Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών
Τομέας Μαθηματικών
Εθνικό Μετσόβειο Πολυτεχνείο

Αλγόριθμοι Στοχαστικής Βελτιστοποίησης με υπερ-γραμμικούς συντελεστές

ΣΤΗΝ ΕΠΙΣΤΗΜΟΝΙΚΗ ΠΕΡΙΟΧΗ: Πιθανότητες και Στοχαστική Ανάλυση

Διπλωματική Εργασία
Μάκρας Νικόλαος
Επιβλέπων: **Σαμπάνης Σωτήριος**, Καθηγητής
Επιτροπή: Βόντα Φιλία, Παπανικολάου Βασίλης

Αθήνα
15 Ιουλίου 2022

Ευχαριστίες

Με το πέρας της διπλωματικής μου εργασίας θα ήθελα να ευχαριστήσω ιδιαίτερα τον επιβλέπων καθηγητή μου, κο. Σωτήρη Σαμπάνη για την συνεργασία μας και την πολύτιμη καθοδήγηση που μου προσέφερε. Μου δόθηκε η ευκαιρία να μελετήσω και να επεκταθώ σε ένα φανταστικό θέμα που πλησιάζει το σύνορο μεταξύ βιβλιογραφικής γνώσης και επιστημονικής έρευνας αίχμης, το οποίο με ωρίμασε ακαδημαϊκά σε όλα τα μέτωπα.
Επίσης θα ήθελα να απευθύνω τις πιο θερμές μου ευχαριστίες στην οικογένεια μου για την αδιάκοπη στήριξη που μου παρείχαν κατά την διάρκεια των σπουδών μου.

Νικόλαος Μάκρας
Αθήνα, Ιούλιος 2022

Νικόλαος Μάκρας, maknik2@gmail.com

Abstract

This dissertation studies how mathematical optimization problems can be solved via Stochastic Differential Equations and most notably Langevin Dynamics.

In the first chapter, we give a brief review of well-known algorithms, discussing the advantages that they offer and their limitations. Hence, making our motivations behind developing and studying the methods presented in the following chapters clear.

It is the purpose of the second chapter to introduce the taming technique to the classical explicit Euler approximations for SDEs, allowing us to relax the conditions called upon their coefficients. Specifically, we show that by taming the superlinearly growing term of the drift on the Euler discretization, we obtain a new numerical explicit scheme which converges to the true solution of the SDE in $\mathcal{L}^p$ sense. To prove the latter its essential to bound the moments of the approximation.

Last but not least, we present the Tamed Unadjusted Langevin Algorithm (TULA) which can be viewed as a Markov Chain Monte Carlo (MCMC) type of algorithm for sampling from a target distribution. Key to this method is to seek a SDE (overdamped Langevin equation) which produces a continuous time process with stationary distribution our desired target and then discretize the process by implementing the techniques mentioned in chapter 2. By doing so we are able to simulate paths for extended lengths of time whose behavior corresponds to the distribution of interest, a limitation of the previous framework which assumed a finite time. The algorithm's capabilities are illustrated through numerical examples.

Keywords: Euler approximations; Tamed schemes; local Lipschitz condition; dissipativity-like condition; superlinear drift; Langevin algorithm; Markov chain Monte Carlo; Total variation; invariant distribution; sampling; optimization

Περίληψη

Κύριος σκοπός του παρόντος πονήματος είναι η μελέτη των Στοχαστικών Διαφορικών Εξισώσεων και ιδιαίτερα των εξισώσεων διάχυσης που προέρχονται απο δυναμικά συστήματα Langevin και το πως μπορούν να χρησιμοποιηθούν για την επίλυση προβλημάτων βελτιστοποίησης.
Το πρώτο κεφάλαιο είναι εισαγωγικό και περιγράφεται το κίνητρο πίσω απο την χρήση των εν λόγω εξισώσεων, τίθεται το πλαίσιο εργασίας των επόμενων κεφαλαίων και δίνεται μια σύντομη χαρτογράφηση του επιστημονικού πεδίου, σχολιάζοντας συγκενικά αποτελέσματα και προβλήματα που παρουσιάζονται σε αυτά.
Στο δεύτερο κεφάλαιο εισάγουμε την έννοια του taming στα κλασσικά αριθμητικά σχήματα Euler για ΣΔΕ, δίνοντας μας την δυνατότητα να χαλαρώσουμε τις συνθήκες που απαιτούμε να ικανοποιούνται απο τους συντελεστές τους. Πιο συγκεκριμένα όταν ο όρος τάσης που συναντάμε είναι υπεργραμμικός, εφαρμόζοντας τις μεθόδους του κεφαλαίου, κατασκευάζουμε ένα νέο σχήμα αριθμητικής επιλύσης ΣΔΕ που καταφέρνει να συγκλίνει ισχυρά στην λύση υπό την έννοια $\mathcal{L}^p$ χώρων.
Οι Markov Chain Monte Carlo αλγόριθμοι που υλοποιούν τις παραπάνω τεχνικές ονομάζονται Tamed Unadjusted Langevin Algorithms (TULA) και αποτελούν τον πυρήνα του τελευταίου κεφαλαίου. Το κεντρικό ζήτημα είναι να δείξουμε πως προσομοιώνοντας στιγμιότυπα της λύσης μιας Langevin ΣΔΕ μέσω των εν λόγο αλγορίθμων, πράγματι αυτά προσεγγίζουν ικανοποιητικά την δειγματοληψία απο την κατανομή ενδιαφέροντος, που δεν είναι άλλη απο την αναλλοίωτη κατανομή της ΣΔΕ. Πέρα απο την θεωρητική θεμελίωση των παραπάνω, προχωράμε στην υλοποίηση των αλγορίθμων και παρέχουμε αριθμητικά παραδείγματα που καταδεικνύουν τα προτερήματα τους.

Contents

Εισαγωγή

Η ολοένα και μεγαλύτερη δίαδοση μεθόδων μηχανικής μάθησης και τεχνητής νοημοσύνης τόσο στους ακαδημαϊκούς χώρους πολλαπλών θετικών επιστημών όσο και στην αγορά έχει συντελέσει στην αναζοπύρωση του ερευνητικού ενδιαφέροντος στην σχεδίαση και μελέτη αποτελεσματικών αλγορίθμων βελτιστοποίησης. Η συντριπτική πλειοψηφία αυτών των μεθόδων εμπεριέχουν ένα βήμα ελαχιστοποίησης, αλλά ενώ το πρόβλημα αυτό έχει μελετηθεί εκτενώς απο την μαθηματική κοινότητα, τα υπάρχοντα εργαλεία απο την κλασσική βιβλιογραφεία δεν επαρκούν για την επίλυση του, καθώς οι συνθήκες που επικαλούνται δεν πληρούνται σε ρεαλιστικά σενάρια. Χαρακτηριστικά, συναρτήσεις που προέρχονται απο την εκπαίδευση νευρωνικών δικτύων αποτυγχάνουν να ικανοποιούν ιδιότητες όπως ισχυρή κυρτότητα, κυρτότητα και γραμμική αύξηση ενώ οι υλοποιήσεις απλών αλγορίθμων συνήθως δεν έχουν καλή συμπεριφορά σε μεγάλες κλίμακες, με την έννοια χώρων μεγάλης διάστασης. Ξεκινάμε με το πιο γενικό πλαίσιο εργασίας, όπου αναζητάμε $x^* \in \mathbb{R}^d$ τέτοιο ώστε

$$x^* = \underset{x\in\mathbb{R}^d}{\text{argmin}} U(x)$$

όπου $U : \mathbb{R}^d \mapsto \mathbb{R}$ είναι η ποσότητα που επιθυμούμε να ελαχιστοποιήσουμε. Αν U είναι τουλάχιστον συνεχώς διαφορίσιμη, τότε λαμβάνουμε τον ακόλουθο απλοϊκό αλγόριθμο καθόδου κλίσης μέσω της διακριτοποίησης του: $dx(t) = -\nabla U(x(t))dt$:

$$x_{n+1} = x_n - h\nabla U(x_n)$$

Μια σχολή σκέψης επισυγκεντρώνεται στο να ενισχύσει τον παραπάνω ντετερμενιστικό αλγόριθμο με τεχνικές τυχαίας εξερεύνησης πάνω στην επιφάνεια της αντικειμενικής συνάρτησης, το οποίο επιτυγχάνεται διαταράζοντας στοχαστικά το κλασσικό σχήμα. Υπο αυτό το πρίσμα, ο αλγόριθμος πλέον δεν κατευθύνεται αποκλειστικά στην κατεύθυνση της πιο απότομης κλίσης και έτσι καθίσταται δυνατή η αποφυγή παγίδευσης του αναδρομικού σχήματος σε τοπικά ελάχστικα ή σαγματικά σημεία. Ωστόσο και πάλι η σύγκλιση στο πραγματικό ελάχιστο δεν είναι εξασφαλισμένη και εν γένη το σημείο που θα καταλήξει ο αλγόριθμος εξαρτάται απο τις συνθήκες αρχικοποίησης. Μια πρώτη στοχαστική παραλλαγή είναι η προσθήκη ενός όρου λευκού θορύβου, έτσι οδηγούμαστε:

$$dX(t) = -\nabla U(X(t))dt + \sigma(t, X(t))dW_t \ , \quad \forall t \in [0, T]$$

με αρχική συνθήκη $X(0) = x_0$ σ.β. πεπερασμένη $\mathcal{F}_0-$μετρήσιμη, όπου $\sigma(t, x) : \mathbb{R} \times \mathbb{R}^d \mapsto \mathbb{R}^{d\times m}$ ο συντελεστής διάχυσης είναι μια $\mathcal{B}(\mathbb{R}_+)\otimes\mathcal{B}(\mathbb{R}^d)$-μετρήσιμη συνάρτηση και $(W(t))_{t\geq 0}$ είναι μια $m-$διάστατη κίνηση Brown. Τότε η διακριτοποίηση Euler-Maruyama ορίζεται ως:

$$X(t_{n+1}) = X(t_n) - h\nabla U(X(t_n)) + \sigma(t_n, X(t_n))\Delta W_n$$

όπου $\Delta W_n = W_{t_{n+1}} - W_{t_n}$ είναι i.i.d. κανονικές $m-$διάστατες τυχαίες μεταβλητές με μέση τιμή μηδέν και κύμανση h, εξού και η εναλλακτική γραφή

$$X(t_{n+1}) = X(t_n) - h\nabla U(X(t_n)) + \sqrt{h}\sigma(t_n, X(t_n))Z_{n+1}$$

με $(Z_n)_{n\geq 1}$ να είναι i.i.d. $m-$δίαστατες τυποποιημένες κανονικές τ.μ.. Το αναδρομικό σχήμα μπορεί να βελτιωθεί περαιτέρω υιοθετόντας μεταβαλλόμενο βήμα h_n ή τεχικές momentum, κατασκευάζοντας έτσι μια οικογένεια αλγορίθμων βελτιστοποίησης που γιατρεύουν αρκετές απο τις παθογένειες του SGD, με διαφορά ο πιο διαδεδομένος τέτοιος αλγόριθμος αυτήν την στιγμή είναι ο ADAM [1]. Παρόλο που για αυτούς τους αλγορίθμους είναι εμπειρικά γνωστό ότι είναι αποδοτικοί για την εκπαίδευση νευρωνικών δικτύων, υπάρχουν ακόμα θεωρητικά κενά στην μαθηματική τους θεμελίωση [2].

Η άλλη κύρια ερευνητική κατεύθυνση επισυγκεντρώνεται στον σχεδιασμό αλγορίθμων τύπου Markov Chain Monte Carlo [3] με στόχο την κατασκευή εργοδικών MC των οποίων η αναλλοίωτη κατανομή επιτρέπει την εκτίμηση της συνάρτησης ενδιαφέροντος. Η δειγματοληψία απο μία τέτοια κατανομή σε συνδιασμό με τον αλγόριθμο simulated annealing καθιστούν ένα αποτελεσματικό πρόγραμμα βελτιστοποίησης, ειδικά στην περίπτωση χώρων υψηλών διαστάσεων. Έστω τώρα μια αντικειμενική συνάρτηση $U(x)$, πάντα μπορούμε να θεωρήσουμε την ακόλουθη συνάρτηση κατανομής ως προς το μέτρο Lebesgue

$$\pi_\beta(x) = e^{-\beta U(x)} / \int_{\mathbb{R}^d} e^{-\beta U(y)} dy$$

όπου $\beta > 0$ είναι μια παράμετρος κλίμακας και θα αναφερόμαστε σε αυτήν ως παράμετρο θερμοκρασίας. Είναι προφανές πως άμα μπορούμε να προσομοιώνουμε τιμές απο την π_β τότε πολύ απλά εντοπίζουμε το $\text{argmin}U(x)$ εκτιμώντας ολόκληρη την συνάρτηση, μια τέτοια ιδέα όμως είναι σπάταλη και υπολογιστικά ασύμφορη. Στην συνέχεια θα δούμε πως παρακάμπτεται αυτό το βήμα, δειγματοληψόντας μόνο απο τις περιοχές της συνάρτησης που έχουν σημασία. Κάτω απο ορισμένες για την $U(x)$ είναι γνωστό απο την θεωρία διάχυσης πως η πυκνότητα π_β σχετίζεται με την στοχαστική εξίσωση διάχυσης Langevin:

$$dY_t = -\nabla U(Y_t)dt + \sqrt{2\beta^{-1}} dW_t$$

υπο την έννοια πως η λύση $(Y_t)_{t>0}$ είναι εκθετικά εργοδική ως προς την αναλλοίωτη κατανομή της, που έχει ως πυκνότητα την π_β. Χρησιμοποιώντας ξανά την μέθοδο Euler-Maruyama οδηγούμαστε στην διακριτό σχήμα της εξίσωσης Langevin, γνωστό ως Unadjusted Langevin Algorithm (ULA)

$$X(t_{n+1}) = X(t_n) - h\nabla U(X(t_n)) + \sqrt{2h\beta^{-1}} Z_{n+1}$$

Η διακριτή διαδικασία $(X_n)_{n\in\mathbb{N}}$ που ορίζεται απο το παραπάνω σχήμα είναι Μαρκοβιανή της οποίας η αναλλοίωτη κατανομή μπορεί να προσεγγίσει την π_β. Παρόλο που οι κατανομές που προέρχονται απο την συνεχή διαδικασία και την διακριτή της έκδοση δεν είναι ταυτόσημες, κάτω από υποθέσεις όπως αυτήν της ολικής Lipschitz συνέχειας για την $\nabla U(x)$, μη-ασυμπτωτικά φράγματα για τις αποστάστεις Wasserstein και την ολική κύμανση μεταξύ των π_β και της κατανομής της $(X_n)_{n\in\mathbb{N}}$, μπορούν να καθιερωθούν. Ακόμα και αυτή όμως αποτελεί μια συνθήκη που συχνά δεν ικανοποιείται και έχει δειχθεί πως ο αλγόριθμος ULA σε αυτήν την περίπτωση είναι ασταθής. Συγκεκριμένα, άμα ο συντελεστής τάσης $\nabla U(x)$ είναι υπεργραμμικός, δηλαδή $\liminf_{\|x\|\to\infty} \|\nabla U(x)\|/\|x\| = +\infty$, είναι γνωστό πως οι ροπές της $(X_n)_{n\in\mathbb{N}}$ αποκλείνουν στο άπειρο [4]. Ένας τρόπος να καταπολεμηθεί αυτή η συμπεριφορά είναι η προσθήκη ενός βήματος Metropolis-Hastings [5]. Πράγματι άμα φανταστούμε τις επαναλήψεις του ULA ως πιθανές κινήσεις εντός ενός τυχαίου περιπάτου με πιθανότητα αποδοχής:

$$a_h(x,y) = 1 \wedge \frac{q_{h,\beta}(y,x)\pi_\beta(y)}{q_{h,\beta}(x,y)\pi_\beta(x)}$$

όπου $q_{h,\beta}(y,x)$ είναι η πυκνότητα μετάβασης και προσθέτοντας αυτό το βήμα απόρριψης-αποδοχής, κατασκευάζεται ο Metropolis-Adjusted Langevin Algorithm (MALA)

$$X^M(t_{n+1}) = \begin{cases} X(t_{n+1}), \text{ if } u_n < a_h(X^M(t_n), X(t_{n+1})) \\ X^M(t_n), \text{ otherwise} \end{cases}$$

όπου $(u_n)_{n\in\mathbb{N}} \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0,1)$. Εκ κατασκευής ο MALA εξασφαλίζει την αντιστρεψιμότητα της αλυσίδας ως προς την π_β διατηρώντας το αναλλοίωτο μέτρο της συνεχής έκδοσης. Αν και βελτίωση απο τον ULA, στην υπεργραμμική περίπτωση ο MALA πάλι μπορεί να αποτυγχάνει να είναι γεωμετρικά εργοδικός[6]. Ως πιθανή λύση έχει προταθεί στην βιβλιογραφεία η παραλλαγή MALTA στην οποία

ο όρος τάσης αντικαθίσταται απο μια truncated εκδοχή του στις περιοχές όπου στο υπόβαθρο το σχήμα Euler εκρήγνυται[6][7]. Αυτό επιτυγχάνεται διατηρώντας την κατεύθυνση του όρου τάσης ενώ η τιμή του κανονικοποιείται. Σε αντίθεση με τον MALA, για τον MALTA μπορεί να αποδειχθεί ότι είναι γεωμετρικά εργοδικός ωστόσο χάνεται η άμεση σχέση με την πρωταρχική διαφορική εξίσωση. Στην σύγχρονη βιβλιογραφεία των προσεγγίσεων Euler έχει προταθεί μια νέα οικογένεια αριθμητικών σχημάτων, όπου στην περίπτωση που δεν είναι διαθέσιμη η συνθήκη της ολίκα Lipschitz συνέχειας, ο όρος τάσης μπορεί να τροποποιηθεί με τέτοιον τρόπο ώστε να είναι ολικά φραγμένος[8][9]. Απο εδώ και στο εξής θα αναφερόμαστε σε αυτήν την τεχνική ως taming. Η υπολογιστική αποδοτικότητα της οικογένειας αυτής και συνδιασμό με το γεγονός πως εξασφαλίζουν $\mathcal{L}^p$-ισχυρή σύγκλιση, ήταν καθοριστικής σημασίας για να αποτελέσουν κίνητρο επέκτασης του taming και σε προβλήματα δειγματοληψίας. Έστω τώρα το σχήμα διακριτοποίησης:

$$X(t_{n+1}) = X(t_n) - hT_h(X(t_n)) + \sqrt{2h\beta^{-1}}Z_{n+1}$$

για κατάλληλη επιλογή taming συνάρτησης $T_h : \mathbb{R}^d \mapsto \mathbb{R}^d$, το άνω σχήμα είναι γνωστό πως παράγει μια μακροβιανή αλυσίδα της οποίας η αναλλοίωτη κατανομή συγκλίνει στην π_β και ταυτόχρονα κληρονομεί την εκθετική εργοδικότητα της συνεχούς έκδοσης ως γεωμετρική, δίνοντας ταχύς ρυθμούς σύγκλισης. Βασιζόμενοι στο [10], υπάρχουν δύο διαθέσιμες επιλογές T_h:

$$G_h(x) = \frac{\nabla U(x)}{1 + h\|\nabla U(x)\|} \text{ and } G_h^c(x) = \left(\frac{\partial_i U(x)}{1 + h|\partial_i U(x)|}\right)_{i\in 1,\dots,d}$$

Η πρώτη επιλογή εφαρμόζει το taming σε όλη την κλίση, οδηγώντας μας στον Tamed Unadjusted Langevin Algorithm (TULA). Εναλλακτικά η δεύτερη taming συνάρτηση μαζί με το άνω διακριτό σχήμα καλούνται, coordinate- wise Tamed Unadjusted Langevin Algorithm (TULAc) που σε αντιδιαστολή με τον TULA, το taming εφαρμόζεται ανεξάρτητα σε κάθε συνιστώστα. Πειραματικά δεδομένα στο [10] δείχνουν πως ο TULAc αποδίδει ανώτερα απο τον TULA όταν συγκρίνονται τα σφάλματα της πρώτης και δεύτερης ροπής των προσομοιωμένων τιμών, ειδικά για μεγάλες επιλογές βήματος h, και προτιμάται για τον σχεδιασμό αλγορίθμων που έχουν ως σκοπό την επίλυση προβλημάτων βελτιστοποίησης σε χώρους μεγάλων διαστάσεων [11]. Επίσης αφήνεται να υπονοηθεί απο την σύγχρονη βιβλιογραφεία, πως ανάλογα το αρχικό πρόβλημα μπορούν να προταθούν παραπάνω taming συναρτήσεις που θα εκμεταλεύονται την δομή της ∇U, πιο συγκεκριμένα κανείς θα μπορούσε να απομονώνει όλες τις υπεργραμμικότητες σε έναν όρο και να εφαρμόζει taming αποκλειστικά σε αυτόν, χωρίς να φράζει ομοιόμορφα όλον τον όρο τάσης. Αυτή είναι η κινητήριος ιδέα πίσω απο τα αποτελέσματα του Κεφαλαίου 2. Σε αυτό το σημείο οφείλουμε να διαλευκάνουμε μια λεπτομέρεια που δεν σχολιάστηκε επαρκώς. Μέχρι στιγμής μιλήσαμε για το πως MCMC αλγόριθμοι μας βοηθάνε να δειγματολήπτουμε απο μια κατανομή της επιλογής μας, αλλά ο αρχικός στόχος μας ήταν να ελαχιστοποιήσουμε μια αντικειμενική συνάρτηση. Αυτό το κενό ανάμεσα στην βελτιστοποίηση και την δειγματοληψία έρχεται να καλύψει το βήμα του simulating annealng. Όπως ήδη αναφέραμε, αν εξαιρέσουμε το ιδεατό σεναρίο όπου όλη η μάζα πιθανότητας της κατανομής βρίσκεται σε μια περιοχή του x^*, προσομοιώνοντας μονοπάτια με τους παραπάνω αλγορίθμους θα χαλάμε πολύτιμο υπολογιστικό χρόνο εξερευνώντας περιοχές χαμηλού ενδιαφέροντος. Υλοποιώντας ωστόσο οποιονδήποτε απο τους αλγορίθμους αυτούς με simulating annealng, θα κατασκευάζουμε μαρκοβιανές αλυσίδες που πλέον η αναλλοίωτη κατανομή τους στο βήμα n δεν θα είναι πλέον $\pi_\beta(x)$ αλλά $\pi_{\beta_n}(x)$ όπου $(1/\beta_n)_{n\in\mathbb{N}}$ είναι ένα φθήνον πρόγραμμα 'ψύξης' τέτοιο ώστε $\lim_{n\to\infty} 1/\beta_n = \infty$. Ο λόγος αυτής της παρέμβασης είναι πως η κατανομή $\lim_{n\to\infty}\pi_{\beta_n}(x)$ συγκεντρώνει ασυμπτωτικά όλην της την μάζα στο ολικό μέγιστο της $\pi(x)$, τουτέστιν το ελάχιστο της $U(x)$. Η βέλτιστη επιλογή της αρχικοποίησης β_0 αλλά και της βηματικής εξέλιξης διαφέρουν ανά εφαρμογή αλλά είναι σύνηθες πως το πρόγραμμα 'ψύξης' πρέπει να είναι αργό π.χ. λογαριθμικό ωστέ να μην επιδρά στην φυσική πορεία των μονοπατιών. Έτσι μέτα απο μία σύντομη περίοδο προσαρμογής

της αλυσίδας στην δυναμική της ΣΔΕ (burn-in period), καταλήγουμε να δειγματοληπτούμε σχεδόν αποκλειστικά απο μια μικρή γειτονία του x^*, που είναι και το επιθυμητό αποτέλεσμα.

