Chapter 6
Estimation of a Mean Vector for Spherically Symmetric Distributions II: With a Residual

6.1 The General Linear Model Case with Residual Vector

In this chapter, we consider the canonical form of the general linear model introduced in Sect. 4.5 when a residual vector U is available. Recall that (X, U) is a random vector around $(\theta, \mathbf{0})$ (such that dim X = dim $\theta = p$ and dim U = dim $\mathbf{0} = k$) with a spherically symmetric distribution, that is, $(X, U) \sim SS_{p+k}(\theta, \mathbf{0})$. Estimation of θ under quadratic loss $\|\delta - \theta\|^2$ parallels the normal situation presented in Sects. 2.3 and 2.4 where $X \sim \mathcal{N}_p(\theta, \sigma^2 I_p)$ (with σ^2 known) and the estimators of θ are of the form $\delta(X) = X + \sigma^2 g(X)$. In the case where σ^2 is unknown (see Sect. 2.4.3), the corresponding estimators are

$$\delta(X) = X + \frac{S}{k+2} g(X)$$

where $S \sim \sigma^2 \chi^2_k$ independent of X. Note that, when $(X^{\mathrm{T}}, U^{\mathrm{T}})^{\mathrm{T}} \sim \mathcal{N}((\theta^{\mathrm{T}}, 0^{\mathrm{T}})^{\mathrm{T}}, \sigma^2 I_{p+k})$, $S = \|U\|^2$. This most basic case of the general linear model suggests considering improved shrinkage estimators of the form

$$\delta(X) = X + \frac{\|U\|^2}{k+2} g(X) \tag{6.1}$$

for some function g from $\mathbb{R}^p$ into $\mathbb{R}^p$. In this section,

$$\sigma^2 = \mathrm{Var}(X_i) = \mathrm{Var}(U_i) = \frac{1}{p} E_\theta[\|X - \theta\|^2] = \frac{1}{k} E_\theta[\|U\|^2] = \frac{1}{p+k} E[R^2],$$

where $R = (\|X - \theta\|^2 + \|U\|^2)^{1/2}$, can be considered as known or unknown. When σ^2 is unknown, $\|U\|^2/k$ is an unbiased estimator of σ^2. Also, when σ^2 is unknown, it is perhaps preferable to use the invariant loss $\|\delta - \theta\|^2/\sigma^2$ since the estimator X

D. Fourdrinier et al., *Shrinkage Estimation*, Springer Series in Statistics,
https://doi.org/10.1007/978-3-030-02185-6_6

has constant risk p and is minimax for this loss provided the variance of X is finite, while the minimax risk for the loss $\|\delta - \theta\|^2$ is infinite. Note that domination of an estimator under one of these losses implies domination under the other.

When σ^2 is known, estimators of the form $\delta(X) = X + \sigma^2 g(X)$ can be used and we will contrast these estimators with estimators (6.1) in the next section. One advantage of the estimators in (6.1) is that they share a striking robustness property, namely that, if $\|g(X)\|^2 + 2\,\mathrm{div}\, g(X) \leq 0$, then $X + g(X)\, \|U\|^2/(k+2)$ dominates X for any spherically symmetric distribution of (X, U). In particular, the form of the density may not be known and indeed there is no need that a density exists. The proof of this robustness property is given below and follows closely that of Cellier and Fourdrinier (1995).

Assuming the risk of X is finite (i.e., $E_\theta[\|X - \theta\|^2] = E_0[\|X\|^2] < \infty$) the risk of $\delta(X)$ is finite if and only if $E_\theta[\|U\|^4 \|g(X)\|^2] < \infty$ and the difference in risk between $\delta(X)$ and X is

$$\begin{aligned} \Delta(\theta) &= R(\theta, \delta) - R(\theta, X) \\ &= E_\theta\left[2\,(X - \theta)^{\mathrm{T}} g(X)\, \frac{\|U\|^2}{k+2} + \|g(X)\|^2 \frac{\|U\|^4}{(k+2)^2} \right]. \end{aligned} \tag{6.2}$$

The cross product term, that is, the first term in the right-hand side of (6.2) will be analyzed as in the normal case. The following is the key adaptation of Stein's identity.

Lemma 6.1 *(Stein type lemma for the general linear model: Cellier and Fourdrinier 1995) Assume that* $(X, U) \sim SS(\theta, \mathbf{0})$ *where* $\dim X = \dim \theta = p$ *and* $\dim U = \dim \mathbf{0} = k$. *Then, for any weakly differentiable function* g *from* $\mathbb{R}^p$ *into* $\mathbb{R}^p$ *such that*

$$E_\theta\big[|(X - \theta)^{\mathrm{T}} g(X)|\big] < \infty,$$

we have

$$E_\theta\big[(X - \theta)^{\mathrm{T}} g(X)\, \|U\|^2\big] = E_\theta\left[\mathrm{div}\, g(X)\, \frac{\|U\|^4}{k+2}\right]. \tag{6.3}$$

Proof We will show that, conditionally on the radius $R = \|X - \theta\|^2 + \|U\|^2$, (6.3) holds. First, conditionally on R, the left-hand side of (6.3) is expressed as (see Corollary 4.2)

$$\begin{aligned} E_{R,\theta}\big[(X - \theta)^{\mathrm{T}} g(X)\, \|U\|^2\big] &= \int_{S_{R,\theta}} (x - \theta)^{\mathrm{T}} g(x)\, \|u\|^2\, d\mathcal{U}_{R,\theta}(x, u) \\ &= \int_{S_{R,\theta}} (x - \theta)^{\mathrm{T}} g(x)\, (R^2 - \|x - \theta\|^2)\, d\mathcal{U}_{R,\theta}(x, u) \\ &= \int_{B_{R,\theta}} (x - \theta)^{\mathrm{T}} g(x)\, C_R^{p,k} (R^2 - \|x - \theta\|^2)^{k/2}\, dx \end{aligned} \tag{6.4}$$

since, according to (4.4), X given R has density

$$\psi_{R,\theta}(x) = C_R^{p,k} (R^2 - \|x-\theta\|^2)^{k/2-1} \, \mathbb{1}_{B_{R,\theta}}(x)$$

with

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

Now, note that

$$(R^2 - \|x-\theta\|^2)^{k/2} \, (x-\theta) = \nabla \gamma(x)$$

where

$$\gamma(x) = \frac{-(R^2 - \|x-\theta\|^2)^{k/2+1}}{k+2} \, .$$

Hence, using the classical identity

$$\big(\nabla \gamma(x)\big)^{\mathrm{T}} g(x) = \operatorname{div}\big(\gamma(x)\, g(x)\big) - \gamma(x) \operatorname{div} g(x) \, ,$$

it follows from (6.4) that

$$E_{R,\theta}\big[(X-\theta)^{\mathrm{T}} g(X) \, \|U\|^2\big] = A + B \tag{6.5}$$

where

$$A = C_R^{p,k} \int_{B_{R,\theta}} \operatorname{div}\big(\gamma(x)\, g(x)\big) dx \tag{6.6}$$

and

$$B = C_R^{p,k} \int_{B_{R,\theta}} -\gamma(x) \operatorname{div} g(x) dx. \tag{6.7}$$

Applying Stokes' theorem to the integral in (6.6) gives

$$A = C_{S_R^{p,k}} \int_{S_{R,\theta}} \gamma(x)\, g(x) \, \frac{x-\theta}{\|x-\theta\|} \, d\sigma_{R,\theta}(x) = 0 \tag{6.8}$$

since, for any $x \in S_{R,\theta}$, $\gamma(x) = 0$. The B term in (6.7) can be expressed as

$$B = \int_{B_{R,\theta}} \operatorname{div} g(x) \, \frac{(R^2 - \|x-\theta\|^2)^2}{k+2} \, \psi_{R,\theta}(x) \, dx = E_{R,\theta}\left[\operatorname{div} g(X) \, \frac{\|U\|^4}{k+2}\right]$$

and, finally, the lemma follows from (6.4), (6.5) and (6.8). □

As a consequence of Lemma 6.1, we can derive a sufficient condition of domination of $\delta(X) = X + \|U\|^2/(k+2)g(X)$ over the usual estimate X.

Theorem 6.1 *Let $(X, U) \sim SS_{p+k}(\theta, 0)$ and the loss be given by $\|\delta - \theta\|^2$. Assume that $E_\theta[\|X\|^2] < \infty$ and $E_\theta[\|U\|^4 \, \|g(X)\|^2] < \infty$. Then an unbiased estimator of the risk difference $\Delta(\theta)$ in (6.2) between $\delta(X) = X + g(X)\,\|U\|^2/(k+2)$ and X is*

$$[2 \operatorname{div} g(X) + \|g(X)\|^2] \, \frac{\|U\|^4}{(k+2)^2}. \tag{6.9}$$

A sufficient condition for domination of $\delta(X)$ over X is that, for any $x \in \mathbb{R}^p$,

$$2 \operatorname{div} g(x) + \|g(x)\|^2 \leq 0 \tag{6.10}$$

with strict inequality on a set a positive measure on $\mathbb{R}^p$.

Proof The proof of (6.9) follows immediately from (6.3) and (6.2). The domination condition (6.10) is a direct consequence of (6.9). □

Remark 6.1 The addition of the residual term U in the estimate yields an interesting and strong robustness property. Note that the hypotheses in Theorem (6.1) are independent of the radial distribution and are consequently valid for any spherically symmetric distribution. This is in contrast with the results of Sect. 6.2 which require conditions on the radial distribution.

Differential expressions that lead to risk domination results, such as in Theorem 6.1, have been extended to spherical and ellipitical location models by several authors (see, for example, Cellier et al. 1989, Chou and Strawderman 1990, Brandwein and Strawderman 1980, Brandwein and Strawderman 1991a, Cellier and Fourdrinier 1995, Fourdrinier et al. 2003, Fourdrinier et al. 2006, Kubokawa 1991, Maruyama 2003a, and Fourdrinier and Strawderman 2008a,b). A notable aspect of many of the papers, in the presence of a residual vector U, is the development of robust estimators in the sense that they are minimax for a wide class of spherically symmetric distributions (see particularly, for example, Cellier et al. 1989, Cellier and Fourdrinier 1995, and Fourdrinier et al. 2006).

The improved estimators in Sect. 5.3,without residual vector, require two critical hypotheses. The first is the superharmonicity condition on an auxillary function h such that $\|g\|^2/2 \leq -h \leq -\operatorname{div} g$. Secondly these estimators require the assumption that the function $R \to R^2 E_{R,\theta}[h]$ is nonincreasing. In contrast, the conditions for improvement of the improved estimator with the residual term included share the same set of hypotheses as the general Stein type estimators in the normal case (see Sect. 2.3). As a result, estimators which dominate X (through the differential inequality) in the normal case dominate X simultaneously for all spherically symmetric distributions (subject to the finiteness of the risk). At this point, we will focus on the so-called robust James-Stein estimators rather than discussing general examples as in Sect. 2.3.

Consider

$$\delta^a_{RJS}(X) = \left(1 - \frac{a}{\|X\|^2}\,\frac{\|U\|^2}{k+2}\right) X$$

where a is a positive constant which is of the form (6.1) with $g(X) = -aX/\|X\|^2$. Note this is the shrinkage in the basic James-Stein estimator in (2.13) with $\sigma^2 = 1$. Using the divergence calculation of this $g(X)$ from (2.16), the unbiased estimator of the risk difference implied by (6.9) is,

$$\left(a^2 - 2\,a(p-2)\right)\frac{1}{\|X\|^2}\,\frac{\|U\|^4}{(k+2)^2}\,,$$

and so it follows that domination occurs for $0 < a < 2(p-2)$, and the optimal constant a (i.e., with minimum risk) is $a = p - 2$. Note that this optimal a is independent of the sampling distribution and yields improvement on X for any spherically symmetric distribution. Hence the best a also has a nice robust optimality property.

