EMPLOYING VAGUE INEQUALITY INFORMATION IN THE ESTIMATION OF NORMAL MEAN VECTORS[1]

(Estimators that shrink to closed convex polyhedra)

by

M. E. Bock
Purdue University

Technical Report #81-42

Department of Statistics
Purdue University

April 1982

[1]This research was supported by National Science Foundation Grants MCS-78-02300 and MCS-81-01670.

EMPLOYING VAGUE INEQUALITY INFORMATION IN THE ESTIMATION OF NORMAL MEAN VECTORS[1]

(Estimators that shrink to closed convex polyhedra)

M. E. Bock

Department of Statistics
Purdue University
West Lafayette, Indiana, U.S.A.

I. INTRODUCTION

Consider the problem of estimating the mean vector θ of a p dimensional normal random vector X with covariance I_p where p is three or more. Let the possible values of θ be all of $\mathbb{R}^p$ and define the loss function for an estimator $\hat{\theta}$ of θ to be

$$L(\theta,\hat{\theta}) = ||\hat{\theta}-\theta||^2,$$

the square of the Euclidean norm.

James and Stein [1] exhibited estimators which shrink to a fixed vector (say zero) in the parameter space yet have the property that they are minimax and dominate the maximum likelihood estimator. Sclove, Morris and Radhakrishnan [3] noted that estimators with this property exist which shrink to linear manifolds when p is greater than or equal to the dimension of the linear manifold plus three. One of the estimators they considered has the form

$$(\hat{\theta}(X))_i = \bar{X} + (X_i-\bar{X})(1-c/s)^+$$

where c is in (0, 2(p-3)) and

[1] This research was supported by National Science Foundation Grants MCS-78-02300 and MCS-81-01670.

$$\bar{X} = \sum_{j=1}^{p} X_j/p \text{ and } s = \sum_{j=1}^{p} (X_j-\bar{X})^2.$$

This estimator shrinks to the linear manifold which consists of vectors whose components are all equal in $\mathbb{R}^p$. The dimension of this manifold is one and the estimator is minimax if p is greater than or equal to four, which is three plus the dimension of the manifold.

A fixed vector or a linear manifold are special cases of a closed convex polyhedron, which is the solution set for a finite system of linear inequality constraints. Suppose that vague inequality information about θ is present in the following form: One suspects that the mean vector θ satisfies the following finite system of linear inequality constraints:

$$A_i^t\theta \leq b_i,$$

where A_i is a known p-dimensional vector and b_i is a known scalar, $i = 1,\ldots,n$. Let Q be the closed convex polyhedron which satisfies these constraints. (We will assume that the system is solveable and that Q is not empty.) A class of estimators is exhibited in Section 2 such that the estimators shrink X to Q in the direction of P(X) the closest point of Q to X. Some of the estimators dominate $\hat{\theta}_0$ the maximum likelihood estimator and are of course minimax. A dominating estimator is exhibited for any estimator $\hat{\theta}$ which assumes the value X when X is not in Q.

The shrinkage estimator is robust in the following sense: If θ is actually in Q, then any estimator which shrinks X to Q by the choice of a value in the interval (X,P(X)] has a smaller loss than the estimator whose value is X when X is not in Q, no matter what the distribution of X. For θ not in Q, the estimators considered improve most in expected loss for values of θ close to Q. Thus one is rewarded most in using these estimators when the vague inequality information that θ is in Q is correct or "nearly correct". The shrinkage is done under the following condition:

Shrink X to Q only if P(X) lies on a face of Q with dimension d_0 such that

$$p \geq d_0 + 3,$$

i.e. the codimension of the face $(p-d_0)$ is three or more.

Section 2 of the paper provides the exact statement of these results in a theorem as well as the proof of the theorem. An appendix follows with proofs of lemmas used in the proof of the theorem.

II. A CLASS OF ESTIMATORS

In the theorem of this section dominating estimators $\hat{\theta}_g$ are given for any estimator $\hat{\theta}_1(X)$ which assumes the value X when X is not inside the convex closed polyhedron Q. If one completely specifies the values of $\hat{\theta}_1$ by defining $\hat{\theta}_1(X)$ to be X for X in Q, then $\hat{\theta}_1$ is the maximum likelihood estimator $\hat{\theta}_0$. The estimators that dominate $\hat{\theta}_0$ are also minimax estimators since $\hat{\theta}_0$ is minimax with constant risk for all values of θ.

THEOREM. *Assume that the* p-*dimensional random vector X is normally distributed with mean vector* θ *in* $\mathbb{R}^p$ *and identity covariance matrix. Let* Q *be the closed convex polyhedron which is the solution to a finite system of linear inequalities, i.e.*

$$Q = \{Y \textit{ in } \mathbb{R}^p\text{:}\ \ A_i{}^tY \leq b_i,\ i = 1,\ldots,n\}$$

where A_i *are known* p-*dimensional vectors and* b_i *are known scalars. Define* $\hat{\theta}_1(X)$ *to be any estimator of* θ *satisfying*

$$\hat{\theta}_1(X) = X \textit{ for } X \textit{ not in } Q.$$

Define P(X) *to be the orthogonal projection of* X *to* Q *and let* d *be the codimension of the face of* Q *in whose relative interior* P(X) *lies. Define*

$$\hat{\theta}_g(X) = \begin{cases} \hat{\theta}_1(X), & \textit{for } X \textit{ in } Q \textit{ and for } X \textit{ not in } Q \textit{ but } d \leq 2; \\ P(X)+g(||X-P(X)||^2)(X-P(X)), & \textit{for } X \textit{ not in } Q \textit{ and } d \geq 3, \end{cases}$$

where g *is a real-valued differentiable function defined on non-negative real values* t, *satisfying*

a) $g(t) \leq 1$;

b) $(1-g(t))\ (2d-t(1-g(t))) \geq 4tg'(t)$ *with strict inequality for a set of* t *values with positive Lebesgue measure;*

c) $\lim_{t\to\infty} t^{\frac{1}{2}}(1-g(t))\exp(-t/2) = 0$ *and*

$\lim_{t\to 0} t^{\frac{1}{2}}(1-g(t+c))\exp(-t/2) = 0$ *for any* $c > 0$.

