Numerical studies of space filling designs: optimization algorithms and subprojection properties

Bertrand Iooss

with Guillaume Damblin & Mathieu Couplet

CEMRACS 2013

July, 30th, 2013

CHANGER L'ÉNERGIE ENSEMBLE

Motivating example: Uncertainties management in simulation of thermal-hydraulic accident

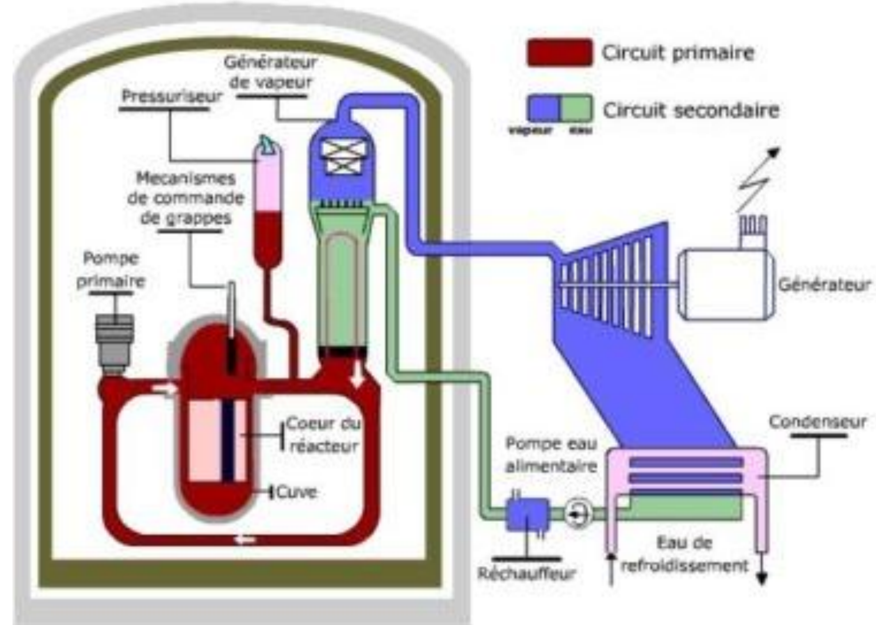
<u>Scenario</u> : Loss of primary coolant accident due to a large break in cold leg
[De Crecy et al., NED, 2008]

<u>p ~ 10-50 input random variables $\mathbf{X}$:</u>
geometry, material properties, environmental conditions, …

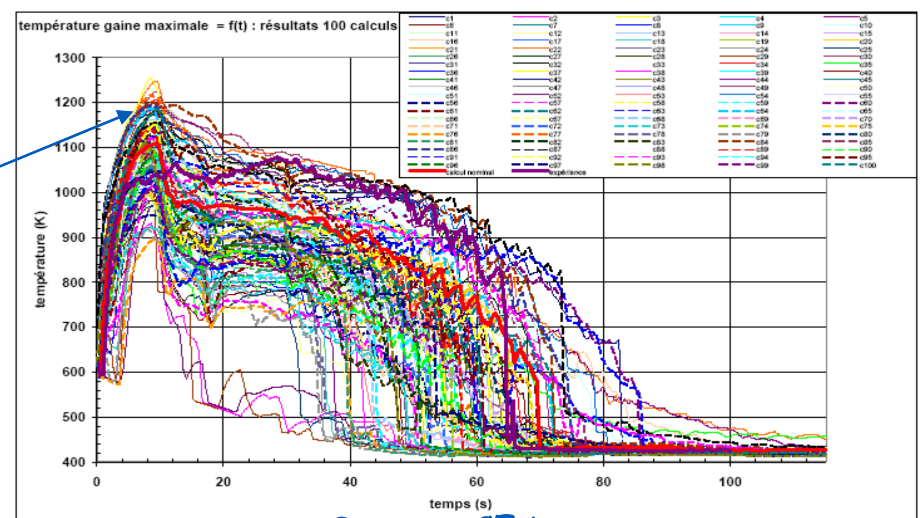
Computer code $Y = f(\mathbf{X})$
Time cost ~ 1-10 h - N ~ 100 - 500

<u>Interest output variable Y :</u>
Peak of cladding temperature

<u>Goal:</u> numerical model exploration via space filling design, then metamodel



Pressurized water nuclear reactor



Source: CEA

EDF

Model exploration goal

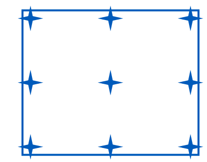
GOAL : explore as best as possible the behaviour of the code

Put some points in the whole input space in order to « maximize » the amount of information on the model output

Contrary to an uncertainty propagation step, it depends on *p*

Regular mesh with n levels → $N = n^p$ simulations

Ex: p =2, n =3 → N =9

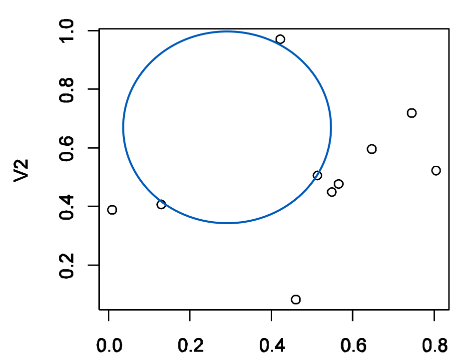


p = 10, n=3 → N = 59049

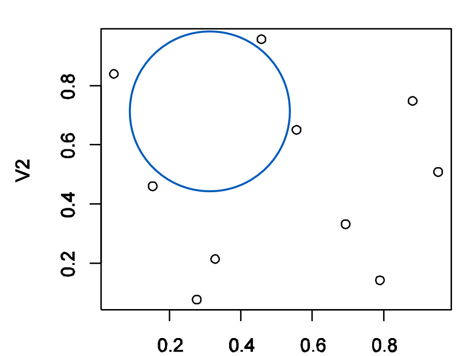
To minimize N, needs to have some techniques ensuring good « coverage » of the input space

