

THÈSES DE L'UNIVERSITÉ PARIS-SUD (1971-2012)

ISABELLE NAGOT

Grandes déviations pour les processus d'apprentissage lent à statistiques discontinues sur une surface, 1995

Thèse numérisée dans le cadre du programme de numérisation de la bibliothèque mathématique Jacques Hadamard - 2016

Mention de copyright :

Les fichiers des textes intégraux sont téléchargeables à titre individuel par l'utilisateur à des fins de recherche, d'étude ou de formation. Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale.

Toute copie ou impression de ce fichier doit contenir la présente page de garde.



UNIVERSITÉ DE PARIS-SUD
U.F.R. SCIENTIFIQUE D'ORSAY

THÈSE présentée pour obtenir

Le GRADE de DOCTEUR EN SCIENCES
de L'UNIVERSITÉ PARIS XI ORSAY

Spécialité: Mathématiques

par

Isabelle NAGOT

Sujet :

*Grandes déviations
pour les processus d'apprentissage lent
à statistiques discontinues sur une surface*

Soutenue le 16 janvier 1995 devant la commission d'examen

M.	Robert	AZENCOTT
M.	Gérard	BEN AROUS
M.	Halim	DOSS
M ^{me}	Nicole	EL KAROUI
M.	Alain	ROUAULT
M.	Ofer	ZEITOUNI

Au terme de ce parcours, je tiens à remercier toutes les personnes qui m'ont permis de mener à bien mes recherches :

Robert Azencott a dirigé cette thèse avec patience et exigence et je lui suis très reconnaissante pour le soutien qu'il m'a toujours porté ; les conversations que nous avons eues ont été source de beaucoup d'enrichissements,

Gérard Ben Arous a guidé mes premiers pas en probabilités à l'Ecole Normale de la rue d'Ulm ; je suis heureuse qu'il intervienne à nouveau dans mon travail,

Halim Doss et Alain Rouault ont lu attentivement mon manuscrit. Je les remercie à la fois pour leur diligence, leur méticulosité et les améliorations qu'ils ont apportées,

Nicole El Karoui m'a témoigné, à chacune de nos rencontres, la plus grande attention ; j'apprécie son intérêt pour mon travail, au moment où j'aborde des sujets dont elle est spécialiste,

Ofer Zeitouni a accepté spontanément de faire partie de mon jury, à l'occasion de son passage en France, et je lui en suis très reconnaissante.

Je remercie aussi
les différentes équipes, à l'Ecole Normale de la rue d'Ulm, l'Université d'Orsay et l'Ecole Normale de Cachan, qui m'ont accueillie avec la plus grande bienveillance,
Elisabeth et Alain pour leur aide finale,
et plus généralement tous ceux qui m'ont soutenue au cours de ce travail.

Je dédie cette thèse à mes parents.

Grandes déviations pour les processus d'apprentissage lent
à statistiques discontinues sur une surface

ABSTRACT

We prove the large deviations principle for a Markov chain having two regions of smooth statistical behavior separated by a boundary across which the transition mechanics change discontinuously. The hypothesis on the two probability fields on either side of the surface are basically those which allow large deviations for a single field. Such dynamics are used in many stochastic algorithms, particularly for learning in neural nets. Two kinds of behavior are then possible for the process suitably normalized, depending on the configuration of the supports of the probabilities near the boundary. In the first one, the number of boundary crossings during a period is not bounded. A new rate function arises, which is a combination of the Cramer's transform of each field and corresponds to a mixture of the two fields. In the second kind, the process crosses the boundary at most one time and each Cramer's transform is integrated one after another to get the rate function. In the last chapter, we give equations for a minimal-cost path between two points. Their resolution enables accelerated simulations of rare events.

Introduction

Nous établissons le principe de grandes déviations pour des chaînes de Markov dont les probabilités de transition sont définies différemment de part et d'autre d'une surface. Les grandes déviations sont observées sur un intervalle de temps fixe pour le processus correctement normalisé. Sur chaque champ de probabilités, on fait les hypothèses de régularité qui permettent "habituellement" d'obtenir le principe de grandes déviations. Cette régularité est perdue sur la frontière.

Précisément, nous étudions les propriétés asymptotiques de familles de processus obtenus de manière suivante : étant donné un champ de probabilités $(\mu_x)_{x \in \mathbb{R}^d}$ sur $\mathbb{R}^d$, on considère une chaîne de Markov dans $\mathbb{R}^d$, $(Y_k^\tau)_{k \in \mathbb{N}}$, paramétrée par un réel strictement positif, τ , telle que, conditionnellement à $Y_k^\tau = x$, l'accroissement $Y_{k+1}^\tau - Y_k^\tau$ soit de la forme τX , où X suit la loi μ_x . On s'intéresse au comportement de cette chaîne quand τ tend vers 0. Cette dynamique est celle de nombreux algorithmes stochastiques (algorithmes de Robbins-Monroe), et nous appelons, à la suite de R. Azencott et G. Ruget [AR], "processus d'apprentissage lent" la famille de processus $(Y_k^\tau)_{k \in \mathbb{N}}$.

Les résultats asymptotiques s'obtiennent pour k tendant vers $+\infty$. Ils peuvent, de façon équivalente, se lire, cette fois pendant une durée finie T , sur le processus re-

normalisé de la façon suivante: appelons γ^τ le chemin aléatoire défini sur $[0, T]$ par

$$\gamma^\tau(t) = Y_k^\tau \text{ si } t \in [k\tau; (k+1)\tau[\quad (0.1)$$

Les modifications éventuelles de la valeur de γ^τ sont espacées du laps de temps τ et proportionnelles à τ . Pour disposer d'un processus continu, on introduit également φ^τ , le chemin aléatoire défini sur $[0, T]$ par :

$$\varphi^\tau(k\tau) = Y_k^\tau \text{ pour } 0 \leq k\tau < T + \tau \text{ et } \varphi^\tau \text{ est affine sur } [k\tau; (k+1)\tau] \quad (0.2)$$

On connaît le comportement moyen du processus quand τ tend vers 0. Notons $m_x = \int y d\mu_x(y)$ la vitesse moyenne adoptée quand le processus se trouve en x . Alors, quand τ tend vers 0, la loi du processus φ^τ , pour $Y_0^\tau = y_0$, tend vers celle du processus déterministe φ solution de

$$\begin{cases} \varphi'(t) = m_{\varphi(t)} \\ \varphi(0) = y_0 \end{cases}$$

On peut également approcher φ^τ par une diffusion. Mais l'effet cumulé des aléas peut, même pour τ petit, amener le processus à suivre des trajectoires arbitraires, phénomène qui n'est pas bien décrit par les deux premières approximations. Les probabilités de ces événements rares sont données par le principe de grandes déviations. Etudier les propriétés de grandes déviations du processus quand τ tend vers 0 revient à étudier l'existence d'une fonctionnelle Λ définie sur l'ensemble des chemins continus de $[0, T]$ dans $\mathbb{R}^d$, $C[0, T]$, à valeurs dans $[0, +\infty]$, telle que, si φ appartient à $C[0, T]$, $\Lambda(\varphi)$ soit approximativement égal, pour τ petit, à $-\tau \log P(\varphi^\tau \sim \varphi)$.

De façon classique (voir par exemple [FW]), l'utilisation de telles estimations permet l'obtention d'expressions asymptotiques de différentes quantités telles que temps et lieux de sortie d'ensembles, probabilités de transition entre voisinages de points stables, mesures invariantes... Une application possible est alors l'étude asymptotique d'algorithmes stochastiques. Ainsi, la prise en compte de discontinuités du champ (μ_x), que nous proposons ici, permet d'étudier des algorithmes pour lesquels l'espace est découpé en plusieurs régions correspondant à des dynamiques différentes, comme par exemple certains algorithmes d'apprentissage pour les réseaux de neurones (citons l'apprentissage compétitif, exposé dans [RZ]). Une suite de ce travail pourrait être d'appliquer les résultats obtenus à ces algorithmes et par exemple d'étudier comment le mode de présentation d'exemples pendant l'apprentissage influe sur la probabilité de passage d'un attracteur à un autre.

Précisons maintenant les estimations étudiées. On dit que le principe de grandes déviations est vérifié si, pour toute partie borélienne G de $C[0, T]$, on a :

$$-\inf_{\varphi \in \overset{\circ}{G}} \Lambda(\varphi) \leq \liminf_{\tau \rightarrow 0} \tau \log P(\varphi^\tau \in G) \leq \limsup_{\tau \rightarrow 0} \tau \log P(\varphi^\tau \in G) \leq -\inf_{\varphi \in \overline{G}} \Lambda(\varphi) \quad (0.3)$$

Ce résultat s'obtient classiquement ([Az],[FW],...) à partir des propriétés suivantes :

- (i) pour tout $I \geq 0$, $\Gamma(I) = \{\varphi \mid \Lambda(\varphi) \leq I\}$ est compact
- (ii) pour tous $I, \varepsilon > 0$, $\limsup_{\tau \rightarrow 0} \tau \log P(d_{0,T}(\varphi^\tau, \Gamma(I)) \geq \varepsilon) \leq -I$
- (iii) pour tous $I, \varepsilon > 0$, $\varphi \in C[0, T]$, $\liminf_{\tau \rightarrow 0} \tau \log P(d_{0,T}(\varphi^\tau, \varphi) < \varepsilon) \geq -\Lambda(\varphi)$,

où $d_{0,T}$ désigne la distance entre les chemins : pour φ_1 et φ_2 chemins de $[0, T]$ dans $\mathbb{R}^d$, on pose

$$d_{0,T}(\varphi_1, \varphi_2) = \sup_{t \in [0, T]} |\varphi_1(t) - \varphi_2(t)|$$

et, pour G ensemble de tels chemins, $d_{0,T}(\varphi_1, G) = \inf_{\varphi_2 \in G} d_{0,T}(\varphi_1, \varphi_2)$.

En général, la fonctionnelle est s.c.i. et la propriété (i) est vérifiée dès que l'ensemble $\{\varphi \mid \Lambda(\varphi) < +\infty\}$ est compact. Il reste alors à établir les propriétés (ii) et (iii), respectivement une majoration et une minoration. D'autre part, dans ce cas, le résultat (0.3) implique :

$$\lim_{\varepsilon \rightarrow 0} \liminf_{\tau \rightarrow 0} \tau \log P(\varphi^\tau_{/[0, T]} \in B(\varphi, \varepsilon) | \mathcal{F}_0) = -\Lambda_{0,T}(\varphi).$$

Ainsi, plus $\Lambda_{0,T}(\varphi)$ est élevé, plus la probabilité que le processus suive le chemin φ est faible. Dans notre étude, nous appelons "fonctionnelle de coût" la fonctionnelle Λ (on dit souvent fonctionnelle d'action). Notons que les conditions initiales restent à préciser ci-dessus. On peut par exemple avoir $Y_0^\tau = y_0$ presque sûrement, on se limite alors dans les résultats à des chemins φ tels que $\varphi(0) = y_0$ (en particulier, on pose $\Gamma(I) = \{\varphi \mid \varphi(0) = y_0 \text{ et } \Lambda(\varphi) \leq I\}$), ou bien, de manière équivalente, on pose $\Lambda(\varphi) = +\infty$ si $\varphi(0)$ n'est pas égal à y_0 .

R. Azencott et G. Ruget dans [AR] (1977) introduisent la notion de processus d'apprentissage lent et démontrent ce résultat de grandes déviations, dans un cadre plus général (sur une variété riemannienne), dans le cas où le champ de probabilités (μ_x) vérifie certaines conditions de régularité ; en particulier ils supposent que l'application $x \mapsto \mu_x$ est continue et, pour certains résultats, que les lignes de niveau de sa transformée de Cramer sont localement lipschitziennes.

P. Dupuis dans [Du88] montre, pour un processus plus général, que l'on peut affaiblir les hypothèses pour écrire les estimations de majoration, en se passant du caractère lipschitzien.

Nous étendons ces résultats au cas où les propriétés de régularité du champ de probabilités (μ_x) sont perdues le long d'une surface. Les résultats de P. Dupuis dans [Du88] et de P. Dupuis, R.S. Ellis et A. Weiss dans [DEW] (1991) donnent des estimations de grandes déviations (une majoration seulement pour les seconds) qui s'appliquent dans ce cas, mais qui ne conduisent pas à un résultat optimal. P. Dupuis et R. Ellis dans [DE] (1992) font cette hypothèse de discontinuité le long d'un hyperplan mais supposent que les champs sont constants de chaque côté de l'hyperplan et énoncent un principe de grandes déviations pour la distribution de $\frac{Y_n^1}{n}$ pour n grand, c'est-à-dire concernant la position moyenne atteinte par la chaîne de Markov -pour τ fixé égal à 1- après un grand nombre de pas et non la valeur de toute la trajectoire. Ils supposent de plus que la position initiale Y_0^τ appartient à l'hyperplan où sont localisées les discontinuités, ce qui rend la loi de $\frac{1}{\tau}(Y_n^\tau - Y_0^\tau)$ indépendante de τ . Nous généralisons leurs résultats au cas où Y_0^τ n'appartient pas nécessairement à l'hyperplan et où les champs ne sont plus constants, puis nous passons des estimations sur la vitesse moyenne aux résultats sur toute la trajectoire.

En conclusion, nous reprenons le cadre d'étude de [AR], pour une chaîne de Markov, en tenant compte de l'affaiblissement possible des hypothèses indiqué par [Du88]; nous supposons que les propriétés de régularité du champ de probabilités (μ_x) sont perdues le long d'une surface, de la même façon que dans [DE] et nous prolongeons [DE] aux propriétés de grandes déviations concernant les trajectoires complètes $(\gamma^\tau)_{\tau>0}$, dans le cas de champs non constants, en gardant un coût optimal comme dans [DE] (le même pour la majoration et la minoration, ou presque...). Nous étudions donc des "processus d'apprentissage lent à statistiques discontinues sur une surface". En fait, l'étude du cas où la surface est un hyperplan permet, grâce à des méthodes de localisation (que nous ne détaillons pas), de déduire les résultats dans le cas d'une surface régulière quelconque. La terminologie processus "à statistiques discontinues" est utilisée dans les travaux de P. Dupuis et R. Ellis pour désigner les processus de Markov dont le générateur dépend de la position de façon non nécessairement continue.

Les méthodes employées s'inspirent essentiellement de celles de [AR] et [DE]. En particulier, [DE] fournit les méthodes permettant d'obtenir des estimations locales (les résultats sur les vitesses moyennes) et [AR] la façon de les intégrer le long des

trajectoires. Notons que nous supposons, comme dans [AR] que les mesures sont à supports bornés. Cette hypothèse rend possible un contrôle fin du processus au voisinage de la frontière. En particulier, elle permet d'écrire le lemme de stabilité 3.1.4 pour montrer l'inégalité de minoration.

La démonstration de l'inégalité de majoration est faite, comme dans [AR], en construisant une trajectoire régularisée à partir des vitesses moyennes. Cependant, la présence de discontinuités ne permet plus de contrôler le coût exact de cette trajectoire. Aussi nous employons, à la différence de [AR], une méthode basée sur l'utilisation d'un coût approché, mieux adaptée à l'étude de discontinuités. Ce coût approché prend en compte en même temps le fait que les champs de probabilités ne sont pas constants, ce qui permet de considérer une trajectoire simplement affine par morceaux (dans [AR], on prend la courbe intégrale d'un champ de vecteurs construit à partir des vitesses moyennes). Cette méthode permet de plus de ne pas faire d'hypothèse de type lipschitzien sur les lignes de niveau de la transformée de Cramer. Pour l'inégalité de minoration, on travaille d'abord avec des tubes d'axe régulier au voisinage de la frontière. On approche ensuite la plupart des chemins continus par de tels chemins réguliers. Les résultats obtenus généralisent ceux de [AR] et de [DE].

Dans le chapitre 1, nous posons précisément le problème, en faisant le bilan des méthodes à notre disposition. Les estimations de majoration et de minoration sont traitées respectivement dans le chapitre 2 et le chapitre 3. Nous observons deux modes de comportement du processus en fonction de la configuration des supports des mesures au voisinage de la frontière.

Le premier mode est obtenu sous l'hypothèse que chaque champ de probabilités associé à un côté de la surface charge toujours l'ensemble des vecteurs qui permettent de passer de l'autre côté de cette surface. Alors, le nombre de traversées de la frontière sur un intervalle de temps de longueur donnée n'est pas borné. Une nouvelle fonction de coût, L_x , apparaît pour x sur la frontière, due à ce mélange des deux champs ; c'est une combinaison des transformées de Cramer des deux champs. Notons que sur les vecteurs parallèles à la frontière, ce coût est inférieur à chacune des transformées de Cramer et que le processus peut rester au voisinage de la frontière alors qu'il la quitterait s'il était soumis à un seul des deux champs. Les fonctionnelles de coût sont calculées sur les chemins φ tels qu'il y a un nombre fini d'intervalles sur lesquels soit φ suit la frontière, soit il est d'un même côté de cette frontière.

Dans le second mode, la frontière n'induit aucun phénomène de ce type. Les estimations font intervenir une fonctionnelle calculée sur les chemins qui traversent au

plus une fois la frontière et qui est une somme des fonctionnelles de Cramer dans chaque demi-espace.

Dans le chapitre 4, nous rassemblons les résultats obtenus dans chacune des deux configurations des supports envisagées. Nous traitons les cas limites où les supports sont tangents à la frontière, puis nous nous intéressons aux chemins optimaux entre deux points. Nous écrivons les équations vérifiées par ces chemins et par le coût qui leur est associé. Nous n'insistons pas sur la résolution de ces équations ; elle peut servir à la mise au point d'algorithmes de simulation rapide d'événements rares, par un changement de probabilité le long de la trajectoire minimisante (voir par exemple [CFM] ou [DK]).

Chapitre 1

Généralités : présentation du problème, outils

Dans ce premier chapitre, nous rappelons la définition des processus d'apprentissage lent proposée par [AR] en présentant les travaux qui permettent d'obtenir des résultats de grandes déviations pour ces processus, notamment les articles de R. Azencott et G. Ruget [AR], et de P. Dupuis [Du88]. La comparaison de ces travaux nous conduit à formuler les hypothèses qui constitueront le cadre de notre étude, en particulier deux hypothèses sur les variations du logarithme de la transformée de Laplace du champ de probabilités dont nous montrons l'équivalence respectivement à la continuité et au caractère lipschitzien des lignes de niveau de la transformée de Cramer du champ de probabilités.

Après la présentation de quelques éléments classiques d'analyse convexe qui nous permettront de faire intervenir des fonctions de coût approchées dans notre étude, nous introduisons dans le paragraphe 1.4 la notion de processus d'apprentissage lent à statistiques discontinues le long d'une surface, qui correspond à un processus d'apprentissage lent pour lequel on relâche les hypothèses habituelles de continuité faites sur le champ de probabilités. Nous proposons, à la suite de P. Dupuis et R.S. Ellis [DE] dans le cas particulier de champs constants, la fonction de coût qui nous permet d'écrire les résultats de grandes déviations pour un tel processus dans les chapitres 2 et 3.

Deux modes de comportement du processus seront distingués dans ces chapitres, en fonction de la configuration, au voisinage de la frontière, des supports des probabilités par rapport à la direction de cette frontière. Une nouvelle fonction de coût, L_x , due à la juxtaposition des deux champs, apparaît, dans un seul des deux modes, qui correspond au cas où chaque champ de probabilités associé à un côté de la

surface charge toujours l'ensemble des vecteurs qui permettent de passer de l'autre côté de cette surface. Cette fonction de coût est une combinaison des transformées de Cramer des deux champs. Nous étudions ses propriétés, ainsi que celles de la fonctionnelle obtenue en intégrant L_x le long des chemins.

Enfin, dans le paragraphe 1.5, nous posons précisément le problème étudié et regroupons les hypothèses faites sur le modèle.

1.1 Deux types de résultats sur les processus d'apprentissage lents

1.1.1 Notations, rappel des résultats de [AR]

On appelle champ de probabilités sur $\mathbb{R}^d$ toute application associant à chaque x de $\mathbb{R}^d$ une probabilité μ_x sur $\mathbb{R}^d$. Etant donné un tel champ $(\mu_x)_{x \in \mathbb{R}^d}$, ^{faiblement continu} On sait ([AR]) qu'on peut construire un espace probabilisé $(\Omega_1, \mathcal{B}_1, P_1)$ et un champ de vecteurs aléatoire X défini sur $(\Omega_1, \mathcal{B}_1, P_1)$ tel que pour tout x de $\mathbb{R}^d$, X_x soit de loi μ_x .

Donnons nous alors $(\Omega, \mathcal{B}, P)$ un espace probabilisé, $(\mathcal{F}_k)_{k \in \mathbb{N}}$ une suite croissante de sous- σ -algèbres de $\mathcal{B}$ et $(X_k(\cdot))_{k \in \mathbb{N}}$, une suite de champs de vecteurs aléatoires indépendants tels que pour tout x de $\mathbb{R}^d$ et tout k de $\mathbb{N}$, $X_k(x)$ soit de loi μ_x et $\mathcal{F}_{k+1}$ -mesurable.

Définition 1.1.1 *On dit que la famille de processus aléatoires $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$ indexée par τ est un processus d'apprentissage lent associé au champ de probabilités (μ_x) si, pour tout $\tau > 0$, $(Y_k^\tau)_{k \in \mathbb{N}}$ vérifie la relation de récurrence :*

$$Y_{k+1}^\tau = Y_k^\tau + \tau X_k(Y_k^\tau) \quad (1.1)$$

(dans [AR], on se place sur une variété riemannienne et cette égalité est vérifiée à un terme du second ordre en τ près).

Dans cette partie, nous rappelons le résultat de grandes déviations concernant un processus d'apprentissage lent $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$ établi par [AR]. On utilise les définitions suivantes :

Notation 1.1.2 *Pour A inclus dans $\mathbb{R}^d$ et ε strictement positif, on pose :*

$$A^\varepsilon = \{x \in \mathbb{R}^d \mid \exists y \in A, |x - y| \leq \varepsilon\} \text{ et } A^{-\varepsilon} = \{x \in A \mid \forall y \notin A, |x - y| > \varepsilon\}.$$

Définition 1.1.3 La distance de Prokhorov entre deux probabilités sur $\mathbb{R}^d$, μ et ν , est telle que, pour tout $\varepsilon > 0$, $d(\mu, \nu)$ est inférieur à ε si et seulement si pour tout borélien A de $\mathbb{R}^d$, on a $\mu(A) \leq \nu(A^\varepsilon) + \varepsilon$ et $\nu(A) \leq \mu(A^\varepsilon) + \varepsilon$.

Définition 1.1.4 Soit φ un chemin absolument continu, de $[0, T]$ dans $\mathbb{R}^d$. On dit que φ est tangent au champ de convexes $(C_x)_{x \in \mathbb{R}^d}$ si, pour presque tout t de $[0, T]$, $\varphi'(t)$ appartient à $C_{\varphi(t)}$.

On suppose que le champ de probabilités (μ_x) possède les propriétés suivantes :
Hypothèses (H) :

- (H1) l'application $x \mapsto \mu_x$ de $\mathbb{R}^d$ dans l'espace des probabilités sur $\mathbb{R}^d$, muni de la distance de Prokhorov, est continue.
- (H2) l'enveloppe convexe fermée du support de μ_x , que l'on note S_x , est compacte et d'intérieur non vide.
- (H3) l'application $x \mapsto S_x$, de $\mathbb{R}^d$ dans l'ensemble des compacts de $\mathbb{R}^d$ muni de la distance de Hausdorff, est continue.
- (H4) pour tout $T > 0$ et tout compact $\mathcal{K}$, il existe un compact contenant tous les $\varphi([0, T])$, dès que φ est un chemin tangent au champ de convexes $(S_x)_{x \in \mathbb{R}^d}$ tel que $\varphi(0)$ appartient à $\mathcal{K}$.

L'hypothèse (H4) permet d'éviter des complications liées à d'éventuels temps d'explosion de trajectoires tangentes au champ de convexes $(S_x)_{x \in \mathbb{R}^d}$. Pour qu'elle soit vérifiée, il suffit d'avoir la propriété suivante : pour tout x_0 de $\mathbb{R}^d$, si on note $C(n)$ la quantité $\sup \{|v| \mid v \in \bigcup_{x \in B(x_0, n)} S_x\}$, alors $\sum_{n=1}^{+\infty} \frac{1}{C(n)}$ vaut $+\infty$.

Notation 1.1.5 Pour x dans $\mathbb{R}^d$, on note m_x le barycentre de la mesure μ_x , H_x le logarithme de sa transformée de Laplace et L_x sa transformée de Cramer.

Précisément, pour α et β dans $\mathbb{R}^d$, on a :

$$\begin{cases} m_x = \int_{\mathbb{R}^d} y \mu_x(dy) \\ H_x(\alpha) = \log \int_{\mathbb{R}^d} e^{\langle \alpha, y \rangle} \mu_x(dy) \\ L_x(\beta) = \sup_{\alpha \in \mathbb{R}^d} \{\langle \alpha, \beta \rangle - H_x(\alpha)\}. \end{cases}$$

Prenons x dans $\mathbb{R}^d$. L'application H_x est convexe, s.c.i. et, d'après la compacité de S_x , finie sur $\mathbb{R}^d$. L'application L_x est convexe et s.c.i.; elle est nulle au barycentre de la mesure, positive dans l'intérieur de S_x et vaut $+\infty$ sur $(S_x)^c$. D'autre part, les hypothèses (H) entraînent que l'application $(x, \beta) \mapsto L_x(\beta)$ est s.c.i., en effet, si on note L_μ la transformée de Cramer de la probabilité μ sur $\mathbb{R}^d$, l'application $(\mu, \beta) \mapsto L_\mu(\beta)$ est s.c.i., en munissant l'ensemble des probabilités sur $\mathbb{R}^d$ de la convergence faible (qui peut être définie par la distance de Prokhorov, cf [AR]), l'hypothèse (H1) permet alors de conclure.

Certains résultats nécessitent une hypothèse supplémentaire, (H'), énoncée ci-dessous. Rappelons que la distance de Hausdorff h entre deux parties A et B de $\mathbb{R}^d$ est définie par :

$$h(A, B) = \inf\{a > 0, A \subset B^a, B \subset A^a\}$$

Définition 1.1.6 Pour L_1 et L_2 deux fonctions sur $\mathbb{R}^d$ à valeurs dans $\mathbb{R}^+$, on appelle "distance entre les lignes de niveau de L_1 et L_2 " le nombre

$$D(L_1, L_2) = \sup_{I \geq 0} h(\{L_1(\cdot) \leq I\}, \{L_2(\cdot) \leq I\}).$$

On vérifie facilement que D est une distance sur l'ensemble des fonctions sur $\mathbb{R}^d$ à valeurs dans $\mathbb{R}^+$.

Définition 1.1.7 On dit que "les lignes de niveau de L_x sont localement lipschitziennes en x sur $\mathbb{R}^d$ " si, pour tout compact $\mathcal{K}$ inclus dans $\mathbb{R}^d$, il existe $C_{\mathcal{K}}$ tel que, quels que soient x, y dans $\mathcal{K}$, on ait $D(L_x, L_y) \leq C_{\mathcal{K}}|x - y|$.

On appelle (H') l'hypothèse que les lignes de niveau de la transformée de Cramer du champ de probabilités sont localement lipschitziennes en x sur $\mathbb{R}^d$, c'est-à-dire que la fonction $x \mapsto L_x$ vérifie la définition précédente.

Soit T un réel strictement positif. On appelle φ^τ le chemin aléatoire continu défini sur $[0, T]$ par :

$$\varphi^\tau(k\tau) = Y_k^\tau \text{ pour } 0 \leq k\tau < T + \tau \text{ et } \varphi^\tau \text{ est affine sur } [k\tau; (k+1)\tau]. \quad (1.2)$$

Soit y_0 un point de $\mathbb{R}^d$; on suppose que Y_0^τ est égal à y_0 , P -presque sûrement. On note $C_{y_0}[0, T]$ l'ensemble des chemins continus φ de $[0, T]$ dans $\mathbb{R}^d$ tels que $\varphi(0) = y_0$ et on le munit de la topologie de la convergence uniforme sur $[0, T]$.

Notation 1.1.8 Pour φ dans $C_{y_0}[0, T]$, on pose

$$\Lambda_{0,T}(\varphi) = \begin{cases} \int_0^T L_{\varphi(t)}(\varphi'(t))dt & \text{si } \varphi \text{ est absolument continu,} \\ +\infty & \text{sinon} \end{cases}$$

$\Lambda_{0,T}$ est la fonctionnelle de Cramer associée au champ de probabilités $(\mu_x)_{x \in \mathbb{R}^d}$. Elle est à valeurs positives et s'annule en un chemin φ si et seulement si presque sûrement en t , $\varphi'(t)$ est égal à $m_{\varphi(t)}$, c'est-à-dire si φ est une courbe intégrale du champ moyen (m_x) .

Définition 1.1.9 On définit une fonctionnelle de Cramer régularisée par :

$$\text{pour } \varphi \in C_{y_0}[0, T], \quad \Lambda_{0,T}^{\text{rég}}(\varphi) = \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \text{ à vitesses intérieures}}} \Lambda_{0,T}(\varphi),$$

où un chemin à vitesses intérieures est un élément ψ de $C_{y_0}[0, T]$, C^1 par morceaux, tel que pour tout t de $[0, T[$, la dérivée à droite de ψ en t appartient à $S_{\psi(t)}^\circ$.

Enfin, tout au long de l'étude, on utilise la notation suivante :

Notation 1.1.10 Si f est une application et A une partie de l'ensemble de définition de f , alors $f(A)$ désigne la borne inférieure de f sur A .

On a le théorème suivant (Azencott/Ruget) :

Considérons $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$ un processus d'apprentissage lent associé au champ de probabilités (μ_x) vérifiant les hypothèses (H). Pour toute partie borélienne G de $C_{y_0}[0, T]$, on a :

1. $-\Lambda^{\text{rég}}(\overset{\circ}{G}) \leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi^\tau \in G)$
2. si l'hypothèse (H') est vérifiée, on a de plus $\Lambda^{\text{rég}} = \Lambda$ et

$$\limsup_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi^\tau \in G) \leq -\Lambda(\overline{G})$$

Quand (H') est vérifiée, la majoration et la minoration s'écrivent avec la même fonction de coût. Pour φ absolument continu, la probabilité qu'une trajectoire du processus issue de $\varphi(0)$ reste, pendant le temps T , uniformément voisine de φ est alors à peu près égale à $\exp(-\frac{1}{\tau} \Lambda_{0,T}(\varphi)) = \exp(-\frac{1}{\tau} \int_0^T L_{\varphi(t)}(\varphi'(t)) dt)$. La fonctionnelle Λ correspond à un "coût" : plus $\Lambda(G)$ est élevé, plus la probabilité que φ^τ appartienne à G est faible.

On a, sous les hypothèses (H) et (H') : pour toute partie borélienne G de $C_{y_0}[0, T]$,

$$-\Lambda(\overset{\circ}{G}) \leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi^\tau \in G) \leq \limsup_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi^\tau \in G) \leq -\Lambda(\overline{G}) \quad (1.3)$$

Exemple: cas de probabilités à supports finis

Considérons le champ de probabilités

$$\mu_x = \sum_{k=1}^K p_k(x) \delta_{a_k(x)}$$

avec, pour x donné, les $p_k(x)$ positifs et de somme 1. Les applications a_k , de $\mathbb{R}^d$ dans $\mathbb{R}^d$, et p_k , de $\mathbb{R}^d$ dans $[0, 1]$, sont supposées continues. Montrons qu'alors (H1) est vérifiée :

pour $\varepsilon > 0$, il existe δ tel que $|x - y| \leq \delta$ et $x, y \in \mathcal{K}$ impliquent, pour chaque k , $|a_k(x) - a_k(y)| \leq \varepsilon$. Alors, pour A borélien de $\mathbb{R}^d$, on a $\mu_x(A) = \sum_{k|a_k(x) \in A} p_k(x)$ donc :

$$\mu_y(A^\varepsilon) \geq \sum_{k|a_k(x) \in A} p_k(y) \geq \mu_x(A) + \varepsilon.$$

(H3) est vérifiée dès que les $p_k(x)$ ne s'annulent pas. Les $\{a_k(x)\}$ doivent engendrer $\mathbb{R}^d$ pour que (H2) soit vraie et il reste une condition à imposer pour obtenir (H4) (par exemple, pour tout x_0 de $\mathbb{R}^d$, si $C(n) = \sup\{|a_k(x)| \mid 1 \leq k \leq K, x \in B(x_0, n)\}$, on a $\sum_{n=1}^{+\infty} \frac{1}{C(n)} = +\infty$). Si, de plus, les applications a_k et p_k sont localement lipschitziennes, alors (H') est vérifiée (voir aussi à la fin du paragraphe 1.2.2) et on a le résultat (1.3). Nous donnons dans le paragraphe 1.4.3 la valeur de la transformée de Cramer pour un tel champ de probabilités.

1.1.2 Rappel des résultats de [Du88]

Dans [Du88], P. Dupuis étudie des propriétés de grandes déviations dans le cas plus général d'un processus à valeurs dans $\mathbb{R}^d$ vérifiant $Y_{k+1}^\tau = Y_k^\tau + \tau b_k(Y_k^\tau, \xi_k)$, où ξ_k est un processus aléatoire et $b_k(\cdot, \cdot)$ une suite de champs de vecteurs aléatoires identiquement distribués. Ses résultats s'appliquent à un processus d'apprentissage lent, comme défini dans le paragraphe précédent, si l'on suppose que b_k ne dépend que de x , avec $P(b_k(x) \in A) = \mu_x(A)$.

Par rapport à l'étude rappelée précédemment, il montre que l'on peut affaiblir les hypothèses et en particulier se passer du caractère lipschitzien pour la majoration (hypothèse (H')) du paragraphe 1.1.1). D'autre part l'application $b_k(\cdot, \cdot)$ peut être discontinue en x , les résultats sont donc applicables dans le cas où le champ de probabilités est discontinu le long d'une surface. Nous commencerons ainsi l'étude de ce cas (paragraphe 1.4.4).

Indiquons brièvement comment l'étude de [Du88] s'applique à un processus d'apprentissage lent. En plus d'hypothèses générales (A1) (qui se réduisent ici à " $|b_k(\cdot)| \leq K$ presque sûrement"), deux hypothèses fondamentales sont utilisées, (A2) pour la

majoration et (A3) pour la minoration, hypothèses qui ne portent que sur certains moments exponentiels (donc sur les trajectoires et non plus directement sur les données du problème; en particulier, il n'y a pas d'hypothèses de régularité de $b_k(\cdot)$ autre que (A2) et (A3)). L'hypothèse (A2) consiste à supposer qu'il existe une fonction réelle $H(x, \alpha)$, convexe en α , vérifiant :

pour tous x dans $\mathbb{R}^d$ et $\gamma > 0$, il existe $\delta > 0$ tel que, pour toute suite (y_k) de $B(x, \delta)$ (boule de centre x et rayon δ) et pour tout α de $\mathbb{R}^d$, on ait :

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{E} \left(\exp \left\langle \alpha, \sum_{k=1}^n X_k(y_k) \right\rangle \right) \leq H(x, \alpha) + \gamma(1 + |\alpha|)$$

tandis que l'hypothèse (A3) suppose qu'il existe une fonction réelle $H(x, \alpha)$ à valeurs dans $\mathbb{R}$, convexe et de classe C^1 en α , vérifiant :

pour tout compact $\mathcal{K}$ de $\mathbb{R}^d$, il existe $C_{\mathcal{K}}$ tel que, pour tous x dans $\mathcal{K}$, $\delta > 0$, (y_k) suite de $B(x, \delta)$ et α dans $\mathbb{R}^d$, on ait :

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{E} \left(\exp \left\langle \alpha, \sum_{k=1}^n X_k(y_k) \right\rangle \right) \leq H(x, \alpha) + C_{\mathcal{K}} \delta (1 + |\alpha|)$$

et

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{E} \left(\exp \left\langle \alpha, \sum_{k=1}^n X_k(y_k) \right\rangle \right) \geq H(x, \alpha) - C_{\mathcal{K}} \delta (1 + |\alpha|)$$

Avec ces hypothèses, on a les résultats classiques de grandes déviations (analogues à (1.3)), avec comme fonctionnelle de coût Λ telle que $\Lambda(\varphi) = \int (H_{\varphi(t)})^*(\varphi'(t)) dt$ si φ est absolument continu, $+\infty$ sinon.

L'inégalité de majoration ($\limsup_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi^\tau \in G) \leq -\Lambda(\overline{G})$, pour toute partie borélienne G de $C_{y_0}[0, T]$) est obtenue dès que les hypothèses (A1) et (A2) sont vérifiées; pour l'inégalité de minoration ($-\Lambda(\overset{\circ}{G}) \leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi^\tau \in G)$), on a besoin, en plus, de l'hypothèse (A3).

1.2 Comparaison de ces résultats, discussion des hypothèses

Dans cette partie, on revient, à la suite de [Du88], sur la comparaison des hypothèses de [AR] et [Du88] afin de faire apparaître celles qui seront adaptées dans le cas où le champ de probabilités est discontinu le long d'une surface. On considère à nouveau un processus d'apprentissage lent (définition 1.1.1), avec les notations 1.1.5.

Ce processus peut être traité selon chacune des deux études exposées précédemment.

On suppose dans toute cette partie que, pour tout x , la log-Laplace H_x est finie sur $\mathbb{R}^d$. On met alors en évidence (paragraphe 1.2.1) deux hypothèses fondamentales, (HC) et (HLip), proches, pour ce processus, des hypothèses (A2) et (A3) de [Du88], et de formulation simple. Ces hypothèses seront celles utilisées dans toute l'étude. Dans le paragraphe 1.2.2, nous faisons le lien entre les hypothèses (HC) et (HLip) et certaines hypothèses sur $x \mapsto L_x$. Nous montrons que (HLip) est équivalente à (H') dans [AR] et que (HC) a, comme (HLip), une formulation sur les lignes de niveau de la transformée de Cramer, le caractère lipschitzien étant remplacé par une continuité simple (théorème 1.2.3). Dans le paragraphe 1.2.3, nous précisons le lien entre (HC) et les hypothèses (H) de [AR] : nous montrons que l'hypothèse (HC) implique (H3) et certaines conséquences de (H1) dans [AR], dont nous avons besoin pour adapter les méthodes de [AR] au cas où le champ est discontinu le long d'un hyperplan, en particulier que l'application $(x, \beta) \mapsto L_x(\beta)$ est s.c.i.. Précisément, nous obtenons (théorème 1.2.8) que l'hypothèse (HC) est équivalente aux deux hypothèses (H3) et $L(\cdot)$ continue sur $\bigcup_x x \times \overset{\circ}{S}_x$.

1.2.1 Formulation de deux hypothèses sur les variations en x de la log-Laplace H_x

Le moment exponentiel apparaissant dans les hypothèses (A2) et (A3) de [Du88] s'écrit :

$$\mathbb{E} \left(\exp \left\langle \alpha, \sum_{k=1}^n X_k(y_k) \right\rangle \right) = \exp \sum_{k=1}^n H_{y_k}(\alpha)$$

puisque

$$\exp \left[\left\langle \alpha, \sum_{k=1}^n X_k(y_k) \right\rangle - \sum_{k=1}^n H_{y_k}(\alpha) \right]$$

est une martingale de moyenne 1. L'hypothèse (A2) est donc vérifiée dès que pour tous x dans $\mathbb{R}^d$ et $\gamma > 0$, il existe $\delta > 0$ tel que, pour tout α de $\mathbb{R}^d$,

$$|H_y(\alpha) - H_x(\alpha)| \leq \gamma |\alpha| \quad \text{si } |y - x| \leq \delta,$$

et (A3) est vérifiée si, pour tout compact $\mathcal{K}$ de $\mathbb{R}^d$, il existe $C_{\mathcal{K}}$ tel que, pour tous x, y dans $\mathcal{K}$, $\delta > 0$ et α dans $\mathbb{R}^d$, on ait :

$$|H_y(\alpha) - H_x(\alpha)| \leq C_{\mathcal{K}} |y - x| |\alpha|.$$

Le logarithme de la transformée de Laplace de μ_x , H_x , joue alors le rôle de la fonction réelle (sous l'hypothèse de compacité des S_x), convexe et de classe C^1 cherchée.

Ainsi, (A2) et (A3) sont vraies dès que les variations de $H(\alpha)$ vérifient les hypothèses suivantes :

Hypothèse (HC) :

$\forall x \in \mathbb{R}^d, \forall \gamma > 0, \exists \delta > 0$ tel que, si $|x - y| \leq \delta, \forall \alpha \in \mathbb{R}^d, |H_x(\alpha) - H_y(\alpha)| \leq \gamma|\alpha|$

Hypothèse (HLip) :

$\forall \mathcal{K}$ compact de $\mathbb{R}^d, \exists C_{\mathcal{K}}$ tel que $\forall x, y \in \mathcal{K}, \forall \alpha \in \mathbb{R}^d, |H_x(\alpha) - H_y(\alpha)| \leq C_{\mathcal{K}}|x - y||\alpha|$

En utilisant [Du88], on peut donc retrouver les résultats (1.3) en se contentant de l'hypothèse (HC) pour l'inégalité de majoration et de (HLip) pour l'inégalité de minoration (en plus de l'hypothèse que les supports des mesures sont bornés), c'est-à-dire qu'on n'a pas besoin du caractère lipschitzien des lignes de niveau de la transformée de Cramer pour obtenir l'inégalité de majoration des grandes déviations. On n'a pas besoin non plus de l'hypothèse " S_x d'intérieur non vide" ; en revanche l'hypothèse de compacité des supports, en plus de (HC), implique que pour tout x dans un compact, on a $|b_k(x)| \leq K$, presque sûrement, ce qui permet d'assurer (A1) dans un compact contenant toutes les trajectoires du processus quand (H4) est vérifiée.

Ce sont ces hypothèses (HC) et (HLip) que nous allons utiliser dans le cas où le champ est discontinu le long d'un hyperplan ; là encore, le caractère lipschitzien (l'hypothèse (HLip)) ne sera nécessaire que pour l'inégalité de minoration des grandes déviations. Les résultats de notre étude engloberont ceux de [AR] et de [Du88] pour un processus d'apprentissage lent en l'absence de discontinuité du champ de probabilités.

1.2.2 Lien entre la distance entre les lignes de niveau de la transformée de Cramer L_x et les variations de la log-Laplace H_x

Rappelons que si f est une fonction convexe sur $\mathbb{R}^d$ à valeurs dans $] -\infty, +\infty]$, on appelle *conjuguée de f* , et on note f^* , la fonction de $\mathbb{R}^d$ dans $\mathbb{R}$ définie par

$$f^*(\beta) = \sup_{\alpha \in \mathbb{R}^d} \{ \langle \alpha, \beta \rangle - f(\alpha) \}.$$

Définition 1.2.1 On dit qu'une fonction convexe, f , de $\mathbb{R}^d$ dans $] -\infty, +\infty]$, est propre si $f(x)$ est fini pour au moins un x .

Si f est convexe, propre et s.c.i., on a :

$$(f^*)^* = f$$

et on dit que f^* est la transformée de Legendre de f ; f et f^* sont duales dans la transformation de Legendre.

En particulier, par définition, pour x donné, la transformée de Cramer L_x et la log-Laplace H_x vérifient $L_x = (H_x)^*$ et elles sont duales dans la transformation de Legendre dès que H_x est finie sur $\mathbb{R}^d$ (car H_x est alors convexe et finie, donc s.c.i.), ce que nous supposons dans cette partie.

Rappelons le résultat suivant ([DE]) :

Si f est convexe et s.c.i. sur $\mathbb{R}^d$, les propriétés suivantes sont équivalentes :

- (i) $\inf_{x, |x|=c} \frac{f(x)}{c}$ tend vers $+\infty$ quand c tend vers $+\infty$
- (ii) f^* est finie sur $\mathbb{R}^d$

On en déduit le lemme :

Lemme 1.2.2

Si f est convexe, s.c.i. sur $\mathbb{R}^d$ et que f^* est finie sur $\mathbb{R}^d$, alors, pour tout I positif, l'ensemble $\{f(\cdot) \leq I\}$ est compact.

D'autre part, on a défini ci-dessus (définition 1.1.6) la distance D entre deux fonctions réelles positives sur $\mathbb{R}^d$ comme la distance de Hausdorff maximale entre leurs lignes de niveau. On a alors les équivalences suivantes :

Théorème 1.2.3

Soit (μ_x) un champ de probabilités sur $\mathbb{R}^d$. On suppose que pour tout x , le logarithme de la transformée de Laplace, H_x , est fini sur $\mathbb{R}^d$.

Alors, l'hypothèse (HC) est équivalente à :

$\forall x \in \mathbb{R}^d, \forall \gamma > 0, \exists \delta > 0$ tel que, pour $|x - y| \leq \delta$, $D(L_x, L_y) \leq \gamma$,
où L_x est la transformée de Cramer de μ_x ce qui revient à dire que
l'application $x \mapsto L_x$ est continue pour la distance D ,

et l'hypothèse (HLip) est équivalente à :

$\forall \mathcal{K}$ compact de $\mathbb{R}^d, \exists C_{\mathcal{K}}$ tel que $\forall x, y \in \mathcal{K}$, $D(L_x, L_y) \leq C_{\mathcal{K}}|x - y|$,
ce qui revient à dire que $x \mapsto L_x$ est localement lipschitzienne.

Démonstration : on commence par montrer le lemme suivant :

Lemme 1.2.4

Soient H_1 et H_2 des fonctions convexes sur $\mathbb{R}^d$, à valeurs dans $[0, +\infty[$ et telles que $H_1(0) = H_2(0) = 0$. On pose $L_i = (H_i)^*$. Alors on a :

$$\sup_{\alpha \in \mathbb{R}^d} \frac{|H_1(\alpha) - H_2(\alpha)|}{|\alpha|} = D(L_1, L_2)$$

Démonstration :

1. on utilise le lemme suivant (adapté du lemme 2.4 de [Du88]) :

Soient H_1 et H_2 des fonctions convexes sur $\mathbb{R}^d$. On suppose qu'il existe $\gamma > 0$ tel que pour tout α , $H_2(\alpha) \leq H_1(\alpha) + \gamma|\alpha|$, et on pose $L_i = (H_i)^*$. Alors, pour tout I dans $[-H_1(0), +\infty[$, on a :

$$\{L_2(\cdot) \leq I\} \subset \{L_1(\cdot) \leq I\}^\gamma$$

On en déduit que, si pour tout α de $\mathbb{R}^d$, $|H_1(\alpha) - H_2(\alpha)|$ est inférieur à $\gamma|\alpha|$, alors, pour tout I positif, $\{L_2(\cdot) \leq I\}$ est inclus dans $\{L_1(\cdot) \leq I\}^\gamma$ et $\{L_1(\cdot) \leq I\}$ dans $\{L_2(\cdot) \leq I\}^\gamma$.

Alors $D(L_1, L_2) = \sup_{I \geq 0} h(\{L_1(\cdot) \leq I\}, \{L_2(\cdot) \leq I\})$ est inférieur à γ . On a donc

$$D(L_1, L_2) \leq \sup_{\alpha \in \mathbb{R}^d} \frac{|H_1(\alpha) - H_2(\alpha)|}{|\alpha|}.$$

2. on utilise pour montrer l'inégalité inverse le fait que les H_i sont s.c.i. (puisque convexes et finies partout) donc qu'on a $H_i = (L_i)^*$.

Supposons $D(L_1, L_2) \leq \gamma$ et considérons α dans $\mathbb{R}^d$. Comme $H_2(\alpha)$ est fini, il existe β' tel que $H_2(\alpha) = \langle \alpha, \beta' \rangle - L_2(\beta')$; or, pour tout I positif, $\{L_2 \leq I\}$ est inclus dans $\{L_1 \leq I\}^\gamma$. On utilise cette propriété pour $I = L_2(\beta')$. Le vecteur β' se trouve dans les deux ensembles correspondants; il existe donc β tel que $|\beta - \beta'| \leq \gamma$ et $L_1(\beta) \leq L_2(\beta')$. On a alors :

$$H_1(\alpha) \geq \langle \alpha, \beta \rangle - L_1(\beta) \geq \langle \alpha, \beta \rangle - L_2(\beta') = \langle \alpha, \beta - \beta' \rangle + H_2(\alpha) \geq H_2(\alpha) - \gamma|\alpha|.$$

D'après la symétrie en H_1 et H_2 , on a donc $|H_1(\alpha) - H_2(\alpha)| \leq \gamma|\alpha|$.

Pour démontrer le théorème 1.2.3, il suffit, pour x, y dans $\mathbb{R}^d$, d'appliquer le lemme 1.2.4 à H_x et H_y qui en vérifient bien les hypothèses. $\square$

A priori, il sera plus facile de vérifier les propriétés (HC) et (HLip) sur les H_x , donc avec leur première formulation. Explicitons toutefois ce que signifie $D(L_1, L_2) \leq \gamma$:

Lemme 1.2.5

Pour β dans $\mathbb{R}^d$ et γ dans $\mathbb{R}^+$, on note $B(\beta, \gamma)$ la boule fermée de centre β et rayon γ . Alors,

$$D(L_1, L_2) \leq \gamma \text{ est équivalent à : } \forall \beta, \begin{cases} \inf_{B(\beta, \gamma)} L_2 \leq L_1(\beta) \\ \inf_{B(\beta, \gamma)} L_1 \leq L_2(\beta) \end{cases}$$

Démonstration : $\{L_1(\cdot) \leq I\} \subset \{L_2(\cdot) \leq I\}^\gamma$ implique que pour tout β de $\mathbb{R}^d$ tel que $L_1(\beta)$ soit fini, il existe β' tel que $|\beta - \beta'| \leq \gamma$ et $L_2(\beta') \leq L_1(\beta)$. On a donc $\inf_{B(\beta, \gamma)} L_2 \leq L_1(\beta)$, et cela reste vrai pour $L_1(\beta) = +\infty$. La réciproque est claire. $\square$

Remarque : on peut aussi montrer que $D(L_1, L_2) \leq \gamma$ implique $L_1(\beta) \leq \sup_{B(\beta, \gamma)} L_2$,

pour tout β (cf lemme 2.4 de [Du88]).

En effet, l'inclusion $\{L_2(\cdot) \leq I\} \subset \{L_1(\cdot) \leq I\}^\gamma$ entraîne $\{L_2(\cdot) \leq I\}^{-\gamma} \subset \{L_1(\cdot) \leq I\}$ (car $\{L_1(\cdot) \leq I\}$ est fermé et convexe), qui est équivalente à $\{L_1(\cdot) \geq I\} \subset \{L_2(\cdot) \geq I\}^\gamma$. Alors, si $\{L_2(\cdot) \leq I\}$ est inclus dans $\{L_1(\cdot) \leq I\}^\gamma$, pour $L_1(\beta) < +\infty$, β appartient à $\{L_2(\cdot) \geq L_1(\beta)\}^\gamma$, donc il existe β' tel que $|\beta - \beta'| \leq \gamma$ et $L_1(\beta) \leq L_2(\beta')$. On a donc $L_1(\beta) \leq \sup_{B(\beta, \gamma)} L_2$.

Si $L_1(\beta) = +\infty$, pour tout I positif, β appartient à $\{L_1(\cdot) \geq I\}$, donc à $\{L_2(\cdot) \leq I\}^\gamma$. Alors $\sup_{B(\beta, \gamma)} L_2 \geq I$. Donc $L_1(\beta) \leq \sup_{B(\beta, \gamma)} L_2$ est encore vrai.

On peut en déduire diverses formulations des hypothèses (HC) et (HLip). On a en particulier :

Théorème 1.2.6

Soit (μ_x) un champ de probabilités sur $\mathbb{R}^d$. On suppose que pour tout x , H_x est finie sur $\mathbb{R}^d$. Alors,

l'hypothèse (HC) est équivalente à : $\forall \mathcal{K}$ compact de $\mathbb{R}^d$, $\forall \gamma, \exists \delta$ tel que

$$\forall x, y \in \mathcal{K}, |x - y| \leq \delta \text{ implique } \forall \beta, \inf_{B(\beta, \gamma)} L_y \leq L_x(\beta)$$

et l'hypothèse (HLip) est équivalente à :

$$\forall \mathcal{K} \text{ compact de } \mathbb{R}^d, \exists C_{\mathcal{K}} \text{ tel que } \forall x, y \in \mathcal{K}, \forall \beta, \exists \beta', \begin{cases} |\beta - \beta'| \leq C_{\mathcal{K}} |x - y| \\ L_y(\beta') \leq L_x(\beta) \end{cases}$$

Démonstration: pour x et y dans $\mathcal{K}$, on utilise le lemme 1.2.5 pour obtenir l'équivalence de

$$“ D(L_x, L_y) \leq \gamma ” \quad \text{et de} \quad “ \forall \beta, \inf_{B(\beta, \gamma)} L_y \leq L_x(\beta) ”$$

L'énoncé équivalent à (HC) correspond à la continuité uniforme de $x \mapsto L_x$ sur tout compact. $\square$

1.2.3 Liens de (HC) avec les hypothèses (H) et avec certaines de leurs conséquences

Théorème 1.2.7

Soit (μ_x) un champ de probabilités sur $\mathbb{R}^d$. On suppose que pour tout x , le support de μ_x est borné (ou S_x compact). Alors l'hypothèse (HC) implique :

- (a) (H3) : l'application $x \mapsto S_x$ de $\mathbb{R}^d$, dans l'ensemble des compacts de $\mathbb{R}^d$ muni de la distance de Hausdorff, est continue.
- (b) l'application $(x, \beta) \mapsto L_x(\beta)$ est s.c.i..

Remarque: pour (b), l'hypothèse “ H_x finie sur $\mathbb{R}^d$, pour tout x ” suffit.

Démonstration: on suppose (HC) vérifiée;

(a) Soit $\varepsilon > 0$ et x dans $\mathbb{R}^d$. On prend δ tel que, pour $|x - y| \leq \delta$, on ait $D(L_x, L_y) \leq \frac{\varepsilon}{2}$. Le convexe S_x engendre un sous-espace F de $\mathbb{R}^d$. On prend I égal à la borne supérieure de L_x sur $A = \{\beta \in S_x \mid \forall \beta' \in F \setminus S_x, |\beta - \beta'| > \frac{\varepsilon}{2}\}$ (A correspond à $(S_x)^{-\frac{\varepsilon}{2}}$ mais par rapport à l'espace F). De $\{L_x(\cdot) \leq I\} \subset \{L_y(\cdot) \leq I\}^{\frac{\varepsilon}{2}}$, on déduit que A est inclus dans $(S_y)^{\frac{\varepsilon}{2}}$ et donc que S_x est inclus dans $(S_y)^\varepsilon$. De même, $\{L_x(\cdot) \leq I\} \subset \{L_y(\cdot) \leq I\}^{\frac{\varepsilon}{2}}$ implique $S_y \subset (S_x)^\varepsilon$. On en déduit la continuité cherchée puisque $|x - y| \leq \delta$ implique $h(S_x, S_y) \leq \varepsilon$.

(b) Soit (x, β) dans $\mathbb{R}^d \times \mathbb{R}^d$, et (x_n, β_n) une suite de $\mathbb{R}^d \times \mathbb{R}^d$ convergeant vers (x, β) quand n tend vers $+\infty$. Montrons $L_x(\beta) \leq \liminf_{n \rightarrow +\infty} L_{x_n}(\beta_n)$.

D'après (HC) et le théorème 1.2.3, pour tout $\varepsilon > 0$, il existe $\delta_\varepsilon > 0$ tel que

$|x - y| \leq \delta_\varepsilon$ implique, pour tout I positif, $\{L_y(\cdot) \leq I\} \subset \{L_x(\cdot) \leq I\}^\gamma$. On prend (ε_n) convergeant vers 0 telle que $|x - x_n| \leq \delta_{\varepsilon_n}$. Alors $\{L_{x_n}(\cdot) \leq L_{x_n}(\beta_n)\}$ est inclus dans $\{L_x(\cdot) \leq L_{x_n}(\beta_n)\}^{\varepsilon_n}$. Le premier ensemble contient β_n , donc il existe β'_n tel que $|\beta_n - \beta'_n| \leq \varepsilon_n$ et $L_x(\beta'_n) \leq L_{x_n}(\beta_n)$. Quand n tend vers $+\infty$, β'_n converge vers β ; or, L_x est s.c.i. (propriété de la transformée de Cramer, vraie même si S_x est d'intérieur

vide), on a donc :

$$L_x(\beta) \leq \liminf_{n \rightarrow +\infty} L_x(\beta'_n) \leq \liminf_{n \rightarrow +\infty} L_{x_n}(\beta_n). \quad \square$$

Remarquons qu'en remplaçant γ par $C_K|x-y|$ dans la démonstration de (a) on peut montrer que l'hypothèse (HLip) entraîne que l'application $x \mapsto S_x$ est localement lipschitzienne.

Théorème 1.2.8

Soit (μ_x) un champ de probabilités sur $\mathbb{R}^d$. On suppose que, pour tout x , S_x est compact et d'intérieur non vide (ie (H2)). Alors,

l'hypothèse (HC) est équivalente à $\begin{cases} \text{(H3)} \\ (x, \beta) \mapsto L_x(\beta) \text{ continue sur } \bigcup_x x \times \mathring{S}_x \end{cases}$

Démonstration :

1. On a vu dans le théorème 1.2.7 que (HC) implique (H3). Montrons que sous l'hypothèse supplémentaire que les supports sont d'intérieur non vide, (HC) entraîne la continuité de $(x, \beta) \mapsto L_x(\beta)$ sur $\bigcup_x x \times \mathring{S}_x$.

Soit x dans $\mathbb{R}^d$, β dans $\mathring{S}_x$ et $\varepsilon > 0$. L'application L_x est continue sur $\mathring{S}_x$ (cf [AR]). On prend δ tel que, pour $|\beta' - \beta| \leq 2\delta$, $|L_x(\beta') - L_x(\beta)| \leq \varepsilon$. On a alors :

$$\sup_{B(\beta, 2\delta)} L_x \leq L_x(\beta) + \varepsilon \quad \text{et} \quad L_x(\beta) - \varepsilon \leq \inf_{B(\beta, 2\delta)} L_x.$$

D'après (HC), il existe δ' tel que, pour $|y - x| \leq \delta'$ et pour tout β' dans $\mathbb{R}^d$,

$$\inf_{B(\beta', \delta)} L_x \leq L_y(\beta') \leq \sup_{B(\beta', \delta)} L_x$$

(d'après le lemme 1.2.5 et la remarque qui le suit). Alors, pour y et β' tels que $|y - x| \leq \delta'$ et $|\beta' - \beta| \leq \delta$, on a :

$$L_y(\beta') \leq \sup_{B(\beta', \delta)} L_x \leq \sup_{B(\beta, 2\delta)} L_x \leq L_x(\beta) + \varepsilon$$

et

$$L_y(\beta') \geq \inf_{B(\beta', \delta)} L_x \geq \inf_{B(\beta, 2\delta)} L_x \geq L_x(\beta) - \varepsilon.$$

On obtient bien le résultat puisque $|y - x| \leq \delta'$ et $|\beta' - \beta| \leq \delta$ impliquent $|L_y(\beta') - L_x(\beta)| \leq \varepsilon$.

2. Pour établir la réciproque, on montre tout d'abord que l'hypothèse

$$“(x, \beta) \mapsto L_x(\beta) \text{ continue sur } \bigcup_x x \times \overset{\circ}{S}_x”$$

implique la continuité de $x \mapsto m_x$, où m_x est le barycentre de la mesure μ_x (les supports sont toujours supposés d'intérieur non vide). Considérons x dans $\mathbb{R}^d$ et $\varepsilon > 0$. L'application L_x est positive sur $B(m_x, \varepsilon)^c$. On appelle δ son minimum sur cet ensemble; il est strictement positif. On prend ε_1 tel que, pour $|x - y| \leq \varepsilon_1$ et pour tout β dans $\overset{\circ}{S}_x$, $|L_x(\beta) - L_y(\beta)|$ soit inférieur à δ . Considérons alors β dans $\overset{\circ}{S}_x$ et n'appartenant pas à $B(m_x, \varepsilon)$. On a, pour $|x - y| \leq \varepsilon_1$:

$$L_y(\beta) \geq L_x(\beta) + \delta \geq 2\delta$$

donc m_y appartient à $B(m_x, \varepsilon)$.

Pour terminer la démonstration du théorème 1.2.8, on a besoin d'un résultat sur une transformation affine (homothétie) des supports, transformation qui sera utilisée plusieurs fois au cours de l'étude et qui a été introduite par [AR].

Définition 1.2.9 Pour $\theta \in [0, 1]$ et $x \in \mathbb{R}^d$, on définit l'application affine :

$$\begin{aligned} h_{x,\theta} : \mathbb{R}^d &\longrightarrow \mathbb{R}^d \\ v &\longmapsto \theta m_x + (1 - \theta)v \end{aligned}$$

D'après [AR] (4.6), pour tout compact $\mathcal{U}$ dans $\mathbb{R}^d$ et tout $\delta > 0$, il existe des constantes strictement positives C_1 et C_2 telles que, pour tous θ de $[0, 1]$, $\delta_1 \geq 0$ et y dans $\mathcal{U}$, on ait :

$$C_1\theta \geq C_2\delta_1 \Rightarrow h_{y,\theta}((S_y)^{\delta_1}) \subset (S_y)^{-(C_1\theta - C_2\delta_1)}. \quad (1.4)$$

[AR] énonce ce résultat sous l'hypothèse supplémentaire (H') sur les lignes de niveau de la transformée de Cramer. En fait, la démonstration n'utilise que la continuité des applications $x \mapsto S_x$ et $x \mapsto m_x$ et le fait que les S_x engendrent $\mathbb{R}^d$. Nous pouvons donc énoncer le lemme plus général suivant :

Lemme 1.2.10

On considère $x \mapsto C_x$ une application de $\mathbb{R}^d$ dans l'ensemble des convexes d'intérieur non vide de $\mathbb{R}^d$, continue pour la distance de Hausdorff sur cet ensemble et $x \mapsto b_x$ une application continue de $\mathbb{R}^d$ dans $\mathbb{R}^d$, telle que, pour tout x , b_x est intérieur à C_x .

Pour $\theta \in [0, 1]$, et $x \in \mathbb{R}^d$, on définit l'application affine

$$\begin{aligned} g_{x,\theta} : \mathbb{R}^d &\rightarrow \mathbb{R}^d \\ v &\mapsto \theta b_x + (1 - \theta)v \end{aligned}$$

Alors, pour tout compact $\mathcal{U}$, il existe des constantes strictement positives C_1 et C_2 telles que, pour tous θ de $[0, 1]$, $\delta_1 \geq 0$ et x dans $\mathcal{U}$, on ait :

$$C_1\theta \geq C_2\delta_1 \Rightarrow g_{x,\theta}((C_x)^{\delta_1}) \subset (C_x)^{-(C_1\theta - C_2\delta_1)}.$$

Démonstration : Pour x dans $\mathcal{U}$, C_x contient une boule de $\mathbb{R}^d$ centrée en b_x , de rayon ρ_x . Il existe δ tel que pour $|x - y| < \delta$, on ait $|b_x - b_y| < \frac{\rho_x}{3}$ et $h(C_x, C_y) < \frac{\rho_x}{3}$. Alors $|x - y| < \delta$ implique $B(b_y, \frac{\rho_x}{3}) \subset B(b_x, \frac{2\rho_x}{3}) \subset (C_x)^{-\frac{\rho_x}{3}} \subset C_y$ (d'après la convexité de C_x pour la dernière inclusion). Il suffit de recouvrir le compact $\mathcal{U}$ par les boules $B(x_i, \delta)$ et de prendre $\rho = \inf \rho_{x_i}$ pour en déduire que, pour tout y dans $\mathcal{U}$, C_y contient $B(b_y, \rho)$.

La fin de la démonstration est exactement celle de [AR](4.6) : on montre qu'il existe α strictement positif tel que, pour tout x de $\mathcal{U}$, pour tout point extrémal v de C_x , C_x contient un cône de révolution de sommet v , et d'angle au sommet α (la base de ce cône est la boule $B(b_x, \rho)$). On en déduit le résultat. $\square$

Reprenons la démonstration du théorème 1.2.8 et montrons que les deux hypothèses (H3) et " $(x, \beta) \mapsto L_x(\beta)$ continue sur $\bigcup_x x \times \overset{\circ}{S}_x$ " entraînent la formulation de (HC) du théorème 1.2.6.

Considérons un compact $\mathcal{K}$ et $\gamma > 0$. Montrons qu'il existe $\delta > 0$ tel que, pour x et y dans $\mathcal{K}$, avec $|x - y|$ inférieur à δ , on a, pour tout β ,

$$\inf_{B(\beta, \gamma)} L_y \leq L_x(\beta). \quad (1.5)$$

Prenons δ_1 tel que pour tous x, y dans $\mathcal{K}$, $|x - y| \leq \delta_1$ implique $|m_x - m_y| \leq \frac{\gamma}{2}$.

D'après (H3) et le fait que les S_x engendrent $\mathbb{R}^d$, quitte à diminuer γ , pour

tout x de $\mathcal{K}$, la boule ouverte $B(m_x, \frac{\gamma}{2})$ est incluse dans l'ensemble $(S_x)^{-\delta_0}$, avec δ_0 ne dépendant que de $\mathcal{K}$ (comme dans le lemme 1.2.10).

Le compact $(S_x)^{-\delta_0} \cap B(m_x, \frac{\gamma}{2})^c$ est inclus dans $\overset{\circ}{S}_x$. D'après la continuité de L sur $\bigcup_x x \times \overset{\circ}{S}_x$, L atteint son minimum sur $\bigcup_{x \in \mathcal{K}} x \times [(S_x)^{-\delta_0} \cap B(m_x, \frac{\gamma}{2})^c]$. On le note ε , il est non nul. D'après la convexité de L_x , on a :

$$\varepsilon = \inf \left\{ L_x(v) \mid x \in \mathcal{K}, v \in B\left(m_x, \frac{\gamma}{2}\right)^c \right\}.$$

Pour β, x dans $\mathbb{R}^d$ et θ dans $[0, 1]$, on pose $\beta_{x,\theta} = h_{x,\theta}(\beta)$. D'après (1.4), pour tout x de $\mathcal{K}$, $\beta_{x,\theta}$ appartient à $(S_x)^{-\delta_\theta}$, avec δ_θ indép de x . D'après la continuité de L sur $\bigcup_{x \in \mathcal{K}} x \times (S_x)^{-\delta_\theta}$, il existe δ'_θ tel que, pour tous x, y dans $\mathcal{K}$, $|x - y| \leq \delta'_\theta$ implique $|L_x(v) - L_y(v)| \leq \theta\varepsilon$ pour tout v de $(S_x)^{-\delta_\theta}$.

Fixons x dans $\mathcal{K}$. On va montrer (1.5) pour tout β . On veillera à ce que le δ obtenu soit indépendant du point x choisi. Il suffit de considérer le cas $L_x(\beta) < +\infty$.

Pour tout β dans $B(m_x, \frac{\gamma}{2})$ et pour tout y dans $B(x, \delta_1)$, m_y appartient à $B(\beta, \gamma)$. Alors (1.5) est vérifié.

On considère β n'appartenant pas à $B(m_x, \frac{\gamma}{2})$. D'après la convexité de L_x , on a $L_x(\beta_{x,\theta}) \leq (1 - \theta)L_x(\beta)$. Donc $L_x(\beta) - L_x(\beta_{x,\theta}) \geq \theta L_x(\beta) \geq \theta\varepsilon$. Le vecteur $\beta_{x,\theta} - \beta$ est égal à $\theta(m_x - \beta)$. On pose $M = \max \{|\beta| \mid \beta \in S_x, x \in \mathcal{K}\}$ et on prend θ tel que $2M\theta \leq \gamma$, alors le vecteur $\beta_{x,\theta}$ appartient à $B(\beta, \gamma)$. Puisque θ est indépendant de x et β , δ'_θ l'est également. Pour x et y dans $\mathcal{K}$, $|x - y| \leq \delta'_\theta$ implique $|L_x(\beta_{x,\theta}) - L_y(\beta_{x,\theta})| \leq \theta\varepsilon$. On a alors $L_y(\beta_{x,\theta}) - L_x(\beta_{x,\theta}) \leq L_x(\beta) - L_x(\beta_{x,\theta})$, donc :

$$L_y(\beta_{x,\theta}) \leq L_x(\beta).$$

(1.5) est donc vérifié pour tout β . La démonstration du théorème 1.2.8 est achevée. $\square$

Les conséquences de (HC) seront suffisantes pour obtenir l'inégalité de majoration du principe de grandes déviations dans le cas de champs de probabilités discontinus le long d'un hyperplan. C'est donc cette hypothèse, pour chacun des deux champs de probabilités considérés, en plus de (H2) également pour chaque champ et d'une hypothèse analogue à (H4), pour l'ensemble des deux champs, qui sera faite dans

le chapitre 2. L'ensemble de ces hypothèses sera appelé (Hmaj).

En revanche, pour l'inégalité de minoration, on aura besoin de deux propriétés plus fortes :

$$(x, \beta) \mapsto \nabla L_x(\beta) \text{ est continue sur } \bigcup x \times \overset{\circ}{S}_x$$

$$\forall \alpha \in \mathbb{R}^d, \forall 1 \leq k \leq d, x \mapsto \frac{\partial^2 H_x}{\partial (\alpha_k)^2}(\alpha) \text{ est continue}$$

où a_k désigne la k^e coordonnée du vecteur α (rappelons que sous (H2), L_x est analytique sur $\overset{\circ}{S}_x$ et H_x sur $\mathbb{R}^d$). Montrons que ces propriétés sont, dès que l'hypothèse (H2) est vérifiée, des conséquences de l'hypothèse (H1): " l'application $x \mapsto \mu_x$ de $\mathbb{R}^d$ dans l'espace des probabilités sur $\mathbb{R}^d$, muni de la distance de Prokhorov, est continue "

Théorème 1.2.11

Soit (μ_x) un champ de probabilités. On suppose (H1) et (H2) vérifiées. Alors :

- (i) $(x, \beta) \mapsto \nabla L_x(\beta)$ est continue sur $\bigcup x \times \overset{\circ}{S}_x$
- (ii) $\forall \alpha \in \mathbb{R}^d, \forall 1 \leq k \leq d, x \mapsto \frac{\partial^2 H_x}{\partial (\alpha_k)^2}(\alpha)$ est continue.

Démonstration: si on pose $\hat{\mu}_x(\alpha) = e^{H_x(\alpha)}$, pour β dans $\overset{\circ}{S}_x$, $\nabla L_x(\beta)$ est l'unique solution de $\frac{d}{d\alpha} \hat{\mu}_x(\cdot) = \hat{\mu}_x(\cdot)\beta$. La propriété (i) vient de la régularité de la solution de cette équation (cf [AR]). La propriété (ii) sur la matrice hessienne de la log-Laplace est immédiate sachant qu'on a :

$$\frac{\partial^2 H_x}{\partial \alpha_i \partial \alpha_j}(\alpha) = \frac{\int y_i y_j e^{(\alpha, y)} \mu_x(dy)}{\int e^{(\alpha, y)} \mu_x(dy)} - \frac{\int y_i e^{(\alpha, y)} \mu_x(dy) \int y_j e^{(\alpha, y)} \mu_x(dy)}{\left[\int e^{(\alpha, y)} \mu_x(dy) \right]^2}. \quad \square$$

Pour obtenir l'inégalité de minoration du principe de grandes déviations, nous ferons dans le chapitre 3 les hypothèses (H2), (HC), (i) et (ii) pour chacun des deux champs de probabilités et une hypothèse globale, analogue à (H4). L'ensemble de ces hypothèses, pour chacun des deux champs de probabilités considérés, sera appelé (Hmin) et est impliqué par les hypothèses (H) de [AR], sauf pour l'hypothèse du type de (H4).

Nous récapitulons, à la fin de ce chapitre, les hypothèses utilisées dans la suite.

Exemple: cas de probabilités à supports finis

Reprenons le champ de probabilités $\mu_x = \sum_{k=1}^K p_k(x) \delta_{a_k(x)}$. On a :

$$H_x(\alpha) = \sum_{k=1}^K p_k(x) e^{\langle \alpha, a_k(x) \rangle}.$$

Si les applications a_k et p_k sont continues, respectivement localement lipschitziennes, alors (HC), respectivement (HLip), est vérifiée. Il est possible que les hypothèses (H) soient équivalentes aux hypothèses (Hmin) dans ce cas (si l'ensemble d'hypothèses (H1),(H3) est équivalent à (HC), dès que (H2) est vraie).

1.3 Outils d'analyse convexe : convexifiée d'un ensemble d'applications

Après ceux du début du paragraphe 1.2.2, rappelons d'autres éléments d'analyse convexe (voir [Roc]) :

Définition 1.3.1 *On appelle enveloppe convexe, ou convexifiée, d'un ensemble d'applications $\{f^i\}$ de $\mathbb{R}^d$ dans $]-\infty, +\infty]$ la plus grande fonction convexe f partout inférieure à chacune des f^i .*

On sait ([Roc], 5) que cette fonction se calcule de la façon suivante :

$$(conv\{f^i\})(\beta) = \inf \left\{ \sum_i \rho^i f^i(\beta^i) \mid \sum_i \rho^i \beta^i = \beta, \rho^i \geq 0, \sum_i \rho^i = 1 \right\} \quad (1.6)$$

où on considère des combinaisons linéaires finies (on peut d'ailleurs se limiter à des combinaisons de $d+1$ éléments d'après le théorème de Carathéodory).

Théorème 1.3.2 ([Roc], 16.5)

La fonction conjuguée de la convexifiée d'un ensemble d'applications propres convexes $f^i, i \in \mathcal{J}$ est égale au sup des fonctions conjuguées des applications :

$$(conv\{f^i \mid i \in \mathcal{J}\})^* = \sup\{(f^i)^* \mid i \in \mathcal{J}\}.$$

Théorème 1.3.3

Si les f^i sont convexes, s.c.i., avec $(f^i)^$ finie sur $\mathbb{R}^d$ et l'ensemble $\mathcal{J}$ fini, on a :*

$$conv\{f^i \mid i \in \mathcal{J}\} = (\sup\{(f^i)^* \mid i \in \mathcal{J}\})^*$$

Démonstration : d'après le théorème 1.3.2, il suffit de montrer que $\text{conv}\{f^i | i \in \mathcal{J}\}$ est s.c.i. ; on aura alors

$$(\text{conv}\{f^i | i \in \mathcal{J}\})^{**} = \text{conv}\{f^i | i \in \mathcal{J}\}.$$

On considère β dans $\mathbb{R}^d$ et (β_n) convergeant vers β . On veut montrer

$$(\text{conv}\{f^i | i \in \mathcal{J}\})(\beta) \leq \liminf_{n \rightarrow +\infty} (\text{conv}\{f^i | i \in \mathcal{J}\})(\beta_n).$$

On peut supposer $\liminf_{n \rightarrow +\infty} (\text{conv}\{f^i | i \in \mathcal{J}\})(\beta_n)$ finie. Soit $\varepsilon > 0$. Pour n entier, prenons (ρ_n^i, β_n^i) tel que (cf (1.6))

$$\sum_i \rho_n^i \beta_n^i = \beta_n, \quad \rho_n^i \geq 0, \quad \sum_i \rho_n^i = 1 \quad \text{et} \quad \sum_i \rho_n^i f^i(\beta_n^i) \leq (\text{conv}\{f^i | i \in \mathcal{J}\})(\beta_n) + \varepsilon$$

A priori, les β_n^i ne varient pas dans un compact. Mais on se ramène à ce cas en utilisant le lemme 1.2.2 qui prouve que pour tout I positif, l'ensemble $\{f^i(\cdot) \leq I\}$ est compact. Pour chaque i de $\mathcal{J}$, (ρ_n^i) admet une valeur d'adhérence ρ^i . Si ρ^i est non nul, $\liminf_{n \rightarrow +\infty} f^i(\beta_n^i)$ est fini. Il existe donc I positif et une sous-suite (β_k^i) qui reste dans le compact $\{f^i(\cdot) \leq I\}$; elle admet alors une valeur d'adhérence β^i . On a nécessairement $\sum_{i|\rho^i \neq 0} \rho^i \beta^i = \beta$ et, en utilisant le fait que chaque f^i est s.c.i. :

$$(\text{conv}\{f^i | i \in \mathcal{J}\})(\beta) \leq \sum_{i|\rho^i \neq 0} \rho^i f^i(\beta^i) \leq \liminf_{n \rightarrow +\infty} (\text{conv}\{f^i | i \in \mathcal{J}\})(\beta_n). \quad \square$$

Remarquons que ce théorème peut aussi être démontré sans utiliser le théorème 1.3.2 (cf [DEW]) :

Soient x et β dans $\mathbb{R}^d$. On a pour tout α de $\mathbb{R}^d$, $\sup\{(f_i)^*\}(\alpha) = \sup_{\substack{\rho_i \geq 0 \\ \sum_i \rho_i = 1}} \sum_i \rho_i (f_i)^*(\alpha)$,

donc

$$\left(\sup_i \{(f_i)^*\} \right)^*(\beta) = \sup_{\alpha} \inf_{\substack{\rho_i \geq 0 \\ \sum_i \rho_i = 1}} \left[\langle \alpha, \beta \rangle - \sum_i \rho_i (f_i)^*(\alpha) \right].$$

D'après [Roc] (37.3.2), on peut échanger sup et inf. Or, pour $\{\rho_i\}$ fixé,

$$\sup_{\alpha} \left[\langle \alpha, \beta \rangle - \sum_i \rho_i (f_i)^*(\alpha) \right] = \inf_{\{\beta_i\}} \left[\sum_i \rho_i f_i(\beta_i) \mid \sum_i \rho_i \beta_i = \beta \right].$$

En effet, le membre de droite est une fonction de β qui est s.c.i. (on utilise que f_i est s.c.i. et que $\mathcal{J}$ est fini), convexe et de transformée de Legendre en α égale à :

$$\sup_{\beta} \sup_{\{\{\beta_i\} | \sum_i \rho_i \beta_i = \beta\}} \left[\langle \alpha, \beta \rangle - \sum_i \rho_i f_i(\beta_i) \right] = \sum_i \rho_i \sup_{\beta_i} \left[\langle \alpha, \beta \rangle - f_i(\beta_i) \right] = \sum_i \rho_i (f_i)^*(\alpha).$$

Alors,

$$\left(\sup_i \{(f_i)^*\} \right)^*(\beta) = \inf_{\{\{\beta_i\}, \{\rho_i\}, \sum_i \rho_i = 1\}} \left[\sum_i \rho_i f_i(\beta_i) \mid \sum_i \rho_i \beta_i = \beta \right] = (\text{conv}\{f_i\})(\beta). \quad \square$$

Notation 1.3.4 Pour F de $\mathbb{R}^d \times \mathbb{R}^d$ dans $] -\infty, +\infty]$: $(x, \beta) \mapsto F_x(\beta)$, et pour h positif, on note ${}^h\overline{F}_x$ l'enveloppe convexe de $\{F_y \mid y \in B(x, h)\}$ (cf définition 1.3.1), où $B(x, h)$ désigne la boule fermée de centre x et rayon h .

Propriétés :

- ${}^h\overline{F}_x$ est partout inférieure à F_x ; ${}^0\overline{F}_x$ est égale à F_x puisqu'on a pris une boule fermée.
- ${}^h\overline{F}_x$ est convexe, de domaine (ensemble des points où l'application est finie) l'enveloppe convexe des domaines des applications F_y , pour $y \in B(x, h)$.
- La fonction conjuguée de ${}^h\overline{F}_x$ est, d'après le théorème 1.3.2, l'application $\sup_{y \in B(x, h)} (F_y)^*$.
- Si pour tout y de $B(x, h)$, F_y est égal à F_x (en particulier si F est constante en la première variable), ${}^h\overline{F}_x$ est égale à F_x .

Cette définition va être utilisée avec $F = L$, champ de transformées de Cramer associé à un champ de probabilités (μ_x) (on reprend les notations 1.1.5). L'application ${}^h\overline{L}_x$ servira à définir une fonctionnelle de coût approché dans le chapitre 2 (utilisée dans les chapitres 2 et 3). Citons quelques propriétés de cette application :

- Le domaine de ${}^h\overline{L}_x$ est l'enveloppe convexe des domaines des L_y pour $y \in B(x, h)$ et elle s'annule sur l'enveloppe convexe des m_y .
- La fonction conjuguée de ${}^h\overline{L}_x$ est l'application $\sup_{y \in B(x, h)} H_y(\cdot)$.
- Quand les champs de probabilités sont constants, ${}^h\overline{L}_x$ est égal à L_x (indépendante de x).

On a, de plus, besoin d'une propriété proche de la semi-continuité de ${}^h\overline{L}_x$. Mais comme on fait varier y dans $B(x, h)$, ${}^h\overline{L}_x$ est la convexifiée d'un nombre infini d'applications et il n'est pas certain que ${}^h\overline{L}_x$ soit s.c.i.. On a probablement besoin d'une régularité de $x \mapsto L_x$ pour avoir une telle propriété. Cependant le résultat asymptotique suivant suffira pour notre propos :

Lemme 1.3.5

On considère un champ de probabilités (μ_x) . On lui associe les notations 1.1.5 et on suppose (H2) et (HC) vérifiées. Alors, pour x et β dans $\mathbb{R}^d$, on a :

$$L_x(\beta) \leq \lim_{(h, \beta') \rightarrow (0, \beta)} {}^h\overline{L}_x(\beta')$$

Démonstration : considérons une suite (h_k, β_k) convergeant vers $(0, \beta)$ donné et montrons qu'alors, on a :

$$L_x(\beta) \leq \liminf_{k \rightarrow +\infty} {}^{h_k}\overline{L}_x(\beta_k).$$

On peut supposer que tous les ${}^{h_k}\overline{L}_x(\beta_k)$ sont finis. Pour k fixé, ${}^{h_k}\overline{L}_x(\beta_k)$ est obtenu comme $\sum_p \rho_{k,p} L_{x_{k,p}}(\beta_{k,p})$ avec $x_{k,p}$ dans $B(x, h_k)$ et $(\rho_{k,p})_p$ famille presque nulle.

On accroît les supports S_x par convolution des probabilités μ_x avec des mesures ν^n à support égal à un compact indépendant de n (cf [AR] lemme 1.21) et convergeant faiblement vers la masse de Dirac à l'origine. On note S_x^n l'enveloppe convexe fermée du support de $\mu_x * \nu^n$. On sait qu'il existe $\delta > 0$ tel que pour tout n et pour tout x , S_x soit inclus dans $(S_x^n)^{-\delta}$. On note L_x^n la transformée de Cramer de $\mu_x * \nu^n$ (mesure qui n'est plus une probabilité). On a (cf [AR]) :

$$\text{pour tous } x, \beta, \quad L_x^n(\beta) \leq L_x(\beta) \quad \text{et} \quad \lim_{n \rightarrow +\infty} L_x^n(\beta) = L_x(\beta)$$

Supposons que l'on ait montré, pour chaque n :

$$L_x^n(\beta) \leq \liminf_{k \rightarrow +\infty} \sum_p \rho_{k,p} L_{x_{k,p}}^n(\beta_{k,p}) \quad (1.7)$$

Alors on a $L_x^n(\beta) \leq \liminf_{k \rightarrow +\infty} {}^{h_k}\overline{L}_x(\beta_k)$ (par $L^n \leq L$) et le membre de gauche de cette dernière inégalité converge vers $L_x(\beta)$ quand n tend vers $+\infty$. On aura donc bien le résultat.

Montrons (1.7) :

on commence par montrer que l'application $L^n(\cdot)$ possède les mêmes propriétés de continuité que $L(\cdot)$.

Posons $H^n(\alpha) = \log \int e^{(\alpha, y)} \nu^n(dy)$ et $H_x^n(\alpha) = \log \int e^{(\alpha, y)} \mu_x * \nu^n(dy)$. Alors H_x^n est égal à $H_x + H^n$. Pour tous x, y, α dans $\mathbb{R}^d$, $|H_x^n(\alpha) - H_y^n(\alpha)|$ est donc égal à $|H_x(\alpha) - H_y(\alpha)|$. D'autre part, l'ensemble S_x^n est compact et d'intérieur non vide. Le champ de mesures $(\mu_x * \nu^n)_x$ vérifie donc les mêmes hypothèses que le champ (μ_x) dans le théorème 1.2.8 (sauf qu'il ne s'agit pas d'un champ de probabilités). On en déduit que l'application $L^n(\cdot)$ est continue sur $\bigcup_y y \times \overset{\circ}{S}_y^n$.

Le vecteur $\beta_{k,p}$ appartient à $(S_{x_{k,p}}^n)^{-\delta}$. Soit $\varepsilon > 0$. L'application $L^n(\cdot)$ est uniformément continue sur $\bigcup_{y \in \mathcal{K}} y \times (S_y^n)^{-\delta}$. On prend δ_1 tel que, pour tous y, y' de $\mathcal{K}$ et pour tout v de $(S_y^n)^{-\delta}$, $|L_y^n(v) - L_{y'}^n(v)|$ soit inférieur à ε , puis K tel que $k \geq K$ implique $h_k < \delta_1$. Alors, pour $k \geq K$, on a pour tout p :

$$L_{x_{k,p}}^n(\beta_{k,p}) \geq L_x^n(\beta_{k,p}) - \varepsilon.$$

On en déduit :

$$\sum_p \rho_{k,p} L_{x_{k,p}}^n(\beta_{k,p}) \geq \sum_p \rho_{k,p} L_x^n(\beta_{k,p}) - \varepsilon \geq L_x^n(\beta_k) - \varepsilon,$$

en utilisant la convexité de L_x^n . Or, L_x^n est s.c.i. donc $L_x^n(\beta)$ est inférieur à $\liminf_{k \rightarrow +\infty} L_x^n(\beta_k)$.

