

MA - No. 43.2

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
Ministère de L'enseignement Supérieur et de la Recherche Scientifique

Université Saad Dahlab 1
Faculté des sciences
Département de mathématiques.

*Mémoire de fin d'étude en vue de l'obtention d'un Master en
Mathématiques Appliqués
Option :*

MODELISATION STOCHASTIQUE ET STATISTIQUE

Méthodologie de la recherche expérimentale.

Space filling designs

Réalisé par :

Proposé et encadré par :

Bouredja Oussama.

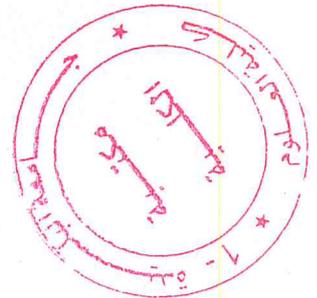
Mr .El Mossaoui.

Soutenance : 20/12/2018.

Promotion 2018-2019.

Examinateur : F. Massoufi

président : D. Rabee .



MA-510-43-2

REMERCIEMENTS

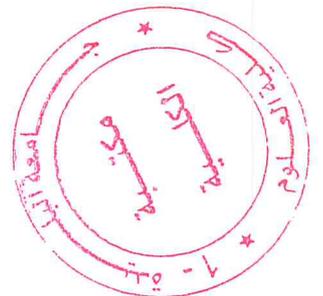
En premier lieu, nos vifs remerciements vont d'emblée à ALLAH qui nous aide et nous donne la patience et le courage qui nous ont permis de réaliser cet humble travail.

Nous tenant à remercier sincèrement Monsieur, El Mossaoui Hichem, en tant que promoteur de mémoire, qu'est toujours montré à l'écoute et très disponible tout au long de la réalisation de ce mémoire, aussi pour l'inspiration, l'aide et le temps qu'il a bien voulu nous consacrer et sans qui ce mémoire n'aurait jamais vu le jour.

On n'oublie pas nos parents pour leur contribution, leur soutien et leur patience.

Enfin, nous adressons nos plus sincères remerciements à tous nos proches et amis, qui nous ont toujours encouragées au cours de la réalisation de ce mémoire.

Merci à tous et à toutes.



DEDICACE

Je dédie ce modeste travail :

A mes très chers parents,

Pour les sacrifices déployés à mes égards, pour leur patience, leur amour et leur confiance en moi.

Qu'ils trouvent dans ce modeste travail, le témoignage de ma profonde affection et mes attachements indéfectibles. Nulle dédicace ne puisse exprimer ce que je leur devons, que dieu leur réserve la bonne santé et une longue vie.

A mes amis et amies,

Qui me sont chers : qu'ils trouvent ici l'expression de mes sentiments les plus dévoués et mes vœux les plus sincères.

A mes frères,

Qui ont tant sacrifié qui ont toujours été présents pour moi dans la joie et le malheur.

A tous les membres de ma famille,

Surtout mes grands-parents, je vous dirais tout simplement, un grand merci.

المخلص

يريد مستخدمي المدونات الرقمية مكلفة في حساب الوقت خفض تكلفة عن طريق الحد من عدد من عمليات المحاكاة بعد الاختيار الحكيم على أساس استخدام التصميم التجريبي التي تتكيف مع البيئة الرقمية و المسمى «تصميم ملء الفراغ». من أجل ملء مساحة المعلمات بشكل أفضل, نقدم في هذا العمل بعض خطط انواع ملء الفضاء مع دراسة مقارنة وفقاً لعدة معايير للاحتياجات كل دراسة .

قد تم إنشاء تطبيق في ماتلاب يسمح ببناء خطط في ملء الفراغ, حساب قيم المعايير المختلفة اختيار وتنفيذ دراسة تجريبية.

RESUME

Les utilisateurs de codes numériques onéreux en temps de calcul souhaitent réduire le coût en limitant le nombre de simulations suivant un choix judicieux fondé sur l'utilisation de plans d'expériences adaptés au contexte numérique et appelés « space filling designs ». Afin de remplir au mieux l'espace des paramètres, nous présentons dans ce travail certains nombre de plans de type spacefilling avec une étude de comparaison et ceci afin de répondre aux besoins de chaque étude.

Une application en matlab a été établie. Elle permet la construction des plans en space filling, le calcul des valeurs des différents critères retenus et de mener à bien une étude expérimentale.

TABLE ES MATIERES

Remerciements	
Dédicace	
Résumé	
Table des matières	
Liste des figures.	
Liste des tableaux	
Liste des indices	
Introduction générale.....	1
1 Chapitre 1 : Introduction à la méthodologie des plans d'expériences.....	3
1.1 introduction.....	3
1.2 Étude d'un phénomène	3
1.2.1 Notion d'espace expérimental.....	4
1.2.2 Notion de Surface de réponse	7
1.2.3 Notion de modélisation mathématique a priori de la réponse.....	8
▪ Modélisation mathématique	8
▪ Modélisation expérimentale	9
▪ Système d'équations.....	10
1.3 Matrice d'expériences.....	11
1.4 Les coordonnées centrées réduites	12
2 Chapitre 2 : Plans de type space filling.....	13
2.1 Hyper cubes latins.....	13
2.1.1 Échantillonnage par hyper cubes latins.....	13
2.2 Suites à faible discrédance.....	15
2.2.1 Les suites de Van Der Corput.....	16
2.2.2 Les suites de Halton	17
2.2.3 Les suites finies de Hammersley.....	20
2.2.4 Les suites de Faure.....	21
2.2.5 Suites de Sobol.....	24
2.3 Plans aléatoires	27
2.3.1Nombres pseudo-aléatoires	28
2.3.2Nombres quasi-aléatoires	28

3	Chapitre 3 : Critère d'optimalité utilisé pour l'évaluation de la qualité des plans de type spacefilling.....	29
3.1	Critères de distance.....	29
3.1.1	La Mesure de recouvrement.....	29
3.1.2	Le rapport de distance.....	30
3.1.3	Distances Maximin et Minimax.....	31
3.2	Critères d'uniformité basés sur la discrédance.....	31
3.2.1	Discrédance en norme 2.....	32
3.2.1.1	Discrédance extrême et à l'origine	32
3.2.1.2	Discrédance modifiée	34
3.2.1.3	Discrédance centrée.....	34
3.2.1.4	Discrédance symétrique.....	36
4	Chapitre 4 : Implémentation et Comparaison.....	38
4.1	Logiciel MATLAB.....	38
4.1.1	Présentation du logiciel MATLAB.....	38
4.1.2	Interface graphique MATLAB	39
4.1.2.1	Création d'une interface graphique sous matlab.....	39
4.1.3	Outils guide	38
4.2	L'étude de comparaison.....	41
4.3	Résultats	41
4.3.1	Résultat pour 10 points et 2 dimensions.....	42
4.3.2	Résultat pour 20 points et 3 dimensions.....	42
4.3.3	Résultat pour 50 points et 7 dimensions.....	43
4.3.4	Résultat pour dimension supérieur à 10.....	44
4.4	Interprétation	45
4.5	Conclusion.....	45
	Conclusion générale.....	46
	Référence	47

LISTE DE FIGURES

Figure 1.1	représente le niveau bas et le niveau haut d'un facteur.....	5
Figure 1.2	désigne la représentation Chaque facteur par un axe gradué et orienté.....	5
Figure 1.3	représente les points expérimentaux aux l'espace expérimental.....	6
Figure 1.4	désign la disposition des points expérimentaux dans le domaine d'études	7
Figure 1.5	Les réponses associées aux points du domaine	8
Figure 2.1	10 points issus d'un échantillonnage par hyper cube latin en dimension 2.....	14
Figure 2.2	illustre la construction d'une suite de Van Der Corput.	17
Figure 2.3	représente les 100, 1000,10000 premiers points de la suite de halton en base 2 et 3.....	19
Figure 2.4	représente les 50,250 et 500 premiers points d'une suite finie de Hammersley en base 2.....	21
Figure 2.5	représente les 100, 1000,10000 premiers points d'une suite de Faure en dimension $s = 2$ et en base $b = 3$	23
Figure 2.6	représente les 100, 1000 et 10000 premiers points d'une suite de sobol.....	26
Figure 3.1	représente un sous-ensemble J pour le calcul de la discrèpance centrée.....	37
Figure 3.2	désigne la représentation le volume total des 2 sous-ensembles	38
Figure 4.1	représente la Fenêtre guide.....	40
Figure 4.2	représente la Proprtey insepector.....	41

LISTE DES TABLEAUX

Tableau 1.1	Matrice d'expériences.....	11
Tableau 2.1	Les 10 premiers nombre de la suite de Halton.....	19
Tableau 2.2	La suite de Faure dans les 3 premières dimensions.....	24
Tableau 2.3	La suite de Sobol dans les 3 premières dimensions.....	27
Tableau 4.1	les valeurs obtenues des différent plans en utilisant différent critère pour $n=10$ et $d=2$	43
Tableau 4.2	les valeurs obtenues des différent plans en utilisant différent critère pour $n=20$ et $d=3$	43
Tableau 4.3	les valeurs obtenues des différent plans en utilisant différent critère pour $n=50$ et $d=7$	43
Tableau 4.4	les valeurs obtenues des différent plans en utilisant différent critère pour $n=200$ et $d=15$	44
Tableau 4.5	les valeurs obtenues des différent plans en utilisant différent critère pour $n=400$ et $d=20$	44

LISTE DES INDICES

a_0 :	Constante d'un modèle.
a_i :	Coefficients des termes de premier degré.
a_i^2 :	Coefficients des termes de deuxième degré.
a_{ii} :	Coefficients des termes carrés.
a_{ij} :	Coefficients des termes rectangles.
X^T	Matrice transposée de X.
$(X^T X)$	Matrice d'information.
$(X^T X)^{-1}$	Matrice de dispersion
\hat{F}_n	désigne la fonction de répartition empirique de n expériences.
F_u	désigne la fonction de répartition de la loi uniforme de n points sur $[0,1]$.
I:	Matrice unité.
d:	Nombre de facteurs.
n:	désigne le nombre des expériences(les points).
Δ :	L'écart d'ajustement ou « Lack of fit ».
e_i :	Ecart résiduel.
e:	Vecteur des écarts.
(x_n)	désigne Une suite <i>aléatoire</i> distribuée sur le domaine d'expérience.
$D_n(x)$	désigne la discrédance qui permet de mesurer l'irrégularité de la distribution de (x_n) .
$\phi_b(i)$	désigne La fonction radicale inverse en base b.
C_p	désigne la suite de Van Der Corput en base b.
$\text{dist}(x^i, x^k)$	désigne la distance euclidienne entre deux points x^i et x^k .
λ	désigne la mesure de recouvrement de l'espace par les points.

INTRODUCTION GENERALE

La simulation numérique est aujourd'hui un outil incontournable de compréhension du réel et d'aide à la conception. Les progrès des ordinateurs permettent la simulation réaliste et, parfois, l'optimisation de systèmes complexes. La simulation numérique est utilisée dans de nombreux domaines de recherche et développement ainsi que dans le secteur financier, où elle cherche à calculer différentes approximations, cette dernière vise à sélectionner une méthode d'approximation en fonctions de critères mathématiques, sans négliger le coût de calcul et précision numérique.