Tamed Euler Approximation Schemes

Απο την κλασσική βιβλιογραφεία γνωρίζουμε πως είναι σύνηθες οι έμμεσοι μέθοδοι να μας προσφέρουν αριθμητικά σχήματα που είναι κατά πολύ πιο πλούσια σε ιδίοτητες ευστάθιας απο τις αντίστοιχες άμεσες εκδοχές τους, πάντα όμως με ένα σημαντικό κόστος στην αυξημένη υπολογιστική πολυπλοκότητα. Στο πεδίο της αριθμητικής επίλυσης ΣΔΕ, αν κανείς επιχειρήσει να εφαρμόσει τα δεύτερα σε εξισώσεις των οποίων οι όροι δεν είναι επαρκώς ομαλές συναρτήσεις δεν είναι καν εξασφαλισμένο πως οι μαρκοβιανές αλυσίδες δεν θα εκρήγνονται προς το άπειρο, όπως μαρτυρούν οι Kloden και Hutzenthaler στο [4]. Πρόσφατα μια νέα οικογένεια άμεσων (tamed) αριθμητικών σχημάτων έχει αναπτυχθεί [9][8] που κατέχουν ταυτόχρονα το προτερήματα της ευστάθειας και απαιτούν μικρό υπολογιστικό κόστος κατα την υλοποίηση τους. Στο [9] αποδυκνύεται πως υπό συνθήκες γραμμικής αύξησης και μορφές τοπικής Lipschitz συνέχειας στους συντελεστές μια ΣΔΕ διάχυσης, τα (tamed) άμεσα σχήματα Euler Maruyama συγκλίνουν στον $\mathcal{L}^p$ στην πραγματική λύση της ΣΔΕ και οι κλασσικοί ρυθμοί σύγκλισης μπορούν να εξασφαλιστούν αμα υποθέσει κανείς περαιτέρω ότι περιορίζεται στην κλάση συναρτήσεων με το πολύ πολυωνυμική αύξηση. Βασικό κλειδί για αυτό το αποτέλεσμα είναι η ομοιόμορφη φραγή των ροπών που παράγονται απο την διακριτοποίηση. Στόχος μας είναι να επεκτείνουμε την ιδέα αυτήν στην περίπτωση όπου ο συντελεστής τάσης μπορεί να εκφραστεί ως το άθροισμα δύο επιμέρους όρων, εκ το οποίων ο πρώτος θα ικανοποιεί την κλασσική συνθήκη Lipschitz συνέχειας ενώ ο δεύτερος θα ακολουθεί τις συνθήκες που διατυπώνονται στο [9]. Αυτό μας δίνει την δυνατότητα να εφαρμόσουμε ένα είδος μερικού taming μόνο στον όρο που εμπεριέχει τις υπεργραμμικότητες. Ακόμα υποθέτουμε πως μόνο ένα πεπερασμένο πλήθος των ροπών της αρχικής συνθήκης είναι φραγμένο, καθώς αυτό είναι το σενάριο που συναντά κανείς σε ρεαλιστικά προβλήματα. Αυτή η μικρή αλλά ουσιαστική διαφοροποίηση απο την βιβλιογραφεία μας αναγκάζει να προσεγγίσουμε την φραγή των ροπών με διαφορετικά τεχνικά εργαλεία.

Πλαίσιο Εργασίας

Έστω $(\Omega, \{\mathcal{F}_t\}_{t\geq 0}, \mathcal{F}, \mathbb{P})$ ένας φιλτραρισμένος χώρος πιθανότητας με τις συνήθεις συνθήκες, θεωρούμε μια ΣΔΕ της μορφής:

$$dX(t) = g(t, X(t))dt + \sigma(t, X(t))dW_t, \qquad \forall t \in [0, T]$$

με αρχική συνθήκη $X(0) = x_0$ σ.β. πεπερασμένη $\mathcal{F}_0-$μετρήσιμη, όπου $\sigma(t, x) : \mathbb{R} \times \mathbb{R}^d \mapsto \mathbb{R}^{d\times m}$ είναι $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B}(\mathbb{R}^d)$-μετρήσιμη συνάρτηση και $(W(t))_{t\geq 0}$ μια τυπική $m-$διάστατη κίνηση Brown. Επιπλέον ο συντελεστής τάσης μπορεί να γραφτεί ως:

$$g(t, x) = b(t, x) + f(x), \qquad \forall (t, x) \in [0, T] \times \mathbb{R}^d$$

όπου έχουμε υποθέσει την $b(t, x) : \mathbb{R} \times \mathbb{R}^d \mapsto \mathbb{R}$ να είναι $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B}(\mathbb{R}^d)$-μετρήσιμη και την $f(\cdot)$ να είναι Lipschitz συνεχής και επομένως $\mathcal{B}(\mathbb{R}^d)$-μετρήσιμη. Τότε υπάρχει θετική σταθερά L_f τέτοια ώστε,

$$|f(x) - f(y)| \leq L_f|x - y|$$

Τώρα ορίζουμε το παρακάτω αριθμητικό σχήμα

$$dX_n(t) = g_n(t, X_n(k_n(t)))dt + \sigma(t, X_n(k_n(t)))dW_t, \quad k_n(t) = [nt]/n, \quad \forall t \in [0, T] \text{ and } \forall n \geq 1$$

όπου ο όρος τάσης είναι μερικώς tamed υπο την έννοια πως $g_n(t,x) = b_n(t,x) + f(x)$, ενώ έχουμε ορίσει

$$b_n(t,x) := \frac{1}{1+n^{-a}|b(t,x)|}b(t,x)$$

$\forall t \in [0,T], x \in \mathbb{R}^d$ και $a \in (0,1/2]$. Επίσης παρατηρείστε πως $|b_n(t,x)| \leq \min(n^a, |b(t,x)|)$

Υποθέσεις

Για λόγους συνέπειας απο εδώ και στο εξής κάθε επίκληση με πρόθεμα 'S' θα θεωρείται απευθείας αναφορά στο [9]. Ξεκινάμε δείχνοντας πως όταν ισχύουν οι υποθέσεις S.A1, S.A2, S.A3, S.A5 τότε λαμβάνουμε αντίστοιχες προτάσεις και για την συνάρτηση $g(t,x)$.

S.A1 Υπάρχει θετική σταθερά K τέτοια ώστε,

$$2xb(t,x) \vee |\sigma(t,x)|^2 \leq K(1+|x|^2)$$

για κάθε $t \in [0,T]$ και $x \in \mathbb{R}^d$.

S.A2 Για κάθε $R > 0$, υπάρχει θετική σταθερά L_R τέτοια ώστε, για κάθε $t \in [0,T]$,

$$2(x-y)(b(t,x)-b(t,y)) \vee |\sigma(t,x)-\sigma(t,y)|^2 \vee L_R|x-y|^2$$

για κάθε $|x|, |y| \leq R$
S.A3 Για κάθε $R \geq 0$, και $p > 0$, υπάρχει $N_R \in \mathbb{L}^p$ τέτοια ώστε,

$$\sup_{|x|\leq R} |b(t,x)| \leq N_R(t)$$

για κάθε $t \in [0,T]$
S.A5 Υπάρχουν θετικές σταθερές ℓ και L τέτοιες ώστε, για κάθε $t \in [0,T]$,

$$(x-y)(b(t,x)-b(t,y)) \vee |\sigma(t,x)-\sigma(t,y)|^2 \leq L|x-y|^2$$

και

$$|b(t,x)-b(t,y)| \leq L(1+|x|^\ell + |y|^\ell)|x-y|$$

για κάθε $x, y \in \mathbb{R}^d$

A4 Για κάθε $p \leq p_0$, $\mathbb{E}[|X(0)|^p] < \infty$ όπου $p_0 \geq 2$
Σημειώστε πώς τότε η $f(\cdot)$ αυξάνεται το πολύ γραμμικά:

$$|f(x)| \leq |f(x)-f(0)| + |f(0)| \leq L_f|x-0| + |f(0)| \leq \max(L_f, |f(0)|)(1+|x|) = K_f(1+|x|)$$

Προσέξτε πως,

$$2xf(x) \leq 2|x||f(x)| \leq 2K_f(|x|+|x|^2) \leq 2K_f(1+|x|)^2 \leq 4K_f(1+|x|^2)$$

Τώρα λόγω της (S.A1) εύκολα φάινεται πως,

$$2xg(t,x) = 2xb(t,x) + 2xf(x) \leq (K+4K_f)(1+|x|^2) = C_1(1+|x|^2)$$

Ακόμα υπό την (S.A2) έχουμε,

$$\begin{aligned}2(x-y)(g(t,x)-g(t,y)) &= 2(x-y)(b(t,x)-b(t,y)) + 2(x-y)(f(x)-f(y)) \\ &\leq L_R|x-y|^2 + 2|x-y|L_f|x-y| \leq (L_R+2L_f)|x-y|^2 = C_R|x-y|^2\end{aligned}$$

Αντίστοιχα απο την (S.A3)

$$\sup_{|x|\leq R}|g(t,x)| \leq \sup_{|x|\leq R}|b(t,x)| + \sup_{|x|\leq R}|f(x)| \leq N_R(t) + K_f(1+R) = G_R(t) \in \mathbb{L}^p$$

Εν τέλη η (S.A5) συνεπάγεται:

$$\begin{aligned}|g(t,x)-g(t,y)| &\leq |b(t,x)-b(t,y)| + |f(x)-f(y)| \\ &\leq L(1+|x|^\ell + |y|^\ell)|x-y| + L_f|x-y| \\ &\leq L(1+|x|^\ell + |y|^\ell)|x-y| + L_f(1+|x|^\ell + |y|^\ell)|x-y| \\ &\leq (L+L_f)(1+|x|^\ell + |y|^\ell)|x-y| = C_2(1+|x|^\ell + |y|^\ell)|x-y|\end{aligned}$$

Παρατήρηση 2.1

Για κάθε $n \geq 1$, η $g_n(t,x)$ αυξάνεται το πολύ γραμμικά. Για να δειχθεί αυτό, πρέπει να αξιοποιηθεί η (S.A2)

$$\begin{aligned}|g_n(t,x)| &\leq |b_n(t,x)| + |f(x)| \leq n^a + K_f(1+|x|) \leq n^a(1+|x|) + K_f(1+|x|) \\ &\leq (n^a+K_f)(1+|x|) \leq (1+\frac{K_f}{n^a})n^a(1+|x|) \leq (1+K_f)n^a(1+|x|) \\ &\leq C_3 n^a(1+|x|)\end{aligned}$$

Βλέπουμε λοιπόν πως και οι δυό νόρμες των $g_n(t,x)$, $\sigma(t,x)$ αυξάνονται το πολύ γραμμικά για κάθε $n \geq 1$. Αυτό εξασφαλίζει την υπάρξη και μοναδικότητα της λύσης για την νέα ΣΔΕ για κάθε σταθερό $n \geq 1$, επιπλέον για κάθε $p \leq p_0$ είναι:

$$\sup_{0\leq t\leq T}\mathbb{E}[|X_n(t)|^p] \leq N := N(n,p,T,\mathbb{E}[|X(0)|^p])$$

υπο την υπόθεση (A4).

Λήμμα 2.2

Θεωρείστε την ΣΔΕ που έχουμε αναπτυξεί παραπάνω. Αν για κάποιο $p \geq 2$,

$$\sup_{n\geq 1}\sup_{0\leq t\leq T}\mathbb{E}\left[|X_n(t)|^p\right] \leq \infty$$

και η (S.A1) ισχύουν, τότε:

$$\sup_{0\leq t\leq T}\mathbb{E}\left[|X_n(t) - X_n(k_n(t))|^p\right] \leq C_4 n^{-p/2}$$

όπου C_4 σταθερά ανεξάρτητη του n.

Παρατήρηση 2.3

Δείτε πως κάτω απο την (S.A1) έχουμε:

$$\begin{aligned}2xg_n(t,x) = 2xb_n(t,x) + 2xf(x) &\leq 2x\frac{b(t,x)}{1+n^{-a}|b(t,x)|} + 2|x||f(x)| \\ &\leq K\frac{1+|x|^2}{1+n^{-a}|b(t,x)|} + 2K_f(|x|+|x|^2) \leq K(1+|x|^2) + 2K_f(1+|x|)^2 \\ &\leq K(1+|x|^2) + 4K_f(1+|x|^2) \leq C_5(1+|x|^2)\end{aligned}$$

Λεμμα 2.4

Υποθέτουμε τις (S.A1) και A4 τότε για κάποια $C_6 := C_6(T, K, K_f, \mathbb{E}\left[|X(0)|^2\right])$

$$\sup_{n\geq 1}\sup_{0\leq t\leq T}\mathbb{E}\left[|X_n(t)|^2\right] \leq C_6$$

Λήμμα 2.5

Υποθέτουμε τις (S.A1) και (S.A4) τότε για κάποια σταθερά $C_8 = C_8\left(T, K, K_f, \mathbb{E}\left[|X(0)|^p\right]\right)$

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|X(t)|^p\right] \vee \sup_{n\geq 1}\mathbb{E}\left[\sup_{0\leq t\leq T}|X_n(t)|^p\right] \leq C_8$$

για κάθε $p \leq p_0$

Θεώρημα 2.6

Υποθέτουμε τις (S.A1)-(S.A3) και A4, τότε το tamed σχήμα Euler συγκλίνει στην ισχυρή λύση της ΣΔΕ στον $\mathcal{L}^p$, τουτέστιν

$$\lim_{n\to\infty}\mathbb{E}\left[\sup_{0\geq t\geq T}|X(t) - X_n(t)|^p\right]$$

για κάθε $p \leq p_0$

Πρόταση 2.7

Υποθέτουμε τις (S.A1),(S.A3),(S.A5) και A4, τότε το tamed σχήμα Euler συγκλίνει για $a = 1/2$ στην ισχυρή λύση της ΣΔΕ στον $\mathcal{L}^p$ με ρυθμό σύγκλισης $1/2$, δηλαδή

$$\mathbb{E}\left[\sup_{0\geq t\geq T}|X(t) - X_n(t)|^p\right] \leq Cn^{-p/2}$$

για κάθε $p \leq p_0$, όπου C είναι σταθερά ανεξάρτητη του n.

Tamed Un-adjusted Langevin Algorithms

Στο παρών κεφάλαιο επιστρέφουμε στην δομή που καθιερώσαμε στην Εισαγωγή και θεωρούμε την περίπτωση όπου ο όρος τάσης ∇U γράφεται ως το άθροισμα δύο όρων. Έτσι όλες οι υπεργραμμικότητες θα αντιμετωπιστούν ταυτόχρονα ελέγχοντας τον πρώτο όρο ενώ ο δεύτερος θα ικανοποιεί την συνθήκη Lipschitz συνέχειας και θα αυξάνεται το πολύ γραμμικά. Προχωράμε υλοποιώντας μια ασθενέστερη μορφή taming απο αυτήν που αναπτύχθηκε στο [10], χωρίς να θυσιάζουμε την απόδοση του αλγορίθμου. Απο εδώ και στο εξής θεωρούμε πως $U = H + F$ είναι συνεχώς παραγωγίσιμη. Προτείνουμε δύο νέες συναρτήσεις μερικού taming $T_h(x)$:

$$G_h(x) = \frac{\nabla H(x)}{1 + h\|\nabla H(x)\|} + \nabla F(x), \text{ και } G_{h,c} = \left(\frac{\partial_i H(x)}{1 + h|\partial_i H(x)|} + \partial_i F(x) \right)_{i \in \{1,\dots,d\}}$$

οι οποίες μαζί με την διακριτοποίηση της Langevin ΣΔΕ αποτελούν αντίστοιχα τους αλγορίθμους Partially Tamed Un-adjusted Algorithm (PTYLA) και (PTULAc).
Θέτουμε τώρα τις παρακάτω υποθέσεις.

H1. Υπάρχουν σταθερές $\ell, L, K \in \mathbb{R}_+$ τέτοιες ώστε για κάθε $x, y \in \mathbb{R}^d$,
(ι) $\|\nabla H(x) - \nabla H(y)\| \leq L \left(1 + \|x\|^\ell + \|y\|^\ell\right) \|x - y\|$
(ιι) $\|\nabla F(x) - \nabla F(y)\| \leq L_f \|x - y\|$
H2.
(ι) $\liminf_{\|x\| \to +\infty} \frac{\|\nabla H(x)\|}{\|x\|} = +\infty$
(ιι) $\liminf_{\|x\| \to +\infty} \left\langle \frac{x}{\|x\|}, \frac{\nabla H(x)}{\|\nabla H(x)\|} \right\rangle > 0$

Κάτω απο αυτές τις συνθήκες αποκτούμε ορισμένες σημαντικές παρατηρήσεις. Βλέπουμε πως η H2 συνεπάγεται $\liminf_{\|x\| \to +\infty} \nabla H(x) = +\infty$, επομένως η H επιδέχεται ελάχιστο x^* τέτοιο ώστε $\nabla H(x^*) = 0$, δίχως βλάβη της γενικότητας μπορούμε να θεωρήσουμε πως $x^* = 0$. Τότε αντικαθιστώντας για $y = x^*$ στην H1(ι) λαμβάνουμε πως για κάθε $x \in \mathbb{R}^d$,

$$\|\nabla H(x)\| \leq 2L(1 + \|x\|^{\ell+1})$$

Επίσης η H1(ιι) δίνει πως για κάθε $x \in \mathbb{R}^d$,

$$\|\nabla F(x)\| \leq \max(L_f, \nabla F(0))(1 + \|x\|) := K(1 + \|x\|)$$

Παρατηρούμε πως υπο την H2(ι) συνεπάγεται πως υπάρχει $M, C \in \mathbb{R}_+^*$ τέτοιο ώστε για κάθε $x \in \mathbb{R}^d,\ \|x\| \geq M$

$$\|\nabla H(x)\| \geq C\|x\|$$

Αντίστοιχα η H2(ιι) συνεπάγεται πως υπάρχει $M, k \in \mathbb{R}_+^*$ τέτοιο ώστε για κάθε $x \in \mathbb{R}^d,\ \|x\| \geq M$

$$x\nabla H(x) \geq k\|x\|\|\nabla H(x)\|$$

όπου το M και στις δύο περιπτώσεις μπορεί να επιλεχθεί αυθαίρετα μεγάλο. Είναι προφανές πως ισχύουν τοπικές συνθήκες Lipschitz και για τους δύο συντελεστες της Langevin ΣΔΕ, άρα απο την

κλασσική βιβλιογραφεία υπάρχει μοναδική ισχυρή λύση $(Y_t)_{t\geq 0}$. Επιπρόσθετα το ισχυρά μαρκοβιανό semigroup $(P_t)_{t\geq 0}$ (βλέπε Θεώρημα 5.4.20 [12]) που κατασκευάζεται ως εξής:

$$P_t(x,A) = P\left(Y_t(x)\in A\right) = \mathbb{E}\left[\mathbb{I}_A(Y_t)|Y_0=x\right] \quad \forall t\geq 0,\ x\in\mathbb{R}^d \text{ και } A\in\mathcal{B}(\mathbb{R}^d)$$

είναι αντιστρέψιμο ως προς την κατανομή π που αντιστοιχεί στην αναλλοίωτη κατανομή του. Επίσης λαμβάνουμε και άλλες σημαντικές ιδιότητες όπως το ότι είναι θετικά επαναληπτική και εκθετικά γεωμετρική, σύμφωνα με τους S.P.Meyn και R.L.Tweedie [13],[14],[6], καταλήγοντας έτσι να ελέγχουμε τις ροπές της διάχυσης μέσω ενός κριτηρίου τύπου Lyapunov-Foster. Ο απειροστικός γεννήτορας $\mathcal{A}$ που συσχετίζεται με τα Langevin Dynamics ορίζεται ως:

$$\mathcal{A}f = \lim_{t\to 0}\frac{1}{t}\left(\mathbb{E}_x\left[f(Y_t)\right]-f(x)\right)$$

για κάθε δοκιμαστική συνάρτηση $f\in C^2(\mathbb{R}^d)$ και $x\in\mathbb{R}^d$. Επειδή η ΣΔΕ είναι της μορφής $dY_t = b(Y_t)dt+\sigma\left(Y_t\right)dB_t$ με $a:=\sigma^\tau\sigma$ υπολογίζουμε επιπλέον πως:

$$\begin{aligned}\mathcal{A}f &= \sum_{i=0}^{d} b_i\frac{\partial f}{\partial x_i}+\frac{1}{2}\sum_{i=0}^{d}\sum_{k=0}^{d}a_{ik}\frac{\partial^2 f}{\partial x_i\partial x_k} = -\sum_{i=0}^{d}\frac{\partial U}{\partial x_i}\frac{\partial f}{\partial x_i}+\frac{1}{2}\sum_{k=0}^{d}a_{kk}\frac{\partial^2 f}{\partial {x_k}^2}+\frac{1}{2}\sum_{i\neq k}a_{ik}\cancelto{0}{\frac{\partial^2 f}{\partial x_i\partial x_k}}\\ &= -\left\langle\nabla U(x)|\nabla f(x)\right\rangle+\Delta f(x)\end{aligned}$$

Ακολουθώντας την βιβλιογραφεία θα χρειαστούμε μια συνάρτηση τύπου-νόρμας που θα είναι πάντα μεγαλύτερη ή ίση της μονάδας, ορίζουμε λοιπόν την συνάρτηση Lyapunov $V_a:\mathbb{R}^d\to[1,\infty)$ για κάθε $x\in\mathbb{R}^d$ για κάθε $a\in\mathbb{R}_+^*$ ως :

$$V_a(x) = \exp\left(a\left(1+\|x\|^2\right)^{1/2}\right)$$

Τότε έχουμε για κάθε $x\in\mathbb{R}^d$

$$\begin{aligned}\mathcal{A}V_a(x) &= -\nabla U(x)\nabla V_a(x)+\Delta V_a(x)\\ &= -\sum_{i=0}^{d}\frac{\partial U(x)}{\partial x_i}\frac{x_i a V_a(x)}{\left(1+\|x\|^2\right)^{1/2}}+\sum_{i=0}^{d}\frac{aV_a(x)}{\left(1+\|x\|^2\right)^{1/2}}+\frac{x_i^2a^2V_a(x)}{1+\|x\|^2}-\frac{x_i^2aV_a(x)}{\left(1+\|x\|^2\right)^{3/2}}\\ &= -\nabla U(x)\frac{axV_a(x)}{\left(1+\|x\|^2\right)^{1/2}}+\frac{adV_a(x)}{\left(1+\|x\|^2\right)^{1/2}}+\frac{a^2\|x\|^2V_a(x)}{1+\|x\|^2}-\frac{a\|x\|^2V_a(x)}{\left(1+\|x\|^2\right)^{3/2}}\end{aligned}$$

