Rapport de stage

Dimitri Faure

8 septembre 2021

1 Contexte du stage

A l'issu de mon Master 1, je suis arrivé à la conclusion que le domaine des mathématiques qui m'intéressait le plus et où je me sentais le plus à l'aise était celui des probabilités. J'ai donc contacté au début du mois de décembre 2020 mon professeur de processus stochastique, Giambattista Giacomin, pour qu'il me donne les coordonnées de chercheurs français travaillant hors de la région parisienne, prêts à me prendre en stage pour quatre mois. Je lui avais en particulier indiqué mon intérêt pour une présentation du chercheur Hubert Lacoin au thé du DMA, où ce dernier parlait de limites d'échelle continues et déterministes pour une dynamique aléatoire et discrète. Giambasttista m'a finalement transmis les coordonnées de plusieurs mathématiciens, dont celles de Cédric Bernardin, professeur à l'Univesité de Nice.

C'est donc ce dernier que j'ai contacté par mail mi-janvier pour lui demander de m'accorder un stage à partir de début mars. Il m'a aussitôt répondu qu'il m'accueillerait très volontier pour un stage encadré par lui et par un collègue spécialisé en physique mathématique, Raphaël Chetrite. Après un mois de repos à l'issue des mes examens de M1, j'ai donc pris le train de six heures pour Nice, profitant du trajet pour commencer à étudier le sujet de mon stage.

Les conditions dans lesquelles s'est déroulé mon stage étaient pour ainsi dire très confortables. Le choix définitif de ma destination fut, en partie, influencé par la présence à Nice d'un appartement avec vue sur le vieux port acheté par mes parents quelques mois auparavant. J'ai donc passé mes quatre mois de stage seul dans un environnement domestique très calme et agréable. A l'Université de Nice, j'ai eu la chance de me voir attribuer un bureau, que je partageais avec deux thésards très sympatiques, à quelques pas de celui de mes maîtres de stage. Mon bureau était plus précisément situé au cinquième étage du bâtiment Fizeau du campus Valrose, heureusement desservi par un ascenseur. Le campus Valrose en lui-même est absolument magnifique tant la végétation y est omniprésente. Je m'y rendais systématiquement en tram, le trajet de mon logement à mon bureau durant en tout et pour tout une trentaine de minutes. Quant à la ville de Nice, elle est toute à fait plaisante à vivre, les commerces en tout genre étant nombreux et il fut très facile d'accéder à des endroits bucoliques à pied depuis mon appartement, en particulier au parc du Mont Boron.

Vis-à-vis de mon organisation personnelle, je me rendais physiquement au laboratoire deux ou trois jours par semaine. J'arrivais en général vers 10h30-11h00 et partait vers 17h30-18h00 (je sais, dure journée). Cependant mes heures de travail réelles étaient assez aléatoires, et je pouvais très bien passer une journée entière au laboratoire sans rien produire avant de travailler sans discontinuité de 21h00 à 02h00 une fois rentré chez moi, travaillant plus quand l'inspiration venait qu'à des horaires fixes. J'ai ainsi obtenu plusieurs idées importantes en me promenant dans la forêt ou en marchant sur la plage. J'ai eu la grande chance d'avoir deux maîtres de stages à l'écoute et disponibles : je les voyais entre une et deux fois par semaine dans des rendez-vous pouvant durer 20 minutes comme 3 heures et obtenus à ma demande souvent du jour pour le lendemain. Au cours de ces discussions, je présentais en général mes nouvelles idées, les calculs que j'avais menés et posais des questions sur les objets avec lesquels je n'étais pas familier.

Enfin, à propos de la bibliographie lue pendant le stage, je dois reconnaître qu'elle est quasiment inexistante. L'objectif du stage fut de montrer un premier théorème, puis un deuxième, et je n'ai jamais eu besoin de chercher un résultat dans un livre ou un article mathématique : les preuves finales de mes résultats sont techniques mais complètement élémentaires. J'ai donc lu en tout et pour tout un cours d'introduction aux équations différentielles stochastiques de Bernard Ycart, des morceaux d'un cours plus conséquent sur le même sujet de Lawrence Evans, quelques chapitre de Multiscale Methods - Averaging and Homogenization de Grigorios Pavlotis et Andrew Stuart, et, évidemment, l'article de mes encadrants cité plus bas.

2 Déroulement du stage

Tous le travail que j'ai effectué durant mon stage visait à généraliser des résultats que mes encadrants avaient obtenus dans un article écrit conjointement avec des collègues de Toulouse : "Emergence of jumps in quantum trajectories via homogeneization". Il s'agit essentiellement de théorèmes d'homogénéisation : on prend une suite de processus stochastiques indexée par un paramètre γ réel positif, et on essaye d'exprimer la dynamique du système quand γ tend vers $+\infty$. Plus précisément, les principaux résultats obtenus au cours du stage sont essentiellement des reformulations dans des cadres particuliers du "méta-théorème" sur un espace d'état général E :

Theorem 1. *Soit $\mathcal{L}^{(0)}$ et $\mathcal{L}^{(1)}$ deux opérateurs, on note $\mathcal{L}_\gamma = \mathcal{L}^{(0)} + \gamma\mathcal{L}^{(1)}$, $\mathcal{P}$ le projecteur sur $\ker \mathcal{L}^{(1)}$ parallèlement à Im $\mathcal{L}^{(1)}$ et $\mathcal{L}_\infty = \mathcal{P}\mathcal{L}^{(0)}\mathcal{P}$. Nous avons alors pour tout $t > 0$:*

$$e^{t\mathcal{L}_\gamma} \underset{\gamma\to+\infty}{\longrightarrow} \mathcal{P}e^{t\mathcal{L}_\infty}\mathcal{P},$$

où $t \mapsto e^{t\mathcal{L}}$ est le semi-groupe du processus stochastique $(X_t)_{t\geq 0}$ de générateur $\mathcal{L}$, i.e. $e^{t\mathcal{L}}f(x) = \mathbb{E}_x(f(X_t))$ pour toute fonction test f et tout $x \in E$.

Ce théorème n'a de sens que si $\mathcal{L}^{(0)}$ et $\mathcal{L}^{(1)}$ ont une forme particulière et nous arrivons ainsi à la première situation étudiée par mes encadrants : si on prend $\mathcal{L}^{(0)}$ et $\mathcal{L}^{(1)}$ deux éléments de $\mathcal{M}_n(\mathbb{R})$ de valeur propre 1 et positifs en dehors de la diagonale, on peut associer à chacun d'entre eux une chaîne de Markov à temps continu qu'on appelle respectivement processus sous-dominant et processus dominant, les matrices $\left(e^{t\mathcal{L}}\right)_{t\geq 0}$ s'identifiant alors au semi-groupe du processus associé à $\mathcal{L}$. Dans ce cadre, mes maîtres de stage ont montré que le théorème 1 était vrai. D'un point de vue probabiliste, ce théorème signifie que quand γ tend vers l'infini, la variable aléatoire X_t^γ "converge en loi" vers $\overline{X}_t$ où le processus stochastique $(\overline{X}_t)_t$ est une chaîne de Markov d'espace d'états $\mathcal{R}_+$, l'ensemble des classes récurrentes positives de la chaîne de Markov associée à $\mathcal{L}^{(1)}$. Ainsi, notre processus complexe s'homogénéise en un processus plus simple quand γ tend vers l'infini.

La question que se posaient mes maîtres de stage était de savoir si le résultat restait vrai si on considérait une chaîne de Markov à espace d'états dénombrable, et ce problème a été présenté comme le premier objectif du stage. On observe d'abord que si la convergence énoncée dans le théorème 1 est selon n'importe quelle norme quand on considère des matrices de taille finie, il est important de determiner pour quelle topologie le théorème est vrai en dimension infinie. La convergence faible est le candidat naturel quand on étudie des semi-groupes, mais encore faut-il trouver quel est le bon espace de fonctions test.

Avant de s'attaquer au théorème pour n'importe quelles matrices de taille dénombrable $\mathcal{L}^{(0)}$ et $\mathcal{L}^{(1)}$, j'ai d'abord considéré le cas où $\mathcal{L}^{(0)}$ est nul, le théorème 1 devenant alors le théorème ergodique classique, mais avec potentiellement plusieurs classes récurrentes (positives ou nulles) et des états transients. En notant $\mathbb{P}_x(T = T_C)$ la probabilité que la première classe récurrente atteinte par un processus de générateur $\mathcal{L}^{(1)}$ démarrant en $x \in E$ soit C et μ_C la mesure invariante associée à la classe récurrente positive C, on a :

Theorem 2. *Soit $\mathcal{L}^{(1)} = \left(\mathcal{L}^{(1)}(x,y)\right)_{x,y\in E}$ un opérateur linéaire et f une fonction test sur l'espace d'état E. Alors pour tout $x \in E$ on a :*

$$e^{\gamma \mathcal{L}^{(1)}} f(x) \underset{\gamma \to +\infty}{\longrightarrow} \sum_{C \in \mathcal{R}_+} \mathbb{P}_x(T = T_C) \langle \mu_C, f \rangle.$$

J'ai alors observé que pour une chaîne de Markov quelconque, le théorème ne pouvait pas être vrai si l'on supposait que l'espace des fonctions test était $\ell^\infty(E)$, mais qu'il était en revanche vrai si l'espace des fonctions tests était $\ell^1(E)$. Ce constat était en réalité un piège qui m'a couté deux semaines de travail à chercher un théorème de convergence qui fonctionnerait pour n'importe quels $\mathcal{L}^{(1)}$ et $\mathcal{L}^{(0)}$ (en supposant juste qu'ils étaient bornés), théorème qui n'existe en fait pas comme je le découvrirai plus tard. En trouvant des contre exemples à ce théorème général s'il existait une classe récurrente nulle, j'ai compris qu'il fallait absolument mettre des hypothèses sur la forme de notre chaîne de Markov dominante. La bonne hypothèse est en faite que pour tout $x \in E$ on a

$$\sum_{x \in E} \mathbb{P}_x(T = T_C) = 1, \tag{1}$$

ou encore que notre processus dominant converge presque sûrement vers une classe récurrente positive en temps fini, ce qui implique en particulier l'absence de classe récurrente nulle. Sous ces hypothèse, on a en particulier que le théorème 2 est vrai pour toute fonction test dans $\ell^\infty(E)$. In fine, sous (1) et avec beaucoup de calculs (qui étaient d'ailleurs très différents de ceux réalisés par mes maîtres de stage dans le cas fini, qui eux reposaient sur de la réduction de Jordan), j'ai prouvé le théoème 1, trouvant en passant un espace d'état E, des opérateurs $\mathcal{L}^{(1)}$ et $\mathcal{L}^{(0)}$ bornés ainsi qu'une fonction $f \in \ell^\infty(E)$ à support fini tels que $e^{t\mathcal{L}_\gamma} f(x)$ ne convergeait pas pour un certain x, brisant là tout espoir d'obtenir le moindre théorème de convergence en affaiblissant (1). Après un mois et demi de stage j'avais donc une preuve du théorème 1 pour les chaînes de Markov à temps continu sous des hypothèses optimales.

C'est là que s'achève la première partie du stage. A l'issue de cette dernière, mes encadrants m'ont indiqué que dans leur article (dont je n'avais lu que les quelques pages sur le cas des chaînes de Markov à espace d'état fini), ils démontraient aussi le théorème 1 pour des opérateurs différentiels particuliers auxquels on associait des processus de diffusion sur des simplexes. Les hypothèses sur les générateurs étaient justifiées par le problème physique qui avait motivé le problème mathématique, mais il n'était pas du tout certain qu'elles étaient nécessaires dans un cadre purement abstrait. Mon travail était de trouver les hypothèses les plus générales possibles sur les opérateurs différentiels $\mathcal{L}^{(1)}$ et $\mathcal{L}^{(0)}$ sous-lesquelles le théorème 1 était encore vrai.

De façon plus précise on considère un compact K de $\mathbb{R}^n$ a priori quelconque et deux opérateurs différentiels $\mathcal{L}^{(1)}$ (dominant) et $\mathcal{L}^{(0)}$ (sous-dominant) sur $\mathcal{C}^\infty(K)$ de la forme :

$$\begin{cases} \mathcal{L}^{(0)} &= \langle b, \nabla_x \rangle + \frac{1}{2}\langle \sigma_0 \sigma_0^\dagger \nabla_x, \nabla_x \rangle \\ \mathcal{L}^{(1)} &= \frac{1}{2}\langle \sigma \sigma^\dagger \nabla_x, \nabla_x \rangle \end{cases} \tag{2}$$

où $\sigma : K \to \mathcal{M}_n(\mathbb{R})$ et $b : K \to \mathbb{R}^n$ sont des fonctions a priori quelconques. On peut associer au générateur $\mathcal{L}_\gamma = \mathcal{L}^{(0)} + \gamma \mathcal{L}^{(1)}$ un processus de diffusion $(X_t^\gamma)_{t \geq 0}$, et la première hypothèse que l'on fait sur ces opérateurs est que X_t^γ reste presque sûrement dans K pour des temps t arbitraires. Cette assomption se reformule explicitement en une condition simple sur les valeurs de b, σ, σ_0 sur les bords de K, la fonction σ etant en particulier nécessairement nulle sur les sommets de K.

Mes encadrants de stage avaient en fait démontré le théorème 1 dans le cas où K était un simplexe convexe (à $n+1$ sommets dans $\mathbb{R}^n$ donc), où b était affine, σ_0, σ étaient quadratiques et où le nombre de zéros de σ était exactement égal à $n+1$. L'interprétation du théorème est alors assez extraordinire : la variable aléatoire X_t^γ sur le simplexe convexe K, "converge en loi" quand γ tend vers l'infini vers $\overline{X}_t$ où $(\overline{X}_t)_{t \geq 0}$ est une chaîne de Markov dont l'espace d'état est l'ensemble des sommets du simplexe. J'ai in fine démontré le théorème 1 en supposant toujours que K était un simplexe convexe et que le cardinal de K_0 était $n+1$, mais en exigeant seulement de σ, σ_0 et b d'être lipschitziens (cette hypothèse est essentielle car sans elle on ne sait même pas si le processus est bien défini).

J'ai commencé par me familiariser pendant quelques semaines avec la théorie des diffusions que j'ai découverte avec de mon stage. On observait rapidement un grand nombre de similitudes entre l'étude du problème pour les chaînes de Markov et pour les diffusions, notamment vis-à-vis du théorème 2 qui se formulait exactement de la même manière dans les deux cas : pour les EDS, l'ensemble des classes récurrentes positives est exactement l'ensemble K_0 des zéros de σ, et les mesures invariantes associées sont les δ_z pour $z \in K_0$. Néanmoins on ne pouvait pas simplement adapter la preuve pour les chaînes de Markov aux diffusions pour une raison simple : dans le cas des chaînes de Markov, l'espace d'état est non compact, mais les opérateurs $\mathcal{L}^{(0)}$ et $\mathcal{L}^{(1)}$ sont bornés, alors que pour les diffusions, l'espace d'état est compact mais les opérateurs sont non-bornés.

Je ne pouvais pas vraiment lire aisément le papier de mes maîtres de stage qui utilisait un formalisme physique avec lequel je n'étais pas familier, mais ces derniers m'avaient indiqué qu'une des étapes cruciales de la preuve du théorème 2 était d'approximer la fonction test f par une fonction affine g ayant les mêmes valeurs sur les sommets du simplexe, c'est à dire sur K_0. L'existence d'une telle fonction g est garantie par le fait que K_0 est composé de $n+1$ points affinement indépendants dans un espace de dimension n. Cette approximation est justifiée car X_t^γ est, avec grande probabilité, dans n'importe quel voisinage de K_0 quand γ tend vers $+\infty$ (résultat que l'on appelle dans l'article le corollaire 1) : on a donc finalement $f(X_t^\gamma) \approx g(X_t^\gamma)$ avec une probabilité proche de 1 quand γ est suffisamment grand. La preuve du corollaire 1 repose sur le fait que, sur des temps très courts (pour s entre $t-h$ et t), X_s^γ se comporte comme le processus dominant de condition initiale X_{t-h}^γ entre 0 et γh. Comme le processus dominant converge presque sûrement (et uniformément dans notre cas) vers les sommets de K, on arrive à la conclusion que X_t^γ est avec une grande probabilité proche des sommets, quelle que soit la valeur de X_{t-h}^γ, pour γ suffisamment grand et h suffisamment petit.

Pour des fonctions test affines et un drift affine, le théorème 1 est complètement trivial à montrer et toute la difficulté réside dans la preuve du corollaire 1. Cependant, comme je supposais seulement que le drift était lipschitzien, je devais trouver un moyen de me ramener à un drift affine. J'ai donc considéré un processus fictif $(Y_t^\gamma)_{t\geq 0}$ de bruits σ et σ_0 mais de drift affine $\widetilde{b}$, la fonction $\widetilde{b}$ étant égal à b sur les sommets de K (nous utilisons encore une fois qu'il n'y a que $n+1$ sommets). J'ai enfin montré que les deux processus X^γ et Y^γ avaient des lois quasiment identiques quand γ était suffisamment grand, ce qui terminait la preuve du théorème 1 pour les diffusions.

J'étais donc parvenu à une preuve complète quand le bruit s'annulait en exactement $n+1$ points affinement indépendants. Mais, si j'utilisais abondamment cette hypothèse, je ne pouvais pas encore prouver qu'elle était nécessaire. Après plusieurs jours de travail, j'ai fini par trouver un contre-exemple au théorème avec un processus de diffusion dans le plan où σ s'annulait en quatre points : le compact K était un triangle convexe et les quatre éléments de K_0 était les trois sommets du triangle et un point intérieur.

