

Partial Coherence Estimation via Spectral Matrix Shrinkage under Quadratic Loss

D. Schneider-Luftman, *Student Member, IEEE* and A. T. Walden, *Senior Member, IEEE*

Abstract—Partial coherence is an important quantity derived from spectral or precision matrices and is used in seismology, meteorology, oceanography, neuroscience and elsewhere. If the number of complex degrees of freedom only slightly exceeds the dimension of the multivariate stationary time series, spectral matrices are poorly conditioned and shrinkage techniques suggest themselves. When true partial coherencies are quite large then for shrinkage estimators of the diagonal weighting kind it is shown empirically that the minimization of risk using quadratic loss (QL) leads to oracle partial coherence estimators superior to those derived by minimizing risk using Hilbert-Schmidt (HS) loss. When true partial coherencies are small the methods behave similarly. We derive two new QL estimators for spectral matrices, and new QL and HS estimators for precision matrices. In addition for the full estimation (non-oracle) case where certain trace expressions must also be estimated, we examine the behaviour of three different QL estimators, the precision matrix one seeming particularly robust and reliable. For the empirical study we carry out exact simulations derived from real EEG data for two individuals, one having large, and the other small, partial coherencies. This ensures our study covers cases of real-world relevance.

Index Terms—partial coherence, quadratic loss, shrinkage, precision matrix, spectral matrix.

I. INTRODUCTION

Consider a p -vector-valued (or multivariate) stationary time series $\{\mathbf{X}_t\}$ where $\mathbf{X}_t = [X_{1,t}, \dots, X_{p,t}]^T \in \mathbb{R}^p$, $t \in \mathbb{Z}$, and T denotes transposition. Without loss of generality we assume $\{\mathbf{X}_t\}$ to have a zero mean. Denote the sample interval by Δ_t . One very important quantity derived from vector-valued time series is the partial coherence between different series. Let $\mathcal{S}(f)$ denote the spectral matrix of $\{\mathbf{X}_t\}$ at frequency f , assumed to exist and be of full rank. With $\mathbf{s}_\tau \stackrel{\text{def}}{=} \text{cov}\{\mathbf{X}_{t+\tau}, \mathbf{X}_t\} = E\{\mathbf{X}_{t+\tau}\mathbf{X}_t^T\}$, we have $\mathcal{S}(f) \stackrel{\text{def}}{=} \Delta_t \sum_{\tau=-\infty}^{\infty} \mathbf{s}_\tau e^{-i2\pi f\tau \Delta_t}$. Denote the $(j, k)^{\text{th}}$ element of the precision matrix $\mathcal{C}(f) \stackrel{\text{def}}{=} \mathcal{S}^{-1}(f)$ by $C_{jk}(f)$. The partial coherence between series j and k can be expressed as, (e.g., [6]), $\gamma_{jk \cdot \{j, k\}}^2(f) = |C_{jk}(f)|^2 / [C_{jj}(f)C_{kk}(f)]$, and is the frequency domain squared correlation coefficient between series j and k after the removal of the linear effects of the remaining series, the remaining series being denoted by $\{j, k\}$. This characteristic has led to partial coherence being used widely in the physical sciences, e.g., in seismology [34], meteorology [11], oceanography [18] and extensively in

neuroscience [10], [19], [25], [28], [23]. Clearly to calculate the partial coherencies we can estimate $\mathcal{S}(f)$ as $\hat{\mathcal{S}}(f)$, and invert it, or we can estimate $\mathcal{C}(f)$ directly.

Consider an estimator $\hat{\mathcal{S}}(f)$, of $\mathcal{S}(f)$. An estimator may be computed by a multitaper scheme involving K tapers (e.g., [26]). (Throughout this paper we assume $K \geq p$ so that the spectral matrix is non-singular with probability one.) The spectral matrices will be non-singular but poorly conditioned if K is only a little larger than p . The derived partial coherencies will reflect this ill-conditioning. This study was motivated by exactly this problem in a neuroscience setting. Due to required low-pass filtering only a small frequency range was available for analysis and K was necessarily kept small — see Section III. K cannot be simply increased because of its connection to the implied smoothing bandwidth: if K is made larger, the required resolution may be lost. (Other estimators such as periodograms smoothed over frequencies have analogous properties.)

Can the mean-square errors of the resulting estimated partial coherencies be reduced? Such a reduction would be very useful: for example, the estimated partial mutual information obtained from the estimated partial coherencies are used in determining brain functional connectivity [28] so that increased precision is scientifically well worthwhile. An obvious approach is to use covariance shrinkage methodology, a subject with a large literature, see e.g., [8], [14], [30], [29], [9], [5].

Ledoit and Wolf (LW) [20] derived ideal shrinkage estimators which are a combination of the standard covariance matrix and a target matrix proportional to the identity; such diagonal up-weighting has a long history [1] and we concentrate on this estimator class in this paper. LW minimize a risk measure — defined in terms of Hilbert-Schmidt (HS) or Frobenius loss — between the true and estimated covariance matrices. The ideal parameter value for shrinking $\hat{\mathcal{S}}(f)$ is easily estimated. While convenient mathematically, Daniels and Kass [7, p. 1174] noted that use of the HS loss function can result in “overshrinkage of the eigenvalues, especially the small ones.” As we shall see, this warning is well-founded in terms of the estimation of partial coherencies: the LW estimator for shrinking $\hat{\mathcal{S}}(f)$ wipes out large partial coherencies. In this paper we study alternatives to the LW-type estimator based on using the quadratic loss (QL) [16] rather than HS loss. QL was exploited [17] in shrinkage estimation for large dimensional covariance matrices but concentrating on singular estimators, with quite different estimators to those considered here.

We also look at estimating the precision matrix directly, rather than first inverting the estimated spectral matrix. [33] used QL for precision matrix shrinkage in the context of large dimensional covariance matrices and derived estimators via

random matrix theory. Here, we address this problem without requiring large dimensionality, and without recourse to random matrix theory. A weighted combination of the estimated inverse covariance matrix and the identity was considered by [8], [13], and forms the basis for our approach, using both the HS and QL losses.

Our simulations are based on real electroencephalogram (EEG) time series data ($p = 10$), which in fact motivated this study. Both large and small partial coherencies are present and the data is perfect for demonstrating the behaviour of the various estimators. The simulations use the exact circulant embedding methodology for vector-valued time series [4].

A. Contributions

Following some background on spectral estimation and the EEG data, the contributions of this paper are:

- 1) We derive two shrinkage estimators, QL_a and QL_b, for $\hat{S}(f)$ under QL loss in closed form. One involves just a single shrinkage parameter, while the other has two parameters.
- 2) The resulting oracle estimates of partial coherencies are compared to those from the LW scheme in terms of the percentage relative improvement in squared error over that of the raw estimator. The LW scheme shrinks partial coherencies towards zero in a manner that renders it unreliable in practice when any significant magnitude partial coherencies are present. The QL-based approaches perform much more robustly.
- 3) We next develop two-parameter shrinkage estimators, HSP and QLP, for the spectral precision matrix $C(f) \stackrel{\text{def}}{=} S^{-1}(f)$ under both HS and QL loss, respectively, both in closed form. The resulting oracle estimates of partial coherencies are compared to each other and to those of the other shrinkage estimators discussed above.
- 4) As a result of 3. above, QL_a, QL_b and QLP estimators are further considered in the real-world — non-oracle — setting where the trace terms have also to be estimated and thence renamed QL_a-est, QL_b-est and QLP-est. It is found that QLP-est behaves in a very appealing robust way for both high and low true coherencies and is recommended as our preferred shrinkage method.

Section II discusses the background spectral estimation, while Section III outlines the necessary preprocessing of the EEG data and our simulation from it, showing the different partial coherence profiles for two individuals utilised in this work. Section IV introduces our two new QL-based shrinkage estimators for the spectral matrix. The standard HS-based shrinkage estimator is contrasted with the new QL-based estimators in Section V in terms of eigenvalue adjustment, shrinkage parameters and accuracy in estimating the partial coherencies. New QL and HS estimators for precision matrices are derived in Section VI and the resultant partial coherence estimation is analysed. Since the QL approach has outperformed HS for the oracle estimators only the three estimators QL_a, QL_b and QLP are examined in Section VII where the full (non-oracle) estimators are examined, leading to QLP-est

being our recommended approach. Detailed derivations of all our new estimators are given in the Appendix.

II. SPECTRAL MATRIX ESTIMATION

A. Multitapering

We make use of a set of $K \geq p$ orthonormal tapers $\{h_{k,t}\}, k = 0, \dots, K-1$. A simple set are the sine tapers [27]. The elements of the k^{th} sinusoidal taper, are given by

$$h_{k,t} = \left[\frac{2}{N+1} \right]^{1/2} \sin \left[\frac{(k+1)\pi(t+1)}{N+1} \right], \quad t = 0, \dots, N-1.$$

For $t = 0, \dots, N-1$, form the product $h_{k,t} \mathbf{X}_t$ of the t th component of the k th taper with the t th component of the p -vector-valued process, and for $k = 0, \dots, K-1$ compute the vector Fourier transform $\mathbf{J}_k(f) \stackrel{\text{def}}{=} \Delta_t^{1/2} \sum_{t=0}^{N-1} h_{k,t} \mathbf{X}_t e^{-i2\pi f t \Delta_t}$. Let $\mathbf{J}(f)$ be the $p \times K$ matrix defined by $\mathbf{J}(f) = [\mathbf{J}_0(f), \dots, \mathbf{J}_{K-1}(f)]$. Then with $\hat{S}_k(f) \stackrel{\text{def}}{=} \mathbf{J}_k(f) \mathbf{J}_k^H(f)$ the multitaper estimator of the $p \times p$ spectral matrix $S(f)$ is

$$\hat{S}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_k(f) = \frac{1}{K} \mathbf{J}(f) \mathbf{J}^H(f). \quad (1)$$

Letting B denote the bandwidth of the spectral window corresponding to the tapering, then $\mathbf{J}_k(f), k = 0, \dots, K-1$, may be taken to be independently and identically distributed as p -vector-valued complex Gaussian with mean zero and covariance matrix $S(f)$:

$$\mathbf{J}_k(f) \stackrel{\text{d}}{=} \mathcal{N}_p^C\{\mathbf{0}, S(f)\}, \quad (2)$$

for $B/2 < |f| < f_N - B/2$ for finite N and Gaussian processes, or $0 < |f| < f_N$ asymptotically [3]. The bandwidth B for sine tapers is given by $B = (K+1)/[(N+1)\Delta_t]$, (e.g., [31]). The choice of K is therefore linked to the bandwidth chosen for the spectral window. It is assumed to have been chosen narrow enough to ensure the components of $S(f)$ are essentially constant across it. $\hat{S}(f)$ in (1) is the maximum-likelihood estimator for $S(f)$, [12].

