Statistical Papers (2024) 65:5835–5849
https://doi.org/10.1007/s00362-024-01610-9

REGULAR ARTICLE

Ridge-type covariance and precision matrix estimators of the multivariate normal distribution

Wessel N. van Wieringen[1,2] · **Gwenaël G. R. Leday**[3]

Received: 11 October 2023 / Revised: 29 August 2024 / Accepted: 5 September 2024 /
Published online: 21 October 2024

Abstract
We consider ridge-type estimation of the multivariate normal distribution's covariance matrix and its inverse, the precision matrix. While several ridge-type covariance and precision matrix estimators have been presented in the literature, their respective inverses are often not considered as precision and covariance matrix estimators even though their estimands are one-to-one related through the matrix inverse. We study which estimator is to be preferred in what case. Hereto we compare the ridge-type covariance matrix estimators and their properties to that of the inverse of the ridge-type precision matrix estimators, and vice versa. The comparison, in which we take all ridge-type estimators along, is limited to a specific case that is illustrative of the difference between the covariance and precision matrix estimators. The comparison addresses the estimators' estimating equation, analytic expression, analytic properties like positive definiteness and penalization limit, mean squared error, consistency, Bayesian formulation, and their loss and potential for marginal and partial correlation screening.

Keywords Bayesian estimation · Consistency · Correlation screening · Mean squared error · Multivariate normal distribution

✉ Wessel N. van Wieringen
w.vanwieringen@amsterdamumc.nl

Gwenaël G. R. Leday
gwenael.leday@gmail.com

[1] Department of Epidemiology and Data Science, Amsterdam UMC, Location Vrije Universiteit Amsterdam, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands

[2] Department of Mathematics, Vrije Universiteit Amsterdam, De Boelelaan 1111, 1081 HV Amsterdam, The Netherlands

[3] Biometris, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

1 Introduction

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be p-variate random variables drawn from the zero-centered multivariate normal distribution $\mathcal{N}(\mathbf{0}_p, \boldsymbol{\Sigma})$ with covariance matrix $\boldsymbol{\Sigma}$. The maximum likelihood estimator $\widehat{\boldsymbol{\Sigma}}$ of the parameter of this distribution is the sample covariance matrix $\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y}_i \mathbf{Y}_i^\top$. If $p > n$, the maximum likelihood estimator is singular. This singularity is typically overcome through either shrinkage, i.e. adding a positive definite matrix to $\mathbf{S}$, or the use of a penalized estimator of the covariance matrix. The latter maximizes the loglikelihood augmented with a penalty. The penalty—be it lasso, ridge or other—is commonly put on the precision matrix, the inverse of the covariance matrix (see e.g. Friedman et al. (2008) and Van Wieringen and Peeters (2016)). Here we investigate whether the ℓ_2 penalty is best put on the covariance or precision matrix or what type of shrinkage estimator is to be preferred.

Penalized covariance matrix estimators result from penalties put on the covariance matrix. While lasso or shrinkage estimators of the covariance matrix exist (cf. Bien and Tibshirani (2011), Ledoit and Wolf (2004), respectively), an ℓ_2-penalized counterpart has not yet been reported. To facilitate our comparison, we address this omission here.

Penalized precision matrix estimators, resulting from a penalty on the precision matrix, have been reported (Friedman et al. 2008; Van Wieringen and Peeters 2016). The ℓ_2-penalized precision matrix estimator maximizes the loglikelihood augmented by an ℓ_2 penalty on the precision matrix:

$$\log(|\boldsymbol{\Omega}|) - \mathrm{tr}(\boldsymbol{\Omega}\mathbf{S}) - \tfrac{1}{2}\lambda_\omega \mathrm{tr}[(\boldsymbol{\Omega} - \mathbf{T}_\omega)^\top(\boldsymbol{\Omega} - \mathbf{T}_\omega)], \tag{1}$$

with inverse covariance matrix a.k.a. precision matrix $\boldsymbol{\Omega}$, penalty parameter λ_ω, and non-random target matrix $\mathbf{T}_\omega \in \mathcal{S}_+^p$, where $\mathcal{S}_+^p$ is the class of nonnegative definite, symmetric, $p \times p$-dimensional matrices. The latter summand of the preceding display is the ℓ_2-penalty as it equals $\frac{1}{2}\lambda_{\boldsymbol{\omega}}\|\boldsymbol{\Omega} - \mathbf{T}_\omega\|_F^2$. For increasing values of λ_ω, the ℓ_2-penalized precision matrix estimator is shrunken towards the target matrix $\mathbf{T}_\omega$. An analytic expression for the maximizer of this loss (1) exists:

$$\widehat{\boldsymbol{\Omega}}(\lambda_\omega) = \{\tfrac{1}{2}(\mathbf{S} - \lambda_\omega\mathbf{T}_\omega) + [\lambda_\omega\mathbf{I}_{pp} + \tfrac{1}{4}(\mathbf{S} - \lambda_\omega\mathbf{T}_\omega)^2]^{1/2}\}^{-1}. \tag{2}$$

Van Wieringen and Peeters (2016) remark that the inverse of this ℓ_2-penalized precision matrix estimator can be used as a covariance matrix estimator. While true, this inversion does not yield the ℓ_2-penalized covariance matrix estimator (as is accidentally suggested in Van Wieringen and Peeters (2016)), which is presented here. This, however, does demand for a comparison between the ridge-type covariance matrix estimators.

The aforementioned ridge-type estimators are 'ridge' as both employ an ℓ_2-type penalty, but other ridge-type estimators based on shrinkage have been proposed. Warton (2008) proposes the ridge-type covariance estimator: $\widehat{\boldsymbol{\Sigma}}(\lambda_a) = \mathbf{S} + \lambda_a\mathbf{I}_{pp}$ with penalty parameter $\lambda_a > 0$. This estimator circumvents the singularity of the sample covariance matrix and provides a stable and positive definite covariance estimate. The ridge covariance estimator of Warton (2008) is 'ridge' as it adds a ridge, i.e. $\lambda_a\mathbf{I}_{pp}$, to the diagonal. In a similar vein, but motivated by the work on James-Stein estimators

(James and Stein 1961), Ledoit and Wolf (2004) present the ridge-type covariance estimator: $\widehat{\mathbf{\Sigma}}(\theta_\ell) = (1-\theta_\ell)\mathbf{S} + \theta_\ell \mathbf{T}_\ell$ with weight parameter $\theta_\ell \in [0, 1]$ and positive definite target matrix $\mathbf{T}_\ell$. Both estimators, $\widehat{\mathbf{\Sigma}}(\lambda_a)$ and $\widehat{\mathbf{\Sigma}}(\theta_\ell)$, and their inverses are taken along in the comparison of ridge-type covariance and precision estimators.

To stress the use of notation we denote covariance and precision matrix estimators by $\widehat{\mathbf{\Sigma}}(\cdot)$ and $\widehat{\mathbf{\Omega}}(\cdot)$. Their arguments and corresponding subscripts determine the particular variant, which have different functional forms. For instance, $\widehat{\mathbf{\Sigma}}(\lambda_a)$ and $\widehat{\mathbf{\Sigma}}(\theta_\ell)$ are both covariance estimators, but the first one is that introduced by Warton (2008) while the second by Ledoit and Wolf (2004) (see the preceding paragraph).

The ridge-type covariance and precision matrix estimators are not only of theoretical interest. They may be used to screen marginal pairwise dependencies (Luo et al. 2014). Hero and Rajaratnam (2011) showed the inflation (i.e. over-estimation) of the sample correlation when dimension p approaches the sample size n or even exceeds it. This may be countered by employing a penalized or shrinkage estimator of the covariance matrix. Both estimator types may improve the screening of marginal pairwise dependencies.

