

Navigation, fonctions harmoniques et contrôle optimal stochastique

Amine Boumaza, Bruno Scherrer

▶ **To cite this version:**

Amine Boumaza, Bruno Scherrer. Navigation, fonctions harmoniques et contrôle optimal stochastique. Cinquièmes Journées Nationales sur Processus Décisionnel de Markov et Intelligence Artificielle - PDMIA 2005, Jun 2005, Lille/France. inria-00000644

HAL Id: inria-00000644
https://inria.hal.science/inria-00000644

Submitted on 10 Nov 2005

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

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

Navigation, fonctions harmoniques et contrôle optimal stochastique

Amine Boumaza (boumaza@loria.fr**)**
Bruno Scherrer (scherrer@loria.fr**)**

Equipe MAIA, LORIA, INRIA-Lorraine,
Campus Scientifique BP 239
54506 Vandœuvre-lès-Nancy, France

RÉSUMÉ. *Nous considérons le problème de la navigation en temps continu dans un environnement stochastique. Nous présentons les approches de planification 1) par fonction harmonique et 2) par contrôle optimal stochastique. Nous montrons que le contrôle par fonction harmonique est un cas très particulier de contrôle optimal stochastique : le contrôle par fonction harmonique est la limite (extrêmement pessimiste) du contrôle optimal quand l'amplitude du bruit de déplacement tend vers l'infini. Nous expliquons que toutes les propriétés intéressantes qui ont motivé (par exemple dans la communauté robotique) l'utilisation de fonctions harmoniques sont également observées dans les modèles de navigation issus du contrôle optimal stochastique. Montrant ainsi que, dans le contexte de la navigation, la théorie du contrôle optimal stochastique est strictement plus générale que celle des fonctions harmonique, nous espérons motiver son utilisation. Nous illustrons notre propos par des simulations numériques.*

ABSTRACT.

MOTS-CLÉS : *navigation en temps continu, contrôle optimal stochastique, fonctions harmoniques*

KEYWORDS:

[e] soumission à *PDMIA 2005*, le 26 avril 2005.

1. La problématique de navigation

Dans ce papier, nous considérons la problématique de navigation. Nous commençons par préciser comment nous modélisons ce problème.

Nous considérons un agent ponctuel naviguant dans $\overline{O}$, la fermeture de O, ouvert connexe et borné de $\mathbb{R}^2$. A chaque instant $t > 0$, la dynamique de ce système est caractérisée par les coordonnées $x(t) = (x_1(t), x_2(t))$ de l'agent et une commande de direction $\theta(t)$. Formellement, nous avons :

$$\frac{dx}{dt} = \overrightarrow{u}(\theta(t)) + \sigma \frac{dw}{dt}$$

où $\overrightarrow{u}(\theta)$ est un vecteur unitaire de direction θ, w est la diffusion (l'aspect stochastique), c'est-à-dire un mouvement brownien de dimension 2, et $\sigma > 0$ est une constante positive qui correspond au niveau de bruit[1]. Il s'agit d'une équation différentielle stochastique (EDS).

On considère qu'une (réalisation de) trajectoire se termine dès lors qu'elle sort de $\overline{O}$. La frontière ∂O de $\overline{O}$ est partitionnée en deux : l'ensemble des obstacles $\mathcal{O}$ et l'ensemble des buts $\mathcal{G}$. Si le point de sortie de la trajectoire est un but (resp. un obstacle) on considère que l'agent a accompli sa tâche (resp. échoué).

Dans ce cadre, le terme navigation correspond à divers problèmes, dont par exemple les suivants :

– trouver, à partir d'une position initiale, une loi de commande qui génère des chemins vers n'importe quel bord but.

– même problème mais à partir d'un ensemble de positions initiales

– trouver, à partir d'un ensemble de positions initiales, une loi de commande qui optimise les chemins (dans un sens qui sera à préciser).

Les approches que nous présentons dans ce qui suit, permettent de répondre à ces problèmes.