Then $\hat{\theta}_g$ *dominates* $\hat{\theta}_1$ *under squared error loss* $||\hat{\theta}-\theta||^2$. *It is assumed that* Q *is not empty.*

Remark. The conditions given in c) assure that the usual "integration by parts" analysis of the risk is legitimate.

Example. If we define a constant c in $(0, 2(d-2))$ and set

$g(t) = 1-c/t$,

then g satisfies the conditions of the theorem. It is a function used by James and Stein (1).

III. OUTLINE OF THE PROOF BY PARTS

A. Because every vector's projection to Q lies in the relative interior of some face of Q and there are a finite number of faces of Q, it suffices to show the following: For each face $\mathcal{T}$ of Q with codimension greater than or equal to three, the expected value of

$$\nabla(X) = ||\hat{\theta}_g(X)-\theta||^2 - ||\hat{\theta}_1(X)-\theta||^2$$

is negative where the expectation is taken over all values of X whose projection to Q lies in the relative interior of that face $\mathcal{J}$ of Q. To exhibit this result, both the relative interior of the face of Q and the projection of X to the face are carefully described. (The description is actually valid for faces with co-dimension less than three.)

B. It is shown that the values of X over which the expected value of $\nabla(X)$ is taken are those whose projections to a certain pointed polyhedron S lie in the relative interior of a certain face F of S. (The convex polyhedron Q is shown to be the sum of a linear subspace L and a pointed polyhedron S (i.e., one with vertices) where S lies in $L^{\perp}$ the orthogonal complement space of L. A face $\mathcal{J}$ in Q may be written as the sum of L and a face F of S. The codimension of $\mathcal{J}$ is equal to the codimension d_F of F in $L^{\perp}$. The vectors exterior to Q whose projections to Q lie in the relative interior of $\mathcal{J}$ are those whose projections to S lie in the relative interior of F.)

C. Assume that one of the vertices in the face F is the zero vector and define N to be the linear subspace N generated by F and define $N^{\perp}$ to be its orthogonal complement space in $L^{\perp}$. (If the assumption is not true, make a translation of the whole problem, replacing S by S_F and F by F_0 which does have the zero vector for a vertex.)

D. A lemma shows that $P_S(X)$ is in F^I if and only if $P_N(X)$ is in F^I and $P_{N^{\perp}}(X)$ is in a certain face C of S^p, the polar cone of S. See Figure 1. Because $P_S(X)$ in F^I implies $P_S(X)$ equals $P_N(X)$, we may replace $P_S(X)$ in $\nabla(X)$ by $P_N(X)$ to obtain $\nabla^*(X)$ which is a function of X only through $P_{N^{\perp}}(X)$. Also, the lemma implies that the expectation of $\nabla(X)$ over S such that $P_S(X)$ is in F^I is equivalent to the expectation of $\nabla^*(X)$ over X such that

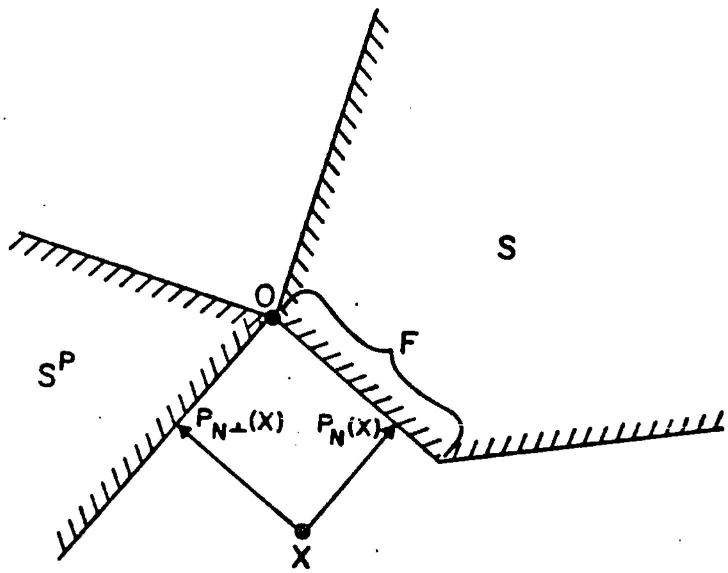

Figure 1.

$P_N(X)$ is in F^I and $P_{N^\perp}(X)$ in C. This is clearly the product of the probability that $P_N(X)$ is in F^I and the expectation of $\nabla^*(X)$ taken over X such that $P_{N^\perp}(X)$ is in C.

E. The face C may be written as the sum of a linear subspace M in $N^\perp$ and a pointed cone C_0 in $M^\perp$, the orthogonal complement of M in $N^\perp$. Thus we can evaluate ∇^* over X such that $P_{M^\perp}(P_{N^\perp}(X))$ is in C_0^I.

F. It is possible to describe C_0 as the union of simplicial cones whose relative interiors are disjoint. (See Definition 7 in the appendix.) So the expectation of $\nabla^*(X)$ is a sum of expectations of $\nabla^*(X)$ over X values in the relative interior of each of these simplicial cones. The original conditions of the theorem on the estimator may then be used to show that each expectation in the summand is negative.

