Experimental design for mixture systems with multicomponent constraints

Article *in* Communication in Statistics- Theory and Methods · January 1979
DOI: 10.1080/03610927908827762

Ron Snee
Snee Associates, LLC

COMMUN. STATIST.-THEOR. METH., A8(4), 303-326 (1979)

EXPERIMENTAL DESIGNS FOR MIXTURE SYSTEMS WITH MULTICOMPONENT CONSTRAINTS

Ronald D. Snee

Engineering Department
E. I. du Pont de Nemours & Co.
Wilmington, DE 19898

Key Words and Phrases: constrained experimentation; extreme vertices; mixture designs; mixture experiments; multicomponent constraints; product formulation.

ABSTRACT

In an earlier paper it was recommended that an experimental design for the study of a mixture system in which the components had lower and upper limits should consist of a subset of the vertices and centroids of the region defined by the limits on the components. This paper extends this methodology to the situation where linear combinations of two or more components (e.g., liquid content = $x_3 + x_4 + x_5 \leq 0.35$) are subject to lower and upper constraints. The CONSIM algorithm, developed by R. E. Wheeler, is recommended for computing the vertices of the resulting experimental region. Procedures for developing linear and quadratic mixture model designs are discussed. A five-component example which has two multiple-component constraints is included to illustrate the proposed methods of mixture experimentation.

1. INTRODUCTION

In mixture experimentation the property or response (y) of interest is a function of the proportions of each of the q components (x_i) in the mixture and not the total amount of the mixture. Hence, the components are subject to the constraints

$$0 \leq x_i \leq 1, \quad \sum_{i=1}^{q} x_i = 1. \tag{1}$$

Some examples of such mixtures are gasoline, flares, plastics, alloys, concrete, paints, textile fibers blends, ceramics, and cake mixes.

In practice, physical and economic considerations often impose lower (L_i) and upper (U_i) bounds

$$0 \leq L_i \leq x_i \leq U_i \leq 1 \ (i=1,2,...,q) \tag{2}$$

on some or all of the components in the mixture. Multiple variable constraints of the form

$$C_j \leq A_{1j}x_1 + A_{2j}x_2 + \cdots + A_{qj}x_q \leq D_j \tag{3}$$

are also encountered in mixture experimentation.

McLean and Anderson (1966) first recommended the extreme vertices design when constraints (1) and (2) are encountered. Snee and Marquardt (1974) and Snee (1975) have discussed algorithms for the construction of extreme vertices designs to estimate the parameters in linear and quadratic models, respectively. Box and Gardiner (1966) and Cornell and Good (1970) developed designs for an ellipsoidal region of interest which was subject to constraints (1) and (2) and the special case of (3) where $C_j = D_j$ for all $j = 1,2...,m$.

The objective of this paper is to extend the methodology developed by Snee and Marquardt (1974) and Snee (1975) to the situation in which the linear combinations of two or more of the mixture components are subject to upper and lower bounds. In most mixture problems, constraints (1), (2), and (3) form an irregular hyperpolyhedron. In the case of the linear model

$$E(y) = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_q x_q \tag{4}$$

the design will consist of a subset of the vertices of the hyper-polyhedron (Elving, 1952; Snee and Marquardt, 1974). In the case of the quadratic model,

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum_{1 \leq i < j}^{q} \beta_{ij} x_i x_j \tag{5}$$

a design with high statistical efficiency can be constructed from a subset of the vertices and centroids of the feasible region (Snee, 1975).

In the following sections an algorithm for the computation of the vertices of a region defined by (1), (2), and (3) is discussed, and the criteria which will be used to judge the statistical properties of the design are described. The algorithms which are recommended for the selection of the vertices and centroids are discussed. A five-component example involving two multicomponent constraints is included to illustrate the use of this methodology in practice.

2. COMPUTATION OF THE VERTICES AND CENTROIDS

In developing a model for a mixture system one first decides which components will be included in the mixture and what constraints or restrictions will be placed on the levels of the components. The next step is to construct the experimental design for the resulting "region of interest". This is accomplished by computing the vertices and centroids of the region.

As an example, consider the following hypothetical three-component region of interest defined by the following constraints:

Code	Constraint
1L,1U	$.1 \leq x_1 \leq .5$
2L,2U	$.1 \leq x_2 \leq .7$
3L,3U	$0 \leq x_3 \leq .7$
C1L,C1U	$90 \leq 85x_1 + 90x_2 + 100x_3 \leq 95$
C2L	$.4 \leq .7x_1 + x_3$

The region defined by these constraints is shown in Figure 2. While it is possible to determine the vertices of a three-component system by graphical methods, a computing algorithm is needed for $q \geq 4$. Several vertices computation algorithms have been discussed in the literature (Balinski, 1961; Burdet, 1974; Motzkin, et al., 1953; Wets and Witzgall, 1967). The algorithm which will be used here is the CONSIM (CONstrained SIMplex) algorithm developed by R. E. Wheeler from the concepts presented by Motzkin, et al. (1953). The basic concepts underlying the CONSIM algorithm are described in the following paragraphs. The computational details are given in Section 3.

CONSIM begins with the q vertices of the simplex and introduces one constraint plane at a time.

. It is determined whether the vertices lie inside or outside the constraint.

. New points are formed by combining (i.e., linear combination) each point outside the constraint with a point which is on the same one-dimensional edge and inside the constraint. The new point is formed in such a way that it is the intersection of the constraint plane and the one-dimensional edge.

