Generalizing the Equation of Defining the Boundary of a Constrained Region for Three Factors and the Preferred Design

By Alyaa Roshdy Zahran[1]

Statistics Department, Faculty of Economics and Political Sciences, Cairo University, Cairo, Egypt

Abstract

Economical, practical, or physical constraints sometimes prevent the factor space of a designed experiment from being a regular p-dimensional hypercube or hypersphere. In this case, standard designs may not be the best choice; and it is desirable to be able to find best designs under these restrictions. This paper extends the work of Zahran (2004) using a more general equation to define the boundary of a modified 2^3 factorial design. The paper considers optimality criteria and the Fraction of design space criterion for a preferred design for this restricted design space..

Keywords: alphabetical optimality criteria, linear models, non-regular design space, fraction of design graph.

[1] *Currently at Department of Mathematics and Physics, College of Arts and Sciences, Qatar Univ. Email: azahran@qu.edu.qa, azahran@vt.edu*

INTRODUCTION

Hypercube or hypersphere operability regions are typical in many designed experiments. Under these regions and the use of first order with/without interaction linear models, factorial designs are the standard designs. These designs have desirable properties including simplicity, straightforward implementation, orthogonality and D- and Q-optimality. Q-optimality is also called V- or IV-optimality in the literature, see Draper and St. John (1977). For hypercube regions, they are also G optimal designs. For more details on the definition and characteristics of these alphabet criteria are given in Myers and Montgomery (2002). However, if some constraints may occur on the factor settings and, hence on the design region (operability region), standard designs may not be feasible. Kennard and Stone (1969) were the first ones in the literature to discuss the problem of irregular experimental regions and suggested computer aided design for selecting an experimental plan. Some case-by-case examples of non-standard design regions are discussed in Snee (1985). Johnson and Nachtsheim (1983) discussed how single-point augmentation procedures are helpful for finding exact D-optimal Designs on Convex Design spaces. Recognizing the importance of computer programs to develop designs when classical designs are not appropriate, Nachtsheim (1987) reviewed and compared the available tools for computer-aided design of experiments. Atkinson and Donev (1992) devoted a short chapter to restricted designs. They used some computer algorithms to find the D-optimum design for certain irregular regions. They emphasize that whatever the shape of the experimental region the principles of the optimality theory remain the same. Montgomery, Loredo, Jearkpaporn, and Testik (2002) gave a brief tutorial on computer-aided methods for constructing designs for irregularly shaped regions.

For the hypercube regions, there are some situations where one combination of the factors (one corner of the cube design space) may not be feasible. For example, in a drug interaction study, where it might be not practical to simultaneously set the factors at high (or low) levels, because it is known a priori that this combination has an undesirable effect. Putting all the factors at the high level might be dangerously potent for the subject, while it may be unethical not to give the subject any effective drug by giving the combination that includes all factors at the low level. Under this scenario and for the two factor setup, Zahran, Anderson-Cook, Myers and Smith (2003) replaced the high-high corner with a quarter of a circle to define the boundary of the restricted region, then different designs are proposed and their efficiencies are studied. Zahran and Anderson-Cook (2003) generalized the defining equation of the constrained region boundary. For the three factor setup, Zahran (2004) replaced the high-high-high corner with a section of a sphere, offered three sensible designs and studied their efficiencies under two different models.

The current paper proposes a general equation to define the boundary of the constrained region for the three factors. With this general form, the practitioner is more flexible to define the most sensible design space for the particular problem under consideration. The appropriate form of the boundary could be chosen according to the available prior information about constrained feasible points or the importance of the interaction terms.

