



**HAL**  
open science

## Un algorithme d'optimisation par exploration sélective

Didier Rullière, Alaeddine Faleh, Frédéric Planchet

► **To cite this version:**

Didier Rullière, Alaeddine Faleh, Frédéric Planchet. Un algorithme d'optimisation par exploration sélective. 2009. hal-00411406v1

**HAL Id: hal-00411406**

**<https://hal.science/hal-00411406v1>**

Preprint submitted on 27 Aug 2009 (v1), last revised 18 Nov 2010 (v2)

**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.

# Un algorithme d'optimisation par exploration sélective<sup>1</sup>

Didier Rullière<sup>2</sup>, Alaeddine Faleh<sup>3</sup> et Frédéric Planchet<sup>4</sup>

## Résumé

Cet article propose un algorithme qui vise à résoudre les problèmes d'optimisation dans le cas où la fonction objectif est estimée à l'aide de simulations stochastiques, en présence de contraintes donnant a priori le domaine d'optimisation de cette fonction. L'algorithme traite ainsi un problème d'optimisation globale d'une fonction non convexe, bruitée, issue de simulations. Il est construit après une étude de critères de compromis entre, d'une part, l'exploration de la fonction objectif en de nouveaux points et d'autre part l'amélioration de la connaissance de celle-ci à travers l'augmentation du nombre de tirages en des points déjà explorés. L'article présente enfin une application numérique illustrative de la conformité du comportement de cet algorithme à celui prévu théoriquement.

**mots clés** : simulation stochastique, optimisation stochastique, zone de recherche, exploration, connaissance, allocation optimale.

## 1 Introduction

**le problème d'optimisation** Nous allons chercher à déterminer le ou les optimums globaux d'une fonction définie d'un sous-ensemble de  $\mathbb{R}^d$  et à valeurs dans  $\mathbb{R}$ , soumise à un bruit, et dont le temps de calcul est coûteux. Supposons que l'on cherche à trouver l'optimum d'une fonction réelle  $f(\theta)$  d'un paramètre  $\theta \in \Theta$ ,  $\Theta \subset \mathbb{R}^d$  (en minimisant par exemple un indicateur synthétique de compromis risque/gain pour une allocation d'actif donnée, ou en maximisant un indicateur de gain pénalisé en fonction du risque). Nous supposons par la suite que  $\Theta$  est une union finie de  $d$ -simplexes.

Il est des cas où la fonction  $f$  n'est pas directement connue, mais est estimée à l'aide de simulations stochastiques. Dans ce cas, les données disponibles sont  $n$  réalisations d'une variable aléatoire réelle  $F(\theta)$  d'espérance  $f(\theta)$ , opérées sur un ensemble  $\Theta_m$  de  $m$  paramètres distincts  $\theta_1, \dots, \theta_m$ . Dans ce cas, le problème revient à trouver un estimateur  $\hat{\theta}^*$  de :

$$\theta^* = \arg \min_{\theta \in \Theta} E [F(\theta)] .$$

Nous supposons que le tirage de chaque valeur de  $F$  est très coûteux en terme de temps de calcul. Par exemple, s'il s'agit de maximiser un ratio de financement dans le contexte de la détermination de l'allocation stratégique optimale d'actifs, une allocation particulière  $\theta$  peut nécessiter de simuler de nombreuses trajectoires de plusieurs actifs, sur de nombreuses périodes, afin de déterminer un unique indicateur de risque ou de gain bruité  $F(\theta)$ . Une très large littérature existe sur les algorithmes stochastiques pour la recherche d'optimum. Lorsque  $d = 1$  des algorithmes tels que celui de Kiefer-Wolfowitz (voir Kiefer, Wolfowitz, 1952) peuvent

---

1. Ce travail a bénéficié d'une aide de l'Agence Nationale de la Recherche portant la référence ANR-08-BLAN-0314-01, et d'une aide de l'ANRT portant la référence 177/2008.

2. Université de Lyon, Université Lyon 1, Laboratoire SAF, Ecole ISFA.

Auteur correspondant : Didier.Rulliere@univ-lyon1.fr, (+33) 4 37 28 74 38, UCBL, ISFA, 50 Avenue Tony Garnier, F-69007 Lyon, France.

3. Université de Lyon, Université Lyon 1, Laboratoire SAF, Ecole ISFA ; Caisse des Dépôts et Consignations

4. Université de Lyon, Université Lyon 1, Laboratoire SAF, Ecole ISFA ; Winter & Associés

être utilisés. Dans le cas  $d > 1$ , une extension due à Blum peut être exploitée (voir Blum, 1954). Strugarek (2006) présente un ensemble d’algorithmes plus récents et plus détaillés portant sur l’optimisation stochastique. S’agissant de la recherche de racines, que nous pourrions également traiter avec l’algorithme ici présenté, des techniques classiques sont celles dérivées de l’algorithme de Robbins-Monro (Robbins, Monro, 1951). De même, l’algorithme de Cohen, Culioli permet de résoudre des problèmes très proches avec une contrainte de contrôle de la probabilité de ruine (voir Cohen, Culioli, 1994). Nous ne détaillerons pas ici l’ensemble des techniques utilisées dans les champs de l’optimisation stochastique (recherche d’optimum), ou de l’approximation stochastique (recherche de racines).

**le risque de se tromper d’optimum** Tout comme les méthodes déterministes de descente de gradient, certains de ces algorithmes peuvent requérir d’appréhender dans un premier temps la forme de la fonction  $f$ , afin de choisir un point de départ suffisamment proche de l’optimum, et d’éviter les plus évidents optima locaux. Une exploration globale de la fonction s’avère donc souvent nécessaire, le risque étant moins de mal estimer la valeur d’un optimum global, que de se tromper d’optimum en choisissant indûment un mauvais optimum local. Un des objectifs poursuivis est donc l’exploration des zones susceptibles de contenir le minimum, puis la proposition d’un estimateur du minimum global (et non la proposition directe d’un minimum local). Nous chercherons donc à explorer sélectivement la fonction  $f$ , de façon à avoir une vision globale de celle-ci, tout en privilégiant certaines zones d’intérêts, comme les minimums locaux, dans un souci global d’économie du nombre de tirages de la fonction  $F$ .

En supposant donnée la valeur  $n$ , qui détermine la précision de la connaissance de  $f$ , et  $\Theta_m$  qui détermine l’ampleur de l’exploration de  $f$ , un estimateur de  $\theta^*$  envisageable est le suivant :

$$\begin{aligned}\widehat{\mu}_n(\theta) &= \frac{1}{n} \sum_{i=1}^n F_i(\theta), \quad \theta \in \Theta_m \\ \widehat{\theta}_{n,m}^* &= \arg \min_{\theta \in \Theta_m} \widehat{\mu}_n(\theta)\end{aligned}$$

La question se pose ici de l’arbitrage entre le choix d’un  $n$  élevé ou d’un cardinal de  $\Theta_m$  élevé. Par la suite, nous ferons également dépendre  $n$  du point  $\theta$  considéré.

**explorer ou connaître** Les algorithmes usuels d’optimisation supposent généralement que l’ensemble des points  $F(\theta)$  sont connus, et ignorent, à notre connaissance, la question pratique de l’optimisation du couple de paramètre  $(n, m)$ . En effet, la détermination d’une réalisation de la variable aléatoire  $F$  peut prendre un temps de calcul important,  $n$  réalisations de  $F$  conduisant, en chaque point de  $\Theta_m$ , à  $n \cdot m$  réalisations de  $F$  en différents points, et à une complexité de la détermination de ces tirages que l’on pourra supposer proportionnelle à  $n \cdot m$ . Pour une complexité fixée, le choix d’un nombre de simulations  $n$  important réduit la variance de  $F$ , et donc en améliore la connaissance, mais empêche également d’explorer la fonction pour d’autres paramètres. Il s’ensuit par exemple que la variance de l’estimateur de  $\theta^*$  pourra varier en fonction de  $m$  et de  $n$ , ce qui introduit la problématique suivante. Nous recherchons un optimum de l’espérance d’une fonction aléatoire de  $\mathbb{R}^d$  dans  $\mathbb{R}$ , dont on ne peut obtenir que des tirages coûteux en terme de temps de calcul. Supposons que l’espérance en question ne soit pas nécessairement convexe, de sorte que la fonction puisse posséder plusieurs optimums locaux (il peut s’agir, par exemple, d’un indicateur de gain pénalisé en fonction du risque à maximiser, d’un indicateur de compromis risque/gain, etc.). Comme nous l’avons vu, trouver ces optimums va requérir d’explorer la fonction en différents points d’une part (nous parlerons d’exploration), et d’opérer également plusieurs tirages de la fonction en chaque point

pour obtenir un estimateur aussi peu dispersé que possible de l'espérance de la fonction (nous parlerons de connaissance ponctuelle). Les tirages étant coûteux et leur nombre étant limité, la question se pose de l'arbitrage entre l'exploration de la fonction en différents points et la connaissance ponctuelle de celle-ci. Nous proposerons dans cet article un algorithme qui vise à répondre à cet arbitrage.

Par la suite, nous nous placerons dans le cadre de la minimisation d'une fonction, et nous désignerons par :

- $\mu(\theta)$  un estimateur de  $f(\theta)$ .
- $\mu^-(\theta)$  un estimateur de la borne inférieure de l'intervalle de confiance de  $f(\theta)$ .
- $m^*$  un estimateur du minimum global de  $f(\theta)$ .

Nous n'évoquerons pas ici la détermination très classique de ces estimateurs. Pour la borne inférieure de l'intervalle de confiance, il est possible d'utiliser de simples quantiles. Une alternative possible est l'usage d'une hypothèse de répartition gaussienne, qui peut permettre de ne pas mémoriser l'intégralité des tirages de  $F(\theta)$ , mais de simplement mémoriser le nombre de tirages précédents, leur somme et la somme de leurs carrés.

