Joint Statistical Meetings 2024

Regression graphs and sparsity-inducing reparametrisations

Heather Battey
Department of Mathematics, Imperial College London

August 6, 2024

INDUCEMENT OF POPULATION-LEVEL SPARSITY

Battey (2023): one old and three "new" examples unified from this perspective.

Main point: in the absence of sparsity on physically-natural scales, preliminary manoeuvres may systematically induce a population-level sparsity on more abstract scales.

Precursor: Cox and Reid (1987).

Battey, H. S. (2023). Inducement of population-level sparsity. *Canad. J. Statist.* (*Festschrift for Nancy Reid*), 51, 760–768.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc. B*, 49, 1–39.

INDUCEMENT OF POPULATION-LEVEL SPARSITY

Two routes: reparametrization; transformations of the data.

Research questions concern:

- Traversal of parametrisation space (not parameter space) with a view to inducing sparsity.
- Traversal of data-transformation space ——— " ———
- Understanding how structure on the physically natural scale relates to sparsity in more abstract domains.

These questions concern structure at the population level and do not involve a notion of a sample.

Benefits of sparsity transfer to any reasonable methodology.

Regression graphs and sparsity-inducing reparametrisations

Jakub Rybak Karthik Bharath

arXiv:2402.09112

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

A^*: . . .

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.

REPARAMETRISATION

Starting from the physically natural representation in terms of $\sigma \in \mathsf{Cone}(p)$, a constrained space, consider reparametrisation to $\alpha \in \mathbb{R}^{p(p+1)/2}$ or $(\alpha, d) \in \mathbb{R}^{p(p+1)/2}$.

$$\begin{array}{ccc} \alpha & \xleftarrow[\phi_{pd}]{\hspace{2cm}} & \sigma \\ \Big\downarrow b_{sym} & & \Big\downarrow \mathbf{hvec}^{-1} \\ L \in \mathsf{Sym}(p) & \xrightarrow{e^L} & \mathsf{PD}(p) \ni \Sigma \end{array}$$

$$\begin{array}{ccc} \alpha & \xleftarrow[\phi_{lt}]{\hspace{2cm}} & \sigma \\ \Big\downarrow b_{lt} & & \Big\downarrow \mathbf{hvec}^{-1} \\ L \in \mathsf{LT}(p) & \xrightarrow{e^L (e^L)^{\mathrm{T}}} & \mathsf{PD}(p) \ni \Sigma \end{array}$$

$$\begin{array}{ccc} (\alpha, d) & \xleftarrow[\phi_{o}]{\hspace{2cm}} & \sigma \\ \Big\downarrow b_{sk} \Big\downarrow & & \Big\downarrow \mathbf{hvec}^{-1} \\ L, D \in \mathsf{Sk}(p) \times \mathsf{D}(p) & \xrightarrow{e^L e^D (e^L)^{\mathrm{T}}} & \mathsf{PD}(p) \ni \Sigma \end{array}$$

$$\begin{array}{ccc} (\alpha, d) & \xleftarrow[\phi_{ltu}]{\hspace{2cm}} & \sigma \\ \Big\downarrow b_{ltu} \Big\downarrow & & \Big\downarrow \mathbf{hvec}^{-1} \\ L, D \in \mathsf{LT_s}(p) \times \mathsf{D}(p) & \xrightarrow{e^L e^D (e^L)^{\mathrm{T}}} & \mathsf{PD}(p) \ni \Sigma \end{array}$$

STARTING POINT FOR ALL FOUR: THE MATRIX LOGARITHM

All four reparametrisations are based on the matrix logarithm L of a constrained matrix.

- L is implicitly defined through the matrix Taylor expansion:

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

- L belongs to to a vector space with a canonical basis $\mathcal{B} = \{B_1, \ldots, B_{\#\mathcal{B}}\}$.
- The unconstrained parameter α is the vector of coefficients in the basis expansion

$$L(\alpha) = \alpha_1 B_1 + \cdots \alpha_{|\mathcal{B}|} B_{|\mathcal{B}|}.$$

- Sparsity in the form $\|\alpha\|_0 = s^* < p$ respects positive definiteness and allows a fruitful analysis.

FOUR INITIAL REPARAMETRISATION MAPS

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

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

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

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

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).

OBJECTIVE OF THE WORK

For the new parametrisations:

- Uncover structure induced on physically natural scales through sparsity on the transformed scale;
- Ascertain the converse result: that matrices encoding such structure are sparse after reparametrisation.
- Ideally ascertain an interpretation for the zeros in α.

