Optimal design of experiments

Session 4: Some theory

Peter Goos

Universiteit Antwerpen

Optimal design theory

- continuous or approximate optimal designs
  - implicitly assume an infinitely large number of observations are available
  - is mathematically convenient
- exact or discrete designs
  - finite number of observations
  - fewer theoretical results

Continuous versus exact designs

- continuous
  - $\xi = \begin{Bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_h \\ w_1 & w_2 & \dots & w_h \end{Bmatrix}$
  - $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_h$: design points or support points
  - $w_1, w_2, \dots, w_h$: weights ($w_i \geq 0$, $\sum_i w_i = 1$)
  - h: number of different points

- exact
  - $\xi = \begin{Bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_h \\ n_1 & n_2 & \dots & n_h \end{Bmatrix}$
  - $n_1, n_2, \dots, n_h$: (integer) numbers of observations at $\mathbf{x}_1, \dots, \mathbf{x}_n$
  - $\sum_i n_i = n$
  - h: number of different points

Information matrix

- all criteria to select a design are based on information matrix
- model matrix

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & +1 & +1 & +1 \\ 1 & +1 & -1 & -1 & +1 & +1 \\ 1 & -1 & +1 & -1 & +1 & +1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(\mathbf{x}_1) \\ \mathbf{f}^T(\mathbf{x}_2) \\ \mathbf{f}^T(\mathbf{x}_3) \\ \vdots \\ \mathbf{f}^T(\mathbf{x}_n) \end{bmatrix}$$

$$\begin{matrix} \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ \mathbf{I} & x_1 & x_2 & x_1 x_2 & x_1^2 & x_2^2 \end{matrix}$$

Information matrix

- (total) information matrix

$$\mathbf{M} = \frac{1}{\sigma^2}\mathbf{X}^T\mathbf{X} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)$$

- per observation information matrix

$$\frac{1}{\sigma^2}\mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)$$

Information matrix industrial example

$$\frac{1}{\sigma^2}\left(\mathbf{X}^T\mathbf{X}\right) = \begin{bmatrix} 11 & 0 & 0 & 0 & 6 & 6 \\ 0 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 6 & 0 & 0 & 0 & 6 & 4 \\ 6 & 0 & 0 & 0 & 4 & 6 \end{bmatrix}$$

Information matrix

- exact designs

$$\mathbf{M} = \sum_{i=1}^{h} n_i \mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)$$

where
h = number of different points
n_i = number of replications of point i

- continuous designs

$$\mathbf{M} = \sum_{i=1}^{h} w_i \mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)$$

D-optimality criterion

- seeks designs that minimize variance-covariance matrix of $\hat{\boldsymbol{\beta}}$
- ... that minimize $\left|\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\right|$
- ... that minimize $\left|(\mathbf{X}^T\mathbf{X})^{-1}\right|$
- ... that maximize $\left|\mathbf{X}^T\mathbf{X}\right|$ or $|\mathbf{M}|$
- D-optimal designs minimize
  - generalized variance of $\hat{\boldsymbol{\beta}}$
  - volume of confidence ellipsoid about unknown $\boldsymbol{\beta}$
- “D” stands for determinant

- useful when the interest is only in a subset of the parameters
- useful when intercept or block effects are of no interest
- seeks designs that minimize determinant of variance-covariance matrix corresponding to subset of $\boldsymbol{\beta}$
- $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
  $= \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$
  where $\boldsymbol{\beta}_1$ is the set of parameters of interest

D_s-optimality criterion

- $\mathbf{M} = (\mathbf{X}^T\mathbf{X}) = \begin{bmatrix} \mathbf{X}_1^T\mathbf{X}_1 & \mathbf{X}_1^T\mathbf{X}_2 \\ \mathbf{X}_2^T\mathbf{X}_1 & \mathbf{X}_2^T\mathbf{X}_2 \end{bmatrix}$
- $\text{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$

  part of the variance-covariance matrix corresponding to $\boldsymbol{\beta}_1$

- minimizing $|\mathbf{A}|$ is the same as minimizing

$$\left|\left(\mathbf{X}_1^T\mathbf{X}_1 - \mathbf{X}_1^T\mathbf{X}_2(\mathbf{X}_2^T\mathbf{X}_2)^{-1}\mathbf{X}_2^T\mathbf{X}_1\right)^{-1}\right|$$

and as maximizing

$$\left|\mathbf{X}_1^T\mathbf{X}_1 - \mathbf{X}_1^T\mathbf{X}_2(\mathbf{X}_2^T\mathbf{X}_2)^{-1}\mathbf{X}_2^T\mathbf{X}_1\right|$$

