COMMUN. STATIST.-THEORY METH., 19(9), 3259-3266 (1990)

DISPERSION-DIMINISHING TRANSFORMATIONS

D. R. Jensen and D. E. Ramirez

Department of Statistics
Virginia Polytechnic Institute
Blacksburg, VA 24061

Department of Mathematics
University of Virginia
Charlottesville, VA 22903

Key Words. *Matrix ordering, congruence, monotone functions, dispersion reduction, linear inference*

0. ABSTRACT

Let $D(\Sigma)$ consist of matrices congruent to and dominated by a given matrix Σ, and let $\tau(\Sigma)$ be the corresponding congruent transformations. These classes are characterized and their properties studied when Σ is positive definite. Dispersion orderings are considered, including dispersion-diminishing linear transformations, concentration properties of which are shown. Arbitrary linear transformations are decomposed into contractions, isometries and dilations on subspaces relative to Mahalanobis norms. Applications are noted in statistical process control and linear inference.

1. INTRODUCTION

Let $(S_k^+, \geqslant)$ be the ordered cone of positive semidefinite real $(k \times k)$ matrices in which $\mathbf{A} \geqslant \mathbf{B}$ whenever $\mathbf{A} - \mathbf{B} \in S_k^+$, with $\mathbf{A} \succ \mathbf{B}$ when $\mathbf{A} - \mathbf{B}$ is positive definite. This ordering pervades much of statistics and applied mathematics, beginning with the work of Loewner (1934). The ordering is preserved under the general linear group $GL(k)$ acting on S_k^+ by congruence, because $\mathbf{A} \geqslant \mathbf{B}$ on $(S_k^+, \geqslant)$ if and only if $\mathbf{G\,A\,G}' \geqslant \mathbf{G\,B\,G}'$ for any $\mathbf{G} \in GL(k)$. Interest often focuses on the ordering of congruent pairs $(\mathbf{A}, \mathbf{G\,A\,G}')$; see Baksalary *et al.* (1983), for example. In this paper we consider the class $D(\Sigma) \subset S_k^+$ of matrices dominated by a given matrix $\Sigma \in S_k^+$, together with the class $\tau(\Sigma)$ of Σ-diminishing congruences such that $\Sigma \geqslant \mathbf{T}\Sigma\mathbf{T}'$ for any $\mathbf{T} \in \tau(\Sigma)$, with $\mathbf{T}$ in $GL(k)$ so that $\mathbf{T}$ is invertible. In particular, we characterize $\tau(\Sigma)$ in terms of singular decompositions, and we establish a basic connection between $D(\Sigma)$ and $\tau(\Sigma)$. General linear transformations on $\mathbb{R}^k$

are decomposed into operators effecting contractions, isometries and dilations on designated subspaces relative to Mahalanobis norms. Developments include the concentration properties of certain probability measures on $\mathbb{R}^k$, and applications are noted in linear inference and the control of multicharacteristic processes.

2. PRELIMINARIES

2.1 Notation. To fix notation $\mathbb{R}^k$ and $\mathbb{R}^k_+$ are Euclidean k-space and its positive orthant; $F_{n \times k}$ and S_k consist of real matrices of order $(n \times k)$ and their $(k \times k)$ symmetric varieties; and $O(k)$ is the real orthogonal group acting on $\mathbb{R}^k$. The linear span of $\{\mathbf{x}_1, \dots, \mathbf{x}_r\}$ in $\mathbb{R}^k$ is denoted by $\mathrm{Sp}(\mathbf{x}_1, \dots, \mathbf{x}_r)$. For $\mathbf{V} \succ 0$ in S_k^+, the squared Mahalanobis norm of $\mathbf{y} \in \mathbb{R}^k$ is $\|\mathbf{y}\|_V^2 = \mathbf{y}'\mathbf{V}^{-1}\mathbf{y}$.

