



## A note on covariance estimation in the unbiased estimator of risk framework



Bala Rajaratnam<sup>a</sup>, Dario Vincenzi<sup>b,\*</sup>

<sup>a</sup> Department of Statistics, 390 Serra Mall - Sequoia Hall, Stanford University, Stanford, CA 94305, USA

<sup>b</sup> Université Nice Sophia Antipolis, CNRS, Laboratoire Jean Alexandre Dieudonné, UMR 7351, 06100 Nice, France

### ARTICLE INFO

#### Article history:

Received 17 February 2015

Received in revised form 15 December 2015

Accepted 18 February 2016

Available online 27 February 2016

#### Keywords:

Covariance estimation

Stein's estimator

Unbiased estimator of risk framework

### ABSTRACT

The estimation of covariance matrices is an important area in multivariate statistics and arises naturally in many applications. Stein's covariance estimator is regarded as a benchmark in the literature, as it generally yields remarkable risk reductions compared to the maximum likelihood estimator (MLE) in small sample sizes. In its original or raw form, however, Stein's estimator becomes unbounded when two sample eigenvalues approach each other. Thus, in some settings, Stein's raw estimator has greater risk than the MLE. This implies that Stein's raw estimator is not uniformly better than the MLE and thus cannot always be used as an alternative. The problem of the unbounded behavior of Stein's raw estimator is often mitigated by employing an ad hoc isotoning algorithm which has no formal statistical basis. By leveraging Stein's unbiased estimator of risk framework, in this paper we propose a general approach that prevents the unbounded behavior of the unbiased estimator of risk as two sample eigenvalues approach each other. We then employ this framework to obtain a proof-of-concept for constructing covariance estimators which retain the attractive properties of Stein's estimator and are simultaneously uniformly better than the MLE.

© 2016 Elsevier B.V. All rights reserved.

### 1. Introduction

The estimation of covariance matrices has many important applications including astrophysics (Pope and Szapudi, 2008; Hamimeche and Lewis, 2009), the environmental sciences (Frei and Kunsch, 2012; Eguchi et al., 2010), genetics (Schäfer and Strimmer, 2005), economics (Ledoit and Wolf, 2004a), etc. For instance, in minimum variance portfolio optimization, the optimal portfolio weights are a function of the covariance matrix (Won et al., 2012). In climate applications, when imputing climate fields and undertaking paleoclimate reconstructions, modeling the covariance of the spatial field requires covariance estimation (Guillot et al., 2015).

The classical estimator for the covariance is the sample covariance matrix, which is also the maximum likelihood estimator (MLE). The performance of the MLE, however, is satisfactory only when the sample size  $n$  is much greater than the dimension of the population covariance matrix,  $p$ . For  $n$  close to  $p$ , the estimates of the larger eigenvalues are biased upwards and those of the smaller eigenvalues are biased downwards (a similar distortion is observed for any  $n$  if the population covariance matrix is close to the identity matrix) (Lin and Perlman, 1985). Therefore, for those applications for which the sample size is limited, more sophisticated estimators are required.

\* Corresponding author.

E-mail address: [dario.vincenzi@unice.fr](mailto:dario.vincenzi@unice.fr) (D. Vincenzi).

A fundamental contribution to the study of covariance estimation was given by Stein (1975, 1986). Stein derived the unbiased estimator of risk with respect to the entropy loss function for a generic orthogonally invariant estimator (Stein, 1986). By minimizing an approximate version of the unbiased estimator of risk, Stein obtained an orthogonally invariant estimator that shrinks the eigenvalues of the sample covariance matrix in order to counteract the aforementioned biases. In its “raw” form, Stein’s estimator can yield negative estimates for the eigenvalues of the covariance matrix or can distort the ordering of the sample eigenvalues. In addition, the risk of Stein’s raw estimator is not defined, because the corresponding unbiased estimator of risk is either infinite or take complex values on an open subset of its domain. Stein, therefore, proposed to use his raw estimator in tandem with an isotonizing algorithm, which preserves the positivity and the ordering of the sample eigenvalues. In practice, Stein’s isotonized estimator generally yields remarkable risk reductions with respect to the MLE; the relative reductions can in fact be as large as 75% in the case when the population covariance matrix is close to the identity (Lin and Perlman, 1985). The performance of this estimator is partly attributable to the fact that it minimizes an optimality criterion. For these and other reasons, Stein’s estimator is regarded as a reference estimator in the literature. However, numerical studies show that there are situations in which the risk of Stein’s estimator can be slightly greater than that of the MLE (Lin and Perlman, 1985; Rajaratnam and Vincenzi, submitted for publication). Hence, Stein’s estimator is not uniformly better than the MLE. Furthermore, the risk reductions obtained with Stein’s estimator are not uniform in the parameter space, but vary significantly with  $p$ ,  $n$ , and the population covariance matrix (Rajaratnam and Vincenzi, submitted for publication). Stein’s estimator does perform considerably better than the MLE when  $n$  is close to  $p$  or when some of the population eigenvalues are close to each other. It does not yield appreciable risk reductions when the population eigenvalues are moderately separated and, simultaneously,  $n$  is sufficiently large compared to  $p$  (Lin and Perlman, 1985; Rajaratnam and Vincenzi, submitted for publication).

Several other covariance estimators have been proposed after Stein’s seminal work (Haff, 1980; Dey and Srinivasan, 1985; Haff, 1991; Yang and Berger, 1994; Daniels and Kass, 1999, 2001; Ledoit and Wolf, 2004b; Rajaratnam et al., 2008; Hoff, 2009; Ledoit and Wolf, 2013) (see also Pourahmadi, 2011 for a review). However, under the entropy loss, Stein’s estimator still remains a contender. Therefore, the question arises as how to leverage Stein’s estimator in order to obtain a covariance estimator that both is uniformly better than the MLE and has risk properties comparable to those of Stein’s estimator. In this paper, we present a general approach for preventing the unbounded behavior of the unbiased estimator of risk of an orthogonally invariant estimator as any two sample eigenvalues approach each other. We then give a proof-of-concept for the construction of new covariance estimators by using Stein’s unbiased estimator of risk. We illustrate this approach through a class of two-dimensional covariance estimators for which we rigorously prove dominance over the MLE. We then analyze the risk properties of these estimators numerically. Our study allows us to characterize the shrinking effect of Stein’s raw estimator and its impact on the risk properties of Stein’s isotonized estimator.

The remainder of the paper is organized as follows. In Section 2, we recall Stein’s approach on orthogonally invariant covariance estimators and describe the two-dimensional case in detail. In Section 3, we develop the unbiased estimator of risk approach and give sufficient conditions on the eigenvalue estimates that ensure that the unbiased estimator of risk of an orthogonally invariant estimator is bounded as two sample eigenvalues approach each other. In Section 4, we apply this approach to the construction of two-dimensional covariance estimators. Section 5 contains a summary of the results and the conclusions.

## 2. Stein’s covariance estimator

Let  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  be a random sample from a  $p$ -dimensional multivariate normal distribution with zero mean and covariance matrix  $\Sigma$ . Throughout this paper, we shall assume  $n \geq p > 1$ . The sample covariance matrix  $S$  (using the same terminology as Stein, 1986 and not dividing  $S$  by  $n$ ) is constructed from the vectors  $\mathbf{X}_i$  in the following way:

$$S := \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^t. \quad (1)$$

The matrix  $S$  is distributed according to the  $p$ -dimensional Wishart distribution with scale parameter matrix  $\Sigma$  and  $n$  degrees of freedom, in symbols,  $S \sim W_p(\Sigma, n)$ . The matrix  $\Sigma$  will also be referred to as the population covariance matrix. In view of its symmetry,  $S$  can be written as

$$S = H \Lambda H^t, \quad (2)$$

where  $H$  is orthogonal and  $\Lambda = \text{diag}(l_1, l_2, \dots, l_p)$ . The  $l_j$  are the sample eigenvalues and are ordered in decreasing order, i.e.,  $l_1 \geq l_2 \geq \dots \geq l_p > 0$ . Using the same terminology as Stein (1986), a covariance estimator  $\widehat{\Sigma}$  is said to be *orthogonally invariant* if it is of the form  $\widehat{\Sigma} = H \Phi H^t$ , where  $\Phi = \text{diag}(\phi_1(\mathbf{I}), \phi_2(\mathbf{I}), \dots, \phi_p(\mathbf{I}))$  and  $H$  is defined in (2). Thus, an orthogonally invariant estimator modifies the spectrum of the sample covariance matrix, but retains its eigenvectors. The estimator of the  $j$ th eigenvalue of  $\Sigma$  can be regarded as a function  $\phi_j$  on the set:

$$D_p := \{\mathbf{I} = (l_1, l_2, \dots, l_p) \in \mathbb{R}^p : l_1 \geq l_2 \geq \dots \geq l_p > 0\}. \quad (3)$$

We shall say that  $\widehat{\Sigma}$  preserves the positivity of the sample eigenvalues if  $\phi_j(\mathbf{l}) > 0$  for all  $j = 1, \dots, p$  and for all  $\mathbf{l} \in D_p$ . Likewise  $\widehat{\Sigma}$  preserves the ordering of the sample eigenvalues if  $\phi_{j+1}(\mathbf{l}) \leq \phi_j(\mathbf{l})$  for all  $j = 1, \dots, p - 1$  and for all  $\mathbf{l} \in D_p$ . In what follows, it will also be useful to consider the functions:

$$\psi_j(\mathbf{l}) = \frac{\phi_j(\mathbf{l})}{l_j}, \quad j = 1, 2, \dots, p. \tag{4}$$

Since  $\phi_j(\mathbf{l}) = \psi_j(\mathbf{l})l_j$ ,  $\psi_j(\mathbf{l})$  determines the extent to which the sample eigenvalue  $l_j$  is amplified or reduced.

To characterize the performance of  $\widehat{\Sigma}$ , Stein used the following loss function (Stein, 1986):

$$L_1(\widehat{\Sigma}, \Sigma) = \text{tr}(\widehat{\Sigma} \Sigma^{-1}) - \ln[\det(\widehat{\Sigma} \Sigma^{-1})] - p \tag{5}$$

with corresponding risk function:

$$R(\widehat{\Sigma}, \Sigma) := \mathbb{E}[L_1(\widehat{\Sigma}, \Sigma)]. \tag{6}$$

Under the assumption that the  $\phi_j$  are continuously differentiable on  $D_p$ , Stein proved that the risk of an orthogonally invariant estimator  $\widehat{\Sigma}$  can be written as (Stein, 1986):

$$R(\widehat{\Sigma}, \Sigma) = \mathbb{E}[F(\mathbf{l})], \tag{7}$$

where

$$F(\mathbf{l}) = \sum_{j=1}^p \sigma_j(\mathbf{l}) - c_{p,n} \tag{8}$$

with

$$\sigma_j(\mathbf{l}) = \left( n - p + 1 + 2l_j \sum_{\substack{i=1 \\ i \neq j}}^p \frac{1}{l_j - l_i} \right) \psi_j(\mathbf{l}) + 2l_j \frac{\partial \psi_j}{\partial l_j} - \ln(\psi_j(\mathbf{l})), \tag{9}$$

$$c_{p,n} = p(1 + \ln 2) + \sum_{j=1}^p \frac{\Gamma'(\frac{1}{2}(n - j + 1))}{\Gamma(\frac{1}{2}(n - j + 1))}.$$

Here  $\Gamma$  denotes the Euler Gamma function and  $\Gamma'$  is its derivative. Eq. (7) indicates that  $F(\mathbf{l})$  is an unbiased estimator of the risk of  $\widehat{\Sigma}$ .