A-optimality criterion

- seeks designs that minimize average variance of parameter estimates
- seeks designs that minimize

$$\sum_{i=1}^{p} \text{var}\left(\hat{\beta}_i\right) / p$$

p = number of model parameters
- seeks designs that minimize

$$\text{trace}\, \sigma^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1}$$

or

$$\text{trace}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$$

V- and G-optimality criteria

- V-optimality (also Q-, IV-, I-optimality)
  - seeks designs that minimize average prediction variance over design region χ
  - ...that minimize

$$\int_{\chi} \sigma^2 \mathbf{f}^T(\mathbf{x})(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})\, d\mathbf{x}$$

- G-optimality
  - seeks designs that minimize maximum prediction variance over design region χ
  - ...that minimize

$$\max_{\chi} \text{var}\{\hat{y}(x)\} = \max_{\chi} \mathbf{f}^T(\mathbf{x})(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})$$

Discussion

- problem: in general, all optimality criteria lead to different designs
- exception: D-optimal continuous designs = G-optimal continuous designs
  - general equivalence theorem
  - prediction variance is maximal in design points
  - maximum = number of model parameters
- what optimality criterion to use?

D-optimality criterion

- most popular criterion
- D-optimal designs are usually quite good w.r.t. other criteria
- D-optimal designs are not affected by linear transformations of levels of the experimental variables
- computational advantages: update formulas

Linear transformations of factor levels

$$
\begin{array}{ccl}
\mathbf{X} & \rightarrow & \mathbf{XA} = \mathbf{Z} \\
\downarrow & & \searrow \\
\max\left|\mathbf{X}^T\mathbf{X}\right| & & \max\left|\mathbf{Z}^T\mathbf{Z}\right| \\
 & & = \left|(\mathbf{XA})^T\mathbf{XA}\right| \\
 & & = \left|\mathbf{A}^T\mathbf{X}^T\mathbf{XA}\right| \\
 & & = \left|\mathbf{A}^T\right|\left|\mathbf{X}^T\mathbf{X}\right||\mathbf{A}| \\
 & & = |\mathbf{A}|^2\left|\mathbf{X}^T\mathbf{X}\right|
\end{array}
$$

Quadratic regression in one variable

- $\chi = [-1, 1]$
- $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$
- D-optimal continuous design

$$
\text{Design 1} = \begin{Bmatrix} -1 & 0 & +1 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}
$$

- information matrix

$$
\mathbf{M}_1 = \begin{bmatrix} 1 & 0 & 2/3 \\ 0 & 2/3 & 0 \\ 2/3 & 0 & 2/3 \end{bmatrix}
$$

Quadratic regression in one variable

- general equivalence theorem implies that D-optimal continuous design is also G-optimal
- D-optimal design minimizes the maximum prediction variance
- maximum is equal to p
- check by plotting prediction variance

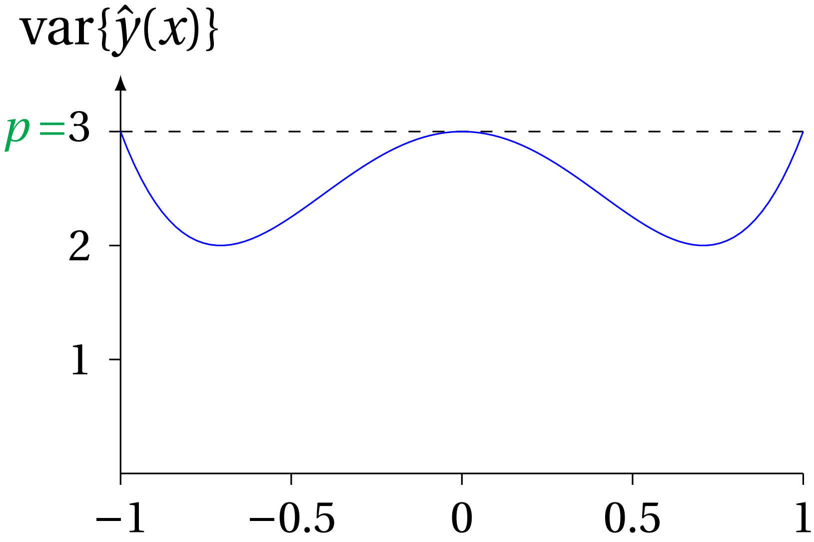
$$\text{var}\{\hat{y}(x)\} = \mathbf{f}^T(x)\mathbf{M}_1^{-1}\mathbf{f}(x)$$