In this work we study ridge-type estimators of the covariance and precision matrix of the multivariate normal distribution. The ridge-type covariance and precision matrix estimators' estimands are one-to-one related through the matrix inverse. The matrix inverse of covariance/precision matrix estimator can thus be considered an estimator of the precision/covariance matrix. Here we are primarily concerned with the assessment of which ridge-type estimator to use in what case. As a secondary goal we develop the hirtherto unreported ℓ_2-penalized covariance matrix estimator (defined in the next section). While the development of the ℓ_2-penalized covariance matrix estimator is the thread in this work, to answer our primary question we compare this estimator throughout its development to the other ridge-type estimators. The development of the ℓ_2-penalized covariance matrix estimator starts from the definition of its loss function and the study of its estimating equation. The sought estimator is then found by solving the estimating equation. In the subsequent section analytic properties of the estimator are proven. We then discuss its mean squared error and consistency and provide a Bayesian formulation. These aspects are contrasted to that of the other ridge-type estimators. Finally, we compare in silico the performance of all ridge-type covariance and precision matrix estimators in terms of Frobenius and quadratic loss and with respect to their marginal and conditional correlation screening behavior. The paper concludes with an overview of the differences between the various covariance and precision matrix estimators.

2 The estimating equation

The ℓ_2-penalized covariance matrix estimator minimizes the log-likelihood augmented with a ridge penalty of the elements of the covariance matrix:

$$\widehat{\mathbf{\Sigma}}(\lambda_\sigma) = \arg \max_{\mathbf{\Sigma} \in \mathcal{S}^p_{++}} -\log(|\mathbf{\Sigma}|) - \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}) - \tfrac{1}{2}\lambda_\sigma \mathrm{tr}[(\mathbf{\Sigma} - \mathbf{T}_\sigma)^\top(\mathbf{\Sigma} - \mathbf{T}_\sigma)], \quad (3)$$

where—for clarity—the penalty parameter λ_σ and target matrix $\mathbf{T}_\sigma$ are now indexed by a σ where $\mathcal{S}_{++}^p$ is the class of positive definite, symmetric, $p \times p$-dimensional matrices. To see that this is indeed an ℓ_2-penalized estimator note that the penalty can be reformulated as $\frac{1}{2}\lambda_\sigma \|\text{vec}(\mathbf{\Sigma}) - \text{vec}(\mathbf{T}_\sigma)\|_2^2$, where the vec-operator stacks the columns of its arguments into a vector. In contrast to the loss function of the ℓ_2-penalized precision matrix estimator, the loss function (3) for its covariance counterpart is not strictly concave. This is due to the fact that the loss function is a sum of a convex and strictly concave function, and an argumentation provided in Bien and Tibshirani (2011). This complicates the search for its maximizer. In contrast, the loss function (1) of the ℓ_2-penalized precision matrix estimator is strictly concave with an analytic solution. While the other ridge-type estimators, $\widehat{\mathbf{\Sigma}}(\lambda_a)$ and $\widehat{\mathbf{\Sigma}}(\theta_\ell)$, have originally not been motivated from a loss function, they do optimize a strictly concave one (see Supplementary Material) with analytic expressions of the optimum provided in the introduction.

The dual formulation of ℓ_2-penalized optimization problem amounts to a constraint estimation problem. The elements of $\mathbf{\Sigma}$ can be expressed in terms of the elements of its inverse $\mathbf{\Omega}$. Moreover, so can the penalty induced parameter constraints of the ℓ_2-penalized covariance and precision matrix estimation problems. This allows us to compare them, which is done in the next toy example.

Example 1 Assume the precision matrix $\mathbf{\Omega}$ to be of the form $(\omega_d - \omega_0)\mathbf{I}_{pp} + \omega_0 \mathbf{1}_{pp}$ with $\mathbf{I}_{pp}$ is the identity matrix and $\mathbf{1}_{pp}$ the $p \times p$-dimensional matrix comprising only ones. The parameters of $\mathbf{\Sigma}$ are then by the Sherman-Morrison formula too expressible in terms of ω_d and ω_0: $\mathbf{\Sigma} = (\omega_d - \omega_0)^{-1}\{\mathbf{I}_{pp} - \omega_0[\omega_d + (p-1)\omega_0]^{-1}\mathbf{1}_{pp}\}$. Moreover, so are the penalty induced parameter constraints:

$$\big\{(\omega_d, \omega_0) \in \mathbb{R}^2 \ : \ \mathbf{\Omega}(\omega_d, \omega_0) \succ 0, \|\mathbf{\Omega}(\omega_d, \omega_0)\|_F^2 = p\omega_d^2 + p(p-1)\omega_0^2 \leq c(\lambda_\omega)\big\}$$

and

$$\big\{(\omega_d, \omega_0) \in \mathbb{R}^2 \ : \ \mathbf{\Sigma}(\omega_d, \omega_0) \succ 0, \|\mathbf{\Sigma}(\omega_d, \omega_0)\|_F^2 = p(\omega_d - \omega_0)^{-2}\{1 - \omega_0[(p-2)\omega_0 + 2\omega_d][(p-1)\omega_0 + \omega_d]^{-2}\} \leq c(\lambda_\sigma)\big\},$$

where for clarity the dependence of $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ on ω_d and ω_0 has temporarily been explicated and the target matrices $\mathbf{T}_\sigma$ and $\mathbf{T}_\omega$ set equal to $\mathbf{0}_{pp}$. Figure 1 shows these sets for $p = 3$ and with $c(\lambda_\omega)$ and $c(\lambda_\sigma)$ scaled to have domains of comparable size. The most striking difference is the shape of the domains. The ℓ_2-penalty of the precision matrix implies a convex domain, while its covariance counterpart is clearly nonconvex. This nonconvexity complicates the maximization of the likelihood.

The estimating equation for the ℓ_2-penalized covariance matrix is found by equating the derivative of its loss function (3) with respect to $\mathbf{\Sigma}$ to zero:

$$-\mathbf{\Sigma}^{-1} + \mathbf{\Sigma}^{-1}\mathbf{S}\mathbf{\Sigma}^{-1} - \lambda_\sigma(\mathbf{\Sigma} - \mathbf{T}_\sigma) = \mathbf{0}_{pp}.$$

For comparison with the other ridge-type estimators, it is more insightful (as it yields a polynomial in $\mathbf{\Sigma}$) to pre- and post-multiply the estimating equation in the preceding display by $\mathbf{\Sigma}$. This gives:

$$\mathbf{0}_{pp} = -\mathbf{\Sigma} + \mathbf{S} - \lambda_\sigma(\mathbf{\Sigma}^3 - \mathbf{\Sigma}\mathbf{T}_\sigma\mathbf{\Sigma}). \tag{4}$$

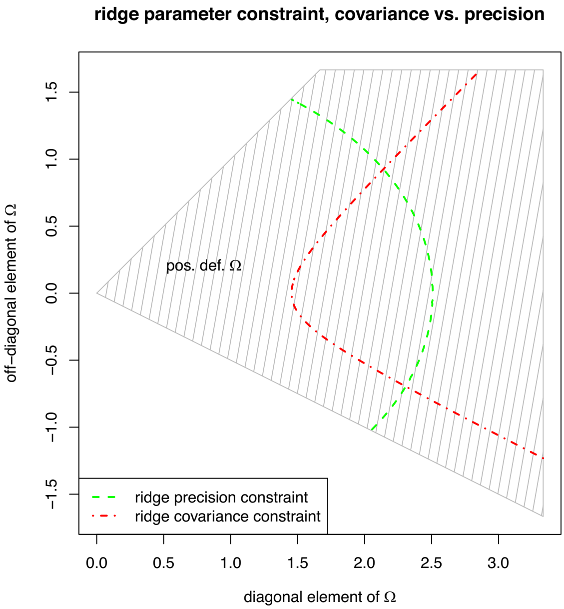

Fig. 1 The gray shaded area demarcates the combinations of ω_d and ω_o that produce a positive definite precision matrix, i.e. the set of feasible estimates. The green dashed and red dashed-dotted lines are the ℓ_2-penalized precision and covariance parameter constraint, respectively. Estimates to the 'right' of these constraints are excluded (color figure online)

To find the ℓ_2-penalized covariance matrix estimator it is thus required to solve a cubic matrix equation. In contrast, the ℓ_2-penalized precision matrix estimator solves a quadratic matrix estimating equation, while the other two ridge-type estimators solve linear ones (see Supplementary Material). As will be obvious from the remainder the computational complexity of solving a cubic matrix equation is substantially more difficult than a linear or quadratic one. In particular, an explicit solution of the former appears to exist only for specific choices of target matrix, while for more general cases one needs to resort to numerical procedures. Both are outlined in either the next section or the Supplementary Material.

