Statistical Papers 37, 1–32 (1996)

Statistical Papers

An overview of design of experiments

N. R. Draper, F. Pukelsheim

Received: July 7, 1994; revised version: October 17, 1995

An overview of the design of statistical experiments is presented, with special emphasis on response surface designs, block designs, neighbor designs. Applications are mentioned for industrial quality improvement, agricultural experiments, biometry. An outlook towards design optimality concludes the survey.

Key words: Response surface designs, Mixtures experiment designs, Block designs, Neighbor Designs, Designs for industrial quality improvement, Designs for agricultural experiments and biometry, Design optimality, Design algorithms

0 Introduction

The topic of statistical design of experiments could well have an entire encyclopedia devoted to it. In our relatively few pages, we discuss some important areas and concepts, and provide references for further reading. These references will, in their turn, lead the reader further into his/her selected areas, and thus enlarge the scope of what we have provided here. We have divided our material into these sections:

1 Some Historical Comments on Experimental Design

Where and when methods of experimental design began is pure speculation. Perhaps cave-dwellers noticed that certain berries made them ill and certain berries did not. Did they conduct a simple experimental design to compare the two? Perhaps. In long established cities, were different building materials or techniques compared to see which would be longer lasting and most resistant to the elements? Perhaps. In ancient agrarian economies, were the effects of various animal feed stocks compared? Was an experiment performed to check the effects of fresh fruits on scurvy, suffered by sailors who were at sea for long periods? All these sorts of experiments were perhaps performed, but we know nothing of them.

One simple experiment to define the unit of a "foot" was recorded by Jacob Köbel in his book *Geometrei* in 1556, published in Frankfurt am Main:

"Sixteen men, small and large, as they happen to come out of church, shall each place a shoe one before the other .... Whence the total length shall be subdivided into sixteen equal parts, using a compass .... Thus is obtained a measuring rod of sixteen equal parts, made up from sixteen unequal shoes."

An old woodcut showing this is reproduced on the dust jacket of Pukelsheim (1993). We note in this statement: the implication of randomization in the words "as they happen to come"; the assurance of a homogeneous population in the specification of "men" rather than women or children; and the use of the average as the best estimate of the population mean. This is of course an *observational* experiment; a natural phenomenon is observed and recorded. Observational experiments have a long history, particularly in astronomy. Stigler (1986, p. 27) writes of Leonhard Euler (1701–83) analyzing observations "made in the years from 1582 through 1745". Well before that however, probably around 2 000 BC, Stonehenge was built. One school of thought believes that its construction was based on certain astronomical observations.

Nowadays we are more concerned with *controlled* experiments where, ahead of the observations, we deliberately decide to treat different experimental units in different ways, with the idea of seeing which of those ways is best for future work. Various writers have received credit for putting forward early ideas of experimental design, including Francis Bacon (1561–1626) who wrote *Novum Organum*, among other works; Gustav Theodor Fechner (1801–87) who wrote the 1860 book *Elemente der Psychophysik* (see Stigler 1986, p. 242); John Stuart Mill (1806–73) whose views were expressed in an

1843 book *System of Logic*; and Charles Sanders Peirce (1839–1914) some of whose work has been republished after his death, for example, Peirce (1957). Bacon's ideas of Nature being led experimentally to reveal her secrets, and Mill's suggestion that the mind provides the possibilities but that the experiment enables us to actually see the phenomenon, are both early expressions of later presentations of the cycle: conjecture, design, experiment, analysis, conjecture ..., see, for example, Box and Draper (1987, p. 8). A later proponent was Sir Karl Popper (1902–94), known for his 1934 Vienna Circle book *Logik der Forschung*, translated into English as the 1959 book *The Logic of Scientific Discovery*. Popper presented science as a way of straining out the wrong concepts by conducting experiments conceived from theoretical conjectures.

Major progress occurred in the 1920s and 1930s when Ronald Aylmer Fisher, later, Sir Ronald Fisher (1890–1962) did pioneering work at Rothamstead Experimental Station in Hertfordshire, England. Ably assisted by others, including Frank Yates (1902–94), Fisher began to write about basic statistical principles that he introduced into crop and animal experiments being performed at Rothamstead. He wrote two crucially important books (Fisher 1925, 1935) which passed through many editions and provided both inspiration (for their important ideas) and frustration (for their sometimes tightly condensed writing) to many followers. An important and valuable book of statistical tables, needed especially before electronic computers became available was also later produced (Fisher and Yates 1938).

Also in the 1930s began the pioneering mathematical work in experimental design in India, led by P.C. Mahalanobis (1893–1972), and R.C. Bose (1901–87).

After World War 2, an explosion of work on experimental design took place, later fueled by government funds in the "Sputnik" fear years of the 1960s. Among important names of that period are G.E.P. Box (1919–), G. Elfving (1908–84), J.C. Kiefer (1924–81), C.R. Rao (1920–), A. Wald (1902–50) and J. Wolfowitz (1910–81).

The rapid expansion also encouraged use of many mathematical tools such as matrix algebra, combinatoric theory, optimization methods and convex analysis. The burgeoning availability of faster and faster computing equipment played an enormous role, and continues to do so.

One might think that the more complicated experimental design methods became, the more difficult it would be to get answers. Fisher (1926) suggested that just the opposite is true: "Nature ... will best respond to a logical and carefully thought out

questionnaire; indeed, if we ask her a single question, she will often refuse to answer until some other topic has been discussed." This is a heartening message to carry into the modern complications associated with designing experiments!

2 Response Surface Designs

2.1 Response Surface Methodology

Response surface methodology consists of the totality of techniques that apply when numerical data from a physical phenomenon are represented by a "best fit" mathematical surface.

Suppose, for example, that we are studying the effect of changing the *predictor variables* "temperature" (x_1), "pressure" (x_2) and "viscosity" (x_3) on the *response* "yield" (y) of a chemical process. Suppose, at least to practical approximation, that x_1, x_2 and x_3 can be set without error, and that y is the result of a "true value" (η) plus a measuring error (ϵ), that is

$$y = \eta + \epsilon.$$

The error ϵ is assumed to come from a specific error distribution, typically a normal distribution with mean 0 and variance σ^2.

The true relationship $\eta = f(\boldsymbol{x})$ between η and the vector $\boldsymbol{x} = (x_1, x_2, x_3)'$ is usually unknown. It is then standard practice to model (approximate) the true relationship by fitting to the data some member from a function class $g(\boldsymbol{x}, \boldsymbol{\beta})$ depending on an unknown parameter vector $\boldsymbol{\beta}$:

$$\eta = g(\boldsymbol{x}, \boldsymbol{\beta}). \tag{1}$$

For example, a *second-order model* uses a quadratic polynomial of the form

$$g(\boldsymbol{x}, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3, \tag{2}$$

with 10 unknown real parameters β_0, β_1, β_2, β_3, β_{11}, β_{22}, β_{33}, β_{12}, β_{13}, β_{23}. If $\beta_{ij} = 0$ for $i \leq j$ is assumed in (2), a *first-order model* is specified that fits a planar surface.

The standard technique in these situations is to carry out an experiment, that is, choose an *experimental design* consisting of units $u = 1, \ldots, n$ of predictor settings $\boldsymbol{x}_u = (x_{1u}, x_{2u}, x_{3u})'$, and observe the corresponding responses y_u. The data are then employed to estimate numerical values for the parameter vector $\boldsymbol{\beta}$, often via the *method of least squares.*

Usually the simpler, planar model is a first choice. A variety of tests on the surface, and on the unexplained portions of the response, called the *residuals*, can be made. If the plane does not fit well, a second-order surface would be tried. One can also attempt to fit higher order models (cubic, quartic, and so on); in general this would not be sensible. It would be more sensible to attempt to find a *transformation* of the observed response y (for example, the logarithmic transformation $\log y$, or the inverse transformation $1/y$) and attempt to fit a first-order or second-order model to that. Methods exist for finding a "best" transformation. If these methods fail, a *nonlinear model* might be tried, that is, a model nonlinear in the *parameters*.

Readers without prior statistics knowledge should consult Box, Hunter and Hunter (1978) for a general introduction into the statistical design of experiments, or Draper and Smith (1981) for a general treatment of least squares regression methods. More sophisticated readers could see Box and Draper (1987) who in Chapter 8 discuss methods of transformations, or Khuri and Cornell (1987) who in their Chapter 8 outline nonlinear response surface models. Also see Bates and Watts (1988), and Seber and Wild (1989).

