Contributions to non-convex stochastic optimization and reinforcement learning

Anas Barakat

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2021IPPAT030

Thèse de doctorat

Contributions to Non-Convex Stochastic Optimization and Reinforcement Learning

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à Télécom Paris

École doctorale n°626 Ecole Doctorale de l'Institut Polytechnique de Paris (ED IP Paris)
Spécialité de doctorat : Mathématiques et Informatique

Thèse présentée et soutenue à Palaiseau, le 07/12/2021, par

ANAS BARAKAT

Composition du Jury :

Sébastien Gadat Professor, Toulouse 1 Capitole University	Président, Rapporteur
Vivek Shripad Borkar Professor, Indian Institute of Technology (IIT) Bombay	Rapporteur
Robert Mansel Gower Assistant Professor, Télécom Paris (LTCI)	Examinateur
Niao He Assistant Professor, ETH Zurich	Examinateur
Edouard Pauwels Assistant Professor, Toulouse 3 Paul Sabatier University	Examinateur
Pascal Bianchi Professor, Télécom Paris (LTCI)	Directeur de thèse
Walid Hachem CNRS Research Director, Gustave Eiffel University	Co-directeur de thèse

À mes parents, Ali et Meryem, à mon frère Yassine

Remerciements

Tout d'abord, je souhaiterai remercier très chaleureusement mes directeurs de thèse Pascal et Walid. Dès ma deuxième année d'école à Télécom, en suivant vos cours, j'ai tout de suite été saisi par votre sérieux, votre rigueur et votre talent pédagogique. Merci pour votre présence, votre soutien et votre encadrement lors de mes années de thèse, même en temps de confinement. Merci pour toutes ces journées de riches échanges mathématiques au tableau (et sur zoom). Merci pour votre investissement dans cette thèse, votre rigueur, votre bienveillance et votre sympathie. J'ai eu la chance de découvrir le monde de la recherche scientifique grâce à vous, j'ai beaucoup appris à vos côtés et je vous en suis infiniment reconnaissant.

I would like to thank Prof. Vivek Borkar and Prof. Sébastien Gadat for having reviewed my thesis. It is really a great honor for me. Prof. Vivek Borkar, I took my first steps in stochastic approximation using your remarkable monograph. Prof. Sébastien Gadat, je voudrai vous remercier pour tous vos commentaires constructifs lors des différents rounds de relecture des articles de journaux constituant cette thèse et qui ont grandement contribué à les améliorer. I also would like to thank Prof. Robert Gower, Prof. Niao He and Prof. Edouard Pauwels for accepting to be part of my jury.

Je tiens aussi à remercier Olivier Fercoq pour ses cours d'optimisation et nos discussions mathématiques dont l'une a notamment inspiré une partie des calculs constituant le quatrième chapitre de cette thèse. Merci aussi à Robert Gower et Alexandre Gramfort qui m'ont inspiré et donné envie de continuer dans le domaine de l'optimisation grâce à leur cours de M2. Merci à Télécom Paris et à l'IMT d'avoir financé ma thèse via le programme "Futur et Ruptures".

Durant ma thèse, j'ai eu la chance de croiser le chemin de merveilleuses personnes. Merci à Adil de m'avoir passé le flambeau de l'approximation stochastique et de la meilleure des manières, j'ai pu suivre de près la fin de ta thèse pendant mon stage et tu as toujours aimablement répondu à mes questions, merci pour tes conseils avisés. A très bientôt j'espère ! Merci à Alex, j'ai eu la chance de découvrir le monde de la recherche avec Romain et toi à travers le monde fascinant des kernels dès mon stage ! Merci à tous les anciens de Télécom avec qui j'ai partagé de bons moments au détour d'une conversation scientifique, autour d'un verre à la Butte-aux-Cailles, d'un Five ou d'un resto à Paris 13, autour d'un repas au self, d'un Baby au foyer ou en conférence à travers le monde. Merci à Mathurin, Pierre A. (qui ont brillamment encadré mes TPs de M2 en Optim !), Alexandre G., Moussab, Pierre L., Kévin, Charles (et ses haches), Eugène, Hamid, Mastane, Anna, Robin, Jérôme-Alexis et Simon.

Je souhaiterai aussi remercier ici les doctorants du laboratoire qui ont adouci mon quotidien. Merci à Sholom pour nos échanges mathématiques stimulants et notre fructueuse collaboration. J'espère qu'il y en aura d'autres ! Merci à Guillaume (et Marie !) pour sa motivation communicative et nos pauses interminables à refaire le monde en semaine ou en weekend ! Nous avons aussi traversé une bonne partie de la pandémie ensemble à Saclay ! Merci aussi à Nidham, Pierre C., Amaury, Emile, Kamélia et Kimia. Merci également à la nouvelle génération saclaisienne qui porte le flambeau à son tour: Anass, Arturo, Dimitri, Emilia, Junjie, Luc, Rémi, Tamim et tous ceux que j'oublie pour les moments de partage ensemble.

Je remercie également mes amis en dehors de la thèse qui m'ont permis de m'évader vers d'autres cieux. Merci à Anas pour nos ballades parisiennes nocturnes improvisées et nos plans vacances. Merci aussi à Ali O. (et la mémorable pastèque niçoise) et Ali M., Simo et Réda pour nos dernières vacances partagées ensemble. A très bientôt pour de nouvelles aventures !

Enfin, je souhaiterai ici remercier toute ma famille pour son soutien et avec qui j'ai du temps à rattraper. J'ai une pensée pour mon grand-père qui m'a toujours encouragé à pousser mes études le plus loin possible et à continuer dans cette voie de la recherche scientifique. Je tiens particulièrement à remercier mes parents et mon frère pour leur soutien indéfectible tout au long de ma vie, y compris dans les moments les plus difficiles de cette thèse. Vous avez toujours su être là pour moi. Vous m'avez donné tous les moyens pour réussir et je vous en suis infiniment reconnaissant.

Abstract

This thesis is focused on the convergence analysis of some popular stochastic approximation methods in use in the machine learning community with applications to optimization and reinforcement learning. The first part of the thesis is devoted to a popular algorithm in deep learning called ADAM used for training neural networks. This variant of stochastic gradient descent is more generally useful for finding a local minimizer of a function. Assuming that the objective function is differentiable and non-convex, we establish the convergence of the iterates in the long run to the set of critical points under a stability condition in the constant stepsize regime. Then, we introduce a novel decreasing stepsize version of ADAM. Under mild assumptions, it is shown that the iterates are almost surely bounded and converge almost surely to critical points of the objective function. Finally, we analyze the fluctuations of the algorithm by means of a conditional central limit theorem.

In the second part of the thesis, in the vanishing stepsizes regime, we generalize our convergence and fluctuations results to a stochastic optimization procedure unifying several variants of the stochastic gradient descent such as, among others, the stochastic heavy ball method, the Stochastic Nesterov Accelerated Gradient algorithm (S-NAG), and the widely used ADAM algorithm. We conclude this second part by an avoidance of traps result establishing the non-convergence of the general algorithm to undesired critical points, such as local maxima or saddle points. Here, the main ingredient is a new avoidance of traps result for non-autonomous settings, which is of independent interest. A chapter of this thesis is also devoted to some non-asymptotic guarantees under a clipping of the effective stepsizes. We control here the expected norm of the gradient along the iterations in the stochastic setting. We also establish convergence rates in terms of the function value gap using the Kurdyka-Łojasiewicz property.

Finally, the last part of this thesis which is independent from the two previous parts, is concerned with the analysis of a stochastic approximation algorithm for reinforcement learning. In this last part, we propose an analysis of an online target-based actor-critic algorithm with linear function approximation in the discounted reward setting. Our algorithm uses three different timescales: one for the actor and two for the critic. Instead of using the standard single timescale temporal difference (TD) learning algorithm as a critic, we use a two timescales target-based version of TD learning closely inspired from practical actor-critic algorithms implementing target networks. First, we establish asymptotic convergence results for both the critic and the actor under Markovian sampling. Then, we provide a finite-time analysis showing the impact of incorporating a target network into actor-critic methods.

Résumé

Cette thèse est centrée autour de l'analyse de convergence de certains algorithmes d'approximation stochastiques utilisés en machine learning appliqués à l'optimisation et à l'apprentissage par renforcement. La première partie de la thèse est dédiée à un célèbre algorithme en apprentissage profond appelé ADAM, utilisé pour entraîner des réseaux de neurones. Cette célèbre variante de la descente de gradient stochastique est plus généralement utilisée pour la recherche d'un minimiseur local d'une fonction. En supposant que la fonction objective est différentiable et non convexe, nous établissons la convergence des itérées au temps long vers l'ensemble des points critiques sous une hypothèse de stabilité dans le régime des pas constants. Ensuite, nous introduisons une nouvelle variante de l'algorithme ADAM à pas décroissants. Nous montrons alors sous certaines hypothèses réalistes que les itérées sont presque sûrement bornées et convergent presque sûrement vers des points critiques de la fonction objective. Enfin, nous analysons les fluctuations de l'algorithme par le truchement d'un théorème central limite conditionnel.

Dans la deuxième partie de cette thèse, dans le régime des pas décroissants, nous généralisons nos résultats de convergence et de fluctuations à une procédure d'optimisation stochastique unifiant plusieurs variantes de descente de gradient stochastique comme la méthode de la boule pesante, l'algorithme stochastique de Nesterov accéléré ou encore le célèbre algorithme ADAM, parmi d'autres. Nous concluons cette partie par un résultat d'évitement de pièges qui établit la non convergence de l'algorithme général vers des points critiques indésirables comme les maxima locaux ou les points-selles. Ici, le principal ingrédient est un nouveau résultat indépendant d'évitement de pièges pour un contexte non-autonome. Un chapitre de cette thèse est également consacré à des garanties non-asymptotiques pour une large classe d'algorithmes adaptatifs sous une hypothèse de clipping des pas effectifs. Nous établissons en particulier une vitesse de convergence de la norme du gradient le long des itérations (ou de son espérance) dans les cas déterministe et stochastique. Nous montrons également des vitesses de convergence des valeurs de la fonction objective en utilisant l'hypothèse faible de Kurdyka-Łojasiewicz.

Enfin, la dernière partie de cette thèse qui est indépendante des deux premières parties est dédiée à l'analyse d'un algorithme d'approximation stochastique pour l'apprentissage par renforcement. Dans cette dernière partie, dans le cadre des processus décisionnels de Markov avec critère de récompense γ-pondéré, nous proposons une analyse d'un algorithme acteur-critique en ligne intégrant un réseau cible et avec approximation de fonction linéraire. Notre algorithme utilise trois échelles de temps distinctes: une échelle pour l'acteur et deux autres pour la critique. Au lieu d'utiliser l'algorithme de différence temporelle (TD) standard à une échelle de temps, nous utilisons une version de l'algorithme TD à deux échelles de temps intégrant un réseau cible inspiré des algorithmes acteur-critique utilisés en pratique. Tout d'abord, nous établissons

des résultats de convergence pour la critique et l'acteur sous échantillonage Markovien. Ensuite, nous menons une analyse à temps fini montrant l'impact de l'utilisation d'un réseau cible sur les méthodes acteur-critique.

Contents

Notation

$:=$	Equal by definition
$\mathbb{R}$	Set of real numbers
$\mathbb{R}_+$	Set of nonnegative real numbers
$\mathbb{R}_+^*$	Set of positive real numbers
$\mathbb{N}$	Set of integers: $\{0, 1, 2, \ldots\}$
$A \cup B$	Set union between the sets A and B
$A \cap B$	Set intersection between the sets A and B

If u, v are vectors of $\mathbb{R}^d$ where $d \in \mathbb{N}$,

$u \odot v$	Coordinatewise product vector with coordinates $u_i v_i$
$u^{\odot 2}$	Coordinatewise square vector with coordinates u_i^2
$\frac{u}{v}$	Coordinatewise quotient vector with coordinates u_i/v_i when $v_i \neq 0$ for every $i \in \mathbb{R}^d$
$\lvert u \rvert$	Vector whose i-th coordinate is given by $\lvert u_i \rvert$
$\sqrt{\lvert u \rvert}$	Vector whose i-th coordinate is given by $\sqrt{\lvert u_i \rvert}$
$\lVert \cdot \rVert$	Standard Euclidean norm
$\lVert x \rVert_v^2$	$:= \sum_i v_i x_i^2$ for every $x \in \mathbb{R}^d$, $v \in (0, +\infty)^d$.
$\mathsf{d}(z, A)$	$:= \inf\{\lVert z - z' \rVert : z' \in A\}$ if $z \in \mathbb{R}^d$ and A is a non-empty subset of $\mathbb{R}^d$
I_n	Identity matrix of size $n \times n$
$M^\top$	Transpose of matrix M
$\mathbb{1}_A$	Characteristic function of a set A, i.e., the function equal to one on that set and to zero elsewhere

1 Introduction

1.1 Motivations

1.1.1 Non-convex stochastic optimization

Nowadays, machine learning is becoming more and more pervasive in society through several applications ranging from machine translation, speech and image recognition to online advertising and even robotics, to name a few. More advanced applications such as personalized medicine and autonomous vehicles in critical societal domains such as healthcare and transportation are also on the horizon. One of the main pillars supporting machine learning is the mathematical field of optimization. Indeed, typically, the numerical implementation of machine learning methods requires the minimization of complex loss functions measuring the inadequacy of a model to available data. Optimal parameters of the model resulting from this optimization step are then used to make decisions based on yet unseen data.

During the last decades, the access to a massive amount of data and the increase in the computing power have revolutionized the field of machine learning. This evolution led to a renewed interest in deep learning and opened the way for a fast development of deep neural networks fueling tremendous advances in machine learning. As a consequence, several new challenges emerged from optimization problems arising in machine learning. First, the large available volume of data induce challenging large-scale optimization problems (see for e.g., Bottou et al. (2018) for a review). This challenge stimulated the design of stochastic algorithms capable of learning online. Second, the success of deep learning brought to the forefront optimization problems which are typically non-convex due to the non-convexity inherited from neural networks as functions of their weights.

More formally, in this thesis, we consider the problem of finding a local minimizer of the expectation $F(x) := \mathbb{E}(f(x, \xi))$ w.r.t. $x \in \mathbb{R}^d$, where $f(\,.\,, \xi)$ is a possibly non-convex function depending on some random variable (r.v.) ξ. The distribution of ξ is assumed unknown, but revealed online by the observation of independent and identically distributed (iid) copies $(\xi_n : n \geq 1)$ of the r.v. ξ. For instance, this general formulation encompasses our motivating learning example: the parameter x then refers to the weights of the neural network at stake whereas the random variable ξ models the data and the function f can be seen as a loss function quantifying the goodness of the prediction model parameterized by x using the data point ξ whereas the function F represents the training loss of the model. The formulation of this stochastic non-convex optimization problem goes beyond this motivating learning problem and finds other applications in diverse sectors such as energy or finance.

1.1.2 From stochastic gradient descent to adaptive gradient methods

In this section, we will gradually introduce the stochastic optimization algorithms at the heart of this thesis, namely adaptive gradient methods.

SGD. Historically introduced by Robbins and Monro (1951), Stochastic Gradient Descent (SGD) is the most classical algorithm to search for a local minimizer of the function F. Given the formulation of our optimization problem and assuming that the function $f(\cdot, \xi)$ is differentiable for every fixed value of ξ, the iterates of SGD initialized at some point $x_0 \in \mathbb{R}^d$ can be written as follows:

$$x_{n+1} = x_n - \gamma_n \nabla f(x_n, \xi_{n+1}) \,, \tag{1.1}$$

where (γ_n) is a sequence of positive stepsizes and $\nabla f(x, \xi)$ denotes the gradient of the mapping $x \mapsto f(x, \xi)$ w.r.t. x for every fixed value of ξ. We briefly highlight that neural networks may involve some activation functions with points where the aforementioned mapping is nondifferentiable (such as the ReLU function, i.e., the positive part function). This setting is beyond the scope of this thesis and we refer to the recent works of Davis et al. (2020); Majewski et al. (2018); Bolte and Pauwels (2021); Bianchi et al. (2020) for this specific case. Nevertheless, smooth activation functions such as the widely used sigmoid function or the smooth ReLU still lead to the differentiable setting under study in this thesis.

Instead of computing the full gradient of the objective function F (which may not even be possible), SGD uses a single stochastic gradient at each iteration. As such, this canonical algorithm is particularly suitable for large-scale machine learning problems involving a large number of data samples and has become the workhorse for many machine learning problems and especially for deep learning. In this method, the update rule (1.1) depends on the parameter γ_n called the *learning rate*, which is generally assumed constant or vanishing.

Although widely used, this algorithm has at least two limitations. First, the choice of the learning rate is generally difficult; large learning rates result in large fluctuations of the estimate, whereas small learning rates induce slow convergence. Second, a common learning rate is used for every coordinate despite the possible discrepancies in the values of the gradient vector's coordinates.

Adaptive stepsizes. To alleviate these limitations, an idea which has made its way into the machine learning community is that of adjusting the learning rate coordinate-wise, as a function of the past values of the squared gradient vectors' coordinates. This modification can be seen as a diagonal preconditioning of the stochastic gradient in SGD based on past observed gradients. The independent works of Duchi et al. (2011) and McMahan and Streeter (2010) in the context of online convex optimization led the way to a new class of algorithms that are sometimes referred to as "adaptive [1] gradient methods". As proposed by Duchi et al. (2011), AdaGrad consists of dividing the learning rate by the square root of the sum of previous gradients squared componentwise. The idea was to give larger learning rates to highly informative but infrequent features instead of using a fixed predetermined schedule. This is particularly relevant in applications such as click through rate prediction for online advertising and text classification where many features only occur rarely with only a few number of non-zero features while few occur very often. We refer to (Duchi et al., 2011, Section 1.3) and (McMahan and Streeter, 2010, Section 1.2) for examples showing how adaptive methods can outperform standard methods when gradients are sparse.

[1] We follow their terminology for the word "adaptive" in this thesis, we bring to the attention of the reader that this same word has been used in the literature in different contexts for different purposes.

However, in practice, the division by the cumulative sum of squared gradients may generate small learning rates, thus freezing the iterates too early. Several works proposed heuristical ways to set the learning rates using a less aggressive policy. Tieleman and Hinton (2012) introduced an unpublished, yet popular, algorithm referred to as RmsProp where the cumulative sum used in AdaGrad is replaced by a moving average of squared gradients.

Momentum. Besides adaptive stepsizes, another popular modification of vanilla SGD is the use of momentum. Introduced by the seminal work of Polyak (1964) (see Section 2 therein) in the deterministic setting, the heavy ball method augments the standard gradient descent method with an additive inertial term corresponding to the difference between two successive iterates of the algorithm. This inertial term which was later called momentum in the modern machine learning community (see for e.g., Sutskever et al. (2013)) led to accelerated convergence rates of the heavy ball method in comparison to deterministic gradient descent for the optimization of twice differentiable strongly convex functions. Later, Nesterov (1983) proposed the Nesterov accelerated gradient descent method for convex optimization with an "optimal" convergence rate of $O(1/k^2)$ of the value function gap, outperforming gradient descent. Since the seminal works of Polyak and Nesterov, momentum methods hold the promise of acceleration and several recent works shed the light on this acceleration phenomenon (Wibisono et al., 2016; Wilson et al., 2021). Although the acceleration benefits of momentum methods in the deterministic optimization setting do not necessarily carry over to the stochastic non-convex optimization setting [2] as also recently advocated by Kidambi et al. (2018), the seminal work of Nesterov together with the renewed interest in first order optimization algorithms for large-scale machine learning fostered the design of similar methods in the context of non-convex stochastic optimization (see for e.g. Sutskever et al. (2013); Gadat et al. (2018)). Apart from possible acceleration, we point out that momentum has also a smoothing effect on the stochastic gradients used in the algorithm.

Adam. We are now ready to introduce the most popular algorithm among the family of adaptive gradient methods we have introduced so far: Adam (for Adaptive Momentum estimation). Proposed by Kingma and Ba (2015), Adam combines the advantages of both AdaGrad, RmsProp and momentum [3] methods. The notorious algorithm thus combines the assets of inertial methods with an adaptive per-coordinate learning rate selection for automatic tuning. Since 2015, the popular Adam algorithm has become widely used in deep learning applications and implemented in massively used deep learning libraries such as TensorFlow (Abadi et al., 2016) and PyTorch (Paszke et al., 2019). As originally proposed in Kingma and Ba (2015), at each timestep $n \in \mathbb{N}$, Adam produces a triplet of iterates $(x_n, m_n, v_n) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$ using the following recursions:

$$
(\text{Adam}) \begin{cases} m_{n+1} &= \alpha m_n + (1-\alpha)\nabla f(x_n, \xi_{n+1}) & \text{(1.2a)} \\ v_{n+1} &= \beta v_n + (1-\beta)\nabla f(x_n, \xi_{n+1})^{\odot 2} & \text{(1.2b)} \\ \hat{m}_{n+1} &= m_{n+1}/(1-\alpha^{n+1}) & \text{(1.2c)} \\ \hat{v}_{n+1} &= v_{n+1}/(1-\beta^{n+1}) & \text{(1.2d)} \\ x_{n+1} &= x_n - \gamma\, \hat{m}_{n+1}/(\varepsilon + \sqrt{\hat{v}_{n+1}}) \ , & \text{(1.2e)} \end{cases}
$$

[2] see though Jain et al. (2018) for an improvement over SGD for least squares regression (in stochastic convex optimization) with their Accelerated SGD algorithm which differs from the stochastic counterparts of the heavy ball method and Nesterov's accelerated gradient method.

[3] sometimes dubbed *inertial*

where $\alpha, \beta \in [0, 1)$ are exponential decay rates hyperparameters for the moving averages (typically $\alpha = 0.9$ and $\beta = 0.999$), $\gamma \in \mathbb{R}_+^*$ is the stepsize, $\varepsilon \in \mathbb{R}_+^*$ is a constant introduced for numerical stability (typically $\varepsilon = 10^{-8}$) and the iterates are initialized with $m_0 = v_0 = 0$ and some $x_0 \in \mathbb{R}^d$. We bring to the attention of the reader that all the operations on vectors are coordinatewise throughout this thesis. If x, y are two vectors on $\mathbb{R}^d$, we denote by $x \odot y$, $x^{\odot 2}$, x/y, $|x|$, $\sqrt{|x|}$ the vectors on $\mathbb{R}^d$ whose i-th coordinates are respectively given by $x_i y_i$, x_i^2, x_i/y_i, $|x_i|$, $\sqrt{|x_i|}$.

As previously described, the algorithm uses momentum as can be seen in steps (1.2a),(1.2e), and computes an adaptive learning rate $(\gamma/(\varepsilon + \sqrt{\hat{v}_{n+1}})$ as can be appreciated in the recursions (1.2b) and (1.2e). Finally, as can be observed in steps (1.2c) and (1.2d) above, the algorithm includes a so-called *bias correction* step. Acting on the current estimate of the gradient vector, this step is especially useful during the early iterations. We shall provide some comments on this step later on in this introduction when stating our contributions and a more in-depth explanation of this step can be found in the subsequent chapters (see Remark 1 in Chapter 2). We highlight that adaptive gradient methods are still first order optimization algorithms in the sense that they only have access to stochastic gradients without resorting to more computationally demanding information such as Hessians. Compared to SGD, adaptive gradient methods have similar computational complexity while requiring less hyperparameter tuning. One additional interesting feature of adaptive gradient methods which is also true for Newton-like algorithms, is that a multiplication of the objective function by a constant automatically impacts the (adaptive) learning rate which is divided by the same constant (see Eq. (1.2e)).

A major part of this thesis is dedicated to the analysis of the algorithms belonging to the class of adaptive gradient methods we have introduced. In particular, AdaGrad, RMSProp and Adam all belong to this class. As we will expand on it later on, our analysis will also encompass stochastic momentum methods including stochastic Nesterov accelerated gradient.

Challenges. As opposed to the vanilla Stochastic Gradient Descent (SGD), the study of such algorithms is more elaborate, for three reasons. First, the update of the iterates involves a so-called *momentum* term, or inertia, which has the effect of "smoothing" the increment between two consecutive iterates. Second, as far as adaptive algorithms are concerned, the update also depends on some additional variable (*a.k.a.* the (adaptive) learning rate) computed online as a function of the history of the computed gradients. As a consequence of this special learning rate, the update rule of the algorithms is not linear as a function of the stochastic gradients. The momentum term and the learning rate together endow the algorithms with a "memory" induced by the use of past stochastic gradient information. In contrast, vanilla SGD is memoryless. Third, the update equation at the time index n is likely to depend on n, making these systems inherently *non-autonomous*. Indeed, apart from the stepsizes depending explicitely on n, the update rules of the algorithms can also feature an additional explicit dependence on n (see (1.2c) and (1.2d) above).

1.1.3 Actor-critic methods in Reinforcement Learning

In this section, we introduce the Reinforcement Learning (RL, Sutton and Barto (2018)) problem we deal with in the last part of this manuscript. This part is independent from the rest of the thesis.

RL has witnessed a huge success in a wide range of applications such as game playing (Mnih et al., 2015), robotic manipulation (Gu et al., 2017; Levine et al., 2018), dialogue generation in natural language processing (Li et al., 2016), data centre cooling regulation (Lazic et al., 2018), traffic signal control (Prashanth and Bhatnagar, 2010, 2011) even in large-scale urban networks (Chen et al., 2020), self-driving cars (Kiran et al., 2021) or clinical decision support in critical care (Liu et al., 2020), to name a few.

In this thesis, we are concerned with sequential decision making problems under uncertainty formulated as Markov Decision Processes (MDP) (Puterman, 2014). In this model, an agent learns how to act optimally by interacting with a possibly unknown environment via trial and error. [4]

We denote by $\mathcal{S} = \{s_1, \cdots, s_n\}$ the finite set of states of the environment and $\mathcal{A}$ the finite set of possible actions the agent can execute. We use the notation $\mathcal{P}(\mathcal{X})$ for the set of probability measures on the set $\mathcal{X}$ where $\mathcal{X}$ can either be $\mathcal{S}$, $\mathcal{A}$ or $\mathcal{S} \times \mathcal{A}$. Let $p : \mathcal{S} \times \mathcal{A} \to \mathcal{P}(\mathcal{S})$ be the state transition probability kernel which gives the probability of moving to the next state given the current state and action. The immediate reward [5] function $R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ provides for every $s \in \mathcal{S}, a \in \mathcal{A}$ the single-stage expected reward $R(s, a)$ when action $a \in \mathcal{A}$ is executed in state $s \in \mathcal{S}$. A randomized stationary policy, which we will simply call a policy, is a mapping $\pi : \mathcal{S} \to \mathcal{P}(\mathcal{A})$ specifying for each $s \in \mathcal{S}, a \in \mathcal{A}$ the probability $\pi(a|s)$ of selecting action a in state s. This policy describes the agent's behavior strategy in the MDP model. At each time step $t \in \mathbb{N}$, the RL agent in a state $S_t \in \mathcal{S}$ executes an action $A_t \in \mathcal{A}$ with probability $\pi(A_t|S_t)$, transitions into a state $S_{t+1} \in \mathcal{S}$ with probability $p(S_{t+1}|S_t, A_t)$ and observes a real random reward R_{t+1} (which we will suppose to be bounded). We denote by $\mathbb{P}_{\rho,\pi}$ the probability distribution of the Markov chain (S_t, A_t) issued from the MDP controlled by the policy π with initial state distribution ρ. The notation $\mathbb{E}_{\rho,\pi}$ refers to the associated expectation. We will use $\mathbb{E}_\pi$ whenever there is no dependence on ρ. The sequence (R_t) is such that (s.t.) $\mathbb{E}_\pi[R_{t+1}|S_t, A_t] = R(S_t, A_t)$. The objective is then to find an optimal policy maximizing the expected cumulative future rewards [6]:

$$\max_\pi J(\pi) := \mathbb{E}_{\rho,\pi}\left[\sum_{t=0}^{\infty} \gamma^t R_{t+1}\right], \tag{1.3}$$

where the discount factor $\gamma \in (0, 1)$ favors immediate rewards over delayed ones. We consider the setting where the dynamic of the environment is not explicitely known which renders the computation of the objective function $J(\pi)$ intractable. This dynamic will be revealed online over time thanks to the observation of the environment's successive states following the executed actions. This is the so-called *model free* approach.

The first fundamental class of RL methods consists of value-based methods using the so-called value function (respectively action-state value function) which quantifies how good is each state (or state-action pair). Given a policy π, the value function $V_\pi : \mathcal{S} \to \mathbb{R}$ and the action-value function (also called Q-function) $Q_\pi : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ are defined for

[4] Following the standard terminology of Sutton and Barto (2018) and the RL community, we also use the the terms *agent, environment, action* instead of *controller, controlled system* and *control signal* as one may encounter in control theory.

[5] We use *reward* as in the RL literature instead of *cost*.

[6] Other performance criteria such as the average reward and the total reward exist in the literature (see, for e.g., Puterman (2014)). In this thesis, we focus on the expected discounted return.

every state $s \in \mathcal{S}$, action $a \in \mathcal{A}$ by:

$$V_\pi(s) := \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t R_{t+1} | S_0 = s \right] \quad \text{and} \quad Q_\pi(s,a) := \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t R_{t+1} | S_0 = s, A_0 = a \right].$$

The optimal value function $V_* : \mathcal{S} \to \mathbb{R}$ and Q-function $Q_* : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ can then be defined for every $s \in \mathcal{S}$, $a \in \mathcal{A}$ by: $V_*(s) := \max_\pi V_\pi(s)$ and $Q_*(s,a) := \max_\pi Q_\pi(s,a)$ .Value-based methods rely on the fundamental property of the optimal policy π_* stipulating that $\pi_*(s) = \arg\max_{a \in \mathcal{A}} Q_*(s,a)$. Therefore, in order to compute the optimal policy, one looks for the optimal Q-value function without requiring the knowledge of the reward or transition probabilities. Since this function Q_* satisfies a dynamic programming (Bellman) equation and can then be seen as the fixed point of the so-called Bellman equation, the function Q_* can be estimated using a stochastic algorithm for fixed point search. This leads to one of the most famous RL algorithms belonging to the class of value-based methods as proposed by Watkins (1989): Q-learning. We refer for instance to (Borkar and Chandak, 2021, Section 2.1) or (Avrachenkov et al., 2021, Section 2.1) for a nice self-contained derivation of the algorithm.

A second class of methods consists of the so-called policy-based methods. In these methods, one directly searchs the optimal policy π_* . The main representant of this class is the policy gradient method. The idea is to parameterize the policy π by a vector parameter $\theta \in \mathbb{R}^d$ (i.e. consider a family of policies $\{\pi_\theta : \theta \in \mathbb{R}^d\}$) and directly maximize the objective function $J(\pi_\theta)$ by performing a stochastic gradient ascent. For this, the gradient of the objective function is then given by the so-called policy gradient theorem[7] (Sutton et al., 1999) as follows:

$$\nabla J(\theta) = \frac{1}{1-\gamma} \cdot \mathbb{E}_{(\tilde{S},\tilde{A}) \sim \mu_{\rho,\theta}} [\Delta_{\pi_\theta}(\tilde{S},\tilde{A}) \nabla \ln \pi_\theta(\tilde{A}|\tilde{S})] , \tag{1.4}$$

for every $\theta \in \mathbb{R}^d$, where the advantage function [8] $\Delta_\pi : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is defined for every policy π and every $(s,a) \in \mathcal{S} \times \mathcal{A}$ by $\Delta_\pi(s,a) := Q_\pi(s,a) - V_\pi(s)$, the initial state S_0 follows an initial probability distribution ρ over states [9] and the gradient is w.r.t. the parameter θ (i.e., the gradient of the function $\theta \mapsto \ln \pi_\theta$). Here, the couple of r.v.s $(\tilde{S},\tilde{A})$ follows the discounted state-action occupancy measure $\mu_{\rho,\theta} \in \mathcal{P}(\mathcal{S} \times \mathcal{A})$ defined for all $(s,a) \in \mathcal{S} \times \mathcal{A}$ by:

$$\mu_{\rho,\theta}(s,a) := d_{\rho,\theta}(s)\, \pi_\theta(a|s) \quad \text{where} \quad d_{\rho,\theta}(s) := (1-\gamma) \sum_{t=0}^{\infty} \gamma^t \mathbb{P}_{\rho,\pi_\theta}(S_t = s) \tag{1.5}$$

is a probability measure over the state space $\mathcal{S}$ known as the discounted state-occupancy measure. We refer the reader to Section 5.3 of the self-contained Chapter 5 for more precisions and rigorous regularity conditions under which this theorem holds. In policy gradient methods, the unknown advantage function Δ_π (or Q-function Q_π) is estimated via Monte Carlo simulation using trajectories generated by the MDP (Monte Carlo

[7] An older version of the policy gradient theorem leading to the so-called REINFORCE algorithm was proposed by Williams (1992). Here, we focus our exposition on this version which is more relevant for our purposes.

[8] We use here the notation Δ_π instead of the more common A_π to avoid notational collision with the sequence of actions.

[9] Note here that in its original form, the theorem uses Q_π instead of Δ_π in Eq. (1.4). This replacement can be straightforwardly shown using the identity "$\nabla \ln \pi_\theta = \nabla \pi_\theta / \pi_\theta$" and noticing that the function V_π does not depend on actions.

rollouts). Denoting by $\hat{\Delta}_{\pi_{\theta_t}}$ such an estimate of $\Delta_{\pi_{\theta_t}}$ at time $t \in \mathbb{N}$ and given Eq. (1.4) and a sequence of positive stepsizes (α_t), we obtain the following recursion:

$$\theta_{t+1} = \theta_t + \alpha_t \frac{1}{1-\gamma} \hat{\Delta}_{\pi_{\theta_t}}(\tilde{S}_t, \tilde{A}_t) \nabla \ln \pi_{\theta_t}(\tilde{A}_t | \tilde{S}_t) \,, \tag{1.6}$$

where the couple $(\tilde{S}_t, \tilde{A}_t)$ is generated following an online procedure (that we do not explicit here for conciseness, see Section 5.4 for details) guaranteeing that the couple of r.v.s will "asymptotically" follow the unknown distribution $\mu_{\rho,\theta}$. One of the shortcomings of such methods is the large variance induced by the gradient estimators. Moreover, a new gradient is estimated regardless of previous estimates as the policy evolves.

Finally, a third class of methods combines value-based and policy-based methods. This is the class of actor-critic methods (Barto et al., 1983; Konda and Borkar, 1999; Konda and Tsitsiklis, 2003) consisting of two coupled iterative algorithms. The first one called the actor updates the current policy using a stochastic gradient ascent strategy with an increment depending on the advantage function (and actually only on the value function V_{π_θ}, see the definition of Δ_{π_θ} and details in Section 5.4.1). This last function being unknown, a second incremental (value-based) algorithm called the critic estimates its value all along the iterations. The critic evolves on a faster timescale using larger stepsizes than for the actor, thereby simulating the effect of nested loops where the estimation of the critic takes place in an inner loop useful for the actor evolving in an outer loop.

In the case of complex MDPs with a large number of states and/or actions, estimating the value at each state-action pair to fill the entire table of state-action values becomes intractable because of the "curse of dimensionality" as computing solutions to dynamic programming equations (that value functions satisfy) is infeasible. Instead of tabular (or look-up table) methods, in the RL literature, we modestly approximate the exact value function by another function belonging to a parameterized family of functions. A simple linear parameterization consists in looking for the approximate value function as a linear combination of some predetermined basis (or feature) functions. For every $\theta \in \mathbb{R}^d$, the state-value function V_{π_θ} is then approximated for every state $s \in \mathcal{S}$ by a linear function of carefully chosen feature vectors as follows:

$$V_{\pi_\theta}(s) \approx V_\omega(s) = \omega^T \phi(s) = \sum_{i=1}^{m} \omega_i \phi^i(s) \,, \tag{1.7}$$

where $\omega = (\omega_1, \cdots, \omega_m)^T \in \mathbb{R}^m$ for some integer $m \ll |\mathcal{S}|$ and $\phi(s) = (\phi^1(s), \cdots, \phi^m(s))^T$ is the feature vector of the state $s \in \mathcal{S}$. We highlight though that the best practical performances are obtained by using nonlinear parameterizations based on deep neural networks. Typically, a neural network takes the state-action pair as an input and outputs the desired approximated value function V_ω. In the case of actor-critic methods, the critic updates the parameters of this neural network whereas a similar neural network parameterization is used for the actor. The textbooks of Sutton and Barto (2018); Bertsekas and Tsitsiklis (1996); Szepesvári (2010) provide nice introductions to RL theory with further details.

Actor-critic methods integrating target networks have exhibited a stupendous empirical success in deep reinforcement learning (Heess et al., 2015; Lillicrap et al., 2016; Fujimoto et al., 2018; Haarnoja et al., 2018). However, if standard actor-critic methods have been well-studied in the literature (see, for e.g., Konda and Borkar (1999); Konda and

Tsitsiklis (2003); Bhatnagar et al. (2009)), a theoretical understanding of the use of target networks in actor-critic methods is largely missing in the literature. We will devote the last part of this thesis to the study of an actor-critic method incorporating a target network in the linear function approximation setting. We will briefly introduce our actor-critic algorithm together with our contributions in Section 1.3 below and we will defer its precise mathematical presentation to Section 5.4 in which the reader can also find an exposition of the idea of target network from the RL literature.

1.1.4 Theoretical context: Stochastic Approximation

The stochastic algorithms considered in this thesis fall under the umbrella of general stochastic approximation algorithms of the form:

$$x_{n+1} = x_n + \gamma_{n+1} h(x_n, \xi_{n+1}) \,, \quad x_0 \in \mathbb{R}^d \,, \tag{1.8}$$

where h is a map from $\mathbb{R}^d \times \mathbb{R}^k$ to $\mathbb{R}^d$, (ξ_n) is a sequence of r.v.s taking their values in $\mathbb{R}^k$ and (γ_n) is a vanishing sequence of positive reals s.t. $\sum_n \gamma_n = +\infty$ and say $\sum_n \gamma_n^2 < +\infty$. Here, the sequence of r.v.s (ξ_n) evolves as follows. Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and considering the σ-field $\mathcal{F}_n$ generated by the r.v.s $x_0, \xi_1, \cdots, \xi_n$, we assume that there exist a family of transition probabilities $\{\Pi_x : \mathbb{R}^k \times \mathcal{B}(\mathbb{R}^k) \to [0,1] \,, x \in \mathbb{R}^d\}$[10] on $\mathbb{R}^k$ s.t. for all $n \in \mathbb{N}$, for all $B \in \mathcal{B}(\mathbb{R}^k)$, almost surely,

$$\mathbb{P}[\xi_{n+1} \in B | \mathcal{F}_n] = \Pi_{x_n}(\xi_n; B) \,, \tag{1.9}$$

where $\mathbb{P}[\xi_{n+1} \in B | \mathcal{F}_n]$ is the conditional probability of the event $\{\xi_{n+1} \in B\}$ given $\mathcal{F}_n$. This property means that the sequence (ξ_n) is a Markov chain controlled by the parameter x. Suppose now that for every x, the Markov chain given by Π_x admits a unique invariant probability μ_x (this is common in applications) and define the function $\bar{h} : \mathbb{R}^d \to \mathbb{R}^d$ for every $x \in \mathbb{R}^d$ by

$$\bar{h}(x) := \int h(x, s) \mu_x(ds) \,. \tag{1.10}$$

We can then rewrite the recursion (1.8) under the following form:

$$x_{n+1} = x_n + \gamma_{n+1} \bar{h}(x_n) + \gamma_{n+1}(h(x_n, \xi_{n+1}) - \bar{h}(x_n)) \,. \tag{1.11}$$

Under a control of the small perturbation induced by the last term in the equation above, one can expect the asymptotic behavior of the algorithm to be closely related to that of the following ordinary differential equation governed by the mean field $\bar{h}$:

$$\dot{x}(t) = \bar{h}(x(t)) \,. \tag{1.12}$$

Considering the discrete stochastic iterates (1.8) as a noisy discretization of the ODE (1.12), the iterates of the algorithm are shown to asymptotically track the trajectory of the solution to the ODE. An analysis of this ODE characterizing its equilibrium points provides the limit points of the algorithm at hand. This is the key idea of the so-called Ordinary Differential Equation (ODE) method from stochastic approximation theory which is at the heart of our analysis.

Initiated by the seminal work of Robbins and Monro (1951), stochastic approximation theory and the ODE method in particular have given rise to a vast literature that we cannot hope to do justice here. We refer though to the historical works of Ljung

[10] $\mathcal{B}(\mathbb{R}^k)$ is the Borel σ-algebra on $\mathbb{R}^k$.

(1977); Kushner and Clark (1978), the contributions of Métivier and Priouret (1984); Métivier and Priouret (1987) studying the general scheme (1.8) presented in this section, Benaïm (1996); Benaïm (1999) and the monographs Benveniste et al. (1990); Duflo (1997); Kushner and Yin (2003); Borkar (2008) for comprehensive treatments of the subject.

Although this thesis is motivated by optimization and RL applications, we briefly mention that the powerful theory of stochastic approximation finds applications in many diverse domains such as adaptive signal processing (see for e.g., Benveniste et al. (1990)), communication networks and economics (as referred to for instance in (Borkar, 2008, Chap. 1)), to name a few. To list a couple of recent applications, we highlight that stochastic approximation theory has also been fruitfully exploited for estimating entropically regularized Wasserstein distances between two probability measures in (semi-discrete) optimal transport (Bercu and Bigot, 2021; Bercu et al., 2021) and for solving min-max problems such as for Generative Adversarial Networks (Hsieh et al., 2021).

Back to the general scheme (1.8), and reusing the notations from Eq. (1.1), we mention that the canonical example of such algorithm is that of SGD which corresponds to the case where $h(x_n, \xi_{n+1}) = \nabla f(x_n, \xi_{n+1})$ and (ξ_n) is a sequence of iid r.v.s. In this particular case, the transition kernel Π coincides with the unknown probability distribution of ξ_0. The corresponding ODE is the so-called gradient flow $\dot{x}(t) = -\nabla F(x(t))$.

In the next two sections, we will independently describe our contributions in both non-convex stochastic optimization and reinforcement learning. With respect to the general stochastic approximation framework of (1.8)-(1.9), the adaptive gradient algorithms we will study will involve a more complex structure of the function h than that of SGD whereas the noise induced by the stochastic algorithms will take the form of a martingale difference sequence. Then, in the last part of this thesis where we tackle a RL problem, the theoretical framework relevant to this setting corresponds to the noise r.v. (ξ_n) being a Markov chain controlled by the sequence of interest (x_n). In this setting, the noise dynamics is more complex than simply iid r.v.s and controlling the Markov noise requires additional theoretical tools such as a decomposition of the perturbation based on the Poisson equation as proposed in Métivier and Priouret (1984). The actor-critic algorithm we will study involves multiple timescales. Accordingly, we will use the multiple timescales stochastic approximation theory developed by Borkar (see for e.g., (Borkar, 2008, Chap. 6))

1.2 Stochastic momentum algorithms for non-convex optimization

If SGD has been well studied in the literature (see for e.g., Benaïm (1996); Delyon (1996); Bertsekas and Tsitsiklis (2000); Moulines and Bach (2011); Bach and Moulines (2013)) and is still the subject of active research Bottou et al. (2018); Gower et al. (2019); Mertikopoulos et al. (2020); Sebbouh et al. (2021), adaptive gradient methods were comparatively much less studied prior to this thesis. Previously known results are either regret bounds in the online convex optimization framework or bounds on the expected gradient norm of the objective function for variants of famous algorithms such as Adam. In particular, these results do not address the question of the convergence of the iterates of the algorithms themselves. We refer the reader to Sections 2.6, 3.3.4, 4.3 and 4.7

in the subsequent chapters for further details on related works and comparison to the literature.

1.2.1 About Adam (Chapter 2)

In this subsection, we describe our contributions related to the study of the Adam algorithm to solve the non-convex stochastic optimization problem we introduced in Section 1.1.1 above. Rigorous statements with precise assumptions can be found in Chapter 2.

Continuous-time system: ODE analysis. We introduce a continuous-time version of the Adam algorithm (see (1.2a)-(1.2e)) under the form of a non-autonomous ODE which can be written as follows:

$$\dot{z}(t) = h(z(t), t)\,, \quad \text{with} \quad h(z,t) = \begin{pmatrix} -\frac{(1-e^{-at})^{-1} m}{\varepsilon + \sqrt{(1-e^{-bt})^{-1} v}} \\ a(\nabla F(x) - m) \\ b(S(x) - v) \end{pmatrix} \tag{1.13}$$

for every $t > 0$, $z = (x, m, v)$ in $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d_+$, where $z(t) = (x(t), m(t), v(t))$, $F(x) := \mathbb{E}[f(x, \xi)]$ and $S(x) := \mathbb{E}[\nabla f(x, \xi)^{\odot 2}]$ for every $x \in \mathbb{R}^d$ [11] and a, b are positive constants defined in Assumption 2.2.4.

Building on the existence of an explicit Lyapunov function for the ODE, we show the existence of a unique global solution to the ODE. This first result turns out to be non-trivial due to the irregularity of the vector field. We then establish the convergence of the continuous-time Adam trajectory to the set of critical points of the objective function F. The proposed continuous-time version of Adam provides useful insights on the effect of the bias correction step. It is shown that, close to the origin, the objective function F is non-increasing along the Adam trajectory, suggesting that early iterations of Adam can only improve the initial guess.

Convergence rates in continuous-time. Under a Łojasiewicz-type condition, we prove that the solution to the ODE converges to a single critical point x^* of the objective function F. In this case, we provide convergence rates in terms of the distance to the critical point $\|x(t) - x^*\|$ as a function of the so-called Łojasiewicz exponent of F at x^* (see Section 2.3.3 of Chapter 2 for a definition).

Constant stepsize Adam. In discrete time, we first analyze the Adam iterates in the constant stepsize regime as originally introduced in Kingma and Ba (2015). Under a stability condition, we prove the asymptotic ergodic convergence of the probability of the discrete-time Adam iterates to approach the set of critical points of the objective function in the doubly asymptotic regime where $n \to \infty$ then $\gamma \to 0$. This long run convergence result stipulates that for every $\delta > 0$,

$$\lim_{\gamma \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{P}(\mathsf{d}(x_k^\gamma, \nabla F^{-1}(\{0\})) > \delta) = 0\,, \tag{1.14}$$

where the iterates x_k^γ of Adam are indexed by the constant stepsize γ and d is the euclidean distance to a set. Note here that the iterates with constant stepsize γ cannot converge in the almost sure sense when only the number of iterations $n \to \infty$.

[11] Assumptions guaranteeing that these functions are well-defined are presented in Chapter 2.

A new decreasing stepsizes version of Adam. Beyond the original constant stepsize Adam, we propose a decreasing stepsize version of the algorithm. From the initial point $z_0 = (x_0, 0, 0)$ where $x_0 \in \mathbb{R}^d$, the algorithm generates a sequence $z_n = (x_n, m_n, v_n)$ as follows:

$$\begin{cases} m_{n+1} &= \alpha_n m_n + (1-\alpha_n)\nabla f(x_n, \xi_{n+1}) & \text{(1.15a)} \\ v_{n+1} &= \beta_n v_n + (1-\beta_n)\nabla f(x_n, \xi_{n+1})^{\odot 2} & \text{(1.15b)} \\ \hat{m}_{n+1} &= m_{n+1}/(1 - \Pi_{i=1}^{n+1}\alpha_i) & \text{(1.15c)} \\ \hat{v}_{n+1} &= v_{n+1}/(1 - \Pi_{i=1}^{n+1}\beta_i) & \text{(1.15d)} \\ x_{n+1} &= x_n - \gamma_{n+1}\,\hat{m}_{n+1}/(\varepsilon + \sqrt{\hat{v}_{n+1}})\ , & \text{(1.15e)} \end{cases}$$

First, compared to the original version of Kingma and Ba (2015) introduced above in (1.2a), the hyperparameters $(\gamma_n, \alpha_n, \beta_n)$ now depend on n. Second, the debiasing steps rescaling the iterates m_n and v_n are different due to the use of time-dependent hyperparameters α_n, β_n. Further details can be found in Chapter 2.

For this decreasing stepsizes version, we provide sufficient conditions ensuring the stability and the almost sure convergence of the iterates towards the critical points of the objective function F as predicted by our continuous-time analysis.

Fluctuations. We establish a convergence rate of the stochastic iterates of the decreasing stepsize algorithm under the form of a conditional central limit theorem. More precisely, for some decreasing stepsizes γ_n, we show that the vector $\sqrt{\gamma_n}^{-1}(x_n - x^*)$ converges in distribution to a zero mean Gaussian distribution with a covariance matrix Σ_1 that we explicitely compute thanks to the Lyapunov equation verified by the covariance matrix.

1.2.2 Generalized momentum algorithm (Chapter 3)

After Chapter 2 which is focused on the specific Adam algorithm, in Chapter 3, we go one step further to study a general stochastic optimization procedure, unifying several variants of stochastic gradient descent. Among others, these variants include the Stochastic Heavy Ball (SHB) method, the stochastic version of Nesterov's Accelerated Gradient method (S-NAG), and the large class of adaptive gradient algorithms, among which Adam is perhaps the most used in practice. We thereby extend the results of Chapter 2 and the work of Gadat et al. (2018) focused on SHB to a general setting.

Continuous-time system. The algorithm we consider is seen as a noisy Euler discretization of a non-autonomous ODE extending (1.13) (which is analyzed in Chapter 2) and concomitantly introduced by Belotto da Silva and Gazeau (2020). This ODE can be written as follows:

$$\dot{\mathsf{z}}(t) = g(\mathsf{z}(t), t)\,, \quad \text{with} \quad g(z,t) = \begin{bmatrix} \mathsf{p}(t)S(x) - \mathsf{q}(t)v \\ \mathsf{h}(t)\nabla F(x) - \mathsf{r}(t)m \\ -m/\sqrt{v+\varepsilon} \end{bmatrix} \tag{1.16}$$

for every $t > 0$, $z = (v, m, x)$ in $\mathbb{R}^d_+ \times \mathbb{R}^d \times \mathbb{R}^d$, where $F : \mathbb{R}^d \to \mathbb{R}$ is a continuously differentiable function, $S : \mathbb{R}^d \to \mathbb{R}^d_+$ is a continuous function, $\mathsf{h}, \mathsf{r}, \mathsf{p}, \mathsf{q} : (0,\infty) \to \mathbb{R}_+$ are four continuous functions, and $\varepsilon > 0$. Moreover, the ODE is initialized by $\mathsf{z}(0) = z_0 = (v_0, m_0, x_0)$ for some $v_0 \in \mathbb{R}^d_+, x_0, m_0 \in \mathbb{R}^d$, and $\mathsf{z}(t) = (\mathsf{v}(t), \mathsf{m}(t), \mathsf{x}(t)) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d_+$ for $t \in \mathbb{R}_+$. A proper choice of the functions $\mathsf{h}, \mathsf{r}, \mathsf{p}, \mathsf{q}$ followed by an Euler discretization of the ODE leads to numerous algorithms used in Machine Learning (see Chapter 3 for

details). In view of applying the ODE method to study the corresponding stochastic algorithm, we analyze this non-autonomous ODE representing the continuous-time versions of the aforementioned set of algorithms. Under some conditions on the functions $\mathsf{h}, \mathsf{r}, \mathsf{p}, \mathsf{q}$ (satisfied for instance for ADAM) together with some regularity assumptions on the objective function F and some mild technical assumptions, we show that there exist a unique bounded global solution to ODE (1.16) and the x component of the latter converges towards the set of critical points of the objective function F.

Almost sure convergence and fluctuations. Then, we introduce the novel general algorithm of interest as a noisy discretization of ODE (1.16). This algorithm uses a sequence of decreasing stepsizes (γ_n) verifying the Robbins Monro conditions. As for the discretization, we define for every integer $n \geq 1$,

$$\tau_n = \sum_{k=1}^{n} \gamma_k, \ h_n = \mathsf{h}(\tau_n), \ r_n = \mathsf{r}(\tau_n), \ p_n = \mathsf{p}(\tau_n), \text{ and } q_n = \mathsf{q}(\tau_n). \tag{1.17}$$

Initialized with $z_0 = (v_0, m_0, x_0) \in \mathbb{R}^d_+ \times \mathbb{R}^d \times \mathbb{R}^d$ and $h_0, r_0, p_0, q_0 \in (0, \infty)$, the algorithm is written as follows [12]:

$$\begin{cases} v_{n+1} &= (1 - \gamma_{n+1} q_n) v_n + \gamma_{n+1} p_n \nabla f(x_n, \xi_{n+1})^{\odot 2} & \text{(1.18a)} \\ m_{n+1} &= (1 - \gamma_{n+1} r_n) m_n + \gamma_{n+1} h_n \nabla f(x_n, \xi_{n+1}) & \text{(1.18b)} \\ x_{n+1} &= x_n - \gamma_{n+1} m_{n+1} / \sqrt{v_{n+1} + \varepsilon} \ , & \text{(1.18c)} \end{cases}$$

Similarly to Chapter 2, the stability and the almost sure convergence of the iterates of our general algorithm towards the set of critical points are established. Compared to Chapter 2, the algorithm above offers some freedom in the choice of the functions $\mathsf{h}, \mathsf{r}, \mathsf{p}, \mathsf{q}$ beyond the specific case of ADAM which corresponds to specific functions $\mathsf{h} \equiv \mathsf{r}$[13] and $\mathsf{p} \equiv \mathsf{q}$ provided in Chapter 3 (see also Belotto da Silva and Gazeau (2020)). This generalization needs some adaptations of the proofs even if they are very similar in spirit. For instance, the Lyapunov function we consider is different from the one considered in Chapter 2. Beyond this generalization, we also note two additional improvements. First, for our almost sure convergence result, we provide noise conditions allowing to choose larger stepsizes. Second, regarding the stability result, we relax an assumption related to the sequence of hyperparameters (α_n) and stepsizes (γ_n) (Assumption 2.5.2-iii)) which is no more needed thanks to a slight modification of the discretized Lyapunov function used in the proof. A noteworthy special case is the convergence proof of S-NAG in a non-convex setting which needs a specific proof. Under some assumptions, the convergence rate is also provided under the form of a Central Limit Theorem for the general algorithm (1.18a)-(1.18c).

Avoidance of traps. In Chapter 3, we also tackle the important question of the avoidance of traps. As stated in the previous paragraph, the iterates of the algorithm are shown to converge almost surely to the set of critical points of the objective function F. However, in a non-convex setting, the set of critical points of the function F is generally larger than the set of local minimizers. A "trap" stands for a critical point at which the Hessian matrix of F has negative eigenvalues, namely, it is a local maximum or saddle

[12] Note here the slight difference with the algorithm in (1.15e) for which ϵ is outside the square root. This slight simplification considered throughout Chapter 3 allows us to avoid some differentiability issues which we have tackled in Chapter 2. Furthermore, debiasing steps are handled via the sequences $(h_n), (r_n), (p_n)$ and (q_n).

[13] this notation means that the functions coincide for every $t \in \mathbb{R}_+$.

point. We establish that the iterates cannot converge to such an undesired point, if the noise is exciting in some directions.

Since the contributions of Pemantle (1990) and Brandière and Duflo (1996), the numerous so-called avoidance of traps results that can be found in the literature (see also for e.g., (Benaïm, 1999, Section 9),(Borkar, 2008, Section 4.3)) deal with the case where the ODE that underlies the stochastic algorithm is an autonomous ODE. Nevertheless, this is not the case of the non-autonomous ODE (1.16). To address this issue, we first state a general avoidance of traps result that extends Pemantle (1990); Brandière and Duflo (1996) to a non-autonomous setting, and which is of independent interest. We loosely describe this result in the rest of this paragraph and refer to Section 3.4 for rigorous statements and Section 3.7 for its proof.

To state our result, consider the stochastic approximation algorithm in $\mathbb{R}^d$ initialized at some $z_0 \in \mathbb{R}^d$:

$$z_{n+1} = z_n + \gamma_{n+1} b(z_n, \tau_n) + \gamma_{n+1}\eta_{n+1} + \gamma_{n+1}\rho_{n+1} \tag{1.19}$$

where $b : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$ is a continuous function, the sequence (γ_n) of nonnegative deterministic stepsizes satisfies the Robbins Monro conditions (i.e., $\sum_n \gamma_n = +\infty$, $\sum_n \gamma_n^2 < +\infty$), and $\tau_n = \sum_{k=1}^n \gamma_k$. Given a filtration $(\mathcal{F}_n)$, assume that the sequence (η_n) is a martingale difference sequence (adapted to $\mathcal{F}_n$) and that the sequence (ρ_n) which is also supposed to be adapted to $\mathcal{F}_n$ has a square summable euclidean norm. Let $z_\star \in \mathbb{R}^d$, and assume that on $\mathcal{V} \times \mathbb{R}_+$ where $\mathcal{V}$ is a certain neighborhood of $z_\star$, the function b can be developed as

$$b(z, t) = D(z - z_\star) + e(z, t), \tag{1.20}$$

where $e(z_\star, \cdot) \equiv 0$, and where the matrix $D \in \mathbb{R}^{d\times d}$ is assumed to have at least one eigenvalue with positive real part. As a consequence, the point $z_\star$ is an unstable equilibrium point of the ODE $\dot{\mathsf{z}}(t) = b(\mathsf{z}(t), t)$, in the sense that the ODE solution will only be able to converge to $z_\star$ along a specific so-called invariant manifold which precise characterization will be deferred to Chapter 3. In other words, the point $z_\star$ is a trap that the algorithm should desirably avoid.

As the reader can observe in Eq. (1.19), the stochastic algorithm is built around the non-autonomous ODE $\dot{\mathsf{z}}(t) = b(\mathsf{z}(t), t)$. The function e can be seen as a non-autonomous perturbation of the autonomous linear ODE $\dot{\mathsf{z}}(t) = D(\mathsf{z}(t) - z_\star)$. Regularity assumptions made on the function e guarantee the existence of a local (around the unstable equilibrium $z_\star$) non-autonomous invariant manifold of the non-autonomous ODE $\dot{\mathsf{z}}(t) = b(\mathsf{z}(t), t)$ with enough regularity properties. For proving this result, we use a non-autonomous version of Poincaré's invariant manifold theorem (Daleckii and Krein, 1974; Kloeden and Rasmussen, 2011) instead of its classical autonomous version.

Under regularity conditions on the function e and mild technical assumptions, we show that the event $\{z_n \to z_\star\}$ is of probability zero if the noise sequence (η_n) is sufficiently exciting in directions of the eigenspace associated with eigenvalues of D with positive real part. Thanks to this exciting noise, the algorithm trajectories will move away from the invariant manifold mentioned above.

We apply this result to both our general stochastic algorithm and S-NAG. Our general avoidance of traps result extends previous works of Gadat et al. (2018) obtained in the context of SHB. This result not only allows to study a broader class of algorithms but also significantly weakens the assumptions (see Section 3.4 in Chapter 3).

1.2.3 Convergence rates of a momentum algorithm with bounded adaptive stepsize (Chapter 4)

In this subsection which summarizes our contributions in Chapter 4 of this thesis, we provide non-asymptotic results complementing our asymptotic analysis in Chapters 2 and 3.

We establish a convergence rate for ADAM in the deterministic case for non-convex optimization under a bounded learning rate and constant stepsize. This algorithm can be seen as a deterministic clipped version of ADAM, which guarantees safe theoretical stepsizes. More precisely, if n is the number of iterations of the algorithm, we show a $O(1/n)$ convergence rate of the minimum of the squared gradients norms by introducing a suitable Lyapunov function. Then, we show a similar convergence result for non-convex stochastic optimization up to the limit of the variance of stochastic gradients under an almost surely bounded learning rate.

Finally, using the Kurdyka-Łojasiewicz (KŁ) property, we propose a convergence rate analysis of the objective function values along the iterates of the algorithm in the deterministic setting.

1.3 Actor-critic with target network for Reinforcement Learning (Chapter 5)

In Chapter 5 of this thesis, we deal with the RL problem we have introduced in subsection 1.1.3 using an algorithm which we will briefly present in what follows using the notations of the aforementioned subsection. Starting from $\theta_0, \omega_0, \bar{\omega}_0 \in \mathbb{R}^d$ and given three different sequences of positive stepsizes (α_t), (β_t), (ξ_t), the following recursions generate the sequence (θ_t) of interest (parameterizing the sought policy) online as follows:

$$\begin{cases} \delta_{t+1} &= R_{t+1} + \gamma\, \phi(S_{t+1})^T \omega_t - \phi(\tilde{S}_t)^T \omega_t & \text{(1.21a)} \\ \bar{\delta}_{t+1} &= R_{t+1} + \gamma\, \phi(S_{t+1})^T \bar{\omega}_t - \phi(\tilde{S}_t)^T \omega_t & \text{(1.21b)} \\ \theta_{t+1} &= \theta_t + \alpha_t \frac{1}{1-\gamma} \delta_{t+1} \nabla \ln \pi_{\theta_t}(\tilde{A}_t | \tilde{S}_t) & \text{(1.21c)} \\ \omega_{t+1} &= \omega_t + \beta_t \bar{\delta}_{t+1} \phi(\tilde{S}_t) & \text{(1.21d)} \\ \bar{\omega}_{t+1} &= \bar{\omega}_t + \xi_t (\omega_{t+1} - \bar{\omega}_t)\,, & \text{(1.21e)} \end{cases}$$

Here, at each timestep $t \in \mathbb{N}$, the state S_{t+1} is generated following the distribution $p(\cdot|\tilde{S}_t, \tilde{A}_t)$ and the action $\tilde{A}_t$ according to $\pi_{\theta_t}(\cdot|\tilde{S}_t)$. In the light of our brief description of actor-critic methods in Section 1.1.3, Eq. (1.21c) is used to update the actor parameter whereas (1.21d) describes the evolution of the so-called critic approximating the unknown value function. Concerning the actor, the temporal difference error δ_{t+1} estimates the unknown advantage function involved in the policy gradient theorem (see Eq. (1.4) and Eq. (1.6)). The sequence (ω_t) corresponds to the critic: as previously mentioned in Eq. (1.7), the quantity $\phi(\tilde{S}_t)^T \omega_t$ is a linear approximation of $V_{\pi_{\theta_t}}(\tilde{S}_t)$. Regarding the critic, if the sequences (ω_t) and $(\bar{\omega}_t)$ coincide (i.e., Eq. (1.21e) is modified accordingly), then the algorithm coincides with a standard actor-critic algorithm where the critic uses a standard value-based method called TD-learning (Sutton (1988)). Instead, the algorithm uses a sequence $(\bar{\omega}_t)$ defining the target network and updated using Eq. (1.21e). We refer the reader to Section 5.4 in Chapter 5 for a detailed presentation of the algorithm together with an explanation of the idea of target network.

Theoretical contributions investigating the use of a target network are very recent and limited to temporal difference (TD) learning for policy evaluation (Lee and He, 2019) and critic-only methods such as Q-learning for control (Zhang et al., 2021b). In particular, these works are not concerned with actor-critic algorithms and leave the question of the finite-time analysis open.

We propose the first theoretical analysis of an online target-based actor-critic algorithm (see Eqs. (1.21a) to (1.21e) above) in the discounted reward setting. We consider the linear function approximation setting where a linear combination of pre-selected feature (or basis) functions estimates the value function in the critic as can be apppreciated in Eqs. (1.21d) and (1.21b). An analysis of this setting is an insightful first step before tackling the more challenging nonlinear function approximation setting aligned with the use of neural networks. Our algorithm uses three different timescales: one for the actor using the stepsizes (α_t) and two for the critic using the stepsizes (β_t) and (ξ_t) s.t. $\alpha_t/\xi_t \to 0$ and $\xi_t/\beta_t \to 0$ as $t \to +\infty$. Instead of using the standard single timescale TD learning algorithm as a critic, as can be seen in Eqs. (1.21b), (1.21d) and (1.21e), we use a two timescales target-based version of TD learning closely inspired from practical actor-critic algorithms implementing target networks.

First, we establish asymptotic convergence results for both the critic and the actor under Markovian sampling. More precisely, as the actor parameter changes slowly compared to the critic one, we show that the critic which uses a target variable tracks a slowly moving target. Then, we show that the actor parameter visits infinitely often a region of the parameter space where the norm of the policy gradient is dominated by a bias induced by linear function approximation.

Second, we conduct a finite-time analysis of our actor-critic algorithm which shows the impact of using a target variable on the convergence rates and the sample complexity. This non-asymptotic analysis provides a more quantitative result bounding the average expected squared gradient norm $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\nabla J(\theta_t)\|^2]$ for a given positive time horizon $T \in \mathbb{N}$. Here, the sequence of interest (θ_t) is produced by our target-based actor-critic algorithm to solve Problem (1.3) with a parameterized family of policies $\{\pi_\theta : \theta \in \mathbb{R}^d\}$ up to the linear function approximation error.

1.4 Structure of the thesis and publications

Our contributions in this thesis resulted in the following publications and preprints listed in chronological order:

- Anas Barakat & Pascal Bianchi (2021). Convergence and Dynamical Behavior of the ADAM Algorithm for Non-Convex Stochastic Optimization. In: *SIAM Journal on Optimization, 31 (1), 244-274.*

- Anas Barakat & Pascal Bianchi (2020). Convergence Rates of a Momentum Algorithm with Bounded Adaptive Stepsize for Non-Convex Optimization. In: *Asian Conference on Machine Learning 2020, Proceedings of Machine Learning Research, 129, 225-240.*

- Anas Barakat, Pascal Bianchi, Walid Hachem & Sholom Schechtman (2021). Stochastic optimization with momentum: convergence, fluctuations, and traps avoidance. In: *Electronic Journal of Statistics 15 (2), 3892-3947.*

- Anas Barakat, Pascal Bianchi & Julien Lehmann (2021). Analysis of a Target-Based Actor-Critic Algorithm with Linear Function Approximation. *ArXiv Preprint: arXiv:2106.07472.*

This manuscript is organized as follows. Each chapter corresponds to one of the aforementioned publications and preprints. Especially devoted to RL, Chapter 5 is independent from the rest of the thesis. The last chapter of this thesis (Chapter 6) contains concluding remarks and some directions of future research.

2

Convergence and Dynamical Behavior of the ADAM Algorithm for Non-Convex Stochastic Optimization

Abstract The purpose of this chapter is to study the ADAM algorithm used for finding a local minimizer of a function. In the constant stepsize regime, assuming that the objective function is differentiable and non-convex, we establish the convergence in the long run of the iterates to a stationary point under a stability condition. The key ingredient is the introduction of a continuous-time version of ADAM, under the form of a non-autonomous ordinary differential equation. The existence and the uniqueness of the solution to the ODE are established. We further show the convergence of the solution towards the critical points of the objective function and quantify its convergence rate under a Łojasiewicz assumption. Then, we introduce a novel decreasing stepsize version of ADAM. Under mild assumptions, it is shown that the iterates are almost surely bounded and converge almost surely to critical points of the objective function. Finally, we analyze the fluctuations of the algorithm by means of a conditional central limit theorem.

2.1 Introduction

Consider the problem of finding a local minimizer of the expectation $F(x) := \mathbb{E}(f(x,\xi))$ w.r.t. $x \in \mathbb{R}^d$, where $f(\,.\,,\xi)$ is a possibly non-convex function depending on some random variable ξ. The distribution of ξ is assumed unknown, but revealed online by the observation of iid copies $(\xi_n : n \geq 1)$ of the r.v. ξ. Stochastic gradient descent (SGD) is the most classical algorithm to search for such a minimizer. Variants of SGD which include an inertial term have also become very popular. In these methods, the update rule depends on a parameter called the *learning rate*, which is generally assumed constant or vanishing. These algorithms, although widely used, have at least two limitations. First, the choice of the learning rate is generally difficult; large learning rates result in large fluctuations of the estimate, whereas small learning rates induce slow convergence. Second, a common learning rate is used for every coordinate despite the possible discrepancies in the values of the gradient vector's coordinates.

To alleviate these limitations, the popular ADAM algorithm (Kingma and Ba, 2015) adjusts the learning rate coordinate-wise, as a function of the past values of the squared gradient vectors' coordinates. The algorithm thus combines the assets of inertial methods with an adaptive per-coordinate learning rate selection. Finally, the algorithm includes a so-called *bias correction* step. Acting on the current estimate of the gradient vector, this step is especially useful during the early iterations.

Despite the growing popularity of the algorithm, only few works investigate its behavior from a theoretical point of view (see the discussion in Section 2.6). The present chapter studies the convergence of ADAM from a dynamical system viewpoint.

Contributions

- We introduce a continuous-time version of the ADAM algorithm under the form of a non-autonomous ordinary differential equation (ODE). Building on the existence of an explicit Lyapunov function for the ODE, we show the existence of a unique global solution to the ODE. This first result turns out to be non-trivial due to the irregularity of the vector field. We then establish the convergence of the continuous-time ADAM trajectory to the set of critical points of the objective function F. The proposed continuous-time version of ADAM provides useful insights on the effect of the bias correction step. It is shown that, close to the origin, the objective function F is non-increasing along the ADAM trajectory, suggesting that early iterations of ADAM can only improve the initial guess.
- Under a Łojasiewicz-type condition, we prove that the solution to the ODE converges to a single critical point of the objective function F. We provide convergence rates in this case.
- In discrete time, we first analyze the ADAM iterates in the constant stepsize regime as originally introduced in Kingma and Ba (2015). In this work, it is shown that the discrete-time ADAM iterates shadow the behavior of the non-autonomous ODE in the asymptotic regime where the stepsize parameter γ of ADAM is small. More precisely, we consider the interpolated process $\mathsf{z}^\gamma(t)$ which consists of a piecewise linear interpolation of the ADAM iterates. The random process z^γ is indexed by the parameter γ, which is assumed constant during the whole run of the algorithm. In the space of continuous functions on $[0, +\infty)$ equipped with the topology of uniform convergence on compact sets, we establish that z^γ converges in probability to the solution to the non-autonomous ODE when γ tends to zero.
- Under a stability condition, we prove the asymptotic ergodic convergence of the probability of the discrete-time ADAM iterates to approach the set of critical points of the objective function in the doubly asymptotic regime where $n \to \infty$ then $\gamma \to 0$.
- Beyond the original constant stepsize ADAM, we propose a decreasing stepsize version of the algorithm. We provide sufficient conditions ensuring the stability and the almost sure convergence of the iterates towards the critical points of the objective function.
- We establish a convergence rate of the stochastic iterates of the decreasing stepsize algorithm under the form of a conditional central limit theorem.

We claim that our analysis can be easily extended to other adaptive algorithms such as e.g. RMSPROP or ADAGRAD (Tieleman and Hinton, 2012; Duchi et al., 2011) and AMSGRAD (see Section 2.6).

Chapter organization. In Section 2.2, we present the ADAM algorithm and the main assumptions. Our main results are stated in Sections 2.3 to 2.5. We provide a review of related works in Section 2.6. The rest of the chapter addresses the proofs of our results (Sections 2.7 to 2.9).

Notations. If x, y are two vectors on $\mathbb{R}^d$ for some $d \geq 1$, we denote by $x \odot y$, $x^{\odot 2}$, x/y, $|x|$, $\sqrt{|x|}$ the vectors on $\mathbb{R}^d$ whose i-th coordinates are respectively given by $x_i y_i$, x_i^2, x_i/y_i, $|x_i|$, $\sqrt{|x_i|}$. Inequalities of the form $x \leq y$ are read componentwise. Denote by

Algorithm 2.1 ADAM$(\gamma, \alpha, \beta, \varepsilon)$.

Initialization: $x_0 \in \mathbb{R}^d, m_0 = 0, v_0 = 0$.
for $n = 1$ **to** n_{iter} **do**
  $m_n = \alpha m_{n-1} + (1-\alpha)\nabla f(x_{n-1}, \xi_n)$
  $v_n = \beta v_{n-1} + (1-\beta)\nabla f(x_{n-1}, \xi_n)^{\odot 2}$
  $\hat{m}_n = m_n/(1-\alpha^n)$ {bias correction step}
  $\hat{v}_n = v_n/(1-\beta^n)$ {bias correction step}
  $x_n = x_{n-1} - \gamma \hat{m}_n/(\varepsilon + \sqrt{\hat{v}_n})$.
end for

$\|\cdot\|$ the standard Euclidean norm. For any vector $v \in (0,+\infty)^d$, write $\|x\|_v^2 = \sum_i v_i x_i^2$. Notation A^T represents the transpose of a matrix A. If $z \in \mathbb{R}^d$ and A is a non-empty subset of $\mathbb{R}^d$, we use the notation $\mathsf{d}(z, A) := \inf\{\|z - z'\| : z' \in A\}$. If A is a set, we denote by $\mathbb{1}_A$ the function equal to one on that set and to zero elsewhere. We denote by $C([0,+\infty), \mathbb{R}^d)$ the space of continuous functions from $[0,+\infty)$ to $\mathbb{R}^d$ endowed with the topology of uniform convergence on compact intervals.

2.2 The ADAM algorithm

2.2.1 Algorithm and Assumptions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $(\Xi, \mathfrak{S})$ denote another measurable space. Consider a measurable map $f : \mathbb{R}^d \times \Xi \to \mathbb{R}$, where d is an integer. For a fixed value of ξ, the mapping $x \mapsto f(x, \xi)$ is supposed to be differentiable, and its gradient w.r.t. x is denoted by $\nabla f(x, \xi)$. Define $\mathcal{Z} := \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$, $\mathcal{Z}_+ := \mathbb{R}^d \times \mathbb{R}^d \times [0,+\infty)^d$ and $\mathcal{Z}_+^* := \mathbb{R}^d \times \mathbb{R}^d \times (0,+\infty)^d$. ADAM generates a sequence $z_n := (x_n, m_n, v_n)$ on $\mathcal{Z}_+$ given by Algorithm 2.1. It satisfies: $z_n = T_{\gamma,\alpha,\beta}(n, z_{n-1}, \xi_n)$, for every $n \geq 1$, where for every $z = (x, m, v)$ in $\mathcal{Z}_+$, $\xi \in \Xi$,

$$T_{\gamma,\alpha,\beta}(n, z, \xi) := \begin{pmatrix} x - \frac{\gamma(1-\alpha^n)^{-1}(\alpha m + (1-\alpha)\nabla f(x,\xi))}{\varepsilon + (1-\beta^n)^{-1/2}(\beta v + (1-\beta)\nabla f(x,\xi)^{\odot 2})^{1/2}} \\ \alpha m + (1-\alpha)\nabla f(x,\xi) \\ \beta v + (1-\beta)\nabla f(x,\xi)^{\odot 2} \end{pmatrix} . \tag{2.1}$$

Remark 1. The iterates z_n form a non-homogeneous Markov chain, because $T_{\gamma,\alpha,\beta}(n, z, \xi)$ depends on n. This is due to the so-called debiasing step, which consists of replacing m_n, v_n in Algorithm 2.1 by their "debiased" versions $\hat{m}_n, \hat{v}_n$. The motivation becomes clear when expanding the expression:

$$\hat{m}_n = \frac{m_n}{1-\alpha^n} = \frac{1-\alpha}{1-\alpha^n} \sum_{k=0}^{n-1} \alpha^k \nabla f(x_k, \xi_{k+1}) .$$

From this equation, it is observed that, $\hat{m}_n$ forms a convex combination of the past gradients. This is unlike m_n, which may be small during the first iterations.

Assumption 2.2.1. The mapping $f : \mathbb{R}^d \times \Xi \to \mathbb{R}$ satisfies the following.

i) For every $x \in \mathbb{R}^d$, $f(x, \,.\,)$ is $\mathfrak{S}$-measurable.
ii) For almost every ξ, the map $f(\,.\,, \xi)$ is continuously differentiable.

iii) There exists $x_* \in \mathbb{R}^d$ s.t. $\mathbb{E}(|f(x_*, \xi)|) < \infty$ and $\mathbb{E}(\|\nabla f(x_*, \xi)\|^2) < \infty$.
iv) For every compact subset $K \subset \mathbb{R}^d$, there exists $L_K > 0$ such that for every $(x, y) \in K^2$, $\mathbb{E}(\|\nabla f(x, \xi) - \nabla f(y, \xi)\|^2) \leq L_K^2 \|x - y\|^2$.

Under Assumption 2.2.1, it is an easy exercise to show that the mappings $F : \mathbb{R}^d \to \mathbb{R}$ and $S : \mathbb{R}^d \to \mathbb{R}^d$, given by:

$$F(x) := \mathbb{E}(f(x, \xi)) \quad \text{and} \quad S(x) := \mathbb{E}(\nabla f(x, \xi)^{\odot 2}) \tag{2.2}$$

are well defined; F is continuously differentiable and by Lebesgue's dominated convergence theorem, $\nabla F(x) = \mathbb{E}(\nabla f(x, \xi))$ for all x. Moreover, ∇F and S are locally Lipschitz continuous.

Assumption 2.2.2. F is coercive.

Assumption 2.2.3. For every $x \in \mathbb{R}^d$, $S(x) > 0$.

It follows from our assumptions that the set of critical points of F, denoted by

$$\mathcal{S} := \nabla F^{-1}(\{0\}),$$

is non-empty. Assumption 2.2.3 means that there is *no* point $x \in \mathbb{R}^d$ satisfying $\nabla f(x, \xi) = 0$ with probability one (w.p.1). This is a mild hypothesis in practice.

2.2.2 Asymptotic regime

We address the constant stepsize regime, where γ is fixed along the iterations (the default value recommended in Kingma and Ba (2015) is $\gamma = 0.001$). As opposed to the decreasing stepsize context, the sequence $z_n^\gamma := z_n$ *cannot* in general converge as n tends to infinity, in an almost sure sense. Instead, we investigate the asymptotic behavior of the family of processes $(n \mapsto z_n^\gamma)_{\gamma > 0}$ indexed by γ, in the regime where $\gamma \to 0$. We use the ODE method (see e.g., Benaïm (1999); Benaïm and Schreiber (2000)). The interpolated process z^γ is the piecewise linear function defined on $[0, +\infty) \to \mathcal{Z}_+$ for all $t \in [n\gamma, (n+1)\gamma)$ by:

$$\mathsf{z}^\gamma(t) := z_n^\gamma + (z_{n+1}^\gamma - z_n^\gamma) \left(\frac{t - n\gamma}{\gamma} \right). \tag{2.3}$$

We establish the convergence in probability of the family of random processes $(\mathsf{z}^\gamma)_{\gamma > 0}$ as γ tends to zero, towards a deterministic continuous-time system defined by an ODE. The latter ODE, which we provide below at Eq. (ODE), will be referred to as the continuous-time version of Adam.

Before describing the ODE, we need to be more specific about our asymptotic regime. As opposed to SGD, Adam depends on two parameters α, β, in addition to the stepsize γ. Kingma and Ba (2015) recommend choosing the constants α and β close to one (the default values $\alpha = 0.9$ and $\beta = 0.999$ are suggested). It is thus legitimate to assume that α and β tend to one, as γ tends to zero. We set $\alpha := \bar{\alpha}(\gamma)$ and $\beta := \bar{\beta}(\gamma)$, where $\bar{\alpha}(\gamma)$ and $\bar{\beta}(\gamma)$ converge to one as $\gamma \to 0$.

Assumption 2.2.4. The functions $\bar{\alpha} : \mathbb{R}_+ \to [0, 1)$ and $\bar{\beta} : \mathbb{R}_+ \to [0, 1)$ are s.t. the following limits exist:

$$a := \lim_{\gamma \downarrow 0} \frac{1 - \bar{\alpha}(\gamma)}{\gamma}, \quad b := \lim_{\gamma \downarrow 0} \frac{1 - \bar{\beta}(\gamma)}{\gamma}. \tag{2.4}$$

Moreover, $a > 0$, $b > 0$, and the following condition holds: $b \leq 4a$.

Note that the condition $b \leq 4a$ is compatible with the default settings recommended by Kingma and Ba (2015). In our model, we shall now replace the map $T_{\gamma,\alpha,\beta}$ by $T_{\gamma,\bar{\alpha}(\gamma),\bar{\beta}(\gamma)}$. Let $x_0 \in \mathbb{R}^d$ be fixed. For any fixed $\gamma > 0$, we define the sequence (z_n^γ) generated by ADAM with a fixed stepsize $\gamma > 0$:

$$z_n^\gamma := T_{\gamma,\bar{\alpha}(\gamma),\bar{\beta}(\gamma)}(n, z_{n-1}^\gamma, \xi_n)\,, \tag{2.5}$$

the initialization being chosen as $z_0^\gamma = (x_0, 0, 0)$.

2.3 Continuous-time system

2.3.1 Ordinary differential equation

In order to gain insight into the behavior of the sequence (z_n^γ) defined by (2.5), it is convenient to rewrite the ADAM iterations under the following equivalent form, for every $n \geq 1$:

$$z_n^\gamma = z_{n-1}^\gamma + \gamma h_\gamma(n, z_{n-1}^\gamma) + \gamma \Delta_n^\gamma\,, \tag{2.6}$$

where we define for every $\gamma > 0$, $z \in \mathcal{Z}_+$,

$$h_\gamma(n, z) := \gamma^{-1}\mathbb{E}(T_{\gamma,\bar{\alpha}(\gamma),\bar{\beta}(\gamma)}(n, z, \xi) - z)\,, \tag{2.7}$$

and where $\Delta_n^\gamma := \gamma^{-1}(z_n^\gamma - z_{n-1}^\gamma) - h_\gamma(n, z_{n-1}^\gamma)$. Note that (Δ_n^γ) is a martingale increment noise sequence in the sense that $\mathbb{E}(\Delta_n^\gamma | \mathcal{F}_{n-1}) = 0$ for all $n \geq 1$, where $\mathcal{F}_n$ stands for the σ-algebra generated by the r.v. $\xi_1, \ldots, \xi_n$. Define the map $h : (0, +\infty) \times \mathcal{Z}_+ \to \mathcal{Z}$ for all $t > 0$, all $z = (x, m, v)$ in $\mathcal{Z}_+$ by:

$$h(t, z) = \begin{pmatrix} -\frac{(1-e^{-at})^{-1} m}{\varepsilon + \sqrt{(1-e^{-bt})^{-1} v}} \\ a(\nabla F(x) - m) \\ b(S(x) - v) \end{pmatrix}\,, \tag{2.8}$$

where a, b are the constants defined in Assumption 2.2.4. We prove that, for any fixed (t, z), the quantity $h(t, z)$ coincides with the limit of $h_\gamma(\lfloor t/\gamma \rfloor, z)$ as $\gamma \downarrow 0$. This remark along with Eq. (2.6) suggests that, as $\gamma \downarrow 0$, the interpolated process z^γ shadows the non-autonomous differential equation

$$\dot{z}(t) = h(t, z(t))\,. \tag{ODE}$$

2.3.2 Existence, uniqueness, convergence

Since $h(\,.\,, z)$ is non-continuous at point zero for a fixed $z \in \mathcal{Z}_+$, and since $h(t, \,.\,)$ is not locally Lipschitz continuous for a fixed $t > 0$, the existence and uniqueness of the solution to (ODE) do not stem directly from off-the-shelf theorems.

Let x_0 be fixed. A continuous map $z : [0, +\infty) \to \mathcal{Z}_+$ is said to be a global solution to (ODE) with initial condition $(x_0, 0, 0)$ if z is continuously differentiable on $(0, +\infty)$, if Eq. (ODE) holds for all $t > 0$, and if $z(0) = (x_0, 0, 0)$.

Theorem 2.1 (Existence and uniqueness)**.** *Let Assumptions 2.2.1 to 2.2.4 hold true. There exists a unique global solution $z : [0, +\infty) \to \mathcal{Z}_+$ to (ODE) with initial condition $(x_0, 0, 0)$. Moreover, $z([0, +\infty))$ is a bounded subset of $\mathcal{Z}_+$.*

On the other hand, we note that a solution may not exist for an initial point (x_0, m_0, v_0) with arbitrary (non-zero) values of m_0, v_0.

Theorem 2.2 (Convergence)**.** *Let Assumptions 2.2.1 to 2.2.4 hold true. Assume that $F(\mathcal{S})$ has an empty interior. Let $z : t \mapsto (x(t), m(t), v(t))$ be the global solution to (ODE) with the initial condition $(x_0, 0, 0)$. Then, the set $\mathcal{S}$ is non-empty and $\lim_{t\to\infty} \mathsf{d}(x(t), \mathcal{S}) = 0$, $\lim_{t\to\infty} m(t) = 0$, $\lim_{t\to\infty} S(x(t)) - v(t) = 0$.*

Lyapunov function. The proof of Th. 2.1 relies on the existence of a Lyapunov function for the non-autonomous equation (ODE). Define $V : (0, +\infty) \times \mathcal{Z}_+ \to \mathbb{R}$ by

$$V(t, z) := F(x) + \frac{1}{2} \|m\|^2_{U(t,v)^{-1}} \,, \tag{2.9}$$

for every $t > 0$ and every $z = (x, m, v)$ in $\mathcal{Z}_+$, where $U : (0, +\infty) \times [0, +\infty)^d \to \mathbb{R}^d$ is the map given by:

$$U(t, v) := a(1 - e^{-at}) \left(\varepsilon + \sqrt{\frac{v}{1 - e^{-bt}}} \right) . \tag{2.10}$$

Then, $t \mapsto V(t, z(t))$ is decreasing if $z(\cdot)$ is the global solution to (ODE).

Cost decrease at the origin. As F itself is not a Lyapunov function for (ODE), there is no guarantee that $F(x(t))$ is decreasing w.r.t. t. Nevertheless, the statement holds at the origin. Indeed, it can be shown that $\lim_{t\downarrow 0} V(t, z(t)) = F(x_0)$ (see Prop. 2.12). As a consequence,

$$\forall t \geq 0, \;\; F(x(t)) \leq F(x_0) \,. \tag{2.11}$$

In other words, the (continuous-time) Adam procedure *can only improve* the initial guess x_0. This is the consequence of the so-called bias correction steps in Adam (see Algorithm 2.1). If these debiasing steps were deleted in the Adam iterations, the early stages of the algorithm could degrade the initial estimate x_0.

Derivatives at the origin. The proof of Th. 2.1 reveals that the initial derivative is given by $\dot{x}(0) = -\nabla F(x_0)/(\varepsilon + \sqrt{S(x_0)})$ (see Lem. 2.9). In the absence of debiasing steps, the initial derivative $\dot{x}(0)$ would be a function of the initial parameters m_0, v_0, and the user would be required to tune these hyperparameters. No such tuning is required thanks to the debiasing step. When ε is small and when the variance of $\nabla f(x_0, \xi)$ is small (*i.e.*, $S(x_0) \simeq \nabla F(x_0)^{\odot 2}$), the initial derivative $\dot{x}(0)$ is approximately equal to $-\nabla F(x_0)/|\nabla F(x_0)|$. This suggests that in the early stages of the algorithm, the Adam iterations are comparable to the *sign* variant of the gradient descent, the properties of which were discussed in previous works, see Balles and Hennig (2018).

2.3.3 Convergence rates

In this paragraph, we establish the convergence to a single critical point of F and quantify the convergence rate, using the following assumption (Łojasiewicz, 1963).

Assumption 2.3.1 (Łojasiewicz property)**.** For any $x^* \in \mathcal{S}$, there exist $c > 0\,, \sigma > 0$ and $\theta \in (0, \frac{1}{2}]$ s.t.

$$\forall x \in \mathbb{R}^d \text{ s.t } \|x - x^*\| \leq \sigma \,, \quad \|\nabla F(x)\| \geq c|F(x) - F(x^*)|^{1-\theta} \,. \tag{2.12}$$

Assumption 2.3.1 holds for real-analytic functions and semialgebraic functions. We refer to Haraux and Jendoubi (2015); Attouch and Bolte (2009); Bolte et al. (2014) for a discussion and a review of applications. We will call any θ satisfying (2.12) for some $c, \sigma > 0$, as a Łojasiewicz exponent of F at x^*. The next result establishes the convergence of the function $x(t)$ generated by the ODE to a single critical point of F, and provides the convergence rate as a function of the Łojasiewicz exponent of F at this critical point. The proof is provided in subsection 2.7.4.

Theorem 2.3. *Let Assumptions 2.2.1 to 2.2.4 and 2.3.1 hold true. Assume that $F(\mathcal{S})$ has an empty interior. Let $x_0 \in \mathbb{R}^d$ and let $z : t \mapsto (x(t), m(t), v(t))$ be the global solution to (ODE) with initial condition $(x_0, 0, 0)$. Then, there exists $x^* \in \mathcal{S}$ such that $x(t)$ converges to x^* as $t \to +\infty$.*

Moreover, if $\theta \in (0, \frac{1}{2}]$ is a Łojasiewicz exponent of F at x^*, there exists a constant $C > 0$ s.t. for all $t \geq 0$,

$$\|x(t) - x^*\| \leq C t^{-\frac{\theta}{1-2\theta}}, \quad \text{if } 0 < \theta < \frac{1}{2},$$
$$\|x(t) - x^*\| \leq C e^{-\delta t}, \quad \text{for some } \delta > 0 \text{ if } \theta = \frac{1}{2}.$$

2.4 Discrete-time system: convergence of ADAM

Assumption 2.4.1. The sequence $(\xi_n : n \geq 1)$ is iid, with the same distribution as ξ.

Assumption 2.4.2. Let $p > 0$. Assume either one of the following conditions.

i) For every compact set $K \subset \mathbb{R}^d$, $\sup_{x \in K} \mathbb{E}(\|\nabla f(x, \xi)\|^p) < \infty$.
ii) For every compact set $K \subset \mathbb{R}^d$, $\exists\, p_K > p$, $\sup_{x \in K} \mathbb{E}(\|\nabla f(x, \xi)\|^{p_K}) < \infty$.

