AISM (2006) 58: 73–92
DOI 10.1007/s10463-005-0001-0

Dominique Fourdrinier · William E. Strawderman · Martin T. Wells

Estimation of a location parameter with restrictions or "vague information" for spherically symmetric distributions

Received: 29 May 2003 / Revised: 2 December 2004 / Published online: 22 February 2006

Abstract In this article we consider estimating a location parameter of a spherically symmetric distribution under restrictions on the parameter. First we consider a general theory for estimation on polyhedral cones which includes examples such as ordered parameters and general linear inequality restrictions. Next, we extend the theory to cones with piecewise smooth boundaries. Finally we consider shrinkage toward a closed convex set K where one has vague prior information that θ is in K but where θ is not restricted to be in K. In this latter case we give estimators which improve on the usual unbiased estimator while in the restricted parameter case we give estimators which improve on the projection onto the cone of the unbiased estimator. The class of estimators is somewhat non-standard as the nature of the constraint set may preclude weakly differentiable shrinkage functions. The technique of proof is novel in the sense that we first deduce the improvement results for the normal location problem and then extend them to the general spherically symmetric case by combining arguments about uniform distributions on the spheres, conditioning and completeness.

Keywords Convex cones · Integration-by-parts · Minimax estimate · Multivariate normal mean · Polyhedral cones · Positively homogeneous set · Quadratic loss · Spherically symmetric distribution · Weakly differentiable function

D. Fourdrinier (✉)
UMR CNRS 6085, Université de Rouen,
Avenue de l'Université, B.P.12, 76801 Saint-Etienne-du-Rouvray, France
E-mail: Dominique.Fourdrinier@univ-rouen.fr

W.E. Strawderman
Department of Statistics, Rutgers University,
New Brunswick, NJ 08903, USA

M.T. Wells
Department of Social Statistics, Cornell University,
364 Ives Hall, Ithaca, NY 14853, USA

1 Introduction

This paper has two principal goals. The first is to demonstrate in some generality that when an unbiased estimator of risk exists in a p-variate normal problem with mean vector θ and known covariance matrix $\sigma^2 I$, then a related unbiased estimator of risk exists in the case of a $p+k$ dimensional vector (X, U) spherically symmetric about $(\theta, 0)$, that is, the distribution of $(X-\theta, U)$ is orthogonally invariant. Hence the model can be viewed as an extension of the canonical form of the general linear model to a spherically symmetric error distribution. Here $\dim X = \dim \theta = p$ and $\dim U = \dim 0 = k$. We use the notation $(X, U) \sim SS_{p+k}(\theta, 0)$ throughout; see also Cellier and Fourdrinier (1995), Fourdrinier and Wells (1995) and Ouassou and Strawderman (2002). The second goal is to apply this technique to the estimation of mean vector θ when θ is restricted to a closed convex cone $\mathcal{C}$. Our goal will be to improve on the "usual" estimator, $\delta_0(X, U) = P_{\mathcal{C}} X$, the projection of X onto $\mathcal{C}$. This estimator is the restricted MLE in the normal case and a natural estimator in general.

The problem of improved estimation of θ when it is restricted to a cone $\mathcal{C}$ is technically challenging, even in the normal case, since the development of unbiased estimators of risk must often deal with non-weakly differentiable functions. Stein's (1981) classical integration-by-parts technique assumes weak differentiability, and hence a replacement or modification to Stein's technique is often needed. The situation in the general spherically symmetric case is at least (and typically more) challenging than in the normal case. Cellier and Fourdrinier (1995) give a basic result which extends Stein's result to this case but they too require weak differentiability.

We give, in Sect. 2, a general technical result which implies that if an unbiased estimator of risk exists in the normal case (regardless of weak differentiability), then a related unbiased estimator of risk also exists in the general spherically symmetric case.

The bulk of the remainder of the paper is devoted to developing unbiased estimators of risk in the normal case (or describing existing ones) and extending the domination results to the general spherically symmetric case. The resulting generalizations have a very strong robustness property. The improved estimators dominate $\delta_0(X, U) = P_{\mathcal{C}} X$ simultaneously and uniformly for all spherically symmetric distributions for which $P_{\mathcal{C}} X$ has finite risk. We will also consider, in Sect. 6, the problem of incorporating vague prior information by shrinking toward a closed convex set.

As an elementary example of a restricted parameter space, suppose X follows a $N_p(\theta, I)$ distribution with $\theta \in \{\theta \mid \theta_i \geq 0,\ i = 1, \ldots, p\} = \mathbb{R}_+^p$. Then the goal would be to study estimates of θ based on the observed value of X. For instance, in the case when $p = 1$, the maximum likelihood estimate $X_+ = \max(X, 0)$ is the most natural estimate of θ. It can be shown that X_+ dominates X in terms of quadratic loss. A simple calculation yields that the risk of X_+ is

$$\mathcal{R}(\theta, X_+) = \theta^2(1 - \Phi(\theta)) - \theta\varphi(\theta) + \Phi(\theta), \tag{1}$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf, respectively. Note that the risk in Eq. (1) is increasing in θ and has its minimum value of $1/2$ at $\theta = 0$ and maximum value 1 at $\theta = +\infty$. Although X_+ does not have constant risk, it is still minimax since (see below) X is itself minimax and is improved upon by X_+. Clearly the estimate X_+ is not a Bayes estimate and hence is not admissible since it is not analytic. However, Katz (1961) showed that the estimate $\delta_K^0(X) = X + \lambda(X)$, where $\lambda(X) = \varphi(X)/\Phi(X)$, is an admissible minimax estimator of θ. This is just the Bayes estimate of θ using the Lebesgue prior on $[0, \infty)$ (see Berger, 1985, p.135).

In the case $\theta \in \mathbb{R}_+^p$ for $p \geq 1$, it can be shown that $X = (X_1, \ldots, X_p)^{\mathrm{T}}$ is minimax on $\mathbb{R}_+^p$ using a limit of translated priors argument. Therefore, as $X_+ = (X_{1+}, \ldots, X_{p+})$ improves on X on $\mathbb{R}_+^p$, it is also minimax (although its risk is not at all constant). Indeed a simple argument shows that an improved estimator can be obtained using the positive part of any estimator (a positive part estimate is reasonable since shrinkage of X_+ may pull the estimate to be negative). Formally, under quadratic loss, the loss difference between any estimator δ and its positive part δ_+ is

$$\|\delta - \theta\|^2 - \|\delta_+ - \theta\|^2 = \sum_{i=1}^{p} \delta_i^2 \, \mathbb{1}_{[\delta_i < 0]} - 2 \sum_{i=1}^{p} \theta_i \delta_i \, \mathbb{1}_{[\delta_i < 0]} \geq 0.$$

Hence δ_+ dominates δ. This gives a simple rationale why one should use the maximum likelihood estimator X_+ rather than X.

Chang (1982) considered domination of a coordinate-wise estimator of the form $\delta^0(X) = (\delta_1^0(X), \ldots, \delta_p^0(X))^{\mathrm{T}}$ with

$$\delta_i^0(X) = X_i + g(X_i) \qquad i = 1, \ldots, p, \tag{2}$$

where g is an arbitrary real-valued function. If $g(x) = -x \, \mathbb{1}_{[x \leq 0]}$, then we recover the X_+ estimate. If $g(x) = \lambda(x)$, for $p = 1$, we get the admissible Katz estimate. In particular, Chang showed that for $p \geq 3$, the estimator $\delta^{\mathrm{c}}(X) = (\delta_1^{\mathrm{c}}(X), \ldots, \delta_p^{\mathrm{c}}(X))^{\mathrm{T}}$ defined coordinate-wise as

$$\delta_k^{\mathrm{c}}(X_i) = \begin{cases} X_i + g(X_i) - \frac{cX_i}{\|X\|^2} & \text{if } X_i \geq 0 \ \forall\, i = 1, \ldots, p \\ X_i + g(X_i) & \text{otherwise} \end{cases},$$

where c is a constant (independent of the data) such that $0 < c < 2(p-2)$, dominates the estimate $\delta^0(X)$ under quadratic loss for $g(x) = -x \, \mathbb{1}_{[x \leq 0]}$. Extensions to the case of spherically symmetric distributions for $g(x) = -x \, \mathbb{1}_{[x \leq 0]}$ are presented in Fourdrinier et al. (2003) and Ouassou and Strawderman (2002).

Sengupta and Sen (1991) give some extensions of Chang's work in the normal model. From the perspective of this paper the important advance of Sengupta and Sen's work is the extension of the domain of shrinkage to all orthants such that the number of positive X_is is at least 3 and the concomitant adjustment of the shrinkage factor on each such orthant. In Chang's setting $(X \sim N_p(\theta, I),\ \theta \in \mathbb{R}_+^p)$ the Sengupta and Sen estimator takes the form

$$\delta^{\mathrm{SS}}(X) = \left(1 - \frac{c_i}{\|X_+\|^2}\right) X_+ \quad \text{if } X \in O_i,\ i = 1, \ldots, 2^p,$$

where $O_i,\ i = 1, \ldots, 2^p$ are the orthants, s_i the number of positive coordinates in O_i and $0 < c_i < 2(s_i - 2)_+$. Alternatively the bounds on c_i may be expressed as $0 < c_i < 2(\sum_{i=1}^{p} \mathbb{1}_{[X_i>0]} - 2)_+$. Hence the bound on the shrinkage constant is itself a complicated random variable (a shifted and truncated convolution of non-iid Bernoulli's). The reader is referred to Sengupta and Sen's paper for details and other developments which are beyond the scope of interest of the present paper.