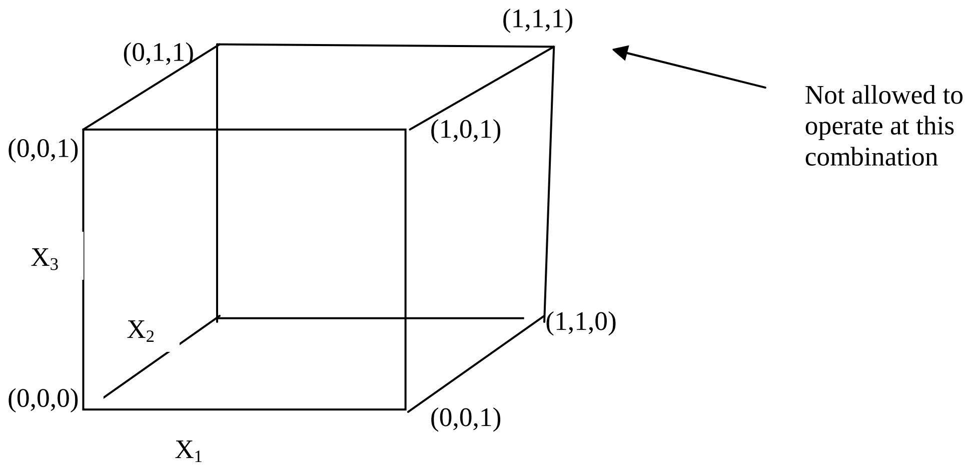
Although, we have considered the problem of excluding the high-high-high combination, all the results hold for excluding any single corner combination of the three factors.

Next section describes briefly the restricted region and optimal design.

RESTRICTED DESIGN SPACE

For the quantitative three-factor case, the standard experimental region is the cube. Consider for some reason the high-high-high (H-H-H) combination of the factors is banned. Possible reasons include impracticability, danger, ethics, or cost. Figure 1 represents the operability region under this scenario. Under this constrained region, the standard factorial is not feasible, and the need for specialized optimal designs arises. Zahran (2004) discussed three sensible designs for this situation. These designs involve changing the factor level combinations to alter the standard factorial design to fit inside the feasible design region. To define the boundary for this case, Zahran (2004) excluded a cube of side length r from the restricted H-H-H corner and replaced it with a portion of a sphere of radius $\rho = r\sqrt{2}$ centered at $(1-r, 1-r, 1-r)$, where r is the fraction of the range of each variable that one wishes to alter. The user would specify what value of r is required to make the design space feasible and of practical interest. Different design regions are obtained by specifying different r values. Once the region is specified, a best design can be selected under the model considered.

Figure 1: Restricted Operability Region



In the following section the first order model with two way interactions in the tree-factor case and the preferred Design are briefly summarized from Zahran (2004).

THE MODEL AND THE PREFERRED DESIGN

The first order model with two way interactions in the three-factor case is under consideration. Equation 1 shows the functional form of the model.

$$y = \beta_0 + \sum_{i=1}^{3} \beta_i x_i + \sum_{i=1}^{3} \sum_{i<j}^{3} \beta_{ij} x_i x_j + \varepsilon \qquad (1)$$

with $\varepsilon \overset{iid}{\sim} N(0, \sigma^2)$. This model would be the more common choice in response surface modeling.

For this model, Zahran (2004) found that the preferred design using the various optimality criteria (D-, G- and Q-optimality) is obtained by replacing the high-high-high point by the point, say (*a*, *a* ,*a*), where $a = 1 - r + \rho / \sqrt{3}$ in the $[0,1]^3$ scale. This is a point in the middle of the replacing portion of the sphere at $\theta = 45°, \phi = 54.7356°$. These two angles are depicted in Figure 2a. The value of *a* depends on the radius of the sphere chosen. This resulted design fills the design space and is often only a minor adjustment from the standard design (see Figure 2b).

Figure 2a: The Definition of ϕ and θ

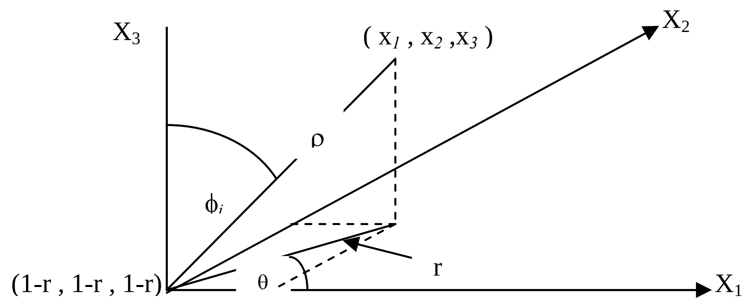
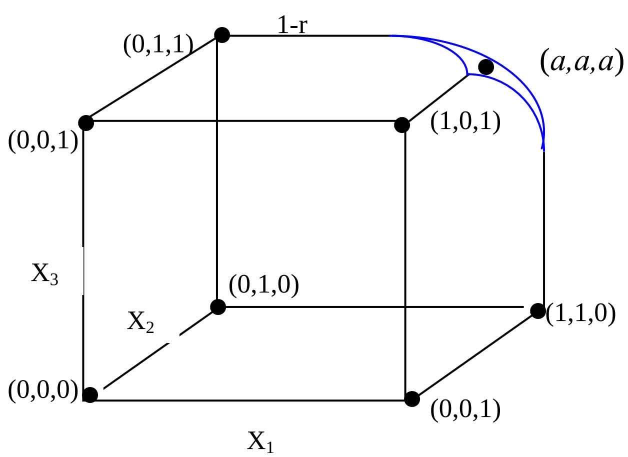


