AStA Adv Stat Anal (2010) 94: 353–366
DOI 10.1007/s10182-010-0142-1

ORIGINAL PAPER

Comparing and generating Latin Hypercube designs in Kriging models

Giovanni Pistone · Grazia Vicario

Received: 31 August 2009 / Accepted: 26 July 2010

Abstract In Computer Experiments (CE), a careful selection of the design points is essential for predicting the system response at untried points, based on the values observed at tried points. In physical experiments, the protocol is based on Design of Experiments, a methodology whose basic principles are questioned in CE. When the responses of a CE are modeled as jointly Gaussian random variables with their covariance depending on the distance between points, the use of the so called space-filling designs (random designs, stratified designs and Latin Hypercube designs) is a common choice, because it is expected that the nearer the untried point is to the design points, the better is the prediction. In this paper we focus on the class of Latin Hypercube (LH) designs. The behavior of various LH designs is examined according to the Gaussian assumption with exponential correlation, in order to minimize the total prediction error at the points of a regular lattice. In such a special case, the problem is reduced to an algebraic statistical model, which is solved using both symbolic algebraic software and statistical software. We provide closed-form computation of the variance of the Gaussian linear predictor as a function of the design, in order to make a comparison between LH designs. In principle, the method applies to any number of factors and any number of levels, and also to classes of designs other than LHs. In our current implementation, the applicability is limited by the high computational complexity of the algorithms involved.

The first author acknowledges the support of DIMAT Politecnico di Torino and of SAMSI, NC, he visited Jan-Mar 2009, where the computational part was developed and some issues were discussed in a working group of the 2008-09 SAMSI Program on Algebraic Methods in Systems Biology and Statistics. The second Author acknowledges the support of DISPEA Politecnico di Torino.

G. Pistone (✉)
Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy
e-mail: giovanni.pistone@gmail.com

G. Vicario
DIMAT Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino Italy
e-mail: grazia.vicario@polito.it

Keywords Computer experiments · Latin hypercube · Mean square prediction error · Gaussian linear prediction

1 Introduction

The official start of Computer Experiments (CE) is the paper by McKay et al. (1979), while the contribution by Sacks et al. (1989b) marked a new step by introducing model-based methods, see the review in this Issue by Levy (2010). A compelling reason for using CE (in a single approach or combined with physical experiments) comes from the fact that physical experimentation may be, in a number of circumstances, expensive or even unapproachable. On the contrary, the use of numerical experiments in product/process development phase is relatively inexpensive and, because of that, has become straightforward. The general availability of comprehensive computing facilities and the recent progresses in software development make numerical simulation of complex systems an attractive alternative to the execution of the expensive and time consuming physical experiments. Standard modern references are Sasena (2002), Santner et al. (2003), Fang et al. (2006).

In this context, a careful selection of the design points or training points is essential for predicting how the unobserved responses depend on the observed ones. In physical experimentation, the researcher is asked to comply to a well set protocol in order to achieve correct inferences. Such a protocol is the Design of Experiments (DoE) methodology, which is an helpful tool in carrying out the mentioned objectives. The design of a CE, when it is used as a surrogate for the physical one, differs in several aspects from designing a physical experiment and the applicability of basic principles of DoE is questioned in CE. The selection of an experimental design in CE is a crucial part of producing an efficient and informative model and cannot be done by merely importing the concepts developed for physical experiments. This means providing efficient strategies for sampling the input space in order to get accurate predictions for untried inputs.

As suggested by the pioneers of the model-based CE, the output can be predicted by assuming joint Gaussian distribution of the responses with a covariance depending on a properly defined distance between the locations, e.g. the Euclidean distance in the original Kriging model, see Krige (1951) and Cressie (1986). The underlying principle is that the nearer an untried point is to the design points, the better the prediction. Based on this view, a good design strategy is to uniformly spread the points across the experimental region. That invites the use of the so called space-filling designs. Random designs, Stratified designs and Latin Hypercube designs are common choices, see e.g. McKay et al. (1979), Fang et al. (2006). The first two designs are not fully satisfactory because they are not marginally space filling, i.e. in individual directions, and, moreover, are quite unsatisfactory for global space filling, especially for a small number of design points. Even if the Latin Hypercube (LH) designs are not very satisfactory for space filling, they are satisfactory for space filling in individual directions, and this statement is valid for any number of design points. For a formal treatment of the subject see Welch et al. (1992), Park (1994), Fang et al. (2000), Butler (2001).

The content of the paper is as follows. We focus on the class of LH designs among the different space-filling ones. We want to investigate which LH designs have the best prediction features.

In Sect. 2, the Gaussian field is assumed to have distance-dependent covariance (stationarity) and is defined on a subset of a regular lattice, i.e. the Cartesian product of uniform one-dimensional lattices. This case corresponds to specific applications, see e.g. Pistone and Vicario (2009), and it is specially fit for the methodology we use. The commonly used Euclidean distance is not natural when considering a regular lattice; therefore we switch to the Manhattan distance, as other authors have suggested, see Santner et al. (2003, p. 138).

In Sect. 3, we assess the behavior of different LH designs by computing in closed form the Mean Square Prediction Error (MSPE) of the linear predictor. It should be noticed that the issue of identifiability, which is of the highest importance in standard DoE, is not relevant here.

Section 4 contains the main results of the paper. Specific study of LH design is performed through an example. We present a step-by-step discussion of LH designs with two factors with four levels each, including a full example of the use of the algebraic software.

In performing the aforementioned comparisons, there are two tricky computational problems: First, the computation of the variance-covariance matrix of the design points and, second, the computation of the closed-form expression of the predictors variance. For the former problem, a solution is presented in a particular case of the exponential correlation function, which is actually one of the most commonly used by CE practitioners. For the latter, the main issue is the computation of the covariances which are the rational functions with respect to a parameter of interest. In order to compute the predictor variances in closed form, we resort to the use of symbolic algebraic software such as CoCoA (Computations in Commutative Algebra), a freely available system for symbolic exact multivariate polynomial computation, see CoCoATeam. Other computations related with the exponential model for covariances are done with the software R, (http://www.R-project.org/). The final result is an example of general methodology to analyze the MSPE efficiency for special training sets for regular lattices and covariances which are exponentials of the Manhattan distance. All LH designs are classified according to the algebraic form of the predictor variance and the algebraic form of the determinant of the correlation matrix (entropy criteria). It is remarkable to observe that, at least in the examples we discuss, the two classifications coincide. This and other conclusions are discussed in Sect. 5.