Τώρα είμαστε έτοιμοι να αποδείξουμε το πρώτο μας θεωρητικό αποτέλεσμα

Πρόταση 3.1

Υποθέτουμε H1,H2. Έστω $a\in\mathbb{R}_+^*$. Τότε υπάρχει $b_a\in\mathbb{R}_+$ τέτοιο ώστε για κάθε $x\in\mathbb{R}^d$

$$\mathcal{A}V_a(x)\leq -aV_a(x)+ab_a$$

και

$$\sup_{t\geq 0}P_tV_a(x)\leq V_a(x)+b_a$$

Αρχικά το Θεώρημα 2.1[13] μας εξασφαλίζει πως η διαδικασία $(Y_t)_{t\geq 0}$ δεν εκρήγνυται, υπο την έννοια

πως $\mathbb{P}_x\{Y_t \to +\infty\} = 0$ για κάθε $x \in \mathbb{R}^d$. Επίσης λόγω του Θεωρήματος 4.2[13] η $(Y_t)_{t\geq 0}$ είναι επαναληπτική και το $\pi(V_a)$ πεπερασμένο. Τότε η διαδικασία διάχυσης είναι θετικά επαναληπτική και το αναλλοίωτο μέτρο π μπορεί να κανονικοποιηθεί σε μέτρο πιθανότητας. Με διαφορά το σημαντικότερο αποτελέσμα έγκειται απο το Θεώρημα 6.1[13] όπου αποκτούμε την σύγκλιση του semigroup $(P_t)_{t\geq 0}$ στην αναλλοίωτη κατανομή π υπο την V_a-νόρμα (εργοδικότητα) και μάλιστα συγκλίνει με εκθετικό ρυθμό, τουτέστιν:

$$\|P_t(x,\cdot) - \pi\|_{V_a} \leq C_a \rho_a^t V_a, \quad t \in \mathbb{R}_+, \quad x \in \mathbb{R}^d$$

με $C_a \in \mathbb{R}_+$, $\rho_a \in [0,1)$ και την V_a-νόρμα ορισμένη ως $\|f_1 - f_2\|_{V_a} := \sup_{|g|\leq V_a} |f_1 g - f_2 g|$. Αν και τετριμμένο είναι σημαντικό να σημειωθεί πως η παραπάνω σχέση είναι αυτή που δικαιολογεί την χρήση εξισώσεων Langevin για δειγματοληψία και αλγόριθμους βελτιστοποιήσης καθώς ο όρος τάσης είναι κατασκευασμένος έτσι ώστε να εξασφαλίζεται η σύγκλιση στην κατανομή ενδιαφέροντος π. Παρόλο που η καλή συμπεριφορά της $(Y_t)_{t\geq 0}$ έχει καθιερωθεί, μία απλοϊκή διακριτοποίηση μπορεί να αποτυγχάνει στην σύγκλιση [6], άρα κρίνεται απαραίτητο να αναπτύξουμε αντίστοιχα αποτελέσματα και για την διακριτή έκδοση $(X_n)_{n\in\mathbb{N}}$. Συγκεκριμένα στο Λήμμα 3.2 σημειώνουμε ιδιότητες των συναρτήσεων $G_h(x)$ που εξασφαλίζουν την γεωμετρική εργοδικότητα των TULA, TULAc ως προς την αναλλοίωτη κατανομή π_h. Επίσης οφείλουμε να τονήσουμε για λόγους πληρότητας πως η κατανομή π_h εν γένη δεν ταυτίζεται με την π, εκτός περιορισμένων εξαιρέσεων. Αυτό το φαινόμενο ωστόσο δεν επιφέρει σοβαρές συνέπειες στην αποτελασματικότητα των αλγορίθμων πέρα απο την εισαγωγή ενός ασυμπτωτικού σφάλματος που μπορεί να ελεγχθεί όσο λαμβάνουμε θεωρητικά αποτελέσματα που φράζουν την απόσταση μεταξύ των δύο αυτών κατανομών, είτε υπό την έννοια αποστάσεων Wasserstein είτε της ολικής κύμανσης. Έστω τώρα μια πολυδίασταση κανονική τ.μ. με μέση τιμή 0 και πίνακα συνδιακύμνασης I_d, τότε το δυναμικό που σχετίζεται με αυτήν την κατανομή είναι: $U(x) = \|x\|^2/2$. Για τον αλγόριθμο ULA λοιπόν και επιλογή βήματος $h = 1$ προκύπτει:

$$X_{n+1} \sim (X_n - \nabla U(X_n), 2I_d) \sim (0, 2I_d) \nsim (0, I_d)$$

Πρέπει να είναι ξεκάθαρο πλέον πως το διακριτό βήμα h αλλά και η επιλογή συνάρτησης taming $T_h(x)$ επηρεάζουν την μορφή της π_h, που δεν θα πρέπει να συγχέεται με την π.

Λήμμα 3.2

Υποθέτουμε τις H1 και H2. Έστω $h > 0$ και πως η T_h ισούται είτε με G_h ή με $G_{h,c}$, τότε τα παρακάτω είναι αληθή:

P1 Τότε υπάρχουν $a \geq 0,\ C_a < +\infty$ τέτοια ώστε για κάθε $h > 0$ και $x \in \mathbb{R}^d$,

$$\|T_h(x) - \nabla U(x)\| \leq h C_a (1 + \|x\|^a)$$

P2 Για κάθε $h > 0$,

$$\liminf_{\|x\|\to+\infty} \left\langle \frac{x}{\|x\|}, T_h(x) \right\rangle - \frac{h}{2\|x\|}\|T_h(x)\|^2 - \left\langle \frac{x}{\|x\|}, \nabla F(x) \right\rangle + \frac{h}{2\|x\|}\|\nabla F(x)\|^2 > 0$$

Πρόταση 3.3

Υποθέτουμε H1,H2 και έστω $h > 0$. Τότε υπάρχουν $M, æ, b \in \mathbb{R}_+^*$ που ικανοποιούν τα παρακάτω για κάθε $x \in \mathbb{R}^d$

$$R_h V_æ(x) \leq e^{-æ^2 h} V_æ(x) + hb\mathbb{I}_{\bar{B}(0,M)}(x)$$

Η παραπάνω συνθήκη σύμφωνα με το Γεωμετρικό Εργοδικό Θεωρήμα (15.0.1) στο [14], επαρκεί ώστε ο R_h να παράγει ένα μοναδικό αναλλοίωτο μέτρο πιθανότητας π_h και ταυτόχρονα να είναι $V_æ-$ γεωμετρικά εργοδικός ως προς αυτό. Αντίστοιχα μια σχέση τύπου Lyapunov-Foster μπορεί να καθιερωθεί για τον R_h^n μέσω της μαθηματικής επαγωγής και της βασικής ανισότητας $1 - e^{-s} \geq se^{-s}$,

$$R_h^n V_æ(x) \leq e^{-æ^2 nh} V_æ(x) + (b/æ^2)e^{æ^2 h}$$

Μέχρι στιγμής έχουμε δείξει πως υποθέτοντας για την κλίση του δυναμικού U ότι αποτελείται απο έναν υπεργραμμικό και Lipschitz συνεχή όρο, καταλήγουμε στην ίδια κλάση δυναμικών που έχει περιγραφτεί στο [10] υπό την έννοια οτι τα αποτελέσματα των Προτάσεων 3.1 και 3.3 δεν διαφέρουν απο τα αναλόγα τους στην άνω βιβλιογραφική αναφορά, μέχρι και κάποιες σταθερές. Επομένως λαμβάνουμε αυτομάτως και τα επιμέρους Θεωρήματα:

Θεώρημα 3.4

Υποθέτουμε H1,H2. Έστω $h_0 > 0$. Τότε υπάρχει $C > 0$ και $\lambda \in (0,1)$ τέτοια ώστε για κάθε $h \in (0, h_0]$, $x \in \mathbb{R}^d$ και $n \in \mathbb{N}$

$$\|\delta_x R_h^n - \pi\|_{V_æ^{1/2}} \leq C(nh\lambda^{nh} V_æ(x) + \sqrt{h})$$

και για κάθε $h \in (0, h_0]$,

$$\|\pi_h - \pi\|_{V_æ^{1/2}} \leq C\sqrt{h}$$

Θεώρημα 3.5

Υποθέτουμε H1,H2 και ισχυρή κυρτότητα στην $\nabla H(x)$.Έστω $h_0 > 0$. Τότε υπάρχει $C > 0$ και $\lambda \in (0,1)$ τέτοιο ώστε για κάθε $h \in (0, h_0]$, $x \in \mathbb{R}^d$ και $n \in \mathbb{N}$

$$W_2^2(\delta_x R_h^n, \pi) \leq C(nh\lambda^{nh} V_æ(x) + \sqrt{h})$$

και για κάθε $h \in (0, h_0]$,

$$W_2^2(\pi_h, \pi) \leq C\sqrt{h}$$

Τα φράγματα του Θεωρήματος 3.5 μπορούν περαιτέρω να βελτιωθούν, άμα απαιτήσει κανείς το δυναμικό να είναι δυο φορές συνεχόμενα διαφορίσιμο και η Λαπλασιανή του να είναι τοπικά Hölder συνεχής.

H3. $U \in C^2(\mathbb{R}^d, \mathbb{R})$ Τότε υπάρχουν $v, L_H \in \mathbb{R}_+$ και $\beta \in [0,1]$τέτοιο ώστε για κάθε $x, y \in \mathbb{R}^d$,

$$\|\nabla^2 H(x) - \nabla^2 H(y)\| \leq L_H \left(1 + \|x\|^v + \|y\|^v\right) \|x - y\|^\beta$$

Θεώρημα 3.6

Υποθέτουμε H1,H2,H3 και ισχυρή κυρτότητα στην $\nabla H(x)$. Έστω $h_0 > 0$. Τότε υπάρχει $C > 0$ και $\lambda \in (0,1)$ τέτοιο ώστε για κάθε $h \in (0, h_0]$, $x \in \mathbb{R}^d$ και $n \in \mathbb{N}$

$$W_2^2(\delta_x R_h^n, \pi) \leq C(nh^{1+\beta}\lambda^{nh} V_{\text{œ}}(x) + h^{1+\beta})$$

και για κάθε $h \in (0, h_0]$,

$$W_2^2(\pi_h, \pi) \leq Ch^{1+\beta}$$

Στην συνέχεια για να μπορέσουμε να αξιοποιήσουμε τις αριθμητικές προσομοιώσεις μας προς σύγκριση των αλγορίθμων, θα χρειαστεί να εκτιμήσουμε την πρώτη και δεύτερη ροπή των μονοπατιών απο τις εμπειρικές μέσες τιμές τους. Σε αυτό το πλαίσιο συμπεριλαμβάνουμε την εξής υπόθεση:

H4. $H \in C^4(\mathbb{R}^d, \mathbb{R})$ και $\|D^i H\| \in C_{\pi o \lambda \psi}(\mathbb{R}^d, \mathbb{R}_+)$ για $i \in \{1, \ldots, 4\}$

Θεώρημα 3.7

Υποθέτουμε H1,H2,H4. Έστω $f \in C^3(\mathbb{R}^d,)$ τέτοιο ώστε $\|D^i f\| \in C_{\pi o \lambda \psi}(\mathbb{R}^d, \mathbb{R}_+)$ για $i \in \{0, \ldots, 3\}$. Έστω επίσης $h_0 > 0$ και $(X_k)_{k \in \mathbb{N}}$ η μαρκοβιανή αλυσίδα όπως ορίζεται απο τους τους tamed αλγορίθμους με αρχικοποίηση στο σημείο $X_0 = 0$. Τότε υπάρχει $C > 0$ τέτοιο ώστε για κάθε $h \in (0, h_0]$ και $n \in \mathbb{N}^*$,

$$\left| \mathbb{E}\left[\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \pi(f) \right]) \right| \leq C(h + \frac{1}{nh})$$

ανδ

$$\mathbb{E}\left[\left(\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \pi(f) \right)^2 \right]) \leq C(h^2 + \frac{1}{nh})$$

Αριθμητικά αποτελέσματα

Ως παράδειγμα για τις αριθμητικές προσομοιώσεις διαλέγουμε το δυναμικό διπλού πηγαδιού $U(x) = (1/4)\|x\|^4 - (1/2)\|x\|^2$ με $\nabla U(x) = \|x\|^2 x - x$ το οποίο λόγω της μορφής του ταιρίαζει απόλυτα στους αλγορίθμους με μερικό taming. Δείχνουμε εν συντομία πως ικανοποιούνται οι υπόθεσεις μας, συγκεκριμένα οι H1,H2 και H4:

$$\|\nabla H(x) - \nabla H(y)\| = \left\| \|x\|^2 x - \|y\|^2 y \right\| \leq (\|x\| + \|y\|)^2 \|x - y\| \leq 2(1 + \|x\|^2 + \|y\|^2)\|x - y\|$$

$$\left\langle \frac{x}{\|x\|}, \frac{\nabla H(x)}{\|\nabla H(x)\|} \right\rangle = \left\langle \frac{x}{\|x\|}, \frac{\|x\|^2 x}{\|x\|^3} \right\rangle = 1 > 0$$

Απο τις Προτάσεις 3.1 και 3.3, η διαδικασία διάχυσης αλλά και η διακριτή της έκδοση που παράγεται απο το μερικώς tamed σχήμα είναι εργοδικές και απο το Θεώρημα 3.4 οι αναλλοίωτες κατανομές τους είναι ικανοποιητικά κοντά υπο την έννοια της ολικής κύμανσης, επομένως εξασφαλίζεται πως θα έχουν κοινή συμπεριφορά για αρκέτα μεγάλα δείγματα. Ταυτόχρονα το Θεώρημα 3.7 μας επιτρέπει να εκτιμήσουμε την πρώτη και δεύτερη ροπή της αναλλοίωτης κατανομής μέσω των εμπειρικών τους μέσων. Σκοπός μας είναι να συγκρίνουμε τους αλγορίθμους PTULA και PTULAc σε σχέση με τον ULA και τους

υπόλοιπους αλγορίθμους που αναπτύχθηκαν στο [10]. Τα αποτελέσματα και τα χαρακτηριστικά των προσομοιώσεων βρίσκονται στο κύριο μέρος της διπλωματικής. Τα κύρια συμπεράσματα μας έχουν ως εξής, όπως αναμέναμε ο ULA σε πολλές περιπτώσεις είναι ασταθής και τα μονοπάτια εκρήγνυνται, ειδικά και μεγάλα διακριτά βήματα h ενώ ο TULA φαίνεται να εισάγει συστηματικά ένα σφάλμα μεροληψίας. Για τον αλγόριθμο PTULA παρατηρούμε πως έχει αντίστοιχη συμπεριφορά και ακρίβεια με τον TULAc. Φαίνεται δηλαδή πως η εφαρμογή του taming είτε κατα συνιστώσα είτε μερικώς επαρκεί για να αφαιρέσει την παθογένεια που συναντάται στον TULA, ωστόσο όταν αυτές οι δύο τεχνικές αξιοποιούνται ταυτόχρονα στον PTULAc δεν έχουμε περαιτέρω βελτίωση.

Introduction

The rapid adoption of machine learning and artificial intelligence from the vast majority of scientific fields (engineering, natural sciences, medicine) but also from the industry (social networking, advertisement, finance) has been a driving force for the development of an efficient optimization algorithm. Most problems in these disciplines end up requiring the global minimization of an objective function. While this mathematical problem is well studied in classical literature, irregular features met in real world tasks such as noisy data, high-dimensional spaces and ill-conditioned objective functions can't be handled trivially. Most notably, objective functions encountered in deep learning applications fail to be convex, a strong property which guarantees global convergence for classical algorithms. The most abstract framework one can get is to seek for $x^* \in \mathbb{R}^d$ such that

$$x^* = \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} U(x)$$

where $U : \mathbb{R}^d \mapsto \mathbb{R}$ is the quantity we want to minimize. If U is at least continuously differentiable, we get the following naive Gradient Decent algorithm by the discretization of $dx(t) = -\nabla U(x(t))dt$:

$$x_{n+1} = x_n - h\nabla U(x_n)$$

One school of thought focuses on augmenting the above deterministic algorithm with random explorations techniques on the surface of the objective function, by modifying the classical recursive scheme with stochastic perturbations to achieve convergence. By doing so, the algorithm is prevented to proceed along the steepest slope and thus it might avoid being trapped in local minima or saddle points. Still, the convergence to the true minimum is not guaranteed and its common for the algorithm to converge to different local minima depending on the initialization. A simple stochastic modification is to add white noise next to the gradient, this results to the formulation of the Stochastic Differential Equation

$$dX(t) = -\nabla U(X(t))dt + \sigma(t, X(t))dW_t \ , \quad \forall t \in [0, T] \tag{1}$$

with initial value $X(0) = x_0$ a.s. finite $\mathcal{F}_0-$measurable, where $\sigma(t, x) : \mathbb{R} \times \mathbb{R}^d \mapsto \mathbb{R}^{d \times m}$ is a $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B}(\mathbb{R}^d)$-measurable function called diffusion coefficient and $(W(t))_{t \geq 0}$ stands for a m-dimensional Wiener Process. Then its Euler-Maruyama discretization is defined as

$$X(t_{n+1}) = X(t_n) - h\nabla U(X(t_n)) + \sigma(t_n, X(t_n))\Delta W_n$$

where $\Delta W_n = W_{t_{n+1}} - W_{t_n}$ are i.i.d normal m-dimensional r.v. with expected value zero and variance h, hence gaining the alternative notation

$$X(t_{n+1}) = X(t_n) - h\nabla U(X(t_n)) + \sqrt{h}\sigma(t_n, X(t_n))Z_{n+1} \tag{2}$$

with $(Z_n)_{n \geq 1}$ being i.i.d standard m-dimensional Gaussian r.v.. This iterative scheme can be further improved by adopting adaptive stepsize and momentum techniques, giving birth to several optimization algorithms which solve many of SGD's pathologies, the most popular among them being ADAM [1]. Although those algorithms are empirically known to be efficient for training neural networks and optimization of non-convex functions there are still significant mathematical gaps to be filled [2].

On another stream of literature the focus is shifted on the development of MCMC-type algorithms [3] on the premise of constructing an ergodic Markov Chain whose invariant distribution allows for

sampling from a target probability measure. Having access to draw such samples, in combination with the simulated annealing algorithm, constitute a computational efficient program for tackling optimization problems, especially if the latter involve computations in high dimensional spaces. More precisely, for any objective function $U(x)$ we can always consider the following probability distribution w.r.t. Lebesgue measure

$$\pi_\beta(x) = e^{-\beta U(x)} / \int_{\mathbb{R}^d} e^{-\beta U(y)} dy \tag{3}$$

where $\beta > 0$ is a parameter refereed to as the inverse temperature. Its apparent that if we have the ability to simulate values from π_β we can simply detect the argmin$U(x)$ by approximating the whole distribution , although that's not computational efficient but we will come back to it later. Under certain regularity conditions on $U(x)$ its known from diffusion theory that the density on (3) is associated with the overdamped Langevin equation

$$dY_t = -\nabla U(Y_t)dt + \sqrt{2\beta^{-1}}dW_t \tag{4}$$

in the sense that the solution $(Y_t)_{t>0}$ of (4) is geometrically ergodic with an invariant probability measure μ that possess the density π_β w.r.t. Lebesgue measure. Invoking once again the Euler-Maruyama method, we derive the following discretization scheme for the Langevin SDE (2) also known as the Unadjusted Langevin Algorithm (ULA)

$$X(t_{n+1}) = X(t_n) - h\nabla U(X(t_n)) + \sqrt{2h\beta^{-1}}Z_{n+1} \tag{5}$$

The discrete process $(X_n)_{n\in\mathbb{N}}$ defined by (5) is a Markov Chain, whose stationary distribution can be used to approximate π_β. Although those distributions are not necessary the same, under the assumption of a globally Lipschitz gradient $\nabla U(x)$, non-asymptotic bounds in total variation and Wasserstein distances between π_β and the distribution of $(X_n)_{n\in\mathbb{N}}$ can be established. Yet, this is a strong condition not regularly met and its has been shown that when violated ULA itself can be unstable. In particular if the drift coefficient of (5), $\nabla U(x)$ is superlinear i.e. $\liminf_{\|x\|\to\infty}\|\nabla U(x)\|/\|x\| = +\infty$, its known that moments of $(X_n)_{n\in\mathbb{N}}$ can diverge to infinity [4]. A way to counteract this behavior, is to introduce an additional Metropolis-Hastings step to ULA [5]. Indeed if we look at the iterates of (5) as proposed moves in a random walk with an acceptance probability of

$$a_h(x,y) = 1 \wedge \frac{q_{h,\beta}(y,x)\pi_\beta(y)}{q_{h,\beta}(x,y)\pi_\beta(x)}$$

where $q_{h,\beta}(y,x)$ is the transition probability density. Adding such an acceptance-rejection step one gets the Metropolis-Adjusted Langevin Algorithm (MALA)

$$X^M(t_{n+1}) = \begin{cases} X(t_{n+1}), \text{ if } u_n < a_h(X^M(t_n), X(t_{n+1})) \\ X^M(t_n), \text{ otherwise} \end{cases}$$

where $(u_n)_{n\in\mathbb{N}} \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0,1)$. By construction MALA ensures reversibility w.r.t. π_β thus preserving the invariant measure. Although an improvement from ULA, in the superlinear case MALA can still fail to be geometrically ergodic[6], a desirable property for a MCMC algorithm as it guarantees central limit-type theorems to hold. This happens because the proposed Markov Chain generated from ULA is often transient, thus leading MALA to lose it's geometrical ergodicity. As a potential solution, MALTA was proposed which modifies MALA by truncating the drift coefficient in regions where the underlying Euler Scheme is explosive[6][7]. This is achieved by preserving the direction

of the drift in ULA while the amplitude gets normalized. In contrary to MALA, MALTA can be shown to be geometrically ergodic but the relation to the original diffusion is not the same. Fortunately, in recent literature on Euler approximations a new class of numerical schemes have been introduced to study the case of non-globally Lipschitz conditions by modifying the drift term in such a way that it become uniformly bounded[8][9]. That's the so called notion of taming. The efficiency of such schemes and their respective desirable properties of strong $\mathcal{L}^p$ convergence create a strong incentive to extended these techniques for the sampling problem. Let us consider the following discretization scheme:

$$X(t_{n+1}) = X(t_n) - hT_h(X(t_n)) + \sqrt{2h\beta^{-1}}Z_{n+1} \tag{6}$$

for an appropriate choice of taming function $T_h : \mathbb{R}^d \mapsto \mathbb{R}^d$, then the iteration rule (6) is known to produce a MC whose stationary distribution converges to the target π_β and most importantly to recover the geometrical ergodicity of (4) for fast rates of convergences. Based on [10], there are two different T_h candidates to select from:

$$G_h(x) = \frac{\nabla U(x)}{1 + h\|\nabla U(x)\|} \text{ and } G_h^c(x) = \left(\frac{\partial_i U(x)}{1 + h|\partial_i U(x)|}\right)_{i\in 1,\dots,d}$$

The first one performs the taming to the whole gradient, resulting to the Tamed Unadjusted Langevin Algorithm (TULA). The second one accompanied with (4), is refereed to as the coordinate-wise Tamed Unadjusted Langevin Algorithm (TULAc), which applies the taming element wise thus scaling the effective stepsize of each coordinate individually. However its experimentally shown in [10] that TULAc outperforms TULA in terms of 1st and 2nd moments errors, especially for large stepsize choices and also, its preferred in the design of algorithms who are meant to solve high dimensional optimization problems [11]. Its further theorized and numerically illustrated that depending on the particular problem, more taming functions can be proposed e.g. uniformly bounding only the non-linearities of $\nabla U(x)$. This will be the driving idea behind our work in Chapter 2.

