Biometrika (2016), **103**, 3, *pp.* 653–666
Printed in Great Britain
doi: 10.1093/biomet/asw030
Advance Access publication 22 August 2016

A theoretical study of Stein's covariance estimator

BY BALA RAJARATNAM

Department of Statistics, Sequoia Hall, 390 Serra Mall, Stanford University, Stanford, California 94305, U.S.A.
brajarat@stanford.edu

AND DARIO VINCENZI

Laboratoire J. A. Dieudonné, Université Nice Sophia Antipolis, CNRS, Parc Valrose, 06100 Nice, France
dario.vincenzi@unice.fr

SUMMARY

Stein proposed an estimator to address the poor performance of the sample covariance matrix for samples of small size. The estimator does not impose sparsity conditions and uses an isotonizing algorithm to preserve the order of the sample eigenvalues. Despite its superior numerical performance, its theoretical properties are not well understood. We demonstrate that Stein's covariance estimator gives modest risk reductions when it is not isotonized, and when it is isotonized the risk reductions are significant. Three broad regimes of the estimator's behaviour are identified.

Some key words: Covariance estimation; Eigenvalue; Shrinkage; Stein's estimator; Unbiased estimator of risk.

1. INTRODUCTION

The covariance matrix is a critical ingredient in statistical procedures such as principal components analysis and discriminant analysis (Ledoit & Wolf, 2004; Schäfer & Strimmer, 2005; Khare & Rajaratnam, 2011; Won et al., 2013). The sample covariance matrix, which is the maximum likelihood estimator, is a poor estimator unless the sample size n is much greater than the dimension p. However, in many modern applications this condition is not satisfied, so standard procedures can lead to poor estimates of the population covariance matrix. Obtaining good estimates is challenging because the number of entries in the population covariance matrix is of $O(p^2)$, and covariance matrices lie in the open convex cone of positive-definite matrices. Various estimators have been proposed; see Rajaratnam et al. (2008) and Pourhamadi (2011) for comprehensive summaries. Several of these estimators shrink sample quantities to a desired target; see Haff (1991), Daniels & Kass (2001), Ledoit & Wolf (2004), Khare & Rajaratnam (2011) or Won et al. (2013), for example.

Stein (1975, 1977, 1986) observed that the larger sample eigenvalues overestimate their population counterparts, whereas the smaller sample eigenvalues underestimate their population counterparts. To address this issue, he proposed shrinking the sample eigenvalues. He considered the class of orthogonally invariant estimators, which modify the sample eigenvalues but leave the sample eigenvectors untouched, and derived the unbiased estimator of risk for the covariance estimation problem. This approach allowed him to choose an optimal way to shrink the sample eigenvalues under the entropy loss function, but doing so can lead to negative eigenvalue estimates or a different ordering of values from the sample spectra. To mitigate these problems, Stein proposed an isotonizing algorithm. Although Stein's estimator requires $n > p$, it does not impose sparsity, which can be unrealistic in some applications. Stein's covariance estimator is considered a gold standard in the literature and numerical studies have demonstrated its superior risk properties (Lin & Perlman, 1985; Daniels & Kass, 2001; Ledoit & Wolf, 2004).

To the best of our knowledge, the theoretical properties of Stein's estimator have not been studied in detail because of the intractability of the isotonizing algorithm, which by itself has no formal statistical basis. Here we study the theoretical properties of Stein's estimator by using the unbiased estimator of risk approach that Stein proposed. The domain of the unbiased estimator of risk of Stein's estimator can be divided into three regimes according to its behaviour relative to the unbiased estimator of risk of the maximum likelihood estimator. The form of the population covariance matrix determines which part of the domain contributes most to the risk. Therefore, by combining knowledge of the values that the population covariance matrix and the unbiased estimator of risk take, it is possible to understand the behaviour of the risk in different parts of the parameter space. This understanding is useful in practice, since broad prior knowledge of covariance regimes is often available. Thus, our study can help a practitioner to understand when Stein's estimator is expected to perform better than the maximum likelihood estimator and what factors are driving any risk reductions.

2. PRELIMINARIES

Consider a sample $X_1, \ldots, X_n$ from a p-dimensional normal distribution $N_p(0, \Sigma)$, with $n \geqslant p$. The sample covariance matrix, up to a multiplicative constant, is $S = \sum_{i=1}^{n} X_i X_i^{\mathrm{T}}$ and can be written as $S = HLH^{\mathrm{T}}$, where $H^{-1} = H^{\mathrm{T}}$ and $L = \mathrm{diag}(l_1, \ldots, l_p)$ with $l_1 \geqslant \cdots \geqslant l_p > 0$. Stein (1975, 1977, 1986) considered orthogonally invariant estimators of Σ of the general form $\hat{\Sigma} = H\Phi(l)H^{\mathrm{T}}$, where $l = (l_1, \ldots, l_p)$ and $\Phi(l) = \mathrm{diag}\{\varphi_1(l), \ldots, \varphi_p(l)\}$. The maximum likelihood estimator S/n corresponds to $\hat{\varphi}_j^{\mathrm{ml}}(l) = l_j/n$. This estimator is known to be significantly biased upwards for larger eigenvalues and downwards for smaller eigenvalues when p/n is not small and when some or all of the l_j are close (Lin & Perlman, 1985). Stein proposed rectifying this problem by using improved eigenvalue estimates within orthogonally invariant estimators. Via a Wishart identity, he obtained the risk of an orthogonally invariant estimator $\hat{\Sigma}$ under the entropy loss function $L_1(\hat{\Sigma}, \Sigma) = \mathrm{tr}(\hat{\Sigma}\Sigma^{-1}) - \log\det(\hat{\Sigma}\Sigma^{-1}) - p$. This loss function is generally tractable, allows calculation of the unbiased estimator of risk, and has a form similar to the Gaussian likelihood. Under this loss function, Stein showed that the risk of $\hat{\Sigma}$ satisfies $R_1(\hat{\Sigma}, \Sigma) = E_\Sigma\{L_1(\hat{\Sigma}, \Sigma)\} = E_\Sigma\{F(l)\}$ with

$$F(l) = \sum_{j=1}^{p} \left\{ \left(n - p + 1 + 2l_j \sum_{i \neq j} \frac{1}{l_j - l_i} \right) \psi_j(l) + 2l_j \frac{\partial \psi_j}{\partial l_j} - \log \psi_j(l) \right\} - c_{p,n}, \tag{1}$$

where $\psi_j(l) = \phi_j(l)/l_j$ and

$$c_{p,n} = E\left(\sum_{j=1}^{p} \log \chi^2_{n-j+1} \right) + p = \sum_{j=1}^{p} \frac{\Gamma'\{(n-j+1)/2\}}{\Gamma\{(n-j+1)/2\}} + p\log 2 + p. \tag{2}$$

Stein observed that $F(l)$ is an unbiased estimator of the risk of $\hat{\Sigma}$. When the derivatives $\partial\psi_j/\partial l_j$ are disregarded, minimizing $F(l)$ with respect to the ψ_j $(j = 1, \ldots, p)$ yields

$$\hat{\psi}_j^{\mathrm{St}}(l) = 1/\alpha_j(l), \quad \alpha_j(l) = n - p + 1 + 2l_j \sum_{i \neq j} (l_j - l_i)^{-1} \quad (j = 1, \ldots, p).$$