2. Navigation par fonction harmonique

2.1. *Principes*

L'utilisation de fonctions harmoniques permet d'aborder le problème suivant : trouver à partir de n'importe quelle position initiale de l'espace libre, un chemin vers un bord but. L'approche que nous présentons dans cette section a été initialement proposée (semble t-il simultanément) par [CON 90] et [AKI 90].

1. L'utilisation d'un coefficient σ plutôt que d'une matrice revient à faire l'hypothèse (naturelle) d'isotropie sur le bruit.

Soit Φ une fonction, deux fois dérivable, à valeurs réelles définie sur le fermé $\overline{O}$ (tel qu'introduit plus haut). On dit que la fonction Φ est harmonique si et seulement si son Laplacien est nul :

$$\Delta\Phi = \frac{\partial^2\Phi}{\partial {x_1}^2} + \frac{\partial^2\Phi}{\partial {x_2}^2} = 0 \qquad [1]$$

Une propriété remarquable des fonctions harmoniques est la suivante : Si Φ est une fonction harmonique dans $\overline{O}$, alors soit Φ atteint son maximum (resp. son minimum) uniquement sur la frontière ∂O, soit Φ est constante. Comme nous allons le voir, cette propriété peut être particulièrement intéressante pour la navigation.

Pour le problème de navigation (introduit à la section 1), on considère une fonction harmonique qui vérifie des conditions au bord particulier, ce qu'on appelle aussi problème de Dirichlet : Soit Φ l'unique fonction $\mathcal{C}^2$ sur $\overline{O}$ telle que

$$\forall x \in O,\ \Delta\Phi\left(x\right) = 0 \text{ avec les conditions au bord } \forall x \in \partial O,\ \Phi\left(x\right) = g\left(x\right). \qquad [2]$$

g, fonction sur $\overline{O}$, représente les valeurs de la fonction de navigation dans les zones obstacles et buts. On choisit souvent, sans perte de généralité[2], $g\left(\mathcal{O}\right) = 1$ et $g\left(\mathcal{G}\right) = 0$. Les seuls minima de cette fonction sont atteints dans les zones buts. La navigation se fait alors en suivant le gradient de la fonction Φ.

En pratique, la fonction de navigation harmonique est appochée, par exemple, par un schéma numérique par différences finies sur une grille régulière. Après une discrétisation de l'environnement et la mise en place des conditions aux bords, on peut montrer que la forme discrète de l'équation de Laplace se traduit par la propriété de la moyenne qui dit que la valeur en chaque point de la grille vaut la moyenne de ses quatre voisins. L'ensemble d'équations linéaires ainsi défini peut être résolu par des méthodes itératives telles Jacobi ou Gauss-Seidel dont le principe est dans ce cas de remplacer à chaque itération la valeur en un point par la moyenne de ses quatre voisins. Si la discrétisation est bien choisie, après convergence, la fonction de navigation discrète ainsi obtenue atteint des minima uniquement dans les zone but (condition de Dirichlet -1). La navigation vers le but se fait alors en suivant le gradient (déterminé par interpolation).

2.2. *Propriétés intéressantes*

La navigation par fonction harmonique possède plusieurs propriétés particulièrement intéressantes, qui ont motivé leur utilisation dans de nombreuses applications, en particulier en robotique [AKI 90, TRA 91, KIM 92, CON 93a] :

– La plus importante est, comme nous l'avons souligné, l'absence de minima locaux. Ainsi, les trajectoires sont complètes c'est-à-dire qu'à partir de n'importe quel

2. Toute transformation linéaire d'une fonction harmonique est encore une fonction harmonique.

point de l'espace libre il existe un chemin (déterministe) vers le but[3]. Si un point et les but n'appartiennent pas au même espace connexe, alors la fonction est constante. Cette propriété est à mettre en contraste avec d'autres approches de navigation, notamment par champs de potentiels (par exemple [KHA 86]), qui admettent dans le cas général des minima locaux.

– Les fonctions harmoniques permettent la mise à jour incrémentale du modèle de l'environnement. Les nouveaux obstacles détectés peuvent être intégrés directement dans la fonction durant le calcul (on met à jour le bord ∂O et la fonction g). Ceci permet de mettre à jour des modèles d'environnements dynamiques ou incomplets et explorés en ligne.