On était arrivé à quelques jours de la fin du stage, et mes encadrants m'ont alors dit qu'avec la preuve du théorème 1 pour les diffusions et le contre-exemple, il était possible de publier un article dans une revue. Ils m'ont juste conseillé de recopier en adaptant les notations une partie de leur article qui montrait que la "convergence en loi" de X_t^γ vers $\overline{X}_t$ pour tout t impliquait la convergence du processus X^γ vers $\overline{X}$ pour une certaine topologie, dite de Meyer-Zheng. C'est alors qu'a commencé le travail de correction de la forme de l'article avant de l'envoyer à un journal, travail qui n'est pas encore complètement fini aujourd'hui. Je joins donc à ce rapport l'article dans sa forme presque finale, le texte jusqu'à la fin de la preuve du théorème 1 ayant été bien corrigé.

3 Conclusion

Je suis extrêmement satisfait de mon stage où j'ai pu découvrir ce qu'était la recherche mathématique sur deux problèmes que j'ai trouvés réellement intéressants, et que j'ai eu la chance de résoudre tous les deux exactement dans les temps. J'ai décidé pour plusieurs raisons de prendre une année de césure en 2021-2022, et mes maîtres de stage m'ont indiqué qu'ils seraient très heureux de tavailler avec moi et avec leurs collègues de Toulouse sur un autre article que nous écririons tous ensemble. Car, si j'ai pu répondre à plusieurs questions lors de mon stage, de nouvelles sont apparues lors de mes recherches. Le "méta-théorème" a déjà été énoncé par d'autres pour certains types de diffusions ou de chaînes de Markov, mais personne ne s'est encore attaqué au théorème dans sa forme la plus générale. Aussi bien dans le cas des chaînes de Markov que des diffusions, j'ai pu trouver des contre-exemples au théorème si on ne faisait pas d'hypothèses supplémentaires, ce qui prouve qu'un théorème complètement général aurait nécessairement une condition forte. Par ailleurs, si la condition que j'ai obtenue pour les chaînes de Markov me satisfait, je ne suis pas encore persuadé que celle que j'ai trouvée pour les diffusions soit optimale : tous mes contre-exemple sont dans des situations où le $(n+2)^{\text{ème}}$ élément de K_0 est dans l'enveloppe convexe des autres, ce qui ne serait pas le cas si on considérait, par exemple, un carré dans $\mathbb{R}^2$. Si l'on veut montrer le théorème général, il faudra donc trouver un moyen de relier le problème sur les chaînes de Markov à celui sur les processus de diffusion. A priori la tâche semble hardue, car les deux preuves que j'ai trouvées ont des structures radicalement différentes. Cependant, s'il est impossible d'adapter la preuve pour les chaîne de Markov à celle pour les diffusions, il n'est pas impossible que la méthode utilisée pour les EDS se généralise à d'autres situations. J'ai récemment isolé dans ma preuve pour les diffusions cinq étapes clés et quatre d'entre elles me semblent facilement réplicables dans le cadre de processus de sauts. La cinquième est de loin la plus technique et celle qui me semble la moins facile à adapter : la généraliser sera sûrement difficile mais c'est une piste qu'il faudra essayer de développer.

Averaging of semigroups associated to diffusion processes on a convex polytope

Dimitri Faure

September 8, 2021

1 Introduction

We consider a set of general markovian processes $(X_t^\gamma)_{\gamma>0}$ on a space E whose generators can be written $\mathcal{L}_\gamma = \mathcal{L}^{(0)} + \gamma\mathcal{L}^{(1)}$. Furthermore, for a given $x \in E$ and $\gamma > 0$, we write $(X_t^\gamma(x))_{t\geq 0}$ the stochastic process of generator $\mathcal{L}_\gamma$ starting from x. What we are looking for is an averaging result when γ tends to infinity, that is to say we want to understand the asymptotic behavior of our process. More precisely, we want to know, for f a test function, $x \in E$ and $t > 0$, if the semigroup of our process

$$e^{t\mathcal{L}_\gamma} f(x) = \mathbb{E}\left(f(X_t^\gamma(x))\right)$$

converges when γ tends to infinity, and in this case to what it converges.

We may actually guess the answer. A general theorem would indeed necessarily be of the form:

$$e^{t\mathcal{L}_\gamma} f(x) \underset{\gamma\to+\infty}{\longrightarrow} \mathcal{P}e^{t\mathcal{L}_\infty}\mathcal{P}f(x) \tag{1}$$

where $\mathcal{P}$ is the projector onto the kernel of $\mathcal{L}^{(1)}$, parallel to the image of $\mathcal{L}^{(1)}$ and $\mathcal{L}_\infty = \mathcal{P}\mathcal{L}^{(0)}\mathcal{P}$. When this quantity is well defined, one has :

$$\mathcal{P} = \lim_{t\to+\infty} \frac{1}{t}\int_0^t e^{t\mathcal{L}^{(1)}}.$$

Let us give a formal proof of (1).

We consider f a test function, and $f_t^\gamma = e^{t\mathcal{L}_\gamma} f$. One has that

$$\frac{\partial}{\partial t} f_t^\gamma = \left(\mathcal{L}^{(0)} + \gamma\mathcal{L}^{(1)}\right) f_t^\gamma \tag{2}$$

We look for a solution of this problem of the form

$$f_t^\gamma = f_t^{(0)} + \frac{1}{\gamma} f_t^{(1)} + O\left(\frac{1}{\gamma^2}\right). \tag{3}$$

By injecting (3) in (2) and equating coefficients in powers of $\frac{1}{\gamma}$ we find that for any $t > 0$:

$$\mathcal{L}^{(1)} f_t^{(0)} = 0 \tag{4}$$

$$\mathcal{L}^{(1)} f_t^{(1)} = \frac{\partial}{\partial t} f_t^{(0)} - \mathcal{L}^{(0)} f_t^{(0)} \tag{5}$$

Equation (4) gives us that $f_t^{(0)}$ is in the kernel of da a$\mathcal{L}^{(1)}$, so eventually

$$f_t^{(0)} = \mathcal{P} f_t^{(0)} \tag{6}$$

by the definition of $\mathcal{P}$.

We have also $\mathcal{P}\mathcal{L}^{(1)} = 0$ by the definition of $\mathcal{P}$, so applying $\mathcal{P}$ on the left on (5), we get, using (6) and $\mathcal{P}^2 = \mathcal{P}$, that:

$$0 = \mathcal{P}\mathcal{L}^{(1)} f_t^{(1)} = \mathcal{P}\frac{\partial}{\partial t} f_t^{(0)} - \mathcal{P}\mathcal{L}^{(0)} f_t^{(0)} = \frac{\partial}{\partial t}\mathcal{P} f_t^{(0)} - (\mathcal{P}\mathcal{L}^{(0)}\mathcal{P})\mathcal{P} f_t^{(0)}.$$

We thus have, using $f_0^{(0)} = f$, that:

$$f_t^{(0)} = \mathcal{P} f_t^{(0)} = \mathcal{P} e^{t\mathcal{L}_\infty} f = \mathcal{P} e^{t\mathcal{L}_\infty} \mathcal{P} f. \tag{7}$$

However, this is only a formal proof, and it is only in some specific situations that we have a rigorous proof of the theorem. For instance, the averaging problem has been widely studied for this stochastic differential equation on the torus $\mathbb{T}^\ell \times \mathbb{T}^{d-\ell}$:

$$\begin{cases} dX_t^\gamma &= f^X\left(X_t^\gamma, Y_t^\gamma\right) \mathrm{d}t + \sigma^X\left(X_t^\gamma, Y_t^\gamma\right) \mathrm{d}B_t \\ dY_t^\gamma &= \gamma f^Y\left(X_t^\gamma, Y_t^\gamma\right) \mathrm{d}t + \sqrt{\gamma}\sigma^Y\left(X_t^\gamma, Y_t^\gamma\right) \mathrm{d}W_t \end{cases} \tag{8}$$

where all the functions involved are smooth and B_t and W_t are independent Brownian motions.

We can here easily identify the subdominant process of generator:

$$\mathcal{L}^{(0)} = \langle f^X, \nabla_x \rangle + \frac{1}{2}\langle \sigma^X (\sigma^X)^T \nabla_x, \nabla_x \rangle$$

and the and the dominant one of generator:

$$\gamma\mathcal{L}^{(1)} = \gamma\langle f^Y, \nabla_y \rangle + \frac{1}{2}\gamma\langle \sigma^Y (\sigma^Y)^T \nabla_y, \nabla_y \rangle.$$

The second equation in (8), when we fix $X_t^\gamma = \xi$, define an SDE on $\mathbb{T}^{d-\ell}$ of invariant distribution $\mu_\xi(\mathrm{d}y) = \rho^\infty(y, \xi)\mathrm{d}y$.

When γ is big enough, for any $t > 0$ and a small amount of time $\mathrm{d}t$, one has that $X_{t+\mathrm{d}t}^\gamma$ is almost equal to $X_t^\gamma = \xi$, while the ergodic theorem on the dominant process gives us that the density of Y_t^γ converges to $\rho^\infty(y, \xi)\mathrm{d}y$ during $\mathrm{d}t$. For any $t > 0$, the density of Y_t^γ is therefore close to $\mu_t^Y(y) = \rho^\infty(y, X_t^\gamma)$ for γ large enough, and we thus only have to know the density of X_t^γ to understand our system.

Yet we have, under some assumptions, that $X_t^\gamma(x)$ converges for the L^p norm to $X_t(x)$ when γ tends to infinity, where $X_t(x)$ is solution of initial condition x of

$$\mathrm{d}X_t = F(X_t)\mathrm{d}t + A(X_t)\mathrm{d}B_t,$$

where

$$\begin{cases} F(\xi) &= \displaystyle\int_{\mathbb{T}^{d-l}} f^X(\xi, y)\rho^\infty(y, \xi)\mathrm{d}y \\ A(\xi)A(\xi)^T &= \displaystyle\int_{\mathbb{T}^{d-l}} \sigma^X(\xi, y)(\sigma^X(\xi, y))^T \rho^\infty(y, \xi)\mathrm{d}y. \end{cases}$$

The generator of this autonomous process is:

$$\overline{\mathcal{L}} = \langle F, \nabla_x \rangle + \frac{1}{2}\langle AA^T \nabla_x, \nabla_x \rangle,$$

therefore the density of X_t^γ is close to $\mu_t^X = \delta_x e^{t\overline{\mathcal{L}}}$ if the process starts from (x, y). We may notice that the starting point on the y-axis does not matter and that $\overline{\mathcal{L}}$ is perfectly defined for functions of domain only $\mathbb{T}^l$.

Hence, the averaging of our diffusion process on $\mathbb{T}^d$ is, in a certain way, a diffusion process on $\mathbb{T}^\ell$. Indeed, by the definition of $\mathcal{P}$ mentioned above, one has for any function f smooth on $\mathbb{T}^d$ that:

$$\mathcal{P}f(x, y) = \int_{\mathbb{T}^{d-l}} f(x, z)\rho^\infty(z, x)\mathrm{d}z,$$

that is independent of y, so eventually, for γ big enough:

$$\begin{aligned}
\mathbb{E}_{(x,y)}(f(X_t^\gamma, Y_t^\gamma)) &= \langle \delta_{(x,y)} e^{t\mathcal{L}_\gamma}, f \rangle \\
&\approx \int_{\mathbb{T}^\ell} \int_{\mathbb{T}^{d-\ell}} f(a, b)\mu_t^X(a)\mu_t^Y(b)\mathrm{d}a\mathrm{d}b \\
&= \int_{\mathbb{T}^\ell} \mu_t^X(a) \int_{\mathbb{T}^{d-\ell}} f(a, b)\rho^\infty(b, a)\mathrm{d}a\mathrm{d}b \\
&= \int_{\mathbb{T}^\ell} \mathcal{P}f(a)\mu_t^X(a)\mathrm{d}a \\
&= \mathbb{E}_x(\mathcal{P}f(X_t)) \\
&= \langle \delta_x e^{t\overline{\mathcal{L}}}, \mathcal{P}f \rangle,
\end{aligned}$$

that is independent of y. Using this formula, we get that we only have to study functions of the form $f = \mathcal{P}g$, so $\mathcal{P}f = f$ and therefore:

$$\begin{aligned}
\mathcal{L}_\infty f &= \mathcal{P}\mathcal{L}^{(0)}\mathcal{P}f \\
&= \mathcal{P}\mathcal{L}^{(0)} f \\
&= \mathcal{P}\langle f^X, \nabla_x \rangle f + \frac{1}{2}\langle \sigma^X (\sigma^X)^T \nabla_x, \nabla_x \rangle f \\
&= \langle \mathcal{P}f^X, \nabla_x \rangle f + \frac{1}{2}\langle \mathcal{P}\sigma^X (\sigma^X)^T \nabla_x, \nabla_x \rangle f \\
&= \langle F, \nabla_x \rangle f + \frac{1}{2}\langle AA^T \nabla_x, \nabla_x \rangle f \\
&= \overline{\mathcal{L}} f.
\end{aligned}$$

This proves (1) for our particular system.

However, in this case, the invariant measures associated to the dominant process have a specific form. Indeed the kernel of $(\mathcal{L}^{(1)})^\dagger$ is generated by the measures $(\rho(\cdot, x))_{x \in \mathbb{T}^l}$ of support $\mathbb{T}^{d-\ell}$, that is to say a continuous set of measures. More qualitatively, we can easily identify here a slow variable and a fast one (respectively X_t^γ and Y_t^γ).

Yet, it is sometimes impossible to identify slow and fast variables. When one considers a diffusion process on a convex polytope K of $\mathbb{R}^n$ where the dominant process is pure noise and is null only on a finite set of points K_0, then $(\mathcal{L}^{(1)})^\dagger$ is generated by the diracs associated to the points of K_0. The effective process we will get when γ approaches infinity cannot thus be a diffusion. In [mettre référence de votre article] is proven that for a specific geometry, assuming the subdominant drift is linear and the dominant and subdominant noises are quadratic, one has that the diffusion process $(X_t^\gamma)_{t\geq 0}$ converges for the Meyer-Zheng topology to a continuous-time Markov chain on K_0. The paper justifies these hypothesis by the fact that they are always fulfilled in the quantum mechanics phenomena that motivates the article.

We may however ask ourselves if these hypothesis are really optimal or if there is room for improvement. In this article we show that under a geometric assumption on K_0, this convergence result still holds if we only assume that the drift and the noises are lipschitz. We also give a counterexample of the theorem if this geometrical hypothesis is not verified, and thus prove the optimality of our conditions.

2 Main result

2.1 Notations

We will use the following notations:

- Let a, b be two vectors of $\mathbb{R}^n$, we write $a \cdot b$ the standart scalar product of a and b in $\mathbb{R}^n$. The euclidean norm associated to it is written $\|\cdot\|_2$.
- For $x \in \mathbb{R}^n$ and $\varepsilon > 0$, we write $B(x, \varepsilon)$ the ball of center x and of radius ε for the norm $\|\cdot\|_2$.
- $\mathcal{M}_n(\mathbb{R})$ is the set of real-valued $n \times n$-matrix.
- For a given matrix A of $\mathcal{M}_n(\mathbb{R})$, the write $A^\dagger$ the transpose of A.
- For a probability distribution μ on $\mathbb{R}^n$ and a bounded function f on $\mathbb{R}^n$, one writes:
$$\langle \mu, f\rangle = \int f \mathrm{d}\mu.$$
- Let $f \in \mathcal{C}^1(\mathbb{R}^n)$, we write $\nabla_x f = (\partial_i f)_{1\leq i\leq n}$ the gradient of f.
- Let $f \in \mathcal{C}^2(\mathbb{R}^n)$, we write $H_f = \left(\partial^2_{i,j} f\right)_{1\leq i,j\leq n}$ the hessian of f.
- For a vector b of $\mathbb{R}^n$, one write $\langle b, \nabla_x\rangle = \sum\limits_{i=1}^n b_i \partial_i$.

- For a matrix A of $\mathcal{M}_n(\mathbb{R})$, one writes $\langle A\nabla_x, \nabla_x\rangle = \sum\limits_{i,j=1}^{n} A_{i,j}\partial^2_{i,j}$

- For a given differential operator $\mathcal{L}$, one writes $t \mapsto e^{t\mathcal{L}}$ the semigroup associated to the Markov process of generator $\mathcal{L}$.

- For a given probability measure μ, a generator $\mathcal{L}$ and a time $t \geq 0$, we write $\mu e^{t\mathcal{L}}$ the law of the process of generator $\mathcal{L}$ at time t.

- Let $A = (a_{i,j})_{1\leq i,j\leq n}$ be in $\mathcal{M}_n(\mathbb{R})$, one writes $\mathbf{tr}(A) = \sum\limits_{k=1}^{n} a_{k,k}$.

- From now on K is a convex polytope of $\mathbb{R}^n$.

- For any continuous function f on K, we write $\|f\|_\infty = \sup\limits_{x\in K}|f(x)|$.

- $\mathbb{D}(\mathbb{R}_+, K)$ is the space of càdlàg functions from $\mathbb{R}_+$ to K.

2.2 Definitions

Let $\gamma \geq 0$, we study the solution $X_t^\gamma := (X_t^\gamma(x))_{t\geq 0}$ of the following stochastic differential equation on K:

$$dX_t^\gamma = \sqrt{\gamma}\sigma(X_t^\gamma)dW_t + b(X_t^\gamma)dt + \sigma_0(X_t^\gamma)dB_t, \quad X_0^\gamma(x) = x \in K \qquad (9)$$

where $\sigma, \sigma_0 : K \to \mathcal{M}_n(\mathbb{R})$ and $b : K \to \mathbb{R}^n$ are lipschitz functions on K and $(W_t)_{t\geq 0}$ and $(B_t)_{t\geq 0}$ are two independent Wiener processes on $\mathbb{R}^n$.