La maximisation d'une fonction peut s'opérer aussi naturellement par minimisation de la fonction opposée, et la recherche de racine peut également être facilement adaptée de l'algorithme présenté, en ne considérant plus l'atteinte possible par  $f$  de l'intervalle  $]-\infty, m^*]$ , mais l'atteinte possible d'un intervalle  $[-h, +h]$ , pour un  $h$  donné supposé petit.

**grille à pas fixe** Une solution simple et très commune est l'utilisation d'une grille de simulation, à pas fixe. Selon cette solution,  $\Theta_m$  est une grille de pas  $\delta$  (toutes les composantes de  $\theta \in \Theta_m$  parcourent l'ensemble des valeurs multiples de  $\delta$  dans  $\Theta$ ), et  $n$  est identique pour chaque  $\theta \in \Theta_m$ . Cette solution est néanmoins très onéreuse en terme de temps de calcul, dans la mesure où  $n$  simulations seront conduites sur chacun des points, y compris ceux très éloignés de l'optimum, pour lesquels la meilleure connaissance de  $F$  n'apporte quasiment rien. Nous chercherons donc à mettre en place une grille à pas variable, où les retirages de  $F$  se feront principalement dans les zones susceptibles d'accueillir le minimum.

**grille à pas variable** L'idée d'une grille à pas variable est relativement simple. Il va s'agir d'envisager d'une part de nouveaux paramètres  $\theta$  (sommets) pour lesquels nous estimerons  $f(\theta)$ , et d'autre part d'envisager la répartition de nouvelles simulations entre les anciens sommets et les nouveaux, en fonction de l'intérêt que peut avoir, sur la connaissance du minimum global, l'ajout de simulation en chacun de ces points. Seront ainsi modifiées les positions des sommets envisagés, qui ne seront plus régulièrement répartis, ainsi que les nombres de simulations conduites en chaque sommet, qui vont différer selon les sommets. Une première difficulté est le choix d'une forme convenable pour les zones de recherche, pour les cellules de la grille. Une difficulté est que l'on a besoin de deviner où l'on va conduire les futures simulations, mais pour cela, il faut avoir une idée de comment va évoluer la fonction aux points interpolés, sans pour autant conduire les simulations : il faut en un sens deviner quel pourra être l'impact de futures simulations avant même de les réaliser.

La connaissance du minimum global dépend principalement de deux éléments :

- la bonne connaissance de  $F$  dans un voisinage de l'estimateur du minimum  $\hat{\theta}^*$
- l'assurance que cet estimateur  $\hat{\theta}^*$  correspond bien à un minimum global, et que le véritable minimum ne se trouve pas dans une autre zone non explorée.

Comme l'on peut aisément l'imaginer, en dimension  $d = 1$ , l'augmentation du nombre de simulations aux sommets connus a relativement peu d'incidence sur la connaissance du minimum lorsque les intervalles de confiance de  $f(\theta)$  en chaque sommet connu  $\theta$  sont très

éloignés du minimum estimé. Au contraire, on imagine une forte incidence lorsque les intervalles de confiance franchissent ce minimum connu.

Compte tenu des informations partielles dont on dispose, la position du minimum peut être envisagée à plusieurs endroits, de sorte que l'on peut supposer une distribution possible de ce minimum sur  $\Theta$ . Dans l'idéal, la répartition des simulations doit conduire à minimiser la dispersion de la position possible de ce minimum. La minimisation de cette dispersion, par l'ajout de simulations en certains sommets, revient à choisir lesquels des intervalles de confiance on va réduire.

Notons par ailleurs l'analogie de la démarche générale avec les algorithmes génétiques, qui reposent sur des règles de croisements et mutations dont la logique est très semblable à celle d'*explorer ou connaître* : la mutation dans ce type d'algorithmes a pour but d'explorer l'espace d'états. Une présentation détaillée de ces techniques est proposée dans Alliot (1996).

**algorithmes présentés et structure du document** Nous allons aborder dans cet article deux algorithmes présentant la réalisation d'une grille à pas variable. Dans la section 2, nous envisagerons le cas où le nombre de tirages en chaque point est fixe, l'algorithme cherchant alors uniquement à déterminer les meilleurs points d'exploration possibles pour la fonction. Dans la section 3, nous introduirons la possibilité d'opérer des retirages en des points déjà explorés, afin d'améliorer l'estimation de l'espérance de la fonction objectif en ces points. Enfin, dans une dernière section 4, nous présenterons des illustrations simples et visuelles du comportement de l'algorithme proposé sur une fonction bimodale élémentaire. Le comportement de l'algorithme dans le cas de la recherche de racine sera également illustré dans cette section.

## 2 Une grille à pas variable par subdivision systématique

### 2.1 forme des zones de recherche

En pratique, par exemple lors de l'optimisation de choix d'investissements de  $d + 1$  actifs, les proportions investies dans chacun des actifs se sommant à un, on est amené à optimiser une fonction de  $\mathbb{R}^d$  dans  $\mathbb{R}$ . Nous évoquerons le cas où davantage de contraintes façonnent le domaine de recherche au paragraphe 2.2.

L'idée d'une "grille à pas variable" est de séparer la zone de recherche de l'optimum en plusieurs zones de différentes tailles. La création d'une nouvelle zone peut se faire en opérant des simulations pour de nouveaux paramètres de la zone, en ajoutant de nouveaux points de  $\Theta$ . Nous supposons que les zones de recherche sont des ensembles convexes, de façon à faciliter d'éventuelles interpolations au cœur de chaque zone. Par exemple, imaginons un pavé de  $\mathbb{R}^d$  dont on connaît les sommets, et à l'intérieur duquel on ajouterait un point  $\theta_c$ . Comment découper rapidement ce pavé en une partition d'ensembles convexes dont l'enveloppe convexe contiendrait  $\theta_c$ ? La réponse n'étant pas si évidente, nous opterons pour le choix suivant :

- Les zones considérées seront délimitées par  $d + 1$  sommets.
- Chaque nouveau point sera ajouté sur un unique segment de l'enveloppe convexe de la zone.
- A chaque étape, chaque zone sera fractionnée en deux ensembles convexes ou laissée telle quelle.

En effet, de par le théorème de Carathéodory, tout point à l'intérieur d'une zone appartiendra à une zone convexe délimitée par  $d + 1$  points, et le choix d'un nombre de sommets égal à  $d + 1$  facilitera par la suite la séparation d'une zone en plusieurs zones. D'autre part, le choix d'un

nombre de sommets fixes délimitant chaque zone sera de nature à faciliter l'implémentation de l'algorithme.

Nous appellerons zone un  $d$ -simplexe, c'est-à-dire un ensemble convexe inclus dans  $\Theta$ , avec  $\Theta \subset \mathbb{R}^d$ , dont l'enveloppe convexe est déterminée par  $d + 1$  sommets distincts (les sommets formant un repère affine de  $\mathbb{R}^d$ ). La zone initiale de recherche de l'optimum est notée  $Z_0$ . A l'étape numéro  $k$ , la zone de recherche est subdivisée en un ensemble de zones recouvrant  $Z_0$ , dont l'intersection est de mesure nulle. Cet ensemble de zones est noté  $\mathfrak{Z}_k$ . Le nombre de zones à l'étape  $k$  est noté  $\nu_k$ .

Considérons un ensemble de sommets  $E$  délimitant une zone  $Z$  et deux points distincts  $\theta_1$  et  $\theta_2$  de cet ensemble. Supposons que  $\theta_c$  soit le barycentre (par exemple équipondéré) entre ces deux points. Lorsque la décision est prise de scinder cette zone en deux zones  $Z_1$  et  $Z_2$ , les ensembles de sommets délimitant ces deux zones seront :

$$\begin{aligned} E_1 &= (E \setminus \theta_1) \cup \{\theta_c\} \\ E_2 &= (E \setminus \theta_2) \cup \{\theta_c\} \end{aligned}$$

Cela revient à créer une nouvelle zone d'enveloppe  $E_2$ , et à remplacer  $\theta_1$  par  $\theta_c$  pour la zone en cours  $Z$ .

**Lemme 2.1 (lemme de séparation)** *Les zones  $Z_1$  et  $Z_2$  comportent  $d + 1$  sommets, sont convexes, recouvrent  $Z$ , et leur intersection forme un hyperplan de  $\mathbb{R}^d$ .*

**Preuve :**

Soit  $P$  un point de  $Z$ , on a  $P = \sum_{j \in J} w_j \theta_j = (\sum_{j \in J, j \neq j_1, j \neq j_2} w_j \theta_j) + w_{j_1} \theta_{j_1} + w_{j_2} \theta_{j_2}$ . En posant  $\theta_c = \frac{1}{2}(\theta_{j_1} + \theta_{j_2})$ , on peut écrire

– Si  $w_{j_1} \geq w_{j_2}$ , alors nécessairement  $w_{j_2} < \frac{1}{2}$  et

$$P = \sum_{j \in J} w_j \theta_j = \sum_{j \in J, j \neq j_1, j \neq j_2} w_j \theta_j + (w_{j_1} - w_{j_2}) \theta_{j_1} + 2w_{j_2} \theta_c.$$

Alors  $P \in Z_2$ .

– Si  $w_{j_2} \geq w_{j_1}$ , alors nécessairement  $w_{j_1} < \frac{1}{2}$  et

$$P = \sum_{j \in J} w_j \theta_j = \sum_{j \in J, j \neq j_1, j \neq j_2} w_j \theta_j + (w_{j_2} - w_{j_1}) \theta_{j_2} + 2w_{j_1} \theta_c.$$

Alors  $P \in Z_1$ .

Il s'ensuit que  $Z \subset Z_1 \cup Z_2$ . Par ailleurs, il est clair que  $Z_1 \cup Z_2 \subset Z$ . Enfin, on établit que  $Z_1 \cap Z_2$  est l'ensemble des points  $P$  tels que  $\alpha_{j_1} = \alpha_{j_2}$ , ce qui définit un hyperplan de  $\mathbb{R}^d$ .  $\square$

Par souci de simplicité, nous nommerons par la suite *sommets* d'une zone  $Z$  l'ensemble des  $d + 1$  sommets délimitant l'enveloppe convexe de la zone  $Z$ .

