Rotatability of variance surfaces and moment matrices

Norman R. Draper

University of Wisconsin–Madison, Madison, WI, USA

Norbert Gaffke and Friedrich Pukelsheim

Universität Augsburg, Augsburg, Germany

Abstract: Box–Hunter (Ann. Math. Statist. 28, 1957) state that the rotatability of a variance surface implies rotatability of the moment matrix $\boldsymbol{M}$ that defines the variance surface. We demonstrate by example that this implication breaks down when the matrix $\boldsymbol{M}$ is positive definite only, but not a moment matrix. We provide a detailed proof that the implication is correct when the matrix $\boldsymbol{M}$ is a positive definite moment matrix.

1. Introduction

In the seminal paper Box–Hunter (1957) the authors present an in-depth study of what has come to be known as response surface methodology. In the present paper we concern ourselves with a single, but important implication out of this 47 page work, namely, that rotatability of a response surface implies rotatability of the corresponding moment matrix. We indicate that this implication has limitations which do not seem to have been noticed before. The key point is that the matrix in question has to be a moment matrix, not just positive definite.

In Section 2 we review rotatability of variance surfaces, and rotatability of moment matrices, for second order models. A rotation $\boldsymbol{R}$ of the vector of experimental conditions $\boldsymbol{t} \in \mathbb{R}^m$ induces a transformation $\boldsymbol{Q}_{\boldsymbol{R}}$ of the corresponding regression vector $\boldsymbol{x}(\boldsymbol{t}) \in \mathbb{R}^k$. In Section 3 we discuss the Box–Hunter (1957) argument—a plain 'Whence'—to establish the implication that rotatability of a variance surface entails rotatability of the corresponding moment matrix $\boldsymbol{M}$.

In Section 4 we illustrate by example that the implication breaks down when the matrix $\boldsymbol{M}$ is positive definite only, rather than also being a moment matrix. In Section 5 we propose a detailed argument that the implication holds true provided $\boldsymbol{M}$ is a moment matrix, besides being positive definite.

Section 6 adds some extensions. The pertinent literature is reviewed in Section 7.

To set the stage, consider a second-order response surface

$$\eta(\xi_1,\dots,\xi_m) = \theta_0 + \sum_i \xi_i\theta_i + \sum_i \xi_i^2\theta_{ii} + \sum_{i<j} \xi_i\xi_j\theta_{ij}. \tag{1}$$

Here $\theta_0, \theta_1, \dots, \theta_m, \theta_{11}, \dots, \theta_{mm}, \theta_{12}, \dots$ are unknown parameters, while $\xi_1, \dots, \xi_m$ are the experimental conditions of m factors which are at the discretion of the experimenter. There is a total of $k = \frac{1}{2}(m+1)(m+2)$ unknown parameters. As usual we collect these parameters to form the column vector $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_m, \theta_{11}, \dots, \theta_{mm}, \theta_{12}, \dots)' \in \mathbb{R}^k$, and we take the experimental conditions to form the column vector $\boldsymbol{t} = (\xi_1, \dots, \xi_m)' \in \mathbb{R}^m$. With *regression vector*

$$\boldsymbol{x}(\boldsymbol{t}) = (1, \xi_1, \dots, \xi_m, \xi_1^2, \dots, \xi_m^2, \xi_1\xi_2, \dots)' \in \mathbb{R}^k, \tag{2}$$

the response surface (1) then takes the simpler form

$$\eta(\boldsymbol{t}) = \boldsymbol{x}(\boldsymbol{t})'\boldsymbol{\theta}. \tag{3}$$

In general the parameter vector $\boldsymbol{\theta}$ is unknown and needs to be estimated from values y_u of the response surface which are observed under a *vector of experimental conditions* $\boldsymbol{t}_u \in \mathbb{R}^m$, for $u = 1, \dots, n$. It is often appropriate to assume that the observations y_u deviate from the true response surface $\eta(\boldsymbol{t}_u) = \boldsymbol{x}(\boldsymbol{t}_u)'\boldsymbol{\theta}$ by random error terms ε_u which have mean zero, a common variance $\sigma^2 > 0$, and are uncorrelated. Thus the statistical model is

$$y_u = x(t_u)'\theta + \varepsilon_u, \quad u = 1, \dots, n. \tag{4}$$

The vectors $\boldsymbol{t}_1, \dots, \boldsymbol{t}_n \in \mathbb{R}^m$ which are chosen for experimental realization are called the *experimental design*. It is well known that their choice determines the variance properties of the least squares estimator $\hat{\theta}$ for the unknown parameter vector $\boldsymbol{\theta}$. The key quantity for this evaluation is the $k \times k$ *moment matrix*

