Some perspectives on separable covariance matrices

J. Virta

University of Turku

ICS and Related Methods
Helsinki, 6.-8. May 2025

Table of Contents

Separability

A random vector $\mathbf{x} \in \mathbb{R}^{p_1 p_2}$ is said to be covariance separable if $\mathbf{\Sigma} := \mathrm{Cov}(\mathbf{x})$ satisfies

$$\mathbf{\Sigma} = \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1,$$

for some positive definite $\mathbf{\Sigma}_1 \in \mathbb{R}^{p_1 \times p_1}$, $\mathbf{\Sigma}_2 \in \mathbb{R}^{p_2 \times p_2}$.

- $\otimes$ is the Kronecker product,

$$\mathbf{B} \otimes \mathbf{A} = \begin{pmatrix} b_{11}\mathbf{A} & \cdots & b_{1p_2}\mathbf{A} \\ \vdots & \ddots & \vdots \\ b_{p_2 1}\mathbf{A} & \cdots & b_{p_2 p_2}\mathbf{A} \end{pmatrix}$$

Matrix-valued data

- The vectorization of a matrix

$$\mathbf{X} = (\mathbf{x}_1 \mid \cdots \mid \mathbf{x}_{p_2}) \in \mathbb{R}^{p_1 \times p_2}$$

is the vector

$$\mathrm{vec}(\mathbf{X}) = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{p_2} \end{pmatrix} \in \mathbb{R}^{p_1 p_2}.$$

- Separability ensues naturally under vectorized matrix data.

Matrix normal

Matrix normal distribution

If the elements of $\mathbf{Z} \in \mathbb{R}^{p_1 \times p_2}$ are i.i.d. from $\mathcal{N}(0,1)$, then $\mathbf{X} = \boldsymbol{\Sigma}_1^{1/2} \mathbf{Z} \boldsymbol{\Sigma}_2^{1/2}$ is said to have the matrix normal distribution

$$\mathcal{N}_{p_1,p_2}(\mathbf{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2).$$

- For $\mathbf{X} \sim \mathcal{N}_{p_1,p_2}(\mathbf{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$, we have

$$\mathrm{Cov}\{\mathrm{vec}(\mathbf{X})\} = \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1.$$

Interpretation of parameters

- The covariance matrix of a column of $\mathbf{X} \sim \mathcal{N}_{p_1,p_2}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$ is

$$\mathrm{Cov}(\mathbf{x}_k) = \sigma_{2kk}\mathbf{\Sigma}_1,$$

and analogously for the rows of $\mathbf{X}$.

Interpretation

$$\mathbf{\Sigma}_1 = \text{ dependencies between the rows of } \mathbf{X},$$
$$\mathbf{\Sigma}_2 = \text{ dependencies between the columns of } \mathbf{X}.$$

Research activity

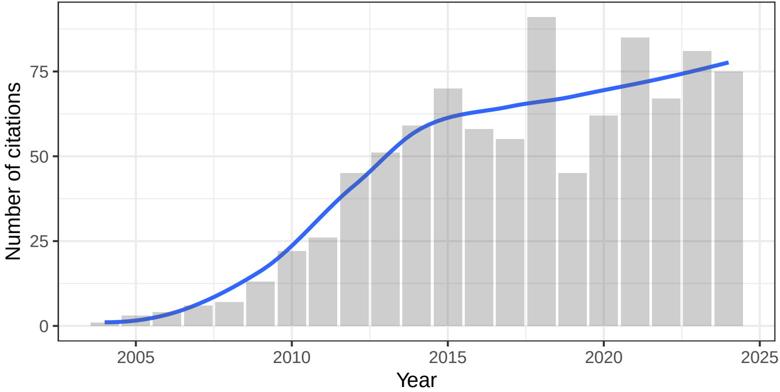


Figure: the aggregated yearly citations to [Dutilleul, 1999, Srivastava et al., 2008, Werner et al., 2008, Wiesel, 2012]

Table of Contents

Non-identifiability of scale

- From $\mathbf{X} = \mathbf{\Sigma}_1^{1/2}\mathbf{Z}\mathbf{\Sigma}_2^{1/2}$ we see that the choices

$$(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \qquad \text{and} \qquad (\lambda\mathbf{\Sigma}_1, \lambda^{-1}\mathbf{\Sigma}_2)$$

  yield the same distribution for $\mathbf{X}$.
- The parameters $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2$ are identifiable only up to scale.

