

# Planification d'expériences numériques à partir du processus ponctuel de Strauss

Jessica Franco, Xavier Bay, Delphine Dupuy, Bernard Corre

► **To cite this version:**

Jessica Franco, Xavier Bay, Delphine Dupuy, Bernard Corre. Planification d'expériences numériques à partir du processus ponctuel de Strauss. 2008. <hal-00260701>

**HAL Id: hal-00260701**

**<https://hal.archives-ouvertes.fr/hal-00260701>**

Submitted on 4 Mar 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# PLANIFICATION D'EXPÉRIENCES NUMÉRIQUES A PARTIR DU PROCESSUS PONCTUEL DE STRAUSS

**Jessica Franco<sup>1,2</sup>, Xavier Bay<sup>1</sup>, Delphine Dupuy<sup>1</sup>, Bernard Corre<sup>3</sup>**

<sup>1</sup>Département 3MI, Ecole Nationale Supérieure des Mines, Saint-Etienne, [nom@emse.fr](mailto:nom@emse.fr)

<sup>2</sup>Total E&P, EP/GSR/TG/G&I, Pau, [jessica.franco@total.com](mailto:jessica.franco@total.com)

<sup>3</sup>Total E&P, EP/GSR/COP/EM, Pau, [bernard.corre@total.com](mailto:bernard.corre@total.com)

---

## Résumé.

Dans le contexte de la planification d'expériences numériques où la relation précise entre les entrées et les sorties du code est inconnue, il est préférable d'utiliser des « Space-Filling Designs » (SFDs).

Dans cet article, nous présentons un nouveau type de SFDs basé sur le processus ponctuel de Strauss. Ces plans obtenus à partir de méthodes Monte-Carlo par chaîne de Markov (McMC) sont facilement modulables et peuvent ainsi répondre à divers objectifs. Signalons en particulier les bonnes propriétés de projection des expériences sur les axes factoriels à l'image des hypercubes latins. Il est possible également de générer des points de manière hétérogène (préférentielle) si l'on dispose d'informations *a priori*.

## Mots clefs.

Plans d'expériences numériques, Space-filling designs, Processus ponctuel, Processus de Strauss, Processus de Gibbs, Méthodes Monte-Carlo par chaîne de Markov.

---

## Abstract.

In the computer experiments setting, Space-Filling Designs (SFDs) are widely used to explore the complex relationship between inputs and outputs.

In this paper, a new SFD is initially defined with the help of the Strauss process. Through Markov chain Monte-Carlo (McMC) methods, more general Gibbs processes can be used to perform different goals. We will see that it is easy to sample over the entire range of each input variable as Latin hypercubes do it. Moreover, non-homogeneous designs can be constructed to take account of *a priori* information.

## Keywords.

Computer experiments, Space-filling designs, Point process, Strauss process, Gibbs process, Markov chain Monte-Carlo.

---

# 1. Introduction

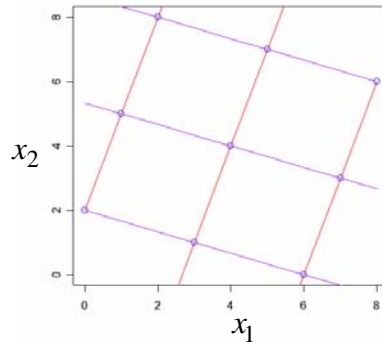
---

Depuis une quinzaine d'années, la théorie des plans d'expériences initiée par Fisher [5] connaît un renouveau avec son utilisation pour l'étude des grands codes de calcul industriels. Ce changement de contexte a induit au moins deux changements majeurs. D'abord, les grands codes de calcul décrivent des phénomènes d'une complexité de plus en plus importante, ce qui implique que les modèles correspondants sont souvent non linéaires. Ensuite, la nature de l'expérience est différente. Contrairement aux expériences physiques, les expériences numériques sont des simulations et fournissent (sauf si le code est stochastique) la même réponse pour des conditions expérimentales identiques (incluant les paramètres algorithmiques et informatiques). La répétition d'une expérience, qui permet usuellement d'apprendre la variabilité des résidus, n'apporte ici aucune information.

Dans ce nouveau contexte, les méthodes de planification d'expériences sont donc différentes. Par exemple lorsque l'objectif est de modéliser le code en phase exploratoire (quand aucune simulation n'a encore été réalisée), on recherche souvent à satisfaire les deux contraintes suivantes. D'une part, répartir les points dans l'espace le plus possible de façon à capter les non-linéarités. D'autre part, faire en sorte que ce remplissage de l'espace subsiste par réduction de la dimension.

La première contrainte est à l'origine des recherches sur les Space-Filling Designs (SFDs). La qualité de la répartition spatiale est mesurée soit à l'aide de critères déterministes comme les distances minimax ou maximin (Johnson et al. [11]), soit à l'aide de critères statistiques comme la discrédance (Niederreiter [16], Hickernell [9] et Fang, Li, Sudjianto [4]).

La deuxième contrainte vient du fait qu'il est fréquent d'observer que le code ne dépend que de quelques variables influentes ou de quelques combinaisons linéaires de ces variables. Il est donc souhaitable que la propriété de remplissage de l'espace soit encore satisfaite en projection sur des sous-espaces. Cette idée a motivé l'utilisation des hypercubes latins (Mc Kay [14]) et des tableaux orthogonaux (Owen [17]) dans le cadre des plans d'expériences numériques. La nature des expériences étant déterministe, la mauvaise répartition en projection se traduit par une perte d'information. Cependant, la seule considération des axes factoriels est insuffisante. Par exemple, l'hypercube latin ci-dessous est très mal adapté si le code est une fonction de  $x_1 + 3x_2$  (ou  $3x_1 - x_2$ ), puisque l'information donnée par le plan d'expériences se réduit à 3 points au lieu de 9.



**Figure 1. Un hypercube latin (plan factoriel orienté).**

Avec un hypercube latin, on peut garantir une bonne répartition des points en projection sur les marges (en plus de la non-répétition), ce qui permet de ne pas perdre d'information si le code ne dépend que d'une seule variable. Les tableaux orthogonaux généralisent cette propriété à la projection sur des sous-espaces marginaux de dimension supérieure (voir par exemple Koehler et Owen [12] et de façon plus générale Santner, Williams, Notz [19]). Cependant, ces plans ne sont pas toujours complètement satisfaisants au sens du remplissage de l'espace.

Dans cet article, une nouvelle famille de plans d'expériences sera présentée. Afin de répondre à la bonne propriété de remplissage, l'idée initiale a été d'utiliser un processus ponctuel de Strauss [22] qui rend compte de répulsion entre des points évoluant dans un domaine spatial. Par la suite, en s'appuyant sur la classe plus générale des processus de Gibbs (Cressie [1], Daley, Vere-Jones [2]) et leur description en termes de potentiels locaux d'interactions, nous proposerons des extensions visant à améliorer la qualité de ces plans au sens des propriétés énoncées ci-dessus. D'une part, une répartition uniforme sur les marges sera imposée en plus de la bonne distribution des points dans le domaine expérimental afin de retrouver les bonnes propriétés des hypercubes latins. D'autre part, les potentiels d'interactions seront également adaptés pour optimiser le remplissage de l'espace et par ce biais le critère de maximin (Johnson et al. [11]). Une application à une fonction déterministe explicite sera donnée. Enfin, nous terminerons par des conclusions et perspectives.

