

1-1-2011

# Self learning strategies for experimental design and response surface optimization

Adel Alaeddini  
*Wayne State University,*

Follow this and additional works at: [http://digitalcommons.wayne.edu/oa\\_dissertations](http://digitalcommons.wayne.edu/oa_dissertations)

 Part of the [Industrial Engineering Commons](#), [Operational Research Commons](#), and the [Statistics and Probability Commons](#)

---

## Recommended Citation

Alaeddini, Adel, "Self learning strategies for experimental design and response surface optimization" (2011). *Wayne State University Dissertations*. Paper 267.

**SELF LEARNING STRATEGIES FOR EXPERIMENTAL DESIGN AND  
RESPONSE SURFACE OPTIMIZATION**

by

**ADEL ALAEDDINI**

**DISSERTATION**

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

**DOCTOR OF PHILOSOPHY**

2011

MAJOR: INDUSTRIAL ENGINEERING

Approved by:

---

Advisor

Date

---

---

---

---

## **DEDICATION**

*For my wife, Sara, and my parents who help me the most to make my wishes true*

## **ACKNOWLEDGEMENTS**

*I acknowledge Dr. Kai Yang, my advisor in the Department of Industrial and Systems Engineering for his supports throughout my study. Also, I am very thankful to Dr. Alper Murat and Dr. Bruce Ankenman for their brilliant comments and ideas during my Dissertation. Finally, I kindly appreciate the support of my family and friends.*

## TABLE OF CONTENTS

Dedication.....	ii
Acknowledgements .....	iii
List of Tables .....	viii
List of Figures .....	x
CHAPTER 1 INTRODUCTION.....	1
1.1 RELEVANT BACKGROUND.....	5
1.1.1 RESPONSE SURFACE METHODOLOGY .....	5
1.1.2 OPTIMAL DESIGNS .....	7
1.1.3 SPACE FILLING DESIGN.....	8
1.1.4 ONE-FACTOR-AT-A-TIME .....	10
1.1.5 ADAPTIVE EXPERIMENTAL DESIGN .....	11
1.1.6 BAYESIAN OPTIMIZATION.....	13
1.1.7 GLOBAL OPTIMIZATION METHODS BASED ON RESPONSE SURFACES.....	14
1.2 PRELIMINARIES.....	14
1.3 LEAST-SQUARE REGRESSION SPLINE (LSRS).....	15
1.4 SEQUENTIAL SIMPLEX OPTIMIZATION.....	16
1.5 MULTIPLE ADAPTIVE REGRESSION SPLINE (MARS).....	18
CHAPTER 2 THE PROPOSED STRATEGY FOR QUADRATIC MODELS WITH TWO VARIABLES.....	22
2.1 INTRODUCTION.....	22
2.2 TERMINOLOGY AND ASSUMPTIONS .....	22

2.3	ALGORITHM AND INITIAL RUN DESIGN .....	24
2.4	NON-PARAMETRIC APPROACH: RANKING STRATEGY .....	28
2.4.1	BEST AT CENTER (BCE) OPTIMAL REGIONS .....	30
2.4.2	BEST AT CORNER (BCO) OPTIMAL REGIONS .....	32
2.5	PARAMETRIC APPROACH: MODEL FITTING STRATEGY .....	33
2.6	DESIGN STRUCTURE FOR THE NEXT RUNS .....	34
2.7	ILLUSTRATIVE EXAMPLE, SIMULATED EXPERIMENTS AND CASE STUDIES .....	35
2.7.1	ILLUSTRATIVE EXAMPLE .....	35
2.7.2	SIMULATED EXPERIMENTS.....	38
2.7.3	PAPER HELICOPTER .....	43
2.7.4	TRAUMATIC BRAIN INJURY (TBI): DESIGN OF CONTROLLED CORTICAL IMPACT MODEL.....	47
2.8	DISCUSSION .....	49
CHAPTER 3 THE PROPOSED STRATEGIES FOR QUADRATIC AND CUBIC FUNCTIONS WITH N-VARIABLES .....		51
3.1	TERMINOLOGY AND ASSUMPTIONS .....	51
3.2	PROPOSED N-ASRSM ALGORITHM.....	53
3.3	DESIGN STRUCTURE OF THE FIRST AND SUBSEQUENT RUNS.....	54
3.4	NON-PARAMETRIC APPROACH: RANKING STRATEGY .....	56
3.4.1	REDUCING FACTOR SPACE (FS).....	57
3.4.2	SELECTING ADDITIONAL CORNER POINTS .....	60
3.4.3	RISK ADJUSTMENT .....	62
3.4.3.1	ESTIMATING THE VARIANCE WITH NOT ENOUGH DATA .....	62

3.4.3.2	CALCULATING THE PROBABILITY OF INCORRECT RANKING AND RISK ADJUSTMENT.....	63
3.5	PARAMETRIC APPROACH: MODEL FITTING STRATEGY .....	65
3.6	EXPANSION OF OR TO CONTAIN EO.....	66
3.7	NUMERICAL EXAMPLES.....	67
3.7.1	QUADRATIC RESPONSE MODELS .....	67
3.7.2	NON-LINEAR RESPONSE MODELS.....	72
3.8	DISCUSSION .....	74
3.9	AN EXTENSION OF THE PROPOSED STRATEGY WITH OPTIMAL DESIGN (N-ASRSM2).....	75
3.9.1	TERMINOLOGY AND ASSUMPTIONS .....	75
3.9.2	ALGORITHM.....	76
3.10	NUMERICAL EXAMPLES.....	79
3.10.1	QUADRATIC RESPONSE MODELS .....	79
3.10.2	NON-LINEAR RESPONSE MODELS.....	82
3.11	DISCUSSION .....	84
CHAPTER 4 THE PROPOSED STRATEGIES FOR BLACK BOX FUNCTIONS WITH N-VARIABLES (B-ASRSM).....		85
4.1	TERMINOLOGY AND ASSUMPTIONS .....	85
4.2	ALGORITHM.....	87
4.3	DESIGN STRUCTURE OF THE FIRST AND SUBSEQUENT RUNS.....	88
4.4	FITTING LEAST SQUARE REGRESSION SPLINE (LSRS).....	90
4.5	IDENTIFY THE SFS WITH MAXIMUM EXPECTED IMPROVEMENT .....	92

4.6	TAKING NEW EXPERIMENTS AND BREAKING DOWN STRUCTURE OF THE METHODOLOGY .....	93
4.7	DESIGN OF THE NEXT RUNS AND STOPPING CRITERIA .....	93
4.8	NUMERICAL EXAMPLES.....	94
4.9	DISCUSSION .....	97
4.10	AN EXTENSION OF THE PROPOSED STRATEGY WITH SPACE FILLING DESIGN AND MULTIPLE ADAPTIVE REGRESSION SPLINES (B-ASRSM2).....	98
4.10.1	ALGORITHM.....	98
4.10.2	NUMERICAL EXAMPLES.....	100
4.10.3	DISCUSSION .....	102
	CHAPTER 5 CONCLUSIONS AND FUTURE DIRECTIONS .....	104
	References.....	105
	Abstract .....	110
	Autobiographical Statement.....	112

## LIST OF TABLES

Table 1 The Runs and experiments of the ASRSM method.....	36
Table 2 The response models used in the simulated experiments .....	38
Table 3 The $R_{adj}^2$ and optimality gap results of simulated experiments with 9 observations determined a priori for CCD and optimal designs .....	40
Table 4 The $R_{adj}^2$ results of simulated experiments beginning with 7 design points and then incrementally adding one design point at a time .....	42
Table 5 The optimality gap results of simulated experiments for the final 9 design points in Table4 .....	42
Table 6 The runs and experiments of ASRSM, CCD, A-, D-, and V-optimal designs for the paper helicopter experiment. $X'_1$ and $X'_2$ correspond to the coded wing length and body length factors, respectively .....	44
Table 7 The fly time responses obtained for the paper helicopter experiment.....	45
Table 8 The prediction results of ASRSM, CCD, and optimal designs for the paper helicopter case study.....	46
Table 9 The runs and experiments of (a) CCD and (b) ASRSM method for the brain trauma case study.....	48
Table 10 The response models used in the simulated experiments of N-ASRSM .....	68
Table 11 The $R_{adj}^2$ for trials 7, 8, 9 of responses with two variables and trials 11, 12 and 13 of the ones with three variables.....	70
Table 12 The optimality gap for trials 7, 8 and 9 of responses with two variables and trials 11, 12, and 13 of the ones with three variables.....	71
Table 13 The non-linear response models used for studying N-ASRSM.....	72
Table 14 The average optimality gap of the proposed N-ASRSM and comparing methods .....	73
Table 15 The average Euclidean distance of the estimated optima from the real optima for N- ASRSM and comparing methods .....	74

Table 16 The quadratic response models used in the simulated experiments of the N-ASRSM2 .....	79
Table 17 The average $R_{adj}^2$ of the N-ASRSM2 an comparing methods for trials 7, 8 and 9 of responses for quadratic functions .....	80
Table 18 The Average optimality gap of the N-ASRSM2 and comparing methods for trials 7, 8 and 9 of responses for quadratic responses.....	82
Table 19 The non-linear response surface models used for studying N-ASRSM2 performance.	82
Table 20 The average optimality gap of the proposed N-ASRSM2 and comparing methods .....	83
Table 21 The average Euclidean distance of the estimated optima from the real optima for the N-ASRSM2 and comparing methods .....	84
Table 22 The response relation considered for the analysis of the B-ASRSM .....	95
Table 23 The average optimality gap and average Euclidean distance to RO of the proposed B-ASRSM methods along with the comparing methods.....	96
Table 24 The average optimality gap and average Euclidean distance to RO of the proposed B-ASRSM2 methods along with the comparing methods.....	101

## LIST OF FIGURES

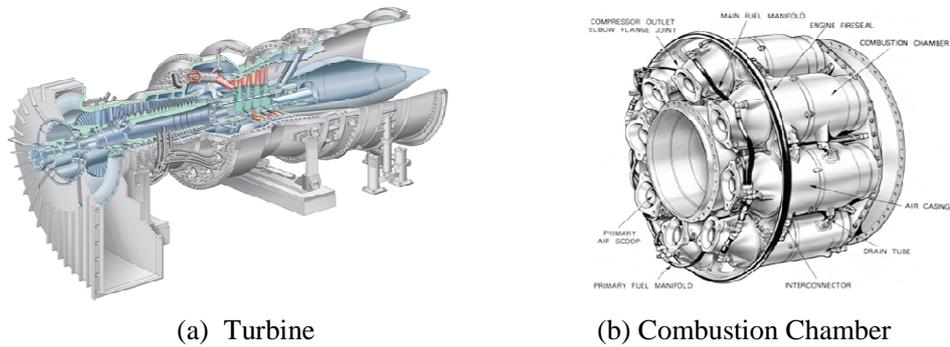
Figure 1 Turbine and aircraft engine .....	2
Figure 2 Simplexes in (A) zero-dimension, (B) one-dimensions, (C) two-dimension and (D) three-dimension (Walters et al. 1991).....	16
Figure 3 The simplex reflection move for (A) one-dimension (B) two-dimension, and (C) three-dimension factor spaces. Dashed line represents the old simplex. Open circle shows the average of the remaining vertices (Walters et al. 1991) .....	17
Figure 4 (a) Possible moves in the variable-size simplex algorithm, (b) logic of the possible moves in the variable-size simplex algorithm (Walters et al. 1991).....	18
Figure 5 An illustration of terminology of the ASRSM on a two dimensional factor space with $E=5$ .....	23
Figure 6 Scheme of the proposed ASRSM methodology.....	25
Figure 7 Illustration of the factor space reduction across runs $r=1, 2,$ and $3$ .....	26
Figure 8 (a) Initial factor space and design structure and (b) Diagonal cross-section of the traditional CCD and proposed ASRSM approach .....	27
Figure 9 (a) Implied optimal region (b) Convex hull envelope of the implied optimal region (c) Rectangular envelope of the implied optimal region based on a two dimensional factor space.....	29
Figure 10 BCE ORs (dotted region: impliedOR, shaded region:OR).....	30
Figure 11 BCO ORs when $N_2$ is at center (dotted region: implied OR, shaded region: OR) .....	32
Figure 12 BCO ORs when $N_3$ is at center (dotted region: implied OR, shaded region: OR) .....	32
Figure 13 Expansion of the ORr when the EOr from model fitting falls outside.....	34
Figure 14 The (a) Actual, (b) CCD, and (c) ASRSM estimated contours of the response $Z$ .....	35
Figure 15 Illustration of the steps of ASRSM for runs 1 and 2 .....	37
Figure 16 Paper helicopter .....	43

Figure 17 The (a) CCD and (b) ASRSM estimated contours of the paper helicopter fly time response .....	47
Figure 18 3D plots of the estimated response for (a) CCD, (b) ASRSM, (c) RBF in the brain trauma case study .....	49
Figure 19 The (a) CCD and (b) ASRSM estimated contours of the response in the brain trauma case study .....	49
Figure 20 An illustration of terminology on a three dimensional factor space .....	52
Figure 21 The general scheme of N-ASRSM .....	54
Figure 22 (a) Initial factor space and design structure and (b) Diagonal cross-section of the traditional CCD and proposed N-ASRSM approach .....	55
Figure 23 (a) Implied optimal region (b) Convex hull envelope of the implied optimal region (c) Rectangular envelope of the implied optimal region based on a three dimensional factor space .....	57
Figure 24 8 sub-regions of a 3 dimensional FS using 1 orthogonal hyper plane (b) 16 sub-regions of a 3 dimensional FS using 2 orthogonal hyper planes .....	59
Figure 25 Cut out regions using the ranking strategy based on two additional points .....	60
Figure 26 The effect of incorrect ranking on the NOR.....	62
Figure 27 Application of Gauss error function on risk adjustment .....	65
Figure 28 Expansion of OR in N-ASRSM when the EO falls outside.....	66
Figure 29 the contour plot of responses 1.1, 1.3.....	68
Figure 30 The contour plot of Responses 1.1 and 2.2 .....	72
Figure 31 An illustration of the terminology of the N-ASRSM2 on a two dimensional factor space .....	76
Figure 32 The general scheme of the proposed N-ASRSM2 strategy.....	77
Figure 33 (a) NOR of the N-ASRSM2 factor space after 6th experiments (b) a grid of 64 sub-regions of a 2 dimensional FS (c) a grid of 16 sub-regions of a 2 dimensional FS in the proposed N-ASRSM strategy .....	78

Figure 34 Cut out regions of N-ASRSM2 using the ranking strategy based on two additional points. ....	78
Figure 35 The graphical representation of the N-ASRSM2 on 2nd and 3rd response surface; the blue cells show OR, greens show OR not robust to miss-ranking (in risk adjustment), Yellows show non-robust NOR and Reds show robust NOR.....	81
Figure 36 The graphical representation of the N-ASRSM2 on 1st and 2nd response surface; the blue cells show OR, greens show OR not robust to miss-ranking (in risk adjustment), Yellows show non-robust NOR and Reds show robust NOR .....	83
Figure 37 An illustration of the terminology of the proposed B-ASRSM strategy .....	86
Figure 38 the general scheme of the proposed B-ASRSM .....	88
Figure 39 The effect of number of levels $k_i$ on the estimated surfaces: (1) $z = x \cdot \exp x^2 - y^2$ with $k=(2,2), (3,3)$ and $(4,44)$ in (a), (b) and (c) and, (2) $z = \sin 0.83\pi x \cdot \cos 1.25\pi y$ with $k=(2,2), (3,3)$ and $(4,44)$ in (e), (f) and (g) .....	89
Figure 40 An example of the fitted LSRS with a spline patch added.....	91
Figure 41 the graphical representations of the proposed B-ASRSM strategy steps .....	94
Figure 42 The contour and function plot of the responses considered for the analysis of the B-ASRSM; a, b, c represents 1st, 2nd and 3rd response relations respectively.....	95
Figure 43 the estimated response of problem 3 at different iteration using (a) RBF and (b) GP .	97
Figure 44 The general scheme of the proposed B-ASRSM(2) .....	99
Figure 45 The graphical representations of the proposed B-ASRSM2 strategy steps.....	100
Figure 46 The graphical representation of the B-ASRSM2 on 3rd response surface; the blue cells show OR, greens show OR not robust to miss-ranking (in risk adjustment), Yellows show non-robust NOR and Reds show robust NOR .....	102

## CHAPTER 1 INTRODUCTION

Most process and design optimization approaches such as the response surface methodology (RSM) require a complete experimental design to be determined prior to the experimentation process (Spendley, Hex and Himsworth, 1962). These preset designs offer ease of implementation and good performance over a wide range of applications. However, they lack the ability to adapt the design based on the characteristics of application and experimental space so as to reduce the number of experiments necessary. This, in particular, constitutes a major disadvantage in many industrial applications where the cost of experimentation is high or when the experimentation resources are limited. These industrial experiments share the following two main characteristics: (1) prior to the experiment, the behavior of the experimental design space is not well known; (2) the cost of each experimental trial is prohibitively high and the experimental budget is limited. An example for such industrial experiments is the combustion test for aircraft engine or turbines where prototypes are very expensive and the behavior of different designs are highly unpredictable (See Figure 1). The computational experimentation approach is commonly resorted as a cost effective alternative to physical testing of complex engineering systems. However, these computational experiments may take 5 to 20 hours per simulation run as in the case of FEA of etc. (Gramacy and Lee 2009). Gu (2001) reports that one crash simulation on a full passenger car takes 36-160 hours at Ford Motor Company.



**Figure 1 Turbine and aircraft engine**

The focus of this dissertation is mainly on the industrial experiments with high experimentation cost, limited experimental resources and requiring high design optimization performance. In designing industrial experiments, the traditional RSM methodologies (CCD, Box-Behnken optimal designs, etc.) are often preferred for various advantages, e.g., rotatability and variance of error estimation. However, these methods rely on “one-shot” designs and thus fall short in providing efficient experimental designs for highly engineered complex systems. This has been pointed out by George E. P. Box as “There should be more studies of statistics from the dynamic point of view” in Box (1999) and that “I think we have spent too much time on one-shot statistical procedures designed to test rather than to learn.” in response to Myers (1999). Further, these methods fit a regression model of the system responses to accurately predict the response curve over the entire domain of feasibility. However, the prediction in the neighborhood of the optimum is often more important than prediction in the domain of feasibility.

In this dissertation, we propose a number of adaptive sequential experimentation strategies based on global optimization concepts, nonparametric regression methods, and response surface methodology for different type’s response surfaces for industrial experiments with, noise, high

experimentation cost, and requiring high design optimization performance. We consider the experimentation as successive series of small data collection efforts. At each step, we learn from the previous results, refine our understanding and develop a new model for the next experiment to reduce uninformative experiments and improve the quality of results. The idea of adaptive experimental design is not new. Beginning with the sequential RSM experimentation with multiple blocks in Box and Wilson (1951), there have been many ideas such as one-factor-at-a-time (OFAT) (Friedman and Savage 1947, Daniel 1973), adaptive OFAT (Frey, Engelhardt, and Greitzer, 2003), adaptive RSM (Wang, Dong and Aitchison 2001, Wang 2003), successive RSM (Stander, 2001), evolutionary operation (Box and Draper, 1969), steepest ascent based methods (Box and Wilson, 1951), and sequential and adaptive approximation methods from the engineering design discipline.

The Adaptive Sequential Response Surface Methodology (ASRSM) approach presented in this dissertation is a local optimization approach for physical experiments where the region of interest is formed by a number of input factors is already determined. Furthermore, in most practical applications, the current settings of the factor values are usually determined and known to produce a stable response and a satisfactory yield. However, due to extraneous changes over time, the current conditions may become less robust and sub-optimal. Hence there might be easy gains in yield by moving in the surrounding region of the design space. We do not make any assumptions regarding the noise in response. Hence our goal is to precisely estimate the relationship between important factors and response and identify the most likely location for the process/product to be optimized in the detailed RSM experimentation stage.

The most salient aspect of the proposed strategies is its experimentation efficiency. Specifically, the proposed approach identifies, for a given response target, the input factor

combination (or containing region) in less number of experiments than its counter parts. This sequential adaptive approach uses the information gained from the previous experiments to design the subsequent experiment by simultaneously reducing the region of interest and identifying factor combinations for new experiments. This reduction is achieved through rank ordering of the responses of preceding experiments and use of polynomial (mostly quadratic) behavior of the underlying function near the real optima. Throughout the process, we consider a fixed design (i.e. two levels or multi-levels factorial design, or space-filling design) which allows inheriting some of the experiments from the previous runs. As a result this method efficiently increases the accuracy and precision of the estimated optimal point by reducing the region of interest.

The strategies in this research differ from earlier approaches on adaptive and sequential RSM in three different ways. First, the reduction of the region of interest is optimal if the relationship between the response and input factors is quadratic (within each small sub-region inside the factor space) and response is deterministic. Specifically, the optimal factor combination is always contained in the reduced region. Second, the proposed strategies typically require fewer experiments in each run as the result of inheriting previous experiments and fixed design structure. Lastly, the proposed strategies identify the reduced region of interest with a combination of nonparametric (ranking based method) and parametric (model based) methods rather than the response levels obtained from each experiment solely. This is indeed similar to using not the value of a parameter but its rank in robust statistics (Hettmansperger and McKean, 1998).

The structure of the dissertation is as follows, in the remainder part of this chapter the preliminaries to the study are briefly discussed followed by a short literature survey on the

advancements in RSM with a special emphasis on the adaptive experimentation methodologies. Chapter 2 presents the proposed ASRSM methodology for the case of quadratic underlying function with two variables, which is the basis of the strategies in the following chapters. Chapter 3 explains the proposed strategies for the case of quadratic and cubic underlying functions with  $n$  variables (N-ASRSM), which is an extension of Chapter 2 methodology into higher dimensions. Chapter 4, describes strategies for noisy black-box functions with  $n$  variables which is gained extending both the complexity and the number of variables in the previous chapters. Finally, Section 5 discusses the results and presents directions for future research.

## **1.1 Relevant Background**

This section presents the relevant literature for the proposed adaptive experimental design methodology. We first review the classical response surface methodologies and then more advanced methods including optimal design, Bayesian design and incomplete design strategies. Finally, we briefly describe other adaptive design methodologies such as steepest ascent, simplex-based methods, evolution operation methods, adaptive OFAT methods, adaptive RSM, sequential RSM, and sequential and adaptive approximation methods from the engineering design domain

### **1.1.1 Response Surface Methodology**

RSM has been used as one of the most effective tools for process and product development since its introduction by Box and Wilson (1951). RSM consists of statistical and numerical/mathematical optimization techniques for examining the relationship between one or more response variables and a set of quantitative experimental variables or factors. Since the

literature on RSM is vast, we herein refer the reader to a number of good review studies. Box (1999) provides a retrospective on the origins of RSM with a general philosophy of sequential learning. Myers, Khuri, and Carter (1989) present a thorough discussion of RSM from 1966 to 1988. Myers (1999) discusses the RSM state in late 90s and gives some directions for future research. Myers, Montgomery, Vining, Borror and Kowalski (2004) presents a retrospective and literature survey on RSM.

Central Composite Design (CCD) and Box and Behnken Design (BBD) are the most popular class of designs used for fitting second order model (Box and Behnken, 1960). Generally, the CCD consists of a  $2^k$  factorial or fractional factorials of resolution  $V$  with  $n_F$  runs,  $2k$  axial or star runs, and  $n_C$  center runs (Figure 2). There are usually two parameters in CCD that must be specified: the distance  $a$  of the axial runs from the design center and the number of center points. It is common to set  $a = (n_F)^{1/4}$  to make the design rotatable. Also three to five runs are recommended in the literature (Montgomery, 2008). The number of runs in CCD increases exponentially with the number of design variables, and hence becomes inefficient for high dimensional design problems. One alternative to CCD is small composite designs (SCD) that consist of a fraction of CCD points. However, the SCD has significant difficulty in estimating linear and interaction coefficients (Myers and Montgomery, 1995). BBD is another design approach which requires  $k \geq 3$  (Box and Behnken, 1960). BBD is formed by combining  $2^k$  factorials with incomplete block designs. This design does not contain any points at the vertices of the region created by the upper and lower limits for each variable.

### 1.1.2 Optimal Designs

Optimal design methodologies select designs which are “best” with respect to some statistical criterion, which is related to the variance-matrix of the estimator. This selection process includes: specifying the model; determining the region of interest, selecting the number of runs to make, specifying the optimality criterion, and choosing the design points from a set of candidate points spaced over the feasible design region. Kiefer (1959, 1961) and Kiefer and Wolfowitz (1959) greatly contribute to the development of the idea of optimal designs.

Optimal designs offer three advantages over sub-optimal experimental designs (Atkinson et al., 2007): (1) Optimal designs reduce the costs of experimentation by allowing statistical models to be estimated with fewer experimental runs. (2) Optimal designs can accommodate multiple types of factors, such as process, mixture, and discrete factors. (3) Optimal designs can be optimized when the design-space is constrained, for example, when the mathematical process-space contains factor-settings that are practically infeasible.

It is known that the least squares estimator minimizes the variance of mean-unbiased estimators. In the estimation theory for statistical models with one real parameter, the reciprocal of the variance of an (efficient) estimator is called the “Fisher information” for that estimator. Because of this reciprocity, minimizing the variance corresponds to maximizing the information.

When the statistical model has several parameters, however, the mean of the parameter-estimator is a vector and its variance is a matrix. The inverse matrix of the variance-matrix is called the “information matrix”. Because the variance of the estimator of a parameter vector is a matrix, the problem of “minimizing the variance” is complicated. Using statistical theory, statisticians compress the information-matrix using real-valued summary statistics; being real-valued functions, these “information criteria” can be maximized. The traditional optimality-

criteria are invariants of the information matrix; algebraically, the traditional optimality-criteria are functions of the Eigen values of the information matrix (Pukelsheim, 2006).

$D$ -optimal design is the most widely used criterion in optimal designs. A design is said to be  $D$ -optimal if  $|(X'X)^{-1}|$  is minimized which is equivalent to minimizing the volume of the joint confidence region of the vector of regression coefficients or equivalently maximizing the differential Shannon information content of the parameter estimates. Andere-Rendon, Montgomery, and Rollier (1997) use  $D$ -optimal design for mixture experiments.

$A$ -optimality seeks to minimize the trace of the inverse of the information matrix ( $Min tr(X'X)^{-1}$ ). This criterion results in minimizing the average variance of the estimates of the regression coefficients.

There are also other types of optimal criterion; for example  $G$ -optimal design minimizes the maximum scaled prediction variance over the design region, and  $V$ -optimal design that minimizes the average prediction variance over the set of  $m$  points of interest. More recently, Ginsburg and Ben-Gal (2006) suggest a new design-of-experiment optimality criterion, termed  $V_s$ -optimal, for the robust design of empirically fitted models. Pukelsheim (2006) provides an excellent source on the optimal design of experiments.

### 1.1.3 Space Filling Design

Space filling designs are part of computerized design of experiments. Unlike classical design which use replication and blocking to control for noise, and randomization to control for bias, in space filling designs and in general computer experiments, blocking and randomization are not considered, since computerized experiments are assumed to be deterministic. In general such designs have the following properties (Santner et al. 2003):

- The only source of error is assumed to be model bias
- Designs should not take more than one observation for any set of inputs
- Designs should allow one to fit a variety of models
- Designs should provide information about all portions of experimental region (Designs should spread points evenly throughout experimental region)

Some of the reasons for using space filling designs are:

- Predictors for response are often based on interpolators
- Prediction error at any point is relative to its distance from closest design point
- Uneven designs can yield predictors that are very inaccurate in sparsely observed parts of experimental region

There are quite a few numbers of spaces filling designs, yet most of them may be clustered into the following groups (Santner et al.2003):

- *Simple Designs* including regular grid, random sampling, stratified random sampling.
- *Latin Hypercube Designs (LHD)* which can be shown (at least under some assumptions) to work better than random sampling, while being applicable to situation where space filling assumptions may be violated
- *Distance-based Designs* such as maxi-min distance design, mini-max distance design, and optimal average distance design, which use measure of spread to assess quality of design
- *Uniform Designs* such as  $L_\infty$  discrepancy and  $L_p$  discrepancy which measure uniformity of design by comparison against uniform distribution using discrepancy measures
- *Designs with multiple criteria*

For more information about space filling design one can refer to (Santner et al. 2003).

### **1.1.4 One-Factor-At-a-Time**

One-Factor-At-a-Time (OFAT) can be considered as the earliest adaptive sequential experimentation approach proposed (Friedman and Savage, 1947). OFAT changes one variable at a time while keeping others constant at fixed values to find the best response. Once a factor is changed, its value is fixed in the remainder of the process. This process is repeated until all the variables are tried. However, OFAT experimentation is generally discouraged in the literature on the experimental design in comparison with factorial design and fractional factorial design. Box, Hunter, Hunter (1978) and Montgomery (2008) talk about advantages of factorial experiment over OFAT experimentation. Czitrom (1999) write in favor of factorial experiment over OFAT experiments in terms of finding the behavior of the system. Frey, Engelhardt and Greitzer (2003) introduce Adaptive One-Factor-At-A-Time (AOFAT) experimentation method. They compare adaptive OFAT (AOFAT) technique with orthogonal arrays through computer simulations and concluded that AOFAT technique tends to achieve greater gains than those of orthogonal arrays when experimental error is small or the interactions among control factors are large. Frey and Jugulum (2006) investigate the mechanisms by which AOFAT technique led to improvement. The parameters that they investigated were conditional main effect, exploitation of an effect, synergistic interaction, anti-synergistic interaction, and overwhelming effect. Frey and Wang (2006) present the models of AOFAT and factor effects and illustrate with theorems that AOFAT method exploits main effects if interactions are small, and exploits two-factor interactions when two-factor interactions are large.

### 1.1.5 Adaptive Experimental Design

The idea of sequential and adaptive experimental design is not new. Box and Wilson (1951) suggest a two-stage sequential CCD where the first stage is a 2-level factorial or fractional factorial design and the axial points constitute the second stage. The axial points are then used if the lack-of-fit test indicates curvature in the system. The method of steepest ascent (Box and Wilson, 1951) is another adaptive sequential experimentation approach in which the experimental points move sequentially along the gradient direction. Evolutionary operation (EVOP), another adaptive experimental approach, iteratively builds a response surface around the optimum from the previous iteration by drifting factorial experiments with center points (Box 1957, Box and Draper 1969). Both these approaches are primarily used for shifting the region of interest close to the optimum and replicate the same experimental design iteratively in different regions of the factor space.

Spendley, Hex and Himsworth (1962) discuss the sequential application of simplex designs in optimization and evolutionary operation. They propose using a simplex pattern instead of a factorial pattern as in Box (1957). A simplex is a  $n + 1$  dimensional form in  $n$  dimensions, e.g. a triangle in two dimensions and a tetrahedron in three dimensions. They present a simplex search method where a sequence of experimental designs in the form of a regular or irregular simplex is used.

Moore et al. (1998) suggest an algorithm, known as *Q2*, for optimizing the expected output of a multi-input noisy continuous function. *Q2* is designed to need only a few experiments and avoids strong assumptions on the form of the function. Their algorithm uses instance-based determination of a convex region of interest for performing experiments. To define a neighborhood, they use a geometric procedure that captures the size and shape of the zone of

possible optimum location/s. Their algorithm also tries to optimize weighted combinations of outputs, and finds inputs to produce target outputs.

Anderson et al. (2000) develop a nonparametric approach called Pairwise Bisection (PB1) for optimizing expensive noisy function with few function evaluations. Their algorithm uses nonparametric reasoning about simple geometric relationships to find minima efficiently. They use nonparametric statistics since for its independence from the traditional assumptions of continuousness and Gaussian noise. They also used pairwise bisection as an attempt to automate the process of robust and efficient experiment design.

Wang, Dong and Aitchison (2001) develop an adaptive RSM methodology, called Adaptive Response Surface Method (ARSM). ARSM is a sequential experimentation method, where, at each iteration, ARSM discards portions of the design space that correspond to the response values worse than a given threshold value. Such elimination reduces the design space gradually to the neighborhood of the global design optimum. ARSM performs a CCD experiment at each iteration and thus the number of required design experiments increases exponentially with the number of design variables. Further ARSM does not inherit any of the previous runs and requires a completely new set of CCD points.

