CHAPTER 10

Response Surface Designs

Response surface methodology (RSM) is, as the name suggests, a collection of tools for fitting a surface to a set of data, and determining optimum factor levels is also part of the methodology. Of course the shape of the surface is determined by the model that is fit and the response values (i.e., the data), and hence the term "response surface." A distinction should of course be made between the fitted surface and the actual surface, which will almost certainly not coincide, just as the estimate of a parameter will rarely be equal to the parameter value. A fitted surface is, in essence, an estimate of the true surface because if an experiment were repeated, the fitted surface would almost certainly differ at least slightly from the first fitted surface.

Typically a full second-order model is fit in trying to determine the optimum combination of factor levels; this is the model with both linear and quadratic terms in each factor and all two-factor interactions. Before that is done, however, it is not uncommon for a first-order model to be fit and then the method of steepest ascent (or descent, see Section 10.4 for both) used to try to zero in on the optimum operating region, and then use a design to accommodate a second-order model for the purpose of characterizing the region.

Thus, the standard approach is to use a three-stage procedure. The procedure can fail, however, if a screening design (i.e., a resolution III design) is used in the first stage in the presence of significant interactions. If interactions are suspected, a resolution IV design should be used in the first stage so that the wrong subset of factors is not identified in that stage. Another possible impediment, as discussed by Steinberg and Bursztyn (2001) and Xu, Cheng, and Wu (2004), is that it is not always possible, or at least practical, to conduct experiments sequentially, although the need to do so has been stressed by G. E. P. Box, in particular, for decades. Thus, experiments beyond the first one may not be possible.

A key component of RSM is response surface designs. Snee (1985) gave a list of desirable properties for response surface designs, the list having first appeared in Box and Draper (1975) and later given in Box and Draper (1987). Most of those

Modern Experimental Design By Thomas P. Ryan

properties are also listed online in, for example, the *NIST/SEMATECH e-Handbook of Statistical Methods* (Croarkin and Tobias, 2002). See http://www.itl.nist.gov/div898/handbook/pri/section3/pri3363.htm and scroll to the bottom of the page.

The 14 properties listed (the last source lists 11) were mostly general, however, and could essentially apply to any experimental design. One of those objectives, "sequential construction of higher order designs from simpler designs," will be indirectly debated in Section 10.1, in which the choice between using one design and more than one design will be discussed.

Similarly, Anderson and Whitcomb (2004) gave their list, which included general properties that any type of good experimental design should possess, in addition to some specific properties such as "behave well when errors occur in the settings of factors (the x's)," and "be insensitive to wild observations."

In this chapter we examine both traditional and nontraditional response surface designs and consider the reasons that have been given during the past few years for moving away from traditional designs and considering other types of designs, especially a uniform design (UD), which is discussed and illustrated in Section 10.3. The designs that have traditionally been used, such as central composite designs (CCDs) (discussed in detail in Section 10.5), permit the fitting of second-order models, so we should view response surface designs as an extension of the designs that were presented in previous chapters. We may also regard a CCD as an enlargement of a factorial design since part of the design is a full or fractional factorial.

We also consider a proposal by Cheng and Wu (2001) and others to move away from the traditional manner in which response surface designs have been used. Typically, a two-level screening design has been used to first identify important factors, then those factors are investigated in a response surface design if the experimenter is interested in fitting a second-order model and perhaps also interested in determining the point/region of optimum operating conditions. We discuss this in Section 10.1.

Published applications of RSM have been primarily in the chemical and food industries, although as stated by Myers (1999), interest in RSM has spread to the biological, biomedical, and biopharmaceutical fields. A small sample of recent applications papers is as follows: Ghadge and Raheman (2006) used RSM for process optimization for biodiesel production; Moberg, Markides, and Bylund (2005) used RSM in a tandem mass spectrometry application; Huang, Lu, Yuan, Lü, and Bie (2006) used RSM in a microbiological application to determine the optimum levels of four protective agents; and Kim and Akoh (2005) used RSM in modeling lipase-catalyzed acidolysis in hexane. Other applications papers include Tuck, Lewis, and Cottrell (1993), who applied RSM in the milling industry.

Although there has been a large number of RSM applications, there are not very many researchers who work in the area of RSM, and the few who do so have lamented their small number (e.g., Khuri, 1999). This is paralleled by the relative paucity of books on RSM, the first of which was Myers (1971), with the "second edition" and now the third edition of that book becoming Myers and Montgomery (2002). The original work on RSM, however, was given in the classic paper by Box and Wilson (1951), and Box and Draper (1987) is another well-known book on RSM, as is Khuri and Cornell (1996). More recently, Khuri (2003) surveyed contemporary modeling approaches and design issues in RSM (see also Khuri, 2005).

10.1 RESPONSE SURFACE EXPERIMENTATION: ONE DESIGN OR MORE THAN ONE?

Assume that there are 20 factors that may be related to a response variable and there is a desire to determine the best combination of levels of the factors that are related to the response variable so as to optimize the (value of) response variable. The first step would obviously be to determine which of the 20 factors are really important. Typically this would be performed with some type of screening design that has a small number of runs relative to the number of factors, and for which main effects would be confounded with two-factor interactions. Of course these screening designs work only if main effects dwarf interactions, as is often the case.

If certain two-factor interactions are significant, we may falsely identify the main effects with which they are confounded as being significant. The consequence of this is that too many factors may be examined in a subsequent response surface design, thus resulting in an inefficient design. Clearly it would be much better if factor screening and response surface fitting and optimization could be performed with a single experiment, and have a high probability of identifying the real effects with the factor screening. Such a strategy has been proposed by Cheng and Wu (2001) and Bursztyn and Steinberg (2001), among others.

The use of such a strategy entails projecting a factor space onto a smaller factor space, namely, the space of factors that seem to be important. This means that the design used for factor screening must have at least three levels since the design will be projected onto a smaller factor space *of the same design*.

Box and Wilson (1951) decried the use of 3^{k-p} designs as second-order designs, pointing out that there is no useful 3^{3-p} design, so a 3^3 design would have to be used. They also pointed out that a 3^{5-1} design (81 runs) is needed for fitting a second-order model with 21 parameters. That is, there are degrees of freedom available for estimating higher-order terms that are not being used, and the 59 degrees of freedom that are available for estimating the error variance are far more than would normally be needed. As pointed out by Cheng and Wu (2001), the Box and Wilson claim that a 3^4 design is necessary for fitting a second-order model with four factors and 15 parameters is incorrect as this could be accomplished with any 3_{IV}^{4-1} design. Cheng and Wu (2001) pointed out that the Box and Wilson argument for not using these designs holds only for $k = 3$ and $k = 5$.

Furthermore, as pointed out by Cheng and Wu (2001), the inefficient run size argument for 3^{k-p} designs clearly doesn't hold when such a design is "doing double duty" by serving as both a screening design and a design for fitting a second-order model. For example, assume that a 3^{10-7} design is used and it is found that there are five important factors. The projected design must of course have the same number of design points, so it would be a 3^{5-2}. Cheng and Wu (2001) gave an example of a 3^{6-3} design projected onto a 3^{5-2} design.

We need to keep in mind, however, that designs such as a 3^{10-7} and a 3^{6-3} are resolution III designs, so if interactions exist, they could undermine this approach. Furthermore, a 3^{5-2} design has the same resolution, so interactions could not be estimated; that is, only pure quadratic terms could be fit, not mixed quadratic terms.

There is no 3_{V}^{k-p} design with 27 runs and the only 3_{V}^{k-p} design with 81 runs (more than most experimenters would probably want to use) is the 3_{V}^{5-1} design.

We should also keep in mind that the idea of using a single design with at least three levels will fail if the design does not cover the optimum region, and furthermore cover it in such a way that the fitted surface is essentially the same as the true surface. This would be a very bold assumption.

Cheng and Wu (2001) used the term *eligible projected design* to designate a second-order design. Since the number and the identity of the important factors are of course not known before the experiment is performed, it is thus desirable to use a design that has a relatively large number of eligible projected designs. For example, assume that there are 12 factors to be examined. There are $\binom{12}{4} = 495$ projected designs onto four factors, but not all of these will be eligible designs.

In general, the idea of letting a 3_{V}^{k-p} design do double duty seems to be not very practical because the design would have to have enough points (at least 81) for the projected design to be resolution V, since all second-order effects should be estimable when a second-order model is fit to characterize the response surface and to seek the optimum combination of factor levels.

However, let's make some comparisons. Assume that there are 10 factors to be screened and we know that five of the factors are important, as well as certain interactions among the 5 factors. We could use a 3_{IV}^{10-6} design and project that into a 3_{V}^{5-1} design, assuming that we have correctly identified the five important factors. We are then able to fit a full second-order model in these five factors with 81 runs, assuming of course that the 3_{V}^{5-1} design covers the optimum region. If a 2_{IV}^{10-k}design with the fewest number of design points were used, this would be a 2_{IV}^{10-5} design, which would project into a 2^5 design for any five of the factors. Converting this into a CCD would require 10 more runs for the axial points, and let's assume that eight centerpoints were used, for a total of 50 runs, 31 runs fewer than with the 3-level design approach. Thus, the latter is not competitive with the two-level, screening design approach for the first stage.

Of course both approaches would fall apart if the optimum region is not covered by the first design, but the 2-level design approach might still be superior if the method of steepest ascent/descent (Section 10.4) had to be employed in addition to the use of two designs, provided that the region in which the second-order design was to be used was reached quickly without much experimentation performed along the path of steepest ascent/descent. Of course the optimum region will generally be unknown and will not likely be covered by a two-level design unless the factor levels practically extend to the limits of the operability region. Since the number of experiments that are necessary along an assumed path of steepest ascent/descent will vary between experiments and will be unknown before any specific experimentation is undertaken, it is almost impossible to compare the suggested one-design approach with the standard sequential approach.

If the response surface is complex with multiple peaks and/or valleys, it can be difficult to determine optimum operating conditions with either approach. For such surfaces, a method of continuous experimentation such as simplex EVOP (Evolutionary Operation) might be employed in searching for the optimum region; then use the standard EVOP approach (Box, 1957) or either of the design approaches described here to determine the optimum conditions after the optimum region seems to have been pinpointed. (Simplex EVOP is discussed in Spendley, Hext, and Himsworth (1962), Lowe (1974), and Hahn (1976).)

10.2 WHICH DESIGNS?

The question of which designs to use in RSM work was also raised in the literature by Tang, Chan, and Lin (2004), who reported that the theoretical results of Fang and Mukerjee (2000) showed that the success of factorial designs in the exploration of response surfaces is due to the fact that the designs uniformly cover the entire design space, rather than being due to the combinatorial or orthogonal properties of the design. Tang et al. (2004) described the successful application of a UD, which they believe has considerable potential in certain situations relative to traditional response surface designs. Uniform designs are covered in detail in Section 13.9.1.

In the same general vein of alternative designs, Hardin and Sloane (1991; http://www.research.att.com/∼njas/doc/doeh.pdf) discussed the computer generation of response surface designs for spherical regions. They stated, "The best designs found have repeated runs at the center and points well spread out over the surface of the sphere." Similarly, Cox and Reid (2000, p. 181) stated that when the relationship between the response variable and factors is highly nonlinear "a space-filling design is then useful for exploring the nature of the response surface." Mee (2004) recommended the use of D-optimal and I-optimal three-level designs for spherical design regions involving three or more factors.

Furthermore, McDaniel and Ankenman (2000) found that a version of the traditional RSM worked best when the objective is to make small factor changes in search of a maximum or minimum. Of course this is also what is done when EVOP is used when seeking optimum conditions (Box and Draper, 1969).

Motivation for using something other than traditional RSM designs also comes from Giesbrecht and Gumpertz (2004, p. 413), who stated in their chapter on response surface designs that "in practice, it is more common to find severe constraints that limit the investigation to irregular regions."

There is thus evidence that we should move away from traditional RSM designs and use either UDs or designs quite similar to UDs over the acceptability or operability region. Nevertheless, we will first review traditional response surface designs but will later return to a discussion of UDs.

10.3 CLASSICAL RESPONSE SURFACE DESIGNS VERSUS ALTERNATIVES

With current computing technology, we can easily see how data look in three dimensions. For example, given in Figure 10.1 is the surface plot for two factors using the data from Vázquez and Martin (1998) that was discussed in Section 6.1.

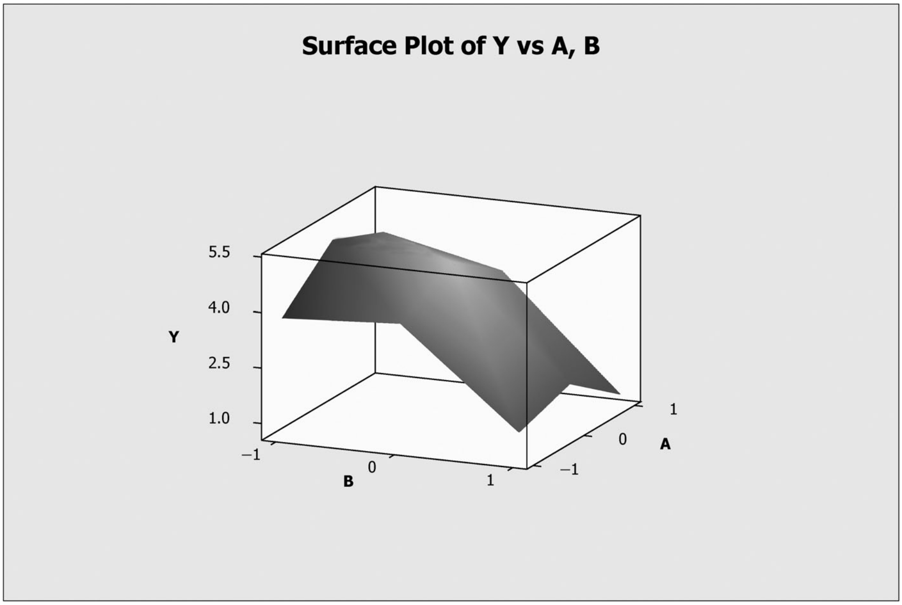

Figure 10.1 Surface plot for the data from Vázquez and Martin (1998).

This surface plot was constructed using Release 14 of MINITAB. The plot was constructed from the data for a 3^2 design, so the nine points were connected to form a surface. The need for a quadratic term in each factor is obvious from the "fold" in the surface on each factor and the interaction is apparent from the bend or kink in the surface. (The surface would have been flat if only main effects were important; the plot suggests that the linear and quadratic terms plus the interaction term explain just about all of the variability in Y, and in fact the model with those terms had an R^2 of .9984.)

We should keep in mind that just as models are unknown, so are response surfaces. The true response surface is almost certainly not as depicted in Figure 10.1, and is very likely irregular with peaks and valleys. We should also keep in mind that the surface in Figure 10.1 is obtained just by connecting points; this is different from a fitted surface that is obtained by fitting a model to data, but such a surface is also random in the sense that the coefficients are realizations of random variables. A fitted surface is discussed and illustrated in Section 10.13.

Obviously the only way we could obtain a very close approximation to the nature of the response surface between the extreme points in the design is to have an extremely large number of points over a fine grid of that region. Unless experimentation is extremely inexpensive, as it is in many computer experiments, the cost of obtaining the necessary number of design points to closely approximate the response surface will generally be prohibitive. The goal then is to approximate the response surface as closely as is needed relative to the objectives of the study, and to do so in an economical manner—not an easy task.

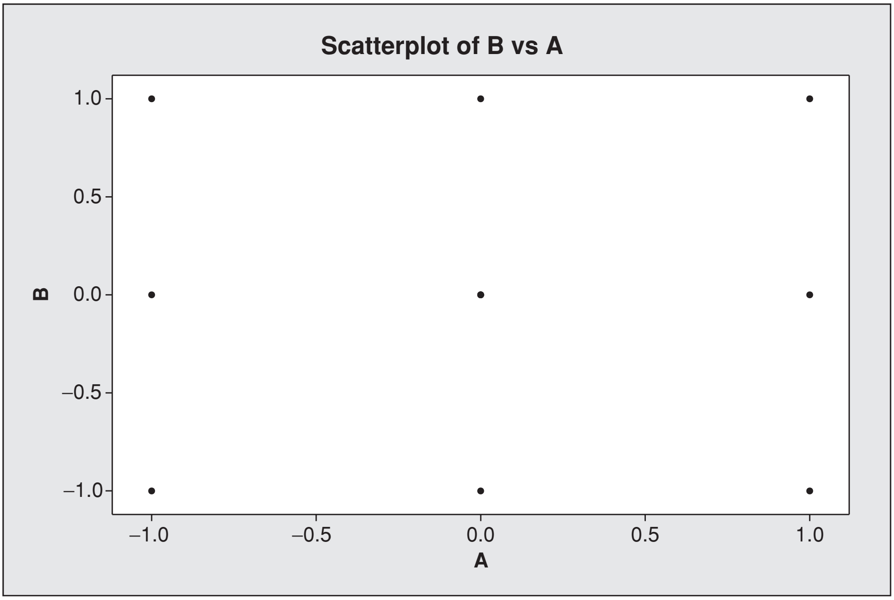

Figure 10.2 3^2 design.

Assume that there are two factors of interest and consider the design layouts depicted in Figures 10.2–10.4.

Figure 10.2 is the design layout for the 3^2 design used by Vázquez and Martin (1998). Although the region appears to be uniformly covered, one weakness is that there is only one point in the interior of the design. This would be an especially poor design if the values of the factors in the original (uncoded) units had been chosen to be extreme values rather than values that were well within the extremes, as with such extreme values the centerpoint might then be the only design point that would be either a point used in practice or in proximity to points used in practice. (Of course here the term "centerpoint" is being used to refer to the design point that is in the center of the design, which of course is different from centerpoints that are added to, say, a 2^k design.)

Figure 10.3 is the layout for a design that is discussed in more detail in Section 10.5: a CCD. This design also has nine unique design points and does not differ greatly from the 3^2 design. As with the latter, the CCD has only one point in the interior of the design. The design was generated by MINITAB using the default values for the axial (star) points, which are the points in proximity to the 2^2 design points. Selection of axial point values is discussed in Section 10.5.

Figure 10.4 is a UD constructed using JMP. (*Note*: This design and the Latin hypercube and sphere-packing designs that are illustrated in Figures 10.6 and 10.7 are random designs in the sense that if JMP, for example, is used to construct the designs, successively generated designs of the same type and with the same design parameters will be different.)

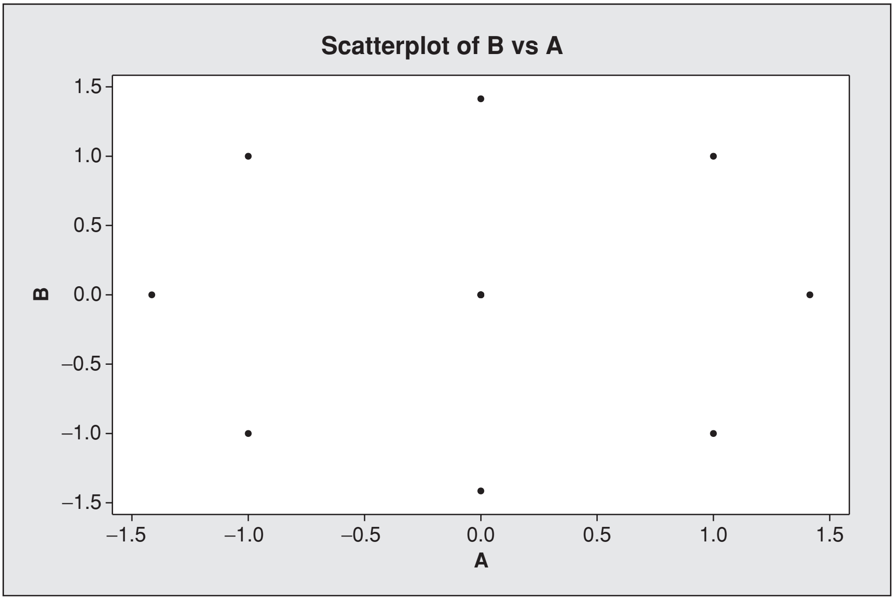

Figure 10.3 Central composite design for two factors.

