清华大学统计学辅修课程

Design and Analysis of Experiments

Lecture 10 –
Response Surface Methods & Designs

周在莹
清华大学统计学研究中心
http://www.stat.tsinghua.edu.cn

Outline

- Response Surface Methodology
  - Three Basic Steps
  - Sequential Procedure
- Three Models
  - Screening Response Model
  - Steepest Ascent Model
  - Optimization Model
- Multiple Responses
- Designs for Fitting Response Surfaces
  - Central Composite Design(CCD,中心复合设计)
  - Box-Behnken Design

Overview of Response Surface Methods

- The primary focus of previous lessons was factor screening
  - Two-level factorials, fractional factorials being widely used
- The objective of Response Surface Methods (RSM) is optimization, finding the best set of factor levels to achieve some goal
  - optimize an underlying process
  - look for the factor level combinations that give us the maximum yield and minimum costs
  - hit a target or aim to match some given specifications
- RSM dates from the 1950's. Early applications were found in the chemical industry
- Modern applications of RSM span many industrial and business settings

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**
  - Discover a proper region to carry out experiment
  - Find the optimal combination of factors
  - Use a small number of experiments
- Challenges
  - The response surface can be high dimensional
  - The shape of the surface is unknown
  - The target region of factors is unknown

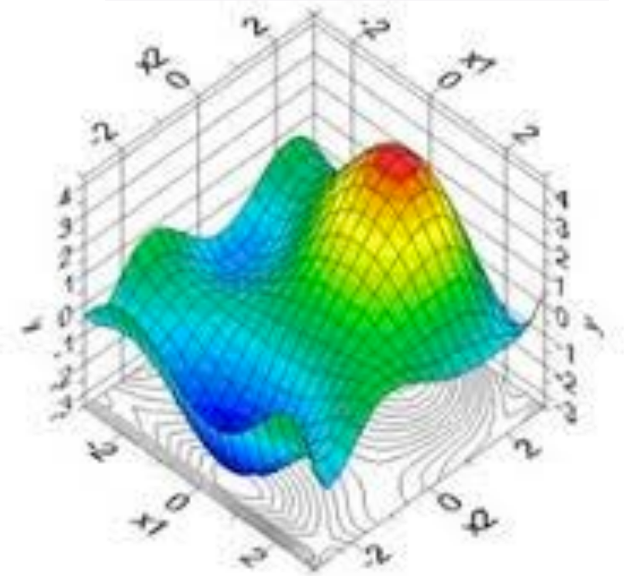
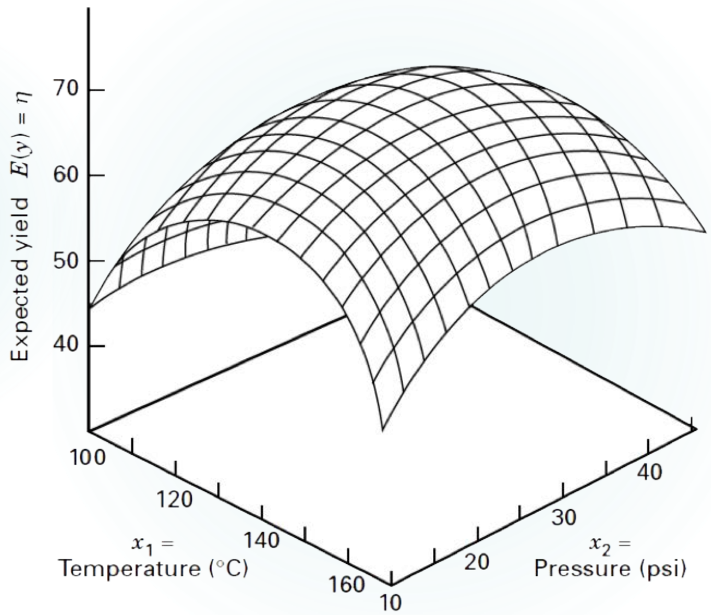


Illustration of a RS

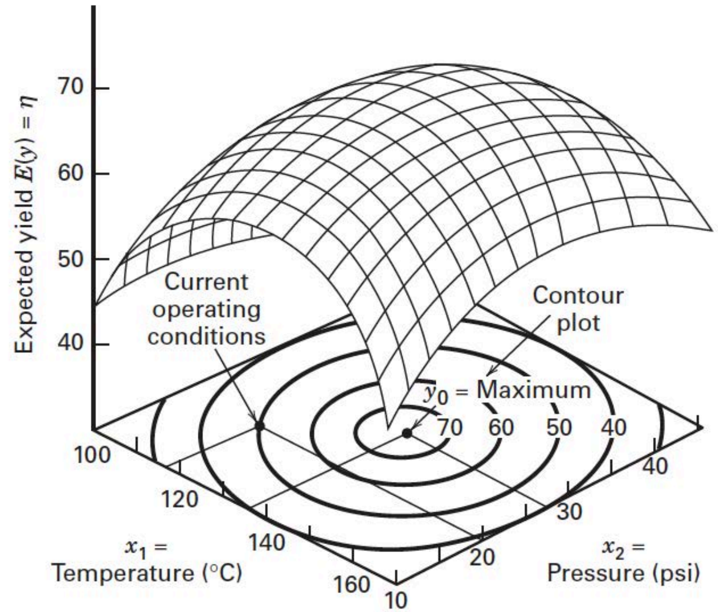
$$y = f(x_1, x_2) + \varepsilon$$

where ε represents the noise or error observed in the response y

- The surface $\mathrm{E}(y) \triangleq \eta = f(x_1, x_2)$ is called a response surface



A three-dimensional response surface showing the expected yield (η) as a function of temperature (x_1) and pressure (x_2)



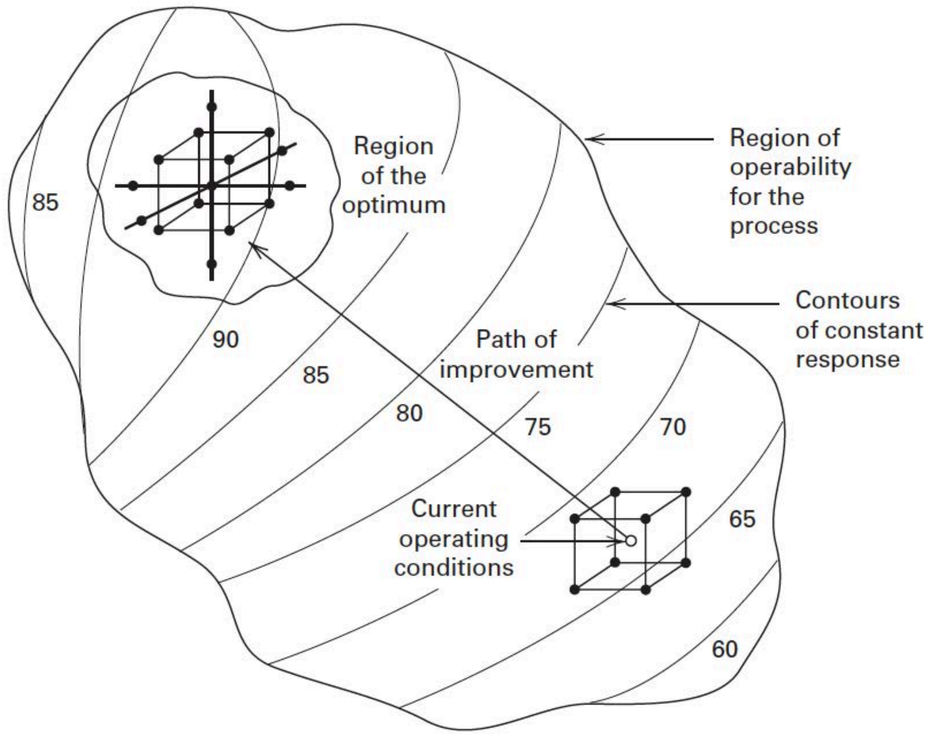
A contour plot of a response surface

Three Basic Steps

- Factor screening - Find $x_1, \dots, x_k$
  - Start with a large number of factors
  - Select a few (≤5) important factors for response surface
- A series of 1st order experiments - Find a suitable approximation for $y = f(x_1, \dots, x_k)$
  - Start from an initial configuration of the few selected factors
  - Move towards the region of the optimal configuration
- A 2nd order experiment - When curvature is found find a new approximation
  - An additional experiment in the neighborhood of the optimal configuration
  - Perform the "Response Surface Analysis"
  - Help to find the optimal configuration

RSM Is a Sequential Procedure

- Sequential exploration of Response Surface
  - Factor screening
  - Finding the region of the optimum
  - Modeling & Optimization of the response



Models Available

- **Screening Response Model**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

The single cross product factor represents the linear × linear interaction component

- **Steepest Ascent Model**

Ignore cross products which gives an indication of the curvature of the response surface

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

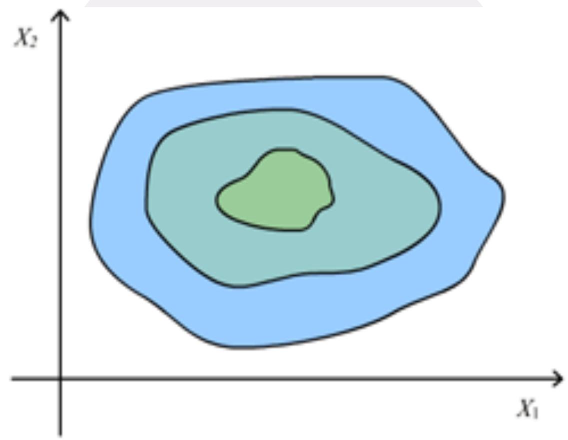
- **Optimization Model**

When we think that we are somewhere near the 'top of the hill' we will fit a second order model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

RSM for 2 Factors

- Look at 2 dimensions - easier to think about and visualize
- Imagine the ideal case where there is actually a 'hill' which has a nice centered peak
- Our quest, to find the values $X_1^{optimum}$ and $X_2^{optimum}$, where the response is at its peak
- We might have a hunch that the optimum exists in certain location. This would be good area to start - some set of conditions
- Take natural units and then center and rescale them to the range from -1 to +1



'Climbing a hill'
or
'Descending into a valley'

Steepest Ascent - The First Order Model

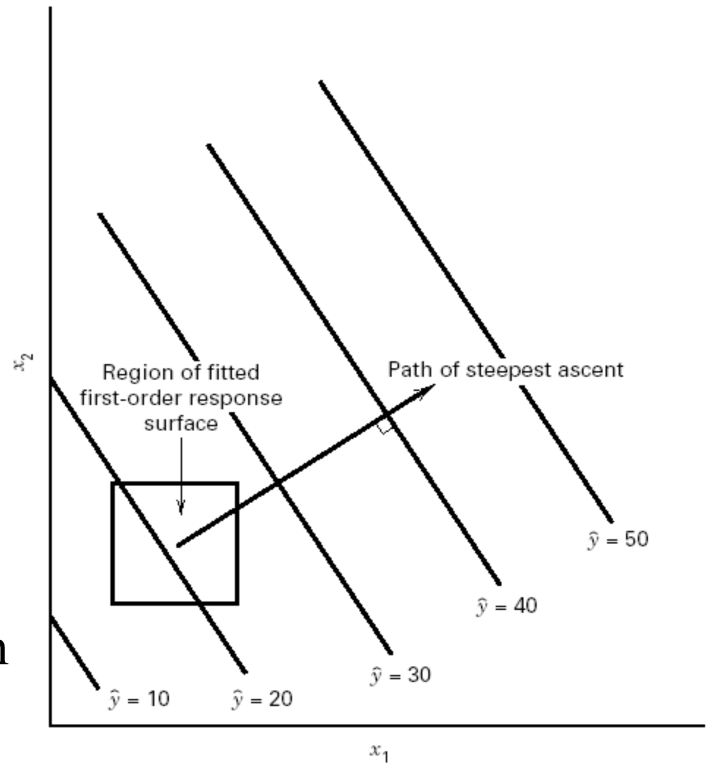
- When we are remote from the optimum, we usually assume that a first-order model is an adequate approximation to the true surface in a small region of the x's
- A procedure for moving sequentially from an initial "guess" towards to region of the optimum
- The 1st order Taylor expansion

$$f(x_1, \dots, x_k) \approx \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

- Steepest ascent is a gradient procedure

$$\frac{\partial f}{\partial x_j} = \beta_j, j = 1, \dots, k$$

- The steps along the path are proportional to the regression coefficients $\{\beta_j\}$



Note on the Steepest Ascent

- Q: Why is gradient the direction of steepest ascent?

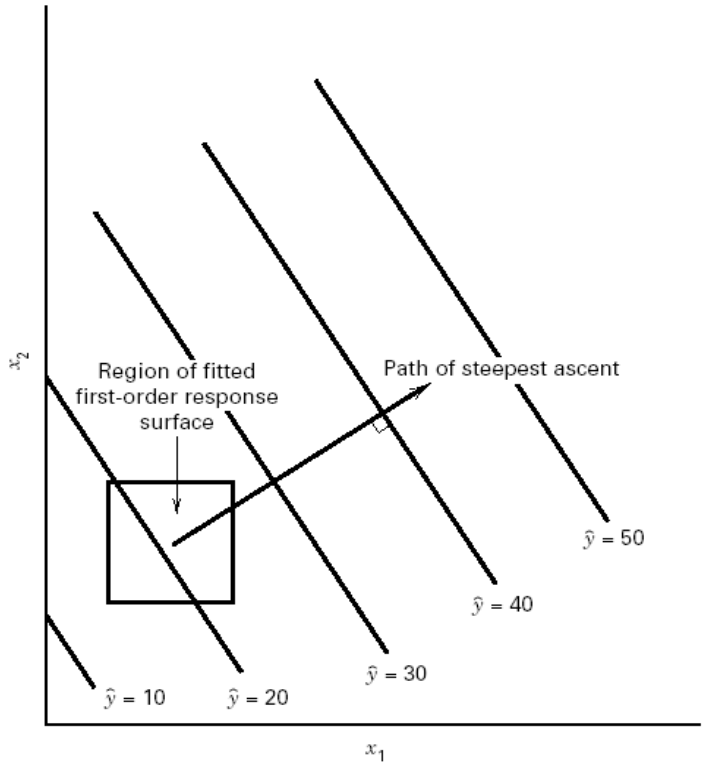
- A: For any arbitrary direction v, the rate of change along v is

$$\lim_{h \to 0} \frac{f(x + hv) - f(x)}{h} \approx \nabla f(x) \cdot v$$

  - We know from linear algebra that the dot product is maximized when the two vectors point in the same direction. This means that the rate of change along an arbitrary vector v is maximized when v points in the same direction as the gradient. In other words, the gradient corresponds to the rate of steepest ascent/descent

Steepest Ascent: Procedure

- Experiments are conducted along the path of steepest ascent until no further increase in response is observed
- Then a new first-order model may be fit, a new path of steepest ascent determined, and the procedure continued
- Eventually, the experimenter will arrive in the vicinity of the optimum. This is usually indicated by lack of fit of a first-order model
- At that time, additional experiments will be conducted to obtain a more precise estimate of the optimum



