } j, \qquad i \longleftarrow j.$$

By contrast, if i and j are separated by a sink node, then conditioning on such a node, indicated by $\boxdot$, is edge inducing, with no direction implied.

$$i \longrightarrow \boxdot \longleftarrow j,$$
$$i \text{ — } j.$$

Edge inducement: an independence statement that applies in the true graph no longer holds in all distributions that are generated over the new graph.

PRECISION MATRICES AND SINK NODES



If edges were undirected: $4 \perp\!\!\!\perp \{1,2,3\} \mid 5$ and $\Sigma^{-1}_{4j} = 0$ for $j = 1,2,3$.

Directed edges: conditioning on common sink node 5 (implicit in interpretation of Σ^{-1}) induces an edge between variable 4 and all other variables.

INTERPRETATION OF ENTRIES OF U IN $\Sigma_{ltu} = U\Psi U^{\mathrm{T}}$

Let $\upsilon_{ij}(\ell)$ denote the effect of Y_j on Y_i along all paths of length ℓ, specified for the three-dimensional example as

$$\begin{aligned} \upsilon_{21}(1) &= \beta_{2.1}, & \upsilon_{31}(1) &= \beta_{3.12}, \\ \upsilon_{32}(1) &= \beta_{3.21}, & \upsilon_{31}(2) &= \beta_{3.21}\beta_{2.1}. \end{aligned}$$

The lower-triangular matrices U and $L = \log(U)$ have the form,

$$U = \begin{pmatrix} 1 & 0 & 0 \\ \upsilon_{21}(1) & 1 & 0 \\ \upsilon_{31}(1) + \upsilon_{31}(2) & \upsilon_{32}(1) & 1 \end{pmatrix}, \quad L = \begin{pmatrix} 0 & 0 & 0 \\ \upsilon_{21}(1) & 0 & 0 \\ \upsilon_{31}(1) + \frac{\upsilon_{31}(2)}{2} & \upsilon_{32}(1) & 0 \end{pmatrix}.$$

INTERPRETATION OF α IN THE Σ_{ltu} PARAMETRISATION

For a given diagonal matrix e^D, the Σ_{ltu} parametrisation is

$$\alpha \mapsto \Sigma_{ltu}(\alpha) = e^{L(\alpha)} e^D (e^{L(\alpha)})^{\mathrm{T}}.$$

Entry L_{ij} and the corresponding coefficient α is equal to the weighted sum of effects of Y_j on Y_i along all paths connecting the two nodes, with weights inversely proportional to the length of the corresponding path.

THE POSSIBILITY OF DIFFERENT SPARSITY ON DIFFERENT SCALES

$$U = \begin{pmatrix} 1 & 0 & 0 \\ \beta_{2.1} & 1 & 0 \\ \beta_{3.12} + \beta_{3.21}\beta_{2.1} & \beta_{3.21} & 1 \end{pmatrix}, \quad L = \begin{pmatrix} 0 & 0 & 0 \\ \beta_{2.1} & 0 & 0 \\ \beta_{3.12} + \frac{\beta_{3.21}\beta_{2.1}}{2} & \beta_{3.21} & 0 \end{pmatrix}.$$

The possibility of increasing sparsity (exact or approximate zeros) through reparametrisation comes either from cancellation (exact or approximate), or from the different weightings of path effects.

WEIGHTING

Let B be the matrix whose entry B_{ij} is the regression coefficient on Y_j in a regression of Y_i on Y_j and on its other causally-ordered predecessors.

- Entries of B represent paths of length 1, i.e. zero weights on longer paths.
- Entries of $U = (I - B)^{-1}$ aggregate contributions along all paths, with weights equal to one, i.e. no discounting of longer paths.
- $L = \log(U)$ weights a path of length ℓ by a factor of $1/\ell$.

INTERPRETATION OF NEAR-ZEROS IN THE LOG DOMAIN UNDER Σ_{ltu}

Interpretation of a near-zero entry L_{ij}: short paths from j to i are associated with small conditional effects, while any large effects are mediated by a string of intermediate variables, where conditioning is on all variables that occur earlier in the causal ordering.

