INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Méthode de gradients en optimisation combinatoire. Applications à un problème de sac à dos et au modèle d'Ising

Stéphane CRÉPEY

N° 2537

Avril 1995

PROGRAMME 1



ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

Méthode de gradients en optimisation combinatoire. Applications à un problème de sac à dos et au modèle d'Ising

Stéphane CRÉPEY*

Programme 1 — Architectures parallèles, bases de données, réseaux et systèmes distribués
Projet MISTRAL

Rapport de recherche n° 2537 — Avril 1995 — 45 pages

Résumé : On présente des heuristiques pour l'optimisation combinatoire en variables binaires. Ces heuristiques s'inspirent de descentes de gradients dans un espace des phases randomisé, coordonnée par coordonnée. On considère donc le problème d'optimisation comme un jeu à deux joueurs, le chercheur contre la $\mathcal{N}$ature. On obtient un algorithme quadratique pour le problème du sac à dos, extrêmement efficace en pratique. De plus on fournit une interprétation probabiliste à un algorithme de P. Rivière pour le problème d'Ising [19], dont on explique ainsi l'efficacité.

Mots-clé : optimisation en variables 0-1, barycentre, gradient, sac à dos, modèle d'Ising.

(Abstract: pto)

*. DEA IMA et ENSAE

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Gradient heuristics in 0-1 optimization

Abstract: We present gradient-like heuristics for combinatorial 0-1 optimization. Considering the optimization problem as a two players game, researcher against $\mathcal{N}$ature, we derive a phase space consisting of the mixed strategies of this game along independant coordinates axes, on which one can define gradients. This leads us to an algorithm for the linear knapsack problem, very efficient indeed. We then give a probabilistic interpretation to P. Rivière's [19] algorithm for the quadratic binary optimization problem.

Key-words: 0-1 optimization, barycentre, gradient, knapsack problem, Ising model.

> "In a discrete manifold, the principle or character of its metric relations is already given in the notion of the manifold, whereas in a continuous manifold this notion has to be found elsewhere, i.e. it has to come from outside. Either, therefore, thereality which underlies space must form a discrete manifold, or we must seek the ground of its metric relations (measure-conditions) outside it, in binding forces which act upon it." *B.Riemann, "Concerning the Hypotheses which lie at the Base of Geometry."*

1 Introduction

On présente une méthode pour la résolution approchée de problèmes d'optimisation combinatoire de la forme

$$\min_{\omega} \varphi(\omega) \quad (\text{s.c.} \quad \phi(\omega) \leqslant C),$$

où ω désigne un vecteur en variables binaires. Il s' agit de problèmes d'allocation optimale tels qu'on en rencontre fréquemment en *ordonnancement* de tâches en recherche opérationnelle, l'assignation d'une variable à 0 ou à 1 correspondant à l'absence ou à la présence d'une arête entre deux sommets d'un graphe; en *imagerie*, les variables binaires modélisant la couleur blanche ou noire d'un pixel de l'écran (voir[22]); ou en *finances*, pour l' élaboration des portefeuilles financiers (voir [4]). De plus tout problème d'optimisation combinatoire peut se reformuler en variables binaires, quitte à augmenter le nombre de variables. Comme les méthodes générales en ce domaine consistent largement en la formalisation de procédés utilisés pour résoudre tel ou tel cas particulier, on se réfère dans la suite à deux problèmes NP-complets bien connus (voir [8]): le *problème d'Ising*, ou minimisation d'une forme quadratique en variables binaires, et le *problème du sac à dos*, ou minimisation d'une forme linéaire sous contrainte linéaire en variables binaires. Ces problèmes apparaissent comme les analogues en variables binaires des programmations linéaires et quadratiques de l'optimisation convexe. Leur étude permet donc d'appréhender les difficultés que soulève le passage de variables continues en variables binaires, dans deux cas de figure contrastés, puisque la difficulté du problème linéaire tient dans sa contrainte, la forme quadratique représentant à elle seule une difficulté importante.

En pratique, on s'intéresse souvent aux politiques ou allocations optimales, plus encore qu'à la valeur correspondante du profit. De plus l'obtention d'une allocation quasi-optimale est également intéressante. Enfin l'obtention d'un vecteur qui minimise absolument φ est en

général un objectif trop difficile à atteindre, le problème étant NP-complet (voir [8]). On s'intéresse donc dans ce travail à des heuristiques conduisant à des algorithmes qui calculent, au prix d'une complexité "raisonnable" (bornée par un polynôme de faible degré en la taille du problème), un vecteur "suffisamment proche" de ce minimum. Le vecteur obtenu peut ensuite être utilisé comme approximation du minimum recherché, ou comme point de départ pour d'autres procédures d'optimisation.

Notre idée directrice est de raisonner statistiquement sur le grand nombre des configurations, source de la difficulté du problème de minimisation *absolue*, afin d'obtenir rapidement une bon minimum *approché*. Étant donnée une fonction en variables binaires φ à minimiser, on introduit donc les moyennes pondérées de φ, vue comme une fonctionnelle sur les champs bernouilliens de paramètre p (tirages à pile ou face indépendants suivant les coordonnées du vecteur ω). On obtient un polynôme multilinéaire $E_p\varphi$, appelé *potentiel* de φ, ayant des propriétés extrémales analogues à celles de φ: ainsi tout minimum local d' $E_p\varphi$ sur l'espace des phases E, muni de sa topologie *cubique* Θ_E, est-il atteint en un point à coordonnées 0 ou 1 (théorème 2.12). La clef de ce théorème est l'harmonicité d' $E_p\varphi$. On définit alors des descentes des gradients d' $E_p\varphi$ dans l'espace des phases, le long desquelles $E_p\varphi$ est décroissante (proposition 3.10). Une formule de Russo (voir [15]) permet d'estimer des gradients, et d'en déduire des algorithmes approchés pour certains problèmes (proposition 3.1). On présente ensuite un algorithme quadratique en espace et linéaire en temps pour le problème du sac à dos, extrêmement efficace en pratique (section 4.2). Enfin on fournit une interprétation probabiliste à un algorithme de P. Rivière pour le problème d'Ising [19], dont on explique ainsi l'efficacité (voir section 4.3).

Voici les principales notations employées dans la suite. $\omega(n)$ est la représentation binaire de l'entier n, $\mathcal{B}$ l'ensemble des représentations binaires d'entiers. $E - F$ est l'ensemble des points de E qui n'appartiennent pas à F (le complémentaire de F dans E, si $F \subset E$). Une partition d'un ensemble E est une famille de parties non vides de E, appelés termes, formant un remplissage sans recouvrement ni lacune de E. Si Π est une partition de E, et $F \subset E$, alors $\Pi_F = (\pi \bigcap F)_{\{\pi\in\Pi\ ;\ \pi\cap F\neq\emptyset\}}$ est une partition, dite Π-induite, de F. La partition induite est dite propre si elle se compose de parties de Π, soit $\forall \pi \in \Pi \quad (\pi \cap F = \emptyset \ \vee \ \pi \subset F)$. Relativement à un minimum, on utilise les termes de *local* et *global* pour une fonction définie sur un continuum, de *relatif* et *absolu* sinon. Enfin dans un espace affine $\mathcal{E}$ on note env($\mathcal{F}$)

l'enveloppe convexe d'une partie $\mathcal{F} \subset \mathcal{E}$, et squel($\mathcal{F}$) le squelette, ou ensemble des sommets de $\mathcal{F}$.

2 Pondérations

Dans cette section on plonge le problème combinatoire dans un problème continu en définissant, à partir des allocations *pures* en variables binaires, des allocations *mixtes*, constituées de loteries indépendantes sur les coordonnées du vecteur ω. Un tel plongement est classique en théorie des jeux. De fait toute optimisation peut être vue comme un jeu à deux joueurs, le chercheur contre la $\mathcal{N}$ature, avec pour fonction de perte φ (voir [1, ch.5]). Supposons que le chercheur choisit sa stratégie de faç aléatoire, ou mixte, avec une distribution p. On obtient ainsi un nouveau jeu, par moyennisation ou randomisation du jeu pur initial, dans lequel celui-ci est plongé: on obtient à nouveau le jeu initial en considérant seulement les stratégies pures du chercheur.

2.1 Espace des phases

La nécessité de raisonner sur des suites infinies de cubes emboités (voir section 4.2) nous conduit aux définitions suivantes.

Définition 2.1 (Configurations, hypercubes, cubes) *On appelle*
• *configuration ω, toute suite presque nulle de 0 et de 1, ou représentation binaire d'entier.*
• *hypercube Ω (resp. cube Γ), toute suite de $\{0\}$, de $\{1\}$, et de $\{0,1\}$ (resp. de $[0,1]$), dont les termes sont $\{0\}$ à partir d'un certain rang.*

L'ensemble des cubes est noté $\mathcal{G}$. On définit également l'union des cubes $I = \cup_{\Gamma \in \mathcal{G}} \Gamma \subset [0,1]^{\mathbb{N}}$. Par espace des phases E, on entend désormais indifféremment un cube Γ, ou l'union des cubes I . De même $\mathcal{D}$ désigne un hypercube Ω, ou l'union des hypercubes $\mathcal{B}$.

Définition 2.2 (Cube engendré) *On appelle*
• *cube engendré par une famille* finie *de points de E (resp. de cubes), l'intersection des cubes contenant tous ces points (resp. ces cubes).*
• *support d'un point $p \in E$, $supp(p) = \{n \in \mathbb{N} \ ; \ p_n \in]0,1[\}$.*
• *dimension de $p \in E$, $dim(p) = Card(supp(p))$.*
• *support (resp. dimension) de E, $supp(E) = \cup_{p \in E} supp(p)$ (resp. $dim(E) = \sup_{p \in E} dim(p)$).*

Le cube engendré par les points p, $q, \cdots r$ est noté $\Gamma(p, q, \cdots, r)$.
Nos résultats s'énoncent naturellement moyennant l'introduction d'une topologie dite *cubique* sur l'espace des phases (voir théorème 2.12), obtenue à partir des topologies euclidiennes sur les cubes de la façon suivante.

Définition 2.3 (Topologie cubique) *A partir des topologies euclidiennes induites sur les cubes par la topologie produit sur $\mathbb{R}^{\mathbb{N}}$, on définit*
• *l' intérieur d'un cube $\overset{\circ}{\Gamma}$, ou $Int(\Gamma)$, et sa frontière $\partial\Gamma$, au sens euclidien.*
• *pour un cube non ponctuel, l'ensemble des faces*

$$\mathcal{F}(\Gamma) = \{\Gamma' \in \mathcal{G} \ ; \ \Gamma' \subset \Gamma \ \wedge \ dim(\Gamma') = dim(\Gamma) - 1\},$$

et la surface, ou union des faces, $\sigma(\Gamma) = \bigcup_{\Gamma' \in \mathcal{F}(\Gamma)} \Gamma'$. Si Γ est ponctuel, $\mathcal{F} \triangleq \sigma(\Gamma) \triangleq \Gamma$.
• *la topologie cubique sur E, engendrée par les topologies euclidiennes sur les cubes $\Lambda \subset E$.*

Soit $\mathcal{D}$ un hypercube (resp. l'union généralisée des hypercubes), E le cube ayant les mêmes sections $\{0\}$, $\{1\}$ que $\mathcal{D}$ (resp. l'union généralisée des cubes). La topologie cubique sur E est notée Θ_E. Pour $n \in \text{supp}(p)$, on note $p_{(n,0)}$ (resp. $p_{(n,1)}$) la suite obtenue à partir de p en changeant en 0 (resp.en 1) la n-ième coordonnée.

Lemme 2.4 (Groupes des échanges) *On a*
i. *Dans l'espace affine produit $\mathbb{R}^{\mathbb{N}}$,*

$$\left\{ \begin{array}{lcl} E & = & env(\mathcal{D}) \\ \mathcal{D} & = & squel(E) \end{array} \right. .$$

ii. *Le cube engendré par un ensemble de points fini de points (resp. de cubes) est un cube, le plus petit*[1] *cube contenant tous ces points (resp. ces cubes).*
iii. *Les transformations $p_n \rightarrow (1 - p_n)$ telles que $n \in supp(E)$ induisent un groupe d'action transitive involutive $\mathcal{S}_{\mathcal{D}}$ sur $\mathcal{D}$, appelé groupe des échanges de $\mathcal{D}$.*

On note $(g, p) \rightarrow gp$ l 'action du groupe des échanges $\mathcal{G}_{\mathcal{D}}$ sur l'espace des phases. Ainsi

$$(gp)_n = \left\{ \begin{array}{cc} 1 - p_n & \text{si} \ \ g_n = 1 \\ p_n & \text{sinon} \end{array} \right. .$$

Preuve
i. Laissé au lecteur

1. pour $\subset$.

ii. Soient $\mathcal{P}$ une famille finie de points de E, $\widehat{\Gamma} = (\widehat{\Gamma}_n)_{n\in\mathbb{N}}$ le cube engendré par $\mathcal{P}$. Alors

$$\widehat{\Gamma}_n = \begin{cases} [0,1] \text{ si } \exists p \in \mathcal{P} \ ; \ p_n \in]0,1[\text{ ou } \exists (p,q) \in \mathcal{P}^2 \ ; \ (p_n,q_n) \in \{(0,1),(1,0)\} \\ p_n = q_n \in \{0,1\} \text{ sinon} \end{cases} .$$

On a donc bien $\widehat{\Gamma} \in \mathcal{G}$. De plus, si $\mathcal{P}$ est une famille finie de cubes, $\widehat{\Gamma}$ est le cube engendré par la famille (finie) des sommets des cubes de $\mathcal{P}$.

iii. Les échanges d'une coordonnée étant involutifs engendrent des groupes involutifs par composition. Soient $(\omega^1, \omega^2) \in \mathcal{D}^2$, $A = \{n \in \mathbb{N} \ ; \ \omega^1_n \neq \omega^2_n\}$, g la composée des échanges $p_n \rightarrow (1-p_n)$ tels que $n \in A$. On a $g\omega_1 = \omega_2$. De plus par définition $A \subset \mathrm{supp}(\Gamma(\omega_1,\omega_2)) \subset \mathrm{supp}(E)$, donc $g \in \mathcal{S}_{\mathcal{D}}$, ce qui montre la transitivité de l'action de $\mathcal{S}_{\mathcal{D}}$ sur $\mathcal{D}$.

□

Commentaire 2.5 (Réalisations géométriques) Notons $\widehat{\Gamma}$ et appelons support géométrique de Γ le produit de ses sections $[0,1]$ plongé dans $\mathbb{R}^d$, où $d = \dim(\Gamma)$. $\widehat{\Gamma}$ est le cube géométrique euclidien usuel de dimension d. Les échanges $p_n \rightarrow 1-p_n$ tels que $n \in \mathrm{supp}(\Gamma)$ correspondent géométriquement aux symétries invariantes de $\widehat{\Gamma}$, par rapport aux hyperplans équidistants de ses faces opposées. $\mathcal{S}_\Gamma$ correspond donc géométriquement à un groupe de symétries invariantes de $\widehat{\Gamma}$. L' énoncé 2.4.iii. signifie donc qu'en composant de telles symétries, on peut appliquer tout sommet de $\widehat{\Gamma}$ sur n'importe quel autre.

La topologie cubique se décrit également à l'aide de la partition $\Pi \triangleq (\{0\},]0,1[,\{1\})^{\mathbb{N}}$ de $[0,1]^{\mathbb{N}}$.