An alternative approach to the results of this section can be based on the approach used in Lemma 5.2 where a density is assumed, that is, $(X, U) \sim f(\|x-\theta\|^2 + \|U\|^2)$. This second approach has been used by many authors in this and more general settings. For spherically symmetric distributions with a density it is essentially related to the above method. A statement of this connection is given at the end of this section. The proof is provided in the Appendix. Thus a straightforward adaptation of the proof of Lemma 5.2 leads to

$$\begin{aligned} E_\theta\left[(X-\theta)^{\mathrm{T}} g(X)\,\|U\|^2\right] &= E_\theta\left[\frac{F(\|X-\theta\|^2 + \|U\|^2)}{f(\|X-\theta\|^2 + \|U\|^2)}\,\mathrm{div}_X g(X)\,\frac{\|U\|^2}{k+2}\right] \\ &= C\,E^*_\theta\left[\mathrm{div}_X g(X)\,\frac{\|U\|^2}{k+2}\right] \qquad (6.11) \end{aligned}$$

where C and E^*_θ are defined in Lemma 5.2. Similarly

$$\begin{aligned} E_\theta\left[\|g(X)\|^2\,\frac{\|U\|^4}{(k+2)^2}\right] &= E_\theta\left[U^{\mathrm{T}}\left(U\,\frac{\|U\|^2}{(k+2)^2}\,\|g(X)\|^2\right)\right] \\ &= E_\theta\left[\frac{F(\|X-\theta\|^2 + \|U\|^2)}{f(\|X-\theta\|^2 + \|U\|^2)}\,\mathrm{div}_U\left(U\|U\|^2\right)\|g(X)\|^2\right] \\ &= E_\theta\left[\frac{F(\|X-\theta\|^2 + \|U\|^2)}{f(\|X-\theta\|^2 + \|U\|^2)}\,\frac{\|U\|^2}{k+2}\,\|g(X)\|^2\right] \\ &= C\,E^*_\theta\left[\frac{\|U\|^2}{k+2}\,\|g(X)\|^2\right]. \qquad (6.12) \end{aligned}$$

Hence the difference in risk between $X+g(X)\,\|U\|^2/(k+2)$ and X can be written as

$$C\,E_\theta^*\left[\left(2\operatorname{div} g(X)+\|g(X)\|^2\right)\frac{\|U\|^2}{k+2}\right]. \tag{6.13}$$

Note that the normalizing constant

$$C=\int_{\mathbb{R}^p\times\mathbb{R}^k} F(\|x-\theta\|^2+\|u\|^2)\,dx\,du\,. \tag{6.14}$$

can be expressed, through a straightforward application of the Fubini theorem, as

$$C=\frac{1}{p+k}\int_0^\infty r^2\,h(r)\,dr \tag{6.15}$$

where $h(r)$ is the radial density. Thus C is the common variance of each coordinate of (X,U). Therefore it follows from (6.13) that condition (6.10) is sufficient for the minimaxity of the estimator $X+g(X)\,\|U\|^2/(k+2)$, provided we treat the density $f(\cdot)$ as fixed and known, which implies implicitly that σ^2 is known. Alternatively, if

$$(X,U)\sim\frac{1}{\sigma^{p+k}}\,f\left(\frac{\|x-\theta\|^2+\|u\|^2}{\sigma^2}\right)$$

where σ^2 is unknown, and the loss is $\|\delta-\theta\|^2/\sigma^2$, then X is minimax simultaneously for all such families where $E_\theta[\|X\|^2]<\infty$. Hence (6.10) implies simultaneous minimaxity for the entire class as well.

6.1.1 More General Estimators

In this section, we give results for a more general class of estimators of θ of the form $\delta=\delta(X,\|U\|^2)$. The loss will be invariant squared error loss, i.e.

$$\eta\,\|\delta-\theta\|^2\,, \tag{6.16}$$

where $\eta=1/\sigma^2$, so that the risk is

$$R(\theta,\eta,\delta)=E_{\theta,\eta}\left[\eta\,\|\delta(X,U)-\theta\|^2\right], \tag{6.17}$$

where $E_{\theta,\eta}$ denotes the expectation with respect to the density (6.33) with $\eta=1/\sigma^2$. For the rest of this section, we assume

$$E_{\theta,\eta}\left[\|X-\theta\|^2\right]<\infty, \tag{6.18}$$

which guarantees that the standard estimator X has finite risk and is minimax. As $\delta(X, \|U\|^2)$ can be written as $\delta(X, \|U\|^2) = X + g(X, \|U\|^2)$, the finiteness of its risk is guaranteed by

$$E_{\theta,\eta}\left[\|g(X, \|U\|^2)\|^2\right] < \infty\,. \tag{6.19}$$

A version of the following lemma can be found in Fourdrinier et al. (2003). Its proof follows closely the pattern of (6.11) and (6.12).

Lemma 6.2 *Assume that the function $g(x, \|u\|^2)$ is weakly differentiable from $\mathbb{R}^{p+k}$ into $\mathbb{R}^p$. Then*

$$\eta\, E_{\theta,\eta}\left[(X-\theta)^{\mathrm{T}} g(X, \|U\|^2)\right] = C\, E^*_{\theta,\eta}\left[\mathrm{div}_X g(X, \|U\|^2)\right], \tag{6.20}$$

where $E^*_{\theta,\eta}$ is the expectation with respect to the density

$$\frac{\eta^{p+k}}{C} F\Big(\eta\,\left(\|x-\theta\|^2 + \|u\|^2\right)\Big), \tag{6.21}$$

provided either of the above expectations exists.

Similarly, for any weakly differentiable function h from $\mathbb{R}^{p+k}$ into $\mathbb{R}^p$,

$$\eta E_{\theta,\eta}\left[U^{\mathrm{T}} h(X, U)\right] = C\, E^*_{\theta,\eta}\left[\mathrm{div}_U h(X, U)\right]\,, \tag{6.22}$$

provided either of these expectations exists.

Thanks to Lemma 6.2, an expression of the risk difference between $\delta(X, \|U\|^2)$ and X is given in the following proposition.

Proposition 6.1 *Assume that $E_{\theta,\eta}\left[\|g(X, U)\|^2\right] < \infty$. The risk difference between $\delta(X, \|U\|^2) = X + g(X, \|U\|^2)$ and X equals*

$$\mathscr{R}(\theta, \eta, \delta) - \mathscr{R}(\theta, \eta, X) = C\, E^*_{\theta,\eta}\left[\mathscr{O} g(X, \|U\|^2)\right],$$

where

$$\mathscr{O} g(X, \|U\|^2) = 2\,\mathrm{div}_X g(X, \|U\|^2) + \frac{k-2}{\|U\|^2}\,\|g(X, \|U\|^2)\|^2 + 2\,\frac{\partial}{\partial S}||g(X, S)||^2\bigg|_{S=\|U\|^2}. \tag{6.23}$$

Proof A straightforward calculation of the risk difference gives

$$\begin{aligned}\triangle(\theta,\eta) &= \eta\, E_{\theta,\eta}\left[2\,(X-\theta)^{\mathrm{T}} g(X,\|U\|^2) + \|g(X,\|U\|^2)\|^2\right]\\ &= \eta\, E_{\theta,\eta}\left[2\,(X-\theta)^{\mathrm{T}} g(X,\|U\|^2) + U^{\mathrm{T}}\frac{U}{\|U\|^2}\|g(X,\|U\|^2)\|^2\right].\end{aligned}$$

Using Lemma 6.2 on each term in the brackets, we obtain

$$\begin{aligned}\triangle(\theta,\eta) &= C\, E^*_{\theta,\eta}\left[2\,\mathrm{div}_X g(X,\|U\|^2) + \mathrm{div}\left(\frac{U}{\|U\|^2}\,\|g(X,\|U\|^2)\|^2\right)\right]\\ &= C\, E^*_{\theta,\eta}\left[2\,\mathrm{div}_X g(X,\|U\|^2) + \frac{k-2}{\|U\|^2}\,\|g(X,\|U\|^2)\|^2\right.\\ &\qquad\left. + \frac{U^{\mathrm{T}}}{\|U\|^2}\,\nabla_U \|g(X,\|U\|^2)\|^2\right]\end{aligned}$$

by the divergence formula. Finally expressing the gradient gives

$$\begin{aligned}\triangle(\theta,\eta) &= C\, E^*_{\theta,\eta}\left[2\,\mathrm{div}_X g(X,\|U\|^2) + \mathrm{div}\left(\frac{U}{\|U\|^2}\,\|g(X,\|U\|^2)\|^2\right)\right]\\ &= C\, E^*_{\theta,\eta}\left[2\,\mathrm{div}_X g(X,\|U\|^2) + \frac{k-2}{\|U\|^2}\,\|g(X,\|U\|^2)\|^2\right.\\ &\qquad\left. +2\,\frac{\partial}{\partial S}||g(X,S)||^2\Big|_{S=\|U\|^2}\right].\end{aligned}$$

□

This result will be used in Sect. 6.3 to develop generalized Bayes minimax estimators. An easy corollary applicable to Baranchik type estimators of the form

$$\left(1 - a\, r\left(\frac{\|X\|^2}{S}\right)\frac{S}{\|X\|^2}\right)X \tag{6.24}$$

is the following. The proof is left to the reader.

Corollary 6.1 *The estimator* (6.24) *dominates* X *simultaneously for all spherically symmetric distributions* $SS_{p+k}(\theta, 0)$ *for which* $E^*_{\theta,\eta}[\|X\|^2] < \infty$ *under loss* (6.16) *provided*

(a) $0 < a \leq 2\,(p-2)$,
(b) $0 \leq r(\cdot) \leq 1$, *and*
(c) $r(\cdot)$ *is nondecreasing.*

6.1.2 A Link Between Expectations with Respect to E^*_{θ,σ^2} and E_{θ,σ^2}

We mentioned above that the two approaches to the results of this section are connected. Here is a lemma, whose proof is postponed to Appendix A.6, which makes explicit this connection thanks to a link between expectations with respect to E^*_{θ,σ^2} and E_{θ,σ^2}.

Lemma 6.3 (Fourdrinier and Strawderman 2015) *For any function γ defined on $\mathbb{R}^p \times \mathbb{R}_+$ and for any $\theta \in \mathbb{R}^p$, we have*

$$\sigma^2 \, C \, E^*_{\theta,\sigma^2}\Big[\gamma\Big(X, \|U\|^2\Big)\Big] = E_{\theta,\sigma^2}\left[\frac{1}{2}\,\frac{1}{\|U\|^{k-2}} \int_0^{\|U\|^2} \gamma(X,s)\, s^{k/2-1}\, ds\right], \tag{6.25}$$

provided these expectations exist, where C is defined in (6.14).

6.2 A Paradox Concerning Shrinkage Estimators

In this section, we contrast the result of the previous section and Sect. 5.2. We continue our study of the problem of estimating the mean vector θ of a spherically symmetric distribution when the scale σ^2 is known but when a residual vector U is available.

In Sect. 5.2, we studied the important class of improved estimators, the James-Stein estimators $\delta^a_{JS}(X) = \big(1 - a\,\sigma^2/\|X\|^2\big)X$. The previous section provided an alternative class of robust James-Stein estimators, that is, $\delta^a_{RJS}(X, U) = \big(1 - a/\|X\|^2 \, \|U\|^2/(k+2)\big)X$. In this section, we show that there often exist situations where $\delta^{p-2}_{RJS}(X, U)$ dominates $\delta^a_{JS}(X)$ simultaneously for all a and hence that the use of the residual vector U to estimate σ^2 may be superior to using its known value. This phenomenon seems paradoxical in the sense that the risk behavior of an estimator may be improved by substituting an estimate for a known quantity. This phenomenon adds to the attractiveness of the robust James-Stein class by demonstrating not only domination of the usual estimator X simultaneously for all spherically symmetric distributions, but also domination of the usual James-Stein estimators in many cases. A similar paradox was found in the context of goodness of fit testing by Wells (1990). The results of this section are Fourdrinier and Strawderman (1996) and Fourdrinier et al. (2004).

Note that the paradox cannot occur in the case of a normal distribution since by the Rao-Blackwell theorem, when σ^2 is known in the normal case, X is a complete sufficient statistic so that the conditional expectation of $\delta^a_{RJS}(X, U)$ given X reduces to $\delta^{ak/(k+2)}_{JS}(X)$ which dominates $\delta^a_{RJS}(X, U)$. Note also that, if the paradox holds for one value of σ^2 for a particular family, it holds for all values of σ^2 by the scale equivariance of $\delta^a_{RJS}(X, U)$ and, therefore, holds for any scale mixture. Hence, as the normal distribution arises as a mixture of uniform distributions on spheres, and also as a mixture of uniform distributions on balls, the paradox cannot occur for these distributions as well.

For ease of presentation, it is convenient to define the general estimator $\delta^a_\alpha(X, U) = \big(1 - a\|U\|^{2\alpha}/\|X\|^2\big)X$ for $\alpha = 0$ or 1 and to assume $\sigma^2 = 1$. Note that, for $\alpha = 0$, $\delta^a_0 = \delta^a_{JS}$ and, for $\alpha = 1$, $\delta^a_1 = \delta^{a/(k+2)}_{RJS}$. As in Sect. 6.1, we assume the finiteness of the risk of X (i.e., $E_0[\|X\|^2] < \infty$) and it is clear that the finiteness of the risk of $\delta^a_\alpha(X, U)$ is guaranteed as soon as $E_\theta\big[\|U\|^{2\alpha}/\|X\|^2\big] < \infty$. Under that condition, the following proposition yields the risk of δ^a_α.