Wang (2003) proposes a modified ARSM where the CCD is substituted with Latin Hypercube Design (LHD). Stander (2001) proposes the successive RSM method (SRSM) which uses a region of interest, a subspace of the design space, to determine an approximate optimum. A range is chosen for each variable to determine its initial size. Then a new region of interest is centrally built on each successive optimum. The improvement in response is attained by moving the center of the region of interest as well as reducing its size through panning and zooming operations, respectively. At each sub-region, a  $D$ -optimal experimental design is used to best

utilize the number of available runs together with over-sampling to maximize the predictive capability.

### 1.1.6 Bayesian Optimization

The mainstream literature on Response Surface Optimization is classical or “frequentist” given that it considers parameters as unknown constants that need to be estimated from data. The sampling variability or experimental error is reflected in the sampling distributions of the estimates. This sampling variability can (and should) be considered in optimization. In contrast, the Bayesian approach to statistical inference considers model parameters (and in fact, any unknowns) as random variables. This has considerable advantages over the classical approach when optimizing a process based on a fitted model, since depending on the estimated parameters different optimal conditions will be determined. In the Bayesian approach, the uncertainty in the model’s parameters is directly incorporated in the analysis. Prior knowledge can be incorporated, if desired, into the optimization process. Otherwise, non-informative priors can be used for optimization purposes Del Castillo (2007).

$$\max_{w \in R} \int_{a_1}^{a_2} p(\tilde{y} | data, w) d\tilde{y} \quad (1)$$

where  $y$  is a vector of future response/s,  $w$  is a vector of controllable factors, and  $a_1$  and  $a_2$  show the specification region  $q$  response/s. Solving such optimization problem provides a solution that satisfies the specifications or tolerances on the responses of interest.

### **1.1.7 Global Optimization Methods Based on Response Surfaces**

Another adaptive and sequential experimentation research stream emerges from the engineering design community. In the engineering design, computation-intensive design analyses are commonly expensive computer “experiments” and thus require experimental optimization for design optimization. Chen, Tsui, Barton and Meckesheimer (2006) provide a review on design, modeling and applications of computer experiments. The response surface models based on computer experiments are called surrogates and commonly used in multidisciplinary design optimization. Sobieszczanski-Sobieski (1988) proposes concurrent subspace optimizations (CSSO) where the multidisciplinary systems are linearly decoupled for concurrent optimization. Renaud and Gabriele (1994) modify this algorithm to build response surface approximations of the objective function and the constraints. Rodríguez, Renaud, and Watson (1998) introduce a general framework for surrogate optimization with a trust region approach. The database for surrogate construction is generated by sampling the linearly decoupled disciplines. Rodríguez, Pérez, Padmanabhan, and Renaud (2001) present two sampling strategies, e.g., variable and medium fidelity samplings. Jones (2001) presents a taxonomy of existing approaches for using response surfaces for global optimization. Two other review studies in this field include Sobieszczanski-Sobieski and Haftka (1997) and Simpson, Booker, Ghosh, Giunta, Koch, and Yang (2002).

## **1.2 Preliminaries**

This Section briefly introduces some of the preliminaries required to comprehend the proposed strategies. In order to increase the performance of the proposed strategies, most of the methods explained here have had some modification before being used in the proposed models.

### 1.3 Least-Square Regression Spline (LSRS)

Least square spline is a way of using spline functions for data smoothing. These estimators have local fitting qualities similar to those for kernel and smoothing spline estimators. However they do not admit kernel or series representations, even asymptotically.

Assuming a basic nonparametric regression model where  $(x_i, y_i), i = 1, \dots, n$  satisfy:

$$y_i = \mu(x_i) + \varepsilon_i, i = 1, \dots, n \quad (2)$$

with  $\varepsilon_i$  representing zero mean, uncorrelated random errors having common variance  $\sigma^2$  and  $0 \leq x_1 \leq \dots \leq x_n \leq 1$ . Then, if  $\mu \in W_2^m[0,1]$  (polynomial of degree  $m$  with first and second derivatives existence) a Taylor expansion allows us to write the regression model as follows:

$$y_i = \sum_{j=1}^m \theta_j x_i^{j-1} + Rem(x_i) + \varepsilon_i, i = 1, \dots, n \quad (3)$$

where

$$Rem(t) = [(m-1)!]^{-1} \int_0^1 \mu^{(m)}(x) (x - \xi)_+^{m-1} d\xi \quad (4)$$

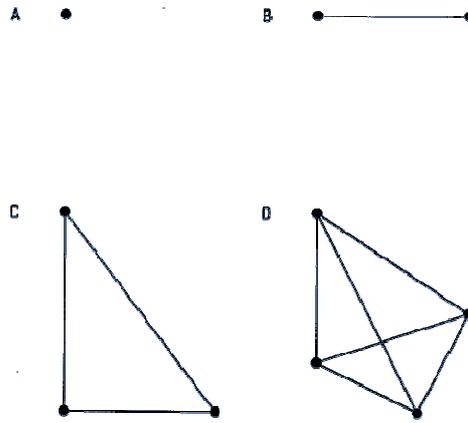
If  $Rem(t_1), \dots, Rem(t_n)$  are uniformly small in magnitude, polynomial regression provides a reasonable methods of analyzing the data. The basic premise is that the integral in equation (4) can be approximated using the quadrature formula  $\sum_{j=1}^k \delta_j (t - \xi_j)_+^{m-1}$  for coefficients  $\delta_1, \dots, \delta_k$  and points  $0 < \xi_1 < \dots < \xi_k < 1$ . Combining this with original polynomial approximation leads to an overall approximation of the general function by (Eubank, 1999):

$$S(x) = \sum_{j=1}^m \theta_j x^{j-1} + \sum_{j=1}^k \delta_j (x - \xi_j)_+^{m-1} \quad (5)$$

Another formulation which is used more often in practice is  $\sum_j w_j (y_j - f(x_j))^2$  where  $w_j$  is weight, and  $f(x_j) = \sum_{i=1}^m \delta_i (x_j - \xi_j)_+^{m-1}$  where  $\xi_j < x_j < \xi_{j+1}$ . The multivariate case can be obtained from univariate splines by the tensor product construct.

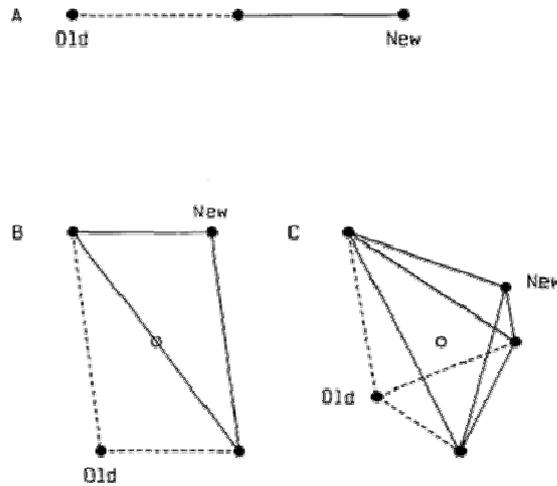
## 1.4 Sequential Simplex Optimization

A simplex is a geometric feature that has a number of vertices (corners) equal to one more than the number of dimensions in the factor space. Simplex can be defined for any number factors: for zero dimension it would be a dot, for one dimension it would be a straight line, for two dimension it would be a triangle, for three dimension it would be a tetrahedron and hyper tetrahedron for higher dimensions (See Figure 2).



**Figure 2 Simplexes in (A) zero-dimension, (B) one-dimension, (C) two-dimension and (D) three-dimension (Walters et al. 1991)**

Simplex can be moved into an adjacent area by rejecting one vertex (usually the vertex that gave the worst response) and projecting it through the average of remaining vertices to create a one new vertex on the opposite side of the simplex (See Figure 3) (Walters et al., 1991). This new vertex corresponds to a new set of experimental conditions that can then be evaluated.



**Figure 3 The simplex reflection move for (A) one-dimension (B) two-dimension, and (C) three-dimension factor spaces. Dashed line represents the old simplex. Open circle shows the average of the remaining vertices (Walters et al. 1991)**

Two fundamental ideas that should be remembered throughout simplex procedure are: (1) the simplex reflection is that of a point through point. It is not a mirror-image reflection across a line, plane or hyper plane. (2) The purpose of simplex is to move rapidly into the region of the optimum. Hence it can be very effective and efficient for this purpose. But when the simplex has located the region of the optimum, it becomes relatively inefficient.

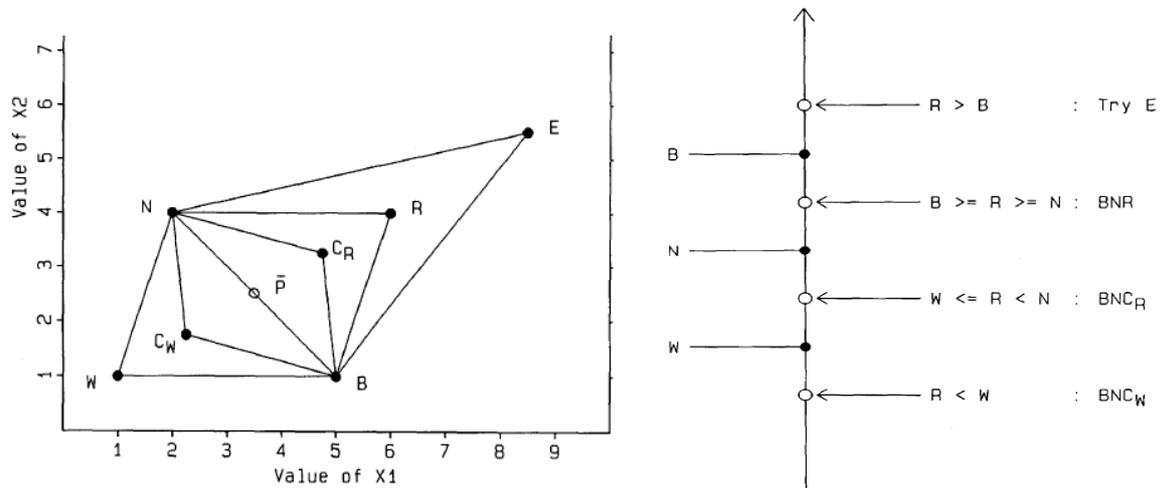
Nelder and Mead (1965) make two modifications to the original simplex algorithm of Spendley, Hex and Homsworth (1962) which allows the simplex to expand in the directions that are favorable and contract in the directions that are unfavorable. The Algorithm below and Figure 4 present the rules of variable-size simplex.

---

**Variable-Size Simplex Algorithm**


---

1. Rank the vertices of the first simplex and find the worst (W)
  2. Calculate the reflection of the worst (R) using Figure 4 as follows:
    - If  $N < R < B$  use simplex B...NR and go to 3
    - If  $R > B$ , calculate and evaluate E:
      - If  $E \geq B$  use simplex B...NE and go to 3
      - If  $E < B$ , use simplex B...NR and go to 3
    - If  $R < N$ 
      - If  $R \geq W$  calculate and evaluate  $C_R$  use simplex R...NC<sub>R</sub> and go to 3
      - If  $R < W$  calculate and evaluate  $C_W$  use simplex B...NC<sub>W</sub> and go to 3
  3. Transfer the current N (Never transfer the current W to the next iteration). rank the remaining retained vertexes in order of decreasing response
- 



**Figure 4 (a) Possible moves in the variable-size simplex algorithm, (b) logic of the possible moves in the variable-size simplex algorithm (Walters et al. 1991)**

### 1.5 Multiple Adaptive Regression Spline (MARS)

Multiple Adaptive Regression Spline (MARS) is a flexible regression modeling of high dimensional data introduced by Jerome Friedman in 1991 (Friedman, 1991). The model take the form of an expansion in product spline basis function, where the number of basis functions as well as the parameters associated with each one (product degree and knot locations) are

automatically determined by the data. The procedure is motivated by the recursive partitioning approach to regression and shares its attractive properties. Unlike recursive partitioning, however, this method produces continuous models with continuous derivatives. It has more power and flexibility to model relationships that are additive or involve interactions in at most a few variables. In addition, the model can be represented in a form that separately identifies the additive contributions and those associated with the different multivariate interactions.

MARS builds models of the form  $\hat{f}(x) = \sum_{i=1}^k c_i B_i(x)$  which is a weighted sum of basis functions  $B_i(x)$  and with constant coefficient  $c_i$ . Each basis function  $B_i(x)$  takes one of the following three forms (Friedman, 1991):

- *Constant* where is just one such term, the intercept
- *Hinge function* with the form  $\max(0, x - \text{const})$  or  $\max(0, \text{const} - x)$  where  $c$  is a constant, called the knot
- *Product of two or more hinge functions*, which can model interaction between two or more variables

MARS builds a model in two phases: the forward and the backward pass. This two stage approach is the same as that used by recursive partitioning trees.

In the forward pass, MARS starts with a model which consists of just the intercept term (which is the mean of the response values). MARS then repeatedly adds basis function in pairs to the model. At each step it finds the pair of basis functions that gives the maximum reduction in sum-of-squares residual error (it is a greedy algorithm). The two basis functions in the pair are identical except that a different side of a mirrored hinge function is used for each function. Each new basis function consists of a term already in the model (which could perhaps be the intercept)

multiplied by a new hinge function. A hinge function is defined by a variable and a knot, so to add a new basis function, MARS must search over all combinations of the following: (1) existing terms (called parent terms in this context); (2) all variables (to select one for the new basis function); and, (3) all values of each variable (for the knot of the new hinge function).

This process of adding terms continues until the change in residual error is too small to continue or until the maximum number of terms is reached. The maximum number of terms is specified by the user before model building starts.

The forward pass usually builds an overfit model. (An overfit model has a good fit to the data used to build the model but will not generalize well to new data.) To build a model with better generalization ability, the backward pass prunes the model. It removes terms one by one, deleting the least effective term at each step until it finds the best submodel. Model subsets are compared using the *GCV* criterion described below. The backward pass has an advantage over the forward pass: at any step it can choose any term to delete, whereas the forward pass at each step can only see the next pair of terms.

The backward pass uses *GCV* to compare the performance of model subsets in order to choose the best subset: lower values of *GCV* are better. The *GCV* is a form of regularization: it trades off goodness-of-fit against model complexity. The raw residual sum-of-squares (*RSS*) on the training data is inadequate for comparing models, because the *RSS* always increases as MARS terms are dropped. In other words, if the *RSS* were used to compare models, the backward pass would always choose the largest model. The formula for the *GCV* is:

$$GCV = \frac{RSS}{N \left(1 - \frac{ENOP}{N}\right)^2} \quad (6)$$

where  $RSS$  is the residual sum-of-squares measured on the training data and  $N$  is the number of observations and  $ENOP$  is the effective number of parameters which is defined in the *MARS* context as:

$$ENOP = P + C \frac{(P-1)}{2} \quad (7)$$

where  $P$  is the number of *MARS* terms and  $C$  is the penalty which is usually set about 2 or 3.

Note that  $\frac{(P-1)}{2}$  is the number of hinge-function knots, so the formula penalizes the addition of knots. Thus the *GCV* formula adjusts the training  $RSS$  to take into account the flexibility of the model.

## CHAPTER 2 THE PROPOSED STRATEGY FOR QUADRATIC MODELS WITH TWO VARIABLES (ASRSM)

### 2.1 Introduction

In this chapter, we discuss the elements of the proposed Adaptive Sequential Response Surface Methodology (ASRSM) for bivariate quadratic functions. This methodology presented here will be used as the basis of approaches proposed for considering more sophisticated models in the following chapters. We start the chapter with terminology and assumptions in Section 2.2. Next, we provide an overview of the methodology in Section 2.3, followed by explaining the two core strategies embedded in ASRSM; (1) Parametric approach in Section 2.4, and (2) Non-parametric approach in Section 2.5. In Section 2.6, we describe how these two strategies are integrated within ASRSM. Finally, in Section 2.7 we provide the result of several numerical examples conducted to evaluate the performance of the proposed strategy.

### 2.2 Terminology and Assumptions

The definitions and terminology used in the proposed ASRSM methodology is as follows. Some of the notation is illustrated in Figure 5 for a two-dimensional factor space with 5 experiments in each run:

$FS_r$  : Factor space at run  $r$  and expressed as Cartesian product of factor ranges in run  $r$

$fs_i$  : Initial range of factor  $i$

$COP$  : A corner point experiment run at the intersection of extrema of factor ranges

$CEP$  : Center point experiment run at the center of gravity of the factor space

$r$  : Index of runs, e.g.  $r = 1, 2, \dots, R$  where  $R$  is the total number of runs

$e$  : Index of experiments in a given run, e.g.  $e = 1, 2, \dots, E$  where  $E$  is the total number of experiments

$B$  : The experiment with the best response level in a given run

$N_k$  : The experiment with the  $k^{th}$  best response level in a given run ( $2ek \leq e - 1$ )

$W$  : The experiment with the worst response level in a given run

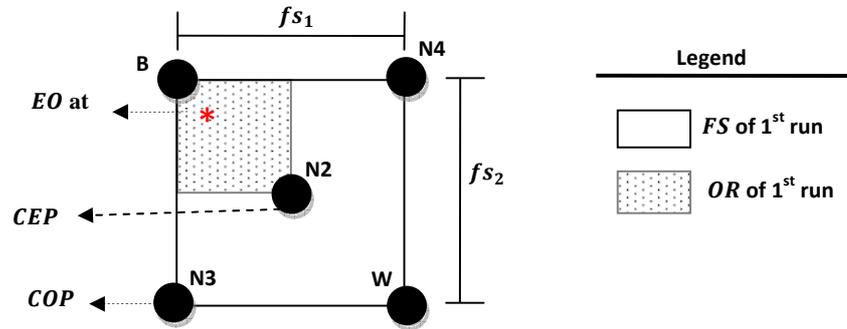
$OR_r$  : Optimal region in run  $r$  containing the estimated optimal experiment,  $OR_r \subseteq FS_r$

$O$  : Optimal experiment, e.g. best experiment in the initial factor space

$EO_r$  : Estimated optimal experiment in run  $r$ , e.g., best incumbent estimation of the optimal experiment

$BCE$  : Best at Center classification of the  $OR_r$  where the location of  $B$  is at  $CEP$

$BCO$  : Best at Corner classification of the  $OR_r$  where the location of  $B$  is at the corner of factor space



**Figure 5 An illustration of terminology of the ASRSM on a two dimensional factor space with  $E=5$**

As in most RSM approaches, the proposed ASRSM methodology relies on a number of simplifying assumptions. The extensions due to the relaxation of these assumptions are beyond the scope of this paper and some of these extensions discussed in the conclusion. For the proposed methodology we consider the following assumptions:

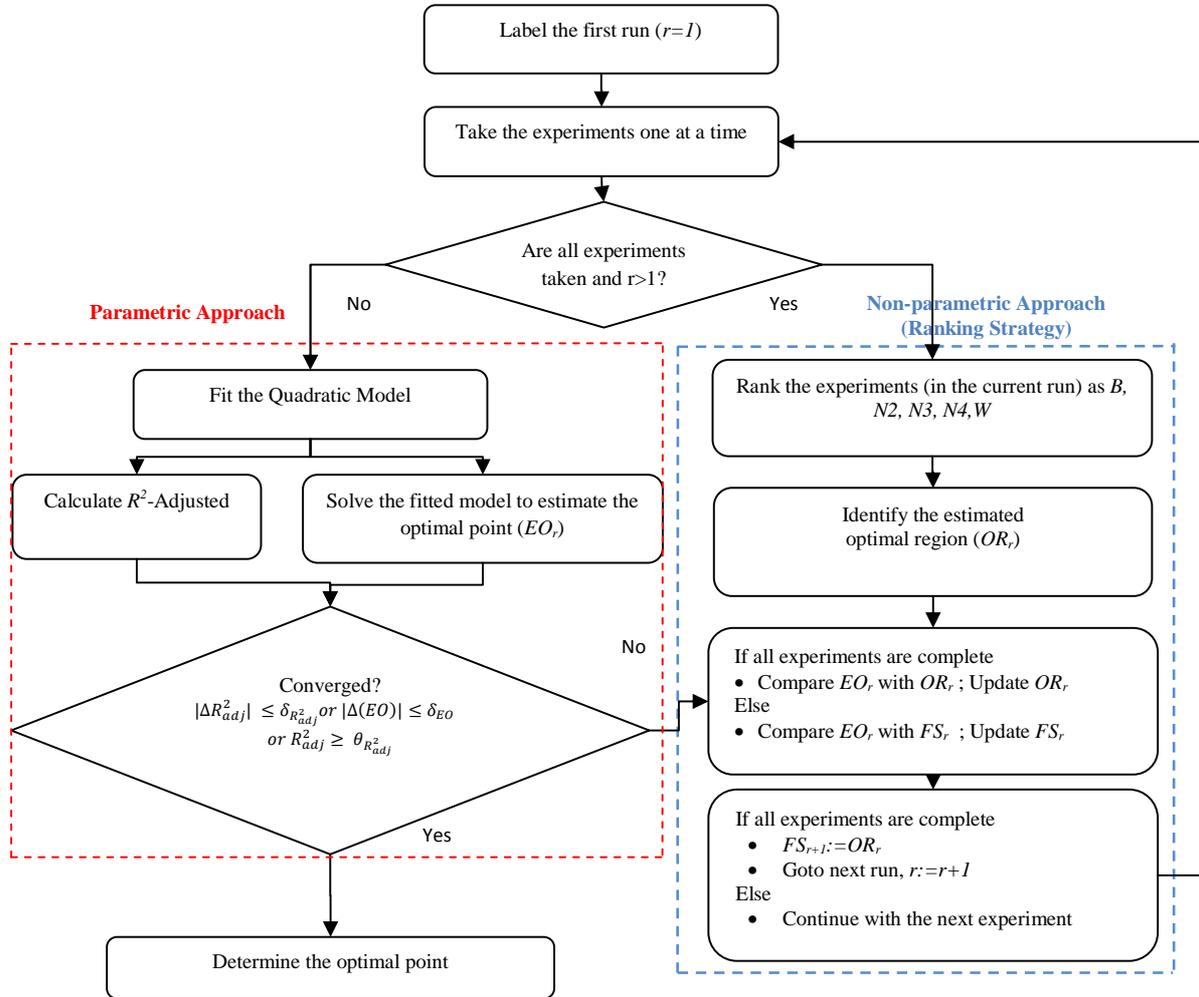
1. *There are two significant and controllable factors.*
2. *The underlying relation between a single response and two factors can be represented by a quadratic model.* RSM models are usually employed in a sufficiently small region around the optimal region. As a result, it is quite common in RSM applications to assume that the underlying model can be approximated via a quadratic function. Such assumption also holds for this study.

3. *The response is convex in the region of interest.* We assume that the region of interest is shifted close to the optimum a priori using an efficient method (e.g., steepest descent). Since for most of the nonlinear minimization problems, the underlying model is locally convex around the optimal solution, we assume that the response is convex in the region of interest. Our empirical test results show that the proposed approach is robust with respect to this assumption such that the proposed method is effective in cases where the underlying model is non-convex.
4. *The factor space in the region of interest is feasible.*

### 2.3 Algorithm and Initial Run Design

Figure 6 illustrates the structure of the proposed ASRSM methodology. The procedure is initialized with a region of interest, e.g., a feasible factor space which is guaranteed to contain the  $O$ . The goal is to reach to the vicinity of  $O$  with a finite set of runs ( $R$ ). Each run  $r$  is set up on a given factor space ( $FS_r$ ) with a specific experimental design ( $D$ ), e.g. a modified version of the factorial design augmented with a center point. The experiments in each run  $r$  are taken one at a time and the  $FS_r$  is not finalized until all experiments are taken. Once an experiment  $e$  is taken, the  $EO_r$  is obtained from the parametric model fitting using the all experiments in all runs. For all but last experiment in run  $r$  (i.e.,  $e \neq E$ ), the  $EO_r$  is tested for belonging to  $FS_r$ . Accordingly,  $FS_r$  is updated (e.g. expanded) if  $EO_r$  is outside  $FS_r$ . For the last experiment (i.e.,  $e = E$ ), the approach follows two concurrent strategies, e.g., non-parametric ranking strategy and parametric model fitting strategy. According to the ranking of experiments and  $EO_r$  from the quadratic model fitting, a reduced factor space containing the  $EO_r$  (i.e.,  $OR_r$ ) is determined for the next run. This procedure continues until the convergence criteria based on estimated optimal experiment or coefficient of determination of the fitted model is attained. The motivation for the dual strategy (e.g., parametric vs. non-parametric) is that, while the information from ranking

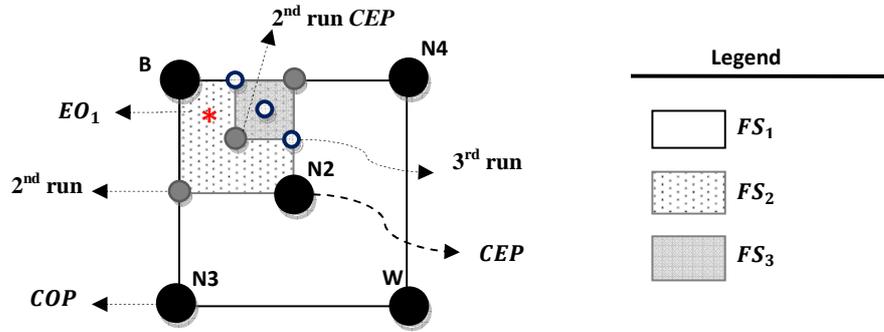
strategy is accurate but not precise, the information from model fitting is precise but not accurate.



**Figure 6 Scheme of the proposed ASRSM methodology**

The factor space of each run ( $FS_r$ ) can be expressed as a mapping ( $\varphi_r$ ) of the factor space of the preceding run ( $FS_{r-1}$ ). In most general form, the proposed methodology generates a series of factor spaces which are nested, e.g.  $FS_r = \varphi_r(\varphi_{r-1}(\dots\varphi_0(FS_1)))$ . The output of this mapping  $\varphi_r$  depends on the current factor space, the experimentation design ( $D$ ), the outcome of ranking of experiments as well as the result of parametric strategy described in the next subsection. The

latter two, the ranking and the parametric strategies, are described in the next two sections, respectively. Before discussing the  $D$  used in each run and the initial factor space, we briefly present the algorithm using the illustration in Figure 7 for a special case. The proposed approach is initialized with  $FS_1$  and the indicated five experiments ( $B, N2, N3, N4, W$ ) are taken from the corresponding design  $D$ . Once the responses are ranked, the non-parametric ranking strategy identifies the  $OR_1$ . Next the parametric model fitting approach determines the  $EO_1$  using the first five experiments. Lastly, the  $OR_1$  and  $EO_1$  are compared to determine the new factor space ( $FS_2$ ). Note that the design in  $r=2$  inherits two experiments from the first run, namely  $B$  and  $N2$ .

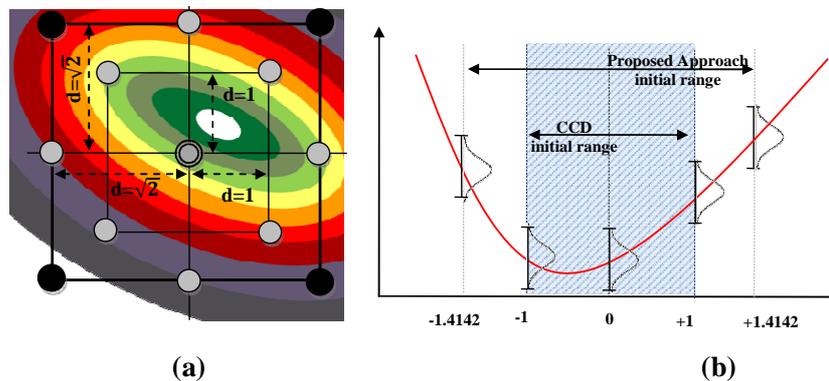


**Figure 7 Illustration of the factor space reduction across runs  $r=1, 2,$  and  $3$**

The proposed ASRSM method uses the same  $D$  in each run which is the factorial design augmented with a center point. Hence we maintain the same experimental design  $D$  and consider a constant number of experiments (e.g.  $E=5$ ) throughout the process. In practice, none of the existing methods for setting the initial point in sequential optimization procedures is superior to the corner initial point as in factorial design (Walters, Parker, Morgan, and Deming, 1991). Furthermore, experiments conducted on corner points benefit from fractional factorial design, especially when the design is orthogonal. In particular, the designs maximize the amount of information gained from each experiment. On the other hand central points are essential for the

modeling the curvature of the underlying function (Montgomery, 2008). Lastly, the five experiments in  $r=1$  are not sufficient to estimate the full quadratic response model. Hence, we estimate the  $EO_1$  by fitting a quadratic response model without the constant term.

The design of the initial factor space in the proposed approach is adapted such that a rational comparison with traditional RSM methods (e.g. CCD) is possible. In the traditional CCD approach, the corner points are taken at  $\pm h$  unit distance from the center point (0,0). In comparison, the proposed methodology starts with a broader initial region around the center point, e.g. at  $\pm\sqrt{2}$  unit distance from the center. Figure 8a illustrates the initial factor space of the traditional CCD and the proposed method with light and dark experiment points, respectively. While beginning with a larger space is initially disadvantageous, experimental results demonstrate that the reduction in the factor space with the same number of experiments far exceeds initial difference. An additional benefit is that this modification may decrease the effect of random error on the initial results. Let's consider the diagonal cross-section of these two designs as illustrated in Figure 8b and assume that the noise is identically distributed on this cross-section. Then, it can be shown that the impact of the noise on prediction of the optimal experiment point is less with the proposed methodology's factor space.



**Figure 8 (a) Initial factor space and design structure and (b) Diagonal cross-section of the traditional CCD and proposed ASRSM approach**

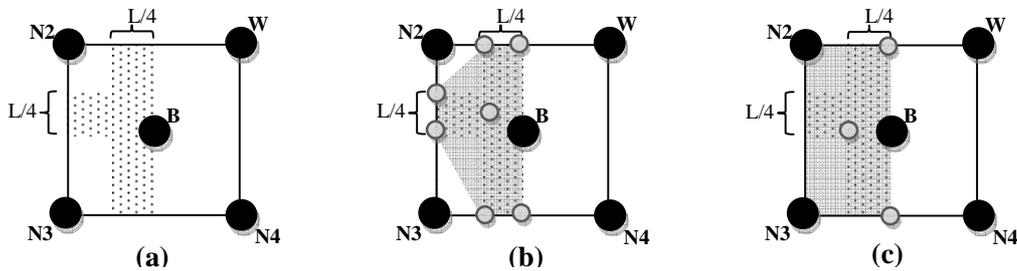
## 2.4 Non-parametric Approach: Ranking Strategy

At each run  $r$  of the proposed ASRSM approach, we first rank the 5 experiments (e.g., 4 factorial and 1 center) as  $B, N2, N3, N4$  and  $W$  according to their response levels. Based on the ranking, we identify the implied optimal region which contains the  $EO_r$ . This region is a polygon contained in  $FS_r$  and can be convex or non-convex in the space of factors. We then identify a rectangle which contains the implied optimal region and denote it as the optimal region ( $OR_r$ ), which determines the factor space of the next run.

This process of encapsulating the implied optimal region with a rectangle is a form of relaxation and is not efficient in terms of factor space reduction. However, there are valid reasons which motivate this relaxation. The foremost reason is the reduced need for new experiments due to the inheritance of experiments from the previous run. Secondly, the rectangular  $FS$  preserves the orthogonality of factorial experimental design. Further, this rectangular form facilitates the recursive characterization of the same rectangular structure throughout the process. In addition, we can use the same experimental design structure, e.g. full factorial with a center point. Specifically, with rectangular envelope, the mapping across runs will be identical, e.g.  $\varphi(\cdot) \equiv \varphi_r(\cdot)$  for  $\forall r$ . This is because we maintain the same experiment design structure and there is a finite number of optimal regions as a result of ranking outcomes. Lastly, the relaxation reduces the risk of selecting an optimal region which excludes the optimal experiment.

An alternative to the rectangular envelope is the convex hull of implied optimal region. Due to its convexity, it also allows for easier tessellation of the  $FS$ . While the convex hull reduces the optimal region more than the rectangular envelope, it does not reduce the number of new experiments as much. Furthermore, the experimental design used in each run will be different

since the convex hulls of the implied optimal regions will vary in shape. Clearly the choice of the right form is a trade-off between the rate of contraction of the optimal region and the total number of experiments conducted. To better illustrate this trade-off, let's consider the implied optimal region in Figure 9a. The convex hull of this implied optimal region is identified in Figure 9b with six vertices (corner points). In contrast, we adopted the rectangular envelope which is illustrated in Figure 9c. Comparison between Figure 9b and 9c reveals that, while convex hull based *OR* leads to the greatest factor space reduction, it also leads to an increased number of new experiments (7 vs. 3 new experiments) and cannot inherit experiments from previous runs. Note that it is not practical to change the design and choose only 3 new experiments (e.g. 2 vertices and one at the center of gravity) for the convex hull in Figure 9b. This is because we assume that the *O* is contained in the current factor space, and, by choosing fewer number of vertices, we would then be implicitly reducing the implied optimal region.