3 The estimator

Here the ℓ_2-penalized covariance matrix estimator is derived, while analytic expressions of the other ridge-type estimators are given in the introduction. First for a diagonal, equivariant target matrix, i.e. $\mathbf{T}_\sigma = \alpha_\sigma \mathbf{I}_{pp}$, and at towards the end of the section the numerical evaluation of the estimator with general $\mathbf{T}_\sigma$ is sketched. Besides practical relevance— often no information on a structured $\mathbf{T}_\sigma$ is available—this former diagonal, invariant case is illustrative of the differences between the ridge-type estimators. Substitution of this diagonal, equivariant target matrix in the reformulated estimating equation (4) yields:

$$\mathbf{0}_{pp} = -\boldsymbol{\Sigma} + \mathbf{S} - \lambda_\sigma(\boldsymbol{\Sigma}^3 - \alpha_\sigma \boldsymbol{\Sigma}^2).$$

The root of this equation must share its eigenspace with $\mathbf{S}$. To see this, let $\mathbf{V}_s \mathbf{D}_s \mathbf{V}_s^{\top}$ be the eigendecomposition of $\mathbf{S}$, where $\mathbf{V}_s$ comprises the eigenvectors as columns diagonal matrix and $\mathbf{D}_s$ has the corresponding eigenvalues on its diagonal. Then, pre- and post-multiply the estimating equation of the preceding display by $\mathbf{V}_s^{\top}$ and $\mathbf{V}_s$, respectively. This yields

$$\mathbf{V}_s^{\top} \boldsymbol{\Sigma} \mathbf{V}_s + \lambda_{\sigma} (\mathbf{V}_s^{\top} \boldsymbol{\Sigma} \mathbf{V}_s \mathbf{V}_s^{\top} \boldsymbol{\Sigma} \mathbf{V}_s \mathbf{V}_s^{\top} \boldsymbol{\Sigma} \mathbf{V}_s - \alpha_{\sigma} \mathbf{V}_s^{\top} \boldsymbol{\Sigma} \mathbf{V}_s \mathbf{V}_s^{\top} \boldsymbol{\Sigma} \mathbf{V}_s) = \mathbf{D}_s.$$

The right-hand side of the preceding display is diagonal and, consequently, so must $\mathbf{V}_s^{\top} \boldsymbol{\Sigma} \mathbf{V}_s$ be. The sought estimator is thus of the form $\widehat{\boldsymbol{\Sigma}}(\lambda_{\sigma}) = \mathbf{V}_s \widehat{\mathbf{D}}_{\sigma} \mathbf{V}_s^{\top}$, where diagonal matrix $\widehat{\mathbf{D}}_{\sigma}$ contains the eigenvalues of the estimator. It rests to find the diagonal elements of $\widehat{\mathbf{D}}_{\sigma}$, which should be strictly positive to warrant the positive definiteness of the ℓ_2-penalized covariance matrix estimator. The solution of the matrix equation is then found by solving the eigenvalue estimating equation (after multiplication by -1 to have a positive lead term):

$$0 = \lambda_{\sigma} d_{\sigma,j}^3 - \alpha_{\sigma} \lambda d_{\sigma,j}^2 + d_{\sigma,j} - d_{s,j}, \tag{5}$$

where $d_{\sigma,j} = (\mathbf{D}_{\sigma})_{jj}$. This is a cubic equation in the unknown $d_{\sigma,j}$, which has—by the fundamental theorem of algebra—three roots. The positive, real roots that yield the maximum of the penalized likelihood constitute the ℓ_2-penalized covariance estimator (their derivation can be found in the Supplementary Material). If $d_{s,j} = 0$,

$$\hat{d}_{\sigma,j} = \max\left\{\tfrac{1}{2}\alpha_{\sigma} - \tfrac{1}{2}\lambda_{\sigma}^{-1/2}\sqrt{\alpha_{\sigma}^2 \lambda_{\sigma} - 4},\ \tfrac{1}{2}\alpha_{\sigma} + \tfrac{1}{2}\lambda_{\sigma}^{-1/2}\sqrt{\alpha_{\sigma}^2 \lambda_{\sigma} - 4}\right\}.$$

If $d_{s,j} > 0$, then define the discriminant of the cubic eigenvalue equation (5)

$$\Delta_j = 18\alpha_{\sigma} \lambda_{\sigma}^2 d_{s,j} - 4\alpha_{\sigma}^3 \lambda_{\sigma}^3 d_{s,j} + \alpha_{\sigma}^2 \lambda_{\sigma}^2 - 4\lambda_{\sigma} - 27\lambda_{\sigma}^2 d_{s,j}^2.$$

Moreover, let $q_j = (-2\alpha_{\sigma}^3 \lambda_{\sigma}^3 + 9\lambda_{\sigma}^2 \alpha_{\sigma} - 27\lambda_{\sigma}^2 d_{s,j})/(27\lambda_{\sigma}^3)$ and $r = (3\lambda_{\sigma} - \alpha_{\sigma}^2 \lambda_{\sigma}^2)/(3\lambda_{\sigma}^2)$. Then, for $\Delta_j < 0$:

$$\hat{d}_{\sigma,j} = \begin{cases} -2\sqrt{\frac{r}{3}}\sinh\left[\frac{1}{3}\mathrm{asinh}\left(\frac{3q_j}{2r}\sqrt{\frac{3}{r}}\right)\right] + \frac{1}{3}\alpha_{\sigma} & \text{if } r > 0, \\ -2\,\mathrm{sign}(q_j)\sqrt{\frac{-r}{3}}\cosh\left[\frac{1}{3}\mathrm{acosh}\left(\frac{3|q_j|}{-2r}\sqrt{\frac{3}{-r}}\right)\right] + \frac{1}{3}\alpha_{\sigma} & \text{if } r < 0 \,\&\, 4r^3 + 27q_j^2 > 0, \\ \sqrt{-r} + \frac{1}{3}\alpha_{\sigma} & \text{if } r < 0 \,\&\, q_j = 0, \\ \sqrt[3]{-q_j} + \frac{1}{3}\alpha_{\sigma} & \text{if } r = 0 \,\&\, q_j < 0, \\ \max\left\{\frac{3q_j}{r}, \frac{-3q_j}{2r}\right\} + \frac{1}{3}\alpha_{\sigma} & \text{if } r \neq 0, q_j \neq 0 \,\&\, 4r^3 + 27q_j^2 = 0. \end{cases}$$

and, for $\Delta_j > 0$:

$$\begin{aligned} \hat{d}_{\sigma,j} = \max\Big\{ & 2\sqrt{\tfrac{-r}{3}}\cos\left[\tfrac{1}{3}\,\mathrm{acos}\left(\tfrac{3q_j}{2r}\sqrt{\tfrac{3}{-r}}\right)\right] + \tfrac{1}{3}\alpha_{\sigma}, 2\sqrt{\tfrac{-r}{3}}\cos\left[\tfrac{1}{3}\,\mathrm{acos}\left(\tfrac{3q_j}{2r}\sqrt{\tfrac{3}{-r}}\right) + \tfrac{2}{3}\pi\right] + \tfrac{1}{3}\alpha_{\sigma}, \\ & 2\sqrt{\tfrac{-r}{3}}\cos\left[\tfrac{1}{3}\,\mathrm{acos}\left(\tfrac{3q_j}{2r}\sqrt{\tfrac{3}{-r}}\right) + \tfrac{4}{3}\pi\right] + \tfrac{1}{3}\alpha_{\sigma}\Big\}. \end{aligned}$$

The ℓ_2-penalized covariance matrix estimator is now fully defined.