Steepest Ascent: Chemical Yield Example

- To maximize the yield of a chemical process
- Two controllable variables: reaction time(A) and reaction temperature(B)
- The region center: (35min, 155°F)
- It is unlikely that this region contains the optimum, so there is little curvature in the system and the first-order model will be appropriate, followed by the method of steepest ascent
- Now the fitted first-order model is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$
$$= 40.44 + 0.775 x_1 + 0.325 x_2$$

Natural Variables		Coded Variables		Response
ξ_1	ξ_2	x_1	x_2	y
30	150	−1	−1	39.3
30	160	−1	1	40.0
40	150	1	−1	40.9
40	160	1	1	41.5
35	155	0	0	40.3
35	155	0	0	40.5
35	155	0	0	40.7
35	155	0	0	40.2
35	155	0	0	40.6

Check the Adequacy of the First-Order Model

- Before exploring along the path of steepest ascent, the adequacy of the first-order model should be investigated
- The 2^2 design with center points allows:
  - *1.* Obtain an estimate of error
    - Use the replicates at the center
  - *2.* Check for interactions (cross-product terms) in the model
    - $F = \frac{SS_{Interaction}}{Pure\ error}$
  - *3.* Check for quadratic effects (curvature)
    - Compare the average response at the four points in the factorial with the average response at the center; the difference is a measure of curvature

Notes on Coef and SS

$$SS_C = \frac{\left(\sum_{i=1}^{a} c_i \bar{y}_{i.}\right)^2}{\frac{1}{n}\sum_{i=1}^{a} c_i^2}$$

> chem

	A	B	y	x1	x2
1	30	150	39.3	-1	-1
2	30	160	40.0	-1	1
3	40	150	40.9	1	-1
4	40	160	41.5	1	1
5	35	155	40.3	0	0
6	35	155	40.5	0	0
7	35	155	40.7	0	0
8	35	155	40.2	0	0
9	35	155	40.6	0	0

- $\hat{\beta}_{12} = \frac{1}{4}(1\times 39.3 + 1\times 41.5 - 1\times 40.0 - 1\times 40.9)$
- $SS_{Interaction} = \frac{(1\times 39.3+1\times 41.5-1\times 40.0-1\times 40.9)^2}{4} = \hat{\beta}_{12}^2 S_{1212}$
- β_{11} and β_{22} are the coefficients of the "pure quadratic" terms ${x_1}^2$ and ${x_2}^2$
  - summary(lm(y ~ x1 + x2 + I(x1^2) + I(x2^2), chem))
- $\bar{y}_F - \bar{y}_C$ is an estimate of $\beta_{11} + \beta_{22}$
- $SS_{Pure\ Quadratic} = \frac{(\bar{y}_F - \bar{y}_C)^2}{\frac{1}{n_F}+\frac{1}{n_C}}$

$$SS_{AB} = \frac{\left(\sum_{i=1}^{4} c_i \bar{y}_{i.}\right)^2}{4/n} = \frac{(ab + (1) - a - b)^2}{4n}$$

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	40.46000	0.08355	484.282	7.13e-13	***
x1	0.77500	0.09341	8.297	0.000415	***
x2	0.32500	0.09341	3.479	0.017671	*
I(x1^2)	-0.03500	0.12532	-0.279	0.791209	
I(x2^2)	NA	NA	NA	NA	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1868 on 5 degrees of freedom
Multiple R-squared: 0.9419, Adjusted R-squared: 0.907
F-statistic: 27.01 on 3 and 5 DF, p-value: 0.001624

Analysis for the First-Order Model

```
> full <- lm(y ~ x1 + x2 + I(x1*x2) + I(x1^2), chem)
> summary(full)
> anova(full)
```

```
Analysis of Variance Table

Response: y
            Df  Sum Sq Mean Sq F value   Pr(>F)
x1           1 2.40250 2.40250 55.8721 0.001713 **
x2           1 0.42250 0.42250  9.8256 0.035030 *
I(x1 * x2)   1 0.00250 0.00250  0.0581 0.821316
I(x1^2)      1 0.00272 0.00272  0.0633 0.813741
Residuals    4 0.17200 0.04300
```

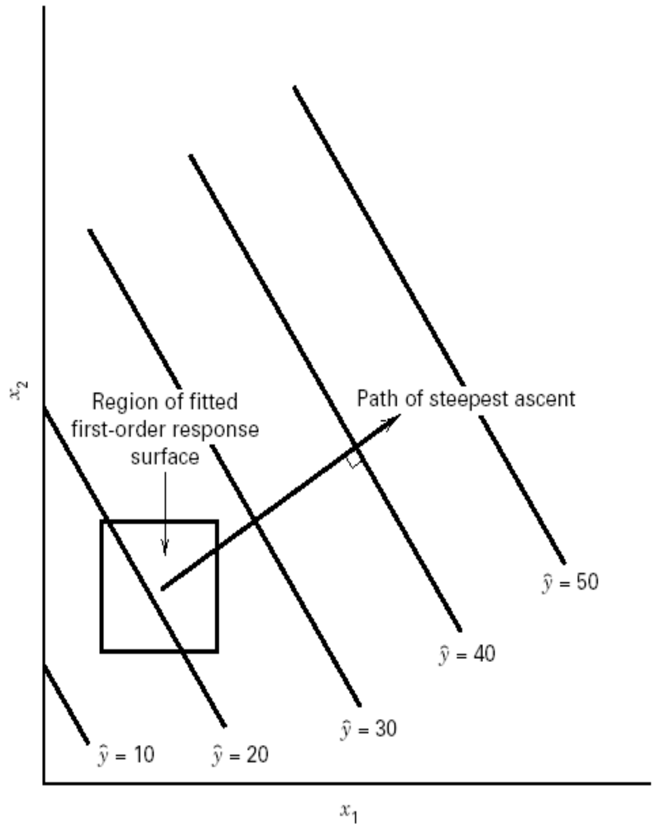
```
Coefficients:
             Estimate  Std. Error t value Pr(>|t|)
(Intercept) 40.46000    0.09274 436.291 1.66e-10 ***
x1           0.77500    0.10368   7.475  0.00171 **
x2           0.32500    0.10368   3.135  0.03503 *
I(x1 * x2)  -0.02500    0.10368  -0.241  0.82132
I(x1^2)     -0.03500    0.13910  -0.252  0.81374

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: 0.009471
```

- Both the interaction and curvature checks are not significant, whereas the *F*-test for the overall regression is significant
- Both regression coefficients are large relative to their standard errors

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	*P*-Value
Model (β_1, β_2)	2.8250	2	1.4125	47.83	0.0002
Residual	0.1772	6			
(Interaction)	(0.0025)	1	0.0025	0.058	0.8215
(Pure quadratic)	(0.0027)	1	0.0027	0.063	0.8142
(Pure error)	(0.1720)	4	0.0430		
Total	3.0022	8			

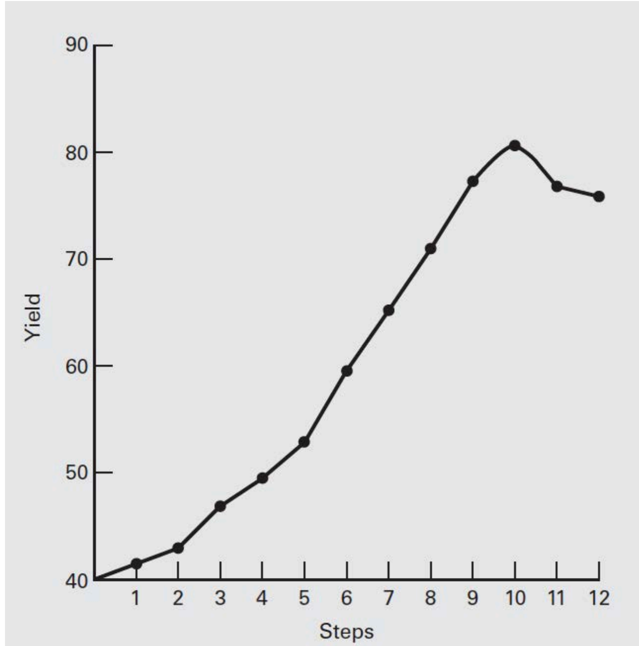
Decide the Direction

- $\hat{y} = 40.44 + 0.775x_1 + 0.325x_2$
- To move away from the design center ($x_1 = x_2 = 0$) 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 ($x_1 = x_2 = 0$) and has a slope 0.325/0.775
- Use 5 minutes of reaction time as the basic step size, that is $\Delta x_1 = 1$ in the coded variable x_1
- Therefore, the steps along the path of steepest ascent are $\Delta x1 = 1.0000$ and $\Delta x_2 = (0.325/0.775) = 0.42$



Get Moving

Steps	Coded Variables x_1	Coded Variables x_2	Natural Variables ξ_1	Natural Variables ξ_2	Response y
Origin	0	0	35	155	
Δ	1.00	0.42	5	2	
Origin + Δ	1.00	0.42	40	157	41.0
Origin + 2Δ	2.00	0.84	45	159	42.9
Origin + 3Δ	3.00	1.26	50	161	47.1
Origin + 4Δ	4.00	1.68	55	163	49.7
Origin + 5Δ	5.00	2.10	60	165	53.8
Origin + 6Δ	6.00	2.52	65	167	59.9
Origin + 7Δ	7.00	2.94	70	169	65.0
Origin + 8Δ	8.00	3.36	75	171	70.4
Origin + 9Δ	9.00	3.78	80	173	77.6
Origin + 10Δ	10.00	4.20	85	175	80.3
Origin + 11Δ	11.00	4.62	90	179	76.2
Origin + 12Δ	12.00	5.04	95	181	75.1



- 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 $x_1 = (\xi_1 - 85)/5$, $x_2 = (\xi_2 - 175)/5$

Second First-Order Model

- Once again, a 2^2 design with five center points is used

Natural Variables		Coded Variables		Response
ξ_1	ξ_2	x_1	x_2	y
80	170	−1	−1	76.5
80	180	−1	1	77.0
90	170	1	−1	78.0
90	180	1	1	79.5
85	175	0	0	79.9
85	175	0	0	80.3
85	175	0	0	80.0
85	175	0	0	79.7
85	175	0	0	79.8

```
Coefficients:
             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 * x2) 0.2500    0.1151   2.172 0.095611 .
I(x1^2)    -2.1900    0.1544 -14.181 0.000144 ***
```

```
Analysis of Variance Table

Response: y
           Df Sum Sq Mean Sq F value    Pr(>F)
x1          1 4.000   4.000  75.472  0.0009664 ***
x2          1 1.000   1.000  18.868  0.0122172 *
I(x1 * x2) 1 0.250   0.250    4.717  0.0956108 .
I(x1^2)     1 10.658 10.658 201.094 0.0001436 ***
Residuals 4  0.212   0.053
```

- The interaction and pure quadratic checks imply that the first-order model is not an adequate approximation
- This curvature in the true surface may indicate that we are near the optimum
- At this point, additional analysis must be done to locate the optimum more precisely

Path for Multiple Predictors

- 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 Δx_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 $\Delta x_i = \dfrac{\hat{\beta}_i}{\hat{\beta}_j / \Delta x_j}, i \neq j$
3. Convert the Δx_i from coded variables to the natural variables

Second-Order Models in RSM

- $y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \sum_{i<j}^{k} \beta_{ij} x_i x_j + \varepsilon$

- 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

2nd Order Approximation

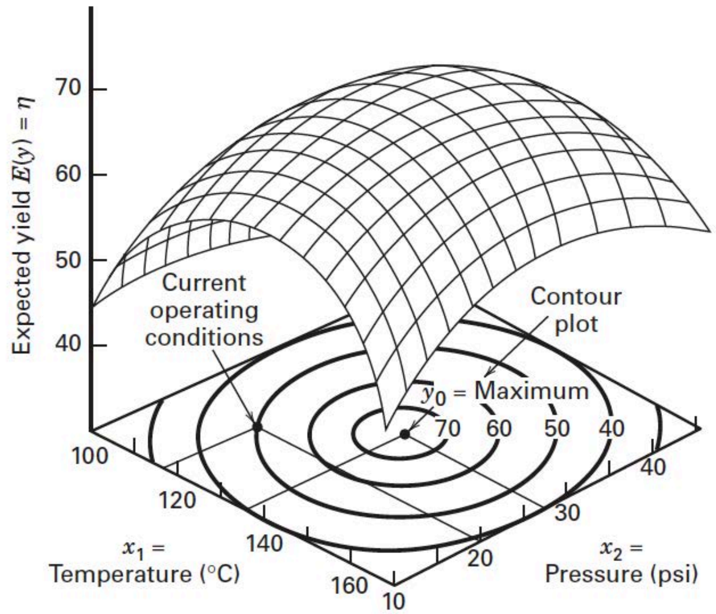
- The 2nd order Taylor expansion

$$f(x_1, \dots, x_k) \approx \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \sum_{i<j}^{k} \beta_{ij} x_i x_j$$

- Total curvature: $\sum_{j=1}^{k} \beta_{jj}$
- Matrix form:

$$b = (\hat{\beta}_1, \dots, \hat{\beta}_k)^T, \; B = \begin{bmatrix} \hat{\beta}_{11} & \frac{1}{2}\hat{\beta}_{21} & \cdots & \frac{1}{2}\hat{\beta}_{1k} \\ \frac{1}{2}\hat{\beta}_{12} & \hat{\beta}_{22} & & \frac{1}{2}\hat{\beta}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}\hat{\beta}_{1k} & \frac{1}{2}\hat{\beta}_{2k} & \cdots & \hat{\beta}_{kk} \end{bmatrix}$$

$$\hat{y}(x) = \hat{\beta}_0 + x^T b + x^T B x$$



2nd Order Approximation

- The 2nd order Taylor expansion

$$\hat{y}(x) = \hat{\beta}_0 + x^T b + x^T B x$$

- SVD of B

$$B = P\Lambda P^T,\ PP^T = P^T P = I, \Lambda = diag\{\lambda_1, \dots, \lambda_k\}$$

- A more convenient form- canonical form of the model

$$\hat{y}(x) = \hat{\beta}_0 + w^T P b + w^T \Lambda w = \hat{\beta}_0 + w^T b^* + \sum_{j=1}^{k} \lambda_j w_j^2$$

where $w = P^T x,\ b^* = Pb = (b_1^*, \dots, b_k^*)^T$

4 Possible Scenarios

- The 2nd order approximation

  $\hat{y}(x) = \hat{\beta}_0 + w^T P b + w^T \Lambda w = \hat{\beta}_0 + w^T b^* + \sum_{j=1}^{k} \lambda_j w_j^2$

- Possible scenarios

  a) Elliptic system: λ_j>0 or <0 for all j

  b) Hyperbolic system: some λ_j>0, some λ_j<0

  c) Stationary ridge system: some $\lambda_j \approx 0$, and the experiment region is close to the center

  d) Rising/falling ridge system: some $\lambda_j \approx 0$, and the experiment region is far away from the center