Lemme 2.6 (Π-partitions) *On a*

i. *Si* $(\Gamma,\Gamma') \in \mathcal{G}^2$

$$\left.\begin{array}{c} \Gamma \subset \Gamma' \\ dim(\Gamma) = dim(\Gamma') \end{array}\right\} \Gamma = \Gamma'.$$

Donc si $\Gamma \subset \Gamma'$, *on a* $\Gamma = \Gamma'$, *ou* $\Gamma \subset \mathcal{F}(\Gamma')$ [2].

ii. *Toute suite décroissante de cubes est stationnaire (l'ensemble ordonné* $(\mathcal{G},\subset)$ *est nœthérien).*

iii. *La surface d'un cube égale sa frontière,* $\sigma(\Gamma) = \partial\Gamma$.

iv. *Les termes de* Π_E *sont les intérieurs des cubes* $\Lambda \subset E$, $\Pi_E = (\mathring{\Lambda})_{\{\Lambda\in\mathcal{G}\ ;\ \Lambda\subset E\}}$. *La* Π*-partition de* E, Π_E, *est propre. La* Π*-partition de la surface d'un cube,* $\Pi_{\sigma(\Gamma)}$, *est propre, et*

2. Les deux sont vrais si $\dim(\Gamma') = 0$.

s'obtient à partir de Π_Γ *en enlevant le terme correspondant à l'intérieur du cube*

$$\begin{cases} \Pi_{\sigma(\Gamma)} \quad \textit{propre} \\ \Pi_{\sigma(\Gamma)} = \Pi_\Gamma - \overset{\circ}{\Gamma} \end{cases} .$$

v. *Si* Λ *est un cube* $\subset E$*, la partition de* Λ *induite par* Π_E *est propre, égale à* Π_Λ*.*

Preuve

i. Laissé au lecteur.

ii. A chaque fois que la suite des cubes décroît strictement, il en va de même de celle des dimensions, d'après **i**. Les dimensions prenant leurs valeurs dans $\mathbb{N}$, ceci ne peut se produire qu'un nombre fini de fois.

iii. Laissé au lecteur.

iv. Un terme $\pi \in \Pi$ admet soit une infinité de sections $]0,1[$, auquel cas aucun point de I n'appartient à π; soit un nombre fini de sections $]0,1[$, auquel cas $\pi \subset I$. Π_I est donc propre, et s'obtient à partir de Π en sélectionnant les termes comportant un nombre fini de sections $]0,1[$, soit $\Pi_I = (\overset{\circ}{\Gamma})_{\Gamma \in \mathcal{G}}$. Montrons maintenant que les Π-partitions des cubes sont propres, par récurrence sur la dimension. Si $\Gamma = \{\omega\}$, alors $\Pi_\Gamma = \{\omega\} \in \Pi$, propre. Supposons $\dim(\Gamma) = d > 0$, et le résultat vrai en dimension $d-1$. Par ce qui précède, $\overset{\circ}{\Gamma} \in \Pi$. De plus, $\Gamma - \overset{\circ}{\Gamma} = \partial\Gamma = \sigma(\Gamma)$. Les faces de Γ sont des cubes de dimension $d-1$, donc leurs partitions Π-induites sont propres, par hypothèse de récurrence. Donc $\Pi'_\Gamma = \{\pi \in \Pi \ ; \ \exists f \in \mathcal{F}(\Gamma) \ \ \pi \in \Pi_f\}$ est constitué de termes de Π. Soit $p \in \Gamma$. Si $p \in \sigma(\Gamma)$, p appartient à au moins une face f de Γ, donc à un terme $\pi \in \Pi_f \subset \Pi'_\Gamma$. De plus Π_f étant propre et π appartenant à Π_f on a $\pi \subset f \subset \sigma(\Gamma)$. On a donc une famille Π'_Γ de termes de Π inclus dans $\sigma(\Gamma)$, recouvrant $\sigma(\Gamma)$. Donc Π'_Γ n'est autre que $\Pi_{\sigma(\Gamma)}$, qui est donc propre. De plus $\overset{\circ}{\Gamma} \in \Pi$. Donc Π_Γ s'obtient à partir de $\Pi_{\sigma(\Gamma)}$ par adjonction de $\overset{\circ}{\Gamma}$, et Π_Γ est également propre.

v. Si Λ est un cube $\subset E$, Π_Λ et Π_E sont propres, d'après **iv**. Donc $\Pi_\Lambda = \{\pi \in \Pi \ ; \ \pi \subset \Lambda\} \subset \{\pi \in \Pi \ ; \ \pi \subset E\} = \Pi_E$. Posons $\Pi'_\Lambda = \{\pi \in \Pi_E \ ; \ \pi \subset \Lambda\}$. Π'_Λ est une famille de termes de Π_E formant une partition de Λ. Donc Π'_Λ n'est autre que la partition de Λ induite par Π_E, qui est donc propre, égale à Π_Λ.

□

Dans la proposition suivante la terminologie employée est celle d'Arnold [24, ch.5].

Proposition 2.7 (Le point de vue des variétés) *On a*

i. *Pour tout* $p \in E$*,* $p \in Int(\Gamma(p))$*. Si* Λ *est un cube* $\subset E$*,* $\overset{\circ}{\Lambda}$ *et* Λ *sont ouverts pour la topologie induite sur* Λ *par* Θ_E*. Cette topologie n'est donc autre que* Θ_Λ*.*

ii. *Paramétrisant le point $p \in E$ par la famille $(p_n \ ; \ n \in supp(p)) \in R^d$, où $d = dim(\Gamma)$, on munit l'espace E d'une structure de variété analytique localement euclidienne séparable dénombrable $\mathcal{V}_E$, un atlas étant constitué par la famille de cartes*

$$\mathcal{A}_E = (\overset{\circ}{\Gamma}, Id_\Gamma,]0,1[^{dim(\Gamma)})_{\overset{\circ}{\Gamma} \in \Theta_E},$$

où Id_Γ désigne l'application

$$\begin{array}{ccc} Id_\Gamma : \overset{\circ}{\Gamma} & \longrightarrow & \mathbb{R}^{dim(\Gamma)} \\ q = (q_n)_{n \in \mathbb{N}} & \longrightarrow & (q_{n_m})_{n_m \in supp(\Gamma)} \end{array} .$$

$dim(p)$ n'est autre que la dimension de la variété localement euclidienne $\mathcal{V}_E$ au point p.

iii. *Etant donné $\epsilon > 0$, on définit $\mathcal{R}^\epsilon$ par*

$$\mathcal{R}^\epsilon_n = \begin{cases} \{p_n\} & si \quad n \notin supp(p) \\ ]p_n - \epsilon, p_n + \epsilon[& sinon \end{cases} .$$

L'ensemble

$$\mathcal{B}_E = \{\mathcal{R}^\epsilon \ \ ; \ \ (p, \epsilon) \in (E \times]0, \epsilon[)\},$$

où $\overline{\epsilon} = \min\{p_n, 1 - p_n; n \in supp(p)\}$ forme une base d'ouverts de (E, Θ_E). On obtient une base dénombrable $\mathcal{D}_E$ si l'on impose les conditions supplémentaires $p \in \mathbb{Q}^{\mathbb{N}}$, $\epsilon \in \mathbb{Q}$.

iv. *La topologie de $(E, \mathcal{V}_E)$ n'est autre que la topologie cubique Θ_E. Tout cube $\Lambda \subset E$ est une sous-variété de $(E, \mathcal{V}_E)$, dont la topologie induite par $(E, \mathcal{V}_E)$ n'est autre que Θ_Λ. Donc on a $(\Lambda, \mathcal{V}_\Lambda, \Theta_\Lambda)$, considéré selon les cas comme espace de référence ou sous-espace de E, mais indépendant du point de vue en tant que variété topologique.*

v. *La Π-partition de E n'est autre que la décomposition en composantes connexes de $(E, \mathcal{V}_E)$, $E = \amalg_{\{\Gamma \in \mathcal{G} \ ; \ \Gamma \subset E\}} \overset{\circ}{\Gamma}$.*

vi. *E σ-compact; E non compact.*

Preuve

i. Soit $p \in E$. Si $p \notin \mathrm{Int}(\Gamma(p))$ alors d'après le lemme 2.6 $p \in \sigma(\Gamma)$, donc $\exists f \in \mathcal{F}(\Gamma)$ tel que $p \in f$. f est donc un sous-cube de $\Gamma(p)$ contenant p, donc $f = \Gamma(p)$. Or $\Gamma(p)$ ne peut coïncider avec l'une de ses faces que s'il est réduit à une configuration ω (lemme 2.6), et pour tout $\omega \in \mathcal{B}$ on a $\omega \in \mathrm{Int}(\Gamma(\omega)) = \{\omega\}$. Ainsi on a bien $p \in \mathrm{Int}(\Gamma(p))$, quel que soit $p \in I$. Soit maintenant $\Gamma \in \mathcal{G}$. Γ est ouvert comme union des termes de sa Π-partition, qui sont des intérieurs de cubes, donc des ouverts (lemme 2.6). Donc si Λ est un cube $\subset E$, la topologie Θ_Λ est constituée des ouverts de Θ_E contenus dans Λ.

ii. L'atlas $\mathcal{A}_E$ munit trivialement l'ensemble E d'une structure de variété analytique $\mathcal{V}_E$. De plus par définition de $\mathcal{A}_E$ la topologie de $(E, \mathcal{V}_E)$ n'est autre que Θ_E. $\mathcal{V}_E$ est séparable, car deux points distincts p et p' étant donnés dans E, soit $\mathrm{Int}(\Gamma(p)) = \mathrm{Int}(\Gamma(p')) \stackrel{\Delta}{=} \overset{\circ}{\Gamma}$, alors les deux points p et p' sont représentés sur la carte Id_Γ; soit $\mathrm{Int}(\Gamma(p))$ et $\mathrm{Int}(\Gamma(p'))$ sont disjoints, alors $Id_{\Gamma(p)}$ et $Id_{\Gamma(p')}$ sont deux cartes de domaines disjoints contenant respectivement p et p'. L'espace topologique (E, Θ_E) est donc séparé. De plus $\mathcal{V}_E$ est dénombrable. En effet si $E = I$ $\mathcal{A}_E$ est dénombrable, et $\mathcal{A}_E$ est fini si E est un cube Γ.

iii. Etant donné un domaine $\mathcal{U} =]0,1[^d$ ouvert dans $\mathbb{R}^d$, tel que $d \leqslant \dim(E)$, on choisit une carte $Id \in \mathcal{A}$ de domaine $\mathcal{U}$, associée à un cube de dimension d dans E. $Id(\mathcal{B}_E)$ constituant classiquement une base de voisinages de $\mathcal{U}$, $\mathcal{B}_E$ est une base d' ouverts de Θ_E. De plus si l'on impose la condition $p \in \mathbb{Q}^{\mathbb{N}}$, $\epsilon \in Q$, on obtient une base $\mathcal{D}_E$ dénombrable. En effet, $\mathcal{D}_E = \cup_{n\in\mathbb{N}} \mathcal{D}_E^n$, où $\mathcal{D}_E^n \stackrel{\Delta}{=} \{\mathcal{R}^\epsilon \in \mathcal{D}_E \ ; \ \mathrm{supp}(p) \subset \mathbb{N}_n\}$ est dénombrable.

iv. Si Λ est un cube $\subset E$, alors d'après **iii.** tout point $p \in \Lambda$ posséde le voisinage $\mathrm{Int}(\Lambda(p)) \in \Theta_E$, sur lequel est défini la carte $Id_{\Lambda(p)} \in \mathcal{A}_E$. Donc Λ est une sous-variété de E. Comme la partition induite sur Λ par Π_E est propre (lemme 2.6), la topologie de variété induite sur Λ par $\mathcal{V}_E$ n'est autre que Θ_Λ.

v. Les domaines des cartes de $\mathcal{A}_E$ partitionnant E, deux points p et q sont reliés par une chaîne de cartes[3] si et seulement si ils appartiennent au même terme $\pi \in \Pi$. La décomposition en composantes connexes de $(E, \mathcal{V}_E, \Theta_E)$ est donc bien la Π-partition de E.

vi. Si E non ponctuel, alors Π_E contient un terme non ponctuel $\pi \in \Pi$. π est ouvert, d'après **i**. Par ailleurs π est non compact, comme image identité inverse du produit non compact $]0,1[^d \subset \mathbb{R}^d$, où $d = \dim(\pi)$. Donc π admet un recouvrement ouvert $\mathcal{R}_\pi$ sans sous-recouvrement fini. Adjoignant à $\mathcal{R}_\pi$ les termes de $\Pi_E - \{\pi\}$, on obtient un recouvrement ouvert $\mathcal{R}$ de E. Soient $\mathcal{Q}$ un sous-recouvrement de $\mathcal{R}$, $\mathcal{Q}_\pi$ le recouvrement induit sur π. Les termes de $\Pi - \{\pi\}$ étant disjoints de π, $\mathcal{Q}_\pi$ est un sous-recouvrement de $\mathcal{R}_\pi$, donc infini, par définition de $\mathcal{R}_\pi$. *A fortiori* il en va de même de $\mathcal{Q}$. Ainsi tout sous-recouvrement de $\mathcal{R}$ est infini, donc E non compact. En revanche chaque $\overset{\circ}{\Gamma}$ est σ-compact, en utilisant la base $\mathcal{D}_\Gamma$, donc E est σ-compact comme union au plus dénombrable d'ensembles σ-compacts.

□

3. Séquence finie de cartes dont chaque suivante intersecte la précédente.

Dans la suite, on identifie les cordonnées $p_n \in]0,1[$ et le paramétrage au voisinage d'un point $p \in E$.

Exemple 2.8 (Laplacien) Le laplacien s'exprime à travers $\mathcal{A}_E$ au voisinage d'un point p^0 par la formule classique

$$\Delta = \sum_{n \in \mathrm{supp}(p^0)} \frac{\partial^2}{\partial p_n^2} \tag{1}$$

2.2 Potentiel

On définit maintenant le *potentiel* de φ sur E, par analogie avec la théorie du potentiel (voir théorème 2.12). Le problème d'optimisation d'une fonction φ sur l'hypercube Ω est alors traité comme le problème d'optimisation de son potentiel $E_p\varphi$ sur l'enveloppe convexe Γ. Cependant le problème continu obtenu n'est pas un problème de programmation convexe. La NP-complétude du problème initial se "réfugie" dans la non-convexité du problème continu, dont la résolution approchée exige une approche originale, fondée sur les symétries au milieu du cube (voir section 4).

Définition 2.9 (Pondérations et potentiels, point de vue global) *On définit*
• *Si $p \in E$, soit X la suite de variables aléatoires sur $[0,1]$ telle que*

$$X_n = \begin{cases} \delta(0) & si \quad p_n = 0 \\ B(1,p_n) & si \quad p_n \in]0,1[\\ \delta(1) & si \quad p_n = 1 \end{cases}, \tag{2}$$

$B(1,p_n)$ désignant la loi de Bernouilli de paramètre p_n. Le champ bernouillien en p est par définition le produit tensoriel $\wedge_p = \otimes_{n\in\mathbb{N}} X_n$. La probabilité de la configuration $\omega \in \mathcal{B}$ sous le champ $\wedge_p$ se note aussi $p \wedge \omega$, soit avec la convention $0^0 = 1$

$$\wedge_p(\omega) = p \wedge \omega = \prod_{n\in\mathbb{N}} p_n^{\omega_n} (1-p_n)^{(1-\omega_n)}.$$

• *L'espace des fonctions $\varphi : \mathcal{B} \longrightarrow \mathbb{R}$ forme un espace vectoriel sur $\mathbb{R}$, pour les lois produits évidentes, composante par composante. On considère en particulier la famille des masses de Dirac $(\delta_n)_{n\in\mathbb{N}}$ telle que $\forall m \in \mathbb{N} \quad \delta_n(\omega(m)) = \delta_{n,m}$. Étant donnée $\varphi : \mathcal{B} \longrightarrow \mathbb{R}$, on note $E_p\varphi$ la moyenne pondérée de la fonction φ sous le champ bernouillien $p \in E$, soit*

$$E_p\varphi = \sum_{\omega\in\mathcal{B}} \varphi(\omega)\, p \wedge \omega. \tag{3}$$

On appelle potentiel de φ , et on note génériquement $E_p\varphi$, la fonction obtenue quand p varie dans E.

