On directed alternatives in linear inference

Abstract

Tests for vector hypotheses $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ in $\mathbb{R}^k$ typically have powers depending on quadratic forms of type $A = (\theta - \theta_0)^\top \Sigma^{-1} (\theta - \theta_0)$. This study examines the case that $(\mu - \mu_0)$ is restricted to subspaces, for example, $(\mu - \mu_0)^\top = c (1,1,0,...0)$ differing only in their first two coordinates. These are called directed alternatives. The spectral decomposition of $\Sigma$ supports the identification of one-dimensional alternatives least likely and most likely to be discerned, to complement conventional data analysis. Applications are drawn in the use of Hotelling’s $T^2$ and of $F$-tests in linear inference. Moreover, it is seen that a given design may be recast so as to reverse the least likely and most likely alternatives. Numerical examples serve to illustrate the findings.

Keywords: linear models; $F$ tests, hotelling’s $T^2$ tests, directed alternatives, reversal designs

Introduction

Power in statistical inference is driven by non-null distributions. For observations in $\mathbb{R}^k$ having dispersion matrix $\Sigma$, noncentrality parameters often emerge as the Mahalanobis’ distance between points $(u,v)$ in $\mathbb{R}^k$, namely,

$$D^2_{\Sigma} (u,v) = (u-v)^\top \Sigma^{-1} (u-v).$$

(1)

This specializes to the Euclidean metric for the case that $\Sigma = c I_k$, in which case the model is called isotropic. In particular, nonparametric and other statistics often have noncentral chi-squared distributions, either in small samples or asymptotically. In addition, pervasive venues in parametric inference, to be reexamined in some detail, include the following. Case (i). Hotelling’s $T^2$ Test: $T^2 = n (\bar{Y} - \mu_0)^\top S^{-1} (\bar{Y} - \mu_0)$ where $(\bar{Y}, S)$ are the sample mean and dispersion matrix of $n$ Gaussian vectors in $\mathbb{R}^k$ having the location–scale parameters $\mu_0, \Sigma$. Then in testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, the power function is $\Psi (\lambda)$ with noncentrality $\lambda = n (\mu - \mu_0)^\top \Sigma^{-1} (\mu - \mu_0)$.

Case (ii). The General Linear Model: $\{ \bar{Y} = X \beta + e \}$ with Gaussian errors having zero means and dispersion matrix $\Sigma^2 I_k$. Then in testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ in $\mathbb{R}^k$, the power function is $\Psi (\lambda)$ with noncentrality $\lambda = (\beta - \beta_0)^\top X'X (\beta - \beta_0) / \sigma^2$.

Classical theory allows for any $(\mu - \mu_0) \in \mathbb{R}^k$ for Case (i), and $(\beta - \beta_0) \in \mathbb{R}^k$ for Case (ii). On the other hand, alternatives lying in designated subspaces may hold substantive interest per se. For example, taking $(\mu - \mu_0)^\top = c (1,1,0,...0)$ allows for discrepancies between $\mu$ and $\mu_0$ in their first two coordinates only, whereas $(\mu - \mu_0)^\top = c (1,1,...,1)$ allows for deviations along the equiangular line in $\mathbb{R}^k$. Both are one-dimensional; subspaces of dimension greater than one are considered subsequently. Alternatives lying in designated subspaces of $\mathbb{R}^k$ are called directed alternatives, and the goal here is to study powers of tests against alternatives of these types.

The present study expands on this as follows. Not only do distinct alternatives differ in importance to users, but so too their probabilities of detection. Here the spectral decomposition of $\Sigma$, if anisotropic, supports the identification of alternatives least likely and most likely to be discovered, as well as intermediate cases. These serve to bracket the effective range of inferences intrinsic to a given study, and thereby complement conventional options in data analysis. Applications are drawn in the use of Hotelling’s $T^2$ in multivariate samples, and of $F$-tests in the analysis of linear models. Moreover, it is shown that a given design may be modified so as to reverse the least likely and most likely alternatives, in the event that this would better serve the objectives of an experiment.

This study is organized as follows. Supporting developments are given next in Section 2, followed by the principal findings of Section 3. Several examples in Section 4 illustrate the essential results. Collateral materials are deferred for completeness to an Appendix.

Preliminaries

Notation

Spaces include $\mathbb{R}^n$ as Euclidean $n$-space; $\mathbb{R}^n_+$ as its positive orthant; $\mathbb{S}_n$ as the real symmetric $(n \times n)$ matrices; $\mathbb{S}_n^+$ as their positive definite varieties; $\mathbb{W}_{n\times k}$ as the real $(n \times k)$ matrices of rank $k \leq n$; and $\mathbb{O}_k$ as the $(k \times k)$ orthogonal group. Vectors and matrices are set in bold type; the transpose, inverse, trace, and determinant of $A$ are $A'$, $A^{-1}$, $tr \ (A)$, and $|A|$; the unit vector in $\mathbb{R}^n$ is $1_n = [1,1,...,1]^\top$; $I_n$ is the $(n \times n)$ identity; and $\text{Diag} \ (\alpha_1,...,\alpha_n)$ is a block-diagonal array. If $B = [b_1, ..., b_k]$ is of order $(n \times k)$ and rank $k < n$, then $S_B (B)$ designates the column span of $B$, i.e., the $k$–dimensional subspace of $\mathbb{R}^n$ spanned by $b_1, ..., b_k$. The ordered eigenvalues of $A \in \mathbb{R}^n_+$ are $\lambda_i (A) = \alpha_i; 1 \leq i \leq n$ with $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_n$, and its spectral decomposition is $A = P D_A P' = \sum_{i=1}^n \alpha_i p_i p_i'$, where $P = [p_1, ..., p_n] \in \mathbb{O}_n$ and $D_A = \text{Diag} \ (\alpha_1, ..., \alpha_n)$. By convention its condition number is $C_A (A) = \alpha_1 / \alpha_n$. The singular decomposition
of \( B \in \mathbb{F}_{n \times k} \) is \( B = PDQ = \sum_{j=1}^{k} \delta_j p_j q_j' \), where the mutually orthogonal columns of \( P = [p_1 \ldots p_k] \) comprise the left-singular vectors; \( D = \text{Diag}(\delta_1, \ldots, \delta_k) \) are its singular values; and columns of \( Q = \mathcal{O}_k \) are the right-singular vectors.