On a alors, pour k assez grand :

$$L_x^n(\beta) \leq L_x^n(\beta_k) + \varepsilon \leq \sum_p \rho_{k,p} L_{x_{k,p}}^n(\beta_{k,p}) + 2\varepsilon,$$

ce qui démontre (1.7) et donc le lemme 1.3.5. $\square$

Comme, de plus, ${}^h\overline{L}_x$ est majorée par L_x , on a en particulier :

$$\text{pour } x \text{ et } \beta \text{ dans } \mathbb{R}^d, L_x(\beta) = \lim_{h \rightarrow 0} {}^h\overline{L}_x(\beta). \quad (1.8)$$

1.4 Champs de probabilités discontinus le long d'une surface

On définit dans un premier paragraphe la famille de processus sujet du reste de l'étude. Il s'agit encore d'un processus d'apprentissage lent mais, par rapport aux hypothèses faites précédemment, toute propriété de continuité de (μ_x) est perdue sur la frontière d'un ouvert régulier. On traite, dans cette étude, le cas où cette frontière est un hyperplan sachant que, par des techniques de localisation (on introduit des coordonnées locales convenables, voir par exemple [AO]), on peut en déduire les résultats concernant une frontière régulière quelconque. Supposer que (μ_x) est continue seulement de part et d'autre d'un hyperplan revient à considérer qu'on a deux champs différents de part et d'autre de cet hyperplan, tous les deux continus. On indique dans les paragraphes suivants les premiers éléments de cette étude : on propose une fonctionnelle de coût "présumée" dont on étudie les propriétés et on évalue ce que donne l'utilisation de résultats existants (dont [Du88]).

1.4.1 Définition du processus

On note π , l'application de $\mathbb{R}^d$ dans $\mathbb{R}$ qui, à un vecteur de $\mathbb{R}^d$, associe sa première coordonnée. On pose

$$\begin{cases} P^0 = \{x \in \mathbb{R}^d, \pi(x) = 0\} \\ P^1 = \{x \in \mathbb{R}^d, \pi(x) < 0\} \\ P^2 = \{x \in \mathbb{R}^d, \pi(x) > 0\} \end{cases}$$

On suppose qu'on a deux champs de probabilités (μ_x^1) et (μ_x^2) sur $\mathbb{R}^d$ (on peut se contenter, pour $i = 1$ ou 2 , de μ_x^i défini au voisinage de P^i). A chaque champ (μ_x^i) , on associe les notations correspondant aux notations 1.1.5 : m_x^i désigne le barycentre de μ_x^i , H_x^i le logarithme de sa transformée de Laplace et L_x^i sa transformée de Cramer. On note S_x^i l'enveloppe convexe fermée du support de μ_x^i .

Soient $(\Omega, \mathcal{B}, P)$ un espace probabilisé, et $(\mathcal{F}_k)_{k \in \mathbb{N}}$ une suite croissante de sous- σ -algèbres de $\mathcal{B}$. On suppose que, pour $i = 1$ ou 2 , les $X_k^i(\cdot)$, $k \in \mathbb{N}$, sont des champs de vecteurs aléatoires indépendants, tels que pour tout x de $\mathbb{R}^d$ et tout k de $\mathbb{N}$, $X_k^i(x)$ est de loi μ_x^i et $\mathcal{F}_{k+1}$ -mesurable.

Définition 1.4.1 Si $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$ est une famille de processus indexée par τ , on dit que $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$ est un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 associé aux champs de probabilités (μ_x^1) et (μ_x^2) si, pour $\tau > 0$, $(Y_k^\tau)_{k \in \mathbb{N}}$ est une chaîne de Markov déterminée par la donnée de Y_0^τ et la relation de récurrence :

$$Y_{k+1}^\tau = Y_k^\tau + \tau \left[\mathbb{I}_{\{Y_k^\tau \in P^1 \cup P^0\}} X_k^1(Y_k^\tau) + \mathbb{I}_{\{Y_k^\tau \in P^2\}} X_k^2(Y_k^\tau) \right] \quad (1.9)$$

Si $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$ est une telle famille, Y_k^τ est $\mathcal{F}_k$ -mesurable et, sous P , la loi de $\frac{1}{\tau}(Y_{k+1}^\tau - Y_k^\tau)$ conditionnellement à Y_k^τ est $\mu_{Y_k^\tau}^1$ si Y_k^τ appartient à $P^1 \cup P^0$ et $\mu_{Y_k^\tau}^2$ si Y_k^τ appartient à P^2 . Il y a donc discontinuité de la loi du champ de vitesses sur l'hyperplan P^0 .

Exemple : un processus de cette forme apparaît dans certains algorithmes d'apprentissage pour les réseaux de neurones, par exemple l'apprentissage compétitif ([RZ]) : on suppose que chacune couche du réseau est partitionnée en plusieurs groupes de cellules et que, au sein d'un même groupe, pour un pattern donné sur la couche précédente, une cellule seulement est activée. Notons W_j le vecteur de poids de la cellule j avec la couche précédente ; dans un même groupe de cellules, si F est le pattern formé sur la couche précédente (les patterns sont présentés aléatoirement sur la couche d'entrée), W_j est modifié pendant l'apprentissage si et seulement si $(F|W_l)$ est maximal pour $l = j$ et il est modifié alors de $\tau(\frac{F}{|F|} - W_j)$. On constate que les frontières de discontinuité de la loi de $(W_l)_l$ sont des hyperplans.

Remarque 1.4.2 En imposant le champ de probabilités μ_x^1 sur P^0 , nous introduisons une dissymétrie du problème, suivant en cela le choix de [DE]. Nous commençons davantage ce choix dans le paragraphe 4.1.3. Nous constaterons que, sous une hypothèse (H_0) sur la position des S_x^i par rapport à $\{\pi(\cdot) = 0\}$, les résultats restent symétriques.

On appelle “pas” les accroissements de la chaîne (Y_k^τ) et on dit qu’un pas est “de type i ” quand il est de la forme τX où X suit une loi μ_x^i .

Pour établir les inégalités de grandes déviations, on veut estimer la probabilité que γ^τ appartienne à des ensembles de chemins donnés. Pour cela, on décompose $[0, T]$ en sous-intervalles de la forme $[jn\tau; (j+1)n\tau[$, de longueur $n\tau$ petite, sur lesquels on utilise des estimations, pour n grand, de la probabilité que la vitesse moyenne soit proche d’un vecteur donné. En fixant $n\tau$ constant, à τ près, faire tendre τ vers 0 est équivalent à faire tendre n vers $+\infty$. Chacun des chapitres 2 et 3 commence ainsi par des estimations sur les vitesses moyennes. Pour pouvoir réunir les résultats concernant les différents sous-intervalles, ces estimations doivent être vraies uniformément pour une valeur initiale Y_0^τ dans un compact donné.

Dans le paragraphe suivant nous donnons de tels résultats sur les vitesses moyennes, ceux démontrés dans [DE] en présence d’un hyperplan de discontinuité, mais avec, de part et d’autre de cet hyperplan, des champs de probabilités constants.

1.4.2 Cas particulier des champs constants, étude des vitesses moyennes asymptotiques ([DE])

Dans [DE], P. Dupuis et R.S. Ellis établissent les résultats suivants, **en supposant constants les deux champs de probabilités** (μ_x^1) et (μ_x^2) (on supprime l’indice x pour toutes les notations associées aux champs de probabilités, les mêmes que celles introduites dans les paragraphes précédents).

On suppose que Y_0^τ est P -presque sûrement égal à y_0 , où y_0 est un point de P^0 . Alors, le paramètre τ n’influe pas sur la position des Y_k^τ par rapport à P^0 , donc sur le champ de mesures gouvernant chaque pas. Les trajectoires pour différents τ sont homothétiques (le vecteur aléatoire $\frac{1}{\tau}(Y_k^\tau - Y_0^\tau)$ est indépendant de τ) et il suffit d’en étudier une. On considère $\tau = 1$ et on omet ce paramètre.

On suppose que, pour $i = 1$ ou 2 , le logarithme de la transformée de Laplace de μ^i , H^i , est fini sur $\mathbb{R}^d$ et que l’hypothèse suivante est vérifiée (hypothèse liée à la présence de la discontinuité) :

Hypothèse (H_0) : pour chaque i , $\{\beta \in \mathbb{R}^d \mid (-1)^i \pi(\beta) < 0\} \cap \text{supp } \mu^i \neq \emptyset$,

où $\text{supp } \mu^i$ désigne le support de la probabilité μ^i . Cette hypothèse consiste à dire que le support de chaque μ^i contient un vecteur sortant strictement de P^i . Elle per-

met aux deux demi-espaces ouverts P^i de communiquer.

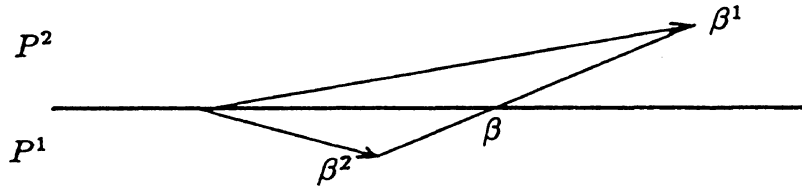
On montre que, sous l'hypothèse que les logtransformées de Laplace H^1 et H^2 sont finies partout, l'hypothèse (H_0) implique, pour chaque i :

$$\{\beta \in \mathbb{R}^d \mid (-1)^i \pi(\beta) < 0\} \cap \text{Int } S^i \neq \emptyset,$$

où $\text{Int } S^i$ désigne l'intérieur de S^i relatif à l'espace vectoriel engendré par S^i .

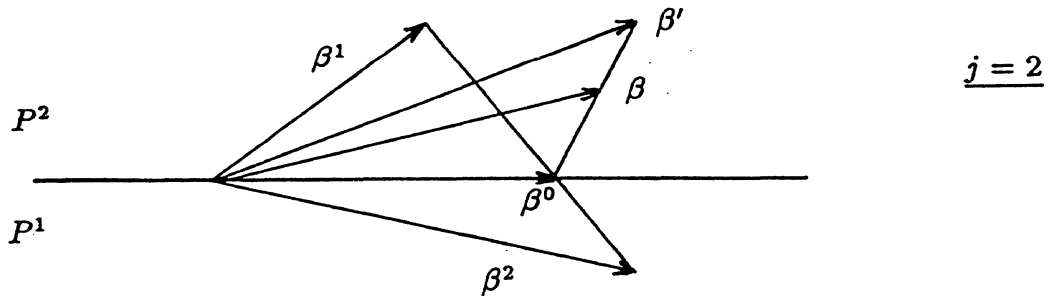
Notation 1.4.3 On pose, pour β dans $\mathbb{R}^d$ de première coordonnée $\pi(\beta)$ nulle :

$$L(\beta) = \inf \left\{ \sum_{i=1}^2 \rho^i L^i(\beta^i) \mid \sum_{i=1}^2 \rho^i = 1, \sum_{i=1}^2 \rho^i \beta^i = \beta, \pi(\beta^1) \geq 0, \pi(\beta^2) \leq 0 \right\}$$



Pour un vecteur β de première coordonnée $\pi(\beta)$ non nulle, on pose, quand on a $(-1)^j \pi(\beta) > 0$:

$$L(\beta) = \inf \left\{ \rho^0 L(\beta^0) + \rho L^j(\beta') \mid \rho^0 + \rho = 1, \rho^0 \beta^0 + \rho \beta' = \beta, \pi(\beta^0) = 0 \right\}.$$



L est une fonction de coût mixte (faisant intervenir L^1 et L^2) correspondant au mélange des deux champs de vecteurs.

On a le résultat suivant (Dupuis/Ellis) :

Si (μ^1) et (μ^2) sont deux champs de probabilités constants tels que (H_0) soit vérifiée et que, pour $i = 1$ ou 2 et α dans $\mathbb{R}^d$, $\int e^{\langle \alpha, y \rangle} \mu^i(dy)$ soit fini, alors, conditionnellement à $Y_0 = y_0$, où y_0 appartient à P^0 , la distribution de $\frac{Y_n}{n}$ suit un principe de grandes déviations avec la fonction de coût L . On a donc, pour toute partie borélienne B de $\mathbb{R}^d$:

$$-L(\overset{\circ}{B}) \leq \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_{y_0} \left(\frac{Y_n}{n} \in B \right) \leq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log P_{y_0} \left(\frac{Y_n}{n} \in B \right) \leq -L(\overline{B}).$$

1.4.3 Retour au cas général, conjectures sur les fonctions de coût et propriétés de ces fonctions

On revient au cas général de deux champs non constants (μ_x^1) et (μ_x^2) . Dans toute cette partie, on suppose que pour tout x , les logarithmes des transformées de Laplace H_x^1 et H_x^2 sont finis sur $\mathbb{R}^d$. En s'inspirant des différents résultats rappelés jusqu'à présent, on cherche une fonction de coût permettant d'écrire des inégalités de grandes déviations pour $(\gamma^\tau)_\tau$. On peut s'attendre à ce qu'intervienne la fonctionnelle $\Lambda_{0,T}$ définie de la façon suivante :

Notation 1.4.4 On a noté, pour x dans $\mathbb{R}^d$, L_x^1 et L_x^2 les transformées de Cramer des probabilités μ_x^1 et μ_x^2 . On pose, pour β dans $\mathbb{R}^d$:

- $L_x^0(\beta) = \begin{cases} L_x^1(\beta) & \text{si } \pi(\beta) \leq 0 \\ L_x^2(\beta) & \text{si } \pi(\beta) > 0 \end{cases}$
- $L_x(\beta) = \inf \left\{ \sum_{i=1}^2 \rho^i L_x^i(\beta^i) + \rho^3 L_x^0(\beta^3) \mid \right.$
 $\left. \sum_{i=1}^3 \rho^i = 1, \sum_{i=1}^3 \rho^i \beta^i = \beta, \pi(\beta^1) \geq 0, \pi(\beta^2) \leq 0, \pi(\rho^1 \beta^1 + \rho^2 \beta^2) = 0 \right\}$ (1.10)

• et

$$\lambda(x, \beta) = \begin{cases} L_x^i(\beta) & \text{si } x \in P^i \\ L_x(\beta) & \text{si } x \in P^0 \end{cases} \quad (1.11)$$

et on définit la fonctionnelle de coût $\Lambda_{0,T}$ par :

Notation 1.4.5 Pour tout chemin continu, φ de $[0, T]$ dans $\mathbb{R}^d$,

$$\Lambda_{0,T}(\varphi) = \begin{cases} \int_0^T \lambda(\varphi(t), \varphi'(t)) dt & \text{si } \varphi \text{ est absolument continu} \\ +\infty & \text{sinon} \end{cases}$$

Remarques :

- la fonction de coût L_x sera utilisée pour x dans P^0 et sera intégrée le long des chemins. D'après la remarque 1.4.9, on peut donc se contenter de retenir simplement la formule (1.12) ci-dessous.
- dans toute l'étude, les coefficients ρ^i envisagés dans les décompositions de la forme $\beta = \sum_{i=1}^3 \rho^i \beta^i$ sont toujours positifs (ils sont donc à valeurs dans $[0, 1]$ puisque leur somme vaut 1 : on regarde des combinaisons linéaires barycentriques).

Dans la suite de ce paragraphe, nous donnons l'interprétation de la fonction L_x puis ses propriétés ainsi que celles de la fonctionnelle $\Lambda_{0,T}$.

I. Interprétation intuitive de la formule du coût L_x

1. si $\pi(\beta) = 0$, $L_x(\beta)$ est calculé avec des combinaisons linéaires de deux termes en β^1 et β^2 . On a précisément :

quand $\pi(\beta)$ vaut 0, $L_x(\beta)$ est égal à :

$$\inf \left\{ \sum_{i=1}^2 \rho^i L_x^i(\beta^i) \mid \sum_{i=1}^2 \rho^i = 1, \sum_{i=1}^2 \rho^i \beta^i = \beta, \pi(\beta^1) \geq 0, \pi(\beta^2) \leq 0 \right\} \quad (1.12)$$

En effet, si ρ^3 est différent de 0, $\pi(\beta^3)$ doit être nul, or L_x^1 est convexe. On peut donc écrire :

$$L_{x/\{\pi(\cdot)=0\}} = \text{conv}(L_{x/\{\pi(\cdot) \geq 0\}}^1, L_{x/\{\pi(\cdot) \leq 0\}}^2).$$

Notre définition de $L_x(\beta)$ coïncide donc parfaitement avec la définition 1.4.3 quand les champs de probabilités sont constants.

Si on enlève les contraintes $\pi(\beta^1) \geq 0$ et $\pi(\beta^2) \leq 0$ dans la formule (1.12) on obtient l'expression de $\text{conv}(L_x^1, L_x^2)(\beta)$, la plus grande fonction convexe minorant simultanément L_x^1 et L_x^2 ; en effet, on a :

$$(\text{conv}(L_x^1, L_x^2))(\beta) = \inf \left\{ \sum_{i=1}^2 \rho^i L_x^i(\beta^i) \mid \sum_{i=1}^2 \rho^i = 1, \sum_{i=1}^2 \rho^i \beta^i = \beta \right\} \quad (1.13)$$

On en déduit que L_x est minoré par $\text{conv}(L_x^1, L_x^2)$ sur $\{\pi(\cdot) = 0\}$.

La formule (1.12), donnant la valeur de $L_x(\beta)$ pour $\pi(\beta) = 0$, s'interprète de la façon suivante :

considérons la chaîne $(Y_k^\tau)_{0 \leq k \leq n}$, avec Y_0^τ appartenant à P^0 ; on note r^1 la proportion d'indices k de $\{0, \dots, n-1\}$ tels que Y_k^τ soit dans $P^0 \cup P^1$ et on pose $r^2 = 1 - r^1$. On appelle w^1 (respectivement w^2) la moyenne des termes $\frac{1}{\tau}(Y_{k+1}^\tau - Y_k^\tau)$ tels que Y_k^τ soit dans $P^0 \cup P^1$ (respectivement dans P^2), c'est-à-dire la moyenne des $\mathbb{I}_{\{Y_k^\tau \in P^1 \cup P^0\}} X_k^1(Y_k^\tau)$ (respectivement $\mathbb{I}_{\{Y_k^\tau \in P^2\}} X_k^2(Y_k^\tau)$). Le vecteur w^i est la vitesse moyenne du processus quand il est soumis au champ (μ_x^i) , sur l'intervalle de longueur $[0, n\tau]$. Pour Y_n^τ proche de P^0 , on a pratiquement $(-1)^i \pi(w^i) \leq 0$ (en effet, si $Y_k^\tau, Y_{k+1}^\tau, \dots, Y_{k+m}^\tau$ sont dans P^i et Y_{k+m+1}^τ dans P^{3-i} , $(-1)^i \pi(Y_{k+m+1}^\tau - Y_k^\tau)$ est négatif et $(-1)^i \pi(w^i)$ est une moyenne de telles quantités, sauf éventuellement pour les derniers pas). On s'intéresse aux valeurs possibles de la vitesse moyenne globale du processus sur un intervalle de longueur $n\tau$. Elle vaut $\frac{Y_n^\tau - Y_0^\tau}{n\tau} = \sum_{i=1}^2 r^i w^i$. Pour qu'elle soit égale à β , quand $\pi(\beta)$ vaut 0, il est donc logique de considérer les décompositions de β sous la forme $\beta = \sum_{i=1}^2 \rho^i \beta^i$ avec $(-1)^i \pi(\beta^i) \leq 0$. Pour n grand, l'opposé du logarithme de la probabilité d'observer $(r^i, w^i) = (\rho^i, \beta^i)$ vaut approximativement $n \sum_{i=1}^2 \rho^i L_x^i(\beta^i)$ (aux variations de la mesure près, sachant qu'elles sont petites pour $n\tau$ petit). La quantité ρ^i correspond au temps passé dans P^i , le vecteur β^i à la moyenne de ceux des pas qui sont contrôlés par la loi μ_x^i . La formule donnant $L_x(\beta)$ correspond au minimum parmi toutes ces combinaisons.

2. Pour β rentrant dans P^i (c'est-à-dire $(-1)^i \pi(\beta) > 0$), on regarde la décomposition de β la moins coûteuse en un vecteur parallèle à P^0 et un vecteur rentrant dans P^i . Le coût correspondant est égal au barycentre des coûts des deux vecteurs, pondéré par les coefficients de la décomposition (formule 1.10).

On constate que L_x est partout minorée (au sens large) par $\text{conv}(L_x^1, L_x^2)$. [DE] donne (pour des champs constants) des exemples où on a $\text{conv}(L^1, L^2) < L$.

II. Interprétation dans le cas de champs de probabilités constants, à support fini

1. Rappelons tout d'abord quel est le coût associé à un seul champ constant de ce type. Considérons le champ de probabilités constant :

$$\mu = \sum_{k=1}^K p_k \delta_{a_k}$$

On sait que la transformée de Cramer de μ en β vaut (voir par exemple [DZ], théorème 2.1.24) :

$$L(\beta) = \inf \left\{ \sum_{k=1}^K \rho_k \log \frac{\rho_k}{p_k} \mid \beta = \sum_{k=1}^K \rho_k a_k, \rho_k \geq 0, \sum_{k=1}^K \rho_k = 1 \right\}. \quad (1.14)$$

Cette expression correspond au minimum de l'information de Kullback, ou entropie relative, de $\sum_{k=1}^K \rho_k \mathbb{I}_{a_k}$ par rapport à μ pour $\beta = \sum_{k=1}^K \rho_k a_k$. Montrons comment on la retrouve d'un point de vue combinatoire en utilisant qu'on doit avoir (théorème de Cramer), si les X_k sont i.i.d. de loi μ :

$$L(\beta) = - \lim_{n \rightarrow +\infty} \frac{1}{n} \log P \left(\frac{X_1 + \dots + X_n}{n} \sim \beta \right)$$

$P \left(\{X_1, \dots, X_n\} = \underbrace{\{a_1, \dots, a_1\}}_{n_1 \text{ fois}}, \dots, \underbrace{\{a_K, \dots, a_K\}}_{n_K \text{ fois}} \right)$ est égal à

$$C_n^{n_1} C_{n-n_1}^{n_2} \dots C_{n_K}^{n_K} \prod_{k=1}^K p_k^{n_k} = \frac{n!}{n_1! \dots n_K!} \prod_{k=1}^K p_k^{n_k}.$$

Supposons que, pour chaque k , $\frac{n_k}{n}$ converge vers une constante strictement positive ρ_k . On a : $\log(n_k!) \sim n_k \log n_k - n_k + O(\log n_k)$.

Alors $\log \left(\frac{n!}{n_1! \dots n_K!} \right) \sim n \log n - \sum_{k=1}^K n_k \log n_k = - \sum_{k=1}^K n_k \log \frac{n_k}{n}$. On a donc

$$-\frac{1}{n} \log P \left(\{X_1, \dots, X_n\} = \underbrace{\{a_1, \dots, a_1\}}_{n_1 \text{ fois}}, \dots, \underbrace{\{a_K, \dots, a_K\}}_{n_K \text{ fois}} \right) \sim \sum_{k=1}^K \frac{n_k}{n} \log \frac{n_k}{n p_k} \sim \sum_{k=1}^K \rho_k \log \frac{\rho_k}{p_k}$$

Pour avoir $\frac{X_1 + \dots + X_n}{n} = \beta$, on regarde toutes les décompositions de la forme $\sum_{k=1}^K \frac{n_k}{n} a_k = \beta$. On retrouve donc $-\frac{1}{n} \log P \left(\frac{X_1 + \dots + X_n}{n} \sim \beta \right) \sim L(\beta)$, où $L(\beta)$ est donné par (1.14).

2. Calculons le coût mixte correspondant aux deux champs constants

$$\mu^1 = \sum_{k=1}^K p_k \delta_{a_k} \text{ et } \mu^2 = \sum_{k=K+1}^{K'} p_k \delta_{a_k}$$

Pour $\pi(\beta) = 0$, $L(\beta)$ vaut :

$$\inf \left\{ \sum_{k=1}^K \rho \rho_k \log \frac{\rho_k}{p_k} + \sum_{k=K+1}^{K'} (1-\rho) \rho_k \log \frac{\rho_k}{p_k} \mid \beta = \rho \sum_{k=1}^K \rho_k a_k + (1-\rho) \sum_{k=K+1}^{K'} \rho_k a_k, \right. \\ \left. \sum_{k=1}^K \rho_k = \sum_{k=K+1}^{K'} \rho_k = 1, 0 \leq \rho \leq 1, \pi \left(\sum_{k=1}^K \rho_k a_k \right) \geq 0, \pi \left(\sum_{k=K+1}^{K'} \rho_k a_k \right) \leq 0 \right\}.$$

Les coefficients ρ_k sont toujours positifs. Sans les deux dernières conditions, on obtient l'expression de $\text{conv}(L^1, L^2)(\beta)$ qui est donc inférieur à $L(\beta)$. La quantité $\text{conv}(L^1, L^2)(\beta)$ peut s'écrire :

$$\inf \left\{ \sum_{k=1}^K \rho_k \log \frac{\rho_k}{p_k \sum_{k=1}^K \rho_k} + \sum_{k=K+1}^{K'} \rho_k \log \frac{\rho_k}{p_k \sum_{k=K+1}^{K'} \rho_k} \mid \beta = \sum_{k=1}^{K'} \rho_k a_k, \sum_{k=1}^{K'} \rho_k = 1 \right\},$$

ou $\inf_{0 \leq \rho \leq 1} \inf \left\{ \sum_{k=1}^K \rho_k \log \frac{\rho_k}{\rho p_k} + \sum_{k=K+1}^{K'} \rho_k \log \frac{\rho_k}{(1-\rho)p_k} \mid \beta = \sum_{k=1}^{K'} \rho_k a_k, \sum_{k=1}^K \rho_k = \rho, \sum_{k=K+1}^{K'} \rho_k = 1 - \rho \right\}.$

On peut chercher à comparer cette dernière fonction à la transformée de Cramer d'une mesure ayant en chaque a_k une masse proportionnelle à p_k . On constate que la fonction $\text{conv}(L^1, L^2)$ est plus grande que le minimum des transformées de Cramer des probabilités $\mu_\rho = \sum_{k=1}^K \rho p_k \delta_{a_k} + \sum_{k=K+1}^{K'} (1-\rho) p_k \delta_{a_k}$, qui s'écrit $\inf_{0 \leq \rho \leq 1} L_\rho$ avec, pour β dans $\mathbb{R}^d$, $L_\rho(\beta)$ égal à

$$\inf \left\{ \sum_{k=1}^K \rho_k \log \frac{\rho_k}{\rho p_k} + \sum_{k=K+1}^{K'} \rho_k \log \frac{\rho_k}{(1-\rho)p_k} \mid \beta = \sum_{k=1}^{K'} \rho_k a_k, \sum_{k=1}^{K'} \rho_k = 1 \right\}$$

On a donc :

$$\inf_{0 \leq \rho \leq 1} L_\rho \leq \text{conv}(L^1, L^2) \leq L$$

Le point de vue combinatoire ci-dessus a mis en évidence le lien entre la valeur de la fonction de coût $L(\beta)$ et les combinaisons observables des vecteurs a_k , appartenant au support de la mesure, dont la moyenne vaut β . La différence entre les fonctions L et $\inf_{0 \leq \rho \leq 1} L_\rho$ correspond, au niveau du processus, au fait de pouvoir ou non mélanger les vecteurs a_k des supports de μ^1 et μ^2 . Pour le coût mixte, l'ordre des a_k conduisant à la valeur de $L(\beta)$ est soumis à la contrainte physique "faire un pas i quand on est dans P^i " : on ne peut échanger entre eux les a_k correspondant à des champs différents. De plus, tous les ordres des a_k correspondant à un même champ ne sont pas observables, par exemple : une combinaison utilisant d'abord tous les vecteurs qui font traverser P^0 , pour chacun des deux champs, est impossible, puisqu'on ne peut plus ensuite observer de mélange des a_k restants, qui éloignent de P^0 . Ces contraintes sur l'ordre des a_k se traduisent dans le calcul de $L(\beta)$ en la contrainte géométrique $(-1)^i \pi(\beta^i) \leq 0$ dans la formule (1.12), contrainte qui est automatiquement vérifiée quand S_x^1 est inclus dans $\{\pi(\cdot) > 0\}$ et S_x^2 dans $\{\pi(\cdot) < 0\}$. La fonction L est alors égale à $\text{conv}(L^1, L^2)$; les vecteurs a_k peuvent être observés dans n'importe quel ordre.

III. Propriétés de la fonction de coût L_x

Proposition 1.4.6

On considère des champs de probabilités (μ_x^1) et (μ_x^2) tels que pour tout x , les logarithmes des transformées de Laplace H_x^1 et H_x^2 sont finis sur $\mathbb{R}^d$. Alors, la fonction de coût L_x vérifie :

- (i) Pour tout x et tout β de $\mathbb{R}^d$, on a $\text{conv}(L_x^1, L_x^2)(\beta) \leq L_x(\beta)$.
 Si $\pi(\beta) = 0$, on a $L_x(\beta) \leq L_x^1(\beta) \wedge L_x^2(\beta)$.
 Si $(-1)^i \pi(\beta) > 0$, on a $L_x(\beta) \leq L_x^i(\beta)$.
- (ii) Dès que $L_x(\beta)$ est fini, la borne inférieure dans la définition (1.10) de $L_x(\beta)$ est atteinte.
- (iii) L'application L_x est s.c.i. sur $\mathbb{R}^d$ et convexe sur $\{\beta | \pi(\beta) \leq 0\}$ et sur $\{\beta | \pi(\beta) \geq 0\}$.

Démonstration :

(i) a été vu.

- (ii) Comme dans la démonstration du théorème 1.3.3, le fait que $(L_x^i)^*$ soit finie permet de faire le même raisonnement que si S_x^i était compact. Supposons $L_x(\beta)$ fini et considérons (ρ_n^i, β_n^i) tel que $\sum_{i=1}^2 \rho_n^i L_x^i(\beta_n^i) \leq L_x(\beta) + \frac{1}{n}$ et vérifiant les conditions apparaissant dans (1.12). Chaque suite (ρ_n^i) admet une valeur d'adhérence ρ^i . Si ρ^i est non nul, (β_n^i) admet une valeur d'adhérence β^i (puisque les ensembles $\{L_x^i(\cdot) \leq I\}$ sont compacts). On a alors : $\beta = \sum_{i|\rho^i \neq 0} \rho^i \beta^i$. D'autre part,

$$\sum_{i|\rho^i \neq 0} \rho^i L_x^i(\beta^i) \leq \lim_{n \rightarrow +\infty} \sum_{i=1}^2 \rho_n^i L_x^i(\beta_n^i) \leq L_x(\beta),$$

donc $L_x(\beta)$ est égal à $\sum_{i|\rho^i \neq 0} \rho^i L_x^i(\beta^i)$.

De même, si $(-1)^i \pi(\beta) > 0$, $L_x(\beta)$ s'écrit $\sum_{i=1}^2 \rho^i L_x^i(\beta^i) + \rho^3 L_x^3(\beta^3)$.

- (iii) La semi-continuité et la convexité peuvent se démontrer en écrivant L_x sous la forme de la conjuguée d'une fonction (cf ci-dessous). On peut également montrer directement ces propriétés. Montrons tout d'abord la semi-continuité. On considère une suite (β_n) convergeant vers un vecteur β donné, dans $\mathbb{R}^d$. La quantité $L_x(\beta_n)$ s'écrit $\sum_{i=1}^2 \rho_n^i L_x^i(\beta_n^i) + \rho_n^3 L_x^0(\beta_n^3)$. La suite $(\rho_n^i, \beta_n^i ; i = 1, 2, 3, n \in \mathbb{N})$ est à valeurs dans un compact donc admet une valeur d'adhérence (ρ^i, β^i) (dans le cas où S_x^i est compact, sinon considérer d'abord ρ^i , une valeur d'adhérence de (ρ_n^i) , et discuter à nouveau suivant que ρ^i est nul

ou pas). On a :

$$\sum_{i=1}^3 \rho^i = 1, \quad \sum_{i=1}^3 \rho^i \beta^i = \beta, \quad \pi(\beta^1) \geq 0, \quad \pi(\beta^2) \leq 0 \quad \text{et} \quad \pi(\rho^1 \beta^1 + \rho^2 \beta^2) = 0,$$

donc :

$$L_x(\beta) \leq \sum_{i=1}^2 \rho^i L_x^i(\beta^i) + \rho^3 L_x^0(\beta^3).$$

Les applications L_x^1 et L_x^2 sont s.c.i.. D'autre part, si $\pi(\beta)$ est non nul, $\pi(\beta^3)$ est également non nul et L_x^0 est alors s.c.i. en β^3 . Dans ce cas, on a bien :

$$L_x(\beta) \leq \lim_{n \rightarrow +\infty} \sum_{i=1}^2 \rho_n^i L_x^i(\beta_n^i) + \rho_n^3 L_x^0(\beta_n^3).$$

Si $\pi(\beta)$ est égal à 0, deux cas sont à envisager :

– soit ρ^3 vaut 0 ; alors, comme $\sum_{i=1}^2 \rho^i \beta^i$ est égal à β , on a :

$$L_x(\beta) \leq \sum_{i=1}^2 \rho^i L_x^i(\beta^i) \leq \lim_{n \rightarrow +\infty} \sum_{i=1}^2 \rho_n^i L_x^i(\beta_n^i) \leq \lim_{n \rightarrow +\infty} \left[\sum_{i=1}^2 \rho_n^i L_x^i(\beta_n^i) + \rho_n^3 L_x^0(\beta_n^3) \right].$$

– si ρ^3 n'est pas égal à 0, c'est $\pi(\beta^3)$ qui est nul. On a alors :

$$(L_x^1 \wedge L_x^2)(\beta^3) \leq \lim_{n \rightarrow +\infty} L_x^0(\beta_n^3).$$

Si, par exemple $(L_x^1 \wedge L_x^2)(\beta^3)$ vaut $L_x^2(\beta^3)$, puisque $\pi\left(\frac{\rho^2 \beta^2 + \rho^3 \beta^3}{\rho^2 + \rho^3}\right) \leq 0$ et $\rho^1 \beta^1 + (\rho^2 + \rho^3) \frac{\rho^2 \beta^2 + \rho^3 \beta^3}{\rho^2 + \rho^3} = \beta$, on a :

$$L_x(\beta) \leq \rho^1 L_x^1(\beta^1) + (\rho^2 + \rho^3) L_x^2\left(\frac{\rho^2 \beta^2 + \rho^3 \beta^3}{\rho^2 + \rho^3}\right).$$

En utilisant la convexité de L_x^2 , on en déduit :

$$L_x(\beta) \leq \sum_{i=1}^2 \rho^i L_x^i(\beta^i) + \rho^3 L_x^2(\beta^3).$$

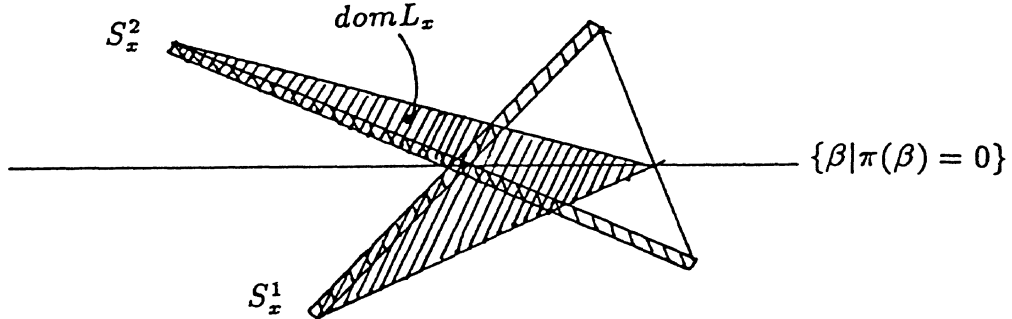
Donc, on a encore

$$L_x(\beta) \leq \lim_{n \rightarrow +\infty} \left[\sum_{i=1}^2 \rho_n^i L_x^i(\beta_n^i) + \rho_n^3 L_x^0(\beta_n^3) \right],$$

ce qui prouve la semi-continuité de L_x .

A cause du terme $\rho^3 L_x^0(\beta^3)$, L_x n'est pas nécessairement convexe. Par exemple

son domaine (ensemble des points où elle est finie) n'est pas convexe dans le cas suivant :



En revanche, L_x est convexe sur chaque $\{(-1)^j \pi(\cdot) \geq 0\}$, $j = 1, 2$. Là encore, l'écriture sous forme de la conjuguée d'une fonction, ci-dessous, donne cette propriété mais on peut écrire une démonstration directe proche de celle de la semi-continuité. En effet, d'après la convexité de L_x^j , pour tout β tel que $\{(-1)^j \pi(\beta) \geq 0\}$, $L_x(\beta)$ s'écrit :

$$\inf \left\{ \sum_{i=1}^2 \rho^i L_x^i(\beta^i) + \rho^3 L_x^3(\beta^3) \mid \sum_{i=1}^3 \rho^i = 1, \sum_{i=1}^3 \rho^i \beta^i = \beta, \pi(\beta^1) \geq 0, \pi(\beta^2) \leq 0, \pi(\rho^1 \beta^1 + \rho^2 \beta^2) = 0 \right\}.$$

On utilise alors la convexité de L_x^1 , L_x^2 et celle des ensembles $\{\beta | \pi(\beta) \leq 0\}$ et $\{\beta | \pi(\beta) \geq 0\}$. La proposition 1.4.6 est donc prouvée. $\square$

Remarquons que, contrairement à L_x^1 et L_x^2 , L_x n'est pas nécessairement strictement convexe sur $\{\beta | \pi(\beta) = 0\}$. Par exemple, si $\pi(m_x^1) = \pi(m_x^2) = 0$, L_x est constamment nulle sur le segment constitué des barycentres de m_x^1 et m_x^2 .

Notation 1.4.7 Si $\{A_i\}$ est un ensemble de parties de $\mathbb{R}^d$, on note $\text{Conv}\{A_i\}$ l'enveloppe convexe de l'ensemble des A_i .

Le domaine de $L_x / \{\beta | \pi(\beta) = 0\}$ est l'intersection de l'ensemble $\{\beta | \pi(\beta) = 0\}$ avec

$$\text{Conv} \left\{ (\text{dom} L_x^1) \cap \{\beta | \pi(\beta) \geq 0\}, (\text{dom} L_x^2) \cap \{\beta | \pi(\beta) \leq 0\} \right\}.$$

Le domaine de $L_x / \{\beta | \pi(\beta) < 0\}$ (respectivement $L_x / \{\beta | \pi(\beta) > 0\}$) est l'enveloppe convexe de $\text{dom} L_x / \{\beta | \pi(\beta) = 0\}$ et de $(\text{dom} L_x^1) \cap \{\beta | \pi(\beta) < 0\}$ (respectivement $(\text{dom} L_x^2) \cap \{\beta | \pi(\beta) > 0\}$).

Nous poursuivons l'étude des propriétés de L_x en donnant son écriture sous forme de la conjuguée d'une fonction et en discutant sa valeur selon la position des supports des probabilités μ_x^i par rapport à la frontière.

Les rappels d'analyse convexe du paragraphe 1.3 permettent, toujours en supposant H_x^1 et H_x^2 finies sur $\mathbb{R}^d$, d'écrire L_x sous la forme de la conjuguée d'une fonction (cf [DE]) :

Théorème 1.4.8

On considère des champs de probabilités (μ_x^1) et (μ_x^2) tels que pour tout x , les logarithmes des transformées de Laplace H_x^1 et H_x^2 soient finis sur $\mathbb{R}^d$. Alors, sur l'ensemble $\{\beta | \pi(\beta) = 0\}$, la fonction L_x est égale à la conjuguée de la fonction $\max\{\widehat{H_x^1}, \widehat{H_x^2}\}$, où on a posé $\widehat{H_x^i}(\alpha) = \inf_{(-1)^i t \geq 0} H_x^i(\alpha - tu_1)$.

Démonstration : sur $\{\pi(\cdot) = 0\}$, L_x est la convexifiée des applications f^1 et f^2 où :

$$f^i(\beta) = \begin{cases} L_x^i(\beta) & \text{si } (-1)^i \pi(\beta) \leq 0 \\ +\infty & \text{sinon} \end{cases}$$

Pour chaque i , f^i est convexe et s.c.i. (car $\{\beta | (-1)^i \pi(\beta) \leq 0\}$ est fermé), avec $(f^i)^* \leq (L_x^i)^*$, donc $(f^i)^*$ finie sur $\mathbb{R}^d$. On peut donc appliquer le théorème 1.3.3 à f^1 et f^2 . On en déduit que, sur $\{\pi(\cdot) = 0\}$, on a :

$$L_x = (\max\{(f^1)^*, (f^2)^*\})^*.$$

Calculons la transformée de Legendre de f^i : on note u_1 le vecteur $(1, 0, \dots, 0)$. On a, pour α dans $\mathbb{R}^d$:

$$(f^i)^*(\alpha) = \sup_{\beta, (-1)^i \pi(\beta) \leq 0} [\langle \alpha, \beta \rangle - L_x^i(\beta)].$$

$$\begin{aligned} \text{Or } \inf_{(-1)^i t \geq 0} H_x^i(\alpha - tu_1) &= \inf_{(-1)^i t \geq 0} \sup_{\beta} [\langle \alpha - tu_1, \beta \rangle - L_x^i(\beta)] \\ &= \inf_{t \geq 0} \sup_{\beta} [\langle \alpha, \beta \rangle - L_x^i(\beta) - t (-1)^i \pi(\beta)]. \end{aligned}$$

Si $(-1)^i \pi(\beta) \geq 0$, la quantité $-t (-1)^i \pi(\beta)$ tend vers $-\infty$ quand t tend vers $+\infty$. On peut donc se limiter aux β tels que $(-1)^i \pi(\beta) \leq 0$. Mais alors, $\inf_{t \geq 0} [-t (-1)^i \pi(\beta)]$ vaut 0. On a donc $\inf_{(-1)^i t \geq 0} H_x^i(\alpha - tu_1) = (f^i)^*(\alpha)$. On en déduit que, sur l'ensemble $\{\beta | \pi(\beta) = 0\}$, L_x est égal à $(\max\{\widehat{H_x^1}, \widehat{H_x^2}\})^*$, d'où le théorème. $\square$

Pour $\pi(\beta) = 0$, $L_x(\beta)$ s'écrit donc :

$$L_x(\beta) = \sup_{\alpha} \left[\langle \alpha, \beta \rangle - \max \left\{ \widehat{H}_x^1(\alpha), \widehat{H}_x^2(\alpha) \right\} \right].$$

Remarquons que cette écriture fournit une nouvelle démonstration du fait que L_x est s.c.i. et convexe sur $\{\pi(\cdot) = 0\}$. Elle permet également de constater que la différence entre $\text{conv}(L_x^1, L_x^2)$ et L_x sur l'ensemble $\{\beta | \pi(\beta) = 0\}$ provient de l'écart entre $\max\{H_x^1, H_x^2\}$ et $\max\{\widehat{H}_x^1, \widehat{H}_x^2\}$. En effet, d'après le théorème 1.3.3, on a :

$$\text{conv}(L_x^1, L_x^2) = (\max\{H_x^1, H_x^2\})^*.$$

Sur l'ensemble $\{\beta | (-1)^i \pi(\beta) \geq 0\}$, L_x est la convexifiée des applications $L_{x/\{\beta | \pi(\beta)=0\}}$ et L_x^i qui vérifient les hypothèses du théorème 1.3.3. Elle s'écrit donc, d'après le théorème 1.4.8 :

$$L_x = \left[\max \left\{ \left(\max \left\{ \widehat{H}_x^1, \widehat{H}_x^2 \right\} \right)^{**}, H_x^i \right\} \right]^*, \text{ sur } \{\beta | (-1)^i \pi(\beta) \geq 0\}.$$

Cette écriture fournit une nouvelle démonstration du fait que L_x est s.c.i. et convexe sur $\{\beta | (-1)^i \pi(\beta) \geq 0\}$.

Etudions plus précisément L_x , en fonction de la position des ensembles S_x^i par rapport à la frontière (par rapport à $\{\beta | \pi(\beta) = 0\}$ exactement) :

Exemple 1 Si, pour un x fixé et pour un i égal à 1 ou 2, S_x^i est inclus dans $\{\beta | (-1)^i \pi(\beta) > 0\}$ (vecteurs strictement rentrants dans P^i), alors L_x est égal à L_x^{3-i} sur $\{\beta | \pi(\beta) = 0\}$. Si cette condition est vérifiée dans une région de $\mathbb{R}^d$, il n'y a pas d'oscillations possibles autour de la frontière dans cette région : une fois que le processus est dans P^i , il ne peut retraverser la frontière. Le processus va pouvoir suivre la frontière seulement en restant dans P^{3-i} . Si la condition $S_x^i \subset \{\beta | (-1)^i \pi(\beta) > 0\}$ est vérifiée pour chaque i , le processus ne reste pas au voisinage de la frontière P^0 .

Cela amènera à discuter suivant que l'hypothèse

$$\text{pour chaque } i, \{\beta \in \mathbb{R}^d \mid (-1)^i \pi(\beta) < 0\} \cap S_x^i \neq \emptyset$$

est vérifiée ou non. Il s'agit de l'hypothèse $H_0(x)$ du paragraphe 1.4.2.

Exemple 2 Si, pour un i , S_x^i est inclus dans $\{\beta \mid (-1)^i \pi(\beta) \leq 0\}$, alors L_x est égal à $\text{conv}(L_x^1, L_x^2)$ sur $\{\beta \mid \pi(\beta) = 0\}$.

Si cela est vérifié dans une région de $\mathbb{R}^d$, on verra dans le paragraphe suivant que la partie majoration des inégalités de grandes déviations est alors déjà prouvée. Par contre, on ne peut rien dire à ce stade sur la minoration (un exemple est toutefois traité dans [DIS]: la majoration et la minoration sont obtenues, en utilisant la théorie des solutions de viscosité aux équations d'Hamilton-Jacobi, et non des méthodes probabilistes).

Exemple 3 Dans le cas de la dimension 1 ($d = 1$), P^0 est réduit à $\{0\}$. Suivant le signe de m_0^i , la frontière peut être stable ou instable:

- si $(-1)^i m_0^i < 0$, pour $i = 1$ ou 2 , $L_0(\beta)$ vaut 0 si et seulement si β est égal à 0 ([DE] montre de plus que dans ce cas on a $\text{conv}(L_x^1, L_x^2) = L_x$ et précise que ce résultat ne se généralise pas à une dimension quelconque, mais nous ne nous intéressons qu'aux vecteurs parallèles à P^0 , cf remarque 1.4.9). Si les applications $x \mapsto m_x^i$ sont continues, pour x dans un voisinage de 0, on a encore $(-1)^i m_x^i < 0$. Dans tout ce voisinage, le processus a tendance à revenir vers la frontière $\{0\}$ (en effet, la fonction $L_x^i(\beta)$ est minimale au point m_x^i , vecteur qui ramène vers la frontière).
- si $(-1)^i m_0^i > 0$, pour $i = 1$ ou 2 , $L_x(0)$ est égal au minimum de $L_0^1(0)$ et $L_0^2(0)$, donc est strictement positif. Pour x dans P^i , L_x^i est minimale pour des vecteurs qui éloignent de P^0 . Au voisinage de 0, dans chaque demi-espace, le processus a tendance à s'éloigner de l'origine.

La suite de notre étude permettra de préciser ces résultats.

Pour terminer l'étude de L_x , nous souhaitons savoir si, dans l'expression (1.12) donnant $L_x(\beta)$ pour $\pi(\beta) = 0$, le minimum est souvent réalisé avec des inégalités strictes $\pi(\beta^1) > 0$ et $\pi(\beta^2) < 0$. Supposons que pour β donné, $\text{conv}(L_x^1, L_x^2)(\beta)$ soit strictement inférieur à $L_x(\beta)$. Nous allons montrer qu'alors, si $L_x(\beta)$ est fini, il s'écrit nécessairement $\sum_{i=1}^2 \rho^i L_x^i(\beta^i)$, avec $\pi(\beta^i) = 0$ (c'est-à-dire que le minimum donnant la valeur de $L_x(\beta)$ est atteint avec des vecteurs parallèles à P^0).

La quantité $\text{conv}(L_x^1, L_x^2)(\beta)$ s'écrit $\sum_{i=1}^2 \rho^i L_x^i(\beta^i)$, avec $\beta = \sum_{i=1}^2 \rho^i \beta^i$ et $(-1)^i \pi(\beta^i) > 0$ (pour avoir $\text{conv}(L_x^1, L_x^2)(\beta) < L_x(\beta)$). Supposons :

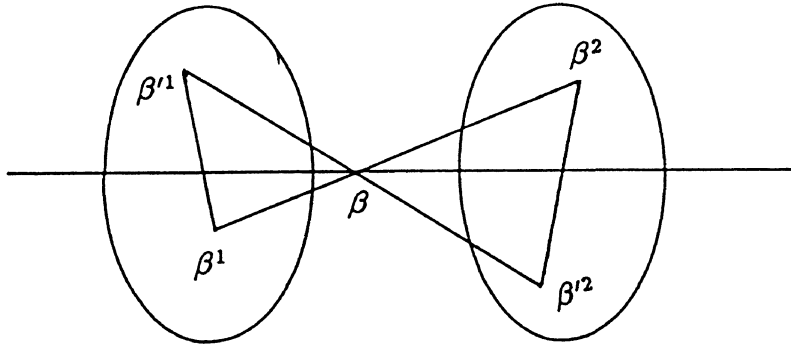
$$L_x(\beta) = \sum_{i=1}^2 \rho^i L_x^i(\beta^i), \text{ avec } \beta = \sum_{i=1}^2 \rho^i \beta^i \text{ et } (-1)^i \pi(\beta^i) \leq 0.$$

Si un des $\pi(\beta'^i)$ est nul, l'autre l'est aussi et on a le résultat. Sinon, $\rho^1\pi(\beta^1)$ est strictement négatif et $\rho^1\pi(\beta'^1)$ strictement positif donc il existe θ non nul tel que $\pi(\theta\rho^1\beta^1 + (1-\theta)\rho'^1\beta'^1) = 0$. Alors $\pi(\theta\rho^2\beta^2 + (1-\theta)\rho'^2\beta'^2)$ vaut 0. On pose

$$\tilde{\beta}^i = \frac{\theta\rho^i\beta^i + (1-\theta)\rho'^i\beta'^i}{\theta\rho^i + (1-\theta)\rho'^i}.$$

Alors $\pi(\tilde{\beta}^i)$ vaut 0 et on a :

$$\sum_{i=1}^2 (\theta\rho^i + (1-\theta)\rho'^i) \tilde{\beta}^i = \theta \sum_{i=1}^2 \rho^i \beta^i + (1-\theta) \sum_{i=1}^2 \rho'^i \beta'^i = \beta.$$



On en déduit que $L_x(\beta)$ est inférieur à $\sum_{i=1}^2 (\theta\rho^i + (1-\theta)\rho'^i) L_x^i(\tilde{\beta}^i)$ donc à :

$$\begin{aligned} & \theta \sum_{i=1}^2 \rho^i L_x^i(\beta^i) + (1-\theta) \sum_{i=1}^2 \rho'^i L_x^i(\beta'^i) \\ &= \theta \operatorname{conv}(L_x^1, L_x^2)(\beta) + (1-\theta) L_x(\beta) < L_x(\beta), \text{ ce qui est absurde} \end{aligned}$$

Cela prouve que dans ce cas, on a :

$$L_x(\beta) = \operatorname{conv}(L_x^1 / \{\pi(\cdot)=0\}, L_x^2 / \{\pi(\cdot)=0\})(\beta). \quad (1.15)$$

Dans le chapitre 3, nous serons gênés par les vecteurs pour lesquels L_x est atteint avec des vecteurs parallèles à P^0 . Nous constatons sur cet exemple que cette situation n'est pas rare.

IV. Propriétés de la fonctionnelle de coût $\Lambda_{0,T}$

Dans la notation 1.4.5, on a défini $\Lambda_{0,T}$ de la façon suivante en un chemin φ absolument continu de $[0, T]$ dans $\mathbb{R}^d$:

$$\Lambda_{0,T}(\varphi) = \int_0^T \left(\sum_{i=1}^2 L_{\varphi(t)}^i(\varphi'(t)) \mathbb{I}_{\varphi(t) \in P^i} + L_{\varphi(t)}(\varphi'(t)) \mathbb{I}_{\varphi(t) \in P^0} \right) dt$$

Le coût mixte $L(\cdot)$ intervient dans cette formule seulement en les t tels que $\varphi(t)$ appartient à P^0 . Ce coût est défini sur les vecteurs parallèles à la frontière comme une fonction convexe et s.c.i. (formule (1.12)). La remarque suivante permet de se limiter à cette formule :

Remarque 1.4.9 *Si φ est un chemin absolument continu, la mesure de Lebesgue de l'ensemble des t tels que $\varphi(t)$ appartient à P^0 et $\pi(\varphi'(t)) \neq 0$ est nulle. Donc, dans le calcul de $\Lambda_{0,T}$, le coût mixte L_x n'intervient que sur des vecteurs parallèles à P^0 . Ainsi, on pourrait se contenter de le définir sur l'ensemble $\{\pi(\cdot) = 0\}$, ensemble où il est convexe. On ne devrait donc pas véritablement être confronté au défaut de convexité souligné dans la démonstration de la proposition 1.4.6, contrairement à ce qui se produit dans [DE] où l'on regarde les vitesses moyennes, sans intégrer la fonction L_x .*

En fait, nous n'utiliserons cette fonctionnelle que sur des chemins suffisamment réguliers au voisinage de la frontière, précisément ceux qui changent un nombre fini de fois d'ensemble P^i . Le coût est, pour un tel chemin φ , l'intégrale de $t \mapsto L_{\varphi(t)}^i(\varphi'(t))$ quand φ est dans P^i , avec $i \neq 0$, et l'intégrale de $t \mapsto L_{\varphi(t)}(\varphi'(t))$ quand φ suit P^0 . On constate que L_x ne sera effectivement utilisé que sur des vecteurs parallèles à P^0 . Pour les autres chemins, les fonctionnelles de coût utilisées seront calculées comme limite de coût de chemins approximants ayant cette régularité.

Nous nous intéressons à d'éventuelles propriétés de semi-continuité de $\Lambda_{0,T}$. Le théorème 3, paragraphe 9.1.4 de [IT] permet de montrer que certaines fonctionnelles sont semi-continues inférieurement. Précisément, adapté aux fonctionnelles de la forme $\varphi \mapsto \int_0^T f(\varphi(t), \varphi'(t)) dt$, il conduit au résultat suivant :

Théorème 1.4.10

Soit $f : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, +\infty]$, mesurable, s.c.i., telle que, pour tout x , $f(x, \cdot)$ soit convexe et $f(x, \cdot)^$ finie sur $\mathbb{R}^d$. On suppose de plus qu'il existe une fonction h de $\mathbb{R}^d$ dans $[0, +\infty]$ telle que h^* soit finie sur $\mathbb{R}^d$ et telle que, pour tout x de $\mathbb{R}^d$, on ait :*

$$f(x, \cdot) \geq h.$$

Alors l'application $\varphi \mapsto \int_0^T f(\varphi(t), \varphi'(t)) dt$ est s.c.i. sur l'ensemble des chemins absolument continus de $[0, T]$ dans $\mathbb{R}^d$ muni de la topologie de la convergence uniforme.

En appliquant par exemple ce résultat à l'application L^i , on en déduit le théorème suivant :

Théorème 1.4.11

On suppose que le champ de probabilités (μ_x^i) vérifie les hypothèses (HC) et (H4) et que les S_x^i sont compacts. Pour tout chemin continu, φ , de $[0, T]$ dans $\mathbb{R}^d$, on pose

$$\Lambda_{0,T}^i(\varphi) = \begin{cases} \int_0^T L_{\varphi(t)}^i(\varphi'(t)) dt & \text{si } \varphi \text{ est absolument continu} \\ +\infty & \text{sinon} \end{cases}$$

La fonctionnelle de coût $\Lambda_{0,T}^i$ ainsi définie est s.c.i. sur $C[0, T]$, l'ensemble des chemins continus sur $[0, T]$.

Démonstration : soit φ un chemin continu sur $[0, T]$ et φ_n une suite de $C[0, T]$ convergant vers φ . On veut montrer $\Lambda_{0,T}^i(\varphi) \leq \liminf_{n \rightarrow +\infty} \Lambda_{0,T}^i(\varphi_n)$. Il suffit de considérer le cas où $\liminf_{n \rightarrow +\infty} \Lambda_{0,T}^i(\varphi_n)$ est fini. Dans ce cas, φ est nécessairement absolument continue ([AR] 4.3) et toutes les applications φ_n , et donc φ , sont à valeurs dans un même compact $\mathcal{U}$, d'après les hypothèses. Quitte à remplacer $L_x^i(\beta)$ par $+\infty$ si x n'appartient pas à $\mathcal{U}$, on peut se contenter de montrer que L^i vérifie les hypothèses du théorème 1.4.10 pour x variant dans $\mathcal{U}$.

Posons $M = \max \{|\beta|, \beta \in S_x^i, x \in \mathcal{U}, i = 1 \text{ ou } 2\}$. Pour x dans $\mathcal{U}$ et α dans $\mathbb{R}^d$, $H_x^i(\alpha)$ est majoré par $\log \int e^{|\alpha|M} \mu_x^i(dy)$ donc par $g(\alpha) = M|\alpha|$. On en déduit que L_x^i est minoré par g^* . Or g^{**} , c'est-à-dire g , est finie sur $\mathbb{R}^d$. Les autres hypothèses sont immédiatement vérifiées (cf théorème 1.2.7 pour la s.c.i. de L^i). Le théorème 1.4.10 prouve alors que $\Lambda_{0,T}^i$ est s.c.i. en φ . $\square$

Remarquons qu'il existe, sous les hypothèses (H), une démonstration directe dans [AR], n'utilisant pas le théorème 1.4.10; elle est basée sur les mêmes éléments que le résultat plus général conduisant au théorème 1.4.10 dans [IT].

Ce résultat ne semble malheureusement pas pouvoir s'étendre à l'application $\Lambda_{0,T}$ car les hypothèses de convexité et de semi-continuité sont perdues au voisinage de la frontière. On ne peut pas non plus montrer ainsi la s.c.i. éventuelle de l'application $\varphi \mapsto \int_0^T {}^h\overline{L}_{\varphi(t)}^i(\varphi'(t))dt$ (à nouveau, problème de la s.c.i. de ${}^h\overline{L}_x^i$, convexifiée d'un nombre infini de fonctions). On peut seulement en déduire que l'application $\varphi \mapsto \int_0^T [\sup\{H_y \mid y \in B(\varphi(t), h)\}]^*(\varphi'(t)) dt$ est s.c.i.; cette propriété est utilisée dans [DEW] (la fonction $[\sup\{H_y \mid y \in B(x, h)\}]^*$ est la convexifiée de ${}^h\overline{L}_x$; elle est inférieure à ${}^h\overline{L}_x$, avec égalité si et seulement si ${}^h\overline{L}_x$ est s.c.i.).

Cependant on peut montrer ainsi que $\Lambda_{0,T}$ est s.c.i. sur l'ensemble des chemins absolument continus qui sont tous à valeurs dans un même P^i :

il suffit de montrer que $\Lambda_{0,T}$ est s.c.i. sur l'ensemble des chemins absolument continus tracés sur P^0 . Pour cela, on utilise que, pour $x \in P^0$, L_x est convexe et s.c.i. sur l'ensemble des vecteurs parallèles à P^0 . De plus on a, sur cet ensemble :

$$L_x = (\max\{\widehat{H}_x^1, \widehat{H}_x^2\})^*$$

donc $(L_x)^*$ finie (inférieure à $\max\{\widehat{H}_x^1, \widehat{H}_x^2\}$ et $\widehat{H}_x^i$ est fini dès que H_x^i l'est) et, pour x dans $\mathcal{U}$, $L_x \geq g^*$, comme ci-dessus. On utilise alors à nouveau le théorème 1.4.10.

1.4.4 Limites de l'utilisation d'estimations antérieures à ce travail

Des résultats de grandes déviations concernant les processus d'apprentissage lent à statistiques discontinues le long d'un hyperplan peuvent être obtenus en utilisant les estimations établies par [Du88] et [DEW] ; nous allons montrer qu'en général, compte-tenu de la fonction de coût que nous conjecturons (paragraphe 1.4.3), ces estimations ne suffisent pas à obtenir les résultats souhaités ici.

1. Les résultats proposés par [Du88] ont été rappelés dans le paragraphe 1.1.2. Pour obtenir l'hypothèse (A2) de [Du88], on cherche une fonction réelle $H(x, \alpha)$, convexe en α , vérifiant :

pour tous x dans $\mathbb{R}^d$ et $\gamma > 0$, il existe $\delta > 0$ tel que, pour toute suite (y_k) de $B(x, \delta)$ (boule de centre x et rayon δ) et pour tout α de $\mathbb{R}^d$, on ait :

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{E} \left(\exp \left\langle \alpha, \sum_{k=1}^n \sum_{i=1}^2 \mathbb{I}_{\{y_k \in P^{i1}\}} X_k^i(y_k) \right\rangle \right) \leq H(x, \alpha) + \gamma(1 + |\alpha|)$$

Le membre de gauche de cette inégalité est majoré par $\sum_{i=1}^2 \frac{1}{n} \sum_k \mathbb{I}_{\{y_k \in P^{i1}\}} H_{y_k}^i(\alpha)$ (où $P^{11} = P^1 \cup P^0$ et $P^{22} = P^2$) mais on cherche un majorant indépendant de la suite (y_k) (puisque l'on veut l'utiliser pour $(y_k) = (Y_k^r)$, cela revient, avec des hypothèses de continuité sur les H^i , à chercher une expression déterministe majorant $\sum_{i=1}^2 r^i H_x^i(\alpha^i)$ où r^i est une variable aléatoire correspondant à la proportion de pas soumis au champ i). La seule façon d'obtenir une telle inégalité dans notre cas est, si l'on se réfère à la démonstration du lemme 2.1.1, de prendre, pour $H(x, \cdot)$, la fonction $\max\{H_x^1, H_x^2\}$ qui est effectivement convexe.

Quand l'hypothèse (A2) est vérifiée, on obtient une majoration de grande déviation utilisant comme fonction de coût l'intégrale de $J(\cdot, \cdot)$ le long des chemins, où $J(x, \cdot)$ est la transformée de Legendre de $\max\{H_x^1, H_x^2\}$, c'est-à-dire $\text{conv}(L_x^1, L_x^2)$. Or, pour obtenir le coût L_x cherché, il aurait fallu écrire (A2) avec $H(x, \alpha)$ égal à $\max\{\widehat{H}_x^1(\alpha), \widehat{H}_x^2(\alpha)\}$ dont la transformée de Legendre est bien L_x , au moins sur $\{\pi(\cdot) = 0\}$ (paragraphe précédent). Le résultat de Dupuis et Ellis est donc insuffisamment précis pour nous, sauf bien sûr lorsque, pour tout x , $\text{conv}(L_x^1, L_x^2)$ et L_x coïncident.

Le remplacement de $\max\{H_x^1, H_x^2\}$ par $\max\{\widehat{H}_x^1, \widehat{H}_x^2\}$ traduit, tout comme la présence des conditions $(-1)^i \pi(\beta^i) \leq 0$ dans le coût, la contrainte physique "faire un pas de type i quand on est dans P^i ". Pour pouvoir majorer par $\max\{\widehat{H}_x^1(\alpha), \widehat{H}_x^2(\alpha)\}$, il faudrait pouvoir traduire cette contrainte sur les pas; elle apparaîtrait ainsi dans le terme de gauche de l'hypothèse (A2).

En fait, $\text{conv}(L_x^1, L_x^2)$ est probablement la plus petite fonction convexe permettant d'écrire l'inégalité de majoration des grandes déviations.

Pour obtenir l'hypothèse (A.3), la fonction convexe en α cherchée, $H(x, \alpha)$, doit vérifier au moins :

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{E} \left(\exp \left\langle \alpha, \sum_{k=0}^{n-1} X_k^i(y_k) \right\rangle \right) \geq H(x, \alpha) - C_K \delta(|\alpha| + 1)$$

Ici, on ne peut obtenir cette inégalité qu'en prenant, pour $H(x, \cdot)$, la fonction convexifiée de $\min\{H_x^1, H_x^2\}$ car elle doit être convexe et inférieure à cette fonction. Sa conjuguée est alors plus grande que $L^1 \vee L^2$, conjuguée de $\min\{H_x^1, H_x^2\}$. Ce coût est très éloigné du coût L_x que nous voulons obtenir.

2. Dans [DEW] (paragraphe 4), on considère un processus d'apprentissage lent comme dans la définition 1.1.1, sans aucune hypothèse de continuité sur le champ de probabilités (μ_x) . On s'intéresse seulement à des résultats de majoration. On pose

$$h(x, \alpha) = \liminf_{\delta \rightarrow 0} \sup_{y \in B(x, \delta)} H_y(\alpha),$$

on prend $l(x, \cdot)$ égale à la transformée de Legendre de $h(x, \cdot)$ et on pose, pour φ absolument continue avec $\varphi(0) = x$:

$$I_x(\varphi) = \int_0^T l(\varphi(t), \varphi'(t)) dt.$$

Si φ n'est pas absolument continue, on prend encore $I_x(\varphi)$ égal à $+\infty$. On a le résultat suivant, en prenant φ^τ comme dans (1.2) :

Théorème 1.4.12 ([DEW])

On considère un processus d'apprentissage lent, $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à un champ de probabilités (μ_x) . On note H_x la log-Laplace de μ_x et on suppose qu'il existe une application $\bar{h}$ finie sur $\mathbb{R}^+$ telle que pour tous x, α dans $\mathbb{R}^d$, on ait $H_x(\alpha) \leq \bar{h}(|\alpha|)$. Alors, pour tout compact $\mathcal{K}$ et tout fermé G de $C[0, T]$, on a, uniformément en y_0 de $\mathcal{K}$:

$$\limsup_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi^\tau \in G) \leq - \inf_{\varphi \in G} I_{y_0}(\varphi)$$

Remarquons que l'hypothèse $H_x(\alpha) \leq \bar{h}(|\alpha|)$ pour tous x, α permet d'utiliser le théorème 1.4.10 pour montrer que la fonctionnelle I_x est s.c.i. (puisque l'on a $l(x, \cdot) \geq [\bar{h}(|\cdot|)]^*$). Pour les processus d'apprentissage lent à statistiques discontinues le long d'un hyperplan que nous considérons, en supposant les probabilités à support compact, cette hypothèse est vérifiée pour x dans un compact, ce qui est suffisant. On a :

$$H(x, \alpha) = \log \int_{\mathbb{R}^d} e^{(\alpha, y)} \mu_x(dy) = \begin{cases} H_x^1(\alpha) & \text{si } x \in P^1 \cup P^0 \\ H_x^2(\alpha) & \text{si } x \in P^2 \end{cases}$$

Pour x dans P^0 , la fonction $l(x, \cdot)$ vérifie sur l'ensemble $\{\beta | \pi(\beta) = 0\}$:

$$l(x, \cdot) = (\max\{H_x^1, H_x^2\})^* = \text{conv}(L_x^1, L_x^2).$$

De nouveau, le résultat est insuffisant pour obtenir la majoration précise cherchée ici et l'amélioration à apporter consiste à remplacer, dans le calcul du coût le long des chemins, $\text{conv}(L_x^1, L_x^2)$ par L_x pour les vitesses parallèles à la frontière.

1.5 Conclusion, bilan des hypothèses faites sur le modèle

Nous exposons dans cette partie les hypothèses précises faites sur le modèle. Dans toute l'étude, on considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , associé à des champs de probabilités (μ_x^1) et (μ_x^2) (définition 1.4.1). Suivant l'inégalité de grandes déviations cherchée, on envisage deux séries d'hypothèses sur ces champs de probabilités :

Hypothèses (Hmaj) :

Pour chaque i , le champ de probabilités (μ_x^i) vérifie :

(H2) l'enveloppe convexe fermée du support de μ_x^i, S_x^i , est compacte et d'intérieur non vide

(HC) $\forall x \in \mathbb{R}^d, \forall \gamma > 0, \exists \delta > 0$ tel que, si $|x - y| \leq \delta, \forall \alpha \in \mathbb{R}^d$,

$$|H_x^i(\alpha) - H_y^i(\alpha)| \leq \gamma |\alpha|$$

D'autre part, on a

(H4) il existe $h > 0$ tel que, pour tout $T > 0$ et tout compact $\mathcal{K}$, il existe un compact contenant tous les $\varphi([0, T])$, dès que φ est un chemin tel que $\varphi(0)$ appartient à $\mathcal{K}$ et qui est tangent au champ de convexes égal à S_x^i si $x \in (P^i)^{-h}$ et $\text{Conv}(S_x^1, S_x^2)$ si $x \in (P^0)^h$.

Hypothèses (Hmin) :

Pour chaque i , le champ de probabilités (μ_x^i) vérifie :

(H2) l'enveloppe convexe fermée du support de μ_x^i, S_x^i , est compacte et d'intérieur non vide

(HC) $\forall x \in \mathbb{R}^d, \forall \gamma > 0, \exists \delta > 0$ tel que, si $|x - y| \leq \delta, \forall \alpha \in \mathbb{R}^d$,

$$|H_x^i(\alpha) - H_y^i(\alpha)| \leq \gamma |\alpha|$$

(i) $(x, \beta) \mapsto \nabla L_x^i(\beta)$ est continue sur $\bigcup x \times \overset{\circ}{S}_x^i$

(ii) $\forall \alpha \in \mathbb{R}^d, \forall 1 \leq k \leq d, x \mapsto \frac{\partial^2 H_x^i}{\partial \alpha_k^2}(\alpha)$ est continue.

D'autre part, on a :

(H4) il existe $h > 0$ tel que, pour tout $T > 0$ et tout compact $\mathcal{K}$, il existe un compact contenant tous les $\varphi([0, T])$, dès que φ est un chemin tel que $\varphi(0)$ appartient à $\mathcal{K}$ et qui est tangent au champ de convexes égal à S_x^i si $x \in (P^i)^{-h}$ et $\text{Conv}(S_x^1, S_x^2)$ si $x \in (P^0)^h$.

Par convention, quand on considère un seul champ de probabilités (μ_x^i) , on dit qu'il vérifie les hypothèses (Hmaj) (ou (Hmin)) quand les premières de ces hypothèses sont vérifiées tandis que (H4) est vraie pour ce seul champ, c'est-à-dire dans son énoncé du paragraphe 1.1.1, ce qui revient à l'hypothèse globale (H4) ci-dessus, considérée pour deux champs égaux à (μ_x^i) . Il n'est pas certain que cette hypothèse globale soit une conséquence de (H4) pour chacun des deux champs.

Rappelons que (HC) implique que l'application $x \mapsto S_x^i$ est continue (c'est-à-dire (H3) pour le champ (μ_x^i)). Les trajectoires du processus (φ^τ) que nous allons étudier ont leur origine dans un compact $\mathcal{K}$ donné, et nous les étudions pendant une durée finie T . Elles sont tangentes, pour h convenable, au champ de convexes apparaissant dans l'hypothèse (H4). Cette hypothèse implique, d'après (H3), que :

$$\left. \begin{array}{l} \text{à tout compact } \mathcal{K} \text{ de } \mathbb{R}^d, \text{ et tout réel } T > 0, \text{ on peut associer un} \\ \text{compact } \mathcal{U} \text{ contenant un voisinage de } \varphi_{[0,T]}^\tau \text{ pour tout } \tau > 0 \text{ et tout} \\ Y_0^\tau \text{ dans } \mathcal{K} \text{ (et des chemins d'origine dans } \mathcal{K} \text{ et tangents au champ de} \\ \text{convexes cité dans (H4)) et un réel } M \text{ tel que, pour tous } x \in \mathcal{U} \text{ et} \\ \beta \in S_x^i, \text{ avec } i = 1 \text{ ou } 2, \text{ on ait : } |\beta| \leq M. \end{array} \right\} \quad (1.16)$$

On utilisera, de plus, pour obtenir certains résultats, l'hypothèse suivante :

$$\begin{aligned} \text{(HLip)} : \quad & \forall \mathcal{K} \text{ compact de } \mathbb{R}^d, \exists C_{\mathcal{K}} \text{ tel que } \forall x, y \in \mathcal{K}, \forall \alpha \in \mathbb{R}^d, \forall i = 1, 2, \\ & |H_x^i(\alpha) - H_y^i(\alpha)| \leq C_{\mathcal{K}} |x - y| |\alpha|. \end{aligned}$$

Comparons ces hypothèses à celle de [Du88] et [AR]. Les hypothèses (Hmaj) impliquent (A1), sur un compact contenant un voisinage des trajectoires d'origine dans un compact, et (A2). L'hypothèse (HLip) implique (A3). A l'inverse, d'après le théorème 1.2.8, les hypothèses (Hmaj) et (Hmin) (sauf éventuellement (H4)) sont impliquées par les hypothèses (H) de [AR] pour chacun des deux champs de probabilités (μ_x^i) . En particulier, (i) et (ii) sont des conséquences de (H1) dès que (H2) est vraie (théorème 1.2.11 ; rappelons que (H1) est la continuité de l'application $x \mapsto \mu_x^i$ et (H3) celle de $x \mapsto S_x^i$).

Pour obtenir l'inégalité de majoration, nous nous inspirons des méthodes de [AR], mais comme dans [Du88], l'hypothèse (HC) suffit ((HLip) n'est pas nécessaire). En revanche, par rapport à [Du88], nous conservons l'hypothèse que les S_x^i sont d'intérieur non vide (on peut s'y ramener dans un cadre plus général en se plaçant sur des ouverts où la dimension de l'espace engendré par chaque S_x^i est constante en x). Pour l'inégalité de minoration, l'hypothèse (HLip) servira seulement à montrer

une égalité du type de $\Lambda^{reg} = \Lambda$ dans [AR]. Enfin, toutes ces propriétés pourront être envisagées seulement sur une partie de $\mathbb{R}^d$, par exemple un ouvert contenant les trajectoires.

Hypothèses liées à la présence d'une frontière de discontinuité du champ de probabilités

Les hypothèses précédentes, sauf (H4), portent sur chacun des deux champs de probabilités et seraient les mêmes en l'absence de la frontière. Cependant, le comportement du processus au voisinage de P^0 dépend fortement de la position des supports des probabilités par rapport à l'hyperplan vectoriel direction de P^0 . Ainsi, on peut distinguer deux modes selon que chaque support S_x^i , pour x sur P^0 , contient des vecteurs strictement rentrants dans P^{3-i} ou non. Nous énonçons alors une hypothèse, analogue à celle de [DE], permettant aux deux demi-espaces de communiquer (cette hypothèse est à considérer seulement pour les x voisins de P^0).

Pour x dans $\mathbb{R}^d$, on note $H_0(x)$ l'hypothèse

$$\text{pour chaque } i, \quad \{\beta \in \mathbb{R}^d \mid (-1)^i \pi(\beta) < 0\} \cap S_x^i \neq \emptyset.$$

Le premier mode de comportement du processus, au voisinage d'un point x de P^0 , correspond au cas où $H_0(x)$ est vérifiée. Il y a alors mélange des deux champs de probabilités (oscillation du processus entre les deux demi-espaces). Nous montrons dans les chapitres qui suivent que c'est seulement dans ce cas qu'apparaît une fonctionnelle de coût mixte, différente de Λ^1 ou Λ^2 , dans les résultats de grandes déviations.

Dans le second mode, obtenu quand $H_0(x)$ est fautive, localement, le processus traverse au plus une fois P^0 . Seules les fonctions L_x^1 et L_x^2 interviennent dans les résultats de grandes déviations et non le coût mixte L_x .

Apport des méthodes de localisation

L'ensemble des points x d'un voisinage ouvert de P^0 pour lesquels l'hypothèse $H_0(x)$ est vérifiée est un ouvert U_0 de $\mathbb{R}^d$. On note, pour $i = 1$ ou 2 , U_i l'ouvert des x pour lesquels S_x^i est inclus dans l'ensemble $\{\beta \mid (-1)^i \pi(\beta) > 0\}$. Dans les chapitres 2 et 3, nous donnons des résultats concernant les trajectoires qui restent dans un des ouverts U_0 , U_1 ou U_2 .

On dit que l'hypothèse (H_0) est vérifiée quand $H_0(x)$ est vérifiée pour tout x de $\mathbb{R}^d$. Dans les chapitres 2 et 3, nous supposons soit que l'hypothèse (H_0) est

vérifiée, soit, avec $i = 1$ ou 2 , que pour tout x de $\mathbb{R}^d$, S_x^i est inclus dans l'ensemble $\{\beta | (-1)^i \pi(\beta) > 0\}$. Le cas où (H_0) est vérifiée est le plus difficile. Dans le chapitre 4, nous étudions les points limites, c'est-à-dire les x appartenant à la frontière de ces ouverts.

Grâce aux méthodes de localisation, on peut en déduire le cas général : entre les instants d'entrée dans un ouvert V tel que $\bar{V}$ est par exemple inclus dans U_0 , on utilise les formules de coût trouvées dans le cas où (H_0) est vérifiée.

Enfin, nous pouvons supposer (quitte à étudier seulement les trajectoires restant $(P^0)^h$) que l'hypothèse (H_4) est vérifiée pour tout h .

Chapitre 2

Grandes déviations trajectorielles : inégalité de majoration

Dans ce chapitre, nous établissons, pour un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (Hmaj), l'inégalité de majoration dans le principe de grandes déviations (1.3). Nous montrons que deux types de résultats sont observables, le premier dans le cas où l'hypothèse (H_0) est vérifiée :

On note $C^{f,1}[0, T]$ l'ensemble des chemins ψ continus pour lesquels il existe une partition finie $0 = t_0 \leq t_1 \leq \dots \leq t_p = T$ telle que $\psi(]t_j, t_{j+1}[)$ est inclus dans P^{i_j} , avec $i_j = 0, 1$ ou 2 et, si $i_j = 0$, ψ est de classe C^1 par morceaux sur $]t_j, t_{j+1}[$.

Théorème 2.0.1

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (Hmaj) et (H_0) . On appelle φ^τ le chemin aléatoire continu défini sur $[0, T]$ par : $\varphi^\tau(k\tau) = Y_k^\tau$ pour $0 \leq k\tau < T + \tau$ et φ^τ est affine sur $[k\tau; (k+1)\tau]$.

Alors, pour toute partie borélienne G de $C[0, T]$, tout compact $\mathcal{K}$ de $\mathbb{R}^d$ et tout $\varepsilon > 0$, il existe τ_0 tel que $\tau \leq \tau_0$ implique la majoration P -ps sur $\{Y_0^\tau \in \mathcal{K}\}$:

$$\tau \log P(\varphi_{/[0, T]}^\tau \in G | \mathcal{F}_0) \leq -\Lambda(\overline{G} \cap C^{f,1}[0, T]) + \varepsilon$$

Le second type de résultat traite le cas où, avec $i = 1$ ou 2 , pour tout x de $\mathbb{R}^d$, S_x^i est inclus dans l'ensemble $\{\beta | (-1)^i \pi(\beta) > 0\}$. Précisément, par exemple dans le cas $i = 1$, nous montrons le résultat suivant :

on note $C^{T_{rav}}$ l'ensemble des chemins tels que $\{t \in [0, T] | \varphi(t) \in P^0\}$ contient au plus un élément ; on a alors :

Théorème 2.0.2

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (H_{maj}) , et tels que, pour tout x de $\mathbb{R}^d$, S_x^1 soit inclus dans l'ensemble $\{\beta | \pi(\beta) < 0\}$. Alors, pour toute partie borélienne G de $C[0, T]$, tout compact $\mathcal{K}$ de $\mathbb{R}^d$ et tout $\varepsilon > 0$, il existe τ_0 tel que $\tau \leq \tau_0$ implique les majorations suivantes :

$$P\text{-ps sur } \{Y_0^\tau \in (P^1 \cup P^0) \cap \mathcal{K}\}, \quad \tau \log P(\varphi_{[0, T]}^\tau \in G) \leq -\Lambda^1(\overline{G}) + \varepsilon$$

$$P\text{-ps sur } \{Y_0^\tau \in P^2 \cap \mathcal{K}\}, \quad \tau \log P(\varphi_{[0, T]}^\tau \in G) \leq -\Lambda(\overline{G} \cap C^{T_{rav}}) + \varepsilon$$

De façon classique, on note, pour x dans $\mathbb{R}^d$, P_x la probabilité étant donné $Y_0^\tau = x$ et $\mathbb{E}_x$ l'espérance correspondante.

Le chapitre est structuré de la façon suivante : l'essentiel de l'étude (parties 2.1 à 2.5) porte sur le cas où l'hypothèse (H_0) est vérifiée et démontre le théorème 2.0.1. Dans la première partie, on étudie la position atteinte après a pas, pour a grand. La deuxième partie énonce un premier théorème sur les trajectoires, qui est démontré dans la troisième partie (on montre d'abord un résultat avec une fonctionnelle approchée puis on passe à la limite). Ce théorème fait intervenir une fonctionnelle de coût régularisée, qui est calculée, dans la quatrième partie, sur l'ensemble $C^{f,1}[0, T]$. Ce calcul permet, dans la cinquième partie, d'écrire le théorème 2.0.1, avec la fonctionnelle de coût $\Lambda(\cdot \cap C^{f,1}[0, T])$ sur les boréliens, plus directement liée à Λ . Une dernière partie étudie le cas où, avec $i = 1$ ou 2 , on a : pour tout x de $\mathbb{R}^d$, S_x^i est inclus dans l'ensemble $\{\beta | (-1)^i \pi(\beta) > 0\}$, et conduit au deuxième type de résultat (théorème 2.0.2).

2.1 Résultats sur les vitesses moyennes

Nous nous intéressons à la moyenne de a pas consécutifs, où a est un entier fixé. On pose, pour $l \geq 1$, $A_l^1 = \{Y_l^\tau \in P^1 \cup P^0\}$ et $A_l^2 = \{Y_l^\tau \in P^2\}$ et on définit

les variables aléatoires suivantes, pour $i = 1$ ou 2 :

$$\begin{cases} r^i &= \frac{1}{a} \sum_{l=0}^{a-1} \mathbb{I}_{A_l^i}, \\ w^i &= \frac{1}{ar^i} \sum_{l=0}^{a-1} \mathbb{I}_{A_l^i} X_l^i(Y_l^\tau) \end{cases}$$

r^i est la proportion, parmi les a pas, de ceux qui sont soumis au champ de probabilités μ^i . Le vecteur w^i est la moyenne de ces pas, ou encore la vitesse moyenne du processus quand il est soumis au champ i (on considère que les a pas sont faits pendant la durée $a\tau$). On a ainsi : $Y_a^\tau = Y_0^\tau + a\tau(r^1 w^1 + r^2 w^2)$. Les variables r^i et w^i sont $\mathcal{F}_a$ -mesurables.

On note $\|\cdot\|_\infty$ la borne supérieure dans $L_\infty(\Omega, P)$. On considère un réel positif T et un compact $\mathcal{K}$ de $\mathbb{R}^d$. On leur associe $\mathcal{U}$ et M comme dans (1.16).

Lemme 2.1.1

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (Hmaj). Soient α^i , $1 \leq i \leq 2$, dans $\mathbb{R}^d$, et a entier. Pour tous $\tau > 0$, y_0 dans $\mathcal{K}$ et pour toute variable aléatoire $\mathcal{F}_{a-1}$ -mesurable $F : \Omega \rightarrow [0, +\infty[$, on a :

$$\mathbb{E}_{y_0} \left(F \exp a \sum_{i=1}^2 r^i \langle \alpha^i, w^i \rangle \right) \leq \left\| F \exp a \sum_{i=1}^2 r^i \sup_{y \in B(y_0, a\tau M)} H_y^i(\alpha^i) \right\|_\infty$$

Démonstration (analogue de [DE] 6.1, adapté au cas de champs de probabilités non constants) :

Pour l dans $\mathbb{N}$, on pose $M_l = \exp \left[\sum_{i=1}^2 \mathbb{I}_{A_l^i} \left(\langle \alpha^i, X_l^i(Y_l^\tau) \rangle - H_{Y_l^\tau}^i(\alpha^i) \right) \right]$

et $Z_l = M_0 M_1 \dots M_{l-1}$.

$$\begin{aligned} \text{Alors, on a : } Z_a &= \exp \sum_{i=1}^2 \left[\left\langle \alpha^i, \sum_{l=0}^{a-1} \mathbb{I}_{A_l^i} X_l^i(Y_l^\tau) \right\rangle - \sum_{l=0}^{a-1} \mathbb{I}_{A_l^i} H_{Y_l^\tau}^i(\alpha^i) \right] \\ &= \exp \sum_{i=1}^2 \langle \alpha^i, ar^i w^i \rangle \cdot \exp - \sum_{i=1}^2 \sum_{l=0}^{a-1} \mathbb{I}_{A_l^i} H_{Y_l^\tau}^i(\alpha^i). \end{aligned}$$

D'autre part, $(Z_l)_{l \in \mathbb{N}}$ est une $(\mathcal{F}_l)$ -martingale de moyenne 1. Le terme à majorer s'écrit alors :

$$\mathbb{E} \left(F Z_a \prod_{l=0}^{a-1} \exp \sum_{i=1}^2 \mathbb{I}_{A_l^i} H_{Y_l^\tau}^i(\alpha^i) \right) = \mathbb{E} \left(F \mathbb{E} (Z_a | \mathcal{F}_{a-1}) \prod_{l=0}^{a-1} \exp \sum_{i=1}^2 \mathbb{I}_{A_l^i} H_{Y_l^\tau}^i(\alpha^i) \right)$$

On le majore donc par :

$$\left\| F \exp \sum_{l=0}^{a-1} \sum_{i=1}^2 \mathbb{I}_{A_l^i} H_{Y_l^\tau}^i(\alpha^i) \right\|_\infty \mathbb{E}(Z_a).$$

Or, pour $0 \leq l < a$, $H_{Y_l^\tau}^i$ est inférieure à $\sup_{y \in B(y_0, a\tau M)} H_y^i$, puisque, d'après le choix de M , les Y_l^τ restent dans $B(y_0, a\tau M)$. $\square$

Remarquons que ce lemme correspond pratiquement à l'hypothèse (A2) de [Du88], avec un majorant qui ne peut être exprimé que comme la borne supérieure d'une variable aléatoire (sinon, cf paragraphe 1.4.4, on n'aboutit pas à la bonne fonctionnelle de coût). On aurait pu prendre la conclusion de ce lemme comme hypothèse. On a choisi de faire des hypothèses explicites sur les variations des applications H^i .