Proposition 2.10 (Pondérations et potentiels, point de vue local) *Pour tout $p \in E$*
i. *$\wedge_p$ est bien défini, de support (de pondération) $supp(\wedge_p) = \Omega(p)$. De plus si $\omega \in \Omega(p)$ alors pour tout $A \supseteq supp(p)$*

$$p \wedge \omega = \prod_{n \in A} p_n^{\omega_n} (1 - p_n)^{(1-\omega_n)}. \tag{4}$$

Le champ de pondérations $(\wedge_p)_{p \in E}$ est donc défini localement.
ii. *Pour tout $\Omega \supseteq \Omega(p)$*

$$E_p\varphi = \sum_{\omega \in \Omega} \varphi(\omega)\, p \wedge \omega. \tag{5}$$

Ainsi $E_p\varphi$ se détermine localement sur $(E, \mathcal{V}_E)$. Si $p \in \overset{\circ}{\Gamma}$, $E_p\varphi$ est un polynôme multilinéaire (affine composante par composante) en les coordonnées locales de la carte Id_Γ.

Preuve
i. En effet si $\omega \notin \Omega(p)$ alors $\exists n \in \mathbb{N}$ tel que $(p_n, \omega_n) = (0,1)$ ou $(1,0)$, donc $p_n^{\omega_n}(1 - p_n)^{(1-\omega_n)} = 0$, donc $p \wedge \omega = 0$, soit $\omega \notin supp(\wedge_p)$. Si maintenant $\omega \in \Omega(p)$ mais $n \notin \mathrm{supp}(p)$, alors $p_n^{\omega_n}(1 - p_n)^{(1-\omega_n)} = 0^0 = 1$. Donc pour tout $\omega \in \Omega(p)$ et $A \supseteq \mathrm{supp}(p)$ on a $p \wedge \omega = \prod_{n \in A} p_n^{\omega_n}(1 - p_n)^{(1-\omega_n)}$.
ii. Car $p \wedge \omega = 0$ si $\omega \notin \Omega(p)$ d'après **i**.

□

Commentaire 2.11 (Interprétation volumique des pondérations) Les plans contenant p parallèles aux faces de $\widehat{\Gamma(p)}$ découpent l'espace $\mathbb{R}^{\dim(p)}$ en boites rectangulaires, appelées cellules de Voronoï, ou boites de Dirichlet. Chaque sommet définit une boite, formée des points de l'espace dont il est le sommet le plus proche, en tenant compte d'une pondération définie par p. Le découpage obtenu s'appelle *diagramme de Voronoï* pondéré associé au cube en p. $p \wedge \omega$ est le volume de la boite opposée à ω (contenant le point diagonalement opposé de $\widehat{\Gamma(p)}$) dans ce découpage. Enfin p est le barycentre de la famille $(\omega)_{\omega \in \Omega(p)}$ munie de la pondération $(p \wedge \omega)_{\omega \in \Omega(p)}$ (Voir figure 1.A). Tout barycentre de sommets d'un cube géométrique est l'intersection d'une famille d'hyperplans canoniquement associée à la pondération qui le définit. Mais ces hyperplans ne sont pas parallèles aux faces du cube, donc on n'a pas l'interprétation volumique précédente, si le champ des pondérations n'est

pas bernouillien (figure 1.B). Les champs bernouilliens sur un hypercube Ω de dimension d forment une variété de dimension d sur $\mathbb{R}$, tandis que la variété de toutes les pondérations sur Ω est de dimension $2^d - 1$ (2^d masses, sommant à 1). L'espace des champs bernouilliens est donc beaucoup plus petit que l'espace de toutes les pondérations. Il contient cependant toutes les stratégies pures. Enfin c'est déjà un continuum. On choisit donc de travailler sur l'espace des phases.

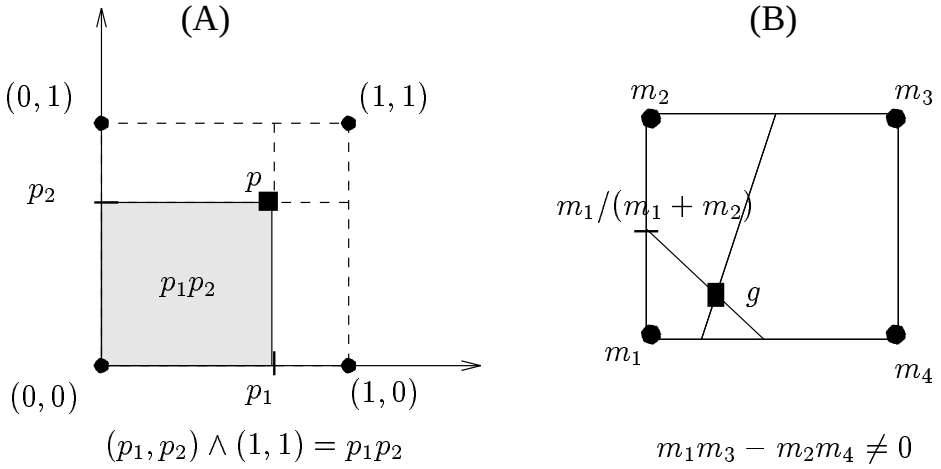


FIG. 1 - *Pondération bernouillienne (A) et non bernouillienne (B) sur les sommets du carré.*

Théorème 2.12 ("Principe du minimum") *Si p^0 est un minimum local d' $E_p\varphi$ sur (E, Θ_E), alors $E_p\varphi$ est constante sur $\Gamma(p^0)$. Donc $E_{p^0}\,\varphi$ est une valeur de φ.*

Preuve. La fonction $E_p\varphi$, localement multilinéaire, est harmonique, d'après l'expression locale du laplacien fournie en exemple 2.8). Soit $\mathcal{O}$ ouvert contenant p^0, sur lequel le potentiel de φ est minimum en p^0. $\mathcal{O}$ contient un ouvert de $\mathbb{R}^d$ à travers l'atlas $\mathcal{A}_E$, où $d = \dim(p^0)$. Une fonction harmonique sur un ouvert de $\mathbb{R}^d$ vérifie le principe du minimum, elle ne peut admettre de minimum local sans être localement constante. Quitte à restreindre $\mathcal{O}$ à un ouvert plus petit on peut donc supposer que $E_p\varphi$ est constante sur $\mathcal{O}$. On peut également prendre $\mathcal{O}$ dans la base d'ouverts $\mathcal{B}_{\Gamma(p^0)}$ définie dans la proposition 2.7. On a donc $\epsilon > 0$ tel que

$$\mathcal{O} = (\mathcal{O}_n)_{n\in\mathbb{N}} \;\; ; \;\; \mathcal{O}_n = \begin{cases} \{p_n^0\} & \text{si} \quad n \notin \mathrm{supp}(p^0) \\ ]p_n^0 - \epsilon, p_n^0 + \epsilon[& \text{sinon} \end{cases} .$$

On montre maintenant par récurrence sur $\dim(p^0)$ le lemme suivant, qui achève la démonstration.

Lemme 2.13 *Avec les notations précédentes,*

$$\forall p \in \mathcal{O} \quad E_p\varphi = E_{p^0}\varphi \quad \Longrightarrow \quad \forall p \in \Gamma(p^0) \quad E_p\varphi = E_{p^0}\varphi.$$

Preuve. Si $\dim(p^0) = 0$, c'est évident. Supposons le lemme vrai si $\dim(p^0) = d-1 \geqslant 0$. Soit p^0 tel que $\dim(p^0) = d$. Fixons $n \in \mathrm{supp}(p^0)$. Comme $\mathcal{O} \subset \Gamma(p^0)$, alors d'après la proposition 2.10, pour tout $p \in \mathcal{O}$

$$E_p\varphi = \sum_{\omega \in \Omega(p^0)} \varphi(\omega) p \wedge \omega.$$

Soit $p \wedge^n \omega \stackrel{\Delta}{=} \prod_{m \in \mathrm{supp}(p^0)-\{n\}} p_m^{\omega_m}(1-p_m)^{(1-\omega_m)}$. Alors

$$\begin{aligned} & p \wedge \omega \\ &= \prod_{m \in \mathrm{supp}(p^0)} p_m^{\omega_m}(1-p_m)^{(1-\omega_m)} \\ &= p_n^{\omega_n}(1-p_n)^{(1-\omega_n)} \prod_{m \in \mathrm{supp}(p^0)-\{n\}} p_m^{\omega_m}(1-p_m)^{(1-\omega_m)}, \end{aligned}$$

soit $p \wedge \omega = p_n^{\omega_n}(1-p_n)^{(1-\omega_n)} p \wedge^n \omega$, où

$$p_n^{\omega_n}(1-p_n)^{(1-\omega_n)} = \left\{ \begin{array}{ccc} p_n & \text{si} & \omega_n = 1 \\ 1-p_n & \text{si} & \omega_n = 0 \end{array} \right. .$$

Notons $\mathcal{H}^\alpha$ l'hyperplan $p_n = \alpha$. Soient $p_{(n,0)}$ (resp. $p_{(n,1)}$) le projeté orthogonal de $p \in \mathcal{H}^\alpha$ sur l'hyperplan $\mathcal{H}^0$ (resp. $\mathcal{H}^1$), et $\Omega^n(p^0)$ le projeté orthogonal de $\Omega(p^0)$ sur $\mathcal{H}^\alpha$. $\Omega^n(p^0)$ est donc identique à un hypercube de dimension $d-1$. De plus si $p \in \Gamma(p^0)$

$$\begin{aligned} & E_p\varphi \\ &= \sum_{\omega \in \Omega(p^0)} \varphi(\omega)\, p \wedge \omega \\ &= \sum_{\omega \in \Omega(p^0) \cap \{\omega_n = 1\}} \varphi(\omega)\, p_n\, p \wedge^n \omega + \sum_{\omega \in \Omega(p^0) \cap \{\omega_n = 0\}} \varphi(\omega)\, (1-p_n)\, p \wedge^n \omega, \end{aligned}$$

soit

$$E_p\varphi = \sum_{\omega \in \Omega^n(p^0)} \left(p_n\varphi(\omega_{n,1}) + (1-p_n)\varphi(\omega_{n,0})\right) p \wedge^n \omega.$$

Introduisons sur $\Omega^n(p^0)$ la fonction φ^α définie par

$$\varphi^\alpha(\omega) = \alpha\varphi(\omega_{(n,1)}) + (1-\alpha)\,\varphi(\omega_{(n,0)}).$$

Notons $E_p^\alpha\varphi$ la restriction à $\mathcal{H}^\alpha$ de $E_p\varphi$. $E_p^\alpha\varphi$ apparaît donc comme le potentiel associé à la fonction combinatoire φ^α. Soit $\mathcal{O}^\alpha = \mathcal{O}\cap\mathcal{H}^\alpha$. $\mathcal{O}^\alpha$ est la trace de l'ouvert $\mathcal{O}$ sur l'hyperplan $\mathcal{H}^\alpha$. $\mathcal{O}^\alpha$ s'identifie donc à un ouvert de $\mathbb{R}^{d-1}$. De plus si $\alpha \in]p_n^0-\epsilon, p_n^0+\epsilon[$ alors $E_p^\alpha\varphi$ est constante et égale à $E_{p^0}\varphi$ sur $\mathcal{O}^\alpha$. Donc d'après l'hypothèse de récurrence appliquée à la fonction $E_p^\alpha\varphi$ sur $\mathcal{O}^\alpha$, notant $\Gamma^n(p^0)$ l'enveloppe convexe de $\Omega^n(p^0)$

$$\forall p \in \Gamma^n(p^0)\quad E_p\varphi^\alpha = E_{p^0}\varphi.$$

En particulier sur $\Omega^n(p^0)$

$$\forall\alpha\in]-\epsilon,\epsilon[\quad \alpha\,\varphi(\omega_{(n,1)}) \;+\; (1-\alpha)\,\varphi(\omega_{(n,0)}) = E_{p^0}\varphi.$$

D'où en dérivant en 0 par rapport à α

$$\varphi(\omega_{(n,1)}) = \varphi(\omega_{(n,0)}). \tag{6}$$

Ceci $\forall\omega\in\Omega^n(p^0)$, donc $\forall\omega\in\Omega(p^0)$; en effet, l'égalité (6) est indépendante des n-ièmes coordonnées. Appliquant le même raisonnement $\forall n\in\mathrm{supp}(p^0)$, on obtient finalement

$$\forall(\omega,n)\in(\Omega(p^0)\times\mathrm{supp}(p^0))\quad \varphi(\omega_{(n,1)}) = \varphi(\omega_{(n,0)}) = E_{p^0}\varphi.$$

Or la famille des échanges d'une coordonnée du support de $\Gamma(p^0)$ est génératrice de $\mathcal{S}_{\Gamma(p^0)}$, groupe d'action transitive sur $\Omega(p^0)$ (lemme 2.4). Donc pour tout $\omega\in\Omega(p^0)$ on a $\varphi(\omega) = E_{p^0}\varphi$, puis $E_p\varphi = E_{p^0}\varphi$ pour tout $p\in\Gamma(p^0)$. Ce qui montre le lemme au rang d.

□

□

Remarque 2.14 (potentiel et fonctions harmoniques) Les potentiels $E_p\varphi$ vérifient donc un principe du minimum, comme les fonctions harmoniques de plusieurs variables réelles sur les ouverts réels. Ceci est dû à leur définition comme *moyennes* de φ (voir [7]). On note également que le potentiel est non convexe, dès que $\dim(E)\geqslant 2$, et ce "d'autant plus" que la dimension est élevée.

3 Dynamiques

On cherche maintenant à définir des trajectoires dans l'espace des phases, "attirées" par les configurations optimales pour φ.

3.1 Gradients

Proposition 3.1 (Formule de Russo) *Pour tout* $p \in (E, \mathcal{V}_E)$
i. *Le potentiel de* φ *admet un gradient au point* p *qui s'exprime localement sous la forme d'un vecteur de dimension* $dim(p)$,

$$\nabla E_p\varphi = (\nabla_n E_p\varphi)_{n\in supp(p)}.$$

ii. *Si* $n \in supp(p)$

$$\nabla_n E_p\varphi = \frac{\partial}{\partial p_n} E_p\varphi = E_{p_{(n,1)}}\varphi - E_{p_{(n,0)}}\varphi. \tag{7}$$

Sinon $\frac{\partial}{\partial p_n} E_p\varphi = 0$.

Ainsi on note également $\nabla E_p\varphi = (\nabla_n E_p\varphi)_{n\in\mathbb{N}}$, avec la convention $\nabla_n E_p\varphi = 0$ si $n \notin \mathrm{supp}(p)$.

Remarque 3.2 L' équation (7) est un avatar d'une formule de L. Russo (voir [15]).

Preuve
i. La fonction $E_p\varphi$, analytique sur la variété localement euclidienne $\mathcal{V}_E$, admet un gradient en tout point. L'espace tangent au point p étant $\mathbb{R}^d$ (son propre dual), où $d = \dim(p)$, le gradient au point p s'exprime sous la forme d'un vecteur de dimension $\dim(p)$.
ii. Soit $p^0 \in E$. D'après la proposition 2.10 appliquée à $A = \mathrm{supp}(p^0)$,

$$\forall q \in \mathrm{Int}(\Gamma(p^0)) \quad q \wedge \omega = \prod_{n\in \mathrm{supp}(p^0)} q_n^{\omega_n}(1-q_n)^{(1-\omega_n)},$$

où $\mathrm{supp}(p^0)$ est localement constant. Donc si $n \in \mathrm{supp}(p^0)$

$$\frac{\partial}{\partial p_n}|_{p^0} p \wedge \omega = p^0_{(n,1)} \wedge \omega - p^0_{(n,0)} \wedge \omega.$$

De même $\Omega(p^0)$ est localement constant, donc

$$\begin{aligned}
&\frac{\partial}{\partial p_n}|_{p^0} E_p\varphi \\
&= \sum_{\omega\in\Omega(p^0)} \varphi(\omega)\, p^0_{(n,1)} \wedge \omega - \sum_{\omega\in\Omega(p^0)} \varphi(\omega)\, p^0_{(n,0)} \wedge \omega \\
&= \sum_{\omega\in\Omega(p^0_{(n,1)})} \varphi(\omega)\, p^0_{(n,1)} \wedge \omega - \sum_{\omega\in\Omega(p^0_{(n,0)})} \varphi(\omega)\, p^0_{(n,0)} \wedge \omega,
\end{aligned}$$

soit

$$\frac{\partial}{\partial p_n}|_{p^0} E_p\varphi = E_{p^0_{(n,1)}}\varphi - E_{p^0_{(n,0)}}\varphi.$$

De plus si $n \notin \mathrm{supp}(p^0)$ $E_p\varphi$ est localement indépendant de p_n, d'après la proposition 2.10, donc $\frac{\partial}{\partial p_n}|_{p^0} E_p\varphi = 0$.

