The James-Stein Estimator of the mean $\underset{\sim}{\mu}$ of $N_p(\underset{\sim}{\mu}, \Sigma)$.

Observe $\underset{\sim}{Y}_1, \dots, \underset{\sim}{Y}_n$ i.i.d. $\sim N_p(\underset{\sim}{\mu}, \Sigma)$. Let $\hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{Y}} \sim N_p(\underset{\sim}{\mu}, \Sigma)$. Then:

(i) $\hat{\underset{\sim}{\mu}}$ is the MLE of $\underset{\sim}{\mu}$.

(ii) $\hat{\underset{\sim}{\mu}}$ is the minimum variance unbiased estimator (M.V.U.E.) of $\hat{\underset{\sim}{\mu}}$:
$\hat{\underset{\sim}{\mu}}$ minimizes the mean square error (MSE) $E_{\underset{\sim}{\mu}} \| \underset{\sim}{\mu}^* - \underset{\sim}{\mu} \|^2$
among all unbiased estimators $\underset{\sim}{\mu}^*$ of $\underset{\sim}{\mu}$. ($\bar{Y}$ is complete & sufficient. Apply Lehmann-Scheffe)

(iii) $\hat{\underset{\sim}{\mu}}$ is the best (MSE) translation-equivariant estimator of $\underset{\sim}{\mu}$:
$\underset{\sim}{\mu}^*(\bar{\underset{\sim}{Y}} + \underset{\sim}{a}) = \underset{\sim}{\mu}^*(\bar{\underset{\sim}{Y}}) + \underset{\sim}{a} \quad \forall \underset{\sim}{a} \in \mathbb{R}^p$. ["equivariant"]

(iv) For $p = 1, 2$, $\hat{\underset{\sim}{\mu}}$ is an admissible estimator of $\underset{\sim}{\mu}$: no other estimator (linear or not) has everywhere better MSE than $\hat{\underset{\sim}{\mu}}$.

However: Stein (1956) showed that for $p \geq 3$, $\hat{\underset{\sim}{\mu}}$ is inadmissible. James & Stein (1962) produced a (biased, non-linear) est. $\hat{\underset{\sim}{\mu}}_{JS}$ whose MSE is everywhere less than the MSE of $\hat{\underset{\sim}{\mu}}$, and substantially less for $\underset{\sim}{\mu}$ near $\underset{\sim}{0}$. [Question: but is $\underset{\sim}{0}$ special?] (The difference in MSE's $\to 0$ as $\|\underset{\sim}{\mu}\| \to \infty$.)

We shall consider only the special case where Σ is known (for simplicity) but similar results hold if Σ is unknown. Thus we can assume that $n = 1$:

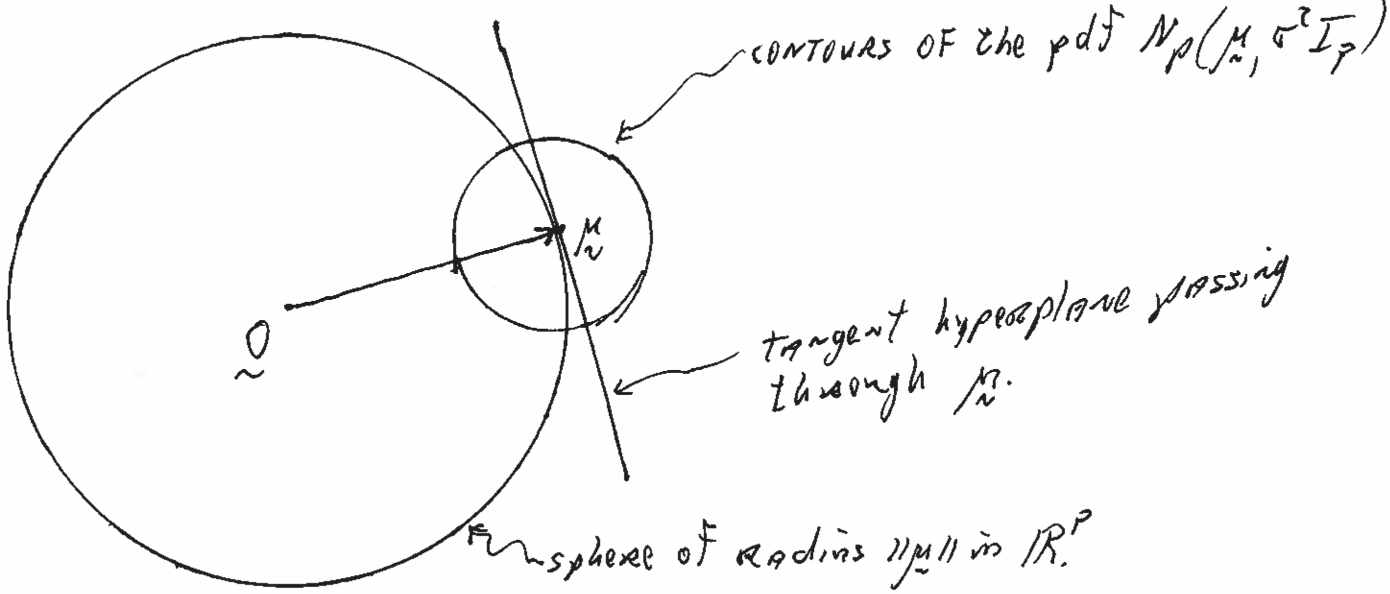
Observe (*) $\underset{\sim}{Y} \sim N_p(\underset{\sim}{\mu}, \sigma^2 I_p)$, σ^2 known. Here $\hat{\underset{\sim}{\mu}} = \underset{\sim}{Y}$.
The MSE of $\hat{\underset{\sim}{\mu}}$ is

$$(1) \qquad E_{\underset{\sim}{\mu}} \| \hat{\underset{\sim}{\mu}} - \underset{\sim}{\mu} \|^2 = E_{\underset{\sim}{\mu}} \| \underset{\sim}{Y} - \underset{\sim}{\mu} \|^2 = E(\sigma^2 \chi^2_p) = \sigma^2 p.$$

The JS estimator has the form $\hat{\underset{\sim}{\mu}}_{JS} = a(\|\underset{\sim}{Y}\|)\underset{\sim}{Y}$, where $a(\|Y\|)$ is a scalar-valued function of $\|\underset{\sim}{Y}\|$. Note that $\hat{\underset{\sim}{\mu}}_{JS}$ is not translation-equivariant, but it is orthogonally equivariant (as is $\hat{\underset{\sim}{\mu}}$):

(2) $\quad \hat{\mu}_{JS}(\Gamma \underset{\sim}{Y}) = \Gamma \hat{\mu}_{JS}(Y) \quad \forall\ \Gamma$: $p \times p$ orthogonal. ["equivariant"]

The multiplier $a(\|Y\|)$ has the form $1 - \frac{d}{\|Y\|^2}$, hence "shrinks" $\underset{\sim}{Y}$ toward $\underset{\sim}{0}$ (Sometimes, it shrinks $\underset{\sim}{Y}$ beyond $\underset{\sim}{0}$, so $\max\{a(\|Y\|), 0\}$ is preferable, but its MSE is hard to calculate analytically.) Here is a simplified version of Stein's heuristic explanation of why shrinking $\hat{\underset{\sim}{\mu}} \equiv \underset{\sim}{Y}$ toward $\underset{\sim}{0}$ might reduce the MSE:



By symmetry, the tangent hyperplane divides $\mathbb{R}^p$ into 2 halfspaces each having probability $\frac{1}{2}$ under $N_p(\mu, \sigma^2 I_p)$.

$\therefore P_{\underset{\sim}{\mu}}[\underset{\sim}{Y} \in$ sphere of radius $\|\underset{\sim}{\mu}\|] < \frac{1}{2}$,

so $P_{\underset{\sim}{\mu}}[\|\underset{\sim}{Y}\| > \|\underset{\sim}{\mu}\|] > \frac{1}{2}$. This implies that $\|\underset{\sim}{Y}\|$ is "too large",

so to estimate $\underset{\sim}{\mu}$ we should "shrink" $\underset{\sim}{Y}$ toward $\underset{\sim}{0}$. [But $\underset{\sim}{0}$ is arbitrary!]