La figure 1 illustre en dimension 2 la façon dont une zone peut être progressivement subdivisée.

## 2.2 zone initiale de recherche

Nous nommerons  $Z_0$  la zone initiale de recherche, de sommets  $(1, 0, \dots, 0)$ ,  $(0, 1, 0, \dots, 0)$ ,  $\dots$ ,  $(0, \dots, 0, 1)$ , ainsi que du sommet  $(0, \dots, 0)$ . Cela correspondra à la situation où la seule

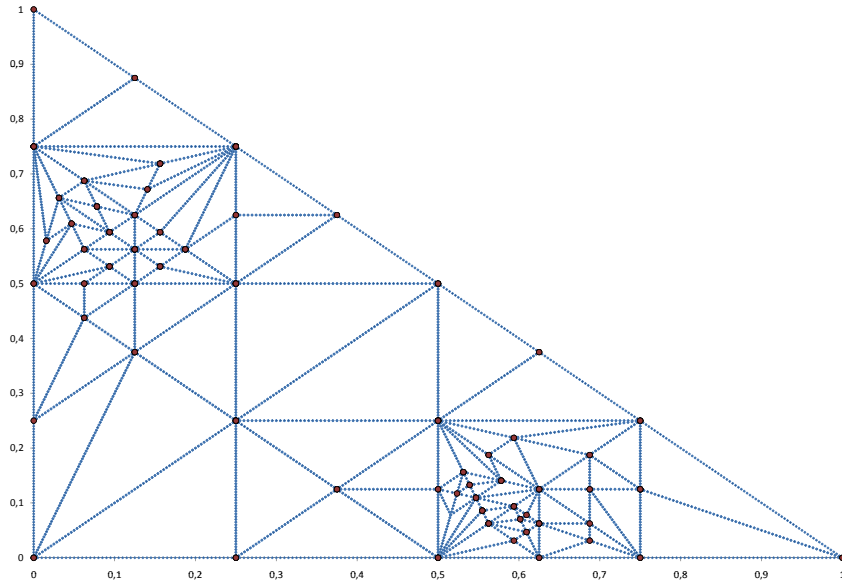


FIGURE 1 – Un exemple de partition d'une zone dans  $\mathbb{R}^2$  en 60 zones. ( $d = 2$ )

contrainte est d'avoir une somme des composantes (par exemple pourcentages d'allocation d'actif) se sommant à un.

Dans les situations où l'optimum est à rechercher sur une zone plus complexe soumise à de nombreuses contraintes, l'algorithme proposé sera applicable si cette zone peut être représentée dès l'initialisation par une union finie de simplexes.

Dans tous les cas, nous supposons que l'ensemble des sommets de la ou des zones initiales ont été explorés, par la réalisation de  $n_0$  tirages de  $F(\theta)$  en chacun des sommets de l'enveloppe de ces zones.

### 2.3 choix de la zone à explorer ou segmenter

Comme le temps de calcul de  $F$  est supposé beaucoup plus important que le temps d'exécution de l'algorithme de choix des zones à explorer, nous choisirons à chaque étape de l'algorithme une unique zone à scinder en deux ou à explorer davantage. L'objectif étant de limiter les risques qu'une zone non explorée contienne un optimum, nous piocherons à chaque étape une zone à explorer, avec une probabilité proportionnelle à un coefficient  $\beta_j$  spécifique à la zone.

Pour chaque zone  $Z_j$ , le coefficient  $\beta_j$  déterminera s'il est plausible que la zone contienne un minimum plus petit que le minimum estimé  $m^*$ . Afin d'approcher ce coefficient sur toute la zone, nous aurons besoin de le déterminer en un unique point. Nous nommerons "potentiel" d'un point un indicateur d'autant plus grand que le point peut contenir un minimum moins élevé que le minimum global observé.

**potentiel d'un point en direction d'un sommet** Considérons un point  $\theta$  d'une zone  $Z_j$ , et un sommet  $\theta_i$  de cette zone. Notons  $p(\theta, \theta_i, m^*)$  la pente requise entre les deux points  $\theta$  et  $\theta_i$  pour passer de l'ordonnée  $\mu^-(\theta_i)$  à l'ordonnée  $m^*$ . On a :

$$p(\theta, \theta_i, m^*) = \frac{\mu^-(\theta_i) - m^*}{d(\theta, \theta_i)},$$

où  $d(\theta_1, \theta_2)$  désigne une mesure de la distance entre deux points  $\theta_1$  et  $\theta_2$ .

Pour la mesure de la vraisemblance de la présence d'un minimum sur le segment  $[\theta, \theta_i]$ , plusieurs choix sont possibles : les choix où cette mesure est une probabilité à valeur dans  $[0, 1]$  par exemple, ou ceux où elle est à valeur dans  $\{0, 1\}$ , qui fixeront ultérieurement une zone "acceptable"  $P_j \subset Z_j$  pour la présence du minimum. On peut ainsi proposer :

- $L_1(\theta, \theta_i) = \mathbb{1}_{p(\theta, \theta_i, m^*) < K}$
- $L_2(\theta, \theta_i) = P \left[ N(0, \sigma_K \sqrt{d(\theta, \theta_i)}) > p(\theta, \theta_i, m^*) \right] = 1 - \Phi \left( \frac{p(\theta, \theta_i, m^*)}{\sigma_K \sqrt{d(\theta_i, \theta)}} \right)$
- $L_3(\theta, \theta_i) = \mathbb{1}_{p(\theta, \theta_i, m^*) < \sigma_K \sqrt{d(\theta, \theta_i)} \Phi^{-1}(\epsilon_K)}$

Il faut tout d'abord remarquer que la pente maximale  $K$  ou la volatilité  $\sigma_K$  sont des constantes que l'on ne cherche pas nécessairement à estimer au vu de la fonction, mais qui servent plutôt à circonscrire les situations que l'on accepte de considérer entre deux points. Elles sont plus semblables en cela à un seuil de confiance qu'à une réelle pente maximale.

Nous excluons le premier choix  $L_1$  de pentes maximales ne dépendant pas de la largeur du segment considéré. Cette première idée de fixer ainsi une pente maximale conduirait à terme à déterminer, au sein de chaque zone  $Z_j$ , la zone  $P_j \subset Z_j$  susceptible de contenir un minimum plus petit que le minimum observé, avec :

$$P_j = \left\{ x \in Z_j, \forall i \in \{1, \dots, d+1\}, \frac{\mu^- - m^*}{d(x, \theta_i)} < K \right\},$$

Les vecteurs  $\theta_i$  désignant les différents sommets de l'enveloppe convexe de  $Z_j$ .  $m^*$  désigne ici un estimateur du minimum global connu de la fonction  $f$ . Les essais numériques impliquant une pente maximale conduisent vite à des points trop uniformément dispersés pour une pente forte, et pour une pente faible à des zones étendues trop rapidement ignorées. Il est en effet plus probable d'observer une pente élevée durant une courte distance que continûment durant une longue distance.

Nous nous pencherons donc sur les deux derniers choix  $L_2$  et  $L_3$ . Nous supposons ici que la fonction  $f$  peut réagir de façon plus ou moins erratique, à la manière d'un mouvement brownien. Partant d'un point origine d'ordonnée  $\mu^-$ , ce mouvement brownien peut, au bout d'une distance  $d$ , se retrouver en dessous d'une valeur  $m^*$  avec une probabilité  $1 - \Phi \left( \frac{p(\theta, \theta_i, m^*)}{\sigma_K \sqrt{d(\theta_i, \theta)}} \right)$ , où  $\sigma_K$  désigne la volatilité du brownien, et  $\Phi$  la fonction de répartition d'une loi normale centrée réduite. Par l'utilisation de  $L_3$ , il est possible de délimiter une zone admissible pour le minimum telle que cette probabilité soit inférieure à un seuil donné. Outre l'ajout d'un nouveau paramètre, nous verrons que l'estimation de la surface d'une telle zone pourra s'avérer relativement coûteuse en terme de temps de calcul. Cette solution reste néanmoins envisageable, notamment lorsque le calcul de  $F$  est très long.

Nous retiendrons donc la deuxième solution  $L_2$ , de sorte que finalement :

$$L(\theta, \theta_i) = L_2(\theta, \theta_i) = 1 - \Phi \left( \frac{p(\theta, \theta_i, m^*)}{\sigma_K \sqrt{d(\theta_i, \theta)}} \right).$$

Par ailleurs, dans le cas particulier où les intervalles de confiance ont été bâtis à l'aide d'une loi normale, il est possible de supprimer le  $\mu^-(\theta_i)$  implicite dans calcul de  $L$  en additionnant simplement les variances liées aux erreurs d'échantillonnage et aux erreurs de variation de la fonction entre les points de la grille, lors de la détermination de la distribution supposée de  $F$  au point  $\theta$ , les deux erreurs étant supposées indépendantes.



A titre de remarque, dans le cas d'une recherche de racine plutôt que d'optimum, il est possible d'adapter ce potentiel, en proposant par exemple la probabilité que le niveau de la fonction, compte tenu de la variabilité  $\sigma_K$ , appartienne à un intervalle  $[-h, +h]$  :

$$\begin{aligned} L^{racine}(\theta, \theta_i) &= |\Phi(p_1) - \Phi(p_2)|, \\ p_1 &= \frac{p(\theta, \theta_i, +h)}{\sigma_K \sqrt{d(\theta_i, \theta)}}, \\ p_2 &= \frac{p(\theta, \theta_i, -h)}{\sigma_K \sqrt{d(\theta_i, \theta)}}. \end{aligned}$$

**potentiel d'un point** Considérons un point  $\theta$  d'une zone  $Z_j$ . On appelle potentiel du point  $\theta$  dans la zone  $Z_j$  le produit des potentiels directionnels de  $\theta$  vers les sommets  $\theta_1, \dots, \theta_{d+1}$  de  $Z_j$  :

$$\beta_j(\theta) = \prod_{i=1}^{d+1} L(\theta, \theta_i).$$

Ce produit semble naturel si aucune direction n'est privilégiée. Pour les choix de  $L(\theta, \theta_i) \in \{0, 1\}$ , ce potentiel indique si la présence du minimum en ce point est ou non admissible. Lorsque  $L(\theta, \theta_i) \in [0, 1]$  ce potentiel sera d'autant plus élevé que le minimum sera susceptible de se trouver au point  $\theta$ .

