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 far 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 appealing. 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 $\{\boldsymbol{X}_t\}$ where $\boldsymbol{X}_t = [X_{1,t}, \ldots, X_{p,t}]^T \in \mathbb{R}^p$, $t \in \mathbb{Z}$, and ^{T} denotes transposition. Without loss of generality we assume $\{\boldsymbol{X}_t\}$ to have a zero mean. Denote the sample interval by Δ_{t} and the Nyquist frequency by $f^\star = 1/(2\Delta_{\mathrm{t}})$. One very important quantity derived from vector-valued time series is the partial coherence between different series. Let $\boldsymbol{S}(f)$ denote the spectral matrix of $\{\boldsymbol{X}_t\}$ at frequency f, assumed to exist and be of full rank. With $\boldsymbol{s}_\tau \stackrel{\mathrm{def}}{=} \mathrm{cov}\{\boldsymbol{X}_{t+\tau}, \boldsymbol{X}_t\} = E\{\boldsymbol{X}_{t+\tau}\boldsymbol{X}_t^T\}$, we have $\boldsymbol{S}(f) \stackrel{\mathrm{def}}{=} \Delta_{\mathrm{t}} \sum_{\tau=-\infty}^{\infty} \boldsymbol{s}_\tau \, \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}}$. Denote the $(j,k)^{\mathrm{th}}$ element of the precision matrix $\boldsymbol{C}(f) \stackrel{\mathrm{def}}{=} \boldsymbol{S}^{-1}(f)$ by $C_{jk}(f)$. The partial coherence between series j and k can be expressed as, (e.g., [6]), $\gamma^2_{jk\bullet\{\backslash jk\}}(f) = |C_{jk}(f)|^2/[C_{jj}(f)C_{kk}(f)]$, 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 $\{\backslash jk\}$. 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 $\boldsymbol{S}(f)$ as $\hat{\boldsymbol{S}}(f)$, and invert it, or we can estimate $\boldsymbol{C}(f)$ directly.

Consider an estimator $\hat{\boldsymbol{S}}(f)$, of $\boldsymbol{S}(f)$. An estimator may be computed by a multitaper scheme involving K tapers (e.g., [26]). Throughout this paper we assume the choice of the number of tapers is such that $K > p$ 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, rather than a large integer multiple of it. 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 but consistent with $K > p$; 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.) So this paper is concerned with the ill-conditioned but non-singular case $K > p, K \simeq p$ with p, K finite.

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 $\hat{\boldsymbol{S}}(f)$ 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 $\hat{\boldsymbol{S}}(f)$ 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,

Deborah Schneider-Luftman and Andrew Walden are both at the Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7 2BZ, UK. (e-mail: deborah.schneider-luftman11@imperial.ac.uk and a.walden@imperial.ac.uk)

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. We examine EEG data for two individuals, one having high partial coherence over some frequencies and the other consistently low partial coherence over all frequencies. These individuals therefore act as practically realistic proxies for cases of high and low partial coherence, and demonstrate the general behaviour of the various estimators for the two cases. 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, QLa and QLb, for $\hat{\boldsymbol{S}}(f)$ under QL loss in closed form. One involves just a single shrinkage parameter, ρ, while the other has two parameters, a shrinkage parameter ρ and a scaling parameter η.
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
$$\hat{\gamma}^2_{jk\bullet\{\backslash jk\}}(f) = |\hat{C}_{jk}(f)|^2/[\hat{C}_{jj}(f)\hat{C}_{kk}(f)], \quad (1)$$
where $\hat{\boldsymbol{C}}(f) \stackrel{\text{def}}{=} \hat{\boldsymbol{S}}^{-1}(f)$ and $\hat{\boldsymbol{S}}(f)$ is the multitaper estimator of $\boldsymbol{S}(f)$. 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 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, QLa, QLb 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 QLa-est, QLb-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 seems a good choice for a practical 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 QLa, QLb 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. Multitaper Estimator

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

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

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

$$\hat{\boldsymbol{S}}(f) \;\; = \;\; \frac{1}{K}\sum_{k=0}^{K-1} \hat{\boldsymbol{S}}_k(f) = \frac{1}{K}\boldsymbol{J}(f)\boldsymbol{J}^H(f). \quad (2)$$

B. Statistical Properties

Let B denote the bandwidth of the spectral window corresponding to the tapering, which for sine tapers is given by $B = (K+1)/[(N+1)\Delta_{\text{t}}]$, (e.g., [31]).

We have the following statistical properties [3]. Suppose $\{\boldsymbol{X}_t\}$ is Gaussian and consider the practical case of finite length N for the time series segment. Then $\boldsymbol{J}_k(f), k = 0, \ldots, K-1$, may be taken to be independently and identically distributed as p-vector-valued complex Gaussian with mean zero and covariance matrix $\boldsymbol{S}(f)$:

$$\boldsymbol{J}_k(f) \stackrel{\text{d}}{=} \mathcal{N}_p^C\{\boldsymbol{0}, \boldsymbol{S}(f)\} \text{ for } B/2 < |f| < f^\star - B/2. \quad (3)$$

The choice of K clearly affects the bandwidth of the spectral window. B is assumed to have been chosen narrow enough to ensure the components of $\boldsymbol{S}(f)$ are essentially constant across it. $\hat{\boldsymbol{S}}(f)$ in (2) is the maximum-likelihood estimator for $\boldsymbol{S}(f)$, [12].

Given (3), and with $K \geq p$, $K\hat{\boldsymbol{S}}(f)$ has the complex Wishart distribution with mean $K\boldsymbol{S}(f)$, written as

$$K\hat{\boldsymbol{S}}(f) \stackrel{\text{d}}{=} \mathcal{W}_p^C\{K, \boldsymbol{S}(f)\}. \quad (4)$$

Therefore,

$$E\{\hat{\boldsymbol{S}}(f)\} = \boldsymbol{S}(f) \quad \text{and} \quad E\{\text{tr}\{\hat{\boldsymbol{S}}(f)\}\} = \text{tr}\{\boldsymbol{S}(f)\}, \quad (5)$$

two simple results which will be made use of. Other properties following from (4) 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 0.5–45 Hz and sample interval of $\Delta_{\text{t}} = 0.01$s. The recorded data $\{\boldsymbol{X}_t\}$ is thus a $p = 10$ vector-valued process, with Nyquist frequency 50 Hz. To fully remove any influence of the highly dominant and contaminating 10 Hz alpha rhythm, which would otherwise cause severe spectral leakage, a 4.6 Hz low-pass filter was applied, followed by resampling to a sample interval of $\Delta_{\text{t}} = 0.05$s giving a new Nyquist frequency of $f^\star = 10$ 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.

Common sources of 'noise' which could perturb partial coherence estimators are spectral line sources such as the alpha rhythm, mentioned above, or power line pick-up from the operating frequency [23, p. 377], although this latter would only be an issue at higher frequencies (50 Hz in the UK, 60 Hz in the USA).