Simple random sampling (Monte Carlo) does not ensure this

Ex: p = 2
N = 10



Monte Carlo



Optimized design

Objectives

When the objectives is to discover what happens inside a numerical model (e.g. non linearities of the model output), we want to build the design $\Xi^N = \left(x_j^{(i)}\right)_{i=1\ldots N,\, j=1\ldots p}$ while respecting the constraints:

1. To « regularly » spread the *N* points over the *p*-dimensional input space χ

2. To ensure that this input space coverage is robust with respect to dimension reduction (because most of the times, only a small number of inputs are influent ⇔ low effective dimension)

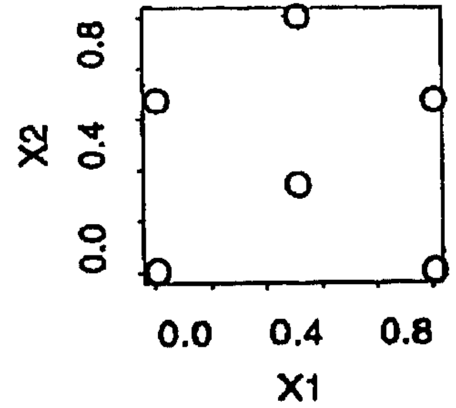
Therefore, we look for some design which insures the « best coverage » of the input space (and its sub-projections)

➡ The class of Space filling Design (SFD) is adequate. It can be:

- Based on an inter-point distance criterion (minimax, **maximin**, …)

- Based on a criterion of uniform distribution of the points (entropy, various discrepancy measures, **L^2 discrepancies**, …)

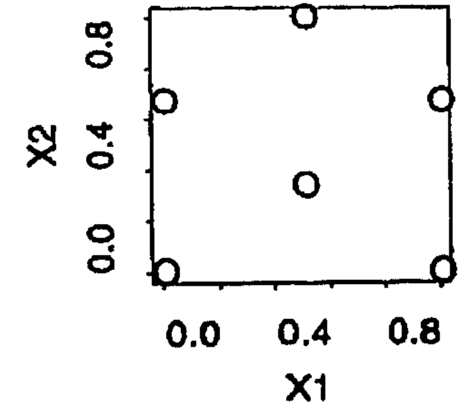
1. Two classical space filling criteria



- **Mindist distance:** $$\phi(\Xi^N) = \min_{x^{(1)}, x^{(2)} \in \Xi^N} d(x^{(1)}, x^{(2)})$$

➡ Maximin design Ξ^N_{Mm} : $$\max_{\Xi^N} \min_{x^{(1)}, x^{(2)} \in \Xi^N} d(x^{(1)}, x^{(2)})$$

1. Two classical space filling criteria

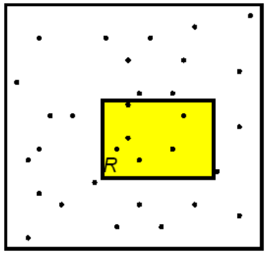


- **Mindist distance:** $$\phi(\Xi^N) = \min_{x^{(1)},x^{(2)} \in \Xi^N} d(x^{(1)}, x^{(2)})$$

Maximin design Ξ^N_{Mm} : $$\max_{\Xi^N} \min_{x^{(1)},x^{(2)} \in \Xi^N} d(x^{(1)}, x^{(2)})$$

- **Discrepancy measure:** Deviation of the sample points distribution from the uniformity

$$D^*\left(\Xi^N\right) = \sup_{\mathbf{t} \in [0,1[^p} \left| \frac{1}{N} \sum_{i=1}^{N} 1_{\mathbf{x}^{(i)} \in Q(\mathbf{t})} - \mathrm{Volume}(Q(\mathbf{t})) \right|$$



L^2 discrepancy allows to obtain analytical formulas

$$D_2^*\left(\Xi^N\right) = \left[\int_{[0,1[^p} \left[\frac{1}{N} \sum_{i=1}^{N} 1_{\mathbf{x}^{(i)} \in Q(\mathbf{t})} - \mathrm{Volume}(Q(\mathbf{t})) \right]^2 d\mathbf{t} \right]^{1/2}$$

Example of discrepancy

Various analytical formulations while considering L² discrepancy and different kind of intervals

[Hickernell 1998]

Modified L_2 discrepancy allows to take into account points uniformity on subspaces of $[0,1[^p$

$$D_2\left(\Xi^N\right) = \left[\sum_{u \neq \emptyset} \int_{C^u} \left[\frac{1}{N} \sum_{i=1}^{N} 1_{\mathbf{x}_u^{(i)} \in Q_u(\mathbf{t})} - \mathrm{Volume}(Q_u(\mathbf{t})) \right]^2 d\mathbf{t} \right]$$

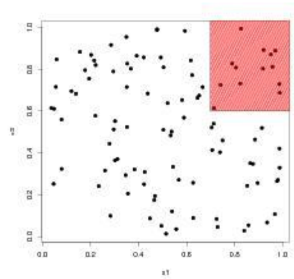
$$\text{with } u \subset \{1, \dots, p\}$$

$$\text{and } Q_u(\mathbf{t}) \text{ the projection of } Q(\mathbf{t}) \text{ on } C^u \text{ (cube unity of coordinates in } u\text{)}$$

Centered L_2-discrepancy (intervals with boundary one vertex of the unit cube)

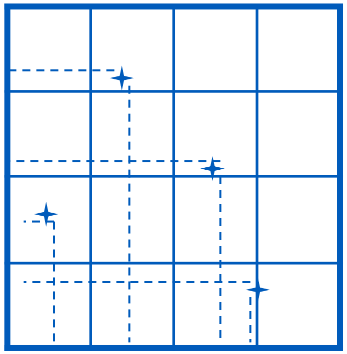
$$C^2(\Xi^N) = \left(\frac{13}{12}\right)^p - \frac{2}{N} \sum_{i=1}^{N} \prod_{k=1}^{p} \left(1 + \frac{1}{2}\left| x_k^{(i)} - \frac{1}{2} \right| - \frac{1}{2}\left| x_k^{(i)} - \frac{1}{2} \right|^2 \right)$$

$$+ \frac{1}{N^2} \sum_{i,j=1}^{N} \prod_{k=1}^{p} \left(1 + \frac{1}{2}\left| x_k^{(i)} - \frac{1}{2} \right| + \frac{1}{2}\left| x_k^{(j)} - \frac{1}{2} \right| - \frac{1}{2}\left| x_k^{(i)} - x_k^{(j)} \right| \right)$$