THE SIMPLEST CASE: Σ_{pd} (Battey, 2017)

$$\Sigma = \exp(L)$$
$$\Sigma^{-1} = \exp(-L)$$

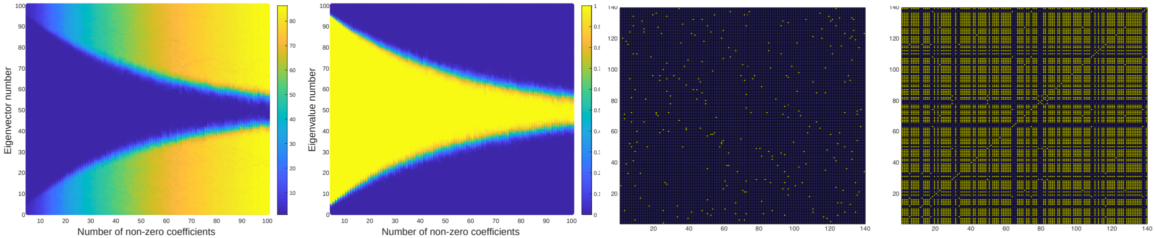
$$\Sigma,\ \Sigma^{-1} \in \mathrm{PD}(p) := \left\{ M \in \mathrm{M}(p) : M = M^T,\ M \succ 0 \right\} \quad \text{(open cone)}$$
$$L \in \mathrm{Sym}(p) := \left\{ M \in \mathrm{M}(p) : M = M^T \right\} \quad \text{(vector space)}.$$

Natural symmetric basis for Sym(p) of the form $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$:

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

THE SIMPLEST CASE: Σ_{pd} (Battey, 2017)

Sparsity of α in $L(\alpha) = \alpha_1 B_1 + \cdots \alpha_{|\mathcal{B}|} B_{|\mathcal{B}|}$ induces structure on Σ via the eigenvectors and eigenvalues.



Left and centre-left: simulation average of $\|\gamma_j\|_0$ (eigenvectors) and $\mathbb{I}\{\lambda_j = 1\}$ (eigenvalues) 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 $p = 100$.

Centre right: zero (blue) and non-zero (yellow) entries of a random sparse L.

Right: zero (blue) and non-zero (yellow) entries of $\exp(L)$.

THE FOUR REPARAMETRISATION MAPS

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

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

- Sparsity of α in the map $\alpha \mapsto \Sigma_{pd}(\alpha)$ has been studied (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.
- An encompassing formulation uncovers further parametrisations with a statistical interpretation.

THE FOUR REPARAMETRISATION MAPS

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

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

- Sparsity of α in the map $\alpha \mapsto \Sigma_{pd}(\alpha)$ has been studied (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.
- An encompassing formulation uncovers further parametrisations with a statistical interpretation.

THE FOUR REPARAMETRISATION MAPS

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

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

- Sparsity of α in the map $\alpha \mapsto \Sigma_{pd}(\alpha)$ has been studied (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.
- An encompassing formulation uncovers further parametrisations with a statistical interpretation.

THE FOUR REPARAMETRISATION MAPS

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

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

- Sparsity of α in the map $\alpha \mapsto \Sigma_{pd}(\alpha)$ has been studied (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.
- An encompassing formulation uncovers further parametrisations with a statistical interpretation.

THE FOUR REPARAMETRISATION MAPS

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

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

- Sparsity of α in the map $\alpha \mapsto \Sigma_{pd}(\alpha)$ has been studied (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.
- An encompassing formulation uncovers further parametrisations with a statistical interpretation.

In the four cases we characterise the structure such that structure $\iff$ α sparse.

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. $M^{\mathrm{T}} M = M M^{\mathrm{T}}$, then $d_r^* = d_c^* = d^*$.

EXAMPLE STRUCTURE OF $M = e^L$

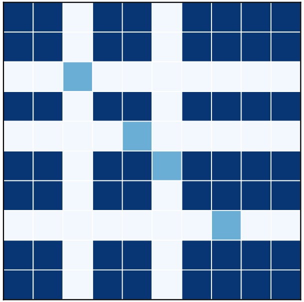


Figure: Example of a structure of M as as described on the last slide with $p = 10$, $d_r^* = 7$, $d_c^* = 8$ and $d^* = 9$. Zero, unit and unconstrained entries are light, medium and dark blue respectively.

The specific vector spaces of interest impose additional constraints.

INTERPRETATION OF SPARSITY-INDUCED STRUCTURES

In terms of the structure induced on the original scale by sparsity of $L(\alpha)$, $\alpha \mapsto \Sigma_{pd}(\alpha)$ and $\alpha \mapsto \Sigma_{ltu}(\alpha)$ represent two extremes...