. After all the new points have been generated, the points outside the constraint are deleted, and the next constraint is introduced.

The process continues until all constraint planes have been considered. The points which remain are the vertices of the region defined by the constraints.

The region defined by the first four constraint planes of the three-component example is shown in Figure 1. The vertices of the simplex are points A, B, and C. The first four constraints are introduced in the following order: 1L ($.1 \leq x_1$), 1U ($x_1 \leq .5$), 2L ($.1 \leq x_2$), and 2U ($x_2 \leq .7$). When constraint 1U is introduced, point B, which is outside 1U, is combined (convex combination) with point A to form point D, and point B is deleted. The

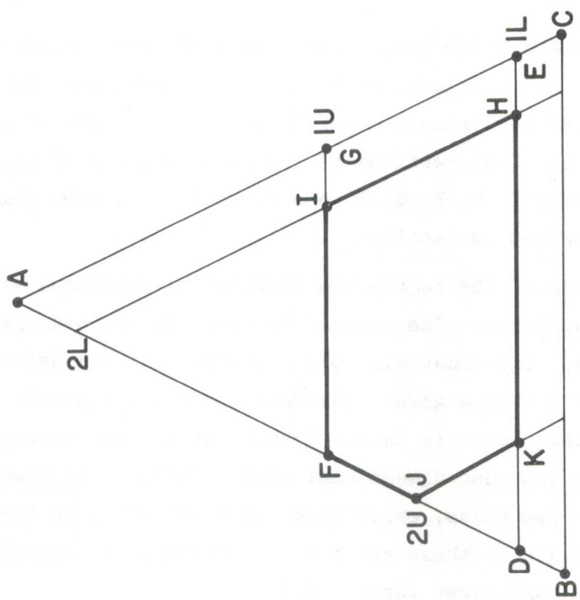

FIG. 1 - THREE-COMPONENT EXAMPLE:
REGION DEFINED BY CONSTRAINTS
ON COMPONENTS x_1 AND x_2

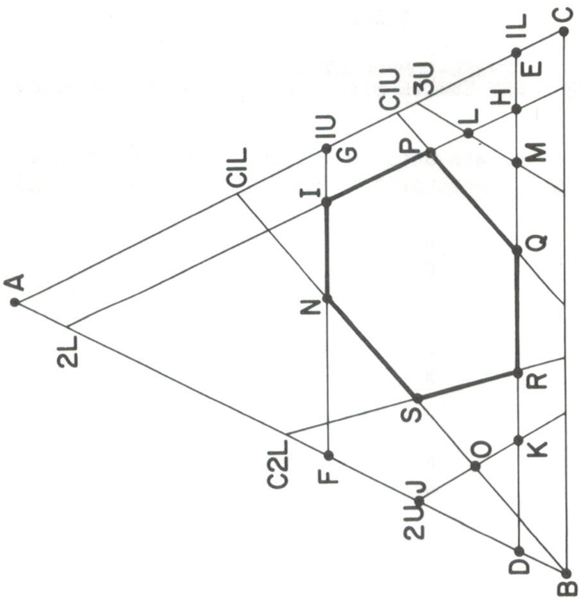

FIG. 2 - THREE-COMPONENT EXAMPLE:
REGION DEFINED BY SINGLE COMPONENT
AND MULTIPLE COMPONENT CONSTRAINTS

TABLE I

THREE-COMPONENT EXAMPLE
RESULTS OF THE FIRST TWELVE STEPS OF THE CONSIM ALGORITHM

Step	Constraint	Points Outside Constraint	Points Combined	Point Added	Points Deleted
1	1L	B	A,B	D	
2		C	A,C	E	
3					B,C
4	1U	A	A,D	F	
5		A	A,E	G	
6					A
7	2L	E	E,D	H	
8		G	G,F	I	
9					E,G
10	2U	D	D,F	J	
11		D	D,H	K	
12					D

results of introducing the next three constraint planes are summarized in Table I. An evaluation of all single-component and multiple-component constraints results in the generation of 19 points (A,B,...,S). The vertices of the resulting region shown in Figure 2 are points I, N, P, Q, R, and S. The associated computations are described in Section 3.

The centroids of the region are computed by averaging various subsets of the vertices. The overall centroid is the average of all the vertices. The constraint plane centroid is the average of all points which lie on a given constraint (e.g., all points with $x_i = U_i$). The edge centroid is the average of any two vertices which lie on the same one-dimensional edge. While it is possible to compute other centroids, experience has shown that, in the case of the quadratic model, these are the only classes of centroids which need to be considered (Snee, 1975).

3. THE CONSIM ALGORITHM

The computational details of the CONSIM algorithm, as developed by R. E. Wheeler, are described in this section and illustrated using the three-component example introduced in Section 2. It should be noted that the CONSIM algorithm is written specifically for the simplex region defined by constraints (1), (2), and (3) described in Section 1. The Motzkin, et al., double-description algorithm is more general and enables one to start with unbounded regions and to colve for "bounding rays" under arbitrary inequalities (see Motzkin, et al. (1953), Section III).

Without loss of generality, constraints (1), (2), and (3) can be written as a set of h inequalities in q variables

$$\Sigma A_{ij}x_i \geq c_j \qquad (j = 1,2,\ldots,h)$$

which can be written as

$$\underline{x}\ \underline{A} \geq \underline{C} \tag{6}$$

where $\underline{x}$ is a (lxq) row vector, $\underline{A}$ is a (qxh) matrix, and $\underline{C}$ is a (hxl) column vector.