### Special distributions

For \( Y \in \mathbb{R}^n \), its distribution, mean, and dispersion matrix are \( \Lambda(Y), \mu(Y), \Sigma(Y) = \Sigma \), say, with variance \( \text{Var}(Y) = \sigma^2 \) on \( \mathbb{R}^n \). Specifically, \( \Lambda(Y) = \mathcal{N}_n(\mu, \Sigma) \) is Gaussian on \( \mathbb{R}^n \) with parameters \((\mu, \Sigma)\). Distributions on \( \mathcal{R}^n \) include the \( \chi^2(n, \lambda) \) with \( n \) degrees of freedom and noncentrality parameter \( \lambda \); the Snedecor–Fisher \( F(n_1, n_2, \lambda) \) with degrees of freedom \((n_1, n_2)\) and noncentrality \( \lambda \); and Hotelling's \( T^2_k(n, \lambda) \) of order \( k \) having \( n \) degrees of freedom and noncentrality \( \lambda \). Recall that \( \Lambda(Y) \) increases stochasticity with increasing \( \lambda \) with other parameters held fixed. Identify \( \sigma \) in context as the upper \( \alpha \)-level rejection rule. The power of a test, to be considered as a function of \( \lambda \), is designated by \( \psi(\lambda) \).

### The principal findings

#### Directed alternatives

Our notation encompasses both (i) Hotelling's \( T^2 \) and (ii) General Linear Models, having location–scale parameters \((\delta, \Xi)\) What distinguishes this study are directed alternatives \( \{ \theta_j;1 \leq j \leq k \} \) aligned with the orthonormal eigenvectors \( Q = [q_1 \ldots q_k] \) of \( \Xi^{-1} \), thus standardized to unit lengths. To continue, as \( \Omega = \Xi^{-1} \) assumes a central role, take

\[
\Omega = \sum_{i=1}^{k} \kappa_i q_i q_i' \in \mathbb{R}^{n \times n}
\]

as its spectral decomposition, with \( \kappa_{i \geq \cdots \geq \kappa_0 > 0} \). As in Appendix A.1, undertake the expansions

\[
\Omega = Q \delta D \Omega' + \sum_{j=1}^{k} \Omega D \Omega' \sum_{i=1}^{k-j} Q D \Omega' = \sum_{j=1}^{k} \delta_j Q D \Omega', \quad j \text{ repeated } s \text{ times},
\]

where elements of \( Q = [q_0, q_1, \ldots, q_k] \) are of orders \( |k \times (r-1)| \) and \( |k \times s|\) for \( d = k - r - s + 1 \), and where \( D = \text{Diag}(D_{11}, D_{12}, D_{13}) \) is partitioned conformably. In regard to quadratic forms of type \( Q(u) = u' \Omega u \) serving as noncentrality parameters, a principal result is the following.

**Theorem 1.** Given is a location–scale model with parameters \((\delta, \Xi)\), together with a test for \( H_0: \delta = \delta_0 \) against \( H_1: \delta \neq \delta_0 \) having power \( \Psi(\lambda) \) increasing monotonically with \( \lambda = \lambda^2(\delta, \delta_0) = [\delta - \delta_0, \Xi^{-1}(\delta - \delta_0)] \) in succession as the eigenvectors \( \{\theta_j; 1 \leq j \leq k\} \) of \( \Xi^{-1} \) with eigenvalues \( \kappa_{1 \geq \cdots \geq \kappa_0 > 0} \).

i. Then powers \( \Psi(\lambda_j) \) of the test at alternatives \( \{[\delta - \delta_0, \Xi^{-1}(\delta - \delta_0)] = \theta_j; 1 \leq j \leq k\} \) depend on the noncentrality parameters \( \lambda_j = \kappa_j; 1 \leq j \leq k \), respectively.

ii. In particular, the alternatives most likely and least likely to be discerned in terms of power are \( \theta_1 \) and \( \theta_k \), having powers \( \Psi(\kappa_1) \) and \( \Psi(\kappa_k) \), respectively.

iii. Consider alternatives \( \{y_j \in \mathcal{S}(\Omega)\} \) standardized to unit lengths. Then bounds on powers at these local alternatives are given by \( \Psi(\kappa_{ij}) \leq \Psi(\lambda_j) \leq \Psi(\kappa_{ij}) \) for every \( y_j \in \mathcal{S}(\Omega) \). The power of a test, \( t(\theta) = \theta' \Xi^{-1} \theta \), is the noncentrality parameter \( \lambda_j = \kappa_j a_j' a_j \), with corresponding power \( \Psi(\lambda_j) \).

**Proof:** Conclusion (i) follows directly from \( \lambda(\theta) = \theta' \Xi^{-1} \theta = \theta' \sum_{i=1}^{k} q_i q_i' \theta = \kappa_j \) since \( \theta = q_j \), whereas \( q_j' q_j = 0 \) and \( q_j' q_j = 1 \) by orthonormality. Conclusion (ii) follows directly from variational properties of Rayleigh quotients as in Lemma A.1(i) of the Appendix. In like manner conclusion (iii) follows from Lemma A.1(ii) as variational properties over subspaces. Conclusion (iv) follows from (iii) since \( q_j' q_j = 0 \).

**Remark 1.** The directed alternatives \( \delta_1 = [1, 1, 0, \ldots, 0] \) and \( \delta_2 = [1, 1, \ldots, 1] \) were featured earlier as discrepancies in the first two coordinates of \( \delta - \delta_0 \), and as deviations about the equiangular line in \( \mathbb{R}^k \). Let \( \Xi^{-1} = \delta_0' \delta_0^{-1} \). Then powers \( \Psi(\lambda) \) at these alternatives will depend on \( \lambda_1 = \sum_{i=1}^{k} \sum_{j=1}^{k} \kappa_{ij} a_i a_j' \) at \( \delta_1 = [1, 1, \ldots, 1] \) and on \( \lambda_2 = \sum_{i=1}^{k} \sum_{j=1}^{k} \kappa_{ij} a_i a_j' \) at \( \delta_2 = [1, 1, \ldots, 1] \).