for all $x \in [-1, 1]$

Prediction variance

$$\begin{aligned}
\text{var}\{\hat{y}(x)\} &= \mathbf{f}^T(x)\mathbf{M}_1^{-1}\mathbf{f}(x) \\
&= \begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} 3 & 0 & -3 \\ 0 & \frac{3}{2} & 0 \\ -3 & 0 & \frac{9}{2} \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} \\
&= \begin{bmatrix} 3-3x^2 & \frac{3}{2}x & -3+\frac{9}{2}x^2 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} \\
&= 3 - 3x^2 + \frac{3}{2}x^2 - 3x^2 + \frac{9}{2}x^4 \\
&= 3 - \frac{9}{2}x^2 + \frac{9}{2}x^4
\end{aligned}$$

Prediction variance



Quadratic regression in one variable

- consider other design

$$\text{Design 2} = \begin{Bmatrix} -1 & 0 & +1 \\ 1/4 & 1/2 & 1/4 \end{Bmatrix}$$

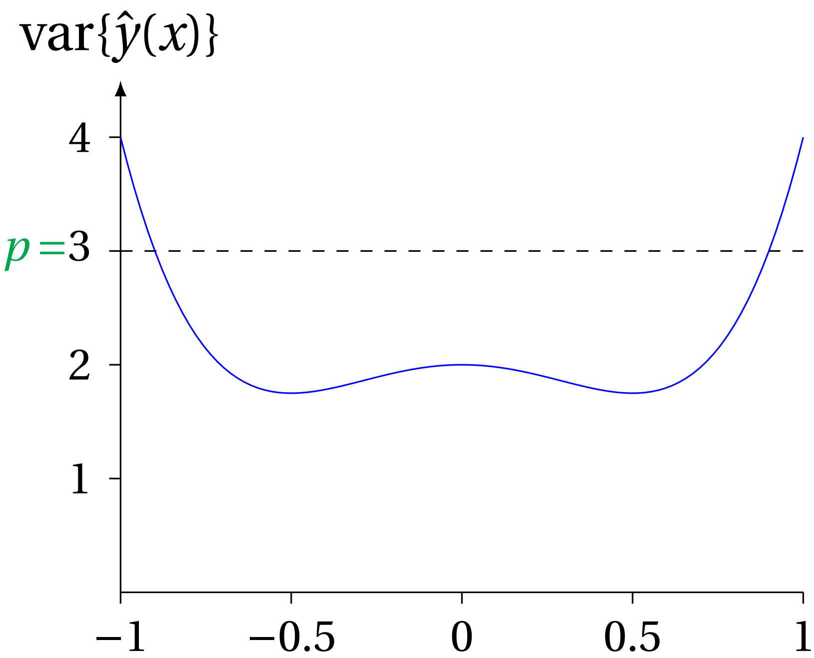
- information matrix

$$\mathbf{M}_2 = \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

Prediction variance

$$\begin{aligned}
\text{var}\{\hat{y}(x)\} &= \mathbf{f}^T(x)\mathbf{M}_2^{-1}\mathbf{f}(x) \\
&= \begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} 2 & 0 & -2 \\ 0 & 2 & 0 \\ -2 & 0 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} \\
&= \begin{bmatrix} 2-2x^2 & 2x & -2+4x^2 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} \\
&= 2-2x^2+2x^2-2x^2+4x^4 \\
&= 2-2x^2+4x^4
\end{aligned}$$

Prediction variance



⇓

Design 2 is not D-optimal

Optimal exact designs

- A-optimal continuous design

$$\xi = \begin{Bmatrix} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{Bmatrix}$$

- what if $n = 4, 8, 12$?
  - multiply weights of continuous design by 4, 8 or 12
  - integer numbers of runs at each point
- what if $n = 5, 6, 7$?
  - multiply weights by 5, 6, or 7
  - non-integer numbers of runs at some of the points
  - not useful in practice