If the set of $\underline{x}$'s are bounded and of dimension r, then the region, $\underline{R}$, is also defined by a set of extreme points

$$\{\underline{a}_k\} \qquad (k = 1,2,\ldots,n) \tag{7}$$

Thus, the region is doubly described by (6) and (7). The algorithm assumes (6) and (7) and finds the vertices of R_1 which are the intersection of R and the half space defined by a new inequality

$$\underline{x}\ \underline{A} \geq c_{h+1} \tag{8}$$

This is accomplished by forming for each point $\underline{a}_k$, k = 1,2,...,n, the linear combination

$$d_k = \underline{a}_k\underline{A} - c_{h+1} = \underline{a}_k\underline{A} - \underline{a}_k c_{h+1} = \underline{a}_k\underline{\theta}$$

where $\theta_i = A_i - c_{h+1}$. If $d_k \geq 0$, then the corresponding a_k may be an extreme point of R_1. If $d_k < 0$, then a_k is not an extreme point of R_1 but can be combined with certain nonnegative points, a_k, (i.e., $d_k \geq 0$) to form extreme points of R_1. In fact, the following theorem holds.

Theorem: If a_k with $d_k \geq 0$ and a_m with $d_m < 0$ are on the same one-dimensional edge of R, then there exists a point $a_w = \alpha a_m + (1 - \alpha)a_k$ with $0 \leq \alpha \leq 1$ which is an extreme point of R_1.

Proof: Since $\alpha = d_k/(d_k - d_m)$, $a_w A = \alpha d_m + (1 - \alpha)d_k = c_{h+1}$, hence, a_w satisfies (8) and is an extreme point of R_1 because it lies on a one-dimensional edge of R_1 at the extreme.

Given a set of inequalities, the algorithm calculates d_k for a new set of inequalities saving all a_k with $d_k \geq 0$ and discarding all a_k with $d_k < 0$ after first adding all convex combinations of negative a_k with nonnegative a_k with common one-dimensional edges. Two points, a_k and a_m, are on the same one-dimensional edge defined by $\underline{A}$ and c_{h+1} if $d_k = d_m = 0$ for at least q-2 constraints. It is noted that d was used in Section 4 to denote the maximum value of $v = \underline{x}(\underline{X}'\underline{X})^{-1}\underline{x}'$ over all candidate points. The letter d with a subscript (i.e., d_k) is used in the CONSIM algorithm to denote criteria for evaluating extreme points. It is hoped that this dual usage of d will not cause confusion.

The three-component mixture space discussed in Section 2 will be used to illustrate the calculations associated with the CONSIM algorithm. The region defined by the constraints in this problem (see Section 2) are shown in Figures 1 and 2. It is helpful to refer to these figures when studying the computational procedure described below.

The computation algorithm begins by expressing all of the constraints in the form

$$\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \geq 0.$$

For example, the constraint

$$c_j \leq A_1 x_1 + A_2 x_2 + A_3 x_3$$

can be rewritten as

$$A_1x_1 + A_2x_2 + A_3x_3 \geq c_j(x_1+x_2+x_3)$$

or $$(A_1-c_j)x_1 + (A_2-c_j)x_2 + (A_3-c_j)X_3 \geq 0.$$

Hence, $\theta_i = A_i-c_j$.

The resulting matrix of θ_i's is

$$\begin{array}{c}
\begin{array}{cccccccc} 1L & 1U & 2L & 2U & 3U & C1L & C1U & C2L \end{array} \\
\underline{\theta} = \begin{bmatrix} .9 & -.5 & -.1 & .7 & .7 & -.05 & .10 & .3 \\ -.1 & .5 & .9 & -.3 & .7 & 0 & .05 & -.4 \\ -.1 & .5 & -.1 & .7 & -.3 & .10 & -.05 & .6 \end{bmatrix}
\end{array}$$

Note that θ_{C1L} and θ_{C1U} have been scaled by multiplying by 0.01. In general, it is recommended that the $\underline{A}$ vector and corresponding c_j be scaled by dividing $\underline{A}$ and c_j by the maximum of A_{ij} and c_j to reduce roundoff errors. To ensure accurate results it is recommended that all calculations be done in double precision and that a tolerance of 10^{-6} be used in determining when $d_k = 0$ (i.e., $d_k = 0$ when $d_k \leq 10^{-6}$).

The first three points are the vertices of the simplex: A = (1,0,0), B = (0,1,0), and C = (0,0,1). The relationship of A, B, and C to the lower constraint on X_1, denoted 1L, is determined by forming the inner product of A, B, and C with the θ's which define the 1L constraint. This results in

$$d_A = .9(1)-.1(0)-.1(0) = .9$$
$$d_B = .9(0)-.1(1)-.1(0) = -.1$$
$$d_C = .9(0)-.1(0)-.1(1) = -.1$$

as shown in column 1L of Table II. It is concluded that, since d_A is positive, and d_B and d_C are negative, A lies inside the 1L constraint and B and C lie outside the 1L constraint. An examination of Figure 1 confirms these results.