Generally, additive noise only increases the diagonal of the spectral matrix. Interestingly low-power additive noise is typically overwhelmed by the shrinkage procedure, which has an estimator that utilizes diagonal loading of the spectral matrix.

B. Simulation Strategy

From this 10-channel series the spectral matrix $\boldsymbol{S}(f)$ was estimated as $\boldsymbol{S}_0(f)$, say, for $|f| \leq f^\star$. Using the vector-valued circulant embedding approach, [4], a large number, M say, of independent Gaussian 10-channel time series were computed, each having $\boldsymbol{S}_0(f)$, $|f| \leq f^\star$, as its true spectral matrix. For each of these time series the estimated spectral matrix $\hat{\boldsymbol{S}}(f)$ was computed using multitaper estimation with some specified number K of sine tapers. These M independent estimates of $\boldsymbol{S}_0(f)$ can then be used to deduce various sampling properties for quantities derived from $\hat{\boldsymbol{S}}(f)$. 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 $\boldsymbol{S}_0(f)$. 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

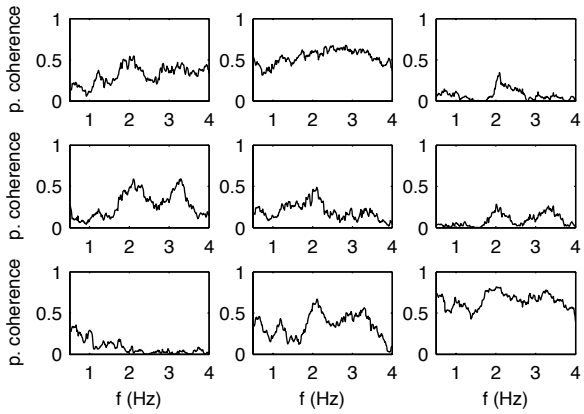

Fig. 1. The first 9 partial coherency plots ($j = 1, k = 2 : 10$) for individual-1, found from the known matrix $\boldsymbol{S}_0(f)$.

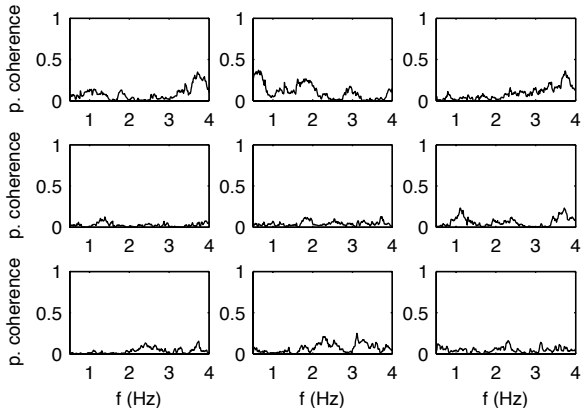

Fig. 2. The first 9 partial coherency plots ($j = 1, k = 2 : 10$) for individual-2, found from the known matrix $\boldsymbol{S}_0(f)$.

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 $\hat{\boldsymbol{S}}(f)$ this would take the form $\boldsymbol{S}^\star(\rho(f)) = (1-\rho(f))\hat{\boldsymbol{S}}(f) + \rho(f)\boldsymbol{T}(f)$, where $\rho(f) \in (0, 1)$ is known as the shrinkage parameter and $\boldsymbol{T}(f)$ is the target matrix. Provided $\hat{\boldsymbol{S}}(f)$ and $\boldsymbol{T}(f)$ 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.

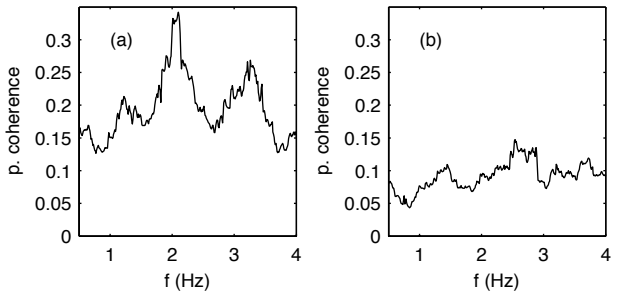

Fig. 3. Averaged true partial coherencies for (a) individual-1, and (b) individual-2, (averaging over all pairs $(j,k), 1 \le j < k \le p$).

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

$$\rho_0 = \underset{\rho\in(0,1)}{\arg\min}\, \mathcal{R}_{\mathcal{L}}(\hat{\boldsymbol{S}}^\star(\rho), \boldsymbol{S}). \tag{6}$$

B. Hilbert-Schmidt loss

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

$$\begin{aligned}\mathcal{R}_{HS}(\hat{\boldsymbol{S}}^\star(\rho), \boldsymbol{S}) &\stackrel{\text{def}}{=} E\{\text{tr}\{(\hat{\boldsymbol{S}}^\star(\rho) - \boldsymbol{S})^2\}\} \\ &= E\{||\boldsymbol{S}^\star(\rho) - \boldsymbol{S}||_{\text{F}}^2\},\end{aligned} \tag{7}$$

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

Using (6) and (7), the target matrix was chosen to be of the form $\boldsymbol{T} = (\text{tr}\{\boldsymbol{S}\}/p)\boldsymbol{I}_p$ in [20] (for real-valued covariance matrices) so that

$$\boldsymbol{S}^\star(\rho) = (1-\rho)\hat{\boldsymbol{S}} + \rho\frac{\text{tr}\{\boldsymbol{S}\}}{p}\boldsymbol{I}_p. \tag{8}$$

For the ill-conditioned case ($p < K, p \simeq K$) the estimator diagonally loads the initial matrix $\hat{\boldsymbol{S}}$ and increases the zero or near-zero eigenvalues. For this choice the solution to (6) and (7) is (e.g. [32, eq. 9]),

$$\rho_0 = \left[1 - \frac{K}{p} + K\frac{\text{tr}\{\boldsymbol{S}^2\}}{\text{tr}^2\{\boldsymbol{S}\}}\right]^{-1}. \tag{9}$$

C. Quadratic Loss

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

$$\mathcal{R}_{QL}(\hat{\boldsymbol{S}}^\star(\rho), \boldsymbol{S}) = E\{\text{tr}\{(\hat{\boldsymbol{S}}^\star(\rho)\boldsymbol{S}^{-1} - \boldsymbol{I}_p)^2\}\}. \tag{10}$$

Lemma 1: With $\boldsymbol{S}^\star(\rho)$ defined as in (8) the form of ρ_0 solving (6) and (10) is

$$\frac{p^2}{(Kp+p^2) - \frac{2}{p^2}\left[Kp\,\text{tr}\{\boldsymbol{S}^{-1}\}\text{tr}\{\boldsymbol{S}\} - \frac{K}{2}\text{tr}^2\{\boldsymbol{S}\}\text{tr}\{\boldsymbol{S}^{-2}\}\right]}. \tag{11}$$

Proof: This is given in Appendix C. ■

For easy identification we shall call this the QLa method.