Figure 2b: Design I



Note that although the user will specify r in the [0,1] range, this value needs to be transformed to the $[-1,1]^3$ scale, since the more standard operability region is usually defined in to the $[-1,1]^3$ scale. As a result, a needs to be transformed also to the $[-1,1]^3$scale, say a^*, where $a^* = 2a - 1$.

In the following section we study this design in a more general defined design space.

GENERAL DESIGN SPACE AND THE PREFERRED DESIGN

The optimality of the above design for all three alphabetical criteria encourages us to check its optimality for other related design spaces. In Zahran (2004), when replacing the high-high-high corner of the cube with a portion of a sphere of radius *r,* the equation in effect was $(x_1 - x_{1_c})^2 + (x_2 - x_{2_c})^2 + (x_3 - x_{3_c})^2 = r^2$, where ($x_{1c}$, x_{2c}, x_{3c}) = (1-*r*,1-*r*,1-*r*) is the center of the replacement sphere. A more general equation that gives flexibility of the design space is

$$| x_1 - x_{1_c} |^d + | x_2 - x_{2_c} |^d + | x_3 - x_{3_c} |^d = r^d \quad \text{for } d > 0. \tag{2}$$

Figure 3 shows the effect of changing the value of *d* on the design space for a fixed *r* value. For *d*=1, a plane is truncating the corner. To truncate less of the square design space, one could use increasing powers of *d*, which gives a convex corner in the adjusted corner (Figure 3b). A value of *d* less than one results in a truncated design space with a concave corner (Figure 3c).

