Input-Output Space Exploration

Kennedy Putra Kusumo

January 2022

1 Introduction

The purpose of the text is to document the investigations and findings that were found in developing an experimental design technique in a specific context that is most relevant in pharmaceutical manufacturing.

The experimental design technique is a model-based one. Meaning that a mathematical model is involved during experimental design and is also the ultimate goal of running the experiments (e.g., to calibrate a model).

The basic idea is that there are two competing objectives associated with two sets of parameters/variables that are of interest, the so-called input and output space. The input space is defined as a set of variables which a practitioner directly (and/or conveniently) manipulate. The output space comprise of variables which the practitioners measure and interested in, but are indirectly manipulated through the input variables. As such, we shall refer to them simply as the input and output objectives.

The input objective is associated with

2 Output Space Exploration Objective

We present alternative criteria for operable space maximization with design of experiments, as introduced by Chen et al. [2018].

2.1 Criteria Which Searches Over Controls

The goal is to find the locations of M points of inputs/controls of dimension n_x $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^{M \times n_x}$ that minimize or maximize a certain metric of the associated M points of the outputs of dimension n_y $\mathbf{y} \in \mathcal{Y} \subset \mathbb{R}^{M \times n_y}$. We assume a model (mapping) $f : \mathcal{X} \to \mathcal{Y}$ is available. The following is a list of formulations populated.

2.1.1 Maximal Spread Design

Spread is defined as the minimal distance between the chosen points. A formulation is as follows:

$$\max_{\mathbf{x}} \eta \tag{1}$$

$$\text{s.t. } \eta \le \|\boldsymbol{y}_i - \boldsymbol{y}_j\|, \quad i = 1 \dots M,\ j = i+1 \tag{2}$$

$$\boldsymbol{y}_i = f(\boldsymbol{x}_i), \quad i = 1 \dots M \tag{3}$$

2.2 Criteria Which Involves Discretization of Controls

We assume that a finite set of candidate points $\mathbf{p}_i \in \mathbb{R}^n,\ i = 1 \dots N$ — possibly rescaled in the unit hypercube $[0,1]^n$. The goal is to select a subset of M points that minimize or maximize a certain metric. The following is a list of formulations and metrics populated.

2.2.1 Maximal Orthogonality Design

This criterion maximizes the *orthogonality* of the chosen points. It involves a

2.2.2 Maximal Covering Design

This criterion selects M centroids among the N candidates in such a way that the maximal distance between any candidate point to their closest centroid is minimized. An (pure) integer programming formulation for this criterion is the following.

$$\min_{\mathbf{y},\mathbf{z},\eta} \eta \tag{4}$$

$$\text{s.t. } \eta \ge \sum_{i=1}^{N} z_{i,j} \|\mathbf{p}_i - \mathbf{p}_j\|, \quad j = 1 \dots N \tag{5}$$

$$1 = \sum_{i=1}^{N} z_{i,j}, \quad j = 1 \dots N \tag{6}$$

$$y_i \ge z_{i,j}, \forall i, j = 1 \dots N \tag{7}$$

$$M \ge \sum_{i=1}^{N} y_i \tag{8}$$

$$y_i, z_{i,j} \in \{0,1\}, \quad i, j = 1 \dots N \tag{9}$$

where the selected points $\mathbf{p}_i$ are those with $y_i = 1$. $z_{i,j} = 1$ indicates the closest centroid $\mathbf{p}_i$ to point $\mathbf{p}_j$. This problem is also known as the uncapacitated vertex k-center problem [Cornejo Acosta et al., 2020], for which heuristics algorithms exist that could be used to warm-start the MIP solution.

2.2.3 Maximal Spread Design

Spread is defined as the minimal distance between the chosen points. M points amongst the N candidates are chosen to maximize spread. A pure integer programming formulation is:

$$\max_{\mathbf{y},\eta} \eta \tag{10}$$

$$\text{s.t. } \eta \leq \|\mathbf{p}_i - \mathbf{p}_j\| + (1 - y_i)D + (1 - y_j)D, \quad i, j = 1 \ldots N \tag{11}$$

$$M \geq \sum_{i=1}^{N} y_i \tag{12}$$

$$y_i \in \{0, 1\}, \quad i = 1 \ldots N \tag{13}$$

where the selected points p_i are those with $y_i = 1$. The scalar D is chosen and should be no smaller than the *diameter* of the set of candidates points. A recommended value is $D = \max_{i,j} \|\mathbf{p}_i - \mathbf{p}_j\|$.

3 Numerical Statistics of Different Alternatives

References

Qi Chen, Remigijus Paulavicius, Claire S. Adjiman, and Salvador Garcia-Munoz. An optimization framework to combine operable space maximization with design of experiments. *AIChE Journal*, 64(11):3944–3957, nov 2018. ISSN 00011541. doi: 10.1002/aic.16214.

José Alejandro Cornejo Acosta, Jesús García Díaz, Ricardo Menchaca-Méndez, and Rolando Menchaca-Méndez. Solving the capacitated vertex k-center problem through the minimum capacitated dominating set problem. *Mathematics*, 8(9), 2020. ISSN 2227-7390. doi: 10.3390/math8091551.

Solver	Grid Size	# of Trials	Computational Time (s)
	Maximal Covering (MIP)		
CPLEX	25	3	0.78
CPLEX	36	3	1.35
CPLEX	49	3	2.73
CPLEX	64	3	5.11
CPLEX	81	3	8.73
CPLEX	100	3	17.12
CPLEX	121	3	27.43
CPLEX	441	3	
GUROBI	25	3	
GUROBI	36	3	
GUROBI	49	3	
GUROBI	64	3	
GUROBI	81	3	
GUROBI	100	3	
GUROBI	121	3	
GUROBI	441	3	
	Maximal Spread (MIP)		
CPLEX	25	3	0.59
CPLEX	36	3	0.91
CPLEX	49	3	1.72
CPLEX	64	3	3.12
CPLEX	81	3	4.89
CPLEX	100	3	7.44
CPLEX	121	3	11.23
CPLEX	441	3	277.75
GUROBI	25	3	0.45
GUROBI	36	3	0.94
GUROBI	49	3	1.86
GUROBI	64	3	3.16
GUROBI	81	3	5.05
GUROBI	100	3	7.85
GUROBI	121	3	13.47
GUROBI	441	3	359.30
Orthogonality			
Convex Hull Volume			
Ellipsoid Approximation			

Table 1: Caption