So far we have seen how MCMC algorithms help us sample from a prescribed distribution of our needs, but our initial goal was to optimize an objection function. The gap between sampling and optimization in this case is bridged via simulating (or temperature) annealing. While simulating unless the distribution has large probability mass around the maximum, computing resources will be wasted exploring areas of no particular interest. By adopting simulated annealing with any of the presented algorithms, we will sample using a Markov chain whose invariant distribution at iteration n is no longer $\pi_\beta(x)$ but rather equal to $\pi_{\beta_n}(x)$ where $(1/\beta_n)_{n\in\mathbb{N}}$ is a decreasing cooling schedule with $\lim_{n\to\infty} 1/\beta_n = \infty$. The reason behind doing this is that $\lim_{n\to\infty}\pi_{\beta_n}(x)$ is a probability density that concentrate itself on the set of global maxima of $\pi(x)$, hence the minima of $\nabla U(x)$. Optimal temperature scheduling and initial β_0 are specific problem dependent, but in general the scheduling should be slow e.g. logarithmic. All in all, as we progress through the burn-in period of a MCMC algorithm, we tend to sample exclusively near the distribution's peaks

Tamed Euler Approximation Schemes

Even from classical literature it known that implicit methods often produce stable numerical schemes in comparison to their explicit counterparts but with the impactful repercussion of extra computational cost. In the stochastic framework a vanilla explicit scheme applied to a not so favorable conditioned SDE can even produce exploding paths [4], a very problematic behavior, that should be avoided. Recently a new family of explicit (tamed) schemes have been developed [9][8], that simultaneity enjoy stability and computational efficiency. In [9] its shown that under linear growth and local Lipschitz conditions on the coefficients of a diffusion based SDE like (2), the tamed implicit Euler-Maruyama scheme is $\mathcal{L}^p$ convergent to the true solution of the SDE and the classical rate of convergence is recovered when globally one-sided Lipschitz condition is demanded. Key to such results is to achieve uniform moments bounds. We aim to extend this work by considering the case in which the drift coefficient consists of two terms, one satisfying the assumptions used in [9] and the other being globally Lipschitz continuous, allowing the latter to be untamed in the explicit scheme. Also we assume that only a finite number of moments of the initial value are bounded, as that's the case in most applications. This saddle but meaningful modification requires a different technical approach for obtaining the moments bounds.

Setup

Let $(\Omega, \{\mathcal{F}_t\}_{t\geq 0}, \mathcal{F}, \mathbb{P})$ be a filtered probability space satisfying the usual conditions and consider the SDE

$$dX(t) = g(t, X(t))dt + \sigma(t, X(t))dW_t, \qquad \forall t \in [0, T] \tag{7}$$

with initial value $X(0) = x_0$ a.s. finite $\mathcal{F}_0-$measurable, where $\sigma(t, x) : \mathbb{R} \times \mathbb{R}^d \mapsto \mathbb{R}^{d\times m}$ is a $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B}(\mathbb{R}^d)$-measurable function and $(W(t))_{t\geq 0}$ stands for a m-dimensional Wiener Process. Additionally the drift coefficient can be written as

$$g(t, x) = b(t, x) + f(x), \qquad \forall (t, x) \in [0, T] \times \mathbb{R}^d$$

where it is assumed that $b(t, x) : \mathbb{R} \times \mathbb{R}^d \mapsto \mathbb{R}$ is $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B}(\mathbb{R}^d)$-measurable and $f(\cdot)$ is globally Lipschitz continuous and thus a $\mathcal{B}(\mathbb{R}^d)$-measurable function as well. Then there exists a positive constant L_f such that,

$$|f(x) - f(y)| \leq L_f|x - y| \tag{8}$$

We now define the following numerical scheme

$$dX_n(t) = g_n(t, X_n(k_n(t)))dt + \sigma(t, X_n(k_n(t)))dW_t, \quad k_n(t) = [nt]/n, \quad \forall t \in [0, T] \text{ and } \forall n \geq 1 \tag{9}$$

where the drift coefficient is partially tamed, in the sense that $g_n(t, x) = b_n(t, x) + f(x)$, while it is assumed that

$$b_n(t, x) := \frac{1}{1 + n^{-a}|b(t, x)|} b(t, x)$$

for any $t \in [0, T], x \in \mathbb{R}^d$ and $a \in (0, 1/2]$. Furthermore we notice that $|b_n(t, x)| \leq \min(n^a, |b(t, x)|)$

Assumptions

Henceforth every call with a 'S' prefix shall be considered a direct reference to [9]. We begin with by showing that if the assumptions S.A1, S.A2, S.A3, S.A5 hold true then we easily obtain identical properties for the function $g(t,x)$.

S.A1 There exists a positive constant K such that,

$$2xb(t,x) \vee |\sigma(t,x)|^2 \leq K(1+|x|^2)$$

for any $t \in [0,T]$ and $x \in \mathbb{R}^d$.

S.A2 For every $R > 0$, there exists a positive constant L_R such that, for any $t \in [0,T]$,

$$2(x-y)(b(t,x)-b(t,y)) \vee |\sigma(t,x)-\sigma(t,y)|^2 \vee L_R|x-y|^2$$

for all $|x|, |y| \leq R$

S.A3 For every $R \geq 0$, and $p > 0$, there exists $N_R \in \mathbb{L}^p$ such that,

$$\sup_{|x|\leq R} |b(t,x)| \leq N_R(t)$$

for any $t \in [0,T]$

S.A5 There exist positive constants ℓ and L such that, for any $t \in [0,T]$,

$$(x-y)(b(t,x)-b(t,y)) \vee |\sigma(t,x)-\sigma(t,y)|^2 \leq L|x-y|^2$$

and

$$|b(t,x)-b(t,y)| \leq L(1+|x|^\ell+|y|^\ell)|x-y|$$

for all $x, y \in \mathbb{R}^d$

A4 For every $p \leq p_0,\ \mathbb{E}[|X(0)|^p] < \infty$ where $p_0 \geq 2$

Note that under (8), $f(\cdot)$ grows at most linearly:

$$|f(x)| \leq |f(x)-f(0)| + |f(0)| \leq L_f|x-0| + |f(0)| \leq \max(L_f, |f(0)|)(1+|x|) = K_f(1+|x|) \quad (10)$$

We observe that,

$$2xf(x) \leq 2|x||f(x)| \leq 2K_f(|x|+|x|^2) \leq 2K_f(1+|x|)^2 \leq 4K_f(1+|x|^2) \quad (11)$$

Now due to (S.A1) and (11) it is easily seen that,

$$2xg(t,x) = 2xb(t,x) + 2xf(x) \leq (K+4K_f)(1+|x|^2) = C_1(1+|x|^2) \quad (12)$$

Also, under (S.A2) and (8) it trivially follows,

$$\begin{aligned} 2(x-y)(g(t,x)-g(t,y)) &= 2(x-y)(b(t,x)-b(t,y)) + 2(x-y)(f(x)-f(y)) \\ &\leq L_R|x-y|^2 + 2|x-y|L_f|x-y| \leq (L_R+2L_f)|x-y|^2 = C_R|x-y|^2 \end{aligned} \quad (13)$$

Similarly due to (S.A3) and (10) we obtain

$$\sup_{|x|\leq R} |g(t,x)| \leq \sup_{|x|\leq R} |b(t,x)| + \sup_{|x|\leq R} |f(x)| \leq N_R(t) + K_f(1+R) = G_R(t) \in \mathbb{L}^p \quad (14)$$

Lastly (S.A5) and (8) imply that

$$\begin{aligned}|g(t,x)-g(t,y)| &\le |b(t,x)-b(t,y)|+|f(x)-f(y)| \\ &\le L(1+|x|^{\ell}+|y|^{\ell})|x-y|+L_f|x-y| \\ &\le L(1+|x|^{\ell}+|y|^{\ell})|x-y|+L_f(1+|x|^{\ell}+|y|^{\ell})|x-y| \\ &\le (L+L_f)(1+|x|^{\ell}+|y|^{\ell})|x-y| = C_2(1+|x|^{\ell}+|y|^{\ell})|x-y| \end{aligned} \tag{15}$$

Remark 2.1

For every $n \ge 1$, $g_n(t,x)$ has at most linear growth. To see this, one shall consider (S.2.4) and (8):

$$\begin{aligned}|g_n(t,x)| &\le |b_n(t,x)|+|f(x)| \le n^a + K_f(1+|x|) \le n^a(1+|x|)+K_f(1+|x|) \\ &\le (n^a+K_f)(1+|x|) \le (1+\frac{K_f}{n^a})n^a(1+|x|) \le (1+K_f)n^a(1+|x|) \\ &\le C_3 n^a(1+|x|) \end{aligned} \tag{16}$$

Due to (16) and (S.A1), both the norms of $g_n(t,x)$ and $\sigma(t,x)$ have at most linear growth for every $n \ge 1$. This implies the existence and uniqueness of a solution to (9) for every $n \ge 1$ and furthermore for all $p \le p_0$:

$$\sup_{0\le t\le T} \mathbb{E}[|X_n(t)|^p] \le N := N(n,p,T,\mathbb{E}[|X(0)|^p]) \tag{17}$$

under the assumption (A4).

Proof.
Let us define the stopping time $\tau_R = \inf\{0 \le t \le T : |X_n(t)| \ge R\}$. Then by (9) written on its integral form, one gets:

$$\begin{aligned}X_n(t\wedge\tau_R) &= X(0)+\int_0^{t\wedge\tau_R} g_n(s,X_n(k_n(s)))ds + \int_0^{t\wedge\tau_R}\sigma(s,X_n(k_n(s)))dW_s \\ &= X(0)+\int_0^{t} g_n(s,X_n(k_n(s)))1\{\tau_R>s\}ds + \int_0^{t}\sigma(s,X_n(k_n(s)))1\{\tau_R>s\}dW_s \\ &= X(0)+\int_0^{t} g_n(s,X_n(k_n(s)\wedge\tau_R))1\{\tau_R>s\}ds + \int_0^{t}\sigma(s,X_n(k_n(s)\wedge\tau_R))1\{\tau_R>s\}dW_s \end{aligned}$$

Taking the supremum and then the p^{th} power we obtain

$$\begin{aligned}\sup_{0\le u\le t}|X_n(u\wedge\tau_R)|^p &\le \Big(|X(0)| + \sup_{0\le u\le t}\Big|\int_0^u g_n(s,X_n(k_n(s)\wedge\tau_R))1\{\tau_R>s\}ds\Big| \\ &\quad + \sup_{0\le u\le t}\Big|\int_0^u \sigma(s,X_n(k_n(s)\wedge\tau_R))1\{\tau_R>s\}dW_s\Big|\Big)^p \\ &\le 3^{p-1}\Big(|X(0)|^p + \sup_{0\le u\le t}\Big|\int_0^u g_n(s,X_n(k_n(s)\wedge\tau_R))1\{\tau_R>s\}ds\Big|^p \\ &\quad + \sup_{0\le u\le t}\Big|\int_0^u \sigma(s,X_n(k_n(s)\wedge\tau_R))1\{\tau_R>s\}dW_s\Big|^p\Big) \end{aligned} \tag{18}$$

Now by Hölder's inequality and (11):

$$\mathbb{E}\left[\sup_{0\leq u\leq t}\mid\int_0^u g_n(s,X_n(k_n(s)\wedge\tau_R))1\{\tau_R>s\}ds\mid^p\right]\leq\mathbb{E}\left[\left(\sup_{0\leq u\leq t}\int_0^u\mid g_n(s,X_n(k_n(s)\wedge\tau_R))\mid 1\{\tau_R>s\}ds\right)^p\right]$$

$$\leq\mathbb{E}\left[\left(\int_0^t\mid g_n(s,X_n(k_n(s)\wedge\tau_R))\mid 1\{\tau_R>s\}ds\right)^p\right]\leq t^{p-1}\mathbb{E}\left[\int_0^t\mid g_n(s,X_n(k_n(s)\wedge\tau_R))\mid^p 1\{\tau_R>s\}ds\right]$$

$$\leq T^{p-1}C_3^p n^{ap}\mathbb{E}\left[\int_0^t(1+\mid X_n(k_n(s)\wedge\tau_R)\mid)^p ds\right] \tag{19}$$

By the BDG inequality applied to an Itô Process, Hölder's inequality and (S.A1):

$$\mathbb{E}\left[\sup_{0\leq u\leq t}\mid\int_0^u\sigma(s,X_n(k_n(s)\wedge\tau_R))1\{\tau_R>s\}dW_s\mid^p\right]\leq C_p\mathbb{E}\left[\left(\int_0^t\mid\sigma(s,X_n(k_n(s)\wedge\tau_R))\mid^2 ds\right)^{p/2}\right]$$
$$\leq C_pT^{(p-2)/2}\mathbb{E}\left[\int_0^t\mid\sigma(s,X_n(k_n(s)\wedge\tau_R))\mid^p ds\right]\leq C_pT^{(p-2)/2}K^p\mathbb{E}\left[\int_0^t(1+\mid X_n(k_n(s)\wedge\tau_R)\mid)^p ds\right] \tag{20}$$

Taking the expectation of (18) and plugging in (19) and (20), we have

$$\mathbb{E}\left[\sup_{0\leq u\leq t}|X_n(u\wedge\tau_R)|^p\right]$$

$$\leq 3^{p-1}\mathbb{E}\left[\mid X(0)\mid^p\right]+3^{p-1}\left(T^{p-1}C_3^p n^{ap}+C_pT^{(p-2)/2}K^p\right)\mathbb{E}\left[\int_0^t(1+\mid X_n(k_n(s)\wedge\tau_R)\mid)^p ds\right]$$

$$\leq 3^{p-1}\mathbb{E}\left[\mid X(0)\mid^p\right]+6^{p-1}\left(T^{p-1}C_3^p n^{ap}+C_pT^{(p-2)/2}K^p\right)\mathbb{E}\left[\int_0^t 1+\mid X_n(k_n(s)\wedge\tau_R)\mid^p ds\right] \tag{21}$$

By Tonelli's theorem one writes

$$\mathbb{E}\left[\int_0^t 1+\mid X_n(k_n(s)\wedge\tau_R)\mid^p ds\right]=t+\int_0^t\mathbb{E}\left[\mid X_n(k_n(s)\wedge\tau_R)\mid^p\right]ds$$
$$\leq T+\int_0^t\mathbb{E}\left[\sup_{0<u<s}\mid X_n(k_n(u)\wedge\tau_R)\mid^p\right]ds\leq T+\int_0^t\mathbb{E}\left[\sup_{0<u<s}\mid X_n(u\wedge\tau_R)\mid^p\right]ds$$

Thus, (21) yields

$$\mathbb{E}\left[\sup_{0\leq u\leq t}|X_n(u\wedge\tau_R)|^p\right]$$

$$\leq\max\left((3^{p-1},6^{p-1}\left(T^pC_3^p n^{ap}+C_pT^{(p/2)}K^p\right)\right)(1+\mathbb{E}\left[\mid X(0)\mid^p\right])$$
$$+(6^{p-1}\left(T^{p-1}C_3^p n^{ap}+C_pT^{(p-2)/2}K^p\right))\int_0^t\mathbb{E}\left[\sup_{0<u<s}\mid X_n(u\wedge\tau_R)\mid^p\right]ds$$
$$\leq N_1(n,T,p)(1+\mathbb{E}\left[\mid X(0)\mid^p\right])+N_2(T,p)\int_0^t\mathbb{E}\left[\sup_{0<u<s}\mid X_n(u\wedge\tau_R)\mid^p\right]ds \tag{22}$$

One notices that the stopped Process $(X_n(t \wedge \tau_R))_{t\geq 0}$ is bounded. Indeed, if $|X(0)| \geq R$ then $|X_n(t \wedge \tau_R)| = |X(0)|$, otherwise $|X_n(t \wedge \tau_R)| \leq R$. Therefore $|X_n(t \wedge \tau_R)| \leq \max(|X(0)|, R)$. It easily follows that:

$$\mathbb{E}\left[\sup_{0\leq u\leq t} |X_n(u \wedge \tau_R)|^p\right] \leq \mathbb{E}\left[\max(|X(0)|^p, R^p)\right] < \infty$$

Now by applying the Gronwall's lemma at the integral inequality (22), we get

$$\mathbb{E}\left[\sup_{0\leq u\leq T} |X_n(u \wedge \tau_R)|^p\right] \leq N_1(n,T,p)(1 + \mathbb{E}\left[|\ X(0)\ |^p\right])e^{TN_2(T,p)} = N(n,p,T,\mathbb{E}[|X(0)|^p]) \quad (23)$$

In order to conclude this Proof, one notices that on the event $\{\tau_R < T\}$ we have $\sup_{0\leq t\leq T} |X_n(t \wedge \tau_R)|^p = R^p$ because with probability 1, $X_n(\cdot)$ has continuous paths. Therefore due to Markov's

inequality

$$\mathbb{E}\left[\sup_{0\le t\le T}|X_n(t\wedge\tau_R)|^p\right]\ge R^p\mathbb{P}\left(\sup_{0\le t\le T}|X_n(t\wedge\tau_R)|^p\ge R^p\right)=R^p\mathbb{P}\left(\tau_R<T\right)$$

$$\iff \mathbb{P}\left(\tau_R<T\right)\le N/R^p\Rightarrow\mathbb{P}\left(\tau_R<T\right)\xrightarrow[R\to\infty]{}0$$

As $R\mapsto\tau_R$ is non decreasing, $\lim_{R\to\infty}\tau_R=T$ almost surely. So

$$\lim_{R\to\infty}\sup_{0\le t\le T}|X_n(t\wedge\tau_R)|^p=\sup_{0\le t\le T}|X_n(t)|^p\qquad\text{a.s.}$$

Due to Fatou's lemma and (23) we get

$$\sup_{0\le t\le T}\mathbb{E}\left[|X_n(t)|^p\right]\le\mathbb{E}\left[\sup_{0\le t\le T}|X_n(t)|^p\right]\le\liminf_R\mathbb{E}\left[\sup_{0\le t\le T}|X_n(t\wedge\tau_R)|^p\right]\le N(n,p,T,\mathbb{E}[|X(0)|^p]) \tag{24}$$

Lemma 2.2

Consider the scheme in (9). If for some $p\ge 2$,

$$\sup_{n\ge1}\sup_{0\le t\le T}\mathbb{E}\left[|X_n(t)|^p\right]\le\infty \tag{25}$$

and (S.A1) hold, then

$$\sup_{0\le t\le T}\mathbb{E}\left[|X_n(t)-X_n(k_n(t))|^p\right]\le C_4n^{-p/2} \tag{26}$$

where C_4 is a positive constant independent of n.

Proof.
One immediately writes

$$\mathbb{E}|X_n(t)-X_n(k_n(t))|^p=\mathbb{E}|\int_{k_n(t)}^t g_n(r,X_n(k_n(r)))dr+\int_{k_n(t)}^t\sigma(r,X_n(k_n(r)))dW_r|^p$$

Due to Clarkson's 1st inequality we get

$$\mathbb{E}|X_n(t)-X_n(k_n(t))|^p\le 2^{p-1}\mathbb{E}|\int_{k_n(t)}^t g_n(r,X_n(k_n(r)))dr|^p+2^{p-1}\mathbb{E}|\int_{k_n(t)}^t\sigma(r,X_n(k_n(r)))dW_r|^p \tag{27}$$

By Hölder's inequality and (16), which stands because of (S.A1), one gets

$$\begin{aligned}\mathbb{E}|\int_{k_n(t)}^t g_n(r,X_n(k_n(r)))dr|^p&\le|t-k_n(t)|^{p-1}\mathbb{E}\int_{k_n(t)}^t|g_n(r,X_n(k_n(r)))|^pdr\\&\le(T/n)^{p-1}\mathbb{E}\int_{k_n(t)}^t C_3^pn^{ap}\left(1+|X_n(k_n(r))|\right)^pdr\\&\le(2T)^{p-1}C_3^pn^{p(a-1)+1}\mathbb{E}\int_{k_n(t)}^t 1+|X_n(k_n(r))|^pdr\end{aligned}$$

Observe that for $\forall r \in [k_n(t), t],\ k_n(r) = k_n(t)$, thus

$$\begin{aligned}\mathbb{E}|\int_{k_n(t)}^{t} g_n(r, X_n(k_n(r)))dr|^p &\leq (2T)^{p-1}C_3^p n^{p(a-1)+1}(t - k_n(t))\mathbb{E}\left[1 + |X_n(k_n(t))|^p\right] \\ &\leq 2^{p-1}T^p C_3^p n^{p(a-1)}\left(1 + \mathbb{E}\left[|X_n(k_n(t))|^p\right]\right) \qquad (28)\end{aligned}$$

Notice that (25) implies that (28) can be written as

$$\mathbb{E}|\int_{k_n(t)}^{t} g_n(r, X_n(k_n(r)))dr|^p \leq M_1 n^{p(a-1)} \qquad (29)$$

where M_1 is a constant independent of n and t.
Now, by (S.A1) and the BDG inequality we obtain once again that

$$\begin{aligned}\mathbb{E}|\int_{k_n(t)}^{t} \sigma(r, X_n(k_n(r)))dW_r|^p &\leq \mathbb{E}\sup_{k_n(t)\leq u\leq t}|\int_{k_n(t)}^{u} \sigma(r, X_n(k_n(r)))dW_r|^p \\ &\leq C_p\mathbb{E}|\int_{k_n(t)}^{t} |\sigma(r, X_n(k_n(r)))|^2 dr|^{p/2} \\ &\leq 2^{(p-2)/2}C_p K^{p/2}(T/n)^{p/2}\left(1 + \mathbb{E}\left[|X_n(k_n(t))|^p\right]\right) \\ &\leq M_2 n^{-p/2} \qquad (30)\end{aligned}$$

where M_2 is a constant independent of n and t.
Substituting (29) and (30) in (27) we write

$$\mathbb{E}|X_n(t) - X_n(k_n(t))|^p \leq 2^{p-1}\max(M_1, M_2)\left(n^{p(a-1)} + n^{-p/2}\right), \qquad \forall t \in [0, T] \qquad (31)$$

because $a \in (0, 1/2]$, (31) immediately yields (26).

