Comparing and generating Latin Hypercube designs in Kriging models

Giovanni Pistone, Grazia Vicario[1]

Politecnico di Torino - Department of Mathematics
Corso Duca degli Abruzzi, 24 – 10129 Torino, Italy

Abstract

Nowadays, any researcher acknowledges that Computer Experiments (CE) are useful in replacing totally or partially physical experiments, because the physical experimentation may be, in a number of circumstances, expensive and unapproachable. In both contests, a careful selection of the design points is mandatory if the aim is to study how the observed responses vary when the input variables are set to the values of interest. In physical experimentation, the protocol is the Design of Experiments (DoE), methodology whose basic principles are questioned in CE. When the output of a CE can be predicted by assuming Gaussian responses with covariance depending parametrically on the distance between the locations the use of the so called space-filling designs (random designs, stratified designs and Latin Hypercube (LH) designs) is a common choice, according to the principle that the nearer an untried point is to the design points the better is the prediction. In this paper we focus on the class of LH designs. According to the mentioned Gaussian assumption, the behavior of each LH designs is examined, in order to minimize the prediction error in the points of the considered regular lattice. Thanks to a symbolic algebraic software and a statistical one, we provide the LH designs with any size (number of factors and levels) and the computation of the variance of the Gaussian linear prediction as a function of the design, in order to allow the user to perform the comparison of the LH designs.

Keywords: Computer Experiments, Latin Hypercube, Mean Square Prediction Error, Gaussian linear prediction.

1. Introduction

The official start of the Computer Experiments (CE) is the paper of McKay et al. (1979) while the contribution by Sachs J. et al. (1989) marked a new step by introducing model based methods. An actual reason for promoting the use of the CE (in a single or a combined approach with the physical experiments) is that the physical experimentation may be, in a number of circumstances, expensive and unapproachable. On the contrary, the use of the numerical experiments in product/process development phase is quite inexpensive and it 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 option 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 contest, a careful selection of the design points or training points is mandatory if the aim is to study how the observed responses vary when the input variables are set to the values of interest. 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 on the mentioned objectives. The design of a CE, when it is used as a surrogate of the physical one, differs in several aspects from designing a physical experiment and the applicability of basic principles of D.o.E. is questioned in CE. The selection of an experimental design in CE is a crucial issue to get to 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 in untried inputs.
As suggested by the pioneers of the model based CE, the output can be predicted by assuming Gaussian responses with covariance depending parametrically on the distance between the locations, as it was in

1 Corresponding Author: grazia.vicario@polito.it

the Kriging modellization (Krige, 1951; Cressie, 1993). The underlying principle is that the nearer an untried point is to the design points the better is the prediction. Based on this view, a good design strategy is to uniformly spread the points across the experimental region. That prompts to the use of the so called space-filling designs. Random designs, Stratified designs and Latin Hypercube designs are common choices. The first two designs are not really satisfactory because they are not space filling marginally, i.e. in individual directions, and, moreover, are quite not satisfactory for global space filling especially for small number of design points. Whereas, 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 Butler (2001), Fang (2000), Park (1994), Welch (1992) etc. This is the reason for focussing on the class of LH designs among the different space-filling ones in this paper: we investigate which ones are the best in the class of LH designs having the same prediction features.

In this paper, the Gaussian field is assumed to have distance-dependent covariance (stationarity) and it is defined on a subset of a regular lattice, i.e. the cartesian product of uniform one-dimensional lattices. This case correspond to specific applications, see Pistone and Vicario (2009), and it is specially adapted to the methodology we use. The commonly used Euclidean distance does not really fit when considering a regular lattice; therefore we switch to the Manhattan distance, as other authors suggest (Santner et al., 2003, pag. 138). We examine the behaviour of different LH design's: the Mean Square Prediction Error (MSPE) of the linear predictor is the statistical index chosen for comparing different LH designs with the same number of training points. It should be noticed that the issue of identifiability which is of the highest importance in standard DoE is not relevant here.

In performing the mentioned comparisons, there are two tricky computational problems: the computation of variance-covariance matrix of the design points and the computation of the closed form expression of the predictors variance. For the former problem, a solution is presented in a particular case of correlation function, even if is one of the most common one used by the CE practitioners. For the latter, the difficulties are in the computation of the rational functions in the covariances with rational coefficients. For computing the predictor variances in closed form, we resort to a symbolic algebraic software such as CoCoA (Computations in Commutative Algebra), a freely available system for symbolic exact multivariate polynomial computation (see http://cocoa.dima.unige.it). Other computations related with the exponential model for covariances are done with the software R (see http://www.R-project.org/). The final result is a general methodology to analyze the MSPE efficiency of the class of interest of training sets for regular lattices and covariances of the form negative exponential of the Manhattan distance.