The value of p will be specified in the sequel, in the statement of the results. Clearly, Assumption 2.4.2 *ii)* is stronger than Assumption 2.4.2 *i)*. We shall use either the latter or the former in our statements.

Theorem 2.4. *Let Assumptions 2.2.1 to 2.2.4 and 2.4.1 hold true. Let Assumption 2.4.2 ii) hold with $p = 2$. Consider $x_0 \in \mathbb{R}^d$. For every $\gamma > 0$, let $(z_n^\gamma : n \in \mathbb{N})$ be the random sequence defined by the ADAM iterations (2.5) and $z_0^\gamma = (x_0, 0, 0)$. Let z^γ be the corresponding interpolated process defined by Eq. (2.3). Finally, let z denote the unique global solution to (ODE) issued from $(x_0, 0, 0)$. Then,*

$$\forall T > 0, \ \forall \delta > 0, \quad \lim_{\gamma \downarrow 0} \mathbb{P}\left(\sup_{t \in [0,T]} \|\mathsf{z}^\gamma(t) - z(t)\| > \delta \right) = 0.$$

Recall that a family of r.v. $(X_\alpha)_{\alpha \in I}$ is called *bounded in probability*, or *tight*, if for every $\delta > 0$, there exists a compact set K s.t. $\mathbb{P}(X_\alpha \in K) \geq 1 - \delta$ for every $\alpha \in I$.

Assumption 2.4.3. There exists $\bar{\gamma}_0 > 0$ s.t. the family of r.v. $(z_n^\gamma : n \in \mathbb{N}, 0 < \gamma < \bar{\gamma}_0)$ is bounded in probability.

Theorem 2.5. *Consider $x_0 \in \mathbb{R}^d$. For every $\gamma > 0$, let $(z_n^\gamma : n \in \mathbb{N})$ be the random sequence defined by the ADAM iterations (2.5) and $z_0^\gamma = (x_0, 0, 0)$. Let Assumptions 2.2.1 to 2.2.4, 2.4.1 and 2.4.3 hold. Let Assumption 2.4.2 ii) hold with $p = 2$. Then, for*

Algorithm 2.2 ADAM- decreasing stepsize $(((\gamma_n, \alpha_n, \beta_n) : n \in \mathbb{N}^*), \varepsilon)$**.**

Initialization: $x_0 \in \mathbb{R}^d, m_0 = 0, v_0 = 0, r_0 = \bar{r}_0 = 0.$
for $n = 1$ **to** n_{iter} **do**
$m_n = \alpha_n m_{n-1} + (1 - \alpha_n)\nabla f(x_{n-1}, \xi_n)$
$v_n = \beta_n v_{n-1} + (1 - \beta_n)\nabla f(x_{n-1}, \xi_n)^{\odot 2}$
$r_n = \alpha_n r_{n-1} + (1 - \alpha_n)$
$\bar{r}_n = \beta_n \bar{r}_{n-1} + (1 - \beta_n)$
$\hat{m}_n = m_n / r_n$ {bias correction step}
$\hat{v}_n = v_n / \bar{r}_n$ {bias correction step}
$x_n = x_{n-1} - \gamma_n \hat{m}_n / (\varepsilon + \sqrt{\hat{v}_n})$.
end for

every $\delta > 0$,

$$\lim_{\gamma \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{P}(\mathsf{d}(x_k^\gamma, \mathcal{S}) > \delta) = 0 \,. \tag{2.13}$$

Convergence in the long run. When the stepsize γ is constant, the sequence (x_n^γ) cannot converge in the almost sure sense as $n \to \infty$. Convergence may only hold in the doubly asymptotic regime where $n \to \infty$ then $\gamma \to 0$.

Randomization. For every n, consider a r.v. N_n uniformly distributed on $\{1, \ldots, n\}$. Define $\tilde{x}_n^\gamma = x_{N_n}^\gamma$. We obtain from Th. 2.5 that for every $\delta > 0$,

$$\limsup_{n \to \infty} \; \mathbb{P}(\mathsf{d}(\tilde{x}_n^\gamma, \mathcal{S}) > \delta) \xrightarrow[\gamma \downarrow 0]{} 0 \,.$$

Relationship between discrete and continuous time ADAM. Th. 2.4 means that the family of random processes $(\mathsf{z}^\gamma : \gamma > 0)$ converges in probability as $\gamma \downarrow 0$ towards the unique solution to (ODE) issued from $(x_0, 0, 0)$. This motivates the fact that the non-autonomous system (ODE) is a relevant approximation to the behavior of the iterates $(z_n^\gamma : n \in \mathbb{N})$ for a small value of the stepsize γ.

Stability. Assumption 2.4.3 ensures that the iterates z_n^γ do not explode in the long run. A sufficient condition is for instance that $\sup_{n,\gamma} \mathbb{E}\|z_n^\gamma\| < \infty$. In theory, this assumption can be difficult to verify. Nevertheless, in practice, a projection step on a compact set can be introduced to ensure the boundedness of the estimates.

2.5 A decreasing stepsize ADAM algorithm

2.5.1 Algorithm

ADAM inherently uses constant stepsizes. Consequently, the iterates (2.5) do not converge in the almost sure sense. In order to achieve convergence, we introduce in this section a decreasing stepsize version of ADAM. The iterations are given in Algorithm 2.2. The algorithm generates a sequence $z_n = (x_n, m_n, v_n)$ with initial point $z_0 = (x_0, 0, 0)$, where $x_0 \in \mathbb{R}^d$. Apart from the fact that the hyperparameters $(\gamma_n, \alpha_n, \beta_n)$ now depend on n, the main difference w.r.t Algorithm 2.1 lies in the expression of the debiasing step. As noted in Remark 1, the aim is to rescale m_n (resp. v_n) in such a way that the rescaled version $\hat{m}_n$ (resp. $\hat{v}_n$) is a convex combination of past stochastic gradients

(resp. squared gradients). While in the constant step case the rescaling coefficient is $(1-\alpha^n)^{-1}$ (resp. $(1-\beta^n)^{-1}$), the decreasing step case requires dividing m_n by the coefficient $r_n = 1 - \prod_{i=1}^n \alpha_i$ (resp. v_n by $\bar{r}_n = 1 - \prod_{i=1}^n \beta_i$), which keeps track of the previous weights:

$$\hat{m}_n = \frac{m_n}{r_n} = \frac{\sum_{k=1}^n \rho_{n,k} \nabla f(x_{k-1}, \xi_k)}{\sum_{k=1}^n \rho_{n,k}},$$

where for every n, k, $\rho_{n,k} = \alpha_n \cdots \alpha_{k+1}(1-\alpha_k)$. A similar equation holds for $\hat{v}_n$.

2.5.2 Almost sure convergence

Assumption 2.5.1 (Stepsizes)**.** The following holds.

i) For all $n \in \mathbb{N}$, $\gamma_n > 0$ and $\gamma_{n+1}/\gamma_n \to 1$,
ii) $\sum_n \gamma_n = +\infty$ and $\sum_n \gamma_n^2 < +\infty$,
iii) For all $n \in \mathbb{N}$, $0 \leq \alpha_n \leq 1$ and $0 \leq \beta_n \leq 1$,
iv) There exist a, b s.t. $0 < b < 4a$, $\gamma_n^{-1}(1-\alpha_n) \to a$ and $\gamma_n^{-1}(1-\beta_n) \to b$.

Theorem 2.6. *Let Assumptions 2.2.1 to 2.2.3, 2.4.1 and 2.5.1 hold. Let Assumption 2.4.2 i) hold with $p = 4$. Assume that $F(\mathcal{S})$ has an empty interior and that the random sequence $((x_n, m_n, v_n) : n \in \mathbb{N})$ given by Algorithm 2.2 is bounded, with probability one. Then, w.p.1, $\lim_{n\to\infty} \mathsf{d}(x_n, \mathcal{S}) = 0$, $lim_{n\to\infty} m_n = 0$ and $lim_{n\to\infty}(S(x_n) - v_n) = 0$. If moreover $\mathcal{S}$ is finite or countable, then w.p.1, there exists $x^* \in \mathcal{S}$ s.t. $\lim_{n\to\infty}(x_n, m_n, v_n) = (x^*, 0, S(x^*))$.*

Th. 2.6 establishes the almost sure convergence of x_n to the set of critical points of F, under the assumption that the sequence $((x_n, m_n, v_n))$ is a.s. bounded. The next result provides a sufficient condition under which almost sure boundedness holds.

Assumption 2.5.2. The following holds.

i) ∇F is Lipschitz continuous.
ii) There exists $C > 0$ s.t. for all $x \in \mathbb{R}^d$, $\mathbb{E}[\|\nabla f(x, \xi)\|^2] \leq C(1 + F(x))$.
iii) We assume the condition: $\limsup_{n\to\infty} \left(\frac{1}{\gamma_n} - \left(\frac{1-\alpha_{n+2}}{1-\alpha_{n+1}} \right) \frac{1}{\gamma_{n+1}} \right) < 2(a - \frac{b}{4})$,
which is satisfied for instance if $b < 4a$ and $1 - \alpha_{n+1} = a\gamma_n$.

Theorem 2.7. *Let Assumptions 2.2.1, 2.2.2, 2.4.1, 2.5.1 and 2.5.2 hold. Let Assumption 2.4.2 i) hold with $p = 4$. Then, the sequence $((x_n, m_n, v_n) : n \in \mathbb{N})$ given by Algorithm 2.2 is bounded with probability one.*

2.5.3 Central limit theorem

Assumption 2.5.3. Let $x^* \in \mathcal{S}$. There exists a neighborhood $\mathcal{V}$ of x^* s.t.

i) F is twice continuously differentiable on $\mathcal{V}$, and the Hessian $\nabla^2 F(x^*)$ of F at x^* is positive definite.
ii) S is continuously differentiable on $\mathcal{V}$.

Define $D := \mathrm{diag}\left((\varepsilon + \sqrt{S_1(x^*)})^{-1}, \cdots, (\varepsilon + \sqrt{S_d(x^*)})^{-1} \right)$. Let P be an orthogonal matrix s.t. the following spectral decomposition holds:

$$D^{1/2} \nabla^2 F(x^*) D^{1/2} = P \mathrm{diag}(\lambda_1, \cdots, \lambda_d) P^{-1},$$

where $\lambda_1, \cdots, \lambda_d$ are the (positive) eigenvalues of $D^{1/2}\nabla^2 F(x^*)D^{1/2}$. Define

$$H := \begin{pmatrix} 0 & -D & 0 \\ a\nabla^2 F(x^*) & -aI_d & 0 \\ b\nabla S(x^*) & 0 & -bI_d \end{pmatrix} \tag{2.14}$$

where I_d represents the $d \times d$ identity matrix and $\nabla S(x^*)$ is the Jacobian matrix of S at x^*. The largest real part of the eigenvalues of H coincides with $-L$, where

$$L := b \wedge \frac{a}{2}\left(1 - \sqrt{\left(1 - \frac{4\lambda_1}{a}\right) \vee 0}\right) > 0\,. \tag{2.15}$$

Finally, define the $3d \times 3d$ matrix

$$Q := \begin{pmatrix} 0 & 0 & 0 \\ 0 & \mathbb{E}\left[\begin{pmatrix} a\nabla f(x^*,\xi) \\ b(\nabla f(x^*,\xi)^{\odot 2} - S(x^*)) \end{pmatrix}\begin{pmatrix} a\nabla f(x^*,\xi) \\ b(\nabla f(x^*,\xi)^{\odot 2} - S(x^*)) \end{pmatrix}^T\right] \\ 0 & \end{pmatrix}. \tag{2.16}$$

Assumption 2.5.4. The following holds.

i) There exist $\kappa \in (0,1]$, $\gamma_0 > 0$, s.t. the sequence (γ_n) satisfies $\gamma_n = \gamma_0/(n+1)^\kappa$ for all n. If $\kappa = 1$, we assume moreover that $\gamma_0 > \frac{1}{2L}$.

ii) The sequences $\left(\frac{1}{\gamma_n}(\frac{1-\alpha_n}{\gamma_n} - a)\right)$ and $\left(\frac{1}{\gamma_n}(\frac{1-\beta_n}{\gamma_n} - b)\right)$ are bounded.

For an arbitrary sequence (X_n) of random variables on some Euclidean space, a probability measure μ on that space and an event Γ s.t. $\mathbb{P}(\Gamma) > 0$, we say that X_n converges in distribution to μ *given* Γ if the measures $\mathbb{P}(X_n \in \cdot\,|\Gamma)$ converge weakly to μ.

Theorem 2.8. *Let Assumptions 2.2.1, 2.2.3, 2.4.1, 2.5.3 and 2.5.4 hold true. Let Assumption 2.4.2 ii) hold with $p = 4$. Consider the iterates $z_n = (x_n, m_n, v_n)$ given by Algorithm 2.2. Set $z^* = (x^*, 0, S(x^*))$. Set $\zeta := 0$ if $0 < \kappa < 1$ and $\zeta := \frac{1}{2\gamma_0}$ if $\kappa = 1$. Assume $\mathbb{P}(z_n \to z^*) > 0$. Then, given the event $\{z_n \to z^*\}$, the rescaled vector $\sqrt{\gamma_n}^{-1}(z_n - z^*)$ converges in distribution to a zero mean Gaussian distribution on $\mathbb{R}^{3d}$ with a covariance matrix Σ which is solution to the Lyapunov equation: $\left(H + \zeta I_{3d}\right)\Sigma + \Sigma\left(H^T + \zeta I_{3d}\right) = -Q$. In particular, given $\{z_n \to z^*\}$, the vector $\sqrt{\gamma_n}^{-1}(x_n - x^*)$ converges in distribution to a zero mean Gaussian distribution with a covariance matrix Σ_1 given by:*

$$\Sigma_1 = D^{1/2}P\left(\frac{C_{k,\ell}}{(1-\frac{2\zeta}{a})(\lambda_k + \lambda_\ell - 2\zeta + \frac{2}{a}\zeta^2) + \frac{1}{2(a-2\zeta)}(\lambda_k - \lambda_\ell)^2}\right)_{k,\ell=1\ldots d} P^{-1}D^{1/2} \tag{2.17}$$

where $C := P^{-1}D^{1/2}\mathbb{E}\left(\nabla f(x^*,\xi)\nabla f(x^*,\xi)^T\right)D^{1/2}P$.

The following remarks are useful.

- The variable v_n has an impact on the limiting covariance Σ_1 through its limit $S(x^*)$ (used to define D), but the fluctuations of v_n and the parameter b have no effect on Σ_1. As a matter of fact, Σ_1 coincides with the limiting covariance matrix that would have been obtained by considering iterates of the form

$$\begin{cases} x_{n+1} &= x_n - \gamma_{n+1} p_{n+1} \\ p_{n+1} &= p_n + a\gamma_{n+1}(D\nabla f(x_n, \xi_{n+1}) - p_n)\,, \end{cases}$$

  which can be interpreted as a preconditioned version of the stochastic heavy ball algorithm (Gadat et al., 2018). Of course, the above iterates are not implementable because the preconditioning matrix D is unknown.

- When a is large, Σ_1 is close to the matrix $\Sigma_1^{(0)}$ obtained when letting $a \to +\infty$ in Eq. (3.10). The matrix $\Sigma_1^{(0)}$ is the solution to the Lyapunov equation

$$(D\nabla^2 F(x^*) - \zeta I_d)\Sigma_1^{(0)} + \Sigma_1^{(0)}(\nabla^2 F(x^*)D - \zeta I_d) = D\mathbb{E}\left(\nabla f(x^*, \xi)\nabla f(x^*, \xi)^T\right) D\,.$$

  The matrix $\Sigma_1^{(0)}$ can be interpreted as the asymptotic covariance matrix of the x-variable in the absence of the inertial term (that is, when one considers RmsProp instead of Adam). The matrix $\Sigma_1^{(0)}$ approximates Σ_1 in the sense that $\Sigma_1 = \Sigma_1^{(0)} + \frac{1}{a}\Delta + O(\frac{1}{a^2})$ for some symmetric matrix Δ which can be explicited. The matrix Δ is neither positive nor negative definite in general. This suggests that the question of the potential benefit of the presence of an inertial term is in general problem dependent.

- In the statement of Th. 2.8, the conditioning event $\{z_n \to z^*\}$ can be replaced by the event $\{x_n \to x^*\}$ under the additional assumption that $\sum_n \gamma_n^2 < +\infty$.

2.6 Related works

Although the idea of adapting the (per-coordinate) learning rates as a function of past gradient values is not new (see *e.g.* variable metric methods such as the BFGS algorithms), AdaGrad (Duchi et al., 2011) led the way to a new class of algorithms that are sometimes referred to as adaptive gradient methods. AdaGrad consists of dividing the learning rate by the square root of the sum of previous gradients squared componentwise. The idea was to give larger learning rates to highly informative but infrequent features instead of using a fixed predetermined schedule. However, in practice, the division by the cumulative sum of squared gradients may generate small learning rates, thus freezing the iterates too early. Several works proposed heuristical ways to set the learning rates using a less aggressive policy. Tieleman and Hinton (2012) introduced an unpublished, yet popular, algorithm referred to as RmsProp where the cumulative sum used in AdaGrad is replaced by a moving average of squared gradients. Adam combines the advantages of both AdaGrad, RmsProp and inertial methods.

As opposed to AdaGrad, for which theoretical convergence guarantees exist (Duchi et al., 2011; Chen et al., 2019; Zhou et al., 2018; Ward et al., 2019a; Traoré and Pauwels, 2021), Adam is comparatively less studied. The initial paper of Kingma and Ba (2015) suggests a $\mathcal{O}(\frac{1}{\sqrt{T}})$ average regret bound in the convex setting, but Reddi et al. (2018) exhibit a counterexample in contradiction with this statement. The latter counterexample implies that the average regret bound of Adam does not converge to

zero. A first way to overcome the problem is to modify the ADAM iterations themselves in order to obtain a vanishing average regret. This led Reddi et al. (2018) to propose a variant called AMSGRAD with the aim to recover, at least in the convex case, the sought guarantees. Balles and Hennig (2018) interpret ADAM as a variance-adapted sign descent combining an update direction given by the sign and a magnitude controlled by a variance adaptation principle. A "noiseless" version of ADAM is considered in Basu et al. (2018). Under quite specific values of the ADAM-hyperparameters, it is shown that for every $\delta > 0$, there exists some time instant for which the norm of the gradient of the objective at the current iterate is no larger than δ. The recent work of Chen et al. (2019) provides a similar result for AMSGRAD and ADAGRAD, but the generalization to ADAM is subject to conditions which are not easily verifiable. Zaheer et al. (2018) provide a convergence result for RMSPROP using the objective function F as a Lyapunov function. However, our work suggests that unlike RMSPROP, ADAM does not admit F as a Lyapunov function. This makes the approach of Zaheer et al. (2018) hardly generalizable to ADAM. Moreover, Zaheer et al. (2018) consider biased gradient estimates instead of the debiased estimates used in ADAM.

In the present work, we study the behavior of an ODE, interpreted as the limit in probability of the (interpolated) ADAM iterates as the stepsize tends to zero. Closely related continuous-time dynamical systems are also studied in Attouch et al. (2000); Cabot et al. (2009). We leverage the idea of approximating a discrete time stochastic system by a deterministic continuous one, often referred to as the ODE method. The recent work of Gadat et al. (2018) fruitfully exploits this method to study a stochastic version of the celebrated heavy ball algorithm. We refer to Davis et al. (2020) for the reader interested in the non-differentiable setting with an analysis of the stochastic subgradient algorithm for non-smooth non-convex objective functions.

Concomitant to the paper on which this chapter is based, Belotto da Silva and Gazeau (31 Oct 2018) (posted only four weeks after the first version of the present work) study the asymptotic behavior of a similar dynamical system as the one introduced here. They establish several results in continuous time, such as avoidance of traps as well as convergence rates in the convex case; such aspects are out of the scope of this chapter. However, the question of the convergence of the (discrete-time) iterates is left open. In the current chapter, we also exhibit a Lyapunov function which allows, amongst others, to draw useful conclusions on the effect of the debiasing step of ADAM. Finally, Belotto da Silva and Gazeau (31 Oct 2018) study a slightly modified version of ADAM allowing to recover an ODE with a locally Lipschitz continuous vector field, whereas the original ADAM algorithm (Kingma and Ba, 2015) leads to an ODE with an irregular vector field. This technical issue is tackled in the present chapter.

2.7 Proofs for Section 2.3

2.7.1 Preliminaries

The results in this section are not specific to the case where F and S are defined as in Eq. (2.2): they are stated for *any* mappings F, S satisfying the following hypotheses.

Assumption 2.7.1. The function $F : \mathbb{R}^d \to \mathbb{R}$ is s.t.: F is continuously differentiable and ∇F is locally Lipschitz continuous.

Assumption 2.7.2. The map $S : \mathbb{R}^d \to [0, +\infty)^d$ is locally Lipschitz continuous.

In the sequel, we consider the following generalization of Eq. (ODE) for any $\eta > 0$:

$$\dot{z}(t) = h(t+\eta, z(t))\,. \tag{ODE$_\eta$}$$

When $\eta = 0$, Eq. (ODE$_\eta$) boils down to the equation of interest (ODE). The choice $\eta \in (0, +\infty)$ will be revealed useful to prove Th. 2.1. Indeed, for $\eta > 0$, a solution to Eq. (ODE$_\eta$) can be shown to exist (on some interval) due to the continuity of the map $h(\,.\,+\eta,\,.\,)$. Considering a family of such solutions indexed by $\eta \in (0, 1]$, the idea is to prove the existence of a solution to (ODE) as a cluster point of the latter family when $\eta \downarrow 0$. Indeed, as the family is shown to be equicontinuous, such a cluster point does exist thanks to the Arzelà-Ascoli theorem. When $\eta = +\infty$, Eq. (ODE$_\eta$) rewrites

$$\dot{z}(t) = h_\infty(z(t))\,, \tag{ODE$_\infty$}$$

where $h_\infty(z) := \lim_{t\to\infty} h(t, z)$. It is useful to note that for $(x, m, v) \in \mathcal{Z}_+$,

$$h_\infty((x, m, v)) = \Big(-m/(\varepsilon + \sqrt{v})\,,\, a(\nabla F(x) - m)\,,\, b(S(x) - v)\Big)\,. \tag{2.18}$$

Contrary to Eq. (ODE), Eq. (ODE$_\infty$) defines an autonomous ODE. The latter admits a unique global solution for any initial condition in $\mathcal{Z}_+$, and defines a dynamical system $\mathcal{D}$. We shall exhibit a strict Lyapunov function for this dynamical system $\mathcal{D}$, and deduce that any solution to (ODE$_\infty$) converges to the set of equilibria of $\mathcal{D}$ as $t \to \infty$. On the otherhand, we will prove that the solution to (ODE) with a proper initial condition is a so-called asymptotic pseudotrajectory (APT) of $\mathcal{D}$. Due to the existence of a strict Lyapunov function, the APT shall inherit the convergence behavior of the autonomous system as $t \to \infty$, which will prove Th. 2.2.

It is convenient to extend the map $h : (0, +\infty) \times \mathcal{Z}_+ \to \mathcal{Z}$ on $(0, +\infty) \times \mathcal{Z} \to \mathcal{Z}$ by setting $h(t, (x, m, v)) := h(t, (x, m, |v|))$ for every $t > 0$, $(x, m, v) \in \mathcal{Z}$. Similarly, we extend h_∞ as $h_\infty((x, m, v)) := h_\infty((x, m, |v|))$. For any $T \in (0, +\infty]$ and any $\eta \in [0, +\infty]$, we say that a map $z : [0, T) \to \mathcal{Z}$ is a solution to (ODE$_\eta$) on $[0, T)$ with initial condition $z_0 \in \mathcal{Z}_+$, if z is continuous on $[0, T)$, continuously differentiable on $(0, T)$, and if (ODE$_\eta$) holds for all $t \in (0, T)$. When $T = +\infty$, we say that the solution is global. We denote by $Z_T^\eta(z_0)$ the subset of $C([0, T), \mathcal{Z})$ formed by the solutions to (ODE$_\eta$) on $[0, T)$ with initial condition z_0. For any $K \subset \mathcal{Z}_+$, we define $Z_T^\eta(K) := \bigcup_{z\in K} Z_T^\eta(z)$.

Lemma 2.9. Let Assumptions 2.7.1 and 2.7.2 hold. Consider $x_0 \in \mathbb{R}^d$, $T \in (0, +\infty]$ and let $z \in Z_T^0((x_0, 0, 0))$, which we write $z(t) = (x(t), m(t), v(t))$. Then, z is continuously differentiable on $[0, T)$, $\dot{m}(0) = a\nabla F(x_0)$, $\dot{v}(0) = bS(x_0)$ and $\dot{x}(0) = \frac{-\nabla F(x_0)}{\varepsilon + \sqrt{S(x_0)}}$.

Proof. By definition of $z(\,.\,)$, $m(t) = \int_0^t a(\nabla F(x(s)) - m(s))ds$ for all $t \in [0, T)$ (and a similar relation holds for $v(t)$). The integrand being continuous, it holds that m and v are differentiable at zero and $\dot{m}(0) = a\nabla F(x_0)$, $\dot{v}(0) = bS(x_0)$. Similarly, $x(t) = x_0 + \int_0^t h_x(s, z(s))ds$, where $h_x(s, z(s)) := -(1 - e^{-as})^{-1} m(s)/(\varepsilon + \sqrt{(1 - e^{-bs})^{-1} v(s)})$. Note that $m(s)/s \to \dot{m}(0) = a\nabla F(x_0)$ as $s \downarrow 0$. Thus, $(1 - e^{-as})^{-1} m(s) \to \nabla F(x_0)$ as $s \to 0$. Similarly, $(1 - e^{-bs})^{-1} v(s) \to S(x_0)$. It follows that $h_x(s, z(s)) \to -(\varepsilon + \sqrt{S(x_0)})^{-1} \nabla F(x_0)$. Thus, $s \mapsto h_x(s, z(s))$ can be extended to a continuous map on $[0, T) \to \mathbb{R}^d$ and the differentiability of x at zero follows. ∎

Lemma 2.10. Let Assumptions 2.2.3, 2.7.1 and 2.7.2 hold. For every $\eta \in [0, +\infty]$, $T \in (0, +\infty]$, $z_0 \in \mathcal{Z}_+$, $z \in Z_T^\eta(z_0)$, it holds that $z((0, T)) \subset \mathcal{Z}_+^*$.

Proof. Set $z(t) = (x(t), m(t), v(t))$ for all t. Consider $i \in \{1, \dots, d\}$. Assume by contradiction that there exists $t_0 \in (0, T)$ s.t. $v_i(t_0) < 0$. Set $\tau := \sup\{t \in [0, t_0] : v_i(t) \geq 0\}$. Clearly, $\tau < t_0$ and $v_i(\tau) = 0$ by the continuity of v_i. Since $v_i(t) \leq 0$ for all $t \in (\tau, t_0]$, it holds that $\dot{v}_i(t) = b(S_i(x(t)) - v_i(t))$ is nonnegative for all $t \in (\tau, t_0]$. This contradicts the fact that $v_i(\tau) > v_i(t_0)$. Thus, $v_i(t) \geq 0$ for all $t \in [0, T)$. Now assume by contradiction that there exists $t \in (0, T)$ s.t. $v_i(t) = 0$. Then, $\dot{v}_i(t) = bS_i(x(t)) > 0$. Thus, $\lim_{\delta \downarrow 0} \frac{v_i(t-\delta)}{-\delta} = bS_i(x(t))$. In particular, there exists $\delta > 0$ s.t. $v_i(t - \delta) \leq -\frac{\delta b}{2} S_i(x(t))$. This contradicts the first point. ∎

Recall the definitions of V and U from Eqs. (2.9) and (2.10). Clearly, $U_\infty(v) := \lim_{t \to \infty} U(t, v) = a(\varepsilon + \sqrt{v})$ is well defined for every $v \in [0, +\infty)^d$. Hence, we can also define $V_\infty(z) := \lim_{t \to \infty} V(t, z)$ for every $z \in \mathcal{Z}_+$.

Lemma 2.11. Let Assumptions 2.7.1 and 2.7.2 hold. Assume that $0 < b \leq 4a$. Consider $(t, z) \in (0, +\infty) \times \mathcal{Z}_+^*$ and set $z = (x, m, v)$. Then, V and V_∞ are differentiable at points (t, z) and z respectively. Moreover, $\langle \nabla V_\infty(z), h_\infty(z) \rangle \leq -\varepsilon \left\| \frac{am}{U_\infty(v)} \right\|^2$ and

$$\langle \nabla V(t, z), (1, h(t, z)) \rangle \leq -\frac{\varepsilon}{2} \left\| \frac{a\, m}{U(t, v)} \right\|^2 .$$

Proof. We only prove the second point, the proof of the first point follows the same line. Consider $(t, z) \in (0, +\infty) \times \mathcal{Z}_+^*$. We decompose $\langle \nabla V(t, z), (1, h(t, z)) \rangle = \partial_t V(t, z) + \langle \nabla_z V(t, z), h(t, z) \rangle$. After tedious but straightforward derivations, we get:

$$\partial_t V(t, z) = -\sum_{i=1}^d \frac{a^2 m_i^2}{U(t, v_i)^2} \left(\frac{e^{-at}\varepsilon}{2} + \left(\frac{e^{-at}}{2} - \frac{b e^{-bt}(1 - e^{-at})}{4a(1 - e^{-bt})} \right) \sqrt{\frac{v_i}{1 - e^{-bt}}} \right) , \tag{2.19}$$

where $U(t, v_i) = a(1 - e^{-at}) \left(\varepsilon + \sqrt{\frac{v_i}{1 - e^{-bt}}} \right)$ and $\langle \nabla_z V(t, z), h(t, z) \rangle$ is equal to:

$$\sum_{i=1}^d \frac{-a^2 m_i^2 (1 - e^{-at})}{U(t, v_i)^2} \left(\varepsilon + (1 - \frac{b}{4a}) \sqrt{\frac{v_i}{1 - e^{-bt}}} + \frac{b S_i(x)}{4a \sqrt{v_i (1 - e^{-bt})}} \right) .$$

Using that $S_i(x) \geq 0$, we obtain:

$$\langle \nabla V(t, z), (1, h(t, z)) \rangle \leq -\sum_{i=1}^d \frac{a^2 m_i^2}{U(t, v_i)^2} \left((1 - \frac{e^{-at}}{2}) \varepsilon + c_{a,b}(t) \sqrt{\frac{v_i}{1 - e^{-bt}}} \right) , \tag{2.20}$$

where $c_{a,b}(t) := 1 - \frac{e^{-at}}{2} - \frac{b}{4a} \frac{1 - e^{-at}}{1 - e^{-bt}}$. Using inequality $1 - e^{-at}/2 \geq 1/2$ in (2.20), the inequality (2.20) proves the lemma, provided that one is able to show that $c_{a,b}(t) \geq 0$, for all $t > 0$ and all a, b satisfying $0 < b \leq 4a$. We prove this last statement. It can be shown that the function $b \mapsto c_{a,b}(t)$ is decreasing on $[0, +\infty)$. Hence, $c_{a,b}(t) \geq c_{a,4a}(t)$. Now, $c_{a,4a}(t) = q(e^{-at})$ where $q : [0, 1) \to \mathbb{R}$ is the function defined for all $y \in [0, 1)$ by $q(y) = y \left(y^4 - 2y^3 + 1 \right) / (2(1 - y^4))$. Hence $q \geq 0$. Thus, $c_{a,b}(t) \geq q(e^{-at}) \geq 0$. ∎

2.7.2 Proof of Th. 2.1

2.7.2.1 Boundedness

Define $\mathcal{Z}_0 := \{(x,0,0) : x \in \mathbb{R}^d\}$. Let $\bar{e} : (0,+\infty) \times \mathcal{Z}_+ \to \mathcal{Z}_+$ be defined for every $t > 0$ and every $z = (x,m,v)$ in $\mathcal{Z}_+$ by:

$$\bar{e}(t,z) := (x, m/(1-e^{-at}), v/(1-e^{-bt}))\,.$$

Proposition 2.12. Let Assumptions 2.2.2, 2.2.3, 2.7.1 and 2.7.2 hold. Assume that $0 < b \leq 4a$. For every $z_0 \in \mathcal{Z}_0$, there exists a compact set $K \subset \mathcal{Z}_+$ s.t. for all $\eta \in [0,+\infty)$, all $T \in (0,+\infty]$ and all $z \in Z_T^\eta(z_0)$, $\left\{\bar{e}(t+\eta, z(t)) : t \in (0,T)\right\} \subset K$. Moreover, choosing z_0 of the form $z_0 = (x_0,0,0)$ and $z(t) = (x(t),m(t),v(t))$, it holds that $F(x(t)) \leq F(x_0)$ for all $t \in [0,T)$.

Proof. Let $\eta \in [0,+\infty)$. Consider a solution $z_\eta(t) = (x_\eta(t), m_\eta(t), v_\eta(t))$ as in the statement, defined on some interval $[0,T)$. Define $\hat{m}_\eta(t) = m_\eta(t)/(1-e^{-a(t+\eta)})$, $\hat{v}_\eta(t) = v_\eta(t)/(1-e^{-b(t+\eta)})$. By Lem. 2.10, $t \mapsto V(t+\eta, z(t))$ is continuous on $[0,T)$, and continuously differentiable on $(0,T)$. By Lem. 2.11, $\dot{V}(t+\eta, z_\eta(t)) \leq 0$ for all $t > 0$. As a consequence, $t \mapsto V(t+\eta, z_\eta(t))$ is non-increasing on $[0,T)$. Thus, for all $t \geq 0$, $F(x_\eta(t)) \leq \lim_{t' \downarrow 0} V(t'+\eta, z_\eta(t'))$. Note that:

$$V(t+\eta, z_\eta(t)) \leq F(x_\eta(t)) + \frac{1}{2}\sum_{i=1}^d \frac{m_{\eta,i}(t)^2}{a(1-e^{-a(t+\eta)})\varepsilon}\,.$$

If $\eta > 0$, every term in the sum in the righthand side tends to zero, upon noting that $m_\eta(t) \to 0$ as $t \to 0$. The statement still holds if $\eta = 0$. Indeed, by Lem. 2.9, for a given $i \in \{1,\dots,d\}$, there exists $\delta > 0$ s.t. for all $0 < t < \delta$, $m_{\eta,i}(t)^2 \leq 2a^2(\partial_i F(x_0))^2 t^2$ and $1 - e^{-at} \geq (at)/2$. As a consequence, each term of the sum is no larger than $4(\partial_i F(x_0))^2 t/\varepsilon$, which tends to zero as $t \to 0$. We conclude that for all $t \geq 0$, $F(x_\eta(t)) \leq F(x_0)$. In particular, $\{x_\eta(t) : t \in [0,T)\} \subset \{F \leq F(x_0)\}$, the latter set being bounded by Assumption 2.2.2.

We prove that $v_{i,\eta}(t)$ is (upper)bounded. Define $R_i := \sup S_i(\{F \leq F(x_0)\})$, which is finite by continuity of S. Assume by contradiction that the set $\{t \in [0,T) : v_{\eta,i}(t) \geq R_i + 1\}$ is non-empty, and denote its infimum by τ. By continuity of $v_{\eta,i}$, one has $v_{\eta,i}(\tau) = R_i + 1$. This by the way implies that $\tau > 0$. Hence, $\dot{v}_{\eta,i}(\tau) = b(S_i(x_\eta(\tau)) - v_{\eta,i}(\tau)) \leq -b$. This means that there exists $\tau' < \tau$ s.t. $v_{\eta,i}(\tau') > v_{\eta,i}(\tau)$, which contradicts the definition of τ. We have shown that $v_{\eta,i}(t) \leq R_i + 1$ for all $t \in (0,T)$. In particular, when $t \geq 1$, $\hat{v}_{\eta,i}(t) = v_{\eta,i}(t)/(1-e^{-bt}) \leq (R_i+1)/(1-e^{-b})$. Consider $t \in (0, 1 \wedge T)$. By the mean value theorem, there exists $\tilde{t}_\eta \in [0,t]$ s.t. $v_{\eta,i}(t) = \dot{v}_{\eta,i}(\tilde{t}_\eta)t$. Thus, $v_{\eta,i}(t) \leq bS_i(x(\tilde{t}_\eta))t \leq bR_i t$. Using that the map $y \mapsto y/(1-e^{-y})$ is increasing on $(0,+\infty)$, it holds that for all $t \in (0, 1 \wedge T)$, $\hat{v}_{\eta,i}(t) \leq bR_i/(1-e^{-b})$. We have shown that, for all $t \in (0,T)$ and all $i \in \{1,\dots,d\}$, $0 \leq \hat{v}_{\eta,i}(t) \leq M$, where $M := (1-e^{-b})^{-1}(1+b)(1+\max\{R_\ell : \ell \in \{1,\dots,d\}\})$.

As $V(t+\eta, z_\eta(t)) \leq F(x_0)$, we obtain: $F(x_0) \geq F(x_\eta(t)) + \frac{1}{2}\left\|m_\eta(t)\right\|^2_{U(t+\eta, v_\eta(t))^{-1}}$. Thus, $F(x_0) \geq \inf F + \frac{1}{2a(\varepsilon+\sqrt{M})}\|m_\eta(t)\|^2$. Therefore, $m_\eta(\,.\,)$ is bounded on $[0,T)$, uniformly in η. The same holds for $\hat{m}_\eta$ by using the mean value theorem in the same way as for $\hat{v}_\eta$. The proof is complete. ∎

Proposition 2.13. Let Assumptions 2.2.2, 2.2.3, 2.7.1 and 2.7.2 hold. Assume that $0 < b \leq 4a$. Let K be a compact subset of $\mathcal{Z}_+$. Then, there exists another compact set $K' \subset \mathcal{Z}_+$ s.t. for every $T \in (0, +\infty]$ and every $z \in Z_T^\infty(K)$, $z([0,T)) \subset K'$.

Proof. The proof follows the same line as Prop. 2.12 and is omitted. ∎

For any $K \subset \mathcal{Z}_+$, define $v_{\min}(K) := \inf\{v_k : (x,m,v) \in K, i \in \{1,\dots,d\}\}$.

Lemma 2.14. Under Assumptions 2.2.2, 2.2.3, 2.7.1 and 2.7.2, the following holds true.

i) For every compact set $K \subset \mathcal{Z}_+$, there exists $c > 0$, s.t. for every $z \in Z_\infty^\infty(K)$, of the form $z(t) = (x(t), m(t), v(t))$, $v_i(t) \geq c \min\left(1, \frac{v_{\min}(K)}{2c} + t\right) \qquad (\forall t \geq 0, \forall i \in \{1,\dots,d\})$.

ii) For every $z_0 \in \mathcal{Z}_0$, there exists $c > 0$ s.t. for every $\eta \in [0, +\infty)$ and every $z \in Z_\infty^\eta(z_0)$, $v_i(t) \geq c \min(1,t) \qquad (\forall t \geq 0, \forall i \in \{1,\dots,d\})$.

Proof. We prove the first point. Consider a compact set $K \subset \mathcal{Z}_+$. By Prop. 2.13, one can find a compact set $K' \subset \mathcal{Z}_+$ s.t. for every $z \in Z_\infty^\infty(K)$, it holds that $\{z(t) : t \geq 0\} \subset K'$. Denote by L_S the Lipschitz constant of S on the compact set $\{x : (x,m,v) \in K'\}$. Introduce the constants $M_1 := \sup\{\|m/(\varepsilon + \sqrt{v})\|_\infty : (x,m,v) \in K'\}$, $M_2 := \sup\{\|S(x)\|_\infty : (x,m,v) \in K'\}$. The constants L_S, M_1, M_2 are finite. Now consider a global solution $z(t) = (x(t), m(t), v(t))$ in $Z_\infty^\infty(K)$. Choose $i \in \{1,\dots,d\}$ and consider $t \geq 0$. By the mean value theorem, there exists $t' \in [0,t]$ s.t. $v_i(t) = v_i(0) + \dot{v}_i(t')t$. Thus, $v_i(t) = v_i(0) + \dot{v}_i(0)t + b(S_i(x(t')) - v_i(t') - S_i(x(0)) + v_i(0))t$, which in turn implies $v_i(t) \geq v_i(0) + \dot{v}_i(0)t - bL_S\|x(t') - x(0)\|t - b|v_i(t') - v_i(0)|t$. Using again the mean value theorem, for every $\ell \in \{1,\dots,d\}$, there exists $t'' \in [0,t']$ s.t. $|x_\ell(t') - x_\ell(0)| = t'|\dot{x}_\ell(t'')| \leq tM_1$. Therefore, $\|x(t') - x(0)\| \leq \sqrt{d}M_1 t$. Similarly, there exists $\tilde{t}$ s.t.: $|v_i(t') - v_i(0)| = t'|\dot{v}_i(\tilde{t})| \leq t'bS_i(x(\tilde{t})) \leq tbM_2$. Putting together the above inequalities, $v_i(t) \geq v_i(0)(1-bt) + bS_i(x(0))t - bCt^2$, where $C := (M_2 + L_S\sqrt{d}M_1)$. For every $t \leq 1/(2b)$, $v_i(t) \geq \frac{v_{\min}}{2} + tbC\left(\frac{S_{\min}}{C} - t\right)$, where we defined $S_{\min} := \inf\{S_i(x) : i \in \{1,\dots,d\}, (x,m,v) \in K\}$. Setting $\tau := 0.5\min(1/b, S_{\min}/C)$,

$$\forall t \in [0,\tau],\ v_i(t) \geq \frac{v_{\min}}{2} + \frac{bS_{\min}t}{2}\,. \tag{2.21}$$

Set $\kappa_1 := 0.5(v_{\min} + bS_{\min}\tau)$. Note that $v_i(\tau) \geq \kappa_1$. Define $S'_{\min} := \inf\{S_i(x) : i \in \{1,\dots,d\}, (x,m,v) \in K'\}$. Note that $S'_{\min} > 0$ by Assumptions 2.7.2 and 2.2.3. Finally, define $\kappa = 0.5\min(\kappa_1, S'_{\min})$. By contradiction, assume that the set $\{t \geq \tau : v_i(t) < \kappa\}$ is non-empty, and denote by τ' its infimum. It is clear that $\tau' > \tau$ and $v_i(\tau') = \kappa$. Thus, $b^{-1}\dot{v}_i(\tau') = S_i(x(\tau')) - \kappa$. We obtain that $b^{-1}\dot{v}_i(\tau') \geq 0.5S'_{\min} > 0$. As a consequence, there exists $t \in (\tau, \tau')$ s.t. $v_i(t) < v_i(\tau')$. This contradicts the definition of τ'. We have shown that for all $t \geq \tau$, $v_i(t) \geq \kappa$. Putting this together with Eq. (2.21) and using that $\kappa \leq v_{\min} + bS_{\min}\tau$, we conclude that: $\forall t \geq 0,\ v_i(t) \geq \min\left(\kappa, \frac{v_{\min}}{2} + \frac{bS_{\min}t}{2}\right)$. Setting $c := \min(\kappa, bS_{\min}/2)$, the result follows.

We prove the second point. By Prop. 2.12, there exists a compact set $K \subset \mathcal{Z}_+$ s.t. for every $\eta \geq 0$, every $z \in Z_\infty^\eta(x_0)$ of the form $z(t) = (x(t), m(t), v(t))$ satisfies $\{(x(t), \hat{m}(t), \hat{v}(t)) : t \geq 0\} \subset K$, where $\hat{m}(t) = m(t)/(1 - e^{-a(t+h)})$ and $\hat{v}(t) = v(t)/(1 -$

$e^{-b(t+h)}$). Denote by L_S the Lipschitz constant of S on the set $\{x : (x,m,v) \in K\}$. Introduce the constants $M_1 := \sup\{\|m/(\varepsilon + \sqrt{v})\|_\infty : (x,m,v) \in K\}$, $M_2 := \sup\{\|S(x)\|_\infty : (x,m,v) \in K'\}$. These constants being introduced, the rest of the proof follows the same line as the proof of the first point. ■

2.7.2.2 Existence

Corollary 2.15. Let Assumptions 2.2.2, 2.2.3, 2.7.1 and 2.7.2 hold. Assume that $0 < b \leq 4a$. For every $z_0 \in \mathcal{Z}_+$,$Z^\infty_\infty(z_0) \neq \emptyset$. For every $(z_0,\eta) \in \mathcal{Z}_0 \times (0,+\infty)$,$Z^\eta_\infty(z_0) \neq \emptyset$.

Proof. We prove the first point (the proof of the second point follows the same line). Under Assumptions 2.7.1 and 2.7.2, h_∞ is continuous. Therefore, Cauchy-Peano's theorem guarantees the existence of a solution to the (ODE) issued from z_0, which we can extend over a maximal interval of existence $[0, T_{\max})$.We conclude that the solution is global ($T_{\max} = +\infty$) using the boundedness of the solution given by Prop. 2.13. ■

Lemma 2.16. Let Assumptions 2.2.2, 2.2.3, 2.7.1 and 2.7.2 hold. Assume that $0 < b \leq 4a$. Consider $z_0 \in \mathcal{Z}_0$. Denote by $(z_\eta : \eta \in (0,+\infty))$ a family of functions on $[0,+\infty) \to \mathcal{Z}_+$ s.t. for every $\eta > 0$, $z_\eta \in Z^\eta_\infty(z_0)$. Then, $(z_\eta)_{\eta>0}$ is equicontinuous.

Proof. For every such solution z_η, we set $z_\eta(t) = (x_\eta(t), m_\eta(t), v_\eta(t))$ for all $t \geq 0$, and define $\hat{m}_\eta$ and $\hat{v}_\eta$ as in Prop. 2.12. By Prop. 2.12, there exists a constant M_1 s.t. for all $\eta > 0$ and all $t \geq 0$, $\max(\|x_\eta(t)\|, \|\hat{m}_\eta(t)\|_\infty, \|\hat{v}_\eta(t)\|) \leq M_1$. Using the continuity of ∇F and S, there exists another finite constant M_2 s.t. $M_2 \geq \sup\{\|\nabla F(x)\|_\infty : x \in \mathbb{R}^d, \|x\| \leq M_1\}$ and $M_2 \geq \sup\{\|S(x)\|_\infty : x \in \mathbb{R}^d, \|x\| \leq M_1\}$. For every $(s,t) \in [0,+\infty)^2$, we have for all $i \in \{1,\dots,d\}$,

$$
\begin{aligned}
|x_{\eta,i}(t) - x_{\eta,i}(s)| &\leq \int_s^t \left| \frac{\hat{m}_{\eta,i}(u)}{\varepsilon + \sqrt{\hat{v}_{\eta,i}(u)}} \right| du \;\leq \frac{M_1}{\varepsilon}|t-s|\,,\\
|m_{\eta,i}(t) - m_{\eta,i}(s)| &\leq \int_s^t a \left| \partial_i F(x_\eta(u)) - m_{\eta,i}(u) \right| du \;\leq a(M_1+M_2)|t-s|\,,\\
|v_{\eta,i}(t) - v_{\eta,i}(s)| &\leq \int_s^t b \left| S_i(x_\eta(u)) - v_{\eta,i}(u) \right| du \;\leq b(M_1+M_2)|t-s|\,.
\end{aligned}
$$

Therefore, there exists a constant M_3, independent from η, s.t. for all $\eta > 0$ and all $(s,t) \in [0,+\infty)^2$, $\|z_\eta(t) - z_\eta(s)\| \leq M_3|t-s|$. ■

Proposition 2.17. Let Assumptions 2.2.2, 2.2.3, 2.7.1 and 2.7.2 hold. Assume that $0 < b \leq 4a$. For every $z_0 \in \mathcal{Z}_0$, $Z^0_\infty(z_0) \neq \emptyset$ *i.e.*, (ODE) admits a global solution issued from z_0.

Proof. By Cor. 2.15, there exists a family $(z_\eta)_{\eta>0}$ of functions on $[0,+\infty) \to \mathcal{Z}$ s.t. for every $\eta > 0$, $z_\eta \in Z^\eta_\infty(z_0)$. We set as usual $z_\eta(t) = (x_\eta(t), m_\eta(t), v_\eta(t))$. By Lem. 2.16, and the Arzelà-Ascoli theorem, there exists a map $z : [0,+\infty) \to \mathcal{Z}$ and a sequence $\eta_n \downarrow 0$ s.t. z_{η_n} converges to z uniformly on compact sets, as $n \to \infty$. Considering some

fixed scalars $t > s > 0$, $z(t) = z(s) + \lim_{n\to\infty} \int_s^t h(u+\eta_n, z_{\eta_n}(u))du$. By Prop. 2.12, there exists a compact set $K \subset \mathcal{Z}_+$ s.t. $\{z_{\eta_n}(t) : t \geq 0\} \subset K$ for all n. Moreover, by Lem. 2.14, there exists a constant $c > 0$ s.t. for all n and all $u \geq 0$, $v_{\eta_n,k}(u) \geq c\min(1,u)$. Denote by $\bar{K} := K \cap (\mathbb{R}^d \times \mathbb{R}^d \times [c\min(1,s), +\infty)^d)$. It is clear that $\bar{K}$ is a compact subset of $\mathcal{Z}_+^*$. Since h is continuously differentiable on the set $[s,t] \times \bar{K}$, it is Lipschitz continuous on that set. Denote by L_h the corresponding Lipschitz constant. We obtain:

$$\int_s^t \|h(u+\eta_n, z_{\eta_n}(u)) - h(u, z(u))\|du \leq L_h \left(\eta_n + \sup_{u\in[s,t]} \|z_{\eta_n}(u) - z(u)\|\right)(t-s)\,,$$

and the righthand side converges to zero. As a consequence, for all $t > s$, $z(t) = z(s) + \int_s^t h(u, z(u))du$. Moreover, $z(0) = z_0$. This proves that $z \in Z_\infty^0(z_0)$. ■

2.7.2.3 Uniqueness

Proposition 2.18. Let Assumptions 2.2.2, 2.2.3, 2.7.1 and 2.7.2 hold. Assume $b \leq 4a$.

i) For every $z_0 \in \mathcal{Z}_0$, $Z_\infty^0(z_0)$ is a singleton *i.e.*, there exists a unique global solution to (ODE) with initial condition z_0.

ii) For every compact subset K of $\mathcal{Z}_+$, there exist nonnegative constants c_1, c_2 s.t. for every $(z, z') \in Z_\infty^\infty(K)^2$,

$$\forall t \geq 0,\ \|z(t) - z'(t)\|^2 \leq \|z(0) - z'(0)\|^2 \exp(c_1 + c_2 t)\,.$$

Proof. i) Consider solutions z and z' in $Z_\infty^0(z_0)$. We denote by $(x(t), m(t), v(t))$ the blocks of $z(t)$, and we define $(x'(t), m'(t), v'(t))$ similarly. For all $t > 0$, we define $\hat{m}(t) := m(t)/(1-e^{-at})$, $\hat{v}(t) := v(t)/(1-e^{-bt})$, and we define $\hat{m}'(t)$ and $\hat{v}'(t)$ similarly. By Prop. 2.12, there exists a compact set $K \subset \mathcal{Z}_+$ s.t. $(x(t), \hat{m}(t), \hat{v}(t))$ and $(x'(t), \hat{m}'(t), \hat{v}'(t))$ are both in K for all $t > 0$. We denote by L_S and $L_{\nabla F}$ the Lipschitz constants of S and ∇F on the compact set $\{x : (x, m, v) \in K\}$. These constants are finite by Assumptions 2.7.1 and 2.7.2. We define $M := \sup\{\|m\|_\infty : (x, m, v) \in K\}$. Define $u_x(t) := \|x(t) - x'(t)\|^2$, $u_m(t) := \|\hat{m}(t) - \hat{m}'(t)\|^2$ and $u_v(t) := \|\hat{v}(t) - \hat{v}'(t)\|^2$. Let $\delta > 0$. Define: $u^{(\delta)}(t) := u_x(t) + \delta u_m(t) + \delta u_v(t)$. By the chain rule and the Cauchy-Schwarz inequality,

$$\begin{aligned}
\dot{u}_x(t) &\leq 2\|x(t) - x'(t)\| \left\| \frac{\hat{m}(t)}{\varepsilon + \sqrt{\hat{v}(t)}} - \frac{\hat{m}'(t)}{\varepsilon + \sqrt{\hat{v}'(t)}} \right\| \\
&\leq 2\|x(t) - x'(t)\| \left(\varepsilon^{-1} \left\| \hat{m}(t) - \hat{m}'(t) \right\| + M\varepsilon^{-2} \left\| \sqrt{\hat{v}(t)} - \sqrt{\hat{v}'(t)} \right\| \right)\,.
\end{aligned}$$

For every $i \in \{1, \ldots, d\}$, $\left|\sqrt{\hat{v}_i(t)} - \sqrt{\hat{v}_i'(t)}\right| = \frac{|\hat{v}_i(t) - \hat{v}_i'(t)|}{|\sqrt{\hat{v}_i(t)} + \sqrt{\hat{v}_i'(t)}|}$. By Lem. 2.14, there exists $c > 0$ s.t. for all $t \geq 0$, for every $i \in \{1, \ldots, d\}$, $\hat{v}_i(t) \wedge \hat{v}_i'(t) \geq c\min(1, t)$. Thus,

$$\dot{u}_x(t) \leq 2\|x(t) - x'(t)\| \left(\varepsilon^{-1} \left\| \hat{m}(t) - \hat{m}'(t) \right\| + \frac{M}{2\varepsilon^2 \sqrt{c\min(1,t)}} \left\| \hat{v}(t) - \hat{v}'(t) \right\| \right)\,.$$

For any $\delta > 0$, $2\|x(t) - x'(t)\| \, \|\hat{m}(t) - \hat{m}'(t)\| \leq \delta^{-1/2}(u_x(t) + \delta u_m(t)) \leq \delta^{-1/2} u^{(\delta)}(t)$. Similarly, $2\|x(t) - x'(t)\| \, \|\hat{v}(t) - \hat{v}'(t)\| \leq \delta^{-1/2} u^{(\delta)}(t)$. Thus, for any $\delta > 0$,

$$\dot{u}_x(t) \leq \left(\frac{1}{\varepsilon\sqrt{\delta}} + \frac{M}{2\varepsilon^2 \sqrt{\delta c \min(1,t)}} \right) u^{(\delta)}(t) \,. \tag{2.22}$$

We now study $u_m(t)$. For all $t > 0$, we obtain after some algebra: $\frac{d}{dt}\hat{m}(t) = a(\nabla F(x(t)) - \hat{m}(t))/(1 - e^{-at})$. Therefore,

$$\begin{aligned} \dot{u}_m(t) &= \frac{2a}{1 - e^{-at}} \langle \hat{m}(t) - \hat{m}'(t), \nabla F(x(t)) - \hat{m}(t) - \nabla F(x'(t)) + \hat{m}'(t) \rangle \\ &\leq \frac{2aL_{\nabla F}}{1 - e^{-at}} \|\hat{m}(t) - \hat{m}'(t)\| \, \|x(t) - x'(t)\| \,. \end{aligned}$$

For any $\theta > 0$, it holds that $2\|\hat{m}(t) - \hat{m}'(t)\| \, \|x(t) - x'(t)\| \leq \theta u_x(t) + \theta^{-1} u_m(t)$. In particular, letting $\theta := 2L_{\nabla F}$, we obtain that for all $\delta > 0$,

$$\delta \dot{u}_m(t) \leq \tfrac{a}{2(1-e^{-at})} \left(4\delta L_{\nabla F}^2 u_x(t) + \delta u_m(t) \right) \leq \left(\tfrac{a}{2} + \tfrac{1}{2t} \right) \left(4\delta L_{\nabla F}^2 u_x(t) + \delta u_m(t) \right) , \tag{2.23}$$

where the last inequality is due to the fact that $y/(1 - e^{-y}) \leq 1 + y$ for all $y > 0$. Using the exact same arguments, we also obtain that

$$\delta \dot{u}_v(t) \leq \left(\frac{b}{2} + \frac{1}{2t} \right) \left(4\delta L_S^2 u_x(t) + \delta u_m(t) \right) . \tag{2.24}$$

We now choose any δ s.t. $4\delta \leq 1/\max(L_S^2, L_{\nabla F}^2)$. Then, Eq. (2.23) and (2.24) respectively imply that $\delta \dot{u}_m(t) \leq 0.5(a + t^{-1}) u^{(\delta)}(t)$ and $\delta \dot{u}_v(t) \leq 0.5(b + t^{-1}) u^{(\delta)}(t)$. Summing these inequalities along with Eq. (2.22), we obtain that for every $t > 0$, $\dot{u}^{(\delta)}(t) \leq \psi(t) u^{(\delta)}(t)$, where: $\psi(t) := \frac{a+b}{2} + \frac{1}{\varepsilon\sqrt{\delta}} + \frac{M}{2\varepsilon^2\sqrt{\delta c \min(1,t)}} + \frac{1}{t}$. From Grönwall's inequality, it holds that for every $t > s > 0$, $u^{(\delta)}(t) \leq u^{(\delta)}(s) \exp\left(\int_s^t \psi(s') ds' \right)$. We first consider the case where $t \leq 1$. We set $c_1 := (a+b)/2 + (\varepsilon\sqrt{\delta})^{-1}$ and $c_2 := M/(\varepsilon^2\sqrt{\delta c})$. With these notations, $\int_s^t \psi(s') ds' \leq c_1 t + c_2\sqrt{t} + \ln \frac{t}{s}$. Therefore, $u^{(\delta)}(t) \leq \frac{u^{(\delta)}(s)}{s} \exp\left(c_1 t + c_2\sqrt{t} + \ln t \right)$. By Lem. 2.9, recall that $\dot{x}(0)$ and $\dot{x}'(0)$ are both well defined (and coincide). Thus,

$$u_x(s) = \|x(s) - x'(s)\|^2 \leq 2\|x(s) - x(0) - \dot{x}(0)s\|^2 + 2\|x'(s) - x'(0) - \dot{x}'(0)s\|^2 \,.$$

It follows that $u_x(s)/s^2$ converges to zero as $s \downarrow 0$. We now show the same kind of result for $u_m(s)$ and $u_v(s)$. Consider $i \in \{1, \dots, d\}$. By the mean value theorem, there exists $\tilde{s}$ (resp. $\tilde{s}'$) in the interval $[0, t]$ s.t. $m_i(s) = \dot{m}_i(\tilde{s})s$ (resp. $m_i'(s) = \dot{m}_i'(\tilde{s}')s$). Thus, $\hat{m}_i(s) = \frac{as}{1-e^{-as}} \left(\partial_i F(x(\tilde{s})) - m_i(\tilde{s}) \right)$, and a similar equality holds for $\hat{m}_i'(s)$. As a consequence,

$$\begin{aligned} |\hat{m}_i(s) - \hat{m}_i'(s)| &\leq \frac{as}{1 - e^{-as}} \left(|\partial_i F(x(\tilde{s})) - \partial_i F(x'(\tilde{s}'))| + |m_i(\tilde{s}) - m_i'(\tilde{s}')| \right) \\ &\leq \frac{as}{1 - e^{-as}} \left(L_{\nabla F} \|x(\tilde{s}) - x'(\tilde{s}')\| + |m_i(\tilde{s}) - m_i'(\tilde{s}')| \right) \\ &\leq \frac{2a(L_{\nabla F} \vee 1)s}{1 - e^{-as}} \|z(\tilde{s}) - z'(\tilde{s}')\| \,, \end{aligned}$$

where we used $\|x(\tilde{s}) - x'(\tilde{s}')\| \leq \|z(\tilde{s}) - z'(\tilde{s}')\|$ and $|m_i(\tilde{s}) - m'_i(\tilde{s}')| \leq \|z(\tilde{s}) - z'(\tilde{s}')\|$ to obtain the last inequality. Using that $\tilde{s} \leq s$ and $\tilde{s}' \leq s$, it follows that:

$$\frac{|\hat{m}_i(s) - \hat{m}'_i(s)|}{s} \leq \frac{2a(L_{\nabla F} \vee 1)s}{1 - e^{-as}} \left(\frac{\|z(\tilde{s}) - z(0)\|}{\tilde{s}} + \frac{\|z'(\tilde{s}') - z'(0)\|}{\tilde{s}'} \right).$$

By Lem. 2.9, z and z' are differentiable at point zero. Then, the above inequality gives $\limsup_{s \downarrow 0} \frac{|\hat{m}_i(s) - \hat{m}'_i(s)|}{s} \leq 4(L_{\nabla F} \vee 1)\|\dot{z}(0)\|$. Thus,

$$\limsup_{s \downarrow 0} \frac{u_m(s)}{s^2} \leq 16d(L^2_{\nabla F} \vee 1)\|\dot{z}(0)\|^2 .$$

Therefore, $u_m(s)/s$ converges to zero as $s \downarrow 0$. By similar arguments, it can be shown that $\limsup_{s \downarrow 0} u_v(s)/s^2 \leq 16d(L^2_S \vee 1)\|\dot{z}(0)\|^2$, thus $\lim u_v(s)/s = 0$. Finally, we obtain that $u^{(\delta)}(s)/s$ converges to zero as $s \downarrow 0$. Letting s tend to zero, we obtain that for every $t \leq 1$, $u^{(\delta)}(t) = 0$. Setting $s = 1$ and $t > 1$, and noting that ψ is integrable on $[1, t]$, it follows that $u^{(\delta)}(t) = 0$ for all $t > 1$. This proves that $z = z'$.

ii) Consider the compact set K, and introduce the compact set $K' \subset \mathcal{Z}_+$ as in Prop. 2.13, and the constant $c > 0$ defined in Lem. 2.14. Define $K'_x = \{x : (x, m, v) \in K'\}$. The set is compact in $\mathbb{R}^d$. Respectively denote by L_S and $L_{\nabla F}$ the Lipschitz constants of S and ∇F on K'_x. Introduce the constant $M := \sup\{\|m\|_\infty : (x, m, v) \in K'\}$. Consider $(z_0, z'_0) \in K^2$ and two global solutions $z(\,.\,)$ and $z'(\,.\,)$ starting at z_0 and z'_0 respectively. We denote by $(x(t), m(t), v(t))$ the blocks of $z(t)$, and we define $(x'(t), m'(t), v'(t))$ similarly. Set $u(t) := \|z(t) - z'(t)\|^2$. Set also $u_x(t) := \|x(t) - x'(t)\|^2$ and define $u_m(t)$ and $u_v(t)$ similarly, hence, $u(t) = u_x(t) + u_m(t) + u_v(t)$. Using the same derivations as above, we establish for all $t \geq 0$ that: $\dot{u}_m(t) \leq aL_{\nabla F} u_x(t) + a(L_{\nabla F} + 2)u_m(t)$. Similarly, $\dot{u}_v(t) \leq bL_S u_x(t) + b(L_S + 2)u_v(t)$. Moreover, $\dot{u}_x(t) \leq (\varepsilon^{-1} + \varepsilon^{-2}MC(t))u_x(t) + \varepsilon^{-1}u_m(t) + \varepsilon^{-2}MC(t)u_v(t)$ where we set $C(t) := \|(\sqrt{v(t)} + \sqrt{v'(t)})^{-1}\|_\infty$. Putting all pieces together, we obtain that there exist nonnegative constants c_1 and c_2, depending on K, s.t.

$$\dot{u}(t) \leq (c_1 + c_2 C(t))u(t) .$$

By Lem. 2.14, there exist two other nonnegative constants c'_1, c'_2 depending on K, s.t. for all $t > 0$, $\dot{u}(t) \leq (c'_1 + c'_2 \max(1, t^{-1/2}))u(t)$. Using Grönwall's lemma, we obtain that for all $t \geq 0$,

$$u(t) \leq u(0) \exp\left(\int_0^t (c'_1 + c'_2 \max(1, s^{-1/2}))ds \right) .$$

It is easy to show that the integral in the exponential is no larger than $2c'_2 + (c'_1 + c'_2)t$. This completes the proof. ∎

We recall that a semiflow Φ on the space (E, d) is a continuous map Φ from $[0, +\infty) \times E$ to E defined by $(t, x) \mapsto \Phi(t, x) = \Phi_t(x)$ such that Φ_0 is the identity and $\Phi_{t+s} = \Phi_t \circ \Phi_s$ for all $(t, s) \in [0, +\infty)^2$.

Proposition 2.19. Let Assumptions 2.2.2, 2.2.3, 2.7.1 and 2.7.2 hold. Assume that $0 < b \leq 4a$. The map Z^∞_∞ is single-valued on $\mathcal{Z}_+ \to C([0, +\infty), \mathcal{Z}_+)$ *i.e.*, there exists a unique global solution to (ODE_∞) starting from any given point in $\mathcal{Z}_+$. Moreover, the following map is a semiflow:

$$\begin{array}{rccc} \Phi : [0, +\infty) \times \mathcal{Z}_+ & \to & \mathcal{Z}_+ \\ (t, z) & \mapsto & Z^\infty_\infty(z)(t) \end{array} \tag{2.25}$$

Proof. The result is a direct consequence of Lem. 2.18. ■

2.7.3 Proof of Th. 2.2

2.7.3.1 Convergence of the semiflow

We first recall some useful definitions and results. Let Ψ represent any semiflow on an arbitrary metric space (E, d). A point $z \in E$ is called an *equilibrium point* of the semiflow Ψ if $\Psi_t(z) = z$ for all $t \geq 0$. We denote by Λ_Ψ the set of equilibrium points of Ψ. A continuous function $\mathsf{V} : E \to \mathbb{R}$ is called a *Lyapunov function* for the semiflow Ψ if $\mathsf{V}(\Psi_t(z)) \leq \mathsf{V}(z)$ for all $z \in E$ and all $t \geq 0$. It is called a *strict Lyapunov function* if, moreover, $\{z \in E : \forall t \geq 0, \mathsf{V}(\Psi_t(z)) = \mathsf{V}(z)\} = \Lambda_\Psi$. If V is a strict Lyapunov function for Ψ and if $z \in E$ is a point s.t. $\{\Psi_t(z) : t \geq 0\}$ is relatively compact, then it holds that $\Lambda_\Psi \neq \emptyset$ and $\mathsf{d}(\Psi_t(z), \Lambda_\Psi) \to 0$, see (Haraux, 1991, Th. 2.1.7). A continuous function $z : [0, +\infty) \to E$ is said to be an asymptotic pseudotrajectory (APT, Benaïm and Hirsch (1996)) for the semiflow Ψ if for every $T \in (0, +\infty)$, $\lim_{t\to+\infty} \sup_{s\in[0,T]} \mathsf{d}(z(t+s), \Psi_s(z(t))) = 0$.

The following result follows from (Benaïm, 1999, Th. 5.7) and (Benaïm, 1999, Prop. 6.4).

Proposition 2.20 (Benaïm (1999))**.**
Consider a semiflow Ψ on (E, d) and a map $z : [0, +\infty) \to E$. Assume the following:

i) Ψ admits a strict Lyapunov function V.
ii) The set Λ_Ψ of equilibrium points of Ψ is compact.
iii) $\mathsf{V}(\Lambda_\Psi)$ has an empty interior.
iv) z is an APT of Ψ.
v) $z([0, \infty))$ is relatively compact.

Then, $\bigcap_{t\geq 0} \overline{z([t, \infty))}$ is a compact connected subset of Λ_Ψ.

For every $\delta > 0$ and every $z = (x, m, v) \in \mathcal{Z}_+$, define:

$$W_\delta(x, m, v) := V_\infty(x, m, v) - \delta\langle \nabla F(x), m\rangle + \delta\|S(x) - v\|^2, \tag{2.26}$$

where we recall that $V_\infty(z) := \lim_{t\to\infty} V(t, z)$ for every $z \in \mathcal{Z}_+$ and V is defined by Eq.(2.9). Consider the set $\mathcal{E} := h_\infty^{-1}(\{0\})$ of all equilibrium points of (ODE$_\infty$), namely: $\mathcal{E} = \{(x, m, v) \in \mathcal{Z}_+ : \nabla F(x) = 0, m = 0, v = S(x)\}$. The set $\mathcal{E}$ is non-empty by Assumption 2.2.2.

Proposition 2.21. Let Assumptions 2.2.2, 2.2.3, 2.7.1 and 2.7.2 hold. Assume that $0 < b \leq 4a$. Let $K \subset \mathcal{Z}_+$ be a compact set. Define $K' := \overline{\{\Phi(t, z) : t \geq 0, z \in K\}}$. Let $\overline{\Phi} : [0, +\infty) \times K' \to K'$ be the restriction of the semiflow Φ to K' *i.e.*, $\overline{\Phi}(t, z) = \Phi(t, z)$ for all $t \geq 0, z \in K'$. Then,

i) K' is compact.
ii) $\overline{\Phi}$ is well defined and is a semiflow on K'.
iii) The set of equilibrium points of $\overline{\Phi}$ is equal to $\mathcal{E} \cap K'$.
iv) There exists $\delta > 0$ s.t. W_δ is a strict Lyapunov function for the semiflow $\overline{\Phi}$.

Proof. The first point is a consequence of Prop. 2.13. The second point stems from Prop. 2.19. The third point is immediate from the definition of $\mathcal{E}$ and the fact that $\overline{\Phi}$ is valued in K'. We now prove the last point. Consider $z \in K'$ and write $\overline{\Phi}_t(z)$ under

the form $\overline{\Phi}_t(z) = (x(t), m(t), v(t))$. For *any* map $\mathsf{W} : \mathcal{Z}_+ \to \mathbb{R}$, define for all $t > 0$, $\mathcal{L}_{\mathsf{W}}(t) := \limsup_{s\to 0} s^{-1}(\mathsf{W}(\overline{\Phi}_{t+s}(z)) - \mathsf{W}(\overline{\Phi}_t(z)))$. Introduce $G(z) := -\langle \nabla F(x), m\rangle$ and $H(z) := \|S(x) - v\|^2$ for every $z = (x, m, v)$. Consider $\delta > 0$ (to be specified later on). We study $\mathcal{L}_{W_\delta} = \mathcal{L}_V + \delta\mathcal{L}_G + \delta\mathcal{L}_H$. Note that $\overline{\Phi}_t(z) \in K' \cap \mathcal{Z}_+^*$ for all $t > 0$ by Lem. 2.10. Thus, $t \mapsto V_\infty(\overline{\Phi}_t(z))$ is differentiable at any point $t > 0$ and the derivative coincides with $\mathcal{L}_V(t) = \dot{V}_\infty(\overline{\Phi}_t(z))$. Define $C_1 := \sup\{\|v\|_\infty : (x, m, v) \in K'\}$. Then, by Lem. 2.11, $\mathcal{L}_V(t) \le -\varepsilon(\varepsilon + \sqrt{C_1})^{-2} \left\|m(t)\right\|^2$. Let $L_{\nabla F}$ be the Lipschitz constant of ∇F on $\{x : (x, m, v) \in K'\}$. For every $t > 0$,

$$\begin{aligned}
\mathcal{L}_G(t) &\le \limsup_{s\to 0} s^{-1}\|\nabla F(x(t)) - \nabla F(x(t+s))\|\|m(t+s)\| - \langle \nabla F(x(t)), \dot{m}(t)\rangle \\
&\le L_{\nabla F}\varepsilon^{-1}\|m(t)\|^2 - a\|\nabla F(x(t))\|^2 + a\langle \nabla F(x(t)), m(t)\rangle \\
&\le -\frac{a}{2}\|\nabla F(x(t))\|^2 + \left(\frac{a}{2} + \frac{L_{\nabla F}}{\varepsilon}\right)\|m(t)\|^2 .
\end{aligned}$$

Denote by L_S the Lipschitz constant of S on $\{x : (x, m, v) \in K'\}$. For every $t > 0$,

$$\begin{aligned}
\mathcal{L}_H(t) &= \limsup_{s\to 0} s^{-1}(\|S(x(t+s)) - S(x(t)) + S(x(t)) - v(t+s)\|^2 - \|S(x(t)) - v(t)\|^2) \\
&= -2\langle S(x(t)) - v(t), \dot{v}(t)\rangle + \limsup_{s\to 0} 2s^{-1}\langle S(x(t+s)) - S(x(t)), S(x(t)) - v(t+s)\rangle \\
&\le -2b\|S(x(t)) - v(t)\|^2 + 2L_S\varepsilon^{-1}\|m(t)\|\|S(x(t)) - v(t)\| .
\end{aligned}$$

Using that $2\|m(t)\|\|S(x(t)) - v(t)\| \le \frac{L_S}{b\varepsilon}\|m(t)\|^2 + \frac{b\varepsilon}{L_S}\|S(x(t)) - v(t)\|^2$, we obtain $\mathcal{L}_H(t) \le -b\|S(x(t)) - v(t)\|^2 + \frac{L_S^2}{b\varepsilon^2}\|m(t)\|^2$. Hence, for every $t > 0$,

$$\mathcal{L}_{W_\delta}(t) \le -M(\delta)\|m(t)\|^2 - \frac{a\delta}{2}\|\nabla F(x(t))\|^2 - \delta b\|S(x(t)) - v(t)\|^2 .$$

where $M(\delta) := \varepsilon(\varepsilon + \sqrt{C_1})^{-2} - \frac{\delta L_S^2}{b\varepsilon^2} - \delta\left(\frac{a}{2} + \frac{L_{\nabla F}}{\varepsilon}\right)$. Choosing δ s.t. $M(\delta) > 0$,

$$\forall t > 0, \quad \mathcal{L}_{W_\delta}(t) \le -c\left(\|m(t)\|^2 + \|\nabla F(x(t))\|^2 + \|S(x(t)) - v(t)\|^2\right) , \tag{2.27}$$

where $c := \min\{M(\delta), \frac{a\delta}{2}, \delta b\}$. It can easily be seen that for every $z \in K'$, $t \mapsto W_\delta(\overline{\Phi}_t(z))$ is Lipschitz continuous, hence absolutely continuous. Its derivative almost everywhere coincides with $\mathcal{L}_{W_\delta}$, which is non-positive. Thus, W_δ is a Lyapunov function for $\overline{\Phi}$. We prove that the Lyapunov function is strict. Consider $z \in K'$ s.t. $W_\delta(\overline{\Phi}_t(z)) = W_\delta(z)$ for all $t > 0$. The derivative almost everywhere of $t \mapsto W_\delta(\overline{\Phi}_t(z))$ is identically zero, and by Eq. (2.27), this implies that $-c\left(\|m_t\|^2 + \|\nabla F(x_t)\|^2 + \|S(x_t) - v_t\|^2\right)$ is equal to zero for every t a.e. (hence, for every t, by continuity of $\overline{\Phi}$). In particular for $t = 0$, $m = \nabla F(x) = 0$ and $S(x) - v = 0$. Hence, $z \in h_\infty^{-1}(\{0\})$. ∎

Corollary 2.22. Let Assumptions 2.2.2, 2.2.3, 2.7.1 and 2.7.2 hold. Assume that $0 < b \le 4a$. For every $z \in \mathcal{Z}_+$, $\lim_{t\to\infty} \mathsf{d}(\Phi(z, t), \mathcal{E}) = 0$.

Proof. Use Prop. 3.16 with $K := \{z\}$. and (Haraux, 1991, Th. 2.1.7). ∎

2.7.3.2 Asymptotic behavior of the solution to (ODE)

Proposition 2.23 (APT)**.** Let Assumptions 2.2.2, 2.2.3, 2.7.1 and 2.7.2 hold true. Assume that $0 < b \leq 4a$. Then, for every $z_0 \in \mathcal{Z}_0$, $Z^0_\infty(z_0)$ is an asymptotic pseudotrajectory of the semiflow Φ given by (2.25).

Proof. Consider $z_0 \in \mathcal{Z}_0$, $T \in (0, +\infty)$ and define $z := Z^0_\infty(z_0)$. Consider $t \geq 1$. For every $s \geq 0$, define $\Delta_t(s) := \|z(t+s) - \Phi(z(t))(s)\|$. The aim is to prove that $\sup_{s\in[0,T]} \Delta_t(s)$ tends to zero as $t \to \infty$. Putting together Prop. 2.12 and Lem. 2.14, the set $K := \overline{\{z(t) : t \geq 1\}}$ is a compact subset of $\mathcal{Z}^*_+$. Define $C(t) := \sup_{s\geq 0} \sup_{z'\in K} \|h(t+s, z') - h_\infty(z')\|$. It can be shown that $\lim_{t\to\infty} C(t) = 0$. We obtain that for every $s \in [0, T]$, $\Delta_t(s) \leq TC(t) + \int_0^s \|h_\infty(z(t+s')) - h_\infty(\Phi(z(t))(s'))\| ds'$. By Lem. 2.14, $K' := \overline{\bigcup_{z'\in\Phi(K)} z'([0,\infty))}$ is a compact subset of $\mathcal{Z}^*_+$. It is immediately seen from the definition that h_∞ is Lipschitz continuous on every compact subset of $\mathcal{Z}^*_+$, hence on $K \cup K'$. Therefore, there exists a constant L, independent from t, s, s.t. $\Delta_t(s) \leq TC(t) + \int_0^s L\Delta_t(s')ds' \qquad (\forall t \geq 1, \forall s \in [0, T])$. Using Grönwall's lemma, it holds that for all $s \in [0, T]$, $\Delta_t(s) \leq TC(t)e^{Ls}$. As a consequence, $\sup_{s\in[0,T]} \Delta_t(s) \leq TC(t)e^{LT}$ and the righthand side converges to zero as $t \to \infty$. ∎

End of the proof of Th. 2.2

By Prop. 2.12, the set $K := \overline{Z^0_\infty(z_0)([0,\infty))}$ is a compact subset of $\mathcal{Z}_+$. Define $K' := \overline{\{\Phi(t, z) : t \geq 0, z \in K\}}$, and let $\overline{\Phi} : [0, +\infty) \times K' \to K'$ be the restriction Φ to K'. By Prop. 3.16, there exists $\delta > 0$ s.t. W_δ is a strict Lyapunov function for the semiflow $\overline{\Phi}$. Moreover, the set of equilibrium points coincides with $\mathcal{E} \cap K'$. In particular, the equilibrium points of $\overline{\Phi}$ form a compact set. By Prop. 2.23, $Z^0_\infty(z_0)$ is an APT of $\overline{\Phi}$. Note that every $z \in \mathcal{E}$ can be written under the form $z = (x, 0, S(x))$ for some $x \in \mathcal{S}$. From the definition of W_δ in (3.33), $W_\delta(z) = W_\delta(x, 0, S(x)) = V_\infty(x, 0, S(x)) = F(x)$. Since $F(\mathcal{S})$ is assumed to have an empty interior, the same holds for $W_\delta(\mathcal{E} \cap K')$. By Prop. 2.20, $\bigcap_{t\geq 0} \overline{Z^0_\infty(z_0)([t,\infty))} \subset \mathcal{E} \cap K'$. The set in the righthand side coincides with the set of limits of convergent sequences of the form $Z^0_\infty(z_0)(t_n)$ for $t_n \to \infty$. As $Z^0_\infty(z_0)([0,\infty))$ is a bounded set, $\mathsf{d}(Z^0_\infty(z_0)(t), \mathcal{E})$ tends to zero.

2.7.4 Proof of Th. 2.3

The proof follows the path of (Haraux and Jendoubi, 2015, Th. 10.1.6, Th. 10.2.3), but requires specific adaptations to deal with the dynamical system at hand. Define for all $\delta > 0$, $t > 0$, and $z = (x, m, v)$,

$$\tilde{W}_\delta(t, (x, m, v)) := V(t, (x, m, v)) - \delta\langle\nabla F(x), m\rangle + \delta\|S(x) - v\|^2 . \tag{2.28}$$

The function $\tilde{W}_\delta$ is the non-autonomous version of the function (3.33). Consider a fixed $x_0 \in \mathbb{R}^d$, and define $w_\delta(t) := \tilde{W}_\delta(t, z(t))$ where $z(t) = (x(t), m(t), v(t))$ is the solution to (ODE) with initial condition $(x_0, 0, 0)$. The proof uses the following steps.

i) Upper-bound on $w_\delta(t)$. From Eq. (2.9), we obtain that for every $t \geq 1$, $V(t, z(t)) \leq |F(x(t))| + \frac{\|m(t)\|^2}{2a\varepsilon(1-e^{-a})}$. Using $\langle\nabla F(x), m\rangle \leq (\|\nabla F(x)\|^2 + \|m\|^2)/2$, we obtain that there exists a constant c_1 (depending on δ) s.t. for every $t \geq 1$,

$$w_\delta(t) \leq c_1 \left(|F(x(t))| + \|m(t)\|^2 + \|\nabla F(x(t))\|^2 + \|S(x(t)) - v(t)\|^2\right) . \tag{2.29}$$

ii) Upper-bound on $\frac{d}{dt}w_\delta(t)$. The function w_δ is absolutely continuous on $[1,+\infty)$.

Moreover, there exist $\delta > 0$, $c_2 > 0$ (both depending on x_0) s.t. for every $t \geq 1$ a.e.,

$$\frac{d}{dt}w_\delta(t) \leq -c_2 \left(\|m(t)\|^2 + \|\nabla F(x(t))\|^2 + \|S(x(t)) - v(t)\|^2 \right) . \tag{2.30}$$

The proof of Eq. (2.30) uses arguments that are similar to the ones used in the proof of Prop. 3.16 (just use Lem. 2.11 to bound the derivative of the first term in Eq. (2.28)). For this reason, it is omitted.

iii) Positivity of $w_\delta(t)$. By Lem. 2.11, the function $t \mapsto V(t, z(t))$ is decreasing. As it is lower bounded, $\ell := \lim_{t\to\infty} V(t, z(t))$ exists. By Th. 2.2, $m(t)$ tends to zero, hence this limit coincides with $\lim_{t\to\infty} F(x(t))$. Replacing F with $F - \ell$, one can assume without loss of generality that $\ell = 0$. By Eq. (2.30), w_δ is non-increasing on $[1,+\infty)$, hence converging to some limit. Using again Th. 2.2, $\langle \nabla F(x(t)), m(t)\rangle \to 0$ and $S(x(t)) - v(t) \to 0$. Thus, $\lim_{t\to\infty} w_\delta(t) = \ell = 0$. Assume that there exists $t_0 \geq 1$ s.t. $w_\delta(t_0) = 0$. Then, w_δ is constant on $[t_0, +\infty)$. By Eq. (2.30), this implies that $m(t) = 0$ on this interval. Hence, $dx(t)/dt = 0$. This means that $x(t) = x(t_0)$ for all $t \geq t_0$. By Th. 2.2, it follows that $x(t_0) \in \mathcal{S}$. In that case, the final result is shown. Therefore, one can assume that $w_\delta(t) > 0$ for all $t \geq 1$.

iv) Putting together (2.29) and (2.30) using the Łojasiewicz condition. By Prop. 2.20 and 2.23, the set $L := \overline{\bigcup_{s\geq 0}\{z(t) : t \geq s\}}$ is a compact connected subset of $\mathcal{E} = \{(x, 0, S(x)) : \nabla F(x) = 0\}$.

The set $\mathcal{U} := \{x : (x, 0, S(x)) \in L\}$ is a compact and connected subset of $\mathcal{S}$. Using Assumption 2.3.1 and (Haraux and Jendoubi, 2015, Lem. 2.1.6), there exist constants $\sigma, c > 0$ and $\theta \in (0, \frac{1}{2}]$, s.t. $\|\nabla F(x)\| \geq c|F(x)|^{1-\theta}$ for all x s.t. $\mathsf{d}(x,\mathcal{U}) < \sigma$. As $\mathsf{d}(x(t),\mathcal{U}) \to 0$, there exists $T \geq 1$ s.t. for all $t \geq T$, $\|\nabla F(x(t))\| \geq c|F(x(t))|^{1-\theta}$. Thus, we may replace the term $\|\nabla F(x(t))\|^2$ in the righthand side of Eq. (2.30) using $\|\nabla F(x(t))\|^2 \geq \frac{1}{2}\|\nabla F(x(t))\|^2 + \frac{1}{2}|F(x(t))|^{2(1-\theta)}$. Upon noting that $2(1-\theta) \geq 1$, we thus obtain that there exists a constant c_3 and some $T' \geq 1$ s.t. for $t \geq T'$ a.e.,

$$\frac{d}{dt}w_\delta(t) \leq -c_3 \left(\|m(t)\|^2 + \|\nabla F(x(t))\|^2 + |F(x(t))| + \|S(x(t)) - v(t)\|^2 \right)^{2(1-\theta)} .$$

Putting this inequality together with Eq. (2.29), we obtain that for some constant $c_4 > 0$ and for all $t \geq T'$ a.e., $\frac{d}{dt}w_\delta(t) \leq -c_4 w_\delta(t)^{2(1-\theta)}$.

v) End of the proof. Following the arguments of (Haraux and Jendoubi, 2015, Th. 10.1.6), by integrating the preceding inequality, over $[T', t]$, we obtain $w_\delta(t) \leq c_5 t^{-\frac{1}{1-2\theta}}$ for $t \geq T'$ in the case where $\theta < \frac{1}{2}$, whereas $w_\delta(t)$ decays exponentially if $\theta = \frac{1}{2}$. From now on, we focus on the case $\theta < \frac{1}{2}$. By definition of (ODE), $\|\dot{x}(t)\|^2 \leq \|m(t)\|^2/((1-e^{-aT'})^2\varepsilon^2)$ for all $t \geq T'$. Since Eq. (2.30) implies $\|m(t)\|^2 \leq -\dot{w}_\delta(t)/c_2$, we deduce that there exists $c, c' > 0$ s.t. for all $t \geq T'$, $\int_t^{2t} \|\dot{x}(s)\|^2 ds \leq c w_\delta(t) \leq c' t^{-\frac{1}{1-2\theta}}$ .Applying (Haraux and Jendoubi, 2015, Lem. 2.1.5), it follows that $\int_t^\infty \|\dot{x}(s)\|^2 ds \leq c t^{-\frac{\theta}{1-2\theta}}$ for some other constant c. Therefore $x^* := \lim_{t\to+\infty} x(t)$ exists by Cauchy's criterion and for all $t \geq T'$, $\|x(t) - x^*\| \leq c t^{-\frac{\theta}{1-2\theta}}$. Finally, since $x(t) \to a$, we remark that, using the same arguments, the global Łojasiewicz exponent θ can be replaced by any Łojasiewicz exponent of f at x^*. When $\theta = \frac{1}{2}$, the proof follows the same line.

2.8 Proofs for Section 2.4

2.8.1 Proof of Th. 2.4

Given an initial point $x_0 \in \mathbb{R}^d$ and a stepsize $\gamma > 0$, we consider the iterates z_n^γ given by (2.5) and $z_0^\gamma := (x_0, 0, 0)$. For every $n \in \mathbb{N}^*$ and every $z \in \mathcal{Z}_+$, we define

$$H_\gamma(n, z, \xi) := \gamma^{-1}(T_{\gamma,\bar\alpha(\gamma),\bar\beta(\gamma)}(n, z, \xi) - z)\,.$$

Thus, $z_n^\gamma = z_{n-1}^\gamma + \gamma H_\gamma(n, z_{n-1}^\gamma, \xi_n)$ for every $n \in \mathbb{N}^*$. For every $n \in \mathbb{N}^*$ and every $z \in \mathcal{Z}$ of the form $z = (x, m, v)$, we define $e_\gamma(n, z) := (x, (1-\bar\alpha(\gamma)^n)^{-1}m, (1-\bar\beta(\gamma)^n)^{-1}v)$, and set $e_\gamma(0, z) := z$.

Lemma 2.24. Let Assumptions 2.2.1, 2.2.4 and 2.4.2 hold true. There exists $\bar\gamma_0 > 0$ s.t. for every $R > 0$, there exists $s > 0$,

$$\sup\left\{\mathbb{E}\left(\left\|H_\gamma(n+1, z, \xi)\right\|^{1+s}\right) : \gamma \in (0, \bar\gamma_0], n \in \mathbb{N}, z \in \mathcal{Z}_+ \text{ s.t. } \|e_\gamma(n, z)\| \le R\right\} < \infty\,. \tag{2.31}$$

Proof. Let $R > 0$. We denote by $(H_{\gamma,\mathsf{x}}, H_{\gamma,\mathsf{m}}, H_{\gamma,\mathsf{v}})$ the block components of H_γ. There exists a constant C_s depending only on s s.t. $\|H_\gamma\|^{1+s} \le C_s(\|H_{\gamma,\mathsf{x}}\|^{1+s} + \|H_{\gamma,\mathsf{m}}\|^{1+s} + \|H_{\gamma,\mathsf{v}}\|^{1+s})$. Hence, it is sufficient to prove that Eq. (2.31) holds respectively when replacing H_γ with each of $H_{\gamma,\mathsf{x}}, H_{\gamma,\mathsf{m}}, H_{\gamma,\mathsf{v}}$. Consider $z = (x, m, v)$ in $\mathcal{Z}_+$. We write: $\|H_{\gamma,\mathsf{x}}(n+1, z, \xi)\| \le \varepsilon^{-1}(\|\frac{m}{1-\bar\alpha(\gamma)^n}\| + \|\nabla f(x, \xi)\|)$. Thus, for every z s.t. $\|e_\gamma(n, z)\| \le R$, there exists a constant C depending only on ε, R and s s.t. $\|H_{\gamma,\mathsf{x}}(n+1, z, \xi)\|^{1+s} \le C(1 + \|\nabla f(x, \xi)\|^{1+s})$. By Assumption 2.4.2, (2.31) holds for $H_{\gamma,\mathsf{x}}$ instead of H_γ. Similar arguments hold for $H_{\gamma,\mathsf{m}}$ and $H_{\gamma,\mathsf{v}}$ upon noting that the functions $\gamma \mapsto (1-\bar\alpha(\gamma))/\gamma$ and $\gamma \mapsto (1-\bar\beta(\gamma))/\gamma$ are bounded under Assumption 2.2.4. ■

For every $R > 0$, and every arbitrary sequence $z = (z_n : n \in \mathbb{N})$ on $\mathcal{Z}_+$, we define $\tau_R(z) := \inf\{n \in \mathbb{N} : \|e_\gamma(n, z_n)\| > R\}$ with the convention that $\tau_R(z) = +\infty$ when the set is empty. We define the map $B_R : \mathcal{Z}_+^{\mathbb{N}} \to \mathcal{Z}_+^{\mathbb{N}}$ given for any arbitrary sequence $z = (z_n : n \in \mathbb{N})$ on $\mathcal{Z}_+$ by $B_R(z)(n) = z_n \mathbb{1}_{n<\tau_R(z)} + z_{\tau_R(z)}\mathbb{1}_{n\ge\tau_R(z)}$. We define the random sequence $z^{\gamma,R} := B_R(z^\gamma)$. Recall that a family $(X_i : i \in I)$ of random variables on some Euclidean space is called *uniformly integrable* if $\lim_{A\to+\infty}\sup_{i\in I}\mathbb{E}(\|X_i\|\mathbb{1}_{\|X_i\|>A}) = 0$.