2.2 Invariant Monotone Functions. A function $g{:}S_k^+ \to \mathbb{R}^1$ is *monotone* on $(S_k^+, \succcurlyeq)$ if $\mathbf{A} \succcurlyeq \mathbf{B}$ implies $g(\mathbf{A}) \geq g(\mathbf{B})$ on $\mathbb{R}^1$, and $g(\cdot)$ is *invariant under unitary congruence* if $g(\mathbf{P}\,\mathbf{A}\,\mathbf{P}') = g(\mathbf{A})$ for any $\mathbf{A} \in S_k^+$ and any real unitary $\mathbf{P} \in O(k)$. Let Φ comprise the functions monotone on $(S_k^+, \succcurlyeq)$; let functions in $\Phi_1 \subset \Phi$ be invariant under unitary congruence; and let $\Phi_0 \subset \Phi_1$ consist of unitarily invariant norms on S_k^+. The class Φ is characterized in Marshall *et al.* (1967), and functions in Φ_1 and Φ_0 arise through composition as follows. Let $\sigma(\mathbf{A}) = (\alpha_1, \dots, \alpha_k)$ map $\mathbf{A} \in S_k^+$ into its ordered eigenvalues $\{\alpha_1 \geq \cdots \geq \alpha_k \geq 0\}$, and let Γ_1 consist of functions on $\mathbb{R}^k$ that are symmetric under the $2^k k!$ reflections and permutations and that are increasing in each argument on $\mathbb{R}^k_+$. Then $\Phi_1 = \{\phi{:}\phi = \gamma\circ\sigma, \gamma \in \Gamma_1\}$, where $(\gamma\circ\sigma)(A) = \gamma(\alpha_1, \dots, \alpha_k)$. Similarly, $\Phi_0 = \{\phi{:}\phi = \gamma\circ\sigma, \gamma \in \Gamma_0\}$ are von Neumann's (1937) unitarily invariant norms restricted to S_k^+, where $\Gamma_0 \subset \Gamma_1$ are the symmetric gauge functions on $\mathbb{R}^k$. For further details see Jensen (1984), Jensen and Mayer (1977), Marshall and Olkin (1965) and Schatten (1970). Some examples follow.

Functions in Γ_0 include $\{\gamma_{(r)}(x_1, \dots, x_k) = \sum_{i=1}^{r} x_{[i]};\ 1 \leq r \leq k\}$, where $\{x_{[1]} \geq \cdots \geq x_{[k]}\}$ are the ordered values of $\{|x_1|, \dots, |x_k|\}$, and the ℓ_p norms $\{\gamma_p(x_1, \dots, x_k) = (\sum_{i=1}^{k} x_{[i]}^p)^{1/p};\ 1 \leq p < \infty\}$. Included are the ℓ_∞ norm $\gamma_{(1)}(x_1, \dots, x_k) = \max\{|x_1|, \dots, |x_k|\} = x_{[1]}$ and the Euclidean norm $\gamma_2(x_1, \dots, x_k) = (\mathbf{x}'\mathbf{x})^{1/2} = \|\mathbf{x}\|_I$ in earlier notation. Functions in Γ_1 but not Γ_0 include $\{\gamma(x_1, \dots, x_k; \lambda) = \prod_{i=1}^{k} x_{[i]}^{\lambda_i};\ 0 \leq \lambda_i < \infty, \lambda = \lambda_1 + \cdots + \lambda_k > 0\}$.

3. PROPERTIES OF $\tau(\Sigma)$ AND $D(\Sigma)$

For fixed $\Sigma \in S_k^+$, a transformation T acting on S_k^+ by congruence is said to be **Σ-*diminishing*** whenever $\Sigma \succcurlyeq T(\Sigma) = \mathbf{T}\Sigma\mathbf{T}'$, where $\mathbf{T}\Sigma\mathbf{T}'$ is the matrix representation of $T(\Sigma)$ with $\mathbf{T} \in GL(k)$. The following theorem characterizes, in terms of singular decompositions, the class $\tau(\Sigma)$ consisting of all such transformations.

THEOREM 1. The transformation $T: \Sigma \to \mathbf{T}\Sigma\mathbf{T}'$ is Σ-diminishing if and only if $\mathbf{T} = \Sigma^{1/2}\mathbf{W}'\Sigma^{-1/2}$, where $\mathbf{W} \in F_{k\times k}$ is some matrix whose singular values are bounded above by unity. In particular, the class $\tau(\Sigma)$ is given by

$$\tau(\Sigma) = \{\mathbf{T} : \mathbf{T} = \Sigma^{1/2}\mathbf{W}'\Sigma^{-1/2}; \|\mathbf{W}\| \leq 1\} \tag{3.1}$$

where $\|\mathbf{A}\| = \sup_{\mathbf{x}} \|\mathbf{A}\,\mathbf{x}\|_l / \|\mathbf{x}\|_l$.

Proof. Clearly $\Sigma - \mathbf{T}\Sigma\mathbf{T}' \succcurlyeq \mathbf{0}$ if and only if $\mathbf{I}_k - \mathbf{W}'\mathbf{W} \succcurlyeq \mathbf{0}$ with $\mathbf{W}' = \Sigma^{-1/2}\mathbf{T}\Sigma^{1/2}$. Apply the singular decomposition $\mathbf{W} = \mathbf{U}\Lambda\mathbf{V}$ such that $\mathbf{U}$ and $\mathbf{V}$ are orthogonal and $\Lambda = \mathrm{Diag}(\delta_1, \ldots, \delta_k)$ consists of the ordered singular values $\{\delta_1 \geq \cdots \geq \delta_k \geq 0\}$ of $\mathbf{W}$. It is clear that $\mathbf{I}_k - \mathbf{W}'\mathbf{W} \succcurlyeq \mathbf{0}$ if and only if $\mathbf{I}_k - \Lambda'\Lambda \succcurlyeq \mathbf{0}$, *i.e.*, $\mathbf{I}_k \succcurlyeq \mathrm{Diag}(\delta_1^2, \ldots, \delta_k^2)$. But this is equivalent to the assertions of the theorem, completing our proof. ■

Essential properties of $\tau(\Sigma)$ are summarized in the following.