Remark 2.3

Observe that if (S.A1) holds, then

$$\begin{aligned}2xg_n(t, x) = 2xb_n(t, x) + 2xf(x) &\leq 2x\frac{b(t, x)}{1 + n^{-a}|b(t, x)|} + 2|x||f(x)| \\ &\leq K\frac{1 + |x|^2}{1 + n^{-a}|b(t, x)|} + 2K_f(|x| + |x|^2) \leq K(1 + |x|^2) + 2K_f(1 + |x|)^2 \\ &\leq K(1 + |x|^2) + 4K_f(1 + |x|^2) \leq C_5(1 + |x|^2) \qquad (32)\end{aligned}$$

Lemma 2.4

Suppose that (S.A1) and A4 hold, then for some $C_6 := C_6(T, K, K_f, \mathbb{E}\left[|X(0)|^2\right])$

$$\sup_{n\geq 1}\sup_{0\leq t\leq T}\mathbb{E}\left[|X_n(t)|^2\right] \leq C_6 \qquad (33)$$

Proof.
Let us define

$$I_n(T) = \mathbb{E}\left[\int_0^T \left(X_n(s) - X_n(k_n(s))\right) g_n(s, X_n(k_n(s)))ds\right]$$
$$= \mathbb{E}\left[\int_0^T \left(\int_{k_n(s)}^s g_n(r, X_n(k_n(r)))dr + \int_{k_n(s)}^s \sigma(r, X_n(k_n(r)))dW_r\right) g_n(s, X_n(k_n(s)))ds\right] \tag{34}$$

We calculate due to (16)

$$I_n^1(T) = \mathbb{E}\left[\int_0^T g_n(s, X_n(k_n(s))) \int_{k_n(s)}^s g_n(r, X_n(k_n(r)))drds\right]$$
$$\leq \mathbb{E}\left[\int_0^T |g_n(s, X_n(k_n(s)))| \int_{k_n(s)}^s |g_n(r, X_n(k_n(r)))|drds\right]$$
$$\leq \mathbb{E}\left[\int_0^T |C_3 n^a(1 + X_n(k_n(s))| \int_{k_n(s)}^s |C_3 n^a(1 + X_n(k_n(r))|drds\right]$$
$$\leq C_3^2 n^{2a}\mathbb{E}\left[\int_0^T (s - k_n(s)) (1 + X_n(k_n(s)))^2 ds\right]$$
$$\leq 2TC_3^2 n^{2a-1}\mathbb{E}\left[T + \int_0^T |X_n(k_n(s))|^2 ds\right] \leq 2TC_3^2\mathbb{E}\left[T + \int_0^T |X_n(k_n(s))|^2 ds\right] \tag{35}$$

Also by the tower property and the fact that an Itô Process is a Martingale

$$I_n^2(T) = \mathbb{E}\left[\int_0^T \left(\int_{k_n(s)}^s \sigma(r, X_n(k_n(r)))dW_r\right) g_n(s, X_n(k_n(s)))ds\right]$$
$$= \mathbb{E}\left[\sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \left(\int_{k_n(s)}^s \sigma(r, X_n(k_n(r)))dW_r\right) g_n(s, X_n(k_n(s)))ds\right]$$
$$= \mathbb{E}\left[\sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \left(\int_{t_j}^s \sigma(r, X_n(t_j))dW_r\right) g_n(s, X_n(t_j))ds\right]$$
$$= \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \mathbb{E}\left[g_n(s, X_n(t_j))\left(\int_{t_j}^s \sigma(r, X_n(t_j))dW_r\right)\right] ds$$
$$= \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \mathbb{E}\left[\mathbb{E}\left[g_n(s, X_n(t_j))\left(\int_{t_j}^s \sigma(r, X_n(t_j))dW_r\right)|\mathcal{F}_{t_j}\right]\right] ds$$
$$= \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \mathbb{E}\left[g_n(s, X_n(t_j))\mathbb{E}\left[\left(\int_{t_j}^s \sigma(r, X_n(t_j))dW_r\right)|\mathcal{F}_{t_j}\right]\right] ds = 0 \tag{36}$$

Thus by substituting (34) and (35) into (36) we get

$$I_n(T) \leq 2TC_3^2 \max(1, T)\left(1 + \mathbb{E}\int_0^T |X_n(k_n(s))|^2 ds\right) \tag{37}$$

One immediately notices from (9) that $d[X_n]_t = \sigma(t, X_n(k_n(t)))^2 dt$ then Itô's formula for the function $h(t,x) = x^2$ gives,

$$\begin{aligned}
|X_n(t)|^2 &= |X(0)|^2 + 2\int_0^t X_n(s) g_n(s, X_n(k_n(s)))ds + \int_0^t |\sigma(s, X_n(k_n(s)))|^2 ds \\
&\quad + 2\int_0^t X_n(s)\sigma(s, X_n(k_n(s)))dW_s \\
&= |X(0)|^2 + 2\int_0^t X_n(k_n(s)) g_n(s, X_n(k_n(s)))ds + \int_0^t |\sigma(s, X_n(k_n(s)))|^2 ds \\
&\quad + 2\int_0^t \left(X_n(s) - X_n(k_n(s))\right) g_n(s, X_n(k_n(s)))ds + 2\int_0^t X_n(s)\sigma(s, X_n(k_n(s)))dW_s \qquad (38)
\end{aligned}$$

Now, by taking the expectation of (38) notice that the last term vanishes. Also due to (S.A1) and (32) we get,

$$\mathbb{E}\left[\int_0^t \sigma(s, X_n(k_n(s)))|^2 ds\right] \leq K\left(T + \mathbb{E}\int_0^t |X_n(k_n(s))|^2 ds\right) \leq K\max(1,T)\left(1 + \mathbb{E}\int_0^t |X_n(k_n(s))|^2 ds\right)$$

and

$$\mathbb{E}\left[\int_0^t 2X_n(k_n(s)) g_n(s, X_n(k_n(s)))ds\right] \leq C_5\max(1,T)\left(1 + \mathbb{E}\int_0^t |X_n(k_n(s))|^2 ds\right)$$

From the above and (37) we conclude that

$$\begin{aligned}
\mathbb{E}|X_n(t)|^2 &\leq \mathbb{E}|X(0)|^2 + (C_5 + K + 2TC_3^2)\max(1,T)\left(1 + \mathbb{E}\int_0^t |X_n(k_n(s))|^2 ds\right) \\
&\leq C_7 + \mathbb{E}|X(0)|^2 + C_7\mathbb{E}\int_0^t |X_n(k_n(s))|^2 ds \\
&\leq C_7 + \mathbb{E}|X(0)|^2 + C_7\int_0^t \mathbb{E}|X_n(k_n(s))|^2 ds \\
&\leq C_7 + \mathbb{E}|X(0)|^2 + C_7\int_0^t \sup_{0\leq u\leq s}\mathbb{E}|X_n(u))|^2 ds \qquad (39)
\end{aligned}$$

but because (39) is true $\forall t \in [0,T]$, it holds in the case of supremum as well. We have

$$\sup_{0\leq u\leq t}\mathbb{E}|X_n(u)|^2 \leq C_7 + \mathbb{E}|X(0)|^2 + C_7\int_0^t \sup_{0\leq u\leq s}\mathbb{E}|X_n(u))|^2 ds \qquad (40)$$

furthermore, by (S.A1) and A4 Remark 1 holds, thus it is guaranteed by (17) that

$$\sup_{0\leq u\leq s}\mathbb{E}|X_n(u))|^2 < \infty,$$

then (40) and the application of Gronwall's lemma yield

$$\sup_{0\leq t\leq T}\mathbb{E}|X_n(t))|^2 \leq \left(C_7 + \mathbb{E}|X(0)|^2\right)e^{TC_7} = C_6(T, K, K_f, \mathbb{E}\left[|X(0)|^2\right]) \qquad (41)$$

Notice that (41) holds $\forall n \in \mathbb{N}$ and that the constant C_6 is independent from n, thus (41) implies (33).

Lemma 2.5

Suppose that (S.A1) and (S.A4) hold, then for some $C_8 = C_8\left(T, K, K_f, \mathbb{E}\left[|X(0)|^p\right]\right)$

$$\mathbb{E}\left[\sup_{0\le t\le T}|X(t)|^p\right] \vee \sup_{n\ge 1}\mathbb{E}\left[\sup_{0\le t\le T}|X_n(t)|^p\right] \le C_8 \tag{42}$$

for every $p \le p_0$

Proof.
The first part of (42) is well known in the literature. By Itô's formula we get:

$$\begin{aligned}
|X_n(t)|^p &= |X(0)|^p + p\int_0^t |X_n(u)|^{p-2}X_n(u)g_n(u, X_n(k_n(u)))du + p\int_0^t |X_n(u)|^{p-2}X_n(u)\sigma(u, X_n(k_n(u)))dW_u \\
&+ \frac{p}{2}\int_0^t |X_n(u)|^{p-2}|\sigma(u, X_n(k_n(u)))|^2du + \frac{p(p-2)}{2}\int_0^t |X_n(u)|^{p-4}|X_n(u)\sigma(u, X_n(k_n(u)))|^2du \\
&\le |X(0)|^p + p\int_0^t |X_n(u)|^{p-2}X_n(u)g_n(u, X_n(k_n(u)))du + p\int_0^t |X_n(u)|^{p-2}X_n(u)\sigma(u, X_n(k_n(u)))dW_u \\
&+ \frac{p(p-1)}{2}\int_0^t |X_n(u)|^{p-2}|\sigma(u, X_n(k_n(u)))|^2du
\end{aligned} \tag{43}$$

Taking the expectation of the supremum of (43)

$$\begin{aligned}
\mathbb{E}\left[\sup_{0\le s\le t}|X_n(s)|^p\right] &\le \mathbb{E}\left[|X(0)|^p\right] + p\mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}X_n(u)g_n(u, X_n(k_n(u)))du\right] \\
&+ p\mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}X_n(u)\sigma(u, X_n(k_n(u)))dW_u\right] \\
&+ \frac{p(p-1)}{2}\mathbb{E}\left[\int_0^t |X_n(u)|^{p-2}|\sigma(u, X_n(k_n(u)))|^2du\right]
\end{aligned} \tag{44}$$

Let us now bound each term in (44). We first write:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}X_n(u)g_n(u, X_n(k_n(u)))du\right] \\
&\le \mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}X_n(k_n(u))g_n(u, X_n(k_n(u)))du\right] \\
&+ \mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}\{X_n(u) - X_n(k_n(u))\}g_n(u, X_n(k_n(u)))du\right]
\end{aligned} \tag{45}$$

Notice that (32) implies that:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}X_n(k_n(u))g_n(u, X_n(k_n(u)))du\right] \le \mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}C_5\left(1 + |X_n(k_n(u))|^2\right)du\right] \\
&\le C_5\mathbb{E}\left[\int_0^t |X_n(u)|^{p-2}\left(1 + |X_n(k_n(u))|^2\right)du\right]
\end{aligned} \tag{46}$$

Young's inequality yields that the LHS of (46) is bounded by,

$$\begin{aligned}
&C_5\mathbb{E}\left[\int_0^t \frac{p-2}{p}|X_n(u)|^p du\right] + C_5\mathbb{E}\left[\int_0^t \frac{2}{p}\Big(1 + |X_n(k_n(u))|^2\Big)^{p/2}du\right] \\
&\le C_5\frac{p-2}{p}\mathbb{E}\left[\int_0^t |X_n(u)|^p du\right] + C_5\frac{2^{p+1}}{p}\mathbb{E}\left[\int_0^t 1 + |X_n(k_n(u))|^p du\right]
\end{aligned} \tag{47}$$

Furthermore, substituting from (9) we get:

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}\{X_n(u)-X_n(k_n(u))\}g_n(u,X_n(k_n(u)))du\right]\\
&\le \mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}\left(\int_{k_n(u)}^u g_n(u,X_n(k_n(r)))dr+\int_{k_n(u)}^u \sigma(u,X_n(k_n(r)))dW_r\right)g_n(u,X_n(k_n(u)))du\right]\\
&\le \mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}\int_{k_n(u)}^u g_n(r,X_n(k_n(r)))dr\ g_n(u,X_n(k_n(u)))du\right]\\
&+\mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}\int_{k_n(u)}^u \sigma(r,X_n(k_n(r)))dW_r\ g_n(u,X_n(k_n(u)))du\right]\\
&\le \mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}\int_{k_n(u)}^u g_n(r,X_n(k_n(r)))g_n(u,X_n(k_n(r)))dr\ du\right]\\
&+\mathbb{E}\left[\int_0^t |X_n(u)|^{p-2}\left|\int_{k_n(u)}^u |g_n(u,X_n(k_n(u)))|\,\sigma(r,X_n(k_n(r)))dW_r\right|\ du\right]
\end{aligned}
$$

By (16), Theorem 7.1 in [15] and Young's inequality we obtain

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}\{X_n(u)-X_n(k_n(u))\}g_n(u,X_n(k_n(u)))du\right]\\
&\le \mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}\int_{k_n(u)}^u C_3^2 n^{2a}(1+|X_n(k_n(r))|)^2 dr\ du\right]\\
&+\mathbb{E}\left[\int_0^t |X_n(u)|^{p-2}\left|\int_{k_n(u)}^u |g_n(u,X_n(k_n(r)))|\,\sigma(r,X_n(k_n(r)))dW_r\right|\ du\right]\\
&\le \mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}C_3^2 n^{2a}(1+|X_n(k_n(u))|)^2\int_{k_n(u)}^u dr\ du\right]\\
&+\mathbb{E}\left[\int_0^t \frac{p}{p-2}|X_n(u)|^p du\right]+\frac{2}{p}\int_0^t \mathbb{E}\left[\left|\int_{k_n(u)}^u g_n(u,X_n(k_n(r)))\sigma(r,X_n(k_n(r)))dr\right|^{p/2}\right]du\\
&\le \mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}C_3^2 n^{2a}(1+|X_n(k_n(u))|)^2\frac{T}{n}\ du\right]\\
&+\frac{p}{p-2}\mathbb{E}\left[\int_0^t |X_n(u)|^p du\right]+C_p\left(\frac{T}{n}\right)^{p/4-1}\int_0^t \mathbb{E}\left[\int_{k_n(u)}^u |g_n(u,X_n(k_n(r)))\sigma(r,X_n(k_n(r)))|^{p/2}dr\right]du\\
&\le C_3^2 T n^{2a-1}\mathbb{E}\left[\int_0^t |X_n(u)|^{p-2}(1+|X_n(k_n(u))|)^2\ du\right]+\frac{p}{p-2}\mathbb{E}\left[\int_0^t |X_n(u)|^p du\right]\\
&+C_p\left(\frac{T}{n}\right)^{p/4-1}\int_0^t \mathbb{E}\left[\int_{k_n(u)}^u (C_3 n^a(1+|X_n(k_n(r))|))^{p/2}\left(\sqrt{K}(1+|X_n(k_n(r))|)\right)^{p/2}dr\right]du
\end{aligned}
$$

$$
\begin{aligned}
&\leq C_3^2 T\frac{pn^{2a-1}}{p-2}\mathbb{E}\left[\int_0^t |X_n(u)|^p du\right] + C_3^2 T\frac{2^{p+1}n^{2a-1}}{p}\mathbb{E}\left[\int_0^t 1+|X_n(k_n(u))|^p du\right]\\
&+\frac{p}{p-2}\mathbb{E}\left[\int_0^t |X_n(u)|^p du\right] + Cn^{1-p/4}n^{ap/2}\int_0^t \mathbb{E}\left[(1+|X_n(k_n(u))|)^p \int_{k_n(u)}^u dr\right] du\\
&\leq C_3^2 T\frac{pn^{2a-1}}{p-2}\mathbb{E}\left[\int_0^t |X_n(u)|^p du\right] + C_3^2 T\frac{2^{p+1}n^{2a-1}}{p}\mathbb{E}\left[\int_0^t 1+|X_n(k_n(u))|^p du\right]\\
&+\frac{p}{p-2}\mathbb{E}\left[\int_0^t |X_n(u)|^p du\right] + Cn^{(a-1/2)(p/2)}\int_0^t 1+\mathbb{E}\left[|X_n(k_n(u))|^p\right] du \qquad (48)
\end{aligned}
$$

Notice that $a \leq 1/2 \Rightarrow n^{2a-1} \leq 1$, also by recalling (SA-1), (45) is simplified to

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{0\leq s\leq t}\int_0^s |X_n(u)|^{p-2}X_n(u)g_n(u,X_n(k_n(u)))du\right]\\
&\leq C\left(1+\int_0^t \mathbb{E}\left[|X_n(u)|^p\right] du + \int_0^t \mathbb{E}\left[|X_n(k_n(u))|^p\right] du\right) \qquad (49)
\end{aligned}
$$

for some arbitrary constant $C = C(p,T,K,C_3)$.
Proceeding with bounding the 3rd term of (44), and using the BDG inequality we get

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{0\leq s\leq t}\int_0^s |X_n(u)|^{p-2}X_n(u)\sigma(u,X_n(k_n(u)))dW_u\right] \leq C\mathbb{E}\left[\left(\int_0^t |X_n(u)|^{2p-4}[X_n(u)\sigma(u,X_n(k_n(u)))]^2 du\right)^{1/2}\right]\\
&\leq C\mathbb{E}\left[\left(\int_0^t |X_n(u)|^{2p-2}|\sigma(u,X_n(k_n(u)))|^2 du\right)^{1/2}\right] \leq C\mathbb{E}\left[\left(\sup_{0\leq r\leq t}|X_n(r)|^{2p-2}\int_0^t |\sigma(u,X_n(k_n(u))|^2 du\right)^{1/2}\right]\\
&\leq C\mathbb{E}\left[\sup_{0\leq r\leq t}|X_n(r)|^{p-1}\left(\int_0^t |\sigma(u,X_n(k_n(u))|^2 du\right)^{1/2}\right]
\end{aligned}
$$

Let V be a positive arbitrary constant, then Young's inequality yields

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{0\leq s\leq t}\int_0^s |X_n(u)|^{p-2}X_n(u)\sigma(u,X_n(k_n(u)))dW_u\right]\\
&\leq C\mathbb{E}\left[\sup_{0\leq r\leq t}|X_n(r)|^{p-1}\left(\int_0^t |\sigma(u,X_n(k_n(u))|^2 du\right)^{1/2}\right]\\
&\leq \frac{C}{2V}\mathbb{E}\left[\sup_{0\leq r\leq t}|X_n(r)|^p\right] + C\mathbb{E}\left[\left(\int_0^t |\sigma(u,X_n(k_n(u)))|^2 du\right)^{p/2}\right]\\
&\leq \frac{C}{2V}\mathbb{E}\left[\sup_{0\leq r\leq t}|X_n(r)|^p\right] + C\mathbb{E}\left[\int_0^t K^{p/2}\left(1+|X_n(k_n(u))|^2\right)^{p/2} du\right]\\
&\leq C\left(1++\int_0^t \mathbb{E}\left[|X_n(k_n(u))|^p\right] du + \frac{1}{2V}\mathbb{E}\left[\sup_{0\leq r\leq t}|X_n(r)|^p\right]\right) \qquad (50)
\end{aligned}
$$

Finally, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0\le s\le t}\int_0^s |X_n(u)|^{p-2}|\sigma(u,X_n(k_n(u)))|^2du\right]\\
&\le \mathbb{E}\left[\int_0^t |X_n(u)|^{p-2}|\sigma(u,X_n(k_n(u)))|^2du\right] \le K\mathbb{E}\left[\int_0^t |X_n(u)|^{p-2}\left(1+|X_n(u,X_n(k_n(u)))|^2\right)du\right]\\
&\le \frac{K(p-2)}{p}\mathbb{E}\left[\int_0^t |X_n(u)|^p du\right]+\frac{2K}{p}\mathbb{E}\left[\int_0^t (1+|X_n(k_n(u))|)^{p/2}du\right]\\
&\le \frac{K(p-2)}{p}\mathbb{E}\left[\int_0^t |X_n(u)|^p du\right]+\frac{2^{p+1}K}{p}\mathbb{E}\left[\int_0^t 1+|X_n(k_n(u))|^p du\right]\\
&\le C\left(1+\int_0^t \mathbb{E}\left[|X_n(u)|^p\right]du+\int_0^t \mathbb{E}\left[|X_n(k_n(u))|^p\right]du\right) \qquad (51)
\end{aligned}$$

In context of (49),(50) and (51), its easy to conclude that

$$\mathbb{E}\left[\sup_{0\le s\le t}|X_n(s)|^p\right]\le C\left(1+\int_0^t \mathbb{E}\left[|X_n(u)|^p\right]du+\int_0^t \mathbb{E}\left[|X_n(k_n(u))|^p\right]du+\frac{1}{2V}\mathbb{E}\left[\sup_{0\le s\le t}|X_n(s)|^p\right]\right)$$

where $C=C(p,T,K,C_3,C_5,\mathbb{E}\left[|X(0)|^p\right])$. Now let us choose $V=C$, then one immediately writes

$$\begin{aligned}
\mathbb{E}\left[\sup_{0\le s\le t}|X_n(s)|^p\right] &\le C\left(1+\int_0^t \mathbb{E}\left[|X_n(u)|^p\right]du+\int_0^t \mathbb{E}\left[|X_n(k_n(u))|^p\right]du\right)\\
&\le C\left(1+\int_0^t \mathbb{E}\left[\sup_{0\le s\le t}|X_n(s)|^p\right]du\right) \qquad (52)
\end{aligned}$$

Due to (24) the quantity inside the integral of (52) is finite and thus Grownwall's lemma is applied and yields

$$\mathbb{E}\left[\sup_{0\le s\le t}|X_n(s)|^p\right]\le C(p,T,K,C_3,C_5,\mathbb{E}\left[|X(0)|^p\right])$$

with C being a constant independent of n, implying that:

$$\sup_{n\ge 1}\mathbb{E}\left[\sup_{0\le s\le t}|X_n(s)|^p\right]\le C(p,T,K,C_3,C_5,\mathbb{E}\left[|X(0)|^p\right]) \qquad (53)$$

The desired result is trivially implied by either Hölder's or Lypanov's inequalities for every $p\le p_0$. Now that moments bounds have been establish we remark that under the assumptions of Lemma 2.2 and Lemma 2.4 the convergence and convergence rate results in [9] immediately hold true. We state those theorems for shake of completeness.