Polynomial regression in one variable

D-optimal design points for the dth order polynomial regression in one variable

d	x_1	x_2	x_3	x_4	x_5	x_6	x_7	weight
1	-1						1	1/2
2	-1			0			1	1/3
3	-1		$-a_3$		a_3		1	1/4
4	-1		$-a_4$	0	a_4		1	1/5
5	-1	$-a_5$	$-b_5$		b_5	a_5	1	1/6
6	-1	$-a_6$	$-b_6$	0	b_6	a_6	1	1/7

$a_3 = \sqrt{1/5}$ $\quad$ $b_5 = \sqrt{(7-2\sqrt{7})/21}$

$a_4 = \sqrt{3/7}$ $\quad$ $a_6 = \sqrt{(15+2\sqrt{15})/33}$

$a_5 = \sqrt{(7+2\sqrt{7})/21}$ $\quad$ $b_6 = \sqrt{(15-2\sqrt{15})/33}$

Quadratic regression in one variable

- Design 1 has covariance matrix

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 3 & 0 & -3 \\ 0 & 3/2 & 0 \\ -3 & 0 & 9/2 \end{bmatrix}$$

  with $\mathrm{trace}(\mathbf{X}^T\mathbf{X})^{-1} = 3 + 3/2 + 9/2 = 9$
- Design 2 has covariance matrix

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 2 & 0 & -2 \\ 0 & 2 & 0 \\ -2 & 0 & 4 \end{bmatrix}$$

  with $\mathrm{trace}(\mathbf{X}^T\mathbf{X})^{-1} = 2 + 2 + 4 = 8$
- Design 2 is A-optimal

2^3 Factorial design

- main-effects model
  $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$
- model matrix

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & +1 & -1 & -1 \\ 1 & -1 & +1 & -1 \\ 1 & +1 & +1 & -1 \\ 1 & -1 & +1 & +1 \\ 1 & +1 & +1 & +1 \\ 1 & -1 & -1 & +1 \\ 1 & +1 & -1 & +1 \end{bmatrix}$$

2^3 factorial design

- (total) information matrix

$$\mathbf{M} = \mathbf{X}^T\mathbf{X} = \begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$$

- per observation information matrix

$$\mathbf{M}^* = \frac{1}{n}(\mathbf{X}^T\mathbf{X}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Prediction variance

- $\text{var}\{\hat{y}(x)\} = \mathbf{f}^T(\mathbf{x})\left\{\mathbf{M}^*\right\}^{-1}\mathbf{f}(\mathbf{x})$

$$= \begin{bmatrix} 1 & x_1 & x_2 & x_3 \end{bmatrix} \left\{\mathbf{M}^*\right\}^{-1} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= 1 + x_1^2 + x_2^2 + x_3^2$$

- if $\chi = [-1, 1]^3$, then this is maximal when

$$x_1 = \pm 1, \quad x_2 = \pm 1, \quad x_3 = \pm 1$$

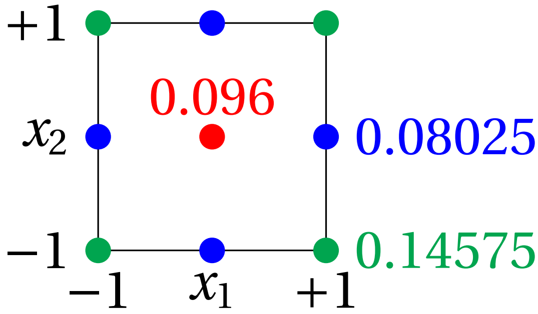
- maximum $= 4 = p$
- general equivalence theorem is satisfied
- 2^3 factorial design is D- and G-optimal

More

- Plackett-Burman designs: optimal for main-effects model
- factorial designs: optimal for main-effects-plus-interactions models
- some remarks
  - off-diagonal elements of information matrix often zero
    (impossible when quadratic terms)
  - optimal designs are often symmetric
  - these properties are more difficult to achieve for exact designs

Quadratic regression in two variables

- $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \epsilon$
- D-optimal continuous design