2. Unidim.-projection robustness via Latin Hypercube Sample

Class of LHS ensures uniform projection on margins



Ex: $p=2$, $N=4$

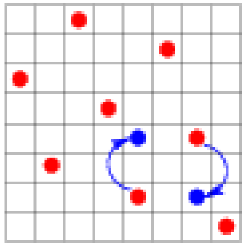
LHS(p,N): - Divide each dimension in N intervals
- Take one point in each stratum
- Random LHS: perturb each point in each stratum

Finding an optimal (SFD) LHS:
impossible exhaustive exploration: $(N!)^p$ different LHS

Methods via optimization algo (ex: minimization of $\phi(.)$ via simulated annealing) :

1. Initialisation of a design Ξ (LHS initial) and a temperature T

2. While $T > 0$:
   1. Produce a neighbor Ξ_{new} of Ξ (permutation of 2 components in a column)
   2. replace Ξ by Ξ_{new} with proba $\min\left(\exp\left[-\frac{\phi(\Xi_{new})-\phi(\Xi)}{T}\right],1\right)$
   3. decrease T



3. Stop criterion => Ξ is the optimal solution

[Park 1993;
Morris & Mitchell 1995]

LHS maximin: regularization of the criterion

- **Mindist criterion :** (to be maximized)

$$\phi(\Xi^N) = \min_{x^{(1)},x^{(2)} \in \Xi^N} d(x^{(1)}, x^{(2)})$$

- **Regularized mindist criterion :** (to be minimized) *[Morris & Mitchell 95]*

$$\phi_q(\Xi^N) = \left[\sum_{i,j=1, i<j}^{N} d(x^{(i)}, x^{(j)})^{-q} \right]^{1/q}$$

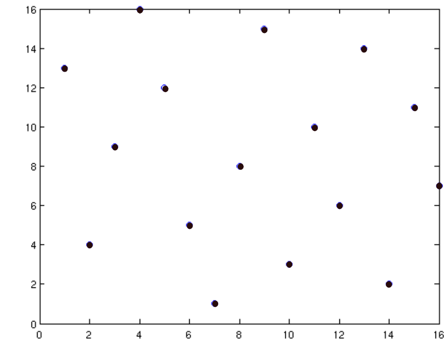
These 2 criteria are equivalent for the optimization when $q \to \infty$

[Pronzato & Müller12]

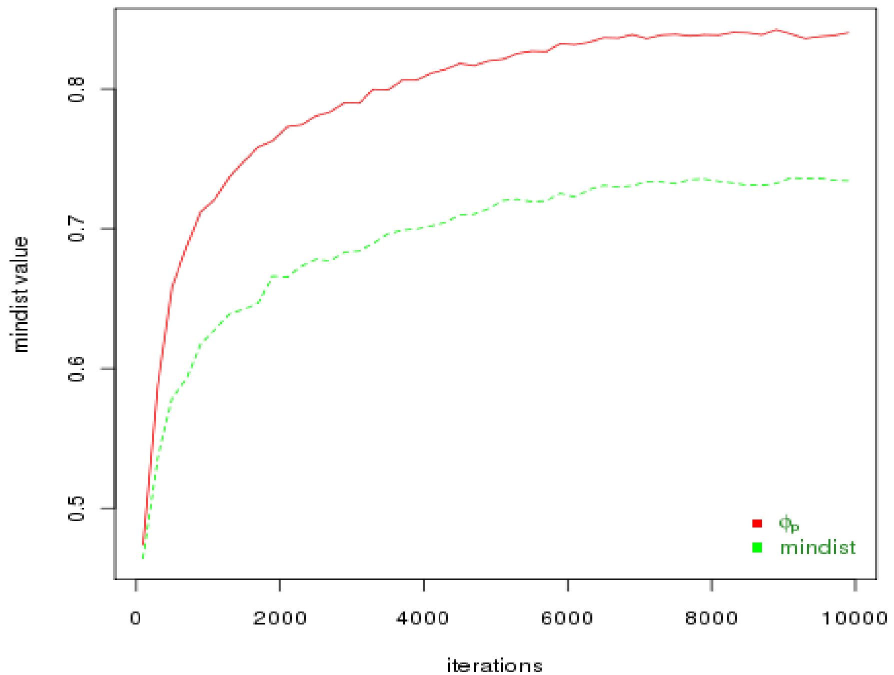
ϕ_q is easier to optimize than mindist

In practice, we take $q = 50$

Example : Maximin LHS(2,16)



Numerical test: N = 100, p = 10



Updating criteria after a LHS perturbation

[Jin et al. 2005]

Between Ξ and Ξ', 2 point coordinates $x^{(i_1)}$ and $x^{(i_2)}$ are modified

- Regularized mindist criterion $\phi_q(\Xi) = \left[\sum_{i,j=1, i<j}^{N} d(x^{(i)}, x^{(j)})^{-q} \right]^{1/q}$
  ($N(N-1)/2$ distances)

$\Rightarrow$ Only recalculate the $2(N-2)$ distances of these 2 points to other points

- L² discrepancy criteria (cost in $O(pN^2)$)

$$C^2(\Xi) = \left(\frac{13}{12}\right)^p - \frac{2}{N}\sum_{i=1}^{N}\prod_{k=1}^{p}\left(1+\frac{1}{2}\left|x_k^{(i)}-\frac{1}{2}\right|-\frac{1}{2}\left|x_k^{(i)}-\frac{1}{2}\right|^2\right) + \frac{1}{N^2}\sum_{i,j=1}^{N}\prod_{k=1}^{p}\left(1+\frac{1}{2}\left|x_k^{(i)}-\frac{1}{2}\right|+\frac{1}{2}\left|x_k^{(j)}-\frac{1}{2}\right|-\frac{1}{2}\left|x_k^{(i)}-x_k^{(j)}\right|\right)$$

$$C^2(\Xi) = \left(\frac{13}{12}\right)^p + \sum_{i,j=1}^{N} c_{ij} \quad ; \quad C^2(\Xi') = \left(\frac{13}{12}\right)^p + \sum_{i,j=1}^{N} c'_{ij}$$