The above could be generalized to target matrices with an eigenspace coinciding with that of the sample covariance matrix: $\mathbf{T}_\sigma = \mathbf{V}_s \mathbf{D}_{t_\sigma} \mathbf{V}_s^\top$. To accommodate such a choice of the target matrix it only requires to replace α_σ by the $(\mathbf{D}_{t_\sigma})_{jj}$ in the j-th eigenvalue estimating equation (5).

No analytic expression of the ℓ_2-penalized covariance matrix estimator for general $\mathbf{T}_\sigma \in \mathcal{S}_+^p$ appears to exist. In the Supplementary Material we sketch how it can be evaluated numerically.

In the remainder we restrict ourselves to the ℓ_2-penalized covariance matrix estimator with an isotropic, i.e. diagonal, equivariant, target $\mathbf{T}_\sigma = \alpha_\sigma \mathbf{I}_{pp}$. In part this is motivated by mathematical convenience. But this particular case is also believed to be illustrative for the properties of the proposed estimator and representative of the difference among the ridge-type covariance and precision matrix estimators, now all equipped with isotropic targets $\mathbf{T}_\omega = \alpha_\omega \mathbf{I}_{pp}$ and $\mathbf{T}_\ell = \alpha_\ell \mathbf{I}_{pp}$.

4 Analytic properties

Analytic—in contrast to stochastic—properties of the ℓ_2-penalized covariance matrix estimator are studied. The proposition below claims the positive definiteness of the estimator and describes its λ_σ-limiting behavior, i.e. no shrinkage and maximum shrinkage.

Proposition 1 *Let $\alpha_\sigma > 0$ and $\lambda_\sigma \geq 0$. If $\boldsymbol{S}$ has one or more zero eigenvalues, assume additionally that $\alpha_\sigma^2 \lambda_\sigma > 4$. Then:*

(i) $\widehat{\boldsymbol{\Sigma}}(\lambda_\sigma) \succ 0$,
(ii) *If $\boldsymbol{S} \succ 0$, then* $\lim_{\lambda_\sigma \downarrow 0} \widehat{\boldsymbol{\Sigma}}(\lambda_\sigma) = \boldsymbol{S}$,
(iii) $\lim_{\lambda_\sigma \to \infty} \widehat{\boldsymbol{\Sigma}}(\lambda_\sigma) = \alpha_\sigma \boldsymbol{I}_{pp}$.

The proof of Proposition 1, like all proofs, is deferred to the Supplementary Material.

Proposition 1 only contains sanity facts on the ℓ_2-penalized covariance matrix estimator. Of main interest is the joint constraint, apart from their individual positiveness, on λ_σ and α_σ to warrant its positive definiteness when it is derived from a singular sample covariance matrix. In contrast, the other ridge-type estimators only require the penalty, weight and shrinkage parameters to be positive to ensure their positive definiteness.

The regularization limits of the ridge-type covariance matrix estimators $\widehat{\boldsymbol{\Sigma}}(\lambda_\sigma)$, $[\widehat{\boldsymbol{\Omega}}(\lambda_\omega)]^{-1}$, $\widehat{\boldsymbol{\Sigma}}(\lambda_a)$ and $\widehat{\boldsymbol{\Sigma}}(\theta_\ell)$ coincide, in the sense that they converge to $\mathbf{S}$ when $\lambda_\sigma, \lambda_\omega, \lambda_a, \theta_\ell \downarrow 0$. Similarly, the regularization limits of the ridge-type precision matrix estimators $[\widehat{\boldsymbol{\Sigma}}(\lambda_\sigma)]^{-1}$, $\widehat{\boldsymbol{\Omega}}(\lambda_\omega)$, $[\widehat{\boldsymbol{\Sigma}}(\lambda_a)]^{-1}$ and $[\widehat{\boldsymbol{\Sigma}}(\theta_\ell)]^{-1}$ coincide, in the sense that they converge to $\mathbf{S}^+$ when $\lambda_\sigma, \lambda_\omega, \lambda_a, \theta_\ell \downarrow 0$. On the other end of the regularization scale the estimators converge towards (the inverse of) their shrinkage target, except for $\widehat{\boldsymbol{\Sigma}}(\lambda_a)$ and its inverse. In the $\lambda_a \to \infty$-limit $\widehat{\boldsymbol{\Sigma}}(\lambda_a)$ is undefined, while its inverse converges to the nonnegative definite $\mathbf{0}_{pp}$. The regularization limits for $\widehat{\boldsymbol{\Sigma}}(\lambda_a)$ and $\widehat{\boldsymbol{\Sigma}}(\theta_\ell)$ and their inverses are self-evident, while that of the ℓ_2-penalized precision matrix and its inverse are provided by Proposition 1 of Van Wieringen and Peeters (2016). An overview of the shrinkage limits is provided in the Supplementary Material.

Finally, a shared shrinkage limit does, however, not imply a common regularization path (see the Supplementary Material for an illustration).

5 Mean squared error

Shrinkage methods like ridge penalized estimation may, for a suitable choice of the penalty parameter, yield estimators that outperform the maximum likelihood estimator in the mean squared error (MSE) sense (of the vectorized estimator). It reveals the inadmissibility of the maximum likelihood estimator. This has been shown for the ridge regression estimator by Theobald (1974). In preceding work (van Wieringen 2017) we proved that this also holds for the ℓ_2-penalized precision matrix estimator, and for its inverse as a covariance matrix estimator, when using an isotropic target matrix. Here we investigate whether the ℓ_2-penalized covariance matrix estimator also exhibits this property.

Before we state the main result of this section, we need the following auxiliary lemma.

Lemma 2 *Let* $\boldsymbol{T}_\sigma = \alpha_\sigma \boldsymbol{I}_{pp}$ *with* $\alpha_\sigma > 0$. *Then:* $\widehat{\boldsymbol{\Sigma}}(\lambda_\sigma) = \boldsymbol{S} + \alpha_\sigma \lambda_\sigma \boldsymbol{S}^2 - \lambda_\sigma \boldsymbol{S}^3 + \lambda_\sigma^2 \boldsymbol{S}^3 (\boldsymbol{I}_{pp} - \boldsymbol{S})(2\alpha_\sigma \boldsymbol{I}_{pp} - 3\boldsymbol{S}) + \mathcal{O}(\lambda_\sigma^3)$ *for* $0 < \lambda_\sigma \ll 1$.

The proof of the lemma is immediate after substitution of the estimating equation into itself, expansion of the right-hand side and grouping of terms. Note that the $\mathcal{O}$-notation in Lemma 2 means that there exists a $C > 0$ such that the largest eigenvalue of the remainder term is bounded by $C\lambda_\sigma^3$ for λ_σ sufficiently close to zero.

Proposition 3 states that there exists $\lambda_\sigma > 0$ such that the ℓ_2-penalized covariance matrix estimator outperforms its maximum likelihood counterpart in the mean squared error sense, when the parameter $\alpha_\sigma > 0$ of the diagonal, equivariant target $\mathbf{T} = \alpha_\sigma \mathbf{I}_{pp}$ matrix is not too large.