## 2. Processus ponctuels et planification d'expériences

---

Supposons que le code prenne en entrée  $d$  facteurs d'intérêts continus ( $d \geq 1$ ) et que l'hypercube unité  $\Omega = [0, 1]^d$  soit le domaine expérimental. L'objectif est de générer un plan  $\mathbf{x} = \{\mathbf{x}^1, \dots, \mathbf{x}^n\}$  constitué de  $n$  expériences  $\mathbf{x}^i \in [0, 1]^d$ , qui vérifie les propriétés vues en introduction.

L'idée principale est de voir chaque expérience comme un point ou une particule chargée de l'espace  $\Omega$ . On assimile ainsi les  $n$  expériences à la réalisation (conditionnelle au nombre  $n$  de points) d'un processus ponctuel (Daley et Vere-Jones [2]). De manière à bien occuper l'espace, il a été naturel de considérer le processus de Strauss [22] qui permet de rendre compte de répulsion entre particules. Ainsi, les  $n$  expériences sont vues comme la photographie à un instant donné de  $n$  particules de même charge électrique, en mouvement dans une boîte  $d$ -dimensionnelle (on néglige la gravité).

Pour générer un tel processus, nous utiliserons les techniques usuelles de simulation par chaîne de Markov (Meyn, Tweedie [14]) et précisément l'algorithme de Metropolis-Hastings (MH) (Metropolis et al. [15], Hastings [8]). C'est essentiellement la propriété de Markov spatiale du processus ponctuel de Strauss qui en pratique permet de simuler de manière efficace ces nouveaux plans d'expériences, notamment en grande dimension.

Pour l'étude de tels plans (aléatoires par nature), le calcul de la moyenne des critères d'uniformité usuels est envisagé. C'est la dynamique markovienne de l'algorithme de MH et ses bonnes propriétés de mélange ou d'ergodicité qui vont permettre ce calcul. Ainsi, on peut parler de planification d'expériences par méthodes de Monte-Carlo par chaîne de Markov (McMC).

### 2.1. Le processus de Strauss

On considère le processus de Strauss avec interaction de type répulsion dans le domaine  $\Omega = [0, 1]^d$  (Ripley, Kelly [18]). La loi  $\pi(\cdot)$  de ce processus conditionnellement au nombre  $n$  de points est donnée par :

$$\pi(\mathbf{x}) = k\gamma^{s(\mathbf{x})} \quad (1)$$

où  $0 < \gamma \leq 1$  est assimilé à un coefficient de répulsion,  $k$  une constante de normalisation et où la fonction  $s(\mathbf{x})$  assimilée à un potentiel global d'énergie est définie par :

$$s(\mathbf{x}) = s(\mathbf{x}^1, \dots, \mathbf{x}^n) = \sum_i \sum_{j>i} 1_{\|\mathbf{x}^i - \mathbf{x}^j\| \leq R}$$

Il s'agit simplement du nombre de paires d'expériences  $\{\mathbf{x}^i, \mathbf{x}^j\}$  telles que la distance entre  $\mathbf{x}^i$  et  $\mathbf{x}^j$  est inférieure ou égale à  $R$ .

En termes de particules chargées, c'est aussi le nombre global d'interactions limité uniquement aux interactions de paires (interactions dites d'ordre 2). Chaque particule a une sphère d'influence de rayon  $R/2$ , l'interaction se produisant exactement lorsque 2 sphères quelconques se rencontrent.

Le choix  $0 < \gamma < 1$  permet de donner une probabilité plus ou moins forte aux plans pour lesquels les expériences sont réparties de manière à interagir plus ou moins les unes par rapport aux autres. Le plan est d'autant plus vraisemblable que le nombre d'interactions  $s(x)$  est faible. Nous ne considérons pas ici le cas limite où  $\gamma = 0$ , qui interdirait toute paire de points de distance inférieure ou égale à  $R$ . Ce modèle, dit de sphères dures, conduit à des points régulièrement espacés et donc à des alignements. Le cas  $\gamma = 1$  correspond exactement à l'indépendance (distribution conjointe uniforme), et dans la version non conditionnelle, à un processus de Poisson d'intensité constante ou homogène (Daley [2]).

Dans toute la suite, on se limitera à des réalisations de processus ponctuels (Strauss, ...) conditionnellement à un nombre  $n$  de points spécifié à l'avance. Ainsi, dans la dynamique de simulation, ce nombre  $n$  n'aura pas la possibilité d'évoluer.

## 2.2. Algorithme de Metropolis-Hastings

Il n'existe pas d'expression analytique explicite pour la constante de normalisation  $k$  intervenant dans la densité du processus (1), même pour des tailles modestes de  $n$  et  $d$ . De la même manière, les méthodes directes de simulation sont difficiles à envisager compte tenu de la relative complexité de cette densité. L'alternative classique est de simuler une chaîne de Markov de loi asymptotique ou stationnaire  $\pi(\cdot)$  (Meyn, Tweedie [14]). Dans ce contexte, un état de la chaîne est un plan de  $n$  expériences (ensemble des positions des  $n$  points). La

souplesse d'une telle approche réside dans le fait qu'il est aisé en général de simuler une transition d'état à état de la chaîne. C'est le cas de l'algorithme de MH qui procède en 2 étapes élémentaires :

- une phase de proposition de changement d'état : le nouvel état ne différant que par une seule expérience
- et une autre d'acceptation ou rejet de ce changement.

---

**Algorithme 1.** Algorithme de MH pour simuler la loi  $\pi$  (processus de Strauss conditionnel)

---

Choisir un plan d'expériences  $x^1, \dots, x^n$  au hasard (loi uniforme par exemple)

**Pour N= 1 à NMC**

**Pour k=1 à n**

- (a) Choisir une expérience  $x^i$  au hasard
- (b) Simuler  $y^i$  uniformément dans  $\Omega = [0, 1]^d$
- (c) Accepter de remplacer  $x^i$  par  $y^i$  avec la probabilité

$$a(x, y) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right) \text{ où } y = (x^1, \dots, x^{i-1}, y^i, x^{i+1}, \dots, x^n)$$

**Fin pour k**

**Fin pour N**

---

On notera qu'il s'agit d'un algorithme itératif, NMC correspondant au nombre d'itérations McMC effectué.