IV. PROOF OF THE THEOREM

A. By Lemma 4 of the appendix, P(X) lies in the relative interior of some face of Q. Let H be the finite collection of the of faces of Q with codimension three or more. Define

$$h(t) = 1-g(t)$$

and note that if P(X) is in $\mathcal{J}^I$, the relative interior of a face $\mathcal{J}$ in H, then

$$\hat{\theta}_g(X) = X-h(||X-P(X)||^2)(X-P(X)).$$

Write the risk difference as

$$R(\theta,\hat{\theta}_g)-R(\theta,\hat{\theta}_1) = E[||\hat{\theta}_g(X)-||\ ^2] - E[||\hat{\theta}_1-\theta||^2]$$

$$= \sum_{\mathcal{J}\text{ in }H} E[I_{\mathcal{J}^I}(P(X))[2(\theta-X)^t(X-P(X))h(||X-P(X)||^2)$$

$$+ h^2(||X-P(X)||^2)||X-P(X)||^2]].$$

It suffices to show that each expectation in the sum is nonnegative. Let $\mathcal{J}$ be a fixed face in H and consider its corresponding expectation in the sum that forms the risk difference.

B. Stoer and Witzgall [4] noted a thoerem of Motzkin [2] shows that a closed convex polyhedron Q can be decomposed into the sum of a linear subspace L and a pointed polyhedron S in $L^\perp$, the orthogonal complement space of L. (See Lemma 6 in the appendix.) The projection of X to Q may be written as

$$P(X) = P_L(X) + P_S(P_{L^\perp}(X))$$

where P_L and $P_{L^\perp}$ and P_S are the projections to L, $L^\perp$ and S respectively. Setting

$$Y = P_{L^\perp}(X),\ \text{and}\ \eta = P_{L^\perp}(\theta),$$

we have that

$$X-P(X) = Y-P_S(Y)$$

and

$$(X-\theta)^t(X-P(X)) = (Y-\eta)^t(Y-P_S(Y)).$$

By Lemma 6 of the appendix, $P(X)$ lies in $\mathcal{J}^I$ if and only if $P_S(Y)$ lies in the relative interior of some face F of S with codimension d_F in $L^\perp$ equal to the codimension of $\mathcal{J}$.

The expectation corresponding to $\mathcal{J}$ in the sum for the risk difference is

$$(*) \quad E[(2(\eta-Y)^T(Y-P_S(Y))h(||Y-P_S(Y)||^2)$$
$$+ h^2(||Y-P_S(Y)||^2)||Y-P_S(Y)||^2)I_{F^I}(P_S(Y))]$$

To prove the theorem it suffices to show that each of these expectations (*) is negative.

C. Let $V_1,\ldots,V_m$ be the vertices of S. Since one or more of them belong to F, define V_F to be the vertex with lowest subscript in F. Define a new polyhedron S_F by subtracting V_F from each vector in S. The set F_0 formed by subtracting V_F from each vector in F is a face in the pointed polyhedron S_F and the codimension of F_0 is equal to that of F. Observe that one of the vertices of F_0 is the zero vector. (Note that now and in the future comments the vectors are all restricted to lie in $L^\perp$ and we speak of the codimension with respect to that space.) Note that $P_S(Y)$ lies in the relative interior of F if and only if $(P_S(Y)-V_F)$ lies in the relative interior of F_0. See, too, that $(P_S(Y)-V_F)$ is $P_{S_F}(Y-V_F)$. Define

$$Z = Y-V_F \text{ and } \rho = \eta-V_F.$$

Thus we can rewrite (*) as

$$(**)\quad E[(2((Z-P_{S_F}(Z))^t(\rho-P_{S_F}(Z))$$
$$-\ ||Z-P_{S_F}(Z)||^2)h(||Z-P_{S_F}(Z)||^2)$$
$$+\ h^2(||Z-P_{S_F}(Z)||^2)||Z-P_{S_F}(Z)||^2)I_{F_0^I}(P_{S_F}(Z))].$$

D. Let N be the linear subspace in $L^\perp$ generated by F_0 and let $N^\perp$ be its orthogonal complement space in $L^\perp$. Then according to Lemma 9, $P_{S_F}(Z)$ is in F_0^I if and only if $P_N(Z)$ is in F_0^I and $P_{N^\perp}(Z)$ is in C where

$$C = N^\perp \cap S_F^p$$

and S_F^p is the polar cone of S_F. The Remark accompanying Lemma 9 implies that C has dimension d_F equal to the codimension of F_0 and C is a face of S_F^p. According to Lemma 8, $P_{S_F}(Z)$ is $P_N(Z)$ if $P_{S_F}(Z)$ is in F_0^I. Noting that $P_{N^\perp}(Z)$ is $(Z-P_{S_F}(Z))$, we may rewrite (**) as

$$-E[(2((||P_{N^\perp}(Z)||^2-(P_{N^\perp}(\rho))^tP_{N^\perp}(Z))h(||P_{N^\perp}(Z)||^2)$$
$$-\ h^2(||P_{N^\perp}(Z)||^2)||P_{N^\perp}(Z)||^2)I_C(P_{N^\perp}(Z))I_{F_0^I}(P_N(Z))].$$

(We have used the fact that $P_N(Z)$ and $P_{N^\perp}(Z)$ are orthogonal as well as the fact that $P_N(\rho)$ and $P_{N^\perp}(Z)$ are orthogonal.) Let $P_L, P_{L^\perp}, P_{N^\perp}$ and P_N be the symmetric idempotent projection matrices to the linear subspaces $L, L^\perp, N^\perp$, and N respectively. Then $P_N(Z)$ is normally distributed with mean $P_N(P_{L^\perp}\theta-V_F)$ and covariance $P_NP_{L^\perp}P_N$. Also $P_N(Z)$ is independent of $P_{N^\perp}(Z)$ which is normally distributed with mean $P_{N^\perp}(P_{L^\perp}\theta-V_F)$ and covariance $P_{N^\perp}P_{L^\perp}P_{N^\perp}$. Because $P_N(Z)$ is independent of $P_{N^\perp}(Z)$, the last expectation is equal to the product of the probability that $P_N(Z)$ is in F_0^I and the expectation

$$E[(2(P_{N^\perp}(\rho)-P_{N^\perp}(Z))^t P_{N^\perp}(Z)h(||P_{N^\perp}(Z)||^2)$$
$$+ h^2(||P_{N^\perp}(Z)||^2)||P_{N^\perp}(Z)||^2)I_C(P_{N^\perp}(Z))].$$

Because d_F is positive the probability is positive and to show that (**) is negative it suffices to show that the last expectation is negative.

