The Annals of Statistics
1995, Vol. 23, No. 2, 571–592

ESTIMATION OF A LOSS FUNCTION FOR SPHERICALLY SYMMETRIC DISTRIBUTIONS IN THE GENERAL LINEAR MODEL

BY DOMINIQUE FOURDRINIER AND MARTIN T. WELLS

Université de Rouen and Cornell University

This paper is concerned with estimating the loss of a point estimator when sampling from a spherically symmetric distribution. We examine the canonical setting of a general linear model where the dimension of the parameter space is greater than 4 and less than the dimension of the sampling space. We consider two location estimators—the least squares estimator and a shrinkage estimator—and we compare their unbiased loss estimator with an improved loss estimator. The domination results are valid for a large class of spherically symmetric distributions and, in so far as the sampling distribution does not need to be precisely specified, the estimates have desirable robustness properties.

1. Introduction.

1.1. Consider the general statistical decision problem of estimating a parameter $\theta \in \mathbb{R}^k$ by some decision procedure φ under a loss function $L(\theta, \varphi)$. Classical decision theory advocates that one should use a decision rule φ^* if it has suitable properties with respect to the frequentist risk $R(\theta, \varphi^*)$. However, once a random vector x has been observed, one would like to calculate the loss incurred by the estimate $\varphi^*(x)$. Now, $L(\theta, \varphi^*(x))$ is not available since it depends on the unknown parameter θ. Hence it is of interest to estimate the loss of $\varphi^*(x)$. Therefore, we wish to construct a *data dependent* measure of the loss incurred for the particular data at hand.

The problem of estimating the loss was first considered by Lehmann and Sheffé [14], who estimated the power of a statistical test. In a series of papers, Kiefer [11–13] addressed the problem of developing conditional and estimated confidence theories to provide frequentist estimates of confidence. Berger [2] compared the Bayesian and frequentist approaches to this problem. Recently Johnstone [10], Rukhin [18], Lu and Berger [16], Casella and discussants [4], Casella, Hwang and Robert [5] and Lele [15] have discussed this problem in a variety of situations.

In this article, we consider estimation of the loss incurred when using the least squares estimator and an improved estimator of the location parameter of a spherically symmetric distribution. We first develop an unbiased estimator of the loss for each location estimator. Next we construct, for each

Received August 1992; revised May 1994.
AMS 1991 *subject classifications*. 62A99, 62C05, 62C15, 62C99, 62F35.
Key words and phrases. Spherical symmetry, loss estimation, shrinkage estimation, conditional inference.

unbiased estimator of the loss, a dominating shrinkage-type estimator, in terms of squared-error loss. These results complement James–Stein-type estimation of a location parameter of a spherically symmetric distribution given by [3] and [6]. An important feature of our results is that the proposed loss estimates dominate the unbiased estimates for the entire class of spherically symmetric distributions. That is, the domination results are robust with respect to spherical symmetry.

The normal distribution has long served as the standard model in the investigation of a location parameter. One of its main attractive feature is that it depends on a small number of parameters which have direct interpretation. The normal distribution has been generalized in two important directions: first as a special case of the exponential family and second as a spherically symmetric distribution. We will consider the latter (see, e.g., [8] for a comprehensive review of that class of distributions). The spherical case has recently received much attention, especially in decision theoretic point estimation with Brandwein and Strawderman [3] and in set estimation with Robert and Casella [17]. These articles refer to specified spherically symmetric distributions while our results are robust with respect to the choice of the spherically symmetric distribution (this fact was already noticed by Cellier, Fourdrinier and Robert [7]).

1.2. Let x be an observation in an n-dimensional Euclidean space $(E, \langle\ ,\ \rangle)$, distributed according to a spherically symmetric distribution P_θ around a location parameter θ. The main hypothesis about P_θ is that θ belongs to a linear subspace Θ of E with dimension k for $0 < k < n$. Suppose we wish to estimate θ, by a decision rule $\varphi(x)$ using the sum of squared-error loss $\|\theta - \varphi(x)\|^2$, where $\|\cdot\|$ denotes the norm connected with the inner product $\langle\ ,\ \rangle$. This loss is unobservable since it depends on θ; hence one may wish to estimate it by $\lambda(x)$ from the data. To study how well λ estimates the loss, a further distance measure is needed; for mathematical convenience, we use squared error to evaluate $\lambda(x)$. Thus the risk incurred by λ is

$$R(\lambda, \theta, \varphi) = E_\theta\Big[\big(\lambda - \|\varphi - \theta^2\big)^2\Big], \tag{1.1}$$

where E_θ denotes the expectation with respect to P_θ. We say that a loss estimator λ' dominates λ if $R(\lambda', \theta, \varphi) \le R(\lambda, \theta, \varphi)$.

Since P_θ is spherically symmetric around θ, for every bounded function f, we have $E_\theta[f] = \int_R E_{R,\theta}[f]\,\rho(dR)$, where $E_{R,\theta}$ denotes the expectation with respect to the uniform distribution $U_{R,\theta}$ on the sphere $S_{R,\theta} = \{x \in E/\|x - \theta\| = R\}$ of radius R and center θ, and ρ is the distribution of the radius, namely, the distribution of the norm $\|\cdot\|$ under P_0. It suffices to prove the domination results working conditionally on the radius, that is to say, to replace P_θ by $U_{R,\theta}$ in the expression (1.1).

As $k < n$, the usual estimator of θ is the orthogonal projector φ_0 from E onto Θ, that is, the regular least squares estimator. A class of competing point estimators which are also considered are of the form $\varphi = \varphi_0 - \|X -$

$\varphi_0\|^2 g \circ \varphi_0$, where X is the identity function in E [for every $x \in E$, $X(x) = x$] and g is a measurable function from Θ into Θ. This class of estimators is closely related to the now classical Stein-like estimators (when estimating the mean of a normal distribution, the square of the residual term $\|x - \varphi_0(x)\|$ is used as an estimate of the unknown variance). Their domination properties are robust with respect to spherical symmetry (cf. [6] and [7]). In the next section, we consider estimation of the loss of the usual least squares estimator φ_0. In Section 3, we consider estimation of the loss of a shrinkage estimator φ. In order to assure the finiteness of the risk of the usual estimator φ_0 and the risk of the shrinkage estimator φ, we need two hypotheses (H1) and (H2) given in [6]. Section 4 contains some concluding remarks. The Appendix provides two technical lemmas that are used repeatedly.

1.3. We wish to emphasize some points on which our article differs from Johnstone's paper (i.e., [10]). In the normal case, we know that the risk of the least squares estimator φ_0 is constant and equal to k, so it is natural to use $\bar{\lambda} = k$ as an unbiased estimator of the loss. Here we use Johnstone's terminology for unbiasedness. The only way we can determine unbiasedness is to take the expected value with respect to P_θ, in which case, the expected value of $\bar{\lambda}$ is equal to the expected value of the loss, both being equal to the risk. Hence, one might speak about the unbiased estimator of the risk rather than unbiased estimators of loss.

In the spherical case, the risk of φ_0 remains constant (with respect to θ) since it is equal to $kE[R^2]/n$, where $E[R^2]$ denotes the expectation of the square of the radius (the second centered moment of P_θ). Thus this risk provides an unbiased estimator of the loss, that is,

$$\bar{\lambda} = \frac{k}{n} E[R^2], \tag{1.2}$$

which is subject to the knowledge of $E[R^2]$. Its properties, as the properties of any improved estimator, may depend on the specific underlying distribution.

A feature of this paper is that we propose an unbiased estimator λ_0 of the loss of φ_0 which is available for every spherically symmetric distribution (with finite fourth moment), that is, $\lambda_0(x) = k\|x - \varphi_0(x)\|^2/(n-k)$. Thus we do not need to know the specific distribution, and we get robustness with an estimator which is no longer constant. In Section 2, we give some examples of some spherically symmetric distributions; these include the Student distribution (and, of course, the normal distribution) and some distributions which are not scale mixtures of the normal distribution.