La force de cet algorithme tient au fait que le calcul du rapport  $\frac{\pi(y)}{\pi(x)}$  se limite à chaque fois à calculer

$$s(x^i) = \sum_{j \neq i} 1_{\|x^i - x^j\| \leq R}$$

Il s'agit du nombre de voisins de l'expérience  $x^i$ , *i.e.* les points contenus dans la sphère centrée en  $x^i$  et de rayon  $R$ . En effet, on a

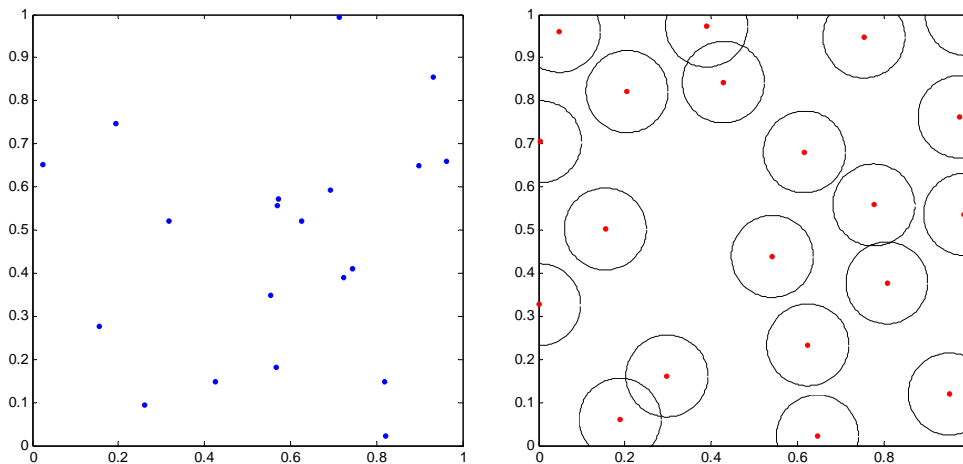
$$\frac{\pi(y)}{\pi(x)} = \gamma^{s(y^i) - s(x^i)}$$

On notera que ces plans sont bien de nature aléatoire comme par exemple les plans uniformes, les plans minimax (Johnson et al. [11]) ou bien encore les plans WSP (Sergent [20], Sergent et al. [21]). Cependant, ces plans sont construits à partir d'une loi de probabilité  $\pi$  spécifiée au préalable. C'est à travers cette loi et ses extensions (processus de Gibbs de potentiel plus général) qu'il sera possible de traduire les propriétés souhaitées. Nous justifierons l'utilisation de l'algorithme de MH pour la simulation de la loi  $\pi$ .

Désormais, nous parlerons de *plans de Strauss*.

### 2.3. Etude des plans de Strauss

La figure de droite montre une distribution spatiale de  $n = 20$  points caractéristique d'un processus de Strauss obtenu après 1000 itérations MCMC. Dans ce cas là, nous avons choisi une distribution initiale uniforme et les paramètres  $R$  et  $\gamma$  sont fixés respectivement à 0.19 et 0.1.



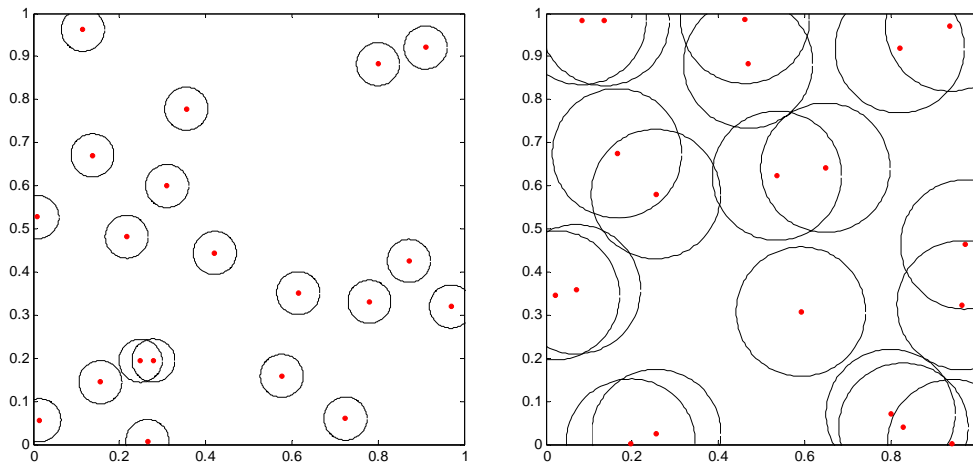
**Figure 2.** A gauche, un plan initial (aléatoire) de 20 points en dimension 2. A droite, le plan de Strauss obtenu à partir du plan de gauche avec la représentation des sphères de rayon  $R/2$ .

Sur les figures suivantes, on matérialisera les interactions entre expériences en traçant les cercles de rayon  $R/2$ , l'intersection de 2 cercles correspondant précisément à une interaction.

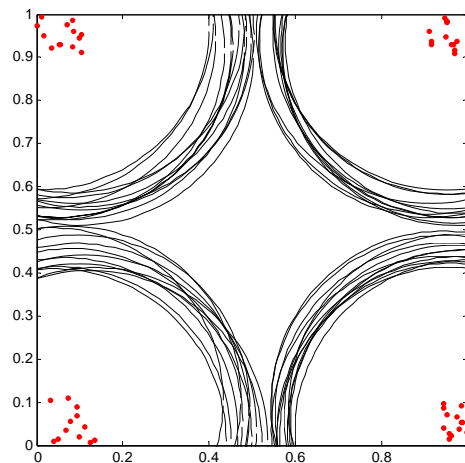


**Influence des paramètres.** Les figures ci-dessous montrent l'influence du paramètre  $R$  sur la distribution finale. Le choix du rayon s'avère être important :

- un rayon trop petit engendre une distribution sans interaction mais avec de nombreuses lacunes (distribution uniforme),
- un rayon trop grand conduit à une distribution avec des agglomérats (clusters).



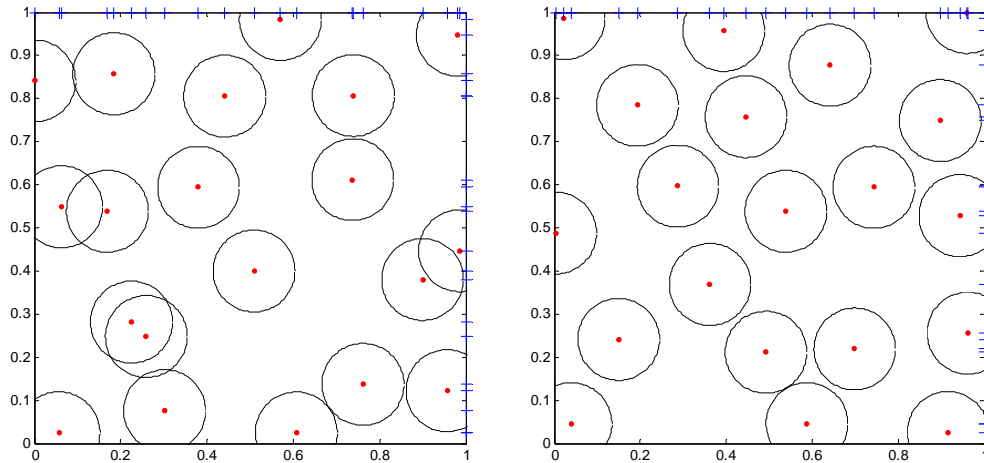
**Figure 3.** A gauche, un plan de Strauss à 20 points en dimension 2 de paramètres  $R=0.1$  et  $\gamma=0.1$ . A droite, un plan de Strauss de paramètres  $R=0.3$  et  $\gamma=0.1$ .



**Figure 4.** Un plan de Strauss à 50 points avec  $R = 0.8$  et  $\gamma=0.1$ .

La figure ci dessus permet d'illustrer et de comprendre, lorsque le rayon choisi est grand, que la loi  $\pi$  rend plus vraisemblable les agglomérats qui correspondent à un nombre global d'interactions compatible avec  $\pi$ .

De même que pour le rayon d'interaction, il est important de bien fixer le paramètre de répulsion  $\gamma$ .