Next points are added to the list by combining pairs of points which lie on the same one-dimensional edge. Two points on the same one-dimensional edge are combined if one point is inside a given constraint ($d_k > 0$) and the other point is outside the constraint ($d_k < 0$). In general, two points lie on the same one-dimensional

TABLE II

THREE COMPONENT EXAMPLE - CONSIM CALCULATIONS

				CONSTRAINT								
Point	x_1	x_2	x_3	1L	1U	2L	2U	3U	C1L	C1U	C2L
A	1.000	0	0	.90	-.50						
B	0	1.000	0	-.10							
C	0	0	1.000	-.10							
D	.100	.90	0	0	.40	.80	-.20				
E	.100	0	.900	0	.40	-.10					
F	.500	.50	0	.40	0	.40	.20	.70	-.25		
G	.500	0	.500	.40	0	-.10					
H	.100	.10	.800	0	.40	0	.60	-.10			
I	.500	.10	.400	.40	0	0	.60	.30	.02	.04	.35
J	.300	.70	0	.20	.20	.60	0	.70	-.02		
K	.100	.70	.200	0	.40	.60	0	.50	.02	.04	-.13
L	.200	.10	.700	.10	.30	0	.60	0	.06	-.01	
M	.100	.20	.700	0	.40	.10	.50	0	.06	-.02	
N	.500	.25	.250	.40	0	.15	.45	.45	0	.05	.20
O	.200	.70	.100	.10	.30	.6	0	.60	0	.05	-.16
P	.267	.10	.633	.17	.23	0	.60	.07	.05	0	.42
Q	.100	.35	.550	0	.40	.25	.35	.15	.05	0	.22
R	.100	.57	.330	0	.40	.47	.13	.37	.03	.02	0
S	.333	.50	.167	.23	.17	.40	.20	.53	0	.05	0

edge if the points have at least q-2 constraint planes in common which is equivalent to having matching $d_k = 0$ for at least q-2 constraint planes. In this example $q-2 = 3-2 = 1$. In Table II we see $d_A > 0$, $d_B < 0$, and $d_C < 0$. Points A and C have matching zeros in the x_2 column. This indicates A and C can be combined to form point E. In a similar fashion A and B have matching zeros in the x_3 column in Table II and are combined to form point D. In order to compute $D = \alpha A+(1-\alpha)B$ it is necessary to determine α. Remembering $D\theta_{1L} = 0$, produces $[\alpha A+(1-\alpha)B]\theta_{1L} = D\theta_{1L} = 0$ which results in $\alpha = B\theta_{1L}/(B-A)\theta_{1L} = d_B/(d_B-d_A)$. Hence, to form D and E, $\alpha = -.1/(-.9-.1)$ and $-.1/(-.9-.1)$, respectively. Now points D = (.1,.9,0) and E = (.1,0,.9) are added to the list, $D\theta_{1L} = E\theta_{1L} = 0$ is recorded in column 1L of Table II and points B and C are discarded. Note that the first q = 3 columns of Table II are the coordinates of the vertices and the remaining columns contains the d_k values for each of the constraints.

The 1U constraint is now introduced and θ_{1U} is dot producted with points A, D, and E (Table II, column 1U). Now $d_A < 0$, $d_D > 0$, $d_E > 0$, and points A and E and A and D have matching zeros in the x_2 and x_3 columns, respectively. Hence, the points to be added to the list are $F = \alpha(A) + (1-\alpha)D$ and $G = \alpha(A) + (1-\alpha)E$. After computing α as described above, points F = (.5,.5,0) and G = (.5,0,.5) are added to the list, $F\theta_{1U} = G\theta_{1U} = 0$, $F\theta_{1L} = .4$, and $G\theta_{1L} = .4$ are recorded in rows F and G of Table II, and point A is deleted because $d_A < 0$.

Next, the 2L constraint is introduced and the dot products $d_D = D\theta_{2L}$, $d_E = E\theta_{2L}$, $d_F = F\theta_{2L}$, and $d_G = G\theta_{2L}$ are recorded in column 2L of Table II. Now d_D and d_F are positive and d_E and d_G are negative; hence, the algorithm adds new points by combining one or more of the following pairs of points: DE, DG, FE, and FG. Points D and E and F and G have matching zeros in columns 1L and 1U of Table II, respectively. In Figure 1 it is apparent that D and E are combined to form point H and F and G are combined to form point I. The cycle is completed by adding points H and I to the list, recording the inner products of H and I with

θ_{1L}, θ_{1U}, and θ_{2L} in rows H and I of Table II, and deleting points E and G from the list.

The process continues by introducing one constraint plane at a time,

- Computing d_K for the new constraint to determine whether the points in the list lie inside ($d_K > 0$) or outside ($d_K < 0$) the constraint.
- Determining which points should be combined to form new points and adding the new points to the list.
- Computing d_k for the new points and all constraints considered up to this point, and
- Deleting all points which lie outside the constraint being considered (i.e., points with $d_k < 0$).

The calculations for the remaining constraints are summarized in Table II and shown graphically in Figures 1 and 2. At the end of the calculation procedure six points remain; namely, points I, N, P, Q, R, and S (Table II, Figure 2). These points are the vertices of the region defined by the constraints.

4. EXPERIMENTAL DESIGN STATISTICAL PROPERTIES

After the vertices and centroids have been computed, one proceeds to construct the design. In order to compare various possible designs it is helpful to have some criteria to evaluate the statistical properties of a design. The linear (4) and quadratic (5) mixture models can be written in matrix notation as $\underline{Y} = \underline{X}\underline{\beta}+\underline{\varepsilon}$ where $\underline{Y}$ is the (nx1) vector of responses, $\underline{X}$ is the (nxp) "expanded" design matrix, $\underline{\beta}$ is the (px1) vector of unknown coefficients, and $\underline{\varepsilon}$ is the (nx1) vector of errors with zero mean and variance $\sigma^2\underline{I}$ where σ^2 is the observation error variance and $\underline{I}$ is the (nxn) identity matrix. In the linear model p = q; in the quadratic model p = q(q+1)/2 and $\underline{X}$ contains q linear terms and q(q-1)/2 quadratic ($x_i x_j$) terms). Hence, n is the number of observations in the design and p is the number of parameters in the model. The variance-covariance matrix for $\hat{\underline{\beta}}$, the least squares estimate of $\underline{\beta}$, is $\sigma^2(\underline{X}'\underline{X})^{-1}$, and the prediction variance at a point $\underline{x}$ is $\sigma^2 v$ where $v = \underline{x}(\underline{X}'\underline{X})^{-1}\underline{x}'$, and $\underline{x}$ is a (1xp) vector.