Lemma 2.25. Let Assumptions 2.2.1, 2.2.4, 2.4.2 and 2.4.1 hold true. There exists $\bar\gamma_0 > 0$ s.t. for every $R > 0$, the family of r.v. $(\gamma^{-1}(z_{n+1}^{\gamma,R} - z_n^{\gamma,R}) : n \in \mathbb{N}, \gamma \in (0, \bar\gamma_0])$ is uniformly integrable.

Proof. Let $R > 0$. As the event $\{n < \tau_R(z^\gamma)\}$ coincides with $\bigcap_{k=0}^n\{\|e_\gamma(k, z_k^\gamma)\| \le R\}$, it holds that for every $n \in \mathbb{N}$,

$$\frac{z_{n+1}^{\gamma,R} - z_n^{\gamma,R}}{\gamma} = \frac{z_{n+1}^\gamma - z_n^\gamma}{\gamma}\mathbb{1}_{n<\tau_R(z^\gamma)} = H_\gamma(n+1, z_n^\gamma, \xi_{n+1})\prod_{k=0}^n \mathbb{1}_{\|e_\gamma(k,z_k^\gamma)\|\le R}\,.$$

Choose $\bar\gamma_0 > 0$ and $s > 0$ as in Lem. 2.24. For every $\gamma \le \bar\gamma_0$,

$$\mathbb{E}\left(\left\|\gamma^{-1}(z_{n+1}^{\gamma,R} - z_n^{\gamma,R})\right\|^{1+s}\right) \le \sup\left\{\mathbb{E}\left(\left\|H_{\gamma'}(\ell+1, z, \xi)\right\|^{1+s}\right) : \gamma' \in (0, \bar\gamma_0], \ell \in \mathbb{N}, z \in \mathcal{Z}_+, \|e_\gamma(\ell, z)\| \le R\right\}.$$

By Lem. 2.24, the righthand side is finite and does not depend on (n, γ). ■

For a fixed $\gamma > 0$, we define the interpolation map $\mathsf{X}_\gamma : \mathcal{Z}^{\mathbb{N}} \to C([0, +\infty), \mathcal{Z})$ as follows for every sequence $z = (z_n : n \in \mathbb{N})$ on $\mathcal{Z}$:

$$\mathsf{X}_\gamma(z) : t \mapsto z_{\lfloor \frac{t}{\gamma} \rfloor} + (t/\gamma - \lfloor t/\gamma \rfloor)(z_{\lfloor \frac{t}{\gamma} \rfloor + 1} - z_{\lfloor \frac{t}{\gamma} \rfloor}) \, .$$

For every $\gamma, R > 0$, we define $\mathsf{z}^{\gamma,R} := \mathsf{X}_\gamma(z^{\gamma,R}) = \mathsf{X}_\gamma \circ B_R(z^\gamma)$. Namely, $\mathsf{z}^{\gamma,R}$ is the interpolated process associated with the sequence $(z_n^{\gamma,R})$. It is a random variable on $C([0, +\infty), \mathcal{Z})$. We recall that $\mathcal{F}_n$ is the σ-algebra generated by the r.v. $(\xi_k : 1 \le k \le n)$. For every γ, n, R, we use the notation: $\Delta_0^{\gamma,R} := 0$ and

$$\Delta_{n+1}^{\gamma,R} := \gamma^{-1}(z_{n+1}^{\gamma,R} - z_n^{\gamma,R}) - \mathbb{E}(\gamma^{-1}(z_{n+1}^{\gamma,R} - z_n^{\gamma,R}) | \mathcal{F}_n) \, .$$

Lemma 2.26. Let Assumptions 2.2.1, 2.2.4, 2.4.2 and 2.4.1 hold true. There exists $\bar{\gamma}_0 > 0$ s.t. for every $R > 0$, the family of r.v. $(\mathsf{z}^{\gamma,R} : \gamma \in (0, \bar{\gamma}_0])$ is tight. Moreover, for every $\delta > 0$, $\mathbb{P}\left(\max_{0 \le n \le \lfloor \frac{T}{\gamma} \rfloor} \gamma \left\| \sum_{k=0}^n \Delta_{k+1}^{\gamma,R} \right\| > \delta \right) \xrightarrow{\gamma \to 0} 0 \, .$

Proof. It is an immediate consequence of Lem. 2.25 and (Bianchi et al., 2019, Lem. 6.2). ■

The proof of the following lemma is omitted.

Lemma 2.27. Let Assumptions 2.2.1 and 2.2.4 hold true. Consider $t > 0$ and $z \in \mathcal{Z}_+$. Let (φ_n, z_n) be a sequence on $\mathbb{N}^* \times \mathcal{Z}_+$ s.t. $\lim_{n \to \infty} \gamma_n \varphi_n = t$ and $\lim_{n \to \infty} z_n = z$. Then, $\lim_{n \to \infty} h_{\gamma_n}(\varphi_n, z_n) = h(t, z)$ and $\lim_{n \to \infty} e_{\gamma_n}(\varphi_n, z_n) = \bar{e}(t, z)$.

End of the Proof of Th. 2.4 Consider $x_0 \in \mathbb{R}^d$ and set $z_0 = (x_0, 0, 0)$. Define $R_0 := \sup \left\{ \|\bar{e}(t, Z_\infty^0(z_0)(t))\| : t > 0 \right\}$. By Prop. 2.12, $R_0 < +\infty$. We select an arbitrary R s.t. $R \ge R_0 + 1$. For every $n \ge 0$, $z \in \mathcal{Z}_+$,

$$z_{n+1}^{\gamma,R} = z_n^{\gamma,R} + \gamma H_\gamma(n+1, z_n^{\gamma,R}, \xi_{n+1}) \mathbb{1}_{\|e_\gamma(n, z_n^{\gamma,R})\| \le R} \, .$$

Define for every $n \ge 1$, $z \in \mathcal{Z}_+$, $h_{\gamma,R}(n, z) := h_\gamma(n, z) \mathbb{1}_{\|e_\gamma(n-1,z)\| \le R}$. Then, $\Delta_{n+1}^{\gamma,R} = \gamma^{-1}(z_{n+1}^{\gamma,R} - z_n^{\gamma,R}) - h_{\gamma,R}(n+1, z_n^{\gamma,R})$. Define also for every $n \ge 0$, $M_n^{\gamma,R} := \sum_{k=1}^n \Delta_k^{\gamma,R} = \gamma^{-1}(z_n^{\gamma,R} - z_0) - \sum_{k=0}^{n-1} h_{\gamma,R}(k+1, z_k^{\gamma,R})$. Consider $t \ge 0$ and set $n := \lfloor t/\gamma \rfloor$. For any $T > 0$, it holds that :

$$\sup_{t \in [0,T]} \left\| \mathsf{z}^{\gamma,R}(t) - z_0 - \int_0^t h_{\gamma,R}(\lfloor s/\gamma \rfloor + 1, \mathsf{z}^{\gamma,R}(\gamma \lfloor s/\gamma \rfloor)) ds \right\| \le \max_{0 \le n \le \lfloor T/\gamma \rfloor + 1} \gamma \|M_n^{\gamma,R}\| \, .$$

By Lem. 2.26,

$$\mathbb{P}\left(\sup_{t \in [0,T]} \left\| \mathsf{z}^{\gamma,R}(t) - z_0 - \int_0^t h_{\gamma,R}\left(\lfloor s/\gamma \rfloor + 1, \mathsf{z}^{\gamma,R}(\gamma \lfloor s/\gamma \rfloor) \right) ds \right\| > \delta \right) \xrightarrow{\gamma \to 0} 0 \, . \quad (2.32)$$

As a second consequence of Lem. 2.26, the family of r.v. $(\mathsf{z}^{\gamma,R} : 0 < \gamma \le \bar{\gamma}_0)$ is tight, where $\bar{\gamma}_0$ is chosen as in Lem. 2.26 (it does not depend on R). By Prokhorov's theorem, there exists a sequence $(\gamma_k : k \in \mathbb{N})$ s.t. $\gamma_k \to 0$ and s.t. $(\mathsf{z}^{\gamma_k,R} : k \in \mathbb{N})$ converges

in distribution to some probability measure ν on $C([0,+\infty),\mathcal{Z}_+)$. By Skorohod's representation theorem, there exists a r.v. z on some probability space $(\Omega',\mathcal{F}',\mathbb{P}')$, with distribution ν, and a sequence of r.v. $(\mathsf{z}_{(k)} : k\in\mathbb{N})$ on that same probability space where for each $k\in\mathbb{N}$, the r.v. $\mathsf{z}_{(k)}$ has the same distribution as the r.v. $\mathsf{z}^{\gamma_k,R}$, and s.t. for every $\omega\in\Omega'$, $\mathsf{z}_{(k)}(\omega)$ converges to $\mathsf{z}(\omega)$ uniformly on compact sets. Now select a fixed $T>0$. According to Eq. (2.32), the sequence

$$\sup_{t\in[0,T]}\left\|\mathsf{z}_{(k)}(t)-z_0-\int_0^t h_{\gamma_k,R}\left(\lfloor s/\gamma_k\rfloor+1,\mathsf{z}_{(k)}(\gamma_k\lfloor s/\gamma_k\rfloor)\right)ds\right\|,$$

indexed by $k\in\mathbb{N}$, converges in probability to zero as $k\to\infty$. One can therefore extract a further subsequence $\mathsf{z}_{(\varphi_k)}$, s.t. the above sequence converges to zero almost surely. In particular, since $\mathsf{z}_{(k)}(t)\to\mathsf{z}(t)$ for every t, we obtain that

$$\mathsf{z}(t)=z_0+\lim_{k\to\infty}\int_0^t h_{\gamma_{\varphi_k},R}\left(\lfloor s/\gamma_{\varphi_k}\rfloor+1,\mathsf{z}_{(\varphi_k)}(\gamma_{\varphi_k}\lfloor s/\gamma_{\varphi_k}\rfloor)\right)ds\quad(\forall t\in[0,T])\,. \tag{2.33}$$

Consider $\omega\in\Omega'$ s.t. the r.v. z satisfies (2.33) at point ω. From now on, we consider that ω is fixed, and we handle z as an element of $C([0,+\infty),\mathcal{Z}_+)$, and no longer as a random variable. Define $\tau:=\inf\{t\in[0,T] : \|\bar{e}(t,\mathsf{z}(t))\|>R_0+\frac{1}{2}\}$ if the latter set is non-empty, and $\tau:=T$ otherwise. Since $\mathsf{z}(0)=z_0$ and $\|z_0\|<R_0$, it holds that $\tau>0$ using the continuity of z. Choose any (s,t) s.t. $0<s<t<\tau$. Note that $\mathsf{z}_{(k)}(\gamma_k\lfloor s/\gamma_k\rfloor)\to\mathsf{z}(s)$ and $\gamma_k(\lfloor s/\gamma_k\rfloor+1)\to s$. Thus, by Lem. 2.27, $h_{\gamma_k}\left(\lfloor s/\gamma_k\rfloor+1,\mathsf{z}_{(k)}(\gamma_k\lfloor s/\gamma_k\rfloor)\right)$ converges to $h(s,\mathsf{z}(s))$ and $e_{\gamma_k}(\lfloor s/\gamma_k\rfloor,\mathsf{z}_{(k)}(\gamma_k\lfloor s/\gamma_k\rfloor))$ converges to $\bar{e}(s,\mathsf{z}(z))$. Since $s<\tau$, $\bar{e}(s,\mathsf{z}(z))\le R_0+\frac{1}{2}$. As $R\ge R_0+1$, there exists a certain $K(s)$ s.t. for every $k\ge K(s)$, $\mathbb{1}_{\|e_{\gamma_k}(\lfloor s/\gamma_k\rfloor,\mathsf{z}_{(k)}(\gamma_k\lfloor s/\gamma_k\rfloor))\|\le R}=1$. As a consequence, $h_{\gamma_k,R}(\lfloor s/\gamma_k\rfloor+1,\mathsf{z}_{(k)}(\gamma_k\lfloor s/\gamma_k\rfloor))$ converges to $h(s,\mathsf{z}(s))$ as $k\to\infty$. Using Lebesgue's dominated convergence theorem, we obtain, for all $t\in[0,\tau]$: $\mathsf{z}(t)=z_0+\int_0^t h\left(s,\mathsf{z}(s)\right)ds$. Therefore $\mathsf{z}(t)=Z^0_\infty(x_0)(t)$ for every $t\in[0,\tau]$. In particular, $\|\mathsf{z}(\tau)\|\le R_0$ and this means that $\tau=T$. Thus, $\mathsf{z}(t)=Z^0_\infty(x_0)(t)$ for every $t\in[0,T]$ (and consequently for every $t\ge 0$). We have shown that for every $R\ge R_0+1$, the sequence of r.v. $(\mathsf{z}^{\gamma,R}:\gamma\in(0,\bar{\gamma}_0])$ is tight and converges in probability to $Z^0_\infty(z_0)$ as $\gamma\to 0$. Therefore, for every $T>0$,

$$\forall\delta>0,\ \lim_{\gamma\to 0}\mathbb{P}\left(\sup_{t\in[0,T]}\left\|\mathsf{z}^{\gamma,R}(t)-Z^0_\infty(x_0)(t)\right\|>\delta\right)=0\,. \tag{2.34}$$

In order to complete the proof, we show that $\mathbb{P}\left(\sup_{t\in[0,T]}\left\|\mathsf{z}^{\gamma,R}(t)-\mathsf{z}^{\gamma}(t)\right\|>\delta\right)\to 0$ as $\gamma\to 0$, for all $\delta>0$. where we recall that $\mathsf{z}^\gamma=\mathsf{X}_\gamma(z^\gamma)$. Note that $\left\|\mathsf{z}^{\gamma,R}(t)\right\|\le\left\|\mathsf{z}^{\gamma,R}(t)-Z^0_\infty(z_0)(t)\right\|+R_0$ by the triangular inequality. Therefore, for every $T,\delta>0$,

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in[0,T]}\left\|\mathsf{z}^{\gamma,R}(t)-\mathsf{z}^{\gamma}(t)\right\|>\delta\right)&\le\mathbb{P}\left(\sup_{t\in[0,T]}\left\|\mathsf{z}^{\gamma,R}(t)\right\|\ge R\right)\\&\le\mathbb{P}\left(\sup_{t\in[0,T]}\left\|\mathsf{z}^{\gamma,R}(t)-Z^0_\infty(z_0)(t)\right\|\ge R-R_0\right).\end{aligned}$$

By Eq. (2.34), the RHS of the above inequality tends to zero as $\gamma\to 0$. The proof is complete.

2.8.2 Proof of Th. 2.5

We start by stating a general result. Consider a Euclidean space X equipped with its Borel σ-field $\mathcal{X}$. Let $\bar{\gamma}_0 > 0$, and consider two families $(P_{\gamma,n} : 0 < \gamma < \bar{\gamma}_0, n \in \mathbb{N}^*)$ and $(\bar{P}_\gamma : 0 < \gamma < \bar{\gamma}_0)$ of Markov transition kernels on X. Denote by $\mathcal{P}(\mathsf{X})$ the set of probability measures on X. Let $X = (X_n : n \in \mathbb{N})$ be the canonical process on X. Let $(\mathbb{P}^{\gamma,\nu} : 0 < \gamma < \bar{\gamma}_0, \nu \in \mathcal{P}(\mathsf{X}))$ and $(\bar{\mathbb{P}}^{\gamma,\nu} : 0 < \gamma < \bar{\gamma}_0, \nu \in \mathcal{P}(\mathsf{X}))$ be two families of measures on the canonical space $(X^{\mathbb{N}}, \mathcal{X}^{\otimes\mathbb{N}})$ such that the following holds:

- Under $\mathbb{P}^{\gamma,\nu}$, X is a non-homogeneous Markov chain with transition kernels $(P_{\gamma,n} : n \in \mathbb{N}^*)$ and initial distribution ν, that is, for each $n \in \mathbb{N}^*$, $\mathbb{P}^{\gamma,\nu}(X_n \in dx | X_{n-1}) = P_{\gamma,n}(X_{n-1}, dx)$.
- Under $\bar{\mathbb{P}}^{\gamma,\nu}$, X is an homogeneous Markov chain with transition kernel $\bar{P}_\gamma$ and initial distribution ν.

In the sequel, we will use the notation $\bar{P}^{\gamma,x}$ as a shorthand notation for $\bar{P}^{\gamma,\delta_x}$ where δ_x is the Dirac measure at some point $x \in \mathsf{X}$. Finally, let Ψ be a semiflow on X. A Markov kernel P is *Feller* if Pf is continuous for every bounded continuous f.

Assumption 2.8.1. Let $\nu \in \mathcal{P}(\mathsf{X})$.

i) For every γ, $\bar{P}_\gamma$ is Feller.
ii) $(\mathbb{P}^{\gamma,\nu} X_n^{-1} : n \in \mathbb{N}, 0 < \gamma < \bar{\gamma}_0)$ is a tight family of measures.
iii) For every $\gamma \in (0, \bar{\gamma}_0)$ and every bounded Lipschitz continuous function $f : \mathsf{X} \to \mathbb{R}$, $P_{\gamma,n} f$ converges to $\bar{P}_\gamma f$ as $n \to \infty$, uniformly on compact sets.
iv) For every $\delta > 0$, for every compact set $K \subset \mathsf{X}$, for every $t > 0$, $\lim_{\gamma \to 0} \sup_{x \in K} \bar{P}^{\gamma,x} \left(\|X_{\lfloor t/\gamma \rfloor} - \Psi_t(x)\| > \delta \right) = 0$.

Let BC_Ψ be the Birkhof center of Ψ *i.e.*, the closure of the set of recurrent points.

Theorem 2.28. *Consider $\nu \in \mathcal{P}(\mathsf{X})$ s.t. Assumption 2.8.1 holds true. Then, for every $\delta > 0$, $\lim_{\gamma \to 0} \limsup_{n \to \infty} \frac{1}{n+1} \sum_{k=0}^{n} \mathbb{P}^{\gamma,\nu} \left(d(X_k, BC_\Psi) > \delta \right) = 0$.*

Proof. For every γ, n, define $\mu_{\gamma,n} := \nu P_{\gamma,1} \cdots P_{\gamma,n}$ with the convention that $\mu_{\gamma,0} = \nu$. Otherwise stated, $\mu_{\gamma,n} = \mathbb{P}_{\gamma,\nu} X_n^{-1}$. Define $\Pi_{\gamma,n} := \frac{1}{n+1} \sum_{k=0}^{n} \mu_{\gamma,k}$ for every $n \in \mathbb{N}$. Assumption 2.8.1 implies that for any fixed γ, $(\Pi_{\gamma,n} : n \in \mathbb{N})$ is tight. By Prokhorov's theorem, it admits a cluster point π_γ. For such a cluster point, consider a subsequence φ_n s.t. $\Pi_{\gamma,\varphi_n} \Rightarrow \pi_\gamma$, where $\Rightarrow$ stands for the weak convergence of probability measures. Consider a bounded Lipschitz continuous function $f : \mathsf{X} \to \mathbb{R}$. It holds that $\Pi_{\gamma,n}(f)$ and $\Pi_{\gamma,n}(\bar{P}_\gamma f)$ respectively converge to $\pi_\gamma(f)$ and $\pi_\gamma(\bar{P}_\gamma f)$ along the subsequence, because $\bar{P}_\gamma$ is Feller. We observe that

$$\left| \Pi_{\gamma,n} \bar{P}_\gamma f - \Pi_{\gamma,n} f \right| \leq \frac{1}{n+1} \sum_{k=0}^{n} |\mu_{\gamma,k}(\bar{P}_\gamma f - P_{\gamma,k+1} f)| + \frac{2\|f\|_\infty}{n+1} .$$

Choose $\delta > 0$ and a compact set $K \subset \mathsf{X}$ s.t. $\sup_k \mu_{\gamma,k}(K^c) < \delta$. For every k, $|\mu_{\gamma,k}(\bar{P}_\gamma f - P_{\gamma,k+1} f)| \leq \sup_{x \in K} |\bar{P}_\gamma f(x) - P_{\gamma,k+1} f(x)| + 2\|f\|_\infty \delta$. By Assumption 2.8.1*iii)*, it holds that $\limsup_n \left| \Pi_{\gamma,n} \bar{P}_\gamma f - \Pi_{\gamma,n} f \right| \leq 2\|f\|_\infty \delta$. As δ is arbitrary, $\Pi_{\gamma,n} \bar{P}_\gamma f - \Pi_{\gamma,n} f \to 0$, which shows that $\pi_\gamma \bar{P}_\gamma f - \pi_\gamma f = 0$. We have shown that every cluster point of $(\Pi_{\gamma,n} : n \in \mathbb{N})$ is an invariant measure of $\bar{P}_\gamma$.

Consider an arbitrary sequence $\gamma_j \downarrow 0$ as $j \to \infty$, and let π_j be an invariant measure of $\bar{P}_{\gamma_j}$ for every j. It is not difficult to show that the sequence (π_j) is also tight, hence converging to some π^* as $j \to \infty$, along some subsequence. We now prove that such a cluster point π^* is an invariant measure for the semiflow Ψ *i.e.*, $\pi^* \Psi_t^{-1} = \pi^*$ for every $t > 0$. Such a proof can be found for instance in Fort and Pagès (1999), we reproduce it here for completeness. Denote by $\bar{\mathbb{E}}^{\gamma,\nu}$ the expectation associated with $\bar{\mathbb{P}}^{\gamma,\nu}$ and by L the Lipschitz constant of f. For an arbitrary $\delta > 0$, consider a compact set K s.t. $\sup_j \pi_j(K^c) < \delta$. For every j and every $t > 0$, using that $\pi_j = \pi_j \bar{P}_{\gamma_j}^{\lfloor \frac{t}{\gamma_j} \rfloor}$, we obtain, by following the same approach as Fort and Pagès (1999),

$$
\begin{aligned}
\left| \int f \circ \Psi_t d\pi_j - \int f d\pi_j \right| &= \left| \mathbb{E}^{\gamma_j, \pi_j} (f(\Psi_t(X_0)) - f(X_{\lfloor \frac{t}{\gamma_j} \rfloor})) \right| \\
&\leq \mathbb{E}^{\gamma_j, \pi_j} \left(|f(\Psi_t(X_0)) - f(X_{\lfloor \frac{t}{\gamma_j} \rfloor})| \mathbb{1}_K(X_0) \right) + 2\|f\|_\infty \delta \\
&\leq \mathbb{E}^{\gamma_j, \pi_j} \left(\left(2\|f\|_\infty \wedge L \|\Psi_t(X_0) - X_{\lfloor \frac{t}{\gamma_j} \rfloor}\| \right) \mathbb{1}_K(X_0) \right) + 2\|f\|_\infty \delta \\
&\leq \mathbb{E}^{\gamma_j, \pi_j} \left(2\|f\|_\infty \mathbb{1}_K(X_0) \mathbb{1}_{\|\Psi_t(X_0) - X_{\lfloor \frac{t}{\gamma_j} \rfloor}\| > \delta} \right) + L\delta + 2\|f\|_\infty \delta \\
&\leq 2\|f\|_\infty \sup_{x \in K} \mathbb{P}^{\gamma_j, x} \left(\|\Psi_t(x) - X_{\lfloor \frac{t}{\gamma_j} \rfloor}\| > \delta \right) + L\delta + 2\|f\|_\infty \delta \,.
\end{aligned}
$$

Thus, $\limsup_j \left| \int f \circ \Psi_t d\pi_j - \int f d\pi_j \right| \leq (L + 2\|f\|_\infty)\delta$, and since δ is arbitrary, the lim sup is equal to zero. Considering the limit along the converging subsequence, it follows that $\int f \circ \Psi_t d\pi^* - \int f d\pi^* = 0$. Hence, π^* is invariant for Ψ. By Poincaré's recurrence theorem, $\pi^*(BC_\Psi) = 1$.

We now conclude the proof of Th. 2.28. For every $\delta > 0$, set $A_\delta := \{x : d(x, BC_\Psi) \geq \delta\}$ By contradiction, assume that there exists $\delta > 0$, a sequence $\gamma_j \downarrow 0$, and, for every j, a sequence $(\varphi_n^j : n \in \mathbb{N})$ s.t. for every n, $\Pi_{\gamma_j, \varphi_n^j}(A_\delta) > \delta$. For every j, as $(\Pi_{\gamma_j, \varphi_n^j} : n \in \mathbb{N})$ is tight, one can extract a subsequence $(\Pi_{\gamma_j, \bar{\varphi}_n^j} : n \in \mathbb{N})$ converging weakly to some measure π_j which is invariant for $\bar{P}_{\gamma_j}$. By the portmanteau theorem, $\pi_j(A_\delta) > \delta$. As (π_j) is tight, it converges weakly along some subsequence to some π^* satisfying $\pi^*(BC_\Psi) = 1$. As $\pi^*(A_\delta) > \delta$, this leads to a contradiction. ∎

End of the Proof of Th. 2.5. We apply Th. 2.28 in the case where $P_{\gamma,n}$ is the kernel of the non-homogeneous Markov chain (z_n^γ) defined by (2.5) and $\bar{P}_\gamma$ is the kernel of the homogeneous Markov chain $(\bar{z}_n^\gamma)$ given by $\bar{z}_n^\gamma = \bar{z}_{n-1}^\gamma + \gamma H_\gamma(\infty, \bar{z}_{n-1}^\gamma, \xi_n)$ for every $n \in \mathbb{N}^*$ and $\bar{z}_0 \in \mathcal{Z}_+$ where $H_\gamma(\infty, \bar{z}_{n-1}^\gamma, \xi_n) := \lim_{k \to \infty} H_\gamma(k, \bar{z}_{n-1}^\gamma, \xi_n)$. The task is merely to verify Assumption 2.8.1*iii)*, the other assumptions being easily verifiable using Th. 2.4, Consider $\gamma \in (0, \bar{\gamma}_0)$. Let $f : \mathcal{Z} \to \mathbb{R}$ be a bounded M-Lipschitz continuous

function and K a compact. For all $z = (x, m, v) \in K$:

$$|P_{\gamma,n}(f)(z) - \bar{P}_\gamma(f)(z)| \leq M\gamma\mathbb{E}\left\| \frac{(1-\alpha^n)^{-1}\tilde{m}_\xi}{\varepsilon + (1-\beta^n)^{-\frac{1}{2}}\tilde{v}_\xi^{1/2}} - \frac{\tilde{m}_\xi}{\varepsilon + \tilde{v}_\xi^{1/2}} \right\|$$

$$\leq \frac{M\gamma\alpha^n}{\varepsilon(1-\alpha^n)} \sup_{x,m} \left(\alpha||m|| + (1-\alpha)\mathbb{E}||\nabla f(x,\xi)|| \right) + \frac{M\gamma\mathbb{E}||\tilde{m}_\xi \odot \tilde{v}_\xi^{1/2}||}{\varepsilon^2} \left(1 - \frac{1}{(1-\beta^n)^{1/2}} \right)$$

where we write $\alpha = \bar{\alpha}(\gamma)$, $\beta = \bar{\beta}(\gamma)$, $\tilde{m}_\xi := \alpha m + (1-\alpha)\nabla f(x,\xi)$ and $\tilde{v}_\xi := \beta v + (1-\beta)\nabla f(x,\xi)^{\odot 2}$. Thus, condition 2.8.1*iii)* follows. Finally, Cor. 2.22 implies $BC_\Phi = \mathcal{E}$.

2.9 Proofs for Section 2.5

In this section, we denote by $\mathbb{E}_n = \mathbb{E}(\cdot|\mathcal{F}_n)$ the conditional expectation w.r.t. $\mathcal{F}_n$. We also use the notation $\nabla f_{n+1} := \nabla f(x_n, \xi_{n+1})$.

The following lemma will be useful in the proofs.

Lemma 2.29. Let the sequence (r_n) be defined as in Algorithm 2.2. Assume that $0 \leq \alpha_n \leq 1$ for all n and that $(1-\alpha_n)/\gamma_n \to a > 0$ as $n \to +\infty$. Then,

i) $\forall n \in \mathbb{N}, r_n = 1 - \prod_{i=1}^n \alpha_i$,
ii) The sequence (r_n) is nondecreasing and converges to 1.
iii) Under Assumption 2.5.4 *i)*, for every $\epsilon > 0$, for sufficiently large n, we have $r_n - 1 \leq e^{-\frac{a\gamma_0}{2(1-\kappa)}n^{1-\kappa}}$ if $\kappa \in (0,1)$ and $r_n - 1 \leq n^{-a\gamma_0/(1+\epsilon)}$ if $\kappa = 1$.

A similar lemma holds for the sequence $(\bar{r}_n)$.

Proof. i) stems from observing that $r_{n+1} - 1 = \alpha_{n+1}(r_n - 1)$ for every $n \in \mathbb{N}$ and iterating this relation ($r_0 = 0$). As a consequence, the sequence (r_n) is nondecreasing. We can write : $0 \leq 1 - r_n \leq \exp(-\sum_{i=1}^n (1-\alpha_i))$. iii) As $\sum_{n\geq 1} \gamma_n = +\infty$ and $(1-\alpha_n) \sim a\gamma_n$, we deduce that $\sum_{i=1}^n (1-\alpha_i) \sim \sum_{i=1}^n a\gamma_i$. The results follow from the fact that $\sum_{i=1}^n \gamma_i \sim \frac{\gamma_0}{1-\kappa} n^{1-\kappa}$ when $\kappa \in (0,1)$ and $\sum_{i=1}^n \gamma_i \sim \gamma_0 \ln n$ for $\kappa = 1$. ∎

2.9.1 Proof of Th. 2.6

We define $\bar{z}_n = (x_{n-1}, m_n, v_n)$ (note the shift in the index of the variable x). We have

$$\bar{z}_{n+1} = \bar{z}_n + \gamma_{n+1} h_\infty(\bar{z}_n) + \gamma_{n+1}\chi_{n+1} + \gamma_{n+1}\varsigma_{n+1}\,,$$

where h_∞ is defined in Eq. (2.18) and where we set

$$\chi_{n+1} = \left(0, \gamma_{n+1}^{-1}(1-\alpha_{n+1})(\nabla f_{n+1} - \nabla F(x_n)), \gamma_{n+1}^{-1}(1-\beta_{n+1})(\nabla f_{n+1}^{\odot 2} - S(x_n))\right)$$

and $\varsigma_{n+1} = (\varsigma_{n+1}^x, \varsigma_{n+1}^m, \varsigma_{n+1}^v)$ with the components defined by: $\varsigma_{n+1}^x = \frac{m_n}{\varepsilon + \sqrt{v_n}} - \frac{\gamma_n}{\gamma_{n+1}} \frac{\hat{m}_n}{\varepsilon + \sqrt{\hat{v}_n}}$, $\varsigma_{n+1}^m = \left(\frac{1-\alpha_{n+1}}{\gamma_{n+1}} - a\right)(\nabla F(x_n) - m_n) + a(\nabla F(x_n) - \nabla F(x_{n-1}))$ and $\varsigma_{n+1}^v = \left(\frac{1-\beta_{n+1}}{\gamma_{n+1}} - b\right)(S(x_n) - v_n) + b(S(x_n) - S(x_{n-1}))$. Therefore, our algorithm has two perturbations: the first one is a martingale increment χ_{n+1}, while the second one is a

negligible perturbation ς_{n+1} converging a.s. to zero. We first prove that $\varsigma_n \to 0$ a.s. Using the triangular inequality,

$$\|\varsigma_n^x\| \leq \left\| \frac{m_n}{\varepsilon + \sqrt{v_n}} - \frac{m_n}{\bar{r}_n^{1/2}\varepsilon + \sqrt{v_n}} \right\| + \left| 1 - \frac{\gamma_n r_n^{-1}}{\gamma_{n+1}\bar{r}_n^{-1/2}} \right| \left\| \frac{m_n}{\bar{r}_n^{1/2}\varepsilon + \sqrt{v_n}} \right\|$$
$$\leq \varepsilon^{-1}|1 - \bar{r}_n^{-1/2}|\|m_n\| + \varepsilon^{-1} \left| \bar{r}_n^{-1/2} - \frac{\gamma_n r_n^{-1}}{\gamma_{n+1}} \right| \|m_n\|,$$

which converges a.s. to zero because of the boundedness of (z_n) combined with Assumption 2.5.1 and Lem. 2.29 for $(\bar{r}_n)$. The components ς_{n+1}^m and ς_{n+1}^v converge a.s. to zero, as products of a bounded term and a term converging to zero. Indeed, note that ∇F and S are locally Lipschitz continuous under Assumption 2.2.1. Hence, there exists a constant C s.t. $\|\nabla F(x_n) - \nabla F(x_{n-1})\| \leq C\|x_n - x_{n-1}\| \leq \frac{C}{\varepsilon}\gamma_n\|m_n\|$. The same inequality holds when replacing ∇F by S. Now consider the martingale increment sequence (χ_n), adapted to $\mathcal{F}_n$. Estimating the second order moments, it is easy to show using Assumption 2.4.2 *i)* that there exists a constant C' s.t. $\mathbb{E}_n(\|\chi_{n+1}\|^2) \leq C'$. Using that $\sum_k \gamma_k^2 < \infty$, it follows that $\sum_n \mathbb{E}_n(\|\gamma_{n+1}\chi_{n+1}\|^2) < \infty$ a.s. By Doob's convergence theorem, $\lim_{n\to\infty} \sum_{k\leq n} \gamma_k\chi_k$ exists almost surely. Using this result along with the fact that ς_n converges a.s. to zero, it follows from usual stochastic approximation arguments (Benaïm, 1999, Remark. 4.5) that the interpolated process $\bar{\mathsf{z}} : [0, +\infty) \to \mathcal{Z}_+$ given by

$$\bar{\mathsf{z}}(t) = \bar{z}_n + (t - \tau_n)\frac{\bar{z}_{n+1} - \bar{z}_n}{\gamma_{n+1}} \qquad \Big(\forall n \in \mathbb{N}\,,\, \forall t \in [\tau_n, \tau_{n+1})\Big)$$

(where $\tau_n = \sum_{k=0}^n \gamma_k$), is almost surely a bounded APT of the semiflow $\bar{\Phi}$ defined by (ODE_∞). The proof is concluded by applying Prop. 2.20 and Prop. 3.16.

2.9.2 Proof of Th. 2.7

As $\inf F > -\infty$, one can assume without loss of generality that $F \geq 0$. In the sequel, C denotes some positive constant which may change from line to line. We define $a_n := (1 - \alpha_{n+1})/\gamma_n$ and $P_n := \frac{1}{2a_n r_n}\langle m_n^{\odot 2}, \frac{1}{\varepsilon + \sqrt{\hat{v}_n}}\rangle$. We have $a_n \to a$ and $r_n \to 1$. By Assumption 2.5.2-*i)*,

$$F(x_n) \leq F(x_{n-1}) - \gamma_n\langle \nabla F(x_n), \frac{\hat{m}_n}{\varepsilon + \sqrt{\hat{v}_n}}\rangle + C\gamma_n^2 P_n. \tag{2.35}$$

We set $u_n := 1 - \frac{a_{n+1}}{a_n}$ and $D_n := \frac{r_n^{-1}}{\varepsilon + \sqrt{\hat{v}_n}}$, so that $P_n = \frac{1}{2a_n}\langle D_n, m_n^{\odot 2}\rangle$. We can write:

$$P_{n+1} - P_n = u_n P_{n+1} + \langle \frac{D_{n+1} - D_n}{2a_n}, m_{n+1}^{\odot 2}\rangle + \langle \frac{D_n}{2a_n}, m_{n+1}^{\odot 2} - m_n^{\odot 2}\rangle. \tag{2.36}$$

We estimate the vector $D_{n+1} - D_n$. Using that (r_n^{-1}) is non-increasing,

$$D_{n+1} - D_n \leq r_n^{-1} \frac{\sqrt{\hat{v}_n} - \sqrt{\hat{v}_{n+1}}}{(\varepsilon + \sqrt{\hat{v}_{n+1}}) \odot (\varepsilon + \sqrt{\hat{v}_n})}\,.$$

Remarking that $v_{n+1} \geq \beta_{n+1} v_n$, recalling that $(\bar{r}_n)$ is nondecreasing and using the update rules of v_n and $\bar{r}_n$, we obtain after some algebra

$$\begin{aligned}\sqrt{\hat{v}_n} - \sqrt{\hat{v}_{n+1}} &= \bar{r}_{n+1}^{-\frac{1}{2}}(1-\beta_{n+1})\frac{v_n - \nabla f_{n+1}^{\odot 2}}{\sqrt{v_n}+\sqrt{v_{n+1}}} + \frac{\bar{r}_{n+1}-\bar{r}_n}{\sqrt{\bar{r}_n}(\sqrt{\bar{r}_n}+\sqrt{\bar{r}_{n+1}})}\sqrt{\frac{v_n}{\bar{r}_{n+1}}} \\ &\leq c_{n+1}\sqrt{\hat{v}_{n+1}} \text{ where } c_{n+1} := \frac{1-\beta_{n+1}}{\sqrt{\beta_{n+1}}}\left(\frac{1}{1+\sqrt{\beta_{n+1}}} + \frac{1-\bar{r}_n}{2\bar{r}_n}\right). \end{aligned} \tag{2.37}$$

It is easy to see that $c_n/\gamma_n \to b/2$. Thus, for any $\delta > 0$, $c_{n+1} \leq (b+2\delta)\gamma_n/2$ for all n large enough. Using also that $\sqrt{\hat{v}_{n+1}}/(\varepsilon + \sqrt{\hat{v}_{n+1}}) \leq 1$, we obtain that $D_{n+1} - D_n \leq \frac{b+2\delta}{2}\gamma_n D_n$. Substituting this inequality in Eq. (3.43), we get

$$P_{n+1} - P_n \leq u_n P_{n+1} + \gamma_n \langle \frac{b+2\delta}{4a_n} D_n, m_{n+1}^{\odot 2} \rangle + \langle \frac{D_n}{2a_n}, m_{n+1}^{\odot 2} - m_n^{\odot 2} \rangle .$$

Using $m_{n+1}^{\odot 2} - m_n^{\odot 2} = 2m_n \odot (m_{n+1} - m_n) + (m_{n+1} - m_n)^{\odot 2}$, and noting that $\mathbb{E}_n(m_{n+1} - m_n) = a_n\gamma_n(\nabla F(x_n) - m_n)$,

$$\mathbb{E}_n \langle \frac{D_n}{2a_n}, m_{n+1}^{\odot 2} - m_n^{\odot 2} \rangle = \gamma_n \langle \nabla F(x_n), \frac{\hat{m}_n}{\varepsilon + \sqrt{\hat{v}_n}} \rangle - 2a_n\gamma_n P_n + \langle \frac{D_n}{2a_n}, \mathbb{E}_n[(m_{n+1} - m_n)^{\odot 2}] \rangle$$

As $a_n \to a$, we have $a_n - \frac{b+2\delta}{4} \geq a - \frac{b+\delta}{4}$ for all n large enough. Hence,

$$\begin{aligned}\mathbb{E}_n P_{n+1} - P_n \leq u_n P_{n+1} - 2(a - \frac{b+\delta}{4})\gamma_n P_n + \gamma_n \langle \nabla F(x_n), \frac{\hat{m}_n}{\varepsilon + \sqrt{\hat{v}_n}} \rangle \\ + \gamma_n^2 \frac{b+2\delta}{2} \langle \nabla F(x_n), \frac{\hat{m}_n}{\varepsilon + \sqrt{\hat{v}_n}} \rangle + C \langle \frac{D_n}{2a_n}, \mathbb{E}_n[(m_{n+1} - m_n)^{\odot 2}] \rangle .\end{aligned}$$

Using the Cauchy-Schwartz inequality and Assumption 2.5.2 *ii)*, it is easy to show the inequality $\langle \nabla F(x_n), \frac{\hat{m}_n}{\varepsilon + \sqrt{\hat{v}_n}} \rangle \leq C(1 + F(x_n) + P_n)$. Moreover, using the componentwise inequality $(\nabla f_{n+1} - m_n)^{\odot 2} \leq 2\nabla f_{n+1}^{\odot 2} + 2m_n^{\odot 2}$ along with Assumption 2.5.2 *ii)*, we obtain

$$\langle \tfrac{D_n}{2a_n}, \mathbb{E}_n[(m_{n+1} - m_n)^{\odot 2}] \rangle \leq 2(1-\alpha_{n+1})^2 \langle \tfrac{D_n}{2a_n}, \mathbb{E}_n[\nabla f_{n+1}^{\odot 2}] + m_n^{\odot 2} \rangle \leq C\gamma_n^2(1 + F(x_n) + P_n) .$$

Putting all pieces together with Eq. (3.49),

$$\mathbb{E}_n(F(x_n) + P_{n+1}) \leq F(x_{n-1}) + P_n + u_n P_{n+1} - 2(a - \tfrac{b+\delta}{4})\gamma_n P_n + C\gamma_n^2(1 + F(x_n) + P_n) . \tag{2.38}$$

Define $V_n := (1 - C\gamma_{n-1}^2)F(x_{n-1}) + (1 - u_{n-1})P_n$ where the constant C is fixed so that Eq. (2.38) holds. Then,

$$\mathbb{E}_n(V_{n+1}) \leq V_n - \left(2a - \frac{b+\delta}{2} - \frac{u_{n-1}}{\gamma_n}\right)\gamma_n P_n + C\gamma_n^2(1 + P_n) + C\gamma_{n-1}^2 F(x_{n-1}) . \tag{2.39}$$

By Assumption 2.5.2, $\limsup_n u_{n-1}/\gamma_n < 2a - b/2$ and for δ small enough, we obtain

$$\mathbb{E}_n(V_{n+1}) \leq V_n + C\gamma_n^2(1 + P_n) + C\gamma_{n-1}^2 F(x_{n-1}) \leq (1 + C'\gamma_n^2)V_n + C\gamma_n^2 .$$

By the Robbins-Siegmund's theorem (Robbins and Siegmund, 1971), the sequence (V_n) converges almost surely to a finite random variable $V_\infty \in \mathbb{R}^+$. In turn, the coercivity of F implies that (x_n) is almost surely bounded. The Robbins-Siegmund's theorem

is sometimes used to prove *at the same time* the stability of an algorithm *and* its convergence. Here, we only use it to establish our stability result. In particular, we do not exploit the "repelling" term in Eq. (2.39) which we upperbound by zero. We now establish the almost sure boundedness of (m_n). Consider the martingale difference sequence $\Delta_{n+1} := \nabla f_{n+1} - \nabla F(x_n)$. We decompose $m_n = \bar{m}_n + \tilde{m}_n$ where $\bar{m}_{n+1} = \alpha_{n+1}\bar{m}_n + (1-\alpha_{n+1})\nabla F(x_n)$ and $\tilde{m}_{n+1} = \alpha_{n+1}\tilde{m}_n + (1-\alpha_{n+1})\Delta_{n+1}$, setting $\bar{m}_0 = \tilde{m}_0 = 0$. We prove that both terms $\bar{m}_n$ and $\tilde{m}_n$ are bounded. Consider the first term: $\|\bar{m}_{n+1}\| \leq \alpha_{n+1}\|\bar{m}_n\| + (1-\alpha_{n+1})\sup_k \|\nabla F(x_k)\|$. By continuity of ∇F, the supremum in the above inequality is almost surely finite. Thus, for every n, the ratio $\|\bar{m}_n\| / \sup_k \|\nabla F(x_k)\|$ is upperbounded by the bounded sequence r_n. Hence, $(\bar{m}_n)$ is bounded w.p.1. Consider now the term $\tilde{m}_n$:

$$\mathbb{E}_n(\|\tilde{m}_{n+1}\|^2) = \alpha_{n+1}^2\|\tilde{m}_n\|^2 + (1-\alpha_{n+1})^2\mathbb{E}_n(\|\Delta_{n+1}\|^2) \leq (1+(1-\alpha_{n+1})^2)\|\tilde{m}_n\|^2 + (1-\alpha_{n+1})^2 C\,,$$

where C is a finite random variable (independent of n) s.t. $\mathbb{E}_n(\|\nabla f_{n+1}\|^2) \leq C$ by Assumption 2.4.2 *i)*. Here, we used $\alpha_{n+1}^2 \leq (1+(1-\alpha_{n+1})^2)$ and the inequality $\mathbb{E}_n(\|\Delta_{n+1}\|^2) \leq \mathbb{E}_n(\|\nabla f_{n+1}\|^2)$. By Assumption 2.5.1, $\sum_n (1-\alpha_{n+1})^2 < \infty$. By the Robbins-Siegmund theorem, it follows that $\sup_n \|\tilde{m}_n\|^2 < \infty$ w.p.1. Finally, it can be shown that (v_n) is almost surely bounded using the same arguments.

2.9.3 Proof of Th. 2.8

We use (Pelletier, 1998, Th. 1). All the assumptions in the latter can be verified in our case, at the exception of a positive definiteness condition on the limiting covariance matrix, which corresponds, in our case, to the matrix Q given by Eq. (2.16). As Q is not positive definite, it is strictly speaking not possible to just cite and apply (Pelletier, 1998, Th. 1). Nevertheless, a detailed inspection of the proofs of Pelletier (1998) shows that only a minor adaptation is needed in order to cover the present case. Therefore, proving the convergence result of Pelletier (1998) from scratch is worthless. It is sufficient to verify the assumptions of (Pelletier, 1998, Th. 1) (except the definiteness of Q) and then to point out the specific part of the proof of Pelletier (1998) which requires some adaptation.

Let $z_n = (x_n, m_n, v_n)$ be the output of Algorithm 2.2. Define $z^* = (x^*, 0, S(x^*))$. Define $\eta_{n+1} := (0, a(\nabla f_{n+1} - \nabla F(x_n)), b(\nabla f_{n+1}^{\odot 2} - S(x_n)))$. We have

$$z_{n+1} = z_n + \gamma_{n+1} h_\infty(z_n) + \gamma_{n+1}\eta_{n+1} + \gamma_{n+1}\epsilon_{n+1}\,, \tag{2.40}$$

where $\epsilon_{n+1} := (\epsilon_{n+1}^1, \epsilon_{n+1}^2, \epsilon_{n+1}^3)$, whose components are given by

$$\epsilon_{n+1}^1 = \tfrac{m_n}{\varepsilon+\sqrt{v_n}} - \tfrac{\hat{m}_{n+1}}{\varepsilon+\sqrt{\hat{v}_{n+1}}};\ \epsilon_{n+1}^2 = \left(\tfrac{1-\alpha_{n+1}}{\gamma_{n+1}} - a\right)\left(\nabla f_{n+1} - m_n\right);\ \epsilon_{n+1}^3 = \left(\tfrac{1-\beta_{n+1}}{\gamma_{n+1}} - b\right)\left(\nabla f_{n+1}^{\odot 2} - v_n\right).$$

Here, η_{n+1} is a martingale increment noise and $\epsilon_{n+1} = (\epsilon_{n+1}^1, \epsilon_{n+1}^2, \epsilon_{n+1}^3)$ is a remainder term. The aim is to check the assumptions (A1.1) to (A1.3) of Pelletier (1998), where the role of the quantities $(h, \varepsilon_n, r_n, \sigma_n, \alpha, \rho, \beta)$ in Pelletier (1998) is respectively played by the quantities $(h_\infty, \eta_n, \epsilon_n, \gamma_n, \kappa, 1, 1)$ of the present chapter.

Let us first consider Assumption (A1.1) for h_∞. By construction, $h_\infty(z^*) = 0$. By Assumptions 2.5.3 and 2.2.3, h_∞ is continuously differentiable in the neighborhood of z^* and its Jacobian at z^* coincides with the matrix H given by Eq. (2.14). As already discussed, after some algebra, it can be shown that the largest real part of the eigenvalues of H coincides with $-L$ where $L > 0$ is given by Eq. (3.9). Hence,

Assumption (A1.1) of Pelletier (1998) is satisfied for h_∞. Assumption (A1.3) is trivially satisfied using Assumption 2.5.4. The crux is therefore to verify Assumption (A1.2). Clearly, $\mathbb{E}(\eta_{n+1}|\mathcal{F}_n) = 0$. Using Assumption 2.4.2*ii)*, it follows from straightforward manipulations based on Jensen's inequality that for any $M > 0$, there exists $\delta > 0$ s.t. $\sup_{n\geq 0} \mathbb{E}_n \left(\|\eta_{n+1}\|^{2+\delta} \right) \mathbb{1}_{\{\|z_n - z^*\| \leq M\}} < \infty$. Next, we verify the condition

$$\lim_{n\to\infty} \mathbb{E} \left(\gamma_{n+1}^{-1} \|\epsilon_{n+1}\|^2 \mathbb{1}_{\{\|z_n - z^*\| \leq M\}} \right) = 0 \,. \tag{2.41}$$

It is sufficient to verify the latter for ϵ_n^i $(i = 1, 2, 3)$ in place of ϵ_n. The map $(m, v) \mapsto m/(\varepsilon + \sqrt{v})$ is Lipschitz continuous in a neighborhood of $(0, S(x^*))$ by Assumption 2.2.3. Thus, for M small enough, there exists a constant C s.t. if $\|z_n - z^*\| \leq M$, then $\|\epsilon_{n+1}^1\| \leq C \left\| r_{n+1}^{-1} m_{n+1} - m_n \right\| + C \left\| \bar{r}_{n+1}^{-1} v_{n+1} - v_n \right\|$. Using the triangular inequality and the fact that $r_{n+1}, \bar{r}_{n+1}$ are bounded sequences away from zero, there exists another constant C s.t.

$$\|\epsilon_{n+1}^1\| \leq C \left\| m_{n+1} - m_n \right\| + C \left\| v_{n+1} - v_n \right\| + C|r_{n+1} - 1| + C|\bar{r}_{n+1} - 1| \,.$$

Using Lem. 2.29 under Assumption 2.5.4 (note that $\gamma_0 > 1/2L \geq 1/a$ when $\kappa = 1$), we obtain that the sequence $|r_n - 1|/\gamma_n$ is bounded, thus $|r_{n+1} - 1| \leq C\gamma_{n+1}$.

The sequence $(1 - \alpha_n)/\gamma_n$ being also bounded, it holds that

$$\|m_{n+1} - m_n\|^2 \mathbb{1}_{\{\|z_n - z^*\| \leq M\}} \leq C\gamma_{n+1}^2 (1 + \|\nabla f_{n+1}\|^2) \mathbb{1}_{\{\|z_n - z^*\| \leq M\}} \,.$$

By Assumption 2.4.2 *ii)*, $\mathbb{E}_n(\|\nabla f_{n+1}\|^2|)$ is bounded by a deterministic constant on $\{\|z_n - z^*\| \leq M\}$. Thus, $\mathbb{E}_n(\|m_{n+1} - m_n\|^2 \mathbb{1}_{\{\|z_n - z^*\| \leq M\}}) \leq C\gamma_{n+1}^2$. A similar result holds for $\|v_{n+1} - v_n\|^2$. We have thus shown that $\mathbb{E}_n \left(\|\epsilon_{n+1}^1\|^2 \mathbb{1}_{\{\|z_n - z^*\| \leq M\}} \right) \leq C\gamma_{n+1}^2$. Hence, Eq. (2.41) is proved for ϵ_{n+1}^1 in place of ϵ_{n+1}. Under Assumption 2.5.4, the proof uses the same kind of arguments for ϵ_{n+1}^2, ϵ_{n+1}^3 and is omitted. Finally, Eq. (2.41) is proved. Continuing the verification of Assumption (A1.2), we establish that

$$\mathbb{E}_n(\eta_{n+1}\eta_{n+1}^T) \to Q \text{ a.s. on } \{z_n \to z^*\} \,. \tag{2.42}$$

Denote by $\bar{Q}(x)$ the matrix given by the righthand side of Eq. (2.16) when x^* is replaced by an arbitrary $x \in \mathcal{V}$. It is easily checked that $\mathbb{E}_n(\eta_{n+1}\eta_{n+1}^T) = \bar{Q}(x_n)$ and by continuity, $\bar{Q}(x_n) \to Q$ a.s. on $\{z_n \to z^*\}$, which proves (2.42). Therefore, Assumption (A1.2) is fulfilled, except for the point mentioned at the beginning of this section : Pelletier (1998) puts the additional condition that the limit matrix in Eq. (2.42) is positive definite. This condition is not satisfied in our case, but the proof can still be adapted. The specific part of the proof where the positive definiteness comes into play is Th. 7 in Pelletier (1998). The proof of (Pelletier, 1998, Th. 1) can therefore be adapted to the case of a positive semidefinite matrix. In the proof of (Pelletier, 1998, Th. 7), we only substitute the inverse of the square root of Q by the Moore-Penrose inverse. Finally, the uniqueness of the stationary distribution μ and its expression follow from (Karatzas and Shreve, 1991, Th. 6.7, p. 357).

Proof of Eq. (3.10). We introduce the $d \times d$ blocks of the $3d \times 3d$ matrix $\Sigma = \left(\Sigma_{i,j}\right)_{i,j=1,2,3}$ where $\Sigma_{i,j}$ is $d \times d$. We denote by $\tilde{\Sigma}$ the $2d \times 2d$ submatrix $\tilde{\Sigma} := \left(\Sigma_{i,j}\right)_{i,j=1,2}$. By Th. 2.8, we have the subsystem:

$$\tilde{H}\tilde{\Sigma} + \tilde{\Sigma}\tilde{H}^T = \begin{pmatrix} 0 & 0 \\ 0 & -a^2\tilde{Q} \end{pmatrix} \qquad \text{where } \tilde{H} := \begin{pmatrix} \zeta I_d & -D \\ a\nabla^2 F(x^*) & (\zeta - a) I_d \end{pmatrix} \tag{2.43}$$

and where $\tilde{Q} := \mathrm{Cov}\left(\nabla f(x^*, \xi)\right)$. The next step is to triangularize the matrix $\tilde{H}$ in order to decouple the blocks of $\tilde{\Sigma}$. For every $k = 1, \dots, d$, set $\nu_k^{\pm} := -\frac{a}{2} \pm \sqrt{a^2/4 - a\lambda_k}$ with the convention that $\sqrt{-1} = \imath$ (inspecting the characteristic polynomial of $\tilde{H}$, these are the eigenvalues of $\tilde{H}$). Set $M^{\pm} := \mathrm{diag}\,(\nu_1^{\pm}, \cdots, \nu_d^{\pm})$ and $R^{\pm} := D^{-1/2} P M^{\pm} P^T D^{-1/2}$. Using the identities $M^+ + M^- = -aI_d$ and $M^+ M^- = a\Lambda$ where $\Lambda := \mathrm{diag}\,(\lambda_1, \cdots, \lambda_d)$, it can be checked that

$$\mathcal{R}\tilde{H} = \begin{pmatrix} DR^+ + \zeta I_d & -D \\ 0 & R^- D + \zeta I_d \end{pmatrix} \mathcal{R}, \text{ where } \mathcal{R} := \begin{pmatrix} I_d & 0 \\ R^+ & I_d \end{pmatrix}.$$

Set $\check{\Sigma} := \mathcal{R}\tilde{\Sigma}\mathcal{R}^T$. Denote by $(\check{\Sigma}_{i,j})_{i,j=1,2}$ the blocks of $\check{\Sigma}$. Note that $\check{\Sigma}_{1,1} = \Sigma_{1,1}$. By left/right multiplication of Eq. (3.55) respectively with $\mathcal{R}$ and $\mathcal{R}^T$, we obtain

$$(DR^+ + \zeta I_d)\Sigma_{1,1} + \Sigma_{1,1}(R^+ D + \zeta I_d) = \check{\Sigma}_{1,2} D + D\check{\Sigma}_{1,2}^T \tag{2.44}$$

$$(DR^+ + \zeta I_d)\check{\Sigma}_{1,2} + \check{\Sigma}_{1,2}(DR^- + \zeta I_d) = D\check{\Sigma}_{2,2} \tag{2.45}$$

$$(R^- D + \zeta I_d)\check{\Sigma}_{2,2} + \check{\Sigma}_{2,2}(DR^- + \zeta I_d) = -a^2\tilde{Q} \tag{2.46}$$

Set $\bar{\Sigma}_{2,2} = P^{-1} D^{1/2} \check{\Sigma}_{2,2} D^{1/2} P$. Define $C := P^{-1} D^{1/2} \tilde{Q} D^{1/2} P$. Eq. (3.58) yields $(M^- + \zeta I_d)\bar{\Sigma}_{2,2} + \bar{\Sigma}_{2,2}(M^- + \zeta I_d) = -a^2 C$. Set $\bar{\Sigma}_{1,2} = P^{-1} D^{-1/2} \check{\Sigma}_{1,2} D^{1/2} P$. Eq. (3.57) rewrites $(M^+ + \zeta I_d)\bar{\Sigma}_{1,2} + \bar{\Sigma}_{1,2}(M^- + \zeta I_d) = \bar{\Sigma}_{2,2}$. We obtain that $\bar{\Sigma}_{1,2}^{k,\ell} = (\nu_k^+ + \nu_\ell^- + 2\zeta)^{-1} \bar{\Sigma}_{2,2}^{k,\ell} = \frac{-a^2 C_{k,\ell}}{(\nu_k^+ + \nu_\ell^- + 2\zeta)(\nu_k^- + \nu_\ell^- + 2\zeta)}$. Set $\bar{\Sigma}_{1,1} = P^{-1} D^{-1/2} \Sigma_{1,1} D^{-1/2} P$. Eq. (3.56) becomes $(M^+ + \zeta I_d)\bar{\Sigma}_{1,1} + \bar{\Sigma}_{1,1}(M^+ + \zeta I_d) = \bar{\Sigma}_{1,2} + \bar{\Sigma}_{1,2}^T$. Thus,

$$\begin{aligned} \bar{\Sigma}_{1,1}^{k,\ell} &= \frac{\bar{\Sigma}_{1,2}^{k,\ell} + \bar{\Sigma}_{1,2}^{\ell,k}}{\nu_k^+ + \nu_\ell^+ + 2\zeta} = \frac{-a^2 C_{k,\ell}}{(\nu_k^+ + \nu_\ell^+ + 2\zeta)(\nu_k^- + \nu_\ell^- + 2\zeta)} \left(\frac{1}{\nu_k^+ + \nu_\ell^- + 2\zeta} + \frac{1}{\nu_k^- + \nu_\ell^+ + 2\zeta} \right) \\ &= \frac{C_{k,\ell}}{(1 - \frac{2\zeta}{a})(\lambda_k + \lambda_\ell - 2\zeta + \frac{2}{a}\zeta^2) + \frac{1}{2(a-2\zeta)}(\lambda_k - \lambda_\ell)^2}, \end{aligned}$$

and the result is proved.

3

Stochastic Optimization with Momentum: Convergence, Fluctuations, and Traps Avoidance

Abstract In this chapter, a general stochastic optimization procedure is studied, unifying several variants of the stochastic gradient descent such as, among others, the stochastic heavy ball method, the Stochastic Nesterov Accelerated Gradient algorithm (S-NAG), and the widely used ADAM algorithm. The algorithm is seen as a noisy Euler discretization of a non-autonomous ordinary differential equation, recently introduced by Belotto da Silva and Gazeau, which is analyzed in depth. Assuming that the objective function is non-convex and differentiable, the stability and the almost sure convergence of the iterates to the set of critical points are established. A noteworthy special case is the convergence proof of S-NAG in a non-convex setting. Under some assumptions, the convergence rate is provided under the form of a Central Limit Theorem. Finally, the non-convergence of the algorithm to undesired critical points, such as local maxima or saddle points, is established. Here, the main ingredient is a new avoidance of traps result for non-autonomous settings, which is of independent interest.

3.1 Introduction

Given a probability space Ξ, an integer $d > 0$, and a function $f : \mathbb{R}^d \times \Xi \to \mathbb{R}$, consider the problem of finding a local minimum of the function $F(x) := \mathbb{E}_\xi[f(x, \xi)]$ w.r.t. $x \in \mathbb{R}^d$, where $\mathbb{E}_\xi$ represents the expectation w.r.t. the random variable ξ on Ξ. The chapter focuses on the case where F is possibly non-convex. It is assumed that the function F is unknown to the observer, either because the distribution of ξ is unknown, or because the expectaction cannot be evaluated. Instead, a sequence $(\xi_n : n \geq 1)$ of i.i.d. copies of the random variable ξ is revealed online.

While the Stochastic Gradient Descent is the most classical algorithm that is used to solve such a problem, recently, several other algorithms became very popular. These include the Stochastic Heavy Ball (SHB), the stochastic version of Nesterov's Accelerated Gradient method (S-NAG) and the large class of the so-called *adaptive* gradient algorithms, among which ADAM (Kingma and Ba, 2015) is perhaps the most used in practice. As opposed to the vanilla Stochastic Gradient Descent, the study of such algorithms is more elaborate, for three reasons. First, the update of the iterates involves a so-called *momentum* term, or inertia, which has the effect of "smoothing" the increment between two consecutive iterates. Second, the update equation at the time index n is likely to depend on n, making these systems inherently *non-autonomous*. Third, as far as adaptive algorithms are concerned, the update also depends on some additional variable (*a.k.a.* the learning rate) computed online as a function of the history of the computed gradients.

In this work, we study in a unified way the asymptotic behavior of these algorithms in the situation where F is a differentiable function which is not necessarily convex, and where the stepsize of the algorithm is decreasing.

Our starting point is a generic non-autonomous Ordinary Differential Equation (ODE) introduced by Belotto da Silva and Gazeau (2020) (see also Chapter 2 for ADAM), depicting the continuous-time versions of the aforementioned florilegium of algorithms. The solutions to the ODE are shown to converge to the set of critical points of F. This suggests that a general provably convergent algorithm can be obtained by means of an Euler discretization of the ODE, including possible stochastic perturbations. Special cases of our general algorithm include SHB, ADAM and S-NAG. We establish the almost sure boundedness and the convergence to critical points. Under additional assumptions, we obtain convergence rates, under the form of a central limit theorem. These results are new. They extend the work of Gadat et al. (2018) and Chapter 2 to a general setting. In particular, we highlight the almost sure convergence result of S-NAG in a non-convex setting, which is new to the best of our knowledge.

Next, we address the question of the avoidance of "traps". In a non-convex setting, the set of critical points of a function F is generally larger than the set of local minimizers. A "trap" stands for a critical point at which the Hessian matrix of F has negative eigenvalues, namely, it is a local maximum or saddle point. We establish that the iterates cannot converge to such a point, if the noise is exciting in some directions. The result extends the previous work of Gadat et al. (2018) obtained in the context of SHB. This result not only allows to study a broader class of algorithms but also significantly weakens the assumptions. In particular, Gadat et al. (2018) use a sub-Gaussian assumption on the noise and a rather stringent assumption on the stepsizes. The main difficulty in the approach of Gadat et al. (2018) lies in the use of the classical autonomous version of Poincaré's invariant manifold theorem. The key ingredient of our proof is a general avoidance of traps result, adapted to non-autonomous settings, which we believe to be of independent interest. It extends usual avoidance of traps results to a non-autonomous setting, by making use of a non-autonomous version of Poincaré's theorem (Daleckii and Krein, 1974; Kloeden and Rasmussen, 2011).

Chapter organization. In Section 3.2, we introduce and study the ODE's governing our general stochastic algorithm. We establish the existence and uniqueness of the solutions, as well as the convergence to the set of critical points. In Section 3.3, we introduce the main algorithm. We provide sufficient conditions under which the iterates are bounded and converge to the set of critical points. A central limit theorem is stated. Section 3.4 introduces a general avoidance of traps result for non-autonomous settings. Next, this result is applied to the proposed algorithm. Sections 3.5, 3.6 and 3.7 are devoted to the proofs of the results of Sections 3.2, 3.3 and 3.4, respectively.

Notations. Given an integer $d \geq 1$, two vectors $x, y \in \mathbb{R}^d$, and a real α, we denote by $x \odot y$, $x^{\odot\alpha}$, x/y, $|x|$, and $\sqrt{|x|}$ the vectors in $\mathbb{R}^d$ whose i-th coordinates are respectively given by $x_i y_i$, x_i^α, x_i/y_i, $|x_i|$, $\sqrt{|x_i|}$. Inequalities of the form $x \leq y$ are to be read componentwise. The standard Euclidean norm is denoted $\|\cdot\|$. Notation M^{T} represents the transpose of a matrix M. For $x \in \mathbb{R}^d$ and $\rho > 0$, the notation $B(x, \rho)$ stands for the open ball of $\mathbb{R}^d$ with center x and radius ρ. We also write $\mathbb{R}_+ = [0, \infty)$. If $z \in \mathbb{R}^d$ and $A \subset \mathbb{R}^d$, we write $\mathrm{dist}(z, A) := \inf\{\|z - z'\| : z' \in A\}$. By $\mathbb{1}_A(x)$, we refer to the function that is equal to one if $x \in A$ and to zero elsewhere. The set of zeros of a function $h : \mathbb{R}^d \to \mathbb{R}^{d'}$ is $\operatorname{zer} h = \{x : h(x) = 0\}$. Let D be a domain in $\mathbb{R}^d$. Given an integer $k \geq 0$, the class $\mathcal{C}^k(D, \mathbb{R})$ is the class of $D \to \mathbb{R}$ maps such that all their partial derivatives up to the order k exist and are continuous. For a function $h \in \mathcal{C}^k(D, \mathbb{R})$ and for every $i \in \{1, \ldots, d\}$, we denote as $\partial_i^k h(x_1, \ldots, x_d)$ the k^{th} partial derivative of the

function h with respect to x_i. When $k = 1$, we just write $\partial_i h(x_1, \dots, x_d)$. The gradient of a function $F : \mathbb{R}^d \to \mathbb{R}$ at a point $x \in \mathbb{R}^d$ is denoted as $\nabla F(x)$, and its Hessian matrix at x is $\nabla^2 F(x)$ as usual. For a function $S : \mathbb{R}^d \to \mathbb{R}^d$, the notation $\nabla S(x)$ stands for the jacobian matrix of S at point x. In this chapter, we bring to the attention of the reader that few notations may slightly differ from Chapter 2: the variables (v, m, x) to be used below were treated as (x, m, v) in Chapter 2, the vector fields are also impacted by this permutation. Moreover, the time variable t occurs as a second variable of the vector field in the non-autonomous ODE of this chapter, whereas it was first in Chapter 2.

3.2 Ordinary differential equations

3.2.1 A general ODE

Our starting point will be a non-autonomous ODE slightly extending the one in Chapter 2 and which is almost identical to the one introduced in Belotto da Silva and Gazeau (2020). Let F be a function in $\mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$, let S be a continuous $\mathbb{R}^d \to \mathbb{R}^d_+$ function, let $\mathsf{h}, \mathsf{r}, \mathsf{p}, \mathsf{q} : (0, \infty) \to \mathbb{R}_+$ be four continuous functions, and let $\varepsilon > 0$. Let $v_0 \in \mathbb{R}^d_+$ and $x_0, m_0 \in \mathbb{R}^d$. Starting at $\mathsf{v}(0) = v_0$, $\mathsf{m}(0) = m_0$, and $\mathsf{x}(0) = x_0$, our ODE on $\mathbb{R}_+$ with trajectories in $\mathcal{Z}_+ := \mathbb{R}^d_+ \times \mathbb{R}^d \times \mathbb{R}^d$ reads

$$\begin{cases} \dot{\mathsf{v}}(t) &= \mathsf{p}(t) S(\mathsf{x}(t)) - \mathsf{q}(t)\mathsf{v}(t) \\ \dot{\mathsf{m}}(t) &= \mathsf{h}(t) \nabla F(\mathsf{x}(t)) - \mathsf{r}(t)\mathsf{m}(t) \\ \dot{\mathsf{x}}(t) &= -\mathsf{m}(t)/\sqrt{\mathsf{v}(t) + \varepsilon} \end{cases} \tag{ODE-1}$$

This ODE can be rewritten compactly as follows. Write $z_0 = (v_0, m_0, x_0)$, and let $\mathsf{z}(t) = (\mathsf{v}(t), \mathsf{m}(t), \mathsf{x}(t)) \in \mathcal{Z}_+$ for $t \in \mathbb{R}_+$. Let $\mathcal{Z} := \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$, and define the map $g : \mathcal{Z}_+ \times (0, \infty) \to \mathcal{Z}$ as

$$g(z, t) = \begin{bmatrix} \mathsf{p}(t) S(x) - \mathsf{q}(t) v \\ \mathsf{h}(t) \nabla F(x) - \mathsf{r}(t) m \\ -m/\sqrt{v + \varepsilon} \end{bmatrix} \tag{3.1}$$

for $z = (v, m, x) \in \mathcal{Z}_+$. With these notations, we can rewrite (ODE-1) as

$$\mathsf{z}(0) = z_0, \quad \dot{\mathsf{z}}(t) = g(\mathsf{z}(t), t) \text{ for } t > 0.$$

By setting $S(x) = \nabla F(x)^{\odot 2}$ when necessary and by properly choosing the functions h, r, p, and q, a large number of iterative algorithms used in Machine Learning can be obtained by an Euler's discretization of this ODE. For instance, choosing $\mathsf{h}(t) = \mathsf{r}(t) = a(t, \lambda, \alpha_1)$ and $\mathsf{p}(t) = \mathsf{q}(t) = a(t, \lambda, \alpha_2)$ with $a(t, \lambda, \alpha) = \lambda^{-1}(1 - \exp(-\lambda\alpha))/(1 - \exp(-\alpha t))$ and $\lambda, \alpha_1, \alpha_2 > 0$, one obtains a version of the Adam algorithm (Kingma and Ba, 2015) (see (Belotto da Silva and Gazeau, 2020, Sections 2.4-4.2) for details). To give another less specific example, if we set $\mathsf{p} = \mathsf{q} \equiv 0$, then the resulting ODE covers a family of algorithms to which the well-known Heavy Ball with friction algorithm (Attouch et al., 2000) belongs. For a comprehensive and more precise view of the deterministic algorithms that can be deduced from (ODE-1) by an Euler's discretization, the reader is referred to (Belotto da Silva and Gazeau, 2020, Table 1).

In this chapter, since we are rather interested in stochastic versions of these algorithms, Eq. (ODE-1) will be the basic building block of the classical "ODE method" which is widely used in the field of stochastic approximation (Benaïm, 1999). In order to analyze

the behavior of this equation in preparation of the stochastic analysis, we need the following assumptions.

Assumption 3.2.1. The function F belongs to $\mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$ and ∇F is locally Lipschitz continuous.

Assumption 3.2.2. F is coercive, *i.e.*, $F(x) \to +\infty$ as $\|x\| \to +\infty$.

Note that this assumption implies that the infimum $F_\star$ of F is finite, and the set $\operatorname{zer} \nabla F$ of zeros of ∇F is nonempty.

Assumption 3.2.3. The map $S : \mathbb{R}^d \to \mathbb{R}^d_+$ is locally Lipschitz continuous.

Assumption 3.2.4. The continuous functions $\mathsf{h}, \mathsf{r}, \mathsf{p}, \mathsf{q} : (0, +\infty) \to \mathbb{R}_+$ satisfy:

i) $\mathsf{h} \in \mathcal{C}^1((0, +\infty), \mathbb{R}_+)$, $\dot{\mathsf{h}}(t) \leq 0$ on $(0, +\infty)$ and the limit $h_\infty := \lim_{t\to\infty} \mathsf{h}(t)$ is positive.
ii) r and q are non-increasing and $r_\infty := \lim_{t\to\infty} \mathsf{r}(t)$, $q_\infty := \lim_{t\to\infty} \mathsf{q}(t)$ are positive.
iii) p converges towards p_∞ as $t \to \infty$.
iv) For all $t \in (0, +\infty)$, $\mathsf{r}(t) \geq \mathsf{q}(t)/4$ and $r_\infty > q_\infty/4$.

These assumptions are sufficient to prove the existence and the uniqueness of the solution to (ODE-1) starting at a time $t_0 > 0$. The following additional assumption extends the solution to $t_0 = 0$.

Assumption 3.2.5. Either $\mathsf{h}, \mathsf{r}, \mathsf{p}, \mathsf{q} \in \mathcal{C}^1([0, +\infty), \mathbb{R}_+)$, or the following holds:

i) For every $x \in \mathbb{R}^d$, we have $S(x) \geq \nabla F(x)^{\odot 2}$.
ii) The functions $\frac{\mathsf{h}}{\mathsf{p}}$, $\frac{\mathsf{h}}{\mathsf{q}-2\mathsf{r}}$, $t \mapsto t\mathsf{h}(t)$, $t \mapsto t\mathsf{r}(t)$, $t \mapsto t\mathsf{p}(t)$, $t \mapsto t\mathsf{q}(t)$ are bounded near zero.
iii) There exists $t_0 > 0$ such that for all $t < t_0$, $2\mathsf{r}(t) - \mathsf{q}(t) > 0$.
iv) There exists $\delta > 0$ such that $\frac{\mathsf{h}}{\mathsf{r}}, \frac{\mathsf{p}}{\mathsf{q}} \in \mathcal{C}^1([0, \delta), \mathbb{R}_+)$.
v) The initial condition $z_0 = (v_0, m_0, x_0) \in \mathcal{Z}_+$ satisfies

$$m_0 = \nabla F(x_0) \lim_{t\downarrow 0} \frac{\mathsf{h}(t)}{\mathsf{r}(t)} \quad \text{and} \quad v_0 = S(x_0) \lim_{t\downarrow 0} \frac{\mathsf{p}(t)}{\mathsf{q}(t)} .$$

Remark 2. The functions $\mathsf{h}, \mathsf{r}, \mathsf{p}, \mathsf{q}$ corresponding to Adam satisfy these conditions. We leave the straightforward verifications to the reader. We just observe here that the function S that will correspond to our stochastic algorithm in Section 3.3 below will satisfy Assumption 3.2.5–*i)* by an immediate application of Jensen's inequality.

The following theorem slightly generalizes the results of (Belotto da Silva and Gazeau, 2020, Th. 3 and Th. 5).

Theorem 3.1. *Let Assumptions 3.2.1 to 3.2.4 hold true. Consider $z_0 \in \mathcal{Z}_+$ and $t_0 > 0$. Then, there exists a unique global solution $\mathsf{z} : [t_0, +\infty) \to \mathcal{Z}_+$ to (ODE-1) with initial condition $\mathsf{z}(t_0) = z_0$. Moreover, $\mathsf{z}([t_0, +\infty))$ is a bounded subset of $\mathcal{Z}_+$. As $t \to +\infty$, $\mathsf{z}(t)$ converges towards the set*

$$\Upsilon := \{z_\star = (p_\infty S(x_\star)/q_\infty, 0, x_\star) \ : \ x_\star \in \operatorname{zer} \nabla F\} . \tag{3.2}$$

If, additionally, Assumption 3.2.5 holds, then we can take $t_0 = 0$.

Remark 3. Th. 3.1 only shows the convergence of the trajectory $\mathsf{z}(t)$ towards a set. Convergence of the trajectory towards a single point is not guaranteed when the set Υ is not countable.

Remark 4. A simpler version of (ODE-1) is obtained when omitting the momentum term. It reads:

$$\begin{cases} \dot{\mathsf{v}}(t) &= \mathsf{p}(t)S(\mathsf{x}(t)) - \mathsf{q}(t)\mathsf{v}(t) \\ \dot{\mathsf{x}}(t) &= -\nabla F(\mathsf{x}(t))/\sqrt{\mathsf{v}(t)+\varepsilon}\,. \end{cases} \tag{ODE-1$'$}$$

This ODE encompasses the algorithms of the family of RMSProp (Tieleman and Hinton, 2012), as shown in Belotto da Silva and Gazeau (2020). The approach for proving the previous theorem can be adapted to (ODE-1$'$) with only minor modifications. In the proofs below, we will point out the particularities of (ODE-1$'$) when necessary.

The following paragraph is devoted to a particular case of (ODE-1), which does not satisfy Assumption 3.2.4, and which requires a more involved treatment than (ODE-1$'$).

3.2.2 The Nesterov case

Cabot et al. (2009), Su et al. (2016) and others studied the ODE

$$\ddot{\mathsf{x}}(t) + \frac{\alpha}{t}\dot{\mathsf{x}}(t) + \nabla F(\mathsf{x}(t)) = 0, \quad \alpha > 0, \quad F \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}),$$

which Euler's discretization generates the well-known Nesterov's accelerated gradient algorithm, see also Attouch et al. (2018); Aujol et al. (2019). This ODE can be rewritten as

$$\begin{cases} \dot{\mathsf{m}}(t) &= \nabla F(\mathsf{x}(t)) - \frac{\alpha}{t}\mathsf{m}(t) \\ \dot{\mathsf{x}}(t) &= -\mathsf{m}(t), \end{cases} \tag{ODE-N}$$

which is formally the particular case of (ODE-1) that is taken for $\mathsf{p}(t) = \mathsf{q}(t) = 0$, $\mathsf{h}(t) = 1$, and $\mathsf{r}(t) = \alpha/t$. Obviously, this case is not covered by Assumption 3.2.4. Moreover, it turns out that, contrary to the situation described in Remark 4 above, this case cannot be dealt with by a straightforward adaptation of the proof of Th. 3.1. The reason for this is as follows. Heuristically, the proof of Th. 3.1 is built around the fact that the solution of (ODE-1) "shadows" for large t the solution of the autonomous ODE

$$\begin{cases} \dot{\mathsf{v}}(t) &= p_\infty S(\mathsf{x}(t)) - q_\infty \mathsf{v}(t) \\ \dot{\mathsf{m}}(t) &= h_\infty \nabla F(\mathsf{x}(t)) - r_\infty \mathsf{m}(t) \\ \dot{\mathsf{x}}(t) &= -\frac{\mathsf{m}(t)}{\sqrt{\mathsf{v}(t)+\varepsilon}}, \end{cases}$$

and the latter can be shown to converge to the set Υ defined in Eq. (3.2), either under Assumption 3.2.4 or for the algorithms covered by Remark 4. This idea does not work anymore for (ODE-N), for its large–t autonomous counterpart

$$\begin{cases} \dot{\mathsf{m}}(t) &= \nabla F(\mathsf{x}(t)) \\ \dot{\mathsf{x}}(t) &= -\mathsf{m}(t). \end{cases}$$

can have solutions that do not converge to the critical points of F. As an example of such solutions, take $d = 1$ and $F(x) = x^2/2$. Then, $t \mapsto (\cos(t), \sin(t))$ is an oscillating solution of the latter ODE.

Yet, we have the following result. Up to our knowledge, the proof of the convergence below as $t \to +\infty$ is new.

Theorem 3.2. *Let Assumptions 3.2.1 and 3.2.2 hold true. Then, for each $x_0 \in \mathbb{R}^d$, there exists a unique bounded global solution $(\mathsf{m}, \mathsf{x}) : \mathbb{R}_+ \to \mathbb{R}^d \times \mathbb{R}^d$ to (ODE-N) with the initial condition $(\mathsf{m}(0), \mathsf{x}(0)) = (0, x_0)$. As $t \to +\infty$, $(\mathsf{m}(t), \mathsf{x}(t))$ converges towards the set*

$$\bar{\Upsilon} := \{(0, x_\star) \, : \, x_\star \in \operatorname{zer} \nabla F\}. \tag{3.3}$$

3.2.3 Related works

The continuous-time dynamical system (ODE-1) we consider was first introduced in (Belotto da Silva and Gazeau, 2020, Eq. (2.1)) with $S = \nabla F^{\odot 2}$. Th. 3.1 above is roughly the same as (Belotto da Silva and Gazeau, 2020, Ths. 3 and 5), with some slight differences regarding the assumptions on the function F, or Assumption 3.2.4-*iv)*. We point out that the main focus of Belotto da Silva and Gazeau (2020) is to study the properties of the *deterministic continous-time* dynamical system (ODE-1). In the present work, we highlight that the purpose of Th. 3.1 is to pave the way to our analysis of the corresponding *stochastic algorithms* in Section 3.3.

Concerning Th. 3.2, the existence and the uniqueness of a global solution to (ODE-N) has been previously shown in the literature, for instance in (Cabot et al., 2009, Prop. 2.1) or in (Su et al., 2016, Th. 1). The convergence statement in Th. 3.2 is new to the best of our knowledge. In particular, we stress that we do not make any convexity assumption on F. The closest result we are aware of is the one of Cabot et al. (2009). In (Cabot et al., 2009, Prop. 2.5), it is shown that if $\mathsf{x}(t)$ converges towards some point $\bar{x}$, then necessarily $\bar{x}$ is a critical point of F. Our result in Th. 3.2 strengthens this statement, by establishing that $\mathsf{x}(t)$ actually converges to the set of critical points.

3.3 Stochastic algorithms

In this section, we discuss the asymptotic behavior of stochastic algorithms that consist in noisy Euler's discretizations of (ODE-1) and (ODE-N) studied in the previous section.

We first set the stage. Let $(\Xi, \mathcal{T}, \mu)$ be a probability space. Denoting as $\mathcal{B}(\mathbb{R}^d)$ the Borel σ-algebra on $\mathbb{R}^d$, consider a $\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{T}$–measurable function $f : \mathbb{R}^d \times \Xi \to \mathbb{R}$ that satisfies the following assumption.

Assumption 3.3.1. The following conditions hold:

i) For every $x \in \mathbb{R}^d$, $f(x, \cdot)$ is μ–integrable.
ii) For every $s \in \Xi$, the map $f(\cdot, s)$ is differentiable. Denoting as $\nabla f(x, s)$ its gradient w.r.t. x, the function $\nabla f(x, \cdot)$ is integrable.
iii) There exists a measurable map $\kappa : \mathbb{R}^d \times \Xi \to \mathbb{R}_+$ s.t. for every $x \in \mathbb{R}^d$:
  a) The map $\kappa(x, \cdot)$ is μ–integrable,
  b) There exists $\varepsilon > 0$ s.t. for every $s \in \Xi$,

$$\forall\, u, v \in B(x, \varepsilon), \ \|\nabla f(u, s) - \nabla f(v, s)\| \leq \kappa(x, s)\|u - v\| \, .$$

Under Assumption 3.3.1, we can define the mapping $F : \mathbb{R}^d \to \mathbb{R}$ as

$$F(x) = \mathbb{E}_\xi[f(x, \xi)] \tag{3.4}$$

for all $x \in \mathbb{R}^d$, where we write $\mathbb{E}_\xi \varphi(\xi) = \int \varphi(\xi)\mu(d\xi)$. It is easy to see that the mapping F is differentiable,

$$\nabla F(x) = \mathbb{E}_\xi[\nabla f(x, \xi)]$$

for all $x \in \mathbb{R}^d$, and ∇F is locally Lipschitz.

Let $(\gamma_n)_{n\geq 1}$ be a sequence of positive real numbers satisfying

Assumption 3.3.2. $\gamma_{n+1}/\gamma_n \to 1$ and $\sum_n \gamma_n = +\infty$.

Define for every integer $n \geq 1$

$$\tau_n = \sum_{k=1}^{n} \gamma_k \,.$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $(\xi_n : n \geq 1)$ be a sequence of iid random variables defined from $(\Omega, \mathcal{F}, \mathbb{P})$ into $(\Xi, \mathcal{T}, \mu)$ with the distribution μ.

3.3.1 General algorithm

Our first algorithm is a discrete and noisy version of (ODE-1).
Let $z_0 = (v_0, m_0, x_0) \in \mathcal{Z}_+$ and $h_0, r_0, p_0, q_0 \in (0, \infty)$. Define for every $n \geq 1$

$$h_n = \mathsf{h}(\tau_n),\ r_n = \mathsf{r}(\tau_n),\ p_n = \mathsf{p}(\tau_n),\ \text{and}\ q_n = \mathsf{q}(\tau_n). \tag{3.5}$$

The algorithm is written as follows.

Algorithm 3.1 (general algorithm)

Initialization: $z_0 \in \mathcal{Z}_+$.
for $n = 1$ **to** n_{iter} **do**
$v_{n+1} = (1 - \gamma_{n+1} q_n) v_n + \gamma_{n+1} p_n \nabla f(x_n, \xi_{n+1})^{\odot 2}$
$m_{n+1} = (1 - \gamma_{n+1} r_n) m_n + \gamma_{n+1} h_n \nabla f(x_n, \xi_{n+1})$
$x_{n+1} = x_n - \gamma_{n+1} m_{n+1} / \sqrt{v_{n+1} + \varepsilon}$.
end for

We suppose throughout the chapter that $1 - \gamma_{n+1} q_n \geq 0$ for all $n \in \mathbb{N}$. This will guarantee that the quantity $\sqrt{v_n + \varepsilon}$ is always well-defined (see Algorithm 3.1). This mild assumption is satisfied as soon as $q_0 \leq \frac{1}{\gamma_1}$ since the sequence (q_n) is non-increasing and the sequence of stepsizes (γ_n) can also be supposed to be non-increasing.

Since this algorithm makes use of the function $\nabla f(x, \xi)^{\odot 2}$, a strengthening of Assumption 3.3.1 is required:

Assumption 3.3.3. In Assumption 3.3.1, Conditions *ii)* and *iii)* are respectively replaced with the stronger conditions

ii') For each $x \in \mathbb{R}^d$, the function $\nabla f(x, \cdot)^{\odot 2}$ is μ -integrable.

iii') There exists a measurable map $\kappa : \mathbb{R}^d \times \Xi \to \mathbb{R}_+$ s.t. for every $x \in \mathbb{R}^d$:

a) The map $\kappa(x, \cdot)$ is μ–integrable.
b) There exists $\varepsilon > 0$ s.t. for every $u, v \in B(x, \varepsilon)$,

$$\|\nabla f(u, s) - \nabla f(v, s)\| \vee \|\nabla f(u, s)^{\odot 2} - \nabla f(v, s)^{\odot 2}\| \leq \kappa(x, s)\|u - v\|.$$

Under Assumption 3.3.3, we can also define the mapping $S : \mathbb{R}^d \to \mathbb{R}^d$ as:

$$S(x) = \mathbb{E}_\xi[\nabla f(x,\xi)^{\odot 2}]$$

for all $x \in \mathbb{R}^d$. Notice that Assumptions 3.2.1 and 3.2.3 are satisfied for F and S.

Assumption 3.3.4. Assume either of the following conditions.

i) There exists $q \geq 2$ s.t. for every compact set $\mathcal{K} \subset \mathbb{R}^d$,

$$\sup_{x\in\mathcal{K}} \mathbb{E}_\xi \|\nabla f(x,\xi)\|^{2q} < \infty \quad \text{and} \quad \sum_n \gamma_n^{1+q/2} < \infty\,.$$

ii) For every compact set $\mathcal{K} \subset \mathbb{R}^d$, there exists a real $\sigma_\mathcal{K} \neq 0$ s.t.

$$\mathbb{E}_\xi \exp\langle u, \nabla f(x,\xi) - \nabla F(x)\rangle \mathbb{1}_{x\in\mathcal{K}} \leq \exp\left(\sigma_\mathcal{K}^2 \|u\|^2/2\right) \text{ and}$$
$$\mathbb{E}_\xi \exp\langle u, \nabla f(x,\xi)^{\odot 2} - S(x)\rangle \mathbb{1}_{x\in\mathcal{K}} \leq \exp\left(\sigma_\mathcal{K}^2 \|u\|^2/2\right)\,,$$

for every $x, u \in \mathbb{R}^d$. Moreover, for every $\alpha > 0$, $\sum_n \exp(-\alpha/\gamma_n) < \infty$.

Remark 5. We make the following comments regarding Assumption 3.3.4.

- Assumption 3.3.4-*i)* allows to use larger stepsizes in comparison to the classical condition $\sum_n \gamma_n^2 < \infty$ which corresponds to the particular case $q = 2$.
- Recall that a random vector X is said to be subgaussian if there exists a real $\sigma \neq 0$ s.t. $\mathbb{E}e^{\langle u, X\rangle} \leq e^{\sigma^2\|u\|^2/2}$ for every constant vector $u \in \mathbb{R}^d$. In Assumption 3.3.4-*ii)*, the subgaussian noise offers the possibility to use a sequence of stepsizes with an even slower decay rate than in Assumption 3.3.4–*i)*.

Assumption 3.3.5. The set $F(\{x : \nabla F(x) = 0\})$ has an empty interior.