2 Correlation function on a lattice

Let $Y(\mathbf{x})$ be a zero mean and covariance stationary Gaussian random field over a design space $\mathcal{X}_d \subset \mathbb{R}^d$, i.e.

$$\mathrm{E}\big(Y(\mathbf{x})\big) = 0,$$

$$\mathrm{Cov}\big(Y(\mathbf{x}), Y(\mathbf{x}+\mathbf{h})\big) = \sigma_Y^2 R(\mathbf{h}; \boldsymbol{\psi}),$$

where σ_Y^2 is the field variance and R is its Stationary Correlation Function (SCF), depending only on the displacement vector $\mathbf{h}$ between any pair of points in $\mathcal{X}_d$ and on a vector parameter $\boldsymbol{\psi}$. A popular choice for the SCF is the power exponential family

$$R(\mathbf{h};\boldsymbol{\psi}) = \prod_{s=1}^{d} \exp\bigl(-\theta_s |h_s|^{p_s}\bigr) = \exp\left(-\sum_{s=1}^{d} \theta_s |h_s|^{p_s}\right), \quad 0 < p_s \le 2, \tag{1}$$

where $\mathbf{h} = (h_1, \ldots, h_d)$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ are positive scale parameters, $\mathbf{p} = (p_1, \ldots, p_d)$ is a vector smoothing parameter, and $\boldsymbol{\psi} = (\boldsymbol{\theta}, \mathbf{p})$. The conditions on $\boldsymbol{\theta}$ and $\mathbf{p}$ are necessary and sufficient for (1) to be positive definite and, therefore, for the existence of a stationary Gaussian field with an SCF of that form, see (Berg et al. 1984). Krige's original idea of a positive correlation between the outputs that decreases when $\sum_{s=1}^{d} \theta_s |h_s|^{p_s}$ increases is true for the given SCF.

Our algebraic methodology is best described if we assume equal scale parameters $\theta_s = \theta$, $s = 1, \ldots, d$. Moreover, it requires an integer valued $\mathbf{p}$ and the Gaussian field to be defined on a regular square lattice $\mathcal{X}_d = \{1, \ldots, l\}^d$. As the Euclidean distance would not really be adapted to the lattice case, we restrict ourselves to the case of a common smoothing parameter $p = p_s = 1$, so that the SCF depends on the Manhattan distance, $\|x - y\|_1 = \sum_{j=1}^{d} |x_j - y_j|$. The SCF we use has the form

$$R(\mathbf{h};\theta) = \prod_{s=1}^{d} \exp\bigl(-\theta |h_s|\bigr) = \exp\left(-\theta \sum_{s=1}^{d} |h_s|\right) = \exp\bigl(-\theta \|\mathbf{h}\|_1\bigr). \tag{2}$$

Let us first consider the univariate case $d = 1$. The distance function is $d_1(i, j) = |i - j|$, and the covariance function in algebraic form is

$$R_1(h;\theta) = \exp\bigl(-\theta d_1(i, j)\bigr) = t^{|i-j|}, \quad h = i - j,\ i, j = 1, \ldots, l,$$

where $t = \exp(-\theta) > 0$.

The matrix of the distances is

$$\mathsf{D}_1 = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ l \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \cdots & l \end{array} \\ \begin{bmatrix} 0 & 1 & \cdots & l-1 \\ 1 & 0 & \cdots & l-2 \\ \vdots & \vdots & \vdots & \vdots \\ l-1 & l-2 & \cdots & 0 \end{bmatrix} \end{array}$$

and the covariance matrix is

$$\Gamma_1 = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ l \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \cdots & l \end{array} \\ \begin{bmatrix} 1 & t & \cdots & t^{l-1} \\ t & 1 & \cdots & t^{l-2} \\ \vdots & \vdots & \vdots & \vdots \\ t^{l-1} & t^{l-2} & \cdots & 1 \end{bmatrix} \end{array} \tag{3}$$

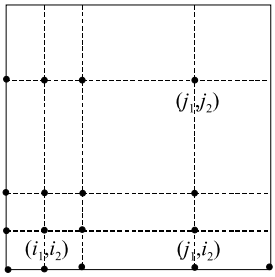


Fig. 1 A representation of the bivariate rectangular lattice $\mathfrak{X}_2 = \{1, 2, \ldots, l\}^2$

If $d = 2$, the distance function is (see Fig. 1)

$$d_2\big((i_1, i_2), (j_1, j_2)\big) = |j_1 - i_1| + |j_2 - i_2| = d_1(i_1, j_1) + d_1(i_2, j_2)$$

and the covariance function for $\mathbf{h} = \mathbf{i} - \mathbf{j} = (i_1 - j_1, i_2 - j_2)$ is:

$$R_2(\mathbf{h}; \theta) = \exp\big(-\theta d_2\big((i_1, i_2), (j_1, j_2)\big)\big) = t^{|i_1 - j_1|} t^{|i_2 - j_2|}$$

It follows that the elements of the distance and covariance matrices, ordering rows and columns in lexicographic order, are given by:

$$D_2\big((i_1, i_2), (j_1, j_2)\big) = D_1(i_1, j_1) + D_1(i_2, j_2), \tag{4}$$

$$\Gamma_2\big((i_1, i_2), (j_1, j_2)\big) = \Gamma_1(i_1, j_1) \times \Gamma_1(i_2, j_2). \tag{5}$$

We write the matrix operations defined in (4) and (5) as $\mathsf{D}_2 = \mathsf{D}_1 \oplus \mathsf{D}_1$ an $\Gamma_2 = \Gamma_1 \otimes \Gamma_1$, respectively. Note that D_2 and Γ_2 are $l^2 \times l^2$ matrices and $\otimes$ denotes the Kronecker product of matrices, see e.g. Ortega (1987).