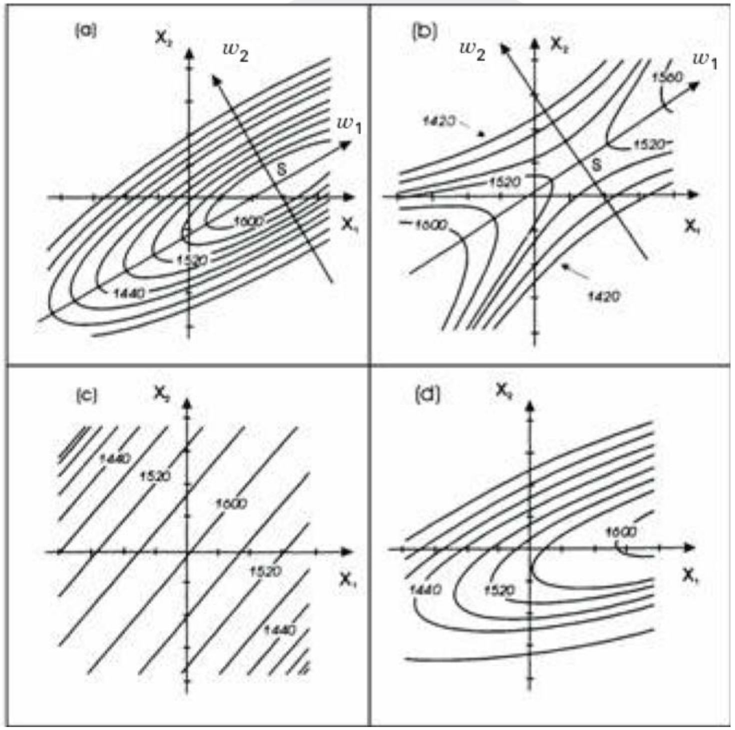
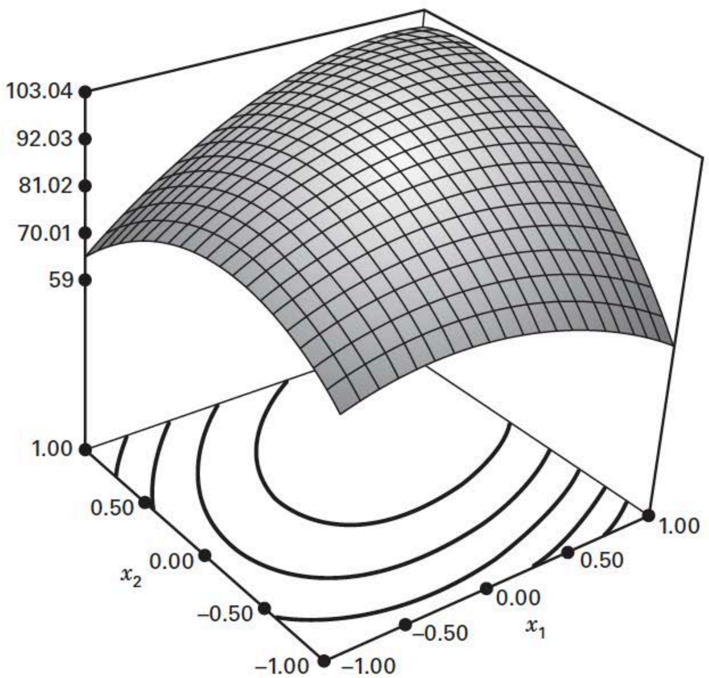
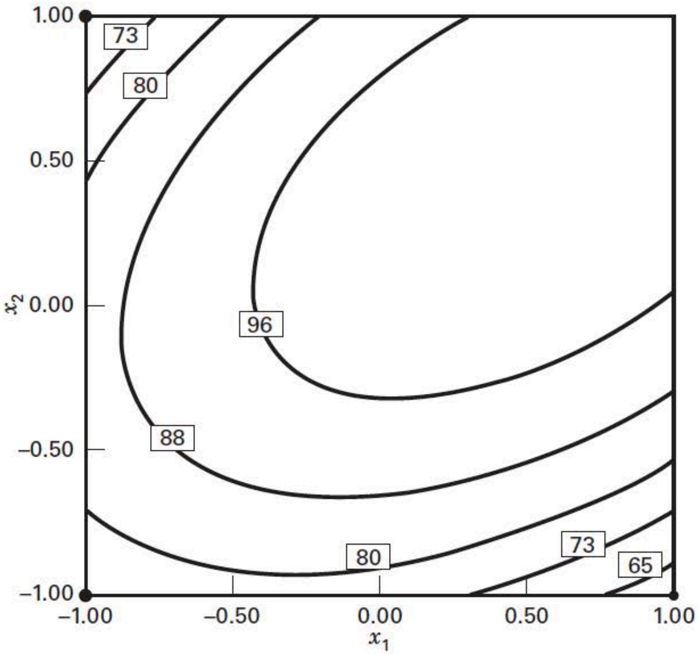


Illustration of a Surface with a Maximum



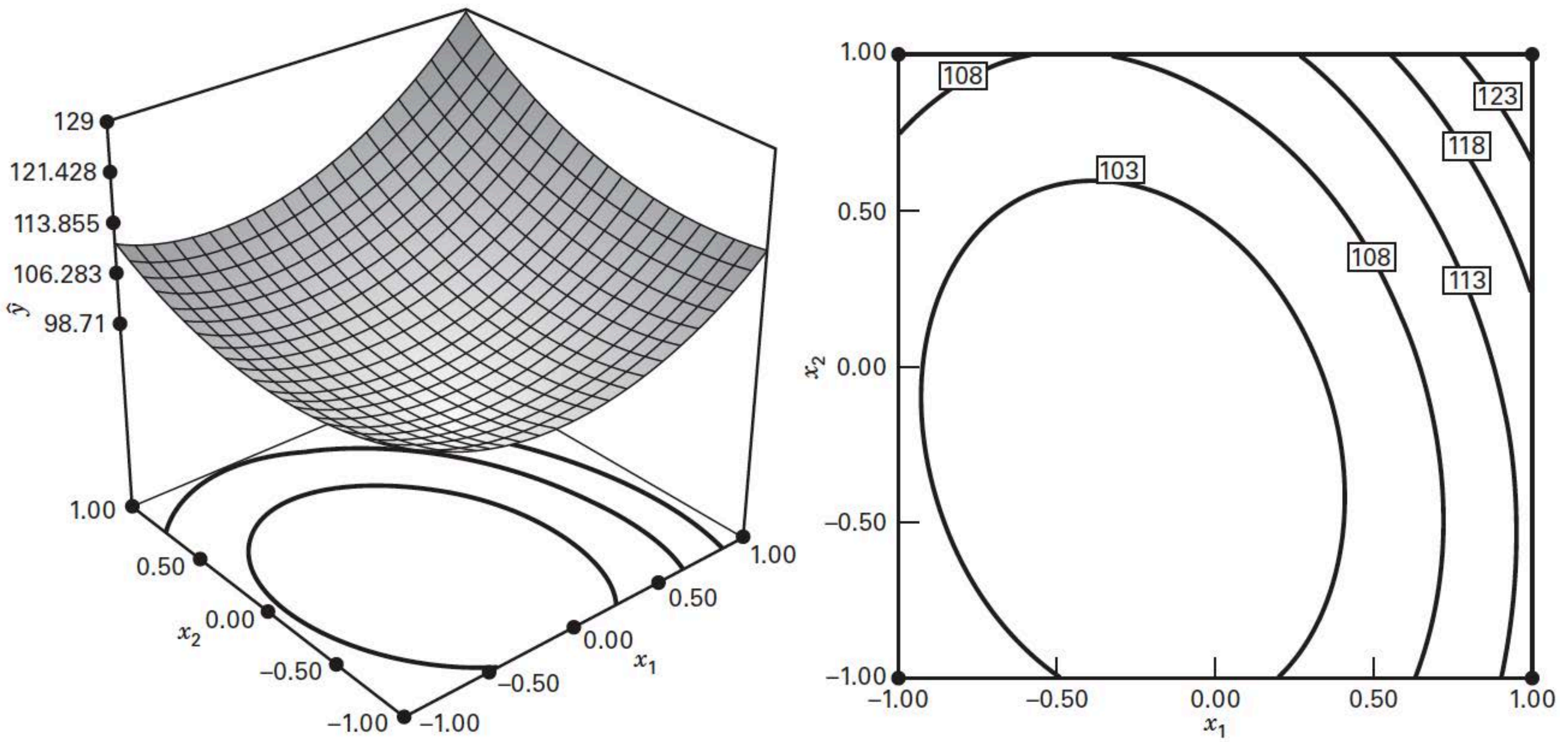
(a) Response surface



(b) Contour plot

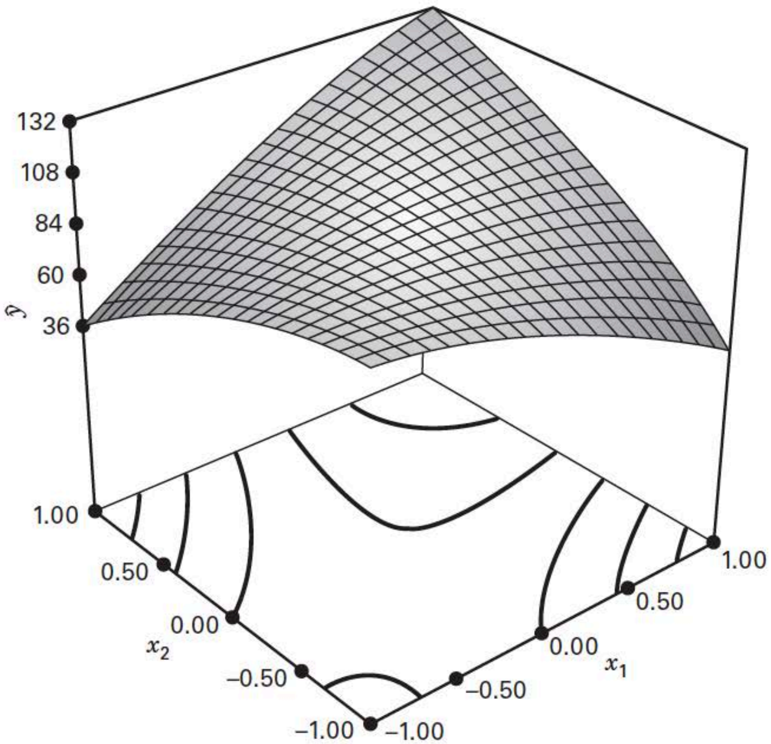
清华大学统计学研究中心

Illustration of a Surface with a Minimum

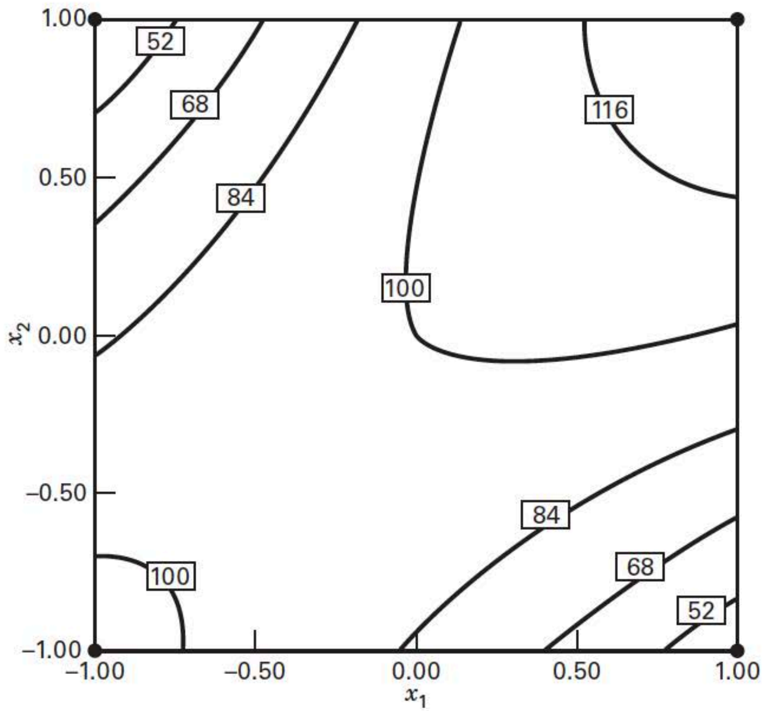


(*a*) Response surface (*b*) Contour plot

Illustrating of a Saddle Point (or Minimax)

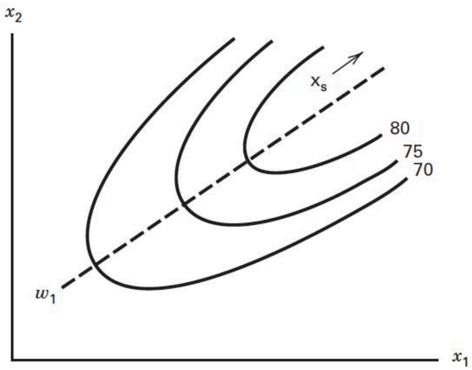


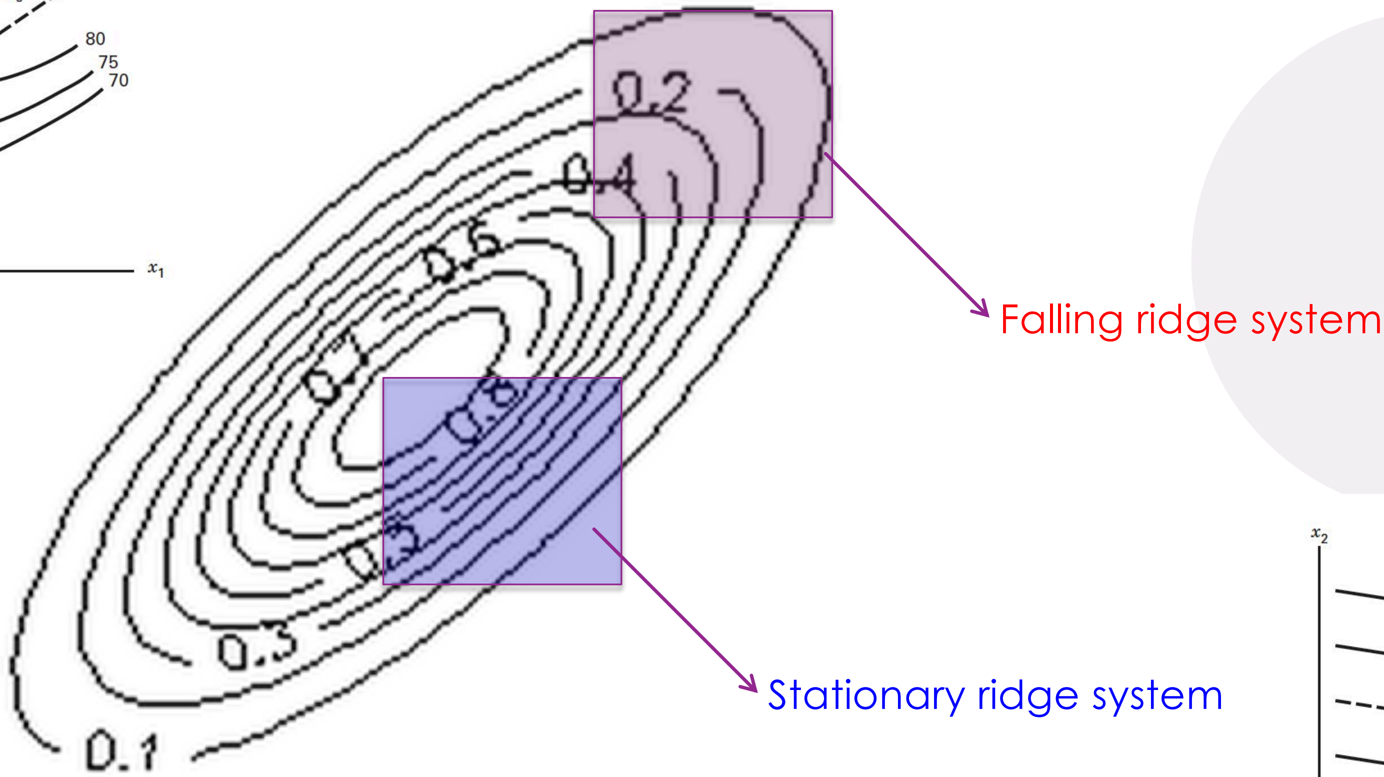
(*a*) Response surface

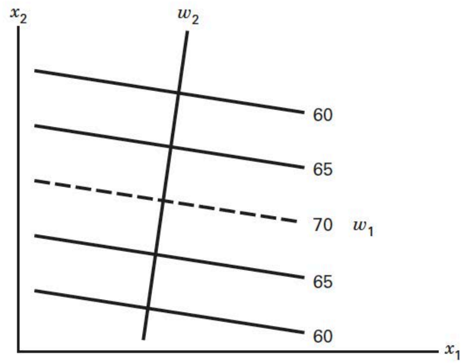


(*b*) Contour plot

A Graphical Illustration of Ridge Systems(岭系统)