Remark 6. Assumption 3.3.5 excludes a pathological behavior of the objective function F at critical points. It is satisfied when $F \in \mathcal{C}^k(\mathbb{R}^d, \mathbb{R})$ for $k \geq d$. Indeed, in this case, Sard's theorem stipulates that the Lebesgue measure of $F(\{x : \nabla F(x) = 0\})$ is zero in $\mathbb{R}$.

Theorem 3.3. *Let Assumptions 3.2.2, 3.2.4, and 3.3.2–3.3.5 hold true. Assume that the random sequence $(z_n = (v_n, m_n, x_n) : n \in \mathbb{N})$ given by Algorithm 3.1 is bounded with probability one. Then, w.p.1, the sequence (z_n) converges towards the set Υ defined in Eq. (3.2). If, in addition, the set of critical points of the objective function F is finite or countable, then w.p.1, the sequence (z_n) converges to a single point of Υ.*

We now deal with the boundedness problem of the sequence (z_n). We introduce an additional assumption for this purpose.

Assumption 3.3.6. The following conditions hold.

i) ∇F is (globally) Lipschitz continuous.
ii) There exists $C > 0$ s.t. for all $x \in \mathbb{R}^d$, $\mathbb{E}_\xi[\|\nabla f(x,\xi)\|^2] \leq C(1 + F(x))$,
iii) $\sum_n \gamma_n^2 < \infty$.

Theorem 3.4. *Let Assumptions 3.2.2, 3.2.4, 3.3.2, 3.3.3, 3.3.4-i) (with $q = 2$) and 3.3.6 hold. Then, the sequence (v_n, m_n, x_n) given by Algorithm 3.1 is bounded with probability one.*

Remark 7. The above stability result requires square summable stepsizes. Showing the same boundedness result under the Assumption 3.3.4 that allows for larger stepsizes is a challenging problem in the general case. In these situations, the boundedness of the iterates can be sometimes ensured by *ad hoc* means.

Remark 8. We can also consider the noisy discretization of (ODE-1′) introduced in Remark 4 above. This algorithm reads

$$\begin{cases} v_{n+1} &= (1-\gamma_{n+1}q_n)v_n + \gamma_{n+1}p_n\nabla f(x_n,\xi_{n+1})^{\odot 2} & \text{(3.6a)} \\ x_{n+1} &= x_n - \gamma_{n+1}\nabla f(x_n,\xi_{n+1})/\sqrt{v_{n+1}+\varepsilon} & \text{(3.6b)} \end{cases}$$

for $(v_0, x_0) \in \mathbb{R}^d_+ \times \mathbb{R}^d$. With only minor adaptations, Th. 3.3 and Th. 3.4 can be shown to hold as well for this algorithm. We refer to the concomitant paper (Gadat and Gavra, 2020, Sec. 2.2) for the link between this algorithm and the seminal algorithms AdaGrad (Duchi et al., 2011) and RMSProp (Tieleman and Hinton, 2012).

3.3.2 Stochastic Nesterov's Accelerated Gradient (S-NAG)

S-NAG is the noisy Euler's discretization of (ODE-N). Given $\alpha > 0$, it generates the sequence (m_n, x_n) on $\mathbb{R}^d \times \mathbb{R}^d$ given by Algorithm 3.2.

Algorithm 3.2 (S-NAG with decreasing steps)

Initialization: $m_0 = 0, x_0 \in \mathbb{R}^d$.
for $n = 1$ **to** n_{iter} **do**
  $m_{n+1} = (1 - \alpha\gamma_{n+1}/\tau_n)m_n + \gamma_{n+1}\nabla f(x_n, \xi_{n+1})$
  $x_{n+1} = x_n - \gamma_{n+1}m_{n+1}$.
end for

Assumption 3.3.7. Assume either of the following conditions.

i) There exists $q \geq 2$ s.t. for every compact set $\mathcal{K} \subset \mathbb{R}^d$,

$$\sup_{x\in\mathcal{K}} \mathbb{E}_\xi \|\nabla f(x,\xi)\|^q < \infty \quad \text{and} \quad \sum_n \gamma_n^{1+q/2} < \infty\,.$$

ii) For every compact set $\mathcal{K} \subset \mathbb{R}^d$, there exists a real $\sigma_{\mathcal{K}} \neq 0$ s.t.

$$\mathbb{E}_\xi \exp\langle u, \nabla f(x,\xi) - \nabla F(x)\rangle \mathbb{1}_{x\in\mathcal{K}} \leq \exp\left(\sigma^2_{\mathcal{K}}\|u\|^2/2\right)\,,$$

for every $x, u \in \mathbb{R}^d$. Moreover, for every $\alpha > 0$, $\sum_n \exp(-\alpha/\gamma_n) < \infty$.

Theorem 3.5. *Let Assumptions 3.2.2, 3.3.1, 3.3.2, 3.3.5 and 3.3.7 hold true. Assume that the random sequence $(y_n = (m_n, x_n) : n \in \mathbb{N})$ given by Algorithm 3.2 is bounded with probability one. Then, w.p.1, the sequence (y_n) converges towards the set $\bar{\Upsilon}$ defined in Eq. (3.3). If, in addition, the set of critical points of the objective function F is finite or countable, then w.p.1, the sequence (y_n) converges to a single point of $\bar{\Upsilon}$.*

The almost sure boundedness of the sequence (y_n) is handled in what follows.

Theorem 3.6. *Let Assumptions 3.2.2, 3.3.1, 3.3.2 and 3.3.6 hold. Then, the sequence $(y_n = (m_n, x_n) : n \in \mathbb{N})$ given by Algorithm 3.2 is bounded with probability one.*

Remark 9. Assumption 3.3.4-*i)* in Th. 3.4 is not needed for Th. 3.6.

3.3.3 Central limit theorem

In this section, we establish a conditional central limit theorem for Algorithm 3.1.

Assumption 3.3.8. Let $x_\star \in \operatorname{zer} \nabla F$. The following holds.

i) F is twice continuously differentiable on a neighborhood of $x_\star$ and the Hessian $\nabla^2 F(x_\star)$ is positive definite.
ii) S is continuously differentiable on a neighborhood of $x_\star$.
iii) There exists $M > 0$ and $b_M > 4$ s.t.

$$\sup_{x \in B(x_\star, M)} \mathbb{E}_\xi[\|\nabla f(x, \xi)\|^{b_M}] < \infty \,. \tag{3.7}$$

Under Assumptions 3.2.4-*i)* to *iii)*, it follows from Eq. (3.5) that the sequences (h_n), (r_n), (p_n) and (q_n) of nonnegative reals converge respectively to $h_\infty, r_\infty, p_\infty$ and q_∞ where h_∞, r_∞ and q_∞ are supposed positive. Define $v_\star := q_\infty^{-1} p_\infty S(x_\star)$. Consider the matrix

$$V := \operatorname{diag}\left((\varepsilon + v_\star)^{\odot - \frac{1}{2}} \right) . \tag{3.8}$$

Let P be an orthogonal matrix s.t. the following spectral decomposition holds:

$$V^{\frac{1}{2}} \nabla^2 F(x_\star) V^{\frac{1}{2}} = P \operatorname{diag}(\pi_1, \cdots, \pi_d) P^{-1} \,,$$

where $\pi_1 \leq \cdots \leq \pi_d$ are the (positive) eigenvalues of $V^{\frac{1}{2}} \nabla^2 F(x_\star) V^{\frac{1}{2}}$. Define

$$\mathcal{H} := \begin{bmatrix} -r_\infty I_d & h_\infty \nabla^2 F(x_\star) \\ -V & 0 \end{bmatrix}$$

where I_d is the $d \times d$ identity matrix. Then the matrix $\mathcal{H}$ is Hurwitz. Indeed, it can be shown that the largest real part of the eigenvalues of $\mathcal{H}$ coincides with $-L$, where

$$L := \frac{r_\infty}{2} \left(1 - \sqrt{\left(1 - \frac{4 h_\infty \pi_1}{r_\infty^2} \right) \vee 0} \right) > 0 \,. \tag{3.9}$$

Assumption 3.3.9. The sequence (γ_n) is given by $\gamma_n = \frac{\gamma_0}{n^\alpha}$ for some $\alpha \in (0, 1]$, $\gamma_0 > 0$. Moreover, if $\alpha = 1$, we assume that $\gamma_0 > \frac{1}{2(L \wedge q_\infty)}$.

Theorem 3.7. *Let Assumptions 3.2.4-i) to iii), 3.3.3, 3.3.8 and 3.3.9 hold. Consider the iterates $z_n = (v_n, m_n, x_n)$ given by Algorithm 3.1. Set $\theta := 0$ if $\alpha < 1$ and $\theta := 1/(2\gamma_0)$ if $\alpha = 1$. Assume that the event $\{z_n \to z_\star\}$, where $z_\star = (v_\star, 0, x_\star)$, has a positive probability. Then, given that event,*

$$\frac{1}{\sqrt{\gamma_n}} \begin{bmatrix} m_n \\ x_n - x_\star \end{bmatrix} \Rightarrow \mathcal{N}(0, \Gamma) \,,$$

where $\Rightarrow$ stands for the convergence in distribution and $\mathcal{N}(0, \Gamma)$ is a centered Gaussian distribution on $\mathbb{R}^{2d}$ with a covariance matrix Γ given by the unique solution to the Lyapunov equation

$$(\mathcal{H} + \theta I_{2d})\Gamma + \Gamma(\mathcal{H} + \theta I_{2d})^T = - \begin{bmatrix} \operatorname{Cov}(h_\infty \nabla f(x_\star, \xi)) & 0 \\ 0 & 0 \end{bmatrix} .$$

In particular, given $\{z_n \to z_\star\}$*, the vector* $\sqrt{\gamma_n}^{-1}(x_n - x_\star)$ *converges in distribution to a centered Gaussian distribution with a covariance matrix given by:*

$$\Gamma_2 = V^{\frac{1}{2}} P \left[\frac{C_{k,\ell}}{\frac{r_\infty - 2\theta}{h_\infty}(\pi_k + \pi_\ell + \frac{2\theta(\theta - r_\infty)}{h_\infty}) + \frac{(\pi_k - \pi_\ell)^2}{2(r_\infty - 2\theta)}} \right]_{k,\ell=1\dots d} P^{-1} V^{\frac{1}{2}} \tag{3.10}$$

where $C := P^{-1} V^{\frac{1}{2}} \mathbb{E}_\xi \left[\nabla f(x_\star, \xi) \nabla f(x_\star, \xi)^T \right] V^{\frac{1}{2}} P$.

A few remarks are in order.

- The matrix Γ_2 coincides with the limiting covariance matrix associated to the iterates

$$\begin{cases} m_{n+1} &= m_n + \gamma_{n+1}(h_\infty V \nabla f(x_n, \xi_{n+1}) - r_\infty m_n) \\ x_{n+1} &= x_n - \gamma_{n+1} m_{n+1} \,. \end{cases}$$

  This procedure can be seen as a preconditioned version of the stochastic heavy ball algorithm (Gadat et al., 2018) although the iterates are not implementable because of the unknown matrix V. Notice also that the limiting covariance Γ_2 depends on $v_\star$ but does not depend on the fluctuations of the sequence (v_n).

- When $h_\infty = r_\infty$ (which is the case for Adam), we recover the expression of the asymptotic covariance matrix previously provided in Chapter 2 Section 2.5.3 and the remarks formulated therein.

- The assumption $r_\infty > 0$ is crucial to establish Th. 3.7. For this reason, Th. 3.7 does not generalize immediately to Algorithm 3.2. The study of the fluctuations of Algorithm 3.2 is left for future works.

3.3.4 Related works

Gadat et al. (2018) study the SHB algorithm, which is a noisy Euler's discretization of (ODE-1) in the situation where $\mathsf{h} = \mathsf{r}$ and $\mathsf{p} = \mathsf{q} \equiv 0$ (*i.e.*, there is no v variable). In this framework, if we set $\mathsf{h} = \mathsf{r} \equiv r > 0$ in Algorithm 3.1 above, then Th. 3.3 above recovers the analogous case in (Gadat et al., 2018, Th. 2.1), which is termed as the exponential memory case. The other important case treated in Gadat et al. (2018) is the case where $\mathsf{h}(t) = \mathsf{r}(t) = r/t$ for some $r > 0$, referred to as the polynomially memory case. Actually, it is known that the ODE obtained for $\mathsf{h}(t) = \mathsf{r}(t) = r/t$ and $\mathsf{p} = \mathsf{q} \equiv 0$ boils down to (ODE-N) after a time variable change (see, *e.g.*, Lem. 3.14 below). Nevertheless, we highlight that the stochastic algorithm that stems from this ODE and that is studied in Gadat et al. (2018) is different from the S-NAG algorithm introduced above which stems from a different ODE (ODE-N). Hence, the convergence result of Th. 3.5 for the S-NAG algorithm we consider is not covered by the analysis of Gadat et al. (2018).

The specific case of the Adam algorithm is analyzed in Chapter 2 in both the constant and vanishing stepsize settings (see Ths. 2.6-2.7 which are the analogues of our Ths. 3.3-3.4). Note that we deal with a more general algorithm in the present chapter. Indeed, Algorithm 3.1 offers some freedom in the choice of the functions h, r, p, q satisfying Assumption 3.2.4 beyond the specific case of the Adam algorithm studied in Chapter 2. Apart from this generalization, we also emphasize some small improvements.

Regarding Th. 3.3, we provide noise conditions allowing to choose larger stepsizes (see Assumption 3.3.4 compared to Assumption 2.4.2 with $p = 2$ in Chapter 2). Concerning the stability result (Th.3.4), we relax Assumption 2.5.2-*iii)* of Chapter 2 which is no more needed in the present chapter (see Assumption 3.3.6) thanks to a modification of the discretized Lyapunov function used in the proof (see the proof in Section 3.6.4 compared to Section 2.9.2 in Chapter 2).

In most generality, the almost sure convergence result of the iterates of Algorithm 3.1 using vanishing stepsizes (Ths. 3.3-3.4) is new to the best of our knowledge. Moreover, while some recent results exist for S-NAG in the constant stepsize and for convex objective functions (see for e.g., Assran and Rabbat (2020)), Ths. 3.5 and 3.6 which tackle the possibly non-convex setting are also new to the best of our knowledge.

In the work of Gadat and Gavra (2020) that was posted on the arXiv repository a few days after our submission, Gadat and Gavra study the specific case of the algorithm described in Eq. (3.6) encompassing both ADAGRAD and RMSPROP, with the possibility to use mini-batches. For this specific algorithm, the authors establish a similar almost sure convergence result to ours (Gadat and Gavra, 2020, Th. 1) for decreasing stepsizes and derive some quantitative results bounding in expectation the gradient of the objective function along the iterations for constant stepsizes (Gadat and Gavra, 2020, Th. 2). We highlight though that they do not consider the presence of momentum in the algorithm. Therefore, their analysis does not cover neither Algorithm 3.1 nor Algorithm 3.2.

In contrast to our analysis, some works in the literature explore the constant stepsize regime for some stochastic momentum methods either for smooth (Yan et al., 2018) or weakly convex objective functions (Mai and Johansson, 2020). Furthermore, concerning ADAM-like algorithms, several recent works control the minimum of the norms of the gradients of the objective function evaluated at the iterates of the algorithm over N iterations in expectation or with high probability (Basu et al., 2018; Zhou et al., 2018; Chen et al., 2018; Zou et al., 2019a; Chen et al., 2019; Zaheer et al., 2018; Alacaoglu et al., 2020a; Défossez et al., 2020; Alacaoglu et al., 2020b) and establish regret bounds in the convex setting (Alacaoglu et al., 2020b).

Similar central limit theorems to Th. 3.7 are established in the cases of the stochastic heavy ball algorithm with exponential memory (Gadat et al., 2018, Th. 2.4) and ADAM (see Th. 2.8 in Chapter 2). In comparison to Gadat et al. (2018), we precise that our theorem recovers their result and provides a closed formula for the asymptotic covariance matrix Γ_2. Our proof of Th. 3.7 differs from the strategies adopted in Gadat et al. (2018) and Chapter 2.

3.4 Avoidance of traps

In Th. 3.3 and Th. 3.5 above, we established the almost sure convergence of the iterates x_n towards the set of critical points of the objective function F for both Algorithms 3.1 and 3.2. However, the landscape of F can contain what is known as "traps" for the algorithm, namely, critical points where the Hessian matrix of F has negative eigenvalues, making these critical points local maxima or saddle points. In this section, we show that the convergence of the iterates to these traps does not take place if the noise is exciting in some directions.

Starting with the contributions of Pemantle (1990) and Brandière and Duflo (1996), the numerous so-called avoidance of traps results that can be found in the literature deal with the case where the ODE that underlies the stochastic algorithm is an autonomous ODE. Obviously, this is neither the case of (ODE-1), nor of (ODE-N). To deal with this issue, we first state a general avoidance of traps result that extends Pemantle (1990); Brandière and Duflo (1996) to a non-autonomous setting, and that has an interest of its own. We then apply this result to Algorithms 3.1 and 3.2.

3.4.1 A general avoidance-of-traps result in a non-autonomous setting

The notations in this subsection and in Sections 3.7.1–3.7.2 are independent from the rest of the chapter. We recall that for a function $h : \mathbb{R}^d \to \mathbb{R}^{d'}$, we denote by $\partial_i^k h(x_1, \ldots, x_d)$ the k^{th} partial derivative of the function h with respect to x_i.

The setting of our problem is as follows. Given an integer $d > 0$ and a continuous function $b : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$, we consider a stochastic algorithm built around the non-autonomous ODE $\dot{\mathsf{z}}(t) = b(\mathsf{z}(t), t)$. Let $z_\star \in \mathbb{R}^d$, and assume that on $\mathcal{V} \times \mathbb{R}_+$ where $\mathcal{V}$ is a certain neighborhood of $z_\star$, the function b can be developed as

$$b(z, t) = D(z - z_\star) + e(z, t), \tag{3.11}$$

where $e(z_\star, \cdot) \equiv 0$, and where the matrix $D \in \mathbb{R}^{d \times d}$ is assumed to admit the following spectral factorization: Given $0 \leq d^- < d$ and $0 < d^+ \leq d$ with $d^- + d^+ = d$, we can write

$$D = Q \Lambda Q^{-1}, \quad \Lambda = \begin{bmatrix} \Lambda^- & \\ & \Lambda^+ \end{bmatrix}, \tag{3.12}$$

where the Jordan blocks that constitute $\Lambda^- \in \mathbb{R}^{d^- \times d^-}$ (respectively $\Lambda^+ \in \mathbb{R}^{d^+ \times d^+}$) are those that contain the eigenvalues λ_i of D for which $\Re\lambda_i \leq 0$ (respectively $\Re\lambda_i > 0$). Since $d^+ > 0$, the point $z_\star$ is an unstable equilibrium point of the ODE $\dot{\mathsf{z}}(t) = b(\mathsf{z}(t), t)$, in the sense that the ODE solution will only be able to converge to $z_\star$ along a specific so-called invariant manifold which precise characterization will be given in Section 3.7.1 below.

We now consider a stochastic algorithm that is built around this ODE. The condition $d^+ > 0$ makes that $z_\star$ is a trap that the algorithm should desirably avoid. The following theorem states that this will be the case if the noise term of the algorithm is omnidirectional enough. The idea is to show that the case being, the algorithm trajectories will move away from the invariant manifold mentioned above.

Theorem 3.8. *Given a sequence (γ_n) of nonnegative deterministic stepsizes such that $\sum_n \gamma_n = +\infty$, $\sum_n \gamma_n^2 < +\infty$, and a filtration $(\mathcal{F}_n)$, consider the stochastic approximation algorithm in $\mathbb{R}^d$*

$$z_{n+1} = z_n + \gamma_{n+1} b(z_n, \tau_n) + \gamma_{n+1}\eta_{n+1} + \gamma_{n+1}\rho_{n+1}$$

where $\tau_n = \sum_{k=1}^n \gamma_k$. Assume that the sequences (η_n) and (ρ_n) are adapted to $\mathcal{F}_n$, and that z_0 is $\mathcal{F}_0$–measurable. Assume that there exists $z_\star \in \mathbb{R}^d$ such that Eq. (3.11) holds true on $\mathcal{V} \times \mathbb{R}_+$, where $\mathcal{V}$ is a neighborhood of $z_\star$. Consider the spectral factorization (3.12), and assume that $d^+ > 0$. Assume moreover that the function e at the right hand side of Eq. (3.11) satisfies the conditions:

i) $e(z_\star, \cdot) \equiv 0$.
ii) *On* $\mathcal{V} \times \mathbb{R}_+$, *the functions* $\partial_2^n \partial_1^k e(z,t)$ *exist and are continuous for* $0 \leq n < 2$ *and* $0 \leq k + n \leq 2$.
iii) *The following convergence holds :*

$$\lim_{(z,t)\to(z_\star,\infty)} \|\partial_1 e(z,t)\| = 0\,. \tag{3.13}$$

iv) *There exist* $t_0 > 0$ *and a neighborhood* $\mathcal{W} \subset \mathbb{R}^d$ *of* $z_\star$ *s.t.*

$$\sup_{z\in\mathcal{W}, t\geq t_0} \left\|\partial_2 e(z,t)\right\| < +\infty \quad \textit{and} \quad \sup_{z\in\mathcal{W}, t\geq t_0} \left\|\partial_1^2 e(z,t)\right\| < +\infty\,.$$

Moreover, suppose that :

v) $\sum_n \|\rho_{n+1}\|^2 \mathbb{1}_{z_n\in\mathcal{W}} < \infty$ *almost surely.*
vi) $\limsup \mathbb{E}[\|\eta_{n+1}\|^4 \,|\, \mathcal{F}_n]\mathbb{1}_{z_n\in\mathcal{W}} < \infty$, *and* $\mathbb{E}[\eta_{n+1} \,|\, \mathcal{F}_n]\mathbb{1}_{z_n\in\mathcal{W}} = 0$.
vii) *Writing* $\tilde{\eta}_n = Q^{-1}\eta_n = (\tilde{\eta}_n^-, \tilde{\eta}_n^+)$ *with* $\tilde{\eta}_n^\pm \in \mathbb{R}^{d^\pm}$, *for some* $c^2 > 0$, *it holds that*

$$\liminf \mathbb{E}[\|\tilde{\eta}_{n+1}^+\|^2 \,|\, \mathcal{F}_n]\mathbb{1}_{z_n\in\mathcal{W}} \geq c^2 \mathbb{1}_{z_n\in\mathcal{W}}\,.$$

Then, $\mathbb{P}([z_n \to z_\star]) = 0$.

Remark 10. Assumptions *i)* to *iv)* of Th. 3.8 are related to the function e defined in Eq. (3.11), which can be seen as a non-autonomous perturbation of the autonomous linear ODE $\dot{\mathsf{z}}(t) = D(\mathsf{z}(t) - z_\star)$. These assumptions guarantee the existence of a local (around the unstable equilibrium $z_\star$) non-autonomous invariant manifold of the non-autonomous ODE $\dot{\mathsf{z}}(t) = b(\mathsf{z}(t), t)$ with enough regularity properties, as provided by Prop. 3.18 and Prop. 3.20 below.

3.4.2 Application to the stochastic algorithms

3.4.2.1 Trap avoidance of the general algorithm 3.1

In Th. 3.3 above, we showed that the sequence (z_n) generated by Algorithm 3.1 converges almost surely towards the set Υ defined in Eq. (3.2). Our purpose now is to show that the traps in Υ (to be characterized below) are avoided by the stochastic algorithm 3.1 under a proper omnidirectionality assumption on the noise.

Our first task is to write Algorithm 3.1 in a manner compatible with the statement of Th. 3.8. The following decomposition holds for the sequence $(z_n = (v_n, m_n, x_n), n \in \mathbb{N})$ generated by this algorithm:

$$z_{n+1} = z_n + \gamma_{n+1} g(z_n, \tau_n) + \gamma_{n+1}\eta_{n+1} + \gamma_{n+1}\tilde{\rho}_{n+1},$$

where $\tilde{\rho}_{n+1} = \left(0\,, 0\,, \frac{m_n}{\sqrt{v_n+\varepsilon}} - \frac{m_{n+1}}{\sqrt{v_{n+1}+\varepsilon}}\right)$, and where η_{n+1} is the martingale increment with respect to the filtration $(\mathcal{F}_n)$ which is defined by Eq. (3.28).

Observe from Eq. (3.2) that each $z_\star \in \Upsilon$ is written as $z_\star = (v_\star, 0, x_\star)$ where $x_\star \in \operatorname{zer} \nabla F$, and $v_\star = q_\infty^{-1} p_\infty S(x_\star)$ (in particular, $x_\star$ and $z_\star$ are in a one-to-one correspondence). We need to linearize the function $g(\cdot, t)$ around $z_\star$. The following assumptions will be required.

Assumption 3.4.1. The functions F and S belong respectively to $\mathcal{C}^3(\mathbb{R}^d, \mathbb{R})$ and $\mathcal{C}^2(\mathbb{R}^d, \mathbb{R}_+^d)$.

Assumption 3.4.2. The functions $\mathsf{h}, \mathsf{r}, \mathsf{p}, \mathsf{q}$ belong to $\mathcal{C}^1((0,\infty),\mathbb{R}_+)$ and have bounded derivatives on $[t_0,+\infty)$ for some $t_0>0$.

Lemma 3.9. Let Assumptions 3.2.4-*i)* to *iii)*, 3.4.1 and 3.4.2 hold. Let $z_\star=(v_\star,0,x_\star)\in\Upsilon$. Then, for every $z\in\mathcal{Z}_+$ and every $t>0$, the following decomposition holds true:

$$g(z,t)=D(z-z_\star)+e(z,t)+c(t),$$

$$\text{where } D=\begin{bmatrix}-q_\infty I_d & 0 & p_\infty\nabla S(x_\star)\\ 0 & -r_\infty I_d & h_\infty\nabla^2F(x_\star)\\ 0 & -V & 0\end{bmatrix},\quad c(t)=\begin{bmatrix}\mathsf{p}(t)S(x_\star)-\mathsf{q}(t)v_\star\\ 0\\ 0\end{bmatrix},$$

and the function $e(z,t)$ (defined in Section 3.7.3.1 below for conciseness) has the same properties as its analogue in the statement of Th. 3.8.

Using this lemma, the algorithm iterate z_{n+1} can be rewritten as an instance of the algorithm in the statement of Th. 3.8, namely,

$$z_{n+1}=z_n+\gamma_{n+1}b(z_n,\tau_n)+\gamma_{n+1}\eta_{n+1}+\gamma_{n+1}\rho_{n+1},\tag{3.14}$$

where in our present setting, $b(z,t)=g(z,t)-c(t)=D(z-z_\star)+e(z,t)$ and $\rho_n=c(\tau_{n-1})+\tilde\rho_n$. In the following assumption, we use the well-known fact that a symmetric matrix H has the same inertia as AHA^{T} for an arbitrary invertible matrix A.

Assumption 3.4.3. Let $x_\star\in\operatorname{zer}\nabla F$, let $v_\star=q_\infty^{-1}p_\infty S(x_\star)$, and define the diagonal matrix $V=\operatorname{diag}((v_\star+\varepsilon)^{\odot-\frac12})$ as in (3.8). Assume the following conditions:

i) $\sum_n\left(q_\infty p_n-p_\infty q_n\right)^2<\infty$,
ii) The Hessian matrix $\nabla^2F(x_\star)$ has a negative eigenvalue.
iii) There exists $\delta>0$ such that $\sup_{x\in B(x_\star,\delta)}\mathbb{E}_\xi[\|\nabla f(x,\xi)\|^8]<\infty$.
iv) Defining Π_{u} as the orthogonal projector on the eigenspace of $V^{\frac12}\nabla^2F(x_\star)V^{\frac12}$ that is associated with the negative eigenvalues of this matrix, it holds that

$$\Pi_{\mathrm{u}}V^{\frac12}\mathbb{E}_\xi(\nabla f(x_\star,\xi)-\nabla F(x_\star))(\nabla f(x_\star,\xi)-\nabla F(x_\star))^{\mathrm{T}}V^{\frac12}\Pi_{\mathrm{u}}\neq 0.$$

Theorem 3.10. *Let Assumptions 3.2.4, 3.3.3, and 3.4.1, 3.4.2 hold true. Let $z_\star\in\Upsilon$ be such that Assumption 3.4.3 holds true for this $z_\star$. Then, the eigenspace associated with the eigenvalues of D with positive real parts has the same dimension as the eigenspace of $\nabla^2F(x_\star)$ associated with the negative eigenvalues of this matrix. Let $(z_n=(v_n,m_n,x_n):n\in\mathbb{N})$ be the random sequence generated by Algorithm 3.1 with stepsizes satisfying $\sum_n\gamma_n=+\infty$ and $\sum_n\gamma_n^2<+\infty$. Then, $\mathbb{P}([z_n\to z_\star])=0$.*

The assumptions and the result call for some comments.

Remark 11. The definition of a trap as regards the general algorithm in the statement of Th. 3.8 is that the matrix D in Eq. (3.11) has eigenvalues with positive real parts. Th. 3.10 states that this condition is equivalent to $\nabla^2F(x_\star)$ having negative eigenvalues. What's more, the dimension of the invariant subspace of D corresponding to the eigenvalues with positive real parts is equal to the dimension of the negative eigenvalue subspace of $\nabla^2F(x_\star)$. Thus, Assumption 3.4.3–*iv)* provides the "largest" subspace where the noise energy must be non zero for the purpose of avoiding the trap.

Remark 12. Assumptions 3.4.2 and 3.4.3-*i)* are satisfied by many widely studied algorithms, among which RMSProp and Adam.

Remark 13. The results of Th. 3.10 can be straightforwardly adapted to the case of (ODE-1′). Assumption 3.4.3-*iv)* on the noise is unchanged.

In the case of the S-NAG algorithm, the assumptions become particularly simple. We state the afferent result separately.

3.4.2.2 Trap avoidance for S-NAG

Assumption 3.4.4. Let $x_\star \in \operatorname{zer} \nabla F$ and let the following conditions hold.

i) The Hessian matrix $\nabla^2 F(x_\star)$ has a negative eigenvalue.
ii) There exists $\delta > 0$ such that $\sup_{x \in B(x_\star, \delta)} \mathbb{E}_\xi[\|\nabla f(x, \xi)\|^4] < \infty$.
iii) $\tilde{\Pi}_{\mathrm{u}} \mathbb{E}_\xi(\nabla f(x_\star, \xi) - \nabla F(x_\star))(\nabla f(x_\star, \xi) - \nabla F(x_\star))^{\mathrm{T}} \tilde{\Pi}_{\mathrm{u}} \neq 0$, where $\tilde{\Pi}_{\mathrm{u}}$ is the orthogonal projector on the eigenspace of $\nabla^2 F(x_\star)$ associated with its negative eigenvalues.

Theorem 3.11. *Let Assumptions 3.2.4, 3.3.1, 3.4.1 and 3.4.4 hold. Define* $y_\star = (0, x_\star)$. *Let* $(y_n = (m_n, x_n) : n \in \mathbb{N})$ *be the random sequence given by Algorithm 3.2 with stepsizes satisfying* $\sum_n \gamma_n = +\infty$ *and* $\sum_n \gamma_n^2 < +\infty$. *Then,* $\mathbb{P}([y_n \to y_\star]) = 0$.

3.4.3 Related works

Up to our knowledge, all the avoidance of traps results that can be found in the literature, starting from Pemantle (1990); Brandière and Duflo (1996), refer to stochastic algorithms that are discretizations of autonomous ODE's (see for e.g., (Benaïm, 1999, Sec. 9) for general Robbins Monro algorithms and (Mertikopoulos et al., 2020, Sec. 4.3) for SGD). In this line of research, a powerful class of techniques relies on Poincaré's invariant manifold theorem for an autonomous ODE in a neighborhood of some unstable equilibrium point. In our work, we extend the avoidance of traps results to a non-autonomous setting, by borrowing a non-autonomous version of Poincaré's theorem from the rich literature that exists on the subject (Daleckii and Krein, 1974; Kloeden and Rasmussen, 2011).

In Gadat et al. (2018), the authors succeeded in establishing an avoidance of traps result for their non-autonomous stochastic algorithm which is close to our S-NAG algorithm (see the discussion at the end of Section 3.3.4 above), at the expense of a sub-Gaussian assumption on the noise and a rather stringent assumption on the stepsizes. The main difficulty in the approach of Gadat et al. (2018) lies in the use of the classical autonomous version of Poincaré's theorem (see (Gadat et al., 2018, Remark 2.1)). This kind of difficulty is avoided by our approach, which allows to obtain avoidance of traps results with close to minimal assumptions. More recently, in the contribution of Gadat and Gavra (2020) discussed in Sec. 3.3.4, the authors establish an avoidance of traps result ((Gadat and Gavra, 2020, Th. 3)) for the algorithm described in Eq. (3.6) using techniques inspired from Pemantle (1990); Benaïm (1999). As previously mentioned, this recent work does not handle momentum and hence neither Algorithm 3.1 nor Algorithm 3.2. Moreover, as indicated in our discussion of Gadat et al. (2018), our strategy of proof is different.

Taking another point of view as concerns the trap avoidance, some recent works (Lee et al., 2019; Du et al., 2017; Jin et al., 2017; Panageas and Piliouras, 2017; Panageas et al., 2019) address the problem of escaping saddle points when the algorithm is deterministic

but when the initialization point is random. In contrast to this line of research, our work considers a stochastic algorithm for which randomness enters into play at each iteration of the algorithm via noisy gradients.

3.5 Proofs for Section 3.2

3.5.1 Proof of Th. 3.1

The arguments of the proof of this theorem that we provide here follow the approach of Belotto da Silva and Gazeau (2020) with some small differences. Close arguments can be found in Chapter 2. We provide the proof here for completeness and in preparation of the proofs that will be related with the stochastic algorithms.

3.5.1.1 Existence and uniqueness

The following lemma guarantees that the term $\sqrt{\mathsf{v}(t)+\varepsilon}$ in (ODE-1) is well-defined.

Lemma 3.12. Let $t_0 \in \mathbb{R}_+$ and $T \in (t_0, \infty]$. Assume that there exists a solution $\mathsf{z}(t) = (\mathsf{v}(t), \mathsf{m}(t), \mathsf{x}(t))$ to (ODE-1) on $[t_0, T)$ for which $\mathsf{v}(t_0) \geq 0$. Then, for all $t \in [t_0, T)$, $\mathsf{v}(t) \geq 0$.

Proof. Assume that $\nu := \inf\{t \in [t_0, T),\, \mathsf{v}(t) < 0\}$ satisfies $\nu < T$. If $\mathsf{v}(t_0) > 0$, Gronwall's lemma implies that $\mathsf{v}(t) \geq \mathsf{v}(t_0)\exp(-\int_{t_0}^t q(t))$ on $[t_0, \nu]$ which is in contradiction with the fact that $\mathsf{v}(\nu) = 0$. If $\mathsf{v}(t_0) = 0$, since $\nu < T$, there exists $t_1 \in (t_0, \nu)$ s.t. $\dot{\mathsf{v}}(t_1) < 0$. Hence, using the first equation from (ODE-1), we obtain $\mathsf{v}(t_1) > 0$. This brings us back to the first case, replacing t_0 by t_1. ∎

Recall that $F_\star = \inf F$ is finite by Assumption 3.2.2. Of prime importance in the proof will be the energy (Lyapunov) function $\mathcal{E} : \mathbb{R}_+ \times \mathcal{Z}_+ \to \mathbb{R}$, defined as

$$\mathcal{E}(h, z) = h(F(x) - F_\star) + \frac{1}{2}\left\| \frac{m}{(v+\varepsilon)^{\odot\frac{1}{4}}} \right\|^2, \tag{3.15}$$

for every $h \geq 0$ and every $z = (v, m, x) \in \mathcal{Z}_+$. This function is slightly different from its analogues that were used in Chapter 2 or in Alvarez (2000); Belotto da Silva and Gazeau (2020).

Consider $(t, z) \in (0, +\infty) \times \mathcal{Z}_+$ and set $z = (v, m, x)$. Then, using Assumption 3.2.1, we can write

$$\begin{aligned}
&\partial_t \mathcal{E}(\mathsf{h}(t), z) + \langle \nabla_z \mathcal{E}(\mathsf{h}(t), z), g(z, t)\rangle \\
&= \dot{\mathsf{h}}(t)(F(x) - F_\star) - \frac{1}{4}\langle \frac{m^{\odot 2}}{(v+\varepsilon)^{\odot\frac{3}{2}}}, \mathsf{p}(t)S(x) - \mathsf{q}(t)v\rangle \\
&\quad + \langle \frac{m}{(v+\varepsilon)^{\odot\frac{1}{2}}}, \mathsf{h}(t)\nabla F(x) - \mathsf{r}(t)m\rangle - \langle \frac{m}{(v+\varepsilon)^{\odot\frac{1}{2}}}, \mathsf{h}(t)\nabla F(x)\rangle \\
&\leq -\left(\mathsf{r}(t) - \frac{\mathsf{q}(t)}{4}\right)\left\| \frac{m}{(v+\varepsilon)^{\odot\frac{1}{4}}} \right\|^2 + \dot{\mathsf{h}}(t)(F(x) - F_\star) - \frac{\mathsf{p}(t)}{4}\langle S(x), \frac{m^{\odot 2}}{(v+\varepsilon)^{\odot\frac{3}{2}}}\rangle.
\end{aligned} \tag{3.16}$$

With the help of this function, we can now establish the existence, the uniqueness and the boundedness of the solution of (ODE-1) on $[t_0, \infty)$ for an arbitrary $t_0 > 0$.

Lemma 3.13. For each $t_0 > 0$ and $z_0 \in \mathcal{Z}_+$, (ODE-1) has a unique solution on $[t_0, \infty)$ starting at $\mathsf{z}(t_0) = z_0$. Moreover, the orbit $\{\mathsf{z}(t) : t \geq t_0\}$ is bounded.

Proof. Let $t_0 > 0$, and fix $z_0 \in \mathcal{Z}_+$. On each set of the type $[t_0, t_0 + A] \times \bar{B}(z_0, R)$ where $A, R > 0$ and $\bar{B}(z_0, R) \subset (-\varepsilon, \infty)^d \times \mathbb{R}^d \times \mathbb{R}^d$, we easily obtain from our assumptions that the function g defined in (3.1) is continuous, and that $g(\cdot, t)$ is uniformly Lipschitz on $t \in [t_0, t_0 + A]$. In these conditions, Picard's theorem asserts that (ODE-1) starting from $\mathsf{z}(t_0) = z_0$ has a unique solution on a certain maximal interval $[t_0, T)$. Lem. 3.12 shows that $\mathsf{v}(t) \geq 0$ on this interval.

Let us show that $T = \infty$. Applying Ineq. (3.16) with $(v, m, x) = (\mathsf{v}(t), \mathsf{m}(t), \mathsf{x}(t))$ and using Assumption 3.2.4, we obtain that the function $t \mapsto \mathcal{E}(\mathsf{h}(t), \mathsf{z}(t))$ is decreasing on $[t_0, T)$. By the coercivity of F (Assumption 3.2.2) and Assumption 3.2.4–*i)*, we get that the trajectory $\{\mathsf{x}(t)\}$ is bounded. Recall the equation $\dot{\mathsf{m}}(t) = \mathsf{h}(t)\nabla F(\mathsf{x}(t)) - \mathsf{r}(t)\mathsf{m}(t)$. Using the continuity of the functions ∇F, h and r along with Gronwall's lemma, we get that $\{\mathsf{m}(t)\}$ is bounded if $T < \infty$. We can show a similar result for $\{\mathsf{v}(t)\}$. Thus, $\{\mathsf{z}(t)\}$ is bounded on $[t_0, T)$ if $T < \infty$ which is a contradiction, see, *e.g.*, (Hartman, 2002, Cor.3.2).

It remains to show that the trajectory $\{\mathsf{z}(t)\}$ is bounded. To that end, let us apply the variation of constants method to the equation $\dot{\mathsf{m}}(t) = \mathsf{h}(t)\nabla F(\mathsf{x}(t)) - \mathsf{r}(t)\mathsf{m}(t)$. Writing $R(t) = \int_{t_0}^t \mathsf{r}(u)\, du$, we get that

$$\frac{d}{dt}\left(e^{R(t)}\mathsf{m}(t)\right) = e^{R(t)}\mathsf{h}(t)\nabla F(\mathsf{x}(t)).$$

Therefore, for every $t \geq t_0$,

$$\mathsf{m}(t) = e^{-R(t)}\mathsf{m}(t_0) + \int_{t_0}^t e^{R(u)-R(t)}\mathsf{h}(u)\nabla F(\mathsf{x}(u))du\,.$$

Using the continuity of ∇F together with the boundedness of x, Assumption 3.2.4 and the triangle inequality, we obtain the existence of a constant $C > 0$ independent of t s.t.

$$\begin{aligned}\left\|\mathsf{m}(t) - \mathsf{m}(t_0)\right\| - \left\|\mathsf{m}(t_0)\right\| &\leq C\mathsf{h}(t_0)\int_{t_0}^t e^{-\int_u^t \mathsf{r}(s)\,ds}du \\ &\leq C\mathsf{h}(t_0)\int_{t_0}^t e^{-r_\infty(t-u)}du \leq \frac{C\mathsf{h}(t_0)}{r_\infty}\,.\end{aligned}$$

The same reasoning applies to $\mathsf{v}(t)$ using the continuity of S and Assumption 3.2.4. This completes the proof. ∎

We can now extend this solution to $t_0 = 0$ along the approach of Belotto da Silva and Gazeau (2020), where the detailed derivations can be found. The idea is to replace $\mathsf{h}(t)$ with $\mathsf{h}(\max(\eta, t))$ for some $\eta > 0$ and to do the same for p, q, and r. It is then easy to see that the ODE that is obtained by doing these replacements has a unique global solution on $\mathbb{R}_+$. By making $\eta \to 0$ and by using the Arzelà-Ascoli theorem along with Assumption 3.2.5, we obtain that (ODE-1) has a unique solution on $\mathbb{R}_+$.

3.5.1.2 Convergence

The first step in this part consists in transforming (ODE-1) into an autonomous ODE by including the time variable into the state vector. More specifically, we start with the following ODE:

$$\begin{bmatrix} \dot{\mathsf{z}}(t) \\ \dot{u}(t) \end{bmatrix} = \begin{bmatrix} g(\mathsf{z}(t), u(t)) \\ 1 \end{bmatrix} \quad \text{with} \quad \begin{bmatrix} \mathsf{z}(0) \\ u(0) \end{bmatrix} = \begin{bmatrix} z_0 \\ t_0 \end{bmatrix},$$

then, we perform the following change of variable in time

$$\begin{bmatrix} z \\ u \end{bmatrix} \mapsto \begin{bmatrix} z \\ s = 1/u \end{bmatrix}$$

allowing the solution to lie in a compact set.

We initialize the above ODE at a time instant $t_0 > 0$. Define the functions $\mathsf{H}, \mathsf{R}, \mathsf{P}, \mathsf{Q} : \mathbb{R}_+ \to \mathbb{R}_+$ by setting $\mathsf{H}(s) = \mathsf{h}(1/s)$, $\mathsf{R}(s) = \mathsf{r}(1/s)$, $\mathsf{P}(s) = \mathsf{p}(1/s)$; $\mathsf{Q}(s) = \mathsf{q}(1/s)$ for $s > 0$; $\mathsf{H}(0) = h_\infty$, $\mathsf{R}(0) = r_\infty$, $\mathsf{P}(0) = p_\infty$ and $\mathsf{Q}(0) = q_\infty$. Our autonomous dynamical system can then be described by the following system of equations:

$$\begin{cases} \dot{\mathsf{v}}(t) &= \mathsf{P}(\mathsf{s}(t))S(\mathsf{x}(t)) - \mathsf{Q}(\mathsf{s}(t))\mathsf{v}(t) \\ \dot{\mathsf{m}}(t) &= \mathsf{H}(\mathsf{s}(t))\nabla F(\mathsf{x}(t)) - \mathsf{R}(\mathsf{s}(t))\mathsf{m}(t) \\ \dot{\mathsf{x}}(t) &= -\frac{\mathsf{m}(t)}{\sqrt{\mathsf{v}(t)+\varepsilon}} \\ \dot{\mathsf{s}}(t) &= -\mathsf{s}(t)^2 \end{cases} \tag{3.17}$$

Since the solution of the ODE $\dot{\mathsf{s}}(t) = -\mathsf{s}(t)^2$ for which $\mathsf{s}(t_0) = 1/t_0$ is $\mathsf{s}(t) = 1/t$, the trajectory $\{\mathsf{s}(t)\}$ is bounded. The three remaining equations are a reformulation of (ODE-1) for which the trajectories have already been shown to exist and to be bounded in Lem. 3.13. In the sequel, we denote by $\Phi : \mathcal{Z}_+ \times \mathbb{R}_+ \to \mathcal{Z}_+ \times \mathbb{R}_+$ the semiflow induced by the autonomous ODE (3.17), *i.e.*, for every $u = (z, s) \in \mathcal{Z}_+ \times \mathbb{R}_+$, $\Phi(u, \cdot)$ is the unique global solution to the autonomous ODE (3.17) initialized at u. Observe that the orbits of this semiflow are precompact. Moreover, the function $\Phi((z, 0), \cdot)$ is perfectly defined for each $z \in \mathcal{Z}_+$ since the associated solution satisfies the ODE (3.19) defined below, which three first equations satisfy the hypotheses of Lem. 3.13.

Consider now a continuous function $V : \mathcal{Z}_+ \times \mathbb{R}_+ \to \mathbb{R}$ defined by:

$$V(u) = \mathcal{E}\left(\mathsf{H}(s), z\right), \quad u = (z, s) \in \mathcal{Z}_+ \times (0, \infty).$$

As for Ineq. (3.16) above, we have here that

$$\frac{d}{dt}V\left(\Phi(u,t)\right) \leq -\left(\mathsf{r}(t) - \frac{\mathsf{q}(t)}{4}\right)\left\|\frac{\mathsf{m}(t)}{(\mathsf{v}(t)+\varepsilon)^{\odot\frac{1}{4}}}\right\|^2 + \dot{\mathsf{h}}(t)(F(\mathsf{x}(t)) - F_\star) - \frac{\mathsf{p}(t)}{4}\langle S(\mathsf{x}(t)), \frac{\mathsf{m}(t)^{\odot 2}}{(\mathsf{v}(t)+\varepsilon)^{\odot\frac{3}{2}}}\rangle$$

if $s > 0$, and the same inequality with $(\dot{\mathsf{h}}(t), \mathsf{p}(t), \mathsf{r}(t), \mathsf{q}(t))$ being replaced with $(0, p_\infty, r_\infty, q_\infty)$ otherwise.

Since $V \circ \Phi(u, \cdot)$ is non-increasing and nonnegative, we can define $V_\infty := \lim_{t\to\infty} V(\Phi(u,t))$. Let $\omega(u) := \bigcap_{s>0} \overline{\bigcup_{t\geq s} \Phi(u,t)}$ be the ω-limit set of the semiflow Φ issued from u. Recall that $\omega(u)$ is an invariant set for the flow $\Phi(u,\cdot)$, and that

$$\operatorname{dist}(\Phi(u,t), \omega(u)) \xrightarrow[t\to\infty]{} 0,$$

see, *e.g.*, (Haraux, 1991, Th. 1.1.8)). In order to finish the proof of Th. 3.1, we need to make explicit the structure of $\omega(u)$.

We know from La Salle's invariance principle that $\omega(u) \subset V^{-1}(V_\infty)$. In particular,

$$\forall y \in \omega(u),\ \forall t \geq 0,\ V(\Phi(y,t)) = V(y) = V_\infty \tag{3.18}$$

by the invariance of $\omega(u)$.

From ODE (3.17), we have that any $y \in \omega(u)$ is of the form $y = (z, 0)$ since $s(t) \to 0$. As a consequence, $\Phi(y,\cdot)$ is a solution to the autonomous ODE

$$\begin{cases} \dot{\mathsf{v}}(t) &= p_\infty S(\mathsf{x}(t)) - q_\infty \mathsf{v}(t) \\ \dot{\mathsf{m}}(t) &= h_\infty \nabla F(\mathsf{x}(t)) - r_\infty \mathsf{m}(t) \\ \dot{\mathsf{x}}(t) &= -\frac{\mathsf{m}(t)}{\sqrt{\mathsf{v}(t)+\varepsilon}} \\ \dot{\mathsf{s}}(t) &= 0\,. \end{cases} \tag{3.19}$$

The three first equations can be written in a more compact form :

$$\dot{\mathsf{z}}(t) = g_\infty(\mathsf{z}(t)) \tag{3.20}$$

where $\mathsf{z}(t) = (\mathsf{v}(t), \mathsf{m}(t), \mathsf{x}(t))$, and

$$g_\infty(z) = \lim_{t\to\infty} g(z,t) = \begin{bmatrix} p_\infty S(x) - q_\infty v \\ h_\infty \nabla F(x) - r_\infty m \\ -m/\sqrt{v+\varepsilon} \end{bmatrix}$$

for each $z \in \mathcal{Z}_+$. Consider $y = (v, m, x, 0) \in \omega(u)$. Using Eq. (3.18), we obtain that $dV(\Phi(y,t))/dt = 0$, which implies that

$$\left(r_\infty - \frac{q_\infty}{4}\right) \left\| \frac{\mathsf{m}(t)}{(\mathsf{v}(t)+\varepsilon)^{\odot\frac{1}{4}}} \right\|^2 + \frac{p_\infty}{4} \langle S(\mathsf{x}(t)), \frac{\mathsf{m}(t)^{\odot 2}}{(\mathsf{v}(t)+\varepsilon)^{\odot\frac{3}{2}}} \rangle = 0$$

for all $(\mathsf{v}(t), \mathsf{m}(t), \mathsf{x}(t), 0) = \Phi(y,t)$. As a consequence, Assumption 3.2.4-*iv)* gives $\mathsf{m}(t) = m = 0$, and then, $\mathsf{x}(t) = x$ for some x s.t. $\nabla F(x) = 0$ using ODE (3.19). We now turn to showing that $\mathsf{v}(t) = v = p_\infty S(x)/q_\infty$. We have proved so far that any element $y \in \omega(u)$ is written $y = (v, 0, x, 0)$ where $\nabla F(x) = 0$. The component $\mathsf{v}(\cdot)$ of $\Phi(y,\cdot)$ is a solution to the ODE $\dot{\mathsf{v}}(t) = p_\infty S(x) - q_\infty \mathsf{v}(t)$ and is thus written

$$\mathsf{v}(t) = \frac{p_\infty S(x)}{q_\infty} + e^{-q_\infty t}\left(v - \frac{p_\infty S(x)}{q_\infty}\right). \tag{3.21}$$

Fixing x, let $\mathcal{S}_x$ be the section of $\omega(u)$ defined by:

$$\mathcal{S}_x\omega(u) = \left\{ y \in \omega(u) \,:\, y = (\tilde{v}, 0, x, 0)\,,\ \tilde{v} \in \mathbb{R}^d_+ \right\}.$$

As $\omega(u)$ is invariant, we have $\mathcal{S}_x\omega(u) = \mathcal{S}_x\Phi(\omega(u),t)$ for all $t \geq 0$. Since the set $\{\tilde{v} \in \mathbb{R}^d_+ \text{ s.t. } (\tilde{v},0,x,0) \in \mathcal{S}_x\omega(u)\}$ lies in a compact, we deduce from Eq. (3.21) that this set is reduced to the singleton $\{p_\infty S(x)/q_\infty\}$ and in particular $v = p_\infty S(x)/q_\infty$. Therefore, the union of ω-limit sets of the semiflow Φ induced by ODE (3.17) coincides with the set of equilibrium points of this semiflow. The latter set itself corresponds to the set of points $(z,0)$ s.t. $z \in \operatorname{zer} g_\infty$. It remains to notice that $\Upsilon = \operatorname{zer} g_\infty$ to finish the proof.

Remark 14. Commenting on Remark 4, the same proof works for (ODE-1′) by using the function $F - F_\star$ as a Lyapunov function. The corresponding limit set (as $t \to +\infty$) is then of the form

$$\{\tilde{z}_\infty = (\tilde{v}_\infty, \tilde{x}_\infty) \in \mathbb{R}^d_+ \times \mathbb{R}^d \,:\, \nabla F(\tilde{x}_\infty) = 0\,, \tilde{v}_\infty = p_\infty S(\tilde{x}_\infty)/q_\infty\}.$$

Similarly, if we set $\mathsf{p} = \mathsf{q} \equiv 0$ in (ODE-1) and we keep what remains in Assumption 3.2.4, the function $\mathsf{h}(t)(F(x) - F_\star) + \frac{1}{2}\|m\|^2$ works as a Lyapunov function, and the limit set has the form $\{(0,x) : \nabla F(x) = 0\}$.

3.5.2 Proof of Th. 3.2

The existence and the uniqueness of the solution to (ODE-N) have been shown in the literature. We refer to (Cabot et al., 2009, Prop. 2.1-2.2.c)) for an identical statement of this result and (Su et al., 2016, Th. 1, Appendix A) for a complete proof. The boundedness of the solution follows immediately from the coercivity of F together with the fact that the function $t \mapsto F(\mathsf{x}(t)) + \frac{1}{2}\|\mathsf{m}(t)\|^2$ is nonincreasing.

Concerning the convergence statement, our proof is based on comparing the solutions of (ODE-N) to the solutions of the ODE in (Gadat et al., 2018, Eq. (2.3)). We first note that under a change of variable, a solution to (ODE-N) gives a solution to (Gadat et al., 2018, Eq. (2.3)).

Lemma 3.14. Let (m,x) be a solution to (ODE-N). Define $\mathsf{y}(t) = \frac{\kappa \mathsf{m}\left(\kappa\sqrt{t}\right)}{2\sqrt{t}}$, $\mathsf{u}(t) = \mathsf{x}\left(\kappa\sqrt{t}\right)$, with $\kappa = \sqrt{2\alpha+2}$ and $\beta = \frac{\kappa^2}{4}$. Then, (y,u) verifies

$$\begin{cases} \dot{\mathsf{y}}(t) &= \frac{\beta}{t}(\nabla F(\mathsf{u}(t))) - \mathsf{y}(t)) \\ \dot{\mathsf{u}}(t) &= -\mathsf{y}(t)\,. \end{cases} \tag{3.22}$$

Proof. By simple differentiation, we get:

$$\dot{\mathsf{y}}(t) = \frac{\beta}{t}\left[\nabla F\left(\mathsf{x}(\kappa\sqrt{t})\right) - \frac{\alpha}{\kappa\sqrt{t}}\mathsf{m}\left(\kappa\sqrt{t}\right)\right] - \frac{\kappa}{4t^{\frac{3}{2}}}\mathsf{m}\left(\kappa\sqrt{t}\right) = \frac{\beta}{t}\left(\nabla F(\mathsf{u}(t)) - \mathsf{y}(t)\right),$$

$$\dot{\mathsf{u}}(t) = -\frac{\kappa}{2\sqrt{t}}\mathsf{m}\left(\kappa\sqrt{t}\right) = -\mathsf{y}(t)\,.$$

■

Consider a solution (m, x) of (ODE-N) starting at $(m_0, x_0) \in \mathbb{R}^d \times \mathbb{R}^d$. As in Section 3.5.1.2, for every $t_0 > 0$, on $[t_0, +\infty)$, we have that $(\mathsf{m}, \mathsf{x}, \mathsf{s})$ is a solution to the autonomous ODE

$$\begin{cases} \dot{\mathsf{m}}(t) &= \nabla F(\mathsf{x}(t)) - \alpha \mathsf{s}(t) \mathsf{m}(t) \\ \dot{\mathsf{x}}(t) &= -\mathsf{m}(t) \\ \dot{\mathsf{s}}(t) &= -\mathsf{s}(t)^2 \,, \end{cases} \tag{3.23}$$

starting at $(m_0, x_0, 1/t_0)$. Denote by $\Phi_N = (\Phi_N^m, \Phi_N^x, \Phi_N^s)$ the semiflow induced by ODE (3.23) and $\omega_N((m_0, x_0, 1/t_0))$ its limit set.

Define (y, u) as in Lem. 3.14. Starting at $(\mathsf{y}(t_0), \mathsf{u}(t_0), 1/t_0)$, we also have that $(\mathsf{y}, \mathsf{u}, \mathsf{s})$ is a solution on $[t_0, +\infty)$ to the "autonomized" Heavy-Ball ODE

$$\begin{cases} \dot{\mathsf{y}}(t) &= \beta \mathsf{s}(t)(\nabla F(\mathsf{u}(t))) - \mathsf{y}(t)) \\ \dot{\mathsf{u}}(t) &= -\mathsf{y}(t) \\ \dot{\mathsf{s}}(t) &= -\mathsf{s}(t)^2 \,. \end{cases} \tag{3.24}$$

Denote by $\Phi_H = (\Phi_H^y, \Phi_H^u, \Phi_H^s)$ the semiflow induced by ODE (3.24) and its limit set $\omega_H((\mathsf{y}(t_0), \mathsf{u}(t_0), 1/t_0))$.

Lemma 3.15. For any compact set $K \subset \mathbb{R}^{2d+1}$ and any $T > 0$, the family of functions $\left\{ \Phi(z, \cdot) : [0, T] \to \mathbb{R}^{2d+1} \right\}_{z \in K}$, where Φ is either Φ_H or Φ_N, is relatively compact in $(\mathcal{C}^0([0, T], \mathbb{R}^{2d+1}), \|\cdot\|_\infty)$.

Proof. The map $\Phi : \mathbb{R}^{2d+1} \times \mathbb{R}_+ \to \mathbb{R}^{2d+1}$ is continuous, hence uniformly continuous on $K \times [0, T]$. The result follows from the application of the Arzelà-Ascoli theorem to the family $\left\{ \Phi(z, \cdot) : [0, T] \to \mathbb{R}^{2d+1} \right\}_{z \in K}$. ∎

Let $(m, x, 0) \in \omega_N((m_0, x_0, 1/t_0))$. There exists a sequence (t_k) of nonnegative reals such that $(m, x, 0) = \lim_{k \to \infty} (\mathsf{m}(t_k), \mathsf{x}(t_k), 1/t_k)$. For any $T > 0$, using Lem. 3.15, up to an extraction, we can say that the sequence of functions $\{\Phi_N((\mathsf{m}(t_k), \mathsf{x}(t_k), 1/t_k), \cdot)\}_k$ converges towards $(\tilde{\mathsf{m}}, \tilde{\mathsf{x}}, 0)$ in $\mathcal{C}^0([0, T], \mathbb{R}^d)$, where $(\tilde{\mathsf{m}}, \tilde{\mathsf{x}})$ is a solution to

$$\begin{cases} \dot{\tilde{\mathsf{m}}}(t) &= \nabla F(\tilde{\mathsf{x}}(t)) \\ \dot{\tilde{\mathsf{x}}}(t) &= -\tilde{\mathsf{m}}(t) \,, \end{cases} \tag{3.25}$$

with $(\tilde{\mathsf{m}}(0), \tilde{\mathsf{x}}(0)) = (m, x)$. Moreover, by Lem. 3.14, we also have that:

$$\begin{aligned} &\sup_{h \in [0, T^2/\kappa^2]} \left\| \tilde{\mathsf{x}}(\kappa\sqrt{h}) - \Phi_N^x((\mathsf{m}(t_k), \mathsf{x}(t_k), 1/t_k), \kappa\sqrt{h}) \right\| \\ &\qquad = \sup_{h \in [0, T^2/\kappa^2]} \left\| \tilde{\mathsf{x}}(\kappa\sqrt{h}) - \Phi_H^u((\mathsf{m}(t_k), \mathsf{x}(t_k), 1/t_k), h) \right\| \xrightarrow[k \to +\infty]{} 0 \,. \end{aligned} \tag{3.26}$$

Using Lem. 3.15, up to an additional extraction, we get on $\mathcal{C}^0([0, T^2/\kappa^2], \mathbb{R}^{2d+1})$ that $\{\Phi_H((\mathsf{x}(t_k), \mathsf{m}(t_k), 1/t_k), \cdot)\}_k$ converges to $(\mathsf{u}, \mathsf{y}, 0)$, where (u, y) is a solution to

$$\begin{cases} \dot{\mathsf{y}}(t) &= 0 \\ \dot{\mathsf{u}}(t) &= -\mathsf{y}(t) \,. \end{cases} \tag{3.27}$$

Therefore, $\mathsf{u}(t) = A + Bt$ for some A and B in $\mathbb{R}^d$. Imagine that $B \neq 0$. We previously proved that x (and therefore u) is bounded by some constant $C > 0$. Let $T' > \frac{C+\|A\|}{\|B\|}$. Up to an extraction, we obtain that $\{\Phi_H((\mathsf{x}(t_k), \mathsf{m}(t_k), 1/t_k), \cdot)\}_k$ converges to u' on $\mathcal{C}^0([0,T'], \mathbb{R}^{2d+1})$, with $\mathsf{u}'(t) = A' + B't$ for some A' and B' in $\mathbb{R}^d$. We then have by uniqueness of the limit that $A' = A$ and $B' = B$. As a consequence, $\left\|\mathsf{u}'(T')\right\| = \left\|A + BT'\right\| > C$ and we obtain a contradiction. Hence $B = 0$.

This implies that u is constant. Then, if we go back to Eqs. (3.26) and (3.25), we get that $\tilde{\mathsf{x}}$ is constant, hence $\tilde{\mathsf{m}} \equiv 0$ and then $\nabla F(\tilde{\mathsf{x}}) \equiv 0$. In particular, this means that $m = \tilde{\mathsf{m}}(0) = 0$ and $\nabla F(x) = \nabla F(\tilde{\mathsf{x}}(0)) = 0$.

3.6 Proofs for Section 3.3

3.6.1 Preliminaries

We first recall some useful definitions and results. Let Ψ represent any semiflow on an arbitrary metric space (E, d). As in the previous section, a point $z \in E$ is called an equilibrium point of the semiflow Ψ if $\Psi(z,t) = z$ for all $t \geq 0$. We denote by Λ_Ψ the set of equilibrium points of Ψ. A continuous function $\mathsf{V} : E \to \mathbb{R}$ is called a Lyapunov function for the semiflow Ψ if $\mathsf{V}(\Psi(z,t)) \leq \mathsf{V}(z)$ for all $z \in E$ and all $t \geq 0$. It is called a *strict* Lyapunov function if, moreover, $\{z \in E : \forall t \geq 0, \mathsf{V}(\Psi(z,t)) = \mathsf{V}(z)\} = \Lambda_\Psi$. If V is a strict Lyapunov function for Ψ and if $z \in E$ is a point s.t. $\{\Psi(z,t) : t \geq 0\}$ is relatively compact, then it holds that $\Lambda_\Psi \neq \varnothing$ and $\mathsf{d}(\Psi(z,t), \Lambda_\Psi) \to 0$, see (Haraux, 1991, Th. 2.1.7). A continuous function $z : [0, +\infty) \to E$ is said to be an asymptotic pseudotrajectory (APT, Benaïm and Hirsch (1996)) for the semiflow Ψ if $\lim_{t\to+\infty} \sup_{s\in[0,T]} \mathsf{d}(z(t+s), \Psi(z(t), s)) = 0$ for every $T \in (0, +\infty)$.

3.6.2 Proof of Th. 3.3

Recall that Φ is the semiflow induced by the autonomous ODE (3.17) which is an "autonomized" version of our initial (ODE-1). In the remainder of this section, the proof will be divided into two main steps : (a) we show that a certain continuous-time linearly interpolated process constructed from the iterates of our algorithm 3.1 is an APT of Φ; (b) we exhibit a strict Lyapunov function for a restriction to a carefully chosen compact set of a well chosen semiflow related to Φ. Then, we characterize the limit set of the APT using (Benaïm, 1999, Th. 5.7) and (Benaïm, 1999, Prop. 6.4). The sequence (z_n) converges almost surely to this same limit set.

(a) APT. For every $n \geq 1$, define $\bar{z}_n = (v_n, m_n, x_{n-1})$ (note the shift in the index of the variable x). We have the decomposition

$$\bar{z}_{n+1} = \bar{z}_n + \gamma_{n+1} g(\bar{z}_n, \tau_n) + \gamma_{n+1}\eta_{n+1} + \gamma_{n+1}\varsigma_{n+1}\,,$$

where g is defined in Eq. (3.1),

$$\eta_{n+1} = \Big(p_n(\nabla f(x_n, \xi_{n+1})^{\odot 2} - S(x_n)),\ h_n(\nabla f(x_n, \xi_{n+1}) - \nabla F(x_n)),\ 0\Big)\,, \qquad (3.28)$$

is a martingale increment and where we set $\varsigma_{n+1} = (\varsigma^v_{n+1}, \varsigma^m_{n+1}, \varsigma^x_{n+1})$ with the components defined by:

$$\begin{cases} \varsigma^v_{n+1} &= p_n(S(x_n) - S(x_{n-1})) \\ \varsigma^m_{n+1} &= h_n(\nabla F(x_n) - \nabla F(x_{n-1})) \\ \varsigma^x_{n+1} &= (\frac{\gamma_n}{\gamma_{n+1}} - 1)\frac{m_n}{\sqrt{v_n+\varepsilon}}\,. \end{cases}$$

We first prove that $\varsigma_n \to 0$ a.s. by considering the components separately. The components ς^m_{n+1} and ς^v_{n+1} converge a.s. to zero by using Assumptions 3.2.1, 3.2.3, together with the boundedness of the sequences (p_n) and (h_n) (which are both convergent). Indeed, since ∇F is locally Lipschitz continuous and the sequence (z_n) is supposed to be almost surely bounded, there exists a constant C s.t. $\|\nabla F(x_n) - \nabla F(x_{n-1})\| \leq C\|x_n - x_{n-1}\| \leq \frac{C}{\varepsilon}\gamma_n\|m_n\|$. The same inequality holds when replacing ∇F by S which is also locally Lipschitz continuous. The component ς^x_{n+1} also converges a.s. to zero by observing that $\|\varsigma^x_{n+1}\| \leq |1 - \frac{\gamma_n}{\gamma_{n+1}}|.\|m_n\|/\sqrt{\varepsilon}$ and using Assumption 3.3.2 together with the a.s. boundedness of (z_n). Now consider the martingale increment sequence (η_n), adapted to $\mathcal{F}_n$. Take $\delta > 0$. Since (z_n) is a.s bounded, there is a constant $C' > 0$ such that $\mathbb{P}(\sup \|x_n\| > C') \leq \delta$. Denoting $\tilde{\eta}_n := \eta_n \mathbb{1}_{\|x_n\| \leq C'}$ and combining Assumptions 3.2.4 with 3.3.4-*i)* we can show using convexity inequalities that

$$\sup_n \mathbb{E}\|\tilde{\eta}_{n+1}\|^q < \infty.$$

Then, we deduce from this result together with the corresponding stepsize assumption from 3.3.4-*i)* and (Benaïm, 1999, Prop. 4.2) (see also (Métivier and Priouret, 1987, Prop. 8)) the key property:

$$\forall T > 0\,, \quad \max\left\{ \left\| \sum_{k=n}^{L-1} \gamma_{k+1}\tilde{\eta}_{k+1} \right\| \;:\; L = n+1, \ldots, J(\tau_n + T) \right\} \xrightarrow[n\to\infty]{\text{a.s.}} 0 \tag{3.29}$$

where $J(t) = \max\{n \geq 0 \,:\, \tau_n \leq t\}$. Hence, for all $T > 0$, with probability at least $1 - \delta$:

$$\max\left\{ \left\| \sum_{k=n}^{L-1} \gamma_{k+1}\eta_{k+1} \right\| \;:\; L = n+1, \ldots, J(\tau_n + T) \right\} \xrightarrow[n\to\infty]{} 0\,. \tag{3.30}$$

Since δ can be chosen arbitrary small, Eq. (3.30) remains true with probability 1. This result also holds under Assumption 3.3.4-*ii)* (instead of 3.3.4-*i)*) by applying (Benaïm, 1999, Prop. 4.4).

Let $\boldsymbol{z} : [0, +\infty) \to \mathcal{Z}_+$ be the continous-time linearly interpolated process given by

$$\boldsymbol{z}(t) = \bar{z}_n + (t - \tau_n)\frac{\bar{z}_{n+1} - \bar{z}_n}{\gamma_{n+1}} \qquad \Big(\forall n \in \mathbb{N}\,,\, \forall t \in [\tau_n, \tau_{n+1})\Big)$$

(where $\tau_n = \sum_{k=1}^n \gamma_k$). Let $t_0 > 0$. Define $\boldsymbol{u} : [t_0, \infty) \to \mathcal{Z} \times (0, 1/t_0]$ by

$$\boldsymbol{u}(t) = \begin{bmatrix} \boldsymbol{z}(t) \\ 1/t \end{bmatrix}, \quad \text{for} \quad t \geq t_0 > 0.$$

Using Eq. (3.30) and the almost sure boundedness of the sequence (z_n) along with the fact that ς_n converges a.s. to zero, it follows from (Benaïm, 1999, Prop. 4.1, Remark 4.5) that $\boldsymbol{u}(t)$ is an APT of the already defined semiflow Φ induced by (3.17). Remark that it

also holds that $\boldsymbol{z}(t)$ is an APT of the semiflow Φ^∞ induced by (3.20). As the trajectory of $\boldsymbol{u}(t)$ is precompact, the limit set

$$\boldsymbol{L}(\boldsymbol{u}) = \bigcap_{t\geq t_0} \overline{\boldsymbol{u}([t,\infty))}$$

is compact. Moreover, it has the form

$$\boldsymbol{L}(\boldsymbol{u}) = \begin{bmatrix} \boldsymbol{S} \\ 0 \end{bmatrix}, \quad \text{where} \quad \boldsymbol{S} := \bigcap_{t\geq t_0} \overline{\boldsymbol{z}([t,\infty))}\,. \tag{3.31}$$

Our objective now is to prove that

$$\boldsymbol{S} \subset \Lambda_{\Phi^\infty}\,. \tag{3.32}$$

In order to establish this inclusion, we study the behavior of the restriction $\Phi|\boldsymbol{L}$ of the semiflow Φ to the set $\boldsymbol{L}$ (which is well-defined since $\boldsymbol{L}$ is Φ-invariant). Remark that

$$\Phi|\boldsymbol{L} = \begin{bmatrix} \Phi^\infty|\boldsymbol{S} \\ 0 \end{bmatrix},$$

where Φ^∞ is the semiflow associated to (3.20). In the second part of the proof, we establish Eq. (3.32) combining item (a) we just proved with (Benaïm, 1999, Th. 5.7) and (Benaïm, 1999, Prop. 6.4). In order to use the latter proposition, we prove a useful proposition in item (b).

(b) Strict Lyapunov function and convergence. For every $\delta > 0$ and every $z = (v, m, x) \in \mathcal{Z}_+$, define:

$$W_\delta(v, m, x) := \mathcal{E}_\infty(z) - \delta\langle \nabla F(x), m\rangle + \delta\|q_\infty v - p_\infty S(x)\|^2\,, \tag{3.33}$$

where, under Assumption 3.2.4-*i)*, the function $\mathcal{E}_\infty$ is defined by

$$\mathcal{E}_\infty(z) := \lim_{t\to+\infty} \mathcal{E}(t,z) = h_\infty(F(x) - F_\star) + \frac{1}{2}\left\| \frac{m}{(v+\varepsilon)^{\odot\frac{1}{4}}} \right\|^2. \tag{3.34}$$

Proposition 3.16. Let $t_0 > 0$ and let Assumptions 3.2.1 to 3.2.4 and 3.3.5 hold true. Let $\boldsymbol{S}$ be the limit set defined in Eq. (3.31). Let $\overline{\Phi}^\infty : \boldsymbol{S}\times[t_0,+\infty) \to \boldsymbol{S}$ be the restriction of the semiflow Φ^∞ to $\boldsymbol{S}$ *i.e.*, $\overline{\Phi}^\infty(z,t) = \Phi^\infty(z,t)$ for all $z \in \boldsymbol{S}, t \geq t_0$.Then,

i) $\boldsymbol{S}$ is compact.
ii) $\overline{\Phi}^\infty$ is a well-defined semiflow on $\boldsymbol{S}$.
iii) The set of equilibrium points of $\overline{\Phi}^\infty$ is equal to $\Lambda_{\Phi^\infty} \cap \boldsymbol{S}$.
iv) There exists $\delta > 0$ s.t. W_δ is a strict Lyapunov function for the semiflow $\overline{\Phi}^\infty$.

Proof. The first point is a consequence of the definition of $\boldsymbol{S}$ and the boundedness of $\boldsymbol{z}$. The second point stems from the definition of Φ^∞. Observing that $\overline{\Phi}^\infty$ is valued in $\boldsymbol{S}$, the third point is immediate from the definition of Λ_{Φ^∞}. We now prove the last point. Consider $z \in \boldsymbol{S}$ and write $\overline{\Phi}^\infty(z,t)$ under the form $\overline{\Phi}^\infty(z,t) = (\mathsf{v}(t), \mathsf{m}(t), \mathsf{x}(t))$. Notice that this quantity is bounded as a function of the variable t. For *any* map $\mathsf{W} : \mathcal{Z}_+ \to \mathbb{R}$, define for all $t \geq t_0$, $\mathcal{L}_\mathsf{W}(t) := \limsup_{s\to 0} s^{-1}(\mathsf{W}(\overline{\Phi}^\infty(z,t+s)) - \mathsf{W}(\overline{\Phi}^\infty(z,t)))$. Introduce $G(z) := -\langle \nabla F(x), m\rangle$ and $H(z) := \|q_\infty v - p_\infty S(x)\|^2$ for every $z = (v,m,x) \in \mathcal{Z}_+$.

Consider $\delta > 0$ (to be specified later on). We study $\mathcal{L}_{W_\delta} = \mathcal{L}_{\mathcal{E}_\infty} + \delta\mathcal{L}_G + \delta\mathcal{L}_H$. Note that $\overline{\Phi}^\infty(z,t) \in \boldsymbol{S} \cap \mathcal{Z}_+$ for all $t \geq t_0$ by an analogous result to Lem. 3.12 for Φ^∞. Thus, $t \mapsto \mathcal{E}_\infty(\overline{\Phi}^\infty(z,t))$ is differentiable at any point $t \geq t_0$ and $\mathcal{L}_{\mathcal{E}_\infty}(t) = \frac{d}{dt}\mathcal{E}_\infty(\overline{\Phi}^\infty(z,t))$. Using similar derivations to Ineq. (3.16), we obtain that

$$\mathcal{L}_{\mathcal{E}_\infty}(t) \leq -\left(r_\infty - \frac{q_\infty}{4}\right)\left\|\frac{\mathsf{m}(t)}{(\mathsf{v}(t)+\varepsilon)^{\odot\frac{1}{4}}}\right\|^2 . \tag{3.35}$$

We now study $\mathcal{L}_G$. For every $t \geq t_0$,

$$\begin{aligned}\mathcal{L}_G(t) &= \limsup_{s\to 0} s^{-1}(-\langle \nabla F(\mathsf{x}(t+s)), \mathsf{m}(t+s)\rangle + \langle \nabla F(\mathsf{x}(t)), \mathsf{m}(t)\rangle)\\ &\leq \limsup_{s\to 0} s^{-1}\|\nabla F(\mathsf{x}(t)) - \nabla F(\mathsf{x}(t+s))\|\|\mathsf{m}(t+s)\| - \langle \nabla F(\mathsf{x}(t)), \dot{\mathsf{m}}(t)\rangle .\end{aligned}$$

Let $L_{\nabla F}$ be the Lipschitz constant of ∇F on the bounded set $\{x : (v,m,x) \in \boldsymbol{S}\}$. Define $C_1 := \sup_t \|\sqrt{\mathsf{v}(t)+\varepsilon}\|$. Then,

$$\begin{aligned}\mathcal{L}_G(t) &\leq L_{\nabla F}\limsup_{s\to 0} s^{-1}\|\mathsf{x}(t) - \mathsf{x}(t+s)\|\|\mathsf{m}(t+s)\| - \langle \nabla F(\mathsf{x}(t)), \dot{\mathsf{m}}(t)\rangle\\ &\leq L_{\nabla F}\|\dot{\mathsf{x}}(t)\|\|\mathsf{m}(t)\| - \langle \nabla F(\mathsf{x}(t)), \dot{\mathsf{m}}(t)\rangle\\ &\leq L_{\nabla F}\|\dot{\mathsf{x}}(t)\|\|\mathsf{m}(t)\| - h_\infty\|\nabla F(\mathsf{x}(t))\|^2 + r_\infty\langle \nabla F(\mathsf{x}(t)), \mathsf{m}(t)\rangle\\ &\leq \left(\frac{L_{\nabla F}C_1^{\frac{1}{2}}}{\varepsilon^{\frac{1}{4}}} + \frac{r_\infty C_1}{2u_1^2}\right)\left\|\frac{\mathsf{m}(t)}{(\mathsf{v}(t)+\varepsilon)^{\odot\frac{1}{4}}}\right\|^2 - \left(h_\infty - \frac{r_\infty u_1^2}{2}\right)\|\nabla F(\mathsf{x}(t))\|^2 \end{aligned}\tag{3.36}$$

where we used the classical inequality $|\langle a,b\rangle| \leq \|a\|^2/(2u^2) + u^2\|b\|^2/2$ for any non-zero real u to derive the last above inequality. We now study $\mathcal{L}_H$. For every $t \geq t_0$,

$$\begin{aligned}\mathcal{L}_H(t) &= \limsup_{s\to 0} s^{-1}(\|q_\infty\mathsf{v}(t+s) - p_\infty S(\mathsf{x}(t+s))\|^2 - \|q_\infty\mathsf{v}(t) - p_\infty S(\mathsf{x}(t))\|^2)\\ &= \limsup_{s\to 0} s^{-1}(p_\infty^2\|S(\mathsf{x}(t)) - S(\mathsf{x}(t+s))\|^2\\ &+ 2p_\infty\langle S(\mathsf{x}(t)) - S(\mathsf{x}(t+s)), q_\infty\mathsf{v}(t+s) - p_\infty S(\mathsf{x}(t))\rangle)\\ &+ \lim_{s\to 0} s^{-1}(\|q_\infty\mathsf{v}(t+s) - p_\infty S(\mathsf{x}(t))\|^2 - \|q_\infty\mathsf{v}(t) - p_\infty S(\mathsf{x}(t))\|^2) .\end{aligned}$$

The second term in the righthand side coincides with

$$-2q_\infty\langle p_\infty S(\mathsf{x}(t)) - q_\infty\mathsf{v}(t), \dot{\mathsf{v}}(t)\rangle = -2q_\infty\|p_\infty S(\mathsf{x}(t)) - q_\infty\mathsf{v}(t)\|^2 .$$

Denote by L_S the Lipschitz constant of S on the set $\{x : (v,m,x) \in \boldsymbol{S}\}$. Note that $s^{-1}(\|S(\mathsf{x}(t+s)) - S(\mathsf{x}(t))\|^2) \leq L_S^2 s\|s^{-1}(\mathsf{x}(t+s) - \mathsf{x}(t))\|^2$ which converges to zero as $s \to 0$. Thus,

$$\begin{aligned}\mathcal{L}_H(t) &= -2q_\infty\|p_\infty S(\mathsf{x}(t)) - q_\infty\mathsf{v}(t)\|^2\\ &+ \limsup_{s\to 0} 2p_\infty s^{-1}\langle S(\mathsf{x}(t)) - S(\mathsf{x}(t+s)), q_\infty\mathsf{v}(t+s) - p_\infty S(\mathsf{x}(t))\rangle\\ &\leq -2q_\infty\|p_\infty S(\mathsf{x}(t)) - q_\infty\mathsf{v}(t)\|^2 + 2p_\infty\|\dot{\mathsf{x}}(t)\|L_S\|q_\infty\mathsf{v}(t) - p_\infty S(x(t))\|\\ &\leq \frac{p_\infty}{\varepsilon^{\frac{1}{2}}u_2^2}\left\|\frac{\mathsf{m}(t)}{(\mathsf{v}(t)+\varepsilon)^{\odot\frac{1}{4}}}\right\|^2 - (2q_\infty - p_\infty u_2^2 L_S^2)\|p_\infty S(\mathsf{x}(t)) - q_\infty\mathsf{v}(t)\|^2 .\end{aligned}\tag{3.37}$$

Recalling that $\mathcal{L}_{W_\delta} = \mathcal{L}_{\mathcal{E}_\infty} + \delta\mathcal{L}_G + \delta\mathcal{L}_H$ and combining Eqs. (3.35), (3.36) and (3.37), we obtain for every $t \geq t_0$,

$$\mathcal{L}_{W_\delta}(t) \leq -M(\delta)\left\|\frac{\mathsf{m}(t)}{(\mathsf{v}(t)+\varepsilon)^{\odot\frac{1}{4}}}\right\|^2 - \delta\left(h_\infty - \frac{r_\infty u_1^2}{2}\right)\|\nabla F(\mathsf{x}(t))\|^2 \\ - \delta\left(2q_\infty - p_\infty u_2^2 L_S^2\right)\|p_\infty S(\mathsf{x}(t)) - q_\infty \mathsf{v}(t)\|^2 . \quad (3.38)$$

where $M(\delta) := r_\infty - \frac{q_\infty}{4} - \delta\left(\frac{r_\infty C_1}{2u_1^2} + \frac{L_{\nabla F} C_1^{\frac{1}{2}}}{\varepsilon^{\frac{1}{4}}} + \frac{p_\infty}{\varepsilon^{\frac{1}{2}} u_2^2}\right)$. Now select u_1, u_2 small enough s.t. $h_\infty - r_\infty u_1^2/2 > 0$ and $2q_\infty - p_\infty u_2^2 L_S^2 > 0$. Then, choose δ in such a way that $M(\delta) > 0$. Thus, there exists a constant c depending on δ s.t. for every $t \geq t_0$,

$$\mathcal{L}_{W_\delta}(t) \leq -c\left(\left\|\frac{\mathsf{m}(t)}{(\mathsf{v}(t)+\varepsilon)^{\odot\frac{1}{4}}}\right\|^2 + \|\nabla F(\mathsf{x}(t))\|^2 + \|p_\infty S(\mathsf{x}(t)) - q_\infty \mathsf{v}(t)\|^2\right) . \quad (3.39)$$

It can easily be seen that for every $z \in \boldsymbol{S}$, $t \mapsto W_\delta(\overline{\Phi}^\infty(z,t))$ is Lipschitz continuous, hence absolutely continuous. Its derivative almost everywhere coincides with $\mathcal{L}_{W_\delta}$, which is nonpositive. Thus, W_δ is a Lyapunov function for $\overline{\Phi}^\infty$. We prove that the Lyapunov function is strict. Consider $z = (v, m, x) \in \boldsymbol{S}$ s.t. $W_\delta(\overline{\Phi}^\infty(z,t)) = W_\delta(z)$ for all $t \geq t_0$. The derivative almost everywhere of $t \mapsto W_\delta(\overline{\Phi}^\infty(z,t))$ is identically zero, and by Eq. (3.39), this implies that

$$-c\left(\left\|\frac{\mathsf{m}(t)}{(\mathsf{v}(t)+\varepsilon)^{\odot\frac{1}{4}}}\right\|^2 + \|\nabla F(\mathsf{x}(t))\|^2 + \|p_\infty S(\mathsf{x}(t)) - q_\infty \mathsf{v}(t)\|^2\right)$$

is equal to zero for every $t \geq t_0$ a.e. (hence, for every $t \geq t_0$, by continuity of $\overline{\Phi}^\infty$). In particular for $t = t_0$, $m = \nabla F(x) = 0$ and $p_\infty S(x) - q_\infty v = 0$. Hence, $z \in \operatorname{zer} g_\infty \cap \boldsymbol{S}$. This concludes the proof since $\Lambda_{\Phi^\infty} = \operatorname{zer} g_\infty$. ■

End of the Proof of Th. 3.3. Finally, Assumption 3.3.5 implies that the set $W_\delta(\Lambda_{\Phi^\infty} \cap \boldsymbol{S})$ is of empty interior. Recall that Assumptions 3.2.1 and 3.2.3 both follow from Assumption 3.3.3 made in Th. 3.3. Given Prop. 3.16, the proof is concluded by applying (Benaïm, 1999, Prop. 6.4) to the restricted semiflow $\bar{\Phi}^\infty$ (with $(M, \Lambda) = (\boldsymbol{S}, \Lambda_{\bar{\Phi}^\infty})$). Note that a Lyapunov function for $\Lambda_{\bar{\Phi}^\infty}$ is what is called a strict Lyapunov function. Such a function is provided by Prop. 3.16. We obtain as a conclusion of (Benaïm, 1999, Prop. 6.4) that $\boldsymbol{S} \subset \Lambda_{\bar{\Phi}^\infty}$. This gives the desired result (Eq. (3.32)) given Prop. 3.16-*iii)*.

The last assertion of Th. 3.3 is a consequence of (Benaïm, 1999, Cor. 6.6).

3.6.3 Proof of Th. 3.5

We can rewrite the iterates from Algorithm 3.2 as follows:

$$\begin{cases} m_{n+1} &= m_n + \gamma_{n+1}(\nabla F(x_n) - \frac{\alpha}{\tau_n} m_n) + \gamma_{n+1}(\nabla f(x_n, \xi_{n+1}) - \nabla F(x_n)) \\ x_{n+1} &= x_n - \gamma_{n+1} m_{n+1} . \end{cases} \quad (3.40)$$

We prove that the sequence $(y_n = (m_n, x_n) : n \in \mathbb{N})$ of iterates of this algorithm converges almost surely towards the set $\bar{\Upsilon}$ defined in Eq. (3.3) if it is supposed to be bounded with probability one. The proof follows a similar path to the proof in Section 3.5.2.

Indeed, denote by X and M the linearly interpolated processes constructed respectively from the sequences (x_n) and (m_n) and let $\mathsf{s}(t) = 1/t$. Recall that $\Phi_N = (\Phi_N^m, \Phi_N^x, \Phi_N^s)$ is the semiflow induced by (3.23). As in Section 3.6.2, we have that $\mathsf{Z} := (\mathsf{M}, \mathsf{X}, \mathsf{s})$ is an APT of (3.23). In particular, this means that

$$\forall T > 0\,, \quad \sup_{h\in[0,T]} \left\| \mathsf{X}(t+h) - \Phi_N^x(\mathsf{Z}(t), h) \right\| \xrightarrow[t\to\infty]{} 0\,. \tag{3.41}$$

By Lem. 3.14, we also have that

$$\begin{aligned} &\sup_{h\in[0,T^2/\kappa^2]} \left\| \mathsf{X}(t+\kappa\sqrt{h}) - \Phi_N^x(\mathsf{Z}(t), \kappa\sqrt{h}) \right\| \\ &\qquad = \sup_{h\in[0,T^2/\kappa^2]} \left\| \mathsf{X}(t+\kappa\sqrt{h}) - \Phi_H^u(\mathsf{Z}(t), h) \right\| \xrightarrow[t\to\infty]{} 0\,. \end{aligned} \tag{3.42}$$

Let (m, x) be a limit point of the sequence (y_n) and let $T > 0$. Using Lem. 3.15, we can proceed in the same manner as in Section 3.5.2 and get a sequence (t_k) such that

$$(\mathsf{M}(t_k + \cdot), \mathsf{X}(t_k + \cdot)) \to (\mathsf{m}, \mathsf{x}) \text{ and } (\Phi_H^y(\mathsf{Z}(t_k), \cdot), \Phi_H^u(\mathsf{Z}(t_k), \cdot)) \to (\mathsf{y}, \mathsf{u})\,,$$

where $(\mathsf{m}(0), \mathsf{x}(0)) = (m, x)$, and (m, x) and (x, u) are respectively solutions to (3.25) and (3.27). As in the end of Section 3.5.2, we obtain that u and x are constant, therefore $\mathsf{m} \equiv 0$ and $\nabla F(\mathsf{x}) \equiv 0$, which finishes the proof.