Figure 3a: Operability Region for d=1

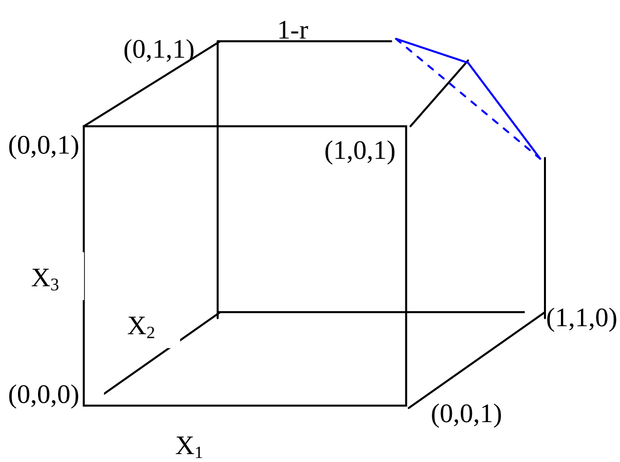


Figure 3b: Operability Region for d=2

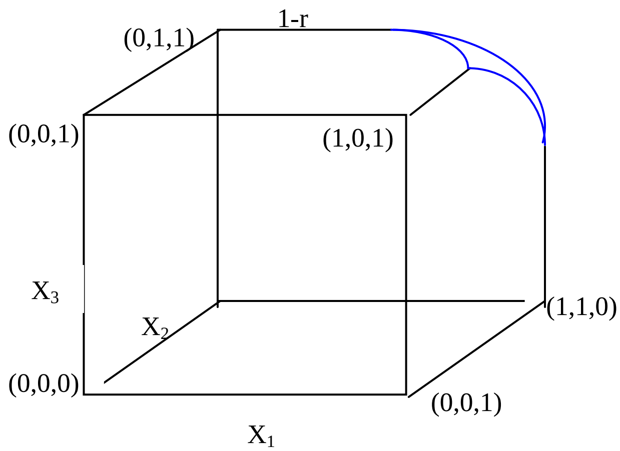
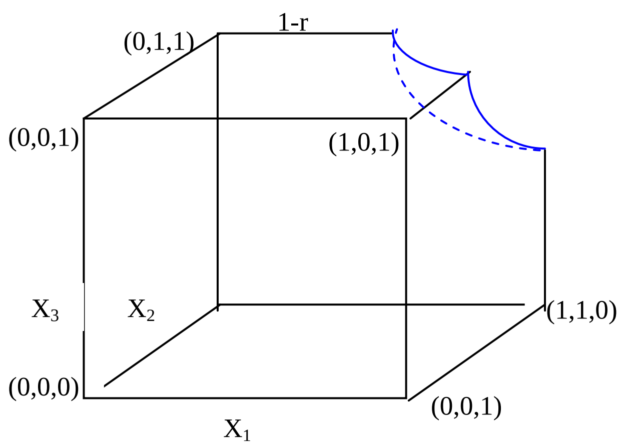


Figure 3c: Operability Region for d=0.5



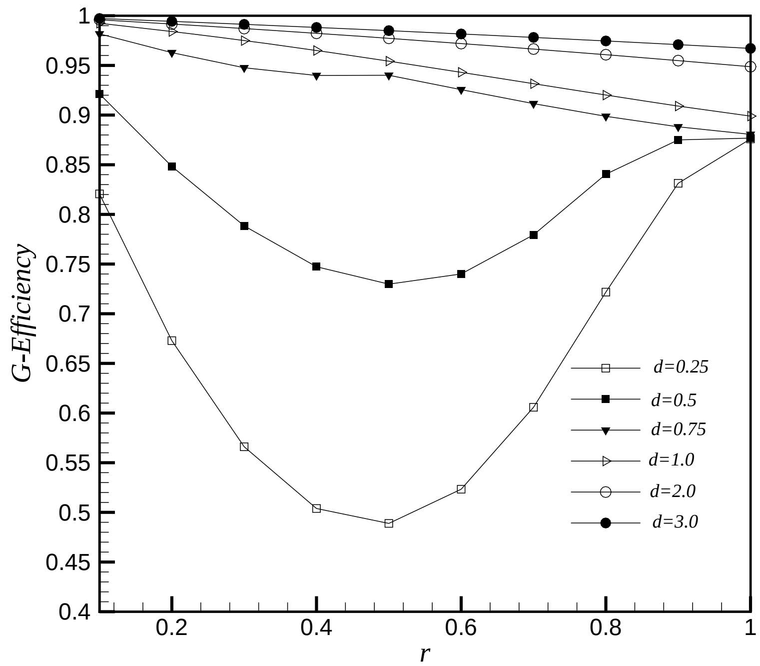
Under the general equation (Equation 2), the preferred design would replace the high-high-high point with a point in the middle on the surface boundary of the restricted design space for the chosen value of *d* (at angles $\theta = 45°, \phi = 54.7356°$). For different combinations of *d* and *r,* Table 1 shows the points *a* and a^* that replaces the high-high-high point.

Table 1: The Point *a* for Different Combinations of *d* and *r*

d	r	*a* (rounded)	*a** (rounded)
.25	**.1**	0.920	0.840
	0.5	0.599	0.196
	1.0	0.198	-0.605
0.5	**0.1**	0.944	0.889
	0.5	0.722	0.444
	1.0	0.444	-0.111
0.75	**0.1**	0.958	0.916
	0.5	0.791	0.582
	1.0	0.582	0.165
1.0	**0.1**	0.967	0.933
	0.5	0.833	0.667
	1.0	0.667	0.333
2.0	**0.1**	0.982	0.964
	0.5	0.908	0.816
	1.0	0.816	0.632
3.0	**0.1**	0.987	0.975
	0.5	0.937	0.874
	1.0	0.874	0.747

