TTK31 - Design of Experiments (DoE), metamodelling and Quality by Design (QbD)
Autumn 2021

Big Data Cybernetics Gang

NTNU

Optimal designs and mixtures (designs with constraints)

Introduction to constrained designs

- Optimal designs
- Mixture designs
- Two examples:
  - Example 1, baking a cake: with long time and high temperature the cake is burned
  - Example 2, cocktail mix: the sum of ingredients is 100% (lime juice, lemon juice, tonic water)

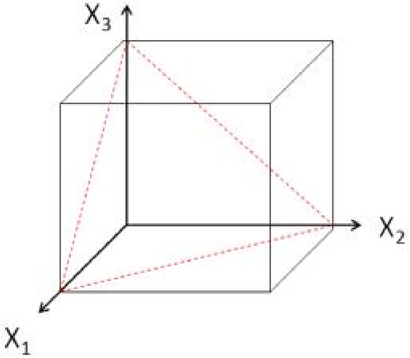
Introduction to constrained designs

- Optimal designs
- Mixture designs
- Two examples:
  - Example 1, baking a cake: with long time and high temperature the cake is burned
  - Example 2, cocktail mix: the sum of ingredients is 100% (lime juice, lemon juice, tonic water)

  Problem: the design variables involved cannot vary completely independently from the others:
  - In case 1, the maximum allowed temperature will depend on the baking time
  - In case 2, the proportions of the cocktail must add up to 100%

Mixture designs

- 3 dimensions collapsed into 2
- For a mixture of 3 ingredients the experimental region becomes flat
- This shape is called a simplex
- The simplex region contains all possible combinations



As there is closure in the design this must be handled in the ANOVA, with so-called Scheffé polynomials

Scheffé polynomials

First order:

$$Y = \sum_{i=1}^{q} \beta_i X_i$$

Second order:

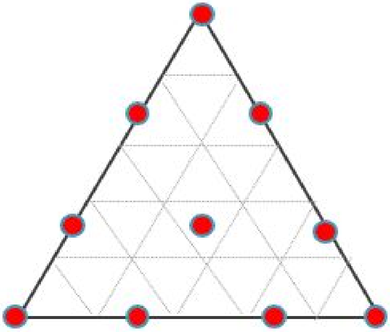
$$Y = \sum_{i=1}^{q} \beta_i X_i + \sum_{i=1}^{q} \sum_{j=i+1}^{q} \beta_{ij} X_i X_j$$

Special cubic:

$$Y = \sum_{i=1}^{q} \beta_i X_i + \sum_{i=1}^{q} \sum_{j=i+1}^{q} \beta_{ij} X_i X_j + \sum^{q} \sum_{i<j<k}^{q} \sum^{q} \beta_{ijk} X_i X_j X_k$$

Simplex Lattice designs

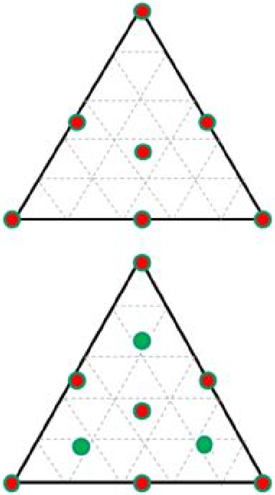
- The Simplex Lattice design has the form $[q,m]$ where q is the number of mixture components and m is the order of the design to be supported.
- Excellent designs for investigating the extremes of a design space and for building prediction models



The [3,3] Simplex Lattice Design

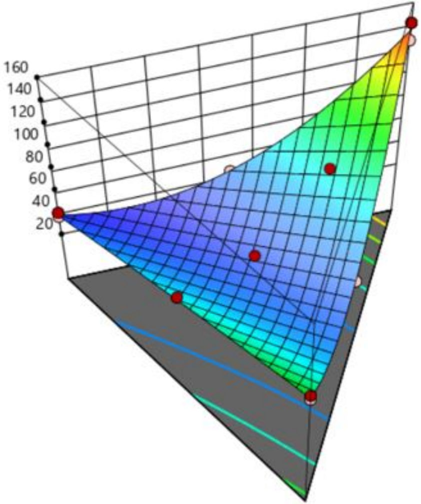
Simplex centroid designs

- In a q component simplex, the number of distinct points is $2q$-1
- Simple design for investigating the entire region economically.
- May be further augmented with axial check blends

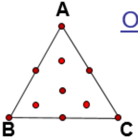


Simplex response surface

- 3D-representation of the mixture triangle
- The three coordinates sum to 100%

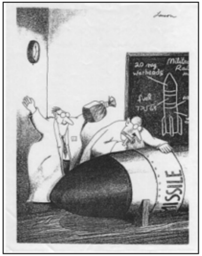


Mixture design - Rocket fuel



Objective: Maximize elasticity of solid rocket fuel.