Linear Model: $\underline{x} = (x_1, x_2, \ldots, x_q)$

Quadratic Model: $\underline{x} = (x_1, x_2, \ldots, x_q, x_1x_2, x_1x_2, \ldots, x_{q-1}, x_q)$

The statistical properties of a design for the linear and quadratic models will be evaluated by computing four design statistics: $|(\underline{X}'\underline{X})^{-1}|$, trace $(\underline{X}'\underline{X})^{-1}$, % G-efficiency = 100p/nd, and maximum prediction variance $\sigma^2 d$, where d is the maximum value of $v = \underline{x}(\underline{X}'\underline{X})^{-1}\underline{x}'$ over all the candidate points. The $|\underline{X}'\underline{X}^{-1}|$ is proportional to the generalized variance of $\hat{\underline{\beta}}$. The trace $(\underline{X}'\underline{X})^{-1}$ is proportional to the average variance of the estimate coefficients. The $|(\underline{X}'\underline{X})^{-1}|$, trace $(\underline{X}'\underline{X})^{-1}$, and maximum prediction variance $\sigma^2 d$ are decreasing functions of n, while G-efficiency (G for global) is calculated on a per-point basis.

Wheeler (1972) has investigated the merits of designs with high efficiencies according to various criteria and concluded that there are few advantages and some disadvantages to insisting on fully-efficient designs. He found that designs which have G-efficiencies $\geq$ 80% also have high efficiencies on most other criteria. He also notes that Atwood (1969) has shown that G-efficiency $\leq$ D-efficiency, which is based on minimizing a limiting form of the generalized variance. As a practical rule of thumb, Wheeler suggests that any design with a G-efficiency $\geq$ 50% be called "good" and shows that pursuit of higher efficiencies is not generally justified in practice.

It is also important that, whenever possible, the maximum prediction variance be $\leq 1.0\sigma^2$ (i.e., $d \leq 1$). Assuming that the model is correct, d = 1 implies that the model predictions are only as precise as the measured responses. In addition to the design providing reasonably efficient prediction of the response over the design space, we also require that "hedge points" be included to enable us to detect model lack of fit. Also, it is advantageous to have sufficient residual degrees of freedom so that abnormal data points can be detected. In addition to these qualities, the design should have other characteristics of a good experimental design as discussed, for example, by Box and Hunter (1957).

5. SELECTING A SUBSET OF THE VERTICES

It is well known that for $q \geq 5$, the extreme vertices design often contains more points than it is possible to run. For example, it is not uncommon for a $q = 10$ problem to have over 1,000 vertices. The introduction of multiple variable constraints does not change the situation, and from a practical viewpoint, it becomes necessary to select a subset of the vertices to be included in the design. The recommended selection procedure depends on whether the model is linear or quadratic.

5.1 Linear Model

Elving's work (1952) indicates that, in the case of the linear mixture model (4), vertices of the region are the only points which need to be included in the design. It is also worthwhile from a practical viewpoint to include the overall centroid in the design as a check for curvature. In the case of multicomponent constraints, Kennard and Stone's CADEX algorithm (Kennard and Stone, 1969; Snee and Marquardt, 1974) or the exchange algorithm (Mitchell, 1974; Wheeler, 1972) which will be discussed in the following section, are recommended for selecting a subset of the vertices. In the case of CADEX, the design statistics ($|(\underline{X}'\underline{X})^{-1}|$, Trace $(\underline{X}'\underline{X})^{-1}$, percent G-efficiency, maximum prediction variance) are computed as each point is added to the design and one chooses a design using one or more of these criteria. In the case of the linear model, Snee and Marquardt (1974) recommend the minimum trace $(\underline{X}'\underline{X})^{-1}$ criterion. As will be discussed later, the exchange algorithm searches for a design with maximum $|\underline{X}'\underline{X}|$.

It should be noted that the XVERT algorithm developed by Snee and Marquardt (1974) can be combined with the CONSIM algorithm to select a subset of the vertices when multiple-component constraints are encountered. The XVERT algorithm which was developed for the construction of linear model designs when single-component constraints are encountered, divides the vertices into a set of core points and a series of candidate subgroups which contain 2 or more points. The core and candidate subgroup points serve as the initial set of starting points for the CONSIM algorithm just as the simplex vertices served as the starting points for the three-component example. The CONSIM algorithm is now employed by introducing the first multiple-variable constraint. Any core point which does not satisfy a multiple-component constraint results in a new candidate subgroup which contains all points generated from the core point. In the case of the candidate subgroups, newly-generated points are added to the candidate subgroup which contained the point used in computing the new point. As with the original XVERT algorithm, the design will consist of the remaining core points and one point from each candidate subgroup. The best design is determined by forming all possible designs consisting of the core points and one point from each subgroup and computing the design statistics. One can then choose a design using the criteria which best suits the needs of the experiment.