If $i, j \neq i_1$ and $i, j \neq i_2$ then $c'_{ij} = c_{ij}$

$$C^2(\Xi') = C^2(\Xi) + c'_{i_1 i_1} - c_{i_1 i_1} + c'_{i_2 i_2} - c_{i_2 i_2} + 2 \sum_{j=1, j\neq i_1, j\neq i_2}^{N} (c'_{i_1 j} - c_{i_1 j} + c'_{i_2 j} - c_{i_2 j})$$

Cost in $O(pN)$

Two different optimization algorithms

1 Morris & Mitchell Simulated Annealing (MMSA) *[Morris & Mitchell 1995]*

Linear profile for the temperature decrease
(geometrical alternative: $T_i = c^i \times T_0$)

Temperature decreases when *B* new LHS do not improve the criterion

Slow convergence but large exploration space

2 Enhanced Stochastic Evolutionary (ESE) *[Jin et al. 2005]*

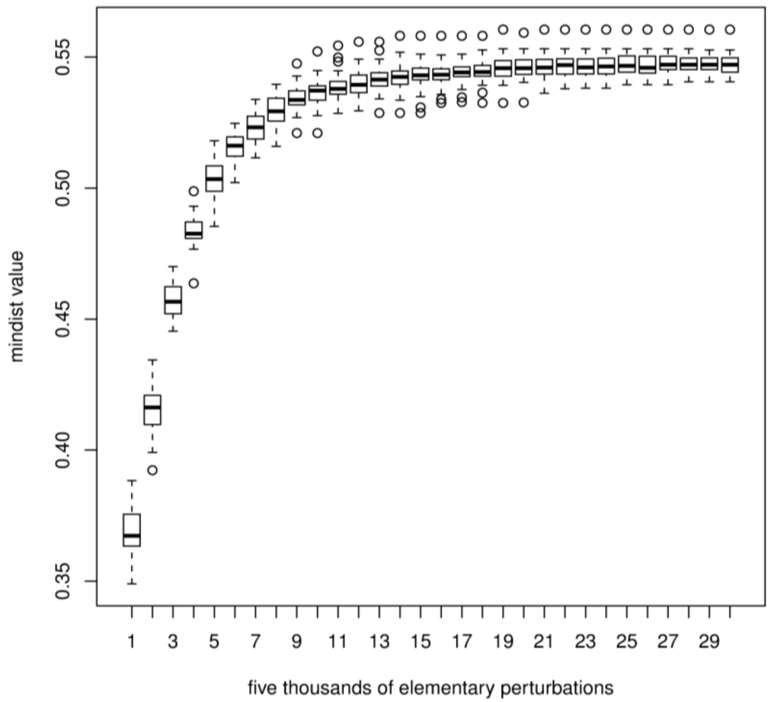
Inner loop (I iterations):
    Proposition of *M* new perturbed LHS at each step

Outer loop to manage the temperature (can decrease or increase)

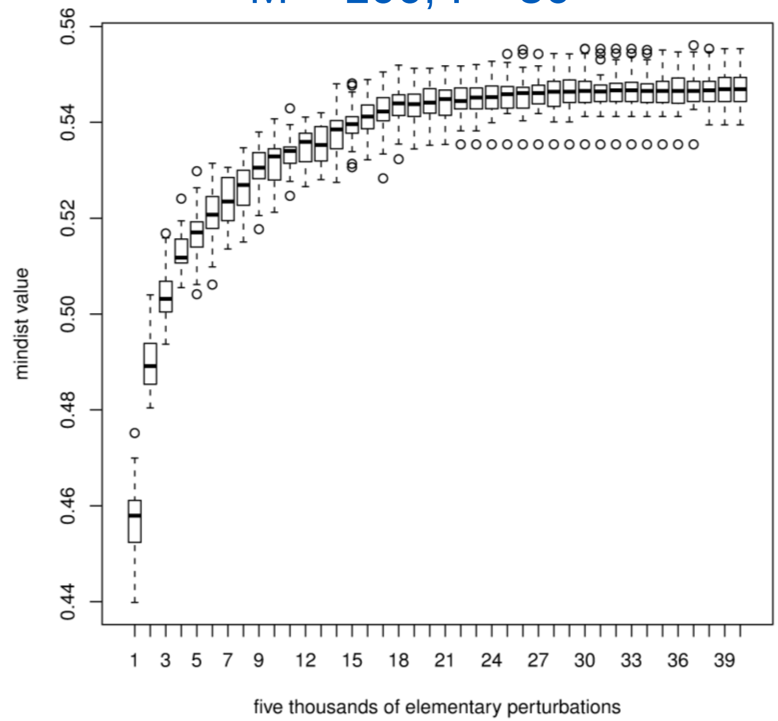
Comparison of optimization algorithms convergence

Numerical tests: $N = 50$, $p = 5$

MMSA - linear profile
$T_0 = 0.1$, B = 300, c = 0.9



ESE
M = 100, I = 50



Both algorithms converge slowly to the same value, after the same iteration numbers

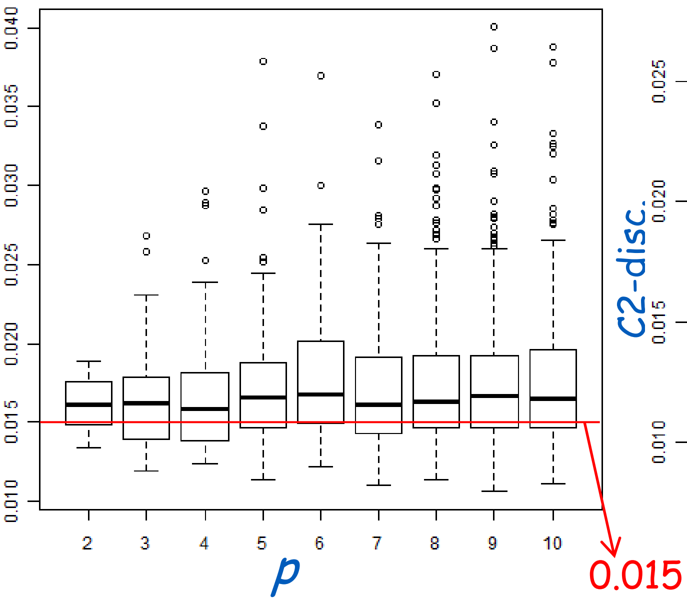
ESE shows a faster convergence at the first iterations than MMSA