Proposition 3 *Let* $\boldsymbol{T}_\sigma = \alpha_\sigma \boldsymbol{I}_{pp}$ *with* $p < n$ *and* $0 \leq \alpha_\sigma < a_2/a_1$ *where*

$$\begin{aligned} a_1 &= n^{-2}\{[\mathrm{tr}(\boldsymbol{\Sigma})]^3 + (2n+3)\mathrm{tr}(\boldsymbol{\Sigma}^2)\mathrm{tr}(\boldsymbol{\Sigma}) + (2n+4)\mathrm{tr}(\boldsymbol{\Sigma}^3)\} \\ a_2 &= n^{-3}\{(2n^2+17n+20)\mathrm{tr}(\boldsymbol{\Sigma}^3)\mathrm{tr}(\boldsymbol{\Sigma}) + (5n+6)[\mathrm{tr}(\boldsymbol{\Sigma})]^2\mathrm{tr}(\boldsymbol{\Sigma}^2) \\ &\qquad + (n^2+4n+5)[\mathrm{tr}(\boldsymbol{\Sigma}^2)]^2 + (3n^2+17n+20)\mathrm{tr}(\boldsymbol{\Sigma}^4)\}. \end{aligned}$$

Then, there exists a $\lambda_\sigma > 0$ *such that* $\mathrm{MSE}\{vec[\widehat{\boldsymbol{\Sigma}}(\lambda_\sigma)]\} < \mathrm{MSE}\{vec[\widehat{\boldsymbol{\Sigma}}(0)]\} = \mathrm{MSE}[vec(\boldsymbol{S})]$.

Analogous results to Proposition 3 can be formulated for the ridge-type covariance matrix estimators. Hereto we provide an overview of the first order expansion of the estimators' MSE in terms of their regularization parameter around zero (which for the ℓ_2-penalized precision estimator and its inverse can be found in van Wieringen (2017)). The overview reveals that all ridge-type covariance matrix estimators but $\widehat{\boldsymbol{\Sigma}}(\lambda_a)$ outperform their maximum likelihood counterpart. However, for $\widehat{\boldsymbol{\Sigma}}(\lambda_\sigma)$ and $\widehat{\boldsymbol{\Sigma}}(\theta_\ell)$ this superior performance depends on the value of the target parameter, whereas $[\widehat{\boldsymbol{\Omega}}(\lambda_\omega)]^{-1}$ outperforms the maximum likelihood estimator for all $a_\omega > 0$.

Similarly, all ridge-type precision matrix estimators outperform their maximum likelihood counterpart, but now $\widehat{\mathbf{\Omega}}(\lambda_\omega)$ only for certain values of its target parameter. The ridge-type estimators thus render the maximum likelihood estimator inadmissable, but they do so among themselves as the behavior of their MSE depends on the model parameter.

It seems counter-intuitive that even for a very poor target parameter, e.g. the inverse of ℓ_2-penalized precision matrix estimator has a better MSE than the maximum likelihood estimator. We therefore illustrate the effect of the target parameter on the mean squared error of the ridge-type covariance and precision matrix estimators in a simple example. Hereto data $\mathbf{Y}_i$ for $i = 1, \ldots, 30$ are drawn from the $\mathcal{N}(\mathbf{0}_5, \mathbf{\Omega}^{-1})$-law with the precision matrix $\mathbf{\Omega}$ as in Sect. 4. From these data the sample covariance matrix and, subsequently, the ridge-type estimators are evaluated over a grid of penalty parameters for various target parameters chosen such that they share the same shrinkage limits. For each estimator the squared Frobenius norm of its difference to $\mathbf{\Sigma} = \mathbf{\Omega}^{-1}$ is calculated. The above is repeated 10, 000 times and the results are averaged. The results are plotted (cf. Supplementary Material). These plots show the mean squared errors versus the regularization parameter over the $[0, \delta]$ domain with $0 < \delta \ll 1$. It confirms that, irrespective of the choice of the target parameter, the mean squared error of, e.g., the inverse of the ℓ_2-penalized precision matrix estimator indeed outperforms that of the maximum likelihood estimator for small enough values of its regularization parameter. But the choice of the target parameter does of course affect the domain of the regularization parameter where the ridge-type estimator is superior to the maximum likelihood one. The plots also shows that for excessive choices of the target parameter, e.g., the ℓ_2-penalized covariance matrix estimator does not outperform the maximum likelihood counterpart, thereby confirming the need for the constraint formulated in Proposition 3.

6 Consistency

The consistency of the ℓ_2-penalized covariance matrix estimator is shown, consistency in the traditional sense with the sample size n tends to infinity while the dimension p is kept fixed. This is deemed appropriate for practical purposes where the dimension of the system under study is fixed from the onset of the analysis.

Proposition 4 *Equip both λ_σ and $\widehat{\mathbf{\Sigma}}(\lambda_\sigma)$ with the novel subscript n and write $\lambda_{\sigma,n}$ and $\widehat{\mathbf{\Sigma}}_n(\lambda_{\sigma,n})$ to stress their dependence on the sample size. Let $\lambda_{\sigma,n}$ converge in probability to zero as n tends to infinity. Then: $p\text{-lim}_{n\to\infty} \widehat{\mathbf{\Sigma}}_n(\lambda_{\sigma,n}) = \mathbf{\Sigma}$.*

Proposition 4 assumes $p\text{-lim}_{n\to\infty} \lambda_{\sigma,n} = 0$. There is no theoretical justification for this assumption. It has been shown in silico to be unproblematic for the ℓ_2-penalized precision matrix estimator and its generalization (Van Wieringen and Peeters 2016; van Wieringen 2019). Intuitively, this is not surprising as for larger n less regularization is required to produce a well-defined estimator.

All ridge-type estimators are, under analogous assumptions on the regularization parameter, consistent estimators. The consistency of $\widehat{\mathbf{\Sigma}}(\lambda_a)$ and $\widehat{\mathbf{\Sigma}}(\theta_\ell)$ for $\mathbf{\Sigma}$ is evident by that of the sample covariance estimator in combination with continuous mapping theorem (van der Vaart 2000), while the consistency of $\widehat{\mathbf{\Omega}}(\lambda_\omega)$ for $\mathbf{\Omega}$ has been shown in Van Wieringen and Peeters (2016). By the continuous mapping theorem, their inverses are also consistent for the inverse of the parameter.

We compare the ℓ_2-penalized covariance matrix estimator and the inverse of the ℓ_2-penalized precision matrix estimator with respect to the amount of regularization that yields the same convergence rate. Hereto consider—in line with the assumption on the penalty parameter of Proposition 4—regularization schemes of the form $\lambda_\sigma = n^{-c_\sigma}$ and $\lambda_\omega = n^{-c_\omega}$ for rates $c_\sigma, c_\omega > 0$. For large n the MSEs of both estimators then are (cf. the proof of Proposition 3 and van Wieringen (2017)):

$$\begin{aligned} \text{MSE}\{\text{vec}[\widehat{\mathbf{\Sigma}}(\lambda_\sigma)]\} &= \text{MSE}[\text{vec}(\mathbf{S})] - 2\lambda_\sigma a_2 + \mathcal{O}(\lambda_\sigma^2), \\ \text{MSE}\{\text{vec}[\widehat{\mathbf{\Omega}}^{-1}(\lambda_\omega)]\} &= \text{MSE}[\text{vec}(\mathbf{S})] - 2\lambda_\omega p(p-1)(n-p+1)^{-1} + \mathcal{O}(\lambda_\omega^2), \end{aligned}$$

where for simplicity both targets α_σ and α_ω have been set to zero. Numerically, for a variety of choices of $\mathbf{\Sigma}$ we have observed that $a_2 > p(p-1)(n-p+1)^{-1}$. This indicates that the ℓ_2-penalized covariance matrix estimator tolerates more regularization to achieve the same MSE asymptotically. Put differently, it can handle more shrinkage, while not comprising the rate of MSE consistency, in comparison to the inverse ℓ_2-penalized precision matrix estimator. A reversed conclusion can be drawn for the inverse ℓ_2-penalized covariance and ℓ_2-penalized precision matrix estimators as estimators of the precision matrix. For large n their MSEs are:

$$\begin{aligned} \text{MSE}\{\text{vec}[\widehat{\mathbf{\Sigma}}^{-1}(\lambda_\sigma)]\} &= \text{MSE}[\text{vec}(\mathbf{S}^{-1})] - 2\lambda_\sigma(p+1)(n-p-1)^{-1} + \mathcal{O}(\lambda_\sigma^2), \\ \text{MSE}\{\text{vec}[\widehat{\mathbf{\Omega}}(\lambda_\omega)]\} &= \text{MSE}[\text{vec}(\mathbf{S}^{-1})] - 2\lambda_\omega \,\text{tr}[\mathbb{E}(\mathbf{S}^{-4} - \mathbf{S}^{-3}\mathbf{\Sigma}^{-1})] + \mathcal{O}(\lambda_\omega^2), \end{aligned}$$

where the former is found by similar means as employed in the proof of Proposition 3 and the latter has been derived in van Wieringen (2017). Numerically, we now observe $(p+1)(n-p-1)^{-1} < \text{tr}[\mathbb{E}(\mathbf{S}^{-4} - \mathbf{S}^{-3}\mathbf{\Sigma}^{-1})]$ for many choices of $\mathbf{\Omega}$. For large n and comparable regularization schemes the MSE of $\widehat{\mathbf{\Omega}}^{-1}(\lambda_\omega)$ is closer to that of $\widehat{\mathbf{\Sigma}}(\lambda_\sigma)$ than that of $\widehat{\mathbf{\Sigma}}^{-1}(\lambda_\sigma)$ is to that of $\widehat{\mathbf{\Omega}}(\lambda_\omega)$. Hence, the ℓ_2-penalized precision matrix estimator is—asymptotically—to be preferred if it is to be used for both covariance and precision matrix estimation.

7 Bayesian interpretation

All ridge-type estimators have a Bayesian interpretation. The ridge covariance matrix estimators $\widehat{\mathbf{\Sigma}}(\lambda_a)$ and $\widehat{\mathbf{\Sigma}}(\theta_\ell)$ arise from a conjugate inverse Wishart prior. Assume $\mathbf{\Sigma}$ to follow an inverse Wishart distribution $\mathcal{IW}(\mathbf{\Psi}, m)$ with scale parameter $\mathbf{\Psi} \succ 0$ and degrees of freedom $m \in \mathbb{N}_+$ such that $m > p - 1$. Then,

$$\mathbb{E}(\mathbf{\Sigma} \,|\, \mathbf{Y}_1, \ldots, \mathbf{Y}_n) = (m + n - p - 1)^{-1}(\mathbf{\Psi} + n\mathbf{S}).$$

Both covariance matrix estimators have this form. Similarly, the precision matrix estimators $[\widehat{\mathbf{\Sigma}}(\lambda_a)]^{-1}$ and $[\widehat{\mathbf{\Sigma}}(\theta_\ell)]^{-1}$ arise from a conjugate Wishart prior. Let $\mathbf{\Sigma} \sim$

$\mathcal{W}(\boldsymbol{\Upsilon}, \nu)$ with scale parameter $\boldsymbol{\Upsilon} \succ 0$ and degrees of freedom $\nu \in \mathbb{N}_+$ such that $n > p - 1$. Then,

$$\mathbb{E}(\boldsymbol{\Sigma}^{-1} \,|\, \mathbf{Y}_1, \ldots, \mathbf{Y}_n) = (\nu + n)(\boldsymbol{\Upsilon}^{-1} + n\mathbf{S})^{-1}.$$

The right-hand side is the same form as the precision matrix estimators.

The ℓ_2-penalized covariance matrix estimator corresponds to the Maximum A Posteriori (MAP) estimator of $\boldsymbol{\Sigma}$. This MAP estimator employs a prior of independent Gaussians on the entries of $\boldsymbol{\Sigma}$. Each of these Gaussians has a mean equal to the corresponding entry of the target matrix $\mathbf{T}_\sigma$, but they have common variance equal to λ_σ^{-1}. Hereto note that

$$-\tfrac{1}{2}\lambda_\sigma \mathrm{tr}[(\boldsymbol{\Sigma} - \mathbf{T}_\sigma)^\top(\boldsymbol{\Sigma} - \mathbf{T}_\sigma)] = -\tfrac{1}{2}\lambda_\sigma \sum\nolimits_{j,j'=1}^{p} [(\boldsymbol{\Sigma})_{j,j'} - (\mathbf{T}_\sigma)_{j,j'}]^2, \quad (6)$$

and that each summand $[(\boldsymbol{\Sigma})_{j,j'} - (\mathbf{T}_\sigma)_{j,j'}]^2$ is proportional to the log-density of the univariate Gaussian distribution with mean $(\mathbf{T}_\sigma)_{j,j'}$ and variance λ_σ^{-1}. Consequently,

$$\widehat{\boldsymbol{\Sigma}}(\lambda_\sigma) = \arg\max_{\boldsymbol{\Sigma}\in\mathcal{S}^p_{++}} \log[p(\boldsymbol{\Sigma} \,|\, \mathbf{Y})] \;=\; \arg\max_{\boldsymbol{\Sigma}\in\mathcal{S}^p_{++}} \log[p(\mathbf{Y} \,|\, \boldsymbol{\Sigma})] + \log[p(\boldsymbol{\Sigma})],$$

with prior $p(\boldsymbol{\Sigma}) \overset{d}{=} \prod_{j,j'=1}^{p} \mathcal{N}[(\boldsymbol{\Sigma})_{j,j'}; (\mathbf{T}_\sigma)_{j,j'}, \lambda_\sigma^{-1}] \times \mathbb{1}_{\{\boldsymbol{\Sigma}\in\mathcal{S}^p_{++}\}}$. In this prior, $\mathcal{N}(x; \mu, \tau^2)$ denotes that the random variable x is normally distributed with mean μ and variance τ^2, and $\mathcal{S}^p_{++}$ is the space of all real symmetric positive definite matrices.

The ℓ_2-penalized precision matrix estimator too corresponds to a MAP estimator of $\boldsymbol{\Omega}$, again under independent Gaussian priors on the entries of $\boldsymbol{\Omega}$ with means equal to the entries of the target matrix $\mathbf{T}_\omega$ and variances equal to λ_ω^{-1}:

$$\widehat{\boldsymbol{\Omega}}(\lambda_\omega) = \arg\max_{\boldsymbol{\Omega}\in\mathcal{S}^p_{++}} \log[p(\boldsymbol{\Omega} \,|\, \mathbf{Y})] \;=\; \arg\max_{\boldsymbol{\Omega}\in\mathcal{S}^p_{++}} \log[p(\mathbf{Y} \,|\, \boldsymbol{\Omega})] + \log[p(\boldsymbol{\Omega})],$$

with $p(\boldsymbol{\Omega}) \overset{d}{=} \prod_{j,j'=1}^{p} \mathcal{N}[(\boldsymbol{\Omega})_{j,j'}; (\mathbf{T}_\omega)_{j,j'}, \lambda_\omega^{-1}] \times \mathbb{1}_{\{\boldsymbol{\Omega}\in\mathcal{S}^p_{++}\}}$.