Figure 4 shows that for fixed value of r, the G-efficiency increase as d increases. For $d \geq 0.75$, the G-efficiency decreases as r increases. This indicates that the more one truncates from the corner the more one would loose in the G-sense. However, the design is still enjoying high G-efficiency (at least 89.9%). Since the G-efficiency is a lower bound for the D-efficiency (see Atwood, 1969), one can deduce that the design is also highly efficient in the D-sense. For $d<0.75$, the G-efficiency curve takes an upward quadratic shape. It is decreasing with increasing values of r as long as $r<0.6$. But for $r\geq 0.6$, the efficiency increases as r increases. This behavior was unexpected, since one would expect the efficiency to decrease as one excludes more space from the altered corner. Figure 5 presents boxplots of the scaled prediction variance (SPV) at *d*=0.5 for each *r* value. Although the maximum SPV appears at r=0.5 and results in the worst G-efficiency there, a more closer look at the graph shows how the largest SPV median, largest first/third quartile and largest IQR appear at *r*=1. This means the SPV distribution is behaving as expected in the sense that as much we extract from the design region prediction capability would decrease. The graph shows how all boxplots for *r*<0.9 are almost similar to each other. These facts were not captured by the G-efficiency criterion, which highlights the need of having other criteria that could judge the design prediction capability not only upon one single value (maximum SPV). The Fraction of design Space (FDS) of Zahran et al. (2003) is one of such criteria that could evaluate the prediction capability of the design more fairly. FDS gives the fraction of design space that is less than or equal to each SPV value throughout the design region. Usually, FDS is used to compare different designs on same design space. However, Figure 6 presents the FDS graph of the same design for different design spaces to be able to study the behavior of the preferred design under different design spaces. At *d*=0.5, the design prediction capability is generally getting worse as r increases. The FDS curve at r=1 has the highest SPV all over the design region. Whereas, the FDS curves of the design at r=0.5 and r=0.1 are similar to each other for almost 80% of the design space of each r value. The graph shows the rapid increase that happens in the SPV values in the remaining 20% of the design space at *r*=0.5, which leads to a low G-efficiency.

Figure 4: G-efficiency for different combinations of *d* and *r*.