COROLLARY 1. Let $\tau(\Sigma)$ be the class of Σ-diminishing congruences. Then

(i) $\tau(\Sigma)$ is closed under composition;

(ii) the n-fold composition $T^{(n)}(\Sigma)$ of $T \in \tau(\Sigma)$ has the matrix representation $\Sigma^{1/2}(\mathbf{W}')^n\Sigma^{-1/2}$;

(iii) $\phi(\Sigma) \geq \phi(\mathbf{T}\Sigma\mathbf{T}')$ for every $\mathbf{T} \in \tau(\Sigma)$ and $\phi \in \Phi$.

(iv) In particular, $\mathrm{tr}(\Sigma) \geq \mathrm{tr}(\mathbf{T}\Sigma\mathbf{T}')$ and $|\Sigma| \geq |\mathbf{T}\Sigma\mathbf{T}'|$.

Proof. (i) With $T_1, T_2 \in \tau(\Sigma)$, observe that the composition $T_2 \circ T_1$ has the representation $\Sigma^{1/2}\mathbf{W}_2'\mathbf{W}_1'\Sigma^{-1/2}$ in $\tau(\Sigma)$. Conclusion (ii) follows similarly. Conclusions (iii) and (iv) follow directly from the monotonicity of functions in Φ, completing our proof. ■

The following theorem establishes a basic connection between $\tau(\Sigma)$ and $D(\Sigma)$, together with further properties of the latter.

THEOREM 2. For each $\Sigma \in S_k^+$, the class $D(\Sigma)$ of matrices dominated by Σ on $(S_k^+, \succcurlyeq)$ has the following properties:

(i) $\Omega \in D(\Sigma)$ if and only if $\Omega = \mathbf{T}\Sigma\mathbf{T}'$ for some $\mathbf{T} \in \tau(\Sigma)$;

(ii) $\Omega \in D(\Sigma)$ if and only if $\Omega = \Sigma^{1/2}\mathbf{W}'\mathbf{W}\Sigma^{1/2}$ with $\mathbf{W}$ as in Theorem 1;

(iii) $\alpha\Sigma \in D(\Sigma)$ for $0 < \alpha \leq 1$;

(iv) $D(\Sigma)$ is convex; and

(v) $\Xi \preccurlyeq \Sigma$ implies $D(\Xi) \subset D(\Sigma)$.

Proof. (i) Suppose $\Omega \in D(\Sigma)$, so that $\Sigma \succcurlyeq \Omega$, and write $\Omega = \Omega^{1/2}\Sigma^{-1/2}\Sigma\Sigma^{-1/2}\Omega^{1/2}$. Because $\Sigma \succcurlyeq \Omega$, it follows that $\mathbf{I}_k \succcurlyeq \Sigma^{-1/2}\Omega\Sigma^{-1/2}$ as in the proof of Theorem 1, and thus $\mathbf{T} = \Omega^{1/2}\Sigma^{-1/2} \in \tau(\Sigma)$. Conversely, if $\mathbf{T} \in \tau(\Sigma)$, then $\mathbf{T}$ is Σ-diminishing and $\Omega = \mathbf{T}\Sigma\mathbf{T}' \preccurlyeq \Sigma$ from the definition of $\tau(\Sigma)$. Conclusion (ii) follows from (i) and Theorem 1. Conclusion (iii) is immediate. Convexity follows on considering Ξ, Ω in $D(\Sigma)$ and $\alpha = 1 - \bar{\alpha} \in [0, 1]$. Then $\Sigma - (\alpha\Xi + \bar{\alpha}\Omega) = \alpha(\Sigma - \Xi) + \bar{\alpha}(\Sigma - \Omega) \succcurlyeq \mathbf{0}$ and thus $\alpha\Xi + \bar{\alpha}\Omega \in D(\Sigma)$, giving conclusion (iv). Conclusion (v) may be seen on writing $\mathbf{T}\Sigma\mathbf{T}' = \Xi$, with $\mathbf{T} = \Sigma^{1/2}\mathbf{W}'\Sigma^{-1/2}$ such that $\mathbf{W}'\mathbf{W} = \Sigma^{-1/2}\Xi\Sigma^{-1/2}$, and using transitivity of the ordering on $(S_k^+, \succcurlyeq)$, thus completing our proof. ∎

4. DISPERSION-DIMINISHING MAPPINGS

Theorem 2 demonstrates that $\Omega \preccurlyeq \Sigma$, and $\Omega = T(\Sigma)$ for $T\in\tau(\Sigma)$, are equivalent. In this section we consider these relations in a probabilistic setting.