It is possible to improve this result, but at a prohibitive cost
(MMSA: T_0=0.01, B=1000, c=0.98; ESE: M=300)

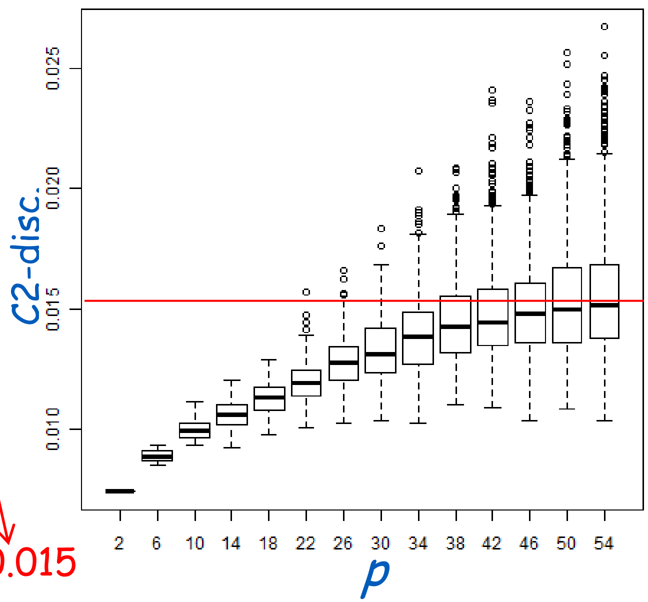
Robustness tests in 2D subprojections of optimal LHS (1/3)

3 types of LHS (n = 100) with increasing p ; *10 replicates for each dimension*
All 2D subprojections are taken into account

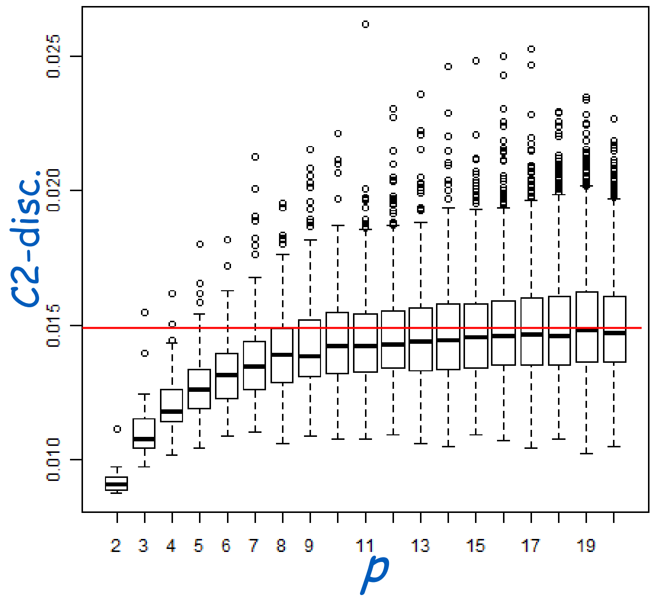
Standard LHS
(reference)



Low C2-discrepancy LHS
(C2 = L^2-centered)



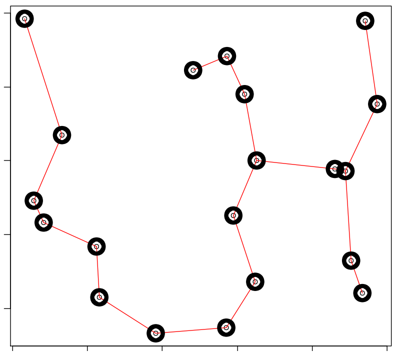
Maximin LHS



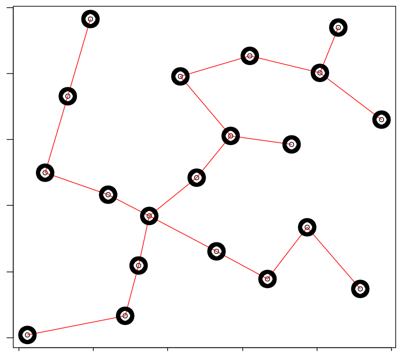
From dimension p=10, the maximin LHS behaves like a standard LHS
From dimension p=40, the low C2-discrepancy LHS behaves like a standard LHS
Another test for the low **L^2-star discrepancy**: convergence for p=10

It confirms the relevance of C2-discrepancy criterion in terms of subprojections

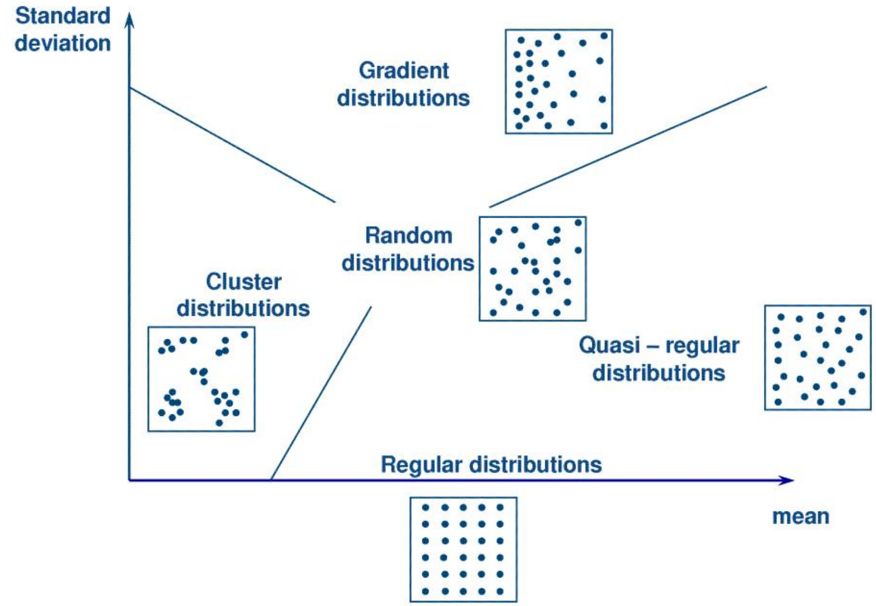
Another space-filling criteria based on Minimal Spanning Tree