**Figure 5. A gauche, un plan de Strauss de paramètres  $R=0.19$  et  $\gamma=0.1$ . A droite, un plan de Strauss de paramètres  $R=0.19$  et  $\gamma=0.001$ .**

La figure ci-dessus montre qu'il est plus aisé de générer une distribution répondant au critère de remplissage de l'espace avec un paramètre de répulsion faible. Cependant, un paramètre de répulsion trop faible contraint généralement les points à être alignés. Ce qui dégrade la répartition des points en projection, notamment sur les frontières du domaine expérimental comme c'est visible sur la Figure 5 pour le facteur  $x_1$  au niveau de la valeur 1.

Ainsi, il est délicat de bien fixer le paramètre de répulsion afin de trouver un compromis entre un remplissage de l'espace optimal et une répartition de points uniforme sur les marges de dimension un. C'est pourquoi, nous proposerons dans la suite une variante qui permet de corriger ce problème en considérant les interactions de paires sur les  $d$  projections.

Le rayon d'interaction est le paramètre le plus sensible à régler. Pour un critère donné, la meilleure solution serait sans doute de tabuler cette valeur selon le nombre de points et la dimension du problème.

Une fois les problèmes de paramétrage résolus, les plans de Strauss permettent de bien répondre à la question du remplissage de l'espace en grande dimension. Pour s'en convaincre, il suffit de calculer les critères intrinsèques tels par exemple le critère d'Arbre de Longueur Minimale (Franco et al. [6]).

## 2.4. Extensions et processus ponctuels de Gibbs

Plusieurs généralisations du processus ponctuel de Strauss sont envisageables. Par exemple, on peut faire intervenir tous les triplets de points distants deux à deux de moins de  $R$ . Nous nous contenterons ici seulement des interactions de paires qui nous paraissent suffisantes pour atteindre nos objectifs.

### 2.4.1. Contraintes sur les marges

Il a été observé que le processus de Strauss ne permet pas toujours de satisfaire à une bonne répartition des points en projection sur les marges de dimension un.

Une première extension naturelle consiste à considérer des interactions globales, *i.e.* en dimension  $d$ , ainsi que des interactions sur chaque projection afin de retrouver les bonnes propriétés des hypercubes latins. Cela fait intervenir deux nouveaux paramètres :  $\gamma_{1D}$  pour l'interaction et  $R_{1D}$  pour l'intervalle d'influence sur les marges ou projections unidimensionnelles (voir le paragraphe 2.4.3 pour l'expression générale de la densité du processus ponctuel associé).

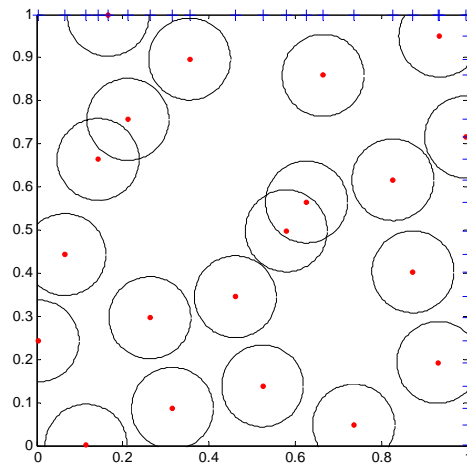


Figure 6. Un plan de Strauss à 20 points en dimension 2 ( $R = 0.19$ ,  $R_{1d} = 0.0375$ ,  $\gamma = 0.1$ ,  $\gamma_{1D} = 0.01$ ).

L'exigence d'une bonne répartition des points en projection sur les axes factoriels permet de corriger l'effet de la grande dimension qui accorde moins de volume au centre du domaine. Il est tout à fait possible d'exiger également de bonnes propriétés de projection sur des sous-espaces de dimension 2 ou plus. Même si l'obtention d'une bonne répartition sur les plans factoriels présentent un intérêt évident (estimation des interactions d'ordre 2 entre les différents facteurs), seules les contraintes sur les marges de dimension un sont présentées ici.

## 2.4.2. Plans de Strauss-Gibbs

Les plans de Strauss sont issus d'une loi de probabilité qui rend plus vraisemblable un nombre relativement faible d'interactions globales (ainsi que sur les marges, voir paragraphe 2.4.1). Pour pallier le phénomène d'agglomérats de points, notamment lorsque le rayon est inadapté (voir Figure 4), on prend en compte la distance réelle entre les points pour mesurer la force de l'interaction et non plus se contenter de savoir s'il y a interaction ou pas (sphères de rayon  $R/2$  qui se rencontrent).

Cela nous amène à considérer la famille plus large des processus ponctuels dont la distribution  $\pi$  est de la forme :

$$\pi(\mathbf{x}) \propto e^{-U(\mathbf{x})}$$

où

- $U(\mathbf{x}) = \beta \sum_{i < j} \varphi(\|\mathbf{x}^i - \mathbf{x}^j\|) + \sum_{k=1}^d \beta_k \sum_{i < j} \varphi_k(|x_k^i - x_k^j|)$
- $\beta = -\ln \gamma > 0$ ;  $\beta_k = -\ln \gamma_k > 0 \quad k = 1, \dots, d$
- et  $\varphi$  de  $[0, +\infty[$  dans  $\mathbb{R}$  est une fonction continue décroissante telle que  $\varphi(0) = 1$  et  $\varphi$  nulle à l'infini. On utilisera la famille ( $\alpha > 0, R > 0$ ) de fonctions suivante

$$\varphi(r) = \left(1 - \frac{r}{R}\right)^\alpha \quad \text{si } 0 \leq r \leq R (\varphi = 0 \text{ pour } r \geq R)$$

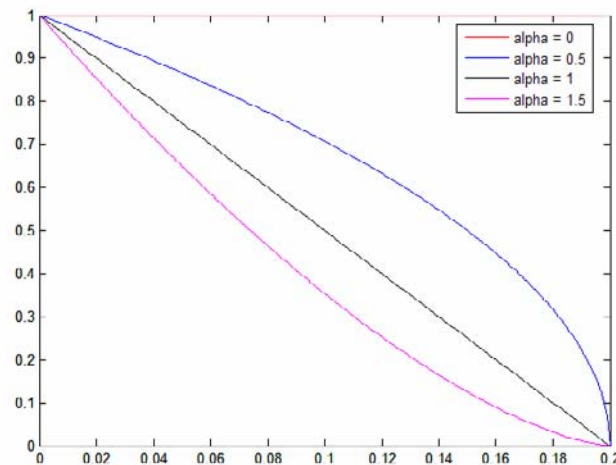
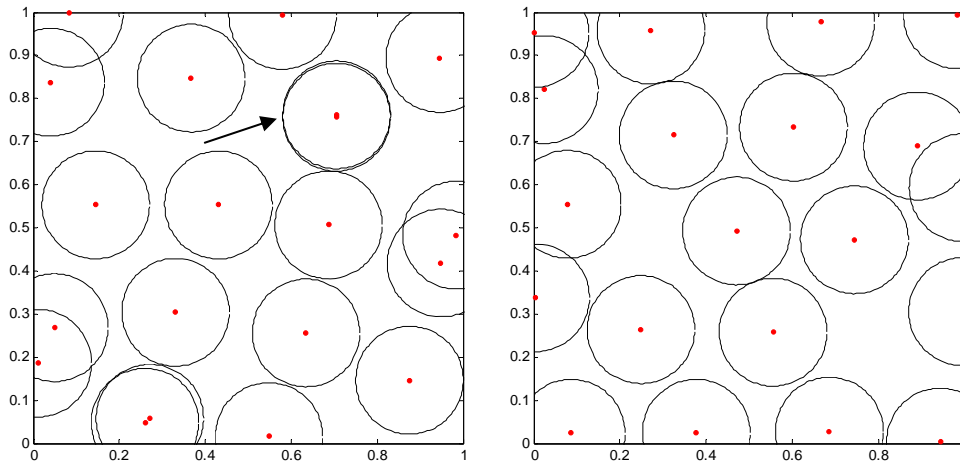


Figure 7. La fonction de potentiel puissance pour différents  $\alpha$  et  $R=0.2$ .