2. Correlation function on a lattice

Let us consider a Gaussian random field $Y(\mathbf{x})$ with zero mean and stationary covariance over a design space $\mathcal{X}_d \subset \mathbb{R}^d$, i.e. $\mathbb{E}[Y(\mathbf{x})] = 0$ and $\mathrm{Cov}(Y(\mathbf{x}), Y(\mathbf{x}+\mathbf{h})) = \sigma_Y^2 R(\mathbf{h};\boldsymbol{\theta})$, where σ_Y^2 is the field variance, R is the Stationary Correlation Function (SCF) depending only on the displacement vector $\mathbf{h}$ between any pair of points in $\mathcal{X}$ and on a vector parameter $\boldsymbol{\theta}$. The original suggestion, the most popular choice for the correlation function among the practitioners of CEs, is within the power exponential family:

$$R(h;\theta) = \prod_{s=1}^{d} exp\left\{-\theta_s |h_s|^p\right\} = exp\left\{-\sum_{s=1}^{d} \theta_s |h_s|^p\right\} \qquad \text{with } 0 < p \le 2 \tag{2.1}$$

where $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_d,p)'$; p is a common smoothing parameter and θ_s, $s = 1, 2, \ldots, d$, are positive scale parameters representing the rapidity of the correlation decays in direction s when increasing distance h_s. The conditions $\theta_s > 0$, $s = 1, 2, \ldots, d$, and p between 0 and 2 for the correlation model in (2.1) are necessary and sufficient for a function of that form to be positive definite on an Euclidean space and, therefore, for the existence of a stationary Gaussian field with that covariance function (Berg et *al*., 1984). The assumption in (2.1) that the positive correlation between outputs diminishes with increasing distance between their input sites is the formalization of the original Krige's idea (Krige, 1951). If $\theta_s=\theta$, $\forall s = 1, 2, \ldots, d$, the correlation depends only on the distance $|\mathbf{h}|$ between any pair of points $\mathbf{x}$ and $\mathbf{x}+\mathbf{h}$.

In the present computations we are going to assume that the Gaussian field is defined on a regular rectangular lattice i.e.

$$\mathcal{X}_d = \{1, \ldots, l\}^d \tag{2.2}$$

We switch to the Manhattan distance, i.e. $|x-y| = \sum_{s=1}^{d} |x_s - y_s|$, because the Euclidean distance is not really adapted to a regular lattice. Let us consider the univariate case $d=1$ and the particular case $p = 1$; the distance function is:

$$d_1(i,j) = |i-j| \qquad i, j = 1, 2, \ldots, l \tag{2.3}$$

and the covariance function is:

$$R_1(h;\theta) = exp\{-\theta d_1(i,j)\} = t^{d_1(i,j)}, \qquad h = i-j \tag{2.4}$$

with: $t = exp(-\theta)$, $t \in (0,+\infty)$; i and j in $\{1, \ldots, l\}$. The $\binom{l}{2}$ values of the distances in (2.3) may be collected in a distance matrix $\mathbf{D}_1$ that in the univariate case is:

$$\mathbf{D}_1 = \begin{pmatrix} 0 & 1 & 2 & . & . & l-1 \\ 1 & 0 & 1 & 2 & . & . \\ 2 & 1 & 0 & . & . & . \\ . & . & . & . & . & 2 \\ . & . & . & . & . & 1 \\ l-1 & . & . & . & . & 0 \end{pmatrix} \tag{2.5}$$

and the corresponding covariance matrix $\mathbf{\Gamma}_1$ is:

$$\mathbf{\Gamma}_1 = \begin{pmatrix} 1 & t & t^2 & . & . & t^{l-1} \\ t & 1 & t & t^2 & . & . \\ t^2 & t & 1 & . & . & . \\ . & . & . & . & . & t^2 \\ . & . & . & . & 1 & t \\ t^{l-1} & . & . & . & t & 1 \end{pmatrix} \tag{2.6}$$

If $d = 2$, the distance function is (see Fig. 2.1):

$$d_2((i_1,i_2),(j_1,j_2)) = |j_1 - i_1| + |j_2 - i_2| \equiv d_1(i_1,j_1) + d_1(i_2,j_2) \tag{2.7}$$

and the covariance function is:

$$R_2(\mathbf{h};\theta) = exp\{-\theta d_2((i_1,i_2),(j_1,j_2))\} = t^{d_1(i_1,j_1)+d_1(i_2,j_2)} = t^{d_1(i_1,j_1)} t^{d_1(i_2,j_2)} \tag{2.8}$$

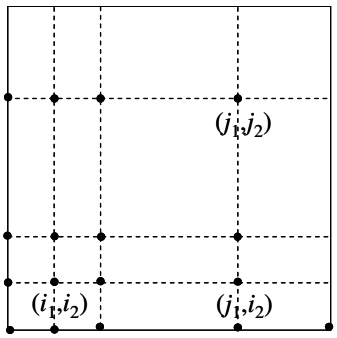


Figure 2.1 – A representation of the bivariate rectangular lattice $\mathcal{X}_2 = \{1, \ldots, l\}^2$.