Définition 2.1.2 Pour $i = 1$ ou 2 , $\theta \in [0, 1]$ et $x \in \mathbb{R}^d$, on définit l'application affine suivante :

$$\begin{aligned} h_{x,\theta}^i : \mathbb{R}^d &\rightarrow \mathbb{R}^d \\ v &\mapsto \theta m_x^i + (1 - \theta)v \end{aligned}$$

Lemme 2.1.3

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, vérifiant les hypothèses (Hmaj). Pour tous ϵ, δ, θ strictement positifs, il existe $\delta_1 > 0$ tel que, pour $R \geq 0$ avec $R + a\tau M \leq \delta_1$ et x tel que $B(x, R) \subset \mathcal{K}$, il existe un entier N tel que, pour $a \geq N$, $I \geq 0$ et y_0 dans $B(x, R)$, on ait la majoration :

$$P_{y_0} \left(\sum_{i=1}^2 r^i L_x^i(h_{x,\theta}^i(w^i)) \geq I \text{ et } h_{x,\theta}^i(w^i) \in (S_x^i)^{-\delta}, \text{ pour } i = 1, 2 \right) \leq e^{-a(I-\epsilon)}.$$

Remarque: le fait de se placer sur $(S_x^i)^{-\delta}$ (notation 1.1.2) permet de contrôler les variations de L^i et de majorer uniformément en I .

Démonstration: soit ϵ et $\delta > 0$; puisque (HC) est vérifiée, d'après le théorème 1.2.8, $L^i(\cdot)$ est uniformément continue sur $\bigcup_{y \in \mathcal{U}} y \times (S_y^i)^{-\frac{\delta}{2}}$. Il existe donc $\delta' > 0$ tel que :

$$\forall y \in \mathcal{U}, \forall \beta, \beta' \in (S_y^i)^{-\delta/2}, |\beta - \beta'| < \delta' \Rightarrow |L_y^i(\beta) - L_y^i(\beta')| < \frac{\epsilon}{4}. \quad (2.1)$$

Considérons x tel que $B(x, R)$ soit inclus dans $\mathcal{K}$. Pour évaluer la moyenne des a pas, on va sommer sur les valeurs prises par r^1 , r^2 , w^1 et w^2 . On recouvre $[0, 1]$ par des intervalles $\mathcal{I}_{j,i} = [p_{j,i}\delta', (p_{j,i} + 1)\delta']$ et $(S_x^i)^{-\delta}$ par des boules ouvertes, B_l^i , de diamètre inférieur à δ' et incluses dans $(S_x^i)^{-\frac{\delta}{2}}$:

$$\begin{cases} [0, 1] \subset \bigcup_j \mathcal{I}_{j,i} \\ (S_x^i)^{-\delta} \subset \bigcup_{1 \leq l \leq M_i} B_l^i \end{cases} \quad (2.2)$$

Notation 2.1.4 On pose $J_x(\delta) = \max \{L_x^i(\beta) \mid \beta \in (S_x^i)^{-\delta}, i = 1 \text{ ou } 2\}$.

Alors, en choisissant les boules B_l^i de manière à ce que chacune rencontre $(S_x^i)^{-\delta}$, on a: $L_x^i(\overline{B_l^i}) \leq J_x(\delta) < +\infty$.

Considérons y_0 dans $\mathcal{K}$ et $\theta > 0$. Le terme

$$\mathcal{A} = P_{y_0} \left(\sum_{i=1}^2 r^i L_x^i(h_{x,\theta}^i(w^i)) \geq I \text{ et } h_{x,\theta}^i(w^i) \in (S_x^i)^{-\delta}, \text{ pour } i = 1, 2 \right)$$

est majoré par la somme finie suivante:

$$\mathcal{A} \leq \sum_{j^1, j^2, l^1, l^2} P_{y_0} \left(\sum_{i=1}^2 r^i L_x^i(h_{x,\theta}^i(w^i)) \geq I, r^i \in \mathcal{I}_{j^i, i} \text{ et } h_{x,\theta}^i(w^i) \in B_{l^i}^i, \text{ pour } i = 1, 2 \right).$$

Dans cette somme, la probabilité correspondant au quadruplet (j^1, j^2, l^1, l^2) est non nulle seulement si (j^1, j^2, l^1, l^2) vérifie les deux conditions suivantes:

$$\begin{cases} \sum_{i=1}^2 (p_{j^i, i} + 1)\delta' \left[L_x^i(\overline{B_{l^i}^i}) + \frac{\varepsilon}{4} \right] > I \\ \sum_{i=1}^2 p_{j^i, i}\delta' \leq 1. \end{cases}$$

En effet, sur $\overline{B_{l^i}^i}$, L_x^i varie, d'après (2.1), d'au plus $\frac{\varepsilon}{4}$ puisque le diamètre de $B_{l^i}^i$ est inférieur à δ' . On appelle $\mathcal{E}$ l'ensemble des (j^1, j^2, l^1, l^2) vérifiant ces deux conditions.

En notant $P_{j^1, j^2, l^1, l^2} = P_{y_0} (r^i \in \mathcal{I}_{j^i, i} \text{ et } w^i \in (h_{x,\theta}^i)^{-1}(B_{l^i}^i), \text{ pour } i = 1, 2)$, on obtient:

$$\mathcal{A} \leq \sum_{(j^1, j^2, l^1, l^2) \in \mathcal{E}} P_{j^1, j^2, l^1, l^2}.$$

On pose $C_{l^i}^i = (h_{x,\theta}^i)^{-1}(B_{l^i}^i)$. La condition $w^i \in C_{l^i}^i$ implique que $C_{l^i}^i \cap S_x^i$ est non vide donc aussi $C_{l^i}^i \cap \overset{\circ}{S}_x^i$. En effet, si β appartient à $C_{l^i}^i \cap S_x^i$, $C_{l^i}^i$, qui est ouvert,

contient un voisinage de β , et comme S_x^i est convexe, il existe une demi-droite issue de β qui rencontre $\overset{\circ}{S}_x^i$ et C_{li}^i . Donc, P_{j^1, j^2, l^1, l^2} est nul dès que $L_x^1(C_{l^1}^1)$ ou $L_x^2(C_{l^2}^2)$ vaut $+\infty$ (puisque L_x^i est fini sur $\overset{\circ}{S}_x^i$).

Il reste donc à majorer P_{j^1, j^2, l^1, l^2} quand $L_x^1(C_{l^1}^1)$ et $L_x^2(C_{l^2}^2)$ sont finis. On se place dans cette situation. Comme dans [DE] (lemme 6.7), on fait intervenir les ensembles suivants :

pour $i = 1$ ou 2 , $\alpha \in \mathbb{R}^d$, $y \in \mathcal{U}$ et $C \subset \mathbb{R}^d$ tel que $L_y^i(C) < +\infty$, on pose :

$$\Psi_y^i(C, \alpha) = \left\{ v \in \mathbb{R}^d \mid \langle \alpha, v \rangle - H_y^i(\alpha) > L_y^i(C) - \frac{\varepsilon}{4} \right\}.$$

Comme $L_y^i(.) = \sup_{\alpha \in \mathbb{R}^d} \{ \langle \alpha, . \rangle - H_y^i(\alpha) \} = \sup_{\alpha \in \mathbb{R}^d \setminus \{0\}} \{ \langle \alpha, . \rangle - H_y^i(\alpha) \}$,

on a alors : $C \subset \left\{ v \in \mathbb{R}^d \mid L_y^i(v) > L_y^i(C) - \frac{\varepsilon}{4} \right\} = \bigcup_{\alpha \in \mathbb{R}^d \setminus \{0\}} \Psi_y^i(C, \alpha).$

On utilise cette propriété pour chaque $\overline{C_{li}^i}$, avec $1 \leq i \leq 2$, qui vérifie $L_x^i(\overline{C_{li}^i}) < +\infty$. Puisque cet ensemble est compact, il existe un nombre fini M_{li} de vecteurs $\alpha_{li, m}^i$, non nuls (dépendant de x) tels que $\overline{C_{li}^i}$ soit contenu dans l'union des $\Psi_x^i(\overline{C_{li}^i}, \alpha_{li, m}^i)$:

$$\overline{C_{li}^i} \subset \bigcup_{1 \leq m \leq M_{li}} \Psi_x^i(\overline{C_{li}^i}, \alpha_{li, m}^i) \quad (2.3)$$

On peut alors écrire :

$$P_{j^1, j^2, l^1, l^2} \leq \sum_{m^1, m^2} P_{y_0} \left(r^i \in \mathcal{I}_{j^i} \text{ et } w^i \in \Psi_x^i(\overline{C_{li}^i}, \alpha_{li, m^i}^i), \text{ pour } i = 1, 2 \right)$$

Or, on a :

$$\begin{aligned} & P_{y_0} \left(r^i \in \mathcal{I}_{j^i} \text{ et } w^i \in \Psi_x^i(\overline{C_{li}^i}, \alpha_{li, m^i}^i), \text{ pour } i = 1, 2 \right) \quad (2.4) \\ &= P_{y_0} \left(r^i \in \mathcal{I}_{j^i} \text{ et } \langle \alpha_{li, m^i}^i, w^i \rangle - H_x^i(\alpha_{li, m^i}^i) - L_x^i(\overline{C_{li}^i}) + \frac{\varepsilon}{4} > 0, \text{ pour } i = 1, 2 \right) \\ &\leq \mathbb{E}_{y_0} \left(\prod_{i=1}^2 \left(\mathbb{I}_{\{r^i \in \mathcal{I}_{j^i}\}} \cdot \exp ar^i \left[\langle \alpha_{li, m^i}^i, w^i \rangle - \left(H_x^i(\alpha_{li, m^i}^i) + L_x^i(\overline{C_{li}^i}) - \frac{\varepsilon}{4} \right) \right] \right) \right). \end{aligned}$$

Comme $\prod_{i=1}^2 \left(\mathbb{I}_{\{r^i \in \mathcal{I}_{j^i}\}} \cdot \exp -ar^i \left[H_x^i(\alpha_{li, m^i}^i) + L_x^i(\overline{C_{li}^i}) - \frac{\varepsilon}{4} \right] \right)$ est $\mathcal{F}_{a-1}$ -mesurable, le lemme 2.1.1 permet de majorer la quantité (2.4) par :

$$\left\| \prod_{i=1}^2 \left(\mathbb{I}_{\{r^i \in \mathcal{I}_{j^i}\}} \cdot \exp -ar^i \left[H_x^i(\alpha_{li, m^i}^i) + L_x^i(\overline{C_{li}^i}) - \frac{\varepsilon}{4} - \sup_{y \in B(y_0, a\tau M)} H_y^i(\alpha_{li, m^i}^i) \right] \right) \right\|_{\infty} \quad (2.5)$$

Supposons que y_0 soit dans $B(x, R)$. Alors

$$H_x^i(\alpha_{l^i, m^i}^i) - \sup\{H_y^i(\alpha_{l^i, m^i}^i) \mid y \in B(y_0, a\tau M)\}$$

est majoré par $\sup\{|H_x^i(\alpha_{l^i, m^i}^i) - H_y^i(\alpha_{l^i, m^i}^i)| \mid y \in B(x, R + a\tau M)\}$, où M est comme dans (1.16) et $a\tau$ inférieur à T . Soit $\varepsilon_1 > 0$. D'après (HC), il existe δ_1 , ne dépendant que du compact $\mathcal{K}$, tel que pour tout α , on ait : $\sup_{y \in B(x, \delta_1)} |H_x^i(\alpha) - H_y^i(\alpha)| \leq \varepsilon_1 |\alpha|$.

Pour $R + a\tau M \leq \delta_1$ (qui assure $a\tau \leq T$ dès que $\delta_1 \leq MT$), on majore alors la quantité (2.5) par :

$$\left\| \prod_{i=1}^2 \left(\mathbb{I}_{\{r^i \in \mathcal{I}_{j^i}\}} \cdot \exp -a r^i \left[L_x^i(\overline{C_{l^i}^i}) - \frac{\varepsilon}{4} - \varepsilon_1 |\alpha_{l^i, m^i}^i| \right] \right) \right\|_{\infty}.$$

On pose $C = \sup_{i, l^i, m^i} |\alpha_{l^i, m^i}^i|$ (indépendant de y_0) et on prend ε_1 assez petit pour que $\varepsilon_1 C$ soit inférieur à $\frac{\varepsilon}{4}$. On a alors, pour tout y_0 de $B(x, R)$:

$$\begin{aligned} P_{j^1, j^2, l^1, l^2} &\leq \sum_{m^1, m^2} \exp \left[a \frac{\varepsilon}{2} - a \sum_{i=1}^2 \mathbb{I}_{\{r^i \in \mathcal{I}_{j^i}\}} L_x^i(\overline{C_{l^i}^i}) \right] \\ &\leq (M_{l^1} + M_{l^2}) \exp -a \left[-\frac{\varepsilon}{2} + \sum_{i=1}^2 p_{j^i} \delta' L_x^i(\overline{C_{l^i}^i}) \right]. \end{aligned}$$

Or, les conditions définissant $\mathcal{E}$ impliquent :

$$\sum_{i=1}^2 p_{j^i} \delta' L_x^i(\overline{C_{l^i}^i}) > I - \frac{\varepsilon}{4} - \delta' \sum_{i=1}^2 \left(L_x^i(\overline{B_{l^i}^i}) + \frac{\varepsilon}{4} \right),$$

car $(h_{x, \theta}^i)^{-1}$ augmente le coût L_x^i donc $L_x^i(\overline{C_{l^i}^i}) \geq L_x^i(\overline{B_{l^i}^i})$ et puisque $\sum_{i=1}^2 p_{j^i} \delta'$ est inférieur à 1. D'autre part, $L_x^i(\overline{B_{l^i}^i})$ est inférieur à $J_x(\delta)$ donc à $\sup_{y \in \mathcal{K}} J_y(\delta)$, quantité qui est finie d'après la continuité uniforme de $L^i(\cdot)$ sur $\cup_{y \in \mathcal{K}} y \times (S_y^i)^{-\delta}$. On prend δ' tel que $\delta'(\sup_{y \in \mathcal{K}} J_y(\delta) + \frac{\varepsilon}{4}) \leq \frac{\varepsilon}{4}$. On a alors, pour (j^1, j^2, l^1, l^2) dans $\mathcal{E}$:

$$\sum_{i=1}^2 p_{j^i} \delta' L_x^i(\overline{C_{l^i}^i}) - \frac{\varepsilon}{2} > I - \varepsilon$$

et donc :

$$P_{j^1, j^2, l^1, l^2} \leq (M_{l^1} + M_{l^2}) e^{-a(I-\varepsilon)}$$

Au total, on a donc :

$$\mathcal{A} \leq \sum_{(j^1, j^2, l^1, l^2) \in \mathcal{E}} (M_{l^1} + M_{l^2}) e^{-a(I-\varepsilon)} \leq C' e^{-a(I-\varepsilon)},$$

où C' est une constante qui dépend des découpages (2.2) et (2.3). En prenant alors a suffisamment grand ($a \geq \frac{4 \log C'}{\epsilon}$), on en déduit, pour tous $I \geq 0$ et y_0 dans $B(x, R)$:

$$\mathcal{A} \leq e^{-a(I-\epsilon)}$$

ce qui achève la démonstration du lemme 2.1.3.

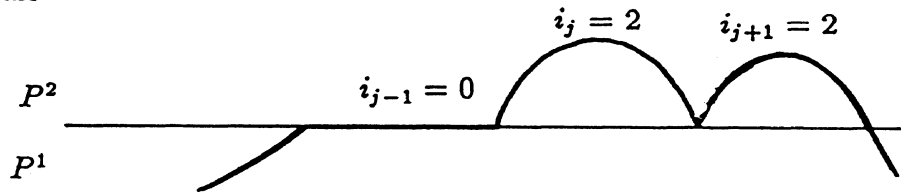
2.2 Enoncé d'un premier théorème sur les trajectoires

Dans ce paragraphe, on démontre le résultat de majoration de la probabilité pour que la trajectoire du processus appartienne à une partie borélienne donnée (théorème 2.2.10) avec une fonction de coût $\mathcal{L}$ qui sera calculée, dans le paragraphe suivant, sur des chemins présentant une certaine régularité (précisément sur l'ensemble $C^{f,1}[0, T]$ défini dans la notation 2.2.2 ci-dessous).

$C[0, T]$ désigne toujours l'ensemble des chemins continus de $[0, T]$ dans $\mathbb{R}^d$ muni de la topologie de la convergence uniforme sur $[0, T]$.

Notation 2.2.1 On note $C^f[0, T]$ l'ensemble des chemins ψ continus pour lesquels il existe une partition finie $0 = t_0 \leq t_1 \leq \dots \leq t_p = T$ telle que $\psi([t_j, t_{j+1}[)$ soit inclus dans P^{i_j} , avec $i_j = 0, 1$ ou 2 .

Remarquons que, dans la notation précédente, deux entiers successifs i_j et i_{j+1} peuvent être égaux :



Notation 2.2.2 On note $C^{f,1}[0, T]$ l'ensemble des applications ψ de $C^f[0, T]$ qui sont de classe C^1 par morceaux sur les intervalles ouverts où elles sont à valeurs dans P^0 .

Remarquons que $C^{f,1}[0, T]$ contient toutes les applications continues ne rencontrant pas P^0 .

2.2.1 Définition d'une fonctionnelle de coût approchée

Avec les notations du paragraphe 1.3, pour h positif et x dans $\mathbb{R}^d$, ${}^h\overline{L}_x^i(\beta)$ est l'enveloppe convexe de $\{L_y^i \mid y \in B(x, h)\}$, où $B(x, h)$ désigne la boule fermée de centre x et rayon h . Donc, pour β dans $\mathbb{R}^d$, on a :

$${}^h\overline{L}_x^i(\beta) = \inf \left\{ \sum_{y \in B(x, h)} \rho_y L_y^i(\beta_y) \mid \sum_{y \in B(x, h)} \rho_y \beta_y = \beta, \rho_y \geq 0, \sum_{y \in B(x, h)} \rho_y = 1 \right\}.$$

Notation 2.2.3 Pour h positif, θ réel, et x et β dans $\mathbb{R}^d$, avec $|\pi(\beta)| \leq \theta$, on pose :

$$L_\theta^h(x, \beta) = \inf \left\{ \sum_{i=1}^2 \rho^i {}^h\overline{L}_x^i(\beta^i) \mid \sum_{i=1}^2 \rho^i = 1, \sum_{i=1}^2 \rho^i \beta^i = \beta, \pi(\beta^1) \geq -\theta \text{ et } \pi(\beta^2) \leq \theta \right\}$$

$L_\theta^h(x, \cdot)$ est une fonction approchant le coût mixte, pour $\pi(\beta) = 0$, donné en (1.12). On va l'utiliser, au voisinage de P^0 , avec θ positif dans ce chapitre et θ négatif dans le chapitre suivant. Dans la suite du présent chapitre, θ est toujours positif.

Remarquons qu'on ne peut pas écrire $L_\theta^h(x, \cdot)$ comme la convexifiée, pour y variant dans $B(x, h)$, des applications $L_\theta^0(y, \cdot)$ définies par :

$$L_\theta^0(y, \beta) = \inf \left\{ \sum_{i=1}^2 \rho^i L_y^i(\beta^i) \mid \sum_{i=1}^2 \rho^i = 1, \sum_{i=1}^2 \rho^i \beta^i = \beta, \pi(\beta^1) \geq -\theta \text{ et } \pi(\beta^2) \leq \theta \right\},$$

puisque $L_\theta^h(x, \beta)$ autorise, pour obtenir ${}^h\overline{L}_x^i(\beta^i)$, des décompositions de β^i en vecteurs β_k^i de coordonnée π quelconque, ce qui ne serait pas le cas de cette convexifiée.

On a défini L_x^0 dans la notation 1.4.4. On a, pour h positif et x, β dans $\mathbb{R}^d$:

$${}^h\overline{L}_x^0(\beta) = \begin{cases} {}^h\overline{L}_x^1(\beta) & \text{si } \pi(\beta) \leq 0 \\ {}^h\overline{L}_x^2(\beta) & \text{si } \pi(\beta) > 0 \end{cases}$$

Notation 2.2.4 On définit, pour θ et h positifs, l'application $\lambda_{\theta, h}$ en posant, pour x et β dans $\mathbb{R}^d$:

$$\lambda_{\theta, h}(x, \beta) = \begin{cases} {}^h\overline{L}_x^i(\beta) & \text{si } x \in (P^i)^{-h}, i = 1 \text{ ou } 2 \\ L_\theta^h(x, \beta) & \text{si } |\pi(\beta)| \leq \theta \\ {}^h\overline{L}_x^i(\beta) & \text{si } |\pi(\beta)| > \theta \end{cases} \quad \text{si } x \in P^i \cap (P^0)^h, i = 0, 1 \text{ ou } 2$$

Remarques:

1. Pour $\theta \geq 0$, si $|\pi(\beta)| \leq \theta$, $L_\theta^h(x, \beta)$ est inférieur à ${}^h\overline{L}_x^1(\beta) \wedge {}^h\overline{L}_x^2(\beta)$.

En particulier,

$$\begin{cases} \text{si } x \in (P^0)^h \text{ et } |\pi(\beta)| \leq \theta, & \lambda_{\theta,h}(x, \beta) \leq {}^h\overline{L}_x^1(\beta) \wedge {}^h\overline{L}_x^2(\beta) \\ \text{pour } i = 0, 1 \text{ ou } 2, \text{ si } x \in P^i, & \lambda_{\theta,h}(x, \cdot) \leq {}^h\overline{L}_x^i(\cdot) \end{cases} \quad (2.6)$$

2. Considérant le cas $(\theta, h) = (0, 0)$, on constate que, pour x dans P^0 , $L_0^0(x, \cdot)$ et L_x coïncident sur l'ensemble $\{\beta | \pi(\beta) = 0\}$ donc $\lambda_{0,0}$ et λ coïncident, sauf en (x, β) tel que x appartient à P^0 et $\pi(\beta)$ est non nul.
3. on aurait pu prendre $\lambda_{\theta,h}$ plus compatible avec la définition de λ en remplaçant, pour x dans $P^i \cap (P^0)^h$ et $i = 0, 1$ ou 2 , dans l'expression de $\lambda_{\theta,h}(x, \beta)$:

$$\begin{cases} L_\theta^h(x, \beta) & \text{si } |\pi(\beta)| \leq \theta \\ {}^h\overline{L}_x^i(\beta) & \text{si } |\pi(\beta)| > \theta \end{cases}$$

par :

$$\inf \left\{ \rho^0 L_\theta^h(x, \beta^0) + \rho {}^h\overline{L}_x^i(\beta') \mid \rho^0 + \rho = 1, \rho^0 \beta^0 + \rho \beta' = \beta, |\pi(\beta^0)| \leq \theta \right\}.$$

Cela aurait permis $L_x(\beta) = L_0^0(x, \beta)$ pour β quelconque quand x appartient à P^0 et donc $\lambda_{0,0} = \lambda$ partout. De plus, $\lambda_{\theta,h}(x, \cdot)$ serait convexe pour x dans $P^i \cap (P^0)^h$, $i = 1$ ou 2 (mais toujours pas pour x dans P^0).

Cependant, la forme choisie est plus maniable et convient également (rappelons que dans le calcul de $\Lambda_{0,T}$, L_x n'intervient que sur les vecteurs parallèles à P^0 , cf Remarque 1.4.9 page 49).

Définition 2.2.5 On définit enfin, pour θ et h positifs, une fonctionnelle de coût approché paramétrée par (θ, h) , en posant, pour tout chemin continu, φ , de $[0, T]$ dans $\mathbb{R}^d$:

$$\Lambda_{0,T}^{\theta,h}(\varphi) = \begin{cases} \int_0^T \lambda_{\theta,h}(\varphi(t), \varphi'(t)) dt & \text{si } \varphi \text{ est absolument continu} \\ +\infty & \text{sinon} \end{cases}$$

Remarquons que, contrairement à $\Lambda_{0,T}$, quand x appartient à $(P^0)^h$, le choix de $\lambda_{\theta,h}(x, \beta)$ pour $\pi(\beta)$ non nul intervient dans le calcul de $\Lambda_{0,T}^{\theta,h}$.

2.2.2 Propriétés de la fonctionnelle de coût approchée $\Lambda_{0,T}^{\theta,h}$

L'application $\Lambda_{0,T}^{\theta,h}$ définie sur $C[0,T]$ ci-dessus possède les propriétés suivantes :

Proposition 2.2.6

- (i) $\Lambda_{0,T}^{\theta,h}$ est décroissante par rapport à chacun des paramètres positifs θ et h .
- (ii) pour tout φ de $C[0,T]$, on a :

$$\Lambda_{0,T}(\varphi) = \Lambda_{0,T}^{0,0}(\varphi) = \lim_{(\theta,h) \rightarrow (0,0)} \Lambda_{0,T}^{\theta,h}(\varphi) \quad (\text{limite croissante})$$

Démonstration :

- (i) Montrons que l'application $\lambda_{\theta,h}$, donnée dans la notation 2.2.4 est décroissante par rapport à θ :

- si $x \in (P^i)^{-h}$, $i = 1$ ou 2 , θ n'influe pas
- si $x \in (P^0)^h$: considérons θ et θ' avec $0 \leq \theta < \theta'$. Les conditions apparaissant dans la définition de $L_{\theta'}^h$ sont moins strictes que celles de la définition de L_{θ}^h donc $L_{\theta'}^h$ est inférieur à L_{θ}^h . Alors, on a :
 - pour $|\pi(\beta)| \leq \theta$, $\lambda_{\theta',h}(x, \beta) \leq \lambda_{\theta,h}(x, \beta)$,
 - pour $\theta < |\pi(\beta)| \leq \theta'$, $L_{\theta'}^h(x, \beta) \leq {}^h\overline{L}_x^i(\beta)$ ($i = 0, 1$ ou 2) donc on a encore $\lambda_{\theta',h}(x, \beta) \leq \lambda_{\theta,h}(x, \beta)$,
 - pour $\theta' < |\pi(\beta)|$, $\lambda_{\theta',h}(x, \beta) = \lambda_{\theta,h}(x, \beta)$.

Montrons la décroissance par rapport à h . Considérons h et h' avec $h < h'$. Pour $i = 1$ ou 2 , on a ${}^{h'}\overline{L}_x^i \leq {}^h\overline{L}_x^i$ donc $L_{\theta}^{h'} \leq L_{\theta}^h$. Alors :

- si $x \in (P^0)^h$, $L_{\theta}^{h'}(x, \beta) \leq L_{\theta}^h(x, \beta)$ pour $|\pi(\beta)| \leq \theta$, et ${}^{h'}\overline{L}_x^i(\beta) \leq {}^h\overline{L}_x^i(\beta)$ pour $|\pi(\beta)| > \theta$ impliquent $\lambda_{\theta,h'}(x, \beta) \leq \lambda_{\theta,h}(x, \beta)$
- si $x \in (P^i)^{-h'}$, $\lambda_{\theta,h'}(x, \cdot) \leq \lambda_{\theta,h}(x, \cdot)$ vient de ${}^{h'}\overline{L}_x^i \leq {}^h\overline{L}_x^i$
- si $x \in (P^i)^{-h} \cap (P^0)^{h'}$, on a $L_{\theta}^{h'}(x, \beta) \leq {}^{h'}\overline{L}_x^i(\beta)$ pour $|\pi(\beta)| \leq \theta$, donc pour tout β , $\lambda_{\theta,h'}(x, \beta) \leq {}^{h'}\overline{L}_x^i(\beta) \leq {}^h\overline{L}_x^i(\beta) = \lambda_{\theta,h}(x, \beta)$.

La monotonie vient donc en particulier de l'inégalité $L_{\theta}^h(x, \beta) \leq {}^h\overline{L}_x^i(\beta)$ pour $|\pi(\beta)| \leq \theta$.

- (ii) pour φ dans $C[0,T]$, $\Lambda_{0,T}(\varphi) = \Lambda_{0,T}^{0,0}(\varphi)$ vient de ce que, presque sûrement en t tel que $\varphi(t)$ appartient à P^0 , $\pi(\varphi'(t))$ vaut 0, et de la remarque 2 page 68.

La deuxième égalité est immédiate si φ n'est pas absolument continu. Considérons un chemin φ absolument continu sur $[0, T]$; comme on a $\Lambda_{0,T}^{\theta,h} \leq \Lambda_{0,T}^{0,0}$, il suffit de montrer

$$\Lambda_{0,T}^{0,0}(\varphi) \leq \lim_{(\theta,h) \rightarrow (0,0)} \Lambda_{0,T}^{\theta,h}(\varphi).$$

On considère t dans $[0, T]$;

— si $\varphi(t)$ appartient à P^i , pour h assez petit, $\varphi(t)$ n'est pas dans $(P^0)^h$, donc $\lambda_{\theta,h}(\varphi(t), \varphi'(t)) = {}^h\overline{L}_{\varphi(t)}^i(\varphi'(t))$ tend vers $L_{\varphi(t)}^i(\varphi'(t))$ quand h tend vers 0, d'après (1.8): $L_x(\beta) = \lim_{h \rightarrow 0} {}^h\overline{L}_x(\beta)$ pour tous x, β de $\mathbb{R}^d$.

— si $\varphi(t)$ appartient à P^0 , presque sûrement en t , $\pi(\varphi'(t))$ vaut 0; dans ce cas, $\lambda_{\theta,h}(\varphi(t), \varphi'(t))$ est égal à $L_{\varphi(t)}^h(\varphi'(t))$ et converge donc vers $L_{\varphi(t)}(\varphi'(t)) = L_0^0(\varphi(t), \varphi'(t))$ quand (θ, h) tend vers $(0, 0)$. En effet, montrons que, pour tous x et β dans $\mathbb{R}^d$, si $\pi(\beta)$ vaut 0, $L_{\varphi(t)}^h(x, \beta)$ converge vers $L_0^0(x, \beta)$ quand (θ, h) tend vers $(0, 0)$:

l'application $L_{\varphi(t)}^h$ est inférieure à L_0^0 ; montrons que, pour (θ_n, h_n) convergeant vers $(0, 0)$, on a:

$$L_0^0(x, \beta) \leq \liminf_{n \rightarrow +\infty} L_{\varphi(t)}^{h_n}(x, \beta) \quad (2.7)$$

On suppose que le membre de droite de (2.7) est fini. Alors, pour n assez grand, la borne inférieure dans la définition de $L_{\varphi(t)}^{h_n}(x, \beta)$ (notation 2.2.3) est atteinte, avec $(\rho_n^1, \beta_n^1, \rho_n^2, \beta_n^2)$. Quand n varie, ce quadruplet reste dans un même compact. Soit $(\rho^1, \beta^1, \rho^2, \beta^2)$ une valeur d'adhérence. On a $\sum_{i=1}^2 \rho^i = 1$, $\sum_{i=1}^2 \rho^i \beta^i = \beta$, $(-1)^i \pi(\beta^i) \leq 0$ et, d'après $\pi(\beta) = 0$ et le lemme 1.3.5:

$$L_0^0(x, \beta) \leq \sum_{i=1}^2 \rho^i L_x^i(\beta^i) \leq \lim_{n \rightarrow +\infty} \sum_{i=1}^2 \rho^i {}^h\overline{L}_x^i(\beta_n^i).$$

Ainsi, $\lambda_{\theta,h}(\varphi(\cdot), \varphi'(\cdot))$ converge presque sûrement vers $\lambda_{0,0}(\varphi(\cdot), \varphi'(\cdot))$ quand (θ, h) tend vers $(0, 0)$. La conclusion vient alors du lemme de Fatou. $\square$

En revanche, l'application $\Lambda_{0,T}^{\theta,h}$ présente un défaut essentiel, c'est que l'on ne sait pas si elle est s.c.i.. L'application $\lambda_{\theta,h}$ n'est pas suffisamment régulière au voisinage de P^0 pour que l'on puisse appliquer le théorème 1.4.10 et les méthodes plus directes échouent: en particulier, pour démontrer la semi-continuité, en un chemin n'appartenant pas à $C^f[0, T]$, c'est-à-dire oscillant une infinité de fois autour de P^0 , on devrait sommer une infinité d'inégalités. Le résultat le plus proche d'une s.c.i. est le résultat asymptotique (2.68) en un chemin φ de $C^{f,1}[0, T]$:

$$\Lambda_{0,T}(\varphi) \leq \lim_{(\theta,h) \rightarrow (0,0)} \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in C^f[0,T]}} \Lambda_{0,T}^{\theta,h}(\psi)$$

Ce résultat nous conduit à régulariser la fonctionnelle de coût en la calculant sur les chemins de $C^f[0, T]$.

2.2.3 Définition d'une fonctionnelle régularisée $\mathcal{L}$. Propriétés. Théorème avec cette fonctionnelle

On introduit d'abord une régularisation de la fonctionnelle approchée, qui va pallier le fait que $\Lambda_{0,T}^{\theta,h}$ n'est pas nécessairement pas s.c.i..

Notation 2.2.7 On pose, pour φ dans $C[0, T]$:

$$\mathcal{L}_{0,T}^{\theta,h}(\varphi) = \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in C^f[0,T]}} \Lambda_{0,T}^{\theta,h}(\psi)$$

Comme Λ^{reg} dans [AR] (définition 1.1.9), ce coût ne fait intervenir que des chemins suffisamment réguliers. L'application $(\theta, h, \varphi) \mapsto \mathcal{L}_{0,T}^{\theta,h}(\varphi)$ possède les propriétés suivantes :

Proposition 2.2.8

- (i) Pour φ dans $C[0, T]$ donné, $\mathcal{L}_{0,T}^{\theta,h}(\varphi)$ est décroissante par rapport à θ et par rapport à h .
- (ii) Pour (θ, h) positifs fixés, l'application $\mathcal{L}_{0,T}^{\theta,h}$ est s.c.i. sur $C[0, T]$.
- (iii) Pour tout compact $\mathcal{K}$ et pour θ, h positifs, l'ensemble $\{\varphi \in C[0, T] \mid \varphi(0) \in \mathcal{K} \text{ et } \mathcal{L}_{0,T}^{\theta,h}(\varphi) < +\infty\}$ est compact.

Démonstration :

(i) est immédiat. On montre facilement (ii), par l'absurde, en utilisant le fait que l'application $\varphi \mapsto \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in C^f[0,T]}} \Lambda_{0,T}^{\theta,h}(\psi)$ est s.c.i. et que $C^f[0, T]$ est dense dans $C[0, T]$.

(iii) (cf démonstration du lemme 4.3 dans [AR]) :

Considérons un chemin φ dans $C[0, T]$ tel qu'on ait $\varphi(0) \in \mathcal{K}$ et $\mathcal{L}_{0,T}^{\theta,h}(\varphi) < +\infty$. Nécessairement, il existe une suite (ψ_n) dans $C^f[0, T]$ convergeant uniformément vers φ et telle que $\lim_{n \rightarrow +\infty} \Lambda_{0,T}^{\theta,h}(\psi_n) < +\infty$. Pour n assez grand, $\psi_n(0)$ appartient à $B(\varphi(0), 1)$ et ps en t , $\psi'_n(t)$ appartient à l'enveloppe convexe de $S_{\psi_n(t)}^1$ et $S_{\psi_n(t)}^2$. Puisque les champs de probabilités vérifient l'hypothèse (H4), on sait que cela implique que les ψ_n sont à valeurs dans un compact. Les ψ'_n sont donc uniformément bornées. On peut donc extraire de (ψ_n) une sous-suite (ψ_k) dont les dérivées convergent pour la topologie faible de $L^2([0, T])$, vers une fonction qui est nécessairement φ' , donc absolument continue (par : $\varphi(t) - \varphi(0) = \lim_{k \rightarrow +\infty} [\psi_k(t) - \psi_k(0)] = \lim_{k \rightarrow +\infty} \int_0^t \psi'_k(s) ds = \int_0^t \varphi'(s) ds$).

Presque sûrement en t , $\psi'_k(t)$ appartient à un compact ne dépendant que de $\mathcal{K}$. L'application φ' est donc bornée par une constante ne dépendant que de $\mathcal{K}$.

En particulier, $\{\varphi \in C[0, T] \mid \varphi(0) \in \mathcal{K} \text{ et } \mathcal{L}_{0,T}^{\theta,h}(\varphi) < +\infty\}$ est équicontinu. Comme d'autre part, il est fermé, d'après (ii), et que tout chemin de cet ensemble est à valeur dans un compact fixe, on en déduit d'après le théorème d'Ascoli que cet ensemble est compact dans $C[0, T]$. $\square$

Remarquons que (iii) revient à dire que les chemins de $C[0, T]$ d'origine dans un compact et de coût fini (que ce soit $\Lambda_{0,T}$, $\mathcal{L}_{0,T}^{\theta,h}$ ou $\mathcal{L}_{0,T}$ ci-dessous) ont des dérivées à droite uniformément bornées (et sont donc à valeurs dans un compact).

Notation 2.2.9 On définit une fonctionnelle de coût régularisée en posant, pour φ dans $C[0, T]$:

$$\mathcal{L}_{0,T}(\varphi) = \lim_{(\theta,h) \rightarrow (0,0)} \mathcal{L}_{0,T}^{\theta,h}(\varphi) = \lim_{(\theta,h) \rightarrow (0,0)} \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in C^1[0,T]}} \Lambda_{0,T}^{\theta,h}(\psi)$$

Avec les notations introduites dans ce paragraphe, nous pouvons énoncer le théorème suivant, démontré dans le prochain paragraphe :

Théorème 2.2.10

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (H_{maj}) et (H_0) . On appelle φ^τ le chemin aléatoire continu défini sur $[0, T]$ par : $\varphi^\tau(k\tau) = Y_k^\tau$ pour $0 \leq k\tau < T + \tau$ et φ^τ est affine sur $[k\tau; (k+1)\tau]$.

Alors, pour toute partie borélienne G de $C[0, T]$, tout compact $\mathcal{K}$ de $\mathbb{R}^d$ et tout $\varepsilon > 0$, il existe τ_0 tel que $\tau \leq \tau_0$ implique, P -ps sur $\{Y_0^\tau \in \mathcal{K}\}$:

$$\tau \log P(\varphi_{[0,T]}^\tau \in G \mid \mathcal{F}_0) \leq -\mathcal{L}(\overline{G}) + \varepsilon$$

Remarquons que ce résultat peut s'énoncer sous la forme : uniformément en x dans $\mathcal{K}$, on a :

$$\limsup_{\tau \rightarrow 0} \tau \log P_x(\varphi_{[0,T]}^\tau \in G) \leq -\mathcal{L}^x(\overline{G})$$

où $\mathcal{L}^x(\overline{G})$ désigne $\inf_{\varphi \in \overline{G}, \varphi(0)=x} \mathcal{L}_{0,T}(\varphi)$.

2.3 Démonstration du théorème avec $\mathcal{L}$

2.3.1 Méthode

Le théorème 2.2.10 donne une majoration de la probabilité qu'un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 soit dans une partie borélienne donnée, quand τ tend vers 0. De façon classique (cf chapitre d'introduction), cela se déduit d'un premier résultat consistant en la majoration de la probabilité que le processus soit à une distance strictement positive de l'ensemble des chemins de coût inférieur à une quantité fixée, toujours quand τ tend vers 0.

Pour établir ce résultat dans le cas d'un seul champ, on peut procéder de la façon suivante (cf théorème 6.3 dans [AR]) : pour $I > 0$, on construit un chemin aléatoire η , régularisé de la trajectoire de la marche aléatoire, restant proche de cette trajectoire et dont le coût est inférieur à I avec une forte probabilité (supérieure à un terme de l'ordre de $1 - e^{-\frac{I}{\tau}}$). Pour construire η , on regroupe les pas effectués pendant un petit intervalle de temps sur lequel on prend η' proche de la vitesse moyenne du processus. Précisément, sur l'intervalle $[kn\tau; (k+1)n\tau[$, on construit η de la façon suivante : on pose

$$w_k = \frac{1}{n} \sum_{l=kn}^{(k+1)n-1} X_l(Y_l^\tau)$$

puis $v_k = h_{Y_{kn}^\tau, \theta}(w_k)$ (cf définition 1.2.9) et on définit un champ de vecteurs $W_k(\cdot)$ tel que, pour y dans $\mathbb{R}^d$, $W_k(y)$ désigne la projection de v_k sur le compact $\{L_y(\cdot) \leq L_{Y_{kn}^\tau}(v_k)\}$. On prend alors pour η sur $[kn\tau; (k+1)n\tau[$ la courbe intégrale du champ W_k . On suppose que le champ (μ_x) vérifie l'hypothèse (HLip) (c'est-à-dire l'hypothèse (H') dans notre paragraphe 1.1.1), ce qui permet de majorer $|\eta'(t) - v_k|$ par $C|\eta(t) - Y_{kn}^\tau|$ et donc de prouver que η reste proche du processus (à une distance inférieure à ε donné) pour θ suffisamment petit. Pour t dans $[kn\tau; (k+1)n\tau[$, $L_{\eta(t)}(W_k(\eta(t)))$ est majoré par $L_{Y_{kn}^\tau}(v_k)$, lui-même contrôlé grâce à un théorème sur les vitesses moyennes du type de notre lemme 2.1.3. On peut alors obtenir, en sommant sur tous les intervalles, une majoration de la probabilité $P(\Lambda_{0,T}(\eta) \geq I)$ et donc une majoration de la probabilité que le processus soit à une distance supérieure à ε de l'ensemble des chemins de coût inférieur à I .

Une construction analogue pourrait être établie dans le cas de deux champs de probabilités (cf remarque 2.3.4). Nous utilisons une autre méthode qui permet de ne pas faire d'hypothèse de type lipschitzien sur les lignes de niveau de la transformée de Cramer. Exposons-la dans le cas d'un seul champ :

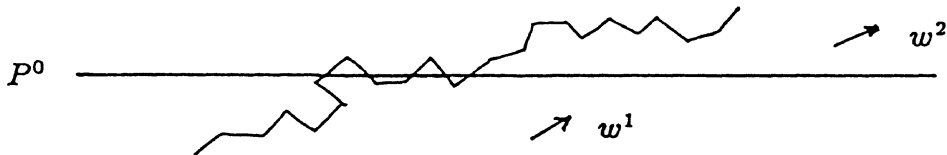
on prend $\eta'(t)$ directement égal à v_k , sans projeter ; alors on ne contrôle plus que

la borne inférieure des $L_y(\eta'(t))$ pour y dans un voisinage de $\eta(t)$. Cela permet cependant de majorer $P(\Lambda_{0,T}^h(\eta) \geq I)$, avec $\Lambda_{0,T}^h(\eta) = \int_0^T h \bar{L}_{\eta(t)}(\eta'(t)) dt$ et il suffit ensuite de passer à la limite quand h tend vers 0. Cette approche est celle utilisée dans [DEW].

Ainsi le premier résultat que nous démontrons consiste en une majoration avec le coût approché $\Lambda_{0,T}^{\theta,h}$ (théorème 2.3.2). Remarquons que la présence de la discontinuité rend de toute façon nécessaire l'usage d'un tel coût. Cela explique la présence d'un autre paramètre : θ .

Pour démontrer le théorème 2.3.2, nous construisons à nouveau un chemin aléatoire η approchant la trajectoire. Etudions ce que doit vérifier η au voisinage de la frontière :

on devra estimer la probabilité que la moyenne des pas soit proche d'un certain vecteur. On sait que, quand la moyenne est constituée de pas du même type i , c'est-à-dire quand le processus reste toujours dans le même demi-espace P^i , cette probabilité est donnée, en un point x , par la valeur de la transformée de Cramer L_x^i en ce vecteur (on se ramène au cas où il y a un seul champ de probabilités). Dans le cas contraire, on doit tenir compte des passages de la frontière P^0 pour regrouper les pas. En effet, supposons que l'on s'intéresse à n pas consécutifs. Si on fait deux moyennes w^1 et w^2 réunissant chacune un type de pas, et qu'on écrit η' en fonction de ces deux vecteurs, on n'arrivera pas à faire apparaître un coût mixte approchant L_x (notation 1.4.4 page 37) car $\pi(w^1)$ et $\pi(w^2)$ sont quelconques : les conditions $(-1)^i \pi(w^i) \leq 0$ ne sont pas vérifiées. On obtiendra seulement le coût $\text{conv}(L_x^1, L_x^2)$:

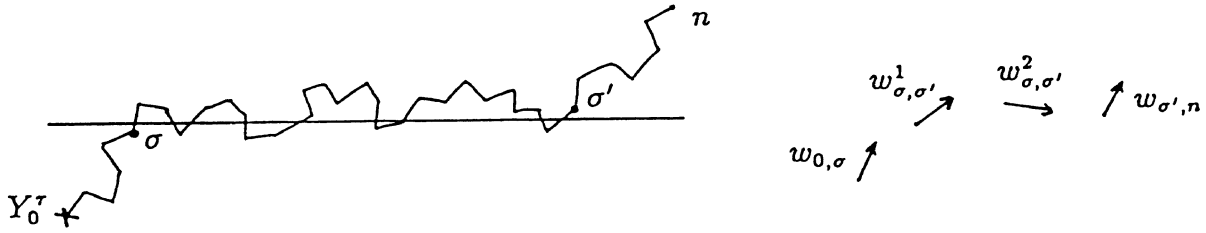


On distingue alors quatre classes de pas parmi ceux considérés :

- ceux antérieurs à σ , instant de première traversée de P^0 , tous de même type, dont on note $w_{0,\sigma}$ la moyenne,
- ceux qui sont effectués après σ' , instant de dernier passage de P^0 , de même type également, dont on note $w_{\sigma',n}$ la moyenne,
- et les pas intermédiaires, qui se font au voisinage de P^0 , et que l'on sépare en deux classes selon leur type, les moyennes correspondantes étant notées respectivement $w_{\sigma,\sigma'}^1$ et $w_{\sigma,\sigma'}^2$.

Cette fois, $w_{\sigma,\sigma'}^i$ est une moyenne de pas faisant passer de P^i à l'autre demi-espace, on a donc $(-1)^i \pi(w_{\sigma,\sigma'}^i) \leq 0$; de plus, la moyenne de tous ces pas intermédiaires, barycentre de $w_{\sigma,\sigma'}^1$ et $w_{\sigma,\sigma'}^2$, est pratiquement parallèle à P^0 (pour τ petit à $n\tau$ fixé),

ce qui justifie l'utilisation du coût mixte sur ce vecteur :



D'autre part, de la même façon que dans le cas d'un seul champ, pour des questions d'uniformité, on doit ramener les vitesses moyennes à l'intérieur des S_x^i par les applications $h_{x,\theta'}^i$; l'écart entre η et la trajectoire va alors être proportionnel à θ' . Or, pour que $h_{x,\theta'}^i(w_{\sigma,\sigma'}^i)$ appartienne à $(S_{Y_0^\tau}^i)^{-\delta}$, on va devoir prendre $n\tau M$ (distance maximale parcourue en n pas) petit devant θ' .

Si par exemple $Y_{j\tau}^\tau(\omega)$ appartient à $P^1 \cup P^0$ et $Y_{(j+1)\tau}^\tau(\omega)$ à P^2 , la construction sur $[jn\tau, (j+1)n\tau]$ pourrait être de prendre η' égal à $h_{Y_{j\tau}^\tau, \theta'}^1(w_{0,\sigma})$ sur $[jn\tau, \sigma]$, à un barycentre de $h_{Y_{j\tau}^\tau, \theta'}^1(w_{\sigma,\sigma'}^1)$ et $h_{Y_{j\tau}^\tau, \theta'}^2(w_{\sigma,\sigma'}^2)$ sur $[\sigma, \sigma']$ et à $h_{Y_{j\tau}^\tau, \theta'}^2(w_{\sigma',n})$ sur $[\sigma', (j+1)n\tau]$. Avec cette construction, η peut passer dans P^2 alors que la trajectoire est encore dans P^1 et avoir une vitesse de première coordonnée élevée. Le coût mixte ne permet pas de prendre en compte le coût 1 d'un tel vecteur, or c'est celui que l'on contrôle puisque la trajectoire est encore soumise au champ 1.

On va donc modifier la construction de façon à maintenir η dans P^1 dans l'exemple précédent : si $\pi(h_{Y_{j\tau}^\tau, \theta'}^1(w_{0,\sigma})) > \theta$, au lieu de $h_{Y_{j\tau}^\tau, \theta'}^1(w_{0,\sigma})$, on prend pour η' un vecteur fixe, parallèle à P^0 . Cette modification devrait permettre de contrôler le coût tout en gardant une distance petite entre η et la trajectoire ; en effet, la durée concernée par cette modification est faible : on obtient un écart en $(\Delta_0 + \sigma\tau 2M\theta')(1 + \frac{C}{\theta})$ si Δ_0 est l'écart initial $|\eta(jn\tau) - Y_{j\tau}^\tau|$ (le terme en $\frac{C}{\theta}$ correspond à la modification).

Malheureusement, cela n'est pas encore suffisant ; en effet, cette modification peut intervenir sur chaque intervalle de longueur $n\tau$, la meilleure majoration de la distance entre les deux chemins est donc en $n\tau 2M\theta'(1 + \frac{C}{\theta})^{\frac{T}{n\tau}}$ qui est, pour $n\tau M \leq \theta'$, plus grand que $n\tau 2Mn\tau M(1 + \frac{C}{\theta})^{\frac{T}{n\tau}}$. Ce majorant tend vers $+\infty$ quand $n\tau$, le pas de découpage de $[0, T]$, tend vers 0.

Ce problème vient du fait que, avec cette construction, η reste à une distance de la trajectoire réelle qui est de l'ordre de la distance parcourue par la marche aléatoire sur l'intervalle concerné et que la position de η par rapport à la frontière P^0 n'est alors pas suffisamment liée à celle de la trajectoire (les instants σ et σ' sont peu significatifs pour η).

La construction qui sera effectivement utilisée dans la démonstration du théorème 2.3.2 consiste à effectuer deux regroupements successifs des pas : pour a et n entiers, on calcule d'abord les moyennes de a accroissements, on regroupe ensuite ces moyennes par paquets de n . Cette construction conduit à une distance entre η et la trajectoire qui est petite devant le pas de découpage $n\tau$. Les instants σ et σ' définis sur un intervalle $[jna\tau, (j+1)na\tau]$ correspondent alors à des instants où η est proche de P^0 .

2.3.2 Théorème de majoration avec coût approché

Le théorème 2.3.2 donne un résultat de majoration avec un coût approché $\Lambda_{0,T}^{\theta,h}$; nous passons à la limite dans le paragraphe suivant (page 98). Rappelons une notation utilisée dans le chapitre d'introduction :

Définition 2.3.1 *On définit la distance entre deux chemins φ_1 et φ_2 de $[0, T]$ dans $\mathbb{R}^d$ par :*

$$d_{0,T}(\varphi_1, \varphi_2) = \sup_{t \in [0, T]} |\varphi_1(t) - \varphi_2(t)|$$

et, si G est un ensemble de tels chemins, on pose $d_{0,T}(\varphi_1, G) = \inf_{\varphi_2 \in G} d_{0,T}(\varphi_1, \varphi_2)$.

Théorème 2.3.2

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (H_{maj}) et (H_0) . Pour $T > 0$, $I \geq 0$, θ et $h > 0$ fixés, on note $\Gamma_{0,T}^{\theta,h}(I)$ l'ensemble des chemins continus, φ , de $[0, T]$ dans $\mathbb{R}^d$, tels que $\Lambda_{0,T}^{\theta,h}(\varphi) \leq I$. Soit γ^τ le chemin aléatoire défini par $\gamma^\tau(t) = Y_k^\tau$ pour $k\tau \leq t < (k+1)\tau$.

Alors, pour tous ε, θ strictement positifs et pour tout compact $\mathcal{K}$ de $\mathbb{R}^d$, il existe h et τ_0 strictement positifs, tels que pour $\tau < \tau_0$ et pour tout y_0 de $\mathcal{K}$, on ait la majoration :

$$P_{y_0} \left(d_{0,T}(\gamma^\tau, \Gamma_{0,T}^{\theta,h}(I)) \geq \varepsilon \right) \leq e^{-\frac{1}{\tau}(I-\varepsilon)}.$$

Remarque : ce théorème sera utilisé seulement sous une forme légèrement plus faible (pour l'obtention de (2.66)).

Démonstration : on considère $\mathcal{K}$ un compact de $\mathbb{R}^d$ et $T > 0$. On leur associe $\mathcal{U}$

et M comme dans (1.16). On note Ω_0 l'ensemble $\{Y_0^\tau \in \mathcal{K}\}$.

On veut disposer, pour tout x de $\mathcal{U}$, d'un vecteur v_x , parallèle à la direction de P^0 et tel que $L_x(v_x)$ soit uniformément borné. On utilise l'hypothèse (H_0) :

2.3.3 Choix de v_x : d'après la continuité de l'application $x \mapsto S_x^i$, il existe π_0 et δ_0 strictement positifs tels que, pour tout x dans $\mathcal{U}$ et pour i égal à 1 ou 2, l'ensemble $\{\beta \in \mathbb{R}^d \mid (-1)^i \pi(\beta) < -\pi_0\} \cap (S_x^i)^{-\delta_0}$ soit non vide. Alors on peut trouver des applications continues sur $\mathbb{R}^d$, $x \mapsto b_x^i$, à valeurs dans $(S_x^i)^{-\delta_0}$, telles que, pour tout x de $\mathcal{U}$, on ait $\pi(b_x^1) > \pi_0$ et $\pi(b_x^2) < -\pi_0$.

On note v_x l'unique vecteur de la forme $\rho b_x^1 + (1-\rho)b_x^2$ tel que $\pi(\rho b_x^1 + (1-\rho)b_x^2) = 0$.

Remarquons qu'on a, avec la notation 2.1.4 :

$$L_x(v_x) \leq \rho L_x^1(b_x^1) + (1-\rho)L_x^2(b_x^2) \leq J_x(\delta_0)$$

et, pour tous θ, h positifs, $L_\theta^h(x, v_x) \leq L_x(v_x) \leq J_x(\delta_0)$.

On se donne θ, θ', h, τ strictement positifs et des entiers n et a . On utilisera :

$$\forall x \in \mathcal{U}, \forall \beta \in S_x^i, |\beta - h_{x,\theta'}^i(\beta)| = \theta' |\beta - m_x^i| \leq 2M\theta' \quad (2.8)$$

On définit les variables aléatoires suivantes :

(avec toujours, pour $l \geq 1$, $A_l^1 = \{Y_l^\tau \in P^1 \cup P^0\}$ et $A_l^2 = \{Y_l^\tau \in P^2\}$)

1. pour $i = 1$ ou 2 , et $k \in \mathbb{N}$,

$$\begin{cases} r_k^i &= \frac{1}{a} \sum_{l=ka}^{(k+1)a-1} \mathbb{I}_{A_l^i} \\ w_k^i &= \frac{1}{ar_k^i} \sum_{l=ka}^{(k+1)a-1} \mathbb{I}_{A_l^i} X_l^i(Y_l^\tau) \\ w_k &= r_k^1 w_k^1 + r_k^2 w_k^2 \end{cases}$$

Le réel r_k^i est la proportion de pas de type i parmi les a pas effectués entre les instants $ka\tau$ et $(k+1)a\tau$. Le vecteur w_k^i est la moyenne de ces pas. On a : $r_k^1 + r_k^2 = 1$ et $Y_{(k+1)a}^\tau = Y_{ka}^\tau + a\tau w_k$. Les variables r_k^i , w_k^i et w_k sont $\mathcal{F}_{(k+1)a\tau}$ -mesurables. Les variables r^i et w^i du paragraphe 2.1 correspondent à r_0^i et w_0^i .

2. pour $i = 1$ ou 2 , et $k < k' \in \mathbb{N}$,

$$\begin{cases} r_{k,k'}^i &= \frac{1}{(k' - k)a} \sum_{l=ka}^{k'a-1} \mathbb{I}_{A_i^i} &= \frac{1}{k' - k} \sum_{l=k}^{k'-1} r_l^i \\ w_{k,k'}^i &= \frac{1}{(k' - k)ar_{k,k'}^i} \sum_{l=ka}^{k'a-1} \mathbb{I}_{A_i^i} X_l^i(Y_l^\tau) &= \frac{\sum_{l=k}^{k'-1} r_l^i w_l^i}{\sum_{l=k}^{k'-1} r_l^i} \\ w_{k,k'} &= \sum_{i=1}^2 r_{k,k'}^i w_{k,k'}^i \end{cases}$$

Le réel $r_{k,k'}^i$ est la proportion de pas de type i entre les instants $ka\tau$ et $k'a\tau$.

Le vecteur $w_{k,k'}^i$ est la moyenne de ces pas. On a :

$$r_{k,k'}^1 + r_{k,k'}^2 = 1, \quad r_k^i = r_{k,k+1}^i, \quad w_k^i = w_{k,k+1}^i \quad \text{et} \quad Y_{k'a}^\tau = Y_{ka}^\tau + (k' - k)a\tau w_{k,k'}.$$

La démonstration commence par la construction d'un chemin aléatoire η régularisant la trajectoire. Si x est un réel, on note $[x]$ sa partie entière. On pose, pour k entier de $[0, \dots, \frac{T}{a\tau}]$:

$$\Delta_k = |\eta(ka\tau) - Y_{ka}^\tau|. \quad (2.9)$$

On construit η successivement sur les intervalles $[jna\tau; (j+1)na\tau[$ recouvrant $[0, T]$. Dans le paragraphe précédent, nous avons suggéré ce qu'allait être la construction de η sur un tel intervalle. Précisons-la maintenant. Supposant $\eta(0)$ connu, nous donnons la construction sur l'intervalle $[0; na\tau]$ et majorons Δ_n en fonction de Δ_0 . La même construction est ensuite appliquée à $[jna\tau; (j+1)na\tau[$ en remplaçant $(\eta(0), (Y_k^\tau)_{0 \leq k \leq na})$ par $(\eta(jna\tau), (Y_k^\tau)_{jna \leq k \leq (j+1)na})$. Nous montrons alors par récurrence que, si on prend $\eta(0)$ égal à $Y^\tau(0)$, pour tout j entier de $[0, \dots, \frac{T}{na\tau}]$, Δ_{jn} est majoré par un terme de la forme $C(na\tau)\theta'$ (résultat (2.27)). Puis nous utilisons le lemme 2.1.3 sur chaque intervalle $[jna\tau; (j+1)na\tau]$ en remplaçant $(Y_k^\tau)_{0 \leq k \leq na}$ par $(Y_k^\tau)_{jna \leq k \leq (j+1)na}$.

Construction de η sur $[0; na\tau]$ et majoration de l'écart avec $(Y_k^\tau)_{0 \leq k \leq na}$

A la chaîne de Markov $(Y_k^\tau)_{0 \leq k \leq na}$, on associe deux variables aléatoires liées à la position de la chaîne par rapport à P^0 :

1. on prend σ égal à $n \wedge \inf \{k \geq 0 \mid |\pi(Y_{ka}^\tau)| \leq a\tau M\}$. L'instant σa est un $(\mathcal{F}_k)$ -temps d'arrêt correspondant, à a près, à la première entrée de $(Y_l^\tau)_{l \in \mathbb{N}}$ dans

$(P^0)^{a\tau M}$. Remarquons que, d'après le choix de M (1.16), $(P^0)^{a\tau M}$ correspond à la zone en dehors de laquelle on ne peut atteindre la frontière en moins de a pas.

2. on appelle σ' le plus grand entier compris entre σ et n , tel que $Y_{\sigma'a}^\tau, \dots, Y_{na}^\tau$ soient dans le même P^i , pour $i = 1$ ou 2 (ce n'est pas un temps d'arrêt). Si Y_{na}^τ appartient à P^0 , on prend σ' égal à n . L'instant $\sigma'a$ correspond, à a près, à la dernière traversée de P^0 de la chaîne $(Y_k^\tau)_{0 \leq k \leq na}$, sauf éventuellement pour $\sigma' = \sigma$, signifiant qu'il n'y a aucune traversée (la valeur commune n'est pas nécessairement n puisqu'on peut entrer dans $(P^0)^{a\tau M}$ sans qu'il y ait de traversée).

On donne la construction de $\eta(\omega)$ pour chaque ω de $\Omega_0 = \{Y_0^\tau \in \mathcal{K}\}$. Si $\sigma(\omega)$ est égal à $\sigma'(\omega)$ (en particulier si $\sigma(\omega)$ est égal à n), les na pas sont tous du même type. Sinon, il y a "mélange" des deux champs (μ_x^1) et (μ_x^2) . Nous supposons par exemple que $Y_0^\tau(\omega)$ appartient à $P^1 \cup P^0$ et, si $\sigma(\omega) < n$, que $Y_{na}^\tau(\omega)$ appartient à P^2 ; les autres cas sont traités de façon analogue. On a alors :

$$\pi(Y_{\sigma a}^\tau) \leq 0 \text{ et, pour } \sigma(\omega) < n, \quad -a\tau M \leq \pi(Y_{\sigma a}^\tau) \quad (2.10)$$

P -presque sûrement sur Ω_0 , la chaîne $(Y_k^\tau)_{0 \leq k \leq na}$ reste dans $\mathcal{K}^{na\tau M}$. On recouvre ce compact par des boules de rayon $a\tau M$:

$$\mathcal{K}^{na\tau M} \subset \bigcup_q B(x_q, a\tau M), \quad (2.11)$$

et pour $0 \leq k < n$, on prend $x(k)$ (variable aléatoire) parmi les centres de ces boules tel que Y_{ka}^τ appartienne à $B(x(k), a\tau M)$ P -ps sur $\{Y_0^\tau \in (P^1 \cup P^0) \cap \mathcal{K}\}$.

On pose, pour $i = 1$ ou 2 et $k \in \mathbb{N}$,

$$v_k^i = h_{x(k), \theta'}^i(w_k^i),$$

et, pour $i = 1$ ou 2 et $k < k' \in \mathbb{N}$,

$$v_{k,k'}^i = \frac{1}{(k' - k)r_{k,k'}^i} \sum_{l=k}^{k'-1} r_l^i v_l^i \text{ et } v_{k,k'} = \sum_{i=1}^2 r_{k,k'}^i v_{k,k'}^i$$

Remarquons que $(k' - k)r_{k,k'}^i$ est égal à $\sum_{l=k}^{k'-1} r_l^i$. D'autre part, le vecteur $v_{k,k'}^i$ est aussi égal au barycentre

$$\theta' \frac{1}{(k' - k)r_{k,k'}^i} \sum_{l=k}^{k'-1} r_l^i m_{x(l)}^i + (1 - \theta') w_{k,k'}^i.$$

La construction de η sur $[0, na\tau]$ est faite ci-dessous. Nous en donnons d'abord un résumé rapide. Les différents choix pour η' sont motivés par le fait qu'on veut, à tout instant, pouvoir majorer $\lambda_{\theta,h}(\eta(t), \eta'(t))$. Le vecteur $\eta'(t)$ est pris égal à des moyennes de pas de la chaîne. Or l'expression de $\lambda_{\theta,h}(\eta(t), \cdot)$ dépend fortement de la position de $\eta(t)$ par rapport à P^0 et on ne contrôle pas toujours sa valeur sur ces moyennes (par exemple quand on a une moyenne de pas de type 1 et que $\eta(t)$ appartient à P^2). Il est alors parfois nécessaire, pendant des durées arbitrairement petites, de remplacer ces moyennes par $v_{Y_0^\tau}$ vecteur pour lequel $\lambda_{\theta,h}(x, v_{Y_0^\tau})$ est parfaitement contrôlé, dès que x est dans $(P^0)^h$ et proche de Y_0^τ .

Le chemin η vérifie, pour un bon choix de h :

1. Sur $[0, \sigma a\tau]$
 - (a) Si $\eta(0) \in P^1$:
sur un premier intervalle $[0, k_1 a\tau]$ tel que $\eta([0, k_1 a\tau])$ est inclus dans P^1 , $\eta'(t)$ vaut v_k^1 pour t dans $[ka\tau; (k+1)a\tau]$ avec $k < k_1$. Sur $[k_1 a\tau, \sigma a\tau]$, η' vaut ensuite $v_{k_1, \sigma}$ ou $v_{Y_0^\tau}$ et $\eta([k_1 a\tau, \sigma a\tau])$ est inclus dans $(P^0)^h$.
 - (b) Si $\eta(0) \in P^0 \cup P^2$:
 η' vaut $v_{0, \sigma}$ ou $v_{Y_0^\tau}$ et $\eta([0, \sigma a\tau])$ est inclus dans $(P^0)^h$.
2. Sur $[\sigma a\tau, \sigma' a\tau]$, η' vaut $v_{\sigma, \sigma'}$, $v_{\sigma, \sigma'}^1$ ou $v_{Y_0^\tau}$. Le chemin η reste dans $(P^0)^h$.
3. Sur $[\sigma' a\tau, na\tau]$, on a une des situations suivantes :
 η' vaut $v_{\sigma', n}$ et η reste dans $(P^0)^h$ ou P^2 , η' vaut $b_{Y_0^\tau}^1$ et η reste dans P^1 ou η' vaut $v_{Y_0^\tau}$ et η reste dans $(P^0)^h$.

Le processus reste presque sûrement dans $\mathcal{U}$ sur $[0, T]$ et on construit η tel que, presque sûrement en t , $\eta'(t)$ appartient à $\bigcup_{y \in \mathcal{U}} \text{Conv}(S_y^1, S_y^2)$ (notation 1.4.7), on sait donc qu'on va avoir, pour tout t de $[0, na\tau]$, $|\eta'(t)| \leq M$ (cf (1.16)) et que η va rester dans $\mathcal{U}$ quitte à agrandir cet ensemble.

On suppose

$$a\tau M \leq \theta' \quad (2.12)$$

et

$$2M\theta' \leq \frac{\theta}{3}. \quad (2.13)$$

On a alors, pour $i = 1$ ou 2 et $k < k' \in \mathbb{N}$:

$$|v_k^i - w_k^i| \leq 2M\theta' \leq \frac{\theta}{3} \quad \text{et} \quad |v_{k,k'}^i - w_{k,k'}^i| \leq 2M\theta' \leq \frac{\theta}{3}. \quad (2.14)$$

Remarque 2.3.4 *L'utilisation de la méthode de [AR] pourrait être envisagée en définissant deux champs de vecteurs $W_{\sigma,\sigma'}^1(\cdot)$ et $W_{\sigma,\sigma'}^2(\cdot)$ de la manière suivante :*

pour $y \in \mathbb{R}^d$, on note $W_{\sigma,\sigma'}^i(y)$ la projection de $v_{\sigma,\sigma'}^i$ sur le compact convexe

$$\left\{ {}^h\overline{L}_y^i(\cdot) \leq \frac{1}{(\sigma' - \sigma)r_{\sigma,\sigma'}^i} \sum_{l=\sigma}^{\sigma'-1} r_l^i L_{Y_l^\tau}^i(v_l^i) \right\}$$

et on prend η comme courbe intégrale du champ $\sum_{i=1}^2 r_{\sigma,\sigma'}^i W_{\sigma,\sigma'}^i$. On aurait alors besoin d'utiliser le fait que les lignes de niveau de l'application $x \mapsto L_x$ sont localement lipschitziennes, ce qui nécessite au moins l'hypothèse (HLip).

• Construction avant l'instant $\sigma a\tau$ (pour $\sigma < 0$):

Remarquons que pour tout $0 \leq k < \sigma$, on a $r_k^1 = 1$ donc $w_k = w_k^1$ et $v_k = v_k^1$.

On distingue deux cas, selon la position de $\eta(0)$ par rapport à P^0 :

1. Si $\eta(0) \in P^1$:

on définit une application $\xi : [0, \sigma a\tau] \rightarrow \mathcal{U}$ par $\xi(0) = \eta(0)$ et $\xi'(t) = v_k$ sur $[ka\tau; (k+1)a\tau]$ pour $0 \leq k < \sigma$. Soit k_1 le plus grand entier tel que $\xi([0, k_1 a\tau])$ soit inclus dans P^1 . On prend η égal à ξ sur $[0, k_1 a\tau]$. D'après (2.14), pour $t \in [ka\tau; (k+1)a\tau]$, on a:

$$|\xi'(t) - w_k| \leq 2M\theta'. \quad (2.15)$$

On en déduit, d'après $Y_{(k+1)a}^\tau = Y_{ka}^\tau + a\tau w_k$ que, pour $0 \leq k < k_1$, on a, avec la notation (2.9), $\Delta_{k+1} \leq \Delta_k + a\tau 2M\theta'$ et donc:

$$\Delta_{k_1} \leq \Delta_0 + k_1 a\tau 2M\theta'. \quad (2.16)$$

Si $k_1 < \sigma$, on distingue deux cas ensuite, selon la valeur de $\pi(w_{k_1,\sigma})$:

– soit $\pi(w_{k_1,\sigma}) \leq$; alors, on prend η' égal à $v_{k_1,\sigma}$ sur $[k_1 a\tau; \sigma a\tau]$. On a d'après (2.14):

$$\Delta_\sigma \leq \Delta_{k_1} + (\sigma - k_1) a\tau 2M\theta' \leq \Delta_0 + \sigma a\tau 2M\theta'.$$

Remarquons qu'on aurait pu également prendre η' égal à v_{0,k_1} sur $[0; k_1 a\tau]$, mais cela ne s'imposait pas comme ici : on aura besoin de $|\pi(\eta'(t))| \leq \theta$ pour contrôler le coût sur l'intervalle $[k_1 a\tau; \sigma a\tau]$, ce qui est assuré par (2.14).

— soit $\pi(w_{k_1,\sigma}) > \frac{2\theta}{3}$; alors, on prend $\eta'(t)$ égal à $v_{Y_0^\tau}$ sur $[k_1 a\tau, \sigma a\tau]$. En majorant simplement les $|w_k - v_{Y_0^\tau}|$ par $2M$, pour $k_1 \leq k < \sigma$, et k_1 par σ , on peut écrire :

$$\Delta_\sigma \leq \Delta_{k_1} + 2M(\sigma - k_1)a\tau \leq \Delta_0 + \sigma a\tau 2M\theta' + 2M(\sigma - k_1)a\tau.$$

Montrons que $(\sigma - k_1)a\tau$ est nécessairement petit : pour cela, on reprend le chemin ξ sur $[0, \sigma a\tau]$. D'après le choix de k_1 et puisque $k_1 < \sigma$, on a $-a\tau M \leq \pi(\xi(k_1 a\tau))$ et donc, d'après (2.12) :

$$-\theta' \leq \pi(\xi(k_1 a\tau)). \quad (2.17)$$

D'après (2.13) et (2.15), $|\xi'(t) - w_k|$ est inférieur à $\frac{\theta}{3}$ pour tout t de l'intervalle $[ka\tau, (k+1)a\tau]$, avec $k_1 \leq k < \sigma$; on en déduit par intégration qu'on a :

$$\left| (\sigma - k_1)a\tau w_{k_1,\sigma} - \int_{k_1 a\tau}^{\sigma a\tau} \xi'(t) dt \right| \leq (\sigma - k_1)a\tau \frac{\theta}{3}.$$

Or, $\pi(w_{k_1,\sigma}) > \frac{2\theta}{3}$ donc

$$(\sigma - k_1)a\tau \frac{\theta}{3} \leq \int_{k_1 a\tau}^{\sigma a\tau} \pi(\xi'(t)) dt = \pi(\xi(\sigma a\tau)) - \pi(\xi(k_1 a\tau)).$$

Comme $|\pi(\xi(\sigma a\tau)) - Y_{\sigma a}^\tau| \leq \Delta_0 + \sigma a\tau 2M\theta'$, on en déduit, d'après (2.17) :

$$(\sigma - k_1)a\tau \frac{\theta}{3} \leq \Delta_0 + \sigma a\tau 2M\theta' + \pi(Y_{\sigma a}^\tau) + \theta',$$

et donc, puisque $\pi(Y_{\sigma a}^\tau) < 0$, on a la majoration suivante de la durée pendant laquelle $\eta'(t)$ vaut $v_{Y_0^\tau}$:

$$(\sigma - k_1)a\tau \leq \frac{3}{\theta}(\Delta_0 + \theta' + \sigma a\tau 2M\theta'). \quad (2.18)$$

On en déduit $\Delta_\sigma \leq \left(1 + \frac{6M}{\theta}\right)(\Delta_0 + \theta' + \sigma a\tau 2M\theta')$.

2. Si $\eta(0) \in P^0 \cup P^2$, on distingue deux cas :

— soit $\pi(w_{0,\sigma}) \leq \frac{2\theta}{3}$; alors on effectue une construction analogue à celle choisie

sur $[k_1 a\tau, \sigma a\tau]$ pour $\pi(w_{k_1, \sigma}) \leq \frac{2\theta}{3}$: on prend η' égal à $v_{0, \sigma}$ sur $[0, \sigma a\tau]$. On a d'après (2.14) :

$$\Delta_\sigma \leq \Delta_0 + \sigma a\tau 2M\theta'$$

– soit $\pi(w_{0, \sigma}) > \frac{2\theta}{3}$; alors on prend η' égal à $v_{Y_0^\tau}$ sur $[0, \sigma a\tau]$.

On a $\Delta_\sigma \leq \Delta_0 + \sigma a\tau 2M$, avec $-\Delta_0 \leq \pi(Y_0^\tau)$, $\pi(Y_0^\tau + \sigma a\tau w_{0, \sigma}) \leq 0$ (d'après (2.10)) et $\pi(w_{0, \sigma}) > \frac{2\theta}{3}$, donc, $\sigma a\tau$ est inférieur à $\frac{3}{2\theta}\Delta_0$. On en déduit :

$$\Delta_\sigma \leq \left(1 + \frac{3M}{\theta}\right)\Delta_0.$$

Nous avons donc dans tous les cas :

$$\Delta_\sigma \leq \left(1 + \frac{6M}{\theta}\right)(\Delta_0 + \theta' + \sigma a\tau 2M\theta'). \quad (2.19)$$

• Construction de η entre les instants $\sigma a\tau$ et $\sigma' a\tau$ (pour $\sigma < \sigma'$) :

Le vecteur $(\sigma' - \sigma)a\tau r_{\sigma, \sigma'}^1, w_{\sigma, \sigma'}^1$, est une somme d'excursions commençant toutes dans $P^1 \cup P^0$ et se terminant dans P^2 (somme des pas entre le σ^e et le $(\sigma' - 1)^e$ qui sont soumis au champ 1) ; donc $\pi(w_{\sigma, \sigma'}^1)$ est positif.

De même, $(\sigma' - \sigma)a\tau r_{\sigma, \sigma'}^2, w_{\sigma, \sigma'}^2$, est une somme d'excursions commençant toutes dans P^2 et se terminant dans $P^1 \cup P^0$, sauf la dernière, qui correspond à au plus a pas. On a donc $(\sigma' - \sigma)a\tau r_{\sigma, \sigma'}^2, \pi(w_{\sigma, \sigma'}^2) \leq a\tau M$. D'où l'implication :

$$\left((\sigma' - \sigma)r_{\sigma, \sigma'}^2 \geq \frac{3M}{2\theta}\right) \Rightarrow \left(\pi(w_{\sigma, \sigma'}^2) \leq \frac{2\theta}{3}\right) \quad (2.20)$$

D'autre part (cf (2.10)) :

$$\pi(Y_{\sigma a}^\tau) \in [-a\tau M, 0] \text{ (car } \sigma < n), \text{ et } \pi(Y_{\sigma' a}^\tau) \in [0, a\tau M] \quad (2.21)$$

donc

$$|\pi(w_{\sigma, \sigma'})| = |\pi(r_{\sigma, \sigma'}^1, w_{\sigma, \sigma'}^1 + r_{\sigma, \sigma'}^2, w_{\sigma, \sigma'}^2)| = \frac{|\pi(Y_{\sigma' a}^\tau - Y_{\sigma a}^\tau)|}{(\sigma' - \sigma)a\tau} \leq \frac{2M}{\sigma' - \sigma} \quad (2.22)$$

Cette quantité est majorée par $\frac{2\theta}{3}$ dès que $\sigma' - \sigma$ est supérieur à $\frac{3M}{\theta}$.

On distingue trois cas pour la construction de η :

1. Si $\pi(w_{\sigma, \sigma'}^2) \leq \frac{2\theta}{3}$ et $\sigma' - \sigma \geq \frac{3M}{\theta}$,

on prend η' égal à $v_{\sigma, \sigma'}$ sur $[\sigma a\tau, \sigma' a\tau]$. Alors $|\eta'(t) - w_{\sigma, \sigma'}| \leq 2M\theta'$; on a donc :

$$\Delta_{\sigma'} \leq \Delta_\sigma + (\sigma' - \sigma)a\tau 2M\theta'.$$

2. Si $\pi(w_{\sigma,\sigma'}^2) > \frac{2\theta}{3}$ et $\pi(w_{\sigma,\sigma'}^1) \leq \frac{2\theta}{3}$,
 alors, d'après (2.20), $(\sigma' - \sigma)\tau_{\sigma,\sigma'}^2 < \frac{3M}{2\theta}$. On prend η' égal à $v_{\sigma,\sigma'}^1$ sur $[\sigma a\tau, \sigma' a\tau]$.
 Par rapport au cas précédent, on a un accroissement supplémentaire de $\Delta_{\sigma'}$
 d'au plus $(\sigma' - \sigma)a\tau \tau_{\sigma,\sigma'}^2 |v_{\sigma,\sigma'}^1 - v_{\sigma,\sigma'}^2|$, quantité qui reste inférieure à $a\tau \frac{3M}{2\theta} 2M$.
 Donc

$$\Delta_{\sigma'} \leq \Delta_{\sigma} + (\sigma' - \sigma)a\tau 2M\theta' + \frac{3M^2}{\theta}a\tau.$$

3. Si $\pi(w_{\sigma,\sigma'}^2) > \frac{2\theta}{3}$ et $\pi(w_{\sigma,\sigma'}^1) > \frac{2\theta}{3}$, ou $\pi(w_{\sigma,\sigma'}^2) \leq \frac{2\theta}{3}$ et $\sigma' - \sigma < \frac{3M}{\theta}$,
 dans les deux cas, $\sigma' - \sigma < \frac{3M}{\theta}$ (d'après (2.22) pour le premier). On prend η' égal
 à $v_{Y_0^{\tau}}$ sur $[\sigma a\tau, \sigma' a\tau]$. Alors, les inégalités $(\sigma' - \sigma)a\tau \leq \frac{3M}{\theta}a\tau$ et $|w_{\sigma,\sigma'} - v_{Y_0^{\tau}}| \leq 2M$ impliquent :

$$\Delta_{\sigma'} \leq \Delta_{\sigma} + \frac{6M^2}{\theta}a\tau$$

Nous avons donc dans tous les cas :

$$\begin{aligned} \Delta_{\sigma'} &\leq \Delta_{\sigma} + (\sigma' - \sigma)a\tau 2M\theta' + \frac{6M^2}{\theta}a\tau \\ &\leq \left(1 + \frac{6M}{\theta}\right)(\Delta_0 + \theta' + \sigma'a\tau 2M\theta') + \frac{6M^2}{\theta}a\tau \quad \text{d'après (2.19).} \end{aligned}$$

• Construction de η après l'instant $\sigma'a\tau$ (pour $\sigma' \neq n$) :

On sait que $Y_{\sigma'a}^{\tau}$ et Y_{na}^{τ} sont dans P^2 . On distingue trois cas selon la valeur de $\pi(w_{\sigma',n})$:

1. Si $|\pi(w_{\sigma',n})| \leq \frac{2\theta}{3}$, on prend η' égal à $v_{\sigma',n}$ sur $[\sigma'a\tau, na\tau]$. D'après (2.14), on a :

$$\Delta_n \leq \Delta_{\sigma'} + (n - \sigma')a\tau 2M\theta'.$$

2. Si $\pi(w_{\sigma',n}) > \frac{2\theta}{3}$,

– soit $\eta(\sigma'a\tau) \in P^2 \cup P^0$, alors on garde la construction précédente, donc la majoration de Δ_n ; on a $\pi(\eta'(t)) > \frac{\theta}{3} > 0$ donc η reste dans P^2 .

– soit $\eta(\sigma'a\tau) \in P^1$, alors on commence par ramener η dans P^2 . On a défini des champs (b_x^i) au moment du choix des v_x (2.3.3). On prend η' égal à $b_{Y_0^{\tau}}^1$ sur l'intervalle $[\sigma'a\tau, t']$, où t' est tel que $\eta(\sigma'a\tau) + (t' - \sigma'a\tau)b_{Y_0^{\tau}}^1$ appartient à P^0 . Sur $[t', na\tau]$, on reprend η' égal à $v_{\sigma',n}$. La trajectoire régularisée η reste

alors dans P^2 sur $[t', n\tau]$. Entre $\sigma'a\tau$ et t' , l'écart entre η et l'application $t \mapsto Y_{ka}^\tau + (t - ka\tau)w_k$ augmente d'au plus $2M(t' - \sigma'a\tau)$. Comme $\pi(b_{Y_0^\tau}^1)$ est plus grand que π_0 , on a, puisque $Y_{\sigma'a}^\tau$ appartient à P^2 :

$$(t' - \sigma'a\tau)\pi_0 \leq -\pi(\eta(\sigma'a\tau)) \leq \Delta_{\sigma'} \quad (2.23)$$

donc $2M(t' - \sigma'a\tau)$ est inférieur à $\Delta_{\sigma'} \frac{2M}{\pi_0}$. Après t' , l'inégalité $|\eta'(t) - w_{\sigma',n}| \leq 2M\theta'$ est de nouveau vérifiée. En majorant $n\tau - t'$ par $(n - \sigma')a\tau$, on obtient :

$$\Delta_n \leq \Delta_{\sigma'} \left(1 + \frac{2M}{\pi_0}\right) + (n - \sigma')a\tau 2M\theta'$$

d'où

$$\Delta_n \leq \left(1 + \frac{2M}{\pi_0}\right)(\Delta_{\sigma'} + (n - \sigma')a\tau 2M\theta').$$

3. Si $\pi(w_{\sigma',n}) < -\frac{2\theta}{3}$, alors, on a $n - \sigma' < \frac{3M}{2\theta}$ d'après les inégalités $\pi(Y_{\sigma'a}^\tau) \leq a\tau M$ et $\pi(Y_{\sigma'a}^\tau) + (n - \sigma')a\tau \pi(w_{\sigma',n}) > 0$. On prend $\eta'(t)$ égal à $v_{Y_0^\tau}$ sur $[\sigma'a\tau, n\tau]$; alors, en majorant $|v_{Y_0^\tau} - w_{\sigma',n}|$ par $2M$, et $(n - \sigma')a\tau$ par $\frac{3M}{2\theta}a\tau$, on obtient :

$$\Delta_n \leq \Delta_{\sigma'} + \frac{3M^2}{\theta}a\tau.$$

Nous avons donc dans tous les cas :

$$\Delta_n \leq \left(1 + \frac{2M}{\pi_0}\right)(\Delta_{\sigma'} + (n - \sigma')a\tau 2M\theta') + \frac{3M^2}{\theta}a\tau$$

donc finalement :

$$\Delta_n \leq \left(1 + \frac{2M}{\pi_0}\right) \left(1 + \frac{6M}{\theta}\right) (\Delta_0 + \theta' + n\tau 2M\theta') + \frac{3M^2}{\theta}a\tau \left(1 + 2\left(1 + \frac{2M}{\pi_0}\right)\right) \quad (2.24)$$

Construction de η sur $[0; T]$ et majoration de l'écart avec $(Y_k^\tau)_{0 \leq k \leq [\frac{T}{\tau}]_\tau}$

La construction précédente peut être faite avec Y_0^τ dans P^2 , on obtient les mêmes résultats. On pose alors $\eta(0) = Y_0^\tau$ et on effectue la construction de η sur $[0, T]$ par récurrence :

soit j_0 l'entier tel que $T \leq j_0 n\tau < T + n\tau$. Pour $j < j_0$, on construit η sur $[jna\tau, (j+1)n\tau]$ de la même façon que sur $[0, n\tau]$. Pour cela, à la chaîne de Markov $(Y_k^\tau)_{jna \leq k \leq (j+1)na}$, on associe des variables σ_j et σ'_j analogues à σ et σ' pour

$(Y_k^\tau)_{0 \leq k \leq na}$ (par exemple σ_j est égal à $[(j+1)n] \wedge \inf \{k \geq jn \mid |\pi(Y_{ka}^\tau)| \leq a\tau M\}$) et on recouvre $\mathcal{U}$ par des boules de rayon $a\tau M$, comme pour $\mathcal{K}^{na\tau M}$ dans (2.11) :

$$\mathcal{U} \subset \bigcup_q B(x_q, a\tau M) \quad (2.25)$$

Enfin, on prend, pour $jn \leq k < (j+1)n$, $x(k)$ parmi les centres de ces boules tel que Y_{ka}^τ appartienne P-ps à $B(x(k), a\tau M)$. La variable $x(k)$ est définie P-ps sur $\{Y_0^\tau \in \mathcal{K}\}$, quitte à agrandir le compact $\mathcal{U}$. On obtient, d'après (2.24), et toujours en supposant (2.12) et (2.13) :

$$\Delta_{(j+1)n} \leq \left(1 + \frac{2M}{\pi_0}\right) \left(1 + \frac{6M}{\theta}\right) (\Delta_{jn} + \theta' + na\tau 2M\theta') + \frac{3M^2}{\theta} a\tau \left(1 + 2\left(1 + \frac{2M}{\pi_0}\right)\right).$$

Cette inégalité s'écrit

$$\Delta_{(j+1)n} \leq C_\theta (\Delta_{jn} + \theta'),$$

où C_θ ne dépend que de θ et des données T , M et π_0 (on a utilisé (2.12) et $na\tau \leq T$). La valeur de θ est elle-même fixe dans le théorème. D'après $\Delta_0 = 0$ et $C_\theta > 1$, on en déduit que pour tout $j \leq j_0$, on a :

$$\Delta_{jn} \leq (C_\theta)^{j_0} \frac{\theta'}{C_\theta - 1}.$$

(On utilise $\Delta_{(j+1)n} \leq \alpha(\Delta_{jn} + \beta) \Rightarrow \Delta_{jn} \leq \alpha^j \Delta_0 + \beta \frac{\alpha^j - 1}{\alpha - 1} \leq \alpha^j \frac{\beta}{\alpha - 1}$, pour $\alpha > 1$.)