We assume (see Remark 3) that σ, σ_0 and b are such that $X_t^\gamma(x) \in K$ for any $t \geq 0$ and any $\gamma \geq 0$ and thus this is also true for $X_t := (X_t(x))_{t\geq 0}$ the solution of:

$$dX_t = \sigma(X_t)dW_t, \quad X_0(x) = x.$$

We furthermore assume for the rest of the article that σ is null only for a finite number of points and we denote $K_0 = \{x \in K,\ \sigma(x) = 0\}$.

The generator of (9) is $\mathcal{L}_\gamma = \gamma\mathcal{L}^{(1)} + \mathcal{L}^{(0)}$ where:

$$\begin{cases} \mathcal{L}^{(0)} & = \langle b, \nabla_x\rangle + \frac{1}{2}\langle \sigma_0\sigma_0^\dagger\nabla_x, \nabla_x\rangle \\ \mathcal{L}^{(1)} & = \frac{1}{2}\langle \sigma\sigma^\dagger\nabla_x, \nabla_x\rangle \end{cases} \qquad (10)$$

respectively the generator of the subdominant and dominant processes.

Remark 1. *Since K is compact and b, σ and σ_0 are lipschitz, we have that these functions are bounded on K.*

Remark 2. *Equation* (9) *is not completely general, since we assumed that the dominant process is pure noise. This implies in particular that the dominant process is a martingale.*

Remark 3. *The fact that for any $x \in K$ one has $X_t^\gamma(x) \in K$ and $X_t(x) \in K$ for arbitrary times implies several constraints on σ, σ_0 and b on the boundary of K. More precisely the process stays in K whatever the starting point is if and only if for any point x of a side of the polytope, $\sigma(x)$ and $\sigma_0(x)$ are parallel to this side and $b(x)$ is null or points to the interior of K. We have thus that σ and σ_0 are null on the vertices of K since they are there parallel to two different sides.*

Before studying X_t^γ we will first consider the dominant process X_t. In Theorem 1 is proved that for all x, the process $(X_t(x))_{t\geq 0}$ converges almost surely to a random variable $X_\infty := X_\infty(x) \in K_0$ as t goes to $+\infty$. We may thus consider for any $z \in K_0$ the function:

$$H_z : x \in K \longmapsto \mathbb{P}(X_\infty(x) = z) \in [0, 1]. \tag{11}$$

that gives the probability that $(X_t(x))_{t\geq 0}$ converges to z.

Now that all our objects are well defined we may state our ergodic theorem:

Theorem 1. *(Uniform Ergodic Theorem)*
We assume that for any $z \in K_0$, the function H_z is continuous.

Then, for any f lipschitz function on K, we have that:

$$e^{t\mathcal{L}^{(1)}} f \underset{t\to+\infty}{\longrightarrow} \mathcal{P}f,$$

where the convergence is uniform in x and the operator $\mathcal{P}$ is defined by:

$$\mathcal{P}f(x) = \sum_{z\in K_0} H_z(x)f(z). \tag{12}$$

The proof of this theorem is long and technical, we therefore postpone it to Appendix. However, the uniformity is fundamental for the derivation of our main theorem.

Remark 4. *(A crucial example)*
The continuity hypothesis in Theorem 1 is for example fulfilled in a specific case that is the one we are interested in in this article.

Let us assume that the set K_0 is composed of at most $n + 1$ points of $\mathbb{R}^n$ that are affinely independent. Then the functions H_z for $z \in K_0$ are affine functions. Indeed, since we have at most $n + 1$ affinely independent points on a n-dimensional vector space, we get that for any $z \in K_0$, there exists an affine function f_z such that for all $y \in K_0$:

$$f_z(y) = \begin{cases} 0 & \text{if } y \neq z, \\ 1 & \text{if } y = z. \end{cases}$$

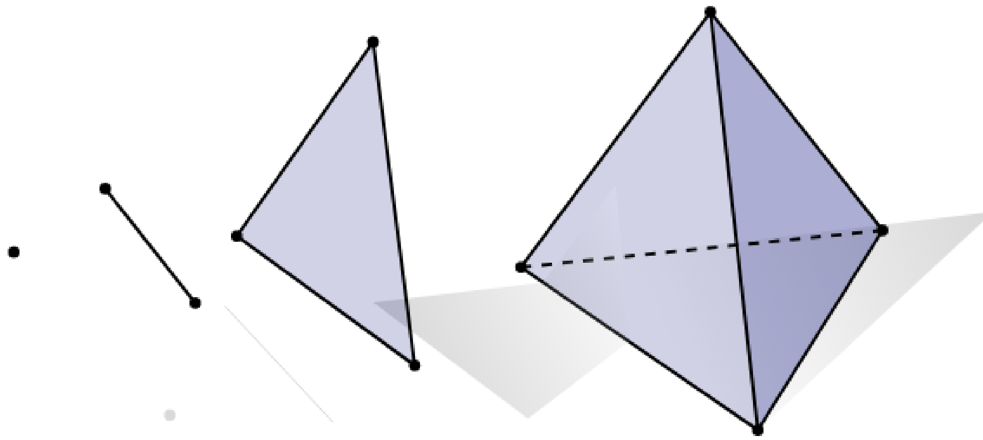


Figure 1: A polytope of $\mathbb{R}^n$ has at least $n+1$ vertices.

Since $(X_t(x))_{t\geq 0}$ is a martingale[1] *living in a bounded space, it converges almost surely to a random variable $X_\infty(x)$ taking its values in K_0 (see the proof of the uniform ergodic theorem in Appendix for more details). Thus for any $x \in K$:*

$$x = \mathbb{E}(X_0(x)) = \mathbb{E}(X_\infty(x)) = \sum_{y \in K_0} \mathbb{P}_x(X_\infty(x) = y)y.$$

Let $z \in K_0$, applying f_z to last equation and using that f_z is affine we get:

$$f_z(x) = f_z\left(\sum_{y \in K_0} \mathbb{P}_x(X_\infty = y)y\right) = \sum_{y \in K_0} \mathbb{P}_x(X_\infty = y) f_z(y) = \mathbb{P}_x(X_\infty = z) = H_z(x).$$

Furthermore, if we suppose that K_0 contains at most $n+1$ affinely independent points on a n-dimension space, we know actually what the points of K_0 are: it is exactly the extremal points of our polytope , since the noise is necessarily null on it (see Remark 3), and a polytope of $\mathbb{R}^n$ has at least $n+1$ extremal points.

2.3 Statements

Now that all the objects of the problem are now well define, we may state our first main theorem.

Theorem 2. *We assume that the cardinal of K_0 is inferior or equal to $n+1$ and the points of this set are affinely independent.*

[1]More exactly, each coordinate of $(X_t(x))_{t\geq 0}$ is a real martingale.

Then for any f lipschitz function on K and any probabilty measure μ on K, one has for $t > 0$:

$$\mathbb{E}_\mu \left(f\left(X_t^\gamma\right)\right) \underset{\gamma\to+\infty}{\longrightarrow} \mathbb{E}_{\overline{\mu}}\left(\overline{f}\left(\overline{X}_t\right)\right),$$

where $\overline{f}$ is the restriction of f to K_0 and $(\overline{X}_t)_{t\geq 0}$ is the jump Markov process of finite state space K_0, of generator $\overline{\mathcal{L}}$ and of initial distribution $\overline{\mu}$, with:

$$\overline{\mathcal{L}} = \left(b(x)\cdot\nabla_x H_z(x)\right)_{x,z\in K_0}, \quad \overline{\mu}(z) = \int_K H_z(x)\mathrm{d}\mu(x) \text{ for } z \in K_0.$$

Remark 5. *The geometrical hypothesis on K_0 may seem restrictive, but it is not clear what should be a more general statement. We will give in the last section a counter example of the theorem where $K \subset \mathbb{R}^2$ and $|K_0| = 4$.*

Theorem 2 provides the convergence of the semigroup $e^{t\mathcal{L}_\gamma}$ to the semigroup $e^{t\overline{\mathcal{L}}}$, that is to say the pointwise convergence in law of $(X_t^\gamma)_{t\geq 0}$ to $(\overline{X}_t)_{t\geq 0}$. It does not say anything about the convergence of $(X_t^\gamma)_{t\geq 0}$ to $(\overline{X}_t)_{t\geq 0}$ at the path level. One has for every $\gamma > 0$ that the paths of both processes $(X_t^\gamma)_{t\geq 0}$ and $(\overline{X}_t)_{t\geq 0}$ belong to $\mathbb{D}(\mathbb{R}_+, K)$. The natural topology of this space is the Skorokhod one, but we cannot in fact expect a weak convergence for this topology. Indeed as underlined in [Billingsey], weak convergence of processes with continuous paths in the Skorokhod topology yields a limiting process with continuous paths, and while for $\gamma > 0$ that $(X_t^\gamma)_{t\geq 0}$ has continuous paths almost surely, we have that $(\overline{X}_t)_{t\geq 0}$ does have discontinuous paths almost surely if b is non-null on every vertex. To overcome this difficulty, we use the idea of [votre papier] which is to replace the Skorokhod topology by the so-called Meyer-Zheng topology.

Let us define the Meyer-Zheng topology:

Definition. *Consider a Euclidean space $(E, \|\cdot\|)$ and denote by $\mathbb{L}^0 = \mathbb{L}^0(\mathbb{R}_+, E)$ the space of E-valued Borel functions on $\mathbb{R}_+$. Given a sequence $(w_\gamma)_{\gamma>0}$ of elements of $\mathbb{L}_0(\mathbb{R}_+, E)$, the following assertions are equivalent and define the convergence in Meyer-Zheng topology of $(w^\gamma)_{\gamma\geq 0}$ to $w \in \mathbb{L}^0(\mathbb{R}_+, E)$:*

- *For all bounded continuous functions $f : \mathbb{R}_+ \times E \to \mathbb{R}$:*

$$\lim_{\gamma\to+\infty}\int_0^{+\infty} f(t, w_t^\gamma)e^{-t}\mathrm{d}t = \int_0^{+\infty} f(t, w_t)e^{-t}\mathrm{d}t.$$

- *For $\lambda(\mathrm{d}t) = e^{-t}\mathrm{d}t$, we have for all $\varepsilon > 0$:*

$$\lim_{\gamma\to+\infty}\lambda(\{s \in \mathbb{R}_+ | \|w_s^\gamma - w_s\| \geq \varepsilon\} = 0.$$

- $\lim\limits_{\gamma\to+\infty} d(w^\gamma, w) = 0$ *where d is defined by:*

$$d(w, w') = \int_0^{+\infty}\left\{1 \wedge \|w_t - w'_t\|\right\}e^{-t}\mathrm{d}t.$$

The distance d metrizes the Meyer-Zheng topology on $\mathbb{L}^0$ and $(\mathbb{L}^0, d)$ is a Polish space.

We may now formulate our second main theorem:

Theorem 3. *Under the geometrical hypothesis of Theorem 2, we have that:*

$$\lim_{\gamma \to +\infty} X^\gamma = \overline{X}, \qquad \text{weakly in } \left(\mathbb{L}^0(\mathbb{R}_+, K), d\right)$$

where $(\overline{X}_t)_{t \geq 0}$ is the continuous-time jump Markov chain on K_0 defined in Theorem 2.

In other words, for all bounded continuous function $F : (\mathbb{L}^0(\mathbb{R}, K), d) \to (\mathbb{R}, |\cdot|)$, one has:

$$\mathbb{E}\left(F\left(X^\gamma\right)\right) \underset{\gamma \to +\infty}{\longrightarrow} \mathbb{E}\left(F\left(\overline{X}\right)\right).$$

3 Proof of Theorem 2 and Theorem 3

3.1 Proof of Theorem 2

Let us start with a quick sketch of the proof of Theorem 2.

We will first prove Lemma 1, a technical result that essentially says that for t small and γ big, our process of generator $\mathcal{L}_\gamma$ is basically the same as the one of generator $\mathcal{L}^{(1)}$. The keystone of our proof is however Corollary 1, that is proved using the uniform erdogic theorem combined with Lemma 1. Corollary 1 says that for a given t strictly positive, X_t^γ de facto lives in small balls centered around the $z \in K_0$ when γ is big enough. Therefore, when γ is big enough, it is like $(X_t^\gamma)_{t \geq 0}$ is jumping from a ball to another.

Corollary 1 is true under the geometrical hypothesis of Theorem 2, since it is based on the ergodic theorem which is only proved true under it (thanks to Remark 4). However if this hypothesis is sufficient, it is not necessary: Corollary 1 would still hold if we only knew that the functions H_z defined by (11) are continuous.

Nevertheless, if we know that there is a jump process on K_0, we a priori do not know its form. Theorem 2 says that, for a specific geometry, our process converges "weakly" to a continuous-time jump process with state space K_0 and generator $\overline{\mathcal{L}}$.

To prove that, we will first simplify our problem by approximating the drift b by an affine drift $\widetilde{b}$ such that $b_{|K_0} = \widetilde{b}_{|K_0}$. This approximation is justified by using Corollary 1. For the stochastic process $\widetilde{X}_t^\gamma$ associated to the linear drift, the proof of our theorem for an affine test function g but for a arbitrary initial condition μ is straightforward since we have thus that $\mathcal{L}_\gamma g = \mathcal{L}^{(0)} g$ is an affine function and $\mathcal{P}$ preserves affine functions. Hence we deduce the theorem for affine test functions even when the drift is not affine.

Then, using Corollary 1 again, we know that $(X_t^\gamma)_{t\geq 0}$ de facto lives in a neighborhood of K_0. Thus, instead of considering a generic test function f, we will consider an affine approximation g of f such that $f_{|K_0} = g_{|K_0}$. The function g exists thanks to our geometrical hypothesis.

Eventually, since the theorem is true for g and that one cannot distinguish $f(X_t^\gamma)$ from $g(X_t^\gamma)$ when γ is big enough, it will still holds for f.

Lemma 1. *Let f be a smooth function on K and μ be a probability measure on K. There exist C_1 and C_2 two positive constants independent of μ such that for any $t, \gamma, h > 0$, we have:*

$$\left| \langle \mu, e^{h\mathcal{L}_{\frac{\gamma}{h}}} f\rangle - \langle \mu, e^{\gamma \mathcal{L}^{(1)}} f\rangle \right| \leq C_1 (h + h^2)^{\frac{1}{2}} e^{C_2 \gamma t}.$$

Proof.
It is sufficient to prove that the inequality is fulfilled for $\mu = \delta_x$ for all $x \in K$.

Let x be be an arbitrary point of K and recall that $X_t^{\frac{\gamma}{h}} := X_t^\gamma(x)$ be the solution of

$$\mathrm{d}X_t^{\frac{\gamma}{h}} = b(X_t^{\frac{\gamma}{h}})\mathrm{d}t + \sigma_0(X_t^{\frac{\gamma}{h}})\mathrm{d}W_t + \sqrt{\frac{\gamma}{h}}\sigma(X_t^{\frac{\gamma}{h}})\mathrm{d}B_t,$$

with initial condition x.

We now fix γ.