**Corollary 1.** On specializing the location–scale parameters \((\delta, \Xi)\), Theorem 1 applies verbatim as follows.

- (i) Hotelling's \( T^2: (\delta, \Xi) = (\mu, \Sigma) \); \( L(T^2) = T^2 \) (\( n - 1, \lambda \)), the power \( \Psi(\lambda) \) depending on \( \lambda = n(\mu - \mu_0) \Sigma^{-1}(\mu - \mu_0) \).

- (ii) General Linear Models:

\[
(\delta, \Xi) = [\beta, X'X]; L(F) = F(\mu, \Sigma), \quad \text{the power } \Psi(\lambda) \text{ depending on } \lambda = [\beta - \beta_0]' X'X(\beta - \beta_0) / \sigma^2.
\]

**Proof:** The noncentral distribution \( F(\mu, \Sigma) \) clearly satisfies the assumptions of Theorem 1 on identifying \( (\delta, \Xi) = [\beta, X'X] \) as claimed. Similarly in testing \( H_0: \mu = \mu_0 \) against \( H_1: \mu = \mu_0 \), Hotelling's \( T^2 \) inherits these properties through the conversion
On directed alternatives in linear inference

are reversed from those of
and with
with inverse
I
as its singular

= 1
where it often suffices to take
tests

1
with

as

and exchangeable

= \{q_j\}_{j=1}^k

) and exchangeable

) calcium. The units

I
, so that
is anisotropic with

jk

∑

ν
( )

2
may be selected in like manner as most likely
ote, however, that these tools are
, i.e., the model is isotropic. Sample evidence regarding
as the sample dispersion matrix from n observations,
against
i
.A
least likely with
with

3
.

of unit lengths
as the residual mean square and with noncentrality
. Our principal reconstruction is articulated in the
utilizes
n addition, any pairs
i ii
.i
us far are predicated in part on the desirability
Variations on
δ δδ −′ ′′= = ++

Σ
is cal
, then the noncentrality parameters

2
as its left–singular vectors,
See,
. If instead the directed
q
jk
. Next construct
and

with
q
be a permutation operator reversing the
Let
Note that the foreg
are realigned.
The most likely and least likely alternatives for design
for example, Re
upper percentile of the central distribution
taking

such that pairs \( \{ \hat{\delta}_j, q_j \} \) and \( \{ \hat{\delta}_i, q_i \} \) are realigned.
Conclusion: The most likely and least likely alternatives for design
X
are reversed from those of
X
, so that \( \theta = q \) now is most likely
with power \( \Psi(\kappa) \) depending on \( \kappa \), and \( \theta = q \) least likely with
power \( \Psi(\kappa) \) depending on \( \kappa \).
Proof: Clearly the conventional reordering of eigenvalues gives

\[
X'X = k\kappa_q'q + \sum_{i=2}^{k-1} \kappa_i q_i'q_i + \kappa_k q_k'q_k
\]
and the conclusion follows on applying Theorem 1(ii) in the context of Corollary 1(ii).
Remark 4. Variations on \( X_x \) are apparent. Any permutation of
\( \{ \hat{\delta}_2, \ldots, \hat{\delta}_4 \} \) gives the same conclusion. In addition, any pairs
\( \{ \hat{\delta}_j, q_j \} \) may be selected in like manner as most likely
and least likely to be discerned. Note, however, that these tools are
available in the case of first–order designs.

Case studies

Studies in Hotelling’s \( T^2 \) and second–order response models are
given, to illustrate Theorem 1 and Corollary 1. Moreover, an example
design serves to illustrates the Theorem 2 reversal of most likely and
least likely alternatives.

Hotelling’s \( T^2 \) tests

We reexamine the role of calcium in the growth of turnip greens,
using data as reported in Kramer et al., in each of 29 experimental
plots the plant calcium \( Y_1 \) was determined, and the soil calcium was
assayed as available \( Y_2 \) and exchangeable \( Y_3 \) calcium. The units
all are milliequivalents per hundred grams. Horticultural specialists
expect these to run at about 15.00, 6.00 and 2.85 units, respectively.
The sample means are \( \bar{Y}_1 = 17.97, 4.39, 2.46 \); and the sample
dispersion matrix is \( S \) with inverse \( S^{-1} = QD Q' \) in spectral form, as
listed in

with \( P = [p_1, \ldots, p_n] \) as its left–singular vectors, \( D_x \) as its singular
values, and columns of \( Q \in \Omega \) as its right–singular vectors. Clearly
\( \kappa = \delta 1 \leq i \leq k \). Our principal reconstruction is articulated in the
following.

Theorem 2. Let \( \pi(\beta) \) be a permutation operator reversing the
ordered array \( \{ \delta_1 \geq \ldots \geq \delta_k \} \) to \( \{ \delta_1, \ldots, \delta_{k-2} \leq \delta_k \} \), and let
D_x = Diag \( \{ \delta_1, \delta_2, \ldots, \delta_{k-2}, \delta_k \} \). Next construct

\[
X'X = k\kappa_q'q + \sum_{i=2}^{k-1} \kappa_i q_i'q_i + \kappa_k q_k'q_k
\]

and the conclusion follows on applying Theorem 1(ii) in the context of Corollary 1(ii).

Remark 3. Note that the foregoing developments are for the general
case that \( \Omega = QD Q' \) is anisotropic with \( \{ \kappa_j \}_{j=1}^k \). If isotropic,
then the following applies.

Definition 1. The model \( \Xi = dI_k \), in which case power functions are directionally invariant,
not depending on directions of alternatives in \( R^k \). Add: Bounds on ARLs
from restricted variation.
Sphericity

The density for \( N_k(\mu, \Sigma) \) has spherical contours for the case
that \( \Sigma = dI_k \), i.e., the model is isotropic. Sample evidence regarding
the isotropy of \( T^2 \) is available. Mauchly1 derived the Likelihood
Ratio test for sphericity, namely, \( H_0 : \Sigma = dI_k \) against \( H_1 : \Sigma \neq dI_k \). A
contemporary test utilizes the modified statistic

\[
LR_M = -\left( 1 - n - \frac{2k^2 + k + 2}{6k} \right) \ln \left( \frac{k!'[S]}{(nS)'^k} \right)
\]
(4)
taking \( S \) as the sample dispersion matrix from n observations,
rejecting at level \( \alpha \) for \( LR_M > c_n \) with \( n = n - 1 \) and with \( c_n \) as the
upper percentile of the central distribution \( \chi^2(k(k+1)/2 - 1, 0) \). See,
for example, Rencher.