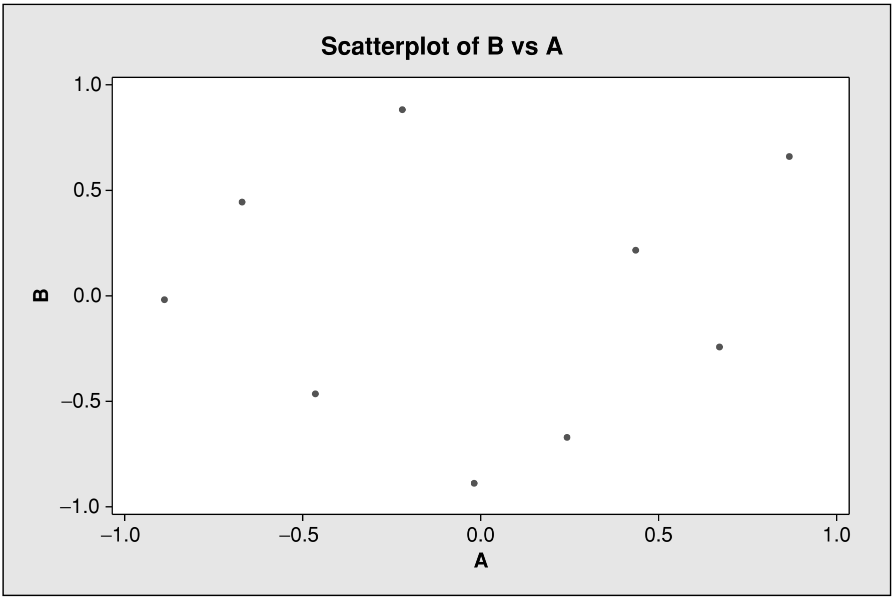

Figure 10.4 Uniform design for two factors.

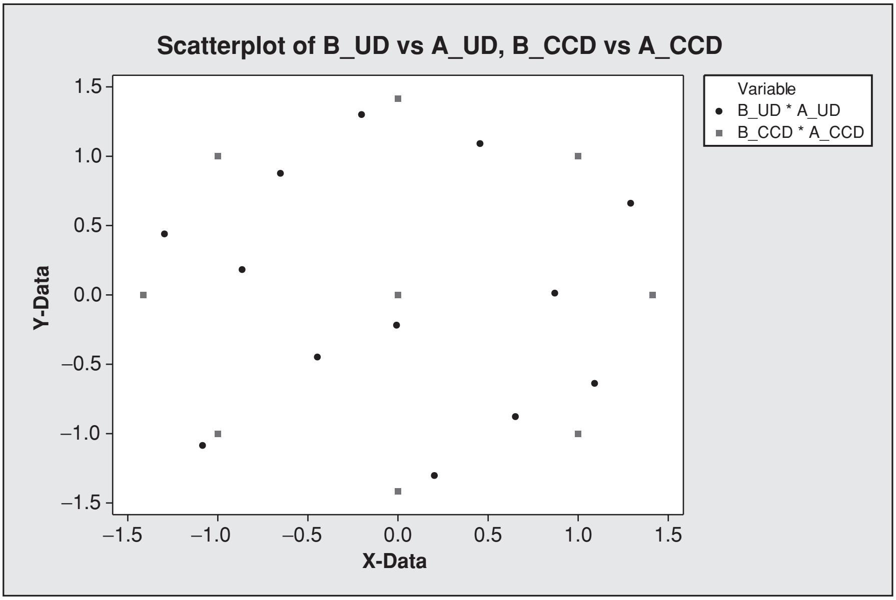

Figure 10.5 Comparison of CCD and UD.

Unlike the first two designs, this design has an irregular configuration, although there is better coverage of the design interior than with the first two designs. Remember that the surface in Figure 10.1 was constructed by connecting the nine design points. Therefore, if serious nonlinearities exist in the interior of the design, a UD would provide a better opportunity for detecting them, although we would probably prefer a few more interior points. That is, if the interior points are not well-fit by the full second-order model, we would like to be able to detect that.

The CCD in Figure 10.3 actually has 13 experimental runs since the centerpoint is replicated five times. Thus, a fairer comparison would be to compare a UD with 13 points to that CCD.

The scaling of the two designs should be the same so as to facilitate a fair comparison, and it would obviously be helpful if the designs could be shown on the same graph. Since the range was $-\sqrt{2}$ to $\sqrt{2}$ for each factor for the CCD (this was not a deliberate choice; this is simply how the design is constructed), this range was also used for the UD.

The comparison is shown in Figure 10.5 and there is obviously no comparison in terms of coverage as a UD provides essentially the same degree of coverage of the periphery of the design space as does the CCD, but has much better coverage of the interior region than the CCD. This of course is due primarily to the fact that there are 13 distinct design points with a UD but only 9 with the CCD, as stated previously.

Some potential UD users might be turned off by the unequal spacing of the design points and by the fact that there is not a unique UD for a given number of factors and

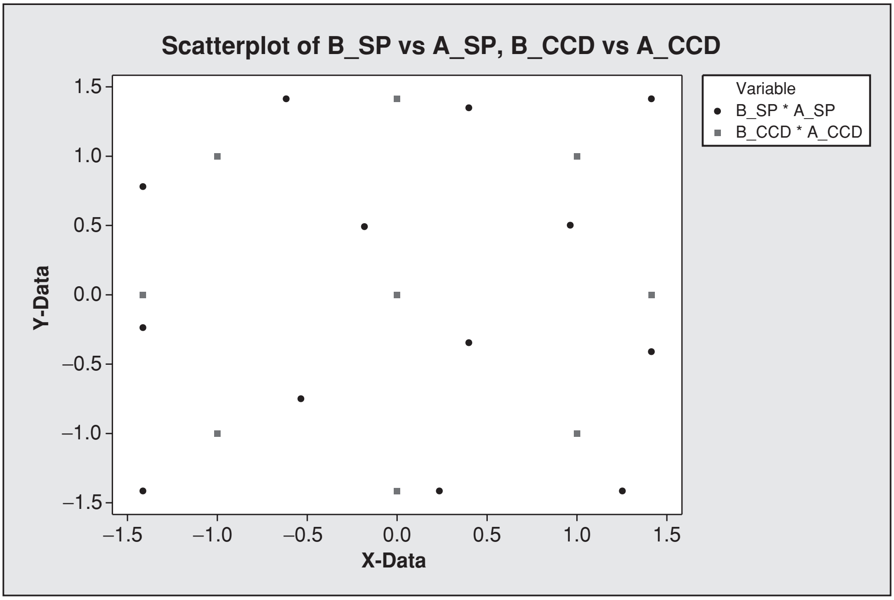

Figure 10.6 Comparison of sphere-packing design and CCD.

design points, as stated previously (see also Section 13.9.1). There are other types of "space-filling" designs, as they are called (Section 13.9), and Figure 10.6 shows the comparison of the sphere-packing design with the CCD.

Finally, Figure 10.7 shows the comparison of a Latin hypercube design with the CCD.

It should be apparent from each of these last three graphs that a space-filling design provides better coverage of the interior part of the design space than does a CCD. There is a price that is paid for this, however, although it is a small one. Specifically, space-filling designs are not orthogonal designs for estimating main effects, although the departure from orthogonality is not great. For example, the correlation between the columns for A and B for the particular Latin hypercube design that was generated is $-.066$, .064 for the sphere-packing design, and $-.041$ for the UD. The superior ability to detect nonlinearities more than offsets this slight departure from orthogonality.

In the general spirit of space-filling designs, Design-Expert can generate "distance-based" response surface designs, which spread the points evenly over the experimental region. These are not random designs as only certain interior points are generated randomly, with the other points having fixed coordinate values of 0, 1, and -1.

10.3.1 Effect Estimates?

Effect estimates were discussed for two-level designs, starting with Section 4.1, and for three-level designs in Section 6.1.1. For the uniform and Latin hypercube designs discussed in Section 10.3, the design points all represent different factor levels, and

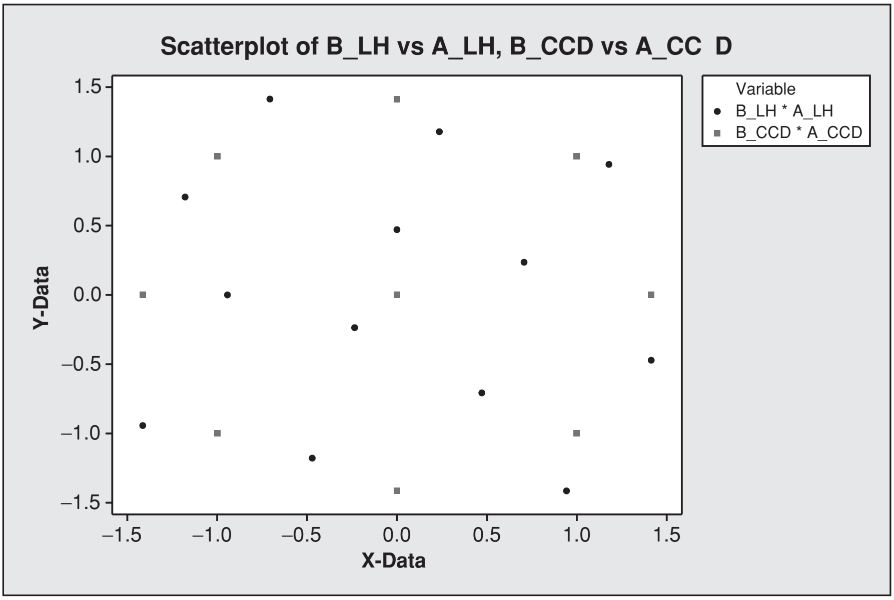

Figure 10.7 Comparison of Latin hypercube with CCD.

indeed the designs are random designs with points that would differ if the designs were successively generated, as was explained in Section 10.3. Therefore, it wouldn't make any sense to think about trying to define effect estimates for such designs, any more than it would make sense to try to do so when a regression model is constructed using a single random regressor (although we do compute regression coefficients, which is different).

We return to UDs and look at some additional statistical properties of them as well as some applications later in the chapter, after first looking at traditional response surface designs. (*Note*: Whereas the designs depicted in Figures 10.5–10.7 were generated using JMP, it isn't possible to use JMP to generate space-filling designs for restricted regions of operability when the region does not form a rectangle. It is, however, possible to do this with SAS Software by using PROC OPTEX, as is illustrated by Giesbrecht and Gumpertz, 2004, p. 415.)

10.4 METHOD OF STEEPEST ASCENT (DESCENT)

This method, due to Box and Wilson (1951), is used to try to identify the region, as narrowly defined as possible, that contains the optimum operating conditions. Methodology for producing a "path of steepest ascent" (or descent) has been available for several decades, but with modern computing capabilities we don't need to think about literally constructing the path.

We will use some simple examples to illustrate the basic idea.

Example 10.1

Assume that there are two factors of interest and for the sake of illustration we will presume that the relationship between them and the response variable is $Y = 2 + 0.4X_1 + 0.3X_2$, with the two factors constrained to lie within $(-1, 1)$, as in a two-level design. The response variable is to be maximized. What is the path of steepest ascent (i.e., path to the maximum), assuming that the current combination of factor levels is (0, 0)?

Looking at it intuitively, it should be apparent that the path will be in the quadrant for which both X_1 and X_2 are positive. What might not be obvious, however, is the relationship between X_1 and X_2 along the path. Since the coefficient of X_1 is larger than the coefficient of X_2, the change in X_2 should be expressed in terms of the change in X_1 and will also be a function of the relationship between the coefficients. This relationship can be easily determined if we think about equal increments in Y, starting with $Y = 2$ since that is the value of Y when $(X_1, X_2) = (0, 0)$. We might think about increments of 1.0 for Y, recognizing that Y cannot exceed 2.7 because of the constraints on X_1 and X_2.

If we fix the increment of X_1, ΔX_1, at 0.1, we can then solve for ΔX_2 after determining the direction of steepest ascent, which is simply $\lambda(0.4, 0.3)$, obtained from the coefficients in the assumed equation, and also what we would obtain using calculus directly. That is, with $\Delta X_1 = 0.1$, $\Delta X_2 = (3/4)\Delta X_1 = 0.075$.

Another way of viewing this is to recognize that for a given value of Y such as $Y = 2$, the slope of the line when X_2 is written as a function of X_1 is -1.33. The slope of a line perpendicular to this line is the negative reciprocal of this slope, namely 0.75. A change in X_1 of 0.1 would thus result in a change in X_2 of 0.075.

The line of steepest ascent would, if the scales were the same, thus be perpendicular to the lines in Figure 10.8. Obviously the maximum occurs at (1, 1) and we don't need the method of steepest ascent to tell us that; this was just a simple illustrative example.

Box and Draper (1987, pp. 190–194) presented a method for determining whether the path of steepest ascent has been determined precisely enough (see also Myers and Montgomery, 1995, pp. 194–198). The method involved the construction of a confidence cone about the steepest ascent/descent direction and is described in detail in Section 5.5.1.2 of Croarkin and Tobias (2002). See http://www.itl.nist.gov/div898/handbook/pri/section5/pri5512.htm#Technical%20Appendix.

The cone is given by the inequality

$$\sum_{i=1}^{k} b_i^2 - \frac{\left(\sum_{i=1}^{k} b_i x_i\right)^2}{\sum_{i=1}^{k} x_i^2} \leq (k-1)\, s_b^2\, F_{\alpha, k-1, n-p}$$

with the b_i denoting the coefficients of the k (linear) model terms and s_b^2 denoting the common variance of the coefficients. Any point with coordinates $(x_1, x_2, \ldots, x_k)$ that

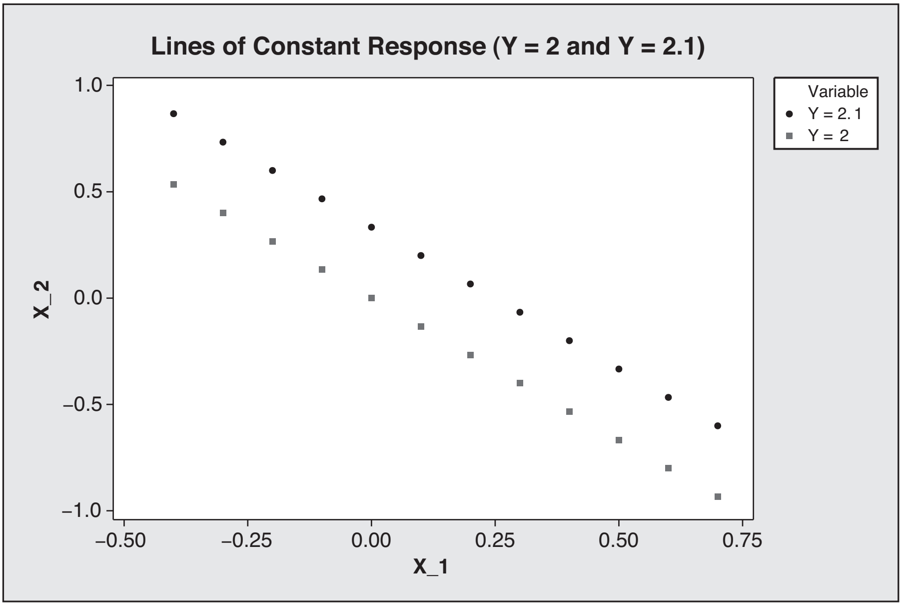

Figure 10.8 Lines of constant response for example.

satisfies this inequality lies within the confidence region provided that $\sum_{i=1}^{k} b_i x_i > 0$ for ascent and $\sum_{i=1}^{k} b_i x_i < 0$ for descent. Box and Draper (1987) gave a measure of how good a particular search direction is by the fraction of directions that are *excluded*. Of course we want that fraction to be as close to 1 as possible. The closer the fraction is to 1, the greater the confidence in the direction that is used. Of course we can't see this in high dimensions but Myers and Montgomery (1995, p. 196) applied the Box and Draper methodology to an example with $k = 2$ to compute the fraction. Sztendur and Diamond (2002) extended the Box and Draper approach to cover heterogeneous variances, nonlinear designs, and generalized linear models. They also used conditional effects in arriving at a path of steepest ascent. This is discussed in Section 10.9.

Whether the path is well determined will obviously depend upon the magnitude of the standard errors of the factor coefficients, which, unlike the example, will generally be unknown. The value of R^2 can also be used as a general indicator, as a low R^2 value for the fitted model would not provide any confidence that the direction has been well determined and low-to-moderate R^2 values will generally correspond to a relatively wide confidence cone.

One of the criticisms of the method of steepest ascent is that the direction is scale dependent. To illustrate, consider Example 10.1. If X_1 is in inches and is divided by 12 to convert to feet, the new coefficient would be $0.4 * 12 = 4.8$ and a different direction would then result since the direction is determined by the relationship between the model coefficients.

This has motivated recent work on scale-independent steepest ascent/descent methods, as given in Kleijnen, den Hertog, and Angün (2004). Del Castillo (1997) discussed stopping rules for the method of steepest ascent and suggested (see http://www.informs-sim.org/wsc04papers/059.pdf) that the experimentation cease when the value of the selected objective function (which might be R^2) is worse than the preceding value. Thus, experimentation might cease when R^2 exhibits a decline. This could indicate that the first-order model is no longer adequate for the region that has been entered, and so experimentation with a second-order model would be necessary. Of course we would also want to see the average response increase or decrease along the path, depending upon whether maximization or minimization is sought.

Thus, we are looking for two things: the need to switch to a second-order model and an indication that the right path (or at least a good path) is being followed. Nicolai, Dekker, Piersma, and van Oortsmarssen (2004) suggested using a simple t-test on responses from successive experiments to determine if the movement in the average response is in the desired direction.

Of course whether or not the method of steepest ascent is used as part of the standard three-stage approach depends on whether or not identification of the important factors and model fitting are to be performed with the same design, as discussed in Section 10.1. This in turn will depend on several factors, including the costs of running multiple experiments and the practicality of doing so, and whether or not the experimenters believe they know the region of optimal response. If offline experimentation is deemed too expensive for multiple experiments, the single-design approach might be used followed by an EVOP-type approach (if deemed possible to use), to check on the declared optimal combination of factor levels. That is, small changes would be made in factor levels, starting from the presumed optimal levels, and the changes in the response values noted and a decision made as to whether the changes are statistically significant. If so, further investigation would be warranted.

10.5 CENTRAL COMPOSITE DESIGNS

The most frequently used response surface design is the central composite design, abbreviated as CCD in the beginning of the chapter and abbreviated the same way in this section. There are various forms of the CCD that will be discussed in this section, which is why the word “designs” is in the section title. Readers interested in the motivation for the invention of the design(s) are referred to Box (1999–2000), who invented the design as given in Box and Wilson (1951). Bisgaard (1997) reported that the move away from traditional factorial and fractional factorial designs to a CCD was quite controversial at the time.

The basic design framework consists of three sections: the points of a 2^{k-p} design with, as usual, k denoting the number of factors and p usually but not necessarily equal to zero, axial (star) points, and centerpoints. All the sections are variable in the sense that although a CCD is generally presented with the factorial portion of the design composed of the points in the 2^k design, a fractional factorial could be used. MINITAB, for example, can be used to construct a "full" CCD (i.e., using the 2^k design in the factorial part) for k between 2 and 7, a half CCD for k between 5 and 8, and a quarter CCD for k equal to 8 or 9, with "half" and "quarter" of course being the half and quarter fractions, respectively. Specifically, for $k = 5$ the half CCD would have 32 points that would consist of the 16 factorial points, 10 axial points, and 6 centerpoints. This is a useful alternative to the full CCD for $k = 5$, since the factorial part of the CCD is resolution V. (The CCD has the same resolution as the factorial portion of it for estimating main effects and interactions since the axial points and centerpoints are not involved in the estimation of interactions.)

Whereas the number of factorial points is determined by the selection of a full or fractional factorial, the number of axial points is always twice the number of factors. There is no hard-and-fast rule for selecting the number of centerpoints, but studies have shown that the best response surface designs have replicated centerpoints. Of course centerpoints aid in the investigation of curvature in the region around the center. These points also aid in stabilizing $\mathrm{Var}(\widehat{Y})$ in and around the center, as illustrated by Myers and Montgomery (1995, pp. 310–311).

A CCD in two factors was shown in Figure 10.3 and compared with other designs in a graphical comparison. The design points are listed below.

Number	A	B
1	-1.000	-1.000
2	1.000	-1.000
3	-1.000	1.000
4	1.000	1.000
5	-1.414	0.000
6	1.414	0.000
7	0.000	-1.414
8	0.000	1.414
9	0.000	0.000
10	0.000	0.000
11	0.000	0.000
12	0.000	0.000
13	0.000	0.000

The star points are $(\alpha, 0)$, $(-\alpha, 0)$, $(0, \alpha)$, and $(0, -\alpha)$ for two factors, with the value of α to be selected. The value of α for the design listed above is $\sqrt{2}$, which is the default value in MINITAB. The reason for this choice will be explained shortly. Notice that if $\alpha = 1$ and the set of full factorial points is used, the resultant design is a 3^2 design when these points are combined with the centerpoints. Otherwise, the design has five levels for each factor, more than enough to estimate the linear and quadratic effects, although the quadratic effect columns are not orthogonal to each other. Of course only the factorial points are used in estimating the AB interaction, as only the first four elements in the column obtained from the product of the A and B columns are nonzero.