### 2.1. The maximum likelihood estimator (MLE)

The MLE of the covariance matrix  $\Sigma$  is:

$$\widehat{\Sigma}^{\text{ml}} := \frac{S}{n}, \tag{10}$$

where the sample covariance matrix  $S$  has been introduced in (1). The MLE obviously is an orthogonally invariant estimator with  $\phi_j^{\text{ml}}(\mathbf{l}) = l_j/n$ . For the MLE, the coefficients  $\psi_j(\mathbf{l})$  are constant and do not depend on  $j$ :  $\psi_j^{\text{ml}}(\mathbf{l}) = 1/n$  for all  $j = 1, \dots, p$ . The corresponding unbiased estimator of risk is also constant (Rajaratnam and Vincenzi, submitted for publication):

$$F^{\text{ml}}(\mathbf{l}) = K_{p,n}^{\text{ml}} := p(1 + \ln n) - c_{p,n}. \tag{11}$$

### 2.2. Stein's isotonized estimator

Stein's covariance estimator is orthogonally invariant. It is derived by disregarding the derivative terms  $\partial \psi_j / \partial l_j$  in the unbiased estimator of risk  $F(\mathbf{l})$  (see (8)) and by minimizing the resulting function with respect to  $\psi_j$ . This approach yields the following estimators of the eigenvalues of  $\Sigma$  (Stein, 1986):

$$\phi_j^{\text{raw}}(\mathbf{l}) := \frac{l_j}{\alpha_j(\mathbf{l})}, \quad \alpha_j(\mathbf{l}) = n - p + 1 + 2l_j \sum_{\substack{i=1 \\ i \neq j}}^p \frac{1}{l_j - l_i} \tag{12}$$

for  $j = 1, \dots, p$ . The estimators  $\phi_j^{\text{raw}}(\mathbf{l})$  assume negative values or violate the original ordering of the sample eigenvalues on the set:

$$Q_{p,n} := \left\{ \mathbf{l} \in D_p : \exists 1 \leq j \leq p \text{ s.t. } \alpha_j(\mathbf{l}) \leq 0 \text{ or } \lim_{\lambda \rightarrow \mathbf{l}} |\alpha_j(\lambda)| = \infty \right\} \\ \cup \left\{ \mathbf{l} \in D_p : \exists 2 \leq j \leq p \text{ s.t. } l_j/\alpha_j(\mathbf{l}) > l_{j-1}/\alpha_{j-1}(\mathbf{l}) \right\}. \tag{13}$$

(For a detailed characterization of  $Q_{p,n}$  as a function of  $p$  and  $n$ , see Rajaratnam and Vincenzi, submitted for publication.) If  $\mathbf{l} \in Q_{p,n}$ , the estimates  $\phi_j^{\text{raw}}(\mathbf{l})$  are then modified by means of an isotonizing algorithm, which consists in repeatedly

pooling adjacent estimates together until all the estimated eigenvalues are positive and ordered in the same way as the  $l_j$ . The estimate obtained by pooling a generic set of  $\phi_j^{\text{raw}}(\mathbf{I})$  is:

$$\phi_i^{\text{iso}}(\mathbf{I}) = \phi_{i+1}^{\text{iso}}(\mathbf{I}) = \dots = \phi_{i+k}^{\text{iso}}(\mathbf{I}) := \frac{l_i + l_{i+1} + \dots + l_{i+k}}{\alpha_i(\mathbf{I}) + \alpha_{i+1}(\mathbf{I}) + \dots + \alpha_{i+k}(\mathbf{I})} \tag{14}$$

with  $1 \leq k \leq p - i$ . The algorithm always produces the desired result of positive-eigenvalues estimate which preserve the order in the sample eigenvalues, because the estimates obtained by pooling all the  $\phi_j^{\text{raw}}(\mathbf{I})$  together are:

$$\phi_1^{\text{iso}}(\mathbf{I}) = \phi_2^{\text{iso}}(\mathbf{I}) = \dots = \phi_p^{\text{iso}}(\mathbf{I}) = \frac{\sum_{j=1}^p l_j}{\sum_{j=1}^p \alpha_j(\mathbf{I})} = \frac{\sum_{j=1}^p l_j}{np}. \tag{15}$$

For more details on the isotonized algorithm and on its implementation, the reader is referred to [Lin and Perlman \(1985\)](#). The explicit form of the estimators  $\phi_j^{\text{iso}}(\mathbf{I})$  for  $p = 2$  and for  $p = 3$  is given in [Rajaratnam and Vincenzi \(submitted for publication\)](#).

Finally, the complete form of Stein's covariance estimator is  $\Sigma^{\text{St}} := H\Phi^{\text{St}}(\mathbf{I})H^t$  with  $\Phi^{\text{St}}(\mathbf{I}) = \text{diag}(\phi_1^{\text{St}}(\mathbf{I}), \phi_2^{\text{St}}(\mathbf{I}), \dots, \phi_p^{\text{St}}(\mathbf{I}))$  and

$$\phi_j^{\text{St}}(\mathbf{I}) := \begin{cases} \phi_j^{\text{raw}}(\mathbf{I}) & \text{if } \mathbf{I} \in D_p \setminus Q_{p,n} \\ \phi_j^{\text{iso}}(\mathbf{I}) & \text{if } \mathbf{I} \in Q_{p,n} \end{cases} \quad (j = 1, \dots, p). \tag{16}$$

The coefficients  $\psi_j^{\text{St}}(\mathbf{I})$  are defined as in (4):

$$\psi_j^{\text{St}}(\mathbf{I}) = \frac{\phi_j^{\text{St}}(\mathbf{I})}{l_j}. \tag{17}$$

In what follows, the unbiased estimator of risk of Stein's estimator will be denoted as  $F^{\text{St}}(\mathbf{I})$ .

It is shown in [Appendix A](#) that the  $\phi_j^{\text{St}}$  are continuous everywhere on  $D_p$  but continuously differentiable only almost everywhere. Nevertheless, (7) holds for Stein's estimator because, in the derivation of (7), all the differentiations are performed under the integral sign ([Stein, 1986](#)).

### 2.3. Stein's estimator for $p = 2$

By writing the ratios  $l_k/l_j$ ,  $1 \leq k < j \leq p$ , as

$$\frac{l_j}{l_k} = \frac{l_j}{l_{j-1}} \frac{l_{j-1}}{l_{j-2}} \dots \frac{l_{k+2}}{l_{k+1}} \frac{l_{k+1}}{l_k}, \tag{18}$$

it can be seen that the  $\psi_j^{\text{St}}$ , and hence  $F^{\text{St}}$ , do not depend on the sample eigenvalues individually, but only on the ratios  $0 < l_{j+1}/l_j \leq 1$ ,  $j = 1, \dots, p - 1$  ([Rajaratnam and Vincenzi, submitted for publication](#)). This fact can be checked by using (12), (14), (17) and (8). In particular, for  $p = 2$ , (17) yields:

$$\psi_j^{\text{St}}(l_1, l_2) = \tilde{\psi}_j^{\text{St}}\left(\frac{l_2}{l_1}\right), \quad j = 1, 2, \tag{19}$$

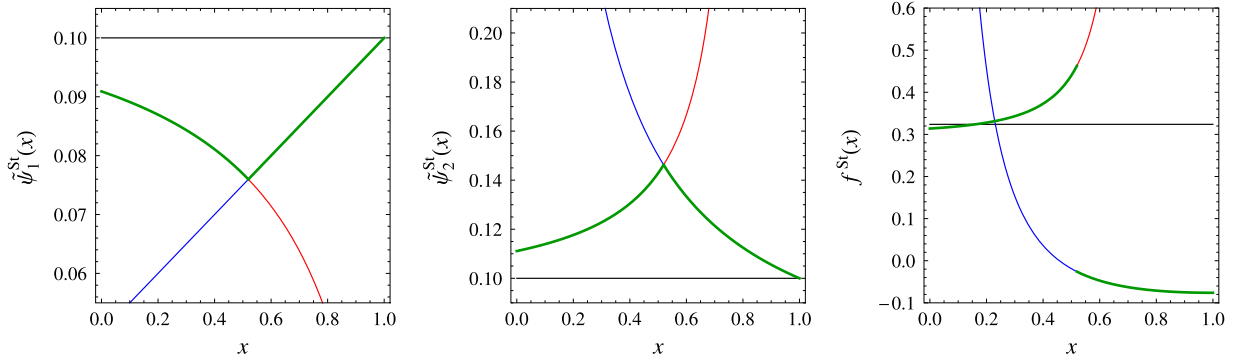
where

$$\tilde{\psi}_1^{\text{St}}(x) = \begin{cases} \tilde{\psi}_1^{\text{raw}}(x) = \frac{1-x}{n+1-(n-1)x} & 0 < x \leq x_n \\ \tilde{\psi}_1^{\text{iso}}(x) = \frac{1+x}{2n} & x_n < x \leq 1 \end{cases} \tag{20}$$

and

$$\tilde{\psi}_2^{\text{St}}(x) = \begin{cases} \tilde{\psi}_2^{\text{raw}}(x) = \frac{1-x}{n-1-(n+1)x} & 0 < x \leq x_n \\ \tilde{\psi}_2^{\text{iso}}(x) = \frac{1+x}{2nx} & x_n < x \leq 1. \end{cases} \tag{21}$$

The point  $x_n$ , which determines the transition from the raw estimator to the isotonized one, is  $x_n = (n + 1 - 2\sqrt{n})/(n - 1)$ . Thus, the set  $Q_{2,n}$  defined in (13) consists of the couples  $(l_1, l_2) \in D_2$  such that  $l_2/l_1 \geq x_n$ .