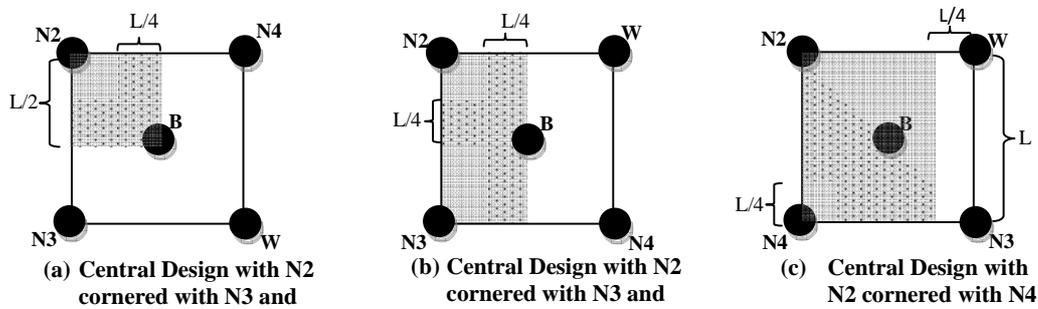


**Figure 9 (a) Implied optimal region (b) Convex hull envelope of the implied optimal region (c) Rectangular envelope of the implied optimal region based on a two dimensional factor space**

In what follows, we present the optimal region alternatives based on the ranking of the experiments and the location of *B* in the current *FS*.

### 2.4.1 Best at Center (BCE) Optimal Regions

An important information obtained from the ranking of experiments is the location of  $B$ . When the  $B$  is located at the center, the current  $FS$  is then classified as having a  $BCE$  optimal region. Depending on the location of  $N2, N3, N4$ , and  $W$ , there are three possible  $OR$ s as illustrated in Figure 10. We first determine the implied optimal region as illustrated as dotted regions in Figure 10. Next we characterize the  $OR$  as the rectangle which contains the this implied optimal region.



**Figure 10 BCE ORs (dotted region: impliedOR, shaded region:OR)**

The implied optimal regions are guaranteed to contain the optimal experiment in the absence of random noise. The mathematical proofs of the optimality of these implied optimal regions is involved and thus excluded. Instead, we provide a general proof sketch of the rectangular optimal regions and illustrate it for the  $OR$  in Figure 10a. The proofs are accomplished through the following steps: (1) Divide the non-optimal region into smaller rectangular sub-regions using factor centerlines; (2) Assume that the optimal point falls in one of these sub-regions; (3) Relocate the origin to that region and formulate the responses at  $B, N_1, \dots$ , and  $W$  based on their displacement from the new origin; (4) Show that at least one pairwise comparison of the responses violate the initial ranking (5) Replicate the steps (2-5) until all sub-regions are evaluated. The proof of  $OR$  in Figure 10a is as follows.

**Proposition:** For BCE optimal region with ranking in Figure 10a, the optimal experiment is located in the optimal region characterized as the quadrant with corners at B and N2 when there is no random noise.

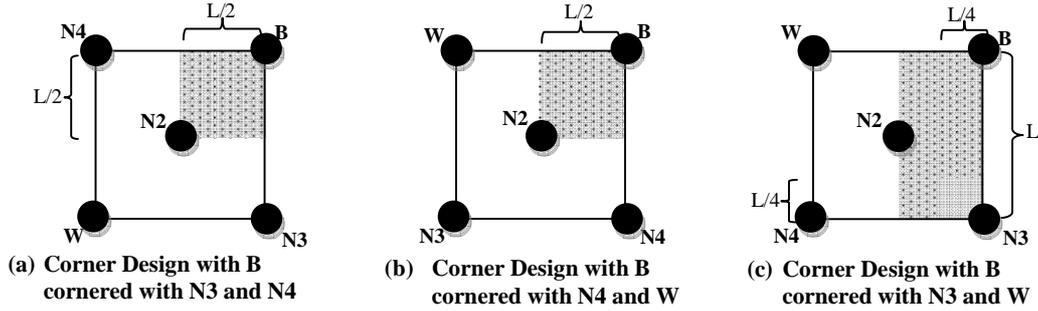
**Proof.** Consider that the FS is divided into equal quadrants (I,II,III,IV) which have (N2,N4,W,N3) as the corners, respectively. Further suppose that O is located in FS outside the OR. For the case, where O is in II, we consider the responses at  $N2 \equiv (x_{N2}, y_{N2})$ ,  $N3 \equiv (x_{N3}, y_{N3})$  and  $N4 \equiv (x_{N4}, y_{N4})$  as  $z_{N2}$ ,  $z_{N3}$  and  $z_{N4}$ , respectively. We have  $z_{N2} = A(dx_2)^2 + B(dy_2)^2 + C(dx_2)(dy_2)$ ,  $z_{N3} = A(dx_3)^2 + B(dy_3)^2 + C(dx_3)(dy_3)$ , and  $z_{N4} = A(dx_4)^2 + B(dy_4)^2 + C(dx_4)(dy_4)$  where  $(dx_2, dy_2) = (x_{N2} - x_O, y_{N2} - y_O)$ ,  $(dx_3, dy_3) = (x_{N3} - x_O, y_{N3} - y_O)$  and  $(dx_4, dy_4) = (x_{N4} - x_O, y_{N4} - y_O)$  and optimal experiment location  $O \equiv (x_O, y_O)$ . We consider  $dz_{N2N4} = z_{N2} - z_{N4} = A[(dx_2)^2 - (dx_4)^2] + B[(dy_2)^2 - (dy_4)^2] + C[(dx_2)(dy_2) - (dx_4)(dy_4)]$ , and  $dz_{N3N4} = z_{N3} - z_{N4} = A[(dx_3)^2 - (dx_4)^2] + B[(dy_3)^2 - (dy_4)^2] + C[(dx_3)(dy_3) - (dx_4)(dy_4)]$ .

Since the response is convex (e.g.,  $A, B > 0$ ), we consider three response scenarios:  $C=0$ ,  $C<0$ ,  $C>0$ . Note that when O is in II, we have  $|dx_2| > |dx_4|$ ,  $dx_2 \leq 0$  and  $dy_2 = dy_4$  making second term  $dz_{N2N4}$  zero. For  $C=0$ , we have the first term in  $dz_{N2N4}$  positive, thus  $dz_{N2N4} > 0$  which is a contradiction to the ranking  $z_{N2} < z_{N4}$ . For  $C<0$ , the third term in  $dz_{N2N4}$  is positive since  $dx_2 \leq 0$  thus making  $dz_{N2N4} > 0$  which is also a contradiction. Lastly, for  $C>0$ , first and second terms in  $dz_{N3N4}$  are positive since  $|dx_3| > |dx_4|$  and  $|dy_3| > |dy_4|$ . Last term in  $dz_{N3N4}$  is also positive since  $dx_3 dy_3 > 0$  and  $|dx_3 dy_3| > |dx_4 dy_4|$ . Thus  $dz_{N3N4} > 0$  which is a contradiction to the ranking  $z_{N3} < z_{N4}$ . For the case, where O is in III, we consider the responses at N2, N3 and W as  $z_{N2}$ ,  $z_{N3}$  and  $z_W$ , respectively. Let's define the  $dz_{N3W}$ ,  $dz_{N2W}$ , and  $dz_{N2N3}$  as before. For case  $C=0$ , it can be shown that  $dz_{N3W} > 0$  which is a contradiction for  $z_{N3} < z_W$ . Similarly, for  $C<0$  and  $C>0$ , we have  $dz_{N2W} > 0$  and  $dz_{N2N3} > 0$  are contradictions for

$z_{N2} < z_W$  and  $z_{N2} < z_{N3}$ . Last case is where  $O$  is in IV. We consider the responses at  $N2, N3, N4$  and  $W$ . Let define  $dz_{N4W}$  as before. For case  $C=0$ , it can be shown that  $dz_{N2N3} > 0$  which is a contradiction for  $z_{N2} < z_{N3}$ . Similarly, for  $C<0$  and  $C>0$ , we have  $dz_{N2N3} > 0$  and  $dz_{N4W} > 0$  are contradictions for  $z_{N2} < z_{N3}$  and  $z_{N4}$ .

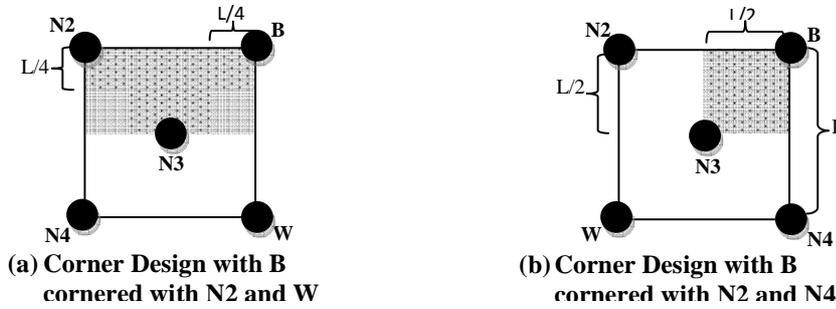
### 2.4.2 Best at Corner (BCO) Optimal Regions

The case when the  $B$  is located at a corner is referred as a  $BCO$  optimal region. In  $BCO$ , either  $N2$  or  $N3$  can occur at the center. For  $N2$  at center, there are three possible  $ORs$  based on the location of  $B, N3, N4$ , and  $W$  (Figure 11).



**Figure 11  $BCO$   $ORs$  when  $N2$  is at center (dotted region: implied  $OR$ , shaded region:  $OR$ )**

In case with  $N3$  at center, there are two possible  $ORs$  based on the location of  $B, N2, N4$ , and  $W$  (Figure 12).



**Figure 12  $BCO$   $ORs$  when  $N3$  is at center (dotted region: implied  $OR$ , shaded region:  $OR$ )**

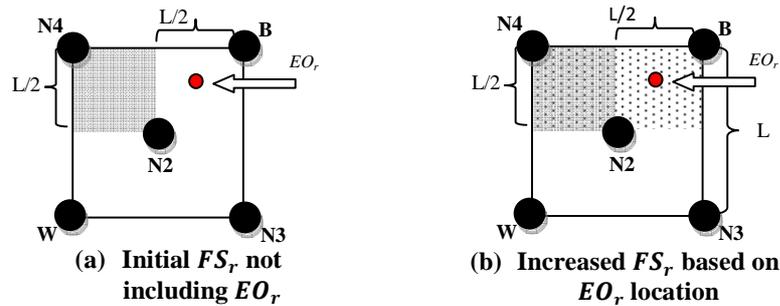
The proving strategy for *BCOs*' *ORs* is the same as *BCE* and is thus excluded. Note that the implied optimal regions are identical to the *ORs* in Figures 8a, 8b, and 9b, thus there is no relaxation due to rectangular envelope.

## 2.5 Parametric Approach: Model Fitting Strategy

We use a parametric approach based on model fitting in addition to the ranking approach. This strategy not only allows us to increase the precision of *EOr* but also supports backtracking through *FSr* correction as explained in previous section. Beginning with the completion of all first run experiments, this parametric approach is used after each experiment. In this approach we fit a quadratic model  $z = [a_1 \quad a_2] \begin{bmatrix} x \\ y \end{bmatrix} + [x \quad y] \begin{bmatrix} q_1 & q_2 \\ q_2 & q_3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + c + u$ , with  $\varepsilon \sim N(0, \sigma^2)$ , to the experimental data to analyze the underlying function of data and efficacy of conducted experiments. In fitting the quadratic model, two objectives are being sought in particular: (1) estimating the estimated optimal experiment *EOr*; (2) calculating the adjusted coefficient of determination ( $R_{adj}^2$ ). *EOr*, the minimum of the fitted model, not only shows the predicted optimal solution, but can also be used for correcting the *FS* of the next run. Furthermore, the change in the *EOr* in consecutive runs is also used as a stopping criterion. In comparison, the  $R_{adj}^2$  shows how well the information gained from the experiments explain the behavior of the underlying system (Seber and Alan, 2003). We also use this measure as a stopping rule in the proposed *ASRSM* methodology and for comparing the explanatory power of different methods.

## 2.6 Design Structure for the Next Runs

Following the characterization of  $OR_r$  through ranking approach and estimation of the  $EO_r$  from parametric approach, we determine the design structure for the next run. In particular, we compare the  $EO_r$  from the model fitting with the  $OR_r$  from the experiment ranking. If  $EO_r$  is contained in the  $OR_r$ , then we use the region as the factor space of the next run. If  $EO_r$  is not contained in the  $OR_r$ , we then expand the optimal region to a larger rectangle envelope containing the  $EO_r$  and use the region as  $FS_r$  of the next run (Figure 13). Next we conduct experiments on the un-experimented corners and the center of the new  $FS_r$ . After each experiment, we fit the quadratic model and check whether  $EO_r$  is contained in  $OR_r$ . If  $EO_r$  is outside  $OR_r$ , then we expand the  $OR_r$  as before. This expansion serves as a backtracking step. These steps are repeated monitor the change in  $R_{adj}^2$  and the  $EO_r$  using the fitted model. The stopping condition for the proposed  $ASRSM$  approach is the convergence of  $R_{adj}^2$  or  $EO_r$  with thresholds  $\delta_{EO}$  and  $\delta_{R_{adj}^2}$ .



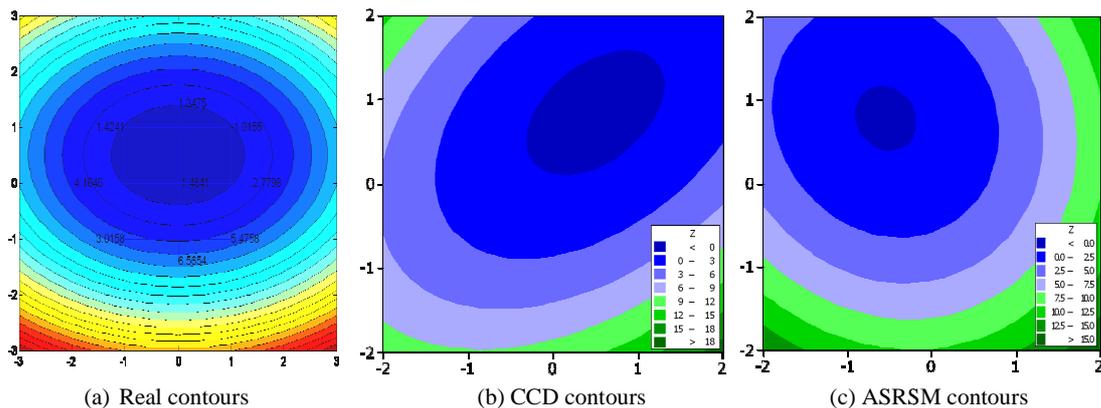
**Figure 13** Expansion of the  $OR_r$  when the  $EO_r$  from model fitting falls outside

## 2.7 Illustrative Example, Simulated Experiments and Case Studies

In this section, we first illustrate the proposed ASRSM approach using a stylized example and compare its performance with the traditional CCD approach. Next, we report on the results of extensive simulation experiments comparing the proposed ASRSM, CCD and three optimal designs using different response models. We then experiment with these approaches using the well-known paper helicopter experiment. Lastly, we report on the results of a rat brain trauma case study comparing ASRSM and CCD approaches.

### 2.7.1 Illustrative Example

We consider the quadratic response model of the form,  $Z = X^2 + 2Y^2 - 2Y + \varepsilon$  with  $\varepsilon \sim N(0, 2^2)$  which is desired to be minimized. The starting region of interest is selected as  $X \in [-3, 3]$  and  $Y \in [-3, 3]$  and the contour plot of the response is presented in Figure 14a. We first conducted a typical CCD with 13 experiments centered at (0,0) and contains the optimal experiment  $O=(0,0.5)$  with mean response  $ZO=-0.5$  (Figure 14a).



**Figure 14** The (a) Actual, (b) CCD, and (c) ASRSM estimated contours of the response  $Z$

Based on 13 experiments, the CCD attains  $R_{adj}^2 = 60.86\%$  with  $(X^*, Y^*) = (0.5, 0.8428)$  as estimation of the optimal experiment. Figure 14b illustrates the estimated contours using CCD design. The reason that CCD could not estimate the orientation of the quadratic response is the large magnitude of the variance of error term in the quadratic surface equation.

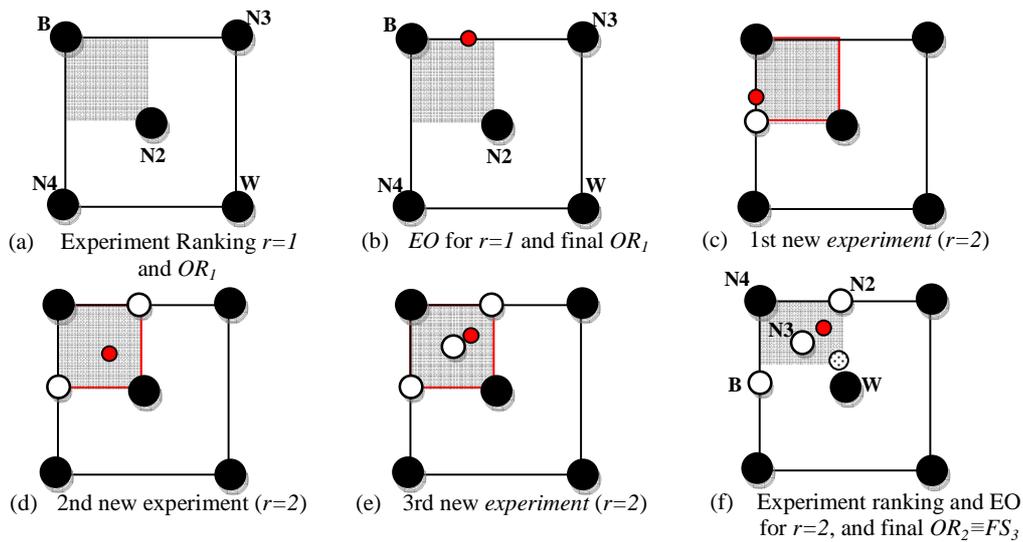
Next we employ the ASRSM and present the results in Table 1. The first 5 rows correspond to the initial run design. Note that the ASRSM is setup as described in previous Section and without any additional information than used in the CCD. We use  $R_{adj}^2 \geq d\theta_{R_{adj}^2} = 85\%$  as the convergence criteria in this example. We now describe each run in detail.

**Table 1 The Runs and experiments of the ASRSM method**

Run No	No of New Experiment	Factor Combination		Response
		X	Y	Z
1	1	-1.4142	1.4142	8.2950
	2	1.4142	-1.4142	10.2316
	3	-1.4142	1.4142	2.1963
	4	1.4142	1.4142	6.8961
	5	0	0	2.2137
2	6	-1.4142	0	-0.4552
	7	1.4142	0	-0.1682
	8	-0.7071	0.7071	0.8719
3	9	0	.3535	0.4117

**Run 1:** Given the initial design, we obtain the  $OR_1$  using the non-parametric approach (Figure 15a). The constrained quadratic fit estimates the optimal experiment ( $EO_1$ ) as  $(-0.3572, 1.4142)$  illustrated with a small point on the edge  $B-N3$  (Figure 15b). Note that since the number of experiments is not sufficient to estimate the full model, the  $EO_1$  is estimated by the quadratic response model without the constant term. Since the  $EO_1$  is contained in the  $OR_1$ , the  $OR_1$  is final. The new factor space  $FS_2$  is determined as  $OR_1$ .

**Run 2:** Figure 15c shows the location of the first new experiment in  $r=2$  which corresponds to  $(-1.4142, 0)$ . Using the constrained quadratic fit as before we estimate the optimal experiment as  $(-1.4142, 0.5930)$  illustrated with a small point (Figure 15c). Since this experiment is contained in  $FS_2$ , there is no update of the factor space. The second new experiment is illustrated in Figure 15d and the corresponding estimate of the optimal experiment  $(-0.7, 0.6428)$  with  $R_{adj}^2 = 51.96\%$ . This estimated optimum is still contained in  $FS_2$ . The third and final experiment of  $r=2$  is shown on Figure 15e together with the estimated optimal experiment  $(-.5571, 0.7857)$  and  $R_{adj}^2 = 84.44\%$ .



**Figure 15 Illustration of the steps of ASRSM for runs 1 and 2**

The second run of the experiment terminates with the estimated optimal experiment in Figure 15e. Continuing to the third run, the final  $OR_2$  and  $FS_3$  are illustrated in Figure 15f. The first experiment in  $r=3$  is the corner point of  $FS_3$  indicated with dotted point. This point is the last experiment in Table 1. The estimated optimal experiment with this experiment is  $(0.5571, 0.7851)$  with  $R_{adj}^2 = 89.06$  which satisfies the termination criteria. The Figure 14c shows

the estimated contours using the proposed approach. Clearly, the response model estimation based on ASRSM is better than CCD.

### 2.7.2 Simulated Experiments

We now describe the simulated experiments performed to compare the performance of the proposed ASRSM approach with those of CCD, A- D- and V-optimal designs. In the simulated experiments, we have considered six response models with varying variance of error and function type (i.e., convex, non-convex). These response models are presented in Table 2. All response models have a quadratic relation, e.g.  $= AX^2 + BY^2 + CX + DY + EXY + F + b$ , with a normal error term  $\varepsilon \sim N(0, \sigma^2)$ .

**Table 2 The response models used in the simulated experiments**

Exp. No.	Response Relation	Error ( $\varepsilon$ )	Response Type	Optimal Experiment	
				$O=(X_o, Y_o)$	$Z_o$
1	$Z=-2x^2+3y^2+2x-y+2xy-1+\varepsilon$	$N(0,0.0.1)$	Nonconvex	(-3.0,1.25)	-29.063
2	$Z=x^2+2y^2-2y+\varepsilon$	$N(0,1)$	Convex	(0.0,0.5)	-0.500
3	$Z=-2x^2+3y^2+2x-y+2xy-1+\varepsilon$	$N(0,2)$	Nonconvex	(-3.0,1.25)	-29.063
4	$Z=-3x^2+2y^2+x-2y+2xy-1+\varepsilon$	$N(0,2)$	Nonconvex	(-3.0,2.0)	-39.000
5	$Z=2x^2+y^2+x+2xy+\varepsilon$	$N(0,2)$	Convex	(0.5,-0.5)	0.750
6	$Z=x^2+2y^2-2y+\varepsilon$	$N(0,2)$	Convex	(0.0,0.5)	-0.500

Whereas the ASRSM is an adaptive sequential method, the CCD, A-, D-, and V-optimal designs are essentially preset designs. In order to understand the effect of this difference, we have carried out two sets of analyses. In the first set, we have fixed the number of observations for each approach and compared the performances in terms of average  $R_{adj}^2$  and average optimality gap (i.e., deviation from the optimal response). Given the optimal response ( $Z_o$ ), the optimality gap of a method is defined as  $(\hat{Z} - Z_o)/Z_o$ , where  $\hat{Z}$  is the mean predicted response

at the estimated optimal point. All simulated experiments are repeated five times and average results are reported. For each response model in Table 2, the design points in A-, D- and V-optimal designs are generated by optimizing the optimality criteria over the starting factor space with a fine grid system spaced with 0.01 intervals. The CCD design consists of 4 corners, 4 axial and 1 center design points as in the illustrative example. Note that, the starting factor space expands the initial region of interest,  $X [-3,3]$  and  $Y [-3,3]$ , by a factor of  $\sqrt{2}$  in all directions as explained before.

Table 3 presents the results of the first set of analyses. The ASRSM has a better  $R_{adj}^2$  in 4 out of the 6 response models and has  $R_{adj}^2 = 94.5\%$  on the average. The A- and D-optimal designs are also competitive and CCD has the worst average performance. With respect to the optimality gap, the proposed ASRSM has the best performance in all but one of the response models with an average gap of 21.2%. The results indicate that increasing the variance of the response decreases the  $R_{adj}^2$  for all approaches, e.g. response models 1 versus 3 and models 2 versus 6. However, this reduction is least with ASRSM. In the case of optimality gap, the increase in the variance of response increases (decreases) the optimality gap in convex (non-convex) response models. While the generalization of these effects requires further analysis, we note that the increase (decrease) in the optimality gap is least (most) with the proposed ASRSM. These results show that the ASRSM is competitive with the three optimal designs and outperforms the classical CCD design. In addition, the performance of ASRSM is more robust with respect to changes in the error variance and convexity of the response model.

**Table 3 The  $R_{adj}^2$  and optimality gap results of simulated experiments with 9 observations determined a priori for CCD and optimal designs**

Exp. No.	Adjusted $R^2$					Optimality Gap				
	ASRSM	CCD	A-Opt.	D-Opt.	V-Opt.	ASRSM	CCD	A-Opt.	D-Opt.	V-Opt.
1	<b>99.98%</b>	99.95%	99.92%	99.94%	99.94%	<b>19.1%</b>	105.9%	19.1%	73.4%	92.3%
2	96.62%	92.00%	<b>96.97%</b>	93.04%	94.52%	19.0%	100.2%	11.7%	432.4%	<b>8.9%</b>
3	92.91%	88.48%	94.30%	<b>93.04%</b>	86.64%	<b>1.4%</b>	77.3%	<b>1.4%</b>	3.1%	84.4%
4	<b>94.17%</b>	82.17%	80.59%	88.96%	85.17%	<b>19.1%</b>	107.0%	<b>19.1%</b>	<b>19.1%</b>	15.9%
5	<b>94.26%</b>	37.03%	80.14%	82.38%	81.46%	<b>45.2%</b>	1119.7%	822.6%	100.0%	139.1%
6	<b>88.92%</b>	49.40%	84.63%	75.88%	64.98%	<b>23.2%</b>	312.9%	592.6%	738.9%	725.0%
Ave.	<b>94.48%</b>	74.84%	89.43%	88.87%	85.45%	<b>21.2%</b>	303.8%	244.4%	227.8%	177.6%

According to Table 3, the ASRSM's average performance improvement over other methods is more significant in optimality gap than in  $R_{adj}^2$ . This is because the ASRSM searches for the optimal design point by sequentially contracting the factor space whereas other approaches select the design points using the initial factor space. Hence, the design points used in ASRSM are more densely distributed than other methods. In order to capture this difference, we have carried another set of analyses for the optimal designs. In this second set, we initially fixed the number of design points at 7 and then incrementally added one design point at a time until we have a total of 9 design points. The initial set of 7 design points is optimally generated as before. Next, each of the additional point is generated by optimizing the optimality criterion given the existing design points and the response model. At each run, we have compared the performances in terms of average  $R_{adj}^2$  and average optimality gap. Note that we have used the earlier results of the ASRSM for consistency. For the CCD, we initially used 7 of the earlier observations by excluding two axial points and then include back in one at a time.

The results of the  $R_{adj}^2$  are presented in Table 4. For ASRSM and CCD, the inclusion of additional design points increases the average  $R_{adj}^2$  across all response models. In comparison, the average  $R_{adj}^2$  decreases with additional design points for the A- and D-optimal designs. Tables 3 and 4 show that applying A- and D-optimal designs sequentially reduces the average  $R_{adj}^2$ . The performance of the V-optimal design is observed to increase. Table 5 presents the optimality gap results of the final 9 design points in Table 4. Tables 3 and 5 show that the optimality gap of A- and D-optimal designs have improved when applied in sequence. In contrast, the optimality gap of V-optimal design has slightly decreased when applied in sequence. The results in Tables 4 and 5 also show that the ASRSM method exhibit monotonous behavior with respect to increasing the number of observations and is thus more suitable for sequential experimentation.

**Table 4 The  $R^2_{adj}$  results of simulated experiments beginning with 7 design points and then incrementally adding one design point at a time**

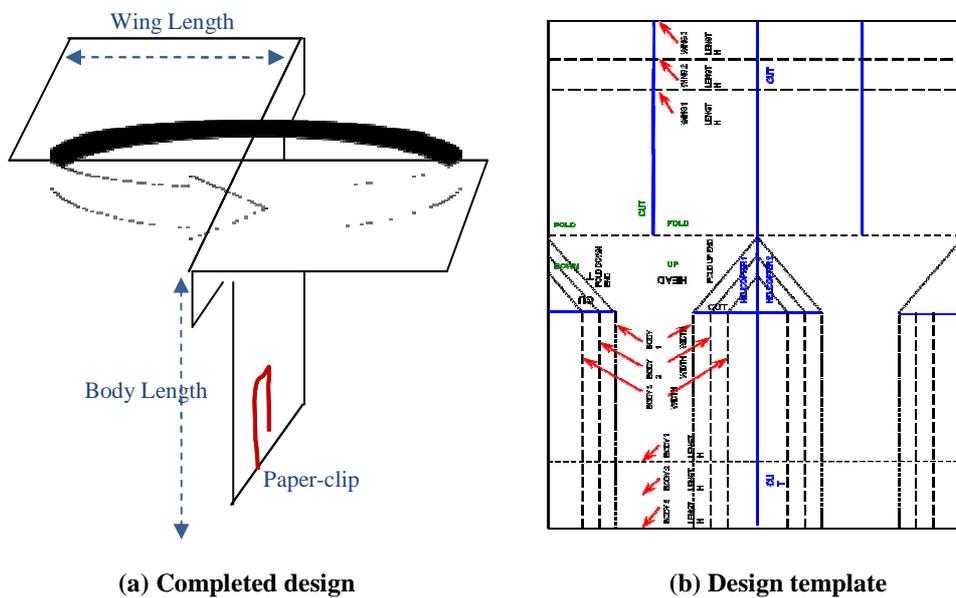
Exp. No.	No Obs.	Adjusted $R^2$				
		ASRSM	CCD	A-Opt.	D-Opt.	V-Opt.
1	7	<b>99.99%</b>	99.96%	99.94%	99.96%	99.92%
	8	<b>99.99%</b>	99.96%	99.95%	99.98%	99.95%
	9	<b>99.98%</b>	99.95%	99.92%	99.97%	99.94%
2	7	93.67%	92.69%	90.69%	<b>95.01%</b>	94.85%
	8	94.97%	92.48%	86.86%	89.20%	<b>95.86%</b>
	9	<b>96.62%</b>	92.00%	86.58%	90.42%	95.42%
3	7	79.57%	69.77%	<b>97.71%</b>	92.96%	72.11%
	8	86.53%	70.60%	79.86%	<b>91.06%</b>	74.92%
	9	<b>92.91%</b>	88.48%	82.00%	89.02%	80.89%
4	7	91.68%	92.87%	<b>94.21%</b>	91.09%	85.12%
	8	<b>92.39%</b>	88.49%	89.27%	91.69%	88.76%
	9	<b>94.17%</b>	82.17%	89.49%	92.58%	85.47%
5	7	<b>90.15%</b>	24.61%	87.19%	87.21%	86.05%
	8	89.16%	35.75%	<b>89.97%</b>	76.74%	71.97%
	9	<b>94.26%</b>	37.03%	86.66%	67.96%	84.95%
6	7	<b>86.18%</b>	35.45%	50.38%	85.31%	85.08%
	8	<b>86.44%</b>	43.17%	57.23%	78.71%	86.04%
	9	<b>88.92%</b>	49.40%	64.49%	67.65%	82.05%
Ave.	7	90.21%	69.22%	86.69%	<b>91.92%</b>	87.19%
	8	<b>91.58%</b>	71.74%	83.86%	87.90%	86.25%
	9	<b>94.48%</b>	74.84%	84.86%	84.60%	88.12%

**Table 5 The optimality gap results of simulated experiments for the final 9 design points in Table 4**

Exp. No.	Optimality Gap				
	ASRSM	CCD	A-Opt.	D-Opt.	V-Opt.
1	19.1%	105.9%	<b>14.0%</b>	45.9%	140.2%
2	19.0%	100.2%	19.0%	<b>17.6%</b>	48.5%
3	<b>1.4%</b>	77.3%	<b>1.4%</b>	3.1%	363.4%
4	<b>19.1%</b>	107.0%	<b>19.1%</b>	<b>19.1%</b>	<b>19.1%</b>
5	<b>45.2%</b>	1119.7%	101.2%	139.0%	434.5%
6	<b>23.2%</b>	312.9%	779.0%	47.7%	95.4%
Ave.	<b>21.2%</b>	303.8%	155.6%	45.4%	183.5%

### 2.7.3 Paper Helicopter

Paper helicopter problem is a simple practical experiment which is frequently used for teaching as well as testing for different methods. Paper helicopter problem consists of studying the effect of a number of factors, i.e. the wing length, body length and etc, on the flying time of a paper helicopter (Figure 16). Using this practical experiment, we now compare the performances of ASRSM, CCD, A-, D-, and V-optimal designs.