MST for
random design

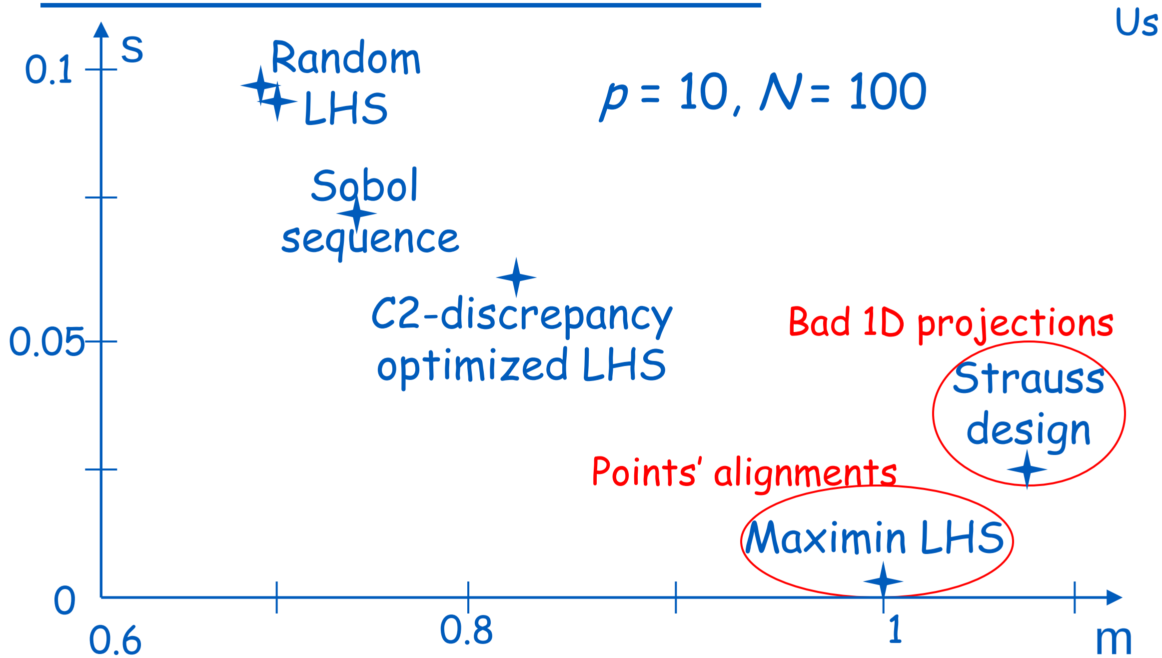


MST for
maximin LHS



Using the Minimal Spanning Tree (MST)
[Franco et al., Chem. Lab., 2009]

Numerical tests on various SFD:



Conclusion

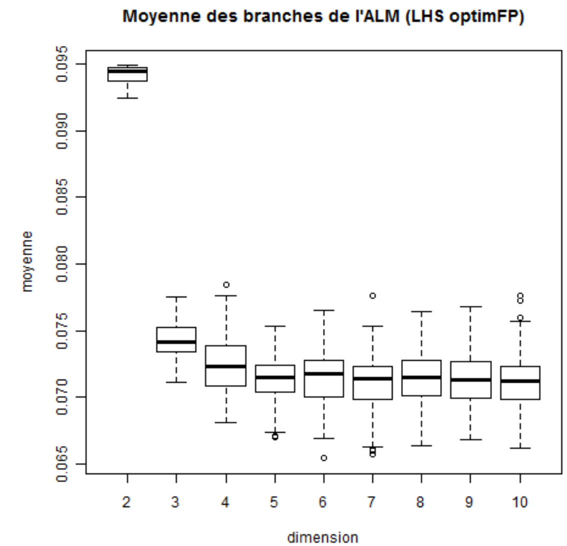
This MST-based graph is a tool to compare designs in terms of regularity in the p-dimensional space