**Fig. 1.** Left and center: Graphs of  $\tilde{\psi}_j^{ml}$  (black), of  $\tilde{\psi}_j^{raw}$  (red), of  $\tilde{\psi}_j^{iso}$  (blue), and of  $\tilde{\psi}_j^{St}$  (green) for  $p = 2, n = 10$ , and  $j = 1, 2$ . The green line is superimposed on the graphs of  $\tilde{\psi}_j^{raw}$  and  $\tilde{\psi}_j^{iso}$  when they are the same as the graph of  $\tilde{\psi}_j^{St}$  ( $j = 1, 2$ ). Note that, for  $n = 10, x_n \approx 0.52$ . Right: Graphs of the unbiased estimator of risk of the MLE (black), of  $f^{raw}$  (red), of  $f^{iso}$  (blue), and of  $f^{St}$  (green) for  $p = 2$  and  $n = 10$ . The green line is superimposed on the graphs of  $f^{raw}$  and  $f^{iso}$  when they are the same as the graph of  $f^{St}$ . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The graphs of  $\tilde{\psi}_j^{St}$  ( $j = 1, 2$ ) are given in Fig. 1 for a representative value of  $n$ . Note that

$$\tilde{\psi}_1^{St}(x) \geq x \tilde{\psi}_2^{St}(x), \quad \forall x \in (0, 1], \tag{22}$$

which can be rewritten as:  $\phi_1^{St}(l_1, l_2) \geq \phi_2^{St}(l_1, l_2)$ . This inequality demonstrates that Stein’s estimator preserves the order of the sample eigenvalues. Moreover,

$$\tilde{\psi}_1^{St}(x) \leq \frac{1}{n} \leq \tilde{\psi}_2^{St}(x) \tag{23}$$

for all  $x \in (0, 1]$  and for all  $n \geq 2$ . Eq. (23) implies that  $\phi_1^{St}(l_1, l_2) \leq l_1/n$  and  $\phi_2^{St}(l_1, l_2) \geq l_2/n$ , and hence shows that Stein’s estimator shrinks the two sample eigenvalues closer together. Also note that the  $\tilde{\psi}_j^{St}$  are continuous on  $(0, 1]$  and continuously differentiable everywhere on  $(0, 1]$  except at  $x = x_n$ . In particular, the derivative of  $\tilde{\psi}_j^{St}$  changes sign at  $x = x_n$ , i.e., as we move from the raw to the isotonized form:

$$\begin{aligned} \lim_{x \rightarrow x_n^+} \frac{d\tilde{\psi}_1^{St}(x)}{dx} &= - \lim_{x \rightarrow x_n^-} \frac{d\tilde{\psi}_1^{St}(x)}{dx} = \frac{1}{2n} \quad \text{and} \\ \lim_{x \rightarrow x_n^+} \frac{d\tilde{\psi}_2^{St}(x)}{dx} &= - \lim_{x \rightarrow x_n^-} \frac{d\tilde{\psi}_2^{St}(x)}{dx} = - \frac{(\sqrt{n} + 1)^2}{2n(\sqrt{n} - 1)^2}. \end{aligned} \tag{24}$$

We can interpret this behavior by saying that the shrinking effect of the raw estimator becomes increasingly stronger as  $l_2$  approaches  $l_1$  and that the isotonizing algorithm mitigates this effect (Fig. 1).

The unbiased estimator of risk of Stein’s estimator can be obtained by substituting  $\psi_j^{St}(\mathbf{l})$  into (8). For  $p = 2$ , a straightforward calculation yields:

$$F^{St}(l_1, l_2) = f^{St} \left( \frac{l_2}{l_1} \right), \tag{25}$$

where

$$f^{St}(x) = \begin{cases} f^{raw}(x) & 0 < x \leq x_n \\ f^{iso}(x) & x_n < x \leq 1 \end{cases} \tag{26}$$

with

$$\begin{aligned} f^{raw}(x) &= 2 + 4x \left\{ \frac{1}{[n + 1 - (n - 1)x]^2} + \frac{1}{[(n + 1)x - (n - 1)]^2} \right\} \\ &\quad + \ln \left( n + \frac{1 + x}{1 - x} \right) + \ln \left( n - \frac{1 + x}{1 - x} \right) - c_{2,n} \end{aligned} \tag{27}$$

and

$$f^{iso}(x) = 1 - \frac{1}{n} + \frac{x}{2} - \frac{3x}{2n} + \frac{n - 3}{2nx} - \ln \left( \frac{1 + x}{2n} \right) - \ln \left( \frac{1 + x}{2nx} \right) - c_{2,n}. \tag{28}$$

The graph of  $f^{St}$  is given in Fig. 1.

A comprehensive study of the function  $f^{St}$  for  $p = 2$  is undertaken in [Rajaratnam and Vincenzi \(submitted for publication\)](#). In particular, the following properties will be used in the remainder of the paper:  $f^{St}(0) < K_{2,n}^{ml}$  and  $f^{St}(1) < K_{2,n}^{ml}$  for all  $n \geq 2$ , i.e., the unbiased estimator of risk of Stein’s estimator is lower than that of the MLE when the sample eigenvalues are well separated or when they are very close to each other (see [Fig. 1](#)). By contrast, for any  $n$ , there exists  $0 < x_n^* < 3 - 2\sqrt{2}$  such that  $f^{St}(x) \geq K_{2,n}^{ml}$  on the interval  $[x_n^*, x_n]$  ([Rajaratnam and Vincenzi, submitted for publication](#)). This fact indicates that there is a subset of  $(0, 1]$  on which the unbiased estimator of risk of Stein’s estimator exceeds that of the MLE ([Fig. 1](#)); on this subset, the sample eigenvalues are moderately separated.

We also note that  $f^{St}$  is not continuous at  $x = x_n$  and drops abruptly from a value greater than  $K_{2,n}^{ml}$  to a value much smaller than  $K_{2,n}^{ml}$  ([Fig. 1](#)). The discontinuity of  $f^{St}$  at  $x = x_n$  is due to the fact that  $f^{St}$  depends on  $d\tilde{\psi}_j^{St}/dx$  (see (8)) and the  $\tilde{\psi}_j^{St}$  are not continuously differentiable at  $x = x_n$ . We observe that each  $\tilde{\psi}_j^{St}$  is continuous and take comparable values on the two intervals  $(0, x_n]$  and  $(x_n, 1]$  and that  $|d\tilde{\psi}_j^{St}/dx|$  behaves analogously. Therefore, the drop in value of  $f^{St}$  should be attributed to the change of sign of the derivative of the  $\tilde{\psi}_j^{St}$  at  $x = x_n$ . In other words, the fact that, as  $x$  approaches  $x_n$ , the raw estimator increasingly shrinks the  $l_j$  induces a rapid increase of  $f^{St}(x)$  beyond the value of the unbiased estimator of risk of the MLE. The isotonizing algorithm produces a drop in value of  $f^{St}(x)$  by counteracting the shrinking effect of the raw estimator.

The above considerations will be used in the next sections to show a proof-of-concept on how to construct covariance estimators that are uniformly better than the MLE.

### 3. Unbiased estimator of risk approach

The unbiased estimator of risk presents a possible framework in which  $\psi_j(\mathbf{I})$  can be appropriately chosen in order to obtain an estimator with good risk properties, an estimator that is potentially better than the MLE or Stein’s estimator and that does not rely on the isotonizing algorithm. In this section, we show that any modified estimator that seeks to improve Stein’s estimator within the unbiased estimator of risk framework should satisfy some general conditions. In addition, we show how the unbiased estimator of risk approach can be used to construct estimators that are uniformly better than the MLE.

Expressing the unbiased estimator of risk in terms of the ratios of the sample eigenvalues is peculiar to Stein’s raw estimator and its isotonized counterpart. We now return to expressing the unbiased estimator of risk in terms of the sample eigenvalues  $l_j$  (see (8)). Clearly, the factor multiplying  $\psi_j(\mathbf{I})$  in  $\sigma_j(\mathbf{I})$  diverges if there exists  $i \neq j$  such that  $l_i = l_j$  (or becomes arbitrarily large as one  $l_i$  approaches an  $l_j$ ). This in turn implies that the unbiased estimator of risk  $F(\mathbf{I})$  may also diverge. For  $F^{raw}(\mathbf{I})$ , this divergence or explosive behavior can be deduced from (27) for  $p = 2$  and was demonstrated in [Rajaratnam and Vincenzi \(submitted for publication\)](#) for a general  $p$ . Hence it is natural to consider a class of estimators  $\psi_j(\mathbf{I})$  such that  $F(\mathbf{I})$  is bounded for all  $\mathbf{I} \in D_p$ . To obtain a bounded unbiased estimator of risk  $F(\mathbf{I})$ , one might be tempted to introduce an estimator  $\psi_j$  that rectifies the aforementioned divergence in the  $\sigma_j(\mathbf{I})$ . [Proposition 1](#) demonstrates, however, that even the simple and natural requirement that the eigenvalues are positive necessitates that the  $\sigma_j(\mathbf{I})$  diverge as  $l_j \rightarrow l_i$ . Thus, the boundedness of  $F$  must be obtained through a combination of different diverging  $\sigma_j(\mathbf{I})$ . [Proposition 1](#) also gives sufficient conditions on the  $\psi_j$  which ensure that  $F(\mathbf{I})$  is bounded for all  $\mathbf{I} \in D_p$ .

**Proposition 1.** *Let  $1 \leq j < k \leq p$  and let  $\tilde{\mathbf{I}} \in D_p$  be such that  $\tilde{l}_j = \tilde{l}_k$  and  $\tilde{l}_i \neq \tilde{l}_r$  for all  $i, r = 1, \dots, p, r \neq i, j, k$ . If  $\psi_i$  is positive and differentiable for all  $i = 1, \dots, p$  and  $\mathbf{I} \in D_p$ , then there exists  $1 \leq q \leq p$  such that*

$$\limsup_{\mathbf{I} \rightarrow \tilde{\mathbf{I}}} |\sigma_q(\mathbf{I})| = \infty.$$

*If in addition  $\psi_i(\mathbf{I})$  is bounded with bounded derivatives for all  $i = 1, \dots, p$  and for all  $\mathbf{I} \in D_p$ , and  $\psi_j$  and  $\psi_k$  satisfy the following condition:*

$$\limsup_{\mathbf{I} \rightarrow \tilde{\mathbf{I}}} \frac{l_j \psi_j(\mathbf{I}) - l_k \psi_k(\mathbf{I})}{l_j - l_k} = \frac{\phi_j(\mathbf{I}) - \phi_k(\mathbf{I})}{l_j - l_k} < \infty, \tag{29}$$

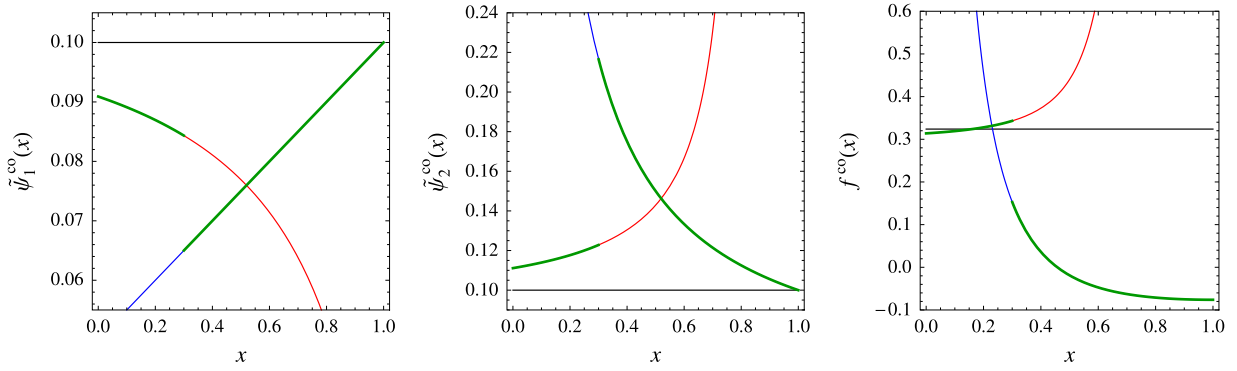
*then  $F(\mathbf{I})$  remains bounded as  $\mathbf{I} \rightarrow \tilde{\mathbf{I}}$ .*