Let $\mathbf{Y}(\omega) \in \mathbb{R}^k$ be a random element on a probability space (Ω, B, P) having the mean $E(\mathbf{Y}) = \mathit{0}$ and dispersion matrix $V(\mathbf{Y}) = \Sigma \succ \mathbf{0}$. Linear transformations $\mathbf{Y} \rightarrow \mathbf{L}\,\mathbf{Y}$ on $\mathbb{R}^k$ induce the transformation $(\mathit{0}, \Sigma) \rightarrow (\mathbf{L}\,\mathit{0}, \mathbf{L}\Sigma\mathbf{L}')$ on $\mathbb{R}^k \times S_k^+$, acting by congruence on S_k^+. Because variance-diminishing scalings on $\mathbb{R}^1$ are given by $Y \rightarrow cY$ with $|c| < 1$, it is tempting to conjecture that reducing each marginal variance should diminish Σ, *i.e.*, that $\mathbf{Y} \rightarrow \mathbf{D}\,\mathbf{Y}$ on $\mathbb{R}^k$ should be Σ-diminishing whenever $\mathbf{D} = \mathrm{Diag}(\delta_1, \dots, \delta_k)$ with $\{|\delta_i| \le 1; 1 \le i \le k\}$. That this conjecture is false is seen for $k = 2$ on letting $\Sigma = [\sigma_{ij}]$ with $\sigma_{11} = 2, \sigma_{12} = \sigma_{21} = \sigma_{22} = 1$, and $\mathbf{D} = \mathrm{Diag}\,(0.1, 0.9)$. Here $\Sigma - \mathbf{D}\Sigma\mathbf{D}'$ has the determinant -0.4519, so that $\mathbf{D}$ is not Σ-diminishing despite a considerable reduction in the variance of the first component. Nonetheless, the class of all dispersion-diminishing linear transformations on $\mathbb{R}^k$ is found from Theorem 1 to be $\{\mathbf{Y} \rightarrow \mathbf{T}\,\mathbf{Y} : \mathbf{T} \in \tau(\Sigma)\}$. Moreover, $\phi[V(\mathbf{Y})] \ge \phi[V(T(\mathbf{Y}))]$ for every $T \in \tau(\Sigma)$ and $\phi \in \Phi$.

We seek connections between ordered scale parameters and the concentration of probabilities. To these ends consider scale families on $\mathbb{R}^k$ generated by a standard probability measure $\mu(\bullet)$ having unit scale, with dispersion matrix $\mathbf{I}_k$ when second moments are defined. Let $M(\mathbb{R}^k, B_k)$, with B_k as the Borel sets in $\mathbb{R}^k$, be the class of all such symmetric measures satisfying $\mu(A) = \mu(-A)$ for each $A\in B_k$ and $\mu \in M(\mathbb{R}^k, B_k)$; let $\mathbb{P}(\mu) = \{\mu_\Lambda(\bullet); \Lambda\in S_k^+\}$ such that $\mu_\Lambda(A) = \mu(\Lambda^{-1/2}A)$ in terms of the symmetric root $\Lambda^{1/2}$ of Λ, with $\mu \in M(\mathbb{R}^k, B_k)$; and let $\mathbb{P}_\Sigma(\mu) = \{\mu_\Omega(\bullet); \Omega\in D(\Sigma)\} \subset \mathbb{P}(\mu)$. Further let $\mathscr{H}(\Sigma)$ consist of sets of the type $A_\Sigma(c) = \{\mathbf{x} \in \mathbb{R}^k : \mathbf{x}'\Sigma^{-1}\mathbf{x} \le c\}$ with $c > 0$. Standard Chebyshev inequalities on $\mathbb{R}^k$ guarantee a lower bound for $\mu_\Omega(A)$ that is larger than for $\mu_\Sigma(A)$ for $A\in\mathscr{H}(\Sigma)$ whenever $\Sigma \succcurlyeq \Omega$, supporting the conjecture that $\mu_\Omega(\bullet)$ is more concentrated than $\mu_\Sigma(\bullet)$ on such sets. That this conjecture does hold is shown next, where the ordering $\mu_\Omega(A) \ge \mu_\Sigma(A)$ is shown precisely for every $A\in\mathscr{H}(\Sigma)$, and a converse is proved.