2.2 Experimental Domain of the Predictor Variables

In general, there may be k real *predictor variables* $x_1, \ldots, x_k$, also called *experimental conditions*, or *factors*. Together they form the vector $\boldsymbol{x} = (x_1, \ldots, x_k)'$. The *experimental domain* over which this vector $\boldsymbol{x}$ is allowed to vary during the experiment is denoted by $\mathcal{X}$, a subset of k-dimensional Euclidean space. A specific vector $\boldsymbol{x} = (x_1, \ldots, x_k)'$ in $\mathcal{X}$ thus defines the settings for the k predictor variables under which the response y is going to be observed. An *experimental design* then defines n *runs*, that is n sets of experimental conditions $\boldsymbol{x}_u = (x_{1u}, \ldots, x_{ku})'$ for $u = 1, \ldots, n$, at which to observe a total of n responses $y_1, \ldots, y_n$.

For ease of interpretation, the predictor variables x_{ju} are often *coded* or standardized so that $x_{ju} = 0$ means standard operating conditions, while $x_{ju} = \pm 1$ means a chosen unit deviation from standard operating conditions, in either direction. Hence while generally the predictor variables x_{ju} may attain quite arbitrary real values, practical situations often feature two-level designs, $x_{ju} \in \{\pm 1\}$, or three-level designs, $x_{ju} \in \{-1, 0, 1\}$, or designs that stay "close" to two-level designs or three-level designs.

2.3 Full Factorial Designs

Suppose we have k factors to examine and two levels of each, coded to -1 and 1, have been selected. We could then consider an experiment with $n = 2^k$ runs (each run being a specific combination of settings such as $-1, 1, 1, -1, \ldots, -1$ for the k factors). The 2^k experiments so generated form *the full factorial design* 2^k. The 2^k observations are usually recomputed into estimates for the grand mean and $2^k - 1$ "effects", this word encompassing main effects, two-factor interactions, three-factor interactions, ..., k-factor interactions.

For two factors, the full factorial design 2^2 consists of the four runs $(x_1, x_2) = (-1, -1)$, $(1, -1)$, $(-1, 1)$, $(1, 1)$, and the points lie at the corners of a square in the (x_1, x_2) space. For the 2^2 design with factors numbered **1** and **2**, and observations y_1, y_2, y_3, y_4 respectively, there exist estimates for two main effects **(1)** and **(2)**, and a two-factor interaction **(12)**. Suppose we were to specify a model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon \tag{3}$$

for this situation, and fit the data by least squares. Then the estimates b_0, b_1, b_2, b_{12} of the corresponding parameters $\beta_0, \beta_1, \beta_2, \beta_{12}$ would be

$$\begin{aligned} b_0 &= \overline{y}, \text{ the average of the four observations,} \\ b_1 &= (-y_1 + y_2 - y_3 + y_4)/4 = \tfrac{1}{2}(\mathbf{1}), \\ b_2 &= (-y_1 - y_2 + y_3 + y_4)/4 = \tfrac{1}{2}(\mathbf{2}), \\ b_{12} &= (y_1 - y_2 - y_3 + y_4)/4 = \tfrac{1}{2}(\mathbf{12}). \end{aligned}$$

The quantities b_1, b_2, and b_{12} estimate half of the corresponding main effects and interactions, as they are usually defined.

Comparing (3) with (2), we see that a two level factorial design permits us to estimate only main effects and interactions but not pure quadratic parameters such as β_{11} and β_{22}. Estimation of these is possible only if additional levels are included in the design in one of the many permissible ways.

Other types of full factorial designs can be constructed. The 3^s series, the mixed $2^k 3^s$ series, and so on, are available. The two-level series is probably the place to start a study of factorial designs, however.

2.4 Fractional Factorial Designs

Suppose we wish to investigate four factors but are restricted (by funds, or by available raw material for the runs) to only eight runs, so that the full factorial design 2^4 which needs 16 runs is out of the question. What can we do? A full 16 run design permits us to fit a factorial model of type

$$\begin{aligned} y = \beta_0 \quad & + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \\ & + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4 \\ & + \beta_{123} x_1 x_2 x_3 + \beta_{124} x_1 x_2 x_4 + \beta_{134} x_1 x_3 x_4 + \beta_{234} x_2 x_3 x_4 \\ & + \beta_{1234} x_1 x_2 x_3 x_4 + \epsilon. \end{aligned} \tag{4}$$

In most practical circumstances we would be surprised if all the terms of (4) were needed. It would generally be thought that the coefficients with three or four subscripts are not relevant, that is, their effects are small enough to neglect. Even if we drop these five terms, however, 11 terms remain. It is clear, then, that a forced reduction to eight runs will not enable us to estimate β_0, all main effects (proportional to $\beta_1, \beta_2, \beta_3, \beta_4$), and all two-factor interactions (proportional to $\beta_{12}, \beta_{13}, \beta_{14}, \beta_{23}, \beta_{24}, \beta_{34}$).

What is the best we *can* do, however? We can get estimates of β_0, all the β_i, and sums of pairs of the β_{ij} by choosing a ***half fraction*** of the 2^4 design, called a 2^{4-1} design. The appropriate choice for our stated objective is to divide the runs of the 2^4 design into two halves, in the following way:

- Design (a) is that half fraction for which the runs have a product $x_1 x_2 x_3 x_4 = 1$. These are $(-1,-1,-1,-1)$, $(-1,-1,1,1)$, $(-1,1,-1,1)$, $(-1,1,1,-1)$, $(1,-1,-1,1)$, $(1,-1,1,-1)$, $(1,1,-1,-1)$ and $(1,1,1,1)$.

- Design (b) consists of those runs which have the product $x_1 x_2 x_3 x_4 = -1$, namely $(-1,1,1,1)$, $(1,-1,1,1)$, $(1,1,-1,1)$, $(1,1,1,-1)$, $(1,-1,-1,-1)$, $(-1,1,-1,-1)$, $(-1,-1,1,-1)$ and $(-1,-1,-1,1)$.

Either half fraction could be used for the design to be performed.

Suppose we pick design (a). Then, it turns out, we can fit the following model with eight coefficients, derived from (4) and from our assumptions and choices:

$$\begin{aligned} y = \beta_0 \quad & + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \\ & + (\beta_{12} + \beta_{34}) x_1 x_2 + (\beta_{13} + \beta_{24}) x_1 x_3 + (\beta_{14} + \beta_{23}) x_1 x_4 + \epsilon. \end{aligned} \tag{5}$$

The implication of (5) is that we can still estimate $\beta_1, \beta_2, \beta_3$, and β_4 with our half fraction designs (a) and (b). But because of $x_1x_2x_3x_4 = 1$ we have $x_1x_2 = x_3x_4$ for every run. Therefore β_{12} and β_{34} are no longer separately estimable, only their sum is, and similarly for the other pairings. We say that the estimates of β_{12} and β_{34} are *confounded*, or that they are *aliases* of each other.

If the other half fraction design (b) is used, a similar argument shows that now $x_1x_2 = -x_3x_4$ and so on, and the estimates obtained correspond to the coefficients in the model

$$\begin{aligned} y = \beta_0 \quad & + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \\ & + (\beta_{12} - \beta_{34}) x_1 x_2 + (\beta_{13} - \beta_{24}) x_1 x_3 + (\beta_{14} - \beta_{23}) x_1 x_4 + \epsilon. \end{aligned} \tag{6}$$

An important feature of a *fractional factorial design* is how it projects into fewer dimensions. Consider the eight run design (a) given above, and indicated by the black dots in Figure 1. What happens if we decide, as a result of examining the response data, that x_4 is not effective in influencing the response? Then, in the other three variables, we have a complete 2^3 design. One must imagine the two cubes in Figure 1 sliding together and coalescing as we abandon x_4 as a variable. There then is a black dot at all eight corners. Remarkably, this is true for any three xs after one x is abandoned. What happens to the black dots if the design is collapsed in the x_1 (or x_2 or x_3) direction? For the x_1 direction collapse, the two cubes of Figure 1 are each compressed from left to right to form two full factorial designs 2^2 in x_2 and x_3, one at each level of x_4. Thus the *projected design* is the complete factorial design 2^3 in the variables 2, 3, and 4.