Proposition 6.2 *Let the loss be* $\|\delta - \theta\|^2$. *The risk of* δ^a_α *equals*

$$R(\delta^a_\alpha, \theta) = E_0[\|X\|^2] + a^2\, E_\theta\left[\frac{\|U\|^{4\alpha}}{\|X\|^2}\right] - 2\,a\,\frac{p-2}{k+2\alpha}\; E_\theta\left[\frac{\|U\|^{2(\alpha+1)}}{\|X\|^2}\right].$$

Proof The risk calculation is a straightforward extension of the one in Lemma 6.1, with $g(x, s) = s^\alpha x/\|x\|^2$. □

It is easy to deduce from Lemma 6.2 that, for any $\theta \in \mathbb{R}^p$, the constant a for which the risk of δ^a_α is minimum is

$$a(\theta) = \frac{p-2}{k+2\alpha}\,\frac{E_\theta\big[\frac{\|U\|^{2(\alpha+1)}}{\|X\|^2}\big]}{E_\theta\big[\frac{\|U\|^{4\alpha}}{\|X\|^2}\big]}\,.$$

The corresponding risk is

$$R\big(\delta^{a(\theta)}_\alpha, \theta\big) = E_0\big[\|X\|^2\big] - \left(\frac{p-2}{k+2\alpha}\right)^2 \frac{\big(E_\theta\big[\frac{\|U\|^{2(\alpha+1)}}{\|X\|^2}\big]\big)^2}{E_\theta\big[\frac{\|U\|^{4\alpha}}{\|X\|^2}\big]}. \tag{6.26}$$

We already noticed in Sect. 6.1 that, for $\alpha = 1$, the optimal a does not depend on θ and equals $\frac{p-2}{k+2}$, which can also be easily seen from the above expression. For $\alpha = 0$, the optimal a depends on θ and equals

$$a(\theta) = \frac{p-2}{k}\,\frac{E_\theta\big[\frac{\|U\|^2}{\|X\|^2}\big]}{E_\theta\big[\frac{1}{\|X\|^2}\big]}. \tag{6.27}$$

Then the paradox will occur if, for any $a \geq 0$, $R\big(\delta_1^{(p-2)/(k+2)}, \theta\big) < R\big(\delta^a_0, \theta\big)$ and will certainly occur if $R\big(\delta_1^{(p-2)/(k+2)}, \theta\big) < R\big(\delta_0^{a(\theta)}, \theta\big)$ with $a(\theta)$ as in (6.27). By (6.26), this is equivalent to

$$\left(\frac{p-2}{k}\right)^2 \frac{\left(E_\theta\left[\frac{\|U\|^2}{\|X\|^2}\right]\right)^2}{E_\theta\left[\frac{1}{\|X\|^2}\right]} < \left(\frac{p-2}{k+2}\right)^2 E_\theta\left[\frac{\|U\|^4}{\|X\|^2}\right],$$

that is, to

$$\frac{\left(E_\theta\left[\frac{\|U\|^2}{\|X\|^2}\right]\right)^2}{E_\theta\left[\frac{\|U\|^4}{\|X\|^2}\right] E_\theta\left[\frac{1}{\|X\|^2}\right]} < \left(\frac{k}{k+2}\right)^2. \tag{6.28}$$

Expression (6.28) is a general condition for the paradox to occur. Fourdrinier and Strawderman (1996) developed a series of bounds for the quantities in the left-hand side of (6.28). However the resulting sufficient condition was complex and could be verified in a limited number of cases, the primary example being the Student Student-t distribution case. Subsequently Fourdrinier et al. (2004) developed an effective approach to deal with the expectations in (6.28) for the case of mixtures of normals.

Assume that (X, U) has a scale mixture of normals distribution with the representation

$$(X, U)|\,(Z = z) \sim \mathcal{N}_{p+k}\big((\theta, 0), z\, I_{p+k}\big) \tag{6.29}$$

where Z is a positive random variable. For model (6.29), expressions of the expectations in (6.28) are given by the following lemma.

Lemma 6.4 *Assume that (X, U) is a scale mixture of normals as in (6.29) and that $p \geq 3$. Let $q > -k/2$ and assume that $E[Z^{q-1}] < \infty$. Then we have*

$$E_\theta\left[\frac{\|U\|^{2q}}{\|X\|^2}\right] = 2^q\, \frac{\Gamma(k/2+q)}{\Gamma(k/2)}\, E\left[Z^{q-1} f_p\left(\frac{\|\theta\|^2}{Z}\right)\right]$$

where $f_p(\gamma) = E[Y^{-1}]$ for a random variable Y having a noncentral chi-square distribution with p degrees of freedom and noncentrality parameter γ.

Proof Note that X and U are independent conditional on Z and $(\|U\|^2/Z)|\,Z \sim \chi^2_k(0)$ and $(\|X\|^2/Z)|\,Z \sim \chi^2_p\big(\|\theta\|^2/Z\big)$. Hence we can write

$$\begin{aligned} E_\theta\left[\frac{\|U\|^{2q}}{\|X\|^2}\,\middle|\, Z\right] &= E[\|U\|^{2q}|\, Z] E_\theta\left[\frac{1}{\|X\|^2}\,\middle|\, Z\right] \\ &= Z^{q-1} E\left[\left(\frac{\|U\|^2}{Z}\right)^q \,\middle|\, Z\right] E_\theta\left[\frac{Z}{\|X\|^2}\,\middle|\, Z\right] \\ &= Z^{q-1} 2^q\, \frac{\Gamma(k/2+q)}{\Gamma(k/2)}\, f_p\left(\frac{\|\theta\|^2}{Z}\right) \end{aligned}$$

since $q > -k/2$. Now use the fact that f_p is bounded if $p \geq 3$ and $E[Z^{q-1}] < \infty$ and uncondition to complete the proof. □

It follows directly from Lemma 6.4 for $q = 0, 1, 2$ that (6.28) is equivalent to

$$H_Z(\lambda) = \frac{\left(E\left[f_p(\lambda^2/Z)\right]\right)^2}{E\left[Zf_p(\lambda^2/Z)\right]E\left[Z^{-1}f_p(\lambda^2/Z)\right]} < \frac{k}{k+2} \tag{6.30}$$

for all $\lambda = \|\theta\| \geq 0$.

Alternatively note that

$$H_Z(\lambda) = \left(E_\lambda[W]E_\lambda[W^{-1}]\right)^{-1} \tag{6.31}$$

where W is a positive random variable with density

$$h_\lambda(w) = c(\lambda) f_p(\lambda^2 w) g(w)$$

where g is the density of $V = Z^{-1}$ and $c(\lambda)$ is a normalizing constant. Then (6.28) can also be expressed as

$$E_\lambda[W]E_\lambda[W^{-1}] > 1 + \frac{2}{k} \tag{6.32}$$

for all $\lambda \geq 0$.

The following main result shows that the paradox occurs for any nondegenerate mixture of normals when the dimension of the residual vector U is sufficiently large.

Theorem 6.2 *Assume that (X, U) is a scale mixture of normals as in (6.29), with Z nondegenerate, $E[Z] < \infty$ and $E[Z^{-1}] < \infty$. Then, for any $p \geq 3$, there exists a positive integer k_0 such that, for any integer $k \geq k_0$, the optimal robust James-Stein estimator $\delta_{RJS}^{(p-2)}$ ($= \delta_1^{(p-2)/(k+2)}$) simultaneously dominates all James-Stein estimators δ_{JS}^a ($= \delta_0^a$).*

Proof Setting $\bar{H} = \sup_{\lambda \geq 0} H_Z(\lambda)$, Condition (6.30) reduces to $k > 2\frac{\bar{H}}{1-\bar{H}}$. From (6.31) we know (by covariance inequality) that $H_Z(\lambda) \leq 1$ with equality if and only if W is degenerate, that is, if and only if Z is degenerate, which corresponds to the normal case. Then $\bar{H} \leq 1$ and we only need to show that $\bar{H} < 1$ since H_Z is continuous, and hence $\bar{H}$ does not depend on k.

Now it can be shown (see Lemma 3 in Fourdrinier et al. 2004) that

$$\begin{aligned}
\lim_{\lambda \to \infty} H_Z(\lambda) &= \left(\lim_{\lambda \to \infty} E_\lambda[W] \lim_{\lambda \to \infty} E_\lambda[W^{-1}]\right)^{-1} \\
&= \left(\frac{1}{E[Z]} \cdot \frac{E[Z^2]}{E[Z]}\right)^{-1} \\
&= \frac{(E[Z])^2}{E[Z^2]} \\
&< 1,
\end{aligned}$$

for $p \geq 3$ and nondegenerate Z. Since $H_Z(\lambda) < 1$ for all λ and $\lim_{\lambda\to\infty} H_Z(\lambda) < 1$, this implies $\bar{H} < 1$. □

The necessity of nondegeneracy of Z is explicit in the proof of Theorem 6.2. Therefore the paradox occurs only in the case of nondegenerate mixtures of normals and not in the normal case, as previously noted.

Outside the class of mixtures of normals little is known. In the case where the radial distribution is concentrated on two points, Fourdrinier and Strawderman (1996) show that the paradox can occur for suitable weights. Showing the existence of the paradox in other families of spherically symmetric distributions is an open question.

6.3 Bayes Estimators

Let (X, U) be a random vector in $\mathbb{R}^p \times \mathbb{R}^k$ with density

$$\frac{1}{\sigma^{p+k}}\, f\left(\frac{\|x-\theta\|^2 + \|u\|^2}{\sigma^2}\right), \tag{6.33}$$

where $\theta \in \mathbb{R}^p$ and $\sigma \in \mathbb{R}_+\backslash\{0\}$ are unknown. We assume throughout that $p \geq 3$.

We consider generalized Bayes estimators of θ for priors of the form

$$\pi(\|\theta\|^2)\, \eta^b\,, \tag{6.34}$$

where $\eta = 1/\sigma^2$, under the invariant quadratic loss in (6.16).

We first show that, under weak moment conditions, such generalized Bayes estimators are robust in the sense that they do not depend on the underlying density f. Furthermore, we exhibit a large class of superharmonic priors π for which these generalized Bayes estimators dominate the usual minimax estimator X for the entire class of densities (6.33). Hence this subclass of estimators has the extended robustness property of being simultaneously generalized Bayes and minimax for the entire class of spherically symmetric distributions.

Note that, paralleling Sect. 4.5, the above model arises as the canonical form of the general linear model $Y = V\beta + \varepsilon$ where V is a $(p+k) \times p$ design matrix, β is a $p \times 1$ vector of unknown regression coefficients, and ε is an $(p+k) \times 1$ error vector with spherically symmetric density $f(\|\varepsilon\|^2/\sigma^2)/\sigma^{p+k}$.

In the following, for a real valued function $g(x, \|u\|^2)$, we denote by $\nabla_x g(x, u)$ and $\Delta_x g(x, \|u\|^2)$ the gradient and the Laplacian of $g(x, \|u\|^2)$ with respect to the variable x. Analogous notations hold with respect to the variable u. When $g(x, \|u\|^2)$ is a vector valued function, $\text{div}_x g(x, \|u\|^2)$ is the divergence with respect to x (here $\dim g(x, \|u\|^2) = \dim\ x$).

As previously noted, Stein (1981) shows that, when the density in (6.33) is normal with known scale, the generalized Bayes estimator corresponding to a prior

$\pi(\theta)$, for which the square root of the marginal density $m(x)$ is superharmonic, is minimax under the loss (6.16). Fundamental to this result is the development of an unbiased estimator of risk based on a differential expression involving $m(x)$ which has become a basic tool in proving minimaxity.

Another line of research pertinent to this section is the development of Bayes and generalized Bayes minimax estimators. In the case of a normal distribution with known scale, see Sect. 3.1, When the scale is unknown, see Sect. 3.4. For variance mixture of normals and, more generally, for spherically symmetric distributions with no residual, see Sect. 5.4.

Maruyama (2003b) showed that, for spherically symmetric distributions with a residual vector U and unknown scale parameter, the generalized Bayes estimator with respect to a prior on θ and η proportioned to $\eta^b \, \|\theta\|^{-a}$ is independent of the density f and is minimax under conditions on a and b and under weak moment conditions (see also Maruyama and Takemura 2008 and Maruyama and Strawderman 2005, 2009).

The goal of this section is to extend the phenomenon in Maruyama (2003b) to a broader class of priors of the form $\pi(\|\theta\|^2)\,\eta^b$ with $\pi(\|\theta\|^2)$ superharmonic. In particular, in Sect. 6.3.1, we show that the generalized Bayes estimators do not depend on the density f under weak moment conditions and, in Sect. 6.3.2, we prove that these generalized Bayes estimators are minimax provided the prior $\pi(\|\theta\|^2)$ is superharmonic and its Laplacian $\Delta\pi(\|\theta\|^2)$ is a nondecreasing function of $\|\theta\|^2$, under conditions on b, p and k.

In the case of a known scale parameter, Fourdrinier and Strawderman (2008a) studied the same class of priors $\pi(\|\theta\|^2)$ and proved minimaxity of generalized Bayes estimators for a large subclass of unimodal densities. We rely strongly on the techniques of that paper, as presented in Sect. 5.4.

6.3.1 Form of the Bayes Estimators

In Sect. 3.2 generalized Bayes estimators for the normal setting with an unknown variance were discussed. In this subsection we extend the normal case to the spherical setting with a residual vector, that is when the sampling distribution is of the form of (6.33). In the normal setting the generalized Bayes estimators in (3.25) were of the form $X - \frac{r(F)}{F} X$ where $F = ||X||^2/\|U\|^2$. In the more general setting of this subsection the shrinkage function is not a function of only F but is a more general function of both X and $\|U\|^2$ as in (3.17).