The objective is generally not to create a 3^2 design, however, but rather to solve for α so as to have a *rotatable* design. The latter is a design for which $\mathrm{Var}(\widehat{Y}_{\mathbf{x}})$ is the same for all points $\mathbf{x}$ (with $\mathbf{x}$ denoting the set of point coordinates) that are equidistant from the center of the design, which is the point (0, 0) when there are two factors, independent of direction from the center. It is well known that this condition is met when $\alpha = \sqrt[4]{F}$, with F denoting the number of points in the factorial part. Thus, $\alpha = \sqrt[4]{4} = \sqrt{2} = 1.414$ when $k = 2$ and the factorial part contains the full factorial. Notice that all the points beyond the centerpoint are equidistant from the origin with this choice of α. (Note that some software packages, such as JMP, will label a CCD with α determined in this manner a *uniform precision design*.)

Since the axial points (which are also called star points since they form a star configuration when combined with the other points) are orthogonal for any choice of α and of course the factorial points are also orthogonal, the design will be orthogonal for estimating the linear factor effects, regardless of the number of centerpoints that are used. The quadratic factor effects will not be orthogonal to each other because that could not happen unless the columns representing the quadratic effects had some negative numbers, but that can't happen since the square of a positive or negative number is a positive number. The columns for the quadratic effects are orthogonal to the linear effects however, and are also orthogonal to the interaction effects.

Speaking of "effects," which were used in only a general way in the preceding paragraph, we can address the question for a CCD. If we consider the CCD for two factors given earlier in this section, we see that observations at the five levels of each factor occur with different frequencies. That is, there is only one observation at each axial point, two observations at each factorial point level and five centerpoints. Computing conditional effects using only the factorial points is potentially useful, however; this is discussed in Section 10.10.

Myers and Montgomery (1995, p. 311) asked the question: "How important is rotatability?" Their view is that a near-rotatable design should be good enough, and the same should probably be said of orthogonality, thus not ruling out a space-filling design on the grounds of nonorthogonality. It should be noted that the quadratic effect estimators as a group are likely to be less correlated with each other with a space-filling design than with a CCD, although this is offset by the fact that the quadratic effect estimators will be correlated with the interaction effect estimators and with the linear effect estimators, and of course the latter two will also be correlated with each other.

Given below is JMP output that shows the correlation between effect estimates when a UD with 19 design points is used to fit a full quadratic model in three factors. Notice that two of the three correlations between the quadratic terms are small and, ignoring correlations with the intercept term, only three of the correlations exceed .2567 in absolute value, with all but one of the other correlations being less than .20 in absolute value. The user would have to decide whether the magnitude of the correlations involving linear and interaction terms offsets the uniform coverage of the design space, but this is not a bad design in terms of the correlations.

Corr

	Intercept	X1	X2	X3	X1*X2	X2*X2	X1*X1	X3*X3	X2*X3	X3*X1
Intercept	1.0000	0.0708	-0.0051	-0.0817	-0.0657	-0.5310	-0.6977	-0.6227	-0.1054	0.0472
X1	0.0708	1.0000	0.0238	-0.0313	0.0440	-0.0895	-0.0108	-0.0501	0.0974	0.0059
X2	-0.0051	0.0238	1.0000	-0.0089	-0.0477	0.0479	0.0382	-0.0723	0.1674	0.1241
X3	-0.0817	-0.0313	-0.0089	1.0000	0.0817	0.1614	0.0086	0.0115	-0.0581	-0.0260
X1*X2	-0.0657	0.0440	-0.0477	0.0817	1.0000	-0.0141	0.0035	0.1592	0.1124	0.2129
X2*X2	-0.5310	-0.0895	0.0479	0.1614	-0.0141	1.0000	0.1679	0.0121	0.0211	0.0886
X1*X1	-0.6977	-0.0108	0.0382	0.0086	0.0035	0.1679	1.0000	0.3239	0.1692	-0.1311
X3*X3	-0.6227	-0.0501	-0.0723	0.0115	0.1592	0.0121	0.3239	1.0000	0.0382	-0.0645
X2*X3	-0.1054	0.0974	0.1674	-0.0581	0.1124	0.0211	0.1692	0.0382	1.0000	0.1851
X3*X1	0.0472	0.0059	0.1241	-0.0260	0.2129	0.0886	-0.1311	-0.0645	0.1851	1.0000

Although rotatability is a desirable property of a response surface design, it loses some of its luster when we think about how the fitted equation that results from the experimentation will likely be used. Once the optimum factor levels have been determined, it would undoubtedly be of interest to obtain a prediction interval for the response. There may not often be a need for a second prediction interval for a combination of factor levels such that the point is the same distance from the center as the point representing the optimum conditions. Furthermore, if one or more factor levels are altered slightly, for whatever reason, from the supposed optimum combination of factor levels, the point representing the combination of levels to be used almost certainly won't be the same distance from the center of the design as the optimum point. Even if the distance were the same, there is no guarantee that the two variances of the fitted values will be the same, since rotatability applies only to points used in the design, not to all points in the region covered by the design that are the same distance from the center as the design points.

Before we look at CCD applications in the next section, we should address the question of whether a CCD gives us too many design points, similar to the discussion at the beginning of Chapter 6 regarding three-level full and fractional factorials. For four factors, a CCD will have 25 points if there is a single centerpoint ($16 + 1 + 8$), whereas a 3^{4-1} design has 27 points—not much difference, and of course CCDs generally have more than one centerpoint. (The effects that can be estimated with the 3^{4-1} design, which is resolution IV, are given by Wu and Hamada, 2000, p. 250.) Wu and Hamada show that all main effects can be estimated, in addition to being able to estimate each two-factor interaction using two degrees of freedom (instead of four df, which each interaction has). Thus, 20 df are used for estimating effects from the 26 df that are available.

Notice that the number of effects to be estimated in a full second-order model is $k + k + \binom{k}{2} = (k^2 + 3k)/2$, whereas the number of points in the CCD is $2^k + 2k + c$, if a full factorial is used, with c denoting the number of centerpoints. This will provide more than enough degrees of freedom for estimating all the effects. In fact, the design is somewhat wasteful in this respect as, for example, there are 20 effects to be estimated for $k = 5$, whereas there would be 43 design points even if only a single centerpoint were used, which is not recommended. Furthermore, we don't need factors with five levels to estimate second-order effects.

We can reduce the number of design points considerably by using either a 2^{k-1} or a 2^{k-2} design in the factorial portion.

10.5.1 CCD Variations

The CCD described in the preceding section is the standard CCD. There are variations of the design that are sometimes useful, however. For example, the region of operability may be defined by the factorial points. If so, no point in the design space could have a coordinate of $\pm\sqrt{2}$, for example. The star points would thus have to be "shortened" to conform to that region, that is, the largest absolute value of any coordinate could not exceed 1. For two factors this means that the factorial and star points constitute a 3^2 design except for the fact that the point (0, 0) is missing, which of course would not be missing from the overall design since the design contains centerpoints. In general, such a design is called a *face center cube* because the axial points occur at the center of the faces of a cuboidal region. We will let this design be designated as CCF so as to distinguish it from the standard CCD.

A face center cube makes sense if the region of interest is cuboidal; if the region of interest is spherical, the standard CCD should be used. The latter should also be used if rotatability is important, since the face center cube is obviously not rotatable since the star points do not have the coordinates that are necessary for rotatability.

If a spherical region is desired but no coordinate can exceed 1 in absolute value, a solution is to "scale down" the design so that the star points fall within the restricted region and the factorial points have smaller coordinates. This can be accomplished by starting with the regular CCD and dividing through by α. For example, for $k = 2$, the nine distinct points would be as follows.

A	B
-0.707	-0.707
0.707	-0.707
-0.707	0.707
0.707	0.707
1	0
-1	0
0	1
0	-1
0	0

This is often called an *inscribed* central composite design, with the notation CCI used in some sources to distinguish this design from the CCD and CCF. Of course the CCI is rotatable since it is just a scaled version of a rotatable design.

A compact summary of the different types of CCDs, including a graphical comparison, is shown in http://www.itl.nist.gov/div898/handbook/pri/section3/pri3361.htm.

10.5.2 Small Composite Designs

There are various ways of constructing composite designs that have fewer runs than a CCD. Such designs date from Hartley (1959), who contended that one could use a resolution III design for the factorial portion, with other methods proposed by Ghosh and Al-Sabah (1996) and Draper and Lin (1990), among others. In addition to being economical, small composite designs have value in that the correlation between the quadratic effect estimates is reduced as the number of factorial points is reduced, as the reader is asked to show in Exercise 10.25. (This follows from the fact that it is the factorial portion of the design that is creating the correlation, which is a perfect correlation for that portion since each column entry is a 1. Therefore, reducing the proportion of factorial-point runs will reduce the correlation.)

10.5.2.1 Draper–Lin Designs

Work on producing composite designs with a fewer number of design points than the composite designs that result from using either a full factorial or a fractional factorial for the factorial portion also includes Westlake (1965), who used irregular fractions of the 2^k system. Draper (1985) used columns of Plackett–Burman designs (see Section 13.4.1) to produce designs for the same number of factors but with fewer design points. Draper and Lin (1990) improved on that set of designs by finding designs for higher values of k and also finding designs that Draper (1985) conjectured did not exist in a singular version. The Draper–Lin designs are not rotatable, however. Another deficiency is that coefficients for terms of the same order, such as linear term coefficients, are not estimated with equal precision. This will be illustrated later with an example. Draper and Lin (1990) did not discuss the choice of α for these designs, but Croarkin and Tobias (2002) indicate that α should be chosen between $(\mathrm{F})^{1/4}$ and $\sqrt{k}$.

Example 10.2

Chapman and Masinda (2003) used what was stated to be a Draper–Lin small composite design with 18 runs for four factors to determine the settings of those factors needed to minimize the effort that is necessary to close a car door on a new prototype design. They also wanted to use the factor settings that were the least sensitive to variations in the process operating conditions and hoped that the same settings would meet both goals.

Although the authors used a randomized run order, the design runs are given, along with the data, in Table 10.1 in an order that allows the nature of the design to be more easily seen.

The authors analyzed these data and arrived at the following fitted equation: $\widehat{Y} = 1.34 - 0.068X_1^2 + 0.021X_2^2 + 0.083X_1X_4 - 0.036X_3X_4$. Notice that there is no main effect in the model, which is as nonhierarchical a model as one could imagine! This would suggest that both the experiment and the data be checked. (Notice also that the sum of the column for A is not zero; the ramifications of this will be discussed later.)

The design given in Table 10.1 is *not* a Draper–Lin design for four factors as described in Draper and Lin (1990). Those designs were constructed by using specific columns from the appropriate Plackett–Burman design, adding to those points the axial points, with centerpoints added if desired. Specifically, for four factors the Plackett–Burman design has eight runs and is a 2_{III}^{4-1} design and they recommend that columns 1, 2, 3, and 6 be chosen so that these runs will constitute the maximum

TABLE 10.1 Data from Chapman and Masinda (1985) Experiment

A	B	C	D	Y
0.5	0	0	0	1.19
0.5	0	0	0	1.18
−0.5	2	2	1	1.33
−0.5	−2	−2	−1	1.31
−0.5	−2	2	−1	1.63
−0.5	2	−2	1	1.40
1.5	2	2	−1	1.14
1.5	−2	−2	1	1.42
1.5	2	−2	−1	1.08
1.5	−2	2	1	1.29
0.5	0	−3.364	0	1.40
0.5	3.364	0	0	1.57
0.5	0	3.364	0	1.29
0.5	−3.364	0	0	1.60
0.5	0	0	−1.682	1.35
0.5	0	0	1.682	1.48
2.182	0	0	0	1.07
−1.182	0	0	0	1.33

"D-value," which they define as the determinant of $X^{\mathrm{T}}X$ divided by n^p. Of course the D-value is computed for the full second-order model that is to be fit. As explained by Draper and Lin (1990), maximizing the D-value causes the design to be the most spread out in the coordinate space.

The 8-run Plackett–Burman (2_{III}^{4-1}) design is as follows.

Run	A	B	C	D	E	F	G
1	+	+	+	−	+	−	−
2	−	+	+	+	−	+	−
3	−	−	+	+	+	−	+
4	+	−	−	+	+	+	−
5	−	+	−	−	+	+	+
6	+	−	+	−	−	+	+
7	+	+	−	+	−	−	+
8	−	−	−	−	−	−	−

When the eight axial points are added to the appropriate columns, this produces a 16-point design that could be used to estimate the four main effects, four quadratic effects, and six interaction effects, in addition to fitting the constant term for a total of 15 model coefficients, although certainly not all the terms would be significant.

A Draper–Lin design can be generated using Design-Expert by specifying that a "small" CCD is to be constructed. The design is given below, in an "unrandomized" order rather than the randomized order given by Design-Expert.

A	B	C	D
-1.00	-1.00	-1.00	-1.00
-1.00	-1.00	1.00	-1.00
-1.00	1.00	-1.00	1.00
1.00	-1.00	1.00	1.00
-1.00	1.00	1.00	1.00
1.00	1.00	-1.00	-1.00
1.00	1.00	1.00	-1.00
1.00	-1.00	-1.00	1.00
1.68	0.00	0.00	0.00
-1.68	0.00	0.00	0.00
0.00	1.68	0.00	0.00
0.00	-1.68	0.00	0.00
0.00	0.00	1.68	0.00
0.00	0.00	-1.68	0.00
0.00	0.00	0.00	1.68
0.00	0.00	0.00	-1.68

This design was first given by Hartley (1959). Notice that the design is constructed using columns 1, 2, 3, and 6 of the Plackett–Burman design. Thus, the design attains the maximum D-value for a fixed Plackett–Burman design from which the columns are selected, and this is also true for the other Draper–Lin designs that are constructed by Design-Expert.

Since 8 is a power of 2, the factorial portion of the design must be a 2^{4-1} design. We can see that $B = -AD$ so $I = -ABD$. Of course this is not the way we would construct a 2^{4-1} design but such designs cannot be used to estimate quadratic effects, and here we are concerned with the best design for estimating the effects in a full quadratic model.

The design employed by Chapman and Masinda (2003), which should have been termed a modification of a Draper–Lin design, can be converted to a Draper–Lin design by dividing the X_2 and X_3 columns by 2 and subtracting 0.5 from each of the X_1 values. This will convert the design into a Draper–Lin design plus two centerpoints. The X_1 values given by the authors do not sum to 0, as they must. Thus the design really isn't a valid design. According to Chapman (personal communication, 2005), the factor-level combinations required by the Draper–Lin design were not feasible, and hence the alterations.

Those alterations created a myriad of problems, however. In addition to the X_1 values not summing to zero, the choice of X_1 values creates a (high) correlation of .707 between X_1 and X_1^2, whereas there is no correlation between linear and pure quadratic terms with a Draper–Lin design. This high correlation probably at least partly explains the existence of the X_1^2 term in the model without the X_1 term. Similarly, the correlation between X_2 and X_1X_4 is $-.641$, which likely helps explain why there is an X_1X_4 term and an X_2^2 term in the model without there being an X_2 term. Finally, when there is a two-factor interaction term in a model (X_3X_4) with neither of the main effects being present, an interaction plot should be constructed, followed by a conditional effects analysis.

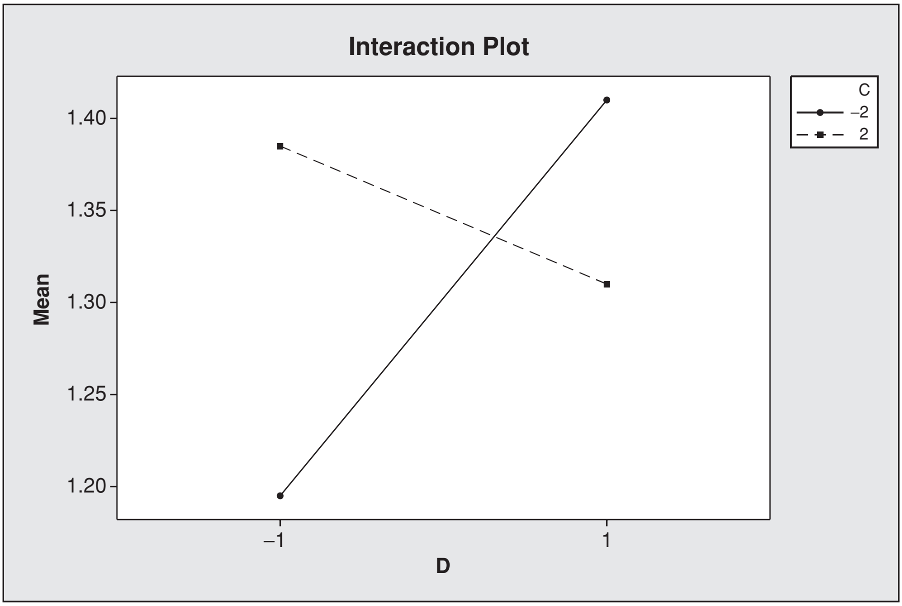

Figure 10.9 Interaction plot using only Plackett–Burman points.

Interaction plots with five levels for each factor do of course create interpretation problems. The advice given later in Section 10.10 is to not use axial points and centerpoints in computing conditional effects. This means that we should also not use such points in constructing interaction plots if the motivation for computing conditional effects is to come from interaction plots. Therefore, the plot in Figure 10.9 is constructed using only the points from the columns of the Plackett–Burman design.

Although each average is computed from only two observations (and thus has a sizable standard error), the strong interaction is apparent, as is the large conditional effect of D at the low (-2) level of C and the large conditional effect of C at the low level of D. Thus, there *are* effects in these two factors; they are simply conditional effects rather than main effects.

Thus, we have shown what resulted in an extremely nonhierarchical model being selected, and how the model is misleading. Analyses of the type given here should *always* be performed whenever a nonhierarchical model is selected.

Example 10.3

We will illustrate the statement made earlier that Draper–Lin designs can have standard errors that differ for terms of the same order. Given below is the Draper–Lin design for four factors constructed using Design-Expert, and using the default value for the number of centerpoints.

A	B	C	D	Y
1.00	-1.00	1.00	1.00	10.1
-1.00	1.00	1.00	1.00	13.1
0.00	0.00	0.00	0.00	12.4
0.00	1.68	0.00	0.00	13.5
1.68	0.00	0.00	0.00	12.7
0.00	0.00	0.00	-1.68	12.6
-1.68	0.00	0.00	0.00	14.2
0.00	-1.68	0.00	0.00	14.3
-1.00	-1.00	-1.00	-1.00	10.8
1.00	-1.00	-1.00	1.00	10.6
0.00	0.00	0.00	0.00	11.9
-1.00	-1.00	1.00	-1.00	12.7
1.00	1.00	1.00	-1.00	13.8
0.00	0.00	0.00	0.00	14.6
0.00	0.00	1.68	0.00	15.2
0.00	0.00	0.00	0.00	16.3
1.00	1.00	-1.00	-1.00	17.1
0.00	0.00	-1.68	0.00	15.5
0.00	0.00	0.00	0.00	16.4
0.00	0.00	0.00	1.68	13.9
-1.00	1.00	-1.00	1.00	11.5

The portion of the output from Design-Expert that shows the standard errors is given below.

Factor	Coefficient Estimate	df	Standard Error
Intercept	14.60	1	0.83
A-A	-0.45	1	0.82
B-B	-0.24	1	0.82
C-C	-0.059	1	0.52
D-D	0.39	1	0.82
AB	1.52	1	1.07
AC	-0.91	1	0.69
AD	-1.65	1	1.07
BC	-0.39	1	0.69
BD	-0.88	1	1.07
CD	0.31	1	0.69
A^2	-0.62	1	0.50
B^2	-0.46	1	0.50
C^2	0.053	1	0.50
D^2	-0.69	1	0.50

Notice that the standard error of the C main effect is less than the standard error of the other three main effects, and the standard errors of the interaction terms also differ. In general, a price has to be paid for small designs of any type and the user must decide if standard errors that differ by this amount are of any concern. (The Draper–Lin design for three factors has equal standard errors for effects of the same type but not the same order as the interaction term coefficients have standard errors that differ from the standard errors of the coefficients of the pure quadratic terms.)