In the above simple version of Sengupta and Sen's setting, our results lead immediately to improved estimators in the spherically symmetric case of the form $(1 - \frac{(s_i-2)_+ U'U}{\|X_+\|^2})X_+$. Section 3 is devoted to a discussion of Stein-type identities for non-weakly differentiable functions in the setting of polyhedral cones. Improved estimation when θ is restricted to a polyhedral cone is the subject of Sect. 4 while Sect. 5 is devoted to several specific examples. As an application of the main result, in Sect. 6 we consider the problem of estimating θ when $\theta \in \mathcal{C}$ is considered as "vague information" rather than as a restriction. Hence θ is unrestricted and the MLE in the normal case is X. It is natural to consider estimators that shrink toward $\mathcal{C}$. Bock (1982) and Kuriki and Takemura (2000) consider such problems. We follow the approach of Kuriki and Takemura in the normal case and extend their results to the general spherically symmetric case. This approach leads also to improved estimators when θ is restricted to a cone with piecewise smooth boundary. The relative ease with which the extension to the spherically symmetric case is obtained is a striking illustration of the utility of the results of Sect. 2. Section 7 gives some comments and indicates possible extension.

2 On Stein-type identities for spherically symmetric distributions

The well-known Stein's (1981) identity

$$E_{(\theta,0)}[(X-\theta)^{\mathrm{T}} g(X)] = \sigma^2 E_{(\theta,0)}[\operatorname{div} g(X)]\,, \tag{3}$$

where $\operatorname{div} g(x) = \sum_{i=1}^{p} \frac{\partial g_i(x)}{\partial x_i}$ holds, if $X \sim N_p(\theta, \sigma^2 I)$ and g is a weakly differentiable function from $\mathbb{R}^p$ into $\mathbb{R}^p$ such that the expectations exist (see Sect. 3 for a precise definition). Cellier and Fourdrinier (1995) extend the identity (3) to the spherically symmetric case in the presence of a residual vector. Specifically, let (X, U) have a spherically symmetric distribution about $(\theta, 0)$ when the dimension of X and the dimension of θ are both equal to p and the dimension of U and the dimension of 0 are both equal to k. We use the notation $(X, U) \sim \mathrm{SS}_{p+k}(\theta, 0)$ throughout this paper. Cellier and Fourdrinier's extension of Eq. (3) is

$$E_{(\theta,0)}\left[\|U\|^2 (X-\theta)^{\mathrm{T}} g(X)\right] = \frac{1}{k+2} E_{(\theta,0)}\left[\|U\|^4 \operatorname{div} g(X)\right]$$

for $(X, U) \sim \mathrm{SS}_{p+k}(\theta, 0)$, again provided g is weakly differentiable and the expectations exist.

However, it can happen that a Stein-type identity holds when g is not weakly differentiable. For example the next section is devoted to such a result in the normal case for functions of the form $g(x)\mathbb{1}_{\mathcal{D}}(x)$. Here, even if g is weakly differentiable, $\mathbb{1}_{\mathcal{D}}(x)$, the indicator function of the restricted set of interest, may make the product non-weakly differentiable.

The purpose of this section is to show that whenever a Stein-type identity holds in the normal case, there is a corresponding result for the spherically symmetric case. An important feature of the result is that g needs not necessarily be weakly differentiable. A practically important corollary of the result is that when Stein's unbiased estimator of risk exists in a particular problem with $X \sim N_p(\theta, \sigma^2 I)$, there is a corresponding risk identity in the spherically symmetric case. It is also important to note that the presence of the residual vector U is essential for the result to hold. Here is the main result of this section.

Theorem 1 *Suppose* $(X, U) \sim \mathrm{SS}_{p+k}(\theta, 0)$. *Assume* g *and* f *are such that in the special case where* $(X, U) \sim N_{p+k}((\theta, 0), \sigma^2 I)$, *we have*

$$E_{(\theta,0)}\left[(X-\theta)^{\mathrm{T}} g(X)\right] = \sigma^2 E_{(\theta,0)}\left[f(X)\right]$$

(*both expected values are assumed to exist*) *for all* $\sigma^2 > 0$. *Then*

$$E_{(\theta,0)}\left[\|U\|^2 (X-\theta)^{\mathrm{T}} g(X)\right] = \frac{1}{k+2} E_{(\theta,0)}\left[\|U\|^4 f(X)\right].$$

Proof Let $X \sim N_p(\theta, \sigma^2 I)$ and $U \sim N_k(0, \sigma^2 I)$ be independent random variables. Then $\|U\|^2 \sim \sigma^2 \chi_k^2$ is also independent of X. Since $E_0\left[\|U\|^2\right] = k\sigma^2$ and $E_0\left[\|U\|^4\right] = k(k+2)\sigma^4$, we have

$$\begin{aligned} E_{(\theta,0)}\left[\|U\|^2 (X-\theta)^{\mathrm{T}} g(X)\right] &= E_0\left[\|U\|^2\right] E_\theta\left[(X-\theta)^{\mathrm{T}} g(X)\right] \\ &= k\sigma^2 E_\theta\left[(X-\theta)^{\mathrm{T}} g(X)\right] \\ &= k\sigma^4 E_\theta\left[f(X)\right] \\ &= \frac{1}{k+2} E_0\left[\|U\|^4\right] E_\theta\left[f(X)\right] \\ &= \frac{1}{k+2} E_{(\theta,0)}\left[\|U\|^4 f(X)\right]. \end{aligned} \tag{4}$$

For each θ (considered fixed), $\|X-\theta\|^2 + \|U\|^2$ is a complete sufficient statistic for the $N_{p+k}\left((\theta, 0), \sigma^2 I\right)$ distribution. Now we have

$$\begin{aligned} &E_{(\theta,0)}[\|U\|^2 (X-\theta)^{\mathrm{T}} g(X)] \\ &\quad = E\left[E_{(\theta,0)}[\|U\|^2 (X-\theta)^{\mathrm{T}} g(X) \mid \|X-\theta\|^2 + \|U\|^2 = R^2]\right] \end{aligned}$$

and

$$\begin{aligned} &\frac{1}{k+2} E_{(\theta,0)}[(\|U\|^4) f(X)] \\ &\quad = \frac{1}{k+2} E\left[E_{\theta,0}[\|U\|^4 f(X) \mid \|X-\theta\|^2 + \|U\|^2 = R^2]\right]. \end{aligned}$$

Hence it follows from Eq. (4) and from the completeness of $\|X-\theta\|^2 + \|U\|^2$ that

$$\begin{aligned} &E_{(\theta,0)}\left[\|U\|^2 (X-\theta)^{\mathrm{T}} g(X) \mid \|X-\theta\|^2 + \|U\|^2 = R^2\right] \\ &\quad = \frac{1}{k+2} E_{(\theta,0)}\left[\|U\|^4 f(X) \mid \|X-\theta\|^2 + \|U\|^2 = R^2\right] \end{aligned}$$

almost everywhere. As shown in the Appendix, these two functions of R are continuous, and therefore they are equal everywhere.

The conditional distribution of (X, U) given $\|X-\theta\|^2+\|U\|^2 = R^2$ is uniform on a sphere centered at $(\theta, 0)$ and of radius R. Hence the above gives equality of the expectations of the functions $\|U\|^2(X-\theta)^{\mathrm{T}}g(X)$ and $\frac{1}{k+2}\|U\|^4 f(X)$ for such uniform distributions.

Now note that if $(X, U) \sim \mathrm{SS}_{p+k}(\theta, 0)$, the conditional distribution of (X, U) given $\|X-\theta\|^2 + \|U\|^2 = R^2$ is uniform on the sphere centered at $(\theta, 0)$ and of radius R. Hence it follows in the general case where $(X, U) \sim \mathrm{SS}_{p+k}(\theta, 0)$ that

$$
\begin{aligned}
&E_{(\theta,0)}\left[\|U\|^2(X-\theta)^{\mathrm{T}}g(X) \mid \|X-\theta\|^2 + \|U\|^2 = R^2\right] \\
&\quad = \frac{1}{k+2}E_{(\theta,0)}\left[\|U\|^4 f(X) \mid \|X-\theta\|^2 + \|U\|^2 = R^2\right]
\end{aligned}
$$

Upon taking the expectation with respect to $\|X-\theta\|^2 + \|U\|^2$ we have the desired results. □

It is interesting to note in this result that the computation for the general spherically symmetric case hinges on the Gaussian result. One can usually deduce integration-by-parts results, for both the Gaussian and spherical cases, via a direct application of Stokes's theorem. Here, since the integrand of interest may not be weakly differentiable, we need the Gaussian result in order to deduce the spherical identity. The conditioning in the proof reduces the distributions of interest to their canonical element as a uniform distribution on the sphere. Once the Gaussian results are reduced to the uniform distribution on the sphere, the corresponding expectations can be extended to the general spherically symmetric distribution via "unconditioning" with respect to the radius of the sphere.