Given (2), and with $K \geq p$, $K\hat{S}(f)$ has the complex Wishart distribution with mean $K S(f)$, written as

$$K\hat{S}(f) \stackrel{\text{d}}{=} \mathcal{W}_p^C\{K, S(f)\}. \quad (3)$$

Therefore,

$$E\{\hat{S}(f)\} = S(f) \quad \text{and} \quad E\{\text{tr}\{\hat{S}\}\} = \text{tr}\{S\}, \quad (4)$$

two simple results which will be made use of. Other properties following from (3) will be introduced where appropriate.

III. APPLICATION TO EEG DATA

We shall illustrate the methodology via data simulated from real electroencephalogram (EEG) data, (resting conditions with eyes closed), [24].

A. Preprocessing

Real EEG signals were recorded on the scalp at 10 sites, using a bandpass filter of 0.5–45Hz and sample interval of $\Delta_t = 0.01$ s. The recorded data $\{\mathbf{X}_t\}$ is thus a $p = 10$ vector-valued process, with Nyquist frequency 50Hz. To fully remove any influence of the highly dominant and contaminating 10Hz alpha rhythm, which would otherwise cause severe spectral leakage, a 4.6Hz low-pass filter was applied, followed by resampling to a sample interval of $\Delta_t = 0.05$ s giving a new Nyquist frequency of $f_N = 10$ Hz. After this downsampling each channel of data had $N = 1024$ time series values. The interesting EEG delta frequency range, $0.5 < f \leq 4$ Hz [24] should be reliably represented after this preprocessing.

B. Simulation Strategy

From this 10-channel series the spectral matrix $\mathbf{S}(f)$ was estimated as $\mathbf{S}_0(f)$, say, for $|f| \leq f_N$. Using the vector-valued circulant embedding approach, [4], a large number, M say, of independent Gaussian 10-channel time series were computed, each having $\mathbf{S}_0(f)$, $|f| \leq f_N$, as its true spectral matrix. For each of these time series the estimated spectral matrix $\hat{\mathbf{S}}(f)$ was computed using multitaper estimation with some specified number K of sine tapers. These M independent estimates of $\mathbf{S}_0(f)$ can then be used to deduce various sampling properties for quantities derived from $\hat{\mathbf{S}}(f)$. As mentioned in the Introduction, we will look at cases where K just exceeds p , so that ill-conditioning is present.

There are $p(p-1)/2 = 45$ partial coherencies as a function of frequency in this case. Fig. 1 shows the first 9 partial coherency plots ($j = 1, k = 2 : 10$) for individual 1 calculated from $\mathbf{S}_0(f)$. These will be referred to as the true partial coherencies (that we are trying to estimate from the simulated data). We can see that there is a good range of values from nearly zero up to around 0.8. We shall also look at true partial coherencies for individual 2 who has most partial coherencies close to zero (the ‘sparse’ case) — see Fig. 2. The two cases are summarized in Fig. 3 which shows that for individual one there are ‘spikes’ of high true partial coherence around 2 and 3.25 Hz. These will be significant for our partial coherence estimators.

Unless stated otherwise, results apply to individual 1.

IV. SHRINKING THE SPECTRAL MATRIX

A. Conventional Approach

The conventional approach to ‘covariance matrix’ regularization which has been extensively studied involves the forming of a convex combination of the sample covariance matrix and some well-conditioned ‘target’ matrix. For an estimated $p \times p$ Hermitian spectral matrix $\hat{\mathbf{S}}(f)$ this would take the form $\mathbf{S}^*(\rho(f)) = (1 - \rho(f))\hat{\mathbf{S}}(f) + \rho(f)\mathbf{T}(f)$, where $\rho(f) \in (0, 1)$ is known as the shrinkage parameter and $\mathbf{T}(f)$ is the target matrix. Provided $\hat{\mathbf{S}}(f)$ and $\mathbf{T}(f)$ are both positive definite, then this convex combination will itself be positive definite.

For notational brevity we shall drop the explicit frequency dependence in most of what follows.

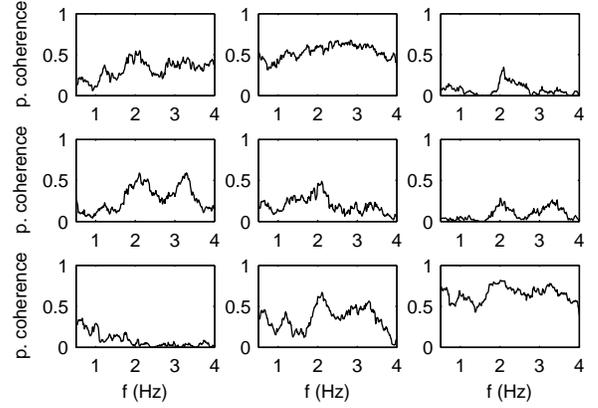


Fig. 1. The first 9 partial coherency plots ($j = 1, k = 2 : 10$) for individual one, found from the known matrix $\mathbf{S}_0(f)$.

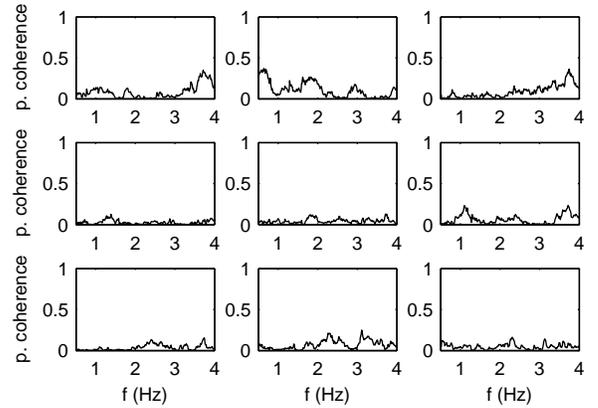


Fig. 2. The first 9 partial coherency plots ($j = 1, k = 2 : 10$) for individual two, found from the known matrix $\mathbf{S}_0(f)$.

The shrinkage coefficient ρ is set such that it minimizes a risk criterion for some given loss \mathcal{L} , say, of $\hat{\mathbf{S}}^*(\rho)$ to the true matrix \mathbf{S} :

$$\rho_0 = \arg \min_{\rho \in (0,1)} \mathcal{R}_{\mathcal{L}}(\hat{\mathbf{S}}^*(\rho), \mathbf{S}). \quad (5)$$

B. Hilbert-Schmidt loss

A common choice is the Hilbert-Schmidt (HS) loss where, with $\mathcal{L} = HS$

$$\begin{aligned} \mathcal{R}_{HS}(\hat{\mathbf{S}}^*(\rho), \mathbf{S}) &\stackrel{\text{def}}{=} E\{\text{tr}\{(\hat{\mathbf{S}}^*(\rho) - \mathbf{S})^2\}\} \\ &= E\{\|\mathbf{S}^*(\rho) - \mathbf{S}\|_F^2\}, \end{aligned} \quad (6)$$

where, for $\mathbf{A} \in \mathbb{C}^{p \times p}$, $\|\mathbf{A}\|_F$ denotes the Frobenius norm $\|\mathbf{A}\|_F = [\text{tr}\{\mathbf{A}\mathbf{A}^H\}]^{1/2}$, $\text{tr}\{\cdot\}$ denotes trace, and H denotes complex-conjugate (Hermitian) transpose.

Using (5) and (6) the target matrix was chosen to be of the form $\mathbf{T} = (\text{tr}\{\mathbf{S}\}/p)\mathbf{I}_p$ in [20] (for real-valued covariance matrices) so that

$$\mathbf{S}^*(\rho) = (1 - \rho)\hat{\mathbf{S}} + \rho \frac{\text{tr}\{\mathbf{S}\}}{p} \mathbf{I}_p. \quad (7)$$

For the ill-conditioned case ($p < K, p \simeq K$) the estimator diagonally loads the initial matrix $\hat{\mathbf{S}}$ and increases the zero or

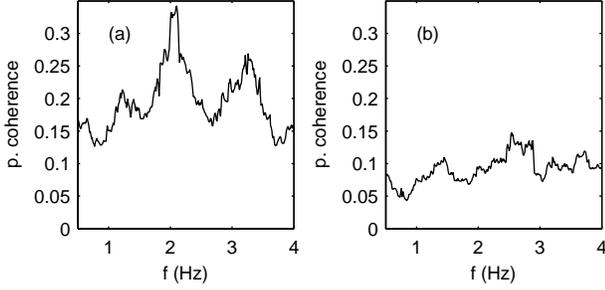


Fig. 3. Averaged true partial coherencies for (a) individual one, and (b) individual two, (averaging over all pairs (j, k) , $1 \leq j < k \leq p$).

near-zero eigenvalues. For this choice the solution to (5) and (6) is (e.g. [32, eq. 9]),

$$\rho_0 = \left[1 - \frac{K}{p} + K \frac{\text{tr}\{\mathbf{S}^2\}}{\text{tr}^2\{\mathbf{S}\}} \right]^{-1}. \quad (8)$$

C. Quadratic Loss

One alternative choice of loss criterion is the quadratic loss function:

$$\mathcal{R}_{QL}(\hat{\mathbf{S}}^*(\rho), \mathbf{S}) = E\{\text{tr}\{(\hat{\mathbf{S}}^*(\rho)\mathbf{S}^{-1} - \mathbf{I}_p)^2\}\}. \quad (9)$$

Lemma 1. With $\mathbf{S}^*(\rho)$ defined as in (7) the form of ρ_0 solving (5) and (9) is

$$\frac{p^2}{(Kp + p^2) - \frac{2}{p^2} [Kp \text{tr}\{\mathbf{S}^{-1}\} \text{tr}\{\mathbf{S}\} - \frac{K}{2} \text{tr}^2\{\mathbf{S}\} \text{tr}\{\mathbf{S}^{-2}\}]}. \quad (10)$$

Proof. This is given in Appendix C. \square

For easy identification we shall call this the QL_a method.