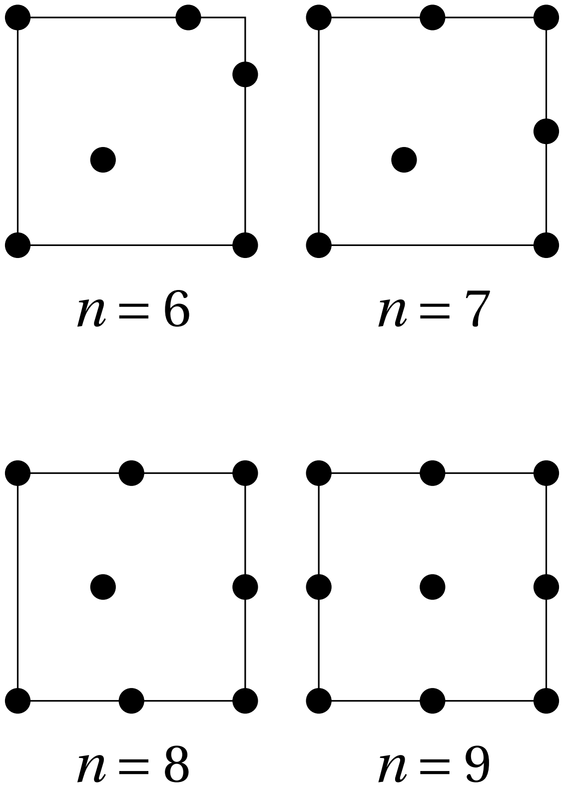


- D-optimal discrete designs
  - $n = 9$: one observation at each point of the continuous D-optimal design
  - $n = 6$: D-optimal exact design does not resemble D-optimal continuous one

Information matrix D-optimal continuous design

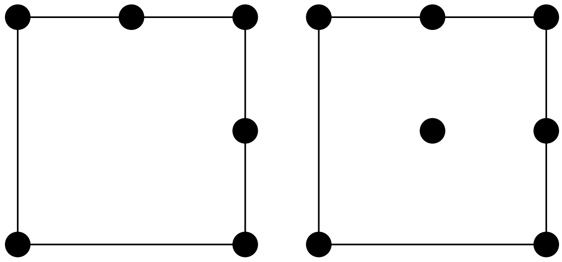
$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0.74 & 0.74 \\ 0 & 0.74 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.74 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.58 & 0 & 0 \\ 0.74 & 0 & 0 & 0 & 0.74 & 0.58 \\ 0.74 & 0 & 0 & 0 & 0.58 & 0.74 \end{bmatrix}$$

Quadratic regression in two variables

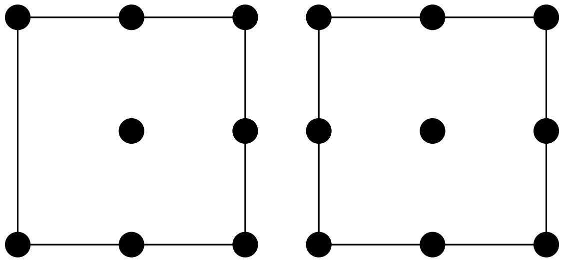


D-optimal exact designs

Quadratic design in two variables



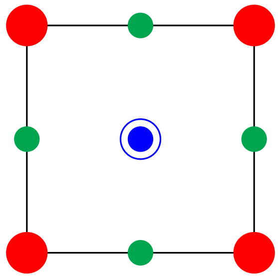
$n = 6$ $n = 7$



$n = 8$ $n = 9$

D-optimal exact three-level designs

Quadratic regression in 2 variables



1 run, 2 runs, 3 runs

D-optimal exact design

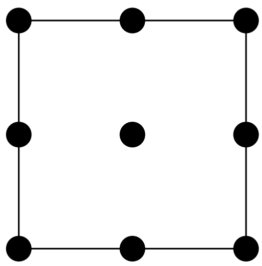
$n = 18$

An industrial example

- experiment to investigate effect of
  - amount of glycerine (%), x_1 $(1 \le x_1 \le 3)$
  - speed temperature reduction (°F/min), x_2 $(1 \le x_2 \le 3)$
- response: amount of surviving biological material (%), y
- context: freeze-dried coffee
  retaining volatile compounds in freeze-dried coffee is important for its smell and taste
- Excel file: `quadratic3.xls`