**Figure 16 Paper helicopter**

The paper helicopter problem is originally designed for the design of experiments. In order to apply alternative approaches to this problem, we first extended it to the RSM domain. For this, we first chose the wing length and the body length as the two controllable factors under study. Next, we conducted a number of experiments to find an initial feasible range for each of the designated variables: wing length ( $4.50 \text{ cm} < X_1 < 10.25 \text{ cm}$ ) and body length ( $3.75 \text{ cm} < X_2 < 9.25 \text{ cm}$ ) which contains the optimal region. Finally, we have applied the ASRSM, CCD, A-, D-, and V-optimal designs to the coded factor space and compared their performances. Note



**Table 7 The fly time responses obtained for the paper helicopter experiment**

Exp. No.	Fly time ( sec.)					
	ASRSM	CCD-9	CCD-13	A-Opt.	D-Opt.	V-Opt.
1	1.47	1.85	1.85	2.41	2.41	2.41
2	2.34	2.19	2.19	1.50	1.5	2.06
3	1.53	1.75	1.75	1.54	1.47	2.75
4	2.5	2.22	2.22	2.60	2.22	1.5
5	2.46	1.59	1.59	2.15	2.32	1.47
6	2.13	2.60	2.6	1.91	1.54	2.22
7	2.41	2.41	2.41	1.60	2.6	2.32
8	2.57	2.13	2.13	2.10	2.63	1.34
9		2.5	2.5			
10			2.28			
11			2.25			
12			1.94			
13			2.46			

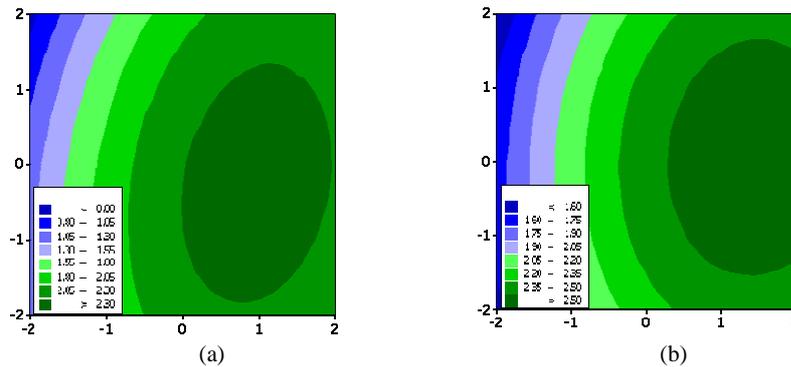
Table 8 presents the prediction results of the paper helicopter experiment. Since the true response model is unknown, we have estimated a quadratic response surface using all 53 experiments. Based on this estimated response, the optimal experiment is identified at  $O=(1.0714,-0.007)$  with mean response  $ZO=2.59487$  seconds. Both the  $R_{adj}^2$  and optimality gap results show that the ASRSM method outperforms other preset and optimal designs. The D- and V-optimal designs are the second and third best performing approaches, respectively. Whereas the differences in the optimality gap is small, the differences in the Euclidean distance between the predicted optimal experiment  $\hat{X}_O$  and optimal experiment  $O$ , e.g.  $\|O - \hat{X}_O\|$ , is more

substantial. This is attributable to the fact that quadratic convex functions are flat around the optimum and the response is relatively insensitive to deviations from the optimum.

**Table 8 The prediction results of ASRSM, CCD, and optimal designs for the paper helicopter case study.**

		ASRSM	CCD-9	CCD-13	A-Opt.	D-Opt.	V-Opt.
<b>Adjusted R<sup>2</sup></b>		86.5%	55.2%	50.8%	62.1%	66.3%	70.8%
<b>Predicted Optimum</b>	<b>X<sub>1</sub></b>	1.0427	0.8624	0.9571	1.1142	1.1635	1.2856
	<b>X<sub>2</sub></b>	0.0429	-0.1021	-0.2428	-0.3842	-0.4124	-0.5451
<b><math>\ O - \hat{X}_0\ </math></b>		0.0576	0.2296	0.2620	0.3796	0.4158	0.5792
<b>Pred. Resp. <math>\hat{Z}</math></b>		2.5931	2.5783	2.5842	2.5904	2.5917	2.5902
<b>Opt. Gap</b>		0.07%	0.64%	0.41%	0.17%	0.12%	0.18%

Based on the 13 experiments, CCD achieves  $R_{adj}^2 = 50.81\%$  with the estimated coded optimal solution (0.957094, -0.02132) and the contours of the estimated fly time response model is shown in Figure 17a. On the other hand, the ASRSM required only 8 experiments in two runs to attain  $R_{adj}^2 = 86.54$  with the estimated coded optimal (1.0427, 0.0429). The contour plot of the estimated fly time response model is shown in Figure 17b. These results clearly show that the ASRSM outperform the traditional RSM CCD method in terms of both the number of experiments and the accuracy of the results.



**Figure 17 The (a) CCD and (b) ASRSM estimated contours of the paper helicopter fly time response**

### **2.7.4 Traumatic Brain Injury (TBI): Design of Controlled Cortical Impact Model**

TBI continues to be a serious societal problem that affects more than 1.4 million Americans each year (Mao, Zhang, Yang, King, and 2006). The controlled cortical impact (CCI) rat model is one of the most frequently used animal models. This model is used to correlate real world injuries with predictions from a validated finite element (FE) model in order to establish injury threshold. In CCI model, the impact depth (potentially ranging 1-3 mm) and the impactor diameter (potentially ranging 2.5-7.5 mm) are believed to be two main factors in determining injury severity. However, the percent of increase/decrease in size of rat brain contributes to variances observed in post-impact tissues. Since the effect of this external parameter is largely unknown, it can be considered as noise. In CCI studies, one common problem is to find the specific levels of factors that result in specific percent of injury in animal brain. However, these experiments are not only very expensive, but are also very time consuming.

In this case study, we used the proposed approach along to find the parameter setting that result in 30% injury in the rat brain. We also conducted CCD experiments to compare the performance with the proposed approach. The technical details of the experiments can be found

in Mao, Zhang, Yang, King, and 2006. Table 9 shows the conducted experiments of CCD and the proposed ASRSM approach at different runs.

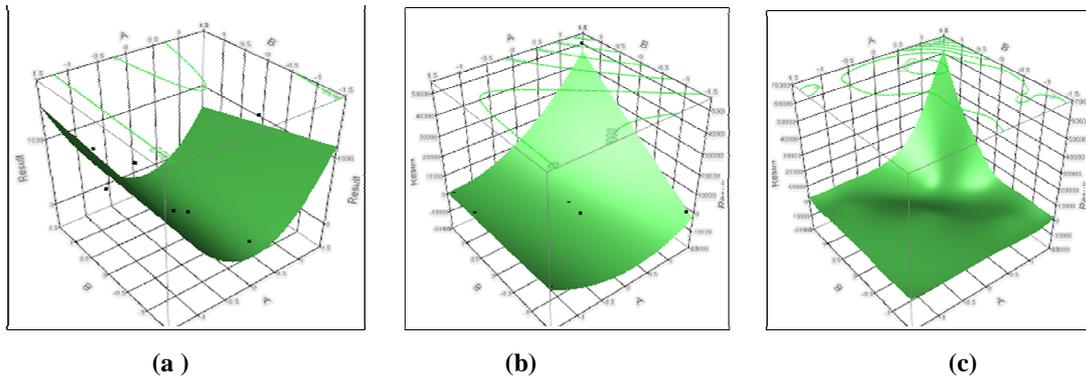
**Table 9 The runs and experiments of (a) CCD and (b) ASRSM method for the brain trauma case study**

(a)						(b)					
Controllable Factors				Random Factor	Response	Controllable Factors				Random Factor	Response
Coded		Original				Coded		Original			
impact depth	impactor diameter	impact depth (mm)	impactor diameter (mm)	brain size variation	Brain Injury	impact depth	impactor diameter	impact depth (mm)	impactor diameter (mm)	brain size variation	Brain Injury
1	-1	0.7	1.8	0%	900.00	-1.41	-1.41	1	2.5	-1%	877.92
1	-1	2.1	1.8	1%	306.37	1.41	1.41	3	7.5	0%	51595.49
-1	1	0.7	5.3	-2%	900.00	1.41	-1.41	3	2.5	0%	78.93
1	1	2.1	5.3	0%	11.54	-1.41	1.41	1	7.5	0%	761.34
-1.41	0	1.0	5.0	1%	894.44	0	0	2	5	1%	10.19
1.41	0	3.0	5.0	1%	1080.03	-1.41	0.71	1	6.25	0%	852.10
0	-1.41	2.0	2.5	1%	280.58	1.41	0.71	3	6.25	1%	19476.43
0	1.41	2.0	7.5	0%	206.08	0	-0.35	2	4.375	1%	51.51
0	0	2.0	5.0	0%	6.57						
0	0	2.0	5.0	-1%	3.40						
0	0	2.0	5.0	1%	10.19						
0	0	2.0	5.0	2%	16.29						
0	0	2.0	5.0	-2%	0.875						

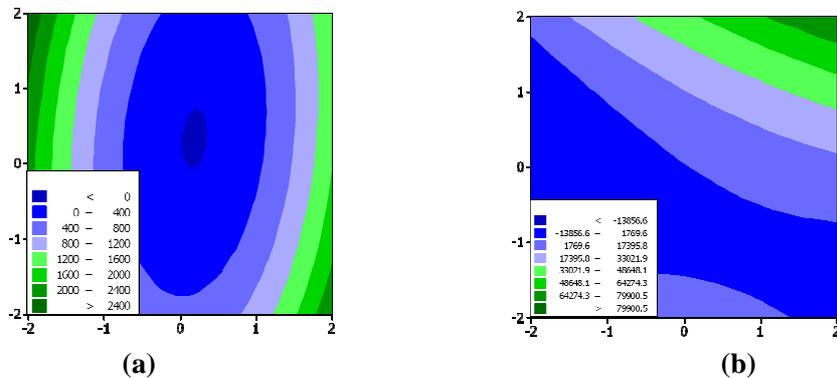
Using 13 experiments, the CCD fits a quadratic surface with  $R_{adj}^2 = 68.31\%$  and identifies  $EO = (0.1857, 0.3286)$ . Figure 18a shows the 3D plot of CCD estimated surface. In comparison, using 8 experiments in 2 runs, the proposed approach fits the quadratic model shown in Figure 18b with  $R_{adj}^2 = 82.59\%$  and  $EO = (0, 0.0505)$ . As shown in the Figure 18, although the estimated optimal experiments of both approaches are close, the estimated response models of the two methods are significantly different. The differences are even more apparent from the contours of the two response model estimates (Figure 19).

To compare the estimated functions, we aggregate the experimental data from both approaches (e.g. Table 9) and used the Radial Basis Function (RBF) to find the best fit. Figure 15c illustrates the model fit using RBF and the aggregated experimental data. From Figure 15 it

can be seen that, the estimated function of the proposed approach is much more similar to RBF with aggregated experimental data. Meanwhile, we also note that the EO of the proposed approach is very close to that of RBF based on the contour plots in Figure 18b and 18c. Given that the underlying response model is potentially highly nonlinear, the results are very encouraging for the effectiveness of the proposed approach in real world applications.



**Figure 18** 3D plots of the estimated response for (a) CCD, (b) ASRSM, (c) RBF in the brain trauma case study



**Figure 19** The (a) CCD and (b) ASRSM estimated contours of the response in the brain trauma case study

## 2.8 Discussion

In this chapter of Dissertation we developed and presented an adaptive methodology for response surface optimization (ASRSM). The proposed approach combines concepts from

nonlinear optimization, design of experiments, and response surface optimization. The ASRSM is a sequential adaptive experimentation approach and uses the information gained from the previous experiments to design the subsequent experiment by simultaneously reducing the region of interest and identifying factor combinations for new experiments. Its major advantage is the experimentation efficiency such that, for a given response target; it identifies the input factor combination (or containing region) in less number of experiments than the classical single-shot RSM designs. It differs from earlier studies in its optimality (under certain assumptions), inheritance of results from previous experiments, and its robustness due to experiment ranking based reduction of the region of interest. Through extensive simulated experiments and real-world case studies, we showed that it outperforms the popular CCD method in terms of both optimality as well as the experimentation efficiency. These results further demonstrate that the ASRSM is competitive with the A-, D-, and V-optimal designs. In particular, the performance of ASRSM is found to be more robust with respect to changes in the error variance and convexity of the response model, and more monotonous with additional experiments. In the following chapters, the proposed methodology will be extended to higher dimensional problems as well as higher order of response functions which can be either convex or non-convex.

## **CHAPTER 3 THE PROPOSED STRATEGIES FOR QUADRATIC AND CUBIC FUNCTIONS WITH N-VARIABLES (N-ASRSM)**

This chapter presents the detailed elements of the proposed adaptive sequential response surface methodology for n-dimensional problems (N-ASRSM). We first describe the terminology, most of which has been kept from previous Chapter, and state the assumptions. Next, we provide an overview of the methodology, and then describe the two core strategies embedded in N-ASRSM: (1) Parametric approach, and (2) Non-parametric approach. Afterwards, we describe how these two strategies are integrated within N-ASRSM. Finally, we compare the performance of the N-ASRSM with a number of popular methods in the literature on various numerical examples. We also extend the proposed methods to N-ASRSM2 which use optimal designs to consider factor spaces with different shapes.

### **3.1 Terminology and Assumptions**

The definitions and terminology used in the proposed N-ASRSM methodology is very similar to ASRSM with a few new items. For more convenience, we show the complete notations below. Figure 20 illustrates some of the notations for a three-dimensional factor space with 5 initial experiments in each run:

$FS_r$  : Factor space at run  $r$  and expressed as Cartesian product of factor ranges in run  $r$

$fs_i$  : Initial range of factor  $i$

$D$  : Design of most current run

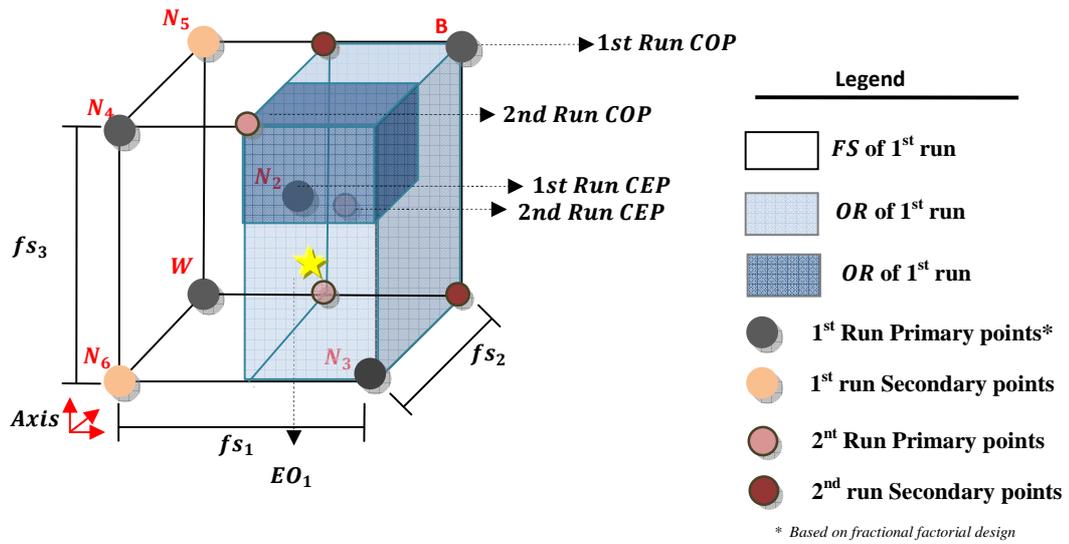
$d$  : minimum number of required points in  $D$

$COP$  : A corner point experiment run at the intersection of extrema of factor ranges

$CEP$  : A center point experiment run at the center of gravity of the factor space

$r$  : Index of runs, e.g.  $r = 1, 2, \dots, R$  where  $R$  is the total number of runs

- $e$  : Index of experiments in a given run, e.g.  $e=1,2,\dots,E$  where  $E$  is the total number of experiments  
 $B$  : The experiment with the best response level in a given run  
 $N_k$  : The experiment with the  $k^{th}$  best response level in a given run ( $2ek \leq e - 1$ )  
 $W$  : The experiment with the worst response level in a given run  
 $OR_r$  : Optimal region in run  $r$  containing the estimated optimal experiment,  $OR_r \subseteq FS_r$   
 $NOR$  : Non-optimal region  
 $O$  : Optimal experiment, e.g. best experiment in the initial factor space  
 $RO$  : Real optimum of the function  
 $EO_r$  : Estimated optimal experiment in run  $r$ , e.g., best incumbent estimation of the optimal experiment  
 $sb$  : index of sub-regions in a given factor , where the total number of sub-regions is  $2^n$   
 $c$  : Number of coefficients of the underlying model  
 $P_l$  : Probability of losing  $RO$  but cutting out



**Figure 20** An illustration of terminology on a three dimensional factor space

The proposed N-ASRSM methodology relies on the simplifying assumption of quadratic relation between the single response and input factors. However in the numerical example we show that the proposed method is acceptably robust to the violation of this assumption.

### 3.2 Proposed N-ASRSM Algorithm

The following algorithm illustrates the general scheme of the proposed methodology. A graphical representation of the algorithm is also shown in Figure 21.

---

#### ***n* – ASRSM Algorithm**

---

##### ***Initialize***

- Label the first run ( $r=1$ )
- Determine the design  $D$  based on the fractional design with smallest resolution

##### ***Repeat until convergence***

- Take experiment(s) one at a time based on  $D$

##### **1. Parametric Strategy (PS)**

- Fit a quadratic model based on experiments,
- Calculate  $R^2$ -adjusted of the model,
- Determine the Estimated Optimal point ( $EO_r$ ),
- Check for convergence ( $r>1$ ):
  - If  $|fR_{adj}^2| \leq \delta_{R_{adj}^2}$  or  $|\Delta(EO)| \leq \delta_{EO}$ , Stop.
- Else, Goto Check\_EO in step 3.

##### **2. Non-parametric Strategy (NPS)**

- Rank the experiments as  $B, N2, \dots, Nk, \dots, W$ ,
- Identify and eliminate non-optimal regions ( $NOR$ ) by solving a system of quadratic inequalities using a Max-Min Optimization approach,
- Check for hyper-rectangular Optimal Region ( $OR_r$ ):
  - If  $OR_r$  can be characterized as a hyper-rectangle,
    - Go to 3 Check\_EO.
  - Else,
    - Iteratively find furthest projections of experiments in an alternating ranking order, beginning with  $W$ , until a hyper-rectangular  $OR$  is obtained. Repeat NPS.

##### **3. Check if $EO_r$ is contained in $OR$ (Check\_EO)**

- If  $EO_r$  is contained in  $OR_r$ ,  $FS_{r+1} := OR_r$
- Else, Expand  $OR_r$  to contain  $EO_r$  and  $FS_{r+1} := OR_r$ .

##### **Return.**

---

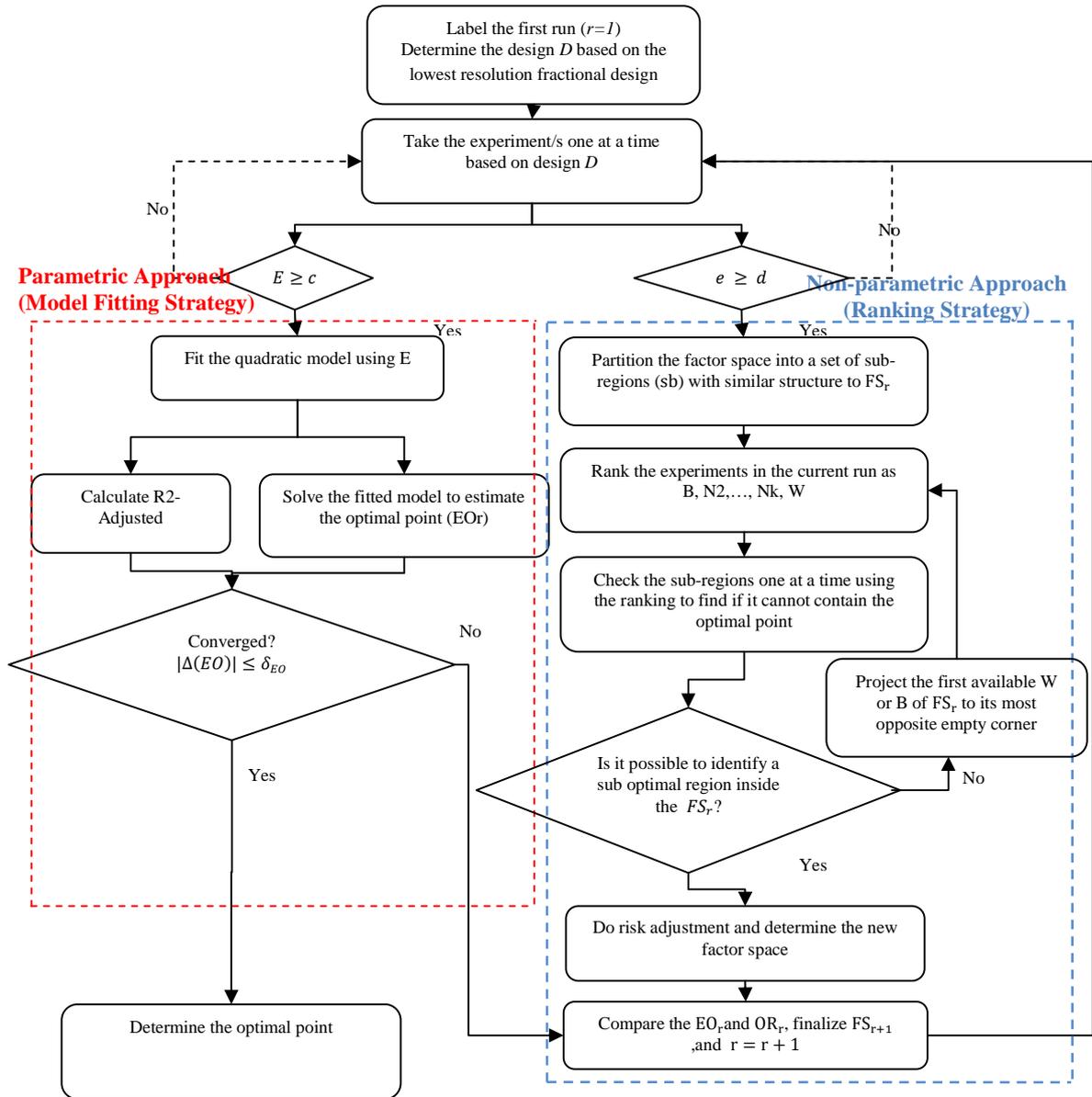


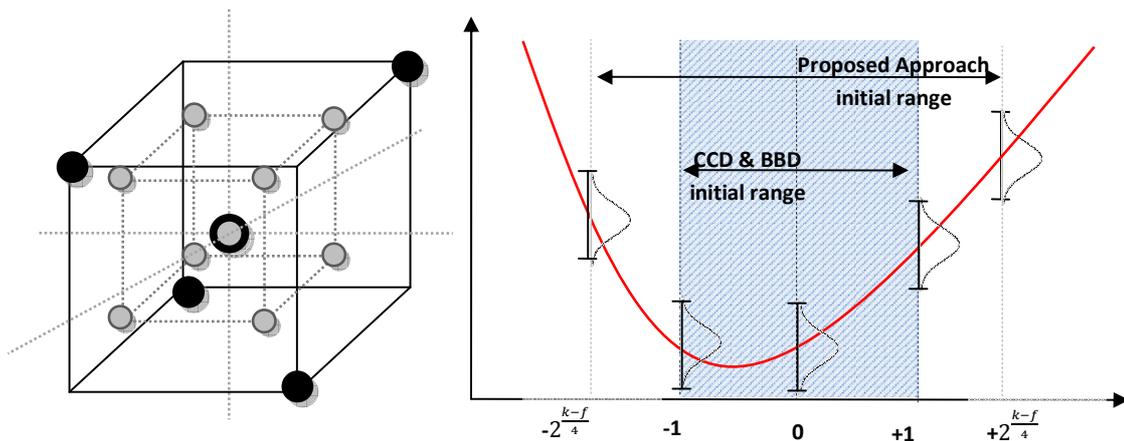
Figure 21 The general scheme of N-ASRSM

### 3.3 Design Structure of the First and Subsequent Runs

The design  $D$  structure of the factor space  $FS_r$  in the proposed approach is adapted from the minimum resolution fractional factorial design augmented with a center point both to minimize the number of experiments and take advantage of designs structure discussed in Chapter 2. This

design may be further augmented with few more experiments on the empty corners of  $FS_r$ , which will be discussed in later.

Similar to Chapter 2, the factor space of each run ( $FS_r$ ) in the proposed approach can be expressed as a mapping ( $\varphi_r$ ) of the factor space of the preceding run ( $FS_{r-1}$ ) maintaining similar design structure, e.g.  $FS_r = \varphi_r(\varphi_{r-1}(\dots \varphi_0(FS_1)))$ . Regarding the  $FS$  size,  $t$  is suggested to start with a broader initial region around the center point in comparison to classic approached, e.g.  $\pm 2^{\frac{k-f}{4}}$  unit distance from the center where  $2^{k-f}$  the number of points is in the chosen fractional factorial design (above relation is based on the calculation axial points in rotatable CCD with a single replicate at all designated points (Montgomery, 2005)). Considering the diagonal cross-section of these two designs in at one dimension as illustrated in Figure 22b and assuming that the noise is identically distributed on this cross-section. Then, it can be shown that the impact of the noise on prediction of the optimal experiment point is less with the proposed methodology's factor space. Figure 22a compares the initial factor space of the traditional CCD and the proposed (See also section 2.3 for more information).

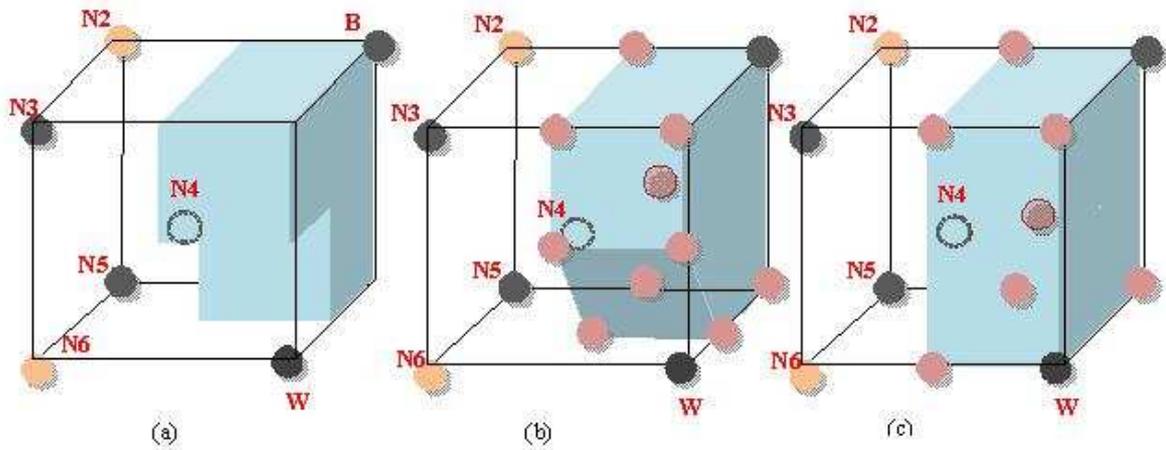


**Figure 22 (a) Initial factor space and design structure and (b) Diagonal cross-section of the traditional CCD and proposed N-ASRSM approach**

### 3.4 Non-parametric Approach: Ranking Strategy

At each run  $r$  of the proposed  $N$ -ASRSM approach, we first rank the experiments (e.g.,  $k$ th point fractional factorial and one center points) as  $N_1$  (we would call  $B$ ),  $N_2, \dots, N_{(k-1)}$  and  $N_{(k-1)}$  (we would call  $W$ ) according to their response levels. Based on the ranking, we identify the implied optimal region which contains the  $EO_r$ . This region is a polygon contained in  $FS_r$  and can be convex or non-convex in the space of factors. We then identify a hyper rectangle which contains the implied optimal region and denote it as the optimal region ( $OR_r$ ), which determines the factor space of the next run.

Similar to Chapter 2 model, this process of encapsulating the implied optimal region with a hyper-rectangle is a form of relaxation and is not efficient in terms of factor space reduction. However, there are valid reasons that motivate this relaxation. Again, the foremost reason is the reduced need for new experiments due to the inheritance of experiments from the previous run. Secondly, the hyper-rectangular  $FS$  preserves the orthogonality of factorial experimental design. Further, this hyper-rectangular form facilitates the recursive characterization of the same rectangular structure throughout the process. In addition, we can use the same experimental design structure, e.g. full factorial with a center point. Specifically, with a hyper-rectangular envelope, the mapping across runs will be identical, e.g.  $\varphi(\cdot) \equiv \varphi_r(\cdot)$  for  $\forall r$ . This is because we maintain the same experiment design structure, and there are a finite number of optimal regions as a result of ranking outcomes. Lastly, the relaxation reduces the risk of selecting an optimal region that excludes the optimal experiment. Figure 23 illustrates the trade-off among implied optimal region, convex hull envelope of the implied optimal region, and rectangular envelope of the implied optimal region.



**Figure 23 (a) Implied optimal region (b) Convex hull envelope of the implied optimal region (c) Rectangular envelope of the implied optimal region based on a three dimensional factor space**

In what follows, we first present the methodology used to reduce the factor space. Next, we describe how to choose additional experiments for characterizing a hyper-rectangle  $OR$ .

### 3.4.1 Reducing Factor Space (FS)

The reduction of the factor space to a sub-region containing  $O$  is achieved through the ranking of experiments of the current run. This reduction is performed by elimination of those sub-regions that do not contain the optimal point, e.g., non-optimal regions ( $NOR$ ). The determination of such sub-regions is exact as per the assumption stated in Chapter 2. Intuitively, the sub-regions in the vicinity of high and low ranking experiments are more simply characterized as a  $NOR$  or not. In particular, the vicinity of  $B$  has a higher probability of containing  $O$  while the other regions, e.g., the vicinity of  $W$ , have considerably less chance of containing  $O$ . Such confidence decreases in the vicinity of less extreme points. This intuition can be formalized in an algorithm as follows:

---

***NORs Elimination Procedure***


---

**Step 1.** Divide  $FS$  into  $2^n$  sub-regions of the same size and structure by bisecting the  $FS$  using  $n$  hyperplanes orthogonal to the  $n$  factor dimensions (See Figure 24).

**Step 2.** For each of the  $2^n$  sub-regions, repeat:

**2.1.** Identify a hypothetical optimal point  $\hat{O}$  in the current sub-region.

**2.2.** For each experiment  $1 \leq e \leq E$ , express the response model in a canonical form as  $Z_e = \sum_{i,j=1}^n A_{i,j} (X_i^e - X_i^{\hat{O}}) (X_j^e - X_j^{\hat{O}}) + R^e$ , where  $A_{i,j} \in R$  and  $R^e$  is a constant term.

**2.3.** Sort the parametric canonical forms of the experiments in ascending order ( $Z_{e_B} < Z_{e_{N_2}} < \dots < Z_{e_W}$ ). (Because the canonical form should comply with empirical ranks of the experiments ( $B < N_2 < \dots < N_k < W$ )).