D. Generalization

The shrinkage concept can be extended to shrinkage estimators with generalised scaled identity targets,

$$\hat{\mathbf{S}}^*(\eta, \rho) = (1 - \rho)\hat{\mathbf{S}} + \rho\eta\mathbf{I}_p, \quad \eta > 0, \rho \in (0, 1) \quad (11)$$

with the optimum shrinkage (ρ_0) and scaling parameter values (η_0) given by

$$(\eta_0, \rho_0) = \arg \min_{\eta > 0, \rho \in (0, 1)} \mathcal{R}_{\mathcal{L}}(\hat{\mathbf{S}}^*(\eta, \rho), \mathbf{S}), \quad (12)$$

for any convex loss criterion \mathcal{L} .

Lemma 2. The solutions of (12) for both the HS and QL losses are

$$(\eta_0, \rho_0)_{\text{HS}} = \left(\frac{\text{tr}\{\mathbf{S}\}}{p}, \left[1 - \frac{K}{p} + K \frac{\text{tr}\{\mathbf{S}^2\}}{\text{tr}^2\{\mathbf{S}\}} \right]^{-1} \right), \quad (13)$$

$$(\eta_0, \rho_0)_{\text{QL}} = \left(\frac{\text{tr}\{\mathbf{S}^{-1}\}}{\text{tr}\{\mathbf{S}^{-2}\}}, \left[1 + \frac{K}{p} - \frac{K \text{tr}^2\{\mathbf{S}^{-1}\}}{p^2 \text{tr}\{\mathbf{S}^{-2}\}} \right]^{-1} \right). \quad (14)$$

Proof. This is given in Appendix D. \square

Remark 1. We see that the choice of target $\mathbf{T} = (\text{tr}\{\mathbf{S}\}/p)\mathbf{I}_p$ is identical to η_0 under the HS loss, i.e., the scaling factor

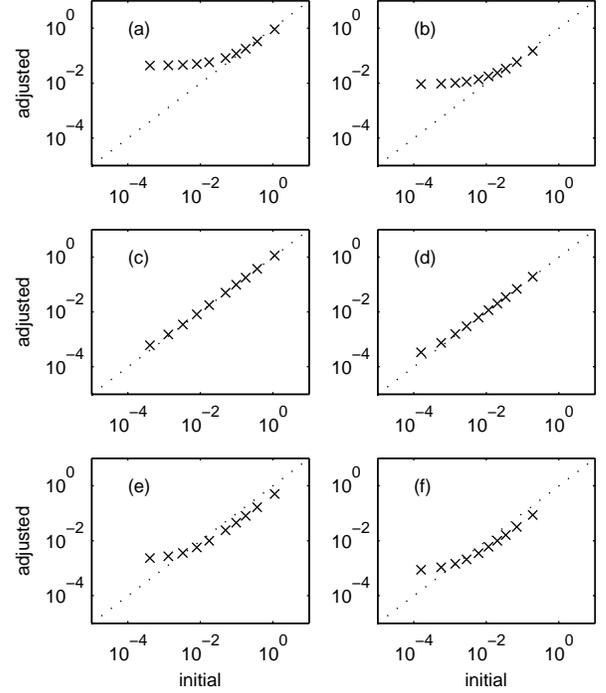


Fig. 4. Average eigenvalue adjustment for frequencies $f = 0.85\text{Hz}$ (left column) and 3.85Hz (right column). The first row is for the HS method, the second for QL_a and the third for QL_b. $K = 12$ here.

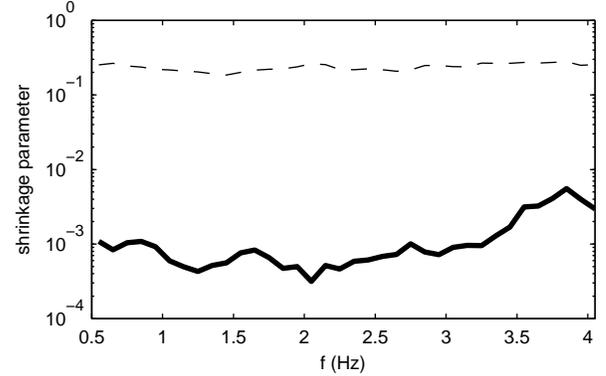


Fig. 5. Oracle value of $\rho(f)$ for HS (dashed line) and QL_a (thick line) for frequencies 0.55 to 4.05 in steps of 0.1.

$(\text{tr}\{\mathbf{S}\}/p)$ is an optimal choice under the HS loss for shrinkage model (11). Since the targets are then the same the form of ρ_0 was already known from (8).

We shall call the method defined by (13) the HS method. We also call the method defined by (14) the QL_b method.

In deriving the optimal forms of η_0 and ρ_0 (Appendix D), we note that the form of η_0 for both losses arises independently of ρ_0 . This leads to the following result.

Lemma 3. If we reparameterize the shrinkage model in (11) to $\hat{\mathbf{S}}^*(\alpha, \beta) = \alpha\hat{\mathbf{S}} + \beta\mathbf{I}_p$, $\alpha, \beta > 0$, it follows that (α_0, β_0) defined by $(\alpha_0, \beta_0) = \arg \min_{\alpha, \beta > 0} \mathcal{R}_{\mathcal{L}}(\hat{\mathbf{S}}^*(\alpha, \beta), \mathbf{S})$, are given by $(\alpha_0, \beta_0)_{\text{HS or QL}} = ((1 - \rho_0), \rho_0\eta_0)_{\text{HS or QL}}$.

Proof. See Appendix E. \square

V. EXAMPLE ORACLE BEHAVIOUR

The above oracle solutions require exact knowledge of the true spectral matrix. Before looking at estimation methods, we examine some aspects of the oracle solutions.

A. Transformation of Eigenvalues

Minimizing the HS and QL risks effectively transforms the raw eigenvalues of $\hat{\mathbf{S}}$. Fig. 4 shows the transformation of the $p = 10$ eigenvalues for HS, QLa and QLb at frequencies 0.8Hz and 3.85Hz. Both the initial and adjusted eigenvalues shown are averages, following sorting, over the $M = 500$ simulations of $\hat{\mathbf{S}}$. The main feature, which is persistent at other frequencies and data sets, is that the HS method leads to a large increase in the small eigenvalues and a very small decrease in the larger eigenvalues. The QLa method increases only the smallest eigenvalues by a small amount, and barely changes the large ones. The QLb method increases the small eigenvalues more than QLa but much less so than HS and decreases the larger eigenvalues by a larger amount than HS. Such behaviour will have a significant effect in the estimation of partial coherencies.

B. Shrinkage Parameters

For methods HS and QLa the shrinkage parameter ρ may be compared directly as the model is the same in both cases. Fig. 5 shows the the oracle shrinkage parameters. From Fig. 4(a),(b) and (c),(d) we are not surprised to see that the shrinkage parameter is orders of magnitude larger for HS than for QLa. For QLa it varies between around 10^{-3} to 10^{-2} while for HS it is around 0.25.

C. Partial coherence

The 45 partial coherencies at each frequency can be quite variable in size from near zero to near unity. Let $\hat{\gamma}_{jk \bullet \{ \setminus jk \}}^2(m; f_l)$ denote the estimate – by any method – of the partial coherence at frequency f_l and replication number m where $m = 1, \dots, M$. To measure the quality of the estimates of partial coherencies we firstly calculated the sum of squared errors over all pairs $(j, k), 1 \leq j < k \leq p$:

$$\mathcal{E}(m; f_l) = \sum_{j,k} \left[\hat{\gamma}_{jk \bullet \{ \setminus jk \}}^2(m; f_l) - \gamma_{jk \bullet \{ \setminus jk \}}^2(f_l) \right]^2.$$

We then averaged these over replications to get $\bar{\mathcal{E}}(f_l) = (1/M) \sum_{m=1}^M \mathcal{E}(m; f_l)$. Finally a percentage relative improvement in squared error (PRISE) was computed as

$$\text{PRISE}(f_l) = 100 \left[\frac{\bar{\mathcal{E}}_B(f_l) - \bar{\mathcal{E}}_M(f_l)}{\bar{\mathcal{E}}_B(f_l)} \right] \%,$$

where $\bar{\mathcal{E}}_B(f)$ denotes the average summed squared errors of the raw or basic estimate, and $\bar{\mathcal{E}}_M(f)$ denotes the same for any of the HS, QLa or QLb shrinkage methods.

We are thus able to determine the improvement in using any of the three shrinkage methods over using just the raw estimates. Results are shown in Fig. 6 with results at frequencies 0.55 to 4.05 in steps of 0.1. These are oracle results: the known quantities

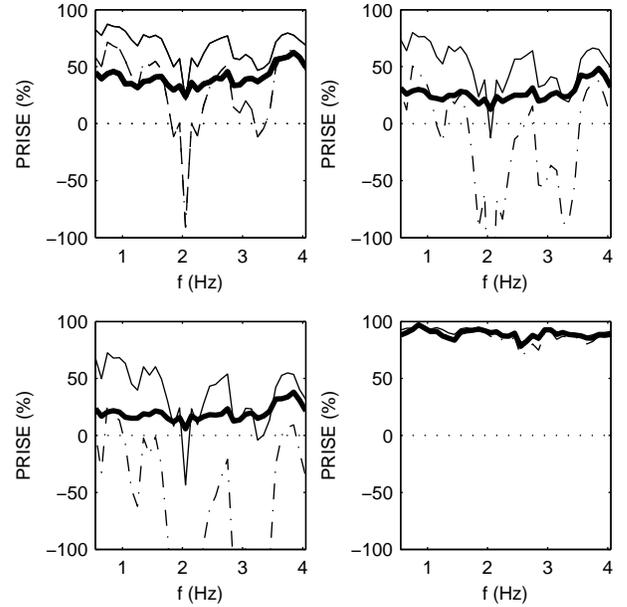


Fig. 6. Oracle PRISE with partial coherencies derived from spectral matrices. HS (dash-dot line), QLa (thick line) and QLb (thin line) for $K = 12$ (top left), $K = 14$ (top right), $K = 16$ (bottom left), all for individual one, and $K = 12$ (bottom right) for individual two.

method	individual one			individual two
	$K = 12$	$K = 14$	$K = 16$	$K = 12$
HS	31	-17	-66	87
QLa	41	27	20	89
QLb	68	51	37	91
HSP	27	7	-7	82
QLP	52	32	23	85
QLa-est	26	18	13	66
QLb-est	65	49	34	87
QLP-est	52	34	25	84

TABLE I
AVERAGE PRISE(%) OVER FREQUENCIES.

$\text{tr}\{\mathbf{S}_0(f)\}, \text{tr}\{\mathbf{S}_0^2(f)\}, \text{tr}\{\mathbf{S}_0^{-1}(f)\}, \text{tr}\{\mathbf{S}_0^{-2}(f)\}$ have been used in the various estimators.