Theorem 2.6

Suppose (S.A1)-(S.A3) and A4 hold, then the tamed Euler scheme (9) converges to the true solution SDE (7) in $\mathcal{L}^p$-sense, i.e.

$$\lim_{n\to\infty}\mathbb{E}\left[\sup_{0\ge t\ge T}|X(t)-X_n(t)|^p\right]$$

for all $p\le p_0$

Corollary 2.7

Suppose (S.A1),(S.A3),(S.A5) and (A4) hold, then the tamed Euler scheme (9) with $a = 1/2$ converges to the true solution of SDE (7) in $\mathcal{L}^p$-sense with order $1/2$, i.e.

$$\mathbb{E}\left[\sup_{0\geq t\geq T}|X(t) - X_n(t)|^p\right] \leq Cn^{-p/2}$$

for all $p \leq p_0$, where C is a constant independent of n.

Tamed Un-adjusted Langevin Algorithms

In this chapter we return to the previous established framework of our Introduction in chapter 1. We consider the case in which the drift coefficient ∇U can be thought of as the sum of two term. This way all the superlinearities will be dealt with simultaneously by controlling the first term, while the second one is assumed to satisfy a global Lipschitz continuity condition and thus grow at most linearly. This framework allows us to implement a weaker taming technique than the one previously developed in [10], without losing practical performance. Henceforth, it is assumed that $U = H + F$ is continuously differentiable. We suggest two different partially taming functions $T_h(x)$:

$$G_h(x) = \frac{\nabla H(x)}{1 + h\|\nabla H(x)\|} + \nabla F(x), \text{ and } G_{h,c} = \left(\frac{\partial_i H(x)}{1 + h|\partial_i H(x)|} + \partial_i F(x) \right)_{i \in \{1,\dots,d\}}$$

which in the scope of (6) result to the Partially Tamed Un-adjusted Algorithm (PTYLA) and its coordinate-wise counterpart (PTULAc) respectively.
We now state the following assumptions.

H1. There exists $\ell, L, K \in \mathbb{R}_+$ such that for all $x, y \in \mathbb{R}^d$,
(i) $\|\nabla H(x) - \nabla H(y)\| \leq L \left(1 + \|x\|^\ell + \|y\|^\ell\right) \|x - y\|$
(ii) $\|\nabla F(x) - \nabla F(y)\| \leq L_f \|x - y\|$

H2.
(i) $\liminf_{\|x\| \to +\infty} \dfrac{\|\nabla H(x)\|}{\|x\|} = +\infty$
(ii) $\liminf_{\|x\| \to +\infty} \left\langle \dfrac{x}{\|x\|}, \dfrac{\nabla H(x)}{\|\nabla H(x)\|} \right\rangle > 0$

Under those hypothesis, we obtain some vital remarks. Observe that H2 implies $\liminf_{\|x\| \to +\infty} \nabla H(x) = +\infty$, hence H has a minimum x^* and $\nabla H(x^*) = 0$, due to transnational invariance assume $x^* = 0$ without loss of generality. Then by substituting $y = x^*$ into H1(i) one immediately gets that for all $x \in \mathbb{R}^d$,

$$\|\nabla H(x)\| \leq 2L(1 + \|x\|^{\ell+1}) \tag{54}$$

Also H1(ii) yields for all $x \in \mathbb{R}^d$,

$$\|\nabla F(x)\| \leq \max(L_f, \nabla F(0))(1 + \|x\|) := K(1 + \|x\|)$$

Notice that H2(i) implies that there exists $M, C \in \mathbb{R}_+^*$ such that for all $x \in \mathbb{R}^d,\ \|x\| \geq M$

$$\|\nabla H(x)\| \geq C\|x\| \tag{55}$$

Respectively H2(ii) implies that there exists $M, k \in \mathbb{R}_+^*$ such that for all $x \in \mathbb{R}^d,\ \|x\| \geq M$

$$x \nabla H(x) \geq k\|x\|\|\nabla H(x)\| \tag{56}$$

where M in both cases can be arbitrary large. Local Lipschitz conditions hold trivially for both the coefficients of (4) and thus by classical literature, (4) has a unique strong solution denoted $(Y_t)_{t \geq 0}$.

Moreover, the strongly Makrovian semigroup $(P_t)_{t\geq 0}$ (consult Theorem 5.4.20 in [12]) constructed by

$$P_t(x,A) = P\left(Y_t(x) \in A\right) = \mathbb{E}\left[\mathbb{I}_A(Y_t)|Y_0 = x\right] \quad \forall t \geq 0,\ x \in \mathbb{R}^d \text{ and } A \in \mathcal{B}(\mathbb{R}^d)$$

is reversible with respect to π and hence it admits a unique invariant measure. To further obtain crucial properties such as positive Harris recurrence and exponential ergodicity we consult the work of S.P.Meyn and R.L.Tweedie in [13],[14] and [6] to control the moments of the diffusion with Foster-Lyapunov like criteria. The infinitesimal generator $\mathcal{A}$ associated with (4) is defined by

$$\mathcal{A}f = \lim_{t\to 0} \frac{1}{t}\left(\mathbb{E}_x\left[f(Y_t)\right] - f(x)\right)$$

for all $f \in C^2(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$. Because (4) is of the form $dY_t = b(Y_t)dt + \sigma\left(Y_t\right)dB_t$ with $a := \sigma^\tau\sigma$ we further calculate that

$$\begin{aligned}\mathcal{A}f &= \sum_{i=0}^{d} b_i \frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i=0}^{d}\sum_{k=0}^{d} a_{ik}\frac{\partial^2 f}{\partial x_i \partial x_k} = -\sum_{i=0}^{d}\frac{\partial U}{\partial x_i}\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{k=0}^{d} a_{kk}\frac{\partial^2 f}{\partial {x_k}^2} + \frac{1}{2}\sum_{i\neq k} a_{ik}\cancelto{0}{\frac{\partial^2 f}{\partial x_i \partial x_k}} \\ &= -\left\langle \nabla U(x) | \nabla f(x)\right\rangle + \Delta f(x)\end{aligned}$$

Following the literature we will need a norm-like function which is always greater or equal than 1, so we define the Lyapunov function $V_a : \mathbb{R}^d \to [1,\infty)$ for all $x \in \mathbb{R}^d$ for any $a \in \mathbb{R}_+^*$ by :

$$V_a(x) = \exp\left(a\left(1 + \|x\|^2\right)^{1/2}\right)$$

We then have for all $x \in \mathbb{R}^d$

$$\begin{aligned}\mathcal{A}V_a(x) &= -\nabla U(x)\nabla V_a(x) + \Delta V_a(x) \\ &= -\sum_{i=0}^{d}\frac{\partial U(x)}{\partial x_i}\frac{x_i a V_a(x)}{\left(1+\|x\|^2\right)^{1/2}} + \sum_{i=0}^{d}\frac{aV_a(x)}{\left(1+\|x\|^2\right)^{1/2}} + \frac{x_i^2 a^2 V_a(x)}{1+\|x\|^2} - \frac{x_i^2 a V_a(x)}{\left(1+\|x\|^2\right)^{3/2}} \\ &= -\nabla U(x)\frac{axV_a(x)}{\left(1+\|x\|^2\right)^{1/2}} + \frac{adV_a(x)}{\left(1+\|x\|^2\right)^{1/2}} + \frac{a^2\|x\|^2 V_a(x)}{1+\|x\|^2} - \frac{a\|x\|^2 V_a(x)}{\left(1+\|x\|^2\right)^{3/2}}\end{aligned}$$

We are now ready to prove our first result.

Proposition 3.1

Assume H1,H2 and let $a \in \mathbb{R}_+^*$. There exists $b_a \in \mathbb{R}_+$ such that for all $x \in \mathbb{R}^d$

$$\mathcal{A}V_a(x) \leq -aV_a(x) + ab_a \tag{57}$$

and

$$\sup_{t\geq 0} P_t V_a(x) \leq V_a(x) + b_a$$

Proof. Consider the fraction

$$\begin{aligned}\frac{\mathcal{A}V_a(x)}{aV_a(x)} &= -\frac{\nabla U(x)x}{\left(1+\|x\|^2\right)^{1/2}} + \frac{d}{\left(1+\|x\|^2\right)^{1/2}} + \frac{a\|x\|^2}{1+\|x\|^2} - \frac{\|x\|^2}{\left(1+\|x\|^2\right)^{3/2}} \\ &\leq -\frac{\nabla H(x)x}{\left(1+\|x\|^2\right)^{1/2}} - \frac{\nabla F(x)x}{\left(1+\|x\|^2\right)^{1/2}} + d + a\end{aligned}$$

Due to remark (56) there exists $M_1,\ k_1 \in \mathbb{R}_+^*,\ \|x\| \geq M_1$:

$$\nabla H(x)x \geq k_1\|x\|\|\nabla H(x)\|$$

Also the linear growth of ∇F grants us

$$\nabla F(x)x \geq -2K(1+\|x\|^2)$$

Combining those two inequalities we get that

$$\frac{\nabla H(x)x}{(1+\|x\|^2)^{1/2}} + \frac{\nabla F(x)x}{(1+\|x\|^2)^{1/2}} \geq \frac{k_1\|x\|\|\nabla H(x)\|}{(1+\|x\|^2)^{1/2}} - \frac{2K(1+\|x\|^2)}{(1+\|x\|^2)^{1/2}}$$

By (55) we further get that for $M_2 > M_1, C \in \mathbb{R}_+^*$ such as for $x \in \mathbb{R}^d,\ \|x\| \geq M_2$

$$\|\nabla H(x)\| \geq C\|x\|$$

notice that C can be arbitrary large, let it be such that $C > \frac{1}{k_1 M_2^2}\{(1+a+d)(1+M_2^2)^{1/2} + 2K + 2KM_2^2\}$, and that the map $s \to s^2/(1+s^2)^{-1/2}$ is non-decreasing, then we obtain

$$\begin{aligned}
\frac{\nabla H(x)x}{(1+\|x\|^2)^{1/2}} + \frac{\nabla F(x)x}{(1+\|x\|^2)^{1/2}} &\geq \frac{k_1 C\|x\|^2}{(1+\|x\|^2)^{1/2}} - \frac{2K(1+\|x\|^2)}{(1+\|x\|^2)^{1/2}} \\
&\geq \frac{\|x\|^2}{(1+\|x\|^2)^{1/2}}(k_1 C - 2K) - \frac{2K}{(1+\|x\|^2)^{1/2}} \\
&\geq \frac{\|x\|^2}{(1+\|x\|^2)^{1/2}}\left(\frac{2K}{M_2^2} + \frac{(1+a+d)(1+M_2^2)^{1/2}}{M_2^2}\right) - \frac{2K}{(1+{M_2}^2)^{1/2}} \\
&\geq \frac{{M_2}^2}{(1+{M_2}^2)^{1/2}}\left(\frac{2K}{M_2^2} + \frac{(1+a+d)(1+M_2^2)^{1/2}}{M_2^2}\right) - \frac{2K}{(1+{M_2}^2)^{1/2}} \\
&\geq 1+a+d
\end{aligned}$$

Therefore, for all $x \in \mathbb{R}^d,\ \|x\| \geq M_2$ we have $\frac{\mathcal{A}V_a(x)}{aV_a(x)} \leq -1 \Rightarrow \mathcal{A}V_a(x) \leq -aV_a(x) + ab_a$. Making use of H1 and its remarks, it's also true that

$$\begin{aligned}
\frac{\mathcal{A}V_a(x)}{aV_a(x)} &\leq \|\nabla H(x)\|\frac{\|x\|}{(1+\|x\|^2)^{1/2}} + \|\nabla F(x)\|\frac{\|x\|}{(1+\|x\|^2)^{1/2}} + d + a \leq \|\nabla H(x)\| + \|\nabla F(x)\| + a + d \\
&\leq 2L(1+\|x\|^{\ell+1}) + K(1+\|x\|) + a + b \leq 2^\ell(2L+K)(1+\|x\|^{\ell+1}) + a + d
\end{aligned}$$

Hence for all $x \in \mathbb{R}^d,\ \|x\| < M_2$ we have

$$\begin{aligned}
\frac{\mathcal{A}V_a(x)}{aV_a(x)} &\leq 2^\ell(2L+K)(1+{M_2}^{\ell+1}) + a + d \leq \frac{V_a(x)}{V_a(x)}\{2^\ell(2L+K)(1+{M_2}^{\ell+1}) + a + d\} \\
&\leq \frac{\exp\left(a\,(1+M_2)^{1/2}\right)}{V_a(x)}\{2^\ell(2L+K)(1+{M_2}^{\ell+1}) + a + d\}
\end{aligned}$$

Equivalently

$$\begin{aligned}\mathcal{A}V_a(x) &\leq a\exp\left(a\left(1+M_2\right)^{1/2}\right)\left\{2^\ell(2L+K)(1+{M_2}^{\ell+1})+a+d\right\}\\ &\leq -aV_a(x)+aV_a(x)+a\exp\left(a\left(1+M_2\right)^{1/2}\right)\left\{2^\ell(2L+K)(1+{M_2}^{\ell+1})+a+d\right\}\\ &\leq -aV_a(x)+a\exp\left(a\left(1+M_2\right)^{1/2}\right)\left\{2^\ell(2L+K)(1+{M_2}^{\ell+1})+1+a+d\right\}\\ &\leq -aV_a(x)+ab_a\end{aligned}$$

where b_a is defined as the positive constant $a\exp\left(a\left(1+M_2\right)^{1/2}\right)\left\{2^\ell(2L+K)(1+{M_2}^{\ell+1})+1+a+d\right\}$. All in all we proved that the Lyapunov-Foster like condition $\mathcal{A}V_a(x)\leq -aV_a(x)+ab_a$ holds true for all $x\in\mathbb{R}^d$. To conclude the proof we apply the Dynkin formula to the function $e^{at}V_a(x)$ and we get

$$\begin{aligned}\mathbb{E}_x\left[e^{at}V_a(Y_t)\right] &= e^{a0}V_a(x)+\mathbb{E}_x\left[\int_0^t \mathcal{A}e^{as}V_a(x)ds\right] = V_a(x)+\mathbb{E}_x\left[\int_0^t e^{as}\left(\mathcal{A}V_a(x)+aV_a(x)\right)ds\right]\\ &\leq V_a(x)+\mathbb{E}_x\left[\int_0^t e^{as}ab_a ds\right] \leq V_a(x)+b_a(e^{at}-1)\end{aligned}$$

Which leads us to

$$P_tV_a(x)e^{at}\leq V_a(x)+b_a(e^{at}-1)\Rightarrow P_tV_a(x)\leq V_a(x)e^{-at}+b_a(1-e^{-at})\Rightarrow \sup_{t\geq 0}P_tV_a(x)\leq V_a(x)+b_a$$

□

First and foremost (57) guarantees by Theorem 2.1 in [13] that the process $(Y_t)_{t\geq 0}$ is non-explosive in the sense that $\mathbb{P}_x\{Y_t\to+\infty\}=0$ for all $x\in\mathbb{R}^d$. Furthermore due to Theorem 4.2 in [13] $(Y_t)_{t\geq 0}$ is Harris recurrent and $\pi(V_a)$ is finite. Therefore the diffusion process is positive Harris recurrent, π can be normalized to a probability measure and thus from now on the invariant measure π can also be referenced as the stationary distribution of (4). By far the strongest result is derived from Theorem 6.1 [13] which yields the convergence of the semigroup $(P_t)_{t\geq 0}$ to the stationary distribution π in the V_a-norm (ergodicity) and does so with an exponentially fast rate, that is:

$$\left\|P_t(x,\cdot)-\pi\right\|_{V_a}\leq C_a\rho_a^t V_a,\quad t\in\mathbb{R}_+,\quad x\in\mathbb{R}^d \tag{58}$$

with $C_a\in\mathbb{R}_+,\ \rho_a\in[0,1)$ and the V_a-norm being defined as $\|f_1-f_2\|_{V_a}:=\sup_{|g|\leq V_a}|f_1g-f_2g|$. Therefore we say that $(Y_t)_{t\geq 0}$ is V_a-exponentially ergodic. While trivial, its important to note that the establishment of (58) justifies the use of the Langevin diffusion in sampling and optimization algorithms, because the drift is designed to ensure convergence to the target π. For some initial distributions $\mu_0,\ \nu_0$ on $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$ satisfying $\mu_0(V_a)+\nu_0(V_a)<+\infty$ we further get:

$$\begin{aligned}\left\|\mu_0P_t(x,\cdot)-\pi\right\|_{V_a} &= \sup_{|g|\leq V_a}|\mu_0P_tg-\pi g| = \sup_{|g|\leq V_a}|\mu_0P_tg-\mu_0\pi g|\\ &\leq \mu_0\sup_{|g|\leq V_a}|P_tg-\pi g| = \int\mu_0(dx)\left\|P_t(x,\cdot)-\pi\right\|_{V_a}\\ &\leq \int\mu_0(dx)V_a(x)C_a\rho_a^t = C_a\rho_a^t\int\mu_0(dx)V_a(x) = C_a\rho_a^t\mu_0(V_a)\end{aligned}$$

Similarly we also have $\left\|\mu_0P_t-\nu_0P_t\right\|_{V_a}\leq C_a\rho_a^t\|\mu_0-\nu_0\|_{V_a}$. Although that the desired behavior of $(Y_t)_{t\geq 0}$ has been ensured, a naive discretization of (4) might not converge even if the diffusion itself

does so [6], thus we need to provide analogous arguments for the Markov chain $(X_n)_{n\in\mathbb{N}}$ produced by the discrete scheme (6). In particular, in Lemma 3.2 properties of $G_h(x)$ are stated which later on guarantee the geometrical ergodicity of (6) with respect to π_h. At this point, it's important in order to avoid confusion, to note that π_h doesn't necessary equals π in most cases but, because our goal is to approximate the behavior of π, this fact doesn't constrain us as long as those two distributions are close to each other, that is in the sense of total variation and Wasserstein distances. To see that consider a multivariate Gaussian variable with mean 0 and convariance matrix I_d, the potential induced from such choice is $U(x) = \|x\|^2/2$. Then under ULA for $h = 1$ we have that

$$X_{n+1} \sim (X_n - \nabla U(X_n), 2I_d) \sim (0, 2I_d) \nsim (0, I_d)$$

It should be clear now that both the stepsize h and the taming function $G_h(x)$ do in fact affect the stationary distribution π_h, which shall not be considered identical to π.

Lemma 3.2

Asssume H1 and H2. Let $h > 0$ and T_h be equal to G_h or $G_{h,c}$, then the following hold true:

P1 There exist $a \geq 0,\ C_a < +\infty$ such that for all $h > 0$ and $x \in \mathbb{R}^d$,

$$\|T_h(x) - \nabla U(x)\| \leq hC_a(1 + \|x\|^a)$$

P2 For all $h > 0$,

$$\liminf_{\|x\|\to+\infty} \left\langle \frac{x}{\|x\|}, T_h(x) \right\rangle - \frac{h}{2\|x\|}\|T_h(x)\|^2 - \left\langle \frac{x}{\|x\|}, \nabla F(x) \right\rangle + \frac{h}{2\|x\|}\|\nabla F(x)\|^2 > 0$$

Proof. Let $h > 0$ we have,

$$\begin{aligned}\|G_h(x) - \nabla U(x)\| &= \left\| \frac{\nabla H(x)}{1 + h\|\nabla H(x)\|} + \nabla F(x) - (\nabla H(x) + \nabla F(x)) \right\| = \left\| \frac{\nabla H(x)}{1 + h\|\nabla H(x)\|} - \nabla H(x) \right\| \\ &\leq \|\nabla H(x)\| \left\| \frac{1 - 1 - h\|\nabla H(x)\|}{1 + h\|\nabla H(x)\|} \right\| \leq h\|\nabla H(x)\|^2\end{aligned}$$

Similarly

$$\begin{aligned}\|G_{h,c}(x) - \nabla U(x)\| &= \left(\sum_{i=1}^d \left[\frac{\partial_i H(x)}{1 + h\partial_i H(x)} - \partial_i H(x) \right]^2 \right)^{1/2} \leq \left(\sum_{i=1}^d |\partial_i H(x)|^2 \left[\frac{-h\partial_i H(x)}{1 + h\partial_i H(x)} \right]^2 \right)^{1/2} \\ &\leq \left(\sum_{i=1}^d h^2 |\partial_i H(x)|^4 \right)^{1/2} \leq h \sum_{i=1}^d |\partial_i H(x)|^2 = h\|\nabla H(x)\|^2\end{aligned}$$

Hence due to remark (54) in both cases we get

$$\|T_h(x) - \nabla U(x)\| \leq h(2L)^2 \left(\left(1 + \|x\|^{\ell+1} \right) \right)^2 \leq hL^2 4^{l+2} \left(1 + \|x\|^{2\ell+2} \right) := hC_{2\ell+2} \left(1 + \|x\|^{2\ell+2} \right)$$

Furthermore we have,

$$\begin{aligned}B_h(x) &:= \left\langle \frac{x}{\|x\|}, T_h(x) \right\rangle - \frac{h}{2\|x\|}\|T_h(x)\|^2 - \left\langle \frac{x}{\|x\|}, \nabla F(x) \right\rangle + \frac{h}{2\|x\|}\|\nabla F(x)\|^2 \\ &= \left\langle \frac{x}{\|x\|}, \frac{\nabla H(x)}{1 + h\|\nabla H(x)\|} \right\rangle - \frac{h}{2\|x\|} \frac{\|\nabla H(x)\|^2}{(1 + h\|\nabla H(x)\|)^2} - \frac{h}{\|x\|} \left\langle \frac{\nabla H(x)}{1 + h\|\nabla H(x)\|}, \nabla F(x) \right\rangle\end{aligned}$$

Notice that $\frac{h\|\nabla H(x)\|}{1+h\|H(x)\|} < 1$, then under H2(ii), for $x \in \mathbb{R}^d$ with $\|x\| > M_1$ there exists $k > 0$ such as :

$$\begin{aligned} B_h(x) &> \frac{k\|x\|\|\nabla H(x)\|}{\|x\|\left(1+h\|\nabla H(x)\|\right)} - \frac{1}{2\|x\|}\frac{\|\nabla H(x)\|}{(1+h\|\nabla H(x)\|)} - \frac{\|\nabla F(x)\|}{\|x\|} \\ &> \frac{\|H(x)\|}{\|x\|\left(1+h\|H(x)\|\right)}\{k\|x\| - \frac{1}{2} - K(1+\|x\|)\frac{1+h\|H(x)\|}{\|H(x)\|}\} \end{aligned}$$