□

Proposition 3.3 (Minimum euclidien et minimum relatif) *Si ω est un minimum local d'$E_p\varphi$ sur le cube Γ, au sens de la topologie* euclidienne *sur Γ (voir définition 2.3), alors*

$$\forall n \in supp(\Omega) \quad \varphi(\omega) = \max\{\varphi(\omega_{(n,1)}), \varphi(\omega_{(n,0)})\}. \tag{8}$$

Preuve. Si $n \in \mathrm{supp}(\Omega)$, on montre comme ci-dessus, la topologie en vigueur étant euclidienne sur Γ

$$\frac{\partial}{\partial p_n} E_\omega\varphi = \varphi(\omega_{(n,1)}) - \varphi(\omega_{(n,0)}). \tag{9}$$

On applique alors la condition de Kuhn et Tucker en ω pour un minimum lié par des contraintes inégalités dans un espace euclidien, soit

$$\begin{cases} \omega_n = 0 & \wedge & \frac{\partial}{\partial p_n} E_\omega\varphi \geqslant 0 \\ & \vee & \\ \omega_n = 1 & \wedge & \frac{\partial}{\partial p_n} E_\omega\varphi \leqslant 0 \end{cases},$$

ou encore

$$\varphi(\omega) = \max\{\varphi(\omega_{(n,1)}), \varphi(\omega_{(n,0)})\}.$$

□

Remarque 3.4 (Minimum relatif et optimum de Pareto) On retrouve une notion habituelle de minimum relatif pour φ. Les minima relatifs de φ sur Ω apparaissent comme des optima de Pareto du jeu où les coordonnées coopèrent contre la $\mathcal{N}$ature pour minimiser $E_p\varphi$, sur Γ muni de sa topologie euclidienne (voir [11, ch.II]).

3.2 Descentes

Lemme 3.5 *(Métrique euclidienne) La distance euclidienne entre p et q dans $\Gamma(p,q)$ définit une métrique $\mathcal{M}_E$ sur E. On a donc un espace métrique $(E, \mathcal{M}_E)$, en plus de l'espace topologique (E, Θ_E). Si Λ est un cube de E, Λ est $\mathcal{M}_E$-compact, ainsi que sa surface $\partial\Lambda$.*

On vérifie aisément que la topologie induite sur un cube $\Lambda \subset E$ par $\mathcal{M}_E$ coïncide avec sa topologie euclidienne, telle qu'on l'a introduite en définition 2.3.

Preuve. Notons $d_\Gamma(.,.)$ la distance euclidienne usuelle sur un cube Γ. Soit $(p,q,r) \in E^3$. Si $n \notin \text{supp}(\Gamma(p,q))$, alors $p_n = q_n$, donc $(p_n - q_n)^2 = 0$. Donc $d(p,q)$ égale la distance euclidienne entre p et q dans $\Gamma(p,q,r)$. Donc

$$\begin{aligned} & d(p,q) + d(q,r) \\ = \;& d_{\Gamma(p,q)}(p,q) + d_{\Gamma(q,r)}(q,r) \\ = \;& d_{\Gamma(p,q,r)}(p,q) + d_{\Gamma(p,q,r)}(q,r) \\ \leqslant \;& d_{\Gamma(p,q,r)}(p,r) = d(p,r), \end{aligned}$$

ce qui donne l'inégalité triangulaire

$$d(p,q) + d(q,r) \leqslant d(p,r).$$

Par suite $d(\cdot,\cdot)$ définit bien une distance sur E. De plus sur un cube $\Lambda \subset E$, $d \equiv d_\Lambda$. Donc Λ et sa surface $\partial\Lambda$ sont $\mathcal{M}_E$-compacts.

□

Si $(E^1,\ E^2) \in E^2$, on définit $d(E^1,E^2) = \inf_{(p^1,\,p^2)\in(E^1\times E^2)} d(p^1,p2)$.

Lemme 3.6 *Si $p^0 \in E$, soit $p :]-S,T[\supset \emptyset \to E$ la solution maximale de l'équation différentielle*

$$\dot{p} = -\nabla E_p \varphi \tag{10}$$

vérifiant la condition initiale $p(0) = p^0$. p est à valeurs dans Int$(\Gamma(p^0))$. De plus si $T < +\infty$

$$\exists\, \mathcal{M}_E \lim_{t\to T^-} E_p\varphi \;=\; p^1 \in \sigma(\Gamma(p^0)).$$

On note (th) le théorème usuel d'existence et d'unicité d'une solution maximale à une équation différentielle ordinaire sur un ouvert de $\mathbb{R}^d$ contenant une condition initiale donnée (voir [24, ch.2,§7]).

Preuve

• À travers sa carte $Id_{\Gamma(p^0)} \in \mathcal{A}_E$, Int$(\Gamma(p^0))$ s'identifie à un ouvert $U^d =]0,1[^d$ de $\mathbb{R}^d$, où $d = \dim(p^0)$. D'après (th), il existe une et une seule solution p à l'équation différentielle (10), définie sur un intervalle de temps maximal $]-S,T[$ contenant 0, telle que $p(0) = p^0$. De plus

p étant dérivable prend nécessairement ses valeurs dans une seule composante connexe de (E, Θ_E). Donc p est à valeurs dans $\mathrm{Int}(\Gamma(p^0))$, composante connexe de sa condition initiale p^0 (proposition 2.7). Enfin comme l'application $E_p\varphi$ est analytique il en va de même de la fonction $p(t)$, d'après une extension bien connue de (th) (voir [24, p.283]).

• Supposons $T < +\infty$. Montrons $\exists\ \mathcal{M}_E \lim_{t\to T^-} E_p\varphi = p^1$. Pour montrer l'existence de la $\mathcal{M}_E$-limite dans le $\mathcal{M}_E$-compact $\Gamma(p^0)$, il suffit de montrer que p présente une unique $\mathcal{M}_E$-valeur d'adhérence en T^-. Supposons

$$\left\{ \begin{array}{lcl} \exists\ \mathcal{M}_E \lim_{s_n\to T^-} p(s_n) & = & p^1 \\ \exists\ \mathcal{M}_E \lim_{t_n\to T^-} p(t_n) & = & p^2 \end{array} \right. ,$$

avec $p^1,\ p^2 \in \Gamma(p^0)$, et $d(p^1, p^2) = \delta > 0$. Soit $\mathcal{O}^1$ (resp. $\mathcal{O}^2$) la $\mathcal{M}_E$-boule fermée de centre p^1 (resp. p^2) , de rayon $d(p^1, p^2)/4$. A partir de s_n et t_n, on forme une nouvelle suite $u_n \to T^-$ telle que $u_{2n} \in \mathcal{O}^1$ et $u_{2n+1} \in \mathcal{O}^2$. On peut construire (u_n), de proche en proche, sinon il existe $N \in \mathbb{N}$ tel que $\forall n > N\ t_n \notin \mathcal{O}^2$, ce qui contredit l'hypothèse $p^2 = \mathcal{M}_E \lim_{t_n\to T^-}\ p(t_n)$; ou il existe $N \in \mathbb{N}$ tel que $\forall n > N\ s_n \notin \mathcal{O}^1$, ce qui contredit l'hypothèse faite sur p^1. D'après la formule (7) et la proposition 2.10, l'application $p \to \nabla E_p\varphi$ s'identifie sur $\Gamma(p^0)$ à la différence de deux polynômes en les coordonnées du support de p^0, donc elle est $\mathcal{M}_E$-continue sur le $\mathcal{M}_E$-compact $\Gamma(p^0)$. Par suite $\nabla E_p\varphi$ est borné sur $\Gamma(p^0)$, et *a fortiori* sur $\mathrm{Int}(\Gamma(p^0)) \subset \Gamma(p^0)$. Soit K un majorant en norme. Alors

$$\forall n \in \mathbb{N}\ \ KT \geqslant \int_{u_0}^{u_n} \parallel \dot{p(t)} \parallel dt \geqslant n\delta/2,$$

ce qui contredit notre hypothèse $\delta > 0$. Donc $d(p_1, p_2) = 0$, soit $p^1 = p^2$.

• Montrons $p^1 \in \sigma(\Gamma(p^0))$. Supposons en effet $p^1 \in \mathrm{Int}(\Gamma(p^0))$. Alors (th;existence) fournit à travers l'atlas $\mathcal{A}_E$ une solution analytique à l'équation (10) au voisinage de p^1, à valeurs dans $\mathrm{Int}(\Gamma(p^0))$, notée q. En prolongeant p par q à droite de T (inclus), on obtient une application r définie sur un intervalle $]-S, T+\epsilon[\supset]-S, T[$. Montrons que r, qui prend ses valeurs dans $\mathrm{Int}(\Gamma(p^0))$, est solution de l'équation (10) sur son domaine de définition. En effet r est solution de (10) sur $]-S, T[$ et sur $]T, T+\epsilon[$. En T, la limite à droite de la dérivée existe, et vaut $-\nabla E_{p^1}\,\varphi = -\nabla E_{r(T)}\,\varphi$. De plus,

Lemme 3.7 *Si u_n prend ses valeurs dans $Int(\Gamma(p^0))$, et $u \in Int(\Gamma(p^0))$, alors*

$$u = \mathcal{M}_E \lim_{\infty} u_n \Rightarrow u = \lim_{\infty} u_n.$$

Preuve. En effet soit $\mathcal{O}$ un ouvert contenant u. D'après la proposition 2.7, on peut supposer $\mathcal{O} \in \mathcal{B}_{\Gamma(p^0)}$. Par ailleurs, d'après le lemme 3.5, on a $d(u_n, u) = d_{\Gamma(p^0)}(u_n, u)$, puisque $u_n,\ u\ \in$

$\Gamma(p^0)$. Donc si $d(u_n, u) = d_{\Gamma(p^0)}(u_n, u)$ tend vers 0, on a nécessairement $u_n \in \mathcal{O}$, pour n assez grand.

□

Montrons qu'alors $\dot{p}$ a une limite à gauche en T, égale à $-\nabla E_{p^1}\,\varphi$. En effet si $p^1 \in \mathrm{Int}(\Gamma(p^0))$, alors la $\mathcal{M}_E$-limite p^1 est une limite, d'après le lemme, d'où $\lim_{t\to T}\dot{p}(t) = -\nabla E_{p^1}\,\varphi$, par continuité de $\nabla E_p\,\varphi$ en p^1. Donc la dérivée de r en T existe et vaut $-\nabla E_{r(T)}\,\varphi$, d'après le théorème du prolongement de la dérivée. Donc r est une solution de l'équation (10) qui prolonge strictement p, contrairement à l'hypothèse de maximalité de p.

□

Comme $\sigma(\Gamma(p^0))$ est l'union des faces de $\Gamma(p^0)$, il existe un unique $\Gamma^1 \in \mathcal{F}(\Gamma(p^0))$ tel que $p^1 \in \mathrm{Int}(\Gamma^1)$. De plus, si $\dim(p^0) > 0$, alors $\dim(p^1) < \dim(p^0)$ (lemme 2.6).

Définition 3.8 (Descentes) *Étant donnée la condition initiale $p^0 \in E$, on définit la descente des potentiels de φ à travers E, depuis le point p^0, par récurrence sur $dim(p^0)$. Si $dim(p^0) = 0$, p est par définition la solution constante sur $\mathbb{R}^+$ à l'équation (10), telle que $\forall t \geqslant 0 \;\; p(t) = p^0$. Les descentes étant supposées définies si $0 \leqslant dim(p^0) \leqslant d$, on définit la descente depuis p^0 tel que $dim(p^0) = d + 1$ ainsi:*
• *Sur $[0, T[$, p est la restriction à $\mathbb{R}^+$ de la solution maximale de (10) telle que $p(0) = p^0$.*
• *Si $T < +\infty$, alors p se prolonge à partir de T sur $\mathbb{R}^+$ tout entier par la descente depuis p^1, où p^1, défini à partir de p^0 dans le lemme 3.6, vérifie $0 \leqslant dim(p^1) < dim(p^0)$.*

Ainsi on définit bien une application $p(t)$ sur $\mathbb{R}^+$, quel que soit $p^0 \in E$.

Remarque 3.9 (Jeux différentiels) En théorie des jeux, il est classique d'étudier la dynamique d'un jeu avec coalition à l'aide du système différentiel régissant le jeu flou (mixte) associé (voir [13, p.177]).

Soit $p(t)$ une descente de potentiels sur $\mathbb{R}^+$, $\mathcal{T}$ la trajectoire associée, image de p.

Proposition 3.10 *On a*
i. *Les cubes sont absorbants pour $\mathcal{T}$. La séquence des cubes engendrés est décroissante, stationnaire. De plus la descente depuis p^0 est indépendante de E contenant p^0.*
ii. *sur $\mathbb{R}^+$, $p(t)$ est analytique par morceaux, qu'on peut choisir en nombre fini; $p(t)$ est $\mathcal{M}_E$-continue.*
iii. *$E_{p(t)}\,\varphi$ est décroissante sur $\mathbb{R}^+$. En particulier $E_{p(t)}\,\varphi$ est constante entre deux instants où sa trajectoire se coupe, ou sur un cycle.*

Preuve

i. En effet $p^1 \in \sigma(\Gamma(p^0))$, donc $\Gamma(p^1) \subset \sigma(\Gamma(p^0)) \subset \Gamma(p^0)$. Alors par récurrence sur $\dim(p^0)$ la séquence des cubes engendrés est décroissante. Elle est donc stationnaire, d'après le lemme 2.6. Donc $\mathcal{T} \subset \Gamma(p^0)$, indépendamment de E contenant p^0.

ii. Reprenant les notations de la preuve du lemme 3.6, $p(t)$ est analytique sur $]0,T[$. La démonstration suit alors par récurrence sur $\dim(p^0)$.

iii. Sur un morceau $]m,M[\neq\emptyset$ où $p(t)$ est analytique, on a $\dot{p} = -\nabla E_p\varphi$, donc

$$\frac{\partial}{\partial t} E_{p(t)}\varphi = \langle \dot{p}\ ,\ \nabla E_{p(t)}\varphi \rangle_p = -\parallel \nabla E_{p(t)}\varphi \parallel_p^2 \leqslant 0 \tag{11}$$

De plus, par $\mathcal{M}_E$-continuité de $p(t)$, $E_{p(t)}\,\varphi$ est continue sur $\mathbb{R}^+$. Donc $E_{p(t)}\,\varphi$ est décroissante sur $\mathbb{R}^+$. En particulier si $E_{p(t_1)}\,\varphi = E_{p(t_2)}\,\varphi$ alors $E_{p(t)}\,\varphi$ est constante sur $[t_1,t_2]$.

□

Remarque 3.11 (Remontées des potentiels) On a des énoncés analogues pour les remontées des potentiels de φ, lorsqu'on remplace $-$ par $+$ dans l'équation (10). $E_{p(t)}\,\varphi$ est alors croissante sur $\mathbb{R}^+$.