The XVERT-CONSIM algorithm works well for one or two multiple-component constraints. It becomes less useful, however, as the number of multiple-component constraints becomes large. Of course, it would be of no use in problems where there were no single-component constraints.

5.2 Quadratic Model

Snee (1975) has shown that a design with high-statistical efficiency for the quadratic model can be constructed from a subset of the vertices and centroids of the feasible region. Specifically, the design for three-component systems will consist of the vertices, overall centroid, and the centroids of any long edges. Designs for four-component systems should consist of the vertices, constraint plane centroids, overall centroid, and the centroids of any long edges. Snee (1975) recommended that designs for five or more components be constructed by using Wheeler's nullification (1972), Wynn (1970), and exchange (Mitchell, 1974; Wheeler, 1972) algorithms to select points from a candidate list consisting of the vertices, edge centroids, constraint plane centroids, and the overall centroid.

Briefly, nullification produces a nonsingular design by starting with the point farthest from the centroid of the candidate list and sequentially adding points from the null space of $\underline{X}'\underline{X}$ developed from the points already in the design. When a nonsingular design is found, points are added sequentially by the Wynn algorithm until the desired number of points in the design is reached. Then the exchange algorithm is used to improve the design if possible. At each step the point added by the Wynn algorithm is that point $\underline{x}$ with the maximum value of $v = \underline{x}(\underline{X}'\underline{X})^{-1}\underline{x}'$ where $(\underline{X}'\underline{X})^{-1}$ is developed for the points already in the design. The exchange algorithm increases $|\underline{X}'\underline{X}|$ of an n-point design (equivalent to minimizing $|(\underline{X}'\underline{X})^{-1}|$) by adding the point with maximum v and then deleting the point from the n+1 point design with the minimum v. In general, at each cycle of the exchange algorithm the design size is increased from n to n+t points by sequentially adding the point at each step with the maximum v and then decreased from n+t to n points by deleting the point with the minimum value of v at each step. Further discussions of these algorithms can be found in the papers by Mitchell (1974), St. John and Draper (1975), Snee (1975), and Wheeler (1972). The use of these algorithms will be illustrated in the following example.

6. ILLUSTRATIVE EXAMPLE

In an experimental program to formulate a plastic product it became necessary to determine the effects of the composition variables on the hardness and other physical properties of the product. It was decided to study five major composition variables which totaled 99.7% of the mixture. The five components were to be studied over the following ranges:

Component	Range
x_1 Polymer A	.50 - .70
x_2 Polymer B	.05 - .15
x_3 Plasticizer 1	.05 - .15
x_4 Plasticizer 2	.10 - .25
x_5 Plasticizer 3	.00 - .15

Physical considerations suggested that the sum of components 4 and 5 should make up between 18% and 26% of the formulation and that the liquid plasticizer content should be less than 35% of the formulation. These constraints, in algebraic terms, are

Components	Range
$x_4 + x_5$	$.18 \leq x_4 + x_5 \leq .26$
Liquid Plasticizer	$x_3 + x_4 + x_5 \leq .35$

The experiment had to be completed within two weeks in order to meet the objectives of the program. It was possible to prepare and test approximately 25 formulations in this time period. Since we had no independent measure of experimental variation, it was decided to use a 20-point design and replicate five of the points. A five-variable quadratic model contains 15 coefficients, hence a 25-point design would result in five degrees of freedom to measure lack of fit.

The single-component and multiple-component constraints defined a region which had 38 vertices, 13 constraint-plane centroids, one overall centroid, and 76 edge centroids, resulting in a candidate list of 128 points. In order to conserve space only the 38 vertices and those centroids which were involved in subsequent calculations are given in Tables III and IV. Wheeler's nullification procedure was used to find an initial 15-point nonsingular design which consisted of points 13, 17, 35, 22, 12, 23, 20, 9, 1, 32, 87, 38, 74, 57 and 40 (Tables III, IV and V). The Wynn algorithm was used to select the next five points, and then the exchange algorithm was used to improve the design. The nullification, Wynn, and exchange algorithms results are given in Table V.

TABLE III

PLASTIC FORMULATION STUDY
VERTICES OF THE EXPERIMENTAL REGION

Point	x_1	x_2	x_3	x_4	x_5	Hardness
1	.700	.050	.050	.197	.000	130, 130
2	.500	.150	.150	.197	.000	
3	.700	.050	.050	.100	.097	
4	.500	.150	.150	.100	.097	
5	.647	.050	.050	.250	.000	
6	.547	.150	.050	.250	.000	
7	.500	.150	.097	.250	.000	
8	.647	.050	.050	.100	.150	
9	.547	.150	.050	.100	.150	
10	.500	.150	.097	.100	.150	
11	.700	.067	.050	.180	.000	
12	.700	.067	.050	.100	.080	300, 230
13	.617	.150	.050	.180	.000	60
14	.617	.150	.050	.100	.080	
15	.700	.050	.067	.180	.000	
16	.700	.050	.067	.100	.080	
17	.617	.050	.150	.180	.000	7
18	.617	.050	.150	.100	.080	
19	.517	.150	.150	.180	.000	
20	.517	.150	.150	.100	.080	5, 4
21	.637	.050	.050	.250	.010	
22	.637	.050	.050	.110	.150	40
23	.537	.150	.050	.250	.010	5
24	.500	.150	.087	.250	.010	
25	.537	.150	.050	.110	.150	
26	.500	.150	.087	.110	.150	
27	.597	.050	.100	.250	.000	
28	.597	.050	.150	.200	.000	
29	.500	.147	.150	.200	.000	
30	.500	.147	.100	.250	.000	
31	.597	.050	.100	.100	.150	
32	.597	.050	.150	.100	.100	7
33	.500	.147	.150	.100	.100	
34	.500	.147	.100	.100	.150	
35	.597	.050	.090	.250	.010	4, 4
36	.500	.147	.090	.250	.010	
37	.597	.050	.090	.110	.150	
38	.500	.147	.090	.110	.150	4, 4