The results of this subsection and the next closely follow the developments in Fourdrinier and Strawderman (2010). We will see that for the sampling distribution in (6.33) and priors of the form (6.34), the generalized Bayes estimators do not depend on the density (6.33); more precisely their expressions depend only on π and b provided that

$$\int_0^{\infty} f(\tau)\, \tau^{(p+k)/2+b+1}\, d\tau < \infty, \tag{6.35}$$

which is equivalent to

$$E_{0,1}\Big[(\|X\|^2 + \|U\|^2)^{2(b+2)}\Big] < \infty\,.$$

Proposition 6.3 *For a prior of the form* (6.34) *and loss* (6.16)*, the generalized Bayes estimator* $\delta(X, \|U\|^2) = X + g(X, \|U\|^2)$ *is such that, for any* $(x, u) \in \mathbb{R}^p \times \mathbb{R}^k$*,*

$$g(x, \|u\|^2) = \frac{\int_{\mathbb{R}^p} \frac{\theta - x}{(\|x-\theta\|^2 + \|u\|^2)^{(p+k)/2+b+2}}\, \pi(\|\theta\|^2)\, d\theta}{\int_{\mathbb{R}^p} \frac{1}{(\|x-\theta\|^2 + \|u\|^2)^{(p+k)/2+b+2}}\, \pi(\|\theta\|^2)\, d\theta}, \tag{6.36}$$

provided (6.35) *holds and* (6.36) *exists and hence* $\delta(X, \|U\|^2)$ *does not depend on* $f(\cdot)$*.*

Note that $g(x, \|u\|^2)$ arises as

$$\frac{\nabla_x M(x, \|u\|^2)}{m(x, \|u\|^2)}\,,$$

where $m(x, \|u\|^2)$ is the marginal associated to π and the density

$$\varphi\left(\|x-\theta\|^2 + \|u\|^2\right) \propto \frac{1}{(\|x-\theta\|^2 + \|u\|^2)^{(p+k)/2+b+2}}\,, \tag{6.37}$$

and M is the marginal associated to ϕ with

$$\phi(t) = \frac{1}{2}\int_t^{\infty} \varphi(v)\, dv\,. \tag{6.38}$$

Therefore, for each fixed u, $\delta(X, u) = X + g(X, u)$ with $g(X, u)$ in (6.36) can be interpreted as the Bayes estimator of θ under the density φ and the prior π for fixed scale parameter $\|u\|$ under the loss$\|\delta - \theta\|^2$. This observation will be important in the next subsection since it will allow us to use results in Sect. 5.4 (Fourdrinier and Strawderman 2008a) which are developed for the case of known scale parameter.

Finally, note that existence of (6.36) will be guaranteed by the stronger finiteness risk condition developed in the proof of Theorem 6.3. More generally, it suffices that π be locally integrable and have tails that do not grow too fast at infinity. In particular, superharmonic priors are locally integrable and have bounded tails.

Proof of Proposition 6.3. The Bayes estimator under loss (6.16) is

$$\delta(X, \|U\|^2) = \frac{E[\eta\,\theta|X, U]}{E[\eta|X, U]} = X + g(X, \|U\|^2),$$

with, for any $(x, u) \in \mathbb{R}^p \times \mathbb{R}^k$,

$$\begin{aligned} g(x, \|u\|^2) &= \frac{E[\eta\,(\theta - x) \mid x, u]}{E[\eta|x, u]} \\ &= \frac{\int_0^\infty \int_{\mathbb{R}^p} \eta\,(\theta - x)\,\eta^{(p+k)/2}\, f(\eta\,(\|x-\theta\|^2 + \|u\|^2))\,\pi(\|\theta\|^2)\,\eta^b\,d\theta\,d\eta}{\int_0^\infty \int_{\mathbb{R}^p} \eta^{(p+k)/2+1}\, f(\eta\,(\|x-\theta\|^2 + \|u\|^2))\,\pi(\|\theta\|^2)\eta^b\,d\theta\,d\eta} \\ &= \frac{\int_{\mathbb{R}^p} \left(\int_0^\infty \eta^{(p+k)/2+b+1}\, f(\eta\,(\|x-\theta\|^2 + \|u\|^2))\,d\eta\right)(\theta - x)\,\pi(\|\theta\|^2)\,d\theta}{\int_{\mathbb{R}^p} \left(\int_0^\infty \eta^{(p+k)/2+b+1}\, f(\eta\,(\|x-\theta\|^2 + \|u\|^2))\,d\,\eta\right)\pi(\|\theta\|^2)\,d\theta}\,, \end{aligned}$$

by Fubini's theorem. Now, through the change of variable $\tau = \eta\,(\|x-\theta\|^2 + \|u\|^2)$ in the innermost integrals, we obtain

$$\begin{aligned} g(x, \|u\|^2) &= \frac{\int_{\mathbb{R}^p} \int_0^\infty \tau^{(p+k)/2+b+1}\, f(\tau)\,d\tau\, \frac{(\theta-x)\,\pi(\|\theta\|^2)}{(\|x-\theta\|^2+\|u\|^2)^{(p+k)/2+b+2}}\,d\theta}{\int_{\mathbb{R}^p} \int_0^\infty \tau^{(p+k)/2+b+1}\, f(\tau)\,d\tau\, \frac{\pi(\|\theta\|^2)}{(\|x-\theta\|^2+\|u\|^2)^{(p+k)/2+b+2}}\,d\theta} \\ &= \frac{\int_{\mathbb{R}^p} \frac{(\theta-x)\,\pi(\|\theta\|^2)}{(\|x-\theta\|^2+\|u\|^2)^{(p+k)/2+b+2}}\,d\theta}{\int_{\mathbb{R}^p} \frac{\pi(\|\theta\|^2)}{(\|x-\theta\|^2+\|u\|^2)^{(p+k)/2+b+2}}\,d\theta} \end{aligned}$$

thanks to (6.35). □

6.3.2 Minimaxity of Generalized Bayes Estimators

According to the expression of $g(X, \|U\|^2)$ in (6.36), we give an expression of the differential operator $\mathscr{O}g(X, \|U\|^2)$ in (6.23). The proof of Proposition 6.4 follows from straightforward calculations.

Proposition 6.4 *For* $g(X, \|U\|^2) = \frac{\nabla_X M(X, \|U\|^2)}{m(X, \|U\|^2)}$, (6.23) *can be expressed as*

$$\begin{aligned} \mathscr{O}g(X, \|U\|^2) = 2\,\frac{\Delta_X M(X, \|U\|^2)}{m(X, \|U\|^2)} - 2\,\frac{\nabla_X m(X, \|U\|^2)^{\mathrm{T}} \nabla_X M(X, \|U\|^2)}{m^2(X, \|U\|^2)} \qquad (6.39) \\ + \frac{k-2}{\|U\|^2} \left\| \frac{\nabla_X M(X, \|U\|^2)}{m(X, \|U\|^2)} \right\|^2 + 2\,\frac{\partial}{\partial s} \left\| \frac{\nabla_X M(X, s)}{m(X, s)} \right\|^2 \Bigg|_{s=\|U\|^2}, \end{aligned}$$

where, for any $(x, u) \in \mathbb{R}^p \times \mathbb{R}^k$,

$$m(x, \|u\|^2) = \int_{\mathbb{R}^p} \varphi(\|x - \theta\|^2 + \|u\|^2)\ \pi(\|\theta\|^2)\, d\theta, \tag{6.40}$$

and

$$M(x, \|u\|^2) = \int_{\mathbb{R}^p} \phi(\|x - \theta\|^2 + \|u\|^2)\, \pi(\|\theta\|^2)\, d\theta \tag{6.41}$$

with φ *and* ϕ *given by* (6.37) *and* (6.38).

In Sect. 5.4, we studied Bayes minimax estimation of a location vector in the case of spherically symmetric distributions with known scale parameter. For a subclass of spherically symmetric densities, we proved minimaxity of generalized Bayes estimators for spherically symmetric priors of the form $\pi(\|\theta\|^2)$ under the following assumptions (see Theorem 5.7 and also Fourdrinier and Strawderman 2008a, 2010).

Assumption 1

(1) $\pi'(\|\theta\|^2) \quad \leq 0$ i.e. $\pi(\|\theta\|^2)$ is unimodal;
(2) $\Delta\pi(\|\theta\|^2) \quad \leq 0$ i.e. $\pi(\|\theta\|^2)$ is superharmonic;
(3) $\Delta\pi(\|\theta\|^2)$ is nondecreasing in $\|\theta\|^2$.

Note that Condition (2) in fact implies Condition (1) by the mean value property of superharmonic functions. Several examples of priors which satisfy Assumption 1 have been given in Sect. 5.4: Examples 5.8, 5.9 and 5.10.

Our main result below is that a generalized Bayes estimator of θ for a density (6.33), a prior (6.34) and the loss (6.16) is minimax under weak moment conditions and conditions on b, provided the prior satisfies the Assumptions above. We remind the reader that, according to Proposition 6.3, the generalized Bayes estimator is independent of the sampling density, f, provided the assumption (6.35) holds. Hence, each such estimator is simultaneously generalized Bayes and minimax for the entire class of spherically symmetric distributions.

Before developing our minimaxity result, we give a theorem which guarantees the risk finiteness of the generalized Bayes estimators.

Theorem 6.3 *Assume that* π *satisfies Assumption 1 and that* $b > -(k/2+1)$. *Then the generalized Bayes estimator associated to* π *has finite risk.*

Proof According to (6.36), the risk finiteness condition (6.17) is satisfied as soon as

$$E_{\theta,\eta}\left[\left\| \frac{\int_{\mathbb{R}^p} (\theta - X)\, \frac{\pi(\|\theta\|^2)}{(\|X-\theta\|^2+\|U\|^2)^{(p+k)/2+b+2}}\, d\theta}{\int_{R^p} \frac{\pi(\|\theta\|^2)}{(\|X-\theta\|^2+\|U\|^2)^{(p+k)/2+b+2}}\, d\theta} \right\|^2\right]$$
$$\leq E_{\theta,\eta}\left[\frac{\int_{R^p} \|\theta - X\|^2\, \frac{\pi(\|\theta\|^2)}{(\|X-\theta\|^2+\|U\|^2)^{(p+k)/2+b+2}} d\theta}{\int_{R^p} \frac{\pi(\|\theta\|^2)}{(\|X-\theta\|^2+\|U\|^2)^{(p+k)/2+b+2}}\, d\theta}\right]$$
$$< \infty\,. \tag{6.42}$$

Note that, for any $(x, u) \in \mathbb{R}^p \times \mathbb{R}^k$ and for any nonnegative function h on $\mathbb{R}_+ \times \mathbb{R}_+$ (see Lemma 1.4),

$$\int_{R^p} \pi(\|\theta\|^2)\, h(\|x-\theta\|^2, \|u\|^2)\, d\theta = \int_0^\infty \int_{S_{R,x}} \pi(\|\theta\|^2)\, d\mathscr{U}_{R,x}(\theta)\, \sigma(S)\, R^{p-1}\, h(R^2, \|u\|^2)\, dR\,, \quad (6.43)$$

where $\mathscr{U}_{R,x}$ is the uniform distribution on the sphere $S_{R,x}$ of radius R and centered at x and $\sigma(S)$ is the area of the unit sphere. Through the change of variable $R = \sqrt{v}$, the right hand side of (6.43) can be written as

$$\int_0^\infty \mathscr{S}_\pi(\sqrt{v}, x)\, v^{p/2-1}\, h(v, \|U\|^2)\, dv,$$

where

$$\mathscr{S}_\pi(\sqrt{v}, x) = \frac{\sigma(S)}{2} \int_{S_{\sqrt{v},x}} \pi(\|\theta\|^2)\, d\mathscr{U}_{\sqrt{v},x}(\theta)$$

is nonincreasing in v by the superharmonicity of $\pi(\|\theta\|^2)$.

Now we can express the last quantity in brackets in (6.42) as

$$\frac{\int_o^\infty \mathscr{S}_\pi(\sqrt{v}, x)\, \frac{v^{p/2}}{(v+\|u\|^2)^{(p+k)/2+b+2}}\, dv}{\int_o^\infty \mathscr{S}_\pi(\sqrt{v}, x)\, \frac{v^{p/2-1}}{(v+\|u\|^2)^{(p+k)/2+b+2}}\, dv} = E_1[v] \le E_2[v], \quad (6.44)$$

where E_1 is the expectation with respect to the density $f_1(v)$ proportional to

$$\mathscr{S}_\pi(\sqrt{v}, x)\, \frac{v^{p/2-1}}{(v + \|u\|^2)^{(p+k)/2+b+2}},$$

and E_2 is the expectation with respect to the density $f_2(v)$ proportional to

$$\frac{v^{p/2-1}}{(v + \|u\|^2)^{(p+k)/2+b+2}}.$$

Indeed the ratio $f_2(v)/f_1(v)$ is nondecreasing by the monotonicity of $\mathscr{S}_\pi(\sqrt{v}, x)$. In (6.44), $E_2[v]$ is

$$E_2[v] = \frac{\int_0^\infty \frac{v^{p/2}}{(v+\|u\|^2)^{(p+k)/2+b+2}}\, dv}{\int_0^\infty \frac{v^{p/2-1}}{(v+\|u\|^2)^{(p+k)/2+b+2}}\, dv}$$

$$= \|u\|^2 \, \frac{\int_0^\infty \frac{v^{p/2}}{(v+1)^{(p+k)/2+b+2}}\, dv}{\int_0^\infty \frac{v^{p/2-1}}{(v+1)^{(p+k)/2+b+2}}\, dv}$$

$$= \|u\|^2 \, \frac{B(p/2+1, k/2+b+1)}{B(p/2, k/2+b+2)},$$

which is finite for $k/2 + b + 1 > 0$.