3.3 Cas où $\dim(p^0) = 2$

On suppose que p^0 est le milieu d'un carré. $\mathcal{T}$ étant indépendante de E contenant p^0, on peut raisonner sur $\widehat{\Gamma(p^0)}$, obtenant la même descente depuis p^0 à travers tout sur-ensemble E. Ajoutant au besoin une constante, sans altération de la dynamique, on se ramène par symétrie au cas où

$$\left\{ \begin{array}{l} \varphi \geqslant 0 \\ \varphi(\mathbf{0}) = 0 \end{array} \right. . \tag{12}$$

Sur $]0,1[^2$, on a alors

$$\begin{aligned} & E_p\varphi \\ &= \varphi(0,1)(1-p_1)p_2 + \varphi(1,0)p_1(1-p_2) + \varphi(1,1)p_1p_2 \\ &= (-\varphi(0,1) - \varphi(1,0) + \varphi(1,1))p_1p_2 + \varphi(1,0)p_1 + \varphi(0,1)p_2, \end{aligned}$$

donc

$$\left\{ \begin{array}{rcl} \dot{p}_1 &=& -(-\varphi(0,1) - \varphi(1,0) + \varphi(1,1))p_2 - \varphi(1,0) \\ \dot{p}_2 &=& -(-\varphi(0,1) - \varphi(1,0) + \varphi(1,1))p_1 - \varphi(0,1) \end{array} \right. . \tag{13}$$

Sur $[0,1]\times\{0\}$, on a

$$E_p\varphi = p_1\,\varphi(1,0),$$

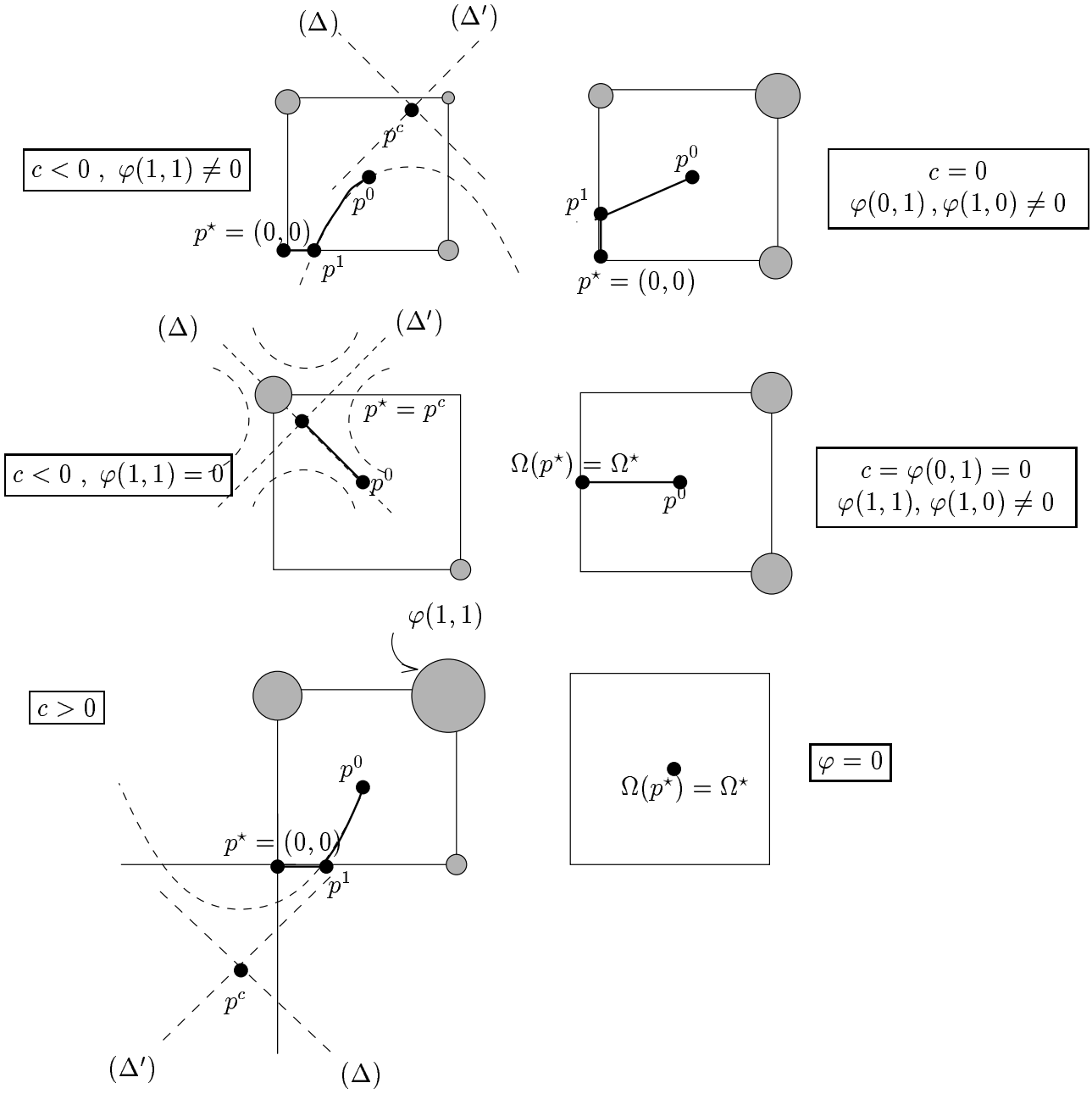


FIG. 2 - *Descentes des potentiels de φ depuis le milieu d'un carré.*

donc

$$\begin{cases} \dot{p}_1 &= -\varphi(1,0) \\ \dot{p}_2 &= 0 \end{cases} . \tag{14}$$

Et ainsi de suite, par symétrie. Soit $c = \varphi(0,1) + \varphi(1,0) - \varphi(1,1)$. Si $c \neq 0$, le système linéaire (13) a un unique point critique sur le plan,

$$p^c = (\varphi(0,1)/c, \varphi(1,0)/c).$$

Les courbes intégrales de (13) forment un champ d'hyperboles sur le plan, orthogonal au champ d'hyperboles des équipotentielles (du prolongement au plan) d' $E_p\varphi$, centré en p^c, admettant pour droites asymptotes les parallèles aux bissectrices des axes de coordonnées passant par p^c (voir [12, ch.XIII] ou [25]). Qualitativement tous les cas possibles sont répertoriés figure 2, page 22. Sur la figure, les superficies des disques gris aux sommets des carrés sont proportionnelles aux valeurs de φ. Le détail est laissé au lecteur.

Commentaire 3.12 (Synthèse de l'étude de cas figure 2, page 22) Soit $\Omega^\star = \{\omega \in \Omega(p^0)\ ;\ \varphi(\omega) = \min_{\Omega^0} \varphi\}$. On fait les constatations suivantes.

- Un point stationnaire $p^\star$ est toujours rejoint en un temps fini.
- On a $\Omega(p^\star) \subset \Omega^\star$, sauf si $\varphi(1,1) = 0,\ \varphi(0,1)$ et $\varphi(1,0) \neq 0$, auquel cas les seuls équilibres stables demeurent les minima $(0,0)$ et $(1,1)$, la trajectoire issue du milieu du carrée étant piégée par le point-selle p^c.
- Dans le cas "générique" où φ est injective, on a d'abord sur $]0,1[^2$, action d'un groupe à un paramètre de diffémorphismes, composée d'une contraction exponentielle suivant (Δ), et d'une dilatation exponentielle suivant (Δ'); puis le long du premier bord atteint, convergence en ligne droite vers le minimum $(0,0)$. La droite (Δ) sépare le carré en deux parties (l'une étant éventuellement vide). La plus grande, celle qui contient le milieu du carré, est le bassin d'attraction du minimum; l'autre, celui du sommet diagonalement opposé dans $\widehat{\Gamma}$.

Ainsi le milieu ρ du cube apparaît comme un point de départ privilégié *a priori* pour la recherche du minimum. C'est en effet le point de départ le plus raisonnable pour une recherche en l'absence d'information extérieure sur φ, toutes les configurations de Ω étant équiprobables sous $\wedge_\rho$.

Remarque 3.13 (Dimensions supérieures d'espace) Si la dimension de $\Gamma(p^0)$ est > 2, la non-linéarité du système différentiel (10) initialisé en p^0 empêche une analyse exhaustive

des différents cas. Cependant on peut construire des contre-exemples où la remontée des potentiels de φ converge vers une configuration $\omega \notin \Omega^\star$.

3.4 Cas des masses de Dirac

Étudions la remontée des potentiels de φ si $\varphi = \delta_\omega$, où ω appartient à un hypercube Ω^0.

Lemme 3.14 *On a*
i. *Si* $(p, g, \omega) \in (E \times \mathcal{S}_E \times \mathcal{D})$, *alors* $p \wedge g\omega = gp \wedge \omega$.
ii. *Les potentiels des masses de Dirac* $(\delta_\omega)_{\omega \in \mathcal{D}}$ *se déduisent les uns des autres par action du groupe* $\mathcal{S}_D$, $E_p \delta_{g\omega} = E_{gp} \delta_\omega$.

Preuve
i. Si $(p, \omega) \in [0,1] \times \{0,1\}$, on vérifie $p^{g\omega}(1-p)^{(1-g\omega)} = (gp)^\omega (1-gp)^\omega$. Pour $p \in E$, on en déduit

$$\begin{aligned} & p \wedge g\omega \\ &= \prod_{n \in \mathbb{N}} p_n^{g\omega_n} (1-p_n)^{(1-g\omega_n)} \\ &= \prod_{n \in \mathbb{N}} (gp_n)^{\omega_n} (1-gp_n)^{(1-\omega_n)}, \end{aligned}$$

soit

$$p \wedge g\omega = gp \wedge \omega.$$

ii. Par définition,

$$E_p \delta_{g\omega} = \begin{cases} p \wedge g\omega & \text{si} \quad g\omega \in \Gamma(p) \\ 0 & \text{sinon} \end{cases}.$$

Comme $\mathcal{S}_E$ est involutif, on a $g\omega \in \Gamma(p) \Leftrightarrow \omega \in g^{-1}\Gamma(p) \Leftrightarrow \omega \in \Gamma(gp)$. D'après **i**, on a donc

$$E_p \delta_{g\omega} = \begin{cases} gp \wedge \omega & \text{si} \quad \omega \in \Gamma(gp) \\ 0 & \text{sinon} \end{cases},$$

soit

$$E_p \delta_{g\omega} = E_{gp} \delta_\omega.$$

□

Proposition 3.15 *Soit* $\omega \in \Gamma^0 \in \mathcal{G}$.
i. *Si* p^0 *est le milieu de* Γ^0, $\mathcal{T}$ *est le segment* $[p^0, \omega]$.

ii. *Si $p \in Int(\Gamma^0)$, soit $\mathcal{F}^0$ l'ensemble des faces de Γ^0 contenant ω, les plus proches de p^0*

$$\mathcal{F}^0 = \{f \in \mathcal{F}(\Gamma^0) \ ; \ \omega \in f \ \wedge \ d(p^0, f) = \min_{g \in \mathcal{F}(\Gamma^0)} d(p^0, g)\}.$$

Alors $T < +\infty$, et $\Gamma(p^1)$ est l'intersection des faces appartenant à $\mathcal{F}^0$: $\Gamma(p^1) = \bigcap_{f \in \mathcal{F}^0} f$. Et ainsi de suite, récursivement. ω est atteint en un temps fini. L'orbite "générique" en $p \in Int(\Gamma^0)$ a donc une structure récursive à l'intérieur d'une chaîne maximale de cubes entre ω et Ω^0 (Voir figure 3).

Preuve. Grâce au lemme 3.14, on se ramène au cas où $\omega_n = 1$ si $n \in \mathrm{supp}(\Gamma^0)$. Donc tant que $p \in \mathrm{Int}(\Gamma^0)$,

$$E_p\delta_\omega = \prod_{n \in \mathrm{supp}(\Gamma^0)} p_n \quad , \quad \frac{\partial}{\partial p_n} E_p\delta_\omega = \begin{cases} \prod_{m \neq n} p_m & \text{si} \quad n \in \mathrm{supp}(\Gamma^0) \\ 0 & \text{sinon} \end{cases} \tag{15}$$

i. On cherche une solution maximale sur $\mathrm{Int}(\Gamma^0)$ à $\dot{p} = \nabla E_p\delta_\omega$, sous la forme $p(t) = f(t)\,\omega$, f réelle. La condition de remontée des potentiels s' écrit $\dot{f} = f^{n-1}$, soit si $n > 2$:

$$1/\dot{\widehat{f^{n-2}}} = -(n-2),$$

qui s'intègre à l'aide de la condition initiale $f(0) = 1/2$ en $f(t) = 1/\sqrt{K - (n-2)t}^{n-2}$, où $K = 2^{n-2}$. Le sommet ω est atteint à l'instant $t_1 = (K-1)/(n-2) > 0$. Donc la solution $p(t)$ est

$$p(t) = \begin{cases} 1/\sqrt{K - (n-2)t}^{n-2}\,\omega & \text{si} \quad 0 \leqslant t_1 \\ \omega & \text{si} \quad t \geqslant t_1 \end{cases} .$$

Si $\dim(p^0) = 2$ la remontée vers ω est exponentielle (section 3.3). Si $\dim(p^0) = 2$ elle est linéaire.

ii. On raisonne par récurrence sur $d = \dim(\Gamma^0)$. C'est trivialement vrai si $d = 0$ où $d = 1$. Supposons le résultat vrai pour $\dim(\Gamma^0) \leqslant d - 1 \in \mathbb{N}$ et donnons-nous Γ^0 contenant ω tel que $\dim(\Gamma^0) = d \geqslant 2$. Soit $\Gamma'^1 = \bigcap_{f \in \mathcal{F}^0} f \quad \in \mathcal{G}$. Par hypothèse de récurrence, il suffit de montrer $T < +\infty$ et $p^1 \in \mathrm{Int}(\Gamma'^1)$, ce qui prouvera $\Gamma(p^1) = \Gamma'^1$. Soit $c = \min_{n \in \mathrm{supp}(\Gamma^0)} p^0_n$. Comme $d > 0$, et $p^0 \in \mathrm{Int}(\Gamma^0)$, c est donc > 0. Soit $\mathcal{C} = \{p \in \Gamma^0 \ ; \ p_n \geqslant p^0_n \text{ si } n \in \mathrm{supp}(\Gamma^0)\}$. En tout point $p \in \mathcal{C}$, les $\nabla_n E_p\delta_\omega$ pour $n \in \mathrm{supp}(\Gamma^0)$ sont > 0, inférieurement bornées par $c^{d-1} > 0$ (équation 15). Comme de plus $p^0 \in \mathcal{C}$, une face contenant ω est atteinte au bout d'un temps fini T, et $p([0,T[) \subset \mathcal{C}$. Soit $\mathcal{H}^{i,j}$ l' hyperplan $p_i = p_j$, où $i, j \in \mathrm{supp}(\Gamma^0)$. En

tout point $p \in \mathcal{H}^{i,j}$, on a $\nabla E_p \delta_\omega \in \mathcal{H}^{i,j}$, donc si $p^0 \in \mathcal{H}^{i,j}$, $p([0,T[) \subset \mathcal{H}^{i,j}$. Enfin sur $\mathbb{R}^+$ on a

$$(p_i \dot{-} p_j)^2 = 2\,(p_i - p_j)\,(p_i \dot{-} p_j) \geqslant 0,$$

d'après l'équation (15). Donc $(p_i - p_j)^2 \geqslant (p_i^0 - p_j^0)^2$, constante strictement positive si $p^0 \notin \mathcal{H}^{i,j}$. La seule solution est donc que $\mathcal{T}$ quitte $\mathrm{Int}(\Gamma^0)$ en un point $p^1 \in \mathrm{Int}(\Gamma'^1)$.