Before leaving this example, a few comments on the computation of quadratic effects seem desirable. The emphasis in this book is on design rather than analysis, as the former cannot be performed by software alone or almost exclusively unless a practitioner has access to expert systems software or software that has a strong expert systems flavor. The computation of quadratic effects is not as simple as the computation of linear effects, partly because there are alternative ways of doing it and there is no computational definition that doesn't have a shortcoming. Therefore, we will not pursue that here. The interested reader is referred to the detailed discussion of Giesbrecht and Gumpertz (2004, pp. 289–298).

10.5.3 Additional Applications

In this section we examine two applications of CCDs and critique the work that was performed.

Example 10.4—Case Study

Wen and Chen (2001) described the application of a CCD after a Plackett–Burman design had been used to identify what seemed to be the important factors. The stated objective was "optimizing EPA production by N. laevis." Of course a screening design is susceptible to the type of problems discussed in Section 10.1.

The CCD that was used by Wen and Chen (2001) was obtained by using a 2^{5-1} design in the factorial part, coupled with six centerpoints and 10 axial points with $(2, -2)$ used as the coordinate values in the star points (i.e., $\alpha = 2$). Note that $\sqrt[4]{16} = 2$, so the condition for rotatability is met with this choice of α. Only the average of the response values for the centerpoints was provided, rather than the individual values, although the individual values were used to estimate σ.

Four response variables were used in the study. The dotplot for the response variable EPA Yield is given in Figure 10.10.

Obviously there is an outlier, which one would reasonably expect to represent an error of some sort, but there was no mention of this observation by Wen and Chen (2001). We will later show a graph that also casts suspicion on the point.

Since the centerpoint observations were not given, it isn't possible to duplicate the analyses by Wen and Chen (2001). It is worth noting, however, that when a full quadratic model in the five factors is fit to the published data, the results in terms of effect significance differ greatly from the results given by Wen and Chen (2001) for each of the response variables. This suggests that the repeated centerpoint values may not be measuring the same thing as the six degrees of freedom that are available for the error term when the average of the values is used. More specifically, the question of whether the response values at the centerpoint are true replicates must be addressed.

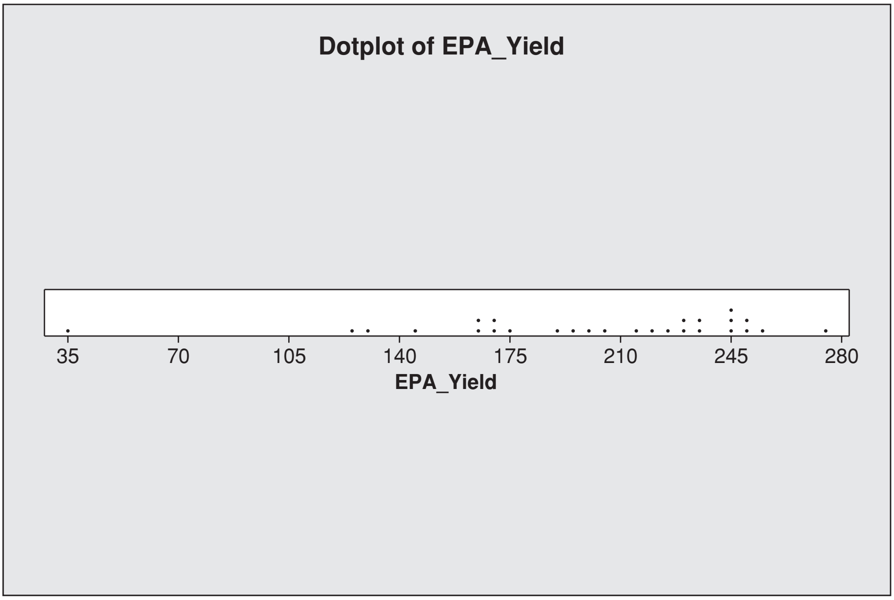

Figure 10.10 Dotplot of EPA yield.

The only factor that shows as being significant for the response EPA Yield when the published data are used is the last factor, temperature, as both the linear and quadratic terms in that factor are significant, the former more so than the latter. The linearity is apparent from Figure 10.11; the quadrature, however, is apparent only if the rightmost point is a valid point. A least squares line fit through the other points in the graph will obviously fall about 100 points above the rightmost point (whose response value is 37.12), suggesting that the first digit should have perhaps been a 1 and the digit is missing. In any event, the point should have been investigated.

As stated, the authors' objective was optimization but we won't pursue our analysis further since some of the data were not published. The moral of this story, however, is that data must be closely examined. "Seek and ye shall find" applies to data analysis as well.

10.6 PROPERTIES OF SPACE-FILLING DESIGNS

Since rotatability is not a relevant issue for a space-filling design because the points are not equidistant from the origin because of the "irregular" configuration of points, and there is also a slight deviation from orthogonality, the designs should be evaluated from other perspectives.

Equileverage designs are discussed extensively in Section 13.6 and have considerable intuitive appeal. An unstated objective in experimental design is that the design points should exert equal, or at least nearly equal, influence on the effect estimates,

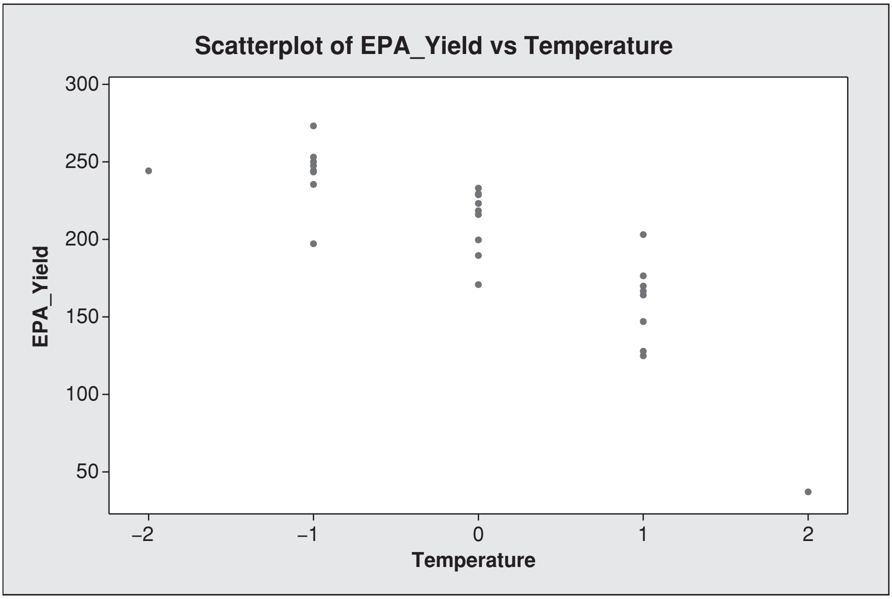

Figure 10.11 Scatterplot of part of Wen and Chen (2001) data.

and equivalently on the regression coefficients in the model representation. That is, the points should have “equal leverage.”

Of course, whether or not an observation is influential will depend in part on the Y-coordinate, but we can improve the chances of not having any influential points by using X-values that are equally influential, that is, by constructing the design so that the design points have equal leverages, remembering that the leverages depend upon the model that is fit.

A leverage value reflects the distance that a point is from the “middle” of the group of points. With this semidefinition in mind, it is easy to see why common designs such as two-level full and fractional factorial designs are equileverage designs since the design points are at the vertices of a hyperrectangular region, such as in Figure 4.5, and are thus equidistant from the center, which would have coordinates of zero for all the factors.

A space-filling design has leverages that can differ considerably, especially in low dimensions, due simply to the “space filling.” That is, the points in the interior of the design have leverages that are much less, relatively speaking, than the points on the periphery of the design space. For example, for the sphere-packing design shown in Figure 10.6, the leverages for fitting the full second-order model range from .266 to .739—a rather wide range. By comparison, the CCD has leverages that are equal (.625) for fitting the full second-order model, except of course for the centerpoint leverage, which is .20.

Since a space-filling design is not orthogonal and differs considerably from the equileverage state because of the interior points, the motivation for using one of these designs must come from the fact that they are much better at detecting nonlinearities, especially complex nonlinearities, than traditional response surface designs. Since the range of possible values for each coordinate must be specified when constructing a space-filling design, very extreme points cannot be constructed, although, as indicated previously, the leverages can still differ considerably.

One property of space-filling designs that will be undesirable in certain applications is that each factor level is constantly being changed since the number of distinct levels of each factor is equal to the size of the design. This is a *major* problem if at least one factor in the experiment is hard to change. Another problem is that the computer-generated levels are not common decimal fractions so the levels will have to be rounded to the nearest feasible values. The multitude of factor-level combinations also increases the likelihood that levels that cannot be used together will be generated (see the discussion of debarred observations in Section 13.8). If the feasible design space is irregular within the permissible range of levels for each factor, there is also an increased likelihood of having points in such regions when many sections of the presumed design space are given points.

Therefore, although space-filling designs are potentially useful, there will be many scenarios when they cannot be used.

10.7 APPLICATIONS OF UNIFORM DESIGNS

Fang and Lin (2003) state that the development of the UD was motivated by three system engineering projects in 1978, as described in Fang (1980) and Wang and Fang (1981). Thus the design was developed to *meet* a requirement and a specific application.

10.8 BOX–BEHNKEN DESIGNS

Since five levels aren't necessary to fit a second-order model, designs with three levels are an alternative. One such design is due to Box and Behnken (1960). These designs can be described as a combination of the balanced incomplete block idea with a 2^2 factorial. Given in Table 10.2 is the Box–Behnken (BB) design for three factors generated by MINITAB, with the number of centerpoints, 3, being the MINITAB default for three factors.

Notice that the 12 points that are not centerpoints consist of the 2^2 factorial designs for all pairs of factors, with a zero filled in for the column in each design that is not used (i.e., the incomplete block idea). Accordingly, the design is orthogonal since it is composed of orthogonal designs. The design is also rotatable because the designs have equal leverages, and the leverages are equal (except for the leverages at the centerpoints) because the design consists of the three possible pairs of 2^2 designs, which are rotatable, ignoring the other factor. Furthermore, the variance of the fitted response is the same at each of the points that are not centerpoints, and those 12 points are all at the same distance, $\sqrt{2}$, from the center. Thus, we should say that the design is rotatable.

TABLE 10.2 Box–Behnken Design for Three Factors and Three Centerpoints

A	B	C
−1	−1	0
1	−1	0
−1	1	0
1	1	0
−1	0	−1
1	0	−1
−1	0	1
1	0	1
0	−1	−1
0	1	−1
0	−1	1
0	1	1
0	0	0
0	0	0
0	0	0

There has been moderate disagreement about the rotatability (or not) of BB designs in the literature, however. Specifically, Myers and Montgomery (1995, p. 321) stated that the designs for four and seven factors are rotatable and the designs for the other numbers of factors (i.e., 3, 5, and 6) are near-rotatable. Wu and Hamada (2000, p. 418) differ slightly, stating "The design for $k = 4$ is known to be rotatable and the other designs are nearly rotatable." At the other extreme, the software user's manual for Statistics Toolbox 4.0 states that "Box–Behnken designs are rotatable designs" (see http://ftp.math.hkbu.edu.hk/help/pdf_doc/stats/rn.pdf and http://www.mathworks.com/access/helpdesk/help/toolbox/stats/f56652.html). If we apply the requirement that the mixed fourth moment must be 3 (see Myers and Montgomery, 1995, p. 309), then we conclude that not all BB designs are rotatable. However, the deciding factor should be whether or not the variance of the fitted response is the same for all design points equidistant from the center. If we adopt that criterion, then the designs are rotatable. The reader is asked to address this issue in Exercise 10.27.

Notice also that there is no design point for which all the factors are at either the high level or the low level. This could be advantageous if there is a restricted region of operability such that such factor-level combinations would not be possible. That is, they would be "debarred observations," a term that has been used in the literature and is also used in Section 13.8. Of course the ABC interaction is not estimable because all the points in the ABC column would be zero, but this could not be termed a weakness of the design because three-factor interactions are generally not significant. (Only when a BB design has at least six factors can any three-factor interaction estimate be obtained, but it is undesirable to compute an interaction estimate from only a few design points, especially when it is a small fraction of all the design points.)

Not only is the BB design devoid of these corner points, but it can also have the same type of restricted operability region advantage over the CCD because the axial points in the CCD may also protrude beyond the operating region. Of course, one way that problem could be avoided would be to use a CCI (see Section 10.5.1), but with that design the factorial points might be farther inside the operability region than desired.

In general, if the region of interest is spherical, a BB design is a good design choice.

It would not be particularly meaningful to compute conditional effects for BB designs for a small number of factors when there are large interactions. This can be explained as follows. Notice from Table 10.2 that all the columns for two-factor interactions will have only four nonzero entries. Thus, each interaction estimate is computed from only 4 of the 15 observations. In general, it is not a good idea to compute *any* effect from only about 1/4 of the design points. Even worse, if a conditional effect were computed by one of the levels involved in the linear main effect estimate, the conditional effect would be computed using only two numbers. Thus, interaction effect estimates would be quite shaky with a BB design with three factors and it would be impractical to compute conditional effects. In fact, interaction effect estimates will be quite shaky with *any* BB design, as is explained in the next paragraph.

For four factors the design will of course have 24 points—$\binom{4}{2} \times 4$—plus the number of centerpoints. Since this as well as all BB designs contain sets of 2^2 factorials with the other column entries being zeros, it follows that *any* two-factor interaction in *any* BB design will be estimated using only four points. Each interaction effect estimate will thus be computed as the difference of two averages that are each computed from only two observations, just as would be the case if the AB interaction were estimated using a 2^2 design. Interaction effect estimates can thus be expected to be *highly variable*, and this is a serious weakness of these designs. This weakness may not have been pointed out previously.

A two-factor interaction could thus be large simply because the estimator has a large variance, which would necessitate computing conditional effects, but these should not be computed from only two numbers. Furthermore, these designs would not be recommended if an experimenter were interested in seeing response values at the vertices of a cuboidal region, as the vertices are not part of a BB design, so the design region is not cuboidal. (Multifactor second-order designs for cuboidal regions were given by Atkinson, 1973.)

Although BB designs have been viewed as being economical relative to CCDs (see, e.g., the comparison at http://www.itl.nist.gov/div898/handbook/pri/section3/pri3363.htm), this is certainly offset by the deficiencies of the design that have been noted earlier in this section.

10.8.1 Application

Example 10.5

An application of a BB design was given by Palamakula, Nutan, and Khan (2004, references). The objective of the study was to determine a model that would produce an optimized self-nanoemulsified capsule dosage form of a highly lipophilic model compound, Coenzyme Q10. There were three factors used in a BB design, so the design was that given in Table 10.2. There were five response variables but four served as constraints relative to the response variable of interest, which was the cumulative percentage of the drug released after 5 minutes.

The authors focused attention on effects whose p-value was less that .05, and those were the following effects: X_2, X_1^2, X_1X_3, X_2^2, and X_3^2. Note that this is very much a nonhierarchical model. Although it has been argued elsewhere that this is not a bad thing (see Section 4.18), it is disturbing that the quadratic terms in the first and third factors are significant without the linear terms being significant. This coupled with the fact that X_1X_3 is significant without the linear terms being significant mandates further analysis, and the reader is asked to perform that analysis in Exercise 10.26, where the data are given.

10.9 CONDITIONAL EFFECTS?

Conditional effects have been discussed and illustrated extensively in previous chapters, especially Chapter 5. Their degree of usefulness for response surface designs depends upon the design. Obviously conditional effects cannot be computed unless levels are repeated, which means that they cannot be used with UDs. Levels would have to be grouped and averages computed, which would not be particularly meaningful.

Conditional effects could be constructed for a CCD. In general, as was emphasized in Section 4.2.1, conditional effects are most meaningful when the averages are computed from more than a few observations, just as effect averages in general are more meaningful when this condition is met. Only the factorial portion of the CCD should be used in computing conditional effects since the centerpoint values cannot be used since each factor has the same coordinate at the centerpoint, so there is no change in coordinate values for any of the factors. Although the axial points could technically be used, each average would actually be just a single observation, so those conditional effects would have large variances. Therefore, it would be practical to compute conditional effects from only the factorial points, and to proceed analogous to the way that conditional effects were used in Chapters 4 and 5.

As stated in Section 10.4, Sztendur and Diamond (2002) did use a conditional effects approach in arriving at a path of steepest ascent. Specifically, an experiment described by Hsieh and Goodwin (1986) involved nine factors and alluded to the analysis of Bisgaard and Fuller (1994–1995) that identified the significant effects as D, F, and one or more of the confounded interactions *BG*, *CJ*, and *EH*. Sztendur and Diamond (2002) assumed that *BG* was the significant interaction (a very bold assumption that is apparently without a basis) and consequently used a model with terms in B, D, F, and BG. Their use of the conditional effect of B that is given by $B + BG$ can be viewed as a way around the problem of having a significant interaction term when the method of steepest ascent is to be employed, as that method cannot be used with interactions. Of course the use of the sum $B + BG$ to produce the conditional effect presupposes that the high level of factor G is the best level in terms of maximizing the response. A clue as to whether or not this is likely to be the case can be found simply by comparing the average response at the high level of G with the average response at the low level of G. Another potential problem is that a conditional effect has a different variance from effects computed using all the data. Thus, summing the regression coefficients, as Sztendur and Diamond (2002) did, produces a regression coefficient that has a smaller standard error than the other coefficients in the model. Furthermore, if the design had been unreplicated, a normal probability plot approach could not have been used because that plot requires effect estimates with equal standard errors.

Their particular application is not justifiable because of (1) their arbitrary selection of one of the three interactions, and (2) their forcing a hierarchical model despite the evidence that nonhierarchical models will often be appropriate (Montgomery, Myers, Carter, and Vining, 2005). Nevertheless, this portion of their article is useful because it reminds us that there is a way out of the problem of having a significant interaction in a model when the method of steepest ascent is to be used.

As with factorial designs, statistical analysis that leads to the selection of a nonhierarchical model could signal the need for a conditional effects analysis, and this does happen in one of the chapter exercises as an interaction plot clearly shows a factor main effect, although the factor was not included in the model.

10.10 OTHER RESPONSE SURFACE DESIGNS

There are many other experimental designs that have been proposed as response surface designs. Such designs might be viewed as secondary response surface designs, and we will look at several of these designs in this section and see whether they should receive greater attention by experimenters.

10.10.1 Hybrid Designs

Roquemore (1976) developed a class of designs that are referred to as *hybrid designs* because the designs are related to CCDs but are constructed to satisfy other criteria. There are multiple designs for each number of factors. For example, for three factors the designs were labeled 310, 311A, and 311B.

Myers and Montgomery (1995, p. 362) stated, “It has been our experience that hybrid designs are not used as much in industrial applications as they should” and go on to indicate that the designs can be useful when experimental runs are expensive. They advise against the use of saturated or near-saturated response surface designs, but reason that such designs are going to be used anyway, and when they are used, a hybrid design would be a good choice.

Therefore, we will consider these designs first, for which some concern has been expressed about possible high leverage values. (A large difference in some leverage values would indicate that, if the response values were very close, some design points would exert noticeably more influence on the model coefficients than other specific points. Leverages are explained fully in Section 13.6.)

One of the undesirable features of the hybrid designs is the odd numbers for many of the factor levels, which will require that the values in the plans be rounded when the designs are applied. For example, in one of the designs for four factors the distinct values for the fourth factor are 1.7844, −1.4945, 0.6444, and −0.9075.

The smallest of these designs is for three factors and has 10 design points. That is a saturated design and leverages cannot be computed for a saturated design. One of the two-factor interactions is computed from only four design points, and if large, the conditional effects would have to be computed from only two points. This of course is the same thing that happens with the BB designs so this particular hybrid design shares that weakness.

One of the designs with three factors and 11 points, labeled the 311A by Roquemore (1976), has equal leverages for most of the design points. The design, constructed using Design-Expert, is given below. (Design-Expert cannot be used to construct either design 310 or design 311B.)

A	B	C
0	0	0
$-\sqrt{2}$	0	$-\sqrt{2}/2$
0	0	$-\sqrt{2}$
-1	-1	$\sqrt{2}/2$
1	1	$\sqrt{2}/2$
$\sqrt{2}$	0	$-\sqrt{2}/2$
0	$\sqrt{2}$	$-\sqrt{2}/2$
0	0	$\sqrt{2}$
1	-1	$\sqrt{2}$
0	$-\sqrt{2}$	$-\sqrt{2}/2$
-1	1	$\sqrt{2}/2$

The third and eighth design points have leverages of 0.75; the other points except the centerpoint all have leverages of 0.9375. Thus, this is close to an equileverage design.