The corresponding matrix of the distances and the covariance matrix are respectively:

$$\mathbf{D}_2 = \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_1+\mathbf{1} & \mathbf{D}_1+\mathbf{2} & . & . & \mathbf{D}_1+\mathbf{l}-\mathbf{1} \\ \mathbf{D}_1+\mathbf{1} & \mathbf{D}_1 & \mathbf{D}_1+\mathbf{1} & \mathbf{D}_1+\mathbf{2} & . & . \\ \mathbf{D}_1+\mathbf{2} & \mathbf{D}_1+\mathbf{1} & \mathbf{D}_1 & . & . & . \\ . & . & . & . & . & \mathbf{D}_1+\mathbf{2} \\ . & . & . & . & . & \mathbf{D}_1+\mathbf{1} \\ \mathbf{D}_1+\mathbf{l}-\mathbf{1} & . & . & . & . & \mathbf{D}_1 \end{pmatrix} \tag{2.9}$$

and:

$$\mathbf{\Gamma}_2 = \mathbf{\Gamma}_1 \otimes \mathbf{\Gamma}_1 \tag{2.10}$$

where: $\mathbf{D}_1$ is (2.5) and $\mathbf{J} = j \times \mathbf{1}$, being $\mathbf{1}$ the matrix of all ones, with $j = 1, 2, \ldots, l-1$, and the symbol $\otimes$ is the Kronecker product between matrices. Exploiting the previous formulas to the more general multivariate case in d dimension, we have:

$$\mathbf{D}_d = \begin{pmatrix} \mathbf{D}_{d-1} & \mathbf{D}_{d-1}+\mathbf{1} & \mathbf{D}_{d-1}+\mathbf{2} & . & . & \mathbf{D}_{d-1}+\mathbf{l}-\mathbf{1} \\ \mathbf{D}_{d-1}+\mathbf{1} & \mathbf{D}_{d-1} & \mathbf{D}_{d-1}+\mathbf{1} & \mathbf{D}_{d-1}+\mathbf{2} & . & . \\ \mathbf{D}_{d-1}+\mathbf{2} & \mathbf{D}_{d-1}+\mathbf{1} & \mathbf{D}_{d-1} & . & . & . \\ . & . & . & . & . & \mathbf{D}_{d-1}+\mathbf{2} \\ . & . & . & . & . & \mathbf{D}_{d-1}+\mathbf{1} \\ \mathbf{D}_{d-1}+\mathbf{l}-\mathbf{1} & . & . & . & . & \mathbf{D}_{d-1} \end{pmatrix} \tag{2.11}$$

$$\mathbf{\Gamma}_d = \mathbf{\Gamma}_{d-1} \otimes \mathbf{\Gamma}_1 \tag{2.12}$$

The formulas (2.11) and (2.12) are fundamental for studying the Gaussian field model, for implementing the computation of variances and covariances in closed form, for comparing different LH designs.

3. Predicting the output in the lattice points

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

$$Y(x) = \boldsymbol{f}'(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}) \tag{3.1}$$

where $\boldsymbol{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}))'$ is a set of specified trend functions, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_m)'$ is a set of (usually unknown) parameters and $Z(\mathbf{x})$ is Gaussian random field with zero mean and stationary covariance over $\mathcal{X}_d$. For the prediction of the response $Y(\mathbf{x}_0)$ at an untried point $\mathbf{x}_0$, most of the practitioners suggest to resort to Bayesian estimators; the prior information on the set $\mathbf{Y}^n = (Y(\mathbf{x}_1), Y(\mathbf{x}_2), \ldots, Y(\mathbf{x}_n))'$ of field variables at $\mathbf{x}^n = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, the training data, is used for predicting the unknown output $Y(\mathbf{x}_0)$. The underlying hypothesis (Santner et al. 2003), consistent with (2.1) and (2.3), assumes that the distribution of the joint random variable $(Y(\mathbf{x}_0), Y(\mathbf{x}_1), Y(\mathbf{x}_2), \ldots, Y(\mathbf{x}_n))$ is normal, $N\left[(\boldsymbol{f}_0', \mathbf{F})'\boldsymbol{\beta}, \sigma_Z^2\boldsymbol{\Sigma}\right]$, with $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \mathbf{r}_0' \\ \mathbf{r}_0 & \mathbf{R} \end{pmatrix}$ and $\mathbf{r}_0$ is the correlation vector $(R(\mathbf{x}_0 - \mathbf{x}_1), \ldots, R(\mathbf{x}_0 - \mathbf{x}_n))'$, $\mathbf{R}$ is the $n \times n$ correlation matrix whose (i,j) element is $R(\mathbf{h}_{ij} = \mathbf{x}_i - \mathbf{x}_j)$; $\mathbf{F}$ is the $n \times m$ matrix $\{f_j(\mathbf{x}_i)\}_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$ of the trend functions evaluated in $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$.