– La nature des trajectoires générées par les fonctions harmoniques peut être très flexible. Elle est liée aux conditions aux bords utilisés. Celles que nous avons présentées (les conditions de Dirichlet) produisent des trajectoires qui ont tendance à s'éloigner des obstacles. Ceci est dû au fait que le flux de potentiel est orthogonal au bord. Une variante consiste à utiliser des conditions de Neumann (conditions limites sur la dérivées de Φ), pour lesquelles le flux est tangentiel au bord et les trajectoires qui en découlent ont tendance à longer les obstacles. En utilisant des combinaisons de différentes conditions aux bords, il est possible de générer une grande variété de comportements intermédiaires.

– Une dernière propriété intéressante de ce type de fonctions est la distribution naturelle du calcul. En effet, les méthodes itératives décrites précédemment ont seulement besoin d'une information locale pour calculer la valeur d'un point de la grille, ce qui les rend massivement parallélisables. Par exemple, une implémentation matérielle très bas niveau a été proposée dans [TRA 91] dans le but de la navigation robotique : une grille de résistances électriques représente la fonction harmonique ; une fois en équilibre, les points de jonctions des résistances ont des voltages moyen des points avoisinants (c'est la propriété de la version discrétisée du problème). De plus, ce caractère distributé font de ce modèle un candidat potentiel pour expliquer le calcul de commande dans le cerveau. Par exemple, Connolly [CON 92, CON 93b] propose un modèle de calcul du noyau gris central fondé dans lequel le striatum serait la structure cérébrale qui calculerait une fonction harmonique pour la commande motrice.

3. Navigation par contrôle optimal stochastique

La modélisation du problème de navigation (introduit à la section 1) dans le cadre de la théorie du contrôle, permet de résoudre le problème suivant : trouver, pour l'ensemble des positions initiales, une loi de contrôle $\theta(t)$ qui optimise dans un certain sens le chemin. Nous commençons par présenter les résultats principaux de la théorie du contrôle. Nous explicitons ensuite comment cela peut s'appliquer au modèle de navigation introduit plus haut.

3. Dans le cas stochastique, la complétude d'un contrôle, c'est-à-dire l'assurance (avec probabilité 1) qu'on atteigne un but avant un obstacle est impossible.

3.1. *Contrôle optimal stochastique*

Le contrôle optimal stochastique considère un système contrôlé défini à l'instant t par son état $x(t) \in \overline{O}$, où $\overline{O}$, la fermeture de O, ouvert borné de $\mathbb{R}^d$, est l'espace d'état. La dynamique d'état du système est régie par l'EDS suivante :

$$\frac{dx}{dt} = f(u(x(t))) + \sigma(x(t), u(x(t)))\frac{dw}{dt}$$

où le deuxième terme correspond à une dérive stochastique, dans lequel σ est en général une matrice de dimension $d \times r$ qui dépend de l'état x et de la commande $u(.)$.

Nous présentons uniquement ici le cas où l'horizon temporel est infini. Pour tout état initial x, toute loi de contrôle $u(.)$, et toute réalisation de la trajectoire $x(t)$ notons $\tau(x, u)$ le temps de sortie de $x(t)$ de $\overline{O}$ (avec la convention que ce taux vaut ∞ si la trajectoire reste indéfiniment dans $\overline{O}$). On définit la fonctionnelle de renforcement actualisée comme suit :

$$J(x, u) = E\left[\int_0^{\tau(x,u)} \gamma^t c(x(t), u(x(t)))dt + \gamma^\tau C(x(\tau(x, u))\right]$$

où c est le coût instantané et C est le coût terminal, et $\gamma \in [0, 1[$ est un facteur de décroissance (qui assure dans le cas de trajectoire infinie la finitude de J).