APPROXIMATION INHERENT TO THRESHOLDING ON THE LOG SCALE

The approximation inherent to any statistical algorithm that sets small values of α to zero is thus as follows: the relation between nodes i and $j < i$ would be declared null if relatively direct regression effects are negligible and other effects manifest through long paths.

COMPARISON OF THRESHOLDING ON DIFFERENT SCALES

Three candidates for thresholding:

$$B, \qquad U = (I - B)^{-1} \qquad \text{and} \qquad L = \log(U).$$

Contain the same information in different guises, B being the most interpretable. Once sparsity is sought, the sparse approximations to B, U and L place emphasis on different aspects.

COMPARISON OF THRESHOLDING ON DIFFERENT SCALES

B: Thresholding retains large direct effects.

U: Entries are sums of effects along all paths. Direct effects are absorbed in a composite.

Thresholding assumes paths of all lengths are equally important.

Potential implication: small number of near-zeros, and recovery of distant effects.

L: Thresholding retains large composite effects weighted by path length.

NUMERICAL EXPLORATION OF SPARSITY REGIMES

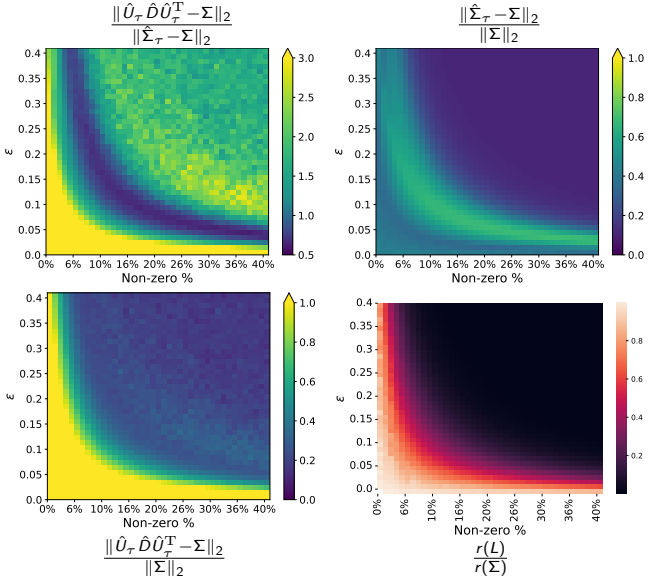
Data from Gaussian DAG with covariance $\Sigma = (I - B)^{-1}\Psi(I - B)^{-1}$.

Generate B by assigning value $\varepsilon > 0$ to a randomly selected prespecified percentage of entries. Other entries 0.

Threshold on different scales. Tuning parameter chosen by cross validation.

$r(A)$ measures (lack of) sparsity in metric used in thresholding literature.

Similar results for thresholding on scale of $L = \log(\Sigma)$.



OPEN QUESTIONS

- How one might test for sparsity across different scales.
- How to choose the sparsity scale empirically: e.g. traversal of parametrization space through convenient parametrized paths.
- More sophisticated estimators, e.g. in the vein of elegant work by Zwiernik (2025).

Thank you for your attention

References

The talk was based on:

- Battey, H. S. (2017). Eigen structure of a new class of structured covariance and inverse covariance matrices. *Bernoulli*, 23, 3166–3177.
- Rybak, J., Battey, H. S. and Bharath, K. (2025). Regression graphs and sparsity-inducing reparametrisations. *arXiv:2402.09112*.

Also cited:

- Cochran, W. G. (1938). The omission or addition of an independent variate in multiple linear regression. *Supplement to the J. Roy. Statist. Soc.*, 5, 171–176.
- Cox, D. R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs. *Statist. Sci.*, 8, 204–218.
- Rybak, J. and Battey, H. S. (2021). Sparsity induced by covariance transformation: some deterministic and probabilistic results. *Proc. Roy. Soc. London*, 477, 20200756.
- Wermuth, N. and Cox, D. R. (2004). Joint response graphs and separation induced by triangular systems. *J. R. Statist. Soc. B*, 66, 687–717.
- Zwiernik, P. (2025). Entropic covariance models. *Ann. Statist.*, to appear.