Primary focus of previous chapters is **factor screening**
- Two-level factorials, fractional factorials are widely used

Objective of **RSM** is optimization

RSM dates from the 1950s; early applications in chemical industry

Modern applications of RSM span many industrial and business settings
11.1 Response Surface Methodology

- Collection of **mathematical and statistical techniques** useful for the modeling and analysis of problems in which a response of interest is influenced by several variables
- Objective is to **optimize the response**
Steps in RSM

1. Find a suitable approximation for $y = f(x)$ using LS {maybe a low – order polynomial}

2. Move towards the region of the optimum

3. When curvature is found find a new approximation for $y = f(x)$ {generally a higher order polynomial} and perform the “Response Surface Analysis”
Response Surface Models

• Screening
  \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon \]

• Steepest ascent
  \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \]

• Optimization
  \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon \]
RSM is a Sequential Procedure

- Factor screening
- Finding the **region** of the optimum
- **Modeling & Optimization** of the response

**FIGURE 11.3** The sequential nature of RSM
11.2 Method of Steepest Ascent

- Text, Section 11.2
- A procedure for moving sequentially from an initial “guess” towards the region of the optimum
- Based on the fitted first-order model
  \[
  \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2
  \]
- Steepest ascent is a gradient procedure
Example 11.1: An Example of Steepest Ascent

| TABLE 11.1 |
| Process Data for Fitting the First-Order Model |

<table>
<thead>
<tr>
<th>Natural Variables</th>
<th>Coded Variables</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \xi_1 )</td>
<td>( \xi_2 )</td>
<td>( x_1 )</td>
</tr>
<tr>
<td>30</td>
<td>150</td>
<td>-1</td>
</tr>
<tr>
<td>30</td>
<td>160</td>
<td>-1</td>
</tr>
<tr>
<td>40</td>
<td>150</td>
<td>1</td>
</tr>
<tr>
<td>40</td>
<td>160</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>155</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>155</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>155</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>155</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>155</td>
<td>0</td>
</tr>
</tbody>
</table>
What is an Appropriate Model?

• Is the first-order model adequate?
  \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \]

• Is there an interaction effect?
  \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon \]

• Is there any quadratic effect?
  \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \epsilon \]
§6.8 Role of Center Points in the $2^k$ Design

1. Obtain an estimate of error
2. Check for interactions (cross-product terms) in the model
3. Check for quadratic effects (curvature)
   • The center points do not affect the estimate of main effects and 2-factor interaction. Why?
   • The replicates at the center can be used to calculate an estimate of error as follows:

\[
\hat{\sigma}^2 = \frac{(40.3)^2 + (40.5)^2 + (40.7)^2 + (40.2)^2 + (40.6)^2 - (202.3)^2/5}{4} = 0.0430
\]
Check for Interaction

\[ H_0 : \beta_{12} = 0 \text{ vs. } H_1 : \beta_{12} \neq 0 \]

\[ \hat{\beta}_{12} = (39.3 + 41.5 - 40.0 - 40.9)/4 = -0.025 \]

\[ se(\hat{\beta}_{12}) = \sqrt{\hat{\sigma}^2 / 4} = \sqrt{0.043/4} = 0.104 \]

\[ t = \frac{\hat{\beta}_{12}}{se(\hat{\beta}_{12})} = \frac{-0.025}{0.104} = -0.240. \]

\[ df = 4 \text{ and } p\text{-value} = 0.822. \]

Accept \( H_0 \). No interaction.
Check for Curvature

• Compare the average response of the four factorial points with the average of the five runs at the center points.

• If the difference is small, then the center points lie on or near the plane passing through the factorial points, and there is no quadratic curvature.

• On the other hand, if the difference is large, then the quadratic curvature is present.

• Use a t or F test to test hypotheses:

\[ H_0 : \beta_{11} + \beta_{22} = 0 \text{ vs. } H_1 : \beta_{11} + \beta_{22} \neq 0 \]
Check for Curvature

\[
\bar{y}_f = \frac{(39.3 + 40.0 + 40.9 + 41.5)}{4} = 40.425
\]

\[
\bar{y}_c = \frac{(40.3 + 40.5 + 40.7 + 40.2 + 40.6)}{5} = 40.46
\]

\[
\hat{\beta}_{11} + \hat{\beta}_{22} = \bar{y}_f - \bar{y}_c = 40.425 - 40.46 = -0.035
\]

\[
se(\hat{\beta}_{11} + \hat{\beta}_{22}) = \sqrt{\frac{\hat{\sigma}^2}{4} + \frac{\hat{\sigma}^2}{5}} = \sqrt{\frac{0.043}{4} + \frac{0.043}{5}} = 0.139
\]

\[
t = \frac{-0.035}{0.139} = -0.252. \text{ p-value} = 0.813.
\]