Using the fact that the map $s \to s/(1+hs)^{-1}$ is non decreasing and that under H2(i), for all $x \in \mathbb{R}^d$ with $\|x\| \geq M_2 > M_1$ there exists $C' > 0$ such that $\|H(x)\| \geq C'\|x\| \geq C'M_2 := C$

$$\begin{aligned} B_h(x) &> \frac{C}{\|x\|\left(1+hC\right)}\{k\|x\| - \frac{1}{2} - K(1+\|x\|)\frac{1+hC}{C}\} \\ &> \frac{C''}{\|x\|}\{\left(k - \frac{K}{C''}\right)\|x\| - \frac{K}{C''} - \frac{1}{2}\} \end{aligned}$$

Let $k = 2K/C''$ and $M_2 > 2 + C''/K$, we eventually get:

$$B_h(x) > \frac{C''}{\|x\|K}\left(\frac{K}{C''}M_2 - \frac{K}{C''} - 1/2\right) > 0 \iff \liminf_{\|x\|\to+\infty} B_h(x) > 0$$

Before we lay out the coordinate-wise case, let us first state some basic, yet useful, norm inequalities

$$h\left\|\left(\frac{\partial_i H(x)}{1+h\partial_i H(x)}\right)_{i\in\{1,\dots,d\}}\right\| = h\left(\sum_{i=i}^{d}\left[\frac{1}{h}\frac{h\partial_i H(x)}{1+h|\partial_i H(x)|}\right]^2\right)^{1/2} \leq h\left(\sum_{i=i}^{d}\frac{1}{h^2}\right)^{1/2} \leq \sqrt{d}$$

$$\left\|\left(\frac{\partial_i H(x)}{1+h\partial_i H(x)}\right)_{i\in\{1,\dots,d\}}\right\| = \left(\sum_{i=i}^{d}\left[\frac{\partial_i H(x)}{1+h|\partial_i H(x)|}\right]^2\right)^{1/2} \leq \frac{\sqrt{d}\|\nabla H(x)\|}{1+h\max|\partial_i H(x)|}$$

and for all $x \in \mathbb{R}^d$ with $\|x\| > M_3$ there exists $k_2 > 0$ such as

$$\left\langle x, \left(\frac{\partial_i H(x)}{1+h\partial_i H(x)}\right)_{i\in\{1,\dots,d\}}\right\rangle \geq \frac{1}{1+h\max|\partial_i H(x)|}\langle x, \nabla H(x)\rangle \geq \frac{k_2\|x\|\|\nabla H(x)\|}{1+h\max|\partial_i H(x)|} \geq \frac{k_2\|x\|\|\nabla H(x)\|}{1+h\|\nabla H(x)\|}$$

Combining these we get that

$$\begin{aligned} B_{h,c}(x) &\geq \left\langle \frac{x}{\|x\|}, \left(\frac{\partial_i H(x)}{1+h\partial_i H(x)}\right)_{i\in\{1,\dots,d\}}\right\rangle - \frac{h}{2\|x\|}\left\|\left(\frac{\partial_i H(x)}{1+h\partial_i H(x)}\right)_{i\in\{1,\dots,d\}}\right\|^2 \\ &\quad - \frac{h\|\nabla F(x)\|}{\|x\|}\left\|\left(\frac{\partial_i H(x)}{1+h\partial_i H(x)}\right)_{i\in\{1,\dots,d\}}\right\| \\ &\geq \frac{k_2\|x\|\|\nabla H(x)\|}{\|x\|(1+h\max|\partial_i H(x)|)} - \frac{d}{2\|x\|}\frac{\|\nabla H(x)\|}{1+h\max|\partial_i H(x)|} - \frac{\sqrt{d}}{\|x\|}\|\nabla F(x)\| \\ &\geq \frac{\|\nabla H(x)\|}{\|x\|(1+h\max|\partial_i H(x)|)}\{k_2\|x\| - d - \sqrt{d}K(1+\|x\|)\frac{1+h\max|\partial_i H(x)|}{\max|\partial_i H(x)|}\} \end{aligned}$$

For all $x \in \mathbb{R}^d$ with $\|x\| > M_4 > M_3$ there exists $C > 0$ such that $\|\nabla H(x)\| > C$. Furthermore for $k_2 = 2\sqrt{d}K\dfrac{1+hC}{C}$ and letting $M_4 > 1 + \dfrac{\sqrt{d}}{K}\dfrac{C}{1+hC}$, we get

$$\begin{aligned} B_{h,c}(x) &\geq \frac{\|\nabla H(x)\|}{\|x\|(1+h\max|\partial_i H(x)|)}\{\sqrt{d}K\frac{1+hC}{C}M_4 - d - \sqrt{d}K\frac{1+hC}{C}\} \\ &\geq \frac{\|\nabla H(x)\|}{\|x\|(1+h\|\nabla H(x)\|)}\{\sqrt{d}K\frac{1+hC}{C}M_4 - d - \sqrt{d}K\frac{1+hC}{C}\} \\ &\geq \frac{C}{\|x\|(1+hC)}\{\sqrt{d}K\frac{1+hC}{C}M_4 - d - \sqrt{d}K\frac{1+hC}{C}\} > 0 \end{aligned}$$

Observe that P1 with remark (54) also implies that

$$\|T_h(x)\| \leq \|T_h(x) - \nabla U(x)\| + \|\nabla U(x)\| \leq hC_a(1+\|x\|^a) + (2L+K)(1+\|x\|^{\ell+1})$$

Before we proceed with the discrete counterpart of Proposition 3.1 we give the Markov kernel R_h associated with (6) for all $h > 0, x \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$::

$$R_h(x,A) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} \mathbb{I}_A\left(x - hT_h(x) + \sqrt{2h}z\right)e^{-\|z\|^2/2}dz$$

Proposition 3.3

Assume H1,H2 and let $h > 0$. There exist $M, æ, b \in \mathbb{R}_+^*$ satisfying for all $x \in \mathbb{R}^d$

$$R_h V_{æ}(x) \leq e^{-æ^2 h}V_{æ}(x) + hb\mathbb{I}_{\bar{B}(0,M)}(x) \tag{59}$$

Proof. Let $h, a \in \mathbb{R}_+^*$. The function $x \to (1+\|x\|)^{1/2}$ is suffiicenty smooth as Lipschitz continuous, by the log-Sobolev inequality for gaussian measures [16], and the Cauchy-Schwarz inequality, we have for all $x \in \mathbb{R}^d$ and $a > 0$:

$$\begin{aligned} R_h V_a(x) &= \int_{\mathbb{R}^d}\exp\left(a(1+\|y\|^2)^{1/2}\right)R_h(x,dy) \leq e^{a^2h}\exp\left\{a\int_{\mathbb{R}^d}(1+\|y\|^2)^{1/2}R_h(x,dy)\right\} \\ &\leq e^{a^2h}\exp\left\{a\left(1+\|x-hT_h(x)\|^2+2hd\right)^{1/2}\right\} \end{aligned} \tag{60}$$

Similarly with the continuous version at Proposition 3.1, we will bound the r.h.s. separately inside and outside an arbitrary ball on $\mathbb{R}^d$

$$\|x - hT_h(x)\|^2 = \|x\|^2 - 2h\left(\langle T_h(x), x\rangle - (h/2)\|T_h(x)\|^2\right)$$

Using P2 from Lemma 2, for all $x \in \mathbb{R}^d$ with $\|x\| > M_1$ there exists $k > 0$ such that

$$\begin{aligned} \|x - hT_h(x)\|^2 &\leq \|x\|^2 - 2hk\|x\| - 2h\langle x, \nabla F(x)\rangle + h^2\|\nabla F(x)\|^2 \\ &\leq \|x\|^2 - 2hk\|x\| + 4hK(1+\|x\|^2) + 2h^2K^2(1+\|x\|^2) \end{aligned}$$

Let us choose now $M = \max(M_1, 2k^{-1}(d+2K+hK^2))$ then for all $x \in \mathbb{R}^d$ with $\|x\| > M$

$$\|x - hT_h(x)\|^2 + 2hd \leq (1+C_{hK})\|x\|^2 - hk\|x\| \tag{61}$$

We have

$$\begin{aligned}\Big(1+\|x-hT_h(x)\|^2+2hd\Big)^{1/2} &\leq \Big(1+(1+C_{hK})\|x\|^2-hk\|x\|\Big)^{1/2}\\ &\leq \Big(1+(1+C_{hK})\|x\|^2\Big)^{1/2}\left(1-\frac{hk\|x\|}{1+(1+C_{hK})\|x\|^2}\right)^{1/2}\end{aligned}$$

It is implied by (61) that $hk\|x\| \leq 1+(1+C_{hK})\|x\|^2$, using for all $s\in[0,1]$, $(1-s)^{1/2}\leq 1-s/2$, we write

$$\begin{aligned}\Big(1+\|x-hT_h(x)\|^2+2hd\Big)^{1/2} &\leq \Big(1+(1+C_{hK})\|x\|^2\Big)^{1/2}\left(1-\frac{1}{2}\frac{hk\|x\|}{1+(1+C_{hK})\|x\|^2}\right)\\ &\leq \Big(1+(1+C_{hK})\|x\|^2\Big)^{1/2}-\frac{1}{2}\frac{hk\|x\|}{\Big(1+(1+C_{hK})\|x\|^2\Big)^{1/2}}\\ &\leq \Big(1+(1+C_{hK})\|x\|^2\Big)^{1/2}-\frac{hkM}{2(1+(1+C_{hK})M^2)^{1/2}}\end{aligned}$$

Plugging this back to (60) shows

$$\begin{aligned}R_hV_a(x) &\leq e^{a^2h}\exp\left\{a\left(1+(1+C_{hK})\|x\|^2\right)^{1/2}-\frac{ahkM}{2(1+(1+C_{hK})M^2)^{1/2}}\right\}\\ &\leq \exp\left\{a^2h-a\frac{hkM}{2(1+(1+C_{hK})M^2)^{1/2}}\right\}\exp\left\{a\left(1+(1+C_{hk})\|x\|^2\right)^{1/2}\right\}\end{aligned}$$

Choosing $æ = \dfrac{kM}{4(1+(1+C_{hK})M^2)^{1/2}}$, results in $R_hV_{æ}(x)\leq e^{-æ^2h}V_{æ}((1+C_{hK})^{1/2}\cdot x)$ and thus the change of variable $x\to x(1+Chk)^{-1/2}$ eventually yields for all $x\in\mathbb{R}^d$ with $\|x\|\geq M$

$$R_hV_{æ}(x)\leq e^{-æ^2h}V_{æ}(x)$$

To get the desired result for $\|x\|<M$, one uses the basic inequality $(1+s_1+s_2)^{1/2}\leq(1+s_1)^{1/2}+s_2/2$ which holds for all $s_1,s_2\geq 0$, hence

$$\begin{aligned}\Big(1+\|x-hT_h(x)\|^2+2hd\Big)^{1/2} &\leq \Big(1+\|x\|^2+2h\|x\|\|T_h(x)\|+h^2\|T_h(x)\|^2+2hd\Big)^{1/2}\\ &\leq (1+\|x\|^2)^{1/2}+h\|x\|\|T_h(x)\|+\frac{h^2}{2}\|T_h(x)\|^2+hd\\ &\leq (1+\|x\|^2)^{1/2}+hM\Big(hC_a(1+M^a)+(2L+K)(1+M^{\ell+1})\Big)\\ &\quad+\frac{h^2}{2}\Big(hC_a(1+M^a)+(2L+K)(1+M^{\ell+1})\Big)^2+hd\\ &\leq (1+\|x\|^2)^{1/2}+hC(a,M,L,\ell,K,h,d)\end{aligned}$$

Setting $c=æ^2+æC(a,M,L,\ell,K,h,d)$ we get that

$$R_hV_{æ}(x)\leq e^{hc}V_{æ}(x)$$

Exploiting the convexity of e^{-s}, we finally have

$$\begin{aligned}R_hV_{æ}(x)-e^{-æ^2h}V_{æ}(x) &\leq (e^{hc}-e^{-æ^2h})V_{æ}(x)\leq e^{hc}(1-e^{-(æ^2+c)h})V_{æ}(x)\\ &\leq he^{hc}(æ^2+c)V_{æ}(x)\leq he^{hc}(æ^2+c)e^{kM/4}:=hb\end{aligned}$$

The condition (59) according to the Geometric Ergodic Theorem (15.0.1) in [14], suffices for R_h to admit a unique invariant probability measure π_h and for it to be $V_{æ}-$ geometrically ergodic with respect to π_h. Additionally, a Lyapunov-Foster condition can be provided when the kernel has acted $n-$times, by using an induction argument and the basic inequality $1 - e^{-s} \geq se^{-s}$, we get

$$R_h^n V_{æ}(x) \leq e^{-æ^2 nh} V_{æ}(x) + (b/æ^2)e^{æ^2 h}$$

Under our framework of assuming the gradient of the potential U to be the sum of a superlinear and Lipschitz continuous terms, we have shown that we end up in the same class of potentials described in [10] in the sense that our adapted versions of Proposition 3.1 and 3.3 from [10] are identical to the latter up to some numerical coefficients and constants. Thus the following results are yielded:

Theorem 3.4

Assume H1,H2. Let $h_0 > 0$. There exist $C > 0$ and $\lambda \in (0, 1)$ such that for all $h \in (0, h_0]$, $x \in \mathbb{R}^d$ and $n \in \mathbb{N}$

$$\|\delta_x R_h^n - \pi\|_{V_{æ}^{1/2}} \leq C(nh\lambda^{nh} V_{æ}(x) + \sqrt{h})$$

and for all $h \in (0, h_0]$,

$$\|\pi_h - \pi\|_{V_{æ}^{1/2}} \leq C\sqrt{h}$$

Theorem 3.5

Assume H1,H2 and strong convexity on $\nabla H(x)$.Let $h_0 > 0$. There exist $C > 0$ and $\lambda \in (0, 1)$ such that for all $h \in (0, h_0]$, $x \in \mathbb{R}^d$ and $n \in \mathbb{N}$

$$W_2^2(\delta_x R_h^n, \pi) \leq C(nh\lambda^{nh} V_{æ}(x) + \sqrt{h})$$

and for all $h \in (0, h_0]$,

$$W_2^2(\pi_h, \pi) \leq C\sqrt{h}$$

The bounds of Theorem 3.5 are further improved by assuming the potential to be twice continuously differentiable and it's Laplacian to be locally $\beta-$ Hölder continuous.

H3. $U \in C^2(\mathbb{R}^d, \mathbb{R})$ There exists $v, L_H \in \mathbb{R}_+$ and $\beta \in [0, 1]$such that for all $x, y \in \mathbb{R}^d$,

$$\|\nabla^2 H(x) - \nabla^2 H(y)\| \leq L_H \left(1 + \|x\|^v + \|y\|^v\right) \|x - y\|^\beta$$

Theorem 3.6

Assume H1,H2,H3 and strong convexity on $\nabla H(x)$.Let $h_0 > 0$. There exist $C > 0$ and $\lambda \in (0, 1)$ such that for all $h \in (0, h_0]$, $x \in \mathbb{R}^d$ and $n \in \mathbb{N}$

$$W_2^2(\delta_x R_h^n, \pi) \leq C(nh^{1+\beta}\lambda^{nh} V_{æ}(x) + h^{1+\beta})$$

and for all $h \in (0, h_0]$,

$$W_2^2(\pi_h, \pi) \leq Ch^{1+\beta}$$

Benchmarking those algorithms in our numerical illustrations, require the estimation of the first and second moments by their empirical average. To this purpose we further assume:

H4. $H \in C^4(\mathbb{R}^d, \mathbb{R})$ and $\|D^i H\| \in C_{\text{poly}}(\mathbb{R}^d, \mathbb{R}_+)$ for $i \in \{1, \dots, 4\}$

Theorem 3.7

Assume H1,H2,H4. Let $f \in C^3(\mathbb{R}^d,)$ be such that $\|D^i f\| \in C_{\text{poly}}(\mathbb{R}^d, \mathbb{R}_+)$ for $i \in \{0, \dots, 3\}$. Let $h_0 > 0$ and $(X_k)_{k\in\mathbb{N}}$ be the Markov chain defined by () and starting at $X_0 = 0$. There exists $C > 0$ such that for all $h \in (0, h_0]$ and $n \in \mathbb{N}^*$,

$$\left| \mathbb{E}\left[\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \pi(f) \right]) \right| \leq C(h + \frac{1}{nh})$$

and

$$\mathbb{E}\left[\left(\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \pi(f) \right)^2 \right]) \leq C(h^2 + \frac{1}{nh})$$

Numerical illustrations

We choose to use the double well for our example because the potential $U(x) = (1/4)\|x\|^4 - (1/2)\|x\|^2$ with $\nabla U(x) = \|x\|^2 x - x$ fits perfectly the partially taming algorithm's framework. We briefly show that it satisfies our hypothesis, in particular H1,H2 and H4:

$$\|\nabla H(x) - \nabla H(y)\| = \left\| \|x\|^2 x - \|y\|^2 y \right\| \leq (\|x\| + \|y\|)^2 \|x - y\| \leq 2(1 + \|x\|^2 + \|y\|^2)\|x - y\|$$

$$\left\langle \frac{x}{\|x\|}, \frac{\nabla H(x)}{\|\nabla H(x)\|} \right\rangle = \left\langle \frac{x}{\|x\|}, \frac{\|x\|^2 x}{\|x\|^3} \right\rangle = 1 > 0$$

So by Propositions 3.1 and 3.3 both the Langevin dynamics and the discrete processes generated by its partially tamed scheme are ergodic and by Theorem 3.4 their invariant distributions will be sufficiently close in the sense of total variation, thus their behavior should be similar in big samples. Furthermore Theorem 3.7 enables us to estimate the first and second moments of the stationary distribution by their empirical means. We benchmark PTULA and PTULAc against ULA and the previously developed tamed algorithms in [10] TULA and TULAc. As noticed in [10] the double well model is coordinate wise exchangeable, i.e. each coordinate evolves independently of each other, and so only data from the first and the last coordinates are saved for calculations. We use 3 different initialization points for all algorithms, at $X_0 = 0, (10, 0^{\otimes d-1}), (100, 0^{\otimes d-1})$ each for 3 different stepsizes $h = 10^{-1}, 10^{-2}, 10^{-3}$. For every unique combination, 100 independently Markov chains are generated at X_0 and are allowed to run for $10^5(10^4)$ iterations in dimension $d = 100(1000$ respectively). In both cases we implement a burn-in period of magnitude 10^4. If a trajectory exceeds a critical value or we encounter instability due to machine precision, the

instance is discarded and the corresponding data are not taken into account in our results. As for the moments, $\int_{\mathbb{R}^d} x_i \pi(x)dx = 0$ for all $i \in \{1, \dots, d\}$ cause of symmetry and its estimated in [10] using RWM of 10^7 samples that $\int_{\mathbb{R}^d} {x_i}^2 \pi(x)dx = 0.104 \pm 0.001$ and $\int_{\mathbb{R}^d} {x_i}^2 \pi(x)dx = 0.032 \pm 0.001$ for all $i \in \{1, \dots, d\}$ in dimension d=100 and d=1000 respectively. Looking at the displayed figures starting by the first one, we remark that TULA has a substantially increased bias in comparison to the rest of the algorithms, which is particularly visible for big stepsizes i.e. $h > 10^{-2}$. Additionally PTULA and PTULAc seem to behave similarly to TULAc. In figure 2 starting further from the origin, we get equivalent results. In figure 3 we finally witness ULA diverging even for the smallest stepsize of 10^{-3} while PTULAc fails in machine precision for big stepsize. Proceeding at dimension $d = 100$, Figures 4 and 5 don't give us any remarkable insight on the behavior of the algorithms, that is the results are analogous with their counterparts in dimension $d = 100$. We do notice however a disturbing pattern on PTULAc for big stepsizes when the starting point is far from the origin, in figure 6. Overall PTULA yields similar results to TULAc, partially taming seems to recover the unwanted bias that TULA produces, which was one of our initial motivations. Still, the sometimes strange behavior of PTULAc needs further elaboration. A naive interpretation could be that either partial taming doesn't mix well with coordinate wise taming, or it's implementation demands more careful operations in terms of machine precision. In any case, further research is needed.