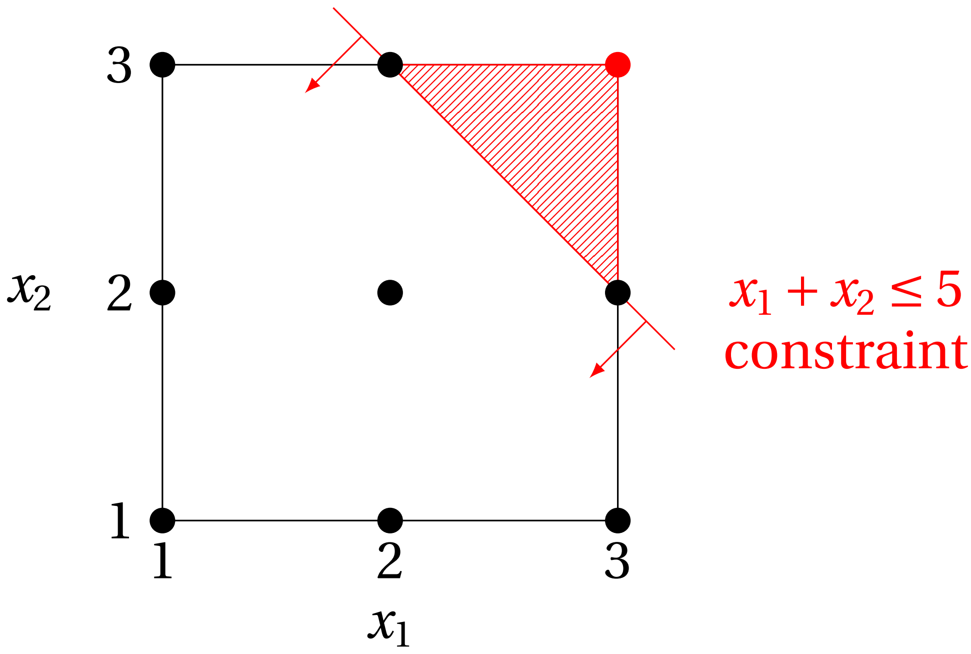
An industrial example

- $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$
  $+\beta_{12} x_{1i} x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \epsilon$
- 9 observations / tests
- 3^2 factorial design is optimal



- what if combination of high x_1 and high x_2 are not allowed?

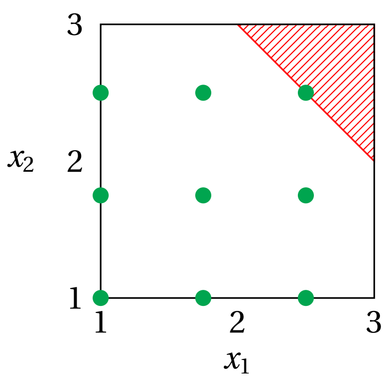
Constrained design region



x_1: glycerine, x_2: temperature reduction

Constrained design region: solution I

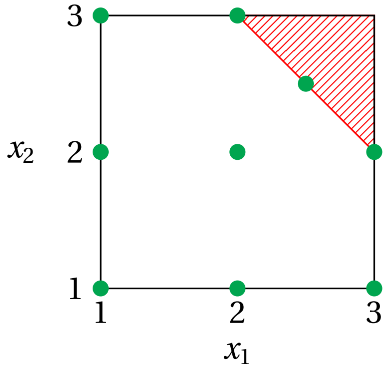
scale 3^2 factorial design down so that it fits in constrained design region



$$\det(\mathbf{X}^T\mathbf{X})^{-1} = 0.0192$$

Constrained design region: solution II

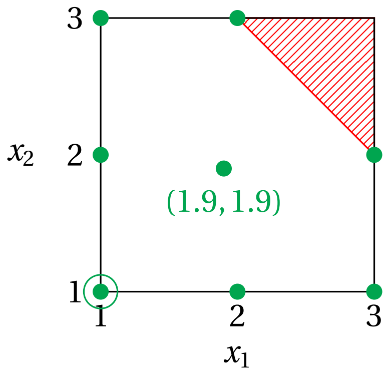
move forbidden point $(3,3)$ inward



$$\det(\mathbf{X}^T\mathbf{X})^{-1} = 0.0006$$

Constrained design region: solution III

use optimal design approach



$$\det(\mathbf{X}^T\mathbf{X})^{-1} = 0.0005$$