Finally the expectations in (6.42) are bounded above by $K\, E_{\theta,\eta}[\|U\|^2]$ where K is a constant, and hence are finite. □

We will need the following result which is essentially a reexpression of Lemma 5.6.

Lemma 6.5 *Let $m(x, \|u\|^2)$ and $M(x, \|u\|^2)$ be as defined in (6.40) and (6.41) and let $\cdot$ be the inner product in $\mathbb{R}^p$. Then we have*

(1)

$$x \cdot \nabla_x m(x, \|u\|^2) = -2 \int_0^\infty H(v, \|x\|^2)\, v^{p/2}\, \varphi'(v + \|u\|^2)\, dv,$$

and

$$x \cdot \nabla_x M(x, \|u\|^2) = \int_0^\infty H(v, \|x\|^2)\, v^{p/2}\, \varphi(v + \|u\|^2)\, dv,$$

where, for $v > 0$,

$$H(v, \|x\|^2) = \lambda(B) \int_{B_{\sqrt{v},x}} x \cdot \theta\, \pi'(\|\theta\|^2)\, d\mathscr{V}_{\sqrt{v},x}(\theta) \tag{6.45}$$

and $\mathscr{V}_{\sqrt{v},x}$ is the uniform distribution on the ball $B_{\sqrt{v},x}$ of radius $\sqrt{v}$ centered at x and $\lambda(B)$ is the volume of the unit ball;

(2) *For any $x \in \mathbb{R}^p$, the function $H(v, \|x\|^2)$ in (6.45) is nondecreasing in v provided that $\Delta\pi(\|\theta\|^2)$ is nondecreasing in $\|\theta\|^2$. (Assumption* 1 (3)*);*

(3) *For any $v > 0$ and any $x \in \mathbb{R}^p$, the function $H(v, \|x\|^2)$ in (6.45) is nonpositive provided $\pi'(\|\theta\|^2) \leq 0$. (Assumption* 1 (1)*).*

Given these preliminaries, we present our main result.

Theorem 6.4 *Suppose that π satisfies Assumption* 1*. Then the generalized Bayes estimator associated to $\pi(\|\theta\|^2)\, \eta^b$ is minimax provided that $b \geq \frac{2p-k-2}{4}$ and the assumptions of Theorem* 6.3 *are satisfied.*

Proof It suffices to show that $\mathscr{O}g(X, \|U\|^2)$ in (6.38), with $m(X, \|U\|^2)$ and $M(X, \|U\|^2)$ given respectively by (6.39) and (6.41), is non positive since the assumptions guarantee that the generalized Bayes estimator δ is of the form $\delta(X, \|U\|^2) = X + \nabla_X M(X, \|U\|^2)/m(X, \|U\|^2)$ and has finite risk.

Due to the superharmonicity of $\pi(\|\theta\|^2)$, for any $(x, u) \in \mathbb{R}^p \times \mathbb{R}^k$, we have $\Delta_x M(x\|u\|^2) \leq 0$ so that

$$\mathscr{O}g(x, \|u\|^2) \leq -2\,\frac{\nabla_x m(x, \|u\|^2)^{\mathrm{T}} \nabla_x M(x, \|u\|^2)}{m^2(x, \|u\|^2)} + \frac{k-2}{\|u\|^2}\left\|\frac{\nabla_x M(x, \|u\|^2)}{m(x, \|u\|^2)}\right\|^2 + 2\frac{\partial}{\partial s}\left\|\frac{\nabla_x M(x, s)}{m(x, s)}\right\|^2\Bigg|_{s=\|u\|^2}.$$

Note that

$$\begin{aligned} m^2(x, s)\,\frac{\partial}{\partial s}\left\|\frac{\nabla_x M(x, s)}{m(x, s)}\right\|^2 &= \frac{\partial}{\partial s}\,\|\nabla_x M(x, s)\|^2 + \|\nabla_x M(x, s)\|^2\, m^2(x, s)\,\frac{\partial}{\partial s}\frac{1}{m^2(x, s)} \\ &\leq \frac{\partial}{\partial s}\,\|\nabla_x M(x, s)\|^2 + (p+k+2b+4)\,\frac{1}{s}\,\|\nabla_x M(x, s)\|^2, \end{aligned}$$

since

$$\begin{aligned} \frac{\partial}{\partial s}\frac{1}{m^2(x, s)} &= \frac{-2}{m^3(x, s)}\int_{\mathbb{R}^p}\frac{-[(p+k)/2+b+2]}{(\|x-\theta\|^2+s)^{(p+k)/2+b+3}}\,\pi(\|\theta\|^2)\,d\theta \\ &= \frac{p+k+2b+4}{m^3(x, s)}\,\frac{1}{s}\int_{\mathbb{R}^p}\frac{s}{\|x-\theta\|^2+s}\,\frac{1}{(\|x-\theta\|^2+s)^{(p+k)/2+b+2}}\,\pi(\|\theta\|^2)\,d\theta \\ &\leq \frac{p+k+2b+4}{m^2(x, s)}\,\frac{1}{s}\,. \end{aligned}$$

Therefore

$$\begin{aligned} m^2(x, s)\,\mathscr{O}g(x, s) \leq{}& -2\,\nabla_x m(x, s)^{\mathrm{T}}\nabla_x M(x, s) \\ &+ \frac{k-2+2(p+k+2b+4)}{s}\,\|\nabla_x M(x, s)\|^2 \\ &+ 2\frac{\partial}{\partial s}\,\|\nabla_x M(x, s)\|^2\,. \end{aligned} \tag{6.46}$$

As $m(x, s)$ and $M(x, s)$ depend on x only through $\|x\|^2$, it is easy to check that (as in Fourdrinier and Strawderman 2008a)

$$\nabla_x m(x, s)^{\mathrm{T}}\nabla_x M(x, s) = \frac{x^{\mathrm{T}}\nabla_x m(x, s)\, x^{\mathrm{T}}\nabla_x M(x, s)}{\|x\|^2}$$

and

$$\|\nabla_x M(x,s)\|^2 = \frac{(x^{\mathrm{T}}\nabla_x M(x,s))^2}{\|x\|^2}\,.$$

Thus the right hand side of (6.46) will be nonpositive as soon as

$$-\,2x^{\mathrm{T}}\nabla_x m(x,s) + \frac{2p+3k+4b+6}{s}\,x^{\mathrm{T}}\nabla_x M(x,s) + 4\frac{\partial}{\partial s}x^{\mathrm{T}}\nabla_x M(x,s) \geq 0, \tag{6.47}$$

since, according to Lemma 6.5, the common factor $x^{\mathrm{T}}\nabla_x M(x,s)$ is nonpositive. Using again Lemma 6.5, the left hand side of (6.47) equals

$$\begin{aligned}
&4\int_0^\infty H(v,\|x\|^2)\,v^{p/2}\,\varphi'(v+s)\,dv \\
&\qquad+\frac{2p+3k+4b+6}{s}\int_0^\infty H(v,\|x\|^2)\,v^{p/2}\,\varphi(v+s)\,dv \\
&\qquad\qquad+4\int_o^\infty H(v,\|x\|^2)v^{p/2}\varphi'(v+s)dv \\
&=\int_0^\infty v^{p/2}\varphi(v+s)\left\{8\,E\left[H(v,\|x\|^2)\,\frac{\varphi'(v+s)}{\varphi(v+s)}\right]\right. \\
&\qquad\left.+\frac{2\,p+3\,k+4\,b+6}{s}E\left[H(v,\|x\|^2)\right]\right\}dv\,,
\end{aligned} \tag{6.48}$$

where E denotes the expectation with respect to the density proportional to $v \longmapsto v^{p/2}\,\varphi(v+s)$.

As

$$\frac{\varphi'(v+s)}{\varphi(v+s)} = \frac{-((p+k)/2+b+2)}{v+s} \tag{6.49}$$

is nondecreasing in v and, according to Lemma 6.5, $H(v,\|x\|^2)$ is also nondecreasing in v, the first expectation in (6.48) satisfies

$$E\Big[H(v,\|x\|^2)\,\frac{\varphi'(v+s)}{\varphi(v+s)}\Big] \geq E\Big[H(v,\|x\|^2)\Big]\,E\left[\frac{\varphi'(v+s)}{\varphi(v+s)}\right]$$

by the covariance inequality. Therefore Inequality (6.47) will be satisfied as soon as

$$8\,E\left[\frac{\varphi'(v+s)}{\varphi(v+s)}\right] + \frac{2p+3k+4b+6}{s} \leq 0, \tag{6.50}$$

since $H(v, \|x\|^2) \leq 0$ by Lemma 6.5.

From (6.49) we have

$$E\left[\frac{\varphi'(v+s)}{\varphi(v+s)}\right] = -\big((p+k)/2+b+2\big)\, E\left[\frac{1}{v+s}\right] \tag{6.51}$$

$$= -\big((p+k)/2+b+2\big) \frac{\int_o^\infty \frac{1}{v+s}\ v^{p/2} \frac{1}{(v+s)^{(p+k)/2+b+2}}\, dv}{\int_0^\infty v^{p/2} \frac{1}{(v+s)^{(p+k)/2+b+2}}\, dv}$$

$$= -\big((p+k)/2+b+2\big)\, \frac{1}{s}\, \frac{\int_0^\infty \frac{z^{p/2}}{(z+1)^{(p+k)/2+b+3}}\, dz}{\int_0^\infty \frac{z^{p/2}}{(z+1)^{(p+k)/2+b+2}}\, dz}$$

$$= -\big((p+k)/2+b+2\big)\, \frac{1}{s}\, \frac{B(p/2+1, k/2+b+2)}{B(p/2+1, k/2+b+1)},$$

where $B(\alpha, \beta)$ is the beta function with parameters $\alpha > 0$ and $\beta > 0$. Then (6.51) becomes

$$E\left[\frac{\varphi'(v+s)}{\varphi(v+s)}\right] = -\frac{((p+k)/2+b+2)}{s}\ \frac{\Gamma((k/2+b+2))}{\Gamma((p+k)/2+b+3)}$$

$$= \frac{\Gamma((p+k)/2+b+2)}{\Gamma(k/2+b+1)}\ \frac{-(k/2+b+1)}{s}. \tag{6.52}$$

It follows from (6.52) that (6.50) reduces to

$$b \geq \frac{2p-k-2}{4},$$

which is the condition given in the theorem. □

The condition on b in Theorem 6.4 can be alternatively expressed as $k \geq 2p - 4b - 2$ which dictates that the dimension, k, of the residual vector, U, increases with the dimension, p, of θ. This dependence can be (essentially) eliminated provided the generalized Bayes estimator in Proposition 6.3 satisfies the following assumption.

Assumption 2 The function $g(x, \|u\|^2)$ in (6.36) can be expressed as

$$g(x, \|u\|^2) = \frac{\nabla_x M(x, \|u\|^2)}{m(x, \|u\|^2)} = -\frac{r(\|x\|^2, \|u\|^2)\, \|u\|^2}{\|x\|^2}\, x,$$

where $r(\|x\|^2, \|u\|^2)$ is nonnegative and nonincreasing in $\|u\|^2$.

Assumption 2 is satisfied, for example, by the generalized Bayes estimator corresponding to the prior on (θ, η) proportional to $\pi(\|\theta\|^2) = \big(1/\|\theta\|^2\big)^{-b/2}\, \eta^a$ for

$0 < b \le p-2$ and $a > -\frac{k}{2} - \frac{b}{2} - 2$, in which case the function $r(\|x\|^2, \|u\|^2) = \phi\left(\|x\|^2/\|u\|^2\right)$, where $\phi(t)$ is increasing in t, and hence $r(\|x\|^2, \|u\|^2)$ is decreasing in $\|u\|^2$ (see, Maruyama 2003b).

We have the following corollary.