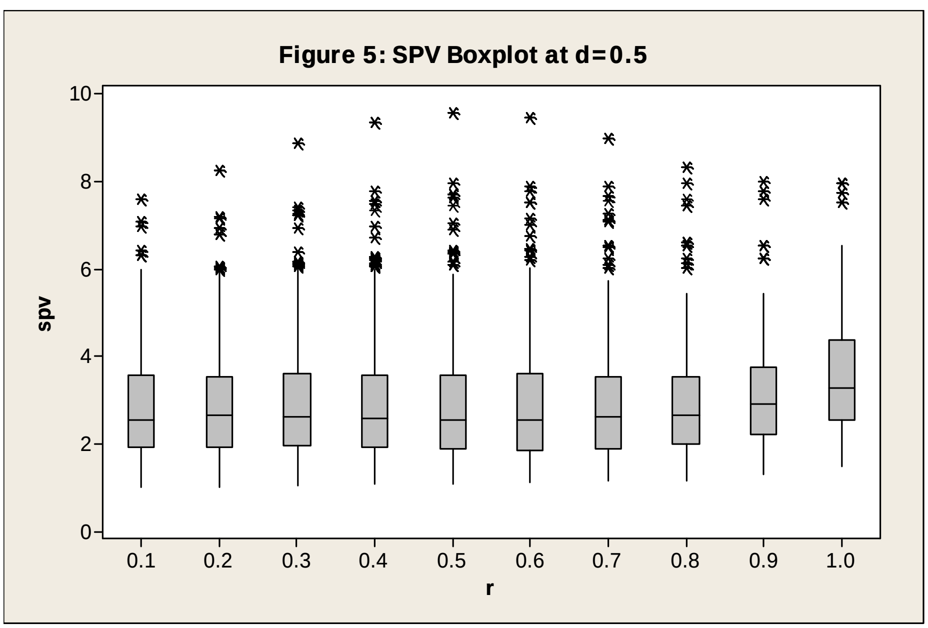
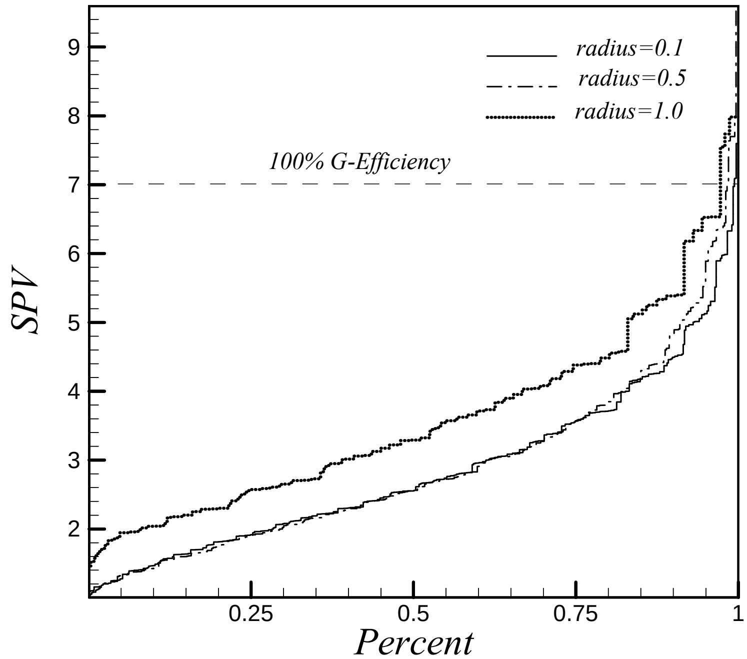
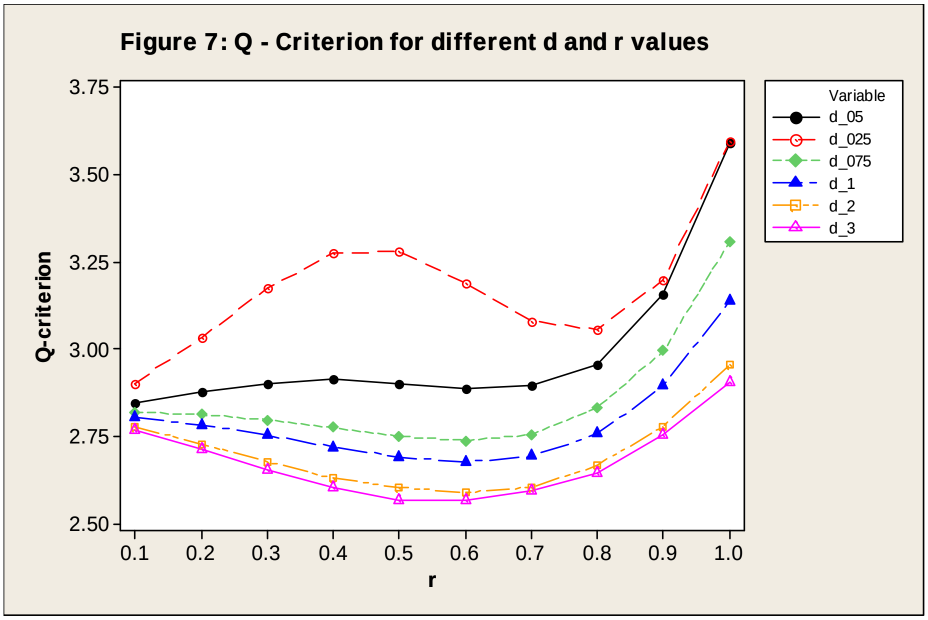


Figure 6: FDS Graph at *d=0.5* for different *r* values



Empirical study shows that the Q-criterion is minimized when the replaced H-H-H point is positioned exactly in the middle, i.e. when using our design. Figure 7 depicts the Q-criterion under different scenarios of design space (different values of *d* and *r)*. For a specific value of *r*, the design loses in the Q-sense as one excludes more from the H-H-H corner (i.e. decreasing *d*). The design performs better with *d* values greater than 1. For $d \geq 0.75$, the design loses its efficiency in the Q-sense when one increases *r* as long as *r*>0.5. However for *r*<0.6, the design performs better in the Q-sense as one excludes more from the altered corner.