E. Because C is a face of S_F^p, it is also a polyhedral cone. By Lemma 4, $P_{N^\perp}(Z)$ in C implies that $P_{N^\perp}(Z)$ is in the relative interior of some face of C. But $P_{N^\perp}(Z)$ is in the relative interior of those faces with dimension less than d_F only for a set of Lebesgue measure zero in the d_F-dimensional space $N^\perp$. So assume that $P_{N^\perp}(Z)$ is in C^I, the only face of C with dimension d_F. Define M to be the lineality space of C in $N^\perp$ with dimension (d_F-r) and let $M^\perp$ be its orthogonal complement space in $N^\perp$. Then the pointed cone

$$C_0 = M^\perp \cap C$$

has dimension r. By Lemma 6, $P_{N^\perp}(Z)$ is in C^I if and only if $P_{M^\perp}(P_{N^\perp}(Z))$ is in C_0^I.

Define U to be $P_M(P_{N^\perp}(Z))$ and ρ_U to be $P_M(P_{N^\perp}(\rho))$ and W to be $P_{M^\perp}(P_{N^\perp}(Z))$ and ρ_W to be $P_{M^\perp}(P_{N^\perp}(\rho))$. Applying the orthogonality of P_M and $P_{M^\perp}$, the last expectation is

$$(***)\quad E[(2((\rho_W-W)^t W + (\rho_U-U)^t U)h(||U||^2+||W||^2)$$
$$+ h^2(||U||^2 + ||W||^2)(||U||^2 + ||W||^2))I_{C_0^I}(W)].$$

Let P_M and $P_{M^\perp}$ be the symmetric idempotent projection matrices to the linear subspaces M and $M^{\perp}$ respectively. Then U is normally distributed and its mean is ρ_U and its covariance Σ_U is $P_M P_{N^{\perp}} P_{L^{\perp}} P_{N^{\perp}} P_M$. Note that U is independent of W which is normally distributed with mean ρ_W and covariance Σ_W equal to $P_{M^{\perp}} P_{N^{\perp}} P_{L^{\perp}} P_{N^{\perp}} P_{M^{\perp}}$. Using the independence of U and W, a standard integration by parts argument implies that

$$(****)\quad E[2(U-\rho_U)^t U h(||U||^2 + ||W||^2) I_{C_0^I}(W)]$$

$$= E[2I_{C_0^I}(W)((d_F - r)h(||U||^2+||W||^2)$$

$$+ 2h'(||U||^2+||W||^2)||U||^2)].$$

Next we will evaluate

$$(*****)\quad E[2(W-\rho_W)^t W h(||U||^2+||W||^2) I_{C_0^I}(W)].$$

F. Let $C_1,\ldots,C_{n_0}$ be a simplicial decomposition of C_0, i.e. the C_i are simplicial cones of dimension r whose union is C_0 and whose relative interiors are disjoint. (See Definition 7 and remark in the appendix.) Note that except for a set of Lebesgue measure zero in $M^{\perp}$, any vector in C_0^I is in C_i^I the relative interior of C_i for some i. Thus $I_{C_0^I}(W)$ may be replaced by

$$\sum_{i=1}^{n_0} I_{C_i^I}(W)$$

in the last expectation since the C_i^I are disjoint. Then the last expectation is the sum of the expectations

$$E[2(W-\rho_W)^t W\, h(||U||^2+||W||^2) I_{C_i^I}(W)],\ i = 1,\ldots,n_0.$$

Let $A_1^i,\ldots,A_r^i$ be the r linearly independent vectors which generate C_i and which form the columns of the matrix A^i. By Lemma 5, W is in C_i^I if and only if W is $A^i\alpha^i$ where α^i is a vector of positive components. Because A_j^i, $j = 1,\ldots,r$, are independent vectors in the r dimensional space $M^\perp$, it is always true that W has a representation in the form $A^i\alpha^i$, for $i = 1,\ldots,n_0$. Thus W is in C_i^I if and only if

$$T = [(A^i)^t A^i]^{-1}(A^i)^t W$$

is a vector of positive components. Now T is normally distributed with mean τ equal to $[(A^i)^t A^i]^{-1}(A^i)^t \rho_W$ and covariance

$$\Sigma = [(A^i)^t A^i]^{-1}(A^i)^t P_{M^\perp} P_{N^\perp} P_{L^\perp} P_{N^\perp} P_{M^\perp} A^i [(A^i)^t A^i]^{-1}.$$

Because the columns of A^i are contained in $M^\perp$ which is contained in $N^\perp$ which is contained in $L^\perp$, we have that A^i is $P_{L^\perp} P_{N^\perp} P_{M^\perp} A^i$. Thus Σ is $[(A^i)^t A^i]^{-1}$ since $P_{L^\perp}$ is idempotent. Thus the last expectation is

$$E[2(\prod_{j=1}^{r} I_{(0,\infty)}(T_j))\{(T-\tau)^t\Sigma^{-1}T\ h(T^t\Sigma^{-1}T+||U||^2)\}].$$

By Lemma 10, this equals

$$E[2(\prod_{j=1}^{r} I_{(0,\infty)}(T_j))\{r\ h(T^t\Sigma^{-1}T+||U||^2)+2h'(T^t\Sigma^{-1}T+||U||^2)\, T^t\Sigma^{-1}T\}].$$