La méthodologie des plans d'expériences permet une recherche expérimentale planifiée appelé plans d'expériences. L'expérimentation ne peut pas être quelconque, elle doit fournir l'information désirée. Cette démarche expérimentale va aider l'expérimentateur à structurer sa recherche de manière différente.

Le but de ce travail est de donner au lecteur un bref aperçu des connaissances actuelles sur les plans d'expériences numériques en phase exploratoire, et à noter que ces différents plans seront souvent cités tout au long de ce mémoire à titre de comparaison pour étudier l'intérêt des Space Filling Designs (en français conception de remplissage de l'espace).

Le premier chapitre aborde la notion générale de planification expérimentale utilisée en expérimentation numérique en utilisant quelques notions mathématiques. Ce chapitre débute par une présentation très générale de la démarche de planification expérimentale ainsi que du vocabulaire de base (facteurs, réponse, domaine expérimental, etc...).

Le deuxième chapitre discute des plans d'expériences de type space filling design utilisés en expérimentation classique, les hypercubes latins, les suites à faible discrétion, ... etc. à noter que les travaux fondateurs, dans ce domaine, sont Goupy, Johnson, Faure, Jessica ...ect

Le troisième chapitre présente les critères d'optimalité pour étudier et qualifier la distribution des points et s'attache à étudier la qualité intrinsèque de ces plans au niveau du remplissage de l'espace.

Le dernière chapitre concerne l'application et l'étude comparative que nous venons d'effectuer sur les principaux plans d'expérience de type space filling qui nous permettront d'ouvrir la discussion sur les plans et de tirer les renseignements pour faire une conclusion.

CHAPITRE 1

INTRODUCTION A LA METHODOLOGIE DES PLANS D'EXPERIENCES

1.1 INTRODUCTION

Les plans d'expériences permettent d'organiser au mieux les essais qui accompagnent une recherche scientifique ou des études industrielles. Ils sont applicables à de nombreuses disciplines et à toutes les industries à partir du moment où l'on recherche le lien qui existe entre une grandeur d'intérêt, et des variables x_i . Il faut penser aux plans d'expériences si l'on s'intéresse à une fonction du type :

$$y = f(x_i) \dots \dots \dots (1.1)$$

Avec les plans d'expériences on obtient le maximum de renseignements avec le minimum d'expériences. Pour cela, il faut suivre des règles mathématiques et adopter une démarche rigoureuse. Il existe de nombreux plans d'expériences adaptés à tous les cas rencontrés par un expérimentateur. Les principes fondamentaux de cette science seront indiqués et les principaux plans seront passés en revue. La compréhension de la méthode des plans d'expériences s'appuie sur deux notions essentielles, celle d'espace expérimental et celle de modélisation mathématique des grandeurs étudiées.

1.2 Étude d'un phénomène

L'étude d'un phénomène revient souvent à s'intéresser à une grandeur particulière comme la consommation d'essence d'une voiture ou comme le prix de revient d'un produit chimique ou encore comme le rendement en blé d'une parcelle de terrain. Cette grandeur, consommation, prix ou rendement, dépend d'un grand nombre de variables. La consommation de la voiture est fonction de la vitesse du véhicule, de la puissance du moteur, de la manière de conduire, de la direction et de la force du vent, du gonflage des pneumatiques, de la présence ou

non d'un porte-bagages, du nombre de personnes transportées, de la marque de la voiture, etc. Le prix du produit chimique dépend de la qualité des matières premières, des rendements des unités de production, des spécifications imposées, des conditions de fabrication, etc. Il en est de même pour le rendement en blé qui est fonction de la nature du terrain, de la quantité d'engrais incorporé, de l'exposition au soleil, du climat, de la variété de blé ensemencé, etc.

Sous une forme mathématique, on peut écrire que la grandeur d'intérêt, y que nous appellerons également réponse par la suite, est une fonction de plusieurs variables x_i (variables que nous appellerons aussi facteurs par la suite). On a :

$$y = f(x_1, x_2, \dots, x_k) \dots \dots (1.2)$$

L'étude du phénomène se ramène à déterminer la fonction $f()$ qui lie la réponse y aux différents facteurs x_1, x_2, \dots, x_k .

Pour approfondir cette approche il faut introduire quelques notions particulières et une terminologie spécifique aux plans d'expériences

1.2.1 Notion d'espace expérimental :

Un expérimentateur qui lance une étude s'intéresse à une grandeur qu'il mesure à chaque essai. Cette grandeur s'appelle la réponse, c'est la grandeur d'intérêt. La valeur de cette grandeur dépend de plusieurs variables. Au lieu du terme «variable» on utilisera le mot facteur. La réponse dépend donc de un ou de plusieurs facteurs. Le premier facteur peut être représenté par un axe gradué et orienté (Figure 1.1). La valeur donnée à un facteur pour réaliser un essai est appelée niveau. Lorsqu'on étudie l'influence d'un facteur, en général, on limite ses variations entre deux bornes. La borne inférieure est le niveau bas. La borne supérieure est le niveau haut.

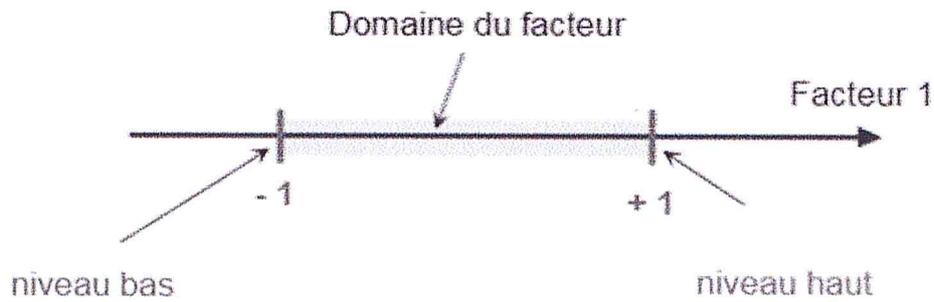


Figure 1.1 : Le domaine de variation du facteur est constitué de toutes les valeurs comprises entre le niveau bas et le niveau haut.

L'ensemble de toutes les valeurs que peut prendre le facteur entre le niveau bas et le niveau haut, s'appelle le domaine de variation du facteur ou plus simplement le domaine du facteur. On a l'habitude de noter le niveau bas par -1 et le niveau haut par $+1$.

S'il y a un second facteur, il est représenté, lui aussi, par un axe gradué et orienté. On définit, comme pour le premier facteur, son niveau haut, son niveau bas et son domaine de variation. Ce second axe est disposé orthogonalement au premier. On obtient ainsi un repère cartésien qui définit un espace euclidien à deux dimensions. Cet espace est appelé l'espace expérimental (Figure 1.2).

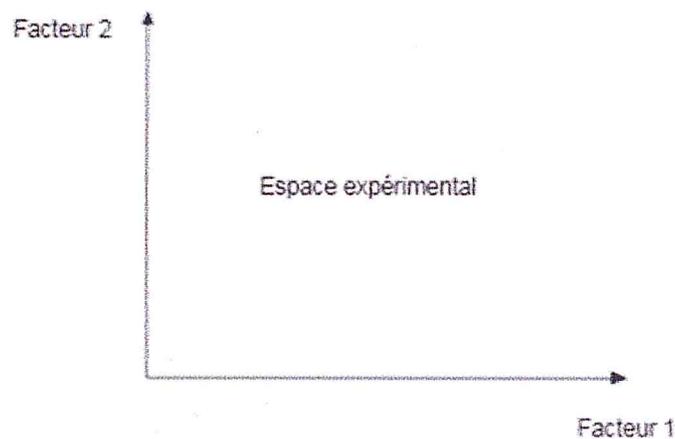


Figure 1.2 : Les axes des facteurs sont orthogonaux entre eux. L'espace ainsi défini est l'espace expérimental.

Le niveau X_1 du facteur 1 et le niveau X_2 du facteur 2 peuvent être considérés comme les coordonnées d'un point de l'espace expérimental (Figure 1.3). Une expérience donnée est alors

représentée par un point dans ce système d'axes. Un plan d'expériences est représenté par un ensemble de points expérimentaux.

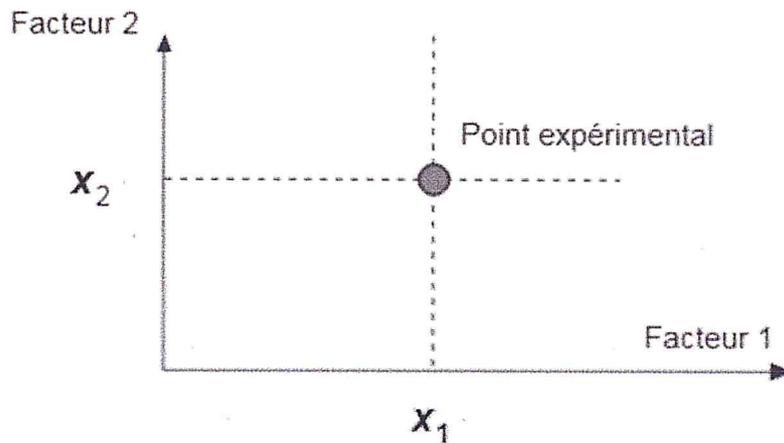


Figure 1.3 : Dans l'espace expérimental, les niveaux des facteurs définissent des points expérimentaux.

Le regroupement des domaines des facteurs définit le «domaine d'étude». Ce domaine d'étude est la zone de l'espace expérimental choisie par l'expérimentateur pour faire ses essais. Une étude, c'est-à-dire plusieurs expériences bien définies, est représentée par des points répartis dans le domaine d'étude (Figure 1.4). Cette façon de représenter une expérimentation par des points dans un espace cartésien est une représentation géométrique de l'étude.

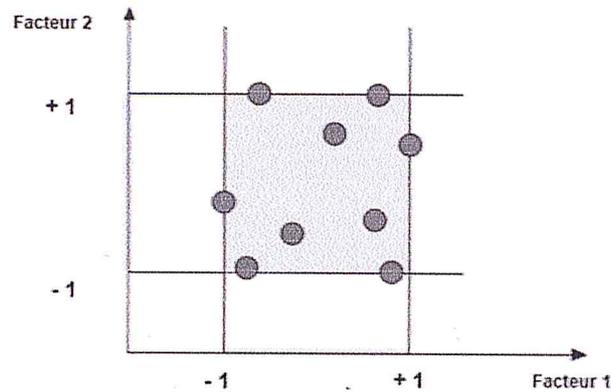


Figure 1.4 : Les points expérimentaux sont disposés dans le domaine d'étude défini par l'expérimentateur.

Les définitions qui ont été données s'appliquent bien aux variables continues. Mais il existe d'autres types de variables. Il y a les variables discrètes comme par exemple des personnes : Julien, Arthur, Louis, Simon et Nathan. On peut encore parler d'espace expérimental mais il n'aura pas les mêmes propriétés que l'espace des variables continues. Il y a également les grandeurs ordonnables comme, par exemple, des distances qui peuvent être courtes, moyennes et longues. Là aussi, la notion d'espace expérimental existe toujours mais cet espace possède des propriétés différentes des deux premiers.

1.2.2 Notion de Surface de réponse :

Les niveaux x_i représentent les coordonnées d'un point expérimental et y est la valeur de la réponse en ce point. On définit un axe orthogonal à l'espace expérimental et on l'attribue à la réponse. La représentation géométrique du plan d'expériences et de la réponse nécessite un espace ayant une dimension de plus que l'espace expérimental. Un plan à deux facteurs utilise un espace à trois dimensions pour être représenté : une dimension pour la réponse, deux dimensions pour les facteurs.