For sake of simplicity, we consider here the *ordinary* Kriging model, i.e. $\boldsymbol{f}'(\mathbf{x})\boldsymbol{\beta} = \beta$:

$$Y(\mathbf{x}) = \beta + Z(\mathbf{x}) \tag{3.2}$$

where the trend is constant even if unknown (opposite to the *universal* Kriging model of (3.1)) and the field variance σ_Y^2 is unit, without affecting our conclusions. This is because we assume that no a-priori

knowledge on the trend function is available to direct the user in its choice.
The prediction is based on the concept of Linear Unbiased Predictor (LUP). In fact, Kriging methodology is a linear spatial interpolation: the random variable $Y(\mathbf{x}_0)$ is predicted by an affine combination of the observed random variables $Y(\mathbf{x}_1), Y(\mathbf{x}_2), ..., Y(\mathbf{x}_n)$:

$$\hat{Y}(\mathbf{x}_0) = a_0 + \sum_{i=1}^{n} a_i Y(\mathbf{x}_i) \tag{3.3}$$

that it is unbiased if and only if $\beta \equiv \mathbb{E}\left[\hat{Y}(\mathbf{x}_0)\right] = a_0 + \sum_{i=1}^{n} a_i \beta$, i.e. $a_0 = 0$ and $\sum_{i=1}^{n} a_i = 1$, for any $\beta \in \mathbb{R}$.

Moreover, the predictor $\hat{Y}(\mathbf{x}_0) = \sum_{i=1}^{n} a_i Y(\mathbf{x}_i)$, with $\sum_{i=1}^{n} a_i = 1$, is the Best (BLUP) if minimizes the Mean Squared Prediction Error (MSPE), a measure of uncertainty of predictions, that means that:

$$\mathrm{MSPE}\left[\hat{Y}_0\right] = \mathbb{E}\left[\left(\hat{Y}_0 - Y(\mathbf{x}_0)\right)^2\right] \tag{3.4}$$

is minimized. In fact, if β and the correlation (2.1) are known, the conditional expectation:

$$\hat{Y}_0 = \mathbb{E}\left(Y(\mathbf{x}_0) \middle| \mathbf{Y}^n\right) = \beta + \mathbf{r}_0' \mathbf{R}^{-1}\left(\mathbf{Y}^n - \beta \mathbf{u}_n\right) \tag{3.5}$$

is an unbiased estimator with minimum MSPE:

$$\mathrm{MSPE}\left[\hat{Y}_0\right] = 1 - \mathbf{r}_0' \mathbf{R}^{-1} \mathbf{r}_0 \tag{3.6}$$

being $\left(Y(\mathbf{x}_0), \mathbf{Y}^n\right)$ a partition of the vector $\left(Y(\mathbf{x}_0), Y(\mathbf{x}_1), Y(\mathbf{x}_2), ..., Y(\mathbf{x}_n)\right)$ and $\mathbf{u}_n$ the unit vector. If β is unknown, its generalized least squares estimator $\hat{\beta} = \left(\mathbf{u}_n' \mathbf{R}^{-1} \mathbf{u}_n\right)^{-1} \mathbf{u}_n' \mathbf{R}^{-1} \mathbf{Y}^n$ must replace β in the predictor (3.5) providing a new predictor. In such a case, the MSPE, usually called *kriging variance*, is larger than the (3.6), because there is an additional uncertainty component in it, and it is:

$$\mathrm{MSPE}\left[\hat{Y}_0\right] = 1 - \mathbf{r}_0' \mathbf{R}^{-1} \mathbf{r}_0 + \mathbf{c}_0' \left(\mathbf{u}_n' \mathbf{R}^{-1} \mathbf{u}_n\right)^{-1} \mathbf{c}_0 \tag{3.7}$$

with $\mathbf{c}_0 = 1 - \mathbf{u}_n' \mathbf{R}^{-1} \mathbf{r}_0$. In the following computations, we use the best linear unbiased predictor and its MSPE. For lowering the size of the matrices to be inverted (the less the size the easier the symbolic computation), we resort to a device. Let be:

$$\begin{cases} \tilde{Y}_1 \equiv -Y(\mathbf{x}_1) \\ \tilde{Y}_2 = Y(\mathbf{x}_2) - Y(\mathbf{x}_1) \\ \tilde{Y}_3 = Y(\mathbf{x}_3) - Y(\mathbf{x}_1) \\ \cdot \\ \cdot \\ \tilde{Y}_n = Y(\mathbf{x}_n) - Y(\mathbf{x}_1) \end{cases} \text{ consistently with: } \begin{cases} Y(\mathbf{x}_1) \equiv -\tilde{Y}_1 \\ Y(\mathbf{x}_2) = \tilde{Y}_2 - \tilde{Y}_1 \\ Y(\mathbf{x}_3) = \tilde{Y}_3 - \tilde{Y}_1 \\ \cdot \\ \cdot \\ Y(\mathbf{x}_n) = \tilde{Y}_n - \tilde{Y}_1 \end{cases} \tag{3.8}$$