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

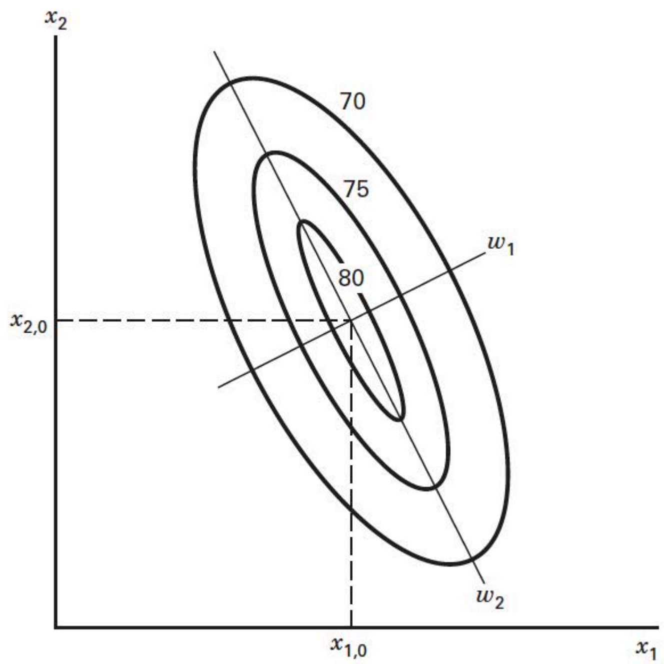
- After fitting a second order model take the partial derivatives with respect to the x_i's and set to zero

$$\delta y / \delta x_1 = \ldots = \delta y / \delta x_k = 0$$

- Stationary point represents
  - Maximum Point
  - Minimum Point
  - Saddle Point
- $\hat{y}(x) = \hat{\beta}_0 + x^T b + x^T B x$
  - $x_s = -\frac{1}{2} B^{-1} b$
  - $\hat{y}_s = \hat{\beta}_0 + \frac{1}{2} x_s^T b$

Canonical Analysis

- Used for sensitivity analysis and stationary point identification
- First to transform the model into a new coordinate system with the origin at the stationary point x_s and then to rotate the axes of this system until they are parallel to the principal axes of the fitted response surface



- Based on the analysis of the transformed model

$$\hat{y}(x) = y_s + \sum_{j=1}^{k} \lambda_j w_j^2$$

- Canonical model

Eigenvalues

- The nature of the response can be determined by the signs and magnitudes of the eigenvalues
  - {e} all positive: a minimum is found
  - {e} all negative: a maximum is found
  - {e} mixed: a saddle point is found
- Eigenvalues can be used to determine the sensitivity of the response with respect to the design factors
- The response surface is steepest in the direction (canonical) corresponding to the largest absolute eigenvalue

Complete Experiment for the Example

- Continue the analysis of the chemical process
- Augment the design with enough points to fit a second-order model

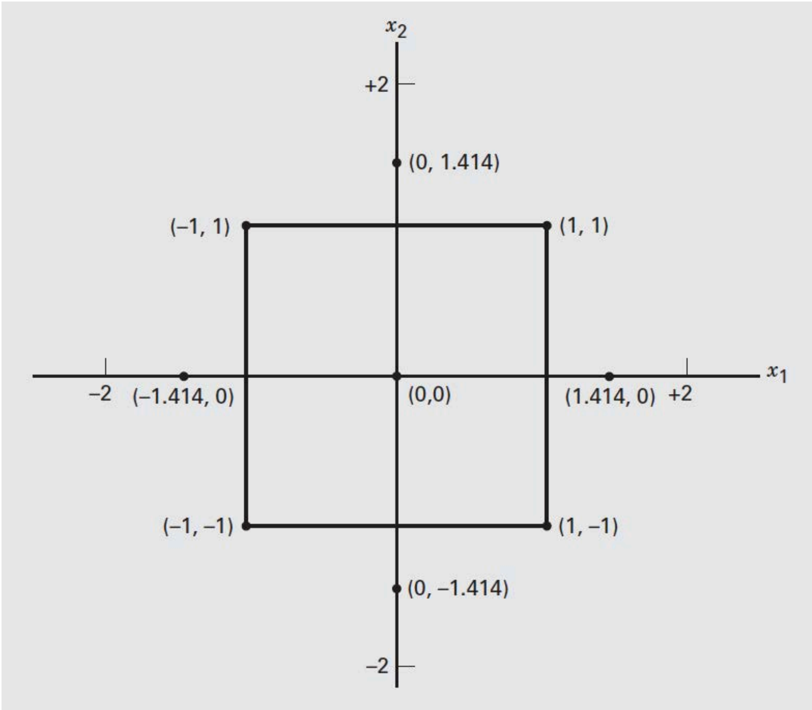
Natural Variables		Coded Variables		Responses		
ξ_1	ξ_2	x_1	x_2	y_1 (Yield)	y_2 (Viscosity)	y_3 (Molecular Weight)
80	170	−1	−1	76.5	62	2940
80	180	−1	1	77.0	60	3470
90	170	1	−1	78.0	66	3680
90	180	1	1	79.5	59	3890
85	175	0	0	79.9	72	3480
85	175	0	0	80.3	69	3200
85	175	0	0	80.0	68	3410
85	175	0	0	79.7	70	3290
85	175	0	0	79.8	71	3500
92.07	175	1.414	0	78.4	68	3360
77.93	175	−1.414	0	75.6	71	3020
85	182.07	0	1.414	78.5	58	3630
85	167.93	0	−1.414	77.0	57	3150

In this second phase of the study, two additional responses were of interest: the viscosity(粘度) and the molecular weight (分子量)of the product

Central Composite Design (CCD)

- Focus on fitting a quadratic model to the yield response y_1

```
> library(rsm)
> rs <- rsm(y1 ~ SO(x1, x2), chem)
> summary(rs)
```



Results

Call:
rsm(formula = y1 ~ SO(x1, x2), data = chem)

```
            Estimate Std. Error  t value  Pr(>|t|)
(Intercept) 79.939955   0.119089 671.2644 < 2.2e-16 ***
x1           0.995050   0.094155  10.5682 1.484e-05 ***
x2           0.515203   0.094155   5.4719  0.000934 ***
x1:x2        0.250000   0.133145   1.8777  0.102519
x1^2        -1.376449   0.100984 -13.6303 2.693e-06 ***
x2^2        -1.001336   0.100984  -9.9158 2.262e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.9827,	Adjusted R-squared:  0.9704
F-statistic: 79.67 on 5 and 7 DF,  p-value: 5.147e-06
```

```
Analysis of Variance Table

Response: y1
            Df  Sum Sq Mean Sq  F value    Pr(>F)
FO(x1, x2)   2 10.0430  5.0215  70.8143 2.267e-05
TWI(x1, x2)  1  0.2500  0.2500   3.5256    0.1025
PQ(x1, x2)   2 17.9537  8.9769 126.5944 3.194e-06
Residuals    7  0.4964  0.0709
Lack of fit  3  0.2844  0.0948   1.7885    0.2886
Pure error   4  0.2120  0.0530

Stationary point of response surface:
       x1        x2
0.3892304 0.3058466
```

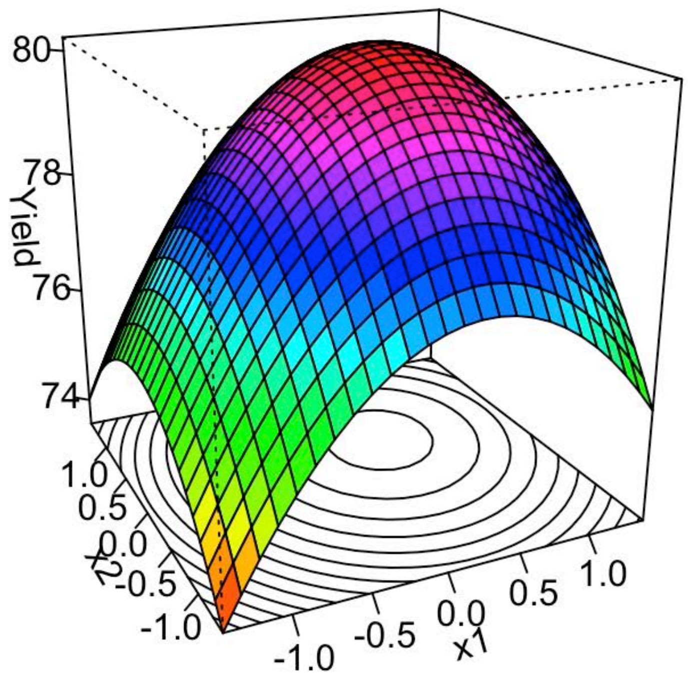
```
Eigenanalysis:
eigen() decomposition
$values
[1] -0.9634986 -1.4142867

$vectors
         [,1]       [,2]
x1 -0.2897174 -0.9571122
x2 -0.9571122  0.2897174
```

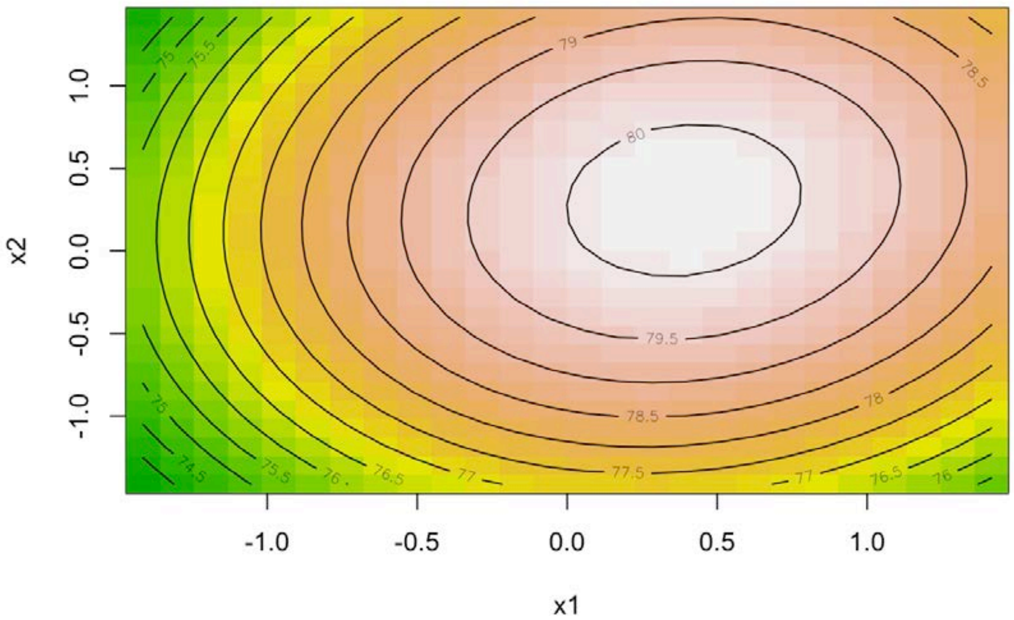
Because both λ_1 and λ_2 are negative and the stationary point is within the region of exploration, we conclude that the stationary point is a maximum

- $x_s = (0.3892304, 0.3058466)$, $\xi 1 = 86.95 \simeq 87$ minutes of reaction time and $\xi 2 = 176.53 \simeq 176.5$°F; $\hat{y}_s = \hat{\beta}_0 + \frac{1}{2} x_s^T b = 80.5$
- The canonical form of the fitted model is $\hat{y}(x) = 80.5 - 0.963 w_1^2 - 1.414 w_2^2$

Response Surface & Contour Plots

- contour(rs,~x1+x2, image = T)
- persp(rs,~x1+x2, col = rainbow(50), zlab='Yield', contours = list(z='bottom'))





Multiple Responses

- Simultaneous consideration of multiple responses involves first building an appropriate response surface model for each response and then trying to find a set of operating conditions that in some sense optimizes all responses or at least keeps them in desired ranges
- We may obtain models for the viscosity and molecular weight responses (y_2 and y_3, respectively) in chemical Example as follows:
  - $\hat{y}_2(x) = 70 - 0.16x_1 - 0.95x_2 - 0.69x_1^2 - 6.69x_2^2 - 1.25x_1x_2$
  - $\hat{y}_3(x) = 3386.2 + 205.1x_1 + 177.4x_2$
- In terms of the natural levels of time (ξ_1) and temperature (ξ_2), these models are
  - $\hat{y}_2(x) = -9030.74 + 13.393\xi_1 + 97.708\xi_2 - 2.75\times10^{-2}\xi_1^2 - 0.26757\xi_2^2 - 5\times10^{-2}\xi_1\xi_2$
  - $\hat{y}_3(x) = -6308.8 + 41.025\xi_1 + 35.473\xi_2$