A chaque point du domaine d'étude correspond une réponse. A l'ensemble de tous les points du domaine d'étude correspond un ensemble de réponses qui se localisent sur une surface appelée la surface de réponse (Figure 5).

Le nombre et de l'emplacement des points d'expériences est le problème fondamental des plan d'expériences. On cherche à obtenir la meilleure précision possible sur la surface de réponse tout en limitant le nombre d'expériences.

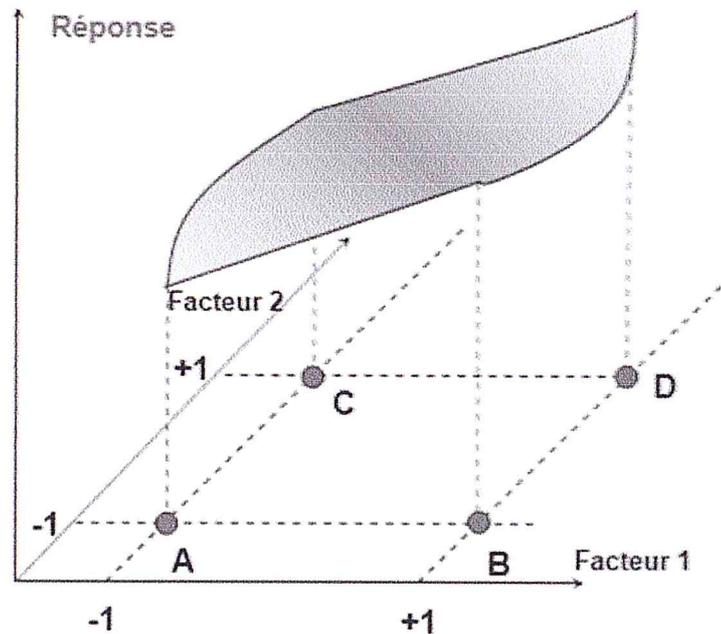


Figure 1.5 :Les quelques réponses mesurées aux points du plan d'expériences.

1.2.3 Notion de modélisation mathématique a priori de la réponse :

- **Modélisation mathématique :**

En l'absence de toute information sur la fonction qui lie la réponse aux facteurs, on se donne a priori une loi d'évolution dont la formulation la plus générale est la suivante :

$$y = f(x_1, x_2, \dots, x_n) \dots \dots \dots (1.3)$$

Cette fonction est trop générale et il est d'usage d'en prendre un développement limité de Taylor-Mac Laurin, c'est-à-dire une approximation. Si les dérivées peuvent être considérées comme des constantes, le développement précédent prend la forme d'un polynôme de degré plus ou moins élevé :

$$y = a_0 + \sum a_i x_i + \sum a_{ij} x_i x_j + \sum a_{ii} x_i^2 + \dots \dots \dots (1.4)$$

Où :

- y est la grandeur à laquelle s'intéresse l'expérimentateur ; c'est la réponse ou la grandeur d'intérêt,
- x_i représente un niveau du facteur i ,
- x_j représenté un niveau du facteur j ;
- a_0, a_{ij}, a_{ii} sont les coefficients du polynôme.

L'intérêt de modéliser la réponse par un polynôme est de pouvoir calculer ensuite toutes les réponses du domaine d'étude sans être obligé de faire les expériences.

Ce modèle est appelé le modèle a priori ou le modèle postulé. Les modèles établis sont des modèles de prévision valables dans le domaine d'étude, domaine que l'on doit toujours préciser. Ce ne sont pas des modèles théoriques basés sur des lois physico-chimiques ou mécaniques. Dans quelques rares cas, il est possible d'utiliser des lois physiques théoriques connues.

▪ **Modélisation expérimentale :**

Deux compléments doivent être apportés au modèle purement mathématique précédemment décrit.

- Le premier complément est le manque d'ajustement. Cette expression traduit le fait que le modèle choisi par l'expérimentateur avant les expériences est probablement un peu différent du modèle réel qui régit le phénomène étudié. Il y a un écart entre ces deux modèles. Cet écart est le manque d'ajustement (lack of fit en anglais), on le note par la lettre Δ .
- Le second complément est la prise en compte de la nature aléatoire de la réponse. En effet, dans le cas général, si l'on mesure plusieurs fois une réponse en un même point expérimental, on n'obtiendra pas exactement le même résultat. Il y a une dispersion des résultats. Les dispersions ainsi constatées sont appelées erreurs aléatoires ou erreurs

expérimentales (pure error en anglais) et on les note par la lettre e . La relation générale (1.4) doit être modifiée ainsi :

$$y = f(x_1, x_2, \dots, x_n) + \Delta + \epsilon \dots \dots \dots (1.5)$$

Δ : L'écart d'ajustement .

ϵ : Erreur expérimental (ou résiduel) .

▪ **Système d'équations :**

Chaque point expérimental permet d'obtenir une valeur de la réponse. Cette réponse est modélisée par un polynôme dont les coefficients sont les inconnues qu'il faut déterminer. A la fin du plan d'expériences, on a un système de n équations (s'il y a n essais) à p inconnues (s'il y a p coefficients dans le modèle choisi a priori). Ce système s'écrit d'une manière simple en notation matricielle :

$$Y = a \cdot x + e \dots \dots \dots (1.6)$$

Où :

- y est le vecteur des réponses.
- X est la matrice de calcul, ou matrice du modèle, qui dépend des points expérimentaux choisis pour exécuter le plan et du modèle postulé.
- a est le vecteur des coefficients.
- e est le vecteur des écarts.

Ce système possède un nombre d'équations inférieur au nombre d'inconnues. Il y a n équations et $p + n$ inconnues. Pour le résoudre, on utilise une méthode de régression basée sur le critère des moindres carrés. On obtient ainsi les estimations des coefficients que l'on note \hat{a} .

Le résultat de ce calcul est :

$$\hat{a} = (X'X)^{-1}X'y$$

Formule dans laquelle la matrice X' est la matrice transposée de X . De nombreux logiciels (comme statistica) exécutent ce calcul et donnent directement les valeurs des coefficients.

1.3 Matrice d'expériences :

La représentation géométrique d'un plan d'expériences est commode pour imaginer la position des points expérimentaux dans le domaine d'étude. Mais elle ne peut plus être employée dès que le nombre de facteurs est supérieur à deux. Pour les espaces multidimensionnels, nous adopterons une représentation sous forme de tableau ou matrice d'expériences. La matrice d'expériences est un objet mathématique qui représente, sous forme codée ou normée, l'ensemble des expériences à réaliser.

C'est un tableau constitué de N lignes correspondant aux N expériences à réaliser et de k colonnes, correspondant aux k variables étudiées. L'élément x_{ij} de la matrice ainsi formée correspond à la valeur du niveau que prend la $j^{\text{ième}}$ variable à la $i^{\text{ième}}$ expérience. La matrice d'expériences (tableau 1.1) définit les essais à réaliser. Le terme essai est l'équivalent de point d'expérience.

N°essai	Facteur 1	Facteur 2
1(A)	-1	-1
2(B)	+1	-1
3(C)	-1	+1
4(D)	+1	+1

Tableau 1.1 : matrice d'expériences.

1.4 Les coordonnées centrées réduites :

Lorsqu'on attribue la valeur - 1 au niveau bas d'un facteur et la valeur +1 au niveau haut, on effectue deux modifications importantes :

- On change l'unité de mesure. Par exemple, si le niveau bas d'un facteur est 10°C et le niveau haut 30°C, il y a 20°C entre ces deux valeurs, soit 20 fois l'unité de température. Entre -1 et +1 il y a deux unités nouvelles. La nouvelle unité vaut 10°C, on lui donne le nom de Pas.
- On déplace l'origine des mesures. Dans l'exemple choisi, le milieu de l'intervalle [-1, +1] correspond à une température de 30°C. La nouvelle origine, notée zéro, diffère donc de l'origine exprimée en unité courante.

Ces deux modifications entraînent l'introduction de nouvelles variables que l'on appelle variables centrées réduites (v.c.r.) ou unités codées. Centrées pour indiquer le changement d'origine et réduites pour signaler la nouvelle unité. Le passage des variables d'origine, A aux variables centrées réduites, X et inversement, est donné par la formule suivante (A_0 est la valeur centrale en unités courantes) :

$$X = \frac{A - A_0}{pas}$$

L'intérêt des v.c.r. est de pouvoir présenter les plans d'expériences de la même manière quels que soient les domaines d'étude retenus et quels que soient les facteurs.

CHAPITRE 2

PLANS DE TYPE SPACE FILLING

Nous discutons dans ce chapitre des différents types des plans space filling qui, par construction, ont des bonnes propriétés en termes de non redondance, et de non alignement sur certains sous-espaces. Cependant, rien ne garantit qu'ils remplissent bien l'espace.

2.1 Hyper cubes latins

2.1.1 Échantillonnage par hyper cubes latins

L'échantillonnage stratifié, dont le but est de réduire la variance, consiste à découper l'espace des variables d'entrée $[0,1]^s$, en sous-espaces disjoints, puis à échantillonner au sein de chacun de ces sous-espaces. L'échantillonnage par hyper cube latin (LHS Latin HyperCube Sampling) est un particulier d'échantillonnage stratifié où on s'assure que le découpage a définis des sous-espaces équiprobables, et que chaque sous-espace est échantillonné de manière uniforme. Plus précisément, chaque intervalle de chacune des variables est partitionné en n sous-intervalles de même probabilité $[\frac{i-1}{n}, \frac{i}{n}]$, $i = 1, \dots, n$. Une valeur est sélectionnée de manière uniforme dans chaque sous-intervalle. Les n valeurs obtenues pour la variable X^1 sont couplés aux n valeurs de la variable X^2 de manière aléatoire et équiprobable pour chaque combinaison. Plus précisément, on se donne deux permutations aléatoires uniformes indépendantes π_1 et π_2 de $\{1, \dots, n\}$. La valeur de X^1 sur le sous-intervalle $[\frac{\pi_1(i-1)}{n}, \frac{\pi_1(i)}{n}]$. Sera couplée à la valeur de X^2 sur le sous-intervalle $[\frac{\pi_2(i-1)}{n}, \frac{\pi_2(i)}{n}]$. Ces n paires sont ensuite combinées de manière aléatoire et équiprobable avec les n valeurs de la variable X^3 pour former n triplets, et ainsi de suite jusqu'à ce que n s -uplets soient formés. On peut donc générer un LHS de n points en dimensions, en prenant s permutations aléatoires uniformes indépendantes de $\{1, \dots, n\}, \pi_j(1), \dots, \pi_j(n), j = 1, \dots, s$.

Les s coordonnées des n points sont alors données par :

$$x_i^j = \frac{\pi_j(i) - U_i^j}{n}, j = 1, \dots, s \text{ et } i = 1, \dots, n.$$

Où $(U_i^j)_{i=1, \dots, n, j=1, \dots, s}$ sont ns variables aléatoires i.i.d de loi $U([0,1]^s)$.