In generic dimension $d \geq 2$, we have the induction formulas

$$\mathsf{D}_d = \mathsf{D}_{d-1} \oplus \mathsf{D}_1, \qquad \Gamma_d = \Gamma_{d-1} \otimes \Gamma_1. \tag{6}$$

The special algebraic structure of Γ_d allows us an algebraic treatment of the prediction problem. In particular, the computation of variances and covariances in closed form may be done by a symbolic algebraic software, so that we can compare the performance of each LH design on the basis of closed-form algebraic expressions.

Indeed, the resulting statistical model is special. In fact, one can show by recursion that $\det \Gamma_1 = (1 - t^2)^{l-1}$ and that Γ_1^{-1} is tri-diagonal, e.g. for $l = 4$

$$\Gamma_1^{-1} = \big(1 - t^2\big)^{-1} \begin{bmatrix} 1 & -t & 0 & 0 \\ -t & 1 + t^2 & -t & 0 \\ 0 & -t & 1 + t^2 & -t \\ 0 & 0 & -t & 1 \end{bmatrix}$$

As Γ_1^{-1} is tri-diagonal, the 1-dimensional case is a Gaussian Markov Chain, or more precisely a discrete Ornstein–Uhlembeck process, (Santner et al. 2003, p. 36). The d-dimensional case is a Kronecker product of Markov chains, therefore it is a Gaussian graphical model. We do not discuss further this interesting point and refer to Lauritzen (1996).

It has been observed by one of the anonymous referees that a Gaussian process with SCF (2) exhibits peculiar symmetries; specifically, the joint distribution over the sites $\mathbf{x} \in \mathcal{X}_d$ depends only on the distances $\|\mathbf{x} - \mathbf{y}\|_1$, $\mathbf{x}, \mathbf{y} \in \mathcal{X}_d$, so that the distribution is invariant under any transformation of $\mathcal{X}_d$ for which the Manhattan distance between points is an invariant. For example, the classification of subsets with four points of a 4×4 grid we shall obtain in Table 3 is based on the action of the dihedral group of the square. The method we present here does not require the preliminary study of such symmetries.

3 Predicting the output on lattice points

Sacks et al. (1989a, 1989b) suggested that the joint use of Kriging model as a metamodel together with Latin Hypercube designs as a training set is the option of choice in CE where no specific model is imposed by the application itself. The model considers the response $Y(\mathbf{x})$, for $\mathbf{x} \in \mathcal{X}_d \subset \mathbb{R}^d$, to be a realization of a Gaussian random field of the form

$$Y(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}),$$

where $\mathbf{f}'(\mathbf{x}) = [f_1(\mathbf{x}) \cdots f_m(\mathbf{x})]$ is a set of specified trend functions, $\boldsymbol{\beta}$ is a vector of parameters, and $Z(\mathbf{x})$ is a Gaussian random field with zero mean and SCF over $\mathcal{X}_d$. According to the Krige's principle, the observed information on the random vector of the field variables $\mathbf{Y}^n = [Y(\mathbf{x}_1) \cdots Y(\mathbf{x}_n)]'$ at the training data set $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is used to predict the unobserved output at $\mathbf{x}_0$. The underlying hypothesis, see (Santner et al. 2003), which is consistent with the covariance function in (1), assumes that the joint distribution of $Y(\mathbf{x}_0)$ and $\mathbf{Y}^n$ is

$$\begin{bmatrix} Y(\mathbf{x}_0) \\ \mathbf{Y}^n \end{bmatrix} \sim \mathrm{N}\left[\begin{bmatrix} \mathbf{f}'(\mathbf{x}_0) \\ \mathsf{F}^n \end{bmatrix} \boldsymbol{\beta}, \sigma_Z^2 \begin{bmatrix} 1 & \mathbf{r}_0' \\ \mathbf{r}_0 & \mathsf{R}^n \end{bmatrix}\right],$$

were F^n is the $n \times m$ matrix with entries $f_j(\mathbf{x}_i)$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, $\mathbf{r}_0' = [R(\mathbf{x}_0 - \mathbf{x}_1) \cdots R(\mathbf{x}_0 - \mathbf{x}_n)$ is the correlation vector, and R^n is the $n \times n$ correlation matrix whose (i, j)-element is $R(\mathbf{x}_i - \mathbf{x}_j)$, $i, j = 1, \ldots, n$.

We consider in this paper the *ordinary* Kriging model, i.e. we assume $\mathbf{f}'(\mathbf{x})\boldsymbol{\beta} = \beta$, so that $Y(\mathbf{x}) = \beta + Z(\mathbf{x})$, where the trend is constant, and we assume the field variance to be the unit, $\sigma_Z^2 = 1$, so that our model reduces to

$$\begin{bmatrix} Y(\mathbf{x}_0) \\ \mathbf{Y}^n \end{bmatrix} \sim \mathrm{N}\left[\beta \mathbf{1}, \begin{bmatrix} 1 & \mathbf{r}_0' \\ \mathbf{r}_0 & \mathsf{R}^n \end{bmatrix}\right].$$

We use the Linear Unbiased Predictor (LUP). When the covariance is known, the Kriging methodology uses a linear spatial interpolation, i.e. the random variable $Y(\mathbf{x}_0)$ is predicted by an affine combination of the observed random variables,

$$\hat{Y}(\mathbf{x}_0) = a_0 + \sum_{i=1}^{n} a_i Y(\mathbf{x}_i). \tag{7}$$

The predictor $\hat{Y}(\mathbf{x}_0)$ is unbiased if and only if, for all β, $\beta = \mathrm{E}(\hat{Y}(\mathbf{x}_0)) = a_0 + \sum_{i=1}^{n} a_i \beta$, i.e. $a_0 = 0$ and $\sum_{i=1}^{n} a_i = 1$. The predictor $\hat{Y}(\mathbf{x}_0) = \sum_{i=1}^{n} a_i Y_i$, $\sum_{i=1}^{n} a_i = 1$ is the Best LUP (BLUP) if the Mean Squared Prediction Error $\mathrm{MSPE}[\hat{Y}(\mathbf{x}_0)] = \mathrm{E}((\hat{Y}(\mathbf{x}_0) - Y(\mathbf{x}_0))^2)$ is minimized. If β is known, the BLUP is the conditional expectation of $Y(\mathbf{x}_0)$ given $\mathbf{Y}^n = [Y(\mathbf{x}_1) \cdots Y(\mathbf{x}_n)]'$,

$$\hat{Y}(\mathbf{x}_0) = \beta + \mathbf{r}_0' \big(\mathsf{R}^n\big)^{-1} \big(\mathbf{Y}^n - \beta \mathbf{1}\big) \tag{8}$$

and the value of the MSPE is $1 - \mathbf{r}_0' (\mathsf{R}^n)^{-1} \mathbf{r}_0$.