Notice λ_0 makes sense because $\dim \Theta < \dim E$. A reasonable goal is to improve on the unbiased estimator of loss and we show, in Section 2, that this can be done while preserving the robustness property. The above considerations of robustness apply to the estimation of the loss of a shrinkage estimator discussed in Section 3. Another major difference between our article and

Johnstone's is that we consider the coordinate-free (cf. [20]) general linear model with $\dim \Theta < \dim E$. Although Johnstone mentioned a type of linear model, he assumed that $\dim \Theta = \dim E$. Hence, his results are not applicable to the general linear model.

2. Estimation of the loss of the least squares estimator.

2.1. Within the framework introduced in Section 1.2, we consider estimation of the loss of the usual least squares estimator φ_0 of θ (the orthogonal projector from E onto Θ) in this section. An unbiased estimator of the loss of φ_0 of θ is given by $\lambda_0 = k\|X - \varphi_0\|^2/(n - k)$. The unbiasedness of λ_0 follows from Lemma A.1 in the Appendix by taking $q = 0$ and $\gamma \equiv 1$. The goal of this section is to prove the domination of the unbiased estimator λ_0 by a competing estimator λ of the form

$$\lambda = \lambda_0 - \|X - \varphi_0\|^4 \gamma \circ \varphi_0, \tag{2.1}$$

where γ is a positive function. It is important to notice that the "residual term" $\|X - \varphi_0\|$ appears explicitly in the shrinkage function. It has been noted by Cellier and Fourdrinier [6] that the use of this term allows fewer assumptions about the distributions than when it does not appear. Specifically, this inclusion gives a robustness property to the results, since they are valid for the entire class of spherically symmetric distributions (under the required moment conditions). Since, for a given observation x, the residual term $\|x - \varphi_0(x)\|^2$ represents the square of the distance between x and its projection on Θ, it is intuitively natural that its consideration strengthens the information we use through the estimator. See Remark 2.1 on a rationale for the choice of the fourth power on the residual term. Johnstone [10] presents some simulation results which point out the great gains one can obtain by using an improved estimate of loss. Similar gains also occur for our loss estimates.

2.2. Before giving our main result, we consider the problem of the finiteness of the risks of the estimators λ_0 and λ. It is easy to check, using the spherical symmetry of P_θ and the proportionality of $E_\theta[\|X - \varphi_0\|^4]$ and $E_\theta[\|\varphi_0 - \theta\|^4]$ [this follows from two applications of Lemma A.1, first with $q = 0$ and $\gamma(t) = \|t - \theta\|^2$, then with $q = 2$ and $\gamma(t) = 1$], that the risk of the unbiased estimator λ_0 is finite if and only if P_θ has a finite fourth moment. If the risk of λ_0 is finite, direct calculation (see the first expression of the risk of λ given at the beginning of the proof of Theorem 2.1) and an application of the Cauchy–Schwarz inequality show that the risk of the shrinkage estimator (2.1) is finite if and only if $E_\theta[\|X - \varphi_0\|^8 \gamma^2 \circ \varphi_0] < \infty$. A straightforward way of showing that this expectation is finite is to assume that there exists a constant $\beta > 0$ such that

$$\gamma(t) \le \beta/\|t\|^2 \quad \text{for every } t \in \theta. \tag{2.2}$$

This condition is often assumed when estimating a location parameter; see [6] for more details and references. Indeed, working conditionally on the radius R, (2.2) implies

$$(2.3)\qquad E_{R,\theta}\left[\|X-\varphi_0\|^8\gamma^2\circ\varphi_0\right]\le\beta^2R^4E_{R,\theta}\left[\left(\frac{\|X-\varphi_0\|^2}{\|\varphi_0\|^2}\right)^2\right].$$

On the right-hand side of (2.3), for $\theta=0$, the expectation is independent of R since it is the second moment of a F-distributed random variable with $n-k$ and k degrees of freedom (up to a multiplicative constant), as the distribution is spherically symmetric around zero. This moment is finite as soon as $k>4$, and remains finite for $\theta\neq 0$ since it can be bounded from above by a constant independent of R (see Section A.1 of the Appendix). When we uncondition with respect to R, under the assumption that P_θ has a finite fourth moment, the right-hand side of (2.3) is finite. Hence, to ensure risk finiteness, we assume $k>4$.

2.3. The following result and the results of Section 3 require the real-valued function γ to be twice weakly differentiable (cf. [23]) in order to include basic examples, which are not twice differentiable. Recall that, for a given multiindex α [i.e., $\alpha=(\alpha_1\dots\alpha_n)$ is an n-tuple of nonnegative integers], a locally integrable function γ is α-weakly differentiable if there exists a locally integrable function δ such that

$$\int_E\varphi(x)\delta(x)\,dx=(-1)^{|\alpha|}\int_E\gamma(x)D^\alpha\varphi(x)\,dx$$

for every infinitely differentiable function φ with compact support, where $|\alpha|=\sum_{i=1}^n\alpha_i$ is the length of α and $D^\alpha=D_1^{\alpha_1}\cdots D_n^{\alpha_n}=\partial^{|\alpha|}/\partial x_1^{\alpha_1}\cdots\partial x_n^{\alpha_n}$ is the higher order derivative operator. Then δ is referred to as the αth weak derivative of γ and is denoted by $\delta=D^\alpha\gamma$. When the αth weak derivative exists for every multiindex α such that $|\alpha|=2$, γ is said to be twice weakly differentiable and the Laplacian of γ is the differential operator given by

$$\Delta\gamma=\sum_{i=1}^nD_i^2\gamma=\sum_{i=1}^n\frac{\partial^2\gamma}{\partial x_i^2}.$$

2.4. We can now state the following theorem.

THEOREM 2.1. *Assume that $k>4$, the distribution P_θ has a finite fourth moment and the function γ is twice weakly differentaible on Θ and satisfies (2.2). A sufficient condition under which the estimator λ given in (2.1) dominates the unbiased estimator λ_0 is that γ satisfies the differential inequality*

$$\gamma^2+\frac{2}{(n-k+4)(n-k+6)}\Delta\gamma\le 0.$$

PROOF. Since P_θ is spherically symmetric around θ, it suffices to obtain the result working conditionally on the radius. Referring to the notations given in Section 1, for $R > 0$ fixed, we can compute using the uniform distribution $U_{R,\theta}$ on the sphere $S_{R,\theta}$. Hence we have

$$E_{R,\theta}\Big[\big(\lambda - \|\varphi_0 - \theta\|^2\big)^2\Big] = E_{R,\theta}\Big[\big(\lambda_0 - \|\varphi_0 - \theta\|^2\big)^2\Big] + E_{R,\theta}\big[\|X - \varphi_0\|^8 \gamma^2 \circ \varphi_0\big] - 2E_{R,\theta}\big[\big(\lambda_0 - \|\varphi_0 - \theta\|^2\big)\|X - \varphi_0\|^4 \gamma \circ \varphi_0\big].$$

Developing the cross-product term and using the form of λ_0, we have

$$\begin{aligned} &E_{R,\theta}\big[\big(\lambda_0 - \|\varphi_0 - \theta\|^2\big)\|X - \varphi_0\|^4 \gamma \circ \varphi_0\big] \\ &\quad = \frac{k}{n-k} E_{R,\theta}\big[\|X - \varphi_0\|^6 \gamma \circ \varphi_0\big] - E_{R,\theta}\big[\|\varphi_0 - \theta\|^2 \|X - \varphi_0\|^4 \gamma \circ \varphi_0\big]. \end{aligned} \tag{2.4}$$