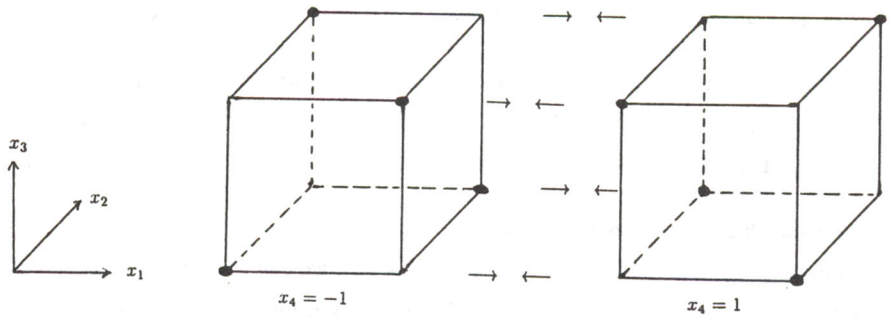

Figure 1. A 2^{4-1} design with $x_1x_2x_3x_4 = 1$. The origin is at the center of the four dimensional space, shown as two three-dimensional slices for $x_4 = -1$ and $x_4 = 1$.

Fractional factorial designs are usually denoted by the notation 2^{k-p}_R, a 2^{-p} fraction of the complete factorial design 2^k. The *resolution* R enables us to assess the degrees of alias involvement; roman numerals are the customary notation. For example the design of Figure 1 is a 2^{4-1}_{IV} design, that is, of resolution four. This means that at least one main effect (in fact *all* in this design) is aliased with at least one three-factor interaction, or that at least one (all in this particular design) two-factor interaction is aliased with another two-factor interaction. If they do not consume too many runs, designs of resolution V that alias main effects with four-factor interactions, or two-factor interactions with three-factor interactions are more desirable than designs of resolution III or IV.

In the branch of combinatorial theory, the 2^{k-p}_R designs are described as *orthogonal arrays*, and resolution then turns into the *strength* of the array, see Raghavarao (1971).

2.5 Central Composite Designs

Central composite designs, which are appropriate for fitting *second-order models*, have a pleasing geometrical structure. First proposed by Box and Wilson (1951), these designs have stood the test of time.

They consist of a combination of three point sets: (a) The two-level full factorial design 2^k, or a 2^{k-p}_R fractional factorial design of resolution R, with points $(\pm, \pm 1, \ldots, \pm 1)$. This set is called a *cube*, even when it literally is not. Usually the resolution is V or higher. (b) The second point set consists of $2k$ axial points $(\pm\alpha, 0, \ldots 0), (0, \pm\alpha, 0, \ldots, 0), \ldots, (0, \ldots, 0, \pm\alpha)$ where α is to be chosen. This set is called a *star*. It can be regarded as a combination of k *vary-one-factor-at-a-time* designs. (c) Finally a number n_0 of center points $(0, 0, \ldots, 0)$ is added. Typically these would be at the most recent "best" conditions, or at some standard operating conditions.

A great advantage of a central composite design is that it permits an initial study made via a 2^{k-p}_R design (perhaps with center points), to be extended to a second-order model exploration with only star points and perhaps more center points added. Moreover, these designs enjoy excellent design properties over a wide range of design criteria. Figure 2 shows a central composite design and its component pieces for $k = 3$.

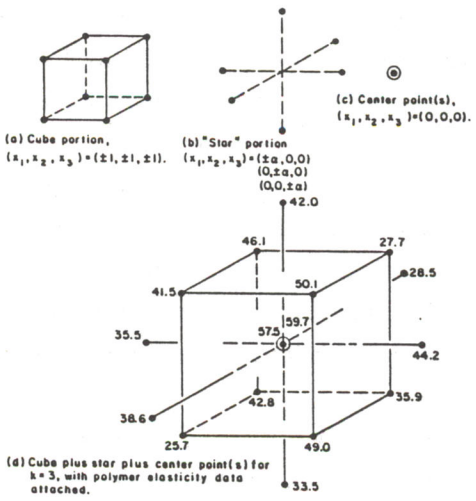

Figure 2. A composite design in three dimensions ($k = 3$). Taken from Box and Draper (1987, p. 306).

2.6 Orthogonal Blocking and Rotatability

Two important desirable features of second-order response surface designs are *orthogonal blocking* and *rotatability*. When the runs of the design must be split up into sections (days, shifts, batches of raw material) that may affect the level of the response from section to section, the model fitting can be safeguarded by dividing the design into orthogonal blocks. This enables the model to be fitted without worrying about blocks, and also permits the removal of a component for block variation from the subsequent analysis of the surface fit.

For central composite designs, orthogonal blocking into two blocks (and often more, depending on the values of k and p of the star portion 2_R^{k-p}) can be achieved by suitable choice of α and n_0. For example, if $k = 4, p = 0$, and $n_0 = 3s$, where s is an integer (typically 1, 2, or 3 in this context), we can make three orthogonal blocks as follows:

- Block (a): A 2_{IV}^{4-1} design with $x_1x_2x_3x_4 = 1$ plus s center points.

- Block (b): A 2_{IV}^{4-1} design with $x_1x_2x_3x_4 = -1$ plus s center points.
- Block (c): Eight axial points at distances $\pm\alpha$, where $\alpha = 2$, plus s center points.

There are many other choices.

Rotatability ensures that the response will be predicted with the same accuracy at all points in the x-space that are equidistant from the origin. For rotatability of the specific design given above, the condition is $\alpha = 2$, and so the design given is both rotatable *and* orthogonally blocked. (This is not achievable in general however.)

For other facets, see Box and Hunter (1957), and Draper and Pukelsheim (1991, 1994). Near rotatability is discussed by Khuri (1988), and Draper and Pukelsheim (1990). Intuitive reasoning and geometric elegance in higher-order models are difficult to envision. This calls for a more formal discussion of rotatability properties as, for example, in Draper, Gaffke and Pukelsheim (1991), and Draper and Pukelsheim (1994).

2.7 An Example

An example of a classic response surface study conducted by Box and Youle (1955) can be found in Box and Draper (1987, p. 358). It is seen that the highest yield of 60 is obtained on a *plane*. This sort of redundancy is most desirable, because it permits many choices of setting time, concentration, and temperature so as to achieve a maximum yield.

3 Mixtures Experiment Designs

3.1 Mixtures Experiments

Response surface experiments often involve *mixtures* of ingredients. For example, fuels can consist of a mixture of petroleum and various additives; fish patties may contain several types of fish; a fruit juice drink may consist of a mixture of orange, pineapple, and grapefruit juices; or a regional wine may be blended from several grape varieties. This adds a restriction to the problem, often of type

$$x_1 + x_2 + \cdots + x_k = 1, \tag{7}$$

where the predictor variables are proportions, $0 \leq x_j \leq 1$.

Suppose there are $k = 3$ ingredients. The second-order model (2) is not feasible in this form, because although we have a three-dimensional x-space, the space of the

experimental domain is only two-dimensional; see Figure 3. In the usual least squares regression fit formulation the estimate is $\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, where $\boldsymbol{X}$ is the model matrix, $\boldsymbol{y}$ is the response vector, and $\boldsymbol{b}$ estimates the parameter vector $\boldsymbol{\beta}$ in the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. In the present case the matrix $\boldsymbol{X}'\boldsymbol{X}$ is singular and its inverse does not exist.

This difficulty can be overcome by either of two (essentially the same) methods:

- Method (a). Transform the k original x-coordinates to $k-1$ new z-coordinates. (This is regarded as somewhat tedious to do, especially for $k \geq 4$.)
- Method (b). Use the restriction (7) to transform the model symmetrically into *canonical form.*

Method (b) was pioneered by Scheffé (1958, 1963), and is now illustrated for $k = 3$. Since $x_1 = 1 - x_2 - x_3$, we have $x_1^2 = x_1 - x_1x_2 - x_1x_3$, and similarly for x_2^2 and x_3^2. Substituting these relations and $\beta_0 = \beta_0x_1 + \beta_0x_2 + \beta_0x_3$ in (2), gathering like terms, and renaming coefficients gives

$$h(\boldsymbol{x}, \boldsymbol{\alpha}) = \alpha_1x_1 + \alpha_2x_2 + \alpha_3x_3 + \alpha_{12}x_1x_2 + \alpha_{13}x_1x_3 + \alpha_{23}x_2x_3. \tag{8}$$

This is the canonical form of the second-order model. For planar models no intercept term is used, and (8) applies with $\alpha_{12} = \alpha_{13} = \alpha_{23} = 0$.

While the usual response surface designs can be employed for mixtures experiments, special designs linked to the canonical form were suggested by Scheffé. For these, and for a comprehensive treatment of the mixtures area in general, see Cornell (1990).