La Figure 2.1 illustre un échantillonnage par hyper cube latin en dimensions = 2. Sur cette figure, on voit que l'intervalle de chaque variable a été découpé en $n = 10$ sous-intervalles de même taille. Pour chaque variable et dans chaque sous-intervalle un point a été généré selon une loi uniforme. Les couples (X^1, X^2) ont été obtenus à l'aide des permutations $\pi_1 = (5;4;7;9;8;2;10;3;1;6)$ et $\pi_2 = (9;4;5;7;10;6;8;1;2;3)$.

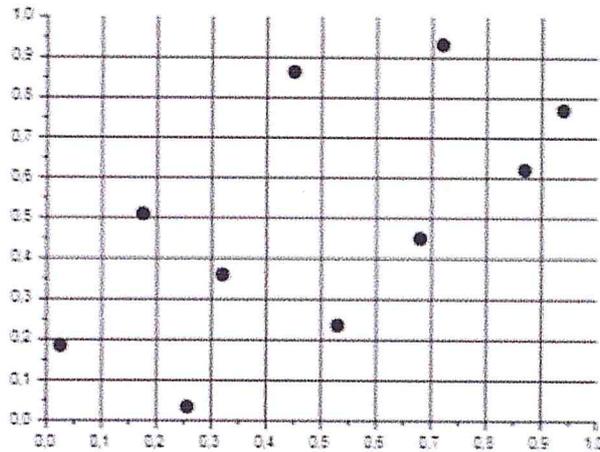


Figure 2.1 : 10 points issus d'un échantillonnage par hyper cube latin en dimension 2.

Discussion :

L'échantillonnage par hyper cube latin présente l'avantage d'échantillonner uniformément chaque intervalle de définition de chaque variable. De plus, sous l'hypothèse de monotonie de la fonction, il existe des preuves théoriques de la réduction de la variance de l'estimateur. Des exemples de fonctions non-monotones ont montré que l'échantillonnage par hyper cube latin pouvait donner des résultats moins bons qu'un échantillonnage aléatoire lorsque le nombre de points est finie.

2.2 Suites à faible discrédance

Discrédance c'est une mesure bien connue de l'irrégularité de la distribution des ensembles de points, elles jouent un rôle majeur dans les méthodes QMC (Quasi Monte Carlo) pour choisir de bons ensembles de points.

Les suites à faible discrédance sont des suites déterministes ayant de meilleures propriétés d'uniformité que les suites aléatoires dans le cube de dimension d , dont la discrédance est d'ordre $O\left(\frac{(\log n)^s}{n}\right)$ pour n assez grand. Ces suites comprennent les suites de Van der corput, Halton, Sobol, Hammersley, et les suites de Faure.

Le caractère aléatoire des points est moins important que leur distribution uniforme. Le caractère aléatoire des points n'est intéressant que dans la mesure où asymptotiquement, il mène au remplissage de l'hyper cube unité. D'où l'idée de remplacer les points aléatoires par des points issus de suites déterministes ayant une distribution plus uniforme.

Pour cela, on utilise des points issus de suites déterministes réparties uniformément dans l'espace. Pour quantifier l'écart à l'uniformité d'une suite finie de points, on utilise un critère appelé discrédance.

Nous présentons ici des plans dont l'objectif est de bien remplir l'espace. Aussi, nous nous attachons à regarder leurs propriétés en projection. Ces suites de points sont initialement utilisées pour remplacer les suites aléatoires dans la méthode de Monte-Carlo, d'où la dénomination de méthodes de quasi Monte-Carlo. La plupart de ces suites à faible discrédance sont générées par des algorithmes déterministes afin d'obtenir des points répartis le plus uniformément possible dans le domaine expérimental.

Définition 2.1 :[5]

La discrédance mesure la déviation de la répartition dans l'espace d'un échantillon de points par rapport à celle d'un échantillon de même taille tire par une loi uniforme. Plusieurs définitions de la discrédance existent et, pour une suite $(x)_{i=1,\dots,N}^{(i)}$ elle peut notamment s'écrire :

$$D(x) = \sup_{y \in [0,1]^K} \left| \frac{1}{N} \sum_{K=1}^N 1_{(x^K \in [0,1])} - \text{Volume}([0,y]) \right|$$

Les suites à faible discrédance sont des suites quasi-aléatoires ayant une faible valeur de la discrédance. Ce sont potentiellement des plans d'expérience très satisfaisants, car elles permettent un bon remplissage de l'espace. Plusieurs suites à faible discrédance existent et nous avons notamment utilise les suites quasi-aléatoires comme de Sobol et les suites de Halton...[5]

2.2.1 Les suites de Van Der Corput

Les suites de Van Der Corput (sont notés VD) sont des suites à discrédance faible en dimension $s = 1$. nous présentons ces suites car elles sont à l'origine de constructions plus générales en dimension s quelconque.

Soit un entier $b \geq 2$. Tout entier $i \in N$ peut se décomposer dans la base b de manière unique :

$$i = \sum_{k=0}^{\infty} C_k(i)b^k \dots\dots (1)$$

Avec $C_k(i) \in \{0,1,\dots,b-1\} \forall k$

On remarque que les coefficients $C_k(i)$ sont nuls pour tout $k > 1 + \lfloor \log_b(i) \rfloor$, où $\lfloor \cdot \rfloor$ désigne la partie entière inférieure. Par conséquent, la somme précédente est finie. A partir de cette décomposition, on définit la fonction radicale inverse par :

$$\varphi_b(i) = \sum_{k=0}^{\infty} C_k(i)b^{-k-1}, \text{ Pour tout } i \in N$$

On a $\varphi_b(i) \in [0,1[$ pour tout $i \in N$.

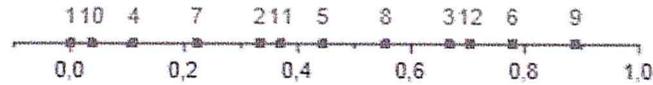
Définition 2.2 :

Soit un entier $b \geq 2$. La suite $VDC_b = (x_0, x_1, \dots) \in [0,1[$ donnée par

$$x_i = \varphi_b(i)$$

est appelée suite de Van Der Corput en base b .

La figure ci-dessous illustre la construction d'une suite de Van Der Corput. Il s'agit dans ce cas particulier de la suite de Van Der Corput en base 3.



La figure 2.2 Les 12 premiers points de la suite de Van Der Corput en base 3.

Exemple 2.1

Supposons $b = 3$ et $i = 19$, on peut écrire 19 en base 3 comme suit :

$i = 19 = 2 \times 3^2 + 0 \times 3^1 + 1 \times 3^0$, donc la suite de Van Der Corput s'écrit comme :

$$\varphi_3(19) = \frac{1}{3} + \frac{0}{9} + \frac{2}{27} = \frac{11}{27}$$

2.2.2 Les suites de Halton :

Halton est le premier à avoir construit des suites dont la discrédance est en $O(\log N)^s$, elles se déduisent de la suite de van der Corput en base b autrement dit les suites de Halton sont une généralisation en dimension $s \geq 1$ des suites de Van Der Corput.

Les suites de Halton sont construites à partir d'une base différente par dimension de l'espace, ces bases étant construites à partir de nombres premiers. Ce choix permet de minimiser la discrédance en réduisant le terme dominant de la majoration de la discrédance, ainsi que d'avoir des suites uniformes, c'est-à-dire uniformément distribuées dans le cube unité.

Définition 2.3:

Soient b_1, \dots, b_s , s entiers positifs premiers entre eux. La suite b_1, \dots, b_s donnée par :

$$x_i = (\phi_{b_1}(i), \dots, \phi_{b_s}(i))$$

est appelée la suite de Halton en bases b_1, \dots, b_s .

Le théorème suivant, donne une majoration de la discrédance infinie d'une suite de Halton.

Théorème 2.1 : [11]

La discrédance infinie à l'origine des $n \in \mathbb{N}^*$ premiers points d'une suite de Halton H_{b_1, \dots, b_s} vérifié:

$$D_{\infty, *} (H_{b_1, \dots, b_s}) \leq \frac{s}{n} + \frac{1}{n} \prod_{j=1}^s \left(\frac{b_j - 1}{2 \log b_j} \log n + \frac{b_j + 1}{2} \right)$$

Cette dernière inégalité peut s'écrire :

$$D_{\infty, *} (H_{b_1, \dots, b_s}) \leq C_s \frac{(\log n)^s}{n} + O\left(\frac{(\log n)^{s-1}}{n}\right)$$

Où

$$C_s = \prod_{j=1}^s \frac{b_j - 1}{2 \log b_j}.$$

Atanassov a, par la suite, amélioré ce résultat en montrant que la constante C_s était en fait égale à :

$$C_s = \frac{1}{s!} \prod_{j=1}^s \frac{b_j - 1}{2 \log b_j}.$$

La constante C_s est croissante par rapport à chaque $b_j, j = 1, \dots, s$, par conséquent, pour qu'elle soit minimale, il faut choisir pour bases les s premiers entiers premiers. Cependant, elle croit de manière exponentielle avec la dimension, ce qui affaiblit l'efficacité de la majoration en grande dimension.

Dans la pratique, on utilise les s premier nombre premier pour b_1, \dots, b_s . Ces suites sont beaucoup moins robustes que les suites de Sobol, notamment des que la dimension est assez grande.

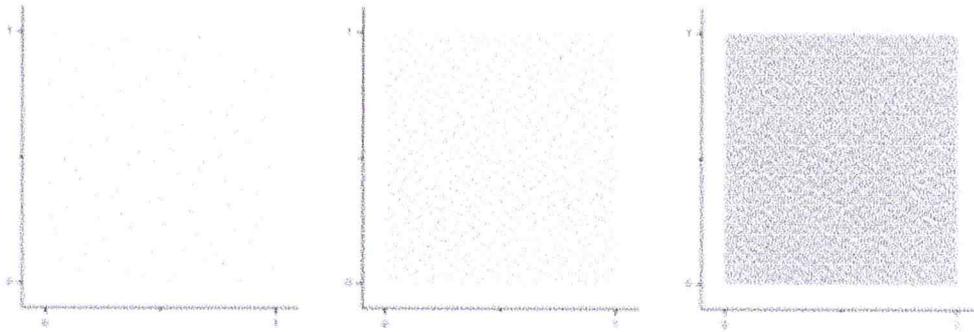


Figure 2.3 : les 100, 1000,10000 premiers points de la suite de halton en base 2 et 3.

Exemple 2.2

Les 10 premiers nombre de la suite de Halton dans les trois premières dimensions sont donnés dans le tableau suivant suivi de leurs représentations graphique pour $d = 1(b = 2)$ et $d = 2(b = 3)$.

	Dim=1(base2)	Dim=2(base3)	Dim=3(base5)
$n = 1$	$1/2$	$1/3$	$1/5$
$n = 2$	$1/4$	$2/3$	$2/5$
$n = 3$	$3/4$	$1/9$	$3/5$
$n = 4$	$1/8$	$4/9$	$4/5$
$n = 5$	$5/8$	$7/9$	$1/25$
$n = 6$	$3/8$	$2/9$	$6/25$
$n = 7$	$7/8$	$5/9$	$11/25$
$n = 8$	$1/16$	$8/9$	$16/25$
$n = 9$	$9/16$	$1/27$	$21/25$
$n = 10$	$5/16$	$10/27$	$2/25$

Tableau 2.1 : Les 10 premiers nombre de la suite de Halton.

2.2.3. Les suites finies de Hammersley

Une suite finie de n points de Hammersley en dimension s est construite à partir d'une suite finie de n points d'une suite de Halton en dimension $n - 1$.