Using Lemma A.1 in the Appendix with $q = 4$, the second integral of the right-hand side becomes

$$\begin{aligned} &E_{R,\theta}\big[\|\varphi_0 - \theta\|^2 \|X - \varphi_0\|^4 \gamma \circ \varphi_0\big] \\ &\quad = \frac{k}{n-k+4} E_{R,\theta}\big[\|X - \varphi_0\|^6 \gamma \circ \varphi_0\big] \\ &\qquad + \frac{1}{(n-k+4)(n-k+6)} E_{R,\theta}\big[\|X - \varphi_0\|^2 \, \Delta\gamma \circ \varphi_0\big]. \end{aligned}$$

Replacing this expression in the cross-product term and combining the terms with $\|X - \varphi_0\|^6$, we get

$$\begin{aligned} &E_{R,\theta}\Big[\big(\lambda - \|\varphi_0 - \theta\|^2\big)^2\Big] \\ &\quad = E_{R,\theta}\Big[\big(\lambda_0 - \|\varphi_0 - \theta\|^2\big)^2\Big] + E_{R,\theta}\big[\|X - \varphi_0\|^8 \gamma^2 \circ \varphi_0\big] \\ &\qquad - \frac{8k}{(n-k)(n-k+4)} E_{R,\theta}\big[\|X - \varphi_0\|^6 \gamma \circ \varphi_0\big] \\ &\qquad + \frac{2}{(n-k+4)(n-k+6)} E_{R,\theta}\big[\|X - \varphi_0\|^8 \, \Delta\gamma \circ \varphi_0\big]. \end{aligned} \tag{2.5}$$

Since, on the right-hand side, the second term is negative (as γ is positive) and we have the same power 8 for the term $\|X - \varphi_0\|$ in the two last integrals, it is clear that $R(\lambda, \theta, \varphi_0) \le R(\lambda_0, \theta, \varphi_0)$ provided that

$$\gamma^2 + \frac{2}{(n-k+4)(n-k+6)} \Delta\gamma \le 0. \qquad \square$$

REMARK. 2.1. The proof of Theorem 2.1 and Lemma A.1 show that the power $q = 4$ chosen for the "residual term" $\|X - \varphi_0\|$ in the expression of λ is the only possible one. Indeed for any arbitrary q we would obtain $\|X - \varphi_0\|^{2q}$ before γ^2 and $\|X - \varphi_0\|^{q+4}$ before $\Delta\gamma$ and the comparison of these two terms is possible only if $2q = q + 4$, that is to say, only if $q = 4$.

EXAMPLE 2.1. As in [10], the example for which $\gamma(t) = d/\|t\|^2$ for all $t \neq 0$ with $d > 0$ works. More precisely it is easy to deduce that $\Delta\gamma(t) = -2d(k-4)/\|t\|^4$ and thus the sufficient condition of the theorem is written as $0 < d \leq 4(k-4)/(n-k+4)(n-k+6)$, which only occurs when $k > 4$. Straightforward calculus shows that the optimal value of d is given by $2(k-4)/(n-k+4)(n-k+6)$. The optimal constant in [10] is equal to $2(k-4)$. The extra terms in the denominator compensate for the $\|X - \varphi_0\|^4$ term in our estimator.

2.5. Straightforward but tedious calculations show that the risk of λ_0 is constant (with respect to θ) and is proportional to the fourth centered moment of the underlying distribution. Indeed, conditionally on the radius R, the risk of λ_0 at θ equals

$$
\begin{aligned}
&E_{R,\theta}\Big[\big(\lambda_0 - \|\varphi_0 - \theta\|^2\big)^2\Big] \\
&\quad = \frac{k^2}{(n-k)^2} E_{R,\theta}\big[\|X - \varphi_0\|^4\big] \qquad (2.6)\\
&\qquad - \frac{2k}{n-k} E_{R,\theta}\big[\|X - \varphi_0\|^2 \|\varphi_0 - \theta\|^2\big] + E_{R,\theta}\big[\|\varphi_0 - \theta\|^4\big]
\end{aligned}
$$

and repeated applications of Lemma A.1 (with different values of q and different functions γ) reduce the right-hand side of (2.6) to a term proportional to $E_{R,\theta}[\|X - \varphi_0\|^4]$, and the use of Lemmas A.1 and A.3 leads to

$$E_{R,\theta}\Big[\big(\lambda_0 - \|\varphi_0 - \theta\|^2\big)^2\Big] = \frac{2k}{(n-k)(n+2)} R^4.$$

Finally, unconditionally, the risk of λ_0 at θ equals

$$R(\lambda_0, \theta, \varphi_0) = \frac{2k}{(n-k)(n+2)} E[R^4].$$

The calculation of the risk of the constant loss estimate $\bar{\lambda}$ [cf. (1.2)] follows the same way. Conditionally on the radius R, we have

$$
\begin{aligned}
E_{R,\theta}\Big[\big(\bar{\lambda} - \|\varphi_0 - \theta\|^2\big)^2\Big] &= \bar{\lambda}^2 - 2\bar{\lambda} E_{R,\theta}\big[\|\varphi_0 - \theta\|^2\big] + E_{R,\theta}\big[\|\varphi_0 - \theta\|^4\big] \qquad (2.7)\\
&= \bar{\lambda}^2 - 2\bar{\lambda}\frac{k}{n} R^2 + \frac{k(k+2)}{n(n+2)} R^4.
\end{aligned}
$$

Thus unconditionally, since $\bar{\lambda} = (k/n)E[R^2]$,

$$R(\bar{\lambda}, \theta, \varphi_0) = \frac{k(k+2)}{n(n+2)} E[R^4] - \left(\frac{k}{n}\right)^2 (E[R^2])^2.$$

Combining the results above, the risk difference between λ_0 and $\bar{\lambda}$ equals, for every θ,

$$\Delta(\theta) = R(\lambda_0, \theta, \varphi_0) - R(\bar{\lambda}, \theta, \varphi_0)$$
$$= \frac{k^2}{n^2}\left(\left(E[R^2]\right)^2 - \frac{n(n-k-2)}{(n-k)(n+2)} E[R^4]\right).$$

When $n - 2 \leq k < n$ (that is, $k = n - 1$ or $k = n - 2$), it is clear that $\Delta(\theta) \geq 0$, so $\bar{\lambda}$ is preferable to λ_0. In the other case, when $1 \leq k < n - 2$, the fact that $\Delta(\theta)$ is positive or negative depends on the distributions of the radius. Thus $\Delta(\theta)$ is nonpositive (and λ_0 is better than $\bar{\lambda}$) if

$$\frac{E[R^4]}{\left(E[R^2]\right)^2} \geq \frac{(n-k)(n+2)}{(n-k-2)n} > 1. \tag{2.8}$$

2.6. *Examples of distributions.* Assume the distribution P_θ is normal with density

$$\frac{1}{(2\pi)^n} \exp\left(-\frac{1}{2}\|x - \theta\|^2\right);$$

thus the density of the radius is equal to

$$\frac{2^{1-n/2}}{\Gamma(n/2)} R^{n-1} \exp\left(-\frac{R^2}{2}\right).$$

It is easy to see that the second and fourth moments of the radius equal

$$E[R^2] = n \quad \text{and} \quad E[R^4] = n(n+2).$$

Therefore, condition (2.8) is not satisfied and $\bar{\lambda}$ $(= k)$ is better than λ_0. Actually, a straightforward calculation of the risks of $\bar{\lambda}$ and λ_0 gives, respectively, $2k$ and $2kn/(n-k)$, so $\Delta(\theta) = 2k^2/(n-k) > 0$.