The centerpoint has a leverage value of 0 for estimating the model effects, which should be apparent since it is at the center of the design. Of course centerpoints are not used in computing effect estimates, as stated previously, so they have no influence whatsoever on those effect estimates, but leverage values, in general, measure *potential* influence, not actual influence. When leverages for this design are computed using the **X** matrix consisting of the first column being 1s followed by the nine columns for the nine effects to be estimated, the leverage for the centerpoint is 1. Thus, it has leverage only in regard to the constant term, which is not of any intrinsic interest. This illustrates why leverages should always be computed from the design matrix, and not from the **X** matrix, as we are interested only in the potential influence of points on the effect estimates corresponding to terms in the model. For any orthogonal design (which these designs are not), the estimate of the constant will simply be the average of all the observations, so every observation is weighted equally in that computation. We know that without looking at any leverage values. For nonorthogonal designs, the constant has no simple interpretation other than being the fitted value at each centerpoint.

The other design, 311B, has the same type of leverage structure as the 311A design as eight of the leverages are 0.925 and two are 0.8. The 311B design has the same weakness as the 310 design, however, as most of the design values are odd numbers to four decimal places.

The leverages for the 310 design were given by Jensen (2000); the design is given below.

A	B	C
0	0	1.2906
0	0	-0.1360
-1	-1	0.6386
1	-1	0.6386
-1	1	0.6386
1	1	0.6386
1.1736	0	-0.9273
-1.1736	0	-0.9273
-1	1.1736	-0.9273
1	-1.1736	-0.9273
0	0	0

If, as before, we ignore the fitting of the constant term, the leverage of the second design point is practically zero and the other leverages are practically 1.0, with the centerpoint leverage of course being zero. (If we fit the constant term, the leverages of the 2nd and 11th points are approximately 0.5, but as stated previously, we should compute leverages from the design matrix.)

The three-factor hybrid designs are near-saturated designs as there are 11 observations for estimating the nine effects in the quadratic model plus the constant. The user of Design-Expert receives the following warning message when selecting the design from the menu.

```
Warning! Hybrid designs are minimal point designs. They are
very sensitive to outliers.
```

Of course any saturated or near-saturated design will be sensitive to outliers.

The 16-point hybrid designs for four factors are more useful as half of the two-factor interactions are computed using 8 observations and the other three are obtained using 10 observations.

If hybrid designs are to be used—and they should certainly be considered—it would be a good idea to avoid the designs with fewer than 16 points. Hybrid designs are not available in software such as MINITAB, JMP, or D. o. E. Fusion, however, so their users would have to key in the design. Some of the published hybrid designs can be created with Design-Expert, as indicated previously, although they are a bit hidden as they must be accessed through the menus, selecting Miscellaneous under Response Surface and then selecting Hybrid under Design Type.

10.10.2 Uniform Shell Designs

These designs have been called Doehlert designs, named after their inventor, David Doehlert (1929–1999). Doehlert (1970) and Doehlert and Klee (1972) presented designs that have been termed *uniform shell designs* because the designs for k factors are spread uniformly over a k-dimensional sphere. Despite the word "uniform," these designs should not be confused with UDs, which were discussed in Section 10.3 and which are discussed in more detail in Section 13.9.1.

Despite not being available in the major statistical software, these designs have been used in many applications, two of which are Dumenil, Mattei, Sergent, Bertrand, Laget, and Phan-Tan-Luu (1988) and De Vansay, Zubrzycki, Sternberg, Raulin, Sergent, and Phan-Tan-Luu (1994).

10.10.3 Koshal Designs

These are designs contributed by Koshal (1933) that are highly simplistic and have the minimum number of design points for estimating effects (i.e., the designs are saturated). The designs are discussed in various places in the literature, including Myers and Montgomery (1995, pp. 357–359). For a second-order design with a single centerpoint, the design is given below.

Row	A	B	C
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1
5	2	0	0
6	0	2	0
7	0	0	2
8	1	1	0
9	1	0	1
10	0	1	1

Rows 2–4 are obviously for estimating the linear effects of a factor (and notice that one factor is being changed at a time), rows 5–7 are for estimating the quadratic effects, and rows 8–10 are for the interaction effects. What may not be obvious when the design is written in the above form is that the design is far from being orthogonal. Specifically, the columns of the design have pairwise correlations of $-.33$. Although supersaturated designs can have pairwise correlations of this magnitude, this is a poor correlation structure for a saturated design. There are better designs.

10.10.4 Hoke Designs

Another class of designs that, like Koshal designs, are also economical were given by Hoke (1974). The properties are almost summarized by the title of the paper and we can add that the designs are also nonorthogonal. The designs generalize those given by Rechtschaffner (1967). Although nonorthogonality will generally be acceptable as long as it is slight, users need to know the extent of the nonorthogonality, preferably by viewing pairwise correlations of the columns of the design matrix. There is no discussion of these correlations in Hoke (1974), however, as the author focused attention on the variances of the estimators of the model parameters. None of the correlations of the Hoke designs are large, although the correlations between the columns of the D2 design for three factors are all $-.184$. This design has only 10 points, which is a small design for fitting a second-order model in three factors. The larger designs have smaller correlations, and the D4 design for five factors has orthogonal columns. The Hoke designs are listed at http://www.math.montana.edu/~jobo/cr/designs.txt.

10.11 BLOCKING RESPONSE SURFACE DESIGNS

It may be necessary to block a response surface design for the same reasons as blocking a factorial design may be necessary, such as not being able to make all the runs in one day. Indeed, this was the case in the experiment that is described in Exercise 14.4. This is especially true because response surface designs generally have more design points than full factorial or fractional factorial two-level designs because of the number of levels used in the designs.

10.11.1 Blocking Central Composite Designs

An obvious way to block a CCD is to use two blocks and, if there were no centerpoints, to have the factorial design runs in one block and the axial points in the other block. Doing so would cause one of the pure quadratic effects to become nonestimable, however, but this can be remedied by adding the appropriate number of centerpoints, which will allow all the quadratic effects to be estimable. As stated previously, however, the quadratic effects are not orthogonal to each other.

When blocking is used, it is desirable to use the number of centerpoints and the value of α that make the blocks orthogonal to columns that represent the factor effects and interactions that are being estimated.

For example, with two factors, it is necessary to have six centerpoints and have $\alpha = \sqrt{2}$ in order to achieve these conditions, with the factorial points plus three centerpoints in one block and the axial points and the other three centerpoints in the other block. With three factors, either two or three blocks could be used. In each case six centerpoints would be used. With two blocks, four of those points would be in the block with the factorial runs and the other two would be in the block with the axial point runs, and the axial point values would all be ± 1.633. If three blocks are used, the six centerpoints are equally divided among the three blocks and the axial point values, which are in one block, are all still ± 1.633. The factorial points occupy the other blocks, and they are split in such a way as to confound the *ABC* interaction between those two blocks, which of course is the preferred way of running a 2^3 design in two blocks.

We might ask why, for example, the CCD for three factors with 20 points has $\alpha = 1.682$ when there is no blocking and $\alpha = 1.633$ when either two or three blocks are used, whereas a CCD for four factors has $\alpha = 2$, regardless of whether there is blocking or not? This is because there are different conditions that must be met when blocking is used. These conditions are discussed in detail by Myers and Montgomery (2002).

One of the conditions is that the sum of squares of the individual factor coordinates must be the same in all blocks when the block sizes are the same. When the block sizes are not the same, as will generally be the case when the number of factors is not 2 or 4, the sum of squares must be proportional to the block size across the blocks.

To illustrate, consider the following MINITAB-generated CCD for three factors to be run in three blocks. Of course the blocks cannot be equal since there are 20 design points.

Block 1			Block 2			Block 3		
A	B	C	A	B	C	A	B	C
-1	-1	-1	1	-1	-1	-1.633	0	0
1	1	-1	-1	1	-1	1.633	0	0
1	-1	1	-1	-1	1	0	-1.633	0
-1	1	1	1	1	1	0	1.633	0
0	0	0	0	0	0	0	0	-1.633
0	0	0	0	0	0	0	0	1.633
						0	0	0
						0	0	0

Notice that the factorial points are in the first two blocks and the axial points are in the last block, with two centerpoints in each block. In order for the sum of squares condition to be met, the axial value α must be such that $2\alpha^2/8 = 4/6$, with $2\alpha^2$ being the sum of squares of the columns in Block #3, and 4 being the sum of the squares of each column in each of the other two blocks. Solving this equation for α produces $\alpha = 1.633$.

Each of the first two blocks contains one of the 2^{3-1} designs, plus two centerpoints, and we can see that *ABC* is confounded between the two blocks because the product of the three columns is -1 in the first block (not counting the centerpoints) and $+1$ in the second block.

The value of α necessary for rotatability when a CCD is blocked differs from the value that is needed for orthogonality. This is apparent from the table given by Box and Hunter (1957), which gave blocking schemes for up to seven factors and the value of α for each scheme that is necessary for orthogonality and rotatability. This table was reproduced by Giesbrecht and Gumpertz (2004, p. 410) and by Myers and Montgomery (2002).

10.11.2 Blocking Box–Behnken Designs

A BB design can be blocked as long as at least four factors are used. MINITAB will construct a blocked BB design for four, five, six, or seven factors. Given below is the MINITAB-generated design for four factors, using the default value of 3 for the number of centerpoints and not randomizing the runs (for presentation purposes).

Block	A	B	C	D
1	-1	-1	0	0
1	1	-1	0	0
1	-1	1	0	0
1	1	1	0	0
1	0	0	-1	-1
1	0	0	1	-1
1	0	0	-1	1
1	0	0	1	1
1	0	0	0	0
2	-1	0	-1	0
2	1	0	-1	0
2	-1	0	1	0
2	1	0	1	0
2	0	-1	0	-1
2	0	1	0	-1
2	0	-1	0	1
2	0	1	0	1
2	0	0	0	0
3	-1	0	0	-1
3	1	0	0	-1
3	-1	0	0	1
3	1	0	0	1
3	0	-1	-1	0
3	0	1	-1	0
3	0	-1	1	0
3	0	1	1	0
3	0	0	0	0

The reader is asked in Exercise 10.20 to explain why the design in Table 10.2 could not be run in three blocks.

10.11.3 Blocking Other Response Surface Designs

A Draper–Lin design can be orthogonally blocked as long as there are three, four, or six factors. The designs for 5, 7, 8, 9, and 10 factors cannot be orthogonally blocked because of factorial runs that are not used, the deletion of which causes the design to be nonorthogonal. For these designs, Design-Expert uses the value of α that minimizes the average squared correlation of the block effect with all second-order model coefficients.

For example, given below is the blocked Draper–Lin design for three factors run in two blocks that is generated by Design-Expert when a "small" design is selected from the CCD menu and the orthogonal blocking option is selected and the default value of three centerpoints in each block is used.

Block	A	B	C
1	1	1	-1
1	1	-1	1
1	-1	1	1
1	-1	-1	-1
1	0	0	0
1	0	0	0
1	0	0	0
2	-1.60	0	0
2	1.60	0	0
2	0	-1.60	0
2	0	1.60	0
2	0	0	-1.60
2	0	0	1.60
2	0	0	0
2	0	0	0
2	0	0	0

This design can be orthogonally blocked because the first block contains a 2^2 design (i.e., an orthogonal design) with centerpoints. For five factors, however, there are 11 runs and since that is an odd number, the block that contains them cannot have orthogonal columns, nor can the columns be orthogonal to a column of 1s.

The (seven) options that Design-Expert gives the user regarding the choice of α are worth noting, as they include rotatable, orthogonal blocks, spherical, orthogonal quadratic, practical, face centered, and other. Each of these choices leads to the construction of the indicated design, but "orthogonal quadratic" and "practical" may need some explanation. The former causes the quadratic terms to be orthogonal to the other terms and the latter is the default value for designs with at least six factors. It is defined as the fourth root of the number of factors.

10.12 COMPARISON OF DESIGNS

Some of the well-known response surface designs have been compared in various studies, with the comparison based on one or more efficiency/optimality measures. These measures are discussed in Section 13.7.1. Ozol-Godfrey, Anderson-Cook, and Montgomery (2005) indicate a preference for the CCD and Hoke D6 designs, with the former preferred for the full second-order model and the latter perhaps being superior for the first-order model and the first-order model plus interaction terms. Indeed, the D6 design is the only Hoke design for three factors whose three columns are orthogonal, but the orthogonality does not extend to the interaction effects.

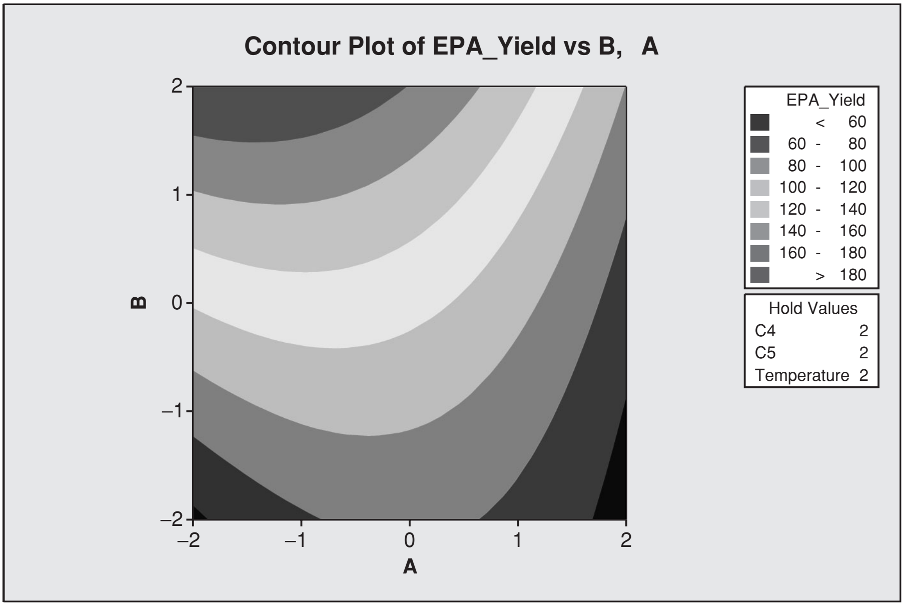

Figure 10.12 Example of a contour plot (MINITAB produced).

10.13 ANALYZING THE FITTED SURFACE

Once a design has been selected, the experiment performed, the data analyzed, and a model fit, the next step is to analyze the fitted surface because optimization is a usual goal of a response surface study. That is, the experimenter is interested in determining the combination of factor settings that will maximize or minimize the expected response, depending upon the objective.

Although it is helpful to see a surface such as Figure 10.1 in which the points are connected, it is generally more useful to look at *contours of constant response*, an example of which is given in Figure 10.12.

This plot is a modification of the plot produced by MINITAB using the RSCONTOUR command. That plot is different from the form of contour plots typically given in textbooks, as the space between the contours is filled in so as to produce a solid graph, and the values associated with the contours are shown in the legend rather than on the graph. Textbooks do not have graphs with several shades of a color, however, so Figure 10.12 is adapted from the multicolor graph produced by MINITAB.

Figure 10.12 is an example of a "rising ridge"; that is, the response continues to increase as one moves toward the upper left portion of the design space. This suggests that the maximum may be outside the design region. There is nothing that can be done in this situation but use a follow-up experiment that has a design center at the largest response on the contour plot and move from there to hopefully envelop the region of the maximum with the design space for the follow-up experiment.

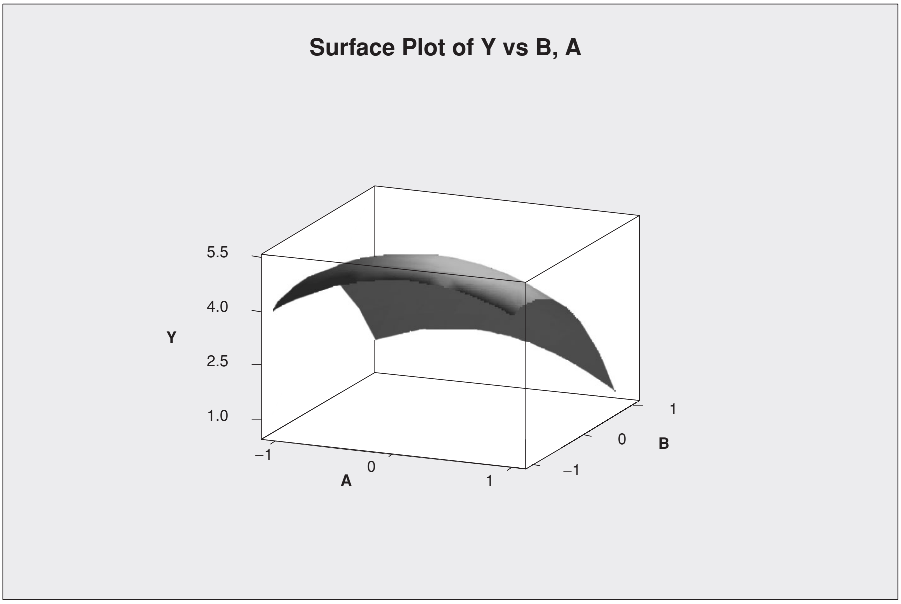

Figure 10.13 Fitted surface that corresponds to Figure 10.1.

The nature of the response surface must be characterized so that the experimenter can determine (1) if the maximum or minimum response appears to lie within the experimental region, and (2) if the maximum or minimum does lie within the experimental region, how much does the response change for small movements in various directions?

Given in Figure 10.13 is the fitted surface for the data from which Figure 10.1 was constructed. Notice that Figure 10.13 differs slightly from Figure 10.1 and of course this is due to the fact that the former is constructed from the following model:

$$\widehat{Y} = 4.847 - 0.017A - 1.550B - 0.918A^2 - 1.318B^2 - 0.325AB \quad (10.1)$$

whereas Figure 10.1 was not constructed from any model. Nevertheless they do not differ by very much and that should certainly be the case since the R^2 value for the fitted model is .998.

The graph suggests that the maximum response lies within the experimental region, and we would guess from the graph that it is in the vicinity of $A = 0$, $B = 0$. In this case it is easy to see from Eq. (10.1) that this is exactly where the maximum occurs because all the coefficients of terms in the model are negative.

In general, it won't be this easy to determine the maximum or minimum in a typical application, so more sophisticated methods must be employed. One problem that must be addressed regardless of the method used is that the coefficients of the squared terms in a second-order model are not independent because the squared terms are not independent (unless they are mean centered, which is not generally done by statistical software).

For example, consider the model

$$Y = 70 + 2A + 3B - A^2 - 2B^2 - AB \tag{10.2}$$

There is obviously no finite global minimum as setting $A = 0$ and $B \to -\infty$ results in $Y \to -\infty$. It appears as though there is a maximum, however, although the maximum is not obvious. What is needed is an analysis of the fitted surface to determine the *stationary point*(s). With only two factors we can make this determination by using calculus and solving a system of equations. That is,

$$\frac{\partial Y}{\partial A} = 2 - 2A - B = 0$$
$$\frac{\partial Y}{\partial B} = 3 - 4B - A = 0$$

Solving these two equations for A and B produces $A = 5/7 = 0.714$ and $B = 4/7 = 0.571$. This produces $Y = 71.5714$, which we can regard as the "true maximum."

Data were generated using the model in Eq. (10.2) and adding an error term $\epsilon \sim N(0, 0.6)$, and the design values of A and B are those from a CCD with five centerpoints and $\alpha = \sqrt{2}$. The fitted equation is

$$\widehat{Y} = 69.810 + 1.973A + 2.985B - 1.159A^2 - 1.908B^2 - 1.414AB \tag{10.3}$$

with $R^2 = .992$. Notice that the coefficients of A^2 and AB are estimated with a large percentage error despite the fact that R^2 is quite high. The correlation between the estimators of the quadratic terms is $-.130$; the estimators of the other terms are uncorrelated.

The fitted surface for the fitted equation (10.3) is given in Figure 10.14.

An analysis of this fitted surface produces

$$\frac{\partial \widehat{Y}}{\partial A} = 1.973 - 2.318A - 1.414B = 0$$
$$\frac{\partial \widehat{Y}}{\partial B} = 2.985 - 3.816B - 1.414A = 0$$

Solving these two equations gives $A = 0.483$ and $B = 0.603$. The value for A differs noticeably from the value of A obtained using Eq. (10.2). The value of $\widehat{Y}$ at this point is 71.19, not far from the true maximum of 71.57 from Eq. (10.2).