Définition 2.4 :

Soient $b_1, \dots, b_{s-1}, s - 1$ entiers premiers positifs distincts. La suite finie de $n \in \mathbb{N}^*$ points $H_{n,b_1,\dots,b_s} \equiv (x_1, \dots, x_n)$ donné par :

$$x_i = \left(\frac{i}{n}, \phi_{b_1}(i), \dots, \phi_{b_{s-1}}(i) \right)$$

est appelée suite finie de Hammersley en bases b_1, \dots, b_{s-1} .

La majoration de la discrétance infinie à l'origine est donnée par le théorème suivant.

Théorème 2.2: [11]

Pour une suite finie de Hammersley de n points $H_{n,b_1,\dots,b_{s-1}}$, on a

$$D_{\infty,*} (H_{n,b_1,\dots,b_{s-1}}) \leq \frac{s}{n} + \frac{1}{n} \prod_{j=1}^{s-1} \left(\frac{b_j-1}{2 \log b_j} \log n + \frac{b_j+1}{2} \right)$$

Cette dernière inégalité peut s'écrire :

$$D_{\infty,*} (H_{n,b_1,\dots,b_s}) \leq C_S \frac{(\log n)^{s-1}}{n} + O\left(\frac{(\log n)^{s-2}}{n}\right)$$

Où

$$C_S = \prod_{j=1}^{s-1} \frac{b_j-1}{2 \log b_j}.$$

Comme pour une suite de Halton, on comprend que la constante C_S de la majoration sera minimale pour les bases correspondant aux $s - 1$ premiers entiers premiers.

Le principe qui permet de passer d'une suite de Halton à une suite finie de Hammersley permet d'obtenir une meilleure majoration. Ce principe est général et peut s'appliquer à

n'importe quelle suite à discrétance faible pour passer d'une suite à discrétance en $O\left(\frac{(\log n)^s}{n}\right)$ à une suite finie de discrétance en $O\left(\frac{(\log n)^{s-1}}{n}\right)$.

Discussion :

Ces suites étant construites à partir de suites de Halton, elles présentent le même phénomène de diagonales successives. De plus, il est impossible de rajouter des points supplémentaires à ces suites sans pour autant perturber la discrétance. Dans le cas où le nombre de points à générer n'est pas connu à l'avance, il est déconseillé d'utiliser une suite de Hammersley. Aussi, ces suites ont perdu l'aspect itératif des suites de Halton qui permettait de rajouter des points facilement.

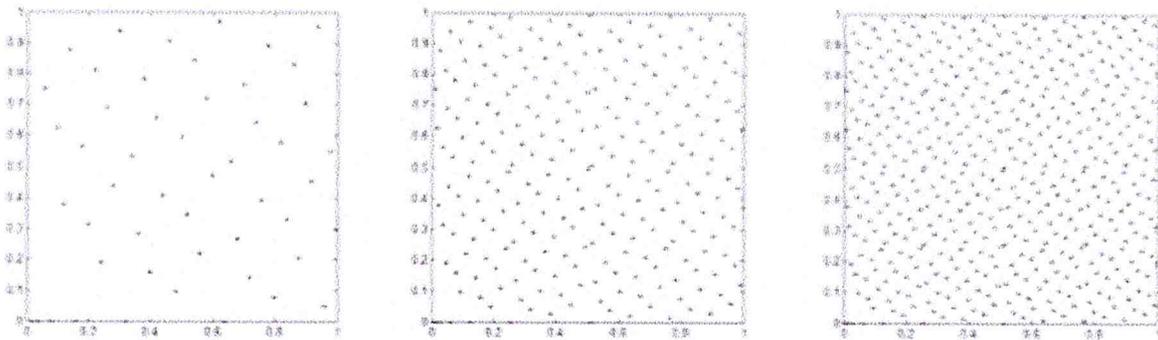


Figure 2.4 : les 50,250 et 500 premiers points d'une suite de hammersley en base 2.

2.2.4 Les suites de Faure

Les suites de Faure sont définies à partir d'une base unique b , où $b \geq s$ est un entier premier.

Définition 2.5 :

Soit $b \geq s$ un entier premier. Soient $B = (b_1, \dots, b_s)$, s éléments de $\{0, \dots, b - 1\}$. La suite de Faure $F_{b,B} = (x_1, x_2, \dots)$ est définie par :

$$x_i^j = \sum_{k=1}^{\infty} x_i^{j,k} b^{-k} \text{ Pour } i \geq 1 \text{ et } j = 1, \dots, s$$

Où

$$x_i^{j,k} = \sum_{p=k}^{\infty} C_p^{k-1} b_j^{p-k+1} c_k(i) \pmod{b}$$

Avec $c_k(i)$ pour $k \geq 1$, les coefficients de la décomposition de i dans la base b (1), et

$$C_p^k = \frac{p!}{k!(p-k)!}$$
 est un coefficient binomial.

Les coefficients $c_k(i)$ étant nuls pour $k > 1 + [\log_b i]$

De même que pour les suites précédentes, il existe un théorème qui fournit une majoration de la discrédance des n premiers points d'une suite de Faure.

Théorème 2.3 : [11]

Pour une suite de Faure $F_{b,B}$ en dimensions, on a :

$$D_{\infty,*}(F_{b,B}) \leq \begin{cases} \frac{3}{16 (\log 2)^2} \frac{(\log n)^s}{n} + O\left(\frac{(\log n)^{s-1}}{n}\right) & \text{si } b = 2 \\ \frac{1}{s!} \left(\frac{b-1}{2 \log b}\right)^s \frac{(\log n)^s}{n} + O\left(\frac{(\log n)^{s-1}}{n}\right) & \text{si } b \geq 3 \end{cases}$$

Contrairement aux suites de Sobol et de Halton, la constante $C_S = \frac{1}{s!} \left(\frac{(b-1)^s}{(2 \log b)}\right)^s$, tend vers 0 lorsque la dimension tend vers l'infini. De plus, elle est inférieure à celles d'une suite de Sobol ou de Halton. C'est pour cette raison que les suites de Faure sont considérées comme meilleures que celles de Halton et de Sobol. Cependant, il est à noter que la constante dans $O\left(\frac{(\log n)^{s-1}}{n}\right)$ explose avec la dimension.

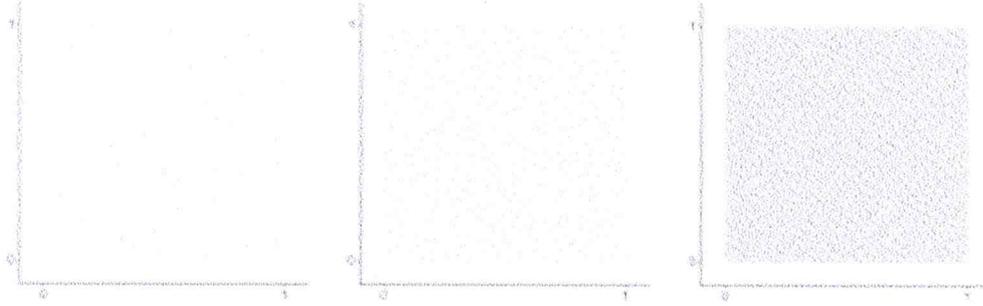


Figure 2.5 : les 100, 1000, 10000 premiers points d'une suite de Faure en dimension $s = 2$ et en base $b = 3$.

Exemple 2.3

Considérons le cas $r = 2$ et $b = 3$. Les matrices $C^{(1)}$ et $C^{(2)}$ sont données par :

$$C^{(1)} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, C^{(2)} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

Pour $k = 1, \dots, 8$, les vecteurs $a(n)$ sont donnés par :

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

Les vecteurs $C^{(1)}a(n) \bmod b$ sont donnés par :

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Les vecteurs $C^{(2)}a(n) \bmod 2$ sont donnés par :

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

En multipliant ces vecteurs par $(\frac{1}{3}, \frac{1}{9})$ et en réarrageant ces fractions en trois lignes, nous obtenons les 9 premiers éléments de la suite de Faure en dimension 3.

	Dim=1 (base 2)	Dim=2 (base 3)	Dim=3(base 3)
$n = 1$	1/3	1/3	1/3
$n = 2$	2/3	2/3	2/3
$n = 3$	1/9	4/9	7/9
$n = 4$	4/9	7/9	1/9
$n = 5$	7/9	1/9	4/9
$n = 6$	2/9	8/9	5/9
$n = 7$	5/9	2/9	8/9

2.2.5 Suites de Sobol

Les suites de Sobol permettent de repartir les différents points dans l'espace en minimisant la distance entre chaque observation. Leur construction est assez compliquée, et en arithmétique modulo 2 elles s'obtiennent par des récurrences linéaires à partir de polynômes primitifs sur le corps $Z_2 = \{0,1\}$. Elles sont nommées quasi-aléatoires car on peut toujours trouver les coordonnées du deuxième point à partir du premier et ainsi de suite.

Pour expliquer leur construction, il faut d'abord rappeler la notion de polynôme primitif :

Définition 2.6:

Un polynôme $p(t)$ de degré m de la forme $t^m + a_1 t^{m-1} + \dots + a_{m-1} t + a_m$ est dit primitif sur Z_2 s'il est irréductible sur Z_2 et si le plus petit entier i , tel qu'il divise $t^i - 1$ ou $t^i + 1$, est égal à $2^m - 1$.

Remarque 2.1 :

La notion d'irréductibilité consiste à vérifier si le polynôme $p(t)$ n'est divisible par aucun autre polynôme de degré inférieur. Le plus petit entier positif i tel que $p(t)$ divise $t^i - 1$ est aussi appelé l'ordre du polynôme. Un polynôme primitif de degré s doit posséder les monômes 1 et t^s , ainsi qu'un nombre impair de termes.

Définition 2.7 :

Ainsi une suite de Sobol $u = \{x^0, x^1, \dots, x^{n-1}\}$ est construite en dimension l de la façon suivante :

$$x^i = \frac{1}{2^s} \left(\sum_{k=1}^m p_k l_k \right)$$

Où :

- $(p_k, 1 \leq k \leq s)$ est la représentation binaire de i .
- $s = \begin{cases} 1 & \text{si } i = 0 \\ 1 + \ln[i] & \text{sinon} \end{cases}$
- symbole $+$ désigne simplement l'addition dans Z_2
- $l_k, 1 \leq k \leq m$ sont des entiers impairs tels que $1 \leq l_k \leq 2$
- Les l_k avec $k > s$ se calculent par la récurrence :

$$l_k = 2a_1 l_{k-1} \oplus 2^2 a_2 l_{k-2} \oplus \dots \oplus 2^{m-1} a_{m-1} l_{k-m+1} \oplus 2^m l_{k-m} \oplus l_{k-m}$$

Où a_k sont les coefficients d'un polynôme primitif de degré m sur Z_2 .