With a multivariate Student distribution, the result differs according to the degrees m of freedom. Indeed consider, for P_θ, the density

$$g(\|x - \theta\|^2) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)(\pi m)^{n/2}} \left[1 + \frac{\|x - \theta\|^2}{m}\right]^{-(m+n)/2};$$

hence, the density of the radius is equal to

$$f(R) = \frac{2\pi^{n/2}}{\Gamma(n/2)} R^{n-1} g(R^2)$$
$$= \frac{2\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)m^{n/2}} \left[1 + \frac{R^2}{m}\right]^{-(m+n)/2} R^{n-1}.$$

After some tedious calculations we get

$$E[R^2] = \frac{nm}{m-2} \quad (\text{for } m > 2) \quad \text{and}$$

$$E[R^4] = \frac{n(n+2)m^2}{(m-4)(m-2)} \quad (\text{for } m > 4).$$

Then we have

$$\frac{E[R^4]}{(E[R^2])^2} = \frac{n+2}{n}\frac{m-2}{m-4}$$

and, according to condition (2.8), λ_0 is better than $\bar{\lambda}$ if and only if $4 < m \le n - k + 2$ (recall that $k < n - 2$). The fact that the Student distribution behaves differently from the normal distribution is well known. Although it gives a good approximation to the normal model, Zellner [22] has shown that a t-distribution leaves, through the choice of m, more freedom to the experimenter (see also [17]).

We give now an example which is not a mixture of normal distributions. Assume P_θ has the density $g(\|x - \theta\|^2)$ with

$$g(s) = \frac{\Gamma(n/2)}{(2\pi)^{n/2}2^p\Gamma(n/2+p)} s^p \exp\left(-\frac{s}{2}\right).$$

When $p \neq 0$, the function g is not completely monotonic [i.e., we do not have $(-1)^m\, d^m g/ds^m \ge 0$, for every m], so the distribution is not a normal mixture (see [1]). The density of the radius is given by

$$f(R) = \frac{2\pi^{n/2}}{\Gamma(n/2)} R^{n-1} g(R^2) = \frac{2^{1-(n/2+p)}}{\Gamma(n/2+p)} R^{n+2p-1} \exp\left(-\frac{R^2}{2}\right).$$

A straightforward calculation gives

$$E[R^2] = n + 2p \quad \text{and} \quad E[R^4] = (n+2p)(n+2+2p) \quad \text{if } n + 2p > 0.$$

Thus, if $p > 0$,

$$\frac{E[R^4]}{(E[R^2])^2} = \frac{n+2+2p}{n+2p} < \frac{n+2}{n} < \frac{n-k}{n-k-2}\frac{n+2}{n}$$

and, according to (2.8), $\bar{\lambda}$ is better than λ_0. When $p < 0$, the difference in risk $\Delta(\theta)$ may be positive or negative.

Recall that this type of comparison only makes sense if we know the distribution P_θ. This underlines the robustness that λ_0 brings; it can be considered even when we do not know P_θ. If P_θ were known and $\bar{\lambda}$ were preferred to λ_0, it would be easy to show that there exist improved versions of $\bar{\lambda}$. However, our primary focus here is on the construction of estimators whose domination properties are valid for the entire class of spherically symmetric distributions.

2.7. It is of interest to investigate the optimality of the unbiased estimator in the class of estimators of the form $\lambda_\alpha = \alpha\|X - \varphi_0\|^2$. As in the proof, it is clear that it suffices to obtain the result in working conditionally on the radius. Referring to the notations given in Section 1, for $R > 0$ fixed, we can compute using the uniform distribution $U_{R,\theta}$ on the sphere $S_{R,\theta}$. Hence by applying Lemma A.1 and some tedious algebra, we have the risk of λ_α equals

$$E_{R,\theta}\Big[\big(\lambda_\alpha - \|\varphi_0 - \theta\|^2\big)^2\Big] = \left[\alpha^2 - 2\alpha\frac{k}{n-k+2} + \frac{k(k+2)}{(n-k)(n-k+2)}\right] \times E_{R,\theta}\big[\|X - \varphi_0\|^4\big].$$

Since this is a quadratic and convex function in α, the minimum occurs at $\alpha^* = k/(n-k+2)$. This shows that the unbiased estimator is *not* optimal in this class. This extra factor of 2 in the denominator is reminiscent of the minimum risk equivalent estimation (MRE) of the variance in the normal case. However, the MRE variance estimate is not optimal since Stein [19] proves that it is improved, using the information in the sample mean in addition to that in the sample variance.

Now consider the more general shrinkage estimator

$$\lambda_\alpha^s = \alpha\|X - \varphi_0\|^2 - \|X - \varphi_0\|^4\gamma \circ \varphi_0.$$

An easy modification of the proof of Theorem 2.1, by replacing $k/(n-k)$ in (2.4) by α, leads to the calculation of the risk of λ_α^s. In this case, the second term on the right-hand side of (2.5) is equal to $-2(\alpha - k/(n-k+4))$ and may still be omitted, as in the proof, if α is chosen so that $k/(n-k+4) \le \alpha$. Note that the optimal α^* satisfies this inequality. Thus, for this value of α and under the conditions of Theorem 2.1, the estimator λ_α^s dominates λ_α.

3. Estimation of a loss of a shrinkage estimator. In this section, we consider the estimation of the loss of a class of shrinkage estimators. The class of location estimators we consider is

$$\varphi = \varphi_0 - \|X - \varphi_0\|^2 g \circ \varphi_0,$$

where g is a weakly differentiable function from Θ into Θ. In [6] it is shown that, if $\|g\|^2 \le 2\operatorname{div} g/(n-k+2)$, φ dominates φ_0, under quadratic loss for all spherically symmetric distributions with a finite second moment. This class of point estimators is reminiscent of the Stein-like estimators (cf. [3]); however it does not contain them.

A general example of a member of this class of estimators is with

$$g(\varphi_0) = r\big(\|\varphi_0\|^2\big)\frac{A(\varphi_0)}{b(\varphi_0)},$$

where r is a positive differentiable and nondecrasing function, A is a symmetric endomorphism whose eigenvalues are positive and b is a positive definite quadratic form on Θ. When r is equal to some constant a, A is the identity on Θ and the quadratic form b is the usual norm, g reduces to $a/\|\varphi_0\|^2$. It can be shown that the optimal choice of a equals $(k-2)/(n-k+2)$. Hence a simple member of the class is

$$\varphi_r = \varphi_0 - (k-2)\frac{\|X-\varphi_0\|^2}{n-k+2}\frac{\varphi_0}{\|\varphi_0\|^2}.$$

This is in contrast to the usual James–Stein estimator

$$\varphi_s = \varphi_0 - (k-2)\frac{\varphi_0}{\|\varphi_0\|^2}$$

in the normal case with known variance. However, φ_r is the James–Stein form used when the variance is unknown.

The difference between the two estimators is the term $\|X-\varphi_0\|^2/(n-k+2)$. An application of Lemma A.3 yields that, in expectation, this term equals $(n-k)E(R^2)/n$. Therefore, the amount of shrinkage depends on the second moment of the radial distribution. In the normal case, this term equals $(n-k)/(n-k+2)$ in expectation; therefore, φ_r and φ_s differ slightly in expectation. Some tedious calculation reveals that, in the normal case (when the variance is known), φ_s has slightly smaller risk than φ_r. However, that comparison is for the sampling distribution *fixed* to be normal. The main attraction of the new estimators is that they are robust, that is, they dominate φ_0 for all spherically symmetric distributions with finite second moment. This is important since one usually makes certain distributional assumptions without full knowledge of the true distribution. When using the robust estimators one needs only assume that the sampling distribution is a member of the large class of spherically symmetric distributions with finite second moment. This loss of efficiency–robustness trade-off is reminiscent of the classical theory of robustness *à la* Huber [9]. Huber's theory basically constructs likelihood estimators over a class of distributions. The estimate for the class is chosen by finding the distribution in the class which has maximal inverse Fisher information and then selecting a minimax estimator. For a fixed distribution of the class, Huber's minimax estimate can be dominated; however, the minimax estimate has desirable properties when the fixed distribution has been misspecified.

