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

- Experiments with a single quantitative factor
  - Regression model vs. fixed effect model
  - Model with non-linear terms
- One Quality Factor & One Quantitative Factor
  - Regression on each level of one factor

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

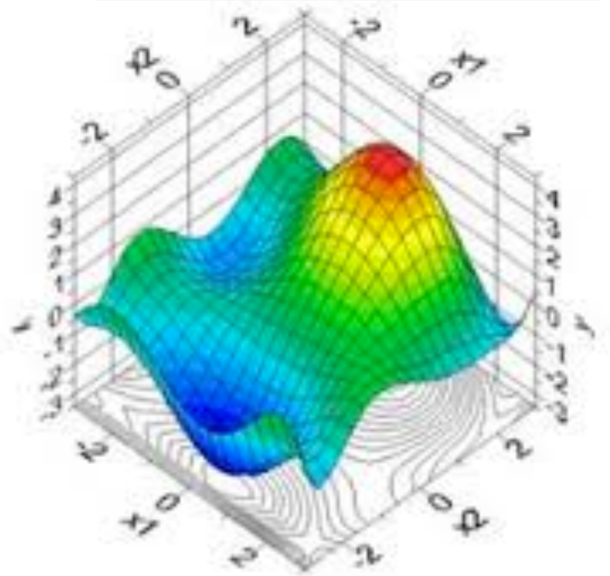
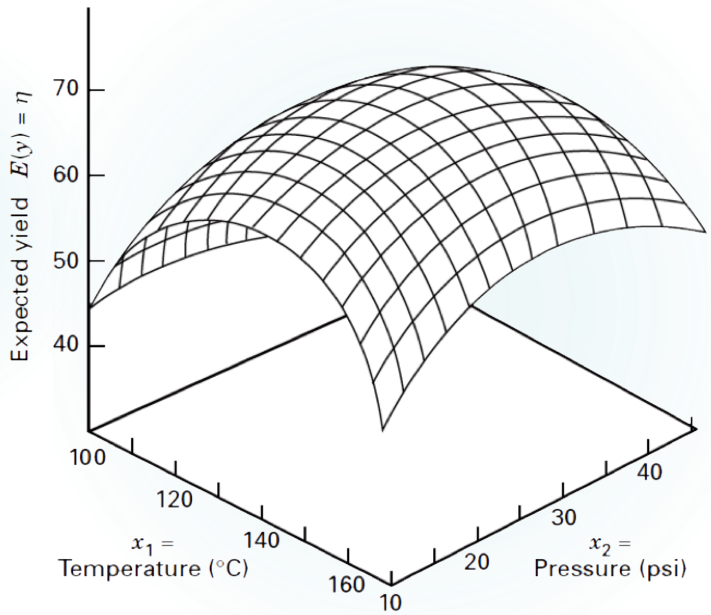


Illustration of a RS

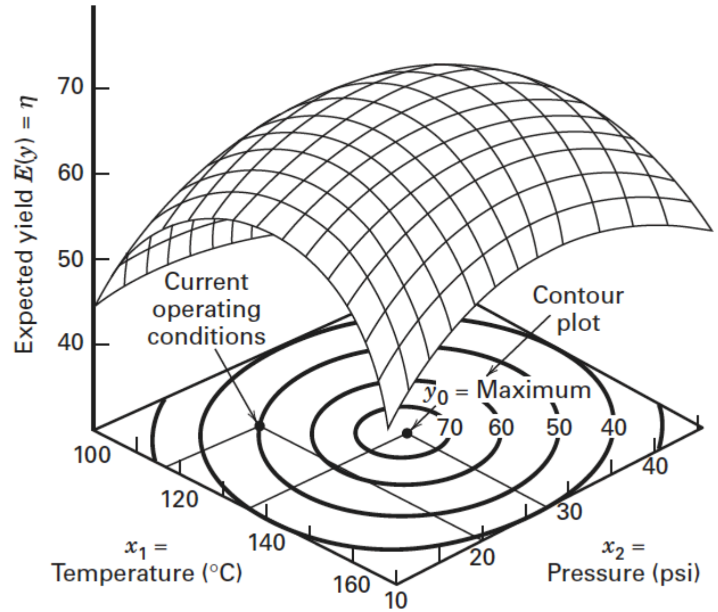
$$y = f(x_1, x_2) + \varepsilon$$

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)