Stein's covariance estimator is therefore given by $\hat{\varphi}_j^{\mathrm{St}}(l) = l_j/\alpha_j(l)$. Note that $\hat{\varphi}_j^{\mathrm{St}}(l) \approx \hat{\varphi}_j^{\mathrm{ml}}(l)$ for large n; in such settings the improvements that Stein's estimator can offer over the maximum likelihood estimator should be modest. Stein's estimator uses adjacent eigenvalue estimates to shrink the sample eigenvalue estimates closer together. That said, there are two problems with Stein's estimator: the $\hat{\varphi}_j^{\mathrm{St}}(l)$ can violate the ordering $l_1 \geqslant \cdots \geqslant l_p$ and can be negative. Stein proposed using an isotonizing algorithm that pools adjacent estimators to eliminate order and sign violations. The pooled estimate using $\hat{\varphi}_j^{\mathrm{St}}(l), \ldots, \hat{\varphi}_{j+s}^{\mathrm{St}}(l)$ is

$$\hat{\varphi}_j^{\mathrm{iso}}(l) = \hat{\varphi}_{j+1}^{\mathrm{iso}}(l) = \cdots = \hat{\varphi}_{j+s}^{\mathrm{iso}}(l) = (l_j + \cdots + l_{j+s})/\{\alpha_j(l) + \cdots + \alpha_{j+s}(l)\}.$$

Hence, Stein's isotonized estimator is defined as

$$\hat{\varphi}_j^{\mathrm{St+iso}}(l) = \begin{cases} \hat{\varphi}_j^{\mathrm{St}}(l), & l_1/\alpha_1(l) \geqslant \cdots \geqslant l_p/\alpha_p(l) > 0, \\ \hat{\varphi}_j^{\mathrm{iso}}(l), & \text{otherwise.} \end{cases}$$

We also define $\hat{\psi}_j^{\mathrm{St+iso}}(l) = \hat{\varphi}_j^{\mathrm{St+iso}}(l)/l_j$. When $n \geqslant p$, Stein's estimator performs well in comparison with other estimators and has established itself as a benchmark (Lin & Perlman, 1985; Daniels & Kass, 2001; Ledoit & Wolf, 2004). To distinguish between Stein's original estimator and the version supplemented by the isotonizing algorithm, we refer to the latter as Stein's isotonized estimator and the former as Stein's raw estimator, unless the context is clear.

3. The unbiased estimator of risk: basic properties

The unbiased estimator of risk of the maximum likelihood estimator, denoted by $F^{\mathrm{ml}}(l)$, is obtained by replacing $\psi_j(l)$ with $\hat{\varphi}_j^{\mathrm{ml}}(l)/l_j = 1/n$ in (1) and using $\sum_{j=1}^{p} \alpha_j(l) = np$:

$$F^{\mathrm{ml}}(l) = K_{p,n}^{\mathrm{ml}}, \quad K_{p,n}^{\mathrm{ml}} = p(1 + \log n) - c_{p,n}. \tag{3}$$

The domain of F^{ml} is $\mathscr{D}_p = \{l \in \mathbb{R}^p : l_1 \geqslant \cdots \geqslant l_p > 0\}$. Likewise, the unbiased estimator of risk of Stein's raw estimator, denoted by $F^{\mathrm{St}}(l)$, is obtained by substituting $\hat{\psi}_j^{\mathrm{St}}(l)$ into (1):

$$F^{\mathrm{St}}(l) = \sum_{j=1}^{p} \left\{ 1 + \frac{4l_j}{\alpha_j^2(l)} \sum_{i \neq j} \frac{l_i}{(l_j - l_i)^2} + \log \alpha_j(l) \right\} - c_{p,n}.$$

Unlike F^{ml}, F^{St} is defined only on the set $\mathscr{E}_{p,n} = \{l \in \mathscr{D}_p : l_i \neq l_j \ (i, j = 1, \ldots, p),\ \alpha_j(l) > 0 \ (j = 1, \ldots, p)\}$, which can be expressed as $\mathscr{E}_{p,n} = \bigcap_{j=2}^{p} \{l \in \mathscr{D}_p : \alpha_j(l) > 0\}$, since if $\tilde{l} \in \mathscr{D}_p$ is such that $\tilde{l}_k = \tilde{l}_j$ for some $1 \leqslant k < j \leqslant p$, some of the $\alpha_j(l)$ must be negative in a neighbourhood

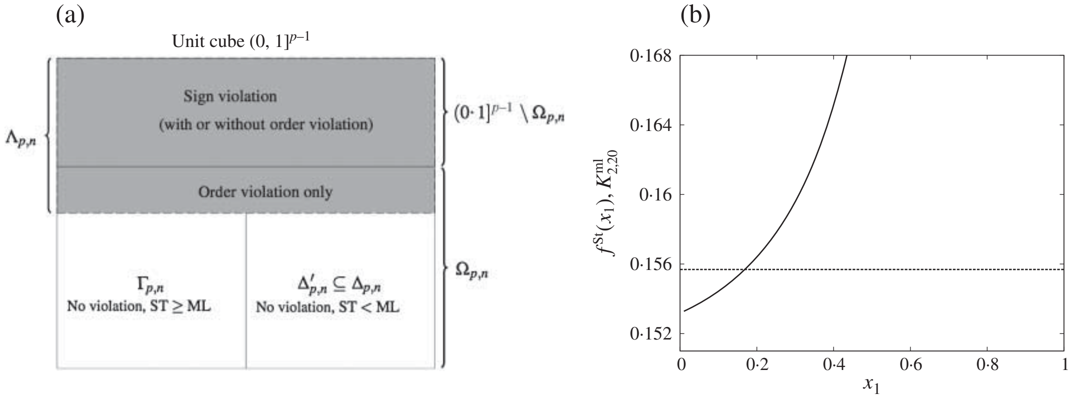

Fig. 1. (a) Schematic representation of the sets $\Omega_{p,n}$, $\Delta_{p,n}$, $\Delta'_{p,n}$, $\Gamma_{p,n}$ and $\Lambda_{p,n}$; (b) Graph of the unbiased estimator of risk for Stein's raw estimator (solid) and the maximum likelihood estimator (dotted) for $p = 2$ and $n = 20$.

of $\tilde{l}$; in addition, the condition $\alpha_1(l) > 0$ is not imposed because it is satisfied for all $l \in \mathscr{D}_p$. As the domain of F^{St} is not the whole of $\mathscr{D}_p$, the risk of Stein's raw estimator is either infinite or complex-valued. It is useful to rewrite $F^{\mathrm{St}}(l)$ as a function of the $p-1$ ratios of the adjacent eigenvalues of S, i.e.,

$$F^{\mathrm{St}}(l) = f^{\mathrm{St}}(x) = \sum_{j=1}^{p}\left[1 + \frac{4}{a_j^2(x)}\sum_{i=1}^{j-1}\frac{\pi_i^j(x)}{\{\pi_i^j(x)-1\}^2} + \frac{4}{a_j^2(x)}\sum_{i=j+1}^{p}\frac{\pi_j^i(x)}{\{1-\pi_j^i(x)\}^2} + \log a_j(x)\right] - c_{p,n}, \tag{4}$$

where $x = (x_1, \ldots, x_{p-1})$ with $x_j = l_{j+1}/l_j$,

$$a_j(x) = n - p + 1 + 2\sum_{i=1}^{j-1}\frac{\pi_i^j(x)}{\pi_i^j(x) - 1} + 2\sum_{i=j+1}^{p}\frac{1}{1-\pi_j^i(x)} \quad (j = 1, \ldots, p) \tag{5}$$

and $\pi_i^j(x) = \prod_{k=i}^{j-1} x_k$ $(1 \leqslant i < j \leqslant p)$; by construction $\pi_i^j(x) \leqslant 1$ for all $1 \leqslant i < j \leqslant p$. The function f^{St} is defined on the following subset of the $(p-1)$-dimensional unit cube: $\Omega_{p,n} = \bigcap_{j=2}^{p}\{x \in (0,1)^{p-1} : a_j(x) > 0\}$. In terms of the variables x, the requirement that the order of the l_j be preserved by Stein's raw estimator is written as $x_j \leqslant a_{j+1}(x)/a_j(x)$ $(j = 1, \ldots, p-1)$.