The points to be replicated were chosen by the Wynn algorithm using the 20-point design developed in the first run as the candidate list. The first five points chosen by the Wynn algorithm were vertices 1, 12, 20, 32, and 38. The resulting 25-point design had a G-efficiency of 60% and a maximum prediction variance of $1.01\sigma^2$ over the 128 candidate points. The overall centroid did not appear in the original design and was included as the 26th point to provide a measure of the response level at the center of the region. The final design consisted of 15 vertices which included five replicates, two constraint plane centroids, the overall centroid, and eight edge centroids.

TABLE IV

PLASTIC FORMULATION STUDY
CENTROIDS USED IN CONSTRUCTING THE EXPERIMENTAL DESIGN

Point	x_1	x_2	x_3	x_4	x_5	Hardness
39*	.579	.103	.091	.162	.062	9
40*	.588	.101	.097	.210	.000	
47*	.588	.101	.097	.100	.110	20
52*	.548	.098	.113	.168	.068	4
54**	.637	.050	.050	.180	.080	10
57	.537	.150	.050	.180	.080	6
59	.567	.150	.100	.180	.000	10
62	.567	.100	.150	.180	.000	5
64	.587	.100	.050	.250	.010	8
74	.658	.050	.108	.100	.080	60
87	.500	.147	.125	.225	.000	Too Soft
92	.582	.150	.050	.100	.115	50

* Point 39 is the overall centroid. Points 40, 47 and 52 are centroids of the following constraint planes shown in parentheses: 40 ($x_5 = 0$), 47 ($x_4 = .10$), 52 ($x_3 + x_4 + x_5 = .35$). For example, point 40 was computed by averaging all vertices with $x_5 = 0$.

** Points 54, 57, ..., 92 are edge centroids formed by averaging the following pairs of vertices shown in parentheses: 54 (21, 22), 57 (23, 25), 59 (13, 19), 62 (17, 19), 64 (21, 23), 74 (16, 18), 87 (29, 30), 92 (9, 14). For example, point 54 was computed by averaging vertices 21 and 22.

The hardness measurements for the 26 formulations are also given in Table III. The 15-term quadratic model with coefficients

$$\hat\beta_1 = 2.97 \quad \hat\beta_{12} = -0.06 \quad \hat\beta_{24} = 6.12$$
$$\hat\beta_2 = 0.66 \quad \hat\beta_{13} = -0.24 \quad \hat\beta_{25} = 5.96$$
$$\hat\beta_3 = -1.72 \quad \hat\beta_{14} = 3.82 \quad \hat\beta_{34} = 5.74$$
$$\hat\beta_4 = -2.19 \quad \hat\beta_{15} = 3.99 \quad \hat\beta_{35} = 5.05$$
$$\hat\beta_5 = -1.33 \quad \hat\beta_{23} = 1.23 \quad \hat\beta_{45} = -3.56$$

fit the 25 observations (observation 19 was missing) with an adjusted R^2 (Marquardt and Snee, 1974) of $R_A^2 = 0.98$. In this

TABLE V

PLASTIC FORMULATION STUDY DESIGN
NULLIFICATION, WYNN, AND EXCHANGE ALGORITHM RESULTS

Algorithm	N	Design Points Added(a)	Design Points Deleted	% G-Efficiency	Maximum $Var(\hat y)/\sigma^2$	Log $\lvert X'X \rvert \times 10^{53}$
Nullification	15	(b)		35	2.83	0.676
Wynn	16	64		39	2.38	1.261
	17	92		43	2.06	1.789
	18	52		57	1.47	2.274
	19	62		56	1.41	2.669
	20	54		56	1.34	3.051
Exchange	20	59	9	63	1.19	3.112
	20	14	40	57	1.31	3.136
	20	47	14	60	1.25	3.147*
Wynn	25	(c)		60	1.01	

(a) See table for description of points.

(b) Nullification design: Points 13, 17, 35, 22, 12, 23, 20, 9, 1, 32, 87, 38, 74, 57, 40 (see Tables III and IV).

(c) Replicated points in order of selection: 1, 20, 38, 12, 35

* Design could not be improved further using a maximum of five-point exchanges.

model the x's were expressed in terms of pseudocomponents (Gorman, 1970) and y = $\log_{10}$ hardness was the response. The regression calculations were made in double precision to reduce computer roundoff errors which can occur when fitting models in constrained mixture spaces (Gorman, 1970). The lack of fit was significant ($F = 5.73$, $p < .05$) when compared with the variation between the five duplicate points, which was unusually small (three of the five pairs of duplicates were identical). The high adjusted R^2 statistic ($R_A^2 = 0.98$) provided assurance that the quadratic model gave a good description of hardness within the region of the data. The quadratic models for the other three responses had no significant lack of fit.