The linear predictor of $Y(\mathbf{x}_0)$ (3.3) is, according the linear transformation (3.8) and the constraint of unbiasdness:

$$\hat{Y}(\mathbf{x}_0) = -\tilde{Y}_1 + \sum_{i=2}^{n} w_i \tilde{Y}_i \tag{3.9}$$

with $\mathbf{x}_0 \in \mathcal{X}_d / \mathbf{x}^n = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$ (i.e. the training set); the MSPE of the LUP (3.9) may be written as:

$$\mathrm{MSPE}\left[\hat{Y}_0\right] = \mathbb{E}\left[\left(Y(\mathbf{x}_0) - \hat{Y}_0\right)^2\right] = \mathbb{E}\left[\left(Y(\mathbf{x}_0) + \tilde{Y}_1 - \sum_{i=2}^{n} w_i \tilde{Y}_i\right)^2\right] = \mathbb{E}\left[\left(\tilde{Y}_0 - \sum_{i=2}^{n} w_i \tilde{Y}_i\right)^2\right] \tag{3.10}$$

being $\tilde{Y}_0 = Y(\mathbf{x}_0) - Y(\mathbf{x}_1)$. It is worthwhile to remark that $\mathbb{E}\left[\left(\tilde{Y}_i\right)\right] = 0$, i = 0, 1, …, n always. Therefore, the conditional expectation $\mathbb{E}\left(\tilde{Y}_0 \middle| \tilde{Y}_2, \tilde{Y}_3, ..., \tilde{Y}_n,\right)$ has minimum MSPE with respect to the coefficients w_i, i = 2, 3, …, n, and according to the (3.10). For this aim, let be $\tilde{\mathbf{R}}$ the covariance matrix of the random variable $\left(Y(\mathbf{x}_0) - Y(\mathbf{x}_1), Y(\mathbf{x}_2) - Y(\mathbf{x}_1), ..., Y(\mathbf{x}_n) - Y(\mathbf{x}_1)\right) \equiv \left(\tilde{Y}_0, \tilde{Y}_2, ..., \tilde{Y}_n\right)$; it is:

$$\tilde{\mathbf{R}} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' \tag{3.11}$$

where $\mathbf{A}$ is the (n, n+1) matrix of the linear transformation generated by (3.8). Therefore, partitioning the matrix $\tilde{\mathbf{R}}$ into:

$$\tilde{\mathbf{R}} = \begin{pmatrix} \tilde{R}_{11} & \tilde{\mathbf{r}}'_0 \\ \tilde{\mathbf{r}}_0 & \tilde{\mathbf{R}}_{22} \end{pmatrix} \tag{3.12}$$

it is: $\mathbf{w} = \tilde{\mathbf{r}}'_0 \tilde{\mathbf{R}}_{22}^{-1}$ and the minimum value of MSPE is:

$$\mathrm{MSPE}\left[\hat{Y}_0\right] = \tilde{R}_{11} - \tilde{\mathbf{r}}'_0 \tilde{\mathbf{R}}_{22}^{-1} \tilde{\mathbf{r}}_0 \tag{3.13}$$

It is worthwhile to note that the covariance matrix $\tilde{\mathbf{R}}_{22}^{-1}$ does not depend on the prediction point $\mathbf{x}_0$: this is a very important feature for the next computations.
The unknown value of the parameter $t = \exp(-\theta)$ may be estimated from the set of the training points, e.g. by the Maximum Likelihood (or the restricted one), cross-validation or the posterior mode, and plugged in into the formula of the estimator. It has to be underlined that the final estimators are no more linear, even if they are still named Empirical Best Linear Unbiased Predictors (EBLUPs). For a thoroughgoing reading, see Santner *et al.*, 2003, pagg. 64-.

4. The class of Lattice Hypercube Designs

It has been frequently observed that some LH designs are not attractive because they do not ensure a sufficient covering of the design space, e.g. Ye et *al.*, 2003. The covering requirement depends on the closeness between the training points and prediction ones: the MSPE (3.13) is large when $\mathbf{x}_0$ is away from the training points and small when it is close to them; it vanishes at the experimental points because the interpolatory property of Kriging. Hence, a reasonable planning may lead to a particular choice of an LH design among the $(l!)^{d-1}$ ones in order to fulfil the afore mentioned demand.
One after the other, we present our steps for the completion of the comparison in the class of the LH designs with d variables and/or factors, each one with l levels:

#1 Permutations of the l integers (number of the levels) and construction of the matrix $l \times (l!)^{d-1}$ containing all the LH designs with d factors. It should be remarked that we resorted to a convenient enumeration of row and column of this matrix to ease the choice of a specific design. This enumeration is coherent with the points (i_1, i_2, …, i_d) of the lattice to facilitate the proximate steps.
In the Table 4.1, we present the possible 24 LH designs relative to two factors each one with four levels. Each column is made of the two coordinates (i_1, i_2) of the lattices in Figure 4.1 and pertaining to an LH design.