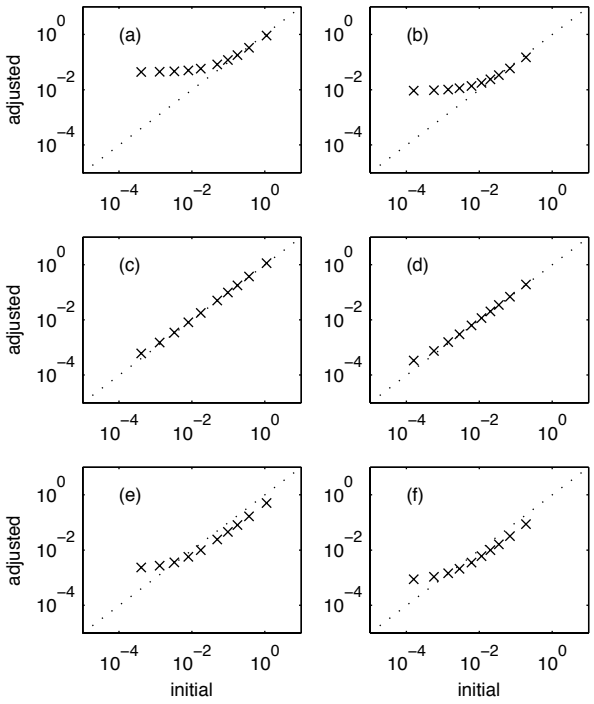

Fig. 4. Average eigenvalue adjustment for frequencies $f = 0.85$ Hz (left column) and 3.85 Hz (right column). The first row is for the HS method, the second for QLa and the third for QLb. $K = 12$ here for individual-1.

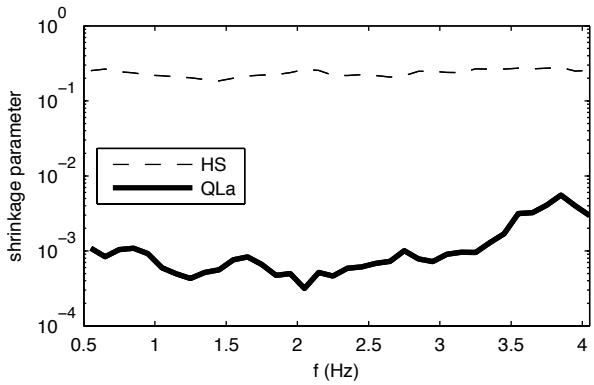

Fig. 5. Oracle value of $\rho(f)$ for HS (dashed line) and QLa (thick line) for frequencies 0.55 to 4.05 Hz in steps of 0.1 for undividual-1.

D. Generalization

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

$$\hat{\boldsymbol{S}}^\star(\eta,\rho) = (1-\rho)\hat{\boldsymbol{S}} + \rho\,\eta\,\boldsymbol{I}_p, \quad \eta > 0, \rho \in (0,1) \tag{12}$$

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

$$(\eta_0, \rho_0) = \underset{\eta>0,\,\rho\in(0,1)}{\arg\min}\, \mathcal{R}_{\mathcal{L}}(\hat{\boldsymbol{S}}^\star(\eta,\rho), \boldsymbol{S}), \tag{13}$$

for any convex loss criterion $\mathcal{L}$.

Lemma 2: The solutions of (13) for both the HS and QL

losses are

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

$$(\eta_0, \rho_0)_{\text{QL}} = \left(\frac{\text{tr}\{\boldsymbol{S}^{-1}\}}{\text{tr}\{\boldsymbol{S}^{-2}\}}, \left[1 + \frac{K}{p} - \frac{K \text{tr}^2\{\boldsymbol{S}^{-1}\}}{p^2 \text{tr}\{\boldsymbol{S}^{-2}\}} \right]^{-1} \right). \quad (15)$$

Proof: This is given in Appendix D. ∎

We shall call the method defined by (14) the HS method and we call the method defined by (15) the QLb method.

Remark 1: We see that the choice of target $\boldsymbol{T} = (\text{tr}\{\boldsymbol{S}\}/p)\boldsymbol{I}_p$ is identical to η_0 under the HS loss, i.e., the scaling factor $(\text{tr}\{\boldsymbol{S}\}/p)$ is an optimal choice under the HS loss for shrinkage model (12). Since the targets are the same, the form of ρ_0 must also be the same and of the form (9).

In deriving the optimal forms of η_0 and ρ_0 (AppendixD), 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 (12) to $\hat{\boldsymbol{S}}^\star(\alpha, \beta) = \alpha \hat{\boldsymbol{S}} + \beta \boldsymbol{I}_p, \quad \alpha, \beta > 0$, it follows that (α_0, β_0) defined by $(\alpha_0, \beta_0) = \arg\min_{\alpha,\beta>0} \mathcal{R}_{\mathcal{L}}(\hat{\boldsymbol{S}}^\star(\alpha, \beta), \boldsymbol{S})$, 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. ∎

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 $\hat{\boldsymbol{S}}$. Fig. 4 shows the transformation of the $p = 10$ eigenvalues for HS, QLa and QLb at frequencies 0.8 Hz and 3.85 Hz. Both the initial and adjusted eigenvalues shown are averages, following sorting, over the $M = 500$ simulations of $\hat{\boldsymbol{S}}$. 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.

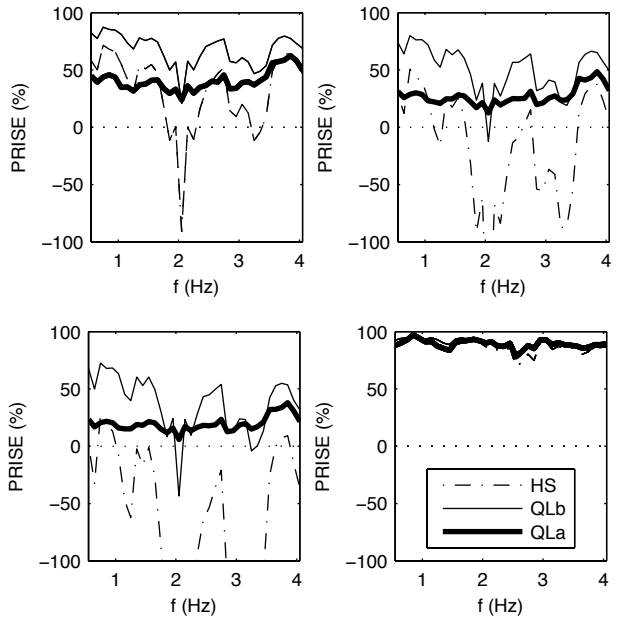

Fig. 6. Oracle PRISE with partial coherencies derived from spectral matrices. HS (dash-dot line), QLa (thick line) and QLb (thin line) for $K = 12$ (top left), $K = 14$ (top right), $K = 16$ (bottom left), all for individual-1, and $K = 12$ (bottom right) for individual-2.