Estimator I

- The moment estimators

$$\mathrm{E}(\mathbf{X}\mathbf{X}') = \mathrm{tr}(\boldsymbol{\Sigma}_2)\boldsymbol{\Sigma}_1$$
$$\mathrm{E}(\mathbf{X}'\mathbf{X}) = \mathrm{tr}(\boldsymbol{\Sigma}_1)\boldsymbol{\Sigma}_2$$

are used in numerous works
  - [Zhang and Zhou, 2005]
  - [Li et al., 2010]
  - [Ding and Cook, 2015]
  - [Virta et al., 2017]
  - [Radojičić et al., 2025]
  - . . .

Estimator I - properties

- Their sample versions

$$\mathbf{C}_1 := \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i' \quad \text{and} \quad \mathbf{C}_2 := \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{X}_i$$

  are

  - conceptually simple,
  - fast to compute,
  - consistent and asymptotically normal,
  - orthogonally equivariant, i.e., for orthogonal $\mathbf{U}, \mathbf{V}$,

$$\mathbf{X}_i \mapsto \mathbf{U}\mathbf{X}_i\mathbf{V}' \quad \text{induces} \quad (\mathbf{C}_1, \mathbf{C}_2) \mapsto (\mathbf{U}\mathbf{C}_1\mathbf{U}', \mathbf{V}\mathbf{C}_2\mathbf{V}')$$

Affine equivariance?

Conjecture in [Virta et al., 2017]

No estimator $\mathbf{S}_1(\mathbf{X}_i)$ exists for which

$$\mathbf{S}_1(\mathbf{A}\mathbf{X}_i\mathbf{B}') \propto \mathbf{A}\mathbf{S}_1(\mathbf{X}_i)\mathbf{A}',$$

for all non-singular $\mathbf{A}, \mathbf{B}$.

Affine equivariance?

Conjecture in [Virta et al., 2017]

No estimator $\mathbf{S}_1(\mathbf{X}_i)$ exists for which

$$\mathbf{S}_1(\mathbf{A}\mathbf{X}_i\mathbf{B}') \propto \mathbf{A}\mathbf{S}_1(\mathbf{X}_i)\mathbf{A}',$$

for all non-singular $\mathbf{A}, \mathbf{B}$.

Not true!

Estimator II

MLE of $(\mathbf{\Sigma}_1, \mathbf{\Sigma_2})$ under matrix normal

The maximum likelihood estimator $(\mathbf{S}_1, \mathbf{S}_2)$ of $(\mathbf{\Sigma}_1, \mathbf{\Sigma_2})$ is any solution to

$$\begin{cases} \mathbf{S}_1 & = \frac{1}{np_2} \sum_{i=1}^n \mathbf{X}_i \mathbf{S}_2^{-1} \mathbf{X}_i', \\ \mathbf{S}_2 & = \frac{1}{np_1} \sum_{i=1}^n \mathbf{X}_i' \mathbf{S}_1^{-1} \mathbf{X}_i. \end{cases}$$

Estimator II - properties

- The sample MLE is
  - affine equivariant (up to scale),
  - a.s. unique for continuous $\mathbf{X}$ when $n > \frac{p_1}{p_2} + \frac{p_2}{p_1} + 1$,
  - a solution to g-convex problem.
- Moreover, under normality
  - the MLE is strictly more efficient than the moment estimator when $\mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 \not\propto \mathbf{I}_{p_1 p_2}$,
  - The scaled estimators

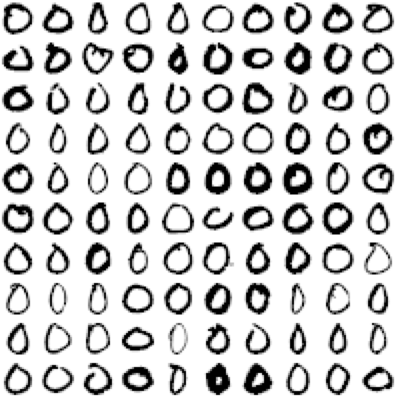
$$\frac{\mathbf{S}_1}{|\mathbf{S}_1|^{1/p_1}} \quad \text{and} \quad \frac{\mathbf{S}_2}{|\mathbf{S}_2|^{1/p_2}}$$

    are asymptotically independent.