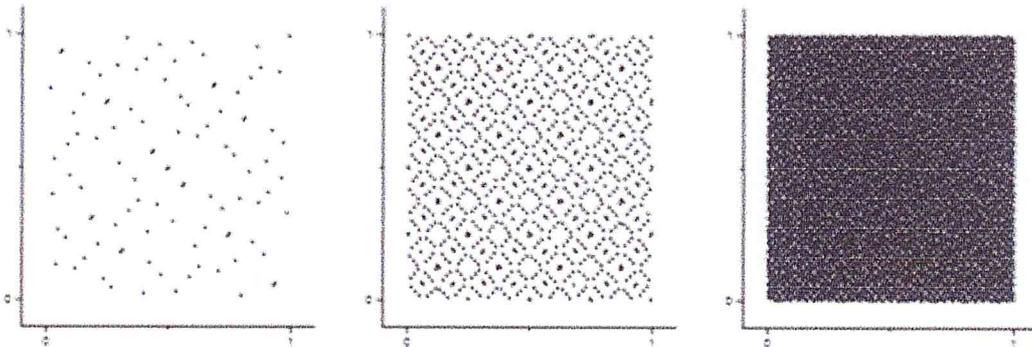


Figure 2.6 : les 100, 1000 et 10000 premiers points d'une suite de sobol.

Discussion :

Pour construire cette suite dans un espace de dimension s , il faut alors prendre d polynômes primitifs distincts.

Rapides à construire, ces suites permettent aussi de garder une bonne disposition dans l'espace malgré l'augmentation en dimension. Un des inconvénients réside dans l'initialisation des entiers l_k . Ces derniers étant aléatoire, la répartition des points dans l'hyper cube unité peut être mal faite. Afin de pallier cet inconvénient, une technique de brouillage peut être utilisée.

La majoration de la discrédance d'une suite de Sobol est donnée par le théorème suivant [11]

Théorème 2.4 :

Pour une suite de Sobol S de n points en dimension s , on :

$$D_{\infty,*}(S) \leq \frac{2^{t_s}}{s! (\log 2)^s} \frac{(\log n)^s}{n} + O\left(\frac{(\log n)^{s-1}}{n}\right)$$

Avec

$$k \frac{s \log s}{\log \log s} \leq t_s \leq \frac{s \log s}{\log 2} + O(s \log \log s)^1 \quad \text{et } k \geq 0.$$

Ce théorème permet tout d'abord de voir que le paramètre t_s d'une suite de Sobol croit approximativement de manière linéaire avec la dimension s . Sobol fournit les valeurs du paramètre t pour plusieurs dimensions s données. De plus, comme pour les suites de Halton, la constante $C_s = \frac{2^t}{s! (\log 2)^s}$ croit de manière exponentielle avec la dimension.

Les suites de Halton sont beaucoup moins robustes que les suites de Sobol, notamment des que la dimension est assez grande.

Exemple 2.4

Ce tableau illustre la suite de Sobol dans les 3 premières dimensions.

	Dim=1(base 2)	Dim=2 (base 2)	Dim=3(base 2)
$n = 1$	1/2	1/2	1/2
$n = 2$	1/4	1/4	3/4
$n = 3$	3/4	3/4	1/4
$n = 4$	3/8	3/8	5/8
$n = 5$	7/8	7/8	1/8
$n = 6$	5/8	1/8	3/8
$n = 7$	1/8	5/8	7/8
$n = 8$	3/16	5/16	5/16

Tableau 2.3 : La suite de Sobol dans les 3 premières dimensions.

2.3 Plans aléatoires

Un Plan Aléatoire c'est un plan construit par tirage aléatoire dans le cube unité.

Le plan classique dans les méthodes de Monte-Carlo consiste à générer des suites de points pseudo-aléatoires dans l'espace des facteurs. L'erreur d'estimation découle du théorème central limite. Elle est en $\frac{1}{\sqrt{n}}$, n étant la taille de l'échantillon. On distingue :

- nombres pseudo-aléatoires
- nombres quasi-aléatoires

2.3.1 Nombres pseudo-aléatoires

- Tirage selon une loi uniforme de n points indépendants de Ω .
- Erreur probable d'estimation en (\sqrt{n}) .
- La vitesse de convergence ne dépend pas de la dimension d de Ω .

Exemple 2.5 :

`r1 = rand(10,10)`

(Sous matlab rand une fonction qui donne un chiffre aléatoire entre 0 et 1)

2.3.2 Nombres quasi-aléatoires

Utilisation de suite à faible discrédance.

CHAPITRE 3

CRITERES D'OPTIMATILE UTILISES POUR L'EVALUATION DE LA QUALITE DES PLANS DE TYPE SPACE FILLING

Dans ce chapitre, nous discutons dans un premier temps les critères de discrédance. Ensuite, nous introduisons la classe des critères de distance.

3.1 Critères de distance

Étudier l'uniformité d'une distribution de points lorsque $d \geq 2$, ne peut se faire visuellement. Il est donc utile de se référer à des critères afin de pouvoir décider si une distribution est uniforme ou non ou/et si elle vérifie également le bon remplissage de l'espace. Il existe pour cela des critères intrinsèques aux plans d'expériences – c'est-à-dire des critères qui ne préjugent en rien de la qualité de la surface de réponse déterminée par la suite à partir de ces plans qui se prêtent particulièrement bien à cet objectif. Nous avons choisi les critères d'uniformité les plus courants.

3.1.1 La Mesure de recouvrement

Elle permet de mesurer l'écart entre les points du plan et ceux d'une grille régulière. Ce critère est nul pour une grille régulière. L'objectif est donc de le minimiser pour se rapprocher d'une grille régulière, et ainsi assurer le remplissage de l'espace, sans toutefois l'atteindre pour respecter une distribution uniforme.

Définition 3.1 :

Soit $x = \{ x^1, \dots, x^n \}$ une suite de n points en d dimensions.

On définit la mesure de recouvrement λ par [6] :

$$\lambda = \frac{1}{\gamma} \left[\frac{1}{n} \sum_{j=1}^n (\gamma_j - \bar{\gamma}) \right]^{1/2}$$

Avec $\gamma_i = \text{dist}(x^i, x^k)$ désignant la distance minimale entre le point x^i et les autres points de la suite et $\bar{\gamma} = \frac{1}{n} \sum_{i=1}^n \gamma_i$ la moyenne des γ_i .

Remarque 3.1 :

- Si les points sont sur une grille régulière, on a $\gamma_1 = \dots = \gamma_i = \gamma$ et donc $\lambda=0$.
- Ainsi, plus λ est faible, et plus les points sont proches de ceux d'une grille régulière. Cette expression fait apparaître explicitement le coefficient de variation de l'échantillon des γ_i , à savoir le rapport de l'écart-type avec la moyenne.

3.1.2 Le rapport de distance

Définition 3.2 :

Soit $X = \{x^1, \dots, x^n\} \in [0,1]^n$ une distribution de n points en d dimensions.

On définit le rapport de la distance par :

$$R = \frac{\max_{i=1, \dots, n} \gamma_i}{\min_{i=1, \dots, n} \gamma_i}$$

Avec $\gamma_i = \min_{k \neq i} \text{dist}(x^i, x^k)$

Dans le cas où les points sont sur une grille régulière, on a $\gamma_1 = \dots = \gamma_i = \gamma$ et donc $R=1$. Ainsi, plus R est proche de 1, et plus les points sont proches de ceux d'une grille régulière.

3.1.3. Distances Maximin et Minimax

Johnson et al. [6] Ont introduit les distances Maximin et minimax afin de construire des plans répondant à la question de remplissage de l'espace. Le critère Maximin consiste à maximiser la distance minimale entre deux points du plan.

Définition 3.3 :

On définit les critères suivants au sens de la distance euclidienne :

$$MinDist = \min \min dist (x^i, x^k)$$

Et

$$avgDist = \frac{1}{n} \sum \min dist (x^i, x^k)$$

Où $X = \{x^1, \dots, x^n\}$ est le plan d'expériences à n points en d dimensions.

Maximiser le critère *MinDist* revient à considérer le critère de distance maximin. Si la valeur de *MinDist* est grande, alors il n'y a aucun point qui soit proche au sens de cette grandeur. Ainsi, plus les critères *MinDist* et *AvgDist* seront grands et plus cela traduira une distribution uniforme.

3.2 Critères d'uniformité basés sur la discrédance

On rappelle que la discrédance mesure l'écart entre une distribution de points donnée et une distribution de points uniforme ; en d'autres termes, elle mesure l'irrégularité de la distribution. On donne ci-dessous la définition permettant de comprendre le principe sur lequel elle est basée.

Contrairement aux deux critères précédents, la discrédance n'est pas basée sur la distance entre les points. Elle permet de mesurer l'écart entre la fonction de répartition empirique des points du plan et celle de la loi uniforme. Il existe différentes mesures de discrédance. Nous retenons la discrédance en norme L_2 .

3.2.1 Discrédance en norme 2

Seule la discrédance en norme L_2 est facilement calculable quelle que soit la dimension. En voici plusieurs. On note X la suite de n points x^1, \dots, x^n sur $[0,1]^d$. Cette section présente dans un premier temps la définition des différentes discrédances en norme 2 ainsi que leurs approches calculatoire.

3.2.1.1 Discrédance extrême et à l'origine :

Définition 3.4 : [6]

La discrédance L^2 d'une suite de n points x^1, \dots, x^n sur $[0,1]^d$ est définie par [6] :

$$DL2_n(X)^2 = \int_{[0,1]^{2d}} (D_n(J, X))^2 da db$$

Où J est l'ensemble des parties de sur $[0,1]^d$ de la forme $\prod_{i=1}^d [a_i, b_i]$

Définition 3.5 :

La discrédance L^2 à l'origine d'une suite de n points x^1, \dots, x^n sur $[0,1]^d$ est définie par [6] :

$$DL2_n^*(X)^2 = \int_{[0,1]^d} (D_n(J, X))^2 db$$

Où J est l'ensemble des parties de sur $[0,1]^d$ de la forme $\prod_{i=1}^d [0, b_i]$

En dimension d , les calculs de $DL2_n(X)^2$ et $DL2_n^*(X)^2$ peuvent se faire par les formules explicites suivante [6] :

$$\begin{aligned} DL2_n(X)^2 &= \left(\frac{1}{12}\right)^d \\ &\quad - \frac{2^{1-d}}{n} \sum_{i=1}^n \prod_{j=1}^d (1 - x_j^i) \\ &\quad + \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \prod_{j=1}^d (1 - \max(x_j^i, x_j^k)) \min(x_j^i, x_j^k) \end{aligned}$$

et

$$\begin{aligned} DL2_n^*(X)^2 &= \left(\frac{1}{3}\right)^d \\ &\quad - \frac{2^{1-d}}{n} \sum_{i=1}^n \prod_{j=1}^d (1 - (x_j^i)^2) \\ &\quad + \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \prod_{j=1}^d (1 - \max(x_j^i, x_j^k)) \end{aligned}$$

3.2.1.2 Discrépance modifiée :

Définition 3.6 :

La discrétance L_2 modifié d'une suite de n points x^1, \dots, x^n sur $[0,1]^d$ définie par :

$$DL2_n^M(X)^2 = \sum_{u \neq \emptyset} \int_{[0:1]} (D_p(J_u, X))^2 db_u$$

Avec $[0,1]^u$ la projection de l'hypercube unité sur les composantes u étant un sous-ensemble de $\{1, \dots, d\}$, $p = \text{Card}(u)$ et J_u est la projection du sous-ensemble J de la forme $\prod_{i=1}^d [0, b_i]$.