**Proof.** See [Appendix B](#).  $\square$

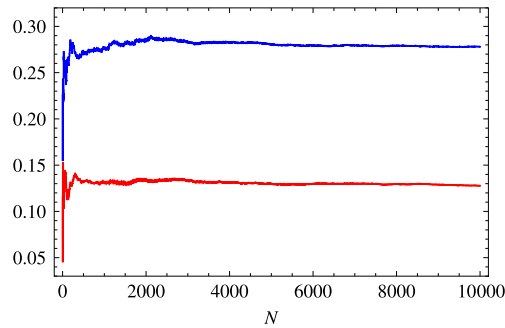
Inequality (29) is a differentiability condition for  $\varphi_j(\mathbf{I})$  as two sample eigenvalues approach one another. This condition essentially implies that, for  $F(\mathbf{I})$  to be bounded,  $\varphi_j(\mathbf{I})$  should converge to  $\varphi_k(\mathbf{I})$  at an equal or faster rate as compared to the rate at which  $l_j$  approaches  $l_k$ . We note that the MLE and Stein’s isotonized estimator satisfy (29).

The assumptions of [Proposition 1](#) are rather broad, hence the requirement that  $F(\mathbf{I})$  is bounded under these assumptions does not yield a sufficiently small and well defined class of estimators with some structure that can be immediately studied further. However, it does specify a basic and important sufficient condition that the  $\varphi_j(\mathbf{I})$  need to satisfy as  $l_j \rightarrow l_k$  for  $j \neq k$  in order for  $F(\mathbf{I})$  to be bounded. Thus, if two orthogonally invariant estimators  $\widehat{\Sigma}, \widehat{\Sigma}'$  satisfy (7) and (29) and their unbiased estimators of risk  $F, F'$  are such that:

$$F(\mathbf{I}) < F'(\mathbf{I}) \tag{30}$$



**Fig. 2.** Left and center: Graphs of  $\tilde{\psi}_j^{ml}$  (black), of  $\tilde{\psi}_j^{raw}$  (red), of  $\tilde{\psi}_j^{iso}$  (blue), and of  $\tilde{\psi}_j^{co}$  (green) for  $p = 2, n = 10, c = 0.3$ , and  $j = 1, 2$ . The green line is superimposed on the graphs of  $\tilde{\psi}_j^{raw}$  and  $\tilde{\psi}_j^{iso}$  when they are the same as the graph of  $\tilde{\psi}_j^{co}$  ( $j = 1, 2$ ). Right: Graphs of the unbiased estimator of risk of the MLE (black), of  $f^{raw}$  (red), of  $f^{iso}$  (blue), and of  $f^{co}$  (green) for  $p = 2, n = 10$ , and  $c = 0.3$ . The green line is superimposed on the graphs of  $f^{raw}$  and  $f^{iso}$  when they are the same as the graph of  $f^{co}$ . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 3.** Empirical means of  $L_1(\hat{\Sigma}^{co}, \Sigma)$  (blue) and of  $F^{co}(\mathbf{I})$  (red) as a function of the number of realizations  $N$  for  $\Sigma = \text{diag}(1, 0.5)$ ,  $n = 10$ ,  $x_n = (11 - 2\sqrt{10})/9 \approx 0.52$ , and  $c = 0.3$ . Note that in this case  $F^{ml}(\mathbf{I}) \approx 0.32$ . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

for Lebesgue almost all  $\mathbf{I} \in D_p$ , then  $\hat{\Sigma}$  is uniformly better than  $\hat{\Sigma}'$ . We shall demonstrate that this approach can be used to construct orthogonally invariant estimators that are uniformly better than the MLE.

In Section 2, we have recalled that, for all  $n$ , there exists a subinterval of  $(0, 1]$  on which  $F^{St}(\mathbf{I}) \geq F^{ml}(\mathbf{I})$ . Therefore, (7) cannot be used to prove that Stein's estimator is uniformly better than the MLE. Nevertheless, a suitable modification of Stein's unbiased estimator of risk may yield an orthogonally invariant estimator that satisfies (30) and hence is uniformly better than the MLE. The following example, however, shows that this strategy must be applied carefully.

Fig. 1 indicates that there exists a subinterval of  $(0, 1]$  on which  $f^{iso}(x) < f^{raw}(x)$  but  $f^{St}(x) = f^{raw}(x)$ , i.e., Stein's estimator uses the raw form even though the isotonized one would yield a lower unbiased estimator of risk. An analogous behavior is observed for other values of  $n$ . In order to obtain a lower unbiased estimator of risk, one may be tempted to construct a new orthogonally invariant estimator that uses the isotonized form not only for  $x > x_n$  but also for smaller values of  $x$ . The form of the resulting "cut-off" estimator is:

$$\psi_j^{co}(l_1, l_2) := \begin{cases} \tilde{\psi}_j^{raw}(l_2/l_1) & \text{if } 0 < l_2/l_1 \leq c \\ \tilde{\psi}_j^{iso}(l_2/l_1) & \text{if } c < l_2/l_1 \leq 1 \end{cases} \quad (31)$$

with  $0 < c < x_n$  and  $j = 1, 2$ . This estimator only differs from Stein's estimator in that the transition from the raw to the isotonized form occurs at smaller  $x$  (Fig. 2). Clearly, the function  $F^{co}(\mathbf{I}) = f^{co}(l_2/l_1)$  that is obtained by substituting  $\psi_j^{co}(\mathbf{I})$  into (8) is lower than the unbiased estimator of risk of Stein's estimator or equal to it everywhere on the domain of definition (compare Figs. 1 and 2).

We now consider a sample of  $N$  symmetric matrices generated numerically from  $W_2(\Sigma, n)$  with  $\Sigma = \text{diag}(1, 1/2)$  and  $n = 10$ . We then compute: (a) the empirical mean of  $L_1(\hat{\Sigma}^{co}, \Sigma)$ , where  $\hat{\Sigma}^{co}$  is the orthogonally invariant estimator constructed from  $\psi_j^{co}(\mathbf{I})$ , and (b) the empirical mean of the function  $F^{co}(\mathbf{I})$ . Fig. 3 shows that the two empirical means converge to different values as  $N$  increases. We conclude that (7) does not hold for the cut-off estimator, i.e.,  $F^{co}$  is not the unbiased estimator of risk of  $\hat{\Sigma}^{co}$ . In this case, the mean of  $F^{co}$  does not represent the risk of the cut-off estimator. The reason for this behavior is that if  $c \neq x_n$ , the  $\psi_j^{co}$  are not continuous functions at  $l_2/l_1 = c$  (Fig. 2) and Stein's theory on the unbiased

estimator of risk does not apply. As a consequence, (7) cannot be used to predict the properties of  $\widehat{\Sigma}^{\text{co}}$ ; we have modified  $\psi_j^{\text{St}}(\mathbf{I})$  in order to obtain an unbiased estimator of risk lower than  $F^{\text{St}}(\mathbf{I})$  or equal to it for all  $\mathbf{I}$ , but there is no guarantee that these modifications result in a corresponding reduction of the risk.

The above example shows that (30) can be used to prove that an orthogonal estimator is uniformly better than another orthogonal estimator only if both estimators satisfy (7). In particular, the latter condition holds if the modified  $\psi_j$  are continuous everywhere and continuously differentiable almost everywhere. This example also suggests that it is difficult to deduce the appropriate modification of Stein’s estimator directly from the form of the unbiased estimator of risk, because  $F$  is a nonlinear functional of the estimates  $\psi_j$ . The next section shows that by working with the estimates themselves, we can render  $F(\mathbf{I})$  lower than the unbiased estimator of risk of the MLE everywhere on its domain of definition.

#### 4. A class of estimators rendering the MLE inadmissible

We now focus on the two-dimensional case and show a proof-of-concept on how to build covariance estimators that are uniformly better than the MLE. For  $p = 2$ , the unbiased estimator of risk of Stein’s estimator reduces to a function of one variable and its behavior can be studied exactly (Rajaratnam and Vincenzi, submitted for publication). In Section 2, we have shown that the unbiased estimator of risk of Stein’s estimator has the following properties:

1.  $f^{\text{St}}(x) \leq K_{2,n}^{\text{ml}}$  both in a neighborhood of  $x = 0$  and of  $x = 1$ ;
2. the shrinking effect of the raw estimator strengthens as  $x$  increases. Moreover, as  $x$  approaches  $x_n$ , this behavior yields a fast increase of  $f^{\text{St}}(x)$  beyond  $K_{2,n}^{\text{ml}}$ ;
3. the isotoning algorithm counteracts the shrinking effect of the raw estimator and yields an unbiased estimator of risk lower than  $K_{2,n}^{\text{ml}}$ .

(We remind the reader that, for given  $p$  and  $n$ ,  $K_{2,n}^{\text{ml}}$  denotes the constant value taken by the unbiased estimator of risk of the MLE—see (11).) In order to construct estimators whose unbiased estimator of risk is lower than that of the MLE, we propose to modify  $\psi_j^{\text{St}}(\mathbf{I})$  in a way as to attenuate the shrinking effect of the raw estimator while retaining properties 1 and 3 above. This approach is demonstrated below by means of four different examples of orthogonally invariant estimators.