Il s'agit de processus ponctuels de Gibbs de potentiel  $U$ , de potentiels locaux les fonctions  $\varphi$  et  $\varphi_k$  qui correspondent uniquement à des interactions de paires (Ripley, Kelly [18]). Le

cas  $\alpha = 0$  (potentiel 0-1 dit du « tout ou rien ») correspond exactement aux processus de Strauss discutés jusqu'ici.

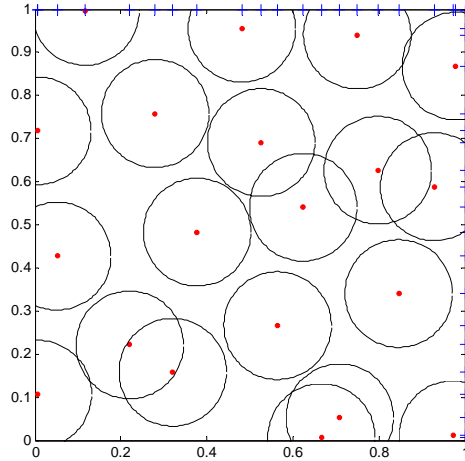
Cette extension réduit le risque de points agglomérés et améliore par exemple un critère de type maximin.



**Figure 8. A gauche. Un plan de Strauss obtenu à partir d'un potentiel tout ou rien. A droite. Un plan de Strauss obtenu à partir d'un potentiel puissance  $\alpha = 0.5$ .**

La Figure 8 permet de bien comprendre les interactions que l'on privilégie. En effet, sur le plan de gauche généré avec un potentiel de type tout ou rien nous observons qu'il y a deux points de mêmes coordonnées (où le cercle est repéré par une flèche). Généralement, ce phénomène n'apparaît pas quand on utilise un potentiel puissance qui favorise des interactions faibles comme le montre la figure de droite. Ainsi, l'utilisation d'un potentiel puissance permet de mieux répondre à la question du remplissage de l'espace.

Cependant, des alignements persistent comme on peut l'observer sur la figure de droite pour le facteur  $x_2$  au niveau de la valeur 0.02. Une solution est de combiner les contraintes sur les marges de dimension un avec le potentiel puissance afin d'obtenir une répartition uniforme sur les marges de dimension un et une répartition optimale au sens du remplissage de l'espace.



**Figure 9. Un plan de Strauss-Gibbs combinant interactions globales et sur les marges avec un potentiel de type puissance**

### 2.4.3. Cas de l'hétérogénéité

En phase exploratoire, l'objectif recherché est de distribuer les points uniformément dans l'hypercube unité. Cependant, si nous avons une connaissance métier sur le phénomène, il peut être souhaitable de distribuer les points plus densément dans une certaine partie de l'hypercube unité. Aussi, il doit être possible de générer des points selon une distribution non plus uniforme mais hétérogène ou bien encore anisotrope.

On définit alors la distribution  $\pi$  sur l'espace  $E$  des plans d'expériences par

$$\pi(\mathbf{x}) \propto \left( \prod_{i=1}^n \lambda(\mathbf{x}^i) \right) e^{-U(\mathbf{x})}$$

où  $\lambda$  est une densité de probabilité que l'on sait simuler facilement, par exemple une densité gaussienne tronquée.

C'est la dernière extension envisagée dans cet article. L'algorithme de MH qui a été retenu est le suivant :

---

**Algorithme 2.** Algorithme de MH pour simuler la loi  $\pi$  (processus de Strauss généralisé conditionnel)

---

Choisir un plan d'expériences  $x^1, \dots, x^n$  au hasard (loi uniforme par exemple)

**Pour N= 1 à NMC**

**Pour k=1 à n**

- (a) Choisir une expérience  $x^i$  au hasard
- (b) Simuler  $y^i$  selon la densité  $\lambda(\cdot)$  dans  $\Omega = [0,1]^d$
- (c) Accepter de remplacer  $x^i$  par  $y^i$  avec la probabilité

$$a(x, y) = \min\left(1, e^{U(x)-U(y)}\right) \text{ où } y = \left(x^1, \dots, x^{i-1}, y^i, x^{i+1}, \dots, x^n\right)$$

**Fin pour k**

**Fin pour N**

---

Deux caractéristiques sont à la base de cet algorithme :

- (i) la proposition de changement d'une expérience prise au hasard, notée  $Q(\mathbf{x}, d\mathbf{y})$
- (ii) la probabilité  $a : E \times E \rightarrow [0,1]$  d'accepter le changement  $\mathbf{x} \mapsto \mathbf{y}$  proposé

De manière précise, on a ici

- (i)  $Q(\mathbf{x}, d\mathbf{y}) = \sum_{i=1}^n \frac{1}{n} \lambda(\mathbf{y}^i) 1_{\mathbf{y}^{(-i)} = \mathbf{x}^{(-i)}} d\mathbf{y}^i$  où  $\mathbf{y}^{(-i)} = \mathbf{x}^{(-i)}$  signifie que  $y_j = x_j$  pour  $j \neq i$
- (ii)  $a(\mathbf{x}, \mathbf{y}) = \min\left(1, e^{U(\mathbf{x})-U(\mathbf{y})}\right)$

Si  $\delta_{\mathbf{x}}$  est la mesure de Dirac en  $\mathbf{x}$ , chaque transition élémentaire  $P : E \times E \rightarrow [0,1]$  de cet algorithme s'écrit :

$$P(x, dy) = a(x, y)Q(x, dy) + \left[ \underbrace{\int_E (1 - a(x, z))Q(x, dz)}_{p_x} \right] \delta_x(dy)$$

On remarquera que  $p_x$  est la probabilité de rester en  $\mathbf{x}$  partant de  $\mathbf{x}$  puisqu'ici  $Q(\mathbf{x}, \{\mathbf{x}\}) = 0$ .

Notons

$$K(\mathbf{x}, d\mathbf{y}) = P^n(\mathbf{x}, d\mathbf{y}) = P(X_{N+1} \in d\mathbf{y}^1 \dots d\mathbf{y}^n | X_N = \mathbf{x})$$

le noyau de transition associé au balayage aléatoire des  $n$  expériences à chaque itération  $N$  de l'algorithme. Par construction, la suite des plans d'expériences  $\mathbf{x}_{(0)}, \mathbf{x}_{(1)}, \dots, \mathbf{x}_{(NMC)}$  ainsi générés est la réalisation d'une chaîne de Markov  $(X_N)_{N \geq 0}$  de probabilité ou noyau de transition  $K$ .

La question essentielle qui se pose est de savoir si la distribution du dernier plan généré  $x_{NMC}$  est proche ou non de la distribution  $\pi(\cdot)$ .