On pose alors

$$C(na\tau) = \frac{e^{(\frac{T}{na\tau} + 1) \log C_\theta}}{C_\theta - 1}. \quad (2.26)$$

On a, pour $j \leq j_0$:

$$\boxed{\Delta_{jn} \leq C(na\tau) \theta'} \quad (2.27)$$

et cela est valable dès que les conditions (2.12) et (2.13) sont vérifiées.

Quitte à augmenter C_θ , on assure en fait :

$$\Delta_{jn} + \theta' + na\tau 2M\theta' \leq C(na\tau) \theta',$$

ce qui permet, en utilisant l'équivalent de (2.18) sur l'intervalle $[jna\tau, (j+1)na\tau]$, d'obtenir :

$$(\sigma_j - k_{1,j})a\tau \leq \frac{3}{\theta} C(na\tau) \theta', \quad (2.28)$$

où $k_{1,j}$ désigne l'équivalent de k_1 sur $[jna\tau, (j+1)n\tau]$. On a également :

$$\Delta_{\sigma'_j} \leq C(na\tau)\theta', \quad (2.29)$$

puisque $\Delta_{\sigma'_j}$ est inférieur au majorant de $\Delta_{(j+1)n}$. Alors, comme dans (2.23) et d'après (2.29), on a :

$$(t'_j - \sigma'_j a\tau)\pi_0 \leq C(na\tau)\theta', \quad (2.30)$$

où t'_j désigne l'équivalent de t' sur $[jna\tau, (j+1)n\tau]$.

Enfin, comme η' est bornée par M , on a, d'après (2.27) :

$$d_{0,T}(\gamma^\tau, \eta) \leq C(na\tau)\theta' + 2na\tau M, \quad (2.31)$$

et, pour tous j, k tels que $0 \leq jna\tau \leq ka\tau < (j+1)n\tau \leq T + na\tau$ et tout t de l'intervalle $[jna\tau, (j+1)n\tau]$,

$$|\eta(t) - Y_{ka}^\tau| \leq C(na\tau)\theta' + 2na\tau M \quad (2.32)$$

Remarquons que la quantité $C(na\tau)$ croît quand $na\tau$ décroît et tend vers $+\infty$ quand $na\tau$ tend vers 0. On prendra θ' petit pour $na\tau$ fixé pour assurer $d_{0,T}(\gamma^\tau, \eta)$ petit.

Majoration d'un coût approché de η

Lemme 2.3.5

Soit $\varepsilon > 0$. Sous certaines conditions sur les différents paramètres, on a : pour h plus grand que $C(na\tau)\theta' + (2n+1)a\tau M$ et pour tout j tel que $jna\tau \leq T + na\tau$,

$$\Lambda_{0,jna\tau}^{\theta,h}(\eta) \leq a\tau \sum_{k=0}^{jn-1} \sum_{i=1}^2 r_k^i L_{x(k)}^i(v_k^i) + \varepsilon.$$

Démonstration : on pose $h = C(na\tau)\theta' + (2n+1)a\tau M$ et on note, pour k et k' entiers de $[0, \dots, \frac{T}{a\tau}]$ tels que $k \leq k'$,

$$A_{k,k'} = \Lambda_{ka\tau, k'a\tau}^{\theta,h}(\eta).$$

D'après (2.27) et la valeur choisie pour h , dès que $\sigma_j < (j+1)n$, c'est-à-dire dès que le processus entre dans $(P^0)^{a\tau M}$ entre $jna\tau$ et $(j+1)n\tau$, $\eta([jna\tau, (j+1)n\tau])$ est

inclus dans $(P^0)^h$ (puisque les vitesses du processus et de η restent bornées par M). Dans ce cas on aura, pour tout t dans $[jna\tau, (j+1)n\tau]$:

$$\lambda_{\theta,h}(\eta(t), \eta'(t)) = \begin{cases} L_{\theta}^h(\eta(t), \eta'(t)) & \text{si } |\pi(\eta'(t))| \leq \theta \\ {}^h\overline{L}_{\eta(t)}^i(\eta'(t)) & \text{si } |\pi(\eta'(t))| > \theta \text{ et } \eta(t) \in P^i \end{cases} \quad (2.33)$$

Remarquons que, d'après (2.6) (page 68), pour $i = 1$ ou 2 , dès que $\eta(t)$ appartient à P^i , on a :

$$\lambda_{\theta,h}(\eta(t), \eta'(t)) \leq {}^h\overline{L}_{\eta(t)}^i(\eta'(t)) \quad (2.34)$$

On utilisera que, toujours pour ce choix de h et pour t dans $[jna\tau, (j+1)n\tau]$, $B(\eta(t), h)$ contient les $x(k)$, $jn \leq k < (j+1)n$. En effet, pour t dans $[jna\tau, (j+1)n\tau]$ et $jn \leq k \leq (j+1)n$, on a

$$|\eta(t) - x(k)| \leq |\eta(t) - Y_{ka}^\tau| + a\tau M \leq C(na\tau)\theta' + 2(n+1)a\tau M,$$

d'après (2.32), donc,

$$|\eta(t) - x(k)| \leq h \quad (2.35)$$

Cela permet d'écrire :

$$\text{pour } t \in [jna\tau, (j+1)n\tau], \quad {}^h\overline{L}_{\eta(t)}^i \leq \text{conv}\{L_{x(k)}^i \mid jn \leq k \leq (j+1)n\}. \quad (2.36)$$

Alors, comme $v_{k,k'}^i = \frac{1}{(k' - k)r_{k,k'}^i} \sum_{l=k}^{k'-1} r_l^i v_l^i$, pour $jn \leq k < k' \leq (j+1)n$, on a :

$${}^h\overline{L}_y^i(v_{k,k'}^i) \leq \frac{1}{(k' - k)r_{k,k'}^i} \sum_{l=k}^{k'-1} r_l^i L_{x(l)}^i(v_l^i). \quad (2.37)$$

Soit $\delta > 0$. D'après les hypothèses (H2) et (HC) (qui entraînent la continuité de $x \mapsto m_x^i$ et de $x \mapsto S_x^i$, cf théorème 1.2.8 et sa démonstration), (1.4) est vérifiée : il existe des constantes strictement positives C_1^i et C_2^i telles que, pour tous θ' de $[0, 1]$, $\delta_2 \geq 0$ et y dans $\mathcal{U}$,

$$C_1^i \theta' \geq C_2^i \delta_2 \Rightarrow \text{pour } i = 1 \text{ ou } 2, \quad h_{y,\theta'}^i((S_y^i)^{\delta_2}) \subset (S_y^i)^{-(C_1^i \theta' - C_2^i \delta_2)}.$$

Alors, pour $C_1 = C_1^1 \vee C_1^2$ et $C_2 = C_2^1 \vee C_2^2$, les conditions

$$C_1 \theta' \geq 2\delta \quad (2.38)$$

et

$$\delta \geq C_2 \delta_2, \quad (2.39)$$

impliquent :

$$\text{pour tout } y \text{ de } \mathcal{U}, \text{ pour } i = 1 \text{ ou } 2, h_{y,\theta'}^i((S_y^i)^{\delta_2}) \subset (S_y^i)^{-\delta} \quad (2.40)$$

D'après la continuité de $x \mapsto S_x^i$, il existe δ_3 tel que

$$|x - y| \leq \delta_3, x \in \mathcal{U} \Rightarrow S_y^i \subset (S_x^i)^{\delta_2}. \quad (2.41)$$

On suppose

$$2a\tau M \leq \delta_3 \quad (2.42)$$

Alors, pour $ka \leq l < (k+1)a$, on a $|Y_l^\tau - x(k)| \leq |Y_l^\tau - Y_{ka}^\tau| + a\tau M \leq 2a\tau M \leq \delta_3$ donc $S_{Y_l^\tau}^i$ est inclus dans $(S_{x(k)}^i)^{\delta_2}$. Le vecteur w_k^i est une moyenne de termes de la forme $X_l^i(Y_l^\tau)$ pour $ka \leq l < (k+1)a$. Comme $X_l^i(Y_l^\tau)$ appartient à $S_{Y_l^\tau}^i$, lui-même inclus dans $(S_{x(k)}^i)^{\delta_2}$, w_k^i appartient à $(S_{x(k)}^i)^{\delta_2}$ et d'après (2.40), on a pour $0 \leq k < n$:

$$v_k^i \in (S_{x(k)}^i)^{-\delta}. \quad (2.43)$$

On majore le coût sur $[0, n\tau]$; on procédera de même sur $[jn\tau, (j+1)n\tau]$. Les seules propriétés utilisées sont (2.27), (2.28), (2.29), (2.30), (2.33), (2.36) et (2.37), qui sont vérifiées sur chaque $[jn\tau, (j+1)n\tau]$, et (2.31) et (2.32).

Nous nous plaçons donc sur $[0, n\tau]$ et reprenons successivement chacun des sous-intervalles considérés précédemment.

• Avant l'instant $\sigma a\tau$:

1. Si $\eta(0)$ appartient à P^1 :

Comme $\eta([0, k_1 a\tau])$ est inclus dans P^1 , pour $k < k_1$ et tout t de $[ka\tau, (k+1)a\tau]$, on a $\lambda_{\theta,h}(\eta(t), \eta'(t)) \leq {}^h\overline{L}_{\eta(t)}^1(\eta'(t)) \leq L_{x(k)}^1(\eta'(t))$ d'après (2.36). Donc :

$$A_{0,k_1} \leq a\tau \sum_{k=0}^{k_1-1} L_{x(k)}^1(v_k).$$

Sur $[k_1 a\tau, \sigma a\tau]$, avec $k_1 < \sigma$, on constate que, comme ξ traverse P^0 et que ξ et η coïncident en $k_1 a\tau$, $\eta([k_1 a\tau, \sigma a\tau])$ est inclus dans $(P^0)^h$, donc (2.33) est vérifié.

— si $\pi(w_{k_1,\sigma}) \leq \frac{2\theta}{3}$; soit $\pi(w_{k_1,\sigma}) \geq -\frac{2\theta}{3}$, alors, d'après (2.14), $|\pi(v_{k_1,\sigma})|$ est inférieur à θ , donc, d'après (2.33),

$$\lambda_{\theta,h}(\eta(t), \eta'(t)) = L_{\theta}^h(\eta(t), v_{k_1,\sigma}) \leq {}^h\overline{L}_{\eta(t)}^1(v_{k_1,\sigma}).$$

soit $\pi(w_{k_1, \sigma}) < -\frac{2\theta}{3}$, alors, comme on a $\pi(Y_{k_1}^\tau) < -a\tau M$ et, si $\sigma < n$, $\pi(Y_\sigma^\tau) \geq -a\tau M$, ce cas est possible seulement avec $\sigma = n$. Comme $\pi(\eta'(t)) < -\frac{\theta}{3}$ et $\eta(k_1 a\tau) \in P^1$, nécessairement $\eta([k_1 a\tau, \sigma a\tau])$ est inclus dans P^1 . Donc, pour t dans $[k_1 a\tau, \sigma a\tau]$, d'après (2.34), on a encore

$$\lambda_{\theta, h}(\eta(t), \eta'(t)) \leq {}^h\overline{L}_{\eta(t)}^1(v_{k_1, \sigma}).$$

Dans les deux cas, on déduit de (2.37) (avec $r_k^1 = 1$ pour $k < \sigma$ et $r_{k_1, \sigma}^1 = 1$):

$$\lambda_{\theta, h}(\eta(t), \eta'(t)) \leq \frac{1}{\sigma - k_1} \sum_{k=k_1}^{\sigma-1} L_{x(k)}^1(v_k).$$

Donc :

$$A_{k_1, \sigma} \leq a\tau \sum_{k=k_1}^{\sigma-1} L_{x(k)}^1(v_k).$$

— si $\pi(w_{k_1, \sigma}) > \frac{2\theta}{3}$; puisque $\pi(v_{Y_0^\tau})$ est nul, on a

$$\lambda_{\theta, h}(\eta(t), \eta'(t)) = L_\theta^h(\eta(t), v_{Y_0^\tau}) \leq L_{\eta(t)}(v_{Y_0^\tau}).$$

Or, pour $i = 1$ ou 2 , et $t \in [0, na\tau]$, on a, d'après (2.32), $|\eta(t) - Y_0^\tau| \leq h$. Donc, $b_{Y_0^\tau}^i \in (S_{Y_0^\tau}^i)^{-\delta_0} \subset (S_{\eta(t)}^i)^{-\frac{\delta_0}{2}}$ pour h assez petit devant δ_0 , condition que l'on écrit

$$h \leq h(\delta_0) \tag{2.44}$$

et que l'on suppose vérifiée. On a :

$$\forall t \in [0, na\tau], b_{Y_0^\tau}^i \in (S_{\eta(t)}^i)^{-\frac{\delta_0}{2}}. \tag{2.45}$$

Alors

$$L_{\eta(t)}(v_{Y_0^\tau}) \leq J_{\eta(t)}\left(\frac{\delta_0}{2}\right) \leq J, \tag{2.46}$$

où $J = \sup_{y \in \mathcal{U}} J_y\left(\frac{\delta_0}{2}\right)$ est fini puisque $L_i^1(\cdot)$ est bornée sur $\bigcup_{y \in U} y \times (S_y^i)^{-\frac{\delta_0}{2}}$. On en déduit, d'après (2.28) :

$$A_{k_1, \sigma} \leq (\sigma - k_1)a\tau J \leq \frac{3}{\theta}C(na\tau)\theta'J$$

Au total, on a donc dès que $\eta(0)$ appartient à P^1 :

$$A_{0, \sigma} \leq a\tau \sum_{k=0}^{\sigma-1} L_{x(k)}^1(v_k) + \frac{3}{\theta}C(na\tau)\theta'J.$$

2. Si $\eta(0)$ appartient à $P^0 \cup P^2$, de nouveau $\eta([0, \sigma a\tau])$ est inclus dans $(P^0)^h$ car $Y_0^\tau \in P^1$ implique $\pi(\eta(0)) \leq C(na\tau)\theta'$, et $C(na\tau)\theta' + na\tau M$ est inférieur à h . On a distingué deux cas :

– soit $\pi(w_{0,\sigma}) \leq \frac{2\theta}{3}$; alors, d'après (2.14), si t appartient à $[0, \sigma a\tau]$, $|\pi(\eta'(t))|$ est inférieur à θ , on a donc :

$$\lambda_{\theta,h}(\eta(t), \eta'(t)) = L_\theta^h(\eta(t), v_{0,\sigma}) \leq {}^h\overline{L^1}_{\eta(t)}(v_{0,\sigma}),$$

et

$$A_{0,\sigma} \leq a\tau \sum_{k=0}^{\sigma-1} L_{x(k)}^1(v_k).$$

– soit $\pi(w_{0,\sigma}) > \frac{2\theta}{3}$, alors $\lambda_{\theta,h}(\eta(t), \eta'(t)) = L_\theta^h(\eta(t), v_{Y_0^\tau}) \leq J$ comme dans (2.46). D'où

$$A_{0,\sigma} \leq \sigma a\tau J \leq \frac{3\Delta_0}{2\theta} J \leq \frac{3}{2\theta} C(na\tau)\theta' J.$$

Au total, nous avons donc dans tous les cas :

$$A_{0,\sigma} \leq a\tau \sum_{k=0}^{\sigma-1} L_{x(k)}^1(v_k) + \frac{3}{\theta} C(na\tau)\theta' J.$$

Dans la suite on regarde le coût sur $[\sigma a\tau, na\tau]$ pour $\sigma < n$. On a alors $\eta([\sigma a\tau, na\tau])$ inclus dans $(P^0)^h$. L'égalité (2.33) est donc vérifiée pour tout t de l'intervalle.

• **Entre les instants $\sigma a\tau$ et $\sigma' a\tau$ (pour $\sigma < \sigma'$) :**

1. Si $\pi(w_{\sigma,\sigma'}) \leq \frac{2\theta}{3}$ et $\sigma' - \sigma \geq \frac{3M}{\theta}$.

Alors, d'après (2.22) (encore vrai sur $[jna\tau, (j+1)na\tau]$), pour t dans $[\sigma a\tau, \sigma' a\tau]$, $|\pi(w_{\sigma,\sigma'})|$ est inférieur à $\frac{2\theta}{3}$ donc $|\pi(\eta'(t))|$ à θ , d'après (2.14). Avec (2.33), on a donc $\lambda_{\theta,h}(\eta(t), \eta'(t)) = L_\theta^h(\eta(t), \eta'(t)) = L_\theta^h(\eta(t), v_{\sigma,\sigma'})$. Or, $\pi(w_{\sigma,\sigma'})$ est positif et $\pi(w_{\sigma,\sigma'})$ inférieur à $\frac{2\theta}{3}$, donc, pour $i = 1$ ou 2 , on a $(-1)^i \pi(v_{\sigma,\sigma'}^i) \leq \theta$.

Alors $L_\theta^h(\eta(t), v_{\sigma,\sigma'}) \leq \sum_{i=1}^2 r_{\sigma,\sigma'}^i {}^h\overline{L^i}_{\eta(t)}(v_{\sigma,\sigma'}^i) \leq \frac{1}{\sigma' - \sigma} \sum_{k=\sigma}^{\sigma'-1} \sum_{i=1}^2 r_k^i L_{x(k)}^i(v_k^i)$, d'après (2.37). Donc,

$$A_{\sigma,\sigma'} \leq a\tau \sum_{k=\sigma}^{\sigma'-1} \sum_{i=1}^2 r_k^i L_{x(k)}^i(v_k^i).$$

2. Si $\pi(w_{\sigma,\sigma'}^2) > \frac{2\theta}{3}$ et $\pi(w_{\sigma,\sigma'}^1) \leq \frac{2\theta}{3}$.

Alors, pour t dans $[\sigma a\tau, \sigma' a\tau]$, on a $|\pi(\eta'(t))| = |\pi(v_{\sigma,\sigma'}^1)| \leq \theta$, donc, d'après (2.33),

$$\lambda_{\theta,h}(\eta(t), \eta'(t)) \leq L_{\theta}^h(\eta(t), \eta'(t)) \leq {}^h\overline{L}^1_{\eta(t)}(v_{\sigma,\sigma'}^1) \leq \frac{1}{\sigma' - \sigma} \sum_{k=\sigma}^{\sigma'-1} L_{x(k)}^1(v_k^1).$$

D'où

$$A_{\sigma,\sigma'} \leq a\tau \sum_{k=\sigma}^{\sigma'-1} L_{x(k)}^1(v_k^1).$$

Ainsi, on a

$$A_{\sigma,\sigma'} \leq a\tau \sum_{k=\sigma}^{\sigma'-1} r_k^1 L_{x(k)}^1(v_k^1) + a\tau \sum_{k=\sigma}^{\sigma'-1} r_k^2 \max_k L_{x(k)}^1(v_k^1)$$

ou encore, d'après (2.43) et $\sum_{k=\sigma}^{\sigma'-1} r_k^2 = (\sigma' - \sigma)r_{\sigma,\sigma'}^2 < \frac{3M}{2\theta}$:

$$A_{\sigma,\sigma'} \leq a\tau \left[\sum_{k=\sigma}^{\sigma'-1} r_k^1 L_{x(k)}^1(v_k^1) + \frac{3M}{2\theta} J(\delta) \right].$$

3. Si $\pi(w_{\sigma,\sigma'}^2) > \frac{2\theta}{3}$ et $\pi(w_{\sigma,\sigma'}^1) > \frac{2\theta}{3}$, ou $\pi(w_{\sigma,\sigma'}^2) \leq \frac{2\theta}{3}$ et $\sigma' - \sigma < \frac{3M}{\theta}$: d'après (2.33) et $\pi(v_{Y_0^r}) = 0$, on a

$$\lambda_{\theta,h}(\eta(t), \eta'(t)) \leq L_{\eta(t)}(v_{Y_0^r}).$$

Avec (2.44), on en déduit :

$$A_{\sigma,\sigma'} \leq (\sigma' - \sigma)a\tau J < a\tau \frac{3M}{\theta} J.$$

On suppose

$$\delta \leq \frac{\delta_0}{2}. \quad (2.47)$$

Alors J est inférieur à $J(\delta)$. Au total, nous avons alors dans tous les cas :

$$\boxed{A_{\sigma,\sigma'} \leq a\tau \left[\sum_{k=\sigma}^{\sigma'-1} \sum_{i=1}^2 r_k^i L_{x(k)}^i(v_k^i) + \frac{3M}{\theta} J(\delta) \right]}$$

• Après l'instant $\sigma'a\tau$ (pour $\sigma' \neq n$):

1. Si $|\pi(w_{\sigma',n})| \leq \frac{2\theta}{3}$, d'après (2.33) et puisque $|\pi(v_{\sigma',n})| \leq \theta$, on a :

$$A_{\sigma',n} \leq a\tau \sum_{k=\sigma'}^{n-1} L_{x(k)}^2(v_k)$$

2. Si $\pi(w_{\sigma',n}) > \frac{2\theta}{3}$,

tant que $\eta(t)$ appartient à P^1 , $\lambda_{\theta,h}(\eta(t), \eta'(t)) \leq L_{\eta(t)}^1(b_{Y_0}^1\tau) \leq J$ d'après (2.45).
Quand $\eta(t)$ appartient à P^2 , d'après (2.34), on a :

$$\lambda_{\theta,h}(\eta(t), \eta'(t)) \leq {}^h\overline{L}_{\eta(t)}^2(\eta'(t)) \leq \frac{1}{n-\sigma'} \sum_{k=\sigma'}^{n-1} L_{x(k)}^2(v_k).$$

D'après (2.30), $t' - \sigma'a\tau$ est inférieur à $\frac{C(na\tau)\theta'}{\pi_0}$ donc

$$A_{\sigma',n} \leq a\tau \sum_{k=\sigma'}^{n-1} L_{x(k)}^2(v_k) + \frac{C(na\tau)\theta'}{\pi_0} J.$$

3. Si $\pi(w_{\sigma',n}) < -\frac{2\theta}{3}$, (2.33), $L_{\theta}^h(\eta(t), v_{Y_0}\tau) \leq J$ et $n - \sigma' < \frac{3M}{2\theta}$ impliquent

$$A_{\sigma',n} \leq (n - \sigma')a\tau J \leq a\tau \frac{3M}{2\theta} J.$$

Au total, nous avons donc dans tous les cas :

$$A_{\sigma',n} \leq a\tau \left[\sum_{k=\sigma'}^{n-1} L_{x(k)}^2(v_k) + \frac{3M}{2\theta} J \right] + \frac{C(na\tau)\theta'}{\pi_0} J$$

En faisant la somme des coûts obtenus sur chacun des sous-intervalles, on peut écrire :

$$\begin{aligned} A_{0,n} &\leq a\tau \sum_{k=0}^{\sigma'-1} L_{x(k)}^1(v_k) + \frac{3}{\theta} C(na\tau)\theta' J + \\ &a\tau \sum_{k=\sigma'}^{\sigma'-1} \sum_{i=1}^2 r_k^i L_{x(k)}^i(v_k^i) + a\tau \frac{3M}{\theta} J(\delta) + a\tau \sum_{k=\sigma'}^{n-1} L_{x(k)}^2(v_k) + a\tau \frac{3M}{2\theta} J + \frac{C(na\tau)\theta'}{\pi_0} J \\ &\leq a\tau \left[\sum_{k=0}^{n-1} \sum_{i=1}^2 r_k^i L_{x(k)}^i(v_k^i) + \frac{9M}{2\theta} J \right] + \theta' C(na\tau) J \left(\frac{3}{\theta} + \frac{1}{\pi_0} \right). \end{aligned}$$

Soit $\varepsilon > 0$. On suppose

$$\frac{9M}{\theta} J \leq n\varepsilon \quad (2.48)$$

et

$$\theta' C(na\tau) J \left(\frac{3}{\theta} + \frac{1}{\pi_0} \right) \leq na\tau \frac{\varepsilon}{2} \quad (2.49)$$

Alors

$$A_{0,n} \leq a\tau \sum_{k=0}^{n-1} \left(\sum_{i=1}^2 r_k^i L_{x(k)}^i(v_k^i) + \varepsilon \right)$$

et on a aussi l'inégalité correspondante pour $A_{j_n, (j+1)n}$, avec j tel que $(j+1)na\tau \leq T$. En sommant ces inégalités, on obtient, pour tout j tel que $(j+1)na\tau \leq T$:

$$A_{0,jn} \leq a\tau \sum_{k=0}^{jn-1} \left(\sum_{i=1}^2 r_k^i L_{x(k)}^i(v_k^i) + \varepsilon \right),$$

et donc, comme $na\tau$ est inférieur à T :

$$\Lambda_{0,jna\tau}^{\theta,h}(\eta) \leq a\tau \sum_{k=0}^{jn-1} \sum_{i=1}^2 r_k^i L_{x(k)}^i(v_k^i) + \varepsilon T,$$

inégalité vérifiée sous les conditions suivantes, pour δ fixé: (2.12) et (2.13) (pour valider la majoration de la distance Δ_{jn} (2.27)), puis (2.38), (2.39), (2.42), (2.44), (2.47), (2.48) et (2.49).

La fonctionnelle $\Lambda_{0,jna\tau}^{\theta,h}$ décroît quand h augmente. On a donc, sous les mêmes conditions et pour tout h supérieur à $C(na\tau)\theta' + (2n+1)a\tau M$:

$$\Lambda_{0,jna\tau}^{\theta,h}(\eta) \leq a\tau \sum_{k=0}^{jn-1} \sum_{i=1}^2 r_k^i L_{x(k)}^i(v_k^i) + \varepsilon T,$$

d'où le lemme 2.3.5. $\square$

Fixons k tel que $0 \leq k < \frac{T}{a\tau}$. On utilise le lemme 2.1.3 sur les moyennes pour le processus $(Y_l^\tau)_{ka \leq l < (k+1)a}$ avec $R = a\tau M$: pour ε, δ et θ' fixés, on sait qu'on peut trouver δ_1 tel que, pour $2a\tau M \leq \delta_1$ et x fixé tel que $B(x, a\tau M)$ soit inclus dans $\mathcal{K}$, il existe $N_{\varepsilon, \delta, \theta'}^x$ tel que, pour $a \geq N_{\varepsilon, \delta, \theta'}^x$, $I \geq 0$ et y_0 dans $B(x, a\tau M)$, on ait:

$$P \left(\sum_{i=1}^2 r_k^i L_x^i(h_{x, \theta'}^i(w_k^i)) \geq I \text{ et } h_{x, \theta'}^i(w_k^i) \in (S_x^i)^{-\delta}, \text{ pour } i = 1 \text{ ou } 2 \mid Y_{ka}^\tau = y_0 \right) \leq e^{-a(I-\varepsilon)}.$$

Rappelons que v_k^i est égal à $h_{x(k),\theta'}^i(w_k^i)$ et définissons la variable $\mathcal{F}_{(k+1)a}$ -mesurable suivante :

$$B_k = \sum_{i=1}^2 r_k^i L_{x(k)}^i(v_k^i).$$

Alors, d'après $v_k^i \in (S_{x(k)}^i)^{-\delta}$ (2.43), pour x fixé, avec $a \geq N_{\varepsilon,\delta,\theta'}$ et $I > \varepsilon$, on a, pour tout k tel que $0 \leq k < \frac{T}{a\tau}$:

$$P(B_k \geq I | \mathcal{F}_{ka}) \leq e^{-a(I-\varepsilon)}, \text{ P-ps sur } \{x(k) = x\}$$

(puisque $\{x(k) = x\}$ implique $\{Y_{ka}^\tau \in B(x, a\tau M)\}$), et cela reste vérifié pour $I \leq \varepsilon$.

On somme sur les valeurs prises par la variable $x(k)$, c'est-à-dire les x_q apparaissant dans le recouvrement (2.25). On obtient, puisqu'ils sont en nombre fini, qu'il existe $N_{\varepsilon,\delta,\theta'}$ tel que

$$2a\tau M \leq \delta_1 \quad (2.50)$$

et

$$a \geq N_{\varepsilon,\delta,\theta'} \quad (2.51)$$

impliquent, pour tout $I \geq 0$:

$$P(a\tau B_k \geq I | \mathcal{F}_{ka}) \leq e^{-\frac{1}{\tau}(I-\varepsilon a\tau)}, \text{ P-ps sur } \Omega_0 = \{Y_0^\tau \in \mathcal{K}\}.$$

On a, sous les conditions (2.50) et (2.51), pour tous $0 \leq k < \frac{T}{a\tau}$ et $I \geq 0$:

$$P(a\tau B_k \geq I | \mathcal{F}_{ka}) \leq e^{-\frac{1}{\tau}(I-\varepsilon a\tau)}, \text{ P-ps sur } \Omega_0 = \{Y_0^\tau \in \mathcal{K}\}. \quad (2.52)$$

Nous reprenons une démonstration de [Rou] (pp. 494-495) pour en déduire une majoration de $P(a\tau(B_0 + \dots B_k) \geq I | \mathcal{F}_0)$. Posons $Z_k = a(B_k - \varepsilon)\mathbb{I}_{B_k \geq \varepsilon}$. La variable Z_k est positive, $\mathcal{F}_{(k+1)a}$ -mesurable et vérifie, pour tous $0 \leq k < \frac{T}{a\tau}$ et v positif :

$$P(Z_k \geq v | \mathcal{F}_{ka}) = P(a\tau B_k \geq v\tau + \varepsilon a\tau | \mathcal{F}_{ka}) \leq e^{-v}, \text{ P-ps sur } \Omega_0. \quad (2.53)$$

Pour une variable réelle quelconque Z , de fonction de répartition F , $-\ln(1 - F(Z))$ suit une loi exponentielle de paramètre 1. Considérons pour k fixé la variable W_k telle que $W_k(\omega)$ soit la valeur au point $(Z_k(\omega), \omega)$ de l'application $(v, \omega) \mapsto -\ln P(Z_k \geq v | \mathcal{F}_{ka})$. La variable W_k est $\mathcal{F}_{(k+1)a}$ -mesurable et suit conditionnellement à $\mathcal{F}_{ka}$ une loi exponentielle. On en déduit que les W_k sont indépendantes et d'après (2.53), pour tout $0 \leq k < \frac{T}{a\tau}$, on a P-presque sûrement sur Ω_0 et pour tout v positif :

$$P(W_k \geq v | \mathcal{F}_{ka}) = e^{-v} \geq P(Z_k \geq v | \mathcal{F}_{ka}).$$

En sommant on en déduit, pour q entier inférieur à $\frac{T}{a\tau}$:

$$P(Z_0 + \dots Z_q \geq v | \mathcal{F}_0) \leq P(W_0 + \dots W_q \geq v | \mathcal{F}_0)$$

Le calcul de la transformée de Cramer de la loi exponentielle permet de majorer le terme de droite de cette dernière inégalité par :

$$e^{-q(\frac{v}{q} - \ln \frac{v}{q} - 1)}$$

Ainsi, on a :

$$P(a\tau(B_0 + \dots B_q) \geq I | \mathcal{F}_0) \leq e^{-v+q \ln \frac{v}{q} + q} \quad (2.54)$$

avec $v = \frac{I}{\tau} - (q+1)a\epsilon$.

Soit $T' < T$. Dès que $na\tau < T - T'$, il existe j entier tel que $T' \leq (j+1)na\tau \leq T$. Prenons alors $q = (j+1)n - 1$. D'après le lemme 2.3.5, pour tout h supérieur à $C(na\tau)\theta' + (2n+1)a\tau M$, on a :

$$\Lambda_{0,(j+1)na\tau}^{\theta,h}(\eta) \leq a\tau(B_0 + \dots B_q) + \epsilon.$$

Donc, d'après (2.54), on a P -presque sûrement sur Ω_0 et pour tout h supérieur à $C(na\tau)\theta' + (2n+1)a\tau M$, l'inégalité

$$P(\Lambda_{0,T'}^{\theta,h}(\eta) \geq I | \mathcal{F}_0) \leq P(a\tau(B_0 + \dots B_q) \geq I - \epsilon | \mathcal{F}_0) \leq e^{-\frac{J}{\tau}}, \quad (2.55)$$

avec $J = I - \epsilon - (q+1)a\tau\epsilon - q\tau \ln \frac{I}{q\tau} - q\tau \geq I - 2\epsilon T - \epsilon$ si on suppose :

$$q\tau \ln \frac{I}{q\tau} + q\tau \leq \epsilon T \quad (2.56)$$

L'inégalité (2.55) est vérifiée dès que les conditions suivantes sont réunies : conditions validant le lemme 2.3.5 (énumérées à la fin de la démonstration de ce lemme), conditions (2.50), (2.51), (2.56) et $na\tau < T - T'$.

Donnons-nous maintenant ϵ_0 et θ strictement positifs, $T' < T$ et ϵ tel que ϵ_0 soit égal à $2\epsilon T + \epsilon$. On choisit les paramètres de façon à vérifier toutes les conditions précédentes :

- on prend $K < T - T'$ et tel que

$$2KM \leq \frac{h(\delta_0)}{3} \quad (2.57)$$

(on imposera à $na\tau$ d'être à-peu-près égal à K),

- on choisit ensuite θ' tel que $2M\theta' \leq \frac{\theta}{3}$ (2.13),

$$\theta' C\left(\frac{K}{2}\right) \leq \frac{h(\delta_0)}{3}, \quad (2.58)$$

$$\theta' C\left(\frac{K}{2}\right) J\left(\frac{3}{\theta} + \frac{1}{\pi_0}\right) \leq \frac{K}{2} \frac{\varepsilon}{2} \quad (2.59)$$

et

$$C\left(\frac{K}{2}\right) \theta' + 2KM \leq \varepsilon_0, \quad (2.60)$$

- puis δ tel que $2\delta \leq C_1\theta'$ (2.38) et $\delta \leq \frac{\delta_0}{2}$ (2.47); on associe δ_2 et δ_3 à δ comme dans (2.39) et (2.41). A partir de $\varepsilon, \delta, \theta'$, on détermine $N_{\varepsilon, \delta, \theta'}$ tel que (2.52) soit vérifié pour $a \geq N_{\varepsilon, \delta, \theta'}$, $0 \leq k < \frac{T}{a\tau}$ et $I \geq 0$.
- On fixe ensuite K' tel que

$$K'M \leq \theta' \wedge \frac{h(\delta_0)}{3} \wedge \frac{\delta_1 \wedge \delta_3}{2}, \quad (2.61)$$

$$\left(\frac{K}{K'} - 1\right) \varepsilon \geq \frac{9M}{\theta} J. \quad (2.62)$$

(on imposera à $a\tau$ d'être à peu près égal à K').

- Soit $S > 0$; on prend τ assez petit pour avoir, en $X = (\frac{2T}{K'} - 1)\tau$:

$$X \ln \frac{S}{X} + X \leq \varepsilon T \quad (2.63)$$

et

$$\tau \leq \frac{K'}{N_{\varepsilon, \delta, \theta'}}. \quad (2.64)$$

- Enfin, on impose $K' - \tau < a\tau \leq K'$ et $K - a\tau < na\tau \leq K$, ce qui, pour τ fixé, détermine a et n de manière unique, en garantissant :

- ◇ $na\tau < T - T'$,
- ◇ (2.49), d'après (2.59) et $\frac{K}{2} \leq na\tau$, qui implique $C(na\tau) \leq C(\frac{K}{2})$,
- ◇ $2a\tau M \leq \delta_1 \wedge \delta_3$, d'après (2.61), alors (2.42) et (2.50) sont vérifiées,
- ◇ $d_{0,T}(\gamma^\tau, \eta) \leq \varepsilon_0$ d'après (2.31) et (2.60),
- ◇ (2.56), pour $I \leq S$, d'après (2.63), $a\tau \leq \frac{K'}{2}$ et $q\tau \leq \frac{T}{a}$.
- ◇ (2.12), d'après (2.61),
- ◇ $n \geq \frac{K-K'}{a\tau} > \frac{K}{K'} - 1$ donc (2.48), d'après (2.62)

◇ et $a > N_{\varepsilon, \delta, \theta'}((2.51))$, d'après (2.64).

On pose $h = C(\frac{K}{2})\theta' + (2K + K')M$. Alors, d'après (2.57), (2.58) et (2.61), (2.44) est vérifiée. De plus, h est supérieur à $C(na\tau)\theta' + (2n+1)a\tau M$ et (2.55) est donc vérifiée.

On obtient ainsi, dès que $\tau < \tau_0$, pour tout $I \leq S$ et tout y_0 de $\mathcal{K}$:

$$\begin{cases} P_{y_0}(\Lambda_{0,T'}^{\theta,h}(\eta) \geq I) \leq e^{-\frac{1}{\tau}(I-\varepsilon_0)} \\ d_{0,T}(\gamma^\tau, \eta) \leq \varepsilon_0 \end{cases}$$

Donc, si $\Gamma_{0,T'}^{\theta,h}(I)$ est l'ensemble des chemins continus ψ tels que $\Lambda_{0,T'}^{\theta,h}(\psi) \leq I$, pour τ comme ci-dessus, on peut écrire pour tout y_0 de $\mathcal{K}$:

$$P_{y_0}(d_{0,T}(\gamma^\tau, \Gamma_{0,T'}^{\theta,h}(I)) > \varepsilon_0) \leq P_{y_0}(\eta \notin \Gamma_{0,T'}^{\theta,h}(I)) \leq P_{y_0}(\Lambda_{0,T'}^{\theta,h}(\eta) \geq I) \leq e^{-\frac{1}{\tau}(I-\varepsilon_0)},$$

ce qui prouve le théorème 2.3.2.

2.3.3 Majoration finale : passage à la limite $(\theta, h) \rightarrow (0, 0)$

Pour montrer le théorème 2.2.10, il reste à passer à la limite quand (θ, h) tend vers $(0, 0)$. Pour T , I , h et θ fixés, on note $G_{0,T}^{\theta,h}(I)$ l'ensemble des chemins continus, φ , de $[0, T]$ dans $\mathbb{R}^d$, tels que $\mathcal{L}_{0,T}^{\theta,h}(\varphi) \leq I$ (notation 2.2.7). Montrons tout d'abord que, dans le théorème 2.3.2, on peut remplacer $\Gamma_{0,T'}^{\theta,h}(I)$ par $G_{0,T}^{\theta,h}(I)$. Pour cela, on reprend la fin de la démonstration de ce théorème (fin du paragraphe précédent). D'après la construction de η , P-ps, η'_ω est constante par morceaux donc η_ω appartient à $C^f[0, T]$; alors $\mathcal{L}_{0,T}^{\theta,h}(\eta) \leq \Lambda_{0,T}^{\theta,h}(\eta)$. Ainsi, on a $P_{y_0}(\eta \notin G_{0,T}^{\theta,h}(I)) \leq P_{y_0}(\eta \notin \Gamma_{0,T'}^{\theta,h}(I)) \leq e^{-\frac{1}{\tau}(I-\varepsilon_0)}$, d'où

$$P_{y_0}(d_{0,T}(\gamma^\tau, G_{0,T}^{\theta,h}(I)) > \varepsilon_0) \leq e^{-\frac{1}{\tau}(I-\varepsilon_0)}. \quad (2.65)$$

Le théorème 2.3.2 est donc encore vrai avec, au lieu de $\Lambda_{0,T}^{\theta,h}$, le coût $\mathcal{L}_{0,T}^{\theta,h}$, c'est-à-dire, en posant $\alpha = (\theta, h)$ et en munissant $(\mathbb{R}^+)^2$ de l'ordre lexicographique, un coût $\mathcal{L}_{0,T}^\alpha$ ayant, d'après la proposition 2.2.8, les propriétés suivantes :

- (1) $\alpha \mapsto \mathcal{L}_{0,T}^\alpha$ est décroissante,
- (2) pour tout φ de $C[0, T]$, $\lim_{\alpha \rightarrow 0} \mathcal{L}_{0,T}^\alpha(\varphi) \leq \lim_{\alpha \rightarrow 0} \liminf_{\psi \rightarrow \varphi} \mathcal{L}_{0,T}^\alpha(\psi)$.
(on a même, pour tout α , $\mathcal{L}_{0,T}^\alpha(\varphi) \leq \liminf_{\psi \rightarrow \varphi} \mathcal{L}_{0,T}^\alpha(\psi)$ puisque $\mathcal{L}_{0,T}^\alpha$ est s.c.i., mais cet énoncé va suffire),

On a noté :

- $\mathcal{L}_{0,T}(\varphi) = \lim_{\alpha \rightarrow 0} \mathcal{L}_{0,T}^\alpha(\varphi)$ (donc (2) s'écrit : $\mathcal{L}_{0,T}(\varphi) \leq \lim_{\alpha \rightarrow 0} \liminf_{\psi \rightarrow \varphi} \mathcal{L}_{0,T}^\alpha(\psi)$).
- pour $I \geq 0$, $G_{0,T}^\alpha(I) = \{\varphi \in C[0,T] \mid \mathcal{L}_{0,T}^\alpha(\varphi) \leq I\}$.

D'après le théorème 2.3.2 (sous une forme plus faible puisque θ est mis ici sur le même plan que h et non plus quelconque), pour I fixé, $\mathcal{K}$ compact de $\mathbb{R}^d$ et $\varepsilon > 0$, il existe α_0 et τ_0 tels que, pour $\alpha \leq \alpha_0$, $\tau \leq \tau_0$ et y_0 dans $\mathcal{K}$,

$$P_{y_0}(d_{0,T}(\gamma^\tau, G_{0,T}^\alpha(I)) \geq \varepsilon) \leq e^{-\frac{1}{\tau}(I-\varepsilon)} \quad (2.66)$$

(la probabilité diminue lorsque α diminue donc (2.66) est vraie dès que cette majoration est vérifiée pour $\alpha = \alpha_0$). Montrons que cet énoncé et les hypothèses (1) et (2) impliquent le résultat du théorème 2.2.10, c'est-à-dire que, si G est une partie borélienne de $C[0,T]$ et $\mathcal{K}$ un compact de $\mathbb{R}^d$, on a :

$$\left. \begin{array}{l} \text{pour tout } \varepsilon > 0, \text{ il existe } \tau_0 \text{ tel que, pour } \tau \leq \tau_0 \text{ et } y_0 \text{ dans } \mathcal{K}, \\ \text{on ait : } \tau \log P_{y_0}(\varphi_{\tau/[0,T]}^\tau \in G) \leq -\mathcal{L}(\overline{G}) + \varepsilon. \end{array} \right\} \quad (2.67)$$

Lemme 2.3.6

On suppose (1) et (2) vérifiées. Soit G une partie borélienne de $C[0,T]$. Pour tout $A < \mathcal{L}(\overline{G})$ et tout compact $\mathcal{K}$ de $\mathbb{R}^d$, il existe α_0 et ε_0 tels que $G \cap \{\varphi \mid \varphi(0) \in \mathcal{K}, \text{ et, ps en } t, |\varphi'(t)| \leq M\}$ soit inclus dans $\{\varphi \in C[0,T] \mid d_{0,T}(\varphi, G_{0,T}^{\alpha_0}(A)) \geq \varepsilon_0\}$.

Démonstration : on considère G une partie borélienne de $C[0,T]$ et on prend ε tel que $A + \varepsilon < \mathcal{L}(\overline{G})$. D'après les hypothèses (Hmaj), nous savons que $\overline{G} \cap \{\varphi \mid \varphi(0) \in \mathcal{K} \text{ et ps en } t, |\varphi'(t)| \leq M\}$ est compact (M est associé au compact $\mathcal{K}$ et à T comme dans (1.16), voir aussi (iii) de la proposition 2.2.8). Notons $G_{\mathcal{K}}$ cet ensemble et $B(\varphi, r)$ l'ensemble $\{\psi \in C[0,T] \mid d_{0,T}(\varphi, \psi) < r\}$. D'après (2), pour tout φ de $G_{\mathcal{K}}$, il existe α_φ tel que $\liminf_{\psi \rightarrow \varphi} \mathcal{L}_{0,T}^{\alpha_\varphi}(\psi)$ soit supérieur à $\mathcal{L}_{0,T}(\varphi) - \frac{\varepsilon}{2}$. Alors il existe ε_φ tel que $\mathcal{L}_{0,T}^{\alpha_\varphi}(B(\varphi, \varepsilon_\varphi))$ soit supérieur à $\mathcal{L}_{0,T}(\varphi) - \varepsilon$. Si $\mathcal{L}_{0,T}(\varphi)$ vaut $+\infty$, on prend simplement $\alpha_\varphi, \varepsilon_\varphi$ tels que $\mathcal{L}_{0,T}^{\alpha_\varphi}(B(\varphi, \varepsilon_\varphi))$ soit strictement supérieur à A . Du recouvrement de $G_{\mathcal{K}}$ par les $B(\varphi, \frac{\varepsilon_\varphi}{2})$, avec φ dans G , on extrait un recouvrement fini $G_{\mathcal{K}} \subset \bigcup_{1 \leq i \leq n} B(\varphi_i, \frac{\varepsilon_i}{2})$ (on note $\varepsilon_i = \varepsilon_{\varphi_i}$ et $\alpha_i = \alpha_{\varphi_i}$). On pose $\varepsilon_0 = \inf_{1 \leq i \leq n} \frac{\varepsilon_i}{2}$. Alors, pour tout ψ dans $(G_{\mathcal{K}})^{c_0}$, il existe i tel que $\psi \in B(\varphi_i, \varepsilon_i)$. On a :

$$\mathcal{L}_{0,T}^{\alpha_i}(\psi) \geq \mathcal{L}_{0,T}(\varphi_i) - \varepsilon, \text{ ou } \mathcal{L}_{0,T}^{\alpha_i}(\psi) > A \text{ pour } \mathcal{L}_{0,T}(\varphi_i) = +\infty,$$

et donc, puisque φ_i appartient à G ,

$$\mathcal{L}_{0,T}^{\alpha_i}(\psi) > A.$$

On pose $\alpha_0 = \inf_{1 \leq i \leq n} \alpha_i$. Pour ψ dans $(G_K)^{\varepsilon_0}$, on a, d'après (1) et pour un i convenable: $\mathcal{L}_{0,T}^{\alpha_0}(\psi) \geq \mathcal{L}_{0,T}^{\alpha_i}(\psi) > A$. L'ensemble $(G_K)^{\varepsilon_0}$ est donc disjoint de $G_{0,T}^{\alpha_0}(A)$. Donc G_K est inclus dans $\{\varphi \mid d_{0,T}(\varphi, G_{0,T}^{\alpha_0}(A)) \geq \varepsilon_0\}$, ce qui démontre le lemme. $\square$

Considérons G une partie borélienne de $C[0, T]$ et $\mathcal{K}$ un compact de $\mathbb{R}^d$. Montrons que (2.67) est vérifié. Considérons $\varepsilon > 0$, posons $A = \mathcal{L}(\overline{G}) - \varepsilon$ et prenons α_0 et ε_0 comme dans le lemme 2.3.6. Pour $\alpha \leq \alpha_0$ et $\varepsilon \leq \varepsilon_0$, $G \cap \{\varphi \mid \varphi(0) \in \mathcal{K} \text{ et ps en } t, |\varphi'(t)| \leq M\}$ est inclus dans $\{\varphi \in C[0, T] \mid d_{0,T}(\varphi, G_{0,T}^{\alpha}(A)) \geq \varepsilon\}$. Prenons $\varepsilon \leq \varepsilon_0$. Il existe α_1 inférieur à α_0 et τ_1 tels que, pour τ inférieur à τ_1 et y_0 dans $\mathcal{K}$, (2.66) soit vérifié pour $\alpha = \alpha_1$, avec $I = A$. On a P-presque sûrement sur $\{Y_0^\tau \in \mathcal{K}\}$:

$$\tau \log P(d_{0,T}(\varphi^\tau, G_{0,T}^{\alpha_1}(A)) \geq \varepsilon | \mathcal{F}_0) \leq -(A - \varepsilon)$$

L'ensemble $\{d_{0,T}(\varphi^\tau, G_{0,T}^{\alpha_1}(A)) \geq \varepsilon\}$ contient $\{\varphi_{/[0,T]}^\tau \in G\} \cap \{Y_0^\tau \in \mathcal{K}\}$. Donc, pour $\tau \leq \tau_1$, P-ps sur $\{Y_0^\tau \in \mathcal{K}\}$, on a:

$$\tau \log P(\varphi_{/[0,T]}^\tau \in G | \mathcal{F}_0) \leq -(A - \varepsilon) = -\mathcal{L}(\overline{G}) + 2\varepsilon,$$

d'où le résultat, ce qui achève la démonstration du théorème 2.2.10 (page 72). $\square$

Remarques:

1. le fait d'avoir supposé (2) et non $\mathcal{L}_{0,T}^\alpha$ s.c.i. pour tout α permet d'appliquer cette démonstration dans le cas d'un seul champ:

alors la fonctionnelle $\varphi \mapsto \int_0^T {}^h\overline{L}_{\varphi(t)}(\varphi'(t))dt$ vérifie (1) et (2). Comme la limite de $\int_0^T {}^h\overline{L}_{\varphi(t)}(\varphi'(t))dt$ quand h tend vers 0 est égale à $\int_0^T L_{\varphi(t)}(\varphi'(t))dt$ pour tout φ de $C[0, T]$, le membre de gauche dans (2) est $\Lambda_{0,T}(\varphi)$. On retrouve donc exactement le résultat de [AR] (qui est par ailleurs contenu dans le nôtre) mais sans avoir supposé que les lignes de niveau de la transformée de Cramer sont localement lipschitziennes. Précisément on obtient le résultat suivant:

Théorème 2.3.7

On considère un processus d'apprentissage lent, $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à un champ de probabilités (μ_x^i) vérifiant les hypothèses $(Hmaj)$. On note $\Lambda_{0,T}^i$ la fonctionnelle de Cramer associé au champ (μ_x^i) et $\Lambda^i(G)$ la borne inférieure de cette fonctionnelle sur un ensemble G . On note φ^τ le chemin aléatoire continu défini sur $[0, T]$ par :

$\varphi^\tau(k\tau) = Y_k^\tau$ pour $0 \leq k\tau < T + \tau$ et φ^τ est affine sur $[k\tau; (k+1)\tau]$.

Alors, pour toute partie borélienne G de $C[0, T]$, tout compact $\mathcal{K}$ de $\mathbb{R}^d$ et tout $\varepsilon > 0$, il existe τ_0 tel que $\tau \leq \tau_0$ implique, P -ps sur $\{Y_0^\tau \in \mathcal{K}\}$:

$$\tau \log P(\varphi_{/[0,T]}^\tau \in G | \mathcal{F}_0) \leq -\Lambda^i(\overline{G}) + \varepsilon$$

2. Si $\Lambda_{0,T}^{\theta,h}$ vérifiait (2), on n'aurait pas besoin d'introduire le coût régularisé (cf les commentaires à la fin du paragraphe 2.2.2) ; mais on a pu montrer

$$\Lambda_{0,T}(\varphi) \leq \lim_{(\theta,h) \rightarrow (0,0)} \liminf_{\psi \rightarrow \varphi} \Lambda_{0,T}^{\theta,h}(\psi)$$

seulement pour φ dans $C^{f,1}[0, T]$ et en se limitant à des chemins ψ variant dans $C^f[0, T]$ (inégalité (2.68)).

2.4 Calcul du coût régularisé

2.4.1 Calcul du coût régularisé $\mathcal{L}$ sur les chemins réguliers

On s'intéresse, dans ce paragraphe, aux chemins pour lesquels le coût régularisé $\mathcal{L}_{0,T}$ prend la même valeur que $\Lambda_{0,T}$ qui est le coût attendu (cf paragraphe 1.4.3).

Théorème 2.4.1

On considère deux champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses $(Hmaj)$. Alors, pour tout chemin φ appartenant à $C^{f,1}[0, T]$, $\mathcal{L}_{0,T}(\varphi)$ est égal à $\Lambda_{0,T}(\varphi)$ (fonctionnelles de coût définies respectivement dans les notations 2.2.9 et 1.4.5).

Ce résultat est vrai, en particulier, pour tous les chemins ne rencontrant pas P^0 .

Démonstration : soit φ dans $C^{f,1}[0, T]$. Le coût $\mathcal{L}_{0,T}^{\theta,h}(\varphi)$ est inférieur à $\Lambda_{0,T}^{\theta,h}(\varphi)$. On a donc :

$$\mathcal{L}_{0,T}(\varphi) = \lim_{(\theta,h) \rightarrow (0,0)} \mathcal{L}_{0,T}^{\theta,h}(\varphi) \leq \lim_{(\theta,h) \rightarrow (0,0)} \Lambda_{0,T}^{\theta,h}(\varphi).$$

D'après le théorème 2.2.6, le membre de droite de l'inégalité est égal à $\Lambda_{0,T}(\varphi)$ qui est donc supérieur à $\mathcal{L}_{0,T}(\varphi)$. Pour prouver l'égalité, il suffit de montrer qu'on a

$$\Lambda_{0,T}(\varphi) \leq \lim_{(\theta,h) \rightarrow (0,0)} \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in C^f[0,T]}} \Lambda_{0,T}^{\theta,h}(\psi) \quad (2.68)$$

Le membre de droite de cette inégalité est une limite croissante quand θ et h décroissent vers 0. Remarquons que la démonstration de (2.68) va être proche de celle du lemme 4.3 de [AR] (il y a alors un seul champ ; on ne fait pas d'hypothèse sur φ et ce qu'on montre alors, c'est que le coût est s.c.i.). Notre démonstration permet d'ailleurs d'obtenir

$$\Lambda_{0,T}(\varphi) \leq \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in C^f[0,T]}} \Lambda_{0,T}(\psi),$$

forme affaiblie de semi-continuité en φ puisqu'on se restreint dans la limite à des suites d'éléments de $C^f[0,T]$.

Considérons une suite (ψ_k) dans $C^f[0,T]$ convergeant uniformément vers φ , avec $\lim_{(\theta,h) \rightarrow (0,0)} \lim_{k \rightarrow +\infty} \Lambda_{0,T}^{\theta,h}(\psi_k) = I$. On veut montrer :

$$\Lambda_{0,T}(\varphi) \leq I \quad (2.69)$$

et il suffit de considérer le cas $I < +\infty$.

Il existe une partition finie $0 = t_0 \leq t_1 \leq \dots \leq t_p = T$ telle que $\varphi([t_j, t_{j+1}[[$ soit inclus dans P^{i_j} , avec $i_j = 0, 1$ ou 2 ; quand i_j vaut 0 , φ est C^1 par morceaux sur $[t_j, t_{j+1}]$. Comme dans la démonstration de la proposition 2.2.8 (iii), on peut extraire une sous-suite (ψ_k) telle que $\lim_{(\theta,h) \rightarrow (0,0)} \lim_{k \rightarrow +\infty} \Lambda_{0,T}^{\theta,h}(\psi_k) = I$ et dont les dérivées convergent sur chaque intervalle $[t_j, t_{j+1}]$ vers φ' , pour la topologie faible de $L^2([t_j, t_{j+1}])$.

Soit t dans $[0, T]$.

1. Considérons le cas où $\varphi(t)$ appartient à P^i , pour un i égal à 1 ou 2 . Alors, il existe $\varepsilon > 0$ et V_t un voisinage de t tel que $\varphi(V_t)$ soit inclus dans $(P^i)^{-\varepsilon}$. Pour k assez grand, $d_{0,T}(\varphi, \psi_k)$ est inférieur à $\frac{\varepsilon}{2}$ donc $\psi_k(V_t)$ est inclus dans $(P^i)^{-\frac{\varepsilon}{2}}$. Alors, pour $h < \frac{\varepsilon}{2}$, $\theta \geq 0$ et s dans V_t , on a :

$$\lambda_{\theta,h}(\psi_k(s), \psi'_k(s)) = {}^h\overline{L}^i_{\psi_k(s)}(\psi'_k(s)).$$

Soit V un voisinage fermé de $\text{Conv}\{S_y^i ; y \in B(\varphi(t), h)\}$ (notation 1.4.7). Pour h petit et k grand, $\Lambda_{0,T}^{\theta,h}(\psi_k)$ est fini, donc, presque sûrement en s dans

V_t , le vecteur $\psi'_k(s)$ appartient à $\text{Conv}\{S_y^i ; y \in B(\psi_k(s), h)\}$. Alors, d'après la continuité de $x \mapsto S_x^i$, V contient $\psi'_k(s)$ pour k assez grand et presque tout s dans V'_t , un voisinage de t inclus dans V_t . Donc $\frac{\psi_k(s_2) - \psi_k(s_1)}{s_2 - s_1}$ appartient à V pour k grand et presque tout $s_2 \neq s_1$ dans V'_t . On en déduit que, pour de tels s_1, s_2 , $\frac{\varphi(s_2) - \varphi(s_1)}{s_2 - s_1}$ appartient à V et donc $\varphi'(t)$ à V .

On a montré que pour h petit, presque sûrement en t de $[0, T]$, $\varphi'(t)$ appartient à $\text{Conv}\{S_y^i ; y \in B(\varphi(t), h)\}$. En faisant tendre h vers 0, on obtient :

$$\varphi'(t) \in S_{\varphi(t)}^i. \quad (2.70)$$

2. Considérons le cas où $\varphi(t)$ appartient à P^0 . Pour $x \in \mathbb{R}^d$, on note $\mathcal{A}_x^{\theta, h}$ l'ensemble :

$$\bigcup_{i=1}^2 \text{Conv}\{S_y^i, y \in B(x, h)\} \cup \left(\text{Conv}\left\{S_{y_1}^1 \cap \{\pi(\cdot) \geq -\theta\}, S_{y_2}^2 \cap \{\pi(\cdot) \leq \theta\} ; y_1, y_2 \in B(x, h)\right\} \cap \{|\pi(\cdot)| \leq \theta\} \right)$$

Pour θ et h petits et k grand, $\Lambda_{0,T}^{\theta, h}(\psi_k)$ est fini, alors pour presque tout s de $[0, T]$, $\psi'_k(s)$ appartient à $\mathcal{A}_{\psi_k(s)}^{\theta, h}$. Considérons V un voisinage fermé de $\mathcal{A}_{\psi_k(t)}^{\theta, h}$; comme dans le cas 1, V contient $\psi'_k(s)$ pour k assez grand et presque tout s dans un voisinage de t , et donc $\varphi'(t)$. On en déduit que $\varphi'(t)$ appartient à $\bigcap_{\theta, h} \mathcal{A}_{\varphi(t)}^{\theta, h}$, ce qui se réduit, puisque $\pi(\varphi'(t))$ vaut 0, à la condition :

$$\varphi'(t) \in \text{Conv}\{S_{\varphi(t)}^1 \cap \{\pi(\cdot) \geq 0\}, S_{\varphi(t)}^2 \cap \{\pi(\cdot) \leq 0\}\} \quad (2.71)$$

On accroît les supports S_x par convolution des probabilités μ_x avec les mesures ν^n comme dans la démonstration du lemme 1.3.5. On a de nouveau, pour tout n et tout x , S_x^i inclus dans $(S_x^{i,n})^{-\delta}$, avec $\delta > 0$ fixé, pour tout x, β , $L_x^{i,n}(\beta) \leq L_x^i(\beta)$ et :

$$\lim_{n \rightarrow +\infty} L_x^{i,n}(\beta) = L_x^i(\beta). \quad (2.72)$$

On indique toujours par un exposant n les notations correspondant au champ de mesures $(\mu_x * \nu^n)$: L_x^n , λ^n , $\lambda_{\theta, h}^n, \dots$

Montrons que pour tout x, β , on a :

$$\lim_{n \rightarrow +\infty} L_x^n(\beta) = L_x(\beta). \quad (2.73)$$

Il suffit de montrer $L_x(\beta) \leq \lim_{n \rightarrow +\infty} L_x^n(\beta)$ et cela dans le cas où le membre de droite est fini. Pour n assez grand, $L_x^n(\beta)$ s'écrit alors $\sum_{i=1}^2 \rho_n^i L_x^{i,n}(\beta_n^i)$ avec β_n^i dans le

compact $S_x^{i,n}$ et les conditions habituelles sur (ρ_n^i, β_n^i) , en particulier $\beta = \sum_{i=1}^2 \rho_n^i \beta_n^i$. Soit (ρ^i, β^i) une valeur d'adhérence de la suite (ρ_n^i, β_n^i) . On a $\beta = \sum_{i=1}^2 \rho^i \beta^i$ et (ρ^i, β^i) vérifie les conditions permettant d'écrire :

$$L_x(\beta) \leq \sum_{i=1}^2 \rho^i L_x^i(\beta^i).$$

Si on note L_μ la transformée de Cramer d'une probabilité μ sur $\mathbb{R}^d$, on sait que l'application $(\mu, \beta) \mapsto L_\mu(\beta)$ est s.c.i. sur $E \times \mathbb{R}^d$ où E désigne l'ensemble des probabilités sur $\mathbb{R}^d$ vérifiant (H2), muni de la topologie de la convergence faible. Puisque $\mu_x * \nu^n$ (normalisée correctement pour avoir une probabilité; le coefficient de normalisation tend vers 1), converge faiblement vers μ_x quand n tend vers $+\infty$, on a :

$$L_x(\beta^i) \leq \lim_{n \rightarrow +\infty} L_x^{i,n}(\beta_n^i).$$

On en déduit :

$$L_x(\beta) \leq \lim_{n \rightarrow +\infty} \sum_{i=1}^2 \rho_n^i L_x^{i,n}(\beta_n^i)$$

et donc (2.73).

Supposons maintenant qu'on ait montré, pour chaque n :

$$\int_0^T \lambda^n(\varphi(t), \varphi'(t)) dt \leq \lim_{(\theta, h) \rightarrow (0,0)} \liminf_{k \rightarrow +\infty} \int_0^T \lambda_{\theta, h}(\psi_k(t), \psi'_k(t)) dt \quad (2.74)$$

Alors on a :

$$\begin{aligned} \Lambda_{0,T}(\varphi) &\leq \liminf_{n \rightarrow +\infty} \int_0^T \lambda^n(\varphi(t), \varphi'(t)) dt \\ &\leq \lim_{(\theta, h) \rightarrow (0,0)} \liminf_{k \rightarrow +\infty} \int_0^T \lambda_{\theta, h}(\psi_k(t), \psi'_k(t)) dt = \lim_{(\theta, h) \rightarrow (0,0)} \liminf_{k \rightarrow +\infty} \Lambda_{0,T}^{\theta, h}(\psi_k) = I, \end{aligned}$$

(2.69) est donc démontré.

(pour la première inégalité, on utilise le lemme de Fatou sachant que, d'après (2.72) et (2.73), $\lambda(\varphi(\cdot), \varphi'(\cdot))$ est la limite simple de $\lambda^n(\varphi(\cdot), \varphi'(\cdot))$ quand n tend vers $+\infty$, et pour la deuxième, (2.74)). $\square$

Dans toute la suite, n est fixé. Pour alléger l'écriture, nous choisissons de le sous-entendre dans les notations (coûts et $S_x^{i,n}$) et de signaler par un $\#$ en exposant les moments où nous utilisons le coût initial (sans convolution avec ν^n). Retenons que

si $L_x^i{}^\#(v)$ est fini, alors v appartient à $(S_x^i)^{-\delta}$ (on l'utilisera pour $v = \psi'_k(t)$) et que, d'après (2.70) et (2.71), si $\varphi(t) \in P^i$, alors $\varphi'(t)$ appartient à $(S_{\varphi(t)}^i)^{-\delta}$ et si $\varphi(t) \in P^0$, alors $\varphi'(t)$ appartient à $\text{Conv}\{(S_{\varphi(t)}^1)^{-\delta} \cap \{\pi(\cdot) \geq 0\}, (S_{\varphi(t)}^2)^{-\delta} \cap \{\pi(\cdot) \leq 0\}\}$.

Montrons (2.74) :

Soit $[a, b]$ un des intervalles $[t_j, t_{j+1}]$. Nous allons montrer :

$$\int_0^T \lambda^n(\varphi(t), \varphi'(t)) dt \leq \lim_{(\theta, h) \rightarrow (0, 0)} \liminf_{k \rightarrow +\infty} \int_0^T \lambda_{\theta, h}{}^\#(\psi_k(t), \psi'_k(t)) dt \quad (2.75)$$

1. Considérons $[a, b]$ tel que $\varphi([a, b])$ soit inclus dans P^i , avec $i = 1$ ou 2 .

Alors $\Lambda_{a, b}^i(\varphi)$ est fini. La mesure de Lebesgue de $\{t \in [a, b] \mid \varphi(t) \in (P^0)^h\}$ tend vers 0 quand h tend vers 0. On en déduit que $\int_{\substack{t \in [a, b] \\ \varphi(t) \in (P^0)^h}} L_{\varphi(t)}^i(\varphi'(t)) dt$ tend vers 0 quand h tend vers 0 (dès qu'une fonction positive f est intégrable, si on a des parties A_n telles que $m(A_n)$ tend vers 0 quand n tend vers 0, alors $\int_{A_n} f dm$ tend vers 0 quand n tend vers 0).

Sur un tel intervalle $[a, b]$, on peut donc trouver t tel que le coût en dehors de $[a + t, b - t]$ soit arbitrairement petit. On pose $\mathcal{I} = [a + t, b - t]$. Considérons h tel que $\varphi(\mathcal{I})$ soit inclus dans $(P^i)^{-2h}$. Pour k assez grand, $\psi_k(\mathcal{I})$ est inclus dans $(P^i)^{-h}$, alors on a :

$$\int_{\mathcal{I}} \lambda_{\theta, h}{}^\#(\psi_k(t), \psi'_k(t)) dt = \int_{\mathcal{I}} {}^h\overline{L}^i_{\psi_k(t)}{}^\#(\psi'_k(t)) dt.$$

Pour obtenir (2.75), il suffit donc de montrer :

$$\int_{\mathcal{I}} L_{\varphi(t)}^i(\varphi'(t)) dt \leq \lim_{h \rightarrow 0} \liminf_{k \rightarrow +\infty} \int_{\mathcal{I}} {}^h\overline{L}^i_{\psi_k(t)}{}^\#(\psi'_k(t)) dt. \quad (2.76)$$

Soit $\varepsilon > 0$. L'application $L(\cdot)$ est uniformément continue sur $\cup_{x \in \mathcal{U}} x \times (S_x^i)^{-\delta}$ et $x \mapsto S_x^i$ est continue donc il existe δ' tel que, si x, y appartiennent à $\mathcal{U}$, $|x - y| < \delta'$ implique $(S_y^i)^{-\delta} \subset (S_x^i)^{-\frac{\delta}{2}}$ et, pour tout v de $(S_y^i)^{-\delta}$, $|L_x^i(v) - L_y^i(v)| < \varepsilon$. On prend h inférieur à $\frac{\delta'}{2}$ et k assez grand pour avoir $d_{a, b}(\varphi, \psi_k) < \frac{\delta'}{2}$. Alors, pour t dans $\mathcal{I}$ et y dans $B(\psi_k(t), h)$, $|y - \varphi(t)|$ est inférieur à δ' , donc pour tout v de $(S_y^i)^{-\delta}$, $|L_y^i(v) - L_{\varphi(t)}^i(v)|$ est inférieur à ε . Presque sûrement en t , ${}^h\overline{L}^i_{\psi_k(t)}{}^\#(\psi'_k(t))$ s'écrit $\sum_y \rho_y^i L_y^i{}^\#(\beta_y^i)$ avec $\sum_y \rho_y^i \beta_y^i = \psi'_k(t)$ et β_y^i dans $(S_y^i)^{-\delta}$. La quantité $L_y^i{}^\#(\beta_y^i) + \varepsilon$ est minorée par $L_y^i(\beta_y^i) + \varepsilon$ donc par $L_{\varphi(t)}^i(\beta_y^i)$. En utilisant la convexité de $L_{\varphi(t)}^i$, on en déduit ${}^h\overline{L}^i_{\psi_k(t)}{}^\#(\psi'_k(t)) + \varepsilon \geq L_{\varphi(t)}^i(\psi'_k(t))$ pour presque tout t de $\mathcal{I}$.

Pour obtenir (2.76), il suffit donc de montrer :

$$\int_{\mathcal{I}} L_{\varphi(t)}^i(\varphi'(t))dt \leq \liminf_{k \rightarrow +\infty} \int_{\mathcal{I}} L_{\varphi(t)}^i(\psi'_k(t))dt. \quad (2.77)$$

Or, la convexité des transformées de Cramer L_x^i permet d'écrire, pour tout t de $[0, T]$ et tout v de $(S_{\varphi(t)}^i)^{-\frac{\delta}{2}}$:

$$L_{\varphi(t)}^i(v) \geq L_{\varphi(t)}^i(\varphi'(t)) + \left\langle \frac{\partial L_{\varphi(t)}^i}{\partial v}(\varphi'(t)), v - \varphi'(t) \right\rangle.$$

On a donc, puisque $\psi'_k(t)$ appartient à $\text{Conv}\{(S_y^i)^{-\delta}; y \in B(\varphi(t), \delta')\}$, qui est inclus dans $(S_y^1)^{-\frac{\delta}{2}}$:

$$\int_{\mathcal{I}} L_{\varphi(t)}^i(\psi'_k(t))dt \geq \int_{\mathcal{I}} L_{\varphi(t)}^i(\varphi'(t))dt + \int_{\mathcal{I}} \left\langle \frac{\partial L_{\varphi(t)}^i}{\partial v}(\varphi'(t)), \psi'_k(t) - \varphi'(t) \right\rangle dt.$$

On conclut en utilisant la convergence dans $L^2(\mathcal{I})$ de (ψ'_k) vers φ' . On a bien (2.77) et donc (2.75).

2. Considérons $[a, b]$ tel que $\varphi([a, b])$ soit inclus dans P^0 .

Soit $\varepsilon > 0$. On découpe $[a, b]$ en intervalles $[a_j, a_{j+1}]$ de façon à ramener le coût des ψ_k à un coût calculé au point $\varphi(a_j)$:

on reprend δ' tel que si x, y appartiennent à $\mathcal{U}$, $|x - y| < \delta'$ implique $(S_y^i)^{-\delta} \subset (S_x^i)^{-\frac{\delta}{2}}$ et, pour tout v de $(S_x^i)^{-\delta}$, $|L_x^i(v) - L_y^i(v)| < \varepsilon$. On choisit les intervalles $[a_j, a_{j+1}]$ assez petits pour avoir $|\varphi(t) - \varphi(a_j)| < \frac{\delta'}{3}$ pour t dans $[a_j, a_{j+1}]$. On prend ensuite k assez grand pour avoir $d_{a,b}(\varphi, \psi_k) < \frac{\delta'}{3}$. Alors, pour t dans $[a_j, a_{j+1}]$, $|\psi_k(t) - \varphi(a_j)|$ est inférieur à $\frac{2\delta'}{3}$. En imposant h inférieur à $\frac{\delta'}{3}$, pour y dans $B(\psi_k(t), h)$ avec $t \in [a_j, a_{j+1}]$, $|y - \varphi(a_j)|$ est inférieur à δ' , on a donc $|L_y^i(v) - L_{\varphi(a_j)}^i(v)| < \varepsilon$ pour tout v de $(S_y^i)^{-\delta}$.

Presque sûrement en t , ${}^h\overline{L}_{\psi_k(t)}^i(\psi'_k(t))$ s'écrit $\sum_y \rho_y^i L_y^i(\beta_y^i)$ avec $\sum_y \rho_y^i \beta_y^i = \psi'_k(t)$ et β_y^i dans $(S_y^i)^{-\delta}$. La quantité $L_y^i(\beta_y^i) + \varepsilon$ est minorée par $L_{\varphi(a_j)}^i(\beta_y^i)$. En utilisant la convexité de $L_{\varphi(a_j)}^i$, on en déduit que pour presque tout t de $[a_j, a_{j+1}]$, on a :

$${}^h\overline{L}_{\psi_k(t)}^i(\psi'_k(t)) + \varepsilon \geq L_{\varphi(a_j)}^i(\psi'_k(t)).$$

De même, $L_{\theta}^h(\psi_k(t), \psi'_k(t))$ s'écrit $\sum_{i=1}^2 \rho^i {}^h\overline{L}_{\psi_k(t)}^i(\beta_t^i)$ et ${}^h\overline{L}_{\psi_k(t)}^i(\beta_t^i) + \varepsilon$ est supérieur à $L_{\varphi(a_j)}^i(\beta_t^i)$, donc, pour presque tout t de $[a_j, a_{j+1}]$, on a :

$$L_{\theta}^h(\psi_k(t), \psi'_k(t)) + \varepsilon \geq L_{\theta}^0(\varphi(a_j), \psi'_k(t)).$$

On peut alors minorer, presque sûrement en t , $\lambda_{\theta,h}^{\sharp}(\psi_k(t), \psi'_k(t)) + \varepsilon$ par :

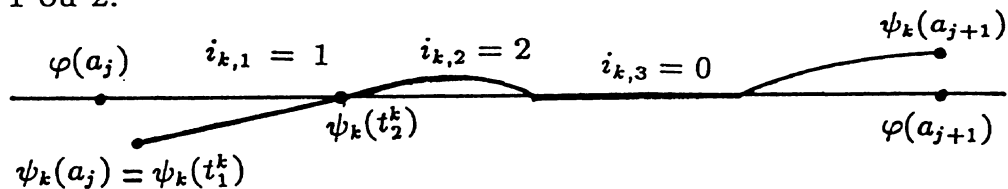
$$\left\{ \begin{array}{l} L_{\theta}^0(\varphi(a_j), \psi'_k(t)) \text{ si } \psi_k(t) \in (P^0)^h \text{ et } |\pi(\psi'_k(t))| \leq \theta, \\ L_{\varphi(a_j)}^i(\psi'_k(t)) \text{ pour } i = 1 \text{ ou } 2 \text{ si } \left\{ \begin{array}{l} \psi_k(t) \in (P^i)^{-h} \\ \text{ou} \\ \psi_k(t) \in (P^0)^h \cap P^i \text{ et } |\pi(\psi'_k(t))| > \theta \end{array} \right. \\ L_{\varphi(a_j)}^0(\psi'_k(t)) \text{ si } \psi_k(t) \in P^0 \text{ et } |\pi(\psi'_k(t))| > \theta \end{array} \right. \quad (2.78)$$

Remarque: le fait de se ramener à des coûts en un même point $\varphi(a_j)$ au lieu de $\varphi(t)$ comme dans le premier cas d'intervalle $[a, b]$ va permettre d'utiliser la convexité de l'application $\beta \mapsto g_{\theta_k}^i(\varphi(a_j), \beta)$ dans (2.80); on remarquera que l'écriture de (2.80) revient à exprimer que pour des champs de probabilités constants, le coût entre deux points pour une durée donnée est minimal pour le chemin affine reliant les deux points en un temps T .

Pour montrer (2.75), considérons une suite (θ_k, h_k) convergeant vers $(0, 0)$ et montrons qu'on a alors :

$$\Lambda_{a,b}(\varphi) \leq \liminf_{k \rightarrow +\infty} \Lambda_{a,b}^{\theta_k, h_k, \sharp}(\psi_k). \quad (2.79)$$

On prend k assez grand pour avoir $d_{a,b}(\varphi, \psi_k)$ et h_k inférieurs à $\frac{\delta'}{3}$. Alors la minoration (2.78) de $\lambda_{\theta,h}^{\sharp}(\psi_k(t), \psi'_k(t)) + \varepsilon$ est vérifiée presque sûrement en t . Pour chaque k , puisque ψ_k appartient à $C^f[0, T]$, on peut découper $[a_j, a_{j+1}]$ en sous-intervalles $[t_l^k, t_{l+1}^k]$ tels que $\psi_k([t_l^k, t_{l+1}^k])$ soit inclus dans $P^{i_{k,l}}$, avec $i_{k,l} = 0, 1$ ou 2 .



Le fait de distinguer les $[t_l^k, t_{l+1}^k]$ va permettre de faire apparaître des mélanges de L^1 et $L_{\theta_k}^0$ ou L^2 et $L_{\theta_k}^0$ mais jamais des trois coûts en même temps, ce qui est fondamental puisqu'on ne veut pas faire apparaître $\text{conv}(L^1, L^2)$.

(a) si $i_{k,l} = 0$; presque sûrement en t , $\pi(\psi'_k(t))$ vaut 0.

L'intégrale $\int_{t_l^k}^{t_{l+1}^k} [\lambda_{\theta_k, h_k}^{\sharp}(\psi_k(t), \psi'_k(t)) + \varepsilon] dt$ est alors minorée, d'après

(2.78), par

$$\int_{t_l^k}^{t_{l+1}^k} L_{\theta_k}^0(\varphi(a_j), \psi'_k(t)) dt$$

donc par

$$(t_{l+1}^k - t_l^k) L_{\theta_k}^0(\varphi(a_j), \beta_{k,l}),$$

où on a posé $\beta_{k,l} = \frac{\psi_k(t_{l+1}^k) - \psi_k(t_l^k)}{t_{l+1}^k - t_l^k}$ (convexité de $L_{\theta_k}^0(\varphi(a_j), \cdot)$).

- (b) si $i_{k,l} = i$, pour $i = 1$ ou 2 , d'après (2.78), $\lambda_{\theta_k, h_k}^{\#}(\psi_k(t), \psi'_k(t)) + \varepsilon$ est minoré par $g_{\theta_k}^i(\varphi(a_j), \psi'_k(t))$, où $g_{\theta}^i(x, \beta)$ est égal à

$$\min \left\{ \rho L_x^i(\beta^i) + (1 - \rho) L_{\theta}^0(x, \xi^i) \mid \beta = \rho \beta^i + (1 - \rho) \xi^i, |\pi(\xi^i)| \leq \theta \right\}$$

(en effet, les deux expressions $L_{\varphi(a_j)}^i(\psi'_k(t))$ et $L_{\theta_k}^0(\varphi(a_j), \psi'_k(t))$ de (2.78) sont possibles, donc on minore par une expression "convexifiée" des deux). $g_{\theta}^i(x, \cdot)$ est convexe puisque L_x^i et $L_{\theta}^0(x, \cdot)$ le sont, ainsi que l'ensemble $\{\beta \mid |\pi(\beta)| \leq \theta\}$. On en déduit, d'après l'inégalité de Jensen :

$$\int_{t_l^k}^{t_{l+1}^k} g_{\theta_k}^i(\varphi(a_j), \psi'_k(t)) dt \geq (t_{l+1}^k - t_l^k) g_{\theta_k}^i(\varphi(a_j), \beta_{k,l}), \quad (2.80)$$

où $\beta_{k,l}$ est défini comme dans le cas précédent : $\beta_{k,l} = \frac{\psi_k(t_{l+1}^k) - \psi_k(t_l^k)}{t_{l+1}^k - t_l^k}$. La quantité $g_{\theta_k}^i(\varphi(a_j), \beta_{k,l})$ est finie ; grâce à la compacité des $S_{\varphi(a_j)}^i$, elle s'écrit :

$$\rho_{k,l}^i L_{\varphi(a_j)}^i(\beta_{k,l}^i) + (1 - \rho_{k,l}^i) L_{\theta_k}^0(\varphi(a_j), \xi_{k,l}^i),$$

avec $\beta_{k,l} = \rho_{k,l}^i \beta_{k,l}^i + (1 - \rho_{k,l}^i) \xi_{k,l}^i$ et $|\pi(\xi_{k,l}^i)| \leq \theta_k$.

On pose, pour k et l fixés, $\alpha_{k,l} = \frac{t_{l+1}^k - t_l^k}{a_{j+1} - a_j}$ et, pour $i = 0, 1$ ou 2 et k entier,

$$\mathcal{T}_k^i = \{l \mid i_{k,l} = i\} \text{ et } \alpha_k^i = \sum_{l \in \mathcal{T}_k^i} \alpha_{k,l} \quad (\text{on a } \sum_{i=0}^2 \alpha_k^i = 1).$$

Pour k fixé, on rassemble les $\beta_{k,l}$ correspondant à $i_{k,l} = 0$: on note β_k^0 leur barycentre, pour les poids $\frac{\alpha_{k,l}}{\alpha_k^0}$:

$$\beta_k^0 = \sum_{l \in \mathcal{T}_k^0} \frac{\alpha_{k,l}}{\alpha_k^0} \beta_{k,l}.$$

Pour $i = 1$ ou 2 , on pose :

$$\rho_k^i = \sum_{l \in \mathcal{T}_k^i} \rho_{k,l}^i, \quad \beta_k^i = \sum_{l \in \mathcal{T}_k^i} \frac{\alpha_{k,l} \rho_{k,l}^i}{\alpha_k^i \rho_k^i} \beta_{k,l}^i \quad \text{et} \quad \xi_k^i = \sum_{l \in \mathcal{T}_k^i} \frac{\alpha_{k,l} (1 - \rho_{k,l}^i)}{\alpha_k^i \rho_k^i} \xi_{k,l}^i.$$

On a, pour $i = 1$ ou 2 :

$$\sum_{l \in \mathcal{T}_k^i} \frac{\alpha_{k,l}}{\alpha_k^i} \beta_{k,l} = \rho_k^i \beta_k^i + (1 - \rho_k^i) \xi_k^i. \quad (2.81)$$

En regroupant tous les sous-intervalles de $[a_j, a_{j+1}]$, on obtient donc :

$$\frac{\psi_k(a_{j+1}) - \psi_k(a_j)}{a_{j+1} - a_j} = \sum_{k,l} \alpha_{k,l} \beta_{k,l} = \alpha_k^0 \beta_k^0 + \sum_{i=1}^2 \alpha_k^i \left[\rho_k^i \beta_k^i + (1 - \rho_k^i) \xi_k^i \right]. \quad (2.82)$$

Or, d'après (a) et (b), on a :

$$\Lambda_{a,b}^{\theta_k, h_k}(\psi_k) + \varepsilon(b-a) = \sum_j \int_{a_j}^{a_{j+1}} [\lambda_{\theta_k, h_k}(\psi_k(t), \psi'_k(t)) + \varepsilon] dt \geq \sum_j (a_{j+1} - a_j) \mathcal{Q}_k^j,$$

où

$$\mathcal{Q}_k^j = \sum_{l \in \mathcal{T}_k^0} \alpha_{k,l} L_{\theta_k}^0(\varphi(a_j), \beta_{k,l}) + \sum_{i=1}^2 \sum_{l \in \mathcal{T}_k^i} \alpha_{k,l} g_{\theta_k}^i(\varphi(a_j), \beta_{k,l}).$$

Puisque

$$\sum_{l \in \mathcal{T}_k^i} \alpha_{k,l} g_{\theta_k}^i(\varphi(a_j), \beta_{k,l}) = \sum_{l \in \mathcal{T}_k^i} \alpha_{k,l} \rho_{k,l}^i L_{\varphi(a_j)}^i(\beta_{k,l}^i) + \sum_{l \in \mathcal{T}_k^i} \alpha_{k,l} (1 - \rho_{k,l}^i) L_{\theta_k}^0(\varphi(a_j), \xi_{k,l}^i),$$

que chaque $L_{\varphi(a_j)}^i$ est convexe et que $L_{\theta_k}^0(\varphi(a_j), \cdot)$ est convexe sur l'ensemble $\{\beta \mid |\pi(\beta)| \leq \theta_k\}$, on en déduit :

$$\mathcal{Q}_k^j \geq \alpha_k^0 L_{\theta_k}^0(\varphi(a_j), \beta_k^0) + \sum_{i=1}^2 \alpha_k^i \left[\rho_k^i L_{\varphi(a_j)}^i(\beta_k^i) + (1 - \rho_k^i) L_{\theta_k}^0(\varphi(a_j), \xi_k^i) \right]. \quad (2.83)$$

Quitte à extraire une sous-suite, on peut supposer que $(\alpha_k^0, \alpha_k^i, \rho_k^i, \beta_k^0, \beta_k^i, \xi_k^i)$ ($i = 1$ ou 2) converge quand k tend vers $+\infty$. Soit $(\alpha^0, \alpha^i, \rho^i, \beta^0, \beta^i, \xi^i)$ sa limite. Alors, on a, d'après (2.82) :

$$\frac{\varphi(a_{j+1}) - \varphi(a_j)}{a_{j+1} - a_j} = \alpha^0 \beta^0 + \sum_{i=1}^2 \alpha^i \left[\rho^i \beta^i + (1 - \rho^i) \xi^i \right]. \quad (2.84)$$

Nous cherchons à minorer, pour k grand, les termes $L_{\varphi(a_j)}^i(\beta_k^i)$, $L_{\theta_k}^0(\varphi(a_j), \beta_k^0)$ et $L_{\theta_k}^0(\varphi(a_j), \xi_k^i)$ apparaissant dans (2.83). Nous traitons le premier terme dans (a) et les deux autres dans (b).

- (a) Considérons (k, l) tel que $i_{k,l} = i$, avec $i = 1$ ou 2 . Alors $\pi(\alpha_{k,l} \beta_{k,l})$ vaut 0, $\frac{-\pi(\psi_k(a_j))}{a_{j+1} - a_j}$ ou $\frac{\pi(\psi_k(a_{j+1}))}{a_{j+1} - a_j}$ donc $\sup_{l \in \mathcal{T}_k^i} |\pi(\alpha_{k,l} \beta_{k,l})|$ tend vers 0 quand k tend vers $+\infty$. D'après (2.81), $|\pi(\rho_k^i \beta_k^i + (1 - \rho_k^i) \xi_k^i)|$ est inférieur à $\frac{1}{\alpha_k^i} \sup_{l \in \mathcal{T}_k^i} |\pi(\alpha_{k,l} \beta_{k,l})|$ donc $|\pi(\rho^i \beta^i + (1 - \rho^i) \xi^i)|$, limite de cette quantité

quand k tend vers $+\infty$, vaut 0 dès que α^i est non nul.

