THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PARIS 13

Discipline : **Mathématiques**

présentée et soutenue publiquement par

Kaouther HAJJI

le 12 Décembre 2014

Accélération de la méthode de Monte Carlo pour des processus de diffusions et applications en Finance

Directeur de thèse : M. Mohamed BEN ALAYA

Co-Directeur de thèse : M. Ahmed KEBAIER

Jury :	M. Gilles Pagès	Université Paris VI
	M. Jean Stéphane Dhersein	Université Paris 13
	M. Denis Talay	INRIA Sophia Antiplois
	Mme. Gersende Fort	Telecom ParisTech
	M. Yueyun Hu	Université Paris 13
	M. Bernard Lapeyre	Ecole des Ponts ParisTech
	M. Mohamed Ben Alaya	Université Paris 13
	M. Ahmed Kebaier	Université Paris 13
Rapporteurs :	M. Denis Talay	INRIA Sophia Antipolis
	M. Bernard Lapeyre	Ecole des Ponts ParisTech

Remerciements

Je tiens tout d'abord à exprimer toute ma gratitude envers Mr. Mohamed Ben Alaya et Mr Ahmed Kebaier qui ont dirigé ma thèse avec une attention constante. Chacun d'eux mérite un remerciement particulier. Je remercie Mohamed Ben Alaya pour avoir accepté de diriger cette thèse avec beaucoup de disponibilité et un soutien continu. Il a fait preuve d'une grande patience et gentillesse et m'a beaucoup encouragé tout au long de cette thèse. Je n'oublierai jamais ses conseils judicieux qu'il m'a prodigués. Je remercie aussi Ahmed Kebaier d'avoir accepté d'encadrer mon stage de Master au sein de l'équipe MATHRISK à Inria Rocquencourt et de m'avoir accordé sa confiance en acceptant de diriger cette thèse et de me donner une seconde fois l'opportunité de travailler avec lui. Sa compétence et son dynamisme m'ont toujours poussé vers l'avant. Je lui suis infiniment reconnaissante pour son soutien, son encouragement et ses précieux conseils dont j'ai tiré un grand profit.

Mes remerciements les plus respectueux vont également à Mr. Denis Talay et Mr. Bernard Lapeyre qui m'ont fait l'honneur de bien vouloir rapporter ma thèse en y prêtant toute l'attention nécessaire. Qu'ils trouvent ici l'expression de ma profonde reconnaissance. Je tiens aussi à remercier Mme. Gersende Fort, Pr. Gilles Pagès, Pr. Jean-Stéphane Dhersein et Pr. Yueyun Hu qui m'ont fait l'honneur de participer à mon jury de thèse.

Je tiens à remercier Mme Eulàlia Nualart de m'avoir accueillie au laboratoire LAGA à l'université Paris 13 et de m'avoir soutenue constamment durant les premières années de cette thèse ainsi que pour sa générosité et sa gentillesse.
Je tiens à remercier l'équipe MATHRISK, notamment Antonino Zanette et Agnès Sulem, de m'avoir permis d'implémenter, dans le logiciel Premia, les méthodes numériques développées dans cette thèse.

Un grand merci à tous les membres du LAGA. Je garde toujours le meilleur souvenir de tous les membres du laboratoire que j'ai côtoyés pendant ces années de thèse. Je pense ainsi en particulier à Mohamed Mrad pour le soutien moral et les conseils qu'il m'a prodigués tout au long de ces années de thèse et aussi à Maher Zerzeri pour son ecouragement constant, ses conseils précieux et sa générosité. Je pense également à mes amis doctorants et docteurs, Amine (mon cher ami avec qui j'ai partagé les meilleurs moments et même les plus difficiles pendant ces années de thèse, je ne cesse de te remercier pour ta confiance, ta bonne humeur, enfin pour tout ce que tu as pu m'apporter durant ces années), Amal (merci pour ta générosité et pour tous les moments agréables qu'on a pu passer ensemble), Abderrahman, Farah, Saida (merci pour la bonne ambiance et les balades à paris), Alexandre, Tran, Bakari (merci pour les moments très sympathiques qu'on a partagé dans notre bureau commun au LAGA). Merci à nos chéres secrétaires Yolande et Isabelle pour leur gentillesse, leur bonne humeur et les nombreux services qu'ils ont rendus toujours avec le sourire. Un grand merci et bravo à l'équipe universitaire de Basketball de Paris 13 où j'ai partagé 3 années de passion pour le grand ballon rond. Merci coach Ludovic et toutes les filles pour l'ambiance familiale et chaleureuse. J'ai passé avec vous de très agréables moments pendant les entrainements, les matchs et surtout le championnat de France universitaire à Montluçon. J'en garde le meilleur souvenir (et oui deux fois championnes d'Ile de France, ce n'est pas rien pour moi!!).

C'est avec plaisir que je remercie maintenant mes amis qui ont joué un grand rôle pour l'aboutissement de cette thèse. Je remercie en particulier ma chère amie Rawya pour son soutien surtout dans la période de grossesse. Tu as été et tu resteras toujours l'une des meilleures amies que j'ai toujours connue. Merci à mes amis, anciens Polytechniciens de Tunisie : Malek, Nagui, Arbi, Amine, Rim. Merci pour les moments qu'on a passé ensemble, la bonne ambiance, les encouragements et le sourire.

A ma famille dont l'amour indéfectible m'a porté et soutenu jour après jour et sans laquelle je ne serai probablement pas arrivé à ce stade aujourd'hui. C'est à mon tour de vous exprimer ma profonde reconnaissance. Tout d'abord à mon père Habib et ma mère Néjiba, je saisis cette heureuse circonstance pour vous exprimer, d'un élan de fierté et de tendresse, mes sentiments de gratitude et d'affection pour votre soutien moral quotidien même si vous êtes de l'autre côté de la Méditérannée. Les mots que je pourrais écrire ne sont pas à la hauteur de ce que je peux ressentir à votre égard. Ce travail vous est dédié.
À mes belles et adorables soeurs pour votre soutien qui a été determinant pour l'aboutissement de ce travail. Merci pour tout ce que vous avez fait pour moi.

Le mot de la fin c'est un affectueux remerciement à mon mari et mon confident Meher pour son soutien constant et inconditionnel et sa présence durant toutes ces années d'étude et surtout dans les moments difficiles... sans toi à mes côtés rien ne serait possible...
Enfin, j'ai une profonde pensée à mon bébé ZAKARYA qui a vu le jour au cours de ces années de thèse et a fait le bonheur de la famille. Je ne serai capable de lui exprimer l'amour que j'éprouverai toujours pour lui et la joie de l'avoir à mes cotés. Je ne cesserai de remercier Dieu pour ce don aussi merveilleux que lui.

À mes anges Gardiens Habib et Néjiba
Et à mes amours Meher et Zakarya...

Résumé

Dans cette thèse, on s'intéresse à la combinaison des méthodes de réduction de variance et de réduction de la complexité de la méthode Monte Carlo. Dans une première partie de cette thèse, nous considérons un modèle de diffusion continu pour lequel on construit un algorithme adaptatif en appliquant l'importance sampling à la méthode de Romberg Statistique Nous démontrons un théorème central limite de type Lindeberg Feller pour cet algorithme. Dans ce même cadre et dans le même esprit, on applique l'importance sampling à la méthode de Multilevel Monte Carlo et on démontre également un théorème central limite pour l'algorithme adaptatif obtenu. Dans la deuxième partie de cette thèse, on développe le même type d'algorithme pour un modèle non continu à savoir les processus de Lévy. De même, nous démontrons un théorème central limite de type Lindeberg Feller. Des illustrations numériques ont été menées pour les différents algorithmes obtenus dans les deux cadres avec sauts et sans sauts.

Mots clés. Algorithmes stochastiques, Robbins-Monro, Théorème limite central, Romberg Statistique, Multilevel Monte Carlo, schéma d'Euler, importance sampling, options exotiques, modèle de Heston, processus de Lévy et approximation, transformation d'Esscher, modèle du CGMY.

Abstract

In this thesis, we are interested in studying the combination of variance reduction methods and complexity improvement of the Monte Carlo method. In the first part of this thesis, we consider a continuous diffusion model for which we construct an adaptive algorithm by applying importance sampling to Statistical Romberg method. Then, we prove a central limit theorem of Lindeberg-Feller type for this algorithm. In the same setting and in the same spirit, we apply the importance sampling to the Multilevel Monte Carlo method. We also prove a central limit theorem for the obtained adaptive algorithm. In the second part of this thesis, we develop the same type of adaptive algorithm for a discontinuous model namely the Lévy processes and we prove the associated central limit theorem. Numerical simulations are processed for the different obtained algorithms in both settings with and without jumps.

Keywords. Stochastic algorithm, Robbins-Monro, Central limit theorem, Statistical Romberg method, Multilevel Monte Carlo, Euler scheme, importance sampling, exotic options, Heston model, Lévy processes and approximation, Esscher transform, CGMY model.

Table des matières

Introduction

Le but de cette thèse est de proposer des techniques d'accélération des méthodes de Monte Carlo utilisées en Finance. Notre approche consiste à combiner deux méthodes de réduction de variance.

- Les méthodes d'importance sampling qui utilisent en particulier les algorithmes stochastiques introduites initialement par Arouna [3] dans sa thèse.
- Les méthodes utilisées pour gagner en complexité en comparaison avec la méthode de Monte Carlo standard à l'exemple de la méthode de Romberg Statistique étudiée dans la thèse de Kebaier [37] et sa généralisation à savoir la méthode de Multilevel Monte Carlo introduite par Giles [28].

Dans ma thèse, plusieurs algorithmes ont été développés et de nouveaux résultats de convergence ont été démontré afin de définir le cadre précis de l'utilisation de ces nouvelles méthodes de réduction de variance. En effet, on s'intéresse au calcul de la quantité $\mathbb{E}f(X_T)$, où $(X_t)_{0\leq t\leq T}, T > 0$ est une diffusion d-dimensionnelle et $f : \mathbb{R}^d \to \mathbb{R}$ est une fonction donnée. Cette quantité représente en finance le prix d'une option sur le marché et f étant la fonction payoff de cette option. À l'aide d'un changement de probabilité en utilisant le théorème de Girsanov, on peut écrire notre quantité d'intérêt sous la forme :

$$\forall\theta \in \mathbb{R}^q, \quad \mathbb{E}f(X_T) = \mathbb{E}g(\theta, X_T^\theta),$$

avec $g : \mathbb{R}^q \times \mathbb{R}^d \to \mathbb{R}$ et $(X_t^\theta)_{0\leq t\leq T}$ est une diffusion dépendante de θ. Dans le but de réduire l'erreur statistique de la méthode de Monte Carlo, il est naturel de chercher à minimiser la variance de l'estimateur $\frac{1}{N}\sum_{i=1}^N g(\theta, X_{T,i}^\theta)$, $(X_{T,i}^\theta)_{1\leq i\leq N}$ est une suite i.i.d de même loi que X_T^θ, ce qui revient à chercher le paramètre optimal qu'on l'a considéré réduit à un singleton dans les cas étudiés

$$\arg\min_{\theta\in\mathbb{R}^q} \mathrm{Var}(g(\theta, X_T^\theta)) = \arg\min_{\theta\in\mathbb{R}^q} \mathbb{E}g^2(\theta, X_T^\theta).$$

En utilisant une autre fois le théorème de Girsanov et en dérivant, ce problème de recherche du minimum d'une fonction peut se ramener à celui de la recherche du zéro d'une fonction. En effet, sous certaines hypothèses, $\arg\min_{\theta\in\mathbb{R}^q} \mathrm{Var}(g(\theta, X_T^\theta))$ peut être solution de

$$\mathbb{E}[H(\theta, X_T)] = 0, \quad \theta \in \mathbb{R}^q, \tag{1}$$

où $H : \mathbb{R}^q \times \mathbb{R}^d \to \mathbb{R}$ est une fonction explicitée à partir des données du problème. Ainsi, le recours aux algorithmes stochastiques de type Robbins Monro semble être adapté pour ce problème d'optimisation. Toutefois, cet algorithme requiert une condition restrictive appelée la condition de non explosion donnée par :

$$\text{(NEC)} \qquad \mathbb{E}\left[|H(\theta, X_T)|^2\right] \leq C(1 + |\theta|^2), \quad \text{pour tout } \theta \in \mathbb{R}^q.$$

Cette condition n'étant pas satisfaite dans notre cadre de travail, nous adoptons ainsi de nouvelles versions de cette procédure récursive permettant de contourner la condition (NEC).

- La méthode de projection aléatoire de l'algorithme de Robbins Monro introduite par Chen et Zhu [18], connue sous le nom de "Projection à la Chen".
- La méthode proposée plus récemment par Lemaire et Pagès dans [47] comme une nouvelle alternative à l'algorithme de Robbins-Monro et qui évite le risque d'explosion sans toutefois avoir recours aux projections. Par un triple changement de probabilité, ils proposent une nouvelle fonction $\tilde{H}$ qui remplace H dans l'équation (1) et vérifiant la condition (NEC). Ils obtiennent ainsi ce qu'ils appellent "a regular Robbins Monro algorithm" qui converge presque sûrement vers $\arg\min_{\theta\in\mathbb{R}^q} \mathrm{Var}(g(\theta, X_T^\theta))$ sans risque d'explosion.

► **Dans une première partie de cette thèse**, on a considéré un modèle continu de l'observation des prix des options dans le marché financier. Ainsi, $(X_t)_{0\leq t\leq T}, T > 0$, est défini comme la solution d'une équation différentielle stochastique dirigée par un mouvement Brownien q-dimensionnel. En dehors du modèle de Black-Scholes, $(X_t)_{0\leq t\leq T}$ n'est pas explicite donc on est souvent ramené à utiliser un schéma d'approximation à savoir le schéma d'Euler $(X_t^n)_{0\leq t\leq T}$ (où n est le nombre de pas de temps de discrétisation). Pour n fixé, Arouna [3] a utilisé l'algorithme de Chen pour déterminer le paramètre optimal qui minimise la variance de $g(\theta, X_T^{n,\theta})$ en construisant une suite θ_i^n qui converge p.s. vers $\arg\min_{\theta\in\mathbb{R}^q} \mathrm{Var}(g(\theta, X_T^{n,\theta}))$ lorsque i tend vers ∞. Plus précisément, il utilise la méthode de Monte Carlo et approxime $\mathbb{E}[f(X_T^n)]$ par

$\frac{1}{N}\sum_{i=1}^{N} g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n})$, $(X_{T,i}^{n,\theta})_{1\leq i\leq N}$ sont des variables aléatoires i.i.d de même loi que $X_T^{n,\theta}$. Ainsi, un algorithme adaptatif a été développé dans lequel la variance est réduite de manière dynamique dans les itérations de Monte Carlo. Le comportement asymptotique de l'algorithme adaptatif proposé par Arouna [3] est donné par

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}) - \mathbb{E}[f(X_T^n)]\right) \xrightarrow[N\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \mathrm{Var}(g(\theta_n^*, X_T^{n,\theta_n^*}))\right), \tag{2}$$

avec $\theta_n^* = \arg\min_{\theta\in\mathbb{R}^q} \mathrm{Var}(g(\theta, X_T^{n,\theta}))$. Un premier résultat qui généralise celui d'Arouna [3] consiste à choisir N en fonction de n en tenant compte de l'erreur faible du schéma d'Euler. On a supposé que

$$\forall\alpha\in[1/2,1],\quad \lim_{n\to\infty} n^{\alpha}\left(\mathbb{E}f(X_T^n) - \mathbb{E}f(X_T)\right) = C_f,\quad C_f\in\mathbb{R}.$$

Puis, nous avons construit une suite doublement indexée $(\theta_i^n)_{i,n\in\mathbb{N}}$ à l'aide des algorithmes stochastiques en utilisant soit la méthode de Projection à la Chen [18] ou la méthode proposée par Lemaire et Pagès [47]. Pour ces deux alternatives, nous avons établi la convergence presque sûre de la suite $(\theta_i^n)_{i,n\in\mathbb{N}}$ (voir Section 3 du chapitre II) et nous avons obtenu

$$\lim_{i,n\to\infty}\theta_i^n = \lim_{i\to\infty}(\lim_{n\to\infty}\theta_i^n) = \lim_{n\to\infty}(\lim_{i\to\infty}\theta_i^n) = \lim_{n\to\infty}\theta_n^* = \arg\min_{\theta\in\mathbb{R}^q}\mathrm{Var}(g(\theta, X_T^{\theta})),\quad \mathbb{P}\text{-}p.s.$$

Par la suite, nous avons démontré un théorème central limite pour la procédure adaptative proposée (voir Théorème 4.4 dans le chapitre II). Nous avons ainsi

$$n^{\alpha}\left(\frac{1}{n^{2\alpha}}\sum_{i=1}^{n^{2\alpha}} g(\theta_{i-1}^n, X_{T,i}^{n,\theta_{i-1}^n}) - \mathbb{E}f(X_T)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(C_f, \mathrm{Var}(g(\theta^*, X_T^{\theta^*}))\right),$$

avec $\theta^* = \arg\min_{\theta\in\mathbb{R}^q}\mathrm{Var}(g(\theta, X_T^{\theta}))$.
Par ailleurs, la méthode de Romberg Statistique, introduite par Kebaier [37] fait intervenir deux schémas d'Euler de nombres de pas de discrétisations différents, un schéma "fin" de pas de discrétisation T/n et un schéma "grossier" de pas de discrétisation $T/\sqrt{n}$. Ainsi, en appliquant les résultats de Kebaier [37] pour $\theta\in\mathbb{R}^q$, $\mathbb{E}f(X_T)$ est approximée par

$$V_n = \frac{1}{N_1}\sum_{i=1}^{N_1} g(\theta, \hat{X}_{T,i}^{\sqrt{n},\theta}) + \frac{1}{N_2}\sum_{i=1}^{N_2}\left(g(\theta, X_{T,i}^{n,\theta}) - g(\theta, X_{T,i}^{\sqrt{n},\theta})\right),$$

où $(X_{T,i}^{\sqrt{n},\theta})_{i\in\mathbb{N}}$ et $(X_{T,i}^{n,\theta})_{i\in\mathbb{N}}$ sont des variables aléatoires i.i.d de même loi que $X_T^{\sqrt{n},\theta}$ et $X_T^{n,\theta}$ simulées à partir de la même trajectoire Brownienne. $(\hat{X}_{T,i}^{\sqrt{n},\theta})_{i\in\mathbb{N}}$ est un autre échantillon indépendant. Cette méthode est contrôlée par un théorème central limite obtenu pour un choix précis des tailles des échantillons en prenant $N_1 = n^{2\alpha}$ et $N_2 = n^{2\alpha-1/2}$ et est donné par

$$n^{\alpha}\left(V_n - \mathbb{E}f(X_T)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(C_f, \Sigma^2(\theta)\right), \tag{3}$$

où $\Sigma^2(\theta)$ est une variance qui est précisée dans le Théorème 1.6 du chapitre I. Dans le cadre de cette thèse, notre idée est de combiner la technique d'importance sampling avec la méthode de Romberg Statistique. On propose alors d'approcher $\mathbb{E}f(X_T)$ par

$$\tilde{V}_n = \frac{1}{n^{2\alpha}}\sum_{i=1}^{n^{2\alpha}} g(\hat{\theta}_{i-1}^{\sqrt{n}}, \hat{X}_{T,i}^{\sqrt{n},\hat{\theta}_{i-1}^{\sqrt{n}}}) + \frac{1}{n^{2\alpha-1/2}}\sum_{i=1}^{n^{2\alpha-1/2}}\left(g(\theta_{i-1}^{n}, X_{T,i}^{n,\theta_{i-1}^{n}}) - g(\theta_{i-1}^{n}, X_{T,i}^{\sqrt{n},\theta_{i-1}^{n}})\right), \tag{4}$$

avec $(\theta_i^n)_{i,n\in\mathbb{N}}$ est une nouvelle suite doublement indexée associée à cet algorithme qui vérifie

$$\lim_{i\to\infty}(\lim_{n\to\infty}\theta_i^n) = \lim_{n\to\infty}(\lim_{i\to\infty}\theta_i^n) = \arg\min_{\theta\in\mathbb{R}^q}\Sigma^2(\theta), \quad \mathbb{P}\text{-}p.s.$$

La double convergence de $(\theta_i^n)_{i,n\in\mathbb{N}}$ pour la technique de Lemaire et Pagès [47] a nécessité une analyse stochastique portant sur la sensibilité des processus $(X_t^\theta)_{0\le t\le T}$ par rapport au paramètre θ. Par la suite, on a prouvé un théorème central limite pour cette nouvelle procédure adaptative (voir Théorème 4.6 dans chapitre II) :

$$n^{\alpha}\left(\tilde{V}_n - \mathbb{E}f(X_T)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(C_f, \Sigma^2(\theta^*)\right),$$

avec $\theta^* = \arg\min_{\theta\in\mathbb{R}^q}\Sigma^2(\theta)$. Rappelons que l'intérêt de la méthode de Romberg Statistique est de réduire la complexité de celle de Monte Carlo classique. En effet, la complexité optimale de la méthode de Romberg Statistique est donnée par $C_{SR} = n^{2\alpha+1/2} << C_{MC} = n^{2\alpha+1}$ (voir la sous-section 1.3 du Chapitre I).

Plus récemment, Giles [28] a introduit la méthode de Multilevel Monte Carlo, une version généralisée de l'approche de Romberg Statistique. Cette méthode permet d'approcher $\mathbb{E}f(X_T)$ en utilisant le schéma d'Euler et qu'on appelle Euler Multilevel Monte Carlo. Cette méthode fait l'objet de plusieurs travaux dans la littérature avec des applications notamment en Finance. Pour plus détails, on peut consulter https://people.maths.ox.ac.uk/gilesm/mlmc_community.html. En particulier, nous citons quelques travaux récents qui portent sur la réduction de la complexité de l'estimateur de Multilevel Monte Carlo à l'exemple du travail de

Giles et Szpruch [29] qui ont construit un nouvel estimateur Multilevel Monte Carlo en utilisant la technique des variables antithétiques. De même, Lemaire et Pagès [48] ont étudié la combinaison de la méthode de Multilevel Monte Carlo avec la technique d'extrapolation de Richardson-Romberg et son extension Multi-Step [53].
La méthode Euler Multilevel Monte Carlo consiste à considérer $L+1$ échantillons i.i.d de tailles respectives $(N_0, ..., N_L)$. Pour chaque échantillon, nous considérons pour $\ell \in \{0, 1, ..., L\}$ un schéma d'Euler "fin" avec un pas de temps T/m^ℓ et un schéma d'Euler "grossier" avec un pas de temps $T/m^{\ell-1}$ simulées à partir de la même trajectoire Brownienne. Ensuite, la quantité d'intérêt $\mathbb{E}f(X_T)$ est approchée par

$$Q_n = \frac{1}{N_0}\sum_{i=1}^{N_0} g(\theta, X_{T,i}^{1,\theta}) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\left(g(\theta, X_{T,i}^{\ell,m^\ell,\theta}) - g(\theta, X_{T,i}^{\ell,m^{\ell-1},\theta})\right), \quad \theta \in \mathbb{R}^q. \tag{5}$$

Il est à noter que ces moyennes empiriques sont indépendantes entre elles. Récemment, Ben Alaya et Kebaier [11] ont démontré un théorème central limite, pour un choix convenable des tailles des échantillons N_ℓ, $\ell \in \{0, 1, ..., L\}$ dépendantes de $n = m^L$, pour la méthode Euler Mutlilevel Monte Carlo donné par

$$\forall \theta \in \mathbb{R}^q, n^\alpha \left(Q_n - \mathbb{E}f(X_T)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(C_f, \tilde{\Sigma}^2(\theta)\right),$$

où $\tilde{\Sigma}^2(\theta)$ est une variance précisée dans le Théorème 1.9 du chapitre I. Dans le cadre de cette thèse, nous avons étendu dans un deuxième travail les résultats obtenus ci-dessus en procédant à la combinaison de la méthode Euler Mutlilevel Monte Carlo et celle de l'importance sampling. Ainsi, nous introduisons l'algorithme adaptatif qui approche la quantité d'intérêt $\mathbb{E}f(X_T)$ par

$$\tilde{Q}_n = \frac{1}{N_0}\sum_{i=1}^{N_0} g(\theta_{i-1}^{m^0}, \hat{X}_{T,i}^{m^0,\theta_{i-1}^{m^0}}) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\left(g(\theta_{i-1}^{\ell,m^\ell}, X_{T,i}^{\ell,m^\ell,\theta_{i-1}^{\ell,m^\ell}}) - g(\theta_{i-1}^{\ell,m^\ell}, X_{T,i}^{\ell,m^{\ell-1},\theta_{i-1}^{\ell,m^\ell}})\right),$$

avec $(\theta_i^n)_{i,n\in\mathbb{N}}$ est une nouvelle suite doublement indexée associée à cet algorithme qui vérifie

$$\lim_{i\to\infty}(\lim_{n\to\infty}\theta_i^n) = \lim_{n\to\infty}(\lim_{i\to\infty}\theta_i^n) = \arg\min_{\theta\in\mathbb{R}^q}\tilde{\Sigma}^2(\theta), \quad \mathbb{P}\text{-}p.s.$$

Par la suite, nous avons prouvé un théorème central limite (voir Théorème 3.2 dans le chapitre III) pour cette nouvelle procédure adaptative à savoir

$$n^\alpha\left(\tilde{Q}_n - \mathbb{E}f(X_T)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(C_f, \tilde{\Sigma}^2(\theta^*)\right),$$

avec $\theta^* = \arg\min_{\theta\in\mathbb{R}^q} \tilde{\Sigma}^2(\theta)$. De même que pour la méthode de Romberg Statistique, l'intérêt de la méthode de Multilevel Monte Carlo est de réduire la complexité de la méthode de Monte Carlo classique. En effet, la complexité optimale associée à la méthode de Multilevel Monte Carlo est donnée par $C_{MLMC} = n^{2\alpha}(\log(n))^2 << C_{RS} = n^{2\alpha+1/2} << C_{MC} = n^{2\alpha+1}$ (voir sous-section 1.4 du chapitre I).

▶ **Dans une deuxième partie de cette thèse**, nous avons considéré des modèles avec sauts. En effet, l'intérêt de ces modèles est qu'empiriquement la trajectoire du sous-jacent n'est pas continue et peut exhiber des sauts à certaines périodes de forte volatilité. Les modèles financiers discontinus, introduits par Merton (1976) [51] et étudiés dans Cont et Tankov [20], Schoutens [61], Applebaum [2] ou encore Asmussen et Glynn [4], permettent de mieux expliquer ce phénomène. D'un point de vue numérique et afin d'améliorer la méthode de Monte Carlo avec des processus de Lévy, Kawai [36] et Lemaire et Pagès [47] ont utilisé les méthodes d'importance sampling de même type que celle d'Arouna [3] pour réduire la variance.

Dans cette partie de la thèse, nous avons considéré le modèle exponentiel de Lévy, où le prix à un instant t d'un actif associé à une valeur initiale "spot" S_0 est donné par

$$S_t = S_0 e^{L_t}, \qquad \text{où } (L_t)_{0\leq t\leq T} \text{ est un processus de Lévy } d\text{-dimensionnel.}$$

Ici, $e^x = (e^{x_1}, ..., e^{x_n}), \forall x = (x_1, ..., x_n) \in \mathbb{R}^d$. Ainsi, notre quantité d'intérêt dans cette partie est de la forme $\mathbb{E}F(L_T)$ avec $F : \mathbb{R}^d \to \mathbb{R}$ est une fonction donnée. Dans ce cadre, nous approchons tout d'abord le processus de Lévy $(L_t)_{0\leq t\leq T}$ par un processus de Lévy simulable $(L_t^\varepsilon)_{0\leq t\leq T}$ avec $\varepsilon > 0$. Par ailleurs, à l'aide de la transformation d'Esscher (voir sous-section 3.5 du chapitre I), on peut trouver une fonction $G : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ et un compact K de $\mathbb{R}^d$ tel que

$$\forall\theta \in K, \quad \mathbb{E}F(L_T^\varepsilon) = \mathbb{E}G(\theta, L_T^{\varepsilon,\theta}), \tag{6}$$

où G est une fonction réelle qui prend ses valeurs dans $\mathbb{R}^d \times \mathbb{R}^d$. Ainsi, nous avons utilisé une version de l'algorithme de Robbins Monro projeté sur un compact (voir sous-section 2.3 du chapitre I) afin de déterminer $\arg\min_{\theta\in K} \mathrm{Var}(G(\theta, L_T^{\varepsilon,\theta}))$. Nous avons construit une suite $(\theta_i^\varepsilon)_{i\in\mathbb{N},0<\varepsilon<1}$ doublement indexée et nous avons démontré

$$\lim_{i\to\infty}(\lim_{\varepsilon\to 0}\theta_i^\varepsilon) = \lim_{\varepsilon\to 0}(\lim_{i\to\infty}\theta_i^\varepsilon) = \theta^* = \arg\min_{\theta\in K} \mathrm{Var}(G(\theta, L_T^\theta)), \quad \mathbb{P}\text{-}p.s.$$

Par la suite, pour tenir compte de l'erreur faible de l'approximation, on a supposé qu'il existe $\upsilon_\varepsilon \underset{\varepsilon\to 0}{\longrightarrow} 0$, pour $\varepsilon > 0$ tel que

$$\lim_{\varepsilon\to 0} \upsilon_\varepsilon^{-1} \left(\mathbb{E}F(L_T) - \mathbb{E}F(L_T^\varepsilon)\right) = C_F, \quad C_F \in \mathbb{R}. \tag{7}$$

Puis, en faisant dépendre N de ε, nous avons démontré un théorème central limite pour la méthode de Monte Carlo adaptative (voir Théorème 5.5 dans le Chapitre IV). Nous avons obtenu

$$\upsilon_\varepsilon^{-1} \left(\frac{1}{N(\varepsilon)} \sum_{i=1}^{N(\varepsilon)} G(\theta_i^\varepsilon, L_{T,i}^{\varepsilon,\theta_{i-1}^\varepsilon}) - \mathbb{E}F(L_T) \right) \xrightarrow[\varepsilon\to 0]{\mathcal{L}} \mathcal{N}\left(C_F, \mathrm{Var}(G(\theta^*, L_T^{\theta^*}))\right), \tag{8}$$

où $(L_{t,i}^{\varepsilon,\theta})_{0\le t\le T}$ est un échantillon i.i.d de même loi que $(L_t^{\varepsilon,\theta})_{0\le t\le T}$. En s'inspirant de l'idée de la méthode de Romberg Statistique dans le cadre du schéma d'Euler, nous avons mis en œuvre la procédure de Romberg Statistique dans le cadre d'approximation des processus de Lévy. En effet, en considérant ε et ε^β avec $\beta \in (0,1)$, nous approchons $\mathbb{E}F(L_T)$ par

$$Q_\varepsilon^{\mathrm{SR}} = \frac{1}{N_1(\varepsilon)} \sum_{i=1}^{N_1(\varepsilon)} G(\theta_1, \hat{L}_{T,i}^{\varepsilon^\beta,\theta_1}) + \frac{1}{N_2(\varepsilon)} \sum_{i=1}^{N_2(\varepsilon)} \left(G(\theta_2, L_{T,i}^{\varepsilon,\theta_2}) - G(\theta_2, L_{T,i}^{\varepsilon^\beta,\theta_2}) \right), \quad \forall \theta_1, \theta_2 \in \mathbb{R}^d, \tag{9}$$

où $(L_{T,i}^{\varepsilon,\theta_2})_{1\le i\le N_2}$ et $(L_{T,i}^{\varepsilon^\beta,\theta_2})_{1\le i\le N_2}$ sont i.i.d de même loi que $L_T^{\varepsilon,\theta_2}$ et $L_T^{\varepsilon^\beta,\theta_2}$ simulées à partir de la même trajectoire. $(\hat{L}_{T,i}^{\varepsilon^\beta,\theta_1})_{1\le i\le N_1}$ est un échantillon indépendant de même loi que $L_T^{\varepsilon^\beta,\theta_1}$. Pour un choix adéquat des tailles des échantillons $N_1(\varepsilon)$ et $N_2(\varepsilon)$, nous avons démontré le théorème central limite suivant

$$\upsilon_\varepsilon^{-1} \left(Q_\varepsilon^{\mathrm{SR}} - \mathbb{E}F(L_T) \right) \xrightarrow[\varepsilon\to 0]{\mathcal{L}} \mathcal{N}\left(C_F, \hat{\Sigma}^2(\theta_1) + \bar{\Sigma}^2(\theta_2)\right).$$

Les quantités $\hat{\Sigma}^2(\theta_1)$ et $\bar{\Sigma}^2(\theta_2)$ sont données dans le Théorème 3.3 du chapitre IV. Pour une large classe de processus de Lévy comme les processus tempérés stables d'indice Y, $Y \in (0,2)$, le gain en complexité de la méthode de Romberg Statistique par rapport à celle de Monte Carlo est de l'ordre de $\varepsilon^{Y(Y/2-1)}$ (voir sous section 3.3 du Chapitre IV). Ensuite, nous avons étudié la combinaison de la méthode de Romberg Statistique avec celle d'importance sampling. Ainsi, nous approchons $\mathbb{E}F(L_T)$ par

$$Q_\varepsilon^{\mathrm{ISSR}} = \frac{1}{N_1(\varepsilon)} \sum_{i=1}^{N_1(\varepsilon)} G(\theta_{1,i-1}^{\varepsilon^\beta}, \hat{L}_{T,i}^{\varepsilon^\beta,\theta_{1,i-1}^{\varepsilon^\beta}}) + \frac{1}{N_2(\varepsilon)} \sum_{i=1}^{N_2(\varepsilon)} \left(G(\theta_{2,i-1}^\varepsilon, L_{T,i}^{\varepsilon,\theta_{2,i-1}^\varepsilon}) - G(\theta_{2,i-1}^\varepsilon, L_{T,i}^{\varepsilon^\beta,\theta_{2,i-1}^\varepsilon}) \right),$$

avec $(\theta^{\varepsilon}_{1,i})_{i\in\mathbb{N},0<\varepsilon<1}$ et $(\theta^{\varepsilon}_{2,i})_{i\in\mathbb{N},0<\varepsilon<1}$ sont deux suites doublement indexées construites en utilisant un algorithme stochastique projeté sur un compact bien déterminé (voir sous-section 2.3 du chapitre I) et vérifiant

$$\begin{cases} \lim\limits_{i\to\infty}(\lim\limits_{\varepsilon\to 0}\theta^{\varepsilon}_{1,i}) = \lim\limits_{\varepsilon\to 0}(\lim\limits_{i\to\infty}\theta^{\varepsilon}_{1,i}) = \theta^*_1 = \arg\min_{\theta\in K}\hat{\Sigma}^2(\theta_1), \quad \mathbb{P}\text{-}p.s. \\ \lim\limits_{i\to\infty}(\lim\limits_{\varepsilon\to 0}\theta^{\varepsilon}_{2,i}) = \lim\limits_{\varepsilon\to 0}(\lim\limits_{i\to\infty}\theta^{\varepsilon}_{2,i}) = \theta^*_2 = \arg\min_{\theta\in K}\bar{\Sigma}^2(\theta_2), \quad \mathbb{P}\text{-}p.s. \end{cases} \tag{10}$$

Nous avons démontré aussi un théorème central limite pour la méthode de Romberg Statistique adaptative (voir Théorème 5.7 dans le Chapitre IV) donné par

$$v_{\varepsilon}^{-1}\left(Q^{\mathrm{ISSR}}_{\varepsilon} - \mathbb{E}F(L_T)\right) \xrightarrow[\varepsilon\to 0]{\mathcal{L}} \mathcal{N}\left(C_F, \hat{\Sigma}^2(\theta^*_1) + \bar{\Sigma}^2(\theta^*_2)\right).$$

Ainsi, nous obtenons dans cette thèse des algorithmes qui réduisent simultanément la variance et la complexité pour des modèles de diffusions avec et sans sauts.
Cette thèse est composée de quatre chapitres.

▶ **Le premier Chapitre** est un chapitre introductif qui rappelle les principaux résultats théoriques qui ont été utiles tout au long de cette thèse. D'abord, nous rappelons le principe de la méthode Euler Monte Carlo. Nous mettons en évidence les nouvelles méthodes qui ont été proposées dans la littérature et qui ont permis un gain en complexité en comparaison avec la méthode de Monte Carlo classique à savoir la méthode de Romberg Statistique et la méthode Euler Multilevel Monte Carlo. Ensuite, nous présentons la théorie de l'approximation stochastique dans l'exemple de Robbins Monro et ses différentes variantes à l'exemple de l'algorithme de Chen et l'algorithme projeté sur un compact. Enfin, nous introduisons les processus de Lévy et leur caractérisation, notamment par la représentation de Lévy-Khintchine et la décomposition d'Itô-Lévy. Nous passons en revue les exemples de modèles de Lévy utilisés dans la littérature financière et nous rappelons le principe de changement de probabilité des mesures de Lévy dont le cas particulier de la transformation d'Esscher.
▶ **Le deuxième Chapitre** étudie la méthode adaptative proposée, dans un modèle continu, comme combinaison de la méthode de Romberg Statistique et l'importance sampling. Un théorème central limite a été prouvé. Des illustrations numériques sont aussi proposées afin de mettre en évidence l'efficacité de cette nouvelle méthode en comparaison avec une méthode de Monte Carlo adaptative. Ce chapitre fait l'objet d'un article accepté pour publication dans le journal de Bernouilli `http://www.bernoulli-society.org/index.php/`

publications/bernoulli-journal/bernoulli-journal-papers.

▶ **Le troisième Chapitre** traite la combinaison de la méthode de Euler Multilevel Monte Carlo avec la méthode de l'importance sampling dans le même cadre continu que le deuxième chapitre. Un théorème central limite a été aussi prouvé. Une comparaison numérique a été effectuée entre les différentes méthodes adaptatives proposées. Ce chapitre fera l'objet d'une prépublication qui est en cours de rédaction.

▶ **Le quatrième Chapitre** étend la combinaison de Romberg Statistique et l'importance sampling à un modèle de proccessus de Lévy. Dans ce cadre, la nouvelle méthode adaptative sera aussi couronnée par la preuve d'un théorème central limite. On traite numériquement le cas du CGMY pour mettre en exergue l'efficacité de la méthode proposée. Ce chapitre a fait l'objet d'une prépublication soumise http://arxiv.org/abs/1408.0898.

Chapitre I

Préliminaires

Soit $W = (W_1, ..., W_q)$ un mouvement Brownien q-dimensionnel défini dans l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On munit cet espace de la filtration canonique $((\mathcal{F}_t)_{0\leq t\leq T}, T > 0)$ qui est la filtration Brownienne naturelle. On considère une diffusion d-dimensionnelle $(X_t)_{0\leq t\leq T}$ solution de l'équation différentielle stochastique :

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0 \in \mathbb{R}^d. \tag{I.1}$$

Les fonctions $b : \mathbb{R}^d \to \mathbb{R}$ et $\sigma : \mathbb{R}^q \times \mathbb{R}^d \to \mathbb{R}$ vérifient :

$$(\mathcal{H}_{b,\sigma}) \quad \exists C_T > 0; \forall x, y \in \mathbb{R}^d, \quad |b(y) - b(x)| + |\sigma(y) - \sigma(x)| \leq C_T|x - y|. \tag{I.2}$$

1 Approximation Euler Monte Carlo

L'évaluation des espérances de fonctions de diffusions multidimensionnelles sous la forme $\mathbb{E}[f(X_T)]$ ne cesse d'être d'un grand intérêt dans différents domaines. En particulier, lorsque l'on s'intéresse au pricing des produits dérivés en finance, on se ramène à calculer de quantités moyennes sous forme d'espérances. Le recours à la méthode de Monte Carlo (voir par exemple Glasserman [30] et Lapeyre, Pardoux et Sentis [6]) est très courant chaque fois que l'on est confronté au problème de calcul de ces quantités. Afin d'évaluer $\mathbb{E}[f(X_T)]$, la méthode de Monte Carlo suppose que l'on sait simuler la loi de la variable aléatoire X_T. Or, en général, on ne peut pas résoudre explicitement l'équation différentielle stochastique associée au processus $(X_t)_{0\leq t\leq T}$. En pratique, on est souvent ramené à discrétiser la diffusion par un schéma d'approximation $(X_t^n)_{0\leq t\leq T}$ tel que le schéma d'Euler, où n désigne le nombre de pas de temps associé à la discrétisation puis à approcher $\mathbb{E}[f(X_T^n)]$ par une méthode de Monte Carlo. La méthode Euler Monte Carlo a deux sources d'erreurs : une erreur liée à la discrétisation et une autre statistique.

En effet, on a

$$\mathbb{E}[f(X_T)] \simeq \mathbb{E}[f(X_T^n)] \simeq \frac{1}{N}\sum_{i=1}^{N} f(X_{T,i}^n).$$

1.1 Erreur de discrétisation dans le schéma d'Euler

On introduit le schéma d'Euler continu $(X_t^n)_{0\leq t\leq T}$, de pas de discrétisation défini par $\delta = T/n$, par :

$$dX_t^n = b(X_{\eta_n(t)})dt + \sigma(X_{\eta_n(t)})dW_t, \quad \eta_n(t) = [t/\delta]\delta. \tag{I.3}$$

On rappelle ici les résultats de convergence connus pour ce schéma.

Vitesse forte. Sous la condition $(\mathcal{H}_{b,\sigma})$, le schéma d'Euler vérifie les propriétés suivantes (voir par exemple Bouleau et Lépingle [15])

$\mathcal{P}1)$ $\forall p > 1, \quad \mathbb{E}\left[\sup_{t\in[0,T]} |X_t - X_t^n|^p\right] \leq \frac{K_p(T)}{n^{p/2}}, \quad K_p(T) > 0.$

$\mathcal{P}2)$ $\forall p > 1, \quad \mathbb{E}\left[\sup_{t\in[0,T]} |X_t|^p\right] + \mathbb{E}\left[\sup_{t\in[0,T]} |X_t^n|^p\right] \leq K_p'(T), \quad K_p'(T) > 0.$

Vitesse faible. Plusieurs travaux ont porté sur l'erreur de discrétisation donnée par :

$$\varepsilon_n = \mathbb{E}f(X_T^n) - \mathbb{E}f(X_T).$$

Initialement, Talay et Tubaro [62] ont démontré que pour des fonctions f, b, σ suffisamment régulières

$$\varepsilon_n = \frac{C}{n} + o(\frac{1}{n}), \quad \text{où } C \in \mathbb{R}. \tag{I.4}$$

Puis, Kloeden et Platen [39] ont démontré que si les fonctions f, b et σ sont $\mathscr{C}_P^4$, c'est à dire la fonction et ses dérivées sont quatre fois dérivables et à croissance polynomiale, alors on a

$$\varepsilon_n = \mathbb{E}f(X_T^n) - \mathbb{E}f(X_T) = O(1/n).$$

Dans le cadre des diffusions non dégénérées avec des conditions de type Hörmander, ces résultats ont été obtenus pour des fonctions mesurables et bornées (Bally et Talay[7, 8]). Plus récemment, Kebaier [37] a démontré le résultat suivant pour des fonctions f de classe $\mathscr{C}^1$.

Proposition 1.1. *Soit f une fonction à valeurs dans $\mathbb{R}^d$ satisfaisant*

$$(\mathcal{H}_f) \quad |f(x) - f(y)| \leq C(1 + |x|^p + |y|^p)|x - y|, \quad \textit{pour } C, p > 0.$$

Supposons que $\mathbb{P}(X_T \notin \mathcal{D}_{\mathbf{f}}) = 0$*, où* $\mathcal{D}_{\mathbf{f}} := \{x \in \mathbb{R}^d \mid f \text{ est dérivable en } x\}$*, alors*

$$\lim_{n\to\infty} \sqrt{n}\,\varepsilon_n = 0.$$

Il a aussi démontré que pour $\alpha \in [1/2, 1]$, il existe une fonction $\mathscr{C}^1$ à dérivées bornées et une diffusion X tel que

$$(\mathcal{H}_{\varepsilon_n}) \qquad \lim_{n\to\infty} n^{\alpha}\,\varepsilon_n = C_f, \quad C_f > 0.$$

Ainsi, l'ordre de l'erreur peut être de l'ordre de $1/n^{\alpha}$ pour tout $\alpha \in [1/2, 1]$.

1.2 Erreur statistique dans la méthode Euler Monte Carlo

Dans ce cadre, pour n fixé, le recours à la méthode de Monte Carlo permet d'approximer $\mathbb{E}f(X_T^n)$ par une moyenne empirique d'un échantillon i.i.d $(X_{T,i}^n)_{1\le i\le N}$ de même loi que X_T. La loi forte des grands nombres est le théorème de base pour la méthode de Monte Carlo et qui assure que pour n fixé et f telle que $\mathbb{E}\,|f(X_T^n)| < \infty$:

$$\frac{1}{N}\sum_{i=1}^{N} f(X_{T,i}^n) \underset{N\to\infty}{\longrightarrow} \mathbb{E}f(X_T^n), \quad p.s. \tag{I.5}$$

Théorème 1.2. *(Théorème central limite) Si on suppose que* $\mathbb{E}\,|f^2(X_T^n)| < \infty$ *alors :*

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} f(X_{T,i}^n) - \mathbb{E}f(X_T^n)\right) \overset{\mathcal{L}}{\underset{N\to+\infty}{\longrightarrow}} \mathcal{N}(0, \mathrm{Var}(f(X_T^n))).$$

Une question naturelle se pose : comment choisir N en fonction de n ?
En vue de répondre à cette question, nous rappelons tout d'abord le théorème central limite de Lindeberg Feller qui sera utile dans toutes les démonstrations des théorèmes centraux limites de cette thèse.

Theorem 1.3 (Théorème central limite de Lindeberg Feller)**.** *Pour tout* $n \in \mathbb{N}^*$*, nous considérons* $Y_{n,1}, Y_{n,2}, ..., Y_{n,k_n}$ *des variables aléatoires indépendantes centrées et de variance finie. Supposons que* $\lim_{n\to\infty} k_n = \infty$ *et que les hypothèses suivantes sont vérifiées :*

A1. Il existe une constante v *strictement positive, telle que*

$$\sum_{i=1}^{k_n} \mathbb{E}(Y_{n,i})^2 \underset{n\to\infty}{\longrightarrow} v.$$

A2. La condition de Lindeberg est vérifiée : à savoir, pour tout $\varepsilon > 0$,

$$\sum_{i=1}^{k_n} \mathbb{E}(|Y_{n,i}|^2 \mathbf{1}_{|Y_{n,i}|\geq \varepsilon}) \underset{n\to\infty}{\longrightarrow} 0.$$

On a alors

$$\sum_{i=1}^{k_n} Y_{n,i} \xrightarrow{\mathcal{L}} \mathcal{N}(0, v), \qquad \text{quand } n \to \infty.$$

Remark 1.4. *L'hypothèse suivante connue sous le nom de la condition de Lyapunov, implique la condition de Lindeberg A2.,*

A3. Il existe un nombre réel $a > 1$, tel que

$$\sum_{k=1}^{k_n} \mathbb{E}\left[|Y_{n,i}|^{2a}\right] \underset{n\to\infty}{\longrightarrow} 0.$$

En utilisant le Théorème 1.3, on démontre ici un théorème central limite de type Lindeberg Feller pour la méthode Euler Monte Carlo classique.

Théorème 1.5. *Soit $f : \mathbb{R}^d \to \mathbb{R}$ une fonction qui satisfait l'hypothèse ($\mathcal{H}_f$) et ($\mathcal{H}_{\varepsilon_n}$) telle que pour $\alpha \in [1/2, 1]$, on a Alors, pour $N = n^{2\alpha}$, on a*

$$n^\alpha \left(\frac{1}{n^{2\alpha}} \sum_{i=1}^{n^{2\alpha}} \Big(f(X^n_{T,i}) - \mathbb{E}f(X_T) \Big) \right) \underset{n\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}\left(C_f, \mathrm{Var}(f(X_T))\right).$$

Preuve. Tout d'abord, on écrit l'erreur totale comme suit :

$$\frac{1}{N}\sum_{i=1}^{N} f(X^n_{T,i}) - \mathbb{E}f(X_T) = \frac{1}{N}\sum_{i=1}^{N} \Big(f(X^n_{T,i}) - \mathbb{E}f(X^n_T) \Big) + \left(\mathbb{E}f(X^n_T) - \mathbb{E}f(X_T)\right). \tag{I.6}$$

L'hypothèse ($\mathcal{H}_{\varepsilon_n}$) assure que $n^\alpha \left(\mathbb{E}f(X^n_T) - \mathbb{E}f(X_T)\right) \underset{n\to\infty}{\longrightarrow} C_f$. Pour étudier le comportement asymptotique du premier terme dans (I.6) donné par $\frac{1}{N}\sum_{i=1}^{N} f(X^n_{T,i}) - \mathbb{E}f(X^n_T)$, on applique le Théorème 1.3. Ainsi, on divise la preuve en deux étapes :

Étape 1. On vérifie l'hypothèse *A1.* : On pose $Y_{n,i} = \frac{n^\alpha}{N}\Big(f(X^n_{T,i}) - \mathbb{E}f(X^n_T)\Big)$.

$$\begin{aligned} \sum_{i=1}^{N} \mathbb{E}(Y^2_{n,i}) &= \frac{n^{2\alpha}}{N}\mathrm{Var}(f(X^n_{T,i})) \quad \text{en prenant } N = n^{2\alpha}, \\ &= \mathrm{Var}(f(X^n_T)). \end{aligned}$$

Sous l'hypothèse $(\mathcal{H}_f)$, il s'ensuit des propriétés $\mathcal{P}1)$ et $\mathcal{P}2)$ que

$$\sum_{i=1}^{N} \mathbb{E}(Y_{n,i}^2) \underset{n\to\infty}{\longrightarrow} \mathrm{Var}(f(X_T)).$$

Étape 2. On vérifie maintenant l'hypothèse de Lyapounov *A3.* : Soit $a > 1$, on écrit :

$$\begin{aligned}\sum_{i=1}^{N} \mathbb{E}(|Y_{n,i}|^{2a}) &= \frac{1}{n^{2\alpha a}} \sum_{i=1}^{N} \mathbb{E}\left[\left|f(X_{T,i}^n) - \mathbb{E}f(X_T^n)\right|^{2a}\right] \\ &\leq 2^{2a-1} \frac{1}{n^{2\alpha(a-1)}} \left(\mathbb{E}|f(X_T^n)|^{2a} + |\mathbb{E}f(X_T^n)|^{2a}\right).\end{aligned}$$

De la même hypothèse $(\mathcal{H}_f)$, il s'ensuit de la propriété $\mathcal{P}2)$ que

$$\sup_{n\in\mathbb{N}} \mathbb{E}|f(X_T^n)|^{2a} < \infty.$$

Ainsi, pour $a > 1$

$$\sum_{i=1}^{n^{2\alpha}} \mathbb{E}(|Y_{n,i}|^{2a}) \underset{n\to\infty}{\longrightarrow} 0.$$

Par conséquent, on obtient le résultat désiré grâce au Théorème 1.3. □

Pour une erreur faible du schéma d'Euler de l'ordre de $1/n^{\alpha}$, la complexité optimale de l'algorithme de Monte Carlo est donnée par :

$$C_{MC} = C \times nN = C \times n^{2\alpha+1}, \text{ avec } C > 0.$$

On rappelle que la complexité d'un algorithme donné est le nombre d'itérations essentielles contenues dans ce dernier. Pour améliorer la convergence de cette méthode, une grande importance est accordée aux techniques de réduction de variance afin de réduire l'erreur statistique de la méthode Euler Monte Carlo. Les méthodes les plus connues sont la méthode de variables de contrôle, la technique des variables antithétiques dont le choix est spécifique au problème traité (Pour plus de détails sur les techniques de réduction de variance nous citons Boyle, Broadie et Glasserman [52] et Lapeyre, Pardoux et Sentis [6]).

1.3 La méthode de Romberg Statistique

Nous considérons, en premier lieu, la méthode de Romberg Statistique développée par Kebaier [37] dans sa thèse, qui s'inspire des méthodes de réduction de variance par variable de contrôle en permettant de réduire la complexité par rapport à une méthode Euler Monte Carlo classique. Le principe de cette méthode est de considérer deux schémas d'Euler, un premier schéma "fin" de pas de discrétisations $\frac{T}{n}$ et un deuxième schéma "grossier" de pas de discrétisations $\frac{T}{m}$ (avec $m << n$). Ainsi, on écrit :

$$\mathbb{E}f(X_T^n) = \mathbb{E}f(X_T^m) + \mathbb{E}(f(X_T^n) - f(X_T^m)).$$

Ensuite, on approche $\mathbb{E}[f(X_T^n)]$ par :

$$V_n := \frac{1}{N_1}\sum_{i=1}^{N_1} f(\hat{X}_{T,i}^m) + \frac{1}{N_2}\sum_{i=1}^{N_2}\left(f(X_{T,i}^n) - f(X_{T,i}^m)\right),$$

où N_1 et N_2 désignent les tailles des échantillons i.i.d des deux estimations Monte Carlo. Les variables aléatoires du premier Monte Carlo sont des copies indépendantes de $\hat{X}_T^m$ et les variables aléatoires du second Monte Carlo sont aussi des copies indépendantes de X_T^n et X_T^m.
Dans sa thèse, Kebaier [37] a démontré un théorème central limite pour la méthode de Romberg Statistique. Il suppose d'abord que les paramètres de la méthode dépendent uniquement de n à savoir

$$m = n^\beta, \beta \in (0,1), \quad N_1 = n^{\gamma_1}, \gamma_1 > 1, \quad N_2 = n^{\gamma_2}, \gamma_2 > 1.$$

La démonstration du théorème central limite repose fondamentalement sur le résultat de Jacod et Protter [35] portant sur la convergence stable de la loi de l'erreur du schéma d'Euler pour les diffusions. Ce résultat est donné par :

$$\sqrt{n}(X^n - X) \Rightarrow U, \tag{I.7}$$

où U est un processus limite donné (voir Théorème 2.1 dans Kebaier [37]).

Théorème 1.6. *(Voir Théorème 3.2 dans Kebaier [37]) Soit $f : \mathbb{R}^d \to \mathbb{R}$ une fonction satisfaisant l'hypothèse $(\mathcal{H}_f)$ et $(\mathcal{H}_{\varepsilon_n})$. Supposons de plus que $\mathbb{P}(X_T \notin \mathcal{D}_\mathbf{f}) = 0$, où $\mathcal{D}_\mathbf{f} := \{x \in \mathbb{R}^d \mid f$ est dérivable en $x\}$. On a alors, pour $\gamma_1 = 2\alpha$ et $\gamma_2 = 2\alpha - \beta$,*

$$n^\alpha\left(V_n - \mathbb{E}(f(X_T))\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(C_f, \Sigma^2\right),$$

où $\Sigma^2 = \mathrm{Var}(f(X_T)) + \mathrm{Var}(\nabla f(X_T).U_T)$.

L'avantage de cette méthode apparaît dans le gain en complexité en comparaison avec la méthode de Monte Carlo classique.

$$\begin{aligned} C_{SR} &= C \times (mN_1 + (n+m)N_2) \quad \text{avec } C > 0, \\ &= C \times (n^{\beta+2\alpha} + (n+n^{\beta})n^{2\alpha-\beta}). \end{aligned} \tag{I.8}$$

Pour le choix optimal de $\beta^* = 1/2$, on obtient une complexité

$$C^*_{SR} = C \times n^{2\alpha+1/2} << C_{MC} = C \times n^{2\alpha+1}.$$

Par ailleurs, pour les options Asiatiques, Kebaier [37] a démontré un théorème central limite pour le schéma des trapèzes et a obtenu une complexité donnée par :

$$C_{SR} = C \times (n^{\beta+2\alpha} + (n+n^{\beta})n^{2\alpha-2\beta}).$$

Pour le choix optimal de $\beta^* = 1/3$, nous obtenons la complexité

$$C^*_{SR} = C \times n^{2\alpha+1/3} << C_{MC} = n^{2\alpha+1}.$$

1.4 La méthode de Multilevel Monte Carlo

Récemment introduite par Giles [28], la méthode de Multilevel Monte Carlo peut être vue comme une généralisation de la méthode de Romberg Statistique (qu'on peut nommer aussi par "two-level Monte Carlo").