C. Partial coherence

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

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

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 estimate (1), 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 (1). Results are shown in Fig. 6 with results at frequencies 0.55 to 4.05 Hz in steps of 0.1. These are oracle results: the known quantities $\text{tr}\{\boldsymbol{S}_0(f)\}, \text{tr}\{\boldsymbol{S}_0^2(f)\}, \text{tr}\{\boldsymbol{S}_0^{-1}(f)\}, \text{tr}\{\boldsymbol{S}_0^{-2}(f)\}$ have been used in the various estimators.

We also found the average and standard error of the PRISEs over frequencies to obtain Table I; the involved parameters are included in the second column of the table. Looking at the top three rows for individual-1, we see that the HS method does very poorly, sometimes worse than using the raw estimates.

method	params.	individual-1			individual-2
		$K=12$	$K=14$	$K=16$	$K=12$
HS	ρ	31 (33)	-17 (54)	-66 (75)	87 (6)
QLa	ρ	41 (9)	27 (8)	20 (7)	89 (4)
QLb	ρ,η	68 (14)	51 (20)	37 (25)	91 (3)
HSP	α,β	27 (30)	7 (33)	-7 (36)	82 (7)
QLP	α,β	52 (9)	32 (9)	23 (8)	85 (6)
QLa-est	$\hat{\rho}$	26 (7)	18 (6)	13 (5)	66 (4)
QLb-est	$\hat{\rho},\hat{\eta}$	65 (14)	49 (21)	34 (28)	87 (3)
QLP-est	$\hat{\alpha},\hat{\beta}$	52 (9)	34 (9)	25 (8)	84 (6)

TABLE I
PRISE(%) TO NEAREST INTEGER. AVERAGE AND STANDARD ERROR (IN BRACKETS) OVER FREQUENCIES. GAUSSIAN MODEL.

QLb generally does well but is very variable and rather unpredictable with frequency (Fig. 6). 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 individual-2, all methods do well.

VI. SHRINKING THE PRECISION MATRIX

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

$$\hat{\boldsymbol{C}}^{\star}(\alpha,\beta) = \alpha\,\hat{\boldsymbol{S}}^{-1} + \beta\,\boldsymbol{I}_p, \quad \alpha,\beta > 0. \tag{16}$$

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

$$\alpha\,\hat{\boldsymbol{S}}^{-1} + \beta/\mathrm{tr}\{\hat{\boldsymbol{S}}\}\boldsymbol{I}_p. \tag{17}$$

Lemma 4: Under the model (16), the HS risk $\mathcal{R}_{HS}(\hat{\boldsymbol{C}}^{\star}(\alpha,\beta),\boldsymbol{C}) = E\{\mathrm{tr}\{(\hat{\boldsymbol{C}}^{\star}(\alpha,\beta) - \boldsymbol{S}^{-1})^2\}\}$ is minimized, for $K > p+1$, by

$$\begin{aligned} \alpha_0 &= \frac{1}{D}\left[p\frac{\mathrm{tr}\{\boldsymbol{S}^{-2}\}}{\mathrm{tr}^2\{\boldsymbol{S}^{-1}\}} - 1\right] \\ \beta_0 &= \frac{c_3\,\mathrm{tr}\{\boldsymbol{S}^{-1}\}}{D}\left[\frac{\mathrm{tr}\{\boldsymbol{S}^{-2}\}}{\mathrm{tr}^2\{\boldsymbol{S}^{-1}\}} + (K-p)\right], \end{aligned} \tag{18}$$

where

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

Proof: See Appendix F. ∎

We shall call this the HSP method.

Lemma 5: Under the model (16), the QL risk $\mathcal{R}_{QL}(\hat{\boldsymbol{C}}^{\star}(\alpha,\beta),\boldsymbol{C}) = E\{\mathrm{tr}\{(\hat{\boldsymbol{C}}^{\star}(\alpha,\beta)\boldsymbol{S} - \boldsymbol{I}_p)^2\}\}$ is minimized, for $K > p+1$, by

$$\alpha_0 = \frac{1}{D}\left[p\frac{\mathrm{tr}\{\boldsymbol{S}^2\}}{\mathrm{tr}^2\{\boldsymbol{S}\}} - 1\right];\ \beta_0 = \frac{p}{D\,\mathrm{tr}\{\boldsymbol{S}\}}\left[c_0 - \frac{K}{K-p}\right] \tag{19}$$

where

$$c_0 = \frac{K^2}{(K-p)^2 - 1};\ D = \left[c_0 p\frac{\mathrm{tr}\{\boldsymbol{S}^2\}}{\mathrm{tr}^2\{\boldsymbol{S}\}} - \frac{K}{K-p}\right]. \tag{20}$$

Proof: See Appendix G. ∎

We shall call this the QLP method.

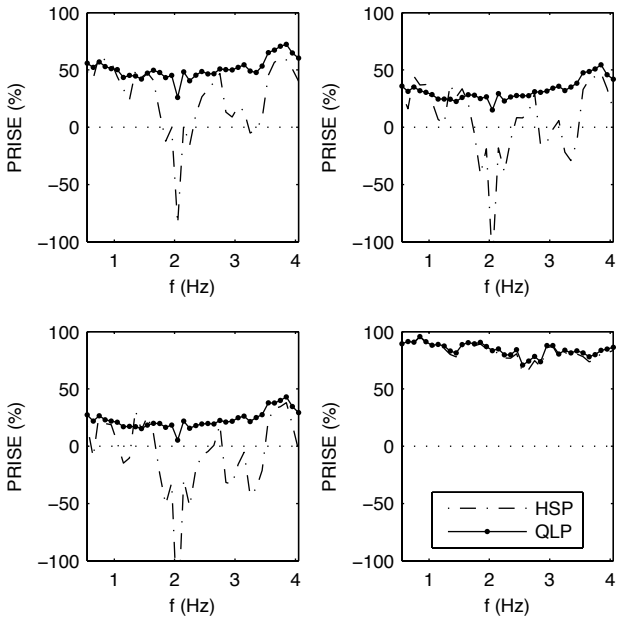

Fig. 7. Oracle PRISE with partial coherencies derived from precision matrices. HSP (dash-dot line) and QLP (line with bold dots) for $K = 12$ (top left), $K = 14$ (top right), $K = 16$ (bottom left), all for individual-1, and $K = 12$ (bottom right) for individual-2.

Remark 2: Under HS risk the optimal parameters α_0 and β_0 for shrinking $\hat{\boldsymbol{S}}^{-1}$ involve $\boldsymbol{S}^{-1}$ and $\boldsymbol{S}^{-2}$, which can be difficult to estimate in applications, whereas under QL risk, they involve $\boldsymbol{S}$ and $\boldsymbol{S}^2$, which are more stably estimated.