We couple the process $(Z_t^h)_{t\geq 0} = (Z_t^h(x))_{t\geq 0}$ solution of:

$$Z_t^h = x + \frac{h}{\gamma}\int_0^t b(Z_u^h)\mathrm{d}u + \sqrt{\frac{h}{\gamma}}\int_0^t \sigma_0(Z_u^h)\mathrm{d}W_u + \int_0^t \sigma(Z_u^h)\mathrm{d}B_u,$$

with the process $(Z_t)_{t\geq 0} := (Z_t(x))_{t\geq 0}$ solution of:

$$Z_t = x + \int_0^t \sigma(Z_u)\mathrm{d}B_u,$$

through the common Brownian motion $(B_t)_t$ is the same in all these processes. By scaling invariance of the Brownian motion, the generator of $(Z_t^h)_{t\geq 0}$ is

$$\frac{h}{\gamma}\langle b(x), \nabla_x\rangle + \frac{1}{2}\frac{h}{\gamma}\langle \sigma_0\sigma_0^\dagger \nabla_x, \nabla_x\rangle + \frac{1}{2}\langle \sigma\sigma^\dagger \nabla_x, \nabla_x\rangle,$$

that is also the generator of $\left(X^{\frac{\gamma}{h}}_{\frac{h}{\gamma}\times t}\right)_{t\geq 0}$. Since they share the same initial distribution, we have for all $h>0$ that:

$$\left(X^{\frac{\gamma}{h}}_{\frac{h}{\gamma}\times t}\right)_{t\geq 0} \overset{\mathcal{L}}{=} (Z^h_t)_{t\geq 0},$$

which implies for $t=\gamma$ that:

$$X^{\frac{\gamma}{h}}_{h} \overset{\mathcal{L}}{=} Z^h_\gamma.$$

Thus using Itô isometry, the fact that b and σ_0 are bounded by a constant M and that σ is k-lipschitz, we have that:

$$\begin{aligned}
\mathbb{E}\left(\|Z^h_\gamma - Z_\gamma\|_2^2\right) &\leq \mathbb{E}\left(\|\int_0^\gamma (\sigma(Z^h_u)-\sigma(Z_u))\mathrm{d}B_u + \sqrt{\frac{h}{\gamma}}\int_0^\gamma \sigma_0(Z^h_u)\mathrm{d}W_u + \frac{h}{\gamma}\int_0^\gamma b(Z^h_u)\mathrm{d}u\|_2^2\right) \\
&\leq 3\mathbb{E}\left(\|\int_0^\gamma (\sigma(Z^h_u)-\sigma(Z_u))\mathrm{d}B_u\|_2^2\right) + 3\mathbb{E}\left(\|\frac{h}{\gamma}\int_0^\gamma b(Z^h_u)\mathrm{d}u\|_2^2\right) \\
&\quad + 3\mathbb{E}\left(\|\sqrt{\frac{h}{\gamma}}\int_0^\gamma \sigma_0(Z^h_u)\mathrm{d}W_u\|_2^2\right) \\
&\leq 3\mathbb{E}\left(\int_0^\gamma \|\sigma(Z^h_u)-\sigma(Z_u)\|_2^2\mathrm{d}u\right) + 3M^2\gamma^2\frac{h^2}{\gamma^2} \\
&\quad + 3\frac{h}{\gamma}\mathbb{E}\left(\int_0^\gamma \|\sigma_0(Z^h_u)\|_2^2\mathrm{d}u\right) \\
&\leq 3k^2\mathbb{E}\left(\int_0^\gamma \|Z^h_u - Z_u\|_2^2\right) + 3M^2(h+h^2).
\end{aligned}$$

From this point, Grönwall's inequality gives us that for all $h>0$ we have:

$$\mathbb{E}\left(\|Z^h_\gamma - Z_\gamma\|_2^2\right) \leq 3M^2(h+h^2)e^{3k^2\gamma}.$$

Thus, for f a smooth function, that is therefore L-Lipschitz on K, one has:

$$\begin{aligned}
\left|e^{h\mathcal{L}_{\frac{\gamma}{h}}}f(x) - e^{\gamma\mathcal{L}^{(1)}}f(x)\right| &= \left|\mathbb{E}(f(X^{\frac{\gamma}{h}}_h)) - \mathbb{E}(f(Z_\gamma))\right| \\
&\leq \mathbb{E}(\left|f(Z^h_\gamma)-f(Z_\gamma)\right|) \\
&\leq L\mathbb{E}\left(\|Z^h_\gamma - Z_\gamma\|_2\right) \\
&\leq L\mathbb{E}\left(\|Z^h_\gamma - Z_\gamma\|_2^2\right)^{\frac{1}{2}} \qquad \text{Using Cauchy-Schwarz} \\
&\leq \sqrt{3}LM(h+h^2)^{\frac{1}{2}}e^{\frac{3}{2}k^2\gamma}.
\end{aligned}$$

Since the constants are independent of x and γ, this proves the theorem for $C_1=\sqrt{3}LM$ and $C_2=\frac{3}{2}k^2$. □

Combining this result with the uniform ergodic theorem gives us this crucial result:

Corollary 1. *For any $t > 0$, $\eta > 0$ and μ probability measure on K, we have that:*

$$\mathbb{P}_\mu \left(X_t^\gamma \in \bigcup_{z \in K_0} B(z, \eta) \right) \underset{\gamma \to +\infty}{\longrightarrow} 1.$$

uniformally in μ.

Proof.
We consider a smooth function $f : K \to [0, 1]$ such that:

- $f = 1$ on $\bigcup_{z \in K_0} B(z, \frac{\eta}{2}) \cap K$.
- $f = 0$ outside of $\bigcup_{z \in K_0} B(z, \eta)$.

Then we have:

$$\langle \mu, e^{t\mathcal{L}_\gamma} f \rangle \leq \mathbb{P}_\mu \left(X_t^\gamma \in \bigcup_{z \in K_0} B(z, \eta) \right),$$

and thus it is sufficient to prove that the term on the left handside of the last display converges to 1.

For a given positive γ and a strictly positive t, we consider $\beta := \beta(\gamma, t), h := h(\gamma, t)$ such that:

$$\gamma = C_1 \beta^2 e^{2C_2 \beta t} \quad \text{and} \quad h = \frac{1}{C_1 \beta e^{2C_2 \beta t}}.$$

Thus one has that $\gamma = \dfrac{\beta}{h}$, that $\beta \underset{\gamma \to +\infty}{\longrightarrow} +\infty$ and that $h \underset{\gamma \to +\infty}{\longrightarrow} 0$.

We recall (11), i.e. that for all $z \in K_0$ and $x \in K$, $H_z(x) = \mathbb{P}(X_\infty(x) = z)$, and therefore $\sum_{z \in K_0} H_z = 1$. We also recall (12), i.e. that for $x \in K$, we have $\mathcal{P}f(x) = \sum_{z \in K_0} f(z) H_z(x)$. Eventually, since f is constant equal to 1 on K_0, we have that:

$$\begin{aligned}
\langle \mu, \mathcal{P}f \rangle &= \sum_{z \in K_0} f(z) \langle \mu, \mathcal{P}f \rangle \\
&= \sum_{z \in K_0} \langle \mu, H_z \rangle \\
&= \langle \mu, \sum_{z \in K_0} H_z \rangle \\
&= \langle \mu, 1 \rangle \\
&= 1,
\end{aligned}$$

for any μ probability measure on K. Thus one has that $\langle \mu e^{(t-h)\mathcal{L}_{\frac{\beta}{h}}}, \mathcal{P}f\rangle = 1$ for all γ so using Markov property:

$$
\begin{aligned}
|\langle \mu, e^{t\mathcal{L}_\gamma} f\rangle - 1| &= \left|\langle \mu, e^{t\mathcal{L}_\gamma} f\rangle - \langle \mu e^{(t-h)\mathcal{L}_{\frac{\beta}{h}}}, \mathcal{P}f\rangle\right| \\
&\leq \left|\langle \mu e^{(t-h)\mathcal{L}_{\frac{\beta}{h}}}, e^{h\mathcal{L}_{\frac{\beta}{h}}} f\rangle - \langle \mu e^{(t-h)\mathcal{L}_{\frac{\beta}{h}}}, e^{\beta\mathcal{L}^{(1)}} f\rangle\right| \\
&\quad + \left|\langle \mu e^{(t-h)\mathcal{L}_{\frac{\beta}{h}}}, e^{\beta\mathcal{L}^{(1)}} f\rangle - \langle \mu e^{(t-h)\mathcal{L}_{\frac{\beta}{h}}}, \mathcal{P}f\rangle\right|
\end{aligned}
$$

We now denote $\mu_\gamma = \mu e^{(t-h)\mathcal{L}_{\frac{\beta}{h}}}$ that is a probability measure. Then, using Lemma 1, we have that there exists C_1 and C_2 strictly positive constants independent of γ and μ such that:

$$
\begin{aligned}
\left|\langle \mu_\gamma, e^{h\mathcal{L}_{\frac{\beta}{h}}} f\rangle - \langle \mu_\gamma, e^{\beta\mathcal{L}^{(1)}} f\rangle\right| &\leq C_1(h+h^2)^{\frac{1}{2}} e^{C_2\beta t} \\
&\leq C_1\sqrt{h}e^{C_2\beta t} + C_1 h e^{C_2\beta t} \\
&\leq C_1\sqrt{\frac{1}{C_1\beta e^{2C_2\beta t}}}e^{C_2\beta t} + \frac{1}{\beta}e^{-C_2\beta t} \\
&\leq \sqrt{\frac{C_1}{\beta}} + \frac{1}{\beta} \\
&\underset{\gamma\to+\infty}{\longrightarrow} 0.
\end{aligned}
$$

On the other hand, since the convergence is uniform in x in the uniform ergodic theorem, we have:

$$
\left|\langle \mu_\gamma, e^{\beta\mathcal{L}^{(1)}} f\rangle - \langle \mu_\gamma, \mathcal{P}f\rangle\right| \leq \sup_{x\in K}\left|e^{\beta\mathcal{L}^{(1)}} f(x) - \mathcal{P}f(x)\right| \underset{\gamma\to+\infty}{\longrightarrow} 0,
$$

uniformally in μ. So eventually:

$$
\langle \mu, e^{t\mathcal{L}_\gamma} f\rangle \underset{\gamma\to+\infty}{\longrightarrow} 1,
$$

and the convergence is uniform in μ. This proves the corollary. □

Finally, to prove our theorem on affine functions for an affine drift, we need a last technical lemma:

Lemma 2. *We assume that $b : x \in \mathbb{R}^n \mapsto Cx + d \in \mathbb{R}^n$ where $C \in \mathcal{M}_n(\mathbb{R})$ and $d \in \mathbb{R}^n$. Then for any affine function $g : x \in \mathbb{R}^n \mapsto v \cdot x + l \in \mathbb{R}$ with $v \in \mathbb{R}^n$ and $l \in \mathbb{R}$, and any $t \geq 0$, one has that $e^{t\mathcal{L}_\gamma} f$ is affine. More precisely:*

$$
e^{t\mathcal{L}_\gamma} g : x \mapsto e^{tC^\dagger} v \cdot x + \left(v \cdot \int_0^t (e^{Cs}d)\mathrm{d}s + l\right) \tag{13}
$$

Proof.

We have that for all $x \in K$:

$$\begin{aligned} e^{t\mathcal{L}_\gamma} g(x) &= \mathbb{E}_x \left(g(X_t^\gamma) \right) \\ &= \mathbb{E}_x \left(v \cdot X_t^\gamma + l \right) \\ &= v \cdot \mathbb{E}_x(X_t^\gamma) + l. \end{aligned} \tag{14}$$

Applying $\mathbb{E}_x$ to the integral formulation of (9), we get:

$$\begin{aligned} \mathbb{E}_x(X_t^\gamma) &= x + \int_0^t \mathbb{E}_x(b(X_t^\gamma))\mathrm{d}s \\ &= x + \int_0^t \mathbb{E}(CX_t^\gamma + d)\mathrm{d}s \\ &= x + dt + C\int_0^t \mathbb{E}_x(X_s^\gamma)\mathrm{d}s. \end{aligned}$$

The associated differential equation is:

$$\frac{\partial}{\partial t}\mathbb{E}_x(X_t^\gamma) = C\mathbb{E}_x(X_s^\gamma) + d, \quad \mathbb{E}_x(X_0^\gamma) = x. \tag{15}$$

We can easily check that the unique solution to (15) is:

$$\mathbb{E}_x(X_t^\gamma) = e^{Ct}x + \int_0^t (e^{Cs}d)\mathrm{d}s \tag{16}$$

Combining (14) and (16), we eventually get:

$$e^{t\mathcal{L}_\gamma} g(x) = v \cdot \left(e^{tC}x\right) + v \cdot \int_0^t (e^{Cs}d)\mathrm{d}s + l$$

□

Remark 6. *Let us consider an affine function g, equation* (13) *gives $e^{t\mathcal{L}^{(1)}} g = g$ for all $t \geq 0$. When t approaches infinity, the uniform ergodic theorem implies therefore that $\mathcal{P}g = g$, that is to say $\mathcal{P}$ preserves linear functions.*

We may now prove our main theorem by following the path we mentioned earlier:

Proof. (Theorem 2)
We have that b is lipschitz. Since the cardinal of K_0 is by hypothesis at most $n+1$ in a n-dimensional vector space, there exists a matrix $C = (c_{i,j})_{1\leq i,j\leq n}$ and a vector d such that the affine maping:

$$\widetilde{b} : x \in \mathbb{R}^n \mapsto Cx + d \in \mathbb{R}^n,$$

satisfies $\widetilde{b}(z) = b(z)$ for all $z \in K_0$.

We will now consider the system of equations:

$$\begin{cases} X_t^\gamma & = Z_0 + \int_0^t b(X_t^\gamma)\mathrm{d}t + \int_0^t \sigma_0(X_t^\gamma)\mathrm{d}B_t + \sqrt{\gamma}\int_0^t \sigma(X_t^\gamma)\mathrm{d}W_t \\ Y_t^\gamma & = Z_0 + \int_0^t \widetilde{b}(Y_t^\gamma)\mathrm{d}t + \int_0^t \sigma_0(Y_t^\gamma)\mathrm{d}B_t + \sqrt{\gamma}\int_0^t \sigma(Y_t^\gamma)\mathrm{d}W_t \end{cases}$$

where the two independent Brownian motions B_t and W_t are the same for the two processes and Z_0 is a random variable of law μ. By hypothesis, X_t^γ stays in K for arbitrary times, so Remark 3 gives us that $b(z)$ points to the interior of K for all $z \in K_0$. But K is a convex polytope, so each side of it is the convex envelope of its vertices, and the value of $\widetilde{b}$ on a point x of this side is a convex combinaison of the values of b on its vertices: $\widetilde{b}(x)$ therefore also points to the interior of K. Thus, Remark 3 gives us that Y_t^γ stays in K for arbitrary times.

We will denote the generators of these two processes by $\mathcal{L}_\gamma = \mathcal{L}^{(0)} + \gamma\mathcal{L}^{(1)}$ and $\mathcal{G}_\gamma = \mathcal{G}^{(0)} + \gamma\mathcal{G}^{(1)}$.

Let us now consider an affine function g. For any $\gamma > 0$ and $t > 0$ we have, using Lemma 2, that $e^{t\mathcal{G}_\gamma}g$ is affine since $\widetilde{b}$ is affine. Hence $\mathcal{G}_\gamma e^{t\mathcal{G}_\gamma}g = \mathcal{G}^{(0)}e^{t\mathcal{G}_\gamma}g$ and:

$$\frac{\partial}{\partial t}e^{t\mathcal{G}_\gamma}g = \mathcal{G}_\gamma e^{t\mathcal{G}_\gamma}g = \mathcal{G}^{(0)}e^{t\mathcal{G}_\gamma}g = \mathcal{P}\mathcal{G}^{(0)}\mathcal{P}e^{t\mathcal{G}_\gamma}g$$

since $\mathcal{P}$ preserves linear functions (see Remark 6).

Thus, writing $\mathcal{G}_\infty = \mathcal{P}\mathcal{G}^{(0)}\mathcal{P}$, we have that:

$$e^{t\mathcal{G}_\gamma}g = e^{t\mathcal{G}_\infty}g = \mathcal{P}e^{t\mathcal{G}_\infty}g.$$

We denote $g_t = e^{t\mathcal{G}_\infty} g$ for $t \geq 0$. We have for $z \in K_0$ that:

$$\begin{aligned}
\frac{\partial}{\partial t} g_t(z) &= \mathcal{P}\mathcal{G}^{(0)}\mathcal{P} g_t(z) \\
&= \sum_{y \in K_0} H_z(y) \mathcal{G}^{(0)} \mathcal{P} g_t(y) \qquad \text{Using (12)} \\
&= \mathcal{G}^{(0)} \mathcal{P} g_t(z) \\
&= \widetilde{b}(z) \cdot \nabla_x \mathcal{P} g_t(z) \\
&= \widetilde{b}(z) \cdot \nabla_x \left(\sum_{y \in K_0} H_x(y) g_t(y) \right)(z) \\
&= \sum_{y \in K_0} \left(\widetilde{b}(z) \cdot \nabla_x H_x(y) \right) g_t(y) \\
&= \left(\widetilde{b}(x) \cdot \nabla_x H_x(y) \right)_{x,y} (\widetilde{g}_t(y))_y(z) \\
&= \overline{\mathcal{G}} g_t(z).
\end{aligned}$$

Eventually we have that $\widetilde{g}_t(z) = e^{t\widetilde{\mathcal{G}}} \overline{g}(z)$ for all $z \in K_0$, where we remind that $\overline{g} = g_{|K_0}$.