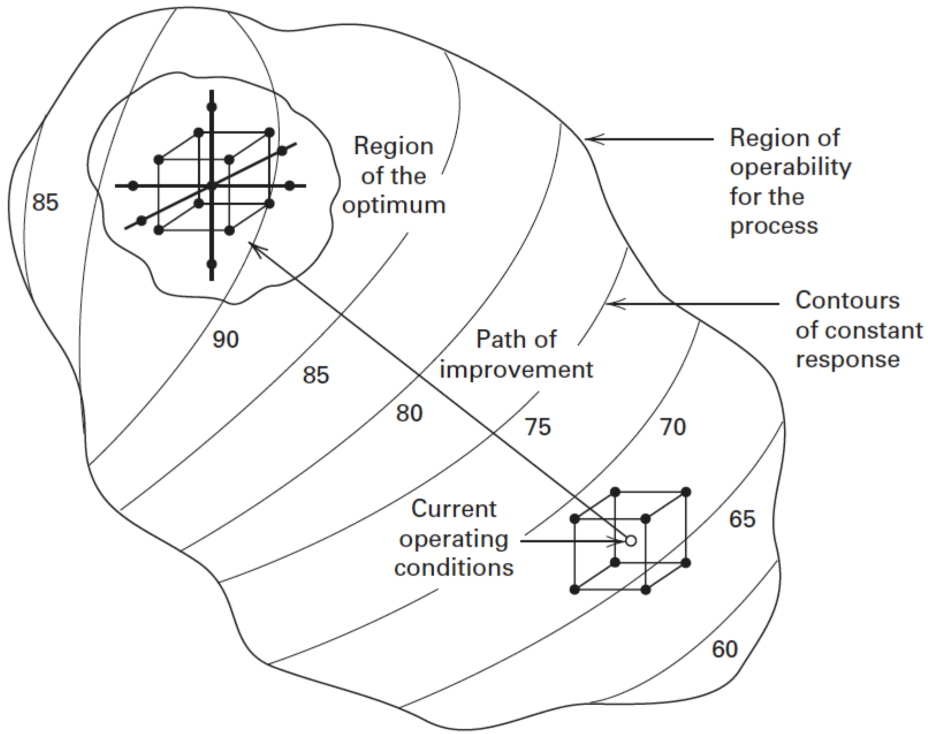
A contour plot of a response surface

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 2st 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

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

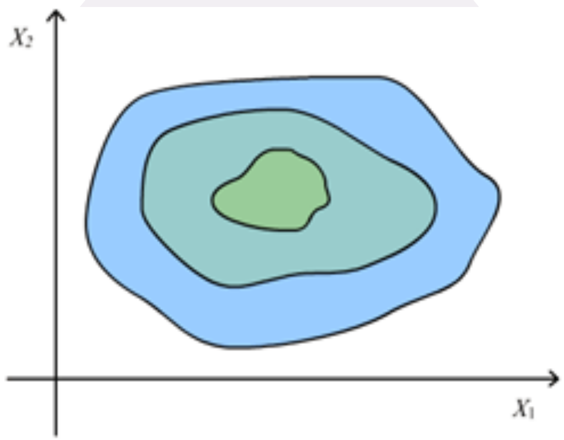
- **Optimization Model**

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

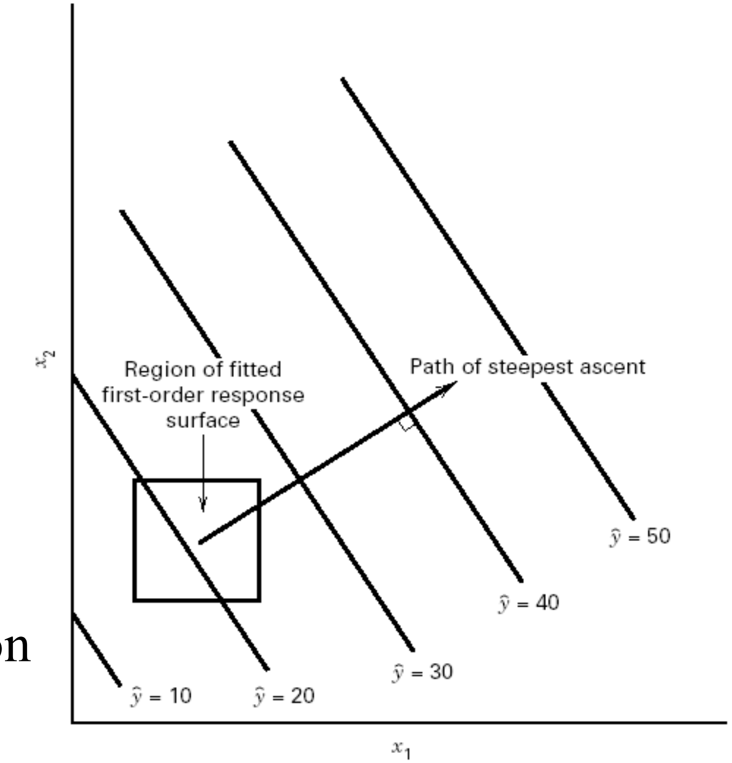
- When we are remote from the optimum, we usually assume that a first-order model is an adequate approximation to the true surface in a small region of the x's
- A procedure for moving sequentially from an initial "guess" towards to region of the optimum
- The 1st order Taylor expansion

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

- 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?