Remark 3: The numerator of α_0 in (18) is essentially the U-statistic for testing for sphericity of $\boldsymbol{C}$ and in (19) for testing for sphericity of $\boldsymbol{S}$.

Repeating the simulations in Section V-C using the oracle estimates $\hat{\boldsymbol{C}}^{\star}(\alpha_0,\beta_0)$ for the precision matrices, and hence the partial coherencies, gave the results in Fig. 7 and lines 4 and 5 of Table I. For individual-1 we see that HSP and HS are comparable for $K = 12$ but for $K = 14$ and 16, HSP does better. QLP performs much better than HSP, in terms of both mean and standard error. For individual-2 HSP is comparable to HS when $K = 12$.

VII. FULL (NON-ORACLE) ESTIMATION

A. Estimators

Each of the estimators we have derived involve the traces of some subset of $\boldsymbol{S}, \boldsymbol{S}^2, \boldsymbol{S}^{-1}, \boldsymbol{S}^{-2}$. 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.

We recall that $\hat{\boldsymbol{S}}$ is the maximum likelihood estimator (MLE) for $\boldsymbol{S}$. By the invariance property of MLEs [2, p. 294] it follows that the MLE of $\mathrm{tr}\{\boldsymbol{S}\}$ is $\mathrm{tr}\{\hat{\boldsymbol{S}}\}$, of $\mathrm{tr}\{\boldsymbol{S}^2\}$ is $\mathrm{tr}\{\hat{\boldsymbol{S}}^2\}$, of $\mathrm{tr}\{\boldsymbol{S}^{-1}\}$ is $\mathrm{tr}\{\hat{\boldsymbol{S}}^{-1}\}$ and of $\mathrm{tr}\{\boldsymbol{S}^{-2}\}$ is $\mathrm{tr}\{\hat{\boldsymbol{S}}^{-2}\}$. The MLE is a strongly consistent estimator [35, p. 129].

- To estimate $\mathrm{tr}\{\boldsymbol{S}\}$ we use $\mathrm{tr}\{\hat{\boldsymbol{S}}\}$ which by (5) is exactly unbiased.

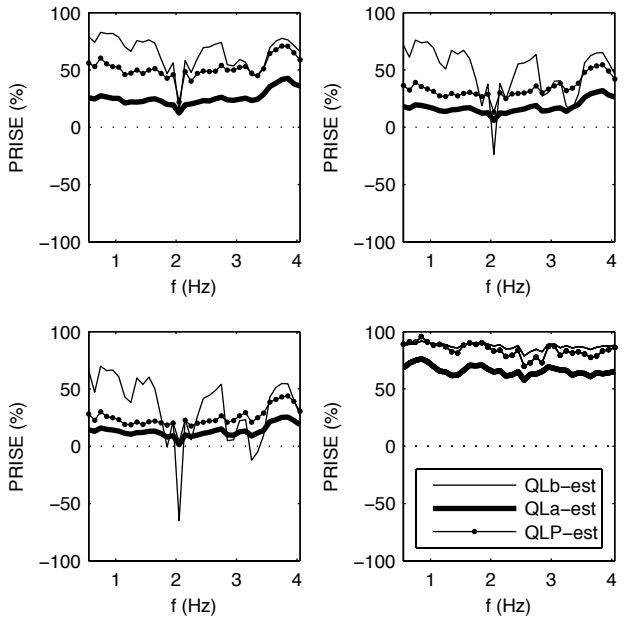

Fig. 8. PRISE with partial coherencies derived from fully estimated spectral or precision matrices. QLa-est (thick line), QLb-est (thin line) and QLP-est (line with bold dots) for $K = 12$ (top left), $K = 14$ (top right), $K = 16$ (bottom left), all for individual-1, and $K = 12$ (bottom right) for individual-2.

- Our estimator for $\text{tr}\{\boldsymbol{S}^2\}$ is $\text{tr}\{\hat{\boldsymbol{S}}^2\} - (1/K)\text{tr}^2\{\hat{\boldsymbol{S}}\}$. Using (22) and (23) in Appendix A we see that
$$E\left\{\text{tr}\{\hat{\boldsymbol{S}}^2\} - \frac{1}{K}\text{tr}^2\{\hat{\boldsymbol{S}}\}\right\} = \left[1 - \frac{1}{K^2}\right]\text{tr}\{\boldsymbol{S}^2\},$$
so the estimator is asymptotically unbiased for $\text{tr}\{\boldsymbol{S}^2\}$.
- To estimate $\text{tr}\{\boldsymbol{S}^{-1}\}$ we use $\left[1 - \frac{p}{K}\right]\text{tr}\{\hat{\boldsymbol{S}}^{-1}\}$. Result (24) in Appendix A shows this estimator is exactly unbiased (under the assumption $K > p$).
- Results (25) and (26) of Appendix A together show that
$$\left[1 - \frac{p}{K}\right]^2 \text{tr}\{\hat{\boldsymbol{S}}^{-2}\} - \frac{1}{K}\left[1 - \frac{p}{K}\right]\text{tr}^2\{\hat{\boldsymbol{S}}^{-1}\} \qquad (21)$$
is unbiased for $\text{tr}\{\boldsymbol{S}^{-2}\}$ 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 $\text{tr}\{\boldsymbol{S}^{-2}\}$ using just the first term of (21), namely $\left[1 - \frac{p}{K}\right]^2 \text{tr}\{\hat{\boldsymbol{S}}^{-2}\}$, which is only asymptotically unbiased.

B. Simulation Results

These estimators of the trace terms were used in the three estimators, now called QLa-est, QLb-est and QLP-est to produce $\hat{\rho}, (\hat{\rho}, \hat{\eta}), (\hat{\alpha}, \hat{\beta})$, respectively, 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.

Let us first consider individual-1. QLb-est performs well in an average-PRISE-over-frequencies sense, but has a relatively high variability (standard error). At some frequencies where the true partial coherence is high (2 and 3.25 Hz) it does poorly. QLP-est seems to deliver very significant improvements over the raw estimates, and has a relatively low variability. QLa-est is generally inferior to the other two (in terms of average PRISE) but appears relatively unaffected by spikes of high true partial coherence and appears to be least variable. For individual-2, QLb-est and QLP-est both do well.

method	params.	individual-3		
		$K = 12$	$K = 14$	$K = 16$
HS	ρ	30 (39)	-23 (68)	-77 (96)
QLa	ρ	54 (12)	40 (11)	31 (10)
QLb	ρ, η	66 (14)	49 (18)	36 (21)
HSP	α, β	35 (31)	22 (33)	15 (33)
QLP	α, β	59 (16)	50 (10)	41 (9)
QLa-est	$\hat{\rho}$	36 (9)	27 (8)	21 (8)
QLb-est	$\hat{\rho}, \hat{\eta}$	63 (15)	46 (20)	32 (24)
QLP-est	$\hat{\alpha}, \hat{\beta}$	56 (18)	48 (13)	40 (12)

TABLE II
PRISE(%) TO NEAREST INTEGER. AVERAGE AND STANDARD ERROR (IN BRACKETS) OVER FREQUENCIES. GAUSSIAN MODEL.