In the rest of the paper, we shall also consider the following subsets of $(0, 1]^{p-1}$; see Fig. 1(a):

$$\Delta_{p,n} = \left\{x \in \Omega_{p,n} : f^{\mathrm{St}}(x) < K^{\mathrm{ml}}_{p,n}\right\}, \tag{6}$$

$$\Delta'_{p,n} = \left\{x \in \Omega_{p,n} : f^{\mathrm{St}}(x) < K^{\mathrm{ml}}_{p,n},\ x_j \leqslant a_{j+1}(x)/a_j(x)\ (j = 1, \ldots, p-1)\right\}, \tag{7}$$

$$\Gamma_{p,n} = \{x \in \Omega_{p,n} : x_j \leqslant a_{j+1}(x)/a_j(x)\ (j = 1, \ldots, p-1),\ f^{\mathrm{St}}(x) \geqslant K^{\mathrm{ml}}_{p,n}\}, \tag{8}$$

$$\Lambda_{p,n} = (0, 1]^{p-1} \setminus (\Delta'_{p,n} \cup \Gamma_{p,n}). \tag{9}$$

4. The unbiased estimator of risk of Stein's raw estimator

4·1. *The case of $p = 2$*

We first study the cases $p = 2, 3, 4$ to gain intuition about the general case, and then we will extend the results to arbitrary p. For $p = 2$, f^{St} is a function of the variable x_1 representing the ratio l_2/l_1:

$$f^{\mathrm{St}}(x_1) = 2 + 4x_1\left[\{n+1-(n-1)x_1\}^{-2} + \{(n+1)x_1-(n-1)\}^{-2}\right] + \log\left\{n + (1+x_1)/(1-x_1)\right\} + \log\left\{n - (1+x_1)/(1-x_1)\right\} - c_{2,n}.$$

The domain of f^{St} is $\Omega_{2,n} = (0, \tilde{x}_1)$ with $\tilde{x}_1 = (n-1)/(n+1)$; hence $\mathrm{vol}(\Omega_{2,n}) = \tilde{x}_1$. Although $\Omega_{p,n}$ does not contain $x = 0$, f^{St} can be defined at $x = 0$ and, to simplify the notation, we shall write $f^{\mathrm{St}}(0)$ instead of $\lim_{\|x\|\to 0} f^{\mathrm{St}}(x)$. The graph of f^{St} is shown in Fig. 1(b) for $n = 20$. To understand the behaviour of f^{St} for an arbitrary n, we note that $f^{\mathrm{St}}(0) = 2 + \log(n^2 - 1) - c_{2,n} < K^{\mathrm{ml}}_{2,n}$, that f^{St} is monotonically increasing, and that $f^{\mathrm{St}}(x_1) = O\{(x_1 - \tilde{x}_1)^{-2}\}$ as $x_1 \to \tilde{x}_1^-$; see the Supplementary Material. As a consequence, there exists one and only one $x_1^\star \in \Omega_{2,n}$ such that $f^{\mathrm{St}}(x_1^\star) = K^{\mathrm{ml}}_{2,n}$; hence $\Delta_{2,n} = (0, x_1^\star)$. Figure 2(b) shows that as $n \to \infty$, $x_1^\star$ converges rapidly to the asymptotic value $3 - 2\surd 2$, which can be obtained by considering the large-n expansion $f^{\mathrm{St}}(x_1) - K^{\mathrm{ml}}_{2,n} = -(x_1^2 - 6x_1 + 1)(x_1 - 1)^{-2} n^{-2} + O(n^{-4})$. This expansion also shows that as $n \to \infty$, the convergence of $f^{\mathrm{St}}(x_1)$ to $K^{\mathrm{ml}}_{2,n}$ is not uniform in x_1, since $\lim_{x_1 \to \tilde{x}_1^-}\{f^{\mathrm{St}}(x_1) - K^{\mathrm{ml}}_{2,n}\} = \infty$ for all $n \geqslant 2$.

As $n \to \infty$, the domain $\Omega_{2,n}$ tends to the whole interval $(0, 1)$. In contrast, $\mathrm{vol}(\Delta_{2,n}) = x_1^\star$ tends to a value much less than 1; therefore, the region on which $f^{\mathrm{St}}(x_1) < K^{\mathrm{ml}}_{2,n}$ is relatively small. The length of $\Delta_{2,n}$ is much less than 1 also for n approaching $p = 2$; see Fig. 2(b). In addition, the ratio $\{K^{\mathrm{ml}}_{2,n} - f^{\mathrm{St}}(0)\}/K^{\mathrm{ml}}_{2,n}$ is a decreasing function of n and is at its maximum when $n = p = 2$, though the reduction is not very large; see Fig. 2(c). This analysis suggests that two competing phenomena are manifested as n increases. First, $f^{\mathrm{St}}(0) - K^{\mathrm{ml}}_{2,n}$ decreases. Second, the size of $\Delta_{2,n}$ increases. Numerical work in the Supplementary Material shows that relative risk reductions compared to the maximum likelihood estimator are higher for smaller n, indicating that the former phenomenon outweighs the latter. Therefore, neither the function values nor the volume of $\Delta_{p,n}$ should be analysed in isolation when studying Stein's estimator in the unbiased estimator of risk framework.

4·2. *The cases of $p = 3$ and $p = 4$*

The graph of $f^{\mathrm{St}}(x)$ defined in (4) for $p = 3$ and $n = 20$ is shown in Fig. 3(a). The domain $\Omega_{3,n}$ is described by the inequalities $a_2(x) > 0$ and $a_3(x) > 0$ and is the union of the sets I and II in Fig. 3(b). In that figure panel, $\Delta_{3,n}$ is denoted by I; it is a connected region containing a neighbourhood of the origin in the first quadrant. The function f^{St} is not defined on set III. In panels (a) and (b) of Fig. 2 the volumes of $\Omega_{3,n}$ and $\Delta_{3,n}$ are plotted as functions of n; the behaviour is similar to that observed for $p = 2$. The function f^{St} tends to infinity as $a_2(x)$ or $a_3(x)$ vanishes, but, unlike in the $p = 2$ case, f^{St} is not monotonic in x_2 for all fixed x_1. Therefore a simple characterization of $\Omega_{3,n}$ and $\Delta_{3,n}$ is not available. In addition, f^{St} is neither convex nor concave; see the Supplementary Material. The percentage reduction in $f^{\mathrm{St}}(0)$ relative to the maximum likelihood estimator decreases approximately linearly with increasing n, as in the $p = 2$ case; see Fig. 2(c). Moreover, for all $x \in \Omega_{p,n}$, $f^{\mathrm{St}}(x) - K^{\mathrm{ml}}_{p,n}$ is $O(1/n^2)$ as $n \to \infty$. This property will be derived for arbitrary p in § 4·4.