In our examples the function f in Theorem 1 is $\operatorname{div} g$ or $\operatorname{div} g\,\mathbb{1}_C$, the latter may be useful if $g\,\mathbb{1}_C$ is not weakly differentiable. As a first example, consider the case of the famous Stein's (1981) identity which holds for weakly differentiable functions g. One of the main uses of this identity is to deduce an unbiased estimator of risk of the estimator $\delta(X) = X + \sigma^2 g(X)$ of θ where $(X, U) \sim N_{p+k}((\theta, 0),\ \sigma^2 I)$, g is weakly differentiable and loss is $\|\delta-\theta\|^2$. This unbiased estimator of risk is deduced (using Stein's identity) as follows

$$
\begin{aligned}
R(\theta, \delta) &= E_\theta\left[\|\delta(X)-\theta\|^2\right] \\
&= E_\theta\left[\|X+\sigma^2 g(X)-\theta\|^2\right] \\
&= E_\theta\left[\|X-\theta\|^2 + \sigma^4\|g(X)\|^2 + 2\sigma^2(X-\theta)^{\mathrm{T}}g(X)\right] \\
&= p\sigma^2 + \sigma^4 E_\theta[\|g(X)\|^2 + 2\operatorname{div} g(X)]\,.
\end{aligned}
$$

Therefore $p\sigma^2 + \sigma^4[\|g(X)\|^2 + 2\operatorname{div} g(X)]$ is an unbiased estimator of $R(\theta, \delta)$.

Similarly, using Theorem 1, if $(X, U) \sim \mathrm{SS}_{p+k}(\theta, 0)$ and $\delta(X) = X + \frac{\|U\|^2}{k+2}g(X)$ then

$$
\begin{aligned}
&E_{(\theta,0)}\left[\left\|X+\frac{\|U\|^2}{k+2}g(X)-\theta\right\|^2\right]\\
&\quad= E_{(\theta,0)}\left[\|X-\theta\|^2+\frac{\|U\|^4}{(k+2)^2}\|g(X)\|^2+2\frac{\|U\|^2}{k+2}(X-\theta)^{\mathrm{T}}g(X)\right]\\
&\quad= E_{(\theta,0)}\left[\frac{p}{k}\|U\|^2+\frac{\|U\|^4}{(k+2)^2}\|g(X)\|^2+2\frac{\|U\|^4}{(k+2)^2}\operatorname{div} g(X)\right]\\
&\quad= E_{(\theta,0)}\left[\frac{p}{k}\|U\|^2+\frac{\|U\|^4}{(k+2)^2}\left(\|g(X)\|^2+2\operatorname{div} g(X)\right)\right].
\end{aligned}
$$

Therefore, when $(X, U) \sim \mathrm{SS}_{p+k}(\theta, 0)$, an unbiased estimator of risk is

$$
\frac{p}{k}\|U\|^2+\frac{\|U\|^4}{(k+2)^2}\left(\|g(X)\|^2+2\operatorname{div} g(X)\right).
$$

Hence in either case, if g is weakly differentiable and $\|g(x)\|^2+2\operatorname{div} g(x)<0$ for any $x\in\mathbb{R}^p$, the estimator [respectively, $X+\sigma^2 g(X)$ or $X+\frac{\|U\|^2}{k+2}g(X)$] dominates X provided the expectations exist. This second domination result is very strong since it implies that $X+\frac{\|U\|^2}{k+2}g(X)$ dominates X under squared error loss simultaneously for all spherically symmetric distributions (for which the expectations exist). This is essentially the main result in Cellier and Fourdrinier (1995).

3 Integration-by-part and non-weakly differentiable functions

In this section we derive some new integration-by-parts formulae that are used in the subsequent sections for certain risk calculations. An important point here is that when the parameter space is restricted, one does not have as much freedom in the choice of shrinkage functions and the classical techniques of Stein's (1981) may no longer be applicable. As noted in Sect. 2, the usual proof of domination by a shrinkage estimator involves the development of an unbiased estimator of risk. The main ingredient of this development is the application of integration-by-parts to the cross product term (i.e., the term involving $(X-\theta)^{\mathrm{T}}g(X)$). A key hypothesis in the integration-by-parts is the same as the key hypothesis in Stokes's Theorem. In other words, the integrand involved needs to be weakly differentiable.

Recall (see Ziemer 1989) that a locally integrable function $g\equiv(g_1,\dots,g_p)$: $\mathbb{R}^p\to\mathbb{R}^p$ is weakly differentiable if there exist locally integrable functions denoted by $\frac{\partial g_i(x)}{\partial x_j}$ such that

$$
\int\frac{\partial g_i(x)}{\partial x_j}\phi(x)\mathrm{d}x=-\int g_i(x)\frac{\partial\phi(x)}{\partial x_j}\mathrm{d}x
$$

for all (i, j) and $\phi\in C^\infty(\mathbb{R})$, where $C^\infty(\mathbb{R})$ is the set of infinitely differentiable functions with compact support. For such a function g, Stein's formula states that when $X\sim N_p(\theta,\sigma^2 I)$,

$$
E_\theta[(X-\theta)^{\mathrm{T}}g(X)]=\sigma^2 E_\theta[\operatorname{div} g(X)], \tag{5}
$$

provided these expectations exist. As highlighted by Stein's (1981) and Johnstone (1988), weak differentiability is fundamental to integration-by-part techniques which lie at the heart of modern (quadratic) risk evaluations. Actually, Johnstone (1988), referring to Morrey (1966), notices that this is equivalent to the statement "g is (equivalent to) a function which is absolutely continuous on almost all line segments parallel to the coordinate axes, and has partial derivatives (existing a.e.) which are locally integrable."

In the context of improved estimation for the general spherically symmetric problem, Cellier and Fourdrinier (1995) link Stein's integration-by-parts technique to Stokes's Theorem which naturally involves weakly differentiable functions. When the restrictions on the underlying parameter space preclude the use of weakly differentiable shrinkage functions one has to resort to new techniques of proof to analyze the estimators under consideration. One of the main contributions of the Sengupta and Sen (1991) paper is to provide such a technique in the normal case. Our contribution is to formalize this technique and to extend it to the spherically symmetric case.

Suppose that $\mathcal{C}$ is a positively homogeneous set, that is, for $x \in \mathcal{C}$ it follows that $ax \in \mathcal{C}$ for all $a > 0$. We assume $\mathcal{C}$ is homogeneous throughout this section unless otherwise stated. Our first result gives an extension of Stein's (1981) classical integration-by-parts results for a particular subclass of non-weakly differentiable functions. Specifically, we consider a function of the form $g(x) = xh(\|x\|^2)\mathbb{1}_{\mathcal{C}}(x)$. Note that such a function may fail to be weakly differentiable even for smooth h due to the presence of the indicator function.

Note also that in order to assure the finiteness of the risk of $X+g(X)$, it suffices to assume that $E_\theta[h^2(\|X\|^2)\|X\|^2] < \infty$. The following lemma is stated under this condition.

Lemma 1 *Let $X \sim N_p(\theta, \sigma^2 I)$ and $\mathcal{C}$ be a positively homogeneous set. Then, for any absolutely continuous function h from $\mathbb{R}_+$ to $\mathbb{R}$ such that $\lim_{y\to 0,\infty} h(y)\, y^{(k+p)/2}\mathrm{e}^{-y/2} = 0$ for all $k \geq 0$ and such that $E_\theta[h^2(\|X\|^2)\|X\|^2] < \infty$, we have*

$$E_\theta[h(\|X\|^2)X^{\mathrm{T}}(X-\theta)\mathbb{1}_{\mathcal{C}}(X)] = \sigma^2 E_\theta[\{2\|X\|^2 h'(\|X\|^2) + p\, h(\|X\|^2)\}\mathbb{1}_{\mathcal{C}}(X)]\,. \tag{6}$$

Proof Note that for $\mathcal{C} = \mathbb{R}^p$, Lemma 1 is Stein's identity (5) with $g(x) = xh(\|x^2\|)$. This immediately implies that the expectations in Eq. (6) exist under the assumption $E_\theta\Big[h^2(\|X\|^2)\|X\|^2\Big] < \infty$. Also note that if $E_\theta\,[|g(X)|] < \infty$ for all $\theta \in \mathbb{R}^p$, then

$$E_0\Big[g(X)\mathrm{e}^{X^{\mathrm{T}}\theta/\sigma^2}\Big] = \sum_{k=0}^{\infty} E_0\left[g(X)\frac{(X^{\mathrm{T}}\theta/\sigma^2)^k}{k\,!}\right]$$

(which can be shown through Dominated Convergence Theorem).