### 3. Etude théorique de la planification d'expériences par méthodes McMC

---

La référence principale ici est Doob [3]. Considérons la chaîne de Markov  $(X_N)_{N \geq 0}$  associée au noyau  $K = P^n$  introduit au paragraphe précédent. Par construction, nous avons les 2 propriétés suivantes :

- $P$  est  $\pi$ -réversible (preuve en annexe)
- $K(\mathbf{x}, d\mathbf{y}) \geq c\lambda(d\mathbf{y})$  où  $c$  est une constante strictement positive

Le mécanisme de transition  $P$  vérifie la condition de  $\pi$ -réversibilité si

$$\pi(d\mathbf{x})P(\mathbf{x}, d\mathbf{y}) = \pi(d\mathbf{y})P(\mathbf{y}, d\mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y} \in E.$$

Cette condition de réversibilité assure  $\pi K = \pi$ , ce qui traduit le fait que  $\pi$  est une loi invariante ou stationnaire pour la chaîne de Markov  $(X_N)_{N \geq 0}$ . En particulier, si la loi initiale  $\nu$  est  $\pi$ , alors  $(X_N)_{N \geq 0}$  est une chaîne de Markov (strictement) stationnaire.

La minoration de  $K$  par une constante strictement positive permet d'assurer la propriété de mélange de la chaîne de Markov. Dans notre cas, on obtient explicitement la minoration suivante en considérant l'évolution élémentaire la moins vraisemblable :

$$K(\mathbf{x}, d\mathbf{y}) \geq c\lambda(d\mathbf{y})$$



où

- $\lambda(d\mathbf{y}) = \left( \prod_{i=1}^n \lambda(\mathbf{y}^i) \right) d\mathbf{y}^1 \dots d\mathbf{y}^n$
- $c = \frac{n!}{n^n} \left( \gamma^{n-1} \prod_{k=1}^d \gamma_k^{n-1} \right)^n$

Ces 2 propriétés conjointes assurent d'excellentes propriétés de la chaîne de Markov. Enonçons le résultat principal qui nous intéresse ici (preuve en annexe).

**Théorème.** La chaîne de Markov  $(X_N)_{N \geq 0}$  est uniformément ergodique, au sens où, pour toute loi initiale  $\nu$ , on a

$$\|\nu K^N - \pi\| \leq 2C^N ; 0 < C = 1 - c < 1$$

où  $\|\mu\| = \sup_{\|f\|_\infty \leq 1} \left| \int f(\mathbf{x}) d\mu(\mathbf{x}) \right|$  désigne la norme sur l'espace des mesures signées de  $E$  (en dualité avec l'espace  $L^\infty$  des fonctions  $f$  mesurables bornées sur  $E$ ). C'est aussi la variation totale de la mesure signée  $\mu$ .

Des résultats classiques de la théorie des chaînes de Markov dans des espaces d'états généraux permettent alors d'envisager la méthode de Monte-Carlo pour estimer la moyenne  $\pi(h)$  de toute fonctionnelle  $h$  sur  $E$  (loi forte des grands nombres ou théorème ergodique). De plus, on dispose d'un théorème central limite pour contrôler la convergence (en  $1/\sqrt{N}$ ).

## 4. Application des plans de Strauss

---

L'objectif de cette étude est de comparer divers plans d'expériences usuels sur une fonction déterministe de 10 variables, considérée « hautement » non linéaire et définie sur  $[-10,10]^{10}$  par (Jin et al. [11]) :

$$f(x_1, \dots, x_{10}) = \sum_{j=1}^{10} \exp(x_j) \left\{ c_j + x_j - \ln \left[ \sum_{k=1}^{10} \exp(x_k) \right] \right\}$$

avec  $(c_1, \dots, c_{10}) = (-6.089; -17.164; -34.054; -5.914; -24.721; -14.986; -24.100; -10.708; -26.662; -22.179)$

La figure suivante représente la coupe à  $x_i = 3$  pour  $i \neq 1, 3$ .

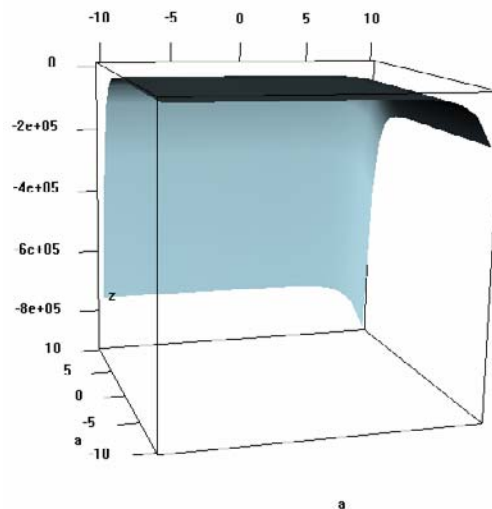


Figure 10. Coupe de la fonction de Jin en fonction de  $x_1$  et  $x_3$  ( $x_i = 3$ ,  $i \neq 1, 3$ )

**Méthodologie.** Les différents plans sont comparés à l'aide de critères – dont le détail est donné ci-dessous – permettant de valider le modèle en ajustement et en prédiction. En pratique, un modèle de surface de réponse de degré 2 calculé à l'aide de la méthode MARS (Friedman [7]) est ajusté pour chaque plan. Afin de calculer les critères de prédiction du modèle, la réponse est prédite sur un plan test de Strauss à 2000 points.

**Critères de validation d'une surface approchée.** Après avoir construit un modèle, la question de la validité du modèle ajusté se pose. Un premier diagnostic consiste à étudier les valeurs du modèle aux points du plan d'expériences : nous parlerons dans ce cas d'**ajustement**. Afin de tester la bonne adéquation entre les observations et les prévisions sur

les données utilisées pour la construction du modèle, on étudiera le coefficient de détermination  $R^2$ . Ce coefficient est défini par :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

où  $\bar{y}$  est une estimation de la moyenne de la réponse et  $n$  est le nombre de points du plan d'expériences.

La surface de réponse ainsi construite et validée va être l'outil de base pour notre étude. Le plan d'expériences représente seulement un petit échantillon des valeurs possibles pour les variables explicatives. Une deuxième étape va permettre d'étudier les qualités de généralisation du modèle proposé, c'est-à-dire son comportement en **prédiction**. L'indicateur proposé ci-dessous mesure l'écart entre la prédiction calculée *via* le modèle ajusté et la valeur réelle de la réponse donnée par le simulateur. La *RMSE* (Root Mean Square Error) correspond à la racine carrée de la moyenne des erreurs de prédiction au carré :

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}$$

où  $m = 2000$  représente le nombre de données de l'ensemble test. Ce critère mesure l'erreur quadratique moyenne commise sur les données tests. Concrètement, une faible valeur de la *RMSE* signifie que les valeurs prédites sont proches des valeurs réelles.

**Les données.** La surface de réponse MARS (Friedman [7]) de degré 2 semblant adaptée à ce cas d'étude et dépendant plus fortement du plan d'expériences que les autres méthodes plus classiques telles que les régressions quadratiques ou bien polyMARS est la méthode envisagée pour cette application.