Accept $H_0$. No curvature.
x1=c(-1,-1,1,1,0,0,0,0,0); x2=c(-1,1,-1,1,0,0,0,0,0)
y=c(39.3, 40, 40.9, 41.5, 40.3, 40.5, 40.7, 40.2, 40.6)  # Table 11.1
table 11.1
> var(y[5:9])  # estimate of sigma^2
[1] 0.043
> g=lm(y~x1+x2+x1*x2+I(x1^2)+I(x2^2)); summary(g)

Coefficients: (1 not defined because of singularities)

        Estimate Std. Error t value   Pr(>|t|)
(Intercept)  40.4600     0.0927   436.3 2.66e-10 ***
  x1         0.7750     0.1037    7.475   0.0017 **
  x2         0.3250     0.1037    3.135   0.0350 *
  I(x1^2)   -0.0350     0.1391   -0.252   0.8137
  I(x2^2)      NA         NA      NA      NA
  x1:x2     -0.0250     0.1037   -0.241   0.8213

Residual standard error: 0.2074 on 4 degrees of freedom
Multiple R-Squared: 0.9427,      Adjusted R-squared: 0.8854
F-statistic: 16.45 on 4 and 4 DF,  p-value: 0.00947
> anova(g)  # Table 11.2

             Df Sum Sq Mean Sq F value   Pr(>F)
  x1            1  2.4025  2.4025 55.8721 0.001713 **
  x2            1  0.4225  0.4225  9.8256 0.035030 *
  I(x1^2)      1  0.0027  0.0027  0.0633 0.813741
  x1:x2        1  0.0025  0.0025  0.0581 0.821316
Residuals     4  0.1720  0.0430  0.0430

Using R
Steep Ascent Direction

The model $\hat{y} = 40.44 + 0.775x_1 + 0.325x_2$ is adequate.

To move away from the design center—the point $(x_1 = 0, x_2 = 0)$—along the path of steepest ascent, we would move 0.775 units in the $x_1$ direction for every 0.325 units in the $x_2$ direction. Thus, the path of steepest ascent passes through the point $(x_1 = 0, x_2 = 0)$ and has a slope 0.325/0.775. The engineer decides to use 5 minutes of reaction time as the basic step size. Using the relationship between $\xi_1$ and $x_1$, we see that 5 minutes of reaction time is equivalent to a step in the coded variable $x_1$ of $\Delta x_1 = 1$. Therefore, the steps along the path of steepest ascent are $\Delta x_1 = 1.0000$ and $\Delta x_2 = (0.325/0.775) = 0.42$.

### Table 11.2

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>$F_0$</th>
<th>$P$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model ($\beta_2$, $\beta_2$)</td>
<td>2.8250</td>
<td>2</td>
<td>1.4125</td>
<td>47.83</td>
<td>0.0002</td>
</tr>
<tr>
<td>Residual</td>
<td>0.1772</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Interaction)</td>
<td>(0.0025)</td>
<td>1</td>
<td>0.0025</td>
<td>0.058</td>
<td>0.8215</td>
</tr>
<tr>
<td>(Pure quadratic)</td>
<td>(0.0027)</td>
<td>1</td>
<td>0.0027</td>
<td>0.063</td>
<td>0.8142</td>
</tr>
<tr>
<td>(Pure error)</td>
<td>(0.1720)</td>
<td>4</td>
<td>0.0430</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3.0022</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 11.3
Steepest Ascent Experiment for Example 11.1