La discrétance modifiée considère les projections sur tous les sous-espaces et est définie par :

$$DL2_n^M(X)^2 = \left(\frac{3}{4}\right)^d - \frac{2^{1-d}}{n} \sum_{i=1}^n \prod_{j=1}^d (3 - (x_j^i)^2) + \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \prod_{j=1}^d (2 - \max(x_j^i, x_j^k)).$$

3.2.1.3 Discrépance centrée

Définition 3.7 :

La discrédance L_2 centrée d'une suite de n points x^1, \dots, x^n sur $[0,1]^d$ définie par :

$$DL_2^c(X)^2 = \sum_{u \neq \emptyset} \int_{[0,1]^u} (D_p(J_u, X))^2 dx_u$$

Avec $[0,1]^u$ la projection de l'hypercube unité sur les composantes u étant un sous-ensemble de $\{1, \dots, d\}$, $p = \text{Card}(u)$ et où J_u est la projection du sous-ensemble construit à partir du point considéré x et de son sommet le plus proche.

Remarque 3.2 :

En dimension 2, l'ensemble J_u peut être de 4 formes différentes dont un est représenté sur la Figure 3.1 . Plus généralement, en dimension d , il existe $2d$ cas.

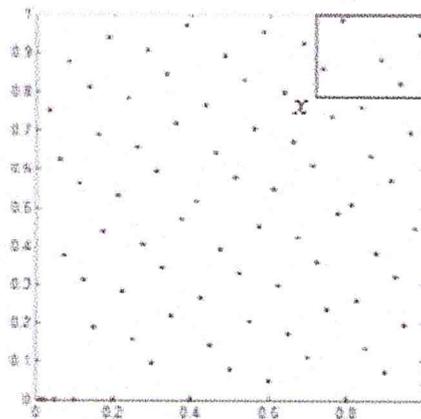


Figure 3.1 : un sous-ensemble J pour le calcul de la discrédance centrée.

Hickernell (1998) donne une expression analytique de la discrédance centrée :

$$\begin{aligned}
DL2_n^c(X)^2 &= \left(\frac{13}{12}\right)^2 - \frac{2}{n} \sum_{i=1}^n \prod_{j=1}^d \left(1 + \frac{1}{2} |x_j^i - 0.5| - \frac{1}{2} |x_j^i - 0.5|^2\right) \\
&+ \frac{2^d}{n^2} \sum_{i=1}^n \sum_{k=1}^n \prod_{j=1}^d \left(1 + \frac{1}{2} |x_j^i - 0.5| + \frac{1}{2} |x_j^k - 0.5| \right. \\
&\quad \left. - \frac{1}{2} |x_j^i - x_j^k|\right)
\end{aligned}$$

3.2.1.4 Discrépance symétrique

Définition 3.8 :

La discr pance L_2 sym trique d'une suite de n points x^1, \dots, x^n sur $[0,1]^d$ d finie par [6] :

$$DL2_n^s(X)^2 = \sum_{u \neq 0} \int_{[0,1]^u} (D_p(J_u, X))^2 dx_u$$

Avec J_u la projection de l'intervalle J sur le sous-espace d fini par les composantes u o  J est la r union des sous-ensembles sym triques i.e. dont la somme des coordonn es des sommets sont  gales.

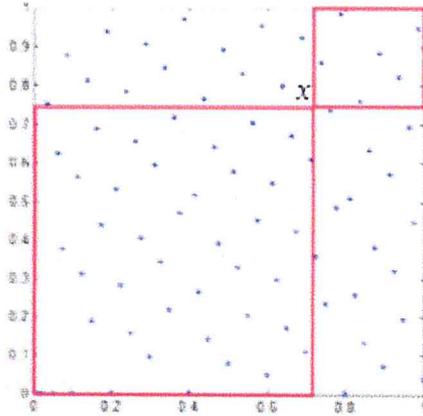


Figure 3.2 : pour $x = (0.7, 0.75)$, le volume total des 2 sous-ensembles J est de 0.6

On dispose également d'une formule analytique pour calculer cette discr pance [6] :

$$DL2_n^s(X)^2 = \left(\frac{3}{4}\right)^d - \frac{2}{n} \sum_{i=1}^n \prod_{j=1}^d (1 + 2x_j^i - 2(x_j^i)^2) + \frac{2^d}{n^2} \sum_{i=1}^n \sum_{k=1}^n \prod_{j=1}^d (1 - |x_j^i - x_j^k|).$$

CHAPITRE 4

IMPLEMENTATION ET COMPARAISON

Dans ce dernier chapitre, nous allons procéder à l'application numérique de quelques suites à faible discrédance (SFD), et les résultats expérimentaux obtenus lors des comparaisons entre les plans SFD en utilisant les bases des critères qui sont déjà présentées dans le chapitre précédant.

La programmation de ces suites et la simulation se fera avec le logiciel MATLAB.

4.1 Logiciel Matlab

4.1.1 Présentation du logiciel MATLAB

De nombreux logiciels interactifs de calcul scientifique existent et sont disponibles sur le marché. Parmi ceux-ci, MATLAB est un produit très intéressant pour développer des calculs numériques. Il intègre un environnement graphique et ses atouts principaux sont : $\frac{3}{4}$ La surprenante simplicité d'utilisation provenant de la vectorisation, $\frac{3}{4}$ La puissance de calcul, $\frac{3}{4}$ La précision des résultats. MATLAB est un outil de base enrichi par ses nombreuses bibliothèques ("toolboxes") qui évitent, dans de nombreux cas, de programmer les techniques numériques les plus courantes.

MATLAB (matrix laboratory) est un langage de programmation de quatrième génération et un environnement de développement ; il est utilisé à des fins de calcul numérique. Développé par la société The MathWorks, MATLAB permet la manipulation de matrice, afficher des courbes et des données, mettre en œuvre des algorithmes, créer des interfaces utilisateurs, et peut s'interfacer avec d'autres langages comme le C, C++, Java, et Fortran. Les utilisateurs de MATLAB (environ un million en 2004) sont de milieux très différents comme l'ingénierie, les sciences et l'économie dans un contexte aussi bien industriel que la recherche.

Matlab peut s'utiliser seul ou bien avec des toolbox (boite à outils)

4.1.2 Interfaces graphiques Matlab

Offre la possibilité de réaliser des Interfaces Graphiques (ou GUI Graphical User Interface) très facilement à l'aide de deux fonctions de bases :

Uicontrol : pour créer dans la fenêtre des boutons, des zones d'affichages, des graphiques, etc ...
et Uimenu : pour créer des menus supplémentaires à ceux de bases : Fichier (File), Edition (Edit), etc ...

La conception de ces interfaces est rendue encore plus simple grâce à la fonction "GUI LayoutTool" du menu fichier.

4.1.2.1 Création d'une interface graphique sous MATLAB :

Une interface graphique permet de contrôler une application interactivement avec la souris, plutôt que par lancement des commandes au clavier.

Une interface graphique comprend des menus, des boutons, des "ascenseurs", des cases à cocher, des listes de choix, des zones de texte.

Première étape : concevoir l'interface

Pour créer une interface graphique, il faut d'abord commencer par un travail de réflexion pour obtenir un croquis

Deuxième étape : coder l'interface

Il existe deux façon de développer les interfaces graphiques : avec le GUIDE ou par programmation (en ligne de code).

4.1.3 Outils guide

Le GUIDE est un constructeur d'interface graphique qui regroupe tous les outils dont le programmeur a besoin pour créer une interface graphique de façon intuitive.

Il s'ouvre, soit en tapant guide dans la commande Windows de MATLAB.

Le placement des objets st réalisé par sélection dans une boite à outils .leur mise en place et leur dimensionnement se font à l'aide de la souris.

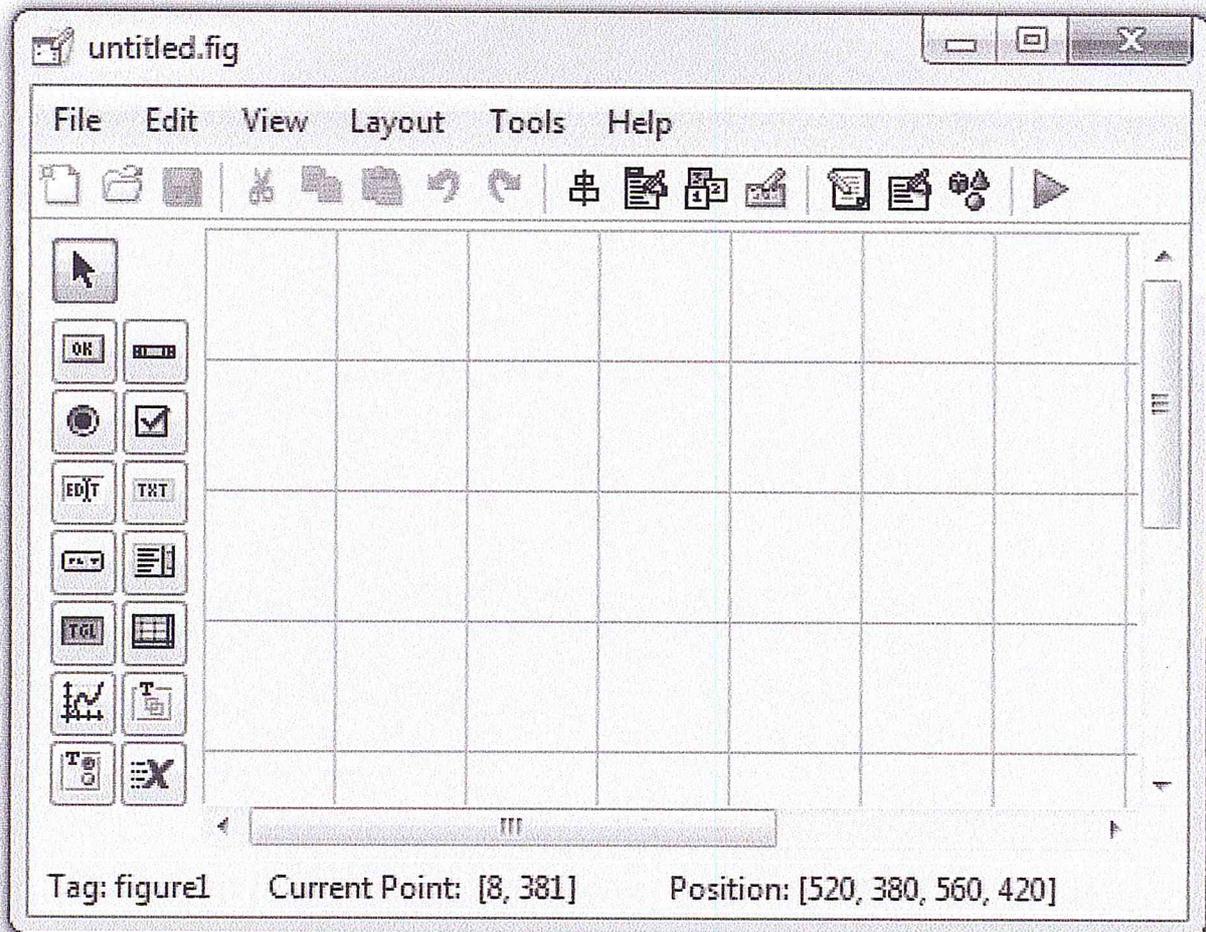


Figure 4.1 : fenêtre principale du Guide

Un double-clic sur un objet permet de faire apparaître de property inspector où les propriétés des objets sont facilement éditables. Leurs modifications et la visualisation de ces modifications sont immédiates.