We also averaged the PRISEs over the frequencies to obtain Table I. Looking at the top three rows for individual one, we see that the HS method does very poorly, being worse than using the raw estimates. QLb generally does well but is very variable (Fig. 6) and rather unpredictable with frequency. QLa seems to behave very nicely giving a fairly frequency-constant improvement over using the raw estimates. In the sparse non-null partial coherence case of patient two, all methods do well.

VI. SHRINKING THE PRECISION MATRIX

Since the partial coherence is derived from the precision matrix $\mathbf{C}(f) = \mathbf{S}^{-1}(f)$ we can also consider shrinkage for this matrix. A possible approach is to take

$$\hat{\mathbf{C}}^*(\alpha, \beta) = \alpha \hat{\mathbf{S}}^{-1} + \beta \mathbf{I}_p, \quad \alpha, \beta > 0. \quad (15)$$

Such a model appears in [8] and [13] with the slight modification that the right side takes the form

$$\alpha \hat{\mathbf{S}}^{-1} + \beta / \text{tr}\{\hat{\mathbf{S}}\} \mathbf{I}_p. \quad (16)$$

Lemma 4. Under the model (15), the HS risk $\mathcal{R}_{HS}(\hat{\mathbf{C}}^*(\alpha, \beta), \mathbf{C}) = E\{\text{tr}\{(\hat{\mathbf{C}}^*(\alpha, \beta) - \mathbf{S}^{-1})^2\}$ is minimized, for $K > p + 1$, by

$$\begin{aligned} \alpha_0 &= \frac{1}{D} \left[p \frac{\text{tr}\{\mathbf{S}^{-2}\}}{\text{tr}^2\{\mathbf{S}^{-1}\}} - 1 \right] \\ \beta_0 &= \frac{c_3 \text{tr}\{\mathbf{S}^{-1}\}}{D} \left[\frac{\text{tr}\{\mathbf{S}^{-2}\}}{\text{tr}^2\{\mathbf{S}^{-1}\}} + (K - p) \right], \end{aligned} \quad (17)$$

where

$$\begin{aligned} c_3 &= \frac{K}{(K - p)^3 - (K - p)} \\ D &= \left[c_3 p (K - p)^2 \frac{\text{tr}\{\mathbf{S}^{-2}\}}{\text{tr}^2\{\mathbf{S}^{-1}\}} + c_3 p (K - p) - \frac{K}{(K - p)} \right]. \end{aligned}$$

Proof. See Appendix F. \square

We shall call this the HSP method.

Lemma 5. Under the model (15), the QL risk $\mathcal{R}_{QL}(\hat{\mathbf{C}}^*(\alpha, \beta), \mathbf{C}) = E\{\text{tr}\{(\hat{\mathbf{C}}^*(\alpha, \beta)\mathbf{S} - \mathbf{I}_p)^2\}$ is minimized, for $K > p + 1$, by

$$\alpha_0 = \frac{1}{D} \left[p \frac{\text{tr}\{\mathbf{S}^2\}}{\text{tr}^2\{\mathbf{S}\}} - 1 \right]; \quad \beta_0 = \frac{p}{D \text{tr}\{\mathbf{S}\}} \left[c_0 - \frac{K}{K - p} \right] \quad (18)$$

where

$$c_0 = \frac{K^2}{(K - p)^2 - 1}; \quad D = \left[c_0 p \frac{\text{tr}\{\mathbf{S}^2\}}{\text{tr}^2\{\mathbf{S}\}} - \frac{K}{K - p} \right]. \quad (19)$$

Proof. See Appendix G. \square

We shall call this the QLP method.

Remark 2. We note that the numerator of α_0 in (17) is essentially the U -statistic for testing for sphericity of \mathbf{C} and in (18) for testing for sphericity of \mathbf{S} .

Repeating the simulations in Section V-C using the oracle estimates $\hat{\mathbf{C}}^*(\alpha_0, \beta_0)$ for the precision matrices, and hence the partial coherencies, gave the results in Fig. 7 and lines 4 and 5 of Table I. We see that HSP and HS are comparable for patient one for $K = 12$ but for $K = 14$ and 16, HSP does better. For individual two HSP is slightly worse than HS when $K = 12$. QLP performs much better than HSP, and for individual one performs intermediate to QLa and QLb.

VII. FULL (NON-ORACLE) ESTIMATION

Each of the estimators we have derived involve the traces of some subset of \mathbf{S} , \mathbf{S}^2 , \mathbf{S}^{-1} , \mathbf{S}^{-2} . For the oracle estimators these were taken as known, but now we turn to the real-world case where these must also be estimated. Since the QL approach has performed uniformly better than HS for the oracle estimators we now only examine the three estimators QLa, QLb and QLP.

Now QLP involves only $\text{tr}\{\mathbf{S}\}$ and $\text{tr}\{\mathbf{S}^2\}$ whereas QLa and QLb involve $\text{tr}\{\mathbf{S}^{-1}\}$ and $\text{tr}\{\mathbf{S}^{-2}\}$, which we would expect to be more problematic to estimate.

We recall that $\hat{\mathbf{S}}$ is the maximum likelihood estimator (MLE) for \mathbf{S} . By the invariance property of MLEs [2, p. 294] it follows that the MLE of $\text{tr}\{\mathbf{S}\}$ is $\text{tr}\{\hat{\mathbf{S}}\}$, of $\text{tr}\{\mathbf{S}^2\}$ is $\text{tr}\{\hat{\mathbf{S}}^2\}$,

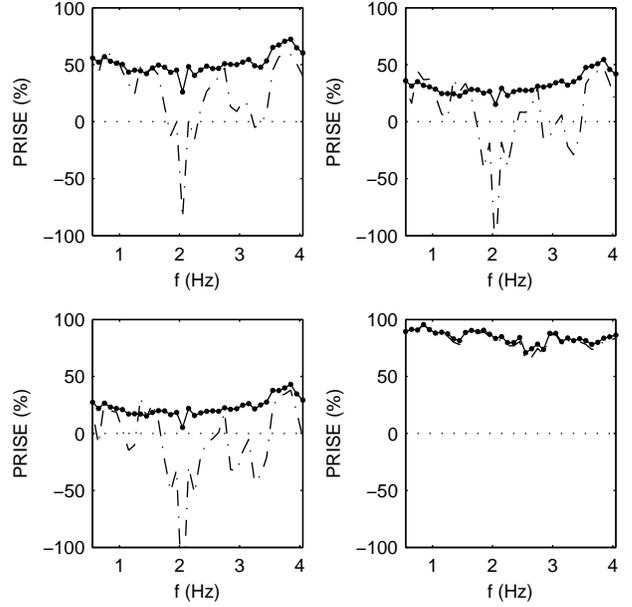


Fig. 7. Oracle PRISE with partial coherencies derived from precision matrices. HSP (dash-dot line) and QLP (line with bold dots) for $K = 12$ (top left), $K = 14$ (top right), $K = 16$ (bottom left), all for individual one, and $K = 12$ (bottom right) for individual two.

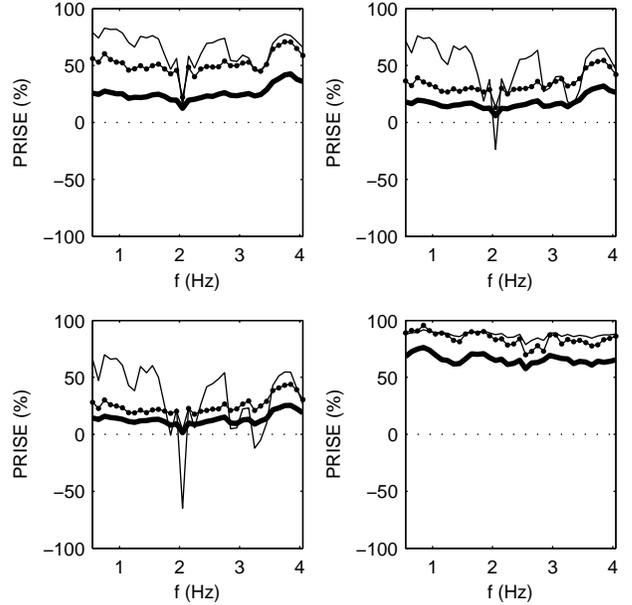


Fig. 8. PRISE with partial coherencies derived from fully estimated spectral or precision matrices. QLa-est (thick line), QLb-est (thin line) and QLP-est (line with bold dots) for $K = 12$ (top left), $K = 14$ (top right), $K = 16$ (bottom left), all for individual one, and $K = 12$ (bottom right) for individual two.

of $\text{tr}\{\mathbf{S}^{-1}\}$ is $\text{tr}\{\hat{\mathbf{S}}^{-1}\}$ and of $\text{tr}\{\mathbf{S}^{-2}\}$ is $\text{tr}\{\hat{\mathbf{S}}^{-2}\}$. The MLE is a strongly consistent estimator [35, p. 129].

- To estimate $\text{tr}\{\mathbf{S}\}$ we use $\text{tr}\{\hat{\mathbf{S}}\}$ which by (4) is exactly unbiased.
- Our estimator for $\text{tr}\{\mathbf{S}^2\}$ is $\text{tr}\{\hat{\mathbf{S}}^2\} - (1/K)\text{tr}^2\{\hat{\mathbf{S}}\}$. Using (21) and (22) we see that

$$E \left\{ \text{tr}\{\hat{\mathbf{S}}^2\} - \frac{1}{K} \text{tr}^2\{\hat{\mathbf{S}}\} \right\} = \left[1 - \frac{1}{K^2} \right] \text{tr}\{\mathbf{S}^2\},$$

so the estimator is asymptotically unbiased for $\text{tr}\{\mathbf{S}^2\}$.

- To estimate $\text{tr}\{\mathbf{S}^{-1}\}$ we use $\left[1 - \frac{p}{K}\right] \text{tr}\{\hat{\mathbf{S}}^{-1}\}$. Result (23) shows this estimator is exactly unbiased (under the assumption $K > p$).
- Results (24) and (25) together show that

$$\left[1 - \frac{p}{K}\right]^2 \text{tr}\{\hat{\mathbf{S}}^{-2}\} - \frac{1}{K} \left[1 - \frac{p}{K}\right] \text{tr}^2\{\hat{\mathbf{S}}^{-1}\} \quad (20)$$

is unbiased for $\text{tr}\{\mathbf{S}^{-2}\}$ with $K > p + 1$. However this estimator was found by simulation to have a very high variance, with occasional negative values. Better results were obtained by estimating $\text{tr}\{\mathbf{S}^{-2}\}$ using just the first term of (20), namely $\left[1 - \frac{p}{K}\right]^2 \text{tr}\{\hat{\mathbf{S}}^{-2}\}$, which is only asymptotically unbiased.