We consider orthogonally invariant estimators that satisfy the differentiability condition (29) and have the form  $\widehat{\Sigma}^{(k)} = H\Phi^{(k)}H^t$ , where  $H$  is defined in (2),  $\Phi^{(k)} = (\phi_1^{(1)}(\mathbf{I}), \phi_2^{(1)}(\mathbf{I}))$ , and  $k = 1, \dots, 4$ . The first estimator is a linear interpolation between  $\widetilde{\psi}_j^{\text{St}}(0)$  and  $\widetilde{\psi}_j^{\text{St}}(1)$ :

$$\phi_j^{(1)}(l_1, l_2) = l_j \psi_j^{(1)}\left(\frac{l_2}{l_1}\right) \quad (j = 1, 2) \tag{32}$$

with

$$\psi_j^{(1)}(x) = \widetilde{\psi}_j^{\text{St}}(0) + [\widetilde{\psi}_j^{\text{St}}(1) - \widetilde{\psi}_j^{\text{St}}(0)]x = \begin{cases} \frac{1}{n+1} + \frac{x}{n(n+1)}, & j = 1, \\ \frac{1}{n-1} - \frac{x}{n(n-1)}, & j = 2. \end{cases} \tag{33}$$

The graph of  $\psi_1^{(1)}$  is given in Fig. 4 for  $n = 10$ . The  $\phi_j^{(1)}$  clearly are continuously differentiable. Hence the unbiased estimator of risk of  $\widehat{\Sigma}^{(1)}$  is obtained by substituting the  $\psi_j^{(1)}(l_2/l_1)$  into (8) and takes the form  $F^{(1)}(l_1, l_2) = f^{(1)}(l_2/l_1)$  with

$$f^{(1)}(x) = 2 - \frac{8x}{n^2 - 1} - \ln\left[\frac{n+x}{n(n+1)}\right] - \ln\left[\frac{n-x}{n(n-1)}\right] - c_{2,n}. \tag{34}$$

The graph of  $f^{(1)}$  is shown in Fig. 5 for  $n = 10$ . Note that the unbiased estimator of risk of  $\widehat{\Sigma}^{(1)}$  is smaller than that of the MLE for all  $x \in (0, 1]$ . Indeed, the first derivative of  $f^{(1)}$  is:

$$\frac{df^{(1)}}{dx} = \frac{2P_n(x)}{(n^2 - 1)(n^2 - x^2)}, \tag{35}$$

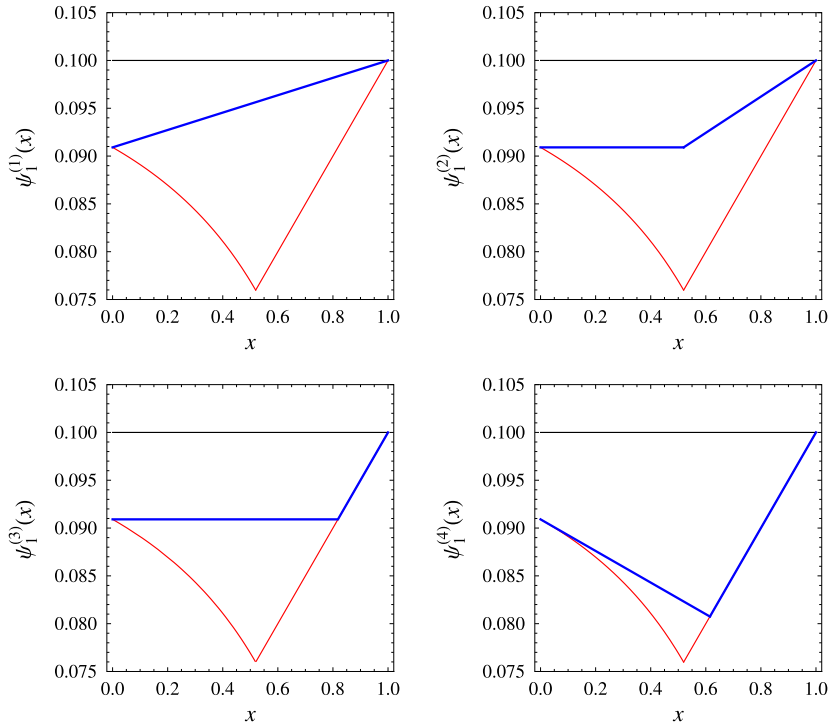
where  $P_n(x) = 4x^2 + (n^2 - 1)x - 4n^2$ . For all  $n$ , the polynomial  $P_n$  has no roots in the interval  $[0, 1]$  and hence has a constant sign on  $[0, 1]$ . In addition,  $P_n(0) = -4n^2 < 0$ . Therefore, the derivative of  $f^{(1)}$  is negative for all  $x \in (0, 1]$ . As  $f^{(1)}(0) = f^{\text{St}}(0) = 2 + \ln(n^2 - 1) - c_{2,n} < K_{2,n}^{\text{ml}}$  for all  $n \geq 2$ , we conclude that  $f^{(1)}(x) < K_{2,n}^{\text{ml}}$  for all  $x \in (0, 1]$ , whence  $\mathbb{E}[F^{(1)}(\mathbf{I})] < \mathbb{E}[F^{\text{ml}}(\mathbf{I})]$ . Consequently,  $R(\widehat{\Sigma}^{(1)}, \Sigma) < R(\widehat{\Sigma}^{\text{ml}}, \Sigma)$ , i.e., the estimator  $\widehat{\Sigma}^{(1)}$  is uniformly better than the MLE.

The second estimator again attenuates the shrinking effect of Stein’s raw estimator by using  $\widetilde{\psi}_j^{\text{St}}(0)$  for  $x \leq x_n$  and then interpolating linearly between the value at  $x = x_n$  and  $\widetilde{\psi}_j^{\text{iso}}(1)$  (Fig. 4):

$$\psi_j^{(2)}(x) = \begin{cases} \widetilde{\psi}_j^{\text{St}}(0), & 0 < x \leq x_n, \\ \frac{\widetilde{\psi}_j^{\text{St}}(0) - \widetilde{\psi}_j^{\text{St}}(1)x_n}{1 - x_n} + \frac{\widetilde{\psi}_j^{\text{St}}(1) - \widetilde{\psi}_j^{\text{St}}(0)}{1 - x_n}x, & x_n < x \leq 1 \end{cases} \tag{36}$$

with  $j = 1, 2$ .





**Fig. 4.** Graphs of the estimators  $\psi_1^{(1)}$  (top, left),  $\psi_1^{(2)}$  (top, right),  $\psi_1^{(3)}$  (bottom, left),  $\psi_1^{(4)}$  (bottom, right) for  $n = 10$ . In each of the plots, the black line is  $\psi_1^{ml}$ , the red line is  $\psi_1^{St}$ , and the blue line is  $\psi_1^{(k)}$ ,  $k = 1, \dots, 4$ . The blue line is superimposed on the graph of  $\psi_1^{St}$  when it is the same as the graph of  $\psi_1^{(k)}$ . The qualitative behavior of the estimators is the same for other values of  $n$ . The functions  $\psi_2^{(k)}$  behave analogously. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Compared to the previous estimators, the third one still attenuates the shrinking effect of Stein’s raw estimator, but also utilizes the isotoning algorithm near  $x = 1$ . More precisely, the third estimator uses  $\tilde{\psi}_j^{St}(0)$  up to  $x_n^*$  such that  $\tilde{\psi}_j^{iso}(x_n^*) = \tilde{\psi}_j^{St}(0)$  and  $\tilde{\psi}_j^{iso}(x)$  beyond (see Fig. 4):

$$\psi_1^{(3)} = \begin{cases} \frac{1}{n+1} & 0 < x \leq x_n^* \\ \frac{1+x}{2n} & x_n^* < x \leq 1 \end{cases} \quad \text{and} \quad \psi_2^{(3)} = \begin{cases} \frac{1}{n-1} & 0 < x \leq x_n^* \\ \frac{1+x}{2nx} & x_n^* < x \leq 1, \end{cases} \tag{37}$$

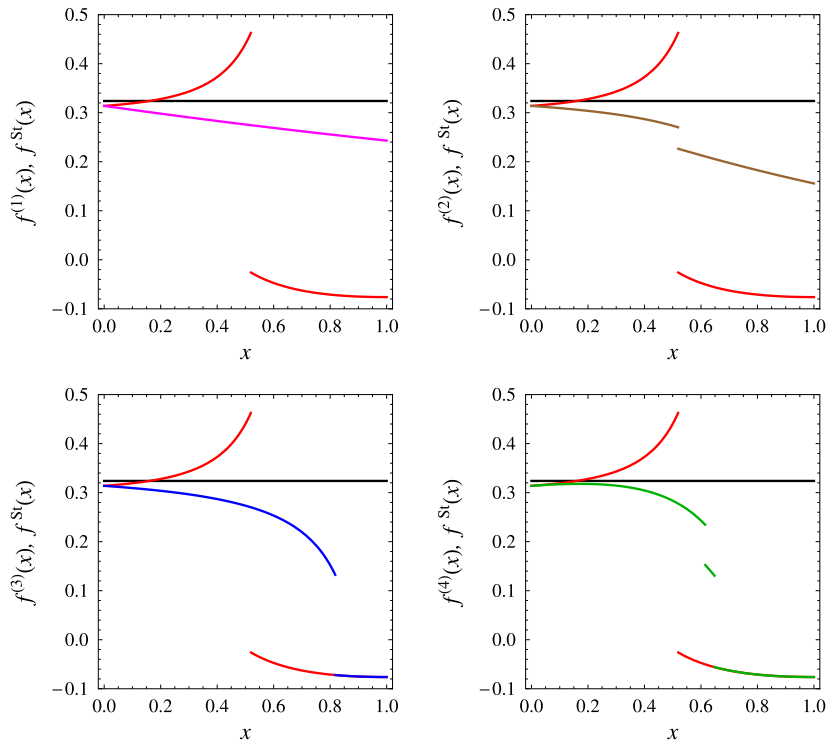
where  $x_n^* = (n - 1)/(n + 1)$ .

The fourth estimator utilizes the isotoning algorithm on a larger portion of the domain compared to the third estimator. As the  $\psi_j$  must be continuous, the reduction of the shrinking effect is weaker for this estimator. The explicit form of the fourth estimator is (Fig. 4):

$$\psi_j^{(4)}(x) = \begin{cases} \tilde{\psi}_j^{St}(0) + \left. \frac{d\tilde{\psi}_j^{St}}{dx} \right|_{x=0} x & 0 < x \leq \bar{x}_{j,n} \\ \tilde{\psi}_j^{St}(x) & \bar{x}_{j,n} < x \leq 1, \end{cases} \tag{38}$$

where  $j = 1, 2$  and  $\bar{x}_{j,n}$  is such that  $\tilde{\psi}_j^{St}(0) + \bar{x}_{j,n} \left. \frac{d\tilde{\psi}_j^{St}}{dx} \right|_{x=0} = \psi_j^{St}(\bar{x}_{j,n})$ , i.e.,

$$\psi_1^{(4)}(x) = \begin{cases} \frac{1}{n+1} - \frac{2x}{(n+1)^2} & 0 < x \leq \bar{x}_{1,n} \\ \frac{1+x}{2n} & \bar{x}_{1,n} < x \leq 1 \end{cases} \quad \text{with} \quad \bar{x}_{1,n} = \frac{n^2 - 1}{n(n + 6) + 1} \tag{39}$$



**Fig. 5.** Graphs of the unbiased estimators of risk of  $\widehat{\Sigma}^{(1)}$  (top, left, magenta),  $\widehat{\Sigma}^{(2)}$  (top, right, brown),  $\widehat{\Sigma}^{(3)}$  (bottom, left, blue),  $\widehat{\Sigma}^{(4)}$  (bottom, right, green) for  $n = 10$ . In all plots, the black curve is the graph of the unbiased estimator of risk of the MLE and the red curve is the graph of  $f^{St}$ . The graphs of the unbiased estimators of risk of two different estimators are superimposed when they are the same. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and

$$\psi_2^{(4)}(x) = \begin{cases} \frac{1}{n-1} + \frac{2x}{(n-1)^2} & 0 < x \leq \bar{x}_{2,n} \\ \frac{1+x}{2nx} & \bar{x}_{2,n} < x \leq 1 \end{cases} \quad \text{with } \bar{x}_{2,n} = \frac{n-1}{8n} \left( \sqrt{n(n+18)+1} - n - 1 \right). \quad (40)$$

Note that  $x_n < \bar{x}_{j,n} < 1$  for all  $n \geq 2$  ( $j = 1, 2$ ).