Sous les bonnes hypothèses de régularité, l'équation différentielle suivante :

$$J(x, u)\ln(\gamma) + c(x, u(x)) + J_x(x, u)f(u(x)) + \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x, u)J_{x_i x_j}(x, u) = 0$$

$$\text{avec les conditions au bord } \forall x \in \partial O, J(x, u) = C(x)$$

est vérifiée. On a noté J_x le gradient de J, $J_{x_i x_j}$ la dérivée partielle de J par rapport à x_i puis x_j, et les $a_{ij}(x, u)$ sont les coefficients de la matrice $a = \sigma(x, u(x))\sigma(x, u(x))^T$.

On définit alors la fonction de valeur optimale J^* comme la valeur minimale de la fonction de valeur pour toutes les lois de contrôle :

$$J^*(x) = \inf_{u(.)} J(x, u)$$

Sous les bonnes hypothèses de régularité, la fonction de valeur optimale J^* est solution de l'équation différentielle suivante :

$$J^*(x)\ln(\gamma) + \inf_u \left\{ c(x, u(x)) + J_x^*(x)f(u(x)) + \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x, u)J_{x_i x_j}^*(x) \right\} = 0 \quad [3]$$

avec les conditions au bord

$$\forall x \in \partial O, J^*(x, u) = C(x)$$

La fonction de valeur optimale est particulièrement intéressante car elle permet de déduire un contrôleur optimal déterministe. En tout état x le contrôle optimal $u^*(x)$ est l'argument u pour lequel l'infimum est atteint dans l'équation 3, ce qu'on appelle aussi le contrôleur gourmand par rapport à J^* :

$$u^*(x) = \arg\inf_u \left\{ c(x, u(x)) + J^*_x(x) f(u(x)) + \frac{1}{2} \sum_{i,j=1}^{n} a_{ij} J^*_{x_i x_j}(x) \right\}$$

En pratique, le calcul approché de la fonction de valeur optimale J^* peut être fait par un schéma numérique de type discrétisation. Par exemple, des schémas comme [MUN 97, MUN 05] impliquent l'introduction d'une version discrète du contrôle optimal, sous la forme d'un processus décisionnel de Markov. Ce dernier peut être résolu par des méthodes de programmation dynamique [PUT 94].

3.2. *Application à la navigation*

Afin d'optimiser un chemin dans le problème de navigation introduit à la section 1, nous proposons la modélisation naturelle suivante :

– $u = \theta, f(u) = \overrightarrow{u}(\theta)$ (le paramètre de contrôle est l'angle de déplacement)

– $\forall x \in O, c(x) = 0$ (pas de coût la plupart du temps)

– $\forall x \in \mathcal{O} \subset \partial O, C(x) = C(\mathcal{O}) > 0$ (un coût positif pour les obstacles) et $\forall x \in \mathcal{G} \subset \partial O, C(x) = C(\mathcal{G}) < 0$ (un coût négatif pour les buts)

– une matrice de diffusion constante de la forme σI (ici σ est la constante correspondant au niveau de bruit, introduite à la section 1) et donc $a = \sigma^2 I$.

Dans ce cas particulier, l'équation 3 qui caractérise la fonction de valeur optimale devient :

$$J^*(x) \ln(\gamma) + \inf_\theta \{ J^*_x(x) \overrightarrow{u}(\theta) \} + \frac{\sigma^2}{2} \Delta J^*(x) = 0 \qquad [4]$$

Nous avons de plus les propriétés suivantes :

– Le contrôleur optimal en un état x est dans la direction qui descend le gradient de J^* en x.

– Pour tout $\gamma \in [0, 1]$ (pour $\gamma = 1$, la fonctionnelle J est ici toujours finie), la fonction de valeur d'un contrôleur u vaut en x

$$C(\mathcal{O}) E[\gamma^{\tau(x,u)} | x(\tau(x, u)) \in \mathcal{O}] + C(\mathcal{G}) E[\gamma^{\tau(x,u)} | x(\tau(x, u)) \in \mathcal{G}] \qquad [5]$$

– Pour tout $\gamma \in [0, 1[$, le contrôleur optimal, qui consiste à descendre le gradient de la fonction de valeur optimale, est celui qui minimise (en espérance) une fonction croissante du temps pour atteindre le but tout en maximisant (en moyenne) une fonction croissante du temps moyen éventuel d'atteindre un obstacle. L'importance relative de ces deux objectifs est liée au rapport $\frac{C(\mathcal{O})}{C(\mathcal{G})}$.