<table>
<thead>
<tr>
<th>Steps</th>
<th>Coded Variables</th>
<th>Natural Variables</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$\xi_1$</td>
</tr>
<tr>
<td>Origin</td>
<td>0</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>1.00</td>
<td>0.42</td>
<td>5</td>
</tr>
<tr>
<td>Origin + $\Delta$</td>
<td>1.00</td>
<td>0.42</td>
<td>40</td>
</tr>
<tr>
<td>Origin + 2$\Delta$</td>
<td>2.00</td>
<td>0.84</td>
<td>45</td>
</tr>
<tr>
<td>Origin + 3$\Delta$</td>
<td>3.00</td>
<td>1.26</td>
<td>50</td>
</tr>
<tr>
<td>Origin + 4$\Delta$</td>
<td>4.00</td>
<td>1.68</td>
<td>55</td>
</tr>
<tr>
<td>Origin + 5$\Delta$</td>
<td>5.00</td>
<td>2.10</td>
<td>60</td>
</tr>
<tr>
<td>Origin + 6$\Delta$</td>
<td>6.00</td>
<td>2.52</td>
<td>65</td>
</tr>
<tr>
<td>Origin + 7$\Delta$</td>
<td>7.00</td>
<td>2.94</td>
<td>70</td>
</tr>
<tr>
<td>Origin + 8$\Delta$</td>
<td>8.00</td>
<td>3.36</td>
<td>75</td>
</tr>
<tr>
<td>Origin + 9$\Delta$</td>
<td>9.00</td>
<td>3.78</td>
<td>80</td>
</tr>
<tr>
<td>Origin + 10$\Delta$</td>
<td>10.00</td>
<td>4.20</td>
<td>85</td>
</tr>
<tr>
<td>Origin + 11$\Delta$</td>
<td>11.00</td>
<td>4.62</td>
<td>90</td>
</tr>
<tr>
<td>Origin + 12$\Delta$</td>
<td>12.00</td>
<td>5.04</td>
<td>95</td>
</tr>
</tbody>
</table>
The engineer computes points along this path and observes the yields at these points until a decrease in response is noted. The results are shown in Table 11.3 in both coded and natural variables. Although the coded variables are easier to manipulate mathematically, the natural variables must be used in running the process. Figure 11.5 plots the yield at each step along the path of steepest ascent. Increases in response are observed through the tenth step; however, all steps beyond this point result in a decrease in yield. Therefore, another first-order model should be fit in the general vicinity of the point ($\xi_1 = 85, \xi_2 = 175$).

**Figure 11.5** Yield versus steps along the path of steepest ascent for Example 11.1
A new first-order model is fit around the point ($\xi_1 = 85, \xi_2 = 175$). The region of exploration for $\xi_1$ is [80, 90], and it is [170, 180] for $\xi_2$. Thus, the coded variables are

$$x_1 = \frac{\xi_1 - 85}{5} \quad \text{and} \quad x_2 = \frac{\xi_2 - 175}{5}$$

Once again, a $2^2$ design with five center points is used. The experimental design is shown in Table 11.4.

The first-order model fit to the coded variables in Table 11.4 is

$$\hat{y} = 78.97 + 1.00x_1 + 0.50x_2$$
• The first-order model is not adequate.
• Call for further experimentation and analysis.
Using R

```r
> x1=c(-1,-1,1,1,0,0,0,0,0); x2=c(-1,1,-1,1,0,0,0,0,0)
> y=c(76.5, 77.0, 78, 79.5, 79.9, 80.3, 80, 79.7, 79.8)  # Table 11.4
> g=lm(y~x1+x2+x1*x2+I(x1^2)+I(x2^2)); summary(g)
Coefficients: (1 not defined because of singularities)
           Estimate Std. Error t value Pr(>|t|)
(Intercept)   79.9400     0.1030 776.446  1.65e-11 ***
x1             1.0000     0.1151   8.687  0.000966 ***
x2             0.5000     0.1151   4.344  0.012217 *
I(x1^2)       -2.1900     0.1544 -14.181  0.000144 ***
I(x2^2)           NA         NA       NA       NA
x1:x2          0.2500     0.1151   2.172  0.095611 .
```

Residual standard error: 0.2302 on 4 degrees of freedom
Multiple R-Squared: 0.9868,   Adjusted R-squared: 0.9737
F-statistic: 75.04 on 4 and 4 DF,  p-value: 0.0005143

```r
> anova(g)  # Table 11.5
                      Df Sum Sq Mean Sq F value    Pr(>F)
(Intercept)             1  0.000  0.000 75.472  0.0009664 ***
x1                      1  4.000  4.000  18.868  0.0122172 *
x2                      1  1.000  1.000  201.094  0.0001436 ***
I(x1^2)                  1 10.658 10.658    NA       NA
x1:x2                   1  0.250  0.250    NA       NA
Residuals               4  0.212  0.053
```

19
• Points on the path of steepest ascent are proportional to the magnitudes of the model regression coefficients
• The direction depends on the sign of the regression coefficient
• Step-by-step procedure:

1. Choose a step size in one of the process variables, say \( \Delta x_j \). Usually, we would select the variable we know the most about, or we would select the variable that has the largest absolute regression coefficient \( |\hat{\beta}_j| \).
2. The step size in the other variables is
   \[
   \Delta x_i = \frac{\hat{\beta}_i}{\hat{\beta}_j/\Delta x_j} \quad i = 1, 2, \ldots, k \quad i \neq j
   \]
3. Convert the \( \Delta x_i \) from coded variables to the natural variables.
11.3 Analysis of Second-Order Response Surface

- Second-order model in RSM

\[ y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i<j} \beta_{ij} x_i x_j + \epsilon \] (11.4)

- These models are used widely in practice
- The Taylor series analogy
- Fitting the model is easy, some nice designs are available
- Optimization is easy
- There is a lot of empirical evidence that they work very well
**Figure 11.6** Response surface and contour plot illustrating a surface with a maximum
Figure 11.7  Response surface and contour plot illustrating a surface with a minimum
**Figure 11.8** Response surface and contour plot illustrating a saddle point (or minimax)
Characterization of the Response Surface

• Find out where our stationary point is
• Find what type of surface we have
  – Graphical Analysis
  – Canonical Analysis
• Determine the sensitivity of the response variable to the optimum value
  – Canonical Analysis
Finding the Stationary Point

• After fitting a second order model take the partial derivatives with respect to the $x_i$’s and set to zero
  - $\frac{\delta y}{\delta x_1} = \ldots = \frac{\delta y}{\delta x_k} = 0$

• Stationary point represents…
  - Maximum Point
  - Minimum Point
  - Saddle Point
Stationary Point

\[
x_s = -\frac{1}{2} B^{-1} b
\]

\[
x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \quad b = \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_k \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \widehat{\beta}_{11}, \widehat{\beta}_{12}/2, \ldots, \widehat{\beta}_{1k}/2 \\ \widehat{\beta}_{22}, \ldots, \widehat{\beta}_{2k}/2 \\ \vdots \\ \text{sym.} \\ \widehat{\beta}_{kk} \end{bmatrix}
\]
• **Example 11.2**: Continue Example 11.1.
• Augment the design to fit a 2nd-order model.
• The complete design is called a **central composite design (CCD)**

![Table 11.6](image)

<table>
<thead>
<tr>
<th>Natural Variables</th>
<th>Coded Variables</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_1$</td>
<td>$\xi_2$</td>
<td>$x_1$</td>
</tr>
<tr>
<td>80</td>
<td>170</td>
<td>−1</td>
</tr>
<tr>
<td>80</td>
<td>180</td>
<td>−1</td>
</tr>
<tr>
<td>90</td>
<td>170</td>
<td>1</td>
</tr>
<tr>
<td>90</td>
<td>180</td>
<td>1</td>
</tr>
<tr>
<td>85</td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td>85</td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td>85</td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td>85</td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td>85</td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td>92.07</td>
<td>175</td>
<td>1.414</td>
</tr>
<tr>
<td>77.93</td>
<td>175</td>
<td>−1.414</td>
</tr>
<tr>
<td>85</td>
<td>182.07</td>
<td>0</td>
</tr>
<tr>
<td>85</td>
<td>167.93</td>
<td>0</td>
</tr>
</tbody>
</table>
additional responses were of interest: the viscosity and the molecular weight of the product. The responses are also shown in Table 11.6.

We will focus on fitting a quadratic model to the yield response \( y_1 \) (the other responses will be discussed in Section 11.3.4). We generally use computer software to fit a response surface and to construct the contour plots. Table 11.7 contains the output from Design-Expert. From examining this table, we notice that this software package first computes the “sequential or extra sums of squares” for the
Fitted second-order model

```
> g=lm(y~x1+x2+x1*x2+I(x1^2)+I(x2^2)); summary(g)

                      Estimate  Std. Error   t value     Pr(>|t|)
(Intercept)             79.94000   0.11909   671.264 < 2e-16 ***
x1                      0.99505    0.09415    10.568  1.48e-05 ***
x2                      0.51520    0.09415     5.472  0.000934 ***
I(x1^2)                 -1.37645    0.10098   -13.630  2.69e-06 ***
I(x2^2)                 -1.00134    0.10098   -9.916  2.26e-05 ***
x1:x2                   0.25000    0.13315     1.878   0.102519

Residual standard error: 0.2663 on 7 degrees of freedom
Multiple R-Squared: 0.9827, Adjusted R-squared: 0.9704
F-statistic: 79.67 on 5 and 7 DF,  p-value: 5.147e-06
```
\textbf{FIGURE 11.11} Contour and response surface plots of the yield response, Example 11.2
11.4 Experimental Designs for Fitting Response Surfaces

Fitting and analyzing response surfaces is greatly facilitated by the proper choice of an experimental design. In this section, we discuss some aspects of selecting appropriate designs for fitting response surfaces.