Design: Augmented Simplex Lattice.



	Fuel 1	Fuel 2	Fuel 3	Elasticity	
Pure Mixtures	1	0	0	323	328
	0	1	0	298	430
	0	0	1	674	737
Binary Blends	1/2	1/2	0	210	307
	1/2	0	1/2	657	816
	0	1/2	1/2	940	850
Centroid	1/3	1/3	1/3	888	1068
Check Blends	2/3	1/6	1/6	713	711
	1/6	2/3	1/6	808	739
	1/6	1/6	2/3	1012	972

Optimal designs - Optimality criteria I

There exists a number of different criteria to replace the initial set of points to find the best subset

- I-optimal
  - The I-optimal criteria is recommended to build response surface designs where the goal is to optimize the factor settings, requiring greater precision in the estimated model.
  - Gives a good distribution inside the design space
- D-optimal
  - A D-optimal algorithm chooses runs that minimizes the determinant of the variance-covariance matrix
  - This minimizes the errors in the regression coefficients
  - The D-optimal criteria is recommended to build factorial designs where the goal is to find factors important to the process

Optimal designs - Optimality criteria II

- A-optimal
  - An A-optimal design minimizes the trace of the variance-covariance matrix.
  - This has the effect of minimizing the average prediction variance of the polynomial model coefficients.
- Modified Distance
  - The modified distance-based point selection algorithm selects model points to obtain a maximum spread throughout the design region while ensuring that adequate runs are chosen to fit the polynomial model
  - Thus, this gives the "best of both worlds"
- Distance (Maximin)
  - This design scatter points throughout the experimental region
  - The algorithm starts from a vertex and then adds points that maximize the minimum Euclidean distance from points already in the design

Optimal exchange methods

Two algorithms are frequently used to decide on the best optimal design. The starting point is a pseudo-random set of points.

- Coordinate exchange
- Point exchange
- The methods require a set of initial candidate points:
  - Vertices are at the extreme limits of the design
  - Center of edges are half-way between two vertices
  - Thirds of edges are one-third and two-thirds the way along an edge between two vertices
  - Constraint plane centroids are the center of a hyper-plane defined by three or more coplanar vertices
  - Axial check points are half-way between the overall centroid and the vertices
  - Interior points are half-way between the centroid and center of edges
  - Overall centroid is the geometric center-of-mass of the design

Coordinate exchange algorithm

1. Select a random initial set of p points, where p is the number of terms in the designed for model
   - Start with a random coordinate (point) within the design space
   - Randomly pick each subsequent design point and evaluate if it increases the rank of the matrix. Continue this process until a full rank matrix is obtained.
2. Randomly select any extra model points
3. Start the coordinate exchange algorithm
   - Calculate the current optimality criterion (OC)
   - Starting with the first point, move it along a set of directions in incremental steps
   - If the OC improves, change the point and move on to the next point in the list. If the OC does not improve, retain the point and move on to the next point
   - Repeat the algorithm several times to improve the odds of finding the globally optimal design
4. Lack-of-fit points and replicates are chosen that best support the optimality criterion

Point exchange

1. Define a candidate set of possible factor combinations
2. Select a random initial set of p points from the above candidate set where p is the number of terms as specified in the model
   - Start with a random point from the candidate set
   - Randomly pick each subsequent design point. If a new point increases the rank of the matrix it is added to the set of points
   - Randomly select any extra model points
3. Perform exchange steps by exchanging first 1 point, then continue up to exchanging 10 points until there is no improvement

Evaluation of the designs

- One important property of the design is the condition number
- It is calculated as the ratio between the largest and smallest eigenvalue
- It measures the multicollinearity of the design
- An orthogonal factorial design has condition number of 1
- There's some guidelines of the size of the condition number w.r.t. the subsequent analysis:
    - $1 < 100$ Multicollinearity does not pose a problem
    - $100 < 1000$ Moderate to strong multicollinearity
    - > 1000 Severe multicollinearity