BACKGROUND: 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.

BACKGROUND: INTERPRETATION OF BLOCKS

Let $Y = (Y_a^{\mathrm{T}}, Y_b^{\mathrm{T}})^{\mathrm{T}}$ be a mean-centred random vector with covariance matrix Σ, $\Pi_{b|a}$ is the matrix of regression coefficients of Y_a in a linear regression of Y_b on Y_a and $\Sigma_{bb.a}$ is the residual 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 marginalizing over the remaining variables in b.

BACKGROUND: MARGINALISATION AND CONDITIONING

Assume now that Y is Gaussian.

Marginalization over a variable in b, indicated by $\not{\not{o}}$, induces an edge between i and j if the marginalized variable is a transition node or a source node.

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

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

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

Σ_{ltu} FROM RECURSIVE BLOCK-TRIANGULARISATION

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 $\Sigma_{ltu} = Ue^{D}U^{\mathrm{T}}$ based on p blocks of size 1×1 where the general form of $U = e^{L}$ ignoring sparsity is

$$U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \beta_{2.1} & 1 & 0 & 0 \\ \beta_{3.1} & \beta_{3.21} & 1 & 0 \\ \beta_{4.1} & \beta_{4.21} & \beta_{4.3[2]} & 1 \end{pmatrix}, \qquad (1)$$

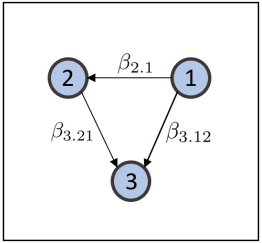
where for $j < i$, $U_{ij} = \beta_{i.j[j-1]}$ is the coefficient on Y_j in a linear regression of Y_i on $Y_1, \ldots, Y_j$. This is not new: it is implicit in Cox and Wermuth (1993, 2004).

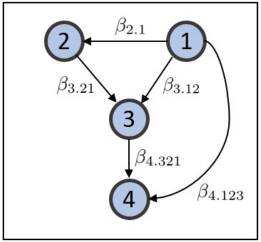
THE ENTRIES OF U IN $\Sigma_{ltu} = Ue^{D}U^{\mathrm{T}}$

Cochran's (1938) formula (stated here for three variables):

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

The total effect decomposes into a sum of partial effects.





For more than three variables, the total effect of variable j on variable i is the sum of effects along all directed paths connecting the two nodes.

STRUCTURE INDUCED ON $U = e^L$ BY SPARSITY OF L

Zeros in L via sparsity $\|\alpha\|_0 = s^*$ induce zero columns of the strictly lower-triangular matrix $U - I$.

Suppose that $U - I$ has a zero jth column. This implies that for every $i > j$, the only source or transition nodes connecting i and j are in the conditioning sets $[j-1]$ (otherwise dependence is induced through marginalization), and that there are no sink nodes among these conditioning variables (as conditioning on sink-nodes is edge-inducing).

STATISTICAL INTERPRETATION OF α in $\Sigma_{ltu}(\alpha)$

- Recall Cochran's (1938) formula for three variables:

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

  The total effect decomposes into a sum of partial effects.

- For more than three variables, the total effect of variable *j* on variable *i* is the sum of effects along all directed paths connecting the two nodes.
- The coefficient α has an interpretation in terms of a length-weighted sum of effects, with the weight inversely proportional to the length of the path.

EXACT VS APPROXIMATE ZEROS

Until now we have focused on exact zeros.

The interpretation of α provides an interpretation for approximate zeros and thereby clarifies the modelling implications of enforcing sparsity after reparametrisation:

In effect the relation between two variables would be declared null if relatively direct regression effects were negligible and other effects manifested through long paths.

FURTHER PARAMETRISATIONS

The paper discusses further parametrisations for which $\alpha \mapsto \Sigma_{pd}(\alpha)$ and $\alpha \mapsto \Sigma_{ltu}(\alpha)$ represent the extreme cases.

The resulting structures corresponds to the chain graph models (Andersson et al., 2001).

References

The talk was based on:

- Rybak, J., Battey, H. S. and Bharath, K. (2024). Regression graphs and sparsity-inducing reparametrisations. *arXiv:2402.09112*.

Also cited:

- Battey, H. S. (2017). Eigen structure of a new class of structured covariance and inverse covariance matrices. *Bernoulli*, 23, 3166–3177.
- Battey, H. S. (2023). Inducement of population sparsity. *Canadian J. Statist. (Festschrift for Nancy Reid)*, 51, 760–768.
- 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 Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc. B*, 19, 1–39.
- 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.