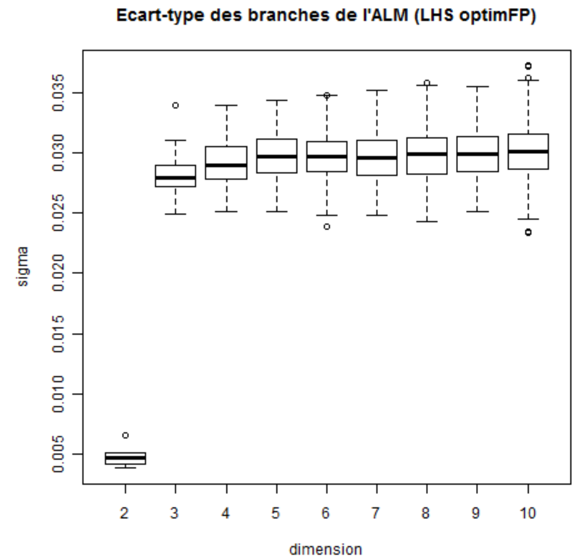
Complementarity with mindist

Robustness tests in 2D subprojections of optimal LHS (3/3)
MST criteria

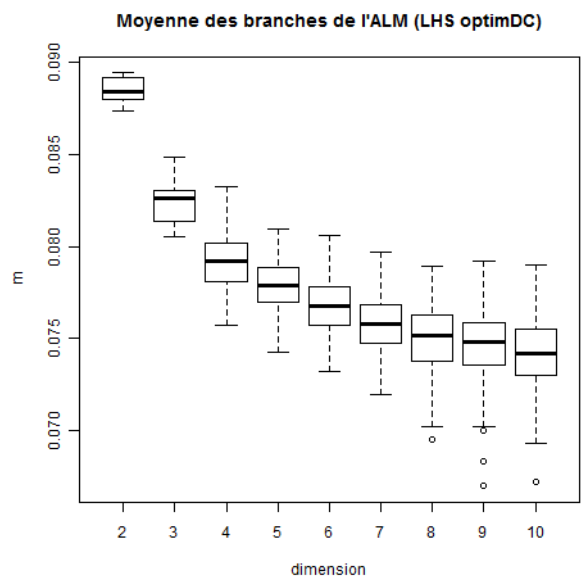
$N = 100$

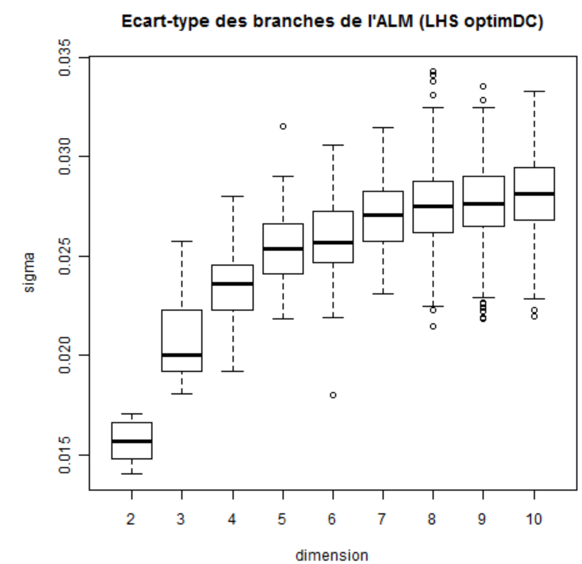
Maximin LHS



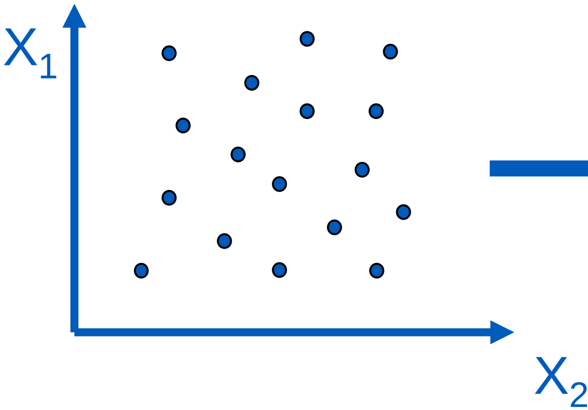


Low C2-discrepancy LHS

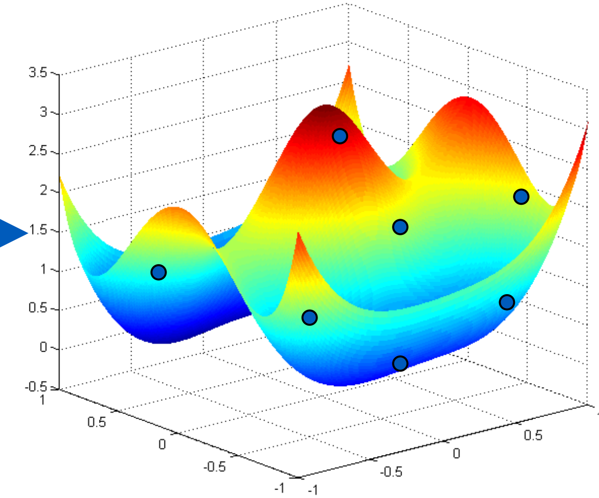




Example: fitting a kriging metamodel



Computer code

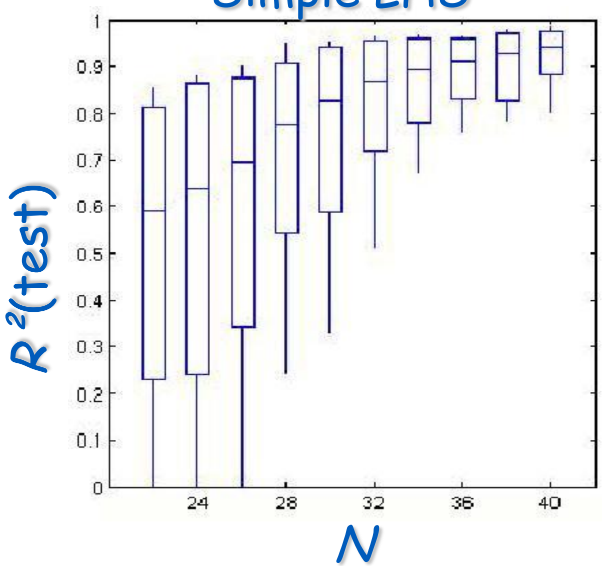


Non monotonic test function (p = 5) : g-function of Sobol

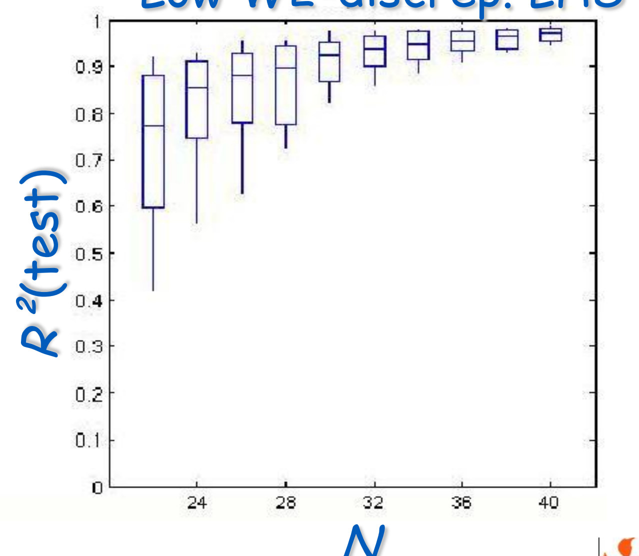
$$g(X_1,\dots,X_5)=\prod_{i=1}^{5}\frac{|4X_i-2|+a_i}{1+a_i}\quad \text{avec } a_i=i \text{ et } X_i\sim U_{[0;1]} \text{ pour } \ i=1\dots 5$$

Metamodel (kriging) is built on a learning sample of sizes N = 22,…,40

Simple LHS



Low W2-discrep. LHS



[Marrel 2008]

Conclusions

1 SFD are useful in an initial exploration step, small *N*, large *p*

2 Algorithms for LHS optimization: ESE seems preferable (faster convergence)
Tuning parameters are difficult to fit; some recommendations are made in refs.