3.2 Mixtures Experiments in Restricted Regions

In many mixtures experiments, exploration of the entire mixtures domain (for example, for $k = 3$, this is the triangle in Figure 3) is not feasible. A common type of restriction is that each ingredient has a lower bound, such as $x_1 \geq a, x_2 \geq b, x_3 \geq c$ in Figure 4(a). Obviously this produces a restricted space of the same shape as the original mixtures space, so that the design considerations can be translated to the smaller space exactly. When upper and lower bounds are specified, $a_2 \geq x_1 \geq a_1$ and so on, the domain shape is different from the original one, as shown in Figure 4(b).

One simple way of choosing some experimental points is to select all or some of the extreme vertices of the domain in an optimum way via some selected criterion. The extreme vertices in our example are the black dots in Figure 4(b), the criterion used might be any of those described in Section 8. Other points, centroids of the boundaries, for example, can be added. For more details, consult Cornell (1990).

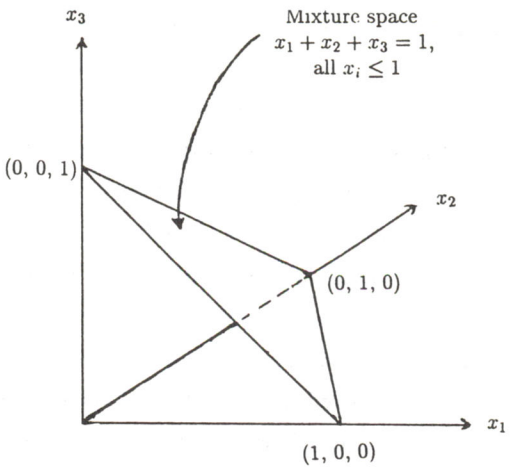

Figure 3. A two-dimensional experimental mixtures space $x_1 + x_2 + x_3 = 1$ embedded in the three-dimensional $(x_1.x_2, x_3)$ space. Only the region for which $0 \leq x_j \leq 1$ is valid, as shown.

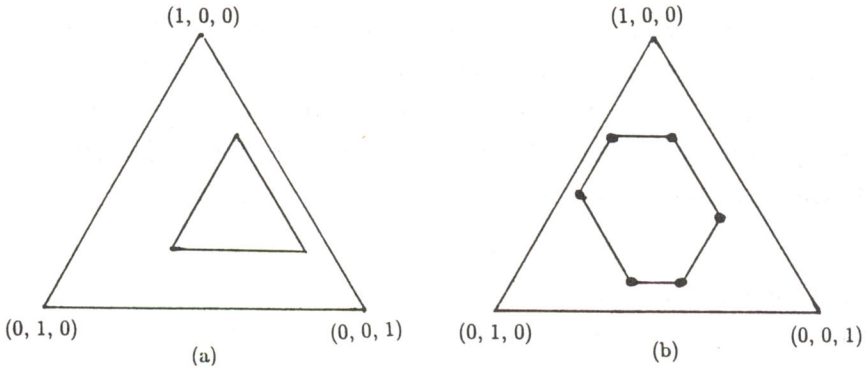

Figure 4. Restrictions in the $k = 3$ mixtures space, pulled from Figure 3 and shown in two dimensions.

4 Designs for Industrial Quality Improvement

4.1 Evolutionary Operation

Response surface designs (including mixtures designs where appropriate) are of basic importance for *industrial experimentation*. The practical choice of a design is guided by many (and partly conflicting) goals, see the discussion in Chapter 15 of Box and Draper (1987).

Designs for first-order models are often used as *screening designs* in situations where many factors need to be examined but few are expected to be of consequence. This has sparked much interest in the geometrical properties of design projections into fewer dimensions. The two-level fractional factorial designs are especially useful in this regard, because they project down to fractional and full factorial designs, perhaps with replications, in the factors that are judged to be the effective ones.

Screening in this way is useful in laboratory or pilot plant work. It is impractical if there is need to study, for example, a major chemical process which cannot be stopped for experimental work but must go on producing salable product. This means that the factors of interest cannot be varied at will; at most only small changes can be made. If only small changes are made, however, it will be impossible to detect the small differences in response that will occur in individual runs. The solution to this is to take advantage of the fact that the standard deviation of a comparison $\overline{y}_1 - \overline{y}_2$, between two independent averages, each based on n observations, is $2\sigma/n^{1/2}$, where σ is the standard deviation of a single run. By choosing a factorial pattern in only two or three factors, and repeating this pattern as many times as necessary, small real effects can be detected. Effective variables are followed up in the factor space to improved response, ineffective variables are dropped and new candidates are tried in their place. This leads to continuous improvement of product without loss of income and at a low cost, typically trivial compared to the improvements received. This system is called *Evolutionary Operation*, and is described by Box and Draper (1969).

4.2 Analysis of Variability

The work of Taguchi (1987), and the methods he suggested to Japanese industry have created some new initiatives, mentioned here and in Sections 4.3 and 4.4. See also Shoemaker, Tsui and Wu (1991), Nair (1992), Abt and Pukelsheim (1995), for example.

Although interest is usually in the mean yield of a process, the size of the *process*

variability is often crucial. If the experiment performed provides for repeated observations on each setting of the factors that appears in the design, a linear model may be fitted to the logarithm of the estimated per setting variance. Thus a properly designed experiment may be analyzed dually, modeling the process mean and the process variability.

The distributional assumptions underlying a linear model will no longer be fully satisfied when applied to the log variances. However, they often hold true to some degree of approximation, since the logarithm serves as a ***variance stabilizing transformation*** when the variance is proportioned to the mean. In a more general context, transformations other than log, as proposed by Box and Cox (1964) may also be applied.

4.3 Signal Factors versus Control Factors

The dual analysis just described leads to a data dependent classification of the predictor variables. The first class, called *signal factors*, consists of those factors that contribute significantly to the process mean while being nonsignificant for process variability. The engineer can then alter the signal factors to bring the process mean on target, or maximize or minimize it, as appropriate, without affecting process variability.

The second class are the *control factors*. They are significant only for process variability, and hence will be set so as to stabilize the process and minimize its variability. Control factors should be clearly designated as such, to prevent their accidental use by process operators and supervisors.

The remaining factors are the *nuisance factors*, used to model the manufacturing process without being significant for the process mean nor process variability. If no other goal intervenes, the nuisance factors are set to minimize production costs. Case studies are presented by Abt, Mayer and Pukelsheim (1995), and Weihs, Berres and Grize (1995), for example.

4.4 Noise Factors

A designed experiment should also aim to incorporate *noise factors* and study their effect on the product. For example, humidity, surrounding temperature or exposure to daylight may be vital for a product to function properly in the user environment, but would not generally occur as predictor variables for the production process. With

noise factors included, the analysis of the experiment hopefully points to settings for the predictor variables that would also perform well over the range of values that the noise factors may attain.

As a result, an experiment that includes *process factors* (that is predictor variables for the process under study) as well as noise factors leads to a product that is robust not just against manufacturing imperfections, but also against random noise at the user's site, such as variations in temperature, humidity and physical treatment. For a review of this topic see Grize (1995).

5 Block Designs

5.1 Balanced Incomplete Block Designs

Block designs are appropriate when the predictor variables do not vary continuously, but attain only a finite number of levels each. For example, in agricultural experiments a finite number of crop *varieties (treatments)*, $i = 1, \ldots, v$, are planted within a finite number of areas of similar fertility (*blocks*), $j = 1, \ldots, b$. A block is called *complete* when it features all treatments, otherwise it is called *incomplete*.

The allocation of treatments to blocks is called a *block design*. A block design with incomplete blocks can be *balanced* (so that any two treatments concur in any blocks the same number of times, λ), or *partially balanced* (when a few different concurrence numbers are admitted, $\lambda_1, \ldots, \lambda_c$, each of them defining an *associate class* of treatments).

Table 1 shows a balanced design in which 5 treatments occur in each of 5 blocks. In fact, the design is derived from 5×5 *Latin square*.

A *balanced incomplete block design* (BIBD) is shown in Table 2, for 21 observations on 7 treatments in 7 blocks. A BIBD is usually quoted with a list of pertinent quantities, the *number of varieties* or treatments, $v = 7$, the *number of blocks*, $b = 7$, the *treatment replication number*, $r = 1$, the *block size*, $k = 3$, and the *concurrence number*, $\lambda = 1$, besides the *number of observations*, $n = 21$. Of these 6 quantities, three are actually sufficient to set all of them, since they satisfy the three identities $n = rv = kb$ and $\lambda = r(k-1)/(v-1)$.