Note on $\hat{y}_2(x)$ and $\hat{y}_3(x)$

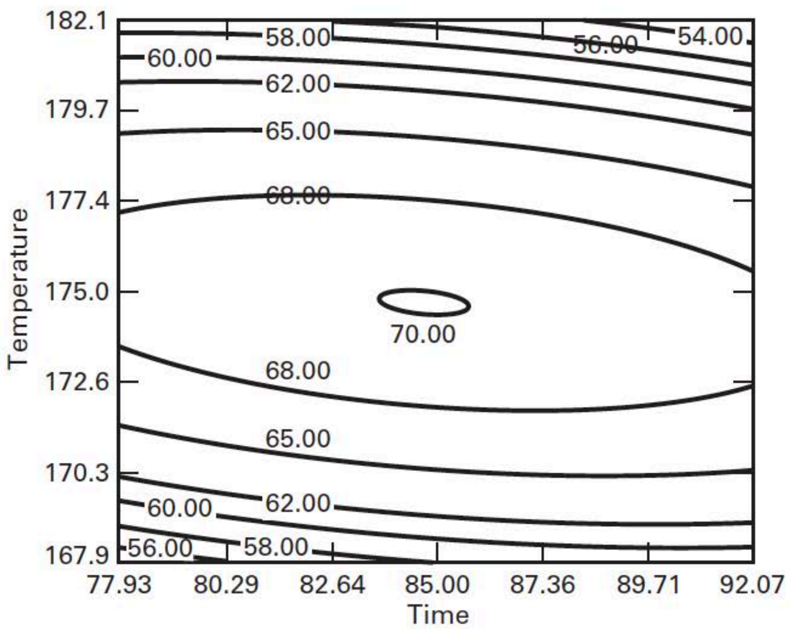
```
> summary( rsm(v ~ SO(x1, x2), chem) ) #model for viscosity
> summary( rsm(m ~ SO(x1, x2), chem)) #model for molecular weight
> summary( rsm(m ~ FO(x1, x2), chem) ) #model for molecular weight
```

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	70.00021	1.01731	68.8093	3.6e-11	***
x1	-0.15527	0.80431	-0.1931	0.8524009	
x2	-0.94839	0.80431	-1.1791	0.2768648	
x1:x2	-1.25000	1.13739	-1.0990	0.3081185	
x1^2	-0.68732	0.86265	-0.7968	0.4517659	
x2^2	-6.68913	0.86265	-7.7542	0.0001112	***

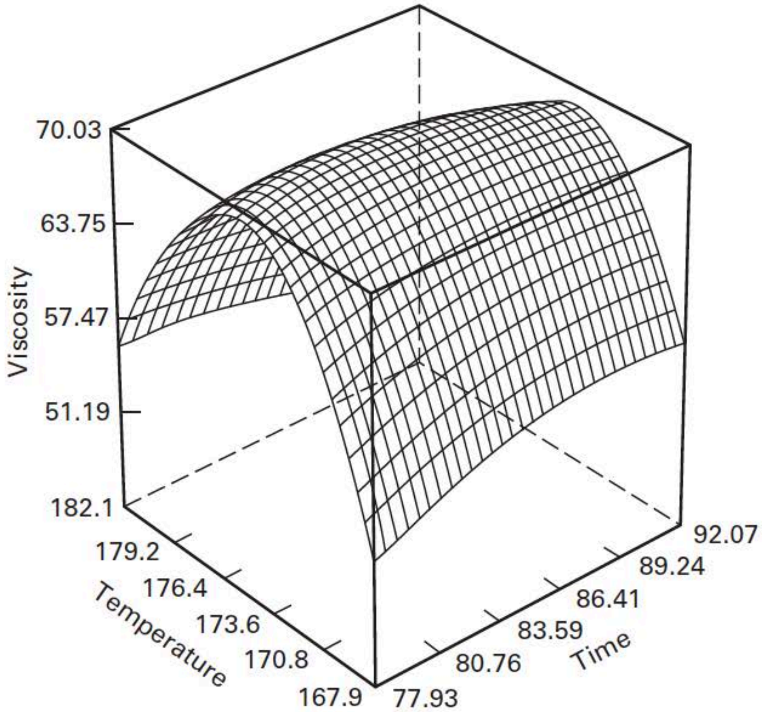
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	3375.975	77.066	43.8064	8.435e-10	***
x1	205.126	60.930	3.3666	0.01198	*
x2	177.367	60.930	2.9110	0.02263	*
x1:x2	-80.000	86.162	-0.9285	0.38406	
x1^2	-41.744	65.350	-0.6388	0.54330	
x2^2	58.286	65.350	0.8919	0.40206	

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	3386.154	45.936	73.7151	5.151e-15	***
x1	205.126	58.561	3.5028	0.0057	**
x2	177.367	58.561	3.0287	0.0127	*

Contour Plot and Response Surface Plot of Viscosity

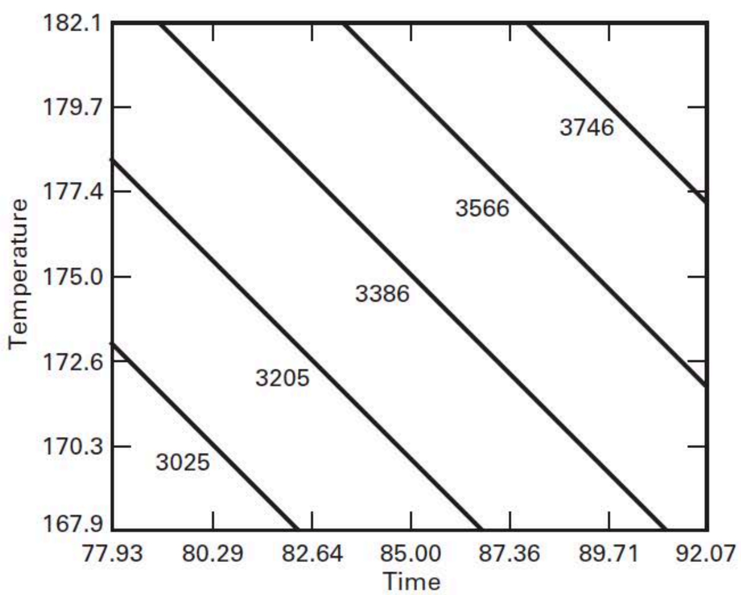


(a) The contour plot

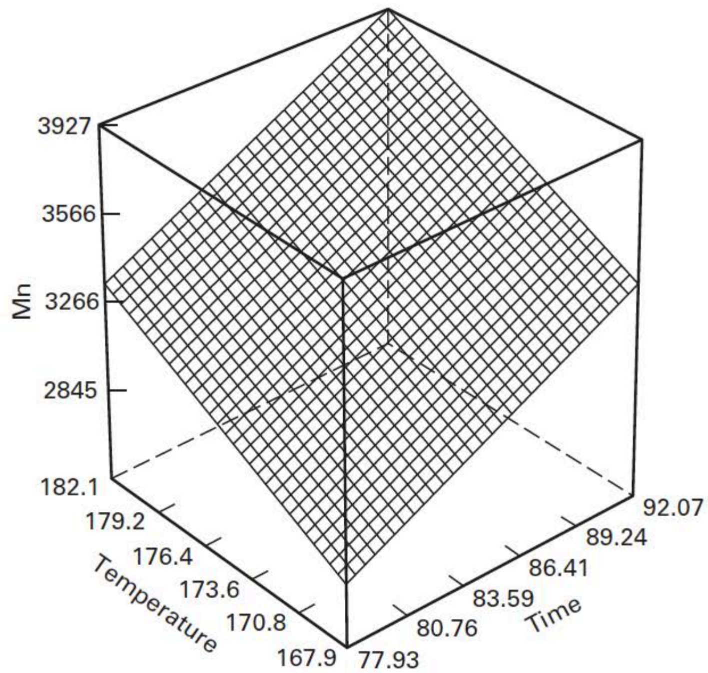


(b) The response surface plot

Contour Plot and Response Surface Plot of Molecular Weight



(a) The contour plot

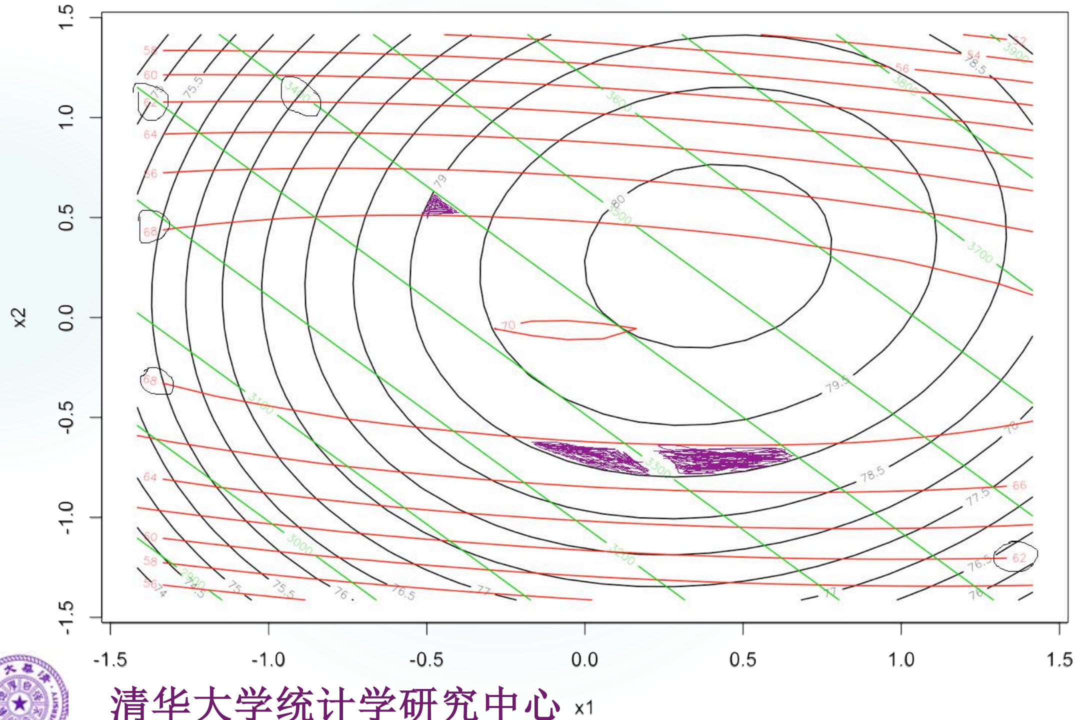


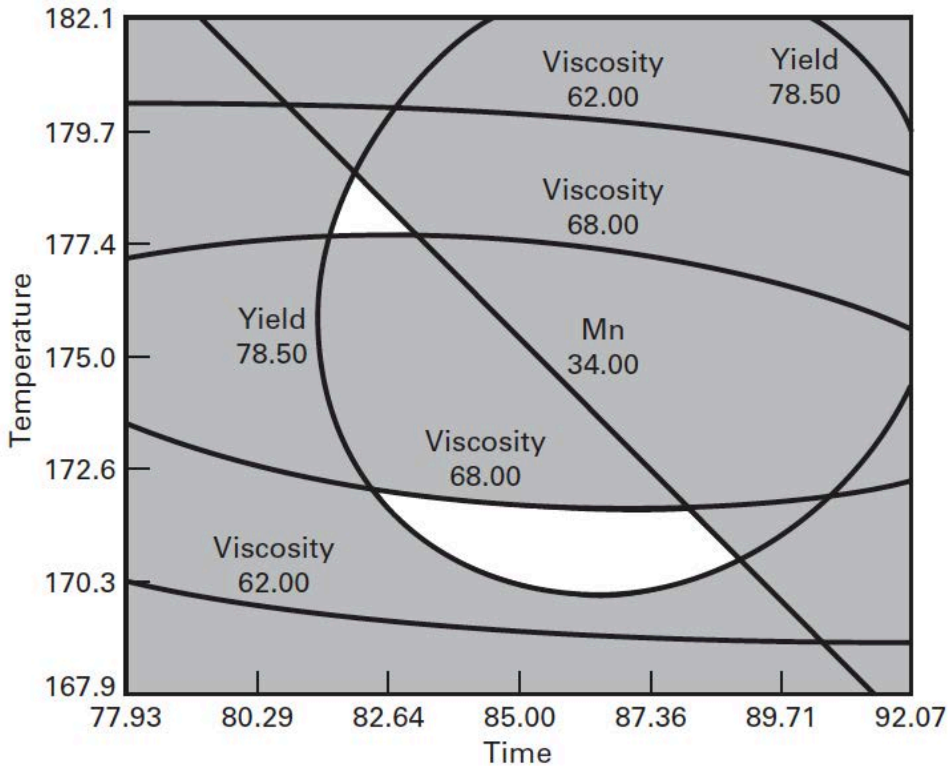
(b) The response surface plot

Approach 1: Overlay Contour Plots(覆盖等高线图)

- Region of the optimum found by overlaying yield, viscosity, and molecular weight response surfaces

```
> contour(rs,~x1+x2)
> contour(rsv,~x1+x2, add = T, col=2)
> contour(rsmw,~x1+x2, add = T, col=3)
```





Approach 2: Mathematical Programming Formulation

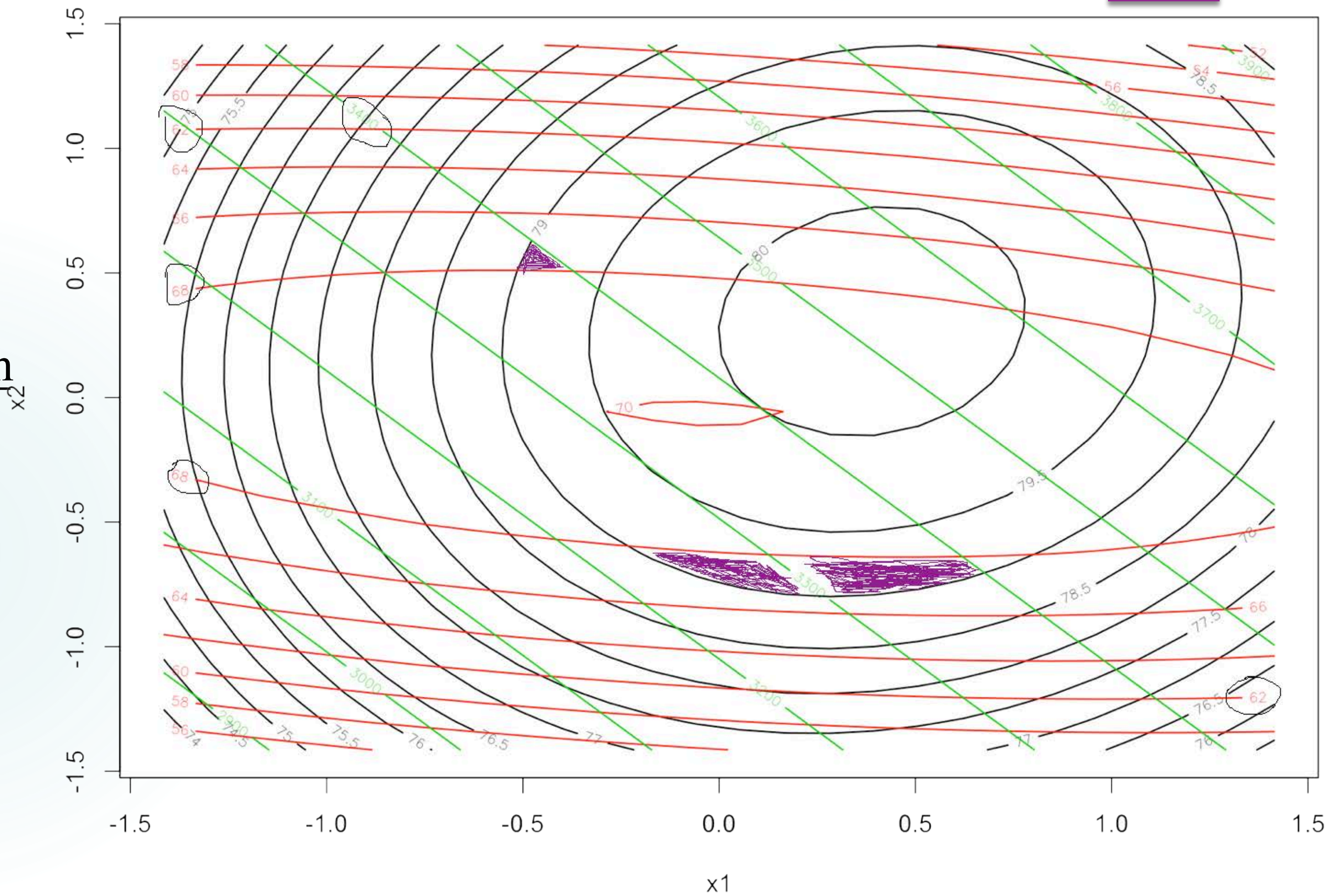
- Formulate and solve the problem as a <u>constrained optimization problem</u>