C. Comments

The QLP-est estimator is derived from (19) and (20) by replacing the trace terms by their estimators. These trace terms involve only $\boldsymbol{S}$ and $\boldsymbol{S}^2$. By way of contrast the QLa-est and QLb-est estimators are derived from (11) and (15), respectively, by estimating their trace terms, but in both cases these trace terms involve $\boldsymbol{S}^{-1}$ and $\boldsymbol{S}^{-2}$, which typically we would expect to be less accurately estimated than $\boldsymbol{S}$ and $\boldsymbol{S}^2$.

Under QL, with β_0 in the form given in (19), there is a $\text{tr}\{\boldsymbol{S}\}$ in the denominator which is playing the role of $\text{tr}\{\hat{\boldsymbol{S}}\}$ in the previously accepted formulation of (17). This adds to the appeal of QLP-est.

Like QLP-est, the estimator QLb-est is a two-parameter formulation. This affords extra flexibility over the one-parameter alternative QLa-est and may account for its superior mean behaviour.

Our results suggest that QLP-est is a good choice for practical estimation, generally having lower variability than QLb-est.

VIII. Consistent Behaviour

We have looked in detail at results for two individuals as proxies for cases of relatively high and low coherency. The same general behaviour of the shrinkage estimators is found for other individuals. By way of illustration, results for another person, individual-3, showing high partial coherencies over some frequencies are shown in Table II. (The results were produced in the same way as for individual-2 in Table I.) As before, the oracle results show that the QL approach is much better than HS, and QLb-est and QLP-est perform better than QLa-est in terms of average over frequencies.

IX. Robustness to Non-Gaussianity

The results presented so far assume a Gaussian probability structure. In the simulation algorithm [4] we replaced the independent Gaussian innovations (with variance unity) driving the simulation by (i) independent innovations having the uniform distribution with variance unity, and (ii) independent

method	individual-1			individual-2
	$K=12$	$K=14$	$K=16$	$K=12$
QLa-est	26 (7)	18 (6)	14 (5)	66 (4)
QLb-est	65 (14)	49 (22)	33 (29)	87 (3)
QLP-est	52 (9)	34 (9)	25 (8)	84 (6)
QLa-est	25 (7)	17 (6)	13 (5)	65 (4)
QLb-est	65 (15)	49 (21)	36 (27)	87 (3)
QLP-est	51 (9)	34 (9)	25 (8)	83 (6)

TABLE III
PRISE(%) TO NEAREST INTEGER. AVERAGE AND STANDARD ERROR (IN BRACKETS) OVER FREQUENCIES. UNIFORM INNOVATIONS (TOP BLOCK) AND t-INNOVATIONS (BOTTOM BLOCK).

innovations having the Student-t distribution (with 5 degrees of freedom) and variance unity. The first distribution has shorter tails than the Gaussian, and the second has longer tails. All the analyses in Table I were re-run for both of these alternative distributions. There was no notable differences from the Gaussian case and none of the conclusions changed. Table III shows results for the (non-oracle) estimation which can be compared with the bottom three rows of Table I. These results suggest robustness to non-Gaussianity, undoubtedly due in part to the central limit theorem effect on Fourier transform terms such as $\boldsymbol{J}_k(f)$, used in (2).

X. CONCLUDING DISCUSSION

We have considered how to improve estimation of partial coherencies from poorly-conditioned matrices. Exact simulations derived from real EEG data, and detailed analyses of results, were carried out 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, based on the results here, QLP-est seems a good choice of method. The methodology and results were also successfully reviewed for robustness to choice of individuals and to non-Gaussianity.

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 $\boldsymbol{A}, \boldsymbol{B}$ be arbitrary complex-valued matrices. Under (4),

$$\begin{aligned}
E\{\mathrm{tr}\{\boldsymbol{A}\hat{\boldsymbol{S}}\boldsymbol{B}\hat{\boldsymbol{S}}\}\} &= \mathrm{tr}\{\boldsymbol{ASBS}\} \\
&\quad +\tfrac{1}{K}\mathrm{tr}\{\boldsymbol{AS}\}\mathrm{tr}\{\boldsymbol{BS}\} && (22)\\
E\{\mathrm{tr}^2\{\hat{\boldsymbol{S}}\}\} &= \mathrm{tr}^2\{\boldsymbol{S}\} + \tfrac{1}{K}\mathrm{tr}\{\boldsymbol{S}^2\} && (23)\\
E\{\mathrm{tr}\{\hat{\boldsymbol{S}}^{-1}\}\} &= \tfrac{K}{K-p}\mathrm{tr}\{\boldsymbol{S}^{-1}\} && (24)\\
E\{\mathrm{tr}\{\boldsymbol{A}\hat{\boldsymbol{S}}^{-1}\boldsymbol{B}\hat{\boldsymbol{S}}^{-1}\}\} &= c_1\left[(K-p)\mathrm{tr}\{\boldsymbol{AS}^{-1}\boldsymbol{BS}^{-1}\}\right. \\
&\quad \left. +\ \mathrm{tr}\{\boldsymbol{AS}^{-1}\}\mathrm{tr}\{\boldsymbol{BS}^{-1}\}\right] && (25)\\
E\{\mathrm{tr}^2\{\hat{\boldsymbol{S}}^{-1}\}\} &= c_1\left[(K-p)\mathrm{tr}^2\{\boldsymbol{S}^{-1}\}\right. \\
&\quad \left. +\ \mathrm{tr}\{\boldsymbol{S}^{-2}\}\right]. && (26)
\end{aligned}$$

Here (22) and (23) hold for $K > p-1$, (24) holds for $K > p$ and (25) and (26) 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, $\boldsymbol{F}(x)$ is an $m \times n$ matrix and $\boldsymbol{G}(x)$ is an $n \times q$ matrix, then [15, p. 301] $\partial(\boldsymbol{FG})/\partial x = \boldsymbol{F}(\partial\boldsymbol{G}/\partial x) + (\partial\boldsymbol{F}/\partial x)\boldsymbol{G}$. Also, [15, p. 304] $\partial\,\mathrm{tr}\{\boldsymbol{F}\}/\partial x = \mathrm{tr}\{\partial\boldsymbol{F}/\partial x\}$. So,

$$\begin{aligned}
\frac{\partial\,\mathrm{tr}\{\boldsymbol{FG}\}}{\partial x} &= \mathrm{tr}\left\{\boldsymbol{F}\frac{\partial\boldsymbol{G}}{\partial x} + \frac{\partial\boldsymbol{F}}{\partial x}\boldsymbol{G}\right\} \\
&= \mathrm{tr}\left\{\boldsymbol{F}\frac{\partial\boldsymbol{G}}{\partial x}\right\} + \mathrm{tr}\left\{\boldsymbol{G}\frac{\partial\boldsymbol{F}}{\partial x}\right\} \\
&= \mathrm{tr}\left\{\frac{\partial\boldsymbol{G}}{\partial x}\boldsymbol{F}\right\} + \mathrm{tr}\left\{\frac{\partial\boldsymbol{F}}{\partial x}\boldsymbol{G}\right\},
\end{aligned}$$