– Dans le cas déterministe (c'est-à-dire quand $\sigma \to 0$)[4], et pour tout $\gamma \in [0, 1[$, le contrôleur optimal, qui consiste à descendre le gradient de la fonction de valeur optimale, est celui qui minimise le temps pour atteindre le but.

La première propriété s'obtient en remarquant que, pour l'équation 4, le seul terme qui intervient dans l'infimum est $J_x^*(x)\overrightarrow{u}(\theta)$, et que ce dernier est minimal lorsque θ est dans la direction qui descend $J_x^*(x)$. La deuxième propriété vient du fait que le coût instantané $c(.)$ est nul. La troisième propriété est la conséquence des deux premières et du fait que $C(\mathcal{G}) < 0 < C(\mathcal{O})$. La dernière propriété vient du fait, que dans le cas déterministe, un chemin optimal atteint nécessairement le but et minimise γ^τ, ce qui équivaut à minimiser le temps τ.

4. Lien entre fonctions harmoniques et contrôle optimal

Cette section contient la contribution essentielle de ce papier. Nous interprétons la navigation par fonction harmonique dans le cadre de la théorie du contrôle optimal stochastique. De plus, nous montrons que les propriétés intéressantes (section 2.2) qui rendent l'utilisation de fonctions harmoniques séduisante sont également obtenues dans le cadre du contrôle optimal stochastique.

4.1. *Un cas particulier de contrôle optimal*

Considérons l'amplitude du bruit σ comme un paramètre ; on peut voir la solution de l'equation différentielle 4 avec les conditions au bord correspondantes comme une fonction $\sigma \mapsto J^*(x, \sigma)$. On peut réécrire l'équation 4 comme suit :

$$\Delta J^*(x,\sigma) = -\frac{2}{\sigma^2}\left(J^*(x,\sigma)\ln(\gamma) + \inf_\theta \{J_x^*(x,\sigma)\overrightarrow{u}(\theta)\}\right)$$

ce qui suggère[5] que

$$\forall x, \Delta \lim_{\sigma\to\infty} J^*(x,\sigma) = 0$$

c'est-à-dire (comme les conditions au bord sont les mêmes pour les deux problèmes) :

$$\forall x, \lim_{\sigma\to\infty} J^*(x,\sigma) = \Phi(x)$$

où $\Phi(x)$ est la fonction harmonique pour la navigation (équation 2). En d'autres termes, le calcul d'un contrôleur par fonction harmonique est un cas particulier de calcul de fonction de valeur : c'est le cas limite (extrêmement pessimiste) où l'on considère que le bruit de navigation est infiniment grand, et que la commande de l'agent est par conséquent négigeable. Ceci explique d'ailleurs pourquoi les trajectoires produites par l'approche harmonique sont souvent qualifiées de "très sûres" [CON 93a, SIN 94].

4. Le contrôle optimal déterministe requiert une analyse un peu plus complexe, impliquant la notion de solution de viscosité (voir [MUN 97, MUN 00]), que nous éludons ici.
5. Une démonstration rigoureuse reste à faire.

4.2. *Propriétés intéresssantes*

A la section 2.2, nous avons argumenté que l'approche harmonique présentait un certain nombre de propriétés intéressantes. Qu'en est-il pour le contrôle optimal ? Nous reprenons ici ces propriétés.

– En général, la fonction de valeur peut admettre des minima locaux. Néanmoins, la loi de contrôle qui en est déduite est complète dans le même sens que pour les fonctions harmoniques : dans le cas déterministe ($\sigma \rightarrow 0$), la loi de contrôle amène toujours au but. Par contre, dans le cas stochastique, la loi de contrôle peut ne pas amener au but si cela s'avère moins intéressant au sens du critère défini par l'équation 5.