THEOREM 3. Let $\Omega, \Sigma\in S_k^+$ be given. The inequality $\mu_\Omega(A) \ge \mu_\Sigma(A)$ holds for every $A\in\mathscr{H}(\Sigma)$, and for every $\mu \in M(\mathbb{R}^k, B_k)$, if and only if $\mu_\Omega \in \mathbb{P}_\Sigma(\mu)$ with $\mu \in M(\mathbb{R}^k, B_k)$.

Proof. (i) Suppose $\mu_\Omega \in \mathbb{P}_\Sigma(\mu)$ for some fixed $\mu \in M(\mathbb{R}^k, B_k)$. Then $\Sigma \succcurlyeq \Omega = \mathbf{T}\Sigma\mathbf{T}' = T(\Sigma)$ with $T\in\tau(\Sigma)$. If $\mathscr{L}(\mathbf{X}) = \mu_\Sigma(\bullet)$ and $\mathscr{L}(\mathbf{Y}) = \mu_\Omega(\bullet)$, with $\mathbf{Y} = \mathbf{T}\,\mathbf{X}$, then we have $\mathbf{y}'\Sigma^{-1}\mathbf{y} = \mathbf{x}'\mathbf{T}'\Sigma^{-1}\mathbf{T}\,\mathbf{x} \le \mathbf{x}'\mathbf{T}'(\mathbf{T}\Sigma\mathbf{T}')^{-1}\mathbf{T}\,\mathbf{x} = \mathbf{x}'\Sigma^{-1}\mathbf{x}$, so that $\mathbf{X}\in A_\Sigma(c)$ implies

$\mathbf{Y} \in A_\Sigma(c)$ and thus $\mu_\Omega(A) \geq \mu_\Sigma(A)$ for every $A \in \mathscr{H}(\Sigma)$. Because $\mu \in M(\mathbb{R}^k, B_k)$ was chosen arbitrarily, the result holds for every $\mu \in M(\mathbb{R}^k, B_k)$.

(ii) To show that $\mu_\Omega \in \mathbb{P}_\Sigma(\mu)$ is necessary, it suffices to assume that $k = 2$, $\Sigma = \mathbf{I}_2$, and $\Omega = \mathrm{Diag}(\lambda_1, \lambda_2)$. By way of contradiction, suppose $\lambda_1 > 1$. Then the set $E = A(c) \cap \Omega^{-1/2}A(c)$ is not empty. For any measure $\nu \in M(\mathbb{R}^2, B_2)$ with support in E, we have $\nu_\Sigma(A(c)) = 1$ while $\nu_\Omega(A(c)) = 0$, the required contradiction which completes our proof. ■

To continue, the inclusion arguments of Theorem 3 may be interpreted in terms of an operator T on the metric space $(\mathbb{R}^k, \|\cdot\|_\Sigma)$ as follows. Because $\mathbf{0} \prec \Sigma^{-1} \preccurlyeq (\mathbf{T}\Sigma\mathbf{T}')^{-1}$ for $T \in \tau(\Sigma)$, we infer that $\|\mathbf{y}\|_\Sigma^2 \leq \|\mathbf{y}\|_{T\Sigma T'}^2 = \|\mathbf{T}^{-1}\mathbf{y}\|_\Sigma^2$, *i.e.*, $\|\mathbf{T}\,\mathbf{x}\|_\Sigma^2 \leq \|\mathbf{x}\|_\Sigma^2$. It follows that every $T \in \tau(\Sigma)$ is a contraction on $(\mathbb{R}^k, \|\cdot\|_\Sigma)$. Similar arguments show that if $T \notin \tau(\Sigma)$ and $\mathbf{T}\Sigma\mathbf{T}' \succcurlyeq \Sigma$, then T is a dilation on $(\mathbb{R}^k, \|\cdot\|_\Sigma)$, *i.e.*, $\|\mathbf{T}\,\mathbf{x}\|_\Sigma^2 \geq \|\mathbf{x}\|_\Sigma^2$. Generally T is neither a contraction nor a dilation on $(\mathbb{R}^k, \|\cdot\|_\Sigma)$, and its properties remain to be studied. We next decompose $\mathbb{R}^k$ into subspaces (F_1, F_2, F_3) in which T is a contraction, an isometry, and a dilation, respectively.