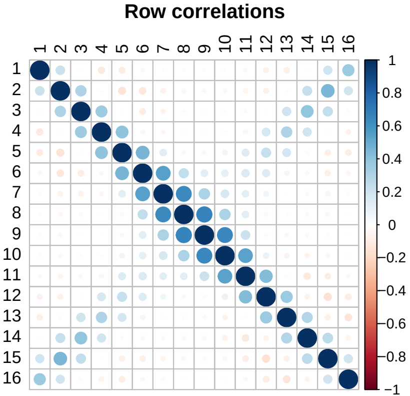
[Wiesel, 2012, Soloveychik and Trushin, 2016, Drton et al., 2021, McCormack and Hoff, 2023]

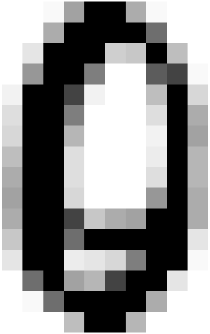
Example data

$\mathbf{X}_i \in [0,1]^{p_1 \times p_2}$ with $n = 1194$, $p_1 = p_2 = 16$.

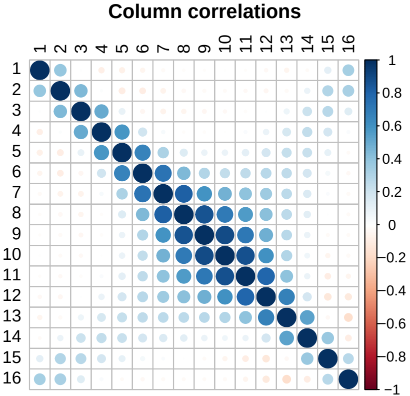


MLE $\mathbf{S}_1$ for the digits





MLE $\mathbf{S}_2$ for the digits



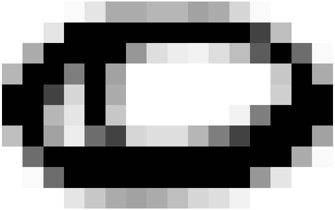


Table of Contents

Motivation

- Under $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1$, the number of parameters is reduced from $\mathcal{O}(p_1^2 p_2^2)$ to $\mathcal{O}(p_1^2 + p_2^2)$.
- Separability testing typically uses Gaussian LRT:
  - [Lu and Zimmerman, 2005]
  - [Mitchell et al., 2005]
  - [Roy and Khattree, 2005]
  - [Simpson, 2010]
  - [Filipiak et al., 2016]
  - $\cdots$

A wider model

Matrix elliptical distributions

We assume

$$\mathbf{X} = \mathbf{\Sigma}_1^{1/2}\mathbf{Z}\mathbf{\Sigma}_2^{1/2},$$

where $\mathbf{Z}$ is matrix spherical, i.e., $\mathbf{Z} \sim \mathbf{UZV}'$ for all orthogonal $\mathbf{U}, \mathbf{V}$.

- All matrix elliptical distributions are separable.
- The family contains matrix normal, matrix-t, ... [Gupta and Nagar, 2018]

Separability test

Test statistic

For the sample covariance matrix $\mathbf{S}$ of the $\mathrm{vec}(\mathbf{X}_i)$, we let

$$\hat{t} := \left\| \left(\frac{\mathbf{S}}{|\mathbf{S}|^{\frac{1}{p_1 p_2}}} \right)^{-1/2} \left(\frac{\mathbf{S}_2}{|\mathbf{S}_2|^{\frac{1}{p_2}}} \otimes \frac{\mathbf{S}_1}{|\mathbf{S}_1|^{\frac{1}{p_1}}} \right) \left(\frac{\mathbf{S}}{|\mathbf{S}|^{\frac{1}{p_1 p_2}}} \right)^{-1/2} - \mathbf{I}_{p_1 p_2} \right\|_F^2 .$$

$\mathrm{Cov}\{\mathrm{vec}(\mathbf{Z}) \otimes \mathrm{vec}(\mathbf{Z})\}$ when $p_1 = p_2 = 2$