Design reversals

Developments thus far are predicated in part on the desirability
to identify alternatives having varying powers of discernment. These
include the most likely and least likely as in Theorem 1(ii). If the least
likely is deemed to be of greatest interest, it remains to ask whether
it might serve instead as the most likely alternative. In the context of
designed experiments the answer is affirmative, as the intrinsic
structure offers a venue for modifying a given design so as to achieve
these ends. Details follow.

Consider the model \( \{ Y = [1, \ldots, X] \alpha, \beta + e \} \) with X centered
such that \( \mathbf{1}'X = 0 \), where location–scale parameters for \( \hat{\beta} \) are \( (\hat{\beta}, \Sigma) \) with
\( \Xi = \{XX\}^{-1} \) as in Corollary 1(ii). In particular, for the test of
\( H_0 : \beta = \beta_o \) against \( H_1 : \beta \neq \beta_o \), utilizes
\( F = (\hat{\beta} - \beta_o)'XX(\hat{\beta} - \beta_o)/ks^2 \)
with \( S^2 \) as the residual mean square and with noncentrality
\( \lambda = (\beta - \beta_o)'XX(\beta - \beta_o)/\sigma^2 \), where it often suffices to take \( \sigma^2 = 1.0 \).
For \( X \in F_{\alpha, n} \), its singular decomposition, followed by \( XX' \), is

\[
X = PD_x Q' = \sum_{i=1}^k \delta_i q_i q'i'X'X = QD_x Q'
\]

(5)

Citation: Jensen D. On directed alternatives in linear inference. Biom Biostat Int J. 2017;6(3):364–371. DOI: 10.15406/bbij.2017.06.00171
On directed alternatives in linear inference

where \( D_k = \text{Diag} (5.61102, 0.04857, 0.00416) \) The data are ill-conditioned, with condition number \( c_1(\Sigma) = 1,348.80 \).

The statistic reported is \( T^2 = n (\bar{\mathbf{y}} - \mu_0)^T \Sigma^{-1} (\bar{\mathbf{y}} - \mu_0) = 24.97 \) with \( \mu_0 = \left[ 15.00, 6.00, 2.85 \right] \), rejecting at level \( \alpha = 0.01 \) the hypothesis \( H_0: \mu = \mu_0 \) in favor of some \( H_i: \mu \neq \mu_0 \). Indeed, the \( p \)-value is \( P(T^2 > 24.97 | H_0) = 0.000751 \) with \( C_\alpha = [14.980] \).

On applying Corollary 1(i) with \( S \) in lieu of \( \Sigma \), the columns of \( \mathbf{Q} = \left[ q_1, q_2, q_3 \right] \) are taken as successive alternatives to \( \left( \mu - \mu_0 \right) \), namely

\[
\begin{bmatrix}
\mu_1 - 15.00 \\
\mu_2 - 6.00 \\
\mu_3 - 2.85
\end{bmatrix}
= 
\begin{bmatrix}
0.01024 \\
0.05973 \\
-0.99916
\end{bmatrix}
\begin{bmatrix}
0.22479 \\
-0.97280 \\
-0.05951
\end{bmatrix}
\]

where the dominant terms are in bold type. The noncentrality parameters \( \left\{ \lambda = n \kappa; 1 \leq i \leq 3 \right\} \) are \( \{162.720, 1.409, 0.121\} \), and taking \( \kappa = 0.05 \) and \( C_\alpha = 9.612, \) powers at these alternatives are

\[
P(T^2 | \lambda) > 9.612 = 1.0000, P(T^2 | \lambda) > 9.612 - 0.1316, P(T^2 | \lambda) > 9.612 - 0.0562.
\]

Accordingly, \( T^2 \) essentially unit power to distinguish the hypothetical deviation \( (\mu_1 - 2.85) \) from -0.99816, since the discrepancies 0.01024 for \( (\mu_1 - 15.00) \) and 0.05973 for \( (\mu_2 - 6.00) \) in \( q_1 \) are negligible. Similarly, \( T^2 \) is marginally able to distinguish \( (\mu_1 - 15.00), (\mu_2 - 6.00) \) from [0.22479, -0.97280] with power 0.1316, but is virtually unable to separate \( (\mu_3 - 15.00), (\mu_2 - 6.00) \) from [-0.97435, -0.22381] with negligible power of 0.0562. In short, the latter suggests [14.03, 5.78] to be plausible values for \( (\mu_1, \mu_2) \).

This is an example, as seen subsequently also, where elements of \( \mathbf{Q} = \left[ q_1, q_2, q_3 \right] \), especially \( q_1 \), convey useful information in regard to the objectives of the study. In summary, directed alternatives, enabled here by Theorem 1 and Corollary 1(i), go beyond conventional usage for \( T^2 \).

**Hotelling's \( T^2 \) charts.**

Multivariate diagnostics figure prominently in Statistical Process Control (SPC), as reviewed subsequently. In monitoring the manufacture of bomb sights during World War II, Hotelling devised \( T^2 \) charts for multivariate means in \( \mathbb{R}^3 \). Here successive values \( \{T^2_i; i = 1,2,...\} \) are charted against time, where the chart signals the process to be out-of-control at level \( \alpha \) whenever \( T^2_i > C_\alpha \).

Moreover, with power \( \psi(\lambda) = P (|T^2_i | \lambda) > C_\alpha \), the Average Run Length (ARL) of time-to-signal is \( ARL = 1/\psi(\lambda) \). To monitor the mean \( \mu \) against its target value \( \mu_0 \), successive samples of size \( n \) yield \( \left\{ \bar{Y}_i, S_i \right\} \), together with \( T^2_i = n (\bar{Y}_i - \mu_0)^T \Sigma^{-1} (\bar{Y}_i - \mu_0) \) having the \( T^2_i \sim \chi^2(n-1,\lambda) \) distribution with \( \lambda = n (\mu_0 - \mu)^T \Sigma^{-1} (\mu_0 - \mu) \). Phase I in SPC is set to establish base line process capabilities, to include parameter estimation, followed in Phase II by implementing the control charts themselves.