If β is to be estimated, the BLUP is given by (8) with β replaced by its generalized least squares estimator

$$\hat{\beta} = \big(\mathbf{1}' \big(\mathsf{R}^n\big)^{-1} \mathbf{1}\big)^{-1} \mathbf{1}' \big(\mathsf{R}^n\big)^{-1} \mathbf{Y}^n.$$

In such a case the MSPE, usually called Kriging variance, is larger because of an additional uncertainty component:

$$\begin{aligned} \mathrm{MSPE}[\hat{Y}(\mathbf{x}_0)] = {} & \big(1 - \mathbf{r}_0' \big(\mathsf{R}^n\big)^{-1} \mathbf{r}_0\big) \\ & + \big(1 - \mathbf{1}' \big(\mathsf{R}^n\big)^{-1} \mathbf{r}_0\big)' \big(\mathbf{1}' \big(\mathsf{R}^n\big)^{-1} \mathbf{1}\big)^{-1} \big(1 - \mathbf{1}' \big(\mathsf{R}^n\big)^{-1} \mathbf{r}_0\big). \end{aligned}$$

In the next section we shall compute in closed form the MSPE as an algebraic function of the parameter $t = \exp(-\theta)$. Since the computational complexity of the symbolic algorithms is high, any reduction of the number of operation—however small—is of interest. To this aim, we derive now a more compact form of the BLUP (7) and of its MSPE, to be used in the following section.

Consider the linear transformation that maps the vector

$$\big[Y(\mathbf{x}_0)\ Y(\mathbf{x}_1) \cdots Y(\mathbf{x}_n)\big]'$$

to the vector $\tilde{\mathbf{Y}}$ whose entries are $\tilde{Y}_1 = Y(\mathbf{x}_0) - Y(\mathbf{x}_1)$, $\tilde{Y}_i = Y(\mathbf{x}_i) - Y(\mathbf{x}_1)$, $i \geq 2$. The generic LUP $\hat{Y}(\mathbf{x}_0) = \sum_{i=1}^{n} a_i Y(\mathbf{x}_i)$, $\sum_{i=1}^{n} a_1 = 1$, can be written as

$$\begin{aligned} \hat{Y}(\mathbf{x}_0) &= a_1 Y(\mathbf{x}_1) + \sum_{i=2}^{n} a_i (\tilde{Y}_i + Y\big(\mathbf{x}_1\big)) \\ &= \left(\sum_{i=1}^{n} a_i\right) Y(\mathbf{x}_1) + \sum_{i=2}^{n} a_i \tilde{Y}_i \\ &= Y(\mathbf{x}_1) + \sum_{i=2}^{n} w_i \tilde{Y}_i, \end{aligned}$$

with unconstrained weights $w_i = a_i$, $i = 2, \ldots, n$. If the w_i's are such that $\mathrm{E}(\tilde{Y}_1 | \tilde{Y}_i, i = 2, \ldots, n) = \sum_{i=2}^{n} w_i \tilde{Y}_i$ the variance

$$\mathrm{Var}\left(\tilde{Y}_1 - \sum_{i=2}^{n} w_i \tilde{Y}_i\right) = \mathrm{Var}\big(Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0)\big)$$

is minimized. Let

$$\tilde{\mathsf{R}} = \begin{bmatrix} \tilde{\mathsf{R}}_{11} & \tilde{\mathsf{R}}_{12} \\ \tilde{\mathsf{R}}_{21} & \tilde{\mathsf{R}}_{22} \end{bmatrix}$$

be the covariance matrix of $\tilde{\mathbf{Y}}$ partitioned $1|n-1$. We have $\mathbf{w} = \tilde{\mathsf{R}}_{12}\tilde{\mathsf{R}}_{22}^{-1}$ and

$$\mathrm{MSPE}[\hat{Y}_0] = \tilde{\mathsf{R}}_{11} - \tilde{\mathsf{R}}_{12}\tilde{\mathsf{R}}_{22}^{-1}\tilde{\mathsf{R}}_{21}. \tag{9}$$

Remark 1 The value of the parameter $t = \exp(-\theta)$ is usually estimated from the set of the training points, e.g. by Maximum Likelihood (or the restricted one), cross-validation or the posterior mode, and plugged in into the formula of the estimator. The estimated value is plugged into the BLUP formula: the final predictor is no longer linear, even if it is still named Empirical Best Linear Unbiased Predictor (EBLUP). For a thoroughgoing reading, see Santner et al. (2003, 64 and pages following). In this paper, we do not consider this issue, because the LH designs are compared at each value of the t parameter in $[0, 1]$. Nevertheless, as the likelihood equations are algebraic, it is possible to use Algebraic Statistics to discuss such a case as done in Catanese et al. (2006) for a different problem.

4 Latin Hypercube designs

A LH design (LHD) is a subset of the l^d lattice with l points that fully projects on each dimension. LHDs were introduced in CE by McKay et al. (1979). For fixed l and d there are $(l!)^{d-1}$ LHDs and their generation reduces to the generation of $d-1$ permutations of $1, \ldots, l$. Other types of fraction are much more difficult to sample from, which partially explains the popularity of LHDs. See Table 1 for an example with $l = 4$ and $d = 2$. The point (i, j) is denoted by ij; the LHD # 11 proceeds from the permutation $1234 \to 3412$. It has been observed that some of the LHDs are not attractive because they do not ensure a sufficient covering of the design space, see e.g. Ye et al. (2000). The covering requirement depends on the closeness between the training points and prediction ones. The MSPE (9) is

1. large when $\mathbf{x}_0$ is away from the training points $\mathbf{x}_1, \ldots, \mathbf{x}_n$;
2. small when it is close to them;
3. zero at the experimental points because of the interpolatory property of Kriging.