$$
\left(\begin{array}{cccc|cccc|cccc|cccc}
* & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * \\
0 & * & 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 & 0 & * & 0 \\
0 & 0 & * & 0 & 0 & 0 & 0 & * & * & 0 & 0 & 0 & 0 & * & 0 & 0 \\
0 & 0 & 0 & * & 0 & 0 & * & 0 & 0 & * & 0 & 0 & * & 0 & 0 & 0 \\
\hline
0 & * & 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 & 0 & * & 0 \\
* & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * \\
0 & 0 & 0 & * & 0 & 0 & * & 0 & 0 & * & 0 & 0 & * & 0 & 0 & 0 \\
0 & 0 & * & 0 & 0 & 0 & 0 & * & * & 0 & 0 & 0 & 0 & * & 0 & 0 \\
\hline
0 & 0 & * & 0 & 0 & 0 & 0 & * & * & 0 & 0 & 0 & 0 & * & 0 & 0 \\
0 & 0 & 0 & * & 0 & 0 & * & 0 & 0 & * & 0 & 0 & * & 0 & 0 & 0 \\
* & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * \\
0 & * & 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 & 0 & * & 0 \\
\hline
0 & 0 & 0 & * & 0 & 0 & * & 0 & 0 & * & 0 & 0 & * & 0 & 0 & 0 \\
0 & 0 & * & 0 & 0 & 0 & 0 & * & * & 0 & 0 & 0 & 0 & * & 0 & 0 \\
0 & * & 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 & 0 & * & 0 \\
* & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & *
\end{array}\right)
$$

The starred elements are parametrized by a total of three values.

Null distribution

Limiting distribution of $\hat{t}$ [Virta and Matsuda, 2025]

Under the matrix elliptical, $\hat{t}$ satisfies

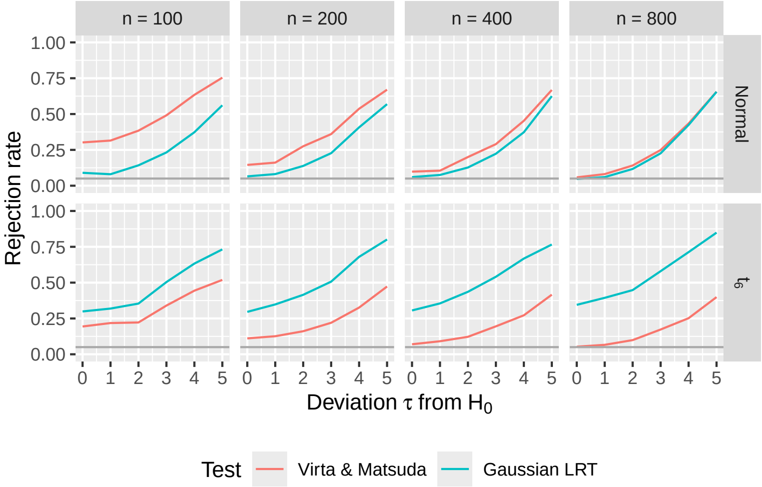
$$n\hat{t} \rightsquigarrow \beta_1 \chi^2_{\frac{1}{4}(p_1+2)(p_1-1)(p_2+2)(p_2-1)} + \beta_2 \chi^2_{\frac{1}{4}p_1 p_2 (p_1-1)(p_2-1)},$$

where $\beta_1, \beta_2 > 0$ are certain fourth moments of $\mathbf{Z}$.

- For matrix normal, $\beta_1 = \beta_2 = 2$, giving just a χ^2-distribution.
- The total degrees of freedom is equal to the difference between the numbers of parameters in $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1$.

Power comparison

- Average rejection rates under local alternatives of the form "$\tau/\sqrt{n}$" when $n = 500$



Bootstrapped null distribution

- If data are unlikely to obey an elliptical distribution, we can simulate the null distribution of $\hat{t}$ as follows:
  1. Whiten the data using $\mathbf{S}^{-1/2}$.
  2. Draw a bootstrap sample.
  3. Backtransform via $(\mathbf{S}_2 \otimes \mathbf{S}_1)^{1/2}$.
  4. Compute $\hat{t}$.

Example

- For the digit data we get $p < 10^{-4}$ regardless of the null distribution (asymptotic or bootstrap).
- We conclude that the distribution is not separable.

Table of Contents

Kronecker covariance

Kronecker covariance

For a random matrix $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2}$ with arbitrary $\mathrm{Cov}\{\mathrm{vec}(\mathbf{X})\}$, [Hoff et al., 2023] define the Kronecker covariance of $\mathbf{X}$ to be

$$k(\mathbf{X}) := \mathbf{\Psi}_2 \otimes \mathbf{\Psi}_1,$$

where $(\mathbf{\Psi}_1, \mathbf{\Psi}_2)$ is the Gaussian MLE evaluated at $\mathbf{X}$.