**2.4.** Rewrite the sorted canonical forms of the experiments in the form of a system of inequalities with  $\frac{E(E-1)}{2}$  pairwise comparisons of experiments as follows:

$$\begin{cases} Z_{e_B} - Z_{e_{N_2}} = \sum_{i,j=1}^n A_{i,j} (X_i^{e_B} - X_i^{\hat{O}}) (X_j^{e_B} - X_j^{\hat{O}}) - \sum_{i,j=1}^n A_{i,j} (X_i^{e_{N_2}} - X_i^{\hat{O}}) (X_j^{e_{N_2}} - X_j^{\hat{O}}) < 0 \\ Z_{e_{N_2}} - Z_{e_{N_3}} = \sum_{i,j=1}^n A_{i,j} (X_i^{e_{N_2}} - X_i^{\hat{O}}) (X_j^{e_{N_2}} - X_j^{\hat{O}}) - \sum_{i,j=1}^n A_{i,j} (X_i^{e_{N_3}} - X_i^{\hat{O}}) (X_j^{e_{N_3}} - X_j^{\hat{O}}) < 0 \\ \vdots \\ Z_{e_{N_k}} - Z_{e_W} = \sum_{i,j=1}^n A_{i,j} (X_i^{e_{N_k}} - X_i^{\hat{O}}) (X_j^{e_{N_k}} - X_j^{\hat{O}}) - \sum_{i,j=1}^n A_{i,j} (X_i^{e_W} - X_i^{\hat{O}}) (X_j^{e_W} - X_j^{\hat{O}}) < 0 \end{cases} \quad (8)$$

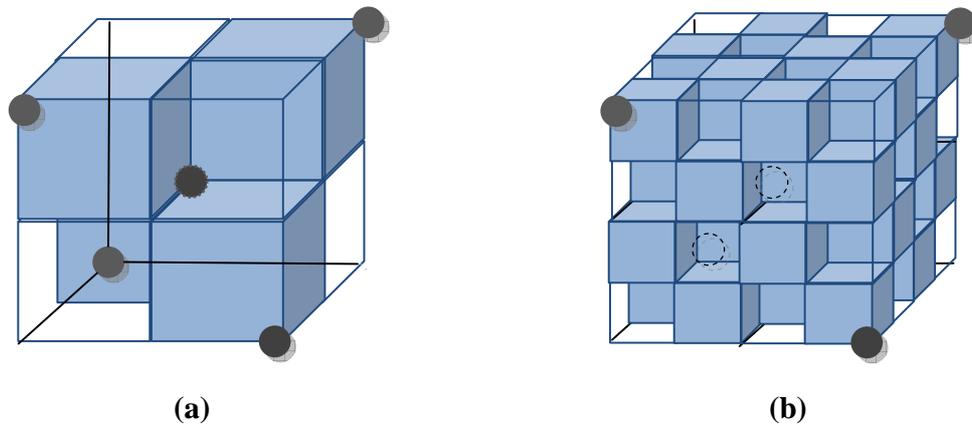
(In above system  $A_{i,j}$  and  $X_j^{\hat{O}}$  are the unknowns, where  $X_j^{\hat{O}}$  is bounded by the borders of the current sub-regions)

**2.5.** Check the feasibility of above system by looking for a negative solution of the following Max-Min optimization model:

$$\begin{aligned} \text{Min Max } Z &= (Z_{e_B} - Z_{e_{N_2}}, Z_{e_B} - Z_{e_{N_2}}, \dots, Z_{e_{N_k}} - Z_{e_W}) \\ \text{Subjected to:} \\ A_{i,j} &\in R, X_j^{\hat{O}} \in \text{Current Subregion} \end{aligned} \quad (9)$$

(Positive solution of above optimization model is equivalent to non-existence of a feasible solution for above system of quadratic inequalities and vice versa).

---



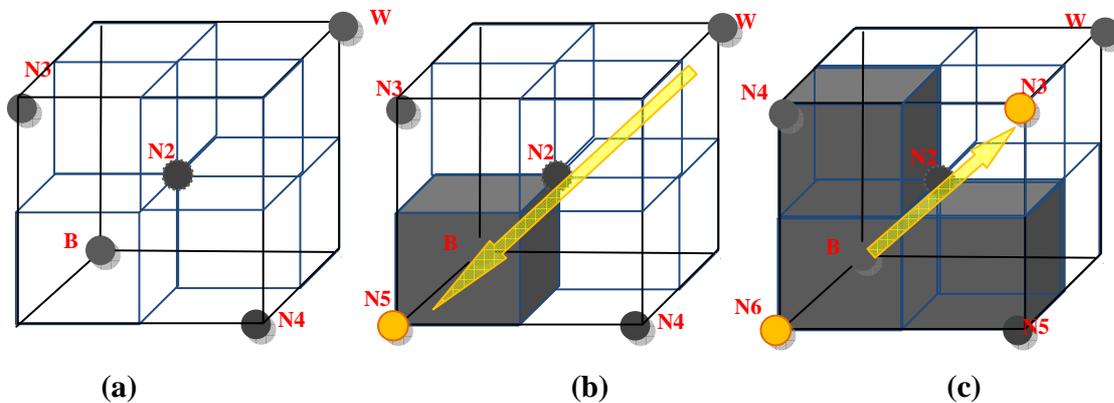
**Figure 24** 8 sub-regions of a 3 dimensional *FS* using 1 orthogonal hyper plane (b) 16 sub-regions of a 3 dimensional *FS* using 2 orthogonal hyper planes

Feasible solution of above system of quadratic inequalities means that the real optima (*RO*) can occur in the sub-region stated by step 2.1 otherwise that sub-region is not feasible and can be eliminated from the *OR*. It can be shown that above procedure eliminate only those sub-regions not containing the optimal point by contradiction as per the assumptions stated in Chapter 2. In particular, we first assume that there exists a sub-region containing the optimal point, which leads to an inconsistent ranking of at least one experiment pair. Next, we show that the  $A_{i,j=1,\dots,n}$  determined for the experiment pair contradicts the convexity assumption of the quadratic response forms.

The *NOR* elimination steps are repeated for all sub-regions until those sub-regions not eliminated or not checked form a hyper-rectangular region inside the *FS*. When such a hyper-rectangular region is obtainable, we then designate it as the *FS* of the next run. If a hyper-rectangular region is not available upon the checking of all sub-regions for *NOR* elimination, then additional corner experiments are necessary. The next section discusses how those additional experiments are determined.

### 3.4.2 Selecting Additional Corner Points

When the *NOR* elimination procedure terminates without a candidate hyper-rectangular *FS* or very small eliminated sub-region, then additional points are needed. These additional points enable eliminating more of the sub-regions in a few ways. First, they increase the number of pair-wise ranking comparisons of experiments such that the likelihood of a previously non-eliminated sub-region becoming a *NOR* is increased. Second, with these additional points, the new ranking of the experiments leads to a better coverage of *FS*. Finally additional points generally result in more reliable ranking of the experiments that potentially allow elimination of more sub-regions. However, since one of the goals of  $n - ASRSM$  is to reduce the total number of experiments, the number of additional points should be kept as small as possible. This can be achieved by selecting the additional points that provide maximum potential for eliminating *NORs*. Figure 25 illustrates the process of selecting two additional points.



**Figure 25** Cut out regions using the ranking strategy based on two additional points

We select additional points one at a time until the next *FS* as a hyper-rectangle can be inferred. The selection strategy employed is based on the simplex optimization method in Walters, Parker, Morgan, and Deming (1962) and aims at maximizing the potential of eliminating more *NORs*. This strategy is executed by using the current ranking information of

the experiments and subsequently identifying those directions with most improvement and worsening of response based on the current experiments. Clearly, the highest ( $B$ ) or lowest ( $W$ ) ranking experiments are ideal candidates for identifying such directions for two reasons. First, the most opposite corner projections of  $B$  and  $W$  provide the most information on the orientation of the diagonals of the underlying function. The second reason is, as in the simplex optimization method, the projection in the opposite of least favorable ( $W$ ) and most favorable ( $B$ ), is likely to produce a new ranking with a more precise range of response orientation. Once the opposite projections of  $B$  and  $W$  are taken as additional points, we continue taking additional points in the opposite reflections of next alternate ranking experiment pairs, e.g.,  $N_{E-2}$  and  $N_2$ , and so on. Figure 25 illustrates the two additional points taken as the opposite projections of first  $W$  and then  $B$ .

The most opposite projections of corner point experiments are determined according to the cosine similarity measure (Tan, Steinbach and Kumar, 2006). To illustrate, the most opposite corner projection of the worst/best experiment is found by:

$$C_k = \arg \min_{C_k} \cos(\alpha), \quad (10)$$

where  $\alpha$  is the angle between  $C_w$  and  $C_k$ , vectors connecting the experiments to the center point. Above procedure works while the candidate experiment is a corner points. If the candidate experiment, e.g.  $B$  or  $W$ , is a center point then opposite projection of the next candidate experiment should be considered.

### 3.4.3 Risk Adjustment

Reducing factor space is exact when there is no noise inside the model, however when data are erroneous, there is a probability that two or more of the rankings be incorrect so the cut off region may be inaccurate or even invalid as shown in Figure 26 (a) and (b).

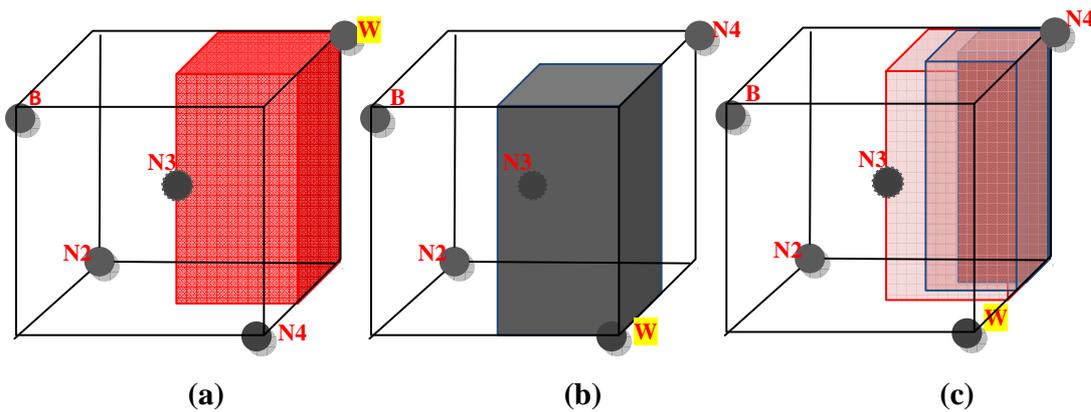


Figure 26 The effect of incorrect ranking on the *NOR*

Knowing the probability of incorrectly ranking the experiments can help to change the size of the *NOR* to adjust the risk of not containing the *RO*. The challenge is that the variance of the noise is unknown and the number of experiments is usually not enough (especially in the early runs) to estimate it. In the rest of this section, first, we present a novel approach for finding a (pessimistic) estimate of variance when there is not enough data to estimate the model parameters. Next, we will show how the estimated variance can be used for adjusting the risk of missing the *RO* when shrinking the *OR*.

#### 3.4.3.1 Estimating the Variance with Not Enough Data

When the number of experiments is not enough to estimate all parameters of the model, we can design a system of two equations for obtaining the pessimistic estimate of variance. The first

equation is from decomposing total sum of square (SST) into sum of square regression (SSR) and sum of square error (SSE) which is shown below:

$$SST = SSE + SSR \quad (11)$$

The second equation is derived based on minimum significance level of the hypothesis testing on meaningfulness of the regression. The testing statistics is  $F = \frac{SSR/k-1}{SSE/n-k}$  where  $k$  the number of parameters in the canonical form of response model, and  $n$  is the total number of experiments. The reason for using canonical model for calculating  $k$  is that since we assumed the optima is occurring in one of the sub-regions, canonical modeling will reduce the number of parameters to be estimated ( $k$ ). The critical value of the hypothesis test is  $F_{\alpha, k-1, n-k}$ , so at the significance level  $\alpha$  considering the minimum value of the statistics which makes the regression meaningful the following equation can be written as:

$$\frac{SSR}{k-1} - F_{\alpha, k-1, n-k} \cdot \frac{SSE}{n-k} = 0 \quad (12)$$

In equations 11 and 12,  $SST$ ,  $k-1$ ,  $n-k$  and  $F_{\alpha, k-1, n-k}$  are known and  $SSR$  and  $SSE$  are unknown, so combining 11 and 12 will result in a system of two equations and two unknowns. One of the solutions of above system would be  $SSE$  which can be used for estimating the variance ( $\widehat{Var} = MSE \leq \frac{SSE}{n-k}$ ).

### 3.4.3.2 Calculating the Probability of Incorrect Ranking and Risk Adjustment

Having a set of ranked experiments  $e_i$  ( $i = B, N_2, \dots, N_{k-1}, W$ ) and the estimate of variance (MSE) available, the probability of incorrect ranking for each pair of experiments  $i, j$  ( $i < j$ ) can

be easily approximated by  $p(e_j < e_i) = \varphi\left(\frac{Z_{e_j} - Z_{e_i}}{MSE}\right)$ . Using this information we want to estimate the probability of losing the  $RO$  by cutting off the sub-region chosen by the proposed approach. While there are many ways to approximate the probability of cutting off the  $RO$  using estimates of incorrect ranking we use the following strategy:

---

***Incorrect Ranking Estimation Procedure***

---

**Step 1.** Set  $r = 1$ .

**Step 2.** For each experiments from  $W$  to  $B$  flip the rank (and observed value) of each experiment with its next  $r^{\text{th}}$  higher and lower rank experiment one at time.

**Step 3.** If  $NOR$  changes, find the probability of rank rotation using above formula and go to step 4. If not go to step 5

**Step 4.** From all possible rotations finds the one with maximum probability ( $P_l$ ) as estimate of cutting off  $RO$ .

**Step 5.** Set  $r=r+1$  and go to step 2

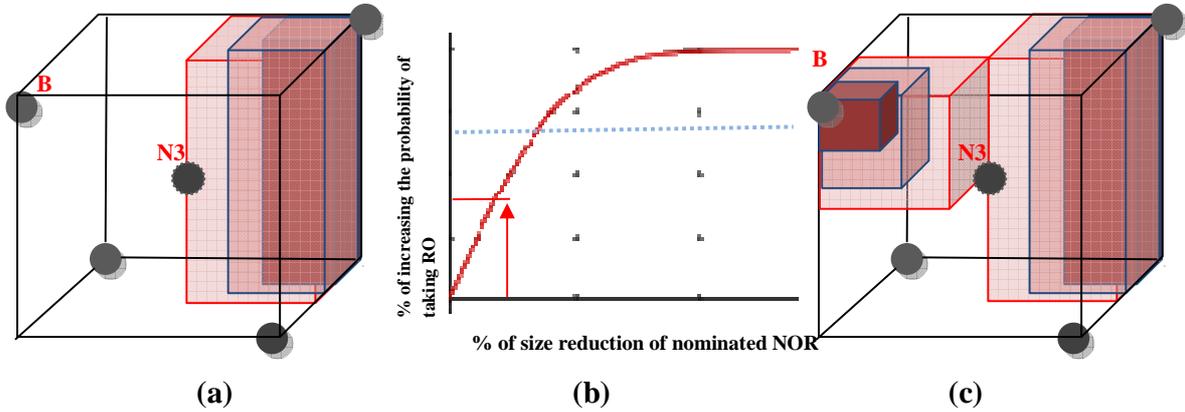
---

Above strategy is based on one factor at a time analysis which has the complexity of  $O(n)$ , though showing good performance in many numerical examples comparing to exhaustive search which has a  $O(m^n)$  complexity. Next, we explain how to incorporate above probability for risk adjustment.

The risk adjustment for cutting off  $NOR$  is based on a simple intuitive idea. If we don't cut any region the probability of cutting off  $RO$  is zero. If we cut of the nominated  $NOR$  the probability of losing  $RO$  is  $P_l$ . So shrinking the size of nominated  $NOR$  will decrease  $P_l$  which is also shown in Figure 27 (c).

To approximate the rate of reduction in  $P_l$  by shrinking the size of nominated  $NOR$ , we incorporate the Gauss error function as shown in Figure 27 and defined in Equation (13):

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (13)$$



**Figure 27 Application of Gauss error function on risk adjustment**

Therefore, by setting an acceptable level of confidence on taking  $RO$ , e.g. 95%, we can easily find the reasonable amount of shrinkage in the cut off region. The final point in risk adjustment of cutting off  $NOR$  is that above steps are applied to every nominated  $NOR$  in section 3.4.1. As a result multiple sub-regions with different  $P_l$  might be nominated for cutting off. In this case we consider the union of candidate sub-regions and maximum of  $P_l$  of different regions for risk adjustment

### 3.5 Parametric Approach: Model Fitting Strategy

We use a parametric approach based on model fitting concurrent to the nonparametric ranking approach described in Section 3.4. This strategy not only allows us to increase the precision of  $EO_r$  but also supports backtracking through the expansion of  $OR_r$  to contain estimated optimal  $EO_r$ . Beginning with the completion of all first run experiments, this parametric approach is used after each experiment. In this approach we fit a quadratic model:

$Z = \sum_{i,j=1}^n Q_{i,j}x_i x_j + \sum_{i=1}^n P_i x_i + R + \varepsilon$  , with  $\varepsilon \sim N(0, \sigma^2)$ , to the experimental data to analyze the underlying function and efficacy of conducted experiments. In fitting the quadratic model, two objectives are being sought in particular: (1) finding the estimated optimal experiment  $EO_r$ ; and (2) calculating the adjusted coefficient of determination ( $R_{adj}^2$ ).  $EO_r$ , the minimum of the fitted model, not only shows the predicted optimal solution, but is also used for the expansion of  $OR_r$ . Furthermore, the change in the  $EO_r$  in consecutive runs is also used as a stopping criterion.  $R_{adj}^2$ , shows how well the information gained from the experiments explain the behavior of the underlying system (Seber and Alan, 2003).

### 3.6 Expansion of OR to contain EO

As described in the N-ASRSM algorithm in the previous section, we check the consistency of the  $EO_r$  obtained from the parametric approach, and the estimated optimal region  $OR_r$  obtained from the non-parametric strategy. When the  $EO_r$  is found to be outside  $OR_r$ , we then expand the  $OR_r$  to contain  $EO_r$  while preserving its hyper-rectangular structure. This expansion is illustrated in Figure 28.

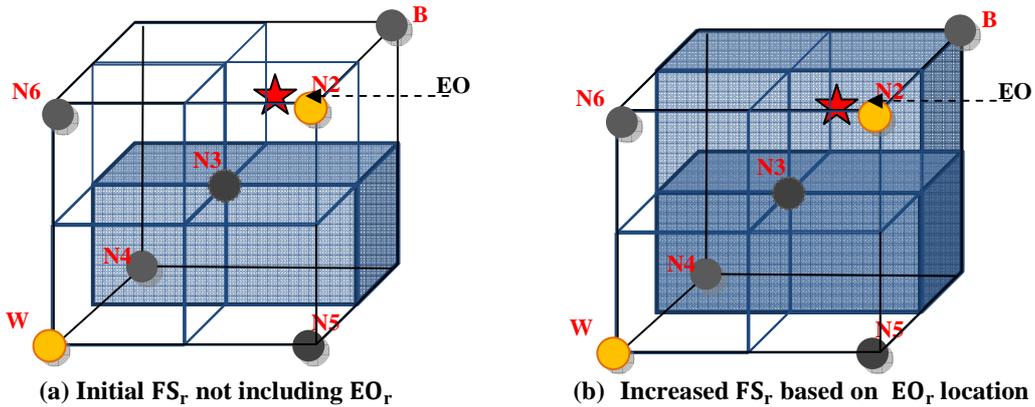


Figure 28 Expansion of  $OR$  in N-ASRSM when the  $EO$  falls outside

### 3.7 Numerical Examples

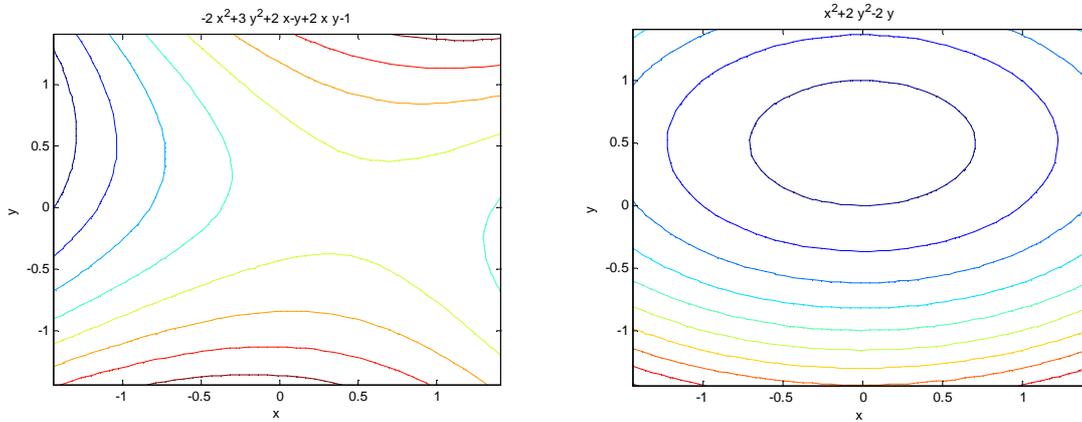
In this section, we describe two sets of simulated experiments performed to evaluate the performance of the proposed N-ASRSM approach. In the first set of simulations N-ASRSM is compared to well-known classical methods including CCD, BBD and A, D, and V optimal designs on different quadratic response models with varying variance of errors. The second set of simulations study the performance of the proposed approach along with classical models, optimal designs and two global optimization methods (Standler et al. (2002), Wang et al. (2003)) on a number of nonlinear response models with various errors.

#### 3.7.1 Quadratic Response Models

We now describe the simulated experiments performed to compare the performance of the proposed ASRSM approach with those of CCD, BBD, A- D- and V-optimal designs on quadratic response models. We have considered 6 problems with two and 5 problems three variables. These problems cover different type functions and various range of standard deviation (See Table 10). As noted earlier, all response models have a quadratic relation with a normal error term  $\varepsilon \sim N(0, \sigma^2)$ . Figure 29 also shows the contour plot of two of the responses.

**Table 10 The response models used in the simulated experiments of N-ASRSM**

No. of variables	Exp. No.	Response Relation	Error ( $\epsilon$ )	Response Type
Two Variable Response	1.1	$W=-2x^2+3y^2+2x-y+2xy-1+\epsilon$	$N(0,0.1)$	Non-convex
	1.2	$W=x^2+2y^2-2y+\epsilon$	$N(0,1)$	Convex
	1.3	$W=-2x^2+3y^2+2x-y+2xy-1+\epsilon$	$N(0,2)$	Non-convex
	1.4	$W=-3x^2+2y^2+x-2y+2xy-1+\epsilon$	$N(0,2)$	Non-convex
	1.5	$W=2x^2+y^2+x+2xy+\epsilon$	$N(0,2)$	Convex
	1.6	$W=x^2+2y^2-2y+\epsilon$	$N(0,2)$	Convex
Three Variable Response	2.1	$W=2x^2-1y^2-2z^2+x-2y+3z-xy+3xz+2yz+2+\epsilon$	$N(0,2)$	Non-convex
	2.2	$W=2x^2+3y^2+5z^2+x+2y+1z-5xy+1xz+1yz+1+\epsilon$	$N(0,3)$	Convex
	2.3	$W=1x^2+1y^2+5z^2+x+2y-5z-5xy+1xz+1yz+1+\epsilon$	$N(0,3)$	Non-convex
	2.4	$W=-1.5x^2-3.5y^2+3z^2+0.5x-3.5y-1.5z-3xy+1.3xz+1.4yz+2+\epsilon$	$N(0,2.5)$	Non-convex
	2.5	$W=2x^2+1.7y^2+1.6z^2-4.4x-5.75y-2.23z-1.2xy+1.3xz-1.1yz+6+\epsilon$	$N(0,2)$	Convex

**Figure 29 the contour plot of responses 1.1, 1.3**

Whereas N-ASRSM is an adaptive sequential method, the CCD, A-, D-, and V-optimal designs are essentially preset designs. In order to evaluate the effect of this difference, we initially fixed the number of design points at 7 for cases with two variables and 11 for cases with

three variables and then incrementally added one design point at a time for two more times. For optimal designs, the initial set of design points is optimally generated by optimizing the optimality criteria over the starting factor space with a fine grid system spaced with 0.01 intervals. Next, each of the additional points is generated by optimizing the optimality criterion given the existing design points and the response model. For the CCD and BBD, we initially used either 7 or 11 of the complete design by excluding some of the points and then re-including them one at a time.

For the analysis, we have studied the performances in terms of average  $R_{adj}^2$ , and average optimality gap (i.e., deviation from the optimal response). All simulated experiments are repeated three times, and average results are reported. The starting factor space is considered with the range of  $[-3,3]$  in all dimensions for both two and three variable examples. Table 11 presents the average  $R_{adj}^2$  performances for the consecutive trials.

**Table 11 The  $R_{adj}^2$  for trials 7, 8, 9 of responses with two variables and trials 11, 12 and 13 of the ones with three variables**

Exp. No.	No Obs.	Adjusted $R^2$					
		CCD	BBD	n-ASRSM	D-Opt.	V-Opt.	A-Opt.
1.1	7	99.96%	N/A	<b>99.99%</b>	99.96%	99.92%	99.94%
	8	99.96%	N/A	<b>99.99%</b>	99.98%	99.95%	99.95%
	9	99.95%	N/A	<b>99.98%</b>	99.97%	99.94%	99.92%
1.2	7	92.69%	N/A	93.67%	<b>95.01%</b>	94.85%	90.69%
	8	92.48%	N/A	94.97%	89.20%	<b>95.86%</b>	86.86%
	9	92.00%	N/A	<b>96.62%</b>	90.42%	95.42%	86.58%
1.3	7	69.77%	N/A	79.57%	92.96%	72.11%	<b>97.71%</b>
	8	70.60%	N/A	86.53%	<b>91.06%</b>	74.92%	79.86%
	9	88.48%	N/A	<b>92.91%</b>	89.02%	80.89%	82.00%
1.4	7	92.87%	N/A	91.68%	91.09%	85.12%	<b>94.21%</b>
	8	88.49%	N/A	<b>92.39%</b>	91.69%	88.76%	89.27%
	9	82.17%	N/A	<b>94.17%</b>	92.58%	85.47%	89.49%
1.5	7	24.61%	N/A	<b>90.15%</b>	87.21%	86.05%	87.19%
	8	35.75%	N/A	89.16%	76.74%	71.97%	<b>89.97%</b>
	9	37.03%	N/A	<b>94.26%</b>	67.96%	84.95%	86.66%
1.6	7	35.45%	N/A	<b>86.18%</b>	85.31%	85.08%	50.38%
	8	43.17%	N/A	<b>86.44%</b>	78.71%	86.04%	57.23%
	9	49.40%	N/A	<b>88.92%</b>	67.65%	82.05%	64.49%
Avg.	7	69.22%	N/A	90.21%	<b>91.92%</b>	87.19%	86.69%
	8	71.74%	N/A	<b>91.58%</b>	87.90%	86.25%	83.86%
	9	74.84%	N/A	<b>94.48%</b>	84.60%	88.12%	84.86%
2.1	11	87.30%	76.11%	96.78%	98.18%	92.66%	<b>98.42%</b>
	12	88.10%	81.32%	97.31%	<b>98.61%</b>	95.60%	96.60%
	13	91.09%	79.87%	96.66%	97.66%	95.67%	<b>97.79%</b>
2.2	11	80.64%	79.93%	91.44%	94.33%	<b>97.16%</b>	87.40%
	12	86.42%	81.86%	<b>93.40%</b>	93.39%	81.34%	90.36%
	13	86.93%	81.16%	<b>94.48%</b>	94.03%	77.13%	88.66%
2.3	11	89.36%	63.23%	<b>98.38%</b>	89.15%	84.79%	93.59%
	12	85.69%	63.75%	<b>96.24%</b>	90.78%	88.26%	91.80%
	13	85.06%	60.24%	<b>93.84%</b>	92.79%	90.96%	91.38%
2.4	11	89.02%	53.12%	<b>99.38%</b>	89.25%	84.88%	90.50%
	12	87.06%	56.04%	<b>94.99%</b>	92.93%	86.88%	91.76%
	13	86.50%	49.48%	<b>94.27%</b>	91.74%	88.27%	90.12%
2.5	11	92.77%	89.21%	97.21%	96.54%	<b>98.08%</b>	96.58%
	12	93.44%	91.03%	97.39%	97.51%	96.99%	<b>97.95%</b>
	13	92.59%	91.61%	<b>97.90%</b>	97.53%	97.32%	97.22%
Ave.	11	87.82%	72.32%	<b>96.64%</b>	93.49%	91.51%	93.30%
	12	88.14%	74.80%	<b>95.87%</b>	94.64%	89.81%	93.69%
	13	88.43%	72.47%	<b>95.43%</b>	94.75%	89.87%	93.03%

Table 12 also presents the average optimality gap of the consecutive trials of comparing methods. The optimality gap is measured as the deviation of the response at the final  $EO$  from the response at true optimal experiment  $RO$ . The experiments show that the optimality gap of the

proposed N-ASRSM is almost the most competitive among all methods. The V- and A-optimal designs are slightly better performing than N-ASRSM in the fifth response model.

**Table 12 The optimality gap for trials 7, 8 and 9 of responses with two variables and trials 11, 12, and 13 of the ones with three variables**

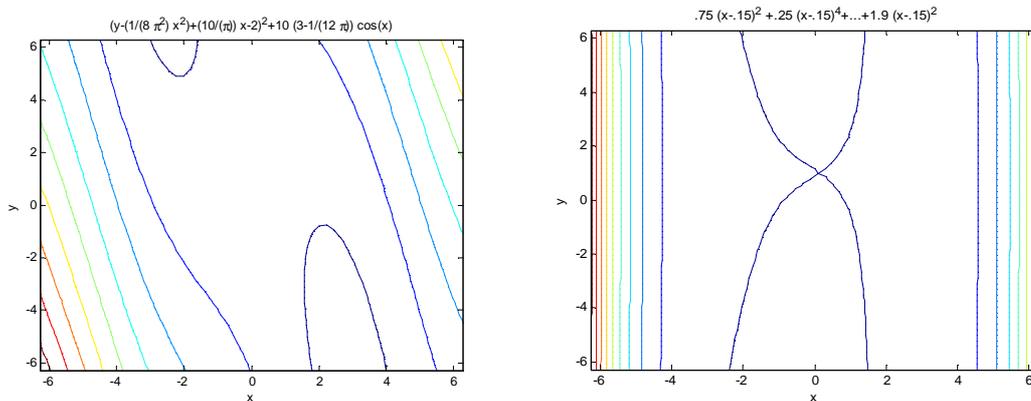
Exp. No.	No Obs.	Optimality gap					
		CCD	BBD	n-ASRSM	D-Opt.	V-Opt.	A-Opt.
1.1	7	3.84	N/A	<b>0.00</b>	<b>0.00</b>	31.76	<b>0.00</b>
	8	3.84	N/A	0.07	<b>0.00</b>	18.35	4.70
	9	0.23	N/A	0.04	0.26	<b>0.00</b>	<b>0.00</b>
1.2	7	70.52	N/A	0.01	<b>0.00</b>	0.61	40.19
	8	70.52	N/A	0.72	<b>0.53</b>	0.76	29.94
	9	70.52	N/A	<b>0.00</b>	7.69	0.83	8.24
1.3	7	909.82	N/A	<b>0.31</b>	<b>0.31</b>	<b>0.31</b>	1260.68
	8	909.82	N/A	<b>0.31</b>	<b>0.31</b>	<b>0.31</b>	492.09
	9	909.82	N/A	<b>0.31</b>	<b>0.31</b>	<b>0.31</b>	5.00
1.4	7	810.23	N/A	<b>30.79</b>	<b>30.79</b>	<b>30.79</b>	<b>30.79</b>
	8	2578.21	N/A	<b>30.79</b>	<b>30.79</b>	<b>30.79</b>	<b>30.79</b>
	9	363.90	N/A	<b>30.79</b>	<b>30.79</b>	<b>30.79</b>	<b>30.79</b>
1.5	7	1.59	N/A	0.01	0.01	0.07	<b>0.00</b>
	8	0.01	N/A	0.03	0.02	<b>0.00</b>	0.29
	9	0.38	N/A	0.02	0.01	<b>0.00</b>	0.21
1.6	7	874.36	N/A	30.79	<b>16.50</b>	<b>16.50</b>	790.25
	8	756.48	N/A	<b>16.50</b>	<b>16.50</b>	<b>16.50</b>	3583.36
	9	900.70	N/A	<b>16.50</b>	696.57	<b>16.50</b>	779.73
Avg.	7	445.06	N/A	10.32	7.94	13.34	353.65
	8	719.81	N/A	8.07	8.03	11.12	690.20
	9	374.26	N/A	<b>7.94</b>	122.61	8.07	137.33
2.1	11	19.24	18.99	<b>9.86</b>	11.95	13.84	13.50
	12	19.46	19.02	<b>10.10</b>	10.88	13.84	12.66
	13	13.09	19.42	<b>9.95</b>	10.56	14.01	11.81
2.2	11	25.00	25.46	<b>21.89</b>	23.99	26.16	32.82
	12	25.05	23.51	<b>22.19</b>	24.06	26.51	33.21
	13	27.84	22.44	<b>22.36</b>	23.17	26.19	32.92
2.3	11	3.64	12.48	<b>0.25</b>	1.01	12.65	0.54
	12	3.64	12.48	<b>0.25</b>	1.43	12.72	1.25
	13	8.59	12.22	<b>0.35</b>	1.40	12.72	2.20
2.4	11	19.30	20.28	<b>0.90</b>	6.39	1.40	1.43
	12	19.30	20.64	<b>1.03</b>	6.37	1.37	1.41
	13	19.53	19.75	<b>1.08</b>	6.40	1.39	1.43
2.5	11	4.13	5.72	1.51	1.48	<b>0.87</b>	1.23
	12	3.71	5.72	1.44	1.46	<b>0.87</b>	1.26
	13	4.60	5.72	1.44	1.58	<b>0.90</b>	1.17
Ave.	11	14.26	16.59	<b>6.88</b>	8.96	10.98	9.90
	12	14.23	16.27	<b>7.00</b>	8.84	11.06	9.96
	13	14.73	15.91	<b>7.04</b>	8.62	11.04	9.91

### 3.7.2 Non-linear Response Models

Here we compare the performance of the proposed N-ASRSM approach with two global optimization methods: Standler et al. (2002), Wang et al. (2003), as well as classical methods CCD, and BBD, and A- D- and V-optimal designs on five nonlinear response models with two and three variables, with different variance and function type. These response models are presented in Table 13. Figure 30 illustrates the contour plot of two of the responses.