Finally we have for all probability measure μ:

$$\langle \mu, e^{t\mathcal{G}_\gamma} g \rangle = \langle \mu\mathcal{P}, e^{t\mathcal{G}_\infty} g \rangle = \langle \overline{\mu}, e^{t\overline{\mathcal{G}}} \overline{g} \rangle$$

However, if we look at the definition of $\overline{\mathcal{L}}$ and $\widetilde{\mathcal{G}}$ in Theorem 2, we notice that only the value of the drift on the $z \in K_0$ matters, therefore we have actually that $\overline{\mathcal{L}} = \widetilde{\mathcal{G}}$ so:

$$\langle \overline{\mu}, e^{t\overline{\mathcal{L}}} \overline{g} \rangle = \langle \overline{\mu}, e^{t\widetilde{\mathcal{G}}} \overline{g} \rangle.$$

Thus, to prove that, for any affine function g, we have:

$$\langle \mu, e^{t\mathcal{L}_\gamma} g \rangle \underset{\gamma \to +\infty}{\longrightarrow} \langle \overline{\mu}, e^{t\overline{\mathcal{L}}} \overline{g} \rangle,$$

we have to show that:

$$\left| \langle \mu, e^{t\mathcal{L}_\gamma} g \rangle - \langle \mu, e^{t\mathcal{G}_\gamma} g \rangle \right| = \left| \langle \mu, e^{t\mathcal{L}_\gamma} g \rangle - \langle \overline{\mu}, e^{t\widetilde{\mathcal{G}}} \overline{g} \rangle \right| \underset{\gamma \to +\infty}{\longrightarrow} 0.$$

First of all, let us use Dynkin's formula for $g : x \mapsto e_i \cdot x$. We have thus $\nabla_x g = e_i$

and $H_g = 0$ so, writing $X_t^\gamma = ([X_t^\gamma]_i)_{1\leq i\leq n}$ and $Y_t^\gamma = ([Y_t^\gamma]_j)_{1\leq j\leq n}$ we have:

$$\begin{aligned}
\mathbb{E}_\mu\left([X_t^\gamma]_i - [Y_t^\gamma]_i\right) &= \mathbb{E}_\mu\left(g(X_t^\gamma)\right) - \mathbb{E}_\mu\left(g(Y_t^\gamma)\right) \\
&= \langle \mu, g\rangle - \langle \mu, g\rangle + \mathbb{E}\left(\int_0^t \left((b(X_s^\gamma)\cdot\nabla_x g(X_s^\gamma) - \widetilde{b}(Y_s^\gamma)\cdot\nabla_x g(Y_s^\gamma)\right)\mathrm{d}s\right) \\
&+ \frac{1}{2}\mathbb{E}\left(\int_0^t (\mathbf{tr}\left(\sigma(X_s^\gamma)\sigma(X_s^\gamma)^\dagger H_g(X_s^\gamma)^\dagger\right) - \mathbf{tr}\left(\sigma(Y_s^\gamma)\sigma(Y_s^\gamma)^\dagger)H_g(Y_s^\gamma)^\dagger\right)\mathrm{d}s\right) \\
&+ \frac{1}{2}\mathbb{E}\left(\int_0^t (\mathbf{tr}\left(\sigma_0(X_s^\gamma)\sigma_0(X_s^\gamma)^\dagger H_g(X_s^\gamma)^\dagger\right) - \mathbf{tr}\left(\sigma_0(Y_s^\gamma)\sigma_0(Y_s^\gamma)^\dagger)H_g(Y_s^\gamma)^\dagger\right)\mathrm{d}s\right) \\
&= \mathbb{E}_\mu\left(\int_0^t (b(X_s^\gamma) - \widetilde{b}(Y_s^\gamma))\cdot e_i \mathrm{d}s\right) \\
&= \mathbb{E}_\mu\left(\int_0^t (\widetilde{b}(X_s^\gamma) - \widetilde{b}(Y_s^\gamma))\cdot e_i \mathrm{d}s\right) + \mathbb{E}_\mu\left(\int_0^t (b(X_s^\gamma) - \widetilde{b}(X_s^\gamma))\cdot e_i \mathrm{d}s\right) \\
&= \mathbb{E}_\mu\left(\int_0^t \sum_{j=1}^n c_{i,j}([X_t^\gamma]_j - [Y_t^\gamma]_j)\mathrm{d}s\right) + \mathbb{E}_\mu\left(\int_0^t (b(X_s^\gamma) - \widetilde{b}(X_s^\gamma))\cdot e_i \mathrm{d}s\right) \\
&= \sum_{j=1}^n c_{i,j}\int_0^t \mathbb{E}_\mu\left([X_t^\gamma]_j - [Y_t^\gamma]_j\right)\mathrm{d}s + \int_0^t \mathbb{E}_\mu\left((b(X_s^\gamma) - \widetilde{b}(X_s^\gamma))\cdot e_i\right)\mathrm{d}s.
\end{aligned}$$

Thus we have:

$$\begin{aligned}
\left|\mathbb{E}_\mu\left([X_t^\gamma]_i - [Y_t^\gamma]_i\right)\right| &\leq \sum_{j=1}^n |c_{i,j}|\int_0^t |\mathbb{E}_\mu([X_s^\gamma]_j - [Y_s^\gamma]_j)|\,\mathrm{d}s + \int_0^t \left|\mathbb{E}_\mu((b(X_s^\gamma) - \widetilde{b}(X_s^\gamma))\cdot e_i)\right| \\
&\leq \sum_{j=1}^n |c_{i,j}|\int_0^t \sup_{1\leq k\leq n}|\mathbb{E}_\mu([X_s^\gamma]_k - [Y_s^\gamma]_k)|\,\mathrm{d}s + \sup_{1\leq k\leq n}\int_0^t \left|\mathbb{E}_\mu((b(X_s^\gamma) - \widetilde{b}(X_s^\gamma))\cdot e_k)\right| \\
&\leq \sum_{l,j=1}^n |c_{i,j}|\int_0^t \sup_{1\leq k\leq n}|\mathbb{E}_\mu([X_s^\gamma]_k - [Y_s^\gamma]_k)|\,\mathrm{d}s + \sup_{1\leq k\leq n}\int_0^t \left|\mathbb{E}_\mu((b(X_s^\gamma) - \widetilde{b}(X_s^\gamma))\cdot e_k)\right|
\end{aligned}$$

We notice that i does not appear in the last expression, thus we have:

$$\sup_{1\leq k\leq n}\left|\mathbb{E}_\mu\left([X_t^\gamma]_k - [Y_t^\gamma]_k\right)\right| \leq \|C\|_1\int_0^t \sup_{1\leq k\leq n}|\mathbb{E}_\mu([X_s^\gamma]_k - [Y_s^\gamma]_k)|\,\mathrm{d}s + \sup_{1\leq k\leq n}\int_0^t \left|\mathbb{E}_\mu((b(X_s^\gamma) - \widetilde{b}(X_s^\gamma)\cdot e_k)\right|.$$

We notice that for $t \in [0, T]$, we have

$$\sup_{1\leq k\leq n}\int_0^t \left|\mathbb{E}_\mu((b(X_s^\gamma) - \widetilde{b}(X_s^\gamma)\cdot e_k)\right| \leq C_T^\gamma$$

where

$$C_T^\gamma = \sup_{1\leq k\leq n}\int_0^T \left|\mathbb{E}_\mu\left((b(X_s^\gamma) - \widetilde{b}(X_s^\gamma))\cdot e_k\right)\right|. \tag{17}$$

Finally, writing $g_\gamma : t \in \mathbb{R}_+ \mapsto \sup_{1\le k\le n} \left|\mathbb{E}_\mu(X^\gamma_{t,k} - Y_{t,k})\right| \in \mathbb{R}_+$, we have, for $t \in [0,T]$, that:

$$g_\gamma(t) \le \|C\|_1 \int_0^t g_\gamma(s)\mathrm{d}s + C^\gamma_T,$$

so Grönwall's lemma gives that for $t \in [0,T]$:

$$0 \le g_\gamma(t) \le C^\gamma_T e^{\|C\|_1 t}. \tag{18}$$

But Corollary 1 gives that for any $s > 0$, we have that X^γ_s lives around points of K_0 when γ is big enough, and we have defined $\widetilde{b}$ in such a way that $\widetilde{b}_{|K_0} = b_{|K_0}$. Eventually we have for all $s > 0$:

$$\left|\mathbb{E}_\mu\left((b(X^\gamma_s) - \widetilde{b}(X^\gamma_s)) \cdot e_k\right)\right| \underset{\gamma\to+\infty}{\longrightarrow} 0.$$

This quantity is bounded by a constant (that is integrable on $[0,T]$), by definition of C^γ_T in (17) the dominated convergence theorem gives us eventually:

$$C^\gamma_T \underset{\gamma\to+\infty}{\longrightarrow} 0.$$

Therefore, by (18) we finally have:

$$g_\gamma(t) \underset{\gamma\to+\infty}{\longrightarrow} 0.$$

Thus, for $g : x \in \mathbb{R}^n \mapsto v \cdot x + l \in \mathbb{R}$, where $v = (v_i)_{1\le i\le n} \in \mathbb{R}^n$ and $e \in \mathbb{R}$, one has:

$$\begin{aligned}
\left|\langle \mu, e^{t\mathcal{L}_\gamma} g\rangle - \langle \mu, e^{t\mathcal{G}_\gamma} g\rangle\right| &= |\mathbb{E}_\mu(g(X^\gamma_t) - \mathbb{E}_\mu(g(X^\gamma_t))| \\
&= |\mathbb{E}(X^\gamma_t - Y^\gamma_t) \cdot v| \\
&= \left|\sum_{1\le i\le n} v_i \mathbb{E}([X^\gamma_t]_i - [Y^\gamma_t]_i)\right| \\
&= \sup_{1\le i\le n} |\mathbb{E}([X^\gamma_t]_i - [Y^\gamma_t]_i)| \times \sum_{1\le j\le n} |v_j| \\
&= g_\gamma(t)\|v\|_1 \\
&\underset{\gamma\to+\infty}{\longrightarrow} 0
\end{aligned} \tag{19}$$

Furthermore, since the result of Corollary 1 is uniform in μ, this convergence result is uniform in μ for g affine and $t > 0$ given.

Eventually, we have that for any affine function g:

$$\begin{aligned}
\langle \mu, e^{t\mathcal{L}_\gamma} g\rangle &\underset{\gamma\to+\infty}{\longrightarrow} \langle \overline{\mu}, e^{t\overline{\mathcal{L}}}\overline{g}\rangle \\
&= \sum_{z\in K_0} \sum_{y\in K_0} \overline{\mu}(y) e^{t\overline{\mathcal{L}}}(y,z)\overline{g}(z).
\end{aligned}$$

We will now use the geometrical hypothesis of the theorem to conclude the proof for a lipschitz test function f. Since we have at most $n+1$ independent points in a n-dimensional vector space, there exists an affine function g such that $f_{|K_0} = g_{|K_0}$.

We thus have $\overline{f} = \overline{g}$, so in particular we have proved that:

$$\langle \mu, e^{t\mathcal{L}_\gamma} g \rangle \underset{\gamma \to +\infty}{\longrightarrow} \langle \overline{\mu}, e^{t\overline{\mathcal{L}}} \overline{g} \rangle = \langle \overline{\mu}, e^{t\overline{\mathcal{L}}} \overline{f} \rangle.$$

Therefore if we prove that:

$$\left| \langle \mu, e^{t\mathcal{L}_\gamma} f \rangle - \langle \mu, e^{t\mathcal{L}_\gamma} g \rangle \right| \underset{\gamma \to +\infty}{\longrightarrow} 0,$$

we will be done with the theorem.

Let $\varepsilon > 0$, we consider $\eta > 0$ sufficiently small in order that the balls $B(z, \eta)$ for $z \in K_0$ are disjoint and such that:

$$\sup_{z \in K_0} \sup_{x \in B(z,\eta)} |f(x) - g(x)| \leq \varepsilon.$$

On the other hand, Corollary 1 gives that for γ big enough

$$\mathbb{P}_\mu \left(X_t^\gamma \notin \bigcup_{z \in K_0} B(z, \eta) \right) \leq \varepsilon.$$

Eventually, we have for γ big enough that:

$$\begin{aligned}
\left| \langle \mu, e^{t\mathcal{L}_\gamma} f \rangle - \langle \mu, e^{t\mathcal{L}_\gamma} g \rangle \right| &= |\mathbb{E}_\mu((f-g)(X_t^\gamma))| \\
&\leq (\|f\|_\infty + \|g\|_\infty) \mathbb{P}_\mu \left(X_t^\gamma \notin \bigcup_{z \in K_0} B(z, \eta) \right) \\
&\quad + \sum_{z \in K_0} \mathbb{P}_\mu \left(X_t^\gamma \in B(z, \eta) \right) \sup_{x \in B(z,\eta)} |f(x) - g(x)| \\
&\leq (\|f\|_\infty + \|g\|_\infty)\varepsilon + \varepsilon.
\end{aligned}$$

This proves the theorem. □

3.2 Proof of Theorem 3

Now that our Theorem 2 is proven, let us give a sketch of the proof of Theorem 3.

We know that for all $\gamma > 0$ our trajectory $(X_t^\gamma)_{t \geq 0}$ are elements of $\mathbb{D}$, so we have a powerful criterion to show that this family of processes is tight in $\mathbb{L}^0(\mathbb{R}, K)$. Thus, Prokhorov's theorem gives us that the set of the laws of theses processes is a relatively compact subset of the space of probability measures on $\mathbb{L}^0(\mathbb{R}, K)$

for the topology of weak convergence.

To prove Theorem 3, we therefore just have to show that if a subsequence of $((X_t^\gamma)_{t\geq 0})_{\gamma>0}$ weakly converges to $\beta \in \mathbb{L}^0$, then its limit is necessarily $(\overline{X}_t)_{t\geq 0}$. It is in fact enough to prove that that for any $r \in \mathbb{N}^*$ and for almost any sequence $0 \leq t_1 \leq \ldots \leq t_r < +\infty$ one has that the laws of $(\overline{X}_{t_1}, \ldots, \overline{X}_{t_r})$ and $(\beta_{t_1}, \ldots, \beta_{t_r})$ are the same.

Let us start with the following result:

Theorem 4. *[mettre référence] Let E be an Euclidean space. If X is an E-valued stochastic process with natural filtration $(\mathcal{F}_t, t \geq 0)$, then for any $\tau \in \mathbb{R}_+$, its conditional variation on $[0,\tau]$ is defined as:*

$$V_\tau(X) := \sup_{0=t_0<t_1<\ldots<t_k=r} \sum_{i=0}^{k-1} \mathbb{E}\left(\|\mathbb{E}(X_{t_{i+1}} - X_{t_i} | \mathcal{F}_{t_i})\|\right)$$

Consider an index set I and a family $(X^\gamma, \gamma > 0)$ of processes living in $\mathbb{D}(\mathbb{R}_+, E)$ which satisfy:

$$\sup_{\gamma>0} \left| V_\tau(X^\gamma) + \mathbb{E}\left(\sup_{0\leq t\leq \tau} X_t^\gamma\right)\right| < +\infty,$$

for all $\tau > 0$. Then the family of laws of the X^γ is tight for the Meyer-Zheng topology and all the limiting points are supported in $\mathbb{D}(\mathbb{R}_+, E)$.

The previous theorem is the main tool for proving:

Lemma 3. *The family of processes:*

$$(X_t^\gamma)_{t\geq 0}, \gamma > 0$$

is tight in the Polish space $(\mathbb{L}^0(\mathbb{R}_+, K), d)$ and all limiting points are supported on càdlàg paths.

Proof.
Since for any $\gamma > 0$ and any $t \geq 0$, one has $X_t^\gamma \in K$ almost surely where K is compact, we only have to prove that the conditional variation in Meyer-Zheng's Theorem 4 is bounded on segments. For fixed $\tau > 0$, we have in our case:

$$V_\tau(X^\gamma) := \sup_{0=t_0<t_1<\ldots<t_k=r} \sum_{i=0}^{k-1} \mathbb{E}\left(\|\mathbb{E}(X^\gamma_{t_{i+1}} - X^\gamma_{t_i} | \mathcal{F}_{t_i})\|\right).$$

Equivalently, thanks to [mettre référence] and following paragraph,

$$V_\tau(X^\gamma) = \sup_{\|\varphi\|_2\leq C_K} \int_0^\tau \mathbb{E}\left(\langle \varphi_t, \mathrm{d}X_t^\gamma\rangle\right)$$

wher the supremum is taken over the simple predictable process taking value in the ball of center 0 and of radius $C_K := \sup_{z\in K}\|z\|_2$. It follows from (9), using that the mean of a brownian motion is equal to 0, that:

$$V_\tau(X^\gamma) = \sup_{\|\varphi\|\leq C_K} \int_0^\tau \mathbb{E}(\langle \varphi_t, b(X_t^\gamma)\rangle)\mathrm{d}t.$$

So Cauchy-Schwarz inequality gives us:

$$V_\tau(X^\gamma) \leq C_K \int_0^\tau \mathbb{E}(\|b(X_t^\gamma)\|_2)\mathrm{d}t.$$

Since b is bounded by M, this quantity is trivially bounded by $C_K \times M \times \tau$ for all $\gamma > 0$, and we have our result using Theorem 4. ☐

Now that we proved that $(X^\gamma)_{\gamma>0}$ is tight, we only have to show that if a subsequence converges, it converges to $\overline{X}$ to conclude. So from now on we assume that $(X^{\gamma_p})_{p\in\mathbb{N}}$ converges weakly to some β, where $(\gamma_p)_{p\in\mathbb{N}}$ is an unbounded sequence in $]0,+\infty[$. Theorem 4 gives us furthermore that the trajectories of β are almost surely càdlàg.

In order to prove that the law of β is the law of $\overline{X}$, we will in fact show that almost all their finite-dimensional distributions are the same, that is to say that for all $r \in \mathbb{N}^*$, for almost all $0 \leq t_1 \leq \ldots \leq t_r < +\infty$ and for all f continuous function on K^r one has:

$$\mathbb{E}\left(f(\beta_{t_1}, \ldots, \beta_{t_r})\right) = \mathbb{E}\left(f(\overline{X}_{t_1}, \ldots, \overline{X}_{t_r})\right).$$

Let us start with a linearization trick:

Lemma 4. *For all f continuous on K^r, there exists $F : (\mathbb{R}^n)^r \longrightarrow \mathbb{R}$ a r-linear function such that:*

$$\lim_{p\to+\infty} \int_{[0,+\infty[^r} \left|\mathbb{E}\left(f(X_{t_1}^{\gamma_p}, \ldots, X_{t_r}^{\gamma_p})\right) - \mathbb{E}\left(F(X_{t_1}^{\gamma_p}, \ldots, X_{t_r}^{\gamma_p})\right)\right| \lambda^{\otimes r}(\mathrm{d}t_1, \ldots, \mathrm{d}t_r) = 0$$

where we remind that $\lambda(\mathrm{d}t) = e^{-t}\mathrm{d}t$.