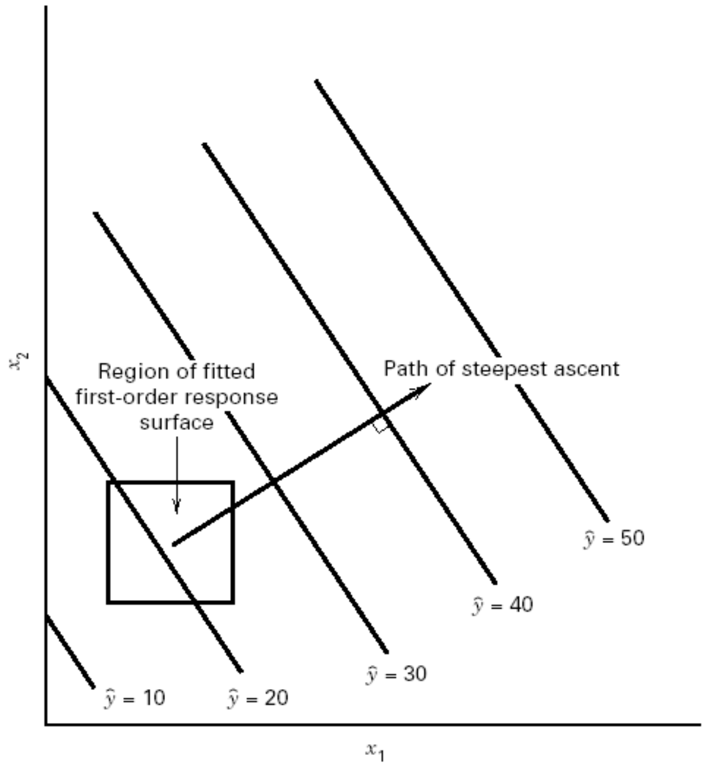
- A: For any arbitrary direction v, the rate of change along v is

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

  - 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)
```

```
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
```

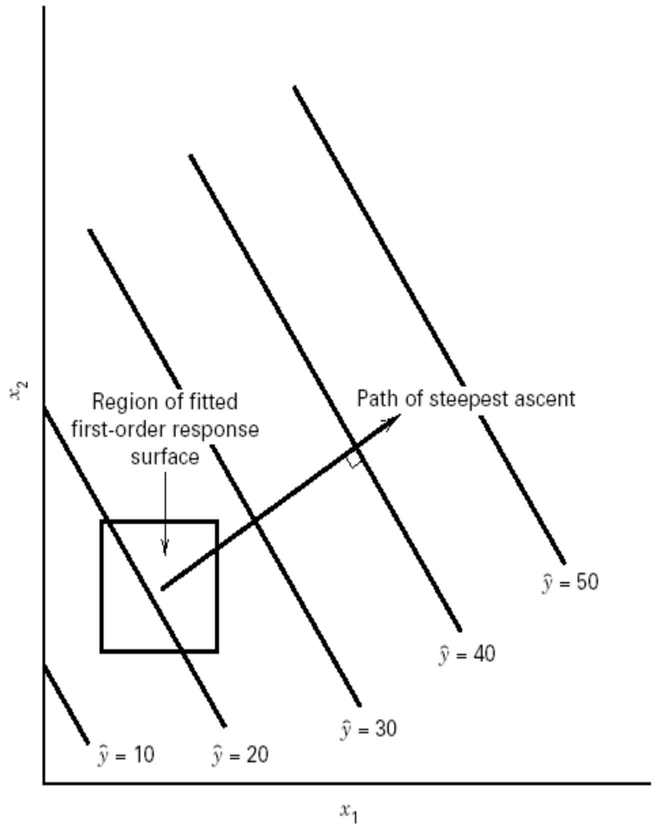
```
Analysis of Variance Table