The Gaussian priors used in the above Bayesian formulations of the ℓ_2-penalized covariance and precision estimators yield penalties that are proportional to the sum of square differences between the parameters and their target. Hence, the ridge prior encourages entries of $\widehat{\boldsymbol{\Sigma}}(\lambda_\sigma)$ (respectively $\widehat{\boldsymbol{\Omega}}(\lambda_\omega)$) to be close to the entries of $\mathbf{T}_\sigma$ (respectively $\mathbf{T}_\omega$) by an amount that is controlled by λ_σ (respectively λ_ω). Remark that if $\mathbf{T}_\sigma = \mathbf{0}_{pp}$ (respectively $\mathbf{T}_\sigma = \mathbf{0}_{pp}$) the ridge prior is centered at zero. Contrast this to $\mathbf{T}_\sigma = \alpha_\sigma \mathbf{I}_{pp}$ (respectively $\mathbf{T}_\sigma = \alpha_\omega \mathbf{I}_{pp}$), where the ridge prior of the diagonal elements of the covariance (respectively precision) matrix is centered at α_σ (respectively α_ω), which is a natural choice as those elements should be positive.

Due to the complex inverse transformation between $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ the priors of the ℓ_2-penalized covariance and precision matrices are not equivalent. This is true even in the simple case where $p = 1$. The left panel of Fig. 2 shows that the inverse transformation affects both the location and shape of the prior. This effect is expected to be accentuated when p is larger. While it is difficult to visualize higher-dimensional priors, the right

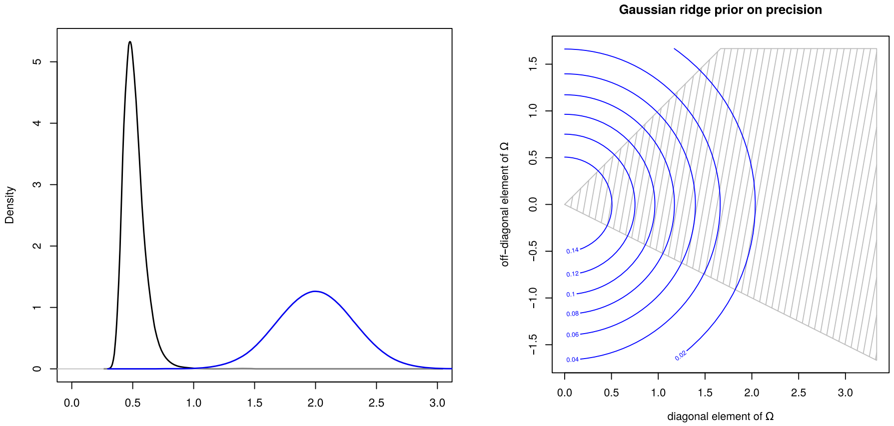

Fig. 2 Illustration of the ridge prior on covariance and precision elements. Left panel: $p = 1$, σ^2 has a truncated normal prior ($t_\sigma = 2$ and $\lambda_\sigma = 10$) and the implied prior on its inverse σ^{-2} is displayed (black line). Right panel: The toy example of Sect. 2 is considered where $\mathbf{\Omega} = (\omega_d - \omega_0)\mathbf{I}_{33} + \omega_0\mathbf{1}_{33}$ is parameterized in terms of two parameters only. The contour plot of the ridge prior where $\mathbf{T}_\omega = \mathbf{0}_{33}$ and $\lambda_\omega = 1$ is displayed along with the positive definite domain of $\mathbf{\Omega}$

panel of Fig. 2 illustrates the prior for the toy example of Sect. 2 the contour plot of the ridge prior when $\mathbf{T}_\omega = \mathbf{0}_{33}$ and $\lambda_\omega = 1$ along with the positive definite domain of $\mathbf{\Omega}$.

Miok et al. (2017) reported that the ℓ_2-penalized precision matrix estimator corresponds to the posterior expectation of $\mathbf{\Sigma}$ (instead of posterior mode) under a Wishart prior with scale matrix $(4n^2\lambda_\omega\mathbf{I}_{pp} + n^2\mathbf{S}^2)^{-1/2}$ and degree of freedom n. However, we prefer the formulation as a Bayesian MAP estimator presented above to the posterior mean one as the former formulation does not require that $n > p - 1$. Moreover, the normal product prior of the MAP estimator emphasizes the symmetry of the ℓ_2-penalty, both on the diagonal and off-diagonal elements of $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ as well as in the involvement of the (prior) matrices $\mathbf{T}_\sigma$ and $\mathbf{T}_\omega$. It is also more in line with the Bayesian formulations of other penalized estimators such as that of the graphical lasso and horseshoe estimators (Wang 2012; Li et al. 2019).

In summary, $\widehat{\mathbf{\Sigma}}(\lambda_a)$ and $\widehat{\mathbf{\Sigma}}(\theta_\ell)$ (and their inverses) are posterior mean estimators that optimize the quadratic loss. They are equipped with conjugate priors and their posterior is a well-known distribution. On the other hand, the ℓ_2-penalized covariance and precision estimators are MAP estimators that optimize a 0/1 loss function. A sampler can be constructed to draw from the posterior distribution.

8 Correlation screening

Through simulation we compare the ridge-type covariance and precision estimators in terms of loss and for screening of either marginal and conditional and partial correlations. The data are sampled from a zero-mean p-variate normal law with a covariance matrix that is either specified directly $\mathbf{\Sigma}$ or indirectly via its inverse, i.e.

the precision matrix $\mathbf{\Omega}$. Five parametrizations of these matrices are adopted: banded (twice), blocked (twice), and striped (details in the Supplementary Material). In these parametrizations the $\mathbf{\Sigma}$ is replaced by $\mathbf{\Omega}$ when considering conditional correlation screening. Furthermore, we consider all pairwise combinations of the sample size $n \in \{10, 50, 100\}$ and dimension $p \in \{50, 100, 250\}$. All combinations represent high-dimensionality as the number of parameters equals $\frac{1}{2}p(p+1)$. In accordance with these specifications $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are drawn from either $\mathcal{N}(\mathbf{0}_p, \mathbf{\Sigma})$ or $\mathcal{N}(\mathbf{0}_p, \mathbf{\Omega}^{-1})$ when investigating properties of the ridge-type covariance and precision matrix estimators, respectively. In either screening case the ridge-type covariance and precision matrices are estimated from the data with their penalty parameters chosen through leave-one-out cross-validation. Subsequently, the performance of the estimators is measured by *i)* the AUC and pAUC, the (partial) Area Under the Curve, obtained from the (partial) correlation matrix estimates with respect to the support of the true (partial) correlation matrix, and more indirectly by *ii)* the Frobenius and quadratic loss, which for e.g., the ℓ_2-penalized covariance matrix estimator are $\|\widehat{\mathbf{\Sigma}}(\lambda_\sigma^{(\text{cv})}) - \mathbf{\Sigma}\|_F^2$, $\|\widehat{\mathbf{\Sigma}}(\lambda_\sigma^{(\text{cv})})\mathbf{\Sigma}^{-1} - \mathbf{I}_{pp}\|_F^2$, $\|\widehat{\mathbf{\Sigma}}^{-1}(\lambda_\sigma^{(\text{cv})}) - \mathbf{\Omega}\|_F^2$, and $\|\widehat{\mathbf{\Sigma}}^{-1}(\lambda_\sigma^{(\text{cv})})\mathbf{\Omega}^{-1} - \mathbf{I}_{pp}\|_F^2$, where the superscript of the regularization parameter explicates that it has been chosen through cross-validation. The above is repeated a hundred times.

The results of the marginal and conditional screening as well as the Frobenius and quadratic loss of the ridge-type covariance and precision matrix estimators are displayed as boxplots (see the Supplementary Material). The results of the (partial) AUC clearly show that marginal correlations screening is best done with the inverse of the ℓ_2-penalized precision matrix, with a small exception almost irrespective of the type of covariance matrix. The estimators $\widehat{\mathbf{\Sigma}}(\lambda_a^{(\text{cv})})$ and $\widehat{\mathbf{\Sigma}}(\theta_\ell^{(\text{cv})})$ are the runners-up for marginal correlation screening, while $\widehat{\mathbf{\Sigma}}(\lambda_\sigma^{(\text{cv})})$ clearly trails behind. For conditional correlation screening, the estimator $[\widehat{\mathbf{\Sigma}}(\lambda_a^{(\text{cv})})]^{-1}$ generally performs best in terms of AUC and pAUC but is closely followed—if not a par—by $[\widehat{\mathbf{\Sigma}}(\theta_\ell^{(\text{cv})})]^{-1}$ and $[\widehat{\mathbf{\Sigma}}(\lambda_\sigma^{(\text{cv})})]^{-1}$. The estimator $\widehat{\mathbf{\Omega}}(\lambda_\omega^{(\text{cv})})$ yields a smaller AUC and pAUC than these estimators but is generally not too far behind.