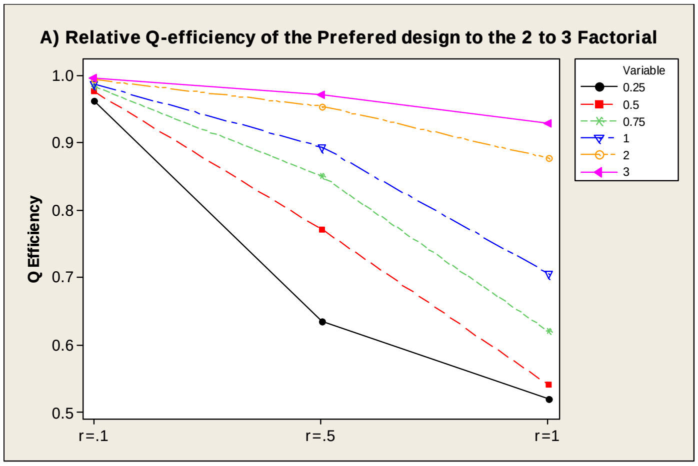
At some situations, there might be an interest in estimating a subset of s parameters from the model parameters as precisely as possible. This is the notion of interest in D_s-criterion. The interest in only a subset of the parameters might occur in situations where the researcher has already done one-at-a-time studies on each factor and now wants to study the interaction of the factors more thoroughly, in which case a D_s-efficient design is preferable. Zahran (2004) shows that the preferred design is also high efficient in the D_s-sense when d=2. Table 2 presents the D_s-efficiency of the design for different values of d and r. The design is efficient in the D_s sense for low values of r and high values of d, i.e. for small deviations from the whole cube. In real life situations, usually r is given to the researcher. Therefore, if he/she is aiming at estimating the interaction terms as precisely as possible, high values of d is to be chosen.

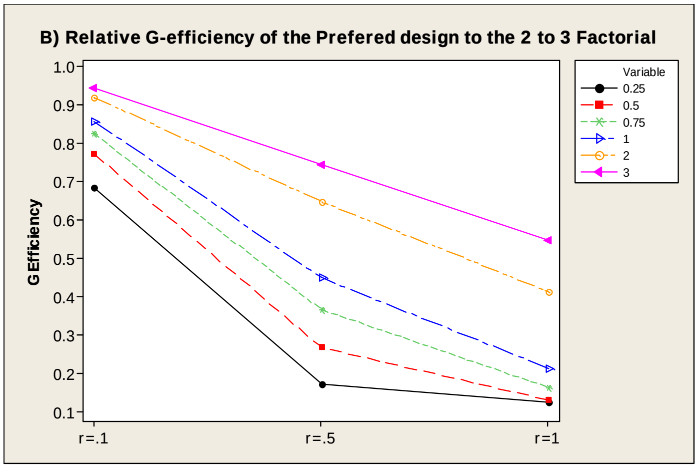
Table 2: D_s efficiency for different combination of *d* and *r*

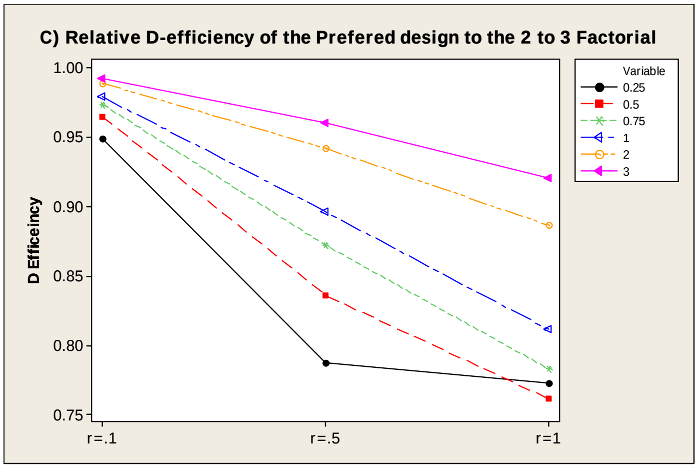
d	***r* =.1**	***r* =.5**	***r* =1**
0.25	0.780003	0.308798	0.257694
0.5	0.843246	0.416998	0.260509
0.75	0.880308	0.515268	0.299976
1	0.903558	0.59014	0.358423
2	0.946025	0.751897	0.558846
3	0.962584	0.823216	0.6724592

For any experiment there are two regions: operability region and the region of interest. The operability (experimental) region is defined on the basis of the capability of the process to operate at certain settings of the independent variables. However, the researcher may have primary interest in another region, which is called the region of interest. Typically, the two regions are the same. Consider now that the region of interest is the whole cube, for which the 2^3 Factorial design is optimal. Figure 8 depicts the relative efficiency of our design to the 2^3 Factorial for different operability regions. This kind of efficiency gives some sense of what we are losing by having the restriction on the design space. The relative efficiency is decreasing with high values of r and low values of d, that is when the excluded portion is increasing. Generally speaking, we do not lose much by implementing the restriction on the region in terms of the D-sense. The worst D-efficiency is about 76% which occurred at d=.25 and r=1 (maximum chopping from the design space). The worst Q-efficiency under same restricted design space is 52%, while the G-efficiency is only about 13%. Thus if the region of interest is the whole cube, excluding much from the design space hurts the prediction capability of the design badly while the estimation capability is not badly affected.