$$\boldsymbol{M} = \frac{1}{n} \sum_{u \leqslant n} \boldsymbol{x}(\boldsymbol{t}_u)\boldsymbol{x}(\boldsymbol{t}_u)', \tag{5}$$

which we assume to be positive definite. The estimated response surface $\widehat{\eta(\boldsymbol{t})} = \boldsymbol{x}(\boldsymbol{t})'\hat{\boldsymbol{\theta}}$ at a general point $\boldsymbol{t}$ then has variance $(\sigma^2/n)\boldsymbol{x}(\boldsymbol{t})'\boldsymbol{M}^{-1}\boldsymbol{x}(\boldsymbol{t})$. Disregarding the common factor σ^2/n, we define the *variance surface* to be

$$v_M(\boldsymbol{t}) = \boldsymbol{x}(\boldsymbol{t})'\boldsymbol{M}^{-1}\boldsymbol{x}(\boldsymbol{t}), \quad \boldsymbol{t} \in \mathbb{R}^m. \tag{6}$$

Thus the variance surface v_M depends on the experimental design $\boldsymbol{t}_1, \dots, \boldsymbol{t}_n \in \mathbb{R}^m$ through the moment matrix $\boldsymbol{M}$.

2. Rotatable variance surfaces, and rotatable moment matrices

Box–Hunter (1957) make a point that a desirable symmetry property of a variance surface is that it depends on the vector of experimental conditions t only through its Euclidean length, that is, that it remains constant for arbitrary rotations of the vector t. Formally, a variance surface v_M is called *rotatable* when

$$v_M(Rt) = v_M(t) \tag{7}$$

for all orthogonal $m \times m$ matrices R and for all $t \in \mathbb{R}^m$. More than 30 years of successful applications of response surface methodology have proved this concept to be a very valuable one.

For a discussion of the implications of rotatability it is expedient to find out what implications it has for the moment matrix M which defines the response surface (6). To this end we first consider an equivariance property of the regression function (2).

Lemma. *For every orthogonal $m \times m$ matrix R there exists a unique nonsingular $k \times k$ matrix Q_R such that for all vectors $t \in \mathbb{R}^m$ we have*

$$x(Rt) = Q_R x(t). \tag{8}$$

Moreover,

$$(Q_R)^{-1} = Q_{R'}. \tag{9}$$

Proof. Let R be an orthogonal $m \times m$ matrix. First we show that there exists a $k \times k$ matrix Q_R which satisfies (8) for all $t \in \mathbb{R}^m$. Partition the regression function (2) according to

$$x(t) = \begin{pmatrix} 1 \\ t \\ \tilde{t} \end{pmatrix}, \tag{10}$$

so that all second order terms are collected in the vector $\tilde{t}$. It is clear that the matrix Q_R must have the pattern

$$Q_R = \begin{pmatrix} 1 & \cdot & \cdot \\ \cdot & R & \cdot \\ \cdot & \cdot & S \end{pmatrix}, \tag{11}$$

a dot indicating zeroes. The remaining block S then is determined from comparing second-order terms,

$$\widetilde{Rt} = S\tilde{t}. \tag{12}$$

Next we establish nonsingularity and uniqueness of Q_R. For every $s \in \mathbb{R}^m$ we insert $t = R's$ in (8) to obtain

$$x(s) = Q_R x(R's). \tag{13}$$

It is well known that we can find k vectors $s_1, \dots, s_k \in \mathbb{R}^m$ such that the second-order regression vectors $x_j = x(s_j)$, $j = 1, \dots, k$, are linearly independent. Thus (13) entails the matrix equality

$$(x_1, \dots, x_k) = Q_R(x(R's_1), \dots, x(R's_k)). \tag{14}$$

Since the $k \times k$ matrix of the left-hand side is nonsingular, so is Q_R. Furthermore the second factor on the right-hand side of (14) is also nonsingular, whence Q_R has the unique representation

$$Q_R = (x_1, \dots, x_k)(x(R's_1), \dots, x(R's_k))^{-1}. \tag{15}$$

Moreover, we have $Q_{R'}Q_R x(t) = Q_{R'}x(Rt) = x(R'Rt) = x(t)$ for all $t \in \mathbb{R}^m$. Insertion of the linearly independent vectors $x_1, \dots, x_k$ shows that $Q_{R'}Q_R = I_k$. This proves (9). □