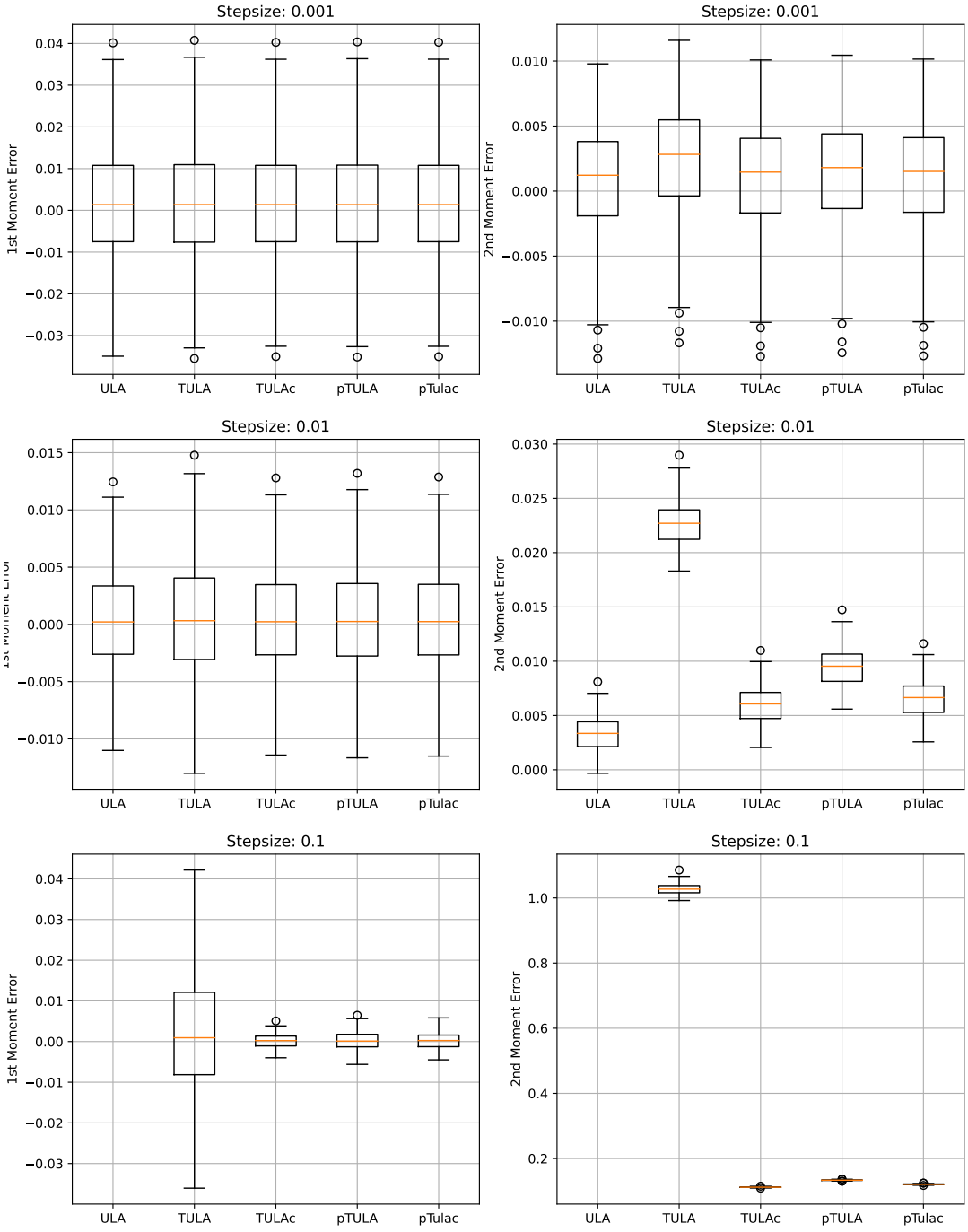

Figure 1: Boxplots of the error on the first and second moment for the Double Well in dimension 100 starting at 0 for different stepsizes

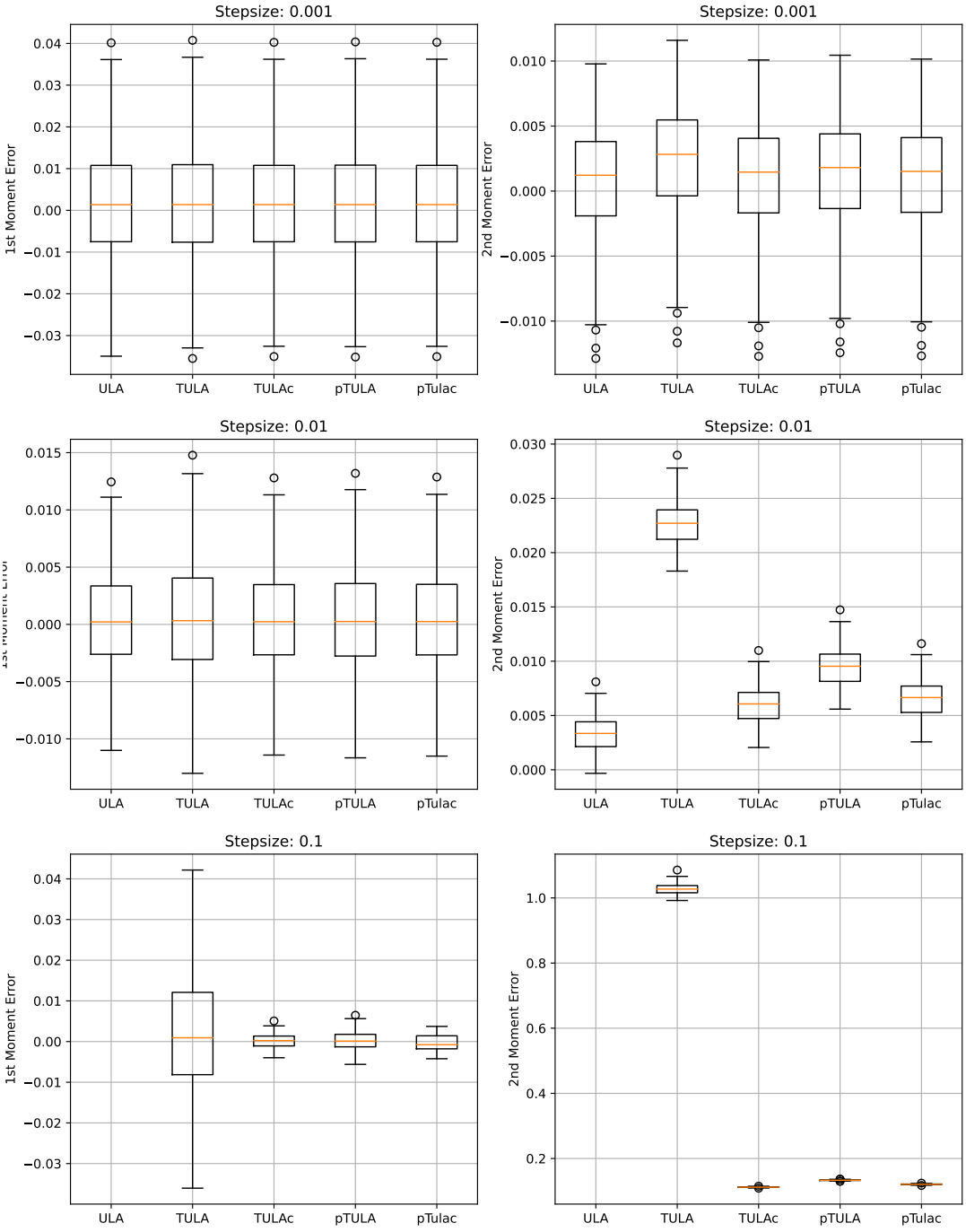

Figure 2: Boxplots of the error on the first and second moment for the Double Well in dimension 100 starting at $(10, 0^{\otimes 99})$ for different stepsizes

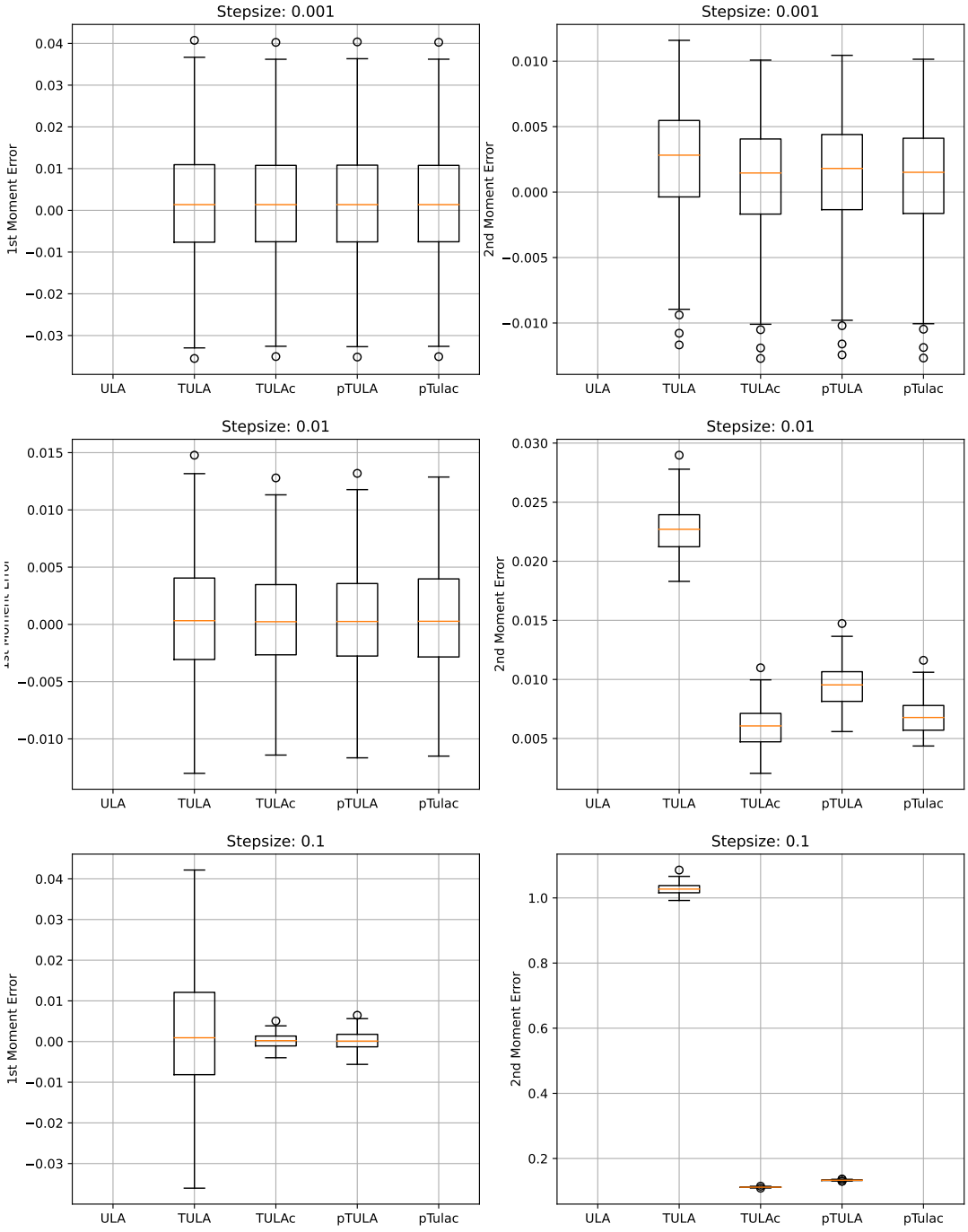

Figure 3: Boxplots of the error on the first and second moment for the Double Well in dimension 100 starting at $(100, 0^{\otimes 99})$ for different stepsizes

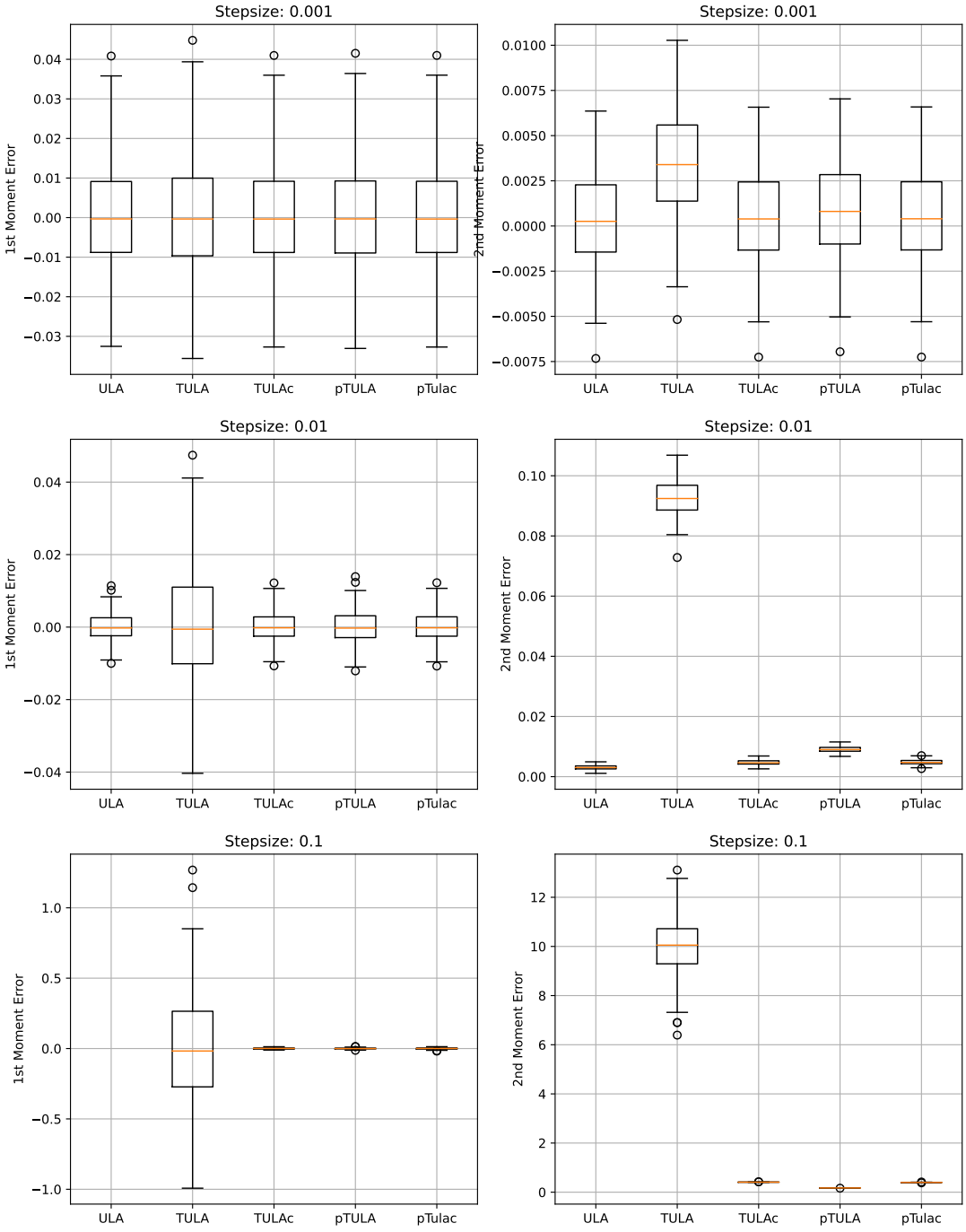

Figure 4: Boxplots of the error on the first and second moment for the Double Well in dimension 1000 starting at 0 for different stepsizes

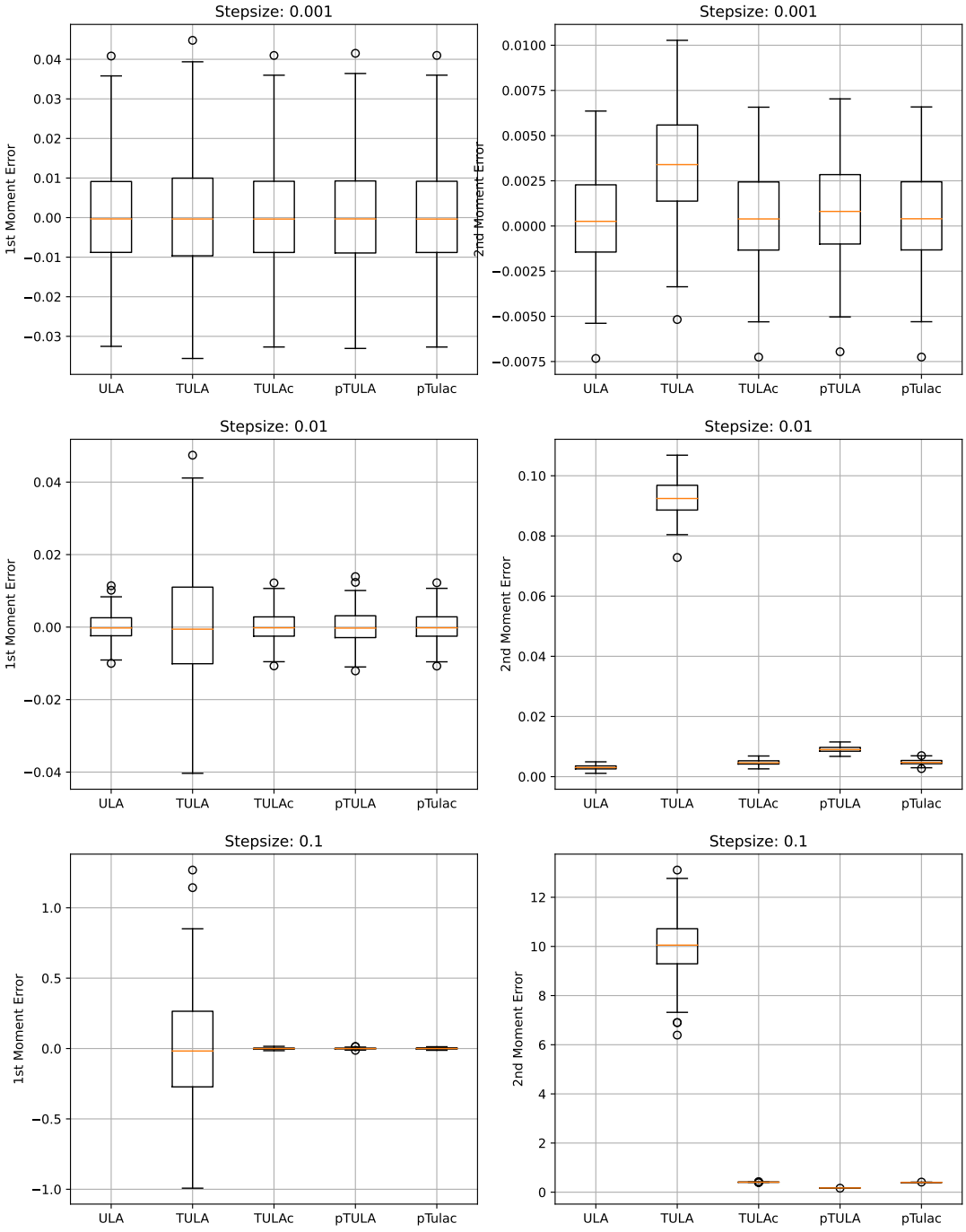

Figure 5: Boxplots of the error on the first and second moment for the Double Well in dimension 1000 starting at $(10, 0^{\otimes 999})$ for different stepsizes

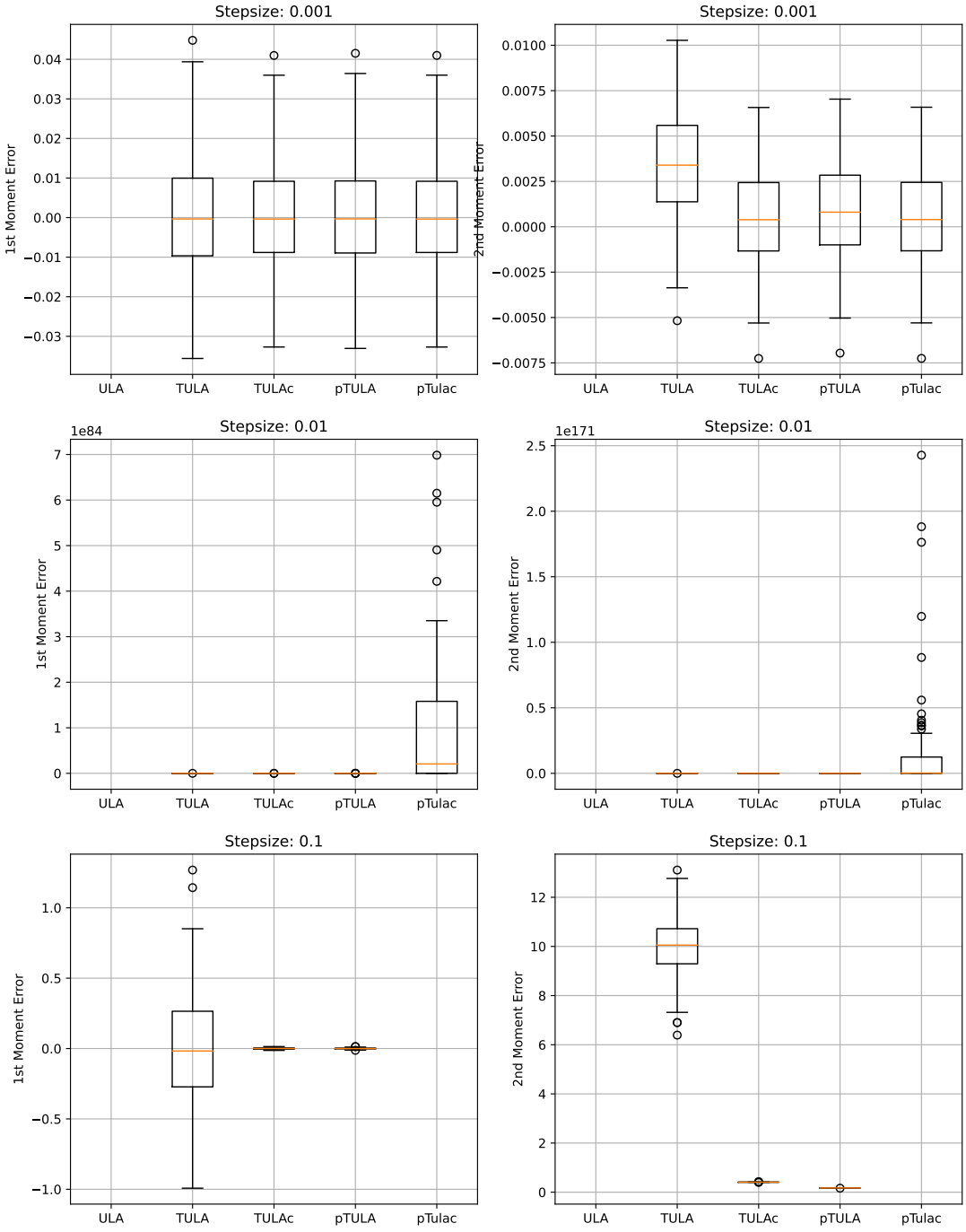

Figure 6: Boxplots of the error on the first and second moment for the Double Well in dimension 1000 starting at $(100, 0^{\otimes 999})$ for different stepsizes

References

[1] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. URL https://arxiv.org/abs/1412.6980.

[2] Oleksandr Borysenko and Maxim Byshkin. Coolmomentum: a method for stochastic optimization by langevin dynamics with simulated annealing. *Scientific Reports*, 11, 05 2021. doi: 10.1038/s41598-021-90144-3.

[3] Christian P Robert, George Casella, and George Casella. *Monte Carlo statistical methods*, volume 2. Springer, 1999.

[4] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. Strong and weak divergence in finite time of euler's method for stochastic differential equations with non-globally lipschitz continuous coefficients. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2130):1563–1576, dec 2010. doi: 10.1098/rspa.2010.0348. URL https://doi.org/10.1098%2Frspa.2010.0348.

[5] Nawaf Bou-Rabee, Martin Hairer, and Eric Vanden-Eijnden. Non-asymptotic mixing of the mala algorithm, 2010. URL https://arxiv.org/abs/1008.3514.

[6] Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996. ISSN 13507265. URL http://www.jstor.org/stable/3318418.

[7] Nawaf Bou-Rabee and Eric Vanden-Eijnden. Pathwise accuracy and ergodicity of metropolized integrators for sdes. *Communications on Pure and Applied Mathematics*, 63(5):655–696, 2010. doi: https://doi.org/10.1002/cpa.20306. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.20306.

[8] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *The Annals of Applied Probability*, 22(4):1611 – 1641, 2012. doi: 10.1214/11-AAP803. URL https://doi.org/10.1214/11-AAP803.

[9] Sotirios Sabanis. A note on tamed Euler approximations. *Electronic Communications in Probability*, 18(none):1 – 10, 2013. doi: 10.1214/ECP.v18-2824. URL https://doi.org/10.1214/ECP.v18-2824.

[10] Nicolas Brosse, Alain Durmus, Éric Moulines, and Sotirios Sabanis. The tamed unadjusted langevin algorithm, 2017. URL https://arxiv.org/abs/1710.05559.

[11] Dong-Young Lim and Sotirios Sabanis. Polygonal unadjusted langevin algorithms: Creating stable and efficient adaptive algorithms for neural networks, 2021. URL https://arxiv.org/abs/2105.13937.

[12] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2012.

[13] Sean P. Meyn and R. L. Tweedie. Stability of markovian processes iii: Foster–lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, 25(3):518–548, 1993. doi: 10.2307/1427522.

[14] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[15] Xuerong Mao. *Stochastic differential equations and applications*. Elsevier, 2007.

[16] Dominique Bakry, Ivan Gentil, Michel Ledoux, et al. *Analysis and geometry of Markov diffusion operators*, volume 103. Springer, 2014.