3 Modified L^2 discrepancies take into account uniformity of the point projections on lower-dimensional subspaces of $[0,1[^p$

In our tests, low L^2-centered discrepancy LHS have shown the best space filling robustness on the projections over 2D subspaces (same effects on 3D subprojections)

⟹ Important property for metamodel fitting and sensitivity indices computation

3 Distance-based designs show stronger space filling regularity but no 2D robustness

Challenge: Building good & robust SFD outside the LHS class

Bibliographie

- **G. Damblin, M. Couplet & B. Iooss, Numerical studies of space filling designs: optimization algorithms and subprojection properties, *submitted***
- **Package in R software: DiceDesign (D. Dupuy, C. Helbert, J. Franco, O. Roustant, G. Damblin, B. Iooss)**
- K-T. Fang, R. Li & A. Sudjianto, *Design and modeling for computer experiments*, Chapman & Hall, 2006
- F.J. Hickernell. A generalized discrepancy and quadrature error bound. *Mathematics of Computation,* 67:299-322, 1998.
- B. Iooss, L. Boussouf, V. Feuillard & A. Marrel. Numerical studies of the metamodel fitting and validation processes. *International Journal of Advances in Systems and Measurements,* 3:11-21, 2010.
- R. Jin, W. Chen & A. Sudjianto. An efficient algorithm for constructing optimal design of computer experiments. *Journal of Statistical Planning and Inference*, 134:268-287, 2005.
- M.E. Johnson, L.M. Moore & D. Ylvisaker. Minimax and maximin distance design. *Journal of Statistical Planning and Inference,* 26:131-148, 1990.
- M. Morris & T. Mitchell. Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference,* 43:381-402, 1995.
- J-S. Park. Optimal Latin-hypercube designs for computer experiments. *Journal of Statistical Planning and Inference*, 39:95-111, 1994.
- L. Pronzato & W. Müller. Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22:681-701, 2012.

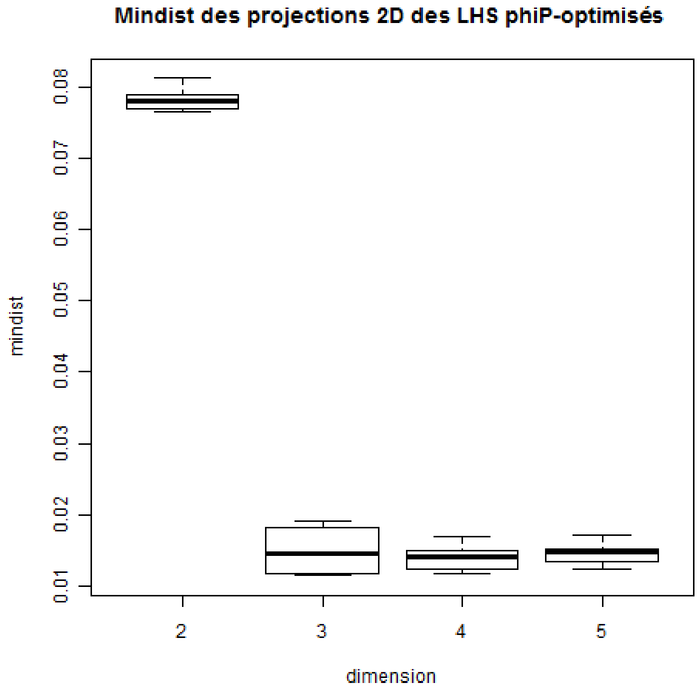
Annexes

Robustness tests in 2D subprojections of optimal LHS (2/3)

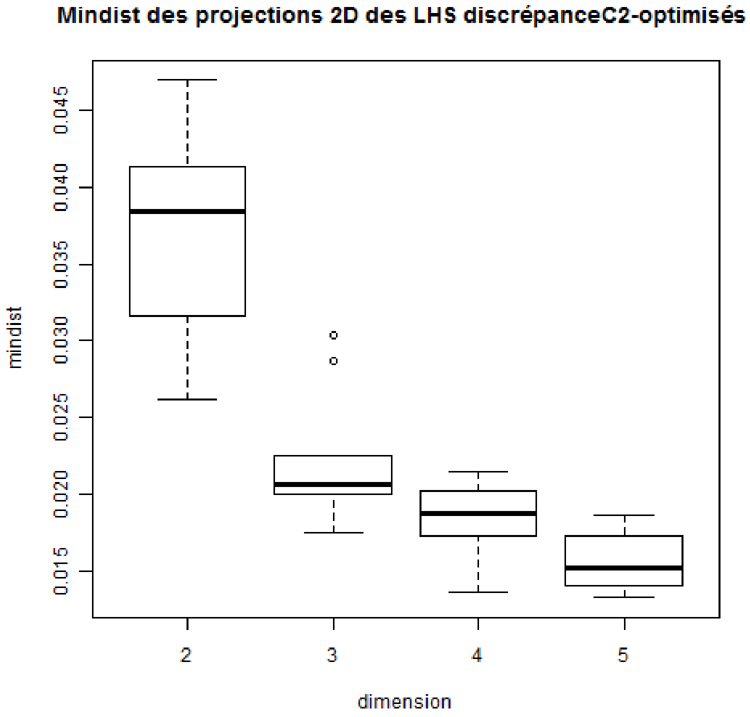
2 types of LHS ($n = 100$) with increasing p ; *10 replicates for each dimension*

All 2D subprojections are taken into account

Maximin LHS



Low C2-discrepancy LHS



It confirms the non-relevance of mindist distance in terms of subprojections