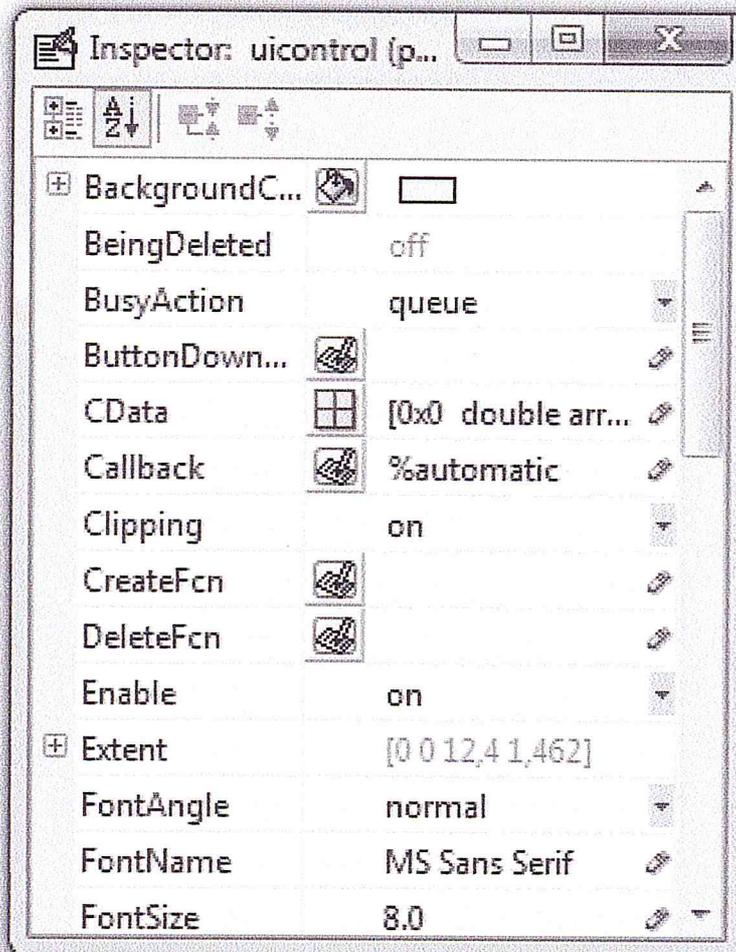


Figure 4.2 : property inspector

Le GUIDE possède également des outils pour gérer l'alignement des objets et pour créer des barres d'outils ou des menus.

Une fois l'interface graphique terminée, son enregistrement donne deux fichiers portant le même nom mais dont les deux extensions sont .fig et .m.

4.2 L'Etude de comparaison

L'objectif de ce chapitre est de comparer la qualité des distributions de points au sens du remplissage de l'espace à l'aide des critères étudiés jusqu'ici.

Le principe consiste à appliquer les critères intrinsèques de chapitre deux à différents plans en dimension 2, puis en dimension 3 et 10, afin d'être dans le contexte des plans d'expériences numériques.

Pour ce faire, on fait un appel aux critères des distances suivant :

- La mesure de recouvrement
- La mesure R
- Mindist
- Maxdist

Les plans confrontés dans cette partie sont généralement les suivants :

- Plans Aléatoire
- Suite de Sobol
- Suite de Faure
- Suite de Halton
- Suite de Hammrely
- Hypercube Latins

4.3 Résultats

Dans cette partie, nous présentons les tables regroupant les résultats d'implémentation des différents algorithmes pour les différents plans considérés. Ces tables sont divisées selon le nombre de points et la dimension de l'espace expérimental.

Les résultats obtenus sont résumés dans les tableaux suivant :

4.3.1 Résultat pour 10 points et 2 dimensions

	Aléatoire	LHS	Sobol	Halton	Hammersley
Mesure de recouvrement	0.5353	0.2068	0.3094	0.2526	0.1096
Le rapport R	7.0170	1.9067	2.6064	1.7990	1.3575
MinDist	0.0601	0.1585	0.1672	0.1419	0.2358
MaxDist	0.4219	0.3022	0.4359	0.2552	0.3202

Tableau 4.1 : les valeurs obtenues des différents plans en utilisant différents critères pour $n = 10$ et $d = 2$.

4.3.2 Résultat pour 20 points et 3 dimensions

	Aléatoire	LHS	Sobol	Halton	Hammersley
Mesure de recouvrement	0.4007	0.2518	0.2876	0.2925	0.2149
Le rapport R	5.3454	2.3135	2.8562	2.2814	2.0433
MinDist	0.0797	0.1785	0.1720	0.1562	0.2199
MaxDist	0.4263	0.4131	0.4912	0.3564	0.4493

Tableau 4.2 : les valeurs obtenues des différents plans en utilisant différents critères pour $n = 20$ et $d = 3$.

4.3.3 Résultat pour 50 points et 7 dimensions

	Aléatoire	LHS	Sobol	Halton	Hammersley
Mesure de recouvrement	0.2243	0.1655	0.1521	0.2711	0.1402 ^
Le rapport R	2.3118	1.7935	2.0504	4.0253	1.6411
MinDist	0.3374	0.4114	0.3759	0.1907	0.4418
MaxDist	0.7799	0.7378	0.7706	0.7676	0.7250 v

Tableau 4.3 : les valeurs obtenues des différents plans en utilisant différents critères pour $n = 50$ et $d = 7$.

4.3.4 Résultat pour dimensions supérieur à 10

- Résultats pour 200 points et 15 dimensions

	Aléatoire	LHS	Sobol	Halton	Hammersley
Mesure de recouvrement	0.1229	0.1026	0.2182	0.1221	0.2229 ^
Le rapport R	2.0519	1.6774	2.2901	2.0470	2.2695
MinDist	0.5835	0.6911	0.5108	0.6032	0.5103
MaxDist	1.1973	1.1594	1.1697	1.2347	1.1582 v

Tableau 4.4 : les valeurs obtenues des différents plans en utilisant différents critères pour $n = 200$ et $d = 15$.

- Résultats pour 400 points et 20 dimensions

	Aléatoire	LHS	Sobol	Halton	Hammersley
Mesure de recouvrement	0.0982	0.0996	0.2819	0.0976	0.2811 ^
Le rapport R	1.8347	1.7674	2.6607	1.8259	2.6107
MinDist	0.7888	0.8131	0.5121	0.7916	0.5109
MaxDist	1.4472	1.4371	1.3624	1.4454	1.3339 v

Tableau 4.5 : les valeurs obtenues des différents plans en utilisant différents critères pour $n = 400$ et $d = 20$.

4.4 Interprétation

On rappelle que :

Plus la mesure de recouvrement est faible, et plus les points sont proches de ceux d'une grille régulière, ainsi que le R rapport est de proche de 1, et plus les points sont proches de ceux d'une grille régulière

On peut remarquer d'après les tableaux obtenus la suite de hammersley atteint une valeur minimale la mesure de recouvrement et le rapport R pour la dimension inférieur ou égale à 10.

Par conséquent les suite à faible discrèpance permet d'améliorer considérablement la qualité.

Donc les tableaux montre que pour la suite de hammersley et hypercube latin donne les meilleurs résultats, et par contre des mauvaise pour un plan aléatoire, ceci implique que la mesure de distance seule ne suffit pas à mesure la qualité des points d'un plans pour dimension élevés , et à noter que cette remarque est valable pour n'importe quel nombre de points. Ainsi, il est nécessaire de choisir plusieurs critères afin de mesurer la bonne répartition des points dans l'hypercube unité.

4.5 Conclusion

En grande dimension la pluparts des plans d'expériences convergentes vers la zone aléatoire, les suites classiques à faible discèpance ne permettent pas d'obtenir des bonnes répartitions de points dans l'espace en grande dimensions.

Considérons à des critères bases valeur sur les mesures de distance pour les différents plans, On peut constater que l'ensemble des plans d'expériences SFD (et par conséquence les suites à faible discèrapnce) se situe dans la zone des distributions quasi-périodiques.

Les résultats obtenus avec ces critères ont permis de qualifier sur ces différents types de plans et d'évaluer leur évolution en fonctions de la dimension

CONCLUSION GENERALE

La Méthodologie de la Recherche Expérimentale, appelée aussi planification des expériences, est une discipline mathématique faisant partie de la statistique inférentielle. La théorie de la méthode des plans d'expériences reste toujours en développement.

Le développement de codes qui modélisent ou simulent des phénomènes complexes sont de plus en plus réalistes et même si la puissance des ordinateurs augmente sans cesse, les temps de calcul demeurent importants et limitent ainsi le recours aux techniques de Monte-Carlo.

Dans le domaine de l'expérimentation numérique, lorsque les relations entre la réponse et les entrées du code de calcul sont complexes, les plans d'expériences SpaceFilling Designs (SFD) sont utilisés pour l'exploration du code ou la construction de métamodèles.

Le développement de métamodèles permet alors de remplacer le simulateur par un outil plus "simple" construit à partir du simulateur complexe. Ces métamodèles ou surfaces de réponse sont en général des fonctions obtenues à l'aide de méthodes d'interpolation ou d'approximation à partir d'un Nombre limité d'exécutions du simulateur sur des jeux de paramètres constituant le plan d'expériences numériques. Du fait des caractéristiques non linéaires et/ou non paramétriques des codes de modélisation ou simulation, il est nécessaire de répartir les points dans l'espace le plus uniformément possible de façon à capter au mieux le comportement du simulateur. C'est le mode de répartition des points que cherchent à proposer les plans d'expériences Space Filling Designs (SFD).

L'étude de l'uniformité d'une distribution de points est difficile en grande dimension et nécessite l'utilisation de plusieurs critères tels la discrédance et les critères basés sur des calculs de distances entre les points. (Nous avons étudié les suites Van Der Corput, Halton, Sobol, Faure qui font partie des familles de suites à faible discrédance).

Cette étude a permis de montrer l'intérêt des Space-Filling Designs en phase exploratoire.

En conclusion, ces travaux permettent de proposer des pistes de recherches pour relier les qualités intrinsèque et extrinsèque des plans SFD.

Références

- [1] Jacques Goupy, "Plans d'expériences pour surface de réponse", Dunod, Paris (2006)
- [2] Jacques Goupy, "Introduction aux plans d'expériences", Dunod 3^e édition (2006).
- [3] Jacques Goupy, "Introduction aux plans d'expériences", Dunod 5^e Edition.
- [4] Goupy, J., "Étude comparative de divers plans d'expériences", revue de statistique appliquée, vol.38(1990).
- [5] Mathieu, Phan-Tan-Luu, R., "Planification d'expériences en formulation", Techniques de l'Ingénieur, traité Génie des procédés(2000).
- [6] Jessica Franco, "Planification d'expériences numériques en phase exploratoire pour la simulation des phénomènes complexes", thèse d'état, l'École Nationale Supérieure des Mines - Saint-Etienne (2012).
- [7] Dreesbeke, J.J., Fine, Saporta, G., "Plans d'expériences. Applications à l'entreprise", édition Technip(1997).
- [8] H. Faure, Discrepance de suites associées à un système de numération (en dimensions), Acta Arith. 41 (1982) pp. 337–351.
- [9] J.H. Halton, "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals", mathematika 2 (1960) pp. 84–90.
- [10] Gille, "Plans d'expériences en gestion industrielle", Lavoisier, Paris (2003).
- [11] Suzanne Varet, "Développement de méthodes statistiques pour la prédiction d'un gabarit de signature infrarouge" (2010).