Corollary 6.2 *Suppose π satisfies Assumptions 1 and the assumptions of Theorem 6.4 and suppose also that the generalized Bayes estimator (which does not depend on the underlying density f) satisfies Assumption 2. Then the generalized Bayes estimator is minimax provided $b \ge -(k+2)/4$.*

Proof Assumption 2 guarantees that

$$\frac{\partial}{\partial s}\left(\frac{1}{s^2}\left\|\frac{\nabla_x M(x,s)}{m(x,s)}\right\|^2\right) = \frac{\partial}{\partial s}\left(\frac{r^2(\|x\|^2, s)}{\|x\|^2}\right) \le 0.$$

Since

$$\begin{aligned}\frac{\partial}{\partial s}\left\|\frac{\nabla_x M(x,s)}{m(x,s)}\right\|^2 &= \frac{\partial}{\partial s}\left(\frac{s^2}{s^2}\left\|\frac{\nabla_x M(x,s)}{m(x,s)}\right\|^2\right)\\ &= \frac{2}{s}\left\|\frac{\nabla_x M(x,s)}{m(x,s)}\right\|^2 + s^2\frac{\partial}{\partial s}\left(\frac{1}{s^2}\left\|\frac{\nabla_x M(x,s)}{m(x,s)}\right\|^2\right),\end{aligned}$$

the inequality for $\mathscr{O}g(X, \|U\|^2)$ in the proof of Theorem 6.4 can be replaced by

$$\mathscr{O}g(x, \|u\|^2) \le -2\,\frac{\nabla_x m(x, \|u\|^2)^{\mathrm{T}} \nabla_x M(x, \|u\|^2)}{m^2(x, \|u\|^2)} + \frac{k+2}{\|u\|^2}\left\|\frac{\nabla_x M(x, \|u\|^2)}{m(x, \|u\|^2)}\right\|^2 .$$

It follows that inequality condition (6.47) becomes

$$-2\,x^{\mathrm{T}}\nabla_x m(x,s) + \frac{k+2}{s}\,x^{\mathrm{T}}\nabla_x M(x,s) \ge 0,$$

and that inequality condition (6.50) becomes

$$4\,E\left[\frac{\varphi'(v+s)}{\varphi(v+s)}\right] + \frac{k+2}{s} \le 0,$$

which, by (6.52), becomes

$$4\left[-\left(\frac{k/2+b+1}{s}\right)\right] + \frac{k+2}{s} \le 0,$$

which is equivalent to $b \ge -(k+2)/4$. □

6.4 The Unknown Covariance Matrix Case

In this section, we consider estimation of the mean vector in the case of elliptically symmetric distribution with an unknown nonsingular scale matrix. Most of the material of this section is taken from Fourdrinier et al. (2003). We assume there is sufficient data in the form of residual vectors to estimate the unknown covariance matrix. In the canonical form of this model, $X, V_1, \ldots, V_{n-1}$ are n random vectors in $\mathbb{R}^p$ with joint density of the form

$$|\Sigma|^{-n/2} f\left((x-\theta)^{\mathrm{T}} \Sigma^{-1}(x-\theta) + \sum_{j=1}^{n-1} V_j^{\mathrm{T}} \Sigma^{-1} V_j\right) \tag{6.53}$$

where the $p \times 1$ location vector θ and the $p \times p$ scale matrix Σ are unknown. Note occasionally we will absorb the normalizing factor $|\Sigma^{-1}|^{n/2}$ in the function f. If both θ and Σ are unknown, X and $S = \sum_{j=1}^{n-1} V_j\, V_j^{\mathrm{T}} = V\, V^{\mathrm{T}}$ are minimal sufficient statistics. Throughout this section, we assume that $p \leq n-1$ so that S is invertible.

The canonical form (6.53) arises through an $n \times n$ orthogonal transformation of

$$(Y_1, \ldots, Y_n) \sim |\Sigma|^{-n/2} f\left(\sum_{j=1}^{n} (Y_i-\theta)^{\mathrm{T}} \Sigma^{-1} (Y_i-\theta)\right)$$

as in the case of an i.i.d. sample of size n from a $\mathcal{N}_p(\theta, \Sigma)$ distribution.

To show this reduction to the canonical form define the $p \times n$ matrices $\mathbf{Y} = (Y_1 : \ldots : Y_n)$ for $Y_i \in \mathbb{R}^p$ and $\Theta = (\theta : \ldots : \theta)$. Let P be an $n \times n$ orthogonal matrix such that the first row of P is $\mathbf{1}_n^{\mathrm{T}}/\sqrt{n}$, where $\mathbf{1}_n^{\mathrm{T}} = (1, \ldots, 1)$ is the $1 \times n$ row vector of ones. Let the $p \times n$ matrices $\mathbf{X} = (X_1 : \ldots : X_n)$ and $\boldsymbol{\nu}_{\mathrm{T}=(\nu_1:\ldots:\nu_n)}$ be defined through $\mathbf{X}^{\mathrm{T}} = P\, \mathbf{Y}^{\mathrm{T}}$ and $\boldsymbol{\nu}^{\mathrm{T}} = P\, \Theta^{\mathrm{T}}$. Then

$$\begin{aligned}
\sum_{i=1}^{n} (Y_i-\theta)^{\mathrm{T}} \Sigma^{-1} (Y_i-\theta) &= \operatorname{tr}\left\{(\mathbf{Y}-\Theta)^{\mathrm{T}} (\mathbf{Y}-\Theta)\, \Sigma^{-1}\right\} \\
&= \operatorname{tr}\left\{(\mathbf{Y}-\Theta)^{\mathrm{T}}\, P\, P^{\mathrm{T}} (\mathbf{Y}-\Theta)\, \Sigma^{-1}\right\} \\
&= \operatorname{tr}\left\{(\mathbf{X}-\boldsymbol{\nu})^{\mathrm{T}} (\mathbf{X}-\boldsymbol{\nu})\, \Sigma^{-1}\right\} \\
&= \sum_{i=1}^{n} (X_i-\nu_i)^{\mathrm{T}} \Sigma^{-1} (X_i-\nu_i) \\
&= (X_1-\theta)^{\mathrm{T}} \Sigma^{-1} (X_1-\theta) + \sum_{i=2}^{n} X_i^{\mathrm{T}} \Sigma^{-1} X_i\,,
\end{aligned}$$

since $\boldsymbol{\nu}^{\mathrm{T}} = P\, \Theta^{\mathrm{T}} = (\theta : 0, \ldots : 0)^{\mathrm{T}}$ because the ith column of Θ is $\theta_i\, \mathbf{1}_n$ and since $P_1^{\mathrm{T}}\, \mathbf{1}_n = 1$ and $P_i^{\mathrm{T}}\, \mathbf{1}_n = 0$ for $i = 2, \ldots, n$, where P_i^{T} is the ith row of P.

Letting $X = X_1$ and $V_{i-1} = X_i$ for $i = 2, \ldots, n$ and noting that the Jacobian of the transformation $\mathbf{Y} \mapsto \mathbf{X} \mapsto (X_1, V_1, \ldots, V_{n-1})$ is 1, the density of $(X_1, V_1, \ldots, V_{n-1})$ is given by (6.53) (see also e.g. Rao 1973; Muirhead 1982 or Anderson 1984).

There is an obvious connection with the canonical form of the general linear model given in Sect. 4.5. Indeed, if $\Sigma = \sigma^2 I_p$, the density (6.53) becomes

$$\sigma^{-pn/2} f\left(\frac{(x-\theta)^{\mathrm{T}}(x-\theta) + \sum_{j=1}^{n-1} V_j^{\mathrm{T}} V_j}{\sigma^2}\right).$$

So, if $U_{ij} = V_{ij}$ and $U = (U_{12}, \ldots, U_{1p}, U_{21}, \ldots, U_{2p}, U_{n-11}, \ldots, U_{n-1p})$ then $(X, U) \sim SS_{p,(n-1)p}(\theta, 0)$. This model is also related to the general (normal) multivariate linear model $Y_{n\times m} = X_{n\times p}\, \beta_{p\times m} + \epsilon_{n\times m}$ where $\epsilon_{i\times m} \sim \mathcal{N}_m(0, \Sigma)$, $i = 1, \ldots, n$ are independent, X is a known design matrix and β is a matrix of unknown regression parameters.

We consider the problem of estimating θ with the invariant loss

$$L(\theta, \delta) = (\delta - \theta)^{\mathrm{T}} \Sigma^{-1} (\delta - \theta). \tag{6.54}$$

Recall that the usual estimator $\delta_0(X) = X$ is minimax provided $E_{0,I}[\|X\|^2] < \infty$ (where $E_{\theta,\Sigma}$ denotes the expectation with respect to the density in (6.53)). Note that, when Σ is a covariance matrix, this expectation is necessarily finite and equal to p. Moreover X is typically admissible when $p \leq 2$ and inadmissible when $p \geq 3$.

We concentrate on the case $p \geq 3$ and construct a class of estimators, depending on the sufficient statistics (X, S), of the form

$$\delta(X, S) = X + g(X, S), \tag{6.55}$$

where $S = \sum_{i=1}^{n-1} V_i V_i^{\mathrm{T}}$, which dominate $\delta_0(X) = X$ simultaneously under loss (6.54), for the entire class of distributions defined in (6.53) such that $E_{0,I}[\|X\|^2] < \infty$. Note that, although the loss in (6.54) is invariant, the estimate in (6.55) may not be equivariant (except for $\delta_0(X)$).

The risk difference $\Delta_{\theta,\Sigma}$ between $\delta(X, S)$ given in (6.55) and $\delta_0(X) = X$ equals

$$\begin{aligned} \Delta_{\theta,\Sigma} &= R\big(\theta, \delta(X, S)\big) - R\big(\theta, \delta_0(X)\big) \\ &= E_{\theta,\Sigma}\big[2 g^{\mathrm{T}}(X, S) \Sigma^{-1} (X - \theta)\big] + E_{\theta,\Sigma}\big[g^{\mathrm{T}}(X, S) \Sigma^{-1} g(X, S)\big], \end{aligned} \tag{6.56}$$

provided $E_{\theta,\Sigma}\big[g^{\mathrm{T}}(X, S)\Sigma^{-1} g(X, S)\big] < \infty$.

We first give a lemma which expresses the two terms in the last expression of (6.56) as expectations $E^*_{\theta,\Sigma}$ with respect to the distribution

$$C^{-1} F\left((x-\theta)^{\mathrm{T}} \Sigma^{-1} (x-\theta) + \sum_{j=1}^{n-1} V_j^{\mathrm{T}} \Sigma^{-1} V_j\right)$$

where F and C are defined as

$$F(t) = \frac{1}{2}\int_t^\infty f(s)ds$$

and

$$C = \int_{\mathbb{R}^p \times \cdots \times \mathbb{R}^p} F\left((x-\theta)^{\mathrm{T}}\Sigma^{-1}(x-\theta) + \sum_{j=1}^{n-1} V_j^{\mathrm{T}}\Sigma^{-1}V_j\right)dx\, dv_1 \cdots dv_{n-1}.$$

To this end, we will use the following notations. For any matrix M, ∇_M is interpreted as the matrix with components $(\nabla_M)_{ij} = \partial/\partial M_{ij}$. The differential operator for a symmetric matrix S is $\mathscr{D}_S = \left(\frac{1}{2}(1+\delta_{ij})(\nabla_S)_{ij}\right)$ and Haff differential operator is defined, for any $p \times p$ matrix function of a symmetric matrix S, say $H(S)$, to be

$$D^*_{1/2}(H(S)) = \mathrm{tr}\big(\mathscr{D}_S\, H(S)\big) = \sum_{i=1}^p \frac{\partial H_{ii}(S)}{\partial S_{ii}} + \frac{1}{2}\sum_{i\neq j}\frac{\partial H_{ij}(S)}{\partial S_{ij}}. \tag{6.57}$$

Lemma 6.6 *Let* $(X, V) = (X, V_1, \ldots, V_{n-1})$ *be a* $p \times n$ *random matrix with density* (6.53) *where* $p \leq n-1$ *and let* $S = V\,V^{\mathrm{T}}$.

(1) *Suppose* $g(x, s)$ *is a weakly differentiable function in* x *for each* s *such that the expectation* $E_{\theta,\Sigma}\big[g^{\mathrm{T}}(X, S)(X, S)\Sigma^{-1}(X-\theta)\big]$ *exists. Then*

$$E_{\theta,\Sigma}\big[g^{\mathrm{T}}(X, S)\Sigma^{-1}(X-\theta)\big] = C\,E^*_{\theta,\Sigma}\big[\mathrm{div}_X\, g(X, S)\big] \tag{6.58}$$

where $\mathrm{div}_x\, g(x, s)$ *is the divergence of* $g(x, s)$ *with respect to* x.

(2) *Suppose* $T(x, s)$ *is a* $p \times p$ *matrix function weakly differentiable in* v_i *(*$i = 1, \ldots, n-1$*) for any* x *and such that the expectation* $E_{\theta,\Sigma}\big[\mathrm{tr}\,\big(T(X, S)\big)\Sigma^{-1}\big]$ *exists. Then*

$$\begin{aligned} &E_{\theta,\Sigma}\big[\mathrm{tr}\,\big(T(X, S)\Sigma^{-1}\big)\big] \\ &\quad = C\,E^*_{\theta,\Sigma}\big[2\,D^*_{1/2}T(X, S) + (n-p-2)\,\mathrm{tr}\,(S^{-1}T(X, S))\big] \\ &\quad = C\,E^*_{\theta,\Sigma}\big[\mathrm{tr}(V\,\nabla_{V^{\mathrm{T}}}\{S^{-1}T(X, S)\}^{\mathrm{T}}) + (n-1)\,\mathrm{tr}\,(S^{-1}T(X, S))\big]. \end{aligned} \tag{6.59}$$

The proof of Lemma 6.6 is given at the end of this section. The two expressions in (6.58) follow from equality between the two integrand terms thanks to the link between the differential operators $D^*_{1/2}$ and $\mathrm{tr}(V\,\nabla_{V^{\mathrm{T}}})$ established in Proposition 6.5 (also given at the end of this section).

Note that, when $X, V_1, \ldots, V_{n-1}$ are independent normal vectors with covariance Σ, then $f = F$ and therefore $E_{\theta,\Sigma}[\quad] = E^*_{\theta,\Sigma}[\quad]$. Hence for Lemma 6.6, the identity in (6.58) essentially reduces to Stein's lemma (Stein 1981), and the identity in (6.59) corresponds to a result of Stein (1977a) and Haff (1979), known as the Stein-Haff identity.