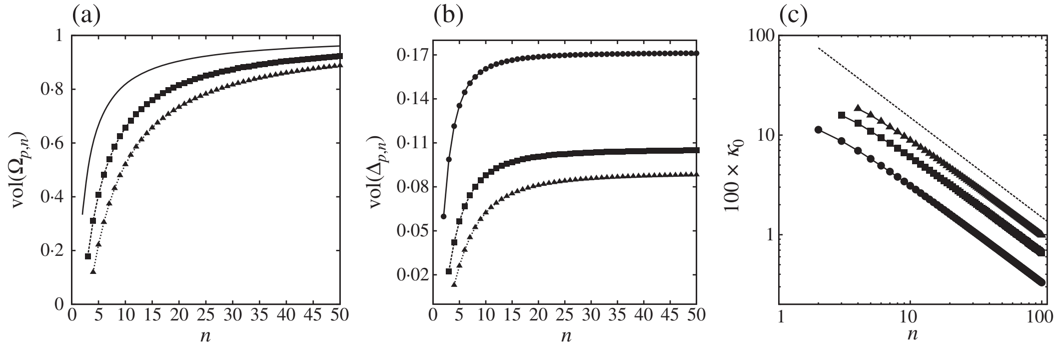

Fig. 2. (a) Volume of $\Omega_{p,n}$, (b) volume of $\Delta_{p,n}$, and (c) $100 \times \kappa_0 = 100 \times [K^{\mathrm{ml}}_{p,n} - f^{\mathrm{St}}(0)]/K^{\mathrm{ml}}_{p,n}$ plotted as a function of n for $p = 2$ (●) and $p = 3$ (■), $p = 4$ (▲). In (c), the vertical and horizontal axes are both scaled logarithmically, and the dashed straight line is proportional to $1/n$.

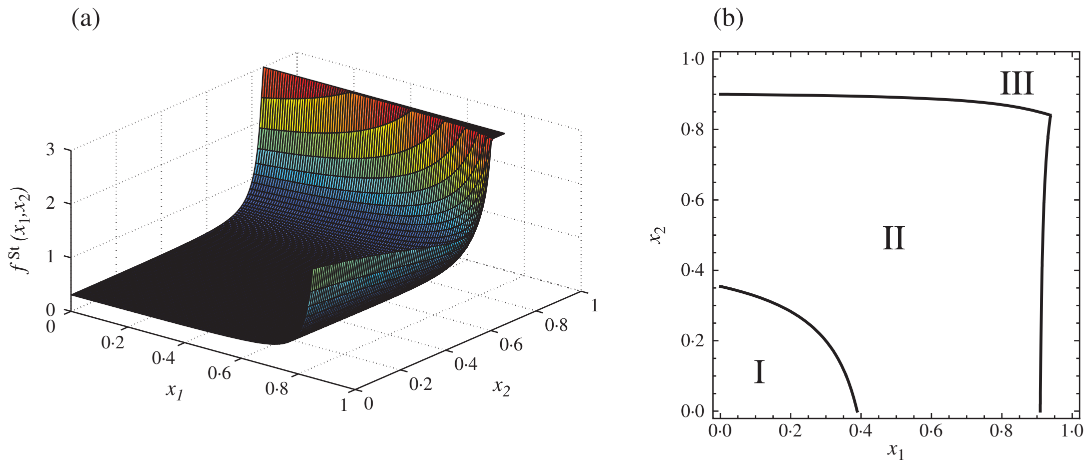

Fig. 3. (a) Graph of f^{St} as a function of (x_1, x_2) for $p = 3$ and $n = 20$. (b) Sets I, II and III in the (x_1, x_2) plane: set I, where $f^{\mathrm{St}}(x) < K^{\mathrm{ml}}_{2,n}$; set II, where $f^{\mathrm{St}}(x) > K^{\mathrm{ml}}_{2,n}$; set III, where f^{St} is not defined.

Results for the $p = 4$ case are similar to those for the $p = 3$ case; see the Supplementary Material for details.

4·3. *Discussion*

The comparison between the unbiased estimator of risk of Stein's raw estimator and that of the maximum likelihood estimator for $p = 2, 3, 4$ is summarized below.

First, f^{St} is not defined for all x, as it contains logarithmic terms whose arguments can be negative in an open subset of its domain. Furthermore, f^{St} diverges as the argument of one of the logarithmic terms vanishes. The risk of Stein's raw estimator is thus either infinite or complex-valued, and the estimator needs to be rectified via the isotonizing algorithm; see § 5. In addition, we need to characterize the region over which f^{St} diverges to assess the severity of the problem.

Second, when the unbiased estimator of risk of Stein's raw estimator is rewritten in terms of the ratios of adjacent eigenvalues of S, its domain $\Omega_{p,n}$ is a strict subset of $(0, 1)^{p-1}$. For a given p, the volume of $\Omega_{p,n}$ increases with increasing n and tends to 1 as $n \to \infty$; it decreases with increasing p when n is held constant; see Fig. 2(a). Hence, the problem of f^{St} not being defined is alleviated as n increases, but is exacerbated as p increases.

Third, $f^{\mathrm{St}}(x) < K^{\mathrm{ml}}_{p,n}$ only when the adjacent eigenvalues of S are well separated. The set $\Delta_{p,n}$ is connected and contains a neighbourhood of $x = 0$. The volume of $\Delta_{p,n}$ increases with n and tends to an asymptote below 1; for a given n, it decreases as p increases; see Fig. 2(b).

Fourth, although $f^{\mathrm{St}}(x) < K^{\mathrm{ml}}_{p,n}$ in the vicinity of $x = 0$, the percentage reduction in $f^{\mathrm{St}}(0)$ relative to the maximum likelihood estimator is modest and decreases linearly with increasing n; see Fig. 2(c). Hence Stein's estimator is not expected to yield substantive gains over the maximum likelihood estimator when either the population eigenvalues are well separated or n is very large. The latter is consistent with the fact that Stein's estimator tends to the maximum likelihood estimator as $n \to \infty$. Moreover, despite diminishing reductions in $f^{\mathrm{St}}(x)$ as n increases, there is a dimension effect; the modest reduction in $f^{\mathrm{St}}(0)$ for fixed n is higher for larger p; see Fig. 2(c).

Fifth, for $n \to \infty$, $f^{\mathrm{St}}(x) - K^{\mathrm{ml}}_{p,n} = O(1/n^2)$ for all $x \in \Omega_{p,n}$, and the convergence is not uniform in x. The leading order of the unbiased estimator of risk of the maximum likelihood estimator is $O(1/n)$; see the Supplementary Material. Therefore, in an asymptotic sense the difference in the unbiased estimator of risk of the two estimators is small, which points to only modest risk reductions when using Stein's estimator instead of the maximum likelihood estimator. We shall see in § 5·3 that this is not the case when the isotonizing algorithm is invoked.

Finally, f^{St} is not convex in general; nor is it monotonic in each of the variables when the other variables are held constant, except in the $p = 2$ case; see § 4·1. The lack of convexity and monotonicity when $p > 2$ implies that the set $\Delta_{p,n}$ is not easily characterizable.

4·4. *The case of arbitrary p*

Several observations reported in § 4·3 for small p can be proved for any $p > 1$. The proofs of the lemmas and propositions are provided in the Supplementary Material.

We have noted that f^{St} may have singularities when $a_j(x)$ vanishes. For $j \geqslant 2$, the last sum in (5) is negative, whereas the remaining terms are positive. Hence, if one of the $\pi_i^j(x)$ for $i < j$ is sufficiently close to 1, then $a_j(x)$ may vanish or become negative. The following lemma shows that there exists some $x \in (0, 1)^{p-1}$ such that at least one of the $a_j(x)$ is equal to zero.

Lemma 1. *For all $n \geqslant p > 1$, there exist $\tilde{x} \in (0, 1)^{p-1}$ and $M \subset \{2, \ldots, p\}$ with $|M| > 1$ such that $a_j(\tilde{x}) = 0$ for all $j \in M$ and $a_j(\tilde{x}) > 0$ for $j \notin M$.*

If $\tilde{x}$ is as in Lemma 1, then it belongs to the boundary of $\Omega_{p,n}$ and is an accumulation point of $\Omega_{p,n}$, owing to the continuity of the a_j. The following proposition asserts that f^{St} becomes unbounded as x approaches one of the points where at least one of the a_j vanishes.