**Table 13 The non-linear response models used for studying N-ASRSM**

No. of variables	Res. No.	Response Relation	Error
Response with Two variable	1.1	$W=(y-(1/(8\pi^2).x^2)+(10/\pi)(x-2)^2+10(3-1/(12\pi))\cos(x)$	$N(0,3.5)$
	1.2	$W=0.75(x-0.15)^2 +.25*(x-0.15)^4+ 1.3.*(x-0.15)^6+1.8(x-0.15)(y-1)^2-2.66(y-1)^2+1.9(y-0.15)^2$	$N(0,2)$
Response with three variable	2.1	$W=(x-0.55)^2 + (y+0.7)^2 + (z-0.33)^2 -\cos(18(x-0.55))-\cos(18(y+0.7))-\cos(18(z-0.33))$	$N(0,2)$
	2.2	$W=(x-1)^3 - 3(y-1)^3+(z+1)^3 -2(x-1)^2 - 2(y-1)^2 +(z+1)^2-(x-1)+5(y-1)+6(z+1)+2(x-1)(y-1)+(x-1)(z+1)-4(y-1)(z+1)+1$	$N(0,1)$
	2.3	$W=x^2 + \exp(y/10 + 10) + \sin(zy)$	$N(0,3)$



**Figure 30 The contour plot of Responses 1.1 and 2.2**

For the following analysis, we have examined the performances based average optimality gap and Euclidian distance of the estimated optima to the real optimal point. All simulated experiments are repeated two times, and average results are reported. To keep the consistency

with the preceding section, the result of trials 7, 8, 9 of the cases with two variables and trials 11, 12, and 13 of the cases with three variables have been reported. Table 14 shows the average optimality gap results of the consecutive trials of the comparing methods.

**Table 14 The average optimality gap of the proposed N-ASRSM and comparing methods**

f(x)	Run	CCD	BBD	N-ASRSM	Standler et al. (2002)	Wang et al. (2003)	D-optimal	A-optimal	V-Optimal
1.1	7	1504.46	N/A	<b>0.00</b>	750.68	702.68	0.05	0.05	0.00
	8	0.00	N/A	0.00	536.36	190.23	0.00	0.03	<b>0.00</b>
	9	0.00	N/A	0.00	536.36	9725.89	0.00	4.47	0.00
1.2	7	464.84	N/A	6.48	471.29	370.10	893.17	<b>4.48</b>	37.33
	8	27.14	N/A	<b>5.11</b>	377.45	503.17	893.17	7.32	9.48
	9	8.37	N/A	<b>4.65</b>	377.45	494.73	893.17	8.49	5.87
Ave.	7	984.65	N/A	<b>3.24</b>	610.99	536.39	446.61	2.26	18.67
	8	13.57	N/A	<b>2.55</b>	456.90	346.70	446.59	3.68	4.74
	9	4.19	N/A	<b>2.32</b>	456.90	5110.31	446.59	6.48	2.93
2.1	11	41.48	3.47	<b>2.93</b>	4.93	25.35	8.24	60.47	4.97
	12	29.81	3.34	3.21	<b>0.55</b>	2.40	11.72	22.77	5.09
	13	6.73	5.78	4.62	<b>0.55</b>	5.26	11.38	18.76	5.39
2.2	11	272.12	348.36	<b>0.08</b>	1114.19	14444.60	6.75	9.83	9.33
	12	366.53	348.36	<b>5.40</b>	1596.96	1137.59	7.06	8.97	9.69
	13	746.63	404.56	9.32	1596.96	730.67	<b>8.29</b>	9.35	8.15
2.3	11	Inf.	Inf.	<b>15.00</b>	Inf.	Inf.	Inf.	Inf.	Inf.
	12	Inf.	Inf.	0.74	Inf.	<b>0.69</b>	Inf.	Inf.	Inf.
	13	Inf.	Inf.	<b>0.37</b>	Inf.	0.38	Inf.	Inf.	Inf.
Ave.	11	Inf.	Inf.	<b>6.00</b>	Inf.	Inf.	Inf.	Inf.	Inf.
	12	Inf.	Inf.	<b>3.12</b>	Inf.	380.00	Inf.	Inf.	Inf.
	13	Inf.	Inf.	<b>4.77</b>	Inf.	245.00	Inf.	Inf.	Inf.

Table 15 also shows the average Euclidean distance of the estimated optima of the comparing methods to the real optima of the underlying functions.

**Table 15 The average Euclidean distance of the estimated optima from the real optima for N-ASRSM and comparing methods**

f(x)	Run	CCD	BBD	N-ASRSM	Standler et al. (2002)	Wang et al. (2003)	D-optimal	A-optimal	V-Optimal
1.1	7	3.99	N/A	<b>0.01</b>	1.88	2.06	0.22	0.01	0.02
	8	0.01	N/A	<b>0.01</b>	2.00	4.61	0.01	0.17	<b>0.00</b>
	9	0.01	N/A	<b>0.00</b>	2.00	5.76	0.01	2.11	<b>0.00</b>
1.2	7	2.92	N/A	<b>0.47</b>	2.09	1.63	2.18	0.44	0.82
	8	0.78	N/A	0.41	1.67	2.15	2.18	0.51	<b>0.33</b>
	9	0.53	N/A	<b>0.38</b>	1.67	2.85	2.18	0.54	0.48
Ave.	7	3.46	N/A	<b>0.24</b>	1.99	1.85	1.20	0.22	0.42
	8	0.39	N/A	0.21	1.83	3.38	1.09	0.34	<b>0.17</b>
	9	0.27	N/A	<b>0.19</b>	1.83	4.31	1.09	1.33	0.24
2.1	11	0.27	0.83	<b>0.19</b>	1.83	4.30	1.09	1.33	0.24
	12	2.46	0.71	<b>0.55</b>	0.94	2.37	0.92	1.94	0.98
	13	1.37	0.63	1.07	1.27	0.36	0.99	1.13	<b>0.26</b>
2.2	11	0.77	0.56	1.09	1.27	0.61	1.25	1.32	<b>0.23</b>
	12	1.82	1.60	<b>0.29</b>	1.98	4.26	0.76	0.85	0.88
	13	1.56	1.60	<b>0.73</b>	2.28	1.62	0.76	0.82	0.90
2.3	11	1.90	1.02	1.01	2.28	1.86	0.83	0.83	<b>0.82</b>
	12	3.33	3.42	<b>1.92</b>	1.95	2.32	3.71	3.71	3.80
	13	4.15	3.43	<b>0.83</b>	1.94	0.23	3.72	3.71	3.80
Ave.	11	3.29	3.10	<b>0.83</b>	1.94	0.59	3.80	3.69	3.80
	12	2.54	1.91	<b>0.92</b>	1.62	2.98	1.80	2.17	1.89
	13	2.36	1.89	<b>0.88</b>	1.83	0.74	1.82	1.89	1.65

### 3.8 Discussion

So far in this Chapter, we have developed and presented an adaptive methodology for n-dimensional quadratic response surface optimization. The proposed approach combines concepts from nonlinear optimization, design of experiments, and response surface optimization. The N-ASRSM is a sequential adaptive experimentation approach and uses the information gained from previous experiments to design the subsequent experiment by simultaneously reducing the region of interest and identifying factor combinations for new experiments. Its major advantage is the experimentation efficiency such that, for a given response target; it identifies the input factor combination (or containing region) in a smaller number of experiments than the classical single-

shot RSM designs. It differs from earlier studies in its optimality (under certain assumptions), inheritance of results from previous experiments, and robustness due to experiment ranking based reduction of the region of interest. Through large set simulated experiments, we showed that in modeling quadratic responses it outperforms the popular CCD, BBD and optimal designs in terms optimality. Based on another set of simulations we also showed that  $n - ASRSM$  performs considerably well in comparison to global optimization approaches in estimating the optima of non-linear responses. In the following section we extend the proposed strategy to work based on optimal designs which provides more flexibility in dealing with non-rectangular factor spaces and constraints.

### **3.9 An Extension of the Proposed Strategy with Optimal Design (N-ASRSM2)**

This section extends the proposed N-ASRSM strategy to work based on optimal design instead of fractional factorial design (N-ASRSM2). For this purpose, we first describe the terminology in Section 3.9.1. Next, we discuss the extended algorithm and its step in Section 3.9.2. Finally we provide a set of numerical examples in Section 3.9.3 to study the performance of the proposed strategy.

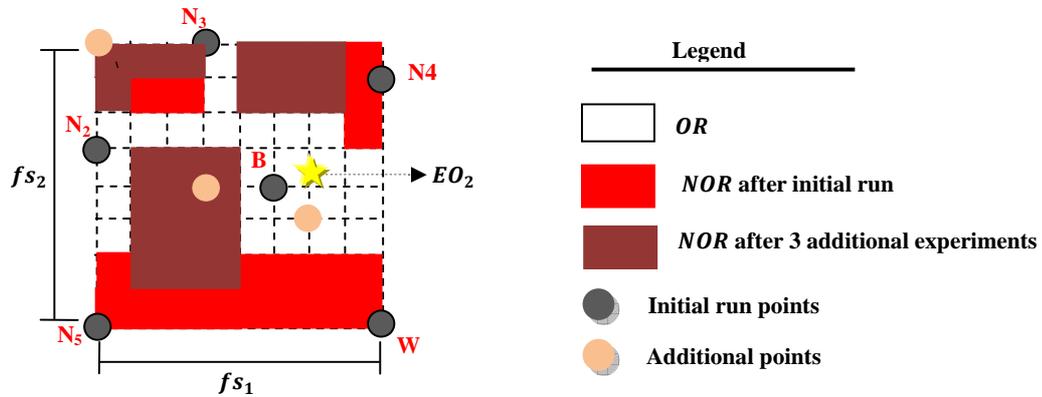
#### **3.9.1 Terminology and Assumptions**

Below we defined the additional notations required for extending the N-ASRSM method to N-ASRSM2. Figure 31 also shows a graphical representation of the notation:

$OR_e$  : Optimal region after  $e^{th}$  experiment

$EO_e$  : Estimated optimal experiment after experiment  $e$

$e^*$  : Index of augmenting experiments , e.g.  $e^*=1,2,\dots,E^*$  where  $E^*$  is the total number of augmenting experiments



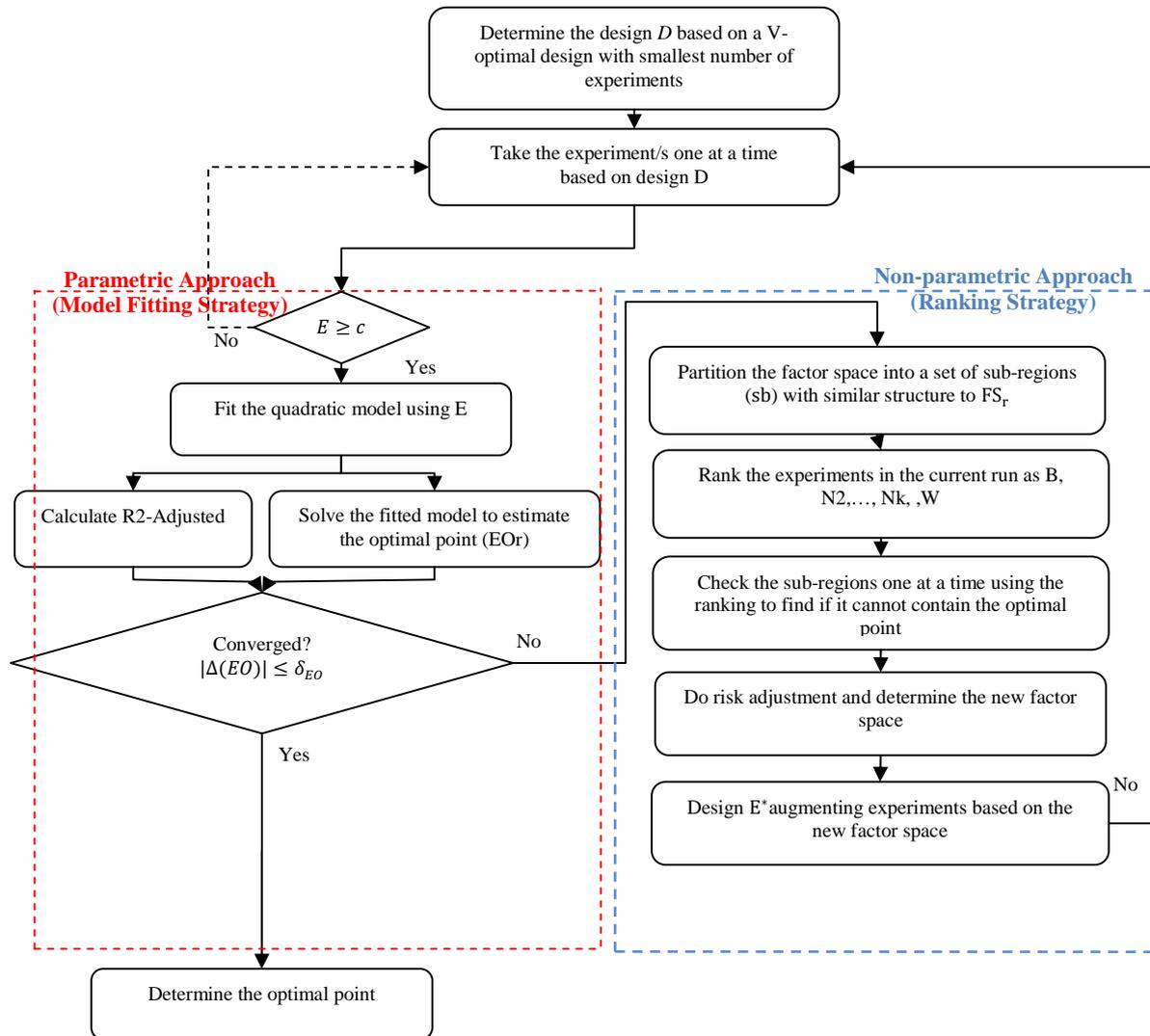
**Figure 31** An illustration of the terminology of the N-ASRSM2 on a two dimensional factor space

Similar to N-ASRSM, the extended methodology is based on the quadratic relation between the response and input factors. Though, the strategy can be built upon cubic assumption as well without any changes. The trade off is that the using cubic function as the underlying function provides the strategy with more flexibility in modeling nonlinearity while requires more experiments (because of its flexibility it is more difficult for cubic function to identify *NOR* with same number of experiments).

### 3.9.2 Algorithm

Figure 32 illustrates the general scheme of the proposed methodology. The initial run is setup with a modified version of optimal design, e.g. *V*- optimal design, augmented with additional experiment in following stages. Once the experimentation is completed, the approach follows two concurrent strategies, e.g., non-parametric ranking strategy and parametric model fitting strategy. Based on the ranking of experiments and the estimated optimal point from quadratic model fitting, a reduced factor space containing the estimated optimal experiment is determined

for the next experiments. This procedure continues until the convergence criteria based on estimated optimal experiment or coefficient of determination of the fitted model is attained.



**Figure 32** The general scheme of the proposed N-ASRSM2 strategy

The details of the steps shown in Figure 30 are similar to what discussed for N-ASRSM (See also Figures 31 and 32). Though, one of the few differences is that there is no separate runs here (each experiment can be considered as a run). Also, for augmenting the design, the new experiments are taken (based on the optimal design) on the  $FS-NOR$  (the remaining factor space

after taking out non optimal regions). This can be easily done by adding a set of constraint to the objective function of optimal design for the following run. For instance if we are using a  $D$ -optimal design and the first few experiments result in the Figure 31(a)  $NOR$ . The optimization model for choosing the next point can be written as follows:

$$\begin{aligned}
 & \text{Min } Z = |(X'X)^{-1}| \\
 & \text{S. T} \\
 & \frac{\max[(X-(-1)),0]}{(X-(-1))} \times \frac{\max[(Y-0.75),0]}{(Y-0.75)} \times \frac{\min[(X-(0.75)),0]}{(X-(0.75))} \times \frac{\min[(Y-(0.5)),0]}{(Y-(0.5))} = 0 \\
 & \frac{\max[(X-0.5),0]}{(X-0.5)} \times \frac{\max[(Y-0.5),0]}{(Y-0.5)} \times \frac{\min[(X-(0.25)),0]}{(X-(0.25))} \times \frac{\min[(Y-(0.25)),0]}{(Y-(0.25))} = 0
 \end{aligned}
 \tag{14}$$

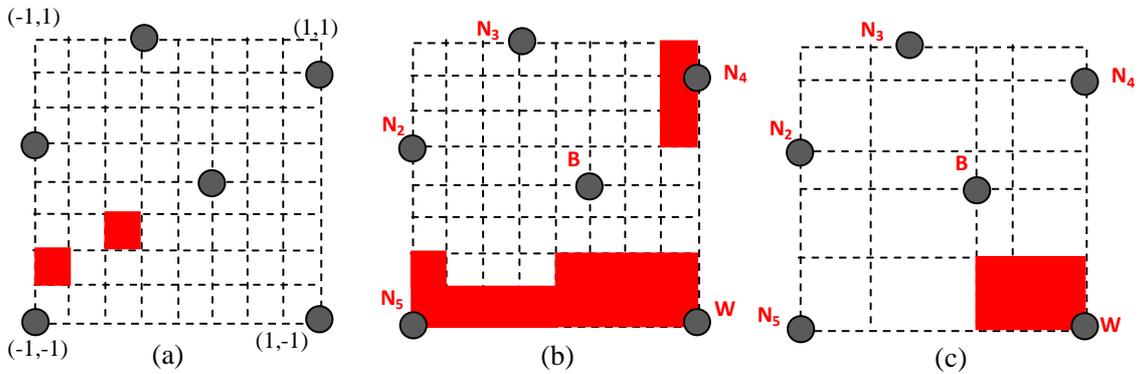


Figure 33 (a)  $NOR$  of the N-ASRSM2 factor space after 6<sup>th</sup> experiments (b) a grid of 64 sub-regions of a 2 dimensional FS (c) a grid of 16 sub-regions of a 2 dimensional FS in the proposed N-ASRSM strategy

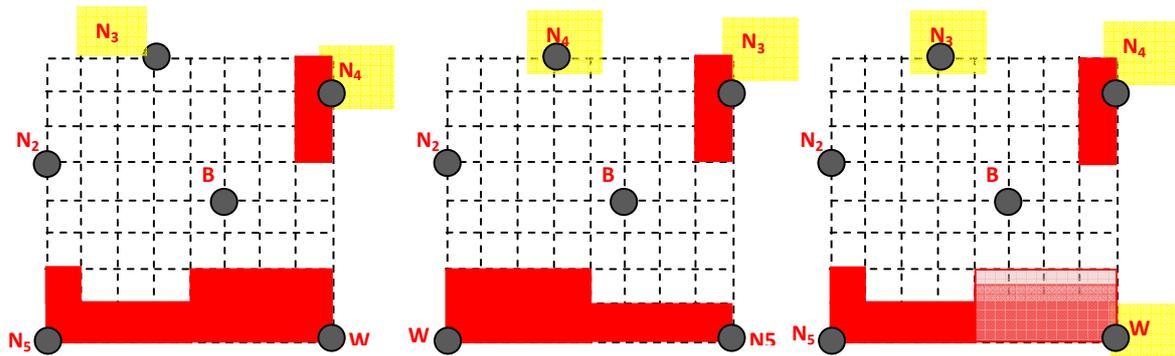


Figure 34 Cut out regions of N-ASRSM2 using the ranking strategy based on two additional points.

### 3.10 Numerical Examples

Here, we present two small sets of simulated experiments to evaluate the performance of the proposed N-ASRSM2 approach. In the first set of simulations N-ASRSM2 is compared to CCD, and A, D, and V optimal designs on quadratic response models. In the second set, the proposed strategy is compared to the classical models, optimal designs and two global optimization methods (Standler et al. (2002), Wang et al. (2003)) on a number of non-linear response models.

#### 3.10.1 Quadratic Response Models

We now describe the simulated experiments performed to compare the performance of the proposed N-ASRSM2 with those of CCD, and optimal designs on quadratic response models. In the simulated experiments, we have considered 3 of response models of the ones studied in the previous section. These response models are presented in Table 16.

**Table 16 The quadratic response models used in the simulated experiments of the N-ASRSM2**

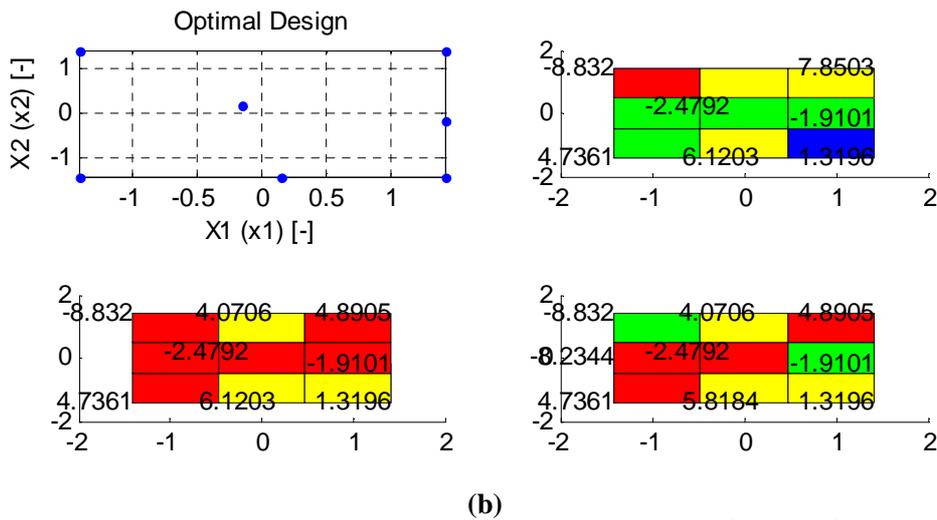
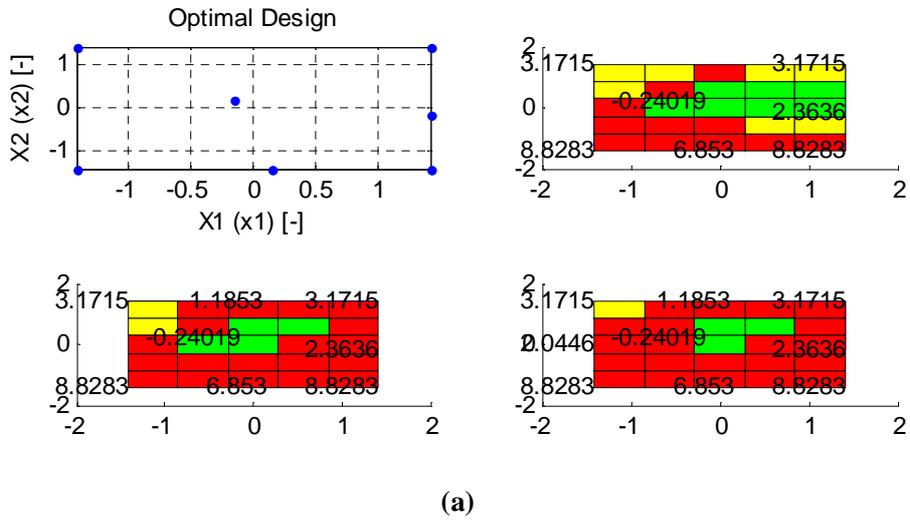
No. of variables	Res. No.	Response Relation	Error ( $\epsilon$ )	Response Type
Two Variable Response	1.1	$W = -2x^2 + 3y^2 + 2x - y + 2xy - 1 + \epsilon$	$N(0, 0.1)$	Non-convex
	1.2	$W = x^2 + 2y^2 - 2y + \epsilon$	$N(0, 1)$	Convex
	1.3	$W = -2x^2 + 3y^2 + 2x - y + 2xy - 1 + \epsilon$	$N(0, 2)$	Non-convex

Table 17 compares the average  $R_{adj}^2$  of the comparing methods. Each of the experiments has been replicated 3 times, so each number in the table is the average of three replicates. Similar to the previous study the bold numbers in each row represents the best of the row. Based on the Table, N-ASRSM2 performs very well on  $R_{adj}^2$  in comparison to other methods. Figure 35

illustrates a graphical representation of the NASRSM2 grid, *FS*, *OR* and *NOR* during 7<sup>th</sup> to 9<sup>th</sup> experiments.

**Table 17 The average  $R_{adj}^2$  of the N-ASRSM2 an comparing methods for trials 7, 8 and 9 of responses for quadratic functions**

Exp. No.	No Obs.	Adjusted R <sup>2</sup>					
		CCD	N-ASRSM2	N-ASRSM	D-Opt.	V-Opt.	A-Opt.
1.1	7	99.96%	<b>99.99%</b>	<b>99.99%</b>	99.96%	99.92%	99.94%
	8	99.96%	<b>99.99%</b>	<b>99.99%</b>	99.98%	99.95%	99.95%
	9	99.95%	<b>99.99%</b>	99.98%	99.97%	99.94%	99.92%
1.2	7	92.69%	<b>99.99%</b>	93.67%	95.01%	94.85%	90.69%
	8	92.48%	<b>99.99%</b>	94.97%	89.20%	95.86%	86.86%
	9	92.00%	<b>98.00%</b>	96.62%	90.42%	95.42%	86.58%
1.3	7	69.77%	<b>99.44%</b>	79.57%	92.96%	72.11%	97.71%
	8	70.60%	<b>99.67%</b>	86.53%	91.06%	74.92%	79.86%
	9	88.48%	92.55%	<b>92.91%</b>	89.02%	80.89%	82.00%
Avg.	7	69.22%	<b>99.81%</b>	90.21%	91.92%	87.19%	86.69%
	8	71.74%	<b>99.88%</b>	91.58%	87.90%	86.25%	83.86%
	9	74.84%	<b>96.85%</b>	94.48%	84.60%	88.12%	84.86%



**Figure 35** The graphical representation of the N-ASRSM2 on 2<sup>nd</sup> and 3<sup>rd</sup> response surface; the blue cells show *OR*, greens show *OR* not robust to miss-ranking (in risk adjustment), Yellows show non-robust *NOR* and Reds show robust *NOR*

Table 18 presents the average optimality gap of the 7<sup>th</sup> to 9<sup>th</sup> trials of the comparing methods gained based on the average of three replicates. The experiments show that the optimality gap of the proposed N-ASRSM2 is comparable to the best performing methods.

**Table 18 The Average optimality gap of the N-ASRSM2 and comparing methods for trials 7, 8 and 9 of responses for quadratic responses**

Exp. No.	No Obs.	Optimality gap					
		CCD	N-ASRSM2	N-ASRSM	D-Opt.	V-Opt.	A-Opt.
1.1	7	3.84	<b>0.00</b>	0.00	<b>0.00</b>	31.76	<b>0.00</b>
	8	3.84	<b>0.00</b>	0.07	<b>0.00</b>	18.35	4.70
	9	0.23	<b>0.00</b>	0.04	0.26	<b>0.00</b>	<b>0.00</b>
1.2	7	70.52	0.23	0.01	<b>0.00</b>	0.61	40.19
	8	70.52	0.86	0.72	<b>0.53</b>	0.76	29.94
	9	70.52	0.41	<b>0.00</b>	7.69	0.83	8.24
1.3	7	909.82	<b>0.31</b>	<b>0.31</b>	<b>0.31</b>	<b>0.31</b>	1260.68
	8	909.82	<b>0.31</b>	<b>0.31</b>	<b>0.31</b>	<b>0.31</b>	492.09
	9	909.82	<b>0.31</b>	<b>0.31</b>	<b>0.31</b>	<b>0.31</b>	5.00
Avg.	7	328.06	0.18	0.11	<b>0.10</b>	10.89	433.62
	8	328.06	0.39	0.37	<b>0.28</b>	6.47	175.58
	9	326.86	0.24	<b>0.12</b>	2.75	0.38	4.41

### 3.10.2 Non-linear Response Models

Here we compare the performance of the proposed N-ASRSM2 approach with the same two dimensional examples as the previous section and the same comparing methods. These response models are presented in Table19. Again all the simulations has been replicated three times and the comparisons are based on the average of three replicates.

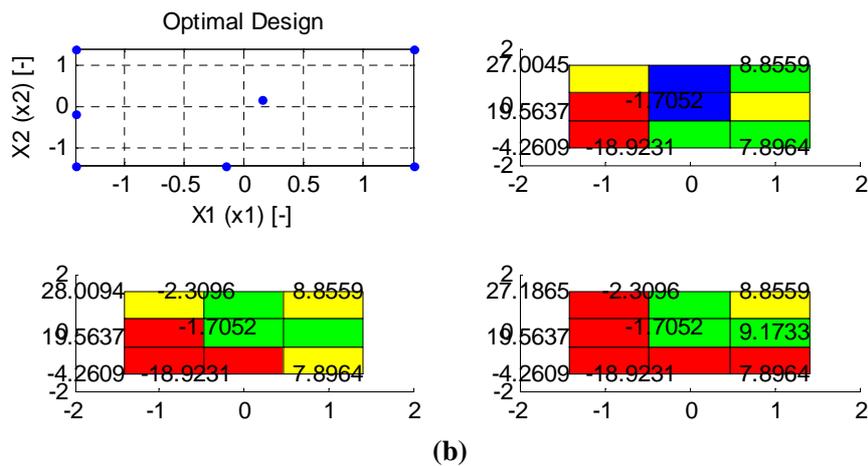
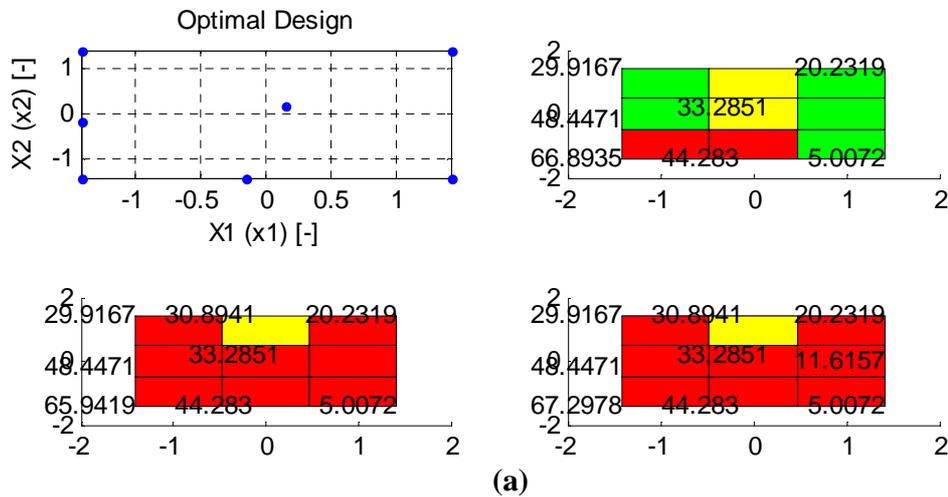
**Table 19 The non-linear response surface models used for studying N-ASRSM2 performance**

Exp. No.	Response Relation	Error
1	$W=(y-(1/(8\pi^2)*x^2)+(10/\pi)(x-2)^2+10(3-1/(12\pi))\cos(x)$	$N(0,2)$
2	$W=0.75(x-0.15)^2+.25*(x-0.15)^4+1.3*(x-0.15)^6+1.8(x-0.15)(y-1)^2-2.66(y-1)^2+1.9(y-0.15)^2$	$N(0,2)$

Table 20 shows the average optimality gap of the consecutive trials of the comparing methods. Based on the results N-ASRSM and N-ASRMS2 are the best performing methods while N-ASRSM performs slightly better. Figure 36 shows the changes in the *OR*, *NOR* of the *FS* in the N-ASRMS2 through different experiments.