To continue, consider the data of Quesenberry\(^7\) to be in Phase I, comprising \( n=30 \) records of 11 quality characteristics indexed in time-order of production. Following Williams et al.\(^4\) dimensions are reduced on selecting the first \( k=5 \) quality characteristics, namely, \( \left\{ Y_1, ..., Y_5 \right\} \) having means \( \left\{ \mu_1, ..., \mu_5 \right\} \), respectively. As in Section Hotelling's \( T^2 \) Tests we take \( S \) in lieu of \( \Sigma \), finding the spectral resolution \( \Sigma^{-1} = \mathbf{Q} \mathbf{Q}^T \) as reported in Table 1. The data are seen to be highly ill-conditioned, with condition number \( c_1(\Sigma) = 3.945.64 \). In keeping with Corollary 1(i), five directed alternatives comprise the columns of \( \mathbf{Q} = \left[ q_1, ..., q_5 \right] \) in Table 1, where dominant elements again are in bold type. In particular, this example shows \( \left\{ q_1, ..., q_5 \right\} \) to be separately informative per se, as each corresponds essentially to deviations in \( (\mu_j - \mu_j) \) for observations \( \left\{ Y_1, Y_2, Y_3, Y_4, Y_5 \right\} \), respectively, since values other than those in bold type are negligible.

Taking \( \{ \lambda = n \kappa; 1 \leq i \leq 5 \} \) with \( n = 30, \alpha = 0.05 \), and critical value \( C_\alpha = 15.097 \), powers against these directed alternatives are listed in Table 1. These show that deviations in the directions of \( \left\{ \mu_1, \mu_2, \mu_3, \mu_4, \mu_5 \right\} \) essentially would be discerned with power at least 0.9982, whereas power in the direction of \( \mu_1 \) would be diminished to 0.2367. Note, however, that these values are inflated by the value \( n=30 \) in Phase I. Samples much smaller in size ordinarily would be taken in Phase II, say \( n=8 \) in this example. Then the powers corresponding to \( \{ \lambda = 8k; 1 \leq i \leq 5 \} \) at \( \alpha=0.05 \) are given subsequently in Table 1. In short, with \( ARL = 1/\psi(\lambda) \) as in the final row of Table 1 with samples of size \( n=8 \), these charts would detect changes in \( \mu_i \) and \( \mu_j \) immediately on average, but less responsive otherwise with \( ARLs \) of \( \{5.41, 5.62, 15.50\} \) for \( \left( \mu_1, \mu_2, \mu_3 \right) \), respectively.

**Remark 5.** This example goes beyond conventional uses of \( T^2 \) charts, in demonstrating that the capacity of a given chart to detect alternatives may differ widely across alternatives in \( \mathbb{R}^5 \) of compelling practical interest. In short, each of five ARLs pertains here to an informative one-dimensional alternative.

**Second-order designs**

Second-order models of type

\[
\mathbf{Y}_i = \mathbf{\beta}_0 + \mathbf{\beta}_i \mathbf{X}_i + \mathbf{\beta}_j \mathbf{X}_j + \mathbf{\beta}_{ij} \mathbf{X}_i \mathbf{X}_j + \mathbf{\beta}_{ijk} \mathbf{X}_i \mathbf{X}_j \mathbf{X}_k + \mathbf{\epsilon}_i; 1 \leq i \leq n
\]

(8)
are considered having zero mean, uncorrelated errors with variance $\sigma^2$. In a typical setting the yield ($Y_i$) of a chemical process is examined at specified reaction time ($X_{1i}$) and temperature ($X_{2i}$). Small designs of historical consequence are the Central Composite (CCD) designs of Box et al., having design points as listed in Table 2.

### Table 1 Spectral values for $S^{-1} - QD_A Q^*$ for the data of Quesenberry (2001) of order (30×5)

| Eigenvalues |
|-------------|
| $\kappa_1$ | $\kappa_2$ | $\kappa_3$ | $\kappa_4$ | $\kappa_5$ |
| 520.4304 | 11.9529 | 1.1404 | 1.0843 | 0.1319 |

| Eigenvectors |
|-------------|
| $q_1$ | $q_2$ | $q_3$ | $q_4$ | $q_5$ |
| 0.99893 | -0.04553 | 0.00490 | 0.00543 | 0.00291 |
| -0.00607 | -0.00445 | 0.14300 | 0.98712 | -0.07126 |
| 0.04535 | 0.99863 | 0.01937 | 0.00322 | 0.01722 |
| -0.00396 | -0.00607 | -0.00445 | 0.14300 | 0.98712 |
| -0.00505 | -0.01907 | 0.98926 | -0.14105 | 0.03287 |

Power at $\alpha = 0.05$ and $n = 30$

| Power at $\alpha = 0.05$ and $n = 30$ |
|--------------------------------------|
| 1.0000 | 1.0000 | 0.9914 | 0.9882 | 0.2367 |

Power at $\alpha = 0.05$ and $n = 8$

| Power at $\alpha = 0.05$ and $n = 8$ |
|--------------------------------------|
| 1.0000 | 1.0000 | 0.1848 | 0.1779 | 0.0645 |

ARLs at $\alpha = 0.05$ and $n = 8$

| ARLs at $\alpha = 0.05$ and $n = 8$ |
|-------------------------------------|
| 1.0000 | 1.0000 | 5.41 | 5.62 | 15.50 |

### Table 2 Regressor vectors for the CCD design $X'$ of order (2×9), where $\alpha = \sqrt{2}$

| $X'$ | -1.00 | -1.00 | $-\alpha$ | $\alpha$ | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
|------|--------|--------|-----------|---------|------|------|------|------|------|
|      | -1.00  | 1.00   | 0.00      | 0.00    | $-\alpha$ | $\alpha$ | -1.00 | 1.00 | 0.00 |

Proceeding as in Corollary 1(ii), we seek spectral values for the Fisher Information Matrix $XX'$, specifically, the eigenvalues and eigenvectors as listed in Table 3, with dominant terms again in bold type. Powers at these directed alternatives follow on taking $\beta \beta = \alpha \alpha < \beta < \alpha \alpha \alpha \alpha \alpha \alpha$ as surrogates in the noncentrality parameter $k_S = \beta \beta / \sigma^2$ with $k = 6$ and $S$ as the residual mean square. Owing to only two degrees of freedom for error, computations replaced $S^2$ by $\sigma^2 = 1.0$, then powers were determined from scaled noncentral $\chi^2(6, k_S)$ distributions with noncentrality parameters as listed in Table 3.