1.4.1 Présentation générale

Dans un cadre général, Giles [28] considère des schémas d'approximation avec des pas de temps différents $h_\ell = m^{-\ell}T$, $\ell \in \{0,..,L\}$. On note par P le payoff $f(X_T)$ et par $\hat{P}_\ell$ l'approximation de P en utilisant une discrétisation numérique avec un pas de temps h_ℓ. Alors, on écrit

$$\mathbb{E}(f(X^n_T)) = \mathbb{E}[\hat{P}_L] = \mathbb{E}[\hat{P}_0] + \sum_{\ell=1}^{L} \mathbb{E}\left[\hat{P}_\ell - \hat{P}_{\ell-1}\right].$$

L'idée de la méthode de Multilevel Monte Carlo est d'estimer indépendamment chacune des espérances de droite dans l'expression ci-dessous de manière à minimiser la complexité. Soit $\hat{Y}_0$ un estimateur de $\mathbb{E}[\hat{P}_0]$ avec N_0 échantillons et soit $\hat{Y}_\ell, \ell > 0$ un estimateur de $\mathbb{E}\left[\hat{P}_\ell - \hat{P}_{\ell-1}\right]$

avec N_ℓ échantillons.

$$\forall \ell > 0, \quad \hat{Y}_\ell = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left(\hat{P}_{\ell,i} - \hat{P}_{\ell-1,i} \right). \tag{I.9}$$

Ici, $(\hat{P}_{\ell,i})_{1 \le i \le N_\ell}$ et $(\hat{P}_{\ell-1,i})_{1 \le i \le N_\ell}$ sont issus de la même trajectoire Brownienne mais avec des pas de discrétisation respectifs h_ℓ et $h_{\ell-1}$. La méthode de Multilevel Monte Carlo exploite le fait que la variance $\mathrm{Var}\left[P_\ell - P_{\ell-1}\right]$ décroit avec ℓ et choisit convenablement les tailles des échantillons N_ℓ afin de minimiser le coût de calcul pour atteindre l'erreur des moindres carrées voulue. Ceci est résumé dans le théorème de complexité ci-dessous.

Théorème 1.7. *(voir Théorème 3.1 dans Giles [28]) S'ils existent des estimateurs indépendants $\hat{Y}_\ell$ basés sur les nombres de Monte Carlo N_ℓ et des constantes $\alpha \ge \frac{1}{2}, \beta, c_1, c_2, c_3$ strictement positives tels que*

i) $\mathbb{E}\left[\hat{P}_\ell - P\right] \le c_1 h_\ell^\alpha$,

ii) $\mathbb{E}\left[\hat{Y}_\ell\right] = \begin{cases} \mathbb{E}\hat{P}_0, & \ell = 0, \\ \mathbb{E}\left(\hat{P}_\ell - \hat{P}_{\ell-1}\right), & \ell > 0, \end{cases}$

iii) $\mathrm{Var}(\hat{Y}_\ell) \le c_2 N_\ell^{-1} h_\ell^\beta$,

iv) C_ℓ,la complexité de $\hat{Y}_\ell$ vérifie

$$C_\ell \le c_3 N_\ell h_\ell^{-1},$$

alors il existe une constante c_4 strictement positive tel que pour tout $\varepsilon < e^{-1}$, on a des valeurs L et N_ℓ pour lequels l'estimateur multilevel

$$\hat{Y} = \sum_{\ell=0}^{L} \hat{Y}_\ell,$$

a une erreur des moindres carrées MSE qui vérifie

$$MSE \equiv \mathbb{E}\left[\left(\hat{Y} - \mathbb{E}[P]\right)^2\right] < \varepsilon^2,$$

avec une complexité C_{ML} avec les bornes

$$C_{ML} \le \begin{cases} c_4 \varepsilon^{-2}, & \beta > 1, \\ c_4 \varepsilon^{-2}(\log(\varepsilon)^2), & \beta = 1, \\ c_4 \varepsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < 1. \end{cases}$$

Ce théorème montre l'importance de la détermination du paramètre β approprié dans *iii*). En effet, $\beta = 1$ correspond au schémas d'ordre 1 à l'exemple du schéma d'Euler. $\beta > 1$

correspond aux shémas d'ordre plus fort à l'exemple du schéma de Milstein.

1.4.2 Approximation Euler Multilevel Monte Carlo

Dans le cadre particulier de la discrétisation avec le schéma d'Euler, on a

$$\mathbb{E}f(X_T^n) = \mathbb{E}f(X_T^{m^0}) + \sum_{\ell=1}^{L} \mathbb{E}\left(f(X_T^{m^\ell}) - f(X_T^{m^{\ell-1}})\right), \quad \text{avec } n = m^L.$$

Ensuite, on approche $\mathbb{E}f(X_T^n)$ par $(L+1)$ moyennes empiriques indépendantes

$$Q_n := \frac{1}{N_0}\sum_{i=1}^{N_0} f(X_{T,i}^1) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\left(f(X_{T,i}^{\ell,m^\ell}) - f(X_{T,i}^{\ell,m^{\ell-1}})\right). \tag{I.10}$$

Ici, pour $\ell \in \{1, ..., L\}$, $(X_{T,i}^{\ell,m^\ell}, X_{T,i}^{\ell,m^{\ell-1}})_{1\leq i\leq N_\ell}$ est une suite i.i.d de même loi que $(X_T^{\ell,m^\ell}, X_T^{\ell,m^{\ell-1}})$. En plus, les simulations X_T^{ℓ,m^ℓ} et $X_T^{\ell,m^{\ell-1}}$ sont issues de la même trajectoire Brownienne mais avec des pas de discrétisation respectivement donnés par $m^{-\ell}T$ et $m^{-(\ell-1)}T$. Pour $\ell = 0$, $(X_{T,i}^1)_{1\leq i\leq N_0}$ est une suite i.i.d de même loi que X_T^1 résultante d'un schéma d'Euler de pas de discrétisation égal à T. Ainsi, la variance de l'estimateur Multilevel (I.10) est donnée par

$$\mathrm{Var}(Q_n) = N_0^{-1}\mathrm{Var}(f(X_T^1)) + \sum_{\ell=1}^{L} N_\ell^{-1}\sigma_\ell^2,$$

avec $\sigma_\ell^2 = \mathrm{Var}(f(X_T^{\ell,m^\ell}) - f(X_T^{\ell,m^{\ell-1}}))$. En particulier, dans le cas où f est une fonction lipschitzienne, sous $(\mathcal{H}_{b,\sigma})$, on a $\sigma_\ell^2 = O(m^{-\ell})$ et $\mathrm{Var}(Q_n) \leq c_2\sum_{\ell=0}^{L} N_\ell^{-1}m^{-\ell}, c_2 > 0$. Afin d'obtenir une erreur des moindres carrés (MSE) de l'ordre de $1/n^\alpha$, Giles [28] obtient des tailles d'échantillons optimales données par :

$$N_\ell = 2c_2 n^{2\alpha}(\frac{\log n}{\log m} + 1)\frac{T}{m^\ell}, \quad \ell \in \{0, ..., L\} \text{ et } L = \frac{\log n}{\log m}.$$

Ainsi, la complexité optimale de l'estimateur Multilevel Monte Carlo est proportionnelle à $n^{2\alpha}(\log n)^2$. Plus récemment, Ben Alaya et Kebaier [11] ont établi un théorème central limite de type Lindeberg-Feller pour la méthode Euler Multilevel Monte Carlo. Tout d'abord, ils ont démontré un théorème de convergence stable de la loi de l'erreur du schéma d'Euler pour deux niveaux consécutifs $m^{\ell-1}$ et m^ℓ (voir Théorème 3 dans Ben Alaya et Kebaier [11]), dans l'esprit du résultat obtenu précédemment par Jacod et Protter [35]. Pour ce faire, ils réécrivent la diffusion (I.1) comme suit

$$dX_t = \varphi(X_t)dY_t = \sum_{j=0}^{q} \varphi_j(X_t)dY_t^j,$$

où φ_j est la j-ème colonne de la matrice σ, pour $1 \leq j \leq q$, $\varphi_0 = b$ et $Y_t := (t, W_t^1, ..., W_t^q)'$. Alors, le schéma d'Euler continu X^n avec pas de temps $\delta = T/n$ devient

$$dX_t^n = \varphi(X_{\eta_n(t)}^n)dY_t = \sum_{j=0}^{q} \varphi_j(X_{\eta_n(t)}^n)dY_t^j, \quad \eta_n(t) = [t/\delta]\delta.$$

Avec ces notations, le théorème de convergence stable est donné par

Théorème 1.8. *Supposons que b et σ sont $\mathscr{C}^1$ et à croissance linéaire alors on a*

$$\text{pour tout } m \in \mathbb{N} \setminus \{0,1\}, \ \sqrt{\frac{m^\ell}{(m-1)T}} \left(X^{m^\ell} - X^{m^{\ell-1}}\right) \stackrel{stably}{\Longrightarrow} U, \quad \text{quand } \ell \to \infty,$$

avec $(U_t)_{0 \leq t \leq T}$ est un processus d-dimensionel vérifiant

$$U_t = \frac{1}{\sqrt{2}} \sum_{i,j=1}^{q} Z_t \int_0^t H_s^{i,j} dB_s^{ij}, \quad t \in [0,T], \tag{I.11}$$

où

$$H_s^{i,j} = (Z_s)^{-1} \dot{\varphi}_{s,j} \bar{\varphi}_{s,i}, \ \text{avec } \dot{\varphi}_{s,j} := \nabla \varphi_j(X_s) \text{ et } \bar{\varphi}_{s,i} := \varphi_i(X_s), \tag{I.12}$$

et $(Z_t)_{0 \leq t \leq T}$ est le processus de $\mathbb{R}^{d \times d}$ solution de l'équation linéaire suivante

$$Z_t = I_d + \sum_{j=0}^{q} \int_0^t \dot{\varphi}_{s,j} dY_s^j Z_s, \quad t \in [0,T].$$

Ici, $\nabla \varphi_j$ est une matrice $d \times d$ avec $(\nabla \varphi_j)_{ik}$ est la dérivée partielle de φ_{ij} respectivement à la k-ème cordonnée et $(B^{ij})_{1 \leq i,j \leq q}$ est un mouvement Brownien standard q^2-dimensionel indépendant de W. Ce processus est défini dans l'extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t \geq 0}, \tilde{\mathbb{P}})$ de l'espace $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$.

La démonstration du théorème central limite repose sur ce résultat de convergence stable. Ils obtiennent un choix optimal des tailles des échantillons explicites données par :

$$N_\ell = \frac{n^2(m-1)T}{m^\ell a_\ell} \sum_{\ell=1}^{L} a_\ell, \quad \ell \in \{0, ..., L\} \text{ et } L = \frac{\log n}{\log m}, \tag{I.13}$$

où $(a_\ell)_{\ell\geq 0}$ est une suite strictement positive satisfaisant pour $p > 2$,

$$\lim_{L\to\infty}\sum_{\ell=1}^{L} a_\ell = \infty \text{ et } \lim_{L\to\infty}\frac{1}{(\sum_{\ell=1}^{L} a_\ell)^{p/2}}\sum_{\ell=1}^{L} a_\ell^{p/2} = 0.$$

Ainsi, l'estimateur de Euler Multilevel Monte Carlo satisfait le théorème central limite suivant :

Théorème 1.9. *(voir Théorème 5 dans Ben Alaya et Kebaier [11]). Supposons que* $\mathbb{P}(X_T \notin \mathcal{D}_{\mathbf{f}}) = 0$ *où* $\mathcal{D}_{\mathbf{f}}$ *est le domaine où* f *est dérivable et que* f *satisfait* $(\mathcal{H}_f)$ *et* $(\mathcal{H}_{\varepsilon_n})$*, alors*

$$n^{\alpha}\left(Q_n - \mathbb{E}f(X_T)\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(C_f, \tilde{\Sigma}^2\right),$$

où $\tilde{\Sigma}^2$ *est la variance limite donnée par* $\mathrm{Var}(\nabla f(X_T).U_T)$.

La complexité de la méthode Euler Multilevel Monte Carlo, utilisant les tailles $(N_\ell, \ell \in \{0,..,L\})$ proposées par la relation (I.13), est donnée par :

$$\begin{aligned} C_{MLMC} &= C \times \left(N_0 + \sum_{\ell=1}^{L} N_\ell (m^\ell + m^{\ell-1})\right) \qquad \text{avec } C > 0, \\ &= C \times \left(\frac{n^{2\alpha}(m-1)T}{a_0}\sum_{\ell=1}^{L} a_\ell + n^{2\alpha}\frac{(m^2-1)T}{m}\sum_{\ell=1}^{L}\frac{1}{a_\ell}\sum_{\ell=1}^{L} a_\ell\right). \end{aligned}$$

Le minimum du deuxième terme dans l'expression de la complexité est atteint pour $a_\ell^* = 1$, $\ell \in \{1, ..., L\}$. Ainsi, la complexité optimale pour la méthode Euler Multilevel Monte Carlo est donnée par

$$C^*_{MLMC} = C \times \left(\frac{(m-1)T}{a_0 \log m} n^{2\alpha}\log n + \frac{(m^2-1)T}{m(\log m)^2} n^{2\alpha}(\log n)^2\right) = O(n^2(\log n)^2).$$

Ceci est en complète concordance avec les résultats de Giles [28] déjà rappelé dans le Théorème 1.7. Par ailleurs, pour les options Asiatiques, Ben Alaya et Kebaier [12] ont démontré un théorème limite central pour les deux schémas d'approximation de Riemann et des trapèzes et ont obtenu une complexité se rapprochant de n^2 selon les choix des $a_\ell, \{\ell = 0, .., L\}$ à savoir $n^2 \log n, n^2 \log\log n, n^2 \log\log\log n$, ce qui peut être d'un point de vue pratique assimilé à l'ordre n^2 obtenu par Giles [28] pour les schémas d'ordre 2.

2 Algorithmes stochastiques de type Robbins Monro

Un algorithme stochastique de type Robbins Monro est une procédure récursive de recherche de zéro d'une fonction moyenne qui admet une représentation sous formes d'espérance. Plus précisément, le but est de trouver l'ensemble :

$$h^{-1}(0) = \{\theta \in \mathbb{R}^q \quad ; \quad h(\theta) = 0\}, \text{ avec } h(x) = \mathbb{E}(H(x, Y)),$$

$H : \mathbb{R}^q \times \mathbb{R}^k \to \mathbb{R}^q$ est une fonction borélienne et Y est un vecteur aléatoire à valeurs dans $\mathbb{R}^k$.

2.1 Algorithme de Robbins Monro

L'étude des algorithmes stochastiques a commencé avec les travaux de Robbins Monro [57] et de Kiefer-Wolfowitz [38] (variante introduisant une méthode de différences finies à pas décroissant pour l'approximation d'une fonction moyenne de type gradiant). En particulier, l'algorithme de Robbins Monro a fait l'objet de nombreuses applications dans différents domaines à l'exemple de la physique, la mécanique et les statistiques. La procédure récursive mise en œuvre s'écrit :

$$\theta_0 \in \mathbb{R}^q \text{ et } \theta_{n+1} = \theta_n - \gamma_{n+1} H(\theta_n, Y_{n+1}), \tag{I.14}$$

avec :

i) $(\gamma_n)_{n\geq 1}$ est une suite strictement positive et décroissante appelée *pas de l'algorithme.*

ii) $(Y_n)_{n\geq 1}$ est une suite de variables aléatoires i.i.d de même loi que Y.

La convergence presque sûre de la suite $(\theta_n)_{n\geq 0}$ vers les zéros de la fonction h a été traité dans plusieurs références. Lorsque h admet un unique zéro, on peut par exemple citer Duflo [25] (voir Théorème 2.2.12 dans Duflo [25]) où on trouve une démonstration du théorème suivant.

Théorème 2.1. *Soit $(\theta_n)_{n\geq 0}$ la suite définie par (I.14). Supposons de plus que la fonction h soit continue. Sous les hypothèses suivantes :*

A1. Il existe un unique $\theta^* \in \mathbb{R}^q$ tel que $h(\theta^*) = 0$ et pour tout $\theta \neq \theta^*$, $(\theta - \theta^*.h(\theta)) > 0$,

A2. $\mathbb{E}(|H(\theta, Y)|^2) \leq K\left(1 + |\theta|^2\right), \quad K > 0$,

A3. $\sum_{n>0} \gamma_n = \infty$ et $\sum_{n>0} \gamma_n^2 < \infty$,

la suite $(\theta_n)_{n\geq 0}$ converge presque sûrement vers θ^*.

Remarque 2.2. *La première hypothèse A1. revient à dire que h dérive d'une fonction strictement convexe. L'hypothèse A2., dite aussi la condition de non explosion de l'algorithme, impose qu'en moyenne la fonction h ait un comportement sous-linéaire. Cette hypothèse peut ne pas être satisfaite en pratique.*

Pour contourner le problème de l'explosion de l'algorithme, des versions plus robustes de l'algorithme de Robbins Monro ont été proposées dans la littérature dont on peut citer l'algorithme de Chen et l'algorithme projeté sur un compact (appelé aussi algorithme contraint).

2.2 Algorithme de Chen

Dans leurs travaux d'amélioration de l'algorithme de Robbins Monro, Chen et Zhu [18] proposent une méthode de projections aléatoires de l'algorithme et démontrent la convergence presque sûre. Le principe de cette méthode est de considérer un ensemble de compacts croissants sur lesquels on projette l'algorithme de Robbins Monro dans le but d'éviter que la suite $(\theta_n)_{n\geq 0}$ n'explose pendant les premières itérations. Plus précisément, on considère une suite de compacts $(\mathcal{K}_j)_{j\geq 0}$ tels que

$$\cup_{j=0}^{\infty}\mathcal{K}_j = \mathbb{R}^d \text{ et pour tout } j, \mathcal{K}_j \subsetneq (\overset{\circ}{\mathcal{K}}_{j+1}), \tag{I.15}$$

et on définit les suites de variables aléatoires $(\bar{\theta}_n)_{n\geq 0}$ et $(\alpha_n)_{n\geq 0}$ avec $\bar{\theta}_0 \in \mathcal{K}_0$, $\alpha_0 = 0$,

$$\begin{cases} \textbf{si} & \bar{\theta}_n - \gamma_{n+1}H(\bar{\theta}_n, Y_{n+1}) \in \mathcal{K}_{\alpha(n)}, \ \textbf{alors} \\ & \bar{\theta}_{n+1} = \bar{\theta}_n - \gamma_{n+1}H(\bar{\theta}_n, Y_{n+1}), \ \text{et } \alpha_{n+1} = \alpha_n \\ \textbf{sinon} & \bar{\theta}_{n+1} = \bar{\theta}_0 \text{ et } \alpha_{n+1} = \alpha_n + 1. \end{cases} \tag{I.16}$$

Ainsi la suite $(\alpha_n)_{n\geq 0}$ compte le nombre de troncatures jusqu'à l'instant n. La convergence presque sûre de l'algorithme tronqué est donnée par le théorème suivant.

Théorème 2.3 (Convergence presque sûre de l'algorithme de Chen)**.** *Supposons que l'hypothèse de convexité A1. et l'hypothèse A3. soient satisfaites et que :*

(A4.) Pour tout $q > 0$, la série $\sum_n \gamma_{n+1}(H(\bar{\theta}_n, Y_{n+1}) - h(\bar{\theta}_n))\mathbf{1}_{\{|\bar{\theta}_n - \theta^*| \leq q\}}$ converge presque sûrement,

alors la suite $(\bar{\theta}_n)_{n\geq 0}$ définie par (I.16) converge presque sûrement vers θ^* et de plus la suite $(\alpha_n)_{n\geq 0}$ est finie p.s. (i.e presque sûrement les projections se produisent un nombre fini de fois).

Remarque 2.4. *La suite $(\bar{\theta}_n)_{n\geq 0}$ (donnée par (I.16)) doit rester p.s. dans un compact donné et lorsqu'elle en sort, l'algorithme est réinitialisé et on considère un compact plus grand.*

Plus récemment, Lelong [46] propose une version améliorée du résultat de Chen et Zhu [18] sur la convergence presque sûre de l'algorithme tronqué en proposant une condition de convergence locale plus facile à vérifier en pratique.

Proposition 2.5. *Sous les hypothèses (A1.), (A3.) et si la fonction $\theta \longmapsto \mathbb{E}(|H(\theta, Y)|^2)$ est* *bornée sur tout compact alors la suite* $(\bar{\theta}_n)_n$ *définie par (I.16) converge presque sûrement vers* θ^* *et de plus la suite* $(\alpha_n)_{n\geq 0}$ *est finie p.s.*

2.3 Algorithme projeté sur un compact

Dans le même cadre d'amélioration de l'algorithme de Robbins Monro, on introduit ici les algorithmes contraints projetés sur un compact. En effet, on considère un compact convexe $K \subset \mathbb{R}^q$ et le but est de déterminer l'ensemble $\{\theta \in K : h(\theta) = \mathbb{E}H(\theta, Y) = 0\}$. Pour $\tilde{\theta}_0 \in K$, on considère la suite $(\tilde{\theta}_n)_{n\geq 0}$ à valeurs dans $\mathbb{R}^q$ définie par

$$\tilde{\theta}_{n+1} = \Pi_K \left[\tilde{\theta}_n + \gamma_{n+1} H(\tilde{\theta}_n, Y_{n+1})\right], \quad n \geq 0, \tag{I.17}$$

où $(Y_n)_{n\geq 1}$ est une suite i.i.d de même loi que Y et Π_K désigne la projection euclidienne sur le compact K. La convergence presque sûre de cet algorithme est donné par le théorème suivant (voir Laruelle, Lehalle et Pagès [44] et Kushner et Yin [41] pour plus de détails).

Théorème 2.6. *Soit* $(\tilde{\theta}_n)_{n\geq 0}$ *une suite définie par (I.17). Supposons qu'il existe un unique* $\tilde{\theta}^* \in K$ *tel que* $h(\tilde{\theta}^*) = 0$ *et que la fonction* h *vérifie* $\forall \theta \neq \tilde{\theta}^* \in K, \quad (\theta - \tilde{\theta}^*).h(\theta) > 0$. *Supposons que la suite* $(\gamma_n)_{n\geq 1}$ *vérifie* $\sum_{n>0} \gamma_n = \infty$ *et* $\sum_{n>0} \gamma_n^2 < \infty$. *De plus, si la fonction* H *vérifie*

$$\forall \theta \in K, \quad \mathbb{E}\left[|H(\theta, Y)|^2\right] \leq C(1 + |\theta|^2), \quad C > 0,$$

alors $\tilde{\theta}_n \xrightarrow[n\to+\infty]{p.s.} \tilde{\theta}^*$.

3 Processus de Lévy

Les modèles discontinus en Finance ont été introduits par Merton [51] afin de prendre en considération la présence de certaines variations brusques du cours de sous-jacents dans la valorisation d'options. Les processus à saut habituellement utilisés dans la modélisation en mathématiques financières sont les processus de Lévy.

3.1 Définition d'un processus de Lévy

Définition 3.1. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité. On appelle un processus de Lévy un processus stochastique à trajectoires càdlàg* $(X_t)_{t\geq 0}$ *à valeurs dans* $\mathbb{R}^d$, *tel que* $X_0 = 0$ *p.s., vérifiant :*

i) Indépendance des accroissements : pour tout $n \geq 1$ *et pour toute suite croissante de temps* $t_0, ..., t_n$*, les variables aléatoires* $X_{t_0}, X_{t_1} - X_{t_0}, ..., X_{t_n} - X_{t_{n-1}}$ *sont indépendantes.*

ii) Accroissements stationnaires : la distribution de $X_{t+h} - X_t$ *ne dépend pas de* t*.*

iii) Continuité stochastique : Pour tout $\varepsilon > 0$ *et pour tout* $t > 0$ $\lim_{h \to 0} \mathbb{P}\left(|X_{t+h} - X_t| \geq \varepsilon\right) = 0$*.*

Un exemple très connu des processus de Lévy est le processus de Poisson et le mouvement Brownien. Soit $(\tau_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes de loi exponentielle et de même paramètre $\lambda > 0$. On pose $T_n = \sum_{i=1}^{n} \tau_i$ pour $n \in \mathbb{N}^*$ et pour tout $t \in \mathbb{R}^+$: $N_t = \sum_{n\geq 1} \mathbf{1}_{T_n \leq t}$. Le processus $(N_t, t \in \mathbb{R}^+)$ est par définition un processus de Poisson de paramètre $\lambda > 0$. Le processus de Poisson est ainsi défini comme un processus de comptage. Pour t donné, N_t compte le nombre de temps aléatoires $(T_n)_{n\geq 1}$ qui ont lieu entre 0 et t. Un processus de Poisson composé $(X_t)_{t\geq 0}$ est défini par $X_t = \sum_{i=1}^{N_t} Y_i$, où $(N_t)_{t\geq 0}$ est un processus de Poisson de paramètre λ et les $(Y_i)_{i\geq 1}$ sont des variables aléatoires i.i.d de loi commune π et indépendantes de $(N_t)_{t\geq 0}$. On a pour tout $t \in \mathbb{R}^+$ et pour $u \in \mathbb{R}^d$:

$$\mathbb{E}\exp(iu \cdot X_t) = \mathbb{E}\left[\mathbb{E}\left[\exp(iu \cdot X_T)|N_t\right]\right] = \exp(t\lambda \int_{\mathbb{R}^d} (e^{iu\cdot x} - 1)\pi(dx)).$$

Le paramètre λ est appelé intensité des sauts et π est la distribution des sauts.
Un mouvement Brownien standard est aussi un processus de Lévy et plus généralement, le processus $\gamma t + A^{1/2} W_t$ l'est aussi, avec $\gamma \in \mathbb{R}$ et A une matrice symétrique définie positive. $\tilde{X}_t = \gamma t + A^{1/2} W_t + X_t$ est un processus de Lévy et sa fonction caractéristique est donnée par

$$\mathbb{E}e^{iz\cdot\tilde{X}_t} = \exp\left[-\frac{1}{2} z \cdot Az + i\gamma \cdot z + \int_{\mathbb{R}^d} (e^{iz\cdot x} - 1)\pi(dx)\right], z \in \mathbb{R}^d.$$

3.2 Représentation de Lévy Khintchine

En général, le processus de Lévy est caractérisé par sa fonction caractéristique donnée par la représentation de Lévy Khintchine. Rappelons tout d'abord la notion de loi infiniment divisible.

Définition 3.2. *Une loi de probabilité* F *dans* $\mathbb{R}^d$ *est dite infiniment divisible si pour tout entier* $n \geq 1$*, il existe* n *variables aléatoires i.i.d* $Y_1, ..., Y_n$ *tel que* $Y_1 + ... + Y_n$ *a la loi* F*.*

Si on note par μ la loi de Y_k dans la définition ci-dessus, alors $F = \mu * \mu * .... * \mu$ est la $n^{\text{ième}}$ convolution de μ.

Proposition 3.3. *(voir Proposition 3.1 dans Cont et Tankov [20]). Soit* $(X_t)_{t\geq 0}$ *un processus de Lévy. Alors pour tout* t*,* X_t *a une loi infiniment divisible. Inversement, si* F *est une loi* *infiniment divisible alors il existe un processus de Lévy $(X_t)_{t\geq 0}$ telle que la loi de X_1 est donnée par F.*

Ainsi, d'après le Théorème 8.1 de Sato [60] sur la représentation de Lévy Khintchine pour les lois infiniment divisibles et le Corollaire 8.3 de Sato [60], on obtient le résultat suivant.

Théorème 3.4. *Soit $(X_t)_{t\geq 0}$ un processus de Lévy dans $\mathbb{R}^d$ alors il existe $\gamma \in \mathbb{R}^d$, une matrice A définie positive et une mesure positive ν sur $\mathbb{R}^d$ vérifiant*

$$\nu(\{0\}) = 0 \quad et \quad \int_{\mathbb{R}^d} (1 \wedge |x|^2)\nu(dx) < \infty, \tag{I.18}$$

telle que $\mathbb{E}\left[e^{iz\cdot X_t}\right] = e^{t\psi(z)}, z \in \mathbb{R}^d$, avec

$$\psi(z) = -\frac{1}{2} z \cdot Az + i\gamma \cdot z + \int_{\mathbb{R}^d} (e^{iz\cdot x} - 1 - iz \cdot x \mathbf{1}_{|x|\leq 1})\nu(dx). \tag{I.19}$$

Inversement, pour tout triplet (γ, A, ν) avec $\gamma \in \mathbb{R}^d$, A est une matrice définie positive et ν une mesure positive vérifiant (I.18), il existe un processus de Lévy X_t tel que $\mathbb{E}\left[e^{iz\cdot X_t}\right] = e^{t\psi(z)}, z \in \mathbb{R}^d$, avec $\psi(z)$ est donné par (I.19).

Ainsi Le triplet (γ, A, ν) est appelé le triplet caractéristique du processus de Lévy, ν est appelée la mesure de Lévy et A est la matrice Gaussienne de covariance. La mesure de Lévy porte des informations utiles sur la structure du processus.

Théorème 3.5. *(Voir Théorème 21.3 dans Sato [60]). Soit $(X_t)_{t\geq 0}$ un processus de Lévy avec le triplet caractéristique $(\gamma,\ A,\ \nu)$.*

- *Si $\nu(\mathbb{R}^d) < \infty$ alors p.s. les trajectoires de X ont un nombre fini de sauts sur tout intervalle compact de $\mathbb{R}_+$. Dans ce cas, on dit que le processus de Lévy X a une activité finie.*
- *Si $\nu(\mathbb{R}^d) = \infty$ alors p.s. les trajectoires de X ont un nombre infini de sauts sur tout intervalle compact de $\mathbb{R}_+$. Dans ce cas, on dit que le processus de Lévy X a une activité infinie.*

Théorème 3.6. *(Voir Théorème 21.9 dans Sato [60]). Soit $(X_t)_{t\geq 0}$ un processus de Lévy avec le triplet caractéristique $(\gamma,\ A,\ \nu)$.*

- *Si $A = 0$ et $\int_{|x|\leq 1} |x|\nu(dx) < \infty$ alors p.s. les trajectoires de X sont à variation finie sur $[0, t], t > 0$.*
- *Si $A \neq 0$ ou $\int_{|x|\leq 1} |x|\nu(dx) = \infty$ alors p.s. les trajectoires de X sont à variation infinie sur $[0, t], t > 0$.*

Remarque 3.7. *Soit* $h : \mathbb{R}^d \to \mathbb{R}^d$ *une fonction mesurable vérifiant pour tout* z, $\int_{\mathbb{R}^d}(e^{iz\cdot x} - 1 - iz\cdot h(x))\nu(dx) < \infty$. *Alors, la fonction caractéristique du processus de Lévy* $(X_t)_{t\geq 0}$ *peut s'écrire comme suit*

$$\begin{aligned}\mathbb{E}\left[e^{iz\cdot X_t}\right] &= e^{t\psi(z)}, z \in \mathbb{R}^d,\\ \textit{avec } \psi(z) &= -\frac{1}{2}z\cdot Az + i\gamma_h\cdot z + \int_{\mathbb{R}^d}(e^{iz\cdot x} - 1 - iz\cdot h(x))\nu(dx),\\ \textit{où } \gamma_h &= \gamma + \int_{\mathbb{R}^d}\left(h(x) - x\mathbf{1}_{|x|\leq 1}\right)\nu(dx).\end{aligned} \tag{I.20}$$

Ainsi, le triplet (γ_h, A, ν) *est appelé le triplet caractéristique du processus de Lévy* $(X_t)_{t\geq 0}$ *relativement à la fonction de troncation* h. *Par exemple, si* $\int_{\mathbb{R}^d}|x|\wedge 1\ \nu(dx) < \infty$, *on peut prendre* $h \equiv 0$ *et la représentation de Lévy Khintchine devient*

$$\psi(z) = -\frac{1}{2}z\cdot Az + i\gamma_0\cdot z + \int_{\mathbb{R}^d}(e^{iz\cdot x} - 1)\nu(dx).$$

γ_0 *est appelé drift du processus* X.

3.3 Décomposition de Lévy-Itô

Quand le processus X a une activité infinie ($\nu(\mathbb{R}^d) = +\infty$), on introduira la décomposition de Lévy-Itô comme suit :

Théorème 3.8. *(Voir Théorème 19.2 dans Sato [60]). Soit* $(X_t)_{t\geq 0}$ *un processus de Lévy dans* $\mathbb{R}^d$ *avec le triplet caractéristique* (γ, A, ν). *Soit*

$$X_t^l = \sum_{s\in(0,t]}\Delta X_s\mathbf{1}_{|\Delta X_s|>1} \quad \textit{et} \quad X_t^\varepsilon = \sum_{s\in(0,t]}\Delta X_s\mathbf{1}_{\varepsilon\leq|\Delta X_s|\leq 1} - t\int_{\varepsilon\leq|x|\leq 1}x\nu(dx), \quad t\geq 0,$$

alors

$$X_t = \gamma t + B_t + X_t^l + \lim_{\varepsilon\to 0}X_t^\varepsilon. \tag{I.21}$$

où $(B_t)_{t\geq 0}$ *est un mouvement Brownien d-dimensionnel de matrice de covariance* A *indépendant de* X_t^l *et* $X_t^\varepsilon, 0 < \varepsilon < 1$. *Les deux derniers termes dans (I.21) sont aussi indépendants entre eux et la convergence du dernier terme est presque sûre et uniforme en* $t \in [0, T], T > 0$.

Ainsi, par la décomposition de Lévy-Itô, le processus de Lévy X est représenté par une somme d'une composante Brownienne caractérisée par la matrice A, un processus de Poisson composé et une limite presque sûre d'un processus de Poisson composé compensé (les petits sauts ont été compensés).

3.4 Exemples de processus de Lévy à activité infinie

a) le modèle Variance-Gamma (VG). Le processus Variance-Gamma (VG) est un processus à saut pur d'activité infinie et de variation finie introduit par Madan et Seneta (1990) [50]. Il est obtenu en procédant à un chagement en temps d'un mouvement Brownien par un subordinateur gamma. Un processus VG de paramètres (θ, σ, κ) (σ et θ sont respectivement la volatilité et le drift du mouvement Brownien, κ est la variance du subordinateur) est défini par son exposant charactéristique :

$$\psi(u) = \log\left(\left(1 - i\theta\kappa u + \frac{\sigma^2}{2}\kappa u^2\right)^{\frac{-1}{\kappa}}\right).$$

Sa mesure de Lévy est donnée par

$$\nu(dx) = \frac{c}{\kappa x}e^{-\lambda_+ x}\mathbf{1}_{x>0} + \frac{c}{\kappa|x|}e^{-\lambda_-|x|}\mathbf{1}_{x<0}.$$

avec $c = \frac{1}{\kappa}, \lambda_+ = \frac{\sqrt{\theta^2+\frac{2\sigma^2}{\kappa}}}{\sigma^2} - \frac{\theta}{\sigma^2}$ et $\lambda_- = \frac{\sqrt{\theta^2+\frac{2\sigma^2}{\kappa}}}{\sigma^2} + \frac{\theta}{\sigma^2}$.

b) le modèle NIG (Normal Inverse Gaussian). Le processus NIG est introduit par Barndorff-Nielsen (1997) [9], il est caractérisé par quatre paramètres $\alpha, \beta, \bar{\delta}, \mu$ avec $0 \leq |\beta| \leq \alpha, \bar{\gamma} > 0$ et $\mu \in \mathbb{R}$. Son exposant caractéristique est donné par, pour tout $u \in \mathbb{R}$:

$$\psi(u) = i\mu u + \bar{\delta}\left(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + iu)^2}\right).$$

Le processus NIG peut être représenté comme un mouvement Brownien de drift β et de variance 1 subordonné par un processus IG (Inverse Gaussian) de paramètres $(\bar{\gamma}, \sqrt{\alpha^2 - \beta^2})$.
Sa mesure de Lévy est donnée par

$$\nu(dx) = \frac{\alpha\bar{\delta}}{\pi|x|}K_1(\alpha|x|)e^{\beta x}dx.$$

où K_1 est la fonction de Bessel de troisième type avec $K_1(z) = \frac{1}{2}\int_{\mathbb{R}^+} \exp(-\frac{1}{2}z(y + \frac{1}{y}))dy$.

c) le modèle CGMY. Le processus CGMY est un processus à saut pur d'activité infinie qui a été introduit par Carr, Madan, German, Yor (2002) [16]. Son exposant caractéristique est donné par

$$\psi(x) = C\Gamma(-Y)\left[M^Y\left((1-\frac{iu}{M})^Y - 1 + \frac{iuY}{M}\right) + G^Y\left((1+\frac{iu}{G})^Y - 1 - \frac{iuY}{G}\right)\right] \text{ si } Y \neq 1.$$

où C, G et M sont strictement positifs et $Y < 2$. Si $Y < 1$ alors le processus CGMY est à variation finie et si $Y > 1$, alors il est à variation infinie. Sa mesure de Lévy s'écrit :

$$\nu(dx) = C\frac{e^{-Mx}}{x^{1+Y}}\mathbf{1}_{x>0}dx + C\frac{e^{-G|x|}}{|x|^{1+Y}}\mathbf{1}_{x<0}dx.$$

3.5 Changement de probabilité des mesures de Lévy

Théorème 3.9. *(Voir Théorème 33.1 et 33.2 dans Sato [60]). Soit* $(X, \mathbb{P})$ *et* $(X, \mathbb{Q})$ *deux processus de Lévy dans* $\mathbb{R}^d$ *générés respectivement par les triplets caractéristique* (γ, A, ν) *et* $(\tilde{\gamma}, A, \tilde{\nu})$*. Si les conditions suivantes sont vérifiées*

i) $A = \tilde{A}$,

ii) les mesures ν *et* $\tilde{\nu}$ *sont équivalentes,*

iii) la fonction $\phi : \mathbb{R}^d \to \mathbb{R}$ *définie par* $e^{\phi(x)} = \frac{d\tilde{\nu}}{d\nu}(x)$ *satisfait*

$$\int_{\mathbb{R}^d} (e^{\phi(x)/2} - 1)^2 \nu(dx) < +\infty.$$

iv) les constantes γ *et* $\tilde{\gamma}$ *satisfont* $\tilde{\gamma} = \gamma + \int_{|x|\leq 1} x(\tilde{\nu} - \nu)(dx) + A\eta, \qquad$ *pour* $\eta \in \mathbb{R}^d$,

alors les mesures de probabilité $\mathbb{P}$ *et* $\mathbb{Q}$ *sont équivalentes avec*

$$\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t} = e^{U_t} \quad \mathbb{P} - p.s.,$$

avec U *est un processus défini dans* $\mathbb{R}$ *par,* $\mathbb{P}$*-p.s.,*

$$U_t := \eta \cdot X^c - \frac{t}{2}\eta \cdot A\eta - t\gamma \cdot \eta + \lim_{\varepsilon \to 0}\left(\sum_{s\in(0,t]} \phi(\Delta X_s)\mathbf{1}_{|\Delta X_s|>\varepsilon} - t\int_{|x|>\varepsilon}\left(e^{\phi(x)} - 1\right)\nu(dx)\right), \quad \text{(I.22)}$$

où X^c *désigne la partie martingale continue du processus* X*. La convergence du dernier terme est presque sûre et uniforme en* $t \in [0, T], T > 0$.

Dans le modèle de Black Scholes, une mesure martingale équivalente peut être obtenue en changeant le drift. Dans les modèles avec sauts, si la composante Gaussienne est absente, on ne peut pas changer le drift mais on peut obtenir une plus grande variété de mesures équivalentes en modifiant la distribution des sauts. Pour ce faire, on introduit la transformation d'Esscher, qui est analogue au changement du drift pour le mouvement Brownien géométrique.

La transformation d'Esscher Soit X un processus de Lévy dans $\mathbb{R}^d$ de triplet caractéristique (γ, A, ν) et soit $\theta \in \mathbb{R}^d$ tel que $\int_{|x|>1} e^{\theta \cdot x} \nu(dx) < \infty$. On applique le changement de probabilité du Théorème 3.9 avec $\eta = \theta$ et $\phi(x) = \theta \cdot x$. On obtient alors une probabilité équivalente sous laquelle X est un processus de Lévy caractérisé par la mesure de Lévy $\tilde{\nu}(dx) = e^{\theta \cdot x} \nu(dx)$ et un terme "drift" $\tilde{\gamma} = \gamma + A\theta + \int_{|x|\leq 1} x(e^{\theta \cdot x} - 1)\nu(dx)$. Ainsi, la densité de Radon-Nikodyn correspondante à ce changement de mesure est donnée par

$$\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t} = \frac{e^{\theta \cdot X_t}}{\mathbb{E}[e^{\theta \cdot X_t}]} = e^{\theta \cdot X_t - \kappa(\theta)t},$$

où $\kappa(\theta) = \ln \mathbb{E}[e^{\theta \cdot X_1}]$.

Chapitre II

Importance Sampling and Statistical Romberg

The efficiency of Monte Carlo simulations is significantly improved when implemented with variance reduction methods. Among these methods we focus on the popular importance sampling technique based on producing a parametric transformation through a shift parameter θ. The optimal choice of θ is approximated using Robbins-Monro procedures, provided that a non explosion condition is satisfied. Otherwise, one can use either a constrained Robbins-Monro algorithm (see e.g. Arouna [3] and Lelong [46]) or a more astute procedure based on an unconstrained approach recently introduced by Lemaire and Pagès in [47]. In this article, we develop a new algorithm based on a combination of the statistical Romberg method and the importance sampling technique. The statistical Romberg method introduced by Kebaier in [37] is known for reducing efficiently the complexity compared to the classical Monte Carlo one. In the setting of discretized diffusions, we prove the almost sure convergence of the constrained and unconstrained versions of the Robbins-Monro routine, towards the optimal shift θ^* that minimizes the variance associated to the statistical Romberg method. Then, we prove a central limit theorem for the new algorithm that we called adaptive statistical Romberg method. Finally, we illustrate by numerical simulation the efficiency of our method through applications in option pricing for the Heston model.

1 Introduction

Monte Carlo methods have proved to be a useful tool for many of numerical computations in modern finance. These includes the pricing and hedging of complex financial products. The general problem is to estimate a real quantity $\mathbb{E}\psi(X_T)$, with $T > 0$ and $(X_t)_{0\leq t\leq T}$ is a given diffusion, defined on $\mathcal{B} := (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, taking values in $\mathbb{R}^d$ and ψ a given function such that $\psi(X_T)$ is square integrable. Since the efficiency of the Monte Carlo simulation considerably depends on the smallness of the variance in the estimation, many variance reduction techniques were developed in the recent years. Among these methods appears the technique of importance sampling very popular for its efficiency. The working of this method is quite intuitive, if we can produce a parametric transformation such that for all $\theta \in \mathbb{R}^q$ we have

$$\mathbb{E}\psi(X_T) = \mathbb{E}g(\theta, X_T).$$

Then it is natural, to implement a Monte Carlo procedure using the optimal θ^* solution to the problem

$$\theta^* = \arg\min_{\theta\in\mathbb{R}^q} \mathbb{E}g^2(\theta, X_T),$$

since the quantity $\mathbb{E}g^2(\theta, X_T)$ denotes the main term of the limit variance in the central limit theorem associated to the Monte Carlo method. But how to compute θ^* ? To solve this problem, one can use the so-called Robbins-Monro algorithm to construct recursively a sequence of random variables $(\theta_i)_{i\in\mathbb{N}}$ that approximate accurately θ^*. Convergence results of this procedure requires a quite restrictive condition known as the non explosion condition (see e.g. [13, 24, 41]) given by

$$\text{(NEC)} \qquad \mathbb{E}\left[g^2(\theta, X_T)\right] \leq C(1 + |\theta|^2), \qquad \text{for all } \theta \in \mathbb{R}^q.$$

To avoid this restrictive condition, two improved versions of this routine are proposed in the literature. The first one, based on a truncation procedure called "Projection à la Chen", is introduced by Chen in [17, 18] and investigated later by several authors (see, e.g. Andrieu, Moulines and Priouret in [1] and Lelong in [46]). The use of this procedure in the context of importance sampling is initially proposed by Arouna in [3] and investigated afterward by Lapeyre and Lelong in [43]. The second alternative, is more recent and introduced by Lemaire and Pagès in [47]. In fact, they proposed an unconstrained procedure by using extensively the regularity of the involved density and they prove the convergence of this algorithm. In what follows, these two methods will be called respectively constrained and unconstrained algorithms. In view of this, a Monte Carlo method that integrates this importance sampling recursion is recommended in practice.

The aim of this paper is to study a new algorithm based on an original combination of the statistical Romberg method and the importance sampling technique. The statistical Romberg method is known for improving the Monte Carlo efficiency when used with discretization schemes and was introduced by Kebaier in [37]. However, the main term of the limit variance in the central limit theorem associated to the statistical Romberg method is quite different from that of the crude Monte Carlo method. It turns out that the optimal θ^*, in this case, is solution to the problem

$$\theta^* = \arg\min_{\theta\in\mathbb{R}^q} \tilde{\mathbb{E}}\left(g^2(\theta, X_T) + (\nabla_x g(\theta, X_T)\cdot U_T)^2\right),$$

where $(U_t)_{t\in[0,T]}$ is a given diffusion associated to the process $(X_t)_{t\in[0,T]}$ defined on an extension $\tilde{\mathcal{B}} = (\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t\geq 0}, \tilde{\mathbb{P}})$ of the initial space $\mathcal{B}$ (see further on). Here, for $\theta\in\mathbb{R}^q$ and $x\in\mathbb{R}^d$, $\nabla_x g(\theta, x)$ denotes the gradient of the function g with respect to the second variable at the point (θ, x). Moreover, we intend to study the discretized version of this problem. More precisely, we denote X_T^n (resp. U_T^n) the Euler scheme, with time step T/n, associated to X_T (resp. U_T and we consider the optimal θ_n^* given by

$$\theta_n^* = \arg\min_{\theta\in\mathbb{R}^q} \tilde{\mathbb{E}}\left(g^2(\theta, X_T^n) + (\nabla_x g(\theta, X_T^n)\cdot U_T^n)^2\right).$$

The convergence of θ_n^* towards θ^* as n tends to infinity is proved in the next section. In section 3 we study the problem of estimating θ_n^* using the Robbins-Monro algorithm. More preciously, we construct recursively a sequence of random variables $(\theta_i^n)_{i,n\in\mathbb{N}}$ using either the constrained or the unconstrained procedure. The aim is to prove that

$$\lim_{i,n\to\infty}\theta_i^n = \lim_{i\to\infty}(\lim_{n\to\infty}\theta_i^n) = \lim_{n\to\infty}(\lim_{i\to\infty}\theta_i^n) = \theta^*,\ \tilde{\mathbb{P}}\text{-}a.s.$$

This assertion is slightly complicated to achieve for the unconstrained procedure. In fact, for fixed $i, n\in\mathbb{N}$, the term θ_{i+1}^n constructed with this latter procedure involves $(X_{T,i+1}^{n,(-\theta_i^n)}, U_{T,i+1}^{n,(-\theta_i^n)})$, a new pair of diffusion, with drift terms containing θ_i^n. To overcome this technical difficulty we make use of the θ-sensitivity process given by $(\frac{\partial}{\partial\theta}X_T^{n,(-\theta)}, \frac{\partial}{\partial\theta}U_T^{n,(-\theta)})$ and we obtain the announced convergence result (see Theorem 3.4 and 3.6 and Corollary 3.8). In section 4, we first introduce the new adaptive algorithm obtained by combining together the importance sampling procedure and the statistical Romberg method. Then, we prove central limit theorems for both adaptive Monte Carlo method (see Theorem 4.4 and the remark below), and adaptive statistical Romberg method (see Theorem 4.6) using the Lindeberg-Feller central limit theorem for martingale array. In Section 5 we proceed to numerical simulations to illustrate the efficiency of this new method with some applications in finance. The last section is devoted to discuss some future openings.

2 General Framework

Let $X := (X_t)_{0\leq t\leq T}$ be the process with values in $\mathbb{R}^d$, solution to

$$dX_t = b(X_t)dt + \sum_{j=1}^{q} \sigma_j(X_t)dW_t^j, \quad X_0 = x \in \mathbb{R}^d \tag{II.1}$$

where $W = (W^1, \ldots, W^q)$ is a q-dimensional Brownian motion on some given filtered probability space $\mathcal{B} = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and $(\mathcal{F}_t)_{t\geq 0}$ is the standard Brownian filtration. The functions $b : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ and $\sigma_j : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, $1 \leq j \leq q$, satisfy condition

$$(\mathcal{H}_{b,\sigma}) \quad \forall x, y \in \mathbb{R}^d \quad |b(x) - b(y)| + \sum_{j=1}^{q} |\sigma_j(x) - \sigma_j(y)| \leq C_{b,\sigma}|x - y|, \quad \text{with } C_{b,\sigma} > 0,$$

where for $x \in \mathbb{R}^d$, $|x|^2 = x.x$ stands for the Euclidean norm associated to the inner product "·". We have also $|x|^2 = x^{tr}x$ where x^{tr} denotes the transpose of x. This ensures strong existence and uniqueness of solution of (II.1). In many applications, in particular for the pricing of financial securities, we are interested in the effective computation by Monte Carlo methods of the quantity $\mathbb{E}\psi(X_T)$, where ψ is a given function. From a practical point of view, we have to approximate the process X by a discretization scheme. So, let us consider the Euler continuous approximation X^n with time step $\delta = T/n$ given by

$$dX_t^n = b(X_{\eta_n(t)})dt + \sum_{j=1}^{q} \sigma_j(X_{\eta_n(t)})dW_t, \quad \eta_n(t) = [t/\delta]\delta. \tag{II.2}$$

It is well known that under condition $(\mathcal{H}_{b,\sigma})$ we have the almost sure convergence of X^n towards X together with the following property (see e.g. Bouleau and Lépingle [15])

$$(\mathcal{P}) \;\; \forall p \geq 1, \quad \sup_{0\leq t\leq T} |X_t|, \sup_{0\leq t\leq T} |X_t^n| \in L^p \quad \text{and} \quad \mathbb{E}\left[\sup_{0\leq t\leq T} |X_t - X_t^n|^p\right] \leq \frac{K_p(T)}{n^{p/2}},$$

where $K_p(T)$ is a positive constant depending only on b, σ, T, p and q.

The weak error is firstly studied by Talay and Tubaro in [62] and now it is well known that if ψ, b and $(\sigma_j)_{1\leq j\leq q}$ are in $\mathscr{C}_P^4$, they are four times differentiable and together with their derivatives at most polynomially growing, then we have (see Theorem 14.5.1 in Kloeden and Platen in [39])

$$\varepsilon_n := \mathbb{E}\psi(X_T^n) - \mathbb{E}\psi(X_T) = O(1/n).$$

The same result was extended in Bally and Talay in [7] for a measurable function ψ but with a non degeneracy condition of Hörmander type on the diffusion. In the context of possibly degenerate diffusions, when ψ satisfies $|\psi(x)-\psi(y)| \leq C(1+|x|^p+|y|^p)|x-y|$ for $C>0$, $p \geq 0$, the estimate $|\varepsilon_n| \leq \frac{c}{\sqrt{n}}$ follows easily from $(\mathcal{P})$. Moreover, Kebaier in [37] proved that in addition of assumption $(\mathcal{H}_{b,\sigma})$, if b and $(\sigma_j)_{1\leq j\leq q}$ are $\mathscr{C}^1$ and ψ satisfies condition

$$(\mathcal{H}_\psi) \quad \mathbb{P}(X_T \notin \mathcal{D}_{\dot{\psi}}) = 0, \text{ where } \mathcal{D}_{\dot{\psi}} := \{x \in \mathbb{R}^d \mid \psi \text{ is differentiable at } x\}$$

then, $\lim_{n\to\infty} \sqrt{n}\,\varepsilon_n = 0$. Conversely, under the same assumptions, he shows that the rate of convergence can be $1/n^\alpha$, for any $\alpha \in [1/2, 1]$. So, it is worth to introduce assumption

$$(\mathcal{H}_{\varepsilon_n}) \quad \text{for } \alpha \in [1/2,1] \quad n^\alpha \varepsilon_n \to C_\psi(T,\alpha), \quad C_\psi(T,\alpha) \in \mathbb{R}.$$

In order to compute the quantity $\mathbb{E}\psi(X_T^n)$, one may use the so-called statistical Romberg method, considered by [37] and which is conceptually related to the Talay-Tubaro extrapolation. This method reduces efficiently the computational complexity of the combination of Monte Carlo method and the Euler discretization scheme. In fact, the complexity in the Monte Carlo method is equal to $n^{2\alpha+1}$ and is reduced to $n^{2\alpha+1/2}$ in the statistical Romberg method. More precisely, for two numbers of discretionary time step n and m such that $m << n$, the idea of the statistical Romberg method is to use many sample paths with a coarse time discretization step $\frac{T}{m}$ and few additional sample paths with a fine time discretization step $\frac{T}{n}$. The statistical Romberg routine approximates our initial quantity of interest $\mathbb{E}\psi(X_T)$ using two empirical means

$$\frac{1}{N_1}\sum_{i=1}^{N_1}\psi(\hat{X}_{T,i}^m) + \frac{1}{N_2}\sum_{i=1}^{N_2}\psi(X_{T,i}^n) - \psi(X_{T,i}^m).$$

The random variables of the first empirical mean are independent copies of $\psi(X_T^m)$ and the random variables in the second empirical mean are also independent copies of $\psi(X_T^n) - \psi(X_T^m)$. The associated Brownian paths $\hat{W}$ and W are independent. Under assumptions $(\mathcal{H}_\psi)$ and $(\mathcal{H}_{\varepsilon_n})$, this method is tamed by a central limit theorem with a rate of convergence equal to n^α. More precisely, for $N_1 = n^{2\alpha}$, $N_2 = n^{2\alpha-1/2}$ and $m = \sqrt{n}$ the global error normalized by n^α converges in law to a Gaussian random variable with bias equal to $C_\psi(T,\alpha)$ and a limit variance equal to

$$\operatorname{Var}(\psi(X_T)) + \tilde{\operatorname{V}}\mathrm{ar}(\nabla\psi(X_T)\cdot U_T),$$

where U is the weak limit process of the error $\sqrt{n}(X^n - X)$ defined on $\tilde{\mathcal{B}}$ an extension of the initial space $\mathcal{B}$ (see Theorem 3.2 in Kebaier [37]). More precisely, the process U is solution to

$$dU_t = \dot{b}(X_t)U_t dt + \sum_{j=1}^{q} \dot{\sigma}_j(X_t)U_t dW_t^j - \frac{1}{\sqrt{2}} \sum_{j,\ell=1}^{q} \dot{\sigma}_j(X_t)\sigma_\ell(X_t) d\tilde{W}_t^{\ell j}, \qquad \text{(II.3)}$$

where $\tilde{W}$ is a q^2-dimensional standard Brownian motion, defined on the extension $\tilde{\mathcal{B}}$, independent of W, and $\dot{b}$ (respectively $(\dot{\sigma}_j)_{1\le j\le q}$) is the Jacobian matrix of b (respectively $(\sigma_j)_{1\le j\le q}$).