Example. For $m = 2$ factors the regression vector has dimension $k = 6$. Consider the orthogonal 2×2 matrix

$$R = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{16}$$

which induces a sign change in the second component. Then (8) reads

$$x(Rt) = \begin{bmatrix} 1 \\ t_1 \\ -t_2 \\ t_1^2 \\ t_2^2 \\ -t_1 t_2 \end{bmatrix} = Q_R \begin{bmatrix} 1 \\ t_1 \\ t_2 \\ t_1^2 \\ t_2^2 \\ t_1 t_2 \end{bmatrix} = Q_R x(t). \tag{17}$$

Clearly the solution is

$$Q_R = \begin{bmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & -1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & -1 \end{bmatrix}. \tag{18}$$

Because of (8) rotation of the design vectors t_u in (5) leads to the rotated moment matrix $Q_R M Q_R'$. A moment matrix M then is said to be *rotatable* when

$$M = Q_R M Q_R' \tag{19}$$

for all orthogonal $m \times m$ matrices R, with Q_R given by (8).

3. The Box–Hunter (1957) 'Whence'

Box–Hunter (1957, pp. 207–208) assume the following result, that rotatability of variance surfaces and of moment matrices is actually the same.

Theorem. *For all positive definite moment matrices M one has that the variance surface v_M is rotatable if and only if the matrix M is rotatable.*

Proof. First we establish the converse part. From (19) and (9) we get $M^{-1} = Q_R'^{-1}M^{-1}Q_R^{-1} = Q_{R'}'M^{-1}Q_{R'}$. Rotatability of the variance surface now follows from

$$\begin{aligned} v_M(t) &= x(t)'M^{-1}x(t) \\ &= x(t)'Q_{R'}'M^{-1}Q_{R'}x(t) \\ &= x(R't)'M^{-1}x(R't) \\ &= v_M(R't). \end{aligned}$$

The direct part is more challenging. It is the theme of the rest of the paper. Box–Hunter (1957), before their Equation (29), condense the direct implication to a plain 'Whence'. Our point is that this 'Whence' must be understood in the right way.

Assume that the variance surface v_M is rotatable. Its value at a vector $t \in \mathbb{R}^m$ rotated by R is

$$v_M(Rt) = x(t)'Q_R'M^{-1}Q_R x(t), \tag{20}$$

by the Lemma. Rotatability of v_M means that

$$\begin{aligned} v_M(t) &= x(t)'M^{-1}x(t) = x(t)'Q_R'M^{-1}Q_R x(t) \\ &= v_M(Rt) \quad \text{for all } t \in \mathbb{R}^m, \end{aligned} \tag{21}$$

and for all orthogonal $m \times m$ matrices R. From this we wish to establish the matrix identity

$$M^{-1} = Q_R'M^{-1}Q_R. \tag{22}$$

As in the direct part of the proof, (22) entails $M = Q_{R'}MQ_{R'}'$, and completes the argument.

Thus the key step is to show that (21) implies (22). The plain Box–Hunter 'Whence' must not be construed to be an appeal to the well known result from matrix algebra that two quadratic forms are identical if and only if the matrices defining these quadratic forms are the same. This reasoning requires

$$x'M^{-1}x = x'Q_R'M^{-1}Q_R x \quad \text{for all } x \in \mathbb{R}^k, \tag{23}$$

rather than the much weaker (21).

Also, the quadratic form argument applies whenever the matrices defining the quadratic forms are symmetric (and nonsingular), and makes no use of the fact that M is a moment matrix. Equating quadratic forms would seem—wrongly as we shall see—to prove that (21) implies (22) whenever M is symmetric and nonsingular. A counterexample will illustrate the failure of this argument.

4. 'Whence' investigated

The following example demonstrates that there exist positive definite matrices M for which (21) holds true while (22) fails. The first such counterexample is due to Koll (1980, pp. 181-187), for $m=3$ and $k=10$. The present instance is the nonsingular version of Counterexample 7.3 of Draper-Gaffke-Pukelsheim (1991).