□

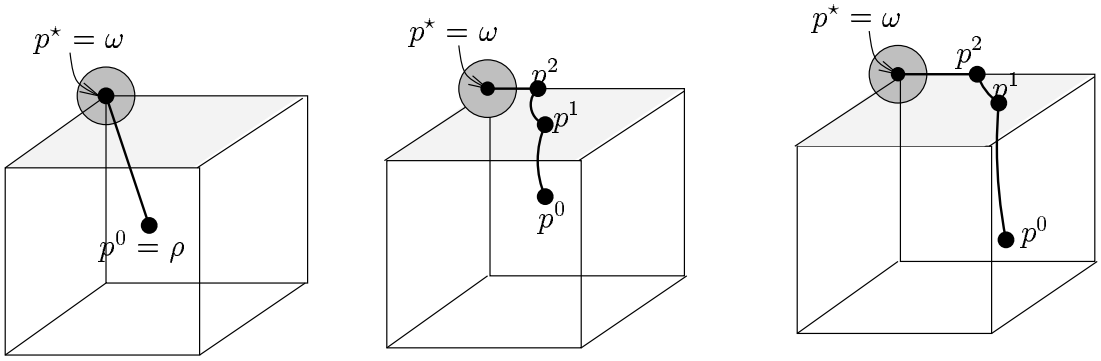


FIG. 3 - *Remontée des potentiels de δ_ω en dimension 3.*

Commentaire 3.16 (Descentes des potentiels et géodésiques) Au regard de la topologie euclidienne et de la distance sur E, les trajectoires génériques sont donc "éloignées" des géodésiques de E (droites euclidiennes) joignant le point initial p^0 au point stationnaire ω. En effet la direction de la trajectoire $\mathcal{T}$ est discontinue à chaque fois que la trajectoire rejoint la surface d'un cube. En théorie de l'inférence paramétrique, Amari (voir [20]) définit des géométries différentielles riemanniennes sur des espaces de distributions de probabilités. Il obtient des algorithmes de remontée des géodésiques.

Remarque 3.17 (Minimisation et maximisation) Concernant les descentes de potentiels de δ_ω, on a les énoncés analogues, relativement au sommet diagonalement opposé à ω dans $\widehat{\Gamma^0}$. La masse ω "attire" les trajectoires de remontée des potentiels de δ_ω, et "repousse" ses trajectoires de descentes.

La proposition 3.15 se généralise de la façon suivante. Soit Ω un hypercube, Γ son enveloppe convexe, $\epsilon = \min_{\Omega - \Omega^\star} \varphi$, où $\Omega^\star$ est le lieu des minima de φ restreinte à Ω, ou $\epsilon = 1$, si φ est constante sur Ω. On définit $\phi : \mathcal{B} \to \mathbb{R}$ par $\phi(\omega) = \varphi(\omega) + \varepsilon \sum_{n \in \mathbb{N}} \omega_n / 2^{n+1}$, et $\psi_A : \mathcal{B} \to \mathbb{R}$ par $\psi_A(\omega) = exp(-A\{\varphi(\omega) - \varphi(\omega^\star)\})$, où $A > 0$. $E_p \psi_A$ est donc, à un facteur de normalisation près, la transformée de Laplace de la variable aléatoire $\phi(\omega)$ au point A, sous la pondération bernouillienne définie par p.

Proposition 3.18 *La fonction $\phi : \mathcal{B} \to \mathbb{R}$ présente un unique minimum $\omega^\star$ sur Ω. On a $\omega^\star \in \Omega^\star$. La trajectoire de remontée des potentiels de ψ_A depuis le milieu de Γ atteint $\omega^\star$ en un temps fini, pour A assez grand.*

Preuve. Par construction la fonction ϕ présente un unique minimum sur Ω, atteint au point $\omega^\star$. De même les fonctions ψ_A présentent un unique maximum sur Ω, atteint au point $\omega^\star$, de valeur nulle. De plus si $p \in \mathrm{Int}(\Gamma)$,

$$\begin{aligned} & \nabla E_p \, exp(-A\{\varphi(\omega) - \varphi(\omega^\star)\}) \\ &= \sum_{\omega \in \Omega} exp(-A\{\varphi(\omega) - \varphi(\omega^\star)\}) \, \nabla E_p \delta_\omega \\ &= \nabla E_p \delta_{\omega^\star} + \sum_{\Omega - \{\omega^\star\}} exp(-A\{\varphi(\omega) - \varphi(\omega^\star)\}) \, \nabla E_p \delta_\omega . \end{aligned}$$

Donc sur $\mathrm{Int}(\Gamma)$, l'équation différentielle régissant la remontée des potentiels de ψ_A est analogue à celle d'une masse en $\omega^\star$, à une perturbation régulière additive près. Pour A assez grand cette perturbation est uniformément bornée sur Γ par toute constante > 0 préalablement fixée. La preuve de la proposition précédente s'adapte alors sans difficulté.

□

Commentaire 3.19 (Séparation des configurations et NP-complétude) La proposition précédente sert à mettre en valeur la relation qui peut exister entre la bonne séparation des valeurs et la NP-complétude de certains problèmes. Ainsi le problème du sac à dos admet un algorithme de résolution polynômial sur entrées entières bornées. Sur entrées rationnelles, ou entières non bornées, la complexité devient exponentielle. En effet le problème devient de moins en moins bien séparé: le rapport d'un entier à son successeur est uniformément $\leqslant n/n+1 < 1$ sur $\mathbb{N}_n$, tandis que sur $\mathbb{N}$ il peut être rendu aussi proche de 1 que l'on veut, de même que peut l' être le rapport entre deux rationnels, même bornés. En termes anthropomorphes, lorsqu'après élimination des configurations les plus défavorables, la comparaison

des configurations restantes inclinerait à hésiter indéfiniment, faute de parvenir à déterminer une préférence, il faut trouver un moyen de trancher, quitte à ce que cette décision comporte une part d'arbitraire. On peut par exemple déformer temporairement la relation de préférence entre les configurations restantes, privilégiant davantage tel ou tel aspect. L'utilité correspondant à la décision est recalculée à l'aide de la fonction initiale. Autrement dit l'utilité associée à une décision doit résulter d' un arbitrage entre les conséquences de cette décision et le temps mis pour l'obtenir. De même on fait intervenir la complexité dans le bilan des joueurs en théorie des jeux itérés (voir [6]).

3.5 Heuristiques: superposition, symétrie

On formule maintenant deux thèses sur lesquelles on s'appuiera plus tard pour former nos algorithmes de minimisation approchée (voir section 4). Nos thèses se fondent sur les identités élémentaires suivantes.

Proposition 3.20 *Si* $\Omega \supseteq \Omega(p)$ *et* $p \in E$

$$E_p\varphi = \sum_{\omega\in\Omega} \varphi(\omega)\, E_p\delta_\omega \tag{16}$$

$$\nabla E_p\varphi = \sum_{\omega\in\Omega} \varphi(\omega)\, \nabla E_p\delta_\omega. \tag{17}$$

Preuve. En effet si $\omega \notin \Omega(p)$ alors $E_p\delta_\omega$ est nul, d'après la proposition 2.10. (17) se déduit alors de (16) par interversion d'un gradient et d'une somme (localement) finie.

□

Nos thèses s' énoncent ainsi (figure 4).

Superposition En un point $p \in E$, l'influence d'un sommet ω est proportionnelle à $\varphi(\omega)$, et inversement proportionnelle à $d(p,\omega)$.

Symétrie Au milieu ρ d'un cube, l'influence d'un sommet ω est proportionnelle à $\varphi(\omega)$.

4 Algorithmes

4.1 Cas général

4.1.1 Difficulté d'estimer les gradients

La difficulté pour les applications est l'estimation des gradients $\nabla E_p\varphi$. En effet ces calculs de gradients se ramènent à des calcul d'espérances, d'après la formule de Russo (équation

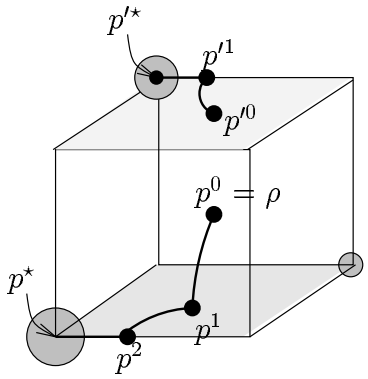


FIG. 4 - *Heuristiques de superposition et de symétrie.*

7). On pourrait donc espérer procéder par simulation. Mais l'augmentation de la dimension quand n croît rend cette approche très irréaliste: on retrouve en effet les problèmes numériques posés par le calcul approché d'intégrales multiples sur des espaces de dimensions trop élevées. Par exemple, dans le cas d'une masse Dirac δ_ω, l'estimation du gradient par la formule de Russo est aussi difficile que la détermination de la configuration optimale ω. De même en l'absence d'information extérieure aucun algorithme ne peut savoir si φ est nulle, avant d'en avoir calculé toutes les valeurs. Plus généralement, il est connu qu'une procédure d'amélioration locale est vouée à des performances médiocres lorsque l'espace de recherche se compose de quelques pics isolés dans de vastes plateaux, ou *Mesa* (voir [17]).

En revanche, si on dispose d'un bon estimateur pour certains gradients, on peut utiliser cette information dans des méthodes directes efficaces. Ainsi dans le cas de δ_ω le gradient au milieu ρ d'un cube Γ contenant ω pointe directement dans la boite de Dirichlet associée à ω. En effet pour $n \in \mathrm{supp}(\rho)$ on a d'après l'équation (7)

$$\frac{\partial}{\partial p_n} E_\rho \varphi = E_{\rho_{(n,1)}} \varphi - E_{\rho_{(n,1)}} \varphi,$$

qui est > 0 si $\omega \in \{p_n = 1\}$, et < 0 si $\omega \in \{p_n = 0\}$. Donc le sommet de la boite de Dirichlet dans laquelle pointe le gradient est à l'intersection des faces contenant ω, c'est donc ω.

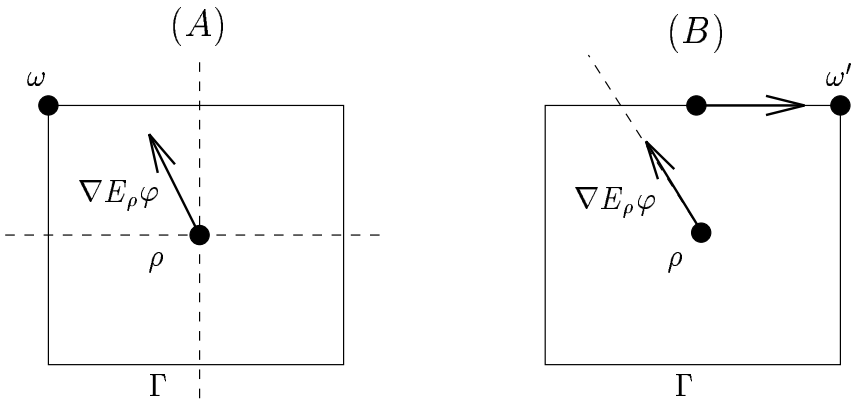


FIG. 5 - *Méthodes directes.*

4.1.2 Oracle

Admettons donc pour l'instant qu'on dispose d'un Oracle, ou procédure de calcul du gradient en temps constant, composante par composante. La complexité en temps de cette procédure est l'unité de mesure pour les algorithmes évoqués dans la suite. Voici quelques idées d'utilisation des gradients dans des algorithmes approchés de minimisation d'une fonction φ sur un hypercube Ω, de milieu ρ.

- Ayant estimé le gradient en ρ, on détermine la (boite de Dirichlet de la) configuration ω vers laquelle pointe ce gradient, en temps linéaire. $\varphi(\omega)$ fournit alors une valeur approchée du minimum de φ sur Ω (figure 5.A).

- Raffinant la procédure précédente, on détermine d'abord la face vers laquelle pointe $\nabla E_\rho\varphi$ [4]. Puis on recommence avec cette face au lieu du cube initial, et ainsi de suite, jusqu'à obtenir une configuration ω' en temps quadratique. $\varphi(\omega')$ est une autre valeur approchée du minimum de φ sur Ω (figure 5.B).

- Pour le choix de la face, un meilleur critère peut être le minimum sur cette face, estimé par la procédure du premier point, à complexité cubique; ou par celle du deuxième point, à complexité quadrique; ou encore par une procédure récursive, qui nous ramène à une énumération. En effet le gradient d' $E_p\varphi$ au milieu d'un segment n' est autre

4. On dit que le gradient en ρ pointe vers Γ' si $\Gamma' = \Gamma(p)$, où $p \stackrel{\Delta}{=} \sigma(\Gamma) \cap (\rho + \mathbb{R}^+ \nabla E_\rho\varphi)$.

que la différence des valeurs de φ entre les extrémités du segment, d'après la formule de Russo (7).

Ainsi on peut imaginer une grande variété d'algorithmes de minimisation de φ si on dispose de gradients.

Remarque 4.1 (Programmation par contraintes) Dans la procédure du deuxième point, récursive suivant les faces, (la plus proche des algorithmes des sections 4.2 et 4.3), on prospecte, en termes de programmation par contraintes, *un pas en avant*. La rentabilité à long terme de passer du temps à prospecter en amont avant d'assigner est un fait bien connu en intelligence artificielle (voir [9]).

4.1.3 Randomisation; parallélisme; commande

La *randomisation* permet l' exploration "au voisinage" d'une des stratégies d'exploration précédentes. Par exemple on peut choisir une face avec une probabilité proportionnelle à la composante du gradient dans sa direction (nulle si le gradient pointe vers la face opposée), au lieu de choisir systématiquement la face vers laquelle pointe le gradient. On obtient des promenades aléatoires sur le cube, dirigées autour de certains gradients. On introduit alors un paramètre de contrôle c, égal au nombre de fois qu'on applique l'algorithme randomisé depuis le milieu du cube. La question devient celle du choix de c (voir [26]). Toutefois il peut exister des situations où ce genre de sous-optimisation ne suffit pas à garantir l'obtention d'un minimum absolu. De plus lors de la première trajectoire de descente il peut être préférable d'appliquer la version dérandomisée de l'algorithme, surtout si $c \ll \dim(\Gamma)$. Ces promenades aléatoires sont qualitativement différentes des processus de recuit simulé. Ainsi ne sont-elles pas transitives. Fondées sur des heuristiques anthropomorphes (voir [2]), elles sont rapides, mais pas infaillibles. A l'opposé, le recuit simulé, fondé sur des principes naturels, atteint lentement mais sûrement ses fins. La valeur renvoyée par nos heuristiques peut aussi servir de bon point de départ pour un recuit simulé ou une recherche tabou.

On peut également voir de tels algorithmes sous l'angle *contrôle avec cible*. On cherche à contrôler une trajectoire issue de ρ afin d'atteindre la cible $\Omega^\star$. Au vu de la formule de Russo, on est tenté d'évoquer la théorie du *feedback*, qui suggère d'appliquer à chaque étape l'entrée qui serait optimale si toutes les quantités inconnues étaient fixées à leur espérance, afin de compenser les incertitudes et le bruit dans le système (voir [4, p. 193]). On retrouve

les liens existant entre stratégie bayésienne et minimax (voir [1, ch.5]).

En dimension d, un gradient a $\Theta(d)$ composantes. Il faut donc $\Theta(d)$ calculs pour le déterminer. Ces calculs sont *a priori* indépendants, donc directement parallélisables: on répartit les coodonnées équitablement entre les processeurs. Pour l' imagerie, ou l'utilisation d'ordinateurs massivement parallèles est pratique courante, la complexité en temps peut donc être divisée par q, où q est le nombre de processeurs.

En pratique l'Oracle s'obtient en se restreignant à des classes bien choisies de fonctions φ, utilisant l' information supplémentaire, de nature algébrique: φ quadratique, φ linéaire sous contrainte linéaire; ou probabiliste: interprétation fréquentiste de la formule de Russo, pour obtenir, dans les bons cas, des estimateurs raisonnables pour certains gradients. Les sections restantes montrent comment on peut estimer des gradients, et en déduire des algorithmes efficaces pour certains problèmes.

4.2 Problème du sac à dos

Pour le problème du sac à dos, l'application du théorème de Lindeberg pour les sommes de variables indépendantes (voir [27, p.262]) rend possible une estimation des gradients aux milieux des cubes.