Table 3 shows a partially balanced incomplete block design in which treatments 1 and 4, 2 and 5, 3 and 6 concur twice each within a block. All other pairs occur together just once. These groupings thus define the two *associate classes* of the design.

1	2	3	4	5
3	4	5	1	2
5	1	2	3	4
2	3	4	5	1
4	5	1	2	3

Table 1. A 5×5 Latin square design, each column being a block. The numbers $1, \ldots, 5$ indicate the five different treatments.

1	2	3	4	5	6	7
2	3	4	5	6	7	1
4	5	6	7	1	2	3

Table 2. A balanced incomplete block design for 7 treatments in 7 blocks of size 3, each column being a block. Any two treatments concur in exactly one block.

1	2	3	4	5	6
2	3	4	5	6	1
4	5	6	1	2	3

Table 3. A partially balanced incomplete block design for 6 treatments in 6 blocks of size 3, each column being a block. Pairs (1, 4), (2, 5), and (3, 6) are first associates, $\lambda_1 = 2$, while all other pairs are second associates, $\lambda_2 = 1$.

5.2 Cyclic Designs

A particular subset of block designs can be generated by the method of cyclic substitution, called *cyclic designs*. Two examples of this have already been given in Tables 2 and 3 where the initial block (1, 2, 4) is developed by adding one to each treatment number and reducing modulo the number of treatments (7 in Table 2, 6 in Table 3) whenever the number of treatments is exceeded. For more on these designs, see John (1987).

5.3 Two-Way Classification Models

The statistical model in which data from block design experiments are often analyzed is

$$y_{ijk} = \tau_i + \beta_j + \epsilon_{ijk}, \qquad k = 1, \ldots, n_{ij}, \tag{9}$$

where n_{ij} is the number of experimental observations taken with variety i in block j. In the examples of subsection 5.1, all designs are *binary*, that is $n_{ij} = 1$ or $n_{ij} = 0$.

The model parameter τ_i is the *treatment effect* of treatment i, the parameter β_j is the *block effect* of block j. Although the total number of parameters is $v + b$, only $v + b - 1$ linear functions of them are estimable.

A linear function $\sum_{i=1}^{v} c_{1i}\tau_i + \sum_{j=1}^{b} c_{2j}\beta_j$ is called a *contrast* when the coefficient sum vanishes, $\sum_{i=1}^{v} c_{1i} + \sum_{j=1}^{b} c_{2j} = 0$. There is a total of $v - 1$ *treatment contrasts* ($\sum_{i=1}^{v} c_{1i} = 0 = c_{21} = \cdots = c_{2b}$), and a total of $b - 1$ *block contrasts* ($\sum_{j=1}^{b} c_{2j} = 0 = c_{11} = \cdots = c_{1v}$). These and the *grand mean* ($c_{1i} = 1 = c_{2j}$) span the space of *estimable functions.*

Two-way classification models are employed when it is thought necessary to incorporate block effects into the model, in addition to the treatment effects that are of primary interest. Accordingly the parameters of interest are formed by the treatment contrasts, while block contrasts are treated as nuisance parameters. The *information matrix for the treatment contrast* is

$$C = \begin{pmatrix} r_1 & & \\ & \ddots & \\ & & r_v \end{pmatrix} - N \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_b \end{pmatrix}^{-1} N', \tag{10}$$

where $N = ((n_{ij}))$ is the $v \times b$ *incidence matrix* that consists of the frequencies n_{ij}, and $r_i = \sum_{j=1}^{b} n_{ij}$ is the replication number for treatment i, while $s_j = \sum_{i=1}^{v} n_{ij}$ is the size of block j. The matrix C is also known as *the C-matrix* of the design N, and is the key quantity to study the statistical properties of a block design.

Multi-way classification models also come under the heading of *models for the elimination of multi-way heterogeneity.* They are discussed, for example, in Cheng (1978), Krafft (1978), Shah and Sinha (1989), Pukelsheim (1993).

6 Neighbor Designs

6.1 Correlation in Time

When observations are taken over time, the assumption that the additive error terms ϵ_{ijk} in (9) are independent (or at least uncorrelated) becomes questionable. A correlation structure that takes into account the time sequential nature would seem more appropriate. As a consequence, a design not only stipulates how often each treatment is going to be observed, but also in which sequence the treatments ought to be arranged.

The more complicated the assumed correlation structure of the observed yield, the more sophisticated is the rule concerning which treatment is allowed to appear as a first-order, second-order or higher-order neighbor of any other treatment. For this reason, designs of this sort often come under the heading of *nearest neighbor designs*. As a rule of thumb, designs with a balanced neighbor structure prove to be better.

Due to the more sophisticated model assumptions, the results that are generally available on nearest neighbor designs are more restrictive. An early paper on the subject is Williams (1952). Many results apply only in some asymptotic sense that varies with the precise nature of the assumptions. Exact results are difficult to come by and are often obtained only under severe combinatorial restrictions (such as two or three treatments, or blocksize two or three, etc.).

6.2 Spatial Correlation

The statistical model becomes yet more demanding when the response on different plots is correlated in space. This can happen in a field experiment where the fertilizer that is applied to one plot may be flushed more into the plots that are neighboring in one direction, rather than those in another direction. The nearest neighbor pattern must then also account for the spatial location of plots. See Martin (1986) for more details and further references.

Another issue arises in mineral exploration, for example. Not only is spatial correlation present but there is also no need for replication, as it makes sense to draw either just one, or else no observations at any one experimental site. (A similar restriction occurs in medical experiments where the effect of a drug on any one patient is observed once, or the patient is in a control group that does not receive this drug.) For a textbook treatment of these and related problems see Cressie (1991). Examples where such deviations from classical design assumptions occur in industrial experiments are given by Grondona and Cressie (1991).

7 Designs for Agricultural Experiments and Biometry

7.1 Historical Roots

Agricultural and biometrical experimentation have been going on for centuries. In terms of fully recorded and continuing agricultural data, the best examples are probably from the *Rothamstead Agricultural Research Station* in England. The first field trial (Broadbalk) was started in 1843 and has continued to the present day, a span of more than 150 years. Other experiments started soon after are also still going.

Formal statistical involvement at Rothamstead dates from the arrival of R.A. Fisher in 1919, when a Statistics Department was formed. *Factorial designs* were developed there in the 1920s and 1930s, largely by R.A. Fisher and his coworkers who included F. Yates.

7.2 Agricultural Experiments

Agricultural experiments may be of several kinds. For example, they may involve the testing of crops which are sown on a series of plots marked out on a large field, or perhaps on several fields, sometimes widely separated. Alternatively, fertilizers or weed killers might be tested rather than the crops themselves. Other possible treatments studied might include pesticides, or methods for applying them, tillage methods, irrigation methods, biocontrol treatments, crop rotation schemes, and so on. The basic objective is usually a comparison of treatments in order to determine which is the best or, at least, which subset is better as a group. Such experiments usually take a "season" which may be a year. For crops (such as raspberries) that fruit only in their second year, more time would be needed to obtain a complete plant cycle.

7.3 Experiments in Biometry

Biometry is the statistical analysis of biological observations and phenomena. *Biometry experiments* could involve animals (diets, diseases, cures for diseases), plants (plant pathology, medical applications) or bacteria. For example, it might be desired to compare several diets fed to pigs or cows, or to test possible cancer cures on mice.

The units of an experiment (the field plots, the fertilizer recipients, the cows, the mice) have their own variation, the sort that would result if all units received the exact same treatment. It is the knowledge (such as it may be) of this variation that affects

the choice of replications, blocking, and randomization, in an attempt to minimize the way unit variation affects differences between treatments.

The treatments applied may include a control (often, no treatment), may be qualitative (and perhaps different levels of a particular treatment, such as different dilutions of a solution) or quantitative (separate wheat varieties, for example). The treatments used may be specific (*fixed effects*) or a sample of treatments from a population (*random effects*). Treatments must usually be applied to a variety of different types of units so that conclusions are valid over a variety of conditions, but they must also be applied in a fashion that will enable differences between treatment effects to be well estimated.

To decide how many units should be allocated to test each treatment requires knowledge of the sizes of differences that need to be detectable between treatments. Many more runs are needed to detect small differences and an experiment which provides too few is useless. Conversely to use a large experiment when only large differences need to be detected is wasteful. Sometimes some comparisons need to be made more accurately than others.