In view to use importance sampling routine, based on the Girsanov transform, we define the family of $\mathbb{P}_\theta$, as all the equivalent probability measures with respect to $\mathbb{P}$ such that

$$L_t^\theta = \frac{d\mathbb{P}_\theta}{d\mathbb{P}}|_{\mathcal{F}_t} = \exp\left(\theta \cdot W_t - \frac{1}{2}|\theta|^2 t\right).$$

Hence, $B_t^\theta := W_t - \theta t$ is a Brownian motion under $\mathbb{P}_\theta$. This leads to

$$\mathbb{E}\psi(X_T) = \mathbb{E}_\theta \left[\psi(X_T) e^{-\theta \cdot B_T^\theta - \frac{1}{2}|\theta|^2 T}\right].$$

Let us introduce the process X_t^θ solution, under $\mathbb{P}$, to

$$dX_t^\theta = \left(b(X_t^\theta) + \sum_{j=1}^{q} \theta_j \sigma_j(X_t^\theta)\right) dt + \sum_{j=1}^{q} \sigma_j(X_t^\theta) dW_t^j, \qquad \text{(II.4)}$$

so that the process $(B_t^\theta, X_t)_{0\le t\le T}$ under $\mathbb{P}_\theta$ has the same law as $(W_t, X_t^\theta)_{0\le t\le T}$ under $\mathbb{P}$. Henceforth, we get

$$\mathbb{E}\psi(X_T) = \mathbb{E}g(\theta, X_T^\theta, W_T), \qquad \text{with } g(\theta, x, y) = \psi(x) e^{-\theta\cdot y - \frac{1}{2}|\theta|^2 T}, \forall x \in \mathbb{R}^d \text{ and } y \in \mathbb{R}^q. \qquad \text{(II.5)}$$

We also introduce the Euler continuous approximation $X^{n,\theta}$ of the process X^θ solution, under $\mathbb{P}$, to

$$dX_t^{n,\theta} = \left(b(X_{\eta_n(t)}^{n,\theta}) + \sum_{j=1}^{q} \theta_j \sigma_j(X_{\eta_n(t)}^{n,\theta})\right) dt + \sum_{j=1}^{q} \sigma_j(X_{\eta_n(t)}^\theta) dW_t^j.$$

Our target now is to use the statistical Romberg method introduced above to approximate $\mathbb{E}\psi(X_T) = \mathbb{E}g(\theta, X_T^\theta, W_T)$ by

$$\frac{1}{N_1}\sum_{i=1}^{N_1} g(\theta, \hat{X}_{T,i}^{m,\theta}, \hat{W}_{T,i}) + \frac{1}{N_2}\sum_{i=1}^{N_2} g(\theta, X_{T,i}^{n,\theta}, W_{T,i}) - g(\theta, X_{T,i}^{m,\theta}, W_{T,i}).$$

Of course the Brownian paths generated by $\hat{W}$ and W have to be independent. According to Theorem 3.2 of Kebaier [37] mentioned above, we have a central limit theorem with limit variance

$$\mathrm{Var}\left(g(\theta, X_T^\theta, W_T)\right) + \tilde{\mathrm{V}}\mathrm{ar}\left(\nabla_x g(\theta, X_T^\theta, W_T) \cdot U_T^\theta\right)$$

where U^θ is the weak limit process of the error $\sqrt{n}(X^{n,\theta} - X^\theta)$ defined on the extension $\tilde{\mathcal{B}}$ and solution to

$$dU_t^\theta = \left(\dot{b}(X_t^\theta) + \sum_{j=1}^q \theta_j \dot{\sigma}_j(X_t^\theta)\right) U_t^\theta dt + \sum_{j=1}^q \dot{\sigma}_j(X_t^\theta) U_t^\theta dW_t^j - \frac{1}{\sqrt{2}} \sum_{j,\ell=1}^q \dot{\sigma}_j(X_t^\theta)\sigma_\ell(X_t^\theta) d\tilde{W}_t^{\ell j}. \quad \text{(II.6)}$$

Therefore, it is natural to choose the optimal θ^* minimizing the associated variance.
As $\mathbb{E}g(\theta, X_T^\theta, W_T) = \mathbb{E}\psi(X_T)$ and $\tilde{\mathbb{E}}(\nabla_x g(\theta, X_T^\theta, W_T) \cdot U_T^\theta) = 0$ (see Proposition 2.1 in Kebaier [37]), it boils down to choose

$$\theta^* = \underset{\theta\in\mathbb{R}^q}{\operatorname{argmin}}\, v(\theta) \quad \text{with} \quad v(\theta) := \tilde{\mathbb{E}}\left(\left[\psi(X_T^\theta)^2 + (\nabla\psi(X_T^\theta) \cdot U_T^\theta)^2\right] e^{-2\theta.W_T - |\theta|^2 T}\right). \quad \text{(II.7)}$$

Note that from a practical point of view the quantity $v(\theta)$ is not explicit, we use the Euler scheme to discretize (X^θ, U^θ) and we choose the associated

$$\theta_n^* := \underset{\theta\in\mathbb{R}^q}{\operatorname{argmin}}\, v_n(\theta) \quad \text{with} \quad v_n(\theta) := \tilde{\mathbb{E}}\left(\left[\psi(X_T^{n,\theta})^2 + (\nabla\psi(X_T^{n,\theta}) \cdot U_T^{n,\theta})^2\right] e^{-2\theta.W_T - |\theta|^2 T}\right) \quad \text{(II.8)}$$

with $U^{n,\theta}$ is the Euler discretization scheme of U^θ, solution to

$$\begin{aligned} dU_t^{n,\theta} = &\left(\dot{b}(X_{\eta_n(t)}^{n,\theta}) + \sum_{j=1}^q \theta_j \dot{\sigma}_j(X_{\eta_n(t)}^{n,\theta})\right) U_{\eta_n(t)}^{n,\theta} dt \\ &+ \sum_{j=1}^q \dot{\sigma}_j(X_{\eta_n(t)}^{n,\theta}) U_{\eta_n(t)}^{n,\theta} dW_t^j - \frac{1}{\sqrt{2}} \sum_{j,\ell=1}^q \dot{\sigma}_j(X_{\eta_n(t)}^{n,\theta})\sigma_\ell(X_{\eta_n(t)}^{n,\theta}) d\tilde{W}_t^{\ell j}. \quad \text{(II.9)} \end{aligned}$$

Through the whole paper, we require $\mathbb{P}(X_T \notin \mathcal{D}_{\dot{\psi}}) = 0$ and $\mathbb{P}(X_T^n \notin \mathcal{D}_{\dot{\psi}}) = 0$, $n \in \mathbb{N}$, that make (II.7) and (II.8) well posed. Also for an integer $k \geq 1$ and $\delta \in [0, 1]$, we denote by $\mathscr{C}_b^{k,\delta}$ the set of functions g in $\mathscr{C}^k$ with k^{th} order partial derivatives globally δ-Hölder and all partial derivatives up to k^{th} order bounded. In case $\delta = 0$ we simply use the usual notation $\mathscr{C}_b^k$.

The following theorem yields estimates on the L^p convergence of $U^{n,\theta}$ towards U^θ.

Theorem 2.1. *Let $p \geq 1$ and $\theta \in \mathbb{R}^q$. If σ and b are in $\mathscr{C}_b^1$, then both processes $\sup_{0\leq t\leq T} |U_t^\theta|$ and $\sup_{0\leq t\leq T} |U_t^{n,\theta}|$ are in L^p. Moreover, if σ and b are in $\mathscr{C}_b^{1,1}$ then we have the almost sure* *convergence of $U^{n,\theta}$ towards U^{θ} together with the following property*

$$(\tilde{\mathcal{P}})\ \ \forall p\geq 1,\quad \sup_{0\leq t\leq T}|U_t^{\theta}|,\ \sup_{0\leq t\leq T}|U_t^{n,\theta}|\in L^p\quad \textit{and}\quad \tilde{\mathbb{E}}\left[\sup_{0\leq t\leq T}|U_t^{\theta}-U_t^{n,\theta}|^p\right]\leq \frac{K_p(T)}{n^{p/2}},$$

where $K_p(T)$ is a positive constant depending on b, σ, θ, T, p and q. Consequently, the above results still hold for the processes U and U^n by taking $\theta=0$.

Proof. Let $\theta\in\mathbb{R}^q$. First, for U^{θ} given by (II.6) we have to prove that $\sup_{0\leq t\leq T}|U_t^{\theta}|$ is in L^p. Using the integral form of the process, if F_t^1, $t\in[0,T]$, denotes the associated first term on the right-hand side of (II.6), then by the Hölder inequality and the boundedness of $\dot{b}$, $\dot{\sigma}_j$, $\{1,\cdots,q\}$, there is $c_1>0$ such that

$$\mathbb{E}\sup_{0\leq s\leq t}|F_s^1|^p\leq \mathbb{E}\Big(\int_0^t\Big(|\dot{b}(X_s^{\theta})|+\sum_{j=1}^q|\theta_j||\dot{\sigma}_j(X_s^{\theta})|\Big)|U_s^{\theta}|\Big)^p ds\leq c_1\int_0^t\mathbb{E}|U_s^{\theta}|^p ds. \tag{II.10}$$

If F_t^2, $t\in[0,T]$, denotes the second term on the right-hand side of (II.6), then by Burkholder-Davis-Gundy's inequality there exists a constant $C_p>0$ depending on p such that

$$\mathbb{E}\sup_{0\leq s\leq t}|F_s^2|^p\leq q^{p-1}\sum_{j=1}^q\mathbb{E}\sup_{0\leq s\leq t}\Big|\int_0^s\dot{\sigma}_j(X_v^{\theta})U_v^{\theta}dW_v^j\Big|^p\leq q^{p-1}C_p\sum_{j=1}^q\mathbb{E}\Big(\int_0^t|\dot{\sigma}_j(X_s^{\theta})|^2|U_s^{\theta}|^2ds\Big)^{p/2}.$$

Thanks to the Hölder inequality and the boundedness of $\dot{\sigma}$, there is a constant $c_2>0$ such that

$$\mathbb{E}\sup_{0\leq s\leq t}|F_s^2|^p\leq c_2\int_0^t\mathbb{E}|U_s^{\theta}|^p ds. \tag{II.11}$$

Now, if F_t^3, $t\in[0,T]$, denotes the third term on the right-hand side of (II.6) then using the same arguments as above together with the linear growth assumption on σ and property $(\mathcal{P})$ we get the existence of $c_3>0$ such that

$$\mathbb{E}\sup_{0\leq s\leq t}|F_s^3|^p\leq \frac{q^{2p-2}}{2^{p/2}}\sum_{j,\ell=1}^q\mathbb{E}\Big(\int_0^t|\dot{\sigma}_j(X_s^{\theta})|^2|\sigma_\ell(X_s^{\theta})|^2ds\Big)^{p/2}\leq c_3. \tag{II.12}$$

So (II.10), (II.11), (II.12), and the inequality $(a+b+c)^p\leq 3^{p-1}(a^p+b^p+c^p)$ tell us that there exists A and B depending on b, σ, θ, p, q and T such that

$$\mathbb{E}\sup_{0\leq s\leq t}|U_s^{\theta}|^p\leq A+B\int_0^t\mathbb{E}\sup_{0\leq v\leq s}|U_v^{\theta}|^p ds.$$

Hence Gronwall's lemma yields $\mathbb{E}\sup_{0\leq s\leq t}|U_s^{\theta}|^p\leq Ae^{Bt}$ for all $t\in[0,T]$ (see e.g. [15] page 269).

Now, the same proof holds for $U^{n,\theta}$, where the constants obtained in the corresponding upper bound do not depend on the parameter n. Hence, we obtain the first assertion of the theorem namely $\sup_{0\le s\le T}|U^\theta_t|$ and $\sup_{0\le s\le T}|U^{n,\theta}_t|$ are in L^p, $p\ge 1$.
We now proceed to control the quantity $\mathbb{E}\sup_{0\le s\le t}|U^\theta_s-U^{n,\theta}_s|^p$ and we write

$$U^\theta_t-U^{n,\theta}_t=G^1_t+G^2_t+G^3_t,\qquad \text{for all } t\in[0,T],$$

with G^1 is the drift term, G^2 is the sum of the stochastic integrals terms with respect to the Brownian motion W and G^3 is the sum of the stochastic integrals terms with respect to the Brownian motion $\tilde{W}$. Concerning the first term G^1, we write it as follows

$$\begin{aligned}G^1_t=&\int_0^t\Big(\dot b(X^\theta_s)+\sum_{j=1}^q\theta_j\dot\sigma_j(X^\theta_s)\Big)(U^\theta_s-U^{n,\theta}_s)ds+\int_0^t\Big(\dot b(X^\theta_s)+\sum_{j=1}^q\theta_j\dot\sigma_j(X^\theta_s)\Big)(U^{n,\theta}_s-U^{n,\theta}_{\eta_n(s)})ds\\&+\int_0^t\Big(\dot b(X^\theta_s)-\dot b(X^{n,\theta}_{\eta_n(s)})+\sum_{j=1}^q\theta_j\big(\dot\sigma_j(X^\theta_s)-\dot\sigma_j(X^{n,\theta}_{\eta_n(s)})\big)\Big)U^{n,\theta}_{\eta_n(s)}dt.\end{aligned}\tag{II.13}$$

If $G^{1,1}_t$, $t\in[0,T]$, denotes the first term on the right-hand side of (II.13), then by the Hölder inequality and the boundedness of $\dot b$, $\dot\sigma$, there is a constant $c_4>0$ such that

$$\mathbb{E}\sup_{0\le s\le t}|G^{1,1}_s|^p\le c_4\int_0^t\mathbb{E}\sup_{0\le v\le s}|U^\theta_v-U^{n,\theta}_v|^pds.\tag{II.14}$$

If $G^{1,2}_t$, $t\in[0,T]$, denotes the second term on the right-hand side of (II.13), then by the same arguments, there is a constant $c_5>0$ such that

$$\mathbb{E}\sup_{0\le s\le t}|G^{1,2}_s|^p\le c_5\int_0^T\mathbb{E}|U^\theta_t-U^{n,\theta}_{\eta_n(t)}|^pdt.$$

Noticing that

$$\begin{aligned}U^{n,\theta}_t-U^{n,\theta}_{\eta_n(t)}=&\Big(\dot b(X^{n,\theta}_{\eta_n(t)})+\sum_{j=1}^q\theta_j\dot\sigma_j(X^{n,\theta}_{\eta_n(t)})\Big)U^{n,\theta}_{\eta_n(t)}(t-\eta_n(t))\\&+\sum_{j=1}^q\dot\sigma_j(X^{n,\theta}_{\eta_n(t)})U^{n,\theta}_{\eta_n(t)}(W^j_t-W^j_{\eta_n(t)})-\frac{1}{\sqrt{2}}\sum_{j,\ell=1}^q\dot\sigma_j(X^{n,\theta}_{\eta_n(t)})\sigma_\ell(X^{n,\theta}_{\eta_n(t)})(\tilde W^{\ell j}_t-\tilde W^{\ell j}_{\eta_n(t)}),\end{aligned}$$

we get thanks to the Cauchy-Schwarz inequality and the boundedness of $\dot b$, $\dot\sigma$, the existence of a constant $c_6 > 0$ such that

$$\mathbb{E}|U_t^{n,\theta} - U_{\eta_n(t)}^{n,\theta}|^p \leq c_6\Big(\mathbb{E}|U_{\eta_n(t)}^{n,\theta}|^p(t-\eta_n(t))^p$$
$$+\sum_{j=1}^{q}\Big(\mathbb{E}|U_{\eta_n(t)}^{n,\theta}|^{2p}\Big)^{\frac{1}{2}}\Big(\mathbb{E}|W_t^j - W_{\eta_n(t)}^j|^{2p}\Big)^{\frac{1}{2}} + \sum_{j,\ell=1}^{q}\Big(\mathbb{E}|\sigma_\ell(X_{\eta_n(t)}^{n,\theta})|^{2p}\Big)^{\frac{1}{2}}\Big(\mathbb{E}|\tilde{W}_t^{\ell j} - \tilde{W}_{\eta_n(t)}^{\ell j}|^{2p}\Big)^{\frac{1}{2}}\Big).$$

Since $\mathbb{E}|W_t^j - W_{\eta_n(t)}^j|^{2p} = \mathbb{E}|\tilde{W}_t^{\ell j} - \tilde{W}_{\eta_n(t)}^{\ell j}|^{2p} = (t-\eta_n(t))^p\frac{2p!}{2^p(p)!}$ and $\sup_{0\leq s\leq T}|X_t^{n,\theta}|$ and $\sup_{0\leq s\leq T}|U_t^{n,\theta}|$ are in L^{2p} , we use the linear growth of σ to deduce the existence of a constant $c_7 > 0$ such that

$$\mathbb{E}\sup_{0\leq s\leq t}|G_s^{1,2}|^p \leq \frac{c_7}{n^{p/2}}. \tag{II.15}$$

If $G_t^{1,3}$, $t \in [0,T]$, denotes the third term on the right-hand side of (II.13), then using the Lipschitz property on $\dot{b}$, $\dot{\sigma}$, and the Cauchy-Schwarz inequality, we deduce the existence of a constant $c_8 > 0$ such that

$$\mathbb{E}\sup_{0\leq s\leq t}|G_s^{1,3}|^p \leq c_8\Big(\mathbb{E}\sup_{0\leq t\leq T}|X_t^\theta - X_{\eta_n(t)}^{n,\theta}|^{2p}\Big)^{\frac{1}{2}}\Big(\mathbb{E}\sup_{0\leq t\leq T}|U_{\eta_n(t)}^{n,\theta}|^{2p}\Big)^{\frac{1}{2}}.$$

Now using property $(\mathcal{P})$, the proposition in page 274 of [15] and $\sup_{0\leq s\leq T}|U_t^{n,\theta}| \in L^{2p}$, we deduce the existence of a constant $c_9 > 0$ such that

$$\mathbb{E}\sup_{0\leq s\leq t}|G_s^{1,3}|^p \leq \frac{c_9}{n^{p/2}}. \tag{II.16}$$

So (II.14), (II.15), (II.16) tell us that there exists A_1 and B_1 depending on b, σ, θ, p, q and T such that

$$\mathbb{E}\sup_{0\leq s\leq t}|G_s^1|^p \leq \frac{A_1}{n^{p/2}} + B_1\int_0^t \mathbb{E}\sup_{0\leq v\leq s}|U_v^\theta - U_v^{n,\theta}|^p ds. \tag{II.17}$$

Concerning the second term G^2, using Burkholder-Davis-Gundy's inequality followed by the Hölder's one we get the existence of a constant $c_{10} > 0$ such that

$$\mathbb{E}\sup_{0\leq s\leq t}|G_s^2|^p \leq c_{10}\sum_{j=1}^{q}\int_0^t \mathbb{E}\Big|\dot{\sigma}_j(X_s^\theta)U_s^\theta - \dot{\sigma}_j(X_{\eta_n(s)}^{n,\theta})U_{\eta_n(s)}^{n,\theta}\Big|^p ds.$$

The expectation term inside the above integral is bounded up to a multiplicative constant by

$$\mathbb{E}\Big|\dot{\sigma}_j(X_s^\theta)(U_s^\theta - U_s^{n,\theta})\Big|^p + \mathbb{E}\Big|\dot{\sigma}_j(X_s^\theta)(U_s^{n,\theta} - U_{\eta_n(s)}^{n,\theta})\Big|^p + \mathbb{E}\Big|\Big(\dot{\sigma}_j(X_s^\theta) - \dot{\sigma}_j(X_{\eta_n(s)}^{n,\theta})\Big)U_{\eta_n(s)}^{n,\theta}\Big|^p.$$

The same evaluations used to get relation (II.17) allow us to handle separately the three terms above. Hence, we deduce in the same manner that there exists A_2 and B_2 depending on b, σ, θ, p, q and T such that

$$\mathbb{E}\sup_{0\leq s\leq t}|G_s^2|^p \leq \frac{A_2}{n^{p/2}} + B_2\int_0^t \mathbb{E}\sup_{0\leq v\leq s}|U_v^\theta - U_v^{n,\theta}|^p ds. \tag{II.18}$$

Concerning the third term G^3, we apply the same arguments again to get the existence of a constant $c_{11} > 0$ such that

$$\mathbb{E}\sup_{0\leq s\leq t}|G_s^3|^p \leq c_{11}\sum_{j,\ell=1}^{q}\mathbb{E}\sup_{0\leq s\leq t}\left|\dot{\sigma}_j(X_s^\theta)\sigma_\ell(X_s^\theta) - \dot{\sigma}_j(X_{\eta_n(s)}^{n,\theta})\sigma_\ell(X_{\eta_n(s)}^{n,\theta})\right|^p.$$

It follows that the expectation term in the above sum is bounded by

$$\mathbb{E}\sup_{0\leq s\leq t}\left|\dot{\sigma}_j(X_s^\theta)\sigma_\ell(X_s^\theta) - \dot{\sigma}_j(X_{\eta_n(s)}^{\theta})\sigma_\ell(X_{\eta_n(s)}^{\theta})\right|^p + \mathbb{E}\sup_{0\leq s\leq t}\left|\dot{\sigma}_j(X_{\eta_n(s)}^\theta)\sigma_\ell(X_{\eta_n(s)}^\theta) - \dot{\sigma}_j(X_{\eta_n(s)}^{n,\theta})\sigma_\ell(X_{\eta_n(s)}^{n,\theta})\right|^p.$$

Since σ is a Lipschitz continuous function with linear growth and $\dot{\sigma}$ is a Lipschitz continuous bounded function, we use again the proposition in page 274 of [15] (respectively property $(\mathcal{P})$) to get a control on the first term (respectively on the second term) of the above expression. Hence, there exists a positive constant A_3 depending on σ, θ, p, q and T such that

$$\mathbb{E}\sup_{0\leq s\leq t}|G_s^3|^p \leq \frac{A_3}{n^{p/2}}. \tag{II.19}$$

Finally putting together relations (II.17), (II.18) and (II.19), we complete the proof by using the Gronwall lemma. □

The existence and uniqueness of θ^* is ensured by the following result.

Proposition 2.2. *Suppose σ and b are in $\mathscr{C}_b^1$ and let ψ satisfying $\mathbb{P}(\psi(X_T)\neq 0) > 0$. If there exists $a > 1$ such that $\mathbb{E}\left[\psi^{2a}(X_T)\right]$ and $\mathbb{E}\left[|\nabla\psi(X_T)|^{2a}\right]$ are finite, then the function $\theta\mapsto v(\theta)$ is $\mathscr{C}^2$ and strictly convex with $\nabla v(\theta) = \tilde{\mathbb{E}}H(\theta, X_T, U_T, W_T)$ where*

$$H(\theta, X_T, U_T, W_T) := (\theta T - W_T)\left[\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2\right]e^{-\theta\cdot W_T + \frac{1}{2}|\theta|^2 T}. \tag{II.20}$$

Moreover, there exists a unique $\theta^*\in\mathbb{R}^q$ such that $min_{\theta\in\mathbb{R}^q}v(\theta) = v(\theta^*)$.

Proof. First of all, note that according to Girsanov theorem, the process (B^θ, X, U) under

$\tilde{\mathbb{P}}_\theta$ has the same law as (W, X^θ, U^θ) under $\tilde{\mathbb{P}}$. So, using a change of probability, we get

$$v(\theta) := \tilde{\mathbb{E}}\left(\left[\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2\right] e^{-\theta.W_T+\frac{1}{2}|\theta|^2T}\right).$$

The function $\theta \mapsto [\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2]\, e^{-\theta.W_T+\frac{1}{2}|\theta|^2T}$ is infinitely continuously differentiable with a first derivative equal to $H(\theta, X_T, U_T, W_T)$. Note that, for $c > 0$ we have

$$\sup_{|\theta|\leq c} |H(\theta, X_T, U_T, W_T)| \leq (cT + |W_T|)\left[\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2\right] e^{c|W_T|+\frac{1}{2}c^2T}.$$

Using Hölder's inequality, it is easy to check that $\tilde{\mathbb{E}}\sup_{|\theta|\leq c}|H(\theta, X_T, U_T, W_T)|$ is bounded by

$$e^{\frac{1}{2}c^2T}\left(\|\psi^2(X_T)\|_a \|e^{c|W_T|}(cT + |W_T|)\|_{\frac{a}{a-1}} + \||\nabla\psi(X_T)|^2\|_a \||U_T|^2\|_{\frac{2a}{a-1}} \|e^{c|W_T|}(cT + |W_T|)\|_{\frac{2a}{a-1}}\right).$$

Since $\mathbb{E}\psi^{2a}(X_T)$ and $\mathbb{E}|\nabla\psi(X_T)|^{2a}$ are finite we conclude, thanks to the first assertion in the above Theorem 2.1, the boundedness of $\tilde{\mathbb{E}}\sup_{|\theta|\leq c}|H(\theta, X_T, U_T, W_T)|$. According to Lebesgue's theorem we deduce that v is $\mathscr{C}^1$ in $\mathbb{R}^q$ and $\nabla v(\theta) = \tilde{\mathbb{E}}H(\theta, X_T, U_T, W_T)$. In the same way, we prove that v is of class $\mathscr{C}^2$ in $\mathbb{R}^q$. So, we have

$$\mathrm{Hess}(v(\theta)) = \tilde{\mathbb{E}}\left[\left((\theta T - W_T)(\theta T - W_T)^{tr} + TI_q\right)(\psi^2(X_T) + (\nabla\psi(X_T)\cdot U_T)^2)e^{-\theta.W_T+\frac{1}{2}|\theta|^2T}\right].$$

Since $\mathbb{P}(\psi(X_T) \neq 0) > 0$, we get for all $u \in \mathbb{R}^q\backslash\{0\}$

$$u^{tr}\ \mathrm{Hess}(v(\theta))\ u = \tilde{\mathbb{E}}\left[T|u|^2 + (u.(\theta T - W_T))^2(\psi^2(X_T) + (\nabla\psi(X_T)\cdot U_T)^2)\ e^{-\theta.W_T+\frac{1}{2}|\theta|^2T}\right] > 0.$$

Hence, v is strictly convex. Consequently, to prove that the unique minimum is attained for a finite value of θ, it will be sufficient to prove that $\lim_{|\theta|\to\infty} v(\theta) = +\infty$. Recall that $v(\theta) = \tilde{\mathbb{E}}\left[(\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2)e^{-\theta.W_T+\frac{1}{2}|\theta|^2T}\right]$. Using Fatou's lemma, we get

$$\begin{aligned}+\infty = \tilde{\mathbb{E}}&\left[\liminf_{|\theta|\to\infty}(\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2)e^{-\theta.W_T+\frac{1}{2}|\theta|^2T}\right] \\ &\leq \liminf_{|\theta|\to+\infty}\tilde{\mathbb{E}}\left[(\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2)e^{-\theta.W_T+\frac{1}{2}|\theta|^2T}\right].\end{aligned}$$

This completes the proof. □

The same results can be obtained for the Euler scheme X^n.

Proposition 2.3. *Suppose σ and b are in $\mathscr{C}_b^1$. Given $n \in \mathbb{N}$, let ψ satisfying $\mathbb{P}(\psi(X_T^n) \neq 0) > 0$.*

If there exists $a > 1$ such that $\mathbb{E}\left[\psi^{2a}(X_T^n)\right]$ and $\mathbb{E}\left[|\nabla\psi(X_T^n)|^{2a}\right]$ are finite, then the function $\theta \mapsto v_n(\theta)$ is $\mathscr{C}^2$ and strictly convex with $\nabla v_n(\theta) = \tilde{\mathbb{E}} H(\theta, X_T^n, U_T^n, W_T)$. Moreover, there exists a unique $\theta_n^* \in \mathbb{R}^q$ such that $min_{\theta\in\mathbb{R}^q} v_n(\theta) = v_n(\theta_n^*)$.

Further, we prove the convergence of θ_n^* towards θ^* as n tends to infinity.

Theorem 2.4. *Suppose σ and b are in $\mathscr{C}_b^{1,1}$. Let ψ satisfying $\mathbb{P}(\psi(X_T) \neq 0) > 0$ and $\mathbb{P}(\psi(X_T^n) \neq 0) > 0$ for all $n \in \mathbb{N}$. If there exists $a > 1$ such that $\mathbb{E}\left[\psi^{2a}(X_T)\right]$, $\sup_{n\in\mathbb{N}} \mathbb{E}\left[\psi^{2a}(X_T^n)\right]$, $\mathbb{E}\left[|\nabla\psi(X_T)|^{2a}\right]$ and $\sup_{n\in\mathbb{N}} \mathbb{E}\left[|\nabla\psi(X_T^n)|^{2a}\right]$ are finite. Then,*

$$\theta_n^* \longrightarrow \theta^*, \qquad \text{as } n \to \infty.$$

Proof. First of all, we will prove that $(\theta_n^*)_{n\in\mathbb{N}}$ is a $\mathbb{R}^q$-bounded sequence. By way of contradiction, let us suppose that there is a subsequence $(\theta_{n_k}^*)_{k\in\mathbb{N}}$ that diverges to infinity, $\lim_{k\to\infty} |\theta_{n_k}^*| = +\infty$. This implies that on the event $\{\psi(X_T) \neq 0\}$ we have

$$\lim_{k\to\infty} \Big(\psi^2(X_T^{n_k}) + (\nabla\psi(X_T^{n_k}).U_T^n)^2\Big) e^{-\theta_{n_k}^* W_T + \frac{1}{2}|\theta_{n_k}^*|^2 T} = +\infty.$$

So, by Fatou's lemma we get $\lim_{k\to\infty} v_{n_k}(\theta_{n_k}^*) = +\infty$ while

$$v_{n_k}(\theta_{n_k}^*) \le v_{n_k}(0) \le \sup_{n\in\mathbb{N}} \mathbb{E}\left[\psi^2(X_T^n)\right] < \infty.$$

This leads to a contradiction and we deduce that there is some $M > 0$ such that $|\theta_n^*| \le M$ for all $n \in \mathbb{N}$. Now, it remains to prove that the set $S = \{x \in \mathbb{R}^q \,:\, \theta_{n_k}^* \to x \text{ for some subsequence } \theta_{n_k}^*\}$ is reduced to the singleton set $\{\theta^*\}$. Let us consider a subsequence $\theta_{n_k}^* \to \theta_\infty^* \in S$ as k tends to infinity. According to Proposition 2.2 above, we have

$$\nabla v_{n_k}(\theta_{n_k}^*) = \tilde{\mathbb{E}}\left[(\theta_{n_k}^* T - W_T)\Big(\psi^2(X_T^{n_k}) + (\nabla\psi(X_T^{n_k}).U_T^{n_k})^2\Big) e^{-\theta_{n_k}^*.W_T + \frac{1}{2}|\theta_{n_k}^*|^2 T}\right] = 0.$$

Now, let $1 < \tilde{a} < a$, using the relation $|x+y|^{\tilde{a}} \le 2^{\tilde{a}-1}(|x|^{\tilde{a}} + |y|^{\tilde{a}})$ and applying Hölder's inequality twice with the boundedness of $\theta_{n_k}^*$ established in the first part of the proof we check easily that there exists $c_1 > 0$ depending on a, T and M such that

$$\tilde{\mathbb{E}}\left[\left|(\theta_{n_k}^* T - W_T)\Big(\psi^2(X_T^{n_k}) + (\nabla\psi(X_T^{n_k}).U_T^{n_k})^2\Big) e^{-\theta_{n_k}^*.W_T + \frac{1}{2}|\theta_{n_k}^*|^2 T}\right|^{\tilde{a}}\right] \le \\ c_1 \left\{ \|\psi^2(X_T^{n_k})\|_a^{\tilde{a}} + \||\nabla\psi(X_T^{n_k})|^2\|_a^{\tilde{a}} \||U_T^{n_k}|^{2\tilde{a}}\|_{\frac{2a}{a-\tilde{a}}} \right\}.$$

Thanks to our assumptions $\sup_{n\in\mathbb{N}} \mathbb{E}\left[\psi^{2a}(X_T^n)\right] < \infty$ and $\sup_{n\in\mathbb{N}} \mathbb{E}\left[|\nabla\psi(X_T^n)|^{2a}\right] < \infty$ and

Theorem 2.1, we get the uniform integrability. Therefore, using the almost sure convergence of $\psi^2(X_T^n)$, $\nabla\psi(X_T^n)$ and U_T^n respectively towards $\psi^2(X_T)$, $\nabla\psi(X_T)$ and U_T which is ensured by $(\mathcal{P})$, $(\tilde{\mathcal{P}})$ and $\mathbb{P}(X_T \notin \mathcal{D}_{\dot{\psi}}) = 0$. So, we obtain

$$\nabla v(\theta_\infty^*) = \tilde{\mathbb{E}}\left[(\theta_\infty^* T - W_T)\Big(\psi^2(X_T) + (\nabla\psi(X_T).U_T)^2\Big)e^{-\theta_\infty^*.W_T + \frac{1}{2}|\theta_\infty^*|^2 T}\right] = 0.$$

We complete the proof using the uniqueness of the minimum ensured by Proposition 2.2. □

3 Robbins-Monro Algorithms

The aim now is to construct for fixed n some sequences $(\theta_i^n)_{i\in\mathbb{N}}$ such that $\lim_{i\to\infty}\theta_i^n = \theta_n^*$ almost surely. It is well known that stochastic algorithms can be used to answer this issue and find an accurate approximation of $\theta_n^* = \arg\min_{\theta\in\mathbb{R}} v_n(\theta)$. Indeed, using the Robbins-Monro algorithm, we construct recursively the sequence of random variables $(\theta_i^n)_{i\in\mathbb{N}}$ in $\mathbb{R}^q$ given by

$$\theta_i^n = \theta_{i-1}^n - \gamma_i H(\theta_{i-1}^n, X_{T,i}^n, U_{T,i}^n, W_{T,i}),\ i \geq 1,\ \theta_0^n \in \mathbb{R}^q, \tag{II.21}$$

where H is given by relation (II.20), the gain sequence $(\gamma_i)_{i\geq 1}$ is a decreasing sequence of positive real numbers satisfying

$$\sum_{i=1}^{\infty}\gamma_i = \infty \quad \text{and} \quad \sum_{i=1}^{\infty}\gamma_i^2 < \infty. \tag{II.22}$$

Here $(X_{T,i}^n, U_{T,i}^n)_{i\geq 1}$ is a sequence of independent copies of the Euler scheme associated to (X_T^n, U_T^n) adapted to the filtration $\tilde{\mathcal{F}}_{T,i} = \sigma(W_{t,l}, \tilde{W}_{t,l}, l \leq i, t \leq T)$, where $(W_i, \tilde{W}_i)_{i\geq 1}$ are independent copies of the pair $(W, \tilde{W})$ introduced before in equation (II.3). To obtain the almost sure convergence of the above algorithm to $\theta_n^* = \arg\min_{\theta\in\mathbb{R}} v_n(\theta)$, we need to check a first condition : $\forall\theta \neq \theta_n^*$, $\langle\nabla v_n(\theta), \theta - \theta_n^*\rangle > 0$, which is satisfied in our context thanks to the convexity property of v_n. Secondly we need also a sub-quadratic assumption known as the non explosion condition

(NEC) $$\tilde{\mathbb{E}}\left[|H(\theta, X_T^n, U_T^n, W_T)|^2\right] \leq C(1 + |\theta|^2), \qquad \text{for all } \theta \in \mathbb{R}^q.$$

Unfortunately, this condition is not satisfied in our context and we will study two different stochastic algorithms using the Robbins-Monro procedure and avoiding the above restriction.

3.1 Constrained stochastic algorithm

The idea of the "Projection à la Chen" is to kill the classic Robbins-Monro procedure when it goes close to explosion and to restart it with a smaller step sequence. This can be described as some repeated truncations when the algorithm leaves a slowly growing compact set waiting for stabilization. Then, the algorithm behaves like the Robbins-Monro algorithm. Formally, for a fixed number of discretization time step $n \geq 1$, the repeated truncations can be written in our context as follows. Let $(\mathcal{K}_i)_{i\in\mathbb{N}}$ denote an increasing sequence of compact sets satisfying $\cup_{i=0}^{\infty} \mathcal{K}_i = \mathbb{R}^d$ and $\mathcal{K}_i \subsetneq \overset{\circ}{\mathcal{K}}_{i+1}, \forall i \in \mathbb{N}$. For $\theta_0^n \in \mathcal{K}_0$, $\alpha_0^n = 0$ and a gain sequence $(\gamma_i)_{i\in\mathbb{N}}$ satisfying (II.22), we define the sequence $(\theta_i^n, \alpha_i^n)_{i\in\mathbb{N}}$ recursively by

$$\begin{cases} \textbf{if} & \theta_{i-1}^n - \gamma_i H(\theta_{i-1}^n, X_{T,i}^n, U_{T,i}^n, W_{T,i}) \in \mathcal{K}_{\alpha_{i-1}^n}, \ \textbf{then} \\ & \theta_i^n = \theta_{i-1}^n - \gamma_i H(\theta_{i-1}^n, X_{T,i}^n, U_{T,i}^n, W_{T,i}), \text{ and } \alpha_i^n = \alpha_{i-1}^n \\ \textbf{else} & \theta_i^n = \theta_0^n \text{ and } \alpha_i^n = \alpha_{i-1}^n + 1, \end{cases} \tag{II.23}$$

where the function H is given above in relation (II.20). For $i \in \mathbb{N}$, α_i^n represents the number of truncations of the first i iterations. In fact, as we can see, if the $(i)^{th}$ iteration of the Robbins-Monro is in the compact set $\mathcal{K}_{\alpha_{i-1}^n}$, then the algorithm will behave like a regular Robbins-Monro. However, if the $(i)^{th}$ iteration outside the compact set $\mathcal{K}_{\alpha_{i-1}^n}$, it will be reinitialized. Then, we increase the domain of projection, so we consider the new compact set $\mathcal{K}_{\alpha_{i-1}^n+1}$.

Theorem 3.1. *Suppose σ and b are in $\mathcal{C}_b^1$. Assume that for all $n \in \mathbb{N}$, $\mathbb{P}(\psi(X_T^n) \neq 0) > 0$ and there exists $a > 1$ such that $\mathbb{E}[\psi^{4a}(X_T^n)]$ and $\mathbb{E}[|\nabla\psi(X_T^n)|^{4a}]$ are finite, then the sequence $(\theta_i^n)_{i\geq 0}$ given by routine (II.23), satisfies*

1. *For all $n \in \mathbb{N}$, we have $\theta_i^n \underset{i\to\infty}{\longrightarrow} \theta_n^*$, almost surely where θ_n^* is given by relation (II.8).*
2. *Reversely, for all $i \in \mathbb{N}$, we have $\theta_i^n \underset{n\to\infty}{\longrightarrow} \theta_i$, almost surely, where the sequence $(\theta_i)_{i\geq 0}$ is obtained by replacing in routine (II.23), $(X_{T,i}^n, U_{T,i}^n)$ by their limit $(X_{T,i}, U_{T,i})$, $i \geq 1$.*

Proof. At the beginning, note that for $n \in \mathbb{N}$ the existence of θ_n^* is ensured by Proposition 2.2. Concerning, the first assertion, we have to check both assumptions of Theorem 3.1 in [43]. The first one given by

$$\forall \theta \neq \theta_n^*, \quad \langle \nabla v_n(\theta), \theta - \theta_n^* \rangle > 0,$$

is satisfied in our context thanks to the convexity property of v_n. So, it remains to check the second assumption given by

$$\forall c > 0, \quad \sup_{|\theta|\leq c} \tilde{\mathbb{E}}\left[|H(\theta, X_T^n, U_T^n, W_T)|^2 \right] < \infty. \tag{II.24}$$

This assumption relaxes the usual (NEC) condition on function H used to run the Robbins-Monro algorithm. Let $c > 0$, we have

$$\sup_{|\theta|\leq c} |H(\theta, X_T^n, U_T^n, W_T)|^2 \leq 2(cT + |W_T|)^2 \left[\psi(X_T^n)^4 + (\nabla\psi(X_T^n)\cdot U_T^n)^4\right] e^{2c|W_T|+c^2T}.$$

Using several times Hölder's inequality together with property $(\tilde{\mathcal{P}})$, it is easy to check assumption (II.24), since $\mathbb{E}\psi^{4a}(X_T^n)$ and $\mathbb{E}|\nabla\psi(X_T^n)|^{4a}$ are finite.

The second assertion follows easily by induction on (θ_i^n, α_i^n), using that for all $i \geq 1$, the pair $(X_{T,i}^n, U_{T,i}^n)$ converges almost surely to $(X_{T,i}, U_{T,i})$ combined with the assumption $\mathbb{P}(X_T \notin \mathcal{D}_{\dot{\psi}}) = 0$. □

Now, by replacing (X_T^n, U_T^n) by their limit (X_T, U_T) in the proof of the first assertion above, we easily get the following result.

Corollary 3.2. *Suppose σ and b are in $\mathscr{C}_b^1$. Assume that $\mathbb{P}(\psi(X_T) \neq 0) > 0$ and there exists $a > 1$ such that $\mathbb{E}\left[\psi^{4a}(X_T)\right]$ and $\mathbb{E}\left[|\nabla\psi(X_T)|^{4a}\right]$ are finite, then the sequence $(\theta_i)_{i\geq 0}$ introduced in the above theorem satisfies*

$$\theta_i \underset{i\to\infty}{\longrightarrow} \theta^* \quad a.s.,$$

where θ^* is given by relation (II.7).

The following corollary follows immediately thanks to theorems 4.4 and 3.1 and Corollary 3.2.

Corollary 3.3. *Under assumptions of Theorem 3.1 and Corollary 3.2, the constrained algorithm given respectively by routine (II.23) satisfies*

$$\lim_{i,n\to\infty} \theta_i^n = \lim_{i\to\infty}(\lim_{n\to\infty} \theta_i^n) = \lim_{n\to\infty}(\lim_{i\to\infty} \theta_i^n) = \theta^*, \quad \tilde{\mathbb{P}}\text{-}a.s.,$$

where θ^* is given by relation (II.7).

3.2 Unconstrained stochastic algorithm

In their recent paper [47], Lemaire and Pagès proposed a new procedure using Robbins-Monro algorithm that satisfies the classical non explosion condition (NEC). In fact, a new expression of the gradient is obtained by a third change of probability. Recall that by Proposition 2.2 we have

$$\nabla v_n(\theta) = \tilde{\mathbb{E}}\left((\theta T - W_T)\left[\psi(X_T^n)^2 + (\nabla\psi(X_T^n)\cdot U_T^n)^2\right] e^{-\theta\cdot W_T + \frac{1}{2}|\theta|^2 T}\right).$$

The aim now is to use their idea in our context. To do so, we apply Girsanov theorem, with the shift parameter $-\theta$. Let $B_t^{(-\theta)} := W_t + \theta t$ and $L_t^{(-\theta)} := \frac{d\mathbb{P}_{(-\theta)}}{d\mathbb{P}}|_{\mathcal{F}_t} = e^{-\theta.W_t - \frac{1}{2}|\theta|^2 t}$, we obtain

$$\nabla v_n(\theta) = \tilde{\mathbb{E}}_{(-\theta)}\left[(2\theta T - B_T^{(-\theta)})\left[\psi(X_T^n)^2 + (\nabla\psi(X_T^n)\cdot U_T^n)^2\right]e^{|\theta|^2 T}\right].$$

As $(B^{(-\theta)}, X^n, U^n)$ under $\tilde{\mathbb{P}}_{(-\theta)}$ has the same law as $(W, X^{n,(-\theta)}, U^{n,(-\theta)})$ under $\tilde{\mathbb{P}}$, we write

$$\nabla v_n(\theta) = \tilde{\mathbb{E}}\left[(2\theta T - W_T)\left[\psi(X_T^{n,(-\theta)})^2 + (\nabla\psi(X_T^{n,(-\theta)})\cdot U_T^{n,(-\theta)})^2\right]e^{|\theta|^2 T}\right].$$

Miming the algorithm proposed by [47], we introduce for a given $\eta > 0$, a new function

$$\tilde{H}_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T) = e^{-\eta|\theta|^2 T}(2\theta T - W_T)\left[\psi(X_T^{n,(-\theta)})^2 + (\nabla\psi(X_T^{n,(-\theta)})\cdot U_T^{n,(-\theta)})^2\right].$$

Then, we introduce for a gain sequence $(\gamma_i)_{i\in\mathbb{N}}$ satisfying (II.22), the algorithm

$$\theta_i^n = \theta_{i-1}^n - \gamma_i \tilde{H}_\eta(\theta_{i-1}^n, X_{T,i}^{n,(-\theta_{i-1}^n)}, U_{T,i}^{n,(-\theta_{i-1}^n)}, W_{T,i}), \quad \theta_0 \in \mathbb{R}^q. \tag{II.25}$$

This algorithm would behave like a classical Robbins-Monro one and does not suffer from the violation of (NEC). Our aim now is to establish the same results satisfied by the constrained routine (II.23) and given by Theorem 3.1. This is splitted into two different theorems. It is worth to note that in this context we will need to control the growth of ψ and its derivatives.

Theorem 3.4. *Suppose σ and b are in $\mathscr{C}_b^1$ and let ψ satisfying $\mathbb{P}(\psi(X_T^n) \neq 0) > 0$, for all $n \in \mathbb{N}$. In addition, assume that for $\lambda > 0$ we have*

$$|\nabla\psi(x)| \leq C_\psi(1 + |x|^\lambda) \quad \textit{for all } x \in \mathcal{D}_{\dot\psi} \textit{ and } C_\psi > 0.$$

Then, the sequence $(\theta_i^n)_{i\geq 0}$ given by routine (II.25), satisfies

$$\forall n \in \mathbb{N}, \quad \theta_i^n \underset{i\to\infty}{\longrightarrow} \theta_n^*, \quad a.s.$$

where θ_n^* is given by relation (II.8).

Proof. To prove the almost sure convergence we will use the classical Robbins-Monro theorem (see Theorem 2.2.12 page 52 in [24]). Let $n \in \mathbb{N}$, under our assumptions the existence of θ_n^* is ensured by Proposition 2.2 and we have to check first that

$$\forall \theta \neq \theta_n^* \quad \langle h_n(\theta), \theta - \theta_n^*\rangle > 0, \qquad \text{where} \qquad h_n(\theta) = \tilde{\mathbb{E}}\, H_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T).$$

This is immediate since $h_n(\theta) = K_\eta(\theta)\nabla v_n(\theta)$ with $K_\eta > 0$ and v_n is a strictly convex function. Now it remains to prove that $\sup_{\theta\in\mathbb{R}^q} \tilde{\mathbb{E}}\left[|H_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T)|^2\right] < \infty$, which guaranties the (NEC) condition. By Cauchy-Schwartz inequality we obtain

$$\tilde{\mathbb{E}}\left[|H_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T)|^2\right] \leq e^{-2\eta|\theta|^2T} \left\||2\theta T - W_T|^2\right\|_2 \times \left(\left\|\psi(X_T^{n,(-\theta)})^2\right\|_2 + \left\|(\nabla\psi(X_T^{n,(-\theta)}) \cdot U_T^{n,(-\theta)})^2\right\|_2\right).$$

Using the polynomial growth assumption on $\nabla\psi$, the second and third term on the right hand side of the above inequality can be bounded respectively up to a standard positive constant by

$$1 + \left\||X_T^{n,(-\theta)}|^{2(\lambda+1)}\right\|_2 \text{ and } 1 + \left\||X_T^{n,(-\theta)}|^{4\lambda}\right\|_2 + \left\||U_T^{n,(-\theta)}|^4\right\|_2.$$

In the following proof, C will denote a positive standard constant that may change from line to line. Let $\lambda_1 = 4\lambda \vee 2(\lambda+1)$, using the identity $(1+x)^\rho \leq C(1+x^\rho)$ for $x \geq 0$ and $\rho \geq 1$, then we have

$$\tilde{\mathbb{E}}\left[|H_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T)|^2\right] \leq Ce^{-2\eta|\theta|^2T}(1+|\theta|^2)\left(1 + \left\||X_T^{n,(-\theta)}|^{\lambda_1}\right\|_2 + \left\||U_T^{n,(-\theta)}|^4\right\|_2\right).$$

As $(B^{(-\theta)}, X^n, U^n)$ under $\tilde{\mathbb{P}}_{(-\theta)}$ has the same law as $(W, X^{n,(-\theta)}, U^{n,(-\theta)})$ under $\tilde{\mathbb{P}}$, we write

$$\tilde{\mathbb{E}}\left|X_T^{n,(-\theta)}\right|^{2\lambda_1} = \tilde{\mathbb{E}}_{(-\theta)}\left|X_T^n\right|^{2\lambda_1} = \tilde{\mathbb{E}}\left(\left|X_T^n\right|^{2\lambda_1} e^{-\theta\cdot W_T - \frac{1}{2}|\theta|^2T}\right)$$

and

$$\tilde{\mathbb{E}}\left|U_T^{n,(-\theta)}\right|^8 = \tilde{\mathbb{E}}_{(-\theta)}\left|U_T^n\right|^8 = \tilde{\mathbb{E}}\left(\left|U_T^n\right|^8 e^{-\theta\cdot W_T - \frac{1}{2}|\theta|^2T}\right).$$

Now using Hölder's inequality, with $\frac{1}{r} + \frac{1}{r'} = 1$, properties ($\mathcal{P}$) and ($\tilde{\mathcal{P}}$) and $\left(\tilde{\mathbb{E}}e^{-r\theta\cdot W_T - \frac{r}{2}|\theta|^2T}\right)^{\frac{1}{r}} = e^{\frac{r-1}{2}|\theta|^2T}$, we obtain

$$\tilde{\mathbb{E}}\left[|H_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T)|^2\right] \leq C(1+|\theta|^2)e^{-(2\eta - \frac{r-1}{4})|\theta|^2T}.$$

Then, one sees immediately that $\sup_{\theta\in\mathbb{R}^q} \tilde{\mathbb{E}}\left[|H_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T)|^2\right]$ is finite by choosing $r \in (1, 1+8\eta)$. This completes the proof. □

In the same way as in the constrained case, we deduce the following result if we replace (X_T^n, U_T^n) by their limit (X_T, U_T) in the above proof.

Corollary 3.5. *Suppose σ and b are in $\mathscr{C}_b^1$. Let ψ satisfying $\mathbb{P}(\psi(X_T) \neq 0) > 0$ and*

$$|\nabla\psi(x)| \leq C_\psi(1+|x|^\lambda) \quad \text{for all } x \in \mathcal{D}_{\dot{\psi}} \text{ and } C_\psi, \lambda > 0.$$

Then, the sequence $(\theta_i)_{i\geq 0}$, obtained when replacing in routine (II.25) $(X^n_{T,i}, U^n_{T,i})_{i\geq 1}$ by their limit $(X_{T,i}, U_{T,i})_{i\geq 1}$, satisfies

$$\theta_i \underset{i\to\infty}{\longrightarrow} \theta^*, \quad a.s.$$

where θ^* is given by relation (II.7).

The aim now is to prove that the same property *2.* in Theorem 3.1, is satisfied by the unconstrained algorithm (II.25). This task looks more complicated to achieve, since for a fixed $i \geq 0$ the stochastic term θ^n_i also appears in the drift part of the pair $(X^{n,(-\theta^n_i)}_{T,i+1}, U^{n,(-\theta^n_i)}_{T,i+1})$. To overcome this technical difficulty we firstly strengthen our hypothesis on the triplet (b, σ, ψ) and secondly make use of the so called θ-sensitivity process given by $(\frac{\partial}{\partial\theta}X^{n,(-\theta)}_T, \frac{\partial}{\partial\theta}U^{n,(-\theta)}_T)$.

Theorem 3.6. *Let b and σ in $\mathscr{C}_b^{2,\delta}$, $\delta > 0$. Assume that ψ is $\mathscr{C}^2$ with polynomial growth as well as all its partial derivatives until order two and satisfies $\mathbb{P}(\psi(X_T) \neq 0) > 0$ and $\mathbb{P}(\psi(X^n_T) \neq 0) > 0$, for all $n \geq 1$. Then, $\forall i \in \mathbb{N}$ and $\forall p \geq 1$, there exists $C > 0$ depending only on i, p, b, σ and T such that*

$$\forall n \in \mathbb{N}^*, \quad \tilde{\mathbb{E}}|\theta^n_{i+1} - \theta_{i+1}|^{2p} \leq \frac{C}{n^p}.$$

Consequently, $\forall i \in \mathbb{N}$ $\theta^n_i \underset{n\to\infty}{\longrightarrow} \theta_i, a.s.$ where the sequence $(\theta_i)_{i\geq 0}$ is introduced in the above corollary.

Proof. We first proceed by induction on $i \in \mathbb{N}$ to prove the first assertion. The case when $i = 0$ is trivial since $\theta^n_0 = \theta_0 \in \mathbb{R}^q$. We now assume the assertion holds for a fixed integer i and show that it also holds for $i+1$. First, we write $\theta^n_{i+1} - \theta_{i+1} = \theta^n_i - \theta_i - \gamma_{i+1}(H_1 + H_2)$ where

$$\begin{aligned} H_1 := {} & e^{-\eta|\theta^n_i|^2 T}(2\theta^n_i T - W_{T,i+1}) \\ & \times \left[\psi(X^{n,(-\theta^n_i)}_{T,i+1})^2 - \psi(X^{(-\theta^n_i)}_{T,i+1})^2 + \left(\nabla\psi(X^{n,(-\theta^n_i)}_{T,i+1}).U^{n,(-\theta^n_i)}_{T,i+1}\right)^2 - \left(\nabla\psi(X^{(-\theta^n_i)}_{T,i+1}).U^{(-\theta^n_i)}_{T,i+1}\right)^2\right] \end{aligned}$$

and

$$\begin{aligned} H_2 := {} & e^{-\eta|\theta^n_i|^2 T}(2\theta^n_i T - W_{T,i+1}) \left[\psi(X^{(-\theta^n_i)}_{T,i+1})^2 + \left(\nabla\psi(X^{(-\theta^n_i)}_{T,i+1}).U^{(-\theta^n_i)}_{T,i+1}\right)^2\right] \\ & - e^{-\eta|\theta_i|^2 T}(2\theta_i T - W_{T,i+1}) \left[\psi(X^{(-\theta_i)}_{T,i+1})^2 + \left(\nabla\psi(X^{(-\theta_i)}_{T,i+1}).U^{(-\theta_i)}_{T,i+1}\right)^2\right]. \end{aligned}$$

Hence, for all $p \geq 1$, we have

$$\tilde{\mathbb{E}}|\theta_{i+1}^n - \theta_{i+1}|^{2p} \leq 3^{2p-1}\tilde{\mathbb{E}}|\theta_i^n - \theta_i|^{2p} + 3^{2p-1}\gamma_{i+1}^{2p}(\tilde{\mathbb{E}}|H_1|^{2p} + \tilde{\mathbb{E}}|H_2|^{2p}). \tag{II.26}$$

Using the induction assumption we only need to control the second and third terms on the right hand side of the inequality (II.26) above.

Term H_1 Using that θ_i^n is $\tilde{\mathfrak{F}}_{T,i}$-measurable, $W_{T,i+1} \perp\!\!\!\perp \tilde{\mathfrak{F}}_{T,i}$ we write $\tilde{\mathbb{E}}|H_1|^{2p} = \tilde{\mathbb{E}}A(\theta_i^n)$ where for all $\theta \in \mathbb{R}^q$

$$A(\theta) := e^{-2p\eta|\theta|^2 T}\tilde{\mathbb{E}}\Big[|2\theta T - W_T|^{2p} \\ \times |\psi(X_T^{n,(-\theta)})^2 - \psi(X_T^{(-\theta)})^2 + (\nabla\psi(X_T^{n,(-\theta)}).U_T^{n,(-\theta)})^2 - (\nabla\psi(X_T^{(-\theta)}).U_T^{(-\theta)})^2|^{2p}\Big].$$

Since $(B^{(-\theta)}, X^n, U^n, X, U)$ under $\tilde{\mathbb{P}}_{(-\theta)}$ has the same law as $(W, X^{n,(-\theta)}, U^{n,(-\theta)}, X^{(-\theta)}, U^{(-\theta)})$ under $\tilde{\mathbb{P}}$ for all $\theta \in \mathbb{R}$, we obtain by a change of probability measure

$$A(\theta) = e^{(-2p\eta-\frac{1}{2})|\theta|^2 T}\tilde{\mathbb{E}}\Big[|\theta T - W_T|^{2p} e^{-\theta.W_T} \\ \times |\psi(X_T^n)^2 - \psi(X_T)^2 + (\nabla\psi(X_T^n).U_T^n)^2 - (\nabla\psi(X_T).U_T)^2\,|^{2p}\Big]$$

By Hölder's inequality, we obtain $\forall r_1 \in (1, \infty)$,

$$A(\theta) \leq e^{(-2p\eta-\frac{1}{2})|\theta|^2 T}\|e^{-\theta.W_T}\|_{r_1}\||\theta T - W_T|^{2p}\|_{\frac{2r_1}{r_1-1}} \\ \times \||\psi(X_T^n)^2 - \psi(X_T)^2 + (\nabla\psi(X_T^n).U_T^n)^2 - (\nabla\psi(X_T).U_T)^2|^{2p}\|_{\frac{2r_1}{r_1-1}}$$

As $e^{(-2p\eta-\frac{1}{2})|\theta|^2 T}\|e^{-\theta.W_T}\|_{r_1}\||\theta T - W_T|^{2p}\|_{\frac{2r_1}{r_1-1}} \leq c_1(1+|\theta|^{2p})e^{(\frac{r_1}{2}-2p\eta-\frac{1}{2})|\theta|^2 T}$, with c_1 is a positive constant depending only on p, r_1 and T. Then, one can choose $r_1 \in (1, 1+4p\eta)$ such that $\sup_{\theta\in\mathbb{R}^q}(1+|\theta|^{2p})e^{(\frac{r_1}{2}-2p\eta-\frac{1}{2})|\theta|^2 T}$ is finite. Hence, we get the existence of a constant c_2 depending only on p, η and T such that

$$A(\theta) \leq c_2\||\psi(X_T^n)^2 - \psi(X_T)^2 + (\nabla\psi(X_T^n).U_T^n)^2 - (\nabla\psi(X_T).U_T)^2|^{2p}\|_{\frac{2r_1}{r_1-1}}. \tag{II.27}$$

Since ψ is $\mathscr{C}^2$ with polynomial growth as well as all its partial derivatives until order two then the function $g(x,y) := \psi^2(x) + \nabla\psi(x).y$, $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$, is $\mathscr{C}^1$ and all its partial derivatives of order one have polynomial growth. Hence, the Taylor expansion on the real-valued function

g yields the existence of a point $(\bar{X}_T^n, \bar{U}_T^n)$ between (X_T^n, U_T^n) and (X_T, U_T) such that

$$g(X_T^n, U_T^n) - g(X_T, U_T) = \nabla g(\bar{X}_T^n, \bar{U}_T^n).(X_T^n - X_T, U_T^n - U_T).$$

Then by the Cauchy-Schwarz inequality, the polynomial growth of g and properties $(\mathcal{P})$ and $(\tilde{\mathcal{P}})$ we get the existence of a constant c_3 depending only on p, η, T, b, σ and ψ such that

$$\begin{aligned} \| |g(X_T^n, U_T^n) - g(X_T, U_T)|^{2p} \|_{\frac{2r_1}{r_1-1}} &\leq \| |\nabla g(\bar{X}_T^n, \bar{U}_T^n)|^{2p} \|_{\frac{4r_1}{r_1-1}} \| |(X_T^n - X_T, U_T^n - U_T)|^{2p} \|_{\frac{4r_1}{r_1-1}} \\ &\leq \frac{c_3}{n^p}. \end{aligned} \tag{II.28}$$

So, (II.27) and (II.28) tell us $A(\theta) \leq \frac{c_2 c_3}{n^p}$, and we deduce the existence of a deterministic constant c_4 depending only on p, η, T, b, σ and ψ such that

$$\tilde{\mathbb{E}} H_1^{2p} \leq \frac{c_4}{n^p}. \tag{II.29}$$

Term H_2 Using that θ_i^n and θ_i are $\tilde{\mathfrak{F}}_{T,i}$-measurable, $W_{T,i+1} \perp\!\!\!\perp \tilde{\mathfrak{F}}_{T,i}$ we write $\tilde{\mathbb{E}}|H_2|^{2p} = \tilde{\mathbb{E}} B(\theta_i^n, \theta_i)$ where for all $(\theta, \theta') \in \mathbb{R}^q \times \mathbb{R}^q$

$$B(\theta, \theta') := \tilde{\mathbb{E}} |e^{-\eta|\theta|^2 T}(2\theta T - W_T) g(X_T^{(-\theta)}, U_T^{(-\theta)}) - e^{-\eta|\theta'|^2 T}(2\theta' T - W_T) g(X_T^{(-\theta')}, U_T^{(-\theta')})|^{2p}. \tag{II.30}$$

According to the study of θ-sensitivity of the processes $(X_t^{(-\theta)})_{t\in[0,T]}$ and $(U_t^{(-\theta)})_{t\in[0,T]}$ given in lemma 3.7 below, we have that for a time $t \in [0, T]$ the function $\theta \mapsto (X_t^{(-\theta)}, U_t^{(-\theta)})$ is almost surely $\mathscr{C}^1$. Hence, we deduce that almost surely the function $\theta \mapsto D(\theta) := e^{-\eta|\theta|^2 T}(2\theta T - W_T) g(X_T^{(-\theta)}, U_T^{(-\theta)})$ is also $\mathscr{C}^1$. This allows us to apply Taylor expansion on each component $D^{\ell'}$ of D, $\ell' \in \{1, \cdots, q\}$, and by standard evaluations we obtain a constant c_5 depending only p and q such that

$$\begin{aligned} B(\theta, \theta') &= \tilde{\mathbb{E}} | \sum_{\ell'=1}^{q} \Big(D^{\ell'}(\theta) - D^{\ell'}(\theta') \Big)^2 |^p = \tilde{\mathbb{E}} | \sum_{\ell'=1}^{q} \Big(\sum_{\ell=1}^{q} (\theta_\ell - \theta'_\ell) \int_0^1 \frac{\partial D^{\ell'}}{\partial \theta_\ell}(t\theta' + (1-t)\theta) dt \Big)^2 |^p \\ &\leq c_5 |\theta - \theta'|^{2p} \sum_{\ell, \ell'=1}^{q} \int_0^1 \tilde{\mathbb{E}} | \frac{\partial D^{\ell'}}{\partial \theta_\ell}(t\theta' + (1-t)\theta)|^{2p} dt. \end{aligned}$$

The term $\tilde{\mathbb{E}}|\frac{\partial D^{\ell'}}{\partial u_\ell}(u)|^{2p}$ is bounded uniformly on $u \in \mathbb{R}^q$. More precisely, we have the following result.