3.6.4 Proof of Th. 3.4

The idea of the proof is to apply Robbins-Siegmund's theorem (Robbins and Siegmund, 1971) to

$$V_n = h_{n-1} F(x_n) + \frac{1}{2} \langle m_n^{\odot 2}, \frac{1}{\sqrt{v_n + \varepsilon}} \rangle$$

(note the similarity of V_n with the energy function (3.15)). Since $\inf F > -\infty$, we assume without loss of generality that $F \geq 0$. In this subsection, we use the notation ∇f_{n+1} as a shorthand notation for $\nabla f(x_n, \xi_{n+1})$ and C denotes some positive constant which may change from line to line. We write $\mathbb{E}_n = \mathbb{E}[\cdot \,|\, \mathcal{F}_n]$ for the conditional expectation w.r.t the σ-algebra $\mathcal{F}_n$. Define $P_n := \frac{1}{2}\langle D_n, m_n^{\odot 2} \rangle$, with $D_n := \frac{1}{\sqrt{v_n+\varepsilon}}$. We have the decomposition:

$$P_{n+1} - P_n = \frac{1}{2}\langle D_{n+1} - D_n, m_{n+1}^{\odot 2} \rangle + \frac{1}{2}\langle D_n, m_{n+1}^{\odot 2} - m_n^{\odot 2} \rangle. \tag{3.43}$$

We estimate the vector

$$D_{n+1} - D_n = \frac{\sqrt{v_n + \varepsilon} - \sqrt{v_{n+1} + \varepsilon}}{\sqrt{v_{n+1} + \varepsilon} \odot \sqrt{v_n + \varepsilon}}\,.$$

Remarking that $v_{n+1} \geq (1-\gamma_{n+1}q_n)v_n$ and using the update rule of v_n, we obtain for a sufficiently large n that

$$\begin{aligned}\sqrt{v_n+\varepsilon} - \sqrt{v_{n+1}+\varepsilon} &= \gamma_{n+1}\frac{q_n v_n - p_n \nabla f_{n+1}^{\odot 2}}{\sqrt{v_n+\varepsilon}+\sqrt{v_{n+1}+\varepsilon}} \\ &\leq \gamma_{n+1}q_n \frac{v_n}{(1+\sqrt{1-\gamma_{n+1}q_n})\sqrt{v_n+\varepsilon}} \\ &= \frac{\gamma_{n+1}q_n}{1+\sqrt{1-\gamma_{n+1}q_n}}\sqrt{v_n}\odot\frac{\sqrt{v_n}}{\sqrt{v_n+\varepsilon}} \\ &\leq c_{n+1}\sqrt{v_{n+1}}\,, \end{aligned} \tag{3.44}$$

where $c_{n+1} := \frac{\gamma_{n+1}q_n}{\sqrt{1-\gamma_{n+1}q_n}(1+\sqrt{1-\gamma_{n+1}q_n})}$. It is easy to see that $c_{n+1}/\gamma_n \to q_\infty/2$. Thus, for any $\delta > 0$, $c_{n+1} \leq (q_\infty + 2\delta)\gamma_n/2$ for all n large enough. Using also that $\sqrt{v_{n+1}}/\sqrt{v_{n+1}+\varepsilon} \leq 1$, we obtain

$$D_{n+1} - D_n \leq \frac{q_\infty + 2\delta}{2}\gamma_n D_n\,. \tag{3.45}$$

Substituting the above inequality in Eq. (3.43), we obtain

$$\begin{aligned} P_{n+1} - P_n &\leq \left(\frac{q_\infty+2\delta}{2}\right)\frac{\gamma_n}{2}\langle D_n, m_{n+1}^{\odot 2}\rangle + \frac{1}{2}\langle D_n, m_{n+1}^{\odot 2} - m_n^{\odot 2}\rangle \\ &\leq \frac{q_\infty+2\delta}{2}\gamma_n P_n + \left(1+\frac{q_\infty+2\delta}{2}\gamma_n\right)\frac{1}{2}\langle D_n, m_{n+1}^{\odot 2} - m_n^{\odot 2}\rangle\,. \end{aligned}$$

Using $m_{n+1}^{\odot 2} - m_n^{\odot 2} = 2m_n \odot (m_{n+1}-m_n) + (m_{n+1}-m_n)^{\odot 2}$, and noting that $\mathbb{E}_n(m_{n+1} - m_n) = \gamma_{n+1}h_n\nabla F(x_n) - \gamma_{n+1}r_n m_n$,

$$\begin{aligned}\mathbb{E}_n\frac{1}{2}\langle D_n, m_{n+1}^{\odot 2} - m_n^{\odot 2}\rangle &= \gamma_{n+1}h_n\langle \nabla F(x_n), \frac{m_n}{\sqrt{v_n+\varepsilon}}\rangle - 2\gamma_{n+1}r_n P_n \\ &\quad + \frac{1}{2}\langle D_n, \mathbb{E}_n[(m_{n+1}-m_n)^{\odot 2}]\rangle\,. \end{aligned}$$

There exists $\delta > 0$ such that $r_\infty - \frac{q_\infty}{4} - \frac{\delta}{2} > 0$ by Assumption 3.2.4-*iv)*. As $\frac{\gamma_{n+1}}{\gamma_n}r_n - \frac{q_\infty}{4} \to r_\infty - \frac{q_\infty}{4}$, for all n large enough, $\frac{\gamma_{n+1}}{\gamma_n}r_n - \frac{q_\infty}{4} > r_\infty - \frac{q_\infty}{4} - \frac{\delta}{2} > 0$. Hence, for all n large enough,

$$\begin{aligned}\mathbb{E}_n P_{n+1} - P_n \leq -2\left(r_\infty - \frac{q_\infty}{4} - \frac{\delta}{2}\right)\gamma_n P_n + \gamma_{n+1}h_n\langle \nabla F(x_n), \frac{m_n}{\sqrt{v_n+\varepsilon}}\rangle \\ + C\gamma_n^2\langle \nabla F(x_n), \frac{m_n}{\sqrt{v_n+\varepsilon}}\rangle + C\langle D_n, \mathbb{E}_n[(m_{n+1}-m_n)^{\odot 2}]\rangle\,. \end{aligned} \tag{3.46}$$

Using the inequality $\langle u, v\rangle \leq (\|u\|^2 + \|v\|^2)/2$ and Assumption 3.3.6-*ii)*, it is easy to show the inequality $\langle \nabla F(x_n), \frac{m_n}{\sqrt{v_n+\varepsilon}}\rangle \leq C(1 + F(x_n) + P_n)$. Moreover, using the componentwise inequality $(h_n\nabla f_{n+1} - r_n m_n)^{\odot 2} \leq 2h_n^2\nabla f_{n+1}^{\odot 2} + 2r_n^2 m_n^{\odot 2}$ along with Assumption 3.3.6-*ii)* and the boundedness of the sequences $(h_n), (r_n)$ and (γ_{n+1}/γ_n), we obtain

$$\langle D_n, \mathbb{E}_n[(m_{n+1}-m_n)^{\odot 2}]\rangle \leq C\gamma_n^2(1 + F(x_n) + P_n)\,. \tag{3.47}$$

Combining Eq. (3.46) and Eq. (3.47), we get

$$\mathbb{E}_n(P_{n+1} - P_n) \leq \gamma_{n+1}h_n\langle\nabla F(x_n), m_n \odot D_n\rangle + C\gamma_n^2(1 + F(x_n) + P_n)\,. \quad (3.48)$$

Denoting by M the Lipschitz coefficient of ∇F, we also have

$$F(x_{n+1}) \leq F(x_n) - \gamma_{n+1}\langle\nabla F(x_n), m_{n+1} \odot D_{n+1}\rangle + \frac{\gamma_{n+1}^2 M}{2}\left\|m_{n+1} \odot D_{n+1}\right\|^2 . \quad (3.49)$$

Using (3.45) and the update rule of m_n, we have

$$\begin{aligned}
&\left\|m_{n+1} \odot D_{n+1} - m_n \odot D_n\right\|^2 \\
&\quad\leq C\left\|(m_{n+1} - m_n) \odot D_n\right\|^2 + C\left\|m_{n+1} \odot (D_{n+1} - D_n)\right\|^2 \\
&\quad\leq C\gamma_{n+1}^2(\left\|\nabla f_{n+1}\right\|^2 + \left\|m_n \odot D_n\right\|^2) + C\gamma_{n+1}^2\left\|m_{n+1} \odot D_n\right\|^2 \\
&\quad\leq C\gamma_{n+1}^2(\left\|m_n \odot D_n\right\|^2 + \left\|\nabla f_{n+1}\right\|^2)\,.
\end{aligned} \quad (3.50)$$

Finally, recalling that $V_n = h_{n-1}F(x_n)+P_n$, (h_n) is decreasing, combining Eq. (3.48),(3.49),(3.50), and using Assumption 3.3.6, we have

$$\begin{aligned}
\mathbb{E}_n[V_{n+1}] &\leq V_n + \gamma_{n+1}h_n\langle\nabla F(x_n), \mathbb{E}_n\left[m_n \odot D_n - m_{n+1} \odot D_{n+1}\right]\rangle \\
&\quad + C\gamma_{n+1}^2\left(1 + F(x_n) + P_n + \left\|m_n \odot D_n\right\|^2\right) \\
&\quad + C\gamma_{n+1}^2\mathbb{E}_n[\left\|m_n \odot D_n - m_{n+1} \odot D_{n+1}\right\|^2] \\
&\leq V_n + C\gamma_n^2\left(1 + F(x_n) + P_n + \left\|m_n \odot D_n\right\|^2 + \mathbb{E}_n\left[\left\|\nabla f_{n+1}\right\|^2\right]\right) \\
&\leq V_n + C\gamma_n^2(1 + F(x_n) + P_n) \\
&\leq (1 + C\gamma_n^2)V_n + C\gamma_n^2\,,
\end{aligned}$$

where we used Cauchy-Schwarz's inequality and the fact that $\left\|m_n \odot D_n\right\|^2 \leq CP_n$. By the Robbins-Siegmund's theorem, the sequence (V_n) converges almost surely to a finite random variable $V_\infty \in \mathbb{R}^+$. Then, the coercivity of F implies that (x_n) is almost surely bounded.

We now establish the almost sure boundedness of (m_n). Assume in the sequel that n is large enough to have $(1 - \gamma_{n+1}r_n) \geq 0$. Consider the martingale difference sequence $\Delta_{n+1} := \nabla f_{n+1} - \nabla F(x_n)$. We decompose $m_n = \bar{m}_n + \tilde{m}_n$ where $\bar{m}_{n+1} = (1 - \gamma_{n+1}r_n)\bar{m}_n + \gamma_{n+1}h_n\nabla F(x_n)$ and $\tilde{m}_{n+1} = (1-\gamma_{n+1}r_n)\tilde{m}_n + \gamma_{n+1}h_n\Delta_{n+1}$, setting $\bar{m}_0 = 0$ and $\tilde{m}_0 = m_0$. We prove that both terms $\bar{m}_n$ and $\tilde{m}_n$ are bounded. Consider the first term: $\|\bar{m}_{n+1}\| \leq (1 - \gamma_{n+1}r_n)\|\bar{m}_n\| + \gamma_{n+1}\sup_k \|h_k\nabla F(x_k)\|$, where the supremum in the above inequality is almost surely finite by continuity of ∇F. We immediately get that if $\left\|\bar{m}_n\right\| \geq \frac{\sup_k \|h_k\nabla F(x_k)\|}{r_\infty}$, then $\left\|\bar{m}_{n+1}\right\| \leq \|\bar{m}_n\|$. Thus

$$\left\|\bar{m}_{n+1}\right\| \leq \frac{\sup_k \|h_k\nabla F(x_k)\|}{r_\infty} + \sup_k \gamma_{k+1}\|h_k\nabla F(x_k)\|\,,$$

which implies that $\bar{m}_n$ is bounded. Consider now the term $\tilde{m}_n$:

$$\begin{aligned}\mathbb{E}_n[\|\tilde{m}_{n+1}\|^2] &= (1-\gamma_{n+1}r_n)^2\|\tilde{m}_n\|^2 + \gamma_{n+1}^2 h_n^2 \mathbb{E}_n[\|\Delta_{n+1}\|^2] \\ &\leq \|\tilde{m}_n\|^2 + \gamma_{n+1}^2 h_n^2 \mathbb{E}_n[\|\Delta_{n+1}\|^2]\,.\end{aligned}$$

Then, the inequality $\mathbb{E}_n[\|\Delta_{n+1}\|^2] \leq \mathbb{E}_n[\|\nabla f_{n+1}\|^2]$ combined with Assumption 3.3.4-*i)* and the a.s. boundedness of the sequence (x_n) imply that there exists a finite random variable $C_{\mathcal{K}}$ (independent of n) s.t. $\mathbb{E}_n[\|\nabla f_{n+1}\|^2] \leq C_{\mathcal{K}}$. As a consequence, since $\sum_n \gamma_{n+1}^2 < \infty$ and the sequence (h_n) is bounded, we obtain that a.s.:

$$\sum_{n\geq 0} \gamma_{n+1}^2 h_n^2 \mathbb{E}_n[\|\Delta_{n+1}\|^2] \leq C C_{\mathcal{K}} \sum_{n\geq 0} \gamma_{n+1}^2 < +\infty\,.$$

Hence, we can apply the Robbins-Siegmund theorem to obtain that $\sup_n \|\tilde{m}_n\|^2 < \infty$ w.p.1. Finally, it can be shown that (v_n) is almost surely bounded using the same arguments, decomposing v_n into $\bar{v}_n + \tilde{v}_n$ as above. Indeed, first, we have:

$$\mathbb{E}_n[\|\tilde{v}_{n+1}\|^2] \leq \|\tilde{v}_n\|^2 + \gamma_{n+1}^2 p_n^2 \mathbb{E}_n[\|\nabla f_{n+1}^{\odot 2} - S(x_n)\|^2]\,.$$

Second, it also holds that:

$$\mathbb{E}_n[\|\nabla f_{n+1}^{\odot 2} - S(x_n)\|^2] \leq \mathbb{E}_n[\|\nabla f_{n+1}^{\odot 2}\|^2] \leq \mathbb{E}_n[\|\nabla f_{n+1}\|^4]\,.$$

Then, using Assumption 3.3.4-*i)* and the a.s. boundedness of the sequence (x_n), there exists a finite random variable $C'_{\mathcal{K}}$ (independent of n) s.t. $\mathbb{E}_n[\|\nabla f_{n+1}\|^4] \leq C'_{\mathcal{K}}$. Moreover, the sequence (p_n) is bounded and $\sum_n \gamma_{n+1}^2 < \infty$. As a consequence, it holds that a.s:

$$\sum_{n\geq 0} \gamma_{n+1}^2 p_n^2 \mathbb{E}_n[\|\nabla f_{n+1}^{\odot 2} - S(x_n)\|^2] \leq C C'_{\mathcal{K}} \sum_{n\geq 0} \gamma_{n+1}^2 < +\infty\,.$$

It follows that the Robbins-Siegmund theorem can be applied to the sequence $\|\tilde{v}_n\|^2$ as for the sequence $\|\tilde{m}_n\|^2$ to obtain that $\sup_n \|\tilde{v}_n\|^2 < \infty$ w.p.1.

3.6.5 Proof of Th. 3.6

The proof of Th. 3.4 easily adapts to Algorithm 3.2 by replacing V_n by

$$\tilde{V}_n := F(x_n) + \frac{1}{2}\left\|m_n\right\|^2\,.$$

The boundedness of (m_n) is an immediate consequence of the convergence of $\tilde{V}_n$.

3.6.6 Proof of Th. 3.7

We shall use the following result.

Theorem 3.17 (adapted from Pelletier (1998), Th. 7)**.** *Let $k \geq 1$. On some probability space equipped with a filtration $\mathcal{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$, consider a sequence of r.v. on $\mathbb{R}^k$ given by*

$$Z_{n+1} = (I + \gamma_{n+1}\bar{H})Z_n + \gamma_{n+1}b_{n+1} + \sqrt{\gamma_{n+1}}\eta_{n+1}$$

and $\mathbb{E}[\|Z_0\|^2] < \infty$, where $\bar{H}$ is a $k \times k$ Hurwitz matrix, (b_n) and (η_n) are random sequences, and $\gamma_n = \gamma_0 n^{-\alpha}$ for some $\gamma_0 > 0$ and $\alpha \in (0,1]$. Let $\Omega_0 \in \mathcal{F}_\infty$ have a positive probability. Assume that the following holds almost surely on Ω_0:

i) $\mathbb{E}[\eta_{n+1}|\mathcal{F}_n] = 0$.
ii) *There exists a constant* $\bar{b} > 2$ *s.t.* $\sup_{n\geq 0} \mathbb{E}[\|\eta_{n+1}\|^{\bar{b}}|\mathcal{F}_n] < \infty$.
iii) $\mathbb{E}[\eta_{n+1}\eta_{n+1}^T|\mathcal{F}_n] = \Sigma + \Delta_n$ *where* $\mathbb{E}[\|\Delta_n\|\mathbb{1}_{\Omega_0}] \to 0$ *and* Σ *is a positive semidefinite matrix.*
iv) *The sequence* (b_n) *is the sum of two sequences* $(b_{n,1})$ *and* $(b_{n,2})$*, adapted to* $\mathcal{F}$*, s.t.* $\sup_{n\geq 0} \mathbb{E}[\|b_{n,1}\|^2] < \infty$, $\mathbb{E}[\|b_{n,1}\|\mathbb{1}_{\Omega_0}] \to 0$ *and* $b_{n,2} \to 0$ *a.s. on* Ω_0.

Then, given Ω_0, (Z_n) *converges in distribution to the unique stationary distribution* $\mu_\star$ *of the generalized Ornstein-Uhlenbeck process*

$$dX_t = \bar{H}X_t dt + \sqrt{\Sigma}dB_t$$

where (B_t) *is the standard Brownian motion and* $\sqrt{\Sigma}$ *is the unique positive semidefinite square root of* Σ*. The distribution* $\mu_\star$ *is the zero mean Gaussian distribution with covariance matrix* Γ *given as the solution to* $(\bar{H} + \frac{\mathbb{1}_{\alpha=1}}{2\gamma_0} I_k)\Gamma + \Gamma(\bar{H} + \frac{\mathbb{1}_{\alpha=1}}{2\gamma_0} I_k)^T = -\Sigma$.

Proof.

The proof is identical to the proof of (Pelletier, 1998, Th. 7), only substituting the inverse of the square root of Σ by the Moore-Penrose inverse. Finally, the uniqueness of the stationary distribution $\mu_\star$ and its expression follow from (Karatzas and Shreve, 1991, Th. 6.7, p. 357). ■

We define $v_n = \bar{v}_n + \delta_n$ where $\delta_0 = 0$, $\bar{v}_0 = v_0$ and

$$\begin{aligned}
\delta_{n+1} &= (1 - \gamma_{n+1}q_n)\delta_n + \gamma_{n+1}(p_n - q_n q_\infty^{-1} p_\infty)S(x_n)\,, \\
\bar{v}_{n+1} &= (1 - \gamma_{n+1}q_n)\bar{v}_n + \gamma_{n+1}q_n q_\infty^{-1} p_\infty S(x_n) \\
&\quad + \gamma_{n+1}p_n(\nabla f(x_n, \xi_{n+1})^{\odot 2} - S(x_n))\,.
\end{aligned}$$

For every $z = (v, m, x) \in \mathcal{Z}_+$ and $\delta \geq 0$, we define

$$r_n(z, \delta) := \begin{bmatrix} q_n q_\infty^{-1} p_\infty (S(x - \gamma_n \frac{m}{\sqrt{v+\delta+\varepsilon}}) - S(x)) \\ h_n(\nabla F(x - \gamma_n \frac{m}{\sqrt{v+\delta+\varepsilon}}) - \nabla F(x)) \\ \frac{\gamma_n}{\gamma_{n+1}}(\frac{1}{\sqrt{v+\varepsilon}} - \frac{1}{\sqrt{v+\delta+\varepsilon}}) \odot m \end{bmatrix}.$$

Moreover, for every $z = (v, m, x) \in \mathcal{Z}_+$ and every $n \in \mathbb{N}$, we set

$$g_n(z) = \begin{bmatrix} q_n q_\infty^{-1} p_\infty S(x) - q_n v \\ h_n \nabla F(x) - r_n m \\ -\frac{\gamma_n}{\gamma_{n+1}} \frac{m}{\sqrt{v+\varepsilon}} \end{bmatrix}.$$

Defining $\zeta_n = (\bar{v}_n, m_n, x_{n-1})$ and recalling the definition of (η_n) from Eq. (3.28), we have the decomposition

$$\zeta_{n+1} = \zeta_n + \gamma_{n+1}g_n(\zeta_n) + \gamma_{n+1}\eta_{n+1} + \gamma_{n+1}r_n(\zeta_n, \delta_n)\,.$$

Define $z_\star := (x_\star, 0, v_\star)$. Note that $g_n(z_\star) = 0$. Evaluating the Jacobian matrix G_n of g_n at $z_\star$, we obtain that there exist constants $C > 0$, $\bar{M} > 0$ and $n_0 \in \mathbb{N}$ s.t. for all $n \geq n_0$,

$$\|g_n(z) - G_n(z - z_\star)\| \leq C\|z - z_\star\|^2 \quad (\forall z \in B(z_\star, \bar{M}))\,, \tag{3.51}$$

where G_n is given by

$$G_n := \begin{bmatrix} -q_n I_d & 0 & q_n q_\infty^{-1} p_\infty \nabla S(x_\star) \\ 0 & -r_n I_d & h_n \nabla^2 F(x_\star) \\ 0 & -\frac{\gamma_n}{\gamma_{n+1}} V & 0 \end{bmatrix},$$

where ∇S is the Jacobian of S and the matrix V is defined in Eq. (3.8). We define

$$G_\infty := \lim_n G_n = \begin{bmatrix} -q_\infty I_d & 0 & p_\infty \nabla S(x_\star) \\ 0 & -r_\infty I_d & h_\infty \nabla^2 F(x_\star) \\ 0 & -V & 0 \end{bmatrix}.$$

One can verify that G_∞ is Hurwitz, and that the largest real part of its eigenvalues is $-L'$, where $L' := L \wedge q_\infty$ and L is defined in Eq. (3.9).

We define $\Omega^{(0)} := \{z_n \to z_\star\}$. We assume $\mathbb{P}(\Omega^{(0)}) > 0$. Using for instance (Delyon et al., 1999, Lem. 4 and Lem. 5), it holds that $\delta_n(\omega) \to 0$ for every $\omega \in \Omega^{(0)}$, and since $x_n(\omega) - x_{n-1}(\omega) \to 0$ on that set, we obtain that $\Omega^{(0)} = \{\zeta_n \to z_\star\}$. Let $M \in (0, \bar{M})$ be a constant, whose value will be specified later on. For every $N_0 \in \mathbb{N}$, define $\Omega^{(0)}_{N_0} := \{\zeta_n \to z_\star \text{ and } \sup_{n \geq N_0} \|\zeta_n - z_\star\| \leq M\}$. We seek to show that $\sqrt{\gamma_n}^{-1}(\zeta_n - z_\star) \Rightarrow \nu$ given $\Omega^{(0)}$, for some Gaussian measure ν, using Th. 3.17. As $\Omega^{(0)}_{N_0} \uparrow \Omega^{(0)}$, it is sufficient to show that the latter convergence holds given $\Omega^{(0)}_{N_0}$, for every N_0 large enough. From now on, we consider that N_0 is fixed. We define the sequence $(\tilde{\zeta}_n)_{n \geq N_0}$ as $\tilde{\zeta}_{N_0} = \zeta_{N_0}$ and for every $n \geq N_0$,

$$\tilde{\zeta}_{n+1} = \tilde{\zeta}_n + \gamma_{n+1} \tilde{g}_n(\tilde{\zeta}_n) + \gamma_{n+1}(\eta_{n+1} + r_n(\tilde{\zeta}_n, \delta_n)) \mathbb{1}_{\mathcal{A}_n}$$

where $\mathcal{A}_n$ is the event defined by

$$\mathcal{A}_n := \bigcap_{k=N_0}^{n} \{\|x_k - x_\star\| \leq M\} \cap \{\|\tilde{\zeta}_n - z_\star\| \leq M\}$$

and

$$\tilde{g}_n(z) := g_n(z) \mathbb{1}_{\|z - z_\star\| \leq M} - K(z - z_\star) \mathbb{1}_{\|z - z_\star\| > M},$$

where $K > 0$ is a large constant which will be specified later on. The sequences $(\tilde{\zeta}_n)_{n \geq N_0}$ and $(\zeta_n)_{n \geq N_0}$ coincide on $\Omega^{(0)}_{N_0}$. Thus, it is sufficient to study the weak convergence of $(\tilde{\zeta}_n)_{n \geq N_0}$.

An estimate of $\|r_n(\tilde{\zeta}_n, \delta_n)\| \mathbb{1}_{\mathcal{A}_n}$. We start by studying the sequence $(\|\delta_n\| \mathbb{1}_{\mathcal{A}_n})$. Unfolding the update rule defining δ_n and using the fact that (q_n) is a sequence of positive reals converging to $q_\infty > 0$, we obtain that

$$\begin{aligned} \|\delta_n\| \mathbb{1}_{\mathcal{A}_n} &\leq \sum_{k=1}^{n} \left[\prod_{j=k+1}^{n} |1 - \gamma_j q_{j-1}| \right] \gamma_k |p_{k-1} - q_{k-1} q_\infty^{-1} p_\infty| \|S(x_{k-1})\| \mathbb{1}_{\mathcal{A}_n} \\ &\leq C \sum_{k=1}^{n} \exp\left(-\beta \sum_{j=k+1}^{n} \gamma_j \right) \gamma_k |p_{k-1} - q_{k-1} q_\infty^{-1} p_\infty| := w_n, \end{aligned}$$

for some $\beta > 0$. The sequence (w_n) is deterministic and converges to zero by (Delyon et al., 1999, Lem. 4). There exists $n_1 \geq n_0$ s.t. $w_n \leq M$. As $v \mapsto \frac{1}{\sqrt{v+\varepsilon}}$ is Lipschitz and ∇F and S are locally Lipschitz, for every $z = (v, m, x)$ and δ s.t. $\|z - z_\star\| \leq M$ and $\|\delta\| \leq M$, we have

$$\begin{aligned}\|r_n(z,\delta)\| &\leq C\gamma_{n+1}\|(v+\delta+\varepsilon)^{\odot -\frac{1}{2}}\|\|m\| \\ &\quad + C\|(v+\delta+\varepsilon)^{\odot -\frac{1}{2}} - (v+\varepsilon)^{\odot -\frac{1}{2}}\|\|m\| \\ &\leq C\gamma_{n+1}\|z - z_\star\| + C\|\delta\|\|z - z_\star\| \,.\end{aligned}$$

This implies that for every $n \geq n_1$,

$$\|r_n(\tilde{\zeta}_n, \delta_n)\|\mathbb{1}_{\mathcal{A}_n} \leq C(\gamma_{n+1} + w_n)\|\tilde{\zeta}_n - z_\star\| \,. \tag{3.52}$$

Tightness of $\sqrt{\gamma_n}^{-1}(\tilde{\zeta}_n - z_\star)$. We decompose

$$\begin{aligned}\tilde{\zeta}_{n+1} - z_\star &= (I_{3d} + \gamma_{n+1}G_n)(\tilde{\zeta}_n - z_\star) + \gamma_{n+1}\left(g_n(\tilde{\zeta}_n) - G_n(\tilde{\zeta}_n - z_\star)\right)\mathbb{1}_{\|\tilde{\zeta}_n - z_\star\| \leq M} \\ &\quad - \gamma_{n+1}(K + G_n)(\tilde{\zeta}_n - z_\star)\mathbb{1}_{\|\tilde{\zeta}_n - z_\star\| > M} + \gamma_{n+1}(\eta_{n+1} + r_n(\tilde{\zeta}_n, \delta_n))\mathbb{1}_{\mathcal{A}_n} \,. \end{aligned} \tag{3.53}$$

For a given $t > 0$, we write $G_\infty = B_t^{-1} G_t B_t$ the Jordan-like decomposition of G_∞, where the ones of the second diagonal of the usual Jordan decomposition are replaced by t, and where B_t is some invertible matrix. We define $S_n := B_t(\tilde{\zeta}_n - z_\star)$. Setting $G_n^{(t)} := B_t G_n B_t^{-1}$, we obtain

$$\begin{aligned}S_{n+1} &= (I_{3d} + \gamma_{n+1}G_n^{(t)})S_n + \gamma_{n+1}B_t\left(g_n(\tilde{\zeta}_n) - G_n(\tilde{\zeta}_n - z_\star)\right)\mathbb{1}_{\|\tilde{\zeta}_n - z_\star\| \leq M} \\ &\quad - \gamma_{n+1}(K + G_n^{(t)})S_n\mathbb{1}_{\|\tilde{\zeta}_n - z_\star\| > M} + \gamma_{n+1}B_t(\eta_{n+1} + r_n(\tilde{\zeta}_n, \delta_n))\mathbb{1}_{\mathcal{A}_n} \,.\end{aligned}$$

Choose $A \in (0, 2L')$ and $A' \in (A, 2L')$. There exists $\bar{\gamma}$ and $t > 0$ s.t. for every $\gamma < \bar{\gamma}$, $\|I + \gamma G_t\|_2 \leq 1 - \gamma(A' + 2L')/2$, where $\|\cdot\|_2$ is the spectral norm. As $G_n^{(t)} \to G^t$, there exists $n_2 \geq n_1$, such that for all $n \geq n_2$, $\|I + \gamma G_n^{(t)}\|_2 \leq 1 - \gamma A'$. Recall the notation $\mathbb{E}_n = \mathbb{E}[\cdot \,|\, \mathcal{F}_n]$. We expand $\|S_{n+1}\|^2$ and use the inequality $\left\|g_n(\tilde{\zeta}_n) - G_n(\tilde{\zeta}_n - z_\star)\right\|^2 \mathbb{1}_{\|\tilde{\zeta}_n - z_\star\| \leq M} \leq C\|S_n\|^2$ to obtain after straightforward algebra

$$\begin{aligned}\mathbb{E}_n\|S_{n+1}\|^2 &\leq (1 - \gamma_{n+1}A')\|S_n\|^2 + C\gamma_{n+1}^2\|S_n\|^2 \\ &\quad + C\gamma_{n+1}^2(\mathbb{E}_n\|\eta_{n+1}\|^2 + \|r_n(\tilde{\zeta}_n, \delta_n)\|^2)\mathbb{1}_{\mathcal{A}_n} \\ &\quad + 2\gamma_{n+1}S_n^* B_t\left(g_n(\tilde{\zeta}_n) - G_n(\tilde{\zeta}_n - z_\star)\right)\mathbb{1}_{\|\tilde{\zeta}_n - z_\star\| \leq M} \\ &\quad - 2\gamma_{n+1}S_n^*(K + G_n^{(t)})S_n\mathbb{1}_{\|\tilde{\zeta}_n - z_\star\| > M} + 2\gamma_{n+1}S_n^* B_t r_n(\tilde{\zeta}_n, \delta_n)\mathbb{1}_{\mathcal{A}_n} \,.\end{aligned}$$

Choose $c := (A' - A)/2$. If M is chosen small enough,

$$\|g_n(\tilde{\zeta}_n) - G_n(\tilde{\zeta}_n - z_\star)\|\mathbb{1}_{\|\tilde{\zeta}_n - z_\star\| \leq M} \leq \frac{c}{2}\|B_t\|^{-1}\|B_t^{-1}\|\|\tilde{\zeta}_n - z_\star\| \,.$$

Moreover, choosing $K > \sup_n \|G_n^{(t)}\|_2$, it holds that $S_n^*(K + G_n^{(t)})S_n \geq 0$. Then,

$$\begin{aligned}\mathbb{E}_n\|S_{n+1}\|^2 &\leq (1 - \gamma_{n+1}(A' - c))\|S_n\|^2 + C\gamma_{n+1}^2\|S_n\|^2 \\ &\quad + C\gamma_{n+1}^2(\mathbb{E}_n\|\eta_{n+1}\|^2 + \|r_n(\tilde{\zeta}_n, \delta_n)\|^2)\mathbb{1}_{\mathcal{A}_n} + 2\gamma_{n+1}\|B_t\|\|S_n\|\|r_n(\tilde{\zeta}_n, \delta_n)\|\mathbb{1}_{\mathcal{A}_n} \,.\end{aligned}$$

Using Eq. (3.52),

$$\mathbb{E}_n\|S_{n+1}\|^2 \le (1-\gamma_{n+1}(A'-c-w_n))\|S_n\|^2 + C\gamma_{n+1}^2(1+w_n^2)\|S_n\|^2 + C\gamma_{n+1}^2\mathbb{E}_n\|\eta_{n+1}\|^2\mathbb{1}_{\mathcal{A}_n}\,.$$

Therefore, there exists $n_3 \ge n_2$ s.t. for all $n \ge n_3$,

$$\mathbb{E}\|S_{n+1}\|^2 \le (1-\gamma_{n+1}A)\mathbb{E}\|S_n\|^2 + C\gamma_{n+1}^2\mathbb{E}(\|\eta_{n+1}\|^2\mathbb{1}_{\|x_n-x_\star\|\le M})\,.$$

The second expectation in the righthand side is bounded uniformly in n by the condition (3.7). Using (Delyon et al., 1999, Lem. 4 and Lem. 5), we conclude that $\sup_n \gamma_n^{-1}\mathbb{E}\|S_n\|^2 < \infty$. Therefore, $\sup_n \gamma_n^{-1}\mathbb{E}\|\tilde\zeta_n - z_\star\|^2 < \infty$, which in turn implies $\sup_n \gamma_n^{-1}\mathbb{E}(\|\zeta_n - z_\star\|^2\mathbb{1}_{\Omega_{N_0}^{(0)}}) < \infty$.

Strongly perturbed iterations. We define $\tilde y_n = \sqrt{\gamma_n}^{-1}(\tilde\zeta_n - z_\star)$. Define

$$\bar G_n := \gamma_{n+1}^{-1}\left(\sqrt{\frac{\gamma_n}{\gamma_{n+1}}} - 1\right) I_{3d} + \sqrt{\frac{\gamma_n}{\gamma_{n+1}}}G_n\,.$$

The sequence $\bar G_n$ converges to $\bar G_\infty := G_\infty + \frac{\mathbb{1}_{\alpha=1}}{2\gamma_0}I_{3d}$. Recalling Eq. (3.53), we can write

$$\tilde y_{n+1} = (I_{3d} + \gamma_{n+1}\bar G_\infty)\tilde y_n + \gamma_{n+1}\bar r_n + \sqrt{\gamma_{n+1}}\bar\eta_{n+1}$$

where $\bar\eta_{n+1} = \eta_{n+1}\mathbb{1}_{\mathcal{A}_n}$ and $\bar r_n = \bar r_{n,1} + \bar r_{n,2} + \bar r_{n,3}$, where

$$\begin{aligned}
\bar r_{n,1} &:= \sqrt{\gamma_{n+1}}^{-1} r_n(\tilde\zeta_n, \delta_n)\mathbb{1}_{\mathcal{A}_n} + (\bar G_n - \bar G_\infty)\tilde y_n \\
\bar r_{n,2} &:= \sqrt{\gamma_{n+1}}^{-1}\left(g_n(\tilde\zeta_n) - G_n(\tilde\zeta_n - z_\star)\right)\mathbb{1}_{\|\tilde\zeta_n - z_\star\|\le M} \\
\bar r_{n,3} &:= -\sqrt{\gamma_{n+1}}^{-1}(K + G_n)(\tilde\zeta_n - z_\star)\mathbb{1}_{\|\tilde\zeta_n - z_\star\| > M}\,.
\end{aligned}$$

We now check that the assumptions of Th. 3.17 are fulfilled. On the event $\Omega_{N_0}^{(0)}$, we recall that $\tilde\zeta_n = \zeta_n$, hence $\bar r_{n,3}$ is identically zero. Moreover, using Eq. (3.52), it holds that for all n large enough,

$$\|\bar r_{n,1}\| \le C\left(\sqrt{\frac{\gamma_n}{\gamma_{n+1}}}(\gamma_{n+1} + w_n) + \|\bar G_n - \bar G_\infty\|\right)\|\tilde y_n\|$$

and therefore, $\mathbb{E}[\|\bar r_{n,1}\|^2] \to 0$. Now consider the term $\bar r_{n,2}$. By Eq. (3.51),

$$\|\bar r_{n,2}\| \le C\sqrt{\gamma_{n+1}}^{-1}\|\tilde\zeta_n - z_\star\|^2\mathbb{1}_{\|\tilde\zeta_n - z_\star\|\le M}\,.$$

Thus, $\|\bar r_{n,2}\|^2 \le C\|\tilde y_n\|^2$ which implies that $\sup_{n\ge N_0}\mathbb{E}[\|r_{n,2}\|^2] < \infty$. Moreover, $\mathbb{E}[\|\bar r_{n,2}\|] \le C\sqrt{\gamma_{n+1}}\mathbb{E}\|\tilde y_n\|^2$ tends to zero. Finally, consider $\bar\eta_{n+1}$. Using condition (3.7), there exist $M > 0$ and $b_M > 4$ s.t.

$$\begin{aligned}
\mathbb{E}_n[\|\bar\eta_{n+1}\|^{b_M/2}] &\le \mathbb{E}_n[\|\eta_{n+1}\|^{b_M/2}]\mathbb{1}_{\|x_n - x_\star\|\le M} \\
&\le C\mathbb{E}_n[\|\nabla f(x_n, \xi_{n+1})\|^{b_M}]\mathbb{1}_{\|x_n - x_\star\|\le M} \le C\,.
\end{aligned}$$

Moreover, $\mathbb{E}_n[\bar{\eta}_{n+1}] = 0$ and finally, almost surely on $\Omega_N^{(0)}$, $\mathbb{E}_n[\bar{\eta}_{n+1}\bar{\eta}_{n+1}^{\mathrm{T}}]$ converges to

$$\Sigma := \begin{bmatrix} \mathbb{E}_\xi \left[\begin{bmatrix} p_\infty(\nabla f(x_\star,\xi)^{\odot 2} - S(x_\star)) \\ h_\infty \nabla f(x_\star,\xi) \end{bmatrix} \begin{bmatrix} p_\infty(\nabla f(x_\star,\xi)^{\odot 2} - S(x_\star)) \\ h_\infty \nabla f(x_\star,\xi) \end{bmatrix}^{\mathrm{T}} \right] & 0 \\ 0 & 0 \end{bmatrix} . \tag{3.54}$$

Therefore, the assumptions of Th. 3.17 are fulfilled for the sequence $\tilde{y}_n$. We obtain the desired result for the sequence (m_n, x_{n-1}). We now show that the same result also holds for the sequence (m_n, x_n). For this purpose, observe that

$$\frac{1}{\sqrt{\gamma_n}} \begin{bmatrix} m_n \\ x_n - x_\star \end{bmatrix} = \frac{1}{\sqrt{\gamma_n}} \begin{bmatrix} m_n \\ x_{n-1} - x_\star \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{\sqrt{\gamma_n}}(x_n - x_{n-1}) \end{bmatrix} .$$

Then, notice that $\|\frac{x_n - x_{n-1}}{\sqrt{\gamma_n}}\| = \sqrt{\gamma_n}\|\frac{m_n}{\sqrt{v_n+\varepsilon}}\| \le \sqrt{\frac{\gamma_n}{\varepsilon}}\|m_n\| \to 0$ as $n \to \infty$ since it is assumed that $z_n \to z_\star$ (which implies in particular that $m_n \to 0$). Hence, it holds that $\sqrt{\gamma_n}^{-1}(x_n - x_{n-1})$ converges a.s. to 0. We conclude by invoking Slutsky's lemma.

Proof of Eq. (3.10). We have the subsystem:

$$\tilde{H}\Gamma + \Gamma\tilde{H}^{\mathrm{T}} = \begin{bmatrix} -h_\infty^2 \mathcal{Q} & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{where } \tilde{H} := \begin{bmatrix} (\theta - r_\infty) I_d & h_\infty \nabla^2 F(x_\star) \\ -V & \theta I_d \end{bmatrix} \tag{3.55}$$

and where $\mathcal{Q} := \mathrm{Cov}\Big(\nabla f(x_\star,\xi)\Big)$. The next step is to triangularize the matrix $\tilde{H}$ in order to decouple the blocks of Γ. For every $k = 1, \dots, d$, set $\nu_k^\pm := -\frac{r_\infty}{2} \pm \sqrt{r_\infty^2/4 - h_\infty \pi_k}$ with the convention that $\sqrt{-1} = \imath$ (inspecting the characteristic polynomial of H, these are the eigenvalues of H). Set $M^\pm := \mathrm{diag}\,(\nu_1^\pm, \cdots, \nu_d^\pm)$ and $R^\pm := V^{-\frac{1}{2}} P M^\pm P^{\mathrm{T}} V^{-\frac{1}{2}}$. Using the identities $M^+ + M^- = -r_\infty I_d$ and $M^+ M^- = h_\infty \,\mathrm{diag}\,(\pi_1, \cdots, \pi_d)$, it can be checked that

$$\mathcal{R}\tilde{H} = \begin{bmatrix} R^- V + \theta I_d & 0 \\ -V & V R^+ + \theta I_d \end{bmatrix} \mathcal{R}, \text{ where } \mathcal{R} := \begin{bmatrix} I_d & R^+ \\ 0 & I_d \end{bmatrix} .$$

Set $\tilde{\Gamma} := \mathcal{R}\Gamma\mathcal{R}^{\mathrm{T}}$. Denote by $(\tilde{\Gamma}_{i,j})_{i,j=1,2}$ the blocks of $\tilde{\Gamma}$. Note that $\tilde{\Gamma}_{2,2} = \Gamma_{2,2}$. By left/right multiplication of Eq. (3.55) respectively by $\mathcal{R}$ and $\mathcal{R}^{\mathrm{T}}$, we obtain

$$(R^- V + \theta I_d)\tilde{\Gamma}_{1,1} + \tilde{\Gamma}_{1,1}(V R^- + \theta I_d) = -h_\infty^2 \mathcal{Q} \tag{3.56}$$

$$(R^- V + \theta I_d)\tilde{\Gamma}_{1,2} + \tilde{\Gamma}_{1,2}(R^+ V + \theta I_d) = \tilde{\Gamma}_{1,1} V \tag{3.57}$$

$$(V R^+ + \theta I_d)\tilde{\Gamma}_{2,2} + \tilde{\Gamma}_{2,2}(R^+ V + \theta I_d) = V\tilde{\Gamma}_{1,2} + \tilde{\Gamma}_{1,2}^{\mathrm{T}} V . \tag{3.58}$$

Set $\bar{\Gamma}_{1,1} = P^{-1} V^{\frac{1}{2}} \tilde{\Gamma}_{1,1} V^{\frac{1}{2}} P$. Define $C := P^{-1} V^{\frac{1}{2}} \mathcal{Q} V^{\frac{1}{2}} P$. Eq. (3.56) yields

$$(M^- + \theta I_d)\bar{\Gamma}_{1,1} + \bar{\Gamma}_{1,1}(M^- + \theta I_d) = -h_\infty^2 C .$$

Set $\bar{\Gamma}_{1,2} = P^{-1} V^{\frac{1}{2}} \tilde{\Gamma}_{1,2} V^{-\frac{1}{2}} P$. Eq. (3.57) is rewritten $(M^- + \theta I_d)\bar{\Gamma}_{1,2} + \bar{\Gamma}_{1,2}(M^+ + \theta I_d) = \bar{\Gamma}_{1,1}$. The component (k, ℓ) is given by

$$\bar{\Gamma}_{1,2}^{k,\ell} = (\nu_k^- + \nu_\ell^+ + 2\theta)^{-1} \bar{\Gamma}_{1,1}^{k,\ell} = \frac{-h_\infty^2 C_{k,\ell}}{(\nu_k^- + \nu_\ell^+ + 2\theta)(\nu_k^- + \nu_\ell^- + 2\theta)} .$$

Set finally $\bar{\Gamma}_{2,2} = P^{-1}V^{-\frac{1}{2}}\Gamma_{2,2}V^{-\frac{1}{2}}P$. Eq. (3.58) becomes

$$(M^+ + \theta I_d)\bar{\Gamma}_{2,2} + \bar{\Gamma}_{2,2}(M^+ + \theta I_d) = \bar{\Gamma}_{1,2} + \bar{\Gamma}_{1,2}^{\mathrm{T}}\,.$$

Thus,

$$\begin{aligned}\bar{\Gamma}_{2,2}^{k,\ell} &= \frac{\bar{\Gamma}_{1,2}^{k,\ell} + \bar{\Gamma}_{1,2}^{\ell,k}}{\nu_k^+ + \nu_\ell^+ + 2\theta} \\ &= \frac{-h_\infty^2 C_{k,\ell}}{(\nu_k^+ + \nu_\ell^+ + 2\theta)(\nu_k^- + \nu_\ell^- + 2\theta)}\left(\frac{1}{\nu_k^- + \nu_\ell^+ + 2\theta} + \frac{1}{\nu_k^+ + \nu_\ell^- + 2\theta}\right).\end{aligned}$$

After tedious but straightforward computations, we obtain

$$\bar{\Gamma}_{2,2}^{k,\ell} = \frac{h_\infty^2 C_{k,\ell}}{(r_\infty - 2\theta)(h_\infty(\pi_k + \pi_\ell) + 2\theta(\theta - r_\infty)) + \frac{h_\infty^2(\pi_k - \pi_\ell)^2}{2(r_\infty - 2\theta)}}\,,$$

and the result is proved.

3.7 Proofs for Section 3.4

3.7.1 Preliminaries

Most of the avoidance of traps results in the stochastic approximation literature deal with the case where the ODE that underlies the stochastic algorithm under study is an autonomous ODE $\dot{\mathsf{z}} = h(\mathsf{z})$. In this setting, a point $z_\star \in \operatorname{zer} h$ is called a trap if $h(z)$ admits an expansion around $z_\star$ of the type $h(z) = D(z - z_\star) + o(\|z - z_\star\|)$, where the matrix D has at least one eigenvalue which real part is (strictly) positive. Initiated by Pemantle (1990) and by Brandière and Duflo (1996), the most powerful class of techniques for establishing avoidance of traps results makes use of Poincaré's invariant manifold theorem for the ODE $\dot{\mathsf{z}} = h(\mathsf{z})$ in a neighborhood of some point $z_\star \in \operatorname{zer} h$. The idea is to show that with probability 1, the stochastic algorithm strays away from the maximal invariant manifold of the ODE where the convergence to $z_\star$ of the ODE flow can take place. As previously mentioned, since we are dealing with algorithms derived from non-autonomous ODEs, we extend the results of Pemantle (1990); Brandière and Duflo (1996) to this setting. The proof of Th. 3.8 relies on a non-autonomous version of Poincaré's theorem. We borrow this result from the rich literature that exists on the subject (Daleckii and Krein, 1974; Kloeden and Rasmussen, 2011).

Let us start by setting the context for the non-autonomous version that we shall need for the invariant manifold theorem. Given an integer $d > 0$ and a matrix $D \in \mathbb{R}^{d\times d}$, consider the linear autonomous differential equation

$$\dot{\mathsf{z}}(t) = D\mathsf{z}(t), \tag{3.59}$$

which solution is obviously $\mathsf{z}(t) = e^{Dt}\mathsf{z}(0)$ for $t \in \mathbb{R}$. Let us factorize D as in (3.12), and write $D = Q\Lambda Q^{-1}$ with $\Lambda = \begin{bmatrix}\Lambda^- & \\ & \Lambda^+\end{bmatrix}$ where we recall that the Jordan blocks that constitute $\Lambda^- \in \mathbb{R}^{d^-\times d^-}$ (respectively $\Lambda^+ \in \mathbb{R}^{d^+\times d^+}$) are those that contain the eigenvalues λ_i of D such that $\Re\lambda_i \le 0$ (respectively $\Re\lambda_i > 0$). Let us assume here that

both d^- and d^+ are positive. It will be convenient to work in the basis of the columns of Q by making the variable change

$$z \mapsto y = \begin{bmatrix} y^- \\ y^+ \end{bmatrix} = Q^{-1} z,$$

where $y^\pm \in \mathbb{R}^{d^\pm}$. In this new basis, the ODE (3.59) is written as

$$\begin{bmatrix} \dot{\mathsf{y}}^- \\ \dot{\mathsf{y}}^+ \end{bmatrix} = \begin{bmatrix} \Lambda^- & \\ & \Lambda^+ \end{bmatrix} \begin{bmatrix} \mathsf{y}^- \\ \mathsf{y}^+ \end{bmatrix}, \tag{3.60}$$

which solution is $\mathsf{y}^\pm(t) = \exp(t\Lambda^\pm)\mathsf{y}^\pm(0)$. One can readily check that for each couple of real numbers α^+ and α^- that satisfy

$$0 < \alpha^- < \alpha^+ < \min\{\Re\lambda_i \,:\, \Re\lambda_i > 0\}, \tag{3.61}$$

there exists a so-called exponential dichotomy of the ODE solutions, which amounts in our case to the existence of two constants $K^-, K^+ \geq 1$ such that

$$\begin{aligned} \|\exp(t\Lambda^-)\| &\leq K^- e^{\alpha^- t} \quad \text{for } t \geq 0, \\ \|\exp(t\Lambda^+)\| &\leq K^+ e^{\alpha^+ t} \quad \text{for } t \leq 0, \end{aligned}$$

see, *e.g.*, Horn and Johnson (1994).

We now consider a non-autonomous perturbation of this ODE, which is represented in the basis of the columns of Q as

$$\dot{\mathsf{y}}(t) = h(\mathsf{y}(t), t) \quad \text{with} \quad h(y,t) = \begin{bmatrix} \Lambda^- & \\ & \Lambda^+ \end{bmatrix} y + \varepsilon(y,t), \tag{3.62}$$

and $\varepsilon : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ is a continuous function. In the sequel, we shall be interested in the asymptotic behavior of this equation for the large values of t, and therefore, restrict our study to the interval $\mathbb{I} = [t_0, \infty)$ for some given $t_0 \geq 0$ that we shall fix later. We assume that $\varepsilon(0, \cdot) = 0$ on $\mathbb{I}$. We denote as $\phi : \mathbb{I} \times \mathbb{I} \times \mathbb{R}^d \to \mathbb{R}^d$ the so-called general solution of (3.62), which is defined by the fact that $\phi(\cdot, t, x)$ is the unique noncontinuable solution of (3.62) such that $\phi(t, t, x) = x$ for $t \in \mathbb{I}$ and $x \in \mathbb{R}^d$, assuming this solution exists and is unique for each $(x, t) \in \mathbb{R}^d \times \mathbb{I}$.

In the linear autonomous case provided by the ODE (3.60), the subspace

$$\mathcal{G} = \left\{ \left(t, \begin{bmatrix} y^- \\ 0 \end{bmatrix} \right) \in \mathbb{R} \times \mathbb{R}^d \;:\; y^- \in \mathbb{R}^{d^-} \right\}$$

is trivially invariant in the sense that if $(t, y) \in \mathcal{G}$, then, $(s, \phi(s, t, y)) \in \mathcal{G}$ for each $s \in \mathbb{R}$. This concept can be generalized to the non-linear and non-autonomous case. We say that the $\mathcal{C}^1$ function $w : \mathbb{R}^{d^-} \times \mathbb{I} \to \mathbb{R}^{d^+}$ defines a global non-autonomous invariant manifold for the ODE (3.62) if $w(0, t) = 0$ for all $t \in \mathbb{I}$, and, furthermore, if for each $t \in \mathbb{I}$ and each $y^- \in \mathbb{R}^{d^-}$, writing $y = (y^-, w(y^-, t))$, the general solution $\phi(s, t, y) = (\phi^-(s, t, y), \phi^+(s, t, y))$ with $\phi^\pm(s, t, y) \in \mathbb{R}^{d^\pm}$ verifies $\phi^+(s, t, y) = w(\phi^-(s, t, y), s)$ for each $s \in \mathbb{I}$. The non-autonomous invariant manifold is the set

$$\mathcal{G} = \left\{ \left(t, \begin{bmatrix} y^- \\ w(y^-, t) \end{bmatrix} \right) \in \mathbb{I} \times \mathbb{R}^d \;:\; y^- \in \mathbb{R}^{d^-} \right\},$$

which obviously satisfies $(t,y) \in \mathcal{G} \;\Rightarrow\; (s,\phi(s,t,y)) \in \mathcal{G}$ for each $s \in \mathbb{I}$.

These invariant manifolds are described by the following proposition, which is a straightforward application of (Pötzsche and Rasmussen, 2006, Th. A.1) (see also (Kloeden and Rasmussen, 2011, Th. 6.3 p. 106, Rem. 6.6 p. 111)). It is useful to note that under the conditions provided in the statement of this proposition, the existence of the general solution ϕ of the ODE (3.62) is ensured by Picard's theorem.

Proposition 3.18. Let $\mathbb{I} = [t_0, \infty)$ for some $t_0 \geq 0$. Assume that the function $\varepsilon(y,t)$ is such that $\varepsilon(0,\cdot) \equiv 0$ on $\mathbb{I}$, the function $\varepsilon(\cdot,t)$ is continuously differentiable for each $t \in \mathbb{I}$, and furthermore, the Jacobian matrix $\partial_1\varepsilon(y,t)$ satisfies

$$|\varepsilon|_1 := \sup_{(y,t)\in\mathbb{R}^d\times\mathbb{I}} \|\partial_1\varepsilon(y,t)\| < \frac{\alpha^+ - \alpha^-}{4K} \tag{3.63}$$

with $K = K^- + K^+ + K^-K^+(K^- \vee K^+)$ and α^-, α^+ chosen as in Eq. (3.61). Then, for each $\delta \in (2K|\varepsilon|_1, (\alpha^+ - \alpha^-)/2)$ and each $\gamma \in (\alpha^- + \delta, \alpha^+ - \delta)$, the set

$$\mathcal{G} = \left\{ (t,y) \in \mathbb{I}\times\mathbb{R}^d \;:\; \sup_{s\geq t} \|\phi(s,t,y)\| \exp(\gamma(t-s)) < \infty \right\}$$

is nonempty, and does not depend on γ. Moreover, this set is a global invariant manifold for the ODE (3.62) that is defined by a continuously differentiable mapping $w : \mathbb{R}^{d^-} \times \mathbb{I} \to \mathbb{R}^{d^+}$. In addition, if the partial derivatives $\partial_1^k\varepsilon : \mathbb{R}^d \times \mathbb{I}$ exist and are continuous for $k \in \{1,\dots,m\}$ with globally bounded partial derivatives

$$|\varepsilon|_k := \sup_{(y,t)\in\mathbb{R}^d\times\mathbb{I}} \|\partial_1^k\varepsilon(y,t)\| < \infty\,, \tag{3.64}$$

under the gap condition

$$m\alpha^- < \alpha^+,\; m \in \mathbb{N}^*, \tag{3.65}$$

the partial derivatives $\partial_1^k w : \mathbb{R}^{d^-} \times \mathbb{I}$ exist and are continuous with

$$\sup_{(y^-,t)\in\mathbb{R}^{d^-}\times\mathbb{I}} \|\partial_1^k w(y^-,t)\| < \infty \quad \text{for all } k \in \{1,\dots,m\}. \tag{3.66}$$

Finally, if $\partial_2^n\partial_1^k\varepsilon$ exist and are continuous for $0 \leq n < m$ and $0 \leq k+n \leq m$, then w is m-times continuously differentiable.

Let us partition the function $h(y,t)$ as

$$h(y,t) = \begin{bmatrix} h^-(y,t) \\ h^+(y,t) \end{bmatrix} = \begin{bmatrix} \Lambda^- y^- + \varepsilon^-(y,t) \\ \Lambda^+ y^+ + \varepsilon^+(y,t) \end{bmatrix}, \tag{3.67}$$

where $h^\pm : \mathbb{R}^d \times \mathbb{I} \to \mathbb{R}^{d^\pm}$, $y^\pm \in \mathbb{R}^{d^\pm}$ and $\varepsilon^\pm : \mathbb{R}^d \times \mathbb{I} \to \mathbb{R}^{d^\pm}$. With these notations, the previous proposition leads to the following lemma.

Lemma 3.19. In the setting of Prop. 3.18, for each t in the interior of $\mathbb{I}$ and each vector $y = (y^-, y^+)$ such that $y^\pm \in \mathbb{R}^{d^\pm}$ and $y^+ = w(y^-,t)$, it holds that

$$h^+(y,t) = \partial_1 w(y^-,t) h^-(y,t) + \partial_2 w(y^-,t)\,. \tag{3.68}$$

Assume that α^- is small enough so that Ineq. (3.65) and Eq. (3.64) hold true with $m = 2$. Assume in addition that $\partial_2^n \partial_1^k \varepsilon$ exists and is continuous for $0 \leq n < 2$ and $0 \leq k + n \leq 2$, and furthermore, that there exists a bounded neighborhood $\mathcal{V} \subset \mathbb{R}^d$ of zero such that

$$\sup_{(y,t)\in\mathcal{V}\times\mathbb{I}} \left\|\partial_2 \varepsilon(y,t)\right\| < +\infty. \tag{3.69}$$

Then, there exists a neighborhood $\mathcal{V}^- \subset \mathbb{R}^{d^-}$ of zero such that

$$\sup_{(y^-,t)\in\mathcal{V}^-\times\mathbb{I}} \left\|\partial_1\partial_2 w(y^-,t)\right\| < +\infty\,, \tag{3.70}$$

$$\sup_{(y^-,t)\in\mathcal{V}^-\times\mathbb{I}} \left\|\partial_2^2 w(y^-,t)\right\| < +\infty\,. \tag{3.71}$$

Proof. By Prop. 3.18, the general solution $\phi(s,t,y)$ of the ODE (3.62) can be written as $\phi(s,t,y) = (\phi^-(s,t,y), \phi^+(s,t,y))$ with $\phi^+(s,t,y) = w(\phi^-(s,t,y),s)$ for each $s \in \mathbb{I}$. Equating the derivatives with respect to s of the two members of this equation and taking $s = t$, we get the first equation.

Writing $g : \mathbb{R}^{d^-} \times \mathbb{I} \to \mathbb{R}^d, (y^-,t) \mapsto (y^-, w(y^-,t))$, Eq. (3.68) can be rewritten as

$$\partial_2 w(y^-,t) = h^+(g(y^-,t),t) - \partial_1 w(y^-,t) h^-(g(y^-,t),t). \tag{3.72}$$

By Prop. 3.18, the function w is twice differentiable, and we can write

$$\partial_2^2 w(y^-,t) = \partial_1 h^+ \partial_2 g + \partial_2 h^+ - (\partial_1\partial_2 w)h^- - (\partial_1 w)(\partial_1 h^- \partial_2 g + \partial_2 h^-), \tag{3.73}$$

where, *e.g.*, h^+ is a shorthand notation for $h^+(g(y^-,t),t)$. It holds from Eq. (3.67) and the assumptions of Prop. 3.18 that for each $(y,t) \in \mathbb{R}^d \times \mathbb{I}$,

$$\|\partial_1 h(y,t)\| \leq \|\Lambda\| + \|\partial_1 \varepsilon(y,t)\| \leq C, \tag{3.74}$$

where the constant $C > 0$ is independent of (y,t) and can change from an inequality to another in the remainder of the proof. By the mean value inequality and Prop. 3.18, we also get that

$$\|w(y^-,t)\| = \|w(y^-,t) - w(0,t)\| \leq \sup_{(u,s)} \|\partial_1 w(u,s)\|\ \|y^-\| \leq C\|y^-\|,$$

thus, $\|g(y^-,t)\| \leq C\|y^-\|$. By the mean value inequality again,

$$\begin{aligned}\left\|h(g(y^-,t),t)\right\| = \left\|h(g(y^-,t),t) - h(0,t)\right\| &\leq \sup_{(u,t)} \left\|\partial_1 h(u,t)\right\| \left\|g(y^-,t)\right\| \\ &\leq C\left\|g(y^-,t)\right\| \leq C\|y^-\|.\end{aligned}$$

By Eq. (3.72) and Prop. 3.18, this implies that

$$\left\|\partial_2 g(y^-,t)\right\| = \left\|\partial_2 w(y^-,t)\right\| = \left\|h^+ - (\partial_1 w)h^-\right\| \leq C\left\|y^-\right\|, \quad \text{and} \tag{3.75}$$

$$\left\|\partial_1\partial_2 w(y^-,t)\right\| = \left\|\partial_1 h^+ \partial_1 g - (\partial_1^2 w)h^- - (\partial_1 w)(\partial_1 h^- \partial_1 g)\right\| \leq C(\left\|y^-\right\| + 1). \tag{3.76}$$

Let $\mathcal{V}^- \subset \mathbb{R}^{d^-}$ be a small enough neighborhood of zero so that $g(y^-,t) \in \mathcal{V}$ for each $y^- \in \mathcal{V}^-$, which is possible by the inequality $\|g(y^-,t)\| \leq C\|y^-\|$. By the assumption on $\|\partial_2 \varepsilon(y,t)\|$ in the statement of Lem. 3.19, we have

$$\forall y^- \in \mathcal{V}^-, \quad \left\|\partial_2 h(g(y^-,t),t)\right\| = \left\|\partial_2 \varepsilon(g(y^-,t),t)\right\| \leq C. \tag{3.77}$$

The bound (3.70) is an immediate consequence of Eq. (3.76). Getting back to Eq. (3.73), the bound (3.71) follows from the inequalities (3.74)–(3.77). ■

Prop. 3.18 deals with the case where the function ε is globally Lipschitz continuous. In practical cases, such a strong assumption is not necessarily verified. In particular, for the ODEs we consider for our application, it is not satisfied (see the function e defined in Subsec. 3.7.3.1 below). Nonetheless, recall that we only need the existence of a *local* non-autonomous invariant manifold, i.e. defined in the vicinity of an arbitrary solution such as the trivial zero solution (since we suppose here $\varepsilon(0,\cdot)=0$) whereas the aforementioned strong assumption provides a global non-autonomous invariant manifold. Indeed, as for the avoidance of traps result we intend to show, we will only need to look at the behavior of our ODE in the neighborhood of a trap $z_\star$. Therefore, in prevision of the proof of Th. 3.8, we localize the ODE (3.62) in the neighborhood of zero. This is the purpose of the next proposition.

Proposition 3.20. Let $\mathbb{I}=[t_0,+\infty)$ for some $t_0\geq 0$ and let $h:\mathbb{R}^d\times\mathbb{I}\to\mathbb{R}^d$ be defined as in Eq. (3.62). Assume that $\varepsilon(0,\cdot)\equiv 0$ on $\mathbb{I}$, that the function $\varepsilon(\cdot,t)$ is continuously differentiable for every $t\in\mathbb{I}$ and that

$$\lim_{(y,t)\to(0,+\infty)} \left\|\partial_1\varepsilon(y,t)\right\| = 0\,. \tag{3.78}$$

Then, there exist $\sigma>0, t_1>0$, a function $\tilde{\varepsilon}:\mathbb{R}^d\times\mathbb{I}_1\to\mathbb{R}^d$ where $\mathbb{I}_1:=[t_1,+\infty)$ and a function $\tilde{h}:\mathbb{R}^d\times\mathbb{I}_1\to\mathbb{R}^d$ defined for every $y\in\mathbb{R}^d, t\in\mathbb{I}_1$ by $\tilde{h}(y,t)=\Lambda y+\tilde{\varepsilon}(y,t)$ s.t. $\tilde{h}$ and $\tilde{\varepsilon}$ verify the assumptions of Prop. 3.18 and for every $(y,t)\in B(0,\sigma)\times\mathbb{I}_1$, we have that $\tilde{h}(y,t)=h(y,t)$ and $\tilde{\varepsilon}(y,t)=\varepsilon(y,t)$. Moreover, for any $\delta>0$, we can choose σ, t_1 respectively small and large enough s.t. the mapping $w:\mathbb{R}^{d^-}\times\mathbb{I}_1\to\mathbb{R}^{d^+}$ obtained from Prop. 3.18 (applied to $\tilde{h}$ and $\tilde{\varepsilon}$) satisfies

$$|w|_1 = \sup_{(y,t)\in\mathbb{R}^{d^-}\times\mathbb{I}_1} \|\partial_1 w(y,t)\| < \delta\,. \tag{3.79}$$

Furthermore, Eq. (3.68) holds for $\tilde{h}$ and w for all $(y,t)\in B(0,\sigma)\times\mathbb{I}_1$. If, additionally, Eq. (3.69) holds for ε, then there exists $\sigma_1\leq\sigma$ such that

$$\sup_{(y^-,t)\in B(0,\sigma_1)\times\mathbb{I}_1} \left\|\partial_1\partial_2 w(y^-,t)\right\| < +\infty\,, \tag{3.80}$$

$$\sup_{(y^-,t)\in B(0,\sigma_1)\times\mathbb{I}_1} \left\|\partial_2^2 w(y^-,t)\right\| < +\infty\,. \tag{3.81}$$

Proof. The idea of the proof is to *localize* the function $h(y,t)$ to a neighborhood of zero in the variable y for the purpose of applying Prop. 3.18. This cut-off technique is known in the non-autonomous ODE literature, see, *e.g.*, (Kloeden and Rasmussen, 2011, Th. 6.10). Let $\psi:\mathbb{R}^d\to[0,1]$ be a smooth function such that $\psi(y)=1$ if $\|y\|\leq 1$, and $\psi(y)=0$ if $\|y\|\geq 2$. Let $C=\max_y\|\nabla\psi(y)\|$ where $\nabla\psi$ is the Jacobian matrix of ψ. Thanks to the convergence (3.78), we can choose $t_1>0$ large enough and $\sigma>0$ small enough so that

$$\sup_{(t,y)\in[t_1,\infty)\times B(0,2\sigma)} \|\partial_1\varepsilon(y,t)\| < \frac{\alpha^+-\alpha^-}{4K(1+2C)},$$

and we set $\mathbb{I}_1 = [t_1, \infty)$. Writing $\tilde{\varepsilon}(y,t) = \psi(y/\sigma)\varepsilon(y,t)$, it holds that for each $(t,y) \in \mathbb{I}_1 \times \mathbb{R}^d$,

$$\begin{aligned}\|\partial_1\tilde{\varepsilon}(y,t)\| &\leq \sigma^{-1}C\mathbb{1}_{\|y\|\leq 2\sigma}\|\varepsilon(y,t)\| + \mathbb{1}_{\|y\|\leq 2\sigma}\|\partial_1\varepsilon(y,t)\| \\ &\leq \left(\max_{\|y\|\leq 2\sigma}\|\partial_1\varepsilon(y,t)\|\right)\left(\sigma^{-1}C\|y\| + 1\right)\mathbb{1}_{\|y\|\leq 2\sigma} \\ &\leq \frac{\alpha^+ - \alpha^-}{4K},\end{aligned}$$

where we used the mean value inequality along with $\varepsilon(0,t) = 0$ to obtain the second inequality. Thus, the function $\tilde{h}(y,t) = \Lambda y + \tilde{\varepsilon}(y,t)$ satisfies all the assumptions of Prop. 3.18. In addition, the function $\tilde{\varepsilon}$ coincides with the function ε on $B(0,\sigma_1) \times \mathbb{I}_1$, and so it is for the functions $\tilde{h}$ and h. Finally, it follows from (Kloeden and Rasmussen, 2011, Th. 6.3) that

$$|w|_1 \leq \frac{2K^2}{\alpha_+ - \alpha_- - 4K|\tilde{\varepsilon}|_1}|\tilde{\varepsilon}|_1$$

(note that L in (Kloeden and Rasmussen, 2011, Th. 6.3) corresponds to $|\tilde{\varepsilon}|_1$ with our notations). Using Eq. (3.78), we can make $|\tilde{\varepsilon}|_1$ as small as needed by choosing σ, t_1 respectively small and large enough, which gives us Eq. (3.79). The proof of the last two equations follows from the application of Lem. 3.19 to $\tilde{h}$ and w. The result is immediate after noticing that for $(y,t) \in \mathbb{R}^d \times \mathbb{I}_1$, we have $\left\|\partial_2\tilde{\varepsilon}(y,t)\right\| \leq \left\|\partial_2\varepsilon(y,t)\right\|$. ■

3.7.2 Proof of Th. 3.8

We shall rely on the following result of Brandière and Duflo. Recall that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space equipped with a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

Proposition 3.21. ((Brandière and Duflo, 1996, Prop. 4)) Given a sequence (γ_n) of deterministic nonnegative stepsizes such that $\sum_k \gamma_k = +\infty$ and $\sum_k \gamma_k^2 < +\infty$, consider the $\mathbb{R}^d$–valued stochastic process $(z_n)_{n\in\mathbb{N}}$ given by

$$z_{n+1} = (I + \gamma_{n+1}H_n)z_n + \gamma_{n+1}\eta_{n+1} + \gamma_{n+1}\rho_{n+1}.$$

Assume that z_0 is $\mathcal{F}_0$–measurable and that the sequences (η_n), (ρ_n) together with the sequence of random matrices (H_n) are $(\mathcal{F}_n)$–adapted. Moreover, on a given event $A \in \mathcal{F}$, assume the following facts:

i) $\sum_n \|\rho_n\|^2 < \infty$.
ii) $\limsup \mathbb{E}[\|\eta_{n+1}\|^{2+a} \,|\, \mathcal{F}_n] < \infty$ for some $a > 0$, and $\mathbb{E}[\eta_{n+1} \,|\, \mathcal{F}_n] = 0$.
iii) $\liminf \mathbb{E}[\|\eta_{n+1}\|^2 \,|\, \mathcal{F}_n] > 0$.

Let $H \in \mathbb{R}^{d\times d}$ be a deterministic matrix such that the real parts of its eigenvalues are all positive. Then,

$$\mathbb{P}\left(A \cap [z_n \to 0] \cap [H_n \to H]\right) = 0.$$

We now enter the proof of Th. 3.8. Recall the development (3.11) of $b(z,t)$ near $z_\star$ and the spectral factorization (3.12) of the matrix D. To begin with, it will be convenient to make the variable change $y = Q^{-1}(z - z_\star)$, and set

$$h(y,t) = Q^{-1}b(Qy + z_\star, t) = \Lambda y + \tilde{e}(y,t),$$

with $\tilde{e}(y,t) = Q^{-1}e(Qy + z_\star, t)$, in such a way that our stochastic algorithm is rewritten as

$$y_{n+1} = y_n + \gamma_{n+1}h(y_n, \tau_n) + \gamma_{n+1}\tilde{\eta}_{n+1} + \gamma_{n+1}\tilde{\rho}_{n+1}$$

where $\tilde{\eta}_n$ is as in the statement of the theorem and $\tilde{\rho}_n = Q^{-1}\rho_n$. Observe that the assumptions on the function e in the statement of the theorem remain true for $\tilde{e}$ with $z_\star$ replaced by zero.

If the matrix Λ has only eigenvalues with (strictly) positive real parts, *i.e.*, $d^- = 0$, then we can apply Prop. 3.21 to the sequence (z_n). Henceforth, we deal with the more complicated case where $d^- > 0$.

Apply Prop. 3.20 to h to obtain $\tilde{h}$ and σ, t_1 respectively small and large enough and $w : \mathbb{R}^{d^-} \times \mathbb{I}_1 \to \mathbb{R}^{d^+}$ where $\mathbb{I}_1 := [t_1, +\infty)$. By Assumption *iv)* of Th. 3.8 and Prop. 3.20 we can choose $\sigma_1 \leq \sigma$ such that Eq. (3.80) and Eq. (3.81) hold. Now, given $p \in \mathbb{N}$, let us define the event

$$E_p = \left[\forall n \geq p,\ \|y_n\| < \sigma_1, \tau_n \in \mathbb{I}_1\right].$$

On E_p, it holds that $h(y_n, \tau_n) = \tilde{h}(y_n, \tau_n)$ and

$$\begin{aligned} \forall n \geq p, \quad y_{n+1} &= y_n + \gamma_{n+1}h(y_n, \tau_n) + \gamma_{n+1}\tilde{\eta}_{n+1} + \gamma_{n+1}\tilde{\rho}_{n+1} \\ &= \begin{bmatrix} y_n^- \\ y_n^+ \end{bmatrix} + \gamma_{n+1}\begin{bmatrix} h^-(y_n, \tau_n) \\ h^+(y_n, \tau_n) \end{bmatrix} + \gamma_{n+1}\begin{bmatrix} \tilde{\eta}_{n+1}^- \\ \tilde{\eta}_{n+1}^+ \end{bmatrix} + \gamma_{n+1}\begin{bmatrix} \tilde{\rho}_{n+1}^- \\ \tilde{\rho}_{n+1}^+ \end{bmatrix} \end{aligned} \tag{3.82}$$

where h is partitioned as in (3.67), and where $\tilde{\eta}_n^\pm, \tilde{\rho}_n^\pm \in \mathbb{R}^{d^\pm}$. Note that, by Prop. 3.20 and Assumptions *vi)* and *vii)* on the sequence (η_n), we can choose σ, t_1 respectively small and large enough s.t.

$$\begin{aligned} \liminf \mathbb{E}[\left\|\tilde{\eta}_{n+1}^+\right\|^2 |\mathcal{F}_n] \mathbb{1}_{E_p}(y_n) \\ - 2\limsup \mathbb{E}[\left\|\partial_1 w(y_n^-, \tau_n)\tilde{\eta}_{n+1}^-\right\|^2 |\mathcal{F}_n] \mathbb{1}_{E_p}(y_n) > \frac{c^2}{2}. \end{aligned} \tag{3.83}$$

This inequality will be important in the end of our proof. Let t be in the interior of $\mathbb{I}_1$, and let $y = (y^-, y^+)$ be in a neighborhood of 0. Make the variable change $(y^-, y^+) \mapsto (u^-, u^+)$ with

$$\begin{aligned} u^+ &= y^+ - w(y^-, t), \\ u^- &= y^-, \end{aligned}$$

where w is the function defined in the statement of Prop. 3.20, and let

$$\begin{aligned} W(u^-, u^+, t) &= h^+(y,t) - \partial_1 w(y^-, t)h^-(y,t) - \partial_2 w(y^-, t) \\ &= h^+((u^-, u^+ + w(u^-, t)), t) \\ &\quad - \partial_1 w(u^-, t)h^-((u^-, u^+ + w(u^-, t)), t) - \partial_2 w(u^-, t). \end{aligned}$$

By Prop. 3.20 and Lem. 3.19, it holds that $W(u^-, 0, t) = 0$. Moreover, $W(u^-, \cdot, t) \in \mathcal{C}^1$ by the assumptions on h. Therefore, writing $y(r) = (u^-, ru^+ + w(u^-, t))$ for $r \in [0, 1]$, and using the decomposition (3.67), we get that

$$\begin{aligned} W(u^-, u^+, t) &= \int_0^1 \partial_2 W(u^-, ru^+, t) u^+ \, dr \\ &= \Lambda^+ u^+ \\ &+ \int_0^1 \left(\partial_1 \varepsilon^+(y(r), t) \begin{bmatrix} 0 \\ I_{d^+} \end{bmatrix} - \partial_1 w(u^-, t) \partial_1 \varepsilon^-(y(r), t) \begin{bmatrix} 0 \\ I_{d^+} \end{bmatrix} \right) u^+ dr. \end{aligned}$$

We can also write $y(r) = (y^-, ry^+ + (1-r)w(y^-, t))$. Recalling that $w(0, t) = 0$ and that $\|\partial_1 w(y^-, t)\|$ is bounded on $\mathbb{R}^{d^-} \times \mathbb{I}$, we get by the mean value inequality that $\left\|w(y^-, t)\right\| \leq C \left\|y^-\right\|$ where $C > 0$ is a constant. Thus, $\left\|y(r)\right\| \leq (1+C) \left\|y\right\|$. Moreover, $\varepsilon(y, t) = Q^{-1} e(Qy, t)$ for $\|y\| < \sigma$. Thus, we get by (3.13) that $\left\|\partial_1 \varepsilon(y(r), t)\right\| \to 0$ as $(y, t) \to (0, \infty)$ uniformly in $r \in [0, 1]$. Using again the boundedness of $\|\partial_1 w(\cdot, \cdot)\|$, we eventually obtain that

$$W(u^-, u^+, t) = \left(\Lambda^+ + \Delta(y, t)\right) u^+, \quad \text{with} \quad \lim_{(y,t) \to (0,\infty)} \Delta(y, t) = 0.$$

On the event E_p, assume that $n \geq p$, and write

$$u_n^+ = y_n^+ - w(y_n^-, \tau_n), \quad u_n^- = y_n^-,$$

(see Eq. (3.82)). Choosing $\alpha_- > 0$ small enough so that the gap condition (3.65) is satisfied with $m = 2$, we have by Taylor's expansion

$$\begin{aligned} &w(y_{n+1}^-, \tau_{n+1}) - w(y_n^-, \tau_n) \\ &= w(y_{n+1}^-, \tau_{n+1}) - w(y_n^-, \tau_{n+1}) + w(y_n^-, \tau_{n+1}) - w(y_n^-, \tau_n) \\ &= \partial_1 w(y_n^-, \tau_{n+1})(y_{n+1}^- - y_n^-) + \gamma_{n+1} \partial_2 w(y_n^-, \tau_n) + \epsilon_{n+1} + \epsilon_{n+1}^\gamma \,, \end{aligned}$$

$$\begin{aligned} \text{with} \quad &\left\|\epsilon_{n+1}\right\| \leq \sup_{y^- \in [y_n^-, y_{n+1}^-]} \left\|\partial_1^2 w(y^-, \tau_{n+1})\right\| \left\|y_{n+1}^- - y_n^-\right\|^2 \,, \\ \text{and} \quad &\left\|\epsilon_{n+1}^\gamma\right\| \leq \sup_{\tau \in [\tau_n, \tau_{n+1}]} \left\|\partial_2^2 w(y_n^-, \tau)\right\| \gamma_{n+1}^2 \,. \end{aligned}$$

Using this equation, we obtain

$$\begin{aligned} u_{n+1}^+ - u_n^+ = \gamma_{n+1} W(u_n^-, u_n^+, \tau_n) &+ \gamma_{n+1} \left(\tilde{\eta}_{n+1}^+ - \partial_1 w(y_n^-, \tau_{n+1}) \tilde{\eta}_{n+1}^-\right) \\ &+ \gamma_{n+1} \left(\tilde{\rho}_{n+1}^+ - \partial_1 w(y_n^-, \tau_{n+1}) \tilde{\rho}_{n+1}^-\right) - \epsilon_{n+1} - \epsilon_{n+1}^\gamma \\ &+ \gamma_{n+1} \left(\partial_1 w(y_n^-, \tau_n) - \partial_1 w(y_n^-, \tau_{n+1})\right) h^-(y_n, \tau_n) \,, \end{aligned}$$

which leads to

$$u_{n+1}^+ = u_n^+ + \gamma_{n+1} \left(\Lambda^+ + \Delta(y_n, \tau_n)\right) u_n^+ + \gamma_{n+1} \bar{\eta}_{n+1} + \gamma_{n+1} \bar{\rho}_{n+1}, \tag{3.84}$$

with $\bar{\eta}_{n+1} = \tilde{\eta}^+_{n+1} - \partial_1 w(y^-_n, \tau_n)\tilde{\eta}^-_{n+1}$ and

$$\bar{\rho}_{n+1} = \tilde{\rho}^+_{n+1} - \partial_1 w(y^-_n, \tau_n)\tilde{\rho}^-_{n+1} - \mathbb{1}_{\gamma_{n+1}>0}\frac{\epsilon_{n+1} + \epsilon^\gamma_{n+1}}{\gamma_{n+1}} + \Big(\partial_1 w(y^-_n, \tau_n) - \partial_1 w(y^-_n, \tau_{n+1})\Big) h^-(y_n, \tau_n)\,. \quad (3.85)$$

To finish the proof, it remains to check that the noise sequence satisfies the assumptions of Prop. 3.21 on the event $A_p = E_p \cap [y_n \to 0]$. In the remainder, C' will indicate some positive constant which can change from an inequality to another one.

First, we verify that $\sum_n \|\bar{\rho}_n\|^2 < \infty$ on A_p by controlling each one of the terms of $\bar{\rho}_n$. Combining the boundedness of $\partial_1 w(\cdot,\cdot)$ with the summability assumption $\sum_n \|\tilde{\rho}_{n+1}\|^2 \mathbb{1}_{z_n\in\mathcal{W}} < +\infty$ a.s., we immediately obtain on A_p that, given our choice of σ, $\sum_n \|\tilde{\rho}^+_{n+1} - \partial_1 w(y^-_n, \tau_n)\tilde{\rho}^-_{n+1}\|^2 < +\infty$. Moreover, it holds that $\Big(\|\epsilon^\gamma_{n+1}\|/\gamma_{n+1}\Big)^2 \leq C'\gamma^2_{n+1}$ by invoking Prop. 3.20. In addition, using the boundedness of $\partial^2_1 w(\cdot,\cdot)$, we can write

$$\mathbb{1}_{\gamma_{n+1}>0}\left\|\frac{\epsilon_{n+1}}{\gamma_{n+1}}\right\|^2 \leq \mathbb{1}_{\gamma_{n+1}>0}\frac{C'}{\gamma^2_{n+1}}\left\|y_{n+1} - y_n\right\|^4 \leq C'\gamma^2_{n+1}\Big(\left\|h(y_n,\tau_n)\right\|^4 + \left\|\tilde{\eta}_{n+1}\right\|^4 + \left\|\tilde{\rho}_{n+1}\right\|^4\Big)\,.$$

A coupling argument (see (Brandière and Duflo, 1996, p. 401)) shows that we can simplify the condition $\limsup \mathbb{E}[\|\eta_{n+1}\|^4 \,|\, \mathcal{F}_n]\mathbb{1}_{z_n\in\mathcal{W}} < \infty$ to $\mathbb{E}[\|\eta_{n+1}\|^4 \,|\, \mathcal{F}_n]\mathbb{1}_{z_n\in\mathcal{W}} < C'$. The latter condition implies that $\mathbb{E}[\mathbb{1}_{A_p}\sum_n \gamma^2_{n+1}\left\|\eta_{n+1}\right\|^4] \leq \sum_n C'\gamma^2_{n+1}$, and therefore $\sum_n \gamma^2_{n+1}\left\|\eta_{n+1}\right\|^4 \mathbb{1}_{A_p} < +\infty$ a.s. As a consequence, noticing also the boundedness of $(h(y_n,\tau_n))$ and $(\tilde{\rho}_n)$ on A_p, we deduce that $\sum_n \mathbb{1}_{\gamma_{n+1}>0}\left\|\frac{\epsilon_{n+1}}{\gamma_{n+1}}\right\|^2 < +\infty$ on A_p. We now briefly control the last term of $\bar{\rho}_n$. By the mean value inequality, we obtain that

$$\left\|\Big(\partial_1 w(y^-_n, \tau_n) - \partial_1 w(y^-_n, \tau_{n+1})\Big) h^-(y_n,\tau_n)\right\| \leq \gamma_{n+1}\sup_{(y^-,t)}\left\|\partial_2\partial_1 w(y^-,t)\right\|\,\|h^-(y_n,\tau_n)\| \leq C'\gamma_{n+1}\,,$$

where the last inequality stems from Prop. 3.20-Eq. (3.80) together with the boundedness of the sequence $(h(y_n,\tau_n))$. In view of Eq. (3.85) and the above estimates, we deduce that $\sum_n \|\bar{\rho}_{n+1}\|^2\mathbb{1}_{A_p} < +\infty$ a.s. on A_p.

We verify the remaining conditions on the noise sequence $(\bar{\eta}_n)$. We can easily remark that $\mathbb{E}[\bar{\eta}_{n+1}|\mathcal{F}_n] = 0$ and $\left\|\bar{\eta}_{n+1}\right\| \leq C'\left\|\eta_{n+1}\right\|$ on A_p. Hence, $\limsup \mathbb{E}[\|\bar{\eta}_{n+1}\|^4 \,|\, \mathcal{F}_n]\mathbb{1}_{z_n\in\mathcal{W}} < \infty$. The last condition, meaning that the noise is exciting enough, stems from noting that

$$\begin{aligned} 2\liminf \mathbb{E}[\left\|\bar{\eta}_{n+1}\right\|^2 |\mathcal{F}_n]\mathbb{1}_{A_p} &\geq \liminf \mathbb{E}[\left\|\tilde{\eta}^+_{n+1}\right\|^2 |\mathcal{F}_n]\mathbb{1}_{A_p} \\ &\quad - 2\limsup \mathbb{E}[\left\|\partial_1 w(y^-_n,\tau_n)\tilde{\eta}^-_{n+1}\right\|^2 |\mathcal{F}_n]\mathbb{1}_{A_p} \\ &> \frac{c^2}{2}\,, \end{aligned}$$

where we used our choice of σ, t_1 and Eq. (3.83).

Noticing that $[y_n \to 0] \subset [\Delta(y_n, \tau_n) \to 0]$, we can now apply Prop. 3.21 to the sequence (u_n^+) (see Eq. (3.84)) with $A = A_p$ to obtain

$$\mathbb{P}\left(A_p \cap [u_n^+ \to 0]\right) = \mathbb{P}\left(A_p \cap [u_n^+ \to 0] \cap [\Delta(y_n, \tau_n) \to 0]\right) = 0\,.$$

We now show that $[y_n \to 0] \subset [u_n^+ \to 0]$, which amounts to prove that $w(y_n^-, \tau_n) \to 0$ given $y_n \to 0$. To that end, upon noting that $w(0, \cdot) \equiv 0$ and that $\partial_1 w(\cdot, \cdot)$ is bounded, it suffices to apply the mean value inequality, writing:

$$\|w(y_n^-, \tau_n)\| = \|w(y_n^-, \tau_n) - w(0, \tau_n)\| \leq \sup_{(y^-, t)} \|\partial_1 w(y^-, t)\|\ \|y_n^-\| \leq K \|y_n^-\|\,.$$

We have shown so far that $\mathbb{P}(A_p) = 0$. Since $y_n = Q^{-1} z_n$ and $[y_n \to 0] \subset \bigcup_{p \in \mathbb{N}} E_p$, we finally obtain that

$$\mathbb{P}[z_n \to 0] = \mathbb{P}[y_n \to 0] = \mathbb{P}\left(\bigcup_{p \in \mathbb{N}} ([y_n \to 0] \cap E_p)\right) = \mathbb{P}\left(\bigcup_{p \in \mathbb{N}} A_p\right) = 0.$$

Th. 3.8 is proven.

3.7.3 Proofs for Section 3.4.2.1

3.7.3.1 Proof of Lem. 3.9

The matrix D coincides with $\nabla g_\infty(z_\star)$, where the function g_∞ is defined in (3.20). As such, its expression is immediate. Recalling that $p_\infty S(x_\star) - q_\infty v_\star = 0$, we get

$$\begin{aligned}
&g(z,t) - D(z - z_\star) \\
&= \begin{bmatrix} \mathsf{p}(t) S(x) - \mathsf{q}(t) v - p_\infty \nabla S(x_\star)(x - x_\star) + q_\infty (v - v_\star) \\ \mathsf{h}(t) \nabla F(x) - \mathsf{r}(t) m - h_\infty \nabla^2 F(x_\star)(x - x_\star) + r_\infty m \\ -m \left((v + \varepsilon)^{-\frac{1}{2}} - (v_\star + \varepsilon)^{-\frac{1}{2}} \right) \end{bmatrix} \\
&= \begin{bmatrix} -\mathsf{q}(t) + q_\infty & 0 & (\mathsf{p}(t) - p_\infty) \nabla S(x_\star) \\ 0 & r_\infty - \mathsf{r}(t) & (\mathsf{h}(t) - h_\infty) \nabla^2 F(x_\star) \\ \frac{m}{2(v_\star + \varepsilon)^{\frac{3}{2}}} & 0 & 0 \end{bmatrix} \begin{bmatrix} v - v_\star \\ m \\ x - x_\star \end{bmatrix} \\
&\quad + \begin{bmatrix} \mathsf{p}(t)(S(x) - S(x_\star) - \nabla S(x_\star)(x - x_\star)) \\ \mathsf{h}(t)(\nabla F(x) - \nabla^2 F(x_\star)(x - x_\star)) \\ -m \odot \left(\frac{1}{\sqrt{v + \varepsilon}} - \frac{1}{\sqrt{v_\star + \varepsilon}} + \frac{v - v_\star}{2(v_\star + \varepsilon)^{\frac{3}{2}}} \right) \end{bmatrix} + \begin{bmatrix} \mathsf{p}(t) S(x_\star) - \mathsf{q}(t) v_\star \\ 0 \\ 0 \end{bmatrix} \\
&:= e(z,t) + c(t).
\end{aligned}$$

Under the assumptions made, it is easy to see that the function $e(z,t)$ has the properties required in the statement of Th. 3.8.

3.7.3.2 Proof of Th. 3.10

Consider the matrix D defined in the statement of Lem. 3.9. A spectral analysis of this matrix as regards its eigenvalues with positive real parts is done in the following lemma.

Lemma 3.22. Let D be the matrix provided in the statement of Lem. 3.9. Each eigenvalue ζ of the matrix D such that $\Re\zeta > 0$ is real, and its algebraic and geometric multiplicities are equal. Moreover, there is a one-to-one correspondence φ between these eigenvalues and the negative eigenvalues of $V^{\frac{1}{2}}\nabla^2 F(x_\star)V^{\frac{1}{2}}$. Let d^+ be the dimension of the eigenspace of $V^{\frac{1}{2}}\nabla^2 F(x_\star)V^{\frac{1}{2}}$ that is associated with its negative eigenvalues, let

$$W = \begin{bmatrix} w_1 \\ \vdots \\ w_{d^+} \end{bmatrix} \in \mathbb{R}^{d^+\times d}$$

be a matrix which rows are independent eigenvectors of $V^{\frac{1}{2}}\nabla^2 F(x_\star)V^{\frac{1}{2}}$ that generate this eigenspace, and denote as $\beta_k < 0$ the eigenvalue associated with w_k. Then, the rows of the rank d^+-matrix

$$A^+ = \left[0_{d^+\times d}, \quad WV^{\frac{1}{2}}, \quad -\operatorname{diag}(r_\infty + \varphi^{-1}(\beta_k))WV^{-\frac{1}{2}}\right] \in \mathbb{R}^{d^+\times 3d}$$

generate the left eigenspace of D, the row k being an eigenvector for the eigenvalue $\varphi^{-1}(\beta_k)$.