Applying (6.58) to the first term in (6.56) and (6.59) to the second term in (6.56) with $T(x, s) = g(x, s)g'(x, s)$, noting that

$$g^{\mathrm{T}}(x, s)\Sigma^{-1}g(x, s) = \mathrm{tr}\left(g(x, s)g^{\mathrm{T}}(x, s)\Sigma^{-1}\right)$$

gives immediately the following theorem.

Theorem 6.5 *Assume that* $g(x, s)$ *and* $T(x, s) = g(x, s)g^{\mathrm{T}}(x, s)$ *satisfy the assumptions of Lemma 6.6. Assume also that* $E_{0,\Sigma}[\|X\|^2] < \infty$ *and* $E_{\theta,\Sigma}\left[g^{\mathrm{T}}(X, S)\Sigma^{-1}g(X, S)\right] < \infty$. *Then the risk difference* $\Delta_{\theta,\Sigma}$ *in* (6.56) *between* $\delta(X, S) = X + g(X, S)$ *and* $\delta_0(X) = X$ *equals*

$$C\, E^*_{\theta,\Sigma}\Big[2\,\mathrm{div}_X g(X, S) + 2\, D^*_{1/2}\left(g(X, S)g^{\mathrm{T}}(X, S)\right) + (n-p-2)\, g^{\mathrm{T}}(X, S)\, S^{-1}g(X, S)\Big]. \tag{6.60}$$

A sufficient condition for $\delta(X, S)$ *to be minimax is that, for all* x *and* s,

$$2\,\mathrm{div}_x g(x, s) + 2\, D^*_{1/2}\big(g(x, s)g^{\mathrm{T}}(x, s)\big) + (n-p-2)\, g^{\mathrm{T}}(x, s)s^{-1}g(x, s) \le 0 \tag{6.61}$$

or, equivalently,

$$2\,\mathrm{div}_x g(x, s) + \mathrm{tr}(v\, \nabla_{v^{\mathrm{T}}}\, \{s^{-1}g(x, s)g^{\mathrm{T}}(x, s)\}^{\mathrm{T}} + (n-1)\, g^{\mathrm{T}}(x, s)s^{-1}\, g(x, s) \le 0\,, \tag{6.62}$$

where $V = (V_1, \ldots, V_{n-1})$ *is a* $p \times (n-1)$ *matrix and* $S = V\,V^{\mathrm{T}}$. *Furthermore* $\delta(X, S)$ *dominates* $\delta_0(X)$ *as soon as* (6.61) *or* (6.62) *is satisfied with strict inequality on a set of positive measure.*

Note that in the normal case $E^*_{\theta,\Sigma}[\quad] = E_{\theta,\Sigma}[\quad]$ so that the left-hand side of (6.61) is an unbiased estimator of the risk difference between $\delta(X, S)$ and $\delta_0(X)$. Perhaps, most importantly, observe that the theorem leads to an extremely strong robustness property for estimators satisfying (6.61). Namely, any such estimator is minimax and, as soon as strict inequality occurs on a set of positive measure in (6.61), dominates $\delta_0(X)$ for the entire class of elliptically symmetric distributions (6.53). This property is analogous to the robustness property mentioned in Sect. 6.1 in the case of spherically symmetric distributions. The following corollary gives a general class of examples of minimax estimates which dominate $\delta_0(X)$ uniformly for densities of the form (6.53).

Corollary 6.3 *Assume that* $E_{0,\Sigma}[\|X\|^2] < \infty$ *and* $E_{\theta,\Sigma}\left[\frac{\|X\|^2}{(X^{\mathrm{T}}S^{-1}X)^2}\right] < \infty$. *Let* $\delta(X, S) = (1 - r(X^{\mathrm{T}}S^{-1}X)/X^{\mathrm{T}}S^{-1}X)\, X$ *where* $r(\cdot)$ *is a nondecreasing function bounded between 0 and* $2(p-2)/(n-p+2)$. *Then* $\delta(X, S)$ *is minimax for any density of the form* (6.53). *Furthermore* $\delta(X, S)$ *dominates* $\delta_0(X)$ *as soon as either* r *is strictly increasing or bounded away for 0 and* $\frac{2(p-2)}{n-p+2}$ *on a set of positive measure.*

Proof Setting

$$g(x,s) = -\frac{r(x^{\mathrm{T}}s^{-1}x)}{x^{\mathrm{T}}s^{-1}x}\,x,$$

we have

$$\operatorname{div}_x g(x,s) = -\left[(p-2)\frac{r(x^{\mathrm{T}}s^{-1}x)}{xs^{-1}x} + 2r'(x^{\mathrm{T}}s^{-1}x)\right]$$

by routine calculations. Now we have

$$\begin{aligned}
&D^*_{1/2}\big(g(x,s)g^{\mathrm{T}}(x,s)\big)\\
&= \sum_{i=1}^{p}\frac{\partial}{\partial s_{ii}}\left[\frac{r^2(x^{\mathrm{T}}s^{-1}x)}{(x^{\mathrm{T}}s^{-1}x)^2}\right]x_i^2 + \frac{1}{2}\sum_{i\neq j}\frac{\partial}{\partial s_{ij}}\left[\frac{r^2(x^{\mathrm{T}}s^{-1}x)}{(x^{\mathrm{T}}s^{-1}x)^2}\right]x_i x_j\\
&= \frac{2(x^{\mathrm{T}}s^{-1}x)^2 r(x^{\mathrm{T}}s^{-1}x)r'(x;s^{-1}x) - 2(x^{\mathrm{T}}s^{-1}x)r^2(x^{\mathrm{T}}s^{-1}x)}{(x^{\mathrm{T}}s^{-1}x)^4}\\
&\qquad\times\left\{\sum_{i=1}^{p}\frac{\partial}{\partial s_{ii}}(x^{\mathrm{T}}s^{-1}x)X_i^2 + \frac{1}{2}\sum_{i\neq j}\frac{\partial}{\partial s_{ij}}(x^{\mathrm{T}}s^{-1}x)x_i x_j\right\}. \qquad (6.63)
\end{aligned}$$

Using the fact that

$$\frac{\partial}{\partial s_{ij}}(x^{\mathrm{T}}s^{-1}x) = -(2-\delta_{ij})(x^{\mathrm{T}}s^{-1})_i(x^{\mathrm{T}}s^{-1})_j$$

it follows that the bracketed term in (6.63) equals

$$\begin{aligned}
&-\left\{\sum_{i=1}^{p}(x^{\mathrm{T}}s^{-1})_i^2 x_i^2 + \frac{1}{2}\sum_{i\neq j}2(x^{\mathrm{T}}s^{-1})_i(x^{\mathrm{T}}s^{-1})_j x_i x_j\right\}\\
&= -\sum_{1\le i,j\le p}(x^{\mathrm{T}}s^{-1})_i(x^{\mathrm{T}}s^{-1})_j x_j\\
&= -\left(\sum_{i=1}^{p}(x^{\mathrm{T}}s^{-1})_i X_i\right)^2\\
&= -(x^{\mathrm{T}}s^{-1}x)^2
\end{aligned}$$

and hence

$$D^*_{1/2}\big(g(x,s)g^{\mathrm{T}}(x,s)\big) = -2\left\{r(x^{\mathrm{T}}s^{-1}x)r'(x^{\mathrm{T}}sx) - \frac{r^2(x^{\mathrm{T}}s^{-1}x)}{x^{\mathrm{T}}s^{-1}x}\right\}.$$

Finally it is clear that

$$g^{\mathrm{T}}(x,s)s^{-1}g(x,s) = \frac{r^2(x^{\mathrm{T}}s^{-1}x)}{x^{\mathrm{T}}s^{-1}x}$$

so that the left-hand side of (6.61) equals

$$\begin{aligned}
&-2\left\{(p-2)\frac{r(x^{\mathrm{T}}s^{-1}x)}{x^{\mathrm{T}}s^{-1}x} + 2r'(x^{\mathrm{T}}s^{-1}x)\right\} + (n-p-2)\frac{r^2(x^{\mathrm{T}}s^{-1}x)}{x^{\mathrm{T}}s^{-1}x} \\
&-4\left\{r(x^{\mathrm{T}}s^{-1}x)r'(x^{\mathrm{T}}s^{-1}x) - \frac{r^2(x^{\mathrm{T}}s^{-1}x)}{x^{\mathrm{T}}s^{-1}x}\right\} \\
&= \frac{r(x^{\mathrm{T}}s^{-1}x)}{x^{\mathrm{T}}s^{-1}x}\left\{-2(p-2) + (n-p+2)r(x^{\mathrm{T}}s^{-1}x)\right\} \\
&\quad -4r'(x^{\mathrm{T}}s^{-1}x)\left\{1 + r(x^{\mathrm{T}}s^{-1}x)\right\} \\
&\le 0,
\end{aligned} \tag{6.64}$$

according to the assumptions on $r(\cdot)$.

Hence the minimaxity of $\delta(X, S)$ follows. The domination result follows as well since strict inequality in (6.64) holds on a set of positive measure under the additional assumptions. □

***Proof of Lemma* 6.6** (Part 1) By definition, we have

$$E_\theta\left[g(X,S)^{\mathrm{T}}\Sigma^{-1}(X-\theta)\right] = \int_{\mathbb{R}^p\times\cdots\times\mathbb{R}^p}\int_{\mathbb{R}^p} g(x,s)^{\mathrm{T}}\Sigma^{-1}(x-\theta)$$
$$f\left((x-\theta)^{\mathrm{T}}\Sigma^{-1}(x-\theta) + \sum_{j=1}^{n-1} v_j^{\mathrm{T}}\Sigma^{-1}v_j\right) dx\, dv_1 \ldots dv_{n-1}\,.$$

Now applying the integration-by-slice in Lemma A.2 in Appendix A.5 with $\varphi(x) = \sqrt{(x-\theta)^{\mathrm{T}}\Sigma^{-1}(x-\theta)}$ to the inner most integral

$$\begin{aligned}
&I(v_1,\ldots,v_{n-1}) \\
&= \int_{\mathbb{R}^p} g(x,s)^{\mathrm{T}}\Sigma^{-1}(x-\theta)\, f\left((x-\theta)^{\mathrm{T}}\Sigma^{-1}(x-\theta) + \sum_{j=1}^{n-1} v_j^{\mathrm{T}}\Sigma^{-1}v_j\right) dx
\end{aligned}$$

gives

$$\nabla\varphi(x) = \frac{\Sigma^{-1}(x-\theta)}{\sqrt{(x-\theta)^{\mathrm{T}}\Sigma^{-1}(x-\theta)}}$$

and

$$
\begin{aligned}
I(v_1, \ldots, v_{n-1}) \\
&= \int_0^{\infty} f\left(R^2 + \sum_{j=1}^{n-1} v_j^{\mathrm{T}} \Sigma^{-1} v_j\right) \int_{[\varphi=R]} \frac{g(x,s)^{\mathrm{T}} \Sigma^{-1}(x-\theta)}{\|\nabla\varphi(x)\|} d\sigma_R(x)\, dR \\
&= \int_0^{\infty} f\left(R^2 + \sum_{j=1}^{n-1} v_j^{\mathrm{T}} \Sigma^{-1} v_j\right) \int_{[\varphi=R]} g(x,s)^{\mathrm{T}} \sqrt{(x-\theta)^{\mathrm{T}} \Sigma^{-1}(x-\theta)} \\
&\quad \times \frac{\nabla\varphi(x)}{\|\nabla\varphi(x)\|}\, d\sigma_R(x)\, dR\,,
\end{aligned}
$$

according to the expression of $\nabla\varphi(x)$. Then, as $\sqrt{(x-\theta)^{\mathrm{T}} \Sigma^{-1}(x-\theta)} = R$ on $[\varphi = R]$, it follows using Stokes' theorem that

$$
\begin{aligned}
& I(v_1, \ldots, v_{n-1}) = \\
& \int_0^{\infty} R\, f\left(R^2 + \sum_{j=1}^{n-1} v_j^{\mathrm{T}} \Sigma^{-1} v_j\right) \int_{[\varphi=R]} g(x,s)\, \frac{\nabla\varphi(x)}{\|\nabla\varphi(x)\|}\, d\sigma_R(x)\, dR = \\
& \int_0^{\infty} R\, f\left(R^2 + \sum_{j=1}^{n-1} v_j^{\mathrm{T}} \Sigma^{-1} v_j\right) \int_{[\varphi \leq R]} \operatorname{div}_x g(x,s)\, dx\, dR\,.
\end{aligned}
$$

Now, using Fubini's theorem gives

$$
\begin{aligned}
& I(v_1, \ldots, v_{n-1}) = \\
& \int_{\mathbb{R}^p} \operatorname{div}_x g(x,s) \int_{\sqrt{(x-\theta)^{\mathrm{T}} \Sigma^{-1}(x-\theta)}}^{\infty} R\, f\left(R^2 + \sum_{j=1}^{n-1} v_j^{\mathrm{T}} \Sigma^{-1} v_j\right) dR\, dx = \\
& \int_{\mathbb{R}^p} \operatorname{div}_x g(x,s)\, \frac{1}{2} \int_{(x-\theta)^{\mathrm{T}} \Sigma^{-1}(x-\theta)}^{\infty} f\left(r + \sum_{j=1}^{n-1} v_j^{\mathrm{T}} \Sigma^{-1} v_j\right) dr\, dx = \\
& \int_{\mathbb{R}^p} \operatorname{div}_x g(x,s) F\left((x-\theta)^{\mathrm{T}} \Sigma^{-1}(x-\theta) + \sum_{j=1}^{n-1} v_j^{\mathrm{T}} \Sigma^{-1} v_j\right) dx\,, \qquad (6.65)
\end{aligned}
$$

through the change of variable $r = R^2$ and by definition of the function F.