Now without loss of generality we assume $\sigma^2 = 1$. Let

$$A_\theta = E_\theta[h(\|X\|^2)X^{\mathrm{T}}(X-\theta)\mathbb{1}_{\mathcal{C}}(X)].$$

Then, using a Taylor series expansion of the exponential function, A_θ can be expressed as

$$A_\theta = (2\pi)^{-p/2}\mathrm{e}^{-\|\theta\|^2/2}\int_{\mathbb{R}^p}\mathrm{e}^{-\|x\|^2/2}\mathrm{e}^{x^{\mathrm{T}}\theta}h(\|x\|^2)(\|x\|^2 - x^{\mathrm{T}}\theta)\mathbb{1}_{\mathcal{C}}(x)\mathrm{d}x$$
$$= \mathrm{e}^{-\|\theta\|^2/2}E_0[h(\|X\|^2)\mathbb{1}_{\mathcal{C}}(X)\sum_{k=0}^{\infty}\frac{(X^{\mathrm{T}}\theta)^k}{k!}(\|X\|^2 - X^{\mathrm{T}}\theta)]\,.$$

Then, as noted above, we have

$$\begin{aligned}A_\theta &= \mathrm{e}^{-\|\theta\|^2/2}\Big\{E_0[h(\|X\|^2)\mathbb{1}_{\mathcal{C}}(X)\|X\|^2] \\ &\quad + \sum_{k=1}^{\infty}\frac{1}{k!}E_0[h(\|X\|^2)\mathbb{1}_{\mathcal{C}}(X)(X^{\mathrm{T}}\theta)^k(\|X\|^2 - k)]\Big\} \\ &= \mathrm{e}^{-\|\theta\|^2/2}\Big\{E_0[h(\|X\|^2)\mathbb{1}_{\mathcal{C}}(X)\|X\|^2] \\ &\quad + \sum_{k=1}^{\infty}\frac{1}{k!}E_0[h(\|X\|^2)\mathbb{1}_{\mathcal{C}}(X)(\frac{X^{\mathrm{T}}\theta}{\|X\|})^k(\|X\|^{2+k} - k\|X\|^k)]\Big\}\,.\end{aligned} \tag{7}$$

By the assumed homogeneity of $\mathcal{C}$ we have $\mathbb{1}_{\mathcal{C}}(X) = \mathbb{1}_{\mathcal{C}}\left(\frac{X}{\|X\|}\right)$. Since $\|X\|$ is independent of $\frac{X}{\|X\|}$ (when $\theta = 0$) it follows that

$$\begin{aligned}A_\theta &= \mathrm{e}^{-\|\theta\|^2/2}\Big\{E_0[\mathbb{1}_{\mathcal{C}}(X)]E_0[h(\|X\|^2)\|X\|^2] \\ &\quad + \sum_{k=1}^{\infty}\frac{1}{k!}E_0[\mathbb{1}_{\mathcal{C}}(X)(\frac{X^{\mathrm{T}}\theta}{\|X\|})^k]E_0[h(\|X\|^2)(\|X\|^{2+k} - k\|X\|^k)]\Big\}.\end{aligned}$$

Now when $\theta = 0$ we have $\|X\|^2 \sim \chi^2_p$ and so, for $d = \frac{1}{2^{p/2}\Gamma(p/2)}$, the last expectation equals

$$\begin{aligned}&d\int_0^\infty h(y)(y^{1+k/2} - ky^{k/2})y^{p/2-1}\mathrm{e}^{-y/2}\mathrm{d}y \\ &= d\int_0^\infty h(y)y^{(k+p)/2}\mathrm{e}^{-y/2}\mathrm{d}y - d\int_0^\infty h(y)ky^{(k+p)/2-1}\mathrm{e}^{-y/2}\mathrm{d}y \\ &= d\int_0^\infty 2\mathrm{e}^{-y/2}[h'(y)y^{(k+p)/2} + \frac{k+p}{2}y^{(k+p)/2-1}h(y)]\mathrm{d}y \\ &\quad - d\int_0^\infty h(y)ky^{(k+p)/2-1}\mathrm{e}^{-y/2}\mathrm{d}y \\ &= d\int_0^\infty \mathrm{e}^{-y/2}[2yh'(y) + p\,h(y)]y^{(k+p)/2-1}\mathrm{d}y \\ &= E_0[(2\|X\|^2h'(\|X\|^2) + p\,h(\|X\|^2))\|X\|^k].\end{aligned}$$

Note that for $k = 0$, we have

$$E_0\left[h(\|X\|^2)\|X\|^2\right] = E_0\left[2\|X\|^2 h'(\|X\|^2) + p\ h(\|X\|^2)\right].$$

Hence

$$\begin{aligned}
A_\theta &= \mathrm{e}^{-\|\theta\|^2/2} \sum_{k=0}^{\infty} \frac{1}{k!} E_0\left[\mathbb{1}_{\mathcal{C}}(X)(\frac{X^{\mathrm{T}}\theta}{\|X\|})^k\right] E_0\Big[(2\|X\|^2 h'(\|X\|^2) \\
&\quad + p\ h(\|X\|^2))\|X\|^k\Big] \\
&= \mathrm{e}^{-\|\theta\|^2/2} \sum_{k=0}^{\infty} E_0\left[(2\|X\|^2 h'(\|X\|^2) + p\ h(\|X\|^2))\mathbb{1}_{\mathcal{C}}(X)\frac{(X^{\mathrm{T}}\theta)^k}{k!}\right] \\
&= E_\theta\left[(2\|X\|^2 h'(\|X\|^2) + p\ h(\|X\|^2))\mathbb{1}_{\mathcal{C}}(X)\right],
\end{aligned}$$

where the final identity follows from the Dominated Convergence Theorem as noted at the beginning of the proof. □

The next result extends Lemma 1 to functions of the form $h(\|PX\|^2)PX\ \mathbb{1}_{\mathcal{D}}(X)$ where P is an orthogonal linear projection of rank s and $\mathcal{D}$ is positively homogeneous in PX for each fixed $P^\perp X$, where $P^\perp = I - P$ is a projection of rank $p - s$ orthogonal to P. In our application in the next section, Lemma 2 will be applied separately for the projection onto each face of a polyhedral cone. Each such projection, restricted to the set, $\mathcal{D}$, which projects onto a particular face, will be equal to an orthogonal linear projection with rank s equal to the dimension of the face and the domain $\mathcal{D}$ will satisfy the positive homogeneity condition.

Lemma 2 *Let $X \sim N_p(\theta, \sigma^2 I)$, P be a linear orthogonal projection of rank s, $3 \le s \le p$. Further let $\mathcal{D}$ be a set such that if $X = PX + P^\perp X$ is in $\mathcal{D}$, then $X' = aPX + P^\perp X$ is in $\mathcal{D}$ for all $a > 0$. Then, for any absolutely continuous function h on $\mathbb{R}_+$ such that $\lim_{y\to 0,\infty} h(y)y^{(j+s)/2}\mathrm{e}^{-y/2} = 0$ for all $j \ge 0$ and such that $E_\theta[h^2(\|PX\|^2)\|PX\|^2] < \infty$,*

$$\begin{aligned}
&E_\theta\left[(X-\theta)^{\mathrm{T}} PXh(\|PX\|^2)\mathbb{1}_{\mathcal{D}}(X)\right] \\
&\quad = \sigma^2 E_\theta\left[\{2\|PX\|^2 h'(\|PX\|^2) + s\ h(\|PX\|^2)\}\mathbb{1}_{\mathcal{D}}(X)\right].
\end{aligned}$$

Proof Note that $(Y_1, Y_2) = (PX, P^\perp X) \sim N((\eta_1, \eta_2), \sigma^2 I)$ where $(P\theta, P^\perp\theta) = (\eta_1, \eta_2)$. Also

$$\begin{aligned}
A(\theta) &= E_\theta[(X-\theta)^{\mathrm{T}} PX\, h(\|PX\|^2)\mathbb{1}_{\mathcal{D}}(X)] \\
&= E_\theta[(PX - P\theta)^{\mathrm{T}} PX\, h(\|PX\|^2)\mathbb{1}_{\mathcal{D}}(X)] \\
&= E_{\eta_1,\eta_2}[(Y_1-\eta_1)^{\mathrm{T}} Y_1\, h(\|Y_1\|^2)\mathbb{1}_{\mathcal{D}'}(Y_1, Y_2)],
\end{aligned}$$

where $\mathcal{D}' = \{(y_1, y_2) \mid (y_1, y_2) = (Px, P^\perp x) \text{ for } x \in \mathcal{D}\}$. Upon conditioning on Y_2,

$$A(\theta) = E_{\eta_2}\left[E_{\eta_1}[\{(Y_1-\eta_1)^{\mathrm{T}} Y_1\, h(\|Y_1\|^2)\mathbb{1}_{\mathcal{D}'}(Y_1, Y_2)\} \mid Y_2]\right].$$

Applying Lemma 1 to $Y_1 \sim N_s(\eta_1, \sigma^2 I)$ (which is independent of $Y_2 \sim N_{p-s}(\eta_2, \sigma^2 I)$) it follows that $A(\theta)$ is equal to

$$\begin{aligned}A(\theta) &= \sigma^2 E_{\eta_2}\left[E_{\eta_1}[\{(2\|Y_1\|^2\, h'(\|Y_1\|^2) + s\ h(\|Y_1\|^2))\mathbb{1}_{\mathcal{D}'}(Y_1, Y_2)\} \mid Y_2]\right] \\ &= \sigma^2 E_\theta[\{2\|PX\|^2\, h'(\|PX\|^2) + s\ h(\|PX\|^2)\}\mathbb{1}_{\mathcal{D}}(X)],\end{aligned}$$

which completes the proof. □

An application of Theorem 1 immediately gives the following general result for spherically symmetric distributions.