Proof. It is obvious that the block lower-triangular matrix D has d eigenvalues equal to $-q_\infty$ and $2d$ eigenvalues which are those of the sub-matrix

$$\widetilde{D} = \begin{bmatrix} -r_\infty I_d & h_\infty \nabla^2 F(x_\star) \\ -V & 0 \end{bmatrix}.$$

Given $\lambda \in \mathbb{C}$, we obtain by standard manipulations involving determinants that

$$\begin{aligned} \det(\widetilde{D} - \lambda) &= \det(\lambda(r_\infty + \lambda) + h_\infty V\nabla^2 F(x_\star)) \\ &= \det(\lambda(r_\infty + \lambda) + h_\infty V^{\frac{1}{2}}\nabla^2 F(x_\star)V^{\frac{1}{2}}). \end{aligned}$$

Denoting as $\{\beta_k\}_{k=1}^d$ the eigenvalues of $h_\infty V^{\frac{1}{2}}\nabla^2 F(x_\star)V^{\frac{1}{2}}$ counting the multiplicities, we obtain from the last equation that the eigenvalues of $\widetilde{D}$ are the solutions of the second order equations

$$\lambda^2 + r_\infty\lambda + \beta_k = 0, \quad k = 1, \dots, d.$$

The product of the roots of such an equation is β_k, and their sum is $-r_\infty \le 0$. Thus, denoting as $\zeta_{k,1}$ and $\zeta_{k,2}$ these roots, it is easy to see that if $\beta_k \ge 0$, then $\Re\zeta_{k,1}, \Re\zeta_{k,2} \le 0$, while if $\beta_k < 0$, then both $\zeta_{k,i}$ are real, and only one of them is positive. Thus, we have so far shown that the eigenvalues of D which real parts are positive are themselves real, and there is a one-to-one map φ from the set of positive eigenvalues of D to the set of negative eigenvalues of $V^{\frac{1}{2}}\nabla^2 F(x_\star)V^{\frac{1}{2}}$. Moreover, the algebraic multiplicity of the eigenvalue $\zeta > 0$ of D is equal to the multiplicity of $\varphi(\zeta)$.

Let us now turn to the left (row) eigenvectors of D that correspond to these eigenvalues. To that end, we shall solve the equation

$$uD = \zeta u \quad \text{with } u = [0, u_1, u_2], \quad u_{1,2} \in \mathbb{R}^{1\times d}, \tag{3.86}$$

for a given eigenvalue $\zeta > 0$ of D. Developing this equation, we get

$$-r_\infty u_1 - u_2 V = \zeta u_1, \quad h_\infty u_1 \nabla^2 F(x_\star) = \zeta u_2.$$

If we now write $\tilde{u}_1 = u_1 V^{-\frac{1}{2}}$ and $\tilde{u}_2 = u_2 V^{\frac{1}{2}}$, this system becomes

$$-r_\infty \tilde{u}_1 - \tilde{u}_2 = \zeta \tilde{u}_1, \quad h_\infty \tilde{u}_1 V^{\frac{1}{2}} \nabla^2 F(x_\star) V^{\frac{1}{2}} = \zeta \tilde{u}_2,$$

or, equivalently,

$$\tilde{u}_2 = -(r_\infty + \zeta)\tilde{u}_1, \quad \tilde{u}_1 \left(\zeta^2 + r_\infty \zeta + h_\infty V^{\frac{1}{2}} \nabla^2 F(x_\star) V^{\frac{1}{2}} \right) = 0,$$

which shows that $\tilde{u}_1$ is a left eigenvector of $V^{\frac{1}{2}} \nabla^2 F(x_\star) V^{\frac{1}{2}}$ associated with the eigenvalue $\varphi(\zeta)$. What's more, assume that r is the multiplicity of $\varphi(\zeta)$, and, without generality loss, that the submatrix $W_{r,\cdot}$ made of the first r rows of W generates the left eigenspace of $\varphi(\zeta)$. Then, the matrix

$$\begin{bmatrix} 0_{r \times d} & W_{r\cdot} V^{\frac{1}{2}} & -(r_\infty + \zeta) W_{r\cdot} V^{-\frac{1}{2}} \end{bmatrix}$$

is a r-rank matrix which rows are independent left eigenvectors that generate the left eigenspace of D for the eigenvalue ζ. In particular, the algebraic and geometric multiplicities of this eigenvalue are equal. The same argument can be applied to the other positive eigenvalues of D. ∎

We now have all the elements to prove Th. 3.10. Recall Eq. (3.14):

$$z_{n+1} = z_n + \gamma_{n+1} b(z_n, \tau_n) + \gamma_{n+1} \eta_{n+1} + \gamma_{n+1} \rho_{n+1},$$

where $b(z,t) = g(z,t) - c(t) = D(z - z_\star) + e(z,t)$ and $\rho_n = c(\tau_{n-1}) + \tilde{\rho}_n$. With these same notations, we check that Assumptions *i)–vi)* in the statement of Th. 3.8 are satisfied. The function $e(z,t)$ satisfies Assumptions *i)–iv)* by Lem. 3.9. We now verify that the sequence (ρ_n) fulfills Assumption *v)*. First, observe that $\sum_n \|c(\tau_n)\|^2 < \infty$ under Assumption 3.4.3-*i)*. Then, we control the second term $(\tilde{\rho}_n)$. After straightforward derivations, one can show the existence of a positive constant C (depending only on ε and a neighborhood $\mathcal{W}$ of $z_\star$) such that

$$\|\tilde{\rho}_{n+1}\|^2 \mathbb{1}_{z_n \in \mathcal{W}} \leq C(\|m_n - m_{n+1}\|^2 + \|v_{n+1} - v_n\|^2) \mathbb{1}_{z_n \in \mathcal{W}}\,. \tag{3.87}$$

Using the boundedness of the sequences (h_n) and (r_n) together with the update rule of m_n and Assumption 3.4.3-*iii)*, there exists a positive constant C' independent of n (which may change from an inequality to another) such that

$$\begin{aligned} \mathbb{E}\left[\|m_n - m_{n+1}\|^2 \mathbb{1}_{z_n \in \mathcal{W}} \right] &\leq \gamma_{n+1}^2 C' \mathbb{E}\left[(1 + \mathbb{E}_\xi\left[\|\nabla f(x_n, \xi)\|^2 \right]) \mathbb{1}_{z_n \in \mathcal{W}} \right] \\ &\leq C' \gamma_{n+1}^2. \end{aligned} \tag{3.88}$$

A similar result holds for $\mathbb{E}\left[\|v_n - v_{n+1}\|^2 \mathbb{1}_{z_n \in \mathcal{W}} \right]$ following the same arguments. In view of Eqs. (3.87)-(3.88), it holds that $\mathbb{E}\left[\sum_n \|\tilde{\rho}_{n+1}\|^2 \mathbb{1}_{z_n \in \mathcal{W}} \right] < +\infty$ given the assumption $\sum_n \gamma_{n+1}^2 < +\infty$. Therefore, $\sum_n \|\tilde{\rho}_{n+1}\|^2 \mathbb{1}_{z_n \in \mathcal{W}} < +\infty$ a.s., which completes our verification of condition *v)* of Th. 3.8. Assumption *vi)* follows from condition 3.4.3-*iii)*. Finally,

let us make Assumption *vii)* of Th. 3.8 more explicit. Partitioning the matrix Q^{-1} as $Q^{-1} = \begin{bmatrix} B^- \\ B^+ \end{bmatrix}$ where $B^\pm$ has $d^\pm$ rows, Lem. 3.22 shows that the row spaces of B^+ and A^+ are the same, which implies that Assumption *vii)* can be rewritten equivalently as $\mathbb{E}[\left\|A^+\eta_{n+1}\right\|^2 \,|\, \mathcal{F}_n]\mathbb{1}_{z_n\in\mathcal{W}} \geq c^2\mathbb{1}_{z_n\in\mathcal{W}}$. By inspecting the form of η_n provided by Eq. (3.28) (written as a column vector), one can readily check that Assumption 3.4.3-*iv)* implies Assumption *vii)* of Th. 3.8 for a small enough neighborhood $\mathcal{W}$, using the continuity of the covariance matrix $V^{\frac{1}{2}}\mathbb{E}_\xi(\nabla f(x,\xi) - \nabla F(x))(\nabla f(x,\xi) - \nabla F(x))^{\mathrm{T}}V^{\frac{1}{2}}$ when x is near $x_\star$.

3.7.4 Proof of Th. 3.11

As mentioned in Section 3.4.2.2, the proof of Th. 3.11 is almost identical to the one of Th. 3.10. We point out the main differences here. In Lem. 3.9, replace D by $\tilde{D} = \begin{bmatrix} 0 & h_\infty\nabla^2 F(x_\star) \\ -I_d & 0 \end{bmatrix}$ and set $c(t) = 0$. Then, in Lem. 3.22, replace the matrix $V^{1/2}\nabla^2 F(x_\star)V^{1/2}$ by the Hessian $\nabla^2 F(x_\star)$.

4

Convergence Rates of a Momentum Algorithm with Bounded Adaptive Stepsize for Non-convex Optimization

Abstract In this chapter, we study the ADAM algorithm for smooth non-convex optimization under a boundedness assumption on the adaptive learning rate. The bound on the adaptive stepsize depends on the Lipschitz constant of the gradient of the objective function and provides safe theoretical adaptive stepsizes. Under this boundedness assumption, we show a novel first order convergence rate result in both deterministic and stochastic contexts. Compared to the previous chapters, the results are of a different flavour. Instead of asymptotic results, we focus on "convergence rates" which are more common in the machine learning community. Furthermore, we establish convergence rates of the function value sequence using the Kurdyka-Łojasiewicz property, borrowing results for gradient-like sequences from the optimization community.

4.1 Contributions

Consider the unconstrained optimization problem $\min_{x\in\mathbb{R}^d} f(x)$, where $f : \mathbb{R}^d \to \mathbb{R}$ is a differentiable map and d is an integer. In this chapter, we focus on the class of adaptive algorithms among which ADAM (Kingma and Ba, 2015) is probably the most popular algorithm for optimizing the weights of neural networks. Recently, Reddi et al. (2018) exhibited a simple convex stochastic optimization problem over a compact set where ADAM fails to converge because of its short-term gradient memory for specific values of its hyperparameters. Moreover, they proposed an algorithm called AMSGRAD to fix this issue. This work opened the way to the emergence of other variants of ADAM (see Section 4.3 for a detailed review). In this chapter, under a bounded stepsize assumption, we propose a convergence rate analysis of ADAM-like algorithms for non-convex optimization.

Our contributions are as follows.

- We establish a convergence rate for ADAM in the deterministic case for non-convex optimization under a bounded stepsize. This algorithm can be seen as a deterministic clipped version of ADAM which guarantees safe theoretical stepsizes. More precisely, if n is the number of iterations of the algorithm, we show a $O(1/n)$ convergence rate of the minimum of the squared gradients norms by introducing a suitable Lyapunov function.

- We show a similar convergence result for non-convex stochastic optimization up to the limit of the variance of stochastic gradients under an almost surely bounded stepsize. In comparison to the literature, the hypothesis of the boundedness of the gradients is relaxed and the convergence result is independent of the dimension d of the parameters.

- We propose a convergence rate analysis of the objective function of the algorithm using the Kurdyka-Łojasiewicz (KŁ) property. To the best of our knowledge, this is the first time such a result is established for an adaptive optimization algorithm.

Table 4.1 – Some famous algorithms.

Algorithm	**Effective stepsize** a_{n+1}	**Momentum**
SGD (Robbins and Monro, 1951)	$a_{n+1} \equiv a$	$b = 1$ (no momentum)
ADAGRAD (Duchi et al., 2011)	$a_{n+1} = a \left(\sum_{i=0}^{n} g_i^2\right)^{-1/2}$	$b = 1$
RMSPROP (Tieleman and Hinton, 2012)	$a_{n+1} = a \left[\epsilon + \left(c \sum_{i=0}^{n} (1-c)^{n-i} g_i^2\right)^{1/2}\right]^{-1}$	$b = 1$
ADAM (Kingma and Ba, 2015)	$a_{n+1} = a \left[\epsilon + \left(c \sum_{i=0}^{n} (1-c)^{n-i} g_i^2\right)^{1/2}\right]^{-1}$	$0 \leq b \leq 1$ (close to 0)

Chapter organization. Section 5.4 introduces the algorithm we analyze. Section 4.3 considers some related works. Section 4.4 establishes first order convergence rates in terms of the minimum of the gradients norms in both deterministic and stochastic settings. Finally, Section 4.5 derives function value convergence rates under the KŁ property. All the proofs are deferred to the last sections of this chapter.

4.2 A momentum algorithm with adaptive stepsize

Notations. All operations between vectors of $\mathbb{R}^d$ are to read coordinatewise. In particular, for two vectors x, y in $\mathbb{R}^d$ and $\alpha \in \mathbb{Z}$, we denote by xy, x/y, x^α the vectors on $\mathbb{R}^d$ whose k-th coordinates are respectively given by $x_k y_k$, x_k/y_k, x_k^α. The vector of ones of $\mathbb{R}^d$ is denoted by $\mathbf{1}$. When a scalar is added to a vector, it is added to each one of its coordinates. Inequalities are also to be read coordinatewise. If $x \in \mathbb{R}^d, x \leq \lambda \in \mathbb{R}$ means that each coordinate of x is smaller than λ.

We investigate the following algorithm defined by two sequences (x_n) and (p_n) in $\mathbb{R}^d$:

$$\begin{cases} x_{n+1} = x_n - a_{n+1} p_{n+1} \\ p_{n+1} = p_n + b\left(\nabla f(x_n) - p_n\right) \end{cases} \tag{4.1}$$

where $\nabla f(x)$ is the gradient of f at point x, (a_n) is a sequence of vectors in $\mathbb{R}^d$ with positive coordinates, b is a positive real constant and $x_0, p_0 \in \mathbb{R}^d$.

Algorithm 5.1 includes the classical Heavy-ball method as a special case, but is much more general. Indeed, we allow the sequence of stepsizes (a_n) to be adaptive: $a_n \in \mathbb{R}^d$ may depend on the past gradients $g_k := \nabla f(x_k)$ and the iterates x_k for $k \leq n$. We stress that the stepsize a_n is a vector of $\mathbb{R}^d$ and that the product $a_{n+1} p_{n+1}$ in (5.1) is read componentwise (this is equivalent to the formulation with a diagonal matrix preconditioner applied to the gradient (McMahan and Streeter, 2010; Gupta et al., 2017; Agarwal et al., 2019; Staib et al., 2019)).

We present in Table 4.1 how to recover some of the famous algorithms with a vector stepsize formulation. In particular, ADAM (Kingma and Ba, 2015) defined by the iterates:

$$\begin{cases} x_{n+1} = x_n - \frac{a}{\epsilon + \sqrt{v_{n+1}}} p_{n+1} \\ p_{n+1} = p_n + b\left(\nabla f(x_n) - p_n\right) \\ v_{n+1} = v_n + c\left(\nabla f(x_n)^2 - v_n\right) \end{cases} \tag{4.2}$$

for constants [1] $a \in \mathbb{R}_+$, $b, c \in [0, 1]$, can be seen as an instance of this algorithm by setting $a_n = \frac{a}{\epsilon + \sqrt{v_n}}$ where the vector v_n, as defined above, is an exponential moving average of the gradient squared. For simplification, we omit bias correction steps for p_{n+1} and v_{n+1}. Their effect vanishes quickly along the iterations.

We introduce the main assumption on the objective function which is standard in gradient-based algorithms analysis.

Assumption 4.2.1. The mapping $f : \mathbb{R}^d \to \mathbb{R}$ is:

(i) continuously differentiable and its gradient ∇f is $L-$Lipschitz continuous,
(ii) bounded from below, i.e., $\inf_{x \in \mathbb{R}^d} f(x) > -\infty$.

4.3 Related works

Adaptive algorithms as Heavy Ball. Thanks to its small per-iteration cost and its acceleration properties (at least in the strongly convex case), the Heavy-ball method, also called gradient descent with momentum, recently regained popularity in large-scale optimization (Sutskever et al., 2013). This speeding up idea dates back to the sixties with the seminal work of Polyak (1964). In order to tackle non-convex optimization problems, Ochs et al. (2014) proposed iPiano, a generalization of the well known heavy-ball in the form of a forward-backward splitting algorithm with an inertial force for the sum of a smooth possibly non-convex and a convex function. In the particular case of the Heavy-ball method, this algorithm writes for two sequences of reals (α_n) and (β_n):

$$x_{n+1} = x_n - \alpha_n \nabla f(x_n) + \beta_n (x_n - x_{n-1}) \,. \tag{4.3}$$

We remark that Algorithm 5.1 can be written in a similar fashion by choosing stepsizes $\alpha_n = b a_{n+1}$ and inertial parameters $\beta_n = (1 - b) a_{n+1}/a_n$. Ochs et al. (2014) only consider the case where α_n and β_n are real-valued. Moreover, the latter does not consider adaptive stepsizes, i.e stepsizes depending on past gradient information. We can show some improvement with respect to Ochs et al. (2014) with weaker convergence conditions in terms of the stepsize of the algorithm (see Section 4.8.5) while allowing adaptive vector-valued stepsizes a_n (see Prop. 4.9).

It is shown in Ochs et al. (2014) that the sequence of function values converges and that every limit point is a critical point of the objective function. Moreover, supposing that the Lyapunov function has the KŁ property at a cluster point, they show the finite length of the sequence of iterates and its global convergence to a critical point of the objective function. Similar results are shown in Wu and Li (2019) for a more general version than iPiano (Ochs et al., 2014) computing gradients at an extrapolated iterate like in Nesterov's acceleration.

Convergence rate. Ochs et al. (2014) determines a $O(1/n)$ convergence rate (where n is the number of iterations of the algorithm) with respect to the proximal residual which boils down to the gradient for noncomposite optimization. Furthermore, a recent work introduces a generalization of the Heavy-ball method (and Nesterov's acceleration) to constrained convex optimization in Banach spaces and provides a non-asymptotic hamiltonian based analysis with $O(1/n)$ convergence rate (Diakonikolas and Jordan, 2019). In the same vein, in Section 4.4, we establish a similar convergence result for an

[1] Please note here that these constants a, b in this chapter do not coincide with the constants a, b of Assumption 2.2.4 in Chapter 2.

adaptive stepsize instead of a fixed predetermined stepsize policy like in the Heavy-ball algorithm (see Th. 4.2).

Convergence rates under the KŁ property. The KŁ property is a powerful tool to analyze gradient-like methods. We elaborate on this property in Section 4.5. Assuming that the objective function satisfies this geometric property, it is possible to derive convergence rates. Indeed, some recent progress has been made to study convergence rates of the Heavy-ball algorithm in the non-convex setting. Ochs (2018) establishes local convergence rates for the iterates and the function values sequences under the KŁ property. The convergence proof follows a general method that is often used in non-convex optimization convergence theory. This framework was used for gradient descent (Absil et al., 2005), for proximal gradient descent (see Attouch and Bolte (2009) for an analysis with the Łojasiewicz inequality) and further generalized to a class of descent methods called *gradient-like descent* algorithms.

KŁ-based asymptotic convergence rates were established for constant Heavy-ball parameters (Ochs, 2018). Asymptotic convergence rates based on the KŁ property were also shown (Johnstone and Moulin, 2017) for a general algorithm solving non-convex nonsmooth optimization problems called Multi-step Inertial Forward-Backward splitting (Liang et al., 2016) which has iPiano and Heavy-ball methods as special cases. In this work, stepsizes and momentum parameter vary along the algorithm run and are not supposed constant. However, specific values are chosen and consequently, their analysis does not encompass adaptive stepsizes i.e. stepsizes that can possibly depend on past gradient information. In the present work, we establish similar convergence rates for methods such as Adam under a bounded stepsize assumption (see Th. 4.6). We also mention Li et al. (2017) which analyzes the accelerated proximal gradient method for non-convex programming (APGnc) and establishes convergence rates of the function value sequence by exploiting the KŁ property. This algorithm is a descent method i.e. the function value sequence is shown to decrease over time. In the present work, we analyze adaptive algorithms which are not descent methods. Note that even Heavy-ball is not a descent method. Hence, our analysis requires additional treatments to exploit the KŁ property: we introduce a suitable Lyapunov function which is not the objective function. We also point out the recent work of Xie et al. (2019) which analyzes the AdaGrad-Norm algorithm under the global Polyak-Łojasiewicz condition. This condition is a particular case of the KŁ property (see Section 4.5).

Theoretical guarantees for Adam-like algorithms. The recent literature on adaptive optimization algorithms is vast. For instance, for AdaGrad-like algorithms, several works cover the non-convex setting (Wu et al., 2018; Ward et al., 2019b; Xie et al., 2019; Li and Orabona, 2019). In the following, we almost exclusively focus on Adam-like algorithms which are different because of the momentum. The first type of convergence results uses the online optimization framework which controls the convergence rate of the average regret. This framework was adopted for AmsGrad, AdamNC (Reddi et al., 2018), AdaBound and AmsBound (Luo et al., 2019). In this setting, it is assumed that the feasible set containing the iterates is bounded by adding a projection step to the algorithm if needed. We do not make such an assumption in our analysis. (Reddi et al., 2018) establishes a regret bound in the convex setting.

The second type of theoretical results is based on the control of the norm of the (stochastic) gradients. We remark that some of these results depend on the dimension of the parameters. Zhou et al. (2018) improve this dependency in comparison to Chen et al. (2019). The convergence result in Basu et al. (2018) is established under quite

specific values of a_{n+1}, b_n and ϵ. Zaheer et al. (2018) show a $O(1/n)$ convergence rate for an increasing mini-batch size. However, the proof is provided for RMSProp and seems difficult to adapt to Adam which involves a momentum term. Indeed, unlike RMSProp, Adam does not admit the objective function as a Lyapunov function.

We also remark that all the available theoretical results assume boundedness of the (stochastic) gradients. We do not make such an assumption. Furthermore, we do not add any decreasing $1/\sqrt{n}$ factor in front of the adaptive stepsize as it is considered in Reddi et al. (2018); Luo et al. (2019) and Chen et al. (2019). Although constant hyperparameters b and c are used in practice, theoretical results are often established for non constant b_n and c_n (Reddi et al., 2018; Luo et al., 2019). We also mention that most of the theoretical bounds depend on the dimension of the parameter (Reddi et al., 2018; Zhou et al., 2018; Chen et al., 2018; Zou et al., 2019a; Chen et al., 2019; Luo et al., 2019).

Other variants of Adam. Recently, several other algorithms were proposed in the literature to enhance Adam. Although these algorithms lack theoretical guarantees, they present interesting ideas and show good practical performance. For instance, AdaShift (Zhou et al., 2019) argues that the convergence issue of Adam is due to its unbalanced stepsizes. To solve this issue, they propose to use temporally shifted gradients to compute the second moment estimate in order to decorrelate it from the first moment estimate. Nadam (Dozat, 2016) incorporates Nesterov's acceleration into Adam, in order to improve its speed of convergence. Moreover, originally motivated by variance reduction, QHAdam (Ma and Yarats, 2019) replaces both Adam's moment estimates by quasi-hyperbolic terms and recovers Adam, Rmsprop and Nadam as particular cases (modulo the bias correction). Guided by the same variance reduction principle, Radam (Liu et al., 2019) estimates the variance of the effective stepsize of the algorithm and proposes a multiplicative variance correction to the update rule.

Stepsize bound. Perhaps, the closest idea to our algorithm is the recent AdaBound (Luo et al., 2019) which considers a dynamic learning rate bound. Luo et al. (2019) show that extremely small and large learning rates can cause convergence issues to Adam and exhibit empirical situations where such an issue shows up. Inspired by the gradient clipping strategy proposed in Pascanu et al. (2013) to tackle the problem of vanishing and exploding gradients in training recurrent neural networks (see Zhang et al. (2019) for recent progress), Luo et al. (2019) apply clipping to the effective stepsize of the algorithm in order to circumvent stepsize instability. More precisely, authors propose dynamic bounds on the learning rate of adaptive methods such as Adam or AmsGrad to solve the problem of extreme learning rates which can lead to poor performance. Initialized respectively at 0 and ∞, lower and upper bounds both converge smoothly to a constant final stepsize following a predetermined formula defined by the user. Consequently, the algorithm resembles an adaptive algorithm in the first iterations and becomes progressively similar to a standard SGD algorithm. Our approach is different: we propose a static bound on the adaptive learning rate which depends on the Lipschitz constant of the objective function. This bound stems naturally from our theoretical derivations.

4.4 First order convergence rate

4.4.1 Deterministic setting

Let $(H_n)_{n\geq 0}$ be a sequence defined for all $n \in \mathbb{N}$ by $H_n := f(x_n) + \frac{1}{2b}\langle a_n, p_n^2\rangle$.

We further assume the following stepsize growth condition.

Assumption 4.4.1. There exists $\alpha > 0$ s.t. $a_{n+1} \leq \frac{a_n}{\alpha}$.

Note that this assumption is satisfied for Adam with $\alpha = \sqrt{1-c}$ where c is the parameter in (4.2). Unlike in AmsGrad (Reddi et al., 2018), the stepsize a_n is not necessarily nonincreasing. Indeed, α can be strictly smaller than 1 in Assumption 4.4.1 as it is the case for Adam.

We provide a proof of the following key lemma in Section 4.8.1.

Lemma 4.1. Let Assumptions 4.2.1 and 4.4.1 hold true. Then, for all $n \in \mathbb{N}$, for all $u \in \mathbb{R}_+$,

$$H_{n+1} \leq H_n - \langle a_{n+1}p_{n+1}^2, A_{n+1}\rangle - \frac{b}{2}\langle a_{n+1}(\nabla f(x_n) - p_n)^2, B\mathbf{1}\rangle\,, \tag{4.4}$$

where $A_{n+1} := 1 - \frac{a_{n+1}L}{2} - \frac{|b-(1-\alpha)|}{2u} - \frac{1-\alpha}{2b}$, and $B := 1 - \frac{|b-(1-\alpha)|u}{b} - (1-\alpha)$.

We now state one of the principal convergence results about Algorithm 5.1. In particular, we establish a sublinear convergence rate for the minimum of the gradients norms until time n.

Theorem 4.2. *Let Assumptions 4.2.1 and 4.4.1 hold true. Suppose that $1-\alpha < b \leq 1$. Let $\varepsilon > 0$ s.t. $a_{\sup} := \frac{2}{L}\left(1 - \frac{(b-(1-\alpha))^2}{2b\alpha} - \frac{1-\alpha}{2b} - \varepsilon\right)$ is nonnegative. Let $\delta > 0$ s.t. for all $n \in \mathbb{N}$,*

$$\delta \leq a_{n+1} \leq \min\left(a_{\sup}, \frac{a_n}{\alpha}\right)\,. \tag{4.5}$$

Then, the sequence (H_n) is nonincreasing and $\sum_n \|p_n\|^2 < \infty$. In particular, $\lim x_{n+1} - x_n \to 0$ and $\lim \nabla f(x_n) \to 0$ as $n \to +\infty$. Moreover, for all $n \geq 1$,

$$\min_{0\leq k\leq n-1} \|\nabla f(x_k)\|^2 \leq \frac{4}{nb^2}\left(\frac{H_0 - \inf f}{\delta\varepsilon} + \|p_0\|^2\right)\,.$$

Sketch of the proof. The key element of the proof is Lem. 4.1 which is a descent lemma on the function H. Indeed, the assumptions of the theorem guarantee that $A_{n+1} \geq \varepsilon$ and $B \geq 0$. Then, the result stems from summing the inequalities of Lem. 4.1. The proof can be found in Section 4.8.3.

We provide some comments on this result.

Dimension dependence. Unlike most of the theoretical results for variants of Adam as gathered in Section 4.7, we remark that the bound does not depend on the dimension d of the parameter x_k.

Comparison to gradient descent. A similar result holds for deterministic gradient descent (see Nesterov (2004, p.28)). If γ is a fix stepsize for gradient descent and there exist $\delta > 0, \varepsilon > 0$ s.t. $\gamma > \delta$ and $1 - \frac{\gamma L}{2} > \varepsilon$, then (see Section 4.8.6) for all $n \geq 1$:

$$\min_{0 \leq k \leq n-1} \|\nabla f(x_k)\|^2 \leq \frac{f(x_0) - \inf f}{n\gamma(1 - \frac{\gamma L}{2})} \leq \frac{f(x_0) - \inf f}{n\delta\varepsilon}.$$

When $p_0 = 0$ (this is the case for Adam), the bound in Th. 4.2 coincides with the gradient descent bound, up to the constant $4/b^2$. We mention however that ε for Algorithm 5.1 is defined by a slightly more restrictive condition than for gradient descent: when $b = 1$, there is no momentum and $a_{\sup} = \frac{1}{L}(1 - 2\varepsilon) < 2/L$. Hence, under the boundedness of the effective stepsize, the algorithm has a similar convergence guarantee to gradient descent. Remark that the stepsize bound almost matches the classical $2/L$ upperbound on the stepsize of gradient descent (see for example Nesterov (2004, Th. 2.1.14)).

Stepsize bound. Condition 4.5 should be seen as a clipping step of the algorithm. Indeed, the lower bound on the effective stepsize has not to be verified a posteriori after running the algorithm. Instead, a clipping of the learning rate would ensure that this boundedness assumption holds. Furthermore, if we drop the lower bound assumption on the effective stepsize a_n from Th. 4.2, we still get the following result (see Prop. 4.9), for all $n \geq 1$,

$$\frac{1}{n}\sum_{k=0}^{n-1} \langle a_{k+1}, \nabla f(x_k)^2 \rangle \leq \frac{2(1+\alpha)}{nb^2\alpha} \left(\frac{H_0 - \inf f}{\varepsilon} + \langle a_0, p_0^2 \rangle \right).$$

Influence of ε and δ. In the specific case of Adam, we obtain $La_{\sup}/2 + \varepsilon = 0.93$ with the recommended default parameters $b = 0.1$ and $c = 0.001$. Hence, we can choose ε of the order of 0.1 without exceeding 0.93. In view of Equation (4.6), the smaller is ε and the larger will be the stepsizes. However, a small ε deteriorates the bounds of Ths. 4.2 and 4.3. Once b, c (and then α) are fixed, ε can be seen as a constant. The clipping parameter δ can also be seen as constant once it is chosen.

4.4.2 Stochastic setting

We establish a similar bound in the stochastic setting. Note that the control of the minimum of the gradients norms is also standard in non-convex stochastic optimization literature (see for e.g., Ghadimi and Lan (2013)). Let $(\Xi, \mathfrak{S})$ denote a measurable space and $d \in \mathbb{N}$. Consider the problem of finding a local minimizer of the expectation $F(x) := \mathbb{E}(f(x, \xi))$ w.r.t. $x \in \mathbb{R}^d$, where $f : \mathbb{R}^d \times \Xi \to \mathbb{R}$ is a measurable map and $f(\,.\,, \xi)$ is a possibly non-convex function depending on some random variable ξ. The distribution of ξ is assumed to be unknown, but revealed online by the observation of iid copies $(\xi_n : n \geq 1)$ of the r.v. ξ. For a fixed value of ξ, the mapping $x \mapsto f(x, \xi)$ is supposed to be differentiable, and its gradient w.r.t. x is denoted by $\nabla f(x, \xi)$. We study a stochastic version of Algorithm 5.1 by replacing the deterministic gradient $\nabla f(x_n)$ by $\nabla f(x_n, \xi_{n+1})$.

Theorem 4.3. *Let Assumption 4.2.1 (for F) and Assumption 4.4.1 hold true. Assume the following bound on the variance in stochastic gradients: $\mathbb{E}\|\nabla f(x, \xi) - \nabla F(x)\|^2 \leq \sigma^2$ for all $x \in \mathbb{R}^d$. Suppose moreover that $1 - \alpha < b \leq 1$. Let $\varepsilon > 0$ s.t. $\bar{a}_{\sup} :=$*

$\frac{2}{L}\left(\frac{3}{4} - \frac{(b-(1-\alpha))^2}{2b\alpha} - \frac{1-\alpha}{2b} - \varepsilon\right)$ *is nonnegative. Let* $\delta > 0$ *s.t. for all* $n \geq 1$*, almost surely,*

$$\delta \leq a_{n+1} \leq \min\left(\bar{a}_{\sup}, \frac{a_n}{\alpha}\right). \tag{4.6}$$

Then,

$$\mathbb{E}[\|\nabla F(x_\tau)\|^2] \leq \frac{4}{nb^2}\left(\frac{H_0 - \inf f}{\delta\varepsilon} + \|p_0\|^2\right) + \frac{4\bar{a}_{\sup}}{\delta\varepsilon b^2}\sigma^2,$$

where x_τ *is an iterate uniformly randomly chosen from* $\{x_0, \cdots, x_{n-1}\}$.

Remark 15. We recover the deterministic bound of Th. 4.2 when the gradients are noiseless ($\sigma = 0$). The complete proof is deferred to Section 4.8.4.

Before proceeding, a few remarks are in order.

SGD as a particular case. By setting $b = 1$ (no momentum) and $a_{n+1} = a_n$ for all n which implies $\alpha = 1$, we recover a known rate for non-convex SGD (Ghadimi and Lan, 2013) with a maximal stepsize here of $\bar{a}_{\sup} = \frac{1}{2L}(1 - 2\varepsilon)$ and note that the proof can be slightly modified to make $\bar{a}_{\sup}$ as close as possible to $1/L$. We highlight though that the Lyapunov function H was especially tailored to handle a momentum algorithm and an analysis with f as a Lyapunov function is largely satisfying for SGD.

RMSProp. In the particular case where there is no momentum in the algorithm (i.e. RMSProp) and assuming that the gradients are bounded, a similar convergence rate is obtained in Zaheer et al. (2018, Thm. 1) (see Section 4.7). Furthermore, although we assume boundedness of the stepsize by Condition (4.6), we do not suppose that $a_1 \leq \frac{\epsilon}{2L}$ (see table in Section 4.7). The latter assumption imposes a very small stepsize ($\epsilon = 10^{-8}$ in Kingma and Ba (2015)) which may result in a slow convergence.

Stepsize lower bound. In the case of Adam ($a_n = \frac{a}{\epsilon + \sqrt{v_n}}$), the uniform lower bound $a_{n+1} \geq \delta$ prevents the exponential moving average v_n of the squared gradients from exploding. This can be guaranteed on the fly by a clipping of a_n. If we drop the uniform lower bound on the effective stepsize, we still obtain the following result (see Remark 17)

$$\mathbb{E}\left[\sum_{k=0}^{n-1}\langle a_{k+1}, \nabla f(x_k, \xi_{k+1})^2\rangle\right] \leq \frac{2(1+\alpha)}{b^2\alpha}\left(\frac{H_0 - \inf f}{\varepsilon} + \langle a_0, p_0^2\rangle + \frac{n\bar{a}_{\sup}\sigma^2}{\varepsilon}\right).$$

Influence of the momentum parameter. Note that ε depends on the momentum parameter b and consequently the bound does not decrease with b. The influence of this parameter is more complex.

4.5 Convergence analysis under the KŁ property

Historically introduced by the fundamental works of Łojasiewicz (1963) and Kurdyka (1998), the KŁ inequality is the key tool of our analysis. We refer to Bolte et al. (2010) for an in-depth presentation of this property. The KŁ inequality is satisfied by a broad class of functions including most nonsmooth deep neural networks. More precisely, as exposed in Davis et al. (2020, Section 5.2, Corollary 5.11) and Castera et al. (2019,

Section 2.2), feedforward neural networks with arbitrary number of layers of arbitrary dimensions, with activations such as sigmoid, ReLU, leaky ReLU, tanh, softplus (and many others), with a loss function such as l_p norm, hinge loss, logistic loss or cross entropy (and many others), belong to this class of so-called *definable* functions in an *o-minimal structure* (Kurdyka, 1998; Attouch et al., 2010; Davis et al., 2020). We refer the interested reader to Zeng et al. (2019, Section 3, Section C) for general conditions for which KŁ inequality holds in the context of deep neural networks training models. The class of *definable* functions is stable under all the typical functional operations in optimization (e.g. sums, compositions, inf-projections) and generalizes the class of semialgebraic functions including objective functions such as $\|\cdot\|_p$ for p rational, real polynomials, rank, etc. (see Bolte et al. (2014, Appendix)).

The KŁ inequality has been used to show the convergence of several first-order optimization methods towards critical points (Attouch and Bolte, 2009; Attouch et al., 2010; Bolte et al., 2014; Li et al., 2017). In this section, we use a methodology exposed in Bolte et al. (2018, Appendix) to show convergence rates based on the KŁ property. Recently developed in Bolte et al. (2014), this abstract convergence mechanism can be used for any *descent* type algorithm. We modify it to encompass momentum methods. Note that although this modification was initiated in Ochs et al. (2014); Ochs (2018), we use a different separable Lyapunov function. The first part of the proof follows these approaches and the second part follows the proof of Johnstone and Moulin (2017, Th. 2).

Consider the function $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined for all $z = (x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ by

$$H(z) = H(x, y) = f(x) + \frac{1}{2b}\|y\|^2 . \tag{4.7}$$

Notice that $H_n = f(x_n) + \frac{1}{2b}\langle a_n, p_n^2\rangle = H(x_n, y_n)$ where $(y_n)_{n\in\mathbb{N}}$ is defined for all $n \in \mathbb{N}$ by $y_n = \sqrt{a_n}p_n$.

Notations and definitions. If (E, d) is a metric space, $z \in E$ and A is a non-empty subset of E, we use the notation $\mathsf{d}(z, A) := \inf\{\mathsf{d}(z, z') : z' \in A\}$. The set of critical points of the function H is defined by $\operatorname{crit} H := \{z \in \mathbb{R}^{2d} \text{ s.t. } \nabla H(z) = 0\}$.

Assumption 4.5.1. f is coercive, that is $f(x) \to +\infty$ as $\|x\| \to +\infty$.

Assumption 4.5.1 will be particularly useful to ensure that the sequence of the iterates $(z_k)_{k\geq 0}$ of Algorithm 5.1 is bounded. Indeed, a coercive function has compact level sets and Lem. 4.1 will guarantee that the iterates lie in a level set of the function H.

We now introduce the limit point set of the sequence $(z_k)_{k\geq 0}$ and exhibit some of its properties.

Definition 4.5.1. (Limit point set) The set of all limit points of $(z_k)_{k\in\mathbb{N}}$ initialized at z_0 is defined by

$$\omega(z_0) := \{\bar{z} \in \mathbb{R}^{2d} : \exists \text{ an increasing sequence of integers } (k_j)_{j\in\mathbb{N}} \text{ s.t } z_{k_j} \to \bar{z} \text{ as } j \to \infty\} .$$

Lemma 4.4. (Properties of the limit point set) Let $(z_k)_{k\in\mathbb{N}}$ be the sequence defined for all $k \in \mathbb{N}$ by $z_k = (x_k, y_k)$ where $y_k = \sqrt{a_k}p_k$ and (x_k, p_k) is generated by Algorithm 5.1 from a starting point z_0. Let Assumptions 4.2.1, 4.4.1, and 4.5.1 hold true. Assume that Condition (4.5) holds. Then,

(i) $\omega(z_0)$ is a nonempty compact set.

(ii) $\omega(z_0) \subset \text{crit}H = \text{crit}f \times \{0\}$.
(iii) $\lim_{k\to+\infty} \mathsf{d}(z_k, \omega(z_0)) = 0$.
(iv) H is finite and constant on $\omega(z_0)$.

We introduce the KŁ inequality in the following. Define $[\alpha < H < \beta] := \{z \in \mathbb{R}^{2d} : \alpha < H(z) < \beta\}$. Let $\eta > 0$ and define Φ_η as the set of continuous functions φ on $[0, \eta)$ which are also continuously differentiable on $(0, \eta)$, concave and satisfy $\varphi(0) = 0$ and $\varphi' > 0$.

Definition 4.5.2. (KŁ property, Bolte et al. (2018, Appendix)) A proper and lower semicontinuous (l.s.c) function $H : \mathbb{R}^{2d} \to (-\infty, +\infty]$ has the KŁ property locally at $\bar{z} \in \text{dom}\, H$ if there exist $\eta > 0$, $\varphi \in \Phi_\eta$ and a neighborhood $U(\bar{z})$ s.t. for all $z \in U(\bar{z}) \cap [H(\bar{z}) < H < H(\bar{z}) + \eta]$:

$$\varphi'(H(z) - H(\bar{z}))\, \|\nabla H(z)\| \geq 1\,. \tag{4.8}$$

When $H(\bar{z}) = 0$, we can rewrite Eq. (4.8) as: $\|\nabla(\varphi \circ H)(z)\| \geq 1$ for suitable z points. This means that H becomes sharp under a reparameterization of its values through the so-called desingularizing function φ.

The function H is said to be a KŁ function if it has the KŁ property at each point of the domain of its gradient. Note that this property can be defined for nonsmooth functions using the Clarke subdifferential in order to encompass nonsmooth neural networks. We limit ourserlves to the simpler differentiable setting. KŁ inequality holds at any non critical point (see Attouch et al. (2010, Remark 3.2 (b))). We introduce now a uniformized version of the KŁ property which will be useful for our analysis.

Lemma 4.5. (Uniformized KŁ property, Bolte et al. (2014, Lemma 6, p 478)) Let Ω be a compact set and let $H : \mathbb{R}^{2d} \to (-\infty, +\infty]$ be a proper l.s.c function. Assume that H is constant on Ω and satisfies the KŁ property at each point of Ω. Then, there exist $\varepsilon > 0, \eta > 0$ and $\varphi \in \Phi_\eta$ such that for all $\bar{z} \in \Omega$, for all $z \in \{z \in \mathbb{R}^d : \mathsf{d}(z, \Omega) < \varepsilon\} \cap [H(\bar{z}) < H < H(\bar{z}) + \eta]$, one has

$$\varphi'(H(z) - H(\bar{z}))\|\nabla H(z)\| \geq 1 \tag{4.9}$$

Definition 4.5.3. (KŁ exponent) If φ can be chosen as $\varphi(s) = \frac{\bar{c}}{\theta} s^\theta$ for some $\bar{c} > 0$ and $\theta \in (0, 1]$ in Def. 4.5.2, then we say that H has the KŁ property at $\bar{z}$ with an exponent of θ [2]. We say that H is a KŁ function with an exponent θ if it has the same exponent θ at any $\bar{z}$.

In the particular case when $\theta = 1/2$, we recover the Polyak-Łojasiewicz condition (see for e.g., Karimi et al. (2016)) satisfied for strongly convex functions. Furthermore, if H is a proper closed semialgebraic function, then H is a KŁ function with a suitable exponent $\theta \in (0, 1]$. The slope of φ around the origin informs about the "flatness" of a function around a point. Hence, the KŁ exponent allows to obtain convergence rates. In the light of this remark, we state one of the main results of this work.

Theorem 4.6. ***(Convergence rates)*** *Let $(z_k)_{k\in\mathbb{N}}$ be the sequence defined for all $k \in \mathbb{N}$ by $z_k = (x_k, y_k)$ where $y_k = \sqrt{a_k} p_k$ and (x_k, p_k) is generated by Algorithm 5.1 from a starting point z_0. Let Assumptions 4.2.1, 4.4.1 and 4.5.1 hold true. Assume that Condition (4.5) holds. Suppose moreover that H is a KŁ function with KŁ exponent θ. Then, the sequence $(H(z_k))_{k\in\mathbb{N}}$ converges to $f(x_*)$ where x_* is a critical point of f and the following convergence rates hold:*

[2] $\alpha := 1 - \theta$ is also defined as the KŁ exponent in other papers (Li and Pong, 2018).

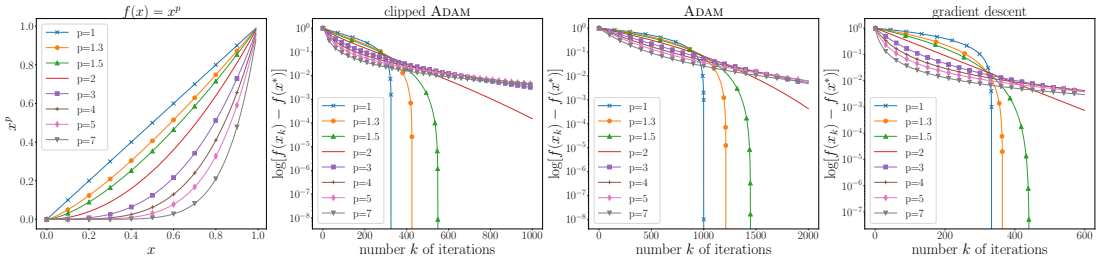

Figure 4.1 – Illustration of KŁ rates for a simple objective function $f(x) = x^p$. From left to right: (i) curves of $f(x) = x^p$, (ii) clipped version of ADAM (see Algorithm 5.1), (iii) ADAM ,(iv) Gradient Descent. Best seen in color.

(i) If $\theta = 1$, then $f(x_k)$ converges in a finite number of iterations.
(ii) If $1/2 \leq \theta < 1$, then $f(x_k)$ converges to $f(x_*)$ linearly i.e. there exist $q \in (0,1), C > 0$ s.t. $f(x_k) - f(x_*) \leq C\, q^k$.
(iii) If $0 < \theta < 1/2$, then $f(x_k) - f(x_*) = O(k^{\frac{1}{2\theta - 1}})$.

The exact same rates hold for gradient descent by supposing that f (instead of H) is KL with exponent θ. Assumption 4.4.1 and condition (4.5) are not needed in this case.

Sketch of the proof. The proof consists of two main steps. The first one is to show that the iterates enter and stay in a region where the KŁ inequality holds. This is achieved using the properties of the limit set (Lem. 4.4) and the uniformized KŁ property (Lem. 4.5). Then, the second step is to exploit this inequality to derive the sought convergence results. We defer the complete proof to Section 4.9.3.

We introduce a lemma in order to make the KŁ assumption on the objective function f instead of the function H.

Lemma 4.7. Let f be a continuously differentiable function satisfying the KL property at $\bar{x}$ with an exponent of $\theta \in (0, 1/2]$. Then the function H defined in Eq. (4.7) has also the KŁ property at $(\bar{x}, 0)$ with an exponent of θ .

The following result derives a convergence rate on the objective function values under a KŁ assumption on this same function instead of an assumption on the Lyapunov function H. The result is an immediate consequence of Lem. 4.7 and Th. 4.6.

Corollary 4.8. Let $(z_k)_{k \in \mathbb{N}}$ be the sequence defined for all $k \in \mathbb{N}$ by $z_k = (x_k, y_k)$ where $y_k = \sqrt{a_k} p_k$ and (x_k, p_k) is generated by Algorithm 5.1 from a starting point z_0. Let Assumptions 4.2.1, 4.4.1, 4.5.1 hold true. Assume that Condition (4.5) holds. Suppose moreover that f is a KŁ function with KŁ exponent $\theta \in (0, 1/2)$. Then, the sequence $(H(z_k))_{k \in \mathbb{N}}$ converges to $f(x_*)$ where x_* is a critical point of f and $f(x_k) - f(x_*) = O(k^{\frac{1}{2\theta - 1}})$.

Toy problem: KŁ rates for $f(x) = x^p$.

KŁ rates are *asymptotic* rates in the sense that the constants cannot be explicited in the convergence rates. As a consequence, the rates can be hardly observable in practice from experiments. However, we can still illustrate these convergence results (Th. 4.6)

in a simple toy example to give more insight. Consider the problem of minimizing the function $f(x) = x^p$ for a real $p \in [1, 7]$. One can easily show that f is a KŁ function with KŁ exponent $\theta = \frac{1}{p}$. Note that the KŁ exponent is difficult to compute in general. This justifies the choice of this toy problem. Moreover, even if the function f is indeed convex, we recall the reader that the KŁ property is a local geometric property of the function that is only interesting at its critical points (since it is automatically verified at any non critical point). Notice that the KŁ analysis is valid in the general non-convex case. The present toy example remains relevant if we modify the objective function f to be non-convex and still keep a x^p shape in a neighborhood of the point zero which is the unique critical point in this example.

The KŁ exponent as shown in the first plot in Figure 4.1 encodes information about the flatness of the function f. Indeed, as p increases, the function f gets flatter around the origin $x = 0$. We run the clipped version of ADAM (see Algorithm 5.1), the ADAM algorithm and gradient descent on the functions f corresponding to different values of the exponent θ, from the same initialization point $x = 1$. As expected from Th. 4.6 for the clipped ADAM, we observe in Figure 4.1 that $f(x_k)$ converges linearly or even in a finite number of iterations for $p \in \{1, 1.3, 1.5, 2\}$. Notice that the linear rate is clearly observable for $p = 2$ corresponding to $\theta = \frac{1}{2}$. Even if we did not establish KŁ rates for original ADAM, Figure 4.1 shows that it presents a very similar behavior to the clipped version of ADAM in terms of KŁ convergence rates in this simple problem. We also represent gradient descent iterates for comparison. Note that KŁ rates are known to hold for gradient descent. Moreover, for $p > 2$, we also observe a slower rate corresponding to the sublinear rate of the function values.

4.6 Conclusion

In this chapter, we provided convergence rates for a clipped version of ADAM which stems from a boundedness assumption on the effective stepsize of the original ADAM. More precisely, similarly to gradient descent, we established a $O(1/n)$ convergence rate of the minimum of the squared gradient norms in the deterministic case. Furthermore, we showed a similar convergence result in the stochastic setting up to the variance of the noisy gradients. Finally, we established function value convergence rates under the same boundedness assumption on the effective stepsizes together with the KŁ geometric property. This property is a powerful tool allowing to address non-convex nonsmooth optimization and covers most deep neural networks.

4.7 About theoretical guarantees of variants of ADAM.

We list most of the existing variants of the ADAM algorithm together with their theoretical convergence guarantees in Table 4.2.

Remark 16. The average regret bound result in the last line of Table 4.2 figures in Luo et al. (2019). Actually, according to Savarese (2019), slightly different assumptions on the bound functions should be considered to guarantee this regret rate.

Table 4.2 – **Theoretical guarantees of variants of ADAM.** The gradient is supposed L-lipschitz continuous in all the convergence results. $g_{1:T,i} = [g_{1,i}, g_{2,i}, \cdots, g_{T,i}]^T$.

Algorithm	**Effective stepsize** a_{n+1}	b_n	c_n	Assumptions	Convergence Result
AMSGRAD$^{(1)}$, ADAMNC$^{(2)}$ (Reddi et al., 2018)	$\frac{a_0}{\sqrt{n}}\frac{1}{\sqrt{\hat{v}_n}}$ $^{(1)}\hat{v}_{n+1} = \max(\hat{v}_n, (1-c_n)v_n + c_n g_n^2)$ $^{(2)}\hat{v}_{n+1} = (1-c_n)\hat{v}_n + c_n g_n^2$	$1 - b_1\lambda^{n-1}$ or $1 - \frac{b_1}{n}$	$c_n \equiv c_1$ $\frac{c_1}{n}$ (for ADAMNC)	• convex functions • bounded gradients • bounded feasible set • $\sum_{i=1}^d \hat{v}_{T,i}^{1/2} \leq d$ (AMSGRAD) • $\sum_{i=1}^d \|g_{1:T,i}\|_2 \leq \sqrt{dT}$ • $b_1 < \sqrt{c_1}$ (ADAMNC)	$R_T/T = O(\sqrt{logT/T})$ $\frac{R_T}{T} = O(1/\sqrt{T})$ (ADAMNC)
ADAM (Basu et al., 2018)	$\frac{4\|g_n\|^2\eta}{3L(1-(1-b)^n)^2(\eta+2\sigma)^3}\frac{1}{\epsilon+\sqrt{\hat{v}_n}}$ $v_{n+1} = (1-c_1)v_n + c_1 g_n^2$	$b_n \equiv b_1$ $= 1 - \frac{n}{\eta+2\sigma}$	$c_n \equiv c_1$	• σ-bounded gradients • $\epsilon = 2\sigma$	$\forall \eta > 0\ \exists n \leq \frac{9L\sigma^2(f(x_2)-f(x_*))}{\eta^6}$ s.t. $\|g_n\| \leq \eta$
PADAM,AMSGRAD (Zhou et al., 2018)	$\frac{1}{\sqrt{N}}\frac{1}{\hat{v}_n^p}$ $\frac{1}{\sqrt{dN}}\frac{1}{\hat{v}_n^p}$ (AMSGRAD) $\hat{v}_n = \max(\hat{v}_{n-1}, (1-c)v_{n-1} + cg_n^2)$	$b_n \equiv b$	$c_n \equiv c$	• bounded gradients For PADAM: • $p \in [0, \frac{1}{4}]$ • $1 - b < (1-c)^{2p}$ • $\sum_{i=1}^d \|g_{1:N,i}\|_2 \leq \sqrt{dN}$ AMSGRAD: $p = \frac{1}{2}$ and $1 - b < 1 - c$	$\mathbb{E}[\|g_\tau\|^2] = O(\frac{1+\sqrt{d}}{\sqrt{N}} + \frac{d}{N})$ $= O(\sqrt{\frac{d}{N}} + \frac{d}{N})$ (AMSGRAD) τ uniform r.v in $\{1, \cdots, N\}$
RMSPROP$^{(1)}$, YOGI$^{(2)}$ (Zaheer et al., 2018)	$\frac{a_1}{\epsilon+\sqrt{v_n}}$ $^{(1)}v_{n+1} = (1-c)v_n + cg_n^2$ $^{(2)}v_n = v_{n-1} - \text{csign}(v_{n-1} - g_n^2)$	$b_n \equiv b$	$c_n \equiv c$	• G-bounded gradients • $a_1 \leq \frac{\epsilon\sqrt{1-c}}{2L}$ (YOGI) • $a_1 \leq \frac{\epsilon}{2L}$ • $c \leq \frac{\epsilon^2}{16G^2}$ • σ^2-bounded variance	$\mathbb{E}[\|g_\tau\|^2] = O(\frac{1}{N} + \sigma^2)$ τ uniform r.v in $\{1, \cdots, N\}$ $O(\frac{1}{N})$ if minibatch $\Theta(N)$
AMSGRAD$^{(1)}$, ADAFOM$^{(2)}$ (Chen et al., 2019)	$\frac{1}{\sqrt{n}}\frac{1}{\sqrt{\hat{v}_n}}$ $^{(1)}\hat{v}_{n+1} = \max(\hat{v}_n, (1-c_n)v_n + c_n g_n^2)$ $^{(2)}\hat{v}_{n+1} = (1-\frac{1}{n})\hat{v}_n + \frac{1}{n}g_n^2$	non-increasing	$c_n \equiv c_1$	• bounded gradients • $\exists c > 0$ s.t. $\|g_{1,i}\| \geq c$	$\min_{n\in[0,N]} \mathbb{E}[\|g_n\|^2] = O(\frac{logN+d^2}{\sqrt{N}})$
GENERIC ADAM (Zou et al., 2019a)	$\frac{\alpha_n}{\sqrt{\hat{v}_n}}$ $v_{n+1} = (1-c_n)v_n + c_n g_n^2$ $\alpha_n = \hat{\alpha}\frac{\sqrt{1-(1-c)^n}}{1-(1-b)^n}$	$b_n \geq b > 0$	$0 < c_n < 1$ non-increasing $\lim c_n = c > b^2$	• bounded gradients in expectation • $d_n \leq \frac{\alpha_n}{\sqrt{c_n}} \leq c_0 d_n$ d_n non-increasing	$\mathbb{E}[\|g_\tau\|_1^{\frac{3}{2}}]^{\frac{2}{3}} \leq \frac{C+C'\sum_{n=1}^N \alpha_n\sqrt{c_n}}{N\alpha_N}$ τ uniform r.v in $\{1, \cdots, N\}$
ADABOUND$^{(1)}$, AMSBOUND$^{(2)}$ (Luo et al., 2019)	$\frac{1}{\sqrt{n}}\text{clip}(\frac{\alpha}{\sqrt{\hat{v}_n}}, \eta_l(n), \eta_u(n))$ $\eta_l(n)$ non-decreasing to α_* $\eta_u(n)$ non-increasing to α_* $^{(1)}v_{n+1} = (1-c)v_n + cg_n^2$ $^{(2)}v_{n+1} = \max(v_n, (1-c)v_n + cg_n^2)$	$1 - (1-b)\lambda^{n-1}$ or $1 - \frac{1-b}{n}$ $b_n \geq b$	$c_n \equiv c$	• bounded gradients • closed convex bounded feasible set • $1 - b < \sqrt{1-c}$	$R_T/T = O(1/\sqrt{T})$

4.8 Proofs for Section 4.4

4.8.1 Proof of Lem. 4.1

Supposing that ∇f is $L-$Lipschitz, using Taylor's expansion and the expression of p_n in the algorithm, we obtain the following inequality:

$$f(x_{n+1}) \leq f(x_n) - \langle \nabla f(x_n), a_{n+1}p_{n+1}\rangle + \frac{L}{2}\|a_{n+1}p_{n+1}\|^2 \tag{4.10}$$

Moreover,

$$\frac{1}{2b}\langle a_{n+1}, p_{n+1}^2\rangle - \frac{1}{2b}\langle a_n, p_n^2\rangle = \frac{1}{2b}\langle a_{n+1}, p_{n+1}^2 - p_n^2\rangle + \frac{1}{2b}\langle a_{n+1} - a_n, p_n^2\rangle. \tag{4.11}$$

Observing that $p_{n+1}^2 - p_n^2 = -b^2(\nabla f(x_n) - p_n)^2 + 2bp_{n+1}(\nabla f(x_n) - p_n)$, we obtain after simplification:

$$H_{n+1} \leq H_n + \frac{L}{2}\|a_{n+1}p_{n+1}\|^2 - \frac{b}{2}\langle a_{n+1}, (\nabla f(x_n) - p_n)^2\rangle - \langle a_{n+1}p_{n+1}, p_n\rangle + \frac{1}{2b}\langle a_{n+1} - a_n, p_n^2\rangle. \tag{4.12}$$

Using again $p_n = p_{n+1} - b(\nabla f(x_n) - p_n)$, we replace p_n :

$$\begin{aligned} H_{n+1} \leq H_n &+ \frac{L}{2}\|a_{n+1}p_{n+1}\|^2 - \frac{b}{2}\langle a_{n+1}, (\nabla f(x_n) - p_n)^2\rangle \\ &- \langle a_{n+1}, p_{n+1}^2\rangle + b\langle a_{n+1}p_{n+1}, \nabla f(x_n) - p_n\rangle + \frac{1}{2b}\langle a_{n+1} - a_n, p_n^2\rangle. \end{aligned}$$

Under Assumption 4.4.1, we write: $\langle a_{n+1} - a_n, p_n^2\rangle \leq (1-\alpha)\langle a_{n+1}, p_n^2\rangle$ and using $p_n^2 = p_{n+1}^2 + b^2(\nabla f(x_n) - p_n)^2 - 2bp_{n+1}(\nabla f(x_n) - p_n)$, it holds that:

$$\begin{aligned} H_{n+1} \leq H_n &- \langle a_{n+1}, p_{n+1}^2\rangle - \frac{b}{2}\langle a_{n+1}, (\nabla f(x_n) - p_n)^2\rangle \\ &+ \frac{L}{2}\|a_{n+1}p_{n+1}\|^2 + (b - (1-\alpha))\langle a_{n+1}p_{n+1}, \nabla f(x_n) - p_n\rangle \\ &+ \frac{1-\alpha}{2b}\langle a_{n+1}, p_{n+1}^2\rangle + \frac{b(1-\alpha)}{2}\langle a_{n+1}, (\nabla f(x_n) - p_n)^2\rangle. \end{aligned}$$

Using the classical inequality $xy \leq \frac{x^2}{2u} + \frac{uy^2}{2}$, we have:

$$(b-(1-\alpha))a_{n+1}p_{n+1}(\nabla f(x_n) - p_n) \leq \frac{|b-(1-\alpha)|}{2u}\langle a_{n+1}, p_{n+1}^2\rangle + \frac{|b-(1-\alpha)|u}{2}\langle a_{n+1}, (\nabla f(x_n) - p_n)^2\rangle. \tag{4.13}$$

Hence, after using this inequality and rearranging the terms, we derive the following inequality:

$$\begin{aligned} H_{n+1} \leq H_n &- \langle a_{n+1}p_{n+1}^2, 1 - \frac{a_{n+1}L}{2} - \frac{|b-(1-\alpha)|}{2u} - \frac{1-\alpha}{2b}\rangle \\ &- \frac{b}{2}\langle a_{n+1}(\nabla f(x_n) - p_n)^2, \left(1 - \frac{|b-(1-\alpha)|u}{b} - (1-\alpha)\right)\mathbf{1}\rangle. \end{aligned}$$

This concludes the proof.

4.8.2 A first result under an upperbound of the stepsize

Proposition 4.9. Let Assumption 4.2.1 hold true. Suppose moreover that $1-\alpha < b \leq 1$. Let $\varepsilon > 0$ s.t. $a_{\sup} := \frac{2}{L}\left(1 - \frac{(b-(1-\alpha))^2}{2b\alpha} - \frac{1-\alpha}{2b} - \varepsilon\right)$ is nonnegative. Assume for all $n \in \mathbb{N}$,

$$a_{n+1} \leq \min\left(a_{\sup}, \frac{a_n}{\alpha}\right).$$

Then, for all $n \geq 1$,

$$\sum_{k=0}^{n-1} \langle a_{k+1}, \nabla f(x_k)^2 \rangle \leq \frac{2(1+\alpha)}{b^2\alpha}\left(\frac{H_0 - \inf f}{\varepsilon} + \langle a_0, p_0^2 \rangle\right)$$

Proof. This is a consequence of Lem. 4.1. Conditions $A_{n+1} \geq \varepsilon$ and $B \geq 0$ write as follow:

$$a_{n+1} \leq \frac{2}{L}\left(1 - \frac{b-(1-\alpha)}{2u} - \frac{1-\alpha}{2b} - \varepsilon\right) \quad \text{and} \quad u \leq \frac{\alpha b}{b-(1-\alpha)}.$$

We get the assumption made in the proposition by injecting the second condition into the first one and adding the assumption $\frac{a_{n+1}}{a_n} \leq \frac{1}{\alpha}$ made in the lemma. Under this assumption, we sum over $0 \leq k \leq n-1$ Eq. (4.4), rearrange it and use $A_{n+1} \geq \varepsilon$, $B \geq 0$ to obtain:

$$\sum_{k=0}^{n-1} \varepsilon \, \langle a_{k+1}, p_{k+1}^2 \rangle \leq H_0 - H_n \,,$$

Then, observe that $H_n \geq f(x_n) \geq \inf f$. Therefore, we derive:

$$\sum_{k=0}^{n-1} \langle a_{k+1}, p_{k+1}^2 \rangle \leq \frac{H_0 - \inf f}{\varepsilon}. \tag{4.14}$$

Moreover, from the Algorithm 5.1 second update rule, we get $\nabla f(x_k) = \frac{1}{b}p_{k+1} - \frac{1-b}{b}p_k$. Hence, we have for all $k \geq 0$:

$$\nabla f(x_k)^2 \leq 2\left(\frac{1}{b^2}p_{k+1}^2 + \frac{(1-b)^2}{b^2}p_k^2\right) \leq \frac{2}{b^2}(p_{k+1}^2 + p_k^2).$$

We deduce that:

$$\begin{aligned}
\sum_{k=0}^{n-1}\langle a_{k+1}, \nabla f(x_k)^2\rangle &\leq \frac{2}{b^2}\sum_{k=0}^{n-1}\langle a_{k+1}, p_{k+1}^2 + p_k^2\rangle \\
&= \frac{2}{b^2}\sum_{k=0}^{n-1}\langle a_{k+1}, p_{k+1}^2\rangle + \frac{2}{b^2}\sum_{k=0}^{n-1}\langle a_{k+1}, p_k^2\rangle \\
&\leq \frac{2}{b^2}\sum_{k=0}^{n-1}\langle a_{k+1}, p_{k+1}^2\rangle + \frac{2}{b^2\alpha}\sum_{k=0}^{n-1}\langle a_k, p_k^2\rangle \\
&\leq \frac{2}{b^2}(1+\frac{1}{\alpha})\sum_{k=0}^{n}\langle a_k, p_k^2\rangle \\
&\leq \frac{2(1+\alpha)}{b^2\alpha}\left(\frac{H_0 - \inf f}{\varepsilon} + \langle a_0, p_0^2\rangle\right).
\end{aligned}$$

■

4.8.3 Proof of Th. 4.2

This is a consequence of Lem. 4.1. Conditions $A_{n+1} \geq \varepsilon$ and $B \geq 0$ write as follow:

$$a_{n+1} \leq \frac{2}{L}\left(1 - \frac{b-(1-\alpha)}{2u} - \frac{1-\alpha}{2b} - \varepsilon\right) \quad \text{and} \quad u \leq \frac{\alpha b}{b-(1-\alpha)}\,.$$

We get the assumption made in the proposition by injecting the second condition into the first one and adding the assumption $\frac{a_{n+1}}{a_n} \leq \alpha$ made in the lemma. Under this assumption, we sum over $0 \leq k \leq n-1$ Eq. (4.4), rearrange it and use $A_{n+1} \geq \varepsilon$, $B \geq 0$ and $a_{k+1} \geq \delta$ to obtain:

$$\sum_{k=0}^{n-1} \delta\,\varepsilon\,\|p_{k+1}\|^2 \leq H_0 - H_n\,,$$

Then, observe that $H_n \geq f(x_n) \geq \inf f$. Therefore, we derive:

$$\sum_{k=0}^{n-1}\|p_{k+1}\|^2 \leq \frac{H_0 - \inf f}{\delta\varepsilon}\,. \tag{4.15}$$

Moreover, from the algorithm 5.1 second update rule, we get $\nabla f(x_k) = \frac{1}{b}p_{k+1} - \frac{1-b}{b}p_k$. Hence, we have for all $k \geq 0$:

$$\|\nabla f(x_k)\|^2 \leq 2\left(\frac{1}{b^2}\|p_{k+1}\|^2 + \frac{(1-b)^2}{b^2}\|p_k\|^2\right) \leq \frac{2}{b^2}(\|p_{k+1}\|^2 + \|p_k\|^2)\,.$$

We deduce that:

$$\sum_{k=0}^{n-1} \|\nabla f(x_k)\|^2 \leq \frac{2}{b^2} \sum_{k=0}^{n-1} (\|p_{k+1}\|^2 + \|p_k\|^2) = \frac{2}{b^2} \left(2 \sum_{k=1}^{n-1} \|p_k\|^2 + \|p_n\|^2 + \|p_0\|^2 \right) \leq \frac{4}{b^2} \sum_{k=0}^{n} \|p_k\|^2 . \quad (4.16)$$

Finally, using Eqs. (4.15)-(4.16), we have:

$$\min_{0 \leq k \leq n-1} \|\nabla f(x_k)\|^2 \leq \frac{1}{n} \sum_{k=0}^{n-1} \|\nabla f(x_k)\|^2 \leq \frac{4}{nb^2} \left(\frac{H_0 - \inf f}{\delta \varepsilon} + \|p_0\|^2 \right) .$$

4.8.4 Proof of Th. 4.3

The proof of this proposition mainly follows the same path as its deterministic counterpart. However, due to stochasticity, a residual term (the last term in Eq. (4.17)) quantifying the difference between the stochastic gradient estimate and the true gradient of the objective function (compare Eq. (4.17) to Lem. 4.1) remains. Following the exact same steps of Section 4.8.1, we obtain by replacing the deterministic gradient $\nabla f(x_n)$ by its stochastic estimate $\nabla f(x_n, \xi_{n+1})$:

$$\begin{aligned} H_{n+1} \leq H_n &- \langle a_{n+1} p_{n+1}^2, 1 - \frac{a_{n+1} L}{2} - \frac{|b - (1-\alpha)|}{2u} - \frac{1-\alpha}{2b} \rangle \\ &- \frac{b}{2} \langle a_{n+1} (\nabla f(x_n, \xi_{n+1}) - p_n)^2, \left(1 - \frac{|b - (1-\alpha)| u}{b} - (1-\alpha) \right) \mathbf{1} \rangle \\ &+ \langle \nabla f(x_n, \xi_{n+1}) - \nabla F(x_n), a_{n+1} p_{n+1} \rangle . \end{aligned} \quad (4.17)$$

Using the classical inequality $xy \leq \frac{x^2}{2\eta} + \frac{\eta y^2}{2}$ with $\eta = 1/2$ and the almost sure boundedness of the stepsize a_{n+1}, we get:

$$\begin{aligned} \langle \nabla f(x_n, \xi_{n+1}) - \nabla F(x_n), a_{n+1} p_{n+1} \rangle &\leq \langle (\nabla f(x_n, \xi_{n+1}) - \nabla F(x_n))^2 + \frac{1}{4} p_{n+1}^2, a_{n+1} \rangle \\ &\leq \bar{a}_{\sup} \|\nabla f(x_n, \xi_{n+1}) - \nabla F(x_n)\|^2 + \frac{1}{4} \langle a_{n+1}, p_{n+1}^2 \rangle . \end{aligned}$$

Therefore, taking the expectation and using the boundedness of the variance, we obtain from Eq. (4.17):

$$\mathbb{E}[H_{n+1}] - \mathbb{E}[H_n] \leq -\mathbb{E} \left[\langle a_{n+1} p_{n+1}^2, \frac{3}{4} - \frac{a_{n+1} L}{2} - \frac{|b - (1-\alpha)|}{2u} - \frac{1-\alpha}{2b} \rangle \right] + \bar{a}_{\sup} \sigma^2 .$$

Then, the proof follows the lines of Section 4.8.2. Hence, we have

$$\mathbb{E}[H_{n+1}] - \mathbb{E}[H_n] \leq -\mathbb{E} \left[\langle a_{n+1} p_{n+1}^2, \varepsilon \mathbf{1} \rangle \right] + \bar{a}_{\sup} \sigma^2 .$$

We sum these inequalities for $k = 0, \cdots, n-1$, inject the assumption $a_{n+1} \geq \delta$ and rearrange the terms to obtain

$$\delta\, \mathbb{E}\left[\sum_{k=0}^{n-1} \|p_{k+1}\|^2\right] \leq \mathbb{E}\left[\sum_{k=0}^{n-1} \langle a_{k+1}, p_{k+1}^2 \rangle\right] \leq \frac{H_0 - \inf f}{\varepsilon} + \frac{n\bar{a}_{\sup}\sigma^2}{\varepsilon}\,. \tag{4.18}$$

Then, using $\nabla f(x_k, \xi_{k+1}) = \frac{1}{b} p_{k+1} - \frac{1-b}{b} p_k$ and a similar upperbound to Eq. (4.16) we show that

$$\sum_{k=0}^{n-1} \|\nabla f(x_k, \xi_{k+1})\|^2 \leq \frac{4}{b^2} \sum_{k=0}^{n} \|p_k\|^2\,. \tag{4.19}$$

Therefore, combining Eqs. (4.19)-(4.18), we establish the following inequality

$$\mathbb{E}\left[\sum_{k=0}^{n-1} \|\nabla f(x_k, \xi_{k+1})\|^2\right] \leq \frac{4}{b^2}\left(\frac{H_0 - \inf f}{\delta\varepsilon} + \|p_0\|^2\right) + \frac{4\bar{a}_{\sup} n}{\delta\varepsilon b^2}\sigma^2\,.$$

Finally, we apply Jensen's inequality to $\|\cdot\|^2$ and divide the previous inequality by n to obtain the sought result

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{E}\left[\|\nabla F(x_k)\|^2\right] \leq \frac{4}{n\delta b^2}\left(\frac{H_0 - \inf f}{\delta\varepsilon} + \|p_0\|^2\right) + \frac{4\bar{a}_{\sup}}{\delta\varepsilon b^2}\sigma^2\,.$$

Remark 17. Following the derivations in Section 4.8.2, note that we also obtain the following result

$$\mathbb{E}\left[\sum_{k=0}^{n-1} \langle a_{k+1}, \nabla f(x_k, \xi_{k+1})^2 \rangle\right] \leq \frac{2(1+\alpha)}{b^2\alpha}\left(\frac{H_0 - \inf f}{\varepsilon} + \langle a_0, p_0^2\rangle + \frac{n\bar{a}_{\sup}\sigma^2}{\varepsilon}\right)\,.$$

4.8.5 Comparison to Ochs et al. (2014)

We recall the conditions satisfied by α_n and β_n in Ochs et al. (2014) in order to traduce them in terms of the algorithm (5.1) at stake. Define:

$$\delta_n := \frac{1}{\alpha_n} - \frac{L}{2} - \frac{\beta_n}{2\alpha_n} \qquad \gamma_n := \delta_n - \frac{\beta_n}{2\alpha_n}.$$

Conditions of Ochs et al. (2014) write: $\alpha_n \geq c_1$ $\beta_n \geq 0$ $\delta_n \geq \gamma_n \geq c_2$ where c_1, c_2 are positive constants and (δ_n) is monotonically decreasing.

One can remark that Algorithm 5.1 can be written as (4.3) with stepsizes $\alpha_n = b a_{n+1}$ and inertial parameters $\beta_n = (1-b)\frac{a_{n+1}}{a_n}$. Conditions on these parameters can be expressed in terms of a_n. Supposing $c_2 = 0$, the condition $\gamma_n \geq c_2$ is equivalent to

$$\frac{a_{n+1}}{a_n} \leq \frac{2}{2 - b(2 - a_n L)}. \tag{4.20}$$

Note that the classical condition $a_n \leq 2/L$ shows up consequently. Moreover, the condition on (δ_n) is equivalent to

$$\frac{1}{a_{n+1}} \leq \frac{3-b}{2}\frac{1}{a_n} - \frac{1-b}{2a_{n-1}} \qquad \text{for} \qquad n \geq 1. \tag{4.21}$$

Note that we get rid of condition (4.21) while allowing adaptive stepsizes a_n (see Prop. 4.9).

4.8.6 Performance of gradient descent in the non-convex setting.

In the non-convex setting, for a smooth function f, we cannot say anything about the convergence rate of the sequences $(f(x_k))$ and (x_k). Nevertheless, as exposed in (Nesterov, 2004, p.28), we can control the minimum of the gradients norms. We prove this result in the following for completeness.

Consider the gradient descent algorithm defined by: $x_{k+1} = x_k - \gamma \nabla f(x_k)$. Assume that $\gamma > 0$ and $1 - \frac{\gamma L}{2} > 0$.

Supposing that ∇f is $L-$Lipschitz, using Taylor's expansion and regrouping the terms, we obtain the following inequality:

$$f(x_{k+1}) \leq f(x_k) - \gamma \left(1 - \frac{\gamma L}{2}\right) \|\nabla f(x_k)\|_2^2.$$

Then, we sum the inequalities for $0 \leq k \leq n-1$, lower bound the gradients norms in the sum by their minimum and we obtain for $n \geq 1$:

$$\min_{0 \leq k \leq n-1} \|\nabla f(x_k)\|_2^2 \leq \frac{f(x_0) - \inf f}{n\gamma(1 - \frac{\gamma L}{2})}.$$

4.9 Proofs for Section 4.5

4.9.1 Three abstract conditions

Inspired from the abstract convergence mechanism of Bolte et al. (2018, Appendix), we show that similar conditions hold in our case. We highlight that these conditions are slightly different here, since we do not deal with *gradient-like descent sequences* (for which the objective function is nonincreasing over the iterations). Conditions below are closer to those of Ochs et al. (2014) which studies a non-descent algorithm. Note however that the Lyapunov function H and the sequence (z_k) we consider are different.

Lemma 4.10. Let $(z_k)_{k\in\mathbb{N}}$ be the sequence defined for all $k \in \mathbb{N}$ by $z_k = (x_k, y_k)$ where $y_k = \sqrt{a_k} p_k$ and (x_k, p_k) is generated by Algorithm 5.1 from a starting point z_0. Let Assumptions 4.2.1 and 4.4.1 hold true. Assume moreover that condition (4.5) holds. Then,

(i) (sufficient decrease property) There exists a positive scalar ρ_1 s.t.:

$$H(z_{k+1}) - H(z_k) \leq -\rho_1 \|x_{k+1} - x_k\|^2 \quad \forall k \in \mathbb{N}.$$

(ii) There exists a positive scalar ρ_2 s.t.:

$$\|\nabla H(z_{k+1})\| \leq \rho_2 \left(\|x_{k+1} - x_k\| + \|x_k - x_{k-1}\|\right) \quad \forall k \geq 1.$$

(iii) (continuity condition) If $\bar{z}$ is a limit point of a subsequence $(z_{k_j})_{j\in\mathbb{N}}$, then $\lim_{j\to+\infty} H(z_{k_j}) = H(\bar{z})$.

Remark 18. Note that the conditions in Lem. 4.10 can be generalized to a nonsmooth objective function. Indeed, in Bolte et al. (2018, Appendix), the Fréchet subdifferential replaces the gradient.

Proof.

(i) From Lem. 4.1 and Th. 4.2, we get for all $k \in \mathbb{N}$:

$$H(z_{k+1}) - H(z_k) \leq -\varepsilon\langle a_{k+1}, p_{k+1}^2\rangle \leq -\varepsilon\langle a_{k+1}, \left(\frac{x_{k+1} - x_k}{-a_{k+1}}\right)^2 \rangle \leq -\frac{\varepsilon}{a_{\sup}} \|x_{k+1} - x_k\|^2.$$

We set $\rho_1 := \frac{\varepsilon}{a_{\sup}}$.

(ii) First, observe that for all $k \in \mathbb{N}$

$$\|\nabla H(z_{k+1})\| \leq \|\nabla f(x_{k+1})\| + \frac{1}{b} \|y_{k+1}\|. \tag{4.22}$$

Now, let us upperbound each one of these two terms. Recall that we can rewrite our algorithm under a "Heavy-ball"-like form as follows:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) + \beta_k (x_k - x_{k-1}) \quad \forall k \geq 1.$$

where $\alpha_k := b a_{k+1}$ and $\beta_k = (1-b)\frac{a_{k+1}}{a_k}$ are vectors.

On the one hand, using the L-Lipschitz continuity of the gradient, we obtain

$$\begin{aligned}\|\nabla f(x_{k+1})\|^2 &\leq 2\left(\|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 + \|\nabla f(x_k)\|^2\right) \\ &\leq 2\left(L^2 \|x_{k+1} - x_k\|^2 + \|\nabla f(x_k)\|^2\right)\end{aligned}$$

Moreover,

$$\begin{aligned}\|\nabla f(x_k)\|^2 &= \left\| \frac{x_k - x_{k+1}}{\alpha_k} + \frac{\beta_k}{\alpha_k}(x_k - x_{k-1}) \right\|^2 \\ &\leq 2 \left\| \frac{x_k - x_{k+1}}{b a_{k+1}} \right\|^2 + 2 \left\| \frac{1-b}{b} \frac{1}{a_k} (x_k - x_{k-1}) \right\|^2 \\ &\leq \frac{2}{b^2\delta^2} \|x_{k+1} - x_k\|^2 + \frac{2(1-b)^2}{b^2\delta^2} \|x_k - x_{k-1}\|^2 \\ &\leq \frac{2}{b^2\delta^2} \left(\|x_{k+1} - x_k\|^2 + \|x_k - x_{k-1}\|^2\right).\end{aligned}$$

Hence,

$$\begin{aligned}\|\nabla f(x_{k+1})\|^2 &\leq 2\left(L^2\|x_{k+1}-x_k\|^2+\|\nabla f(x_k)\|^2\right)\\ &\leq 2\left(L^2+\frac{2}{b^2\delta^2}\right)\|x_{k+1}-x_k\|^2+\frac{4}{b^2\delta^2}\|x_k-x_{k-1}\|^2\\ &\leq 2\left(L^2+\frac{2}{b^2\delta^2}\right)(\|x_{k+1}-x_k\|^2+\|x_k-x_{k-1}\|^2)\,.\end{aligned}$$

Therefore, the following inequality holds:

$$\|\nabla f(x_{k+1})\| \leq \sqrt{2\left(L^2+\frac{2}{b^2\delta^2}\right)}(\|x_{k+1}-x_k\|+\|x_k-x_{k-1}\|)\,.$$

On the otherhand,

$$\|y_{k+1}\| = \|\sqrt{a_{k+1}}p_{k+1}\| = \left\|\frac{x_{k+1}-x_k}{\sqrt{a_{k+1}}}\right\| \leq \frac{1}{\sqrt{\delta}}\|x_{k+1}-x_k\|\,.$$

Finally, combining the inequalities for both terms in Eq. (4.22), we obtain

$$\|\nabla H(z_{k+1})\| \leq \rho_2(\|x_{k+1}-x_k\|+\|x_k-x_{k-1}\|) \quad \forall k\geq 1\,.$$

with $\rho_2 := \left(\sqrt{2\left(L^2+\frac{2}{b^2\delta^2}\right)}+\frac{1}{b\sqrt{\delta}}\right)$.

(iii) This is a consequence of the continuity of H.

■

4.9.2 Proof of Lem. 4.4

(i) By Th. 4.2, the sequence $(H(z_n))_{n\in\mathbb{N}}$ is nonincreasing. Therefore, for all $n\in\mathbb{N}$, $H(z_n)\leq H(z_0)$ and hence $z_n\in\{z : H(z)\leq H(z_0)\}$. Since f is coercive, H is also coercive and its level sets are bounded. As a consequence, $(z_n)_{n\in\mathbb{N}}$ is bounded and there exist $z_*\in\mathbb{R}^d$ and a subsequence $(z_{k_j})_{j\in\mathbb{N}}$ s.t. $z_{k_j}\to z_*$ as $j\to\infty$. Hence, $\omega(z_0)\neq\emptyset$. Furthermore, $\omega(z_0)=\bigcap_{q\in\mathbb{N}}\overline{\bigcup_{k\geq q}\{z_k\}}$ is compact as an intersection of compact sets.

(ii) First, $\mathrm{crit}H = \mathrm{crit}f\times\{0\}$ because $\nabla H(z)=(\nabla f(x), y/b)^T$. Let $z_*\in\omega(z_0)$. Recall that $x_{k+1}-x_k\to 0$ as $k\to\infty$ by Th. 4.2. We deduce from the second assertion of Lem. 4.10 that $\nabla H(z_k)\to 0$ as $k\to\infty$. As $z_*\in\omega(z_0)$, there exists a subsequence $(z_{k_j})_{j\in\mathbb{N}}$ converging to z_*. Then, by Lipschitz continuity of ∇H, we get that $\nabla H(z_{k_j})\to\nabla H(z_*)$ as $j\to\infty$. Finally, $\nabla H(z_*)=0$ since $\nabla H(z_k)\to 0$ and $(\nabla H(z_{k_j}))_{j\in\mathbb{N}}$ is a subsequence of $(\nabla H(z_n))_{n\in\mathbb{N}}$.

(iii) This point stems from the definition of limit points. Every subsequence of the sequence $(\mathsf{d}(z_k, \omega(z_0)))_{k\in\mathbb{N}}$ converges to zero as a consequence of the definition of $\omega(z_0)$.

(iv) The sequence $(H(z_n))_{n\in\mathbb{N}}$ is nonincreasing by Th. 4.2. It is also bounded from below because $H(z_k) \geq f(x_k) \geq \inf f$ for all $k \in \mathbb{N}$. Hence we can denote by l its limit. Let $\bar{z} \in \omega(z_0)$. There there exists a subsequence $(z_{k_j})_{j\in\mathbb{N}}$ converging to $\bar{z}$ as $j \to \infty$. By the third assertion of Lem. 4.10, $\lim_{j\to+\infty} H(z_{k_j}) = H(\bar{z})$. Hence this limit equals l since $(H(z_n))_{n\in\mathbb{N}}$ converges towards l. Therefore, the restriction of H to $\omega(z_0)$ equals l.

4.9.3 Proof of Th. 4.6

The first step of this proof follows the same path as Bolte et al. (2018, Proof of Th. 6.2, Appendix). Since f is coercive, H is also coercive. The sequence $(H(z_k))_{k\in\mathbb{N}}$ is nonincreasing. Hence, (z_k) is bounded and there exists a subsequence $(z_{k_q})_{q\in\mathbb{N}}$ and $\bar{z} \in \mathbb{R}^{2d}$ s.t. $z_{k_q} \to \bar{z}$ as $q \to \infty$. Then, since $(H(z_k))_{k\in\mathbb{N}}$ is nonincreasing and lowerbounded by $\inf f$, it is convergent and we obtain by continuity of H,

$$\lim_{k\to+\infty} H(z_k) = H(\bar{z}). \tag{4.23}$$

Using Th. 4.2, observe that the sequence (y_k) converges to zero since (a_k) is bounded and $p_k \to 0$. If there exists $\bar{k} \in \mathbb{N}$ s.t. $H(z_{\bar{k}}) = H(\bar{z})$, then $H(z_{\bar{k}+1}) = H(\bar{z})$ and by the first point of Lem. 4.10, $x_{\bar{k}+1} = x_{\bar{k}}$ and then $(x_k)_{k\in\mathbb{N}}$ is stationary and for all $k \geq \bar{k}$, $H(z_k) = H(\bar{z})$ and the results of the theorem hold in this case (note that $\bar{z} \in \mathrm{crit}H$ by Lem. 4.4). Therefore, we can assume now that $H(\bar{z}) < H(z_k) \forall k > 0$ since $(H(z_k))_{k\in\mathbb{N}}$ is nonincreasing and Eq. (4.23) holds. One more time, from Eq. (4.23), we have that for all $\eta > 0$, there exists $k_0 \in \mathbb{N}$ s.t. $H(z_k) < H(\bar{z}) + \eta$ for all $k > k_0$. From Lem. 4.4, we get $\mathsf{d}(z_k, \omega(z_0)) \to 0$ as $k \to +\infty$. Hence, for all $\varepsilon > 0$, there exists $k_1 \in \mathbb{N}$ s.t. $\mathsf{d}(z_k, \omega(z_0)) < \varepsilon$ for all $k > k_1$. Moreover, $\omega(z_0)$ is a nonempty compact set and H is finite and constant on it. Therefore, we can apply the uniformization Lem. 4.5 with $\Omega = \omega(z_0)$. Hence, for any $k > l := \max(k_0, k_1)$, we get

$$\varphi'(H(z_k) - H(\bar{z}))^2 \|\nabla H(z_k)\|^2 \geq 1. \tag{4.24}$$

This completes the first step of the proof which is illustrated in Figure 4.2.

In the second step, we follow the proof of Johnstone and Moulin (2017, Th. 2). Using Lem. 4.10 .(i)-(ii), we can write for all $k \geq 1$,

$$\|\nabla H(z_{k+1})\|^2 \leq 2\rho_2^2 \left(\|x_{k+1} - x_k\|^2 + \|x_k - x_{k-1}\|^2\right) \leq \frac{2\rho_2^2}{\rho_1}(H(z_{k-1}) - H(z_{k+1})).$$

Injecting the last inequality in Eq. (4.24), we obtain for all $k > k_2 := \max(l, 2)$,

$$\frac{2\rho_2^2}{\rho_1} \varphi'(H(z_k) - H(\bar{z}))^2 (H(z_{k-2}) - H(z_k)) \geq 1.$$

Now, use $\varphi'(s) = \bar{c}s^{\theta-1}$ to derive the following for all $k > k_2$:

$$[H(z_{k-2}) - H(\bar{z})] - [H(z_k) - H(\bar{z})] \geq \frac{\rho_1}{2\rho_2^2 \bar{c}^2}[H(z_k) - H(\bar{z})]^{2(1-\theta)}. \tag{4.25}$$

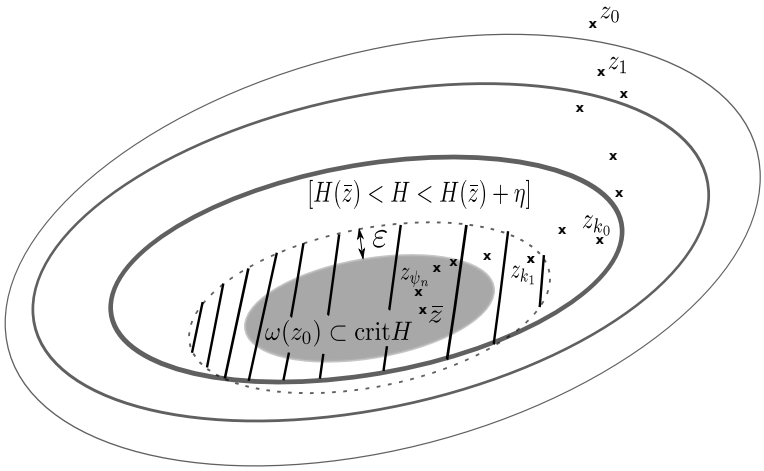

Figure 4.2 – Illustration of the first step of the proof of Th. 4.6

Let $r_k := H(z_k) - H(\bar{z})$ and $C_1 = \frac{\rho_1}{2\rho_2^2 \bar{c}^2}$. Then, we can rewrite Eq. (4.25) as

$$r_{k-2} - r_k \geq C_1 r_k^{2(1-\theta)} \quad \forall k > k_2 \,. \tag{4.26}$$

We distinguish three different cases to obtain the sought results.

(i) $\underline{\theta = 1}$:
Suppose $r_k > 0$ for all $k > k_2$. Then, since we know that $r_k \to 0$ by Eq. (4.23), C_1 must be equal to 0. This is a contradiction. Therefore, there exist $k_3 \in \mathbb{N}$ s.t. $r_k = 0$ for all $k > k_3$ (recall that $(r_k)_{k\in\mathbb{N}}$ is nonincreasing).

(ii) $\underline{\theta \geq \frac{1}{2}}$:
As $r_k \to 0$, there exists $k_4 \in \mathbb{N}$ s.t. for all $k \geq k_4$, $r_k \leq 1$. Observe that $2(1-\theta) \leq 1$ and hence $r_{k-2} - r_k \geq C_1 r_k$ for all $k > k_2$ and then

$$r_k \leq (1 + C_1)^{-1} r_{k-2} \leq (1 + C_1)^{-p_1} r_{k_4} \,. \tag{4.27}$$

where $p_1 := \lfloor \frac{k-k_4}{2} \rfloor$. Notice that $p_1 > \frac{k-k_4-2}{2}$. Thus, the linear convergence result follows. Note also that if $\theta = 1/2$, $2(1-\theta) = 1$ and Eq. (4.27) holds for all $k > k_2$.