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
```

- 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			

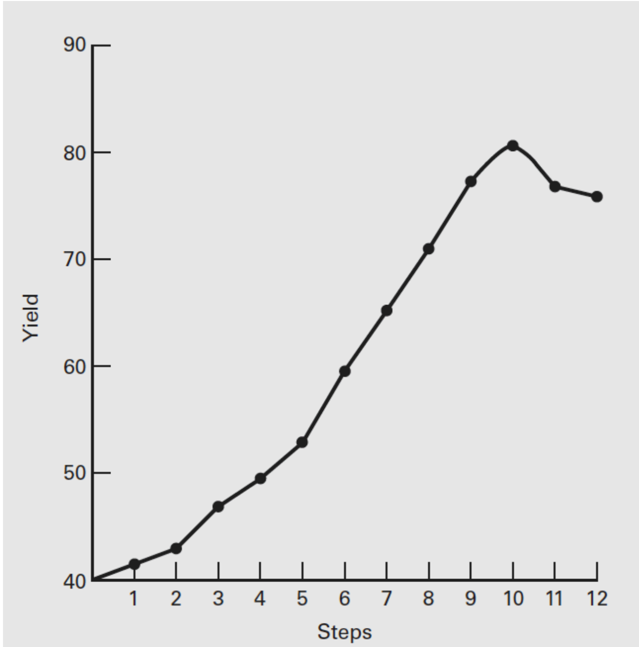
Decide the Direction

- $\hat{y} = 40.44 + 0.775x_1 + 0.325x_2$
- 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 $\Delta x_i = \dfrac{\widehat{\beta}_i}{\widehat{\beta}_j / \Delta x_j}, i \neq j$
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

2^{nd} Order Approximation