**potentiel d'une zone** En pratique, il faudrait estimer au sein de chaque zone  $Z_j$  une valeur

$$\bar{\beta}_j = \iint_{\theta \in Z_j} \beta_j(\theta) d\theta.$$

Nous noterons  $V(Z_j)$  le volume de la zone  $Z_j$  (hypervolume dans le cas  $d > 3$ ). Afin d'éviter le calcul délicat de la valeur  $\bar{\beta}_j$ , il est envisageable de procéder de plusieurs façons.

- Il est possible de considérer le point  $B_j$  isobarycentre des sommets de la zone  $Z_j$  comme représentatif du potentiel de la zone, auquel cas on peut poser

$$\beta_{j,1} = V(Z_j) \cdot \beta_j(B_j).$$

Le tirage arbitraire d'un point dans la zone peut se faire en calculant un barycentre des sommets de la zone, de coefficients indépendants et identiquement distribués, de loi uniforme sur  $[0, 1]$ .

- Il est possible de piocher à chaque étape  $n_S$  points  $P_1, \dots, P_{n_S}$ , aléatoirement dans  $Z_j$ , puis de calculer le coefficient :

$$\beta_{j,2} = V(Z_j) \cdot \frac{1}{n_S} \sum_{i=1}^{n_S} \beta_j(P_i).$$

La première solution offre l'avantage de ne pas être aléatoire, et les calculs de  $\beta_j$  menés sur de précédentes zones pourront être très facilement conservés. Cette solution a le mérite de la simplicité, et nous a paru présenter un comportement correct dans nos applications numériques, toutefois limitées au cas  $d = 2$ . Ce sera la solution que nous retiendrons :

$$\begin{aligned} \beta_j &= V(Z_j) \cdot \beta_j(B_j). \\ B_j &\text{ isobarycentre des sommets de la zone } Z_j. \end{aligned}$$

S'agissant du calcul du volume  $V(Z_j)$ , dans le cas où  $d = 2$ , les  $Z_j$  sont des triangles et le volume  $V(Z_j)$  est donné par exemple par la formule de Héron :

$$V(Z_j) = \sqrt{s(s-a)(s-b)(s-c)},$$

avec  $a = d(\theta_1, \theta_2)$ ,  $b = d(\theta_1, \theta_3)$ ,  $c = d(\theta_2, \theta_3)$ ,  $s = \frac{1}{2}(a + b + c)$ . Dans le cas général, le déterminant de Cayley-Menger donne le volume exact de la zone. Une distance moyenne élevée à la puissance  $d$  peut également fournir un substitut si le calcul du déterminant s'avère trop long. Plus simplement, dans le cas d'une séparation d'une zone en deux volumes égaux, ce qui sera ici le cas, la simple mémorisation du volume de la zone à subdiviser permet de déduire immédiatement le demi-volume de chaque zone fille.

## 2.4 choix du sommet de séparation

Une fois une zone  $Z^+$  choisie parmi l'ensemble des zones  $\{Z_j\}_{j=1,2,\dots}$ , il va s'agir de scinder la zone en ajoutant un point à l'intérieur de celle-ci. De par nos choix précédents, ce point se situera sur l'un des segments de l'enveloppe convexe de la zone.

Pour la zone  $Z^+$ , nous noterons  $C = \{(\theta_{j_1}, \theta_{j_2})\}_{j_1, j_2}$  l'ensemble des couples de sommets distincts de  $Z^+$ . Le sommet de séparation retenu peut être choisi de différentes façons. Ainsi, parmi l'ensemble des couples  $C = \{(\theta_{j_1}, \theta_{j_2})\}_{j_1, j_2}$ , il est possible de sélectionner pour le couple  $(\theta_{j_1}, \theta_{j_2})$

- celui qui sera choisi aléatoirement de façon uniforme.
- celui dont l'isobarycentre  $\theta_c = \frac{1}{2}\theta_{j_1} + \frac{1}{2}\theta_{j_2}$  minimise la quantité estimée  $\mu^-(\theta_c)$ .
- celui qui maximise la largeur du segment  $d(\theta_{j_1}, \theta_{j_2})$ .

Le choix peut dépendre en pratique de la complexité comparée de l'algorithme par rapport au tirage en lui-même de chaque valeur  $F(\theta)$ . Si ce tirage est particulièrement long, il peut être intéressant de dépenser un peu de temps pour sélectionner aussi judicieusement que possible le point de segmentation. Dans le cas contraire, le simple choix aléatoire d'un segment peut être suffisant.

La solution que nous retiendrons consiste ici à sélectionner, parmi l'ensemble des couples possibles, celui qui maximise la largeur du segment. Si  $\theta_1, \dots, \theta_{d+1}$  désignent les sommets de la zone  $Z^+$ , nous choisirons comme sommet de séparation de la zone le sommet  $\theta^+$ , isobarycentre du segment  $(\theta_{j_1}, \theta_{j_2})$  tel que :

$$\begin{aligned} (\theta_{j_1}, \theta_{j_2}) &= \arg \max_{(\theta_i, \theta_j) \in C} d(\theta_i, \theta_j), \\ C &= \{(\theta_i, \theta_j)\}_{i \in \{1, \dots, d\}, j \in \{i+1, \dots, d+1\}}. \end{aligned}$$

Cette solution a le mérite de limiter l'apparition de simplexes très déséquilibrés dans leurs longueurs de segment : en effet, la connaissance de l'appartenance de l'optimum à une zone est moins utile si cette zone est très allongée dans certaines directions. Le choix d'un sommet conduisant également à la partition des zones adjacentes, l'apparition de simplexes très déséquilibrés, si elle est ainsi limitée, n'est toutefois pas absolument exclue. Une idée de la forme des zones ainsi obtenues en dimension 2 est illustrée par la figure 1, qui a été construite par scission du segment de plus grande longueur de la zone choisie.

Lorsque des simulations sont opérées sur le sommet du segment choisi de séparation, la scission des zones se fait ainsi :

- l'ensemble des zones  $\mathfrak{Z}^+$  contenant ce segment est déterminé (celles qui contiennent les deux sommets à la fois).
- chacune des zones  $Z$  de l'ensemble  $\mathfrak{Z}^+$  est scindée en deux zones  $Z_1$  et  $Z_2$  comme indiqué dans le paragraphe 2.1.

## 2.5 schéma de l'algorithme de scission systématique

### présentation du schéma à l'étape $k$

- Pour chaque zone  $Z_j$  de  $\mathfrak{Z}_k$ ,  $j \in \{1, \dots, \nu_k\}$ 
  - le calcul du coefficient  $\beta_j$  de la zone est mené (ou simplement mis à jour s'il est modifié).
  - le coefficient total  $\beta$  est mis à jour en conséquence.
- une zone  $Z^+$  est piochée aléatoirement parmi l'ensemble des zones  $\{Z_j\}_{j \in \{1, \dots, \nu_k\}}$ , la probabilité que la zone  $Z_j$  soit piochée étant  $\beta_j/\beta$ .
- un segment de la zone  $Z^+$  est choisi suivant la procédure indiquée au paragraphe 2.4, le centre  $\theta^+$  de ce segment désignant le sommet à partir duquel la zone sera scindée en deux parties.
- toutes les zones contenant ce segment sont séparées en deux comme indiquée au paragraphe 2.3.
- $n_0$  simulations sont opérées au sommet  $\theta^+$ . Les valeurs de  $\mu$  et  $\mu^-$  sont estimées en ce point  $\theta^+$ . La valeur de  $m^*$  est modifiée en conséquence.

## 3 Une grille à pas variable avec retraitage possible

Jusqu'à présent, chaque zone était systématiquement subdivisée en deux. Or, il peut être plus avantageux, plutôt que de subdiviser une zone, d'ajouter des simulations en des points déjà explorés. Ce sera l'objet de cette section.

### 3.1 Critères de segmentation

Supposons que l'on estime une espérance et un intervalle de confiance de la fonction  $F(\theta)$ , aux points  $\theta_1$  et  $\theta_2$ . Supposons que  $\theta_c$  soit le barycentre équipondéré (isobarycentre) entre ces deux points (il serait également envisageable de ne pas scinder le segment en son milieu, mais en un autre barycentre).

La question que l'on se pose est la suivante : au vu des intervalles de confiance de  $F(\theta)$ , connus aux points  $\theta_1$  et  $\theta_2$ , et connaissant par exemple une pente maximale admissible pour  $f(\theta)$  sur ce segment, nous cherchons un critère permettant de déterminer la décision à prendre entre les suivantes :

- augmenter le nombre de tirages de  $F$  aux points  $\theta_1$  et  $\theta_2$ , de façon à rétrécir les intervalles de confiance de  $F$  en ces deux points.
- scinder le segment en deux parties, en opérant des tirages de  $F$  au nouveau point  $\theta_c$ .

Plusieurs critères peuvent être retenus pour opérer ce choix. Le choix du critère retenu au final dépend essentiellement de l'objectif poursuivi : s'assurer que le minimum ne peut être présent dans une zone si la fonction objectif ne varie pas trop brusquement, minimiser le maximum des potentiels d'une zone, etc. Ce choix pouvant dépendre du problème considéré, nous détaillerons ci-dessous quelques-uns des critères qui pourraient être envisagés.