Lemma 3.7. *The solutions $(X_t^{(-\theta)})_{t\in[0,T]}$ and $(U_t^{(-\theta)})_{t\in[0,T]}$ of respectively Itô's stochastic differ-*

ential equations (II.4) and (II.6) have modifications of $\mathscr{C}^1$ with respect to the parameter θ and their partial derivatives are L^p-bounded for all $p \geq 1$. Further, there exists a positive constant c depending only on p, q, b, σ, ψ and T such that

$$\tilde{\mathbb{E}}|\frac{\partial D^{\ell'}}{\partial u_\ell}(u)|^{2p} \leq c \quad \forall u \in \mathbb{R}^q \text{ and } \ell, \ell' \in \{1, \cdots, q\}.$$

For the reader convenience, the proof of this lemma is postponed to the end of the current subsection. Thus, thanks to Lemma 3.7 above there is a constant c_6 depending only on p, q, b, σ, ψ and T such that $B(\theta, \theta') \leq c_6|\theta - \theta'|^{2p}$, and it follows from $\tilde{\mathbb{E}}|H_2|^{2p} = \tilde{\mathbb{E}}B(\theta_i^n, \theta_i)$ that

$$\tilde{\mathbb{E}}H_2^{2p} \leq c_6\mathbb{E}|\theta_i^n - \theta_i|^{2p}. \tag{II.31}$$

So, (II.26), (II.29) and (II.31) show that

$$\tilde{\mathbb{E}}|\theta_{i+1}^n - \theta_{i+1}|^{2p} \leq 3^{2p-1}(1 + c_6\gamma_{i+1}^{2p})\tilde{\mathbb{E}}|\theta_i^n - \theta_i|^{2p} + 3^{2p-1}\gamma_{i+1}^{2p}\frac{c_4}{n^p}.$$

Using the induction assumption for stage i, we deduce for $p \geq 1$ the existence of a positive constant C depending only on p, q, b, σ, ψ, T and i such that

$$\forall n \in \mathbb{N}^*, \quad \tilde{\mathbb{E}}|\theta_{i+1}^n - \theta_{i+1}|^{2p} \leq \frac{C}{n^p}.$$

Finally, for all $i \in \mathbb{N}$, the almost sure convergence, of θ_i^n towards θ_i as n tends to ∞ is a classical and immediate consequence of the first assertion shown above, based on the Borel-Cantelli lemma. □

The following corollary follows immediately thanks to theorems 4.4, 3.4 and 3.6 and Corollary 3.5.

Corollary 3.8. *Under assumptions of Theorem 3.6 and $\mathbb{P}(\psi(X_T) \neq 0) > 0$, the unconstrained algorithm given respectively by routine (II.25) satisfies*

$$\lim_{i,n\to\infty} \theta_i^n = \lim_{i\to\infty}(\lim_{n\to\infty} \theta_i^n) = \lim_{n\to\infty}(\lim_{i\to\infty} \theta_i^n) = \theta^*, \quad \tilde{\mathbb{P}}\text{-a.s.},$$

where θ^* is given by relation (II.7).

Our task now is to show the result given by Lemma 3.7 and used in the proof of Theorem 3.6.

Proof.(Proof of Lemma 3.7). It is worth to note that all theoretical results known on the differentiation of the solution of Itô's stochastic differential equation with respect to its initial value, can be extended to any parameter. Thus, thanks to Theorem 4.6.5 in [40], our assumptions on b and σ ensures the differentiability of the processes $(X_t^{(-u)})_{0\leq t\leq T}$. Further, if we denote by $\partial_\ell X_t^{(-u)}$ the processes where we take the partial derivatives of all components of $(X_t^{(-u)})_{0\leq t\leq T}$ with respect to the ℓ^{th} variable u_l then the $\mathbb{R}^d$-valued process $(\partial_\ell X_t^{(-u)})_{0\leq t\leq T}$ satisfies the stochastic differential equation

$$\partial_\ell X_t^{(-u)} = \Big(\dot{b}(X_t^{(-u)})\partial_\ell X_t^{(-u)} - \sigma_\ell(X_t^{(-u)})\Big)dt + \sum_{j=1}^{q} \dot{\sigma}_j(X_t^{(-u)})\partial_\ell X_t^{(-u)}(dW_t^j - u_j dt). \tag{II.32}$$

Moreover, Corollary 4.6.7 in [40] ensures the L^p boundedness of the random variable $\partial_\ell X_t^{(-u)}$, $t\in[0,T]$ and $p\geq 1$. Concerning the process $(U_t^{(-u)})_{0\leq t\leq T}$, we need a more general result to study its u-sensitivity, we apply Theorem 4.6.4 in [40] to obtain its differentiability with respect to u. The process $(\partial_\ell U_t^{(-u)})_{0\leq t\leq T}$ is defined similarly and for $i\in\{1,\cdots,d\}$, we denote by $(\partial_\ell(U_t^{(-u)})^i)_{0\leq t\leq T}$ its i^{th} component satisfying the stochastic differential system

$$\begin{aligned}\partial_\ell(U_t^{(-u)})^i &= \Big((\partial_\ell X_t^{(-u)})^{tr}\ddot{b}_i(X_t^{(-u)})U_t^{(-u)} + \dot{b}_i(X_t^{(-u)})\partial_\ell U_t^{(-u)} - \dot{\sigma}_\ell(X_t^{(-u)})U_t^{(-u)}\Big)dt \\ &+ \sum_{j=1}^{q}\Big((\partial_\ell X_t^{(-u)})^{tr}\ddot{\sigma}_{ij}(X_t^{(-u)})U_t^{(-u)} + \dot{\sigma}_{ij}(X_t^{(-u)})\partial_\ell U_t^{(-u)}\Big)(dW_t^j - u_j dt) \\ &- \frac{1}{\sqrt{2}}\sum_{j,j'=1}^{q}\Big((\partial_\ell X_t^{(-u)})^{tr}\ddot{\sigma}_{ij}(X_t^{(-u)})\sigma_{j'}(X_t^{(-u)}) + \dot{\sigma}_{ij}(X_t^{(-u)})\dot{\sigma}_{j'}(X_t^{(-u)})\partial_\ell X_t^{(-u)}\Big)d\tilde{W}_t^{j'j}.\end{aligned} \tag{II.33}$$

Moreover, the same Theorem 4.6.4 in [40] ensures that these components are also L^p bounded, for all $p\geq 1$.

Now, let us recall that the $\mathbb{R}^q$-valued function D is defined by $D(u) = e^{-\eta|u|^2T}(2uT - W_T)g(X_T^{(-u)}, U_T^{(-u)})$. For $\ell,\ell'\in\{1,\cdots,q\}$, the partial derivative of component $D^{\ell'}$ with respect to u_ℓ is given by

$$\begin{aligned}\frac{\partial D^{\ell'}}{\partial u_\ell}(u) &= \Big(2T\delta_{\ell\ell'} - 2u_\ell T\eta(2u_{\ell'}T - W_T^{\ell'})\Big)e^{-\eta|u|^2T}g(X_T^{(-u)}, U_T^{(-u)}) \\ &+ e^{-\eta|u|^2T}(2u_{\ell'}T - W_T^{\ell'})\Big(\nabla_x g(X_T^{(-u)}, U_T^{(-u)}).\partial_\ell X_T^{(-u)} + \nabla_y g(X_T^{(-u)}, U_T^{(-u)}).\partial_\ell U_T^{(-u)}\Big).\end{aligned} \tag{II.34}$$

Here, for the function $g:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$, we denote by $\nabla_x g$ (resp. $\nabla_y g$) the gradient with respect to the first variable x (resp. the second variable y), and the notation $\delta_{\ell\ell'}$ stands for the

Kronecker symbol. Now, let Y and Z be solution to the following stochastic differential system

$$dY_{t,\ell} = \Big(\dot{b}(X_t)Y_{t,\ell} - \sigma_\ell(X_t)\Big)dt + \sum_{j=1}^{q} \dot{\sigma}_j(X_t)Y_{t,\ell}dW_t^j.$$

and

$$\begin{aligned} d(Z_{t,\ell})^i &= \Big((Y_{t,\ell})^{tr}\ddot{b}_i(X_t)U_t + \dot{b}_i(X_t)Z_{t,\ell} - \dot{\sigma}_\ell(X_t)U_t\Big)dt \\ &+ \sum_{j=1}^{q}\Big((Y_{t,\ell})^{tr}\ddot{\sigma}_{ij}(X_t)U_t + \dot{\sigma}_{ij}(X_t)Z_{t,\ell}\Big)dW_t^j \\ &-\frac{1}{\sqrt{2}}\sum_{j,j'=1}^{q}\Big((Y_{t,\ell})^{tr}\ddot{\sigma}_{ij}(X_t)\sigma_{j'}(X_t) + \dot{\sigma}_{ij}(X_t)\dot{\sigma}_{j'}(X_t)Y_{t,\ell}\Big)d\tilde{W}_t^{j'j}. \end{aligned}$$

These both processes can be seen as solutions of respectively (II.32) and (II.33) at point $u = 0$, consequently they are L^p bounded, $p \geq 1$. Note that (II.4), (II.6), (II.32) and (II.33) allow us to apply Girsanov theorem and deduce that $(W, X^{(-u)}, U^{(-u)}, \partial_\ell X^{(-u)}, \partial_\ell U^{(-u)})$ under $\tilde{\mathbb{P}}$ has the same law as $(B^{(-u)}, X, U, Y_{.,\ell}, Z_{.,\ell})$ under $\tilde{\mathbb{P}}_{(-u)}$. Hence, using relation (II.34) followed by a change of a probability measure, we obtain

$$\begin{aligned} \tilde{\mathbb{E}}|\frac{\partial D^{\ell'}}{\partial u_\ell}(u)|^{2p} = \tilde{\mathbb{E}}\Big[|\Big(2T\delta_{\ell\ell'} - 2u_\ell T\eta(u_{\ell'}T - W_T^{\ell'})\Big)e^{-\eta|u|^2T}g(X_T, U_T) \\ + e^{-\eta|u|^2T}(u_{\ell'}T - W_T^{\ell'})\Big(\nabla_x g(X_T,U_T).Y_{T,\ell} + \nabla_y g(X_T,U_T).Z_{T,\ell}\Big)|^{2p}e^{-u.W_T - \frac{1}{2}|u|^2T}\Big]. \end{aligned}$$

Rearranging the terms in the above inequality, we get by Hölder's inequality $\forall r_1 \in (1,\infty)$

$$\begin{aligned} \tilde{\mathbb{E}}|\frac{\partial D^{\ell'}}{\partial u_\ell}(u)|^{2p} \leq e^{(-2p\eta-\frac{1}{2})|u|^2T}\|e^{-u.W_T}\|_{r_1}\Big\|(2T+1+2|u_\ell|T\eta)^{2p}|u_{\ell'}T - W_T^{\ell'}|^{2p}\Big\|_{\frac{2r_1}{r_1-1}} \\ \times \Big\|(|g(X_T,U_T)| + |\nabla_x g(X_T,U_T).Y_{T,\ell} + \nabla_y g(X_T,U_T).Z_{T,\ell}|)^{2p}\Big\|_{\frac{2r_1}{r_1-1}}. \end{aligned}$$

As $e^{(-2p\eta-\frac{1}{2})|u|^2T}\|e^{-u.W_T}\|_{r_1}\Big\|(2T+1+2|u_\ell|T\eta)^{2p}|u_{\ell'}T-W_T^{\ell'}|^{2p}\Big\|_{\frac{2r_1}{r_1-1}} \leq c_1(1+|u|^{4p})e^{(\frac{r_1}{2}-2p\eta-\frac{1}{2})|u|^2T}$, with c_1 is a positive constant depending only on p, r_1 and T. Then, one can choose $r_1 \in (1, 1+4p\eta)$ such that $\sup_{u\in\mathbb{R}^q}(1+|u|^{2p})e^{(\frac{r_1}{2}-2p\eta-\frac{1}{2})|u|^2T}$ is finite. Hence, we get the existence of a constant c_2 depending only on p, η and T such that

$$\tilde{\mathbb{E}}|\frac{\partial D^{\ell'}}{\partial u_\ell}(u)|^{2p} \leq c_2\Big\|(|g(X_T,U_T)| + |\nabla_x g(X_T,U_T).Y_{T,\ell} + \nabla_y g(X_T,U_T).Z_{T,\ell}|)^{2p}\Big\|_{\frac{2r_1}{r_1-1}}.$$

Since ψ is $\mathscr{C}^2$ with polynomial growth as well as all its partial derivatives until order two then the function g mapping the couple (x, y) into $\psi^2(x) + \nabla\psi(x).y$ is $\mathscr{C}^1$ and all its first partial derivatives have polynomial growth. The proof is completed, thanks to properties $(\mathcal{P})$, $(\tilde{\mathcal{P}})$ and using the L^p boundedness of Y_T and Z_T for all $p \geq 1$. □

4 Central limit theorem for the adaptive procedure

In this section, we prove a central limit theorem for both adaptive Monte Carlo and adaptive statistical Romberg methods. Let us recall that the adaptive importance sampling algorithm for the statistical Romberg method approximates our initial quantity of interest $\mathbb{E}\psi(X_T) = \mathbb{E}\left[\psi(X_T^\theta)e^{-\theta\cdot W_T - \frac{1}{2}|\theta|^2 T}\right]$ by

$$\frac{1}{N_1}\sum_{i=1}^{N_1} g(\hat{\theta}_i^m, \hat{X}_{T,i+1}^{m,\hat{\theta}_i^m}, \hat{W}_{T,i+1}) + \frac{1}{N_2}\sum_{i=1}^{N_2}\Big(g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - g(\theta_i^n, X_{T,i+1}^{m,\theta_i^n}, W_{T,i+1})\Big), \quad \text{(II.35)}$$

where for all $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^q$, $g(\theta, x, y) = \psi(x)e^{-\theta\cdot y - \frac{1}{2}|\theta|^2 T}$. Here the paths generated by W and $\hat{W}$ are of course independent. In order to prove a central limit theorem for this algorithm, we need to study independently each of the above empirical means. This is the aim of subsections 4.2 and 4.3. We need first to recall some useful results.

4.1 Technical results

Let us recall the Central Limit Theorem for martingales array (see e.g. [24]).

Theorem 4.1. *Suppose that $(\Omega, \mathbb{F}, \mathbb{P})$ is a probability space and that for each n, we have a filtration $\mathbb{F}_n = (\mathcal{F}_k^n)_{k\geq 0}$, a sequence $k_n \longrightarrow \infty$ as $n \longrightarrow \infty$ and a real square integrable vector martingale $M^n = (M_k^n)_{k\geq 0}$ which is adapted to $\mathbb{F}_n$ and has quadratic variation denoted by $(\langle M\rangle_k^n)_{k\geq 0}$. We make the following two assumptions.*

A1. There exists a deterministic symmetric positive semi-definite matrix Γ, such that

$$\langle M\rangle_{k_n}^n = \sum_{k=1}^{k_n}\mathbb{E}\left[|M_k^n - M_{k-1}^n|^2|\mathcal{F}_{k-1}^n\right] \xrightarrow[n\to\infty]{\mathbb{P}} \Gamma.$$

A2. Lindeberg's condition holds : that is, for all $\varepsilon > 0$,

$$\sum_{k=1}^{k_n}\mathbb{E}\left[|M_k^n - M_{k-1}^n|^2 1_{\{|M_k^n - M_{k-1}^n|>\varepsilon\}}|\mathcal{F}_{k-1}^n\right] \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Then

$$M^n_{k_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma) \quad \text{as } n \to \infty.$$

Remark 4.2. *The following assumption known as the Lyapunov condition, implies the Lindeberg's condition A2.,*

A3. There exists a real number $a > 1$, such that

$$\sum_{k=1}^{k_n} \mathbb{E}\left[|M^n_k - M^n_{k-1}|^{2a} | \mathcal{F}^n_{k-1}\right] \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

As a prelude to the results of this subsection, we give a double indexed version of the Toeplitz lemma that will be very helpful in the sequel.

Lemma 4.3. *Let $(a_i)_{1\le i\le k_n}$ a sequence of real positive numbers, where $k_n \uparrow \infty$ as n tends to infinity, and $(x^n_i)_{i\ge1,n\ge1}$ a double indexed sequence such that*

(i) $\lim\limits_{n\to\infty} \sum_{1\le i\le k_n} a_i = \infty$

(ii) $\lim\limits_{i,n\to\infty} x^n_i = \lim\limits_{i\to\infty}(\lim\limits_{n\to\infty} x^n_i) = \lim\limits_{n\to\infty}(\lim\limits_{i\to\infty} x^n_i) = x < \infty.$

Then,

$$\lim_{n\to+\infty} \frac{\sum_{i=1}^{k_n} a_i x^n_i}{\sum_{i=1}^{k_n} a_i} = x.$$

Proof. For all $\varepsilon > 0$, there exists $N_1(\varepsilon)$ such that for all $n \ge N_1(\varepsilon)$ and $i \ge N_1(\varepsilon)$, we have that :

$$|x^n_i - x| \le \frac{\varepsilon}{2}.$$

As k_n goes to infinity, there exists $N_2(\varepsilon)$ such that for all $n \ge N_2(\varepsilon)$, we have $k_n \ge N_1(\varepsilon)$. Therefore, for all $n \ge sup(N_1(\varepsilon), N_2(\varepsilon)) = N(\varepsilon)$, we can write :

$$\sum_{i=1}^{k_n} a_i |x^n_i - x| = \sum_{i=1}^{N_1(\varepsilon)-1} a_i |x^n_i - x| + \sum_{i=N_1(\varepsilon)}^{k_n} a_i |x^n_i - x|.$$

For the second term of the expression above, we have :

$$\sum_{i=N_1(\varepsilon)}^{k_n} a_i |x^n_i - x| \le \frac{\varepsilon}{2} \sum_{i=N_1(\varepsilon)}^{k_n} a_i \le \frac{\varepsilon}{2} \sum_{i=1}^{k_n} a_i.$$

On the other hand, by assumptions (i) and (ii) there exists $\tilde{N}(\varepsilon)$ such that for all $n \ge \tilde{N}(\varepsilon)$

$$\frac{\sup_{1\le i\le N_1(\varepsilon)-1} \sup_{n\ge1} |x^n_i - x| \sum_{1\le i\le N_1(\varepsilon)-1} a_i}{\sum_{1\le i\le k_n} a_i} \le \frac{\varepsilon}{2}.$$

Therefore, for all $n \geq \tilde{N}(\varepsilon)$

$$\left| \frac{\sum_{i=1}^{k_n} a_i x_i^n}{\sum_{i=1}^{k_n} a_i} - x \right| \leq \varepsilon.$$

This completes the proof. □

Let $\tilde{\mathcal{B}} = (\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t\geq 0}, \tilde{\mathbb{P}})$ be the extension probability space introduced in Section 2 endowed with the filtration $\tilde{\mathcal{F}}_{T,i} = \sigma(W_{t,l}, \tilde{W}_{t,l}, l \leq i, t \leq T)$ given in the very beginning of Section 3. In what follows, let $(\theta_i^n)_{i\geq 0}$, $n \in \mathbb{N}$ and $(\theta_i)_{i\geq 0}$ a family of sequences satisfying

$$(\mathcal{H}_\theta) \quad \begin{cases} \text{For each } n \in \mathbb{N},\ (\theta_i^n)_{i\geq 0} \text{ and } (\theta_i)_{i\geq 0} \text{ are } (\tilde{\mathcal{F}}_{T,i})_{i\geq 0}\text{-adapted} \\ \\ \lim_{i\to\infty}\big(\lim_{n\to\infty} \theta_i^n\big) = \lim_{i\to\infty} \theta_i = \lim_{n\to\infty}\big(\lim_{i\to\infty} \theta_i^n\big) = \lim_{n\to\infty} \theta_n^* = \theta^*, \quad \tilde{\mathbb{P}}\text{-}a.s., \end{cases}$$

with deterministic limits θ^* and θ_n^*.

4.2 The adaptive Monte Carlo method

Let us recall that the statistical Romberg algorithm (II.35) runs successively two independent empirical means. The first one is a crude Monte Carlo simply depending on the Euler scheme with the coarse time step T/m. However, the second empirical mean involves the functional difference between the fine Euler scheme with time step T/n and the coarse one constructed from the same Brownian path. The task now is to prove a central limit theorem for the first empirical mean.

Theorem 4.4. *Let $(\theta_i^n)_{i\geq 0}$, $n \in \mathbb{N}$ and $(\theta_i)_{i\geq 0}$ be a family of sequences satisfying $(\mathcal{H}_\theta)$. Moreover, assume that b and σ satisfy the global Lipschitz condition $(\mathcal{H}_{b,\sigma})$ and the function ψ is a real valued function satisfying assumption $(\mathcal{H}_{\varepsilon_n})$, with $\alpha \in [1/2, 1]$ and $C_\psi \in \mathbb{R}$, such that $|\psi(x) - \psi(y)| \leq C(1 + |x|^p + |y|^p)|x - y|$, for some $C, p > 0$, then the following convergence holds*

$$n^\alpha \left(\frac{1}{n^{2\alpha}} \sum_{i=1}^{n^{2\alpha}} g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - \mathbb{E}\psi(X_T) \right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(C_\psi, \sigma^2\right).$$

where $\sigma^2 := \mathbb{E}\left(\psi(X_T)^2 e^{-\theta^*.W_T - \frac{1}{2}|\theta^*|^2 T}\right) - [\mathbb{E}\psi(X_T)]^2$ and for all $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^q$, $g(\theta, x, y) = \psi(x) e^{-\theta\cdot y - \frac{1}{2}|\theta|^2 T}$. Furthermore, we have also for all $\alpha, \beta > 0$

$$n^\alpha \left(\frac{1}{n^{2\alpha}} \sum_{i=1}^{n^{2\alpha}} g(\theta_i^{n^\beta}, X_{T,i+1}^{n^\beta,\theta_i^{n^\beta}}, W_{T,i+1}) - \mathbb{E}\psi(X_T^{n^\beta}) \right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma^2\right).$$

Proof. At first, we rewrite the total error as follows

$$\frac{1}{n^{2\alpha}}\sum_{i=1}^{n^{2\alpha}} g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - \mathbb{E}\psi(X_T) =$$
$$\frac{1}{n^{2\alpha}}\sum_{i=1}^{n^{2\alpha}} \left(g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - \mathbb{E}g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1})\right) + \mathbb{E}\psi(X_T^n) - \mathbb{E}\psi(X_T).$$

Note that $\tilde{\mathbb{E}}g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) = \tilde{\mathbb{E}}\left(\tilde{\mathbb{E}}\left(g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1})|\tilde{\mathfrak{F}}_{T,i}\right)\right) = \mathbb{E}\psi(X_T^n)$. Assumption $(\mathcal{H}_{\varepsilon_n})$ ensures that $n^\alpha(\mathbb{E}\psi(X_T^n) - \mathbb{E}\psi(X_T)) \longrightarrow C_\psi$ as $n \to \infty$. Consequently, it remains to study the asymptotic behavior of the martingale array $(M_k^n)_{k\geq 1}$ given by

$$M_k^n := \frac{1}{n^\alpha}\sum_{i=1}^{k}\left(g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - \mathbb{E}g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1})\right).$$

To do so, we split the proof into two steps devoted to apply the central limit theorem for martingales array (see Theorem 4.1 and comments their).
Step 1. We need first to study the asymptotic behavior of the quadratic variation of the martingale array $(M_k^n)_{k\geq 1}$ given by

$$\langle M\rangle_{n^{2\alpha}}^n = \frac{1}{n^{2\alpha}}\sum_{i=1}^{n^{2\alpha}}\tilde{\mathbb{E}}\left[\left(g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - \tilde{\mathbb{E}}g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1})\right)^2|\tilde{\mathfrak{F}}_{T,i}\right].$$

Since θ_i^n is $\tilde{\mathfrak{F}}_{T,i}$-measurable and $W_{T,i+1} \perp\!\!\!\perp \tilde{\mathfrak{F}}_{T,i}$, we obtain easily that

$$\langle M\rangle_{n^{2\alpha}}^n = \frac{1}{n^{2\alpha}}\sum_{i=1}^{n^{2\alpha}}\nu_n(\theta_i^n) - \left[\mathbb{E}\psi(X_T^n)\right]^2, \tag{II.36}$$

where for all $\theta \in \mathbb{R}^q$

$$\nu_n(\theta) := \mathbb{E}\left(\psi(X_T^{n,\theta})^2 e^{-2\theta.W_T - |\theta|^2 T}\right) = \mathbb{E}\left(\psi(X_T^n)^2 e^{-\theta\cdot W_T + \frac{1}{2}|\theta|^2 T}\right).$$

It is clear that by assumption $(\mathcal{H}_{\varepsilon_n})$, the last term on the right hand side of the relation (II.36) converges to $\left[\mathbb{E}\psi(X_T)\right]^2$, as n tends to infinity. Concerning the first term, we introduce $\nu(\theta) := \mathbb{E}\left(\psi(X_T)^2 e^{-\theta\cdot W_T + \frac{1}{2}|\theta|^2 T}\right)$ and we get for all $\theta \in \mathbb{R}^q$

$$|\nu_n(\theta) - \nu(\theta)| \leq \mathbb{E}\left(\left|\psi(X_T^n)^2 - \psi(X_T)^2\right| e^{-\theta\cdot W_T + \frac{1}{2}|\theta|^2 T}\right) \leq e^{\frac{3}{2}|\theta|^2 T}\|\psi(X_T^n)^2 - \psi(X_T)^2\|_2.$$

Under the condition on ψ together with property $(\mathcal{P})$, there exists $C > 0$ such that

$$|\nu_n(\theta) - \nu(\theta)| \le \frac{C}{\sqrt{n}} e^{\frac{3}{2}|\theta|^2 T}, \quad \forall \theta \in \mathbb{R}^q.$$

By similar calculations, we check easily the equicontinuity of the family functions $(\nu_n)_{n\ge1}$ and we deduce thanks to property $(\mathcal{H}_\theta)$

$$\lim_{i,n\to\infty} \nu_n(\theta_i^n) = \nu(\theta^*) \quad \tilde{\mathbb{P}}\text{-}a.s.$$

Therefore, Lemma 4.3 applies and we deduce that $\langle M\rangle^n_{n^{2\alpha}} \underset{n\to\infty}{\longrightarrow} \sigma^2$.

Step 2. We will check now the Lyapunov condition, that is assumption *A3.*, which implies the Lindeberg condition *A2.* Let $a > 1$, we have

$$\begin{aligned}
\sum_{i=1}^{n^{2\alpha}} \tilde{\mathbb{E}}\left[|M_i^n - M_{i-1}^n|^{2a} | \tilde{\mathcal{F}}_{T,i-1}\right] &= \frac{1}{n^{2a\alpha}} \sum_{i=1}^{n^{2\alpha}} \tilde{\mathbb{E}}\left[\left|g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - \mathbb{E}\psi(X_T^n)\right|^{2a} |\tilde{\mathcal{F}}_{T,i}\right] \\
&\le \frac{2^{2a-1}}{n^{2a\alpha}} \sum_{i=1}^{n^{2\alpha}} \nu_{a,n}(\theta_i^n) + \frac{2^{2a-1}}{n^{2\alpha(a-1)}} [\mathbb{E}\psi(X_T^n)]^{2a}
\end{aligned}$$

where for all $\theta \in \mathbb{R}^q$, $\nu_{a,n}(\theta) = \mathbb{E}\left(\psi(X_T^n)^{2a} e^{-(2a-1)\theta\cdot W_T - (a-\frac{3}{2})|\theta|^2 T}\right)$. Following the same arguments detailed in the first step, we prove that

$$\frac{1}{n^{2\alpha}} \sum_{i=1}^{n^{2\alpha}} \nu_{a,n}(\theta_i^n) \underset{n\to\infty}{\longrightarrow} \nu_a(\theta^*) \quad \tilde{\mathbb{P}}\text{-}a.s.$$

where for all $\theta \in \mathbb{R}^q$, $\nu_a(\theta) = \mathbb{E}\left(\psi(X_T)^{2a} e^{-(2a-1)\theta\cdot W_T - (a-\frac{3}{2})|\theta|^2 T}\right)$. The second assertion is easily obtained following the above proof with $\alpha, \beta > 0$. This completes the proof. □

Remark 4.5. *If one have in mind to reduce the variance by using an adaptive crude Monte Carlo method, it appears clear that the natural choice is*

$$\theta^* = \arg\min_{\theta\in\mathbb{R}^q} \tilde{\mathbb{E}}\left(g^2(\theta, X_T)\right) \quad \text{and} \quad \theta_n^* = \arg\min_{\theta\in\mathbb{R}^q} \tilde{\mathbb{E}}\left(g^2(\theta, X_T^n)\right) \text{ for } n \ge 1.$$

Under suitable conditions on ψ, b and σ, one can of course construct sequences $(\theta_i^n)_{i\ge0}$, $n \in \mathbb{N}$ and $(\theta_i)_{i\ge0}$ satisfying $(\mathcal{H}_\theta)$ by either the constrained or the unconstrained Robbins-Monro algorithm.

4.3 The adaptive statistical Romberg method

As we pointed out at the beginning of the above subsection, the statistical Romberg algorithm (II.35) consists of two empirical means. So our task now is to study the asymptotic behavior of the second one in view to establish a central limit theorem for the method.

Theorem 4.6. *Let $(\theta_i^n)_{i\geq 0}$, $n \in \mathbb{N}$ and $(\theta_i)_{i\geq 0}$ be a family of sequences satisfying $(\mathcal{H}_\theta)$. Moreover, assume that b and σ are $\mathcal{C}^1$ functions satisfying the global Lipschitz condition $(\mathcal{H}_{b,\sigma})$ and ψ is a real valued function satisfying assumptions $(\mathcal{H}_\psi)$, $(\mathcal{H}_{\varepsilon_n})$, with constants $\alpha \in (1/2, 1]$ and $C_\psi \in \mathbb{R}$, such that*

$$|\psi(x) - \psi(y)| \leq C(1 + |x|^p + |y|^p)|x - y|, \quad \text{for some } C, p > 0.$$

If we choose $N_1 = n^{2\alpha}$, $N_2 = n^{2\alpha-\beta}$ and $m = n^\beta$, $0 < \beta < 1$ then the statistical Romberg algorithm denoted by V_n in (II.35) satisfies

$$n^\alpha \left(V_n - \mathbb{E}\psi(X_T)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(C_\psi, \sigma^2 + \tilde{\sigma}^2\right) \qquad \text{as } n \to \infty,$$

where $\sigma^2 = \mathbb{E}\left[\psi(X_T)^2 e^{-\theta^*\cdot W_T - \frac{1}{2}|\theta^*|^2 T}\right] - \left[\mathbb{E}\psi(X_T)\right]^2$, $\tilde{\sigma}^2 := \tilde{\mathbb{E}}\left[\left[\nabla\psi(X_T)\cdot U_T\right]^2 e^{-\theta^*\cdot W_T - \frac{1}{2}|\theta^*|^2 T}\right]$ and U is the process introduced from the beginning by relation (II.3).

Proof. First of all, note that we can rewrite the normalized total error as follows

$$n^\alpha \left(V_n - \mathbb{E}\psi(X_T)\right) := A_1^n + A_2^n$$

with $A_1^n := n^\alpha \left(V_n - \mathbb{E}\psi(X_T^{n^\beta}) - \mathbb{E}[\psi(X_T^n) - \psi(X_T^{n^\beta})]\right)$, and $A_2^n := n^\alpha \left(\mathbb{E}[\psi(X_T^n) - \psi(X_T)]\right)$. So, assumption $(\mathcal{H}_{\varepsilon_n})$ yields the convergence of the second term A_2^n towards the discretization constant C_ψ, as n tends to infinity. The first term A_1^n can be also rewritten as follows $A_1^n := A_{1,1}^n + A_{1,2}^n$, where

$$\begin{aligned}
A_{1,1}^n &:= \frac{1}{n^\alpha}\sum_{i=1}^{n^{2\alpha}} \left(g(\hat{\theta}_i^{n^\beta}, \hat{X}_{T,i+1}^{n^\beta,\hat{\theta}_i^{n^\beta}}, W_{T,i+1}) - \mathbb{E}\psi(X_T^{n^\beta})\right),\\
A_{1,2}^n &:= \frac{1}{n^{\alpha-\beta}}\sum_{i=1}^{n^{2\alpha-\beta}} \left(g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - g(\theta_i^n, X_{T,i+1}^{n^\beta,\theta_i^n}, W_{T,i+1}) - \mathbb{E}[\psi(X_T^n) - \psi(X_T^{n^\beta})]\right).
\end{aligned}$$

Using the independence between $A_{1,1}^n$ and $A_{1,2}^n$, we study separately their asymptotic behavior. Concerning the first term, the second assertion in Theorem 4.4 applies and gives the asymptotic normality of $A^n_{1,1}$,

$$A^n_{1,1} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\sigma^2\right), \qquad \text{as } n\to\infty. \tag{II.37}$$

Now, concerning the second term $A^n_{1,2}$ we introduce the martingale arrays $(M^n_k)_{k\geq 1}$

$$M^n_k := \frac{1}{n^{\alpha-\beta}}\sum_{i=1}^{k}\left(g(\theta^n_i, X^{n,\theta^n_i}_{T,i+1}, W_{T,i+1}) - g(\theta^n_i, X^{n^\beta,\theta^n_i}_{T,i+1}, W_{T,i+1}) - \mathbb{E}[\psi(X^n_T)-\psi(X^{n^\beta}_T)]\right),$$

in view to apply Theorem 4.1. To do so, we will verify both assumptions *A1.* and *A3.* in the following two steps.

• **Step 1.** The quadratic variation of M evaluated at $n^{2\alpha-\beta}$ is given by

$$\langle M\rangle^n_{n^{2\alpha-\beta}} = \frac{1}{n^{2\alpha-\beta}}\sum_{i=1}^{n^{2\alpha-\beta}} n^\beta \xi_n(\theta^n_i) - \left(n^{\frac{\beta}{2}}[\mathbb{E}\psi(X^n_T) - \mathbb{E}\psi(X^{n^\beta}_T)]\right)^2, \tag{II.38}$$

where $\forall\theta\in\mathbb{R}^q$, $\xi_n(\theta) := \mathbb{E}\left([\psi(X^n_T)-\psi(X^{n^\beta}_T)]^2 e^{-\theta\cdot W_T+\frac{1}{2}|\theta|^2 T}\right)$. Now, assumption $(\mathcal{H}_{\varepsilon_n})$ with $1/2<\alpha\leq 1$ ensures that the second term on the right hand side of relation (II.38) vanishes as n tends to infinity. We focus now on the asymptotic behavior of $n^\beta\xi_n(\theta)$. Under assumption $(\mathcal{H}_\psi)$, we apply the Taylor expansion theorem twice to get for all $\theta\in\mathbb{R}^q$

$$n^{\frac{\beta}{2}}[\psi(X^n_T)-\psi(X^{n^\beta}_T)]e^{-\frac{1}{2}\theta\cdot W_T+\frac{1}{4}|\theta|^2 T} = n^{\frac{\beta}{2}}\nabla\psi(X_T)\cdot[X^n_T - X^{n^\beta}_T]e^{-\frac{1}{2}\theta\cdot W_T+\frac{1}{4}|\theta|^2T} + R_n,$$

where

$$R_n := n^{\frac{\beta}{2}}(X^n_T - X_T)\varepsilon(X_T, X^n_T - X_T) - n^{\frac{\beta}{2}}(X^{n^\beta}_T - X_T)\varepsilon(X_T, X^{n^\beta}_T - X_T)$$

with $\varepsilon(X_T, X^n_T - X_T)\xrightarrow{\mathbb{P}\text{-}a.s.} 0$ and $\varepsilon(X_T, X^{n^\beta}_T - X_T)\xrightarrow{\mathbb{P}\text{-}a.s.} 0$ as $n\to\infty$, since the global Lipschitz condition $(\mathcal{H}_{b,\sigma})$ is satisfied. Further, as b and σ are $\mathscr{C}^1$ functions then according to Theorem 3.2 in [35] we have the tightness of $n^{\frac{\beta}{2}}(X^n_T - X_T)$ and $n^{\frac{\beta}{2}}(X^{n^\beta}_T - X_T)$ and we deduce the convergence in probability of the remaining term R_n to zero as n tends to infinity. Once again, by the same theorem in [35], we get for all $\theta\in\mathbb{R}^q$

$$n^{\frac{\beta}{2}}[\psi(X^n_T)-\psi(X^{n^\beta}_T)]e^{-\frac{1}{2}\theta\cdot W_T+\frac{1}{4}|\theta|^2T} \overset{stably}{\Longrightarrow} \nabla\psi(X_T)\cdot U_T e^{-\frac{1}{2}\theta\cdot W_T+\frac{1}{4}|\theta|^2T}. \tag{II.39}$$

Otherwise, $\forall\theta\in\mathbb{R}^q$ and $a'>1$ we have by Cauchy-Schwarz inequality

$$\mathbb{E}\left|n^{\frac{\beta}{2}}[\psi(X^n_T)-\psi(X^{n^\beta}_T)]e^{-\frac{1}{2}\theta\cdot W_T+\frac{1}{4}|\theta|^2T}\right|^{2a'} \leq n^{\beta a'}\left[\mathbb{E}\left|\psi(X^n_T)-\psi(X^{n^\beta}_T)\right|^{4a'}\right]^{\frac{1}{2}} e^{\frac{a'(2a'+1)}{2}|\theta|^2T}.$$

Thanks to the assumption on ψ together with property $(\mathcal{P})$, we obtain

$$\sup_n \mathbb{E}\left|n^{\frac{\beta}{2}}[\psi(X_T^n)-\psi(X_T^{n^\beta})]e^{-\frac{1}{2}\theta\cdot W_T+\frac{1}{4}|\theta|^2T}\right|^{2a'}<\infty. \tag{II.40}$$

Hence, by the stable convergence obtained in (II.39) and the uniform integrability property given by (II.40) we deduce $\forall\theta\in\mathbb{R}^q$

$$\lim_{n\to\infty} n^\beta\xi_n(\theta)=\tilde{\mathbb{E}}\left([\nabla\psi(X_T)\cdot U_T]^2\,e^{-\theta\cdot W_T+\frac{1}{2}|\theta|^2T}\right):=\xi(\theta). \tag{II.41}$$

Using property $(\mathcal{P})$ with assumption on ψ, it is easy to check by standard evaluations the equicontinuity of the family functions $(n^\beta\xi_n)_{n\geq1}$. So under assumption $(\mathcal{H}_\theta)$, we get

$$\lim_{i,n\to\infty} n^\beta\xi_n(\theta_i^n)=\xi(\theta^*)\quad \tilde{\mathbb{P}}\text{-}a.s.$$

Then, Lemma 4.3 yields $\lim_{n\to\infty}\langle M\rangle^n_{n^{2\alpha-\beta}}=\xi(\theta^*)$, $\quad\tilde{\mathbb{P}}$-*a.s.*

• **Step 2.** The second step consists on checking Lyapunov assumption *A3.* Let $a>1$,

$$\sum_{i=1}^{n^{2\alpha-\beta}}\tilde{\mathbb{E}}\left[|M_i^n-M_{i-1}^n|^{2a}|\tilde{\mathcal{F}}_{T,i-1}\right]\leq\frac{2^{2a-1}}{n^{a(2\alpha-\beta)}}\sum_{i=1}^{n^{2\alpha-\beta}}n^{\beta a}\xi_{a,n}(\theta_i^n)+\frac{2^{2a-1}n^{\beta a}}{n^{(2\alpha-\beta)(a-1)}}|\mathbb{E}\psi(X_T^n)-\mathbb{E}\psi(X_T^{n^\beta})|^{2a}$$

where for all $\theta\in\mathbb{R}^q$, $\xi_{a,n}(\theta):=\mathbb{E}\left(|\psi(X_T^n)-\psi(X_T^{n^\beta})|^{2a}e^{-(2a-1)\theta\cdot W_T-(a-\frac{3}{2})|\theta|^2T}\right)$. Miming the same arguments used in the first step, we prove under assumption $(\mathcal{H}_\theta)$ using relations (II.39) and Lemma 4.3, that

$$\frac{1}{n^{2\alpha-\beta}}\sum_{i=1}^{n^{2\alpha-\beta}}n^{\beta a}\xi_{a,n}(\theta_i^n)\underset{n\to\infty}{\longrightarrow}\xi_a(\theta^*):=\tilde{\mathbb{E}}\left(|\nabla\psi(X_T)\cdot U_T|^{2a}e^{-(2a-1)\theta\cdot W_T-(a-\frac{3}{2})|\theta|^2T}\right),\ \tilde{\mathbb{P}}\text{-a.s.}$$

Consequently, since $a>1$, we conclude using assumption $(\mathcal{H}_{\varepsilon_n})$ that *A3.* holds. This gives the asymptotic normality of $A^n_{1,2,2}$ so that we have $A^n_{1,2}\xrightarrow{\mathcal{L}}\mathcal{N}(0,\tilde{\sigma}^2)$, as $n\to\infty$. This completes the proof. □

Remark 4.7. *We recall that for the adaptive statistical Romberg method the optimal choice of θ^* and θ_n^* is given respectively by relations (II.7) and (II.8). According to Corollary 3.3 (resp. Corollary 3.8), the sequences $(\theta_i^n)_{i\geq0}$, $n\in\mathbb{N}$ and $(\theta_i)_{i\geq0}$ obtained by the constrained Robbins-Monro algorithm (resp. the unconstrained Robbins-Monro algorithm) satisfy $(\mathcal{H}_\theta)$ under some regularity conditions on ψ, b and σ.*

Complexity analysis According to the main theorems of this section, we deduce that for a total error of order $1/n^\alpha$, $\alpha \in (1/2, 1]$, the minimal computational effort necessary to run the adaptive statistical Romberg algorithm is obtained for $N_1 = n^{2\alpha}$, $N_2 = n^{2\alpha-\beta}$ and $m = n^\beta$. This leads to a time complexity given by $C_{SR} = C \times (n^{2\alpha+\beta} + (n + n^\beta)n^{2\alpha-\beta})$, with $C > 0$. So the time complexity reaches its minimum for the optimal choice of $\beta = 1/2$. Hence, the optimal parameters to run the method are given by $m = \sqrt{n}$, $N_1 = n^{2\alpha}$ and $N_2 = n^{2\alpha-1/2}$. Then the optimal complexity of the adaptive statistical Romberg algorithm is given by $C_{SR} \simeq C \times n^{2\alpha+\frac{1}{2}}$. However, for the same error of order $1/n^\alpha$, the optimal complexity of the adaptive Monte Carlo algorithm is given by $C_{MC} = C \times (N \times n) = C \times n^{2\alpha+1}$. We conclude that the adaptive statistical Romberg method is more efficient in terms of time complexity.

5 Numerical results for the Heston model

Stochastic volatility models are increasingly important in practical derivatives pricing applications. In this section we show, throughout the problem of option pricing with a stochastic volatility model, the efficiency of the importance sampling statistical Romberg method compared to the importance sampling Monte Carlo one. The popular stochastic volatility model in finance is the Heston model introduced by Heston in [33] as solution to

$$\begin{cases} dS_t = rS_t dt + \sqrt{V_t} S_t dW_t^1 \\ dV_t = \kappa(\bar{v} - V_t)dt + \sigma\sqrt{V_t}\rho dW_t^1 + \sigma\sqrt{V_t}\sqrt{1-\rho^2}dW_t^2, \end{cases}$$

where W^1 and W^2 are two independent Brownian motions. Parameters κ, σ, $\bar{v}$ and r are strictly positive constants and $|\rho| \leq 1$. In this model, κ is the rate at which V_t reverts to $\bar{v}$, $\bar{v}$ is the long run average price variance, σ is the volatility of the variance, r is the interest rate and ρ is a correlation term.

5.1 European options

Our aim is to use the importance sampling method in order to reduce the variance when computing the price of an European option, with strike K, under the Heston model. The payoff of the option is $\psi(S_T) = (S_T - K)_+$ for European call option and $\psi(S_T) = (K - S_T)_+$ for European put option. Then, the price is $e^{-rT}\mathbb{E}\psi(S_T)$. After a density transformation, given by

Girsanov theorem, the price will be defined by :

$$e^{-rT}\mathbb{E}\left[g(\theta, S_T^{\theta})\right] = e^{-rT}\mathbb{E}\left[\psi(S_T^{\theta}) \quad e^{-\theta.W_T - \frac{1}{2}|\theta|^2 T}\right], \qquad \theta \in \mathbb{R}^2.$$

For more details on definitions of the function g and S_T^{θ}, see relation (II.5) and related results given in the same page. To approximate S_T^{θ}, we consider the step T/n and we discretize the stochastic process using the Euler scheme. For $i \in [\![0, n-1]\!]$ and $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$,

$$\begin{cases} S_{t_{i+1}}^{n,\theta} = S_{t_i}^{n,\theta}\left(1 + (r + \theta_1\sqrt{V_{t_i}^{n,\theta}})\dfrac{T}{n} + \sqrt{V_{t_i}^{n,\theta}\dfrac{T}{n}}Z_{1,i+1}\right), \\ V_{t_{i+1}}^{n,\theta} = \left|V_{t_i}^{n,\theta} + \left(\kappa(\bar{v} - V_{t_i}^{n,\theta}) + \sigma\sqrt{V_{t_i}^{n,\theta}}(\rho\theta_1 + \sqrt{1-\rho^2}\theta_2)\right)\dfrac{T}{n} + \sigma\sqrt{V_{t_i}^{n,\theta}\dfrac{T}{n}}Z_{2,i+1}\right|, \end{cases}$$

with $(Z_{1,i}, Z_{2,i})_{1\leq i\leq n}$ is a sequence of a standard Gaussian random vectors taking values in $\mathbb{R}^2$. Hence, the price of the European call option is firstly approximated by

$$e^{-rT}\mathbb{E}\left[g(\theta, S_T^{n,\theta})\right] = e^{-rT}\mathbb{E}\left[\psi(S_T^{n,\theta}) \quad e^{-\theta.W_T - \frac{1}{2}|\theta|^2 T}\right], \qquad \theta \in \mathbb{R}^2.$$

The choice of θ depends on using the classical Monte Carlo method or the statistical Romberg one. The optimal θ for the first method is given by

$$\theta_n^* = \underset{\theta\in\mathbb{R}^2}{\arg\min}\mathbb{E}\left[\psi^2(S_T^{n,\theta}) \quad e^{-2\theta.W_T - |\theta|^2 T}\right].$$

However, The optimal θ for the second one is

$$\tilde{\theta}_n^* = \underset{\theta\in\mathbb{R}^2}{\arg\min}\mathbb{E}\left[\left(\psi^2(S_T^{n,\theta}) + (\nabla\psi(S_T^{n,\theta}).U_T^{n,\theta})^2\right) e^{-2\theta.W_T - |\theta|^2 T}\right],$$

where $U^{n,\theta}$ denotes the Euler discretization scheme obtained when we replace coefficients b and σ of relation (II.9) by the corresponding parameters in the Heston model. Here, we have also the choice of the algorithm approximating both θ_n^* and $\tilde{\theta}_n^*$. We can use either the constrained or the unconstrained stochastic algorithms studied in section 3 above.

- **Approximation of θ_n^* by**

  i. Constrained algorithm : let $(\mathcal{K}_i)_{i\in\mathbb{N}}$ denote an increasing sequence of compact sets satisfying $\cup_{i=0}^{\infty}\mathcal{K}_i = \mathbb{R}^d$ and $\mathcal{K}_i \subsetneq \overset{\circ}{\mathcal{K}}_{i+1}, \forall i \in \mathbb{N}$. For $\theta_0^n \in \mathcal{K}_0$, $\alpha_0^n = 0$ and a gain sequence $(\gamma_i)_{i\in\mathbb{N}}$

satisfying (II.22), we define the sequence $(\theta_i^n, \alpha_i^n)_{i\in\mathbb{N}}$ recursively by

$$\begin{cases} \textbf{if} & \theta_i^n - \gamma_{i+1} H(\theta_i^n, S_{T,i+1}^n, U_{T,i+1}^n, W_{T,i+1}) \in \mathcal{K}_{\alpha_i^n}, \ \textbf{then} \\ & \theta_{i+1}^n = \theta_i^n - \gamma_{i+1} H(\theta_i^n, S_{T,i+1}^n, U_{T,i+1}^n, W_{T,i+1}), \ \text{and } \alpha_{i+1}^n = \alpha_i^n \\ \textbf{else} & \theta_{i+1}^n = \theta_0^n \text{ and } \alpha_{i+1}^n = \alpha_i^n + 1, \end{cases} \tag{II.42}$$

where $H(\theta_i^n, S_{T,i+1}^n, W_{T,i}) = (\theta_i^n T - W_{T,i+1}) \ \ \psi^2(S_{T,i+1}^n) \ \ e^{-\theta_i^n . W_{T,i+1} + \frac{1}{2}|\theta_i^n|^2 T}$.

ii. Unconstrained algorithm : $\theta_{i+1}^n = \theta_i^n - \gamma_{i+1}(2\theta_i^n T - W_{T,i+1})\psi^2(S_{T,i+1}^{n,-\theta_i^n}) e^{-\eta|\theta_i^n|^2 T}$, with $\eta > 0$.

- **Approximation of $\tilde{\theta}_n^*$ by**

i. Constrained algorithm : we use the same routine (II.42) with

$$H(\theta_i^n, S_{T,i+1}^n, W_{T,i}) = (\theta_i^n T - W_{T,i+1}) \left(\psi^2(S_{T,i+1}^n) + (\nabla\psi(S_{T,i+1}^n).U_{T,i+1}^n)^2\right) e^{-\theta_i^n . W_{T,i+1} + \frac{1}{2}|\theta_i^n|^2 T}.$$

ii. Unconstrained algorithm : we use the routine

$$\theta_{i+1}^n = \theta_i^n - \gamma_{i+1}(2\theta_i^n T - W_{T,i+1}) \left(\psi^2(S_{T,i+1}^{n,-\theta_i^n}) + (\nabla\psi(S_{T,i+1}^{n,-\theta_i^n}) \cdot U_{T,i+1}^n)^2\right) e^{-\eta|\theta_i^n|^2 T}.$$

To compare these different routines we run a number of iterations $M = 500\,000$. The parameters in the Heston model are chosen as follows : $S_0 = 100$, $V_0 = 0.01$, $K = 100$, the free interest rate $r = \log(1.1)$, $\sigma = 0.2$, $k = 2$, $\bar{v} = 0.01$, $\rho = 0.5$ and maturity time $T = 1$. Table II.1 (respecyively Table II.2) gives the obtained values of the two-dimensional vectors θ_n^* and $\tilde{\theta}_n^*$ for the European call option (respectively for the European put option).

	Constrained algorithm	**Unconstrained algorithm**
θ_n^*	$(0.7906, 0.0516)$	$(0.7904, 0.0532)$
$\tilde{\theta}_n^*$	$(0.7884, 0.0587)$	$(0.7898, 0.0576)$

Tableau II.1 – *Estimation of θ_n^* and $\tilde{\theta}_n^*$ for European call option*

	Constrained algorithm	**Unconstrained algorithm**
θ_n^*	$(-1.1908, 0.6707)$	$(-0.9574, 0.4742)$
$\tilde{\theta}_n^*$	$(-1.1116, 0.5760)$	$(-0.9482, 0.4463)$

Tableau II.2 – *Estimation of θ_n^* and $\tilde{\theta}_n^*$ for European put option*

In Figure II.1, we test the robustness of both routines, for the computation of $\tilde{\theta}_n^*$ for the European Call option, using the averaged algorithm “à la Ruppert & Poliak" (see e.g. [54]) known to give optimal rate for convergence. We implement this averaged algorithm using both constrained and unconstrained procedures. So, we proceed as follows,

i. first, we choose a slowly decreasing step : $\gamma_i = \gamma_0/i^\alpha$, for $\alpha \in (\frac{1}{2}, 1)$ and $\gamma_0 > 0$.

ii. Then, we compute the empirical mean of all the previous observations,

$$\bar{\theta}_{i+1}^n := \frac{1}{i+1} \sum_{k=0}^{i} \tilde{\theta}_k^n.$$

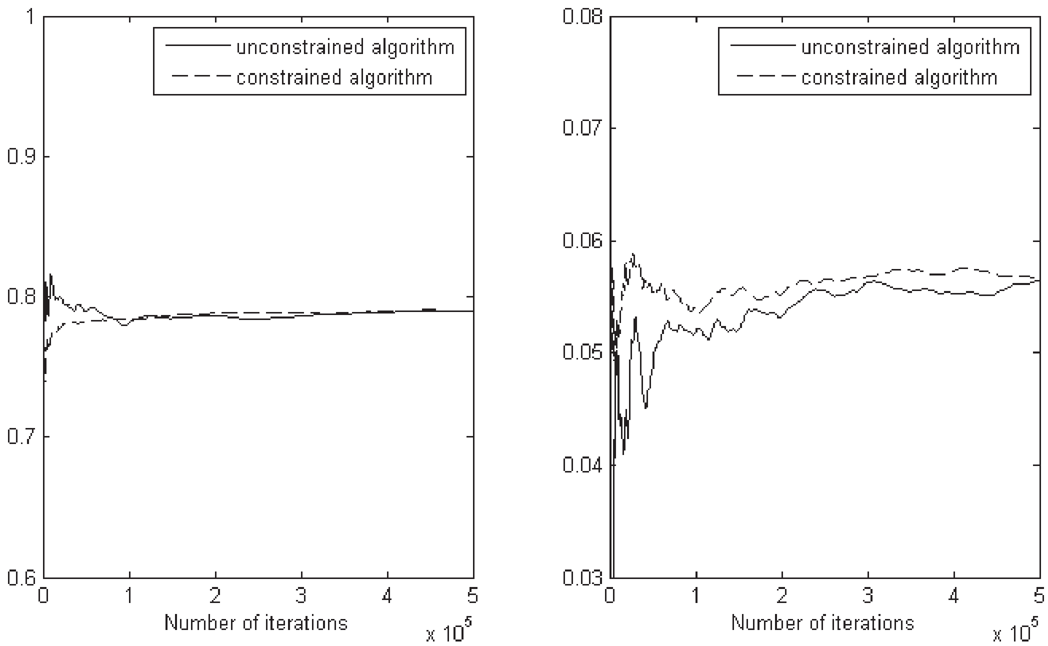

Figure II.1 – *Values of* $(\bar{\theta}_i^n)_{1\leq i\leq M}$ *obtained with* $n = 100$, $\gamma_0 = 0.01$ *and* $\alpha = 0.75$.