The unbiased estimators of risk of the estimators  $\widehat{\Sigma}^{(k)}$  are shown in Fig. 5. For all  $k = 2, 3, 4$  and for all  $n \geq 2$ , the unbiased estimator of risk of  $\widehat{\Sigma}^{(k)}$  is smaller than that of the MLE everywhere on  $(0, 1]$  (the proof of this fact amounts to comparing functions of one variable and is omitted). Thus, by using the same strategy used to prove that  $\widehat{\Sigma}^{(1)}$  is uniformly better than the MLE, we conclude that the estimators  $\widehat{\Sigma}^{(k)}$ ,  $k = 2, 3, 4$ , also are uniformly better than the MLE. The above results are now stated formally in the following proposition.

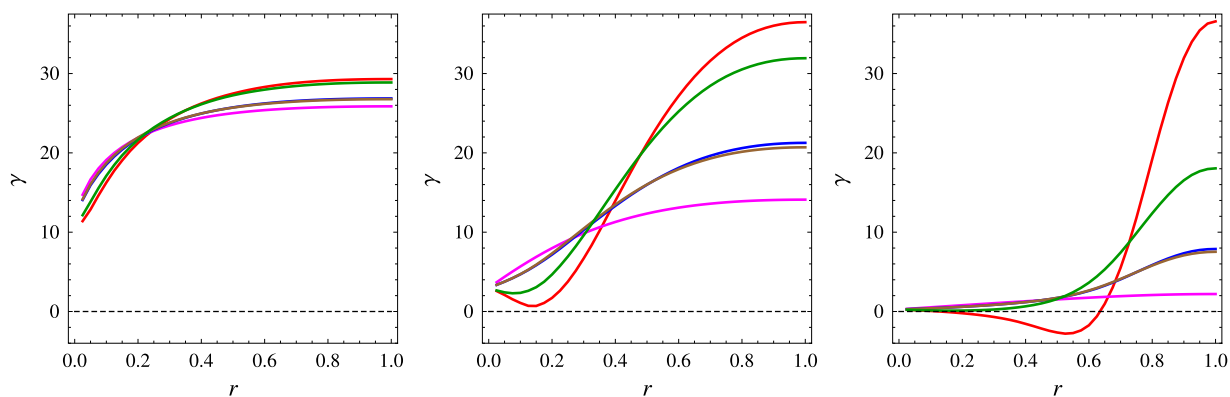
**Proposition 2.** Let  $X_1, \dots, X_n$  be an i.i.d. sample from a bivariate normal distribution with zero mean and covariance matrix  $\Sigma$  ( $n \geq 2$ ). Let  $\widehat{\Sigma}^{(k)}$ ,  $k = 1, \dots, 4$ , denote the orthogonal estimators of  $\Sigma$  as defined in Eqs. (32), (33), and (36)–(40). Then

$$R(\widehat{\Sigma}^{(k)}, \Sigma) \leq R(\widehat{\Sigma}^{ml}, \Sigma) \quad \forall \Sigma \in \mathbb{P}_+ \quad (k = 1, \dots, 4) \quad (41)$$

for all  $n \geq 2$ , and where  $\mathbb{P}_+$  denotes the cone of symmetric positive definite matrices.

We observe that all the estimators  $\widehat{\Sigma}^{(k)}$  prevent their unbiased estimator of risk from becoming greater than that of the MLE by attenuating the increasingly shrinking effect of Stein’s raw estimator. This behavior is more pronounced for the estimators  $\widehat{\Sigma}^{(1)}$ ,  $\widehat{\Sigma}^{(2)}$ ,  $\widehat{\Sigma}^{(3)}$ . For intermediate values of the ratio  $l_2/l_1$ , the shrinkage operated by  $\widehat{\Sigma}^{(4)}$  is comparable to that of Stein’s estimator; consequently, the reduction in the unbiased estimator of risk is less for intermediate  $l_2/l_1$ . We also note that, for all estimators, a considerable reduction of the unbiased estimator of risk is obtained only when the estimator makes use of the isotonizing algorithm (Fig. 5).

Although we have proved that the risk of  $\widehat{\Sigma}^{(k)}$  is always smaller than that of the MLE, it is important to quantify the risk reductions and compare the performance of  $\widehat{\Sigma}^{(k)}$  with that of Stein’s estimator. In Section 2, we observed that the unbiased estimator of risk of Stein’s estimator only depends on the ratios of the adjacent sample eigenvalues and not on the sample eigenvalues individually. By construction, the estimators  $\widehat{\Sigma}^{(k)}$  also have this property. This fact suggests that the dependence



**Fig. 6.** Percentage risk reductions for Stein’s estimator (red),  $\widehat{\Sigma}^{(1)}$  (magenta),  $\widehat{\Sigma}^{(2)}$  (brown),  $\widehat{\Sigma}^{(3)}$  (blue),  $\widehat{\Sigma}^{(4)}$  (green) and  $n = 2$  (left panel),  $n = 10$  (center panel),  $n = 100$  (right panel). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the risk of  $\widehat{\Sigma}^{(k)}$  upon the parameters can be studied by considering a diagonal population covariance matrix and by varying the ratio of its adjacent eigenvalues (the same approach is used in Rajaratnam and Vincenzi, submitted for publication to study the properties of Stein’s covariance estimator). For a generic covariance estimator  $\widehat{\Sigma}$ , we define the percentage risk reduction with respect to the MLE as:

$$\gamma := \frac{R(\widehat{\Sigma}^{ml}, \Sigma) - R(\widehat{\Sigma}, \Sigma)}{R(\widehat{\Sigma}^{ml}, \Sigma)} \times 100. \tag{42}$$

We then consider a sample of  $N = 10^4$  matrices generated numerically from  $W_2(\Sigma, n)$  with  $\Sigma = \text{diag}(1, r)$ ,  $0 < r \leq 1$ , and  $n = 2, 10, 100$ . The statistical sample is generated by following the same procedure as in Lin and Perlman (1985) and Rajaratnam and Vincenzi (submitted for publication). Finally, we compute  $\gamma$  as a function of  $r$  and  $n$  for the estimators  $\widehat{\Sigma}^{(k)}$ ,  $k = 1, \dots, 4$  and for Stein’s estimator (see Fig. 6). The analysis of the results leads to the following conclusions:

- i. The risks of the estimators  $\widehat{\Sigma}^{(k)}$  are obviously smaller than that of the MLE for all  $r \in (0, 1]$  and for all the values of  $n$  that we have considered. By contrast, for large  $n$ , the risk of Stein’s estimator can be greater than that of the MLE (for  $n = 100$  and for intermediate values of  $r$ ,  $\gamma$  is negative—see Fig. 6). Thus, the estimators  $\widehat{\Sigma}^{(k)}$  have a smaller risk compared to Stein’s estimator in the regime in which  $n/p$  is large and the eigenvalues of  $\Sigma$  are moderately or well separated. This is indeed the regime in which the performance of Stein’s estimator is worse (Rajaratnam and Vincenzi, submitted for publication);
- ii. For Stein’s estimator,  $\gamma$  varies considerably as a function of  $r$ , i.e., the performance of Stein’s estimator depends rather strongly on the parameters. By contrast, for the estimators  $\widehat{\Sigma}^{(k)}$  the variation of  $\gamma$  with  $r$  is smaller; hence, the performance of the estimators  $\widehat{\Sigma}^{(k)}$  is more uniform over the parameter space;
- iii. Stein’s estimator yields greater risk reductions when  $r$  is close to 1, i.e., when the population covariance matrix is close to the identity. In this regime, if  $n$  is large, none of the estimators  $\widehat{\Sigma}^{(k)}$  yields risk reductions comparable to those given by Stein’s estimator. This fact can be understood by considering the behavior of the unbiased estimator of risk as a function of the ratio of the sample eigenvalues. Recall that the risk is the average of the unbiased estimator of risk [(7)] and note that if  $\Sigma$  is close to the identity, the set on which  $l_2/l_1$  is near 1 contributes relatively more to the average. Furthermore, we have recalled that the unbiased estimator of risk is considerably smaller than that of the MLE only when the isotoning algorithm is used. Stein’s estimator and the estimators  $\widehat{\Sigma}^{(k)}$  make use of the isotoning algorithm when  $l_2/l_1$  is near 1; however, the estimators  $\widehat{\Sigma}^{(k)}$  employ the isotoning algorithm to a lesser extent. Therefore, when the population covariance matrix is near the identity, the estimators  $\widehat{\Sigma}^{(k)}$  benefit less than Stein’s estimator from the use of the isotoning algorithm and thus yield smaller risk reductions;
- iv. The risk of  $\widehat{\Sigma}^{(4)}$  is lower than that of the other estimators  $\widehat{\Sigma}^{(k)}$  when the eigenvalues of  $\Sigma$  are near each other (Fig. 5). Indeed,  $\widehat{\Sigma}^{(4)}$  utilizes the isotoning algorithm over a larger portion of the domain compared to  $\widehat{\Sigma}^{(1)}$ ,  $\widehat{\Sigma}^{(2)}$ ,  $\widehat{\Sigma}^{(3)}$ . Furthermore, for  $r$  near 1,  $\widehat{\Sigma}^{(2)}$  and  $\widehat{\Sigma}^{(3)}$  exhibit a faster shrinkage reduction than  $\widehat{\Sigma}^{(1)}$ . The result of this behavior is that  $\widehat{\Sigma}^{(2)}$  and  $\widehat{\Sigma}^{(3)}$  yield greater risk reductions than  $\widehat{\Sigma}^{(1)}$  for  $r$  near 1;
- v. The risks of  $\widehat{\Sigma}^{(1)}$ ,  $\widehat{\Sigma}^{(2)}$ ,  $\widehat{\Sigma}^{(3)}$  are smaller than that of  $\widehat{\Sigma}^{(4)}$  when the eigenvalues of  $\Sigma$  are moderately or well separated. This fact can again be understood by recalling that the risk is the average of the unbiased estimator of risk. For moderately or well separated population eigenvalues, the region in which Stein’s unbiased estimator of risk is greater than that of the MLE contributes significantly to the risk. We have seen that those estimators that attenuate the shrinking effect of Stein’s raw estimator slow down or even eliminate the increase of the unbiased estimator of risk beyond the value of the MLE. The estimators  $\widehat{\Sigma}^{(1)}$ ,  $\widehat{\Sigma}^{(2)}$ ,  $\widehat{\Sigma}^{(3)}$  significantly attenuate the shrinking effect of Stein’s raw estimator as the ratio of the sample eigenvalues increases, whereas the shrinking effect of  $\widehat{\Sigma}^{(4)}$  is comparable to that of Stein’s estimator (see Fig. 4). Therefore, if the eigenvalues of  $\Sigma$  are moderately or well separated, the risk of  $\widehat{\Sigma}^{(1)}$ ,  $\widehat{\Sigma}^{(2)}$ ,  $\widehat{\Sigma}^{(3)}$  is smaller than the risk of  $\widehat{\Sigma}^{(4)}$ .