More precisely:

$\|\underset{\sim}{Y}\|^2 \sim \sigma^2 \underbrace{\chi_p^2(\|\underset{\sim}{\mu}\|^2/\sigma^2)}_{}$, so $E\|\underset{\sim}{Y}\|^2 = \sigma^2[p + \frac{\|\underset{\sim}{\mu}\|^2}{\sigma^2}] = p\sigma^2 + \|\underset{\sim}{\mu}\|^2 > \|\underset{\sim}{\mu}\|^2$.

(noncentral χ^2)

Thus $\|Y\|^2$ is too large an estimate of $\|\mu\|^2$ – in fact, $\|\underset{\sim}{Y}\|^2 - p\sigma^2$ is an unbiased estimator of $\|\underset{\sim}{\mu}\|^2$.

The JSE can be motivated by an "empirical Bayesian" argument. Suppose we make the Bayesian assumption that $\underset{\sim}{\mu}$ is also random, with prior distribution

(3) $$\underset{\sim}{\mu} \sim N_p(\underset{\sim}{0}, \lambda^2 I_p).$$ [We could replace $\underset{\sim}{0}$ by any $\underset{\sim}{\mu_0}$!]

Then we should rewrite (*) as $\underset{\sim}{Y} \mid \underset{\sim}{\mu} \sim N(\underset{\sim}{\mu}, \sigma^2 I_p)$. From these two facts, (the marginal dist'n of $\underset{\sim}{\mu}$ and the conditional dist'n of $\underset{\sim}{Y} \mid \underset{\sim}{\mu}$), we can deduce that the joint distribution of $(\underset{\sim}{Y}, \underset{\sim}{\mu})$ is the following:

(4) $$\begin{pmatrix} \underset{\sim}{Y} \\ \underset{\sim}{\mu} \end{pmatrix} \sim N_{2p}\left[\begin{pmatrix} \underset{\sim}{0} \\ \underset{\sim}{0} \end{pmatrix}, \begin{bmatrix} (\sigma^2+\lambda^2) I_p & \lambda^2 I_p \\ \lambda^2 I_p & \lambda^2 I_p \end{bmatrix}\right].$$ [Exercise]

Therefore the posterior distribution of $\underset{\sim}{\mu} \mid \underset{\sim}{Y}$ (i.e., the conditional dist'n) is

(5) $$\underset{\sim}{\mu} \mid \underset{\sim}{Y} \sim N_p\left[\left(\frac{\lambda^2}{\sigma^2+\lambda^2}\right)\underset{\sim}{Y}, \left(\frac{\sigma^2\lambda^2}{\sigma^2+\lambda^2}\right) I_p\right]$$ [Exercise]

The usual Bayes estimator (w.r.t. $\underbrace{\text{quadratic loss}}_{\equiv MSE}$) is the posterior mean

(6) $$\hat{\underset{\sim}{\mu}}_\lambda \equiv E[\underset{\sim}{\mu} \mid \underset{\sim}{Y}] = \left(\frac{\lambda^2}{\sigma^2+\lambda^2}\right)\underset{\sim}{Y}.$$

This is a linear estimator of the form $a\underset{\sim}{Y}$ with $a = \frac{\lambda^2}{\sigma^2+\lambda^2} < 1$, so $E\|\hat{\underset{\sim}{\mu}}_\lambda - \underset{\sim}{\mu}\|^2 = E\|a\underset{\sim}{Y} - \underset{\sim}{\mu}\|^2 = a^2 E\|\underset{\sim}{Y} - \underset{\sim}{\mu}\|^2 + (a-1)^2\|\underset{\sim}{\mu}\|^2 \to \infty$ as $\|\underset{\sim}{\mu}\| \to \infty$, hence $\hat{\underset{\sim}{\mu}}_\lambda$ cannot dominate the MLE $\hat{\underset{\sim}{\mu}} \equiv \underset{\sim}{Y}$. [The MSE of $\hat{\underset{\sim}{\mu}}_\lambda$ is less than that of $\hat{\underset{\sim}{\mu}}$ for $\underset{\sim}{\mu}$ near $\underset{\sim}{0}$, i.e., for $\underset{\sim}{\mu}$ near the mean of the assumed prior dist'n (3).]