The left curve (resp. the right curve) is the representation, for $1 \leq i \leq M$, of the first component (resp. the second component) of the two-dimensional vector $\bar{\theta}_i^n$. The same results can be obtained for the European put option. The Figure II.5 illustrates the trajectories of the two components of the two-dimensional vector $\bar{\theta}_i^n$ using the constrained and unconstrained algorithms.

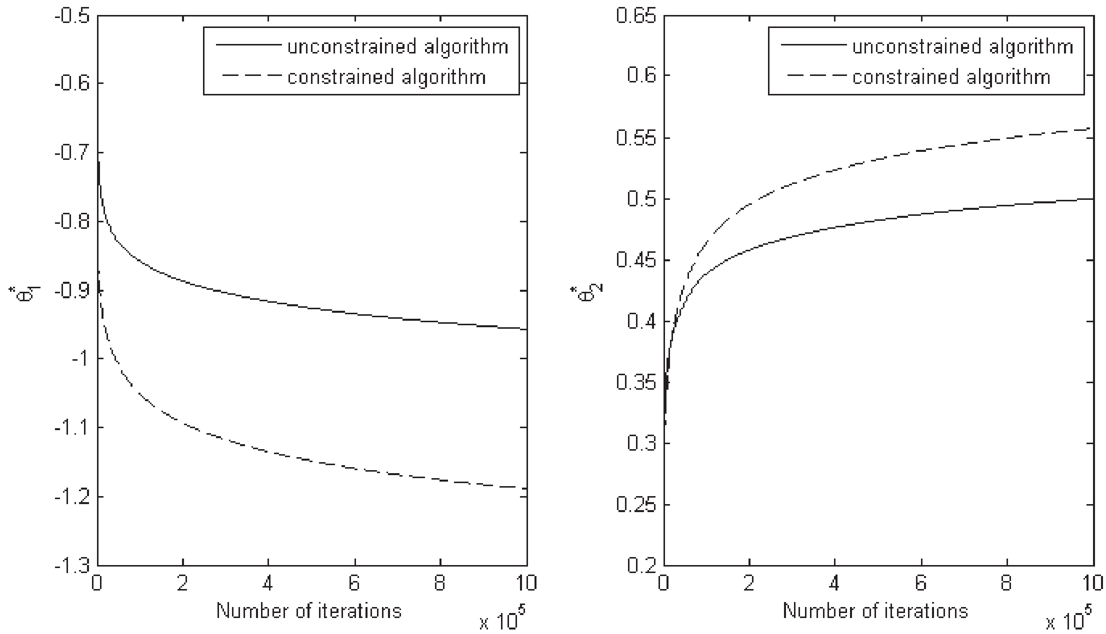

Figure II.2 – *Values of* $(\bar{\theta}_i^n)_{1\leq i\leq M}$ *obtained with* $n = 100$, $\gamma_0 = 0.01$ *and* $\alpha = 0.75$.

The trajectories obtained using the constrained or the unconstrained algorithm are comparable. Consequently, since we did not notice any major difference between the two methods we have chosen to only use the constrained algorithm for approximating θ_n^* (resp. $\tilde{\theta}_n^*$) by θ_M^n (resp $\tilde{\theta}_M^n$). Our aim now, is to compare both importance sampling Monte Carlo method (denoted by MC+IS) and importance sampling statistical Romberg (denoted by SR+IS).

- MC+IS method : European option price approximation with $N = n^2$

$$\frac{e^{-rT}}{N}\sum_{i=1}^{N} g(\theta_M^n, S_{T,i+1}^{n,\theta_M^n}) = \frac{e^{-rT}}{N}\sum_{i=1}^{N}\psi(S_{T,i+1}^{n,\theta_M^n})e^{-\theta_M^n.W_{T,i+1}-\frac{1}{2}|\theta_M^n|^2 T}. \tag{II.43}$$

- SR+IS method : European option price approximation method with $N_1 = n^2$ and $N_2 = n^{\frac{3}{2}}$

$$\frac{e^{-rT}}{N_1}\sum_{i=1}^{N_1} g(\tilde{\theta}_M^n, \hat{S}_{T,i+1}^{\sqrt{n},\tilde{\theta}_M^n}) + \frac{e^{-rT}}{N_2}\sum_{i=1}^{N_2}\left(g(\tilde{\theta}_M^n, S_{T,i+1}^{n,\tilde{\theta}_M^n}) - g(\tilde{\theta}_M^n, S_{T,i+1}^{\sqrt{n},\tilde{\theta}_M^n})\right). \tag{II.44}$$

The first method (II.43) is already implemented and available in the free online version of Premia platform (`https://www.rocq.inria.fr/mathfi/Premia/index.html`) and our method (II.44) is now added in the latest premium version. In Table II.3 (resp. Table II.4), we compare for each given number of time step n, the obtained European call price (resp. the sensitivity European call price parameter Δ) with the corresponding length of the 95%-confidence interval and the CPU time (per second) for both methods (II.43) and (II.44). It is worth to note that the number of time step n needed to achieve a given accuracy depends on the choice of the method.

Method	n	**Price**	**Confidence**	**time**
		Interval length		
	400	9.641444	0.060094	10.38
MC+IS	900	9.661192	0.029409	91.5
	1600	9.656892	0.016538	512.29
	600	9.659409	0.057454	3.36
SR+IS	1600	9.660062	0.019933	26.79
	3600	9.65673	0.008584	194.6

Tableau II.3 – *Call Price for the Heston model*

Method	n	**Price**	**Confidence**	**time**
		Interval length		
	400	0.863968	0.00721	9.39
MC+IS	900	0.863291	0.003151	91.58
	1600	0.863766	0.001774	515.31
	600	0.867441	0.007249	3.27
SR+IS	1600	0.864213	0.002541	27.02
	3600	0.862589	0.001095	202.2

Tableau II.4 – *Delta call price for the Heston model*

In Table II.5 (resp. Table II.6), we compare for each given number of time step n, the obtained European put price (resp. the sensitivity European call price parameter Δ) with the corresponding length of the 95%-confidence interval and the CPU time (per second) for both methods (II.43) and (II.44). It is worth to note that the number of time step n needed to achieve a given accuracy depends on the choice of the method.

Method	n	Price Interval length	Confidence	time
MC+IS	400	0.5712	0.0073	7.98
	900	0.5693	0.0032	95.08
	1600	0.5686	0.0018	698.23
SR+IS	625	0.5719	0.0089	2.46
	1600	0.5659	0.0032	25.65
	3600	0.5666	0.0014	243.59

Tableau II.5 – *Put Price for the Heston model*

Method	n	Price Interval length	Confidence	time
MC+IS	400	−0.1369	0.0021	7.98
	900	−0.137	0.0009	91.08
	1600	−0.1364	0.0005	698.23
SR+IS	625	−0.1369	0.003	2.46
	1600	−0.1361	0.0014	25.65
	3600	−0.1360	0.00047	243.59

Tableau II.6 – *Delta put price for the Heston model*

We also compare both methods (II.43) and (II.44) for a large range of time step numbers n. Then, we make a simple log-log scale plot of CPU time versus the corresponding 95%-confidence interval length. Computations are done on a PC with a 2.5 GHz Intel core i5 processor. In Figure II.3, we represent in the left curve (respectively the right curve) the CPU time versus the 95%-confidence interval length for the European call option (respectively for the European put option). The line marked by circles denotes MC+IS method and the line marked by squares denotes SR+IS method. The values mentioned near the points correspond to the chosen number of steps n. Clearly, the SR+IS curve is lower than the MC+IS one for both curves, which means that MC+IS method spends more time than SR+IS method to achieve the same given error when computing the call option price and put option price. For example for an error of 0.06, the SR+IS method reduces time by a factor of 3.33 compared to a MC+IS one when computing European call option price. Note that, the more the imposed error is small, the better improvement is. For example for a small error 0.02, the time reduction exceeds a factor of 10 when computing European call option price. For the European put option, for an error of 0.009, the SR+IS method reduces time by a factor of 3 compared to a MC+IS. For a small error 0.002, the time reduction exceeds a factor of 7.5.

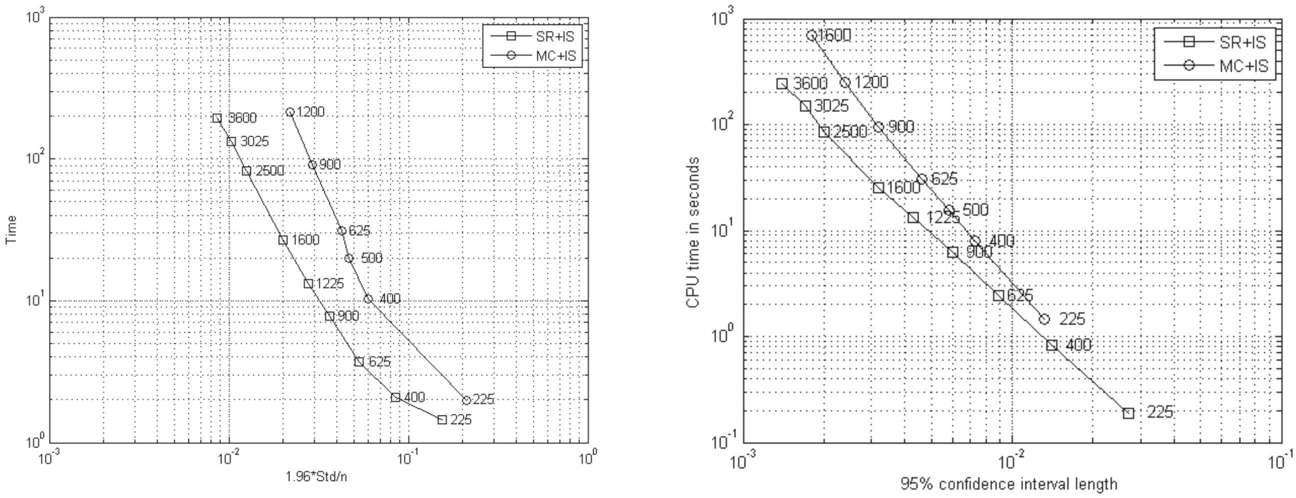

Figure II.3 – CPU time versus the 95%-confidence interval length for European options

5.2 Asian option

Now, our goal is to use the importance sampling method in order to reduce the variance when computing the price of an Asian option, with strike K, under the Heston model. The payoff of the option is $\psi(S_t, t \leq T) = (\frac{1}{T}\int_0^T S_t dt - K)_+$ for Asian call option and $\psi(S_t, t \leq T) = (K - \frac{1}{T}\int_0^T S_t dt)_+$ for European put option. Then, the price is $e^{-rT}\mathbb{E}\psi(S_T)$. Then the approximate price of the option is given by $e^{-rT}\mathbb{E}\left[G(S)\right]$. The function G is computed by using the discretization of the mean $A(T,S) = \frac{1}{T}\int_0^T S_t dt$, then it will be given by $G(S) = (\hat{A}(T,S) - K)_+$ for Asian call option and $G(S) = (K - \hat{A}(T,S))_+$ for Asian put option. After a density transformation, given by Girsanov theorem, the price will be defined by :

$$e^{-rT}\mathbb{E}\left[g(\theta, S_T^\theta)\right] = e^{-rT}\mathbb{E}\left[G(S_T^\theta) \quad e^{-\theta.W_T - \frac{1}{2}|\theta|^2 T}\right], \qquad \theta \in \mathbb{R}^2.$$

We approximate S_t^θ in the same way as in the previous section and the Asian option price is approximated by

$$e^{-rT}\mathbb{E}\left[g(\theta, S_T^{n,\theta})\right] = e^{-rT}\mathbb{E}\left[G(S_T^{n,\theta}) \quad e^{-\theta.W_T - \frac{1}{2}|\theta|^2 T}\right], \qquad \theta \in \mathbb{R}^2.$$

The optimal θ for the first method is given by

$$\theta_n^* = \arg\min_{\theta \in \mathbb{R}^2} \mathbb{E}\left[G^2(S_T^{n,\theta}) \quad e^{-2\theta.W_T - |\theta|^2 T}\right].$$

However, The optimal θ for the second one is

$$\tilde{\theta}_n^* = \arg\min_{\theta \in \mathbb{R}^2} \mathbb{E}\left[\left(G^2(S_T^{n,\theta}) + (\nabla G(S_T^{n,\theta}).U_T^{n,\theta})^2\right) e^{-2\theta.W_T - |\theta|^2 T}\right],$$

where $U^{n,\theta}$ denotes the Euler discretization scheme obtained when we replace coefficients b and σ of relation (II.9) by the corresponding parameters in the Heston model.
Now,we implement the averaged algorithm "à la Ruppert & Poliak" given by 5.1 to test the robustness of both routines constrained and unconstrained algorithms.

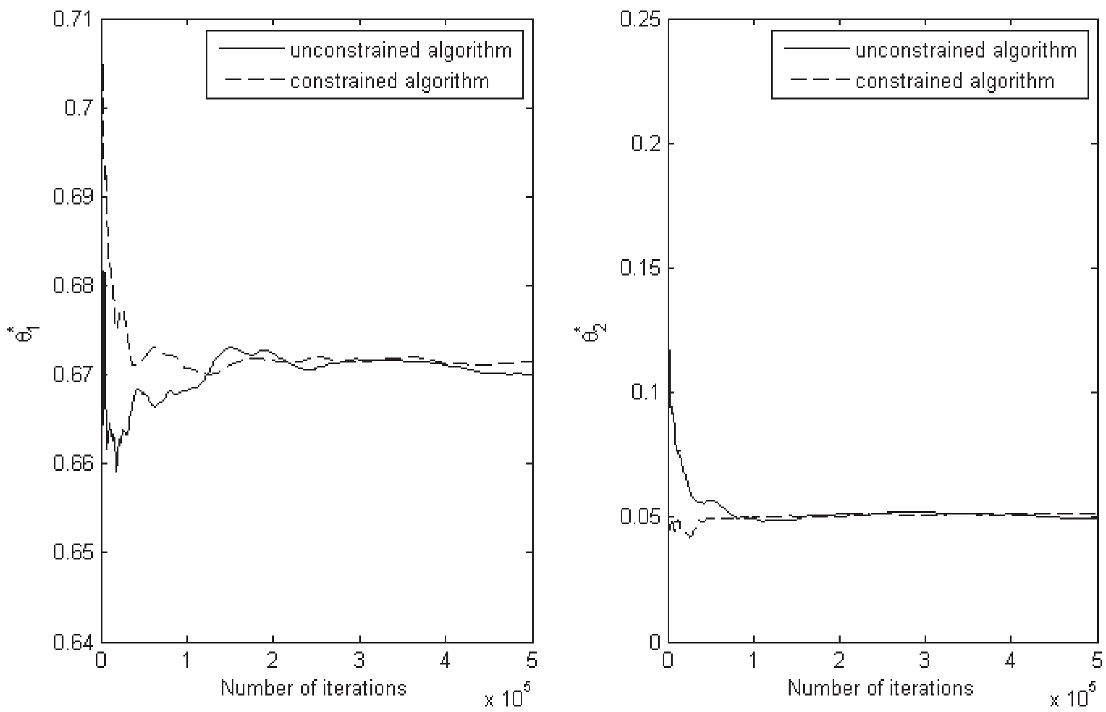

Figure II.4 – *Values of* $(\bar{\theta}_i^n)_{1\leq i\leq M}$ *obtained with* $n = 100$, $\gamma_0 = 0.01$ *and* $\alpha = 0.75$.

The left curve (resp. the right curve) is the representation, for $1 \leq i \leq M$, of the first component (resp. the second component) of the two-dimensional vector $\bar{\theta}_i^n$. The same results can be obtained for the asiatic put option. The Figure II.4 illustrates the trajectories of the two components of the two-dimensional vector $\bar{\theta}_i^n$ using the constrained and unconstrained algorithms.

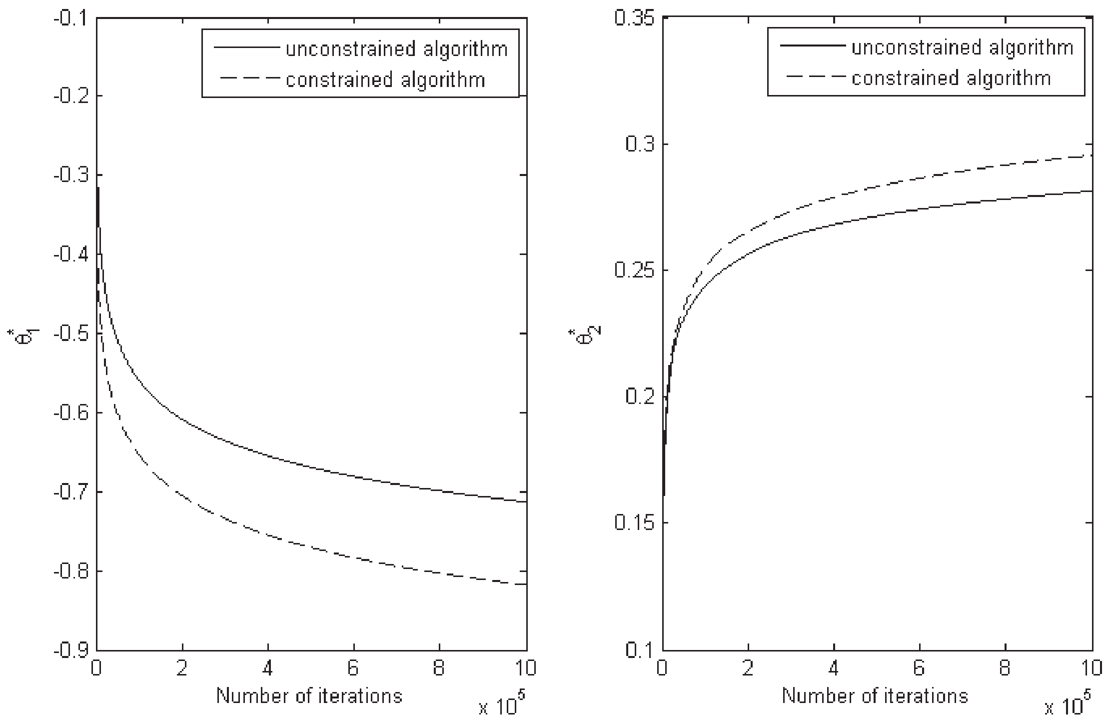

Figure II.5 – *Values of* $(\bar{\theta}_i^n)_{1\leq i\leq M}$ *obtained with* $n = 100$*,* $\gamma_0 = 0.01$ *and* $\alpha = 0.75$*.*

Here, we use the constrained algorithm to approximate the optimal θ for the classical Monte Carlo method and the Statistical Romberg one.

	Constrained algorithm
θ_n^*	$(0.6734, 0.0511)$
$\tilde{\theta}_n^*$	$(0.6685, 0.0471)$

Tableau II.7 – *Estimation of* θ_n^* *and* $\tilde{\theta}_n^*$ *for Asian call option*

	Constrained algorithm
θ_n^*	$(-0.8136, 0.3079)$
$\tilde{\theta}_n^*$	$(-0.7064, 0.2762)$

Tableau II.8 – *Estimation of* θ_n^* *and* $\tilde{\theta}_n^*$ *for Asian put option*

Our aim now is to compare both importance sampling Monte Carlo method (MC+IS) and importance sampling statistical Romberg method (SR+IS). In Table II.9 (resp. Table II.10), we compare for each given number of time step n, the obtained Asian call price (resp. the sensitivity Asian call price parameter Δ) with the corresponding length of the 95%-confidence interval and the CPU time (per second) for both methods (II.43) and (II.44). It is worth to note that the number of time step n needed to achieve a given accuracy depends on the choice of the method.

Method	n	**Price** **Interval length**	**Confidence**	**time**
MC+IS	400	4.9435	0.041	8.02
	900	4.94	0.0182	127.24
	1600	4.9325	0.0102	860.06
SR+IS	625	4.9506	0.038	2.68
	1600	4.9332	0.0146	31.58
	3600	4.9374	0.0064	293.13

Tableau II.9 – *Asian Call Price for the Heston model*

Method	n	**Price** **Interval length**	**Confidence**	**time**
MC+IS	400	0.7746	0.0062	8.02
	900	0.7743	0.0028	127.24
	1600	0.7737	0.0016	860.06
SR+IS	625	0.7733	0.0076	2.68
	1600	0.7741	0.0029	31.58
	3600	0.7742	0.0012	293.13

Tableau II.10 – *Delta asiatic call price for the Heston model*

In Table II.11 (resp. Table II.12), we compare for each given number of time step n, the obtained Asian put price (resp. the sensitivity Asian call price parameter Δ) with the corresponding length of the 95%-confidence interval and the CPU time (per second) for both methods (II.43) and (II.44). It is worth to note that the number of time step n needed to achieve a given accuracy depends on the choice of the method.

Method	n	Price Interval length	Confidence	time
MC+IS	400	0.459	0.0074	9.27
	900	0.4595	0.0033	107.98
	1600	0.4592	0.0018	872.08
SR+IS	625	0.4648	0.0077	2.53
	1600	0.4607	0.0029	31.74
	3600	0.4582	0.0012	331.07

Tableau II.11 – *Asian Put Price for the Heston model*

Method	n	Price Interval length	Confidence	time
MC+IS	400	-0.1797	0.0029	9.27
	900	-0.1796	0.0013	107.98
	1600	-0.1796	0.0007	872.08
SR+IS	625	-0.1803	0.0036	2.53
	1600	-0.1794	0.0013	31.74
	3600	-0.1793	0.0005	331.07

Tableau II.12 – *Delta asiatic put price for the Heston model*

We also compare both methods (II.43) and (II.44) for a large range of time step numbers n. Then, we make a simple log-log scale plot of CPU time versus the corresponding 95%-confidence interval length. Computations are done on a PC with a 2.5 GHz Intel core i5 processor. In Figure II.6, we represent in the left curve (respectively the right curve) the CPU time versus the 95%-confidence interval length for the Asian call option (respectively for the Asian put option). The line marked by circles denotes MC+IS method and the line marked by squares denotes SR+IS method. The values mentioned near the points correspond to the chosen number of steps n. Clearly, the SR+IS curve is lower than the MC+IS one for both curves, which means that MC+IS method spends more time than SR+IS method to achieve the same given error when computing the Asian call option price and put option price. For example for an error of 0.06, the SR+IS method reduces time by a factor of 3.125 compared to a MC+IS one when computing Asian call option price. Note that, the more the imposed error is small, the better improvement is. For example for a small error 0.01, the time reduction exceeds a factor of 11.5 when computing call option price. For the Asian put option, for an error of 0.01, the SR+IS method reduces time by a factor of 2.5 compared to a MC+IS. For a small error 0.002, the time reduction exceeds a factor of 8.75.

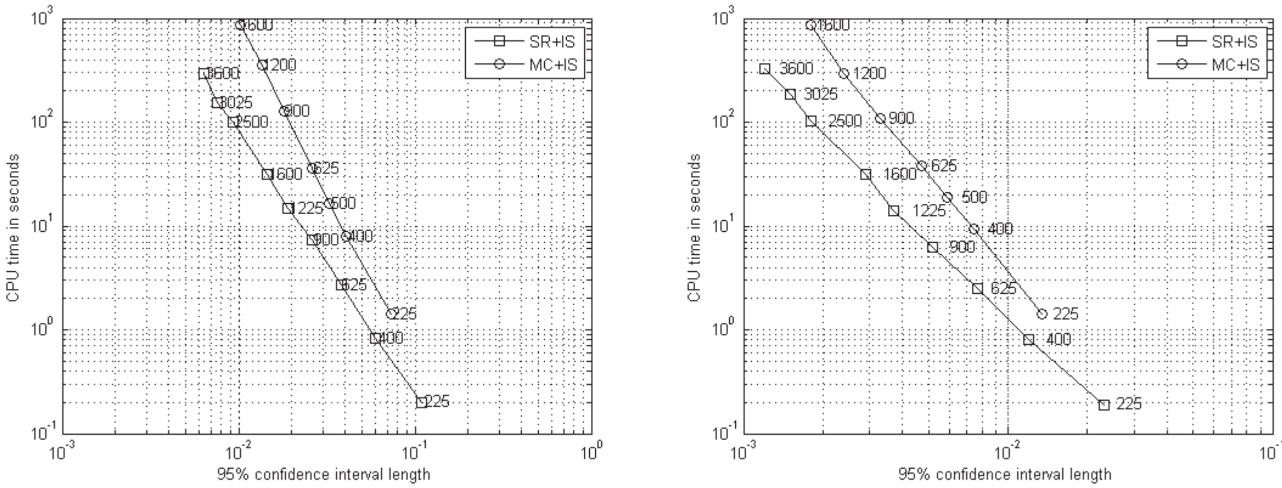

Figure II.6 – CPU time versus the 95%-confidence interval length for Asian options

Chapitre III

Importance Sampling and Euler Multilevel Monte Carlo

The aim of this chapter is to study a new algorithm based on an original combination of the Euler Multilevel Monte Carlo and the importance sampling technique. This algorithm will extend the work given in Chapter II of this thesis by using the Euler Multilevel Monte Carlo method instead of the Statistical Romberg one.

1 Introduction

We consider the same framework as the Chapter II. Recall that we are interested in estimating the expected payoff value $\mathbb{E}\psi(X_T)$ in option pricing problems, with $T > 0$ and $(X_t)_{0\leq t\leq T}$ is the process with values in $\mathbb{R}^d$ solution to the diffusion

$$dX_t = b(X_t)dt + \sum_{j=1}^{q} \sigma_j(X_t)dW_t^j, \quad X_0 = x \in \mathbb{R}^d, \tag{III.1}$$

where $W = (W^1, \ldots, W^q)$ is a q-dimensional Brownian motion on some given filtered probability space $\mathcal{B} = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and $(\mathcal{F}_t)_{t\geq 0}$ is the standard Brownian filtration. The functions $b : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ and $\sigma_j : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, $1 \leq j \leq q$, satisfy condition

$$(\mathcal{H}_{b,\sigma}) \quad \forall x, y \in \mathbb{R}^d \quad |b(x) - b(y)| + \sum_{j=1}^{q} |\sigma_j(x) - \sigma_j(y)| \leq C_{b,\sigma}|x - y|, \quad \text{with } C_{b,\sigma} > 0.$$

This ensures strong existence and uniqueness of solution of (III.1).
In view of reducing the variance in the estimation, we introduce the importance sampling

technique. Then, by applying the Girsanov theorem, we obtain for all $\theta \in \mathbb{R}^q$,

$$\mathbb{E}\psi(X_T) = \mathbb{E}g(\theta, X_T^\theta, W_T), \qquad \text{where } g : \mathbb{R}^q \times \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R},$$

where $(X_t^\theta)_{0\leq t\leq T}$ is solution to

$$dX_t^\theta = \left(b(X_t^\theta) + \sum_{j=1}^q \theta_j \sigma_j(X_t^\theta) \right) dt + \sum_{j=1}^q \sigma_j(X_t^\theta) dW_t^j. \tag{III.2}$$

Our goal now is to approximate this quantity by using the Multilevel Monte Carlo method with a given optimal parameter θ.

The Multilevel Monte Carlo method is popularized for financial applications by M. Giles in [28]. This method can be seen as a generalized version of the Statistical Romberg method introduced by Kebaier in [37]. It has been extensively applied to various fields of numerical probability (Brownian stochastic differential equations, Lévy-driven stochastic differential equations and more general numerical analysis problems (high dimensional parabolic SPDEs, etc). For more references, we refer to the website `https://people.maths.ox.ac.uk/gilesm/mlmc_community.html`. The idea of the Multilevel Monte Carlo method is to apply the Monte Carlo method for several nested levels of time step sizes and to compute different numbers of paths on each level, from a few paths when the time step size is small to many paths when the step size is large.

In practice, we consider the Euler continuous approximation X^n of the process X, with time step $\delta = T/n$ given by

$$dX_t^n = b(X_{\eta_n(t)})dt + \sum_{j=1}^q \sigma_j(X_{\eta_n(t)})dW_t, \quad \eta_n(t) = [t/\delta]\delta. \tag{III.3}$$

The Euler Multilevel Monte Carlo method uses information from a sequence of computations with decreasing step sizes and approximates the quantity $\mathbb{E}\psi(X_T)$ by

$$Q_n = \frac{1}{N_0}\sum_{i=1}^{N_0} \psi(X_{T,i}^{m^0}) + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left(\psi(X_{T,i}^{\ell,m^\ell}) - \psi(X_{T,i}^{\ell,m^{\ell-1}}) \right), \quad m \in \mathbb{N} \setminus \{0,1\}.$$

The time step sizes is defined by T/m^ℓ, $\ell = 0, 1, ..., L$ where $L = \frac{\log n}{\log m}$ and m is the refinement factor. For the first empirical mean, the random variables $(X_{T,i}^{m^0})_{1\leq i\leq N_0}$ are independent copies of $X_T^{m^0}$ which denotes the Euler scheme with time step T. For $\ell \in \{1, ..., L\}$, the couples $(X_{T,i}^{\ell,m^\ell}, X_{T,i}^{\ell,m^{\ell-1}})_{1\leq i\leq N_\ell}$ are independent copies of $(X_T^{\ell,m^\ell}, X_T^{\ell,m^{\ell-1}})$ whose components denote the

Euler schemes with tim e steps T/m^ℓ and $T/m^{\ell-1}$. However, for fixed ℓ, the simulation of X_T^{ℓ,m^ℓ} and $X_T^{\ell,m^{\ell-1}}$ has to be based on the same Brownian path. Giles [28] proved that the Multilevel Monte Carlo method reduces efficiently the computational complexity of the combination of Monte Carlo method and the Euler discretization scheme. In fact, the complexity in the Monte Carlo method is equal to $n^{2\alpha+1}$ and is reduced to $n^{2\alpha}(\log n)^2$ in the Multilevel Monte Carlo method where $\alpha \in [1/2, 1]$ is the order of the rate of convergence of the weak error given by $\varepsilon_n = \mathbb{E}\psi(X_T^n) - \mathbb{E}\psi(X_T)$. So, it is worth to introduce the following assumption

$$(\mathcal{H}_{\varepsilon_n}) \quad \text{for } \alpha \in [1/2, 1] \quad n^\alpha \varepsilon_n \to C_\psi, \quad C_\psi \in \mathbb{R}.$$

In their recent work, Ben Alaya and Kebaier [11], [12] proved a central limit theorem for the Euler Multilevel Monte Carlo Euler method with a rate of convergence equal to n^α, $\alpha \in [1/2, 1]$ (see Theorem 4 of [11]). At first, they assumed that the sample size N_ℓ is given by this relation

$$N_\ell = \frac{n^{2\alpha}(m-1)T}{m^\ell a_\ell} \sum_{\ell=1}^{L} a_\ell, \quad \ell \in \{0, ..., L\} \text{ and } L = \frac{\log n}{\log m}, \tag{III.4}$$

where $(a_\ell)_{\ell \in \mathbb{N}}$ is a real sequence of positive terms satisfying

$$(\mathcal{W}) \quad \lim_{L\to\infty} \sum_{\ell=1}^{L} a_\ell = \infty \text{ and } \lim_{L\to\infty} \frac{1}{(\sum_{\ell=1}^{L} a_\ell)^{p/2}} \sum_{\ell=1}^{L} a_\ell^{p/2} = 0, \text{ for } p > 2.$$

Then, under assumptions $(\mathcal{H}_{\varepsilon_n})$ and $(\mathcal{H}_\psi)$ which is given by

$$(\mathcal{H}_\psi) \quad \mathbb{P}(X_T \notin \mathcal{D}_{\dot\psi}) = 0, \text{ where } \mathcal{D}_{\dot\psi} := \{x \in \mathbb{R}^d \mid \psi \text{ is differentiable at } x\},$$

they proved that this method is tamed by a central limit theorem (see Theorem 1.9 in Chapter I). More precisely, the global error normalized by n^α converges in law to a Gaussian random variable with bias equal to C_ψ and a limit variance equal to

$$\tilde{\mathrm{V}}\mathrm{ar}\left(\nabla\psi(X_T) \cdot U_T\right), \tag{III.5}$$

where U is the weak limit process of the error $\sqrt{\frac{m^\ell}{(m-1)T}}(X^{m^\ell} - X^{m^{\ell-1}})$ defined on $\tilde{\mathcal{B}}$ an extension of the initial space $\mathcal{B}$ (see Theorem 3 in Ben Alaya and Kebaier [11]). The process U is solution to

$$dU_t = \dot b(X_t)U_t dt + \sum_{j=1}^{q} \dot\sigma_j(X_t)U_t dW_t^j - \frac{1}{\sqrt{2}} \sum_{j,\ell=1}^{q} \dot\sigma_j(X_t)\sigma_\ell(X_t) d\tilde W_t^{\ell j}, \tag{III.6}$$

where $\tilde{W}$ is a q^2-dimensional standard Brownian motion, defined on the extension $\tilde{\mathcal{B}}$, independent of W, and $\dot{b}$ (respectively $(\dot{\sigma}_j)_{1\leq j\leq q}$) is the Jacobian matrix of b (respectively $(\sigma_j)_{1\leq j\leq q}$). Our target now is to use the Euler Multilevel Monte Carlo method introduced above to approximate $\mathbb{E}\psi(X_T) = \mathbb{E}g(\theta, X_T^\theta, W_T), \theta \in \mathbb{R}^q$ by

$$\frac{1}{N_0}\sum_{i=1}^{N_0} g(\theta, \hat{X}_{T,i}^{m^0,\theta}, \hat{W}_{T,i}) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\left(g(\theta, X_{T,i}^{\ell,m^\ell,\theta}, W_{T,i}^\ell) - g(\theta, X_{T,i}^{\ell,m^{\ell-1},\theta}, W_{T,i}^\ell)\right), \tag{III.7}$$

where $X_T^{n,\theta}$ is the Euler scheme associated to X_T^θ (III.2) with time step T/n and $g(\theta,x,y) = \psi(x)e^{-\theta\cdot y-\frac{1}{2}|\theta|^2T}, \forall x \in \mathbb{R}^d$ and $y \in \mathbb{R}^q$, obtained by using Girsanov theorem in view to use importance sampling. According to relation (III.5), our limit variance is given by

$$\tilde{\mathrm{V}}\mathrm{ar}\left(\nabla_x g(\theta, X_T^\theta, W_T)\cdot U_T^\theta\right),$$

where U^θ is the weak limit process of the error $\sqrt{\frac{m^\ell}{(m-1)T}}(X^{m^\ell,\theta} - X^{m^{\ell-1},\theta})$ defined on the extension $\tilde{\mathcal{B}}$ and solution to

$$dU_t^\theta = \left(\dot{b}(X_t^\theta) + \sum_{j=1}^{q}\theta_j\dot{\sigma}_j(X_t^\theta)\right)U_t^\theta dt + \sum_{j=1}^{q}\dot{\sigma}_j(X_t^\theta)U_t^\theta dW_t^j - \frac{1}{\sqrt{2}}\sum_{j,\ell=1}^{q}\dot{\sigma}_j(X_t^\theta)\sigma_\ell(X_t^\theta)d\tilde{W}_t^{\ell j}. \tag{III.8}$$

Note that U and U^θ are the same processes obtained for the asymptotic behavior of $(X^{m^\ell} - X)$ and $(X^{m^\ell,\theta} - X^\theta)$ (see chapter II).

From a practical point of view, it is natural to choose the optimal θ^* minimizing the associated variance. As $\tilde{\mathbb{E}}(\nabla_x g(\theta, X_T^\theta, W_T)\cdot U_T^\theta) = 0$ (see Proposition 2.1 in Kebaier [37]), it boils down to choose

$$\begin{aligned}\theta^* &= \arg\min_{\theta\in\mathbb{R}^q} v^{ML}(\theta),\\ &\text{with } v^{ML}(\theta) := \tilde{\mathbb{E}}\left[(\nabla_x g(\theta, X_T^\theta, W_T)\cdot U_T^\theta)^2\right] = \tilde{\mathbb{E}}\left[(\nabla\psi(X_T)\cdot U_T)^2\, e^{-\theta\cdot W_T+\frac{1}{2}|\theta|^2T}\right].\end{aligned} \tag{III.9}$$

The last relation is obtained by using a change of probability. In fact, according to Girsanov theorem, the process (B^θ, X, U) under $\tilde{\mathbb{P}}^\theta$ has the same law as (W, X^θ, U^θ) under $\tilde{\mathbb{P}}$. Note that from a practical point of view the quantity $v^{ML}(\theta)$ is not explicit, we use the Euler scheme to discretize (X,U) and we introduce the following quantity

$$\theta_n^* := \arg\min_{\theta\in\mathbb{R}^q} v_n^{ML}(\theta) \quad \text{with} \quad v_n^{ML}(\theta) := \tilde{\mathbb{E}}\left[(\nabla\psi(X_T^n)\cdot U_T^n)^2\, e^{-\theta\cdot W_T+\frac{1}{2}|\theta|^2T}\right], \tag{III.10}$$

where U^n is the Euler discretization scheme of U.
In the next section, we give the results on the convergence of θ_n^* towards θ^* as n tends to infinity which is proved in the same way as in Chapter II. Moreover, we establish the convergence results of the problem of estimating θ_n^* using the constrained version of Robbins-Monro algorithm (see Subsection 2.3 in Chapter I). More precisely, we construct recursively a sequence of random variables $(\theta_i^n)_{i,n\in\mathbb{N}}$ using a stochastic algorithm with projection and we prove that

$$\lim_{i,n\to\infty} \theta_i^n = \lim_{i\to\infty}\big(\lim_{n\to\infty} \theta_i^n\big) = \lim_{n\to\infty}\big(\lim_{i\to\infty} \theta_i^n\big) = \theta^*,\ \tilde{\mathbb{P}}\text{-}a.s.$$

In section 3, we first introduce the new adaptive algorithm obtained by combining together the importance sampling procedure and the Euler Multilevel Monte Carlo method. Then, we prove a central limit theorem for this new algorithm (see Theorem 3.2). In section 4, we give a complexity analysis of the adaptive Euler Multilevel Monte Carlo algorithm and we proceed to numerical simulations to illustrate the efficiency of this new method for pricing an European call option under the Heston model.

2 Stochastic algorithm with projection

At first, let us recall that in Chapter II, the variance associated to the Statistical Romberg method is equal to $\tilde{\mathbb{E}}\left(\left[\psi(X_T^\theta)^2 + (\nabla\psi(X_T^\theta)\cdot U_T^\theta)^2\right] e^{-2\theta.W_T - |\theta|^2 T}\right)$. Therefore, the first three propositions below can be proved by following step by step the proofs of the results obtained in the previous chapter. The existence and uniqueness of θ^* is ensured by the following result.

Proposition 2.1. *Suppose σ and b are in $\mathscr{C}_b^1$ ⋆ and let ψ satisfying $\mathbb{P}(\nabla\psi(X_T)\cdot U_T \neq 0) > 0$. If there exists $a > 1$ such that $\mathbb{E}\left[|\nabla\psi(X_T)|^{2a}\right]$ is finite, then the function $\theta \mapsto v^{ML}(\theta)$ is $\mathscr{C}^2$ and strictly convex with $\nabla v^{ML}(\theta) = \tilde{\mathbb{E}} H^{ML}(\theta, X_T, U_T, W_T)$ where*

$$H^{ML}(\theta, X_T, U_T, W_T) := (\theta T - W_T)(\nabla\psi(X_T)\cdot U_T)^2 e^{-\theta\cdot W_T + \frac{1}{2}|\theta|^2 T}. \tag{III.11}$$

Moreover, there exists a unique $\theta^* \in \mathbb{R}^q$ such that $min_{\theta\in\mathbb{R}^q} v^{ML}(\theta) = v^{ML}(\theta^*)$.

Proof. Here, we follow the proof of Proposition 2.2 in Chapter II to obtain the result. □

The same results can be obtained for the Euler scheme X^n.

⋆. for an integer $k \geq 1$, $\mathscr{C}_b^k$ denotes the set of functions g in $\mathscr{C}^k$ with k^{th} all partial derivatives up to k^{th} order bounded.

Proposition 2.2. *Suppose σ and b are in $\mathscr{C}_b^1$. Given $n \in \mathbb{N}$, let ψ satisfying $\mathbb{P}(\nabla\psi(X_T^n)\cdot U_T^n \neq 0) > 0$. If there exists $a > 1$ such that $\mathbb{E}\left[|\nabla\psi(X_T^n)|^{2a}\right]$ is finite, then the function $\theta \mapsto v_n^{ML}(\theta)$ is $\mathscr{C}^2$ and strictly convex with $\nabla v_n^{ML}(\theta) = \tilde{\mathbb{E}}H^{ML}(\theta, X_T^n, U_T^n, W_T)$. Moreover, there exists a unique $\theta_n^* \in \mathbb{R}^q$ such that $min_{\theta\in\mathbb{R}^q} v_n^{ML}(\theta) = v_n^{ML}(\theta_n^*)$.*

Further, we have the convergence of θ_n^* towards θ^* as n tends to infinity.

Proposition 2.3. *Suppose σ and b are in $\mathscr{C}_b^{1,1}$ ⋆. Let ψ satisfying $\mathbb{P}(\nabla\psi(X_T)\cdot U_T \neq 0) > 0$ and $\mathbb{P}(\nabla\psi(X_T^n)\cdot U_T^n \neq 0) > 0$ for all $n \in \mathbb{N}$. If there exists $a > 1$ such that $\mathbb{E}\left[|\nabla\psi(X_T)|^{2a}\right]$ and $\sup_{n\in\mathbb{N}}\mathbb{E}\left[|\nabla\psi(X_T^n)|^{2a}\right]$ are finite. Then,*

$$\theta_n^* \longrightarrow \theta^*, \qquad \text{as } n \to \infty.$$

Proof. This result is obtained by following the proof of Theorem 4.4 in Chapter II. □

Now, we focus our interest on the effective approximation of the above quantities θ_n^* and θ^*. Our aim is to construct a sequence $(\theta_i)_{i\in\mathbb{N}}$ (resp. for fixed n, $(\theta_i^n)_{i\in\mathbb{N}}$) such that $\lim_{i\to\infty}\theta_i = \theta^*$ (resp. $\lim_{i\to\infty}\theta_i^n = \theta_n^*$) almost surely. Let $K \subset \mathbb{R}^q$ be a compact convex subset containing 0. For $\theta_0 \in K$, we introduce the sequences $(\theta_i)_{i\in\mathbb{N}}$ and $(\theta_i^n)_{i\in\mathbb{N}}$ defined recursively by

$$\begin{cases} \theta_{i+1} &= \Pi_K\left[\theta_i - \gamma_{i+1}H^{ML}(\theta_i, X_{T,i+1}, U_{T,i+1}, W_{T,i+1})\right] \\ \theta_{i+1}^n &= \Pi_K\left[\theta_i^n - \gamma_{i+1}H^{ML}(\theta_i^n, X_{T,i+1}^n, U_{T,i+1}^n, W_{T,i+1})\right], \end{cases} \tag{III.12}$$

where Π_K is the Euclidean projection onto the constraint set K, H^{ML} is given by relation (III.11) and the gain sequence $(\gamma_i)_{i\in\mathbb{N}}$ is a decreasing sequence of positive real numbers satisfying

$$\sum_{i=1}^{\infty}\gamma_i = \infty \quad \text{and} \quad \sum_{i=1}^{\infty}\gamma_i^2 < \infty. \tag{III.13}$$

Theorem 2.4. *Suppose σ and b are in $\mathscr{C}_b^1$. Assume that $\mathbb{P}((\nabla\psi(X_T)\cdot U_T) \neq 0) > 0$ and for all $n \in \mathbb{N}$, $\mathbb{P}((\nabla\psi(X_T^n)\cdot U_T^n) \neq 0) > 0$ and there exists $a > 1$ such that $\mathbb{E}\left[|\nabla\psi(X_T)|^{4a}\right]$ and $\mathbb{E}\left[|\nabla\psi(X_T^n)|^{4a}\right]$ are finite, then the following assertions hold.*

1) *Behavior of the stochastic algorithms when $i \to \infty$*
   - *If $\theta^* = \arg\min v^{ML}(\theta)$ is unique s.t $\nabla v^{ML}(\theta^*) = 0$ and $\theta^* \in K$ then $\theta_i \xrightarrow[i\to\infty]{a.s} \theta^*$.*
   - *If $\theta_n^* = \arg\min v_n^{ML}(\theta)$ is unique s.t $\nabla v_n^{ML}(\theta_n^*) = 0$ and $\theta_n^* \in K$ then $\theta_i^n \xrightarrow[i\to\infty]{a.s} \theta_n^*$.*

⋆. for an integer $k \geq 1$ and $\delta \in [0,1]$, we denote by $\mathscr{C}_b^{k,\delta}$ the set of functions g in $\mathscr{C}^k$ with k^{th} order partial derivatives globally δ-Hölder and all partial derivatives up to k^{th} order bounded

2) *Behavior of the stochastic algorithm when $n \to \infty$*
For all $i \in \mathbb{N}$, we have $\theta_i^n \underset{n\to\infty}{\longrightarrow} \theta_i$ a.s.

Proof. Concerning the behavior of the stochastic algorithm when $i \to \infty$, both assertions can be proved in the same way, so we choose to give the proof only for the first one. According to Theorem A.1. in Laruelle, Lehalle and Pagès [44] (See Theorem 2.3 in Chapter I) on Robbins Monro algorithm with projection : to prove that $\theta_i \underset{n\to\infty}{\longrightarrow} \theta^*$, we need to check firstly that

$$\forall \theta \neq \theta^*, \quad \langle \nabla v^{ML}(\theta), \theta - \theta^* \rangle > 0.$$

This is satisfied using $\nabla v^{ML}(\theta^*) = 0$ and thanks to the convexity of v^{ML} ensured by Proposition 2.1. Secondly, we have to check the non explosion condition given by

$$\exists C > 0 \text{ such that } \forall \theta \in K, \quad \mathbb{E}\left[|H^{ML}(\theta, X_T, U_T, W_T)|^2\right] < C(1 + |\theta^2|).$$

Using Hölder's inequality, we can check that

$$\mathbb{E}\left[|H^{ML}(\theta, X_T, U_T, W_T)|^2\right] \leq e^{|\theta|^2 T}\mathbb{E}^{1/a}\left[|\nabla\psi(X_T)|^{4a}\right]$$
$$\mathbb{E}^{\frac{a-1}{2a}}\left[|U_T|^{\frac{8a}{a-1}}\right]\mathbb{E}^{\frac{a-1}{2a}}\left[\left|e^{-\theta\cdot W_T}(\theta T - W_T)\right|^{\frac{4a}{a-1}}\right]. \quad \text{(III.14)}$$

Since $\mathbb{E}\left[|\nabla\psi(X_T)|^{4a}\right]$ is finite and $\mathbb{E}\left[|U_T|^{\frac{4a}{a-1}}\right]$ is finite thanks to $(\tilde{\mathcal{P}})$ (see Chapter II), there exists $C > 0$ such that

$$\mathbb{E}\left[|H^{ML}(\theta, X_T, U_T, W_T)|^2\right] \leq Ce^{|\theta|^2 T}\mathbb{E}^{\frac{a-1}{2a}}\left[\left|e^{-\theta\cdot W_T}(\theta T - W_T)\right|^{\frac{4a}{a-1}}\right]. \quad \text{(III.15)}$$

Using $\theta \in K$, we conclude that $\sup_{\theta\in K}\mathbb{E}\left[|H^{ML}(\theta, X_T, U_T, W_T)|^2\right] < \infty$. Concerning the behavior of the stochastic algorithm when n goes to infinity, we proceed by induction. The base case is trivial and for the inductive step we suppose that for $i \in \mathbb{N}$, θ_i^n converges to θ_i a.s as n goes to infinity and we prove the statement for $n + 1$. We have $\theta_{i+1}^n = \Pi_K\left[\theta_i^n - \gamma_{i+1}H^{ML}(\theta_i^n, X_{T,i+1}^n, U_{T,i+1}^n, W_{T,i+1})\right]$. By the continuity of function H^{ML} and the projection function Π_K, we use the almost sure convergence of the pair $(X_{T,i}^n, U_{T,i}^n)$ to $(X_{T,i}, U_{T,i})$ as n goes to infinity combined with assumption $(\mathcal{H}_\psi)$ to deduce that θ_{i+1}^n converges to θ_{i+1} almost surely when n goes to infinity. □

The following corollary follows immediately thanks to Proposition 2.3 and Theorem 2.4.

Corollary 2.5. *Under assumptions of Proposition 2.3 and Theorem 2.4, the constrained algo-*

rithm given by (III.12) satisfies

$$\lim_{i,n\to\infty} \theta_i^n = \lim_{i\to\infty}\big(\lim_{n\to\infty} \theta_i^n\big) = \lim_{n\to\infty}\big(\lim_{i\to\infty} \theta_i^n\big) = \theta^*, \quad \tilde{\mathbb{P}}\text{-}a.s.,$$

where θ^* is given by relation (III.9).

3 Adaptive Euler Multilevel Monte Carlo procedure

Let $\tilde{\mathcal{B}} = (\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t\geq 0}, \tilde{\mathbb{P}})$ be the extension probability space introduced above endowed with the filtration $\tilde{\mathcal{F}}_{T,i} = \sigma(W_{t,j}, \tilde{W}_{t,j}, j \leq i, t \leq T)$ given in the beginning of this chapter. In what follows, let $(\theta_i^n)_{i\geq 0}$, $n \in \mathbb{N}$ and $(\theta_i)_{i\geq 0}$ be two families of sequences satisfying

$$(\mathcal{H}'_\theta) \quad \begin{cases} \text{For each } n \in \mathbb{N},\ (\theta_i^n)_{i\geq 0} \text{ and } (\theta_i)_{i\geq 0} \text{ are } (\tilde{\mathcal{F}}_{T,i})_{i\geq 0}\text{-adapted,} \\ (\theta_i)_{i\in\mathbb{N}} \in K \subset \mathbb{R}^q \text{ and } (\theta_i^n)_{i\in\mathbb{N}} \in K \subset \mathbb{R}^q, \\ \lim_{i\to\infty}\big(\lim_{n\to\infty} \theta_i^n\big) = \lim_{i\to\infty} \theta_i = \lim_{n\to\infty}\big(\lim_{i\to\infty} \theta_i^n\big) = \lim_{n\to\infty} \theta_n^* = \theta^*, \quad \tilde{\mathbb{P}}\text{-}a.s., \end{cases}$$

with deterministic limits θ^* and θ_n^*.

In this section we prove a central limit theorem for the adaptive Euler Multilevel Monte Carlo method which approximates our initial quantity of interest $\mathbb{E}\psi(X_T) = \mathbb{E}\left[\psi(X_T^\theta)e^{-\theta\cdot W_T - \frac{1}{2}|\theta|^2 T}\right]$ by

$$\frac{1}{N_0}\sum_{i=1}^{N_0} g(\theta_{i-1}^{m^0}, X_{T,i}^{m^0,\theta_{i-1}^{m^0}}, \hat{W}_{T,i}) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\left(g(\theta_{i-1}^{\ell,m^\ell}, X_{T,i}^{\ell,m^\ell,\theta_{i-1}^{\ell,m^\ell}}, W_{T,i}^\ell) - g(\theta_{i-1}^{\ell,m^\ell}, X_{T,i}^{\ell,m^{\ell-1},\theta_{i-1}^{\ell,m^\ell}}, W_{T,i}^\ell)\right), \tag{III.16}$$

where for all $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^q$, $g(\theta, x, y) = \psi(x)e^{-\theta\cdot y - \frac{1}{2}|\theta|^2 T}$, $L = \frac{\log n}{\log m}$. The sequences $(\theta_i^{\ell,m^\ell}, i \in \mathbb{N})_{\ell\in\mathbb{N}}$ are independent copies of $(\theta_i^{m^\ell}, i \in \mathbb{N})$ satisfying $(\mathcal{H}'_\theta)$. Note that the $(L+1)$ empirical means are independent, the Brownian path as well as $(\theta_i^{m^\ell}, i \in \mathbb{N})$ is independent in each empirical mean in the Euler Multilevel Monte Carlo.

Theorem 3.1. *For the above setting, assume that b and σ are $\mathscr{C}^1$ functions satisfying $(\mathcal{H}_{b,\sigma})$ and ψ satisfying assumptions $(\mathcal{H}_{\varepsilon_n})$, $(\mathcal{H}_\psi)$ and*

$$|\psi(x) - \psi(y)| \leq C(1 + |x|^p + |y|^p)|x - y|, \quad \textit{for some } C, p > 0. \tag{III.17}$$

Then, for the choice of $N_\ell, \ell \in \{0,1,...,L\}$ given by (III.4), the following convergence holds

$$\text{for } \alpha \in [1/2,1], \quad \lim_{n\to\infty} n^{2\alpha}\mathbb{E}\left[|Q_n - \mathbb{E}\psi(X_T)|^2\right] = \tilde{\sigma}^2 + C_\psi^2,$$

where C_ψ and α are given by relation $(\mathcal{H}_{\varepsilon_n})$ and $\tilde{\sigma}^2 := \tilde{\mathbb{E}}\left[[\nabla\psi(X_T)\cdot U_T]^2 e^{-\theta^*\cdot W_T + \frac{1}{2}|\theta^*|^2 T}\right]$.

Proof. At first, we rewrite the total error as follows

$$n^\alpha\left(Q_n - \mathbb{E}\psi(X_T)\right) = Q_n^1 + Q_n^2 + n^\alpha(\mathbb{E}\psi(X_T^n) - \mathbb{E}\psi(X_T)), \tag{III.18}$$

where Q_n^1 and Q_n^2 are given by the following expressions

$$Q_n^1 := n^\alpha\left(\frac{1}{N_0}\sum_{i=1}^{N_0} g(\hat{\theta}_{i-1}^{m^0}, \hat{X}_{T,i}^{m^0,\hat{\theta}_{i-1}^{m^0}}, W_{T,i}) - \mathbb{E}\psi(X_T^1)\right). \tag{III.19}$$

$$Q_n^2 := n^\alpha\left[\sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\left(g(\theta_{i-1}^{\ell,m^\ell}, X_{T,i}^{\ell,m^\ell,\theta_{i-1}^{\ell,m^\ell}}, W_{T,i}) - g(\theta_{i-1}^{\ell,m^\ell}, X_{T,i}^{\ell,m^{\ell-1},\theta_{i-1}^{\ell,m^\ell}}, W_{T,i}) - \mathbb{E}\left[\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})\right]\right)\right].$$

Since Q_n^1 and Q_n^2 are independent and centered, we write also

$$n^{2\alpha}\mathbb{E}\left[|Q_n - \mathbb{E}\psi(X_T)|^2\right] = \mathbb{E}\left[|Q_n^1|^2\right] + \mathbb{E}\left[|Q_n^2|^2\right] + n^{2\alpha}(\mathbb{E}\psi(X_T^n) - \mathbb{E}\psi(X_T))^2.$$

Using assumption $(\mathcal{H}_{\varepsilon_n})$, the last term of the previous expression converges towards the discretization constant C_ψ^2 as n goes to infinity.