– Les fonctions de valeur permettent également la mise à jour incrémentale du modèle de l'environnement. Les nouveaux buts/obstacles détectés peuvent être intégrés dans la fonction durant le calcul. La mise à jour d'un modèle incomplet ou dynamique dans le cadre du contrôle correspond alors à des problèmes dits d'apprentissage par renforcement (voir [SUT 98] pour une présentation générale du domaine, voir [MUN 97] pour l'étude plus particulière du cas à temps continu).

– Dans la théorie du contrôle optimal, les lois de contrôle peuvent être affinées de multiples façons. Contrairement au cas harmonique, on peut expliciter, via la définition des fonctions coûts c et C, la gravité relative de toucher des obstacles les uns par rapport aux autres, l'importance relative d'atteindre les buts les uns par rapport aux autres, et l'importance relative d'atteindre les buts plutôt que de toucher un obstacle. De même, les lois de la dynamique f, et le niveau de bruit σ, sont autant de paramètres qui peuvent explicitement influencer la nature des trajectoires produites.

– Le calcul de la fonction de valeur est un candidat naturel pour la parallélisation. Les méthodes itératives (Gauss-Seidel, Jacobi) qui permettent de calculer les versions discrétisées du problème, peuvent être mises en œuvre de manière complètement asynchrone, ce qui permet de construire des architectures massivement parallèles[BER 89]. Ainsi, il est possible d'implémenter ces algorithmes sur microprocesseurs parallèles avec délai de communications et sans synchronisation[BER 83]. Enfin, comme dans le cas harmonique, ces implémentations massivement parallèles ont motivé des analogies avec ce qui se passe dans le cerveau[SCH 04].

Ainsi, la théorie du contrôle optimal est non seulement plus générale que celle des fonctions, mais elle présente également les propriétés (mentionnées à la section 2.2) qui rendent l'approche harmonique a priori intéressante.

4.3. *Travaux liés dans la littérature*

Dans [CON 94], l'auteur montre le lien entre fonction harmonique et probabilité de collision : il en déduit une relation proche de l'équation 5 (avec $\gamma = 1$). Par rapport à ce que nous présentons ici, l'auteur raisonne uniquement sur le schéma numérique aux différences finies correspondant, alors que nous avons ici raisonné directement

sur le modèle continu. Enfin, s'il fait le lien avec les probabilités de collision, il ne poursuit pas le lien avec la théorie du contrôle optimal.

Dans [SIN 94], les auteurs présentent une architecture basée à la fois sur les fonctions harmoniques et l'apprentissage par renforcement : les fonctions harmoniques produisent des trajectoires "sûres" à partir desquelles des algorithmes d'apprentissage par renforcement calculent des solutions optimales. Bien que l'apprentissage par renforcement soit intrinsèquement lié au contrôle optimal, les auteurs n'identifient pas le lien étroit que nous présentons dans ce papier.

5. Simulations

Cette dernière section présente, à l'aide de simulations, des illustrations de planification par fonction harmonique comme par contrôle optimal, et illustre la relation entre les deux (comme décrit à la section 4.1).

Pour les schémas numériques, nous avons utilisé une discrétisation continue linéaire par morceaux, sur une triangulation régulière de l'espace $\overline{O}$. Pour les fonctions harmoniques, on considère le schéma décrit par exemple dans [CON 93a] ; pour le contrôle optimal, on utilise le schéma aux différences finies présenté dans [MUN 97]. Le pas de discrétisation de l'espace est $ds = 1$, celui du temps est $dt = 0.1$. Pour le contrôle optimal, le pas de discrétisation des actions est $da = \frac{\pi}{8}$, et on a choisi $\gamma = 0.99$. Nous utilisons la méthode itérative de Jacobi pour résoudre le système d'équations qui résulte de la discrétisation.

Nous avons considéré un environnement qui contient un certain nombre de murs et deux buts (voir figure 1) : le but 1 se situe dans le coin haut gauche (on y accède par un couloir relativement étroit), le but 2 se situe dans une zone plus dégagée (dans la partie inférieure gauche). Nous avons représenté, pour 8 valeurs croissantes de σ, les trajectoires moyennes suivies à partir de points de départ équi-répartis. La 9^{eme} et dernière représentation correspond au calcul par fonction harmonique Φ.