$$\text{Max } y_1$$

subject to

$$62 \le y_2 \le 68$$

$$y_3 \le 3400$$



Nonlinear Programming in R


```
> library('nloptr')
> x0 <- c(0, -.6)  # x0 <- c(-.4, .5)
> fn <- function(x) -79.94-.99*x[1]-.52*x[2]-.25*x[1]*x[2]+1.38*x[1]^2+x[2]^2
> hin <- function(x)
    c (70-.16*x[1]-.95*x[2]-.69*x[1]^2-6.69*x[2]^2-1.25*x[1]*x[2]-62,
     68-70+.16*x[1]+.95*x[2]+.69*x[1]^2+6.69*x[2]^2+1.25*x[1]*x[2],
     3400- 3386.2-205.1*x[1]-177.4*x[2] ) ## hin >= 0

> auglag(x0, fn, gr = NULL, hin = hin)
```

```
$par
[1] 0.2768991 -0.6389103

$value
[1] -79.32365

$par
[1] -0.3720258  0.5078785

$value
[1] -79.33962
```

Since $x_1 = (\xi_1 - 85)/5$, $x_2 = (\xi_2 - 175)/5$,

$(\xi_1, \xi_2) = (86.4, 171.8)$
$(\xi_1, \xi_2) = (83.1, 177.5)$

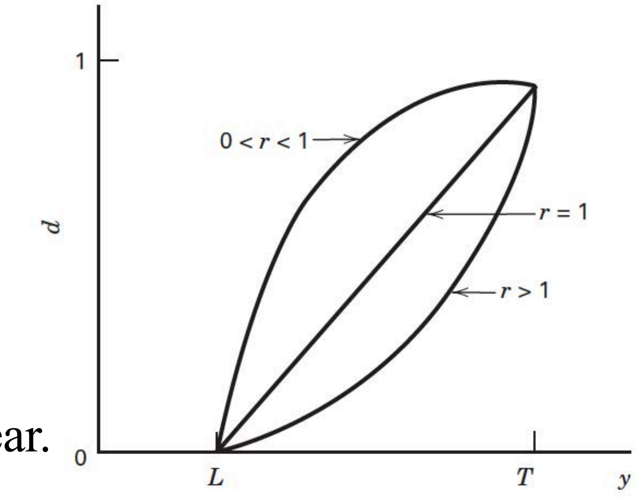
Approach 3: Desirability Function(渴求函数) Method

- First convert each response y_i into an individual desirability function d_i that varies over the range $0 \le d_i \le 1$ where if the response y_i is at its goal or target, then $d_i = 1$ and if the response is outside an acceptable region, $d_i = 0$, $1 \le i \le m$
- Then the design variables are chosen to maximize the overall desirability

$$D = (d_1 d_2 \cdots d_m)^{1/m}$$

- Larger-the-better (望大特征)

$$d = \begin{cases} 0 & y < L \\ \left(\dfrac{y-L}{T-L}\right)^r & L \le y \le T \\ 1 & y > T \end{cases}$$



(a) Objective (target) is to maximize y

when the weight $r = 1$, the desirability function is linear. Choosing $r > 1$ places more emphasis on being close to the target value and choosing $0 < r < 1$ makes this less important

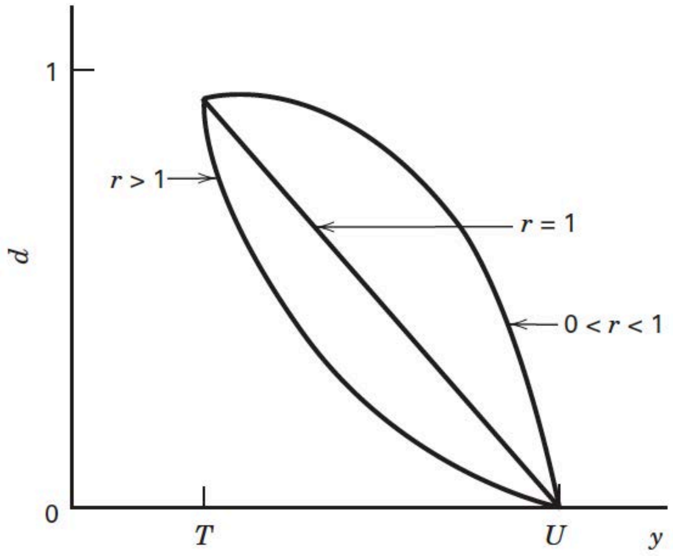
- Smaller-the-better (望小特征)

$$d=\begin{cases}1 & y<T\\ \left(\dfrac{U-y}{U-T}\right)^{r} & T\le y\le U\\ 0 & y>U\end{cases}$$

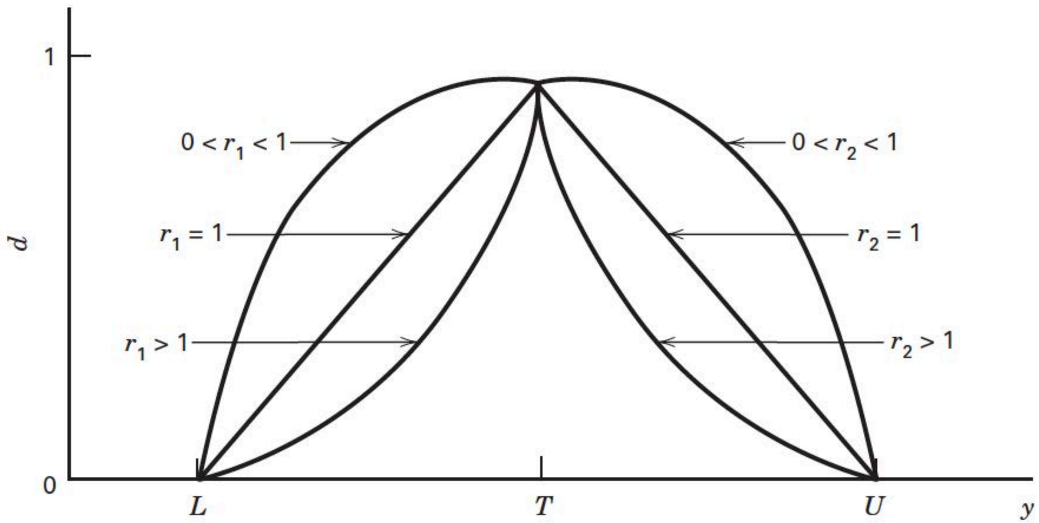
- Nominal-the-best (望目特征)

$$d=\begin{cases}0 & y<L\\ \left(\dfrac{y-L}{T-L}\right)^{r_1} & L\le y\le T\\ \left(\dfrac{U-y}{U-T}\right)^{r_2} & T\le y\le U\\ 0 & y>U\end{cases}$$

The two-sided desirability function assumes that the target is located between the lower (L) and upper (U) limits



(b) Objective (target) is to minimize y



(c) Objective is for y to be as close as possible to the target

The 'desirability' Package in R

- For multivariate optimization using the desirability function approach of Harrington (1965)
- Chose T = 80 as the target for the yield response with L = 70 and set the weight for this individual desirability equal to unity
- Set T = 65 for the viscosity response with L = 62 and U = 68 (to be consistent with specifications), with both weights $r_1 = r_2 = 1$
- Any molecular weight between 3200 and 3400 was acceptable
- Two solutions were found:
  - Solution 1

    Time = 81.71 Temp = 179.19 $D = 0.939$

    $\hat{y}1 = 78.3 \quad \hat{y}2 = 65 \quad \hat{y}3 = 3400$
  - Solution 2

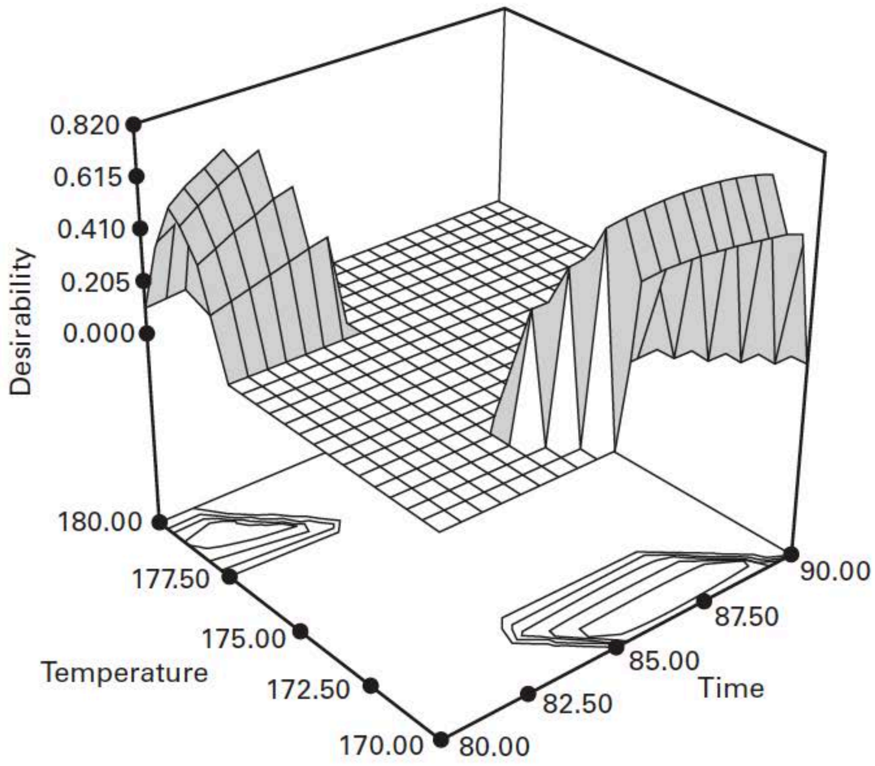
    Time = 86.1 Temp = 170.2 $D = 0.952$

    $\hat{y}1 = 78.63 \quad \hat{y}2 = 65 \; \hat{y}3 = 3260.86$

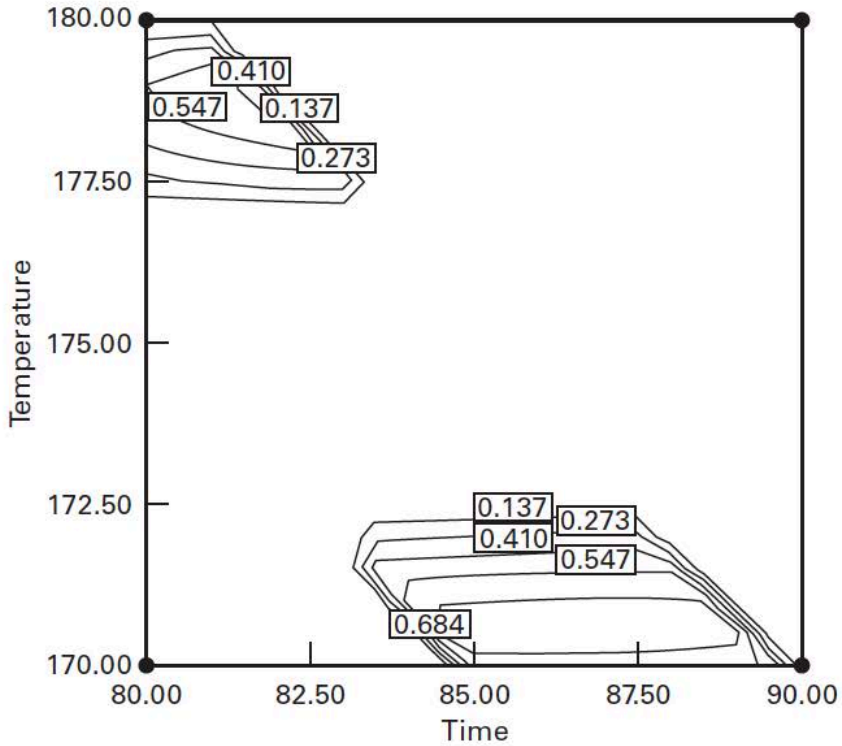
```
res <- optim(x0, D)
-D(res$par)
res$par*5+c(85, 175)
f1(res$par)
f2(res$par)
f3(res$par)
```

```
f1 <- function(x) 79.94+.99*x[1]+.52*x[2]+.25*x[1]*x[2]-1.38*x[1]^2-x[2]^2
f2 <- function(x) 70-.16*x[1]-.95*x[2]-.69*x[1]^2-6.69*x[2]^2-1.25*x[1]*x[2]
f3 <- function(x) 3386.2+205.1*x[1]+177.4*x[2]

library(desirability)
d1 <- dMax(low=70, high=80)
d2 <- dTarget(low=62, target=65, high=68)
d3 <- dBox(low=3200, high=3400)
D <-  function(x) -(predict(d1, f1(x))*predict(d2, f2(x))*predict(d3, f3(x)))^(1/3)
```