Counterexample. For $m=2$ input variables the regression vector has dimension $k=6$. Our example starts from the rotatable moment matrix

$$M = \begin{pmatrix} 1 & \cdot & \cdot & 1 & 1 & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & 3 & 1 & \cdot \\ 1 & \cdot & \cdot & 1 & 3 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix}, \tag{24}$$

given in Box-Draper (1987, p. 486). For instance, the central composite design with 8 uniformly spaced points on the circle of radius 2 and 8 center points has M for its moment matrix. The matrix M is positive definite and has inverse

$$M^{-1} = \begin{pmatrix} 2 & \cdot & \cdot & -\frac{1}{2} & -\frac{1}{2} & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ -\frac{1}{2} & \cdot & \cdot & \frac{1}{2} & \cdot & \cdot \\ -\frac{1}{2} & \cdot & \cdot & \cdot & \frac{1}{2} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix}. \tag{25}$$

Nonsingularity is maintained if we perturb the entries by a small amount ε. Indeed, direct verification shows that for $\varepsilon \in [0, 1)$ the matrix

$$A(\varepsilon) = M^{-1} + \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & -\varepsilon \\ \cdot & \cdot & \varepsilon & \cdot & \cdot & \cdot \\ \cdot & \varepsilon & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -\varepsilon & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \tag{26}$$

induces the quadratic form, with $x = (x_1, \ldots, x_6)' \in \mathbb{R}^6$,

$$x'A(\varepsilon)x = (1-\varepsilon)(x_1^2+x_2^2+x_3^2+x_6^2)+\tfrac{1}{2}((x_1-x_4)^2 + (x_1-x_5)^2)+\varepsilon((x_1-x_6)^2+(x_2+x_3)^2).$$

Hence $A(\varepsilon)$ is positive definite.

Now define $M(\varepsilon)=A(\varepsilon)^{-1}$. Its variance surface is easily found to be

$$v_{M(\varepsilon)}(t) = 2+\tfrac{1}{2}(t't)^2 \quad \text{for all } t\in \mathbb{R}^m, \tag{27}$$

which is visibly rotatable, for all $\varepsilon\in[0,1)$.

Except for $\varepsilon=0$, however, $M(\varepsilon)$ does not solve equation (22). To see this it suffices to exhibit a rotation R so that (22) is violated. Choosing R and Q_R as in (16) and (18), we obtain

$$Q_R'A(\varepsilon)Q_R = A(-\varepsilon). \tag{28}$$

Clearly we have $A(-\varepsilon)\neq A(\varepsilon)$ whenever $\varepsilon\neq 0$. □

It is not hard to convince oneself that dimension $k=6$ is the smallest dimension for which a counterexample may be found. The reason is that the problem becomes trivial for a single factor ($m=1$), or for two factors ($m=2$) and a first order model.

In the next section we show that the implication from (21) to (22) holds true for positive definite moment matrices M. In retrospect we may then conclude for our counterexample that the matrices $M(\varepsilon)$ for $\varepsilon\neq 0$ fail to be moment matrices. In general, it is difficult to recognize whether a given positive definite matrix is a moment matrix or not.

5. 'Whence' affirmed

We now prove that (21) implies (22), for positive definite moment matrices M. In contrast to the argument of equating quadratic forms we make use of the fact that M is a moment matrix.

In a first step we claim that

$$\operatorname{trace} MQ_R'M^{-1}Q_R = k. \tag{29}$$

This follows from the moment matrix representation (5) and assumption (21) via

$$\begin{aligned}\operatorname{trace} MQ_R'M^{-1}Q_R &= (1/n)\sum_{u\leqslant n} v_M(Rt_u)\\ &= (1/n)\sum_{u\leqslant n} v_M(t_u)\\ &= \operatorname{trace} MM^{-1}\\ &= k.\end{aligned} \tag{30}$$

In a second step we claim that

$$\operatorname{trace}(MQ_R'M^{-1}Q_R)^{-1} = k. \tag{31}$$

Using (9), (21), and $RR'=I_m$, we get as in (30)

$$\begin{aligned}\operatorname{trace}(MQ_R'M^{-1}Q_R)^{-1} &= \operatorname{trace}(Q_{R'}MQ_{R'}'M^{-1})\\ &= (1/n)\sum_{u\leqslant n} \upsilon_M(R't_u)\\ &= (1/n)\sum_{u\leqslant n} \upsilon_M(RR't_u)\\ &= k.\end{aligned}$$

For the third and final step we introduce the positive definite matrix

$$A = M^{1/2}Q_R'M^{-1}Q_RM^{1/2}. \tag{32}$$

Then we know from (29) and (31) that

$$\operatorname{trace} A = \operatorname{trace} A^{-1} = k. \tag{33}$$

We claim that A must be the identity matrix. One way to see this follows the argument of Lemma 7.4 of Draper-Gaffke-Pukelsheim (1991). Namely, let $\lambda_1, \ldots, \lambda_k$ be the eigenvalues of A. Then we have $\sum_{j\leqslant k}(\lambda_j + 1/\lambda_j) = \operatorname{trace} A + \operatorname{trace} A^{-1} = 2k$. But for $\lambda > 0$ the function $\lambda + 1/\lambda$ attains the minimum value 2 uniquely at $\lambda = 1$. Therefore all eigenvalues are unity, and $A = I_k$.