Proof. Let us first introduce the function F. We write $z_0, \ldots, z_n$ the elements of K_0 and for all $\mathbf{i} = (i_1, \ldots, i_r) \in [\![0,n]\!]^r$ we write $F_\mathbf{i} = f(z_{i_1}, \ldots, z_{i_r})$. We notice that for all $(x_1, \ldots, x_r) \in \{z_i,\ i \in [\![0,n]\!]\}^r$ one has:

$$\begin{aligned} f(x_1, \ldots, x_r) &= \sum_{\mathbf{i}\in[\![0,n]\!]^r} F_\mathbf{i} \prod_{k=1}^r \mathbb{1}_{z_{i_k}}(x_k) \\ &= \sum_{\mathbf{i}\in[\![0,n]\!]^r} F_\mathbf{i} \prod_{k=1}^r f_{z_{i_k}}(x_k) \\ &=: F(x_1, \ldots, x_r), \end{aligned}$$

where the functions $f_{z_{i_k}}$ are the one mentioned in Remark 4. Since they are all linear, the function F is clearly a continuous r-linear map. Now [mettre référence] applied to f and F says exactly that:

$$\lim_{p\in\mathbb{N}} \int_{[0,+\infty[^r} |\mathbb{E}(f(X^{\gamma_p}_{t_1},\dots,X^{\gamma_p}_{t_r})) - \mathbb{E}(f(\beta_{t_1},\dots,\beta_{t_r}))| \lambda^{\otimes r}(\mathrm{d}t_1,\dots,\mathrm{d}t_r) = 0 \tag{20}$$

and:

$$\lim_{p\in\mathbb{N}} \int_{[0,+\infty[^r} |\mathbb{E}(F(X^{\gamma_p}_{t_1},\dots,X^{\gamma_p}_{t_r})) - \mathbb{E}(F(\beta_{t_1},\dots,\beta_{t_r}))| \lambda^{\otimes r}(\mathrm{d}t_1,\dots,\mathrm{d}t_r) = 0 \tag{21}$$

We will now invoke an immediate consequence of Corollary 1: for any sequence $0 \le t_1 \le \ldots \le t_r < +\infty$ and for any $\varepsilon > 0$ one has:

$$\begin{aligned}
&\mathbb{P}_\mu\left(X^\gamma_{t_1} \in \bigcup_{z\in K_0} B(z,\eta),\dots,X^\gamma_{t_r} \in \bigcup_{z\in K_0} B(z,\varepsilon)\right)\\
&\ge \mathbb{P}_\mu\left(X^\gamma_{t_1} \in \bigcup_{z\in K_0} B(z,\varepsilon),\dots,X^\gamma_{t_{r-1}} \in \bigcup_{z\in K_0} B(z,\varepsilon)\right)\\
&\qquad + \mathbb{P}_\mu\left(X^\gamma_{t_r} \in \bigcup_{z\in K_0} B(z,\varepsilon)\right) - 1 \qquad \text{since } \mathbb{P}(A\cap B) \ge \mathbb{P}(A) + \mathbb{P}(B) - 1\\
&\ge \left(\sum_{k=1}^r \mathbb{P}_\mu\left(X^\gamma_{t_k} \in \bigcup_{z\in K_0} B(z,\varepsilon)\right)\right) - (r-1) \qquad \text{by recurrence}\\
&\underset{\gamma\to+\infty}{\longrightarrow} 1.
\end{aligned}$$

Thus, since X^{γ_p} converges weakly to β, one has that $(\beta_{t_1},\dots,\beta_{t_r}) \in K_0^r$ almost surely, so eventually one has almost surely:

$$f(\beta_{t_1},\dots,\beta_{t_r}) = F(\beta_{t_1},\dots,\beta_{t_r}).$$

Since this is true almost surely, the expectancy of these two quantities are equal, and combining (20) and (21) we find:

$$\lim_{p\in\mathbb{N}} \int_{[0,+\infty[^r} |\mathbb{E}(f(X^{\gamma_p}_{t_1},\dots,X^{\gamma_p}_{t_r})) - \mathbb{E}(F(X^{\gamma_p}_{t_1},\dots,X^{\gamma_p}_{t_r}))| \lambda^{\otimes r}(\mathrm{d}t_1,\dots,\mathrm{d}t_r) = 0 \tag{22}$$

Therefore, in the large p limit, every continuous function f in r variables can be replaces by its r-linearization. □

The r-linearization of f is a sum of terms of the form $\prod_{k=1}^{r} f_{z_{i_k}}$, so we will first study these elementary bricks. One has for $F = \prod_{k=1}^{r} f_{z_{i_k}}$ that:

$$\begin{aligned}
\mathbb{E}_{\overline{\mu}}\left(F(\overline{X}_{t_1},\dots,\overline{X}_{t_r})\right) &= \mathbb{E}_{\overline{\mu}}\left(\prod_{i=1}^{r} f_{z_{i_k}}(\overline{X}_{t_k})\right) \\
&= \mathbb{E}_{\overline{\mu}}\left(\mathbb{E}_{\overline{\mu}}\left(\prod_{i=1}^{r} f_{z_{i_k}}(\overline{X}_{t_k})|\overline{X}_{t_{r-1}} = z_{i_{r-1}}\right)\right) \\
&= \mathbb{E}_{\overline{\mu}}(f_{z_r}(\overline{X}_{t_r})|\overline{X}_{t_{r-1}} = z_{i_{r-1}}) \times \mathbb{E}_{\overline{\mu}}\left(\prod_{i=1}^{r-1} f_{z_{i_k}}(\overline{X}_{t_k})\right) \\
&= \mathbb{E}_{\overline{\mu}}\left(f_{z_{i_1}}(X_{t_1})\right) \times \prod_{k=2}^{r} \mathbb{E}_{\mu}\left(f_{z_k}(\overline{X}_{t_k}|\overline{X}_{t_{k-1}} = z_{i_{k-1}})\right) \\
&= \langle\overline{\mu}, e^{t_1} f_{z_1}\rangle \prod_{k=2}^{r}\langle\delta_{z_{i_{k-1}}}, e^{(t_k-t_{k-1})\overline{\mathcal{L}}} f_{z_{i_k}}\rangle \quad \text{by Theorem 3} \\
&= \sum_{i=0}^{n}\overline{\mu}(z_i)\prod_{k=1}^{r} e^{(t_k-t_{k-1})\overline{\mathcal{L}}} f_{z_k}(z_{i_{k-1}})
\end{aligned}$$

where we implicitely assumed that $i_0 = i$.

To prove that this is the limit of $\mathbb{E}\left(\prod_{k=1}^{r} f_z(X_{t_k}^{\gamma_p})\right)$ for almost any sequence $0 \leq t_1 \leq \ldots \leq t_r$, we will prove by induction that:

$$0 = \lim_{p\to+\infty}\int_{0=t_0\leq t_1\leq\ldots\leq t_r<+\infty}\left|\mathbb{E}\left(\prod_{k=1}^{r} f_{z_{i_k}}(X_{t_k}^{\gamma_p})\right) - \sum_{i=0}^{n}\overline{\mu}(z_i)\prod_{k=1}^{r} e^{(t_k-t_{k-1})\overline{\mathcal{L}}} f_{z_k}(z_{i_{k-1}})\right|\lambda^{\otimes r}(\mathrm{d}t_1,\dots,\mathrm{d}t_r) \tag{23}$$

For $r = 1$, this is trivial since the integrand is bounded by 2 that is integrable on $((\mathbb{R}_+)^r, \lambda^{\otimes r})$ and it converges to 0 pointwisely using Theorem 2, we conclude using the dominated convergence theorem.

We assume that (23) is proven for $r \in \mathbb{N}^*$ given. Let $(\mathcal{F}_t, t \geq 0)$ be the natural filtration. For all $0 \leq t_1 \leq \ldots \leq t_{r+1} < +\infty$, the tower property of conditional expectation an then the Markov property imply:

$$\begin{aligned}
\mathbb{E}\left(\prod_{k=1}^{r+1} f_{z_{i_k}}(X_{t_k}^{\gamma_p})\right) &= \mathbb{E}\left(\prod_{k=1}^{r} f_{z_{i_k}}(X_{t_k}^{\gamma_p}) \times \mathbb{E}(f_{z_{i_{r+1}}}(X_{t_{r+1}}^{\gamma_p})|\mathcal{F}_{t_r})\right) \\
&= \mathbb{E}\left(\prod_{k=1}^{r} f_{z_{i_k}}(X_{t_k}^{\gamma_p}) \times \langle\delta_{X_{t_r}^{\gamma_p}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle\right)
\end{aligned}$$

We have therefore:

$$
\begin{aligned}
&\left|\mathbb{E}\left(\prod_{k=1}^{r+1} f_{z_{i_k}}(X_{t_k}^{\gamma_p})\right) - \mathbb{E}\left(\prod_{k=1}^{r} f_{z_{i_k}}(X_{t_k}^{\gamma_p}) \times \langle \delta_{z_{i_r}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle\right)\right| \\
&= \left|\mathbb{E}\left(\prod_{k=1}^{r} f_{z_{i_k}}(X_{t_k}^{\gamma_p}) \times \langle \delta_{X_{t_r}^{\gamma_p}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle\right) - \mathbb{E}\left(\prod_{k=1}^{r} f_{z_{i_k}}(X_{t_k}^{\gamma_p}) \times \langle \delta_{z_{i_r}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle\right)\right| \\
&= \left|\mathbb{E}\left(\prod_{k=1}^{r} f_{z_{i_k}}(X_{t_k}^{\gamma_p}) \times \left(\langle \delta_{X_{t_r}^{\gamma_p}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle - \langle \delta_{z_{i_r}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle\right)\right)\right| \\
&\leq \mathbb{E}\left(\left| f_{z_{i_r}}(X_{t_r}^{\gamma_p}) \times \left(\langle \delta_{X_{t_r}^{\gamma_p}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle - \langle \delta_{z_{i_r}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle\right)\right|\right)
\end{aligned}
\tag{24}
$$

where the penultimate inequality follows the fact that $|f_{z_{i_k}}(X_{t_k}^{\gamma_p})| \leq 1$. We want to show that (24) converges to 0. In order to do that, we will consider approximations of $\langle \delta_{X_{t_r}^{\gamma_p}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle$ and $\langle \delta_{z_{i_r}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle$ where γ_p only appears in the bras.

The key idea is the one used in the proof of Theorem 2: instead of considering a complex drift b, we approximate our process of generator $\mathcal{L}_\gamma$ with a one of generator $\mathcal{G}_\gamma$ whose drift $\widetilde{b}$ is linear. Since $f_{z_{i_{r+1}}}$ is affine and the convergence in (19) is uniform in μ, one has:

$$\left|\langle \delta_{X_{t_r}^{\gamma_p}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle - \langle \delta_{X_{t_r}^{\gamma_p}}, e^{(t_{r+1}-t_r)\mathcal{G}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle\right| \underset{p\to+\infty}{\longrightarrow} 0,$$

and

$$\left|\langle \delta_{z_{i_r}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle - \langle \delta_{z_{i_r}}, e^{(t_{r+1}-t_r)\mathcal{G}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle\right| \underset{p\to+\infty}{\longrightarrow} 0,$$

where the first convergence is uniform in ω in the sample space.

We can therefore replace $\mathcal{L}_\gamma$ by $\mathcal{G}_\gamma$ in (24), but since $\widetilde{b}$ is linear, Lemma 2 gives us:

$$\langle \delta_{z_{i_r}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle = \langle \delta_{z_{i_r}}, e^{(t_{r+1}-t_r)\widetilde{\mathcal{L}^{(0)}}} f_{z_{i_{r+1}}}\rangle$$

and

$$\langle \delta_{X_{t_r}^{\gamma_p}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle = \langle \delta_{X_{t_r}^{\gamma_p}}, e^{(t_{r+1}-t_r)\widetilde{\mathcal{L}^{(0)}}} f_{z_{i_{r+1}}}\rangle$$

We eventually integrate (24) and use the last two equalities:

$$\begin{aligned}
&\limsup_{p\to+\infty}\int_{0=t_0\le t_1\le\ldots\le t_{r+1}<+\infty}\left|\mathbb{E}\left(\prod_{k=1}^{r+1} f_{z_{i_k}}(X^{\gamma_p}_{t_k})\right)\right.\\
&\qquad\left.-\ \mathbb{E}\left(\prod_{k=1}^{r} f_{z_{i_k}}(X^{\gamma_p}_{t_k})\times\langle\delta_{z_{i_r}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle\right)\right|\lambda^{\otimes r}(\mathrm{d}t_1,\ldots,\mathrm{d}t_{r+1})\\
&\le \limsup_{p\to+\infty}\int_0^{+\infty}\mathbb{E}\left(\left|f_{z_{i_r}}(X^{\gamma_p}_{t_r})\times\left(\langle\delta_{X^{\gamma_p}_{t_r}}, e^{(t_{r+1}-t_r)\widetilde{\mathcal{L}^{(0)}}} f_{z_{i_{r+1}}}\rangle-\langle\delta_{z_{i_r}}, e^{(t_{r+1}-t_r)\widetilde{\mathcal{L}^{(0)}}} f_{z_{i_{r+1}}}\rangle\right)\right|\right)\lambda(\mathrm{d}t_{r+1})\\
&=\int_0^{+\infty}\mathbb{E}\left(\left|f_{z_{i_r}}(\beta_{t_r})\times\left(\langle\delta_{\beta_{t_r}}, e^{(t_{r+1}-t_r)\widetilde{\mathcal{L}^{(0)}}} f_{z_{i_{r+1}}}\rangle-\langle\delta_{z_{i_r}}, e^{(t_{r+1}-t_r)\widetilde{\mathcal{L}^{(0)}}} f_{z_{i_{r+1}}}\rangle\right)\right|\right)\\
&=0,
\end{aligned}$$

the last equality being a consequence of $f_{z_{i_r}}(\beta_{t_r}) = \mathbb{1}_{z_{i_r}}(\beta_{t_r})$ for all t_r, so that the product with the other term vanished necessarily. Now invoking the induction hypothesis with r:

$$\begin{aligned}
0 =&\limsup_{p\to+\infty}\int_{0=t_0\le t_1\le\ldots\le t_{r+1}<+\infty}\left|\mathbb{E}\left(\prod_{k=1}^{r} f_{z_{i_k}}(X^{\gamma_p}_{t_k})\times\langle\delta_{z_{i_r}}, e^{(t_{r+1}-t_r)\mathcal{L}_{\gamma_p}} f_{z_{i_{r+1}}}\rangle\right)\right.\\
&\qquad\qquad\left.\sum_{i=0}^{n}\overline{\mu}(z_i)\prod_{k=1}^{r+1} e^{(t_k-t_{k-1})\overline{\mathcal{L}}} f_{z_k}(z_{i_{k-1}})\right|\lambda^{\otimes r}(\mathrm{d}t_1,\ldots,\mathrm{d}t_{r+1})
\end{aligned}$$

Combining the last limits, we prove the claim for $r+1$.

Using that F is a sum of terms of the form $\prod_{k=1}^{r} f_{z_{i_k}}$, equation (23) implies:

$$\begin{aligned}
0 = &\lim_{p\to+\infty}\int_{0=t_0\le t_1\le\ldots\le t_r<+\infty}\left|\mathbb{E}\left(F(X^{\gamma_p}_{t_1},\ldots,X^{\gamma_p}_{t_r})\right)\right.\\
&\qquad\left.-\sum_{\mathbf{i}\in[\![0,n]\!]^r} F_{\mathbf{i}}\sum_{i=0}^{n}\overline{\mu}(z_i)\prod_{k=1}^{r} e^{(t_k-t_{k-1})\overline{\mathcal{L}}} f_{z_k}(z_{i_{k-1}})\right|\lambda^{\otimes r}(\mathrm{d}t_1,\ldots,\mathrm{d}t_r) \quad (25)
\end{aligned}$$

This last result combined with equations (21) and (22) implies that for $\lambda^{\otimes r}$-almost every (so for Lebesgue-almost every) sequence $0\le t_1\le\ldots\le t_r$ and for any function f continuous, we have:

$$\begin{aligned}
&\lim_{p\to+\infty}\mathbb{E}\left(f(X^{\gamma_p}_{t_1},\ldots,X^{\gamma_p}_{t_r})\right)\\
&=\lim_{p\to+\infty}\mathbb{E}\left(F(X^{\gamma_p}_{t_1},\ldots,X^{\gamma_p}_{t_r})\right)\\
&=\mathbb{E}\left(F(\beta_{t_1},\ldots,\beta_{t_r})\right)\\
&=\lim_{p\to+\infty}\mathbb{E}\left(F(X^{\gamma_p}_{t_1},\ldots,X^{\gamma_p}_{t_r}\right)\\
&=\sum_{\mathbf{i}\in[\![0,n]\!]^r} F_{\mathbf{i}}\sum_{i=0}^{n}\overline{\mu}(z_i)\prod_{k=1}^{r} e^{(t_k-t_{k-1})\overline{\mathcal{L}}} f_{z_k}(z_{i_{k-1}})
\end{aligned}$$

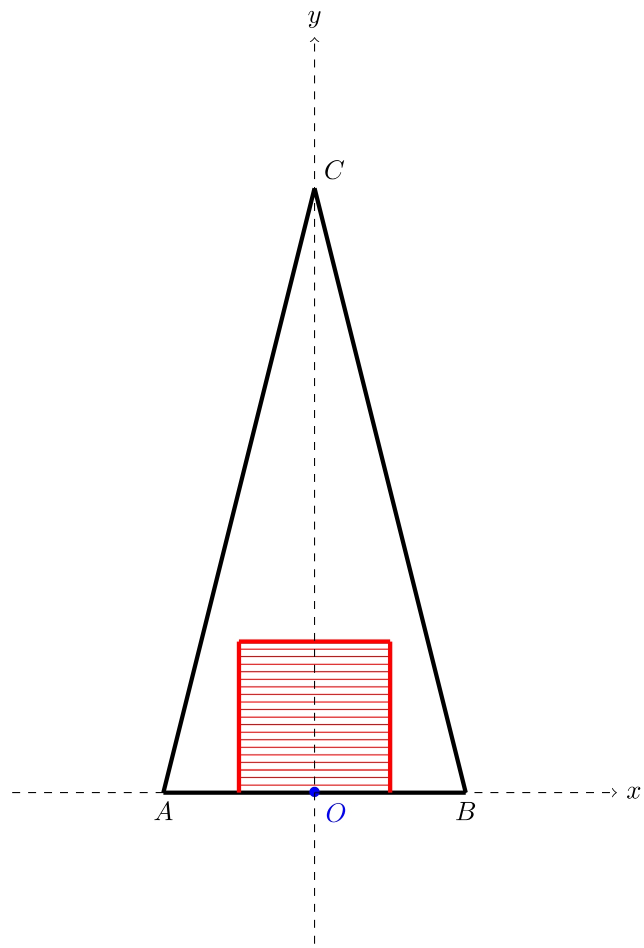
Thus, almost all the finite-dimensional distributions of β and $\overline{X}$ are the same. Finally, [mettre référence] yields that β has the same law as the processus $\overline{X}$ where $\overline{X}$ is the Markov process on K_0 of generator $\overline{\mathcal{L}}$ and of initial condition $\overline{\mu}$. This proves Theorem 3

4 Discussion on the geometrical hypothesis

The proof of our theorem massively use affine approximations, first of the drift then of our test functions, and thus falls apart if we do not assume that K_0 is composed of at most $n+1$ affinely independent points.