Table 1 The 24 LHDs for $l = 4$ and $d = 2$

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
11	11	11	14	14	11	11	11	13	13	13	14	14	13	13	13	12	12	12	14	14	12	12	12
22	22	24	21	21	24	23	23	21	21	24	23	23	24	22	22	23	23	24	22	22	24	21	21
33	34	32	32	33	33	34	32	32	34	31	31	32	32	34	31	31	34	33	33	31	31	34	33
44	43	43	43	42	42	42	44	44	42	42	42	41	41	41	44	44	41	41	41	43	43	43	44

Hence, a good planning should look for a specific choice of an LHD among the $(l!)^{d-1}$ that are available. The present paper is aimed to discuss a specific algebraic methodology to assess the properties of each LHD. See also the discussion by (Jourdan and Franco 2010) and (Petelet et al. 2010) in this Issue.

The computation of a BLUP from an LHD requires the inversion of a sub-matrix of Γ_d defined in (6). The sub-matrix has a special structure that we illustrate in the case $l = 4$, $d = 2$. The full $l \times l$ grid covariance matrix is

$$\Gamma_2 = \begin{array}{c} \\ 11 \\ 12 \\ 13 \\ 14 \\ 21 \\ 22 \\ 23 \\ 24 \\ 31 \\ 32 \\ 33 \\ 34 \\ 41 \\ 42 \\ 43 \\ 44 \end{array} \begin{array}{c} \begin{array}{cccccccccccccccc} 11 & 12 & 13 & 14 & 21 & 22 & 23 & 24 & 31 & 32 & 33 & 34 & 41 & 42 & 43 & 44 \end{array} \\ \begin{bmatrix} 1 & t & t^2 & t^3 & t & t^2 & t^3 & t^4 & t^2 & t^3 & t^4 & t^5 & t^3 & t^4 & t^5 & t^6 \\ t & 1 & t & t^2 & t^2 & t & t^2 & t^3 & t^3 & t^2 & t^3 & t^4 & t^4 & t^3 & t^4 & t^5 \\ t^2 & t & 1 & t & t^3 & t^2 & t & t^2 & t^4 & t^3 & t^2 & t^3 & t^5 & t^4 & t^3 & t^4 \\ t^3 & t^2 & t & 1 & t^4 & t^3 & t^2 & t & t^5 & t^4 & t^3 & t^2 & t^6 & t^5 & t^4 & t^3 \\ t & t^2 & t^3 & t^4 & 1 & t & t^2 & t^3 & t & t^2 & t^3 & t^4 & t^2 & t^3 & t^4 & t^5 \\ t^2 & t & t^2 & t^3 & t & 1 & t & t^2 & t^2 & t & t^2 & t^3 & t^3 & t^2 & t^3 & t^4 \\ t^3 & t^2 & t & t^2 & t^2 & t & 1 & t & t^3 & t^2 & t & t^2 & t^4 & t^3 & t^2 & t^3 \\ t^4 & t^3 & t^2 & t & t^3 & t^2 & t & 1 & t^4 & t^3 & t^2 & t & t^5 & t^4 & t^3 & t^2 \\ t^2 & t^3 & t^4 & t^5 & t & t^2 & t^3 & t^4 & 1 & t & t^2 & t^3 & t & t^2 & t^3 & t^4 \\ t^3 & t^2 & t^3 & t^4 & t^2 & t & t^2 & t^3 & t & 1 & t & t^2 & t^2 & t & t^2 & t^3 \\ t^4 & t^3 & t^2 & t^3 & t^3 & t^2 & t & t^2 & t^2 & t & 1 & t & t^3 & t^2 & t & t^2 \\ t^5 & t^4 & t^3 & t^2 & t^4 & t^3 & t^2 & t & t^3 & t^2 & t & 1 & t^4 & t^3 & t^2 & t \\ t^3 & t^4 & t^5 & t^6 & t^2 & t^3 & t^4 & t^5 & t & t^2 & t^3 & t^4 & 1 & t & t^2 & t^3 \\ t^4 & t^3 & t^4 & t^5 & t^3 & t^2 & t^3 & t^4 & t^2 & t & t^2 & t^3 & t & 1 & t & t^2 \\ t^5 & t^4 & t^3 & t^4 & t^4 & t^3 & t^2 & t^3 & t^3 & t^2 & t & t^2 & t^2 & t & 1 & t \\ t^6 & t^5 & t^4 & t^3 & t^5 & t^4 & t^3 & t^2 & t^4 & t^3 & t^2 & t & t^3 & t^2 & t & 1 \end{bmatrix} \end{array} . \tag{10}$$

A generic 2-dimensional l-elements LHD has the form $[1\sigma(1), \ldots, l\sigma(l)]$ and the element of the sub-matrix with row name $i\sigma(i)$ and column name $j\sigma(j)$ is $t^{|j-i|+|\sigma(j)-\sigma(i)|} = t^{|j-i|} \times t^{|\sigma(j)-\sigma(i)|}$ therefore the sub-matrix is $\Gamma_1 \circ \sigma(\Gamma_1)$ where $\circ$ denotes the component-wise (Hadamard) product of matrices and $\sigma(\Gamma_1)$ is the σ-permutation of rows and columns. For example, the correlation sub-matrix of the LHD # 11 is

$$\begin{array}{c} \\ 13 \\ 24 \\ 31 \\ 42 \end{array} \begin{array}{c} \begin{array}{cccc} 13 & 24 & 31 & 42 \end{array} \\ \begin{bmatrix} 1 & t^2 & t^4 & t^4 \\ t^2 & 1 & t^4 & t^4 \\ t^4 & t^4 & 1 & t^2 \\ t^4 & t^4 & t^2 & 1 \end{bmatrix} \end{array} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 1 & t & t^2 & t^3 \\ t & 1 & t & t^2 \\ t^2 & t & 1 & t \\ t^3 & t^2 & t & 1 \end{bmatrix} \end{array} \circ \begin{array}{c} \\ 3 \\ 4 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cccc} 3 & 4 & 1 & 2 \end{array} \\ \begin{bmatrix} 1 & t & t^2 & t \\ t & 1 & t^3 & t^2 \\ t^2 & t^3 & 1 & t \\ t & t^2 & t & 1 \end{bmatrix} \end{array} .$$

4.1 The method

We present below a step-by-step description of our algebraic way to perform a comparison in the class of LHDs with d variables and/or factors, each one with l levels. Illustration is given for the case $d = 2$, $l = 4$. Computations are organized in a batch which calls for the statistical software R (for the manipulation of the matrices) and the algebraic software CoCoA (for symbolic computations).