Using the definition of T, this may be written as

$$E[2I_{C_i^I}(W)(rh(||U||^2+||W||^2)+2h'(||U||^2+||W||^2)||W||^2)].$$

Summing the last expectation over $i = 1,\ldots,n_0$, we may represent (*****) as

$$E[2I_{C_0^I}(W)(rh(||U||^2+||W||^2)+2h'(||U||^2+||W||^2)||W||^2)].$$

Incorporating the final representations for (****) and (*****), we may write (***) as

$$E[(-2d_F h(||U||^2+||W||^2)-4h'(||U||^2+||W||^2)(||U||^2+||W||^2)$$
$$+ h^2(||U||^2+||W||^2)(||U||^2+||W||^2))I_{C_0^I}(W)].$$

Recalling that h(t) is (1-g(t)) and assumptions a) and b) of the theorem, the integrand is negative for a set of values with positive Lebesgue measure. Thus the expectation (***) is negative as was to be shown.

V. APPENDIX

Definitions. Every closed convex polyhedron Q in $\mathbb{R}^p$ is described by a finite system of linear inequalities

$$Q = \{Y \text{ in } \mathbb{R}^p\text{:} \quad A_i{}^t Y \leq b_i,\ i = 1,\ldots,n\}$$

where A_i is a vector and b_i is a scalar, $i = 1,\ldots,n$. A linear inequality in the description is defined to be a *nonsingular inequality* if at least one vector in the polyhedron satisfies it strictly, i.e. if there is a vector Y_0 in Q such that

$$A_i^t Y_0 < b_i,$$

then

$$\{A_i^t Y \leq b_i\}$$

is a nonsingular inequality for Q.

A *relative interior point* of the polyhedron Q is a point of the polyhedron Q which strictly satisfies every nonsingular inequaltiy.

LEMMA 1. *Suppose there is a vector* P(X) *in the convex set* K *which is the orthogonal projection of the vector* X *to* K. *Then for all* Y *in* K,

$$(X-P(X))^t(Y-P(X)) \leq 0.$$

Proof. Suppose that $(X-P(X))^t(Y-P(X))$ is positive for some Y in K and define

$$M = \max\{||X-P(X)||^2,\ ||Y-P(X)||^2\}.$$

Then, by the Cauchy-Schwartz inequality and the definition of M,

$$0 < \rho \leq 1$$

where

$$\rho = (X-P(X)))^t(Y-P(X))/M.$$

Thus Y* is a member of K where

$$Y^* = \rho Y+(1-\rho)P(X).$$

Note that

$$\begin{aligned} ||X-Y^*||^2 &= ||(X-P(X))-\rho(Y-P(X))||^2 \\ &= ||X-P(X)||^2 + \rho^2||Y-P(X)||^2-2\rho(X-P(X))^t(Y-P(X)) \\ &= ||X-P(X)||^2 + \rho^2\{||Y-P(X)||^2-2M\}. \end{aligned}$$

The term multiplying ρ^2 is negative by the definition of M. This implies that Y* is closer to X than P(X), a contradiction to the definition of P(X). Thus, the original supposition leads to a contradiction.

LEMMA 2. *Let zero be contained in the convex set* K *which is contained in the convex set* Q. *Assume that* P(X) *the projection* *of the vector* X *to* Q *lies in the relative interior of* K. *Then*

$$(X-P(X))^t P(X) = 0.$$

Proof. We assume that P(X) is not zero. The following lemma (3.2.9, page 90) is given in Stoer and Witzgall [4]:

LEMMA. [*Let* K *be a convex set. If* X *is a vector in the relative interior of* K, *then for each vector* Y *in the linear manifold generated by* K *there exists a positive scalar* ε *such that the vectors* $(X+\varepsilon(Y-X))$ *and* $(X-\varepsilon(Y-X))$ *are in* K.]

Because P(X) is a relative interior point of K and because zero is in K, the LEMMA cited above implies that the line segment $[0,(1+\varepsilon)P(X)]$ is in K where ε is a positive constant.

Let Y be the projection of X to the linear subspace spanned by the vector P(X):

$$Y = \beta\, P(X)$$

where

$$\beta = (P(X))^t X / ||P(X)||^2.$$

We know that β is greater than or equal to one since zero in Q implies

$$(X-P(X))^t(0-P(X)) \leq 0,$$

by Lemma 1. Because $\beta\, P(X)$ and $(X-\beta P(X))$ are orthogonal, so are $(\beta-1)P(X)$ and $(X-\beta\, P(X))$; thus

$$||X-P(X)||^2 = ||X-\beta P(X)||^2 + (\beta-1)^2||P(X)||^2.$$

If β is not one, then $\beta\, P(X)$ is closer to X than P(X). If $\beta\, P(X)$ is in Q, then this is a contradiction to the definition of P(X). Assume that $\beta\, P(X)$ is not in Q. Then

$$\begin{aligned} ||X-P(X)||^2 &= ||(X-(1+\varepsilon)P(X))+\varepsilon\, P(X)||^2 \\ &= ||X-(1+\varepsilon)P(X)||^2 + \varepsilon^2||P(X)||^2 \\ &\qquad + 2(X-(1+\varepsilon)P(X))^t P(X)\varepsilon. \end{aligned}$$

The last term in the above equality may be written as

$$2(X-\beta P(X)+(\beta-1-\varepsilon)P(X))^t P(X)\varepsilon$$

$$= 2(\beta-1-\varepsilon)\varepsilon||P(X)||^2$$

since the orthogonality of β P(X) and (X-β P(X)) implies the orthogonality of ε P(X) and (X-β P(X)). Thus

$$||X-P(X)||^2 = ||X-(1+\varepsilon)P(X)||^2+[(\beta-1)^2-(\beta-1-\varepsilon)^2]||P(X)||^2.$$

Since β is greater than or equal to one and β P(X) is not in Q we have

$$\beta > 1+\varepsilon.$$

This implies along with the last equality that $(1+\varepsilon)$P(X) is closer to X than P(X), which is a contradiction to the definition of P(X). So β equals one. The definition of β implies the conclusion of the lemma.