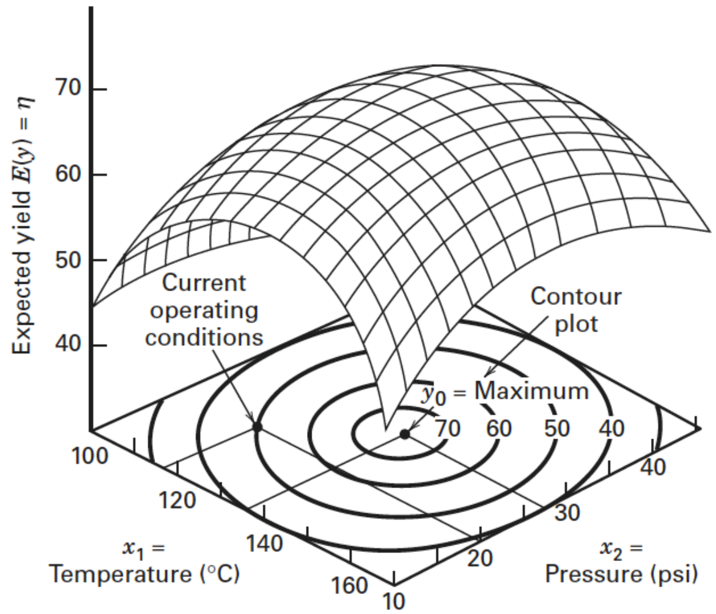
- The 2^{nd} order Taylor expansion

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

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

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

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



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

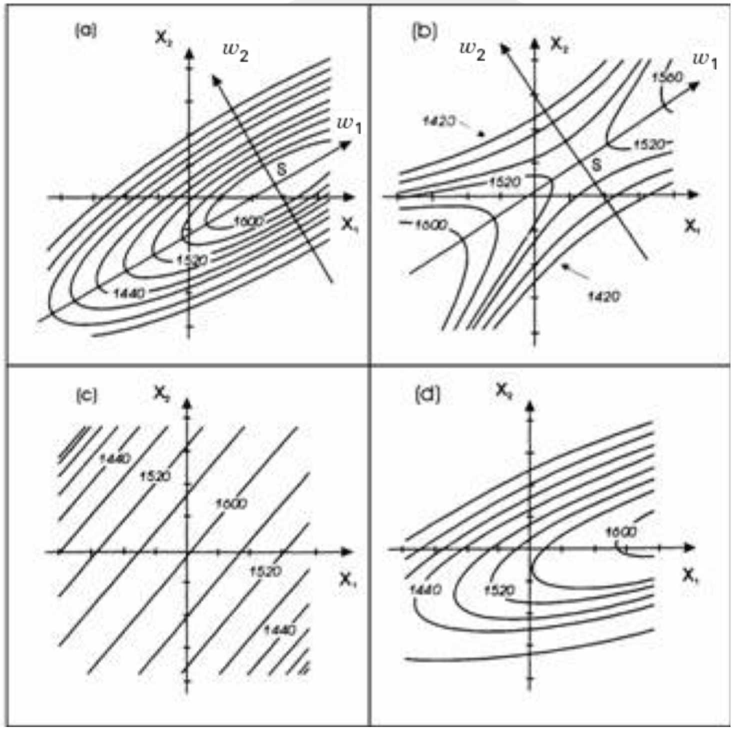