Global covariance matrix. The covariance matrix $\Gamma_d = \Gamma_1^{\otimes d}$, where Γ_1 is given in (3), is computed by Kronecker products, e.g. (10).

Table 2 The 24 LHDs for $l = 4$ and $d = 2$ classified according the TMSPE criterion (bottom line)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
11	11	11	14	14	11	11	11	13	13	13	14	14	13	13	13	12	12	12	14	14	12	12	12
22	22	24	21	21	24	23	23	21	21	24	23	23	24	22	22	23	23	24	22	22	24	21	21
33	34	32	32	33	33	34	32	32	34	31	31	32	32	34	31	31	34	33	33	31	31	34	33
44	43	43	43	42	42	42	44	44	42	42	42	41	41	41	44	44	41	41	41	43	43	43	44
1	2	3	4	3	4	3	5	3	6	7	2	1	2	3	4	3	4	3	5	3	6	7	2

Generation of the LHDs. Generation of permutations of the integers $1, 2, \ldots, l$ produces a $l \times (l!)^{d-1}$ table containing all the LHDs, e.g. Table 1.

Sub-setting. For each LHD and each $\mathbf{x}_0$ in its complement the joint covariance matrix is computed by sub-setting Γ_d; MSPE for each $\mathbf{x}_0$ is symbolically computed as a rational function of t. The Total of the MSPEs over all points to be predicted (TMSPE) is obtained for all LHDs as rational functions of t and subsequently TMSPEs are clustered according to the algebraic form of the rational function. The TMSPE index replaces the Integrated MSPE in the discrete case. For example, with $l = 4$ and $d = 2$, the following output is obtained from CoCoA. It is a list of rational functions, each one representing the MSPE of one of the seven classes.

```
C1:
(2t^8 - 4t^7 - 12t^6 + 4t^5 + 8t^4 + 20t^3 + 12t - 30)/(t^4 - t^2 - 2)
C2:
(t^13 + t^12 + 1/2t^11 - 11/2t^10 - 31/2t^9 - 19/2t^8 - 5/2t^7 + 71/2t^6
+ 43/2t^5 + 71/2t^4 + 13t^3 - 27t^2 - 18t - 30)/
(t^7 + t^6 - 2t^4 - 2t^3 - 4t^2 - 2t - 2)
C3:
(1/2t^15 + 1/2t^14 + t^13 - 4t^12 + 3/2t^11 - 17/2t^10 - 47/2t^9 + 9/2t^8
- 7t^7 + 48t^6 + 59/2t^5 + 49/2t^4 + 16t^3 - 35t^2 - 18t - 30)/
(t^8 + t^7 - 4t^4 - 2t^3 - 5t^2 - 2t - 2)
C4:
(1/2t^14 - 1/2t^12 - 2t^11 + 3t^10 - 4t^9 - 18t^8 + 12t^7 - 27/2t^6
+ 40t^5 + 3/2t^4 + 38t^3 - 39t^2 + 12t - 30)/(t^6 - t^4 - 4t^2 - 2)
C5:
(t^15 + t^14 + t^13 - 7t^12 - 5t^11 - 5t^10 - 23t^9 + 13t^8 + 2t^7
+ 42t^6 + 20t^5 + 20t^4 + 22t^3 - 34t^2 - 18t - 30)/
(t^9 + t^8 + 2t^7 - 2t^6 - 4t^4 - t^3 - 5t^2 - 2t - 2)
C6:
(t^9 - 3t^8 + 2t^7 + 2t^6 - 22t^5 + 8t^4 + 28t^3 + 8t^2 - 9t - 15)/
(t^3 - t^2 - t - 1)
C7:
(3t^8 - 2t^7 + 6t^6 - 9t^5 - 5/2t^4 - 10t^3 + 10t^2 - 3t + 15/2)/
(t^2 + 1/2),
```

which, in turn, gives rise to the classification shown in the last row of Table 2.

Performance evaluation Each rational function is studied to assess the performance in each class of LHD. For example, in graphical form as in Fig. 2; the classification obtained is best appreciated by looking at a graphical representation of the 24 LHDs in Table 3. Each of the seven classes is an orbit of the action of the dihedral group of the square on the set of all LHDs.

4.2 Discussion of the example

In this case there is not much difference between the LHDs if the criteria is TMSPE. Some difference can be seen in the second graph showing a relative difference near