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

$$\frac{\partial\,\mathrm{tr}\{\boldsymbol{F}^2\}}{\partial x} = 2\,\mathrm{tr}\left\{\frac{\partial\boldsymbol{F}}{\partial x}\boldsymbol{F}\right\} = 2\,\mathrm{tr}\left\{\boldsymbol{F}\frac{\partial\boldsymbol{F}}{\partial x}\right\}. \qquad (27)$$

C. Proof of Lemma 1

From (10) the quadratic loss is $\mathcal{R}_{QL}(\hat{\boldsymbol{S}}^{\star}(\rho), \boldsymbol{S}) = E\{\mathrm{tr}\{(\hat{\boldsymbol{S}}^{\star}\boldsymbol{S}^{-1} - \boldsymbol{I}_p)^2\}\}$. So

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

We now use (27) to find the derivative by setting

$$\boldsymbol{F} = \hat{\boldsymbol{S}}^{\star}\boldsymbol{S}^{-1} - \boldsymbol{I}_p = [(1-\rho)\hat{\boldsymbol{S}} + \frac{\rho}{p}\mathrm{tr}\{\boldsymbol{S}\}\boldsymbol{I}_p]\boldsymbol{S}^{-1} - \boldsymbol{I}_p.$$

Then,

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

Using (22),

$$\begin{aligned}
E\{\mathrm{tr}\{\boldsymbol{S}^{-1}\hat{\boldsymbol{S}}\boldsymbol{S}^{-1}\hat{\boldsymbol{S}}\}\} &= [\mathrm{tr}\{\boldsymbol{S}^{-1}\boldsymbol{S}\boldsymbol{S}^{-1}\boldsymbol{S}\} + \frac{1}{K}\mathrm{tr}^2\{\boldsymbol{S}^{-1}\boldsymbol{S}\}] \\
&= \left[p + \frac{p^2}{K}\right]. \qquad (28)
\end{aligned}$$

Using (28) we obtain

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

setting the result to zero and tidying gives

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

which gives the expression (11) for ρ.

The second derivative is given by

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

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

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

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

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

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 $\boldsymbol{S}$. So $p\geq\text{tr}^2\{\boldsymbol{S}^{-1}\}/\text{tr}\{\boldsymbol{S}^{-2}\}$, and so

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

which proves that the turning point is a minimum.

D. Proof of Lemma 2

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

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

where we have used (27) in AppendixB and $E\{\hat{\boldsymbol{S}}\}=\boldsymbol{S}$. Then setting to zero gives

$$\eta_0=\text{tr}\{\boldsymbol{S}\}/p. \tag{29}$$

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

$$\begin{aligned}&E\{\text{tr}\{2(\eta\boldsymbol{I}_p-\hat{\boldsymbol{S}})[(1-\rho)\hat{\boldsymbol{S}}+\eta\rho\boldsymbol{I}_p-\boldsymbol{S}]\}\}\\ &\quad=E\{\text{tr}\{2(\eta\boldsymbol{I}_p-\hat{\boldsymbol{S}})[\rho(\eta\boldsymbol{I}_p-\hat{\boldsymbol{S}})+\hat{\boldsymbol{S}}-\boldsymbol{S}]\}\}\\ \Rightarrow\rho_0\stackrel{\text{def}}{=}\rho_0(\eta_0) &= \frac{E\{\text{tr}\{(\eta_0\boldsymbol{I}_p-\hat{\boldsymbol{S}})(\boldsymbol{S}-\hat{\boldsymbol{S}})\}\}}{E\{\text{tr}\{(\eta_0\boldsymbol{I}_p-\hat{\boldsymbol{S}})^2\}\}}.\end{aligned} \tag{30}$$

The optimal shrinkage coefficient is dependent on the form of the target matrix, namely $\eta\boldsymbol{I}_p$. Look first at the numerator, a say, of (30). Expanding gives

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

Now $K\hat{\boldsymbol{S}}$ has the complex Wishart distribution with mean $K\boldsymbol{S}$. Since $E\{\hat{\boldsymbol{S}}\}=\boldsymbol{S}$ the first two terms cancel. Using (23) we get

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

Let the deniminator of (30) be denoted by b. Expanding gives, with $\eta_0=\text{tr}\{\boldsymbol{S}\}/p$,

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

The ratio a/b then has the form (9) or (14).

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

$$\begin{aligned}&E\{\text{tr}\{2\rho\boldsymbol{S}^{-1}([(1-\rho)\hat{\boldsymbol{S}}+\eta\rho\boldsymbol{I}_p]\boldsymbol{S}^{-1}-\boldsymbol{I}_p)\}\}\\ &\qquad=-2\rho^2\text{tr}\{\boldsymbol{S}^{-1}\}+2\rho^2\,\eta\,\text{tr}\{\boldsymbol{S}^{-2}\},\\ &\Rightarrow\eta_0 = \text{tr}\{\boldsymbol{S}^{-1}\}/\text{tr}\{\boldsymbol{S}^{-2}\}.\end{aligned} \tag{31}$$

Here we have again used (27) in AppendixB and $E\{\hat{\boldsymbol{S}}\}=\boldsymbol{S}$. This is very different in form to the scaling parameter (29) under HS loss. Then $(\partial/\partial\rho)\mathcal{R}_{QL}(\hat{\boldsymbol{S}}^\star(\eta,\rho),\boldsymbol{S})$ is given by $(\partial/\partial\rho)E\{\text{tr}\{[(1-\rho)\hat{\boldsymbol{S}}+\eta\rho\boldsymbol{I}_p]\boldsymbol{S}^{-1}-\boldsymbol{I}_p)^2\}\}$ which is

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

where we have again used (27) in AppendixB. Setting to zero gives

$$\rho_0\stackrel{\text{def}}{=}\rho_0(\eta_0)=\frac{E\{\text{tr}\{(\eta_0\boldsymbol{I}_p-\hat{\boldsymbol{S}})\boldsymbol{S}^{-1}(\boldsymbol{S}-\hat{\boldsymbol{S}})\boldsymbol{S}^{-1}\}\}}{E\{\text{tr}\{(\eta_0\boldsymbol{I}_p-\hat{\boldsymbol{S}})\boldsymbol{S}^{-1}(\eta_0\boldsymbol{I}_p-\hat{\boldsymbol{S}})\boldsymbol{S}^{-1}\}\}}. \tag{32}$$

Let the numerator of (32) be denoted by a. Expanding gives

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

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

From (28) we know that $E\{\text{tr}\{\hat{\boldsymbol{S}}\boldsymbol{S}^{-1}\hat{\boldsymbol{S}}\boldsymbol{S}^{-1}\}\}=(p+(p^2/K))$. So $a=-p+(p+(p^2/K))=p^2/K$.