These estimators of the trace terms were used in the three estimators, now called QLa-est, QLb-est and QLP-est for this full estimation situation. Repeating the simulations in Section V-C for just the three estimators gave the results in Fig. 8 and lines 6 to 8 of Table I. For individual one QLb-est performs best but at frequencies where the true partial coherence is high (2 and 3.25 Hz) it does poorly. It is generally quite variable. QLP-est seems to deliver very significant improvements over the raw estimates, and whilst generally slightly inferior to QLb-est, is relatively frequency invariant and behaves better for high partial coherencies. For individual two QLb-est slightly outperforms QLP-est, but both do very well in this sparse case. QLa-est is generally inferior to the other two but appears relatively unaffected by spikes of high true partial coherence.

In closing, we note that (18) and (19) involve only \mathbf{S} and \mathbf{S}^2 which are more accurately estimated than their inverses which appear in QLa-est and QLb-est; this undoubtedly adds greatly to the reliability and robustness of QLP-est. We also note that with β_0 in the form given in (18) that there is a $\text{tr}\{\mathbf{S}\}$ in the denominator which is playing the role of $\text{tr}\{\hat{\mathbf{S}}\}$ in the previously accepted formulation of (16). Hence we recommend QLP-est as our estimator of choice.

VIII. CONCLUDING DISCUSSION

We have considered how to improve estimation of partial coherencies from poorly-conditioned matrices. Our study carried out exact simulations derived from real EEG data for two individuals, one having large, and the other small, partial coherencies. When true partial coherencies are quite large then for shrinkage estimators of the diagonal weighting kind — for spectral matrices or precision matrices — minimization of risk using QL leads to oracle partial coherence estimators superior to HS equivalents. When true partial coherencies are small the methods behave similarly. For the full estimation (non-oracle) case, QLP-est seems particularly robust and reliable and based on the results here is our recommended approach.

APPENDIX

To simplify notation we drop explicit frequency dependence.

A. Useful Expectations

The first three of the following results are given in [22]. The third and fourth can be deduced from [21, p. 308]. Let \mathbf{A}, \mathbf{B} be arbitrary complex-valued matrices. Under (3),

$$E\{\text{tr}\{\mathbf{A}\hat{\mathbf{S}}\mathbf{B}\hat{\mathbf{S}}\}\} = \text{tr}\{\mathbf{A}\mathbf{S}\mathbf{B}\mathbf{S}\} + \frac{1}{K} \text{tr}\{\mathbf{A}\mathbf{S}\}\text{tr}\{\mathbf{B}\mathbf{S}\} \quad (21)$$

$$E\{\text{tr}^2\{\hat{\mathbf{S}}\}\} = \text{tr}^2\{\mathbf{S}\} + \frac{1}{K} \text{tr}\{\mathbf{S}^2\} \quad (22)$$

$$E\{\text{tr}\{\hat{\mathbf{S}}^{-1}\}\} = \frac{K}{K-p} \text{tr}\{\mathbf{S}^{-1}\} \quad (23)$$

$$E\{\text{tr}\{\mathbf{A}\hat{\mathbf{S}}^{-1}\mathbf{B}\hat{\mathbf{S}}^{-1}\}\} = c_1 [(K-p)\text{tr}\{\mathbf{A}\mathbf{S}^{-1}\mathbf{B}\mathbf{S}^{-1}\} + \text{tr}\{\mathbf{A}\mathbf{S}^{-1}\}\text{tr}\{\mathbf{B}\mathbf{S}^{-1}\}] \quad (24)$$

$$E\{\text{tr}^2\{\hat{\mathbf{S}}^{-1}\}\} = c_1 [(K-p)\text{tr}^2\{\mathbf{S}^{-1}\} + \text{tr}\{\mathbf{S}^{-2}\}]. \quad (25)$$

Here (21) and (22) hold for $K > p - 1$, (23) holds for $K > p$ and (24) and (25) hold for $K > p + 1$. c_1 is given by

$$c_1 = \frac{K^2}{(K-p)^3 - (K-p)}.$$

B. Derivative of trace of squared matrix

If x is a scalar, $\mathbf{F}(x)$ is an $m \times n$ matrix and $\mathbf{G}(x)$ is an $n \times q$ matrix, then [15, p. 301] $\partial(\mathbf{F}\mathbf{G})/\partial x = \mathbf{F}(\partial\mathbf{G}/\partial x) + (\partial\mathbf{F}/\partial x)\mathbf{G}$. Also, [15, p. 304] $\partial \text{tr}\{\mathbf{F}\}/\partial x = \text{tr}\{\partial\mathbf{F}/\partial x\}$. So,

$$\begin{aligned} \frac{\partial \text{tr}\{\mathbf{F}\mathbf{G}\}}{\partial x} &= \text{tr}\left\{\mathbf{F}\frac{\partial\mathbf{G}}{\partial x} + \frac{\partial\mathbf{F}}{\partial x}\mathbf{G}\right\} \\ &= \text{tr}\left\{\mathbf{F}\frac{\partial\mathbf{G}}{\partial x}\right\} + \text{tr}\left\{\mathbf{G}\frac{\partial\mathbf{F}}{\partial x}\right\} \\ &= \text{tr}\left\{\frac{\partial\mathbf{G}}{\partial x}\mathbf{F}\right\} + \text{tr}\left\{\frac{\partial\mathbf{F}}{\partial x}\mathbf{G}\right\}, \end{aligned}$$

so if $\mathbf{F} = \mathbf{G}$,

$$\frac{\partial \text{tr}\{\mathbf{F}^2\}}{\partial x} = 2 \text{tr}\left\{\frac{\partial\mathbf{F}}{\partial x}\mathbf{F}\right\} = 2 \text{tr}\left\{\mathbf{F}\frac{\partial\mathbf{F}}{\partial x}\right\}. \quad (26)$$

C. Proof of Lemma 1

From (9) the quadratic loss is $\mathcal{R}_{QL}(\hat{\mathbf{S}}^*(\rho), \mathbf{S}) = E\{\text{tr}\{(\hat{\mathbf{S}}^*\mathbf{S}^{-1} - \mathbf{I}_p)^2\}\}$. So

$$\frac{\partial}{\partial \rho} \mathcal{R}_{QL}(\hat{\mathbf{S}}^*(\rho), \mathbf{S}) = E\left\{\frac{\partial}{\partial \rho} \text{tr}\{(\hat{\mathbf{S}}^*\mathbf{S}^{-1} - \mathbf{I}_p)^2\}\right\}.$$

We now use (26) to find the derivative by setting

$$\mathbf{F} = \hat{\mathbf{S}}^*\mathbf{S}^{-1} - \mathbf{I}_p = [(1-\rho)\hat{\mathbf{S}} + \frac{\rho}{p}\text{tr}\{\mathbf{S}\}\mathbf{I}_p]\mathbf{S}^{-1} - \mathbf{I}_p.$$

Then,

$$\begin{aligned} \frac{\partial\mathbf{F}}{\partial\rho}\mathbf{F} &= \left(-\hat{\mathbf{S}}\mathbf{S}^{-1} + \frac{1}{p}\text{tr}\{\mathbf{S}\}\mathbf{S}^{-1}\right) \\ &\times \left([(1-\rho)\hat{\mathbf{S}} + \frac{\rho}{p}\text{tr}\{\mathbf{S}\}\mathbf{I}_p]\mathbf{S}^{-1} - \mathbf{I}_p\right) \\ &= -(1-\rho)\hat{\mathbf{S}}\mathbf{S}^{-1}\hat{\mathbf{S}}\mathbf{S}^{-1} - \frac{\rho}{p}\text{tr}\{\mathbf{S}\}\hat{\mathbf{S}}\mathbf{S}^{-2} + \hat{\mathbf{S}}\mathbf{S}^{-1} \\ &+ \frac{(1-\rho)}{p}\text{tr}\{\mathbf{S}\}\mathbf{S}^{-1}\hat{\mathbf{S}}\mathbf{S}^{-1} + \frac{\rho}{p^2}\text{tr}^2\{\mathbf{S}\}\mathbf{S}^{-2} \\ &- \frac{1}{p}\text{tr}\{\mathbf{S}\}\mathbf{S}^{-1}. \end{aligned}$$

Using (21),

$$\begin{aligned} E\{\text{tr}\{\mathbf{S}^{-1}\hat{\mathbf{S}}\mathbf{S}^{-1}\hat{\mathbf{S}}\}\} &= [\text{tr}\{\mathbf{S}^{-1}\mathbf{S}\mathbf{S}^{-1}\mathbf{S}\} + \frac{1}{K}\text{tr}^2\{\mathbf{S}^{-1}\mathbf{S}\}] \\ &= \left[p + \frac{p^2}{K}\right]. \end{aligned} \quad (27)$$

Using (27) we obtain

$$\begin{aligned} \frac{\partial}{\partial \rho} \mathcal{R}_{QL}(\hat{\mathbf{S}}^*(\rho), \mathbf{S}) &= E\left\{2\text{tr}\left\{\frac{\partial \mathbf{F}}{\partial \rho} \mathbf{F}\right\}\right\} \\ &= [-2 + 2\rho]\left[p + \frac{p^2}{K}\right] - \frac{4\rho}{p}\text{tr}\{\mathbf{S}\}\text{tr}\{\mathbf{S}^{-1}\} \\ &\quad + 2\frac{\rho}{p^2}\text{tr}^2\{\mathbf{S}\}\text{tr}\{\mathbf{S}^{-2}\} + 2p. \end{aligned}$$

setting the result to zero and tidying gives

$$\rho\left(\left[p + \frac{p^2}{K}\right] - \frac{2}{p}\text{tr}\{\mathbf{S}^{-1}\}\text{tr}\{\mathbf{S}\} + \frac{1}{p^2}\text{tr}^2\{\mathbf{S}\}\text{tr}\{\mathbf{S}^{-2}\}\right) = \frac{p^2}{K}$$

which gives the expression (10) for ρ .