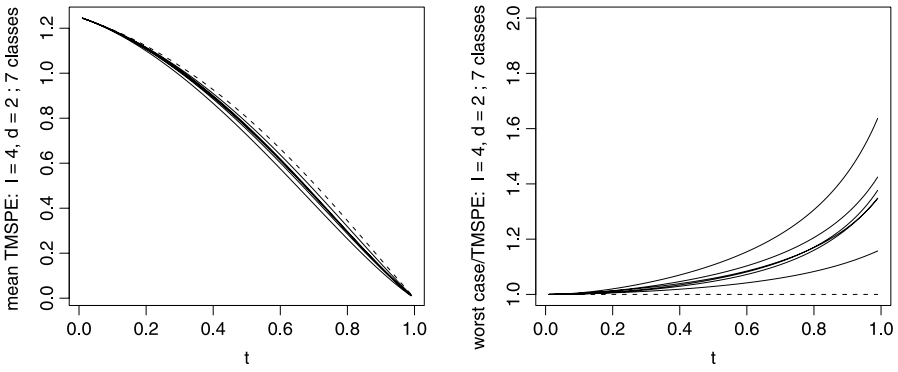
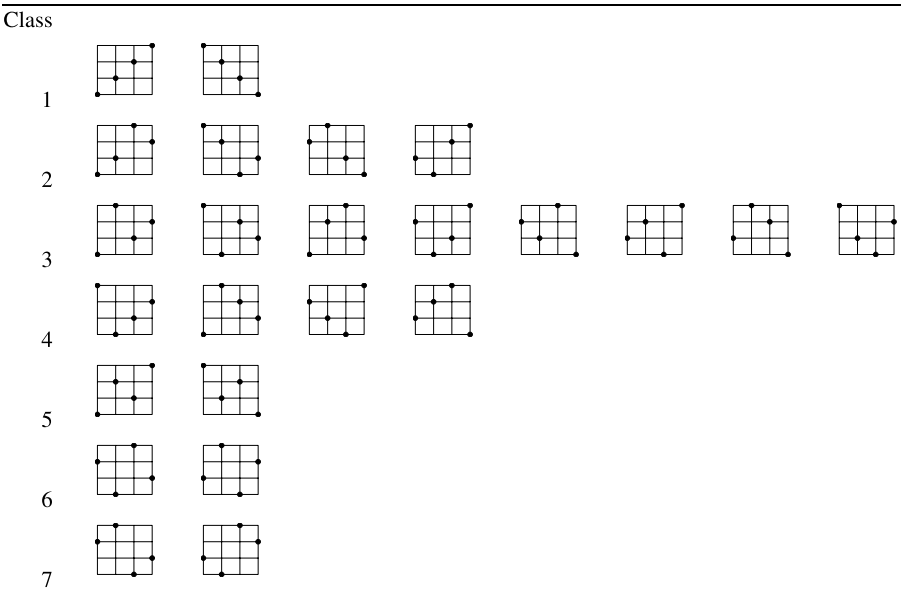


Fig. 2 Performance of LHDs for $l = 4$ and $d = 2$

Table 3 24 LHDs classified according the TMSPE criteria



$t = 1$, that is, $\theta = 0$ (constant covariance). As the computations are symbolic there is no risk of numerical errors near the critical point. Other criteria have been suggested, for example based on entropy, see Shewry and Wynn (1987). Our methodology, when applied to the entropy instead of TMSPE, produces the same classification of LHDs as in Table 2, because of the invariance argument, and the same ranking of classes.

Class 6 shows the best performance. It consists of LHDs which are maximin in the L^1 distance, see Morris and Mitchell (1995); it also consists of "tilted" 2^2 designs; the designs in this class are called U-design in Tang (1993). Classes 3, 4, 5 are essentially equivalent and worse than class 6. Designs in class 5 are called cyclic designs in Bates et al. (1996). Class 2 is the second worst. Classes 1 and 4 consist of regular fractions 4^{2-1}. Class 3 is made of regular fractions 2^{4-2} (pseudo factors).

4.3 Other examples

The methodology we have illustrated in the case $l = 4$ and $d = 2$ is, in principle, of general applicability. Figure 3 shows the results in the case $l = 4$ and $d = 3$, and Fig. 4 the results for $l = 6$ and $d = 2$. The choice of these dimensions for levels and number of factors is merely indicative of the type of comparisons that are available. The dashed lines represent the TMSPE of the LHDs whose points lie on the diagonals of the grid. These designs are considered to not be properly space filling, even if they have nice marginal properties and the largest value of the respective TMSPE confirm their poor capability of prediction. Other remarkable features shown are the minimum value of the TMSPE corresponding to maximum correlation $t = 1$ and the maximum value of the TMSPE corresponding to the null correlation $t = 0$ between the training points (i.e. independence, according to the assumption of normality). This is coherent with the Kriging prediction methodology: predicting at an untried location, observations closer to it should have more influence on the prediction because of the existing correlation. In the right-hand graphs, we resort to a different representation of the comparison index relating the TMSPEs to the worst case, in order to magnify the

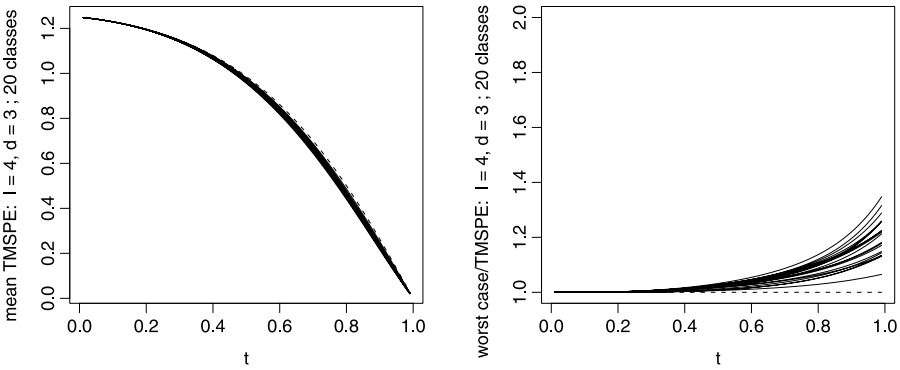


Fig. 3 Performance of LHDs for $l = 4$ and $d = 3$

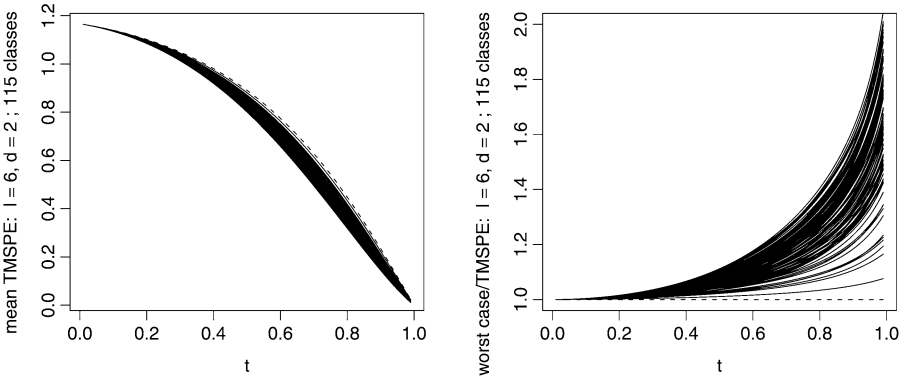


Fig. 4 Performance of LHDs for $l = 6$ and $d = 2$

differences: the ratios between the TMSPE of the diagonal LH designs and the ones corresponding to each of the other classes are plotted against $t = \exp(-\theta)$.