We may however ask ourselves if it is only a technical hypothesis, or if it is completely essential, that is to say Theorem 2 is false when it is not fulfilled. In this section, we actually construct a counterexample of our theorem with a compact set K of $\mathbb{R}^2$ and a diffusion process on it where the cardinal of K_0 is equal to $4 > 2+1$.

Let us give a graphic representation of our counterexample :



Here, our compact K is the solid triangle of vertices $A(-1,0)$, $B(1,0)$, and $C(0,4)$. Let us construct a partial definition of our dominant noise σ. For $(x,y) \in [-\frac{1}{2}, \frac{1}{2}]\times]0,1]$, we take :

$$\sigma(x,y) = (x^2(1-x)(1+x)+y^2)\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}$$

If we furthemore take for $x \in [-1,1]$:

$$\sigma(x,0) = (x^2(1-x)(1+x)+y^2)\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}$$

we have defined a smooth function on $[-1,1]\times\{0\}\cup]-\frac{1}{2},\frac{1}{2}]\times]0,1]$ that is null in A, B and O. While it is not very interesting to write it down explicitely, it is really not hard to find a smooth extension of σ that will be null only on A, B, C and O. We have thus $K_0 = \{A, B, C, O\}$.

To understand where the problem is, it may be intersting to consider the functions H_z for $z \in K_0$. The key hypothesis of our ergodic theorem was that the functions H_z were continuous on K. Yet, this is false in our example, and so is our ergodic theorem.

Let us consider H_O: on the one hand, one has obviously that $H_O(O) = 1$. On the other hand, let us consider a stochastic process X_t, of pure noise σ, starting from $(0, y)$ with $y > 0$. Then, for ω an element of the sample space, if $X_t(\omega)$ converges to O when t approaches infinity, for any $\varepsilon > 0$, one has $X_t(\omega) \in B(O, \varepsilon)$ for all t big enough. But in fact, the red lines of our drawing are insurmountable obstacles: since the noise on the y-axis in null on it, if $X_t(\omega)$ is on one of them at a given time t, in order to decrease in ordinate it has to get out of the red box and enter it again. Yet we know that $X_t(\omega) \in B(O, \frac{1}{2})$ for t big enough, and thus $X_t(\omega)$ cannot get out of the $[-\frac{1}{2}, \frac{1}{2}] \times [0, 1]$ box. Therefore, we have

$$X_t(\omega) \in \left[-\frac{1}{2}, \frac{1}{2}\right] \times \{0\}$$

for t big enough, that is to say we reached the bottom line in finite time. But since we have to quit the box to get to the line, there exists T such that

$$X_T(\omega) \in \left(\left[-1, \frac{1}{2}\right] \cup \left[\frac{1}{2}, 1\right]\right) \times \{0\}.$$

Starting from any point of this set, the probability of converging to O is inferior to $\frac{1}{2}$, so using Markov strong property (the process in homogeous in time), we have that the probability of converging to O starting from $(0, y)$ is inferior to $\frac{1}{2}$ for any $y > 0$: our function H_O cannot be continuous.

Remark 7. *If the noise is lipschitz, it is in fact impossible to reach the border in finite time, so we have actually $H_z(x, y) = 0$ for all $y > 0$: the state O is unreachable except if one starts from the bottom line.*

We easily see that Theorem 2 cannot hold in this situation, since our objects are not even well defined: to compute matrix $\overline{\mathcal{L}}$, one has to consider the quantity:

$$b(O) \cdot \nabla H_O(O),$$

that is not well defined if b is non null on the y-axis.

Furthermore, even if our process nevertheless converged to a stochastic process, there are situations where we know that the latter could not be a continuous-time Markov chain of state space K_0. We consider the noise σ defined above (that is only null on A, B, C and O), and a drift b that checks :

$$\begin{cases} b(x, y) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \text{if } (x, y) \in [-\frac{1}{2}, \frac{1}{2}] \times [0, 1] \\ b \in \mathbb{R}_+ \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \\ b = 0 & \text{outside the box } [-\frac{5}{9}, \frac{5}{9}] \times [0, \frac{10}{9}] \\ b \text{ is smooth} & \end{cases}$$

We assume that the stochastic process $(X_u^\gamma)_{u\geq 0}$ starting from O of generator $\mathcal{L}_\gamma$ converges in the weak sense of Theorem 2 to the Markov chain $(\widetilde{X}_u)_{u\geq 0}$ starting from O of state space K_0 and of generator $\overline{\mathcal{L}}$, and we will find a contradiction.

The idea of the proof is the following : since $(\widetilde{X}_u)_{u\geq 0}$ is a Markov process starting from O, we have $\widetilde{X}_t \in O$ with a probability approaching 1 when $t > 0$ is small enough. Let us now consider the process $(X_u^\gamma)_{u\geq 0}$ when γ is big enough. During $[0,t]$ it first enters the box $[-\frac{1}{2}, \frac{1}{2}]\times]0,1]$ pushed by the drift in O, then, since the noise inside is very strong, it quits the red box and, from the exit point, finally approaches A, B or C with a probability greater than a strictly positive quantity. Since this one is independent of γ and t when γ is big enough, it will contradict the fact that it should converges to 0 when γ approaches infinity and t approaches 0.

We consider a smooth function f such that $f(O) = 1$. We assume furthermore that $0 \leq f \leq 1$ and that f is null outside $B(O, \frac{1}{8})$. We know that there exists $t > 0$ small enough such that :

$$\langle \delta_O, e^{t\overline{\mathcal{L}}}\overline{f}\rangle \geq 1 - 2^{-10}.$$

so for γ big enough:

$$\mathbb{P}\left(X_t^\gamma \in B(O, \frac{1}{8})\right) \geq \langle \delta_O, e^{t\mathcal{L}_\gamma} f\rangle \geq 1 - 2^{-9} \tag{26}$$

Let us now consider the process $(X_t^\gamma)_{t\geq 0}$ starting from O of generator $\mathcal{L}_\gamma$. We write T^γ the hitting time of the boundary of our red square :

$$T_\gamma = \inf\left\{t > 0, X_t^\gamma \in \left\{-\frac{1}{2}\right\} \times [0,1] \bigcup \left\{\frac{1}{2}\right\} \times [0,1] \bigcup \left[-\frac{1}{2}, \frac{1}{2}\right] \times \{1\}\right\}.$$

We may now prove that it is arbitrarily small with a high probability when γ approaches infinity. Indeed let g_ε be a smooth function such that $g_\varepsilon(A) = g_\varepsilon(B) = g_\varepsilon(C) = g_\varepsilon(O) = 1$ and we furthermore assumer that $0 \leq g_\epsilon \leq 1$ and that g_ε is null outside of the balls of center $z \in K_0$ and of radius ε. Then for any $s > 0$ and for any ε one has that

$$\mathbb{P}\left(X_s^\gamma \in \bigcup_{z\in K_0} B(z,\varepsilon)\right) \geq \langle \delta_O, e^{s\mathcal{L}_\gamma} g_\varepsilon\rangle \underset{\gamma\to+\infty}{\longrightarrow} \langle \delta_O, e^{s\overline{\mathcal{L}}}\overline{g_\varepsilon}\rangle = 1 \tag{27}$$

Now let us assume that there exists a sequence $(\gamma_k)_k$ approaching infinity such that

$$\mathbb{P}(T_{\gamma_k} > s) \geq \alpha > 0.$$

We notice that for ω in the sample space such that $T_\gamma(\omega) > s$ one has using that $\sigma \cdot e_y = 0$ into the red box that:

$$X_s^\gamma(\omega) \cdot e_y = \int_0^s b(X_u^\gamma(\omega))\mathrm{d}u \cdot e_y = s$$

and since $X_s^\gamma(\omega)$ is on the red box, we have thus:

$$\mathbb{P}\left(X_s^{\gamma_k} \notin \bigcup_{z\in K_0} B(z,\frac{s}{2})\right) \geq \mathbb{P}(T_{\gamma_k} > s) \geq \alpha.$$

for all $k \in \mathbb{N}$, but it contradicts (27) when k approaches infinity.

In conclusion, for all $s > 0$, one has for γ large enough that $\mathbb{P}(T_\gamma \leq s) \geq \frac{1}{2}$, and we notice that for ω such that $T_\gamma(\omega) \leq s$, one has

$$X^\gamma_{T_\gamma(\omega)}(\omega)\cdot e_y = \int_0^{T_\gamma(\omega)} b(X_u^\gamma(\omega))\cdot e_y \mathrm{d}u = T_\gamma(\omega) \leq s. \tag{28}$$

We now consider the two points $R_{-1}(-\frac{1}{2},0)$ and $R_1(\frac{1}{2},0)$. We just proved that choosing γ big enough, we can make sure that $(X_u^\gamma)_{u\geq 0}$ reaches a point on the border of the red box arbitrarily close to R_{-1} or R_1 during $[0,t]$ with a probability at least $\frac{1}{2}$. We will thus make two approximations: we will first approach during a short time our complex diffusion starting from the point $x_\gamma = X^\gamma_{T_\gamma}$ with the dominant one starting from the same point (using Lemma 1), and then approach the dominant process starting from x_γ with the one starting from R_{-1} or R_1.

Since both cases are symmetric, from now on we assume that $x_\gamma \in \{\frac{1}{2}\}\times[0,1]$. We consider $(X_u)_{u\geq 0}(z)$ the process of generator $\mathcal{L}^{(1)}$ starting from $z \in K$. We have with a probability $\frac{1}{2}$ that $X_u(R_1) \underset{u\to+\infty}{\longrightarrow} B$, so there exists $\beta > 0$ such that:

$$\mathbb{P}\left(X_\beta(R_1) \in B(B,\frac{1}{32}) \mid X_\infty = B\right) \geq \frac{3}{4}.$$

and therefore:

$$\begin{aligned}\mathbb{P}\left(X_\beta(R_1) \in B(B,\frac{1}{32})\right) &\geq \mathbb{P}\left(X_\beta(R_1) \in B(B,\frac{1}{32}) \mid X_\infty = B\right)\mathbb{P}(X_\infty(R_1) = B)\\ &\geq \frac{3}{4}\times\frac{1}{2}\\ &\geq \frac{3}{8}\end{aligned} \tag{29}$$

We know use the dependence on initial conditions: there exists $\eta \in]0,\frac{t}{2}[$ such that for all $x \in B(R_1,\eta)$, one has:

$$\sup_{u\in[0,\beta]} \mathbb{E}(|X_u(R_1) - X_u(x)|) \leq \frac{1}{128}.$$

so by Markov's inequality, for all $x \in B(R_1,\eta)$:

$$\mathbb{P}\left(\|X_\beta(R_1) - X_\beta(x)\|_2 \geq \frac{1}{32}\right) \leq \frac{1}{8}. \tag{30}$$

We now use (27): for γ large enough, taking $s = \eta$, we have, with a probability at least $\frac{1}{2}$, that $T_\gamma \leq \eta$ and thus $X^\gamma_{T_\gamma} \in B(R_1, \eta)$ (respectively $B(R_{-1}, \eta)$).

We now write $\gamma = \frac{\beta}{h}$, and we take γ big enough so that, using Lemma 1, we have for all $x \in B(R_1, \eta)$:

$$\mathbb{E}\left(\|X_h^{\frac{\beta}{h}}(x) - X_\beta(x)\|_2 \right) \leq \frac{1}{128}$$

and therefore:

$$\mathbb{P}\left(\|X_h^{\frac{\beta}{h}}(x) - X_\beta(x)\|_2 \geq \frac{1}{16} \right) \leq \frac{1}{8}. \tag{31}$$

Finally we have using (29), (30) and (31):

$$\begin{aligned}
&\mathbb{P}\left(X_h^\gamma(x) \notin B(B, \frac{1}{8}) \right) \\
&\leq \mathbb{P}\left(\{X_\beta(R_1) \notin B(B, \frac{1}{32})\} \bigcup \{|X_\beta(R_1) - X_\beta(x)| \geq \frac{1}{32}|\} \bigcup \{|X_\beta(x) - X_h^{\frac{\gamma}{h}}(x)| \geq \frac{1}{16}|\} \right) \\
&\leq \frac{5}{8} + \frac{1}{8} + \frac{1}{8} \\
&\leq \frac{7}{8}.
\end{aligned}$$

Thus we have:

$$\begin{aligned}
&\mathbb{P}\left(\{X_{T_\gamma + h} \in B(A, \frac{1}{8}) \cup B(B, \frac{1}{8})\} \bigcap \{T_\gamma \leq \frac{t}{2}\} \right) \\
&\geq \mathbb{P}\left(\{X_{T_\gamma + h} \in B(A, \frac{1}{8}) \cup B(B, \frac{1}{8})\} \bigcap \{T_\gamma \leq \eta\} \right) \\
&\geq \mathbb{P}\left(X_{T_\gamma + h} \in B(A, \frac{1}{8}) \cup B(B, \frac{1}{8}) | X_{T_\gamma} \leq \eta \right) \mathbb{P}\left(T_\gamma \leq \eta\right) \\
&\geq \frac{1}{8} \times \frac{1}{2} \\
&\geq \frac{1}{16}
\end{aligned} \tag{32}$$

We thus have with a probability at least $\frac{1}{16}$ that at a time $T_\gamma + h < t$ our process with be in balls of radius $\frac{1}{8}$ of centers A or B, this result being true for all γ big enough. We just have to prove that during $[T_\gamma + h, t]$, our process stays, with a fix strictly positive probability, far from the ball $B(O, \frac{1}{8})$ to find a contradiction with (26). The idea of the proof is to show that if one is close to B, the probability that it stays close to B forever is strictly positive.

We consider S the stopping defined by:

$$S(x) = \inf\{u \geq 0, X_u(x) \notin B(B, \frac{1}{4})\}.$$

We notice that the definition S involves the dominant process and not the general one, and for a good reason: as long as our process lives outside the box $[-\frac{5}{9}, \frac{5}{9}] \times [0, \frac{10}{9}]$, the drift is null and our complex process acts exactly like the dominant one. We have for all $x \in B(B, \frac{1}{8})$ that $(X_{S(x)\wedge u}(x))_{u\geq 0}$ is a martingale, so we have that:

$$x = \mathbb{E}(X_0(x)) = \mathbb{E}(X_{S(x)}(x)),$$

with $S(x)$ being potentially infinite. Since $(X_u(x))_{u\geq 0}$ converges (it is a bounded martingale), if $S(x)(\omega) = +\infty$ then $X_\infty(x)(\omega) = B$. We have therefore, rewriting the last equation, that:

$$x = \mathbb{P}(S(x) = +\infty)B + \mathbb{P}(S(x) < +\infty)\mathbb{E}(X_{S(x)}(x)|S(x) < +\infty). \tag{33}$$

Using basic geometry we have that the point $W(x) = \mathbb{E}(X_{S(x)}(x)|S(x) < +\infty)$, being the barycenter of points from $\partial B(B, \frac{1}{4}) \cap K$, checks

$$\|W(x) - B\|_2 \geq \frac{\sqrt{2}}{8}. \tag{34}$$

Since $x \in B(B, \frac{1}{8})$, one has using (33) and (34) that:

$$\mathbb{P}(S(x) = +\infty) \geq \frac{\sqrt{2} - 1}{\sqrt{2}} > \frac{1}{4} \tag{35}$$

Our dominant process starting from $x \in B(B, \frac{1}{8})$ does not quit $B(B, \frac{1}{4})$ with a probability at least $\frac{1}{4}$. Yet, if it never quits $B(B, \frac{1}{4})$, it will never reach $B(O, \frac{1}{8})$, so using (32) and (35) we get for γ big enough:

$$\begin{aligned}
&\mathbb{P}\left(X_t^\gamma \notin B(O, \frac{1}{8})\right)\\
&\geq \mathbb{P}\left(\{X_t^\gamma \notin B(O, \frac{1}{8})\} \bigcap \{X_{T_\gamma+h}^\gamma \in B(A, \frac{1}{8}) \cup B(B, \frac{1}{8})\} \bigcap \{T_\gamma \leq \frac{t}{2}\}\right)\\
&\geq \mathbb{P}\left(X_t^\gamma \notin B(O, \frac{1}{8}) | X_{T_\gamma+h}^\gamma \in B(A, \frac{1}{8}) \cup B(B, \frac{1}{8}), T_\gamma \leq \frac{t}{2})\right)\\
&\qquad \mathbb{P}\left(\{X_{T_\gamma+h}^\gamma \in B(A, \frac{1}{8}) \cup B(B, \frac{1}{8})\} \bigcap \{T_\gamma \leq \frac{t}{2}\})\right)\\
&\geq \mathbb{P}\left(S(X_{T_\gamma+h}^\gamma) = +\infty | X_{T_\gamma+h}^\gamma \in B(A, \frac{1}{8}) \cup B(B, \frac{1}{8}), T_\gamma \leq \frac{t}{2}\right)\\
&\qquad \mathbb{P}\left(\{X_{T_\gamma+h}^\gamma \in B(A, \frac{1}{8}) \cup B(B, \frac{1}{8})\} \bigcap \{T_\gamma \leq \frac{t}{2}\})\right)\\
&\geq \frac{1}{4} \times \frac{1}{16}\\
&\geq 2^{-6}
\end{aligned} \tag{36}$$

Finally (26) gives us $\mathbb{P}\left(X_t^\gamma \notin B(O,\frac{1}{8})\right) \leq 2^{-9}$ and (36) gives us $\mathbb{P}\left(X_t^\gamma \notin B(O,\frac{1}{8})\right) \geq 2^{-6}$: Contradiction.