The second derivative is given by

$$\begin{aligned} \frac{\partial^2}{\partial \rho^2} \mathcal{R}_{QL}(\hat{\mathbf{S}}^*(\rho), \mathbf{S}) &= 2\left[p + \frac{p^2}{K}\right] - \frac{4}{p}\text{tr}\{\mathbf{S}\}\text{tr}\{\mathbf{S}^{-1}\} \\ &\quad + \frac{2}{p^2}\text{tr}^2\{\mathbf{S}\}\text{tr}\{\mathbf{S}^{-2}\}. \end{aligned}$$

We now write this as a quadratic in $\text{tr}\{\mathbf{S}\}$, i.e., as $a\text{tr}^2\{\mathbf{S}\} - b\text{tr}\{\mathbf{S}\} + c$, where

$$a = \frac{2}{p^2}\text{tr}\{\mathbf{S}^{-2}\}; \quad b = \frac{4}{p}\text{tr}\{\mathbf{S}^{-1}\}; \quad c = 2\left[p + \frac{p^2}{K}\right].$$

Next complete the square to obtain $(\sqrt{a}\text{tr}\{\mathbf{S}\} - b/[2\sqrt{a}])^2 - b^2/(4a) + c$. This will be positive if $c - (b^2/4a)$ is positive; i.e.,

$$2\left[p + \frac{p^2}{K}\right] - 2\frac{\text{tr}^2\{\mathbf{S}^{-1}\}}{\text{tr}\{\mathbf{S}^{-2}\}} > 0.$$

By Chebyshev's inequality we have $p\sum_{j=1}^p \lambda_j^{-2} \geq \left(\sum_{j=1}^p \lambda_j^{-1}\right)^2$, where the λ_j 's are eigenvalues of \mathbf{S} . So $p \geq \text{tr}^2\{\mathbf{S}^{-1}\}/\text{tr}\{\mathbf{S}^{-2}\}$, and so

$$2\left[p + \frac{p^2}{K}\right] - 2\frac{\text{tr}^2\{\mathbf{S}^{-1}\}}{\text{tr}\{\mathbf{S}^{-2}\}} \geq 2\frac{p^2}{K} > 0,$$

which proves that the turning point is a minimum.

D. Proof of Lemma 2

For the HS loss we have that $(\partial/\partial\eta)\mathcal{R}_{HS}(\hat{\mathbf{S}}^*(\eta, \rho), \mathbf{S})$ is given by $(\partial/\partial\eta)E\{\text{tr}\{[(1-\rho)\hat{\mathbf{S}} + \eta\rho\mathbf{I}_p - \mathbf{S}]^2\}\}$ which is

$$\begin{aligned} E\{\text{tr}\{2\rho\mathbf{I}_p[(1-\rho)\hat{\mathbf{S}} + \eta\rho\mathbf{I}_p - \mathbf{S}]\}\} &= 2\rho\text{tr}\{\eta\rho\mathbf{I}_p - \rho\mathbf{S}\} \\ &= 2\rho^2(\eta p - \text{tr}\{\mathbf{S}\}), \end{aligned}$$

where we have used (26) in Appendix B and $E\{\hat{\mathbf{S}}\} = \mathbf{S}$. Then setting to zero gives

$$\eta_0 = \text{tr}\{\mathbf{S}\}/p. \quad (28)$$

Next, $(\partial/\partial\rho)\mathcal{R}_{HS}(\hat{\mathbf{S}}^*(\eta, \rho), \mathbf{S})$ is given by $(\partial/\partial\rho)E\{\text{tr}\{[(1-\rho)\hat{\mathbf{S}} + \eta\rho\mathbf{I}_p - \mathbf{S}]^2\}\}$ which is

$$\begin{aligned} &E\{\text{tr}\{2(\eta\mathbf{I}_p - \hat{\mathbf{S}})[(1-\rho)\hat{\mathbf{S}} + \eta\rho\mathbf{I}_p - \mathbf{S}]\}\} \\ &= E\{\text{tr}\{2(\eta\mathbf{I}_p - \hat{\mathbf{S}})[\rho(\eta\mathbf{I}_p - \hat{\mathbf{S}}) + \hat{\mathbf{S}} - \mathbf{S}]\}\} \\ \Rightarrow \rho_0 &\stackrel{\text{def}}{=} \rho_0(\eta_0) = \frac{E\{\text{tr}\{(\eta_0\mathbf{I}_p - \hat{\mathbf{S}})(\mathbf{S} - \hat{\mathbf{S}})\}\}}{E\{\text{tr}\{(\eta_0\mathbf{I}_p - \hat{\mathbf{S}})^2\}\}}. \end{aligned} \quad (29)$$

The optimal shrinkage coefficient is dependent on the form of the target matrix, namely $\eta\mathbf{I}_p$. Look first at the numerator, a say, of (29). Expanding gives

$$a = E\{\text{tr}\{\eta_0\mathbf{S} - \eta_0\hat{\mathbf{S}} - \hat{\mathbf{S}}\mathbf{S} + \hat{\mathbf{S}}^2\}\}.$$

Now $K\hat{\mathbf{S}}$ has the complex Wishart distribution with mean $K\mathbf{S}$. Since $E\{\hat{\mathbf{S}}\} = \mathbf{S}$ the first two terms cancel. Using (22) we get

$$a = -\text{tr}\{\mathbf{S}^2\} + \text{tr}\{\mathbf{S}^2\} + \frac{1}{K}\text{tr}^2\{\mathbf{S}\} = \frac{1}{K}\text{tr}^2\{\mathbf{S}\}.$$

Expanding the denominator, b say, of (29) gives, with $\eta_0 = \text{tr}\{\mathbf{S}\}/p$,

$$\begin{aligned} b &= E\{\text{tr}\{\eta_0^2\mathbf{I}_p - 2\eta_0\hat{\mathbf{S}} + \hat{\mathbf{S}}^2\}\} \\ &= \frac{1}{p}\text{tr}^2\{\mathbf{S}\} - \frac{2}{p}\text{tr}^2\{\mathbf{S}\} + \text{tr}\{\mathbf{S}^2\} + \frac{1}{K}\text{tr}^2\{\mathbf{S}\} \\ &= \left[\frac{1}{K} - \frac{1}{p}\right]\text{tr}^2\{\mathbf{S}\} + \text{tr}\{\mathbf{S}^2\}. \end{aligned}$$

The ratio a/b then has the form (8) or (13).

For the QL loss, we have that $(\partial/\partial\eta)\mathcal{R}_{QL}(\hat{\mathbf{S}}^*(\eta, \rho), \mathbf{S})$ is given by $(\partial/\partial\eta)E\{\text{tr}\{[(1-\rho)\hat{\mathbf{S}} + \eta\rho\mathbf{I}_p]\mathbf{S}^{-1} - \mathbf{I}_p\}^2\}$ which is

$$\begin{aligned} &E\{\text{tr}\{2\rho\mathbf{S}^{-1}([(1-\rho)\hat{\mathbf{S}} + \eta\rho\mathbf{I}_p]\mathbf{S}^{-1} - \mathbf{I}_p)\}\} \\ &= -2\rho^2\text{tr}\{\mathbf{S}^{-1}\} + 2\rho^2\eta\text{tr}\{\mathbf{S}^{-2}\}, \\ \Rightarrow \eta_0 &= \text{tr}\{\mathbf{S}^{-1}\}/\text{tr}\{\mathbf{S}^{-2}\}. \end{aligned} \quad (30)$$

Here we have again used (26) in Appendix B and $E\{\hat{\mathbf{S}}\} = \mathbf{S}$. This is very different in form to the scaling parameter (28) under HS loss. Then $(\partial/\partial\rho)\mathcal{R}_{QL}(\hat{\mathbf{S}}^*(\eta, \rho), \mathbf{S})$ is given by $(\partial/\partial\rho)E\{\text{tr}\{[(1-\rho)\hat{\mathbf{S}} + \eta\rho\mathbf{I}_p]\mathbf{S}^{-1} - \mathbf{I}_p\}^2\}$ which is

$$\begin{aligned} &E\{\text{tr}\{2(\eta\mathbf{I}_p - \hat{\mathbf{S}})\mathbf{S}^{-1}([(1-\rho)\hat{\mathbf{S}} + \eta\rho\mathbf{I}_p]\mathbf{S}^{-1} - \mathbf{I}_p)\}\} \\ &= 2E\{\text{tr}\{\rho(\eta\mathbf{I}_p - \hat{\mathbf{S}})\mathbf{S}^{-1}(\eta\mathbf{I}_p - \hat{\mathbf{S}})\mathbf{S}^{-1} \\ &\quad + (\eta\mathbf{I}_p - \hat{\mathbf{S}})\mathbf{S}^{-1}(\hat{\mathbf{S}} - \mathbf{S})\mathbf{S}^{-1}\}\}, \end{aligned}$$

where we have again used (26) in Appendix B. Setting to zero gives

$$\rho_0 \stackrel{\text{def}}{=} \rho_0(\eta_0) = \frac{E\{\text{tr}\{(\eta_0\mathbf{I}_p - \hat{\mathbf{S}})\mathbf{S}^{-1}(\mathbf{S} - \hat{\mathbf{S}})\mathbf{S}^{-1}\}\}}{E\{\text{tr}\{(\eta_0\mathbf{I}_p - \hat{\mathbf{S}})\mathbf{S}^{-1}(\eta_0\mathbf{I}_p - \hat{\mathbf{S}})\mathbf{S}^{-1}\}\}}. \quad (31)$$

Look first at the numerator, a say, of (31). Expanding gives

$$a = E\{\text{tr}\{\eta_0\mathbf{S}^{-1} - \eta_0\mathbf{S}^{-1}\hat{\mathbf{S}}\mathbf{S}^{-1} - \hat{\mathbf{S}}\mathbf{S}^{-1} + \hat{\mathbf{S}}\mathbf{S}^{-1}\hat{\mathbf{S}}\mathbf{S}^{-1}\}\}.$$

Again, we use that $K\hat{\mathbf{S}}$ has the complex Wishart distribution with mean $K\mathbf{S}$. Since $E\{\hat{\mathbf{S}}\} = \mathbf{S}$ the first two terms cancel. Then $E\{\text{tr}\{\hat{\mathbf{S}}\mathbf{S}^{-1}\}\} = \text{tr}\{\mathbf{I}_p\} = p$.

From (27) we know that $E\{\text{tr}\{\hat{\mathbf{S}}\mathbf{S}^{-1}\hat{\mathbf{S}}\mathbf{S}^{-1}\}\} = (p + (p^2/K))$. So $a = -p + (p + (p^2/K)) = p^2/K$.