If the prior variance λ^2 is <u>unknown</u>, however, then we cannot use $\hat{\underset{\sim}{\mu}}_\lambda$. This raises the question: can we <u>estimate</u> λ, based on $\underset{\sim}{Y}$? If so, then we can replace $\hat{\underset{\sim}{\mu}}_\lambda$ by $\hat{\underset{\sim}{\mu}}_{\hat{\lambda}}$, where $\hat{\lambda}$ is the estimated value. The estimator $\hat{\underset{\sim}{\mu}}_{\hat{\lambda}}$ is called an "<u>empirical</u> <u>Bayes</u> <u>estimator</u>"

because the prior distribution is estimated from the data.

To obtain $\hat{\lambda}^2$, note from (4) that the marginal dist'n of $\underset{\sim}{Y}$ is

(7) $$\underset{\sim}{Y} \sim N_p\left[\underset{\sim}{0}, (\sigma^2+\lambda^2) I_p\right],$$

so

(8) $$\|\underset{\sim}{Y}\|^2 \sim (\sigma^2+\lambda^2)\chi_p^2.$$

Now, we can rewrite $\hat{\underset{\sim}{M}}_\lambda$ as

(9) $$\hat{\underset{\sim}{M}}_\lambda = \left(1 - \frac{\sigma^2}{\sigma^2+\lambda^2}\right)\underset{\sim}{Y},$$

so it will suffice to estimate $\frac{1}{\sigma^2+\lambda^2}$. But

(10) $$\frac{1}{\|\underset{\sim}{Y}\|^2} \sim \left(\frac{1}{\sigma^2+\lambda^2}\right)\frac{1}{\chi_p^2},$$

so

(11) $$E\left(\frac{1}{\|\underset{\sim}{Y}\|^2}\right) = \left(\frac{1}{\sigma^2+\lambda^2}\right)E\left(\frac{1}{\chi_p^2}\right) = \left(\frac{1}{\sigma^2+\lambda^2}\right)\left(\frac{1}{p-2}\right)$$ provided $p \geq 3$.

Thus, when $p \geq 3$, we may estimate $\left(\frac{1}{\sigma^2+\lambda^2}\right)$ by $\frac{p-2}{\|\underset{\sim}{Y}\|^2}$; if we replace $\frac{1}{\sigma^2+\lambda^2}$ by this estimate in (9), we obtain the James-Stein estimator

(12) $$\hat{\underset{\sim}{M}}_{JS} \equiv \left[1 - \frac{(p-2)\sigma^2}{\|\underset{\sim}{Y}\|^2}\right]\underset{\sim}{Y}.$$

[When $p=1$ or 2, $E\left(\frac{1}{\chi_p^2}\right) = \infty$ so this motivation of (12) fails. The estimator on the right of (12) is still well defined, but it will not dominate]

THEOREM: $E_{\underset{\sim}{M}}\|\hat{\underset{\sim}{M}}_{JS} - \underset{\sim}{M}\|^2 < E_{\underset{\sim}{M}}\|\hat{\underset{\sim}{M}} - \underset{\sim}{M}\|^2 \equiv p\sigma^2$ for every $\underset{\sim}{M} \in \mathbb{R}^p$ $(p \geq 3)$.

Proof: We will prove a more general result. Define

(13) $$\hat{\underset{\sim}{M}}_c = \left[1 - \frac{c\,\sigma^2}{\|\underset{\sim}{Y}\|^2}\right]\underset{\sim}{Y}, \qquad [c \text{ a constant}].$$

We shall show that for every $0 < c < 2(p-2)$,

(14) $$E_{\underset{\sim}{M}}\|\hat{\underset{\sim}{M}}_c - \underset{\sim}{M}\|^2 < p\sigma^2 \text{ for every } \underset{\sim}{M} \in \mathbb{R}^p,$$

and that the left side of (14) is minimized when $c = p-2$, i.e., when $\hat{\underset{\sim}{M}}_c = \hat{\underset{\sim}{M}}_{JS}$.