Proposition 1. *Let $n \geqslant p > 1$ and $M \subseteq \{2, \ldots, p\}$. If $\tilde{x} \in (0, 1)^{p-1}$ is such that $a_j(\tilde{x}) = 0$ for all $j \in M$ and $a_j(\tilde{x}) > 0$ for all $j \notin M$, then* $\lim_{x \to \tilde{x}} f^{\mathrm{St}}(x) = +\infty$.

To obtain the estimator $\hat{\psi}_j^{\mathrm{St}}$, Stein disregarded the terms $2l_j \partial\psi_j/\partial l_j$ in the unbiased estimator of risk. The proof of Proposition 1 demonstrates that these derivative terms determine the behaviour of f^{St} near the boundary of its domain and are therefore not negligible.

The large-n behaviour of f^{St} is described by the following proposition.

PROPOSITION 2. *For any $p > 1$ and $x \in \Omega_{p,n}$,*

$$f^{\mathrm{St}}(x) = K^{\mathrm{ml}}_{p,n} + \left(\frac{1}{n^2}\right) \sum_{j=1}^{p} \left[4 \sum_{i=1}^{j-1} \frac{\pi_i^j(x)}{\{\pi_i^j(x) - 1\}^2} + 4 \sum_{i=j+1}^{p} \frac{\pi_j^i(x)}{\{1 - \pi_j^i(x)\}^2} - \frac{1}{2} \left\{ 1 - p + 2 \sum_{i=1}^{j-1} \frac{\pi_i^j(x)}{\pi_i^j(x) - 1} + 2 \sum_{i=j+1}^{p} \frac{1}{1 - \pi_j^i(x)} \right\}^2 \right] + O\left(\frac{1}{n^3}\right) \quad (10)$$

as $n \to \infty$. Furthermore, the convergence is not uniform in x.

The analysis in the cases $p = 2, 3, 4$ shows that when the ratios l_{j+1}/l_j are sufficiently small, i.e., when $\|x\|$ is close to 0, $f^{\mathrm{St}}(x) < K^{\mathrm{ml}}_{p,n}$. The next proposition states that this is true for arbitrary p.

PROPOSITION 3. *For any $n \geqslant p > 1$, there exists an open ball $B_\delta(0)$ with radius $\delta > 0$ centred at $x = 0$ such that for all $x \in B_\delta(0) \cap \Omega_{p,n}$, $f^{\mathrm{St}}(x) < K^{\mathrm{ml}}_{p,n}$ and the order of the sample eigenvalues is preserved by Stein's raw estimator. Moreover, for a given n, $K^{\mathrm{ml}}_{p,n} - f^{\mathrm{St}}(0)$ is a monotonically increasing function of p for $1 < p \leqslant n$.*

Finally, the lemma below describes the large-n behaviour of the percentage reduction in $f^{\mathrm{St}}(0)$ relative to the maximum likelihood estimator and asserts that it is most pronounced for small n.

LEMMA 2. *Consider $K^{\mathrm{ml}}_{p,n}$ and $f^{\mathrm{St}}(x)$ as defined in (3) and (4), respectively. Then, for $p > 1$ and as $n \to \infty$, $\kappa_0 = \{K^{\mathrm{ml}}_{p,n} - f^{\mathrm{St}}(0)\}/K^{\mathrm{ml}}_{p,n} \sim (p-1)/3n$.*

The analysis conducted earlier in this section showed that the domain of f^{St} diminishes as p increases; see Fig. 2(a). However, Proposition 3 and Lemma 2 demonstrate that, for large n, the behaviour of f^{St} at $x = 0$ improves with increasing p. Therefore, studying only the local behaviour of f^{St} can lead to misleading conclusions regarding the risk gains of Stein's estimator, which does not perform uniformly well in all parts of the parameter space. Hence the risk gains seen in previous numerical work should be interpreted in this context, i.e., global statements have been inferred regarding the performance of the estimator based on risk gains for specific parameter values. Furthermore, if prior information is known on how well separated the eigenvalues are, Stein's raw estimator can be used effectively to yield risk gains.

It is useful to understand the reduction in $f^{\mathrm{St}}(0)$ when p grows and $n = p$. Further calculations show that in this regime the relative risk reduction at $x = 0$ is bounded as $p \to \infty$.

5. THE UNBIASED ESTIMATOR OF RISK OF STEIN'S ISOTONIZED ESTIMATOR

5·1. *The case of $p = 2$*

In practice, Stein's raw estimator performs very well when coupled with the isotonizing algorithm. Therefore, a natural question concerns the role played by this algorithm in the risk reductions enjoyed by Stein's estimator.

Recall definitions (6)–(9). For $p = 2$, the set $\Lambda_{2,n}$ consists of the points $x_1 \in (0, 1]$ such that $(n-1)x_1^2 - 2(n+1)x_1 + (n-1) < 0$, order violation, or $x_1 \geqslant \tilde{x}_1$, violation of positivity, where $\tilde{x}_1$ has been defined in § 4·1. Thus, $\Lambda_{2,n} = (x_1^{\star\star}, 1]$ with $x_1^{\star\star} = (n + 1 - 2\surd n)/(n-1)$. Note that

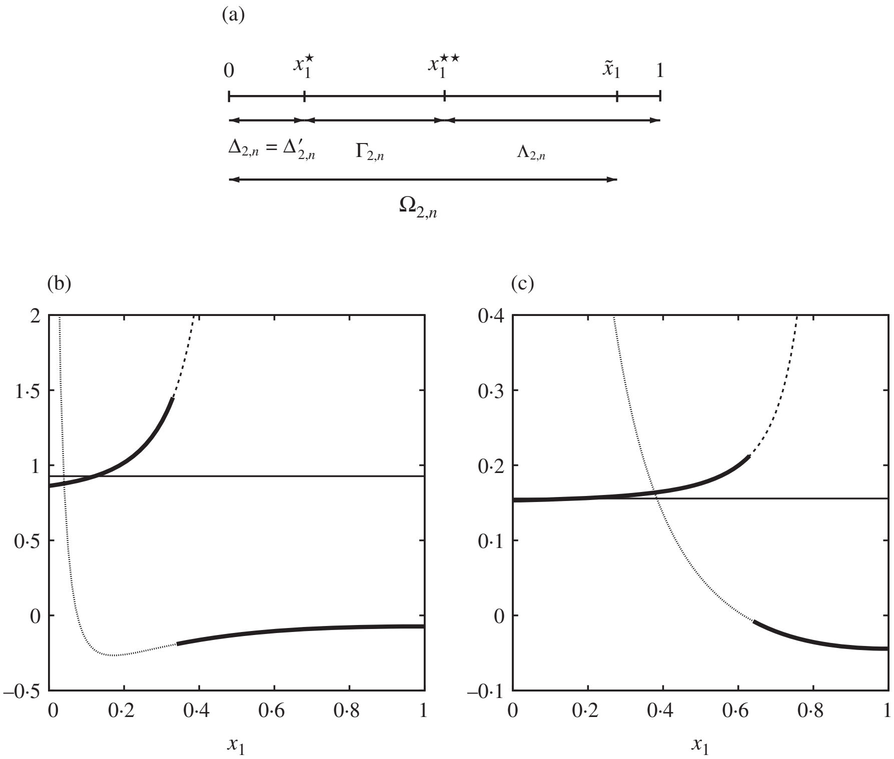

Fig. 4. (a) Schematic representation of the $p = 2$ case; see § 3 for the definitions of $\Omega_{2,n}$, $\Delta_{2,n}$, $\Gamma_{2,n}$ and $\Lambda_{2,n}$. (b, c) Graphs of f^{St} (dashed), f^{iso} (dotted), $f^{\mathrm{St+iso}}$ (thick solid) and $K_{2,n}^{\mathrm{ml}}$ (thin solid flat line) for (b) $n = 4$ and (c) $n = 20$; a thick solid line is superimposed on the graphs of f^{St} and f^{iso} when they are the same as the graph of $f^{\mathrm{St+iso}}$.