Arranged in decreasing order of their powers, these are $q_1$ with power 0.9765; $q_2$, $q_3$, $q_4$ each with power 0.5307; and $q_5$ and $q_6$ as alternatives with powers 0.2698 and 0.0775. Here elements of $Q$ are separately informative: $q_1$ for discrepancies between $(\beta_1, \beta_2, \beta_{12})$ and their hypothetical values; and $(q_2, q_3, q_4)$ for discrepancies between $(\beta_1, \beta_2, \beta_{12})$ and their hypothetical values, respectively. To continue, observe that the eigenvalue $\kappa_4 = 8.0000$ is repeated three times. On applying Theorem 1(iv) in the context of Corollary 1(ii), we see that all standardized elements in $S_p(q_2, q_3, q_4)$ have power 0.5307. These include, in addition to $q_1$ for $\beta_1$ and $q_4$ for $\beta_2$, the standardized sums

$$
\theta_1 = (q_3 + q_4) / \sqrt{2} = [0, 1, -1, 0, 0, 0] / \sqrt{2}
$$

(9)

$$
\theta_2 = (q_2 + q_3 + q_4) / \sqrt{3} = [0, 1, -1, 0, 7071, 0.7071, 0] / \sqrt{3}
$$

(10)

for example, with $\theta$ for the discrepancy between $(\beta_1 - \beta_1)$, $(\beta_2 - \beta_2)$ and $[1, -1] / \sqrt{2}$.

In summary, the directed second–order alternatives treated here are innovations not found in classical linear inference. Instead, these are enabled by Theorem 1 and Corollary 1(ii). Again the elements of $Q$, especially $|q_2, ..., q_4|$, are separately informative about coefficients of the model (8). Their simple and revealing structure may be attributed to the symmetry and balance of CCD designs.
Design reversals

Begin with \( \{ \mathbf{y} = [1_n, \mathbf{X}] (\mathbf{\beta}_0', \mathbf{\beta}^T) + \epsilon \} \) with \( X_0 = \left[ I_n \cdot X \right] \) and \( X \) in centered form as in Section 3.2, with \( \mathbf{\beta} = [\mathbf{\beta}_1, \mathbf{\beta}_2, \mathbf{\beta}_3] \), having the design \( X = P \text{Diag} (\delta_1, \delta_2, \delta_3) Q' \) as listed in Table 4. Construct \( X_1 = P \text{Diag} (\delta_1, \delta_2, \delta_3) Q' \) on permuting singular values, but retaining the left and right singular vectors.

To continue, take the right–singular vectors, now \( Q = [\mathbf{\theta}, \mathbf{\theta}, \mathbf{\theta}] \), as directed alternatives to \( (\mathbf{\beta} - \mathbf{\beta}_0)' = [(\mathbf{\beta}_1 - \mathbf{\beta}_{10}'), (\mathbf{\beta}_2 - \mathbf{\beta}_{20}'), (\mathbf{\beta}_3 - \mathbf{\beta}_{30}')] \), together with \( k' = [14.5909, 4.4066, 0.2828] \) as \( \{ k_i = \delta_i^2; 1 \leq i \leq 3 \} \). Here the level 0.05 critical value is 6.5914; then powers are determined in turn from \( F (\cdot , 4, k_i) \) together with the power functions \( \{ \psi (x) = P(F > 6.5914/k_i); 1 \leq i \leq 3 \} \) having values listed in the final row of Table 4.

Table 3 Spectral values for the Fisher Information Matrix \( X'X \) for the CCD design

| Eigenvalues |
|-------------|
| \( \kappa_1 \) | \( \kappa_2 \) | \( \kappa_3 \) | \( \kappa_4 \) | \( \kappa_5 \) | \( \kappa_6 \) |
| 24.3427 | 8.0000 | 8.0000 | 8.0000 | 4.0000 | 0.6573 |

| Eigenvectors |
|-------------|
| \( q_1 \) | \( q_2 \) | \( q_3 \) | \( q_4 \) | \( q_5 \) | \( q_6 \) |
| 0.59349 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.80484 |
| 0.00000 | 0.00000 | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| 0.00000 | 0.00000 | 0.00000 | -1.00000 | 0.00000 | 0.00000 |
| 0.56911 | 0.70711 | 0.00000 | 0.00000 | 0.00000 | -0.41966 |
| 0.56911 | -0.70711 | 0.00000 | 0.00000 | 0.00000 | 0.41966 |
| 0.00000 | 0.00000 | 0.00000 | 0.00000 | 1.00000 | 0.00000 |

Power at \( \alpha = 0.05 \)

0.9765

0.5307

0.5307

0.5307

0.2698

0.0775

Table 4 Design matrix \( X = PDQ' \) and the modified \( X_1 = [P \text{Diag} (\delta_1, \delta_2, \delta_3) Q'] \) the left \( (P') \) and right \( (Q') \) singular vectors; and the singular values \( \delta^* = [3.8198, 2.0992, 0.5318] \).