#2 Construction of the distance matrix between any pair of points in the lattice according to the equation (2.12). Referring to the 4×4 lattice in Figure 4.1, the distance matrix is:

$$\mathbf{D}_2 = \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_1^1 & \mathbf{D}_1^2 & \mathbf{D}_1^3 \\ \mathbf{D}_1^1 & \mathbf{D}_1 & \mathbf{D}_1^1 & \mathbf{D}_1^2 \\ \mathbf{D}_1^2 & \mathbf{D}_1^1 & \mathbf{D}_1 & \mathbf{D}_1^1 \\ \mathbf{D}_1^3 & \mathbf{D}_1^2 & \mathbf{D}_1^1 & \mathbf{D}_1 \end{pmatrix}, \text{ where: } \mathbf{D}_1 = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{pmatrix} \text{ and } \mathbf{D}_1^j = \mathbf{D}_1 + \begin{pmatrix} j & j & j & j \\ j & j & j & j \\ j & j & j & j \\ j & j & j & j \end{pmatrix}, \text{ with } j = 1, 2, 3.$$

#3 Implementation of the Kronecker product between any pair of matrices, so the computing of the covariance matrix (2.13) between any pair of points of the lattice $\{1, ..., l\}^d$ is available. Referring to the

same example, we have: $\mathbf{\Gamma}_1 = \begin{pmatrix} 1 & t & t^2 & t^3 \\ t & 1 & t & t^2 \\ t^2 & t & 1 & t \\ t^3 & t^2 & t & 1 \end{pmatrix}$ and $\mathbf{\Gamma}_2 = \begin{pmatrix} \mathbf{\Gamma}_1 & t\mathbf{\Gamma}_1 & t^2\mathbf{\Gamma}_1 & t^3\mathbf{\Gamma}_1 \\ t\mathbf{\Gamma}_1 & \mathbf{\Gamma}_1 & t\mathbf{\Gamma}_1 & t^2\mathbf{\Gamma}_1 \\ t^2\mathbf{\Gamma}_1 & t\mathbf{\Gamma}_1 & \mathbf{\Gamma}_1 & t\mathbf{\Gamma}_1 \\ t^3\mathbf{\Gamma}_1 & t^2\mathbf{\Gamma}_1 & t\mathbf{\Gamma}_1 & \mathbf{\Gamma}_1 \end{pmatrix}$

#4 Computation of the statistical index chosen for the comparison. In comparing different LH designs with the same number of training points, we resort to one of the most popular optimality criteria for the LH designs: the Total Mean Squared Prediction Error (TMSPE), i.e. the total of the MSPEs in the single design points. The results in this paper refer to this criterion.

#5 Since the number of the LH designs is very large, the LH designs are clustered according to the same value of the index in the previous step because of the symmetry properties. For example, the 6 LH designs with 2 factors and 3 levels have been clustered in 5 groups, the 36 LH designs with 3 factors and 3 levels into 10 groups, the 24 LH designs with 2 factors and 4 levels into 17 groups, the 576 LH designs with 3 factors and 4 levels into 110 groups, … and so on. The number of the groups overincreases: consider that for 2 factors with 6 levels there are 394 different TMSPE!

Table 4.1 – The 24 LH designs with two factors each one with four levels.

LH	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*
Training points	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

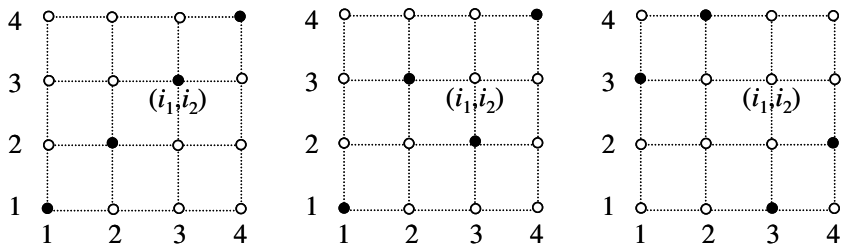


Figure 4.1 A representation of the bivariate lattice $\mathcal{X}_2$ and of three LH designs in the Table 4.1 (design in columns 1, 8 and 11). The four filled dots represent the training points, the remaining sixteen dashed the prediction ones.

In the next section, we discuss the behaviour of different LH designs in the two classes of the $(4!)^{2-1} = 24$ and $(4!)^{3-1} = 576$ LH designs according the TMSPE mentioned criterium.
Certainly, other criteria are available for the users: Entropy, the Minimax Distance and Maxmin Distance. The criterion based on the entropy, a very well known index both in statistics and in information theory, has been proposed by Shannon in 1948 and other practitioners (Shewry and Wynn, 1987; Currin et al., 1991; Mitchell and Scott, 19879 suggested its use by in selecting and finding designs. The maximum entropy criterion demands maximizing $\det\left(\sigma_Z^2\mathbf{R}\right)$ if $\boldsymbol{\beta}$ and the correlation (2.1) are known and $\det\left(\sigma_Z^2\hat{\mathbf{R}}\right)\times\det\left(\mathbf{\Sigma}_{\hat{\boldsymbol{\beta}}}\mathbf{F}'\left(\sigma_Z^2\hat{\mathbf{R}}\right)^{-1}\mathbf{F}+\mathrm{I}_p\right)$, with $\mathbf{\Sigma}_{\hat{\boldsymbol{\beta}}}$ covariance matrix of the generalized least squares estimator of $\boldsymbol{\beta}$ in the opposite case. The third criterion is the based on distances between points in the domain and training points in order to provide designs training points not so far from the point of the domain. We may have minimax design, i.e. designs that minimize the maximum distance $\min_{\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n\}}\max_{\mathbf{x}\in lattice} d\left(\mathbf{x},\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n\}\right)$ (4.5) or maxmin design, i.e. designs that maximize the minimum distance $\max_{\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n\}}\min_{\mathbf{u},\mathbf{v}\in\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n\}} d\left(\mathbf{u},\mathbf{v}\right)$. Their implementation is not difficult at all, since the program provide all the items for the computation of the indices.