Illustration of a Surface with a Maximum

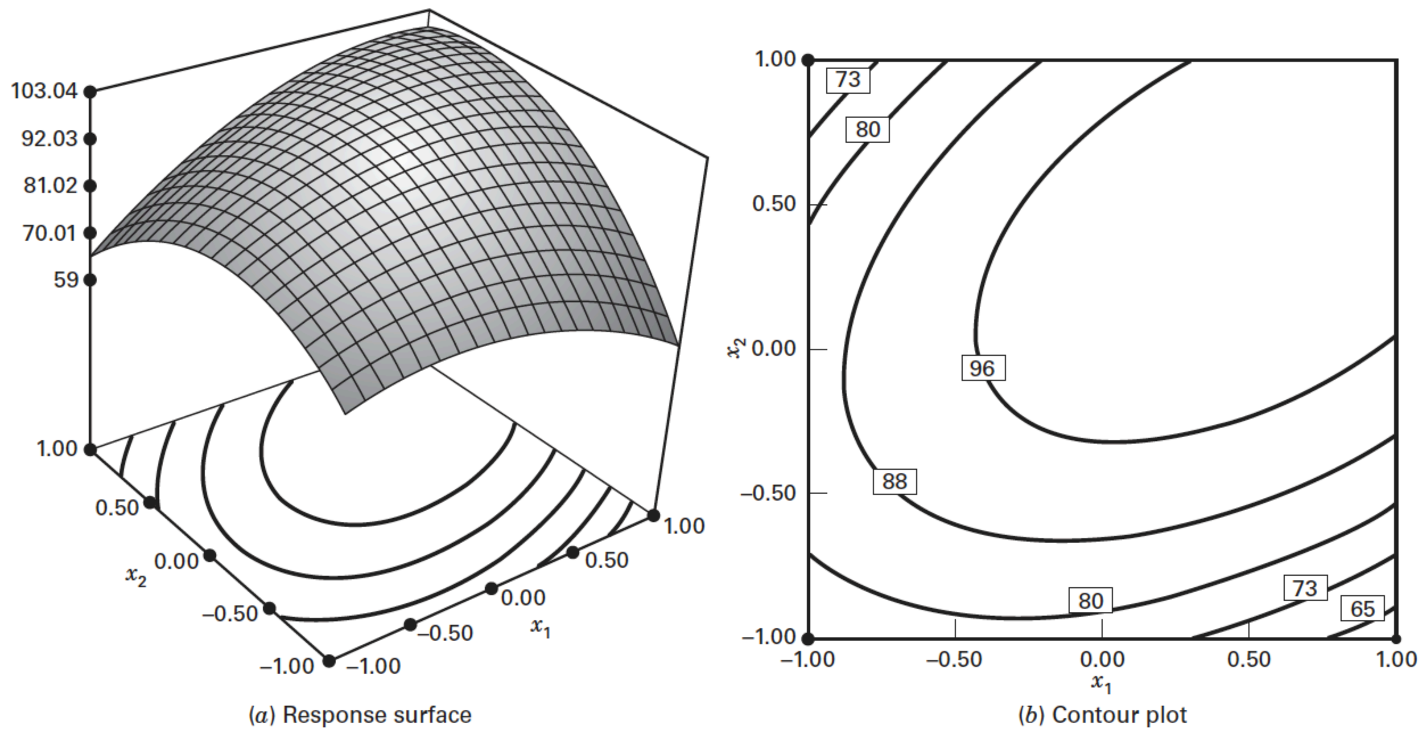
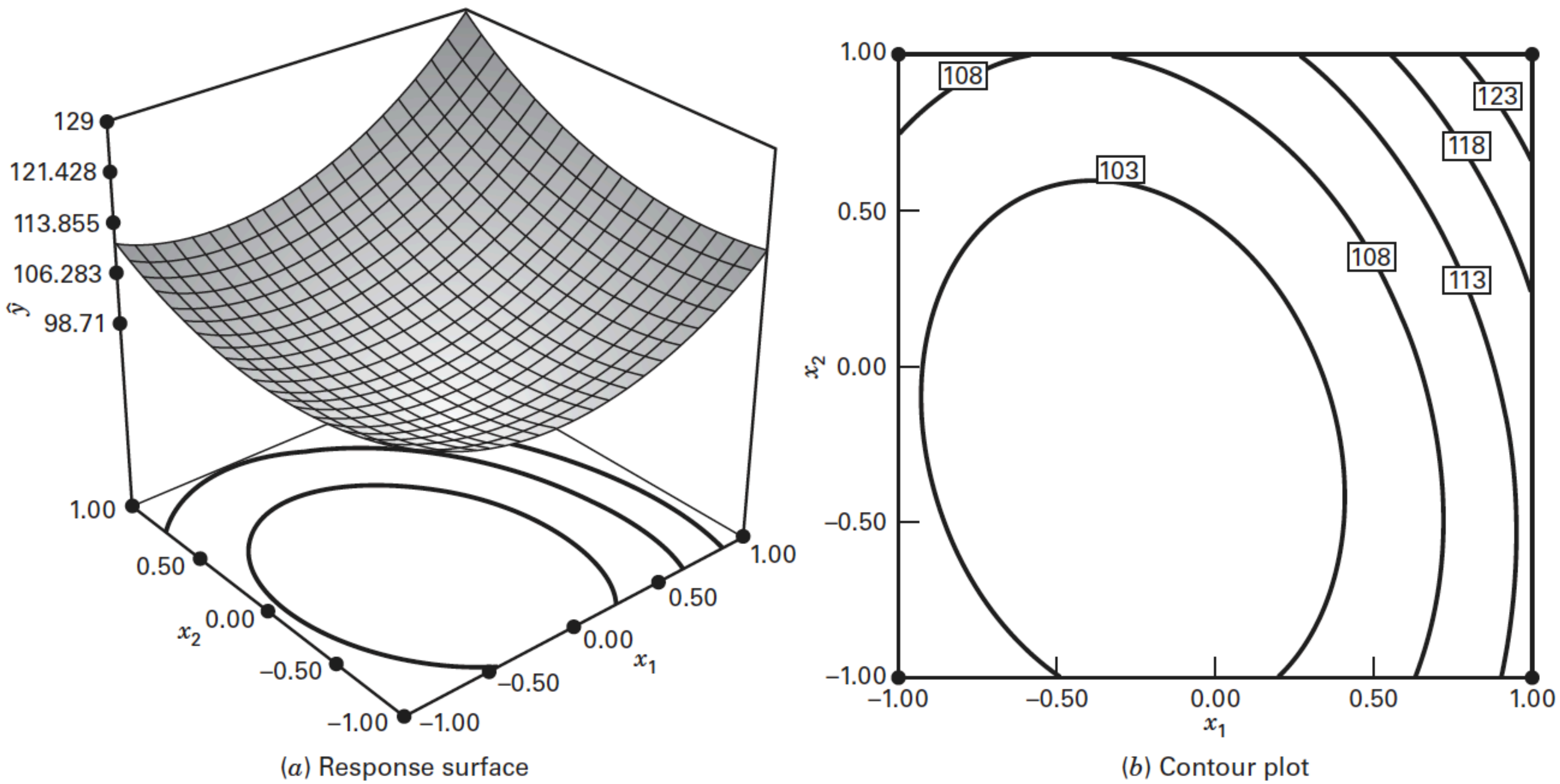
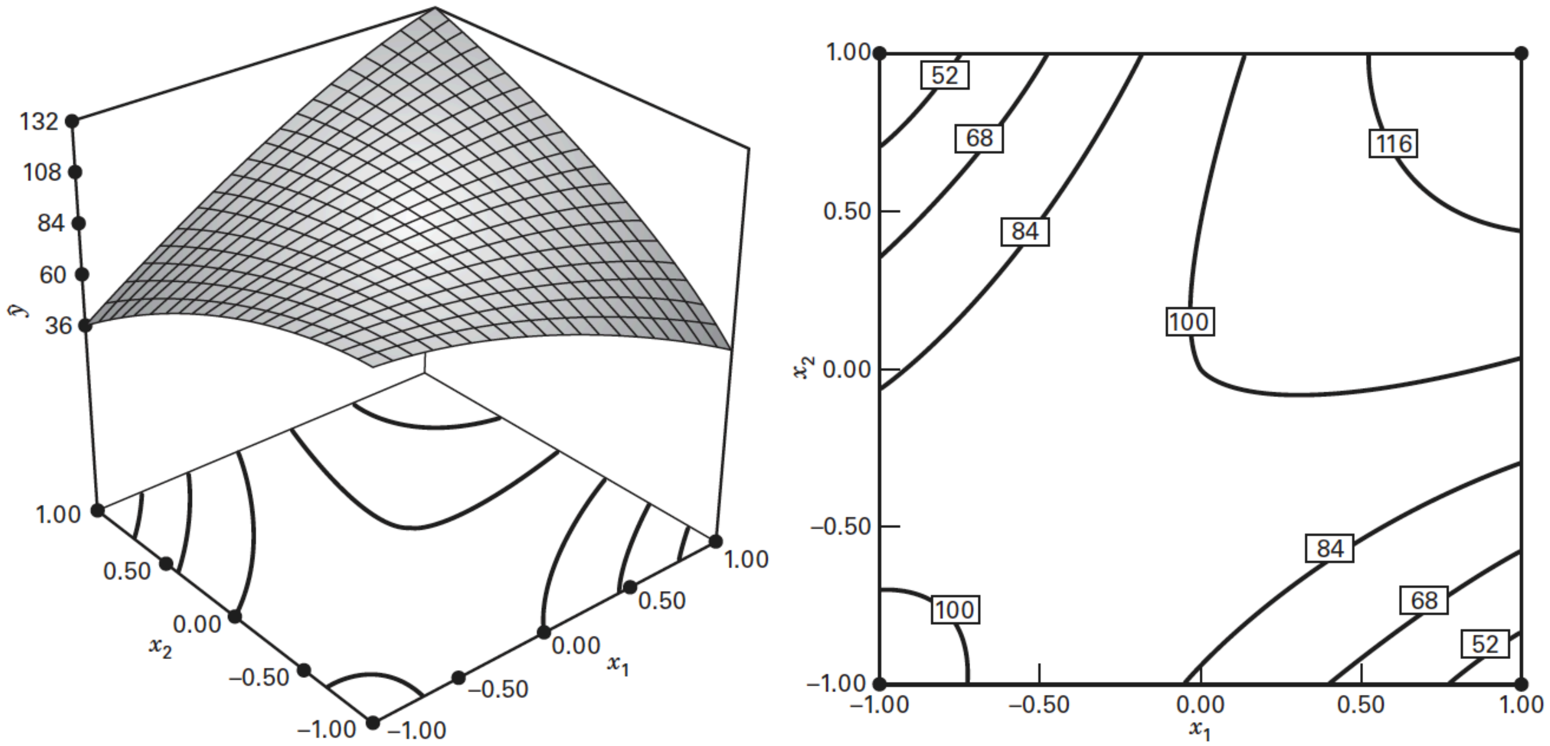


Illustration of a Surface with a Minimum

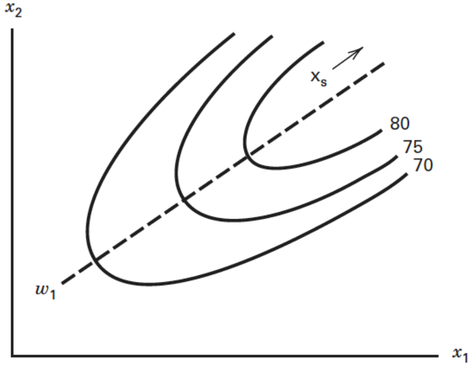


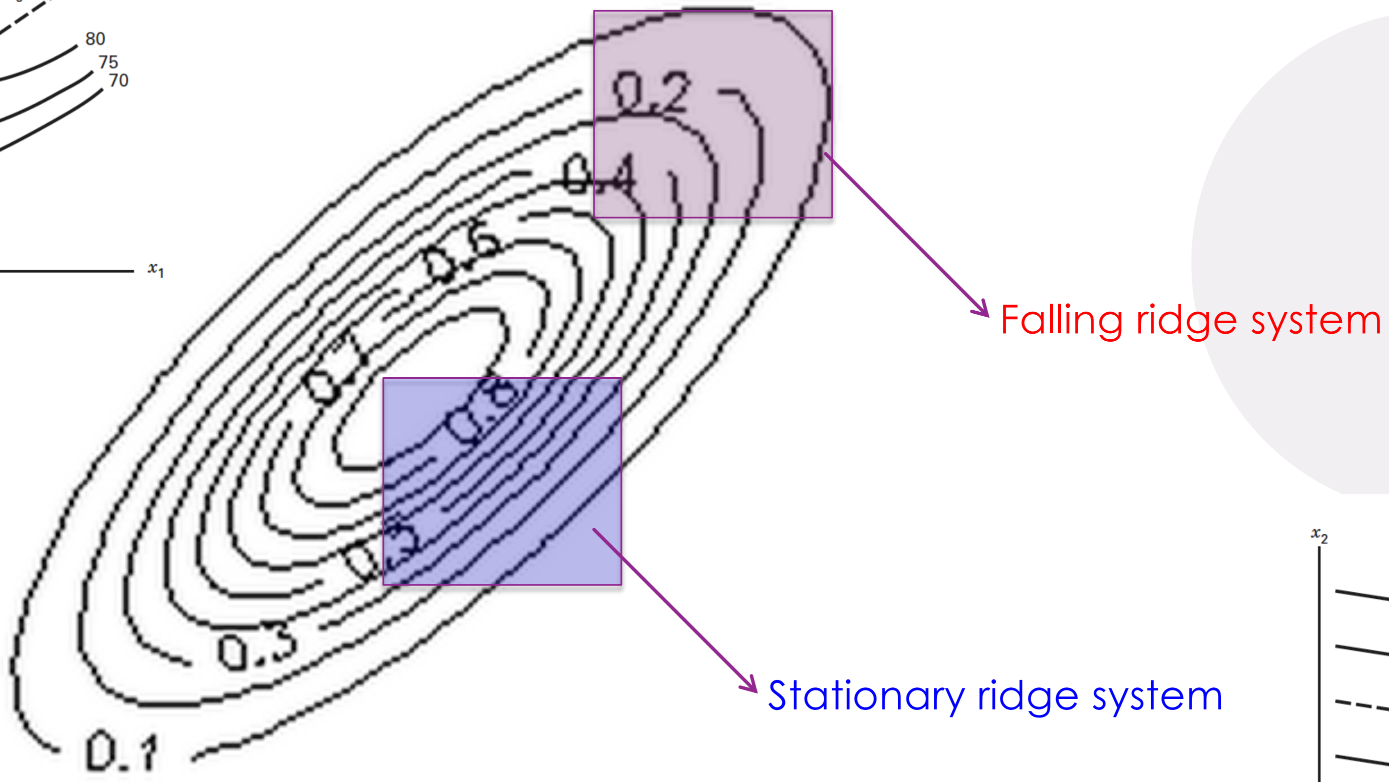
Illustrating of a Saddle Point (or Minimax)

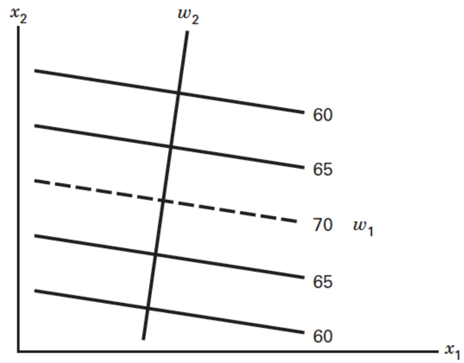


(a) Response surface (b) Contour plot

A Graphical Illustration of Ridge Systems