$x_1^\star < x_1^{\star\star} < \tilde{x}_1$ for all $n \geqslant 2$, where $x_1^\star$ has been defined in § 4·1. Therefore, $\Gamma_{2,n} = [x_1^\star, x_1^{\star\star}]$ and the set on which $f^{\mathrm{St}}(x_1) > K_{2,n}^{\mathrm{ml}}$ but where the isotonizing correction does not apply is nonempty. The inequality $x_1^\star < x_1^{\star\star} < \tilde{x}_1$ for all $n \geqslant 2$ also implies that $\Delta'_{2,n} = \Delta_{2,n}$ for all $n \geqslant 2$; hence the order of the sample eigenvalues is never violated when $f^{\mathrm{St}}(x_1) < K_{2,n}^{\mathrm{ml}}$. A schematic description of the $p = 2$ case is given in Fig. 4(a).

The isotonized estimator thus takes the form $\hat{\varphi}_j^{\mathrm{St+iso}}(l) = l_j \hat{\psi}_j^{\mathrm{St+iso}}(l_2/l_1)$, with

$$\hat{\psi}_1^{\mathrm{St+iso}}(x_1) = \begin{cases} (1 - x_1)/\{n + 1 - (n - 1)x_1\}, & 0 < x_1 \leqslant x_1^{\star\star}, \\ (1 + x_1)/(2n), & x_1^{\star\star} < x_1 \leqslant 1, \end{cases}$$

and

$$\hat{\psi}_2^{\mathrm{St+iso}}(x_1) = \begin{cases} (1 - x_1)/\{n - 1 - (n + 1)x_1\}, & 0 < x_1 \leqslant x_1^{\star\star}, \\ (1 + x_1)/(2nx_1), & x_1^{\star\star} < x_1 \leqslant 1. \end{cases}$$

Correspondingly, the unbiased estimator of risk of Stein's isotonized estimator is

$$f^{\mathrm{St+iso}}(x_1) = \begin{cases} f^{\mathrm{St}}(x_1), & 0 < x_1 \leqslant x_1^{\star\star}, \\ f^{\mathrm{iso}}(x_1), & x_1^{\star\star} < x_1 \leqslant 1, \end{cases}$$

where $f^{\text{iso}}(x_1)$ is the unbiased estimator of risk when the isotonizing algorithm is used:

$$f^{\text{iso}}(x_1) = 1 - \frac{1}{n} + \frac{x_1}{2} - \frac{3x_1}{2n} + \frac{n-3}{2nx_1} - \log\left(\frac{1+x_1}{2n}\right) - \log\left(\frac{1+x_1}{2nx_1}\right) - c_{2,n}.$$

The function f^{iso} is well-defined on (0, 1] regardless of its use within Stein's estimator; see Figs. 4(b) and (c). It is easily shown that $f^{\text{iso}}(1) = 2(1 - 2/n + \log n) - c_{2,n} < K^{\text{ml}}_{2,n}$. Furthermore, for $n = 2, 3$, $f^{\text{iso}}(x_1) < K^{\text{ml}}_{2,n}$ for all $x_1 \in (0, 1]$, while for $n \geqslant 4$ there exists one and only one $\bar{x}_1 \in (0, 1)$ such that $f^{\text{iso}}(x_1) > K^{\text{ml}}_{2,n}$ for $0 < x_1 < \bar{x}_1$ and $f^{\text{iso}}(x_1) < K^{\text{ml}}_{2,n}$ for $\bar{x}_1 < x_1 \leqslant 1$.

The study of $f^{\text{St+iso}}$ for $p = 2$ thus yields the following insights. First, there exists an interval on which $f^{\text{St+iso}}(x_1) > K^{\text{ml}}_{2,n}$. Second, the isotonizing algorithm is invoked on a large subset of (0, 1] containing the region where x_1 is near 1. When the isotonizing algorithm applies, the reduction in the unbiased estimator of risk compared to the maximum likelihood estimator is substantial and much greater than when isotonizing is not used; see Figs. 4(b) and (c). This observation can be made more precise by considering the quantities

$$\kappa_1 = \frac{K^{\text{ml}}_{p,n} - f^{\text{St+iso}}(1, \ldots, 1)}{K^{\text{ml}}_{p,n} - f^{\text{St}}(0, \ldots, 0)}, \quad \kappa_2 = \frac{K^{\text{ml}}_{p,n} - f^{\text{St+iso}}(1, \ldots, 1)}{K^{\text{ml}}_{p,n}}. \tag{11}$$

The first quantity compares the difference between $K^{\text{ml}}_{2,n}$ and $f^{\text{St}}(x)$ at the origin with the same difference at $x = (1, \ldots, 1)$. The ratio κ_1, therefore, is one simple way to quantify the relative efficacy of the isotonizing algorithm in reducing the unbiased estimator of risk in comparison with Stein's raw estimator. The second quantity is the relative reduction in the unbiased estimator of risk at $x = (1, \ldots, 1)$, where the reduction is entirely due to the isotonizing algorithm and is analogous to κ_0; see Lemma 2. For $p = 2$, $\kappa_1 = 4[n \log\{n^2/(n^2 - 1)\}]^{-1} \sim 4n$ and $\kappa_2 = 4n^{-1}\{2(1 + \log n) - c_{2,n}\}^{-1} \sim 4/3 - 26/(27n)$ as $n \to \infty$. These expressions affirm the fact that the relative reduction in the unbiased estimator of risk afforded by the isotonizing algorithm is nonnegligible and is a complex function of n.

Third, there exist points $x_1 < x_1^{\star\star}$ such that $f^{\text{iso}}(x_1) < K^{\text{ml}}_{2,n}$ but $f^{\text{St+iso}}(x_1) = f^{\text{St}}(x_1) > K^{\text{ml}}_{2,n}$. On this set, although an isotonizing correction is not required, using it would reduce the unbiased estimator of risk. In other words, the shrinkage given by the isotonizing correction is useful even before its purpose of order preservation becomes relevant. These insights may be useful for constructing new estimators which can combine the strengths of Stein's raw estimator and the isotonizing algorithm (Rajaratnam & Vincenzi, 2016).

Fourth, Fig. 4(b) and (c) show three regimes in the behaviour of $f^{\text{St+iso}}$ in comparison to $K^{\text{ml}}_{2,n}$. The first is near $x_1 = 0$, where the eigenvalues are well separated and $f^{\text{St+iso}}(x_1) = f^{\text{St}}(x_1) < K^{\text{ml}}_{2,n}$, though the reduction is small. In the second regime, the eigenvalues are moderately separated and $f^{\text{St+iso}}(x_1) = f^{\text{St}}(x_1) > K^{\text{ml}}_{2,n}$. In the third, the eigenvalues are close and the isotonizing algorithm is invoked due to sign or order violations. In this regime, $f^{\text{St+iso}}(x_1)$ is much lower than $K^{\text{ml}}_{p,n}$.