Point 19 could not be made because the resulting formulation was too soft to work with. The quadratic model predicted a hardness of 2.5 units, which is smaller than any other of the observed hardness values. This gave us additional confidence in the prediction accuracy of the quadratic model. The omission of point 19 results in a considerable increase in the prediction variance over the region (at point 19 the prediction variance as $2.32\sigma^2$). The $|\underline{X}'\underline{X}|$ (pseudocomponent model: $n = 25$, $\log_{10} |\underline{X}'\underline{X}| = -7.54197$; $n = 26$, $\log_{10} |\underline{X}'\underline{X}| = -5.22432$) increased by approximately 30% as measured by the p^{th} root of the ratio of the $|\underline{X}'\underline{X}|$ of the two designs (Atwood, 1969).

Before leaving this example we should also point out that vertices which are pseudoreplicates (i.e., near neighbors) often occur in constrained mixture problems (Gorman, 1966). This is particularly true when multiple component constraints are present. For example, in Table III points 21, 22, 23, 25, 29, 30, 33, 34, 36, and 38 are pseudoreplicates for points 5, 8, 6, 9, 2, 7, 4, 10, 24, and 26. Pseudoreplicates do not present a problem for CADEX (Kennard and Stone, 1969) because of the nature of the selection algorithm. It has also been our experience that pseudoreplicates do not present a problem in the use of the Wynn and exchange algorithms. The number of pseudoreplicates is also a function of the tolerance used in the CONSIM calculations (see Section 3).

7. CONCLUDING REMARKS

The techniques presented in this paper make it possible to experiment with any mixture system in which constraints have been placed on linear combinations of the components. In general, constraints on the component levels make it more difficult to obtain precise estimates of the coefficients in the model. In some instances unnecessary constraints may be imposed; however, it is the author's opinion that one should not necessarily attempt to talk the experimenter out of imposing constraints, for it is the experimenter who has the technical responsibility for the problem. One point which is often difficult for experimenters to understand that in order to obtain a good model we often have to obtain data points over a wider region than "the process" would typically be operated over. Of course, one would not recommend experiments in regions where it is physically difficult or unreasonable to obtain data. The final design will almost always be a compromise between a set of constraints the experimenter would like to impose on the experiment and the type of data which is needed to obtain a good model for the system.

8. ACKNOWLEDGMENTS

The author wishes to express his appreciation to R. E. Wheeler for his permission to discuss the use of his previously-unpublished CONSIM algorithm. The author would also like to thank L. W. Yeatman for her computer programming assistance during the course of this work and R. J. Kempf for his technical assistance with the example.

BIBLIOGRAPHY

Atwood, C. L. (1969). Optimal and efficient designs of experiments. Ann. Math. Stat., 40, 1570-1602.

Balinski, M. L. (1961). An algorithm for finding all vertices of convex polyhedral sets. J. Soc. Indust. Appl. Math., 9, 72-88.

Box, G.E.P. and Gardiner, C. J. (1966). Constrained designs - Part I. First-order designs. University of Wisconsin Department of Statistics Technical Report No. 89.

Box, G.E.P. and Hunter, J. S. (1957). Multifactor experimental designs for exploring response surfaces. Ann. Math. Stat., 28, 195-241.

Burdet, C. A. (1974). Generating all faces of a polyhedron. SIAM J. Appl. Math., 26, 479-489.

Cornell, J. A. and Good, I. J. (1970). The mixture problem for categorized components. J. Am. Stat. Assn., 65, 339-355.

Elfving, G. (1952). Optimum allocation in linear regression theory, Ann. Math. Stat., 23, 255-262.

Gorman, J. W. (1966). Discussion of "Extreme vertices design of mixture experiments" by R. A. McLean and V. L. Anderson, Technometrics, 8, 455-456.

Gorman, J. W. (1970). Fitting equations to mixture data with restraints on compositions. J. Quality Technology, 2, 186-194.

Kennard, R. W. and Stone, L. (1969). Computer-aided design of experiments, Technometrics, 11, 137-148.

Marquardt, D. W. and Snee, R. D. (1974). Test statistics for mixture models. Technometrics, 16, 533-537.

McLean, R. A. and Anderson, V. L. (1966). Extreme vertices design of mixture experiments. Technometrics, 8, 447-454.

Mitchell, T. J. (1974). An algorithm for the construction of "D-optimal" experimental designs. Technometrics, 16, 203-210.

Motzkin, T. S., Raiffa, H., Thompson, G. L., and Thrall, R. M. (1953). The double description method, Contributions to the Theory of Games II, Princeton University Press, Princeton, NJ.

St. John, R. C. and Draper, N. R. (1975). D-optimality for regression designs: A review. Technometrics, 17, 15-23.

Snee, R. D. (1975). Experimental designs for quadratic models in constrained mixture spaces, Technometrics, 17, 149-159.

Snee, R. D. and Marquardt, D. W. (1974). Extreme vertices designs for linear mixture models. Technometrics, 16, 399-408.

Wets, R.J.B. and Witzgall, C. (1967). Algorithms for frames and lineality spaces of cones. J. of Res., Nat'l Bureau of Standards - B. Math. and Math. Physics, 71B, 1-7.

Wheeler, R. E. (1972). Efficient experimental design. Presented at the Annual Meeting of the American Statistical Association, Montreal, Canada, August 1972.

Wynn, H. P. (1970). The sequential generation of D-optimum experimental designs. Ann. Math. Stat., 41, 1655-1664.

Received July, 1978; Revised December, 1978.

Recommended by D. W. Marquardt, E. I. du Pont de Nemours & Co., Inc., Wilmington, DE.

Refereed by J. W. Gorman, Amoco Oil Company, Naperville, IL.