To these ends consider a canonical form in which $(\Sigma, \mathbf{T}\Sigma\mathbf{T}') \rightarrow (\mathbf{I}_k, \mathbf{W}'\mathbf{W})$ with $\mathbf{W}' = \Sigma^{-1/2}\mathbf{T}\Sigma^{1/2}$; write the spectral equations $\{(\mathbf{W}'\mathbf{W})^{-1}\mathbf{u}_i = \gamma_i\mathbf{u}_i;\ 1 \leq i \leq k\}$; let $\{\mathbf{x}_i = \Sigma^{1/2}\mathbf{u}_i;\ 1 \leq i \leq k\}$; and suppose that neither $\mathbf{W}'\mathbf{W} \preccurlyeq \mathbf{I}_k$ nor $\mathbf{W}'\mathbf{W} \succcurlyeq \mathbf{I}_k$. Then at least one of two integers (r, s) exist such that

$$\{\gamma_1 \geq \cdots \geq \gamma_r > \gamma_{r+1} = 1 = \cdots = \gamma_{r+s} > \gamma_{r+s+1} \geq \cdots \geq \gamma_k > 0\}.$$

Accordingly, decompose $(\mathbb{R}^k, \|\cdot\|_{I_k})$ into $\mathbb{R}^k = V_1 \oplus V_2 \oplus V_3$ such that $V_1 = \mathrm{Sp}(\mathbf{u}_1, \ldots, \mathbf{u}_r)$, $V_2 = \mathrm{Sp}(\mathbf{u}_{r+1}, \ldots, \mathbf{u}_{r+s})$ and $V_3 = \mathrm{Sp}(\mathbf{u}_{r+s+1}, \ldots, \mathbf{u}_k)$. Further let $E_1 = \Sigma^{1/2}V_1$, $E_2 = \Sigma^{1/2}V_2$ and $E_3 = \Sigma^{1/2}V_3$, and finally let $F_1 = T^{-1}(E_1)$, $F_2 = T^{-1}(E_2)$ and $F_3 = T^{-1}(E_3)$. A principal result is the following.

THEOREM 4. Let T be a nonsingular linear transformation on $(\mathbb{R}^k, \|\cdot\|_\Sigma)$, and consider subspaces $\{F_1, F_2, F_3\}$ of $\mathbb{R}^k$ as defined. Then

(i) T is a contraction on $(F_1, \|\cdot\|_\Sigma)$, *i.e.*, $\|\mathbf{T}\,\mathbf{z}\|_\Sigma^2 \leq \|\mathbf{z}\|_\Sigma^2$ for $\mathbf{z} \in F_1$;

(ii) T is isometric on $(F_2, \|\cdot\|_\Sigma)$, *i.e.*, $\|\mathbf{T}\,\mathbf{z}\|_\Sigma^2 = \|\mathbf{z}\|_\Sigma^2$ for $\mathbf{z} \in F_2$; and

(iii) T is a dilation on $(F_3, \|\cdot\|_\Sigma)$, *i.e.*, $\|\mathbf{T}\,\mathbf{z}\|_\Sigma^2 \geq \|\mathbf{z}\|_\Sigma^2$ for $\mathbf{z} \in F_3$.

Proof. We provide details for conclusion (i), the other conclusions following similarly. On writing $\{(\mathbf{W}'\mathbf{W})^{-1}\mathbf{u}_i = \gamma_i\mathbf{u}_i; \gamma_i > 1\}$ and $\{\mathbf{x}_i = \Sigma^{1/2}\mathbf{u}_i;\ 1 \leq i \leq r\}$, we infer that $\Sigma^{1/2}V_1 = \mathrm{Sp}(\mathbf{x}_1, \ldots, \mathbf{x}_r) \subset \mathbb{R}^k$. Moreover, we compute

$$\begin{aligned}\|\mathbf{x}_i\|_\Sigma^2 &= \mathbf{x}_i'\Sigma^{-1}\mathbf{x}_i = \mathbf{u}_i'\mathbf{u}_i \leq \gamma_i\mathbf{u}_i'\mathbf{u}_i = \mathbf{u}_i'(\mathbf{W}'\mathbf{W})^{-1}\mathbf{u}_i \\ &= \mathbf{x}_i'\Sigma^{-1/2}(\mathbf{W}'\mathbf{W})^{-1}\Sigma^{-1/2}\mathbf{x}_i = \mathbf{x}_i'(\mathbf{T}\Sigma\mathbf{T}')^{-1}\mathbf{x}_i = \|\mathbf{x}_i\|_{T\Sigma T'}^2 = \|\mathbf{T}^{-1}\mathbf{x}_i\|_\Sigma^2.\end{aligned}$$

This inequality passes to finite linear combinations of $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ since for $i \neq j$, $\mathbf{x}_i'\Sigma^{-1}\mathbf{x}_j = \mathbf{u}_i'\mathbf{u}_j = 0$ and $\mathbf{x}_i'(\mathbf{T}\Sigma\mathbf{T}')^{-1}\mathbf{x}_j = \gamma_j\mathbf{u}_i'\mathbf{u}_j = 0$. This shows that T^{-1} is a dilation on $\Sigma^{1/2}V_1$, and thus T is a contraction on $(F_1, \|\cdot\|_\Sigma)$, as claimed. This completes our proof. ■

5. APPLICATIONS

The foregoing developments bear heavily on statistical practice in significant fields of application. Details pertaining to two such fields follow.