Corollary 1 *Let* $(X, U) \sim \mathrm{SS}_{p+k}(\theta, 0)$*, P a linear orthogonal projection (from* $\mathbb{R}^p$*) of rank s,* $3 \le s \le p$*. Further let* $\mathcal{D} \in \mathbb{R}^p$ *be as in Lemma 2. Then, for any absolutely continuous function h on* $\mathbb{R}_+$ *such that* $\lim_{y\to 0,\infty} h(y)\, y^{(j+s)/2} \mathrm{e}^{-y/2} = 0$ *for all* $j \ge 0$ *and such that*

$$E_{(\theta,0)}[h^2(\|PX\|^2)\|PX\|^2\|U\|^4] < \infty,$$

we have

$$\begin{aligned}&E_{(\theta,0)}[\|U\|^2(X-\theta)^{\mathrm{T}} PX\, h(\|PX\|^2)\,\mathbb{1}_{\mathcal{D}}(X)] \\ &\quad = E_{(\theta,0)}\left[\frac{\|U\|^4}{k+2}\{2\|PX\|^2\, h'(\|PX\|^2) + sh(\|PX\|^2)\}\mathbb{1}_{\mathcal{D}}(X)\right]. \qquad (8)\end{aligned}$$

Remark 1 Note that when $P = I$, then $s = p$ and so Corollary 1 also gives an extension of Lemma 1. Furthermore, if $h(t) = \frac{r(t)}{t}$, then Eq. (8) becomes (since $(X-\theta)^{\mathrm{T}} PX = (PX-\theta)^{\mathrm{T}} PX$)

$$\begin{aligned}&E_{(\theta,0)}[\|U\|^2(PX-\theta)^{\mathrm{T}} PX \frac{r\left(\|PX\|^2\right)}{\|PX\|^2}\mathbb{1}_{\mathcal{D}}(X)] \\ &\quad = E_{(\theta,0)}\left[\frac{\|U\|^4}{k+2}\{\frac{(s-2)\, r(\|PX\|^2)}{\|PX\|^2} + 2r'(\|PX\|^2)\}\mathbb{1}_{\mathcal{D}}(X)\right]. \qquad (9)\end{aligned}$$

This form of the identity will be extensively used in the next section. The function $\frac{r(\|PX\|^2)}{\|PX\|^2} PX\,\mathbb{1}_{\mathcal{D}}(X)$ may fail to be weakly differentiable because of the presence of the indicator function $\mathbb{1}_{\mathcal{D}}(X)$. For example, if $p = 2$, $P = I$ and $\mathcal{D} = \{(x_1, x_2) \mid x_1 \ge 0, x_2 \ge 0, x_1 \le x_2\}$, then $x\,\mathbb{1}_{\mathcal{D}}(x)$ is not weakly differentiable. A notable exception to this non-weak differentiability is when $p = 1$, the case $\mathcal{D} = \mathbb{R}_+$ and the function is $x\,\mathbb{1}_{\mathbb{R}_+}(x)$.

4 Improved estimation when θ is restricted to a polyhedral cone

In this section we again study the model $(X, U) \sim \mathrm{SS}_{p+k}(\theta, 0)$ where θ is restricted to lie in a polyhedral cone, $\mathcal{C}$. We take as the usual estimator of θ

$$\delta_0(X, U) = P_{\mathcal{C}} X, \qquad (10)$$

where $P_{\mathcal{C}}X$ is the projection of X onto the cone $\mathcal{C}$. Our goal will be to find estimators which improve on $P_{\mathcal{C}}X$ with respect to the loss $L(\theta, \delta) = \|\delta - \theta\|^2$. Chang (1981, 1982) and Sengupta and Sen (1991) studied estimation of θ restricted to $\mathcal{C}$ when $X \sim N_p(\theta, \sigma^2 I)$. In this case the estimator $\delta_0(X) = P_{\mathcal{C}}X$ is the MLE. In the case where $\theta \in \mathcal{C}$ is "vague information" rather than a true restriction, the usual estimator (the unrestricted MLE) is X, as in Bock (1982). We will discuss this point further in Sect. 6. Our class of improved estimators will be basically Baranchik (1970) type shrinkage estimators of the form

$$\delta(X, U) = \left(1 - \frac{\|U\|^2}{k+2} \frac{(s(X) - 2)_+ r(\|\delta_0(X)\|^2)}{\|\delta_0(X)\|^2}\right) \delta_0(X) \,. \tag{11}$$

Here $s(X) = \dim\ F^{\mathrm{o}}(X)$ where $F^{\mathrm{o}}(X)$ is the relative interior of the face in which $\delta_0(X)$ lies, $0 \le r \le 2$ and r is non-decreasing. (Actually we will allow, in Eq. (12), r to be a different function on different faces.) Hence the shrinkage factor will depend on the dimension of the face onto which X is projected. No shrinkage occurs if the dimension of the face is less than or equal to 2.

We next describe the properties of polyhedral cones which we will use. See Stoer and Witzgall (1970) and Robertson et al. (1988) for extended discussion of polyhedral cones. A polyhedral cone is defined as the intersection of a finite number of half spaces

$$\mathcal{C} = \{x \mid a_i^{\mathrm{T}} x \le 0, i = 1, \dots, m\}.$$

The cone $\mathcal{C}$ is positively homogeneous, closed and convex and for each $x \in \mathbb{R}^p$ there exists a unique point, $P_{\mathcal{C}}x$, in $\mathcal{C}$ such that $\|P_{\mathcal{C}}x - x\| = \inf_{y \in \mathcal{C}} \|y - x\|$.

For simplicity, we assume throughout that $\mathcal{C}$ has a non-empty interior. In this case $\mathcal{C}$ may be partitioned into $\{\mathcal{C}_i,\ i = 0, \dots, n\}$ where $\mathcal{C}_0 = \mathcal{C}^{\mathrm{o}}$, the interior of $\mathcal{C}$, and $\mathcal{C}_i$, $i = 1, \dots, n$, are the relative interiors of the proper faces of $\mathcal{C}$. Further for each set $\mathcal{C}_i$, let $\mathcal{D}_i = P_{\mathcal{C}}^{-1}\mathcal{C}_i$ (the pre-image of C_i under the projection operator) and $s_i = \dim\ \mathcal{C}_i$. Then $\{\mathcal{D}_i,\ i = 0, \dots, n\}$ form a partition of $\mathbb{R}^P$ and $\mathcal{D}_0 = \mathcal{C}_0$.

For each $x \in \mathcal{D}_i$, $P_{\mathcal{C}}x = P_i x$ where P_i is the orthogonal linear projection onto the linear space L_i of dimension s_i spanned by $\mathcal{C}_i$. Also for such x, $P_i^{\perp}x$, the orthogonal linear projection onto $L_i^{\perp}$ is equal to $P_{\mathcal{C}^*}x$ where $\mathcal{C}^*$ is the polar cone corresponding to $\mathcal{C}$. Also if $x \in \mathcal{D}_i$ then $a P_i x + P_i^{\perp} x \in \mathcal{D}_i$ for all $a > 0$, so that $\mathcal{D}_i$ is positively homogeneous in $P_i x$ for each fixed $P_i^{\perp} x$ (see Robertson et al., 1988, Theorem 8.2.7).

The specific form of the estimator (11) that we will consider is the following:

$$\delta(X, U) = \sum_{i=0}^{n} 1\!\!1_{\mathcal{D}_i}(X) \left(1 - \frac{\|U\|^2 (s_i - 2)_+ r_i(\|P_i X\|^2)}{(k+2)\|P_i X\|^2}\right) P_i X \,. \tag{12}$$

Note that since $X \in \mathcal{D}_i$ implies $P_i X = P_{\mathcal{C}} X$ [and $s_i = s(X)$], the estimator in Eq. (12) reduces to Eq. (11) provided $r_i = r$ for all i.

The main result of this section is the following theorem:

Theorem 2 *Let* $(X, U) \sim \mathrm{SS}_{p+k}(\theta, 0)$ *and let* θ *be restricted to a polyhedral cone,* $\mathcal{C}$*, with non-empty interior. Assume also that* r_i *is absolutely continuous and satisfies* $0 \le r_i \le 2$ *and* $r_i' \ge 0$ *for all* i *such that* $s_i > 2$*. Then* $\delta(X, U)$ *in Eq.* (12) *dominates* $\delta_0(X)$*.*

Proof Note that $\delta_0(X) = \sum_{i=0}^{n} 1\!\!1_{\mathcal{D}_i}(X) P_i(X)$. The difference in risk can therefore be expressed as

$$\begin{aligned}\Delta(\theta) &= E_{(\theta,0)}\left[\|\delta(X,U)-\theta\|^2\right] - E_{(\theta,0)}\left[\|\delta_0(X)-\theta\|^2\right] \\ &= E_{(\theta,0)}\left[\sum_{i=0}^{n} 1\!\!1_{\mathcal{D}_i}(X)\left\{\frac{\|U\|^4[(s_i-2)_+]^2 r_i^2(\|P_iX\|^2)}{(k+2)^2\|P_iX\|^2}\right.\right. \\ &\quad \left.\left. -2\frac{\|U\|^2(s_i-2)_+ r_i(\|P_iX\|^2)}{(k+2)\|P_iX\|^2}(P_iX)^{\mathrm{T}}(P_iX-\theta)\right\}\right].\end{aligned}$$

Using Corollary 1 in the form of Remark 1 (and noting that each $\mathcal{D}_i$ satisfies the condition of the corollary) we have that

$$\begin{aligned}\Delta(\theta) &= E_{(\theta,0)}\left[\sum_{i=0}^{n} 1\!\!1_{\mathcal{D}_i}(X)\left\{\frac{\|U\|^4[(s_i-2)_+]^2 r_i^2(\|P_iX\|^2)}{(k+2)^2\|P_iX\|^2)}\right.\right. \\ &\quad \left.\left. -2\left[\frac{\|U\|^4}{(k+2)^2}\left(\frac{[(s_i-2)_+]r_i^2(\|P_iX\|^2)}{\|P_iX\|^2} + 2(s_i-2)_+ r_i'(\|P_iX\|^2)\right)\right]\right\}\right] \\ &\le 0.\end{aligned}$$

□

5 Examples of improved estimation under polyhedral cones restrictions

Example 1 Suppose $\mathcal{C} = \mathbb{R}_+^p$. Then we may describe $\delta_{\mathcal{C}}^0(X) = X_+ = (X_{1+}, \ldots, X_{p+})^{\mathrm{T}}$. Also $\delta_{\mathcal{C}}^0(X) = \sum_{i=1}^{2^p} 1\!\!1_{\mathcal{D}_i}(X) P_i X$ where $\mathcal{D}_i$ is the ith orthant and P_i is the projection onto the nearest face of $\mathcal{C}$. If the orthant $\mathcal{D}_i$ has positive entries in position $(i_1, \ldots, i_s)$, $1 \le i_1 \cdots \le i_s \le p$ and $s \ge 0$, P_i is the linear projection operator such that $(P_iX)_j = X_j 1\!\!1[j \in (i_1, \ldots, i_s)]$ and is of rank $p_i = s$. In this case the estimator in Eq. (12), with $r_i = r$, $i = 1, \ldots, 2^p$, equals

$$\delta_{\mathcal{C}}(X) = \left(1 - \frac{\|U\|^2\, r(\|X_+\|^2)}{(k+2)\,\|X_+\|^2}(q-2)_+\right) X_+,$$

where q equals the number of positive components of X.

