Chemometrics and Intelligent Laboratory Systems 113 (2012) 10–18

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Capability of high intrinsic quality Space Filling Design for global sensitivity analysis and metamodelling of interference optical systems

O. Vasseur [*], A. Azarian, V. Jolivet, P. Bourdon

Onera - The French Aerospace Lab, F-97761 Palaiseau, France

ARTICLE INFO

Article history:
Received 4 February 2011
Received in revised form 9 May 2011
Accepted 10 May 2011
Available online 16 May 2011

Keywords:
Computer experiments
Space-Filling Designs
Minimum Spanning Tree
Minimal Spanning Tree
Thin film coatings
Optical filter
Sensitivity analysis
Complex systems

ABSTRACT

In the field of computer experiments, when the relations between the outputs and the inputs of the computer code are complex, the Space Filling Designs (SFD) are used to study the outputs all over the domain or to build metamodels. It is necessary that the points of these designs are distributed as uniformly as possible in the domain of interest. The methodology of Minimum Spanning Tree (MST) was proposed to evaluate the quality of the distribution of the points of an experimental design in a multidimensional space. We introduce here the results obtained in high dimensional case (dimension higher than 20) and we point out the bad quality of the classical designs or the designs based on low discrepancy sequences but mainly the advantages of WSP designs. Empirical results point out the need to qualify SFD in the original space due to the insufficiency of the conclusions obtained using projections onto 2D subspaces. We define the intrinsic quality of SFD as the characteristic of the distribution of the points in the space. Likewise, the extrinsic quality of SFD describes the performances of the results obtained by the use of SFD on an application case. The results obtained with different SFD in the case of interference optical system sensitivity analysis are presented. Finally, the intrinsic quality of SFD is assessed by MST criterion and is proved to be in accordance with the results of the interference optical system sensitivity analysis. So, this study can be considered as an empirical step to connect intrinsic quality and extrinsic quality of SFD designs.

1. Introduction

In the field of computer experiments, models or simulations of complex phenomena is getting more and more realistic. Although computer power has highly increased over the last years, models or simulations including numerous parameters, are still very time consuming. The use of numerical designs is an effective method to explore codes with a high dimensional space of parameters and to build metamodels, which approximate the code and the relationship between inputs and outputs of this code. Due to the computation time of the original model or simulator, the number of runs for building the metamodel must be limited and the set of computer experiments must provide the maximum of information of the code. In general, when the relationship between the response and inputs of the code is not explicit, the purpose is to use designs that uniformly spread the points all over the domain. These designs are called Space-Filling Designs (SFD).

In the case of the dimension of the input parameter space less than 10, Franco et al.[9] pointed out that classical criteria are insufficient to conclude about the uniform distribution of points and proposed to use Minimum Spanning Tree (MST) (Beardwood et al.[4]) criterion, which is more informative than the classical methods. Indeed, Dussert et al. [6] showed that the construction of a MST on a set of points in dimension 2 allows to qualify the distribution of the points (ordered, random, cluster…) and Wallet et al.[39] compared the five most used methods of topographical analysis (nearest neighbour distribution, radial distribution, Voronoï paving, quadratic count and minimal spanning tree graph) to conclude that the method using the MST offers the best discrimination power and stability. By the use of MST criterion, Franco et al.[9] pointed out that the designs which uniformly spread the points all over the domain belong to the quasi-regular distributions. So, we define the intrinsic quality of SFD as the characteristic of the distribution of the points in the space and a good intrinsic quality corresponds to a quasi-regular point distribution.

We firstly summarize the MST methodology and we expose the first intrinsic qualification of SFD designs in the case of high dimension space (dim$\geq$20). Using additional results obtained by projections onto 2-D subspaces, we point out the complementarities of MST methodology and the Radial Scanning Statistic (RSS) (Roustant et al.[27]) on a first hand and the need to analyse the SFD designs in the original space to qualify them precisely on the second hand.

In a second part, we apply the different designs which have been sorted by the MST criterion to the study of sensitivity analyses (Saltelli et al.[28]) and metamodels of interference optical systems. By this way, we define the extrinsic quality of SFD which describes the performances of the results obtained by the use of SFD on an application case. These empirical results on applications cases confirm the intrinsic quality of the designs determined by the MST criterion.

---

[*] Corresponding author. Tel.: +33 0 1 80 38 63 53; fax: +33 0 1 80 38 63 45.
E-mail address: Olivier.Vasseur@onera.fr (O. Vasseur).

0169-7439/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2011.05.008

2. Intrinsic quality analysis of high dimensional numerical designs “Space Filling Designs” by minimal spanning tree

2.1. MST methodology

Consider a set of N points which are randomly distributed in a d-dimensional region with a volume of V. An edge-weighted linear graph $G=(X,E)$ is composed of a set of points $X=\{x_1, x_2, \ldots x_N\}$ called nodes and a set of node pairs $E=\{(x_i, x_j)\}$ called edges, with a number called weight (the Euclidean distance in our case) assigned to each edge. A graph is connected if there is a path connecting every pair of nodes and a tree is a connected graph without closed loops. A Minimal Spanning Tree (MST) is a tree which contains all the nodes with a minimal sum of the edge weights (Zhan [40]). It can be noticed that for a set of N points (corresponding to a number of edges of $N-1$) which are randomly distributed d-dimensional region, the total length of the MST is asymptotically given by $L_{G_\infty}=\alpha_d(VN^{d-1})^{1/d}$ where α_d depends on the problem solved (Beardwood et al.[4]). So the mean length of an edge is given by:

$$M=\alpha_d\frac{(VN)^{1/2}}{N-1} \tag{1}$$

In our study, the points of all designs are generated in the domain $[0;1]^d$ and the normalized values of m and σ of the MST (mean m and the standard deviation σ of the edge length) constructed from a given set of data in our case are obtained by dividing the original lengths by the expression following the normalization process proposed by Hoffman and Jain [12]:

$$\frac{N^{(d-1)/d}}{N-1} \tag{2}$$

Algorithms, as those by Kruskal [17] or Prim [26], allow calculating the MST. In Prim's algorithm, the MST is grown from a single node by adding the closest node to current tree at each stage along with the edge corresponding to that closest distance. Depending on the starting point there may be more than one MST for a given set of points, but all of the MST's have the same length-edge histogram (Zahn [40], Dussert et al.[6]). The normalized values of the mean m and the standard deviation σ of the edge length can be used to characterize the distribution of points (ordered, random, cluster…) (Dussert et al.[6]) as shown on the Fig. 1. In the field of topographical analysis, this method presents the advantages of a high discrimination power and stability to characterize spatial point patterns (Wallet et al.[39]).

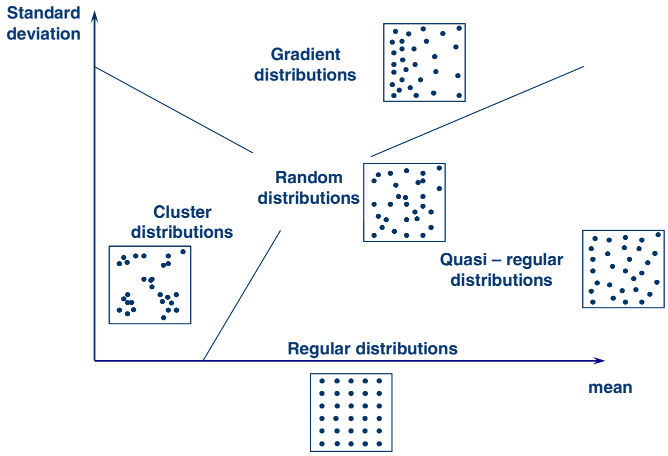

Fig. 1. Points distributions in the (m, σ) plane.

From the MST, the mean m and the standard deviation σ of the edges length may be used as characteristics for the corresponding distribution of the SFD points. On this graphical representation, it is possible to distinguish different areas and mainly the quasi-regular area which is the best area for Space Filling Designs due to the best compromise between a high average length of MST branches to fill the space and a small standard deviation to obtain a sufficient regularity: for example random designs have points too closed and grid designs ($\sigma=0$) are too regular. Moreover with perfect ordered structure ($\sigma=0$), the points are not evenly spread across the projection of the experimental space onto all subspaces (Franco et al.[9]).

2.2. Qualification of SFD in high dimensional spaces

The quality of a large part of numerical designs in dimensions less than 10 was presented by Franco et al.[9]: in this case, all the designs are located in the quasi-periodic area when the space dimension is lower than 5 but only WSP designs offer a good robustness. We present here the results obtained in higher dimension case (20-D).

The experimental designs studied here are:

- Latin Hypercube Design (LHD) (Mc Kay et al. [22])
- WSP design (Sergent et al.[30]). The aim of this algorithm is to find, among a set of possible points, a subset of points situated according to a uniform disposition. The points are selected such as they must be at a minimal distance of every point already included in the design and as near as possible to the center of gravity of the included points.
- Low Discrepancy Sequences (LDS): Faure [8], Halton [11], Sobol [32]
- LDS wrap-around design (Marrel [23])

The LHD and LDS designs are built with the lhs, fOptions and DiceDesign packages in R-software [13,14].

All the designs are compared with a random distribution and with cluster arrangements built using the Neyman and Scott [24] process. In dimension 20 the LHDs, LDS wrap-around sequences and Sobol's design are very close to the random area (cf. Fig. 2). The quality of LDS wrap around sequences does not differ than the other LHD and LDS designs so we will neither study them nor apply them on our application cases.

The other classical LDS (except Sobol) have a particular representation in the (m, σ) plane. The histogram analysis of the edge lengths of the MST explains the high values of the standard deviation. For example, the Faure sequence produces a MST whose values of edge lengths are distributed on two values: ~0.35 and ~1.38. Criticisms relating to the LDS designs often underlined the specific characteristics obtained by projections of the points on subspaces (Tan [34]). The insufficiencies of these LDS designs are here highlighted independently of any projection, the MST being built directly on the points of the original space.

The classical space-filling designs that had good properties (quasi-regular distribution) in low dimension are no more in the quasi-regular area, and are, in the best case, closed to the random distribution area as Sobol design.

Only the WSP designs have simultaneously a higher average length of branch and a lower standard deviation than those of the random designs and correspond to the characteristics of the quasi-regular distributions needed to build high quality SFD.

As our application study corresponds to 18-D and 29-D spaces, we present on Fig. 3 results obtained in both dimensions for few designs.

These results are similar to those presented in 20-D and only WSP designs are located in the quasi-regular area in the (m, σ) plane. Thus, for dimensions of space higher than 20, only the WSP designs seem to have an intrinsic good quality to explore the space due to an important edge length between points without perfect order. In conclusion, these results highlight that only the WSP designs present good properties to be used as SFD.