5·2. *The cases of $p = 3$ and $p = 4$*

The form of $\hat{\psi}^{\text{St+iso}}_j(x)$ for $p = 3$ is described in the Supplementary Material; only the main properties of $f^{\text{St+iso}}$ are presented here. The domain of $f^{\text{St+iso}}$ is $(0, 1] \times (0, 1]$ and can be divided into four subsets, A^{I}, A^{II}, A^{III} and A^{IV}, on each of which Stein's isotonized estimator takes a different form. In A^{I}, all eigenvalue estimates are positive and the order is preserved, so

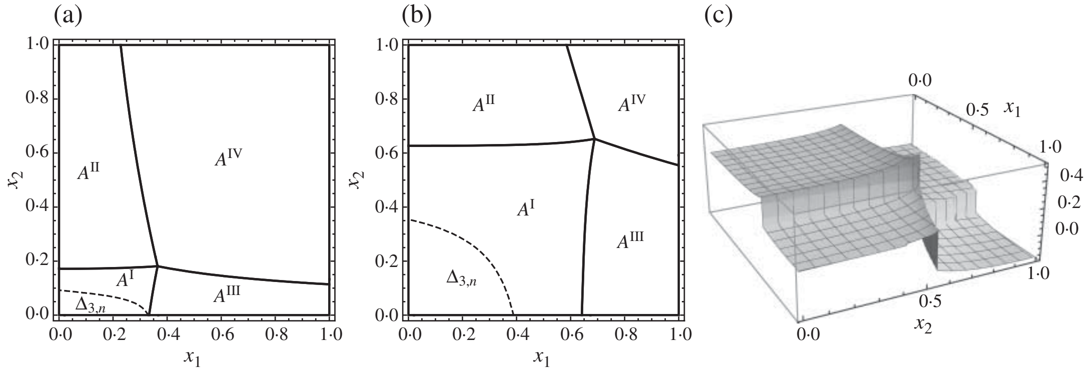

Fig. 5. The subsets A^{IV}, A^{II}, A^{III} and A^{IV} for (a) $n = 3$ and (b) $n = 20$. (c) Graph of $f^{\mathrm{St+iso}}$ for $p = 3$ and $n = 20$; note that $K^{\mathrm{ml}}_{3,20} = 0{\cdot}318$.

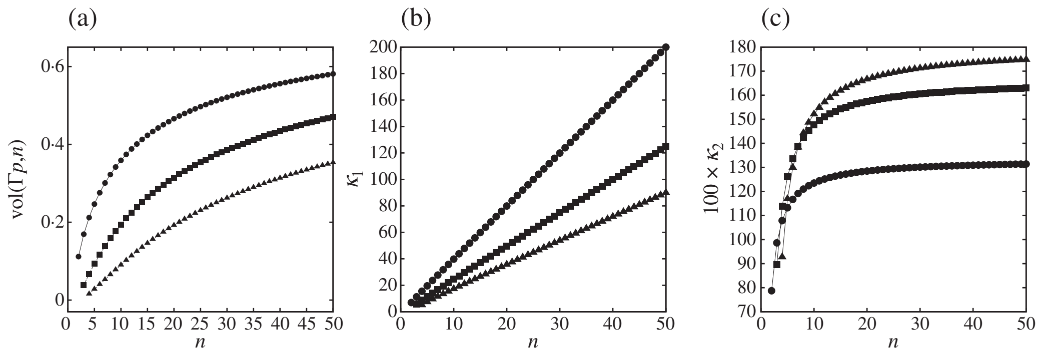

Fig. 6. (a) Volume of $\Gamma_{p,n}$ as a function of n for $p = 2$ (●), $p = 3$ (■) and $p = 4$ (▲); (b) κ_1 as a function of $n \in [p, 50]$ for $p = 2$ (●), $p = 3$ (■) and $p = 4$ (▲); (c) $100 \times \kappa_2$ as a function of $n \in [p, 50]$ for $p = 2$ (●), $p = 3$ (■) and $p = 4$ (▲).

$\hat{\varphi}^{\mathrm{St+iso}}_j(l) = \hat{\varphi}^{\mathrm{St}}_j(l)$ $(j = 1, 2, 3)$. In A^{II}, $\hat{\varphi}^{\mathrm{St}}_2(l)$ and $\hat{\varphi}^{\mathrm{St}}_3(l)$ are pooled together because of either order reversal or negative values, while the estimator of the first eigenvalue is $\hat{\varphi}^{\mathrm{St}}_1(l)$. In A^{III}, $\hat{\varphi}^{\mathrm{St}}_1(l)$ and $\hat{\varphi}^{\mathrm{St}}_2(l)$ are pooled together and the third estimate is $\hat{\varphi}^{\mathrm{St}}_3(l)$. In A^{IV}, $\hat{\varphi}^{\mathrm{St}}_1(l)$, $\hat{\varphi}^{\mathrm{St}}_2(l)$ and $\hat{\varphi}^{\mathrm{St}}_3(l)$ are all pooled together.

The subsets A^{I}, A^{II}, A^{III} and A^{IV} are shown in Fig. 5(a) and (b) for $n = p = 3$ and for $n = 20$. As seen in § 4, the portion of the domain where the isotonizing correction applies, $\Lambda_{3,n} = A^{\mathrm{II}} \cup A^{\mathrm{III}} \cup A^{\mathrm{IV}}$, is substantial, especially when n is close to p. The set A^{IV} contains $\Delta_{3,n}$. Similarly to the $p = 2$ case, $\Delta'_{3,n}$ and $\Delta_{3,n}$ are indistinguishable, i.e., on the domain where $f^{\mathrm{St}}(x) < K^{\mathrm{ml}}_{3,n}$, Stein's estimator retains positivity and the ordering of the sample eigenvalues. The difference $A^{\mathrm{I}} \setminus \Delta_{3,n}$ is $\Gamma_{3,n}$. Figure 5 shows that, for $n = 3$ and $n = 20$, $\Gamma_{3,n}$ is nonempty. The same results hold for other values of n; see Fig. 6(a). Therefore, when the sample eigenvalues are moderately separated, $f^{\mathrm{St+iso}}(x) = f^{\mathrm{St}}(x) \geqslant K^{\mathrm{ml}}_{p,n}$. In fact, more than 40% of the total volume is covered by $\Gamma_{3,n}$.

As $\hat{\psi}^{\mathrm{St+iso}}_j(x)$ takes different forms on different subsets of the domain, $f^{\mathrm{St+iso}}$ is piecewise defined and continuous within each part; see Fig. 5(c) for the $n = 20$ case. Clearly, on A^{II}, A^{III} and A^{IV} the isotonizing correction produces a significant reduction in $f^{\mathrm{St+iso}}(x)$. Similar behaviour is observed for other values of n. As in the $p = 2$ case, the behaviour of the unbiased estimator of risk of Stein's estimator in comparison to that of the maximum likelihood estimator is characterized by three regimes corresponding to the sets $\Delta_{3,n}$, $\Gamma_{3,n}$ and $\Lambda_{3,n}$.

Even in the $p = 3$ case, there exists a subset of the domain on which $f^{\text{iso}}(x)$ is much less than $K^{\text{ml}}_{3,n}$ and simultaneously Stein's estimator does not require the isotonizing correction, though $f^{\text{St}}(x) > K^{\text{ml}}_{3,n}$; see also the Supplementary Material.

As in the $p = 2$ case, the effect of the isotonizing algorithm on the unbiased estimator of risk can be quantified by considering the quantities κ_1 and κ_2 in (11). Both quantities are plotted in Fig. 6(b) and (c) as functions of n for fixed p; κ_1 increases approximately linearly, while κ_2 tends to a finite value. Thus, the reduction in the unbiased estimator of risk due to the isotonizing algorithm compared to that of Stein's raw estimator increases with n, as can be seen in Fig. 6(b), and these increases are more pronounced for lower p. In contrast, as Fig. 6(c) shows, the reduction in the unbiased estimator of risk due solely to the isotonizing algorithm is robust and improves as p gets larger.

Analysis of the $p = 4$ case gives similar results to those observed for $p = 3$; see Fig. 6.