4.2.1 Définition du problème et applications

Étant données les suites réelles $u = (u_n)_{n\in\mathbb{N}}$ et $v = (v_n)_{n\in\mathbb{N}}$, la contrainte $V^0 \in \mathbb{R}$, et l' hypercube Ω^0, le problème du sac à dos simple en variables binaires est

$$\left\{ \begin{array}{rcl} \max & & u^T\omega, \\ v^T\omega & \leqslant & V^0 \\ \omega & \in & \Omega^0 \end{array} \right. \tag{18}$$

$u^T\omega$ désignant le "produit scalaire" de la suite u par la suite presque nulle ω. Ce problème admet la version imagée bien connue du remplissage optimal d'un sac à dos sous contrainte de capacité, en présence d'objets de tailles et d'utilités marginales diverses. Quitte à changer x_i en $1 - x_i$, on peut supposer, à des assignations triviales près, les $u,\ v,\ V^0\ > 0$. On est donc bien dans le cadre défini en section 2, avec $\varphi(\omega) = u^T\omega\, 1_{(v^T\omega\leqslant V^0)}$. La contrainte a pour effet de filtrer certains termes dans $E_p\varphi$.

Ce problème a valeur de référence en programmation entière, à cause de la grande simplicité de son énoncé, et en tant que sous-problème de situations plus complexes. De plus il se pose directement dans des applications industrielles de grande taille. Enfin il intervient en cryptographie (voir [5]). Aussi a-t' il fait l'objet de nombreuses recherches (voir [23]).

4.2.2 Estimation des gradients aux milieux des cubes

Soit (Γ^k) suite de cubes telle que $\Gamma^k \in \mathcal{F}(\Gamma^{k+1})$. Pour $n, m, \cdots \notin s^k$, on note $s^k(n,m,\cdots) = \mathrm{supp}(\Gamma^k) - \{n,m,\cdots\}$, $V^k(n,m,\cdots) = V^0 - \sum_{l \notin s^k} \rho_l^k v^l - v_n - v_m - \cdots$, et $W^k(n,m,\cdots) = (\sum_{l \in s^k(n,m,\cdots)} v_l^2)^{1/2}$. Soit $n \in \mathbb{N}$. On suppose $n \in s^0$. Donc $n \in s^k$ si $k \geqslant 0$, d'où d'après la formule de Russo

$$\begin{aligned}
&\nabla_n E_{\rho^k} \varphi \\
&= E_{\rho^k_{(n,1)}} \varphi - E_{\rho^k_{(n,0)}} \varphi \\
&= E_{\rho^k_{(n,1)}} (u^T \omega \, 1_{(v^T\omega \leqslant V)}) \; - \; E_{\rho^k_{(n,0)}} (u^T \omega \, 1_{(v^T\omega \leqslant V)}) \\
&= u^T \, (E_{\rho^k_{(n,1)}} - E_{\rho^k_{(n,0)}}) \{\omega \, 1_{(v^T\omega \leqslant V)}\}.
\end{aligned}$$

Or si $m \in s^k(n)$,

$$\begin{aligned}
&(E_{\rho^k_{(n,1)}} - E_{\rho^k_{(n,0)}}) \{\omega_m \, 1_{(v^T\omega \leqslant V)}\} \\
&= \frac{1}{2} \{\wedge_{\rho^k} (\sum_{l \in s^k(n,m)} v_l \, \omega_l \; \leqslant \; V^k(n,m)) - \wedge_{\rho^k} (\sum_{l \in s^k(n,m)} v_l \, \omega_l \; \leqslant \; V^k(m))\}.
\end{aligned}$$

De même,

$$(E_{\rho^k_{(n,1)}} - E_{\rho^k_{(n,0)}}) \{\omega_n \, 1_{(v^T\omega \leqslant V)}\} = \wedge_{\rho^k} (\sum_{l \in s^k(n)} v_l \, \omega_l \; \leqslant \; V^k(n)).$$

Enfin si $m \notin s^k$

$$(E_{\rho^k_{(n,1)}} - E_{\rho^k_{(n,0)}}) \{\omega_m \, 1_{(v^T\omega \leqslant V)}\} = 0.$$

Donc

$$\begin{aligned}
&\nabla_n E_{\rho^k} \varphi \\
&= u_n \wedge_{\rho^k} (\sum_{l \in s^k(n)} v_l \, \omega_l \; \leqslant \; V^k(n)) + \frac{1}{2} \sum_{m \in s^k(n)} u_m \\
&\quad \{\wedge_{\rho^k} (\sum_{l \in s^k(n,m)} v_l \, \omega_l \; \leqslant \; V^k(n,m)) - \wedge_{\rho^k} (\sum_{l \in s^k(n,m)} v_l \, \omega_l \; \leqslant \; V^k(m))\}
\end{aligned}$$

$$= \quad u_n \wedge_{\rho^k} \Big(\sum_{l \in s^k(n)} \frac{v_l(2\,\omega_l - 1)}{W^k(n)} \leqslant \frac{2\,V^k(n) - \sum_{l \in s^k(n)} v_l}{W^k(n)} \Big) + \frac{1}{2} \sum_{m \in s^k(n)} u_m$$

$$\Big\{ \wedge_{\rho^k} \Big(\sum_{s^k(n,m)} \frac{v_l(2\omega_l - 1)}{W^k(n,m)} \leqslant \frac{2V^k(n,m) - \sum_{s^k(n,m)} v_l}{W^k(n,m)} \Big)$$

$$- \wedge_{\rho^k} \Big(\sum_{s^k(n,m)} \frac{v_l(2\omega_l - 1)}{W^k(n,m)} \leqslant \frac{2V^k(m) - \sum_{s^k(n,m)} v_l}{W^k(n,m)} \Big) \Big\}.$$

Si de plus la suite v est à valeurs dans un compact $[c, C]$, où $c > 0$, alors quand $k \to \infty$, la distribution de la somme normalisée $\sum_{l \in s^k(n)} \frac{v_l(2\,\omega_l - 1)}{W^k(n)}$ (resp. $\sum_{l \in s^k(n,m)} \frac{v_l(2\,\omega_l - 1)}{W^k(n)}$) tend vers la distribution normale standard, d'après le théorème de Lindeberg (voir [27, p.262]). On choisit donc comme estimateur pour la composante d'indice n du gradient

$$\widehat{\nabla_n E_{\rho^k}} \varphi = u_n \, \phi\Big(\frac{2\,V^k(n) - \sum_{l \in s^k(n)} v_l}{W^k(n)} \Big) + \frac{1}{2} \sum_{m \in s^k(n)} u_m \Big(\phi\Big(\frac{2\,V^k(n,m) - \sum_{l \in s^k(n,m)} v_l}{W^k(n,m)} \Big) - \phi\Big(\frac{2\,V^k(m) - \sum_{l \in s^k(n,m)} v_l}{W^k(n,m)} \Big) \Big), \tag{19}$$

où ϕ est la distribution normale standard. De plus, substituant à l'estimateur (19) son développement de Taylor au premier ordre en v/V^k, on obtient pour $\nabla_n E_{\rho^k} \varphi$ un estimateur sous forme close $\widetilde{\nabla_n E_{\rho^k}} \varphi$, calculable en temps constant à partir de $\widetilde{\nabla_n E_{\rho^{k+1}}} \varphi$, équivalent en pratique à l'estimateur (19).

4.2.3 Algorithme du sac à dos

On utilise l'estimateur $\widetilde{\nabla_n E_{\rho^k}} \varphi$ dans un algorithme "récursif suivant les faces" (voir remarque 4.1). On se donne un hypercube Ω^0 de dimension d, sur lequel on cherche une allocation quasi-optimale pour φ. On suppose $u,\ v,\ V^0 > 0$, ce qui n'est pas restrictif, à un *preprocessing* linéaire trivial près (voir section 4.2.1). Dans l'algorithme suivant, la variable Ω désigne un hypercube de centre ρ et de support $\mathrm{supp}(\Omega)$, V étant une variable réelle.

1. $\Omega \leftarrow \Omega^0$, $k \leftarrow 0$, $V \leftarrow 0$
2. Calculer les $\widetilde{\nabla_n E_\rho} \varphi$ pour $n \in \mathrm{supp}(\Omega)$.
3. Tant que $k < d - \underline{d}$,
   (a) Pour tout $n \in \mathrm{supp}(\Omega)$ tel que $V + v_n > V^0$, faire
      i. $\Omega \leftarrow \Omega \ \cap \ \{\omega_n = 0\}$

ii. Mise à jour des $\widetilde{\nabla_n E_\rho \varphi}$ tels que $n \in \mathrm{supp}(\Omega)$

iii. $k \leftarrow k+1$

(b) Choisir n qui maximise $|\widetilde{\nabla_n E_\rho \varphi}|$ sur $\mathrm{supp}(\Omega)$.

(c) Faire

i. $\Omega \leftarrow \Omega \cap \{p_n = \omega_n\}$, où $\omega_n = \begin{cases} 1 & \text{si} \quad \widetilde{\nabla_n E_\rho \varphi} > 0 \\ 0 & \text{sinon} \end{cases}$

ii. $V \leftarrow V + v_n$

iii. Mise à jour des $\widetilde{\nabla_n E_\rho \varphi}$ tels que $n \in \mathrm{supp}(\Omega)$

iv. $k \leftarrow k+1$

4. Compléter la recherche par une procédure exhaustive de type *branch and bound*, sur l'hypercube Ω de dimension $\underline{d}$ atteint à l'issue des étapes précédentes.

5. Retourner l'allocation obtenue, et la valeur correspondante pour φ.

(3.a) garantit que l'allocation renvoyée par l'algorithme vérifie la contrainte. Lorsque la dimension de Ω devient trop faible pour l'estimation asymptotique (19), on procède à l' exploration *Branch and Bound* des configurations restantes (4). Ainsi une configuration ω satisfaisant la contrainte dans le problème (18), telle que $\varphi(\omega)$ est une valeur approchée du maximum de φ sur Ω.

4.2.4 Expériences numériques

On a testé numériquement l'algorithme sur données u et v indépendantes uniformément distribuées dans $\mathbb{N}_{1000}$, avec $V = 1/2 \sum_{n \in \mathrm{supp}\rho^0} v_n$, pour des valeurs de d comprises entre 20 et 5000, la profondeur de l'exploration finale $\underline{d}$ étant fixée à 12. On obtient un algorithme de complexité en temps quadratique en d, où la durée de l'exploration systématique est négligeable dès que $d \geqslant 20$. Les données u et v sont stockées sous forme de vecteurs d'entiers. La complexité en espace est donc linéaire. Confrontées à l'algorithme pseudo-polynômial de la programmation dynamique, les expériences font part d'une erreur relative inférieure au pourcent dès les plus basses valeurs de d, décroissant rapidement vers 0 quand d croît. Ces erreurs sont extrêmement faibles. A titre de comparaison, on renvoie à [23], dont on a emprunté le protocole de tests. Les résultats de nos expérimentations sur station Sun 20 sont résumés dans le tableau 1, page 36. Quand n croît, l'algorithme exact de vérification *a posteriori* par programmation dynamique devient coûteux en temps, d'où la diminution du nombre des expériences.

d	N	CPU	m	max	σ
20	100	0.01	0.00000	0.00000	0.00000
50	100	0.023	0.00830	0.26247	0.05125
100	200	0.0615	0.00440	0.12192	0.02282
200	200	0.24	0.00252	0.04064	0.00994
500	100	1.29	0.00135	0.01285	0.00381
750	100	3.27	0.00174	0.01715	0.00351
1000	50	5.77	0.00242	0.01044	0.00350
2000	10	23.09	0.00140	0.00419	0.00177
3000	5	51.59	0.00112	0.00155	0.00076
5000	3	144.85	0.00051	0.00090	0.00040

TAB. 1 - *Résultats d'expériences sur le problème du sac à dos.*

$$\begin{cases} d, & \text{le nombre de variables initial;} \\ N, & \text{le nombre d'expériences;} \\ CPU, & \text{le temps CPU moyen, en secondes;} \\ m, & \text{l'erreur relative moyenne, en \%;} \\ \max, & \text{l'erreur relative maximum, en \%;} \\ \sigma, & \text{l'écart-type de l'erreur relative.} \end{cases}$$

Remarque 4.2 (Où l'on retrouve la séparation des valeurs) On a choisi de procéder à des expériences numériques sur données entières dans $\mathbb{N}_{1000}$ afin de pouvoir vérifier les solutions obtenues par l'algorithme pseudo-polynômial de la programmation dynamique, et pour permettre la comparaison avec les résultats décrits dans [23]. Si on choisit les données dans $\mathbb{N}_N$, lorsque N croît, d'une part la durée de la vérification par programmation dynamique croît exponentiellement vite, d'autre part il faut augmenter (par exemple logarithmiquement) $\underline{d}$, si l'on veut éviter que les performances de notre algorithme ne finissent par se dégrader. En effet la convergence des estimateurs du gradient (19) devient de plus en plus lente. C'est ainsi que se manifeste à nouveau le lien entre la difficulté du problème et la séparation des valeurs de φ (voir commentaire 3.19), tant au niveau de la procédure exacte de la programmation dynamique qu' à celui de notre algorithme d'approximation. Par ailleurs si on introduit une corrélation entre u et v, le problème devient de plus en plus difficile (voir [23]), comme on peut s'en rendre compte en essayant de le résoudre "à la main". De fait on observe alors une détérioration des performances de notre algorithme. Ces

performances restent toutefois supérieures à celles rapportées dans [23] pour des algorithmes de même complexité, mais dans une moindre mesure que sur données u et v indépendantes.

4.3 Problème d'Ising

On donne une interprétation dans le cadre de la théorie développée ici à l'algorithme approché de P. Rivière pour le problème d'Ising (voir [19]). L' interprétation probabiliste de cet algorithme (voir section 4.3.3) est le point de départ de notre travail.

4.3.1 Classe de φ, modèle d'Ising

On cherche à minimiser $\varphi(\omega)$ sur un hypercube $\Omega^0 \subset \mathcal{B}$, où si $\omega \in \mathcal{B}$

$$\varphi(\omega) = \omega^T A \omega, \tag{20}$$

A étant fixée dans $\mathcal{M}_{\mathbb{N}}(\mathbb{R})$. Dans la suite on suppose A symétrique, sans restriction. En effet la contribution à φ d'une matrice quelconque se réduit à celle de sa partie symétrique.

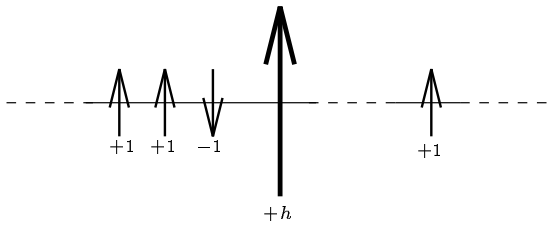


FIG. 6 - *Modèle d'Ising.*

La physique statistique considère des systèmes de verres de spins (voir figure 6), particules pointant vers le haut: $\sigma_i = +1$, ou vers le bas: $\sigma_i = -1$, dont l'énergie d'ensemble, donnée par l'hamiltonien $\mathcal{H}$ du système, s'écrit

$$\mathcal{H} = -\sigma^T J \sigma + \sigma^T h. \tag{21}$$

Dans l'expression (21), $J \in \mathcal{M}_{\mathbb{N}}(\mathbb{R})$ est une matrice strictement supérieure modélisant les interactions de court terme entre spins, tandis que $h \in \mathbb{R}^{\mathbb{N}}$ est un champ magnétique externe. Le but du problème est de déterminer un état stable du système, qui minimise l'énergie $\mathcal{H}$. Pour un système de spins dénombrable, le problème n'est résolu analytiquement qu'en

l'absence de champ magnétique, si le graphe des interactions est planaire[5](voir [18]). Par ailleurs il n'existe pas d'algorithme exact performant pour le problème combinatoire fini. On se limite donc souvent à l'étude du problème fini à d variables, pour lequel on procéde à des simulations. Ainsi les heuristiques sont-elles particulièrement précieuses pour ce problème, qui admet en outre une application importante en *imagerie* (voir [3]). Les problèmes en variables $\{-1,+1\}$ et $\{0,1\}$ étant reliés par le changement de variables trivial $\sigma_n = 2\omega_n - 1$, les mêmes algorithmes s'appliquent à l'un ou à l'autre indifféremment.