Finally integrating (6.65) with respect to the v_j gives an expression for the expectation $E_\theta[g(X,S)^{\mathrm{T}}\Sigma^{-1}(X-\theta)]$ and yields (6.58).

(Part 2) First note that, setting $G = S^{-1}T(X,S)$, we have

$$\mathrm{tr}\left(T(X,S)\Sigma^{-1}\right) = \mathrm{tr}\left(\Sigma^{-1}S\,G(X,S)\right).$$

Then, as $V = (V_1, \ldots, V_{n-1})$ and $S = V V^{\mathrm{T}}$, we have

$$\begin{aligned}
\mathrm{tr}\Big(\Sigma^{-1}S\,G(X,S)\Big) &= \mathrm{tr}\Big(G(X,S)\,\Sigma^{-1}S\Big) \\
&= \mathrm{tr}\left(G(X,S)\,\Sigma^{-1}\sum_{i=1}^{n-1} V_i\,V_i^{\mathrm{T}}\right) \\
&= \sum_{i=1}^{n-1} \mathrm{tr}\Big(V_i^{\mathrm{T}}\,G(X,S)\,\Sigma^{-1}\,V_i\Big) \\
&= \sum_{i=1}^{n-1} V_i^{\mathrm{T}}\,G(X,S)\,\Sigma^{-1}\,V_i. \qquad (6.66)
\end{aligned}$$

Now, according to Part 1 of Lemma 6.6 where the roles of X and θ are played by V_i and 0 respectively, it follows from (6.66) that

$$\begin{aligned}
E_{\theta,\Sigma}\left[\mathrm{tr}\Big(\Sigma^{-1}S\,G(X,S)\Big)\right] &= C\sum_{i=1}^{n-1} E^*_{\theta,\Sigma}\left[\mathrm{div}_{V_i}(G^{\mathrm{T}}(X,S)\,V_i)\right] \\
&= C\,E^*_{\theta,\Sigma}[A_1 + A_2], \qquad (6.67)
\end{aligned}$$

where

$$\begin{aligned}
A_1 &= \sum_{i=1}^{n-1}\sum_{j=1}^{p}\sum_{m=1}^{p} \frac{\partial V_{mi}}{\partial V_{ji}}\,G^{\mathrm{T}}_{jm} \\
&= \sum_{i=1}^{n-1}\sum_{j=1}^{p}\sum_{m=1}^{p} \delta_{jm}\,G^{\mathrm{T}}_{jm}(X,S) \\
&= (n-1)\sum_{j=1}^{p} G^{\mathrm{T}}_{jj}(X,S) \\
&= (n-1)\,\mathrm{tr}(G(X,S)) \qquad (6.68)
\end{aligned}$$

and

$$\begin{aligned} A_2 &= \sum_{i=1}^{n-1} \sum_{j=1}^{p} \sum_{m=1}^{p} V_{mi} \, \frac{\partial G^{\mathrm{T}}_{jm}(X, S)}{\partial V_{ji}} \\ &= \sum_{i=1}^{n-1} \sum_{m=1}^{p} V_{mi} \sum_{j=1}^{p} \frac{\partial G^{\mathrm{T}}_{jm}(X, S)}{\partial V_{ji}} \\ &= \sum_{i=1}^{n-1} \sum_{m=1}^{p} V_{mi} (\nabla_{V^{\mathrm{T}}} G^{\mathrm{T}}(X, S))_{im} \\ &= \sum_{m=1}^{p} (V \nabla_{V^{\mathrm{T}}} G^{\mathrm{T}}(X, S))_{mm} \\ &= \mathrm{tr}(V \, \nabla_{V^{\mathrm{T}}} \, G^{\mathrm{T}}(X, S)) \, . \end{aligned} \tag{6.69}$$

Finally, combining (6.67), (6.68) and (6.69), we obtain the second formula in (6.59).

As for the first formula in (6.59), it follows directly from the link between the differential expressions $D^*_{1/2} S \, G(X, S)$ and $\mathrm{tr}\big(V \, \nabla_{V^{\mathrm{T}}} \, G^{\mathrm{T}}(X, S)\big)$ given in Proposition 6.5 below, whose proof is given in Appendix A.7. □

Proposition 6.5 (Fourdrinier et al. 2016) *For any $p \times p$ matrix function $G(x, s)$ weakly differentiable with respect to s for any x,*

$$2 \, D^*_{1/2}\big(S \, G(X, S)\big) = (p + 1) \, \mathrm{tr}(G(X, S)) + \mathrm{tr}\big(V \, \nabla_{V^{\mathrm{T}}} \, G^{\mathrm{T}}(X, S)\big) \, . \tag{6.70}$$

6.5 Shrinkage Estimators for Concave Loss in the Presence of a Residual Vector

In this section, we consider the case of concave loss and illustrate that certain classes of shrinkage estimators which properly use the residual vector have the strong robustness property of dominating the usual unbiased estimator uniformly over the class of spherically symmetric distributions, simultaneously for a broad class of concave loss functions. It extends and broadens the results of Sect. 5.5 to the residual vector case. We follow closely the development in Brandwein and Strawderman (1991a).

Specifically, let (X, U) be a $p + k$ dimensional vector with mean vector $(\theta, 0)$, where the dimensions of X and θ are equal to p and the dimensions of the residual vector U and its mean vector, 0, are equal to k, that is, $(X, U) \sim SS_{p+k}(\theta, 0)$. The loss function we consider is

$$L(\theta, \delta) = \ell(||\theta - \delta||^2), \tag{6.71}$$

for $\ell(t)$ a nonnegative concave monotone nondecreasing function.

The estimators we consider will be of the now familiar form

$$\delta(X, \|U\|^2) = X + a(S/(k+2))g(X), \tag{6.72}$$

where $S = ||U||^2$, and $g(\cdot)$ maps $\mathbb{R}^p$ into $\mathbb{R}^p$.

The following result, extracted from the development in Theorem 5.5 due to Brandwein and Strawderman (1991a) is basic to the development of this section.

Lemma 6.7 (Brandwein and Strawderman 1991a) *Let $X \sim SS_p(\theta)$, for $p \geq 4$ and let $g(X)$ map R^p into $\mathbb{R}^p$ be weakly differentiable, and such that*

(1) $||g(X)||^2/2 \leq -h(X) \leq -\nabla^{\mathrm{T}} g(X)$,
(2) *$-h(X)$ is superharmonic and $E_\theta[R^2 h(W)|R]$ is a nondecreasing function of R, where W has a uniform distribution on the sphere of radius R centered at θ.*

Then $E_\theta[||X + ag(X) - \theta||^2 - ||X - \theta||] \leq E[(-2a^2/r^2 + 2a/p)E_\theta[r^2 h(W)|r^2]]$, where $r^2 = ||X - \theta||^2$.

We will also need the following well known result (see e.g. the discussion at the end of Sect. 1.2).

Lemma 6.8 *Suppose $(X, U) \sim SS_{p+k}(\theta, 0)$. Then the random variable $\beta = ||X - \theta||^2/(||X - \theta||^2 + S)$ has a $Beta(p/2, k/2)$ distribution, independent of $R^2 = ||X - \theta||^2 + S$, where $S = ||U||^2$.*

The main result is the following.

Theorem 6.6 *Suppose $(X, U) \sim SS_{p+k}(\theta, 0)$, that loss is given by loss (6.71) and that the estimator $\delta(X, S)$ is given by (6.72). Then $\delta(X, S)$ dominates the unbiased estimator X, provided that*

(1) *$g(X)$ satisfies assumptions (1) and (2) of Lemma 6.7,*
(2) *the concave nondecreasing function $\ell(t)$ also satisfies $t^\alpha \ell'(t)$ is nondecreasing,*
(3) $0 < a \leq (p - 2 - 2\alpha)/p$.

Note first, by concavity of $\ell(\cdot)$, that $\ell(t) \leq \ell(y) + (t - y)\ell'(y)$. Hence the risk satisfies

$$\begin{aligned}
R(\theta, \delta) &= E[\ell(||X + \frac{aSg(X)}{k+2} - \theta||^2)] \\
&\leq E[\ell(||X - \theta||^2) + \ell'(||X - \theta||^2)(||X + \frac{aSg(X)}{k+2} - \theta||^2 - ||X - \theta||^2)] \\
&= R(\theta, X) + E[\ell'(||X - \theta||^2)(||X + \frac{aSg(X)}{k+2} - \theta||^2 - ||X - \theta||^2)].
\end{aligned}$$

It suffices to prove the second term in the above expression is negative. Now, let $r^2 = ||X - \theta||^2$, $R^2 = ||X - \theta||^2 + S$ (where $S = ||U||^2 = R^2 - r^2$), and note that the conditional distribution of X given r and R is $SS_p(\theta)$. Then it follows, using

Lemma 6.7 that

$$\begin{aligned}
&E[\ell'(||X-\theta||^2)(||X+\frac{aSg(X)}{k+2}-\theta||^2-||X-\theta||^2)] \\
&= E[\ell'(r^2)E[||X+\frac{aSg(X)}{k+2}-\theta||^2-||X-\theta||^2|R,r]] \\
&\leq E[\ell'(r^2)E[(2(\frac{aS}{(k+2)r})^2-2\frac{aS}{(k+2)p})E_\theta[-r^2h(W)|r^2]|R,r]].
\end{aligned}$$

Now using Lemma 6.8, this last expression may be written as

$$\begin{aligned}
&2\,E\left[\ell'(\beta R^2)\left(\left\{\frac{a(1-\beta)R^2}{k+2}\right\}^2\frac{1}{\beta R^2}-\frac{a(1-\beta)R^2}{(k+2)p}\right)\right. \\
&\qquad\qquad\left.\times\, E_\theta[-\beta R^2h(W)|\beta R^2]|R]\right] \\
&=\frac{2a}{k+2}E\left[(R^2(\beta R^2)^\alpha\ell'(\beta R^2)(\beta R^2)^{-\alpha}(1-\beta)\left(\frac{(1-\beta)a}{\beta(k+2)}-\frac{1}{p}\right)\right. \\
&\qquad\qquad\left.\times\, E_\theta[-\beta R^2h(W)|\,\beta R^2]|\;R]\right].
\end{aligned}$$

Next, for fixed R, by assumption (2) of Lemma 6.7 $E_\theta[-\beta R^2h(W)\mid\beta R^2]$ is nonnegative and nondecreasing in β and by assumption (6.71) so is $\beta^\alpha\ell'(\beta R^2)$. Also $(1-\beta)/\beta$ is decreasing in β. Hence it follows from the covariance inequality (and independence of β and R) that the previous expression is less than or equal to

$$\begin{aligned}
&\frac{2a}{k+2}E\left[[E_\theta[-\beta R^2h(W)|\beta R^2]R^2(R^2\beta)^\alpha\ell'(\beta R^2)|R]E[\beta^\alpha(1-\beta)]\right. \\
&\qquad\qquad\left.\times\left(\frac{a(1-\beta)}{\beta(k+2)}-\frac{1}{p}\right)\right].
\end{aligned}$$

Since the first expectation in this term is nonnegative, it suffices that the second expectation is negative. But this is equivalent to

$$0\leq a\leq\frac{k+2}{p}E[\beta^\alpha(1-\beta)]/E[(\beta^\alpha(1-\beta)^2)/\beta]=(p-2-2\alpha)/p,$$

which completes the proof. □

For the loss $L(\theta,\delta)=||\theta-\delta||^q$, $\ell(t)=t^{q/2}$, it follows that $t^\alpha\ell'(t)=(q/2)t^{\alpha+q/2-1}$ is nondecreasing for $\alpha\geq 1-q/2$. Thus, the following corollary is immediate.

Corollary 6.4 *Under the loss* $L(\theta, \delta) = ||\theta - \delta||^q$, *for* $p > 4$ *and* $0 < q \leq 2$, *the estimator in Theorem 6.6 dominates* X *for* $0 < a \leq (p-4+2q)/p$ *simultaneously for all spherically symmetric distributions with finite second moment. It does so simultaneously for all such losses for* $0 < a \leq (p-4)/p$.

Note that the range of shrinkage constants for which domination holds includes $a = 1/2$ as soon as $p \geq 8$. For the usual James-Stein estimator,

$$\delta(X) = (1 - a(2(p-2)S)/((k+2)||X||^2))X, \tag{6.73}$$

the uniformly optimal constant for quadratic loss ($\ell(\cdot) = 1$) is $a = 1/2$ and hence this optimal estimator improves for all such l_q losses simultaneously for $p \geq 8$.