5·3. *The case of arbitrary p*

The following lemma evaluates the impact of the isotonizing correction on the unbiased estimator of risk using κ_1 and κ_2 as before; see the Supplementary Material for the proof.

Lemma 3. *For $p > 1$ and as $n \to \infty$,*

$$\kappa_1 \sim \frac{6\{p(p+1)-2\}n}{p(p^2-1)}, \quad \kappa_2 = 2\left\{1 - \frac{2}{p(p+1)}\right\}\left[1 - \frac{2\{p(2p+3)-1\}}{12(p+1)n}\right] + O\left(\frac{1}{n^2}\right).$$

The behaviour of κ_2 demonstrates that when all the eigenvalues are pooled together, the isotonizing algorithm leads to a large reduction in $f^{\text{St+iso}}(x)$ relative to $K^{\text{ml}}_{p,n}$, which persists even as n increases. Moreover, the horizontal asymptotes shown in Fig. 6(c) are theoretically explained, and the dimension effect seen in Fig. 6(c) is apparent from the expression for κ_2.

The behaviour of κ_1 shows that the reduction in the unbiased estimator of risk of Stein's estimator is orders of magnitude greater when the sample eigenvalues are close to each other than when they are well separated. In other words, the isotonizing algorithm reduces the unbiased estimator of risk much more than Stein's raw estimator does. This suggests that some of the substantial risk reductions seen in Stein's estimator may be attributable to the isotonizing algorithm. Moreover, the convergence of $f^{\text{St+iso}}(x)$ to $K^{\text{ml}}_{p,n}$ is no faster than $O(1/n)$; see the Supplementary Material. This behaviour differs from the analogous rate of $O(1/n^2)$ for Stein's raw estimator; see Proposition 2. The isotonizing correction has the effect of shifting Stein's estimator away from the maximum likelihood estimator by an entire order of magnitude. The ratio of the two rates yields the linear function of n that was observed for small p in Fig. 6(b).

Lemma 3 quantifies the maximal reduction in the unbiased estimator of risk that is achievable by applying the isotonizing algorithm. Analysing the unbiased estimator of risk for the sample covariance structure corresponding to the full-multiplicity case $x = (1, \ldots, 1)$ is convenient as the isotonizing algorithm is being used in this setting. Considering this particular setting yields analytical expressions that can be theoretically examined without resorting to numerical calculations. The effect of the isotonizing algorithm can, however, be investigated under more general covariance structures. Our study of the cases $p = 2, 3, 4$ addressed this issue broadly. The numerical study undertaken in the Supplementary Material affirms the insights from the unbiased estimator of risk framework by quantifying the risks of Stein's isotonized estimator.

A theoretical analysis similar to that above as $p(n), n \to \infty$ gives qualitatively similar results.

6. Discussion

In this paper, three different regimes of behaviour were identified for Stein's unbiased estimator of risk. These findings provide a theoretical means of understanding how the risk of Stein's estimator depends on Σ. In the first regime, when all the eigenvalues of Σ are well separated, the sample eigenvalues are mainly distributed in the part of the domain where $f^{\mathrm{St+iso}}(x) = f^{\mathrm{St}}(x) < K^{\mathrm{ml}}_{p,n}$. Thus, the risk of Stein's estimator is less than that of the maximum likelihood estimator. In this case, the isotonizing correction is not required and the reduction in the unbiased estimator of risk, and hence in the risk, is to be attributed only to Stein's raw estimator. The reduction, however, is not very large and decreases as n increases. In the second regime, some of the eigenvalues of Σ are close together or are moderately separated but n is close to p. In this regime, Stein's estimator requires extensive use of the isotonizing correction, which reduces the unbiased estimator of risk considerably. Here the risk of Stein's isotonized estimator is much lower than that of the maximum likelihood estimator. Our analysis shows that the size of the set in which the isotonizing algorithm applies increases as n decreases and attains its maximum at $n = p$. Lastly, in the third regime, when the eigenvalues of Σ are only moderately separated and n is sufficiently greater than p, the unbiased estimator of risk of Stein's raw estimator is greater than that of the maximum likelihood estimator. Here the isotonizing algorithm does not apply, although were it to be applied, it would lead to a reduction in the unbiased estimator of risk. As a consequence, the risk of Stein's estimator is comparable to that of the maximum likelihood estimator, or even greater than it. In conclusion, the unbiased estimator of risk approach enables us to gain important insights into the theoretical workings of Stein's isotonized estimator.

Some open issues that can be explored in the future warrant mention. First, deriving a singular Wishart identity in the sample-starved setting may help in extending the estimator to modern high-dimensional settings. Second, it would be useful to derive a Wishart identity when various structures on the covariance are imposed, such as a graphical model structure or constraints on the condition number. An example of this would be a hyper-Wishart identity. Third, deriving the unbiased estimator of risk for other loss functions such as the operator norm loss may also be useful for high-dimensional applications.

Acknowledgement

We thank Charles Stein for his encouragement and Brett Naul for helping to draw Fig. 1(a). We acknowledge the support of the France–Stanford Center for Interdisciplinary Studies. The first author was partially supported by the U.S. Air Force Office of Sponsored Research, the National Science Foundation, the Defense Advanced Research Projects Agency, the United Parcel Service Fund, and the Stanford Management Company.

Supplementary material

Supplementary material available at *Biometrika* online includes additional details on the unbiased estimator of risk of Stein's raw estimator for $p = 2, 3, 4$, the proofs of the lemmas and propositions, validation of the unbiased estimator of risk approach via risk calculations, the proof of the spectral norm convergence of Stein's isotonized estimator, and a comparison of Stein's estimator with the Ledoit–Wolf shrinkage estimator.

References

Daniels, M. J. & Kass, R. E. (2001). Shrinkage estimators for covariance matrices. *Biometrics* **57**, 1173–84.
Haff, L. R. (1991). The variational form of certain Bayes estimators. *Ann. Statist.* **19**, 1163–90.

Khare, K. & Rajaratnam, B. (2011). Wishart distributions for decomposable covariance graph models. *Ann. Statist.* **39**, 514–55.

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

Lin, S. & Perlman, M. (1985). A Monte Carlo comparison of four estimators of a covariance matrix. In *Multivariate Analysis VI*, P. R. Krishnaiah, ed. Amsterdam: North-Holland, pp. 411–29.

Pourahmadi, M. (2011). Covariance estimation: The GLM and regularization perspectives. *Statist. Sci.* **26**, 369–87.

Rajaratnam, B., Massam, H. & Carvalho, C. M. (2008). Flexible covariance estimation in graphical Gaussian models. *Ann. Statist.* **36**, 2818–49.

Rajaratnam, B. & Vincenzi, D. (2016). A note on covariance estimation in the unbiased estimator of risk framework. *J. Statist. Plan. Infer.* **175**, 25–39.

Schäfer, J. & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statist. Appl. Genet. Molec. Biol.* **4**, article 32.

Stein, C. (1975). Estimation of a covariance matrix. In *Riesz Lecture, 39th Annual Meeting*. Atlanta, Georgia: Institute of Mathematical Statistics.

Stein, C. (1977). Lectures on the theory of estimation of many parameters (in Russian). In *Studies in the Statistical Theory of Estimation, Part I*, I. A. Ibragimov & M. S. Nikulin, eds. Proceedings of Scientific Seminars of the Steklov Institute, Leningrad Division, vol. 74, pp. 4–65.

Stein, C. (1986). Lectures on the theory of estimation of many parameters. *J. Math. Sci.* **34**, 1373–403.

Won, J., Kim, S. J., Lim, J. & Rajaratnam, B. (2013). Condition number-regularized covariance estimation. *J. R. Statist. Soc.* B **75**, 427–50.

[*Received August* 2014. *Revised June* 2016]