4.3.2 Algorithme [19]

Dans [19], P. Rivière présente un algorithme approché pour le problème (20). Reformulé dans le vocabulaire introduit ici, cet algorithme est une descente récursive de face en face, depuis le milieu ρ^0 d'un hypercube Ω^0, qui aboutit en un temps quadratique à une configuration $\omega \in \Omega^0$. Dans l'algorithme suivant la variable Ω désigne à nouveau un hypercube de centre ρ, les α_n étant des variables réelles. On définit également $s(n,m,\cdots) = \mathrm{supp}(\Omega) - \{n,m,\cdots\}$, $X_n = \alpha_n + \sum_{m\in s(n)} a_{m,n}$ et $Q_n = (\alpha_n + \sum_{m\in s(n)} a_{m,n})/(\sum_{m\in s(n)} |a_{m,n}|)$. L'algorithme [19] est alors le suivant, étant donnée la matrice symétrique A indicée par $\mathrm{supp}(\Omega^0)$

1. $k \leftarrow 0$, $\Omega \leftarrow \Omega^0$, $\alpha_n \leftarrow a_{n,n}$ si $n \in \mathrm{supp}(\Omega)$.
2. Calculer les Q_n pour $n \in \mathrm{supp}(\Omega)$.
3. Tant que $k < d$
   (a) Choisir n qui maximise $|Q_n|$ sur $\mathrm{supp}(\Omega)$.
   (b) Faire
      i. $\Omega \leftarrow \Omega \cap \{p_n = \omega_n\}$, où $\omega_n = \begin{cases} 1 & \text{si} \quad X_n < 0 \\ 0 & \text{sinon} \end{cases}$
      ii. Mettre à jour les α_n puis les Q_n pour $n \in \mathrm{supp}(\Omega)$
      iii. $k \leftarrow k+1$
4. Retourner l'allocation correspondant à l'hypercube de dimension nulle obtenue à l'issue des étapes précédentes, et la valeur correspondante pour φ.

5. interactions limitées aux plus proches voisins.

4.3.3 Interprétation probabiliste

Soit (Γ^k) suite de cubes de centres ρ^k, de supports s^k, tel que $\Gamma^k \in \mathcal{F}(\Gamma^{k+1})$. Pour $n \in s^k$, notons $s^k(n,m,\cdots) = \text{supp}(\Gamma^k) - \{n,m,\cdots\}$, $a^k_{n,n} = a_{n,n} + 2\sum_{m \notin s^k(n)} a_{m,n}\rho^k_m$, $X^k_n = a^k_{n,n} + 2\sum_{m\in s^k(n)} a_{m,n}\omega_m$, et $Q^k_n = (|a^k_{n,n} + \sum_{m\in s^k(n)} a_{m,n}|)/(\sum_{m\in s^k(n)} |a_{m,n}|)$. Isolant la contribution à φ de la variable ω_n, il vient, grâce à l'identité $\omega_n^2 = \omega_n$:

$$\varphi(\omega) = \omega_n(2\sum_{m\neq n} a_{m,n}\omega_m + a_{n,n}) + \sum_{m,l\neq n} a_{m,l}\,\omega_m\omega_l.$$

Par suite, $\sum_{m,l\neq n} a_{m,l}\omega_m\omega_l$ étant indépendant de la n-ième composante de ρ^k

$$\begin{aligned}
&\nabla_n E_{\rho^k}\varphi \\
&= (E_{\rho^k_{(n,1)}} - E_{\rho^k_{(n,0)}})\,\omega^T A\omega \\
&= (E_{\rho^k_{(n,1)}} - E_{\rho^k_{(n,0)}})\,\omega_n(2\sum_{m\neq n} a_{m,n}\omega_m + a_{n,n})
\end{aligned}$$

Soit

$$E_{\rho^k_{(n,1)}} X^k_n = a^k_{n,n} + \sum_{m\in s^k(n)} a_{m,n}$$

De plus on a

$$\sigma_{\rho^k_{(n,1)}}(2\sum_{m\neq n} a_{m,n}\omega_m + a_{n,n}) = (\sum_{m\in s^k(n)} a^2_{m,n})^{1/2} \approx \sum_{m\in s^k(n)} |a_{m,n}|.$$

D'après le théorème de Lindeberg (voir section 4.2) , la variable X^k_n est asymtotiquement normale en k, de moyenne et écart-type

$$E_{\rho^k_{(n,1)}} X^k_n = a^k_{n,n} + \sum_{m\in s^k(n)} a_{m,n} \quad , \quad \sigma_{\rho^k_{(n,1)}} X^k_n \approx \sum_{m\in s^k(n)} |a_{m,n}|.$$

Donc l'indice $n \in s^k$ qui maximise $|Q^k_n|$ correspond à la moindre incertitude sur le signe de $\nabla_n E_{\rho^k}\varphi$, asymptotiquement en k. Comme les variables X^k_n sont asymptotiquement gaussiennes, elles ont "peu de chances" de s'écarter notablement de leur moyenne, et la plus dissymétrique d'entre elle par rapport à 0 a "encore moins de chance" de prendre une valeur de signe opposé à celui de sa moyenne.

4.3.4 Expériences numériques

Dans [19], l'auteur teste l'efficacité de son heuristique sur des problèmes à $\leqslant 50$ variables; il obtient des erreurs moyennes de l'ordre du pourcent. À partir de 50 variables, l'absence de

procédure de vérification en temps raisonnable empêche de tester numériquement l'heuristique. Cependant on peut confronter les valeurs renvoyées par l'algorithme à une conjecture de la physique statistique. En effet si $h = \mathbf{0}$, et pour tout $k < l \in \text{supp}(\rho^0)$ $J_{kl} \rightsquigarrow \mathcal{N}_{(0,1/d)}$, une conjecture de la théorie de la réplication veut que $\mathcal{H}_0^d/d \overset{(loi)}{\sim} \mathcal{N}_{(\mathcal{H}_0,1/d)}$, où $\mathcal{H}_0 \simeq 0,7633$ à 10^{-4} près (voir [16]). Reprenant l' heuristique [19] randomisée (voir section 4.1.3), on obtient le tableau 4.3.4, correspondant à des statistiques sur une centaine d'expériences (matrices A) différentes. Pour 500 variables, l'erreur par rapport à la loi limite du minimum semble donc de l'ordre de 5%.

N	m	σ^2
1	-0.7055	$1.65\ 10^{-4}$
30	-0.7278	$6.01\ 10^{-5}$

TAB. 2 - *Résultats d'expériences sur le problème d'Ising à 500 variables.*

$\left\{ \begin{array}{ll} N, & \text{le nombre de descentes, la première étant dérandomisée;} \\ m, & \text{la moyenne des meilleures valeurs sur N descentes;} \\ \sigma^2, & \text{la variance de la meilleure valeur sur N descentes.} \end{array} \right.$

5 Conclusion

L'obstacle majeur à notre méthode est l'absence de technique générale pour estimer nos gradients: on doit procéder au cas par cas, classe de fonctions φ par classe de fonctions φ. Cependant cette difficulté ne fait que refléter la NP-complétude des problèmes traités, également appelée en programmation dynamique "the curse of dimensionality" (voir [21]). Aussi cela ne doit-il pas nous surprendre, mais plutôt nous inciter à envisager davantage encore la résolution pratique de problèmes combinatoires comme un arbitrage entre complexité et qualité de la solution obtenue. De plus les bons résultats sur les problèmes d'Ising et du sac à dos incitent à penser qu'on peut étendre la classe des problèmes traitables.

La modélisation des problèmes de décision finis sur l' *hypercube* semble préférable aux modèles *arborescents* usuels. Les probabilités s' introduisent rigoureusement, par convexification; lorsqu'on désire incorporer la notion de choix de la variable, là où le modèle arborescent est obligé de recourir à des arbres non déterministes, dans lesquels un niveau n'est plus assujetti à une variable en particulier, sur le cube l'ordre de choix des variables est simplement

celui des faces rencontrées par une trajectoire. Il ne faut pas confondre cette convexification avec une méthode de relaxation continue dans un sens plus habituel. Ainsi le problème auquel on est ramené est complètement *non* convexe, ce que traduit le grand nombre des dimensions, et le fait que les minima locaux sont atteints aux sommets du cube. De plus on procède par moyennisation, non par approximations successives.

L' introduction du *hasard* pour la résolution de problèmes déterministes peut sembler surprenante, mais il en existe d'autres exemples, par exemple en contrôle optimal (voir [10, p.81-90]). De plus, en ce qui concerne les problèmes combinatoires, leur "calcul des variations", ou programmation dynamique, n'est pas assez universel en pratique. Il faut donc essayer d'autres voies. Souhaitant obtenir une bonne valeur approchée dès la première valeur de φ calculée, en assignant les variables une par une, selon un ordre et des valeurs à déterminer, n'est-il pas raisonnable, en l'absence d'information supplémentaire, de raisonner pour l'assignation suivante en termes probabilistes, conditionnellement à la seule certitude des assignations déjà fixées? C'est là en effet l' interprétation de la dynamique de descente des potentiels, au vu de la formule de Russo.

Annexe. Utilité économique et espérance mathématique

> "We have practically defined numerical utility as being that thing for which the calculus of mathematical expectations is legitimate." *J. Von Neumann et O. Morgenstern, "Theory of Games and Economic Behavior."*

Soit $\mathcal{P}$ l'ensemble des distributions de probabilités sur un espace probabilisable Ω. Toute relation de préférence $\sim$ sur $\mathcal{P}$ vérifiant l' axiomatique de Von Neumann Morgenstern s'exprime à l'aide d'une fonction numérique définie sur *l'ensemble* Ω (voir [14, p.26]). Le fait que la relation de préférence soit définie sur $\mathcal{P}$ et non sur Ω signifie qu'on sait choisir entre deux biens ω et ω', mais aussi décider si on préfère ω à ω' plus qu'on ne préfère ω' à ω'', ou réciproquement. L' axiomatique de Von Neumann Morgenstern signifie qu'il n'y a pas d'intérêt spécifique à jouer. Par exemple la relation de préférence suivante ne satisfait pas à

cette condition, car il existe une loterie entre deux biens préférée à chacun de ces deux biens pris séparément.

Exemple 5.1 Soit $\Omega = \{0,1\}$. L'ensemble $\mathcal{P}$ des distributions de probabilités sur $(\Omega, \mathcal{P}(\Omega))$ est muni de la distance de la variation totale,

$$d(\mu,\nu) \stackrel{\Delta}{=} \sup_{A\in\mathcal{P}(\Omega)} |\mu(A)-\nu(A)| = \inf_{X\rightsquigarrow\mu\ ,\ Y\rightsquigarrow\nu} P(X\neq Y).$$

On considère la masse de Dirac en 0 (resp. en 1) δ_0 (resp. δ_1), et la distribution uniforme $\mathcal{U}$. Alors

$$\left\{\begin{array}{ccccc} d(\delta_0,\mathcal{U}) & = & d(\delta_1,\mathcal{U}) & = & 1/2 \\ d((\delta_0+\delta_0)/2,\mathcal{U}) & = & d(\delta_0,\mathcal{U}) & = & 1/2 \\ d((\delta_0+\delta_1)/2,\mathcal{U}) & = & d(\mathcal{U},\mathcal{U}) & = & 0. \end{array}\right.$$

Ainsi une loterie entre les biens 0 et 1 est jugée préférable aux biens 0 et 1 pris séparément, au sens de la moindre distance à $\mathcal{U}$. Donc la relation de préférence sur $\mathcal{P}$ définie par la moindre distance à $\mathcal{U}$ ne vérifie pas l'axiomatique de Von Neumann Morgenstern.

Pour le détail de l'axiomatique, on renvoie à $[14, p.26]$. Alors, avec les notations précédentes

Théorème 5.2 ([14, p.17]) *Si $\sim$ vérifie l'axiomatique de Von Neumann Morgenstern, alors il existe une fonction* numérique $\varphi : \Omega \to \mathbb{R}$ *telle que*

$$\forall P \in \mathcal{P} \quad \sim_P \equiv E_P\varphi.$$

La fonction numérique φ, définie sur l'ensemble des *biens* Ω (au lieu de l'ensemble des *distributions* $\mathcal{P}$), mesure les préférences sous la relation $\sim$, à transformation strictement croissante affine près, comme une température mesure la chaleur.

Remerciements

Ce travail, issu d'une réflexion initiée à l'ENSAE à la *Formation par la recherche* dirigée par M. A. Trognon, est le résultat d'un stage de DEA IMA réalisé à l'INRIA Sophia-Antipolis, au sein du projet Mistral de M. F. Baccelli. Je suis particulièrement redevable, d'abord à l'ENSAE à M. P. Rivière de m'avoir formé à ces problèmes (voir [19]), il est à l'origine de nombreuses idées développées ici; puis au projet Mistral à M. S. Zuyev, d'avoir rendu possible le présent travail en déplaçant la problématique sur le terrain d'un "système dynamique" approprié. Enfin je remercie les membres de Mistral de leur accueil extrêmement sympathique.

Références

[1] A.Borovkov. *Statistique mathématique.* Mir, 1987.

[2] H.Simon A.Newell. *Human problem solving.* Prentice-Hall, 1972.

[3] R. Azencott. *Simulated annealing: parallelization techniques.* Wiley, 1992.

[4] D.P. Bertsekas. *Dynamic programming and stochastic control.* Mathematics in science and engineering, 1976.

[5] B.Z.Chor. *Two issues in public key cryptography : RSA bit security and a new knapsack system.* ACM distinguished dissertation. MIT press, 1986.

[6] M. Girshick D.Blackwell. *Theory of games and statistical decisions.* Wiley, 1954.

[7] J. Faraut. *Théorie du potentiel et analyse harmonique.* Springer, 1974.

[8] D.Johnson G.Garey. *Computers and intractability: a guide to the theory of NP-Completness.* San-Francisco, 1979.

[9] P.Van Hentenryck. *Constraint satisfaction in logic programming.* MIT press, 1989.

[10] G.Dupuis H.Kushner. *Numerical methods for stochastic control problems in continuous time.* Springer, 1992.

[11] I.Ekeland. *La théorie des jeux et ses applications à l'économie mathématique.* PUF, 1974.

[12] J.Dieudonné. *Calcul infinitésimal.* Méthodes, 1968.

[13] J.P.Aubin. *L'analyse non linéaire et ses motivations économiques.* Masson, 1984.

[14] O. Morgenstern J.Von Neumann. *Theory of games and economic behavior.* Princeton, 3ème edition, 1953.

[15] L.Russo. On the critical percolation probabilities. *Z. Wahrsch. Verw. Geb.*, 56:229–237, 1981.

[16] G.Parisi M.Mézard. A replica analysis of the traveling salesman problem. *J.Physique*, 47:1285–1296, août 1986.

[17] M.Minsky. Steps towards artificial intelligence. In A.Feigenbaum, editor, *Computers and thought*. MacGraw-Hill, 1963.

[18] L. Onsager. *Phys. Rev.*, 65:117–149, 1964

[19] P.Rivière. Minimization of a quadratic boolean form: three algorithms based on constraint programming. *soumis à RAIRO*, 1994.

[20] S.Amari. Information geometry of boltzmann machines. *IEEE transactions on neural networks*, 3(2), 1992.

[21] R. Bellman S.Dreyfus. *Applied Dynamic Programming*. Princeton University Press, 1962.

[22] D.Geman S.Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE transactions on pattern analysis and machine intelligence*, 6(6), 1984.

[23] P. Toth S.Martello. *Knapsack Problems, Algorithms and Computer Implementations*. Wiley, 1990.

[24] V.Arnold. *Equations différentielles ordinaires*. Mir, 1974.

[25] S.Smale W. Hirsch. *Differential equations, dynamical systems, and linear algebra*. Academic-Press, 1974.

[26] A. Wald. *Statistical decision functions*. Wiley, 1950.

[27] W.Feller. *An introduction to probability theory and its applications*, volume II. Wiley, 3 edition, 1968.



Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399