LEMMA 3. *Let* F *be a convex set containing zero. Assume that* F *is contained in the convex set* Q *and that* P(X) *the projection of the vector* X *to* Q *lies in the relative interior of* F. *Then* (X-P(X)) *is the projection of* X *to the polar cone* Q^p *of* Q. *Furthermore,* P(X) *is also the projection of* X *to* Q^{pp} *the polar cone of the polar cone of* Q.

Proof. We make use of a lemma (2.7.5, page 51) of Stoer and Witzgall [4]:

LEMMA: [*Let* C *in* $\mathbb{R}^n$ *be a cone. If* X *in* $\mathbb{R}^n$ *admits an orthogonal decomposition* X = Y+Z *with* Y *in* C, Z *in* C^p *and* $Y^tZ = 0$, *then* Y *and* Z *are the projections of* X *into* C *and* C^p.]

Since P(X) is in Q, we have P(X) in Q^{pp} which contains Q. By Lemma 2, P(X) and (X-P(X)) are orthogonal. Using the LEMMA cited above it only remains to show that (X-P(X)) is in Q^p.

Because Lemma 1 implies that for any Y in Q,

$$(X-P(X))^t(Y-P(X)) \leq 0,$$

we have that for any Y in Q,

$$(X-P(X))^tY \leq 0$$

since P(X) and (X-P(X)) are orthogonal. Thus (X-P(X)) is in Q^P.

LEMMA 4. *Any point in a closed convex polyhedron is a relative interior point of some face of the polyhedron.*

Proof. Suppose that the point X is in the closed convex polyhedron Q. Suppose X is not a relative interior point of Q. (Since Q is a face of itself, then X would be a relative interior point of a face if X were a relative interior point of Q.) Then the set I_Q is nonempty where

$$Q = \{Z:\ A_i^tZ \leq b_i,\ i = 1,\ldots,n\} \text{ and}$$

$$I_Q = \{i:\ A_i^tX=b_i \text{ and } \{A_i^tZ \leq b_i\} \text{ is a nonsingular inequality for } Q\}.$$

Note I_Q is contained in $\{1,\ldots,n\}$. Then X is in the face F of Q where F is the intersection of Q and the hyperplanes of the form

$$\{Y:\ A_i^tY = b_i\}$$

and i is in I_Q. Because the inequalities are nonsingular for Q, F is a proper face of Q. If X is a relative interior point of F, we are done. Suppose X is not a relative interior point of F (which is also a closed convex polyhedron). Then the set I_F is nonempty where

$$F = \{Z:\ D_i^tZ \leq c_i,\ i = 1,\ldots,m\} \text{ and}$$

$$I_F = [i:\ D_i^tX=c_i \text{ and } \{D_i^tZ \leq c_i\} \text{ is a nonsingular inequality for } F].$$

We can conclude X is also in a proper face G of F which is a proper face of Q distinct from F. If X is not a relative interior point of G, we can continue in the same fashion to find another proper face of Q which contains X. Because Q has a finite number of faces, the process must end.

LEMMA 5. *Let* C *be a polyhedral cone generated by* d *independent vectors* A_i *which form the columns of a matrix* A. *The vector* X *lies in the relative interior* C^I *of* C *if and only if the components of the vector*

$$\alpha = (A^tA)^{-1}A^tX$$

are positive and X *is* $A\alpha$.

Proof. Note that X is in C^I if and only if X has the form $A\beta$ where the components of the vector β are positive. Suppose X is in C^I, i.e. X is $A\beta$ where the components of β are positive. Then $A(A^tA)^{-1}A^tX$ is $A(A^tA)^{-1}A^tA\beta$ which equals $A\beta$ which is X. So X equals $A\alpha$. Now $(A^tA)^{-1}A^tX$ equals $(A^tA)^{-1}A^tA\beta$ which equals β, a vector with positive components. Suppose the components of α are positive and X is $A\alpha$. Then X is in C^I.

LEMMA 6. *For* p-*dimensional vectors* A_i *and scalars* b_i, *define* Q *to be a closed convex polyhedron given by*

$$Q = \{X \text{ in } \mathbb{R}^p\colon\ A_i^tX \leq b_i,\ i = 1,\ldots,n\}.$$

Define the linear subspace (called the lineality space of Q*)*

$$L = \{X \text{ in } \mathbb{R}^p\colon\ A_i^tX=0,\ i = 1,\ldots,n\}$$

and let $L^\perp$ *be its orthogonal complement space. Define the pointed polyhedron* S *by*

$$S = Q \cap L^\perp.$$

Then the codimension of a face F_I *of* Q *is equal to the codimension in* $L^\perp$ *of the face*

$$F_I^* = F_I \cap S$$

of S. Furthermore the vector X lies in the relative interior of F_I if and only if $P_{L^\perp}(X)$ lies in the relative interior of F_I^*, where $P_{L^\perp}$ is the projection to $L^\perp$.

Remark. F_I is $L \oplus F_I^*$, the direct sum of L and F_I^*.