Dans le cas où la dynamique est déterministe ($\sigma = 0$), les trajectoires optimisent la longueur du chemin vers les buts ; la plupart des trajectoires partant de la partie haute de l'environnement se dirigent vers le but 1 tandis que toutes les autres vont vers le but 2. Quand on augmente un peu le bruit ($\sigma = 0.4$), les trajectoires s'éloignent légèrement des murs et s'arrondissent légèrement. Avec un tout petit plus de bruit ($\sigma = 0.5$), le couloir amenant au but 1 devient plus risqué et la plupart des trajectoires se dirigent alors vers le but 1. A partir de $\sigma = 1$, les trajectoires sont de plus en plus arrondies, et on commence à voir apparaître des zones de départs qui ne vont plus vers les buts : les chemins à travers le couloir devenant trop dangereux, le système préfère maintenir l'agent dans des zones sûres (loin des obstacles). Ce phénomène s'aggrave significativement pour $\sigma = 5$ et $\sigma = 20$. A partir de $\sigma = 50$, une nouvelle phase commence : le bruit étant trop fort, la stratégie consistant à rester dans des zones sûres n'est même plus efficace (la dérive amène inévitablement vers un mur). Perdu pour perdu, la seule solution (pour optimiser le critère) consiste à reprendre de nouveau

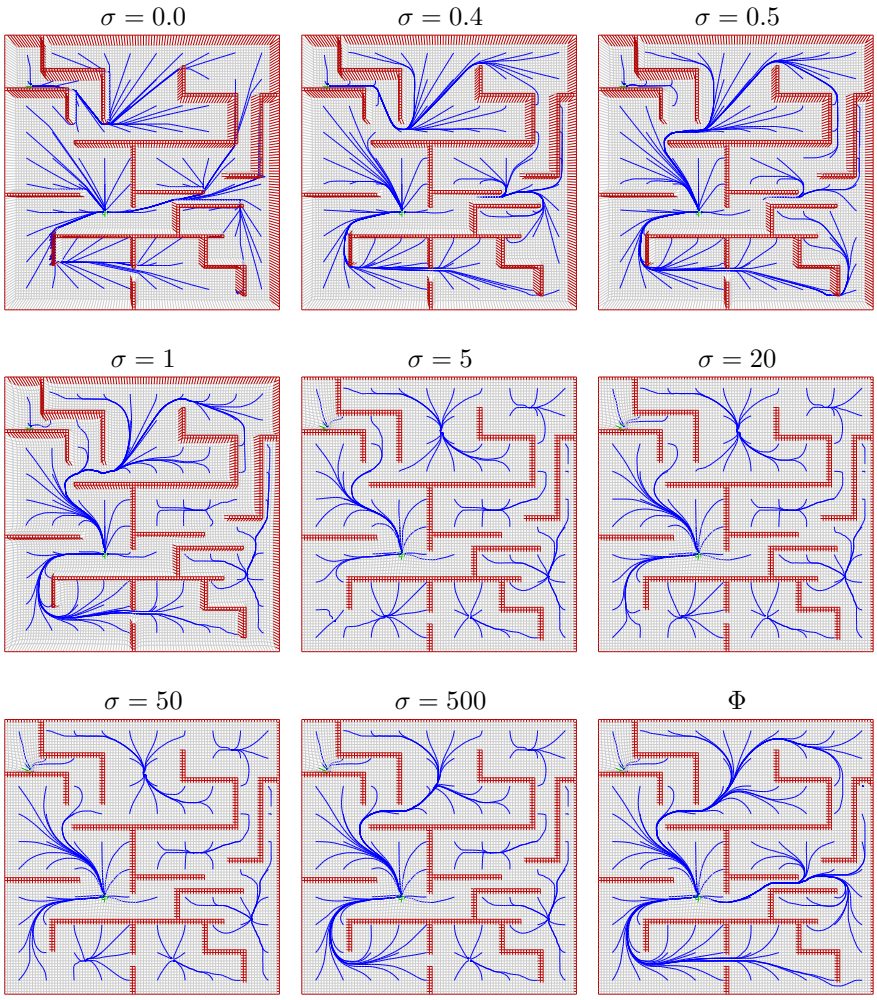


Figure 1. ***Evolution des trajectoires en fonction de*** σ**.** *Voir le texte pour les détails*

des risques, donc à se déplacer de nouveau. Cette deuxième phase prend nettement de l'ampleur quand on augmente encore le bruit ($\sigma = 500$). On observe que cette avant-dernière simulation ($\sigma = 500$) se rapproche maintenant nettement de la limite quand $\sigma \to \infty$ (c'est-à-dire de Φ).