Or $|\pi(\xi_{k,l}^i)| \leq \theta_k$ implique $|\pi(\xi_k^i)| \leq \theta_k$ et donc $\pi(\xi^i) = 0$ (en montrant cela, on montre plus ou moins qu'un chemin minimisant le coût localement suit P^0 , ce qui tend à prouver que les chemins minimisant le coût $\Lambda_{0,T}$ ne peuvent avoir une infinité d'oscillations autour de P^0). On en déduit $\pi(\beta^i) = 0$ et c'est ce qui va permettre la minoration par le bon coût mixte; en effet cela implique $L_{\varphi(a_j)}(\beta^i) \leq L_{\varphi(a_j)}^i(\beta^i)$ et comme $L_{\varphi(a_j)}^i$ est s.c.i., on a :

$$L_{\varphi(a_j)}(\beta^i) \leq \liminf_{k \rightarrow +\infty} L_{\varphi(a_j)}^i(\beta_k^i).$$

Ce résultat sera utilisé pour minorer $\mathcal{Q}_k^j$ à partir de (2.83) si α^i est non nul. On a réussi à minorer avec $L_{\varphi(a_j)}$ et non $\text{conv}(L_{\varphi(a_j)}^1, L_{\varphi(a_j)}^2)$.

(b) Pour traiter les deux autres termes, montrons :

$$\begin{cases} L_{\varphi(a_j)}(\beta^0) \leq \liminf_{k \rightarrow +\infty} L_{\theta_k}^0(\varphi(a_j), \beta_k^0) \\ L_{\varphi(a_j)}(\xi^i) \leq \liminf_{k \rightarrow +\infty} L_{\theta_k}^0(\varphi(a_j), \xi_k^i) \end{cases}$$

$\pi(\beta^0)$ et $\pi(\xi^i)$ sont nuls (vu pour le second et d'après $\pi(\beta_{k,l}) = 0$ quand $i_{k,l}$ vaut 0 pour le premier). Or, pour tout β tel que $\pi(\beta) = 0$, si (β_k) converge vers β , on a pour tout x de P^0

$$L_x(\beta) \leq \liminf_{k \rightarrow +\infty} L_{\theta_k}^0(x, \beta_k)$$

En effet, d'après (2.7), $L_0^0(x, \beta)$ est inférieur à $\liminf_{k \rightarrow +\infty} L_{\theta_k}^0(x, \beta_k)$ or, d'après la remarque 2 page 64, $L_0^0(x, \cdot)$ et L_x coïncident sur $\{\beta \mid \pi(\beta) = 0\}$.

On en déduit

$$\liminf_{k \rightarrow +\infty} \mathcal{Q}_k^j \geq \alpha^0 L_{\varphi(a_j)}(\beta^0) + \sum_{i=1}^2 \alpha^i \left[\rho^i L_{\varphi(a_j)}(\beta^i) + (1 - \rho^i) L_{\varphi(a_j)}(\xi^i) \right]$$

et les vecteurs β^0, β^i et ξ^i sont tous de première coordonnée nulle (si $\alpha^i \neq 0$ pour le deuxième). La convexité de $L_{\varphi(a_j)}$ sur $\{\beta \mid \pi(\beta) = 0\}$ permet alors de minorer le membre de droite par $L_{\varphi(a_j)}\left(\frac{\varphi(a_{j+1}) - \varphi(a_j)}{a_{j+1} - a_j}\right)$, d'après (2.84). Finalement, on a :

$$\liminf_{k \rightarrow +\infty} \Lambda_{a,b}^{\theta_k, h_k}(\psi_k) + \varepsilon(b - a) \geq \sum_j (a_{j+1} - a_j) L_{\varphi(a_j)}\left(\frac{\varphi(a_{j+1}) - \varphi(a_j)}{a_{j+1} - a_j}\right).$$

Pour conclure, on utilise que φ est de classe C^1 par morceaux sur $[a, b]$ (cette hypothèse sert seulement à cet endroit) : quitte à diminuer les intervalles $[a_j, a_{j+1}]$, on peut supposer que φ' varie d'au plus $\frac{\varepsilon}{2}$ sur chacun de ces intervalles (parmi les a_j on prend tous les points de discontinuité de φ'). Pour j fixé, il existe c dans $[a_j, a_{j+1}]$ tel que $\frac{\varphi(a_{j+1}) - \varphi(a_j)}{a_{j+1} - a_j} = \varphi'(c)$. Ce vecteur appartient à $B(\varphi'(a_j), \frac{\varepsilon}{2})$, ensemble inclus dans $(S_{\varphi(a_j)}^i)^{-\frac{\varepsilon}{2}}$, donc, pour t dans $[a_j, a_{j+1}]$,

$$L_{\varphi(a_j)}\left(\frac{\varphi(a_{j+1}) - \varphi(a_j)}{a_{j+1} - a_j}\right) \geq L_{\varphi(t)}\left(\frac{\varphi(a_{j+1}) - \varphi(a_j)}{a_{j+1} - a_j}\right) - \varepsilon.$$

On a donc :

$$\liminf_{k \rightarrow +\infty} \Lambda_{a,b}^{\theta_k, h_k}(\psi_k) + 2\varepsilon(b-a) \geq \sum_j \int_{a_j}^{a_{j+1}} L_{\varphi(t)}\left(\frac{\varphi(a_{j+1}) - \varphi(a_j)}{a_{j+1} - a_j}\right) dt.$$

Posons alors $\chi(t) = \frac{\varphi(a_{j+1}) - \varphi(a_j)}{a_{j+1} - a_j}$ pour t dans $[a_j, a_{j+1}[$. On prend la limite sur une suite de partitions (a_j) dont le pas tend vers 0. Soit χ_m la fonction correspondant à la partition d'indice m . La suite χ_m converge presque partout vers φ' . La fonction $L_{\varphi(t)}$ est s.c.i. donc on a, presque sûrement en t , $L_{\varphi(t)}(\varphi'(t)) \leq \liminf_{m \rightarrow +\infty} L_{\varphi(t)}(\chi_m(t))$. De plus, elle est positive donc, d'après le lemme de Fatou, on a :

$$\Lambda_{a,b}(\varphi) = \int_a^b L_{\varphi(t)}(\varphi'(t)) dt \leq \liminf_{m \rightarrow +\infty} \int_a^b L_{\varphi(t)}(\chi_m(t)) dt.$$

Or, on a montré, pour tout m ,

$$\liminf_{k \rightarrow +\infty} \Lambda_{a,b}^{\theta_k, h_k}(\psi_k) + 2\varepsilon(b-a) \geq \int_a^b L_{\varphi(t)}(\chi_m(t)) dt.$$

Donc

$$\liminf_{k \rightarrow +\infty} \Lambda_{a,b}^{\theta_k, h_k}(\psi_k) + 2\varepsilon(b-a) \geq \Lambda_{a,b}(\varphi).$$

En faisant varier ε , on obtient

$$\liminf_{k \rightarrow +\infty} \Lambda_{a,b}^{\theta_k, h_k}(\psi_k) \geq \Lambda_{a,b}(\varphi).$$

On a montré (2.79). L'inégalité (2.75) est donc vraie également pour ce type d'intervalle $[a, b]$.

Il y a un nombre fini d'intervalles $[a, b]$ à considérer ; on a donc bien montré le théorème 2.4.1. $\square$

Ce théorème prouve qu'on a amélioré les résultats de [Du88] et [DEW] : notre majorant $-\mathcal{L}(\overline{G})$ dans le théorème 2.2.10 est meilleur que $-\Lambda(\overline{G})$. En effet, pour φ de classe C^1 tracée sur P^0 , $-\Lambda_{0,T}(\varphi)$ peut être strictement plus petit que $-\int_0^T (\text{conv}(L_{\varphi(t)}^1, L_{\varphi(t)}^2))(\varphi'(t))$, majorant obtenu en utilisant les résultats de [Du88] ou [DEW] (paragraphe 1.4.4). Cette amélioration est claire dans le théorème final 2.5.1 où la fonctionnelle $\mathcal{L}$ n'apparaît plus.

2.4.2 Simplification de l'expression du coût régularisé

Pour φ n'appartenant pas à $C^{f,1}[0, T]$, $\mathcal{L}_{0,T}(\varphi)$ est calculé comme

$$\lim_{(\theta_k, h) \rightarrow (0,0)} \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in C^f[0,T]}} \Lambda_{0,T}^{\theta_k, h}(\psi).$$

On peut essayer de l'écrire $\liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in C^f[0,T]}} \Lambda_{0,T}(\psi)$ mais, en tout cas, on devra garder

l'approximation par des éléments de $C^f[0, T]$: le coût se calcule sur des chemins suffisamment réguliers au voisinage de P^0 . On peut simplifier l'écriture de $\mathcal{L}_{0,T}$ de la façon suivante :

Théorème 2.4.2

Pour tout φ de $C[0, T]$, dans l'expression du coût

$$\mathcal{L}_{0,T}(\varphi) = \lim_{(\theta, h) \rightarrow (0,0)} \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in C^f[0,T]}} \int_0^T \lambda_{\theta, h}(\psi(t), \psi'(t)) dt,$$

la fonction $\lambda_{\theta, h}$ (notation 2.2.4) peut être remplacée par $\nu_{\theta, h}$ où, pour x et β dans $\mathbb{R}^d$, on a :

$$\nu_{\theta, h}(x, \beta) = \begin{cases} L_x^i(\beta) & \text{si } x \in (P^i)^{-h}, i = 1 \text{ ou } 2 \\ L_\theta^0(x, \beta) & \text{si } |\pi(\beta)| \leq \theta \\ L_x^i(\beta) & \text{si } |\pi(\beta)| > \theta \end{cases} \quad \text{si } x \in P^i \cap (P^0)^h, i = 0, 1 \text{ ou } 2$$

Rappelons qu'on a posé :

$$\lambda_{\theta,h}(x, \beta) = \begin{cases} {}^h\overline{L}_x^i(\beta) & \text{si } x \in (P^i)^{-h}, i = 1 \text{ ou } 2 \\ L_\theta^h(x, \beta) & \text{si } |\pi(\beta)| \leq \theta \\ {}^h\overline{L}_x^i(\beta) & \text{si } |\pi(\beta)| > \theta \end{cases} \quad \text{si } x \in P^i \cap (P^0)^h, i = 0, 1 \text{ ou } 2$$

et que $L_\theta^0(x, \beta)$ est égal à :

$$\inf \left\{ \sum_{i=1}^2 \rho^i L_x^i(\beta^i) \mid \sum_{i=1}^2 \rho^i = 1, \sum_{i=1}^2 \rho^i \beta^i = \beta, \pi(\beta^1) \geq -\theta \text{ et } \pi(\beta^2) \leq \theta \right\}.$$

Démonstration : soit φ dans $C[0, T]$. La fonction $\lambda_{\theta,h}$ est inférieure à $\nu_{\theta,h}$, donc il suffit de montrer :

$$\lim_{(\theta,h) \rightarrow (0,0)} \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in C^f[0,T]}} \int_0^T \nu_{\theta,h}(\psi(t), \psi'(t)) dt \leq \lim_{(\theta,h) \rightarrow (0,0)} \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in C^f[0,T]}} \int_0^T \lambda_{\theta,h}(\psi(t), \psi'(t)) dt.$$

Pour cela, on reprend le début de la démonstration du théorème 2.4.1. Il suffit de considérer le cas où $\mathcal{L}_{0,T}(\varphi)$ est fini et de montrer le résultat en prenant dans le terme de gauche les coûts associés à $(\mu_x * \nu^n)_x$ pour n fixé. Il s'agit de transformer les ${}^h\overline{L}^i$ en $L^{i,n}$ et on utilise pour cela le raisonnement qui a permis de montrer (2.76). $\square$

Nous ne développons pas davantage ce point puisqu'il suffit de s'intéresser à $\mathcal{L}_{0,T}(\varphi)$ pour φ dans $C^{f,1}[0, T]$, comme nous le montrons dans le paragraphe suivant.

2.5 Nouvel énoncé

Le théorème 2.4.1 permet de donner une autre formulation du résultat de majoration (par rapport au théorème 2.2.10), celle que nous avons annoncée au début de ce chapitre. Nous reprenons la démonstration du lemme 2.3.6, en montrant que l'on peut remplacer $\mathcal{L}(\overline{G})$ par $\Lambda(\overline{G} \cap C^{f,1}[0, T])$.

Puisque $C^{f,1}[0, T]$ est dense dans $C[0, T]$, on peut recouvrir G par les $B(\varphi, \frac{\varepsilon}{2})$ où φ varie seulement dans $G \cap C^{f,1}[0, T]$ au lieu de $G \cap C[0, T]$. On peut donc remplacer $\mathcal{L}(\overline{G})$ par $\mathcal{L}_{0,T}(\overline{G} \cap C^{f,1}[0, T])$ (le raisonnement est inchangé) qui est bien égal à $\Lambda(\overline{G} \cap C^{f,1}[0, T])$ puisque $\mathcal{L}_{0,T}$ et $\Lambda_{0,T}$ coïncident sur $C^{f,1}[0, T]$ d'après le théorème 2.4.1.

Remarquons d'autre part, qu'on a nécessairement $\Lambda(\overline{G} \cap C^{f,1}[0, T]) \geq \mathcal{L}(\overline{G})$. L'énoncé

suivant améliore donc le théorème 2.2.10 (il est possible que pour tout fermé G on ait $\Lambda(\overline{G} \cap C^{f,1}[0, T]) = \mathcal{L}(\overline{G})$ auquel cas les deux énoncés sont équivalents).

Enfin, comme $\Lambda(\overline{G} \cap C^{f,1}[0, T])$ est supérieur à $\Lambda(\overline{G})$, on obtient également la majoration avec le coût Λ . On peut avoir $\Lambda(\overline{G} \cap C^{f,1}[0, T]) > \Lambda(\overline{G})$, par exemple pour $G = \{\varphi\}$ avec φ n'appartenant pas à $C^{f,1}[0, T]$ et $\Lambda_{0,T}(\varphi)$ fini.

Théorème 2.5.1

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses $(Hmaj)$ et (H_0) . On appelle φ^τ le chemin aléatoire continu défini sur $[0, T]$ par: $\varphi^\tau(k\tau) = Y_k^\tau$ pour $0 \leq k\tau < T + \tau$ et φ^τ est affine sur $[k\tau; (k+1)\tau]$. Alors, pour toute partie borélienne G de $C[0, T]$, tout compact $\mathcal{K}$ de $\mathbb{R}^d$ et tout $\varepsilon > 0$, il existe τ_0 tel que, $\tau \leq \tau_0$ implique, P -ps sur $\{Y_0^\tau \in \mathcal{K}\}$:

$$\tau \log P(\varphi_{/[0,T]}^\tau \in G | \mathcal{F}_0) \leq -\Lambda(\overline{G} \cap C^{f,1}[0, T]) + \varepsilon \leq -\Lambda(\overline{G}) + \varepsilon$$

Application: d'après (2.68) et $\Lambda_{0,T}^{0,0} = \Lambda_{0,T}$ (Proposition 2.2.6), on a:

$$\Lambda_{0,T}(\varphi) \leq \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in C^{f,1}[0,T]}} \Lambda_{0,T}(\psi)$$

on en déduit que pour tout φ de $C^{f,1}[0, T]$, $\Lambda(B(\varphi, \varepsilon) \cap C^{f,1}[0, T])$ tend vers $\Lambda_{0,T}(\varphi)$ quand ε tend vers 0, on a donc:

$$\lim_{\varepsilon \rightarrow 0} \liminf_{\tau \rightarrow 0} \tau \log P(\varphi_{/[0,T]}^\tau \in B(\varphi, \varepsilon) | \mathcal{F}_0) \leq -\Lambda_{0,T}(\varphi).$$

Le théorème précédent doit être précisé en tenant compte du fait que, si G est un ensemble de chemins à valeurs dans P^i pour $i = 1$ ou 2 , il n'en est pas nécessairement de même pour $\overline{G}$. Dans ce cas, tandis que $\Lambda(G \cap C^{f,1}[0, T])$ et $\Lambda^i(G)$ coïncident, ce n'est pas toujours le cas de $\Lambda(\overline{G} \cap C^{f,1}[0, T])$ et $\Lambda^i(\overline{G})$. Or, le résultat de grandes déviations doit faire intervenir dans ce cas la fonctionnelle Λ^i et non $\Lambda(\cdot \cap C^{f,1}[0, T])$, qui risque de prendre en compte un coût mixte le long des chemins de $\overline{G}$ qui suivent P^0 sur certains intervalles. $\Lambda(\overline{G} \cap C^{f,1}[0, T])$ serait alors plus petit que $\Lambda^i(\overline{G})$.

Pour tenir compte de cette éventualité, il convient de remplacer $\Lambda(\overline{G} \cap C^{f,1}[0, T])$ par:

$$\inf\{J_{0,T}(\varphi) \mid \varphi \in \overline{G} \cap C^{f,1}[0, T]\}$$

où $J_{0,T}(\varphi)$ est égal à $\Lambda_{0,T}(\varphi)$ sauf si il existe $[t_1, t_2]$ tel que $\varphi([t_1, t_2])$ soit inclus dans P^0 et tel que pour toute suite φ_n dans G convergeant vers φ , on ait, pour chaque n , $\varphi_n([t_1, t_2])$ inclus dans $P^1 \cup P^2$. Alors, sur $[t_1, t_2]$, on remplace $\Lambda_{t_1, t_2}(\varphi)$ par $\Lambda_{t_1, t_2}^1(\varphi) \wedge \Lambda_{t_1, t_2}^2(\varphi)$, et si les $\varphi_n([t_1, t_2])$ sont tous inclus dans P^i , on remplace $\Lambda_{t_1, t_2}(\varphi)$ par $\Lambda_{t_1, t_2}^i(\varphi)$.

Exemple: supposons $0 < t_1 < t_2 < T$ et appelons G l'ouvert des chemins de $C[0, T]$ tels que $\varphi([t_1, t_2])$ est inclus dans P^i . Alors, on remplace le résultat du théorème 2.5.1 par :

$$\begin{aligned} \tau \log P(\varphi_{/[0, T]}^\tau \in G | \mathcal{F}_0) \leq & -\inf\{\Lambda_{0, t_1}(\varphi) \mid \varphi \in \overline{G} \cap C^{f,1}[0, T]\} \\ & -\inf\{\Lambda_{t_1, t_2}^i(\varphi) \mid \varphi \in \overline{G}\} \\ & -\inf\{\Lambda_{t_2, T}(\varphi) \mid \varphi \in \overline{G} \cap C^{f,1}[0, T]\} + \varepsilon \end{aligned}$$

où le majorant peut être strictement inférieur à $-\Lambda(\overline{G} \cap C^{f,1}[0, T]) + \varepsilon$.

2.6 Cas où pour tout x , S_x^i est inclus dans l'ensemble $\{\beta \mid (-1)^i \pi(\beta) > 0\}$

Jusqu'à présent dans ce chapitre, nous avons supposé que l'hypothèse $H_0(x)$ était vérifiée pour tout x de $\mathbb{R}^d$. Nous étudions maintenant le cas où, pour au moins un des deux indices $i = 1$ ou 2 , on a : " pour tout x de $\mathbb{R}^d$, S_x^i est inclus dans l'ensemble des vecteurs strictement rentrants dans P^i ". Nous montrons qu'alors, il n'y a pas de mélange des champs c'est-à-dire pas d'oscillations du processus entre les deux demi-espaces (dès que le processus entre dans P^i , il y reste). Seuls les coûts L_x^1 et L_x^2 interviennent et non le coût mixte L_x .

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (Hmaj). Supposons que pour tout x de $\mathbb{R}^d$, S_x^1 soit inclus dans l'ensemble $\{\beta \mid \pi(\beta) < 0\}$. Alors, pour tout chemin φ absolument continu, $\Lambda_{0,T}(\varphi)$ est égal à

$$\int_0^T \left(L_{\varphi(t)}^1(\varphi'(t)) \mathbb{I}_{\{\varphi(t) \in P^1\}} + L_{\varphi(t)}^2(\varphi'(t)) \mathbb{I}_{\{\varphi(t) \in P^2 \cup P^0\}} \right) dt$$

Si le processus a son origine Y_0^τ dans $P^1 \cup P^0$, presque sûrement, sa trajectoire reste dans $P^1 \cup P^0$ et le champ μ^1 gouverne tous les pas. Alors le principe de grandes déviations (majoration et minoration) est vérifié avec la fonctionnelle de coût Λ^1 (les chemins de coût fini ayant leur origine dans $P^1 \cup P^0$ sont à valeurs dans $P^1 \cup P^0$) quels que soient les supports S_x^2 .

On suppose maintenant que le processus a son origine Y_0^τ dans P^2 . Presque sûrement, le processus traverse au plus une fois P^0 .

Notation 2.6.1 On note $C^{T_{rav}}$ l'ensemble des chemins φ qui traversent au plus une fois P^0 , c'est-à-dire tels que $\{t \in [0, T] \mid \varphi(t) \in P^0\}$ contient au plus un élément.

Remarquons que pour les chemins de cet ensemble on n'a pas besoin de définir de fonction de coût L_x pour x dans P^0 .

Théorème 2.6.2

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (H_{maj}) et tels que, pour tout x de $\mathbb{R}^d$, S_x^1 soit inclus dans l'ensemble $\{\beta \mid \pi(\beta) < 0\}$. Alors, pour toute partie borélienne G de $C[0, T]$, tout compact $\mathcal{K}$ et tout $\varepsilon > 0$, il existe τ_0 tel que $\tau \leq \tau_0$ implique les majorations suivantes :

$$P\text{-ps sur } \{Y_0^\tau \in (P^1 \cup P^0) \cap \mathcal{K}\}, \quad \tau \log P(\varphi_{/[0, T]}^\tau \in G \mid \mathcal{F}_0) \leq -\Lambda^1(\overline{G}) + \varepsilon$$

$$P\text{-ps sur } \{Y_0^\tau \in P^2 \cap \mathcal{K}\},$$

$$\tau \log P(\varphi_{/[0, T]}^\tau \in G \mid \mathcal{F}_0) \leq -\Lambda(\overline{G} \cap C^{T_{rav}}) + \varepsilon \leq -\Lambda(\overline{G}) + \varepsilon$$

Démonstration : on a vu que si le processus a son origine Y_0^τ dans $P^1 \cup P^0$, le principe de grandes déviations est vérifié avec la fonctionnelle de coût Λ^1 . Soit G une partie borélienne de $C[0, T]$ et $\mathcal{K}$ un compact. Il suffit donc de montrer que P -ps sur $\{Y_0^\tau \in P^2 \cap \mathcal{K}\}$, on a, pour $\tau \leq \tau_0$ convenable :

$$\tau \log P(\varphi_{/[0, T]}^\tau \in G \mid \mathcal{F}_0) \leq -\Lambda(\overline{G} \cap C^{T_{rav}}) + \varepsilon$$

Considérons φ dans le compact $\overline{G \cap C^{T_{rav}}}$, en supposant d'abord que φ appartient à $C^{T_{rav}}$ et que $\Lambda_{0, T}(\varphi)$ est fini. Alors, pour tout $\varepsilon > 0$, il existe α et deux instants t et t' tels que le tube d'axe $\varphi_{/[0, t]}$ et de rayon α soit à valeurs dans P^2 et le tube d'axe $\varphi_{/[t', T]}$ et rayon α à valeurs dans P^i , pour $i = 1$ ou 2 , avec, de plus, $\Lambda_{t, t'}(\varphi)$

inférieur à ε . En utilisant les résultats concernant un seul champ (théorème 2.3.7), on sait qu'il existe τ_φ tel que, pour $\tau \leq \tau_\varphi$, on ait P-ps sur $\{Y_0^\tau \in P^2 \cap \mathcal{K}\}$:

$$\begin{aligned} \tau \log P(\varphi_{/[0,T]}^\tau \in B(\varphi, \alpha) | \mathcal{F}_0) &\leq \tau \log P(\varphi_{/[0,t]}^\tau \in B(\varphi_{/[0,t]}, \alpha) | \mathcal{F}_0) \\ &\quad + \tau \log P(\varphi_{/[t',T]}^\tau \in B(\varphi_{/[t',T]}, \alpha) | Y_{[t']_T}^\tau \in B(\varphi(t'), \alpha)) \\ &\leq -\Lambda_{0,t}^2(\varphi) - \Lambda_{t',T}^1(\varphi) + \varepsilon \leq -\Lambda_{0,T}(\varphi) + 2\varepsilon \end{aligned}$$

($[t']_\tau$ désigne le plus grand multiple de τ inférieur à t').

Si $\Lambda_{0,T}(\varphi)$ est infini, on peut trouver α tel qu'on ait, P-ps sur $\{Y_0^\tau \in P^2 \cap \mathcal{K}\}$:

$$\liminf_{\tau \rightarrow 0} \tau \log P(\varphi_{/[0,T]}^\tau \in B(\varphi, \alpha)) \leq -\infty$$

Considérons maintenant φ dans $\overline{G \cap C^{Trau}}$ n'appartenant pas à C^{Trau} et (φ_n) une suite d'éléments de $G \cap C^{Trau}$ convergeant vers φ . Soit $[t_1, t_2]$ l'ensemble des t tels que $\varphi(t)$ appartient à P^0 . Pour α_φ assez petit, l'événement $\{\varphi_{/[0,T]}^\tau \in B(\varphi, \alpha_\varphi)\}$ implique que le processus reste dans P^2 jusqu'à un instant t_3 arbitrairement proche de t_2 (avec $t_3 \leq t_2$). En effet, dès qu'il entre dans $P^0 \cup P^1$, le processus quitte presque sûrement le voisinage de la frontière. Pour n suffisamment grand, $B(\varphi_n, \frac{\alpha_\varphi}{2})$ est inclus dans $B(\varphi, \alpha_\varphi)$ et $\varphi_n([0, t_3])$ dans P^2 , ce qui permet de majorer P-ps sur $\{Y_0^\tau \in P^2 \cap \mathcal{K}\}$ $\tau \log P(\varphi_{/[0,T]}^\tau \in B(\varphi, \alpha_\varphi) | \mathcal{F}_0)$ par $-\Lambda_{0,T}(\varphi_n)$, à ε près. On obtient, pour $\tau \leq \tau_\varphi$, P-ps sur $\{Y_0^\tau \in P^2 \cap \mathcal{K}\}$:

$$\tau \log P(\varphi_{/[0,T]}^\tau \in B(\varphi, \alpha_\varphi) | \mathcal{F}_0) \leq - \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in G \cap C^{Trau}}} \Lambda_{0,T}(\psi) + \varepsilon$$

On en déduit (en utilisant $[\overline{G \cap C^{Trau}} \cap C^{Trau}] \cup [G \cap C^{Trau}] \subset \overline{G \cap C^{Trau}}$) que, pour tous les chemins φ de $\overline{G \cap C^{Trau}}$, pour α_φ convenable et $\tau \leq \tau_\varphi$, on a P-ps sur $\{Y_0^\tau \in P^2 \cap \mathcal{K}\}$:

$$\tau \log P(\varphi_{/[0,T]}^\tau \in B(\varphi, \alpha_\varphi) | \mathcal{F}_0) \leq - \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in \overline{G \cap C^{Trau}}}} \Lambda_{0,T}(\psi) + \varepsilon \leq -\Lambda(\overline{G \cap C^{Trau}}) + \varepsilon$$

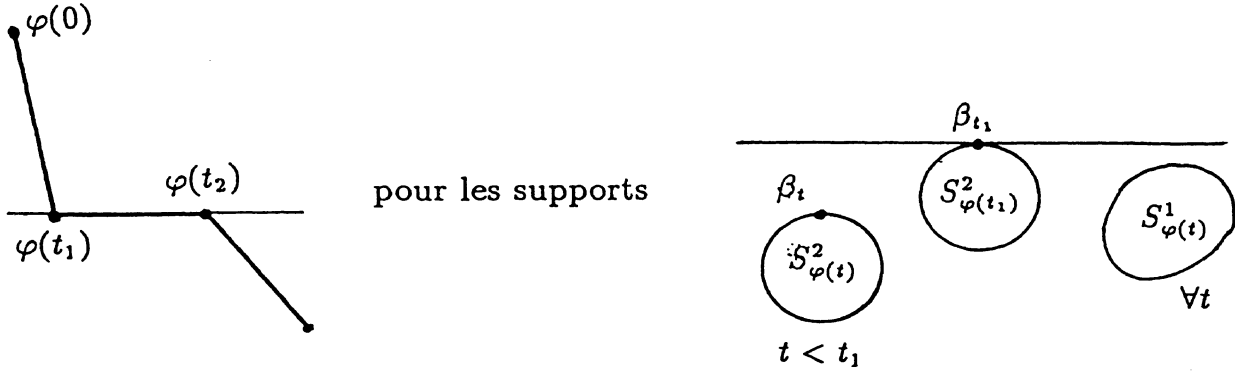
L'ensemble $\overline{G \cap C^{Trau}} \cap \{\varphi \in C[0, T] | \varphi(0) \in \mathcal{K} \text{ et, ps en } t, |\varphi'(t)| \leq M\}$ est un compact. On le recouvre par une union de boules ouvertes de centre φ_i dans $\overline{G \cap C^{Trau}}$ et de rayon $\alpha_i = \alpha_\varphi$ choisi comme ci-dessus. On en déduit, puisque φ^τ appartient à l'ensemble C^{Trau} (parce qu'on a choisi le champ (μ_x^1) sur P^0), que pour τ assez petit, on a P-ps sur $\{Y_0^\tau \in P^2 \cap \mathcal{K}\}$:

$$\tau \log P(\varphi_{/[0,T]}^\tau \in G | \mathcal{F}_0) \leq -\Lambda(\overline{G \cap C^{Trau}}) + \varepsilon$$

d'où le théorème. $\square$

La quantité $-\Lambda(\overline{G} \cap C^{Trau})$ est inférieure à $-\Lambda(\overline{G})$, mais ce dernier terme ne serait pas un majorant optimal comme nous le constatons dans l'exemple suivant :

Considérons un chemin φ tel qu'on ait :

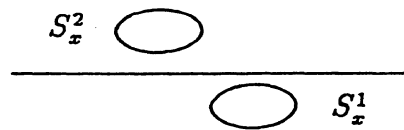


(2.85)

pour tout t , $\varphi'(t) = \beta_t$ avec $\pi(\beta_t)$ maximal parmi les $\pi(\beta)$ pour β dans $S^2_{\varphi(t)}$ et $L^2_{\varphi(t)}(\beta_t) < +\infty$. Un processus issu de $\varphi(0)$ atteint presque sûrement $P^1 \cup P^0$ à une date proche de t_1 (et supérieure à t_1) et quitte alors le voisinage de P^0 . Au voisinage de φ la bonne fonctionnelle de coût est $\inf\{\Lambda_{0,T}(\psi) \mid \psi \text{ a une seule intersection avec } P^0\}$. Elle vaut $+\infty$ sur le voisinage de φ (c'est-à-dire que $\Lambda(B(\varphi, \varepsilon) \cap C^{Trau})$ vaut $+\infty$ pour ε assez petit), alors que $\Lambda_{0,T}(\varphi)$ est fini et n'intervient pas dans les résultats.

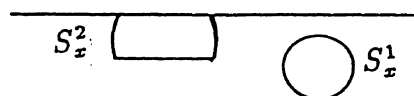
Indiquons maintenant des cas où, au contraire, $\Lambda(G \cap C^{Trau})$ et $\Lambda(G)$ coïncident pour toute partie borélienne G de chemins d'origine dans P^2 :

1. pour tout x de $\mathbb{R}^d$, S^2_x est inclus dans l'ensemble $\{\beta \mid \pi(\beta) \geq 0\}$:



presque sûrement le processus reste dans P^2 . On a alors un résultat de grandes déviations avec la fonctionnelle de coût Λ^2 (égale à la fonctionnelle Λ sur les chemins d'origine appartenant à P^2).

2. pour tout x de $\mathbb{R}^d$, S^2_x est inclus dans $\{\beta \mid \pi(\beta) \leq 0\}$ avec $\mu_x^2(\{\pi(\cdot) = 0\}) = 0$: un chemin de coût fini ne peut suivre P^0 et appartient à C^{Trau} , donc $\Lambda(G \cap C^{Trau}) = \Lambda(G)$, pour toute partie G .



Remarquons que si on considère la situation symétrique par rapport à la précédente (échange des indices $i = 1$ et $i = 2$: on suppose que pour tout x de $\mathbb{R}^d$, S_x^2 est inclus dans l'ensemble $\{\beta | \pi(\beta) > 0\}$), les résultats ne sont pas tout à fait les symétriques des précédents. En effet, φ^τ n'appartient pas nécessairement à l'ensemble $C^{T_{rav}}$: si pour tout x , $\mu_x^1(\{\pi(\cdot) = 0\})$ est strictement positif, certains processus peuvent suivre la frontière P^0 ; on se contente alors d'une majoration avec la fonctionnelle $\Lambda_{0,T}(\cdot \cap C^f[0,T])$ sur les boréliens pour Y_0^τ dans $P^1 \cup P^0$. Pour Y_0^τ dans P^2 , la majoration est obtenue avec $\Lambda_{0,T}^2$.

Retenons que quand l'hypothèse (H_0) n'est plus vérifiée, le comportement du processus dépend fortement de la condition initiale et du choix du champ sur P^0 et que la symétrie en les indices $i = 1$ et $i = 2$ est perdue (cf Remarque 1.4.2).

Dans le paragraphe 4.1.3, nous étudions les cas limites des points x pour lesquels S_x^1 (respectivement S_x^2) est inclus dans l'ensemble $\{\pi(\cdot) \geq 0\}$ (respectivement $\{\pi(\cdot) \leq 0\}$) mais tangent à $\{\pi(\cdot) = 0\}$.

Chapitre 3

Grandes déviations trajectorielles : inégalité de minoration

Dans ce chapitre, nous établissons, pour un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (Hmin) (cf fin du chapitre 1), l'inégalité de minoration du principe de grandes déviations (1.3). Précisément, nous montrons le théorème suivant : notons $C^{f,r}[0, T]$ l'ensemble des chemins φ de $C[0, T]$ tels que

- (1) $\Lambda_{0,T}^{-\theta,0}(\varphi)$ est fini pour un $\theta > 0$
- (2) il existe une partition finie $0 = t_0 \leq t_1 \leq \dots \leq t_p = T$ telle que, pour tout j , $\varphi([t_j, t_{j+1}[)$ soit inclus dans P^{i_j} , avec $i_j = 0, 1$ ou 2 . De plus, pour $i_j \neq 0$ et $\varphi(t_j)$ sur P^0 , $S_{\varphi(t_j)}^1$ et $S_{\varphi(t_j)}^2$ contiennent des vecteurs strictement rentrants dans P^{i_j} .

Théorème 3.0.1

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (Hmin) et (HLip). On appelle φ^τ le chemin aléatoire continu défini par :

$\varphi^\tau(k\tau) = Y_k^\tau$ pour $0 \leq k\tau < T + \tau$ et φ^τ est affine sur $[k\tau; (k+1)\tau]$.

Alors, pour tout y_0 de $\mathbb{R}^d$ et pour toute partie borélienne G de $C_{y_0}[0, T]$ (ensemble des chemins φ continus sur $[0, T]$ tels que $\varphi(0) = y_0$), on a :

$$-\Lambda(\overset{\circ}{G} \cap C^{f,r}[0, T]) \leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0,T]}^\tau \in G)$$

Dans le paragraphe 3.4.2 on étudie l'existence de θ réalisant la condition (1) dans la définition de $C^{f,r}[0, T]$. La condition (2) est peu contraignante (cf remarques après

le théorème 3.3.3) pour un chemin qui est déjà dans $C^f[0, T]$ et a un coût fini.

Nous complétons ce résultat par un second qui permet de prendre en compte la configuration des supports considérée dans le paragraphe 2.6, pour laquelle le mino- rant dans le premier résultat vaut $-\infty$. On note $C^{T_{rav+}}$ l'ensemble des chemins φ qui traversent au plus une fois P^0 et tels que si $\{t \in [0, T] \mid \varphi(t) \in P^0\}$ est égal à $\{t\}$, alors pour i tel que $\varphi(T)$ appartient à P^i , les ensembles $S_{\varphi(t)}^1$ et $S_{\varphi(t)}^2$ contiennent des vecteurs strictement rentrants dans P^i .

Théorème 3.0.2

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses $(Hmin)$ et $(HLip)$ et tels que, pour tout x de $\mathbb{R}^d$, S_x^1 soit inclus dans l'ensemble $\{\beta \mid \pi(\beta) < 0\}$. Alors, pour tout x de P^2 et pour toute partie borélienne G de $C_{y_0}[0, T]$, on a :

$$-\Lambda(\overset{\circ}{G} \cap C^{T_{rav+}}) \leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0, T]}^\tau \in G)$$

Remarquons que dans ce second résultat, on ne calcule des fonctionnelles de coût que le long de chemins qui ne suivent jamais la frontière. C'était déjà le cas pour la majoration. Il n'y a donc pas lieu, sous cette hypothèse sur les supports, de définir une fonction de coût L_x avec x dans P^0 .

La minoration est, en général, la partie la plus délicate à établir dans un résultat de grandes déviations. Une difficulté notable ici est d'obtenir le coût L_x le long de la frontière, (1.12), qui fait intervenir les contraintes $\pi(\beta^1) \geq 0$ et $\pi(\beta^2) \leq 0$, et non simplement $conv(L_x^1, L_x^2)$ (1.13). Cette difficulté a été résolue par P. Dupuis et R.S. Ellis dans [DE], pour l'étude des vitesses moyennes, mais seulement dans le cas particulier de champs de probabilités constants et avec une origine sur la frontière. Dans le paragraphe 3.1.2, nous étendons leur démonstration au cas où les champs de probabilités sont variables et où Y_0^τ n'est pas nécessairement dans P^0 , ce qui nous conduit au résultat concernant les vitesses moyennes (théorème 3.1.8 page 139).

Nous en déduisons une minoration de la probabilité que φ^τ appartienne à un tube d'axe un chemin suffisamment régulier ; on approxime ensuite "la plupart" des chemins par de tels axes, ce qui conduit à la minoration pour des tubes d'axes plus généraux. On passe alors à des parties boréliennes quelconques.

Nous traitons à part le cas où les champs de probabilités sont constants. L'utilisation de la fonctionnelle de coût $\Lambda(\mathring{G} \cap C^{f,r}[0, T])$ au lieu de $\Lambda(\mathring{G} \cap C^f[0, T])$ dans l'inégalité de minoration des grandes déviations résulte en effet d'une difficulté qui disparaît dans ce cas.

Enfin dans une dernière partie, nous abordons, comme pour la majoration, le cas où, pour au moins un des deux indices i , comme dans le paragraphe 2.6, S_x^i est inclus dans l'ensemble des vecteurs strictement rentrants dans P^i , pour tout x de $\mathbb{R}^d$ et montrons le deuxième théorème énoncé ci-dessus.

Dans tout le chapitre, on se donne un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (Hmin). On suppose que ce processus a sa valeur initiale dans un compact $\mathcal{K}$ et est à valeurs, pour $n\tau \leq T$, dans un compact $\mathcal{U}$, qui contient également tous les chemins issus de $\mathcal{K}$ et de coût $\Lambda_{0,T}(\varphi)$ fini. Les vecteurs des $S_x^i, x \in \mathcal{U}$ sont alors de norme inférieure à un réel M (cf (1.16)).

3.1 Résultats sur les vitesses moyennes

3.1.1 Introduction

Dans cette partie, on minore, pour une valeur initiale Y_0^τ proche de P^0 , la probabilité que la moyenne $\frac{Y_n^\tau - Y_0^\tau}{n\tau}$ soit proche d'un vecteur donné, parallèle à P^0 (théorème 3.1.8) et la probabilité que Y_n^τ soit dans un demi-espace P^i fixé (paragraphe 3.1.5). Au passage, on retrouve (lemme 3.1.7) le cas où tous les pas sont gouvernés par le même champ de probabilités, qui sera utilisé pour Y_0^τ éloigné de P^0 .

Pour le premier résultat, comme pour la majoration, on décompose la moyenne $\frac{Y_n^\tau - Y_0^\tau}{n\tau}$ en plusieurs termes : les pas effectués avant de franchir P^0 , puis ceux effectués de part et d'autre de P^0 .

On appelle σ'' l'instant de première traversée de P^0 et on commence par établir un lemme (lemme 3.1.2) qui s'applique P-ps sur $\{|\pi(Y_0^\tau)| \leq \tau M\}$ et qui sera utilisé pour étudier $(Y_k^\tau)_{k \geq \sigma''}$; τM est la longueur maximale d'un pas fait dans $\mathcal{U}$; on s'intéresse donc au processus à partir d'un instant où il est susceptible de traverser P^0 . Cette contrainte $|\pi(Y_0^\tau)| \leq \tau M$ permet d'obtenir le lemme de stabilité 3.1.4 sur la chaîne $(\frac{1}{\tau}\pi(Y_k^\tau))_{k \leq n}$.

On en déduit le théorème 3.1.8 qui s'applique P-ps sur $\{|\pi(Y_0^\tau)| \leq n\tau\alpha\}$, en utilisant le lemme 3.1.2 pour traiter $(Y_k^\tau)_{k \geq \sigma''}$ et le lemme 3.1.7 pour les σ'' premiers pas de (Y_k^τ) , qui obéissent à un même champ de probabilités.

On prend α petit ; alors, P-ps sur $\{|\pi(Y_0^\tau)| \leq n\tau\alpha\}$, σ'' est petit avec une probabilité en e^{-nI} , où I est petit ; le résultat obtenu convient pour obtenir la minoration concernant des tubes d'axe régulier dans le paragraphe 3.2.

Dans les lemmes 3.1.2, 3.1.7 et 3.1.9, et dans le théorème 3.1.8, on se ramène à des probabilités presque constantes en supposant Y_0^τ dans $B(x_0, \varepsilon_1)$ avec ε_1 petit et en prenant $n\tau$ suffisamment petit pour que les probabilités gouvernant n pas successifs varient très peu.

Notation 3.1.1 Pour x_0 dans P^0 et $\delta_0 > 0$, on note $\mathcal{E}(x_0, \delta_0)$ l'ensemble des $(\rho^1, \rho^2, \beta^1, \beta^2)$ tels que

$$(-1)^i \pi(\beta^i) < 0, \beta^i \in (S_{x_0}^i)^{-\delta_0}, \rho^i \geq 0, \sum_{i=1}^2 \rho^i = 1 \quad \text{et} \quad \pi\left(\sum_{i=1}^2 \rho^i \beta^i\right) = 0.$$

Remarquons que l'hypothèse $H_0(x_0)$ est vérifiée si et seulement si il existe $\delta_0 > 0$ tel que $\mathcal{E}(x_0, \delta_0)$ soit non vide, et que dans ce cas, $\overline{\mathcal{E}(x_0, \delta_0)}$, l'adhérence de $\mathcal{E}(x_0, \delta_0)$, est l'ensemble des $(\rho^1, \rho^2, \beta^1, \beta^2)$ vérifiant les conditions de la notation 3.1.1, sauf la première, changée en l'inégalité large $(-1)^i \pi(\beta^i) \leq 0$.

3.1.2 Minoration de la probabilité de suivre un vecteur parallèle à P^0 pour des points de départ Y_0^τ très proches de P^0

On remarquera que pour que $\frac{Y_n^\tau - Y_0^\tau}{n\tau}$ soit proche de $\sum_{i=1}^2 \rho^i \beta^i$ parallèle à P^0 , quand $\pi(\beta^1) > 0$ et $\pi(\beta^2) < 0$, avec, pour chaque i , la moyenne des pas de type i égale à β^i , le nombre de passages de P^0 par le processus doit tendre vers $+\infty$ quand n tend vers $+\infty$. En effet, si N_i est le nombre de pas de type i qui traversent P^0 , pour $\pi(Y_0^\tau) = 0$ et par exemple $\pi(Y_n^\tau) \geq 0$, $N_1\tau M$ doit être supérieur à $n\tau\pi(\beta^1)$.

Lemme 3.1.2

Soit $\varepsilon, \delta_0 > 0$. Il existe ε_1, K tels que, pour tout $\varepsilon_2 > 0$, pour tout x_0 dans $P^0 \cap \mathcal{K}$ tel que $\mathcal{E}(x_0, \delta_0)$ soit non vide et tout $(\rho^1, \rho^2, \beta^1, \beta^2)$ dans $\overline{\mathcal{E}(x_0, \delta_0)}$, on peut trouver N tel que $n \geq N$ implique, quels que soient τ tel que $n\tau \leq K$ et y_0 dans $B(x_0, \varepsilon_1) \cap \{|\pi(\cdot)| \leq \tau M\}$:

$$\frac{1}{n} \log P_{y_0} \left(\frac{Y_n^\tau - Y_0^\tau}{n\tau} \in B \left(\sum_{i=1}^2 \rho^i \beta^i, \varepsilon_2 \right) \right) \geq - \sum_{i=1}^2 \rho^i L_{x_0}^i(\beta^i) - \varepsilon.$$

Démonstration : nous montrons tout d'abord comment on passe du résultat avec $\mathcal{E}(x_0, \delta_0)$ au résultat avec $\overline{\mathcal{E}(x_0, \delta_0)}$. Après avoir rappelé les résultats obtenus par [DE] pour des champs constants, nous exposons ensuite les grandes lignes de leur extension au cas présent : champs non constants, Y_0^τ pas nécessairement sur P^0 .

1. Montrons comment on passe du résultat avec $\mathcal{E}(x_0, \delta_0)$ au résultat avec $\overline{\mathcal{E}(x_0, \delta_0)}$: il suffit de montrer que, pour tout $\varepsilon > 0$ et tout $(\rho^1, \rho^2, \beta^1, \beta^2)$ dans $\overline{\mathcal{E}(x_0, \delta_0)}$, il existe $(\tilde{\rho}^1, \tilde{\rho}^2, \tilde{\beta}^1, \tilde{\beta}^2)$ dans $\mathcal{E}(x_0, \delta_0)$ tel que :

$$\left| \sum_{i=1}^2 \tilde{\rho}^i \tilde{\beta}^i - \sum_{i=1}^2 \rho^i \beta^i \right| \leq \varepsilon \quad \text{et} \quad \sum_{i=1}^2 \tilde{\rho}^i L_{x_0}^i(\tilde{\beta}^i) \leq \sum_{i=1}^2 \rho^i L_{x_0}^i(\beta^i) + \varepsilon.$$

Pour chaque i , on considère b^i appartenant à $(S_{x_0}^i)^{-\delta_0}$ tel que $\pi(b^1) = -\pi(b^2) > 0$ (possible puisque $\mathcal{E}(x_0, \delta_0)$ est non vide). On distingue alors trois cas :

1. si, pour chaque i , $(-1)^i \pi(\beta^i) < 0$, alors $(\rho^1, \rho^2, \beta^1, \beta^2)$ est déjà dans $\mathcal{E}(x_0, \delta_0)$.
2. si $\pi(\beta^1) = \pi(\beta^2) = 0$ et $\rho^1 = 0$ ou 1 :

si par exemple ρ^1 vaut 0, on pose $\tilde{\beta}^1 = b^1$ et $\tilde{\beta}^2 = \theta b^2 + (1 - \theta)\beta^2$ et on prend $\tilde{\rho}^1$ et $\tilde{\rho}^2$ tels que $\tilde{\rho}^1 + \tilde{\rho}^2 = 1$ et $\pi(\sum_{i=1}^2 \tilde{\rho}^i \tilde{\beta}^i) = 0$. Quand on fait tendre θ vers 0, $\sum_{i=1}^2 \tilde{\rho}^i \tilde{\beta}^i$ et $\sum_{i=1}^2 \tilde{\rho}^i L_{x_0}^i(\tilde{\beta}^i)$ sont arbitrairement proches respectivement de $\sum_{i=1}^2 \rho^i \beta^i$ et $\sum_{i=1}^2 \rho^i L_{x_0}^i(\beta^i)$, en particulier parce que, pour chaque i , β^i appartient à $(S_{x_0}^i)^{-\delta_0}$.

3. si $\pi(\beta^1) = \pi(\beta^2) = 0$ et $\rho^1 \in]0, 1[$:

on pose $\tilde{\beta}^i = \theta \rho^{3-i} b^i + (1 - \theta \rho^{3-i}) \beta^i$; alors, $(-1)^i \pi(\beta^i)$ est strictement négatif et $\pi(\sum_{i=1}^2 \rho^i \tilde{\beta}^i)$ vaut 0. La modification est arbitrairement petite et on conclut comme dans le cas précédent. $\square$

2. Rappels des résultats de [DE] : pour des champs de probabilités constants, minoration de la probabilité que la moyenne de n pas consécutifs soit proche d'un vecteur donné parallèle à P^0

Rappelons l'étape essentielle conduisant au résultat de minoration établi par Dupuis et Ellis dans [DE] et redonné dans le paragraphe 1.4.2 :

les deux champs de probabilités (μ_x^1) et (μ_x^2) sont supposés constants (égaux respectivement à μ^1 et μ^2). Soit x un point de P^0 ; on considère la chaîne (S_k) définie de la façon suivante :

$$\begin{cases} S_0 = x \\ S_{k+1} = S_k + \mathbb{I}_{\{S_k \in P^1 \cup P^0\}} X_k^1 + \mathbb{I}_{\{S_k \in P^2\}} X_k^2 \end{cases} \quad (3.1)$$

Alors, pour tous $(\rho^1, \rho^2, \beta^1, \beta^2)$ tels que $0 < \rho^1 < 1$, $\rho^1 + \rho^2 = 1$, $\pi(\beta^1) < 0$, $\pi(\beta^2) > 0$ et $\pi(\rho^1 \beta^1 + \rho^2 \beta^2) = 0$, et pour tout $\varepsilon > 0$, on a :

$$-\rho^1 L^1(\beta^1) - \rho^2 L^2(\beta^2) \leq \liminf_{n \rightarrow +\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \in B(\rho^1 \beta^1 + \rho^2 \beta^2; \varepsilon)\right) \quad (3.2)$$

Ce résultat est établi de la façon suivante :

□ On considère $(\rho^1, \beta^1, \rho^2, \beta^2)$ comme ci-dessus. On pose :

$$\begin{aligned} \text{pour } k \geq 1, A_k^1 &= \{S_k \in P^1 \cup P^0\} \text{ et } A_k^2 = \{S_k \in P^2\}, \\ \text{et, pour } i = 1 \text{ ou } 2, r_n^i &= \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{I}_{A_k^i} \text{ et } w_n^i = \frac{1}{nr_n^i} \sum_{k=0}^{n-1} \mathbb{I}_{A_k^i} X_k^i. \end{aligned}$$

Pour chaque i , on prend α_i tel que $L^i(\beta^i) = \langle \alpha_i, \beta^i \rangle - H^i(\alpha_i)$ (on a $\alpha_i = \nabla L^i(\beta^i)$). Si P_n est la probabilité induite par la distribution marginale des $(X_j^i)_{j \leq n}$ par rapport à P , on fait un changement de probabilité tel que la nouvelle probabilité donne un poids maximal à l'événement "avoir la vitesse β^i dans P^i ", c'est-à-dire qu'on prend :

$$\frac{d\tilde{P}_n}{dP_n}(\omega) = \exp \sum_{j=1}^n \sum_{i=1}^2 \left[\langle \alpha_i, X_j^i(\omega) \rangle - H^i(\alpha_i) \right]$$

Les deux éléments fondamentaux qui permettent alors de conclure sont les convergences en probabilité, sous $\tilde{P}$, extension de Kolmogorov de $(\tilde{P}_n)$, de r_n^i et $r_n^i(w_n^i - \beta^i)$ respectivement vers ρ^i et vers 0, quand n tend vers $+\infty$. □

Nous allons utiliser une démarche analogue dans notre cas, qui est plus complexe techniquement. On notera qu'un élément important de ce raisonnement est le fait que les conditions $(-1)^i \pi(\beta^i) < 0$ impliquent une stabilité de la première coordonnée du processus sous la condition qu'il ait la vitesse β^i quand il est dans P^i (lemme 3.1.4).

3. Grandes lignes de l'extension au cas des champs non constants; démonstration du lemme 3.1.2 pour $(\rho^1, \beta^1, \rho^2, \beta^2)$ variant dans $\mathcal{E}(x_0, \delta_0)$

Revenons au cas général de champs de probabilités non constants. On considère la chaîne $(\frac{1}{\tau} Y_k^\tau)_k$: elle a des accroissements en X_k^i comme la chaîne (S_k) considérée dans (3.1); on va alors lui appliquer une généralisation de (3.2): cette fois, les champs sont non constants et l'origine de la marche aléatoire n'est plus tout à fait sur P^0 mais reste à une distance bornée de P^0 (on cherche un résultat vrai P-ps sur $\{|\frac{1}{\tau} \pi(Y_0^\tau)| \leq M\}$).

Cette généralisation doit être uniforme en τ . En effet, quand on fait tendre n vers $+\infty$ en gardant $n\tau \leq K$, τ tend vers 0. Or, contrairement au cas des champs de probabilités constants, la loi des accroissements varie avec τ . Les résultats doivent donc s'appliquer à toutes les chaînes correspondant aux différentes valeurs du paramètre τ .

Pour conserver, dans le cas de champs de probabilités non constants, les convergences établies dans [DE], on utilise, pour $i = 1$ ou 2 , les propriétés (i) et (ii):

- (i) $(x, \beta) \mapsto \nabla L_x^i(\beta)$ est continue sur $\bigcup x \times \overset{\circ}{S}_x^i$
- (ii) pour tout α de $\mathbb{R}^d$ et tout $1 \leq k \leq d$, $x \mapsto \frac{\partial^2 H_x^i}{\partial^2 \alpha_k}(\alpha)$ est continue

Soit $\delta_0, \varepsilon > 0$. D'après (H3) et (i), et puisque $L^i(\cdot)$ est continue sur $\bigcup x \times \overset{\circ}{S}_x^i$, il existe $\varepsilon_1(\delta_0, \varepsilon)$ tel que pour tout x dans $\mathcal{K}$, on ait, pour $i = 1$ ou 2 :

$$|x - y| \leq 2\varepsilon_1 \Rightarrow \begin{cases} (S_x^i)^{-\delta_0} \subset (S_y^i)^{-\frac{\delta_0}{2}} \\ \forall \beta \in (S_x^i)^{-\delta_0}, \quad \begin{cases} |L_x^i(\beta) - L_y^i(\beta)| < \varepsilon \\ |\nabla L_x^i(\beta) - \nabla L_y^i(\beta)| < \varepsilon \end{cases} \end{cases} \quad (3.3)$$

On considère x_0 dans $P^0 \cap \mathcal{K}$ tel que $\mathcal{E}(x_0, \delta_0)$ soit non vide. On suppose $KM \leq \varepsilon_1$ et $n\tau \leq K$ et on pose $\Omega_0 = \{Y_0^\tau \in B(x_0, \varepsilon_1)\} \cap \{|\pi(Y_0^\tau)| \leq \tau M\}$. Alors on a:

$$P\text{-ps sur } \Omega_0, \text{ pour tout } j \leq n, Y_j^\tau \text{ appartient à } B(x_0, 2\varepsilon_1). \quad (3.4)$$

Dans toute la suite de ce paragraphe, nous reprenons les principes de la démonstration de [DE], en les adaptant à notre cas :

- les champs ne sont plus constants
- le point initial est dans un voisinage de P^0 et non plus sur P^0 .

On considère $(\rho^1, \rho^2, \beta^1, \beta^2)$ dans $\mathcal{E}(x_0, \delta_0)$. On utilise une suite de changements de probabilité, associés à β^1 et β^2 , obtenus de la façon suivante :

pour $i = 1$ ou 2 et x tel que β^i appartienne à $\overset{\circ}{S}_x^i$, on pose $\alpha_x^i = \nabla L_x^i(\beta^i)$. Alors on a $L_x^i(\beta^i) = \langle \alpha_x^i, \beta^i \rangle - H_x^i(\alpha_x^i)$. D'après (3.3), (3.4) et $\beta^i \in (S_{x_0}^i)^{-\delta_0}$, P -ps sur Ω_0 , pour tout $j \leq n-1$, β^i appartient à $(S_{Y_j^\tau}^i)^{-\frac{\delta_0}{2}}$; alors le vecteur $\alpha_{Y_j^\tau}^i$ est bien défini et d'après la propriété (i), il reste dans un compact.

Les variables X_j^i sont définies sur un espace de probabilité $(\Omega, \mathcal{F}, P)$. On appelle P_n^τ la probabilité sur $(\Omega, \mathcal{F})$ induite par la distribution des $\{X_j^i(Y_j^\tau); j = 0, \dots, n-1, i = 1, 2\}$ par rapport à P . Pour chaque (τ, n) , on définit une nouvelle probabilité Q_n^τ sur Ω de dérivée de Radon-Nikodym par rapport à P_n^τ égale à :

$$\frac{dQ_n^\tau}{dP_n^\tau}(\omega) = \exp \sum_{j=0}^{n-1} \sum_{i=1}^2 \left[\langle \alpha_{Y_j^\tau}^i, X_j^i(Y_j^\tau)(\omega) \rangle - H_{Y_j^\tau}^i(\alpha_{Y_j^\tau}^i(\omega)) \right] \quad (3.5)$$

$\left(\exp \sum_{j=0}^{n-1} \sum_{i=1}^2 \left[\langle \alpha_{Y_j^\tau}^i, X_j^i(Y_j^\tau) \rangle - H_{Y_j^\tau}^i(\alpha_{Y_j^\tau}^i) \right] \right)_n$ est une martingale de moyenne 1, donc Q_n^τ est une probabilité. De plus, pour $j \leq n-1$, $Q_n^\tau(X_j^i(Y_j^\tau) \in dx | Y_j^\tau = y)$ est égal à $\exp[\langle \alpha_y^i, x \rangle - H_y^i(\alpha_y^i)] \mu_y^i(dx)$. En particulier, $X_j^i(Y_j^\tau)$ a pour moyenne β^i par rapport à Q_n^τ .

On appelle Q^τ la probabilité sur $(\Omega, \mathcal{F})$ extension de Kolmogorov de la suite $(Q_n^\tau)_{n \in \mathbb{N}}$ (existe, puisque les hypothèses du théorème d'extension sont vérifiées).

On pose, pour $j \geq 0$, $A_j^1 = \{Y_j^\tau \in P^1 \cup P^0\}$ et $A_j^2 = \{Y_j^\tau \in P^2\}$. Remarquons que ces événements dépendent de τ dès que Y_0^τ n'appartient pas à P^0 . Enfin, on pose :

$$\begin{cases} r_{n,\tau}^i &= \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{I}_{A_j^i}, \\ w_{n,\tau}^i &= \frac{1}{nr_{n,\tau}^i} \sum_{j=0}^{n-1} \mathbb{I}_{A_j^i} X_j^i(Y_j^\tau) \end{cases} \quad (3.6)$$

On va montrer qu'on a, comme dans le cas des champs constants, les deux convergences en probabilité, par rapport à Q^τ : P-ps sur Ω_0 , pour $i = 1$ ou 2

$$\begin{cases} r_{n,\tau}^i(w_{n,\tau}^i - \beta^i) & \xrightarrow{n \rightarrow +\infty} 0 \\ r_{n,\tau}^i & \xrightarrow{n \rightarrow +\infty} \rho^i, \end{cases} \quad (3.7)$$

Précisément, pour tout $\varepsilon > 0$ et tout y_0 dans $B(x_0, \varepsilon_1) \cap \{|\pi(\cdot)| \leq \tau M\}$, nous montrons que, pour chaque i , $Q_{y_0}^\tau(|r_{n,\tau}^i(w_{n,\tau}^i - \beta^i)| \geq \varepsilon)$ et $Q_{y_0}^\tau(|r_{n,\tau}^i - \rho^i| \geq \varepsilon)$ tendent vers 0 quand n tend vers $+\infty$, et cela uniformément en τ tel que $n\tau \leq K$ et en la condition initiale y_0 .

Pour établir la première convergence, on utilise, pour $j \leq n$:

$$\mathbb{E}^{Q^\tau} \left(\mathbb{I}_{A_j^i} X_j^i(Y_j^\tau) \mid \mathcal{F}_j \right) = \mathbb{I}_{A_j^i} \beta^i.$$

On appelle π^k l'application qui associe à un vecteur sa k^e coordonnée et on pose $W_j^i = \mathbb{I}_{A_j^i} (X_j^i(Y_j^\tau) - \beta^i)$; c'est une variable aléatoire $\mathcal{F}_{j+1}$ -mesurable. Soit y_0 dans $\mathbb{R}^d$, on a $\mathbb{E}_{y_0}^{Q^\tau}(W_j^i) = 0$ et donc :

$$n^2 \mathbb{E}_{y_0}^{Q^\tau} \left(|r_{n,\tau}^i(w_{n,\tau}^i - \beta^i)|^2 \right) = \sum_{k=1}^d \mathbb{E}_{y_0}^{Q^\tau} \left\{ \sum_{j=0}^{n-1} [\pi^k(W_j^i)]^2 + 2 \sum_{0 \leq j < m \leq n-1} \pi^k(W_j^i) \pi^k(W_m^i) \right\},$$

$$\text{avec, pour } j < m, \quad \mathbb{E}_{y_0}^{Q^\tau} \left(\pi^k(W_j^i) \pi^k(W_m^i) \right) = \mathbb{E}_{y_0}^{Q^\tau} \left(\pi^k(W_j^i) \mathbb{E}^{Q^\tau}(\pi^k(W_m^i) \mid \mathcal{F}_m) \right) = 0.$$

D'autre part, on a :

$$\mathbb{E}_{y_0}^{Q^\tau} \left([\pi^k(W_j^i)]^2 \right) \leq \mathbb{E}_{y_0}^{Q^\tau} \left(\mathbb{E}^{Q^\tau} \left([\pi^k(X_j^i(Y_j^\tau) - \beta^i)]^2 \mid \mathcal{F}_j \right) \right)$$

et, d'après $\beta^i = \nabla H_x^i(\alpha_x^i)$:

$$\begin{aligned} \frac{\partial^2 H_x^i}{\partial \pi^k(\alpha) \partial \pi^{k'}(\alpha)}(\alpha_x^i) &= \int \pi^k(y - \beta^i) \pi^{k'}(y - \beta^i) e^{(\alpha_x^i, y) - H_x^i(\alpha_x^i)} \mu_x^i(dy) \\ &= \mathbb{E}^{Q^\tau} \left(\pi^k(X_j^i(Y_j^\tau) - \beta^i) \pi^{k'}(X_j^i(Y_j^\tau) - \beta^i) \mid Y_j^\tau = x \right) \end{aligned}$$

D'après (3.4), pour y_0 dans $B(x_0, \varepsilon_1)$, on peut donc majorer $\mathbb{E}_{y_0}^{Q^\tau}([\pi^k(W_j^i)]^2)$ par

$$M_k = \sup_{x \in B(x_0, 2\varepsilon_1)} \frac{\partial^2 H_x^i}{\partial^2 \pi^k(\alpha)}(\alpha_x^i)$$

Cette quantité est finie d'après les propriétés (i) et (ii) : $\frac{\partial^2 H_x^i}{\partial^2 \pi^k(\alpha)}(\nabla L_x^i(\beta^i))$ est continue en x sur $B(x_0, 2\varepsilon_1)$, puisque, pour x dans cette boule, β^i appartient à $(S_x^i)^{-\frac{\varepsilon_0}{2}}$

(cf (3.3)). On obtient :

$$n^2 \mathbb{E}_{y_0}^{Q^\tau} \left(|r_{n,\tau}^i(w_{n,\tau}^i - \beta^i)|^2 \right) \leq n \sum_{k=1}^d M_k,$$

et on en déduit d'après l'inégalité de Tchebichev que, pour tout $\varepsilon_3 > 0$, $Q_{y_0}^\tau(|r_{n,\tau}^i(w_{n,\tau}^i - \beta^i)| \geq \varepsilon_3)$ tend vers 0 quand n tend vers $+\infty$, uniformément en τ , pour $n\tau \leq K$, et en la condition initiale, pour y_0 dans $B(x_0, \varepsilon_1)$, ce qui prouve la première convergence.

Pour établir la deuxième convergence, on associe à la chaîne de Markov $(\frac{1}{\tau}Y_k^\tau)_k$ son générateur infinitésimal $\mathcal{L}^\tau$:

pour τ fixé, $(\frac{1}{\tau}Y_k^\tau)_{k \leq n}$ est, sous Q^τ , une chaîne de Markov de probabilité de transition $\tilde{\pi}$ telle que $\tilde{\pi}(x, x + dy) = \sum_{i=1}^2 \mathbb{I}_{P^i}(x) e^{(\alpha_{\tau x}^i, y) - H_{\tau x}^i(\alpha_{\tau x}^i)} \mu_{\tau x}^i(dy)$, où $P^1 = P^1 \cup P^0$ et $P^2 = P^2$; on pose, pour toute fonction continue bornée $f : \mathbb{R}^d \rightarrow \mathbb{R}$:

$$\mathcal{L}^\tau f(x) = \sum_{i=1}^2 \mathbb{I}_{P^i}(x) \int [f(x + y_i) - f(x)] e^{(\alpha_{\tau x}^i, y_i) - H_{\tau x}^i(\alpha_{\tau x}^i)} \mu_{\tau x}^i(dy^i). \quad (3.8)$$

Alors, on a $\mathcal{L}^\tau f\left(\frac{1}{\tau}Y_{j-1}^\tau\right) = \mathbb{E}^{Q^\tau} \left[f\left(\frac{1}{\tau}Y_j^\tau\right) - f\left(\frac{1}{\tau}Y_{j-1}^\tau\right) \mid Y_{j-1}^\tau \right]$.

Remarque : quand les champs sont constants, si $f(x)$ ne dépend que de $\pi(x)$, alors $\mathcal{L}^\tau f(x)$ dépend seulement de $\pi(x)$. Cela n'est plus vrai pour les champs de probabilités non constants considérés ici.

Pour τ donné, on définit une mesure d'occupation en posant :

$$r_n^\tau(dx) = r_n^\tau(\omega, dx) = \frac{1}{n} \sum_{j=1}^n \delta_{\frac{1}{\tau}Y_{j-1}^\tau(\omega)} ;$$

pour ω fixé, r_n^τ est une probabilité sur $\mathbb{R}^d$.

Lemme 3.1.3

Pour toute fonction continue bornée f de $\mathbb{R}^d$ dans $\mathbb{R}$ et pour tout y_0 de $\mathbb{R}^d$, $\int \mathcal{L}^\tau f(x) r_n^\tau(dx)$ tend vers 0 quand n tend vers $+\infty$, en probabilité par rapport à $Q_{y_0}^\tau$, et cela uniformément en τ et en y_0 .

Démonstration : on utilise celle du lemme 6.14 de [DE] : on veut montrer que, pour tout $\varepsilon > 0$, $Q_{y_0}^\tau(|\int \mathcal{L}^\tau f(x) r_n^\tau(dx)| \geq \varepsilon)$ tend vers 0 quand n tend vers $+\infty$, uniformément en τ et en y_0 . Or, $\int \mathcal{L}^\tau f(x) r_n^\tau(dx)$ est égal à $\frac{1}{n} \sum_{j=1}^n \mathcal{L}^\tau f(\frac{1}{\tau} Y_{j-1}^\tau)$; pour tout τ , on majore donc

$$\frac{1}{n} \mathbb{E}_{y_0}^{Q^\tau} \left(\left[f\left(\frac{1}{\tau} Y_n^\tau\right) - f\left(\frac{1}{\tau} Y_0^\tau\right) - \sum_{j=1}^n \mathcal{L}^\tau f\left(\frac{1}{\tau} Y_{j-1}^\tau\right) \right]^2 \right)$$

par $\frac{4}{n^2} \|f\|_\infty^2$. $\square$

On pose

$$r_{n,1}^\tau(dx_1) = \frac{1}{n} \sum_{j=1}^n \delta_{\frac{1}{\tau} \pi(Y_{j-1}^\tau)}(dx_1)$$

Pour ω fixé, $r_{n,1}^\tau$ est une probabilité sur $\mathbb{R}$ et on a $r_{n,\tau}^1 = r_{n,1}^\tau(\mathbb{R}^-)$ (où $r_{n,\tau}^1$ est la variable aléatoire définie en (3.6)).

Les contraintes $\pi(\beta^1) > 0$ et $\pi(\beta^2) < 0$ garantissent, pour le processus $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$ sous la probabilité $Q_{y_0}^\tau$, avec y_0 dans $B(x_0, \varepsilon_1) \cap \{|\pi(\cdot)| \leq \tau M\}$, le résultat de stabilité suivant sur la chaîne $(\frac{1}{\tau} \pi(Y_k^\tau))_{k \leq n}$:

Lemme 3.1.4

Pour tout λ de $[0, 1]$, il existe $\mathcal{K}_\lambda$, compact de $\mathbb{R}$ tel que, quels que soient $\tau > 0$, $n \in \mathbb{N}$ et y_0 dans $B(x_0, \varepsilon_1) \cap \{|\pi(\cdot)| \leq \tau M\}$, on ait :

$$Q_{y_0}^\tau \left(r_{n,1}^\tau(\mathcal{K}_\lambda) \geq 1 - \lambda \right) > 1 - \lambda$$

Démonstration : (analogue de [DE] lemme 6.15)

1. Grâce à $\pi(\beta^1) > 0$ et $\pi(\beta^2) < 0$, le générateur $\mathcal{L}^\tau$ vérifie, pour $V(x) = |\pi(x)|$:

$$\exists C, \lambda_0 > 0 \text{ tels que } \forall \tau, |\pi(x)| > \lambda_0 \Rightarrow \mathcal{L}^\tau V(x) \leq -C \quad (3.9)$$

En effet, considérons x dans $\mathbb{R}^d$ et posons $x_1 = \pi(x)$. Si x_1 est négatif, d'après (3.8), on a :

$$\mathcal{L}^\tau V(x) = \int_{\mathbb{R}^d} (|\pi(x+y)| - |\pi(x)|) e^{\langle \alpha_{\tau x}^1, y \rangle - H_{\tau x}^1(\alpha_{\tau x}^1)} \mu_{\tau x}^1(dy)$$

Notation 3.1.5 On note $\nu_x^i(dy_1)$ la mesure sur $\mathbb{R}$ obtenue en intégrant $e^{\langle \alpha_x^i, y \rangle - H_x^i(\alpha_x^i)} \mu_x^i(dy)$ sur les $d-1$ dernières coordonnées de y .

On a, toujours avec $x_1 = \pi(x)$ négatif,

$$\begin{aligned}\mathcal{L}^\tau V(x) &= \int_{\mathbb{R}} (|x_1 + y_1| + x_1) \nu_{\tau x}^1(dy_1) \\ &= \int_{\mathbb{R}} [x_1 - (x_1 + y_1)] \nu_{\tau x}^1(dy_1) + \int_{\mathbb{R}} [(x_1 + y_1) + |x_1 + y_1|] \nu_{\tau x}^1(dy_1)\end{aligned}$$

La première intégrale de cette somme vaut $-\pi(\nabla H_x^i(\alpha_x^i)) = -\pi(\beta^1)$, la seconde est égale à $\int_{-x_1}^{+\infty} 2(x_1 + y_1) \nu_{\tau x}^1(dy_1)$, quantité qui est inférieure à $\frac{\pi(\beta^1)}{2}$ pour $|x_1|$ assez grand. Donc $\mathcal{L}^\tau V(x)$ est inférieur à $-\frac{\pi(\beta^1)}{2}$. On montre de même que, pour x_1 positif assez grand, $\mathcal{L}^\tau V(x)$ est majoré par $\frac{\pi(\beta^2)}{2}$, d'où (3.9) avec $C = \frac{\pi(\beta^1) \wedge |\pi(\beta^2)|}{2}$, constante indépendante de τ (elle peut également être indépendante des $\pi(\beta^i)$ si on se limite aux β^i tels que $(-1)^i \pi(\beta^i) \leq -\theta$, cf remarque après l'énoncé du théorème 3.1.8). $\square$

2. La condition (3.9) permet de prouver, grâce à une extension d'un résultat classique (par exemple [Ku] p.146), que pour tout λ de $[0, 1]$, il existe $\mathcal{K}_\lambda$ compact de $\mathbb{R}$ tel que, pour tout y_0 de $B(x_0, \varepsilon_1) \cap \{|\pi(\cdot)| \leq \tau M\}$, on ait :

$$\inf_{\tau, j} Q_{y_0}^\tau \left(\frac{1}{\tau} \pi(Y_{j-1}^\tau) \in \mathcal{K}_\lambda \right) \geq 1 - \lambda \quad (3.10)$$

c'est-à-dire que le processus $(\frac{1}{\tau} \pi(Y_j^\tau))_j$ est tendu, uniformément en τ .

$\square$ La démonstration de Kushner concerne un unique processus de diffusion X . Nous allons la généraliser au cas où on a une famille de processus réels paramétrés par τ , (X^τ) , vérifiant une condition telle que (3.9), donc uniforme en τ . Supposons qu'il existe C et $\lambda_0 > 0$ tels que :

$$\text{pour tout } \tau, V(x) > \lambda_0 \text{ implique } \mathcal{L}^\tau V(x) \leq -C \quad (3.11)$$

et montrons que cette propriété implique que pour tout compact B de $\mathbb{R}$ et tout λ de $[0, 1]$, il existe $\mathcal{K}_\lambda$ compact de $\mathbb{R}$ tel que, pour tout x_0 dans B , on ait :

$$\inf_{\tau, j} Q_{x_0}^\tau (X^\tau \in \mathcal{K}_\lambda) \geq 1 - \lambda. \quad (3.12)$$

Nous rappelons brièvement les étapes essentielles de la démonstration de Kushner, en vérifiant qu'il y a uniformité en τ . Posons $\mathcal{V}_0 = \{x | V(x) \leq \lambda_0\}$. D'après (3.11), si x n'appartient pas à $\mathcal{V}_0$, pour tout τ , on a $\mathcal{L}^\tau V(x) \leq -C$. On pose, pour $\lambda_1 > \lambda_0$, $\mathcal{V}_1 = \{x | V(x) \leq \lambda_1\}$ et on définit (Fr désigne la

frontière d'une partie de $\mathbb{R}$) :

$$\begin{cases} \sigma_1^\tau = \min\{t | X^\tau(t) \in \mathcal{V}_0\}, \\ \dots \\ \sigma_{2k}^\tau = \min\{t > \sigma_{2k-1}^\tau | X^\tau(t) \in Fr\mathcal{V}_1\}, \\ \sigma_{2k+1}^\tau = \min\{t > \sigma_{2k}^\tau | X^\tau(t) \in Fr\mathcal{V}_0\}, \\ \dots \end{cases}$$

La démonstration est basée sur les résultats suivants, ici vérifiés uniformément en τ :

$$\begin{aligned} \mathbb{E}_{\sigma_{2k}^\tau}(\sigma_{2k+1}^\tau - \sigma_{2k}^\tau) \mathbb{I}_{\{\sigma_{2k}^\tau < +\infty\}} &\leq \frac{\lambda_1}{C} \mathbb{I}_{\{\sigma_{2k}^\tau < +\infty\}}, \\ \mathbb{I}_{\{\sigma_{2k}^\tau < +\infty\}} P_{\sigma_{2k}^\tau} \left\{ \sup_{[\sigma_k^\tau, \sigma_{k+1}^\tau]} V(X^\tau(t)) \right\} &\xrightarrow{\lambda \rightarrow +\infty} 0, \end{aligned}$$

et il existe $\delta_1, \delta_2 > 0$ tels que

$$\mathbb{I}_{\{\sigma_{2k-1}^\tau < +\infty\}} P_{\sigma_{2k-1}^\tau} \{\sigma_{2k}^\tau - \sigma_{2k-1}^\tau \geq \delta_1\} \geq \delta_2 \mathbb{I}_{\{\sigma_{2k-1}^\tau < +\infty\}}. \quad \square$$

Remarquons que l'origine $X^\tau(0)$ doit rester dans un compact, ce qui est vérifié dans notre cas puisqu'on suppose $\frac{1}{\tau} |\pi(Y_0^\tau)| \leq M$ P -ps.

3. De (3.10), on déduit le lemme de stabilité 3.1.4 de la façon suivante :

pour j donné dans $\mathbb{N}$, on considère $(W_j^k)_{k \in \mathbb{N}}$ une suite de vecteurs i.i.d. de même distribution que $\frac{1}{\tau} \pi(Y_{j-1}^\tau)$. On note $r_{n,1}^{\tau,k}(dx_1) = \frac{1}{n} \sum_{j=1}^n \delta_{W_j^k}(dx_1)$ la mesure d'occupation de $(W_j^k)_{1 \leq j \leq n}$. On suppose que le lemme 3.1.4 est faux. Alors il existe λ dans $[0, 1]$ tel que, pour tout compact $\mathcal{K}$ dans $\mathbb{R}$, on puisse trouver τ , n et y_0 dans $B(x_0, \varepsilon_1) \cap \{|\pi(\cdot)| \leq \tau M\}$ tels qu'on ait :

$$Q_{y_0}^\tau \left(r_{n,1}^\tau(\mathcal{K}^c) \geq \lambda \right) \geq \lambda.$$

Dans ce cas, il existe $M_n(\mathcal{K})$ tel que, pour $m \geq M_n(\mathcal{K})$, le cardinal de l'ensemble $\{k \in \{1, \dots, m\} \mid r_{n,1}^{\tau,k}(\mathcal{K}^c) \geq \lambda\}$ est plus grand que $\frac{\lambda m}{2}$. On prend γ dans $[0, \frac{\lambda^2}{2}]$ et $\mathcal{K}_\gamma$ le compact associé tel que (3.10) soit vérifié. On a :

$$\sum_{k=1}^m \sum_{j=1}^n \mathbb{I}_{\{W_j^k \in (\mathcal{K}_\gamma)^c\}} \geq \sum_{\{k \mid r_{n,1}^{\tau,k}(\mathcal{K}^c) \geq \lambda\}} \sum_{j=1}^n \mathbb{I}_{\{W_j^k \in (\mathcal{K}_\gamma)^c\}}$$

Dans le terme de droite, $\sum_{j=1}^n \mathbb{I}_{\{W_j^k \in (\mathcal{K}_\gamma)^c\}}$ est strictement plus grand que λn pour les k considérés. Alors, pour $m \geq M_n(\mathcal{K}_\gamma)$, on a :

$$\sum_{k=1}^m \sum_{j=1}^n \mathbb{I}_{\{W_j^k \in (\mathcal{K}_\gamma)^c\}} > \frac{\lambda^2 m n}{2}.$$

On en déduit :

$$\liminf_{m \rightarrow +\infty} \frac{1}{m} \sum_{k=1}^m \sum_{j=1}^n \mathbb{I}_{\{W_j^k \in (\mathcal{K}_\gamma)^c\}} > \frac{\lambda m}{2} \lambda n.$$

Or, d'après le choix de $\mathcal{K}_\gamma$ et des (W_j^k) , on a aussi :

$$\lim_{m \rightarrow +\infty} \frac{1}{m} \sum_{k=1}^m \sum_{j=1}^n \mathbb{I}_{\{W_j^k \in (\mathcal{K}_\gamma)^c\}} = \sum_{j=1}^n Q_{y_0}^\tau \left(\frac{1}{\tau} \pi(Y_{j-1}^\tau) \in (\mathcal{K}_\gamma)^c \right) \leq n\gamma$$

ce qui est absurde puisque $n\gamma$ est strictement plus petit que $\frac{\lambda^2 n}{2}$. Le lemme 3.1.4 est donc démontré. $\square$

Considérons, pour $\delta > 0$, g_δ , une fonction impaire bornée croissante sur $\mathbb{R}$, de classe C^1 , égale à l'identité sur $[0, \frac{1}{\delta}]$ et telle qu'on ait $g_\delta(x) \leq x$ pour tout $x \geq 0$, et g'_δ bornée uniformément en δ . On définit ensuite f_δ de $\mathbb{R}^d$ dans $\mathbb{R}$, en posant $f_\delta(x) = (g_\delta(\pi(x)), 0, \dots, 0)$. La famille de fonctions f_δ correspond au type de familles considérées dans [DE], paragraphe 6.b : ce sont des applications de classe C^1 , bornées, et telles que, pour chaque δ , $f_\delta(x)$ ne dépend que de $\pi(x)$ et vaut $\pi(x)$ quand $|\pi(x)|$ est inférieur à $\frac{1}{\delta}$; on a également $|f_\delta(x)| \leq |\pi(x)|$ et $\sup_{\delta > 0} \sup_x |\nabla f_\delta(x)| \leq C_1$ avec $C_1 > 0$.

On pose $\beta(x) = \mathbb{I}_{P^1 \cup P^0}(x)\pi(\beta^1) + \mathbb{I}_{P^2}(x)\pi(\beta^2)$. Cette quantité coïncide avec $\mathcal{L}^\tau \pi(x)$ et on a :

$$\int \beta(x) r_n^\tau(dx) = \sum_{i=1}^2 r_{n,\tau}^i \pi(\beta^i) \quad (3.13)$$

On veut comparer $\int \mathcal{L}^\tau f_\delta(x) r_n^\tau(dx)$ et $\int \beta(x) r_n^\tau(dx)$. On a :

$$|\mathcal{L}^\tau f_\delta(x)| \leq \sup_{\mathbb{R}^d} |\nabla f_\delta| \sum_{i=1}^2 \int |y_i| e^{(\alpha_{\tau x}^i, y_i) - H_{\tau x}^i(\alpha_{\tau x}^i)} \mu_{\tau x}^i(dy_i),$$

donc, pour τx dans $B(x_0, 2\varepsilon_1)$ (ensemble inclus dans $\mathcal{U}$), $|\mathcal{L}^\tau f_\delta(x)|$ est majoré par $M \sup_{\mathbb{R}^d} |\nabla f_\delta|$ donc par MC_1 . Soit $\lambda > 0$. On considère le compact $\mathcal{K}_\lambda$ associé à λ comme dans le lemme 3.1.4 et y_0 dans $B(x_0, \varepsilon_1) \cap \{|\pi(\cdot)| \leq \tau M\}$. On a :

$$Q_{y_0}^\tau \left(r_{n,1}^\tau((\mathcal{K}_\lambda)^c) \geq \lambda \right) < \lambda \quad (3.14)$$

Si τx appartient à $B(x_0, 2\varepsilon_1)$, on majore $|\mathcal{L}^\tau f_\delta(x) - \beta(x)|$ par $MC_1 + |\pi(\beta^1)| \vee |\pi(\beta^2)|$ donc par $MC_1 + 2M$, que l'on note C' .

Si $\pi(x)$ appartient à $\mathcal{K}_\lambda$, on écrit une majoration plus précise: $\mathcal{L}^\tau f_\delta(x) - \beta(x)$ est égal à $\mathcal{L}^\tau f_\delta(x) - \mathcal{L}^\tau \pi(x)$ donc à:

$$\sum_{i=1}^2 \mathbb{I}_{P^i}(x) \int \left(\left[f_\delta(x + y_i) - f_\delta(x) \right] - \left[\pi(x + y^i) - \pi(x) \right] \right) \nu_{\tau x}^i(dy_1^i)$$

(cf notation 3.1.5). On prend δ tel que $\mathcal{K}_\lambda$ soit inclus dans $[-\frac{1}{2\delta}, \frac{1}{2\delta}]$; alors, si $\pi(x)$ appartient à $\mathcal{K}_\lambda$, on a $f_\delta(x) = \pi(x)$ et, pour $|\pi(y)| < \frac{1}{2\delta}$, $f_\delta(x + y) = \pi(x + y)$, donc:

$$|\mathcal{L}^\tau f_\delta(x) - \beta(x)| \leq \sum_{i=1}^2 \int_{|\pi(y^i)| > \frac{1}{2\delta}} 2|\pi(x) + \pi(y^i)| \nu_{\tau x}^i(dy_1^i).$$

$\mathcal{K}_\lambda$ est fixé; on fait tendre δ vers 0 ($\mathcal{K}_\lambda$ reste dans $[-\frac{1}{2\delta}, \frac{1}{2\delta}]$). Le majorant précédent tend vers 0 (l'ensemble $\bigcup_{x \in \mathcal{K}} S_{\tau x}^i$ ne rencontre pas $\{|\pi(\cdot)| > \frac{1}{2\delta}\}$ dès que δ est assez petit); il est donc inférieur à $\lambda C'$ pour δ assez petit. On a ainsi, pour τx dans $B(x_0, 2\varepsilon_1)$:

$$\begin{cases} |\mathcal{L}^\tau f_\delta(x) - \beta(x)| \leq C' & \text{pour } \pi(x) \notin \mathcal{K}_\lambda \\ \leq \lambda C' & \text{pour } \pi(x) \in \mathcal{K}_\lambda \end{cases}$$

On en déduit:

$$\begin{aligned} & Q_{y_0}^\tau \left(\left| \int \mathcal{L}^\tau f_\delta(x) r_n^\tau(dx) - \int \beta(x) r_n^\tau(dx) \right| \geq 2\lambda C' \right) \\ & \leq Q_{y_0}^\tau \left(\left| \int_{\pi(x) \notin \mathcal{K}_\lambda} [\mathcal{L}^\tau f_\delta(x) - \beta(x)] r_n^\tau(dx) \right| \geq \lambda C' \right) \leq Q_{y_0}^\tau \left(C' r_{n,1}^\tau((\mathcal{K}_\lambda)^c) \geq \lambda C' \right), \end{aligned}$$

d'où, d'après (3.14):

$$Q_{y_0}^\tau \left(\left| \int \mathcal{L}^\tau f_\delta(x) r_n^\tau(dx) - \int \beta(x) r_n^\tau(dx) \right| \geq 2\lambda C' \right) < \lambda \quad (3.15)$$

Or, d'après le lemme 3.1.3, il existe N tel que, pour n supérieur à N et pour tous τ et y_0 , on ait $Q_{y_0}^\tau(|\int \mathcal{L}^\tau f_\delta(x) r_n^\tau(dx)| \geq \lambda C') \leq \lambda$; alors, d'après (3.13) et (3.15), pour $n \geq N$, on a:

$$Q_{y_0}^\tau \left(\left| \sum_{i=1}^2 r_{n,\tau}^i \pi(\beta^i) \right| \geq 3\lambda C' \right) \leq 2\lambda.$$

On en déduit que pour n assez grand, on a, sur un ensemble de mesure $Q_{y_0}^\tau$ au moins égale à $1 - 2\lambda$:

$$r_{n,\tau}^1 + r_{n,\tau}^2 = 1 \quad \text{et} \quad r_{n,\tau}^1 \pi(\beta^1) + r_{n,\tau}^2 \pi(\beta^2) = c$$

avec $|c| \leq 3\lambda C'$. On a alors

$$r_{n,\tau}^1 = \rho^1 + \frac{c}{\pi(\beta^1) - \pi(\beta^2)} \quad \text{et} \quad r_{n,\tau}^2 = \rho^2 - \frac{c}{\pi(\beta^1) - \pi(\beta^2)}.$$

Puisque λ peut être pris arbitrairement petit, on en déduit que $r_{n,\tau}^i$ converge vers ρ^i en probabilité par rapport à $Q_{y_0}^\tau$, uniformément en τ , pour $n\tau \leq K$, et en la condition initiale pour y_0 dans $B(x_0, \varepsilon_1) \cap \{|\pi(\cdot)| \leq \tau M\}$, c'est-à-dire la deuxième convergence cherchée. Notons que le fait que, pour $(-1)^i \pi(\beta^i) < 0$, il existe un unique (ρ^1, ρ^2) tel que $\sum_{i=1}^2 \rho^i = 1$ et $\pi(\sum_{i=1}^2 \rho^i \beta^i) = 0$ est essentiel à la conclusion.