| Design |
|--------|
| \( X' \) |
| \( -1.0000 \) | \( 1.0000 \) | \( -1.4142 \) | \( 1.4142 \) | \( -1.0000 \) | \( 1.0000 \) | \( 0.0000 \) | \( 0.0000 \) |
| \( 0.0000 \) | \( 0.0000 \) | \( 1.0000 \) | \( 1.0000 \) | \( 0.0000 \) | \( 0.0000 \) | \(-1.0000 \) | \(-1.0000 \) |
| \(-0.8000 \) | \( 0.6000 \) | \(-0.8000 \) | \( 2.0000 \) | \(-0.8000 \) | \( 0.6000 \) | \(-0.8000 \) | \( 0.0000 \) |

| Design |
|--------|
| \( X_1' \) |
| \( 0.4279 \) | \(-1.0391 \) | \( 0.2809 \) | \( 0.7126 \) | \( 0.4279 \) | \(-1.0391 \) | \(-1.0391 \) | \( 1.3370 \) |
| \(-0.0193 \) | \(-0.3197 \) | \(-1.3211 \) | \(-0.3556 \) | \(-0.0193 \) | \(-0.3197 \) | \( 0.4993 \) | \( 1.8556 \) |
| \(-0.1811 \) | \( 0.8999 \) | \( 0.2484 \) | \(-1.1704 \) | \(-0.1811 \) | \( 0.8999 \) | \( 1.1798 \) | \(-1.6954 \) |

| Left–singular vectors \( P' \) |
| \(-0.3308 \) | \( 0.2946 \) | \(-0.3736 \) | \( 0.6594 \) | \(-0.3308 \) | \( 0.2946 \) | \(-0.1792 \) | \(-0.0342 \) |
| \( 0.0960 \) | \(-0.1138 \) | \( 0.6024 \) | \( 0.3785 \) | \( 0.0960 \) | \(-0.1138 \) | \(-0.5082 \) | \(-0.4372 \) |
| \( 0.0996 \) | \(-0.3618 \) | \(-0.1302 \) | \( 0.2927 \) | \( 0.0996 \) | \(-0.3618 \) | \(-0.3435 \) | \( 0.7055 \) |

| Right–singular vectors |
| \(-0.7099 \) | \(-0.1307 \) | \(-0.6921 \) |
| \( 0.3508 \) | \(-0.9177 \) | \(-0.1865 \) |
| \( 0.6108 \) | \( 0.3752 \) | \(-0.6973 \) |

Detection probabilities: design \( X \)

0.4947

0.1828

0.0577

Citation: Jensen D. On directed alternatives in linear inference. Biom Biostat Int J. 2017;6(3):364–371. DOI: 10.15406/bbij.2017.06.00171
In particular, discovering alternatives \((\beta - \beta_0)'\) in the direction \(\theta = 3\) is seen to be unlikely, with power 0.0577. On the other hand, suppose instead that it is critical in context to discover alternatives in the negative orthant of \(\mathbb{R}^3\). Then the design \(X_1\) serves to reverse these so that alternatives in the directions of \(\{\theta_1, \theta_2, \theta_3\}\) are now detected with probabilities \([0.4947, 0.1828, 0.0577]\), respectively.

Further properties of the design \(X\) and its reversal \(X_1\) deserve mention. Observe that \(X_0'X_0 = \text{Diag}(n, XX'\lambda)\), whereas \(X_1'X_1 = \text{Diag}(n, XX'\lambda_1)\), where now \(XX' = Q \text{Diag}(\kappa_1, \kappa_2, \kappa_3)X'\) and \(X_1'X_1 = Q_\lambda \text{Diag}(\kappa_1, \kappa_2, \kappa_3)\). Then, the design \(X_1\) serves to reverse these so that alternatives in the directions of \(\{\theta_1, \theta_2, \theta_3\}\) remain ordered. Clearly \(V(\beta | X) = \sigma^2 (XX')^{-1}\) and \(V(\beta | X_1) = \sigma^2 (X_1'X_1)^{-1}\) differ, where their diagonal elements are listed as variances in Table 5. However, their eigenvalues are identical by construction, as are their \(\{A, D, E\}\) efficiency indices as the trace, determinant, and largest eigenvalues of \((X_0'X_0)^{-1}\) under both designs \(X\) and \(X_1\).

| Design Characteristics | \(X\) | \(X_1\) | \(X\) | \(X_1\) |
|-------------------------|-----|-----|-----|-----|
| Estimates \(\hat{\beta}_0\) | 0.12500 | 0.12500 | 3.53636 | 3.53636 |
| \(\hat{\beta}_1\) | 1.38170 | 1.83546 | 0.22694 | 0.22694 |
| \(\hat{\beta}_2\) | 0.69002 | 0.26120 | 0.12500 | 0.12500 |
| \(\hat{\beta}_3\) | 1.76012 | 1.73518 | 0.06854 | 0.06854 |
| Diagnostic | A | D | E |
| ADE | 3.95684 | 0.00688 | 3.53636 |

Summary and discussion

This study reexamines the concept of directional invariance, or isotropy, for distributions on \(\mathbb{R}^k\) having location–scale parameters \((\delta, \Sigma)\). Powers of tests for \(H_0: \delta = 0\) against \(H_0: \delta = 0\) often depend on noncentrality parameters of type \(\lambda = (\delta - 0)/\Sigma^{-1} (\delta - 0)\). The spectral decomposition of \(\Sigma\) supports the identification of directed alternatives in directions determined by the eigenvectors of \(\Sigma\), to encompass the alternatives most likely and least likely in a given study. Powers of these types are independent of direction if and only if \(\Sigma = \sigma^2 I_k\). Applications are drawn in the use of Hotelling’s \(T^2\) in multivariate samples, and of \(F\)-tests in linear models. Case studies are given where this approach leads to the discovery of further insight regarding the natural parameters of a problem.

One concept of directional invariance figures prominently in the SPC literature. The following is excerpted from Linna et al. \(^{13}\)

"It is well known that the ARL performance of multivariate SPC procedures depends heavily on the covariance structure of the observed data. See, for example, Mason et al. \(^{11}\) Further, it has been noted by Pignatiello et al. \(^{12}\) that many multivariate procedures, including the \(\chi^2\) chart, Hotellings \(T^2\) chart, and most of the multivariate CUSUM charts, are directionally invariant. The performance of the multivariate Ewma chart proposed in Lowry et al. \(^{13}\) is also directionally invariant. Lowry et al. \(^{14}\) and others also note the directional invariance of many of these multivariate control charting methods. Directional invariance means that the performance of a procedure does not depend on the specific direction in \(p\)-space of a shift in the mean vector of the process variables being monitored. Instead, performance of a directionally invariant procedure depends only on the statistical (or Mahalanobis) distance between the in-control mean vector \(\mu_0\) and the out-of-control mean vector under consideration, \(\mu\). (Italicics supplied.) That is, \(D_\Sigma (\mu_0, \mu) = (\mu_0 - \mu) \Sigma^{-1} (\mu_0 - \mu)\).