**Table 20 The average optimality gap of the proposed N-ASRSM2 and comparing methods**

f(x)	Run	CCD	N-ASRSM2	N-ASRSM	Standler et al. (2002)	Wang et al. (2003)	D-optimal	A-optimal	V-Optimal
1.1	7	1504.460	<b>0.000</b>	<b>0.000</b>	750.680	702.682	0.050	0.051	0.001
	8	0.004	<b>0.000</b>	<b>0.000</b>	536.355	190.227	0.003	0.029	<b>0.000</b>
	9	0.004	<b>0.000</b>	<b>0.000</b>	536.355	9725.895	0.003	4.470	<b>0.000</b>
1.2	7	464.841	9.662	6.478	471.292	370.099	893.174	<b>4.478</b>	37.333
	8	27.141	8.303	<b>5.109</b>	377.449	503.172	893.174	7.322	9.480
	9	8.367	5.347	<b>4.650</b>	377.449	494.735	893.174	8.488	5.866
Ave.	7	984.650	4.831	<b>3.239</b>	610.986	536.391	446.612	2.264	18.667
	8	13.573	4.152	<b>2.555</b>	456.902	346.699	446.589	3.676	4.740
	9	4.186	2.674	<b>2.325</b>	456.902	5110.315	446.589	6.479	2.933



**Figure 36 The graphical representation of the N-ASRSM2 on 1<sup>st</sup> and 2<sup>nd</sup> response surface; the blue cells show *OR*, greens show *OR* not robust to miss-ranking (in risk adjustment), Yellows show non-robust *NOR* and Reds show robust *NOR***

Table 21 presents the average Euclidean distance of the estimated optima from the real optima of the comparing methods which have almost similar result to Table 20. Again on average both proposed methods works on a par with or better than optimal designs and CCD.

**Table 21 The average Euclidean distance of the estimated optima from the real optima for the N-ASRSM2 and comparing methods**

f(x)	Run	CCD	N-ASRSM2	N-ASRSM	Standler et al. (2002)	Wang et al. (2003)	D-optimal	A-optimal	V-Optimal
1.1	7	3.994	0.008	<b>0.005</b>	1.877	2.063	0.224	0.007	0.016
	8	0.006	<b>0.004</b>	0.005	1.996	4.606	0.007	0.171	<b>0.004</b>
	9	0.006	<b>0.004</b>	<b>0.004</b>	1.996	5.758	0.007	2.114	<b>0.004</b>
1.2	7	2.92	0.56	<b>0.47</b>	2.09	1.63	2.18	0.44	0.82
	8	0.78	<b>0.32</b>	0.41	1.67	2.15	2.18	0.51	<b>0.33</b>
	9	0.53	0.42	<b>0.38</b>	1.67	2.85	2.18	0.54	0.48
Ave.	7	3.46	0.28	<b>0.24</b>	1.99	1.85	1.20	0.22	0.42
	8	0.39	<b>0.162</b>	0.21	1.83	3.38	1.09	0.34	0.17
	9	0.27	0.21	<b>0.19</b>	1.83	4.31	1.09	1.33	0.24

### 3.11 Discussion

Here, we have extended the N-ASRSM to consider optimal design instead of fractional factorial design. Compared to N-ASRSM, the proposed N-ASRSM2 approach provides more flexibility in working with non-rectangular factor spaces, and grids with different resolution. In addition to getting benefited from optimal designs advantage N-ASRSM2 can employ cubic underlying function instead of quadratic function which can enable it to model nonlinearity better (although there the trade of increasing the number of experiments required for eliminating non-optimal regions). Based on the simulation studies, the proposed N-ASRSM2 strategy works robustly well on both quadratic and non-linear responses, in comparison to classical and optimal designs. However its performance on quadratic responses is slightly worse than its counterparts while on nonlinear responses N-ASRSM works slightly superior.

## CHAPTER 4 THE PROPOSED STRATEGIES FOR BLACK BOX FUNCTIONS WITH N-VARIABLES (B-ASRSM)

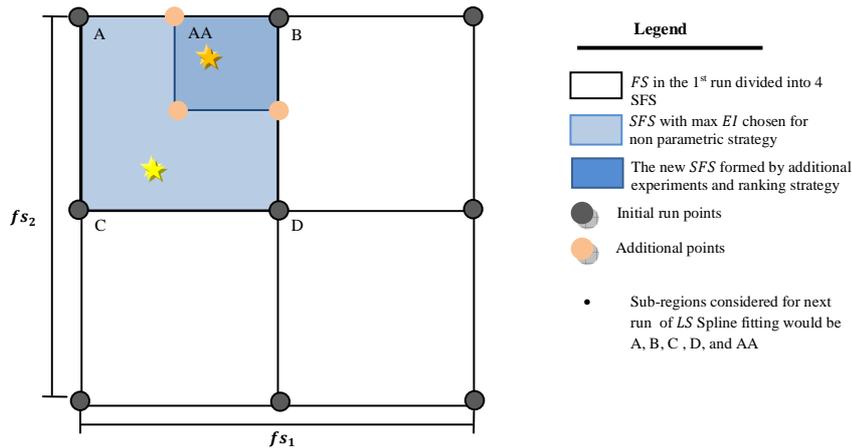
This chapter presents the detailed elements of the proposed adaptive sequential strategies for noisy expensive black-box function (B-ASRSM). For this purpose, we first describe the terminology that is mostly borrowed from previous Chapters. Next, we provide an overview of the methodology and its core components. Then, we explain how these components are integrated together to form B-ASRSM. Finally, we provide a small numerical example to examine the performance of the proposed strategy. We also extend the proposed strategy (B-ASRSM2) to consider multiple adaptive regression splines which provides the proposed model more flexibility in dealing with high dimension.

### 4.1 Terminology and Assumptions

As expected the definitions and terminology used in the proposed strategy has many commonalities with previous Sections. For more convenience we provide the complete notations below. Figure 37 illustrates some of the notations on a two-dimensional factor space with 9 initial experiments. The only assumption that we consider here is the normality of error, e.g.  $\varepsilon \sim (0, \sigma^2)$ .

- $di$  : Number of factors (dimensions) in the  $FS$
- $k_{a_i} - 1$  : Number of knots in each factor
- $EI$  : Expected improvement
- $FS_r$  : Factor space at run  $r$  and expressed as Cartesian product of factor ranges in run  $r$
- $SFS_r^s$  :  $s^{\text{th}}$  sub-factor space of at run  $r$
- $fs_i$  : Initial range of factor  $i$
- $D$  : Design of most current run

- $D$  : minimum number of required points in  $D$
- $COP$  : A corner point experiment run at the intersection of extrema of factor ranges
- $CEP$  : A center point experiment run at the center of gravity of the factor space
- $r$  : Index of runs, e.g.  $r = 1, 2, \dots, R$  where  $R$  is the total number of runs
- $e$  : Index of experiments in a given run, e.g.  $e = 1, 2, \dots, E$  where  $E$  is the total number of experiments
- $B$  : The experiment with the best response level in a  $SFS$  and a given run
- $N_k$  : The experiment with the  $k^{th}$  best response level in a  $SFS$  and a given run ( $2ek \leq e - 1$ )
- $W$  : The experiment with the worst response level in a  $SFS$  and a given run
- $OR_r$  : Optimal region/s in run  $r$  containing the estimated optimal experiment,  $OR_r \subseteq FS_r$
- $NOR$  : Non-optimal region/s
- $O$  : Optimal experiment, e.g. best experiment in the initial factor space
- $RO$  : Real optima of the function
- $EO_r$  : Estimated optimal experiment in run  $r$ , e.g., best incumbent estimation of the optimal experiment
- $sb_i^{(r,s)}$  : sub-regions  $i$  in  $s^{th}$   $SFS$  of the  $r$ th run in a given factor, e.g.  $sb = 1, 2, \dots, mn$ , where  $m^n$  is the total number of sub-regions
- $c$  : Number of coefficients of the underlying model
- $P_l$  : Probability of losing  $RO$  but cutting out



**Figure 37** An illustration of the terminology of the proposed B-ASRSM strategy

## 4.2 Algorithm

Figure 38 illustrates the general scheme of the proposed strategy. The initial run is setup with a modified version of  $k$  level factorial design which divides the factor space into  $k_1 \times \dots \times k_{d_i}$  sub (hyper) rectangular regions. Once the experimentation is completed, considering each of the divisions as a spline sub-region, a quadratic least square regression spline (LSRS) is fitted. The next step is considering each divisions as sub-factor space (*SFS*) and candidate the *SFS* with maximum expected improvement for further breaking. This step is followed by applying chapter 3 non-parametric and parametric approaches to find the optimal and non optimal sub-regions in the candidate *SFS*, and taking experiment on its empty corners to make a new sub-regions for the next run of the algorithm (fitting LSRS). This procedure continues until the convergence criteria based changes in the expected improvement  $|\Delta(EI)| \leq \delta_{EI}$  is attained.

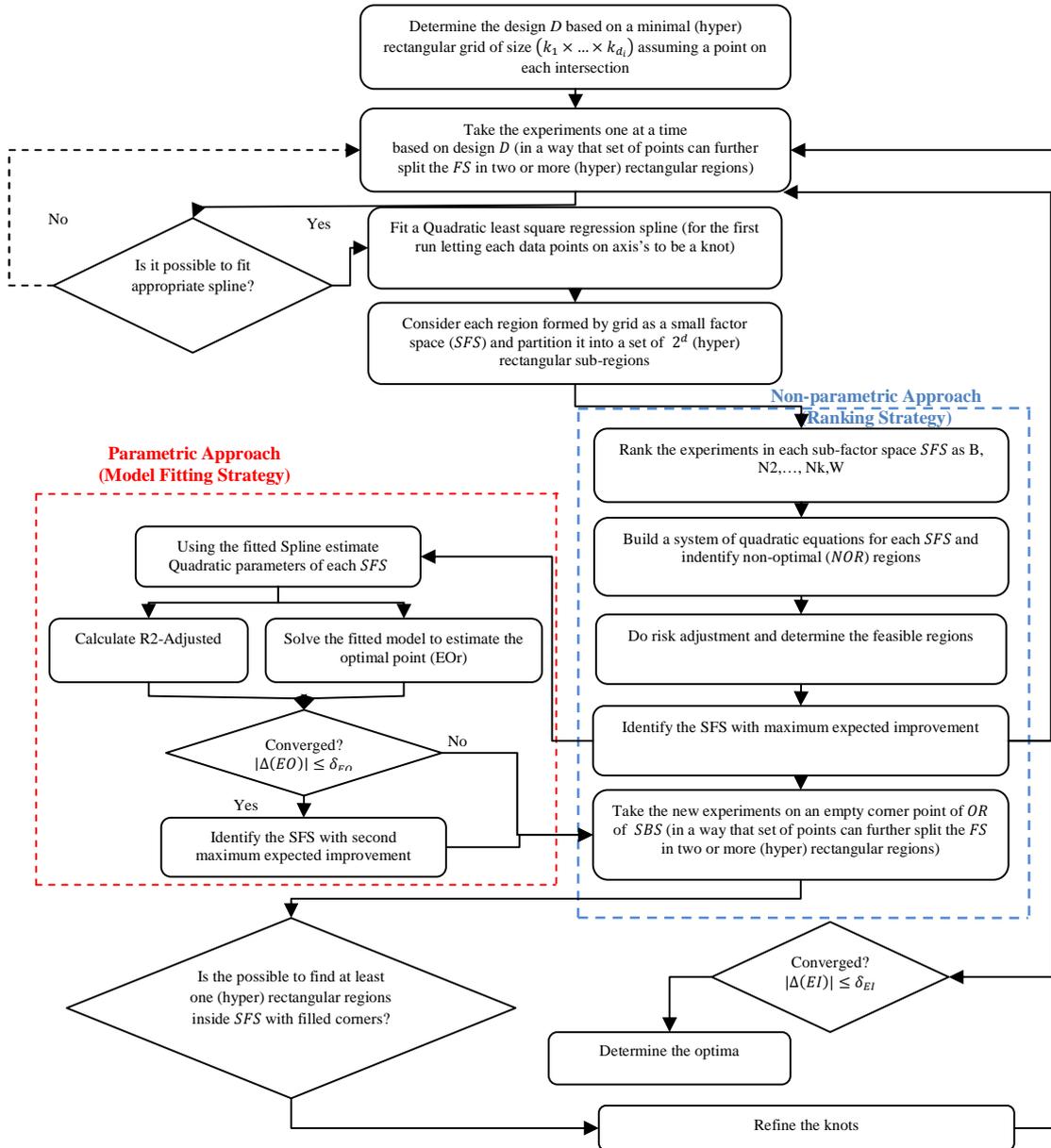
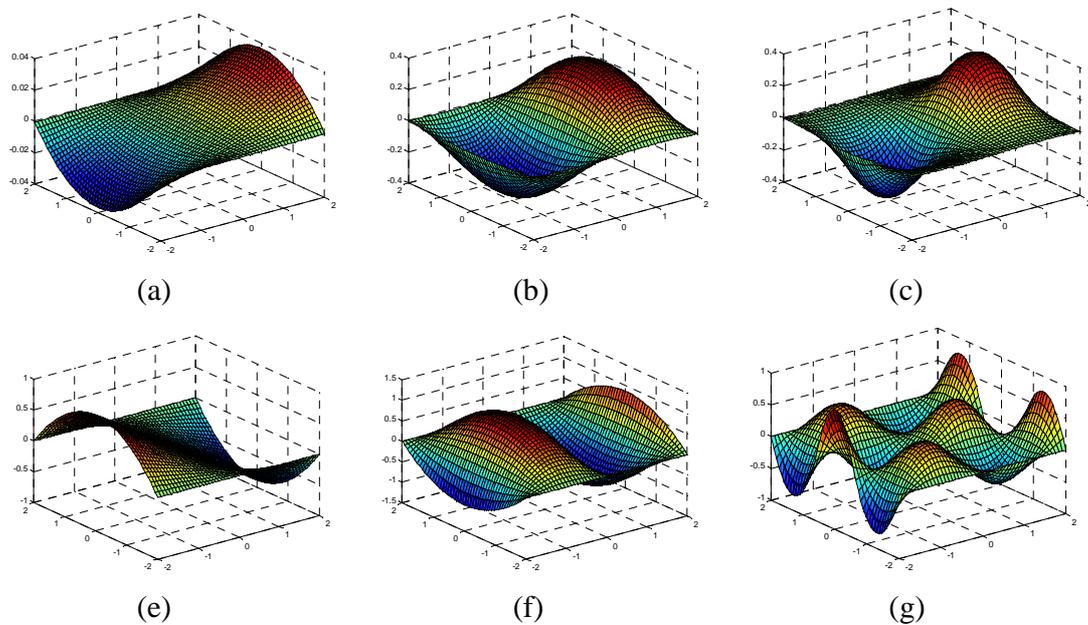


Figure 38 the general scheme of the proposed B-ASRSM

### 4.3 Design structure of the first and subsequent runs

The design  $D$  structure of the factor space  $FS_r$  in the proposed approach is adapted from  $k$  level factorial design (here we use  $k_1, \dots, k_{d_i}$  to differentiate among the number of levels along different factors (dimensions)). The choice  $k_i$  is important and contains tradeoff between

accuracy and expense. Setting each  $k_i$  to a very small value, e.g. 2 or 3, save a lot of experimentation cost while may result in not capturing true curvature of the underlying function and as consequence exploring wrong region/s. in contrast setting  $k_i$  to a larger value, e.g. 5 or more, provides enough confidence about the underlying curvature but with a lot of cost of experimentation. Based on the simulated experiments 4 and 5 can be good choice when there is no information about the underlying function bumpiness. Figure 35 shows the effect of different choice of  $k$  on the estimation of surface for two examples.



**Figure 39** The effect of number of levels  $k_i$  on the estimated surfaces: (1)  $z = x \cdot \exp(x^2 - y^2)$  with  $k=(2,2)$ ,  $(3,3)$  and  $(4,4)$  in (a), (b) and (c) and, (2)  $z = \sin(0.83\pi x) \cdot \cos(1.25\pi y)$  with  $k=(2,2)$ ,  $(3,3)$  and  $(4,4)$  in (e), (f) and (g)

The initial design may be further augmented with few more experiments inside of the *SFS* which is more probable to contain the *RO* to break it down further and provide better estimation.

This result in a nested set of factor spaces in different runs like before, e.g.

$$FS_r = \varphi_r \left( \varphi_{r-1} \left( \dots \varphi_0 (FS_1) \right) \right).$$

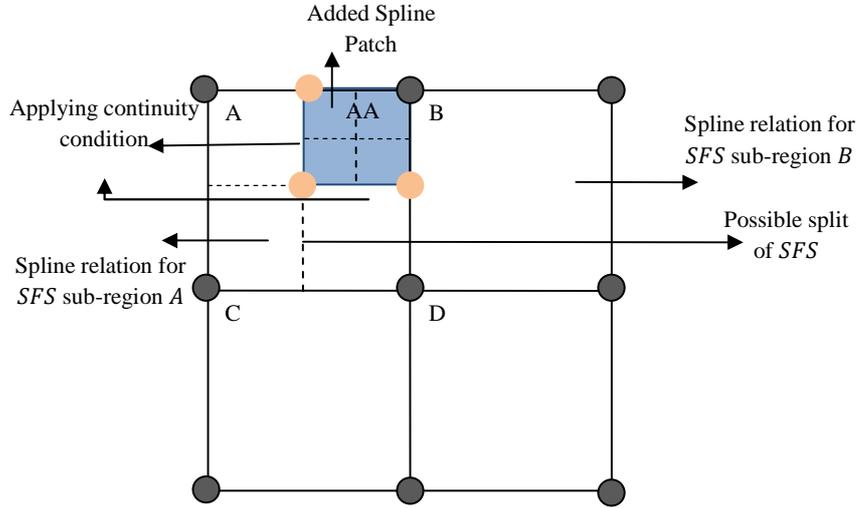
Regarding the  $FS$  size, since in general spline functions do not work very well on boundaries, also to alleviate the effect of noise, similar to former strategies, it is suggested to start with a broader initial region where applicable. Based on the simulated experiments a  $FS$  of %5-%10 larger than the typical setting can result in better accuracy.

#### 4.4 Fitting Least Square Regression Spline (LSRS)

After taking the experiments based on the design  $D$ , considering each (hyper) rectangular region inside the  $FS$  as a spline sub-region, a multivariate quadratic least square regression spline is fitted to the data (See Chapter 1 for detail formulation of LSRS). Since each (hyper) rectangular region inside the  $FS$  is considered as a spline sub-region, in the first run, the breaks in each factor (dimension) would be equal to the factor levels in that dimension. This combination is changed in the subsequent runs by adding a patch after breaking down one or more of the  $SFS$  (See Figure 40). As a consequence, the quadratic regression model fitted to the patch, estimate the system behavior within the patch boundaries while Least Square Regression Spline ( $LSRS$ ) fitted to rest of  $FS$  estimate for the remaining part. The patch model which is formulated by fitting a quadratic function to data within its boundaries gets connected to the rest of spline by applying continuity conditions at their intersections which is done by making the estimated function values a first well as the first and second derivative agrees at the intersections:

$$\begin{cases} F_{patch}(X_{intersection}) = F_{Spline}(X_{intersection}) \\ \frac{\partial F_{patch}(X_{intersection})}{\partial X} = \frac{\partial F_{Spline}(X_{intersection})}{\partial X} \\ \frac{\partial^2 F_{patch}(X_{intersection})}{(\partial X)^2} = \frac{\partial^2 F_{Spline}(X_{intersection})}{(\partial X)^2} \end{cases} .$$

(15)



**Figure 40** An example of the fitted LSRS with a spline patch added

In order to reduce the number of experiments (points) and comply further with the original quadratic regression function, in the proposed model the terms with total order more than two in the regression spline (which occurs in the problems with two or more factors) can be taken out. For example in a two dimensional factor space with set of  $K$  knots  $\{\xi_1, \dots, \xi_K\}$  and the regression spline relation  $\sum_{k=1}^K \sum_{i=0}^2 \sum_{j=0}^2 c_{ijk} (x_1 - \xi_k)_+^i (x_2 - \xi_k)_+^j$  the terms  $c_{22k} (x_1 - \xi_k)_+^2 (x_2 - \xi_k)_+^2$  and  $c_{21k} (x_1 - \xi_k)_+^2 (x_2 - \xi_k)_+^1$  and  $c_{12k} (x_1 - \xi_k)_+^1 (x_2 - \xi_k)_+^2$  may be omitted from the further consideration without affecting the model accuracy very much (the reason is that all these term are related to interactions of high order which are typically of very small effect). This consideration can be easily added to the LSRS by adding constraint imposing zero value for the coefficients of the above terms.

#### 4.5 Identify the SFS with Maximum Expected Improvement

One of the advantages of fitting LSRS is having a set simple quadratic regression function for each *SFS* as a result. Based on the fitted quadratic regression function, the response at each point (of each *SFS*) will have the following *t*-student distribution:

$$\hat{y}_0 \sim t_{n-p}(X'_0 \hat{\beta}, \hat{\sigma}^2 X'_0 (X'X)^{-1} X_0) \quad (16)$$

Where  $X$  is the matrix of experiments input in the quadratic form (only the point fallen in each *SFS* contribute to the  $X$  matrix of that *SFS*),  $X_0$  is the query input which should also be in quadratic form,  $\hat{\beta}$  is the fitted regression parameters, and  $\hat{\sigma}$  is the estimated standard deviation (*MSE*).

Knowing the distribution of response, the expected improvement of each point can be calculated as follows:

$$EI(X) = \int_0^{\infty} d \cdot p(y < d) dd \quad (17)$$

where  $d = \min(f(x)) - \alpha(F_{max} - F_{min})$  and  $\min(f(x))$  is the minimum of the fitted spline over the whole *FS*,  $\alpha$  is coefficient related to the rate of reduction (suggested  $\alpha$  values are 0.1 to 0.25), and  $F_{max}, F_{min}$  are the minimum and maximum of the observations (for more information about  $d$  See Jones (2001)). In our proposed strategy, the *SFS* which contains the point with maximum expected improvement is the candidate for further break down (It should be noted that other measures such as estimated minimum *EO* or maximum variance can also be used for identifying the *SFS* for breaking down).

## 4.6 Taking New Experiments and Breaking down Structure of the Methodology

After finding the *SFS* with maximum expected improvement, it should be furthered explored to:

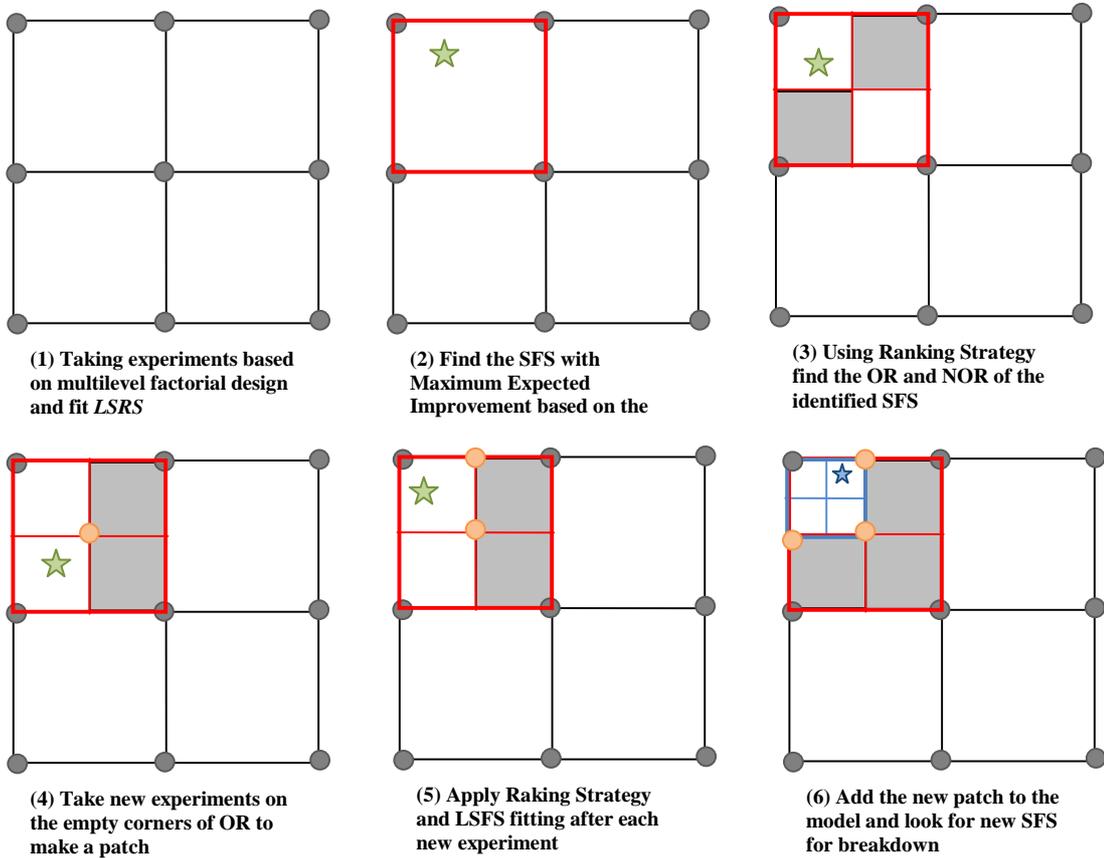
1. Check if the speculation on the location of the minima is correct
2. Monitor the convergence of the algorithm
3. Improve the estimation of the underlying function, especially around the global minimum

This is typically done by adding one or more points to the *SFS*. However, the location of the points should be chosen with the goal of quickly locating at least one smaller (hyper) rectangular region containing the *OR* inside the candidate *SFS* (the region is used as a new spline patch in the fitted *LSRS*). For this purpose, the first additional point would be chosen at the center of the candidate *SFS*, and the next points would be chosen based on the ranking strategy in Chapter 2. In this sense the algorithm for the breaking down the *FS* to *SFSs* towards the *RO* can be seen as a form of big square small square optimization (Plastria 1992) (See Also Figure 41).

## 4.7 Design of the Next Runs and Stopping Criteria

Completing the initial run of the algorithm after adding each new point the *LSRS* is fitted to the data to check the location of *SFS* with maximum improvement and to form a spline patch around the *EO*. This procedure continues until the expected improvement at two consecutive points gets lower than a threshold  $|\Delta(EI)| \leq \delta_{EI}$  is attained or the pre-specified number of

experiments is reached. Figure 41 presents a simple graphical representation of the proposed strategy steps.



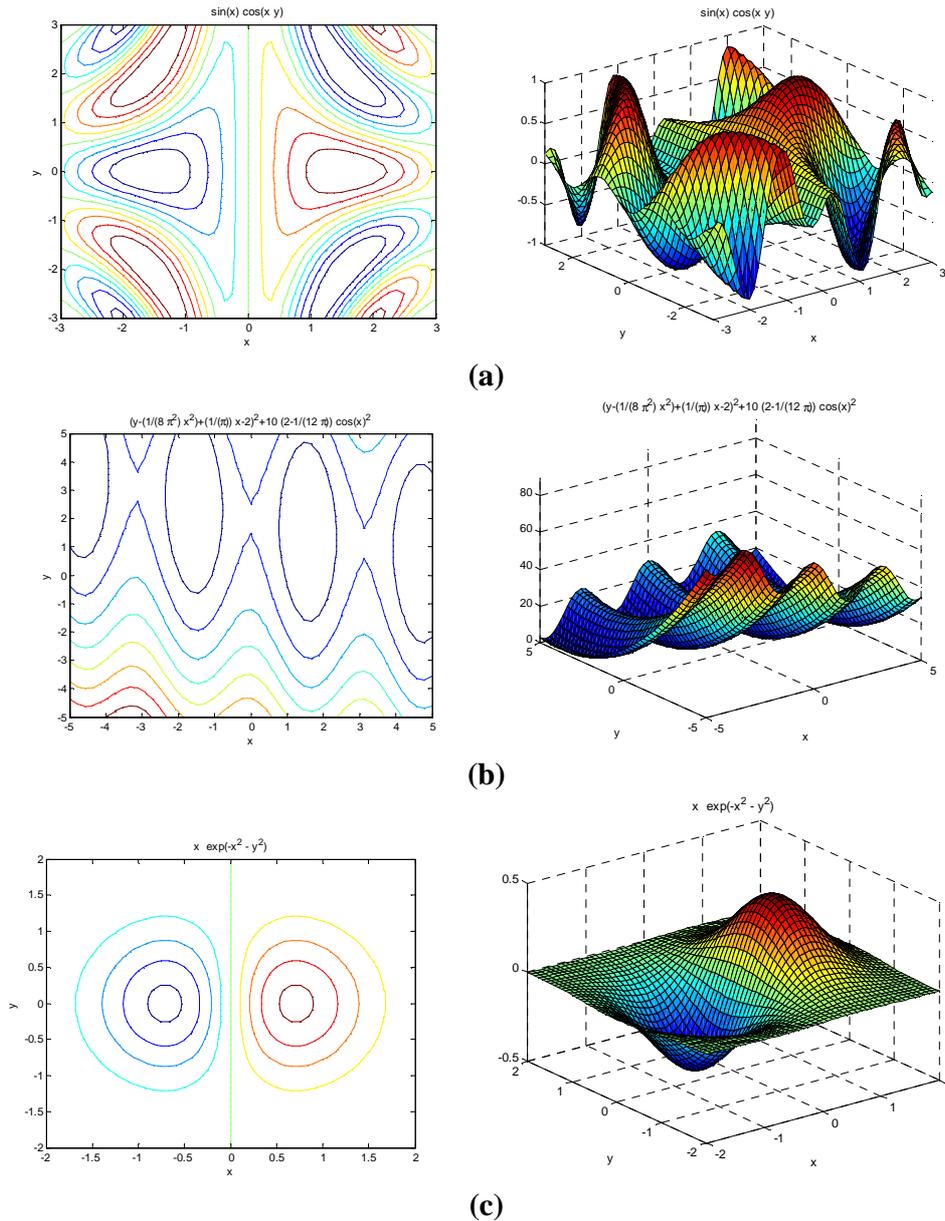
**Figure 41** the graphical representations of the proposed B-ASRSM strategy steps

## 4.8 Numerical Examples

In this section, we describe a small set of simulated experiments performed to evaluate the performance of the proposed strategy along with two popular global optimization methods Radial Basis Function (RBF) and Gaussian Process (GP). For modeling RBF we used ARESLab toolbox version 1.5 for Matlab/Octave written by Jekabsons. Also for modeling GP we use Gaussian Toolbox version 3.1 for Matlab/Octave written by Rasmussen and Nickisch. The response models are presented in Table 16 and Figure 42.