It is apparent from Figure 10.14 that the maximum occurs when both A and B are between 0 and 1, and if we look closely at the figure, we can see that the maximum occurs when A and B are each close to 0.5. So the point (0.483, 0.603) thus "looks right."

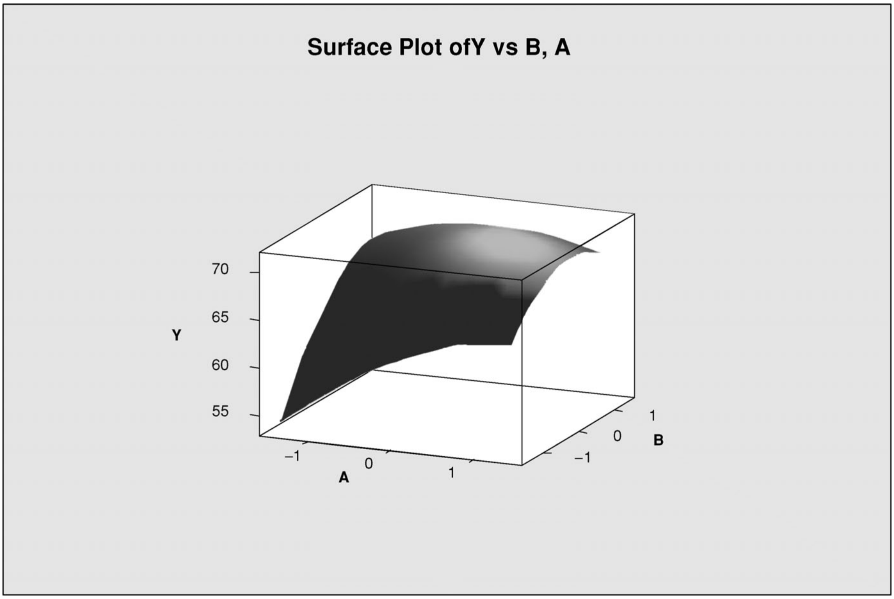

Figure 10.14 Fitted surface for Eq. (10.3).

10.13.1 Characterization of Stationary Points

It is easy to determine whether a stationary point is a maximum or a minimum when there are only two factors since the surface can be easily graphed. This is not the case when there are more than two factors, however. Consequently, there is a need to determine whether a stationary point is a maximum, minimum, or saddle point. (The latter will be illustrated shortly.)

This is done by looking at the eigenvalues of the matrix of coefficients of the second-order terms. We will let W denote that matrix. From Eq. (10.3) we have

$$W = \begin{bmatrix} -1.159 & -1.414 \\ -1.414 & -1.908 \end{bmatrix}$$

(Notice that the coefficient of AB occupies the off-diagonal positions.) Both eigenvalues of this matrix are negative, as can be determined, for example, by using the EIGEN command in MINITAB. Since the eigenvalues are negative, the point must be a maximum. If all the eigenvalues had been positive, the point would have been a minimum, and a mixture of signs means that the point is a saddle point.

We will now illustrate a saddle point. Consider the equation

$$Y = 70 + 2A + 3B - A^2 + 3B^2 - AB \tag{10.4}$$

which is a slight modification of Eq. (10.2), as the coefficient of B^2 is now positive. This causes the signs of the eigenvalues to be mixed, so that using Eq. (10.4) to

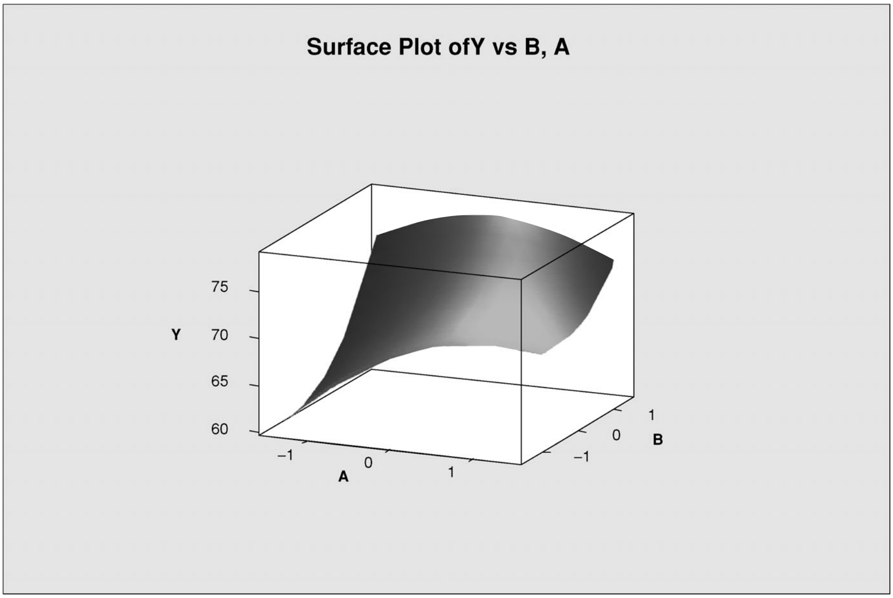

Figure 10.15 Fitted surface for Eq. (10.5).

generate data should result in a fitted surface that resembles a saddle. Generating data with random errors that have the same distribution as before results in the fitted equation

$$\widehat{Y} = 70.1339+2.0672A+2.988B - 0.9301A^2 + 2.9032B^2 - 0.9590AB \qquad (10.5)$$

with the parameter estimates differing only slightly from the corresponding parameters.

The fitted surface for Eq. (10.5) is given in Figure 10.15.

The saddle may not be obvious from Figure 10.15, but should become apparent when the contours of constant response are displayed, analogous to Figure 10.12. This is given in Figure 10.16.

Figure 10.16 isn't a perfect saddle; rather, it looks like a "twisted saddle with one side missing." The point to be made is that the response increases as one starts from the center and moves either up or down, whereas the response decreases if one moves to the left.

10.13.2 Confidence Regions on Stationary Points

In trying to determine optimum factor levels to maximize or minimize the response, it would be helpful to know how well determined the optimum is and how sensitive the response is to slight changes in the factor levels away from the optimum. Just as confidence intervals on parameters are constructed in the application of basic statistical

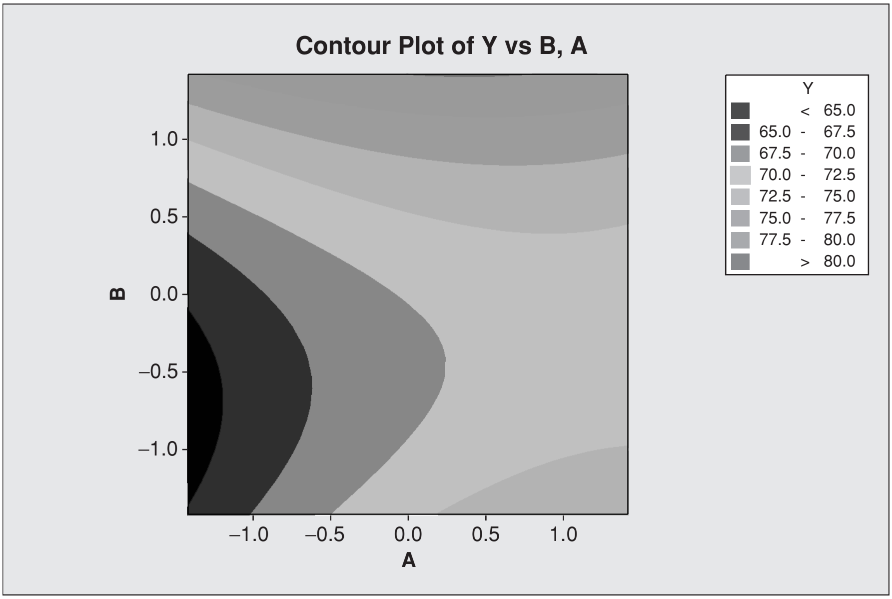

Figure 10.16 Contours of constant response for Eq. (10.5).

procedures, it would be desirable to have a confidence region on the optimum point. (Notice the use of the word "region" instead of "interval" since at least two dimensions will be involved.)

Unfortunately, it is not practical to attempt to compute a confidence region on the optimum point without software. Accordingly, del Castillo, and Cahya (2001) described their software program, accessible from StatLib (http://lib.stat.cmu.edu/TAS) and to be used with MAPLE, and illustrated its use in their article.

10.13.3 Ridge Analysis

Contour plotting is quite useful when there is a small number of factors, but is of very limited use when there is a large number of factors. Ridge analysis, due to Hoerl (1959), is a technique that can be used in determining the optimum combination of factor levels for *any* number of factors. The objective is to determine a path toward the optimum region by using a constrained optimization approach. (Note that this is different from the path of steepest ascent/descent. The latter is determined from a fitted first-order model whereas ridge analysis is generally applied to a second-order model, such as the model given in Eq. (10.1).)

Specifically, the optimum is sought subject to the constraint that $\sum_{i=1}^{k} X_i^2 = R^2$. The method of Lagrangian multipliers is often used to determine the X_i that are the coordinates of the optimum point subject to the constraint, although Peterson, Cahya, and del Castillo (2002) gave a method that does not require the use of Lagrangian multipliers. Their method and the methodology of Gilmour and Draper (2003) are discussed and debated in Peterson, Cahya, and del Castillo (2004) and Gilmour and Draper (2004).

Confidence bands on the path toward the optimum response were introduced by Carter, Chinchilli, Myers, and Campbell (1986), with an improved approach given by Peterson (1993).

Consider the rising ridge shown in Figure 10.12. Since the optimum point appears to lie outside the experimental region, it would be helpful to know the path to take toward the optimum point so that future experimentation could occur along that path. We can make a rough guess of that path from Figure 10.12, but if we had three factors instead of two, we would need multiple contour plots and would have to try to merge the information from those plots.

10.13.3.1 Ridge Analysis with Noise Factors

Peterson and Kuhn (2005) provided an approach for performing a ridge analysis and optimizing a response surface in the presence of noise variables, which extended the work of Peterson (1993). The authors showed how to construct an optimal path for the root mean squared error about a process target. They repeat the message of Hoerl (1985) that contour plots may not be sufficient for understanding high-dimensional response surfaces, especially when there are noise variables. They recommend using an overlaid ridge trace plot to indicate how the process mean differs from the target value as there is movement along the ridge path.

10.13.4 Optimum Conditions and Regions of Operability

It is entirely possible that the methods for determining optimum operating conditions will not be applicable because the region of feasible operating conditions is not hyperrectangular. When this occurs, different methods will have to be employed. Giesbrecht and Gumpertz (2004) discuss this in some detail, especially relative to SAS PROC OPTEX, which permits the selection of a specific subset of points, with the other points spread uniformly over the feasible region.

10.14 RESPONSE SURFACE DESIGNS FOR COMPUTER SIMULATIONS

The use of experimental designs in computer simulation experiments is increasing considerably and response surface designs are among the most frequently used designs for such applications, especially in the field of engineering. For example, Unal, Wu, and Stanley (1997) used a CCD to study a tetrahedral truss for a scientific space platform. Zink, Mavris, Flick, and Love (1999) used a face-centered CCD to improve the wing design of a lightweight fighter jet by assessing a new active aeroelastic wing technology that could only be simulated through physics-based finite element analysis on high-powered computers. See, for example, http://www.statease.com/pubs/rsmsimpexcerpts–chap10.pdf for additional information on the use of response surface designs in computer simulation experiments.

10.15 ANOM WITH RESPONSE SURFACE DESIGNS?

The use of Analysis of Means (ANOM) with various designs has been discussed in earlier chapters. It would not be practical to use ANOM in conjunction with designs like the CCD for various reasons however. First, the levels of a factor are not equally spaced when the CCD is used, and there is only one observation at each of the star points. Therefore, an average computed at a star point wouldn't make any sense. Nor would it make any sense to consider ANOM for use with a UD since the design points generally do not repeat. Although ANOM might be used with a BB design, doing so would present problems since the number of observations at the middle level of each factor differs from the number of observations at each of the other two levels, for each factor. Thus, straightforward use of ANOM would not be possible, although ANOM might be applied by using the data as if there were a single factor with the number of levels equal to three times the number of factors, with an unbalanced ANOM approach then applied.

10.16 FURTHER READING

Response surface methodology, including designs, is a broad field that books have been written about and there is not sufficient space to describe all the methodological advances and interesting applications herein.

Other applications and methodological papers that may be of interest include the following. Huang et al. (2006), mentioned at the start of this chapter, first used a Plackett–Burman design (see Section 13.4.1) to identify significant factors, then used a full factorial CCD in a biological application. Allen, Yu, and Bernshteyn (2000) eschewed the use of a CCD because sufficient resources were not available to carry it out, and instead used a design that the reader is asked to critique in Exercise 10.28. Vining, Kowalski, and Montgomery (2005) discussed the use of a CCD and a BB design run under a split-plot structure and Kowalski, Borror, and Montgomery (2005) gave a modified path of steepest ascent for split-plot experiments.

Morris (2000) presented a new method of constructing composite designs and Gilmour (2006) developed a new class of three-level response surface designs for use in biological applications where the run-to-run variation is generally much greater than it is in engineering experiments. Gilmour (2004) extended the work of Edmondson (1991) in developing four-level response surface designs based on irregular two-level fractional factorial designs.

Draper and John (1998) gave response surface designs for hard-to-change factors, as did Trinca and Gilmour (2001), and response surface designs for both quantitative and qualitative factors were given by Wu and Ding (1998). Akhtar and Prescott (1986) gave designs that were robust to missing observations. Ankenman, Liu, Karr, and Picka (2002) introduced a new class of designs, which they termed *split factorial* designs for response surface applications. Mays and Easter (1997) gave optimal response surface designs in the presence of dispersion effects. Mee (2001) presented noncentral composite designs, which consist of a pair of two-level designs with different centers, as an alternative to the method of steepest ascent. Gilmour and Trinca (2003) gave row–column response surface designs, which are designs that result from blocking. Block and Mee (2001a) presented some new second-order designs and Block and Mee (2001b) presented a table of second-order designs, including those due to Notz (1982).

Box and Draper (1982) gave measures of lack of fit for response surface designs and predictor transformations. Giovannitti-Jensen and Myers (1989) discussed graphical assessment of the prediction capability of response surface designs and Zahran, Anderson-Cook, and Myers (2003) proposed using a fraction of the design space for assessing the prediction capability.

Vining and Myers (1990) introduced the dual response problem, where the focus is on both mean and variance, in the context of RSM, and other papers followed, including more recent papers of Ding, Lin, and Wei (2004) and Jeong, Kim, and Chang (2005).

Myers, Montgomery, Vining, Borror, and Kowalski (2004) gave a response surface literature review.

10.17 THE PRESENT AND FUTURE DIRECTION OF RESPONSE SURFACE DESIGNS

As with statistical methodology in general, the frequency of usage of any "improved" response surface designs will depend upon whether or not these designs are incorporated in software. One relatively recent movement in that direction was the addition of space-filling designs to JMP a few years ago. In his paper "Response surface methodology—current status and future directions," Myers (1999) summarized the then current state of RSM and discussed future directions. One movement in a new direction has been the application of response surface designs and methodology when the appropriate model is a generalized linear model rather than a linear model.

There should be other considerations as well. In particular, designs should be constructed to take into consideration debarred (i.e., impossible) factor-level combinations (Section 13.8) and restricted regions of operability (also Section 13.8). These issues motivate the use of designs other than traditional response surface designs, including optimal designs. Recent work on optimal response surface designs includes Drain, Carlyle, Montgomery, Borror, and Anderson-Cook (2004).

10.18 SOFTWARE

Comments were made in the chapter sections about software availability for the various designs that were covered. In this section we summarize those comments and provide some additional information.

Unfortunately, some of the better designs are not included in major statistical software packages so users of those designs will have to manually construct the designs or find special-purpose software. For example, hardly any software has the capability for Hoke designs and this is true in general for less-known designs.

MINITAB has the capability for CCDs—full, half, and quarter, blocked and unblocked. Box–Behnken designs are the only other response surface designs available, however, and they also can be either blocked or unblocked. The numerical analysis capabilities are the same type as are available with regression analyses; the graphical analyses include surface and contour plots, examples of which were given in the chapter figures. When there are more than two responses, the contour plots can be overlaid. This is discussed in Section 12.1, although it is stated that contour plots start to lose some value when there are more than two factors.

A 55-page discussion of MINITAB's response surface design capabilities, as well as good practical advice on the use of these designs, is available at http://www.minitab.com/support/docs/rel14/14helpfiles/DOE/ResponseSurfaceDesigns.pdf. One potentially problematic feature of MINITAB, which it shares with other statistical software, is that it will not allow the fitting of nonhierarchical models, such as fitting a quadratic term without fitting the linear term. As discussed by Montgomery et al. (2005), there are conditions under which nonhierarchical models can be justified. Specifically, they state that physical and chemical mechanisms are unlikely to be hierarchical. An example of an (extreme) nonhierarchical model was given in a research paper that is cited in Exercise 10.17. As was shown in Example 10.2, however, in which an *extreme* nonhierarchical model found in the literature was discussed, evidence can often be found that does not support the model. As can be inferred from the end of Section 4.18, a nonhierarchical model should *always* be viewed with suspicion and an investigation performed and additional calculations, including conditional effects, made to determine if the model is justifiable.

Design-Expert has the best capability for response surface designs of the software packages that are discussed here. For example, with Design-Expert a BB design can be created for up to 21 factors, whereas MINITAB and JMP do not go beyond 7 factors. Designs with at most 12 factors can be blocked using Design-Expert. The latter allows the user to specify a "user-defined" design by indicating whether or not each category of points (axial check points, interior points, centroid points, vertices, centers of edges, etc.) is to be included in the design. For example, a design with 21 factors can be constructed with 1024 vertices, 5120 centers of edges, 20 constraint plan centroids, 1024 axial check points, 5140 interior points, and 1 overall centroid, for a total of 12,329 design points! Of course only in something like a computer experiment would such a design ever have a chance of being used . . . and computer experiments aren't free! The response surface design capabilities in Design-Expert 7 can be read in the tutorial at http://www.statease.com/x70ug/DX7-04C-MultifactorRSM-P1.pdf.

D. o. E. Fusion guides the user who has no design in mind, as well as allowing the knowledgeable user to reach a menu from which to select a response surface design. Its capabilities for response surface designs fall between MINITAB and Design-Expert. For example, D. o. E. Fusion allows the user to construct a BB design for up to 10 factors. CCDs of resolution IV and resolution V can be constructed, with these being labeled Central Composite IV and Central Composite V, respectively, on the menu. In addition to CCD and BB designs, "Star" designs can also be constructed. These are designs that consist of runs in which each factor except one is at the middle level, plus centerpoints.

JMP can be used to generate the type of space-filling designs discussed in this chapter, including UDs. It will also generate a standard CCD as well as CCD-Orthogonal, CCD-Orthogonal Blocks, and CCD-Uniform Precision. It will also generate both blocked and unblocked BB designs, but as with MINITAB, it will not generate a BB design for more than seven factors. This is perhaps due to the fact that most sources do not discuss or illustrate BB designs for more than seven factors. An exception is Giesbrecht and Gumpertz (2004), who give BB designs for up to 10 factors. Of course the number of design points becomes quite large with that many factors; the design for 10 factors given by Giesbrecht and Gumpertz (2004) has 170 points.

No response surface design can be constructed for more than eight factors with JMP. For three-level designs, only the L_{27} Taguchi design can be constructed when there are more than four factors, and the factors have to be categorical, not continuous. Thus, it would be difficult to use JMP in the "one-design approach" discussed in Section 10.1 because of the limitation on the number of factors that can be handled.

Statgraphics can be used to create Draper–Lin designs, BB designs, and CCDs, and response surface designs that are robust against time trends can be created with the Gendex DOE toolkit.

10.19 CATALOGS OF DESIGNS

In addition to software, there are also catalogs of response surface designs, just as there are catalogs of 3^{k-p} designs as discussed in Section 6.10. Many second-order designs can be found at http://stat.bus.utk.edu/techrpts/2001/2nd_orderdesigns.htm#hartley2. This includes some response surface designs that were not discussed in this chapter.

10.20 SUMMARY

Classical and modern approaches to RSM have been presented in this chapter. Historically, the standard approach has been to use a three-stage procedure: (1) Use a two-level design as a screening design to identify the important factors, (2) conduct experiments along a path of steepest ascent/descent in an effort to identify the region of optimum response, and (3) once that region appears to have been located, use a response surface design to characterize the nature of the surface (assuming that some second-order model is fit) to try to identify the optimum factor settings.