In Proposition 2.3.1 of Section 2.3 of [6] it is shown that an unbiased estimator of the loss of the shrinkage estimator φ is given by

$$\lambda_0^s = \frac{k}{n-k}\|X-\varphi_0\|^2 + \left(\|g\circ\varphi_0\|^2 - \frac{2}{n-k+2}\operatorname{div} g\circ\varphi_0\right)\|X-\varphi_0\|^4.$$

As in the previous section, we prove that the unbiased estimator of the loss can be improved by a shrinkage estimator of the loss. Thus the competing estimator we consider is

$$\lambda^s = \lambda_0^s - \|X-\varphi_0\|^4\gamma\circ\varphi_0, \tag{3.1}$$

where γ is a positive function. Note that (3.1) is a true shrinkage estimator, while Johnstone's [10] optimal loss estimate for the normal case is an expanding estimator. This is not contradictory since we are using a different estimator than Johnstone and he is only dealing with the normal case. The remarks about the risk finiteness preceding Theorem 2.1 are also valid for this new loss estimator.

The main result of this section is Theorem 3.1. It is easy to see that for $g \equiv 0$ its statement reduces to Theorem 2.1. For clarity of exposition we prefer to give two separate results.

THEOREM 3.1. *Assume that $k > 4$, the distribution P_θ has a finite fourth moment and the function γ is twice weakly differentiable on Θ and satisfies* (2.2). *A sufficient condition under which the estimator λ^s given in* (3.1) *dominates the unbiased estimator λ_0^s is that γ satisfies the differential inequality*

$$\gamma^2 + \frac{4}{n-k+2}\gamma \operatorname{div} g - \frac{4}{n-k+6}\operatorname{div}(\gamma g) + \frac{2}{(n-k+4)(n-k+6)}\Delta\gamma \le 0.$$

PROOF. As in the proof of Theorem 2.1, we work conditionally on the radius R. Hence, we can write

$$\begin{aligned} &E_{R,\theta}\Big[\big(\lambda^s - \|\varphi - \theta\|^2\big)^2\Big] \\ &\quad = E_{R,\theta}\Big[\big(\lambda_0^s - \|\varphi - \theta\|^2\big)^2\Big] + E_{R,\theta}\big[\|X - \varphi_0\|^8 \gamma^2 \circ \varphi_0\big] \\ &\qquad - 2E_{R,\theta}\big[\|X - \varphi_0\|^4 \gamma \circ \varphi_0\big(\lambda_0^s - \|\varphi - \theta\|^2\big)\big]. \end{aligned} \tag{3.2}$$

It is clear, using Lemma A.1 as in the previous theorem, that

$$\begin{aligned} &E_{R,\theta}\big[\|X - \varphi_0\|^4 \gamma \circ \varphi_0 \|\varphi_0 - \theta\|^2\big] \\ &\quad = \frac{k}{n-k+4} E_{R,\theta}\big[\|X - \varphi_0\|^6 \gamma \circ \varphi_0\big] \\ &\qquad - \frac{1}{(n-k+4)(n-k+6)} E_{R,\theta}\big[\|X - \varphi_0\|^8 \Delta\gamma \circ \varphi_0\big] \end{aligned}$$

and, using Lemma A.2, that

$$\begin{aligned} &2E_{R,\theta}\big[\|X - \varphi_0\|^6 \langle \varphi_0 - \theta, \gamma \circ \varphi_0 \cdot g \circ \varphi_0 \rangle\big] \\ &\quad = \frac{2}{n-k+6} E_{R,\theta}\big[\|X - \varphi_0\|^8 \operatorname{div}(\gamma \circ \varphi_0 \cdot g \circ \varphi_0)\big]. \end{aligned}$$

Thus it follows that (3.2) equals

$$E_{R,\theta}\Big[\big(\lambda_0^s - \|\varphi - \theta\|^2\big)^2\Big] + E_{R,\theta}\big[\|X - \varphi_0\|^8 \gamma^2 \circ \varphi_0\big]$$
$$- \frac{8k}{(n-k)(n-k+4)} E_{R,\theta}\big[\|X - \varphi_0\|^6 \gamma \circ \varphi_0\big]$$
$$+ E_{R,\theta}\bigg[\|X - \varphi_0\|^8 \bigg\{ \frac{4}{n-k+2} \gamma \circ \varphi_0 \operatorname{div} g \circ \varphi_0 - \frac{4}{n-k+6} \operatorname{div}(\gamma \circ \varphi_0 \cdot g \circ \varphi_0) \bigg\}\bigg]$$
$$+ \frac{2}{(n-k+4)(n-k+6)} E_{R,\theta}\big[\|X - \varphi_0\|^8 \, \Delta\gamma \circ \varphi_0\big].$$

Since the function γ is positive, the third term on the right-hand side is negative. The same term $\|X - \varphi_0\|^8$ occurs in the other expressions; hence, it is clear it is sufficient that

$$\gamma^2 + \frac{4}{n-k+2} \gamma \operatorname{div} g - \frac{4}{n-k+6} \operatorname{div}(\gamma \circ g) + \frac{2}{(n-k+4)(n-k+6)} \Delta\gamma \le 0$$

in order that the inequality $R(\lambda^s, \theta, \varphi) \le R(\lambda_0^s, \theta, \varphi)$ holds. □

REMARK 3.1. We can give the same comment we do in the remark following Theorem 2.1. The power q occurring for the residual term $\|X - \varphi_0\|$ is necessarily $q = 4$.

EXAMPLE 3.1. Let us consider the usual shrinkage estimator φ of θ with the shrinkage factor g defined by $g(t) = ct/\|t\|^2$, where c is a positive constant, and the shrinkage loss estimator used in the previous example with the shrinkage function γ defined by $\gamma(t) = d/\|t\|^2$, where d is a positive constant. For every $t \in \Theta$, it is easy to check that

$$\begin{aligned} \operatorname{div} g(t) &= c(k-2)/\|t\|^2, \\ \nabla\gamma(t) &= -2\, dt/\|t\|^4, \\ \Delta\gamma(t) &= -2d(k-4)/\|t\|^4 \end{aligned}$$

and

$$\begin{aligned} \operatorname{div}(\gamma \cdot g)(t) &= \gamma(t) \operatorname{div} g(t) + \langle \nabla\gamma(t), g(t) \rangle \\ &= cd(k-4)/\|t\|^4. \end{aligned}$$

Hence the sufficient condition of Theorem 3.1 becomes (simplifying by $\|t\|^4$)

$$d^2 + \frac{4}{n-k+2} cd(k-2) - \frac{4}{(n-k+6)} cd(k-4) - \frac{2}{(n-k+4)(n-k+6)} 2d(k-4) \leq 0.$$

The optimal d can be determined as in Example 2.1; it is given by

$$d^* = \frac{2(k-4)}{(n-k+4)(n-k+6)} - 2c\left(\frac{k-2}{n-k+2} - \frac{k-4}{n-k+6}\right).$$

It is clear that the optimal shrinkage factor of the loss estimator depends on the shrinkage factor of the point estimate. We found that with the optimal c, that is, $c = (k-2)/(n-k+2)$, the optimal d is still positive for reasonable values of n and k.

4. Concluding remarks.

4.1. *About the correction γ given by the improved estimator.* In the normal case, the correction to the unbiased estimator of loss is downward (that is, the corresponding function γ is positive) when the loss of the least squares estimator is considered and upward (γ is negative) when the loss of a shrinkage estimator is estimated. Here, in the spherical case, we always need γ to be positive. A finer analysis of Johnstone's proof indicates that, when using the classical James–Stein point estimate, in the normal case, dominating loss estimators may be constructed with both positive and negative shrinkage factors. It turns out that the optimal factor, for his special case, is positive.