Les deux convergences (3.7) sont prouvées ; pour montrer le lemme 3.1.2, on a besoin d'un dernier résultat, extension au cas des champs de probabilités non constants du lemme 6.12 de [DE] :

Lemme 3.1.6

Soit $G : (\mathbb{R}^d)^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable, bornée. On pose, pour x dans $\mathbb{R}^d$ et β application de $\mathbb{R}^d$ dans $\mathbb{R}^d$: $f_x^i(\beta) = -[\langle \alpha_x^i, \beta(x) \rangle - H_x^i(\alpha_x^i)]$. Alors, pour tout y_0 de $\mathbb{R}^d$, $\int G(w_{n,\tau}^1, w_{n,\tau}^2, r_{n,\tau}^1, r_{n,\tau}^2) dP_{y_0}^\tau$ est égal à :

$$\int G(w_{n,\tau}^1, w_{n,\tau}^2, r_{n,\tau}^1, r_{n,\tau}^2) \exp \left[\sum_{j=0}^{n-1} \sum_{i=1}^2 \mathbb{I}_{A_j^i} f_{Y_j^\tau}^i(X_j^i) \right] dQ_{y_0}^\tau$$

Démonstration : d'après (3.5), il suffit de montrer que, pour tout ensemble d'indices $(i_k)_{k=1,\dots,n}$,

$$\int \left(\prod_{k=1}^n \mathbb{I}_{A_k^{i_k}} \right) G(w_{n,\tau}^1, w_{n,\tau}^2, r_{n,\tau}^1, r_{n,\tau}^2) \exp \left[\sum_{j=0}^{n-1} \sum_{i=1}^2 f_{Y_j^\tau}^i(X_j^i(Y_j^\tau)) \right] dQ_{y_0}^\tau$$

est égal à

$$\int \left(\prod_{k=1}^n \mathbb{I}_{A_k^{i_k}} \right) G(w_{n,\tau}^1, w_{n,\tau}^2, r_{n,\tau}^1, r_{n,\tau}^2) \exp \left[\sum_{j=0}^{n-1} \sum_{i=1}^2 \mathbb{I}_{A_j^i} f_{Y_j^\tau}^i(X_j^i) \right] dQ_{y_0}^\tau.$$

Si y_0 appartient à $P^1 \cup P^0$ (resp. P^2), on a P-ps $\mathbb{I}_{A_1^1} = 1$ (resp. $\mathbb{I}_{A_1^1} = 0$). Pour i fixé, on utilise le fait que $(\prod_{k=1}^n \mathbb{I}_{A_k^{i_k}}) G(w_{n,\tau}^1, w_{n,\tau}^2, r_{n,\tau}^1, r_{n,\tau}^2) \exp \sum_{k=0}^{n-1} f_{Y_k^\tau}^{i_k}(X_k^{i_k}(Y_k^\tau))$ est une fonction mesurable de $\{X_k^{i_k}(Y_k^\tau), 0 \leq k \leq n-1\}$, et que, pour une telle fonction Γ , on a :

$$\int \Gamma(\{X_k^{i_k}(Y_k^\tau)\}) \exp \sum_{k=0}^{n-1} f_{Y_k^\tau}^{3-i_k}(X_k^{3-i_k}(Y_k^\tau)) dQ_{y_0}^\tau = \int \Gamma(\{X_k^{i_k}(Y_k^\tau)\}) dQ_{y_0}^\tau$$

Cette dernière égalité est obtenue en utilisant le fait que $X_k^{i_k}(Y_k^\tau)$ et $X_k^{3-i_k}(Y_k^\tau)$ sont indépendants conditionnellement à $\mathcal{F}_k$ et que $\int \exp \sum_{k=0}^{n-1} f_{Y_k^\tau}^{3-i_k}(X_k^{3-i_k}(Y_k^\tau)) dQ_{y_0}^\tau$ vaut

1. On a alors :

$$\int \left(\prod_{k=1}^n \mathbb{I}_{A_k^{i_k}} \right) G(w_{n,\tau}^1, w_{n,\tau}^2, r_{n,\tau}^1, r_{n,\tau}^2) \exp \left[\sum_{j=0}^{n-1} \sum_{i=1}^2 f_{Y_j^\tau}^i(X_j^i(Y_j^\tau)) \right] dQ_{y_0}^\tau =$$

$$\int \left(\prod_{k=1}^n \mathbb{I}_{A_k^{i_k}} \right) G(w_{n,\tau}^1, w_{n,\tau}^2, r_{n,\tau}^1, r_{n,\tau}^2) \exp \sum_{k=0}^{n-1} f_{Y_k^\tau}^{i_k}(X_k^{i_k}(Y_k^\tau)) dQ_{y_0}^\tau.$$

On conclut en utilisant que, sur l'ensemble $\left\{ \prod_{k=1}^n \mathbb{I}_{A_k^{i_k}} = 1 \right\}$, $\sum_{k=0}^{n-1} f_{Y_k^\tau}^{i_k}(X_k^{i_k}(Y_k^\tau))$ est égal à $\sum_{j=0}^{n-1} \sum_{i=1}^2 \mathbb{I}_{A_j^i} f_{Y_j^\tau}^i(X_j^i)$. $\square$

Soit y_0 dans $B(x_0, \varepsilon_1) \cap \{|\pi(\cdot)| \leq \tau M\}$. Nous utilisons les deux convergences (3.7) et le lemme 3.1.6 pour obtenir l'inégalité du lemme 3.1.2 pour le quadruplet choisi $(\rho^1, \rho^2, \beta^1, \beta^2)$:

pour ε_3 assez petit (fonction de ε_2 et de M), d'après $\frac{Y_n^\tau - Y_0^\tau}{n\tau} = \sum_{i=1}^2 r_{n,\tau}^i w_{n,\tau}^i$, on a :

$$P_{y_0} \left(\frac{Y_n^\tau - Y_0^\tau}{n\tau} \in B \left(\sum_{i=1}^2 \rho^i \beta^i, \varepsilon_2 \right) \right) \geq \int_{E_{y_0}} \frac{dP_n^\tau}{dQ_n^\tau}(\omega) Q_n^\tau(d\omega), \quad (3.16)$$

où on a posé :

$$E_{y_0} = \{Y_0^\tau = y_0\} \cap \bigcap_{i=1,2} \left(\{r_{n,\tau}^i(w_{n,\tau}^i - \beta^i) \in B(0, \varepsilon_3)\} \cap \{r_{n,\tau}^i \in B(\rho^i, \varepsilon_3)\} \right)$$

Sur cet ensemble, d'après le lemme 3.1.6, $\log \frac{dP_n^\tau}{dQ_n^\tau}$ s'écrit :

$$\log \frac{dP_n^\tau}{dQ_n^\tau} = - \sum_{j=0}^{n-1} \sum_{i=1}^2 \mathbb{I}_{A_j^i} \left[\langle \alpha_{Y_j^\tau}^i, X_j^i(Y_j^\tau) \rangle - H_{Y_j^\tau}^i(\alpha_{Y_j^\tau}^i) \right]$$

D'après (3.3), pour y dans $B(x_0, 2\varepsilon_1)$, on a :

$$|L_y^i(\beta^i) - L_{x_0}^i(\beta^i)| < \varepsilon \quad \text{et} \quad |\alpha_y^i - \alpha_{x_0}^i| < \varepsilon,$$

et donc :

$$|H_y^i(\alpha_y^i) - H_{x_0}^i(\alpha_{x_0}^i)| \leq |\langle \alpha_y^i - \alpha_{x_0}^i, \beta^i \rangle| + |L_y^i(\beta^i) - L_{x_0}^i(\beta^i)| \leq \varepsilon(M+1).$$

D'après (3.4) et le fait que $|X_j^i(Y_j^\tau)|$ est borné P-presque sûrement par M , on minore donc $\log \frac{dP_n^\tau}{dQ_n^\tau}$ sur E_{y_0} par :

$$- \sum_{j=0}^{n-1} \sum_{i=1}^2 \mathbb{I}_{A_j^i} \left[\langle \alpha_{x_0}^i, X_j^i(Y_j^\tau) \rangle - H_{x_0}^i(\alpha_{x_0}^i) \right] - n\varepsilon(2M+1)$$

ou encore par :

$$-n \left(\sum_{i=1}^2 \left[\langle \alpha_{x_0}^i, r_{n,\tau}^i w_{n,\tau}^i \rangle - r_{n,\tau}^i H_{x_0}^i(\alpha_{x_0}^i) \right] + \varepsilon(2M+1) \right),$$

égal à

$$-n \left(\sum_{i=1}^2 \left[\langle \alpha_{x_0}^i, r_{n,\tau}^i (w_{n,\tau}^i - \beta^i) \rangle + r_{n,\tau}^i L_{x_0}^i(\beta^i) \right] + \varepsilon(2M+1) \right). \quad (3.17)$$

D'après les convergences, de $r_{n,\tau}^i (w_{n,\tau}^i - \beta^i)$ vers 0 et de $r_{n,\tau}^i$ vers ρ^i , en probabilité par rapport à $Q_{y_0}^\tau$, on a donc, uniformément en τ tel que $n\tau \leq K$ et en y_0 de $B(x_0, \varepsilon_1) \cap \{|\pi(\cdot)| \leq \tau M\}$:

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_{y_0} \left(\frac{Y_n^\tau - Y_0^\tau}{n\tau} \in B \left(\sum_{i=1}^2 \rho^i \beta^i, \varepsilon_2 \right) \right) \geq - \sum_{i=1}^2 \rho^i L_{x_0}^i(\beta^i) - (2M+1)\varepsilon \quad (3.18)$$

ce qui achève la démonstration du lemme 3.1.2.

3.1.3 Minoration pour les vitesses moyennes en présence d'un seul champ de probabilités

En présence d'un seul champ de probabilités, la démonstration précédente se simplifie beaucoup ; la première convergence de (3.7) suffit pour obtenir le résultat suivant :

Lemme 3.1.7

On considère un processus d'apprentissage lent, $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à un champ de probabilités (μ_x^i) vérifiant les hypothèses (Hmin) et $\mathcal{K}$ un compact de $\mathbb{R}^d$. Soit $\varepsilon, \delta_0 > 0$. Alors il existe ε_1, K tels que, pour tous $\varepsilon_2 > 0$, x_0 dans $\mathcal{K}$ et β dans $S_{x_0}^i$, on puisse trouver N tel que $n \geq N$ implique, pour tout τ tel que $n\tau \leq K$ et tout y_0 de $B(x_0, \varepsilon_1)$:

$$\frac{1}{n} \log P_{y_0} \left(\frac{Y_n^\tau - Y_0^\tau}{n\tau} \in B(\beta, \varepsilon_2) \right) \geq -L_{x_0}^i(\beta) - \varepsilon.$$

Démonstration : Soit $\varepsilon > 0$. On prend δ_0 tel que β appartienne à $(S_{x_0}^i)^{-\delta_0}$. A ε et δ_0 , on associe ε_1 tel que (3.3) soit vérifié et K tel que $KM \leq \varepsilon_1$. On pose $\Omega_0 = \{Y_0^\tau \in B(x_0, \varepsilon_1)\}$. On effectue le changement de probabilité associé à β comme

dans (3.5) : pour chaque (τ, n) , on définit une nouvelle probabilité R_n^τ sur Ω de dérivée de Radon-Nikodym

$$\frac{dR_n^\tau}{dP_n^\tau}(\omega) = \exp \sum_{j=0}^{n-1} \left[\langle \nabla L_{Y_j^\tau(\omega)}^i(\beta), X_j^i(Y_j^\tau(\omega))(\omega) \rangle - H_{Y_j^\tau(\omega)}^i(\nabla L_{Y_j^\tau(\omega)}^i(\beta)) \right] \quad (3.19)$$

On appelle R^τ la probabilité sur $(\Omega, \mathcal{F})$ extension de Kolmogorov de la suite $(R_n^\tau)_{n \in \mathbb{N}}$. On pose $w_{n,\tau} = \frac{1}{n} \sum_{j=0}^{n-1} X_j^i(Y_j^\tau) = \frac{Y_n^\tau - Y_0^\tau}{n\tau}$ et on considère y_0 dans $B(x_0, \varepsilon_1)$. Alors $w_{n,\tau}$ converge vers β par rapport à $R_{y_0}^\tau$ quand n tend vers $+\infty$ (cf (3.7)). On en déduit :

$$P_{y_0} \left(\frac{Y_n^\tau - Y_0^\tau}{n\tau} \in B(\beta, \varepsilon_2) \right) \geq \int_{E_{y_0}} \frac{dP_n^\tau}{dR_n^\tau}(\omega) R_n^\tau(d\omega),$$

où E_{y_0} désigne l'ensemble $\{Y_0^\tau = y_0\} \cap \{|w_{n,\tau} - \beta| \leq \varepsilon_2\}$.

On conclut comme dans le paragraphe précédent : on a obtenu l'équivalent de (3.16), on utilise de nouveau le lemme 3.1.6 ; d'après le choix de ε_1 et de K , pour $n\tau$ inférieur à K , les variations de $L^i(\beta)$ et de $\nabla L^i(\beta)$ sur $B(x_0, \varepsilon_1 + n\tau M)$ sont inférieures à ε . De nouveau, sur E_{y_0} , $\log \frac{dP_n^\tau}{dR_n^\tau}$ est minoré par (3.17), avec $\beta = \beta^i$. Cela permet de minorer $\liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_{y_0} \left(\frac{Y_n^\tau - Y_0^\tau}{n\tau} \in B(\beta, \varepsilon_2) \right)$ par $-L_{x_0}^i(\beta)$ à un terme en ε près, comme dans (3.18), d'où le résultat. $\square$

3.1.4 Minoration de la probabilité de suivre un vecteur parallèle à P^0 avec une contrainte plus faible sur $\pi(Y_0^\tau)$

Dans ce paragraphe, nous déduisons du lemme 3.1.2 un énoncé équivalent avec une contrainte plus faible sur $\pi(Y_0^\tau)$, résultat qui permettra d'obtenir la minoration pour des tubes de trajectoires. On utilise le lemme 3.1.7 pour traiter les pas effectués avant la traversée de P^0 .

La nouvelle contrainte sur $\pi(Y_0^\tau)$ est $|\pi(Y_0^\tau)| \leq n\tau\alpha$. Pour déduire du résultat la minoration pour des tubes, on ajoute dans la probabilité à minorer la condition $|\pi(Y_n^\tau)| \leq n\tau\alpha$, ce qui permettra d'utiliser le théorème sur tous les intervalles $[jn\tau, (j+1)n\tau]$ successifs. Rappelons que, pour x_0 dans P^0 et $\delta_0 > 0$, $\mathcal{E}(x_0, \delta_0)$ est l'ensemble des $(\rho^1, \rho^2, \beta^1, \beta^2)$ tels que (notation 3.1.1) :

$$(-1)^i \pi(\beta^i) < 0, \beta^i \in (S_{x_0}^i)^{-\delta_0}, \rho^i \geq 0, \sum_{i=1}^2 \rho^i = 1 \text{ et } \pi \left(\sum_{i=1}^2 \rho^i \beta^i \right) = 0.$$

Théorème 3.1.8

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (Hmin) et $\mathcal{K}$ un compact de $\mathbb{R}^d$. Soit $\varepsilon, \delta_0 > 0$. Il existe ε_1, K tels que, pour tout $\varepsilon_2 > 0$, il existe α tel que, pour tout x_0 dans $P^0 \cap \mathcal{K}$ tel que $\mathcal{E}(x_0, \delta_0)$ soit non vide et tout $(\rho^1, \rho^2, \beta^1, \beta^2)$ dans $\overline{\mathcal{E}(x_0, \delta_0)}$, on peut trouver N tel que $n \geq N$ implique, pour tout τ tel que $n\tau \leq K$ et tout y_0 de $B(x_0, \varepsilon_1) \cap \{|\pi(\cdot)| \leq n\tau\alpha\}$:

$$\frac{1}{n} \log P_{y_0} \left(\frac{Y_n^\tau - Y_0^\tau}{n\tau} \in B \left(\sum_{i=1}^2 \rho^i \beta^i, \varepsilon_2 \right) \text{ et } |\pi(Y_n^\tau)| \leq n\tau\alpha \right) \geq - \sum_{i=1}^2 \rho^i L_{x_0}^i(\beta^i) - \varepsilon.$$

Remarquons que K et α sont indépendants de x_0 et de $(\rho^1, \rho^2, \beta^1, \beta^2)$ variant respectivement dans $P^0 \cap \mathcal{K}$ et dans $\overline{\mathcal{E}(x_0, \delta_0)}$. Une uniformité analogue est sans doute possible pour N , en se limitant, pour θ fixé, aux β^i tels que $(-1)^i \pi(\beta^i) \leq -\theta$, ainsi que dans le lemme 3.1.2, mais on ne l'utilisera pas.

Démonstration : on note σ'' l'instant de première traversée de P^0 :

$$\sigma'' = n \wedge \inf \{k \geq 0 \mid \pi(Y_0^\tau), \dots, \pi(Y_k^\tau) \text{ ne sont pas du même signe}\}$$

(NB : cette définition est différente de celle de σ ; ici, pour $\sigma'' < n$, l'accroissement $X_{\sigma''-1}^i(Y_{\sigma''-1}^\tau)$ fait franchir P^0). La variable σ'' est un temps d'arrêt.

Soit $\mathcal{K}$ un compact. On prend M comme dans (1.16) puis ε_1 et K comme dans le lemme 3.1.2. Quitte à diminuer K , on peut supposer $KM \leq \frac{\varepsilon_1}{2}$. On impose $n\tau \leq K$. On considère $(\rho^1, \rho^2, \beta^1, \beta^2)$ dans $\overline{\mathcal{E}(x_0, \delta_0)}$ et on pose, pour $\alpha > 0$ donné :

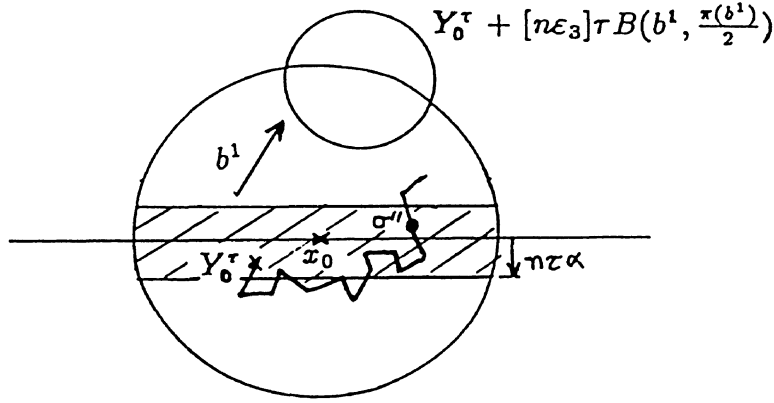
$$\Omega_0 = \left\{ Y_0^\tau \in B \left(x_0, \frac{\varepsilon_1}{2} \right) \right\} \cap \{ |\pi(Y_0^\tau)| \leq n\tau\alpha \}.$$

P-ps sur Ω_0 , on a $Y_{\sigma''}^\tau \in B(x_0, \varepsilon_1) \cap \{|\pi(\cdot)| \leq \tau M\}$, ce qui permet d'utiliser le lemme 3.1.2 pour $(Y_k^\tau)_{\sigma'' \leq k \leq n}$. On choisit le rayon $\frac{\alpha}{2}$:

soit $\varepsilon_3 \in]0, 1[$. Il existe N tel que, pour $(1 - \varepsilon_3)n \geq N$, P-ps sur $\Omega_0 \cap \{\frac{\sigma''}{n} \leq \varepsilon_3\}$, on ait pour tout τ tel que $n\tau \leq K$:

$$\frac{1}{n} \log P \left(\frac{Y_n^\tau - Y_{\sigma''}^\tau}{n\tau} \in B \left(\sum_{i=1}^2 \rho^i \beta^i, \frac{\alpha}{2} \right) \mid \mathcal{F}_{\sigma''} \right) \geq - \sum_{i=1}^2 \rho^i L_{x_0}^i(\beta^i) - \varepsilon \quad (3.20)$$

On s'intéresse maintenant à la partie $(Y_k^\tau)_{k \leq \sigma''}$. Si x est un réel, on note toujours $[x]$ sa partie entière. On pose $\Omega_0^1 = \Omega_0 \cap \{Y_0^\tau \in P^1 \cup P^0\}$ et $\Omega_0^2 = \Omega_0 \cap \{Y_0^\tau \in P^2\}$. Puisque $\mathcal{E}(x_0, \delta_0)$ est non vide, il existe b^1 dans $(S_{x_0}^1)^{-\delta_0}$ tel que $\pi(b^1)$ soit strictement positif. Grâce à la continuité de $x \mapsto S_x^1$, pour tout x_0 de $\mathcal{K}$, on peut choisir b^1 tel que $\pi(b^1)$ soit minoré par une constante π_0 strictement positive, indépendante de x_0 . On a aussi $L_{x_0}^1(b^1) \leq \sup_{x \in \mathcal{K}} J_x(\delta_0)$. On prend $\alpha < \varepsilon_3 \frac{\pi(b^1)}{2}$. Alors, P-ps sur Ω_0^1 , $Y_0^\tau + [n\varepsilon_3]\tau B(b^1, \frac{\pi(b^1)}{2})$ est inclus dans P^2 (pour n grand à cause de la partie entière).



Donc P-ps sur Ω_0^1 ,

$$P\left(\frac{\sigma''}{n} \leq \varepsilon_3 \mid \mathcal{F}_0\right) \geq P\left(\frac{X_0^1(W_0^\tau) + \dots + X_{[n\varepsilon_3]-1}^1(W_{[n\varepsilon_3]-1}^\tau)}{[n\varepsilon_3]} \in B\left(b^1, \frac{\pi(b^1)}{2}\right) \mid \mathcal{F}_0\right), \quad (3.21)$$

où (W_k^τ) est une chaîne de même valeur initiale que (Y_k^τ) et d'accroissement $X_k^1(W_k^\tau)$ (donc contrôlée par le champ de probabilités (μ_x^1)). D'après le lemme 3.1.7, on a alors, dès que n est assez grand, P-ps sur Ω_0^1 :

$$P\left(\frac{\sigma''}{n} \leq \varepsilon_3 \mid \mathcal{F}_0\right) \geq e^{-[n\varepsilon_3] (L_{x_0}^1(b^1) + \varepsilon)} \quad (3.22)$$

uniformément en τ et en la condition initiale.

Considérons maintenant la chaîne complète $(Y_k^\tau)_{0 \leq k \leq n}$. On pose $\beta = \sum_{i=1}^2 \rho^i \beta^i$ et on prend $\varepsilon_2 > 0$. La norme de $\frac{Y_n^\tau - Y_0^\tau}{n\tau}$ est majorée P-ps par $\frac{\sigma''}{n} M$, donc, pour ε_3 et α petits (indépendamment de x_0 et β) et n grand, l'ensemble

$$\left\{ \frac{Y_n^\tau - Y_0^\tau}{(n - \sigma'')\tau} \in B\left(\beta, \frac{\alpha}{2}\right) \right\} \cap \left\{ \frac{\sigma''}{n} \leq \varepsilon_3 \right\} \cap \Omega_0$$

est inclus dans l'ensemble

$$\left\{ \frac{Y_n^\tau - Y_0^\tau}{n\tau} \in B(\beta, \varepsilon_2) \right\} \cap \left\{ |\pi(Y_n^\tau)| \leq n\tau\alpha \right\}.$$

En effet l'événement définissant le premier ensemble implique les conditions suivantes:

$$\left| \frac{Y_{\sigma''}^\tau - Y_0^\tau}{n\tau} \right| \leq \frac{\varepsilon_2}{2}, \text{ pour } \varepsilon_3 M \leq \frac{\varepsilon_2}{2},$$

$$|\pi(Y_n^\tau)| \leq |\pi(Y_{\sigma''}^\tau)| + (n - \sigma'')\tau \frac{\alpha}{2} \leq \tau M + n\tau \frac{\alpha}{2} \leq n\tau\alpha, \text{ pour } M \leq n\frac{\alpha}{2}$$

et donc :

$$\left| \frac{Y_n^\tau - Y_0^\tau}{n\tau} - \beta \right| \leq \frac{\varepsilon_2}{2} + \left| \frac{Y_n^\tau - Y_{\sigma''}^\tau}{n\tau} - \beta \right|$$

et, pour ε_3 et α assez petits,

$$\frac{Y_n^\tau - Y_{\sigma''}^\tau}{n\tau} \in B\left(\left(1 - \frac{\sigma''}{n}\right)\beta, \left(1 - \frac{\sigma''}{n}\right)\frac{\alpha}{2}\right) \subset B\left(\beta, \frac{\varepsilon_2}{2}\right).$$

On a donc, P-ps sur Ω_0^1 ,

$$P\left(\frac{Y_n^\tau - Y_0^\tau}{n\tau} \in B(\beta, \varepsilon_2) \text{ et } |\pi(Y_n^\tau)| \leq n\tau\alpha \mid \mathcal{F}_0\right) \geq$$

$$P\left(\frac{Y_n^\tau - Y_{\sigma''}^\tau}{n\tau} \in B\left(\beta, \frac{\alpha}{2}\right) \mid \mathcal{F}_{\sigma''}\right) \times P\left(\frac{\sigma''}{n} \leq \varepsilon_3 \mid \mathcal{F}_0\right) \geq$$

$$\exp\left[-n \sum_{i=1}^2 \rho^i L_{x_0}^i(\beta^i) - n\varepsilon - [n\varepsilon_3] (L_{x_0}^1(b^1) + \varepsilon)\right], \quad (3.23)$$

d'après (3.20) et (3.22). On prend $\varepsilon_3 \sup_{x \in \mathcal{K}} J_x(\delta_0) < \varepsilon$ (ε_3 , et donc α , est indépendant de x_0 et $(\rho^1, \beta^1, \rho^2, \beta^2)$). On minore alors (3.23) par :

$$\exp\left[-n \sum_{i=1}^2 \rho^i L_{x_0}^i(\beta^i) - n\varepsilon - 2n\varepsilon\right]$$

On fait de même sur Ω_0^2 . Pour ε_2 fixé, on obtient l'existence de α tel que, pour tout x_0 de $\mathcal{K}$, tout $(\rho^1, \beta^1, \rho^2, \beta^2)$ de $\mathcal{E}(x_0, \delta_0)$ et pour n assez grand avec $n\tau \leq K$, on ait P-ps sur Ω_0 :

$$P\left(\frac{Y_n^\tau - Y_0^\tau}{n\tau} \in B(\beta, \varepsilon) \text{ et } |\pi(Y_n^\tau)| \leq n\tau\alpha \mid \mathcal{F}_0\right) \geq \exp -n\left(\sum_{i=1}^2 \rho^i L_{x_0}^i(\beta^i) + 3\varepsilon\right),$$

d'où le théorème 3.1.8.

3.1.5 Probabilité d'être dans un demi-espace donné après n pas

Le théorème 3.1.8 indique la probabilité que la vitesse moyenne sur un intervalle de longueur $n\tau$ soit proche d'un vecteur parallèle à P^0 pour Y_0^τ voisin de P^0 . Un deuxième résultat sur les vitesses moyennes est nécessaire avant de passer à l'étude des tubes.

On veut pouvoir étudier la probabilité que le chemin γ^τ soit proche, sur un intervalle $[a, b]$, d'un chemin φ tel que $\varphi(]a, b[)$ est inclus dans un des demi-espaces, avec $\varphi(a)$ sur la frontière. On s'occupe de ce qui se passe au voisinage de a (pour le reste de l'intervalle, on prendra un rayon suffisamment petit pour que seul un des champs de probabilités soit concerné et on utilisera le théorème 3.2.3).

On a besoin de minorer la probabilité qu'une moyenne de pas issue d'un voisinage de P^0 soit à valeurs dans un P^i donné. Il suffit de minorer cette probabilité par une quantité de la forme e^{-nB} avec B fini (localement, il s'agit du coût maximal d'une traversée). En effet cette minoration sera utilisée sur un intervalle $[a, a + t]$ avec t arbitrairement petit.

Soit x_0 dans $\mathbb{R}^d$ et pour $i = 1$ ou 2 , c^i un vecteur de $S_{x_0}^i$.

On va minorer la probabilité que $\frac{Y_n^\tau - Y_0^\tau}{n\tau}$ appartienne à $\bigcup_{\beta \in A} B(\beta, \varepsilon)$, où A est l'enveloppe convexe de c^1 et c^2 . En prenant c^1 et c^2 rentrants dans P^2 , c'est-à-dire tels que $\pi(c^1) > 0$ et $\pi(c^2) > 0$, pour Y_0^τ assez proche de P^0 (à $n\tau$ fixé) et ε petit, cette condition implique que Y_n^τ appartient à une boule incluse dans P^2 .

Cela permettra d'en déduire la minoration pour le tube d'axe φ si, pour x dans un voisinage de $\varphi(a)$, chaque S_x^i contient un vecteur rentrant dans le demi-espace contenant $\varphi(]a, b[)$.

Le prochain lemme correspond à l'extension au cas de champs de probabilités non constants du lemme 6.17 de [DE].

Lemme 3.1.9

On considère x_0 dans $\mathbb{R}^d$ et, pour $i = 1$ ou 2 , c^i dans $S_{x_0}^i$. On note A l'enveloppe convexe de c^1 et c^2 . Alors, pour tout $\varepsilon > 0$, il existe ε_1 , K et N tels que $n \geq N$ implique, pour tout τ tel que $n\tau \leq K$ et tout y_0 de $B(x_0, \varepsilon_1)$, on ait :

$$\frac{1}{n} \log P_{y_0} \left(\frac{Y_n^\tau - Y_0^\tau}{n\tau} \in \bigcup_{\beta \in A} B(\beta, \varepsilon) \right) \geq -L_{x_0}^1(c^1) \vee L_{x_0}^2(c^2) - \varepsilon.$$

Démonstration : comme pour montrer le lemme 3.1.7, on reprend la démonstration de la première convergence dans le lemme 3.1.2. Considérons $\varepsilon > 0$ et δ_0 tel que, pour chaque i , c^i appartienne à $(S_{x_0}^i)^{-\delta_0}$. A ε et δ_0 , on associe ε_1 tel que (3.3) soit vérifié et K tel que $KM \leq \varepsilon_1$. On pose $\Omega_0 = \{Y_0^\tau \in B(x_0, \varepsilon_1)\}$. On effectue le changement de probabilité associé à c^1 et c^2 comme dans (3.5) : pour chaque (τ, n) , on définit une nouvelle probabilité R_n^τ sur Ω de dérivée de Radon-Nikodym

$$\frac{dR_n^\tau}{dP_n^\tau}(\omega) = \exp \sum_{j=0}^{n-1} \sum_{i=1}^2 \left[\langle \nabla L_{Y_j^\tau(\omega)}^i(c^i), X_j^i(Y_j^\tau(\omega))(\omega) \rangle - H_{Y_j^\tau(\omega)}^i(\nabla L_{Y_j^\tau(\omega)}^i(c^i)) \right] \quad (3.24)$$

On appelle R^τ la probabilité sur $(\Omega, \mathcal{F})$ extension de Kolmogorov de la suite $(R_n^\tau)_{n \in \mathbb{N}}$. On définit $r_{n,\tau}^i$ et $w_{n,\tau}^i$ comme dans (3.6) et on considère y_0 dans $B(x_0, \varepsilon_1)$.

La convergence de $r_{n,\tau}^i(w_{n,\tau}^i - c^i)$ vers 0 par rapport à $R_{y_0}^\tau$ quand n tend vers $+\infty$ est encore vraie (cf (3.7)). En revanche, on n'a pas de résultat de convergence des $r_{n,\tau}^i$. Le vecteur $r_{n,\tau}^1 c^1 + r_{n,\tau}^2 c^2$ appartient P-ps à A , donc $B(r_{n,\tau}^1 c^1 + r_{n,\tau}^2 c^2, \varepsilon)$ est inclus dans $\bigcup_{\beta \in A} B(\beta, \varepsilon)$. D'autre part, $\frac{Y_n^\tau - Y_0^\tau}{n\tau}$ est égal à $r_{n,\tau}^1 w_{n,\tau}^1 + r_{n,\tau}^2 w_{n,\tau}^2$. On en déduit :

$$P_{y_0} \left(\frac{Y_n^\tau - Y_0^\tau}{n\tau} \in \bigcup_{\beta \in A} B(\beta, \varepsilon) \right) \geq \int_{E_{y_0}} \frac{dP_n^\tau}{dR_n^\tau}(\omega) R_n^\tau(d\omega),$$

où E_{y_0} désigne l'ensemble $\{Y_0^\tau = y_0\} \cap \{\forall i, |r_{n,\tau}^i(w_{n,\tau}^i - c^i)| \leq \frac{\varepsilon}{2}\}$.

On conclut comme précédemment : on a obtenu l'équivalent de (3.16), on utilise de nouveau le lemme 3.1.6 ; d'après le choix de ε_1 et de K , pour $n\tau$ inférieur à K , les variations de $L^i(c^i)$ et de $\nabla L^i(c^i)$ sur $B(x_0, \varepsilon_1 + n\tau M)$ sont inférieures à ε . De nouveau, sur E_{y_0} , $\log \frac{dP_n^\tau}{dR_n^\tau}$ est minoré par (3.17), avec c^i au lieu de β^i . Les variables aléatoires $r_{n,\tau}^i$ ne convergent pas vers ρ^i , mais on majore $\sum_{i=1}^2 r_{n,\tau}^i L_{x_0}^i(c^i)$ par $L_{x_0}^1(c^1) \vee L_{x_0}^2(c^2)$. Cela permet de minorer $\liminf_{n \rightarrow +\infty} \frac{1}{n} \log P_{y_0} \left(\frac{Y_n^\tau - Y_0^\tau}{n\tau} \in \bigcup_{\beta \in A} B(\beta, \varepsilon) \right)$ par $-L_{x_0}^1(c^1) \vee L_{x_0}^2(c^2)$ à un terme en ε près, comme dans (3.18), avec uniformité en τ tel que $n\tau \leq K$ et en $y_0 \in B(x_0, \varepsilon_1)$, d'où le résultat. $\square$

3.2 Minoration de la probabilité d'un tube de chemins d'axe régulier tracé sur P^0

Un chemin régulier (par rapport aux deux champs de convexes (S_x^1) et (S_x^2)) va jouer le même rôle qu'un chemin à vitesses intérieures dans [AR] (cf notre définition 1.1.9).

Définition 3.2.1 On dit qu'un chemin ψ de $C[0, T]$, tracé sur P^0 , est régulier si ψ est différentiable et si, pour tout t de $[0, T]$, $\psi'(t)$ s'écrit $\sum_{i=1}^2 \rho_i^i \beta_i^i$, avec, pour chaque i , $t \mapsto \beta_i^i$ continue par morceaux, $\beta_i^i \in S_{\psi(t)}^i$ et $(-1)^i \pi(\beta_i^i) < 0$.

Remarquons qu'un tel chemin ψ est nécessairement C^1 par morceaux (la continuité de $t \mapsto \beta_i^i$ implique celle de $t \mapsto \rho_i^i$ d'après $\rho_i^i = \frac{\pi(\beta_i^{3-i})}{\pi(\beta_i^{3-i}) - \pi(\beta_i^i)}$). D'autre part, nécessairement, pour tout t de $[0, T]$, $H_0(\psi(t))$ est vérifiée.

Rappelons que le chemin γ^τ est défini sur $[0, T]$ par :

$$\gamma^\tau(t) = Y_k^\tau \text{ si } t \in [k\tau; (k+1)\tau[\quad (3.25)$$

Théorème 3.2.2

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (Hmin) et ψ un chemin de $C[0, T]$ à valeurs dans P^0 , qui est régulier : pour tout t de $[0, T]$, $\psi'(t)$ s'écrit $\sum_{i=1}^2 \rho_i^i \beta_i^i$, avec, pour chaque i , $t \mapsto \beta_i^i$ continue par morceaux, β_i^i dans $S_{\psi(t)}^i$ et $(-1)^i \pi(\beta_i^i) < 0$. Alors, pour tout $\varepsilon > 0$, il existe α, τ_0 tels que, pour $\tau < \tau_0$, on ait :

$$\tau \log P(d_{0,T}(\gamma^\tau, \psi) \leq \varepsilon | \mathcal{F}_0) \geq - \int_0^T \sum_{i=1}^2 \rho_i^i L_{\psi(t)}^i(\beta_i^i) dt - \varepsilon$$

P -ps sur $\{Y_0^\tau \in B(\psi(0), \alpha)\}$.

Remarque : puisque le théorème 3.1.8 prend en compte $\overline{\mathcal{E}(x_0, \delta_0)}$, on pourrait montrer ce théorème pour ψ dérivable telle que $\psi'(t)$ s'écrit $\sum_{i=1}^2 \rho_i^i \beta_i^i$, avec, pour chaque i , $t \mapsto \beta_i^i$ et $t \mapsto \rho_i^i$ continues par morceaux, β_i^i appartenant à $S_{\psi(t)}^i$ et $(-1)^i \pi(\beta_i^i) \leq 0$. Ce cas sera contenu dans le théorème 3.4.9 (remarquons qu'alors pour tout t de $[0, T]$, $H_0(\psi(t))$ est vérifiée).

Démonstration: puisque $t \mapsto \beta_t^i$ est continue par morceaux, il existe δ_0 tel que, ps en $t \in [0, T]$,

$$\beta_t^i \in (S_{\psi(t)}^i)^{-\delta_0}.$$

Alors, on sait que

$$(\rho_t^1, \rho_t^2, \beta_t^1, \beta_t^2) \text{ appartient à } \mathcal{E}(\psi(t), \delta_0) \quad (3.26)$$

On passe des résultats sur les vitesses moyennes du paragraphe 3.1 à ce théorème de façon standard (cf [AR]) en utilisant la propriété de Markov pour réunir les intervalles $[jn\tau, (j+1)n\tau]$:

on prend $\mathcal{K}$ compact égal à $B(\psi(0), 1)$, M associé à $\mathcal{K}$ et T comme dans (1.16) et $\varepsilon > 0$. A δ_0 et ε , on associe ε_1 tel que (3.3) soit vérifié et on prend K tel que $KM \leq \varepsilon_1$. Soit $\varepsilon_2 \leq \frac{\varepsilon_1}{2}$. On suppose d'abord que ψ est de classe C^1 sur $[0, T]$. Alors, il existe K_1 tel que, pour $n\tau \leq K_1$, on ait :

$$|\psi(t) - \psi(jn\tau)| \leq \varepsilon_2 \text{ et } |\psi'(t) - \psi'(jn\tau)| \leq \frac{\varepsilon_2}{2T}, \text{ pour } jn\tau \leq t < (j+1)n\tau \leq T \quad (3.27)$$

A ε_2 , on associe α comme dans le théorème 3.1.8 (ε_1 et K ont été déterminés eux-mêmes comme dans ce théorème).

On pose $\Omega_0 = \{Y_0^\tau \in B(\psi(0), \varepsilon_2)\} \cap \{|\pi(Y_0^\tau)| \leq n\tau\alpha\}$ et, pour $j_0 n\tau \leq T$,

$$\Omega_{j_0} = \Omega_0 \cap \left\{ \frac{Y_{(j+1)n}^\tau - Y_{jn}^\tau}{n\tau} \in B\left(\psi'(jn\tau), \frac{\varepsilon_2}{2T}\right) \text{ et } |\pi(Y_{(j+1)n}^\tau)| \leq n\tau\alpha, 0 \leq j < j_0 \right\}.$$

Pour $n\tau \leq K \wedge K_1$, P-ps sur Ω_{j_0} , on a $|\psi(jn\tau) - Y_{jn}^\tau| \leq \varepsilon_2 + 2\frac{\varepsilon_2}{2T}jn\tau \leq 2\varepsilon_2$ pour $j \leq j_0$ et $j_0 n\tau \leq T$. Donc P-ps, l'événement Ω_{j_0} implique $d_{0, j_0 n\tau}(\psi, \gamma^\tau) \leq 3\varepsilon_2$, pour $j_0 n\tau \leq T$, d'après (3.27).

D'autre part, P-ps sur Ω_j , d'après $|\psi(jn\tau) - Y_{jn}^\tau| \leq 2\varepsilon_2 \leq \varepsilon_1$ et $|\pi(Y_{jn}^\tau)| \leq n\tau\alpha$, le théorème 3.1.8 (avec ce choix de ε_1 , K et α) implique, grâce à (3.26) en $t = jn\tau$, qu'il existe N_j tel que, pour $n \geq N_j$ et τ tel que $n\tau \leq K$, on ait :

$$\begin{aligned} P\left(\frac{Y_{(j+1)n}^\tau - Y_{jn}^\tau}{n\tau} \in B\left(\psi'(jn\tau), \frac{\varepsilon_2}{2T}\right) \text{ et } |\pi(Y_{(j+1)n}^\tau)| \leq n\tau\alpha \mid \mathcal{F}_{jn}\right) \mathbb{I}_{\Omega_j} \\ \geq \exp\left[-n\left(\sum_{i=1}^2 \rho_{jn\tau}^i L_{\psi(jn\tau)}^i(\beta_{jn\tau}^i) + \varepsilon\right)\right] \end{aligned}$$

On en déduit $P(\Omega_{j_0} | \mathcal{F}_0) \geq \exp\left[-n \sum_{j=0}^{j_0-1} \sum_{i=1}^2 \rho_{jn\tau}^i L_{\psi(jn\tau)}^i(\beta_{jn\tau}^i) - nj_0\varepsilon\right]$ P-ps sur Ω_0 , pour $j_0 n\tau \leq T$ et $n \geq \sup N_j$.

On se donne $S < T$, on suppose $3\varepsilon_2 \leq \varepsilon$, $n\tau \leq T - S$ et on choisit j_0 tel que

$j_0 n \tau \in [S, T]$. Alors Ω_{j_0} est P-ps inclus dans $\{d_{0,S}(\gamma^\tau, \psi) \leq \varepsilon\}$. Cela permet d'obtenir

$$P(d_{0,S}(\gamma^\tau, \psi) \leq \varepsilon | \mathcal{F}_0) \geq \exp \left[-\frac{1}{\tau} \left(n\tau \sum_{j=0}^{j_0-1} \sum_{i=1}^2 \rho_{jn\tau}^i L_{\psi(jn\tau)}^i(\beta_{jn\tau}^i) + T\varepsilon \right) \right]$$

Les applications $t \mapsto \rho_t^i$ et $t \mapsto \beta_t^i$ sont continues et $(\psi(\cdot), \beta^i)$ est à valeurs dans le domaine où $L^i(\cdot)$ est continue, donc il existe K_2 tel que, pour $n\tau \leq K_2$, on ait :

$$\left| n\tau \sum_{j=0}^{j_0-1} \sum_{i=1}^2 \rho_{jn\tau}^i L_{\psi(jn\tau)}^i(\beta_{jn\tau}^i) - \int_0^S \sum_{i=1}^2 \rho_t^i L_{\psi(t)}^i(\beta_t^i) dt \right| \leq \varepsilon.$$

On en déduit que, P-ps sur Ω_0 , on a :

$$\tau \log P(d_{0,S}(\gamma^\tau, \psi) \leq \varepsilon | \mathcal{F}_0) \geq - \int_0^S \sum_{i=1}^2 \rho_t^i L_{\psi(t)}^i(\beta_t^i) dt - (T+1)\varepsilon,$$

inégalité valable pour tout $n\tau \leq K \wedge K_1 \wedge K_2 \wedge (T-S)$ et $n \geq \sup N_j$, donc pour τ assez petit une fois n correctement choisi. En prenant S tel que $(T-S)M \leq \varepsilon$, on obtient, sous les mêmes conditions :

$$\tau \log P(d_{0,T}(\gamma^\tau, \psi) \leq 2\varepsilon | \mathcal{F}_0) \geq - \int_0^T \sum_{i=1}^2 \rho_t^i L_{\psi(t)}^i(\beta_t^i) dt - (T+1)\varepsilon,$$

ce qui démontre le théorème 3.2.2 quand ψ est de classe C^1 sur $[0, T]$. Si ψ est seulement C^1 par morceaux, on applique le résultat précédent sur les différents sous-intervalles où ψ est de classe C^1 en choisissant convenablement les rayons (cf le passage de (3.35) au résultat du théorème 3.3.3 dans la démonstration de ce théorème). $\square$

Remarquons que l'utilisation du lemme 3.1.7 permet d'énoncer un théorème analogue en présence d'un seul champ de probabilités. Rappelons qu'un chemin à vitesses intérieures par rapport au champ (μ_x^i) est (définition 1.1.9) un élément ψ de $C[0, T]$, C^1 par morceaux, tel que pour tout t de $[0, T[$, la dérivée à droite de ψ en t appartient à $S_{\psi(t)}^{\circ}$ (à rapprocher de β_t^i appartient à $S_{\psi(t)}^{\circ}$ dans la définition d'un chemin régulier 3.2.1).

En itérant (de la même façon que dans la démonstration précédente) le lemme 3.1.7, on retrouve le résultat suivant, correspondant au théorème 7.3 de [AR] sous des hypothèses légèrement différentes (et contenu également dans [DE]) :

Théorème 3.2.3

On considère un processus d'apprentissage lent, $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à un champ de probabilités (μ_x^i) vérifiant les hypothèses (Hmin) et ψ un chemin de $C[0, T]$ à vitesses intérieures par rapport au champ (μ_x^i) . Alors, pour tout $\varepsilon > 0$, il existe α, τ_0 tels que, pour $\tau < \tau_0$, on ait :

$$\tau \log P(d_{0,T}(\gamma^\tau, \psi) \leq \varepsilon | \mathcal{F}_0) \geq - \int_0^T L_{\psi(t)}^i(\psi'(t)) dt - \varepsilon$$

P -ps sur $\{Y_0^\tau \in B(\psi(0), \alpha)\}$.

3.3 Minoration pour un tube d'axe un chemin continu, avec un coût approché

3.3.1 Etude des lignes de niveau du coût L , sous la condition (HLip)

Dans le paragraphe précédent, nous avons établi une minoration de la probabilité d'un tube d'axe régulier. Nous allons en déduire dans le paragraphe 3.3.2 une minoration avec des axes plus généraux en les approchant par de tels chemins réguliers. Nous reprenons ainsi la démarche proposée dans [AR], où la minoration est d'abord montrée pour des tubes dont l'axe est un chemin à vitesses intérieures (théorème 3.2.3 ci-dessus), ce qui conduit, dans le cas d'un seul champ, au résultat suivant, où on a posé (définition 1.1.9) :

$$\Lambda_{0,T}^{\text{reg}}(\varphi) = \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \text{ à vitesses intérieures}}} \Lambda_{0,T}(\varphi) :$$

pour tout y_0 de $\mathbb{R}^d$ et pour toute partie borélienne G de $C_{y_0}[0, T]$ (chemins continus d'origine en y_0), on a :

$$- \Lambda^{\text{reg}}(\overset{\circ}{G}) \leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi^\tau \in G) \quad (3.28)$$

On fait alors l'hypothèse que les lignes de niveau de la transformée de Cramer du champ de probabilités sont localement lipschitziennes (définition 1.1.7). Cette hypothèse (H') permet, pour tout chemin continu φ et tout ε , d'approcher φ par un chemin à vitesses intérieures ψ avec $\Lambda_{0,T}(\psi)$ plus petit que $\Lambda_{0,T}(\varphi) + \varepsilon$, de la façon

suivante :

pour t dans $[0, T]$, on pose $v_t = h_{\varphi(t), \theta}(\varphi'(t))$ (cf définition 1.2.9) et on obtient un premier chemin ψ_1 par intégration de manière à ce que pour presque tout t de $[0, T]$, $\psi'_1(t)$ soit la projection de v_t sur le compact convexe $\{L_{\psi_1(t)}(\cdot) \leq L_{\varphi(t)}(v_t)\}$. La distance entre ψ_1 et φ est majorée par une quantité proportionnelle à θ grâce à (H') qui implique $|\psi'_1(t) - v_t| \leq C |\psi_1(t) - \varphi(t)|$. On a $\Lambda_{0,T}(\psi_1) \leq \Lambda_{0,T}(\varphi)$. De plus, d'après (1.4), il existe $\delta > 0$ tel que, presque sûrement en t , $\psi'_1(t)$ appartient à $(S_{\psi_1(t)})^{-\delta}$. On peut alors approcher ψ_1 (qui, a priori, n'est pas de classe C^1) par un chemin à vitesses intérieures ψ tel que $\Lambda_{0,T}(\psi) \leq \Lambda_{0,T}(\varphi) + \varepsilon$ (on approche d'abord ψ'_1 par une fonction étagée puis par une fonction constante par morceaux, cf lemme 4.4 de [AR]). On montre ainsi que Λ^{reg} est majoré par Λ (les deux fonctionnelles sont alors égales puisqu'on a $\Lambda \leq \Lambda^{reg}$ d'après la s.c.i. de Λ). On peut donc remplacer Λ^{reg} par Λ dans (3.28). On obtient, sous les hypothèses (H) et (H') :

pour tout y_0 de $\mathbb{R}^d$ et pour toute partie borélienne G de $C_{y_0}[0, T]$, on a :

$$-\Lambda(\overset{\circ}{G}) \leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi^\tau \in G) \quad (3.29)$$

L'apport de l'hypothèse (H') est que, pour x, y dans un compact $\mathcal{K}$ donné et pour tout β de $\mathbb{R}^d$, il existe $\tilde{\beta}$ à une distance de β inférieure à $C_{\mathcal{K}}|x - y|$ tel que $L_y(\tilde{\beta}) \leq L_x(\beta)$. Pour pouvoir reprendre ce type d'approche dans le cas du mélange de deux champs, on doit disposer d'une propriété analogue vis-à-vis du coût mixte défini en 1.4.4. On cherche donc une propriété de la forme suivante :

pour x, y dans un compact $\mathcal{K}$, pour $(\rho^1, \rho^2, \beta^1, \beta^2)$ tels que

$$\sum_{i=1}^2 \rho^i = 1, \quad \pi\left(\sum_{i=1}^2 \rho^i \beta^i\right) = 0, \quad (-1)^i \pi(\beta^i) \leq 0,$$

on peut trouver $(\tilde{\rho}^1, \tilde{\rho}^2, \tilde{\beta}^1, \tilde{\beta}^2)$ tel que

$$\sum_{i=1}^2 \tilde{\rho}^i = 1, \quad \pi\left(\sum_{i=1}^2 \tilde{\rho}^i \tilde{\beta}^i\right) = 0, \quad (-1)^i \pi(\tilde{\beta}^i) \leq 0, \quad L_y^i(\tilde{\beta}^i) \leq L_x^i(\beta^i)$$

et $\left| \sum_{i=1}^2 \tilde{\rho}^i \tilde{\beta}^i - \sum_{i=1}^2 \rho^i \beta^i \right| \leq C'_{\mathcal{K}} |x - y|$, où $C'_{\mathcal{K}}$ ne dépend que du compact $\mathcal{K}$.

Il faut au minimum supposer que (HLip) (correspondant à (H')) pour chacun des deux champs) est vérifiée mais cela ne suffit pas. En effet, cette propriété est réalisée en $(\rho^1, \rho^2, \beta^1, \beta^2)$ de façon certaine seulement quand on a $(-1)^i \pi(\beta^i) < 0$; cette

condition permet en effet d'effectuer séparément une modification de chaque β^i , et de prendre l'unique barycentre des deux nouveaux vecteurs qui est parallèle à 0, barycentre voisin de $\sum_{i=1}^2 \rho^i \beta^i$ quand $\pi(\beta^1)$ et $\pi(\beta^2)$ ne sont pas trop proches de 0. Nous le traduisons dans la proposition suivante :

Proposition 3.3.1

On considère des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant (HLip). Soit $\theta_0 > 0$. Alors, pour tous x et y de $\mathcal{K}$ on a : pour tout $(\rho^1, \rho^2, \beta^1, \beta^2)$ tel que $\sum_{i=1}^2 \rho^i = 1$, $\pi\left(\sum_{i=1}^2 \rho^i \beta^i\right) = 0$ et $(-1)^i \pi(\beta^i) \leq -\theta_0$, il existe $(\tilde{\rho}^1, \tilde{\rho}^2, \tilde{\beta}^1, \tilde{\beta}^2)$ tel que $\sum_{i=1}^2 \tilde{\rho}^i = 1$, $\pi\left(\sum_{i=1}^2 \tilde{\rho}^i \tilde{\beta}^i\right) = 0$, $(-1)^i \pi(\tilde{\beta}^i) \leq 0$, $L_y^i(\tilde{\beta}^i) \leq L_x^i(\beta^i)$ et

$$\left| \sum_{i=1}^2 \tilde{\rho}^i \tilde{\beta}^i - \sum_{i=1}^2 \rho^i \beta^i \right| \leq \left(1 + \frac{M}{\theta_0}\right) C_K |x - y|$$

Démonstration : supposons $(-1)^i \pi(\beta^i) \leq -\theta_0$. Il existe $\tilde{\beta}^i$ tel que $|\tilde{\beta}^i - \beta^i| \leq C_K |x - y|$ et $L_y^i(\tilde{\beta}^i) \leq L_x^i(\beta^i)$. Pour x et y assez proches, on a $(-1)^i \pi(\tilde{\beta}^i) < 0$. Il existe alors $\tilde{\rho}^1$ et $\tilde{\rho}^2$ uniques tels que $\sum_{i=1}^2 \tilde{\rho}^i = 1$ et $\pi(\sum_{i=1}^2 \tilde{\rho}^i \tilde{\beta}^i) = 0$. La quantité $\tilde{\rho}^1 - \rho^1$ est égale à

$$\tilde{\rho}^1 + \frac{\pi(\beta^2)}{\pi(\beta^1) - \pi(\beta^2)} = \frac{\tilde{\rho}^1 \pi(\beta^1) + \tilde{\rho}^2 \pi(\beta^2)}{\pi(\beta^1) - \pi(\beta^2)}.$$

D'après $\pi(\sum_{i=1}^2 \tilde{\rho}^i \tilde{\beta}^i) = 0$, on a donc :

$$\tilde{\rho}^1 - \rho^1 = \frac{\tilde{\rho}^1 [\pi(\beta^1) - \pi(\tilde{\beta}^1)] + \tilde{\rho}^2 [\pi(\beta^2) - \pi(\tilde{\beta}^2)]}{\pi(\beta^1) - \pi(\beta^2)}$$

et enfin, puisque $\tilde{\rho}^1 - \rho^1 = \tilde{\rho}^2 - \rho^2$, pour $i = 1$ ou 2 :

$$|\tilde{\rho}^i - \rho^i| \leq \frac{\max_i |\tilde{\beta}^i - \beta^i|}{\pi(\beta^1) - \pi(\beta^2)}. \quad (3.30)$$

La quantité $\pi(\beta^1) - \pi(\beta^2)$ est minorée par $2\theta_0$, donc on a $|\tilde{\rho}^i - \rho^i| \leq \frac{C_K}{2\theta_0} |x - y|$ et

$$\left| \sum_{i=1}^2 \tilde{\rho}^i \tilde{\beta}^i - \sum_{i=1}^2 \rho^i \beta^i \right| \leq \sum_{i=1}^2 \tilde{\rho}^i |\tilde{\beta}^i - \beta^i| + \sum_{i=1}^2 |\beta^i| |\tilde{\rho}^i - \rho^i| \leq \left(1 + \frac{M}{\theta_0}\right) C_K |x - y|. \quad \square$$

Nous nous contentons des quadruplets vérifiant les conditions de la proposition 3.3.1 pour poursuivre : la minoration pour les tubes d'axe un chemin continu va

faire intervenir un coût approché $J_{0,T}^{\theta_0}(\varphi)$ correspondant au fait qu'on se limite aux décompositions de la forme $\varphi'(t) = \sum_{i=1}^2 \rho^i \beta^i$ avec $(\rho^1, \rho^2, \beta^1, \beta^2)$ vérifiant ces conditions. On passe ensuite à la limite (quand θ_0 tend vers 0), comme pour la majoration. Dans le cas de champs de probabilités constants, on ne se limite pas à ces quadruplets ; la minoration est alors obtenue avec $\Lambda_{0,T}(\varphi)$ (paragraphe 3.7).

3.3.2 Application aux chemins de $C^f[0, T]$

Rappelons qu'on a noté (notation 2.2.1) $C^f[0, T]$ l'ensemble des applications ψ continues pour lesquelles il existe une partition finie $0 = t_0 \leq t_1 \leq \dots \leq t_p = T$ telle que :

$$\psi(]t_j, t_{j+1}[) \subset P^{i_j}, \text{ avec } i_j = 0, 1 \text{ ou } 2. \quad (3.31)$$

Pour θ_0 positif, on reprend la définition 2.2.5 du coût approché, pour $h = 0$ et $\theta = -\theta_0$; pour tout chemin continu, φ , de $[0, T]$ dans $\mathbb{R}^d$, on a :

$$\Lambda_{0,T}^{-\theta_0,0}(\varphi) = \begin{cases} \int_0^T \lambda_{-\theta_0,0}(\varphi(t), \varphi'(t)) dt & \text{si } \varphi \text{ est absolument continu} \\ +\infty & \text{sinon} \end{cases}$$

avec, pour x et β dans $\mathbb{R}^d$:

$$\lambda_{-\theta_0,0}(x, \beta) = \begin{cases} L_x^i(\beta) & \text{si } x \in P^i, i = 1 \text{ ou } 2 \\ L_{-\theta_0}^0(x, \beta) & \text{si } |\pi(\beta)| \leq \theta_0 \\ L_x^i(\beta) & \text{si } |\pi(\beta)| > \theta_0 \end{cases} \quad \text{si } x \in P^0$$

où $L_{-\theta_0}^0(x, \beta)$ est égal à

$$\inf \left\{ \sum_{i=1}^2 \rho^i L_x^i(\beta^i) \mid \sum_{i=1}^2 \rho^i = 1, \sum_{i=1}^2 \rho^i \beta^i = \beta, \pi(\beta^1) \geq \theta_0, \pi(\beta^2) \leq -\theta_0 \right\}.$$

Notation 3.3.2 Pour θ_0 positif, pour simplifier, on remplace, dans ce paragraphe, la notation $\Lambda_{0,T}^{-\theta_0,0}$ par $J_{0,T}^{\theta_0}$.

Soit φ un chemin de $C^f[0, T]$, absolument continu. Pour $\varphi(t)$ dans P^0 , ps en t , $\pi(\varphi'(t))$ est nul. Donc $J_{0,T}^{\theta_0}(\varphi)$ est égal à l'intégrale de 0 à T de l'application

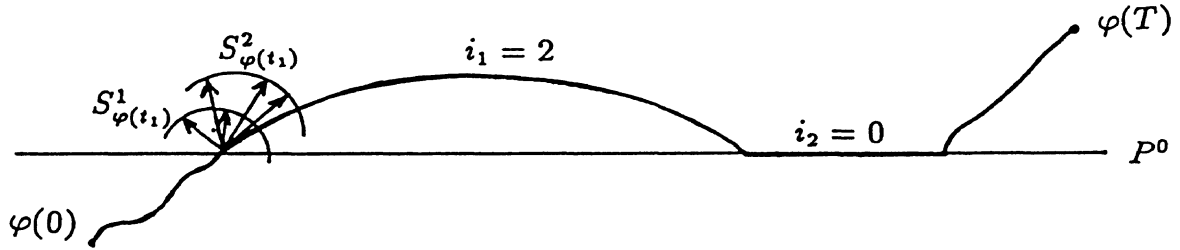
$$t \mapsto \begin{cases} L_{\varphi(t)}^i(\varphi'(t)) & \text{si } \varphi(t) \in P^i \\ L_{-\theta_0}^0(\varphi(t), \varphi'(t)) & \text{si } \varphi(t) \in P^0 \end{cases}$$

Théorème 3.3.3

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses $(Hmin)$ et $(HLip)$. Soit φ un chemin de $C^f[0, T]$. On lui associe $\{t_j, i_j\}$ comme dans (3.31). On suppose que, quand on a $i_j \neq 0$ avec $\varphi(t_j) \in P^0$, $S_{\varphi(t_j)}^1$ et $S_{\varphi(t_j)}^2$ contiennent des vecteurs strictement rentrants dans P^{i_j} .

Alors, pour tous $\theta_0, \varepsilon > 0$, il existe α, τ_0 tels que, pour tout $\tau \leq \tau_0$ et tout y_0 de $B(\varphi(0), \alpha)$, on ait :

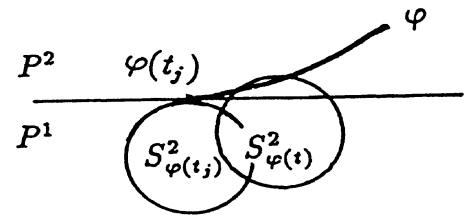
$$\tau \log P_{y_0}(d_{0,T}(\gamma^\tau, \varphi) \leq \varepsilon) \geq -J_{0,T}^{\theta_0}(\varphi) - \varepsilon$$



Remarques :

1. le cas $i_j \neq 0$ et $\varphi(t_j) \notin P^0$ ne se produit éventuellement que pour $j = 0$. Si $i_j \neq 0$ avec $\varphi(t_j) \in P^0$, dans la plupart des cas, l'hypothèse sur $S_{\varphi(t_j)}^1$ et $S_{\varphi(t_j)}^2$ est peu contraignante. Supposons par exemple i_j égal à 2. Puisque $\varphi([t_j, t_{j+1}[)$ est inclus dans P^2 , l'hypothèse sur $S_{\varphi(t_j)}^2$ est en général vérifiée, sauf dans des cas particuliers, par exemple :

$\sup\{\pi(\beta) | \beta \in S_{\varphi(t)}^2\}$ est strictement positif pour $t > t_j$, mais s'annule en $t = t_j$.

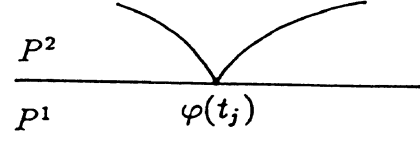


Pour l'hypothèse sur $S_{\varphi(t_j)}^1$ (pour $j \neq 0$) :

— si $i_{j-1} = 0$, alors, si $J_{t_{j-1}, t_j}^{\theta_0}(\varphi)$ vaut $+\infty$, la minoration est triviale ; sinon, d'après la continuité de l'application $x \mapsto S_x^1$, on sait que $S_{\varphi(t_j)}^1$ rencontre l'ensemble $\{\pi(\cdot) \geq \theta_0\}$.

— si $i_{j-1} = 1$, $\sup\{\pi(\beta) | \beta \in S_{\varphi(t)}^1\}$ est strictement positif à gauche de t_j ; l'hypothèse exclut le cas où cette quantité s'annule en $t = t_j$.

– si $i_{j-1} = 2$, on ne sait rien sur $S_{\varphi(t_j)}^1$.



Si l'hypothèse n'est pas vérifiée, au voisinage de $\varphi(t_j)$, le processus peut franchir la frontière et rester ensuite dans P^1 . L'hypothèse implique que le processus revient dans P^2 avec une probabilité non nulle.

– de même, si $j = 0$, on ne sait rien sur $S_{\varphi(0)}^1$ (donc l'hypothèse est nécessaire si $\varphi(0)$ appartient à P^0 , à moins de travailler avec y_0 dans $B(\varphi(0), \alpha) \cap P^2$).

2. en fait, si $i_j \neq 0$, l'hypothèse sur $S_{\varphi(t_j)}^{3-i_j}$ est vérifiée dès que $H_0(\varphi(t_j))$ est vraie.
3. $J_{0,T}^{\theta_0}(\varphi)$ est fini seulement si $H_0(\varphi(t))$ est vérifiée pour tout t de $[t_j, t_{j+1}]$ quand i_j vaut 0.
4. il suffit que l'hypothèse (HLip) soit vérifiée dans un voisinage de $\varphi([0, T])$.

Démonstration : la démonstration consiste à approcher sur chaque $[t_j, t_{j+1}]$ le chemin φ par un chemin régulier (si $i_j = 0$) ou à vitesses intérieures par rapport au champ (μ_x^i) (si $i_j = i$) en utilisant le lemme 3.1.9 à droite de t_j tel que $i_j \neq 0$ et $\varphi(t_j) \in P^0$ grâce à l'hypothèse faite dans ce cas. Cette démonstration fait l'objet de la partie 3.6.

3.4 Minoration finale pour un tube : passage à la limite $\theta \rightarrow 0$

3.4.1 Résultat général

Théorème 3.4.1

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses $(Hmin)$ et $(HLip)$.

Soit φ un chemin de $C^f[0, T]$. On lui associe $\{t_j, i_j\}$ comme dans (3.31). On suppose que, quand on a $i_j \neq 0$ avec $\varphi(t_j) \in P^0$, $S_{\varphi(t_j)}^1$ et $S_{\varphi(t_j)}^2$ contiennent des vecteurs strictement rentrants dans P^{i_j} . Alors, pour tout ε , il existe α et τ_0 tels que, pour tout $\tau \leq \tau_0$ et tout y_0 de $B(\varphi(0), \alpha)$, on ait :

$$\tau \log P_{y_0}(d_{0,T}(\gamma^\tau, \varphi) \leq \varepsilon) \geq - \lim_{\substack{\theta \rightarrow 0 \\ \theta > 0}} J_{0,T}^\theta(\varphi) - \varepsilon$$

Dans la suite, on note $\lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi)$ pour $\lim_{\substack{\theta \rightarrow 0 \\ \theta > 0}} J_{0,T}^\theta(\varphi)$.

Démonstration : considérons φ tel que $J_{0,T}^{\theta_0}(\varphi)$ soit fini pour un $\theta_0 > 0$ (sinon rien à montrer). Le coût $L_{-\theta}^0(x, \cdot)$ est un minimum calculé sur un ensemble conditionné par $\pi(\beta^1) \geq \theta$ et $\pi(\beta^2) \leq -\theta$. Ces contraintes sont moins fortes quand θ diminue, alors $L_{-\theta}^0(x, \cdot)$ diminue. On a donc $J_{0,T}^\theta(\varphi) \leq J_{0,T}^{\theta'}(\varphi)$ pour $\theta \leq \theta'$. Considérons $\varepsilon > 0$ et θ_0 tel que $J_{0,T}^{\theta_0}(\varphi)$ soit inférieur à $\lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi) + \varepsilon$. D'après le théorème 3.3.3, il existe α et τ_0 tels que, pour $\tau \leq \tau_0$ et y_0 dans $B(\varphi(0), \alpha)$, on ait :

$$\tau \log P_{y_0}(d_{0,T}(\gamma^\tau, \varphi) \leq \varepsilon) \geq -J_{0,T}^{\theta_0}(\varphi) - \varepsilon \geq - \lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi) - 2\varepsilon. \quad \square$$

Dans la suite de cette partie, nous calculons $\lim_{\theta \rightarrow 0} J_{0,T}^{\theta_0}(\varphi)$ (dans la proposition 2.2.6, on a montré, pour tout chemin φ de $C[0, T]$, $\Lambda_{0,T}(\varphi) = \lim_{(\theta, h) \rightarrow (0, 0)} \Lambda_{0,T}^{\theta, h}(\varphi)$ mais la limite portait sur les θ positifs alors qu'ici, on a : $J_{0,T}^\theta(\varphi) = \Lambda_{0,T}^{-\theta, 0}(\varphi)$).

Nous montrons que $\lim_{\theta \rightarrow 0} J_{0,T}^{\theta_0}(\varphi)$ vaut soit $\Lambda_{0,T}(\varphi)$, soit $+\infty$. Nous cherchons ensuite à caractériser les configurations de supports des mesures pour lesquelles le deuxième cas peut se produire.

Théorème 3.4.2

On considère des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (Hmin) et φ un chemin de $C^1[0, T]$. On suppose qu'il existe θ_0 tel que $J_{0,T}^{\theta_0}(\varphi)$ soit fini. Alors on a :

$$\lim_{\theta \rightarrow 0} J_{0,T}^{\theta}(\varphi) = \Lambda_{0,T}(\varphi)$$

La condition "il existe θ_0 tel que $J_{0,T}^{\theta_0}(\varphi)$ soit fini" est étudiée dans le paragraphe 3.4.2. Remarquons que si, pour tout θ positif, $J_{0,T}^{\theta}(\varphi)$ est égal à $+\infty$, $\lim_{\theta \rightarrow 0} J_{0,T}^{\theta}(\varphi)$ est égal à $+\infty$ et le minorant dans le théorème 3.4.1 vaut $-\infty$. Dans certains cas cependant (champs de probabilités constants dans le paragraphe 3.6, ou exemples du paragraphe 4.1), on pourra améliorer cette minoration en remplaçant $\lim_{\theta \rightarrow 0} J_{0,T}^{\theta}(\varphi)$ par $\Lambda_{0,T}(\varphi)$ dans le théorème 3.4.1.

Démonstration du théorème (on n'a pas besoin de toutes les hypothèses (Hmin)) : les fonctionnelles J^{θ} et Λ coïncident sur les chemins restant dans un des P^i . On peut donc se contenter d'étudier le cas où $\varphi([0, T])$ est inclus dans P^0 .

Notation 3.4.3 On pose, pour x et β dans $\mathbb{R}^d$:

$$L_+(x, \beta) = \inf \left\{ \sum_{i=1}^2 \rho^i L_x^i(\beta^i) \mid \sum_{i=1}^2 \rho^i = 1, \sum_{i=1}^2 \rho^i \beta^i = \beta, (-1)^i \pi(\beta^i) < 0 \right\}.$$

On utilisera cette notation pour β tel que $\pi(\beta)$ vaut 0.

Lemme 3.4.4

Avec les hypothèses du théorème 3.4.2, si de plus φ est tracé sur P^0 , on a :

$$\lim_{\theta \rightarrow 0} J_{0,T}^{\theta}(\varphi) = \int_0^T L_{\varphi(t)}^+(\varphi'(t)) dt$$

Démonstration : on suppose $J_{0,T}^{\theta_0}(\varphi) < +\infty$ et $\varphi([0, T])$ inclus dans P^0 . $J_{0,T}^{\theta}(\varphi)$ s'écrit $\int_0^T f_{\theta}(t) dt$, où $f_{\theta}(t) = L_{-\theta}^0(\varphi(t), \varphi'(t))$. Pour $\theta \leq \theta_0$, on a $f_{\theta} \leq f_{\theta_0}$, avec f_{θ_0} intégrable, donc, d'après le théorème de convergence dominée :

$$\lim_{\theta \rightarrow 0} J_{0,T}^{\theta}(\varphi) = \int_0^T \lim_{\theta \rightarrow 0} f_{\theta}(t) dt,$$

où $\lim_{\theta \rightarrow 0} f_{\theta}$ désigne la limite simple (avec $\theta > 0$). Or, $L_{\varphi(t)}^+(\varphi'(t))$ est un minimum calculé sur un ensemble qui est la réunion, pour θ variant dans $]0, +\infty[$, des ensembles

$$\left\{ (\rho^1, \rho^2, \beta^1, \beta^2) \mid \sum_{i=1}^2 \rho^i = 1, \sum_{i=1}^2 \rho^i \beta^i = \varphi'(t), \pi(\beta^1) \geq \theta, \pi(\beta^2) \leq -\theta \right\} \quad (3.32)$$

Pour θ fixé, l'ensemble (3.32) est celui sur lequel on calcule le minimum donnant $f_\theta(t)$. En raison de la monotonie, on a donc, pour tout t :

$$L_{\varphi(t)}^+(\varphi'(t)) = \lim_{\theta \rightarrow 0} f_\theta(t),$$

d'où le lemme. $\square$

Etudions maintenant le lien entre L_x^+ et L_x :

Lemme 3.4.5

Pour tous x et β dans $\mathbb{R}^d$ tels que $L_x^+(\beta)$ soit fini, on a :

$$L_x^+(\beta) = L_x(\beta).$$

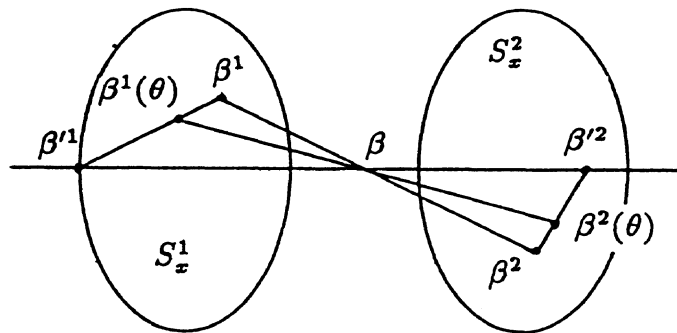
Démonstration: pour x et β quelconques, on a $L_x(\beta) \leq L_x^+(\beta)$. Supposons $L_x^+(\beta)$ fini et considérons (ρ^i, β^i) tel que $\sum_{i=1}^2 \rho^i L_x^i(\beta^i) \leq L_x^+(\beta) + 1$ avec

$$\sum_{i=1}^2 \rho^i = 1, \quad \sum_{i=1}^2 \rho^i \beta^i = \beta \quad \text{et} \quad (-1)^i \pi(\beta^i) < 0 \quad (3.33)$$

$L_x(\beta)$ s'écrit $\sum_{i=1}^2 \rho^i L_x^i(\beta^i)$ avec les mêmes conditions que (3.33) sauf la dernière, qui est remplacée par une inégalité large. Posons, pour chaque i ,

$$\beta^i(\theta) = \frac{\theta \rho^i \beta^i + (1 - \theta) \rho^i \beta^i}{\theta \rho^i + (1 - \theta) \rho^i}$$

Alors on a $(-1)^i \pi(\beta^i(\theta)) < 0$ et $\beta = \sum_{i=1}^2 [\theta \rho^i + (1 - \theta) \rho^i] \beta^i(\theta)$.



On a donc: $L_x^+(\beta) \leq \sum_{i=1}^2 [\theta \rho^i + (1 - \theta) \rho^i] L_x^i(\beta^i(\theta))$.

D'après la convexité des applications L_x^i , on en déduit que, pour tout θ de $]0, 1[$, on a :

$$L_x^+(\beta) \leq \sum_{i=1}^2 (\theta \rho^i L_x^i(\beta^i) + (1 - \theta) \rho'^i L_x^i(\beta'^i)) \leq \theta(L_x^+(\beta) + 1) + (1 - \theta)L_x(\beta).$$

En faisant tendre θ vers 0, on obtient :

$$L_x^+(\beta) \leq L_x(\beta), \text{ d'où le lemme. } \square$$

Pour démontrer le théorème 3.4.2 pour φ tracé sur P^0 , on remarque que dès que $J_{0,T}^{\theta_0}(\varphi)$ est fini, presque sûrement en t , $L_{\varphi(t)}^+(\varphi'(t))$ est fini donc égal à $L_{\varphi(t)}(\varphi'(t))$ d'après le lemme 3.4.5. Alors $\int_0^T L_{\varphi(t)}^+(\varphi'(t)) dt$ est égal à $\Lambda_{0,T}(\varphi)$ et le lemme 3.4.4 permet de conclure. $\square$

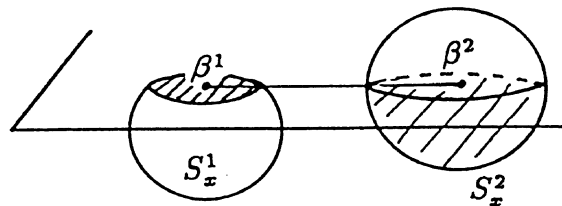
Pour compléter le lemme 3.4.5, nous étudions plus précisément, dans la suite de ce paragraphe, dans quels cas $L_x^+(\beta)$ est infini alors que $L_x(\beta)$ est fini. Pour β tel que $\pi(\beta) = 0$, supposons $L_x(\beta)$ fini égal à $\sum_{i=1}^2 \rho^i L_x^i(\beta^i)$, avec $\beta = \sum_{i=1}^2 \rho^i \beta^i$, $\sum_{i=1}^2 \rho^i = 1$ et $(-1)^i \pi(\beta^i) \leq 0$ et précisons à quelles conditions $L_x^+(\beta)$ peut être fini, donc égal à $L_x(\beta)$.

Pour assurer l'existence des exemples qui suivent, rappelons que chacun des β^i précédents (on a $L_x^i(\beta^i) < +\infty$) peut être sur la frontière de S_x^i quelle que soit sa forme. En effet, considérons une mesure μ sur un ouvert S , un vecteur β_0 sur la frontière de S et, pour p dans $]0, 1[$, posons $\mu = p\delta_{\beta_0} + (1 - p)\mu^1$. Alors, la transformée de Cramer de μ est finie en β_0 (si on note L cette transformée de Cramer, L^1 celle de μ^1 et H^1 la logtransformée de Laplace de μ^1 , on a :

$$L(\beta_0) = \sup_{\alpha} \{ \langle \alpha, \beta_0 \rangle - \ln [p e^{\langle \alpha, \beta_0 \rangle} + (1 - p) e^{H^1(\alpha)}] \} = \ln \frac{1}{p + (1 - p) e^{-L^1(\beta_0)}} \leq -\log p.$$

Etudions alors les différents cas possibles (tous les exemples qui suivent sont observables avec $\sum_{i=1}^2 \rho^i L_x^i(\beta^i) < +\infty$) :

Si $H_0(x)$ n'est pas vérifiée, $L_+(x, \beta)$ vaut $+\infty$.

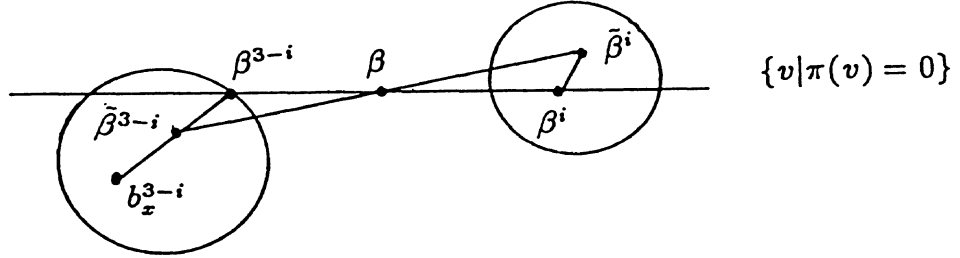


On suppose $H_0(x)$ vérifiée. Si un des $\pi(\beta^i)$ est non nul, l'autre est non nul. $L_x^+(\beta)$ est alors égal à $L_x(\beta)$. Etudions le cas où $\pi(\beta^1) = \pi(\beta^2) = 0$:

1. si, pour un i , β^i appartient à $\overset{\circ}{S}_x^i$ (avec $\rho^i \neq 0$).

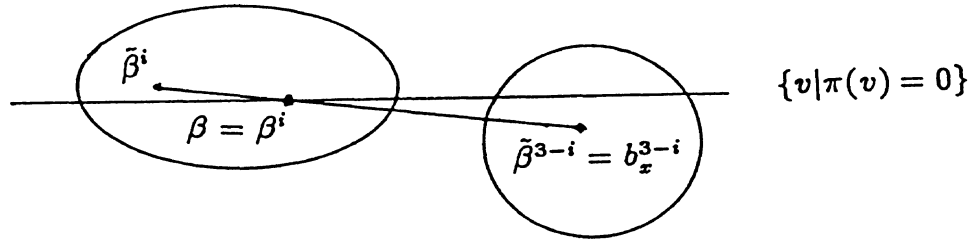
On considère un vecteur b_x^{3-i} dans $S_x^{3-i} \cap \{(-1)^{3-i}\pi(\cdot) < 0\}$ (d'après $H_0(x)$).

- (a) si $\rho^{3-i} \neq 0$



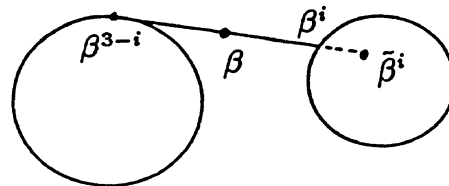
On pose $\tilde{\beta}^{3-i} = \theta b_x^{3-i} + (1 - \theta)\beta^{3-i}$ et $\tilde{\beta}^i = \beta^i - \frac{\theta \rho^{3-i}}{\rho^i}(b_x^{3-i} - \beta^{3-i})$. On a alors $(-1)^i \pi(\tilde{\beta}^i) < 0$, $\beta = \sum_{i=1}^2 \rho^i \tilde{\beta}^i$ et $L_x^i(\tilde{\beta}^i)$ est fini pour θ assez petit. On en déduit $L_+(x, \beta) < +\infty$ et donc $L_+(x, \beta) = L_x(\beta)$.

- (b) si β^{3-i} n'est pas défini ($\rho^{3-i} = 0$).



On pose $\tilde{\beta}^{3-i} = b_x^{3-i}$ et $\tilde{\beta}^i = (1 + \theta)\beta^i - \theta b_x^{3-i}$. Alors on a $(-1)^i \pi(\tilde{\beta}^i) < 0$, et $\beta = \beta^i = \frac{1}{1+\theta}\tilde{\beta}^i + \frac{\theta}{1+\theta}\tilde{\beta}^{3-i}$, avec $L_x^i(\tilde{\beta}^i)$ fini pour θ assez petit, donc $L_+(x, \beta) < +\infty$ et $L_+(x, \beta) = L_x(\beta)$.

2. si chaque β^i est extremal dans S_x^i (pour chaque i tel que ρ^i soit non nul et toujours avec $\pi(\beta^i) = 0$): supposons que ρ^i appartienne à $]0; 1[$ et que la droite $(\beta^1 \beta^2)$ rencontre $\overset{\circ}{S}_x^1$ ou $\overset{\circ}{S}_x^2$:



(la configuration de la figure est possible si $L_x^{3-i}(\beta^{3-i})$ est élevé). Si $(\beta^1 \beta^2)$

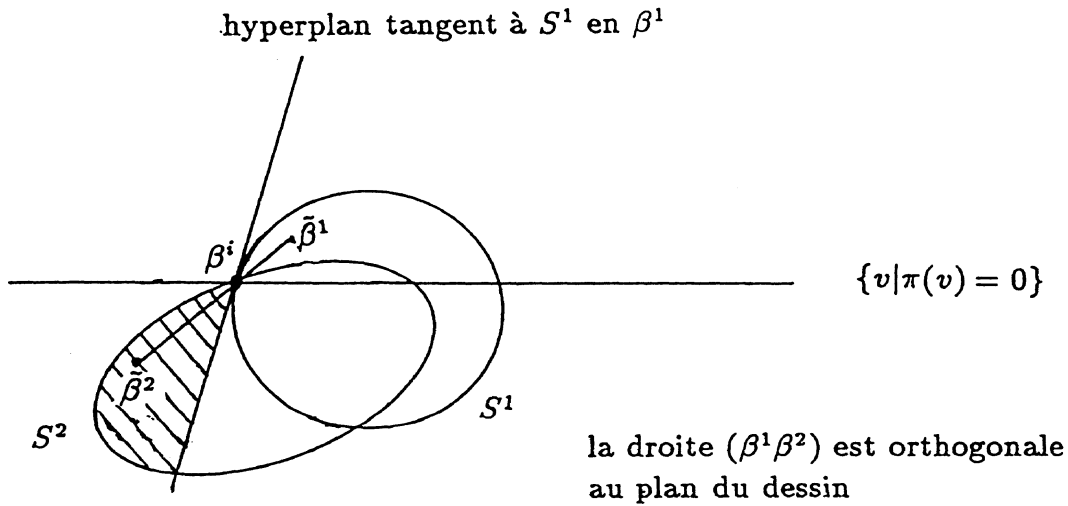
rencontre $\overset{\circ}{S}_x^i$, par convexité, $\overset{\circ}{S}_x^i \cap (\beta^1 \beta^2)$ contient des vecteurs de première coordonnée nulle, arbitrairement proches de β^i . Le vecteur β n'est pas égal à β^i donc on peut prendre $\tilde{\beta}^i$ sur la demi-droite $\beta + \mathbb{R}^+[\beta^i - \beta]$ et dans $\overset{\circ}{S}_x^i$, avec $L_x^i(\tilde{\beta}^i)$ arbitrairement proche de $L_x^i(\beta^i)$. Le vecteur β est barycentre de $\tilde{\beta}^i$ et β^{3-i} avec des coefficients arbitrairement proches de ρ^i et ρ^{3-i} . Comme $\tilde{\beta}^i$ appartient à $\overset{\circ}{S}_x^i$, on peut appliquer la construction de 1.(a). On a donc $L_+(x, \beta) < +\infty$ et donc $L_+(x, \beta) = L_x(\beta)$.