5.1 Statistical Process Control. Ongoing procedures are usually in place for tightening the process variance in statistical process control. Such tightening results not only in greater homogeneity of product, but also in greater power of Shewhart's (1931) $\bar{X}$ charts to detect a given shift in the mean of a single quality characteristic.

Benefits of tightening a multicharacteristic process are less clear. Mean vectors of such processes are routinely monitored using Hotelling's (1947) T^2 charts, where $T^2(\Sigma)$ identifies the T^2 statistic from a process having dispersion matrix Σ. Consider two different processes having a common mean vector but different dispersion matrices, Σ and Ω. In order that $T^2(\Omega)$ should be uniformly more powerful than $T^2(\Sigma)$, it is necessary and sufficient that $\Sigma \succcurlyeq \Omega$. If tightening a process serves merely to reduce each marginal variance, then such tightening need not be Σ-diminishing, as noted in Section 3. Thus tightening each marginal variance does not necessarily give greater power to T^2 charts. On the other hand, Theorems 1 and 2 characterize the class $D(\Sigma)$ of dispersion matrices resulting in uniformly greater power than $T^2(\Sigma)$ to detect a given shift in the means when the process is not in control. Thus, by Theorem 2, the matter of tightening a multicharacteristic process is tantamount to discovering mechanisms for adjusting the process so as to emulate one of the Σ-diminishing transformations of Theorem 1.

5.2 Experimental Design. The choice of $\mathbf{X} \in F_{n\times k}$ determines the dispersion matrix $V[\hat{\beta}(\mathbf{X})] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ of the Gauss-Markov estimator $\hat{\beta}(\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ in a linear model $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$. Here $\mathbf{y} \in \mathbb{R}^n$, $n > k$, and $\mathbf{e} \in \mathbb{R}^n$ is random having $E(\mathbf{e}) = \mathbf{0}$ and $V(\mathbf{e}) = \sigma^2\mathbf{I}_n$. Various criteria have been advanced for choosing among designs, including $C_A(\mathbf{X}) = \mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1}$, $C_D(\mathbf{X}) = |\mathbf{X}'\mathbf{X}|^{-1}$ and $C_E(\mathbf{X}) = \xi_1$, with $\{\xi_1 \geq \cdots \geq \xi_k > 0\}$ as the ordered eigenvalues of $(\mathbf{X}'\mathbf{X})^{-1}$. These, in turn, support notions of A-optimal, D-optimal and E-optimal designs within a given class. It is well known that different criteria may entail different optimal designs.

Suppose that $\mathbf{X} \in F_{n\times k}$ may be redesigned as $\mathbf{X} \to \mathbf{X}\,\mathbf{G}$ for some $\mathbf{G} \in GL(k)$, resulting in the model $\mathbf{u} = \mathbf{Z}\beta + \mathbf{e}$ with $\mathbf{Z} = \mathbf{X}\,\mathbf{G}$. Then $\mathbf{G} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}$, and $\mathbf{X}$ and $\mathbf{Z}$ have the same column span. Conversely, if $\mathbf{X}$ and $\mathbf{Z}$ have the same column span, then with $\mathbf{G} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}$, we have that $\mathbf{X}\,\mathbf{G} = [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Z} = \mathbf{Z}$, since $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the projection matrix for $\mathbf{X}$.

The Gauss-Markov estimators using designs $\mathbf{X}$ and $\mathbf{Z}$ are $\hat{\beta}(\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $\hat{\beta}(\mathbf{Z}) = \mathbf{G}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$, respectively. We can assure that the design $\mathbf{Z}$ dominates $\mathbf{X}$ by choosing $\mathbf{G}$ so that $\mathbf{G}^{-1} = \mathbf{T} \in \tau((\mathbf{X}'\mathbf{X})^{-1})$ since Section 4 shows that $V[\hat{\beta}(\mathbf{X})] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \succcurlyeq \sigma^2\mathbf{T}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{T}' = \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1} = V[\hat{\beta}(\mathbf{Z})]$. Moreover, because $C_A(\mathbf{X})$,

$C_D(\mathbf{X})$ and $C_E(\mathbf{X})$ are all in Φ, it follows from Corollary 2 that modifying $\mathbf{X} \to \mathbf{X}\,\mathbf{G}$, with $\mathbf{G}^{-1} \in \tau((\mathbf{X}'\mathbf{X})^{-1})$, simultaneously diminishes all three criteria $C_A(\mathbf{X})$, $C_D(\mathbf{X})$ and $C_E(\mathbf{X})$. These facts are helpful in comparing alternative designs and in improving existing designs.