Kronecker covariance properties

- The Kronecker covariance $k(\mathbf{X})$
  - is equal to $\mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1$ when $\mathbf{X}$ is covariance separable,
  - is a.s. unique under mild conditions,
  - is a solution to g-convex problem,
  - is affine equivariant in the sense that

$$k(\mathbf{AXB}') = (\mathbf{A} \otimes \mathbf{B}) k(\mathbf{X}) (\mathbf{A}' \otimes \mathbf{B}').$$

Core covariance

Core covariance

The core covariance of $\mathbf{X}$ is defined as

$$c(\mathbf{X}) = k(\mathbf{X})^{-1/2}\mathrm{Cov}\{\mathrm{vec}(\mathbf{X})\}k(\mathbf{X})^{-1/2}.$$

- The core covariance $c(\mathbf{X})$ is
  - equal to $\mathbf{I}_{p_1 p_2}$ when $\mathbf{X}$ is covariance separable,
  - affine invariant in the sense that

$$c(\mathbf{AXB}') = (\mathbf{U} \otimes \mathbf{V})c(\mathbf{X})(\mathbf{U}' \otimes \mathbf{V}'),$$

for some orthogonal $\mathbf{U}, \mathbf{V}$.

Kronecker-core decomposition

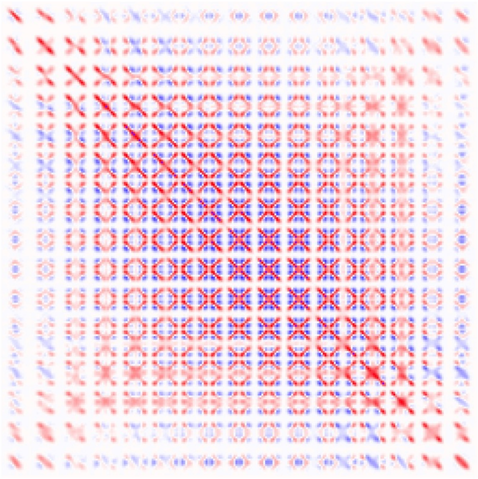
- The Kronecker-core decomposition (KCD) is

$$\mathrm{Cov}\{\mathrm{vec}(\mathbf{X})\} = k(\mathbf{X})^{1/2} c(\mathbf{X}) k(\mathbf{X})^{1/2}.$$

Interpretation

$$\begin{aligned} k(\mathbf{X}) &= \text{ row-column variation / “main effects”}, \\ c(\mathbf{X}) &= \text{ residual variation / “interactions”}. \end{aligned}$$

$\mathrm{Cov}\{\mathrm{vec}(\mathbf{X})\}$ for the digits



K and **C** for the digits

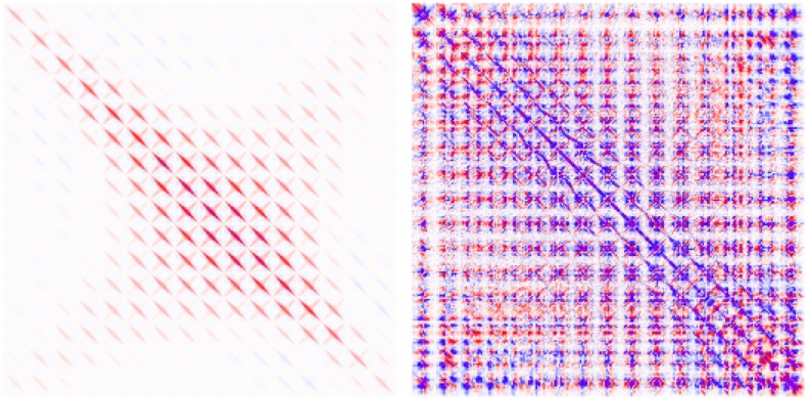
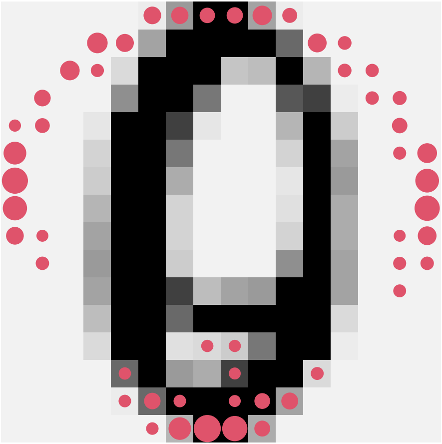


Figure: **K** (left) and **C** (right)