**Table 22 The response relation considered for the analysis of the B-ASRSM**

No. of variables	Res. No.	Response Relation	Range	Error
Response with Two variable	1	$W=(y-(1/(8*\pi.^2)*x.^2)+(1/\pi)).*x-2).^2+10*(2-1/(12*\pi)).*\cos(x)^2$	[-5, 5]	$N(0,3)$
	2	$W=\sin(x) \cos(x y)$	[-3 3]	$N(0,0.05)$
	3	$W=x.* \exp(-x.^2 - y.^2)$	[-2 2]	$N(0,0.1)$

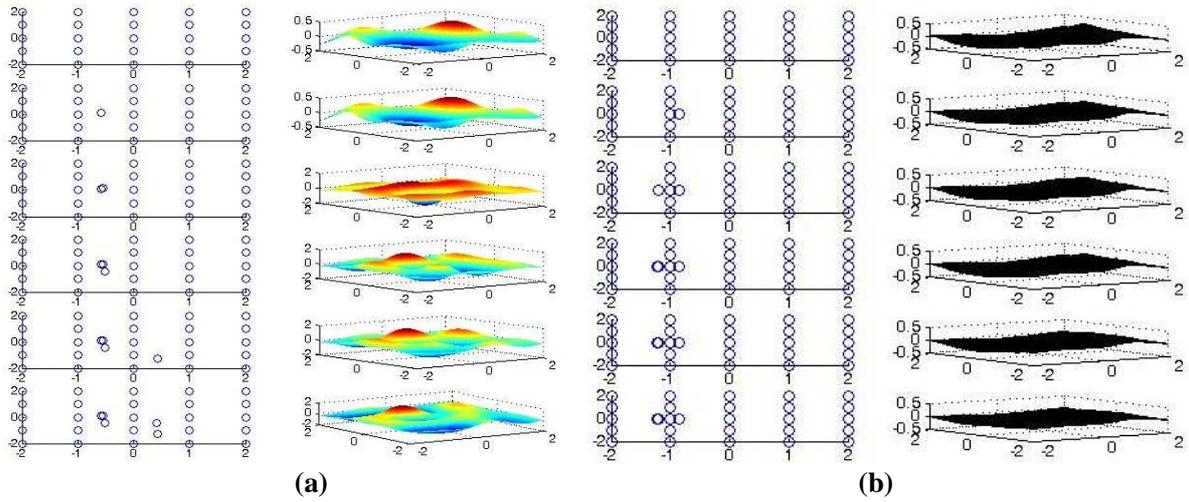


**Figure 42 The contour and function plot of the responses considered for the analysis of the B-ASRSM; a, b, c represents 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> response relations respectively**

For the following analysis, we have compared the performances based on average optimality gap and Euclidian distance of the estimated optima to the real optimal point. All simulated experiments are repeated two times, and the average of the results is reported. To have a fair comparison, same initial design structure (multilevel factorial design) has been considered for all three comparing methods. Figure 43 illustrates the location of the experiments and the estimated surface of RBF and GP methods in different iterations. Table 23 shows the average optimality gap results and average Euclidian distance to the *RO* of the consecutive trials of the comparing methods. As can be seen from the Table the proposed method not only have an acceptable rate of convergence toward the real optima, but also it has comparable results to those of RBF and GP.

**Table 23 The average optimality gap and average Euclidean distance to *RO* of the proposed B-ASRSM methods along with the comparing methods**

Exp. No.	No Obs.	Optimality gap of Response			Euclidean distance to <i>RO</i>		
		B-ASRSM	RBF	GP	B-ASRSM	RBF	GP
1	25	8.72	4.99	<b>2.63</b>	<b>1.75</b>	3.44	6.93
	26	<b>3.30</b>	11.53	3.80	<b>1.26</b>	3.80	7.21
	27	39.36	10.10	<b>2.86</b>	1.54	<b>0.91</b>	0.92
	28	12.95	29.56	<b>2.26</b>	1.34	1.46	<b>0.98</b>
	29	<b>2.30</b>	6.27	6.27	<b>0.75</b>	1.86	0.98
	30	1.29	4.23	<b>0.24</b>	<b>0.34</b>	2.68	0.70
2	141	0.69	0.53	<b>0.04</b>	1.89	2.00	<b>0.33</b>
	142	0.90	0.50	<b>0.03</b>	3.74	1.68	<b>0.29</b>
	143	0.17	<b>0.07</b>	0.08	1.56	2.27	<b>0.08</b>
	144	3.63	0.59	<b>0.01</b>	<b>0.10</b>	2.31	0.12
	145	<b>0.07</b>	<b>0.07</b>	0.08	<b>0.09</b>	0.44	0.10
	146	0.14	0.55	<b>0.01</b>	0.10	0.79	<b>0.09</b>
3	25	1.65	10.02	<b>0.63</b>	1.64	0.39	<b>0.02</b>
	26	<b>0.06</b>	15.20	5.79	0.85	0.44	<b>0.12</b>
	27	<b>0.27</b>	4.43	7.17	0.99	0.42	<b>0.20</b>
	28	0.09	<b>0.08</b>	7.67	0.54	0.59	<b>0.26</b>
	29	<b>0.13</b>	10.50	2921.03	0.53	<b>0.41</b>	0.61
	30	<b>0.12</b>	0.32	1.37	<b>0.07</b>	0.66	<b>0.07</b>



**Figure 43** the estimated response of problem 3 at different iteration using (a) RBF and (b) GP

## 4.9 Discussion

So far in this chapter we developed a strategy based on a hybrid of quadratic least square regression spline and the ranking strategy, discussed in previous chapters, for expensive noisy black-box function optimizations. In a set of iteration, the proposed strategy breaks down the factor space of the black-box function into a set of small regions (smaller than the factor space) in a way that each of them can be accurately approximated using quadratic functions. It will then identify the most promising sub-region/s, assuming that the real optimum is located there. This step may be followed by conducting new experiments and further breaking down the identified sub-region/s to improve the estimation of the location of the real optimum. Using three numerical examples we examine the performance of the proposed strategy along with two popular global optimizations methods, Gaussian process and radial basis functions, and shows B-ASRSM effectiveness both in terms of number of experiments and accuracy. The proposed method is easy to understand and implement. It is also flexible with using different

implementation criteria. In the remaining part of this chapter we will extend the proposed strategy to work based on space filling design and multiple adaptive regression splines (MARS).

#### **4.10 An Extension of the Proposed Strategy with Space Filling Design and Multiple Adaptive Regression Splines (B-ASRSM2)**

Here we extend the proposed strategy to a more flexible configuration which is specifically useful in modeling problems of high dimension by applying some changes to the original algorithm: (1) Using space filling designs instead of multi-level factorial design, (2) Using multiple adaptive regression splines (MARS) instead of LSRS, and finally (3) Using optimal design for augmentation instead of factorial design. In the following Sections we will first briefly describe the extended algorithm. Next, we examine the performance of the proposed B-ASRSM2 strategy on the same numerical examples as for the B-ASRSM strategy.

##### **4.10.1 Algorithm**

Figure 44 illustrates the algorithm of the proposed B-ASRSM2 strategy based on the few changes mentioned above. As can be seen, the general structure of the algorithms has not been modified however comparing Figure 45 to Figure 41, which is a pictorial view of the implementation of B-ASRSM2 and B-ASRSM, reveals the difference between the two proposed strategies in practice.

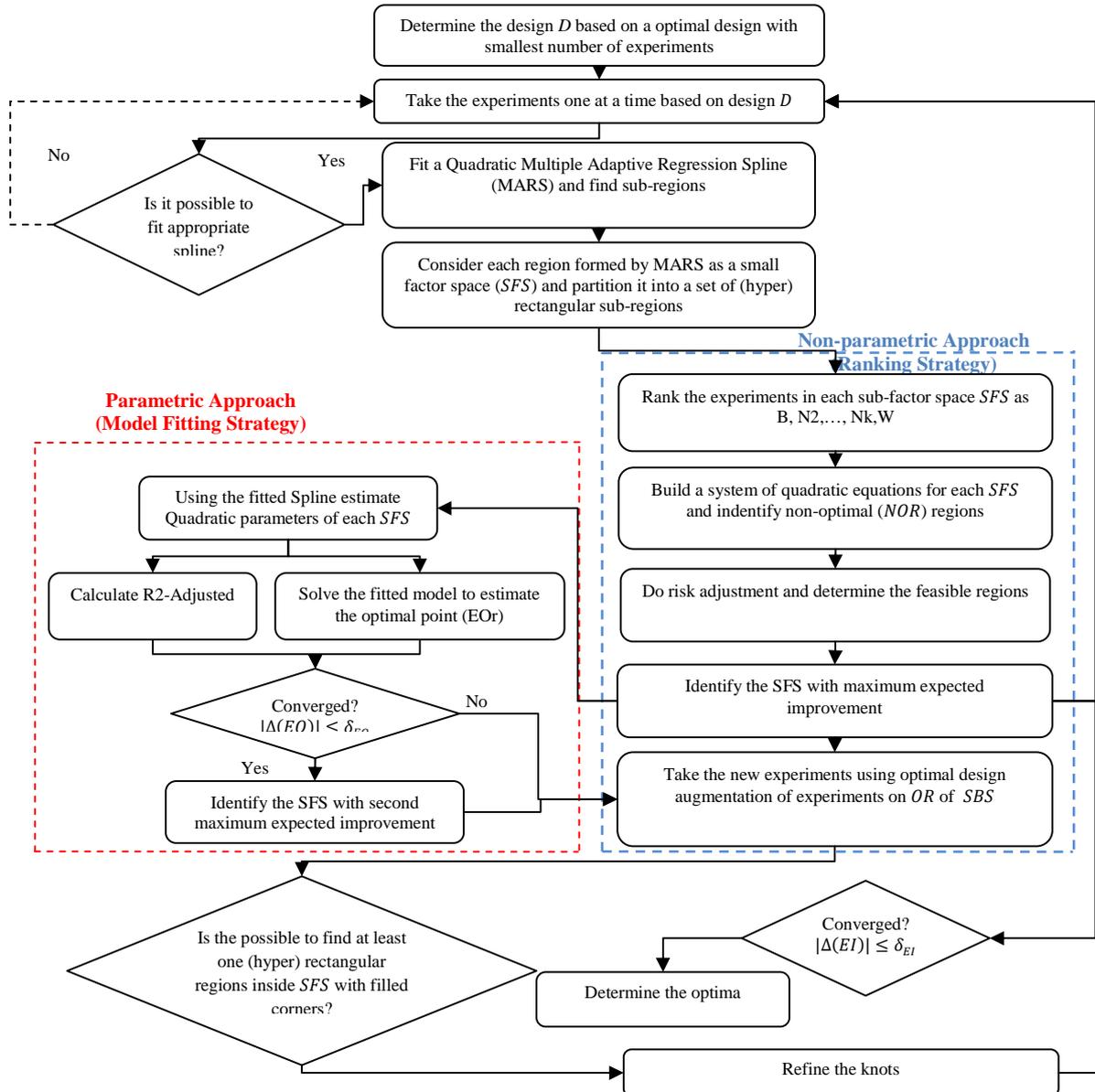
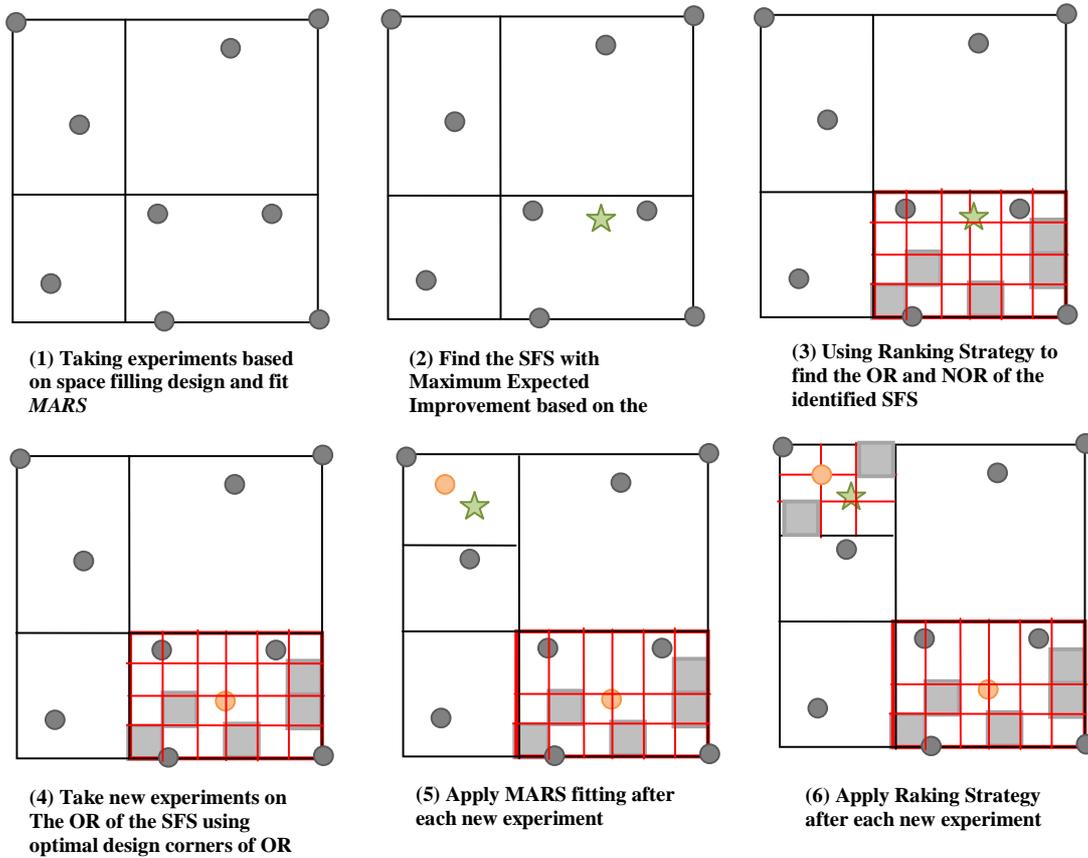


Figure 44 The general scheme of the proposed B-ASRSM(2)



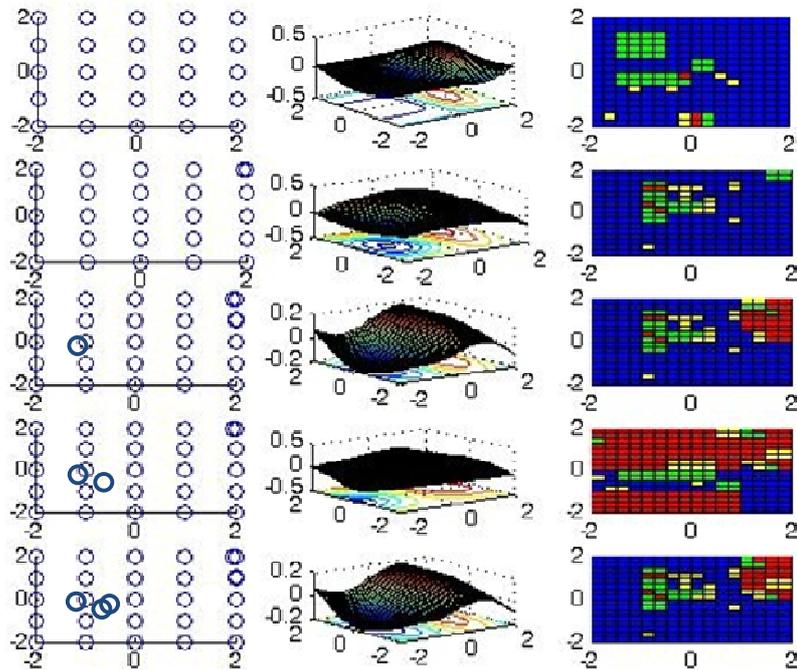
**Figure 45** The graphical representations of the proposed B-ASRSM2 strategy steps

#### 4.10.2 Numerical Examples

Here, we evaluate the proposed B-ASRSM2 on the numerical examples in section 4.8. Table 24 shows the average optimality gap of the consecutive trials of the comparing methods. Figure 45 also illustrates the distribution of the points, the estimated surface as well as the result of ranking strategy at different runs of the B-ASRSM2 for the 3<sup>rd</sup> function. As can be seen from the table, like B-ASRSM, the performance of B-ASRSM2 is acceptable for both optimality gap and Euclidian distance to the real optima, and comparable to RBF and GP.

**Table 24 The average optimality gap and average Euclidean distance to *RO* of the proposed B-ASRSM2 methods along with the comparing methods**

Exp. No.	No Obs.	Optimality gap of Response				Euclidean distance to <i>RO</i>			
		B-ASRSM2	B-ASRSM	RBF	GP	B-ASRSM2	B-ASRSM	RBF	GP
1	25	96.36	8.72	4.99	<b>2.63</b>	4.21	<b>1.75</b>	3.44	6.93
	26	3.70	<b>3.30</b>	11.53	3.80	<b>1.22</b>	1.26	3.80	7.21
	27	<b>2.20</b>	39.36	10.10	<b>2.86</b>	2.61	1.54	<b>0.91</b>	0.92
	28	17.10	12.95	29.56	<b>2.26</b>	1.72	1.34	1.46	<b>0.98</b>
	29	<b>2.17</b>	2.30	6.27	6.27	1.57	<b>0.75</b>	1.86	0.98
	30	0.45	1.29	4.23	<b>0.24</b>	0.93	<b>0.34</b>	2.68	0.70
2	141	0.32	0.69	0.53	<b>0.04</b>	<b>0.65</b>	1.89	2.00	<b>0.33</b>
	142	0.19	0.90	0.50	<b>0.03</b>	0.71	3.74	1.68	<b>0.29</b>
	143	<b>0.07</b>	0.17	<b>0.07</b>	0.08	1.43	1.56	2.27	<b>0.08</b>
	144	0.71	3.63	0.59	<b>0.01</b>	1.40	<b>0.10</b>	2.31	0.12
	145	0.21	<b>0.07</b>	<b>0.07</b>	0.08	0.24	<b>0.09</b>	0.44	0.10
	146	0.12	0.14	0.55	<b>0.01</b>	0.19	0.10	0.79	<b>0.09</b>
3	25	3076.37	1.65	10.02	<b>0.63</b>	0.79	1.64	0.39	<b>0.02</b>
	26	3127.79	<b>0.06</b>	15.20	5.79	0.85	0.85	0.44	<b>0.12</b>
	27	<b>0.23</b>	<b>0.27</b>	4.43	7.17	0.30	0.99	0.42	<b>0.20</b>
	28	<b>0.08</b>	0.09	<b>0.08</b>	7.67	<b>0.16</b>	0.54	0.59	<b>0.26</b>
	29	3195.78	<b>0.13</b>	10.50	2921.03	<b>0.19</b>	0.53	0.41	0.61
	30	0.15	<b>0.12</b>	0.32	1.37	<b>0.07</b>	<b>0.07</b>	0.66	<b>0.07</b>



**Figure 46** The graphical representation of the B-ASRSM2 on 3<sup>rd</sup> response surface; the blue cells show *OR*, greens show *OR* not robust to miss-ranking (in risk adjustment), Yellows show non-robust *NOR* and Reds show robust *NOR*

### 4.10.3 Discussion

Here we extended the strategy developed in the earlier part of this chapter to a more flexible hybrid system of space filling design and multiple adaptive regression splines (MARS) to estimate the minimum of expensive noisy black-box function. The proposed B-ASRSM2 starts with a space filling design and based on the results, partition the factor space into a set of small sub-regions. Assuming quadratic behavior of the function in each small sub-region, B-ASRSM2 fits a MARS to the factor space. Next, it identifies the most promising sub-regions and augments the design with a single point in that sub-region using optimal design and previously discussed ranking strategy. This procedure continues until B-ASRSM2 gets to the vicinity of the real optima. Similar to B-ASRSM, we compare the performance of the B-ASRSM2 along with RBF

and GP on three numerical examples and show its effectiveness both in terms of number of experiments and accuracy.

## CHAPTER 5 CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation we have developed and presented a number of adaptive sequential strategies for response surface optimization (ASRSM). The proposed approaches combine the concept of nonlinear optimization, non-parametric regression and response surface optimization. The proposed strategies uses the information gained from the previous experiments to design the subsequent experiment by simultaneously reducing the region of interest and identifying factor combinations for new experiments. Its major advantage is the experimentation efficiency such that, for a given response target; it identifies the input factor combination (or containing region) in less number of experiments than the classical counterparts. It differs from earlier studies in its optimality, inheritance of results from previous experiments, and its robustness due to experiment ranking based reduction of the region of interest. Through extensive simulated experiments and real-world case studies, we showed that the strategies clearly outperform the classical methods such as BBD and CCD method in terms of both optimality as well as the experimentation efficiency. These results also reveal that the proposed strategies on average perform superior to A-, D-, and V-optimal designs. Further analysis demonstrates that the ASRSM is competitive with popular global optimization methods such as RBF and Gaussian Process. In particular, the performance of ASRSM is found to be very robust with respect to changes in the error variance and convexity of the response model, and more monotonous with additional experiments. For future studies, the proposed methodology will be further extended by adding Bayesian inference capability into it.

## REFERENCES

1. Andere-Rendon, J., Montgomery, D. C., and Rollier, D. A. (1997), "Design of Mixture Experiments Using Bayesian D-Optimality", *Journal of Quality Technology*, 29 (4), 451-463.
2. Anderson, B.S, Moore A.W. and Cohn D., (2000), "A Non-Parametric Approach to Noisy and Costly Optimization", *ICML 2000 Proceedings of the Seventeenth International Conference on Machine Learning*.
3. Atkinson, A.C., Donev, A.N., and Tobias, R.D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press. pp. 511+xvi. ISBN 978-0-19-929660-6.
4. Box, G.E.P. (1957) "Evolutionary Operation: A Method for Increasing Industrial Productivity", *Applied Statistics*, 6, 81-101.
5. Box, G. E. P. (1999), "Statistics as a Catalyst to Learning by Scientific Method: Part II – A Discussion", *Journal of Quality Technology*, 31(1):16-29.
6. Box. G.E.P. and Behnken. D.W. (1960), "Some New Three Level Designs for the Study of Quantitative Variables". *Technometrics*, 2, 455-476.
7. Box, G.E.P. and Draper, N. R. (1969). *Evolutionary Operation: A Method for Increasing Industrial Productivity*. NY: John Wiley and Sons.
8. Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, NY: John Wiley and Sons.
9. Box. G.E.P. and Wilson. K. B. (1951), "On the Experimental Attainment of Optimum Conditions". *Journal of the Royal Statistical Society*, 13, 1-15.
10. Chen V. C. P., Tsui, K.L. , Barton R.R. and Meckesheimer M. (2006), " A review on design, modeling and applications of computer experiments" *IIE Transactions*, 38 (4), 273-291.
11. Czitrom V. (1999), "One-Factor-at-a-Time versus Designed Experiments", *The American Statistician*, 53 (2), 126-131.

12. Daniel, C. (1973), "One-at-a-Time Plans," *Journal of the American Statistical Association*, 68, 353–360.
13. Del Castillo, E. (2007), *Process optimization: A statistical approach*, International series in operations research and management science, 1st ED, NY: Springer
14. Eubank, R. L. (1999), *Nonparametric Regression and Spline Smoothing (Statistics: A Series of Textbooks and Monographs)*, 2<sup>nd</sup> ED, NY: Marcel Dekker.
15. Frey, D. D., Engelhardt, and F., Greitzer, E. M., (2003), "A Role for One Factor at a Time Experimentation in Parameter Design", *Research in Engineering Design*, 14, 65-74.
16. Frey, D. D., and Jugulum, R. (2006), "The Mechanisms by which Adaptive One-Factor-at-a-Time Experimentation Leads to Improvement," *American Society of Mechanical Engineers Journal of Mechanical Design*, 128,1050-1060.
17. Frey, D. D., and Wang, H., (2006), "Adaptive One-Factor-at-a-Time Experimentation and Expected Value of Improvement", *Technometrics*, 48(3), 418-31.
18. Friedman, M., and Savage. L. J. (1947), Planning Experiments Seeking Maxima. *In Techniques of Statistical Analysis*, edited by C. Eisenhart, M. Hastay, and W. A. Wallis, 363-72. NY: McGraw-Hill.
19. Gramacy, R.B. and Lee, H. K. H. (2009). "Adaptive Design and Analysis of Supercomputer Experiments". *Technometrics*, 51 (2), 130-145.
20. Ginsburg H. and Ben-Gal I., (2006), "Designing experiments for robust-optimization problems: the Vs-optimality criterion" *IIE Transactions*, 38 (6), 445-461.
21. Gu, L., (2001), "A Comparison of Polynomial Based Regression Models in Vehicle Safety Analysis", *American Society of Mechanical Engineers Design Engineering Technical Conferences—Design Automation Conference*, A. Diaz, ed., American Society of Mechanical Engineers, New York, Paper No. DETC2001/DAC-21063.

22. Hettmansperger T.P., and McKean, J.W., (1998), *Robust Nonparametric Statistical Methods*, London and New York, N.Y.: Arnold/Wiley.
23. Jones, D. R. (2001), “A Taxonomy of Global Optimization Methods Based on Response Surfaces”, *Journal of Global Optimization*, 21 (4), 345-383.
24. Kiefer, J. (1959). “Optimum Experimental Designs” *Journal of Royal Statistical Society B*, 21, 272-304.
25. Kiefer, J. (1961). “Optimum Designs in Regression Problems” *Annals of Mathematical Statistics*, 32, 298-325.
26. Kiefer, J., and Wolfowitz, J., (1959), “Optimum Designs in Regression Problems”, *Annals of Mathematical Statistics*, 30, 271-294.
27. Mao, H., Zhang, L., Yang, K.H., King, A.I. (2006), “Application of a Finite Element Model of the Brain to Study Traumatic Brain Injury Mechanisms in the Rat”, *Stapp Car Crash Journal*, 50, 583-600.
28. Montgomery, D.C., (2008). *Design and Analysis of Experiments*, 7<sup>th</sup> Ed, NJ: John Wiley& Sons.
29. Moore A.W., Schneider, J. M, Boyan, J, and Lee, M. S., (1998), “Q2: A Memory Based Active Learning Algorithm for Black box Noisy Optimization” *Proceedings of the Fifteenth International Conference on Machine Learning*, 386-394 Morgan Kaufmann.
30. Myers. R. H. (1999). “Response Surface Methodology- Current Status and Future Directions”, *Journal of Quality Technology*, 31, 30-44.
31. Myers. R. H., Khuri. A. I., and Carter. W. H. (1989). “Response Surface Methodology: 1966-1988” *Technometrics*, 31, 137-157.
32. Myers, R., and Montgomery, D. C. (1995), *Response Surface Methodology*, NY: John Wiley& Sons.

33. Myers, R. H. Montgomery, D. C., Vining, C. G. Borror C. M and Kowalski S. M. (2004). "Response Surface Methodology: A Retrospective and Literature Survey", *Journal of Quality Technology*, 36, 53-77.
34. Nelder, J. A. and Mead, R., (1965), "A Simplex Method for Function Minimization" *Computer J.* 308-313.
35. Plastria, F., (1992), "GBSSS: The generalized big square small square method for planar single-facility location", *European Journal of Operational Research.*, 62(2), 163-174.
36. Pukelsheim, F. (2006), "Optimal Design of Experiments. Society for Industrial and Applied Mathematics", (SIAM), Philadelphia, PA.
37. Renaud, J.E., and Gabriele, G.A. (1994), "Approximation in Nonhierarchical System Optimization", *American Institute of Aeronautics and Astronautics Journal*, 32(1), 198-205.
38. Rodriguez, J.F., Perez, V.M., Padmanabhan, D., Renaud, J.E., (2001) "Sequential Approximate Optimization Using Variable Fidelity Response Surface Approximations", *Structural Optimization*, 22(1), 24-34.
39. Rodriguez, J. F., Renaud, J. E., and Watson, L. T. (1998), "Convergence of Trust Region Augmented Lagrangian Methods using Variable Fidelity Approximation Data", *Structural and Multidisciplinary Optimization*, 15, 141-156.
40. Seber G. A. F. Alan J. L. (2003), *Linear Regression Analysis*, NJ :Wiley Interscience.
41. Simpson, T. W., Booker, A. J., Ghosh, D., Giunta, A. A., Koch, P. N., and Yang, R.J. (2004), "Approximation Methods in Multidisciplinary Analysis and Optimization: A Panel Discussion", *Structural and Multidisciplinary Optimization*, 27(5), 302-313.
42. Sobieszczanski-Sobieski, J. (1988), "Optimization by decomposition: A step from hierarchic to nonhierarchic systems", *Second NASA/Air Force Symposium on Recent Advances in Multidisciplinary Analysis and Optimization*, Hampton, VA, NASA CP-3031, Part 1. Also NASA TM-101494.

43. Sobieszczanski-Sobieski J., and Haftka RT, (1997), "Multidisciplinary Aerospace Design Optimization: Survey of Recent Developments", *Structural and Multidisciplinary Optimization* 14(1), 1-23.
44. Spendley, G. R., Hex G. R. and Himsworth, F. R. (1962), "Sequential Application of Simplex Designs in Optimization and Evolutionary Operation" *Technometrics*, 4, 441-461.
45. Stander, N. (2001), "The Successive Response Surface Method Applied to Sheet-Metal Forming", *Proceedings of the First MIT Conference on Computational Fluid and Solid Mechanics*, Boston, June 12-14, 2001. Elsevier Science Ltd., Oxford.
46. Santner, T.J., Williams, B. J., Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, NY: Springer.
47. Tan P.N., Steinbach M., and Kumar V., (2006), *Introduction to Datamining*, NY: Addison-Wesley.
48. Walters, F. S., Parker Jr., L. R., Morgan, S. L., and Deming, S. N. (1991), *Sequential Simplex Optimization for Quality & Productivity in Research, Development, and Manufacturing*, FL: CRC Press.
49. Wang, G., (2003), "Adaptive Response Surface Method Using Inherited Latin Hypercube Designs," American Society of Mechanical Engineers Journal of Mechanical Design, 125(2), 210-220.
50. Wang, G., Dong, Z., and Aitchison, P., (2001), "Adaptive Response Surface Method - A Global Optimization Scheme for Computation-intensive Design Problems," Journal of Engineering Optimization, 33(6), 707-734.

**ABSTRACT****SELF LEARNING STRATEGIES FOR EXPERIMENTAL DESIGN AND  
RESPONSE SURFACE OPTIMIZATION**

by

**ADEL ALAEDDINI****August 2011****Advisor:** Dr. Kai Yang**Major:** Industrial Engineering**Degree:** Doctor of Philosophy

Most preset RSM designs offer ease of implementation and good performance over a wide range of process and design optimization applications. These designs often lack the ability to adapt the design based on the characteristics of application and experimental space so as to reduce the number of experiments necessary. Hence, they are not cost effective for applications where the cost of experimentation is high or when the experimentation resources are limited. In this dissertation, we present a number of self-learning strategies for optimization of different types of response surfaces for industrial experiments with noise, high experimentation cost, and requiring high design optimization performance. The proposed approach is a sequential adaptive experimentation approach which combines concepts from nonlinear optimization, non-parametric regression, statistical analysis, and response surface optimization. The proposed strategies uses the information gained from the previous experiments to design the subsequent experiment by simultaneously reducing the region of interest and identifying factor combinations

for new experiments. Its major advantage is the experimentation efficiency such that, for a given response target, it identifies the input factor combination (or containing region) in less number of experiments than the classical designs. Through extensive simulated experiments and real-world case studies, we show that the proposed ASRSM method clearly outperforms the classical CCD and BBD methods, works superior to optimal A- D- and V- optimal designs on average and compares favorably with global optimizations methods including Gaussian Process and RBF.

## AUTOBIOGRAPHICAL STATEMENT

Adel Alaeddini is a Ph.D. candidate of Industrial Engineering at Wayne State University. He used to work as a lecturer in the Departments of Industrial Engineering, and Information Technology of Azad University, Tehran, Iran. His main areas of research are Statistical Learning, Global Optimization, Response Surface Methodology and Healthcare Operations Management. He has two nominations for the best paper award in top conferences, three co-authored grants, more than ten published journal papers and several refereed conference papers. Based on his Ph.D. dissertation work, in addition to five working papers, he has published/presented the following articles:

### Awards

- i. Best Paper Award of Industrial Engineering Research Conference (IERC 2010), Cancun, Mexico (2010).

### Journal Publications

- i. A. Alaeddini, K. Yang, (2009), "Adaptive Sequential Experiment Methodology for Response Surface Optimization", *International Journal Quality Technology and Engineering*, 1 20-61.
- ii. A. Alaeddini, K. Yang, A. Murat, (2011), "ASRSM: A Sequential Experimental Design for Response Surface Optimization", *Quality and Reliability Engineering International*, To Appear.

### Conference Proceedings/ Presentations

- i. A. Alaeddini, S. Shirinkam, K. Yang, (2011), "An Adaptive Sequential Bayesian Methodology for Process Optimization", IERC 2010, Reno, Nevada, USA.
- ii. A. Alaeddini, Kai Yang, Alper Murat, (2010), "An Adaptive Sequential Methodology for n-Dimensional Quadratic Response Surface Optimization", INFORMS 2010, Texas, USA.
- iii. A. Alaeddini, K. Yang, A. Muart, (2010), "Adaptive Sequential Experimentation Methodology for Response Surface Optimization", IERC 2010, Cancun, Mexico.