Desirability Function Response Surface & Contour Plot

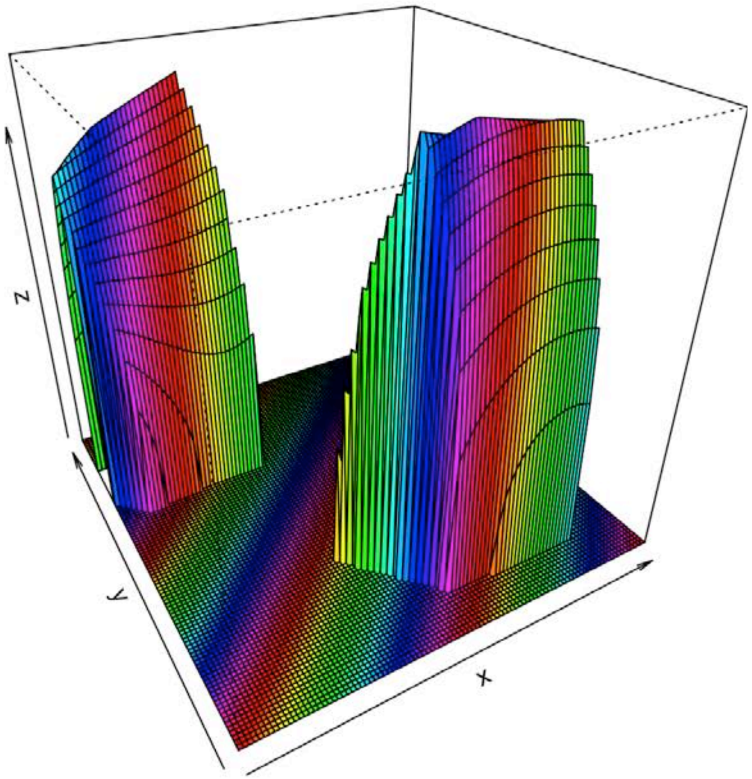


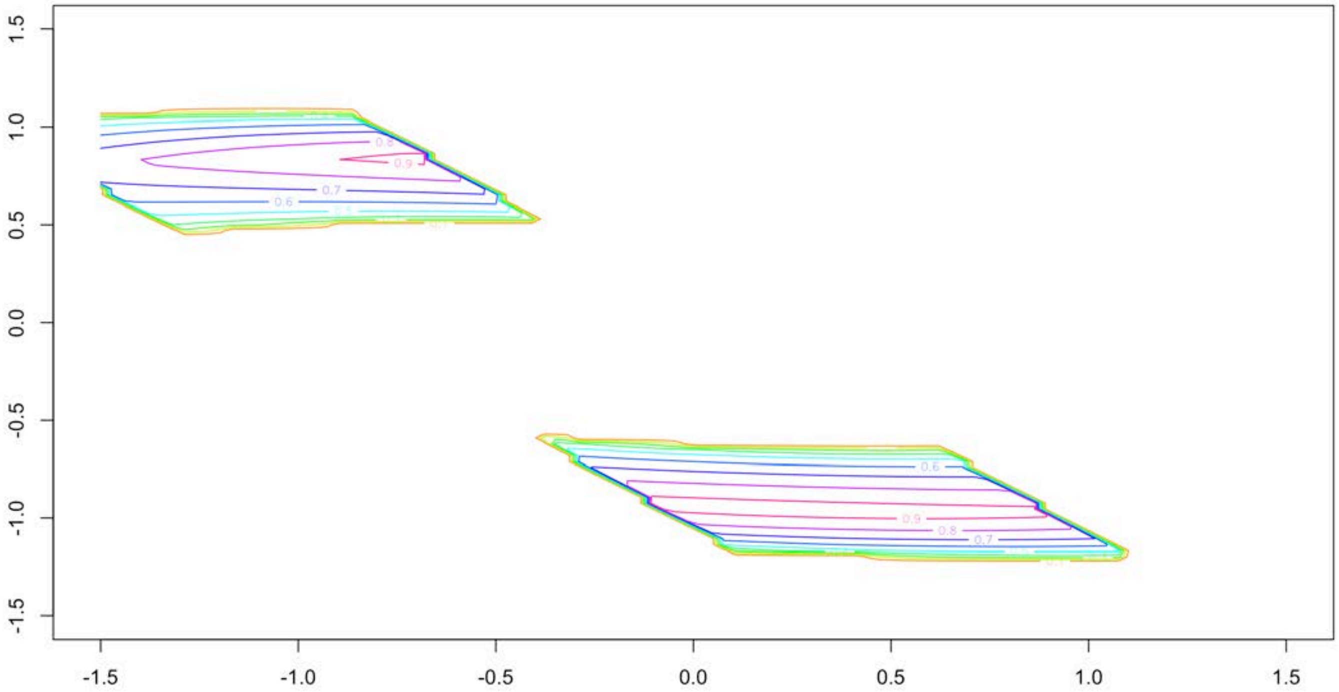
(a) Response surface



(b) Contour plot

Plots in R





In Short

- Multiple responses are common in practice
- Typically, we want to simultaneously optimize all responses, or find a set of conditions where certain product properties are achieved
- Approaches:
  - 1. Model all responses & overlay the contour plots
  - 2. Optimized the most import factors while setting constraints to other factors
  - 3. Combine multiple responses into a “integrated response”

Experimental Designs for Fitting Response Surfaces

- When selecting a RS design, some of the features of a desirable design:
  - *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

Designs for Fitting the First-Order Model

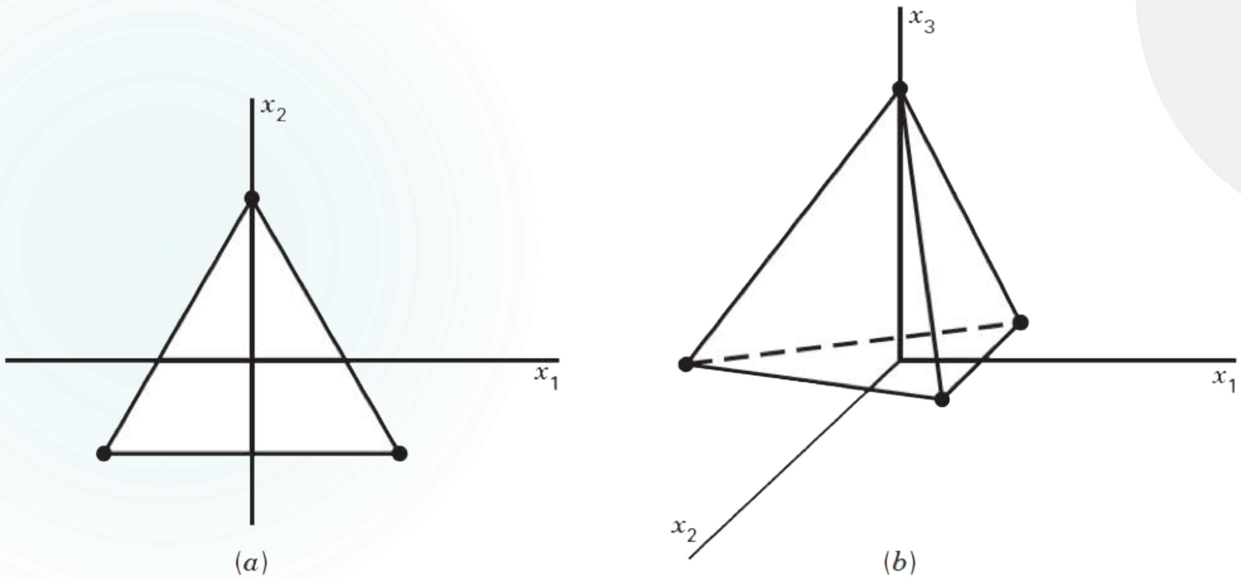
- With k variables

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon$$

- The orthogonal first-order designs is a unique class of designs that minimize the variance of the regression coefficients $\{\hat{\beta_i}\}$
- A design is orthogonal if the off-diagonal elements of the (X′X) matrix are all zero. It 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 ± 1 levels

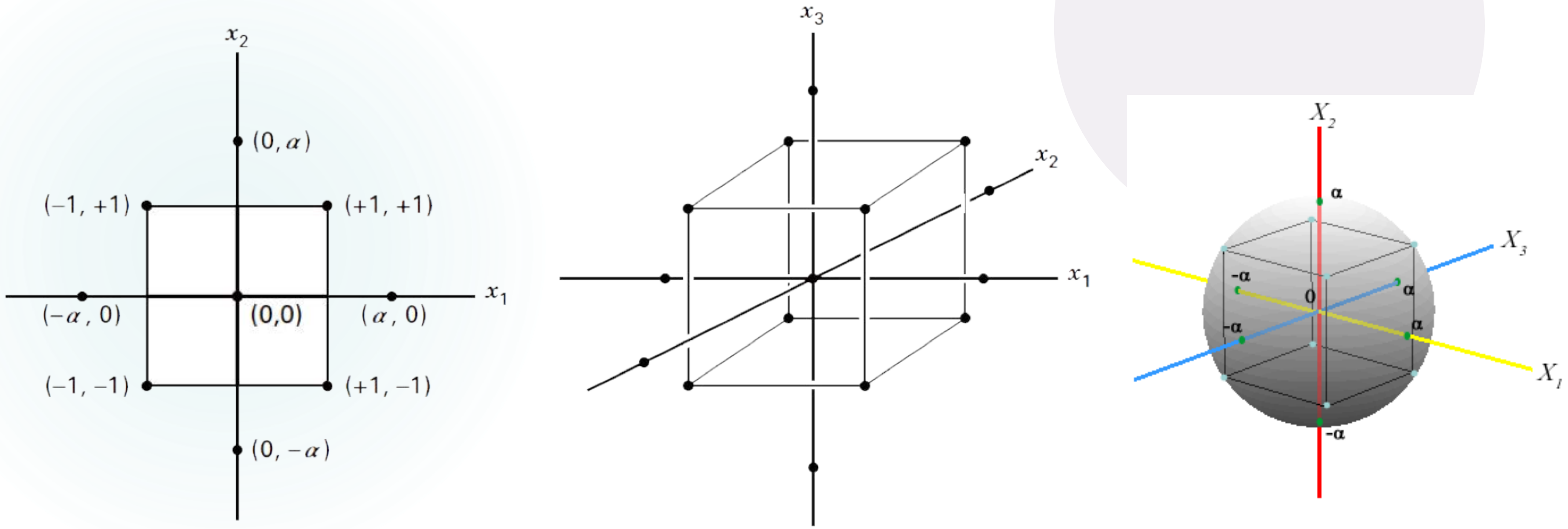
Note on Simplex(单纯形)

- Another orthogonal first-order design is the simplex
- It is a regularly sided figure with $k + 1$ vertices in k dimensions
  - for $k = 2$ it is an equilateral triangle(等边三角形)
  - for $k = 3$ it is a regular tetrahedron(四面体)



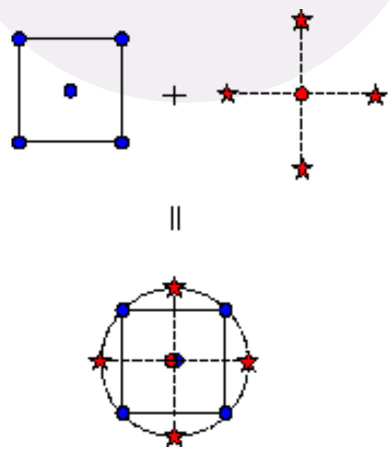
Designs for Fitting the Second-Order Model

- Box-Wilson Central Composite Design(CCD, 中心复合设计) is the most popular class of designs used for fitting these models
- The CCD consists of a 2^k factorial (or fractional factorial of resolution V) with n_F factorial runs, $2k$ axial or star runs(星点/轴点), and n_C center runs



Keys of CCD

- The practical deployment of a CCD often arises through sequential experimentation
- That is, a 2^k has been used to fit a first-order model, this model has exhibited lack of fit, and the axial runs are then added to allow the quadratic terms to be incorporated into the model
- The CCD is a very efficient design for fitting the second-order model. Two parameters must be specified:
  - the distance α of the axial runs from the design center
  - the number of center points n_C
- We now discuss the choice of these two parameters

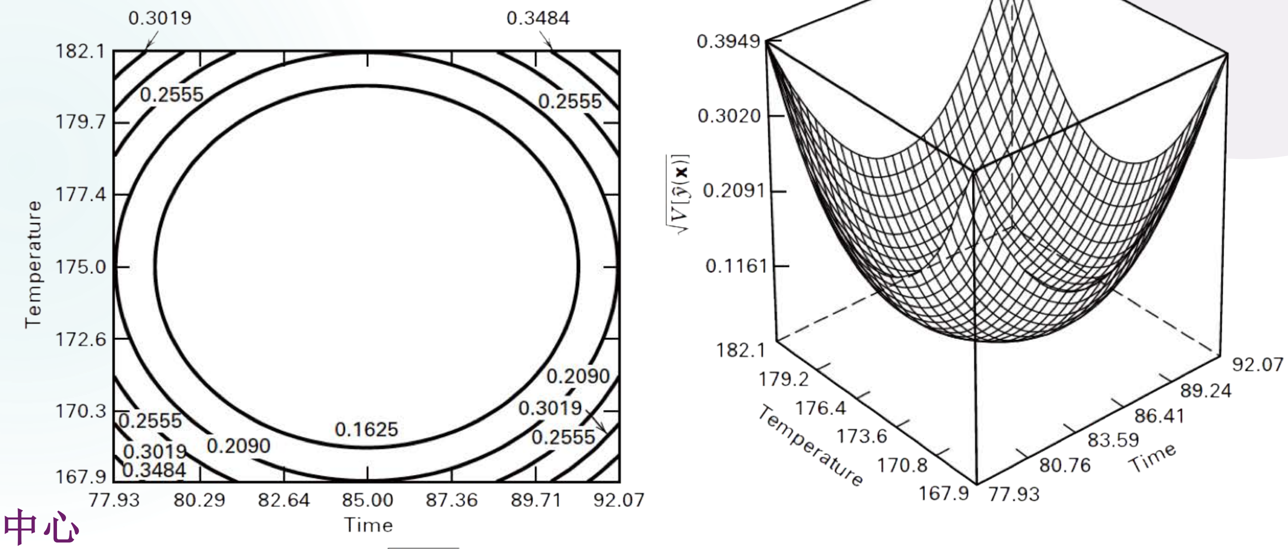


The Rotatable CCD

- A ‘good’ model has <u>rotatability</u>. That means:
  - It has a reasonably consistent and stable variance of the predicted response at points of interest x
  - $Var\left(\hat{y}(x)\right) = \sigma^2 x'(X'X)^{-1}x$

    is the same at all points x that are at the same distance from the design center
  - The variance of predicted response is constant on spheres

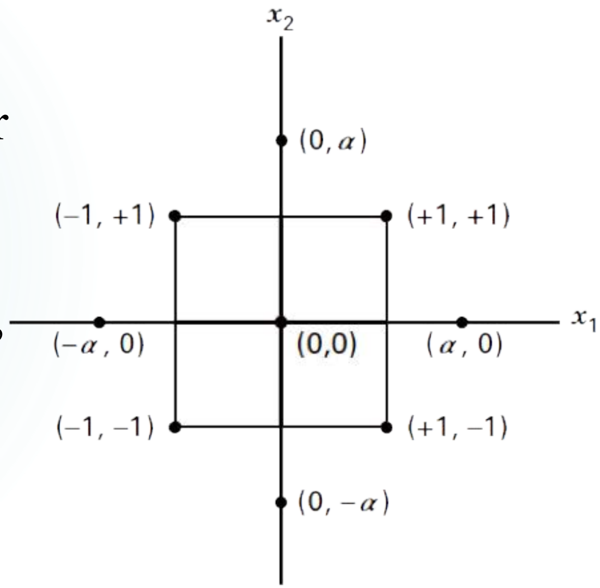
A design with this property will leave the variance of $\hat{y}$ unchanged when the design is rotated about the center (0, 0, . . . , 0), hence the name rotatable design