Expanding the denominator, b say, of (31) gives

$$b = E\{\text{tr}\{\eta_0^2 \mathbf{S}^{-2} - \eta_0 \mathbf{S}^{-1} \hat{\mathbf{S}} \mathbf{S}^{-1} - \eta_0 \hat{\mathbf{S}} \mathbf{S}^{-2} + \hat{\mathbf{S}} \mathbf{S}^{-1} \hat{\mathbf{S}} \mathbf{S}^{-1}\}\}, \quad (32)$$

Then

$$\begin{aligned} b &= \frac{\text{tr}^2\{\mathbf{S}^{-1}\}}{\text{tr}^2\{\mathbf{S}^{-2}\}} \text{tr}\{\mathbf{S}^{-2}\} - 2 \frac{\text{tr}^2\{\mathbf{S}^{-1}\}}{\text{tr}\{\mathbf{S}^{-2}\}} + (p + (p^2/K)) \\ &= -\frac{\text{tr}^2\{\mathbf{S}^{-1}\}}{\text{tr}\{\mathbf{S}^{-2}\}} + (p + (p^2/K)). \end{aligned}$$

On tidying up the ratio a/b is of the form (14). That η_0, ρ_0 thus defined correspond to a minimum point is more easily shown via the proof of Lemma 3.

E. Proof of Lemma 3

Proceeding as before, and using (22), for the HS loss

$$\begin{aligned} (\partial/\partial\alpha)\mathcal{R}_{HS}(\hat{\mathbf{S}}^*(\alpha, \beta), \mathbf{S}) &= 2[\alpha C + \beta \text{tr}\{\mathbf{S}\} - \text{tr}\{\mathbf{S}^2\}] \\ (\partial/\partial\beta)\mathcal{R}_{HS}(\hat{\mathbf{S}}^*(\alpha, \beta), \mathbf{S}) &= 2[\alpha \text{tr}\{\mathbf{S}\} + \beta p - \text{tr}\{\mathbf{S}\}], \end{aligned}$$

where $C = \text{tr}\{\mathbf{S}^2\} + (1/K)\text{tr}^2\{\mathbf{S}\}$. Setting to zero and solving for β gives

$$\beta_0 = \frac{\text{tr}^3\{\mathbf{S}\}}{Kp \text{tr}\{\mathbf{S}^2\} + [p - K]\text{tr}^2\{\mathbf{S}\}} = \rho_0 \eta_0.$$

Similarly, for α ,

$$\alpha_0 = \frac{p \text{tr}\{\mathbf{S}^2\} - \text{tr}^2\{\mathbf{S}\}}{p[\text{tr}\{\mathbf{S}^2\} + \frac{1}{K}\text{tr}^2\{\mathbf{S}\}] - \text{tr}^2\{\mathbf{S}\}} = 1 - \rho_0.$$

For the determinant of the Hessian matrix we have

$$4 \begin{vmatrix} C & \text{tr}\{\mathbf{S}\} \\ \text{tr}\{\mathbf{S}\} & p \end{vmatrix}_{\alpha_0, \beta_0} = 4p \text{tr}\{\mathbf{S}^2\} - 4[1 - \frac{p}{K}]\text{tr}^2\{\mathbf{S}\}.$$

This is positive if $p \text{tr}\{\mathbf{S}^2\} > [1 - \frac{p}{K}]\text{tr}^2\{\mathbf{S}\}$, i.e., $p > [1 - \frac{p}{K}]\text{tr}^2\{\mathbf{S}\}/\text{tr}\{\mathbf{S}^2\}$, since $\text{tr}\{\mathbf{S}^2\} > 0$. But by Chebyshev's inequality we know that $p \geq \text{tr}^2\{\mathbf{S}\}/\text{tr}\{\mathbf{S}^2\}$, and so we know that $p > [1 - \frac{p}{K}]\text{tr}^2\{\mathbf{S}\}/\text{tr}\{\mathbf{S}^2\}$, as required. Furthermore,

$$(\partial^2/\partial\beta^2)\mathcal{R}_{HS}(\hat{\mathbf{S}}^*(\alpha, \beta), \mathbf{S}) \Big|_{\alpha_0, \beta_0} = 2p > 0,$$

and so we can conclude that the solution is a minimum. For the QL loss,

$$\begin{aligned} (\partial/\partial\alpha)\mathcal{R}_{QL}(\hat{\mathbf{S}}^*(\alpha, \beta), \mathbf{S}) &= 2[\alpha D + \beta \text{tr}\{\mathbf{S}^{-1}\} - p] \\ (\partial/\partial\beta)\mathcal{R}_{QL}(\hat{\mathbf{S}}^*(\alpha, \beta), \mathbf{S}) &= 2[\alpha \text{tr}\{\mathbf{S}^{-1}\} + \beta \text{tr}\{\mathbf{S}^{-2}\} \\ &\quad - 2\text{tr}\{\mathbf{S}^{-1}\}], \end{aligned}$$

where $D = p[1 + (p/K)]$. For the determinant of the Hessian matrix,

$$4 \begin{vmatrix} D & \text{tr}\{\mathbf{S}^{-1}\} \\ \text{tr}\{\mathbf{S}^{-1}\} & \text{tr}\{\mathbf{S}^{-2}\} \end{vmatrix}_{\alpha_0, \beta_0} = 4D \text{tr}\{\mathbf{S}^{-2}\} - 4\text{tr}^2\{\mathbf{S}^{-1}\}.$$

This is positive if $D > \text{tr}^2\{\mathbf{S}^{-1}\}/\text{tr}\{\mathbf{S}^{-2}\}$, since $\text{tr}\{\mathbf{S}^{-2}\} > 0$. We know from Chebyshev's inequality that $p \geq \text{tr}^2\{\mathbf{S}^{-1}\}/\text{tr}\{\mathbf{S}^{-2}\}$, so $p[1 + (p/K)] > \text{tr}^2\{\mathbf{S}^{-1}\}/\text{tr}\{\mathbf{S}^{-2}\}$. Also we note that

$$(\partial^2/\partial\alpha^2)\mathcal{R}_{QL}(\hat{\mathbf{S}}^*(\alpha, \beta), \mathbf{S}) \Big|_{\alpha_0, \beta_0} = 2D > 0,$$

and so we can conclude that the solution is a minimum.

F. Proof of Lemma 4

With $\hat{\mathbf{C}}^*(\alpha, \beta)$ defined in (15) and $\mathcal{R}_{HS}(\hat{\mathbf{C}}^*(\alpha, \beta), \mathbf{C})$ we take

$$\mathbf{F} = \hat{\mathbf{C}}^* \mathbf{S} - \mathbf{S}^{-1} = \alpha \hat{\mathbf{S}}^{-1} \mathbf{S} + \beta \mathbf{S} - \mathbf{S}^{-1}.$$

Using (23) and (24) we find that $(\partial/\partial\alpha)\mathcal{R}_{HS}(\hat{\mathbf{C}}^*(\alpha, \beta), \mathbf{C})$ is given by

$$\begin{aligned} &2 \left[\alpha c_1 [(K-p)\text{tr}\{\mathbf{S}^{-2}\} + \text{tr}^2\{\mathbf{S}^{-1}\}] \right. \\ &\quad \left. + \beta \frac{K}{K-p} \text{tr}\{\mathbf{S}^{-1}\} - \frac{K}{K-p} \text{tr}\{\mathbf{S}^{-2}\} \right]. \end{aligned} \quad (33)$$

Using (23) we find $(\partial/\partial\beta)\mathcal{R}_{HS}(\hat{\mathbf{C}}^*(\alpha, \beta), \mathbf{C})$ is given by

$$2 \left[\alpha \frac{K}{K-p} \text{tr}\{\mathbf{S}^{-1}\} + \beta p - \text{tr}\{\mathbf{S}^{-1}\} \right]. \quad (34)$$

Setting (33) and (34) to zero and solving the simultaneous equations gives α_0 and β_0 as stated in the lemma. The determinant of the Hessian (divided by 4) is

$$\delta = \begin{vmatrix} c_1 [(K-p)\text{tr}\{\mathbf{S}^{-2}\} + \text{tr}^2\{\mathbf{S}^{-1}\}] & \frac{K}{K-p} \text{tr}\{\mathbf{S}^{-1}\} \\ \frac{K}{K-p} \text{tr}\{\mathbf{S}^{-1}\} & p \end{vmatrix}_{\alpha_0, \beta_0},$$

which is

$$\delta = c_1 p [(K-p)\text{tr}\{\mathbf{S}^{-2}\} + \text{tr}^2\{\mathbf{S}^{-1}\}] - \left[\frac{K}{K-p} \right]^2 \text{tr}^2\{\mathbf{S}^{-1}\}.$$

Now $p \text{tr}\{\mathbf{S}^{-2}\} \geq \text{tr}^2\{\mathbf{S}^{-1}\}$, so

$$\begin{aligned} \delta &\geq c_1 (K-p) \text{tr}^2\{\mathbf{S}^{-1}\} + c_1 p \text{tr}^2\{\mathbf{S}^{-1}\} - \left[\frac{K}{K-p} \right]^2 \text{tr}^2\{\mathbf{S}^{-1}\} \\ &= \left(c_1 K - \left[\frac{K}{K-p} \right]^2 \right) \text{tr}^2\{\mathbf{S}^{-1}\} \\ &= \left(\frac{pK^2(K-p) + K^2}{(K-p)^4 - (K-p)^2} \right) \text{tr}^2\{\mathbf{S}^{-1}\} > 0, \end{aligned}$$

since $K > p+1$. Combining the positive determinant with the fact that $(\partial^2/\partial\beta^2)\mathcal{R}_{HS}(\hat{\mathbf{C}}^*(\alpha, \beta), \mathbf{C}) = 2p > 0$ we see that the turning point is indeed a minimum.