## 5. Conclusions

Note that if an unbiased estimator of risk of a covariance estimator  $\widehat{\Sigma}$  is pointwise lower than the unbiased estimator of risk of the MLE, then  $\widehat{\Sigma}$  is uniformly better than the MLE. The unbiased estimator of risk of Stein's estimator is greater than that of the MLE on an open subset of its domain (Rajaratnam and Vincenzi, submitted for publication). This behavior is due to the fact that the shrinkage effect of the raw estimator rapidly becomes stronger and stronger before the isotonizing algorithm attenuates it (in the absence of isotonization, the unbiased estimator of risk of the raw estimator would actually grow in an unbounded way). As a consequence, the unbiased estimator of risk cannot be used to prove whether or not Stein's estimator is uniformly better than the MLE. We derive sufficient conditions on the eigenvalue estimates that ensure that the unbiased estimator of risk of an orthogonally invariant estimator is bounded as two sample eigenvalues approach each other. In addition, we show that within these conditions it is possible to modify Stein's estimator in order to reduce its shrinkage rate and thus render its unbiased estimator of risk lower than that of the MLE everywhere on the domain of definition. The estimators that are obtained by means of this approach are uniformly better than the MLE.

We have illustrated this technique for  $p = 2$ , because in this case the functional properties of the unbiased estimator of risk of Stein's estimator can be studied thoroughly. The two-dimensional estimators  $\widehat{\Sigma}^{(k)}$  that we have introduced in Section 4 have two properties in common:

- i. for moderately separated sample eigenvalues, they attenuate the shrinking effect of Stein's raw estimator;
- ii. they make use of the isotonizing algorithm to a lesser extent compared to Stein's estimator.

Property (i) implies that the unbiased estimator of risk of  $\widehat{\Sigma}^{(k)}$  does not exceed the value of the unbiased estimator of risk of the MLE. Thus, if the eigenvalues of  $\Sigma$  are moderately or well separated, the risk of the estimators  $\widehat{\Sigma}^{(k)}$  is smaller than that of Stein's estimator.

Property (ii) implies that the estimators  $\widehat{\Sigma}^{(k)}$  do not take full advantage of the reduction of the unbiased estimator of risk given by the isotonizing algorithm. Hence, if the population covariance matrix is close to the identity, they yield smaller risk reductions compared to Stein's estimator.

Our analysis identifies two effects that reduce the unbiased estimator of risk: (a) the attenuation of the shrinkage action of Stein's raw estimator and (b) isotonization, with the latter effect yielding considerably greater reductions than the former. The unbiased-estimator-of-risk approach relies on (7) and requires that the  $\phi_j$  are continuously differentiable. This requirement constraints the way in which the estimators  $\phi_j^{\text{St}}$  can be modified in order to reduce the unbiased estimator of risk and hence the risk of the estimator. In particular, it is not possible to take full advantage of the isotonizing algorithm if we want to maintain the form of  $\phi_j^{\text{St}}(\mathbf{I})$  for well separated  $l_j$  and simultaneously reduce the strong increase of the shrinkage as the  $l_j$  approach each other. The form of  $\phi_j^{\text{St}}(\mathbf{I})$  for well separated  $l_j$  is fixed by Stein's minimization theory, while the attenuation of the shrinkage effect is necessary to prove that the estimator is uniformly better than the MLE. Therefore, the estimators  $\widehat{\Sigma}^{(k)}$  cannot fully benefit of the isotonizing algorithm when the  $l_j$  are close to each other, so in the degenerate case ( $\Sigma \approx \mathbf{I}$ ) they yield smaller risk reductions compared to Stein's estimator. In conclusion, in the unbiased-estimator-of-risk approach, dominance over the MLE can only be obtained by sacrificing the considerable risk reductions given by the isotonizing algorithm in the degenerate case.

In our study, we only considered simple modifications of the  $\phi_j^{\text{St}}(\mathbf{I})$  (such as linear interpolations), because our purpose was to illustrate the unbiased-estimator-of-risk approach through some analytically solvable examples. More efficient estimators could be obtained by finding an optimal combination of the attenuation of the shrinking effect of the raw estimator and the use of the isotonizing algorithm. Furthermore, the same approach as used in the  $p = 2$  case can be used in higher dimension. In particular, properties 1–3 in Section 4 also hold in higher dimension (Rajaratnam and Vincenzi, submitted for publication) and can be used to modify the estimates  $\psi_j^{\text{St}}(\mathbf{I})$  and render the unbiased estimator of risk lower than that of the MLE. What makes this approach more tedious in higher dimension is the fact that, for large  $p$ , the unbiased estimator of risk is a function of many variables and its properties are more cumbersome to analyze compared to the  $p = 2$  case. Here, we do not specifically construct estimators in the  $p \geq 3$  case, as our goal was to show a proof of concept on how to use the unbiased-estimator-of-risk framework to retain the properties of Stein's estimator and still achieve dominance over the MLE. However, we remark that taking the  $p = 2$  approach is also useful from a practical point of view. The class of estimators we propose does not explicitly use the isotonizing algorithm in an ad hoc manner. In practice one can use the  $p = 2$  case to fix the unboundedness of any pairs of sample eigenvalues which are too close to one another, and do so without relying on the isotonizing algorithm in an ad hoc way. Doing so leads to a class of estimators that retain the original functional form of Stein's estimator in higher dimensions within a principled framework.

## Acknowledgments

The authors thank an anonymous referee for a very careful reading of the paper and for excellent suggestions. The authors also acknowledge the financial support of the France–Stanford Center for Interdisciplinary Studies. B.R. was partially supported by grant funding from: US Air Force Office of Scientific Research (AFOSR) (grant no. FA9550-13-1-0043), US National Science Foundation (NSF) (grant nos. DMS-0906392, DMS-CMG 1025465, AGS-1003823, DMS-1106642, DMS-CAREER-1352656), Defense Advanced Research Projects Agency (DARPA) (grant no. YFAN66001-11-1-4131), the UPS fund and SMC.

**Appendix A. Differentiability of Stein’s estimator**

The estimates  $\phi_j^{St}(\mathbf{I})$  are bounded functions of  $\mathbf{I}$  for all  $\mathbf{I} \in D_p$  and for all  $j = 1, \dots, p$ , because  $\phi_j^{St}(\mathbf{I}) \leq \phi_1^{St}(\mathbf{I})$  and  $\phi_1^{St}(\mathbf{I})$  has the form:

$$\phi_1^{St}(\mathbf{I}) = \frac{\sum_{j=1}^k l_j}{\sum_{j=1}^k \alpha_j(\mathbf{I})} \quad (1 \leq k \leq p), \tag{A.1}$$

where

$$\begin{aligned} \sum_{j=1}^k \alpha_j(\mathbf{I}) &= k(n - p + 1) + 2 \sum_{j=1}^k \left[ l_j \left( \sum_{i=1}^{j-1} \frac{1}{l_j - l_i} + \sum_{i=j+1}^k \frac{1}{l_j - l_i} \right) + l_j \sum_{i=k+1}^p \frac{1}{l_j - l_i} \right] \\ &= k(n - p + 1) + k(k - 1) + 2 \sum_{j=1}^k l_j \sum_{i=k+1}^p \frac{1}{l_j - l_i} > 0. \end{aligned} \tag{A.2}$$

The value of  $k$  in (A.1) is the number of sample eigenvalues that are pooled together to estimate  $\phi_1^{St}(\mathbf{I})$ .

The domain  $D_p$  can be partitioned into a finite number of subsets according to which  $\phi_j^{raw}(\mathbf{I})$  are pooled together in order to preserve the positivity and the ordering of the sample eigenvalues. This partition includes for example: (a) the set  $D \setminus Q_{p,n}$ , in which no  $\phi_j^{raw}$  are pooled together; (b) the subset of  $D_p$  in which  $\phi_p^{raw}(\mathbf{I})$  and  $\phi_{p-1}^{raw}(\mathbf{I})$  are pooled together, while the other  $\phi_j^{raw}(\mathbf{I})$  are left intact; (c) the subset in which all the  $\phi_j^{raw}(\mathbf{I}), j = 1, \dots, p$ , are pooled together, etc. Everywhere in the interior of each of the subsets of the partition, all the  $\phi_j^{St}$  are continuously differentiable. This property is a straightforward consequence of the definitions of  $\phi_j^{raw}$  and  $\phi_j^{iso}$ .

Consider now a generic subset of the aforementioned partition. Its boundary consists of the points  $\mathbf{I}$  such that there exist  $q, r, s \in \mathbb{N}, 1 \leq q \leq p, 0 \leq r \leq p - q - 1, 0 \leq s \leq p - q - r$  and:

$$\frac{l_q + \dots + l_{q+r}}{\alpha_q(\mathbf{I}) + \dots + \alpha_{q+r}(\mathbf{I})} = \frac{l_{q+r+1} \dots + l_{q+r+s}}{\alpha_{q+r+1}(\mathbf{I}) + \dots + \alpha_{q+r+s}(\mathbf{I})}. \tag{A.3}$$

The estimators  $\phi_j^{St}$  are continuous also on this boundary, because if  $\mathbf{I}$  belongs to it, the estimate obtained by pooling the left-hand and the right-hand sides of Eq. (A.3) is equal to the two original estimates:

$$\begin{aligned} \frac{l_q + \dots + l_{q+r+s}}{\alpha_q(\mathbf{I}) + \dots + \alpha_{q+r+s}(\mathbf{I})} &= \frac{\left[ \sum_{i=1}^s \alpha_{q+r+i}(\mathbf{I}) \right] \left( \sum_{j=0}^r l_{q+j} \right) + \left[ \sum_{i=1}^s \alpha_{q+r+i}(\mathbf{I}) \right] \left( \sum_{j=1}^s l_{q+r+j} \right)}{\left[ \sum_{i=1}^s \alpha_{q+r+i}(\mathbf{I}) \right] \left[ \sum_{j=0}^{r+s} \alpha_{q+j}(\mathbf{I}) \right]} \\ &= \frac{\left[ \sum_{i=0}^r \alpha_{q+i}(\mathbf{I}) \right] \left( \sum_{j=1}^s l_{q+r+j} \right) + \left[ \sum_{i=1}^s \alpha_{q+r+i}(\mathbf{I}) \right] \left( \sum_{j=1}^s l_{q+r+j} \right)}{\left[ \sum_{i=1}^s \alpha_{q+r+i}(\mathbf{I}) \right] \left[ \sum_{j=0}^{r+s} \alpha_{q+j}(\mathbf{I}) \right]} \\ &= \frac{l_{q+r+1} \dots + l_{q+r+s}}{\alpha_{q+r+1}(\mathbf{I}) + \dots + \alpha_{q+r+s}(\mathbf{I})} = \frac{l_q + \dots + l_{q+r}}{\alpha_q(\mathbf{I}) + \dots + \alpha_{q+r}(\mathbf{I})}. \end{aligned} \tag{A.4}$$

As the  $\phi_j^{St}$  are continuous everywhere on  $D_p$  and their partial derivatives are bounded on the interior part of each subset of the partition, the partial derivatives of the  $\phi_j^{St}$  may only have jump discontinuities on the boundaries of these subsets, which however form a measure-zero set. Therefore,  $\phi_j^{St}$  is continuously differentiable at least almost everywhere on  $D_p$  for all  $j = 1, \dots, p$ .