(iii) $\underline{\theta < \frac{1}{2}}$:
Define the function h by $h(t) = \frac{D}{1-2\theta} t^{2\theta-1}$ where $D > 0$ is a constant. Then,

$$h(r_k) - h(r_{k-2}) = \int_{r_{k-2}}^{r_k} h'(t)dt = D \int_{r_k}^{r_{k-2}} t^{2\theta-2} dt \geq D\,(r_{k-2} - r_k)\, r_{k-2}^{2\theta-2} \,.$$

We disentangle now two cases:

(a) Suppose $2r_{k-2}^{2\theta-2} \geq r_k^{2\theta-2}$. Then, by Eq. (4.26), we get

$$h(r_k) - h(r_{k-2}) = D\,(r_{k-2} - r_k)\, r_{k-2}^{2\theta-2} \geq \frac{C_1\, D}{2} \,. \tag{4.28}$$

(b) Suppose now the opposite inequation $2r_{k-2}^{2\theta-2} < r_k^{2\theta-2}$. We can suppose without loss of generality that r_k are all positive. Otherwise, if there exists p such that $r_p = 0$, the sequence $(r_k)_{k\in\mathbb{N}}$ will be stationary at 0 for all $k \geq p$. Observe that $2\theta - 2 < 2\theta - 1 < 0$, thus $\frac{2\theta-1}{2\theta-2} > 0$. As a consequence, we can write in this case $r_k^{2\theta-1} > q\, r_{k-2}^{2\theta-1}$ where $q := 2^{\frac{2\theta-1}{2\theta-2}} > 1$. Therefore, using moreover that the sequence $(r_k)_{k\in\mathbb{N}}$ is nonincreasing and $2\theta - 1 < 0$, we derive the following

$$\begin{aligned} h(r_k) - h(r_{k-2}) &= \frac{D}{1-2\theta}(r_k^{2\theta-1} - r_{k-2}^{2\theta-1}) \\ &> \frac{D}{1-2\theta}\,(q-1) r_{k-2}^{2\theta-1} > \frac{D}{1-2\theta}\,(q-1) r_{k_2}^{2\theta-1} := C_2\,. \end{aligned} \tag{4.29}$$

Combining Eq. (4.28) and Eq. (4.29) yields $h(r_k) \geq h(r_{k-2}) + C_3$ where $C_3 := \min(C_2, \frac{C_1 D}{2})$. Consequently, $h(r_k) \geq h(r_{k-2\,p_2}) + p_2\, C_3$ where $p_2 := \lfloor \frac{k-k_2}{2} \rfloor$. We deduce from this inequality that

$$h(r_k) \geq h(r_k) - h(r_{k-2\,p_2}) \geq p_2\, C_3\,.$$

Therefore, rearranging this inequality using the definition of h, we obtain $r_k^{1-2\theta} \leq \frac{D}{1-2\theta}(C_3\, p_2)^{-1}$. Then, since $p_2 > \frac{k-k_2-2}{2}$,

$$r_k \leq C_4\, p_2^{\frac{1}{2\theta-1}} \leq C_4 \left(\frac{k-k_2-2}{2}\right)^{\frac{1}{2\theta-1}}\,.$$

where $C_4 := \left(\frac{C_3\,(1-2\theta)}{D}\right)^{\frac{1}{2\theta-1}}$.

We conclude the proof by observing that $f(x_k) \leq H(z_k)$ and recalling that $\bar{z} \in \mathrm{crit} H$.

4.9.4 Proof of Lem. 4.7

Since f has the KŁ property at $\bar{x}$ with an exponent $\theta \in (0, 1/2]$, there exist c, ε and $\nu > 0$ s.t.

$$\|\nabla f(x)\|^{\frac{1}{1-\theta}} \geq c(f(x) - f(\bar{x})) \tag{4.30}$$

for all $x \in \mathbb{R}^d$ s.t. $\|x - \bar{x}\| \leq \varepsilon$ and $f(x) < f(\bar{x}) + \nu$ where condition $f(\bar{x}) - f(x)$ is dropped because Eq. (4.30) holds trivially otherwise. Let $z = (x, y) \in \mathbb{R}^{2d}$ be s.t. $\|x - \bar{x}\| \leq \varepsilon$, $\|y\| \leq \varepsilon$ and $H(\bar{x}, 0) < H(x, y) < H(\bar{x}, 0) + \nu$. We assume that $\varepsilon < b$ (ε can be shrunk if needed). We have $f(x) \leq H(x, y) < H(\bar{x}, 0) + \nu = f(\bar{x}) + \nu$. Hence Eq. (4.30) holds for these x.

By concavity of $u \mapsto u^{\frac{1}{2(1-\theta)}}$, we obtain

$$\|\nabla H(x, y)\|^{\frac{1}{1-\theta}} \geq C_0 \left(\|\nabla f(x)\|^{\frac{1}{1-\theta}} + \left\|\frac{y}{b}\right\|^{\frac{1}{1-\theta}}\right)$$

where $C_0 := 2^{\frac{1}{2(1-\theta)}-1}$.

Hence, using Eq. (4.30), we get

$$\|\nabla H(x,y)\|^{\frac{1}{1-\theta}} \geq C_0 \left(c\left(f(x) - f(\bar{x})\right) + \left\| \frac{y}{b} \right\|^{\frac{1}{1-\theta}} \right) .$$

Observe now that $\frac{1}{1-\theta} \geq 2$ and $\left\| \frac{y}{b} \right\| \leq \frac{\varepsilon}{b} \leq 1$. Therefore, $\left\| \frac{y}{b} \right\|^{\frac{1}{1-\theta}} \geq \|y/b\|^2$. Finally,

$$\begin{aligned}
\|\nabla H(x,y)\|^{\frac{1}{1-\theta}} &\geq C_0 \left(c\left(f(x) - f(\bar{x})\right) + \frac{2}{b}\frac{1}{2b}\|y\|^2 \right) \\
&\geq C_0 \min\left(c, \frac{2}{b}\right) \left(f(x) - f(\bar{x}) + \frac{1}{2b}\|y\|^2 \right) \\
&= C_0 \min\left(c, \frac{2}{b}\right) \Big(H(x,y) - H(\bar{x},0) \Big) .
\end{aligned}$$

This completes the proof.

5

Analysis of a Target-Based Actor-Critic Algorithm with Linear Function Approximation

Abstract Actor-critic methods integrating target networks have exhibited a stupendous empirical success in deep reinforcement learning. However, a theoretical understanding of the use of target networks in actor-critic methods is largely missing in the literature. In this chapter, we bridge this gap between theory and practice by proposing the first theoretical analysis of an online target-based actor-critic algorithm with linear function approximation in the discounted reward setting. Our algorithm uses three different timescales: one for the actor and two for the critic. Instead of using the standard single timescale temporal difference (TD) learning algorithm as a critic, we use a two timescales target-based version of TD learning closely inspired from practical actor-critic algorithms implementing target networks. First, we establish asymptotic convergence results for both the critic and the actor under Markovian sampling. Then, we provide a finite-time analysis showing the impact of incorporating a target network into actor-critic methods.

5.1 Introduction

Actor-critic algorithms (Barto et al., 1983; Konda and Borkar, 1999; Konda and Tsitsiklis, 2003; Peters and Schaal, 2008; Bhatnagar et al., 2009) are a class of reinforcement learning (RL) (Sutton and Barto, 2018; Bertsekas and Tsitsiklis, 1996) methods to find an optimal policy maximizing the total expected reward in a stochastic environment modelled by a Markov Decision Process (MDP) (Puterman, 2014). In this type of algorithms, two main processes interplay: the actor and the critic. The actor updates a parameterized policy in a direction of performance improvement whereas the critic estimates the current policy of the actor by estimating the unknown state-value function. In turn, the critic estimation is used to produce the update rule of the actor. Combined with deep neural networks as function approximators of the value function, actor-critic algorithms witnessed a tremendous success in a range of challenging tasks (Heess et al., 2015; Lillicrap et al., 2016; Mnih et al., 2016; Fujimoto et al., 2018; Haarnoja et al., 2018). Apart from using neural networks for function approximation (FA), one of the main features underlying their remarkable empirical achievements is the use of target networks for the critic estimation of the value function. Introduced by the seminal work of Mnih et al. (2015) to stabilize the training process, this target innovation consists in using two neural networks maintaining two copies of the estimated value function: A so-called target network tracking a main network with some delay computes the target values for the value function update.

Despite their resounding empirical success in deep RL, a theoretical understanding of the use of target networks in actor-critic methods is largely missing in the literature. Theoretical contributions investigating the use of a target network are very recent and limited to temporal difference (TD) learning for policy evaluation (Lee and He, 2019) and

critic-only methods such as Q-learning for control (Zhang et al., 2021b). In particular, these works are not concerned with actor-critic algorithms and leave the question of the finite-time analysis open.

In the present work, we bridge this gap between theory and practice by proposing the first theoretical analysis of an online target-based actor-critic algorithm in the discounted reward setting. We consider the linear FA setting where a linear combination of pre-selected feature (or basis) functions estimates the value function in the critic. An analysis of this setting is an insightful first step before tackling the more challenging nonlinear FA setting aligned with the use of neural networks. We conduct our study in the multiple timescales framework. In the standard two timescales actor-critic algorithms (Konda and Tsitsiklis, 2003; Bhatnagar et al., 2009), at each iteration, the actor and the critic are updated simultaneously but the actor evolves more slowly than the critic, using smaller stepsizes than the latter. We face two main challenges due to the integration of the target variable mechanism. First, in contrast to standard two timescales actor-critic algorithms, our algorithm uses three different timescales: one for the actor and two for the critic. Instead of using the single timescale TD learning algorithm as a critic, we use a two timescales target-based version of TD learning closely inspired from practical actor-critic algorithms implementing target networks. Second, incorporating a target variable into the critic results in the intricate interplay between three processes evolving on three different timescales. In particular, the use of a target variable significantly modifies the dynamics of the actor-critic algorithm and deserves a careful analysis accordingly.

Our main contributions are summarized as follows. First, we prove asymptotic convergence results for both the critic and the actor. More precisely, as the actor parameter changes slowly compared to the critic one, we show that the critic which uses a target variable tracks a slowly moving target corresponding to the well-known TD solution (Tsitsiklis and Van Roy, 1997). Our development is based on the classical ordinary differential equation (ODE) method of stochastic approximation (see, for e.g., Benveniste et al. (1990); Borkar (2008)). Then, we show that the actor parameter visits infinitely often a region of the parameter space where the norm of the policy gradient is dominated by a bias due to linear FA. Second, we conduct a finite-time analysis of our actor-critic algorithm which shows the impact of using a target variable on the convergence rates and the sample complexity. Loosely speaking, up to a FA error, we show that our target-based algorithm converges in expectation to an ϵ-approximate stationary point of the non-concave performance function using at most $\mathcal{O}(\epsilon^{-3}\log^3\frac{1}{\epsilon})$ samples compared with $\mathcal{O}(\epsilon^{-2}\log(\frac{1}{\epsilon}))$ for the best known complexity for two timescales actor-critic algorithms without a target network. All the proofs are deferred to Sections 5.7 and 5.8.

5.2 Related works

In this section, we briefly discuss the most relevant related works to ours. Existing theoretical results in the literature can be divided into two classes.

Asymptotic results. Almost sure convergence results are referred to as asymptotic. Konda & Tsitsiklis (Konda and Tsitsiklis, 2003; Konda, 2002) provided almost sure (with probability one) convergence results for a two timescales actor-critic algorithm in which the critic estimates the action-value function via linear FA. Our algorithm is closer to an actor-critic algorithm introduced by Bhatnagar et al. (2009) in the average reward setting. However, unlike Bhatnagar et al. (2009), we consider the discounted reward setting and integrate a target variable mechanism into our critic. Moreover, as previously

mentioned, the target variable for the critic adds an additional timescale in comparison to Konda and Tsitsiklis (2003); Bhatnagar et al. (2009) which only involve two different timescales. Regarding theoretical results considering target networks, Lee and He (2019) proposed a family of single timescale target-based TD learning algorithms for policy evaluation. Our critic corresponds to a two timescales version of the single timescale target-based TD learning algorithm of (Lee and He, 2019, Algorithm 2) called Averaging TD. In (Lee and He, 2019, Th. 1), this single timescale algorithm is shown to converge with probability one (w.p.1) towards the standard TD solution solving the projected Bellman equation (see Tsitsiklis and Van Roy (1997) for a precise statement). Besides the timescales difference with Lee and He (2019), in this article, we are concerned with a control setting in which the policy changes at each timestep via the actor update. Yang et al. (2019) proposed a bilevel optimization perspective to analyze Q-learning with a target network and an actor-critic algorithm without any target network. More recently, Zhang et al. (2021b) investigated the use of target networks in Q-learning with linear FA and a target variable with Ridge regularization. Their analysis covers both the average and discounted reward settings and establishes asymptotic convergence results for both policy evaluation and control. This recent work of Zhang et al. (2021b) focuses on the critic-only Q-learning method with a target network update rule, showing the role of the target network in the off-policy setting. In particular, this work is not concerned with actor-critic algorithms.

Finite-time analysis. The second type of results consists in establishing time-dependent bounds on some error or performance quantities such as the average expected norm of the gradient of the performance function. These are referred to as finite-time analysis. In the last few years, several works proposed finite-time analysis for TD learning (Bhandari et al., 2018; Srikant and Ying, 2019) for two timescales TD methods (Xu et al., 2019) and even more generally for two timescales linear stochastic approximation algorithms (Gupta et al., 2019; Dalal et al., 2018; Kaledin et al., 2020). These works opened the way to the recent development of a flurry of nonasymptotic results for actor-critic algorithms (Yang et al., 2018; Qiu et al., 2019; Kumar et al., 2019; Hong et al., 2020; Xu et al., 2020; Wang et al., 2020; Wu et al., 2020; Shen et al., 2020). Regarding online one-step actor-critic algorithms, Wu et al. (2020) provided a finite-time analysis of the standard two timescales actor-critic algorithm (Bhatnagar et al., 2009, Algorithm 1) in the average reward setting with linear FA. Shen et al. (2020) conducted a similar study for a revisited version of the asynchronous advantage actor-critic (A3C) algorithm in the discounted setting. None of the mentioned works uses a target network. In this work, we conduct a finite-time analysis of our target-based actor-critic algorithm. Such new results are missing in all theoretical results investigating the use of a target network (Lee and He, 2019; Zhang et al., 2021b).

The summary table 5.1 compiles some key features of our work to situate it in the literature and highlights our contributions with respect to (w.r.t.) the closest related works.

5.3 Preliminaries

Notation. For every finite set $\mathcal{X}$, we use the notation $\mathcal{P}(\mathcal{X})$ for the set of probability measures on $\mathcal{X}$. The cardinality of a finite set $\mathcal{Y}$ is denoted by $|\mathcal{Y}|$. For two sequences of nonnegative reals (x_n) and (y_n), the notation $x_n = \mathcal{O}(y_n)$ means that there exists a constant C independent of n such that $x_n \leq C y_n$ for all $n \in \mathbb{N}$. For any integer p, the

Table 5.1 – Comparison to closest related works.

	Discounted reward	Actor critic	Markovian sampling[1]	Target variable	Asymptotic results	Finite-time analysis	Timescales
Lillicrap et al. (2016)	✓	✓	✗	✓	✗	✗	1
Lee and He (2019)	✓	✗	✗	✓	✓	✓[2]	1
Wu et al. (2020)	✗	✓	✓	✗	✗	✓	3
Shen et al. (2020)	✓	✓	✓	✗	✗	✓	2
Zhang et al. (2021b)	✓	✗	✓	✓	✓	✗	2
Ours	✓	✓	✓	✓	✓	✓	3

[1] refers to the use of samples generated from the MDP and the acting policy, this excludes experience replay as in Lillicrap et al. (2016) and identically independently distributed (i.i.d.) samples used in theoretical analysis.

[2] Lee & and He Lee and He (2019) provide a finite-time analysis for a target-based TD-learning algorithm (for policy evaluation) based on the periodic update style of the target variable used in Mnih et al. (2015) involving two loops. They highlight that a finite-time analysis of the Polyak-averaging style update rule Lillicrap et al. (2016) is an open question. Here, we address this question in the control setting.

euclidean space $\mathbb{R}^p$ is equipped with its usual inner product $\langle \cdot, \cdot \rangle$ and its corresponding 2-norm $\| \cdot \|$. For any integer d and any matrix $A \in \mathbb{R}^{d \times p}$, we use the notation $\|A\|$ for the operator norm induced by the euclidean vector norm. For a symmetric positive semidefinite matrix $B \in \mathbb{R}^{p \times p}$ and a vector $x \in \mathbb{R}^p$, the notation $\|x\|_B^2$ refers to the quantity $\langle x, Bx \rangle$. The transpose of the vector x is denoted by x^T and I_p is the identity matrix.

5.3.1 Markov decision process and problem formulation

Consider the RL setting (Sutton and Barto, 2018; Bertsekas and Tsitsiklis, 1996; Szepesvári, 2010) where a learning agent interacts with an environment modeled as an infinite horizon discrete-time discounted MDP. We denote by $\mathcal{S} = \{s_1, \cdots, s_n\}$ the finite set of states and $\mathcal{A}$ the finite set of actions. Let $p : \mathcal{S} \times \mathcal{A} \to \mathcal{P}(\mathcal{S})$ be the state transition probability kernel and $R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ the immediate reward function. A randomized stationary policy, which we will simply call a policy in the rest of the chapter, is a mapping $\pi : \mathcal{S} \to \mathcal{P}(\mathcal{A})$ specifying for each $s \in \mathcal{S}, a \in \mathcal{A}$ the probability $\pi(a|s)$ of selecting action a in state s. At each time step $t \in \mathbb{N}$, the RL agent in a state $S_t \in \mathcal{S}$ executes an action $A_t \in \mathcal{A}$ with probability $\pi(A_t|S_t)$, transitions into a state $S_{t+1} \in \mathcal{S}$ with probability $p(S_{t+1}|S_t, A_t)$ and observes a random reward $R_{t+1} \in [-U_R, U_R]$ where U_R is a positive real. We denote by $\mathbb{P}_{\rho,\pi}$ the probability distribution of the Markov chain (S_t, A_t) issued from the MDP controlled by the policy π with initial state distribution ρ. The notation $\mathbb{E}_{\rho,\pi}$ refers to the associated expectation. We will use $\mathbb{E}_\pi$ whenever there is no dependence on ρ. The sequence (R_t) is such that (s.t.) $\mathbb{E}_\pi[R_{t+1}|S_t, A_t] = R(S_t, A_t)$. Let $\gamma \in (0, 1)$ be a discount factor. Given a policy π, the long-term expected cumulative discounted reward is quantified by the state-value function $V_\pi : \mathcal{S} \to \mathbb{R}$ and the action-value function $Q_\pi : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ defined for all $s \in \mathcal{S}, a \in \mathcal{A}$ by $V_\pi(s) := \mathbb{E}_\pi[\sum_{t=0}^{\infty} \gamma^t R_{t+1}|S_0 = s]$and $Q_\pi(s, a) := \mathbb{E}_\pi[\sum_{t=0}^{\infty} \gamma^t R_{t+1}|S_0 = s, A_0 = a]$ .We also define the advantage function $\Delta_\pi : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ by $\Delta_\pi(s, a) := Q_\pi(s, a) - V_\pi(s)$. Given an initial probability distribution ρ over states for the initial state S_0, the goal of the agent is to find a policy π maximizing the expected long-term return $J(\pi) := \sum_{s \in \mathcal{S}} \rho(s) V_\pi(s)$. For this purpose, the agent has only access to realizations of the random variables S_t, A_t and R_t whereas the state transition kernel p and the reward function R are unknown.

5.3.2 Policy Gradient framework

From now on, we restrict the policy search to the set of policies π parameterized by a vector $\theta \in \mathbb{R}^d$ for some integer $d > 0$ and optimize the performance criterion J over this family of parameterized policies $\{\pi_\theta : \theta \in \mathbb{R}^d\}$. The policy dependent function J can also be seen as a function of the parameter θ. We use the notation $J(\theta)$ for $J(\pi_\theta)$ by abuse of notation. The problem that we are concerned with can be written as: $\max_{\theta\in\mathbb{R}^d} J(\theta)$ .Whenever it exists, define for every $\theta \in \mathbb{R}^d$ the function $\psi_\theta : \mathcal{S} \times \mathcal{A} \to \mathbb{R}^d$ for all $(s, a) \in \mathcal{S} \times \mathcal{A}$ by:

$$\psi_\theta(s, a) := \nabla \ln \pi_\theta(a|s)\,,$$

where ∇ denotes the gradient w.r.t. θ. We introduce an assumption on the regularity of the parameterized family of policies which is a standard requirement in policy gradients (see, for eg., (Zhang et al., 2020a, Assumption 3.1)(Konda and Tsitsiklis, 2003, Assumption 2.1)). In particular, it ensures that ψ_θ is well defined .

Assumption 5.3.1. The following conditions hold true for every $(s, a) \in \mathcal{S} \times \mathcal{A}$.

(a) For every $\theta \in \mathbb{R}^d$, $\pi_\theta(a|s) > 0$.

(b) The function $\theta \mapsto \pi_\theta(a|s)$ is continuously differentiable and L_π-Lipschitz continuous.

(c) The function $\theta \mapsto \psi_\theta(s, a)$ is bounded and L_ψ-Lipschitz.

Assumption 5.3.1 is satisfied for instance by the Gibbs (or softmax) policy and the Gaussian policy (see (Zhang et al., 2020a, Section 3) and the references therein for details). Under Assumption 5.3.1 , the policy gradient theorem (Sutton et al., 1999)(Konda, 2002, Th. 2.13) with the state-value function as a baseline provides an expression for the gradient of the performance metric J w.r.t. the policy parameter θ given by:

$$\nabla J(\theta) = \frac{1}{1-\gamma} \cdot \mathbb{E}_{(\tilde{S},\tilde{A})\sim\mu_{\rho,\theta}}[\Delta_{\pi_\theta}(\tilde{S}, \tilde{A})\, \psi_\theta(\tilde{S}, \tilde{A})]\,. \tag{5.1}$$

Here, the couple of random variables $(\tilde{S}, \tilde{A})$ follows the discounted state-action occupancy measure $\mu_{\rho,\theta} \in \mathcal{P}(\mathcal{S} \times \mathcal{A})$ defined for all $(s, a) \in \mathcal{S} \times \mathcal{A}$ by:

$$\mu_{\rho,\theta}(s, a) := d_{\rho,\theta}(s)\, \pi_\theta(a|s) \quad \text{where} \quad d_{\rho,\theta}(s) := (1-\gamma) \sum_{t=0}^{\infty} \gamma^t \mathbb{P}_{\rho,\pi_\theta}(S_t = s) \tag{5.2}$$

is a probability measure over the state space $\mathcal{S}$ known as the discounted state-occupancy measure. Note that under Assumption 5.3.1 , the policy gradient ∇J is Lipschitz continuous (see (Zhang et al., 2020a, Lem. 4.2)).

5.4 Target-based actor-critic algorithm

In this section, we gradually present our actor-critic algorithm.

5.4.1 Actor update

First, we need an estimate of the policy gradient $\nabla J(\theta)$ of Eq. (5.1) in view of using stochastic gradient ascent to solve the maximization problem . Given Eq. (5.1) and following previous works, we recall how to sample according to the distribution $\mu_{\rho,\theta}$. As described in (Konda, 2002, Section 2.4), the distribution $\mu_{\rho,\theta}$ is the stationary distribution

of a Markov chain $(\tilde{S}_t, \tilde{A}_t)_{t\in\mathbb{N}}$ issued from the artificial MDP whose transition kernel $\tilde{p} : \mathcal{S} \times \mathcal{A} \to \mathcal{P}(\mathcal{S})$ is defined for every $(s, a) \in \mathcal{S} \times \mathcal{A}$ by

$$\tilde{p}(\cdot|s, a) := \gamma\, p(\cdot|s, a) + (1 - \gamma)\, \rho(\cdot)\,, \tag{5.3}$$

and which is controlled by the policy π_θ generating the action sequence $(\tilde{A}_t)$. We will later state conditions to ensure its existence and uniqueness. Therefore, under suitable conditions, the distribution of the Markov chain $(\tilde{S}_t, \tilde{A}_t)_{t\in\mathbb{N}}$ will converge geometrically towards its stationary distribution $\mu_{\rho,\theta}$. This justifies the following sampling procedure. Given a state $\tilde{S}_t$ and an action $\tilde{A}_t$, we sample a state $\tilde{S}_{t+1}$ according to this artificial MDP by sampling from $p(\cdot|\tilde{S}_t, \tilde{A}_t)$ with probability γ and from ρ otherwise. For this purpose, at each time step t, we draw a Bernoulli random variable $B_t \in \{0, 1\}$ with parameter γ which is independent of all the past random variables generated until time t.

Then, using the definition of the advantage function, Eq. (5.1) becomes:

$$\nabla J(\theta) = \frac{1}{1-\gamma} \cdot \mathbb{E}_{(\tilde{S},\tilde{A})\sim\mu_{\rho,\theta},S\sim p(\cdot|\tilde{S},\tilde{A})}[(R(\tilde{S}, \tilde{A}) + \gamma V_{\pi_\theta}(S) - V_{\pi_\theta}(\tilde{S}))\, \psi_\theta(\tilde{S}, \tilde{A})]\,. \tag{5.4}$$

From this equation, it is natural to define for every $V \in \mathbb{R}^n$ the temporal difference (TD) error

$$\delta^V_{t+1} = R_{t+1} + \gamma\, V(S_{t+1}) - V(\tilde{S}_t)\,, \tag{5.5}$$

where S_{t+1} is drawn from the distribution $p(\cdot|\tilde{S}_t, \tilde{A}_t)$ and $(\tilde{S}_t, \tilde{A}_t)_{t\in\mathbb{N}}$ is the Markov chain induced by the artificial MDP described in Eq. (5.3) and controlled by the policy π_θ. Notice here from Eq. (5.4) that we need two different sequences (S_t) and $(\tilde{S}_t)$ respectively sampled from the kernels p and $\tilde{p}$. In our discounted reward setting, using only the sequence $(\tilde{S}_t)$ issued from the artificial kernel $\tilde{p}$ would result in a bias with a sampling error of the order $1 - \gamma$ (see (Shen et al., 2020, Eq. (14) and Lem. 7)).

Supposing for now that the value function V_{π_θ} is known, it stems from Eq. (5.4) that a natural estimator of the gradient $\nabla J(\theta)$ is $\delta^{V_{\pi_\theta}}_{t+1}\psi_\theta(\tilde{S}_t, \tilde{A}_t)/(1 - \gamma)$. This estimator is only biased because the distribution of our sampled Markov chain $(\tilde{S}_t, \tilde{A}_t)_t$ is not exactly $\mu_{\rho,\theta}$ but converges geometrically to this one. However, the state-value function V_{π_θ} is unknown. Given an estimate $V_{\omega_t} \in \mathbb{R}^n$ of $V_{\pi_{\theta_t}}$ and a positive stepsize α_t, the actor updates its parameter as follows:

$$\theta_{t+1} = \theta_t + \alpha_t \frac{1}{1-\gamma} \delta^{V_{\omega_t}}_{t+1} \psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t)\,. \tag{5.6}$$

5.4.2 Critic update

The state-value function V_{π_θ} is approximated for every state $s \in \mathcal{S}$ by a linear function of carefully chosen feature vectors as follows: $V_{\pi_\theta}(s) \approx V_\omega(s) = \omega^T\phi(s) = \sum_{i=1}^m \omega_i\phi^i(s)$,where $\omega = (\omega_1, \cdots, \omega_m)^T \in \mathbb{R}^m$ for some integer $m \ll n = |\mathcal{S}|$ and $\phi(s) = (\phi^1(s), \cdots, \phi^m(s))^T$ is the feature vector of the state $s \in \mathcal{S}$. We compactly represent the feature vectors as a matrix of features Φ of size $n \times m$ whose ith row corresponds to the row vector $\phi(s)^T$ for some $s \in \mathcal{S}$.

Now, before completing the presentation of our algorithm, we motivate the use of a target variable for the critic. As previously mentioned, instead of a standard TD learning algorithm (Sutton, 1988) for the critic, we use a target-based TD learning algorithm. We follow a similar exposition to (Lee and He, 2019, Secs. 2.3, 2.4 and 3) to introduce the target variable for the critic. Let us introduce some additional notations for this purpose.

Fix $\theta \in \mathbb{R}^d$. Let P_θ be the transition matrix over the finite state space associated to the Markov chain (S_t), i.e., the matrix of size $n \times n$ defined for every $s, s' \in \mathcal{S}$ by $P_\theta(s'|s) := \sum_{a\in\mathcal{A}} p(s'|s,a)\pi_\theta(a|s)$ .Consider the vector $R_\theta = (R_\theta(s_1), \cdots R_\theta(s_n))$ whose ith coordinate is provided by $R_\theta(s_i) = \sum_{a\in\mathcal{A}} \pi_\theta(a|s_i)R(s_i,a)$. Let $D_{\rho,\theta}$ be the diagonal matrix with elements $d_{\rho,\theta}(s_i)$, $i = 1, \cdots n$ along its diagonal. Define also the Bellman operator $T_\theta : \mathbb{R}^n \mapsto \mathbb{R}^n$ for every $V \in \mathbb{R}^n$ by $T_\theta V := R_\theta + \gamma P_\theta V$. The true value function V_{π_θ} satisfies the celebrated Bellman equation $V_{\pi_\theta} = T_\theta V_{\pi_\theta}$. This naturally leads to minimize the mean-square Bellman error (MSBE) (Sutton et al., 2009b, Section 3) defined for every $\omega \in \mathbb{R}^m$ by $\mathcal{E}_\theta(\omega) := \frac{1}{2}\|T_\theta V_\omega - V_\omega\|^2_{D_{\rho,\theta}}$ where $V_\omega = \Phi\omega$. The gradient of the MSBE w.r.t. ω can be written as $\nabla_\omega \mathcal{E}_\theta(\omega) = \mathbb{E}_{\tilde{S}\sim d_{\rho,\theta}}[(T_\theta V_\omega(\tilde{S}) - V_\omega(\tilde{S}))(\mathbb{E}_{S\sim P_\theta(\cdot|\tilde{S})}[\gamma\nabla_\omega V_\omega(S)] - \nabla_\omega V_\omega(\tilde{S}))]$. As explained in (Bertsekas and Tsitsiklis, 1996, p. 369), omitting the gradient term $\nabla_\omega T_\theta V_\omega(\tilde{S}) = \mathbb{E}_{S\sim P_\theta(\cdot|\tilde{S})}[\gamma\nabla_\omega V_\omega(S)]$ in $\nabla_\omega \mathcal{E}_\theta(\omega)$ yields the standard TD learning update rule $\omega_{t+1} = \omega_t + \delta_{t+1}\phi(\tilde{S}_t)$. The TD learning update does not coincide with a stochastic gradient descent on the MSBE or even any other objective function (see (Barnard, 1993, Appendix 1) for a proof). The idea of target-based TD learning is to consider a modified version of the MSBE $\tilde{\mathcal{E}}_\theta(\omega, \bar{\omega}) := \frac{1}{2}\|T_\theta V_{\bar{\omega}} - V_\omega\|^2_{D_{\rho,\theta}}$. Observe that the term $T_\theta V_\omega$ depending on ω in the MSBE is now freezed in $\tilde{\mathcal{E}}_\theta(\omega, \bar{\omega})$ thanks to the target variable $\bar{\omega}$. We now need to introduce a new sequence $\bar{\omega}_t$ to define a sample-based version of $T_\theta V_{\bar{\omega}} - V_\omega$ which will be a modified version of the standard TD-error

$$\bar{\delta}_{t+1} = R_{t+1} + \gamma\phi(S_{t+1})^T\bar{\omega}_t - \phi(\tilde{S}_t)^T\omega_t \,. \tag{5.7}$$

Then, a stochastic gradient descent on $\tilde{\mathcal{E}}$ w.r.t. ω yields the critic update

$$\omega_{t+1} = \omega_t + \beta_t\bar{\delta}_{t+1}\phi(\tilde{S}_t) \,. \tag{5.8}$$

The target variable sequence $\bar{\omega}_t$ needs to be a slowed down version of the critic parameter ω_t. For this purpose, instead of using a periodical synchronization of the target variable $\bar{\omega}_t$ with ω_t through a copy as in DQN, we use the Polyak-averaging update rule proposed by Lillicrap et al. (2016)

$$\bar{\omega}_{t+1} = \bar{\omega}_t + \xi_t(\omega_{t+1} - \bar{\omega}_t) \,, \tag{5.9}$$

where ξ_t is a positive stepsize chosen s.t. the sequence $(\bar{\omega}_t)$ evolves on a slower timescale than the sequence (ω_t) to track it. The update rules of the actor and the critic collected together from Eqs. (5.5) to (5.8) give rise to Algorithm 5.1. We will use the shorthand notation $\delta_{t+1} := \delta_{t+1}^{V_{\omega_t}}$ from now on.

5.5 Convergence analysis

In this section, we provide asymptotic convergence guarantees for the critic and the actor of Algorithm 5.1 successively . For every $\theta \in \mathbb{R}^d$, let $\tilde{K}_\theta \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|\times|\mathcal{S}||\mathcal{A}|}$ be the transition matrix over the state-action pairs defined for every $(s,a),(s',a') \in \mathcal{S}\times\mathcal{A}$ by $\tilde{K}_\theta(s',a'|s,a) = \tilde{p}(s'|s,a)\pi_\theta(a'|s')$. Let $\mathcal{K} := \{\tilde{K}_\theta : \theta \in \mathbb{R}^d\}$ and let $\bar{\mathcal{K}}$ be its closure. Every element of $\bar{\mathcal{K}}$ defines a Markov chain on the state-action space. We make the following assumption (see also Zhang et al. (2021b); Marbach and Tsitsiklis (2001)).

Assumption 5.5.1. For every $K \in \bar{\mathcal{K}}$, the Markov chain induced by K is ergodic.

Algorithm 5.1 Actor-critic algorithm $(\gamma, (\alpha_t), (\beta_t), (\xi_t))$.

Initialization: $\theta_0, \omega_0 \in \mathbb{R}^d$.
for $t = 0, 1, 2, \cdots, T-1$ **do**
$\tilde{A}_t \sim \pi_{\theta_t}(\cdot|\tilde{S}_t)$; $S_{t+1} \sim p(\cdot|\tilde{S}_t, \tilde{A}_t)$
$\delta_{t+1} = R_{t+1} + \gamma\, \phi(S_{t+1})^T \omega_t - \phi(\tilde{S}_t)^T \omega_t$ {classical TD error}
$\bar{\delta}_{t+1} = R_{t+1} + \gamma\, \phi(S_{t+1})^T \bar{\omega}_t - \phi(\tilde{S}_t)^T \omega_t$ {target-based TD error}
$\theta_{t+1} = \theta_t + \alpha_t \frac{1}{1-\gamma} \delta_{t+1} \psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t)$ {actor update}
$\omega_{t+1} = \omega_t + \beta_t \bar{\delta}_{t+1} \phi(\tilde{S}_t)$ {critic update}
$\bar{\omega}_{t+1} = \bar{\omega}_t + \xi_t(\omega_{t+1} - \bar{\omega}_t)$ {target variable update}
$S^{\rho}_{t+1} \sim \rho$; $B_{t+1} \sim \mathcal{B}(\gamma)$
$\tilde{S}_{t+1} = B_{t+1} S_{t+1} + (1 - B_{t+1}) S^{\rho}_{t+1}$
end for
Output: Policy and value function parameters θ_T and ω_T.

In particular, it ensures the existence of a unique invariant distribution $\mu_{\rho,\theta}$ for the kernel $\tilde{K}_\theta$ for every $\theta \in \mathbb{R}^d$. Note that we can replace $\tilde{p}$ by p in Assumption 5.5.1.

Algorithm 5.1 involves three different timescales. The actor parameter θ_t is updated on a slower timescale (i.e., with smaller stepsizes) than the target variable $\bar{\omega}_t$ which itself uses smaller stepsizes than the main critic parameter ω_t. This is guaranteed by a specific choice of the three stepsize schedules. The following assumption is a three timescales version of the standard assumption used for two timescales stochastic approximation (Borkar, 2008, Chap. 6) and plays a pivotal role in our analysis.

Assumption 5.5.2 (stepsizes)**.** The sequences of positive stepsizes $(\alpha_t), (\beta_t)$ and (ξ_t) satisfy:

(a) $\sum_t \alpha_t = \sum_t \beta_t = \sum_t \xi_t = +\infty\,, \quad \sum_t (\alpha_t^2 + \beta_t^2 + \xi_t^2) < \infty\,,$
(b) $\lim_{t\to\infty} \alpha_t/\xi_t = \lim_{t\to\infty} \xi_t/\beta_t = 0\,.$

We also need the following stability assumption.

Assumption 5.5.3. $\sup_t \|\omega_t\| < +\infty$ and $\sup_t \|\theta_t\| < +\infty$ *w.p.*1.

The almost sure boundedness assumption is classical (Konda and Borkar, 1999; Borkar, 2008; Bhatnagar et al., 2009; Karmakar and Bhatnagar, 2018). The stability question could be addressed in a look-up table representation setting (for e.g., $m = n$). Nevertheless, this question seems out of reach in the FA setting without any modification of the algorithm. Indeed, as discussed in (Bhatnagar et al., 2009, p. 2478-2479), FA makes it hard to find a Lyapunov function to apply the stochastic Lyapunov function method (Kushner and Yin, 2003) whereas the function J can be readily used in the tabular case. We highlight however that Assumption 5.5.3 is indeed strong. The almost sure boundedness of the sequence (ω_t) could be probably relaxed by using a generalization to three timescales of the rescaling technique of Borkar and Meyn (2000) which was extended by Lakshminarayanan and Bhatnagar (2017) to two timescales stochastic approximation in the case of i.i.d. samples. Relaxing this assumption via using the latter result would also require to handle the Markov noise. We leave this technical question to future work. Concerning the sequence (θ_t), as previously mentioned, it seems out of reach without modifying the algorithm, Lakshminarayanan and Bhatnagar (2017) (see

their Section 6) propose for example to regularize the objective function J by adding a quadratic penalty $\epsilon \|\theta\|^2$ leading to an additional term $\epsilon\,\theta_t$ (for any positive ϵ) in the actor update of the actor-critic algorithm 1 of Bhatnagar et al. (2009). It is also worth mentioning that several works enforce the boundedness via a projection of the iterates on some compact set (Bhandari et al., 2018; Wu et al., 2020; Shen et al., 2020). The drawback of this procedure is that it modifies the dynamics of the iterates and could possibly introduce spurious equilibria.

First, we will analyze the critic before investigating the convergence properties of the actor.

5.5.1 Critic analysis

The following assumption regarding the family of basis functions is a standard requirement (Bhatnagar et al., 2009; Konda and Tsitsiklis, 2003; Tsitsiklis and Van Roy, 1997).

Assumption 5.5.4 (critic features)**.** The matrix Φ has full column rank.

We follow the strategy of (Borkar, 2008, Chap. 6, Lem. 1) for the analysis of multi-timescale stochastic approximation schemes based on the ODE method. We start by analyzing the sequence (ω_t) evolving on the fastest timescale, i.e., with the slowly vanishing stepsizes β_t (see Assumption 5.5.2). The main idea behind the proofs is that $\theta_t, \bar{\omega}_t$ can be considered as quasi-static in this timescale. Then, loosely speaking (see Section 5.7.1.1 for a rigorous statement and proof), we can show from its update rule Eq. (5.8) that (ω_t) is associated to the ODE

$$\frac{d}{ds}\omega(s) = \bar{h}(\theta(s), \bar{\omega}(s)) - \bar{G}(\theta(s))\,\omega(s)\,, \quad \frac{d}{ds}\theta(s) = 0\,, \quad \frac{d}{ds}\bar{\omega}(s) = 0\,, \qquad \text{(ODE-}\omega\text{)}$$

where $\bar{h} : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^m$ and $\bar{G} : \mathbb{R}^d \to \mathbb{R}^{m\times m}$ are defined for every $\theta \in \mathbb{R}^d, \bar{\omega} \in \mathbb{R}^m$ by

$$\bar{h}(\theta, \bar{\omega}) := \Phi^T D_{\rho,\theta}(R_\theta + \gamma P_\theta \Phi\, \bar{\omega}) \quad \text{and} \quad \bar{G}(\theta) := \Phi^T D_{\rho,\theta}\Phi\,. \tag{5.10}$$

Recall that the matrices $D_{\rho,\theta}, P_\theta$ and the vector R_θ are defined in Section 5.4.2.

Remark 19. Under Assumptions 5.5.1 and 5.5.4 , the matrix $-\bar{G}(\theta)$ is Hurwitz for every $\theta \in \mathbb{R}^d$, i.e., all its eigenvalues have negative real parts. In particular, it is invertible.

The matrix $-\bar{G}(\theta)$ being Hurwitz, it follows from (ODE-ω) that ω_t tracks a slowly moving target $\omega_*(\theta_t, \bar{\omega}_t)$ governed by the slower iterates θ_t and $\bar{\omega}_t$. The detailed proof in Section 5.7.1.1 makes use of a result from Karmakar and Bhatnagar (2018) to handle the Markovian noise.

Proposition 5.1. Under Assumptions 5.3.1 and 5.5.1 to 5.5.4 , the linear equation $\bar{G}(\theta)\omega = \bar{h}(\theta, \bar{\omega})$ has a unique solution $\omega_*(\theta, \bar{\omega})$ for every $\theta \in \mathbb{R}^d, \bar{\omega} \in \mathbb{R}^m$ and it holds that $\lim_t \|\omega_t - \omega_*(\theta_t, \bar{\omega}_t)\| = 0$ *w.p.*1.

In a second step, we analyze the target variable sequence $(\bar{\omega}_t)$ which is evolving on a faster timescale than the sequence (θ_t) and slower than the sequence (ω_t). At the timescale ξ_t, everything happens as if the quantity ω_t in Eq. (5.9) could be replaced by $\omega_*(\theta_t, \bar{\omega}_t)$ thanks to Prop. 5.1. Thus, in a sense that is made precise in Section 5.7.1.2, we can show from Eq. (5.9) that $(\bar{\omega}_t)$ is related to the ODE

$$\frac{d}{ds}\bar{\omega}(s) = \bar{G}(\theta(s))^{-1}(h(\theta(s)) - G(\theta(s))\bar{\omega}(s))\,, \quad \frac{d}{ds}\theta(s) = 0\,, \qquad \text{(ODE-}\bar{\omega}\text{)}$$

where $h : \mathbb{R}^d \to \mathbb{R}^n$ and $G : \mathbb{R}^d \to \mathbb{R}^{m\times m}$ are defined for every $\theta \in \mathbb{R}^d$ by

$$h(\theta) := \Phi^T D_{\rho,\theta} R_\theta \quad \text{and} \quad G(\theta) := \Phi^T D_{\rho,\theta}(I_n - \gamma P_\theta)\Phi\,. \tag{5.11}$$

We show in Section 5.7.1.2 that the matrix $-G(\theta)$ is Hurwitz. This result differs from (Bertsekas and Tsitsiklis, 1996, Lem. 6.6. p.300) or (Tsitsiklis and Van Roy, 1997, Lem. 9) because the matrix $D_{\rho,\theta}$ corresponds to the stationary distribution associated to the artificial kernel $\tilde{p}$ and the policy π_θ in lieu of the original transition kernel p. Then, we prove that $-\bar{G}(\theta)^{-1}G(\theta)$ is also stable, which suggests from (ODE-$\bar{\omega}$) that $\bar{\omega}_t$ tracks an other slowly moving target $\bar{\omega}_*(\theta_t)$. This is established in the next proposition.

Proposition 5.2. Under Assumptions 5.3.1 and 5.5.1 to 5.5.4, for every $\theta \in \mathbb{R}^d$, the linear equation $G(\theta)\bar{\omega} = h(\theta)$ has a unique solution $\bar{\omega}_*(\theta)$ and $\lim_t \|\bar{\omega}_t - \bar{\omega}_*(\theta_t)\| = 0$ $w.p.1$. Moreover, for every $\theta \in \mathbb{R}^d$, $\Phi\,\bar{\omega}_*(\theta)$ is a fixed point of the projected Bellman operator, i.e., $\Pi_\theta T_\theta(\Phi\,\bar{\omega}_*(\theta)) = \Phi\,\bar{\omega}_*(\theta)$, where $\Pi_\theta = \Phi(\Phi^T D_{\rho,\theta}\Phi)^{-1}\Phi^T D_{\rho,\theta}$ is the projection matrix on the space $\{\Phi\,\omega : \omega \in \mathbb{R}^m\}$ of all vectors of the form $\Phi\,\omega$ for $\omega \in \mathbb{R}^m$ w.r.t. the norm $\|\cdot\|_{D_{\rho,\theta}}$.

Combining the results from Props. 5.1 and 5.2, we prove that ω_t tracks the same target $\bar{\omega}_*(\theta_t)$.

Theorem 5.3. *Let Assumptions 5.3.1, and 5.5.1 to 5.5.4 hold true. Then, we have*

$$\lim_t \|\omega_t - \bar{\omega}_*(\theta_t)\| = 0 \ w.p.1\,.$$

Moreover, this limit implies the following: $\lim_t \|\Pi_{\theta_t} T_{\theta_t}(\Phi\,\omega_t) - \Phi\,\omega_t\| = 0 \quad w.p.1\,.$

Remark 20. When the actor parameter θ_t is fixed (i.e., we are back to a policy evaluation problem), the second part of the above convergence result coincides with the widely known interpretation of the limit of the TD learning algorithm provided in Tsitsiklis and Van Roy (1997) (see also (Bertsekas and Tsitsiklis, 1996, p. 303-304)).

5.5.2 Actor analysis

Theorem 5.4. *Let Assumptions 5.3.1 and 5.5.1 to 5.5.4 hold true. Then, w.p.1*

$$\liminf_t \Big(\|\nabla J(\theta_t)\| - \|b(\theta_t)\|\Big) \le 0\,,$$

where for every $\theta \in \mathbb{R}^d, (s,a) \in \mathcal{S}\times\mathcal{A}$, $b(\theta) = \frac{1}{1-\gamma}\mathbb{E}_{\mu_{\rho,\theta}}[\psi_\theta(\tilde{S},\tilde{A})(\hat{Q}_\theta(\tilde{S},\tilde{A}) - Q_{\pi_\theta}(\tilde{S},\tilde{A}))]$ *and* $\hat{Q}_\theta(s,a) = R(s,a) + \gamma\sum_{s'\in\mathcal{S}} p(s'|s,a)\phi(s')^T\bar{\omega}_*(\theta)$.

Th. 5.4 is analog to (Konda, 2002, Th. 5.5) which is established for the standard on-policy actor-critic in the average reward setting and (Zhang et al., 2020b, Th. 3) for an off-policy actor-critic without any target network. The result states that the sequence (θ_t) generated by our actor-critic algorithm visits any neighborhood of the set $\{\theta \in \mathbb{R}^d : \|\nabla J(\theta)\| \le \|b(\theta)\|\}$ infinitely often. The bias $b(\theta)$ corresponds to the difference between the gradient $\nabla J(\theta)$ and the steady state expectation of the actor's update direction. The estimate used to update the actor in Eq. (5.6) is only a biased estimate of $\nabla J(\theta)$ because of linear FA.

Remark 21. The bias $b(\theta)$ disappears in the tabular setting ($m = |\mathcal{S}|$ and the features spanning $\mathbb{R}^{|\mathcal{S}|}$) when we do not use FA and in the linear FA setting when the value function belongs to the class of linear functions spanned by the pre-selected feature (or basis) functions. Beyond these particular settings, considering compatible features as introduced in Sutton et al. (1999); Konda and Tsitsiklis (2003) can be a solution to cancel the bias $b(\theta)$ incurred by Algorithm 5.1. We do not investigate this direction in this work.

5.6 Finite-time analysis

Our analysis in this section should be valid for a continuous state space $\mathcal{S}$ (and still finite action space) upon supposing that the feature map ϕ defined in Section 5.4.2 has bounded norm (i.e., $\|\phi(\cdot)\| \leq 1$) and slightly adapting our notations and definitions to this more general setting (see also for e.g., Wu et al. (2020)). To stay concise and consistent with the first part of our analysis in Section 5.5, we restrict ourselves to the finite state space setting.

5.6.1 Critic analysis

For every $\theta \in \mathbb{R}^d$, we suppose that the Markov chain $(\tilde{S}_t)$ induced by the policy π_θ and the transition kernel $\tilde{p}$ mixes at a geometric rate.

Assumption 5.6.1. There exist constants $c > 0$ and $\sigma \in (0, 1)$ s.t.

$$\sup_{s \in \mathcal{S}} d_{TV}(\mathbb{P}(\tilde{S}_t \in \cdot | \tilde{S}_0 = s, \pi_\theta), d_{\rho,\theta}) \leq c\sigma^t \,, \ \forall\, t \in \mathbb{N}\,, \forall\, \theta \in \mathbb{R}^d,$$

where $d_{TV}(\cdot, \cdot)$ denotes the total-variation distance between two probability measures.

This assumption is used to control the Markovian noise induced by sampling transitions from the MDP under a dynamically changing policy[1]. It was considered first in Bhandari et al. (2018) in a policy evaluation setting for the finite-time analysis of TD learning. It was later used for instance in Zou et al. (2019b); Wu et al. (2020); Shen et al. (2020).

We have seen in Section 5.5.1 that the dynamics of the critic is driven by two key matrices $-\bar{G}(\theta)$ and $-\bar{G}(\theta)^{-1}G(\theta)$. While we only need these matrices to be stable for our asymptotic results, we actually show in Section 5.7.1.1 that $-\bar{G}(\theta)$ is even negative definite uniformly in θ. We suppose that the second matrix $-\bar{G}(\theta)^{-1}G(\theta)$ is also negative definite uniformly in θ.

Assumption 5.6.2. There exists $\zeta > 0$ s.t. for every $\theta \in \mathbb{R}^d$, $\omega \in \mathbb{R}^m$,

$$\omega^T \bar{G}(\theta)^{-1} G(\theta) \omega \geq \zeta \|\omega\|^2 \,.$$

We are now ready to state our critic convergence rate.

Theorem 5.5. *Let Assumptions 5.3.1, 5.5.1 and 5.5.3 to 5.6.2 hold. Let $c_1, c_2, c_3, \alpha, \xi, \beta$ be positive constants s.t. $0 < \beta < \xi < \alpha < 1$. Set $\alpha_t = \frac{c_1}{(1+t)^\alpha}$, $\xi_t = \frac{c_2}{(1+t)^\xi}$ and $\beta_t = \frac{c_3}{(1+t)^\beta}$. Then, the sequences (ω_t) and (θ_t) generated by Algorithm 5.1 satisfy for*

[1] Note that the sup in the assumption is useful for more general (nondiscrete or infinite countable) state spaces.

every integer $T \geq 1$,

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\omega_t - \bar{\omega}_*(\theta_t)\|^2] = \mathcal{O}\left(\frac{1}{T^{1-\xi}}\right) + \mathcal{O}\left(\frac{\log T}{T^{\beta}}\right) + \mathcal{O}\left(\frac{1}{T^{2(\alpha-\xi)}}\right) + \mathcal{O}\left(\frac{1}{T^{2(\xi-\beta)}}\right) .$$

The bound of Th. 5.5 shows the impact of using a target variable. First, the last two terms impose the conditions $\alpha > \xi$ and $\xi > \beta$. At least with linear FA, this may provide a theoretical justification to the common practice of updating the target network at a slower rate compared to the main network for the critic. Second, compared to (Wu et al., 2020, Th. 4.7) which is concerned with the standard actor-critic in the average reward setting, we have the slower $\mathcal{O}(T^{\xi-1})$ instead of $\mathcal{O}(T^{\beta-1})$ and our bound comprises four error terms. These are also consequences of the use of a target variable.

5.6.2 Actor analysis

Assumption 5.6.3. There exists ϵ_{FA} s.t. for every $\theta \in \mathbb{R}^d$, $\|V_{\pi_\theta} - \Phi\,\bar{\omega}_*(\theta)\|_{D_{\rho,\theta}} \leq \epsilon_{\text{FA}}$.

Theorem 5.6. *Let Assumptions 5.3.1 , 5.5.1 , 5.5.3 to 5.6.1 and 5.6.3 hold. Let* $c_1, c_2, c_3, \alpha, \xi, \beta$ *be positive constants s.t.* $0 < \beta < \xi < \alpha < 1$. *Set* $\alpha_t = \frac{c_1}{(1+t)^{\alpha}}$, $\xi_t = \frac{c_2}{(1+t)^{\xi}}$ *and* $\beta_t = \frac{c_3}{(1+t)^{\beta}}$. *Then, for every integer* $T \geq 1$,

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\nabla J(\theta_t)\|^2] = \mathcal{O}\left(\frac{1}{T^{1-\alpha}}\right) + \mathcal{O}\left(\frac{\log^2 T}{T^{\alpha}}\right) + \mathcal{O}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\omega_t - \bar{\omega}_*(\theta_t)\|^2]\right) + \mathcal{O}\left(\epsilon_{FA}\right) .$$

Combining Th. 5.5 and Th. 5.6, we obtain the following result.

Corollary 5.7. Under the setting and the assumptions of Ths. 5.5 and 5.6 , we have for every $T \geq 1$,

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\nabla J(\theta_t)\|^2] = \mathcal{O}\left(\frac{1}{T^{1-\alpha}}\right) + \mathcal{O}\left(\frac{\log T}{T^{\beta}}\right) + \mathcal{O}\left(\frac{1}{T^{2(\alpha-\xi)}}\right) + \mathcal{O}\left(\frac{1}{T^{2(\xi-\beta)}}\right) + \mathcal{O}\left(\epsilon_{\text{FA}}\right) .$$

Moreover, if we set $\alpha = \frac{2}{3}$, $\xi = \frac{1}{2}$ and $\beta = \frac{1}{3}$ to define the stepsizes (α_t), (ξ_t) and (β_t), the actor parameter sequence (θ_t) generated by Algorithm 5.1 within $T = \mathcal{O}(\epsilon^{-3}\log^3(\frac{1}{\epsilon}))$ steps, satisfies

$$\min_{0\leq t\leq T}\mathbb{E}[\|\nabla J(\theta_t)\|^2] \leq \mathcal{O}(\epsilon_{\text{FA}}) + \epsilon\,.$$

As a consequence, since Algorithm 5.1 uses a single sample from the MDP per iteration, its sample complexity is $\mathcal{O}(\epsilon^{-3}\log^3(\frac{1}{\epsilon}))$. This is to compare with the best $\mathcal{O}(\epsilon^{-2}\log(\frac{1}{\epsilon}))$ sample complexity known in the literature (to the best of our knowledge) for actor-critic algorithms up to the linear FA error (Xu et al., 2020, Th. 2). Although the use of a target variable seems to deteriorate the sample complexity w.r.t. the best known result for target-free actor-critic methods, note that it is still aligned with the complexity reported in Qiu et al. (2019) (up to logarithmic factors), better than the $\mathcal{O}(\epsilon^{-4})$ sample complexity obtained in Kumar et al. (2019) with i.i.d. sampling and that we do not make use of mini-batching of samples (even from a single sample path) or nested loops as in Xu et al. (2020). We refer to (Wu et al., 2020, Section 4.4) and (Xu et al., 2020, Table 1) for further discussion. We briefly comment on the origin of this deteriorated

sample complexity stemming from our finite-time bounds. Due to the use of a target variable, instead of the $O(T^{2(\alpha-\beta)})$ error term of the standard actor-critic (see (Wu et al., 2020, Cor. 4.9) or (Shen et al., 2020, Ths.3-4)), we have two error terms $\mathcal{O}(T^{2(\alpha-\xi)})$ and $\mathcal{O}(T^{2(\xi-\beta)})$ slowing down the convergence because of the condition $\beta < \xi < \alpha$. Interestingly, at least in the linear FA setting, this corroborates the practical intuition that the use of a target network may slow learning as formulated for instance in (Lillicrap et al., 2016, Section 3) (even if constant stepsizes are used in practice).

Remark 22. Remark 21 also applies to the function approximation error ϵ_{FA}.

5.7 Proofs for Section 5.5

5.7.1 Proof of Th. 5.3

The objective of this section is to prove Th. 5.3. First, we recall the outline of the proof. Our actor-critic algorithm features three different timescales associated to three different stepsizes converging to zero with different rates, each one associated to one of the sequences $(\theta_t), (\bar{\omega}_t)$ and (ω_t). In spirit, we follow the strategy of (Borkar, 2008, Chap. 6, Lem. 1) for the analysis of two timescales stochastic approximation schemes. We make use of the results of Karmakar and Bhatnagar (2018) which handles controlled Markov noise. The proof is divided into three main steps:

(i) We start by analyzing the sequence (ω_t) evolving on the fastest timescale, i.e., with the stepsizes β_t which are converging the slowest to zero (see Assumption 5.5.2). We rewrite the slower sequences $(\theta_t), (\bar{\omega}_t)$ with the stepsizes β_t. In this timescale, $(\theta_t), (\bar{\omega}_t)$ are quasi-static from the point of view of the evolution of the sequence (ω_t). We deduce from this first step that ω_t tracks a slowly moving target $\omega_*(\theta_t, \bar{\omega}_t)$ governed by the slower iterates θ_t and $\bar{\omega}_t$. This is the purpose of Prop. 5.1 which is proved in Section 5.7.1.1 below.

(ii) In a second step, we analyze the sequence $(\bar{\omega}_t)$ which is evolving in a faster timescale than the sequence (θ_t) and slower than the sequence (ω_t). Similarly, we show that $\bar{\omega}_t$ tracks an other slowly moving target $\bar{\omega}_*(\theta_t)$. This is established in the proof of Prop. 5.2 in Section 5.7.1.2.

(iii) We conclude in Section5.7.1.3 by combining the results from the first two steps, proving that the sequence ω_t tracks the same target $\bar{\omega}_*(\theta_t)$.

5.7.1.1 Proof of Prop. 5.1

Let $\mathcal{F}_t$ be the σ-field generated by the random variables $S_l, \tilde{S}_l, \tilde{A}_l, \theta_l, \bar{\omega}_l, \omega_l$ for $l \leq t$. For each time step t, let $Z_t = (\tilde{S}_t, \tilde{A}_t)$. Our objective here is to show that the critic sequence (ω_t) tracks the slowly moving target $\omega_*(\theta_t, \bar{\omega}_t)$ defined in Prop. 5.1. From the update rule of the sequence (ω_t), we have

$$\begin{aligned}\omega_{t+1} &= \omega_t + \beta_t \bar{\delta}_{t+1} \phi(\tilde{S}_t) \\ &= \omega_t + \beta_t (R_{t+1} + \gamma \phi(S_{t+1})^T \bar{\omega}_t - \phi(\tilde{S}_t)^T \omega_t) \phi(\tilde{S}_t) \\ &= \omega_t + \beta_t w(\bar{\omega}_t, \omega_t, Z_t) + \beta_t \eta_{t+1}^{(1)}\,, \end{aligned} \tag{5.12}$$

where for every $\bar{\omega}, \omega \in \mathbb{R}^m, z = (s,a) \in \mathcal{S} \times \mathcal{A}$,

$$w(\bar{\omega}, \omega, z) := \left(R(s,a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s,a) \phi(s')^T \bar{\omega} \right) \phi(s) - \phi(s)\phi(s)^T \omega \tag{5.13}$$

and $\eta_{t+1}^{(1)}$ is a martingale difference sequence defined as

$$\eta_{t+1}^{(1)} = (R_{t+1} - R(\tilde{S}_t, \tilde{A}_t))\phi(\tilde{S}_t) + \gamma \bar{\omega}_t^T (\phi(S_{t+1}) - \mathbb{E}[\phi(S_{t+1})|\mathcal{F}_t])\, \phi(\tilde{S}_t)\,. \tag{5.14}$$

As can be seen in Eq. (5.12), the sequence (ω_t) can be written as a linear stochastic approximation scheme controlled by the slowly varying Markov chains (θ_t) and $(\bar{\omega}_t)$. In view of characterizing its asymptotic behavior, we compute for fixed $\bar{\omega}, \omega \in \mathbb{R}^m$ the expectation of the quantity $w(\bar{\omega}, \omega, Z)$ (see Eq. (5.13)) where $Z = (\tilde{S}, \tilde{A})$ is a random variable (on $\mathcal{S} \times \mathcal{A}$) following the stationary distribution $\mu_{\rho,\theta}$ (see Eq. (5.2)) of the Markov chain (Z_t). Recall the definitions of $\bar{h} : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^m$ and $\bar{G} : \mathbb{R}^d \to \mathbb{R}^{m \times m}$ from Eq. (5.10), for every $\theta \in \mathbb{R}^d, \bar{\omega} \in \mathbb{R}^m$

$$\bar{h}(\theta, \bar{\omega}) := \Phi^T D_{\rho,\theta}(R_\theta + \gamma P_\theta \Phi\, \bar{\omega}) \quad \text{and} \quad \bar{G}(\theta) := \Phi^T D_{\rho,\theta} \Phi\,.$$

Lemma 5.8. Under Assumption 5.5.1 , for every $\bar{\omega}, \omega \in \mathbb{R}^m$, we have

$$\mathbb{E}_{Z \sim \mu_{\rho,\theta}}[w(\bar{\omega}, \omega, Z)] = \bar{h}(\theta, \bar{\omega}) - \bar{G}(\theta)\omega\,.$$

Proof. We obtain from the definitions of w in Eq. (5.13) and $\mu_{\rho,\theta}$ in Eq. (5.2) that

$$\begin{aligned}
\mathbb{E}_{Z \sim \mu_{\rho,\theta}}[w(\bar{\omega}, \omega, Z)] &= \mathbb{E}_{Z \sim \mu_{\rho,\theta}} \left[\left(R(\tilde{S}, \tilde{A}) + \gamma \sum_{s' \in \mathcal{S}} p(s'|\tilde{S}, \tilde{A}) \phi(s')^T \bar{\omega} \right) \phi(\tilde{S}) - \phi(\tilde{S})\phi(\tilde{S})^T \omega \right] \\
&= \sum_{s \in \mathcal{S}, a \in \mathcal{A}} \mu_{\rho,\theta}(s,a) \left(R(s,a) + \gamma \sum_{s' \in \mathcal{S}} p(s'|s,a) \phi(s')^T \bar{\omega} \right) \phi(s) - \phi(s)\phi(s)^T \omega \\
&= \sum_{s \in \mathcal{S}} d_{\rho,\theta}(s) \left(R_\theta(s)\phi(s) + \gamma \sum_{s' \in \mathcal{S}} p_\theta(s'|s) \phi(s')^T \bar{\omega} \phi(s) - \phi(s)\phi(s)^T \omega \right) \\
&= \bar{h}(\theta, \bar{\omega}) - \bar{G}(\theta)\omega\,,
\end{aligned}$$

where the penultimate equation stems from recalling that $R_\theta(s) = \sum_{a \in \mathcal{A}} R(s,a)\pi_\theta(a|s)$ and $p_\theta(s'|s) = \sum_{a \in \mathcal{A}} p(s'|s,a)\pi_\theta(a|s)$ for every $s \in \mathcal{S}$. ∎

Defining $\chi_t = (\theta_t, \bar{\omega}_t)$, we obtain from the update rules of (θ_t) and $(\bar{\omega}_t)$ that

$$\chi_{t+1} = \chi_t + \beta_t \varepsilon_t\,, \tag{5.15}$$

where $\varepsilon_t = \left(\frac{\alpha_t}{\beta_t} \frac{1}{1-\gamma} \delta_{t+1} \psi_{\theta_t}(Z_t), \frac{\xi_t}{\beta_t}(\omega_{t+1} - \bar{\omega}_t) \right)$. Notice that $\epsilon_t \to 0$ as $t \to \infty$. This is because $\frac{\alpha_t}{\beta_t} \to 0$, $\frac{\xi_t}{\beta_t} \to 0$ by Assumption 5.5.2 , (ω_t) and (hence) $(\bar{\omega}_t)$ are a.s. bounded by Assumption 5.5.3 , (R_t) is bounded by U_R, $\theta \mapsto \psi_\theta(s,a)$ is bounded by Assumption 5.3.1 and $\mathcal{S}, \mathcal{A}$ are finite.

Let $\zeta_t = (\chi_t, \omega_t)$, $\zeta = (\theta, \bar{\omega}, \omega) \in \mathbb{R}^{d+2m}$, $W(\zeta, z) = (0, w(\bar{\omega}, \omega, z))$, $\varepsilon'_t = (\varepsilon_t, 0)$ and $\tilde{\eta}^{(1)}_{t+1} = (0, \eta^{(1)}_{t+1})$. Then, we can write Eqs. (5.15) and (5.12) in the framework of (Karmakar and Bhatnagar, 2018, Section 3, Eq.(14), Lem. 9), i.e., as a single timescale controlled Markov noise stochastic approximation scheme:

$$\zeta_{t+1} = \zeta_t + \beta_t[W(\zeta_t, Z_t) + \varepsilon'_t + \tilde{\eta}^{(1)}_{t+1}]\,, \tag{5.16}$$

with $\epsilon'_t \to 0$. Under the assumptions of Karmakar and Bhatnagar (2018) that we will verify at the end of the proof, we obtain that the sequence (ζ_t) converges to an internally chain transitive set (i.e., a compact invariant set which has no proper attractor, see definition in (Karmakar and Bhatnagar, 2018, Section 2.1) or (Benaïm, 1996, Section 1 p. 439)) of the ODE

$$\frac{d}{ds}\zeta(s) = \bar{W}(\zeta(s)) \quad \text{where} \quad \bar{W}(\zeta) = (0, \bar{h}(\chi) - \bar{G}(\theta)\omega)\,,$$

i.e.,

$$\begin{cases} \frac{d}{ds}\chi(s) &= 0\,, \\ \frac{d}{ds}\omega(s) &= \bar{h}(\chi(s)) - \bar{G}(\theta(s))\omega(s)\,. \end{cases} \tag{5.17}$$

As we will show that the second ODE governing ω has a unique asymptotically stable equilibrium $\omega_*(\theta, \bar{\omega})$ for every constant function $\chi(t) = \chi = (\theta, \bar{\omega})$, it follows that (χ_t, ω_t) converges a.s. towards the set $\{(\chi, \omega_*(\chi)) : \chi \in \mathbb{R}^{d+m}\}$. In other words, $\lim_t \|\omega_t - \omega_*(\theta_t, \bar{\omega}_t)\| = 0$, which is the desired result.

We now conclude the proof by verifying among (A1) to (A7) of Karmakar and Bhatnagar (2018) the assumptions under which (Karmakar and Bhatnagar, 2018, Lems. 9 and 10) hold.

(i) (A1): (Z_t) takes values in a compact metric space. Note that it is a finite state-action Markov chain controlled by the sequence (θ_t).

(ii) (A2): It is easy to see from Eq. (5.13) that the drift function w is Lipschitz continuous w.r.t. the variables $\bar{\omega}, \omega$ uniformly w.r.t. the last variable z because p is a probability kernel and the set of states $\mathcal{S}$ is finite.

(iii) (A3): $(\tilde{\eta}^{(1)}_{t+1})$ is a martingale difference sequence w.r.t. the filtration $(\mathcal{F}_t)$. Moreover, since (R_t) is bounded, there exists $K > 0$ s.t. $\mathbb{E}[\|\tilde{\eta}^{(1)}_{n+1}\|^2 | \mathcal{F}_t] \leq K(1 + \|\omega_t\|^2 + \|\bar{\omega}_t\|^2)$.

(iv) (A4): The stepsizes (β_t) satisfy $\sum_t \beta_t = +\infty$ and $\sum_t \beta_t^2 < \infty$ as formulated in Assumption 5.5.2.

(v) (A5): The transition kernel associated to the controlled Markov process (Z_t) is continuous w.r.t. the variables $z \in \mathcal{S} \times \mathcal{A}$, $\chi \in \mathbb{R}^{d+m}$, $\omega \in \mathbb{R}^m$. Continuity (w.r.t. to the metric of the weak convergence of probability measures) is a consequence of the fact that we have a finite-state MDP.

(vi) (A6'): We first note that the inverse of the matrix $\bar{G}(\theta)$ exists thanks to Assumptions 5.5.1 and 5.5.4. For all $\chi = (\theta, \bar{\omega}) \in \mathbb{R}^{d+m}$, we now show that the ODE $\frac{d}{ds}\omega(s) = \bar{h}(\chi) - \bar{G}(\theta)\omega(s)$ has a unique globally asymptotically stable equilibrium $\omega_*(\chi) = \bar{G}(\theta)^{-1}\bar{h}(\chi)$. The aforementioned ODE is stable if and only if the matrix $\bar{G}(\theta)$ is Hurwitz. We actually show that we have a stronger result in Lem. 5.9 under

Assumptions 5.5.1 and 5.5.4. We briefly explicit why the assumption as formulated in the rest of (A6') holds. Define the function $L(\chi,\omega) = \frac{1}{2}\|\bar{G}(\theta)\omega - \bar{h}(\chi)\|^2$. For every $\chi = (\theta,\bar{\omega}) \in \mathbb{R}^{d+m}$, the function $L(\chi,\cdot)$ is a Lyapunov function for ODE (5.17). Indeed, using Lem. 5.9 below, we can write

$$\frac{d}{ds}L(\chi,\omega(s)) = -\langle \bar{h}(\chi)-\bar{G}(\theta)\omega(s), \bar{G}(\theta)(\bar{h}(\chi)-\bar{G}(\theta)\omega(s))\rangle \leq -\varepsilon\|\bar{G}(\theta)\omega(s)-\bar{h}(\chi)\|^2 .$$

(vii) (A7): The stability Assumption 5.5.3 ensures that $\sup_t(\|\omega_t\| + \|\theta_t\|) < +\infty$ w.p.1. As a consequence, it also follows from the update rule of $(\bar{\omega}_t)$ that $\sup_t \|\bar{\omega}_t\| < +\infty$.

Lemma 5.9. Under Assumptions 5.5.1 and 5.5.4, there exists $\varepsilon > 0$ s.t. for all $\theta \in \mathbb{R}^d, \omega \in \mathbb{R}^m$,

$$\omega^T \bar{G}(\theta)\omega \geq \varepsilon\|\omega\|^2 .$$

In particular, it holds that $\sup_{\theta\in\mathbb{R}^d} \|\bar{G}(\theta)^{-1}\| < \infty$.

Proof. Recall that $\mathcal{K} := \{\tilde{K}_\theta : \theta \in \mathbb{R}^d\}$ where for every $\theta \in \mathbb{R}^d$, $\tilde{K}_\theta \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|\times|\mathcal{S}||\mathcal{A}|}$ is the transition matrix over the state-action pairs defined for every $(s,a),(s',a') \in \mathcal{S}\times\mathcal{A}$ by $\tilde{K}_\theta(s',a'|s,a) = \tilde{p}(s'|s,a)\pi_\theta(a'|s')$. We also denoted by $\bar{\mathcal{K}}$ the closure of $\mathcal{K}$. Under Assumption 5.5.1, there exists a unique stationary distribution $\mu_K \in \mathbb{R}^{\mathcal{S}\times\mathcal{A}}$ for every $K \in \bar{\mathcal{K}}$.

We first show that the map $K \mapsto \mu_K$ is continuous over the set $\bar{\mathcal{K}}$. The proof of this fact is similar to the proofs of (Zhang et al., 2021b, Lem. 9) and (Marbach and Tsitsiklis, 2001, Lem. 1). We reproduce a similar argument here for completeness. Observe first that μ_K satisfies:

$$M(K)\mu_K = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{where} \quad M(K) := \begin{bmatrix} K^T - I \\ \mathbb{1} \end{bmatrix} .$$

As a consequence, since $M(K)$ has full column rank thanks to Assumption 5.5.1, the matrix $M(K)^T M(K)$ is invertible and we obtain a closed form expression for μ_K given by:

$$\mu_K = (M(K)^T M(K))^{-1} M(K)^T \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{\mathrm{com}(M(K)^T M(K))^T}{\det(M(K)^T M(K))} M(K)^T \begin{bmatrix} 0 \\ 1 \end{bmatrix} ,$$

where $\mathrm{com}(A)$ stands for the comatrix of the matrix A. Then, it can be seen from this expression that the map $K \mapsto \mu_K$ is continuous. Note for this that the entries of the comatrix are polynomial functions of the entries of $M(K)^T M(K)$, and the determinant operator is continuous.

It follows from Assumption 5.5.1 that for every $K \in \bar{\mathcal{K}}$ and every $(s,a) \in \mathcal{S}\times\mathcal{A}$, $\mu_K(s,a) > 0$. We deduce from the continuity of the map $K \mapsto \mu_K$ over the compact set $\bar{\mathcal{K}}$ that $\inf_{K\in\bar{\mathcal{K}}} \mu_K(s,a) > 0$. Since $\tilde{K}_\theta \in \bar{K}$ for every $\theta \in \mathbb{R}^d$, we obtain that $\inf_\theta \mu_{\rho,\theta}(s,a) > 0$ where we recall that $\mu_{\rho,\theta}$ is the unique stationary distribution of the Markov chain induced by $\tilde{K}_\theta$. As a consequence, since $d_{\rho,\theta}(s) = \sum_{a\in\mathcal{A}} \mu_{\rho,\theta}(s,a)$, it also holds that

$$\inf_\theta d_{\rho,\theta}(s) > 0 .$$

Therefore, for every $\theta \in \mathbb{R}^d$, $\omega \in \mathbb{R}^m$:

$$\omega^T \bar{G}(\theta)\omega = (\Phi\,\omega)^T D_{\rho,\theta}(\Phi\,\omega) \geq \min_{s\in\mathcal{S}} \inf_\theta d_{\rho,\theta}(s)\|\Phi\omega\|^2 \geq \min_{s\in\mathcal{S}} \inf_\theta d_{\rho,\theta}(s)\lambda_{\min}(\Phi^T\Phi)\|\omega\|^2 ,$$

where $\lambda_{\min}(\Phi^T\Phi) > 0$ corresponds to the smallest eigenvalue of the symmetric positive definite matrix $\Phi^T\Phi$ which is invertible thanks to Assumption 5.5.4. The proof is concluded by setting $\varepsilon := \lambda_{\min}(\Phi^T\Phi)\cdot\min_{s\in\mathcal{S}} \inf_\theta d_{\rho,\theta}(s) > 0$ which is independent of θ. ■

5.7.1.2 Proof of Prop. 5.2

Recall the definitions of the vector $h(\theta)$ and the matrix $G(\theta)$ from Eq. (5.11):

$$h(\theta) := \Phi^T D_{\rho,\theta} R_\theta \quad \text{and} \quad G(\theta) := \Phi^T D_{\rho,\theta}(I_n - \gamma P_\theta)\Phi\,. \tag{5.18}$$

We begin the proof by showing the existence of a unique solution $\bar{\omega}_*(\theta)$ to the linear system $G(\theta)\bar{\omega} = h(\theta)$. The following lemma establishes the uniform positive definiteness of the matrix $G(\theta)$ Note that we do not include symmetry in our definition of positive definiteness as in Bertsekas and Tsitsiklis (1996). As a matter of fact, the matrix $G(\theta)$ is not symmetric in general.

Lemma 5.10. If Assumptions 5.5.1 and 5.5.4 hold, there exists $\kappa > 0$ s.t. for all $\theta \in \mathbb{R}^d$ and $\omega \in \mathbb{R}^m$,

$$\omega^T G(\theta)\omega \geq \kappa\|\omega\|^2\,.$$

In particular, the matrix $G(\theta)$ is invertible.

Proof. First, we have for every $\theta \in \mathbb{R}^d$, $\omega \in \mathbb{R}^m$,

$$\omega^T G(\theta)\omega = (\Phi\omega)^T D_{\rho,\theta}(I_n - \gamma P_\theta)\Phi\omega = (\Phi\omega)^T D_{\rho,\theta}(\Phi\omega) - \gamma(\Phi\omega)^T D_{\rho,\theta} P_\theta(\Phi\omega)\,. \tag{5.19}$$

Then, the Cauchy-Schwarz inequality yields

$$(\Phi\omega)^T D_{\rho,\theta} P_\theta(\Phi\omega) = (\Phi\omega)^T D_{\rho,\theta}^{\frac{1}{2}} D_{\rho,\theta}^{\frac{1}{2}} P_\theta(\Phi\omega) \leq \|\Phi\omega\|_{D_{\rho,\theta}} \|P_\theta\Phi\omega\|_{D_{\rho,\theta}}\,. \tag{5.20}$$

Notice now that we cannot use the classical result (Tsitsiklis and Van Roy, 1997, Lem. 1) to obtain that $\|P_\theta V\|_{D_{\rho,\theta}} \leq \|V\|_{D_{\rho,\theta}}$ for any $V \in \mathbb{R}^n$ because $D_{\rho,\theta}$ is not the stationary distribution of the kernel P_θ but it is instead associated to the artificial kernel $\tilde{P}_\theta$. Nevertheless, the following lemma provides an analogous result with a similar proof.

Lemma 5.11. For every $\theta \in \mathbb{R}^d$, $V \in \mathbb{R}^n$, we have

$$\|P_\theta V\|^2_{D_{\rho,\theta}} \leq \frac{1}{\gamma}\|V\|^2_{D_{\rho,\theta}} - \frac{1-\gamma}{\gamma}\|V\|^2_\rho \leq \frac{1}{\gamma}\|V\|^2_{D_{\rho,\theta}}\,.$$

Proof. It follows from Jensen's inequality that

$$\|P_\theta V\|^2_{D_{\rho,\theta}} = \sum_{i=1}^n d_{\rho,\theta}(s_i)\left(\sum_{j=1}^n P_\theta(s_j|s_i)V_j\right)^2 \leq \sum_{i=1}^n d_{\rho,\theta}(s_i)\sum_{j=1}^n P_\theta(s_j|s_i)V_j^2.$$

Then, observe that $\tilde{P}_\theta = \gamma P_\theta + (1-\gamma)\mathbb{1}\rho^T$ as a consequence of Eq. (5.3). By plugging this formula and then using the fact that $d_{\rho,\theta}^T\tilde{P}_\theta = d_{\rho,\theta}^T$, we obtain

$$\begin{aligned}
\sum_{i=1}^n d_{\rho,\theta}(s_i)\sum_{j=1}^n P_\theta(s_j|s_i)V_j^2 &= \frac{1}{\gamma}\left[\left(\sum_{j=1}^n\sum_{i=1}^n d_{\rho,\theta}(s_i)\tilde{P}_\theta(s_j|s_i)V_j^2\right) - (1-\gamma)\sum_{j=1}^n \rho(s_j)V_j^2\right] \\
&= \frac{1}{\gamma}\left[\sum_{j=1}^n d_{\rho,\theta}(s_j)V_j^2 - (1-\gamma)\sum_{j=1}^n \rho(s_j)V_j^2\right] \\
&= \frac{1}{\gamma}\|V\|^2_{D_{\rho,\theta}} - \frac{1-\gamma}{\gamma}\|V\|^2_\rho\,,
\end{aligned}$$

which concludes the proof of Lem. 5.11 . ■

We now complete the proof of Lem. 5.10 . From Eq. (5.20), Lem. 5.11 with $V = \Phi\omega$ yields

$$(\Phi\omega)^T D_{\rho,\theta} P_\theta(\Phi\omega) \leq \frac{1}{\sqrt{\gamma}} \|\Phi\omega\|^2_{D_{\rho,\theta}} = \frac{1}{\sqrt{\gamma}} (\Phi\omega)^T D_{\rho,\theta}(\Phi\omega) \,.$$

Whence, we obtain from Eq. (5.19) that

$$\omega^T G(\theta)\omega \geq (1-\sqrt{\gamma})(\Phi\omega)^T D_{\rho,\theta}(\Phi\omega) \geq \varepsilon(1-\sqrt{\gamma})\|\omega\|^2 \,,$$

where the last inequality stems from Lem. 5.9. ■

We now prove the remaining convergence results. We start with the first result showing that the sequence $(\bar{\omega}_t)$ tracks $\bar{\omega}_*(\theta_t)$. From the update rules of the sequences $(\bar{\omega}_t)$ and (ω_t) (Eqs. (5.8)-(5.9)), we can introduce the quantity $\omega_*(\theta_t, \bar{\omega}_t)$ as defined in Prop. 5.1 to obtain

$$\begin{aligned}
\bar{\omega}_{t+1} &= \bar{\omega}_t + \xi_t(\omega_{t+1} - \bar{\omega}_t) \\
&= \bar{\omega}_t + \xi_t(\omega_t + \beta_t w(\bar{\omega}_t, \omega_t, Z_t) + \beta_t \eta^{(1)}_{t+1} - \bar{\omega}_t) \\
&= \bar{\omega}_t + \xi_t(\omega_*(\theta_t, \bar{\omega}_t) - \bar{\omega}_t) + \xi_t(\omega_t - \omega_*(\theta_t, \bar{\omega}_t) + \beta_t w(\bar{\omega}_t, \omega_t, Z_t)) + \xi_t\beta_t\eta^{(1)}_{t+1} \,.
\end{aligned} \tag{5.21}$$

Then, using the expressions of $\bar{h}, \bar{G}$ in Eq. (5.10) and h, G in Eq. (5.11), we can write

$$\omega_*(\theta_t, \bar{\omega}_t) - \bar{\omega}_t = \bar{G}(\theta_t)^{-1}(\bar{h}(\theta_t, \bar{\omega}_t) - \bar{G}(\theta_t)\bar{\omega}_t) = \bar{G}(\theta_t)^{-1}(h(\theta_t) - G(\theta_t)\bar{\omega}_t) \,.$$

As a consequence,

$$\bar{\omega}_{t+1} = \bar{\omega}_t + \xi_t \bar{G}(\theta_t)^{-1}(h(\theta_t) - G(\theta_t)\bar{\omega}_t) + \xi_t(\omega_t - \omega_*(\theta_t, \bar{\omega}_t) + \beta_t w(\bar{\omega}_t, \omega_t, Z_t)) + \xi_t\beta_t\eta^{(1)}_{t+1} \,. \tag{5.22}$$

Therefore, the sequence $(\bar{\omega}_t)$ satisfies a linear stochastic approximation scheme driven by the slowly varying Markov chain (θ_t) evolving on a slower timescale than the iterates $(\bar{\omega}_t)$. We proceed similarly to the proof of Prop. 5.1 .

Recall the notation $\chi_t = (\theta_t, \bar{\omega}_t)$. Let $\chi = (\theta, \bar{\omega}) \in \mathbb{R}^{d+m}$, $U(\chi) = (0, \bar{G}(\theta)^{-1}(h(\theta) - G(\theta)\bar{\omega}))$. Then,

$$\chi_{t+1} = \chi_t + \xi_t[U(\chi_t) + \tilde{\varepsilon}_t] \,, \tag{5.23}$$

where $\tilde{\varepsilon}_t = (\frac{\alpha_t}{\xi_t}\frac{1}{1-\gamma}\delta_{t+1}\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t), \omega_t - \omega_*(\theta_t, \bar{\omega}_t) + \beta_t w(\bar{\omega}_t, \omega_t, Z_t) + \beta_t\eta^{(1)}_{t+1})$.

It can be shown that $\tilde{\varepsilon}_t \to 0$ as $t \to +\infty$. Note for this that $\alpha_t/\xi_t \to 0$ and $\beta_t \to 0$ by Assumption 5.5.2 , $\omega_t - \omega_*(\theta_t, \bar{\omega}_t) \to 0$ as proved in Prop. 5.1 and $\delta_{t+1}\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t), w(\bar{\omega}_t, \omega_t, Z_t)$ are bounded by Assumptions 5.3.1-*(c)* , 5.5.3, the boundedness of the reward function R and the fact that the sets $\mathcal{S}, \mathcal{A}$ are finite. Moreover, Assumption 5.5.2 ensures that $\sum_t \xi_t = +\infty$ and $\sum_t \xi_t^2 < +\infty$.

Furthermore, one can show that the function U is Lipschitz continuous. For this, remark that:

(a) The function U is affine in $\bar{\omega}$.

(b) The functions $\theta \mapsto R_\theta$ and $\theta \mapsto P_\theta$ are Lipschitz continuous as $P_\theta(s'|s) = p(s'|s,a)\pi_\theta(a|s)$, $R_\theta(s) = \sum_{a\in\mathcal{A}} R(s,a)\pi_\theta(a|s)$ and Assumption 5.3.1-*(b)* guarantees that $\theta \mapsto \pi_\theta(a|s)$ is Lipschitz continuous for every $(s,a) \in \mathcal{S}\times\mathcal{A}$.

(c) The function $\theta \mapsto D_{\rho,\theta}$ is Lipschitz continuous. We refer to (Zhang et al., 2021b, Lem. 9) for a proof.

(d) The function $\theta \mapsto \bar{G}(\theta)^{-1}$ is Lipschitz continuous. Observe for this that for every $\theta,\theta' \in \mathbb{R}^d$, $\bar{G}(\theta)^{-1} - \bar{G}(\theta')^{-1} = \bar{G}(\theta)^{-1}(\bar{G}(\theta') - \bar{G}(\theta))\bar{G}(\theta')^{-1}$ and that $\sup_\theta \|\bar{G}(\theta)^{-1}\| < \infty$ using Lem. 5.9.

(e) The reward function R is bounded and the entries of the matrices $D_{\rho,\theta}$ and P_θ are bounded by one.

Using classical stochastic approximation results (see, for e.g., (Benaïm, 1996, Th.1.2)), we obtain that the sequence (χ_t) converges a.s. towards an internally chain transitive set of the ODE $\frac{d}{ds}\chi(s) = U(\chi(s))$, i.e.,

$$\begin{cases} \frac{d}{ds}\theta(s) &= 0\,, \\ \frac{d}{ds}\bar{\omega}(s) &= \bar{G}(\theta(s))^{-1}(h(\theta(s)) - G(\theta(s))\bar{\omega}(s))\,. \end{cases} \tag{5.24}$$

We conclude by showing that for every $\theta \in \mathbb{R}^d$, the ODE $\frac{d}{ds}\bar{\omega}(s) = \bar{G}(\theta)^{-1}(h(\theta) - G(\theta)\bar{\omega}(s))$ has a globally asymptotically stable equilibrium $\bar{\omega}_*(\theta)$. This result holds if the matrix $-\bar{G}(\theta)^{-1}G(\theta)$ is Hurwitz, i.e., all its eigenvalues have negative real parts. We show this result in Lem. 5.12 below. Then, it follows that $\chi_t = (\theta_t, \bar{\omega}_t)$ converges a.s. towards the set $\{(\theta, \bar{\omega}_*(\theta)) : \theta \in \mathbb{R}^d\}$. This yields the desired result $\lim_t \|\bar{\omega}_t - \bar{\omega}_*(\theta_t)\| = 0$.

Lemma 5.12. For every $\theta \in \mathbb{R}^d$, the matrix $-\bar{G}(\theta)^{-1}G(\theta)$ is Hurwitz.

Proof. We first recall Lyapunov's theorem which characterizes Hurwitz matrices (see, for e.g., (Horn and Johnson, 1994, Th.2.2.1 p. 96)). A complex matrix A is Hurwitz if and only if there exists a positive definite matrix $M = M^*$ s.t. $A^*M + MA$ is negative definite, where M^* and A^* are the complex conjugate transposes of M and A. We use this theorem with $A = -\bar{G}(\theta)^{-1}G(\theta)$ and $M = \bar{G}(\theta)$ which is symmetric by definition and positive definite thanks to Lem. 5.9. Then, we obtain that

$$A^*M + MA = -G(\theta)^T\bar{G}(\theta)^{-1}\bar{G}(\theta) - \bar{G}(\theta)\bar{G}(\theta)^{-1}G(\theta) = -(G(\theta)^T + G(\theta))\,.$$

We conclude the proof by showing that $G(\theta)^T + G(\theta)$ is a (symmetric) positive definite matrix. For that, observe that for every nonzero vector $\omega \in \mathbb{R}^m$, it holds that $\omega^T(G(\theta)^T + G(\theta))\omega = 2\omega^T G(\theta)\omega > 0$ where the positivity stems from Lem. 5.10. ∎

The last result states that for every $\theta \in \mathbb{R}^d$, $\Phi\bar{\omega}_*(\theta)$ is a fixed point of the projected Bellman operator $\Pi_\theta T_\theta$. This is a consequence of the following derivations:

$$\begin{aligned} \Pi_\theta T_\theta(\Phi\bar{\omega}_*(\theta)) &= \Phi\bar{G}(\theta)^{-1}\Phi^T D_{\rho,\theta}T_\theta(\Phi\bar{\omega}_*(\theta)) \\ &= \Phi\bar{G}(\theta)^{-1}\Phi^T D_{\rho,\theta}(R_\theta + \gamma P_\theta\Phi\bar{\omega}_*(\theta)) \\ &= \Phi\bar{G}(\theta)^{-1}h(\theta) + \Phi\bar{G}(\theta)^{-1}(\bar{G}(\theta) - G(\theta))G(\theta)^{-1}h(\theta) \\ &= \Phi\bar{G}(\theta)^{-1}h(\theta) + \Phi G(\theta)^{-1}h(\theta) - \Phi\bar{G}(\theta)^{-1}h(\theta) \\ &= \Phi G(\theta)^{-1}h(\theta) \\ &= \Phi\bar{\omega}_*(\theta)\,, \end{aligned} \tag{5.25}$$

where the first equality uses the expression of the projection Π_θ, the second one uses the definition of the Bellman operator T_θ and the third one stems from the definitions of the matrices $\bar{G}(\theta)$ and $G(\theta)$ (see Eqs. (5.10) and (5.11)).