Violations 1

- Pixels whose variances most violate separability



Violations 2

- Pixel pairs whose positive correlations most violate separability

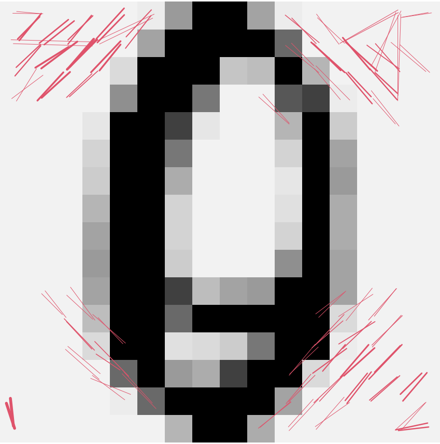


Table of Contents

List of R-packages

- Gaussian MLE / hypothesis testing
  - `MixMatrix`
  - `tensr`
  - `MN`
  - `sEparaTe`
- Non-Gaussian matrix-variate distributions
  - `MixMatrix`
- Kronecker-core decomposition
  - `covKCD`
- General operations for matrix/tensor data
  - `tensorBSS`
  - `tensr`

Future prospects

- General theory of scatter matrices for matrix data
- Matrix ICS
- M-estimators of scatter for matrix data
- Robust analysis of matrix data
- ...

Thank you for your attention!

References I

Ding, S. and Cook, R. D. (2015).
Tensor sliced inverse regression.
Journal of Multivariate Analysis, 133:216–231.

Drton, M., Kuriki, S., and Hoff, P. (2021).
Existence and uniqueness of the kronecker covariance MLE.
The Annals of Statistics, 49(5):2721–2754.

Dutilleul, P. (1999).
The MLE algorithm for the matrix normal distribution.
Journal of Statistical Computation and Simulation, 64(2):105–123.

Filipiak, K., Klein, D., and Roy, A. (2016).
Score test for a separable covariance structure with the first component as compound symmetric correlation matrix.
Journal of Multivariate Analysis, 150:105–124.

References II

Gupta, A. K. and Nagar, D. K. (2018).
Matrix Variate Distributions.
Chapman and Hall/CRC.

Hoff, P., McCormack, A., and Zhang, A. R. (2023).
Core shrinkage covariance estimation for matrix-variate data.
Journal of the Royal Statistical Society Series B: Statistical Methodology, 85(5):1659–1679.

Li, B., Kim, M. K., and Altman, N. (2010).
On dimension folding of matrix-or array-valued statistical objects.
Annals of Statistics, 38(2):1094–1121.

Lu, N. and Zimmerman, D. L. (2005).
The likelihood ratio test for a separable covariance matrix.
Statistics & Probability Letters, 73(4):449–457.

References III

McCormack, A. and Hoff, P. (2023).
Information geometry and asymptotics for kronecker covariances.
arXiv preprint arXiv:2308.02260.

Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2005).
Testing for separability of space–time covariances.
Environmetrics: The Official Journal of the International Environmetrics Society, 16(8):819–831.

Radojičić, U., Lietzén, N., Nordhausen, K., and Virta, J. (2025).
Order determination for tensor-valued observations using data augmentation.
Journal of Computational and Graphical Statistics.
To appear.

References IV

Roy, A. and Khattree, R. (2005).
On implementation of a test for Kronecker product covariance structure for multivariate repeated measures data.
Statistical Methodology, 2(4):297–306.

Simpson, S. L. (2010).
An adjusted likelihood ratio test for separability in unbalanced multivariate repeated measures data.
Statistical Methodology, 7(5):511–519.

Soloveychik, I. and Trushin, D. (2016).
Gaussian and robust Kronecker product covariance estimation: Existence and uniqueness.
Journal of Multivariate Analysis, 149:92–113.

References V

Srivastava, M. S., von Rosen, T., and von Rosen, D. (2008).
Models with a kronecker product covariance structure: estimation and testing.
Mathematical Methods of Statistics, 17:357–370.

Virta, J., Li, B., Nordhausen, K., and Oja, H. (2017).
Independent component analysis for tensor-valued data.
Journal of Multivariate Analysis, 162:172–192.

Virta, J. and Matsuda, T. (2025).
Asymptotic test of covariance separability for matrix elliptical data.
Work under progress.

References VI

Werner, K., Jansson, M., and Stoica, P. (2008).
On estimation of covariance matrices with Kronecker product structure.
IEEE Transactions on Signal Processing, 56(2):478–491.

Wiesel, A. (2012).
Geodesic convexity and covariance estimation.
IEEE Transactions on Signal Processing, 60(12):6182–6189.

Zhang, D. and Zhou, Z.-H. (2005).
$(2D)^2$2PCA: Two-directional two-dimensional pca for efficient face representation and recognition.
Neurocomputing, 69(1-3):224–231.