Since $\|\underset{\sim}{z}\|^2 = \underset{\sim}{z}'\underset{\sim}{z}$, we have

$$E_{\underset{\sim}{\mu}}\|\hat{\underset{\sim}{\mu}}_c - \underset{\sim}{\mu}\|^2 = E_{\underset{\sim}{\mu}}\left\|(\underset{\sim}{Y}-\underset{\sim}{\mu}) - \frac{c\sigma^2}{\|\underset{\sim}{Y}\|^2}\underset{\sim}{Y}\right\|^2$$

(15)

$$= \underbrace{E_{\underset{\sim}{\mu}}\|\underset{\sim}{Y}-\underset{\sim}{\mu}\|^2}_{\equiv p\sigma^2} - 2c\sigma^2 E_{\underset{\sim}{\mu}}\left[(\underset{\sim}{Y}-\underset{\sim}{\mu})'\frac{\underset{\sim}{Y}}{\|\underset{\sim}{Y}\|^2}\right] + c^2\sigma^4 E_{\underset{\sim}{\mu}}\left[\frac{1}{\|\underset{\sim}{Y}\|^2}\right].$$

Now $\underset{\sim}{Y} \sim N_p(\underset{\sim}{\mu}, \sigma^2 I_p)$, so

(16) $$E_{\underset{\sim}{\mu}}\left[(\underset{\sim}{Y}-\underset{\sim}{\mu})'\frac{\underset{\sim}{Y}}{\|\underset{\sim}{Y}\|^2}\right] = \sum_{i=1}^{p} E\left[(Y_i-\mu_i)\frac{Y_i}{\|\underset{\sim}{Y}\|^2}\right] = \sum_{i=1}^{p} \int\cdots\int_{\mathbb{R}^p} (Y_i-\mu_i)\frac{Y_i}{\|\underset{\sim}{Y}\|^2}\left[\prod_{j=1}^{p} f(Y_j-\mu_j)\,dY_j\right],$$

where

(17) $$f(Y_j-\mu_j) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(Y_j-\mu_j)^2/2\sigma^2}. \qquad [\text{i.e., } Y_j \sim N_1(\mu_j, \sigma^2)].$$

Now use integration by parts:

$$\int_{-\infty}^{\infty} \underbrace{\frac{Y_i}{\|\underset{\sim}{Y}\|^2}}_{\equiv u} \cdot \underbrace{(Y_i-\mu_i)\, e^{-(Y_i-\mu_i)^2/2\sigma^2} dY_i}_{\equiv dv}$$

$$\boxed{u = \frac{Y_i}{Y_1^2+\cdots+Y_p^2}, \quad du = \frac{1}{Y_1^2+\cdots+Y_p^2} - \frac{2Y_i^2}{(Y_1^2+\cdots+Y_p^2)^2} \quad V = -\sigma^2 e^{-(Y_i-\mu_i)^2/2\sigma^2}}$$

(18)

$$= \underbrace{\frac{Y_i}{\|\underset{\sim}{Y}\|^2}\left(-\sigma^2 e^{-(Y_i-\mu_i)^2/2\sigma^2}\right)\Big]_{-\infty}^{\infty}}_{\equiv 0} + \sigma^2\int_{-\infty}^{\infty}\left(\frac{1}{\|\underset{\sim}{Y}\|^2} - \frac{2Y_i^2}{\|\underset{\sim}{Y}\|^4}\right) e^{-(Y_i-\mu_i)^2/2\sigma^2},$$

so by (16),

$$E_{\underset{\sim}{\mu}}\left[(\underset{\sim}{Y}-\underset{\sim}{\mu})'\frac{\underset{\sim}{Y}}{\|\underset{\sim}{Y}\|^2}\right] = \sum_{i=1}^{p}\sigma^2\int\cdots\int_{\mathbb{R}^p}\left(\frac{1}{\|\underset{\sim}{Y}\|^2} - \frac{2Y_i^2}{\|\underset{\sim}{Y}\|^4}\right)\prod_{j=1}^{p} f(Y_j-\mu_j)\,dY_j$$

(19)

$$= \sigma^2 E_{\underset{\sim}{\mu}}\left[\frac{p}{\|\underset{\sim}{Y}\|^2} - \frac{2}{\|\underset{\sim}{Y}\|^2}\right] \qquad = \sigma^2(p-2)\, E_{\underset{\sim}{\mu}}\left(\frac{1}{\|\underset{\sim}{Y}\|^2}\right).$$