Figure 8: Alphabetical Relative Efficiency of the preferred Design to the 2^3 Factorial Design







CONCLUSIONS AND DISCUSSION

Under prohibiting the high-high-high combination in a three factor factorial design, the equation defining the boundary of this restricted design space is generalized. This general form is intuitive since it gives the practitioner more flexibility to define the design space. Depending on prior information about restrictions of feasible points, the interest in subsets of the parameters, or the nature of region of interest, one can choose the power, *d*, in the general form equation. Moreover the value of *r* specifies what fraction of design factor ranges the practitioner may want to alter.

For the first order with two-way interactions model in the three-factor case, the preferred design offered in Zahran (2004) to accommodate the cube with a banned corner, is studied under this general equation of defining the design boundary. Using the FDS plots, the prediction capability of the design is more stable and better for small deviations from the cube (high *d* and small *r* values). The design performs well in the Q- and G-sense for *d* values greater than or equal to 1. In addition, the design is high efficient in the D_s sense for small deviation from the cube, where s is the set of all two way interactions. If the region of interest is the whole cube, we found that the design is highly efficient in the D-sense for any values of (*r*,*d*). However, the Q- and G-efficiency are hurt when increasing the size of the excluded corner.

Although, we have considered the problem of excluding the high-high-high combination, all the results hold for the problem of excluding any corner combination of the three factors.

References

Atkinson, A.C. and Donev, A.N. (1992). *Optimum Experimental Designs*. Oxford University Press, Oxford.

Atwood, C. L.(1969), "Optimal and Efficient Designs of Experiments", *The Annals of Mathematical Statistics*, 40, 1570-1602.

Draper, N.R. and St John, R.C. (1977), "Designs in Three and Four components for Mixtures Models with Inverse Terms". *Technometrics*, 19, pp. 117-130.

Johnson, M.E. and Nachtsheim, C. J. (1983). "Some guidelines for Constructing Exact D-Optimal Designs on Convex Design Spaces". *Technometrics*, 25, pp. 271-277.

Kennard, R.W. and Stone, L. (1969). "Computer Aided Design of Experiments". *Technometrics*, 11, pp. 298-325.

Kiefer, J. and Wolfowitz, J. (1960). "The Equivalence of Two Extremum Problems". *Canadian Journal of Mathematics*,12, pp. 363-366.

Montgomery, D. C. , Loredo, E. N., Jearkpaporn, D., Testik, M.C. (2002). "Experimental Designs for constrained Regions". To appear in Quality Engineering.

Myers, R.H. and Montgomery, D. C. (2002), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd edition, Wiley.

Nachtsheim, C. J. (1987). "Tools for Computer-Aided Design of Experiments". *Journal of Quality Technology*, 19, pp. 132-160.

Snee, R.D. (1985). "Computer-Aided Design of Experiments – Some Practical Experiences". *Journal of Quality Technology*, 17, pp. 222-236.

Zahran, A.R (2004) "Modifying 2^3 Factorial to Accommodate a Restricted Design Space", the Proceedings of the 16th Annual Conference of Statistics Department and Center of Information and Computing Systems, Cairo Univ., Egypt.

Zahran, A.R & Anderson-Cook, C. (2003) "A General Equation and Optimal Design for a 2-factor Restricted Region", Statistics and Probability Letters, 64, 9-16.

Zahran, A.R, Anderson-Cook, C. & Myers, R. (2003). "Fraction of Design Space to Assess the Prediction Capability of Response Surface Designs", JQT, 35, 377-386.

Zahran, A.R, Anderson-Cook, C., Myers, R. & Smith, E.P (2003) "Modifying 2^2 Factorial Designs to Accommodate a Restricted Design Space", JQT, 35, 387-392.