**critère déterministe linéaire de scission d'une zone** Dans un premier temps nous supposons que nous ignorons combien de points vont être ajoutés. Dans le cas contraire, des critères plus fins peuvent être envisagés : si  $F$  est approchée par une loi normale, par exemple, alors la largeur de l'intervalle de confiance en un point évolue de façon connue en fonction de  $\sqrt{n}$ , où  $n$  désigne le nombre de simulations en ce point.

S'agissant de l'*erreur d'échantillonnage*, cette erreur liée à la mauvaise connaissance de la fonction lors de l'estimation du minimum global de la fonction dépend essentiellement de la probabilité de présence du minimum au point  $\theta_1$  ou  $\theta_2$  du fait de trop larges intervalles de confiance en ces points. En l'absence de doute lié à l'évolution de la fonction entre ces deux points, cette erreur peut-être estimée au point  $\theta_c$  par la quantité suivante, dans le cas d'un barycentre équipondéré :

$$m_1(\theta_c) = \frac{1}{2}\mu^-(\theta_1) + \frac{1}{2}\mu^-(\theta_2)$$

où  $\mu^-(\theta)$  désigne la borne inférieure de l'intervalle de confiance de  $F$  au point  $\theta$ . Cette erreur représente la position minimale admissible pour le minimum au point  $\theta_c$ , compte tenu de la seule mauvaise connaissance de la fonction  $f$  aux extrémités du segment.

S'agissant de l'*erreur de grille*, cette erreur liée à la largeur du segment sur l'estimation du minimum global de la fonction dépend essentiellement de la possibilité que la fonction évolue brusquement sur ce segment, sans laquelle il serait inutile de scinder le segment. En l'absence de doute lié à la mauvaise connaissance de la fonction aux extrémités du segment, cette erreur peut-être estimée au point  $\theta_c$  par la quantité suivante, dans le cas d'un barycentre équipondéré :

$$m_2(\theta_c) = \max(\mu(\theta_1), \mu(\theta_2)) - g_K(d(\theta_1, \theta_2)) , .$$

Cette erreur représente la position minimale admissible pour le minimum au point  $\theta_c$ , compte tenu de la seule mauvaise exploration de la fonction.

On pourrait imaginer prendre, par exemple,  $g_K(d) = \sigma_K \cdot \sqrt{\frac{d}{2}}\Phi^{-1}(1 - \epsilon_K)$  : si partant du niveau le plus élevé aux extrémités du segment  $\max(\mu(\theta_1), \mu(\theta_2))$ , un brownien de volatilité  $\sigma_K$  parcourrait la distance  $\frac{1}{2}d(\theta_1, \theta_2)$ , alors la borne inférieure l'intervalle de confiance pour la position de ce brownien au point  $\theta_c$  serait  $m_2(\theta_c)$ , pour un seuil de confiance  $\epsilon_K$  fixé.

Ce choix d'un mouvement brownien se conformerait bien au choix effectué lors du calcul du potentiel d'une zone. Nous allons toutefois l'écarter, car nous avons ici choisi de séparer les erreurs liées à l'échantillonnage et à la largeur de la grille, ce qui conduit à une erreur de grille que l'on conçoit moins facilement brownienne, et car il ne s'agit plus ici de comparer des coefficients entre eux, mais de comparer deux erreurs de natures différentes. En pratique, comme nous le verrons au paragraphe 4, le choix de la subdivision est notamment important lorsque les zones de recherches sont suffisamment petites. Il s'agit alors de comparer l'incertitude liée à la largeur de la cellule de la grille à l'incertitude liée à la profondeur potentielle de la fonction sur cette grille. Le choix d'une évolution brownienne, sur de courts intervalles, en l'absence d'erreur d'échantillonnage nous a paru excessif : cela revient à systématiquement privilégier la scission sur des intervalles très courts, la dérivée de la racine carrée tendant alors vers  $+\infty$ , cela nous a semblé peu réaliste en l'absence ici d'erreur d'échantillonnage, et les essais conduisent, pour des niveaux de variabilités comparables à ceux d'autres choix que nous détaillerons, à des scissions quasi-systématiques.

Aussi, dans nos applications, nous avons choisi de prendre  $g_K(d) = K \cdot \frac{d}{2}$ . de sorte que le choix finalement retenu est :

$$\begin{aligned} m_2(\theta_c) &= \max(\mu(\theta_1), \mu(\theta_2)) - \frac{K}{2}d(\theta_1, \theta_2) , \\ \mu^-(\theta_c) &= m_2(\theta_c) . \end{aligned}$$

Il est bien sûr possible d'apporter de très nombreux raffinements à ces mesures d'incertitudes liées soit à l'erreur d'échantillonnage, soit à l'erreur de grille. Rappelons que le minimum actuellement estimé de  $f(\theta)$  est noté  $m^*$ . Si  $m_2(\theta_c) < m_1(\theta_c)$ , la scission peut sembler plus pertinente

que l'ajout éventuel de simulations aux extrémités du segment, dans la mesure où l'erreur de grille conduit à la présence d'un minimum potentiel inférieur à l'erreur d'échantillonnage.

Par souci de simplicité, l'algorithme de scission systématique nous ayant paru donner de bons résultats, comme cela sera évoqué au paragraphe 4, nous scinderons la zone en deux parties si :

$$m_2(\theta_c) < m_1(\theta_c)$$

Dans le cas contraire, nous pouvons par exemple répartir les nouvelles simulations entre les sommets à l'extrémité du segment (par exemple par moitiés, mais d'autres possibilités seront évoquées par la suite).

**critère déterministe de scission selon le meilleur potentiel** Nous envisageons ici le cas où la répartition se fait en fonction de l'amélioration du potentiel de la zone, scindée ou non. On suppose que  $n_0$  points seront ajoutés sur un sommet  $S^+$ .

- Si l'on envisage d'ajouter  $n_0$  points sur un sommet existant  $\theta$  de la zone choisie, il est aisé d'estimer l'amélioration de l'intervalle de confiance autour de  $\theta$ , et donc une estimation  $\mu_{ajout}^-(\theta)$  de la nouvelle borne inférieure de l'intervalle de confiance en cas d'ajout.
- Si l'on envisage d'ajouter  $n_0$  points sur le barycentre  $\theta_c$  du segment choisi, on peut alors estimer par interpolation la valeur de  $\mu(\theta_c)$  ainsi que l'écart-type autour de cette valeur, et donc l'intervalle de confiance obtenu après  $n_0$  simulations, et donc une valeur estimée  $\mu_{ajout}^-(\theta_c)$  de la borne inférieure de l'intervalle de confiance en cas d'ajout.

Quel que soit le sommet choisi pour l'ajout parmi les  $d + 1$  sommets existants, on peut calculer la nouvelle valeur estimée globale de  $\beta(Z^+)$  sur la zone choisie. Si un sommet existant doit être privilégié pour l'ajout, il semble logique de convenir d'un ajout sur celui de ces  $d + 1$  points qui conduit à minimiser cette valeur estimée, ce minimum étant noté  $\beta_{\min}$ . Si au contraire une scission est envisagée, deux zones  $Z_1$  et  $Z_2$  seront formées. La scission pourra être privilégiée, par exemple, si les potentiels estimés sur ces deux zones sont tels que :

$$\max(\beta(Z_1), \beta(Z_2)) < \beta_{\min},$$

L'esprit étant alors qu'une scission ne doit pas conduire à augmenter le maximum des  $\beta(Z_i)$  sur l'ensemble des zones  $\{Z_i\}$  à l'étape considérée : l'ajout se porte ainsi sur le sommet dont on estime qu'il minimise le maximum des potentiels sur l'ensemble des zones.

L'avantage de cette répartition est qu'elle ne nécessite pas la fixation d'un nouveau seuil, qu'elle est pleinement cohérente avec la définition choisie des coefficients  $\beta$ , et qu'elle conduit, par construction, à une diminution du potentiel maximum estimé sur l'ensemble des zones, ce qui est de nature à faciliter les futures démonstrations de la convergence de l'algorithme.

**répartition stochastique des points ajoutés** La question est finalement de répartir les simulations à opérer entre  $\theta_1$ ,  $\theta_2$  et  $\theta_c$ . Cette répartition peut se faire de façon déterministe, par l'application de critères comme ceux présentés ci-dessus, ou de façon stochastique, en piochant par exemple un sommet avec une probabilité proportionnelle à un indicateur donné.

On pourrait ainsi opérer une répartition en fonction d'une estimation de  $\mu^-$  ou encore d'un critère  $\alpha(\theta) = \epsilon + (m^* - \mu^-(\theta))_+$ , la valeur  $\epsilon > 0$  visant à éviter les cas où chaque  $\alpha(\theta)$  serait nul. On dispose par exemple pour chaque sommet d'un ensemble de valeurs  $\alpha(\theta) = (m^* - \mu^-(\theta))_+$ . Dans un premier temps, et afin d'éviter un tri de ces valeurs pour les différents sommets, on se fixera une valeur  $U$  uniforme sur  $]0, 1]$ . Nous pourrons ajouter  $n_0$  simulations au sommet dont l'indice  $i$  sera le premier pour lequel

$$S(i) \geq U\alpha,$$

avec  $S(i) = \sum_{r=1}^i \alpha(\theta_r)$ , et  $\alpha = \sum_i \alpha(\theta_i)$  le cumul de tous les coefficients sur l'ensemble des sommets. Cela revient à piocher aléatoirement un sommet parmi ceux possibles, avec une probabilité  $\gamma(i) = \alpha(i)/\alpha$ .

On suppose que le calcul des  $\alpha(i)$  et de cette répartition s'opère en un temps négligeable devant celui des simulations. Dans le cas contraire, il peut être utile d'ajouter des simulations en plusieurs points à la fois.