7.4 Blocking and Randomization

In agriculture and biometry, *blocking* is a basic tool and various types of block designs are in standard use (see Section 5). The experimenter attempts to group units which are expected to have small unit-to-unit variation together in a block and to allocate all the treatments to that block. For example, a litter of animals could form a block, or a group of apple trees close to one another could. If the block size is too small to accommodate all the treatments, incomplete block designs could be used. A basic rule in all experimental design work is to block the design to take care of known variation and then, when no further blocking is possible, to allocate the treatments randomly within a block.

While blocking controls heterogeneities that are known, *randomization* helps to guard against effects of unknown heterogeneity. So, for example, having chosen several litters of animals for the blocks, one could allocate the diets to be tested randomly within the litter. This randomization has important aspects for the analysis of such experiments as it provides the randomization distribution which, while seldom used itself, leads to the justification for applying tests related to assumptions that the errors are normally distributed. See Bailey (1983, 1985) for more details on randomization.

8 Optimality of Designs

8.1 Symmetry versus Optimality

The choice of an experimental design is made according to the properties desired by the experimenter. One way of selecting a design is to concentrate on its symmetry properties, such as *balancedness* in the case of block designs, or *rotatability* in the case of response surface designs.

Another approach is to choose a design according to whether it maximizes or minimizes a specific design criterion, within a class of competing designs. Such criteria often involve some function of the variance of the predicted value $\hat{y}$ of η at any specified value of the predictor vector $\boldsymbol{x}$, or some function of the variances of the estimates of the parameter vector $\boldsymbol{\beta}$. It is also possible to modify such criteria by allowing for the effect of *bias* introduced by fitting a model that is (perhaps only slightly) wrong.

Bandemer (1977) and Bandemer and Näther (1980) are encyclopedic volumes, with extensive tabulations of many important design families. Pázman (1986) emphasizes the underlying optimality aspects that are common to the various criteria. Atkinson and Donev (1992) is concerned with determinant optimal designs in practical circumstances. Pukelsheim (1993) treats design optimality for a general family of criteria called information functions.

8.2 Optimal Design Problem

All optimality criteria relate in some way or other the performance of a design to the model within which the data are evaluated. For a linear regression model we have

$$y(\boldsymbol{x}) = f(\boldsymbol{x})'\boldsymbol{\beta} + \epsilon, \tag{11}$$

with uncorrelated and homoscedastic errors ϵ. The pertinent quantity is the regression function f that maps the *experimental domain* $\mathcal{X}$, a subset of the m-dimensional Euclidean space, into the k-dimensional Euclidean space. Thus the components of $\boldsymbol{x}$ in (11) represent m predictor variables or factors, in a model where the parameter vector $\boldsymbol{\beta}$ has k components.

A design that calls for n_i replicates of the predictor vector $\boldsymbol{x}_i$ for $i = 1, \ldots, \ell$, is evaluated through the per observation *moment matrix*,

$$M = \sum_{i=1}^{\ell} \frac{n_i}{n} f(\boldsymbol{x}_i)\, f(\boldsymbol{x}_i)', \tag{12}$$

where $n = \sum_{i=1}^{\ell} n_i$ is the total number of observations. In the terminology of Section 2.2, a possible enumeration of the runs $u = 1, \ldots, n$ is given by

$$\underbrace{x_1, \ldots, x_1}_{n_1 \text{ times}}, \quad \ldots, \quad \underbrace{x_\ell, \ldots, x_\ell}_{n_\ell \text{ times}}. \tag{13}$$

The problem of finding an *optimal design for n observations* (rather, an optimal moment matrix in the set of moment matrices originating from such designs) leads to a discrete optimization problem for which the solution may change from n to $n+1$.

A smoother version of the problem, with stronger results, is obtained by considering *(approximate) designs* ξ (also called measure designs) which, by definition, are probability distributions with finite support on the experimental domain $\mathcal{X}$. The set of all approximate designs on $\mathcal{X}$ is denoted by Ξ.

A design ξ with ℓ support points x_i has a moment matrix M given by

$$M = \sum_{i=1}^{\ell} \xi(x_i) f(x_i) f(x_i)' = \int_{\mathcal{X}} f(x) f(x)' \, d\xi. \tag{14}$$

This matrix consists of the uncentered second-order moments of the regression function f that determines the model, under the design ξ that defines the experiment.

The search for an *optimal design* concentrates on the determination of an *optimal moment matrix*, within a class of competing moment matrices $\mathcal{M}$. In general, the optimum is sought in a subclass $\mathcal{M}$ of the class of all moment matrices $M(\Xi)$ that arises from the approximate designs in Ξ. The set $\mathcal{M}$ is called the *set of competing designs*.

The final step to specify an instance of the design problem is to single out a criterion function ψ, which maps the nonnegative definite moment matrices into the real numbers in such a way as is in accordance with the underlying statistical model. Given a criterion function ψ, and given a set $\mathcal{M}$ of competing moment matrices, the optimal design problem then reads:

$$\text{Maximize} \quad \psi(M) \quad \text{subject to} \quad M \in \mathcal{M}.$$

A moment matrix $M \in \mathcal{M}$ that attains the optimum is called *ψ-optimal in $\mathcal{M}$*. If M is the moment matrix of ξ, then the design ξ itself is called ψ-optimal within the set of those designs that have a moment matrix lying in $\mathcal{M}$.

8.3 Optimality Criteria

The optimality criteria ψ that are of statistical interest mostly concentrate on a *parameter system of interest* $K'\beta$, defined by a known $k \times s$ coefficient matrix K of full column rank s. Accordingly the $k \times k$ moment matrix M is reduced to $C_K(M)$, the $s \times s$ *information matrix for* $K'\beta$, and the criterion factorizes, $\psi(M) = \phi(C_K(M))$. The design problem thus takes the form:

$$\text{Maximize} \quad \phi\big(C_K(M)\big) \quad \text{subject to} \quad M \in \mathcal{M}. \tag{15}$$

The *information matrix mapping* C_K is studied in detail in Chapter 3 of Pukelsheim (1993).

It remains to select a criterion ϕ all of which aim in some way or other to make the information matrix $C = C_K(M)$ "large". This captures the idea that an experiment is designed so as to achieve maximum information on the underlying unknown parameters. Alternatively these criteria can be interpreted so as to minimize the variances of the least squares estimates of the parameter system of interest, $K'\beta$.

The most prominent criterion is the *determinant criterion*, $\phi(C) = (\det C)^{1/s}$, minimizing the volume of the joint confidence ellipsoid for $K'\beta$. Other criteria are the *average-variance criterion* $\phi(C) = (\frac{1}{s} \operatorname{trace} C^{-1})^{-1}$, minimizing the sum of the variances of the estimates for $K'\beta$, or the smallest-eigenvalue criterion, $\phi(C) = \lambda_{\min}(C)$, minimizing the largest principal axis of the confidence ellipsoid.

For the purpose of Bayes designs, and designs that simultaneously perform well in a number of alternative models, it is useful to admit more general criteria that form a class closed under averaging and composition. This is achieved with *information functions* as discussed in Chapter 5 of Pukelsheim (1993).

8.4 Kiefer–Wolfowitz Equivalence Theorem

The type of results that characterize the solutions to the optimal design problem through necessary and sufficient conditions for design optimality are generally called *Equivalence Theorem*. This follows the lead of Kiefer and Wolfowitz (1960) who presented the first such result for the important case of determinant optimality, for the full parameter vector β, in the maximal set $M(\Xi)$ of all moment matrices. Their result states that a candidate matrix M is determinant optimal for β in $M(\Xi)$ if and only if

$$f(x)'M^{-1}f(x) \le k \qquad \text{for all } x \in \mathcal{X}. \tag{16}$$

Although this is a statement on *moment matrices* it implies important information about *experimental designs* that attain the optimal moment matrix, in that equality holds whenever x is a support point of an optimal design, and that the weights of the optimal design are bounded by $1/k$.

Equivalence theorems that pertain to the general optimal design problem are complicated by the fact that the optimality criterion may fail to be differentiable, and that optimal moment matrices need not be nonsingular. Results of a varying degree of sophistication are available in Bandemer (1977), Pázman (1986), and Pukelsheim (1993).