We think that the sign of the correction in the spherical case is due to the fact that we correct λ_0 instead of $\bar{\lambda}$. Actually, the possible correction of $\bar{\lambda}$ for the least squares estimator seems difficult to obtain (except for the normal case) and is less interesting since, as previously noted, it means a lack of robustness. Indeed conditionally on the radius R, the risk of λ is equal to

$$E_{R,\theta}\left[\left(\lambda - \|\varphi_0 - \theta\|^2\right)^2\right] = E_{R,\theta}\left[\left(\bar{\lambda} - \|\varphi_0 - \theta\|^2\right)^2\right] + E_{R,\theta}\left[\|X - \varphi_0\|^8 \gamma^2 \circ \varphi_0\right] - 2E_{R,\theta}\left[\left(\bar{\lambda} - \|\varphi_0 - \theta\|^2\right)\|X - \varphi_0\|^4 \gamma \circ \varphi_0\right].$$

Using Lemma A.1 as in the proof of Theorem 2.1, we get that the risk difference between λ and $\bar{\lambda}$ is

$$-2\bar{\lambda} E_{R,\theta}\left[\|X - \varphi_0\|^4 \gamma \circ \varphi_0\right] + \frac{2k}{n-k+4} E_{R,\theta}\left[\|X - \varphi_0\|^6 \gamma \circ \varphi_0\right] + E_{R,\theta}\left[\|X - \varphi_0\|^8 \gamma^2 \circ \varphi_0\right] + \frac{2}{(n-k+4)(n-k+6)} E_{R,\theta}\left[\|X - \varphi_0\|^8 \Delta\gamma \circ \varphi_0\right].$$

The difference with the expression of the risk difference between $\lambda_0 - \|X - \varphi_0\|^4 \gamma \circ \varphi_0$ and λ_0 comes from the fact that, here, we get one term in $\|X - \varphi_0\|^4$ and one term in $\|X - \varphi_0\|^6$ when, in the previous case, we got one term in $\|X - \varphi_0\|^6$. In the proof of Theorem 2.1 we needed to assume that γ is positive so that the nonhomogenous term in (2.5) could be deleted. It is important to notice that, in the normal case, the sum of these two terms vanishes. Hence we do not need to assume γ is positive. We can show that the corresponding risk difference, in the normal case, equals

$$(n-k+6)(n-k+4)(n-k+2)(n-k)E_\theta[\gamma^2 \circ \varphi_0] + 2(n-k+2)(n-k)E_\theta[\Delta \circ \varphi_0].$$

This is the general linear model analog of Johnstone's formula. Here the calculations were possible because we used the independence between $\|X - \varphi_0\|$ and $\gamma \circ \varphi_0$. For other distributions, an improvement on $\bar{\lambda}$ seems difficult to get. We expect that a similar rationale applies to the estimation of the loss of shrinkage estimators. However, in this case the computations become unwieldy.

4.2 *Improvement using positive rules.* A possible problem with the estimators (2.1) and (3.1) is that they may be negative, which should not happen since we are estimating a nonnegative quantity. A simple remedy to this problem is to use the positive-part estimators. Indeed it is easy to see this solution works for every location estimator φ of θ, for every nonnegative loss function $L(\cdot,\cdot)$ and for every loss estimator δ. More generally, if we define the positive-part estimator δ^+ as $\delta^+ = \max\{\delta, 0\}$, the loss difference between δ^+ and δ is

$$\begin{aligned}(\delta - L(\theta, \varphi(x)))^2 &- (\delta^+ - L(\theta, \varphi(x)))^2 \\ &= (\delta^2 - 2\delta(x)L(\theta, \varphi(x)))\mathbb{1}_{[\delta \le 0]}(x);\end{aligned}$$

hence it is always nonnegative. Therefore the risk difference is positive, which implies that δ^+ dominates δ. This type of result is in the same spirit of the positive-part James–Stein estimator of location. Note that the proof of this result does not depend on the underlying distribution, only on the form of the risk function.

APPENDIX

A.1. This Appendix yields two technical lemmas that are used repeatedly. Before giving them in Section A.2, we state in this section some preliminary results related to the uniform distribution on a sphere.

With the notations introduced in Section 1, recall the fact that although $U_{R,\theta}$ is a singular distribution, its image by the orthogonal projector φ_0:

$(E \to \Theta)$ has a density with respect to Lebesgue measure on Θ, which is equal to

$$C_R^{n,k}\left(R^2 - \|t-\theta\|^2\right)^{(n-k)/2-1} \mathbb{1}_{B_{R,\theta}}(t),$$

where

$$C_R^{n,k} = \frac{\Gamma(n/2)R^{2-n}}{\Gamma((n-k)/2)\pi^{k/2}}$$

and

$$B_{R,\theta} = \{t \in \Theta / \|t-\theta\| \le R\}.$$

The use of the density of the projection is a powerful tool when dealing with expectations, with respect to $U_{R,\theta}$, of random variables which depend on the observation only through the projectors φ_0 and $X - \varphi_0$. It can be shown that the second moment of the noncentral generalized F-distributed random variable $E_{R,\theta}[(\|X-\varphi_0\|^2/\|\varphi_0\|^2)^2]$ (up to the degrees of freedom $n-k$ and k) is bounded from above by a constant independent of R as soon as $k > 4$. Indeed, using the radial density of the image distribution $\varphi_0(U_{R,\theta})$, we have

$$\text{(A.1)}\qquad \begin{aligned} m_{R,\theta} &= E_{R,\theta}\left[\left(\frac{\|X-\varphi_0\|^2}{\|\varphi_0\|^2}\right)^2\right] \\ &= \frac{2\Gamma(n/2)R^{2-n}}{\Gamma((n-k)/2)\pi^{k/2}} \int_0^R r^{k-1}(R^2 - r^2)^{(n-k)/2+1} E_{r,\theta}\left[\|\cdot\|^{-4}\right] dr. \end{aligned}$$

When $\theta = 0$, the equality $E_{r,\theta}[\|\cdot\|^{-4}] = r^{-4}$ leads to the fact that $m_{R,0}$ is proportional to the beta function $B((k-4)/2, (n-k+4)/2)$ and finally equals

$$m_{R,0} = \frac{(n-k)(n-k+2)}{(k-2)(k-4)},$$

which requires $k > 4$ and proves that $m_{R,0}$ is independent of R.

When $\theta \neq 0$, we have

$$E_{r,\theta}\left[\|\cdot\|^{-4}\right] = \int_{\langle\theta\rangle}\left[\int_{\langle\theta\rangle^\perp} \frac{1}{\left(\|u\|^2 + \|v\|^2\right)^2} U_{r,\theta}(du/\pi_\theta = v)\right]\pi_\theta(U_{r,\theta})(dv),$$

where $\langle\theta\rangle$ is the one-dimensional linear subspace spanned by θ, $\langle\theta\rangle^\perp$ is its $(k-1)$-dimensional orthogonal subspace, π_θ is the orthogonal projector onto $\langle\theta\rangle$ and $U_{r,\theta}(\cdot/\pi_\theta = v)$ is the conditional probability of $U_{r,\theta}$ given $\pi_\theta = v$. Because

$$U_{r,\theta}(\cdot/\pi_\theta = v) = U_{\sqrt{r^2 - \|v-\theta\|^2},0} \otimes \delta_v$$

(where δ_v is the Dirac measure at v), the expression of the density of

$\pi_\theta(U_{R,\theta})$ yields

$$E_{r,\theta}\big[\|\cdot\|^{-4}\big] = \frac{\Gamma(k/2)r^{2-k}}{\Gamma((k-1)/2)\pi^{1/2}}\int_{B_{R,\theta}} \frac{\big(r^2 - \|v - \theta\|^2\big)^{(k-1)/2-1}}{\big(r^2 - \|v-\theta\|^2 + \|v\|^2\big)^2}\,dv$$

$$\leq \frac{\Gamma(k/2)r^{2-k}}{\Gamma((k-1)/2)\pi^{1/2}}\int_{B_{R,\theta}} \big(r^2 - \|v-\theta\|^2\big)^{[(k-4)-1]/2-1}\,dv.$$

Noticing that the integrand of the last integral is (up to a multiplicative constant) the density of the orthogonal projector from a $(k-4)$-dimensional space onto the one-dimensional space $\langle\theta\rangle$, it is clear, after simplification, that

$$\text{(A.2)}\qquad E_{r,\theta}\big[\|\cdot\|^{-4}\big] \leq \frac{(k-2)(k-4)}{(k-3)(k-5)}r^{-4},$$

which is valid for $k > 5$.