• **Step 1.** For the first term $\mathbb{E}|Q_n^1|^2$, noticing that $\left(\sum_{i=1}^k g(\theta_{i-1}^{m^0}, X_{T,i}^{m^0,\theta_{i-1}^{m^0}}, W_{T,i}) - \mathbb{E}\psi(X_T^{m^0}), k \geq 1\right)$ is a martingale with respect to the filtration $\tilde{\mathcal{F}}_{T,k}$, then we write

$$\begin{aligned}\mathbb{E}|Q_n^1|^2 &= \mathbb{E}\left(\frac{n^{2\alpha}}{N_0^2}\sum_{i=1}^{N_0}\mathbb{E}\left[\left|g(\theta_{i-1}^{m^0}, X_{T,i}^{m^0,\theta_{i-1}^{m^0}}, W_{T,i}) - \mathbb{E}\psi(X_T^{m^0})\right|^2 |\tilde{\mathcal{F}}_{T,i-1}\right]\right)\\ &= \frac{n^{2\alpha}}{N_0^2}\sum_{i=1}^{N_0}\left(\mathbb{E}\left(\psi(X_T^{m^0})^2 e^{-\theta_{i-1}^{m^0}\cdot W_T + \frac{1}{2}|\theta_{i-1}^{m^0}|^2 T}\right) - \left[\mathbb{E}\psi(X_T^{m^0})\right]^2\right).\end{aligned}$$

Since $X_{T,i}^{m^0} \perp\!\!\!\perp \tilde{\mathcal{F}}_{T,i-1}$ and $\theta_{i-1}^{m^0}$ is $\tilde{\mathcal{F}}_{T,i-1}$-measurable and thanks to Girsanov theorem, we obtain the last equality by introducing a new couple of random variables $(X_T^{m^0}, W_T)$ independent of

$\tilde{\mathfrak{F}}_T = \cup_{i\geq 1}\tilde{\mathfrak{F}}_{T,i}$. As $N_0 = \frac{n^2(m-1)T}{a_0\sum_{\ell=1}^L a_\ell}$, we write

$$\mathbb{E}|Q_n^1|^2 = \frac{n^{2\alpha-2}a_0}{(m-1)T\sum_{\ell=1}^L a_\ell}\frac{1}{N_0}\sum_{i=1}^{N_0}\left(\mathbb{E}\left(\psi(X_T^{m^0})^2 e^{-\theta_{i-1}^{m^0}\cdot W_T+\frac{1}{2}|\theta_{i-1}^{m^0}|^2T}\right) - \left[\mathbb{E}\psi(X_T^{m^0})\right]^2\right). \quad \text{(III.20)}$$

Since $(\theta_{i-1}^{m^0})_{i\in\mathbb{N}} \in K$, we have

$$\exists c>0, \quad \sup_{i\in\mathbb{N}}\left|\psi(X_T^{m^0})^2 e^{-\theta_{i-1}^{m^0}\cdot W_T+\frac{1}{2}|\theta_{i-1}^{m^0}|^2T}\right| \leq \psi(X_T^{m^0})^2 e^{c|W_T|+\frac{c^2}{2}T}.$$

This upper bound is clearly integrable using property $(\mathcal{P})$ and assumption (III.17) together with the Hölder's inequality. Therefore, by the dominated convergence theorem and under assumption $(\mathcal{H}'_\theta)$, we obtain that

$$\lim_{i\to\infty}\mathbb{E}\left(\psi(X_T^{m^0})^2 e^{-\theta_{i-1}^{m^0}\cdot W_T+\frac{1}{2}|\theta_{i-1}^{m^0}|^2T}\right) = \mathbb{E}\left(\psi(X_T^{m^0})^2 e^{-\theta^*_{m^0}\cdot W_T+\frac{1}{2}|\theta^*_{m^0}|^2T}\right).$$

Thus, by applying Cesaro's lemma, we obtain that

$$\lim_{i\to\infty}\frac{1}{N_0}\sum_{i=1}^{N_0}\left(\mathbb{E}\left(\psi(X_T^{m^0})^2 e^{-\theta_{i-1}^{m^0}\cdot W_T+\frac{1}{2}|\theta_{i-1}^{m^0}|^2T}\right) - \left[\mathbb{E}\psi(X_T^{m^0})\right]^2\right)$$
$$= \mathbb{E}\left(\psi(X_T^{m^0})^2 e^{-\theta^*_{m^0}\cdot W_T+\frac{1}{2}|\theta^*_{m^0}|^2T}\right) - \mathbb{E}^2(\psi(X_T^{m^0})). \quad \text{(III.21)}$$

As $\sum_{\ell=1}^L a_\ell \underset{\ell\to\infty}{\longrightarrow}\infty$ and $\alpha\in[1/2,1]$, we conclude thanks to relation (III.20) that

$$\mathbb{E}|Q_n^1|^2 \underset{n\to\infty}{\longrightarrow} 0. \quad \text{(III.22)}$$

So it remains now to study the asymptotic behavior of $\mathbb{E}|Q_n^2|^2$.

• **Step 2.** We consider the second term $\mathbb{E}|Q_n^2|^2$ and we write

$$\mathbb{E}|Q_n^2|^2 = \mathbb{E}\left[\left(\sum_{\ell=1}^L\frac{n^\alpha}{N_\ell}\sum_{i=1}^{N_\ell} Z_{T,i}^{m^\ell,m^{\ell-1},\theta_{i-1}^{\ell,m^\ell}}\right)^2\right],$$

where

$$Z_{T,i}^{m^\ell,m^{\ell-1},\theta_{i-1}^{\ell,m^\ell}} = \Big(g(\theta_{i-1}^{\ell,m^\ell}, X_{T,i}^{\ell,m^\ell,\theta_{i-1}^{\ell,m^\ell}}, W_{T,i}^\ell) - g(\theta_{i-1}^{\ell,m^\ell}, X_{T,i}^{\ell,m^{\ell-1},\theta_{i-1}^{\ell,m^\ell}}, W_{T,i}^\ell) - \mathbb{E}[\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})]\Big). \quad \text{(III.23)}$$

As $(Z_{T,i}^{m^\ell,m^{\ell-1},\theta_{i-1}^{\ell,m^\ell}})_{1\leq\ell\leq L}$ are independent, we have

$$\mathbb{E}|Q_n^2|^2 = \sum_{\ell=1}^{L} \frac{n^{2\alpha}}{N_\ell^2} \mathbb{E}\left(\sum_{i=1}^{N_\ell} Z_{T,i}^{m^\ell,m^{\ell-1},\theta_{i-1}^{\ell,m^\ell}}\right)^2.$$

By the same argument as in the first step, noticing that for each $\ell \in \{1,...,L\}$, $\left(\sum_{i=1}^k Z_{T,i}^{m^\ell,m^{\ell-1},\theta_{i-1}^{\ell,m^\ell}}, k\geq 1\right)$ is $\tilde{\mathfrak{F}}_{T,k}$ martingale, we write

$$\begin{aligned}\mathbb{E}|Q_n^2|^2 &= \mathbb{E}\left(\sum_{\ell=1}^{L} \frac{n^{2\alpha}}{N_\ell^2} \sum_{i=1}^{N_\ell} \mathbb{E}\left[\left(Z_{T,i}^{m^\ell,m^{\ell-1},\theta_{i-1}^{m^\ell}}\right)^2 |\tilde{\mathfrak{F}}_{T,i-1}\right]\right) \\ &= \sum_{\ell=1}^{L} \frac{n^{2\alpha}}{N_\ell}\left[\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\left(\mathbb{E}\left([\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})]^2 e^{-\theta_{i-1}^{m^\ell}\cdot W_T + \frac{1}{2}|\theta_{i-1}^{m^\ell}|^2 T}\right)\right.\right. \\ &\qquad \left.\left. - \left(\mathbb{E}\psi(X_T^{m^\ell}) - \mathbb{E}\psi(X_T^{m^{\ell-1}})\right)^2\right)\right].\end{aligned} \quad \text{(III.24)}$$

Since for each $\ell \in \{1,...,L\}$, $X_{T,i}^{\ell,m^\ell} \perp\!\!\!\perp \tilde{\mathfrak{F}}_{T,i-1}$ and $\theta_{i-1}^{m^\ell}$ is $\tilde{\mathfrak{F}}_{T,i-1}$-measurable and thanks to Girsanov theorem, we obtain the last equality by introducing a new couple of random variables $(X_T^{m^\ell}, W_T)$ independent of $\tilde{\mathfrak{F}}_T = \cup_{i\geq 1}\tilde{\mathfrak{F}}_{T,i}$. As $N_\ell = \frac{n^{2\alpha}(m-1)T}{m^\ell a_\ell}\sum_{\ell=1}^L a_\ell$, $\ell \in \{0,...,L\}$ (see relation (III.4)), we write

$$\begin{aligned}\mathbb{E}|Q_n^2|^2 = &\frac{1}{\sum_{\ell=1}^L a_\ell}\sum_{\ell=1}^{L} a_\ell \frac{m^\ell}{(m-1)T}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\mathbb{E}\left([\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})]^2 e^{-\theta_{i-1}^{m^\ell}\cdot W_T + \frac{1}{2}|\theta_{i-1}^{m^\ell}|^2 T}\right) \\ &- \frac{1}{\sum_{\ell=1}^L a_\ell}\sum_{\ell=1}^{L} a_\ell\left(\sqrt{\frac{m^\ell}{(m-1)T}}\left[\mathbb{E}\psi(X_T^{m^\ell}) - \mathbb{E}\psi(X_T^{m^{\ell-1}})\right]\right)^2.\end{aligned} \quad \text{(III.25)}$$

Now, for the last term of relation (III.24), under assumption $(\mathcal{H}_\psi)$, we apply the Taylor's expansion theorem twice and we get

$$\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}}) = \nabla\psi(X_T).U_T^{m^\ell,m^{\ell-1}} \\ + (X_T^{m^\ell} - X_T)\varepsilon(X_T, X_T^{m^\ell} - X_T) - (X_T^{m^{\ell-1}} - X_T)\varepsilon(X_T, X_T^{m^{\ell-1}} - X_T).$$

The function ε is given by the Taylor-young expansion, so it satisfies $\varepsilon(X_T, X_T^{m^\ell} - X_T) \xrightarrow[\ell\to\infty]{\mathbb{P}} 0$ and $\varepsilon(X_T, X_T^{m^{\ell-1}} - X_T) \xrightarrow[\ell\to\infty]{\mathbb{P}} 0$. By property $(\mathcal{P})$, we get the tightness of $\sqrt{\frac{m^\ell}{(m-1)T}}(X_T^{m^\ell} - X_T)$ and $\sqrt{\frac{m^\ell}{(m-1)T}}(X_T^{m^{\ell-1}} - X_T)$ and we deduce

$$\sqrt{\frac{m^\ell}{(m-1)T}}\Big((X_T^{m^\ell} - X_T)\varepsilon(X_T, X_T^{m^\ell} - X_T) - (X_T^{m^{\ell-1}} - X_T)\varepsilon(X_T, X_T^{m^{\ell-1}} - X_T)\Big) \xrightarrow[\ell\to\infty]{\mathbb{P}} 0.$$

So, according to the Theorem of stable convergence (see Theorem 1.8 in Chapter I) we conclude that

$$\sqrt{\frac{m^\ell}{(m-1)T}}\Big(\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})\Big) \overset{stably}{\Longrightarrow} \nabla\psi(X_T).U_T, \quad \text{as } \ell \to \infty. \tag{III.26}$$

Using (III.17), it follows from property $(\mathcal{P})$ that

$$\forall a > 1, \sup_\ell \mathbb{E}\left|\sqrt{\frac{m^\ell}{(m-1)T}}\Big(\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})\Big)\right|^{2a} < \infty.$$

Hence, we deduce using relation (III.26) that

$$\mathbb{E}\left(\sqrt{\frac{m^\ell}{(m-1)T}}\Big(\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})\Big)\right)^k \to \tilde{\mathbb{E}}\left(\nabla\psi(X_T).U_T\right)^k < \infty \text{ for } k \in \{1,2\}. \tag{III.27}$$

Then, $\sqrt{\frac{m^\ell}{(m-1)T}}\Big(\mathbb{E}(\psi(X_T^{m^\ell})) - \mathbb{E}(\psi(X_T^{m^{\ell-1}}))\Big)$ converges to $\tilde{\mathbb{E}}\left(\nabla\psi(X_T).U_T\right) = 0$ as ℓ goes to infinity. Consequently, by applying Toeplitz lemma, we obtain that

$$\lim_{\ell\to\infty} \frac{1}{\sum_{\ell=1}^L a_\ell}\sum_{\ell=1}^L a_\ell\left(\sqrt{\frac{m^\ell}{(m-1)T}}\Big[\mathbb{E}\psi(X_T^{m^\ell}) - \mathbb{E}\psi(X_T^{m^{\ell-1}})\Big]\right)^2 = 0.$$

We focus now on the asymptotic behavior of the first term on the right hand side on relation (III.24). Since $\theta_{i-1}^{m^\ell} \in K$, we have

$$\exists c > 0,\ \sup_{i\in\mathbb{N}} \left| \left(\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})\right)^2 e^{-\theta_{i-1}^{\ell,m^\ell}\cdot W_T + \frac{1}{2}|\theta_{i-1}^{\ell,m^\ell}|^2 T} \right| \leq \left(\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})\right)^2 e^{c|W_T| + \frac{c^2}{2}T}.$$

This upper bound is clearly integrable using property ($\mathcal{P}$) and assumption (III.17) together with the Hölder's inequality. Therefore, by the dominated convergence and under assumption ($\mathcal{H}'_\theta$), we obtain that

$$\lim_{i\to\infty} \mathbb{E}\left([\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})]^2 e^{-\theta_{i-1}^{\ell,m^\ell}\cdot W_T + \frac{1}{2}|\theta_{i-1}^{\ell,m^\ell}|^2 T} \right) = \mathbb{E}\left([\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})]^2 e^{-\theta^*_{m^\ell}\cdot W_T + \frac{1}{2}|\theta^*_{m^\ell}|^2 T} \right).$$

Hence, by using Cesaro lemma, we obtain that

$$\lim_{i\to\infty} \frac{1}{N_\ell}\sum_{i=1}^{N_\ell} \mathbb{E}\left([\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})]^2 e^{-\theta_{i-1}^{\ell,m^\ell}\cdot W_T + \frac{1}{2}|\theta_{i-1}^{\ell,m^\ell}|^2 T} \right) = \mathbb{E}\left([\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})]^2 e^{-\theta^*_{m^\ell}\cdot W_T + \frac{1}{2}|\theta^*_{m^\ell}|^2 T} \right).$$

Otherwise, thanks to (III.26), we have

$$\sqrt{\frac{m^\ell}{(m-1)T}}[\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})] e^{-\frac{1}{2}\theta^*_{m^\ell}\cdot W_T + \frac{1}{4}|\theta^*_{m^\ell}|^2 T} \underset{\ell\to\infty}{\overset{stably}{\Longrightarrow}} \nabla\psi(X_T)\cdot U_T e^{-\frac{1}{2}\theta^*\cdot W_T + \frac{1}{4}|\theta^*|^2 T}. \quad \text{(III.28)}$$

Moreover, for $a > 1$ we have by Cauchy-Schwartz inequality

$$\mathbb{E}\left|\sqrt{\frac{m^\ell}{(m-1)T}}[\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})] e^{-\frac{1}{2}\theta^*_{m^\ell}\cdot W_T + \frac{1}{4}|\theta^*_{m^\ell}|^2 T}\right|^{2a} \leq \left(\frac{m^\ell}{(m-1)T}\right)^a \left[\mathbb{E}\left|\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})\right|^{4a}\right]^{\frac{1}{2}} e^{\frac{a(2a+1)}{2}|\theta^*_{m^\ell}|^2 T}. \quad \text{(III.29)}$$

Since $(\theta^*_{m^\ell})_{1\leq\ell\leq L} \in K$ and thanks to (III.17) together with property ($\mathcal{P}$) (see Chapter II) , we obtain thanks to (III.29)

$$\sup_\ell \mathbb{E}\left|\sqrt{\frac{m^\ell}{(m-1)T}}[\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})] e^{-\frac{1}{2}\theta^*_{m^\ell}\cdot W_T + \frac{1}{4}|\theta^*_{m^\ell}|^2 T}\right|^{2a} < \infty. \quad \text{(III.30)}$$

Then, by the stable convergence obtained in (III.28) and the uniform integrability property given by (III.30) and under the assumption $(\mathcal{H}'_\theta)$, we deduce

$$\lim_{\ell\to\infty} \frac{m^\ell}{(m-1)T}\mathbb{E}\left([\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})]^2 e^{-\theta^*_{m^\ell}\cdot W_T+\frac{1}{2}|\theta^*_{m^\ell}|^2T}\right) = \mathbb{E}\left([\nabla\psi(X_T)\cdot U_T]^2\, e^{-\theta^*\cdot W_T+\frac{1}{2}|\theta^*|^2T}\right). \quad \text{(III.31)}$$

Now, considering once again the first term on the right hand side of relation (III.24) and using Toeplitz lemma, we obtain that

$$\lim_{\ell\to\infty} \frac{1}{\sum_{\ell=1}^L a_\ell}\sum_{\ell=1}^L a_\ell \frac{m^\ell}{(m-1)T}\mathbb{E}\left([\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}})]^2 e^{-\theta^*_{m^\ell}\cdot W_T+\frac{1}{2}|\theta^*_{m^\ell}|^2T}\right) = \mathbb{E}\left([\nabla\psi(X_T)\cdot U_T]^2\, e^{-\theta^*\cdot W_T+\frac{1}{2}|\theta^*|^2T}\right). \quad \text{(III.32)}$$

Hence, combining this last result together with relation (III.22), we conclude thanks to (III.24) that

$$\lim_{\ell\to\infty} \mathbb{E}|Q_n^2|^2 = \tilde{\mathbb{E}}\left([\nabla\psi(X_T)\cdot U_T]^2\, e^{-\theta^*\cdot W_T+\frac{1}{2}|\theta^*|^2T}\right), \quad \tilde{\mathbb{P}}\text{-}a.s. \quad \text{(III.33)}$$

This completes the proof. □

Our aim now is to prove a central limit theorem for the adaptive Euler Multilevel Monte Carlo method.

Theorem 3.2. *Under assumptions of Theorem 3.1 and for the choice of $N_\ell, \ell \in \{0,1,...,L\}$ given by (III.4), the following convergence holds*

$$n^\alpha\left(Q_n - \mathbb{E}\psi(X_T)\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(C_\psi, \tilde{\sigma}^2\right),$$

where $\tilde{\sigma}^2 := \tilde{\mathbb{E}}\left[[\nabla\psi(X_T)\cdot U_T]^2\, e^{-\theta^*\cdot W_T+\frac{1}{2}|\theta^*|^2T}\right]$.

Proof. We consider the same decomposition given by relation (III.18) in the beginning of the proof of Theorem 3.1. Under assumption $(\mathcal{H}_{\varepsilon_n})$, the last term on the right hand side of this relation converges to C_ψ as n goes to ∞. For the convergence of the term Q_n^1, we use the result (III.22) obtained in Theorem 3.1 and we deduce that $Q_n^1 \xrightarrow[n\to\infty]{\mathbb{P}} 0$. Concerning the convergence of the term Q_n^2, we plan to use the Lindeberg-Feller central limit theorem (see Theorem 1.3 in Chapter I) with the Lyapunov condition. We introduce the independent random variables

$$(S_{\ell,N_\ell}, \ell\in\{1,...,L\}) \text{ where } (S_{\ell,k} = \frac{n^\alpha}{N_\ell}\sum_{i=1}^k Z_{T,i}^{m^\ell,m^{\ell-1},\theta_{i-1}^{\ell,m^\ell}}, k\geq 1) \text{ for } \ell\in\{1,...,L\}, \quad \text{(III.34)}$$

and we need to check the assertions A1. and A3. in Theorem 1.3 in Chapter I. More precisely, we will prove

A1. $\lim_{L\to\infty} \sum_{\ell=1}^{L} \mathbb{E}(S^2_{\ell,N_\ell}) = \tilde{\mathbb{E}}\left([\nabla\psi(X_T)\cdot U_T]^2\, e^{-\theta^*\cdot W_T+\frac{1}{2}|\theta^*|^2T}\right)$.
A3. For $p>2$, $\lim_{L\to\infty} \sum_{\ell=1}^{L} \mathbb{E}(S^p_{\ell,N_\ell}) = 0$.

Concerning the first assertion A1., we have that $\sum_{\ell=1}^{L} \mathbb{E}(S^2_{\ell,N_\ell}) = \mathbb{E}(Q_n^2)^2$. Therefore, using relation (III.33) in the proof of Theorem 3.1, we get

$$\lim_{\ell\to\infty} \sum_{\ell=1}^{L} \mathbb{E}(S^2_{\ell,N_\ell}) = \tilde{\mathbb{E}}\left([\nabla\psi(X_T)\cdot U_T]^2\, e^{-\theta^*\cdot W_T+\frac{1}{2}|\theta^*|^2T}\right), \quad \tilde{\mathbb{P}}\text{-}a.s.$$

Now, it remains to verify the assertion A3. We get by Burkholder's inequality (see Theorem 2.10 in [32]) : for $p>2$, there exists $C_p>0$ such that

$$\mathbb{E}\,|S_{\ell,N_\ell}|^p \le C_p \frac{n^{\alpha p}}{N_\ell^{p/2}} \mathbb{E}\left[\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}(Z_{T,i}^{m^\ell,m^{\ell-1},\theta_i^{\ell,m^\ell}})^2\right]^{p/2}. \tag{III.35}$$

Moreover, by using Jensen's inequality, we obtain for $p>2$

$$\sum_{\ell=1}^{L} \mathbb{E}\,|S_{\ell,N_\ell}|^p \le \sum_{\ell=1}^{L} C_p \frac{n^{\alpha p}}{N_\ell^{p/2}} \mathbb{E}\left[\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}(Z_{T,i}^{m^\ell,m^{\ell-1},\theta_i^{\ell,m^\ell}})^p\right].$$

Using that $N_\ell = \frac{n^{2\alpha}(m-1)T}{m^\ell a_\ell}\sum_{\ell=1}^{L} a_\ell$ (see relation (III.4)) and by conditioning, we get

$$\sum_{\ell=1}^{L} \mathbb{E}\,|S_{\ell,N_\ell}|^p \le \sum_{\ell=1}^{L} C_p \frac{a_\ell^{p/2}}{(\sum_{\ell=1}^{L} a_\ell)^{p/2}} \mathbb{E}\left(\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\mathbb{E}\left[\frac{m^{\ell p/2}}{((m-1)T)^{p/2}}(Z_{T,i}^{m^\ell,m^{\ell-1},\theta_i^{\ell,m^\ell}})^p|\tilde{\mathfrak{F}}_{T,i-1}\right]\right)$$

We write also

$$\sum_{\ell=1}^{L} \mathbb{E}\,|S_{\ell,N_\ell}|^p \le 2^{p-1}C_p \sum_{\ell=1}^{L} \frac{a_\ell^{p/2}}{(\sum_{\ell=1}^{L} a_\ell)^{p/2}} A_\ell + 2^{p-1}C_p \sum_{\ell=1}^{L} \frac{a_\ell^{p/2}}{(\sum_{\ell=1}^{L} a_\ell)^{p/2}} B_\ell. \tag{III.36}$$

where

$$A_\ell = \frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\mathbb{E}\left[\left|\sqrt{\frac{m^\ell}{((m-1)T)}}(\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}}))\right|^p e^{-(p-1)\theta_{i-1}^{m^\ell}\cdot W_T-(p/2-3/2)|\theta_{i-1}^{m^\ell}|^2T}\right],$$
$$B_\ell = \left(\mathbb{E}\left(\sqrt{\frac{m^\ell}{((m-1)T)}}(\psi(X_T^{m^\ell}) - \psi(X_T^{m^{\ell-1}}))\right)\right)^p.$$

Since for each $\ell \in \{1,...,L\}$, $X^{\ell,m^\ell}_{T,i} \perp\!\!\!\perp \tilde{\mathcal{F}}_{T,i-1}$ and $\theta^{m^\ell}_{i-1}$ is $\tilde{\mathcal{F}}_{T,i-1}$-measurable and thanks to Girsanov theorem, we obtain the last equality (III.36) by introducing a new couple of random variables $(X^{m^\ell}_T, W_T)$ independent of $\tilde{\mathcal{F}}_T = \cup_{i\geq 1}\tilde{\mathcal{F}}_{T,i}$. By using relation (III.27) together with assumption $(\mathcal{W})$, we deduce that the last term on the right hand side in (III.36) converges to 0 as ℓ goes to infinity

$$\lim_{\ell\to\infty}\sum_{\ell=1}^{L}\frac{a_\ell^{p/2}}{(\sum_{\ell=1}^{L}a_\ell)^{p/2}}B_\ell = 0. \tag{III.37}$$

Now, we focus on the convergence of the first term on the right hand side in (III.36). Since $\theta^{m^\ell}_{i-1} \in K$, there exists $c > 0$ such that

$$\begin{aligned}\sup_{i\in\mathbb{N}}\left|\left|\sqrt{\frac{m^\ell}{((m-1)T)}}(\psi(X^{m^\ell}_T)-\psi(X^{m^{\ell-1}}_T))\right|^p e^{-(p-1)\theta^{m^\ell}_{i-1}\cdot W_T-(p/2-3/2)|\theta^{m^\ell}_{i-1}|^2T}\right| \\ \leq \left|\sqrt{\frac{m^\ell}{((m-1)T)}}(\psi(X^{m^\ell}_T)-\psi(X^{m^{\ell-1}}_T))\right|^p e^{(p-1)c|W_T|+(p/2-3/2)c^2T}.\end{aligned}$$

This upper bound is clearly integrable using property $(\mathcal{P})$ and assumption (III.17) together with the Hölder's inequality. Therefore, by the dominated convergence and under assumption $(\mathcal{H}'_\theta)$, we obtain that

$$\begin{aligned}\lim_{i\to\infty}\mathbb{E}\left[\left|\sqrt{\frac{m^\ell}{((m-1)T)}}(\psi(X^{m^\ell}_T)-\psi(X^{m^{\ell-1}}_T))\right|^p e^{-(p-1)\theta^{m^\ell}_{i-1}\cdot W_T-(p/2-3/2)|\theta^{m^\ell}_{i-1}|^2T}\right] \\ = \mathbb{E}\left[\left|\sqrt{\frac{m^\ell}{((m-1)T)}}(\psi(X^{m^\ell}_T)-\psi(X^{m^{\ell-1}}_T))\right|^p e^{-(p-1)\theta^*_{m^\ell}\cdot W_T-(p/2-3/2)|\theta^*_{m^\ell}|^2T}\right].\end{aligned}$$

Then, by applying Cesaro lemma, we have that

$$\begin{aligned}\lim_{i\to\infty}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\mathbb{E}\left[\left|\sqrt{\frac{m^\ell}{((m-1)T)}}(\psi(X^{m^\ell}_T)-\psi(X^{m^{\ell-1}}_T))\right|^p e^{-(p-1)\theta^{m^\ell}_{i-1}\cdot W_T-(p/2-3/2)|\theta^{m^\ell}_{i-1}|^2T}\right] \\ = \mathbb{E}\left[\left|\sqrt{\frac{m^\ell}{((m-1)T)}}(\psi(X^{m^\ell}_T)-\psi(X^{m^{\ell-1}}_T))\right|^p e^{-(p-1)\theta^*_{m^\ell}\cdot W_T-(p/2-3/2)|\theta^*_{m^\ell}|^2T}\right].\end{aligned}$$

Moreover, we obtain under assumption $(\mathcal{H}'_\theta)$ and thanks to relations (III.28) and (III.29) that

$$\lim_{\ell\to\infty}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\mathbb{E}\left[\left|\sqrt{\frac{m^\ell}{((m-1)T)}}(\psi(X_T^{m^\ell})-\psi(X_T^{m^{\ell-1}}))\right|^p e^{-(p-1)\theta_{i-1}^{m^\ell}\cdot W_T-(p/2-3/2)|\theta_{i-1}^{m^\ell}|^2T}\right]$$
$$=\lim_{\ell\to\infty}\mathbb{E}\left[\left|\sqrt{\frac{m^\ell}{((m-1)T)}}(\psi(X_T^{m^\ell})-\psi(X_T^{m^{\ell-1}}))\right|^p e^{-(p-1)\theta^*_{m^\ell}\cdot W_T-(p/2-3/2)|\theta^*_{m^\ell}|^2T}\right]$$
$$=\tilde{\mathbb{E}}\left([\nabla\psi(X_T)\cdot U_T]^p\, e^{-\theta^*\cdot W_T+\frac{1}{2}|\theta^*|^2T}\right). \quad \text{(III.38)}$$

Once again, by using assumption $(\mathcal{W})$ ($\lim_{\ell\to\infty}\frac{1}{(\sum_{\ell=1}^L a_\ell)^{p/2}}\sum_{\ell=1}^L a_\ell^{p/2}=0$), we conclude that

$$\lim_{\ell\to\infty}\sum_{\ell=1}^{L}\mathbb{E}\,|S_{\ell,N_\ell}|^p=0.$$

This completes the proof. □

Now, we prove a Berry-Essen type bound on our central limit theorem. This improves the relevance of the above result. We set $s_n^2=\mathbb{E}|Q_n^1|^2+\sum_{\ell=1}^L\mathbb{E}|S_{\ell,N_\ell}|^2$ and $\rho_n=\mathbb{E}|Q_n^1|^3+\sum_{\ell=1}^L\mathbb{E}|S_{\ell,N_\ell}|^3$, where Q_n^1 and S_{ℓ,N_ℓ} are respectively given by relations (III.19) and (III.34).

Proposition 3.3. *Under assumptions of Theorem 3.1, we define by F_n the distribution function of $n^\alpha(Q_n-\mathbb{E}\psi(X_T^n))/s_n$. Then, we have*

$$|F_n(x)-G(x)|\le\frac{6\rho_n}{s_n^3},$$

where G is the distribution function of a standard Gaussian random variable. Moreover, if ψ is continuous Lipschitz function, there exists a constant $C>0$ such that

$$|F_n(x)-G(x)|\le\frac{1}{s_n^3}\frac{C}{(\sum_{\ell=1}^L a_\ell)^{3/2}}\sum_{\ell=1}^L a_\ell^{3/2}.$$

For the optimal choice $a_\ell=1$, the obtained Berry-Essen type bound is of order $1/\sqrt{\log n}$.

Proof. Using Theorem 2 page 544 in [27] and since $n^\alpha(Q_n-\mathbb{E}(\psi(X_T^n)))=Q_n^1+Q_n^2=\sum_{\ell=0}^L X_{n,\ell}$, by taking $p=3$ in both inequalities (III.35) and (III.36) in the proof of Theorem 3.2, we deduce a Berry-Essen bound on our central limit theorem.

$$|F_n(x)-G(x)|\le\frac{6\rho_n}{s_n^3}. \quad \text{(III.39)}$$

When ψ is Lipschitz, using property $(\mathcal{P})$, there exists a positive constant C depending on b,σ, T and ψ such that

$$\rho_n \leq \frac{C}{(\sum_{\ell=1}^{L} a_\ell)^{3/2}} \sum_{\ell=1}^{L} a_\ell^{3/2}.$$

Hence, the Berry-Essen type bound on our central limit theorem is given by

$$|F_n(x) - G(x)| \leq \frac{1}{s_n^3} \frac{C}{(\sum_{\ell=1}^{L} a_\ell)^{3/2}} \sum_{\ell=1}^{L} a_\ell^{3/2},$$

and s_n^3 are asymptotically constant. □

4 Complexity analysis and numerical results

According to Theorem 3.2, we deduce that for a total error of order $1/n^\alpha$, $\alpha \in (1/2, 1]$, the minimal computational effort necessary to run the adaptive Multilevel Monte Carlo algorithm is obtained for a sequence of sample sizes specified by relation (III.4) This leads to a time complexity given by

$$\begin{aligned} C_{MLMC} &= C \times \left(N_0 + \sum_{\ell=1}^{L} N_\ell (m^\ell + m^{\ell-1}) \right) \quad \text{with } C > 0 \\ &= C \times \left(\frac{n^{2\alpha}(m-1)T}{a_0} \sum_{\ell=1}^{L} a_\ell + n^{2\alpha} \frac{(m^2-1)T}{m} \sum_{\ell=1}^{L} \frac{1}{a_\ell} \sum_{\ell=1}^{L} a_\ell \right) \quad \text{with } C > 0. \end{aligned}$$

So the time complexity reaches its minimum for the choice of weights $a_\ell^* = 1,\ \ell \in \{1, ..., L\}$. The optimal complexity of the Multilevel Monte Carlo Euler method is given by

$$C_{MLMC} = C \times \left(\frac{(m-1)T}{a_0 \log m} n^{2\alpha} \log n + \frac{(m^2-1)T}{m(\log m)^2} n^{2\alpha} (\log n)^2 \right) = O(n^{2\alpha} (\log n)^2).$$

We conclude that the adaptive Multilevel Monte Carlo method is more efficient in terms of time complexity in comparison with the Monte Carlo method one $C_{MC} = O(n^{2\alpha+1})$.
We consider the problem of the an option pricing under the Heston model. We remind that the the Heston model is a popular stochastic volatility model in finance introduced by Heston in [33] solution to

$$\begin{cases} dS_t = rS_t dt + \sqrt{V_t} S_t dW_t^1 \\ dV_t = \kappa(\bar{v} - V_t)dt + \sigma\sqrt{V_t}\rho dW_t^1 + \sigma\sqrt{V_t}\sqrt{1-\rho^2} dW_t^2, \end{cases}$$

where W^1 and W^2 are two independent Brownian motions. Parameters κ, σ, $\bar{v}$ and r are strictly positive constants and $|\rho| \leq 1$. In this model, κ is the rate at which V_t reverts to $\bar{v}$, $\bar{v}$ is the long run average price variance, σ is the volatility of the variance, r is the interest rate and ρ is a correlation term. Firstly, our aim is to use the importance sampling method in order to reduce the variance when computing the price of an European option, with strike K, under the Heston model. The price is $e^{-rT}\mathbb{E}\psi(S_T)$ where $\psi(S_T) = (S_T - K)_+$ is the payoff of the option. After a density transformation, given by Girsanov theorem, the price will be defined by :

$$e^{-rT}\mathbb{E}\left[\psi(S_T^\theta) \quad e^{-\theta.W_T - \frac{1}{2}|\theta|^2 T}\right], \qquad \theta \in \mathbb{R}^2.$$

The price of the European call option is approximated by

$$e^{-rT}\mathbb{E}\left[g(\theta, S_T^{n,\theta})\right] = e^{-rT}\mathbb{E}\left[\psi(S_T^{n,\theta}) \quad e^{-\theta.W_T - \frac{1}{2}|\theta|^2 T}\right], \qquad \theta \in \mathbb{R}^2.$$

The optimal $\hat{\theta}^*$ of the Multilevel Monte Carlo method is given by

$$\hat{\theta}_n^* = \underset{\theta\in\mathbb{R}^2}{\arg\min}\mathbb{E}\left[\left(\nabla\psi(S_T^{n,\theta}).U_T^{n,\theta}\right)^2 e^{-2\theta.W_T - |\theta|^2 T}\right],$$

where $U^{n,\theta}$ denotes the Euler discretization scheme obtained when we replace coefficients b and σ of relation (III.8) by the corresponding parameters in the Heston model. Here, in order to compare the adaptive Euler Multilevel Monte Carlo algorithm with the adaptive Statistical Romberg and adaptive Monte Carlo ones (studied in Chapter II), we use the constrained algorithm to approximate $\hat{\theta}^*$ which is given by this routine let $(\mathcal{K}_i)_{i\in\mathbb{N}}$ denote an increasing sequence of compact sets satisfying $\cup_{i=0}^{\infty} \mathcal{K}_i = \mathbb{R}^d$ and $\mathcal{K}_i \subsetneq \mathring{\mathcal{K}}_{i+1}, \forall i \in \mathbb{N}$. For $\theta_0^n \in \mathcal{K}_0$, $\alpha_0^n = 0$ and a gain sequence $(\gamma_i)_{i\in\mathbb{N}}$ satisfying (III.13), we define the sequence $(\theta_i^n, \alpha_i^n)_{i\in\mathbb{N}}$ recursively by

$$\begin{cases} \textbf{if} & \theta_{i-1}^n - \gamma_i H^{ML}(\theta_{i-1}^n, S_{T,i}^n, U_{T,i}^n, W_{T,i}) \in \mathcal{K}_{\alpha_{i-1}^n}, \ \textbf{then} \\ & \theta_i^n = \theta_{i-1}^n - \gamma_i H^{ML}(\theta_{i-1}^n, S_{T,i}^n, U_{T,i}^n, W_{T,i}), \text{ and } \alpha_i^n = \alpha_{i-1}^n \\ \textbf{else} & \theta_i^n = \theta_0^n \text{ and } \alpha_i^n = \alpha_{i-1}^n + 1, \end{cases} \tag{III.40}$$

where $H^{ML}(\theta_{i-1}^n, S_{T,i}^n, U_{T,i}^n, W_{T,i}) = (\theta_i^n T - W_{T,i}) \quad (\nabla\psi(S_{T,i}^n).U_{T,i}^n)^2 \quad e^{-\theta_{i-1}^n.W_{T,i} + \frac{1}{2}|\theta_{i-1}^n|^2 T}$. We choose the parameters in the Heston model : $S_0 = 100$, $V_0 = 0.01$, $K = 100$, the free interest rate $r = \log(1.1)$, $\sigma = 0.2$, $k = 2$, $\bar{v} = 0.01$, $\rho = 0.5$ and maturity time $T = 1$. We run a number of iterations $M = 500\,000$ and we obtain the two-dimensional vector of the optimal theta : $\hat{\theta}_n^* = (0.4398, 0.5902)$. Our aim now, is to compare the importance sampling Multilevel Monte Carlo method (denoted MLMCIS) with the two methods already studied in Chapter II, the importance sampling statistical Romberg method (SRIS) and importance sampling Monte Carlo method (MCIS) to compute the price of the European call option :

- MLMC+IS method : European option price approximation with N_ℓ given by expression (III.4)

$$\frac{e^{-rT}}{N_0}\sum_{i=1}^{N_0} g(\hat{\theta}^n_M, S^{m^0,\hat{\theta}^n_M}_{T,i}, U^n_{T,i}, W_{T,i}) + \sum_{\ell=1}^{L}\frac{e^{-rT}}{N_\ell}\sum_{i=1}^{N_\ell}\Big(g(\hat{\theta}^n_M, S^{\ell,m^\ell,\hat{\theta}^n_M}_{T,i}, U^n_{T,i}, W_{T,i}) - g(\hat{\theta}^n_M, S^{\ell,m^{\ell-1},\hat{\theta}^n_M}_{T,i}, U^n_{T,i}, W_{T,i})\Big). \quad \text{(III.41)}$$

- SR+IS method : European option price approximation method with $N_1 = n^2$ and $N_2 = n^{\frac{3}{2}}$

$$\frac{e^{-rT}}{N_1}\sum_{i=1}^{N_1} g(\tilde{\theta}^n_M, \hat{S}^{\sqrt{n},\tilde{\theta}^n_M}_{T,i}, U^n_{T,i}, W_{T,i}) + \frac{e^{-rT}}{N_2}\sum_{i=1}^{N_2}\Big(g(\tilde{\theta}^n_M, S^{n,\tilde{\theta}^n_M}_{T,i}, U^n_{T,i}, W_{T,i}) - g(\tilde{\theta}^n_M, S^{\sqrt{n},\tilde{\theta}^n_M}_{T,i}, U^n_{T,i}, W_{T,i})\Big). \quad \text{(III.42)}$$

- MC+IS method : European option price approximation with $N = n^2$

$$\frac{e^{-rT}}{N}\sum_{i=1}^{N} g(\theta^n_M, S^{n,\theta^n_M}_{T,i}) = \frac{e^{-rT}}{N}\sum_{i=1}^{N}\psi(S^{n,\theta^n_M}_{T,i})e^{-\theta^n_M\cdot W_{T,i}-\frac{1}{2}|\theta^n_M|^2 T}. \quad \text{(III.43)}$$

We consider a set of values $N = 30$, we compute the different estimators of the European call option (III.41), (III.42) and (III.43), for different values of the discretization steps n. For each value n and for each method, we compute the CPU time and the mean squared-error which is given by

$$MSE = \frac{1}{M}\sum_{i=1}^{M}(\text{Real value} - \text{Simulated value}). \quad \text{(III.44)}$$

Hence, for each given n and for each method, we provide a couple of points (MSE, CPU time) which are plotted on Figure III.1.

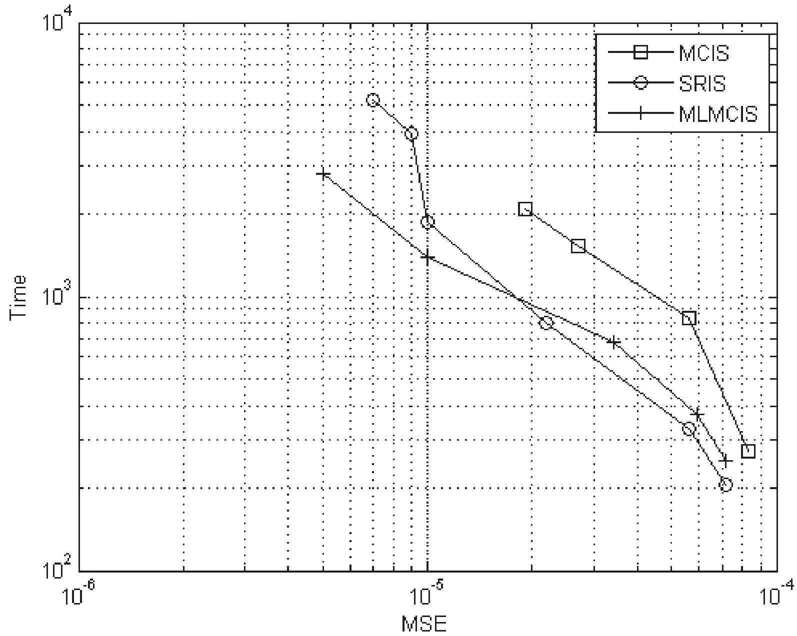

Figure III.1 – *CPU time versus MSE for European call option*

Now let us interpret Figure III.1. The curves of the importance sampling Statistical Romberg (SRIS) and the importance sampling Multilevel Monte Carlo (MLMCIS) methods are displaced below the curve of the importance sampling Monte Carlo (MCIS) method. Therefore, for a given error, the number of values computed in one second by the two methods is larger than the values computed by the MCIS procedure. For a Mean Square Error (MSE) lower than 10^{-5} , we observe that the importance sampling multilevel Monte Carlo procedure becomes more effective than both methods.

5 Conclusion

The central limit theorem derived in this paper confirms the superiority of the adaptive Euler Multilevel Monte Carlo approach. It may be interest to develop analogous results for path dependent options. A next natural question consists on studying central limit theorems for adaptive Multilevel Monte Carlo when using high order discretization schemes for a general setting of stochastic differential equation for a given payoff function.

Chapitre IV

Importance Sampling and Statistical Romberg for Lévy processes

An important family of stochastic processes arising in many areas of applied probability is the class of Lévy processes. Generally, such processes are not simulatable especially for those with infinite activity. In practice, it is common to approximate them by truncating the jumps at some cut-off size ε ($\varepsilon \searrow 0$). This procedure leads us to consider a simulatable compound Poisson process. This paper first introduces, for this setting, the statistical Romberg method to improve the complexity of the classical Monte Carlo one. Roughly speaking, we use many sample paths with a coarse cut-off ε^β, $\beta \in (0,1)$, and few additional sample paths with a fine cut-off ε. Central limit theorems of Lindeberg-Feller type for both Monte Carlo and statistical Romberg method for the inferred errors depending on the parameter ε are proved. This leads to an accurate description of the optimal choice of parameters with explicit limit variances. Afterwards, the authors propose a stochastic approximation method of finding the optimal measure change by Esscher transform for Lévy processes with Monte Carlo and statistical Romberg importance sampling variance reduction. Furthermore, we develop new adaptive Monte Carlo and statistical Romberg algorithms and prove the associated central limit theorems. Finally, numerical simulations are processed to illustrate the efficiency of the adaptive statistical Romberg method that reduces at the same time the variance and the computational effort associated to the effective computation of option prices when the underlying asset process follows an exponential pure jump CGMY model.

1 Introduction

Lévy processes arise in many areas of applied probability and specially in mathematical finance, where they become very fashionable since they can describe the observed reality of financial markets in a more accurate way than models based on Brownian motion (see e.g. Cont and Tankov [20] and Schoutens [61]). In particular in the pricing of financial securities we are interested in the computation of the real quantity $\mathbb{E}F(L_T)$, $T>0$, where $(L_t)_{0\leq t\leq T}$ is a $\mathbb{R}^d$-valued pure jump Lévy process, $d\geq 1$ and $F:\mathbb{R}^d\mapsto\mathbb{R}$ is a given function. In the literature, the computation of this quantity involves three types of methods : Fourier transform methods, numerical methods for partial integral differential equations and Monte Carlo methods. It is well known that the two first methods can not cope with high dimensional problems. This gives a competitive edge for Monte Carlo methods in this setting. Therefore, the focus of this work is to study improved Monte Carlo methods using the statistical Romberg algorithm and the importance sampling technique. The statistical Romberg method is known for reducing the time complexity and the importance sampling technique is aimed at reducing the variance. The Monte Carlo method consists of two steps. In the first step, we approximate the Lévy process $(L_t)_{0\leq t\leq T}$ by a simulatable Lévy process $(L^\varepsilon_t)_{0\leq t\leq T}$ with $\varepsilon>0$. If ν denotes the Lévy measure of the Lévy process under consideration, then it is common to take $(L^\varepsilon_t)_{0\leq t\leq T}$ with Lévy measure $\nu_{|\{|x|\geq\varepsilon\}}$ and $\varepsilon\searrow 0$. This approximation is nothing but a compound Poisson process. In the second step, we approximate $\mathbb{E}\,F\left(L^\varepsilon_T\right)$ by $\frac{1}{N}\sum_{i=1}^N F(L^\varepsilon_{T,i})$, where $(L^\varepsilon_{T,i})_{1\leq i\leq N}$ is a sample of N independent copies of L^ε_T. Therefore, this Monte Carlo method (MC) is affected respectively by an approximation error and a statistical one

$$\mathcal{E}_1(\varepsilon):=\mathbb{E}\left(F(L^\varepsilon_T)-F(L_T)\right) \text{ and } \mathcal{E}_2(N):=\frac{1}{N}\sum_{i=1}^N F(L^\varepsilon_{T,i})-\mathbb{E}F(L^\varepsilon_T).$$

On one hand, for a Lipschitz function F we have $\mathcal{E}_1(\varepsilon)=O(\sigma(\varepsilon))$, where $\sigma^2(\varepsilon)=\mathbb{E}|L_1-L^\varepsilon_1|^2$ (see relation (IV.6) for more details). On the other hand, the statistical error is controlled by the central limit theorem with order $1/\sqrt{N}$. Hence, optimizing the choice of the sample size N in the Monte Carlo method leads to $N=O(\sigma^{-2}(\varepsilon))$. Moreover, if we choose $N=\sigma^{-2}(\varepsilon)$ we prove a central limit theorem of Lindeberg-Feller type (see Theorem 3.1). Therefore, if we denote by $\mathcal{K}(\varepsilon)$ the cost of a single simulation of L^ε_T, then the total time complexity necessary to achieve the precision $\sigma(\varepsilon)$ is given by $C_{MC}=O(\mathcal{K}(\varepsilon)\sigma^{-2}(\varepsilon))$ (see subsection 3.3).

In order to improve the performance of this method we use the idea of the statistical Romberg method introduced by Kebaier [37] in the setting of Euler Monte Carlo methods for stochastic differential equations driven by a standard Brownian Motion which is also related to the well known Romberg's method introduced by Talay and Tubaro in [62]. Inspired by this technique, we introduce a novel method for the computation of our initial target. The main idea of this new method is to consider two cut-off sizes ε and ε^β, $\beta\in(0,1)$ and then approximate

$\mathbb{E}F(L_T)$ by

$$\frac{1}{N_1}\sum_{i=1}^{N_1} F(\hat{L}^{\varepsilon^\beta}_{T,i}) + \frac{1}{N_2}\sum_{i=1}^{N_2} F(L^{\varepsilon}_{T,i}) - F(L^{\varepsilon^\beta}_{T,i}).$$

The samples $(L^{\varepsilon}_{T,i})_{1\le i\le N_2}$ and $(L^{\varepsilon^\beta}_{T,i})_{1\le i\le N_2}$ have to be independent of $(\hat{L}^{\varepsilon}_{T,i})_{1\le i\le N_1}$. Moreover, for $1 \le i \le N_2$, the process $(L^{\varepsilon}_{t,i})_{0\le t\le T}$ is nothing else the sum of $(L^{\varepsilon^\beta}_{t,i})_{0\le t\le T}$ and an independent Lévy process $(L^{\varepsilon,\varepsilon^\beta}_{t,i})_{0\le t\le T}$ with Lévy measure $\nu_{|\{\varepsilon\le|x|\le\varepsilon^\beta\}}$ which is also simulatable as a compound Poisson process. This new method will be referred as the statistical Romberg method (SR). Additionally, like for the MC method, we prove a central limit theorem of Lindeberg-Feller type for the SR algorithm with $N_1 = \sigma^{-2}(\varepsilon)$ and $N_2 = \sigma^{-2}(\varepsilon)\sigma^2(\varepsilon^\beta)$ (see Theorem 3.3). Then, according to subsection 3.3, the total time complexity necessary to achieve the precision $\sigma(\varepsilon)$ is given by $C_{SR} = \left(\mathcal{K}(\varepsilon^\beta) + \mathcal{K}(\varepsilon)\sigma^2(\varepsilon^\beta)\right)\sigma^{-2}(\varepsilon)$. It turns out that the complexity ratio C_{SR}/C_{MC} vanishes as ε goes to zero.

Since the efficiency of the Monte Carlo simulation considerably depends on the smallness of the variance in the estimation, many variance reduction techniques were developed in the recent years. Among these methods appears the technique of importance sampling very popular for its efficiency. For the Gaussian setting, the importance sampling technique was studied by Arouna [3], Galasserman, Heidelberger and Shahabuddin [31] for MC method and by Ben Alaya, Hajji and Kebaier [10] for SR method. Concerning Lévy process without a Brownian component, Kawai [36] has already applied this technique for MC algorithm using the Esscher transform which is nothing but the well known exponential tilting of laws. From a practical point of view, his approach is exploitable only when the Lévy process $(L_t)_{0\le t\le T}$ is simulatable without any approximation. Note also that in his study there is no results on the rate of convergence of the obtained algorithm.

The main aim of the present work is to apply the idea of [36] to the approximation Lévy process $(L^{\varepsilon}_t)_{0\le t\le T}$ for both MC and SR algorithms and to study the inferred error in terms of the cut-off ε ; a question which has not been addressed in previous research. Roughly speaking, thanks to the Esscher transform we produce a parametric transformation such that for all $\theta \in K$ we have $\mathbb{E}F(L^{\varepsilon}_T) = \mathbb{E}G(\theta, L^{\varepsilon}_T)$, where K is a suitable subset of $\mathbb{R}^d$ and $(\theta, x) \mapsto G(\theta, x)$ is a real function taking values in $\mathbb{R}^d \times \mathbb{R}^d$. Concerning the MC method it looks natural to implement the method with $\theta^*_{1,\varepsilon} = \arg\min_{\theta\in K} \mathbb{E}G^2(\theta, L^{\varepsilon}_T)$. However, for the SR method the inferred error is controlled by $\mathrm{Var}(G(\theta, L^{\varepsilon}_T)) + T\mathbb{E}(|\nabla_x G(\theta, L^{\varepsilon}_T)|^2)$. Then, in this case, it is natural to implement the first (resp. the second) empirical mean appearing in the SR estimator with $\theta^*_{1,\varepsilon}$ (resp. $\theta^*_{2,\varepsilon} = \arg\min_{\theta\in K} \mathbb{E}(|\nabla_x G(\theta, L^{\varepsilon}_T)|^2)$. But what about the effective computation of $(\theta^*_{i,\varepsilon})_{i\in\{1,2\}}$? To answer this question, we use a constrained version of the well-known stochastic approximation Robbins-Monro. All these ideas led us to introduce two new methods based on adaptive approximations. The first method concerns a combination of an adaptive importance sampling technique and the MC method that will be called Importance Sampling Monte Carlo method (ISMC) (see relation (IV.22)). The second one concerns an original combination of an adaptive importance sampling technique with the SR algorithm that will be referred as Importance Sampling Statistical Romberg method (ISSR) (see relation (IV.26)). The main point in favor of the ISSR method is that it inherits the variance reduction from the Importance sampling procedure and the complexity reduction from the SR method. A complexity analysis is also provided.

The rest of the paper is organized as follows. Section 2 introduces the general framework and recalls some useful results. In section 3, the central limit theorems of Lindeberg-Feller type are proved for both MC and SR methods (see Theorems 3.1 and 3.3). Similar results are derived for the setting of an exponential Lévy model (see Corollaries 3.2 and 3.4). A complexity analysis is included. In section 4, we recall the Esscher transform and the principle of importance sampling technique for the SR method. For $i \in \{1,2\}$ and $\varepsilon \searrow 0$, we prove the convergence of the optimal choice $\theta^*_{i,\varepsilon}$ to the optimal choice associated to the limit model (see Theorem 4.4). In section 5, we first study, for $i \in \{1,2\}$, the almost sure convergence of the stochastic recursive constrained Robbins-Monro algorithm given by the double indexed sequence $\theta_{i,\varepsilon,n}$ as $\varepsilon \searrow 0$ and $n \nearrow \infty$ (see Theorems 5.1 and 5.2 and Corollary 5.3). The rest of this section is devoted to prove the central limit theorems of Lindeberg-Feller type for both adaptive ISMC and ISSR methods (see Theorems 5.5 and 5.7). Section 6 illustrates the superiority of the ISSR method over all the other ones via numerical examples for both one and two-dimensional Carr, Geman, Madan and Yor (CGMY) process [16]. Finally, the last Section is devoted to discuss some future openings.

2 General Framework

We denote by $(\Omega, \mathcal{F}, \mathbb{P})$ our underlying probability space. A stochastic process $(L_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^d$ such that $L_0 = 0$ is a Lévy process if it has independent and stationary increments. We endow the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the canonical filtration $(\mathcal{F}_t)_{0\leq t\leq T}$ where $\mathcal{F}_t = \sigma(L_s, s \leq t)$. The characteristic function of a Lévy process L with generating triplet (γ, A, ν) is given by the well known Lévy Kintchine representation

$$\mathbb{E}e^{iu.L_t} = \exp\left\{t\left(i\gamma.u - \frac{1}{2}u.Au + \int_{\mathbb{R}^d}(e^{iu.x} - 1 - iu.x\mathbf{1}_{|x|\leq 1})\nu(dx)\right)\right\}, \quad u \in \mathbb{R}^d,$$

where $\gamma \in \mathbb{R}^d$, A is a symmetric non-negative-definite $d \times d$ matrix and ν is a Lévy measure on $\mathbb{R}^d \setminus \{0\}$ verifying $\int_{\mathbb{R}^d\setminus\{0\}}(|x|^2 \wedge 1)\nu(dx) < \infty$. (Given vectors x and $y \in \mathbb{R}^d$, $x.y$ denotes the inner product of x and y associated to the Euclidean norm $|\cdot|$). In this paper, we are interested in studying pure-jump Lévy processes, that is, we set $A \equiv 0$ throughout all the paper.Then, $(L_t)_{t\geq 0}$ is a Lévy process with generating triplet $(\gamma, 0, \nu)$. The simulation of a Lévy process with infinite Lévy measure is not straightforward. From the Lévy-Itô decomposition (see e.g. Theorem 19.2 in Sato [60]), we know that L can be represented as a sum of a compound Poisson process and an almost sure limit of compensated compound Poisson process $L_t = \lim_{\varepsilon\to 0} L_t^\varepsilon$ *a.s.* where for $0 < \varepsilon < 1$

$$L_t^\varepsilon = \gamma t + \sum_{0<s\leq t} \Delta L_s \mathbf{1}_{|\Delta L_s|>1} + (\sum_{0<s\leq t} \Delta L_s \mathbf{1}_{\varepsilon\leq|\Delta L_s|\leq 1} - t\int_{\varepsilon\leq|x|\leq 1} x\nu(dx)), \quad t \geq 0. \tag{IV.1}$$

Note that without the compensation $t\int_{\varepsilon\leq|x|\leq 1} x\nu(dx)$, the sum of jumps $\sum_{0<s\leq t} \Delta L_s \mathbf{1}_{\varepsilon\leq|\Delta L_s|\leq 1}$ may not converge as ε goes to zero. We denote the approximation error by

$$R^\varepsilon = L - L^\varepsilon. \tag{IV.2}$$

The process R^ε is also a Lévy process independent of L^ε with characteristic function

$$\mathbb{E}e^{iu.R_t^\varepsilon} = \exp\left\{t\int_{|x|\leq\varepsilon}(e^{iu.x} - 1 - iu.x)\nu(dx)\right\}.$$

Consequently, $\mathbb{E}[R_t^\varepsilon] = 0$ and the variance-covariance matrix $\mathbb{E}[R_t^\varepsilon(R_t^\varepsilon)'] = t\Sigma_\varepsilon$ where

$$\Sigma_\varepsilon = \int_{|x|\leq\varepsilon} xx'\nu(dx).$$

(A' denotes the transpose of a matrix A). The asymptotic behavior of the distribution of R^ε is firstly studied by Asmussen and Rosiński [5] in the one dimensional case and later extended to the multidimensional case by Cohen and Rosiński [19]. Throughout this paper $W = (W_t)_{t\geq 0}$ is a standard Brownian motion in $\mathbb{R}^d$ independent of $(L_t)_{t\geq 0}$.