Such an approach will fail if there are significant interactions as then the wrong set of factors will be identified in the first stage. To guard against this possibility, especially if interactions are suspected, a single 3-level resolution V design might be used both for factor screening and for fitting a second-order model, with the design projected onto the factors that seem to be significant. As pointed out in Section 10.1, however, such a design would require a large number of runs.

It is quite likely that newer designs such as uniform designs for irregular regions will play a greater role in response surface methodology applications in the future, once they become widely available in software.

REFERENCES

Akhtar, M. and P. Prescott (1986). Response surface designs robust to missing observations. *Communications in Statistics: Simulation and Computation*, **15**, 345–363.

Allen, T., L. Yu, and M. Bernshteyn (2000). Low-cost response surface methods applied to the design of plastic fasteners. *Quality Engineering*, **12**(4), 583–591.

Anderson, M. J. and P. J. Whitcomb (2004). *RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments*. University Park, IL: Productivity Press.

Ankenman, B. E., H. Liu, A. F. Karr, and J. D. Picka (2002). A class of experimental designs for estimating a response surface and variance components. *Technometrics*, **44**(1), 45–54.

Atkinson, A. C. (1973). Multifactor second order designs for cuboidal regions. *Biometrika*, **60**, 15–19.

Bisgaard, S. (1997). Why three-level designs are not so useful for technological applications. *Quality Engineering*, **9**(3), 545–550.

Bisgaard, S. and H. T. Fuller (1994–1995). Analysis of factorial experiments with defects or defectives as the response. *Quality Engineering*, **7**(2), 429–443.

Block, R. M. and R. W. Mee (2001a). Some new second-order designs. Technical Report 2001-3, Department of Operations, Statistics, and Management Science, University of Tennessee. (http://stat.bus.utk.edu/techrpts)

Block, R. M. and R. W. Mee (2001b). Table of second order designs. Technical Report 2001-1, Department of Operations, Statistics, and Management Science, University of Tennessee. (http://stat.bus.utk.edu/techrpts)

Box, G. E. P. (1957). Evolutionary operation: A method for increasing industrial productivity. *Applied Statistics*, **6**(2), 81–101.

Box, G. E. P. (1999–2000). The invention of the composite design. *Quality Engineering*, **12**(1), 119–122.

Box, G. E. P. and D. W. Behnken (1960). Some new three-level designs for the study of quantitative variables. *Technometrics*, **2**, 455–475.

Box, G. E. P. and N. R. Draper (1969). *Evolutionary Operation*. New York: Wiley.

Box, G. E. P. and N. R. Draper (1975). Robust designs. *Biometrika*, **62**, 347–352.

Box, G. E. P. and N. R. Draper (1982). Measures of lack of fit for response surface designs and predictor variable transformations. *Technometrics*, **24**, 1–8.

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

Box, G. E. P. and J. S. Hunter (1957). Multi-factor experimental designs for exploring response surfaces. *Annals of Mathematical Statistics*, **28**, 195–241.

Box, G. E. P. and K. B. Wilson (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society, Series B*, **13**, 1–45.

Bursztyn, D. and D. M. Steinberg (2001). Rotation designs for experiments in high bias situations. *Journal of Statistical Planning and Inference*, **97**, 399–414.

Carter, W. H., V. M. Chinchilli, R. H. Myers, and E. D. Campbell (1986). Confidence intervals and an improved ridge analysis of response surfaces. *Technometrics*, **28**, 339–346.

Chapman, R. E. and K. Masinda (2003). Response surface designed experiment for door closing effort. *Quality Engineering*, **15**(4), 581–585.

Cheng, S.-W. and C. F. J. Wu (2001). Factor screening and response surface exploration. *Statistica Sinica*, **11**, 553–580; discussion: 581–604.

Cox, D. R. and N. Reid (2000). *The Theory of the Design of Experiments*. Boca Raton, FL: CRC Press.

Croarkin, C. and P. Tobias, eds. (2002). *NIST/SEMATECH e-Handbook of Statistical Methods* (http://www.itl.nist.gov/div898/handbook), joint effort of the National Institute of Standards and Technology and International SEMATECH.

De Vansay, E., S. Zubrzycki, R. Sternberg, F. Raulin, M. Sergent, and R. Phan-Tan-Luu (1994). Gas chromatography of Titan's atmosphere. V. Determination of permanent gases in the presence of hydrocarbons and nitriles with a molecular sieve micropacked column and optimization of the GC parameters using a Doehlert experimental design. *Journal of Chromatography A*, **668**, 161–170.

Del Castillo, E. (1997). Stopping rules for steepest ascent in experimental optimization. *Communications in Statistics: Simulation and Computation*, **26**(4), 1599–1615.

Del Castillo, E. and S. Cahya (2001). A tool for computing confidence regions on the stationary point of a response surface. *The American Statistician*, **55**(4), 358–365.

Ding, R., D. K. J. Lin, and D. Wei (2004). Dual-response surface optimization: A weighted MSE approach. *Quality Engineering*, **16**(3), 377–385.

Doehlert, D. H. (1970). Uniform shell designs. *Applied Statistics*, **19**, 231–239.

Doehlert, D. H. and V. L. Klee (1972). Experimental designs through level reduction of a d-dimensional cuboctahedron. *Discrete Mathematics*, **2**, 309–334.

Drain, D., W. M. Carlyle, D. C. Montgomery, C. Borror, and C. Anderson-Cook (2004). A genetic algorithm hybrid for constructing optimal response surface designs. *Quality and Reliability Engineering International*, **20**, 637–650.

Draper, N. R. (1985). Small composite designs. *Technometrics*, **27**, 173–180.

Draper, N. R. and J. A. John (1998). Response surface designs where levels of some factors are difficult to change. *Australian and New Zealand Journal of Statistics*, **40**, 487–495.

Draper, N. R. and D. K. J. Lin (1990). Small response-surface designs. *Technometrics*, **32**, 187–194.

Dumenil, G., G. Mattei, M. Sergent, J. C. Bertrand, M. Laget, and R. Phan-Tan-Luu (1988). Application of a Doehlert experimental design to the optimization of microbial degradation of crude oil in sea water by continuous culture. *Applied Microbiology and Biotechnology*, **27**, 405–409.

Edmondson, R. N. (1991). Agricultural response surface experiments based on four-level factorial designs. *Biometrics*, **47**, 1435–1448.

Fang, K.-T. (1980). The uniform design: Application of number-theoretic methods in experimental design. *Acta Mathematicae Applicatae Sinica*, **3**, 363–372.

Fang, K.-T. and D. K. J. Lin (2003). Uniform experimental designs and their applications in industry. In *Handbook of Statistics*, Vol. 22, Chap. 4 (R. Khattree and C. R. Rao, eds.). Amsterdam: Elsevier Science B.V.

Fang, K.-T. and R. Mukerjee (2000). A connection between uniformity and aberration in regular fractions of two-level factorials. *Biometrika*, **87**, 193–198.

Ghadge, S. V. and H. Raheman (2006). Process optimization for biodiesel production from mahua (*Madhuca indica*) oil using response surface methodology. *Bioresource Technology*, **97**(3), 379–384.

Gheshlaghi, R., J. M. Scharer, M. Moo-Young, and P. L. Douglas (2005). Medium optimization for hen egg white lysozyme production by recombinant *Aspergillus niger* using statistical methods. *Biotechnology and Bioengineering*, **90**(6), 754–760.

Ghosh, S. and W. S. Al-Sabah (1996). Efficient composite designs with small number of runs. *Journal of Statistical Planning and Inference*, **53**(1), 117–132.

Giesbrecht, F. G. and M. L. Gumpertz (2004). *Planning, Construction, and Statistical Analysis of Comparative Experiments*. Hoboken, NJ: Wiley.

Gilmour, S. G. (2004). Irregular four-level response surface designs. *Journal of Applied Statistics*, **31**(9), 1043–1048.

Gilmour, S. G. (2006). Response surface designs for experiments in bioprocessing. *Biometrics*, **62**, 323–331.

Gilmour, S. G. and N. R. Draper (2003). Confidence intervals around the ridge of optimal response on fitted second-order response surfaces. *Technometrics*, **45**, 333–339.

Gilmour, S. G. and N. R. Draper (2004). Response. *Technometrics*, **46**, 358.

Gilmour, S. G. and L. A. Trinca (2003). Row–column response surface designs. *Journal of Quality Technology*, **35**(2), 184–193.

Giovannitti-Jensen, A. and R. H. Myers (1989). Graphical assessment of the prediction capability of response surface designs. *Technometrics*, **31**, 159–171.

Gorenflo, V. M., J. B. Ritter, D. S. Aeschliman, H. Drouin, B. D. Bowen, and J. M. Piret (2005). Characterization and optimization of acoustic filter performance by experimental design methodology. *Biotechnology and Bioengineering*, **90**(6), 746–753.

Hahn, G. J. (1976). Process improvement through Simplex EVOP. *Chemtech*, **6**, 343–345.

Hardin, R. H. and N. J. A. Sloane (1991). Computer-generated minimal (and larger) response surface designs: (I) The sphere. (available at http://www.research.att.com/~njas/doc/doeh.pdf)

Hartley, H. O. (1959). Smallest composite designs for response surfaces. *Biometrics*, **15**, 611–624.

Hoerl, A. E. (1959). Optimum solution to many variables equations. *Chemical Engineering Progress*, **55**, 69–78.

Hoerl, R. W. (1985). Ridge analysis 25 years later. *The American Statistician*, **39**, 186–192.

Hoke A. T. (1974). Economical second-order designs based on irregular fractions of 3^n factorial. *Technometrics*, **16**, 375–423.

Hsieh, P. and W. Goodwin (1986). Sheet molded compound process improvement. In *Fourth Symposium on Taguchi Methods*, pp. 13–21. Dearborn, MI: American Supplier Institute.

Huang, L. Z., Y. Lu, Y. Yuan, F. Lü, and X. Bie (2006). Optimization of a protective medium for enhancing the viability of freeze-dried *Lactobacillus delbrueckii subsp. bulgaricus* based on response surface methodology. *Journal of Industrial Microbiology and Biotechnology*, **33**(1), 55–61.

Jensen, D. R. (2000). The use of standardized diagnostics in regression. *Metrika*, **52**, 213–223.

Jeong, I.-J., K.-J. Kim, and Y. C. Chang (2005). Optimal weighting of bias and variance in dual response surface optimization. *Journal of Quality Technology*, **37**(3), 236–247.

Khuri, A. I. (1999). Discussion. *Journal of Quality Technology*, **31**(1), 58–60.

Khuri, A. I. (2003). Current modeling and design issues in response surface methodology: GLMs and models with block effects. In *Handbook of Statistics*, Vol. 22, Chap. 6 (R. Khattree and C. R. Rao, eds.). Amsterdam: Elsevier Science B.V.

Khuri, A. I., ed. (2005). *Response Surface Methodology and Related Topics.* Washington, DC: World Scientific.

Khuri, A., and J. Cornell (1996). *Response Surface: Design and Analysis*, 2nd ed. New York: Marcel Dekker.

Kim, B. H. and C. C. Akoh (2005). Modeling of lipase-catalyzed acidolysis of sesame oil and caprylic acid by response surface methodology: Optimization of reaction conditions by considering both acyl incorporation and migration. *Journal of Agricultural and Food Chemistry*, **53**(20), 8033–8037.

Kleijnen, J. P. C., D. den Hertog, and E. Angün (2004). Response surface methodology's steepest ascent and step size revisited. *European Journal of Operational Research*, **159**, 121–131.

Koshal, R. S. (1933). Application of the method of maximum likelihood to the improvement of curves fitted by the method of moments. *Journal of the Royal Statistical Society, Series A*, **96**, 303–313.

Kowalski, S. M., C. M. Borror, and D. C. Montgomery (2005). A modified path of steepest ascent for split-plot experiments. *Journal of Quality Technology*, **37**(1), 75–83.

Lowe, C. W. (1974). Evolutionary operation in action. *Applied Statistics*, **23**(2), 218–226.

Lu, W.-K., T.-Y. Chiu, S.-H. Hung, I.-L. Shih, and Y.-N. Chang (2004). Use of response surface methodology to optimize culture medium for production of poly-γ-glutamic acid by *Bacillus licheniformis*. *International Journal of Applied Science and Engineering*, **2**, 49–58.

Mays, D. P. and S. M. Easter (1997). Optimal response surface designs in the presence of dispersion effects. *Journal of Quality Technology*, **29**, 59–70.

McDaniel, W. R. and B. E. Ankenman (2000). A response surface test bed. *Quality and Reliability Engineering International*, **16**, 363–372.

Mee, R. W. (2001). Noncentral composite designs. *Technometrics*, **43**(1), 34–43.

Mee, R. W. (2004). Optimal three-level designs for response surfaces in spherical experimental regions. Technical Report 2004-3, Department of Statistics, Operations, and Management Science, University of Tennessee (http://stat.bus.utk.edu/techrpts).

Moberg, M., K. E. Markides and D. Bylund (2005). Multi-parameter investigation of tandem mass spectrometry in a linear ion trap using response surface modelling. *Journal of Mass Spectrometry*, **40**(3), 317–324.

Montgomery, D. C., R. H. Myers, W. H. Carter, Jr., and G. G. Vining (2005). The hierarchy principle in designed industrial experiments. *Quality and Reliability Engineering International*, **21**, 197–201.

Morris, M. D. (2000). A class of three-level experimental designs for response surface modeling. *Technometrics*, **42**, 111–121.

Myers (1971). *Response Surface Methodology*. Boston: Allyn and Bacon.

Myers, R. H. (1999). Response surface methodology—current status and future directions. *Journal of Quality Technology*, **31**(1), 30–44; discussion: 45–74.

Myers, R. H. and D. C. Montgomery (1995). *Response Surface Methodology: Process and Product Optimization using Designed Experiments*. New York: Wiley.

Myers, R. H. and D. C. Montgomery (2002). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd ed. New York: Wiley.

Myers, R. H., D. C. Montgomery, G. G. Vining, C. M. Borror, and S. M. Kowalski (2004). Response surface methodology: A retrospective and literature survey. *Journal of Quality Technology*, **36**(1), 53–77.

Nicolai, R. P., R. Dekker, N. Piersma, and G. J. van Oortmarssen (2004). Automated response surface methodology for stochastic optimization models with known variance. In *Proceedings of the 2004 Winter Simulation Conference*, pp. 491–499. (R. G. Ingalls, M. D. Rosetti, J. S. Smith, and B. H. Peters, eds.), The Society for Computer Simulation International, San Diego, CA.

Notz, W. (1982). Minimal point second order designs. *Journal of Statistical Planning and Inference*, **6**, 47–58.

Ozol-Godfrey, A., C. M. Anderson-Cook, and D. C. Montgomery (2005). Fraction of design space plots for examining model robustness. *Journal of Quality Technology*, **37**(3), 223–235.

Palamakula, A., M. T. H. Nutan, and M. A. Khan (2004). Response surface methodology for optimization and characterization of limonene-based Coenzyme Q-10 self-nanoemulsified capsule dosage form. *AAPS PharmSciTech*, **5**(4), Article 66. (This article is available at http://www.aapspharmscitech.org/articles/pt0504/pt050466/pt050466.pdf.)

Peterson, J. J. (1993). A general approach to ridge analysis with confidence intervals. *Technometrics*, **35**, 204–214.

Peterson, J. J., S. Cahya, and E. del Castillo (2002). A general approach to confidence regions for optimal factor levels of response surfaces. *Biometrics*, **58**, 422–431.

Peterson, J. J., S. Cahya, and E. del Castillo (2004). Letter to the editor. *Technometrics*, **46**(3), 355–357.

Peterson, J. J. and A. M. Kuhn (2005). Ridge analysis with noise variables. *Technometrics*, **47**(3), 274–283.

Rechtschaffner, R. (1967). Saturated fractions of 2^n and 3^n factorial designs. *Technometrics*, **9**, 569–575.

Roquemore K. G. (1976). Hybrid designs for quadratic response surfaces. *Technometrics*, **18**, 419–424.

Snee, R. D. (1985). Computer-aided design of experiments—some practical experiences. *Journal of Quality Technology*, **17**(4), 222–236.

Spendley, W., G. R. Hext, and F. R. Himsworth (1962). Sequential applications of simplex designs in optimization and EVOP. *Technometrics*, **4**(4), 441–461.

Steinberg, D. M. and D. Bursztyn (2001). Discussion of "Factor screening and response surface exploration" by S.-W. Cheng and C. F. J. Wu. *Statistica Sinica*, **11**, 596–599.

Sztendur, E. M. and N. T. Diamond (2002). Extensions to confidence region calculations for the path of steepest ascent. *Journal of Quality Technology*, **34**(3), 289–296.

Tang, M., J. Li, L.-Y. Chan, and D. K. J. Lin (2004). Application of uniform design in the formation of cement mixtures. *Quality Engineering*, **16**(3), 461–474.

Tuck, M. G., S. M. Lewis, and J. I. L. Cottrell (1993). Response surface methodology and Taguchi: A quality improvement study from the milling industry. *Journal of the Royal Statistical Society, Series C*, **42**, 671–676.

Trinca, L. A. and S. G. Gilmour (2001). Multistratum response surface designs. *Technometrics*, **43**(1), 25–33.

Unal, R., K. C. Wu and D. O. Stanley (1997). Structural design optimization for a space truss platform using response surface methods. *Quality Engineering*, **9**, 441–447.

Vázquez, M. and A. M. Martin (1998). Optimization of *Phaffia rhodozyma* continuous culture through response surface methodology. *Biotechnology and Bioengineering*, **57**(3), 314–320.

Vining, G. G. and R. H. Myers (1990). Combining Taguchi and response surface philosophies: A dual response approach. *Journal of Quality Technology*, **22**, 38–45.

Vining, G. G., S. M. Kowalski, and D. C. Montgomery (2005). Response surface designs within a split-plot structure. *Journal of Quality Technology*, **37**(2), 115–129.

Wang, Y. and K.-T. Fang (1981). A note on uniform distribution and experimental design. *KeXue TongBao*, **26**, 485–489.

Wen, Z.-Y. and F. Chen (2001). Application of statistically based experimental designs for the optimization of eicosapentaenoic acid production by the diatom *Nitzschia laevis*. *Biotechnology and Bioengineering*, **75**(2), 159–169.

Westlake, W. J. (1965). Composite designs based on irregular fractions of factorials. *Biometrics*, **21**, 324–335.

Wu, C. F. J. and Y. Ding (1998). Construction of response surface designs for qualitative and quantitative factors. *Journal of Statistical Planning and Inference*, **71**, 331–348.

Wu, C. F. J. and M. Hamada (2000). *Experiments: Planning, Analysis, and Parameter Design Optimization*. New York: Wiley.

Xu, H., S. W. Cheng, and C. F. J. Wu (2004). Optimal projective three-level designs for factor screening and interaction detection. *Technometrics*, **46**, 280–292.

Zahran, A. R., C. M. Anderson-Cook, and R. H. Myers (2003). Fraction of design space to assess prediction capability of response surface designs. *Journal of Quality Technology*, **34**, 377–386.

Zink, P. S., D. N. Mavris, P. M. Flick, and M. H. Love (1999). Impact of active aeroelastic wing technology on wing geometry using response surface methodology. Talk given at the *Langley International Forum on Aeroelasticity and Structural Dynamics*, Williamsburg, VA, June 22–25, 1999.

EXERCISES

10.1 Construct and interpret the surface plot for the example given in Section 6.2. Does the plot suggest that something is awry? Explain.

10.2 In their article "Main and interaction effects of acetic acid, furfural, and *p*-hydroxybenzoic acid on growth of ethanol productivity of yeasts," E. Palmqvist, H. Grage, N. Q. Meinander, and B. Hahn-Hägerdal (*Bioengineering* and *Biotechnology*, **63**(1), 46–55, 1999) stated that they used a modified CCD. The factorial portion was a 2^3 design that was replicated three times, with each replicate run as a block, with the blocks corresponding to days. A reference point, which was a point without any added compounds, was used in each block. The axial points were run in separate blocks, and two blocks were used since two replicates of the axial points were used. A reference point was also used in each axial block, so the total number of reference points was 5, one per block.

The full design, including replicates, consisted of 45 runs. The authors gave the design values in the original units, not in coded units. One of the blocks for the factorial portion (they are all the same, except for the run order) is given below, in juxtaposition to one of the axial blocks.