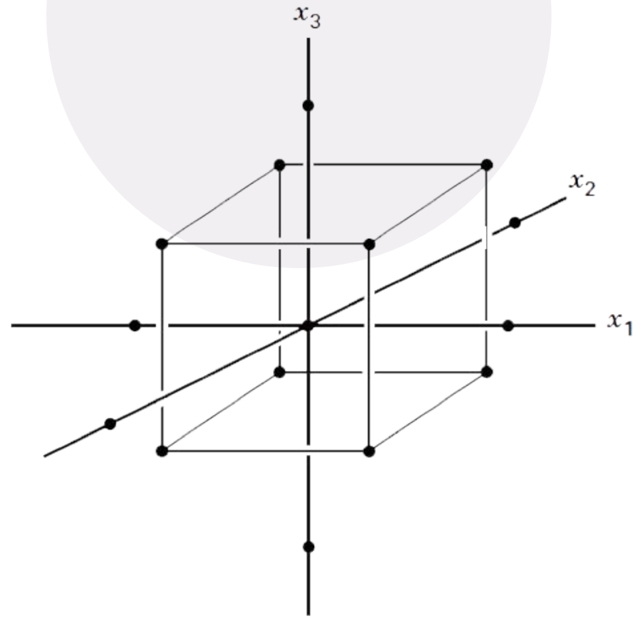


(a) Contours of $\sqrt{V[\hat{y}(\mathbf{x})]}$ (b) The response surface plot

The Spherical CCD

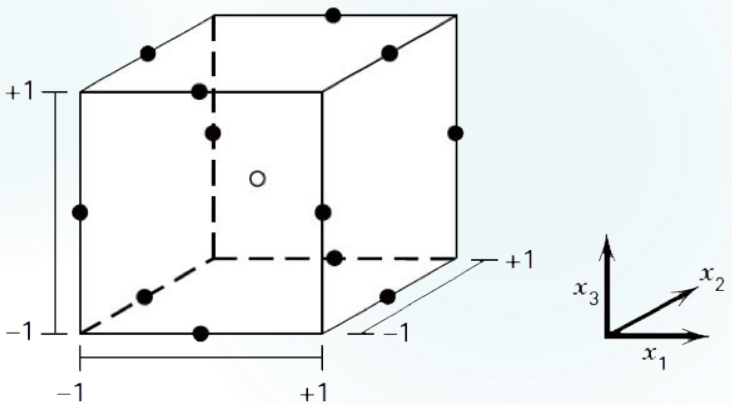
- Rotatability is a spherical property; that is, it makes the most sense as a design criterion when the region of interest is a sphere.
- The best choice of α is to set $\alpha = \sqrt{k}$, called a spherical CCD, puts all the factorial and axial design points on the surface of a sphere of radius $\sqrt{k}$
- When this region is a sphere, the design must include center runs to provide reasonably stable variance of the predicted response. Generally, three to five center runs are recommended





The Box–Behnken Design

- Box and Behnken (1960) have proposed some three-level designs for fitting response surfaces
  - formed by combining 2^k factorials with incomplete block designs
  - are usually very efficient in terms of the number of required runs
  - are either rotatable or nearly rotatable

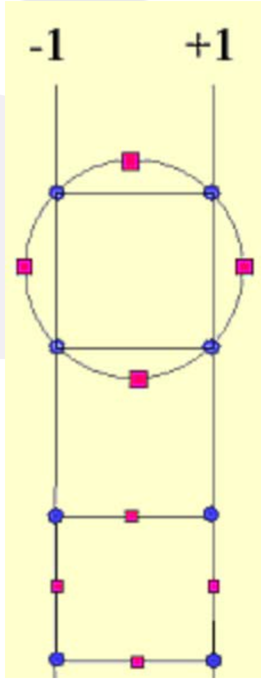


- A spherical design, with all points lying on a sphere of radius $\sqrt{2}$
- Does not contain any points at the vertices of the cubic region

Run	x_1	x_2	x_3
1	−1	−1	0
2	−1	1	0
3	1	−1	0
4	1	1	0
5	−1	0	−1
6	−1	0	1
7	1	0	−1
8	1	0	1
9	0	−1	−1
10	0	−1	1
11	0	1	−1
12	0	1	1
13	0	0	0
14	0	0	0
15	0	0	0

A Design on A Cube – The Face-Centered CCD

- When the region of interest is cuboidal rather than spherical, a useful variation of the CCD is the face-centered CCD/cube, in which $\alpha = 1$
- Locate the star/axial points on the centers of the faces of the cube
- Also used in practice when it is difficult to change factor levels
- Not rotatable
- Does not require as many center points as the spherical CCD. In practice, $n_C = 2$ or 3 is sufficient to provide good variance of prediction throughout the experimental region



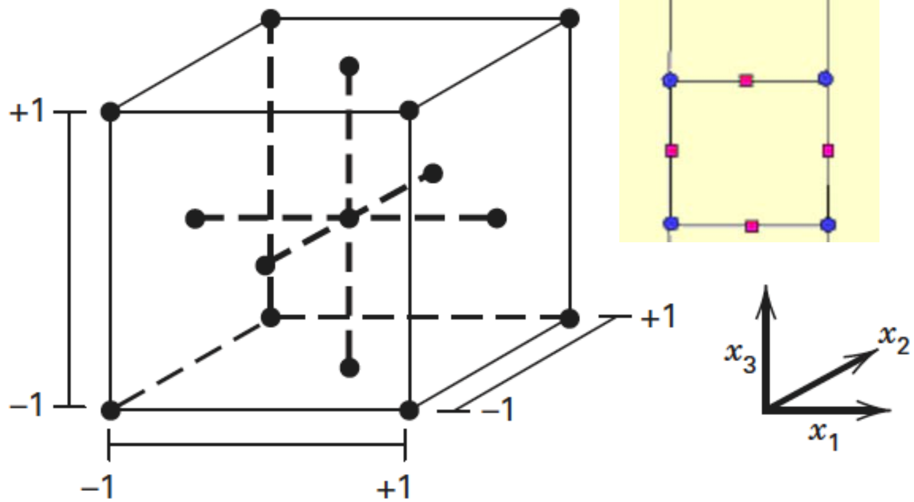
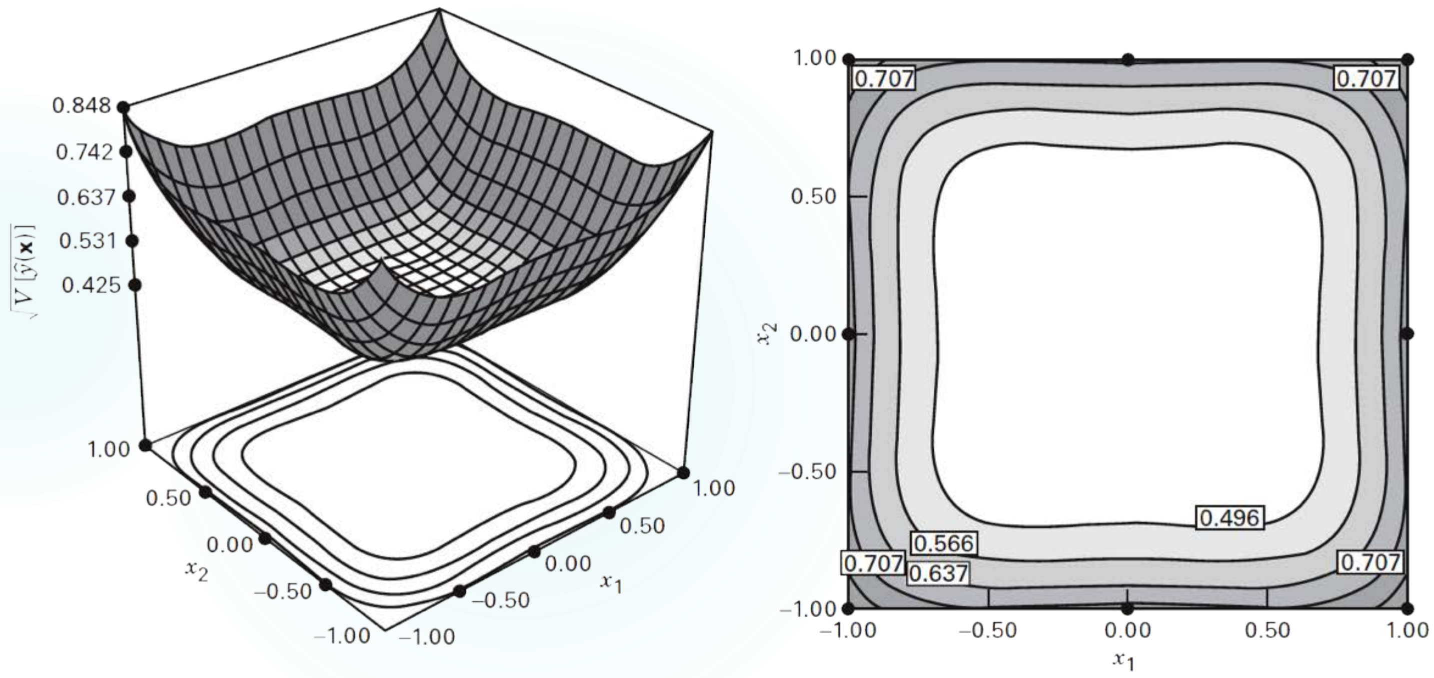


Illustration of Non-constant Prediction Variance

- The square root of prediction variance $\sqrt{Var(\hat{y}(x))}$ for the face-centered cube for $k = 3$ with $n_C = 3$ center points

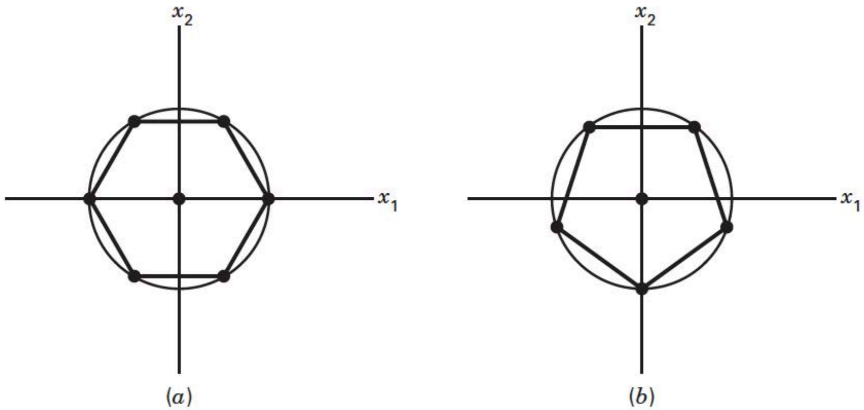


(*a*) Response surface (*b*) Contour plot

The standard deviation of predicted response is reasonably uniform over a relatively large portion of the design space

Other Designs

- Equiradial designs ($k = 2$ only): regular polygons
- The small composite design (SCD)
  - Consists of a fractional factorial in the cube of resolution III and the usual axial and center runs
  - Not a great choice because of poor prediction variance properties
- Hybrid designs
  - Excellent prediction variance properties
  - Unusual factor levels
- Computer-generated designs



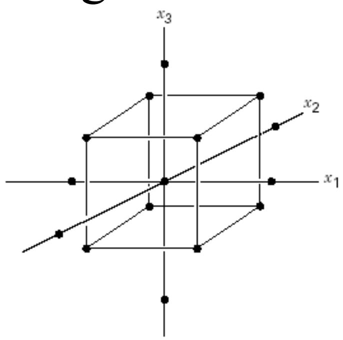
Standard Order	x_1	x_2	x_3
1	1.00	1.00	−1.00
2	1.00	−1.00	1.00
3	−1.00	1.00	1.00
4	−1.00	−1.00	−1.00
5	−1.73	0.00	0.00
6	1.73	0.00	0.00
7	0.00	−1.73	0.00
8	0.00	1.73	0.00
9	0.00	0.00	−1.73
10	0.00	0.00	1.73
11	0.00	0.00	0.00
12	0.00	0.00	0.00
13	0.00	0.00	0.00
14	0.00	0.00	0.00

Ranitidine(呋喃硝胺) Experiment Example

- Consider an experiment to study three quantitative factors with up to 5 levels

Factor	Levels
A. pH	2, 3.42, 5.5, 7.58, 9
B. voltage (kV)	9.9, 14, 20, 26, 30.1
C. a-CD (mM)	0, 2, 5, 8, 10

- The design matrix and the data are ->
- The CCD design differs from 2^{k-p} design in two respects :
  - 6 replicates at the center
  - 6 runs along the three axes



Run	Factor A	Factor B	Factor C	CEF	ln CEF
1	−1	−1	−1	17.293	2.850
2	1	−1	−1	45.488	3.817
3	−1	1	−1	10.311	2.333
4	1	1	−1	11757.084	9.372
5	−1	−1	1	16.942	2.830
6	1	−1	1	25.400	3.235
7	−1	1	1	31697.199	10.364
8	1	1	1	12039.201	9.396
9	0	0	−1.67	7.474	2.011
10	0	0	1.67	6.312	1.842
11	0	−1.68	0	11.145	2.411
12	0	1.68	0	6.664	1.897
13	−1.68	0	0	16548.749	9.714
14	1.68	0	0	26351.811	10.179
15	0	0	0	9.854	2.288
16	0	0	0	9.606	2.262
17	0	0	0	8.863	2.182
18	0	0	0	8.783	2.173
19	0	0	0	8.013	2.081
20	0	0	0	8.059	2.087

A question to think about

- Why do we need the “star points” in the central composite design?

Answer

- Corner points help to estimate main & interaction effects
- Center points help to estimate noise term σ^2
- Star points help to estimate quadratic effects $\{\beta_{ii}\}$
  - Without star points, the quadratic effects are confounded together

Number of Runs Summary

- If we have k factors, then we have, 2^k factorial points, $2*k$ axial points and n_c center points. Below is a table that summarizes these designs and compares them to 3^k designs

		$k=2$	$k=3$	$k=4$	$k=5$
Central Composite Designs	Factorial points 2^k	4	8	16	32
	Star points 2^k	4	6	8	10
	Center points n_c (varies)	5	5	6	6
	Total	13	19	30	48
3^k Designs		9	27	81	243
Choice of α	Spherical design ($\alpha = \sqrt{k}$)	1.4	1.73	2	2.24
	Rotatable design ($\alpha = (n_F)^{\frac{1}{4}}$)	1.4	1.68	2	2.38

Final Slide: Design Selection Guideline

	Current State of Possessed Knowledge			
Knowledge	Low ←-----------+--------------+--------------+--------→ High			
Type of Design	Screening	Fractional Factorials	Factorials	Response Surface
Usual # of Factors	>4	3-15	1-7	<8
Purpose: Identification	Most important factors (vital few)	Some interactions	Relationships among all factors	Optimal factor setting
Purpose: Estimation	Crude direction for improvement (linear effects)	All main effects and some interactions	All main effects and all interactions	Curvature in response, empirical models