During the workshop, Professors Ingram Olkin and George P.H. Styan pointed out an alternative argument based on the fact that

$$\operatorname{trace}(A - I_k)A^{-1}(A - I_k) = \operatorname{trace}(A - I_k - I_k + A^{-1}) \tag{34}$$

vanishes, by (33), and that this forces $A = I_k$.

In any case, (32) and $A = I_k$ together are the same as (22), which completes the proof of the Theorem. □

6. Extensions

Professor Jerzy K. Baksalary proposed the question of finding necessary and sufficient conditions for a positive definite matrix M such that (21) implies (22). A closer analysis reveals that our arguments carry over to matrices

$$M = \sum_{u\leqslant n} \mu_u x(t_u)x(t_u)', \tag{35}$$

with arbitrary scalars $\mu_u \in \mathbb{R}$, subject to making M positive definite. Hence (21) implies (22) provided M is a positive definite member of the linear space of matrices that are spanned by the rank one matrices $x(t)x(t)'$, $t \in \mathbb{R}^m$.

Furthermore we may place our arguments in a more general context than rotatable designs. Namely, let M_1 and M_2 be two positive definite moment matrices that define identical variance surfaces,

$$v_{M_1}(t) = v_{M_2}(t) \quad \text{for all } t \in \mathbb{R}^m. \tag{36}$$

Again we obtain trace $M_1 M_2^{-1} = k = \text{trace}\, M_1^{-1} M_2$, giving $M_1 = M_2$.

In other words, the variance surface uniquely determines the moment matrix that defines it. In Section 5 we were looking at the particular cases $M_1 = M$ and $M_2 = Q_{R'} M Q'_{R'}$.

Dr. Kenneth Nordström remarked that, for our arguments to carry over to the more general context, an assumption weaker than (36) is sufficient. Namely, if the moment matrix M_1 originates from an experimental design $t_1, \dots, t_n$, and if M_2 originates from $s_1, \dots, s_m$, then all we need to require is equality of the sum of the variances over each of the two designs,

$$\sum_{u \leqslant n} v_{M_1}(t_u) = \sum_{u \leqslant n} v_{M_2}(t_u),$$

$$\sum_{w \leqslant m} v_{M_1}(s_w) = \sum_{w \leqslant m} v_{M_2}(s_w).$$

7. Conclusion

In conclusion we wish to emphasize that the Box–Hunter (1957) 'Whence' serves its purpose perfectly well. Only Koll (1980), Professors O. Krafft and J.N. Srivastava (who were aware of the problem) and the present authors seem to fear that the brief 'Whence' can possibly cause a troublesome misunderstanding.

To the best of our knowledge all previously published literature either reproduces the Box–Hunter (1957) argument, as in Bandemer (1977, p. 398), Myers (1971, p. 221), or otherwise they refer directly to the original publication of Box-Hunter (1957), as do Khuri–Cornell (1987, pp. 60–61), Box–Draper (1987, p. 486).

References

Bandemer, H. (1977). *Theorie und Anwendung der optimalen Versuchsplanung I.* Akademie-Verlag, Berlin.

Box, G.E.P. and N.R. Draper (1987). *Empirical Model-Building and Response Surfaces.* John Wiley, New York.

Box, G.E.P. and J.S. Hunter (1957). Multi-factor experimental designs for exploring response surfaces. *Ann. Math. Statist.* **28**, 195–241.

Draper, N.R., N. Gaffke and F. Pukelsheim (1991). First and second order rotatability of experimental designs, moment matrices, and information functions. *Metrika* **38**, 129–161.

Khuri, A.I. and J.A. Cornell (1987). *Response Surfaces—Designs and Analyses.* Dekker, New York.

Koll, K. (1980). Drehbare Versuchspläne erster und zweiter Ordnung. Diplomarbeit, RWTH Aachen.

Myers, R.H. (1971). *Response Surface Methodology.* Allyn and Bacon, Boston.