When selecting a response surface design, some of the features of a desirable design are as follows:

1. Provides a reasonable distribution of data points (and hence information) throughout the region of interest
2. Allows model adequacy, including lack of fit, to be investigated
3. Allows experiments to be performed in blocks
4. Allows designs of higher order to be built up sequentially
5. Provides an internal estimate of error
6. Provides precise estimates of the model coefficients
7. Provides a good profile of the prediction variance throughout the experimental region
8. Provides reasonable robustness against outliers or missing values
9. Does not require a large number of runs
10. Does not require too many levels of the independent variables
11. Ensures simplicity of calculation of the model parameters
11.4.1 Designs for Fitting the First-Order Model

Suppose we wish to fit the first-order model in \( k \) variables

\[
y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \epsilon
\]  

(11.14)

There is a unique class of designs that minimizes the variance of the regression coefficients \( \hat{\beta}_i \). These are the **orthogonal first-order designs**. A first-order design is orthogonal if the off-diagonal elements of the \((X'X)\) matrix are all zero. This implies that the cross products of the columns of the \( X \) matrix sum to zero.

The class of orthogonal first-order designs includes the \( 2^k \) factorial and fractions of the \( 2^k \) series in which main effects are not aliased with each other. In using these designs, we assume that the low and high levels of the \( k \) factors are coded to the usual \( \pm 1 \) levels.

Addition of center points is usually a good idea
11.4.2 Designs for Fitting the Second-Order Model

We have informally introduced in Example 11.2 (and even earlier, in Example 6.6) the central composite design or CCD for fitting a second-order model. This is the most popular class of designs used for fitting these models. Generally, the CCD consists of a $2^k$ factorial (or fractional factorial of resolution V) with $n_F$ factorial runs, $2k$ axial or star runs, and $n_C$ center runs. Figure 11.20 shows the CCD for $k = 2$ and $k = 3$ factors.

![Diagram of central composite designs for $k = 2$ and $k = 3$]

**Figure 11.20** Central composite designs for $k = 2$ and $k = 3
The Rotatable CCD

\[ \alpha = F^{1/4} \]

**Figure 11.21** Contours of constant standard deviation of predicted response for the rotatable CCD, Example 11.2
Computer-Generated (Optimal) Designs

• These designs are good choices whenever
  – The experimental region is irregular
  – The model isn’t a standard one
  – There are unusual sample size or blocking requirements

• These designs are constructed using a computer algorithm and a specified “optimality criterion”

• Many “standard” designs are either optimal or very nearly optimal
Optimality Criteria

Linear Model: \( y = X\beta + \varepsilon \)

D-optimality: \( \min \ |(X^T X)^{-1}| \)

A-optimality: \( \min \ \text{trace}(X^T X)^{-1} \)

- D-optimal design minimizes the volume of the joint confidence region on the vector of regression coefficients
- A-optimal design minimizes the sum of the variances of the regression coefficients
G- and I-optimality

G – optimality: \( \min \max_{x \in R} V(\hat{y}(x)) \)

I – optimality: \( \min \frac{1}{A} \int_{R} V(\hat{y}(x)) \, dx \)

• G-optimal design minimizes the maximum prediction variance over design region \( R \)
• I-optimal design minimizes the average or integrated variance over design space \( R \)
Which Criterion Should I Use?

• For fitting a first-order model, $D$ is a good choice
  – Focus on estimating parameters
  – Useful in screening
• For fitting a second-order model, $G$ or $I$ is a good choice
  – Focus on response prediction
  – Appropriate for optimization
• The $2^k$ design is D, A, G, I-optimal for fitting the first-order model in $k$ variables (w. or w/o) interactions
Algorithms

• Point exchange
  – Requires a candidate set of points
  – The design is chosen from the candidate set
  – Random start, several (many) restarts to ensure that a highly efficient design is found

• Coordinate exchange
  – No candidate set required
  – Search over each coordinate one-at-a-time
  – Many random starts used