On remarque par ailleurs que la largeur de l'intervalle de confiance en chaque sommet, ancien ou nouveau, peut être estimée après ajout de  $n_0$  simulations. Si le temps de calcul de chaque simulation est vraiment très important, il peut être utile d'optimiser le vecteur des points ajoutés en fonction des améliorations estimées des  $\alpha(i)$  en fonction de ce vecteur d'ajout.

Par souci de simplicité, ces répartitions stochastiques n'ont pas été développées dans ce papier, la seule composante aléatoire résidant alors dans le choix des zones à éventuellement scinder.

## 3.2 schéma général de l'algorithme

### présentation du schéma à l'étape $k$

- Pour chaque zone  $Z_j$  de  $\mathfrak{Z}_k$ ,  $j \in \{1, \dots, \nu_k\}$ 
  - le calcul du coefficient  $\beta_j$  de la zone est mené (ou simplement mis à jour).
  - le coefficient total  $\beta$  est mis à jour en conséquence.
- une zone  $Z^+$  est piochée aléatoirement parmi l'ensemble des zones  $\{Z_j\}_{j \in \{1, \dots, \nu_k\}}$ , la probabilité que la zone  $Z_j$  soit piochée étant  $\beta_j/\beta$ , comme indiqué au paragraphe 2.3.
- un segment de la zone  $Z^+$  est choisi suivant la procédure indiquée au paragraphe 2.4, le centre  $\theta^+$  de ce segment désignant le sommet à partir duquel la zone sera peut-être scindée en deux parties.
- Le choix de scinder la zone est opéré conformément au critère exposé au paragraphe 3.1
  - en cas de segmentation :
    - $n_0$  simulations sont opérées au sommet  $\theta^+$ . Les valeurs de  $\mu$  et  $\mu^-$  sont estimées en ce point  $\theta^+$ . La valeur de  $m^*$  est modifiée en conséquence.
    - toutes les zones contenant ce segment sont séparées en deux comme indiquée au paragraphe 2.1.
  - sinon :
    - une extrémité  $\theta_0$  du segment est choisie.
    - $n_0$  simulations supplémentaires sont opérées au sommet  $\theta_0$ . Les valeurs estimées de  $\mu$  et  $\mu^-$  en ce point  $\theta_0$  sont corrigées en conséquence, ainsi que la valeur de  $m^*$ .

L'algorithme présenté ici vise simplement à montrer l'intérêt et la faisabilité technique d'une grille à pas variable, susceptible de répondre au problème du choix entre exploration et connaissance d'une fonction aléatoire. De nombreuses améliorations de cet algorithme sont actuellement à l'étude et devraient permettre d'améliorer son efficacité, notamment dans les cas où la fonction  $F$  est très lourde à calculer (dépendant par exemple de simulations stochastiques de trajectoires de nombreux actifs par exemple).

## 4 Applications numériques

### 4.1 recherche d'optimum

**Comportement attendu de l'algorithme de scission systématique** Nous présenterons ici une application de l'algorithme sur une fonction connue, afin notamment de vérifier le bon comportement de l'algorithme. Par souci de lisibilité, cette application sera présentée en dimension  $d = 2$  (mais l'algorithme présenté dans ce papier est proposé pour toute dimension  $d \in \mathbb{N}^*$ )

Pour l'algorithme de scission systématique, l'algorithme vise essentiellement à répartir des points d'exploration, en délaissant certaines zones non susceptibles de contenir le minimum. Selon les paramètres choisis et la fonction optimisée, le comportement attendu est le suivant :

- Lorsque le bruit est important devant les variations de la fonction, ou lorsque l'incertitude sur la variabilité de la fonction  $\sigma_K$  est élevée, l'algorithme doit explorer assez uniformément la fonction, de façon similaire aux traditionnelles grilles de recherche à pas fixe évoquées précédemment.
- Lorsque le bruit n'est pas trop élevé et lorsque la variabilité de la fonction  $\sigma_k$  est faible, l'algorithme doit se concentrer sur les zones contenant les optimums, le comportement de la fonction étant supposé sans surprise entre les points déjà explorés.

Le choix de la variabilité de la fonction  $\sigma_K$  dépend donc du but poursuivi : trop élevé, l'exploration de la fonction sera très poussée, et le temps de calcul pourra être élevé si l'on vise une bonne connaissance d'un optimum de la fonction. Pour un  $\sigma_K$  trop faible, l'algorithme se focalisera très vite sur un optimum local, donc permettra un gain notable en terme de temps de calcul, à précision égale, mais le risque de ne pas explorer une zone susceptible de contenir un autre optimum, éventuellement meilleur, sera plus élevé.

**Comportement attendu de l'algorithme avec retraitage** Pour l'algorithme avec retraitements possibles, le comportement attendu est le suivant :

- Lorsque  $K$  est faible, le comportement de  $f$  est jugé sans surprise entre les points de la zone sélectionnée : l'algorithme doit rajouter des simulations exclusivement sur les sommets connus de cette zone.
- Lorsque le bruit est faible, ou à plus forte raison lorsque  $F$  est déterministe, la seule incertitude réside dans le comportement de la fonction entre les points, et l'algorithme doit systématiquement créer de nouveaux points de tirage.

Nous allons voir que les résultats obtenus se conforment bien à ces comportements attendus.

**résultats obtenus** Nous considérerons ici la fonction suivante dans le cas  $d = 2$  (correspondant par exemple à 3 poids d'allocation d'actifs se sommant à 1).

$$\begin{aligned} F(\theta) &= (\min(x, y) - 0.1)^2 + (\max(x, y) - 0.6)^2 + \sigma_B(U - 0.5), \\ \theta &= (x, y) \end{aligned}$$

avec  $U$  une variable aléatoire uniforme sur  $[0, 1]$ . L'espérance  $f$  de la fonction  $F$  admet deux minimums, l'un en  $\theta_1^* = (0.1, 0.6)$ , l'autre en  $\theta_2^* = (0.6, 0.1)$ , la valeur de  $f$  étant alors 0. A titre indicatif, afin d'imaginer les variations possibles de cette fonction, la fonction  $f$  atteint son maximum sur  $Z_0$  en  $\theta = (0, 0)$  et l'on a alors  $f(\theta) = 0.37$  (mais l'on cherche ici le minimum, non le maximum). Un bruit pouvant conduire à des variations d'amplitude de 0.1 entre deux points proches est donc assez élevé au regard du domaine de variation  $[0, 0.37]$  de la fonction sur la zone initiale de recherche.

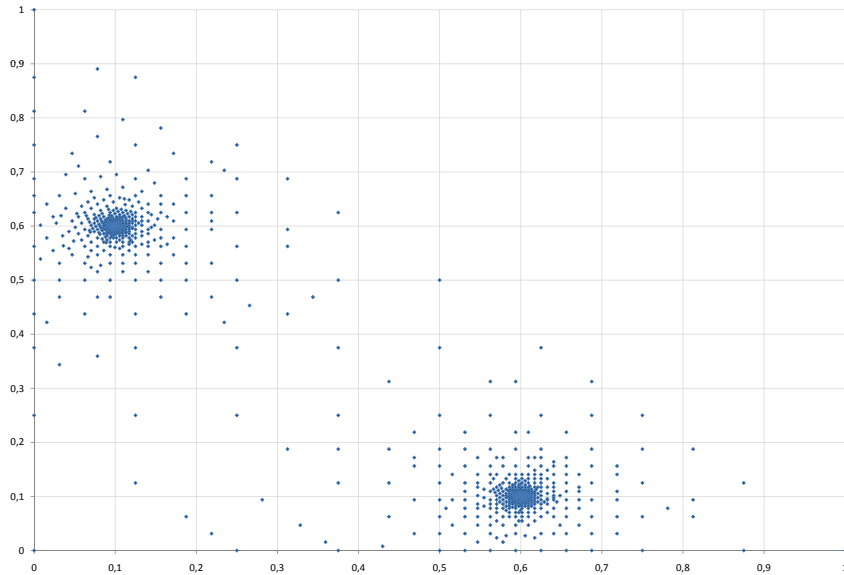


FIGURE 2 – Ensemble des points de simulations de  $F$  pour une variabilité  $\sigma_K = 2$  et un niveau de bruit  $\sigma_B = 0$  (scission automatique)

Dans le cadre de l’algorithme à scission automatique, nous utiliserons ici toujours les paramètres suivants :

- le nombre de tirages en chaque point sera  $n_0 = 10$ .
- le nombre de zones créées sera égal à 1000.
- le niveau du bruit  $\sigma_B$  sera précisé pour chaque illustration.
- le niveau de variabilité  $\sigma_K$  sera précisé pour chaque illustration.

A titre de remarque, la figure 1 présentée précédemment a été obtenue avec la fonction  $F$  évoquée ci-dessus, à partir des paramètres  $\sigma_K = 1$  et  $\sigma_B = 0.1$ .

Le comportement de la recherche de l’optimum global est bien celui attendu. Pour un niveau de variabilité supposée de la fonction  $\sigma_K = 2$ , illustré dans les figures 2, 3 et 4), lorsque le niveau de bruit passe respectivement par  $\sigma_B = 0$ ,  $\sigma_B = 0.1$  et  $\sigma_B = 0.3$ , on observe bien l’exploration préférentielle des zones à proximité des optima locaux. Bien entendu, cette exploration est plus étendue lorsque le niveau de bruit augmente, puisque du fait de ce bruit, l’assurance de l’absence d’autres optima locaux nécessite l’exploration d’une zone plus vaste. Il faut noter que dans la pratique, ce paramètre est une donnée exogène sur laquelle il n’est pas possible d’agir.