Therefore, from (15),

(20) $$\Delta(\underset{\sim}{\mu}) \equiv p\sigma^2 - E_{\underset{\sim}{\mu}}\|\hat{\underset{\sim}{\mu}}_c - \underset{\sim}{\mu}\|^2 = \sigma^4 c\left[2(p-2) - c\right] E_{\underset{\sim}{\mu}}\left(\frac{1}{\|\underset{\sim}{Y}\|^2}\right),$$

so $\Delta(\underset{\sim}{\mu}) > 0$ when $0 < c < 2(p-2)$, and $\Delta(\underset{\sim}{\mu})$ is maximized (for $\underline{\underline{\text{all}}}$ $\underset{\sim}{\mu}$) when $c = (p-2)$. This completes the proof of the Theorem.

Remarks: A. When $c = p-2$ so $\hat{\underset{\sim}{\mu}}_c = \hat{\underset{\sim}{\mu}}_{JS}$, the reduction in total MSE provided by $\hat{\underset{\sim}{\mu}}_{JS}$ is

(21) $$\Delta(\underset{\sim}{\mu}) = (p-2)^2\sigma^2 E_{\underset{\sim}{\mu}}\left(\frac{1}{\|\underset{\sim}{Y}/\sigma\|^2}\right) = (p-2)^2\sigma^2 E\left[\frac{1}{\chi^2_p(\delta)}\right],$$

where $\delta = \|\underset{\sim}{\mu}\|^2/\sigma^2$. Since the noncentral $\chi^2_p(\delta)$-distribution can be shown to have monotone likelihood ratio in δ, the expectation $E[1/\chi^2_p(\delta)]$ is decreasing in δ, so $\Delta(\underset{\sim}{\mu})$ is decreasing in $\|\underset{\sim}{\mu}\|$. Thus the maximum reduction in total MSE provided by $\hat{\underset{\sim}{\mu}}_{JS}$ occurs when $\underset{\sim}{\mu} = 0$ and is given by

(22) $$\Delta(\underset{\sim}{0}) = (p-2)^2\sigma^2 E\left[\frac{1}{\chi^2_p}\right] = (p-2)\sigma^2,$$

which is quite large for large p. On the other hand, as $\delta \to \infty$, $\chi^2_p(\delta) \to \infty$ and $E[1/\chi^2_p(\delta)] \to 0$ [this requires verification], so

(23) $$\Delta(\underset{\sim}{\mu}) \to 0 \quad \text{as } \|\underset{\sim}{\mu}\| \to \infty.$$

Thus, the total MSE of $\hat{\underset{\sim}{\mu}}_{JS}$, i.e., $E_{\underset{\sim}{\mu}}\|\hat{\underset{\sim}{\mu}}_{JS} - \underset{\sim}{\mu}\|^2$, depends on $\|\underset{\sim}{\mu}\|$ as follows:

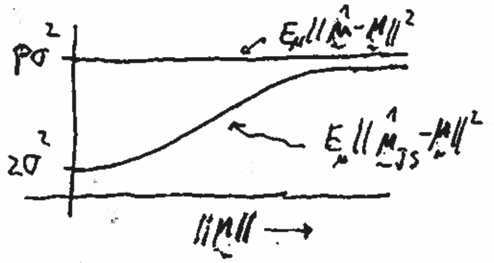


FIGURE 1.

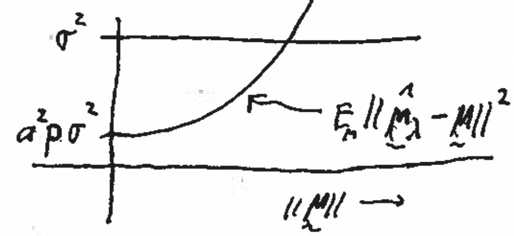


FIGURE 2.