We get that one cannot write the limit of our stochastic process of generator $\mathcal{L}_\gamma$, as γ approaches infinity, as a Markov chain of state space K_0.

Remark 8. *In this counterexample, we notice that O belongs to the segment $[A,B]$. More generally, it is easy to build a counterexample of the theorem when the $(n+2)^{th}$ element of K_0 belongs to the convex envelope of the other points: if our extra point is in the interior of K, we just have to squeeze it between two red boxes like the one of our counterexample to find similar contradictions. We have however not yet found a counterexample of our theorem when there is at least $n+2$ points but we assumed that the the points of K_0 are exactly the extremal points of K.*

Remark 9. *One may say that our counterexample is a pathological one since our point O is unreachable if one does not start from a point of the boundary, that is say unreachable from almost any starting point. We may indeed imagine that our limit process could be, from almost any starting point, a Markov chain on the points $\{A,B,C\}$. However, and it is quite surprising, a point that almost surely does not exist for the dominant process, may in fact exists for the complex process. Indeed, if one takes on the compact K of our counterexample:*

$$b(A) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},\ b(B) = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix},\ b(C) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},\ b(O) = 0.$$

and the same σ as above, then we observe in the limit process an increasing accumulation of the mass on O, whatever the starting point is. We cannot thus make as if O did not exist.

Appendix: The proof of the Ergodic Theorem

To prove this we will use several times the following elementary lemma:

Lemma 5. *Let A,B,C be three events.*

One has that:

$$\mathbb{P}(B) \leq \mathbb{P}(A\cap B\cap C) + \mathbb{P}(A^{\complement}\cap B) + \mathbb{P}(C^{\complement}\cap B),$$

that is to say:

$$\mathbb{P}(A\cap B\cap C) \geq \mathbb{P}(B) - \mathbb{P}(A^{\complement}\cap B) - \mathbb{P}(C^{\complement}\cap B),$$

This is completely trivial if one draws the Venn diagram.

Proof. (Theorem 1)
We will first prove the simple convergence of our process.

Since $\mathcal{L}^{(1)}$ is just noise, one has for all x and for all $i \in [\![1, n]\!]$ that

$$X_{t,i}(x) = X_t(x) \cdot e_i$$

is a martingale that lives in the compact space K and is therefore bounded.

The martingale convergence theorem thus assures us that for any x, we have $X_{t,i}(x)$ converges almost surely and in L^2 to $[X_\infty(x)]_i$. Let us prove that $X_{t,i}(x)$ converges necessarily to one of the $z \in K_0$. From now on the x will be implicit.

Using Levy's theorem, there exists an extension $\widetilde{\Omega}$ of our probability space Ω and a Brownian motion on this space β such that:

$$X_{t,i} = \beta_{\langle X_i \rangle_t}$$

where:

$$\begin{aligned}\langle X_i \rangle_t &= \int_0^t \sum_{j=1}^n \sigma_{i,j} \left(X_t\right)^2 \mathrm{d}t \\ &= \sum_{j=1}^n \int_0^t \sigma_{i,j} \left(X_t\right)^2 \mathrm{d}t\end{aligned}$$

Thus, since for almost any $\omega \in \widetilde{\Omega}$ one has that $X_{t,i}(\omega)$ converges when t tends to infinity, we have that $\langle X_i \rangle_t(\omega)$ converges.

But $\langle X_i \rangle_t$ is a sum of n integrals from 0 to t of a positive function, that is to say n functions increasing with t and thus for all $i, j \in [\![1, n]\!]$ one has:

$$t \mapsto \int_0^t \sigma_{i,j} \left(X_t(\omega)\right)^2 \mathrm{d}t$$

converges and thus

$$\sigma_{i,j} \left(X_t(\omega)\right) \underset{t \to +\infty}{\longrightarrow} 0.$$

Eventually one has that:

$$\sigma \left(X_\infty(\omega)\right) = 0,$$

and thus $X_\infty(\omega) \in K_0$.

Therefore we get that for all $x \in K$, one has:

$$\begin{aligned}\mathbb{E}\left(f(X_\infty)\right) &= \sum_{z \in K_0} \mathbb{P}_x(X_\infty = z) f(z) \\ &= \mathcal{P} f(x).\end{aligned}$$

Finally if f is K-lipschitz:

$$\begin{aligned}\left|\mathbb{E}_x\left(f(X_t)\right)-\mathcal{P}f(x)\right| &\leq K\mathbb{E}\left(\|X_t(x)-X_\infty(x)\|_2\right)\\ &\leq K\mathbb{E}\left(\|X_t(x)-X_\infty(x)\|_2^2\right)^{\frac{1}{2}}\\ &\underset{t\to+\infty}{\longrightarrow} 0,\end{aligned}$$

using the theorem of convergence of martinagales in L^2.

We thus proved the pointwise convergence but not the uniform one.

We notice that for any $t>0$, $x,x'\in K$, one has:

$$\begin{aligned}\left|e^{t\mathcal{L}^{(1)}}f(x')-\mathcal{P}f(x')\right| &\leq K\mathbb{E}\left(\|X_t(x)-X_t(x')\|_2\right)+K\mathbb{E}\left(\|X_t(x)-X_\infty(x)\|_2\right)\\ &\quad+K\mathbb{E}\left(\|X_\infty(x)-X_\infty(x')\|_2\right)\\ &\leq K\mathbb{E}\left(\|X_t(x)-X_t(x')\|_2^2\right)^{\frac{1}{2}}+K\mathbb{E}\left(\|X_t(x)-X_\infty(x)\|_2^2\right)^{\frac{1}{2}}\\ &\quad+K\mathbb{E}\left(\|X_\infty(x)-X_\infty(x')\|_2^2\right)^{\frac{1}{2}}\\ &\leq K\mathbb{E}\left(\|X_t(x)-X_\infty(x)\|_2^2\right)^{\frac{1}{2}}+2K\mathbb{E}\left(\|X_\infty(x)-X_\infty(x')\|_2\right)^{\frac{1}{2}}\end{aligned}$$

where we used in the last inequality that $\|X_t(x)-X_\infty(x)\|_2^2$ is a submartingale, and thus its expectation is increasing with t.

From now on we fix an $\varepsilon\in]0,\frac{1}{2}[$.

Using the last inequation, we have that if there is $\eta_x>0$ such that for all $x'\in B(x,\eta_x)$ one has:

$$\mathbb{E}\left(\|X_\infty(x')-X_\infty(x)\|_2^2\right)\leq\varepsilon^2, \tag{37}$$

then for $t>0$ such that

$$\mathbb{E}\left(\|X_t(x)-X_\infty(x)\|_2^2\right)^{\frac{1}{2}}\leq\varepsilon,$$

we will have for all $x'\in B(x,\eta_x)$:

$$\left|e^{t\mathcal{L}^{(1)}}f(x')-\mathcal{P}f(x')\right|\leq 3\varepsilon.$$

Since K is a compact set, it can be covered by a finite amount of balls of the form $B(x,\eta_x)$, yet we have for all x that:

$$\mathbb{E}\left(\|X_t(x)-X_\infty(x)\|_2^2\right)\underset{t\to+\infty}{\longrightarrow}0,$$

so eventually there exists $M>0$ such that for all $t>M$ and for all $x'\in K$ we have:

$$\left|e^{t\mathcal{L}^{(1)}}f(x')-\mathcal{P}f(x')\right|\leq 3\varepsilon.$$

We will now prove (37). The idea of the proof is the following: we want to prove that for η_x small enough, one has for $\omega \in \Omega$ and $x' \in B(x, \eta_x)$ that if $X_\infty(x)(\omega) = z$, then we have $X_\infty(x')(\omega) = z$ with a high probability. However, these are asymptotic events, and we cannot use results such as dependence on the initial condition.

We will therefore invoke an event that is close to $\{X_\infty(x) = z\}$, but that only depend of the process until time $N > 0$ with high probability. This event is "among the balls of center $y \in K_0$ and radius a, the first one that $X_t(x)$ reaches is the ball of center z".

On this event, we may use dependence on initial condition and conclude.

Let us start with some definitions. We write for $a > 0$

$$T^a = \inf \left\{ t \geq 0, X_t(x) \in \bigcup_{y \in K_0} B\left(y, \frac{a}{2}\right) \right\}.$$

We first notice that the definition of this random variable involves explicitely x (and not x'). It is a stopping time that is almost surely finite since X_t converges to one of the y almost surely.

Using the fact that the functions $x \mapsto \mathbb{P}_x(T^a = T^a_y)$ are continuous, we can take $a > 0$ such that:

$$\inf_{y \in K_0} \inf_{w \in B(y,a)} \mathbb{P}_w(T^a = T^a_y) \geq 1 - \varepsilon^2. \tag{38}$$

We may furthermore take a small enough so that the balls of center z and radius a are separated. From now on we fix a such a and we write $T = T_a$.

Since T is almost surely finite there exists $M > 0$ such that:

$$\mathbb{P}(T > M) \leq a^2 \varepsilon^2.$$

We know use the dependence on initial conditions: there exists $\eta_x > 0$ such that for all $x' \in B(x, \eta_x)$, one has:

$$\sup_{t \in [0,M]} \mathbb{E}(|X_t(x) - X_t(x')|) \leq a^2 \varepsilon^2.$$

Eventually we have:

$$\begin{aligned} \mathbb{E}\left(|X_T(x) - X_T(x')|^2\right) &\leq L\mathbb{E}(\mathbb{1}_{T>M}) + \mathbb{E}\left(|X_T(x) - X_T(x')|^2 \mathbb{1}_{T \leq M}\right) \\ &\leq (L+1) a^2 \varepsilon^2, \end{aligned}$$

where $L = \sup_{z \in K} \|z\|_2^2$.

Thus we have using Markov inequality that:

$$\mathbb{P}(|X_T(x) - X_T(x')| \geq \frac{a}{2}) \leq \frac{\mathbb{E}\left(|X_T(x) - X_T(x')|^2\right)}{\left(\frac{a}{2}\right)^2} \leq 4(L+1)\varepsilon^2. \tag{39}$$

We will now consider four events:

$$A = \{X_\infty(x) = z\}, C = \{X_{\infty(x')} = z\},$$

and:

$$B = \{X_T(x) \in B(z, \frac{a}{2})\}, B' = \{X_T(x) \in B(z, a)\}.$$

To conclude we have to prove that:

$$\mathbb{P}(C|A) \geq 1 - \varepsilon^2 \tag{40}$$

since we have thus:

$$\begin{aligned}
\mathbb{E}(\|X_\infty(x) - X_\infty(x')\|_2^2) &\leq \sum_{y \in K_0} \mathbb{E}(\|X_\infty(x) - X_\infty(x')\|_2^2 | X_\infty(x) = y)\mathbb{P}(X_\infty(x) = y) \\
&\leq \sum_{y \in K_0} 4C\mathbb{P}(X_\infty(x') \neq y | X_\infty(x) = y)\mathbb{P}(X_\infty(x) = y) \\
&\leq 4L\varepsilon^2.
\end{aligned}$$

Let us prove (40). To do that we will first study $A \cap B$ and $C \cap B$.

If we write $\mu_t = \delta_x e^{t\mathcal{L}^{(1)}}$ using the homogeneity of our process and the fact that $X_\infty \circ \theta_T = X_\infty$, we have:

$$\begin{aligned}
\mathbb{P}(A \cap B) &= \int_{B(z, \frac{a}{2}} \mathbb{P}_v(X_\infty = z)\mathrm{d}\mu_T(v) \\
&\geq \int_{B(z, \frac{a}{2}} (1 - \epsilon^2)\mathrm{d}\mu_T(v) \qquad \text{using (38).} \\
&\geq (1 - \epsilon^2)\mathbb{P}(B).
\end{aligned} \tag{41}$$

On the one hand we have that if $X_T(x) \in B(z, \frac{a}{2})$, then

$$\{X_T(x') \notin B(z, a)\} \subset \{\|X_T(x) - X_T(x')\| > \frac{a}{2}\}$$

So eventually using (39) we have:

$$\begin{aligned}
\mathbb{P}(B \cap B') &= \mathbb{P}(B'|B)\mathbb{P}(B) \\
&= (1 - \mathbb{P}(X_T(x') \notin B(z, a)))\mathbb{P}(B) \\
&\geq (1 - \varepsilon^2)\mathbb{P}(B)
\end{aligned} \tag{42}$$

so in particular

$$\mathbb{P}(B') \geq (1-\varepsilon^2)\mathbb{P}(B) \tag{43}$$

On the other hand, we know that $X_T(x)$ belongs to one of the balls of center z and radius $\frac{a}{2}$, so if $X_T(x')$ belongs to the ball of center z and radius a we will have:

$$\left\{\|X_T(x)-X_T(x')\|_2 \leq \frac{a}{2}\right\} \subset \left\{X_T(x) \in B(z,\frac{a}{2})\right\}$$

so eventually using (39) we get:

$$\begin{aligned}
\mathbb{P}(B) &\geq \mathbb{P}(B\cap B') \\
&\geq \mathbb{P}(B|B')\mathbb{P}(B') \\
&\geq \mathbb{P}(\|X_T(x)-X_T(x')\|_2 \leq \frac{a}{2})\mathbb{P}(B') \\
&\geq (1-\varepsilon^2)\mathbb{P}(B')
\end{aligned}$$

So since we have chosen $\varepsilon < \frac{1}{2}$ we have:

$$2\mathbb{P}(B) \geq \frac{1}{1-\varepsilon^2}\mathbb{P}(B) \geq \mathbb{P}(B') \tag{44}$$

Finally we have using the same reasoning as in the proof of (41) that:

$$\mathbb{P}(B'\cap C) \geq (1-\varepsilon^2)\mathbb{P}(B'). \tag{45}$$

So eventually we have using Lemma 5, that:

$$\begin{aligned}
\mathbb{P}(B\cap C) &\geq \mathbb{P}(B\cap B'\cap C) \\
&\geq \mathbb{P}(B') - (\mathbb{P}(B') - \mathbb{P}(B'\cap B)) - (\mathbb{P}(B') - \mathbb{P}(B'\cap C)) \\
&\geq \mathbb{P}(B') - \varepsilon^2\mathbb{P}(B') - \varepsilon^2\mathbb{P}(B') \qquad \text{Using (42) and (45)} \\
&\geq (1-\varepsilon^2)\mathbb{P}(B) - 2\varepsilon^2 \times 2\mathbb{P}(B) \qquad \text{Using (43) and (44)} \\
&\geq (1-5\varepsilon^2)\mathbb{P}(B). \qquad (46)
\end{aligned}$$

Finally we have using Lemma 5 again that:

$$\begin{aligned}
\mathbb{P}(A\cap C) &\geq \mathbb{P}(A\cap B\cap C) \\
&\geq \mathbb{P}(B) - (\mathbb{P}(B) - \mathbb{P}(A\cap B)) - (\mathbb{P}(B) - \mathbb{P}(B\cap C)) \\
&\geq \mathbb{P}(B) - \varepsilon^2\mathbb{P}(B) - 5\varepsilon^5\mathbb{P}(B) \qquad \text{Using (41) and (46)} \\
&\geq (1-6\varepsilon^2)\mathbb{P}(B). \\
&\geq (1-6\varepsilon^2)\mathbb{P}(A)\mathbb{P}(B|A). \qquad (47)
\end{aligned}$$

We have almost proved (40) and thus the theorem. We just need to show that B is a very likely event knowing A.

We consider the random variable

$$S = \inf_{t\geq 0} \inf_{\substack{y\neq z \\ y\in K_0}} \|X_t(x)-y\|_2$$

If $\omega \in A$, then $X_t(x)(\omega) \underset{t\to+\infty}{\longrightarrow} z$, so $X_t(x)(\omega) \in B(z,a)$ for t large enough and using the continuity of $t \mapsto X_t(x)(\omega)$, we get that $S(\omega) > 0$.

Yet the sequence of events $\{S < q\}_{q>0}$ is decreasing with q and

$$\bigcap_{q>0} \{S < q\} = \{S = 0\}.$$

But since S is strictly positive for $\omega \in A$ one has:

$$\mathbb{P}(S = 0 | A) = 0$$

so the theorem of monotone convergence theorem gives us:

$$\mathbb{P}(M < q | A) \underset{q\to 0}{\longrightarrow} 0.$$

Therefore, there exists $q > 0$ such that $\mathbb{P}(M < q|A) \leq \epsilon^2$ so if we take $a < q$ in the definition of $T = T_a$, we get:

$$\mathbb{P}\left(B^{\complement}\right) = \mathbb{P}\left(X_T(x) \in \bigcap_{\substack{y\neq z \\ y\in K_0}} B(y,a) \mid A\right) \leq \mathbb{P}(M < a) \leq \varepsilon^2.$$

Therefore, we have finally that:

$$\mathbb{P}(B|A) \geq (1 - \varepsilon^2).$$

so we can rewrite (47):

$$\mathbb{P}(A \cap C) \geq (1 - 6\varepsilon^2)(1 - \varepsilon^2)\mathbb{P}(A)$$

and eventually:

$$\mathbb{P}(C|A) \geq (1 - 7\varepsilon^2).$$

This proves the theorem. □