5. Comparing indices for different LH design

In this section, we compare all the LH designs with d =2, 3 factors and l = 4 levels (the choice of these

LH designs is merely indicative of the comparisons that are available). For computing the predictor variances, we resort to a symbolic computation with multivariate polynomials, since the predictor variance is a rational function in the covariance parameter, with integer coefficients and it is suitable for symbolic exact computation. Note that the symbolic computation avoids any concern of numerical errors in limit situations, i.e. as $t \to 1$, and allows for an exact classification of the different behaviours.
The plots in Fig. 5.1 show the TMSPEs of all the mentioned LH designs: the TMSPEs are plotted vs $t = \exp(-\theta)$. The dashed lines in both pictures represent the TMSPE of the LHs whose points lie on the diagonals. These design are considered not properly space-filling, even if have nice marginal properties. And the largest value of the respective TMSPEs confirm their poor capability in prediction. Other remarkable features are the minimum value of the TMSPEs corresponding to maximum correlation and the maximum value of the TMSPEs corresponding to the null correlation 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 influence more the prediction because of the existing correlation. In Fig. 5.2, we resort to a different representation of the comparison index relating the TMSPEs to the worst TMSPE for enlightening the differences: the ratios between the TMSPE of the diagonal LH designs and the ones corresponding to the single groups of LH designs are plotted vs $t = \exp(-\theta)$. And indeed the differences are relevant! Depending on the number of factors levels, the best solution for the training points design may have a TMSPE that is approximately the 60% less than the TMSPE of the diagonal LH designs.

6. Conclusions

The present paper considers a special case of ordinary kriging which is suitable for symbolic computations. We assume a discrete lattice of sampling locations and covariance depending of the L^1 distance. Symbolic computations are done with the academic CoCoA system, but the same results could be obtained with commercial systems such as Maple or Matematica. Closed form of the kriging variance and other performance indices can be computed in order to gain a better understanding of general features of the design problem, i.e. the optimization of the training set. We discuss here a specific example of training set, i.e. LH designs. Given the number of levels and the number of factors, all LH designs are generated and for each of them a relevant performance index, e.g. Total Mean Squared Prediction Error is computed in the form of a rational function of the covariance parameter. This procedures allows a neat classification of various levels of performance and, consequently, a clear discussion of the performance of each LH design. This can lead to a sound rational for the choice of good training sets. The symbolic computation procedures are computationally complex with respect to the standard numerical procedures; therefore our methodology must be considered of exploratory nature.

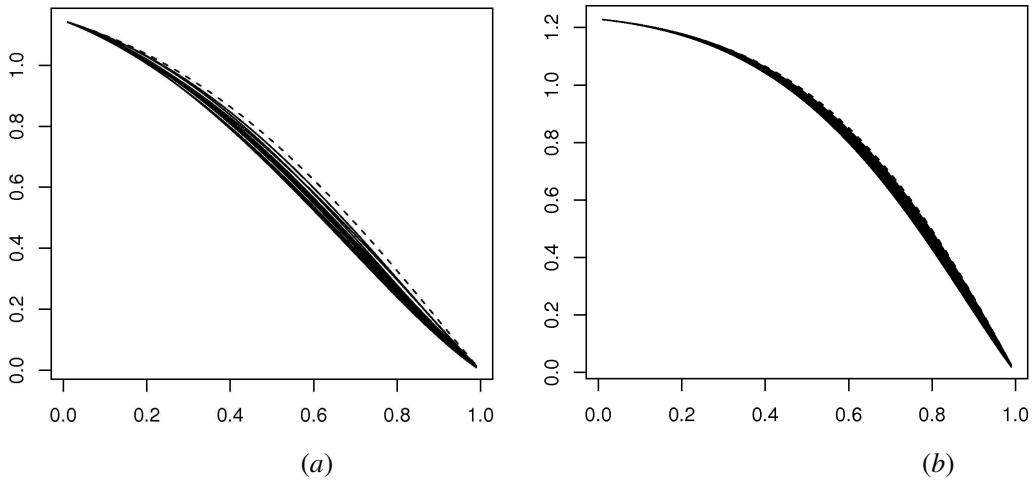
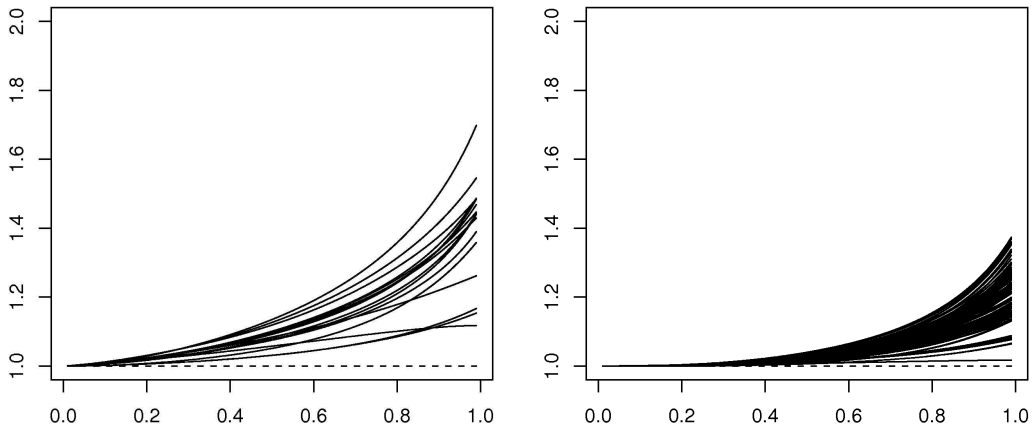