The following example shows that in fact there exist points in which the partial derivatives of  $\phi_j^{St}$  have jump discontinuities. Consider the set  $A \subset D_p$  in which all the  $\phi_j^{raw}, j = 1, \dots, p$ , are pooled together and the set  $B \subset D_p$  in which only  $\phi_2^{raw}(\mathbf{I}), \phi_3^{raw}(\mathbf{I}), \dots, \phi_p^{raw}(\mathbf{I})$  are pooled together, while the first eigenvalue is estimated as  $\phi_1^{St} = l_1/\alpha_1(\mathbf{I})$ . Then, the limit of  $\partial_2 \phi_1^{St}$  for  $\mathbf{I}$  approaching  $\partial A \cap \partial B$  from inside  $A$  is:  $1/(np) > 0$ . The limit of the same quantity for  $\mathbf{I}$  approaching  $\partial A \cap \partial B$  from inside  $B$  is:  $-2l_1^2/[\alpha_1^2(\mathbf{I})(l_1 - l_2)^2] < 0$ . Therefore, we conclude that the  $\phi_j^{St}$  are continuously differentiable only almost everywhere on  $D_p$ .

## Appendix B. General properties of the unbiased estimator of risk

**Proof of Proposition 1.** We prove the results in two cases.

Case I.  $\psi_j(\mathbf{I}) \sum_{i \neq j}^p \frac{1}{l_j - l_i} = O(1)$  as  $\mathbf{I} \rightarrow \tilde{\mathbf{I}}$ .

Then,  $\psi_j(\mathbf{I}) = O(\|\mathbf{I} - \tilde{\mathbf{I}}\|)$  as  $\mathbf{I} \rightarrow \tilde{\mathbf{I}}$ , whence  $2l_j \partial \psi_j / \partial l_j = O(1)$  as  $\mathbf{I} \rightarrow \tilde{\mathbf{I}}$  and  $\lim_{\mathbf{I} \rightarrow \tilde{\mathbf{I}}} \ln(\psi_j(\mathbf{I})) = -\infty$ . Hence, taking the limit of the sum of the terms that form  $\sigma_j(\mathbf{I})$  leads to its divergence. The proposition is thus proved by letting  $q = j$ .

Case II.  $\psi_j(\mathbf{I}) \sum_{i \neq j}^p \frac{1}{l_j - l_i} \neq O(1)$  as  $\mathbf{I} \rightarrow \tilde{\mathbf{I}}$ .

If case I does not hold, we can proceed by contradiction that  $\sigma_q$  is bounded. Since  $\psi_j(\mathbf{I}) > 0$ , by assumption the only alternative to case I is:

$$\limsup_{\mathbf{I} \rightarrow \tilde{\mathbf{I}}} \psi_j(\mathbf{I}) \sum_{i \neq j}^p \frac{1}{l_j - l_i} = +\infty. \quad (\text{B.1})$$

If we assume to the contrary that  $\sigma_q(\mathbf{I})$  is bounded for all  $q = 1, \dots, p$ , then (B.1) implies:

$$\liminf_{\mathbf{I} \rightarrow \tilde{\mathbf{I}}} [(n - p + 1)\psi_j(\mathbf{I}) + 2l_j \partial_j \psi_j(\mathbf{I}) - \ln(\psi_j(\mathbf{I}))] = -\infty. \quad (\text{B.2})$$

Since  $\psi_j(\mathbf{I})$  is positive for all  $\mathbf{I}$ , neither  $\psi_j(\mathbf{I})$  nor  $\partial_j \psi_j(\mathbf{I})$  can tend to  $-\infty$  as  $\mathbf{I} \rightarrow \tilde{\mathbf{I}}$ . Eq. (B.2) then implies:  $\limsup_{\mathbf{I} \rightarrow \tilde{\mathbf{I}}} \ln(\psi_j(\mathbf{I})) = +\infty$  or, equivalently,

$$\limsup_{\mathbf{I} \rightarrow \tilde{\mathbf{I}}} \psi_j(\mathbf{I}) = +\infty. \quad (\text{B.3})$$

We now observe that either  $\partial_j \psi_j(\mathbf{I})$  is bounded or it diverges positively as  $\mathbf{I} \rightarrow \tilde{\mathbf{I}}$ . If  $\partial_j \psi_j(\mathbf{I})$  is bounded, then  $\psi_j$  cannot satisfy (B.2) since (B.3) implies  $\limsup_{\mathbf{I} \rightarrow \tilde{\mathbf{I}}} \ln(\psi_j(\mathbf{I})) / \psi_j(\mathbf{I}) = 0$ . If  $\partial_j \psi_j(\mathbf{I})$  diverges positively, Eq. (B.3) and De L'Hôpital's rule imply

$$\limsup_{\mathbf{I} \rightarrow \tilde{\mathbf{I}}} \frac{\ln(\psi_j(\mathbf{I}))}{\psi_j(\mathbf{I}) + \partial_j \psi_j(\mathbf{I})} = 0,$$

and hence not even in this case can  $\psi_j$  satisfy (B.2). Consequently,  $\sigma_j(\mathbf{I})$  cannot be bounded for  $\mathbf{I} \rightarrow \tilde{\mathbf{I}}$ , and the proposition holds with  $q = j$ .

The second part of the proposition follows from the fact that  $F(\mathbf{I})$  can be written as:

$$F(\mathbf{I}) = \sum_{\substack{r=1 \\ r \neq j, k}}^p \left[ \left( n - p + 1 + 2l_r \sum_{i \neq r}^p \frac{1}{l_r - l_i} \right) \psi_r(\mathbf{I}) + 2l_r \frac{\partial \psi_r}{\partial l_r} - \ln \psi_r(\mathbf{I}) \right] \\ + (n - p + 1)[\psi_j(\mathbf{I}) + \psi_k(\mathbf{I})] + 2l_j \frac{\partial \psi_j}{\partial l_j} + 2l_k \frac{\partial \psi_k}{\partial l_k} - \ln(\psi_j(\mathbf{I})\psi_k(\mathbf{I})) + 2 \frac{l_j \psi_j(\mathbf{I}) - l_k \psi_k(\mathbf{I})}{l_j - l_k},$$

where all the terms are bounded in the limit  $\mathbf{I} \rightarrow \tilde{\mathbf{I}}$ .  $\square$

## References

- Daniels, M.J., Kass, R.E., 1999. Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *J. Amer. Statist. Assoc.* 94, 1254–1263.
- Daniels, M.J., Kass, R.E., 2001. Shrinkage estimators for covariance matrices. *Biometrics* 57, 1173–1184.
- Dey, D.K., Srinivasan, C., 1985. Estimation of a covariance matrix under Stein's loss. *Ann. Statist.* 13, 1581–1591.
- Eguchi, N., Saito, R., Saeki, T., Nakatsuka, Y., Belikov, D., Maksyutov, S., 2010. A priori covariance estimation for CO2 and CH4 retrievals. *J. Geophys. Res.* 115, D10215.
- Frei, M., Kunsch, H.R., 2012. Sequential state and observation noise covariance estimation using combined ensemble Kalman and particle filters. *Mon. Weather Rev.* 140, 1476–1495.
- Guillot, D., Rajaratnam, B., Emile-Geay, J., 2015. Statistical paleoclimate reconstructions via Markov random fields. *Ann. Appl. Stat.* 9, 324–352.
- Haff, L.R., 1980. Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.* 8, 586–597.
- Haff, L.R., 1991. The variational form of certain Bayes estimator. *Ann. Statist.* 19, 1163–1190.
- Hamimeche, S., Lewis, A., 2009. Properties and use of CMB power spectrum likelihoods. *Phys. Rev. D* 79, 083012.
- Hoff, P.D., 2009. A hierarchical eigenmodel for pooled covariance estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 71, 971–992.
- Ledoit, O., Wolf, M., 2004a. Honey, I Shrank the sample covariance matrix. *J. Portfolio Manage.* 30, 110–119.
- Ledoit, O., Wolf, M., 2004b. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.* 88, 365–411.
- Ledoit, O., Wolf, M., 2013. Optimal estimation of a large-dimensional covariance matrix under Stein's loss. in: Working Paper ECON 122, Department of Economics, University of Zurich.
- Lin, S., Perlman, M., 1985. A Monte-Carlo comparison of four estimators of a covariance matrix. In: Krishnaiah, P.R. (Ed.), *Multivariate Analysis VI*. North Holland, Amsterdam, pp. 411–429.
- Pope, A.C., Szapudi, I., 2008. Shrinkage estimation of the power spectrum covariance matrix. *Mon. Not. R. Astron. Soc.* 389, 766–774.
- Pourahmadi, M., 2011. Covariance estimation: The GLM and regularization perspectives. *Statist. Sci.* 26, 369–387.
- Rajaratnam, B., Massam, H., Carvalho, C.M., 2008. Flexible covariance estimation in graphical Gaussian models. *Ann. Statist.* 36, 2818–2849.
- Rajaratnam, B., Vincenzi, D., 2014. A theoretical study of Stein's covariance estimator, submitted for publication.

- Schäfer, J., Strimmer, K., 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* 4, 32.
- Stein, C., 1975. Estimation of a covariance matrix. In: *Riesz Lecture. 39th Annual Meeting. IMS, Atlanta, GA.*
- Stein, C., 1986. *Lectures on the theory of estimation of many parameters.* J. Math. Sci. 34, 1373–1403.
- Won, J.H., Lim, J., Kim, S.J., Rajaratnam, B., 2012. Condition-number-regularized covariance estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 75, 427–450.
- Yang, R., Berger, J.O., 1994. Estimation of a covariance matrix using the reference prior. *Ann. Statist.* 22, 1195–1211.