Theorem 2.1. *Under the above notation, suppose that Σ_ε is invertible for every $\varepsilon \in (0,1]$. Then as $\varepsilon \to 0$,*

$$\Sigma_\varepsilon^{-1/2} R^\varepsilon \Rightarrow W,$$

if and only if for each $k > 0$

$$\lim_{\varepsilon \to 0} \int_{\langle \Sigma_\varepsilon^{-1} x, x \rangle > k} \langle \Sigma_\varepsilon^{-1} x, x \rangle \mathbf{1}_{|x| \leq \varepsilon} \nu(dx) = 0. \tag{IV.3}$$

Here "$\Rightarrow$*" stands for the convergence in distribution.*

If ν is given in polar coordinates by $\nu(dr, du) = \mu(dr|u)\lambda(du), \quad r > 0, u \in S^{d-1}$, where $\{\mu(\cdot|u) : u \in S^{d-1}\}$ is a measurable family of Lévy measures on $(0, \infty)$ and λ is a finite measure on the unit sphere S^{d-1}, then

$$\Sigma_\varepsilon = \int_{S^{d-1}} \int_0^\varepsilon r^2 u u' \mu(dr|u) \lambda(du).$$

If we define $\sigma^2(\varepsilon, u) := \int_0^\varepsilon r^2 \mu(dr|u)$ and $\sigma^2(\varepsilon) := \int_{S^{d-1}} \sigma^2(\varepsilon, u)\lambda(du)$, then

$$\mathbb{E}|L_t - L_t^\varepsilon|^2 = t\mathrm{Tr}(\Sigma_\varepsilon) = t\sigma^2(\varepsilon). \tag{IV.4}$$

Remark 2.2. *In the one dimensional case Assmussen and Rosiński [5] have obtained the convergence of* $\sigma^{-1}(\varepsilon)R^\varepsilon$ *to a standard Brownian motion if and only if for each* $k > 0$, $\sigma(k\sigma(\varepsilon) \wedge \varepsilon) \sim \sigma(\varepsilon)$ *which is satisfied as soon as* $\lim_{\varepsilon \to 0} \frac{\sigma(\varepsilon)}{\varepsilon} = \infty$ *(see Theorem 2.1 and Proposition 2.1 in [5]). An extension to this sufficient condition in the multidimensional case is given by Theorem 2.5 in Cohen and Rosiński [19]. Suppose that the support of the measure* λ *is not contained in any proper linear subspace of* $\mathbb{R}^d$*, they proved that if*

$$\lim_{\varepsilon \to 0} \frac{\sigma(\varepsilon, u)}{\varepsilon} = \infty, \lambda - a.e. \tag{IV.5}$$

then Σ_ε *is invertible and condition (IV.3) of Theorem 2.1 holds.*

On the other hand, according to Proposition 2.1 of Dia [21], we have a L^q-upper bound of the error approximation in the one dimensional case for any real $q > 0$. This result on the strong error approximation remains valid for the multidimensional case. More precisely, if we consider the d-dimensional error Lévy process R^ε given by relation (IV.2), then we can easily deduce that

$$\mathbb{E}|R_t^\varepsilon|^q \leq K_{q,T}\sigma_0(\varepsilon)^q, \quad \text{where } K_{q,T} > 0 \text{ and } \sigma_0(\varepsilon) = \sigma(\varepsilon) \vee \varepsilon. \tag{SE}$$

Concerning the weak error, if F denotes a real valued Lipschitz continuous function with Lipschitz constant $C > 0$, then it is easy to see that

$$|\mathbb{E}F(L_T) - \mathbb{E}F(L_T^\varepsilon)| \leq C\sqrt{T}\sigma(\varepsilon) \tag{IV.6}$$

Moreover, under some regularity conditions on function F we can obtain an expansion of the weak error as in Proposition 2.2 and Remark 2.3 of [21]. So, it is worth to introduce the following assumption : there exist $C_F \in \mathbb{R}$ and $\upsilon_\varepsilon \searrow 0$ as $\varepsilon \searrow 0$ such that

$$\upsilon_\varepsilon^{-1}\left(\mathbb{E}F(L_T) - \mathbb{E}F(L_T^\varepsilon)\right) \to C_F \quad \text{as } \varepsilon \searrow 0. \tag{WE$_{\upsilon_\varepsilon}$}$$

We recall, in what follows, an important moment property of Lévy processes. For this, we introduce before the below definition.

Definition 2.3. *A function $f : \mathbb{R}^d \mapsto [0, \infty)$ is said to be submultiplicative if there exists a positive constant c such that $f(x + y) \leq cf(x)f(y)$ for $x, y \in \mathbb{R}^d$. The product of two submultiplicative functions is also submultiplicative.*

Theorem 2.4 (Sato [60], Theorem 25.3)**.** *Let f be a submultiplicative, locally bounded, measurable function on $\mathbb{R}^d$, and let $(L_t)_{t\geq 0}$ be a Lévy process in $\mathbb{R}^d$ with Lévy measure ν. Then, $\mathbb{E}f(L_t)$ is finite for every $t > 0$ if and only if $\int_{|z|\geq 1} f(z)\nu(dz) < +\infty$.*

3 Statistical Romberg method and Lévy process

In this section, we establish two central limit theorems of Lindeberg-Feller type, for the inferred errors associated to MC and SR algorithms, in terms of the cut-off ε. Similar results are derived for the setting of an exponential Lévy model. We also provide a complexity analysis for both algorithms.

3.1 Central limit theorem for the MC method

Theorem 3.1. *Let $F : \mathbb{R}^d \to \mathbb{R}$ be a continuous function satisfying assumption (WE$_{\upsilon_\varepsilon}$). If $\sup_{0<\varepsilon\leq 1} \mathbb{E}\left[F^{2a}(L_T^\varepsilon)\right] < +\infty$ for $a > 1$, then for $N = \upsilon_\varepsilon^{-2}$ we have*

$$\upsilon_\varepsilon^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} F(L_{T,i}^\varepsilon) - \mathbb{E}F(L_T)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(C_F, \mathrm{Var}(F(L_T))) \quad \text{as } \varepsilon \searrow 0. \tag{IV.7}$$

Proof. At first, we write the total error as follows

$$\frac{1}{N}\sum_{i=1}^{N}F(L^{\varepsilon}_{T,i})-\mathbb{E}F(L_T)=\frac{1}{N}\sum_{i=1}^{N}F(L^{\varepsilon}_{T,i})-\mathbb{E}F(L^{\varepsilon}_T)+\left(\mathbb{E}F(L^{\varepsilon}_T)-\mathbb{E}F(L_T)\right).$$

Assumption ($\mathrm{WE}_{\upsilon_\varepsilon}$) ensures that $\lim_{\varepsilon\to 0}\upsilon_\varepsilon^{-1}\mathbb{E}\left(F(L^{\varepsilon}_T)-F(L_T)\right)=C_F$. Concerning the first term on the right hand side of the above relation, as N depends on ε we plan to apply the Lindeberg-Feller central limit theorem (see Theorem 1.3). In order to do that, we set $X_{i,\varepsilon}:=\frac{\upsilon_\varepsilon^{-1}}{N}\left(F(L^{\varepsilon}_{T,i})-\mathbb{E}F(L^{\varepsilon}_T)\right)$ and we check assumptions *A1* and *A3* of Theorem 1.3. Thus, the proof is divided into two steps.

Step 1. For assumption *A1*, it is straightforward that $\sum_{i=1}^{N}\mathbb{E}(X^2_{i,\varepsilon})=\mathrm{Var}(F(L^{\varepsilon}_T))$. Then, by the almost sure convergence of L^{ε}_T toward L_T, the continuity of function F and the uniform integrability condition given by $\sup_{0<\varepsilon\le 1}\mathbb{E}\left[F^{2a}(L^{\varepsilon}_T)\right]<+\infty$, we obtain

$$\lim_{\varepsilon\to 0}\sum_{i=1}^{N}\mathbb{E}(X^2_{i,\varepsilon})=\lim_{\varepsilon\to 0}\mathrm{Var}(F(L^{\varepsilon}_T))=\mathrm{Var}(F(L_T)). \tag{IV.8}$$

Step 2. Concerning the Lyapunov condition *A3*, for $1<\tilde{a}<a$, we have

$$\sum_{i=1}^{N}\mathbb{E}\left[|X_{i,\varepsilon}|^{2\tilde{a}}\right]=\upsilon_\varepsilon^{2(\tilde{a}-1)}\mathbb{E}\left|F(L^{\varepsilon}_T)-\mathbb{E}F(L^{\varepsilon}_T)\right|^{2\tilde{a}}.$$

Once again by the same arguments used in the previous step we prove the convergence of $\mathbb{E}\left|F(L^{\varepsilon}_T)-\mathbb{E}F(L^{\varepsilon}_T)\right|^{2\tilde{a}}$ toward $\mathbb{E}\left|F(L_T)-\mathbb{E}F(L_T)\right|^{2\tilde{a}}$ as ε tends to zero. Since $\upsilon_\varepsilon^{2(\tilde{a}-1)}\underset{\varepsilon\to 0}{\longrightarrow}0$, we obtain

$$\lim_{\varepsilon\to 0}\sum_{i=1}^{N}\mathbb{E}\left[|X_{i,\varepsilon}|^{2\tilde{a}}\right]=0. \tag{IV.9}$$

By (IV.8) and (IV.9), we obtain thanks to Theorem 1.3 the desired convergence in law. □

In the corollary below, we will treat the special case where $F(x)=f(e^{x_1},\cdots,e^{x_d})$ for all $x=(x_1,\cdots,x_d)\in\mathbb{R}^d$ and $f:\mathbb{R}^d_+\to\mathbb{R}$ is a Lipschitz continuous function. In finance this model is well known as an exponential Lévy model.

Corollary 3.2. *Assume that* $\int_{|z|>1}e^{2a|z|}\nu(dz)$ *is finite for* $a>1$*. Then, in the setting of an exponential Lévy model there is* $C>0$ *such that* $|\mathbb{E}F(L_T)-\mathbb{E}F(L^{\varepsilon}_T)|\le C\sigma(\varepsilon)$*. Moreover, if we choose* $N=\sigma^{-2+\eta}(\varepsilon)$*, with* $0<\eta<2$*, then*

$$\sigma^{-1+\eta/2}(\varepsilon)\left(\frac{1}{N}\sum_{i=1}^{N}F(L^{\varepsilon}_{T,i})-\mathbb{E}F(L_T)\right)\xrightarrow{\mathcal{L}}\mathcal{N}(0,\mathrm{Var}(F(L_T)))\quad \text{as } \varepsilon\searrow 0. \tag{IV.10}$$

Proof. We denote by e^x the exponential function element-wise of the vector $x = (x_1, \cdots, x_d) \in \mathbb{R}^d$, $e^x = (e^{x_1}, \cdots, e^{x_d})$. Let C_f denote the Lipschitz constant of function f, since L_T^1 and $(L_T - L_T^1, L_T^\varepsilon - L_T^1)$ are independent we obtain by standard calculations

$$\begin{aligned} |\mathbb{E}F(L_T) - \mathbb{E}F(L_T^\varepsilon)| &\leq C_f \mathbb{E} e^{\left|L_T^1\right|} \mathbb{E}\left|L_T - L_T^\varepsilon\right| \left(e^{\left|L_T - L_T^1\right|} + e^{\left|L_T^\varepsilon - L_T^1\right|}\right) \\ &\leq C_f \sigma(\varepsilon) \mathbb{E} e^{\left|L_T^1\right|} \left(\left\|e^{|L_T - L_T^1|}\right\|_2 + \left\|e^{|L_T^\varepsilon - L_T^1|}\right\|_2\right). \end{aligned}$$

Now, on the one hand thanks to Theorem 2.4, the assumption $\int_{|z|>1} e^{2a|z|}\nu(dz) < +\infty$ ensures the finiteness of $\mathbb{E}e^{\left|L_T^1\right|}$. On the other hand by virtue of Lemmas 25.6 and 25.7 in Sato [60] we have the boundedness of $\left\|e^{|L_T - L_T^1|}\right\|_2$. Concerning the term $\left\|e^{|L_T^\varepsilon - L_T^1|}\right\|_2$, we have $e^{|x|} \leq \prod_{j=1}^d (e^{x_j} + e^{-x_j})$, this last upper bound can be written as a sum of finite number of exponential functions evaluated at points which are a linear combination of the components of the vector x. Therefore there exists a family of $\mathbb{R}^d$-valued vectors, $(b_j)_{1\leq j\leq 2^d}$ such that

$$\left\|e^{|L_T^\varepsilon - L_T^1|}\right\|_2^2 \leq \sum_{j=1}^{2^d} \exp\left\{T \int_{\varepsilon\leq|x|\leq 1} (e^{b_j.x} - 1 - b_j.x)\nu(dx)\right\}.$$

Note that the finiteness of the above upper bound is once again ensured by Lemmas 25.6 and 25.7 in Sato [60]. Since its limit exists we deduce that $\sup_{0<\varepsilon\leq 1}\left\|e^{|L_T^\varepsilon - L_T^1|}\right\|_2$ is finite. Now, thanks to the linear growth of f and using the same arguments as above we check in the same manner the property $\sup_{0<\varepsilon\leq 1} \mathbb{E}\left[F^{2a}(L_T^\varepsilon)\right] < +\infty$. Hence, if we choose $\upsilon_\varepsilon = \sigma^{1-\eta/2}(\varepsilon)$ then Theorem 3.1 applies and this completes the proof. □

3.2 Central limit theorem for the SR method

We use the SR method to approximate $\mathbb{E}[F(L_T)]$ by

$$Q_\varepsilon = \frac{1}{N_1}\sum_{i=1}^{N_1} F(L_{T,i}^{\varepsilon^\beta}) + \frac{1}{N_2}\sum_{i=1}^{N_2}\left(F(L_{T,i}^\varepsilon) - F(L_{T,i}^{\varepsilon^\beta})\right)$$

Theorem 3.3. *Let $F : \mathbb{R}^d \to \mathbb{R}$ be a $\mathscr{C}^1$ function satisfying assumption ($\mathrm{WE}_{\upsilon_\varepsilon}$) and such that $\sup_{0<\varepsilon\leq 1}\mathbb{E}F^{2a}(L_T^\varepsilon)$ and $\sup_{0<\varepsilon\leq 1}\mathbb{E}\left|\sigma^{-1}(\varepsilon)(F(L_T^\varepsilon) - F(L_T))\right|^{2a}$ are finite, for $a > 1$. Moreover, assume that*

H1. Condition (IV.3) in Theorem 2.1 holds and there exists a definite positive matrix Σ such that $\lim_{\varepsilon\to 0}\sigma^{-2}(\varepsilon)\Sigma_\varepsilon = \Sigma$.

H2. For $0 < \beta < 1$, we have $\lim_{\varepsilon\to 0}\sigma(\varepsilon)\sigma^{-1}(\varepsilon^\beta) = 0$ and $\lim_{\varepsilon\to 0}\upsilon_\varepsilon\sigma^{-1}(\varepsilon^\beta) = 0$.

If we choose $N_1 = \upsilon_\varepsilon^{-2}$ *and* $N_2 = \upsilon_\varepsilon^{-2}\sigma^2(\varepsilon^\beta)$*, then*

$$\upsilon_\varepsilon^{-1}\left(Q_\varepsilon - \mathbb{E}F(L_T)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\Big(C_F, \mathrm{Var}(F(L_T)) + T\mathbb{E}(\nabla F(L_T).\Sigma\nabla F(L_T))\Big), \quad \text{as } \varepsilon \searrow 0.$$

Proof. At first we write the total error as $Q_\varepsilon - \mathbb{E}F(L_T) = Q_\varepsilon^1 + Q_\varepsilon^2 + \mathbb{E}F(L_T) - \mathbb{E}F(L_T^\varepsilon)$, with

$$Q_\varepsilon^1 = \frac{1}{N_1}\sum_{i=1}^{N_1} F(L_{T,i}^{\varepsilon^\beta}) - \mathbb{E}F(L_T^{\varepsilon^\beta}) \text{ and } Q_\varepsilon^2 = \frac{1}{N_2}\sum_{i=1}^{N_2} F(L_{T,i}^{\varepsilon}) - F(L_{T,i}^{\varepsilon^\beta}) - \mathbb{E}\left[F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta})\right].$$

So, assumption ($\mathrm{WE}_{\upsilon_\varepsilon}$) yields the convergence of $\upsilon_\varepsilon^{-1}\left(\mathbb{E}F(L_T) - \mathbb{E}F(L_T^\varepsilon)\right)$ toward C_F as ε goes to zero and following step by step the proof of Theorem 3.1 the convergence law of $\upsilon_\varepsilon^{-1}Q_\varepsilon^1$ to the normal distribution $\mathcal{N}(0, \mathrm{Var}(F(L_T)))$ is easily obtained. Concerning the term Q_ε^2, we plan to use Theorem 1.3 and we set $X_{i,\varepsilon} := \frac{\upsilon_\varepsilon^{-1}}{N_2}\left(F(L_{T,i}^\varepsilon) - F(L_{T,i}^{\varepsilon^\beta}) - \left(\mathbb{E}F(L_T^\varepsilon) - \mathbb{E}F(L_T^{\varepsilon^\beta})\right)\right)$. In the following two steps, we will check assumptions *A1* and *A3* of Theorem 1.3.

Step 1. It is straightforward that $\sum_{i=1}^{N_2}\mathbb{E}(X_{i,\varepsilon}^2) = \sigma^{-2}(\varepsilon^\beta)\mathrm{Var}(F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta}))$. Now applying Taylor-Young's expansion to the real valued $\mathcal{C}^1$ function F we get

$$F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta}) = \nabla F(L_T^{\varepsilon^\beta}).(L_T^\varepsilon - L_T^{\varepsilon^\beta}) + (L_T^\varepsilon - L_T^{\varepsilon^\beta}).\epsilon(L_T^\varepsilon - L_T^{\varepsilon^\beta}),$$

where $\epsilon(L_T^\varepsilon - L_T^{\varepsilon^\beta}) \xrightarrow{a.s.} 0$ as $\varepsilon \to 0$. Now, by applying twice Theorem 2.1 to $L_T^\varepsilon - L_T$ and $L_T - L_T^{\varepsilon^\beta}$ and thanks to assumption *H2* we obtain $\sigma^{-1}(\varepsilon^\beta)\Big(L_T^\varepsilon - L_T^{\varepsilon^\beta}\Big) \xrightarrow[\varepsilon\to 0]{\mathcal{L}} \Sigma^{1/2}W_T$. Since $L_T^{\varepsilon^\beta}$ is independent from $L_T^\varepsilon - L_T^{\varepsilon^\beta}$ and $\nabla F(L_T^{\varepsilon^\beta}) \xrightarrow[\varepsilon\to 0]{a.s.} \nabla F(L_T)$, we obtain

$$\sigma^{-1}(\varepsilon^\beta)\left(F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta})\right) \xrightarrow[\varepsilon\to 0]{\mathcal{L}} \nabla F(L_T).\Sigma^{1/2}W_T \qquad \text{(IV.11)}$$

For the second term, using the tightness of $\sigma^{-1}(\varepsilon^\beta)\Big(L_T^\varepsilon - L_T^{\varepsilon^\beta}\Big)$ we deduce that $\sigma^{-1}(\varepsilon^\beta)\Big(L_T^\varepsilon - L_T^{\varepsilon^\beta}\Big)\epsilon(L_T^\varepsilon - L_T^{\varepsilon^\beta}) \xrightarrow[\varepsilon\to 0]{a.s.} 0$. Thanks to the inequality $|x+y|^{2a} \le 2^{2a-1}(|x|^{2a} + |y|^{2a})$, for any $x, y \in \mathbb{R}$, $\sup_{0<\varepsilon\le 1}\mathbb{E}\left|\sigma^{-1}(\varepsilon)(F(L_T^\varepsilon) - F(L_T))\right|^{2a} < +\infty$ and $\lim_{\varepsilon\to 0}\sigma(\varepsilon)\sigma^{-1}(\varepsilon^\beta) = 0$, we deduce the uniform integrability of $\sigma^{-2}(\varepsilon^\beta)|F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta})|^2$. Therefore, we obtain the first condition

$$\lim_{\varepsilon\to 0}\sum_{i=1}^{N_2}\mathbb{E}(X_{i,\varepsilon})^2 = \mathrm{Var}(\nabla F(L_T).\Sigma^{1/2}W_T) = T\mathbb{E}(\nabla F(L_T).\Sigma\nabla F(L_T)).$$

Step 2. For the Lyapunov condition, let $1 < a' < a$, we get by standard evaluations

$$\sum_{i=1}^{N_2} \mathbb{E}|X_{i,\varepsilon}|^{2a'} \leq 2^{2a'} \upsilon_\varepsilon^{2(a'-1)} \sigma^{-2(a'-1)}(\varepsilon^\beta) \mathbb{E}\left|\sigma^{-1}(\varepsilon^\beta)(F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta}))\right|^{2a'}.$$

Once again we use the convergence in distribution given by relation (IV.11) and the uniform integrability property $\sup_{0<\varepsilon\leq 1} \mathbb{E}\left|\sigma^{-1}(\varepsilon^\beta)(F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta}))\right|^{2a} < +\infty$ to deduce the convergence of $\mathbb{E}\left|\sigma^{-1}(\varepsilon^\beta)(F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta}))\right|^{2a'}$ toward $\mathbb{E}\left|\nabla F(L_T).\Sigma^{1/2}W_T\right|^{2a'}$. Finally, since $\lim_{\varepsilon\to 0} \upsilon_\varepsilon \sigma^{-1}(\varepsilon^\beta) = 0$, we conclude that $\lim_{\varepsilon\to 0} \sum_{i=1}^{N_2} \mathbb{E}|X_{i,\varepsilon}|^{2a'} = 0$ with $a' > 1$. This gives the asymptotic normality of Q_ε^2 and completes the proof. □

Now, we get back to the exponential Lévy model setting introduced before Corollary 3.2 where $F(x) = f(e^{x_1}, \cdots, e^{x_d})$ for a given $\mathscr{C}^1$ Lipschitz continuous function f. Our aim is to deduce in this setting a central limit theorem for SR method.

Corollary 3.4. *Assume that $\int_{|z|>1} e^{2a|z|}\nu(dz)$ is finite for $a > 1$. In the setting of an exponential Lévy model there is $C > 0$ such that $|\mathbb{E}F(L_T) - \mathbb{E}F(L_T^\varepsilon)| \leq C\sigma(\varepsilon)$. Moreover, assume that for $0 < \beta < 1$ there exists $0 < \eta < 2$ such that $\lim_{\varepsilon\to 0} \sigma^{1-\eta/2}(\varepsilon)\sigma^{-1}(\varepsilon^\beta) = 0$, $\sigma(\varepsilon) > \varepsilon$ for all $0 < \varepsilon < 1$ and condition H1 of Theorem 3.3 is satisfied. Then, if we choose $N_1 = \sigma^{-2+\eta}(\varepsilon)$ and $N_2 = \sigma^{-2+\eta}(\varepsilon)\sigma^{-1}(\varepsilon^\beta)$ we obtain*

$$\sigma^{-1+\eta/2}\left(Q_\varepsilon - \mathbb{E}F(L_T)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \mathrm{Var}(F(L_T)) + T\mathbb{E}(\nabla F(L_T).\Sigma\nabla F(L_T))\Big), \quad \textit{as } \varepsilon \searrow 0.$$

Proof. According to Theorem 3.3 and Corollary 3.2 we only need to check that assumption $\sup_{0<\varepsilon\leq 1} \mathbb{E}\left|\sigma^{-1}(\varepsilon)(F(L_T^\varepsilon) - F(L_T))\right|^{2a} < \infty$ is satisfied. Since f is Lipschitz it is sufficient to find an upper bound for $\mathbb{E}\left|e^{L_T^\varepsilon} - e^{L_T}\right|^{2a}$. To do so, we use the independence of L_T^1 and the couple $(L_T - L_T^1, L_T^\varepsilon - L_T^1)$ and Cauchy-Schwartz's inequality to get

$$\mathbb{E}\left|e^{L_T^\varepsilon} - e^{L_T}\right|^{2a} \leq \mathbb{E}e^{2a|L_T^1|}\left\||L_T - L_T^\varepsilon|^{2a}\right\|_2 \left(\left\|e^{2a|L_T - L_T^1|}\right\|_2 + \left\|e^{2a|L_T^\varepsilon - L_T^1|}\right\|_2\right).$$

By the same arguments given in the proof of Corollary 3.2 we have the finiteness of $\mathbb{E}e^{2a|L_T^1|}$, $\left\|e^{2a|L_T - L_T^1|}\right\|_2$ and $\sup_{0<\varepsilon\leq 1}\left\|e^{2a|L_T^\varepsilon - L_T^1|}\right\|_2$. Combining all these results together with assumption (SE) we deduce the existence of a constant $C > 0$ not depending on ε such that

$$\mathbb{E}\left|\sigma^{-1}(\varepsilon)(F(L_T^\varepsilon) - F(L_T))\right|^{2a} \leq C\sigma^{-2a}(\varepsilon)\sigma_0^{2a}(\varepsilon).$$

This completes the proof since $\sigma_0(\varepsilon) = \sigma(\varepsilon)$, for $0 < \varepsilon < 1$. □

3.3 Complexity Analysis

Thanks to the above limit results we are able now to provide a complexity analysis for both MC and SR algorithm. To keep things simple, we consider the particular case $d = 1$, $v_\varepsilon = \sigma(\varepsilon)$ and we assume that the measure ν has a density of the form $L(x)/|x|^{Y+1}$ for a small x, where $L(x)$ is a slowly varying as $x \to 0$ and $Y \in (0, 2)$. Observe that the positive (resp. negative) part of the approximation $(L_t^\varepsilon)_{0\leq t\leq T}$ is essentially a compound Poisson process with intensity $\nu([\varepsilon, +\infty))$ (resp. $\nu((-\infty, -\varepsilon]))$. Then, the cost necessary of a single simulation is random, with expectation of order $\mathcal{K}(\varepsilon) = \nu(|x| \geq \varepsilon)$. Hence, according to Theorem 3.1 the time complexity of the MC method necessary to achieve a total error of order $\sigma(\varepsilon)$ is random with expectation of order

$$C_{MC} = \mathcal{K}(\varepsilon)N = \mathcal{K}(\varepsilon)\sigma^{-2}(\varepsilon).$$

In the same way, thanks to Theorem 3.3 the time complexity of the SR method necessary to achieve a total error of order $\sigma(\varepsilon)$ is random with expectation of order

$$C_{SR} = \mathcal{K}(\varepsilon^\beta)N_1 + \mathcal{K}(\varepsilon)N_2 = \left(\mathcal{K}(\varepsilon^\beta) + \mathcal{K}(\varepsilon)\sigma^2(\varepsilon^\beta)\right)\sigma^{-2}(\varepsilon).$$

By Karamata's theorem (see e.g. Bingham, Goldie and Teugels [14] or Feller [27])

$$\sigma^2(\varepsilon) = \int_{-\varepsilon}^{\varepsilon} |x|^{1-Y} L(x)dx \sim \frac{L(\varepsilon) + L(-\varepsilon)}{2 - Y}\varepsilon^{2-Y}.$$

Similarly we have

$$\mathcal{K}(\varepsilon) \sim \frac{L(\varepsilon) + L(-\varepsilon)}{Y}\varepsilon^{-Y}.$$

Consequently, we compute the time complexity ratio given by

$$\frac{C_{SR}}{C_{MC}} = \frac{L(\varepsilon^\beta) + L(-\varepsilon^\beta)}{L(\varepsilon) + L(-\varepsilon^\beta)}\varepsilon^{Y(1-\beta)} + \frac{L(\varepsilon^\beta) + L(-\varepsilon^\beta)}{2 - Y}\varepsilon^{\beta(2-Y)}.$$

If $L(\varepsilon)$ is constant in the neighborhood of zero, like for the CGMY model (see relation (IV.28)), then we easily get

$$\frac{C_{SR}}{C_{MC}} = O\left(\varepsilon^{Y(1-\beta)} + \varepsilon^{\beta(2-Y)}\right).$$

Optimizing the order of this last quantity yields $\beta = Y/2$ which leads us to a gain of a complexity of order $\varepsilon^{Y(Y/2-1)}$ that asymptotically increases as soon as ε becomes small.

4 Importance Sampling and Statistical Romberg method

Let $\{L_t; t \geq 0\}$ be a Lévy process in $\mathbb{R}^d$ under the probability $\mathbb{P}$ with generating triplet $(\gamma, 0, \nu)$. We define the set

$$\Theta_1 := \Big\{\theta \in \mathbb{R}^d : \mathbb{E}[e^{\theta.L_t}] < +\infty\Big\} = \Big\{\theta \in \mathbb{R}^d : \int_{|x|>1} e^{\theta.x}\nu(dx) < \infty\Big\}, \tag{IV.12}$$

where the second equality holds by Theorem 2.4. Thanks to the convexity of the exponential function it is straightforward that the set Θ_1 is convex. In view to use importance sampling routine, based on exponential tilting, we define the family of $\{\mathbb{P}_\theta, \theta \in \Theta_1\}$, as all the equivalent probability measures with respect to $\mathbb{P}$ such that

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}}\Big|_{\mathcal{F}_t} = \frac{e^{\theta.L_t}}{\mathbb{E}[e^{\theta.L_t}]} = e^{\theta.L_t - t\kappa(\theta)}$$

whee κ denotes the cumulant generating function given by $\kappa(\theta) = \ln \mathbb{E}\left[e^{\theta.L_1}\right]$. Under $\mathbb{P}_\theta$, the stochastic process $\{L_t; t \geq 0\}$ is still a Lévy process with the exponential tilted triplet $(\gamma_\theta, 0, \nu_\theta)$ where $\gamma_\theta = \gamma + \int_{|x|\leq 1} x(\nu_\theta - \nu)(dx)$ and $\nu_\theta(dx) = e^{\theta.x}\nu(dx)$ (see e.g. Cont and Tankov [20]). Hence, we obtain $\mathbb{E}\left[F(L_T)\right] = \mathbb{E}_\theta\left[F(L_T)e^{-\theta.L_T + T\kappa(\theta)}\right]$. If we introduce the Lévy process $\{L_t^\theta; t \geq 0\}$ with generating triplet $(\gamma_\theta, 0, \nu_\theta)$ under $\mathbb{P}$, then the random variable L_T under $\mathbb{P}_\theta$ has the same law as L_T^θ under $\mathbb{P}$ and we get

$$\mathbb{E}\left[F(L_T)\right] = \mathbb{E}\left[F(L_T^\theta)e^{-\theta.L_T^\theta + T\kappa(\theta)}\right].$$

Further, one can use this importance sampling twice in the SR algorithm with considering θ_1 and θ_2 in $\mathbb{R}^d$ and approximate $\mathbb{E}[F(L_T)]$ by

$$\frac{1}{N_1}\sum_{k=1}^{N_1} F(L_{T,k}^{\varepsilon^\beta,\theta_1})e^{-\theta_1.L_{T,k}^{\varepsilon^\beta,\theta_1} + T\kappa(\theta_1)} + \frac{1}{N_2}\sum_{k=1}^{N_2}(F(L_{T,k}^{\varepsilon,\theta_2}) - F(L_{T,k}^{\varepsilon^\beta,\theta_2}))e^{-\theta_2.L_{T,k}^{\varepsilon,\theta_2} + T\kappa(\theta_2)}.$$

Miming the proof of Theorem 3.3 we establish a central limit theorem with limit variance $\mathrm{Var}(F(L_T^{\theta_1})e^{-\theta_1 L_T^{\theta_1} + T\kappa(\theta_1)}) + T\mathbb{E}((\nabla F(L_T^{\theta_2}).\Sigma\nabla F(L_T^{\theta_2}))e^{-2\theta_2 L_T^{\theta_2} + 2T\kappa(\theta_2)})$. Since $L_T^{\theta_1}$ (resp. $L_T^{\theta_2}$) under $\mathbb{P}$ has the same law as L_T under $\mathbb{P}_{\theta_1}$ (resp. $\mathbb{P}_{\theta_2}$) we rewrite this variance using once again the Esscher transform as

$$\mathbb{E}\left[F^2(L_T)e^{-\theta_1.L_T + T\kappa(\theta_1)}\right] - \left[\mathbb{E}F(L_T)\right]^2 + T\mathbb{E}\left[(\nabla F(L_T).\Sigma\nabla F(L_T))e^{-\theta_2 L_T + T\kappa(\theta_2)}\right].$$

Hence, let us introduce for $i \in \{1, 2\}$,

$$v_i(\theta) := \mathbb{E}\left[F_i(L_T)e^{-\theta L_T + T\kappa(\theta)}\right], \text{ with } F_1 \equiv F^2 \text{ and } F_2 \equiv \nabla F.\Sigma\nabla F. \tag{IV.13}$$

Our aim now is to minimize separately these two quantities. To do so, for $i \in \{1, 2\}$, we introduce a first set

$$\Theta_{i,2} := \Theta_1 \cap \left\{\theta \in \mathbb{R}^d : \mathbb{E}\left[F_i(L_T)e^{-\theta.L_T}\right] < +\infty\right\}$$

to ensure the existence of $v_i(\theta)$ and a second set

$$\Theta_{i,3} := \Theta_{i,2} \cap \left\{\theta \in \mathbb{R}^d : \mathbb{E}\left[|L_T|^2 F_i(L_T)e^{-\theta.L_T}\right] < +\infty\right\}$$

to make sens for the first and second derivatives of $v_i(\theta)$. For $i \in \{1, 2\}$, if we assume that $Leb(\Theta_{i,3}) > 0$, then the convexity of sets $\Theta_{i,2}$ and $\Theta_{i,3}$ can be proved in a similar manner to the proof of Lemma 2.2 in [36]. Moreover, we prove the convexity of v_i, $i \in \{1, 2\}$.

Proposition 4.1. *Let $i \in \{1, 2\}$. Assume $\mathbb{P}(F_i(L_T) \neq 0) > 0$. Then, $\theta \mapsto v_i(\theta)$ is a $\mathscr{C}^2$ strictly convex function on $\Theta_{i,3}$ and $\nabla v_i(\theta) = \mathbb{E}[H_i(\theta, L_T)]$ where*

$$H_i(\theta, L_T) = (T\nabla\kappa(\theta) - L_T)F_i(L_T)\exp(-\theta.L_T + T\kappa(\theta)). \tag{IV.14}$$

Proof. For a fixed $i \in \{1, 2\}$, the function $\theta \mapsto F_i(L_T)e^{-\theta L_T + T\kappa(\theta)}$ is almost surely differentiable on Θ_1 with a first derivative equal to $H_i(\theta, L_T)$. Further, according to the properties of the moment generating function, the function $\theta \mapsto v_i(\theta)$ is finite for $\theta \in \Theta_{i,2}$ and is differentiable with $\nabla v_i(\theta) = \mathbb{E}[H_i(\theta, L_T)]$ provided that $\mathbb{E}[|H_i(\theta, L_T)|]$ is finite. Using Hölder's inequality, this last condition is satisfied as soon as $\theta \in \Theta_{i,3}$. In the same way, we prove that v_i is of class $\mathscr{C}^2$ on $\Theta_{i,3}$ and we get for all $u \in \mathbb{R}^d \setminus \{0\}$,

$$u.\text{Hess}(v_i(\theta))u = \mathbb{E}\left[\left(u.\text{Hess}(\kappa(\theta))u + (u.(T\nabla\kappa(\theta) - L_T))^2\right) F_i(L_T)e^{-\theta.L_T + T\kappa(\theta)}\right].$$

Note that $\text{Hess}(\kappa(\theta))$ is nothing but the variance-covariance matrix of the random vector L_T under the probability measure $\mathbb{P}_\theta$ and it is clearly definite positive. Finally, since $\mathbb{P}(F_i(L_T) \neq 0) > 0$, we conclude that v_i is strictly convex on $\Theta_{i,3}$. □

For $\varepsilon > 0$, the same result holds for the approximated Lévy process $(L_t^\varepsilon)_{t\geq 0}$ by considering the associated sets Θ_1^ε, $\Theta_{i,2}^\varepsilon$ and $\Theta_{i,3}^\varepsilon$ and functions κ_ε and $v_{i,\varepsilon}$, $i \in \{1, 2\}$, with the canonical filtration $(\mathcal{F}_t^\varepsilon)_{0\leq t\leq T}$ defined by $\mathcal{F}_t^\varepsilon = \sigma(L_s^\varepsilon, s \leq t)$.

Proposition 4.2. *Let $i \in \{1,2\}$. Assume $\mathbb{P}(F_i(L_T^\varepsilon) \neq 0) > 0$ then the function $v_{i,\varepsilon}(\theta) = \mathbb{E}\left[F_i(L_T^\varepsilon)e^{-\theta L_T^\varepsilon + T\kappa_\varepsilon(\theta)}\right]$ is of class $\mathscr{C}^2$ and strictly convex on $\Theta_{i,3}^\varepsilon$ with $\nabla v_{i,\varepsilon}(\theta) = \mathbb{E}\left[H_i(\theta, L_T^\varepsilon)\right]$.*

Now, let us introduce for $i \in \{1,2\}$

$$\theta_{i,\varepsilon}^* := \underset{\theta\in\Theta_{i,3}^\varepsilon}{\arg\min}\, v_{i,\varepsilon}(\theta) \quad \text{and} \quad \theta_i^* := \underset{\theta\in\Theta_{i,3}}{\arg\min}\, v_i(\theta). \tag{IV.15}$$

Our aim now is to study for $i \in \{1,2\}$ the convergence of $\theta_{i,\varepsilon}^*$ toward θ_i^* as ε tends to zero. For $q > 1$, we define the set

$$\Theta_q := \left\{\theta \in \mathbb{R}^d : \int_{|x|>1} |x|^{2q} e^{-q\theta.x}\nu(dx) < +\infty\right\}. \tag{IV.16}$$

Remark 4.3. *1. It is worth to note that for $0 \leq q' \leq 2q$ and $\theta \in \Theta_q$ we have $\int_{|x|>1} |x|^{q'} e^{-q\theta.x}\nu(dx) < +\infty$. We also have $\Theta_{q_2} \subset \Theta_{q_1}$ for all $q_1 \leq q_2$.*

2. Further, for $i \in \{1,2\}$, if $\mathbb{E}\left[F_i^a(L_T)\right]$, $a > 1$, is finite then by Hölder's inequality we easily get $\Theta_q \subset \Theta_{i,3}$ for all $q \geq a/a-1$. The same result holds for the approximated Lévy process. Indeed, for $\varepsilon > 0$, we have $\Theta_q \subset \Theta_{i,3}^\varepsilon$ provided that $\mathbb{E}\left[F_i^a(L_T^\varepsilon)\right] < \infty$.

According the above remark, choosing $\theta \in \Theta_q$ with $q \geq a/a-1$ ensures that θ will belong to the domain of convexity of both v_i and $v_{i,\varepsilon}$. On the other hand it also guarantees the finiteness of the quantity $\int_{|x|>1} |x|^q e^{-q\theta.x}\nu(dx)$ which will be needed in each proof assuming condition $\theta \in \Theta_q$.

In what follows, let $\mathring{E}$ denote the set of all interior points of a given set E. We have the following result.

Theorem 4.4. *Let $i \in \{1,2\}$. Suppose that $x \mapsto F_i(x)$ is continuous, that is for the case $i = 1$ the function F is continuous and for $i = 2$ the function F is of class $\mathscr{C}^1$. Moreover, assume $\mathbb{P}(F_i(L_T) \neq 0) > 0$, $\mathbb{P}(F_i(L_T^\varepsilon) \neq 0) > 0$ for all $\varepsilon > 0$ and there exists $a > 1$ such that $\mathbb{E}\left[F_i^a(L_T)\right]$ and $\sup_{\varepsilon>0} \mathbb{E}\left[F_i^a(L_T^\varepsilon)\right]$ are finite. Let K be a compact set such that $K \subset \mathring{\Theta}_q$ with $q > \frac{a}{a-1}$ and assume that the sequence $(\theta_{i,\varepsilon}^*)_{\varepsilon>0} \in K$. Then,*

$$\theta_{i,\varepsilon}^* \longrightarrow \theta_i^* \in K, \qquad \text{as } \varepsilon \to 0.$$

We prove Theorem 4.4 after the following technical lemma.

Lemma 4.5. *Let K be a compact subset of Θ_q with $q > 1$, we have $\sup_{\theta\in\Theta_q} \mathbb{E}\left[|L_T^\varepsilon|^q e^{-q\theta.L_T^\varepsilon}\right]$ is uniformly bounded in ε.*

Proof. Let us consider the two independent Lévy processes L^1 and $\tilde{L}^\varepsilon := L^\varepsilon - L^1$ and the submultiplicative function $g_\theta(x) := (|x| \vee 1)^q e^{-q\theta.x}$. There exists $c_q > 0$ depending only on q such that $g_\theta(x+y) \leq c_q g_\theta(x) g_\theta(y)$ for any $\theta \in \mathbb{R}^d$ and

$$\mathbb{E}\Big[|L_T^\varepsilon|^q e^{-q\theta.L_T^\varepsilon}\Big] \leq c_q \mathbb{E}\Big[g_\theta(\tilde{L}_T^\varepsilon)\Big]\mathbb{E}\Big[g_\theta(L_T^1)\Big].$$

Since the function $\theta \mapsto \mathbb{E}\Big[g_\theta(L_T^1)\Big]$ is continuous on Θ_q the second expectation on the right hand side is uniformly bounded on $\theta \in K$. Concerning the first expectation, we start by establishing the uniform convergence of $\tilde{\kappa}_\varepsilon$ toward $\tilde{\kappa}$, where $\tilde{\kappa}_\varepsilon$ and $\tilde{\kappa}$ denote the cumulant generating functions of respectively $\tilde{L}^\varepsilon = L^\varepsilon - L^1$ and $\tilde{L} = L - L^1$. According to the Lévy Kintchine decomposition, we have $\tilde{\kappa}(\theta) - \tilde{\kappa}_\varepsilon(\theta) = \int_{|x|<\varepsilon}(e^{\theta.x} - 1 - \theta.x)\nu(dx)$ and thanks to Taylor's expansion we get

$$|\tilde{\kappa}(\theta) - \tilde{\kappa}_\varepsilon(\theta)| \leq \frac{|\theta|^2}{2} e^{|\theta|}\sigma^2(\varepsilon). \tag{IV.17}$$

This ensures the uniform convergence of the family functions $(\tilde{\kappa}_\varepsilon)_{0<\varepsilon<1}$ on any compact set of $\mathbb{R}^d$. Note that for all $x = (x_1, \cdots, x_d) \in \mathbb{R}^d$ we have $(|x| \vee 1)^q \leq ce^{|x|} \leq c\prod_{j=1}^d (e^{x_j} + e^{-x_j})$ with some $c > 0$ depending only on q. This last upper bound can be written as a sum of finite number of exponential functions evaluated at points which are a linear combination of the components of the vector x. Therefore there exists a family of deterministic $\mathbb{R}^d$-valued vectors, $(b_j)_{1\leq j\leq 2^d}$ such that

$$\mathbb{E}\Big[g_\theta(\tilde{L}_T^\varepsilon)\Big] \leq c\sum_{j=1}^{2^d} \mathbb{E}\Big[e^{(b_j - q\theta).L_T^\varepsilon}\Big].$$

Each term in the above sum is nothing else $\exp(\tilde{\kappa}_\varepsilon(b_j - q\theta))$ which in turn converges to $\exp(\tilde{\kappa}(b_j - q\theta))$ as ε tends to zero. This gives us the desired claim. □

Proof.[Proof of Theorem 4.4] Let $i \in \{1, 2\}$ and $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence decreasing to zero. Note that $(\theta^*_{i,\varepsilon_n})_{n\in\mathbb{N}}$ is a $\mathbb{R}^d$-bounded sequence. So, we only need to prove that for any subsequence $(\theta^*_{i,\varepsilon_{n_k}})_{k\in\mathbb{N}}$, if $\theta^*_{i,\varepsilon_{n_k}} \to \theta^*_{i,\infty} \in \mathbb{R}^d$ then $\theta^*_{i,\infty} = \theta^*_i$. According to Proposition 4.2 above we have

$$\nabla v_{i,\varepsilon_{n_k}}(\theta^*_{i,\varepsilon_{n_k}}) = \mathbb{E}\left[(\theta^*_{i,\varepsilon_{n_k}}T - L_T^{\varepsilon_{n_k}})F_i(L_T^{\varepsilon_{n_k}})e^{-\theta^*_{i,\varepsilon_{n_k}}.L_T^{\varepsilon_{n_k}} + T\kappa_{\varepsilon_{n_k}}(\theta^*_{i,\varepsilon_{n_k}})}\right] = 0.$$

Now, let $\tilde{a} = \frac{aq}{a+q}$, it is easy to check that $1 < \tilde{a} < a$, so by applying Hölder's inequality we get

$$\mathbb{E}\left[\left|(\theta^*_{i,\varepsilon_{n_k}}T - L_T^{\varepsilon_{n_k}})F_i(L_T^{\varepsilon_{n_k}})e^{-\theta^*_{i,\varepsilon_{n_k}}.L_T^{\varepsilon_{n_k}} + T\kappa_{\varepsilon_{n_k}}(\theta^*_{i,\varepsilon_{n_k}})}\right|^{\tilde{a}}\right] \leq$$
$$\mathbb{E}^{(a-\tilde{a})/a}\left[\left|(\theta^*_{i,\varepsilon_{n_k}}T - L_T^{\varepsilon_{n_k}})e^{-\theta^*_{i,\varepsilon_{n_k}}.L_T^{\varepsilon_{n_k}} + T\kappa_{\varepsilon_{n_k}}(\theta^*_{i,\varepsilon_{n_k}})}\right|^{\tilde{a}a/(a-\tilde{a})}\right]\mathbb{E}^{\tilde{a}/a}\left[F_i^a(L_T^{\varepsilon_{n_k}})\right].$$

Note that $\sup_{\varepsilon>0} \mathbb{E}\left[F_i^a(L_T^\varepsilon)\right] < \infty$. Hence, to get the uniform integrability it is sufficient to prove that the first expectation on the right hand side of the above inequality is uniformly bounded on ε_{n_k} and $\theta^*_{i,\varepsilon_{n_k}}$. Indeed, using the almost sure convergence of L_T^ε toward L_T and the continuity of function F_i, we easily get

$$\nabla v_i(\theta^*_{i,\infty}) = \mathbb{E}\left[(\theta^*_{i,\infty}T - L_T)F_i(L_T)e^{-\theta^*_{i,\infty}.L_T+T\kappa(\theta^*_{i,\infty})}\right] = 0$$

and then we complete the proof using the uniqueness of the minimum ensured by Proposition 4.1. Consequently, noticing that $q = \tilde{a}a/(a-\tilde{a})$, it remains now to prove the uniform boundedness of the quantity $\mathbb{E}\left[\left|(\theta^*_{i,\varepsilon_{n_k}}T - L_T^{\varepsilon_{n_k}})e^{-\theta^*_{i,\varepsilon_{n_k}}.L_T^{\varepsilon_{n_k}}+T\kappa_{\varepsilon_{n_k}}(\theta^*_{i,\varepsilon_{n_k}})}\right|^q\right]$. To do so, we establish first the uniform convergence of κ_ε toward κ. According to the decomposition given by relation (IV.2), we have that $\kappa(\theta)-\kappa_\varepsilon(\theta) = \int_{|x|<\varepsilon}(e^{\theta.x}-1-\theta.x)\nu(dx)$. By Taylor's expansion we deduce

$$|\kappa(\theta)-\kappa_\varepsilon(\theta)| \leq \frac{|\theta|^2}{2}e^{|\theta|}\sigma^2(\varepsilon). \tag{IV.18}$$

Hence, the family functions $(\kappa_\varepsilon)_{0<\varepsilon<1}$ is equicontinuous on any compact subset of Θ_1 and we deduce the convergence of $\kappa_{\varepsilon_{n_k}}(\theta^*_{i,\varepsilon_{n_k}})$ toward $\kappa(\theta^*_{i,\infty})$ when k tends to infinity. Noticing that $-qK \subset \Theta_1$, we use once again the equicontinuity of $(\kappa_\varepsilon)_{0<\varepsilon<1}$ on the compact set $-qK$ to get $\lim_{k\to\infty}\kappa_{\varepsilon_{n_k}}(-q\theta^*_{i,\varepsilon_{n_k}}) = \kappa(-q\theta^*_{i,\infty})$ and then the problem is reduced to prove the uniform boundedness of $\mathbb{E}\left[|L_T^{\varepsilon_{n_k}}|^q e^{-q\theta^*_{i,\varepsilon_{n_k}}.L_T^{\varepsilon_{n_k}}}\right]$ which is ensured by Lemma 4.5. □

5 The adaptive procedure

5.1 Stochastic algorithms

The aim now is to construct family sequences converging almost surely to the optimal limits $\theta^*_{1,\varepsilon}$ and $\theta^*_{2,\varepsilon}$ of the previous section. For this, let $(L_{T,n})_{n\geq1}$ (resp. $(L^\varepsilon_{T,n})_{n\geq1}$, $\varepsilon>0$), be i.i.d copies of the $\mathbb{R}^d$-valued random variable L_T (resp. L_T^ε). Let K be a compact convex subset of $\Theta_1 \subset \mathbb{R}^d$ with $\{0\} \in K$. For fixed $i \in \{1,2\}$ and $\theta_{i,0} \in K$, we construct recursively the sequences of $\mathbb{R}^d$-valued random variables $(\theta_{i,n})_{n\in\mathbb{N}}$ and $(\theta_{i,\varepsilon,n})_{n\in\mathbb{N}}$ defined by the system

$$\begin{cases} \theta_{i,n+1} &= \Pi_K\left[\theta_{i,n} - \gamma_{n+1}H_i(\theta_{i,n}, L_{T,n+1})\right] \\ \theta_{i,\varepsilon,n+1} &= \Pi_K\left[\theta_{i,\varepsilon,n} - \gamma_{n+1}H_i(\theta_{i,\varepsilon,n}, L^\varepsilon_{T,n+1})\right] \end{cases} \tag{IV.19}$$

where Π_K is the Euclidean projection onto the constraint set K, H_1 and H_2 are given by relation (IV.14) and the gain sequence $(\gamma_n)_{n\geq1}$ is a decreasing sequence of positive real numbers satisfying

$$\sum_{n=1}^{\infty} \gamma_n = \infty \text{ and } \sum_{n=1}^{\infty} \gamma_n^2 < \infty \tag{IV.20}$$

Theorem 5.1. *Let $i \in \{1,2\}$. Assume $\mathbb{P}(F_i(L_T) \neq 0) > 0$, $\mathbb{P}(F_i(L_T^{\varepsilon}) \neq 0) > 0$ for all $\varepsilon > 0$ and there exists $a > 1$ such that $\mathbb{E}\left[F_i^{2a}(L_T)\right]$ and $\sup_{\varepsilon>0} \mathbb{E}\left[F_i^{2a}(L_T^{\varepsilon})\right]$ are finite. Let K be a compact set such that $K \subset \mathring{\Theta}_{2a/(a-1)}$ then the following assertions hold.*

– *If the unique $\theta_i^* = \underset{\theta \in \Theta_{i,3}}{\arg\min}\, v_i(\theta)$ satisfies $\theta_i^* \in K$ then the sequence $\theta_{i,n} \underset{n\to+\infty}{\longrightarrow} \theta_i^*$ a.s.*

– *If the unique $\theta_{i,\varepsilon}^* = \underset{\theta \in \Theta_{i,3}^{\varepsilon}}{\arg\min}\, v_{i,\varepsilon}(\theta)$ satisfies $\theta_{i,\varepsilon}^* \in K$ then the sequence $\theta_{i,\varepsilon,n} \underset{n\to+\infty}{\longrightarrow} \theta_{i,\varepsilon}^*$ a.s.*

Proof. Both items can be proved in the same way, so we choose to give the proof only for the first one. According to Theorem A.1. in Laruelle, Lehalle and Pagès [45] on truncated Robbins Monro algorithm (see also Kushner and Yin [42] for more details) : in order to prove that $\theta_{i,n}^{\varepsilon} \underset{n\to+\infty}{\longrightarrow} \theta_{i,\varepsilon}^*$ *a.s.*, we need to check firstly the mean-reverting property, namely

$$\forall \theta \neq \theta_i^* \in K, \quad \langle \nabla v_i(\theta), \theta - \theta_i^* \rangle > 0.$$

This is satisfied using $\nabla v_i(\theta_i^*) = 0$ and the convexity of v_i ensured by Proposition 4.1. Secondly, we have to check the non explosion assumption given by

$$\exists C > 0 \text{ such that } \forall \theta \in K, \quad \mathbb{E}\left[|H_i(\theta, L_T)|^2\right] < C(1 + |\theta|^2).$$

In fact, using Hölder's inequality with the couple a and $a/(a-1)$, we obtain

$$\mathbb{E}|H_i(\theta, L_T)|^2 \leq \mathbb{E}^{\frac{1}{a}}\left[F_i^{2a}(L_T)\right] \mathbb{E}^{\frac{a-1}{a}}\left[|T\nabla\kappa(\theta) - L_T|^{2a/(a-1)} e^{-2a/(a-1)\theta.L_T}\right] e^{2T\kappa(\theta)}$$

Since $\mathbb{E}\left[F_i^{2a}(L_T)\right]$ is finite and $\theta \in K \subset \Theta_{2a/(a-1)}$, we deduce that $\sup_{\theta\in K} \mathbb{E}|H_i(\theta, L_T)|^2 < \infty$ which completes the proof. □

Theorem 5.2. *Considering the sequences given by relation (IV.19), for $i \in \{1,2\}$, we have for all $n \in \mathbb{N}$*

$$\theta_{i,\varepsilon,n} \underset{\varepsilon\to 0}{\longrightarrow} \theta_{i,n} \quad a.s.$$

Proof. We proceed by induction. The base case is trivial and for the inductive step we suppose that for $i \in \{1,2\}$, $n \in \mathbb{N}$, $\theta_{i,\varepsilon,n}$ converges to $\theta_{i,n}$ *a.s.* as ε goes to 0 and we prove the statement for $n+1$. We have $\theta_{i,\varepsilon,n+1} = \Pi_K\left[\theta_{i,\varepsilon,n} - \gamma_{i+1} H_i(\theta_{i,\varepsilon,n}, L_{T,n+1}^{\varepsilon})\right]$. By the continuity of the function H_i given by (IV.14), the almost sure convergence of $L_{T,n+1}^{\varepsilon}$ to $L_{T,n+1}$ and the continuity of the projection function Π_K, we deduce that $\theta_{i,\varepsilon,n+1}$ converges to $\theta_{i,n+1}$ *a.s.* as ε goes to 0. □

The following corollary follows immediately thanks to theorems 4.4, 5.1 and 5.2.

Corollary 5.3. *Under assumptions of Theorem 5.1, the constrained algorithm given by routine* (IV.19) *satisfies for* $i \in \{1,2\}$

$$\lim_{\substack{\varepsilon\to 0\\ n\to\infty}} \theta_{i,\varepsilon,n} = \lim_{\varepsilon\to 0}(\lim_{n\to\infty}\theta_{i,\varepsilon,n}) = \lim_{n\to\infty}(\lim_{\varepsilon\to 0}\theta_{i,\varepsilon,n}) = \theta_i^*, \quad \mathbb{P}\text{-}a.s. \tag{IV.21}$$

Remark 5.4. *Suppose for a while that we omit assumptions* $\theta_i^* \in K$ *and* $\theta_{i,\varepsilon}^* \in K$ *in Theorem 5.1 above. According to Theorem 3.2. of Kawai [36] based on Theorem 2.1 of Kushner and Yin [42] there exist* $\bar\theta_i$ *and* $\bar\theta_{i,\varepsilon}$ *in* K *such that* $\theta_{i,n} \underset{n\to+\infty}{\longrightarrow} \bar\theta_i$ *a.s. and* $\theta_{i,\varepsilon,n} \underset{n\to+\infty}{\longrightarrow} \bar\theta_{i,\varepsilon}$ *a.s. Moreover,* $v_i(\bar\theta_i) \le v_i(\theta)$ *and* $v_{i,\varepsilon}(\bar\theta_{i,\varepsilon}) \le v_{i,\varepsilon}(\theta)$ *for all* $\theta \in K$*. In this case we can prove that the constrained algorithm given by routine* (IV.19) *satisfies relation* (IV.21) *with* $\bar\theta_i$ *instead of* θ_i^*.

5.2 Central limit theorems

In what follows, we consider the filtration $\mathcal{F}_{T,k} = \sigma(L_{t,\ell}, L^\varepsilon_{t,\ell}, 0 < \varepsilon < 1, t \le T, \ell \le k)$, where $(L_\ell, L^\varepsilon_\ell)_{\ell\ge 1}$ are independent copies of (L, L^ε). Let us assume that there exists a family of sequences $(\theta^\varepsilon_k)_{k\ge 0, 0<\varepsilon\le 1}$ and $(\theta_k)_{k\ge 0}$ satisfying

$$(\mathcal{H}_\theta) \quad \begin{cases} \text{For each } \varepsilon > 0,\ (\theta^\varepsilon_k)_{k\ge 0} \text{ and } (\theta_k)_{k\ge 0} \text{ are } (\mathcal{F}_{T,k})_{k\ge 0}\text{-adapted} \\[2ex] \lim\limits_{k\to\infty}(\lim\limits_{\varepsilon\to 0}\theta^\varepsilon_k) = \lim\limits_{k\to\infty}\theta_k = \lim\limits_{\varepsilon\to 0}(\lim\limits_{k\to\infty}\theta^\varepsilon_k) = \lim\limits_{\varepsilon\to 0}\theta^*_\varepsilon = \theta^*, \quad \mathbb{P}\text{-}a.s., \end{cases}$$

with deterministic limits θ^* and θ^*_ε.