It is worth noting, as Sengupta and Sen (1991) point out in their case, that shrinking only on $C_1 = C$ is strictly worse than our estimator. Each additional set C_i with positive Lebesgue measure adds to the reduction in risk. In particular the above estimator strictly dominates the estimator of Ouassou and Strawderman (2002) for this case if $p \ge 4$ since that estimator shrinks only when $X \in \mathcal{C}$.

Example 2 (The Suborthant Model) Suppose $\mathcal{C} = \mathbb{R}_+^s \otimes \mathbb{R}^{s-p} = \{\theta \mid \theta_i \ge 0, i = 1, \ldots, s\}$, the suborthant model. The usual estimator is

$$\delta_{\mathcal{C}}^0(X) = \begin{pmatrix} X_{1+} \\ X_2 \end{pmatrix},$$

where dim $X_1 = s$ and dim $X_2 = p - s$.

Our competing estimator may be described succinctly as

$$\delta_C(X) = \left(1 - (q + p - s - 2)_+ \frac{\|U\|^2 r(\|X_{1+}\|^2 + \|X_2\|^2)}{(k+2)(\|X_{1+}\|^2 + \|X_2\|^2)}\right) \delta^0_C(X),$$

where q equals the number of positive components of X_{1+}.

Example 3 Occasionally a cone $\mathcal{C}$ can be transformed into an orthant or suborthant model via an orthogonal transformation B. For example, the cone $\mathcal{B}_1 = \{\theta \mid \theta_1 \leq \frac{\theta_1+\theta_2}{2} \leq \cdots \leq \frac{\theta_1+\cdots+\theta_p}{p}\}$, the increasing on average cone and $\mathcal{B}_2 = \{\theta \mid \theta_1 \leq \frac{\theta_1+\theta_2}{2} \leq \cdots \leq \frac{\theta_1+\cdots+\theta_p}{p}$ and $0 \leq \frac{\theta_1+\cdots+\theta_p}{p}\}$ can be transformed into $\mathbb{R}^{p-1}_+ \otimes \mathbb{R}^1$ and $\mathbb{R}^p_+$, respectively, by the Helmert orthogonal transformation

$$B = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & \cdots & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & 0 & \cdots & 0 \\ -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & \frac{3}{\sqrt{13}} & \cdots & 0 \\ -\frac{1}{\sqrt{p(p-1)}} & -\frac{1}{\sqrt{p(p-1)}} & -\frac{1}{\sqrt{p(p-1)}} & -\frac{1}{\sqrt{p(p-1)}} & \cdots & \frac{p-1}{\sqrt{p(p-1)}} \\ \frac{1}{\sqrt{p}} & \frac{1}{\sqrt{p}} & \frac{1}{\sqrt{p}} & \frac{1}{\sqrt{p}} & \cdots & \frac{1}{\sqrt{p}} \end{pmatrix}.$$

Example 4 The cone $\mathcal{C} = \{\theta | \theta_1 \leq \theta_2 \leq \theta_3 \leq \cdots \leq \theta_p\}$ cannot be transformed into orthant via an orthogonal transformation if $p \geq 3$ since for $p \geq 3$ this cone is strictly contained in $\mathcal{B}_1$ of Example 3. It can be transformed into $\mathbb{R}^+_{p-1} \otimes \mathbb{R}_1$ via a non-orthogonal matrix. Spherical symmetry of the distribution of X will not be preserved under such a transformation, however. Here, direct determination of the $\mathcal{D}_i$ and P_i of Eq. (12) is more complicated. See Robertson et al. (1988), and references therein, for further discussion of this issue.

6 Improved estimation in the presence of vague information

In this section we again demonstrate the utility of Theorem 1 in a setting where the shrinkage function may not be differentiable. We consider estimation of the mean vector when there is "vague prior information" that the mean μ is in a closed convex body K, with piecewise smooth boundary but where μ is not restricted to lie in K. Note that in this section only the mean vector of X will be μ and not θ (θ will be reserved for another use) and the inner product $Y^{\mathrm{T}}Z$ will be denoted by $\langle Y, Z\rangle$.

Recently Kuriki and Takemura (KT) (2000) considered this problem for $X \sim N_p(\mu, I)$ following the work of Bock on this same model but where K is a polytope. They showed their estimator can be considered as natural extensions of Bock's estimators.

In this class of problems one has "vague knowledge" that μ is in or near the specified convex set and consequently wishes to shrink toward the set but does not wish to restrict the estimator to lie in the set. KT show, however, that the problem of shrinkage toward a closed convex set, $\mathcal{C}$, is essentially equivalent to the problem of estimating μ when μ is restricted to the dual cone, $\mathcal{C}^*$, of $\mathcal{C}$. Note that since $\mathcal{C}^*$ is positively homogeneous the results of the previous section apply. We discuss this further at the end of this section.

We briefly review the results of Kuriki and Takemura and indicate an extension to the case where X has a p-variate normal distribution with mean μ and covariance $\sigma^2 I$. We closely follow their notation in this section (which is why μ, not θ, is the mean). We then give an extension using Theorem 1 to the general spherically symmetric case. We emphasize that the shrinkage functions involved may not be weakly differentiable and that the use of Theorem 1 in such settings seems quite useful. The reader is referred to KT for details of the development below. For reasons of space, our development is greatly condensed. Our primary purpose is to give an example of the utility of Theorem 1.

Let K be a closed convex set in $\mathbb{R}^p$, and for each $X \in \mathbb{R}^p$, $X = X_K + (X - X_K)$ when X_K is the unique closest point in K to X. Let ∂K be the boundary of K. For fixed $s \in \partial K$, the normal cone of K at s is $\mathcal{N}(K, s) = \{y - s | y_K = s\}$ and is such that $X - X_K \in \mathcal{N}(K, X_K)$. Partition the boundary of K as $\partial K = D_1(\partial K) \cup D_2(\partial K) \cup \cdots \cup D_p(\partial K)$ where $D_m(\partial K) = \{s \in \partial K | \dim \mathcal{N}(K, s) = m\}$. Let $E_m(\partial K) = \{X \in \mathbb{R}^p \backslash K | X_K \in D_m(\partial K)\}$ and note that $E_1(\partial K), \ldots, E_p(\partial K)$ form a partition of $\mathbb{R}^p \backslash K$. We need the boundary set, ∂K, to be "piecewise smooth", that is, each $D_m(\partial K)$ is a $p - m$ dimensional C^2 manifold consisting of a finite number of relatively open connected components. Further ∂K is called smooth if it is piecewise smooth and each $D_m(\partial K)$ is empty for $m \geq 2$.

KT introduce a C^2 local coordinate system $\theta = (\theta^1, \ldots, \theta^{p-m})$ for $X_K = s = s(\theta^1, \ldots, \theta^{p-m}) \in D_m(\partial K)$ in a neighborhood of s. The tangent space $T_{s(\theta)}$ of $D_m(\partial K)$ at $s(\theta)$ is spanned by $\{b_a(\theta) = \frac{\partial s(\theta)}{\partial \theta^a},\ a = 1, \ldots, p - m\}$ and the space $T^{\perp}_{s(\theta)}$ is spanned by an orthonormal basis $\{\eta_\alpha(\theta), \alpha = 1, \ldots, m\}$ where $\langle b_a(\theta), \eta_\alpha(\theta)\rangle = 0$ and $\langle \eta_\alpha(\theta), \eta_\beta(\theta)\rangle = \delta_{\alpha\beta}$ (Kronecker's delta) ($\langle , \rangle$ is the inner product). Any element of $\mathcal{N}(K, s)$ can be expressed as $\sum t^\alpha \eta_\alpha(\theta)$ where $t = (t^1, \ldots, t^m)$ and hence we have the decomposition $X = s(\theta) + n(\theta, t)$ where $n(\theta, t) = \sum t^\alpha \eta_\alpha(\theta)$.

KT show that the Jacobian of the local 1–1 transformation $X \leftrightarrow (\theta, t)$ is expressed through (Lemma 2.1 of KT) $\mathrm{d}X = \pm |I_{p-m} + H(\theta, t)| \mathrm{d}s\, \mathrm{d}t$ where $H(\theta, t) = \sum_{\alpha=1}^m t^\alpha H_\alpha(\theta)$,

$$\mathrm{d}s = \sqrt{|G(\theta)|}\, \mathrm{d}\theta^1 \mathrm{d}\theta^2 \cdots \mathrm{d}\theta^{p-m}.$$

(The volume element of $D_m(\partial K)$ and $\mathrm{d}x = \mathrm{d}x_1 \cdots \mathrm{d}x_p$, $\mathrm{d}t = \mathrm{d}t^1 \cdots \mathrm{d}t^m$.)