5.3 Numerical Example. We next consider a numerical example in which $\mathbf{X}$ and $\mathbf{Z}$ have different column spans, there being no transformation $\mathbf{G}$ mapping $\mathbf{X}$ to $\mathbf{Z}$. Nonetheless, we are able to compare the performance characteristics of the two designs and to demonstrate a Σ-diminishing congruence as in Theorem 1. The model we study is a second-order response function having linear and pure quadratic effects, *i.e.*,

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + e \tag{5.1}$$

in the variables X_1 and X_2. Take $\mathbf{X}$ to be the matrix

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 0 & 1 & 0 \\ 1 & -1 & 1 & 1 & 1 \\ 1 & 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{5.2}$$

Next rotate points in the (X_1, X_2)-space clockwise through 45 degrees to form in the (Z_1, Z_2)-space the design matrix

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & -\sqrt{2} & 0 & 2 \\ 1 & -\sqrt{1/2} & -\sqrt{1/2} & 1/2 & 1/2 \\ 1 & -\sqrt{2} & 0 & 2 & 0 \\ 1 & \sqrt{1/2} & -\sqrt{1/2} & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & -\sqrt{1/2} & \sqrt{1/2} & 1/2 & 1/2 \\ 1 & \sqrt{2} & 0 & 2 & 0 \\ 1 & \sqrt{1/2} & \sqrt{1/2} & 1/2 & 1/2 \\ 1 & 0 & \sqrt{2} & 0 & 2 \end{bmatrix}. \tag{5.3}$$

The covariance structures for designs $\mathbf{X}$ and $\mathbf{Z}$ are $V[\hat{\beta}(\mathbf{X})] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \Sigma$, and $V[\hat{\beta}(\mathbf{Z})] = \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1} = \Omega$. To compare the two designs we compute the eigenvalues of $\Sigma^{1/2}\Omega^{-1}\Sigma^{1/2}$ which are $\{4, 1, 1, 1, 1\}$. Thus $\Sigma \succcurlyeq \Omega$, and the Gauss-Markov estimators $\hat{\beta}(\mathbf{Z})$ are uniformly more efficient than $\hat{\beta}(\mathbf{X})$. It follows from Theorem 1 that there is a Σ-diminishing congruence $\mathbf{T}$ taking $(\mathbf{X}'\mathbf{X})^{-1}$ into $(\mathbf{Z}'\mathbf{Z})^{-1}$ given by $\mathbf{T} = \Sigma^{1/2}\mathbf{W}'\Sigma^{-1/2}$ with $\mathbf{W}'\mathbf{W} = \Sigma^{-1/2}\Omega\Sigma^{-1/2}$. We set $\mathbf{W} = (\Sigma^{-1/2}\Omega\Sigma^{-1/2})^{1/2}$ and compute $\mathbf{T} = \Sigma^{1/2}\mathbf{W}'\Sigma^{-1/2}$ to be

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 0 & 1/4 & 3/4 \end{bmatrix}. \tag{5.4}$$

BIBLIOGRAPHY

Baksalary, J. K., Kala, R. and Klaczyński, K. (1983). "The matrix inequality $M \geq B^*MB$." *Linear Algebra and Its Applications* 54: 77-86.

Hotelling, H. (1947). "Multivariate quality control, illustrated by the air testing of sample bombsights." In *Techniques of Statistical Analysis,* C. Eisenhart *et al.*, eds. McGraw-Hill, New York.

Jensen, D. R. (1984). "Invariant ordering and order preservation." In *Inequalities in Statistics and Probability,* Y. L. Tong, ed., pp. 26-34. IMS Lecture Notes/Monograph Series, Volume 5, Hayward, CA.

Jensen, D. R. and Mayer, L. S. (1977). "Some variational results and their applications in multiple inference." *Ann. Statist.* 5: 922-931.

Loewner, C. (1934). "Über monotone Matrix Funktionen." *Math. Z.* 38: 177-216.

Marshall, A. W. and Olkin, I. (1965). "Norms and inequalities for condition numbers." *Pacific J. Math.* 15: 241-247.

Marshall, A. W., Walkup, D. W. and Wets, R. J.-B. (1967). "Order-preserving functions; applications to majorization and order statistics." *Pacific J. Math.* 23: 569-584.

Schatten, R. (1970). *Norm Ideals of Completely Continuous Operators.* Springer-Verlag, New York.

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product.* Van Nostrand, New York.

von Neumann, J. (1937). "Some matrix inequalities and metrization of matrix space." *Tomsk Univ. Rev.* 1: 286-300.

Received July 1989; Revised July 1990.

Recommended by R. P. Gupta, Dalhousie University, CANADA.

Refereed Anonymously.