The loss results of the estimators reveal that $\widehat{\mathbf{\Sigma}}(\lambda_a^{(\text{cv})})$ and $[\widehat{\mathbf{\Omega}}(\lambda_\omega^{(\text{cv})})]^{-1}$ are the best covariance matrix estimators, both in the Frobenius and quadratic loss sense. These estimators are followed at some distance by $\widehat{\mathbf{\Sigma}}(\lambda_\sigma^{(\text{cv})})$, while $\widehat{\mathbf{\Sigma}}(\theta_\ell^{(\text{cv})})$ performs worst, especially high-dimensionally. The picture is more or less similar for the estimation of the precision matrix, with $\widehat{\mathbf{\Omega}}(\lambda_\omega^{(\text{cv})})$ now slightly preferably in both losses over $[\widehat{\mathbf{\Sigma}}(\lambda_a^{(\text{cv})})]^{-1}$. But $[\widehat{\mathbf{\Sigma}}(\lambda_\sigma^{(\text{cv})})]^{-1}$ has comparable but worse loss, while $[\widehat{\mathbf{\Sigma}}(\theta_\ell^{(\text{cv})})]^{-1}$ trails behind dramatically in the high-dimensional setting.

To assess how the performance is affected if we relax the normality assumption, we have repeated the simulation now drawing the data from the multivariate t-distribution with 10 degrees of freedom. The above parametrizations of the covariance and precision matrix are then that of the multivariate's t-distribution's scale parameter or its inverse, respectively. Conclusions are more or less similar to the normal case: 'marginal correlation' screening is best done on the basis of either scale parameter estimator $\widehat{\mathbf{\Sigma}}(\lambda_a^{(\text{cv})})$ or $[\widehat{\mathbf{\Omega}}(\lambda_\omega^{(\text{cv})})]^{-1}$, with the latter performing best with higher dimensionality. These two estimators also yield the lowest loss among the scale parameter estimators, with preference depending on the topology. Again notable is the poor loss of $\widehat{\mathbf{\Sigma}}(\theta_{\ell}^{(\mathrm{cv})})$. Turning to the inverse scale parameter estimates, for 'conditional correlation' screening $[\widehat{\mathbf{\Sigma}}(\lambda_a^{(\mathrm{cv})})]^{-1}$ is slightly preferable but $[\widehat{\mathbf{\Sigma}}(\theta_{\ell}^{(\mathrm{cv})})]^{-1}$ and $[\widehat{\mathbf{\Sigma}}(\lambda_{\sigma}^{(\mathrm{cv})})]^{-1}$ are good competitors while $\widehat{\mathbf{\Omega}}(\lambda_{\omega}^{(\mathrm{cv})})$ follows suit. With respect to the loss of the inverse scale estimators, their performance is—apart from $[\widehat{\mathbf{\Sigma}}(\theta_{\ell}^{(\mathrm{cv})})]^{-1}$—more or less on a par.

9 Conclusion

A first contribution of this work can be found in the presentation of the ℓ_2-penalized estimator of the covariance matrix, which was hitherto unavailable in the literature. An analytic solution of the estimator's estimating equation has been found for a specific but practically relevant case, while a numerical procedure for the general case has been outlined. Moving forward with the specific case, analytic and stochastic properties, like its positive definiteness, mean squared error, consistency, the Bayesian perspective and correlation screening, of the ℓ_2-penalized covariance matrix estimator have been studied.

The other contribution lies in the comparison of the various ridge-type covariance and precision matrix estimators to shed light on which estimator best to use for the estimation of the covariance or the precision matrix. While in the two-dimensional case the estimators are equivalent, in higher dimensions they exhibit clear differences. Their differences originate in their estimating equations, a nonconvex cubic vs. a convex quadratic vs. linear matrix equation for the ℓ_2-penalized covariance and precision matrix estimators and shrinkage covariance matrix estimators, respectively. Analytic expressions of the solutions to the estimating equations of $\widehat{\mathbf{\Sigma}}(\lambda_a)$, $\widehat{\mathbf{\Sigma}}(\theta_{\ell})$, and $\widehat{\mathbf{\Omega}}(\lambda_{\omega})$ exist. For $\widehat{\mathbf{\Sigma}}(\lambda_{\sigma})$ it exists only for isotropic targets. Moreover, the latter comes with an additional constraint on the penalty parameter and target if the dimension p exceeds the sample size n. All estimators have been shown to outperform—for suitably chosen regularization parameter values—their maximum likelihood counterpart in the MSE sense. Similarly, under assumptions, the estimators are all consistent for both $\mathbf{\Sigma}$ and $\mathbf{\Omega}$. From a Bayesian viewpoint, we have shown that the ℓ_2-penalized estimators of the covariance and precision matrices are Maximum A Posteriori estimators when employing independent Gaussian priors on the entries of the matrices and that the priors are not equivalent, even when $p = 1$. The other ridge-type estimators are posterior mean estimator that stem from a conjugate Wishart-type prior. Finally, the estimators (and their inverses) have been compared in simulation with respect to their capacity of marginal and partial correlation screening and their Frobenius and quadratic loss. This revealed that, while the preferred estimator depends on the objective and context, overall the (inverse of the) ℓ_2-penalized precision matrix estimator has a (small) edge over its counterparts. Only for conditional correlation screening in high-dimensional settings the other ridge-type estimators exhibit slightly favorable behavior.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00362-024-01610-9.

References

Bien J, Tibshirani RJ (2011) Sparse estimation of a covariance matrix. Biometrika 98(4):807–820

Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9(3):432–441

Hero A, Rajaratnam B (2011) Large-scale correlation screening. J Am Stat Assoc 106(496):1540–1552

James W, Stein C (1961) Estimation with quadratic loss. In: Proceedings of the Berkeley symposium on mathematical statistics and probability, p 361

Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. J Multivar Anal 88:365–411

Li Y, Craig BA, Bhadra A (2019) The graphical horseshoe estimator for inverse covariance matrices. J Comput Graph Stat 28(3):747–757

Luo S, Song R, Witten D (2014) Sure screening for gaussian graphical models. arXiv preprint arXiv:1407.7819

Miok V, Wilting SM, Wieringen WN (2017) Ridge estimation of the var (1) model and its time series chain graph from multivariate time-course omics data. Biom J 59(1):172–191

Theobald CM (1974) Generalizations of mean square error applied to ridge regression. J R Stat Soc Ser B 36(1):103–106

van der Vaart AW (2000) Asymptotic statistics, vol 3. Cambridge University Press, Cambridge

Van Wieringen WN, Peeters CFW (2016) Ridge estimation of the inverse covariance matrix from high-dimensional data. Comput Stat Data Anal 103:284–303

Wang H (2012) Bayesian graphical lasso models and efficient posterior computation. Bayes Anal 7(4):867–886

Warton DI (2008) Penalized normal likelihood and ridge regularization of correlation and covariance matrices. J Am Stat Assoc 103(481):340–349

Wieringen WN (2017) On the mean squared error of the ridge estimator of the covariance and precision matrix. Stat Probab Lett 123:88–92

Wieringen WN (2019) The generalized ridge estimator of the inverse covariance matrix. J Comput Graph Stat 28(4):932–942