Proof. A theorem of Motzkin [2] cited in Stoer and Witzgall [4] shows that Q is $L \oplus S$. For I a subset of $\{1,\ldots,n\}$, define the face F_I to be

$$F_I = Q \cap \{X\colon A_i^t X = b_i, \text{ for i in I}\}.$$

The definition of F_I^* implies that

$$F_I^* = \{X \text{ in } S\colon \ A_i^t X = b_i, \text{ for i in I}\}.$$

It is clear that F_I^* is a face of S when we write

$$S = \{X \text{ in } L^\perp\colon \ A_i^t X \le b_i, \ i = 1,\ldots,n\}.$$

(Proof of Remark: Clearly F_I is contained in $(P_L(F_I) \oplus P_{L^\perp}(F_I))$. It suffices to show that $P_L(F_I)$ is L and $(L \oplus P_{L^\perp}(F_I))$ is contained in F_I and $P_{L^\perp}(F_I)$ is F_I^*. Assume that F_I is nonempty. Let X_0 be in F_I. Then for any X in L, and each i in I,

$$A_i^t(X+P_{L^\perp}(X_0)) \le b_i$$

since $A_i^t X$ is zero and $A_i^t P_{L^\perp}(X_0)$ is $A_i^t X_0$. The same sort of reasoning shows that for each i not in I,

$$A_i^t(X+P_{L^\perp}(X_0)) \le b_i.$$

Thus $(X+P_{L^\perp}(X_0))$ is in F_I for any X in L and any X_0 in F_I. This implies $P_L(F_I)$ is L and $(L \oplus P_{L^\perp}(F_I))$ is contained in F_I.

Now we show that $P_{L^\perp}(F_I)$ is $(F_I \cap S)$: Let Z be in F_I. Then for each i in I,

$$b_i = A_i^t Z$$

$$= A_i^t (P_L(Z) + P_{L^\perp}(Z))$$

$$= 0 + A_i^t P_{L^\perp}(Z).$$

Also, for each i not in I,

$$b_i \leq A_i^t Z = 0 + A_i^t P_{L^\perp}(Z).$$

Thus $P_{L^\perp}(Z)$ is in $(F_I \cap S)$ which implies that $P_{L^\perp}(F_I)$ is in $(F_I \cap S)$. Since $(F_I \cap S)$ is contained in F_I and is also contained in $L^\perp$, we have that $(F_I \cap S)$ is contained in $P_{L^\perp}(F_I)$. Thus $P_{L^\perp}(F_I)$ is $(F_I \cap S)$. q.e.d. Remark.)

Returning to the proof of the lemma, we note that by definition the dimension of F_I is the dimension of the linear manifold generated by F_I. Since F_I is $(L \oplus F_I^*)$ by the Remark, the dimension of F_I is the dimension of L plus the dimension of F_I^*. The codimension of F_I is defined to be p minus the dimension of F_I. This must be the dimension of $L^\perp$ minus the dimension of F_I^*, i.e. the codimension of F_I^* in $L^\perp$.

The inequalities which describe F_I are

$$\{X: \ A_i^t X \leq b_i, \text{ for } i = 1,\ldots,n; \ -A_j^t X \leq -b_j \text{ for } j \text{ in } I\}.$$

The possible nonsingular inequalities of F_I are of the form

$$A_i^t X \leq b_i$$

where i is not in I.

The inequalities which describe F_I are the same inequalities which describe F_I^* plus the restriction that the vectors of F_I^* lie in $L^{\perp}$. So the possible nonsingular inequalities for F_I^* are also those of F_I where the vectors are constrained to lie in $L^{\perp}$. But for i not in I

$$b_i > A_i^t X$$

$$= A_i^t (P_L(X) + P_{L^{\perp}}(X))$$

$$= A_i^t P_{L^{\perp}}(X).$$

Thus an inequality is nonsingular for F_I if and only if it is a nonsingular inequality for F_I^*. Thus a vector X is in the relative interior of F_I if and only if $P_{L^{\perp}}(X)$ is in the relative interior of F_I^*.

Definition 7. (Simplicial decomposition of a pointed cone). Define a simplicial cone of dimension d to be the pointed cone generated by d linearly independent vectors. A simplicial decomposition of a pointed cone of dimension d is the description of the pointed cone as the union of a finite number of simplicial cones of dimension d whose relative interiors are pairwise disjoint.

Remark. A simplicial decomposition exists for any pointed cone. (To see this, note that a pointed cone is generated by a finite number of vectors. Consider a hyperplane which passes through the interior of the cone and bisects it in such a fashion that the intersection of the cone and hyperplane is a pointed polyhedron whose vertices are among the generators of the cone. The pointed polyhedron would be a face of the polytope with vertex at zero created by the bisection. A simplicial decomposition of the polyhedron exists and generates a simplicial decomposition for the cone.)

LEMMA 8. *Let* F *be a convex set containing zero and suppose that* F *is contained in the convex set* K. *Let* N *be the linear* *subspace generated by* F. *If* $P_K(X)$ *is the projection of the vector* X *to* K *and* $P_K(X)$ *lies in the relative interior* F^I *of* F, *then* $P_K(X)$ *equals* $P_N(X)$, *the projection of* X *to* N.

Proof. Because F is contained in K and $P_K(X)$ lies in F, we have that $P_K(X)$ equals $P_F(X)$ the projection of X to F. Because zero is in F, the linear manifold generated by F is also the linear subspace generated by F. Clearly $P_N(X)$ lies in N. By the LEMMA of Stoer and Witzgall stated in the proof of Lemma 2 of this appendix, there is a positive scalar ε such that $(P_F(X)+\varepsilon(P_N(X)-P_F(X)))$ is in F. Then F contains

$$Z = P_F(X)+\varepsilon/(1+\varepsilon)(P_N(X)-P_F(X))$$

$$= 1/(1+\varepsilon)(P_F(X)+\varepsilon(P_N(X)-P_F(X)) + \varepsilon/(1+\varepsilon)P_F(X)$$

since F is convex.