Pour un niveau de bruit de la fonction fixé à  $\sigma_B = 0.1$ , illustré dans les figures 3, 5 et 6), lorsque la variabilité supposée de la fonction passe successivement par  $\sigma_K = 2$ ,  $\sigma_K = 20$  et  $\sigma_K = 100$ , on observe là encore un résultat parfaitement logique : si le comportement attendu de la fonction entre les points de tirage est supposé peu variable, l’algorithme se concentre sur les zones proches des optima locaux. Dans le cas inverse, une variabilité supposée extrême, dans le cas  $\sigma_K = 100$ , conduit à une répartition quasi uniforme des points d’exploration : avec une telle variabilité, l’optimum global de la fonction peut en effet se trouver dans n’importe quelle zone. Il faut toutefois remarquer que la fixation d’un seuil de variabilité  $\sigma_K$  trop faible conduirait par définition à supposer que le comportement de la fonction est globalement connu entre les points explorés, ce qui conduit à négliger des zones pourtant susceptibles de contenir un optimum. L’arbitrage entre la rapidité de convergence vers les optimums globaux et le risque



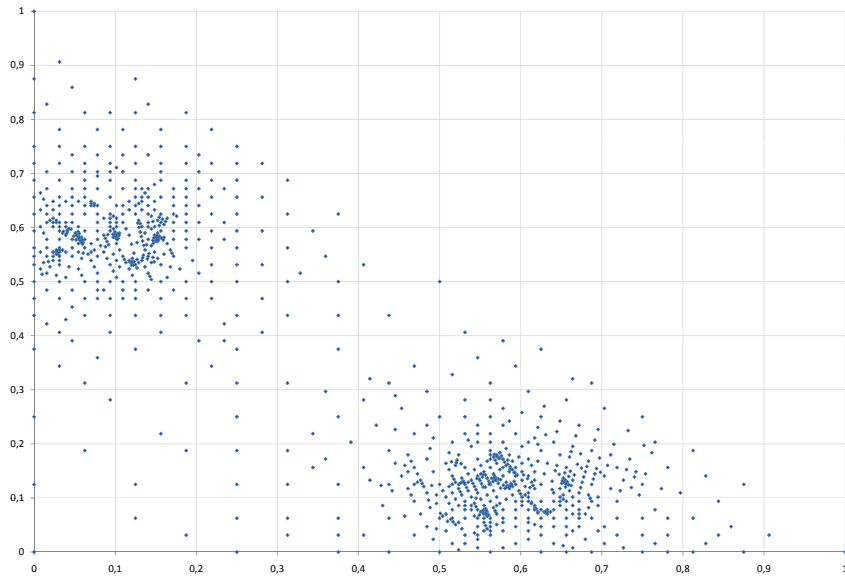


FIGURE 3 – Ensemble des points de simulations de  $F$  pour une variabilité  $\sigma_K = 2$  et un niveau de bruit  $\sigma_B = 0.1$  (scission automatique)

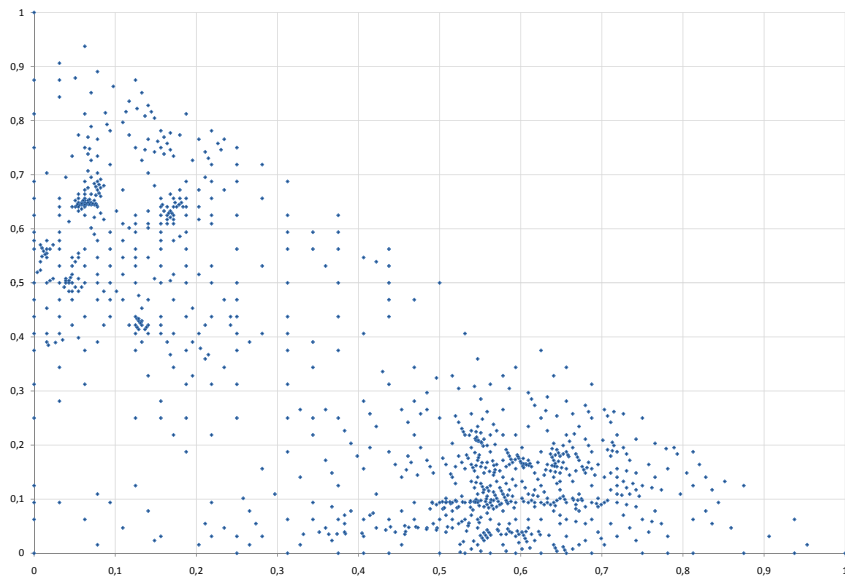


FIGURE 4 – Ensemble des points de simulations de  $F$  pour une variabilité  $\sigma_K = 2$  et un niveau de bruit  $\sigma_B = 0.3$  (scission automatique)

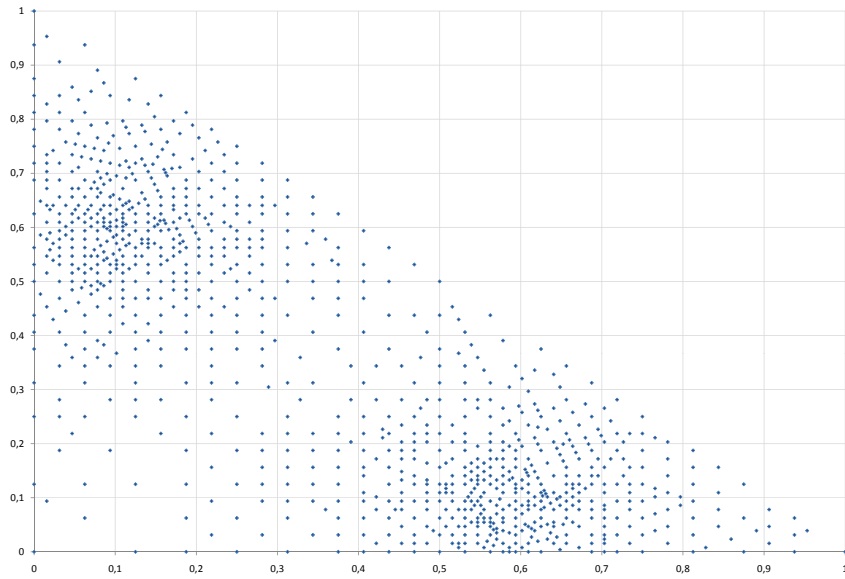


FIGURE 5 – Ensemble des points de simulations de  $F$  pour une variabilité  $\sigma_K = 20$  et un niveau de bruit  $\sigma_B = 0.1$  (scission automatique)

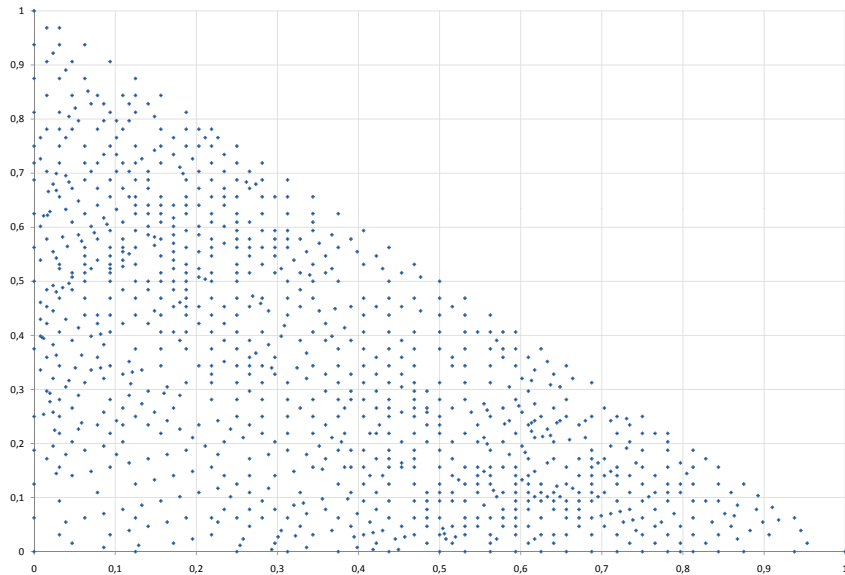


FIGURE 6 – Ensemble des points de simulations de  $F$  pour une variabilité  $\sigma_K = 100$  et un niveau de bruit  $\sigma_B = 0.1$  (scission automatique). La recherche de l'optimum d'une fonction constante bruitée conduit à un nuage de point d'allure proche de celui-ci.

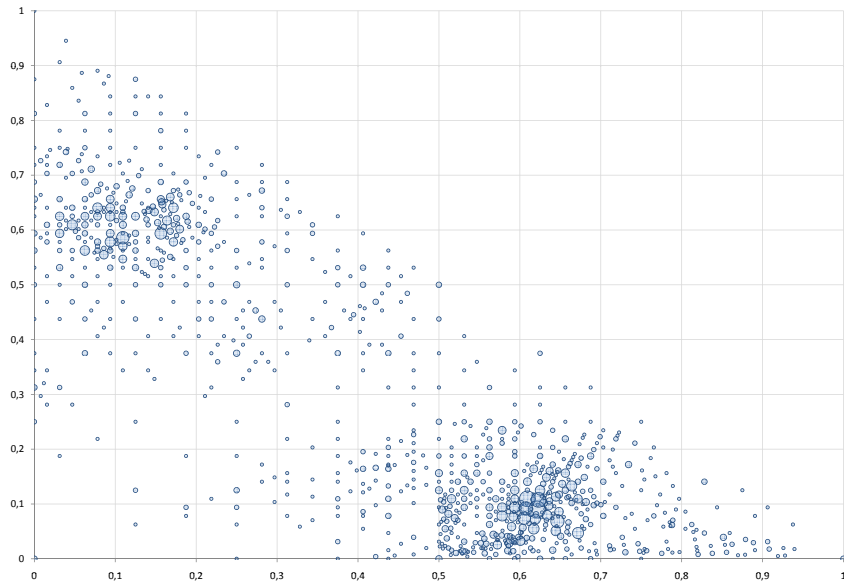


FIGURE 7 – Ensemble des points de simulations de  $F$  pour un niveau de bruit  $\sigma_B = 0.1$ , avec scission sur critère déterministe linéaire (pour une variabilité  $\sigma_K = 1$ , une pente de choix de scission  $K = 1$  et un nombre minimal de tirage  $n_0 = 10$ ). La surface des bulles est proportionnelle au nombre de tirage de  $F$  en chaque point.

d’avoir négligé une zone dépendra donc largement de ce paramètre  $\sigma_K$ .