Let the denominator of (32) be denoted by b. Expanding,

$$b=E\{\text{tr}\{\eta_0^2\boldsymbol{S}^{-2}-\eta_0\boldsymbol{S}^{-1}\hat{\boldsymbol{S}}\boldsymbol{S}^{-1}-\eta_0\hat{\boldsymbol{S}}\boldsymbol{S}^{-2}+\hat{\boldsymbol{S}}\boldsymbol{S}^{-1}\hat{\boldsymbol{S}}\boldsymbol{S}^{-1}\}\}. \tag{33}$$

Then

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

On tidying up the ratio a/b is of the form (15). 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 (23), for the HS loss

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

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

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

Similarly, for α,

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

For the determinant of the Hessian matrix we have

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

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

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

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

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

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

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

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

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

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

F. Proof of Lemma 4

With $\hat{\boldsymbol{C}}^{\star}(\alpha,\beta)$ defined in (16) and $\mathcal{R}_{HS}(\hat{\boldsymbol{C}}^{\star}(\alpha,\beta),\boldsymbol{C})$ we take

$$\boldsymbol{F}=\hat{\boldsymbol{C}}^{\star}\boldsymbol{S}-\boldsymbol{S}^{-1}=\alpha\,\hat{\boldsymbol{S}}^{-1}\boldsymbol{S}+\beta\,\boldsymbol{S}-\boldsymbol{S}^{-1}.$$

Using (24) and (25) we find that $(\partial/\partial\alpha)\mathcal{R}_{HS}(\hat{\boldsymbol{C}}^{\star}(\alpha,\beta),\boldsymbol{C})$ is given by

$$\begin{aligned}2\Big[&\alpha c_1[(K-p)\text{tr}\{\boldsymbol{S}^{-2}\}+\text{tr}^2\{\boldsymbol{S}^{-1}\}]\\ &+\ \beta\tfrac{K}{K-p}\text{tr}\{\boldsymbol{S}^{-1}\}-\tfrac{K}{K-p}\text{tr}\{\boldsymbol{S}^{-2}\}\Big].\end{aligned}\tag{34}$$

Using (24) we find $(\partial/\partial\beta)\mathcal{R}_{HS}(\hat{\boldsymbol{C}}^{\star}(\alpha,\beta),\boldsymbol{C})$ is given by

$$2\left[\alpha\tfrac{K}{K-p}\text{tr}\{\boldsymbol{S}^{-1}\}+\beta p-\text{tr}\{\boldsymbol{S}^{-1}\}\right].\tag{35}$$

Setting (34) and (35) 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

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

which is

$$\delta=c_1p[(K-p)\text{tr}\{\boldsymbol{S}^{-2}\}+\text{tr}^2\{\boldsymbol{S}^{-1}\}]-\left[\tfrac{K}{K-p}\right]^2\text{tr}^2\{\boldsymbol{S}^{-1}\}.$$

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

$$\begin{aligned}\delta &\geq c_1(K-p)\text{tr}^2\{\boldsymbol{S}^{-1}\}+c_1p\text{tr}^2\{\boldsymbol{S}^{-1}\}-\left[\tfrac{K}{K-p}\right]^2\text{tr}^2\{\boldsymbol{S}^{-1}\}\\ &= \left(c_1K-\left[\tfrac{K}{K-p}\right]^2\right)\text{tr}^2\{\boldsymbol{S}^{-1}\}\\ &= \left(\frac{pK^2(K-p)+K^2}{(K-p)^4-(K-p)^2}\right)\text{tr}^2\{\boldsymbol{S}^{-1}\}>0,\end{aligned}$$

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

G. Proof of Lemma 5

With $\hat{\boldsymbol{C}}^{\star}(\alpha,\beta)$ defined in (16) and $\mathcal{R}_{QL}(\hat{\boldsymbol{C}}^{\star}(\alpha,\beta),\boldsymbol{C})$ we take

$$\boldsymbol{F}=\hat{\boldsymbol{C}}^{\star}\boldsymbol{S}-\boldsymbol{I}_p=\alpha\,\hat{\boldsymbol{S}}^{-1}\boldsymbol{S}+\beta\,\boldsymbol{S}-\boldsymbol{I}_p.$$

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

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

Setting $\boldsymbol{A}=\boldsymbol{B}=\boldsymbol{S}$ in (25),

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

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

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

Putting results together we see that

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

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

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

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

$$\alpha c+\beta\text{tr}\{\boldsymbol{S}\}-p = 0\tag{36}$$

$$\alpha\frac{K}{K-p}\text{tr}\{\boldsymbol{S}\}+\beta\text{tr}\{\boldsymbol{S}^2\}-\text{tr}\{\boldsymbol{S}\} = 0\tag{37}$$

Multiply (36) by $K\text{tr}\{\boldsymbol{S}\}/(K-p)$ and (37) by c. Subtracting, with $c = c_0 p$, gives β_0 in (19) and (20). Multiply (36) by $\text{tr}\{\boldsymbol{S}^2\}$ and (37) by $\text{tr}\{\boldsymbol{S}\}$. Subtracting, gives α_0 in (19) and (20).

The determinant of the Hessian (divided by 4) is

$$\left| \begin{array}{cc} \frac{K}{K-p}c & \frac{K}{K-p}\text{tr}\{\boldsymbol{S}\} \\ \frac{K}{K-p}\text{tr}\{\boldsymbol{S}\} & \text{tr}\{\boldsymbol{S}^2\} \end{array} \right|_{\alpha_0,\beta_0} = \frac{K}{K-p}[c\text{tr}\{\boldsymbol{S}^2\} - \frac{K}{K-p}\text{tr}^2\{\boldsymbol{S}\}].$$

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

$$\frac{pK^2}{(K-p)^2-1}\text{tr}\{\boldsymbol{S}^2\} - \frac{K}{K-p}\text{tr}^2\{\boldsymbol{S}\} \tag{38}$$

We know that $p\text{tr}\{\boldsymbol{S}^2\} \geq \text{tr}^2\{\boldsymbol{S}\}$, so (38) 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 $(\partial^2/\partial\beta^2)\mathcal{R}_{QL}(\hat{\boldsymbol{C}}^\star(\alpha,\beta),\boldsymbol{C}) = 2\text{tr}\{\boldsymbol{S}^2\} > 0$ we see that the turning point is indeed a minimum.

ACKNOWLEDGMENT

The work of Deborah Schneider-Luftman was supported by EPSRC (UK). The authors thank the referees for their many useful comments and suggestions.

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 Yok: 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," *IEE 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.