Hence, the JS estimator $\hat{\underset{\sim}{\mu}}_{JS}$ "dominates" the LSE $\hat{\underset{\sim}{\mu}}$: its total MSE is always smaller. On the other hand, the MSE of the Bayes estimator $\hat{\underset{\sim}{\mu}}_\lambda$ in (6) (shown in Figure 2) exceeds that of $\hat{\underset{\sim}{\mu}}$ for sufficiently large $\|\underset{\sim}{\mu}\|$:

(24) $$E_{\underset{\sim}{\mu}}\|\hat{\underset{\sim}{\mu}}_\lambda - \underset{\sim}{\mu}\|^2 = E_{\underset{\sim}{\mu}}\|a\underset{\sim}{Y} - \underset{\sim}{\mu}\|^2 = E_{\underset{\sim}{\mu}}\|a(\underset{\sim}{Y}-\underset{\sim}{\mu}) + (a-1)\underset{\sim}{\mu}\|^2$$
$$= a^2 E_{\underset{\sim}{\mu}}\|\underset{\sim}{Y}-\underset{\sim}{\mu}\|^2 + (a-1)^2\|\underset{\sim}{\mu}\|^2 = a^2 p\sigma^2 + (a-1)^2\|\underset{\sim}{\mu}\|^2 \to \infty$$

where $a = \frac{\lambda^2}{\sigma^2+\lambda^2} < 1$, so $E_{\underset{\sim}{\mu}}\|\hat{\underset{\sim}{\mu}}_\lambda - \underset{\sim}{\mu}\|^2 \to \infty$ as $\|\underset{\sim}{\mu}\| \to \infty$.

B. The JS estimator $\hat{\mu}_{JS}$ does not dominate the MLE $\hat{\mu}$ in the strong sense

$$E_\mu(\hat{\mu}-\mu)(\hat{\mu}-\mu)^t - E_\mu(\hat{\mu}_{JS}-\mu)(\hat{\mu}_{JS}-\mu)^t \equiv \Sigma_{MLE} - \Sigma_{JS}$$

is NOT positive semidefinite. The THEOREM on p. 4.4 states that $tr(\Sigma_{MLE}) > tr(\Sigma_{JS})$, but it can be shown that some diagonal elements of Σ_{JS} (i.e., $E_\mu(\hat{\mu}_{JS,i}-\mu_i)^2$) can exceed the corresponding diagonal elements of Σ_{MLE} (i.e., $E_\mu(\hat{\mu}_i-\mu_i)^2$).

C. The JS estimate has the form $a(\|Y\|)\cdot Y$, where $a(\|Y\|) = \left[1-\frac{(p-2)\sigma^2}{\|Y\|^2}\right]$. Since $a(\|Y\|) < 1$, $\hat{\mu}_{JS}$ "shrinks" Y toward 0 (mimicking the behavior of the Bayes estimator $\hat{\mu}_B$, which shrinks Y toward $0 \equiv$ the prior mean). However, it can occur that $a(\|Y\|) < 0$, (whenever $\|Y\|^2 < (p-2)\sigma^2$), in which case $\hat{\mu}_{JS}$ "shrinks Y" too far. It can be shown that the modified JS estimator $\hat{\mu}_{JS}^+ \equiv a^+(\|Y\|)\cdot Y$ dominates $\hat{\mu}_{JS}$ (hence $\hat{\mu}_{JS}^+$ dominates $\hat{\mu}$), where $a^+(\|Y\|) = \max\{a(\|Y\|), 0\}$.

D. If σ^2 is unknown but we have an estimator $\hat{\sigma}^2$ of σ^2 such that $\hat{\sigma}^2$ is independent of Y and $\hat{\sigma}^2 \sim \sigma^2 \cdot \chi^2_k/k$ for some k, then estimators of μ of the form $[1-\frac{c\hat{\sigma}^2}{\|Y\|^2}]\cdot Y$ will dominate $\hat{\mu}$ for suitable choices of c.

E. As already mentioned, if we replace the prior (3) by $N(\mu_0, \lambda^2 I_p)$, we are led to the "JS" estimator that shrinks Y toward μ_0, namely,

$$(25)\quad \hat{\mu}_{JS}(\mu_0) \equiv \left[1-\frac{(p-2)\sigma^2}{\|Y-\mu_0\|^2}\right](Y-\mu_0) + \mu_0 .$$

The choice of μ_0 should not be made arbitrarily, but should be based on the prior information available. If no such info. is available, the JS estimator should not be used (this is my personal opinion).

⊛ F. It may be preferable to shrink toward a linear subspace, e.g. toward $(\bar{Y},\dots,\bar{Y})$. Use p-3.