Therefore, following the same arguments used in the calculation of $m_{R,0}$, (A.1) and (A.2) show that

$$m_{R,\theta} = \frac{(n-k)(n-k+2)}{(k-3)(k-5)},$$

as soon as $k > 5$.

The case $k = 5$ is somewhat particular and we need to compute the exact expression of $m_{R,\theta}$ in (A.1). Although this calculation is tedious, it can be done. It rests on the fact that the moment $E_{r,\theta}[\|\cdot\|^{-4}] = r^{-4}$ can be expressed in term of a hypergeometric function and behaves as an r^{-4} term. Hence the previous conclusion holds for $k = 5$. Therefore, (2.3) is finite for $k > 4$.

A.2. The proof of the two lemmas below repeatedly uses the divergence theorem. Recall it is valid in the context of weak differentiability of a vector-valued function $\delta = (\delta_1, \cdots, \delta_n)$ from E into E (i.e., δ_i is weakly differentiable for every i); see [21] for a full account.

Following the notations introduced in Section 2.3, we denote by $\operatorname{div}\delta = \sum_{i=1}^n D_i\delta_i = \sum_{i=1}^n \partial\delta_i/\partial x_i$ the divergence of δ.

LEMMA A.1. *For every twice weakly differentiable function $\gamma(\Theta \to \mathbb{R}_+)$ and for every integer q,*

$$E_{R,\theta}\big[\|X - \varphi_0\|^q\|\varphi_0 - \theta\|^2\gamma\circ\varphi_0\big]$$

$$= \frac{k}{n-k+q}E_{R,\theta}\big[\|X-\varphi_0\|^{q+2}\gamma\circ\varphi_0\big]$$

$$+ \frac{1}{(n-k+q)(n-k+q+2)}E_{R,\theta}\big[\|X-\varphi_0\|^{q+4}\,\Delta\gamma\circ\varphi_0\big].$$

PROOF. We have

$$E_{R,\theta}\left[\|X - \varphi_0\|^q \|\varphi_0 - \theta\|^2 \gamma \circ \varphi_0\right]$$
$$= C_R^{n,k} \int_{B_{R,\theta}} \left(R^2 - \|t - \theta\|^2\right)^{q/2} \|t - \theta\|^2 \gamma(t) \left(R^2 - \|t - \theta\|^2\right)^{(n-k)/2 - 1} dt$$
$$= C_R^{n,k} \int_{B_{R,\theta}} \left\langle \left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2-1} (t - \theta), \gamma(t)(t - \theta) \right\rangle dt.$$

Now we can notice that, for every $t \in \Theta$,

$$\left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2-1} (t - \theta) = \nabla f(t)$$

with

$$f(t) = -\frac{\left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2}}{n - k + q}.$$

Since, for every $t \in \Theta$,

$$\langle \nabla f(t), \gamma(t)(t - \theta) \rangle = \operatorname{div}(f(t)\gamma(t)(t - \theta)) - f(t)\operatorname{div}(\gamma(t)(t - \theta)),$$

then

$$E_{R,\theta}\left[\|X - \varphi_0\|^q \|\varphi_0 - \theta\|^2 \gamma \circ \varphi_0\right]$$
$$= \frac{C_R^{n,k}}{n - k + q} \int_{B_{R,\theta}} \operatorname{div}(\gamma(t)(t - \theta)) \left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2} dt$$
$$+ C_R^{n,k} \int_{B_{R,\theta}} \operatorname{div}(f(t)\gamma(t)(t - \theta))\, dt.$$

Now the last integral is null since, by the divergence theorem,

$$\int_{B_{R,\theta}} \operatorname{div}(f(t)\gamma(t)(t - \theta))\, dt = \int_{S_{R,\theta}} \left\langle f(t)\gamma(t)(t - \theta), \frac{t - \theta}{\|t - \theta\|} \right\rangle \sigma_{R,\theta}(dt),$$

where $\sigma_{R,\theta}$ is the area measure on $S_{R,\theta}$, the nullity coming from the fact that the function f is equal to zero on $S_{R,\theta}$. Therefore, upon applying, as above, the expression of the divergence of a product, namely, for every $t \in \Theta$,

$$\operatorname{div}(\gamma(t)(t - \theta)) = k\gamma(t) + \langle \nabla\gamma(t), t - \theta \rangle,$$

we obtain

$$E_{R,\theta}\left[\|X - \varphi_0\|^q \|\varphi_0 - \theta\|^2 \gamma \circ \varphi_0\right]$$
$$= \frac{kC_R^{n,k}}{n - k + q} \int_{B_{R,\theta}} \gamma(t) \left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2} dt$$
$$+ \frac{C_R^{n,k}}{n - k + q} \int_{B_{R,\theta}} \langle \nabla\gamma(t), (t - \theta) \rangle \left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2} dt.$$

The last integral can be treated as in the beginning of the proof. Indeed if, for every $t \in \Theta$, we denote by

$$F(t) = -\frac{\left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2+1}}{n - k + q + 2},$$

then

$$\nabla F(t) = \left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2}(t - \theta)$$

and we have

$$\int_{B_{R,\theta}} \langle \nabla\gamma(t), t - \theta\rangle\left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2} dt = \int_{B_{R,\theta}} \langle \nabla\gamma(t), \nabla F(t)\rangle\, dt.$$

Now using again the expression of the divergence of a product and noticing that, for every $t \in \Theta$, $\operatorname{div}(\nabla\gamma)(t) = \Delta\gamma(t)$, we obtain

$$\int_{B_{R,\theta}} \langle \nabla\gamma(t), \nabla F(t)\rangle\, dt = \int_{B_{R,\theta}} \operatorname{div}(F\cdot\nabla\gamma)(t)\, dt - \int_{B_{R,\theta}} F(t)\,\Delta\gamma(t)\, dt.$$

As before, by the divergence theorem,

$$\int_{B_{R,\theta}} \operatorname{div}(F\cdot\nabla\gamma)(t)\, dt = \int_{S_{R,\theta}} \left\langle F(t)\,\nabla\gamma(t), \frac{t - \theta}{\|t - \theta\|}\right\rangle \sigma_{R,\theta}(dt) = 0$$

since, for every$t \in S_{R,\theta}$, $F(t) = 0$.