Let M be a moment matrix that maximizes a criterion of the form $\phi \circ C_K$, where $K'\beta$ determines an s-dimensional system of parameters of interest. Then there exists a design ξ with no more than

$$\frac{1}{2}s(s+1) + s(\text{rank } M - s) \tag{17}$$

support points that achieves the given moment matrix M. This theoretical bound on the size of the optimal support grows quadratically in s. For many practical settings the bound can be further improved.

8.5 Efficient Design Apportionment

In general the weights w_i of an optimal design need not be rational. Hence if a design is implemented for sample size n, the fair shares nw_i need not be integers. Numerical rounding of the fair shares easily violates the goal of allocating a total of n observations, for the reason that the quantities nw_i when numerically rounded no longer sum to n.

Instead an *efficient design apportionment* is described in Chapter 12 of Pukelsheim (1993). The fair shares nw_i are replaced by pseudoquotas νw_i, for some arbitrary positive multiplier $\nu > 0$. The pseudoquotas νw_i are rounded up to the next integer, $\lceil \nu w_i \rceil$. Finally the multiplier ν is adjusted so that the replication frequencies $\lceil \nu w_i \rceil$ sum up to the prescribed sample size n,

$$\sum_{i=1}^{\ell} \lceil \nu w_i \rceil = n. \tag{18}$$

For more results on the properties and practical implementation of this efficient rounding procedure see Chapter 12 in Pukelsheim (1993).

9 Design Algorithms

9.1 Computer Packages

Many of the current statistical computer packages include an option to select or specify *practical* designs, for various models. They are mostly accompanied by the appropriate module to evaluate the experimental data and analyze the model. Nachtsheim (1987) reviews a wide variety or commercial programs and research software. Noticeable codes to calculate *optimal* designs, in the exact and/or the approximate theory, include `ACED` of Welch (1984), `Gosset` of Hardin and Sloane (1992), or `OptDes` of Wilhelm (1994).

The numerical challenge of the optimal design problem stems from the fact that they usually have a *flat optimum*. Small perturbations of the support points or the weights do not change the objective function in any great manner. This is reassuring from a practical point of view, in that the experimenter can adjust the design a little bit without giving away too much in terms of most optimality criteria. However, it renders standard numerical procedures inefficient as they proceed to the optimum very slowly only.

9.2 Gradient Algorithms

The majority of algorithms are based on the techniques from differentiable optimization. The general idea is to use directional derivatives to find a direction of improvement, and then employ a line search to determine an optimal step length. For the design problem special tasks are to delete "superfluous" support points, and add new ones.

Specific algorithms are given in Fedorov (1972), Silvey, Titterington and Torsney (1978), or Wu and Wynn (1978). An overview over existing methods and a unifying approach to them is presented in Gaffke and Mathar (1992). See also the discussion paper by Cook and Fedorov (1994).

Use of directional derivatives runs into difficulties for *nondifferentiable criteria* (such as the smallest-eigenvalue criterion $\lambda_{min}(C)$), or when convergence occurs against a moment matrix that is singular as may easily be the case when interest concentrates on parameter subsets $K'\beta$.

9.3 Subgradient Algorithms

Recent advances in *nondifferentiable, nonconvex optimization* may be used to overcome some of the specific difficulties of optimal design algorithms. Wilhelm (1994) uses a ***bundle trust methods*** to develop the `OptDes` algorithm. The idea is to carry along a "bundle" of subgradients from previous iteration points, and to construct a polyhedral approximation to the objective function in a local "trust" region.

Besides giving up the overly restrictive assumption of differentiability, bundle trust methods also relax the requirement that the objective function must be concave (or convex). Indeed, as a function of the support points when weights are fixed, the objective function $\psi = \phi \circ C_K$ of the design problem generally fails to be concave or convex.

The solution is to generalize the subgradient concept for convex functions to also cover sufficiently smooth nonconvex functions, the required property being Lipschitz continuity. The task then is one of calculating, in each iteration step, one generalized subgradient. Generalized subgradients for the design problem are given in Wilhelm (1995).

Acknowledgment

We are grateful for partial support from the Alexander von Humboldt-Stiftung through a Max-Planck-Award for cooperative research.

References and Bibliography

Abt M (1992). Some exact optimal designs for linear covariance functions in one dimension. Communications in Statistics, Theory and Methods **21** 2059–2069.

Abt M, Mayer R, Pukelsheim F (1995). Improving manufacturing quality through planned experiments: pressure governor case study. Surveys on Mathematics for Industry **5** 35–47.

Abt M, Pukelsheim F (1995). Improving manufacturing quality through planned experiments: statistical methodology. Surveys on Mathematics for Industry **5** 27–33.

Atkinson AC, Donev AN (1992). Optimum Experimental Designs. Oxford: Clarendon Press.

Bailey RA (1983). Randomization. pp. 9–26 in: Recent Trends in Statistics. Proceeding of the Dortmund 1982 Anglo–German Statistical Meeting (Heiler S, ed.). Göttingen: Vandenhoeck & Ruprecht.

Bailey RA (1985). Restricted randomization versus blocking. International Statistical Review **53** 171–182.

Bailey RA (1990). Cyclic designs and factorial designs. pp. 51–74 in: Probability, Statistics and Design of Experiments. Proceedings of the Delhi 1988 Symposium in Honour of RC Bose (Bahadur RR, ed.). New Delhi: Wiley Eastern.

Bandemer H (ed.) (1977). Theorie und Anwendung der optimalen Versuchsplanung I. Handbuch zur Theorie. Berlin: Akademie-Verlag.

Bandemer H, Näther W (1980). Theorie und Anwendung der optimalen Versuchsplanung II. Handbuch zur Anwendung. Berlin: Akademie-Verlag.

Banks D (1993). Is industrial statistics out of control? With discussion. Statistical Science **88** 356–409.

Bates DM, Watts DG (1988). Nonlinear Regression Analysis and Its Applications. New York: Wiley.

Berenblut II, Webb GI (1974). Experimental design in the presence of autocorrelated errors. Biometrika **61** 427–437.

Bischoff W (1993). On D-optimal designs for linear models under correlated observations with an application to a linear model with multiple response. Journal of Statistical Planning and Inference **37** 69–80.

Box GEP (1985). The Collected Works of George E.P. Box (Tiao GC, Granger CWJ, Guttman I, Margolin BH, Snee RD, Stigler SM, eds.). Belmont, CA: Wadsworth.

Box GEP, Cox DR (1964). An analysis of transformations. With discussion. Journal of the Royal Statistical Society, Series B **26** 211–252.

Box GEP, Draper NR (1969). Evolutionary Operation. New York: Wiley.

Box GEP, Draper NR (1987). Empirical Model-Building and Response Surfaces. New York: Wiley.

Box GEP, Hunter JS (1957). Multifactor experimental designs for exploring response surfaces. Annals of Mathematical Statistics **28** 195–241.

Box GEP, Hunter WG, Hunter JS (1978). Statistics for Experimenters. New York: Wiley.

Box GEP, Wilson KB (1951). On the experimental attainment of optimum conditions. With discussion. Journal of the Royal Statistical Society, Series B **13** 1–45. Reprinted in: Breakthroughs in Statistics **II** (Kotz S, Johnson NL, eds.). New York: Wiley.

Box GEP, Youle PV (1955). The exploration and exploitation of response surfaces: An example of the link between the fitted surface and the basic mechanism of the system. Biometrics **11** 287–323.

Cheng CS (1978). Optimal designs for the elimination of multi-way heterogeneity. Annals of Statistics **6** 1262–1272.

Cheng CS (1983). Construction of optimal balanced incomplete block designs for correlated observations. Annals of Statistics **11** 240–246.

Clatworthy WH (1956). Contributions on Partially Balanced Incomplete Block Designs with Two Associate Classes. National Bureau of Standards Applied Mathematics Series **47**. Superintendent of Documents, US Government Printing Office.

Clatworthy WH (1973). Tables of Two-Associate-Class Partially Balanced Designs. National Bureau of Standards Applied Mathematics Series **63**. Superintendent of Documents, US Government Printing Office.

Constantine GM (1987). Combinatorial Theory and Statistical Design. New York: Wiley.

Cook DV, Fedorov VV (1994). Constrained optimization of experimental design. With discussion. Statistics **26** 129–178.

Cornell JA (1990). Experiments with Mixtures, Second Edition. New York: Wiley.

Cressie N (1991). Statistics for Spatial Data. New York: Wiley.

Daniel C (1976). Applications of Statistics to Industrial Experimentation. New York: Wiley.

Deming WE (1986). Out of the Crisis. Cambridge, MA: MIT Center for Advanced Engineering Study.