Unfortunately, this notion of directional invariance is grossly misleading, is antithetical to the very concept of invariance as in our Definition 1, at best is a misnomer, and in any event deserves to be clarified in the SPC literature. In fact, such essentials as power functions and ARLs do indeed depend on directions of alternatives, as seen in Section 4.2 as counter examples, unless the model is isotropic.

On the other hand, a disclaimer of Tsui et al. \(^{15}\) should be noted: ‘‘There is no reason in practice, however, for a shift to \(\mu = \mu^*\) to be always considered as important as a shift to \(\mu = \mu^*\) just because the corresponding values of the noncentrality parameters are equal.’’ Our study represents a substantial elaboration on this point.

Citation: Jensen D. On directed alternatives in linear inference. Biom Biostat Int J. 2017;6(3):364–371. DOI: 10.15406/bbij.2017.06.00171
A Appendix

A.1 rayleigh quotients.

At issue are variational properties of quadratic forms of type
\( Q(u) = u' \Omega u / u' u \), known as Rayleigh Quotients; see Bellman.\(^{16}\)

Write \( \Omega = \sum_{i=1}^{k} Q_i \) as its spectral form with \( Q \in \mathbb{R}^k \) so that
\[ \langle q, \Omega q \rangle = \kappa_i; \quad 1 \leq i \leq k. \]
Further partition \( Q = [Q_1, Q_2, Q_3] \) of orders \( (k \times (r-1)), (k \times s), (k \times d) \) with \( d = k - r - s + 1 \), and
\[ \Omega = Q_1 + Q_2 + Q_3 \]
conformably. Then
\[ \Omega = \text{Diag}(D_{\kappa_1}, D_{\kappa_2}, D_{\kappa_3}) \]
\[ \text{Diag}(\kappa_1, \ldots, \kappa_3) \]
where \( \kappa = \{ q, \ldots, q \} \) and \( D = \text{Diag}(\kappa, \ldots, \kappa) \).

Essentials follow.

Lemma A.1 Consider the positive definite form \( \nu \Omega \nu / \nu \nu \) with \( \Omega \)
as in expression (11).

(i) Variational properties of \( Q(u) \) as \( u \) varies over \( \mathbb{R}^k \) are
\[ \kappa_1 \leq \nu \Omega \nu \nu \leq \kappa_2 \]
where the lower and upper limits are attained at \( \nu = q \) and \( \nu = q_1 \), respectively.

(ii) Variational properties of \( Q(u) \) as \( u \) varies over \( S_{p} (Q_2) \) are
\[ \kappa_{r+s} \leq \nu \Omega \nu \nu \leq \kappa_r \]
where the lower and upper limits are attained at \( \nu = q_{r+s} \) and \( \nu = q_r \), respectively.

Proof: Conclusion (i) is given in Bellman,\(^{16}\) where the limits are attained as given since \( \kappa_1 = q_1' \Omega q_1 \) and \( \kappa_2 = q_2' \Omega q_2 \).
To see conclusion (ii), \( u \in S_{p} (Q_2) \) may be represented as
\[ \nu = q_{r+s} a \]
with \( a = [a_1, \ldots, a_{r+s}] \). Then for \( \nu \in S_{p} (Q_2) \), (11) gives
\[ \nu \Omega \nu = a' Q_2 a \]
\[ = \sum_{i=1}^{r+s} \kappa_i a_i^2. \]
The lower and upper limits for \( \nu \in S_{p} (Q_2) \) follow as in conclusion (i) for \( \nu \in \mathbb{R}^k \), except that now these are attained as given since \( q_{r+s}' \Omega q_{r+s} = \kappa_{r+s} \) and
\[ q_r' \Omega q_r' = \kappa_r. \]

Acknowledgement

The author is indebted to Professor Donald E. Ramirez for substantial contributions, including computations using the MINITAB and MAPLE software packages.

References

1. Mahalanobis PC. On the generalised distance in statistics. Proc Nat Acad Sci India. 1936;2(1):49–55.
2. Hotelling H. The generalization of Student’s ratio. Ann Math Statist. 1931;2:360–378.
3. Mauchly JW. Significance test for sphericity of a normal n-variate distribution. Ann Math Statist. 1940;11(2):204–209.
4. Rencher AC. Methods of Multivariate Analysis, 2nd ed. Wiley, USA: New York; 2002.
5. Kramer CY, Jensen DR. Fundamentals of Multivariate Analysis, Part I. Inference about Means. J Quality Technology. 1969;1:120–133.
6. Hotelling H. Multivariate quality control. In: Eisenhart C, editor. Techniques of Statistical Analysis, McGraw-Hill, USA: New York; 1947. p. 111–184.
7. Quesenberry CP. The multivariate short-run snapshot Q chart. Quality Engineering. 2001;13(4):679–683.
8. Williams JD, Woodall WH, Birch JB, et al. Distribution of Hotelling’s \( T^2 \) statistic based on successive differences covariance matrix estimator. J Quality Technology. 2005;38:217–229.
9. Box GEP, Wilson KB. On the experimental attainment of optimum conditions. J Royal Statist Soc Ser B. 1951;13(1):1–45.
10. Linna KW, Woodall WH, Busby KL. Performance of multivariate control charts in the presence of measurement error. J Quality Technology. 2001;33:349–355.
11. Mason RL, Champ CW, Tracy ND, et al. Assessment of multivariate process control techniques. J Quality Technology. 1997;29:140–143.
12. Pignatiello JJ, Runger GC. Comparisons of multivariateCUSUM charts. J Quality Technology. 1990;22(3):173–186.
13. Lowry CA, Woodall WH, Champ CW, et al. A multivariate exponentially weighted moving average control chart. Technometrics. 1992;34(1):46–53.
14. Lowry CA, Montgomery DC. A review of multivariate control charts. IEE Transactions. 1995;27(6):800–810.
15. Tsui KL, Woodall WH. Multivariate control charts based on loss functions. Sequential Analysis. 1993;12(1):79–92.
16. Bellman R. Introduction to Matrix Analysis. McGraw-Hill, USA: New York; 1960.

**Citation:** Jensen D. On directed alternatives in linear inference. Biom Biostat Int J. 2017;6(3):364–371. DOI: 10.15406/bbij.2017.06.00171