Factorial Block			Axial Block		
0	0	0	0	0	0
2	0.6	0.4	0	1.5	1
8	0.6	0.4	10	1.5	1
2	0.6	1.6	5	1.5	0
8	0.6	1.6	5	1.5	2
2	2.4	0.4	5	0	1
8	2.4	0.4	5	3.0	1
2	2.4	1.6	5	1.5	1
8	2.4	1.6	5	1.5	1

(a) Determine the points in the axial block in coded units.

(b) Assess the design. (Remember that the authors stated that this is a "modified central composite design.") Do you consider this to be a useful design?

10.3 Use appropriate software, such as JMP, to construct a uniform design for four factors in 16 runs, using $(-2, 2)$ as the range for each factor. Compare this with the corresponding (rotatable) CCD with 16 runs for four factors. Comment.

10.4 Assume that an experimenter uses a CCD for three factors, using a 2^3 design for the factorial part and using five centerpoints. The latter were used for estimating σ, which produced a result that was much smaller than the estimate obtained from all the available degrees of freedom for obtaining the estimate. Give one possible reason for this discrepancy.

10.5 In their article, Allen, Yu, and Bernshteyn (2000, references) described a scenario in which engineers decided to use response surface methods to select the design parameters, in the absence of "accurate engineering models." A CCD was eschewed, however, since it required 25 runs and management was willing to guarantee only enough resources to perform 12 experimental runs. Instead, a "low cost" design with only 14 runs was used and the design is given below.

Run	A	B	C	D
1	-0.5	-1.0	-0.5	1.0
2	1.0	1.0	-1.0	1.0
3	-1.0	1.0	1.0	1.0
4	1.0	-1.0	-0.5	-0.5
5	0.0	0.0	-1.0	0.0
6	0.0	1.0	0.0	0.0
7	-0.5	-1.0	1.0	-0.5
8	-1.0	0.0	0.0	0.0
9	1.0	1.0	1.0	-1.0
10	-1.0	1.0	-1.0	-1.0
11	0.0	0.0	0.0	-1.0
12	0.5	-0.5	0.5	0.5
13	0.5	-0.5	0.5	0.5
14	0.5	-0.5	0.5	0.5

(a) Is this a sufficient number of design points to fit a second-order model in four factors? Explain.

(b) What are the properties of the design? Specifically, are there any effects that can be estimated orthogonally? Are there any effects for which the correlation between the estimated effects is undesirable? Explain.

(c) Would you recommend that this design be used? Explain.

10.6 In the article cited in the previous exercise, the authors also gave the results of a study with only 12 experimental runs, with the design and data given below.

Row	A	B	C	D	Y
1	1.25	1.7	12.5	10.00	55.95
2	2.00	2.1	10.0	10.00	101.76
3	1.00	2.1	20.0	10.00	101.23
4	2.00	1.7	12.5	6.25	52.93
5	1.50	1.9	10.0	7.50	59.93
6	1.50	2.1	15.0	7.50	80.54
7	1.25	1.7	20.0	6.25	60.87
8	1.00	1.9	15.0	7.50	72.02
9	2.00	2.1	20.0	5.00	102.70
10	1.00	2.1	10.0	5.00	51.36
11	1.50	1.9	15.0	5.00	59.42
12	1.75	1.8	17.5	8.75	81.94

(a) Notice that the design values were given in raw units rather than coded units. Convert these to coded units.

(b) Can the full second-order model be fit to these data? If not, how might one proceed to determine the effects to estimate?

(c) The authors fit a model with 10 terms. How would you determine significant effects? Would you use ANOVA or some other approach?

(d) Two of the terms in the model fit by the authors were A and A^2. Compute the correlation between the terms for the settings used for A. Now subtract the mean of A from each value of A, square those values, and then compute the correlation between the mean-centered values, and the square of those values. Comment. What would you recommend?

(e) Would you recommend that a different design be used that would be essentially as economical as this one? Explain.

10.7 Response surface methodology has long been used in the food industry. An example is the article "A response surface methodology approach to the optimization of whipping properties of an ultrafiltered soy product" by Carol L. Lah, Munir Cheryan, and Richard E. DeVor (*Journal of Food Science*, **45**, 1720–1726). The objective was to determine optimum conditions for whipping a full-fat soy protein product produced by ultrafiltration. The initial design and the data are given below.

Row	A	B	C	D	E	F	G	Y_1	Y_2
1	-1	-1	-1	-1	-1	-1	-1	97.0	60
2	1	-1	-1	-1	1	1	-1	47.0	100
3	-1	1	-1	-1	1	1	1	76.5	93
4	1	1	-1	-1	-1	-1	1	50.2	0
5	-1	-1	1	-1	1	-1	1	195.0	57
6	1	-1	1	-1	-1	1	1	60.4	93
7	-1	1	1	-1	-1	1	-1	68.3	90
8	1	1	1	-1	1	-1	-1	64.5	100
9	-1	-1	-1	1	-1	1	1	-6.1	100
10	1	-1	-1	1	1	-1	1	16.0	55
11	-1	1	-1	1	1	-1	-1	83.0	100
12	1	1	-1	1	-1	1	-1	9.3	100
13	-1	-1	1	1	1	1	-1	40.3	68
14	1	-1	1	1	-1	-1	-1	24.9	98
15	-1	1	1	1	-1	-1	1	101.0	98
16	1	1	1	1	1	1	1	-10.4	100

The two response variables are overruns (Y_1) and stability (Y_2) and the factors are variables like time, speed, and temperature.

(a) Notice that this is not a CCD. What type of design is it? Be specific. What is the likely purpose of using such a design in view of the title of the article?

(b) Fifteen additional runs were made, using only four of the factors, so that a CCD in those four factors is formed. Do you agree that three of the factors should be dropped from the model for each of the two response variables? If so, which three do you believe must have been dropped?

(c) Would you have proceeded differently? If so, how?

10.8 List the design points for an inscribed CCD (CCI) for four factors. Is the design rotatable? Explain. If not, could the design be made rotatable? Explain.

10.9 What could be the motivation for using a CCI design instead of a standard CCD?

10.10 Explain why a surface plot constructed from a CCD with four factors almost certainly does not represent the true surface.

10.11 Assume that $k = 3$ and a confidence region is constructed on a point that is apparently optimum. Explain why this is termed a "region" instead of an "interval." What will influence the size of the region?

10.12 We consider again the data that were partially analyzed in Exercise 5.39 in Chapter 5 and continue the analysis from the standpoint of response surface methodology. In the experiments described in that exercise, two 2^{5-1} designs were used, with centerpoints used with the first design. The method of steepest ascent was employed after the first experiment, with the second 2^{5-1} design centered at the best combination of steepest ascent design points. Would you have used a different type of design for the second experiment? In particular, do you believe that the experimenters appropriately characterized the nature of the response surface with the results of the second experiment in view of the fact that there were significant interaction effects when the first experiment was performed? (They recognized the significance of one interaction.) Could the method of steepest ascent be used with interactions? Read the article and comment on what they did. What would you have done differently, if anything. Explain.

10.13 In a recent article, Gheshlaghi, Scharer, Moo-Young, and Douglas (2005, references) used a conventional approach to optimization using RSM in a three-stage operation. The first stage consisted of the use of a 2_{V}^{5-1} design with five centerpoints to try to identify significant factors and interactions; the second stage consisted of experimental runs made along what was believed to be the path of steepest ascent; and the final stage consisted of the use of a CCD in three of the five original factors in the region that was believed to contain the optimum combination of factor levels, with the objective being to maximize lysozyme concentration.

Of interest is the following quote from their paper: "The traditional method of optimization involves varying one factor at a time, while keeping the others constant" (p. 754). Whether or not such an inefficient approach is still traditional may be debatable, but perhaps it is still traditional in the authors' field(s).

(a) When only significant terms were included in the model, the model was $Y = 88.25 + 5.5X_1 + 15.38X_2 + 7.13X_3$. Seven design points were used along what was hoped was the path of steepest ascent. Two points are given below but only one coordinate is given for each point. Fill in the missing coordinates.

Point number	X_1	X_2	X_3
22	0.50	—	—
23	—	2.8	—

(b) The final model equation chosen by the experimenters was

$$Y = 208.8 + 7.49X_1 + 3.49X_2 - 7.65X_1^2$$
$$-7.65X_2^2 - 8.71X_3^2 + 4.63X_1X_2 - 4.62X_2X_3$$

Notice that the equation contains the quadratic term in X_3 as well as an interaction term in that variable, without the linear term also being included. Without seeing the data, what would you be inclined to investigate about X_3?

(c) The response values at the high level of factor C in the factorial portion of the design were 182, 187, 170, and 192, and the response values at the low level of C were 171, 181, 176, and 206. The sum of the first pair of four numbers is 731 and the sum of the second pair is 734, the closeness of the two sums suggesting that there is no C effect. The treatment combinations for these eight numbers were, in order, c, ac, bc, abc, (1), a, b, and ab. If you do an appropriate analysis, using only these numbers and ignoring the two axial points on C and the centerpoints, would you conclude that there is no C effect? Explain.

(d) Notice that the stated coefficients for the two interaction terms differ by .01. Using the data in part (b) and thinking about which data points in the CCD are used in computing interaction effects, what should the two coefficients be, recalling the relationship between effect estimates and model coefficients for factorial designs that was given in Appendix B to Chapter 4?

(e) After the CCD was run, 48 experimental runs had been made, yet only the data from the CCD runs were used in computing the model equation. Thinking about the projective properties of fractional factorial designs, as discussed in Section 5.11, would it be possible, or at least practical, to combine those runs with the runs from the CCD? Explain. There were seven points made along the path of steepest ascent, the first two of which were the focus of part (b). Could the response values for those points be used in any way to obtain model coefficients? Explain.

10.14 Gorenflo, Ritter, Aeschliman, Drouin, Bowen, and Piret (2005, references) stated "According to the principle of hierarchy, nonsignificant terms were kept in the model if they were contained in other interaction terms that were found to be significant . . . "

(a) Do you agree with this policy? Could this decision pose any special problems in RSM? Explain. (Note from the equation in part (b) of Exercise 10.13 that such a policy is not followed by all experimenters.)

(b) The following table is from Gorenflo et al. (2005), which lists the p-values of significant effects and one nonsignificant effect for a full quadratic model that is fit using a CCD for five factors that was a 2^{5-1} design in the factorial part. Notice that the table illustrates the above quote.

Parameter	Value	SE	P-value
Intercept	14.56	0.09	<.0001
BX	0.54	0.05	<.0001
PR	1.20	0.03	<.0001
PI	0.44	0.02	<.0001
ST	0.004	0.025	0.86
RR	0.15	0.02	<.0001
PR2	0.40	0.03	<.0001
BX*PR	0.23	0.03	<.0001
BX*PI	0.11	0.02	<.0001
BX*RR	0.16	0.03	<.0001
PR*PI	0.27	0.02	<.0001
PR*ST	0.14	0.03	<.0001
PR*RR	0.13	0.03	0.0001
PI*ST	0.09	0.02	<.0001
ST*RR	0.11	0.03	0.0005

Note: ST main effect kept in model due to principle of hierarchy (significant higher order terms contain respective main effect, e.g., PR ST, ST RR, PI ST); BX—Bioreactor cell concentration; PR—Perfusion rate; PI—Power input; ST—Stop time; RR—Recirculation ratio.

There would be 10 two-factor interactions if all of them were included in the model; notice that 8 of them have very small p-values. Similarly, four of the five main effects are highly significant. The raw data were not given by Gorenflo et al., so the data would have to be obtained to do further analyses. If you had the data, what would you look for/suspect? If after careful analyses, you were in agreement with the numbers given in the table above, would you include ST in the model and would you be concerned about the magnitude of the interactions? Explain.

10.15 The case study given in Section 10.5.3 utilized one of the response variables given by Wen and Chen (2001, references). Another response variable was cell dry weight (DW). The values of that response variable and the corresponding treatment combinations are given below for the factorial part of the design, with the factors simply labeled as A, B, C, D, and E. The axial points were in the order -2, 2 and ordered on the factors A–E. The corresponding response values for the 10 axial points and the average of the six centerpoints were 8.00, 7.70, 7.48, 8.06, 7.51, 7.49, 7.97, 6.15, 1.60, and 7.69 for the average.

Treatment comb.:	e	*d*	*c*	*cde*	*b*	*bde*	*bce*	*a*	*ade*
Response value:	4.90	7.74	7.23	6.13	6.84	7.03	5.57	6.55	7.05
Treatment comb.:	*ace*	*acd*	*abe*	*ab*	*abd*	*abc*	*abcde*		
Response value:	5.80	4.32	7.40	4.25	6.40	5.11	5.75		

(a) Recognizing the limitations caused by the fact that only the average of the centerpoint observations is given, analyze the data and arrive at a fitted model.

(b) Are there any outliers for this response variable as was the case for the response variable EPA Yield, as was shown in Figure 10.10? If so, remove the outlier(s), making the assumption that this is simply bad data, and reanalyze the data. If you conclude that there are no outliers, give appropriate graphical and/or other support for your conclusion.

10.16 Consider part (c) of Exercise 10.13 and construct the *BC* interaction plot. What would you tell the experimenters based on what you see in that plot?

10.17 Lu, Chiu, Hung, Shih, and Chang (2004, references) used RSM to study the effects of L-Glutamic acid, citric acid, glycerol, and NH_4Cl on the production of poly-γ-glutamic acid under certain conditions. The strategy that they employed was to first use a 2^4 factorial design for the purpose of determining the path of steepest ascent. Experiments were conducted along that path as long as the first-order model was adequate. A CCD was then used for those four factors with ± 2 as the axial point values, and four centerpoints were used. In this study, the experimenters knew what factors they wanted to study; hence, there was no need for a screening design.

(a) The axial point values combined with the centerpoints means that each factor was measured at five equispaced levels. Accordingly, could the design be the equivalent of some 5^{4-k} design? Why, or why not?

(b) The fitted equation given by the authors was $\widehat{Y} = 35.34 - 3.82X_1^2 - 4.34X_2^2 - 3.83X_3^2 - 7.20X_4^2$, with X_1 through X_4 denoting the four factors and only significant terms used in the model. Notice that there are no linear (or interaction) terms, so the model is (strongly) nonhierarchical, and the coefficient of each term is negative. Without necessarily knowing anything about the subject matter, would you be inclined to question this equation just on general grounds? Why, or why not?

(c) The data and the design as they appeared in the paper are given below, except that here the factors are arbitrarily labeled as *A–D*. Note that although the first column is labeled "Run," there is no indication of the actual run order that was used.

Run	A	B	C	D	Y
1	31.0	18.0	140.0	6.0	22.17
2	31.0	18.0	140.0	16.0	23.16
3	31.0	18.0	152.0	6.0	23.67
4	31.0	18.0	152.0	16.0	23.03
5	31.0	34.0	140.0	6.0	23.52
6	31.0	34.0	140.0	16.0	28.97
7	31.0	34.0	152.0	6.0	27.11
8	31.0	34.0	152.0	16.0	22.78
9	37.0	18.0	140.0	6.0	36.66
10	37.0	18.0	140.0	16.0	28.67
11	37.0	18.0	152.0	6.0	18.87
12	37.0	18.0	152.0	16.0	23.16
13	37.0	34.0	140.0	6.0	28.14
14	37.0	34.0	140.0	16.0	21.09
15	37.0	34.0	152.0	6.0	27.23
16	37.0	34.0	152.0	16.0	26.77
17	28.0	26.0	146.0	11.0	27.67
18	40.0	26.0	146.0	11.0	29.41
19	34.0	10.0	146.0	11.0	23.94
20	34.0	42.0	146.0	11.0	31.07
21	34.0	26.0	134.0	11.0	24.98
22	34.0	26.0	158.0	11.0	32.07
23	34.0	26.0	146.0	1.0	22.26
24	34.0	26.0	146.0	21.0	21.32
25	34.0	26.0	146.0	11.0	36.86
26	34.0	26.0	146.0	11.0	34.82
27	34.0	26.0	146.0	11.0	34.86
28	34.0	26.0	146.0	11.0	34.80

(d) Do you agree with the model that was selected by the authors? They stated that their model had an R^2 value of .707. Is that correct?

(e) Use whatever supplementary graphs and/or numerical analyses that are necessary to explain why the coefficients of the quadratic terms are negative. After looking at these results, what questions, if any, would you ask of the experimenters/authors?

10.18 If you were going to determine the significant factors from a group of candidate factors and then try to determine the optimum combination of levels of those factors so as to maximize the response, would you do so by starting with a three-level design or with a two-level design? Explain.

10.19 Explain why the method of steepest ascent cannot be applied to a fitted model that contains an interaction term.

10.20 (Harder problem) Explain why the Box–Behnken design in Table 10.2 cannot be run in three blocks with one centerpoint in each block.

10.21 (Harder problem) There is no Box–Behnken design for two factors. Explain why such a design could not be constructed.

10.22 (Harder problem) Explain why it is not possible to construct a Box–Behnken design with only two nonzero numbers in each row of the design.

10.23 As discussed in Section 10.5.2.1, specific columns of a Plackett–Burman design are used in constructing Draper–Lin designs. Determine the difference in the D-values for the 16-run designs for four factors constructed using columns 1, 2, 3, and 6 of the 8-run Plackett–Burman design and columns 1, 2, 3, and 4 of that design.

10.24 Assume that a Draper–Lin design for four factors is used and the 16 response values are 10.2, 10.7, 11.5, 12.6, 13.4, 12.2, 15.1, 10.9, 11.9, 11.3, 12.2, 14.1, 15.8, 13.6, 13.9, and 15.0. Fit the full second-order model and compare the standard errors of the estimates. Is the difference in precision of any concern? Explain.

10.25 Consider a full CCD for five factors with 48 points: 32 factorial points, 10 axial points, and 6 centerpoints. Compute the correlation between the quadratic effect columns, then do the same for a half CCD for the same number of factors with 32 points: 16 factorial points, 10 axial points, and 6 centerpoints. Compare the two sets of correlation coefficients and comment.

10.26 The results of an experiment in which a Box–Behnken design for three factors was used was described in Example 10.5. The response values and the factors in raw units are given below.

Limonene (mg)	Cremophor (mg)	Capmul GMO-50 (mg)	%dissolved in 5 minutes
81	57.6	7.2	44.4
81	7.2	7.2	6.0
18	57.6	7.2	3.75
18	7.2	7.2	1.82
81	32.4	12.6	18.2
81	32.4	1.8	57.8
18	32.4	12.6	68.4
18	32.4	1.8	3.95
49.5	57.6	12.6	58.4
49.5	57.6	1.8	24.8
49.5	7.2	12.6	1.60
49.5	7.2	1.8	12.1
49.5	32.4	7.2	81.2
49.5	32.4	7.2	72.1
49.5	32.4	7.2	82.06

(a) Perform an analysis of the data, using conditional effects and any supplementary computations if necessary, and compare your conclusions with those mentioned in Example 10.5 and with the conclusions reached in the article, which is available at http://www.aapspharmscitech.org/articles/pt0504/pt050466/pt050466.pdf.

(b) Consider Figure 3 in Palamakula, Nutan, and Khan (2004), which the authors claim shows the "quadratic effect of interactions." Do you agree with that conclusion? Why, or why not?

10.27 Consider the Box–Behnken design for three factors that was given in Table 10.2. Simulate a set of response values and determine the variance at each of the design points that is not a centerpoint. What do you observe and what do you conclude about the rotatability (or not) of the design?

10.28 Allen et al. (2000, references) were faced with an experimental situation for which there were four factors to be investigated but there were sufficient resources for only 12 experimental runs. The design that they used is given below.

A	B	C	D
-0.5	-1.0	-0.5	1.0
1.0	1.0	-1.0	1.0
-1.0	1.0	1.0	1.0
1.0	-1.0	-0.5	-0.5
0.0	0.0	-1.0	0.0
0.0	1.0	0.0	0.0
-0.5	-1.0	1.0	-0.5
-1.0	0.0	0.0	0.0
1.0	1.0	1.0	-1.0
-1.0	1.0	-1.0	-1.0
0.0	0.0	0.0	-1.0
0.5	-0.5	0.5	0.5

(a) Would you recommend that this design be used for fitting a full quadratic model if there were *no* cost constraints? Why, or why not?

(b) Is this an orthogonal design? If not, is the degree of nonorthogonality likely to be of any consequence? Explain.

(c) How would one determine what effects in a full quadratic model are significant with this design? (*Hint*: The discussion in Section 13.7 may be helpful.)