Since F is contained in N, we have $P_F(X)$ equal to $P_F(P_N(X))$. But

$$||P_N(X)-Z||^2 = (1-\varepsilon/(1+\varepsilon))^2||P_N(X)-P_F(X)||^2,$$

which is less than $||P_N(X)-P_F(X)||^2$ unless $P_N(X)$ equals $P_F(X)$. This would be a contradiction to the definition of $P_F(X)$ (equal to $P_F(P_N(X))$) since it would imply that Z is closer to $P_N(X)$ than $P_F(X)$. Thus we must have $P_F(X)$ equal to $P_N(X)$.

LEMMA 9. *Let zero be contained in the convex set* K *which is contained in the convex set* Q. *Define* N *to be the linear subspace generated by* K *and let* $N^\perp$ *be its orthogonal complement space. Define*

$$C_K = N^\perp \cap Q^P$$

where Q^P *is the polar cone of* Q *and let* P_Q, P_N, *and* $P_{N^\perp}$ *be the projections to* Q, N, *and* $N^\perp$ *respectively. Then* $P_Q(X)$ *is in the* *relative interior* K^I *of* K *if and only if* $P_N(X)$ *is in* K^I *and* $P_{N^\perp}(X)$ *is in* C_K.

Remark. In the case that Q is a pointed convex polyhedron with a vertex at zero, then Q^p and Q^{pp} are convex polyhedral cones. As noted by H. P. Wynn [5], C_K is a face of Q^p whose dimension is equal to the codimension of K.

Proof. Assume that $P_N(X)$ is in K^I and $P_{N^\perp}(X)$ is in C_K. Then $P_N(X)$ is in Q^{pp} since K^I is in Q which is in Q^{pp}. Also $P_{N^\perp}(X)$ is in Q^p by the definition of C_K. Furthermore $P_N(X)$ and $P_{N^\perp}(X)$ are orthogonal since N is a linear subspace. By the LEMMA cited in the proof of Lemma 2, $P_N(X)$ equals $P_{Q^{pp}}(X)$ the projection of X to Q^{pp}. Because $P_N(X)$ is in Q, then $P_{Q^{pp}}(X)$ is in Q. Because Q is contained in Q^{pp} we have that $P_Q(X)$ equals $P_{Q^{pp}}(X)$. Thus $P_Q(X)$ equals $P_N(X)$ and is in K^I.

Assume that $P_Q(X)$ is in K^I. Then by Lemma 8, $P_Q(X)$ is $P_N(X)$. Thus $P_N(X)$ is in K^I. By Lemma 3, $(X-P_Q(X))$ is $P_{Q^p}(X)$. But $P_{N^\perp}(X)$ is $(X-P_N(X))$ which is $(X-P_Q(X))$. Thus $P_{N^\perp}(X)$ is in both N and Q^p, which implies that $P_{N^\perp}(X)$ is in C_K.

LEMMA 10. *Let* T *be an* r *dimensional normal random vector with mean* τ *and covariance* Σ. *Assume that* h *is a real-valued differentiable function defined on* $(0,\infty)$ *satisfying*

(*) $\lim_{t\to\infty} t^{\frac{1}{2}}h(t)\exp(-t/2) = 0$

(**) $\lim_{t\to 0} t^{\frac{1}{2}}h(t+c)\exp(-t/2) = 0$ *for any* $c > 0$.

Then

$$E[(\prod_{j=1}^{r} I_{(0,\infty)}(T_j))T^t\Sigma^{-1}(T-\tau)h(T^t\Sigma^{-1}T)]$$

$$= E[(\prod_{j=1}^{r} I_{(0,\infty)}(T_j))(rh(T^t\Sigma^{-1}T)+2h'(T^t\Sigma^{-1}T)T^t\Sigma^{-1}T)].$$

Proof. Let u be $(T-\tau)^t\Sigma^{-1}(T-\tau)$, and let v be $T^t\Sigma^{-1}T$. Note that

$$\frac{d}{dT_j}\{\exp(-\tfrac{1}{2}u)T_jh(v)\} = (\Sigma^{-1}(T-\tau))_j\exp(-\tfrac{1}{2}u)T_jh(v)$$

$$+ \exp(-\tfrac{1}{2}u)h(v)+2(\Sigma^{-1}T)_jT_j\exp(-\tfrac{1}{2}u)h'(v).$$

Thus

$$\left[\exp(-\tfrac{1}{2}u)T_jh(v)\right]_{T_j=0}^{T_j=\infty} = \int_0^\infty[-(\Sigma^{-1}(T-\tau))_jT_jh(v)+h(v)$$

$$+ 2(\Sigma^{-1}T)_jT_jh'(v)]\exp(-\tfrac{1}{2}u)dT_j.$$

Using (*) and (**) we have

$$\int_0^\infty(\Sigma^{-1}(T-\tau))_jT_jh(v)\exp(-\tfrac{1}{2}u)dT_j$$

$$= \int_0^\infty h(v)\exp(-\tfrac{1}{2}u)dT_j + \int_0^\infty 2(\Sigma^{-1}T)_jT_jh'(v)dT_j.$$

Now multiply both sides of the equation by $\prod_{i\neq j} I_{(0,\infty)}(T_i)$ and integrate both sides of the equation with respect to the other T_i's. Then summing over j yields the appropriate result.

REFERENCES

[1] James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* 1. Univ. of California Press, Berkeley, 361-379.

[2] Motzkin, T. (1936). *Beitrage zur Theorie der linearen Ungleichungen.* Azriel, Jerusalem.

[3] Sclove, S. L., Morris, C., and Radhakrishnan, R. (1972). Non-optimality of preliminary-test estimators for the mean of a multivariate normal distribution. *Ann. Math. Statist.* 43, 1481-1490.

[4] Stoer, J. and Witzgall, (1970). *Convexity and Optimizationn in Finite Dimensions I* Springer-Verlag, Berlin.

[5] Wynn, H. P. (1975). Integrals for one-sided confidence bounds: A general result. *Biometrika* 62, 393-396.