La fixation du paramètre  $\sigma_K$  reste donc un problème ouvert, elle pourra se baser en particulier sur les explorations précédentes de la fonction, sur d’autres connaissances de celles-ci, ainsi que sur les contraintes d’implémentation et de niveau de risque accepté. Il faut noter à ce sujet que rien dans le modèle n’oblige à fixer cette valeur constante sur l’ensemble de la zone de recherche, et qu’il est également possible de tenir compte d’une variabilité particulière de la fonction sur certaines zones.

Les deux figures 7 et 8 illustrent les retirages qui peuvent s’opérer sur des sommets déjà explorés, afin d’améliorer la connaissance de la fonction  $f$  en ces points. Lorsque le niveau de bruit est  $\sigma_B = 0.1$ , la figure 7 montre que les retirages ont essentiellement lieu autour des points minimaux, lorsqu’il n’est plus seulement nécessaire d’avoir une idée de la zone où se situe l’optimum, mais qu’il faut également connaître le niveau précis de cet optimum. Lorsque le bruit augmente, cela tend naturellement à disperser un peu les points d’exploration, et favorise le retraitage, dans la mesure où l’incertitude en certains points, du fait du bruit, devient plus forte que l’incertitude liée à la variation de la fonction entre les points de la grille.

La figure 9 illustre un choix de scission sur critère déterministe de meilleur potentiel : elle donne une idée de l’impact du choix du critère de scission, qui conduit à un comportement toutefois proche de retirages sur des sommets existants lorsque la présence du minimum est suffisamment vraisemblable et que l’incertitude pesant sur les sommets de la zone doit être diminuée.

Pour les critères déterministes abordés ici, il faut enfin noter que pour un niveau de bruit nul, aucun retraitage n’intervient (les graphiques illustrant ce résultat ont toutefois été omis). Cela est logique dans la mesure où ces retirages n’apporteraient rien à la connaissance de la fonction  $f$ .

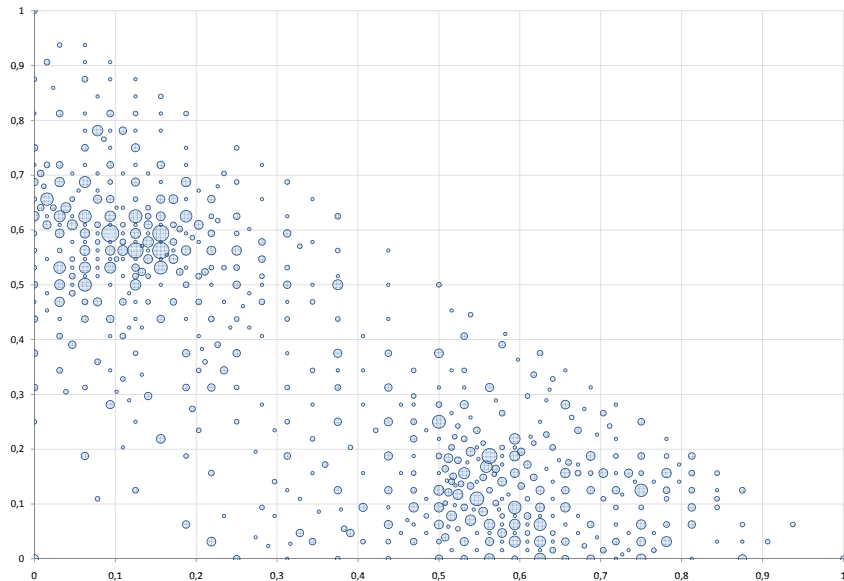


FIGURE 8 – Ensemble des points de simulations de  $F$  pour un niveau de bruit  $\sigma_B = 0.2$ , avec scission sur critère déterministe linéaire (pour une variabilité  $\sigma_K = 1$ , une pente de choix de scission  $K = 1$  et un nombre minimal de tirage  $n_0 = 10$ ). La surface des bulles est proportionnelle au nombre de tirage de  $F$  en chaque point.

## 4.2 recherche de racines

Nous avons enfin rapidement testé la recherche de racine à l'aide de l'adaptation proposée de l'algorithme du paragraphe 2.3, dans le cadre d'un algorithme à scission systématique. Pour une meilleure visualisation, la fonction  $F_c$  employée a été un cercle bruité, de centre  $(0.3, 0.3)$ , de rayon 0.1.

$$\begin{aligned} F_c(\theta) &= (x - 0.3)^2 + (y - 0.3)^2 - (0.1)^2 + \sigma_B(U - 0.5), \\ \theta &= (x, y) \end{aligned}$$

avec  $U$  une variable aléatoire uniforme sur  $[0, 1]$ . Les résultats sont illustrés dans la figure 10, et font bien apparaître un tirage préférentiel aux environs du cercle solution.

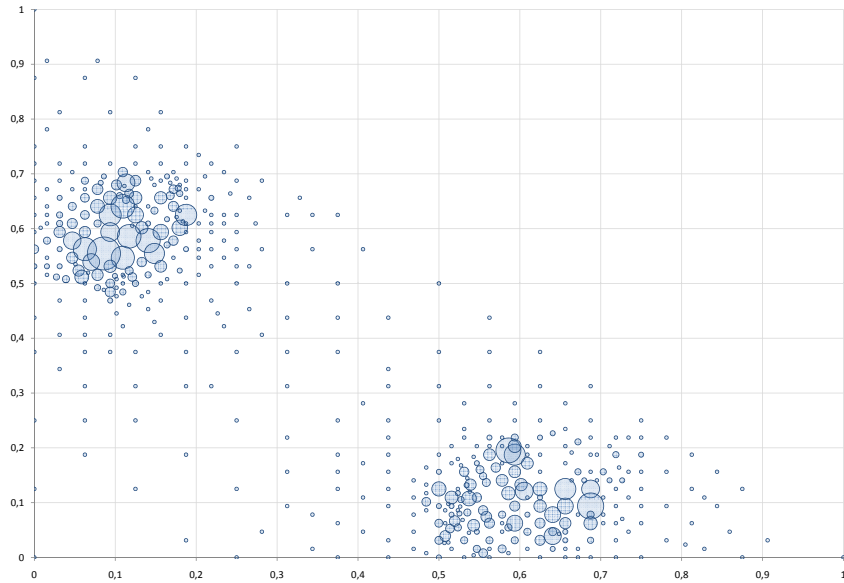


FIGURE 9 – Ensemble des points de simulations de  $F$  pour un niveau de bruit  $\sigma_B = 0.1$ , avec scission sur critère déterministe selon le meilleur potentiel (pour une variabilité  $\sigma_K = 5$  et un nombre minimal de tirage  $n_0 = 10$ ). La surface des bulles est proportionnelle au nombre de tirage de  $F$  en chaque point.

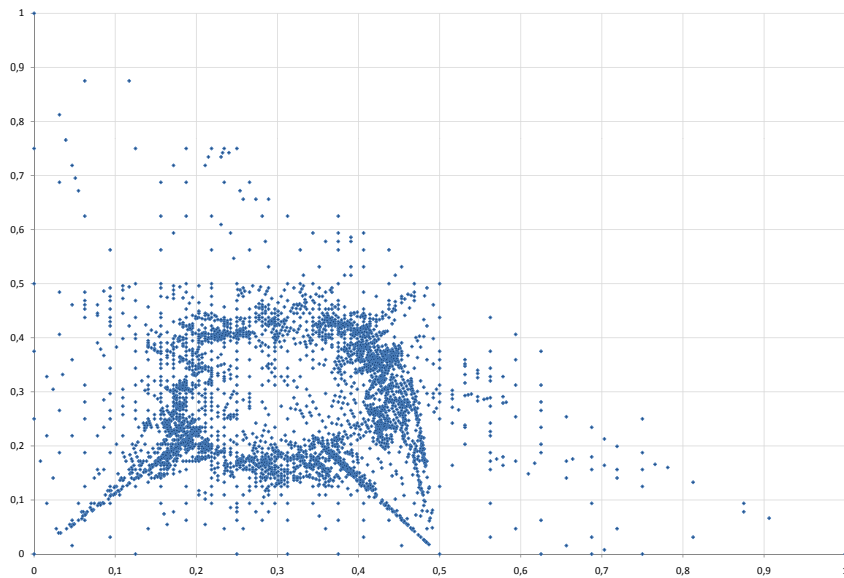


FIGURE 10 – Ensemble des points de simulations de  $F_c$  pour un niveau de bruit  $\sigma_B = 0.05$ , avec scission systématique (pour une variabilité  $\sigma_K = 1$  et un nombre minimal de tirage  $n_0 = 10$ ).

## Références

- Alliot, J.M. (1996), *Techniques d'optimisation stochastique appliquées aux problèmes du contrôle aérien*. INPT, Habilitation à Diriger des Recherches.
- Blum, J.R., (1954), *Multidimensional stochastic approximation methods*. Annals of Mathematical Statistics, 25, 737-744.
- Cohen, G., Culioli, J.-C. (1994), *Optimisation stochastique sous contraintes en espérance*. Rapport interne Centre Automatique et Systèmes, Ecole des Mines de Paris, no. A-288.
- Kiefer, J., Wolfowitz, J. (1952), *Stochastic estimation of the maximum of a regression function*. Annals of Mathematical Statistics, 23, 462-466.
- Robbins, H., Monro, S. (1951), *A Stochastic approximation method*. Annals of Mathematical Statistics, 22, 400-407.
- Strugarek, C. (2006), *Approches variationnelles et autres contributions en optimisation stochastique*. ENPC, Thèse de doctorat.

version de ce document en date du 27 août 2009, 13:05.