5 Discussion

The main result of this paper is the illustration of symbolic computations in a particular model of Kriging. It shows the potential applicability, in this area, of ideas from Algebraic Statistics, see. the review in Gibilisco et al. (2009). Interesting exploratory results are obtained, for example the neat classification and ranking of the 4×4 LHDs shown in Table 2, Fig. 2, and Table 3. The same classification and ranking was obtained for other optimality criteria, e.g. entropy. While the classes depend on the symmetry of the problem under the dihedral group of the square, it is notable that it is the result of a symbolic computation. The algebraic form is also of use in evaluating the relative error near the critical point $t = 1$.

The methodology depends on the algebraic form of the SCF and could be implemented for any class of training set, LHDs being an example. Some of the other popular performance criteria are based on non-algebraic operations, typically max MSPE at untried points, and cannot be solved algebraically in a standard way.

This methodology has limitations when the number of levels and/or factors is large. Indeed, the very high complexity of the symbolic computations limits considerably the dimension of the problems that can be practically solved. In this paper, we have not tried any optimization of the program we have used, which consists of a sequence on runs from standard packages, nor we considered writing any ad-hoc software. Such an optimization would require the study of efficient ways to generate the training sets from the class of interest, which is a general problem not depending on the algebraic methodology, and the study of algorithms for the symbolic computation of the inverse and the determinant of a correlation matrix with polynomial entries.

Acknowledgements A preliminary version of the present paper was presented at the EMBIS-EMSE 2009 Conference, July 1–3, Saint-Étienne, France. We thank the editors and anonymous referees for comments and suggestions.

References

Bates, R.A., Buck, R.J., Riccomagno, E., Wynn, H.P.: Experimental designs and observations for large systems. J. R. Stat. Soc. B **58**, 77–94 (1996)

Berg, C., Christensen, J.P.R., Ressel, P.: Harmonic Analysis on Semigroups. Theory of Positive Definite and Related Functions. Graduate Texts in Mathematics, vol. 100. Springer, New York (1984)

Butler, N.A.: Optimal and orthogonal Latin Hypercube designs for computer experiments. Biometrika **88**(3), 847–857 (2001)

Catanese, F., Hoşten, S., Khetan, A., Sturmfels, B.: The maximum likelihood degree. Am. J. Math. **128**(3), 671–697 (2006)

CoCoATeam (online) CoCoA: a system for doing computations in commutative algebra. Available at cocoa.dima.unige.it

Cressie, N.: Kriging nonstationary data. J. Am. Stat. Assoc. **81**(395), 625–634 (1986)

Fang, K.T., Lin, D.K.J., Winker, P., Zhang, Y.: Uniform design: theory and application. Technometrics **42**(3), 237–248 (2000)

Fang, K.T., Li, R., Sudjianto, A.: Design and Modeling for Computer Experiments. Computer Science and Data Analysis Series. Chapman & Hall/CRC, Boca Raton, FL (2006)
Gibilisco, P., Riccomagno, E., Rogantin, M., Wynn, H.P. (eds.): Algebraic and Geometric Methods in Statistics. Cambridge University Press, Cambridge (2009)
Jourdan, A., Franco, J.: Optimal Latin hypercube designs for the Kullback–Leibler criterion. Adv. Stat. Anal. **94**(4), 341–351 (2010)
Krige, D.G.: A statistical approach to some basic mine valuation problems on the witwatersrand. J. Chem. Metall. Min. Soc. S. Afr. **52**(6), 119–139 (1951)
Lauritzen, S.L.: Graphical Models. Clarendon/Oxford University Press, New York (1996),
Levy, S. Steinberg, D.M.: Computer experiments: a review. Adv. Stat. Anal. **94**(4), 311–324 (2010)
McKay, M.D., Beckman, R.J., Conover, W.J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics **21**, 239–245 (1979)
Morris, M.D., Mitchell, T.J.: Exploratory designs for computational experiments. J. Stat. Plan. Inference **43**(3), 381–402 (1995)
Ortega, J.M.: Matrix Theory. The University Series in Mathematics. Plenum, New York (1987), a second course
Park, J.S.: Optimal Latin-hypercube designs for computer experiments. J. Stat. Plan. Inference **39**(1), 95–111 (1994)
Petelet, M., Iooss, B., Asserin, O., Loredo, A.: Latin hypercube sampling with inequality constraints. Adv. Stat. Anal. **94**(4), 325–339 (2010)
Pistone, G., Vicario, G.: Random generation of designs in computer experiments. In: Statistics for Innovation. Statistical Design of "Continuous" Product Innovation. Springer, Berlin, pp. 91–102 (2009)
Sacks, J., Schiller, S.B., Welch, W.J.: Designs for computer experiments. Technometrics **31**(1), 41–47 (1989a)
Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P.: Design and analysis of computer experiments (with discussion). Stat. Sci. **4**, 409–435 (1989b)
Santner, T.J., Williams, B.J., Notz, W.I.: The Design and Analysis of Computer Experiments. Springer Series in Statistics. Springer, New York (2003)
Sasena, M.J.: Flexibility and efficiency enhancements for constrained global design optimization with Kriging approximations. PhD thesis. University of Michigan (2002)
Shewry, M.C., Wynn, H.P.: Maximum entropy sampling. J. Appl. Stat. **14**, 165–170 (1987)
Tang, B.: Orthogonal array-based Latin hypercubes. J. Am. Stat. Assoc. **88**(424), 1392–1397 (1993)
Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Mitchell, T.J., Morris, M.D. Screening, predicting, and computer experiment. Technometrics **34**(1), 15–25 (1992)
Ye, K.Q., Li, W., Sudjianto, A.: Algorithmic construction of optimal symmetric Latin hypercube designs. J. Stat. Plan. Inference **90**(1), 149–159 (2000)