G. Proof of Lemma 5

With $\hat{\mathbf{C}}^*(\alpha, \beta)$ defined in (15) and $\mathcal{R}_{QL}(\hat{\mathbf{C}}^*(\alpha, \beta), \mathbf{C})$ we take

$$\mathbf{F} = \hat{\mathbf{C}}^* \mathbf{S} - \mathbf{I}_p = \alpha \hat{\mathbf{S}}^{-1} \mathbf{S} + \beta \mathbf{S} - \mathbf{I}_p.$$

Proceeding as before we find that $(\partial/\partial\alpha)\mathcal{R}_{QL}(\hat{\mathbf{C}}^*(\alpha, \beta), \mathbf{C})$ is given by

$$2[\alpha E\{\text{tr}\{\hat{\mathbf{S}}^{-1} \mathbf{S} \hat{\mathbf{S}}^{-1} \mathbf{S}\}\} + \beta E\{\text{tr}\{\hat{\mathbf{S}}^{-1} \mathbf{S}^2\}\} - E\{\text{tr}\{\hat{\mathbf{S}}^{-1} \mathbf{S}\}\}].$$

Setting $\mathbf{A} = \mathbf{B} = \mathbf{S}$ in (24),

$$E\{\text{tr}\{\hat{\mathbf{S}}^{-1} \mathbf{S} \hat{\mathbf{S}}^{-1} \mathbf{S}\}\} = \frac{K}{K-p} \frac{pK^2}{(K-p)^2 - 1} = \frac{K}{K-p} c,$$

where $c = pK^2/[(K-p)^2 - 1]$. Next we use the result that $E\{\hat{\mathbf{S}}^{-1}\} = (K/(K-p))\mathbf{S}^{-1}$, which means that

$$E\{\text{tr}\{\hat{\mathbf{S}}^{-1} \mathbf{S}^2\}\} = \frac{K}{K-p} \text{tr}\{\mathbf{S}\} \text{ and } E\{\text{tr}\{\hat{\mathbf{S}}^{-1} \mathbf{S}\}\} = \frac{Kp}{K-p}.$$

Putting results together we see that

$$\frac{\partial}{\partial\alpha} \mathcal{R}_{QL}(\hat{\mathbf{C}}^*(\alpha, \beta), \mathbf{C}) = 2 \frac{K}{K-p} [\alpha c + \beta \text{tr}\{\mathbf{S}\} - p].$$

Next, $(\partial/\partial\beta)\mathcal{R}_{QL}(\hat{\mathbf{C}}^*(\alpha, \beta), \mathbf{C})$ is given by

$$\begin{aligned} & 2[\alpha E\{\text{tr}\{\mathbf{S}\hat{\mathbf{S}}^{-1}\}\} + \beta E\{\text{tr}\{\mathbf{S}^2\}\} - E\{\text{tr}\{\mathbf{S}\}\}] \\ &= 2[\alpha \frac{K}{K-p} \text{tr}\{\mathbf{S}\} + \beta \text{tr}\{\mathbf{S}^2\} - \text{tr}\{\mathbf{S}\}]. \end{aligned}$$

Setting both partial derivatives to zero and removing constant multipliers gives the simultaneous equations

$$\alpha c + \beta \text{tr}\{\mathbf{S}\} - p = 0 \quad (35)$$

$$\alpha \frac{K}{K-p} \text{tr}\{\mathbf{S}\} + \beta \text{tr}\{\mathbf{S}^2\} - \text{tr}\{\mathbf{S}\} = 0 \quad (36)$$

Multiply (35) by $K\text{tr}\{\mathbf{S}\}/(K-p)$ and (36) by c . Subtracting, with $c = c_0p$, gives β_0 in (18) and (19). Multiply (35) by $\text{tr}\{\mathbf{S}^2\}$ and (36) by $\text{tr}\{\mathbf{S}\}$. Subtracting, gives α_0 in (18) and (19).

The determinant of the Hessian (divided by 4) is

$$\begin{vmatrix} \frac{K}{K-p}c & \frac{K}{K-p}\text{tr}\{\mathbf{S}\} \\ \frac{K}{K-p}\text{tr}\{\mathbf{S}\} & \text{tr}\{\mathbf{S}^2\} \end{vmatrix}_{\alpha_0, \beta_0} = \frac{K}{K-p} [c\text{tr}\{\mathbf{S}^2\} - \frac{K}{K-p}\text{tr}^2\{\mathbf{S}\}].$$

The term in the square bracket on the right-side can be rewritten as

$$\frac{pK^2}{(K-p)^2 - 1} \text{tr}\{\mathbf{S}^2\} - \frac{K}{K-p} \text{tr}^2\{\mathbf{S}\} \quad (37)$$

We know that $p\text{tr}\{\mathbf{S}^2\} \geq \text{tr}^2\{\mathbf{S}\}$, so (37) will be positive if $K^2/[(K-p)^2 - 1] - K/(K-p) > 0$. This term can be written as $[Kp(K-p) + K]/[(K-p)^3 - (K-p)] > 0$ since we are assuming $K > p+1$. Combining the positive determinant with the fact that $(\partial^2/\partial\beta^2)\mathcal{R}_{QL}(\hat{\mathbf{C}}^*(\alpha, \beta), \mathbf{C}) = 2\text{tr}\{\mathbf{S}^2\} > 0$ we see that the turning point is indeed a minimum.

ACKNOWLEDGMENT

The work of Deborah Schneider-Luftman was supported by EPSRC (UK).

REFERENCES

- [1] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 24, pp. 397–401, 1988.
- [2] G. Casella and R. L. Berger, *Statistical Inference*. Belmont, CA: Duxbury, 1990.
- [3] S. Chandna and A. T. Walden, "Statistical properties of the estimator of the rotary coefficient," *IEEE Trans. Signal Process.*, vol. 59, pp. 1298–1303, 2011.
- [4] S. Chandna and A. T. Walden, "Simulation methodology for inference on physical parameters of complex vector-valued signals," *IEEE Trans. Signal Process.*, vol. 60, pp. 5260–5269, 2013.
- [5] X. Chen, Z. Jane Wang, and M. J. McKeown, "Shrinkage-to-tapering estimation of large covariance matrices," *IEEE Trans. Signal Process.*, vol. 60, pp. 5640–5656, 2012.
- [6] R. Dahlhaus, "Graphical interaction models for multivariate time series," *Metrika*, vol. 51, pp. 157–172, 2000.
- [7] M. J. Daniels and R. E. Kass, "Shrinkage estimators for covariance matrices," *Biometrics*, vol. 57, pp. 1173–1184, 2001.
- [8] B. Efron and C. Morris, "Multivariate empirical Bayes and estimation of covariance matrices," *The Annals of Statistics*, vol. 4, pp. 22–32, 1976.
- [9] T. J. Fisher and X. Sun, "Improved Stein-type shrinkage estimators for the high-dimensional normal covariance matrix," *Computational Statistics and Data Analysis*, vol. 55, pp. 1909–18, 2011.
- [10] M. Fiecas, H. Ombao, C. Linkletter, W. Thompson and J. Sanes, "Functional connectivity: shrinkage estimation and randomization test," *NeuroImage*, vol. 49, pp. 3005–3014, 2010.
- [11] G. J. Fraser, "The 5-day wave and ionospheric absorption," *J. Atmospheric and Terrestrial Physics*, vol. 39, pp. 121–124, 1977.
- [12] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, pp. 152–77, 1963.
- [13] L. R. Haff, "Estimation of the inverse covariance matrix: random mixtures of the inverse Wishart matrix and the identity," *The Annals of Statistics*, vol. 7, pp. 1264–1276, 1979.
- [14] L. R. Haff, "Empirical Bayes estimation of the multivariate normal covariance matrix," *The Annals of Statistics*, vol. 8, pp. 586–597, 1980.
- [15] D. A. Harville, *Matrix Algebra From A Statistician's Perspective*. New York: Springer-Verlag, 1997.
- [16] W. James and C. Stein, "Estimation with quadratic loss," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, Berkeley, CA: University of California Press, pp. 361–379, 1961.
- [17] Y. Konno, "Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss," *Journal of Multivariate Analysis*, vol. 100, pp. 2237–2253, 2009.
- [18] V. G. Koutitonsky, N. Navarro and D. Booth, "Descriptive physical oceanography of Great-Entry Lagoon, Gulf of St. Lawrence," *Estuarine, Coastal and Shelf Science*, vol. 54, pp. 833–847, 2002.
- [19] Larsen, P. D., Lewis, C. D., Gebber, G. L. & Zhong, S., "Partial spectral analysis of cardiac-related sympathetic nerve discharge," *J. Neurophysiol.*, vol. 84, pp. 1168–79, 2000.
- [20] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis* vol. 88, pp. 365–411, 2004.
- [21] G. Letac and H. Massam, "All invariant moments of the Wishart distribution," *Scand. J. of Stat.*, vol. 31, pp. 295–318, 2004.
- [22] D. Maiwald and D. Kraus, "Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices," *IEEE Proceedings Radar, Sonar Navigation*, vol. 147, pp. 162–168, 2000.
- [23] T. Medkour, A. T. Walden and A. Burgess, "Graphical modelling for brain connectivity via partial coherence," *J. Neurosci. Meth.* vol. 180, pp. 374–383, 2009.
- [24] T. Medkour, A. T. Walden, A. P. Burgess & V. B. Strelets, "Brain connectivity in positive and negative syndrome schizophrenia," *Neuroscience*, vol. 169, pp. 1779–88, 2010.
- [25] T. Mima, T. Matsuoka and M. Hallett, "Functional coupling of right and left cortical motor areas demonstrated with partial coherence analysis," *Neurosci. Lett.*, vol. 287, pp. 93–6, 2000.
- [26] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications*. Cambridge, UK: Cambridge University Press, 1993.
- [27] K. S. Riedel and A. Sidorenko, "Minimum bias multiple taper spectral estimation," *IEEE Trans. Signal Process.* vol. 43, pp. 188–195.
- [28] R. Salvador, J. Suckling, C. Schwarzbauer and E. Bullmore, "Undirected graphs of frequency-dependent functional connectivity in whole brain networks," *Phil. Trans. Roy. Soc. Lond., Ser. B*, vol. 360, pp. 937–46, 2005.
- [29] J. Schäfer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statist. Appl. Genet. Molec. Biol.*, vol. 4, no. 1, 2005.
- [30] C. Stein, "Estimation of a covariance matrix." Rietz lecture, 39th Annual Meeting IMS. Atlanta, GA, 1975.
- [31] A. T. Walden, E. J. McCoy and D. B. Percival, "The effective bandwidth of a multitaper spectral estimator," *Biometrika*, vol. 82, 201–214, 1995.
- [32] A. T. Walden and D. Schneider-Luftman, "Random matrix derived shrinkage of spectral precision matrices," *IEEE Trans. Signal Process.*, vol. 63, pp. 4689–4699, 2015.
- [33] C. Wang, G. Pan, T. Tong and L. Zhu, "Shrinkage estimation of large dimensional precision matrix using random matrix theory," *Statistica Sinica*, vol. 25, pp. 993–1008, 2015.
- [34] R. E. White, "Signal and noise estimation from seismic reflection data using spectral coherence methods," *Proc. IEEE*, vol. 72, pp. 1340–1356, 1984.
- [35] G. A. Young and R. L. Smith, *Essentials of Statistical Inference*. Cambridge UK: Cambridge University Press, 2005.