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

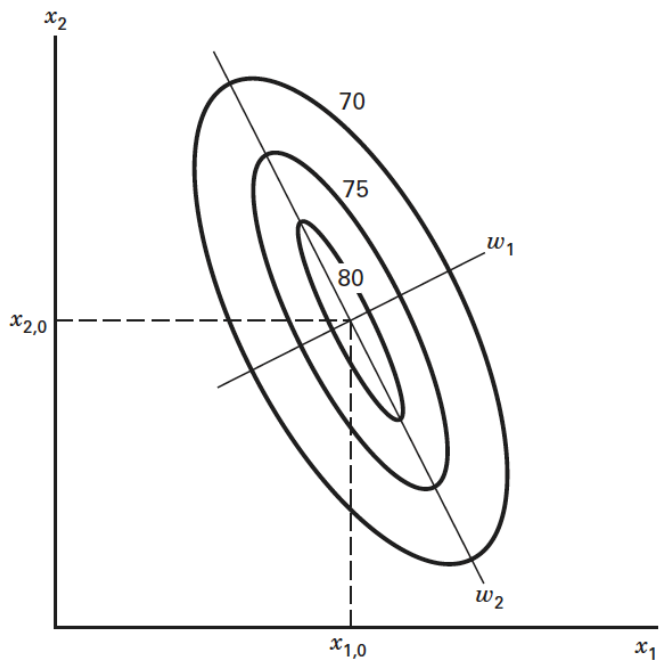
- After fitting a second order model take the partial derivatives with respect to the x_i's and set to zero

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

- 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

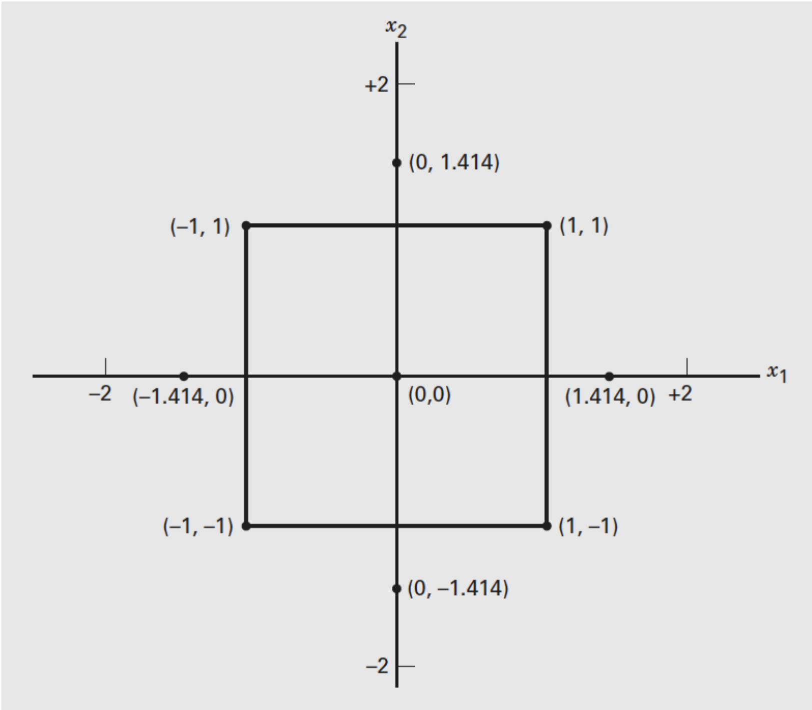
Natural Variables		Coded Variables		Responses		
ξ_1	ξ_2	x_1	x_2	y_1 (Yield)	y_2 (Viscosity)	y_3 (Molecular Weight)
80	170	−1	−1	76.5	62	2940
80	180	−1	1	77.0	60	3470
90	170	1	−1	78.0	66	3680
90	180	1	1	79.5	59	3890
85	175	0	0	79.9	72	3480
85	175	0	0	80.3	69	3200
85	175	0	0	80.0	68	3410
85	175	0	0	79.7	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 $\hat{y}(x) = 80.5 - 0.963w_1^2 - 1.414w_2^2$

Response Surface & Contour Plots

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