Les plans testés comptent tous le même nombre de points fixé à 100. En voici la liste :

- Plans aléatoires (*Aléatoire*)
- Plans minimax (*Minimax*)
- Hypercubes latins maximin (*mLHs*)
- Plans de Strauss (*Strauss*)
- Plans de Strauss généralisés (potentiel puissance à 1/2) (*Strauss\_g*)
- Plans de Strauss avec des contraintes 1D (*Strauss1D*)
- Plans de Strauss généralisés avec des contraintes 1D (*Strauss1D\_g*)

Les plans étudiés introduisant tous de l'aléa stochastique, 100 plans de chaque type sera généré afin de donner un sens aux résultats. Le plan test utilisé pour calculer les critères de qualité du modèle est un plan de Strauss à 2000 points.

**Les résultats.** Les figures ci-dessous représentent les boxplots obtenus à la suite de cette étude.

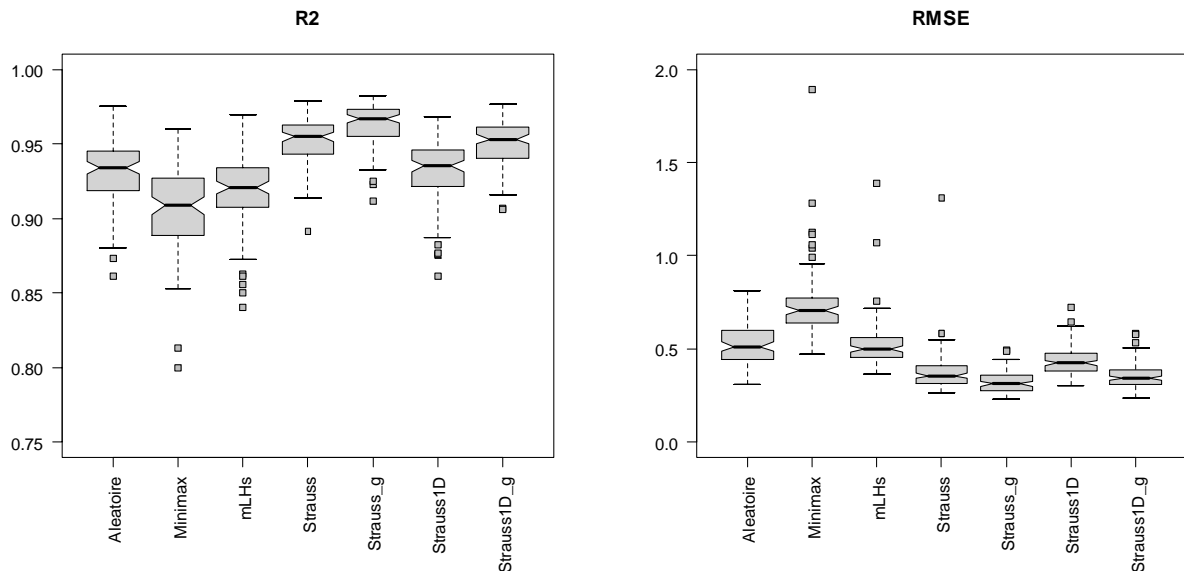


Figure 11. Boxplots des critères de qualité d'ajustement (à gauche) et de validation (à droite) des plans.

**Conclusions.** Même si ces résultats sont très dépendants du cas étudié, ils montrent que les plans de Strauss (potentiel tout ou rien et potentiel puissance) offrent une qualité d'ajustement et de validation du modèle meilleure que celle obtenue avec les autres plans.

D'autre part, les plans minimax générés par la routine *cover.design* du package *Fields* de R ne permet pas, en grande dimension, de générer de bons SFD.

Nous pouvons également remarquer que les hypercubes latins tant utilisés en expérimentation numérique sont de qualité semblable entre eux et similaire à celle d'un plan aléatoire. Ce qui pourrait confirmer que le fait de considérer les projections en dimension 1 est inenvisageable en grande dimension. Si l'on doit réellement considérer des bonnes projections en 1D, les plans de Strauss avec des contraintes 1D sont légèrement meilleurs que les hypercubes latins pourtant améliorés à l'aide du critère de maximin.

Les résultats liés aux plans de Strauss considérant un potentiel généralisé confirme que celui-ci permet de mieux remplir l'espace.

## 5. Conclusions et perspectives

---

L'utilisation des méthodes McMC dans le contexte des plans d'expériences numériques permet de construire de nouveaux plans spécifiés à partir d'une loi de Strauss et même plus généralement d'une loi de Gibbs. Cette approche offre une grande souplesse dans la mesure où l'on peut facilement jouer sur cette loi *via* sa représentation en termes de potentiels locaux d'énergie pour imposer des propriétés (par exemple de remplissage).

Par ailleurs, la méthodologie McMC constitue une alternative intéressante à l'approche statistique classique qui consiste à travailler sur des réalisations indépendantes d'une même loi. Ainsi, si l'on souhaite étudier un critère d'optimalité du plan, il est possible de calculer l'intervalle de confiance sur la valeur du critère à partir d'une seule simulation McMC (théorème central limite pour une chaîne de Markov).

Enfin, de nombreuses extensions pour la planification d'expériences numériques sont envisageables. Comme par exemple, la considération de contraintes spatiales ou bien encore l'utilisation de métriques différentes de la métrique euclidienne pour rendre compte d'hétérogénéité (notamment en grande dimension où les plans uniformes sont naturellement concentrés sur le bord du domaine).

## 6. Annexe

---

### Preuve de la $\pi$ -réversibilité du mécanisme P.

Soit un borélien de ExE. On a

$$\begin{aligned}
 \int 1_B(\mathbf{x}, \mathbf{y}) \pi(\mathbf{x}) a(\mathbf{x}, \mathbf{y}) Q(\mathbf{x}, d\mathbf{y}) d\mathbf{x} &= \int 1_B(\mathbf{x}, \mathbf{y}) \left( \sum_{i=1}^n \frac{1}{n} \pi(\mathbf{x}) \min(1, e^{U(\mathbf{x})-U(\mathbf{y})}) \lambda(\mathbf{y}^i) 1_{\mathbf{y}^{(-i)}=\mathbf{x}^{(-i)}} d\mathbf{y}^i \right) d\mathbf{x} \\
 &= \int 1_B(\mathbf{x}, \mathbf{y}) \left( \sum_{i=1}^n \frac{1}{n} k \prod_{l=1}^n \lambda(\mathbf{x}^l) e^{-U(\mathbf{x})} \min(1, e^{U(\mathbf{x})-U(\mathbf{y})}) \lambda(\mathbf{y}^i) 1_{\mathbf{y}^{(-i)}=\mathbf{x}^{(-i)}} d\mathbf{y}^i \right) d\mathbf{x} \\
 &= \int 1_B(\mathbf{x}, \mathbf{y}) \left( \sum_{i=1}^n \frac{1}{n} k \prod_{\substack{l=1 \\ l \neq i}}^n \lambda(\mathbf{x}^l) \lambda(\mathbf{x}^i) \lambda(\mathbf{y}^i) e^{-U(\mathbf{x})} \min(1, e^{U(\mathbf{x})-U(\mathbf{y})}) 1_{\mathbf{y}^{(-i)}=\mathbf{x}^{(-i)}} d\mathbf{y}^i \right) d\mathbf{x} \\
 &= \int 1_B(\mathbf{x}, \mathbf{y}) \left( \sum_{i=1}^n \frac{1}{n} k \prod_{l=1}^n \lambda(\mathbf{y}^l) \lambda(\mathbf{x}^i) e^{-U(\mathbf{y})} \min(1, e^{U(\mathbf{y})-U(\mathbf{x})}) 1_{\mathbf{y}^{(-i)}=\mathbf{x}^{(-i)}} d\mathbf{y}^i \right) d\mathbf{x} \\
 &= \int 1_B(\mathbf{x}, \mathbf{y}) \pi(\mathbf{y}) \min(1, e^{U(\mathbf{y})-U(\mathbf{x})}) \sum_{i=1}^n \frac{1}{n} \lambda(\mathbf{x}^i) 1_{\mathbf{x}^{(-i)}=\mathbf{y}^{(-i)}} d\mathbf{x}^i d\mathbf{y}
 \end{aligned}$$

Finalement,

$$\int 1_B(\mathbf{x}, \mathbf{y}) \pi(d\mathbf{x}) P(\mathbf{x}, d\mathbf{y}) = \int 1_B(\mathbf{x}, \mathbf{y}) \pi(d\mathbf{y}) P(\mathbf{y}, d\mathbf{x}).$$

C'est la condition de  $\pi$ -réversibilité du mécanisme de transition P.

