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

1.1 Output Space Exploration Objective

We assume a discretized formulation where a discrete set of samples in the input space are generated through a sampling procedure and the model has been used to generate the mapped locations of these samples in the output space.

A key feature of the optimization problem is the presence of the binary variables which defined whether a point sample is chosen as part of the final solution, let's call them p with an index i to refer to the point samples. Thus, notations like $p_1 = 1$ means that point sample number 1 is part of the final design and $p_1 = 0$ means it is NOT chosen.

1.1.1 Centroid Formulation

We first compute the centroid of the chosen points. Let's assume a simple 2D example, and let's call the centroid C with two elements C_x and C_y for the x and y axes, respectively. Let x_i and y_i be the x-, and y-coordinates of the i-th point sample.

$$C_x = \frac{\sum_i p_i x_i}{\sum_i p_i} \tag{1}$$

$$C_y = \frac{\sum_i p_i y_i}{\sum_i p_i} \tag{2}$$

$$\tag{3}$$

The centroid helps us compute the "volume" or "spread" (V) of the points in the output space, which we define in this subsection as follows.

$$V = \sum_i \left(p_i \left((x_i - C_x)^2 + (y_i - C_y)^2 \right) \right) \tag{4}$$

Notice that the p_i appears as cubed terms when expanded out. Using sympy also confirms that the p_i terms come as cubed terms when expanded out, while x_i and y_i terms come as quadratic terms.

1.1.2 Centroid of the Feasible Space Formulation

An alternative and similar formulation is to precompute an appropriate replacement for the centroid to simply the expression from before. The idea is very simple, we compute the centroid of the precomputed response points first. This will mean that we will use the centroid of the feasible space in place of the centroid of the actual computed selected points.

In this formulation, because the centroid is precomputed, the p_i terms appear in order 1 in the final optimization problem. It takes exactly the same expression as (4) but the centroids are given parameters.

A problem though, is that the optimization will simply yield repeated points. It will simply determine the point which is furthest away (euclidean distance) and repeat it for all the selected experimental points. There is simply no incentive for the points to not be repeated; unlike the previous formulation where centroid of the selected points are used.

1.1.3 Augmented HR Criterion

Another possible criterion formulation is

2 Numerical Statistics of Different Alternatives

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