Dette H (1992). Optimal designs for a class of polynomials of odd or even degree. Annals of Statistics **20** 238–259.

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

Draper NR, Pukelsheim F (1990). Another look at rotatability. Technometrics **32** 195–202.

Draper NR, Pukelsheim F (1994). On third order rotatability. Metrika **41** 137–161.

Draper NR, Smith H (1981). Applied Regression Analysis. New York: Wiley.

Dyke GV (1974). Comparative Experiments with Field Crops. London: Butterworths.

Efron B (1971). Forcing a sequential experiment to be balanced. Biometrika **58** 403–417.

Elfving G (1952). Optimum allocation in linear regression theory. Annals of Mathematical Statistics **23** 255–262.

Federer WT, Balaam LN (1972). Bibliography on Experiment and Treatment Design Pre-1968. Edinburgh: Oliver & Boyd.

Fedorov VV (1972). Theory of Optimal Experiments. New York: Academic Press.

Fisher RA (1925). Statistical Methods for Research Workers, Edinburgh: Oliver & Boyd. (10th edition, 1948)

Fisher RA (1926). The arrangement of field experiments. Journal of the Ministry of Agriculture **33** 503–513.

Fisher RA (1935). Design of Experiments, Edinburgh: Oliver & Boyd. (8th edition, 1966)

Fisher RA (1966). Design of Experiments, Eighth Edition. Edinburgh: Oliver & Boyd.

Fisher RA, Yates F (1938). Statistical Tables for Biological, Agricultural and Medical Research. (5th edition, 1957)

Fisher RA, Yates F (1963). Statistical Tables for Biological, Agricultural and Medical Research, Sixth Edition. New York: Hafner.

Gaffke N (1987). On D-optimality of exact linear regression designs with minimum support. Journal of Statistical Planning and Inference **15** 189–204.

Gaffke N, Mathar R (1992). On a class of algorithms from experimental design theory. Optimization **24** 91–126.

Ghosh S (1990). Statistical Design and Analysis of Industrial Experiments. New York: Dekker.

Gill PS, Shukla GK (1985). Efficiency of nearest neighbour balanced block designs for correlated observations. Biometrika **72** 539–544.

Grize YL (1995). Review of robust process design approaches. Journal of Chemometrics **9** 239–262.

Grondona MO, Cressie N (1991). Using spatial considerations in the analysis of experiments. Technometrics **33** 381–392.

Grove D, Davis TP (1992). Engineering, Quality, and Experimental Design. Harlow, Essex, UK: Longman.

Hardin RH, Sloane NJA (1992). A new approach to the construction of optimal designs. Journal of Statistical Planning and Inference **37** 339–369.

Hedayat AS, Jacroux M, Majumdar D (1988). Optimal designs for comparing test treatments with a control. With discussion. Statistical Science **3** 462–491.

Heiligers B (1991). A note on the connectedness of block designs. Metrika **38** 377–381.

John JA (1995). Cyclic Designs, Second Edition. London: Chapman and Hall.

John PWM (1990). Statistical Methods in Engineering and Quality Assurance. New York: Wiley.

Kempthorne O (1952). The Design and Analysis of Experiments. New York: Wiley.

Khuri AI (1988). A measure of rotatability for response-surface designs. Technometrics **30** 95–104.

Khuri AI, Cornell JA (1987). Response Surfaces. New York: Marcel Dekker.

Kiefer JC (1985). Collected Papers **III** (Brown LD, Olkin I, Sacks J, Wynn HP, eds.). New York: Springer.

Kiefer JC, Wynn HP (1984). Optimum and minimax exact treatment designs for one-dimensional autoregressive error processes. Annals of Statistics **12** 431–450.

Krafft O (1978). Lineare statistische Modelle und optimale Versuchspläne. Göttingen: Vandenhoeck & Ruprecht.

Kunert J (1985). Optimal repeated measurements designs for correlated observations and analysis by weighted least squares. Biometrika **72** 375–389.

Kunert J (1987). Neighbour balanced block designs for correlated errors. Biometrika **74** 717–724.

Kunert J, Martin RJ (1987). On the optimality of finite Williams II(a) designs. Annals of Statistics **15** 1604–1628.

Kurotschka V (1971). Optimale Versuchspläne bei zweifach klassifizierten Beobachtungsmodellen. Metrika **17** 215–232.

Logothetis N, Wynn HP (1989). Quality Through Design. Oxford: Clarendon Press.

Martin RJ (1986). On the design of experiments under spatial correlation. Biometrika **73** 247–277.

Mead R, Curnow RN, Hasted AM (1993). Statistical Methods in Agriculture and Experimental Biology, Second Edition. London: Routledge, Chapman and Hall.

Morgan JP, Chakravarty IM (1988). Block designs for first and second order neighbor correlations. Annals of Statistics **16** 1206–1224.

Myers RH, Khuri AI, Carter WH (1989). Response surface methodology: 1966–1988. Technometrics **31** 137–157.

Nachtsheim CJ (1987). Tools for computer-aided design of experiments. Journal of Quality Technology **19** 132–160.

Nair VJ (1992). Taguchi's parameter design: A panel discussion. Technometrics **34** 127–161.

Pázman A (1986). Foundations of Optimum Experimental Design. Dordrecht: Reidel.

Peirce CS (1957). Essays in the Philosophies of Science (Tomas V, ed.). Indianapolis: Bobbs-Merrill.

Pilz J (1991). Estimation and Experimental Design in Linear Regression Models. New York: Wiley.

Plackett RL (1946). Some generalizations in the multifactorial design. Biometrika **33** 328–332.

Plackett RL, Burman JP (1946). The design of optimum multifactorial experiments. Biometrika **33** 305–325.

Pukelsheim F (1993). Optimal Design of Experiments. New York: Wiley.

Raghavarao D (1971). Constructions and Combinatorial Problems in Design of Experiments. New York: Wiley.

Raktoe BL, Hedayat A, Federer WT (1981). Factorial Designs. New York: Wiley.

Rasch D, Guiard V, Nürnberg G (1992). Statistische Versuchsplanung. Stuttgart: Gustav Fischer.

Scheffé H (1958). Experiments with mixtures. Journal of the Royal Statistical Society, Series B **20** 344–360.

Scheffé H (1963). The simplex-centroid design for experiments with mixtures. Journal of the Royal Statistical Society, Series B **25** 235–263.

Seber GAF, Wild CJ (1989). Nonlinear Regression. New York: Wiley.

Shah KS, Sinha BK (1989). Theory of Optimal Designs. Lecture Notes in Statistics **54**. New York: Springer.

Shoemaker AC, Tsui KL, Wu CFJ (1991). Economical experimentation methods for robust design. Technometrics **33** 415–427.

Silvey SD (1980). Optimal Design. London: Chapman and Hall.

Silvey SD, Titterington MD, Torsney B (1978). An algorithm for optimal design on a finite design space. Communications in Statistics, Theory and Methods **A7** 1379–1389.

Stigler SM (1986). The History of Statistics. Cambridge, MA: Belknap Press.

Street AP, Street DJ (1987). Combinatorics of Experimental Design. Oxford: Clarendon Press.

Studden WJ (1982). Some robust-type *D*-optimal designs in polynomial regression. Journal of the American Statistical Association **77** 916–921.

Taguchi G (1987). System of Experimental Design. New York: Kraus.

Titterington DM (1980). Aspects of optimal design in dynamic systems. Technometrics **22** 287–299.

Weihs C, Berres M, Grize YL (1995). Statistical design of experiments in industrial practice. Surveys on Mathematics for Industry **5** 49–73.

Welch WJ (1984). Algorithms for the construction of experimental designs. User's Guide Version 1.5.2. Vancouver: Faculty of Commerce and Business Administration, University of British Columbia.

Wilhelm A (1994). Computing optimal designs by bundle trust methods. Report 314, Institut für Mathematik, Universität Augsburg.

Wilhelm A (1995). Subdifferentiability and Lipschitz continuity in optimal design problems. Metrika, forthcoming.

Williams RM (1952). Experimental design for serially correlated observations. Biometrika **39** 151–167.

Wu CFJ, Wynn HP (1978). The convergence of general step-length algorithms for regular optimum design criteria. Annals of Statistics **6** 1273–1285.

Norman R. Draper
Department of Statistics
University of Wisconsin–Madison
MADISON, WI 53706
USA

Friedrich Pukelsheim
Institut für Mathematik
Universität Augsburg
D-86135 AUGSBURG
GERMANY