Dans tous les cas précédents, $L_+(x, \beta)$ est égal à $L_x(\beta)$.

En revanche, si ρ^i appartient à $]0; 1[$ alors que $(\beta^1 \beta^2)$ ne rencontre ni $\overset{\circ}{S}_x^1$ ni $\overset{\circ}{S}_x^2$, ou si ρ^i est égal à 0 ou 1, on ne peut pas conclure de façon générale. Les deux situations $L_x(\beta) = L_+(x, \beta) < +\infty$ et $L_x(\beta) < L_+(x, \beta) < +\infty$ sont observables :

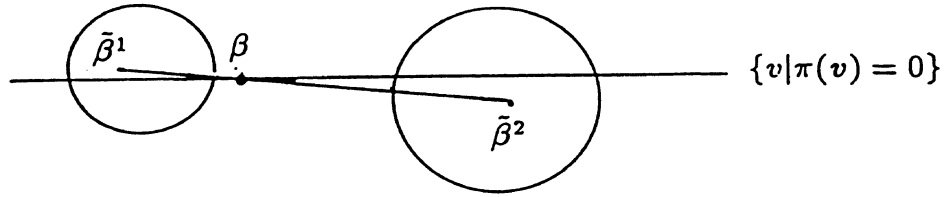
Exemples où $L_x(\beta) = L_+(x, \beta)$:

•



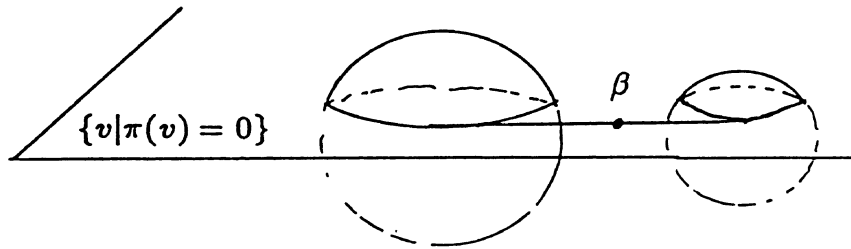
S_x^1 admet un hyperplan tangent en β^1 et $S_x^2 \cap \{\pi < 0\}$ contient un vecteur de l'autre côté de cet hyperplan par rapport à S_x^1 (nécessairement cet hyperplan contient β^1 et β^2) : alors, on trouve un vecteur $\tilde{\beta}^2$ dans $\overset{\circ}{S}_x^2 \cap \{\pi < 0\}$ arbitrairement proche de β^2 . La droite $(\tilde{\beta}^2 \beta)$ contient des points de $\overset{\circ}{S}_x^1 \cap \{\pi > 0\}$ arbitrairement proches de β^1 . On conclut comme dans le cas 2. ci-dessus où $(\beta^1 \beta^2)$ rencontre un des $\overset{\circ}{S}_x^i$.

- ρ^1 vaut 1 et il existe une droite contenant β et rencontrant $S_x^1 \cap \{\pi > 0\}$ et $S_x^2 \cap \{\pi < 0\}$:

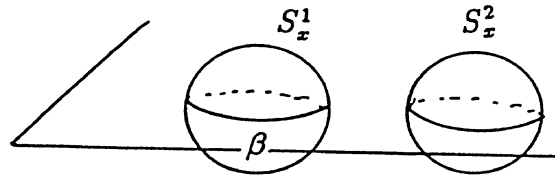


Exemples où $L_x(\beta) < L_+(x, \beta) = +\infty$:

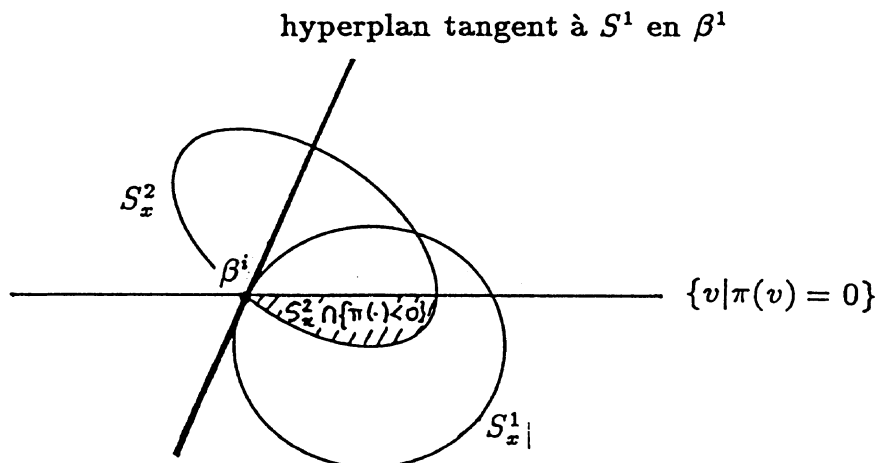
- β s'écrit $\sum_{i=1}^2 \rho^i \beta^i$ avec $L_x(\beta) = \sum_{i=1}^2 \rho^i L_x(\beta^i) < +\infty$ tandis que $L_+(x, \beta)$ vaut $+\infty$:



éventuellement avec $\rho^1 = 1$:



- S_x^1 admet un hyperplan tangent en β^1 et $S_x^2 \cap \{\pi < 0\}$ est du même côté que S_x^1 par rapport à cet hyperplan.



3.4.2 Cas où $\lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi) = \Lambda_{0,T}(\varphi)$

Soit φ un chemin continu tracé sur P^0 . On a $\Lambda_{0,T}(\varphi) \leq \lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi)$ et il y a égalité dès qu'il existe θ_0 tel que $J_{0,T}^{\theta_0}(\varphi)$ soit fini; on sait qu'alors on peut minorer avec $\Lambda_{0,T}(\varphi)$ dans le théorème 3.4.1. Etudions dans quels cas cette condition est réalisée. D'après la monotonie en θ cela correspond aux cas où $\lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi)$ est fini.

Nécessairement, l'hypothèse $H_0(\varphi(t))$ doit être vérifiée presque sûrement pour t dans $[0, T]$. D'autre part, le coût $\Lambda_{0,T}(\varphi)$ doit être fini. En particulier, dans ce cas, presque sûrement en t , il existe $(\rho_t^1, \rho_t^2, \beta_t^1, \beta_t^2)$ dans $[0, 1]^2 \times S_{\varphi(t)}^1 \times S_{\varphi(t)}^2$ tel que

$$\sum_{i=1}^2 \rho_t^i \beta_t^i = \varphi'(t), \quad \sum_{i=1}^2 \rho_t^i = 1 \quad \text{et} \quad (-1)^i \pi(\beta_t^i) \leq 0 \quad (3.34)$$

(en effet, presque sûrement en t , $L_{\varphi(t)}(\varphi'(t))$ s'écrit $\sum_{i=1}^2 \rho_t^i L_{\varphi(t)}^i(\beta_t^i)$, avec ces conditions).

Lemme 3.4.6

Soit φ un chemin absolument continu tracé sur P^0 tel que, presque partout en t , $\varphi'(t)$ s'écrive $\sum_{i=1}^2 \rho_t^i \beta_t^i$, avec $(\rho_t^1, \rho_t^2, \beta_t^1, \beta_t^2)$ dans $[0, 1]^2 \times S_{\varphi(t)}^1 \times S_{\varphi(t)}^2$, $\sum_{i=1}^2 \rho_t^i = 1$ et $(-1)^i \pi(\beta_t^i) \leq 0$.

Dès qu'on peut découper $[0, T]$ en un nombre fini d'intervalles sur lesquels :

- soit il existe θ_0 tel que, ps en t , $(-1)^i \pi(\beta_t^i) \leq -\theta_0$,*
- soit il existe δ_0 tel que, ps en t , $\beta_t^i \in (S_{\varphi(t)}^i)^{-\delta_0}$,*

alors il existe θ_0 tel que $J_{0,T}^{\theta_0}(\varphi) < +\infty$ et donc $\lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi) = \Lambda_{0,T}(\varphi)$.

Remarquons que d'après l'étude conduisant à (1.15) page 48, le cas "il existe θ_0 tel que, ps en t , $(-1)^i \pi(\beta_t^i) \leq -\theta_0$ " n'est pas nécessairement fréquent.

Démonstration : il suffit d'étudier le deuxième cas. Supposons donc que, pour presque tout t de $[0, T]$, β_t^i appartienne à $(S_{\varphi(t)}^i)^{-\delta_0}$. Alors, comme $(-1)^i \pi(\beta_t^i) \leq 0$, $H_0(\varphi(t))$ est vérifiée presque sûrement en t , donc H_0 est vérifiée dans un voisinage de $\varphi([0, T])$. On va modifier (ρ_t^i, β_t^i) de façon à obtenir, pour un θ_0 convenable, indépendant de t , $(\tilde{\rho}_t^i, \tilde{\beta}_t^i)$ tel que

$$\varphi'(t) = \sum_{i=1}^2 \tilde{\rho}_t^i \tilde{\beta}_t^i, \quad \sum_{i=1}^2 \tilde{\rho}_t^i = 1, \quad (-1)^i \pi(\tilde{\beta}_t^i) \leq -\theta_0 \quad \text{et} \quad \int_0^T \sum_{i=1}^2 \tilde{\rho}_t^i L_{\varphi(t)}^i(\tilde{\beta}_t^i) dt < +\infty.$$

Considérons $\theta_0, \rho_0 > 0$.

1. Si $(-1)^i \pi(\beta_t^i) \leq -\theta_0$ pour $i = 1$ ou 2 , on pose $\tilde{\beta}_t^i = \beta_t^i$ et $\tilde{\rho}_t^i = \rho_t^i$.

2. Si pour un i , on a $(-1)^i \pi(\beta_t^i) > -\theta_0$, avec $\rho_t^i \in [\rho_0, 1 - \rho_0]$. L'hypothèse H_0 est vérifiée dans un voisinage de $\varphi([0, T])$. Pour tout x de ce voisinage, on peut donc considérer un vecteur b_x^i défini comme dans le chapitre 2 (choix du vecteur v_x 2.3.3). On a $(-1)^i \pi(b_x^i) \leq -\pi_0$. On pose :

$$\tilde{\beta}_t^{3-i} = \theta b_{\varphi(t)}^{3-i} + (1 - \theta) \beta_t^{3-i} \quad \text{et} \quad \tilde{\beta}_t^i = \beta_t^i - \frac{\theta \rho_t^{3-i}}{\rho_t^i} (b_{\varphi(t)}^{3-i} - \beta_t^{3-i})$$

On a alors $(-1)^{3-i} \pi(\tilde{\beta}_t^{3-i}) \leq -\theta \pi_0$ et

$$\pi(\tilde{\beta}_t^i) = \pi(\beta_t^i) + \frac{\theta}{\rho_t^i} \pi(\rho_t^{3-i} \beta_t^{3-i}) - \frac{\theta \rho_t^{3-i}}{\rho_t^i} \pi(b_{\varphi(t)}^{3-i}) = \pi(\beta_t^i)(1 - \theta) - \frac{\theta \rho_t^{3-i}}{\rho_t^i} \pi(b_{\varphi(t)}^{3-i}).$$

Donc

$$(-1)^i \pi(\tilde{\beta}_t^i) \leq (-1)^{3-i} \frac{\theta \rho_t^{3-i}}{\rho_t^i} \pi(b_{\varphi(t)}^{3-i}) \leq -\theta \rho_0 \pi_0.$$

D'autre part, on a, pour $j = 1$ ou 2 , $|\tilde{\beta}_t^j - \beta_t^j| \leq \frac{2M\theta}{\rho_0} \leq \frac{\delta_0}{2}$ pour θ assez petit, à ρ_0 fixé, et $\varphi'(t)$ est égal à $\sum_{i=1}^2 \rho_t^i \tilde{\beta}_t^i$.

3. Si pour un i , $(-1)^i \pi(\beta_t^i) > -\theta_0$, avec $\rho_t^i \notin [\rho_0, 1 - \rho_0]$; par exemple $\rho_t^1 \leq \rho_0$. Alors $\pi(\beta_t^2)$ appartient à $] -\theta_0, 0]$ (sinon, $\pi(\beta_t^1)$ serait dans $[0, \theta_0[$; on aurait $-\pi(\beta_t^2) = \frac{\rho_t^1 \pi(\beta_t^1)}{\rho_t^2} \leq \frac{\rho_0 \theta_0}{1 - \rho_0} < \theta_0$, donc c'est vrai de toutes façons). On suppose θ inférieur à ρ_0 et on pose

$$\tilde{\beta}_t^1 = b_{\varphi(t)}^1 \quad \text{et} \quad \tilde{\beta}_t^2 = \beta_t^2 + \frac{\theta(\beta_t^2 - b_{\varphi(t)}^1) + \rho_t^1(\beta_t^1 - b_{\varphi(t)}^1)}{\rho_t^2 - \theta}$$

Alors on a $\pi(\tilde{\beta}_t^1) \geq \pi_0$ et, en posant $\tilde{\rho}_t^1 = \rho_t^1 + \theta$ et $\tilde{\rho}_t^2 = \rho_t^2 - \theta$, $\varphi'(t) = \sum_{i=1}^2 \tilde{\rho}_t^i \tilde{\beta}_t^i$. D'autre part, on a

$$\pi(\tilde{\beta}_t^2) = -\frac{\tilde{\rho}_t^1 \pi(\tilde{\beta}_t^1)}{\tilde{\rho}_t^2} \leq -\frac{\tilde{\rho}_t^1 \pi_0}{\tilde{\rho}_t^2} \leq -\tilde{\rho}_t^1 \pi_0 \leq -\theta \pi_0$$

et

$$|\tilde{\beta}_t^2 - \beta_t^2| \leq \frac{2M(\theta + \rho_0)}{1 - \rho_0 - \theta} \leq \frac{\delta_0}{2},$$

pour θ et ρ_0 assez petits.

Prenons alors $\theta_1 = \theta \rho_0 \pi_0 \wedge \theta_0$. Pour chaque i , dans chacun des cas précédents, $(-1)^i \pi(\tilde{\beta}_t^i)$ est inférieur à $-\theta_1$. De plus, presque sûrement en t , $|\tilde{\beta}_t^i - \beta_t^i|$ est inférieur à $\frac{\delta_0}{2}$ (ou bien $\tilde{\beta}_t^i = b_{\varphi(t)}^i$) donc le vecteur $\tilde{\beta}_t^i$ appartient à $(S_{\varphi(t)}^i)^{-\frac{\delta_0}{2}}$.

On a donc :

$$\int_0^T \sum_{i=1}^2 \tilde{\rho}_i^i L_{\varphi(t)}^i(\tilde{\beta}_i^i) dt \leq T J\left(\frac{\delta_0}{2}\right) < +\infty,$$

avec $J(\frac{\delta_0}{2}) = \sup\{L_x^i(v) \mid x \in \varphi([0, T]), v \in (S_x^i)^{-\frac{\delta_0}{2}}, i = 1 \text{ ou } 2\}$. Donc $J_{0,T}^{\theta_1}(\varphi)$ est fini. On conclut en utilisant le théorème 3.4.2. $\square$

Pour certaines configurations des supports au voisinage de la frontière, $\lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi) = \Lambda_{0,T}(\varphi)$ est vrai avec des conditions plus faibles sur φ , parfois même pour tout chemin continu φ :

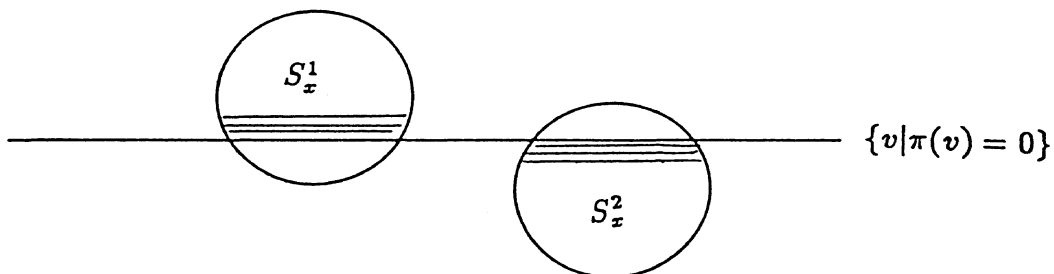
Exemple Si pour tout x d'un compact $\mathcal{K}$, chaque S_x^i est inclus dans $\{(-1)^i \pi(\cdot) < 0\}$: il existe θ_0 tel que, pour y dans $\mathcal{K}$ et $i = 1$ ou 2 , S_y^i soit inclus dans $\{(-1)^i \pi(\cdot) \leq -\theta_0\}$. Alors, pour φ dans $C[0, T]$ à valeurs dans $\mathcal{K}$, $\lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi)$ et $\Lambda_{0,T}(\varphi)$ sont égaux (on revient sur cet exemple dans le chapitre 4).

Notation 3.4.7 Pour $i = 1$ ou 2 et x dans $\mathbb{R}^d$, on note q_x^i l'application qui, à un réel c , associe la projection de $S_x^i \cap \{\pi(\cdot) = -(-1)^i c\}$ sur l'ensemble $\{\pi(\cdot) = 0\}$.

Lemme 3.4.8

Soit φ un chemin absolument continu sur $[0, T]$ à valeurs dans P^0 . Supposons que, pour x dans $\varphi([0, T])$, les applications q_x^1 et q_x^2 soient strictement croissantes, pour l'inclusion, sur un intervalle à droite de 0 et qu'il existe ρ_0 tel que, presque sûrement en t , $\varphi'(t)$ s'écrive $\varphi'(t) = \sum_{i=1}^2 \rho_i^i \beta_i^i$ avec $(\rho_i^1, \rho_i^2, \beta_i^1, \beta_i^2)$ dans $[\rho_0, 1 - \rho_0]^2 \times S_{\varphi(t)}^1 \times S_{\varphi(t)}^2$, $\sum_{i=1}^2 \rho_i^i = 1$ et $(-1)^i \pi(\beta_i^i) \leq 0$. Alors, il existe θ_0 tel que $J_{0,T}^{\theta_0}(\varphi) < +\infty$, et donc $\lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi) = \Lambda_{0,T}(\varphi)$.

Exemple :



Démonstration : Posons, pour $\theta > 0$, $i = 1$ ou 2 et avec $u_1 = (1, 0, \dots, 0)$:

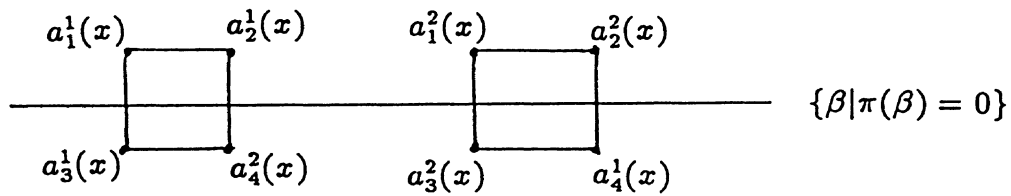
$$\beta_i^i(\theta) = \beta_i^i - \theta(-1)^i \rho_i^{3-i} u_1.$$

D'après l'hypothèse sur les applications q_x^i , il existe $\delta > 0$ tel que, pour θ assez petit, presque sûrement en t , le vecteur $\beta_t^i(\theta)$ appartienne à $(S_{\varphi(t)}^i)^{-\delta}$. Alors, on a :

$$\varphi'(t) = \sum_{i=1}^2 \rho_i^i \beta_t^i(\theta), \quad (-1)^i \pi(\beta_t^i(\theta)) \leq -\theta \rho_i^{3-i} \leq -\theta \rho_0 \quad \text{et} \quad L_{\varphi(t)}^i(\beta_t^i(\theta)) \leq J(\delta)$$

Donc $J_{0,T}^{\theta \rho_0}(\varphi)$ est fini. On conclut en utilisant le théorème 3.4.2. $\square$

Remarquons que ce résultat peut être vrai avec q_x^1 et q_x^2 constantes : par exemple, pour $d = 2$, considérons les probabilités $\mu_x^i = \sum_{k=1}^4 p_k^i(x) \delta_{a_k^i(x)}$, pour $i = 1$ ou 2 , avec la configuration suivante au voisinage de P^0 :



Alors, on peut faire la même modification que dans la démonstration précédente. On n'a plus $\beta_t^i(\theta) \in (S_{\varphi(t)}^i)^{-\delta}$ mais le fait que L_x^i soit bornée sur S_x^i (par $-\log \min_{1 \leq k \leq 4} p_k^i(x)$) assure $J_{0,T}^{\theta \rho_0}(\varphi) < +\infty$.

3.4.3 Conclusion

Les lemmes 3.4.6 et 3.4.8 mettent en évidence des chemins φ pour lesquels le théorème 3.4.1 peut s'écrire avec $\Lambda_{0,T}(\varphi)$ au lieu de $\lim_{\theta \rightarrow 0} J_{0,T}^{\theta}(\varphi)$ (les deux quantités sont égales). En rassemblant ces résultats, on obtient le théorème suivant pour les tubes :

Théorème 3.4.9

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses $(Hmin)$ et $(HLip)$. Soit φ un chemin de $C^f[0, T]$ et $\{t_j, i_j\}$ tels que $\psi([t_j, t_{j+1}[) \subset P^{i_j}$. On suppose :

- si $i_j \neq 0$ avec $\varphi(t_j) \in P^0$, alors $S_{\varphi(t_j)}^1$ et $S_{\varphi(t_j)}^2$ contiennent des vecteurs strictement rentrants dans P^{i_j} .
- si $i_j = 0$ et que φ est absolument continu sur $[t_j, t_{j+1}]$, presque partout en t , $\varphi'(t)$ s'écrit $\sum_{i=1}^2 \rho_i^t \beta_i^t$, avec $(\rho_i^1, \rho_i^2, \beta_i^1, \beta_i^2)$ dans $[0, 1]^2 \times S_{\varphi(t)}^1 \times S_{\varphi(t)}^2$, $\sum_{i=1}^2 \rho_i^t = 1$, $(-1)^i \pi(\beta_i^t) \leq 0$ et on peut découper $[t_j, t_{j+1}]$ en un nombre fini d'intervalles sur lesquels :
 - soit il existe θ_0 tel que, ps en t , $(-1)^i \pi(\beta_i^t) \leq -\theta_0$,
 - soit il existe δ_0 tel que, ps en t , $\beta_i^t \in (S_{\varphi(t)}^i)^{-\delta_0}$,
 - soit les applications q_x^1 et q_x^2 sont, pour tout x de $\varphi([0, T])$, strictement croissantes sur un intervalle à droite de 0 et il existe ρ_0 strictement positif tel que, presque sûrement en t , ρ_i^t appartient à $[\rho_0, 1 - \rho_0]$.

Alors, pour tout ε , il existe α et τ_0 tels que, pour tout $\tau \leq \tau_0$ et tout y_0 de $B(\varphi(0), \alpha)$, on ait :

$$\tau \log P_{y_0}(d_{0,T}(\gamma^\tau, \varphi) \leq \varepsilon) \geq -\Lambda_{0,T}(\varphi) - \varepsilon$$

Exemple: ce résultat est vérifié pour ψ comme dans la remarque qui suit le théorème 3.2.2: dérivable et telle que pour tout t de $[0, T]$, $\psi'(t)$ s'écrit $\sum_{i=1}^2 \rho_i^t \beta_i^t$, avec, pour chaque i , $t \mapsto \beta_i^t$ et $t \mapsto \rho_i^t$ continues par morceaux, β_i^t appartenant à $S_{\psi(t)}^i$ et $(-1)^i \pi(\beta_i^t) \leq 0$. En effet, alors, il existe δ_0 tel que, ps en t , β_i^t appartient à $(S_{\psi(t)}^i)^{-\delta_0}$.

3.5 Minoration pour un borélien quelconque

On a défini (notation 2.2.1) $C^f[0, T]$ comme l'ensemble des chemins ψ continus pour lesquels il existe une partition de $[0, T]$ en un nombre fini de sous-intervalles à l'intérieur desquels ψ reste dans un même P^i .

Définition 3.5.1 Pour y_0 dans $\mathbb{R}^d$, on appelle $C_{y_0}[0, T]$ l'ensemble des chemins continus $\varphi : [0, T] \mapsto \mathbb{R}^d$ tels que $\varphi(0) = y_0$. On munit cet ensemble de la topologie

de la convergence uniforme.

On veut établir une relation du type $-\Lambda_{\min}(\mathring{G}) \leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi^\tau \in G)$ pour toute partie borélienne G de $C_{y_0}[0, T]$. On a obtenu des résultats de minoration d'un tube quand son axe est dans $C^f[0, T]$. On ne sait pas approcher un chemin quelconque par un élément de $C^f[0, T]$ en modifiant peu le coût. Le coût de minoration $\Lambda_{\min}$ va donc se lire sur les chemins de $C^f[0, T]$. D'autre part, il est inutile de prendre en compte les chemins φ pour lesquels $J_{0,T}^\theta(\varphi)$ vaut $+\infty$ pour tout θ puisqu'alors $\lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi)$ est infini. Pour les autres chemins, quand ils vérifient les hypothèses du théorème 3.4.1, dans ce théorème, le minorant est égal à $-\Lambda_{0,T}(\varphi) - \varepsilon$.

Notation 3.5.2 On note $C^{f,r}[0, T]$ l'ensemble des chemins φ de $C^f[0, T]$ tels que

- (1) $J_{0,T}^\theta(\varphi)$ est fini pour un $\theta > 0$
- (2) si $\{t_j, i_j\}$ est associé à φ comme dans (3.31) ($\psi([t_j, t_{j+1}[) \subset P^{i_j}$), pour $i_j \neq 0$ et $\varphi(t_j)$ sur P^0 , $S_{\varphi(t_j)}^1$ et $S_{\varphi(t_j)}^2$ contiennent des vecteurs strictement rentrants dans P^{i_j} .

Remarque: la condition (2) est peu contraignante (cf remarques après le théorème 3.3.3) et on a vu dans le paragraphe précédent des conditions pour que (1) soit vérifiée.

On a le résultat final suivant :

Théorème 3.5.3

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses $(H_{\min})$ et $(HLip)$. On appelle φ^τ le chemin aléatoire continu défini par $\varphi^\tau(k\tau) = Y_k^\tau$ pour $0 \leq k\tau < T + \tau$ et φ^τ est affine sur $[k\tau; (k+1)\tau]$.

Alors, pour tout y_0 de $\mathbb{R}^d$ et pour toute partie borélienne G de $C_{y_0}[0, T]$, on a :

$$-\Lambda(\mathring{G} \cap C^{f,r}[0, T]) \leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0,T]}^\tau \in G)$$

Démonstration : on suppose $\Lambda(\mathring{G} \cap C^{f,r}[0, T])$ fini et on considère, pour $\varepsilon > 0$ donné, un chemin φ dans $\mathring{G} \cap C^{f,r}[0, T]$ tel que $\Lambda_{0,T}(\varphi)$ soit inférieur à $\Lambda(\mathring{G} \cap C^{f,r}[0, T]) + \varepsilon$.

Pour $\varepsilon_1 > 0$ assez petit, $\mathring{G}$ contient le tube d'axe φ et rayon ε_1 . D'après le théorème 3.4.1, on a :

$$\liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(d_{0,T}(\gamma^\tau, \varphi) \leq \varepsilon_1) \geq -\Lambda_{0,T}(\varphi)$$

($|\gamma^\tau - \varphi^\tau|$ est P-ps majoré par τM , petit devant ε_1) d'où le résultat. $\square$

Remarquons que la condition (1) ne peut être supprimée dans la définition de $C^{f,r}[0, T]$ (notation 3.5.2) :

nous donnons à la fin du paragraphe 4.1.3 un exemple où $\lim_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi^\tau_{/[0,T]} \in G)$ vaut $-\infty$ alors que $\mathring{G}$ contient un chemin φ vérifiant la condition (2) de la notation 3.5.2 (en fait φ est à valeurs dans P^0) et tel que $\Lambda_{0,T}(\varphi) < +\infty$. Il s'agit d'un exemple dans lequel (H_0) n'est pas vérifiée. Nous n'avons pas d'exemple quand (H_0) est vérifiée au voisinage des $\varphi([0, T])$, $\varphi \in G$. Il est possible que $\Lambda(\mathring{G} \cap C^{f,r}[0, T])$ soit toujours égal à $\Lambda(\mathring{G} \cap C^f[0, T])$ dans ce cas. Nous le montrons pour les champs de probabilités constants dans le paragraphe 3.7.

3.6 Démonstration du théorème 3.3.3

Revenons à la démonstration du théorème 3.3.3 concernant les chemins continus. Il suffit de montrer que, sous les hypothèses (Hmin) et (HLip), pour tout φ de $C[0, T]$ tel que $\varphi([0, T])$ est inclus dans P^i , pour $i = 0, 1$ ou 2 , avec $S^1_{\varphi(0)}$ et $S^2_{\varphi(0)}$ rencontrant $\{(-1)^i \pi(\cdot) > 0\}$ si $i \neq 0$ et $\varphi(0) \in P^0$, on a :

$$\left. \begin{array}{l} \text{pour tous } \theta_0, \varepsilon > 0, \text{ il existe } \alpha, \tau_0 \text{ tels que, pour tout } \tau \leq \tau_0 \text{ et tout } y_0 \\ \text{de } B(\varphi(0), \alpha), \text{ on ait : } \tau \log P_{y_0}(d_{0,T}(\gamma^\tau, \varphi) \leq \varepsilon) \geq -J_{0,T}^{\theta_0}(\varphi) - \varepsilon \end{array} \right\} \quad (3.35)$$

Nous démontrerons (3.35) un peu plus bas. Supposons ce résultat obtenu. Alors, pour φ comme dans le théorème, on l'utilise sur chaque intervalle $[t_j, t_{j+1}]$ avec $0 \leq j \leq p-1$: θ_0 et ε sont fixés ; on pose $\alpha_p = \varepsilon$ et, par récurrence de $j = p-1$ à $j = 0$, on prend α_j et τ_j tels que, pour $\tau \leq \tau_j$, on ait, P-ps sur $\{Y_0^\tau \in B(\varphi(t_j), \alpha_j)\}$:

$$\tau \log P \left(\sup_{t_j \leq t \leq t_{j+1}} |\gamma^\tau(t - t_j) - \varphi(t)| \leq \frac{\alpha_{j+1}}{2} \mid \mathcal{F}_0 \right) \geq -J_{t_j, t_{j+1}}^{\theta_0}(\varphi) - \varepsilon.$$

On utilise ce résultat pour la chaîne $(Y_k^\tau)_{t_j^\tau \leq k\tau \leq t_{j+1}}$ où, pour j et τ fixés, on note t_j^τ l'unique multiple de τ tel que $t_j^\tau - \tau < t_j \leq t_j^\tau$. Pour $\tau \leq \inf_j \tau_j$, on a, P-ps sur $\{\gamma^\tau(t_j^\tau) \in B(\varphi(t_j), \alpha_j)\}$:

$$\tau \log P \left(\sup_{t_j \leq t \leq t_{j+1}} |\gamma^\tau(t_j^\tau + t - t_j) - \varphi(t)| \leq \frac{\alpha_{j+1}}{2} \mid \mathcal{F}_{t_j^\tau} \right) \geq -J_{t_j, t_{j+1}}^{\theta_0}(\varphi) - \varepsilon.$$

On suppose $J_{0,T}^{\theta_0}(\varphi)$ fini sinon il n'y a rien à montrer. Alors, pour presque tout t , on a $|\varphi'(t)| \leq M$, où M est associé à $\mathcal{K} = B(\psi(0), 1)$ et à T comme dans (1.16). Donc, pour $\tau \leq \inf_j(\tau_j \wedge \frac{\alpha_j}{4M})$, pour chaque j , on a, P-ps sur $\{\gamma^\tau(t_j^\tau) \in B(\varphi(t_j), \alpha_j)\}$:

$$\tau \log P\left(\sup_{t_j^\tau \leq t \leq t_{j+1}} |\gamma^\tau(t) - \varphi(t)| \leq \frac{3\alpha_{j+1}}{4} \mid \mathcal{F}_{t_j^\tau}\right) \geq -J_{t_j, t_{j+1}}^{\theta_0}(\varphi) - \varepsilon. \quad (3.36)$$

Pour conclure, on utilise la propriété de Markov en conditionnant aux différents instants t_j^τ . On note E_j l'événement $\{\sup_{t_j^\tau \leq t \leq t_{j+1}} |\gamma^\tau(t) - \varphi(t)| \leq \frac{3\alpha_{j+1}}{4}\}$. On a, pour tout y_0 dans $B(\varphi(0), \alpha_0)$:

$$P_{y_0}(d_{0,T}(\gamma^\tau, \varphi) \leq \varepsilon) \geq P_{y_0}\left(\cup_{0 \leq j \leq p-1} E_j\right) = P_{y_0}(E_0) \prod_{0 \leq j < p-1} P(E_{j+1} | E_j).$$

Pour $\tau \leq \frac{\alpha_{j+1}}{4M}$, E_j est inclus dans $\{\gamma^\tau(t_{j+1}^\tau) \in B(\varphi(t_{j+1}), \alpha_{j+1})\}$, on a donc, d'après (3.36), pour $\tau \leq \inf_j(\tau_j \wedge \frac{\alpha_j}{4M})$:

$$\tau \log P(E_{j+1} | E_j) \geq -J_{t_j, t_{j+1}}^{\theta_0}(\varphi) - \varepsilon.$$

La somme des différents $J_{t_j, t_{j+1}}^{\theta_0}(\varphi)$ est égale à $J_{0,T}^{\theta_0}(\varphi)$; on en déduit le résultat du théorème 3.3.3 pour φ .

Démonstration de (3.35) :

1. Montrons (3.35) quand $\varphi([0, T])$ est inclus dans P^0 :

On suppose $J_{0,T}^{\theta_0}(\varphi)$ fini. Le réel M est associé à $\mathcal{K} = B(\psi(0), 1)$ et à T comme dans (1.16). Alors, $L_{-\theta_0}^0(\varphi(t), \varphi'(t))$ est fini presque sûrement en t . Sa valeur est obtenue comme le minimum, sur un compact, d'une fonction s.c.i. donc presque sûrement en t , il existe $(\rho_t^1, \rho_t^2, \beta_t^1, \beta_t^2)$ dans $[0, 1]^2 \times S_{\varphi(t)}^1 \times S_{\varphi(t)}^2$ tel qu'on ait $L_{-\theta_0}^0(\varphi(t), \varphi'(t)) = \sum_{i=1}^2 \rho_t^i L_{\varphi(t)}^i(\beta_t^i)$, avec $(-1)^i \pi(\beta_t^i) \leq -\theta_0$. Remarquons que les ρ_t^i sont à valeurs dans un compact inclus dans $]0, 1[$.

Soit $\varepsilon > 0$. On souhaite approcher le chemin φ par un chemin régulier (définition 3.2.1). Pour cela, on modifie ρ_t^i et β_t^i de manière à construire des applications $t \mapsto \tilde{\rho}_t^i$ et $t \mapsto \tilde{\beta}_t^i$, C^0 par morceaux, telles que, si on pose

$$\psi(t) = \varphi(0) + \int_0^t \sum_{i=1}^2 \tilde{\rho}_u^i \tilde{\beta}_u^i du$$

alors le chemin ψ vérifie les conditions suivantes :

$$\sup_{t \in [0, T]} (-1)^i \pi(\tilde{\beta}_t^i) < 0 \quad (3.37)$$

$$\inf_{t \in [0, T]} d(\tilde{\beta}_t^i, Fr(S_{\psi(t)}^i)) > 0 \quad (3.38)$$

$$d_{0,T}(\varphi, \psi) \leq \varepsilon \quad (3.39)$$

$$\int_0^T \sum_{i=1}^2 \tilde{\rho}_t^i L_{\psi(t)}^i(\tilde{\beta}_t^i) dt \leq J_{0,T}^{\theta_0}(\varphi) + \varepsilon T. \quad (3.40)$$

Les deux premières conditions permettront, à l'aide du théorème 3.2.2, de minorer la probabilité pour que γ^τ appartienne à un tube d'axe ψ ; on en déduira la minoration pour un tube d'axe φ , faisant intervenir $J_{0,T}^{\theta_0}(\varphi)$, en utilisant les deux dernières (remarquons qu'il n'est pas nécessaire que le coût de ψ soit donné par $\int_0^T \sum_{i=1}^2 \tilde{\rho}_t^i L_{\psi(t)}^i(\tilde{\beta}_t^i) dt$: dans le théorème 3.2.2, on n'a pas supposé $L_{\psi(t)}(\psi'(t)) = \sum_{i=1}^2 \tilde{\rho}_t^i L_{\psi(t)}^i(\tilde{\beta}_t^i)$).

Une modification de ρ_t^i et β_t^i pour un t fixé est facile à obtenir. La difficulté ici vient de ce qu'on veut ramener les vecteurs β_t^i à l'intérieur des supports tout en gardant proche de $\varphi'(t)$ celui de leurs barycentres qui est parallèle à P^0 , avec une uniformité en t .

On considère $\theta > 0$ et on pose $v_t^i = h_{\varphi(t), \theta}^i(\beta_t^i)$ (cf définition 2.1.2). On a $|v_t^i - \beta_t^i| \leq 2M\theta$ et

$$(-1)^i \pi(v_t^i) = \theta(-1)^i \pi(m_{\varphi(t)}^i) + (1 - \theta)(-1)^i \pi(\beta_t^i) \leq -(1 - \theta)\theta_0 + M\theta. \quad (3.41)$$

D'autre part, $v_t^i \in (S_{\varphi(t)}^i)^{-C_1\theta}$, d'après (1.4) du chapitre 1, et $L_{\varphi(t)}^i(v_t^i) \leq L_{\varphi(t)}^i(\beta_t^i)$.

Pour x dans $\mathbb{R}^d$ et $I \geq 0$, on note $p_{x,I}^i$ la projection sur le convexe compact $\{L_x^i(\cdot) \leq I\}$ comme dans [AR] 4.6. On pose $I^i(t) = L_{\varphi(t)}^i(v_t^i)$. Si on a $(-1)^i \pi(p_{x,I^i(t)}^i(v_t^i)) < 0$ pour chaque i , on note $F(t, x)$ l'unique barycentre de $p_{x,I^1(t)}^1(v_t^1)$ et $p_{x,I^2(t)}^2(v_t^2)$ tel que $\pi(F(t, x)) = 0$. Alors, $F(t, x)$ s'écrit $\sum_{i=1}^2 \rho^i(t, x) p_{x,I^i(t)}^i(v_t^i)$, avec $\sum_{i=1}^2 \rho^i(t, x) = 1$.

L'application $(t, x) \mapsto p_{x,I^i(t)}^i(v_t^i)$ est localement lipschitzienne en x , uniformément en $t \in [0, T]$. Sur un intervalle où $(-1)^i \pi(p_{x,I^i(t)}^i(v_t^i))$ est majoré par une constante négative donnée, notée $-C$, la fonction $\rho^i(t, x)$ l'est également (cf (3.30): on a $|\rho^i(t, x) - \rho^i(t, y)| \leq \frac{\max_i |p_{x,I^i(t)}^i(v_t^i) - p_{y,I^i(t)}^i(v_t^i)|}{2C}$); alors $F(t, x)$ est une fonction borélienne du couple (t, x) qui est localement lipschitzienne en x , uniformément en t appartenant à cet intervalle.

On considère un chemin ψ_1 défini par $\psi_1(0) = \varphi(0)$ et $\psi_1'(t) = F(t, \psi_1(t))$ sur un intervalle à droite de 0, tant que $(-1)^i \pi(p_{\psi_1(t), I^i(t)}^i(v_t^i))$ est majoré par une constante

négative. On a sur cet intervalle :

$$\psi_1(t) = \varphi(0) + \int_0^t \sum_{i=1}^2 \rho^i(u, \psi_1(u)) p_{\psi_1(u), I^i(u)}^i(v_u^i) du.$$

On pose $\bar{\rho}_t^i = \rho^i(t, \psi_1(t))$ et $\bar{\beta}_t^i = p_{\psi_1(t), I^i(t)}^i(v_t^i)$ (projection de v_t^i sur le compact $\{L_{\psi_1(t)}^i(\cdot) \leq L_{\varphi(t)}^i(v_t^i)\}$). Alors $\psi_1'(t)$ est égal à $\sum_{i=1}^2 \bar{\rho}_t^i \bar{\beta}_t^i$. D'après l'hypothèse (HLip) sur les lignes de niveau des L_x^i , il existe une constante C_U telle que, presque sûrement en t , on ait $|\bar{\beta}_t^i - v_t^i| \leq C_U |\psi_1(t) - \varphi(t)|$, donc $|\beta_t^i - \bar{\beta}_t^i| \leq 2M\theta + C_U |\psi_1(t) - \varphi(t)|$. On reprend les calculs faits dans la démonstration de la proposition 3.3.1. On a :

$$|\psi_1'(t) - \varphi'(t)| = \left| \sum_{i=1}^2 \bar{\rho}_t^i \bar{\beta}_t^i - \sum_{i=1}^2 \rho_t^i \beta_t^i \right| \leq \sum_{i=1}^2 \bar{\rho}_t^i |\bar{\beta}_t^i - \beta_t^i| + \sum_{i=1}^2 |\beta_t^i| |\bar{\rho}_t^i - \rho_t^i|,$$

avec $|\bar{\rho}_t^i - \rho_t^i| \leq \frac{\max_i |\bar{\beta}_t^i - \beta_t^i|}{\pi(\beta_t^1) - \pi(\beta_t^2)}$ (cf (3.30)) et $(-1)^i \pi(\beta_t^i) \leq -\theta_0$, donc

$$|\psi_1'(t) - \varphi'(t)| \leq \left(1 + \frac{M}{\theta_0}\right) \max_i |\bar{\beta}_t^i - \beta_t^i| \leq \left(1 + \frac{M}{\theta_0}\right) C_U |\psi_1(t) - \varphi(t)| + M'\theta,$$

où $M' = 2M(1 + \frac{M}{\theta_0})$. Par intégration, on en déduit pour t dans $[0, T]$:

$$|\psi_1(t) - \varphi(t)| \leq \left(1 + \frac{M}{\theta_0}\right) C_U \int_0^t |\psi_1(u) - \varphi(u)| du + M\theta'T$$

et, en utilisant le lemme de Gronwall :

$$d_{0,T}(\varphi, \psi_1) \leq M'\theta T \exp\left[\left(1 + \frac{M}{\theta_0}\right) C_U T\right] = C(T)\theta$$

pour $C(T)$ convenable. Pour chaque i , presque sûrement en t , $\bar{\beta}_t^i$ vérifie les propriétés suivantes :

$|\bar{\beta}_t^i - v_t^i| \leq C_U C(T)\theta$, donc, d'après (3.41), pour $(M + \theta_0 + C_U C(T))\theta \leq \frac{\theta_0}{2}$:

$$(-1)^i \pi(\bar{\beta}_t^i) \leq -(1 - \theta)\theta_0 + M\theta + C_U C(T)\theta \leq -\frac{\theta_0}{2} \quad (3.42)$$

et

$$L_{\psi_1(t)}^i(\bar{\beta}_t^i) \leq L_{\varphi(t)}^i(v_t^i) \leq L_{\varphi(t)}^i(\beta_t^i). \quad (3.43)$$

D'autre part, comme on a (cf remarque après la démonstration du théorème 1.2.7 : d'après (HLip), $x \mapsto S_x^i$ est localement lipschitzienne), quitte à augmenter C_U :

$$h(S_{\varphi(t)}^i, S_{\psi_1(t)}^i) < C_U |\psi_1(t) - \varphi(t)| \leq C_U C(T)\theta,$$

de $v_t^i \in (S_{\varphi(t)}^i)^{-C_1\theta}$ et $|\bar{\beta}_t^i - v_t^i| \leq C_U C(T)\theta$, on déduit :

$$\bar{\beta}_t^i \in (S_{\psi_1(t)}^i)^{-\delta} \text{ où } \delta = C_1\theta - 2C_U C(T)\theta.$$

Quand T tend vers 0, $C(T) = M'T \exp[(1 + \frac{M}{\theta_0})C_u T]$ tend vers 0. On prend T assez petit pour avoir

$$2C_u C(T) \leq \frac{C_1}{2}; \quad (3.44)$$

on a alors

$$\bar{\beta}_t^i \in (S_{\psi_1(t)}^i)^{-\frac{C_1\theta}{2}} \quad (3.45)$$

Il reste cependant à rendre $t \mapsto \bar{\beta}_t^i$ et $t \mapsto \bar{\rho}_t^i$ continues par morceaux, en conservant un chemin à vitesses parallèles à P^0 . Soit $\varepsilon_1 > 0$. Pour chaque i , on approche $t \mapsto \bar{\beta}_t^i$ par une fonction étagée puis par une fonction constante par morceaux $t \mapsto \tilde{\beta}_t^i$ telle que ps en t , $|\tilde{\beta}_t^i - \bar{\beta}_t^i| \leq \varepsilon_1 \theta$ (cf la démonstration du lemme 4.4 dans [AR]). On prendra ε_1 petit devant toutes les autres données pour conserver à $\tilde{\beta}_t^i$ les propriétés de $\bar{\beta}_t^i$. Supposons déjà $\varepsilon_1 \theta \leq \frac{\theta_0}{4}$; on a, d'après (3.42), pour presque tout t :

$$(-1)^i \pi(\tilde{\beta}_t^i) \leq (-1)^i \pi(\bar{\beta}_t^i) + \varepsilon_1 \theta \leq -\frac{\theta_0}{4}.$$

La condition (3.37) est donc vérifiée. Il existe alors $\tilde{\rho}_t^1$ et $\tilde{\rho}_t^2$ tels que $\tilde{\rho}_t^1 + \tilde{\rho}_t^2 = 1$ et $\pi\left(\sum_{i=1}^2 \tilde{\rho}_t^i \tilde{\beta}_t^i\right) = 0$. On définit un chemin ψ par :

$$\psi(t) = \varphi(0) + \int_0^t \sum_{i=1}^2 \tilde{\rho}_u^i \tilde{\beta}_u^i du.$$

Alors, pour presque tout t , on a

$$|\psi'(t) - \psi_1'(t)| \leq \sum_{i=1}^2 \tilde{\rho}_t^i |\tilde{\beta}_t^i - \bar{\beta}_t^i| + \sum_{i=1}^2 |\tilde{\rho}_t^i - \bar{\rho}_t^i| M,$$

avec, d'après (3.42) :

$$|\tilde{\rho}_t^i - \bar{\rho}_t^i| \leq \frac{\max_i |\tilde{\beta}_t^i - \bar{\beta}_t^i|}{\pi(\bar{\beta}_t^1) - \pi(\bar{\beta}_t^2)} \leq \frac{\varepsilon_1 \theta}{\theta_0}$$

On en déduit $d_{0,T}(\psi_1, \psi) \leq (1 + \frac{2M}{\theta_0})\varepsilon_1 \theta T$ et donc

$$d_{0,T}(\varphi, \psi) \leq \left(1 + \frac{2M}{\theta_0}\right)\varepsilon_1 \theta T + C(T)\theta \leq 2C(T)\theta,$$

pour ε_1 petit. En prenant θ assez petit, on obtient la condition (3.39) :

$$d_{0,T}(\varphi, \psi) \leq \varepsilon. \quad (3.46)$$

D'après les hypothèses (Hmin), il existe δ_θ tel que les conditions

$$|x - y| \leq \delta_\theta, \quad |v - v'| \leq \delta_\theta, \quad x \in \mathcal{U} \quad \text{et} \quad v \in (S_x^i)^{-\frac{c_1\theta}{2}}$$

impliquent

$$v' \in (S_y^i)^{-\frac{c_1\theta}{4}} \quad \text{et} \quad |L_x^i(v) - L_y^i(v')| \leq \frac{\varepsilon}{2}. \quad (3.47)$$

On prend ε_1 assez petit pour avoir :

$$\varepsilon_1 \theta \left(1 \wedge \left(1 + \frac{2M}{\theta_0} \right) T \right) \leq \delta_\theta.$$

Alors $d_{0,T}(\psi_1, \psi) \leq (1 + \frac{2M}{\theta_0})\varepsilon_1 \theta T$, $|\tilde{\beta}_t^i - \bar{\beta}_t^i| \leq \varepsilon_1 \theta$ et (3.45) impliquent, pour presque tout t :

$$\tilde{\beta}_t^i \in (S_{\psi(t)}^i)^{-\frac{c_1\theta}{4}}, \quad (3.48)$$

ce qui permet de vérifier la condition (3.38), et $|L_{\psi_1(t)}^i(\bar{\beta}_t^i) - L_{\psi(t)}^i(\tilde{\beta}_t^i)| \leq \frac{\varepsilon}{2}$. En utilisant (3.43), on en déduit :

$$\sum_{i=1}^2 \rho_t^i L_{\psi(t)}^i(\tilde{\beta}_t^i) \leq \sum_{i=1}^2 \rho_t^i L_{\varphi(t)}^i(\beta_t^i) + \frac{\varepsilon}{2}.$$

Or, on a

$$|\tilde{\rho}_t^i - \rho_t^i| \leq \frac{\max_i |\tilde{\beta}_t^i - \beta_t^i|}{2\theta_0} \leq \frac{2M\theta + C_u C(T)\theta + \varepsilon_1 \theta}{2\theta_0}$$

et

$$\rho_t^i = \frac{(-1)^i \pi(\beta_t^{3-i})}{\pi(\beta_t^1) - \pi(\beta_t^2)} \geq \frac{\theta_0}{2M},$$

donc

$$|\tilde{\rho}_t^i - \rho_t^i| \leq \rho_t^i \frac{M}{\theta_0^2} \left[2M + C_u C(T) + \varepsilon_1 \right] \theta \leq \rho_t^i M'' \theta, \quad \text{où } M'' \text{ est une constante.}$$

On a $\tilde{\rho}_t^i \leq \rho_t^i (1 + M'' \theta)$, donc $\sum_{i=1}^2 \tilde{\rho}_t^i L_{\psi(t)}^i(\tilde{\beta}_t^i) \leq (1 + M'' \theta) \left[\sum_{i=1}^2 \rho_t^i L_{\varphi(t)}^i(\beta_t^i) + \frac{\varepsilon}{2} \right]$. On en déduit :

$$\int_0^T \sum_{i=1}^2 \tilde{\rho}_t^i L_{\psi(t)}^i(\tilde{\beta}_t^i) dt \leq (1 + M'' \theta) [J_{0,T}^{\theta_0}(\varphi) + \frac{\varepsilon}{2} T] \leq J_{0,T}^{\theta_0}(\varphi) + \varepsilon T, \quad (3.49)$$

pour θ petit, d'où la condition (3.40).

Remarque : le fait de devoir prendre T petit devant certaines constantes du problème (condition (3.44)) n'est pas gênant : en effet, la construction précédente peut être

faite à partir de $\psi(0) = z$ du moment que z est suffisamment proche de $\varphi(0)$. On peut alors construire ψ successivement sur un nombre fini de sous-intervalles de longueur T_1 suffisamment petite. Les $\psi(jT_1)$ successifs sont arbitrairement proches des $\varphi(jT_1)$.

On utilise alors le théorème 3.2.2: il existe α et τ_0 tels que, pour $\tau < \tau_0$, P-ps sur $\{Y_0^\tau \in B(\psi(0), \alpha)\}$ on ait :

$$\tau \log P\left(d_{0,T}(\gamma^\tau, \psi) \leq \varepsilon \mid \mathcal{F}_0\right) \geq - \int_0^T \sum_{i=1}^2 \tilde{\rho}_t^i L_{\psi(t)}^i(\tilde{\beta}_t^i) dt - \varepsilon.$$

On a alors, d'après (3.46) et (3.49), pour $\tau < \tau_0$:

$$\begin{aligned} \tau \log P\left(d_{0,T}(\gamma^\tau, \varphi) \leq 2\varepsilon \mid \mathcal{F}_0\right) &\geq - \int_0^T \sum_{i=1}^2 \tilde{\rho}_t^i L_{\psi(t)}^i(\tilde{\beta}_t^i) dt - \varepsilon \\ &\geq -J_{0,T}^{\theta_0}(\varphi) - \varepsilon(T+1), \end{aligned}$$

P-ps sur $\{Y_0^\tau \in B(\varphi(0), \alpha)\}$, ce qui démontre (3.35) pour φ .

2. Montrons maintenant (3.35) lorsqu'on suppose que $\varphi([0, T[)$ est inclus dans un P^i ($i = 1$ ou 2), et que $S_{\varphi(0)}^1$ et $S_{\varphi(0)}^2$ rencontrent $\{(-1)^i \pi(\cdot) > 0\}$ pour cet i :

on suppose que $\varphi(0)$ appartient à P^0 , sinon le résultat est évident (il suffit de minorer la probabilité d'être dans un tube d'axe φ sur $[0, T']$ avec T' arbitrairement proche de T et de rayon suffisamment petit pour que les chemins du tube soient à valeurs dans P^i ; un seul champ de probabilités est alors concerné).

Supposons par exemple i égal à 2. Alors $J_{0,T}^{\theta_0}(\varphi)$ est égal à $\Lambda_{0,T}^2(\varphi)$. L'hypothèse (HLip) étant vérifiée, pour tout $\varepsilon > 0$, on peut approcher φ par un chemin à vitesses intérieures, ψ , dont le coût soit proche de celui de φ . On souhaite que le chemin ψ soit à valeurs dans P^2 . Pour cela on modifie légèrement la construction faite dans le cas d'un seul champ, rappelée dans le paragraphe 3.3.1 :

d'après la condition sur $S_{\varphi(0)}^2$, pour y dans un voisinage de $\varphi(0)$, on peut trouver un vecteur c_y^2 appartenant à $(S_y^2)^{-\delta_0}$ tel que $\pi(c_y^2)$ soit supérieur à π_0 , où δ_0 et π_0 sont strictement positifs et indépendants de y . Pour t dans $[0, T]$, au lieu de $\theta m_{\varphi(t)}^2 + (1 - \theta)\varphi'(t)$, on prend v_t égal à $\theta c_{\varphi(t)}^2 + (1 - \theta)\varphi'(t)$. On a encore $|v_t - \varphi'(t)| \leq 2M\theta$. D'après le lemme 1.2.10, il existe une constante C_1 telle que, presque sûrement en t , v_t appartient à $(S_{\varphi(t)}^2)^{-C_1\theta}$. D'autre part, on a $L_{\varphi(t)}^2(v_t^2) \leq L_{\varphi(t)}^2(\varphi'(t)) + \theta J(\delta_0) \leq L_{\varphi(t)}^2(\varphi'(t)) + \varepsilon$ pour θ petit.

Comme dans le paragraphe 3.3.1, on obtient un premier chemin ψ_1 par intégration de manière à ce que, presque sûrement en t , $\psi'_1(t)$ soit la projection de v_t sur $\{L_{\psi_1(t)}(\cdot) \leq L_{\varphi(t)}(v_t)\}$; on a $|\psi'_1(t) - \varphi'(t)| \leq C_U |\psi_1(t) - \varphi(t)| + 2M\theta$ donc $d_{0,T}(\varphi, \psi_1) \leq 2M\theta T e^{C_U T}$. Ensuite on approche ψ_1 par un chemin à vitesses intérieures ψ tel que $d_{0,T}(\psi'_1, \psi') \leq \varepsilon_1 \theta$. On prend $\psi(0) = \psi_1(0) = \varphi(0)$. Pour θ et ε_1 petits, la distance entre ψ et φ est inférieure à ε et on a $\Lambda_{0,T}(\psi) \leq \Lambda_{0,T}(\varphi) + 2\varepsilon T$. D'autre part, on a pour tout $t \leq T$:

$$\pi(\psi'(t)) \geq \pi_0 \theta + (1 - \theta) \pi(\varphi'(t)) - |\psi'_1(t) - v_t| - \varepsilon_1 \theta,$$

et $|\psi'_1(t) - v_t|$ est majoré par $C_U d_{0,T}(\varphi, \psi_1)$ donc par $C_U 2M\theta T e^{C_U T}$. On considère alors T_1 tel que $C_U 2MT_1 e^{C_U T_1}$ soit inférieur à $\frac{\pi_0}{3}$ et on suppose ε_1 inférieur à $\frac{\pi_0}{3}$. On construit ψ sur l'intervalle $[0, T_1]$. On a, pour tout t de $[0, T_1]$:

$$\pi(\psi'(t)) \geq \frac{\pi_0}{3} \theta + (1 - \theta) \pi(\varphi'(t))$$

donc

$$\pi(\psi(t)) \geq \frac{\pi_0}{3} \theta t + (1 - \theta) \pi(\varphi(t)) > 0.$$

Donc ψ est à valeurs dans P^2 sur l'intervalle $[0, T_1]$. D'autre part, puisque $|\psi'(t) - v_t|$ est majoré par $C_U 2M\theta T e^{C_U T} + \varepsilon_1 \theta$ et $|\psi(t) - \varphi(t)|$ par $2M\theta T e^{C_U T} + \varepsilon_1 \theta T$, de $v_t \in (S_{\varphi(t)}^2)^{-C_1 \theta}$, on déduit que $\psi'(t)$ appartient à $(S_{\psi(t)}^2)^{-\delta}$ avec:

$$\delta = C_1 \theta - [C_U 2MT_1 e^{C_U T_1} + \varepsilon_1] \theta - C_U [2MT_1 e^{C_U T_1} + \varepsilon_1 T_1] \theta \geq \frac{C_1 \theta}{2},$$

pour ε_1 et T_1 petits devant C_1 . Le chemin ψ est alors un chemin à vitesses intérieures sur l'intervalle $[0, T_1]$.

Pour T' arbitrairement proche de T et ε assez petit, on sait minorer la probabilité que γ^τ soit dans un tube d'axe $\varphi_{/[T_1, T']}$ et rayon ε . Le résultat fait intervenir le coût $\Lambda_{T_1, T'}^2(\varphi) = J_{T_1, T'}^{\theta_0}(\varphi)$. Il suffit donc de montrer (3.35) pour $\varphi_{/[0, T_1]}$ et quitte à remplacer φ par ψ ci-dessus, on peut supposer que φ est un chemin à vitesses intérieures à valeurs dans P^2 sur $]0, T]$. On prend alors δ tel que presque sûrement en t , $\varphi'(t)$ appartient à $(S_{\varphi(t)}^2)^{-\delta}$. On veut minorer la probabilité d'un ensemble de trajectoires, pour cela, on va tenir compte seulement des trajectoires qui sont dans P^2 après une certaine durée $n\tau$, petite devant T . Par hypothèse, il existe c^1 dans $S_{\varphi(0)}^1$ et c^2 dans $S_{\varphi(0)}^2$ rentrants dans P^2 , c'est-à-dire tels que $\pi(c^i) > 0$ pour chaque i .

Soit ε dans $]0, \frac{\pi(c^1) \wedge \pi(c^2)}{3}]$. On prend ε_1 , K et N comme dans le lemme 3.1.9, avec K tel que $KL_{\varphi(0)}^1(c^1) \vee L_{\varphi(0)}^2(c^2) < \varepsilon$ et $K \leq 1$; on considère $K_1 \leq K$ et α inférieur à

$\frac{K_1}{3}\pi(c^1) \wedge \pi(c^2)$ et à ε_1 . Pour $n\tau \leq K$ et $n \geq N$, on a :

$$\begin{aligned} \tau \log P\left(\frac{Y_n^\tau - Y_0^\tau}{n\tau} \in \bigcup_{\beta \in A} B(\beta, \varepsilon) \mid Y_0^\tau \in B(\varphi(0), \alpha)\right) \\ \geq -n\tau L_{\varphi(0)}^1(c^1) \vee L_{\varphi(0)}^2(c^2) - n\tau\varepsilon \geq -\varepsilon(1 + K) \geq -2\varepsilon \end{aligned}$$

Or, pour $K_1 \leq n\tau$, comme ε est inférieur à $\frac{\pi(c^1) \wedge \pi(c^2)}{3}$, P-ps sur

$$\{Y_0^\tau \in B(\varphi(0), \alpha)\} \cap \left\{ \frac{Y_n^\tau - Y_0^\tau}{n\tau} \in \bigcup_{\beta \in A} B(\beta, \varepsilon) \right\},$$

Y_n^τ appartient à $\{\pi(\cdot) \geq \pi_1\}$, avec $\pi_1 > 0$. En effet, on a, P-ps sur cet ensemble :

$$\pi(Y_n^\tau) \geq -\alpha + [\pi(c^1) \wedge \pi(c^2) - \varepsilon] n\tau \geq -\alpha + \frac{2K_1}{3}\pi(c^1) \wedge \pi(c^2) \geq \frac{K_1}{3}\pi(c^1) \wedge \pi(c^2),$$

d'où le résultat avec $\pi_1 = \frac{K_1}{3}\pi(c^1) \wedge \pi(c^2)$. D'autre part, $Y_n^\tau - Y_0^\tau$ est P-ps majoré par $n\tau M$ donc, pour $K_1 \leq n\tau \leq K$ et $n \geq N$, on a :

$$\tau \log P(Y_n^\tau \in \mathcal{E} \mid Y_0^\tau \in B(\varphi(0), \alpha)) \geq -2\varepsilon, \quad (3.50)$$

où $\mathcal{E}$ est l'ensemble $\{\pi(\cdot) \geq \pi_1\} \cap B(\varphi(0), \alpha + MK)$ (boule fermée).

Pour y dans $\mathcal{E}$, on note φ_y le chemin $y - \varphi(0) + \varphi$. On prend K et α assez petits pour avoir, pour tous x, y dans $\mathcal{U}$ tels que $|x - y| \leq \alpha + MK$ et pour tout v dans $(S_x^2)^{-\delta}$, $|L_x^2(v) - L_y^2(v)|$ inférieur à ε . Alors, puisque $|y - \varphi(0)| \leq \alpha + MK$, $\Lambda_{0,T}^2(\varphi_y)$ est inférieur à $\Lambda_{0,T}^2(\varphi) + \varepsilon$. Les chemins du tube d'axe φ_y et rayon $\frac{\pi_1}{2}$ sont à valeurs dans P^2 sur $[0, T]$ (car $\pi(y) - \pi(\varphi(0)) \geq \pi_1$ et $\pi(\varphi(t)) \geq 0$). D'après le théorème 3.2.3, il existe alors α_y et τ_y tels que, pour $\tau \leq \tau_y$, on ait, P-ps sur $\{Y_0^\tau \in B(y, \alpha_y)\}$:

$$\tau \log P\left(d_{0,T}(\gamma^\tau, \varphi_y) \leq \frac{\pi_1}{2} \mid \mathcal{F}_0\right) \geq -\Lambda_{0,T}^2(\varphi_y) - \varepsilon.$$

On recouvre $\mathcal{E}$ par une union finie de boules ouvertes $B(y_i, \alpha_{y_i})$, avec y_i dans $\mathcal{E}$. Alors, pour $\tau \leq \inf \tau_{y_i}$, P-ps sur $\{Y_0^\tau \in \mathcal{E}\}$, on a :

$$\tau \log P\left(\text{pour un } i, d_{0,T}(\gamma^\tau, \varphi_{y_i}) \leq \frac{\pi_1}{2} \mid \mathcal{F}_0\right) \geq -\Lambda_{0,T}^2(\varphi) - 2\varepsilon.$$

On prend K, K_1 et α assez petits pour que $\alpha + MK + \frac{\pi_1}{2}$ soit inférieur à ε . Si pour un i , $d_{0,T}(\gamma^\tau, \varphi_{y_i})$ est inférieur à $\frac{\pi_1}{2}$, alors $d_{0,T}(\gamma^\tau, \varphi)$ est inférieur à ε . Donc, pour $\tau \leq \inf \tau_{y_i}$, P-ps sur $\{Y_0^\tau \in \mathcal{E}\}$, on a :

$$\tau \log P(d_{0,T}(\gamma^\tau, \varphi) \leq \varepsilon \mid \mathcal{F}_0) \geq -\Lambda_{0,T}^2(\varphi) - 2\varepsilon. \quad (3.51)$$

On utilise (3.50) pour $(Y_k^\tau)_{k \leq n}$ et (3.51) pour $(Y_k^\tau)_{k \geq n}$. La quantité $n\tau$ appartient à $[K_1, K]$. Prendre n grand revient donc à prendre τ petit. Il existe donc τ_0 tel que, pour $\tau \leq \tau_0$, l'inégalité (3.50) est vérifiée: P-ps sur $\{Y_0^\tau \in B(\varphi(0), \alpha)\}$, on a $\tau \log P(Y_n^\tau \in \mathcal{E} \mid \mathcal{F}_0) \geq -2\varepsilon$. D'autre part, d'après (3.51), pour $\tau \leq \inf \tau_{y_i}$, P-ps sur $\{Y_n^\tau \in \mathcal{E}\}$, on a:

$$\tau \log P\left(\sup_{n\tau \leq t \leq T} |\gamma^\tau(t) - \varphi(t - n\tau)| \leq \varepsilon \mid \mathcal{F}_n\right) \geq -\Lambda_{0,T}^2(\varphi) - 2\varepsilon.$$

On prend K assez petit pour que $t, t' \in [0, T]$ et $|t - t'| \leq K$ impliquent $|\varphi(t) - \varphi(t')| \leq \varepsilon$. Alors, pour $\tau \leq \inf \tau_{y_i}$, P-ps sur $\{Y_n^\tau \in \mathcal{E}\}$, on a:

$$\tau \log P(d_{n\tau, T}(\gamma^\tau, \varphi) \leq 2\varepsilon \mid \mathcal{F}_n) \geq -\Lambda_{0,T}^2(\varphi) - 2\varepsilon.$$

Donc, pour $\tau \leq \tau_0 \wedge \inf \tau_{y_i}$, P-ps sur $\{Y_0^\tau \in B(\varphi(0), \alpha)\}$, on a:

$$\tau \log P(d_{0,T}(\gamma^\tau, \varphi) \leq 2\varepsilon \mid \mathcal{F}_0) \geq$$

$$\tau \log P(Y_n^\tau \in \mathcal{E} \mid \mathcal{F}_0) + \tau \log P(d_{n\tau, T}(\gamma^\tau, \varphi) \leq 2\varepsilon \mid Y_n^\tau \in \mathcal{E}) \geq -\Lambda_{0,T}^2(\varphi) - 4\varepsilon,$$

ce qui démontre (3.35) pour φ .

3.7 Cas des champs de probabilités constants, sous l'hypothèse (H_0)

Dans le cas des champs de probabilités constants, les hypothèses (Hmin) sont vérifiées dès qu'on a $\mu_x^i = \mu^i$ avec S^i compact d'intérieur non vide (on note sans indice x les données relatives aux champs de probabilités) et (HLip) est toujours vraie. D'autre part, si $\{t_j, i_j\}$ est associé comme dans (3.31) à φ appartenant à $C_x^f[0, T]$, avec $\Lambda_{0,T}(\varphi)$ fini, quand i_j est non nul, S^{i_j} rencontre $\{(-1)^{i_j} \pi(\cdot) > 0\}$. Dans tout ce paragraphe, l'hypothèse (H_0) est supposée vérifiée. La condition (2) dans la définition de $C^{f,r}[0, T]$ (notation 3.5.2) est alors satisfaite pour tout chemin φ de coût fini. On a donc, pour toute partie borélienne G :

$$\Lambda(\overset{\circ}{G} \cap C^{f,r}[0, T]) = \inf\{\Lambda_{0,T}(\varphi) \mid \varphi \in G \text{ et } \exists \theta, J_{0,T}^\theta(\varphi) < +\infty\}$$

Nous allons montrer qu'en fait, même si, pour tout θ , $\Lambda_{0,T}^\theta(\varphi)$ vaut $+\infty$, on peut remplacer $\lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi)$ par $\Lambda_{0,T}(\varphi)$ dans le théorème 3.4.1. Grâce à ce résultat, la minoration avec $\Lambda_{0,T}(\varphi)$ pour le tube d'axe φ est obtenue pour tout φ dans $C^f[0, T]$ et

non plus seulement dans $C^{f,\tau}[0, T]$. On obtient alors le théorème 3.5.3, où $C^{f,\tau}[0, T]$ est remplacé par $C^f[0, T]$. Nous énonçons dans le chapitre 4 le résultat complet de grandes déviations dans ce cas.

Montrons ici le théorème suivant :

Théorème 3.7.1

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 associé à des champs de probabilités constants (μ^1) et (μ^2) dont les supports sont bornés, engendrent $\mathbb{R}^d$ et vérifient l'hypothèse (H_0) . Soit φ un chemin de $C^f[0, T]$. Alors, pour tout ε , il existe α et τ_0 tels que, pour tout $\tau \leq \tau_0$ et tout y_0 de $B(\varphi(0), \alpha)$, on ait :

$$\tau \log P_{y_0}(d_{0,T}(\gamma^\tau, \varphi) \leq \varepsilon) \geq -\Lambda_{0,T}(\varphi) - \varepsilon$$

Démonstration : on n'est pas obligé de se limiter aux $(\rho^1, \rho^2, \beta^1, \beta^2)$ vérifiant les hypothèses de la proposition 3.3.1 : montrons que pour tout chemin φ , quand $\Lambda_{0,T}(\varphi)$ s'écrit $\int_0^T \sum_{i=1}^2 \rho_i^i L^i(\beta_i^i) dt < +\infty$, avec les conditions (3.34) pour presque tout t , on peut construire un chemin ψ vérifiant les hypothèses du théorème 3.2.2 (avec $\psi'(t) = \sum_{i=1}^2 \tilde{\rho}_i^i \tilde{\beta}_i^i$) et tel que ψ est arbitrairement proche de φ et $\int_0^T \sum_{i=1}^2 \tilde{\rho}_i^i L^i(\tilde{\beta}_i^i) dt$ arbitrairement proche de $\Lambda_{0,T}(\varphi)$. Nous en déduisons la minoration avec $\Lambda_{0,T}(\varphi)$.

Considérons donc un chemin φ tracé sur P^0 tel que $\Lambda_{0,T}(\varphi)$ soit fini. P-ps en t , il existe $(\rho_t^1, \rho_t^2, \beta_t^1, \beta_t^2)$ dans $[0, 1]^2 \times S^1 \times S^2$ tel que $L(\varphi'(t))$ soit égal à $\sum_{i=1}^2 \rho_t^i L^i(\beta_t^i)$, avec (ρ_t^i, β_t^i) vérifiant les conditions (3.34).

La construction étend celle de la démonstration de (3.35) quand $\varphi([0, T])$ est inclus dans P^0 : dans cette démonstration, nous nous sommes limités aux écritures $\varphi'(t) = \sum_{i=1}^2 \rho_t^i \beta_t^i$ avec $(-1)^i \pi(\beta_t^i) \leq -\theta_0$, ce qui a conduit à une minoration avec $J_{0,T}^{\theta_0}(\varphi)$. Ici la contrainte n'est plus que $(-1)^i \pi(\beta_t^i) \leq 0$ qui amène à une minoration en $\Lambda_{0,T}(\varphi)$. On distingue trois cas pour la construction de $(\tilde{\rho}_t^i, \tilde{\beta}_t^i)$. Le premier ($t \in \mathcal{A}_1$) est utilisé pour les t tels que $(-1)^i \pi(\beta_t^i) \leq -\theta_0$ est encore vérifié ; la construction est alors celle de la démonstration du théorème 3.3.3. Les deux suivants ($t \in \mathcal{A}_2$ et $t \in \mathcal{A}_3$) aboutissent seulement dans le cas présent où les champs de probabilités sont supposés constants. Ces trois cas sont les mêmes que dans la démonstration du lemme 3.4.6 mais on n'a plus la condition $\beta_t^i \in (S_{\varphi(t)}^i)^{-\delta_0}$.

Soit ρ_0, θ_0 et θ dans $[0, \frac{1}{2}]$ et $\varepsilon > 0$. La construction de $\tilde{\rho}_t^i$ et $\tilde{\beta}_t^i$ va être faite en fonction des valeurs de ρ^i et $(-1)^i \pi(\beta_t^i)$ par rapport respectivement à ρ_0 et à $-\theta_0$. Rappelons qu'on a de toute façon $(-1)^i \pi(\beta_t^i) \leq 0$. On prendra ρ_0 et θ_0 petits, pour

ε fixé. Pour ρ_0 et θ_0 fixés, la modification sera proportionnelle à θ .

On note

- $\mathcal{A}_1$ l'ensemble des t pour lesquels, pour $i = 1$ et 2 , $(-1)^i \pi(\beta_i^i) \leq -\theta_0$,
- $\mathcal{A}_2$ l'ensemble des t pour lesquels un des $(-1)^i \pi(\beta_i^i)$ est dans $[-\theta_0, 0]$ et ρ_i^1 dans $[\rho_0, 1 - \rho_0]$ (dans la construction, on supposera $\pi(\beta_i^1) \leq \theta_0$), et
- $\mathcal{A}_3$ l'ensemble des t pour lesquels un des $(-1)^i \pi(\beta_i^i)$ est dans $[-\theta_0, 0]$ et ρ_i^1 n'appartient pas à $[\rho_0, 1 - \rho_0]$. On suppose, par exemple, ρ_i^1 inférieur à ρ_0 ; alors $\pi(\beta_i^2)$ appartient à $[-\theta_0, 0]$ (β_i^1 n'est pas toujours défini).

1. si t appartient à $\mathcal{A}_1$, on pose $v_i^i = h_\theta^i(\beta_i^i)$ (définition 2.1.2, avec des champs constants). On a les propriétés suivantes: $|v_i^i - \beta_i^i| \leq 2M\theta$, et $(-1)^i \pi(v_i^i) = \theta(-1)^i \pi(\beta_i^i) + (1 - \theta)(-1)^i \pi(\beta_i^i) \leq M\theta - (1 - \theta)\theta_0 \leq -\frac{\theta_0}{2}$, pour $(M + \theta_0)\theta \leq \frac{\theta_0}{2}$; d'autre part, $v_i^i \in (S^i)^{-C_1\theta}$, avec $C_1 > 0$, enfin $L^i(v_i^i) \leq L^i(\beta_i^i)$.
2. si t appartient à $\mathcal{A}_2$. Pour chaque i , on considère b^i dans $\overset{\circ}{S}^i$ tel que $(-1)^i \pi(b^i) < 0$. On prend π_0 et δ_0 strictement positifs tels que, $\pi(b^1) \geq \pi_0$, $\pi(b^2) \leq -\pi_0$ et $b^i \in (S^i)^{-\delta_0}$. Pour $i = 1$ ou 2 et θ dans $[0, 1]$, on note g_θ^i l'application définie sur $\mathbb{R}^d$ par $g_\theta^i(v) = \theta b^i + (1 - \theta)v$. Quitte à diminuer C_1 , on a $g_\theta^i(S^i) \subset (S^i)^{-C_1\theta}$.

Pour chaque i , on a $(-1)^i \rho_i^i \pi(b^i) < 0$; il existe donc α_i^1 dans $]0, 1[$ tel que, si on pose $\alpha_i^2 = 1 - \alpha_i^1$, on ait $\pi(\sum_{i=1}^2 \alpha_i^i \rho_i^i b^i) = 0$ (si les supports le permettent, par exemple s'ils rencontrent tous les deux $\{\pi(\cdot) = 0\}$, on prend $\pi(b^1) = -\pi(b^2) = \pi_0$ et on remplace α_i^i par ρ_i^{3-i}). Alors α_i^1 est égal à $\frac{-\rho_i^2 \pi(b^2)}{\rho_i^1 \pi(b^1) - \rho_i^2 \pi(b^2)}$. Comme $\pi(b^1)$ et $-\pi(b^2)$ appartiennent à $[\pi_0, M]$, α_i^1 est supérieur à $\alpha_0 = \frac{\rho_0 \pi_0}{M}$, de même pour α_i^2 . On a donc $\alpha_i^i \in [\alpha_0, 1 - \alpha_0]$.

On pose $v_i^i = g_{\alpha_i^i \theta}^i(\beta_i^i) = \alpha_i^i \theta b^i + (1 - \alpha_i^i \theta) \beta_i^i$. Alors, on a $|v_i^i - \beta_i^i| \leq 2M\theta$, $(-1)^i \pi(v_i^i) \leq \alpha_i^i \theta (-1)^i \pi(b^i) \leq -\theta \alpha_0 \pi_0$, $v_i^i \in (S^i)^{-C_1 \alpha_i^i \theta} \subset (S^i)^{-C_1 \alpha_0 \theta}$ et

$$L^i(v_i^i) \leq \alpha_i^i \theta J(\delta_0) + (1 - \alpha_i^i \theta) L^i(\beta_i^i) \leq L^i(\beta_i^i) + \varepsilon,$$

quand on suppose

$$\theta \leq \frac{\varepsilon}{J(\delta_0)}. \quad (3.52)$$

Rappelons que $J(\delta_0)$ est égal à $\max \{L^i(\beta) \mid \beta \in (S^i)^{-\delta_0}, i = 1 \text{ ou } 2\}$.

3. si t appartient à $\mathcal{A}_3$, on a supposé $\rho_i^1 < \rho_0$ qui implique $\pi(\beta_i^2) \in [-\theta_0, 0]$ (avec, éventuellement $\rho_i^1 = 0$ et β_i^1 non défini). On pose $v_i^1 = b^1$ et $v_i^2 = g_\theta^2(\beta_i^2)$. Alors, on a, d'une part, si β_i^1 est défini, $\rho_i^1 |v_i^1 - \beta_i^1| \leq 2M\rho_i^1 < 2M\rho_0$ et

$\pi(v_t^1) \geq \pi_0$, et d'autre part, $|v_t^2 - \beta_t^2| \leq 2M\theta$ et $\pi(v_t^2) \leq -\theta\pi_0$.

Enfin, quitte à supposer $C_1\theta \leq \delta_0$, v_t^i appartient à $(S^i)^{-C_1\theta}$ pour chaque i et

$$L^2(v_t^2) \leq \theta J(\delta_0) + (1 - \theta)L^2(\beta_t^2)$$

donc

$$\sum_{i=1}^2 \rho_t^i L^i(v_t^i) \leq \rho_t^1 J(\delta_0) + \rho_t^2 \left(\theta J(\delta_0) + L^2(\beta_t^2) \right) \leq \rho_t^2 L^2(\beta_t^2) + \varepsilon,$$

pour $(\rho_0 + \theta)J(\delta_0) \leq \varepsilon$.

Dnas tous les cas, les vecteurs v_t^i vérifient presque sûrement en t :

1. $\sum_{i=1}^2 \rho_t^i |v_t^i - \beta_t^i| \leq 2M\theta + 2M\rho_0$,
2. $(-1)^i \pi(v_t^i) \leq -\theta\alpha_0\pi_0$ (dès que $\theta\alpha_0\pi_0$ est inférieur à $\frac{\theta_0}{2}$),
3. $v_t^i \in (S^i)^{-C_1\alpha_0\theta}$ et
4. $\sum_{i=1}^2 \rho_t^i L^i(v_t^i) \leq \sum_{i=1}^2 \rho_t^i L^i(\beta_t^i) + \varepsilon$.

Soit $\varepsilon_1 > 0$. Pour chaque i , on approche $t \mapsto v_t^i$ par une fonction constante par morceaux $t \mapsto \tilde{\beta}_t^i$ telle que, presque sûrement en t , on ait $|\tilde{\beta}_t^i - v_t^i| \leq \varepsilon_1\theta$ (cf [1] p.33). Supposons $\varepsilon_1 \leq \frac{\alpha_0}{2}(\pi_0 \wedge C_1)$. En utilisant les propriétés 2 et 3 de v_t^i , on montre :

$$(-1)^i \pi(\tilde{\beta}_t^i) \leq -\frac{\alpha_0\pi_0}{2}\theta. \quad (3.53)$$

$$\tilde{\beta}_t^i \in (S^i)^{-\theta\frac{C_1\alpha_0}{2}} \quad (3.54)$$

L^i est uniformément continue sur $(S^i)^{-\theta C_1\alpha_0}$; pour ε_1 assez petit, $|v - v'| \leq \varepsilon_1\theta$ et $v \in (S^i)^{-\theta C_1\alpha_0}$ impliquent $|L^i(v) - L^i(v')| \leq \varepsilon$. Alors on a $|L^i(v_t^i) - L^i(\tilde{\beta}_t^i)| \leq \varepsilon$. En utilisant la propriété 4 des v_t^i , on en déduit :

$$\sum_{i=1}^2 \rho_t^i L^i(\tilde{\beta}_t^i) \leq \sum_{i=1}^2 \rho_t^i L^i(\beta_t^i) + 2\varepsilon. \quad (3.55)$$

(3.53) permet de définir $\tilde{\rho}_t^1$ tel que $\pi(\sum_{i=1}^2 \tilde{\rho}_t^i \tilde{\beta}_t^i) = 0$, avec $\tilde{\rho}_t^2 = 1 - \tilde{\rho}_t^1$. Alors $\tilde{\rho}_t^i$ appartient à $]0, 1[$. On pose

$$\psi(t) = \varphi(0) + \int_0^t \sum_{i=1}^2 \tilde{\rho}_u^i \tilde{\beta}_u^i du$$

Le chemin ψ est à valeurs dans P^0 . On a :

$$|\psi'(t) - \varphi'(t)| \leq \sum_{i=1}^2 \left(|\tilde{\beta}_t^i| |\tilde{\rho}_t^i - \rho_t^i| + \rho_t^i |\tilde{\beta}_t^i - \beta_t^i| \right). \quad (3.56)$$

Etudions $|\tilde{\rho}_t^i - \rho_t^i|$ dans chacun des cas :

1. si $t \in \mathcal{A}_1$, on a (comme pour (3.30)) :

$$|\tilde{\rho}_t^i - \rho_t^i| \leq \frac{\max_i |\tilde{\beta}_t^i - \beta_t^i|}{\pi(\beta_t^1) - \pi(\beta_t^2)} \leq \frac{2M\theta + \varepsilon_1\theta}{2\theta_0}.$$

D'autre part,

$$\rho_t^i = \frac{(-1)^i \pi(\beta_t^{3-i})}{\pi(\beta_t^1) - \pi(\beta_t^2)} \geq \frac{\theta_0}{2M}.$$

On en déduit

$$|\tilde{\rho}_t^i - \rho_t^i| \leq \frac{M}{\theta_0^2} (2M\theta + \varepsilon_1\theta) \rho_t^i.$$

2. si $t \in \mathcal{A}_2$, on a $(-1)^i \pi(v_t^i) < 0$, il existe donc ν_t^1 dans $]0, 1[$ tel que $\pi(\sum_{i=1}^2 \nu_t^i v_t^i) = 0$ pour $\nu_t^2 = 1 - \nu_t^1$. La quantité $|\nu_t^i - \rho_t^i|$ est petite grâce au choix de α_t^i :

si un des $\pi(\beta_t^i)$ est nul, l'autre l'est aussi et $\pi(v_t^i)$ vaut $\theta \alpha_t^i \pi(b^i)$ donc ν_t^i et ρ_t^i sont égaux. Sinon, on a : $\rho_t^1 - \nu_t^1 = \frac{\rho_t^1 \pi(v_t^1) + \rho_t^2 \pi(v_t^2)}{\pi(v_t^1) - \pi(v_t^2)}$, avec $\pi(\sum_{i=1}^2 \alpha_t^i \rho_t^i b^i) = 0$ et $(-1)^{3-i} \pi(v_t^i) \geq \alpha_0 \pi_0 \theta$. On en déduit

$$|\rho_t^i - \nu_t^i| \leq \frac{1}{2\alpha_0 \pi_0 \theta} |(1 - \alpha_t^1 \theta) \rho_t^1 \pi(\beta_t^1) + (1 - \alpha_t^2 \theta) \rho_t^2 \pi(\beta_t^2)|.$$

$\pi(\sum_{i=1}^2 \rho_t^i \beta_t^i) = 0$ et $\pi(\beta_t^1) \leq \theta_0$ impliquent alors

$$|\rho_t^i - \nu_t^i| \leq \frac{\theta |\alpha_t^2 - \alpha_t^1| \rho_t^1 \pi(\beta_t^1)}{2\alpha_0 \pi_0 \theta} \leq \frac{\theta_0}{2\alpha_0 \pi_0}$$

Or on a $|\tilde{\rho}_t^i - \nu_t^i| \leq \frac{\max_i |\pi(\tilde{\beta}_t^i) - \pi(v_t^i)|}{\pi(v_t^1) - \pi(v_t^2)} \leq \frac{\varepsilon_1}{2\alpha_0 \pi_0}$ donc

$$|\tilde{\rho}_t^i - \rho_t^i| \leq \frac{\theta_0 + \varepsilon_1}{2\alpha_0 \pi_0} = \frac{\theta_0 + \varepsilon_1}{2\rho_0 \pi_0^2} M.$$

En supposant $M(\theta_0 + \varepsilon_1) \leq 2\rho_0^3 \pi_0^2$, on en déduit

$$|\tilde{\rho}_t^i - \rho_t^i| \leq \rho_0^2 \leq \rho_0 \rho_t^i.$$

3. si $t \in \mathcal{A}_3$, de $\pi(v_t^1) \geq \pi_0$, en supposant $\varepsilon_1\theta \leq \frac{\pi_0}{2}$, on déduit $\pi(\tilde{\beta}_t^1) \geq \frac{\pi_0}{2}$. Alors $\tilde{\rho}_t^1 = \frac{-\pi(\tilde{\beta}_t^2)}{\pi(\tilde{\beta}_t^1) - \pi(\tilde{\beta}_t^2)}$ est inférieur à $\frac{-2\pi(\tilde{\beta}_t^2)}{\pi_0}$.
 Or, on a $-\pi(\tilde{\beta}_t^2) \leq -\pi(\beta_t^2) + |\tilde{\beta}_t^2 - \beta_t^2| \leq \theta_0 + (2M + \varepsilon_1)\theta \leq \frac{3\theta_0}{2}$, pour $(2M + \varepsilon_1)\theta \leq \frac{\theta_0}{2}$; alors $\tilde{\rho}_t^1$ est majoré par $\frac{3\theta_0}{\pi_0}$.
 On a donc $|\tilde{\rho}_t^1 - \rho_t^1| \leq \frac{3\theta_0}{\pi_0} \vee \rho_0$. On suppose $3\theta_0 \leq \rho_0(1 - \rho_0)\pi_0$, alors on a $|\tilde{\rho}_t^1 - \rho_t^1| \leq \rho_0$ et $|\tilde{\rho}_t^2 - \rho_t^2| \leq \rho_0\rho_t^2$.