6. Bibliographie

[AKI 90] AKISHITA S., KAWAMURA S., HAYASHI K., « Laplace potential for moving obstacle avoidance and approach of a mobile robot », *Japan-USA Symposium on Flexible Automation, A Pacific Rim Conference*, 1990, p. 139–142.

[BER 83] BERTSEKAS D., « Distributed asynchronous computation of fixed points. », *Math. Programming*, vol. 27, 1983, p. 107–120.

[BER 89] BERTSEKAS D., TSITSIKLIS J. N., *Parallel and Distributed Computation : Numerical Methods*, Prentice hall, 1989.

[CON 90] CONNOLLY C. I., BURNS J. B., WEISS R., « Path Planning Using Laplace's Equation », *The Proceedings of IEEE Internetioal Conference on Robotics and Automation, ICRA'90*, May 1990, p. 2102-2106.

[CON 92] CONNOLLY C. I., BURNS J. B., « The Basal Ganglia and the Planning of Actions », rapport n° 92-18, Feb 1992, COINS Department, University of Massachusetts.

[CON 93a] CONNOLLY C. I., GRUPEN R., « On the Applications of Harmonic Functions to Robotics », *Journal of Robotic and Systems*, vol. 10, n° 7, 1993, p. 931-946.

[CON 93b] CONNOLLY C., BURNS J. B., « A New Striatal Model and its Relationship to Basal Ganglia Diseases », *Neuroscience Research*, vol. 16, 1993, p. 271–274.

[CON 94] CONNOLLY C. I., « Harmonic Functions and Collision Probabilities », *The Proceedings of the IEEE International Conference on Robotics and Automation, ICRA'94*, San Diego, CA, may 1994, IEEE, p. 3015-3019.

[KHA 86] KHATIB O., « Real Time Obstacle Avoidance for Manipulators and Mobile Robots », *The International Journal of Robotics Research*, vol. 5, n° 1, 1986, p. 90-99.

[KIM 92] KIM J.-O., KHOSLA P., « Real-Time Obstacle Avoidance Using Harmonic Potential Functions », *IEEE Transactions on Robotics and Automation*, , 1992.

[MUN 97] MUNOS R., « Apprentissage par renforcement, étude du cas continu », PhD thesis, Ecole des Hautes Etudes en Sciences Sociales, 1997.

[MUN 00] MUNOS R., « A Study of Reinforcement Learning in the Continuous Case by the Means of Viscosity Solutions », *Machine Learning*, vol. 40, n° 3, 2000, p. 265-299.

[MUN 05] MUNOS R., ZIDANI H., « Consistency of a simple multidimensional scheme for HJB equations », *CRAS*, 2005.

[PUT 94] PUTERMAN M., *Markov Decision Processes*, Wiley, New York, 1994.

[SCH 04] SCHERRER B., « Asynchronous Neurocomputing for optimal control and reinforcement learning with large state space », *Neurocomputing*, , To appear, 2004.

[SIN 94] SINGH S., BARTO A., GRUPEN R., CONNOLLY C., « Robust Reinforcement Learning in Motion Planning », *Advances in Neural Information Processing Systems 6 (NIPS)*, 1994, p. 655-662.

[SUT 98] SUTTON R., BARTO A., *Reinforcement Learning, An introduction*, BradFord Book. The MIT Press, 1998.

[TRA 91] TRASSENKO L., BLAKE A., « Analogue Computation of Collision-free Paths », *International Conference on Robotics and Automation*, IEEE, April 1991, p. 540–545.