Example: Mixture design

- Objective: Make a system for testing air condition systems in cars
- Technology: Multivariate calibration with six gas sensors that are partly selective for the individual gases
- Gases and their concentration ranges (anonymized):

Component	Name	Minimum	Maximum
A	G1	75	89
B	Air	0	15
C	Propane	0	10
D	G2	0	5
E	G3	0	6
F	G4	0	6

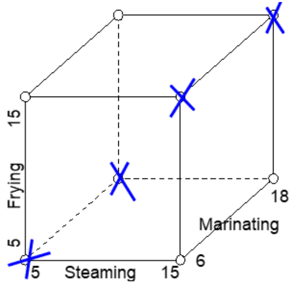
Example of a constrained situation - I

- Optimizing the quality of cooked Meat
  - Design variables: Marinating Time, Steaming Time, Frying Time
  - Responses: Sensory measurements
- Original idea: Full Factorial design
- 8 experiments combining the low and high levels of the variables

Sample	Marinating	Steaming	Frying
1	6	5	5
2	18	5	5
3	6	15	5
4	18	15	5
5	6	5	15
6	18	5	15
7	6	15	15
8	18	15	15

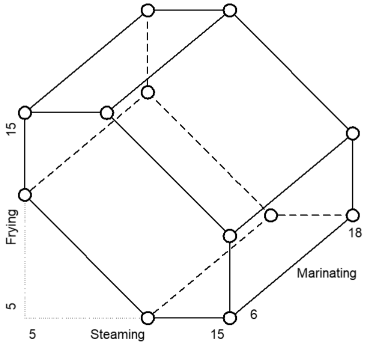
Example of a constrained situation - II

- Extreme combinations should be avoided:
  - Steaming + Frying $< 16 \implies$ raw meat
  - Steaming + Frying $> 24 \implies$ overcooked
- Full factorial does not apply:
  - 4 out of 8 cube samples are excluded
  - The remaining 4 are not enough to explore the region of interest



Visualizing the constrained regions

- Multi-linear constraints:
  - Steaming + Frying $\geq$ 16 min
  - Steaming + Frying $\leq$ 24 min
- The experimental region becomes a polyhedron
- Orthogonal designs are no longer possible
- $\Longrightarrow$ optimal design



How to define constraints

- Multi-factor constraints must be entered as an equation
- Constraints points (CPs) need to be defined
- The equation describes the boundaries for the experimental region
- The following equation applies:

$$1 \leq \frac{A - LL_A}{CP_A - LL_A} + \frac{B - LL_B}{CP_B - LL_B}$$

A small example of setting constraints

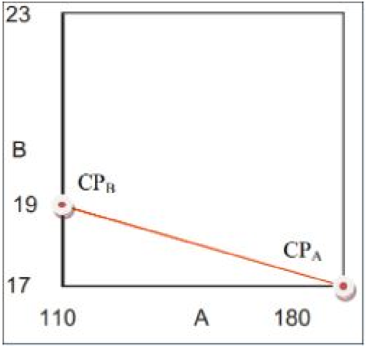
- Assume to design factors, A and B
- When A is at low level, B must be higher than 19

In this case:

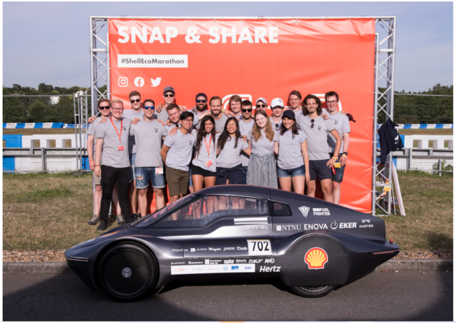
$$1 \leq \frac{A - 110}{180 - 110} + \frac{B - 17}{19 - 17}$$

which leads to:

$$775 \leq A + 35B$$



Fuel fighter - the NTNU team



- Problem: What influences the vertical acceleration on the front and rear of the car when driving over speed bumps?
- ... and how to find the optimal settings given a maximum value for the acceleration?
- Design factors: Height of bump, center of gravity, weight and speed
- Will most likely be presented as an exercise in two weeks' time!