In the above $G(\theta) = (g_{\alpha\beta}(\theta))_{1 \leq \alpha, \beta \leq p-m}$ with $g_{\alpha\beta}(\theta) = \langle b_\alpha(\theta), b_\beta(\theta)\rangle$ and $H(\theta, t) = \sum_{\alpha=1}^m t^\alpha H_\alpha(\theta)$ where $H_\alpha(\theta) = (h^b_{a\alpha}(\theta))_{1 \leq a, b \leq p-m}$ and where $h^b_{a\alpha}(\theta) = \sum_{c=1}^{p-m} h_{ac\alpha}(\theta) g^{cb}(\theta)$, $h_{ab\alpha}(\theta) = \langle -\frac{\partial^2 s(\theta)}{\partial \theta^a \partial \theta^b},\ \eta_\alpha(\theta)\rangle$ and $G^{-1}(\theta) = (g^{\alpha\beta}(\theta))$.

Let ℓ denote the length of the orthogonal projection given by $\ell = \|X - X_K\| = \|n(\theta, t)\| = \sqrt{\sum (t^\alpha)^2}$ and let $u = \ell^{-1} t \in S^{m-1}$ (the unit sphere in $\mathbb{R}^m$). KT give two lemmas (2.2 and 2.3) which express the conditional distribution of $t = (t^1, \ldots, t^m)$ given $X_K = s(\theta) \in D_m(\partial K)$ and the conditional density of ℓ given $X_K = s(\theta) \in D_m(\partial K)$ and u such that $n(\theta, u) \in N(K, s(\theta))$ when $X \sim N_p(\mu, I)$. These lemmas extend in obvious ways when $X \sim N_p(\mu, \sigma^2 I)$.

The following lemma represents the key to our extension of KT results to the general spherically symmetric case.

Lemma 3 *Let $X \sim N_p(\mu, \sigma^2 I)$ and let ∂K be piecewise smooth. Assume for each $X \in E_m(\partial K)$, $c(X) = c(\theta, \ell u)$ is a continuous and piecewise differentiable* *function in ℓ for fixed (θ, u) and satisfies the boundary condition*

$$\lim_{\ell \to 0, +\infty} \frac{c(\theta, \ell u)}{\ell} f(\ell|\theta, u) = 0,$$

where $f(\ell|\theta, u)$ is the conditional distribution of ℓ given (θ, u).

(a) Then

$$E\left\{c(\theta, \ell u)\left[\frac{\langle n(\theta, \ell u), \mu - s(\theta)\rangle}{\ell^2} - 1\right]\Bigg|\, \theta, u\right\}$$
$$= \sigma^2 E\left\{-\frac{1}{\ell}\frac{\partial c(\theta, \ell u)}{\partial \ell} - \frac{c(\theta, \ell u)}{\ell^2}(d-2)\Bigg|\, \theta, u\right\},$$

where $d = d(x) = d(\theta, \ell u) = m + tr\ H(\theta, t)(I + H(\theta, t))^{-1} = m + \ell\, tr\ H(\theta, u)(I + \ell H(\theta, u))^{-1}$.

(b) Therefore

$$E\left\{c(X)\left(\frac{\langle X - X_K, \mu - X_K\rangle}{\|X - X_K\|^2} - 1\right)\right\}$$
$$= \sigma^2 E\left\{-\frac{1}{\|X - X_K\|}\frac{\partial c(X)}{\partial \|X - X_K\|} - \frac{c(X)}{\|X - X_K\|^2}(d(X) - 2)\right\},$$

where we use the notation

$$\frac{\partial c(X)}{\partial \|X - X_K\|} = \left.\frac{\partial c(\theta, \ell u)}{\partial \ell}\right|_{\theta = \theta(X), \ell = \ell(X), u = u(X)}.$$

Proof The proof of (a) is essentially the same as that of Lemma 1 of KT adapted to the case of $X \sim N_p(\mu, \sigma^2 I)$. The second part follows directly from (a) on taking the expectation with respect to (θ, u). □

Corollary 2 *Suppose $c(X)$ satisfies the conditions of Lemma 3. If the distribution of (X, U) is $\mathrm{SS}_{p+k}(\mu, 0)$ and all expected values are finite*

$$E\left\{\|U\|^2 c(X)\left(\frac{\langle X - X_K, \mu - X_K\rangle}{\|X - X_K\|^2} - 1\right)\right\}$$
$$= E\left\{\frac{\|U\|^4}{k+2}\left(-\frac{1}{\|X - X_K\|}\frac{\partial c(X)}{\partial \|X - X_K\|} - \frac{c(X)}{\|X - X_K\|^2}(d(X) - 2)\right)\right\}.$$

Proof Through an elementary calculation, we have

$$E\left\{c(X)\left(\frac{\langle X - X_K, \mu - X_K\rangle}{\|X - X_K\|^2} - 1\right)\right\} = E\left\{\left\langle X - \mu, \frac{-c(X)}{\|X - X_K\|^2}(X - X_K)\right\rangle\right\}$$
$$= \sigma^2 E\left\{-\frac{1}{\|X - X_K\|}\frac{\partial c(X)}{\partial \|X - X_K\|} - \frac{c(X)}{\|X - X_K\|^2}(d(X) - 2)\right\}$$

according to Lemma 3 (b). Then the result follows from Theorem 1 with

$$f(X) = \frac{-c(X)}{\|X - X_K\|^2}(X - X_K).$$

□

We now apply Corollary 2 to estimation of μ when we have "vague" prior information that $\mu \in K$ and we wish to shrink toward K but not to restrict the estimator (or parameter) to lie in K. We consider estimators similar to those in KT adapted to the general spherically symmetric case. In particular, let $\delta(X, U) = X_K + (1 - \frac{\phi(X)U'U}{k+2})(X - X_K)$ where $\phi(X) = \frac{c(X)}{\|X-X_K\|^2}$. Loss will be $L(\mu, \delta) = \|\mu - \delta\|^2$. We compare the risk of $\delta(X, U)$ and $\delta_0(X) = X$.

Theorem 3 *Suppose* $(X, U) \sim SS_{p+k}(\mu, 0)$ *and that* K *is a closed convex set such that its boundary* ∂K *be piecewise smooth. Suppose also that* $c(X)$ *satisfies the assumption of Lemma 3 and that all expectations are finite.*

(a) The risk difference given by $\Delta R = E_\mu\{L(\theta, \delta(X, U))\} - E\{L(\theta, X)$ *equals*

$$E_\mu\left\{\left(\frac{\|U\|^2}{k+2}\right)^2\left[\frac{c^2(X)}{\|X-X_K\|^2} - 2\left(\frac{c(X)(d(X)-2)}{\|X-X_K\|^2}\right.\right.\right.$$
$$\left.\left.\left. + \frac{1}{\|X-X_K\|}\frac{\partial c(X)}{\partial\|X-X_K\|}\right)\right]\right\}.$$

(b) Hence $\delta(X, U)$ *dominates* X *provided the term in braces is everywhere non-positive and strictly negative on a set of positive measure.*

Proof ΔR is the difference between

$$E_\mu\left\{\|X_K + \left(1 - \frac{c(X)\|U\|^2}{(k+2)\|X-X_K\|^2}\right)(X - X_K) - \mu\|^2\right\}$$

and

$$E_\mu\{\|X_K + (X - X_K) - \mu\|^2\}.$$

Expanding the squared norm, ΔR reduces to

$$E_\mu\left\{\left(\frac{\|U\|^2}{k+2}\right)^2\frac{c^2(X)}{\|X-X_K\|^2} - 2\frac{c(X)\|U\|^2}{(k+2)\|X-X_K\|^2}\langle X - X_K, X_K + (X - X_K) - \mu\rangle\right\}$$

and finally equals

$$E_\mu\left\{\left(\frac{\|U\|^2}{k+2}\right)^2\frac{c^2(X)}{\|X-X_K\|^2}\right.$$
$$\left. - \frac{2}{k+2}c(X)\|U\|^2 - \frac{2}{k+2}\frac{c(X)\|U\|^2}{\|X-X_K\|^2}\langle X - X_K, X_K - \mu\rangle\right\},$$

which gives the desired result by applying Corollary 2.

Part (b) follows directly from part (a). □

Note that the theorem says, in effect, that if $\delta(X) = X_K + (1 - \sigma^2\phi(X))(X - X_K)$ dominates X in the case where $X \sim N_p(\mu, \sigma^2 I)$, then $\delta(X, U) = X_K + (1 - \frac{U'U}{k+2}\phi(X))(X - X_K)$ dominates X in the general spherically symmetric case.

KT give two examples of functions $c(X)$ for which domination holds (recall that m is the dimension of the normal cone at X_K). These are:

$$c(X) = \begin{cases} d(X) - 2 & \text{if } m \geq 2 \\ d(X) - 2 + \frac{1}{|I_{p-1} + H(X)|} & \text{if } m = 1 \\ 0 & \text{if } X \in K \end{cases}$$

and

$$c(X) = \begin{cases} \max(d(X) - 2, 0) & \text{if } m \geq 1 \\ 0 & \text{if } X \in K \end{cases}.$$

They note that when K is polyhedral $H(\theta, t) = 0$ and $d(X) = m$. They also note, as is clear from Theorem 3 that a Baranchik-type estimator where $c(X)$ is replaced by $c(X)r(\|X - X_K\|^2)$ for $0 < r(\cdot) < 1$ and $r'(\cdot) > 0$ will give a dominating estimator as well.