Figure 5.1 The TMSPE of all the LH designs with two (*a*) and three (*b*) factors, both with four levels, are plotted vs $t = \exp(-\theta)$. The dashed lines in both the pictures represent the TMSPE of the LHs whose points lie on the diagonals; the filled lines correspond to the 16 (a) and 109 (b) different TMSPEs of the remaining LH designs.



(*a*) (*b*)

Figure 5.2 The ratios between the Total Mean Square Prediction Error of the diagonal LH design and the one of all the LH designs with two (*a*) and three (*b*) factors, both having four levels are plotted vs $t = \exp(-\theta)$. The dashed lines in both the pictures represent the ratio corresponding to the LHs whose points lie on the diagonals.

Acknowledgement: Giovanni Pistone has been supported by DIMAT, Politecnico di Torino, and by SAMSI, RCP NC, USA.

References

Berg, C., Christensen, J.P.R, Ressel, P. (1984). Harmonic analysis on semigroups: theory of positive definite and related functions. Springer, New York.

Butler, N.A. (2001). Optimal and Orthogonal Latin Hypercube Designs for Computer Experiments. *Biometrika* **88**: 847-857.

CoCoATeam, CoCoA: a system for doing computations in Commutative Algebra, Version 4.7.4 2008 Available at http://cocoa.dima.unige.it.

Cressie, N. (1993). *Statistics for spatial data*, Wiley, New York.

Currin, C., Mitchell T.J., Morris, M.D. and Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association* **86:** 953-963.

Fang, K.-T., Li R., Sudjianto, A. (2006). Design and Modelling for Computer Experiments Chapman&Hall Boca Raton, FL.

Fang, K.-T., Lin D.K., Winker, P., Zhang, Y. (2000). Uniform Design: Theory and Applications. *Technometrics* **42**: 237–248.

Krige, D.G. (1951). A statistical approach to some mine valuations and allied problems at the Witwatersrand. Master's thesis of the University of Witwatersrand.

McKay, M.D. Conover, W.J., and Beckman, R.J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code". *Technometrics* **21**: 239–245.

Mitchell T.J. and Scott D.S. (1987). A computer program for the design of group testing experiments. *Communication in Statistics – Theory and Methods*, **16**: 2943-2955.

Park, J.-S. (1994) Optimal latin-hypercube designs for computer experiments. *J. Statist. Plann. Inference* **43:** 381-402.

Pistone G., Vicario G. (2009). *Design for Computer Experiments: comparing and generating designs in Kriging models*, In: Erto P., Statistics for Innovation. (pp. 91-102). Springer-Verlag Italia. ISBN: 978-88-470-0815.

Sachs J., Welch W.J., Mitchell T.J., Wynn H.P. (1989). Design and analysis of computer experiments. *Statistical Science*, **4**: 409-423.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003). The Design and Analysis of Computer Experiments. Springer-Verlag, New York.

Sasena, M. J. (2002). Flexibility and Efficiency Enhacements for Costrained Global Design Optinmization with Kriging Approximations, PhD Thesis University of Michigan.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**: 379-423.
Shewry M.C. and Wynn H.P. (1987). Maximum entropy sampling. *Journal of Applied Statistics*, **14**: 165-170.
Ye, Q., Li, W. and Sudjianto, A. (2000). Algorithmic construction of optimal symmetric latin hypercube designs, *J. Statist. Plann. Inference*, **90**: 145-159
Welch W. J., Buck, R.J., Sacks J., Wynn H. P., Mitchell T. J. and Morris, M.D. (1992). Sreening, predicting, computer experiments. *Technometrics* **34**: 15–25.