Finally we get

$$\begin{aligned} &E_{R,\theta}\left[\|X - \varphi_0\|^q\|\varphi_0 - \theta\|^2\gamma\circ\varphi_0\right] \\ &\quad = \frac{kC_R^{n,k}}{n - k + q}\int_{B_{R,\theta}} \gamma(t)\left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2} dt \\ &\qquad + \frac{C_R^{n,k}}{(n-k+q)(n-k+q+2)}\int_{B_{R,\theta}} \Delta\gamma(t)\left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2+1} dt, \end{aligned}$$

which is the desired result. □

LEMMA A.2. *For every twice weakly differentible function $h(\Theta \to \Theta)$ and for every integer q,*

$$E_{R,\theta}\left[\|X - \varphi_0\|^q\langle \varphi_0 - \theta, h\circ\varphi_0\rangle\right] = \frac{1}{n - k + q}E_{R,\theta}\left[\|X - \varphi_0\|^{q+2}\operatorname{div} h\circ\varphi_0\right].$$

PROOF. We have

$$\begin{aligned} &E_{R,\theta}\left[\|X - \varphi_0\|^q\langle \varphi_0 - \theta, h\circ\varphi_0\rangle\right] \\ &\quad = C_R^{n,k}\int_{B_{R,\theta}} \left\langle\left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2-1}(t - \theta), h(t)\right\rangle dt. \end{aligned}$$

As in Lemma A.1, taking, for every $t \in \Theta$,

$$f(t) = -\frac{\left(R^2 - \|t - \theta\|^2\right)^{n-k+q}}{n - k + q},$$

we obtain

$$\begin{aligned} E_{R,\theta}&\left[\|X - \varphi_0\|^q \langle \varphi_0 - \theta, h \circ \varphi_0 \rangle\right] \\ &= \frac{C_R^{n,k}}{n-k+q} \int_{B_{R,\theta}} \operatorname{div} h(t) \left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2} dt \\ &\quad + C_R^{n,k} \int_{B_{R,\theta}} \operatorname{div}(f \cdot h)(t)\, dt. \end{aligned}$$

For a reason given in the proof of Lemma A.1, the second integral on the right-hand side is null, so

$$\begin{aligned} E_{R,\theta}&\left[\|X - \varphi_0\|^q \langle \varphi_0 - \theta, h \circ \varphi_0 \rangle\right] \\ &= \frac{C_R^{n,k}}{n-k+q} \int_{B_{R,\theta}} \operatorname{div} h(t) \left(R^2 - \|t - \theta\|^2\right)^{(n-k+q)/2} dt, \end{aligned}$$

which is the desired result. □

LEMMA A.3. *For every integer $j \geq 1$,*

$$E_{R,\theta}\left[\|X - \varphi_0\|^{2j}\right] = R^{2j} \prod_{i=1}^{j} \frac{(n-k)/2 + j - i}{n/2 + j - i}.$$

PROOF. We have

$$\begin{aligned} E_{R,\theta}\left[\|X - \varphi_0\|^{2j}\right] &= C_R^{n,k} \int_{B_{R,\theta}} \left(R^2 - \|t - \theta\|^2\right)^{(n-k)/2+j-1} dt \\ &= C_R^{n,k} \int_{B_{R,\theta}} \left(R^2 - \|t - \theta\|^2\right)^{((n+2j)-k)/2-1} dt \\ &= C_R^{n,k} \times \left(C_R^{n+2j,k}\right)^{-1} \\ &= \frac{\Gamma(n/2) R^{2-n}}{\Gamma((n-k)/2)\pi^{k/2}} \times \frac{\Gamma((n+2j-k)/2)\pi^{k/2}}{\Gamma((n+2j)/2) R^{2-n-2j}} \\ &= R^{2j} \frac{\Gamma(n/2)}{\Gamma(n/2+j)} \times \frac{\Gamma((n-k)/2+j)}{\Gamma((n-k)/2)} \\ &= R^{2j} \prod_{i=1}^{j} \frac{(n-k)/2 + j - i}{n/2 + j - i}. \end{aligned}$$

□

Acknowledgments. The authors wish to thank George Casella, Gene Hwang and Bill Strawderman for helpful discussions; the first author wishes particularly to thank them for their pleasurable hospitality. We are grateful to Christian Robert for a careful reading of the last version of this paper. We also thank the Associate Editor and the two referees for their insightful comments.

REFERENCES

[1] BERGER, J. O. (1976). Inadmissibility results for generalized Bayes estimators of coordinates of a location parameter. *Ann. Statist.* **4** 302–333.

[2] BERGER, J. O. (1985). The frequentist viewpoint and conditioning. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. Le Cam and R. Olshen, eds.) 15–44. Wadsworth, Belmont, CA.

[3] BRANDWEIN, A. C. and STRAWDERMAN, W. E. (1991). Generalizations of James–Stein estimators under spherical symmetry. *Ann. Statist.* **19** 1639–1650.

[4] CASELLA, G. (1988). Conditionally acceptable frequentist solutions (with discussion). In *Statistical Decision Theory and Related Topics 4* (S. S. Gupta and J. O. Berger, eds.) **1** 73–117. Springer, New York.

[5] CASELLA, G., HWANG, J. T. G. and ROBERT, C. (1994). Loss functions for set estimation. In *Statistical Decision Theory and Related Topics 5* (J. O. Berger and S.S. Gupta, eds.) 237–252. Springer, New York.

[6] CELLIER, D. and FOURDRINIER, D. (1995). Shrinkage estimators under spherical symmetry for the general linear model. *J. Multivariate Anal.* To appear.

[7] CELLIER, D., FOURDRINIER, D. and ROBERT, C. (1989). Robust shrinkage estimators of the location parameter for elliptically symmetric distribution. *J. Multivariate Anal.* **29** 39–52.

[8] FANG, T. K. KOTZ, S. and NG, K. W. (1990). *Symmetric Multivariate and Related Distributions.* Chapman and Hall, New York.

[9] HUBER, P. (1981). *Robust Statistics.* Wiley, New York.

[10] JOHNSTONE, I. (1988). On inadmissibility of some unbiased estimates of loss. In *Statistical Decision Theory and Related Topics 4* (S. S. Gupta and J. O. Berger, eds.) **1** 361–379. Springer, New York.

[11] KIEFER, J. (1975). Conditional confidence approach in multi-decision problems. In *Multivariate Analysis 4* (P. R. Krishnaiah, ed.). Academic Press, New York.

[12] KIEFER, J. (1976). Admissibility of conditional confidence procedures. *Ann. Statist.* **4** 836–865.

[13] KIEFER, J. (1977). Conditional confidence statements and confidence estimators. *J. Amer. Statist. Assoc.* **72** 789–827.

[14] LEHMANN, E. L. and SCHEFFÉ, H. (1950). Completeness, similar regions and unbiased estimates. *Sankhyā* **10** 305–340.

[15] LELE, C. (1993). Admissibility results in loss estimation. *Ann. Statist.* **21** 378–390.

[16] LU, K. L. and BERGER, J. O. (1989). Estimation of normal means: Frequentist estimation of loss. *Ann. Statist.* **17** 890–906.

[17] ROBERT, C. and CASELLA, G. (1990). Improved confidence sets for spherically symmetric distributions. *J. Multivariate Anal.* **32** 84–94.

[18] RUKHIN, A. L. (1988). Estimated loss and admissible loss estimators. In *Statistical Decision Theory and Related Topics 4* (S. S. Gupta and J. O. Berger, eds.) **1** 409–418. Springer, New York.

[19] STEIN, C. (1964). Inadmissibility of the usual estimator of the variance of the normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160.

[20] STONE, M. (1987). *Coordinate-Free Multivariate Statistics.* Oxford Univ. Press.

[21] TEMAM, R. (1979). *Navier-Stokes Equations.* North-Holland, Amsterdam.

[22] ZELLNER, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error term. *J. Amer. Statist. Assoc.* **71** 400–405.

[23] ZIEMER, W. P. (1989). *Weakly Differentiable Functions—Sobolev Spaces and Functions of Bounded Variation.* Springer, New York.

UNIVERSITÉ DE ROUEN
UFR DES SCIENCES–MATHÉMATIQUES
URA CNRS 1378
76821 MONT SAINT AIGNAN CEDEX
FRANCE

STATISTICS CENTER
CORNELL UNIVERSITY
ITHACA, NEW YORK