At first, we start with studying the MC setting. We use the adaptive importance sampling algorithm for the MC method to approximate our initial quantity of interest $\mathbb{E}F(L_T)$ by

$$Q_\varepsilon^{\text{ISMC}} = \frac{1}{N}\sum_{k=1}^N F(L_{T,k}^{\varepsilon,\theta^\varepsilon_{k-1}}) e^{-\theta^\varepsilon_{k-1}\cdot L_{T,k}^{\varepsilon,\theta^\varepsilon_{k-1}} + T\kappa_\varepsilon(\theta^\varepsilon_{k-1})}. \tag{IV.22}$$

Our task now is to establish a central limit theorem for the adaptive importance sampling Monte Carlo method (ISMC).

Theorem 5.5. *Let* $F : \mathbb{R}^d \to \mathbb{R}$ *be a continuous function satisfying assumption* ($\mathrm{WE}_{\upsilon_\varepsilon}$) *and such that* $\sup_{0<\varepsilon\le 1}\mathbb{E}[F^{2a}(L^\varepsilon_T)] < +\infty$ *for* $a > 1$*. Moreover, assume that* $Leb(\Theta_q) > 0$ *with* $q > a/(a-1)$ *and there exists a double indexed family* $(\theta^\varepsilon_k)_{k\in\mathbb{N},\varepsilon>0}$ *satisfying* ($\mathcal{H}_\theta$) *and belonging* *to some compact subset $K \subset \mathring{\Theta}_q$. Then, if we choose $N = \upsilon_\varepsilon^{-2}$, the following convergence holds*

$$\upsilon_\varepsilon^{-1}\left(Q_\varepsilon^{\text{ISMC}} - \mathbb{E}F(L_T)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(C_F, \sigma^2), \qquad \text{as } \varepsilon \searrow 0, \tag{IV.23}$$

where $\sigma^2 := \mathbb{E}\left[F^2(L_T)e^{-\theta^*.L_T^\varepsilon + T\kappa(\theta^*)}\right] - (\mathbb{E}[F(L_T)])^2$.

Proof. By assumption ($\text{WE}_{\upsilon_\varepsilon}$) we only need to study the asymptotic behavior of the martingale arrays $(M_k^\varepsilon)_{k\geq 1}$ given by $M_k^\varepsilon := \upsilon_\varepsilon \sum_{i=1}^k \left(F(L_{T,i}^{\varepsilon,\theta_{i-1}^\varepsilon})e^{-\theta_{i-1}^\varepsilon.L_{T,i}^{\varepsilon,\theta_{i-1}^\varepsilon} + T\kappa_\varepsilon(\theta_{i-1}^\varepsilon)} - \mathbb{E}F(L_T^\varepsilon)\right)$. To do so, we plan to apply the Lindeberg-Feller central limit theorem for martingales arrays (see Theorem 4.1 in Chapter II). The proof is divided into two steps.

Step 1. The quadratic variation of the martingale arrays $(M_k^\varepsilon)_{k\geq 1}$ is given by

$$\langle M^\varepsilon\rangle_N = \frac{1}{N}\sum_{k=1}^N \mathbb{E}\Big[F^2(L_{T,k}^{\varepsilon,\theta_{k-1}^\varepsilon})e^{-2\theta_{k-1}^\varepsilon.L_{T,k}^{\varepsilon,\theta_{k-1}^\varepsilon} + 2T\kappa_\varepsilon(\theta_{k-1}^\varepsilon)}|\mathcal{F}_{T,k-1}\Big] - (\mathbb{E}F(L_T^\varepsilon))^2. \tag{IV.24}$$

Since θ_{k-1}^ε is $\mathcal{F}_{T,k-1}$-measurable and $(L_{T,k}^{\varepsilon,\theta})_{\theta\in\Theta_q} \perp\!\!\!\perp \mathcal{F}_{T,k-1}$, by Esscher transform we obtain

$$\langle M^\varepsilon\rangle_N = \frac{1}{N}\sum_{k=1}^N \gamma_\varepsilon(\theta_{k-1}^\varepsilon)e^{T\kappa_\varepsilon(\theta_{k-1}^\varepsilon)} - (\mathbb{E}F(L_T^\varepsilon))^2,$$

where for all $\theta \in \Theta_q$, $\gamma_\varepsilon(\theta) = \mathbb{E}\left[F^2(L_T^\varepsilon)e^{-\theta.L_T^\varepsilon}\right]$. On the one hand, using assumption ($\text{WE}_{\upsilon_\varepsilon}$), we have $\lim_{\varepsilon\to 0}\mathbb{E}F(L_T^\varepsilon) = \mathbb{E}F(L_T)$. On the other hand, thanks to relation (IV.18) we have the uniform equicontinuity of the family $(\kappa_\varepsilon)_{\varepsilon>0}$ on the compact subset K. So, we only need to check this last property for the family $(\gamma_\varepsilon)_{\varepsilon>0}$ in view to use after that Lemma 4.3 in Chapter II and then deduce the convergence of $\langle M^\varepsilon\rangle_N$ toward $\gamma(\theta^*) - (\mathbb{E}F(L_T))^2$ as $\varepsilon \searrow 0$, where $\gamma(\theta) := \mathbb{E}\left[F^2(L_T)e^{-\theta.L_T}\right]$.

Thus, it remains to prove the uniform equicontinuity of the family functions $(\gamma_\varepsilon)_{\varepsilon>0}$ defined on the compact set K. Using Hölder's inequality and the assumption $\sup_{\varepsilon>0}\mathbb{E}\left[F^{2a}(L_T^\varepsilon)\right] < +\infty$, there exists $c_1 > 0$ not depending on ε such that

$$\begin{aligned} |\gamma_\varepsilon(\theta) - \gamma_\varepsilon(\theta')| &\leq \mathbb{E}\left[F^2(L_T^\varepsilon)\left|e^{-\theta.L_T^\varepsilon} - e^{-\theta'.L_T^\varepsilon}\right|\right] \\ &\leq c_1\mathbb{E}^{1/q}\left[\left|e^{-\theta.L_T^\varepsilon} - e^{-\theta'.L_T^\varepsilon}\right|^q\right]. \end{aligned}$$

By Taylor's expansion and standard calculations we easily get

$$|e^{-\theta.L_T^\varepsilon} - e^{-\theta'.L_T^\varepsilon}|^q \leq |\theta - \theta'|^q \int_0^1 |L_T^\varepsilon|^q e^{-q(u\theta + (1-u)\theta').L_T^\varepsilon}du.$$

Therefore, we have

$$|\gamma_\varepsilon(\theta)-\gamma_\varepsilon(\theta')|\le c_1|\theta-\theta'|\sup_{\theta\in\Theta_q}\mathbb{E}^{1/q}\Big[|L_T^\varepsilon|^q e^{-q\theta.L_T^\varepsilon}\Big].$$

Hence, according to Lemma 4.5 there exists a constant $c_2>0$ also not depending on and ε such that

$$|\gamma_\varepsilon(\theta)-\gamma_\varepsilon(\theta')|\le c_2|\theta-\theta'|. \tag{IV.25}$$

This completes the proof of the first step.

Step 2. We check now the Lyapunov condition given by assumption *B3* in Theorem 4.1 (see Chapter II). So, let $\tilde a=\frac{aq+a}{2a+q}$, it is easy to check that $1<\tilde a<a$. Once again using the mesurability properties of the family $(L_{T,k}^{\varepsilon,\theta})_{\theta\in\Theta_q}$ and the sequence $(\theta_k^\varepsilon)_{k\ge0}$, we get using the Esscher transform

$$\begin{aligned}\sum_{k=1}^N\mathbb{E}\left[\left|M_k^\varepsilon-M_{k-1}^\varepsilon\right|^{2\tilde a}|\mathcal{F}_{T,k-1}\right] &= \frac{1}{N^{\tilde a}}\sum_{k=1}^N\mathbb{E}\Big[\Big|F(L_{T,k}^{\varepsilon,\theta_{k-1}^\varepsilon})e^{-\theta_{k-1}^\varepsilon.L_{T,k}^{\varepsilon,\theta_{k-1}^\varepsilon}+T\kappa_\varepsilon(\theta_{k-1}^\varepsilon)}-\mathbb{E}F(L_T^\varepsilon)\Big|^{2\tilde a}|\mathcal{F}_{T,k-1}\Big]\\ &\le \frac{2^{2\tilde a-1}}{N^{\tilde a}}\sum_{k=1}^N\gamma_{\tilde a,\varepsilon}(\theta_{k-1}^\varepsilon)e^{(2\tilde a-1)T\kappa_\varepsilon(\theta_{k-1}^\varepsilon)}+\frac{2^{2\tilde a-1}}{N^{\tilde a}}\Big|\mathbb{E}F(L_T^\varepsilon)\Big|^{2\tilde a}\end{aligned}$$

where for all $\theta\in\Theta_q$, $\gamma_{\tilde a,\varepsilon}(\theta)=\mathbb{E}\left[F^{2\tilde a}(L_T^\varepsilon)e^{-(2\tilde a-1)\theta.L_T^\varepsilon}\right]$. Then, by Hölder's inequality we get

$$\gamma_{\tilde a,\varepsilon}(\theta)\le\mathbb{E}^{\tilde a/a}\Big[F^{2a}(L_T^\varepsilon)\Big]\mathbb{E}^{(a-\tilde a)/a}\Big[e^{-(2\tilde a-1)a/(a-\tilde a)\theta.L_T^\varepsilon}\Big].$$

Noticing that $q=(2\tilde a-1)a/(a-\tilde a)$, it results from assumption $\sup_{0<\varepsilon\le1}\mathbb{E}\left[F^{2a}(L_T^\varepsilon)\right]<+\infty$ that $\gamma_{\tilde a,\varepsilon}$ is uniformly bounded on the compact subset $K\subset\Theta_q$. Moreover, using once again relation (IV.18) we deduce the uniform boundedness of the family $(\kappa_\varepsilon)_{\varepsilon>0}$ on the compact subset K. Hence, combining all these results together with assumption ($\mathrm{WE}_{\upsilon_\varepsilon}$), we deduce the existence of $c_3>0$ not depending on ε such that $\sum_{k=1}^N\mathbb{E}\left[\left|M_k^\varepsilon-M_{k-1}^\varepsilon\right|^{2\tilde a}|\mathcal{F}_{T,k-1}\right]\le\frac{c_3}{N^{\tilde a-1}}$. This completes the proof. □

Remark 5.6. *If one have in mind to reduce the variance by using an adaptive crude Monte Carlo method, it appears clear that the natural choice is*

$$\theta_1^*=\arg\min_{\theta\in\Theta_{1,3}}v_1(\theta)\quad\textit{and}\quad\theta_{1,\varepsilon}^*=\arg\min_{\theta\in\Theta_{1,3}^\varepsilon}v_{1,\varepsilon}(\theta)\ \textit{for}\ \varepsilon>0,$$

where v_1 and $v_{1,\varepsilon}$ are presented in section 4. The construction of stochastic sequences converging *almost surely to these desired targets and satisfying* $(\mathcal{H}_\theta)$ *is ensured by Corollary 5.3.*

Now, we use the adaptive importance sampling statistical Romberg method (ISSR) to approximate our initial quantity of interest $\mathbb{E}F(L_T)$ by

$$Q_\varepsilon^{\text{ISSR}} := \frac{1}{N_1}\sum_{k=1}^{N_1} F(L_{T,k}^{\varepsilon^\beta,\theta_{1,k-1}^{\varepsilon^\beta}})e^{-\theta_{1,k-1}^{\varepsilon^\beta}.L_{T,k}^{\varepsilon^\beta,\theta_{1,k-1}^{\varepsilon^\beta}}+T\kappa_{\varepsilon^\beta}(\theta_{1,k-1}^{\varepsilon^\beta})} + \frac{1}{N_2}\sum_{k=1}^{N_2}\left(F(L_{T,k}^{\varepsilon,\theta_{2,k-1}^{\varepsilon}}) - F(L_{T,k}^{\varepsilon^\beta,\theta_{2,k-1}^{\varepsilon}})\right)e^{-\theta_{2,k-1}^{\varepsilon}.L_{T,k}^{\varepsilon,\theta_{2,k-1}^{\varepsilon}}+T\kappa_{\varepsilon}(\theta_{2,k-1}^{\varepsilon})} \quad \text{(IV.26)}$$

Our second result is a central limit theorem for the adaptive ISSR method

Theorem 5.7. *Let* $F:\mathbb{R}^d\to\mathbb{R}$ *be a* $\mathscr{C}^1$ *function satisfying assumption* $(\text{WE}_{\upsilon_\varepsilon})$ *and such that* $\sup_{0<\varepsilon\le 1}\mathbb{E}F^{2a}(L_T^\varepsilon)$ *and* $\sup_{0<\varepsilon\le 1}\mathbb{E}\left|\sigma^{-1}(\varepsilon)(F(L_T^\varepsilon)-F(L_T))\right|^{2a}$ *are finite, for* $a>1$. *Suppose also that the following assumptions are satisfied.*

H1. Condition (IV.3) in Theorem 2.1 holds and there exists a definite positive matrix Σ *such that* $\lim_{\varepsilon\to 0}\sigma^{-2}(\varepsilon)\Sigma_\varepsilon = \Sigma$.

H2. For $0<\beta<1$, *we have* $\lim_{\varepsilon\to 0}\sigma(\varepsilon)\sigma^{-1}(\varepsilon^\beta)=0$ *and* $\lim_{\varepsilon\to 0}\upsilon_\varepsilon\sigma^{-1}(\varepsilon^\beta)=0$.

Moreover, assume that $Leb(\Theta_q)>0$ *with* $q>a/(a-1)$ *and for* $i\in\{1,2\}$ *there exists a double indexed family* $(\theta_{i,k}^\varepsilon)_{k\in\mathbb{N},\varepsilon>0}$ *satisfying* $(\mathcal{H}_\theta)$ *and belonging to some compact subset* $K_i\subset\mathring{\Theta}_q$. *If we choose* $N_1=\upsilon_\varepsilon^{-2}$ *and* $N_2=\upsilon_\varepsilon^{-2}\sigma^2(\varepsilon^\beta)$, *then*

$$\upsilon_\varepsilon^{-1}\left(Q_\varepsilon^{\text{ISSR}}-\mathbb{E}F(L_T)\right)\xrightarrow{\mathcal{L}}\mathcal{N}\left(C_F,\sigma^2+\tilde{\sigma}^2\right),\quad \text{as } \varepsilon\to 0,$$

where

$$\sigma^2=\mathbb{E}\left[F^2(L_T)e^{-\theta^*.L_T+T\kappa(\theta^*)}\right]-\left[\mathbb{E}F(L_T)\right]^2 \text{ and } \tilde{\sigma}^2=T\mathbb{E}\left[(\nabla F(L_T).\Sigma\nabla F(L_T))e^{-\theta^*.L_T+T\kappa(\theta^*)}\right].$$

Proof. By assumption $(\text{WE}_{\upsilon_\varepsilon})$ we only need to study the asymptotic behavior of $\upsilon_\varepsilon^{-1}Q_{1,\varepsilon}^{\text{ISSR}}+\upsilon_\varepsilon^{-1}Q_{2,\varepsilon}^{\text{ISSR}}$ with

$$Q_{1,\varepsilon}^{\text{ISSR}} = \frac{1}{N_1}\sum_{k=1}^{N_1}\left(F(L_{T,k}^{\varepsilon^\beta,\theta_{1,k-1}^{\varepsilon^\beta}})e^{-\theta_{1,k-1}^{\varepsilon^\beta}.L_{T,k}^{\varepsilon^\beta,\theta_{1,k-1}^{\varepsilon^\beta}}+T\kappa_{\varepsilon^\beta}(\theta_{1,k-1}^{\varepsilon^\beta})}-\mathbb{E}F(L_T^{\varepsilon^\beta})\right)$$

and

$$Q_{2,\varepsilon}^{\text{ISSR}} = \frac{1}{N_2}\sum_{k=1}^{N_2}\left(\left[F(L_{T,k}^{\varepsilon,\theta_{2,k-1}^{\varepsilon}})-F(L_{T,k}^{\varepsilon^\beta,\theta_{2,k-1}^{\varepsilon}})\right]e^{-\theta_{2,k-1}^{\varepsilon}.L_{T,k}^{\varepsilon,\theta_{2,k-1}^{\varepsilon}}+T\kappa_{\varepsilon}(\theta_{2,k-1}^{\varepsilon})}-\mathbb{E}\Big[F(L_T^\varepsilon)-F(L_T^{\varepsilon^\beta})\Big]\right).$$

An application of Theorem 5.5 yields $\upsilon_\varepsilon^{-1}Q_{1,\varepsilon}^{\text{ISSR}}\xrightarrow{\mathcal{L}}\mathcal{N}(0,\sigma^2)$, as $\varepsilon\to 0$. For the second term, we aim to apply Theorem 4.1 in Chapter II. So, we introduce the martingale arrays $(M_k^\varepsilon)_{k\geq 1}$

$$M_k^\varepsilon := \frac{\upsilon_\varepsilon^{-1}}{N_2}\sum_{\ell=1}^{k}\Big(\Big(F(L_{T,\ell}^{\varepsilon,\theta_{2,\ell-1}^\varepsilon}) - F(L_{T,\ell}^{\varepsilon^\beta,\theta_{2,\ell-1}^\varepsilon})\Big)e^{-\theta_{2,\ell-1}^\varepsilon . L_{T,\ell}^{\varepsilon,\theta_{2,\ell-1}^\varepsilon}+T\kappa_\varepsilon(\theta_{2,\ell-1}^\varepsilon)} - \mathbb{E}\Big[F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta})\Big]\Big).$$

Step 1. Thanks to assumption $(\mathcal{H}_\theta)$ and the Esscher transform, the quadratic variation of M evaluated at N_2 is equal to

$$\langle M^\varepsilon\rangle_{N_2} = \frac{1}{N_2}\sum_{k=1}^{N_2}\xi_\varepsilon(\theta_{2,k-1}^\varepsilon)e^{T\kappa_\varepsilon(\theta_{2,k-1}^\varepsilon)} - \Big(\mathbb{E}\Big[\sigma^{-1}(\varepsilon^\beta)(F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta}))\Big]\Big)^2,$$

where for all $\theta \in \Theta_q$, $\xi_\varepsilon(\theta) = \sigma^{-2}(\varepsilon^\beta)\mathbb{E}\left(\left|F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta})\right|^2 e^{-\theta . L_T^\varepsilon}\right)$. Using the convergence in law given by relation (IV.11), the assumption $\sup_{0<\varepsilon\leq 1}\mathbb{E}\left|\sigma^{-1}(\varepsilon)(F(L_T^\varepsilon) - F(L_T))\right|^{2a} < +\infty$ and the independence of L_T and W_T, we deduce that the second term on the right hand side of the above equation vanishes when ε tends to zero. Concerning the first one, we aim to use Lemma 4.3 in Chapter II. So, we only need to prove the equicontinuity of the family $(\xi_\varepsilon)_{\varepsilon>0}$ on any compact subset of Θ_q. First, we prove the simple convergence of ξ_ε to ξ with $\xi(\theta) = \mathbb{E}\Big(\left|\nabla F(L_T).\Sigma^{\frac{1}{2}}W_T\right|^2 e^{-\theta . L_T}\Big)$. For this, we can proceed analogously to the proof of relation (IV.11). More precisely, we use Taylor-Young's expansion with function F, the convergence in law given by (IV.11), the independence of $L_T^\varepsilon - L_T^{\varepsilon^\beta}$ and $L_T^{\varepsilon^\beta}$ and Slutsky's theorem to get

$$\sigma^{-2}(\varepsilon^\beta)\left|F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta})\right|^2 e^{-\theta . L_T^\varepsilon} \xrightarrow[\varepsilon\to 0]{\mathcal{L}} \left|\nabla F(L_T).\Sigma^{\frac{1}{2}}W_T\right|^2 e^{-\theta . L_T}.$$

Now, applying Hölder's inequality with $\tilde{a} = \frac{aq}{a+q}$ yields

$$\mathbb{E}\left|\sigma^{-2}(\varepsilon^\beta)\Big(F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta})\Big)^2 e^{-\theta . L_T^\varepsilon}\right|^{\tilde{a}} \leq \mathbb{E}^{\tilde{a}/a}\left|\sigma^{-1}(\varepsilon^\beta)\Big(F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta})\Big)\right|^{2a}\mathbb{E}^{(a-\tilde{a})/a}e^{-\frac{\tilde{a}a}{a-\tilde{a}}\theta . L_T^\varepsilon}.$$

Using assumptions *H2* and $\sup_{0<\varepsilon\leq 1}\mathbb{E}\left|\sigma^{-1}(\varepsilon)(F(L_T^\varepsilon) - F(L_T))\right|^{2a} < +\infty$, it is easy to check the uniform boundedness with respect to ε of the first term on the right hand side of the above inequality. Concerning the second one, since $q = \frac{\tilde{a}a}{a-\tilde{a}}$ we use relation (IV.18) to deduce the same result. Hence, we have the simple convergence of ξ_ε toward ξ when ε tends to zero. Therefore, it remains to prove the equicontinuity of the family functions $(\xi_\varepsilon)_{\varepsilon>0}$ on any compact subset $K \subset \Theta_q$. Replacing $F(L_T^\varepsilon)$ by $\sigma^{-1}(\varepsilon^\beta)\Big(F(L_T^\varepsilon) - F(L_T^{\varepsilon^\beta})\Big)$ in the steps of the proof of relation (IV.25) and using assumptions *H2* and $\sup_{0<\varepsilon\leq 1}\mathbb{E}\left|\sigma^{-1}(\varepsilon)(F(L_T^\varepsilon) - F(L_T))\right|^{2a} < +\infty$ we prove the existence of a constant $c > 0$ not depending on ε such that

$$|\xi_\varepsilon(\theta) - \xi_\varepsilon(\theta')| \leq c|\theta - \theta'|. \tag{IV.27}$$

Thus, under assumption $(\mathcal{H}_\theta)$, we get the almost sure convergence of $\xi_\varepsilon(\theta^\varepsilon_{2,k})$ toward $\xi(\theta^*)$ as k goes to infinity and ε vanishes. We complete the proof of the first step using the almost sure convergence of $\kappa_\varepsilon(\theta^\varepsilon_{2,k})$ toward $\kappa(\theta^*)$ as k goes to infinity and ε vanishes. This last convergence is obtained thanks to relation (IV.18).

Step 2. The second step of this proof consists on checking the Lyapunov condition *B3* of Theorem 4.1 (see Chapter II). We proceed in the same way as in the second step of the proof of Theorem 5.5. We take $\tilde{a} = \frac{aq+a}{2a+q}$ and we get using the same arguments that $\sum_{k=1}^{N_2} \mathbb{E}\Big[\Big|M^\varepsilon_k - M^\varepsilon_{k-1}\Big|^{2\tilde{a}} |\mathcal{F}_{T,k-1}\Big]$ is bounded by

$$\frac{2^{2\tilde{a}-1}}{N^{\tilde{a}}}\sum_{k=1}^{N_2}\xi_{a,\varepsilon}(\theta^\varepsilon_{2,k-1})e^{(2\tilde{a}-1)T\kappa_\varepsilon(\theta^\varepsilon_{2,k-1})} + \frac{2^{2\tilde{a}-1}}{N^{\tilde{a}}}\Big|\mathbb{E}\Big[\sigma^{-1}(\varepsilon^\beta)(F(L^\varepsilon_T) - F(L^{\varepsilon^\beta}_T))\Big]\Big|^{2a}$$

where for all $\theta \in \Theta_q$, $\xi_{\tilde{a},\varepsilon}(\theta) = \mathbb{E}\Big[\Big|\sigma^{-1}(\varepsilon^\beta)(F(L^\varepsilon_T) - F(L^{\varepsilon^\beta}_T))\Big|^{2\tilde{a}} e^{-(2\tilde{a}-1)\theta.L^\varepsilon_T}\Big]$. Then replacing $F(L^\varepsilon_T)$ by $\sigma^{-1}(\varepsilon^\beta)\Big(F(L^\varepsilon_T) - F(L^{\varepsilon^\beta}_T)\Big)$ in the second step of the proof of Theorem 5.5, the same arguments remain valid thanks to assumptions *H2* and $\sup_{0<\varepsilon\leq 1}\mathbb{E}\,|\sigma^{-1}(\varepsilon)(F(L^\varepsilon_T) - F(L_T))|^{2a} < +\infty$. So, we deduce the existence of $c > 0$ not depending on ε such that

$$\sum_{k=1}^{N_2}\mathbb{E}\left[\Big|M^\varepsilon_k - M^\varepsilon_{k-1}\Big|^{2\tilde{a}} |\mathcal{F}_{T,k-1}\right] \leq \frac{c}{N_2^{\tilde{a}-1}}.$$

This completes the proof. □

Remark 5.8. *Similarly as in the MC case, we still have in mind to reduce the variance associated now to the SR method. This goes back to optimize separately v_1 and v_2. Hence, the optimal choice corresponds to*

$$\theta^*_i = \arg\min_{\theta\in\Theta_{1,3}} v_i(\theta) \quad \textit{and} \quad \theta^*_{i,\varepsilon} = \arg\min_{\theta\in\Theta^\varepsilon_{i,3}} v_{i,\varepsilon}(\theta) \textit{ for } \varepsilon > 0 \quad \textit{and } i \in \{1,2\},$$

where v_i and $v_{i,\varepsilon}$ are presented in section 4. In the same way, the construction of stochastic sequences converging almost surely to these desired targets and satisfying $(\mathcal{H}_\theta)$ is ensured by Corollary 5.3.

6 Numerical results

Now, we present numerical simulations that illustrate the efficiency of the ISSR method throughout the pricing of vanilla options with an underlying asset following an exponential pure jump CGMY model. The CGMY process has been introduced by Carr, Geman, Madan and Yor [16] with the aim to develop a model for the dynamic of equity log-returns which is rich enough to accommodate jumps of finite or infinite activity, and finite or infinite variation. Monte Carlo simulation of the CGMY process has been tackled in the literature specifically by Madan and Yor [49], Poirot and Tankov [55] and Rosinski [59]. A CGMY process is a pure jump process with generating triplet $(0, 0, \nu)$ where for $C > 0, G > 0, M > 0$ and $Y < 2$

$$\nu(dx) = C\frac{e^{-Mx}}{x^{1+Y}}\mathbf{1}_{x>0}dx + \frac{Ce^{-G|x|}}{|x|^{1+Y}}\mathbf{1}_{x<0}dx. \tag{IV.28}$$

Following the notations of [55], we consider the Lévy-Kintchine representation with a truncation function h and a characteristic exponent given by

$$\psi(u) = i\gamma_h u + \int_{\mathbb{R}}(e^{iux} - 1 - iuh(x))\nu(dx) \text{ with } \gamma_h = \int_{\mathbb{R}}(h(x) - x\mathbf{1}_{\{|x|\leq 1\}})\nu(dx),\ u \in \mathbb{R}.$$

- For $1 < Y < 2$ and $h(x) = x$, we have $\gamma_h = \int_{|x|\geq 1} x\nu(dx)$ and

$$\psi(u) = iu\gamma_h + C\Gamma(-Y)\left[M^Y\left((1 - \frac{iu}{M})^Y - 1 + \frac{iuY}{M}\right) + G^Y\left((1 + \frac{iu}{G})^Y - 1 - \frac{iuY}{G}\right)\right].$$

- For $0 < Y < 1$ and $h(x) = 0$, we have $\gamma_h = \int_{|x|\leq 1} x\nu(dx)$ and

$$\psi(u) = iu\gamma_h + C\Gamma(-Y)\left[M^Y\left((1 - \frac{iu}{M})^Y - 1\right) + G^Y\left((1 + \frac{iu}{G})^Y - 1\right)\right].$$

In what follows, we consider the risk neutral model with jumps generalizing the Black Scholes model by replacing the Brownian motion by $(L_t)_{0\leq t\leq T}$ the CGMY process with generating triplet $(\gamma, 0, \nu)$, $\gamma \in \mathbb{R}$ and define the asset price

$$S_t = S_0 \exp(rt + L_t), \text{ where } r > 0 \text{ is the interest rate and } S_0 > 0.$$

To guarantee that $e^{-rt}S_t$ is a martingale we have to impose the condition $\int_{|x|\geq 1} e^x\nu(dx) < \infty$ (which is satisfied as soon as $M > 1$) and the condition

$$\gamma + \int_{\mathbb{R}}(e^y - 1 - y\mathbf{1}_{\{|y|\leq 1\}})\nu(dy) = 0, \tag{IV.29}$$

or in other words $\gamma = -\psi(-i)$.

Now, let us recall that for $0 < \varepsilon < 1$, the approximation $(L_t^\varepsilon)_{t\geq 0}$ of $(L_t)_{t\geq 0}$ is a Lévy process with generating triplet $(\gamma, 0, \nu_\varepsilon)$ where $\nu_\varepsilon(dx) := \mathbf{1}_{\{|x|\geq\varepsilon\}}\nu(dx)$. It is worth to note that $(L_t^\varepsilon)_{t\geq 0}$ can be seen as a compound Poisson process with drift $\gamma_\varepsilon := \gamma - \int_{\varepsilon\leq|x|\leq 1} x\nu(dx)$, see (IV.1). This compound Poisson process can be represented as the difference of two independent processes namely the positive part and the negative one. More precisely, the positive part (resp. the negative part) is a compound Poisson process with jump size $\nu_\varepsilon^+ = \mathbf{1}_{\{x\geq\varepsilon\}}\frac{\nu(dx)}{\nu([\varepsilon,+\infty[)}$ (resp. $\nu_\varepsilon^- = \mathbf{1}_{\{x\leq-\varepsilon\}}\frac{\nu(dx)}{\nu(]-\infty,-\varepsilon])}$) and intensity $\nu([\varepsilon, +\infty[)$ (resp. $\nu(]-\infty, -\varepsilon])$). To simulate these compound Poisson processes, we can use either the classical rejection method as described in Cont and Tankov [20] or an improved method used by Madan and Yor [49]. Indeed, when we simulate the positive part we choose $\nu_{0,\varepsilon}^+$ so that $\frac{d\nu_\varepsilon^+}{d\nu_{0,\varepsilon}^+}(x) = e^{-Mx}\mathbf{1}_{\{x>\varepsilon\}} \leq 1$. Then, according to Rosinski [58] we may simulate the paths of ν_ε^+ from those of $\nu_{0,\varepsilon}^+$ by only accepting all jumps x in the paths of $\nu_{0,\varepsilon}^+$ for which $\frac{d\nu_\varepsilon^+}{d\nu_{0,\varepsilon}^+}(x) > u$ where u is an independent draw from uniform distribution. Hence, we use following algorithm

Algorithm 1 Simulating the positive jump size Z of the CGMY process using Rosinski's rejection

Require: U_1 and U_2 are uniform random variables and $Z = \varepsilon U_1^{-\frac{1}{Y}}$
  if $U_2 > \exp -M.Z$ **then**
    $Z = 0$
  end if
  return Z

In the same way, we simulate the negative jump part by replacing in the above algorithm the parameter M by G.

Our aim is to test our approximation methods for computing the price of a vanilla option with payoff F. To do so, we use the importance sampling technique, introduced in section 4, to approximate the price $e^{-rT}\mathbb{E}F(S_T)$ by

$$e^{-rT}\mathbb{E}\left[F(S_T^{\varepsilon,\theta})e^{-\theta.L_t^{\varepsilon,\theta}+T\kappa_\varepsilon(\theta)}\right], \text{ with } S_T^{\varepsilon,\theta} = S_0\exp(rt + L_t^{\varepsilon,\theta}) \tag{IV.30}$$

where $L_T^{\varepsilon,\theta}$ is also a Lévy process with generating triplet $(\gamma_{\varepsilon,\theta}, 0, \nu_{\varepsilon,\theta})$, where $\nu_{\varepsilon,\theta} = e^{\theta.x}\nu_\varepsilon(dx)$ and $\gamma_{\theta,\varepsilon} = \gamma_\varepsilon + \int_{-1}^1 x(e^{\theta.x}-1)\nu_\varepsilon(dx)$. The choice of θ depends on using the classical MC method or the SR one. According to relation (IV.15), $\theta_{1,\varepsilon}^*$ is the optimal choice for the MC method. However, for the SR method, we omptize separately each quantity appearing in the associated variance and the optimal choice is given by the couple $(\theta_{1,\varepsilon}^*, \theta_{2,\varepsilon}^*)$ (see relation (IV.15)) . To compute these optimal terms, we use the constrained algorithms introduced in the system (IV.19). It is worth to note that in practice it is easier to use $\kappa(\theta)$ instead of $\kappa_\varepsilon(\theta)$.

6.1 One-dimensional CGMY process

In this setting we consider the European call option with payoff $F(x) = (x - \text{Strike})_+$. The parameters of the CGMY model are chosen as follows : $S0 = 100, \text{Strike} = 100, C = 0.0244, G = 0.0765, M = 7.5515, Y = 1.2945$, the free interest rate $r = \log(1.1)$ and maturity time $T = 1$. We run 50000 iteration for the constrained algorithm with the compact set $[-G, M]$. The obtained optimal values are given by $(\theta^*_{1,\varepsilon}, \theta^*_{2,\varepsilon}) = (5.3, 2.5)$ (see Figure IV.1).

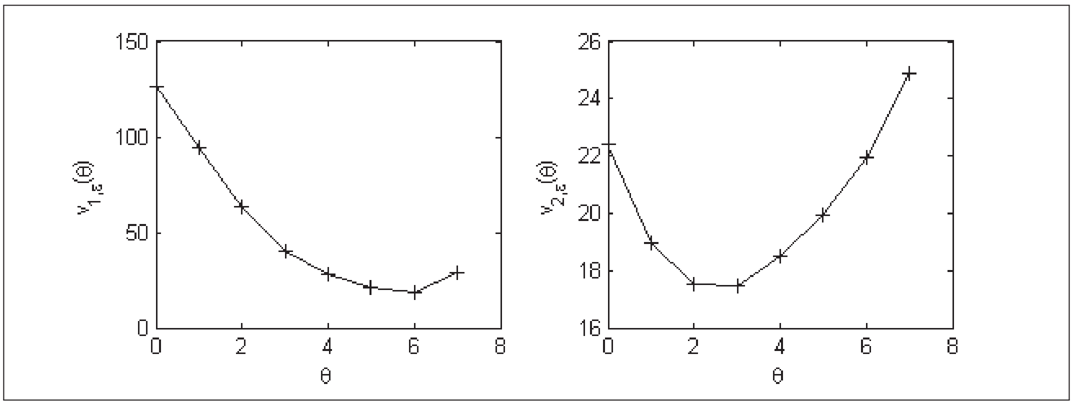

Figure IV.1 – *Variances $v_{1,\varepsilon}$ and $v_{2,\varepsilon}$ versus θ in the one-dimensional setting.*

In order to compare the ISMC algorithm (IV.22) and the ISSR one (IV.26) we use the couple $(\theta^*_{1,\varepsilon}, \theta^*_{2,\varepsilon})$ computed above. For this, we compute for each method the CPU time (per second) (the computations are done on a PC with a 2.5 GHz Intel core i5 processor) and an error measure given by the mean squared error (MSE) which is defined by

$$\text{MSE} = \frac{1}{30}\sum_{i=1}^{30}(\text{Real value} - \text{Simulated value})^2. \tag{IV.31}$$

The real value is obtained using the Fourier-cosine method introduced by Fang and Oosterlee [26] for a one-dimensional CGMY with an accuracy of order 10^{-10}. This method is available in the free online version of Premia platform (`https://www.rocq.inria.fr/mathfi/Premia/index.html`). For this setting, our ISSR algorithm (IV.26) is now available in the latest premium version of Premia.

For different values of ε, we give in Figure IV.2 below the log-log plot of the obtained MSE versus the CPU time for the classical Monte Carlo (MC), the statistical Romberg (SR), the importance sampling Monte Carlo (ISMC) and the importance sampling statistical Romberg (ISSR) methods.

According to Table IV.1 and for a fixed MSE of order $6 \cdot 10^{-3}$, the ISSR method reduces the CPU time by a factor of $8,73$ compared to the ISMC one.

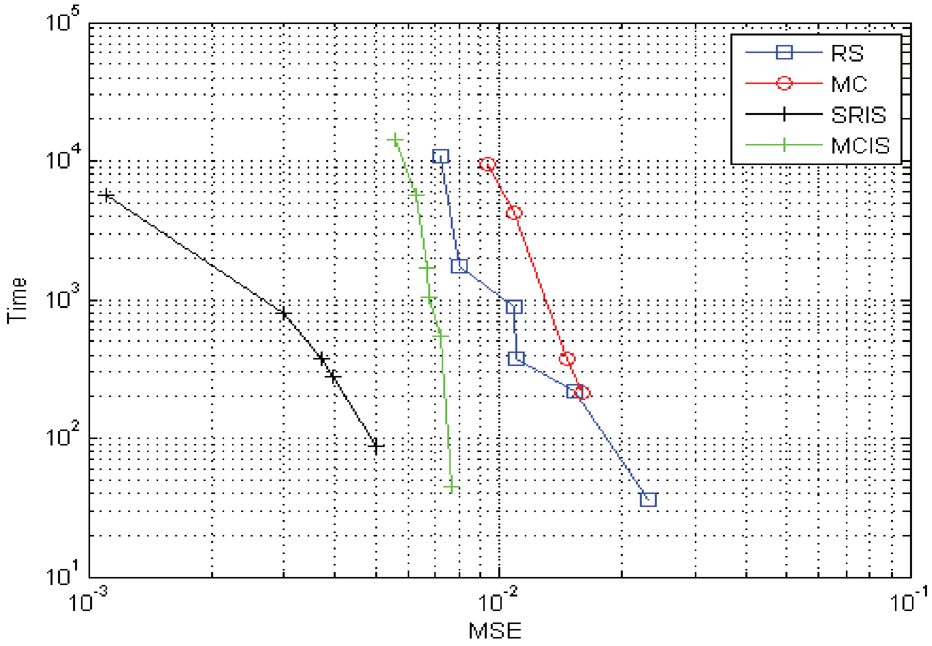

Figure IV.2 – CPU time versus MSE in the one-dimensional setting.

Clearly the ISSR method is the most efficient compared to the other ones.

Time complexity reduction		
MSE	ISMC CPU time	ISSR CPU time
$7 \cdot 10^{-3}$	$7 \cdot 10^2$	$5 \cdot 10^2$
$6,5 \cdot 10^{-3}$	$2 \cdot 10^3$	$6 \cdot 10^2$
$6 \cdot 10^{-3}$	$5,5 \cdot 10^3$	$6,3 \cdot 10^2$
$5,5 \cdot 10^{-3}$	$15 \cdot 10^3$	$7 \cdot 10^2$

Tableau IV.1 – Time complexity reduction (ISSR versus ISMC).

6.2 Two-dimensional CGMY process

We focus now on the computation of a price of the form $e^{-rT}\mathbb{E}F(S_T^1, S_T^2)$, where $F(x,y) = (x + y - \text{Strike})_+$ and the couple $(S_t^1, S_t^2)_{0\leq t\leq T}$ denotes the underlying asset process. In this setting we choose $(S_t^1, S_t^2) = (S_0 e^{rt+L_t^1}, S_0 e^{rt+L_t^2})$ where $(L_t^1)_{0\leq t\leq T}$ and $(L_t^2)_{0\leq t\leq T}$ are two independent CGMY processes with generating triplets $(\gamma_1, 0, \nu_1)$ and $(\gamma_2, 0, \nu_2)$ such that the processes $(e^{-rt}S_t^1)_{0\leq t\leq T}$ and $(e^{-rt}S_t^2)_{0\leq t\leq T}$ are two martingales. So, it amounts to selectγ_1 and γ_2 as in relation (IV.29).

Since the Fourier-cosine method with high accuracy is no more available for the two-dimensional setting, the "Benchmark" price is obtained by running the classical MC algorithm with a very small value of ε. Indeed, for $\varepsilon = 10^{-6}$ the "Benchmark" price is 21.0782 with a CPU time of 24718 seconds. The parameters of the considered two CGMY processes defined by (C, G_1, M_1, Y) and (C, G_2, M_2, Y) are chosen as follows : $C = 0.0244, G_1 = 0.0765, M_1 = 7.55015, G_2 = 2, M_2 = 5, Y = 0.9$, $S0 = 100$, Strike $= 200$, $r = \log(1.1)$ and the maturity time $T = 1$. Using the constrained algorithms (IV.19), we obtain the values of the optimal two-dimensional vectors given by relation (IV.15) and we get $\theta_{1,\varepsilon}^* = (4, 3.5)$ and $\theta_{2,\varepsilon}^* = (3.5, 1.1)$. In Figure IV.3, we plot the evolution of both variances $v_{1,\varepsilon}$ and $v_{2,\varepsilon}$ in terms of $\theta = (\theta_1, \theta_2) \in [-G_1, M_1] \times [-G_2, M_2]$.

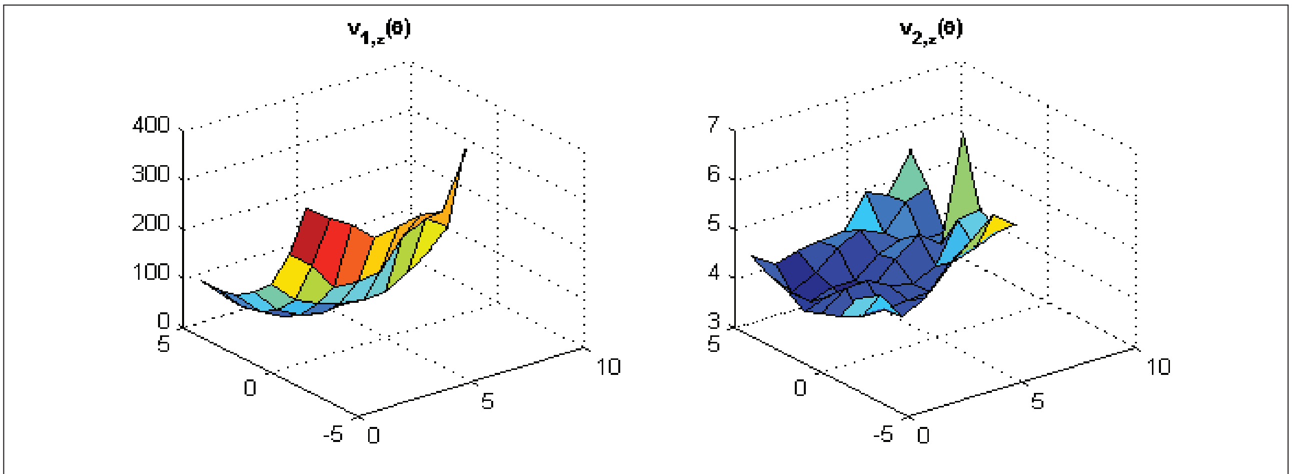

Figure IV.3 – Variances $v_{1,\varepsilon}$ and $v_{2,\varepsilon}$ versus θ in the two-dimensional setting.

Now we proceed as in the one-dimensional case to compare the different methods. Figure IV.4 confirms the superiority of the ISSR method over the other ones and this holds even when we compare it to the ISMC method. Indeed, for a given MSE, the ISSR spends less time than the other methods to compute the desired option price. The difference in terms of computational time becomes more significant as soon as the MSE becomes very small, which corresponds to low values of ε (see Figure IV.4 below).

According to Table IV.2 and for a fixed MSE of order 10^{-3}, the ISSR reduces the CPU time of the considered option price by a factor 2 in comparison to the ISMC method. Moreover, this factor becomes more important when we consider a smaller MSE. In fact, for a fixed MSE of order $3 \cdot 10^{-4}$, the ISSR reduces the CPU time by a factor > 5 in comparison to the ISMC one.

Time complexity reduction		
MSE	ISMC CPU time	ISSR CPU time
10^{-3}	40	20
$6 \cdot 10^{-4}$	100	30
$4 \cdot 10^{-4}$	250	60
$3 \cdot 10^{-4}$	450	80

Tableau IV.2 – Time complexity reduction ISSR versus ISMC.

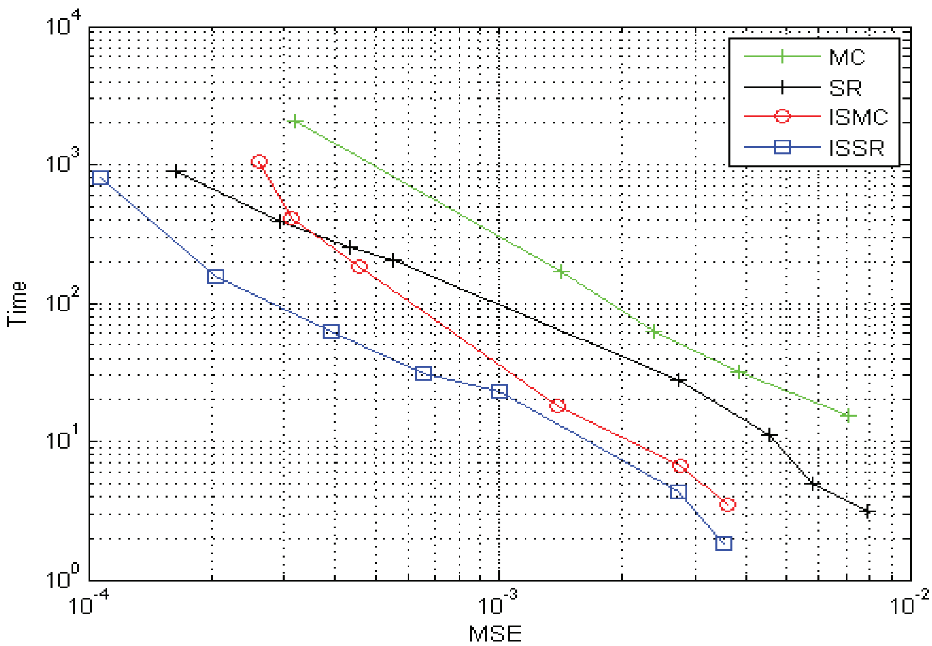

Figure IV.4 – CPU time versus MSE in the two-dimensional setting.

7 conclusion

In this paper, we highlight the superiority of the ISSR method over the classical Monte Carlo approach for the setting of Lévy processes. It may be of interest to extend this study to the setting of Euler discretization schemes for Lévy driven diffusions developed by Protter and Talay [56] and Jacod, Kurtz, Méléard and Protter [34]. Also, a next natural question consists on developing analogous results for path dependent options in exponential Lévy models in the spirit of the works of Dia and Lamberton [22, 23]. These two points will be the object of a forthcoming works.

Bibliographie

[1] C. Andrieu, E. Moulines, and P. Priouret. Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.*, 44(1) :283–312, 2005.

[2] David Applebaum. *Lévy processes and stochastic calculus*, volume 116 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2009.

[3] B. Arouna. Adaptative Monte Carlo method, a variance reduction technique. *Monte Carlo Methods Appl.*, 10(1) :1–24, 2004.

[4] S. Asmussen and P.W. Glynn. *Stochastic simulation : algorithms and analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007.

[5] S. Asmussen and J. Rosiński. Approximations of small jumps of Lévy processes with a view towards simulation. *J. Appl. Probab.*, 38(2) :482–493, 2001.

[6] E. Pardoux B. Lapeyre and R. Sentis. *Méthodes de Monte-Carlo pour les équations de transport et de diffusion*, volume 29 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 1998.

[7] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. *Probab. Theory Related Fields*, 104(1) :43–60, 1996.

[8] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density. *Monte Carlo Methods Appl.*, 2(2) :93–128, 1996.

[9] O.E. Barndorff-Nielsen. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24 :1–13, 1997.

[10] M. Ben Alaya, K. Hajji, and A. Kebaier. Importance Sampling and Statistical Romberg method. *Accepted at Bernoulli Journal*, 2014.

[11] M. Ben Alaya and A. Kebaier. Central limit theorem for the Multilevel Monte Carlo Euler method. *Accepted at Annals of Applied Probability*, 2014.

[12] M. Ben Alaya and A. Kebaier. Multilevel Monte Carlo for Asian options and limit theorems. *Monte Carlo Methods Appl.*, 20(3) :181–194, 2014.

[13] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1990. Translated from the French by Stephen S. Wilson.

[14] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1987.

[15] N. Bouleau and D. Lépingle. *Numerical methods for stochastic processes*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1994. A Wiley-Interscience Publication.

[16] P. Carr, H. Geman, D. Madan, and M. Yor. The fine structure of asset returns : An empirical investigation. *The Journal of Business*, 75 :305–332, 2002.

[17] H. F. Chen, G. Lei, and A. J. Gao. Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.*, 27(2) :217–231, 1988.

[18] H. F. Chen and Y. M. Zhu. Stochastic approximation procedures with randomly varying truncations. *Sci. Sinica Ser. A*, 29(9) :914–926, 1986.

[19] S. Cohen and J. Rosiński. Gaussian approximation of multivariate Lévy processes with applications to simulation of tempered stable processes. *Bernoulli*, 13(1) :195–210, 2007.

[20] R. Cont and P. Tankov. *Financial modelling with jump processes*. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2004.

[21] E. H. A. Dia. Error bounds for small jumps of Lévy processes. *Adv. in Appl. Probab.*, 45(1) :86–105, 2013.

[22] E. H. A. Dia and D. Lamberton. Connecting discrete and continuous lookback or hindsight options in exponential Lévy models. *Adv. in Appl. Probab.*, 43(4) :1136–1165, 2011.

[23] E. H. A. Dia and D. Lamberton. Continuity correction for barrier options in jump-diffusion models. *SIAM J. Financial Math.*, 2(1) :866–900, 2011.

[24] M. Duflo. *Algorithmes stochastiques*, volume 23 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 1996.

[25] M. Duflo. *Random iterative models*, volume 34 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997.

[26] F. Fang and C. W. Oosterlee. A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM J. Sci. Comput.*, 31(2) :826–848, 2008/09.

[27] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[28] M. B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3) :607–617, 2008.

[29] M. B. Giles and L. Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Ann. Appl. Probab.*, 24(4) :1585–1620, 2014.

[30] P. Glasserman. *Monte Carlo methods in financial engineering*, volume 53 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 2004. Stochastic Modelling and Applied Probability.

[31] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Math. Finance*, 9(2) :117–152, 1999.

[32] P. Hall and C. C. Heyde. *Martingale limit theory and its application.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980. Probability and Mathematical Statistics.

[33] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6 :327–343, 1993.

[34] J. Jacod, T. G. Kurtz, S. Méléard, and P. Protter. The approximate Euler method for Lévy driven stochastic differential equations. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(3) :523–558, 2005.

[35] J. Jacod and P. Protter. Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.*, 26(1) :267–307, 1998.

[36] R. Kawai. Optimal importance sampling parameter search for Lévy processes via stochastic approximation. *SIAM J. Numer. Anal.*, 47(1) :293–307, 2008/09.

[37] A. Kebaier. Statistical Romberg extrapolation : a new variance reduction method and applications to option pricing. *Ann. Appl. Probab.*, 15(4) :2681–2705, 2005.

[38] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *Ann. Math. Statistics*, 23 :462–466, 1952.

[39] P. E. Kloeden and E. Platen. Numerical methods for stochastic differential equations. In *Nonlinear dynamics and stochastic mechanics*, CRC Math. Model. Ser., pages 437–461. CRC, Boca Raton, FL, 1995.

[40] H. Kunita. *Stochastic flows and stochastic differential equations*, volume 24 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997. Reprint of the 1990 original.

[41] H. J. Kushner and G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.

[42] H. J. Kushner and G. G. Yin. *Stochastic approximation algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1997.

[43] B. Lapeyre and J. Lelong. A framework for adaptive Monte Carlo procedures. *Monte Carlo Methods Appl.*, 17(1) :77–98, 2011.

[44] S. Laruelle, C. Lehalle, and Gilles G. Pagès. Optimal split of orders across liquidity pools : a stochastic algorithm approach. *SIAM J. Financial Math.*, 2(1) :1042–1076, 2011.

[45] S. Laruelle, C. Lehalle, and G. Pagès. Optimal posting price of limit orders : learning by trading. *Math. Financ. Econ.*, 7(3) :359–403, 2013.

[46] J. Lelong. Almost sure convergence for randomly truncated stochastic algorithms under verifiable conditions. *Statist. Probab. Lett.*, 78(16) :2632–2636, 2008.

[47] V. Lemaire and G. Pagès. Unconstrained recursive importance sampling. *Ann. Appl. Probab.*, 20(3) :1029–1067, 2010.

[48] V. Lemaire and G. Pagès. Multilevel Richardson-Romberg extrapolation. *ArXiv preprint : 1401.1177*, 2013.

[49] D. B. Madan and M. Yor. Representing the CGMY and Meixner Lévy processes as time changed Brownian motions. *J. Comput. Finance*, 12(1) :27–47, 2008.

[50] D.B. Madan and E. Seneta. The variance-gamma model for share market returns. *The journal of Business*, 63 :511–524, 1990.

[51] R.C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1-2) :125–144, 1976.

[52] M. Broadie P. Boyle and P. Glasserman. Monte carlo methods for security pricing. *Journal of Economic Dynamics and Control*, 21(8-9) :1267–1321, June 1997.

[53] G. Pagès. Multi-step Richardson-Romberg extrapolation : remarks on variance control and complexity. *Monte Carlo Methods Appl.*, 13(1) :37–70, 2007.

[54] M. Pelletier. Asymptotic almost sure efficiency of averaged stochastic algorithms. *SIAM J. Control Optim.*, 39(1) :49–72 (electronic), 2000.

[55] J. Poirot and P. Tankov. Monte carlo option pricing for tempered stable (cgmy) processes. *J. Asia-Pacific Financial Markets*, 13(4) :327–344, 2006.

[56] P. Protter and D. Talay. The Euler scheme for Lévy driven stochastic differential equations. *Ann. Probab.*, 25(1) :393–423, 1997.

[57] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statistics*, 22 :400–407, 1951.

[58] J. Rosiński. Series representations of Lévy processes from the perspective of point processes. In *Lévy processes*, pages 401–415. Birkhäuser Boston, Boston, MA, 2001.

[59] J. Rosiński. Tempering stable processes. *Stochastic Process. Appl.*, 117(6) :677–707, 2007.

[60] K. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2013. Translated from the 1990 Japanese original, Revised edition of the 1999 English translation.

[61] W. Schoutens. *Levy Processes in Finance : Pricing Financial Derivatives*. Wiley Series in Probability and Statistics. Wiley, 2003.

[62] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4) :483–509, 1990.

Accélération de la méthode Monte Carlo pour des processus de diffusions et applications en Finance

Résumé. Dans cette thèse, on s'intéresse à la combinaison des méthodes de réduction de variance et de réduction de la complexité de la méthode Monte Carlo. Dans une première partie de cette thèse, nous considérons un modèle de diffusion continu pour lequel on construit un algorithme adaptatif en appliquant l'importance sampling à la méthode de Romberg Statistique Nous démontrons un théorème central limite de type Lindeberg Feller pour cet algorithme. Dans ce même cadre et dans le même esprit, on applique l'importance sampling à la méthode de Multilevel Monte Carlo et on démontre également un théorème central limite pour l'algorithme adaptatif obtenu. Dans la deuxième partie de cette thèse, on développe le même type d'algorithme pour un modèle non continu à savoir les processus de Lévy. De même, nous démontrons un théorème central limite de type Lindeberg Feller. Des illustrations numériques ont été menées pour les différents algorithmes obtenus dans les deux cadres avec sauts et sans sauts.

Mots clés. Algorithmes stochastiques, Robbins-Monro, Théorème limite central, Romberg Statistique, Multilevel Monte Carlo, schéma d'Euler, importance sampling, options exotiques, modèle de Heston, processus de Lévy et approximation, transformation d'Esscher, modèle du CGMY.

Improved Monte Carlo method for diffusion processes and applications in Finance

Abstract. In this thesis, we are interested in studying the combination of variance reduction methods and complexity improvement of the Monte Carlo method. In the first part of this thesis, we consider a continuous diffusion model for which we construct an adaptive algorithm by applying importance sampling to Statistical Romberg method. Then, we prove a central limit theorem of Lindeberg-Feller type for this algorithm. In the same setting and in the same spirit, we apply the importance sampling to the Multilevel Monte Carlo method. We also prove a central limit theorem for the obtained adaptive algorithm. In the second part of this thesis, we develop the same type of adaptive algorithm for a discontinuous model namely the Lévy processes and we prove the associated central limit theorem. Numerical simulations are processed for the different obtained algorithms in both settings with and without jumps.

Keywords. Stochastic algorithm, Robbins-Monro, Central limit theorem, Statistical Romberg method, Multilevel Monte Carlo, Euler scheme, importance sampling, exotic options, Heston model, Lévy processes and approximation, Esscher transform, CGMY model.