On suppose $(2M + \varepsilon_1)\theta \leq M\rho_0$. Alors, d'après la propriété 1 des v_t^i , dans tous les cas, $\sum_{i=1}^2 \rho_t^i |\tilde{\beta}_t^i - \beta_t^i|$ est majoré par $3M\rho_0$. Si on suppose de plus $\frac{M}{\theta_0^2}(2M + \varepsilon_1)\theta \leq \rho_0$, presque sûrement en t , $|\tilde{\rho}_t^i - \rho_t^i|$ est majoré par ρ_0 . On peut alors écrire, d'après (3.56) :

$$|\psi'(t) - \varphi'(t)| \leq 2M\rho_0 + 3M\rho_0 \leq 5M\rho_0.$$

On a donc $d_{0,T}(\psi, \varphi) \leq 5M\rho_0 T \leq \varepsilon$ pour ρ_0 assez petit.

Sauf pour $t \in \mathcal{A}_3$ et $i = 1$, presque sûrement en t , $|\tilde{\rho}_t^i - \rho_t^i|$ est majoré par $\rho_0\rho_t^i$. On prend $\varepsilon_1\theta \leq \frac{\theta_0}{2}$; alors, pour tout t de $\mathcal{A}_3$, $\tilde{\beta}_t^1$ appartient à $(S^1)^{-\frac{\theta_0}{2}}$. On a donc : $\tilde{\rho}_t^1 L^1(\tilde{\beta}_t^1) \leq \rho_0 J(\frac{\theta_0}{2}) \leq \varepsilon$ pour ρ_0 petit. En utilisant $\tilde{\rho}_t^i \leq (1 + \rho_0)\rho_t^i$ pour $t \notin \mathcal{A}_3$ ou $i \neq 1$, on obtient dans tous les cas :

$$\sum_{i=1}^2 \tilde{\rho}_t^i L^i(\tilde{\beta}_t^i) \leq (1 + \rho_0) \sum_{i=1}^2 \rho_t^i L^i(\tilde{\beta}_t^i) + \varepsilon \leq (1 + \rho_0) \left[\sum_{i=1}^2 \rho_t^i L^i(\beta_t^i) + 2\varepsilon \right] + \varepsilon,$$

d'après (3.55). On suppose ρ_0 choisi tel que $\rho_0\Lambda_{0,T}(\varphi) \leq \varepsilon T$. On a alors :

$$\int_0^T \sum_{i=1}^2 \tilde{\rho}_t^i L^i(\tilde{\beta}_t^i) dt \leq (1 + \rho_0) \left[\Lambda_{0,T}(\varphi) + 2\varepsilon T \right] + \varepsilon T \leq \Lambda_{0,T}(\varphi) + 6\varepsilon T$$

On fixe successivement ρ_0 , θ_0 , θ et ε_1 vérifiant toutes les conditions énoncées. Alors, d'après (3.53) et (3.54), ψ est un chemin régulier. Le théorème 3.2.2 implique donc qu'il existe α et τ_0 tels que, pour $\tau < \tau_0$, on ait, P-ps sur $\{Y_0^\tau \in B(\psi(0), \alpha)\}$:

$$\tau \log P(d_{0,T}(\gamma^\tau, \psi) \leq \varepsilon | \mathcal{F}_0) \geq - \int_0^T \sum_{i=1}^2 \tilde{\rho}_t^i L^i(\tilde{\beta}_t^i) dt - \varepsilon T.$$

On a alors, pour $\tau < \tau_0$, P-ps sur $\{Y_0^\tau \in B(\varphi(0), \alpha)\}$:

$$\tau \log P(d_{0,T}(\gamma^\tau, \varphi) \leq 2\varepsilon | \mathcal{F}_0) \geq - \int_0^T \sum_{i=1}^2 \rho_t^i L^i(\beta_t^i) dt - 7\varepsilon T$$

qui est bien le résultat annoncé. $\square$

3.8 Cas où l'hypothèse (H_0) n'est pas vérifiée

Dans les cinq premiers paragraphes de ce chapitre, nous n'avons pas fait l'hypothèse (H_0) . Cependant, les résultats obtenus sont triviaux si cette hypothèse est fausse (on minore toujours par $-\infty$; en particulier, $\Lambda_{/C^1, r[0, T]}$ vaut $+\infty$). Dans ce paragraphe, nous précisons les résultats dans le cas où les trajectoires restent dans un ouvert où l'un des S_x^i est toujours inclus dans l'ensemble des vecteurs strictement rentrants dans P^i , comme dans le paragraphe 2.6. Nous trouvons des fonctionnelles de coût très proches de celles obtenues pour la majoration.

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (Hmin) et (HLip).

Notation 3.8.1 On note $C^{T_{rav}+}$ l'ensemble des chemins φ de $C[0, T]$ qui traversent au plus une fois P^0 et tels que si $\{t \in [0, T] \mid \varphi(t) \in P^0\}$ est égal à $\{t\}$, pour i tel que $\varphi(T)$ appartient à P^i , $S_{\varphi(t)}^1$ et $S_{\varphi(t)}^2$ contiennent des vecteurs strictement rentrants dans P^i .

Remarquons que, comme $C^{T_{rav}}$, cet ensemble contient tous les chemins continus qui restent dans un même P^i .

Théorème 3.8.2

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses (Hmin) et (HLip) et tels que, pour tout x de $\mathbb{R}^d$, S_x^1 soit inclus dans l'ensemble $\{\beta \mid \pi(\beta) < 0\}$. Alors, pour tout y_0 de $\mathbb{R}^d$ et pour toute partie borélienne G de $C_{y_0}[0, T]$, on a :

$$-\Lambda(\overset{\circ}{G} \cap C^{T_{rav}+}) \leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0, T]}^\tau \in G)$$

Démonstration : si le processus a son origine Y_0^τ dans $P^1 \cup P^0$, le principe de grandes déviations est vérifié avec la fonctionnelle de coût Λ^1 . Or, si y_0 appartient à $P^1 \cup P^0$, dans le théorème ci-dessus, on a bien $\Lambda(\overset{\circ}{G} \cap C^{T_{rav}+}) = \Lambda^1(\overset{\circ}{G})$ puisque tous les chemins de coût fini et d'origine en y_0 restent dans $P^1 \cup P^0$. Remarquons de la même façon que, si par exemple, pour tout x de $\mathbb{R}^d$, S_x^2 est inclus dans l'ensemble $\{\beta \mid \pi(\beta) \geq 0\}$ et que le processus a son origine Y_0^τ dans P^2 , le principe de grandes déviations est vérifié avec la fonctionnelle de coût Λ^2 et on a bien $\Lambda(\overset{\circ}{G} \cap C^{T_{rav}+}) = \Lambda^2(\overset{\circ}{G})$ pour y_0 dans P^2 .

Soit φ un chemin de $C^{T_{rav}+}$ tel que $\varphi(0)$ appartienne à P^2 ; supposons par exemple que $\varphi(T)$ appartienne à P^1 . Le chemin φ a une seule intersection avec P^0 , à un instant t , avec $S_{\varphi(t)}^1 \cap \{\pi(\cdot) < 0\}$ et $S_{\varphi(t)}^2 \cap \{\pi(\cdot) < 0\}$ non vides. On minore la probabilité que le processus appartienne à un tube d'axe φ en utilisant le théorème 3.2.3 dans chaque demi-espace et le lemme 3.1.9 au voisinage de l'intersection avec P^0 , sur un intervalle de longueur arbitrairement petite (comme dans la démonstration du théorème 3.3.3 au voisinage de t_j tel que $i_j \neq 0$ et $\varphi(t_j) \in P^0$, sachant que $S_{\varphi(t)}^1$ et $S_{\varphi(t)}^2$ rencontrent tous deux l'ensemble $\{\pi(\cdot) < 0\}$). La fonctionnelle de coût à utiliser est bien $\Lambda_{0,T}$, c'est-à-dire qu'on obtient le résultat du théorème 3.3.3 mais avec $\Lambda_{0,T}(\varphi)$ dans le minorant.

Soit G une partie borélienne de $C_{y_0}[0, T]$ pour laquelle $\Lambda(\overset{\circ}{G} \cap C^{T_{rav}+})$ est fini. On prend, pour $\varepsilon > 0$ donné, φ dans $\overset{\circ}{G} \cap C^{T_{rav}+}$ vérifiant $\Lambda_{0,T}(\varphi) < \Lambda(\overset{\circ}{G} \cap C^{T_{rav}+}) + \varepsilon$. Pour ε_1 assez petit, $\overset{\circ}{G}$ contient le tube d'axe φ et rayon ε_1 . On a donc :

$$\liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(d_{0,T}(\varphi^\tau, \varphi) \leq \varepsilon_1) \geq -\Lambda_{0,T}(\varphi)$$

On en déduit :

$$-\Lambda(\overset{\circ}{G} \cap C^{T_{rav}+}) \leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi^\tau_{/[0,T]} \in G) \quad \square$$

On a vu dans le paragraphe 2.6 deux cas où $\Lambda(G \cap C^{T_{rav}})$ est toujours égal à $\Lambda(G)$, cela reste vrai avec $C^{T_{rav}+}$ puisque dans ces deux cas $C^{T_{rav}+} \cap \{\Lambda_{0,T}(\cdot) < +\infty\}$ et $C^{T_{rav}} \cap \{\Lambda_{0,T}(\cdot) < +\infty\}$ coïncident. On a vu également un cas (exemple 2.85) où la minoration serait fautive avec $\Lambda(\overset{\circ}{G})$ (pour $G = B(\varphi, \varepsilon)$ comme dans l'exemple). Le chemin φ alors considéré n'appartient pas à $C^{T_{rav}+}$; on a $\Lambda(B(\varphi, \varepsilon) \cap C^{T_{rav}+}) = +\infty$. Le chemin φ n'est donc pris en compte ni dans le calcul du majorant ni dans le calcul du minorant.

Remarquons que l'on peut remplacer $\Lambda(\overset{\circ}{G} \cap C^{T_{rav}+})$ par la borne inférieure des $\Lambda_{0,T}(\varphi)$ pour φ variant dans $\overset{\circ}{G}$ et tel que, si $[t_1, t_2]$ est l'ensemble des t tels que $\varphi(t)$ appartient à P^0 , alors $S_{\varphi(t_1)}^2 \cap \{\pi(\cdot) > 0\}$, $S_{\varphi(t_2)}^1 \cap \{\pi(\cdot) < 0\}$ et $S_{\varphi(t_2)}^2 \cap \{\pi(\cdot) < 0\}$ sont non vides. En effet, avec ces conditions sur φ , on peut approcher $\varphi_{/[0,t_2]}$ par un chemin ψ à vitesses intérieures par rapport au champ (μ_x^2) , à valeurs dans P^2 et de coût proche de $\Lambda_{0,t_2}^2(\varphi)$ (comme dans la démonstration de (3.35) quand on suppose que $\varphi(]0, T[)$ est inclus dans un P^i). En construisant correctement ψ sur $[t_2, T]$ on obtient un élément de $C^{T_{rav}+}$ tel que $d_{0,T}(\varphi, \psi)$ et $|\Lambda_{0,T}(\psi) - \Lambda_{0,T}(\varphi)|$ sont arbitrairement petits.

Dans le cas symétrique où S_x^2 est toujours inclus dans l'ensemble $\{\beta | \pi(\beta) < 0\}$, on peut constater que le résultat du théorème 3.8.2 est encore vrai. On a en particulier le résultat de grandes déviations avec la fonctionnelle Λ^2 pour y_0 dans P^2 , et avec la fonctionnelle Λ^1 pour y_0 dans $P^1 \cup P^0$ si pour tout x , S_x^1 est inclus dans l'ensemble $\{\beta | \pi(\beta) \leq 0\}$.

Les résultats complets de grandes déviations pour les cas envisagés dans ce paragraphe sont énoncés dans le chapitre 4.

Chapitre 4

Conclusion et compléments

Dans la première partie de ce chapitre, nous rassemblons les résultats de grandes déviations démontrés pour les processus d'apprentissage lent à statistiques discontinues le long d'un l'hyperplan. Dans la seconde partie, nous indiquons comment sont obtenues les équations vérifiées par les trajectoires minimisantes entre deux points et le coût optimal associé.

4.1 Résultats complets

Nous rassemblons, dans les deux premiers paragraphes, les résultats obtenus dans les chapitres 2 et 3, en fonction de l'hypothèse faite sur la configuration des supports des mesures au voisinage de la frontière.

Dans le premier cas, c'est-à-dire sous l'hypothèse (H_0) , le principe de grandes déviations fait intervenir au voisinage de la frontière un coût qui mélange les transformées de Cramer des deux champs de probabilités. Cela correspond à un comportement du processus tel que le nombre de traversées de la frontière sur un intervalle de temps de longueur donnée n'est pas borné. Notons que sur les vecteurs parallèles à la frontière, ce coût est inférieur à chacune des transformées de Cramer et que le processus peut rester au voisinage de la frontière alors qu'il la quitterait s'il était soumis à un seul des deux champs.

Au contraire dans le deuxième cas, localement, le processus traverse au plus une fois la frontière et la fonctionnelle de coût fait intervenir l'une ou l'autre des transformées de Cramer des deux champs. On se contente alors de calculer les fonctionnelles de coût le long de chemins qui ne suivent P^0 sur aucun intervalle et il n'est pas nécessaire de définir de fonction de coût L_x pour x dans P^0 .

Dans le troisième paragraphe, nous étudions les “cas limites” de configurations des supports exclues dans les deux premiers cas.

On constatera que dans tous les cas nous ne calculons des fonctionnelles de coût que le long des chemins de $C^f[0, T]$ c'est-à-dire changeant un nombre fini de fois d'ensemble P^i , pour $i = 0, 1$ ou 2 , comme nous l'avons indiqué après la remarque 1.4.9. En l'absence de frontière, cette restriction disparaît et nous retrouvons les résultats de [AR]. D'autre part, si l'on doit être prudent dans le choix des chemins sur lesquels on calcule le minorant, la majoration est toujours vraie avec la fonctionnelle Λ : pour tout y_0 de $\mathbb{R}^d$ et toute partie borélienne G de $C_{y_0}[0, T]$, on a

$$\limsup_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0, T]}^\tau \in G) \leq -\Lambda(\overline{G})$$

Rappelons enfin que l'énoncé complet des hypothèses figure à la fin du chapitre 1.

4.1.1 Cas où l'hypothèse (H_0) est partout vérifiée

Dans le cas où l'hypothèse (H_0) est partout vérifiée, on a le résultat final suivant :

Théorème 4.1.1

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses $(Hmin)$, $(HLip)$ et (H_0) . On appelle φ^τ le chemin aléatoire continu défini sur $[0, T]$ par :

*$\varphi^\tau(k\tau) = Y_k^\tau$ pour $0 \leq k\tau < T + \tau$ et φ^τ est affine sur $[k\tau; (k+1)\tau]$.
Pour tout y_0 de $\mathbb{R}^d$ et toute partie borélienne G de $C_{y_0}[0, T]$, on a :*

$$\begin{aligned} -\Lambda(\overset{\circ}{G} \cap C^{f, \tau}[0, T]) &\leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0, T]}^\tau \in G) \\ &\leq \limsup_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0, T]}^\tau \in G) \leq -\Lambda(\overline{G} \cap C^{f, 1}[0, T]) \end{aligned}$$

Rappelons que les hypothèses (i) et (ii) dans $(Hmin)$, ainsi que $(HLip)$, sont nécessaires seulement pour obtenir l'inégalité de minoration.

Dans ce résultat de grandes déviations, les coûts de majoration et de minoration sont tous deux minorés par Λ et proches en général. Ils coïncident dans certains cas :

Exemple 1 Si pour tout x de $\mathbb{R}^d$, chaque S_x^i est inclus dans $\{(-1)^i \pi(\cdot) < 0\}$:

$$\frac{\textcircled{S_x^1}}{\textcircled{S_x^2}}$$

Alors, pour toute partie borélienne G de $C[0, T]$, les quantités $\Lambda(G)$, $\mathcal{L}(G)$ et $\Lambda(G \cap C^{f,1}[0, T])$ coïncident (en effet, $\Lambda_{0,T}^{\theta,h}$ et $\Lambda_{0,T}$ sont égaux, tandis que $\Lambda_{0,T}$ est s.c.i. d'après $L_x = \text{conv}(L_x^1, L_x^2)$ sur l'ensemble $\{\beta \mid \pi(\beta) = 0\}$ -fonction convexe- et le théorème 1.4.10). D'autre part, pour φ dans $C[0, T]$, on a $\lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi) = \Lambda_{0,T}(\varphi)$ (cf paragraphe 3.4.2) et en fait, pour toute partie borélienne G , on a $\Lambda(G \cap C^{f,r}[0, T]) = \Lambda(G)$ (les hypothèses (1) et (2) de la définition de $C^{f,r}[0, T]$ sont toujours vérifiées : on n'a jamais $i_j \neq 0$ avec $\Lambda_{0,T}(\varphi)$ fini, puisque les seuls chemins de coût fini sont à valeurs dans P^0).

Dans ce cas, le processus reste au voisinage de la frontière et oscille entre les deux demi-espaces. Précisément, le nombre de traversées de P^0 par le processus pendant une durée fixée tend vers $+\infty$ quand τ tend vers 0. Si les hypothèses (Hmin), (HLip) et (H_0) sont vérifiées, pour tout y_0 de $\mathbb{R}^d$ et toute partie borélienne G de $C_{y_0}[0, T]$, on a :

$$-\Lambda(\overset{\circ}{G}) \leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0,T]}^\tau \in G) \leq \limsup_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0,T]}^\tau \in G) \leq -\Lambda(\overline{G}). \quad (4.1)$$

Exemple 2 Cas des champs de probabilités constants :

On suppose qu'on a $\mu_x^i = \mu^i$, probabilité à support borné engendrant $\mathbb{R}^d$ et (H_0) vérifiée ((HLip) l'est nécessairement ainsi que les hypothèses (Hmin)). On a montré (paragraphe 3.7) que dans ce cas, on a encore le théorème 3.5.3, avec $C^{f,r}[0, T]$ remplacé par $C^f[0, T]$. On peut donc énoncer le théorème suivant :

Théorème 4.1.2

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 associé à des champs de probabilités constants (μ^1) et (μ^2) dont les supports sont bornés, engendrent $\mathbb{R}^d$ et vérifient l'hypothèse (H_0) . Alors, pour tout y_0 de $\mathbb{R}^d$ et toute partie borélienne G de $C_{y_0}[0, T]$, on a :

$$\begin{aligned} -\Lambda(\overset{\circ}{G} \cap C^f[0, T]) &\leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0,T]}^\tau \in G) \\ &\leq \limsup_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0,T]}^\tau \in G) \leq -\Lambda(\overline{G} \cap C^{f,1}[0, T]) \end{aligned}$$

Exemple 3 Cas où les deux champs de probabilités sont égaux :

On suppose que pour tout x , on a $\mu_x^1 = \mu_x^2$. Alors Λ est égal à Λ^1 . Sous l'hypothèse (H_0) , la condition (2) dans la définition de $C^{f,r}[0, T]$ (notation 3.5.2) est toujours vraie. Comme dans le cas des champs constants, le théorème 3.4.1 est vrai avec $\Lambda_{0,T}(\varphi)$ au lieu de $\lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi)$. Le théorème de l'exemple précédent est donc encore vrai. Notre majorant est au moins aussi bon que celui obtenu dans [AR] ; le minorant peut être moins bon. En fait, $\Lambda(\overset{\circ}{G} \cap C^{f,r}[0, T])$ et $\Lambda(\overline{G} \cap C^{f,1}[0, T])$ coïncident sans doute dans de nombreux cas avec respectivement $\Lambda(\overset{\circ}{G})$ et $\Lambda(\overline{G})$.

4.1.2 Cas où pour tout x , S_x^i est inclus dans l'ensemble $\{\beta | (-1)^i \pi(\beta) > 0\}$

Dans le second cas, on a le résultat suivant :

Théorème 4.1.3

On considère un processus d'apprentissage lent à statistiques discontinues le long de l'hyperplan P^0 , $\{(Y_k^\tau)_{k \in \mathbb{N}}\}$, associé à des champs de probabilités (μ_x^1) et (μ_x^2) vérifiant les hypothèses $(Hmin)$ et $(HLip)$, et tels que, pour tout x de $\mathbb{R}^d$, S_x^1 soit inclus dans l'ensemble $\{\beta | \pi(\beta) < 0\}$. Pour tout y_0 de $\mathbb{R}^d$ et toute partie borélienne G de $C_{y_0}[0, T]$, on a :

$$\begin{aligned} -\Lambda(\overset{\circ}{G} \cap C^{Trav+}) &\leq \liminf_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0,T]}^\tau \in G) \\ &\leq \limsup_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0,T]}^\tau \in G) \leq -\Lambda(\overline{G} \cap C^{Trav}) \end{aligned}$$

Remarquons que, si y_0 appartient à $P^1 \cup P^0$, le principe de grandes déviations est vérifié avec la fonctionnelle de coût Λ^1 ; or on a bien $\Lambda(\overset{\circ}{G} \cap C^{Trav+}) = \Lambda^1(\overset{\circ}{G})$ et $\Lambda(\overline{G} \cap C^{Trav}) = \Lambda^1(\overline{G})$ pour toute partie borélienne G de $C_{y_0}[0, T]$. D'autre part, si par exemple, pour tout x de $\mathbb{R}^d$, S_x^2 est inclus dans l'ensemble $\{\beta | \pi(\beta) \geq 0\}$ et que y_0 appartient à P^2 , le principe de grandes déviations est vérifié avec la fonctionnelle de coût Λ^2 , ce qui est là encore le résultat du théorème.

Dans le cas symétrique où S_x^2 est toujours inclus dans l'ensemble $\{\beta | \pi(\beta) < 0\}$, on a un résultat analogue, sauf pour la dernière inégalité, à remplacer par :

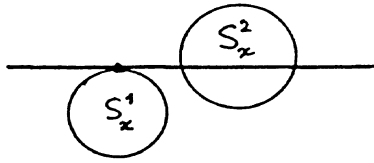
$$\limsup_{\tau \rightarrow 0} \tau \log P_{y_0}(\varphi_{/[0,T]}^\tau \in G) \leq -\Lambda(\overline{G}) \quad \text{pour } y_0 \text{ dans } P^1 \cup P^0$$

4.1.3 Cas limites (S_x^i tangent à $\{\beta | \pi(\beta) = 0\}$)

Dans les chapitres 2 et 3, nous nous sommes placés sur des ouverts où aucune des configurations suivantes n'était possible :

$$S_x^1 \subset \{\pi(\cdot) \leq 0\} \quad \text{et}$$

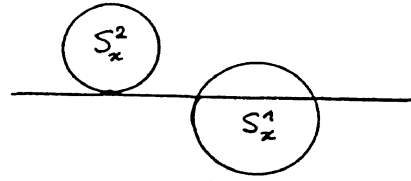
$$S_x^2 \cap \{\pi(\cdot) \leq 0\} \neq \emptyset$$



configuration 1

$$S_x^2 \subset \{\pi(\cdot) \geq 0\} \quad \text{et}$$

$$S_x^1 \cap \{\pi(\cdot) \geq 0\} \neq \emptyset$$



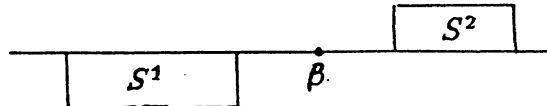
configuration 2

c'est-à-dire que nous avons donné les résultats de grandes déviations concernant les portions de trajectoires restant dans l'un de ces ouverts. Pour compléter cette étude, il conviendrait de donner des résultats permettant de relier ces différentes "portions" de trajectoires (en faisant intervenir les temps d'arrêt correspondant à l'entrée dans chacun de ces ouverts, ce que nous ne détaillons pas) mais aussi de décrire les portions de trajectoires restant dans un ouvert non vide où l'on a une des deux configurations précédentes.

Nous donnons, dans la suite de ce paragraphe quelques exemples de résultats et de situations éventuellement pathologiques (c'est-à-dire où la fonctionnelle de coût n'est pas Λ). Elles sont liées à la présence de masses de Dirac sur $\{\beta | \pi(\beta) = 0\}$ quand un des S_x^i est tangent à cet ensemble. Dans tous les cas, les méthodes exposées jusqu'à présent permettent de conclure. On remarque surtout que le comportement du processus devient très sensible au choix du champ de probabilités sur P^0 (alors que celui-ci est sans incidence quand (H_0) est vérifiée) et aux conditions initiales (loi de Y_0^r), et que la symétrie en l'indice i n'est pas toujours assurée. Remarquons également que les points où un des S_x^i est tangent à $\{\beta | \pi(\beta) = 0\}$ sont ceux où il peut y avoir changement de fonctionnelle, c'est-à-dire passage d'un type de résultat à l'autre.

- Points isolés : remarquons que si par exemple (H_0) est fautive en un point isolé x (et vraie autour de x) et si 0 n'appartient pas à l'enveloppe convexe de S_x^1 et S_x^2 , le processus quitte presque sûrement un petit voisinage de x en un temps (et avec un coût) arbitrairement petit, et on garde les résultats pour (H_0) vraie.

• Configuration 1 : si Y_0^τ appartient à $P^1 \cup P^0$, presque sûrement, le processus reste dans $P^1 \cup P^0$ et le champ μ^1 gouverne tous les pas. Alors le principe de grandes déviations (majoration et minoration) est vérifié avec la fonctionnelle de coût Λ^1 quels que soient les supports S_x^2 (les chemins de coût fini ayant leur origine dans $P^1 \cup P^0$ sont à valeurs dans $P^1 \cup P^0$). Le long de P^0 , Λ peut être différent de Λ^1 (quand les S_x^2 sont à cheval sur P^0) et n'est donc pas la bonne fonctionnelle de coût. Exemple :



On suppose que les champs de probabilités sont constants, avec les configurations du dessin pour les S^i , et que $\mu^1(\{\pi(\cdot) = 0\})$ et $\mu^2(\{\pi(\cdot) = 0\})$ sont strictement positifs. Si Y_0^τ appartient à $P^1 \cup P^0$ (respectivement P^2), le processus reste dans $P^1 \cup P^0$ (respectivement P^2). On en déduit le résultat de grandes déviations avec Λ^1 (respectivement Λ^2). Considérons alors φ tel que $\varphi(0)$ appartient à P^0 et, pour tout t , $\varphi'(t)$ est égal à β (cf dessin). $\Lambda_{0,T}^1(\varphi)$ et $\Lambda_{0,T}^2(\varphi)$ sont infinis alors que $\Lambda_{0,T}(\varphi)$ est fini et n'intervient pas dans les résultats. En fait, dans ce cas, les masses de chacune des mesures sur $\{\pi(\cdot) = 0\}$ ne se mélangent pas.

Revenons sur la remarque 1.4.2. Dans la démonstration des résultats quand (H_0) est vérifiée, le choix de la probabilité en un point de P^0 n'est pas intervenu. Il apparaît clairement dans cet exemple que ce choix devient significatif quand (H_0) n'est plus vraie : on aurait pu choisir au lieu de μ^1 un mélange aléatoire des deux probabilités, alors le coût $\Lambda_{0,T}(\varphi)$ interviendrait dans les estimations, ce qui correspondrait au mélange des masses de chacune des mesures sur $\{\pi(\cdot) = 0\}$.

Nous pouvons obtenir dans cet exemple des résultats très proches de ceux du théorème 4.1.3. En effet on peut encore appliquer la méthode du paragraphe 3.8 qui conduit à la minoration avec $\Lambda(\overset{\circ}{G} \cap C^{\tau_{rav^+}})$; pour la majoration, on avait utilisé dans le paragraphe 2.6 que le processus quittait le voisinage de la frontière dès qu'il était dans $P^1 \cup P^0$. Cela n'est plus vrai ici et l'on obtient plutôt une majoration avec la fonction $L_x^1 \wedge L_x^2$ le long de P^0 .

Avec la configuration 2, on obtient des résultats comparables.

4.2 Calcul du coût minimal entre deux points et des trajectoires minimisantes

Pour une partie borélienne G de $C_{x_0}[0, T]$, les quantités $\Lambda(\overset{\circ}{G} \cap C^{f,r}[0, T])$ et $\Lambda(\overline{G} \cap C^{f,1}[0, T])$ sont proches de

$$\inf\{\Lambda_{0,T}(\varphi) \mid \varphi \in G \cap C_{x_0}^f[0, T]\}. \quad (4.2)$$

Rappelons que $\Lambda_{0,T}(\varphi)$ vaut :

$$\begin{cases} \int_0^T \left[\sum_{i=1}^2 L_{\varphi(t)}^i(\varphi'(t)) \mathbb{I}_{\{\varphi(t) \in P^i\}} + L_{\varphi(t)}(\varphi'(t)) \mathbb{I}_{\{\varphi(t) \in P^0\}} \right] dt & \text{si } \varphi \text{ est absolument continu} \\ +\infty & \text{sinon} \end{cases}$$

et que d'après les résultats précédents, on peut se contenter d'envisager des chemins φ tels que, pour deux instants t_1 et t_2 quelconques, $\varphi([t_1, t_2])$ est inclus dans P^0 seulement si $(H_0(\varphi(t)))$ est fausse au plus pour un nombre fini de t de $[t_1, t_2]$.

Dans cette partie, on s'intéresse à la quantité (4.2), quand G est l'ensemble des chemins reliant en temps T deux points x_0 et y_0 . Si on identifie φ telle que $\Lambda_{0,T}(\varphi)$ soit égal à ce minimum, il conviendra alors de regarder si φ appartient à $C_{x_0}^{f,r}[0, T]$ ou à $C_{x_0}^{f,1}[0, T]$.

4.2.1 Cas particulier des champs de probabilités constants

Nous énonçons quelques résultats simples concernant la trajectoire optimale entre deux points dans le cas où les champs de probabilités sont constants.

Remarquons d'abord qu'en présence d'un seul champ constant, le coût optimal à durée T fixée entre deux points est obtenu, quand il est fini, pour le segment reliant ces deux points. On peut le montrer en utilisant l'inégalité de Jensen :

$$\frac{1}{T} \int_0^T L^i(\varphi'(t)) dt \geq L^i\left(\frac{1}{T} \int_0^T \varphi'(t) dt\right) = L^i\left(\frac{\varphi(T) - \varphi(0)}{T}\right).$$

On en déduit que dans le cas de deux champs, il y a au plus un segment où la trajectoire optimale suit P^0 . En effet, un chemin φ de la forme

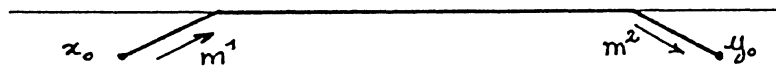


n'est pas optimal : entre x_0 et y_0 , le segment (parcouru pendant la même durée) est

moins coûteux (car L^i , fonction utilisée le long de φ , est convexe et supérieure à L sur $\{\pi(\cdot) = 0\}$).

En particulier, la trajectoire optimale entre deux points pendant la durée T appartient à $C^{f,1}[0, T]$ et si les deux points appartiennent à P^0 , elle est à valeurs dans P^0 .

On constate d'autre part que la présence de la frontière modifie le problème de façon importante : entre deux points de P^i le chemin optimal n'est plus nécessairement le segment qui les joint. Par exemple si on a $\pi(m^1) > 0$ et $\pi(m^2) < 0$, où m^i est le barycentre de μ^i , on peut suivre P^0 avec un coût nul. Le chemin



peut alors être moins coûteux que le segment joignant x_0 et y_0 . Cela est lié au fait qu'on peut avoir deux points x_0 et y_0 sur P^0 avec $L^1(\frac{y_0 - x_0}{T})$ et $L^2(\frac{y_0 - x_0}{T})$ élevés (voire infinis) alors que $L(\frac{y_0 - x_0}{T})$ est faible (voire nul comme sur le dessin).

Pour traiter le cas de champs de probabilités variables, nous donnons dans le prochain paragraphe un argument topologique permettant de prouver l'existence d'une trajectoire optimale entre deux points et dans les deux paragraphes suivants les méthodes habituelles permettant d'écrire des équations vérifiées par ce chemin optimal ou par le coût optimal lui-même.

4.2.2 Existence de chemins minimisants

Dans ce paragraphe, on donne un argument topologique permettant de prouver, quand F est une fonctionnelle sur $C[0, T]$, l'existence de chemins φ tels que $F(\varphi)$ soit minimale (on l'utilisera pour $F = \Lambda_{0,T}^i$ ou $F = \mathcal{L}_{0,T}$).

On a le résultat suivant :

Lemme 4.2.1

Soit F une fonctionnelle sur $C[0, T]$, à valeurs réelles. Si, pour tout $I \geq 0$, l'ensemble $\{F(\cdot) \leq I\}$ est compact, alors F atteint son minimum sur tout fermé non vide de $C[0, T]$.

Remarquons que F est nécessairement s.c.i.. L'hypothèse du lemme est par exemple vérifiée si F est s.c.i., avec, de plus, $\{\varphi \mid F(\varphi) < +\infty\}$ compact, ce qui est le cas

(toujours sous l'hypothèse de compacité des S_x^i) pour F égale à la restriction de $\Lambda_{0,T}^i$ à $\{\varphi | \varphi(0) \in \mathcal{K}\}$, avec $i = 1$ ou 2 et $\mathcal{K}$ un compact donné.

Démonstration: voir par exemple [FW].

On en déduit le théorème suivant sur le premier coût de majoration $\mathcal{L}_{0,T}$ apparaissant dans le théorème 2.2.10, même s'il n'est pas s.c.i..

Théorème 4.2.2

Sous les hypothèses (Hmaj) sur les champs de probabilités (μ_x^1) et (μ_x^2) , pour tout compact $\mathcal{K}$ de $\mathbb{R}^d$, la fonctionnelle $\mathcal{L}_{0,T}$ atteint son minimum sur tout fermé non vide de $\{\varphi \in C[0,T] \mid \varphi(0) \in \mathcal{K}\}$.

Démonstration: soit G un fermé non vide inclus dans $\{\varphi \in C[0,T] \mid \varphi(0) \in \mathcal{K}\}$ tel qu'on ait $\mathcal{L}(G) = \inf_{\varphi \in G} \mathcal{L}_{0,T}(\varphi) < +\infty$. Rappelons que la fonctionnelle $\mathcal{L}_{0,T}$ est définie sur un chemin φ absolument continu par :

$$\mathcal{L}_{0,T}(\varphi) = \lim_{(\theta,h) \rightarrow (0,0)} \mathcal{L}_{0,T}^{\theta,h}(\varphi) = \lim_{(\theta,h) \rightarrow (0,0)} \liminf_{\substack{\psi \rightarrow \varphi \\ \psi \in \mathcal{C}_{0,T}^f}} \Lambda_{0,T}^{\theta,h}(\psi).$$

On considère une suite $(\theta_n, h_n, \varphi_n)$ dans $\mathbb{R}^2 \times \mathbb{R}^2 \times G$ telle que, pour tout n , $\mathcal{L}_{0,T}^{\theta_n, h_n}(\varphi_n)$ soit inférieur à $\mathcal{L}(G) + \frac{1}{n}$. L'ensemble $G_n = \{\mathcal{L}_{0,T}^{\theta_n, h_n}(\cdot) \leq \mathcal{L}(G) + \frac{1}{n}\} \cap G$ est non vide puisqu'il contient φ_n et compact puisque, d'après la proposition 2.2.8, $\mathcal{L}_{0,T}^{\theta,h}$ est s.c.i. et que les éléments de G_n ont des dérivées uniformément bornées. L'intersection $\bigcap_n G_n$ contient un chemin φ de G tel que, pour tout n , on ait : $\mathcal{L}_{0,T}^{\theta_n, h_n}(\varphi) \leq \mathcal{L}(G) + \frac{1}{n}$. Alors, puisque $\mathcal{L}_{0,T}(\varphi)$ est la limite croissante, quand (θ, h) tend vers $(0, 0)$, de $\mathcal{L}_{0,T}^{\theta,h}(\varphi)$, on a :

$$\mathcal{L}_{0,T}(\varphi) = \lim_{n \rightarrow +\infty} \mathcal{L}_{0,T}^{\theta_n, h_n}(\varphi) \leq \lim_{n \rightarrow +\infty} \left[\mathcal{L}(G) + \frac{1}{n} \right] = \mathcal{L}(G). \quad \square$$

Ce dernier théorème prouve que, pour toute partie borélienne G , le coût de majoration $\mathcal{L}(\overline{G})$ dans le théorème 2.2.10 est atteint en un chemin de $\overline{G}$.

En revanche, on ne sait pas montrer que l'ensemble $\{\Lambda_{0,T}(\cdot) \leq I\}$ est compact (on n'a pas montré la s.c.i. éventuelle de Λ) donc le lemme 4.2.1 ne s'applique pas à $\Lambda_{0,T}$. Mais il s'applique pour un ensemble fermé de chemins à valeurs dans un même P^i . En particulier, $\Lambda_{0,T}$ atteint son minimum sur tout ensemble fermé de chemins à valeurs dans P^0 et d'origine dans un compact donné.

Par ailleurs, la quantité $\Lambda(\overline{G} \cap C^{f,1}[0, T])$ ne s'écrit pas nécessairement sous la forme $\Lambda_{0,T}(\varphi)$ avec φ dans $\overline{G} \cap C^{f,1}[0, T]$. Cela vient cette fois du fait que $C^f[0, T]$ n'est pas fermé. Considérons l'exemple où, en (x, y) de $\mathbb{R}^2$, on a la mesure :

$$\mu_{(x,y)} = \sum_{k=1}^K p_k(x, y) \delta_{a_k(x,y)}$$

avec $\sum_{k=1}^K p_k(x, y) = 1$ et $p_k(x, y) \geq 0$. On peut trouver des applications p_k et a_k continues telles que pour tout (x, y) , le barycentre $m_{(x,y)} = \sum_{k=1}^K p_k(x, y) a_k(x, y)$ soit :

$$m_{(x,y)} = \left(x, 3x^2 \sin \frac{1}{x} - x \sin \frac{1}{x} \right).$$

La courbe intégrale (φ, ψ) du champ moyen telle que $(\varphi(0), \psi(0)) = (0, 0)$ vérifie alors :

$$(\varphi(t), \psi(t)) = \left(t, t^3 \sin \frac{1}{t} \right).$$

Elle oscille donc une infinité de fois entre P^1 et P^2 au voisinage de 0. Prenons alors les deux champs $\mu_{(x,y)}^1$ et $\mu_{(x,y)}^2$ égaux à $\mu_{(x,y)}$ dans cet exemple ; si un ouvert G contient le chemin (φ, ψ) précédent, G contient un chemin de $C^{f,1}[0, T]$ de coût $\Lambda_{0,T}$ arbitrairement petit (pour des S_x^i suffisamment grands : il suffit de modifier légèrement (φ, ψ) au voisinage de 0). Donc $\Lambda(G \cap C^{f,1}[0, T])$ est nul tandis que $\Lambda_{0,T}$ ne s'annule pas sur $G \cap C^{f,1}[0, T]$, car la trajectoire du champ moyen n'appartient pas à $C^f[0, T]$ (donc le problème est lié au choix du majorant : $\Lambda(\overline{G})$ aurait convenu au lieu de $\Lambda(\overline{G} \cap C^{f,1}[0, T])$, et dans ce cas, le minimum est atteint en un point de G).

Remarquons qu'en remplaçant $t^3 \sin \frac{1}{t}$ par $t^5 \sin \frac{1}{t}$, le vecteur $m_{(x,y)}$ peut correspondre à $\sum_{k=1}^K p_k(x, y) a_k(x, y)$ avec p_k et a_k localement lipschitziennes, donc au barycentre pour un champ de probabilités vérifiant l'hypothèse (HLip).

Dans la fin de ce chapitre, nous établissons les équations vérifiées par les trajectoires minimisantes entre deux points et le coût optimal associé. Nous énonçons des conditions nécessaires, sans entrer dans le détail des résultats d'existence ou d'unicité de solutions pour lesquels nous donnons simplement des références.

4.2.3 Equations vérifiées par les chemins minimisants entre deux points : équations d'Euler

Soient x_0 et y_0 dans $\mathbb{R}^d$. L'ensemble des chemins φ de $C[0, T]$ tels que $\varphi(0) = x_0$ et $\varphi(T) = y_0$ est un fermé non vide. On sait donc que $\mathcal{L}_{0,T}$ atteint son minimum sur cet ensemble (théorème 4.2.2). Ce minimum correspond au coût minimal pour relier x_0 à y_0 en temps T fixé (problème à horizon fini). Un autre problème est de chercher à minimiser le coût entre ces deux points sans fixer le temps (problème à horizon infini). Comme dans le cas des champs constants, nous étudions le premier problème et commençons par le traiter dans le cas où il y a un seul champ de probabilités μ^i .

Rappelons que le calcul de trajectoires minimisantes peut servir à la mise au point d'algorithmes de simulation rapide d'événements rares, par exemple la sortie d'un bassin d'attraction de l'équation moyenne. Pour cela, on effectue un changement de probabilité le long de la trajectoire minimisante, ce qui transforme l'événement rare en événement non rare et permet d'accélérer sa simulation ([CFM], [DK]).

Cas d'un seul champ

On s'intéresse au problème de minimisation suivant :

$$\inf \{ \Lambda_{0,T}^i(\varphi) \mid \varphi \in C[0, T], \varphi(0) = x_0, \varphi(T) = y_0 \}, \quad (4.3)$$

où

$$\Lambda_{0,T}^i(\varphi) = \begin{cases} \int_0^T L^i(\varphi(t), \varphi'(t)) dt & \text{si } \varphi \text{ est absolument continu,} \\ +\infty & \text{sinon} \end{cases}$$

(on remplace $L_x^i(\beta)$ par $L^i(x, \beta)$ pour des commodités d'écriture.) Il s'agit d'un problème variationnel classique (cf [Ca] pour le cas le plus élémentaire où l'on recherche des fonctions φ de classe C^1 ou [Li] pour φ absolument continue). On sait le résoudre dans le cas où la fonction L^i est de classe C^1 . Supposons donc que le minimum (4.3) est atteint en un chemin φ tel que $(\varphi(\cdot), \varphi'(\cdot))$ est à valeurs dans un ouvert où L^i est de classe C^1 . On peut par exemple supposer L^i de classe C^1 sur $\mathbb{R}^d \times S_{\varphi(t)}^{\circ}$, ce qui est vérifié dans l'exemple suivant : considérons le champ de probabilités sur $\mathbb{R}^d$

$$\mu_x = \sum_{k=0}^d p_k(x) \delta_{a_k(x)}$$

avec a_k, p_k de classe C^1 . Pour simplifier, on suppose que les $d+1$ points $a_k(x)$ forment une base. Alors tout vecteur de l'enveloppe convexe des $a_k(x)$ est repéré par

ses coordonnées barycentriques. Supposons

$$\beta = \sum_{k=0}^d \rho_k(x, \beta) a_k(x), \text{ avec } \sum_{k=0}^d \rho_k(x, \beta) = 1$$

La transformée de Cramer de μ_x vaut alors en β :

$$L_x(\beta) = \sum_{k=0}^d \rho_k(x, \beta) \log \frac{\rho_k(x, \beta)}{p_k(x)}$$

Donc L est de classe C^1 en (x, β) sur $\bigcup_{\bar{x}} x \times \overset{\circ}{S}_x^i$ puisque les ρ_k le sont ($\rho_k(x, \beta)$ est une combinaison linéaire de β et des $a_k(x)$).

Si le minimum (4.3) est fini, on doit avoir :

pour tout chemin ψ de $C[0, T]$, de classe C^1 et tel que $\psi(0) = \psi(T) = 0$, pour tout $\varepsilon > 0$,

$$\int_0^T L^i(\varphi(t) + \varepsilon\psi(t), \varphi'(t) + \varepsilon\psi'(t)) dt \geq \int_0^T L^i(\varphi(t), \varphi'(t)) dt.$$

Pour toutes les fonctions ψ considérées, on doit donc avoir :

$$\int_0^T \left[\left\langle \frac{\partial L^i}{\partial \beta}(\varphi(t), \varphi'(t)), \psi'(t) \right\rangle + \left\langle \frac{\partial L^i}{\partial x}(\varphi(t), \varphi'(t)), \psi(t) \right\rangle \right] dt = 0.$$

On en déduit que φ doit vérifier l'équation :

$$\boxed{-\frac{d}{dt} \left[\frac{\partial L^i}{\partial \beta}(\varphi(t), \varphi'(t)) \right] + \frac{\partial L^i}{\partial x}(\varphi(t), \varphi'(t)) = 0,}$$

pour tout t de $[0, T]$, si φ est de classe C^1 , et au sens des distributions, si φ est seulement absolument continue. Cette équation est l'équation d'Euler.

Remarquons que l'on retrouve, pour des champs de probabilités constants, le fait que les chemins optimaux entre deux points sont des segments : l'équation entraîne que $\frac{\partial L^i}{\partial \beta}(\varphi'(t))$ est constant or $\frac{\partial L^i}{\partial \beta}$ est une bijection de $\overset{\circ}{S}^i$ dans $\mathbb{R}^d$.

Avec davantage d'hypothèses sur L^i , on a des propriétés de régularité sur les solutions : on sait que si L^i est de classe C^r avec $r \geq 2$, puisque $\frac{\partial^2 L^i}{\partial \beta^2}(x, \beta)$ est définie positive pour tout (x, β) de $\bigcup_{\bar{x}} x \times \overset{\circ}{S}_x^i$, alors φ est de classe C^r et l'équation d'Euler est vérifiée en tout t de $[0, T]$ ([Li] ou [FR]). Les résultats d'existence seraient donc simples si le domaine de chaque L^i était tout $\mathbb{R}^d$ et non un compact.

Cas de deux champs

On cherche φ dans $C^1[0, T]$ minimisant $\Lambda_{0,T}(\varphi)$ en présence d'une frontière de discontinuité du champ de probabilités. Remarquons qu'il n'est pas rare en général qu'une trajectoire optimale suive cette frontière (cf cas des champs de probabilités constants).

Si on prend (t_j, i_j) associé à φ comme dans (3.31), on peut écrire une équation d'Euler sur chaque $[t_j, t_{j+1}]$, les instants t_j et les extrémités $x_j = \varphi(t_j)$ et $x_{j+1} = \varphi(t_{j+1})$ étant fixés. Si i_j vaut 0, on écrit que $\Lambda_{t_j, t_{j+1}}(\varphi)$ doit être minimale pour φ variant seulement dans l'ensemble des chemins tracés sur P^0 (le long desquels $\Lambda_{0,T}$ est égal à $\int \text{conv}(L_{\varphi(t)/\{\pi(\cdot) \geq 0\}}^1, L_{\varphi(t)/\{\pi(\cdot) \leq 0\}}^2) \varphi'(t) dt$) et parmi les $\Lambda_{t_j, t_{j+1}}^i(\psi)$ avec ψ à valeurs dans P^i et $\psi(t_j) = x_j$ et $\psi(t_{j+1}) = x_{j+1}$. Les dérivées $\frac{\partial L}{\partial \beta}$ et $\frac{\partial L}{\partial x}$ (on suppose qu'elles existent) sont prises dans $\mathbb{R}^{d-1}$ ($d-1$ dernières coordonnées). On fait ensuite varier les x_j et les t_j . En plus des équations d'Euler sur les $[t_j, t_{j+1}]$, on peut alors obtenir des conditions limites aux différents points t_j où φ est dérivable à gauche et à droite.

• Nous détaillons cette étude dans le cas où le minimum est atteint pour un chemin ayant une seule intersection avec P^0 . Considérons x_0 dans P^1 et y_0 dans P^2 et supposons que le minimum

$$\inf\{\Lambda_{0,T}(\varphi) \mid \varphi \in C[0, T], \varphi(0) = x_0, \varphi(T) = y_0\} \quad (4.4)$$

soit fini et atteint en φ ayant un seul point d'intersection avec P^0 , cela à un instant t .

La quantité

$$\inf\{\Lambda_{0,t}^i(\psi) \mid \psi \in C[0, t], \psi(0) = x_0, \psi(t) = \varphi(t)\}$$

est égale à $\Lambda_{0,t}^i(\varphi)$. On suppose que $s \mapsto (\varphi(\cdot), \varphi'(\cdot))$ est à valeurs dans un ouvert U tel que, pour $i = 1$ ou 2 , L^i soit de classe C^1 sur $U \cap \{(x, \beta) \mid x \in P^i\}$. Le chemin φ doit alors vérifier :

$$-\frac{d}{ds} \left[\frac{\partial L^1}{\partial \beta}(\varphi(s), \varphi'(s)) \right] + \frac{\partial L^1}{\partial x}(\varphi(s), \varphi'(s)) = 0, \text{ pour presque tout } s \text{ de } [0, t]. \quad (4.5)$$

De même, on doit avoir :

$$-\frac{d}{ds} \left[\frac{\partial L^2}{\partial \beta}(\varphi(s), \varphi'(s)) \right] + \frac{\partial L^2}{\partial x}(\varphi(s), \varphi'(s)) = 0, \text{ pour presque tout } s \text{ de } [t, T]. \quad (4.6)$$

Ces équations d'Euler sont obtenues en faisant varier φ à t et $\varphi(t)$ fixés. En déformant maintenant φ sans plus garder $\varphi(t)$ fixe, on va obtenir une équation au point t .

On ajoute $\varepsilon\psi$ à φ en prenant pour ψ l'application définie par $\psi(s) = s(T-s)\beta$, avec β fixé dans $\mathbb{R}^d$. Elle est de classe C^1 et vérifie $\psi(0) = \psi(T) = 0$.

Considérons d'abord un vecteur β tel que $\pi(\beta)$ soit nul. Cela revient à faire varier le point d'intersection de φ avec P^0 , en supposant que ce point reste atteint à l'instant t . Pour que le coût soit minimal en φ , la quantité

$$\int_0^t L^1(\varphi(s) + \varepsilon\psi(s), \varphi'(s) + \varepsilon\psi'(s))ds + \int_t^T L^2(\varphi(s) + \varepsilon\psi(s), \varphi'(s) + \varepsilon\psi'(s))ds \quad (4.7)$$

doit être minimale en $\varepsilon = 0$. Cette quantité a pour dérivée en 0 :

$$C_\beta(t) = \int_0^t \left[\left\langle \frac{\partial L^1}{\partial \beta}(\varphi(s), \varphi'(s)), \psi'(s) \right\rangle + \left\langle \frac{\partial L^1}{\partial x}(\varphi(s), \varphi'(s)), \psi(s) \right\rangle \right] ds \\ + \int_t^T \left[\left\langle \frac{\partial L^2}{\partial \beta}(\varphi(s), \varphi'(s)), \psi'(s) \right\rangle + \left\langle \frac{\partial L^2}{\partial x}(\varphi(s), \varphi'(s)), \psi(s) \right\rangle \right] ds.$$

Si on pose $A^i(s) = \frac{\partial L^i}{\partial x}(\varphi(s), \varphi'(s))$ et $B^i(s) = \frac{\partial L^i}{\partial \beta}(\varphi(s), \varphi'(s))$, d'après les deux premières équations (4.5) et (4.6), on a $A^1 = (B^1)'$ presque sûrement sur $[0, t]$ et $A^2 = (B^2)'$ presque sûrement sur $[t, T]$. La quantité $C_\beta(t)$ est donc égale à :

$$\int_0^t \left[\langle B^1(s), \psi'(s) \rangle + \langle (B^1)'(s), \psi(s) \rangle \right] ds + \int_t^T \left[\langle B^2(s), \psi'(s) \rangle + \langle (B^2)'(s), \psi(s) \rangle \right] ds = \\ \left\langle \lim_{\substack{s \rightarrow t \\ s < t}} B^1(s) - \lim_{\substack{s \rightarrow t \\ s > t}} B^2(s), \psi(t) \right\rangle,$$

d'après $\psi(0) = \psi(T) = 0$, d'où, pour tout β :

$$C_\beta(t) = \left\langle \frac{\partial L^1}{\partial \beta}(\varphi(t), \varphi'_g(t)) - \frac{\partial L^2}{\partial \beta}(\varphi(t), \varphi'_d(t)), \psi(t) \right\rangle \quad (4.8)$$

avec $\pi(\psi(t)) = 0$ pour tout t .

En considérant β tel que $\pi(\beta) = 0$, on montre donc que toutes les coordonnées de $\frac{\partial L^1}{\partial \beta}(\varphi(t), \varphi'_g(t))$ et $\frac{\partial L^2}{\partial \beta}(\varphi(t), \varphi'_d(t))$ sauf la première doivent être égales :

$$\text{pour } 2 \leq j \leq d, \quad \pi^j \left(\frac{\partial L^1}{\partial \beta}(\varphi(t), \varphi'_g(t)) \right) = \pi^j \left(\frac{\partial L^2}{\partial \beta}(\varphi(t), \varphi'_d(t)) \right) \quad (4.9)$$

En considérant β tel que $\pi(\beta) \neq 0$, on peut éventuellement obtenir d'autres équations. Supposons que φ est de classe C^1 à gauche et à droite de t et que $(\varphi(s), \varphi'(s))$ est à valeurs dans $P^1 \times \{\pi(\cdot) > 0\}$ pour presque tout s d'un intervalle $]t - \varepsilon_1, t[$ et dans $P^2 \times \{\pi(\cdot) < 0\}$ pour presque tout s dans $]t, t + \varepsilon_1[$ (et toujours $s \mapsto (\varphi(\cdot), \varphi'(\cdot))$ à valeurs dans un ouvert U convenable). Cette dernière hypothèse permet d'avoir

$\pi[\varphi'_d(t)]$ et $\pi[\varphi'_g(t)]$ strictement positifs.

Pour ε petit et sur un voisinage de t , $\varphi + \varepsilon\psi$ coupe une seule fois P^0 , à un instant $t + f(\varepsilon)$. Alors f est monotone et dérivable à gauche et à droite en t ; $f(0)$ vaut 0. Pour $\pi(\beta) < 0$, f est croissante, négative à gauche de 0 positive à droite; pour $\pi(\beta) > 0$, f est décroissante, positive à gauche de 0 et négative à droite. On a pour tout ε :

$$\pi[\varphi(t + f(\varepsilon))] + \varepsilon\pi[\psi(t + f(\varepsilon))] = 0.$$

D'après les hypothèses sur φ et ψ , on en déduit que, pour ε petit, $\varepsilon\pi[\psi(t)]$ est équivalent à $-\pi[\varphi'_d(t)]f(\varepsilon)$ pour $f(\varepsilon)$ positif et à $-\pi[\varphi'_g(t)]f(\varepsilon)$ pour $f(\varepsilon)$ négatif. Donc, pour $\pi(\beta) < 0$, $f'_g(0)$ vaut $-\frac{\pi[\psi(t)]}{\pi[\varphi'_g(t)]}$ et $f'_d(0)$ vaut $-\frac{\pi[\psi(t)]}{\pi[\varphi'_d(t)]}$. Ces deux valeurs sont échangées pour $\pi(\beta) > 0$. Pour que le coût soit minimal en φ , la quantité

$$\int_0^{t+f(\varepsilon)} L^1(\varphi(s) + \varepsilon\psi(s), \varphi'(s) + \varepsilon\psi'(s))ds + \int_{t+f(\varepsilon)}^T L^2(\varphi(s) + \varepsilon\psi(s), \varphi'(s) + \varepsilon\psi'(s))ds \quad (4.10)$$

doit être minimale en $\varepsilon = 0$. Cela est vérifié pour tout β tel que $\pi(\beta) < 0$ dès qu'on a $\varphi'_g(t) \notin S_{\varphi(t)}^2$ et pour tout β tel que $\pi(\beta) > 0$ si $\varphi'_d(t) \notin S_{\varphi(t)}^1$. Si ces deux conditions sont vraies, par exemples si $S_{\varphi(t)}^1$ et $S_{\varphi(t)}^2$ sont disjoints, on n'obtient pas de nouvelles équations.

Supposons $\varphi'_g(t) \in S_{\varphi(t)}^2$ et $\varphi'_d(t) \in S_{\varphi(t)}^1$ (cas où l'on obtient le plus d'équations). Alors la quantité (4.10) est dérivable à gauche et à droite en 0. Sa dérivée à gauche doit être négative. Elle vaut :

$$f'_g(0) \left[L^1(\varphi(t), \varphi'_g(t)) - L^2(\varphi(t), \varphi'_g(t)) \right] + C_\beta(t), \text{ pour } \pi(\beta) < 0$$

et

$$f'_g(0) \left[L^1(\varphi(t), \varphi'_d(t)) - L^2(\varphi(t), \varphi'_d(t)) \right] + C_\beta(t), \text{ pour } \pi(\beta) > 0$$

Sa dérivée à droite doit être positive. Elle vaut :

$$f'_d(0) \left[L^1(\varphi(t), \varphi'_d(t)) - L^2(\varphi(t), \varphi'_d(t)) \right] + C_\beta(t), \text{ pour } \pi(\beta) > 0$$

et

$$f'_d(0) \left[L^1(\varphi(t), \varphi'_g(t)) - L^2(\varphi(t), \varphi'_g(t)) \right] + C_\beta(t), \text{ pour } \pi(\beta) > 0.$$

Du cas $\pi(\beta) < 0$, on déduit que l'on doit donc avoir :

$$-\frac{\pi[\psi(t)]}{\pi[\varphi'_g(t)]} \left[L^1(\varphi(t), \varphi'_g(t)) - L^2(\varphi(t), \varphi'_g(t)) \right] + C_\beta(t) \leq 0$$

et

$$-\frac{\pi[\psi(t)]}{\pi[\varphi'_d(t)]} \left[L^1(\varphi(t), \varphi'_d(t)) - L^2(\varphi(t), \varphi'_d(t)) \right] + C_\beta(t) \geq 0,$$

alors que le cas $\pi(\beta) > 0$ conduit aux inégalités inverses. On en déduit que les deux quantités précédentes doivent être nulles. En reprenant la valeur de $C_\beta(t)$ dans (4.8), on obtient :

$$\frac{\pi[\psi(t)]}{\pi[\varphi'_g(t)]} \left[L^1(\varphi(t), \varphi'_g(t)) - L^2(\varphi(t), \varphi'_g(t)) \right] = \left\langle \frac{\partial L^1}{\partial \beta}(\varphi(t), \varphi'_g(t)) - \frac{\partial L^2}{\partial \beta}(\varphi(t), \varphi'_d(t)), \psi(t) \right\rangle$$

et

$$\frac{\pi[\psi(t)]}{\pi[\varphi'_d(t)]} \left[L^1(\varphi(t), \varphi'_d(t)) - L^2(\varphi(t), \varphi'_d(t)) \right] = \left\langle \frac{\partial L^1}{\partial \beta}(\varphi(t), \varphi'_g(t)) - \frac{\partial L^2}{\partial \beta}(\varphi(t), \varphi'_d(t)), \psi(t) \right\rangle.$$

On en déduit, en simplifiant par $\pi[\psi(t)]$, que l'on doit avoir, en plus de (4.9) :

$$\begin{aligned} \frac{L^1(\varphi(t), \varphi'_g(t)) - L^2(\varphi(t), \varphi'_g(t))}{\pi[\varphi'_g(t)]} &= \frac{L^1(\varphi(t), \varphi'_d(t)) - L^2(\varphi(t), \varphi'_d(t))}{\pi[\varphi'_d(t)]} \\ &= \pi \left[\frac{\partial L^1}{\partial \beta}(\varphi(t), \varphi'_g(t)) - \frac{\partial L^2}{\partial \beta}(\varphi(t), \varphi'_d(t)) \right] \end{aligned} \quad (4.11)$$

Les conditions (4.9) et (4.11) représentent au total $d + 1$ équations sur $\varphi(t)$, $\varphi'_g(t)$ et $\varphi'_d(t)$.

Dans le cas des champs constants les conditions s'obtiennent plus simplement, de la façon suivante : d'après les deux équations d'Euler, sur $[0, t]$ et $[t, T]$, on doit chercher

$$\inf \left\{ t L^1 \left(\frac{x - x_0}{t} \right) + (T - t) L^2 \left(\frac{y_0 - x}{T - t} \right) \mid x = (0, x_1, \dots, x_d), 0 \leq t \leq T \right\}$$

La dérivation par rapport à x donne (4.9). La dérivation par rapport à t (en supposant que les β^i sont à l'intérieur des supports) ajoute l'équation

$$L^1(\beta^1) - \left\langle \beta^1, \frac{\partial L^1}{\partial \beta}(\beta^1) \right\rangle - L^2(\beta^2) - \left\langle \beta^2, \frac{\partial L^2}{\partial \beta}(\beta^2) \right\rangle = 0$$

où on a posé

$$\beta^1 = \frac{x - x_0}{t} \quad \text{et} \quad \beta^2 = \frac{y_0 - x}{T - t}.$$

On doit donc avoir

$$H^1 \left(\frac{\partial L^1}{\partial \beta}(\beta^1) \right) = H^2 \left(\frac{\partial L^2}{\partial \beta}(\beta^2) \right).$$

Remarquons que l'on trouve (et c'est normal) à peu près les mêmes conditions quand on cherche, pour β donné, les vecteurs β^1 et β^2 en lesquels le coût $L_x(\beta)$ est atteint (en supposant que les β^i sont à l'intérieur des supports). En effet rappelons la valeur de $L_x(\beta)$:

$$L_x(\beta) = \inf \left\{ \sum_{i=1}^2 \rho^i L_x^i(\beta^i) \mid \sum_{i=1}^2 \rho^i = 1, \sum_{i=1}^2 \rho^i \beta^i = \beta, \pi(\beta^1) \geq 0, \pi(\beta^2) \leq 0 \right\}$$

On cherche le minimum de $(\rho, \beta^1) \mapsto \rho L_x^1(\beta^1) + (1 - \rho) L_x^2(\frac{\beta - \rho \beta^1}{1 - \rho})$ en supposant qu'il est atteint à l'intérieur de $[0, 1] \times S_x^1$. La dérivation par rapport à β^1 conduit à

$$\frac{\partial L^1}{\partial \beta}(x, \beta^1) = \frac{\partial L^2}{\partial \beta}(x, \beta^2)$$

et la dérivation par rapport à ρ à

$$H^1\left(\frac{\partial L^1}{\partial \beta}(\beta^1)\right) = H^2\left(\frac{\partial L^2}{\partial \beta}(\beta^2)\right).$$

On doit donc résoudre $H^1(\frac{\partial L^1}{\partial \beta}(\beta^1)) = H^2(\frac{\partial L^2}{\partial \beta}(\beta^2))$, en se limitant aux solutions β^1 pour lesquelles $\pi(\beta^1)$ est positif. Ensuite, si on pose $\beta^2 = (\frac{\partial L^2}{\partial \beta})^{-1} \circ (\frac{\partial L^1}{\partial \beta}(\beta^1))$, on doit avoir $\pi(\beta^2)$ négatif et $\beta - \beta^1$ et $\beta^2 - \beta^1$ colinéaires.

- Pour étudier un chemin de $C^1[0, T]$ qui minimise le coût entre deux points avec $\varphi([0, t])$ inclus dans P^1 et $\varphi([t, T])$ dans P^0 , on applique la même méthode. Pour obtenir l'équation d'Euler sur les intervalles où ce chemin suit la frontière on a besoin de la différentiabilité de la fonction L_x sur $\{\pi(\cdot) = 0\}$ pour x dans P^0 (on peut regarder si le théorème des fonctions implicites est vérifié aux points où le minimum de $(\rho, \beta^1) \mapsto \rho L_x^1(\beta^1) + (1 - \rho) L_x^2(\frac{\beta - \rho \beta^1}{1 - \rho})$ ci-dessus est atteint à l'intérieur de $[0, 1] \times S_x^1$). Dans ce cas, (4.9) est conservée entre L^1 et L sur $\{\pi(\cdot) = 0\}$. On peut ensuite chercher de la même façon que ci-dessus des conditions aux points où le chemin change d'ensemble P^i .

- Dans le cas de plusieurs traversées, on étudierait les équations d'Euler sur chaque intervalle où le chemin est dans un même P^i et les équations au niveau de chaque point de passage dans un autre P^i . Il conviendrait d'étudier la convexité du coût en fonction de ces points de passage. Les résultats obtenus pour des champs constants permettent de supposer qu'une trajectoire minimisante traverse un nombre fini de fois la frontière.

4.2.4 Equations vérifiées par le coût minimal entre deux points : équations d'Hamilton-Jacobi

On reprend le problème précédent, en s'intéressant cette fois aux équations vérifiées par le coût minimal entre deux points.

Cas d'un seul champ

A nouveau on étudie d'abord le cas d'un seul champ de probabilités (et horizon fini). On a le résultat suivant :

Théorème 4.2.3

On considère un champ de probabilités (μ_x) vérifiant les hypothèses (H). Supposons qu'il existe u de $\mathbb{R}^d \times \mathbb{R}$ dans $\mathbb{R}^+$, de classe C^1 , nul en $(x_0, 0)$ et vérifiant l'équation d'Hamilton-Jacobi suivante :

$$\frac{\partial u}{\partial t}(x, t) + H\left(x, \frac{\partial u}{\partial x}(x, t)\right) = 0.$$

Alors, on a :

$$u(x, t) = \inf\{\Lambda_{0,t}(\varphi) \mid \varphi \in C[0, T], \varphi(0) = x_0, \varphi(t) = x\}.$$

Démonstration (d'après [FW]) :

considérons un chemin absolument continu φ tel que $\varphi(0) = x_0$ et $\varphi(t) = x$. En utilisant que $L(\varphi(s), \varphi'(s))$ est supérieur à $\langle \alpha, \varphi'(s) \rangle - H(\varphi(s), \alpha)$ au point $\alpha = \frac{\partial u}{\partial x}(\varphi(s), s)$, on peut écrire :

$$\Lambda_{0,t}(\varphi) \geq \int_0^t \left\langle \frac{\partial u}{\partial x}(\varphi(s), s), \varphi'(s) \right\rangle ds - \int_0^t H\left(\varphi(s), \frac{\partial u}{\partial x}(\varphi(s), s)\right) ds$$

le membre de droite est égal à $\int_0^t \frac{d}{ds}[u(\varphi(s), s)] ds$ d'après l'équation d'Hamilton-Jacobi. On en déduit :

$$\Lambda_{0,t}(\varphi) \geq u(\varphi(t), t) - u(\varphi(0), 0) = u(x, t).$$

Considérons alors φ solution du système suivant :

$$\begin{cases} \varphi'(s) = \frac{\partial H}{\partial \alpha}\left(\varphi(s), \frac{\partial u}{\partial x}(\varphi(s), s)\right) \\ \varphi(0) = x_0 \end{cases}$$

Cette solution existe, d'après le théorème de Cauchy-Peano (par exemple dans [CL]), puisque l'application $(x, s) \mapsto \frac{\partial H}{\partial \alpha}(x, \frac{\partial u}{\partial x}(x, s))$ est continue (grâce à l'hypothèse (H1)).

D'après $H(\varphi(s), \cdot) = L(\varphi(s), \cdot)^*$ et les propriétés de différentiabilité des fonctions, on a :

$$H(\varphi(s), \alpha) = \left\langle \alpha, \frac{\partial H}{\partial \alpha}(\varphi(s), \alpha) \right\rangle - L\left(\varphi(s), \frac{\partial H}{\partial \alpha}(\varphi(s), \alpha)\right).$$

En utilisant cette relation pour $\alpha = \frac{\partial u}{\partial x}(\varphi(s), s)$, on obtient :

$$\begin{aligned} \frac{d}{ds} \left[u(\varphi(s), s) \right] &= \left\langle \frac{\partial u}{\partial x}(\varphi(s), s), \varphi'(s) \right\rangle + \frac{\partial u}{\partial s}(\varphi(s), s) \\ &= \left\langle \frac{\partial u}{\partial x}(\varphi(s), s), \frac{\partial H}{\partial \alpha} \left(\varphi(s), \frac{\partial u}{\partial x}(\varphi(s), s) \right) \right\rangle + \frac{\partial u}{\partial s}(\varphi(s), s) \\ &= L(\varphi(s), \varphi'(s)) + H\left(\varphi(s), \frac{\partial u}{\partial x}(\varphi(s), s)\right) + \frac{\partial u}{\partial s}(\varphi(s), s) \\ &= L(\varphi(s), \varphi'(s)) \end{aligned}$$

$$\text{Donc } \Lambda_{0,t}(\varphi) = \int_0^t \frac{d}{ds} \left[u(\varphi(s), s) \right] ds = u(\varphi(t), t).$$

On a prouvé successivement que $u(x, t)$ est majoré et minoré par $\inf\{\Lambda_{0,t}(\varphi) \mid \varphi \in C[0, T], \varphi(0) = x_0, \varphi(t) = x\}$; on en déduit le théorème. $\square$

Remarquons que pour le problème à horizon infini, l'équation devient :

$$H\left(x, \frac{\partial u}{\partial x}(x)\right) = 0$$

c'est-à-dire que si u est solution de cette équation, on a :

$$u(x) = \inf\{\Lambda_{T_1, T_2}(\varphi) \mid \varphi \in C[T_1, T_2], \varphi(T_1) = x_0, \varphi(T_2) = x, -\infty \leq T_1 < T_2 \leq +\infty\}.$$

On peut énoncer un résultat local ([Li],[Be]) de la façon suivante :

on prend u telle que $u(x, t) = \inf\{\Lambda_{t, T}(\varphi) \mid \varphi \in C[0, T], \varphi(t) = x, \varphi(T) = y_0\}$. Alors, si L et H sont continues en (t, x) et si u est différentiable en (t, x) , u vérifie l'équation d'Hamilton-Jacobi au point (t, x) . Cette équation est obtenue alors par un argument de programmation dynamique, en remarquant que, pour tout s de $[t, T]$, on a :

$$u(x, t) = \inf\{\Lambda_{t, s}(\varphi) + u(\varphi(s), s) \mid \varphi \in C[0, T], \varphi(t) = x, \varphi(T) = y_0\}.$$

L'existence (et l'unicité) de solutions aux deux équations d'Hamilton-Jacobi précédentes est assurée dès que H possède certaines propriétés ([Li],[Be]). L'étude de ces équations fait intervenir la théorie des solutions de viscosité, que nous n'aborderons pas. Dans notre cas, les hypothèses ne seront pas vérifiées au bord des S_x et

on aura probablement des résultats seulement si la trajectoire minimisante vérifie $\varphi'(t) \in S_{\varphi(t)}$.

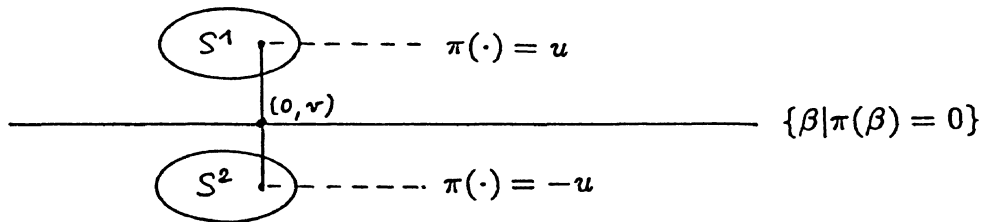
Cas de deux champs

On reprend le problème avec une frontière de discontinuité. Comme pour l'équation d'Euler, on écrit l'équation d'Hamilton-Jacobi entre des points appartenant à un même P^i . Entre deux points de P^0 , on utilise le même raisonnement que pour montrer le théorème 4.2.3, pour le coût mixte L , avec $(x_0, 0)$ et (x, t) sur P^0 et en se limitant aux chemins qui suivent P^0 . On doit prendre H telle que pour tout y de P^0 , on ait $H(y, \cdot) = L(y, \cdot)^*$ dans $\mathbb{R}^{d-1}$, c'est-à-dire d'après le théorème 1.4.8, $H(y, \cdot) = (\max\{\widehat{H}_y^1, \widehat{H}_y^2\})^{**}$. On a besoin de $H(y, \cdot)$ et $L(y, \cdot)$ strictement convexes, de classe C^1 sur l'intérieur de leur domaine avec $(y, \alpha) \mapsto \frac{\partial H}{\partial \alpha}(y, \alpha)$ continue, conditions qui ne sont pas toujours vérifiées. Si elles le sont, la solution de l'équation d'Hamilton-Jacobi

$$\frac{\partial u}{\partial t}(x, t) + \left(\max\{\widehat{H}_x^1, \widehat{H}_x^2\} \right)^{**} \left(\frac{\partial u}{\partial x}(x, t) \right) = 0. \quad (4.12)$$

donne le coût minimal.

Exemple: nous donnons un exemple où H vérifie toutes les propriétés. Supposons que les champs de probabilités sont constants avec S^1 inclus dans $\{\pi(\cdot) > 0\}$ et pour tout (u, v) de $\mathbb{R} \times \mathbb{R}^{d-1}$, $\mu^1(u, v) = \mu^2(-u, v)$ (il s'agit d'un cas où le coût mixte est égal à $\text{conv}(L^1, L^2)$; alors H est égal à $\max(H^1, H^2)$).



Alors on peut montrer que le coût mixte vaut, en un vecteur parallèle à P^0 , $L(0, v) = \inf_u L^1(u, v)$ et $+\infty$ ailleurs, et que $L(0, \cdot)$ est donc la transformée de Cramer de la probabilité ν^1 sur $\mathbb{R}^{d-1}$ définie par $d\nu^1(v) = \int_u d\mu^1(u, v)$. Sur $\{\pi(\cdot) = 0\}$, H est alors égale au logarithme de la transformée de Laplace de ν^1 ; elle a donc les mêmes propriétés que les H^i . On peut alors écrire le même résultat que dans le théorème 4.2.3, pour l'équation (4.12).

Enfin, comme dans le paragraphe précédent, les équations d'Hamilton-Jacobi permettent d'obtenir des équations aux points de passage entre différents P^i . Supposons que le coût minimal entre $x_0 \in P^1$ et $y_0 \in P^2$ soit atteint en φ ayant un seul point d'intersection avec P^0 , à un instant t . Si on note $V(x_0, y_0)$ ce coût, on a :

$$V(x_0, y_0) = \inf_{\substack{x \in P^0 \\ t \in [0, T]}} \left[u(x_0; x, t) + v(y_0; x, t) \right], \quad (4.13)$$

où on a posé :

$$u(x_0; x, t) = \inf \{ \Lambda_{0,t}^1(\varphi) \mid \varphi \in C[0, T], \varphi(0) = x_0, \varphi(t) = x \} \quad (4.14)$$

et

$$v(y_0; x, t) = \inf \{ \Lambda_{t,T}^2(\varphi) \mid \varphi \in C[0, T], \varphi(t) = x, \varphi(T) = y_0 \}. \quad (4.15)$$

On cherche des conditions pour que le minimum $V(x_0, y_0)$ soit atteint en un point (x, t) . Si $u(x_0; \cdot)$ et $v(y_0; \cdot)$ sont différentiables au point (x, t) , (4.14) et (4.15) impliquent :

$$\frac{\partial u}{\partial t}(x_0; x, t) + H^1\left(x, \frac{\partial u}{\partial x}(x_0; x, t)\right) = 0$$

et

$$-\frac{\partial v}{\partial t}(y_0; x, t) + H^2\left(x, -\frac{\partial v}{\partial x}(y_0; x, t)\right) = 0$$

D'autre part, pour que (4.13) soit minimale en (x, t) , il faut :

$$\begin{cases} \frac{\partial u}{\partial t}(x_0; x, t) + \frac{\partial v}{\partial t}(y_0; x, t) = 0 \\ \frac{\partial u}{\partial x}(x_0; x, t) + \frac{\partial v}{\partial x}(y_0; x, t) = 0 \end{cases}$$

L'ensemble des conditions au point (x, t) s'écrit donc :

$$\begin{cases} \frac{\partial u}{\partial x}(x_0; x, t) = -\frac{\partial v}{\partial x}(y_0; x, t) = \alpha \\ H^1(x, \alpha) = H^2(x, \alpha) \\ \frac{\partial u}{\partial t}(x_0; x, t) + \frac{\partial v}{\partial t}(y_0; x, t) = 0 \\ \frac{\partial u}{\partial t}(x_0; x, t) + H^1(x, \alpha) = 0 \end{cases}$$

Pour plusieurs traversées, à nouveau, on doit rassembler les équations correspondant aux intervalles où le chemin est dans un même ensemble P^i et les conditions aux points de passage entre les différents ensembles.

Table des matières

Introduction	5
1 Généralités : présentation du problème, outils	11
1.1 Deux types de résultats sur les processus d'apprentissage lents	12
1.1.1 Notations, rappel des résultats de [AR]	12
1.1.2 Rappel des résultats de [Du88]	16
1.2 Comparaison de ces résultats, discussion des hypothèses	17
1.2.1 Formulation de deux hypothèses sur les variations en x de la log-Laplace H_x	18
1.2.2 Lien entre la distance entre les lignes de niveau de la transfor- mée de Cramer L_x et les variations de la log-Laplace H_x	19
1.2.3 Liens de (HC) avec les hypothèses (H) et avec certaines de leurs conséquences	23
1.3 Outils d'analyse convexe : convexifiée d'un ensemble d'applications . .	29
1.4 Champs de probabilités discontinus le long d'une surface	33
1.4.1 Définition du processus	33
1.4.2 Cas particulier des champs constants, étude des vitesses moyennes asymptotiques ([DE])	35
1.4.3 Retour au cas général, conjectures sur les fonctions de coût et propriétés de ces fonctions	37
1.4.4 Limites de l'utilisation d'estimations antérieures à ce travail .	51
1.5 Conclusion, bilan des hypothèses faites sur le modèle	53
2 Grandes déviations trajectorielles : inégalité de majoration	59
2.1 Résultats sur les vitesses moyennes	60
2.2 Enoncé d'un premier théorème sur les trajectoires	66
2.2.1 Définition d'une fonctionnelle de coût approchée	67
2.2.2 Propriétés de la fonctionnelle de coût approchée $\Lambda_{0,T}^{\theta,h}$	69
2.2.3 Définition d'une fonctionnelle régularisée $\mathcal{L}$. Propriétés. Théo- rème avec cette fonctionnelle	71

2.3	Démonstration du théorème avec $\mathcal{L}$	73
2.3.1	Méthode	73
2.3.2	Théorème de majoration avec coût approché	76
2.3.3	Majoration finale: passage à la limite $(\theta, h) \rightarrow (0, 0)$	98
2.4	Calcul du coût régularisé	101
2.4.1	Calcul du coût régularisé $\mathcal{L}$ sur les chemins réguliers	101
2.4.2	Simplification de l'expression du coût régularisé	112
2.5	Nouvel énoncé	113
2.6	Cas où pour tout x , S_x^i est inclus dans l'ensemble $\{\beta (-1)^i \pi(\beta) > 0\}$	115
3	Grandes déviations trajectorielles: inégalité de minoration	121
3.1	Résultats sur les vitesses moyennes	123
3.1.1	Introduction	123
3.1.2	Minoration de la probabilité de suivre un vecteur parallèle à P^0 pour des points de départ Y_0^τ très proches de P^0	124
3.1.3	Minoration pour les vitesses moyennes en présence d'un seul champ de probabilités	138
3.1.4	Minoration de la probabilité de suivre un vecteur parallèle à P^0 avec une contrainte plus faible sur $\pi(Y_0^\tau)$	139
3.1.5	Probabilité d'être dans un demi-espace donné après n pas	143
3.2	Minoration de la probabilité d'un tube de chemins d'axe régulier tracé sur P^0	145
3.3	Minoration pour un tube d'axe un chemin continu, avec un coût approché	148
3.3.1	Etude des lignes de niveau du coût L , sous la condition (HLip)	148
3.3.2	Application aux chemins de $C^f[0, T]$	151
3.4	Minoration finale pour un tube: passage à la limite $\theta \rightarrow 0$	154
3.4.1	Résultat général	154
3.4.2	Cas où $\lim_{\theta \rightarrow 0} J_{0,T}^\theta(\varphi) = \Lambda_{0,T}(\varphi)$	161
3.4.3	Conclusion	164
3.5	Minoration pour un borélien quelconque	165
3.6	Démonstration du théorème 3.3.3	167
3.7	Cas des champs de probabilités constants, sous l'hypothèse (H_0)	176
3.8	Cas où l'hypothèse (H_0) n'est pas vérifiée	182
4	Conclusion et compléments	185
4.1	Résultats complets	185
4.1.1	Cas où l'hypothèse (H_0) est partout vérifiée	186

4.1.2	Cas où pour tout x , S_x^i est inclus dans l'ensemble $\{\beta (-1)^i \pi(\beta) > 0\}$	188
4.1.3	Cas limites (S_x^i tangent à $\{\beta \pi(\beta) = 0\}$)	189
4.2	Calcul du coût minimal entre deux points et des trajectoires minimisantes	191
4.2.1	Cas particulier des champs de probabilités constants	191
4.2.2	Existence de chemins minimisants	192
4.2.3	Equations vérifiées par les chemins minimisants entre deux points: équations d'Euler	195
4.2.4	Equations vérifiées par le coût minimal entre deux points: équations d'Hamilton-Jacobi	202

Bibliographie

- [AO] R. Anderson et S. Orey. *Small random perturbations of dynamical systems with reflecting boundary*. *Nagoya Math*, 5, 60, 189-216 (1976).
- [Az] R. Azencott. *Cours sur les grandes déviations à l'école d'été de St-Flour*. Springer-Verlag, Lecture notes in Mathematics, 774, 1-176 (1978).
- [AR] R. Azencott et G. Ruget. *Mélange d'équations différentielles et grands écarts à la loi des grands nombres*. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 38, Springer-Verlag, 1-54 (1977).
- [Be] S.H. Benton. *The Hamilton-Jacobi equation : a global approach*. Academic Press, New York (1977).
- [Ca] H. Cartan. *Formes différentielles, applications élémentaires au calcul des variations et à la théorie des courbes et des surfaces*. Coll. Méthodes. Hermann, Paris (1967).
- [CL] E.A. Coddington and N. Levinson. *Theory of Ordinary Differential Equations*. McGraw-Hill Book, New York (1955).
- [CFM] M. Cottrell, J.C. Fort et G. Malgouyres. *Large deviations and rare events in the study of stochastic algorithms*. I.E.E.E. Transactions on Automatic Control, Vol AC-28, N°9, September, 907-920 (1983).
- [DZ] A. Dembo, O. Zeitouni. *Large Deviations Techniques and Applications*. Jones and Bartlett Publishers (1992).
- [DS] J.-D. Deuschel and D. W. Stroock. *Large deviations*. Academic Press, San Diego, CA (1989).
- [Du88] P. Dupuis. *Large deviations analysis of some recursive algorithms with state dependent noise*. *The Annals of Probability*, Vol. 16, N° 4, 1509-1536 (1988).

- [Du92] P. Dupuis. *A Numerical Method for a Calculus of Variations Problem with Discontinuous Integrand*. In I. Karatzas and D. Ocone, editors, *Proceedings of the Rutgers Conference on Applied Stochastic Analysis*, Lecture Notes in Control and Information Science, Springer-Verlag, 90-107 (1992).
- [DE] P. Dupuis and R.S. Ellis. *Large deviations for Markov processes with discontinuous statistics, II: Random walks*. *Probab. Theory Relat. Fields*, 91, 153-194 (1992).
- [DEW] P. Dupuis, R.S. Ellis and Alan Weiss. *Large deviations for Markov processes with discontinuous statistics, I: General upper bounds*. *The Annals of Probability*, Vol. 19, N° 3, 1280-1297 (1991).
- [DIS] P. Dupuis, H. Ishii and H.M. Soner. *A viscosity solution approach to the asymptotic analysis of queuing systems*. *The Annals of Probability*, Vol. 18, N° 1, 226-255 (1990).
- [DK] P. Dupuis, H. Kushner. *Stochastic systems with small noise, analysis and simulation; a phase locked loop examples*. *Preprint* (1985).
- [FR] W.H. Fleming and R. Rishel. *Deterministic and stochastic optimal control*. Springer-Verlag, Berlin (1975).
- [FW] M.I. Freidlin and A.D. Wentzell. *Random Perturbations of Dynamical Systems*. Grundlehren der mathematischen Wissenschaften, 260, Springer-Verlag (1984).
- [IT] A.D. Ioffe and V.M. Tihomirov. *Theory of extremal problems*. Studies in Mathematics and its Applications, Vol. 6. North-Holland, Amsterdam, New York (1979).
- [Ku] H.J. Kushner. *Approximation and Weak convergence Methods for Random Processes with Applications to Stochastic System Theory*. MIT Press, Cambridge, MA (1984).
- [Li] P.L. Lions. *Generalized solutions of Hamilton-Jacobi equations*. Research Notes in Mathematics, 69. Pitman, Boston (1982).
- [Roc] R.T. Rockafellar. *Convex analysis*. Princeton University Press, Princeton (1970).
- [Rou] A. Rouault. *Espérances et majorations pour un processus de branchement spatial markovien*. *Annales de l'Institut Henri Poincaré*, Vol. 23, N°3, 459-497 (1987).

- [RZ] D.E. Rumelhart and D. Zipser. *Feature discovery by competitive learning*. In : D.E. Rumelhart and J.L. MacClelland, Eds., *Parallel Distributed Processing*, Chap. 5. MIT Press, Cambridge, MA (1986).
- [Va] S.R.S. Varadhan. *Large deviations and Applications*. S.I.A.M., Philadelphia (1984).
- [We,1] A.D. Wentzell. *Rough limit theorems on large deviations for Markov stochastic processes*. I. *Theory Probab. Appl.*, 21, 227-242 (1976).
- [We,2] A.D. Wentzell. *Rough limit theorems on large deviations for Markov stochastic processes*. II. *Theory Probab. Appl.*, 21, 499-512 (1976).
- [We,3] A.D. Wentzell. *Rough limit theorems on large deviations for Markov stochastic processes*. III. *Theory Probab. Appl.*, 24, 675-692 (1979).
- [We,4] A.D. Wentzell. *Rough limit theorems on large deviations for Markov stochastic processes*. IV. *Theory Probab. Appl.*, 27, 215-534 (1982).