KT also point out a connection between estimation in the presence of "vague" prior information, and estimation in a restricted parameter space. In particular we have the following analog to the discussion in their Section 3.3.

Let $\mathcal{C}$ be a closed convex cone and let $\mathcal{C}^*$ be the dual cone which is assumed to have a piecewise smooth boundary. Let $(X, U) \sim \mathrm{SS}_{p+k}(\mu, 0)$ and suppose μ is restricted to lie in $\mathcal{C}$. We consider shrinkage estimators of the form

$$\delta(X, U) = \left(1 - \frac{c(X)U'U}{\|X_C\|^2}\right) X_C = (1 - \phi(X)U'U)X_C.$$

Since $\mathcal{C}^*$ is the dual cone of $\mathcal{C}$, $X = X_C + X_{C^*}$, $\langle X_C, X_{C^*}\rangle = 0$ and $X_C = \|X - X_C\|^2$. We compare $\delta(X, U)$ to $\delta_0(X) = X_C$ and have that the difference in loss between the two estimators is given by

$$\begin{aligned} \Delta L &= \|\delta(X, U) - \mu\|^2 - \|X_C - \mu\|^2 \\ &= \|(1 - \phi(X)U'U)X_C - \mu\|^2 - \|X_C - U\|^2 \\ &= \|X_{C^*} + (1 - \phi(X)U'U)X_C - \mu\|^2 - \|X_{C^*} + X_C - \mu\|^2 \\ &= \|X_{C^*} + (1 - \phi(X)U'U)X_C - \mu\|^2 - \|X - \mu\|^2. \end{aligned}$$

Hence by taking $K = \mathcal{C}^*$ in Theorem 3 we obtain estimators which dominate X_C when μ is restricted to $\mathcal{C}$.

7 Extensions and comments

This article has two main goals. The first is to give a different proof of a result of Cellier and Fourdrinier (1995) which provides unbiased estimates of risk for quite general estimators of the location vector of a spherically symmetric distribution when a residual vector is present. The current result reproduces that result but is also applicable in certain cases where the original result was inapplicable—in particular when the estimator is not weakly differentiable but has a particular form.

The second is to apply the method to estimation of the location parameter when the parameter space is restricted. In this setting, some of the useful alternative estimators may not necessarily be weakly differentiable. This appears to be the case, for example, if the cone $\mathcal{C}$ has faces which do not coincide with any of the orthant faces.

The application of our results in these cases gives improved estimators with a very strong robustness property, namely they improve over the classical estimator uniformly in θ simultaneously for all spherically symmetric distributions.

We note in passing that the techniques of this paper also work in the context of elliptically symmetric distributions with a known correlation structure. In this case it is possible to transform to spherical symmetry (with an unknown scale) and a result analogous to Theorem 1 and consequently the other results of this paper will also hold.

We also note the papers of Fourdrinier et al. (1998, 2003) in which we derive results for general elliptical distributions in the unrestricted mean case. It is quite plausible that techniques similar to those in this paper would work in that context as well but we have not yet pursued them.

8 Appendix

Lemma 4 *Let $(X, U) \sim \mathrm{SS}_{p+k}(\theta, 0)$ and let $\alpha \in \mathbb{N}$. Assume $\varphi(X)$ is such that for any $R \geq 0$, the conditional expectation*

$$f(R) = E_{(\theta,0)}[\|U\|^{\alpha}\varphi(X) \mid \|X-\theta\|^2 + \|U\|^2 = R^2]$$

exists. Then the function f is continuous on $\mathbb{R}_+$.

Proof Clearly it suffices to prove the result when $\theta = 0$ and we assume without loss of generality that the function φ is non-negative. As the conditional distribution of (X, U) given $\|X\|^2 + \|U\|^2 = R^2$ is the uniform distribution U_R on the sphere $S_R = \{y \in \mathbb{R}^{p+k} \mid \|y\| = R\}$ in $\mathbb{R}^{p+k}$ centered at 0 and of radius R, we have

$$f(R) = \int_{S_R} \|u\|^{\alpha}\varphi(x)\mathrm{d}U_R(x, u).$$

Since for $(x, u) \in S_R$, we have $\|u\|^2 = R^2 - \|x\|^2$ and X has distribution concentrated on the ball $B_R = \{x \in \mathbb{R}^p \mid \|x\| \leq R\}$ in $\mathbb{R}^p$ with density proportional to $R^{2-(p+k)}(R^2 - \|x\|^2)^{k/2-1}$, we have that $R^{p+k-2}f(R)$ is proportional to

$$g(R) = \int_{B_R} (R^2 - \|x\|^2)^{(k+\alpha)/2-1}\varphi(x)\mathrm{d}x.$$

Now, through the area measure σ_r on the sphere S_r, we can write

$$\begin{aligned} g(R) &= \int_0^R \int_{S_r} (R^2 - \|x\|^2)^{(k+\alpha)/2-1}\varphi(x)\mathrm{d}\sigma_r(x)\mathrm{d}r \\ &= \int_0^R (R^2 - r^2)^{(k+\alpha)/2-1} H(r)\mathrm{d}r \end{aligned}$$

with

$$H(r) = \int_{S_r} \varphi(x)\mathrm{d}\sigma_r(x).$$

Since H and $(k+\alpha)/2-1$ are non-negative, the family of integrable functions$r \to K(R,r) = (R^2 - r^2)^{(k+\alpha)/2-1} H(r) \mathbb{1}_{[0,R]}(r)$, indexed by R, is non-decreasing in R and bounded above (for $R < R_0$) by the integrable function $K(R_0, r)$. Then the continuity of $g(R)$, and hence of $f(R)$, is guaranteed by the Dominated Convergence Theorem. □

Acknowledgements The authors express their gratitude to two referees who read the original manuscript with care and made a number of suggestions for improving the presentation. The support of NSF Grant DMS 97-04524 is gratefully acknowledged by W.E. Strawderman. The support of NSF Grant DMS 99-71586 is gratefully acknowledged by M.T. Wells

References

Baranchik, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *The Annals of Mathematical Statistics*, *41*, 642–645.

Berger, J. O. (1985). *Statistical decision theory and bayesian analysis* (2nd ed) Berlin Heidelberg, New York: Springer.

Bock, M. E. (1982). Employing vague inequality information in the estimation of normal mean vectors (estimators that shrink to closed convex polyhedral) In: S. S. Gupta & J. O. Berger (Eds.), *Statistical decision theory and related topics III* (Vol. 1 169–193). Academic Press, New York.

Cellier, D., Fourdrinier, D. (1995). Shrinkage estimators under spherical symmetry for the general linear model. *Journal of Multivariate Analysis*, *52*, 338–351.

Chang, Y. -T. (1981). Stein-type estimators for parameters restricted by linear inequalities. *Keio Science and Technology Reports*, *34*, 83–95.

Chang, Y. -T. (1982). Stein-type estimators for parameters in truncated spaces. *Keio Science and Technology Reports*, *35*, 185–193.

Fourdrinier, D., Ouassou, I., Strawderman, W. E. (2003). Estimation of a parameter vector when some components are restricted. *Journal of Multivariate Analysis*, *86*, 14–27.

Fourdrinier, D., Strawderman, W. E., Wells, M. T. (1998). Estimation robuste pour les lois à symétrie elliptique à matrice de covariance inconnue. *Note aux Comptes Rendus de l'Académie des Sciences*, *Paris, Série I*, *326*, 1135–1140.

Fourdrinier, D., Strawderman, W. E., Wells, M. T. (2003). Robust shrinkage estimation for elliptically symmetric distributions with unknown covariance matrix. *Journal of Multivariate Analysis*, *85*, 24–39.

Fourdrinier, D., Wells, M. T. (1995). Estimation of a loss function for spherically symmetric distributions in the general linear model. *The Annals of Statistics*, *23*, 571–592.

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

Katz, M. W. (1961). Admissible and minimax estimates of parameters in truncated spaces. *The Annals of Mathematical Statistics*, *32*, 136–142.

Kuriki, S., Takemura, A. (2000). Shrinkage estimation towards a closed convex set with a smooth boundary. *Journal of Multivariate Analysis*, *75*, 79–111.

Morrey, C. B. (1966). *Multiple integrals in the calculus of variations*. Berlin-Heidelberg-New York: Springer.

Ouassou, I., Strawderman, W. E. (2002). Estimation of a parameter vector restricted to a cone. *Statistics and Probability Letters*, *56*, 121–129.

Robertson, T., Wright, F. T., Dykstra, R. L. (1988). *Order restricted statistical inference*. New York: Wiley.

Sengupta, D., Sen, P. K. (1991). Shrinkage estimation in a restricted parameter space. *Sankhyā*, *53*, 389–411.

Stein, C. (1981). Estimation of a multivariate normal mean. *The Annals of Statistics*, *9*, 1135–1151.

Stoer, J., Witzgall, C. (1970). *Convexity and optimization in finite dimensions I*. Berlin Heidelberg, New York: Springer.

Ziemer, W. P. (1989). *Weakly differentiable functions—Sobolev spaces and functions of bounded variation*. Berlin Heidelberg, New York: Springer.