### Preuve du théorème.

Soient  $\nu_1$  et  $\nu_2$  deux lois initiales quelconques. Notons  $\nu$  la mesure signée  $\nu_1 - \nu_2$  sur  $E$  de sorte que

$$\nu(E) = \nu_1(E) - \nu_2(E) = 0$$

Soit  $\mu = \nu_1 K - \nu_2 K$  la mesure signée

$$\mu(d\mathbf{y}) = \int \nu_1(d\mathbf{x}) K(\mathbf{x}, d\mathbf{y}) - \int \nu_2(d\mathbf{x}) K(\mathbf{x}, d\mathbf{y})$$

Pour toute constante  $c$  et  $f$  mesurable bornée sur E, considérons la décomposition

$$\int f(\mathbf{y}) \mu(d\mathbf{y}) = \int f(\mathbf{y}) (\nu_1(d\mathbf{x}) - \nu_2(d\mathbf{x})) (K(\mathbf{x}, d\mathbf{y}) - c\lambda(d\mathbf{y})) + \int f(\mathbf{y}) \nu_1(d\mathbf{x}) c\lambda(d\mathbf{y}) - \int f(\mathbf{y}) \nu_2(d\mathbf{x}) c\lambda(d\mathbf{y})$$

En utilisant le fait que  $\nu(E) = 0$ , on a simplement

$$\int f(\mathbf{y})\mu(d\mathbf{y}) = \int f(\mathbf{y})(\nu_1(d\mathbf{x}) - \nu_2(d\mathbf{x}))(K(\mathbf{x}, d\mathbf{y}) - c\lambda(d\mathbf{y}))$$

On pose  $g(\mathbf{x}) = \int_E f(\mathbf{y})(K(\mathbf{x}, d\mathbf{y}) - c\lambda(d\mathbf{y}))$ . On a clairement  $g$  mesurable bornée sur  $E$  et

$$\int f(\mathbf{y})\mu(d\mathbf{y}) = \int g(\mathbf{x})\nu(d\mathbf{x})$$

On en déduit, avec le choix particulier d'une constante  $c$  telle que  $K(\mathbf{x}, d\mathbf{y}) \geq c\lambda(d\mathbf{y})$ ,

$$|g(\mathbf{x})| \leq \int |f(\mathbf{y})|(K(\mathbf{x}, d\mathbf{y}) - c\lambda(d\mathbf{y})) \leq \|f\|_\infty \times C \quad \text{où } C = 1 - c < 1$$

Ainsi,

$$\|g\|_\infty \leq C \|f\|_\infty$$

et, pour  $f$  bornée par 1,

$$\left| \int f(\mathbf{y})\mu(d\mathbf{y}) \right| = |\nu(g)| \leq \|\nu\| \|g\|_\infty \leq C \|f\|_\infty \|\nu\| \leq C \|\nu\|$$

Finalement on a,

$$\|\nu_1 K - \nu_2 K\| \leq C \|\nu_1 - \nu_2\|.$$

Ce qui traduit que l'« opérateur »  $\mu \longrightarrow \mu K$  est strictement contractant de l'espace des probabilités sur  $E$  dans lui-même.

Il suffit pour conclure de se souvenir de la relation  $\pi = \pi K$  qui traduit que la probabilité  $\pi$  est un point fixe de  $K$ .

## 7. Références

---

- [1] Cressie N. Statistics for Spatial Data, New York, Wiley, (1993) 11.
- [2] Daley D.J., Vere-Jones D. An Introduction to The Theory of Point processes. Elementary Theory and Methods, 2<sup>nd</sup> edn, New York, Springer, **I** (2003) 4.
- [3] Doob, J. L., Stochastic Processes, John Wiley & Sons, 1953
- [4] Fang K.-T., Li R., Sudjianto A. Design and Modeling for Computer Experiments. Chapman & Hall (2006).
- [5] Fisher R.A. The arrangement of field experiments. J. Ministry Agric. **33** (1926) 503-513.
- [6] Franco J., Vasseur O., Corre B., Sergent M. Minimum Spanning Tree : a new approach to assess the quality of design of computer experiments. Chemolab (soumis en 2007)
- [7] Friedman J. Multivariate Adaptive Regression Splines (invited paper), Annals of Statistics, **10** (1991) 1-141.
- [8] Hastings W.K. Monte Carlo Sampling methods using Markov chains and their applications, Biometrika, **57** (1970) 97-109.
- [9] Hickernell, F. A generalized discrepancy and quadrature error bound, Mathematics of computation, **67** (1998) 299-322.
- [10] Jin R., Chen W., Simpson T.W. Comparative studies of metamodelling techniques under multiple modelling criteria. Struct Multidisc Optim, Springer-Verlag, **23** (2001) 1-13.
- [11] Johnson M.E., Moore L.M., Ylvisaker D. Minimax and maximin distance designs. J. of Statis. Planning and Inference, **26** (1990) 131-148.
- [12] Koehler J.R., Owen A.B. Computer Experiments, Handbook of Statistics, **13** (1996) 261-308.
- [13] Mc Kay 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** (1979) 239-245.
- [14] Meyn S.P., Tweedie R.L. Markov Chain and Stochastic Stability, New York, Springer (1993).
- [15] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E. Equation of state calculations by fast computing machines, Journal of Chamental Physics, **21** (1953) 1087-1092.
- [16] Niederreiter, H. Low-Discrepancy and Low-Dispersion Sequences, Journal of number theory, **30** (1987) 51-70.
- [17] Owen A.B. Orthogonal arrays for computer experiments, integration and visualization. Statistica Sinica **2** (1992) 439-452.
- [18] Ripley B.D., Kelly F.P. Markov point processes, Journal of the London Mathematical Society, **15** (1977) 188-192.
- [19] Santner T.J., Williams B.J., Notz W.I. The Design and Analysis of Computer Experiments, Springer, (2003) 121-161.



- [20] Sergent M. Contribution de la Méthodologie de la Recherche Expérimentale à l'élaboration de matrices uniformes : Application aux effets de solvants et de substituants, Ph D Thesis (1989) Université Aix Marseille III.
- [21] Sergent M, Phan-Tan-Luu R., Elguero J. Statistical Analysis of Solvent Scales. Part 1, Anales de Quimica Int. Ed., **93** (1997) 3-6.
- [22] Strauss D.J. A model for clustering, *Biometrika*, **62** (1975) 467-475.