5.7.1.3 Proof of Th. 5.3

The proof of Th. 5.3 uses both Prop. 5.1 and Prop. 5.2 .

In order to show that $\lim_t \|\omega_t - \bar{\omega}_*(\theta_t)\| = 0$ $w.p.1$,we prove the two following results:

(a) $\lim_t \|\omega_t - \omega_*(\theta_t, \bar{\omega}_*(\theta_t))\| = 0$ $w.p.1$.

(b) $\omega_*(\theta, \bar{\omega}_*(\theta)) = \bar{\omega}_*(\theta)$ for all $\theta \in \mathbb{R}^d$.

(a) We have the decomposition

$$\begin{aligned}
\omega_t - \omega_*(\theta_t, \bar{\omega}_*(\theta_t)) &= [\omega_t - \omega_*(\theta_t, \bar{\omega}_t)] + [\omega_*(\theta_t, \bar{\omega}_t) - \omega_*(\theta_t, \bar{\omega}_*(\theta_t))], \\
&= [\omega_t - \omega_*(\theta_t, \bar{\omega}_t)] + \bar{G}(\theta_t)^{-1}(\bar{h}(\theta_t, \bar{\omega}_t) - \bar{h}(\theta_t, \bar{\omega}_*(\theta_t))) \\
&= [\omega_t - \omega_*(\theta_t, \bar{\omega}_t)] + \bar{G}(\theta_t)^{-1}\Phi^T D_{\rho,\theta_t} P_{\theta_t} \Phi(\bar{\omega}_t - \bar{\omega}_*(\theta_t)) \\
&= [\omega_t - \omega_*(\theta_t, \bar{\omega}_t)] + \bar{G}(\theta_t)^{-1}(\bar{G}(\theta_t) - G(\theta_t))(\bar{\omega}_t - \bar{\omega}_*(\theta_t)) \\
&= [\omega_t - \omega_*(\theta_t, \bar{\omega}_t)] + (I_m - \bar{G}(\theta_t)^{-1}G(\theta_t))(\bar{\omega}_t - \bar{\omega}_*(\theta_t)) \,. \qquad (5.26)
\end{aligned}$$

It follows from Prop. 5.1 that the first term in the above decomposition goes to zero. Then, observe that $\sup_\theta \|\bar{G}(\theta)^{-1}\| < \infty$ given Lem. 5.9 and $\sup_\theta \|G(\theta)\| < \infty$ thanks to the boundedness of the matrices P_θ and $D_{\rho,\theta}$ uniformly in θ. As a consequence, the second term also converges to zero using Prop. 5.2 .

(b) Using the definitions of the functions ω_* and $\bar{\omega}_*$, we can write for every $\theta \in \mathbb{R}^d$,

$$\begin{aligned}
\omega_*(\theta, \bar{\omega}_*(\theta)) &= \bar{G}(\theta)^{-1}\bar{h}(\theta, \bar{\omega}_*(\theta)) \\
&= \bar{G}(\theta)^{-1}\Phi^T D_{\rho,\theta}(R_\theta + \gamma P_\theta \Phi G(\theta)^{-1} h(\theta)) \\
&= \bar{G}(\theta)^{-1}(h(\theta) + \gamma \Phi^T D_{\rho,\theta} P_\theta \Phi G(\theta)^{-1} h(\theta)) \\
&= \bar{G}(\theta)^{-1}(I_n + \gamma \Phi^T D_{\rho,\theta} P_\theta \Phi G(\theta)^{-1}) h(\theta) \\
&= \bar{G}(\theta)^{-1}(G(\theta) + \gamma \Phi^T D_{\rho,\theta} P_\theta \Phi) G(\theta)^{-1} h(\theta) \\
&= \bar{G}(\theta)^{-1}\bar{G}(\theta) G(\theta)^{-1} h(\theta) \\
&= \bar{\omega}_*(\theta) \,.
\end{aligned}$$

For the last result, we write

$$\begin{aligned}
\|\Pi_{\theta_t} T_{\theta_t}(\Phi\omega_t) - \Phi\omega_t\| &= \|\Phi\left(\bar{G}(\theta_t)^{-1}\Phi^T D_{\rho,\theta_t} T_{\theta_t}(\Phi\omega_t) - \omega_t\right)\| \\
&= \|\Phi\left(\bar{G}(\theta_t)^{-1}\Phi^T D_{\rho,\theta_t}(T_{\theta_t}(\Phi\omega_t) - \Phi\omega_t)\right)\| \\
&= \|\Phi\left(\bar{G}(\theta_t)^{-1}(h(\theta_t) - G(\theta_t)\omega_t)\right)\| \\
&= \|\Phi\bar{G}(\theta_t)^{-1}G(\theta_t)(\omega_t - \bar{\omega}_*(\theta_t))\| \\
&\leq \|\Phi\|\|\bar{G}(\theta_t)^{-1}\|\|G(\theta_t)\|\|\omega_t - \bar{\omega}_*(\theta_t)\| \,. \qquad (5.27)
\end{aligned}$$

Then, as previously mentioned in the proof, observe that $\sup_\theta \|\bar{G}(\theta)^{-1}\| < \infty$ and $\sup_\theta \|G(\theta)\| < \infty$. Since $\bar{\omega}_t - \bar{\omega}_*(\theta_t) \to 0$ as $t \to \infty$, the result follows.

5.7.2 Proof of Th. 5.4

In this subsection, we present a proof of Th. 5.4 which is similar in spirit to the proof in (Konda and Tsitsiklis, 2003, Section 6). Recall the notation $Z_t = (\tilde{S}_t, \tilde{A}_t)$. Note that (Z_t) is a Markov chain. The actor parameter θ_t iterates as follows:

$$\begin{aligned}\theta_{t+1} &= \theta_t + \alpha_t \frac{1}{1-\gamma} \delta_{t+1} \psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t) \\ &= \theta_t + \alpha_t \frac{1}{1-\gamma} (R_{t+1} + (\gamma \phi(S_{t+1}) - \phi(\tilde{S}_t))^T \omega_t) \psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t) \\ &= \theta_t + \alpha_t \frac{1}{1-\gamma} (R(\tilde{S}_t, \tilde{A}_t) \psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t) + H_{\theta_t}(Z_t)\omega_t) + \alpha_t \frac{1}{1-\gamma} \tilde{\eta}_{t+1}\,,\end{aligned}$$

where for every $\theta \in \mathbb{R}^d, z = (s,a) \in \mathcal{S} \times \mathcal{A}$,

$$H_\theta(z) = \psi_\theta(s,a) \left(\gamma \sum_{s' \in \mathcal{S}} p(s'|s,a) \phi(s') - \phi(s) \right)^T ,$$

and $(\tilde{\eta}_{t+1})$ is an $\mathbb{R}^d$-valued $\mathcal{F}_t$-martingale difference sequence defined by

$$\tilde{\eta}_{t+1} = (R_{t+1} - \mathbb{E}[R_{t+1}|\mathcal{F}_t]) \psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t) \; + \gamma \psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t)(\phi(S_{t+1}) - \mathbb{E}[\phi(S_{t+1})|\mathcal{F}_t])^T \omega_t\,. \tag{5.28}$$

We now introduce the steady-state expectation of the main term $H_\theta(Z_t)\omega_t + R(\tilde{S}_t, \tilde{A}_t)\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t)$. Recall that $\mu_{\rho,\theta}$ is the stationary distribution of the Markov chain (Z_t). Define the functions $\bar{H} : \mathbb{R}^d \to \mathbb{R}^{d \times m}$ and $u : \mathbb{R}^d \to \mathbb{R}^d$ for every $\theta \in \mathbb{R}^d$ by

$$\bar{H}(\theta) = \mathbb{E}_{Z \sim \mu_{\rho,\theta}}[H_\theta(Z)]\,, \tag{5.29}$$

$$u(\theta) = \mathbb{E}_{Z \sim \mu_{\rho,\theta}}[R(\tilde{S}, \tilde{A}) \psi_\theta(\tilde{S}, \tilde{A})]\,, \tag{5.30}$$

where $Z = (\tilde{S}, \tilde{A})$ is a random variable following the distribution $\mu_{\rho,\theta}$.

Then, we introduce the quantity $\bar{\omega}_*(\theta_t)$ which approximates well ω_t for large t (in the sense of Th. 5.3) and only depends on the actor parameter θ_t. We obtain the following decomposition

$$\theta_{t+1} = \theta_t + \alpha_t f(\theta_t) + \alpha_t \frac{1}{1-\gamma} (\tilde{\eta}_{t+1} + e_t^{(1)} + e_t^{(2)})\,, \tag{5.31}$$

where the function $f : \mathbb{R}^d \to \mathbb{R}^d$ and the error terms $e_t^{(1)}$ and $e_t^{(2)}$ are defined as follows

$$f(\theta) = \frac{1}{1-\gamma} (\bar{H}(\theta)\, \bar{\omega}_*(\theta) + u(\theta))\,, \tag{5.32}$$

$$e_t^{(1)} = (R(\tilde{S}_t, \tilde{A}_t)\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t) + H_{\theta_t}(Z_t)\bar{\omega}_*(\theta_t)) - (\bar{H}(\theta_t)\, \bar{\omega}_*(\theta_t) + u(\theta_t))\,, \tag{5.33}$$

$$e_t^{(2)} = H_{\theta_t}(Z_t)(\omega_t - \bar{\omega}_*(\theta_t))\,. \tag{5.34}$$

The bias induced by the approximation of $\nabla J(\theta)$ by our actor-critic algorithm is defined for every $\theta \in \mathbb{R}^d$ by

$$b(\theta) := f(\theta) - \nabla J(\theta)\,. \tag{5.35}$$

This bias is due to the linear FA of the true state-value function. It is defined as the difference between the steady-state expectation of the actor update given by the function f defined in Eq. (5.32) and the gradient $\nabla J(\theta)$ we are interested in. The following lemma provides a more explicit and interpretable expression for the bias $b(\theta)$. The state-value function V_{π_θ} will be seen as a vector of $\mathbb{R}^{|\mathcal{S}|}$.

Lemma 5.13. For every $\theta \in \mathbb{R}^d$,

$$b(\theta) = \frac{\gamma}{1-\gamma} \sum_{s\in\mathcal{S}, a\in\mathcal{A}} \mu_{\rho,\theta}(s,a)\psi_\theta(s,a) \sum_{s'\in\mathcal{S}} p(s'|s,a)(\phi(s')^T \bar{\omega}_*(\theta) - V_{\pi_\theta}(s'))\,.$$

Proof. The expression follows from using the definition of $b(\theta)$ and computing both the function $\bar{H}$ defined in Eq. (5.29) and the gradient of the function J.

First, we explicit the function $\bar{H}$, writing

$$\begin{aligned}
\bar{H}(\theta) = \mathbb{E}_{Z\sim\mu_{\rho,\theta}}[H_\theta(Z)] &= \mathbb{E}_{Z\sim\mu_{\rho,\theta}}\left[\psi_\theta(\tilde{S},\tilde{A})\left(\gamma \sum_{s'\in\mathcal{S}} p(s'|\tilde{S},\tilde{A})\phi(s') - \phi(\tilde{S})\right)^T\right] \\
&= \sum_{s\in\mathcal{S}, a\in\mathcal{A}} \mu_{\rho,\theta}(s,a)\psi_\theta(s,a)\left(\gamma \sum_{s'\in\mathcal{S}} p(s'|s,a)\phi(s')^T - \phi(s)^T\right) \\
&= \gamma \sum_{s\in\mathcal{S}, a\in\mathcal{A}} \mu_{\rho,\theta}(s,a)\psi_\theta(s,a) \sum_{s'\in\mathcal{S}} p(s'|s,a)\phi(s')^T\,, \qquad (5.36)
\end{aligned}$$

where the last equality stems from remarking that $\sum_{a\in\mathcal{A}} \mu_{\rho,\theta}(s,a)\psi_\theta(s,a) = 0$.

Then, the policy gradient theorem as formulated in Eq. (5.1) and the definition of the advantage function provide

$$\begin{aligned}
(1-\gamma)\nabla J(\theta) &= \mathbb{E}_{Z\sim\mu_{\rho,\theta}}[\Delta_{\pi_\theta}(\tilde{S},\tilde{A})\psi_\theta(\tilde{S},\tilde{A})] \\
&= \mathbb{E}_{Z\sim\mu_{\rho,\theta}}[(R(\tilde{S},\tilde{A}) + \gamma \sum_{s'\in\mathcal{S}} p(s'|\tilde{S}_t,\tilde{A}_t)V_{\pi_\theta}(s') - V_{\pi_\theta}(\tilde{S}))\psi_\theta(\tilde{S},\tilde{A})] \\
&= \sum_{s,a} \mu_{\rho,\theta}(s,a)(R(s,a) + \gamma \sum_{s'\in\mathcal{S}} p(s'|s,a)V_{\pi_\theta}(s') - V_{\pi_\theta}(s))\psi_\theta(s,a) \\
&= u(\theta) + \gamma \sum_{s\in\mathcal{S}, a\in\mathcal{A}} \mu_{\rho,\theta}(s,a)\psi_\theta(s,a) \sum_{s'\in\mathcal{S}} p(s'|s,a)V_{\pi_\theta}(s')\,. \qquad (5.37)
\end{aligned}$$

The result stems from using the definition of $b(\theta)$ together with Eqs. (5.36) and (5.37). ■

Using a second-order Taylor expansion of the $\tilde{L}$-Lipschitz function ∇J (again see (Zhang et al., 2020a, Lem. 4.2)) together with Eq. (5.31), we can derive the following inequalities

$$\begin{aligned}
J(\theta_{t+1}) &\geq J(\theta_t) + \langle \nabla J(\theta_t), \theta_{t+1} - \theta_t\rangle - L\|\theta_{t+1} - \theta_t\|^2\,, \\
&\geq J(\theta_t) + \alpha_t \langle \nabla J(\theta_t), f(\theta_t)\rangle \\
&\quad + \frac{\alpha_t}{1-\gamma}\langle \nabla J(\theta_t), \tilde{\eta}_{t+1} + e_t^{(1)} + e_t^{(2)}\rangle - \tilde{L}\frac{\alpha_t^2}{(1-\gamma)^2}\|\delta_{t+1}\psi_{\theta_t}(\tilde{S}_t,\tilde{A}_t)\|^2\,. \qquad (5.38)
\end{aligned}$$

The above inequality consists of a main term involving the function f and noise terms. The following lemma controls these noise terms which are shown to be negligible.

Lemma 5.14. *(a)* $\sum_{t=0}^{\infty} \alpha_t \langle \nabla J(\theta_t), e_t^{(1)}\rangle < \infty$ w.p.1 ,

(b) $\sum_{t=0}^{\infty} \alpha_t \langle \nabla J(\theta_t), \tilde{\eta}_{t+1}\rangle < \infty$ w.p.1 ,

(c) $\lim_{t\to\infty} e_t^{(2)} = 0$, w.p.1 ,

(d) $\sum_{t=0}^{\infty} \alpha_t^2 \|\delta_{t+1}\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t)\|^2 < \infty$ w.p.1 .

Proof.

(a) The proof is based on the classical decomposition of the Markov noise term $e_t^{(1)}$ using the Poisson equation (Benveniste et al., 1990, p. 222-229). We refer to (Zhang et al., 2020b, Lem. 7 and Section A.8.3) for a detailed proof using this technique. The proof of our result here follows the same line. For conciseness, we only describe the necessary tools, pointing out the differences with (Zhang et al., 2020b, Lem. 7 and Section A.8.3) which is concerned with a different algorithm.

Let $\mathcal{Z} := \mathcal{S} \times \mathcal{A}$. First, define the functions $g_\theta^* : \mathcal{Z} \to \mathbb{R}^d$ and $\bar{g} : \mathbb{R}^d \to \mathbb{R}^d$ by:

$$g_\theta^*(z) := R(z)\psi_\theta(z) + H_\theta(z)\bar{\omega}_*(\theta) , \tag{5.39}$$

$$\bar{g}(\theta) := u(\theta) + \bar{H}(\theta)\bar{\omega}_*(\theta) , \tag{5.40}$$

for every $z = (s, a) \in \mathcal{Z}, \theta \in \mathbb{R}^d$. Observe in particular that $e_t^{(1)} = g_{\theta_t}^*(\tilde{S}_t, \tilde{A}_t) - \bar{g}(\theta_t)$. Recall that for every $\theta \in \mathbb{R}^d$, the kernel transition $\tilde{K}_\theta$ is defined for every $(s, a), (s', a') \in \mathcal{S} \times \mathcal{A}$ by $\tilde{K}_\theta(s', a') = \tilde{p}(s'|s, a)\pi_\theta(a'|s')$ (see Assumption 5.5.1). The idea of the proof is to introduce for each integer $i = 1, \cdots, d$ a Markov Reward Process (MRP) (Puterman, 2014, Section 8.2) on the space $\mathcal{Z}$ induced by the transition kernel $\tilde{K}_\theta$ and the reward function $g_{\theta,i}^*$ (ith coordinate of the function g_θ^*). As a consequence, the corresponding average reward is given by $\bar{g}_i(\theta)$ (ith coordinate of $\bar{g}(\theta)$). Then, the differential value function of the MRP is provided by $v_{\theta,i} := (I - \tilde{K}_\theta + \mathbb{1}\mu_{\rho,\theta}^T)^{-1}(I - \mathbb{1}\mu_{\rho,\theta}^T)g_{\theta,i}^*$ as shown for instance in (Puterman, 2014, Section 8.2). The functions $v_{\theta,i}$ for $i = 1, \cdots, d$ define together a vector valued function $v_\theta : \mathcal{Z} \to \mathbb{R}^d$. Under Assumption 5.5.1, using similar arguments to the proof of Lem. 5.9 (see also (Zhang et al., 2021b, Proof of Lem. 4, p. 26)), we can show that the function $K \in \bar{\mathcal{K}} \mapsto (I - K + \mathbb{1}\mu_K^T)^{-1}(I - \mathbb{1}\mu_K^T)$ is continuous on the compact set $\bar{\mathcal{K}}$. It follows that $\sup_{\theta,z} \|v_\theta(z)\| < \infty$ because $\tilde{K}_\theta \in \bar{\mathcal{K}}$ for every $\theta \in \mathbb{R}^d$ and $g_{\theta,i}^*$ is uniformly bounded w.r.t. θ under our assumptions . Moreover, the differential value function satisfies the crucial Bellman equation:

$$v_\theta(z) = g_\theta^*(z) - \bar{g}(\theta) + \sum_{z' \in \mathcal{Z}} \tilde{K}_\theta(z'|z)v_\theta(z) ,$$

for every $z \in \mathcal{Z}$. We use the above Poisson equation to express $e_t^{(1)} = g_{\theta_t}^*(\tilde{S}_t, \tilde{A}_t) - \bar{g}(\theta_t)$ using v_θ. The rest of the proof follows the same line as (Zhang et al., 2020b, Lem. 7 and Section A.8.3).

(b) First, recall that $(\tilde{\eta}_t)$ is a martingale difference sequence adapted to $\mathcal{F}_t$ and so is $(\langle \nabla J(\theta_t), \tilde{\eta}_{t+1} \rangle)$. Using the boundedness of the function $\theta \to \psi_\theta(s, a)$ guaranteed by Assumption 5.3.1-*(c)* with the boundedness of the rewards sequence (R_t), the sequence (ω_t) (Assumption 5.5.3) and the gradient ∇J , one can show by Cauchy-Schwarz inequality that there exists a constant $C > 0$ s.t. $\mathbb{E}[|\langle \nabla J(\theta_t), \tilde{\eta}_{t+1} \rangle|^2 | \mathcal{F}_t] \leq C$ a.s. Then, using that $\sum_t \alpha_t^2 < \infty$ (Assumption 5.5.2), it follows that $\sum_t \mathbb{E}[|\alpha_t \langle \nabla J(\theta_t), \tilde{\eta}_{t+1} \rangle|^2 | \mathcal{F}_t] < \infty$ a.s. We deduce from Doob's convergence theorem that item *(b)* holds.

(c) As for item *(c)*, we first observe that $\bar{H}(\theta_t)$ is bounded since $\theta \mapsto \psi_\theta(s,a)$ is bounded for every $(s,a) \in \mathcal{S} \times \mathcal{A}$ thanks again to Assumption 5.3.1-*(c)*. Then, item *(c)* stems from the fact that $\omega_t - \bar{\omega}_*(\theta_t) \to 0$ as shown in Th. 5.3 .

(d) Similarly to $\bar{H}(\theta_t)$, upon noticing that the reward sequence (R_t) is bounded by U_R and the sequence (ω_t) is a.s. bounded by Assumption 5.5.3 , the quantity $\delta_{t+1}\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t)$ is also a.s. bounded. Then, item *(d)* is a consequence of the square summability of the stepsizes α_t ($\sum_t \alpha_t^2 < \infty$) as guaranteed by Assumption 5.5.2.

■

The end of the proof follows the same line as (Konda and Tsitsiklis, 2003, p. 1163) (see also (Konda, 2002, p. 86)). We reproduce the argument here for completeness. Let $T > 0$. Define a sequence k_t by

$$k_0 = 0\,, \quad k_{t+1} = \min \left\{ k \geq k_t : \sum_{i=k_t}^{k} \alpha_i \geq T \right\} \quad \text{for } t > 0\,.$$

Using Eq. (5.38) together with the Cauchy-Schwartz inequality and Eq. (5.35), we can write

$$J(\theta_{k_{t+1}}) \geq J(\theta_{k_t}) + \sum_{k=k_t}^{k_{t+1}-1} \alpha_k (\|\nabla J(\theta_k)\|^2 - \|b(\theta_k)\| \cdot \|\nabla J(\theta_k)\|) + \upsilon_t\,,$$

where υ_t is defined by

$$\upsilon_t = \sum_{k=k_t}^{k_{t+1}-1} \left(\frac{\alpha_k}{1-\gamma} \langle \nabla J(\theta_k), \tilde{\eta}_{k+1} + e_k^{(1)} + e_k^{(2)} \rangle - \tilde{L} \frac{\alpha_k^2}{(1-\gamma)^2} \|\delta_{k+1} \psi_{\theta_k}(\tilde{S}_k, \tilde{A}_k)\|^2 \right)\,.$$

It stems from Lem. 5.14 that $\upsilon_t \to 0$ as $t \to +\infty$. By contradiction, if the result does not hold, the sequence $J(\theta_k)$ would increase indefinitely. This contradicts the boundedness of the function J (note that $\theta \mapsto V_{\pi_\theta}$ is bounded since the rewards are bounded).

5.8 Proofs for Section 5.6

Throughout our finite-time analysis, we will not track the constants although these can be precisely determined. The universal constant C may change from line to line and from inequality to inequality. It may depend on constants of the problem s.t. the Lipschitz constants of the functions $J, \theta \mapsto \psi_\theta$, $\theta \mapsto \pi_\theta$, upperbounds of the rewards and the score function ψ_θ or the cardinal $|\mathcal{A}|$ of the action space .

5.8.1 Proof of Th. 5.5

The proof is inspired from the recent works Wu et al. (2020); Shen et al. (2020). However, it significantly deviates from these works because of the use of a target variable $\bar{\omega}$ in Algorithm 5.1 . In particular, as previously mentioned, Algorithm 5.1 involves three different timescales whereas the actor-critic algorithms considered in Wu et al. (2020);

Shen et al. (2020) only use two different timescales respectively associated to the critic and the actor.

We follow a similar strategy to our asymptotic analysis of the critic. Indeed, our non-asymptotic analysis consists of two main steps based on the following decomposition:

$$\begin{aligned}\omega_t - \bar{\omega}_*(\theta_t) &= \omega_t - \omega_*(\theta_t, \bar{\omega}_t) + \omega_*(\theta_t, \bar{\omega}_t) - \bar{\omega}_*(\theta_t) \\ &= \omega_t - \omega_*(\theta_t, \bar{\omega}_t) + \omega_*(\theta_t, \bar{\omega}_t) - \omega_*(\theta_t, \bar{\omega}_*(\theta_t)) \\ &= \omega_t - \omega_*(\theta_t, \bar{\omega}_t) + \bar{G}(\theta_t)^{-1}(\bar{h}(\theta_t, \bar{\omega}_t) - \bar{h}(\theta_t, \bar{\omega}_*(\theta_t)))\,. \end{aligned} \tag{5.41}$$

Hence, it is sufficient to obtain a control of the convergence rates of the quantities $\omega_t - \omega_*(\theta_t, \bar{\omega}_t)$ and $\bar{\omega}_t - \bar{\omega}_*(\theta_t)$. We already know that these quantities converge a.s. to zero thanks to Props. 5.1 and 5.2. We conduct a finite-time analysis of each of the terms separately in the subsections below and combine the obtained results to conclude the proof.

We start by introducing a few useful shorthand notations. Let $\tilde{x}_t := (\tilde{S}_t, \tilde{A}_t, S_{t+1})$. Define for every $\tilde{x} = (\tilde{s}, \tilde{a}, s) \in \mathcal{S} \times \mathcal{A} \times \mathcal{S}$ and every $\bar{\omega}, \omega \in \mathbb{R}^m$:

$$\bar{\delta}(\tilde{x}, \bar{\omega}, \omega) = R(\tilde{s}, \tilde{a}) + \gamma \phi(s)^T \bar{\omega} - \phi(\tilde{s})^T \omega\,, \tag{5.42}$$

$$g(\tilde{x}, \bar{\omega}, \omega) = \bar{\delta}(\tilde{x}, \bar{\omega}, \omega)\phi(\tilde{s})\,. \tag{5.43}$$

Finally, define for every $\theta \in \mathbb{R}^d$ the steady-state expectation:

$$\bar{g}(\theta, \bar{\omega}, \omega) = \mathbb{E}_{\tilde{s} \sim d_{\rho,\theta}, \tilde{a} \sim \pi_\theta, s \sim p(\cdot|\tilde{s}, \tilde{a})}[g(\tilde{x}, \bar{\omega}, \omega)]\,. \tag{5.44}$$

5.8.1.1 Control of the first error term $\omega_t - \omega_*(\theta_t, \bar{\omega}_t)$

We introduce an additional shorthand notation for brevity:

$$\nu_t := \omega_t - \omega_*(\theta_t, \bar{\omega}_t)\,.$$

Decomposition of the error. Using the update rule of the critic gives

$$\begin{aligned}\|\nu_{t+1}\|^2 &= \|\omega_t + \beta_t g(\tilde{x}_t, \bar{\omega}_t, \omega_t) - \omega_*(\theta_{t+1}, \bar{\omega}_{t+1})\|^2 \\ &= \|\nu_t + \beta_t g(\tilde{x}_t, \bar{\omega}_t, \omega_t) + \omega_*(\theta_t, \bar{\omega}_t) - \omega_*(\theta_{t+1}, \bar{\omega}_{t+1})\|^2\,. \end{aligned}$$

Then, we develop the squared norm and use the classical inequality $\|a+b\|^2 \leq 2\|a\|+2\|b\|^2$ to obtain

$$\begin{aligned}\|\nu_{t+1}\|^2 \leq \|\nu_t\|^2 + 2\beta_t \langle \nu_t, g(\tilde{x}_t, \bar{\omega}_t, \omega_t)\rangle + 2\langle \nu_t, \omega_*(\theta_t, \bar{\omega}_t) - \omega_*(\theta_{t+1}, \bar{\omega}_{t+1})\rangle \\ + 2\|\omega_*(\theta_t, \bar{\omega}_t) - \omega_*(\theta_{t+1}, \bar{\omega}_{t+1})\|^2 + 2C\beta_t^2\,. \end{aligned} \tag{5.45}$$

Now, we decompose the first inner product into a main term generating a repelling effect and a second Markov noise term as follows

$$\langle \nu_t, g(\tilde{x}_t, \bar{\omega}_t, \omega_t)\rangle = \langle \nu_t, \bar{g}(\theta_t, \bar{\omega}_t, \omega_t)\rangle + \Lambda(\theta_t, \bar{\omega}_t, \omega_t, \tilde{x}_t)\,, \tag{5.46}$$

where we used the shorthand notation

$$\Lambda(\theta, \bar{\omega}, \omega, \tilde{x}) := \langle \omega - \omega_*(\theta, \bar{\omega}), g(\tilde{x}, \bar{\omega}, \omega) - \bar{g}(\theta, \bar{\omega}, \omega)\rangle\,. \tag{5.47}$$

We control the first term in Eq. (5.46) as follows

$$\langle \nu_t, \bar{g}(\theta_t, \bar{\omega}_t, \omega_t)\rangle = \langle \nu_t, \bar{g}(\theta_t, \bar{\omega}_t, \omega_t) - \bar{g}(\theta_t, \bar{\omega}_t, \omega_*(\theta_t, \bar{\omega}_t))\rangle = -\langle \nu_t, \bar{G}(\theta_t)\nu_t\rangle \leq -\kappa\|\nu_t\|^2 . \tag{5.48}$$

We used the fact that $\bar{g}(\theta_t, \bar{\omega}_t, \omega_*(\theta_t, \bar{\omega}_t)) = 0$ for the first equality and Lem. 5.9 for the inequality. Then, it can be shown that

$$\|\omega_*(\theta_t, \bar{\omega}_t) - \omega_*(\theta_{t+1}, \bar{\omega}_{t+1})\| \leq C(\|\theta_t - \theta_{t+1}\| + \|\bar{\omega}_t - \bar{\omega}_{t+1}\|) \leq C(\alpha_t + \xi_t) . \tag{5.49}$$

Combining Eqs. (5.45) to (5.49) leads to

$$\|\nu_{t+1}\|^2 \leq (1-2\kappa\beta_t)\|\nu_t\|^2 + 2\beta_t\Lambda(\theta_t, \bar{\omega}_t, \omega_t, \tilde{x}_t) + C(\alpha_t+\xi_t)\|\nu_t\| + C(\alpha_t^2+\xi_t^2+\beta_t^2) . \tag{5.50}$$

Control of the Markov noise term $\Lambda(\theta_t, \bar{\omega}_t, \omega_t, \tilde{x}_t)$. We decompose the noise term using a similar technique to Zou et al. (2019b) which was then used in Wu et al. (2020); Shen et al. (2020). Let $T > 0$. Define the mixing time

$$\tau_T := \min\{t \in \mathbb{N}, t \geq 1 : c\sigma^{t-1} \leq \min\{\alpha_T, \xi_T, \beta_T\}\} . \tag{5.51}$$

In the remainder of the proof, we will use the notation τ for τ_T (interchangeably). In order to control the difference between the update rule of the critic and its steady-state expectation, we introduce an auxiliary chain which coincides with $\tilde{x}_t$ except for the τ last steps where the policy is fixed to $\pi_{\theta_{t-\tau}}$. The auxiliary chain will be denoted by $\check{x}_t := (\check{S}_t, \check{A}_t, S_{t+1})$ where $S_{t+1} \sim p(\cdot|\check{S}_t, \check{A}_t)$ and $(\check{S}_t, \check{A}_t)$ is generated as follows:

$$\tilde{S}_{t-\tau} \xrightarrow{\theta_{t-\tau}} \tilde{A}_{t-\tau} \xrightarrow{\tilde{p}} \tilde{S}_{t-\tau+1} \xrightarrow{\theta_{t-\tau}} \check{A}_{t-\tau+1} \xrightarrow{\tilde{p}} \check{S}_{t-\tau+2} \xrightarrow{\theta_{t-\tau}} \check{A}_{t-\tau+2} \xrightarrow{\tilde{p}} \cdots \xrightarrow{\tilde{p}} \check{S}_t \xrightarrow{\theta_{t-\tau}} \check{A}_t \xrightarrow{\tilde{p}} \check{S}_{t+1} .$$

Compared to this chain, the original chain has a drifting policy, i.e., at each time step, the actor parameter θ_t is updated and so is the policy π_{θ_t} and we recall that it is given by:

$$\tilde{S}_{t-\tau} \xrightarrow{\theta_{t-\tau}} \tilde{A}_{t-\tau} \xrightarrow{\tilde{p}} \tilde{S}_{t-\tau+1} \xrightarrow{\theta_{t-\tau+1}} \tilde{A}_{t-\tau+1} \xrightarrow{\tilde{p}} \tilde{S}_{t-\tau+2} \xrightarrow{\theta_{t-\tau+2}} \tilde{A}_{t-\tau+2} \xrightarrow{\tilde{p}} \cdots \xrightarrow{\tilde{p}} \tilde{S}_t \xrightarrow{\theta_t} \tilde{A}_t \xrightarrow{\tilde{p}} \tilde{S}_{t+1} .$$

Using the shorthand notation $z_t := (\bar{\omega}_t, \omega_t)$, the Markov noise term can be decomposed as follows:

$$\begin{aligned}\Lambda(\theta_t, \bar{\omega}_t, \omega_t, \tilde{x}_t) = (\Lambda(\theta_t, z_t, \tilde{x}_t) - \Lambda(\theta_{t-\tau}, z_{t-\tau}, \tilde{x}_t)) + (\Lambda(\theta_{t-\tau}, z_{t-\tau}, \tilde{x}_t) - \Lambda(\theta_{t-\tau}, z_{t-\tau}, \check{x}_t)) \\ + \Lambda(\theta_{t-\tau}, z_{t-\tau}, \check{x}_t) .\end{aligned} \tag{5.52}$$

We control each one of the terms successively.

(a) **Control of $\Lambda(\theta_t, z_t, \tilde{x}_t) - \Lambda(\theta_{t-\tau}, z_{t-\tau}, \tilde{x}_t)$:** Using that ω_* and $\bar{g}$ are Lipschitz in all their arguments, g is Lipschitz in its two last arguments and ω_t, ω_*, g and $\bar{g}$ are all bounded, one can show after tedious decompositions that

$$|\Lambda(\theta_t, z_t, \tilde{x}_t) - \Lambda(\theta_{t-\tau}, z_{t-\tau}, \tilde{x}_t)| \leq C(\|\theta_t - \theta_{t-\tau}\| + \|\bar{\omega}_t - \bar{\omega}_{t-\tau}\| + \|\omega_t - \omega_{t-\tau}\|) . \tag{5.53}$$

Then, recalling that the sequence (α_t) is nonincreasing, remark that

$$\|\theta_t - \theta_{t-\tau}\| \leq \sum_{t-\tau}^{t-1} \|\theta_{j+1} - \theta_j\| \leq C \sum_{t-\tau}^{t-1} \alpha_j \leq C\tau\alpha_{t-\tau} .$$

Similarly, we have $\|\bar{\omega}_t - \bar{\omega}_{t-\tau}\| \leq C\tau\xi_{t-\tau}$, $\|\omega_t - \omega_{t-\tau}\| \leq C\tau\beta_{t-\tau}$ and we can therefore deduce from Eq. (5.53) that

$$|\Lambda(\theta_t, z_t, \tilde{x}_t) - \Lambda(\theta_{t-\tau}, z_{t-\tau}, \tilde{x}_t)| \leq C\tau(\alpha_{t-\tau} + \beta_{t-\tau} + \xi_{t-\tau}) . \tag{5.54}$$

(b) **Control of $\Lambda(\theta_{t-\tau}, z_{t-\tau}, \tilde{x}_t) - \Lambda(\theta_{t-\tau}, z_{t-\tau}, \check{x}_t)$:** following similar arguments to Wu et al. (2020); Shen et al. (2020), we upperbound the conditional expectation of this error term w.r.t. $\tilde{S}_{t-\tau+1}, \bar{\omega}_{t-\tau}, \omega_{t-\tau}$ and $\theta_{t-\tau}$. Note that our definition of $\check{x}_t$ is slightly different from the ones used in the two aforementioned references because of the third component of $\check{x}_t$ (and also $\tilde{x}_t$) which is generated according to the original kernel p instead of the artificial kernel $\tilde{p}$. We have

$$\begin{aligned}
&\mathbb{E}[\Lambda(\theta_{t-\tau}, z_{t-\tau}, \tilde{x}_t) - \Lambda(\theta_{t-\tau}, z_{t-\tau}, \check{x}_t)|\tilde{S}_{t-\tau+1}, \theta_{t-\tau}] && (5.55)\\
&= \mathbb{E}[\langle \nu_{t-\tau}, g(\tilde{x}_t, z_{t-\tau}) - g(\check{x}_t, z_{t-\tau})\rangle|\tilde{S}_{t-\tau+1}, \theta_{t-\tau}]\\
&\leq C d_{TV}(\mathbb{P}(\tilde{x}_t \in \cdot|\tilde{S}_{t-\tau+1}, \theta_{t-\tau}), \mathbb{P}(\check{x}_t \in \cdot|\tilde{S}_{t-\tau+1}, \theta_{t-\tau}))\\
&\leq \frac{C}{2}|\mathcal{A}|L_\pi \sum_{i=t-\tau}^{t} \mathbb{E}[\|\theta_i - \theta_{t-\tau}\||\tilde{S}_{t-\tau+1}, \theta_{t-\tau}]\,, && (5.56)
\end{aligned}$$

where the first equality stems from the definition of Λ, the first inequality uses the definition of the total variation distance d_{TV} between two probability measures and the last inequality is a consequence of (Wu et al., 2020, Lem. B.2, p.17) (see also (Shen et al., 2020, Lem. 2 p.12)).

Then, we have

$$\begin{aligned}
\sum_{i=t-\tau}^{t} \mathbb{E}[\|\theta_i - \theta_{t-\tau}\||\tilde{S}_{t-\tau+1}, \theta_{t-\tau}] &\leq \sum_{i=t-\tau}^{t}\sum_{j=t-\tau}^{i-1} \mathbb{E}[\|\theta_{j+1} - \theta_j\||\tilde{S}_{t-\tau+1}, \theta_{t-\tau}]\\
&\leq C\sum_{i=t-\tau}^{t}\sum_{j=t-\tau}^{i-1} \alpha_j \leq C\alpha_{t-\tau}\sum_{i=0}^{\tau} i \leq C\alpha_{t-\tau}(\tau+1)^2\,.
\end{aligned}$$

As a consequence of these derivations, Eq. (5.55) yields

$$\mathbb{E}[\Lambda(\theta_{t-\tau}, z_{t-\tau}, \tilde{x}_t) - \Lambda(\theta_{t-\tau}, z_{t-\tau}, \check{x}_t)|\tilde{S}_{t-\tau+1}, \theta_{t-\tau}] \leq C\alpha_{t-\tau}(\tau+1)^2\,, \qquad (5.57)$$

(c) **Control of $\Lambda(\theta_{t-\tau}, z_{t-\tau}, \check{x}_t)$:** Define $\bar{x}_t := (\bar{S}_t, \bar{A}_t, S_{t+1})$ where $\bar{S}_t \sim d_{\rho,\theta_{t-\tau}}$, $\bar{A}_t \sim \pi_{\theta_{t-\tau}}$ and $S_{t+1} \sim p(\cdot|\bar{S}_t, \bar{A}_t)$. Observing that $\mathbb{E}[\Lambda(\theta_{t-\tau}, z_{t-\tau}, \bar{x}_t)|\tilde{S}_{t-\tau+1}, \theta_{t-\tau}] = 0$, we obtain

$$\begin{aligned}
\mathbb{E}[\Lambda(\theta_{t-\tau}, z_{t-\tau}, \check{x}_t)|\tilde{S}_{t-\tau+1}, \theta_{t-\tau}] &= \mathbb{E}[\Lambda(\theta_{t-\tau}, z_{t-\tau}, \check{x}_t) - \Lambda(\theta_{t-\tau}, z_{t-\tau}, \bar{x}_t)|\tilde{S}_{t-\tau+1}, \theta_{t-\tau}]\\
&= \mathbb{E}[\langle \nu_{t-\tau}, g(\check{x}_t, z_{t-\tau}) - g(\bar{x}_t, z_{t-\tau})\rangle|\tilde{S}_{t-\tau+1}, \theta_{t-\tau}]\\
&\leq C d_{TV}(\mathbb{P}(\check{x}_t \in \cdot|\tilde{S}_{t-\tau+1}, \theta_{t-\tau}), \mathbb{P}(\bar{x}_t \in \cdot|\tilde{S}_{t-\tau+1}, \theta_{t-\tau}))\\
&= C d_{TV}(\mathbb{P}(\tilde{S}_t \in \cdot|\tilde{S}_{t-\tau+1}, \theta_{t-\tau}), d_{\rho,\theta_{t-\tau}})\\
&\leq C\sigma^{\tau-1}\\
&\leq \alpha_T\,, && (5.58)
\end{aligned}$$

where the first inequality stems again from the definition of the total variation norm and the last two ones follow from Assumption 5.6.1 and the definition of the mixing time $\tau = \tau_T$ (see Eq. (5.51)).

Given the decomposition of Eq. (5.52), collecting Eqs.(5.54), (5.57), (5.58) and taking total expectation leads to the conclusion of this subsection

$$\mathbb{E}[\Lambda(\theta_t, z_t, \tilde{x}_t)] \leq C(\tau(\alpha_{t-\tau} + \beta_{t-\tau} + \xi_{t-\tau}) + \alpha_{t-\tau}(\tau+1)^2 + \alpha_T)\,. \qquad (5.59)$$

Derivation of the convergence rate of the mean error term $\frac{1}{T}\sum_{t=1}^{T}\|\nu_t\|^2$. We obtain from taking the total expectation in Eq. (5.50) together with Eq. (5.59) that

$$\mathbb{E}[\|\nu_{t+1}\|^2] \leq (1-2\kappa\beta_t)\mathbb{E}[\|\nu_t\|^2] + 2C\beta_t(\tau(\alpha_{t-\tau}+\beta_{t-\tau}+\xi_{t-\tau})+\alpha_{t-\tau}(\tau+1)^2+\alpha_T) + C(\alpha_t+\xi_t)\mathbb{E}[\|\nu_t\|] + C(\alpha_t^2+\xi_t^2+\beta_t^2)\,. \quad (5.60)$$

Rearranging the inequality and summing for t between τ_T and T, we get

$$2\kappa\sum_{t=\tau_T}^{T}\mathbb{E}[\|\nu_t\|^2] \leq I_1(T)+I_2(T)+I_3(T)+I_4(T)\,, \quad (5.61)$$

where

$$I_1(T) := \sum_{t=\tau_T}^{T}\frac{1}{\beta_t}(\mathbb{E}[\|\nu_t\|^2]-\mathbb{E}[\|\nu_{t+1}\|^2])\,, \quad (5.62)$$

$$I_2(T) := \sum_{t=\tau_T}^{T} 2C(\tau(\alpha_{t-\tau}+\beta_{t-\tau}+\xi_{t-\tau})+\alpha_{t-\tau}(\tau+1)^2+\alpha_T) \quad (5.63)$$

$$I_3(T) := C\sum_{t=\tau_T}^{T}\left(\frac{\alpha_t}{\beta_t}+\frac{\xi_t}{\beta_t}\right)\mathbb{E}[\|\nu_t\|] \quad (5.64)$$

$$I_4(T) := C\sum_{t=\tau_T}^{T}\frac{\alpha_t^2}{\beta_t}+\frac{\xi_t^2}{\beta_t}+\beta_t\,. \quad (5.65)$$

We derive estimates of each one of the terms $I_i(T)$ for $i=1,2,3,4$.

(1) Since (ν_t) is a bounded sequence,

$$I_1(T) = \sum_{t=\tau_T}^{T}\left(\frac{1}{\beta_t}-\frac{1}{\beta_{t-1}}\right)\mathbb{E}[\|\nu_t\|^2]+\frac{1}{\beta_{\tau_T-1}}\mathbb{E}[\|\nu_{\tau_T}\|^2]-\frac{1}{\beta_{\tau_T}}\mathbb{E}[\|\nu_{T+1}\|^2] \leq C\left[\sum_{t=\tau_T}^{T}\left(\frac{1}{\beta_t}-\frac{1}{\beta_{t-1}}\right)+\frac{1}{\beta_{\tau_T-1}}\right]=\frac{C}{\beta_T}=\mathcal{O}(T^{\beta})\,. \quad (5.66)$$

Then, since $\tau_T=\mathcal{O}(\log T)$, it follows that

$$\frac{1}{1+T-\tau_T}I_1(T) \leq \frac{1}{1+T-\tau_T}\frac{C}{\beta_T} = \frac{1}{T(\frac{1}{T}+1-\frac{\tau_T}{T})}\frac{C}{\beta_T}=\mathcal{O}(T^{\beta-1})\,.$$

(2) Using the inequality $\sum_{k=a}^{b}k^{-\beta}\leq\frac{b^{1-\beta}}{1-\beta}$ for $1\leq a<b$ and the fact that $\tau_T=\mathcal{O}(\log T)$, we have

$$I_2(T) \leq C\left(\tau_T\sum_{t=0}^{T-\tau}(\alpha_t+\beta_t+\xi_t)+(\tau+1)^2\sum_{t=0}^{T-\tau}\alpha_t+(1+T-\tau)\alpha_T\right) \quad (5.67)$$

$$\leq C(\tau(1+T)^{1-\beta}+(\tau+1)^2(1+T)^{1-\alpha}) \quad (5.68)$$

$$= \mathcal{O}((\log T)T^{1-\beta})+\mathcal{O}(\log^2(T)T^{1-\alpha}) = \mathcal{O}((\log T)T^{1-\beta})\,, \quad (5.69)$$

where we recall for the second inequality that $0<\beta<\xi<\alpha<1$. As a consequence,

$$\frac{1}{1+T-\tau_T}I_2(T)=\mathcal{O}((\log T)T^{-\beta})\,.$$

(3) Using the Cauchy-Schwartz inequality, we can write:

$$I_3(T) = \sum_{t=\tau_T}^{T} C\left(\frac{\alpha_t}{\beta_t} + \frac{\xi_t}{\beta_t}\right)\mathbb{E}[\|\nu_t\|] \leq C^2\sqrt{\sum_{t=\tau_T}^{T}\left(\frac{\alpha_t}{\beta_t} + \frac{\xi_t}{\beta_t}\right)^2}\sqrt{\sum_{t=\tau_T}^{T}\mathbb{E}[\|\nu_t\|^2]}\,. \tag{5.70}$$

Then, observing that the sequences $(\frac{\alpha_t}{\beta_t})$ and $(\frac{\xi_t}{\beta_t})$ are nonincreasing, we have:

$$\begin{aligned}
\frac{1}{1+T-\tau_T}\sum_{t=\tau_T}^{T}\left(\frac{\alpha_t}{\beta_t} + \frac{\xi_t}{\beta_t}\right)^2 &\leq \frac{2}{1+T-\tau_T}\sum_{t=\tau_T}^{T}\left(\left(\frac{\alpha_t}{\beta_t}\right)^2 + \left((\frac{\xi_t}{\beta_t}\right)^2\right) \\
&= \frac{2}{1+T-\tau_T}\sum_{t=0}^{T-\tau_T}\left(\left(\frac{\alpha_{t+\tau_T}}{\beta_{t+\tau_T}}\right)^2 + \left(\frac{\xi_{t+\tau_T}}{\beta_{t+\tau_T}}\right)^2\right) \\
&\leq \frac{2}{T-\tau_T+1}\sum_{t=0}^{T-\tau_T}\left(\left(\frac{\alpha_t}{\beta_t}\right)^2 + \left(\frac{\xi_t}{\beta_t}\right)^2\right) \\
&\leq \frac{(T-\tau_T+1)^{-2(\alpha-\beta)}}{1-2(\alpha-\beta)} + \frac{(T-\tau_T+1)^{-2(\xi-\beta)}}{1-2(\xi-\beta)} \\
&= \mathcal{O}(T^{-2(\alpha-\beta)} + T^{-2(\xi-\beta)})\,.
\end{aligned} \tag{5.71}$$

(4) Similarly to item (3), to control the fourth term, we write:

$$\begin{aligned}
\frac{1}{1+T-\tau_T}\sum_{t=\tau_T}^{T}\left(\frac{\alpha_t^2}{\beta_t} + \frac{\xi_t^2}{\beta_t} + \beta_t\right) &\leq \frac{1}{1+T-\tau_T}\sum_{t=0}^{T-\tau_T}\left(\frac{\alpha_t^2}{\beta_t} + \frac{\xi_t^2}{\beta_t} + \beta_t\right) \\
&\leq \frac{(1+T-\tau_T)^{-(2\alpha-\beta)}}{1-(2\alpha-\beta)} + \frac{(1+T-\tau_T)^{-(2\xi-\beta)}}{1-(2\xi-\beta)} \\
&\quad + \frac{(1+T-\tau_T)^{-\beta}}{1-\beta} \\
&= \mathcal{O}(T^{-(2\alpha-\beta)} + T^{-(2\xi-\beta)} + T^{-\beta})\,.
\end{aligned} \tag{5.72}$$

Hence,

$$\frac{1}{1+T-\tau_T}I_4(T) = \mathcal{O}(T^{-(2\alpha-\beta)} + T^{-(2\xi-\beta)} + T^{-\beta})\,. \tag{5.73}$$

Define:

$$N(T) := \frac{1}{1+T-\tau_T}\sum_{t=\tau_T}^{T}\mathbb{E}[\|\nu_t\|^2]\,, \tag{5.74}$$

$$F(T) := \frac{1}{1+T-\tau_T}\sum_{t=\tau_T}^{T}\left(\left(\frac{\alpha_t}{\beta_t}\right)^2 + \left(\frac{\xi_t}{\beta_t}\right)^2\right)\,, \tag{5.75}$$

$$G(T) := \frac{1}{1+T-\tau_T}(I_1(T) + I_2(T) + I_4(T))\,. \tag{5.76}$$

Using items (1) to (4), we have:

$$F(T) = \mathcal{O}(T^{-2(\alpha-\beta)} + T^{-2(\xi-\beta)}) \,, \tag{5.77}$$

$$G(T) = \mathcal{O}(T^{\beta-1}) + \mathcal{O}((\log T)T^{-\beta}) + \mathcal{O}(T^{-(2\alpha-\beta)} + T^{-(2\xi-\beta)} + T^{-\beta}) \,. \tag{5.78}$$

From Eq. (5.61) and items (1) to (4) above, we have:

$$2\kappa N(T) \leq C\sqrt{F(T)}\sqrt{N(T)} + G(T) \,.$$

Solving this inequality yields:

$$N(T) = \mathcal{O}(F(T) + G(T)) \,.$$

Remarking that $0 < 2(\alpha - \beta) < 2\alpha - \beta$ and $0 < 2(\xi - \beta) < 2\xi - \beta$, we obtain:

$$N(T) = \mathcal{O}(T^{\beta-1}) + \mathcal{O}((\log T)T^{-\beta}) + \mathcal{O}(T^{-2(\alpha-\beta)}) + \mathcal{O}(T^{-2(\xi-\beta)}) \,.$$

Then, we conclude that:

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\|\nu_t\|^2] = \mathcal{O}((\log T)T^{-1}) + \mathcal{O}(N(T)) = \mathcal{O}(N(T)) \,.$$

5.8.1.2 Control of the second error term $\bar{\omega}_t - \bar{\omega}_*(\theta_t)$

Consider the shorthand notation $\bar{\nu}_t := \bar{\omega}_t - \bar{\omega}_*(\theta_t)$.

Using the update rules of $(\bar{\omega}_t)$, (ω_t) and developing the squared norm gives:

$$\begin{aligned}
\|\bar{\nu}_t\|^2 &= \|\bar{\omega}_t + \xi_t(\omega_{t+1} - \bar{\omega}_t) - \bar{\omega}_*(\theta_{t+1})\|^2 \\
&= \|\bar{\nu}_t + \xi_t(\omega_t + \beta_t g(\tilde{x}_t, \bar{\omega}_t, \omega_t) - \bar{\omega}_t) + \bar{\omega}_*(\theta_t) - \bar{\omega}_*(\theta_{t+1})\|^2 \\
&= \|\bar{\nu}_t + \Big(\xi_t(\nu_t + \beta_t g(\tilde{x}_t, \bar{\omega}_t, \omega_t) + \omega_*(\theta_t, \bar{\omega}_t) - \bar{\omega}_t) + \bar{\omega}_*(\theta_t) - \bar{\omega}_*(\theta_{t+1})\Big) \|^2 \\
&= \|\bar{\nu}_t\|^2 + 2\langle \bar{\nu}_t, \xi_t(\nu_t + \beta_t g(\tilde{x}_t, \bar{\omega}_t, \omega_t) + \omega_*(\theta_t, \bar{\omega}_t) - \bar{\omega}_t) + \bar{\omega}_*(\theta_t) - \bar{\omega}_*(\theta_{t+1})\rangle \\
&+ \|\xi_t(\nu_t + \beta_t g(\tilde{x}_t, \bar{\omega}_t, \omega_t) + \omega_*(\theta_t, \bar{\omega}_t) - \bar{\omega}_t) + \bar{\omega}_*(\theta_t) - \bar{\omega}_*(\theta_{t+1})\|^2 \,.
\end{aligned} \tag{5.79}$$

Since $(\nu_t), (\bar{\omega}_t), g, \omega_*$ are bounded and the function $\bar{\omega}_*$ is Lipschitz continuous, the last squared norm term can be bounded by: $C(\xi_t^2\beta_t^2 + \xi_t^2 + \alpha_t^2)$.

We now control the scalar product in Eq. (5.79). We decompose this term into four different terms:

(a) Using Assumption 5.6.2, it holds that:

$$2\xi_t\langle \bar{\nu}_t, \omega_*(\theta_t, \bar{\omega}_t) - \bar{\omega}_t\rangle = -2\xi_t\langle \bar{\nu}_t, \bar{G}(\theta_t)^{-1}G(\theta_t)\bar{\nu}_t\rangle \leq -2\zeta\xi_t\|\bar{\nu}_t\|^2 \,.$$

(b) The boundedness of the function g implies that:

$$2\xi_t\beta_t\langle \bar{\nu}_t, g(\tilde{x}_t, \bar{\omega}_t, \omega_t)\rangle \leq 2\xi_t\beta_t c\|\bar{\nu}_t\| \,.$$

(c) Applying the Cauchy-Schwarz inequality gives:

$$2\xi_t\langle \bar{\nu}_t, \nu_t\rangle \leq 2\xi_t\|\bar{\nu}_t\| \cdot \|\nu_t\| \,.$$

(d) Since $\bar{\omega}_*$ is Lipschitz continuous, we can write:

$$2\langle \bar{\nu}_t, \bar{\omega}_*(\theta_t) - \bar{\omega}_*(\theta_{t+1})\rangle \leq C\alpha_t \|\bar{\nu}_t\| \,.$$

Collecting the bounds from items (a) to (d) and incorporating them into Eq. (5.79), we obtain:

$$\|\bar{\nu}_{t+1}\|^2 \leq (1-2\zeta\xi_t)\|\bar{\nu}_t\|^2 + C(\xi_t\beta_t+\alpha_t)\|\bar{\nu}_t\| + 2\xi_t\|\bar{\nu}_t\|\cdot\|\nu_t\| + C(\xi_t^2\beta_t^2+\xi_t^2+\alpha_t^2)\,. \tag{5.80}$$

Rearranging Ineq. (5.80) leads to:

$$2\zeta\|\bar{\nu}_t\|^2 \leq \frac{1}{\xi_t}(\|\bar{\nu}_t\|^2 - \|\bar{\nu}_{t+1}\|^2) + C\left(\beta_t + \frac{\alpha_t}{\xi_t}\right)\|\bar{\nu}_t\| + 2\|\bar{\nu}_t\|\cdot\|\nu_t\| + C\left(\xi_t\beta_t^2 + \xi_t + \frac{\alpha_t^2}{\xi_t}\right). \tag{5.81}$$

Summing this inequality for t between 1 and T and taking total expectation yield:

$$\frac{2\zeta}{T}\sum_{t=1}^{T}\mathbb{E}[\|\bar{\nu}_t\|^2] \leq \Sigma_1(T) + \Sigma_2(T) + \Sigma_3(T) + \Sigma_4(T)\,, \tag{5.82}$$

where

$$\Sigma_1(T) := \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\xi_t}(\mathbb{E}[\|\bar{\nu}_t\|^2] - \mathbb{E}[\|\bar{\nu}_{t+1}\|^2])\,, \tag{5.83}$$

$$\Sigma_2(T) := \frac{C}{T}\sum_{t=1}^{T}\left(\beta_t + \frac{\alpha_t}{\xi_t}\right)\mathbb{E}[\|\bar{\nu}_t\|]\,, \tag{5.84}$$

$$\Sigma_3(T) := \frac{2}{T}\sum_{t=1}^{T}\mathbb{E}[\|\bar{\nu}_t\|\cdot\|\nu_t\|]\,, \tag{5.85}$$

$$\Sigma_4(T) := \frac{C}{T}\sum_{t=1}^{T}\left(\xi_t\beta_t^2 + \xi_t + \frac{\alpha_t^2}{\xi_t}\right). \tag{5.86}$$

Similarly to Section 5.8.1.1, we control each one of the terms $\Sigma_i, i = 1,2,3,4$ successively.

(i) First, using the boundedness of $(\bar{\nu}_t)$, we estimate Σ_1 as follows:

$$\Sigma_1(T) = \frac{1}{T}\left[\sum_{t=1}^{T}\left(\frac{1}{\xi_t} - \frac{1}{\xi_{t-1}}\right)\mathbb{E}[\|\bar{\nu}_t\|^2] + \frac{1}{\xi_0}\mathbb{E}[\|\bar{\nu}_1\|^2] - \frac{1}{\xi_T}\mathbb{E}[\|\bar{\nu}_{T+1}\|^2]\right] \leq \frac{C}{T\xi_T} = \mathcal{O}(T^{\xi-1})\,.$$

(ii) Cauchy-Schwarz inequality implies:

$$\Sigma_2(T) \leq \frac{C}{T}\sqrt{\sum_{t=1}^{T}\left(\beta_t + \frac{\alpha_t}{\xi_t}\right)^2}\sqrt{\sum_{t=1}^{T}\mathbb{E}[\|\bar{\nu}_t\|^2]} \leq C\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\beta_t^2 + \left(\frac{\alpha_t}{\xi_t}\right)^2\right)}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\bar{\nu}_t\|^2]}\,.$$

Moreover,

$$\frac{1}{T}\sum_{t=1}^{T}\left(\beta_t^2 + \left(\frac{\alpha_t}{\xi_t}\right)^2\right) \leq \frac{1}{T}\left(\frac{(T+1)^{1-2\beta}}{1-2\beta} + \frac{(T+1)^{1-2(\alpha-\xi)}}{1-2(\alpha-\xi)}\right) = \mathcal{O}(T^{-2\beta}) + \mathcal{O}(T^{-2(\alpha-\xi)})\,.$$

(iii) Invoking the Cauchy-Schwarz inequality again yields:

$$\Sigma_3(T) \leq 2\sqrt{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\bar{\nu}_t\|^2]}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\nu_t\|^2]}\,.$$

(iv) Similarly to item (ii), we obtain

$$\Sigma_4(T) = \mathcal{O}(T^{-\xi-2\beta}) + \mathcal{O}(T^{-\xi}) + \mathcal{O}(T^{\xi-2\alpha})\,.$$

Define for every $T > 0$ the following quantities:

$$W(T) := \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\nu_t\|^2]\,, \tag{5.87}$$

$$X(T) := \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\bar{\nu}_t\|^2]\,, \tag{5.88}$$

$$Y(T) := \frac{1}{T}\sum_{t=1}^{T}\left(\beta_t^2 + \left(\frac{\alpha_t}{\xi_t}\right)^2\right)\,, \tag{5.89}$$

$$Z(T) := \Sigma_1(T) + \Sigma_4(T)\,. \tag{5.90}$$

It follows from items (i) to (iv) and Section 5.8.1.1 (for the last estimate) that

$$Y(T) = \mathcal{O}(T^{-2\beta}) + \mathcal{O}(T^{-2(\alpha-\xi)})\,, \tag{5.91}$$

$$Z(T) = \mathcal{O}(T^{\xi-1}) + \mathcal{O}(T^{-\xi-2\beta}) + \mathcal{O}(T^{-\xi}) + \mathcal{O}(T^{\xi-2\alpha})\,, \tag{5.92}$$

$$W(T) = \mathcal{O}(T^{\beta-1}) + \mathcal{O}((\log T)T^{-\beta}) + \mathcal{O}(T^{2(\beta-\alpha)}) + \mathcal{O}(T^{2(\beta-\xi)})\,. \tag{5.93}$$

Eq. (5.82) can be written:

$$2\zeta X(T) \leq C\left(\sqrt{Y(T)} + \sqrt{W(T)}\right)\sqrt{X(T)} + Z(T)\,.$$

Solving this inequality implies:

$$X(T) = \mathcal{O}(Y(T) + W(T) + Z(T))\,. \tag{5.94}$$

Since $0 < \beta < \xi < \alpha < 1$, we obtain:

$$X(T) = \mathcal{O}(T^{\xi-1}) + \mathcal{O}((\log T)T^{-\beta}) + \mathcal{O}(T^{-2(\alpha-\xi)}) + \mathcal{O}(T^{-2(\xi-\beta)})\,. \tag{5.95}$$

5.8.1.3 End of Proof of Th. 5.5

We conclude our finite-time analysis of the critic by combining both previous sections (5.8.1.1 and 5.8.1.2):

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\omega_t-\bar{\omega}_*(\theta_t)\|^2] &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\nu_t+\omega_*(\theta_t,\bar{\omega}_t)-\bar{\omega}_*(\theta_t)\|^2]\\ &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\nu_t+\omega_*(\theta_t,\bar{\omega}_t)-\omega_*(\theta_t,\bar{\omega}_*(\theta_t))\|^2]\\ &\leq 2W(T)+CX(T)\\ &= \mathcal{O}(X(T))\\ &= \mathcal{O}(T^{\xi-1})+\mathcal{O}((\log T)T^{-\beta})+\mathcal{O}(T^{-2(\alpha-\xi)})+\mathcal{O}(T^{-2(\xi-\beta)})\,,\end{aligned} \tag{5.96}$$

where the second equality follows from using the identity $w_*(\theta,\bar{\omega}_*(\theta))=\bar{\omega}_*(\theta)$ for every $\theta\in\mathbb{R}^d$, the inequality stems from using the classical inequality $\|a+b\|^2\leq 2(\|a\|^2+\|b\|^2)$ together with the fact that ω_* is Lipschitz continuous, the penultimate equality is a consequence of Eq. (5.94) and the last equality is the result of the previous section (see Eq. (5.95)).

5.8.2 Proof of Th. 5.6

Recall the notation $\tilde{x}_t:=(\tilde{S}_t,\tilde{A}_t,S_{t+1})$. In this section, we overload this notation with the reward sequence (R_t), i.e., $\tilde{x}_t:=(\tilde{S}_t,\tilde{A}_t,S_{t+1},R_{t+1})$. Let us fix some additional convenient notations. Define for every $\tilde{x}=(\tilde{s},\tilde{a},s,r)\in\mathcal{S}\times\mathcal{A}\times\mathcal{S}\times[-U_R,U_R]$, and every $\omega\in\mathbb{R}^m,\theta\in\mathbb{R}^d$:

$$\hat{\delta}(\tilde{x},\omega):=r+\gamma\phi(s)^T\omega-\phi(\tilde{s})^T\omega \tag{5.97}$$

$$\delta(\tilde{x},\theta)=r+\gamma V_{\pi_\theta}(s)-V_{\pi_\theta}(\tilde{s})\,. \tag{5.98}$$

Note that the TD error δ_{t+1} used in Algorithm 5.1 coincides with $\hat{\delta}(\tilde{x}_t,\omega_t)$.

Since the function ∇J is L_J-Lipschitz continuous, a classical Taylor inequality combined with the update rule of (θ_t) yields:

$$J(\theta_{t+1})\geq J(\theta_t)+\alpha_t\langle\nabla J(\theta_t),\hat{\delta}(\tilde{x}_t,\omega_t)\psi_{\theta_t}(\tilde{S}_t,\tilde{A}_t)\rangle-\frac{L_J}{2}\alpha_t^2\|\hat{\delta}(\tilde{x}_t,\omega_t)\psi_{\theta_t}(\tilde{S}_t,\tilde{A}_t)\|^2\,. \tag{5.99}$$

Recalling that $\theta\mapsto\psi_\theta(s,a)$ is bounded by Assumption 5.3.1-*(c)*, (R_t) and (ω_t) are bounded (see Assumption 5.5.3) and $\mathcal{S},\mathcal{A}$ are finite, we obtain from Eq. (5.99) that there exists a constant C s.t.:

$$J(\theta_{t+1})\geq J(\theta_t)+\alpha_t\langle\nabla J(\theta_t),\hat{\delta}(\tilde{x}_t,\omega_t)\psi_{\theta_t}(\tilde{S}_t,\tilde{A}_t)\rangle-CL_J\alpha_t^2\,. \tag{5.100}$$

Now, we decompose the TD error by introducing both the moving target $\bar{\omega}_*(\theta_t)$ and the TD error $\delta(\tilde{x}_t,\theta_t)$ associated to the true value function $V_{\pi_{\theta_t}}$:

$$\hat{\delta}(\tilde{x}_t,\omega_t)=[\hat{\delta}(\tilde{x}_t,\omega_t)-\hat{\delta}(\tilde{x}_t,\bar{\omega}_*(\theta_t))]+[\hat{\delta}(\tilde{x}_t,\bar{\omega}_*(\theta_t))-\delta(\tilde{x}_t,\theta_t)]+\delta(\tilde{x}_t,\theta_t)\,. \tag{5.101}$$

Incorporating this decomposition (5.101) into Eq. (5.100) gives:

$$\begin{aligned} J(\theta_{t+1}) \geq J(\theta_t) &+ \alpha_t \langle \nabla J(\theta_t), (\hat{\delta}(\tilde{x}_t, \omega_t) - \hat{\delta}(\tilde{x}_t, \bar{\omega}_*(\theta_t)))\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t)\rangle \\ &+ \alpha_t \langle \nabla J(\theta_t), (\hat{\delta}(\tilde{x}_t, \bar{\omega}_*(\theta_t)) - \delta(\tilde{x}_t, \theta_t))\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t)\rangle \\ &+ \alpha_t \langle \nabla J(\theta_t), \delta(\tilde{x}_t, \theta_t)\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t) - \nabla J(\theta_t)\rangle + \alpha_t \|\nabla J(\theta_t)\|^2 - C L_J \alpha_t^2 \,. \end{aligned} \tag{5.102}$$

In Eq. (5.102), the first inner product corresponds to the bias introduced by the critic. The second one represents the linear FA error and the third translates the Markovian noise. Our task now is to control each one of these error terms in Eq. (5.102).

For the first term, observing that $\hat{\delta}(\tilde{x}_t, \omega_t) - \hat{\delta}(\tilde{x}_t, \bar{\omega}_*(\theta_t)) = (\gamma\phi(S_{t+1}) - \phi(\tilde{S}_t))^T(\omega_t - \bar{\omega}_*(\theta_t))$, the Cauchy-Schwarz inequality leads to:

$$\mathbb{E}[\langle \nabla J(\theta_t), \hat{\delta}(\tilde{x}_t, \omega_t) - \hat{\delta}(\tilde{x}_t, \bar{\omega}_*(\theta_t))\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t)\rangle] \leq -C\sqrt{\mathbb{E}[\|\nabla J(\theta_t)\|^2]}\sqrt{\mathbb{E}[\|\omega_t - \bar{\omega}_*(\theta_t)\|^2]} \,. \tag{5.103}$$

Then, we control each one of the second and third terms in Eq. (5.102) in the following sections successively.

5.8.2.1 Control of the Markovian bias term

We introduce a specific convenient notation for the second term:

$$\Gamma(\tilde{x}_t, \theta_t) := \langle \nabla J(\theta_t), \delta(\tilde{x}_t, \theta_t)\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t) - \nabla J(\theta_t)\rangle \,.$$

Recall from Section 5.8.1.1 the auxiliary Markov chain $(\check{x}_t)$, the Markov chain $(\bar{x}_t)$ induced by the stationary distribution and the mixing time τ defined in Eq. (5.51).

Similarly to Section 5.8.1.1, we introduce the following decomposition:

$$\begin{aligned} \mathbb{E}[\Gamma(\tilde{x}_t, \theta_t)] = \mathbb{E}[\Gamma(\tilde{x}_t, \theta_t) - \Gamma(\tilde{x}_t, \theta_{t-\tau})] &+ \mathbb{E}[\Gamma(\tilde{x}_t, \theta_{t-\tau}) - \Gamma(\check{x}_t, \theta_{t-\tau})] \\ &+ \mathbb{E}[\Gamma(\check{x}_t, \theta_{t-\tau}) - \Gamma(\bar{x}_t, \theta_{t-\tau})] + \mathbb{E}[\Gamma(\bar{x}_t, \theta_{t-\tau})] \,. \end{aligned} \tag{5.104}$$

We address each term of this decomposition successively.

(a) For this first term, we write:

$$\begin{aligned} \Gamma(\tilde{x}_t, \theta_t) - \Gamma(\tilde{x}_t, \theta_{t-\tau}) = &\langle \nabla J(\theta_t) - \nabla J(\theta_{t-\tau}), \delta(\tilde{x}_t, \theta_t)\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t) - \nabla J(\theta_t)\rangle \\ &+ \langle \nabla J(\theta_{t-\tau}), (\delta(\tilde{x}_t, \theta_t) - \delta(\tilde{x}_t, \theta_{t-\tau}))\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t)\rangle \\ &+ \langle \nabla J(\theta_{t-\tau}), \delta(\tilde{x}_t, \theta_{t-\tau})(\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t) - \psi_{\theta_{t-\tau}}(\tilde{S}_t, \tilde{A}_t))\rangle \\ &+ \langle \nabla J(\theta_{t-\tau}), \nabla J(\theta_{t-\tau}) - \nabla J(\theta_t)\rangle \,. \end{aligned}$$

Moreover, note that:

$$\delta(\tilde{x}_t, \theta_t) - \delta(\tilde{x}_t, \theta_{t-\tau}) = \gamma(V_{\pi_{\theta_t}}(S_{t+1}) - V_{\pi_{\theta_{t-\tau}}}(S_{t+1})) + V_{\pi_{\theta_{t-\tau}}}(\tilde{S}_t) - V_{\pi_{\theta_t}}(\tilde{S}_t) \,.$$

Remark that $\nabla J, \theta \mapsto \psi_\theta$ and $\theta \mapsto V_{\pi_\theta}$ are bounded functions under Assumption 5.3.1 . Since $\nabla J, V_{\pi_\theta}, \psi_\theta$ are in addition Lipschitz continuous as functions of θ (see, for e.g., (Shen et al., 2020, Lem. 3) for a proof for V_{π_θ}) under Assumption 5.3.1 , one can show after tedious inequalities that:

$$|\Gamma(\tilde{x}_t, \theta_t) - \Gamma(\tilde{x}_t, \theta_{t-\tau})| \leq C\|\theta_t - \theta_{t-\tau}\| \,. \tag{5.105}$$

(b) For the second term, we have:

$$
\begin{aligned}
&|\mathbb{E}[\Gamma(\tilde{x}_t, \theta_{t-\tau}) - \Gamma(\check{x}_t, \theta_{t-\tau})]| \\
&= |\mathbb{E}[\langle \nabla J(\theta_{t-\tau}), \delta(\tilde{x}_t, \theta_{t-\tau})\psi_{\theta_{t-\tau}}(\tilde{S}_t, \tilde{A}_t) - \delta(\check{x}_t, \theta_{t-\tau})\psi_{\theta_{t-\tau}}(\check{S}_t, \check{A}_t)\rangle]| \\
&= |\mathbb{E}[\langle \nabla J(\theta_{t-\tau}), \delta(\tilde{x}_t, \theta_{t-\tau})\psi_{\theta_{t-\tau}}(\tilde{S}_t, \tilde{A}_t) - \delta(\check{x}_t, \theta_{t-\tau})\psi_{\theta_{t-\tau}}(\check{S}_t, \check{A}_t)\rangle | \tilde{S}_{t-\tau+1}, \theta_{t-\tau}]| \\
&\leq C\mathbb{E}[d_{TV}(\mathbb{P}(\tilde{x}_t \in \cdot | \tilde{S}_{t-\tau+1}, \theta_{t-\tau}), \mathbb{P}(\check{x}_t \in \cdot | \tilde{S}_{t-\tau+1}, \theta_{t-\tau}))] \\
&\leq C \sum_{i=t-\tau}^{t} \mathbb{E}[\|\theta_i - \theta_{t-\tau}\|] \,. \qquad (5.106)
\end{aligned}
$$

Here, the first inequality is a consequence of the definition of the total variation distance whereas the second inequality follows from applying (Wu et al., 2020, Lem. B.2). Indeed, using this last lemma, to show the last inequality, it is sufficient to write:

$$
\begin{aligned}
&d_{TV}(\mathbb{P}(\tilde{x}_t \in \cdot | \tilde{S}_{t-\tau+1}, \theta_{t-\tau}), \mathbb{P}(\check{x}_t \in \cdot | \tilde{S}_{t-\tau+1}, \theta_{t-\tau})) \\
&= d_{TV}(\mathbb{P}((\tilde{S}_t, \tilde{A}_t) \in \cdot | \tilde{S}_{t-\tau+1}, \theta_{t-\tau}), \mathbb{P}((\check{S}_t, \check{A}_t) \in \cdot | \tilde{S}_{t-\tau+1}, \theta_{t-\tau})) \\
&\leq d_{TV}(\mathbb{P}(\tilde{S}_t \in \cdot | \tilde{S}_{t-\tau+1}, \theta_{t-\tau}), \mathbb{P}(\check{S}_t \in \cdot | \tilde{S}_{t-\tau+1}, \theta_{t-\tau})) + \frac{1}{2}|\mathcal{A}| L_\pi \mathbb{E}[\|\theta_t - \theta_{t-\tau}\|] \,.
\end{aligned}
$$

Iterating this inequality gives the desired result of Eq. (5.106) . We conclude from this item that:

$$
\mathbb{E}[\Gamma(\tilde{x}_t, \theta_{t-\tau}) - \Gamma(\check{x}_t, \theta_{t-\tau})] \geq -C \sum_{i=t-\tau}^{t} \mathbb{E}[\|\theta_i - \theta_{t-\tau}\|] \,.
$$

(c) Regarding the third term, similarly to item (b), we can write:

$$
\begin{aligned}
\mathbb{E}[\Gamma(\check{x}_t, \theta_{t-\tau}) - \Gamma(\bar{x}_t, \theta_{t-\tau})] &\geq -C\mathbb{E}[d_{TV}(\mathbb{P}(\check{x}_t \in \cdot | \tilde{S}_{t-\tau+1}, \theta_{t-\tau}), \mathbb{P}(\bar{x}_t \in \cdot | \tilde{S}_{t-\tau+1}, \theta_{t-\tau}))] \\
&= -C\mathbb{E}[d_{TV}(\mathbb{P}(\check{x}_t \in \cdot | \tilde{S}_{t-\tau+1}, \theta_{t-\tau}), d_{\rho,\theta_{t-\tau}} \otimes \pi_{\theta_{t-\tau}} \otimes p)] \\
&= -C\mathbb{E}[d_{TV}(\mathbb{P}(\check{S}_t \in \cdot | \tilde{S}_{t-\tau+1}, \theta_{t-\tau}), d_{\rho,\theta_{t-\tau}})] \\
&\geq -C\sigma^{\tau-1} \,, \qquad (5.107)
\end{aligned}
$$

where the equalities follow from the definitions of $\check{x}_t, \bar{x}_t$ and the last inequality stems from Assumption 5.6.1 .

(d) Since the Markov chain $\bar{x}_t$ is built s.t. $\bar{S}_t \sim d_{\rho,\theta_{t-\tau}}, \bar{A}_t \sim \pi_{\theta_{t-\tau}}, S_{t+1} \sim p(\cdot|\bar{S}_t, \bar{A}_t)$, one can see that $\mathbb{E}[\Gamma(\bar{x}_t, \theta_{t-\tau})] = 0$.

We conlude this section from Eq. (5.104) by collecting Eqs. (5.105) to (5.107) (items (a) to (d)) to obtain:

$$\begin{aligned}
\mathbb{E}[\Gamma(\tilde{x}_t, \theta_t)] &\geq -C\mathbb{E}[\|\theta_t - \theta_{t-\tau}\|] - C \sum_{i=t-\tau+1}^{t} \mathbb{E}[\|\theta_i - \theta_{t-\tau}\|] - C\sigma^{\tau-1} \\
&\geq -C \sum_{i=t-\tau+1}^{t} \mathbb{E}[\|\theta_i - \theta_{i-1}\|] - C \sum_{i=t-\tau+1}^{t} \sum_{j=t-\tau+1}^{i} \mathbb{E}[\|\theta_j - \theta_{j-1}\|] - C\sigma^{\tau-1} \\
&\geq -C \sum_{i=t-\tau+1}^{t} \mathbb{E}[\|\theta_i - \theta_{i-1}\|] - C \sum_{i=t-\tau+1}^{t} \sum_{j=t-\tau+1}^{t} \mathbb{E}[\|\theta_j - \theta_{j-1}\|] - C\sigma^{\tau-1} \\
&\geq -C(\tau+1) \sum_{i=t-\tau+1}^{t} \mathbb{E}[\|\theta_i - \theta_{i-1}\|] - C\sigma^{\tau-1} \\
&\geq -C(\tau+1)^2 \alpha_{t-\tau} - C\alpha_T \,, \qquad (5.108)
\end{aligned}$$

where the last inequality uses the definition of the mixing time τ and the fact that the sequence (α_t) is nonincreasing.

5.8.2.2 Control of the linear FA error term

Recall that $\theta \mapsto \psi_\theta$ is Lipschitz continuous, ∇J is bounded and remark that the quantity $\hat{\delta}(\tilde{x}_t, \bar{\omega}_*(\theta_t)) - \delta(\tilde{x}_t, \theta_t)$ is bounded. Therefore, using the Cauchy-Schwarz inequality, we have:

$$\begin{aligned}
&\mathbb{E}[\langle \nabla J(\theta_t), (\hat{\delta}(\tilde{x}_t, \bar{\omega}_*(\theta_t)) - \delta(\tilde{x}_t, \theta_t))\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t)\rangle] \\
&= \mathbb{E}[\langle \nabla J(\theta_t), (\hat{\delta}(\tilde{x}_t, \bar{\omega}_*(\theta_t)) - \delta(\tilde{x}_t, \theta_t))(\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t) - \psi_{\theta_{t-\tau}}(\tilde{S}_t, \tilde{A}_t))\rangle] \\
&+ \mathbb{E}[\langle \nabla J(\theta_t), (\hat{\delta}(\tilde{x}_t, \bar{\omega}_*(\theta_t)) - \delta(\tilde{x}_t, \theta_t))\psi_{\theta_{t-\tau}}(\tilde{S}_t, \tilde{A}_t)\rangle] \\
&\geq -C\mathbb{E}[\|\theta_t - \theta_{t-\tau}\|] + \mathbb{E}[\langle \nabla J(\theta_t), (\hat{\delta}(\tilde{x}_t, \bar{\omega}_*(\theta_t)) - \delta(\tilde{x}_t, \theta_t))\psi_{\theta_{t-\tau}}(\tilde{S}_t, \tilde{A}_t)\rangle] \,. \qquad (5.109)
\end{aligned}$$

Let us introduce the shorthand notation:

$$\Delta(\tilde{x}, \theta) := \langle \nabla J(\theta), (\hat{\delta}(\tilde{x}, \bar{\omega}_*(\theta)) - \delta(\tilde{x}, \theta))\psi_{\theta_{t-\tau}}(\tilde{S}, \tilde{A})\rangle$$

Note here that $\theta_{t-\tau}$ is fixed for $\psi_{\theta_{t-\tau}}$ in the above notation. The following decomposition holds:

$$\begin{aligned}
\Delta(\tilde{x}_t, \theta_t) = (\Delta(\tilde{x}_t, \theta_t) - \Delta(\tilde{x}_t, \theta_{t-\tau})) + (\Delta(\tilde{x}_t, \theta_{t-\tau}) - \Delta(\check{x}_t, \theta_{t-\tau})) \\
+ (\Delta(\check{x}_t, \theta_{t-\tau}) - \Delta(\bar{x}_t, \theta_{t-\tau})) + \Delta(\bar{x}_t, \theta_{t-\tau}) \,. \qquad (5.110)
\end{aligned}$$

Similar derivations to the previous section allow us to control each one of the error terms.

(i) Using that $\theta \mapsto \nabla J, \theta \mapsto V_{\pi_\theta}$ and $\theta \mapsto \bar{\omega}_*(\theta)$ are Lipschitz continuous, we obtain:

$$\Delta(\tilde{x}_t, \theta_t) - \Delta(\tilde{x}_t, \theta_{t-\tau}) \geq -C\|\theta_t - \theta_{t-\tau}\| \,. \qquad (5.111)$$

Using similar manipulations to the previous section, we get:

(ii)

$$\mathbb{E}[\Delta(\tilde{x}_t, \theta_{t-\tau}) - \Delta(\check{x}_t, \theta_{t-\tau})] \geq -C \sum_{i=t-\tau}^{t} \mathbb{E}[\|\theta_i - \theta_{t-\tau}\|] \,. \tag{5.112}$$

(iii)

$$\mathbb{E}[\Delta(\check{x}_t, \theta_{t-\tau}) - \Delta(\bar{x}_t, \theta_{t-\tau})] \geq -C\sigma^{\tau-1} \,. \tag{5.113}$$

(iv) For the last term, we can write:

$$\mathbb{E}[\Delta(\bar{x}_t, \theta_{t-\tau})|\theta_{t-\tau}] \geq -C\|\nabla J(\theta_{t-\tau})\| \cdot \mathbb{E}[|\hat{\delta}(\bar{x}_t, \bar{\omega}_*(\theta_{t-\tau})) - \delta(\bar{x}_t, \theta_{t-\tau})||\theta_{t-\tau}] \,. \tag{5.114}$$

Then, recall that $\bar{x}_t = (\bar{S}_t, \bar{A}_t, S_{t+1})$ where $S_{t+1} \sim p(\cdot|\bar{S}_t, \bar{A}_t)$ and observe that:

$$\begin{aligned}\hat{\delta}(\bar{x}_t, \bar{\omega}_*(\theta_{t-\tau})) - \delta(\bar{x}_t, \theta_{t-\tau}) = \gamma(\phi(S_{t+1})^T \bar{\omega}_*(\theta_{t-\tau}) - V_{\pi_{t-\tau}}(S_{t+1})) \\ + (V_{\pi_{t-\tau}}(\bar{S}_t) - \phi(\bar{S}_t)^T \bar{\omega}_*(\theta_{t-\tau})) \,.\end{aligned} \tag{5.115}$$

Recalling that $\tilde{p} = \gamma p + (1-\gamma)\rho$ and using Assumption 5.6.3, one can then easily show that:

$$\mathbb{E}[|\hat{\delta}(\bar{x}_t, \bar{\omega}_*(\theta_{t-\tau})) - \delta(\bar{x}_t, \theta_{t-\tau})||\theta_{t-\tau}] \leq C\epsilon_{\text{FA}} \,.$$

As a consequence, noticing again that ∇J is Lipschitz continuous, we obtain from Eq. (5.114):

$$\mathbb{E}[\Delta(\bar{x}_t, \theta_{t-\tau})] \geq -C\epsilon_{\text{FA}}\mathbb{E}[\|\nabla J(\theta_{t-\tau})\|] \geq -C\epsilon_{\text{FA}}\mathbb{E}[\|\theta_t - \theta_{t-\tau}\|] - C\epsilon_{\text{FA}}\mathbb{E}[\|\nabla J(\theta_t)\|] \,.$$

Combining items (i) to (iv) with the boundedness of the function ∇J, we conclude from this section that:

$$\begin{aligned}&\mathbb{E}[\langle \nabla J(\theta_t), (\hat{\delta}(\tilde{x}_t, \bar{\omega}_*(\theta_t)) - \delta(\tilde{x}_t, \theta_t))\psi_{\theta_t}(\tilde{S}_t, \tilde{A}_t)\rangle] \\ &\geq -C\mathbb{E}[\|\theta_t - \theta_{t-\tau}\|] - C\sum_{i=t-\tau}^{t} \mathbb{E}[\|\theta_i - \theta_{t-\tau}\|] - C\sigma^{\tau-1} - C\epsilon_{\text{FA}} \\ &\geq -C((\tau+1)^2\alpha_{t-\tau} + \alpha_T + \epsilon_{\text{FA}}) \,,\end{aligned} \tag{5.116}$$

where the last inequality has already been established in Section 5.8.1 with the choice of the mixing time $\tau = \tau_T$.

5.8.2.3 End of the proof of Th. 5.6

Combining Eq. (5.102) with Eqs. (5.103), (5.108) and (5.116) yields:

$$\begin{aligned}\mathbb{E}[J(\theta_{t+1})] \geq \mathbb{E}[J(\theta_t)] + \alpha_t\mathbb{E}[\|\nabla J(\theta_t)\|^2] - C\alpha_t\sqrt{\mathbb{E}[\|\nabla J(\theta_t)\|^2]}\sqrt{\mathbb{E}[\|\omega_t - \bar{\omega}_*(\theta_t)\|^2]} \\ - C\alpha_t((\tau+1)^2\alpha_{t-\tau} + \alpha_T + \epsilon_{\text{FA}}) - C\alpha_t^2 \,.\end{aligned} \tag{5.117}$$

Rearranging and summing this inequality for $t = \tau_T$ to T lead to:

$$\frac{1}{T - \tau_T + 1}\sum_{t=\tau_T}^{T} \mathbb{E}[\|\nabla J(\theta_t)\|^2] \leq U_1(T) + U_2(T) + U_3(T) + C\epsilon_{\text{FA}} \,, \tag{5.118}$$

where

$$U_1(T) := \frac{1}{T-\tau_T+1}\sum_{t=\tau_T}^{T}\frac{1}{\alpha_t}(\mathbb{E}[J(\theta_{t+1})]-\mathbb{E}[J(\theta_t)])\,, \tag{5.119}$$

$$U_2(T) := \frac{C}{T-\tau_T+1}\sum_{t=\tau_T}^{T}((\tau+1)^2\alpha_{t-\tau_T}+\alpha_T+\alpha_t)\,, \tag{5.120}$$

$$U_3(T) := \frac{C}{T-\tau_T+1}\sum_{t=\tau_T}^{T}\sqrt{\mathbb{E}[\|\nabla J(\theta_t)\|^2]}\sqrt{\mathbb{E}[\|\omega_t-\bar{\omega}_*(\theta_t)\|^2]}\,. \tag{5.121}$$

Let us now provide estimates of each one of the quantities $U_i(T)$ for $i=1,2,3$.

1. Noticing that the function J is bounded and the sequence (α_t) is nonincreasing, the first term can be controlled as follows:

$$\begin{aligned}U_1(T) &= \frac{1}{T-\tau_T+1}\left(\frac{1}{\alpha_T}\mathbb{E}[J(\theta_{T+1})]-\frac{1}{\alpha_{\tau_T-1}}\mathbb{E}[J(\theta_{\tau_T})]+\sum_{t=\tau_T}^{T}\left(\frac{1}{\alpha_{t-1}}-\frac{1}{\alpha_t}\right)\mathbb{E}[J(\theta_t)]\right)\\ &\le \frac{C}{T-\tau_T+1}\left(\frac{1}{\alpha_T}+\frac{1}{\alpha_{\tau_T-1}}+\frac{1}{\alpha_T}-\frac{1}{\alpha_{\tau_T-1}}\right)\\ &\le \frac{C}{T-\tau_T+1}\frac{2}{\alpha_T}\\ &= \mathcal{O}(T^{\alpha-1})\,.\end{aligned} \tag{5.122}$$

2. Recalling that the sequence of stepsizes (α_t) is nonincreasing and that $\tau_T = \mathcal{O}(\log T)$, the second term can be estimated by the following derivations:

$$\begin{aligned}U_2(T) &= \frac{C}{T-\tau_T+1}\left((\tau_T+1)^2\sum_{t=\tau_T}^{T}\alpha_{t-\tau_T}+(T-\tau_T+1)\alpha_T+\sum_{t=\tau_T}^{T}\alpha_t\right)\\ &\le \frac{C}{T-\tau_T+1}\left((\tau_T+1)^2\sum_{t=0}^{T-\tau_T}\alpha_t+(T-\tau_T+1)\alpha_T+\sum_{t=0}^{T-\tau_T}\alpha_t\right)\\ &\le \frac{C}{T-\tau_T+1}\left(((\tau_T+1)^2+1)\frac{(T-\tau_T+1)^{1-\alpha}}{1-\alpha}+(T-\tau_T+1)\alpha_T\right)\\ &= \mathcal{O}(\log^2(T)T^{-\alpha})\,.\end{aligned} \tag{5.123}$$

3. Using the Cauchy-Schwarz inequality, we have:

$$U_3(T) \le \frac{C}{T-\tau_T+1}\sqrt{\sum_{t=\tau_T}^{T}\mathbb{E}[\|\nabla J(\theta_t)\|^2]}\sqrt{\sum_{t=\tau_T}^{T}\mathbb{E}[\|\omega_t-\bar{\omega}_*(\theta_t)\|^2]}\,. \tag{5.124}$$

Define the quantities:

$$F(T) := \frac{1}{T - \tau_T + 1} \sum_{t=\tau_T}^{T} \mathbb{E}[\|\nabla J(\theta_t)\|^2] \,, \tag{5.125}$$

$$E(T) := \frac{1}{T - \tau_T + 1} \sum_{t=\tau_T}^{T} \mathbb{E}[\|\omega_t - \bar{\omega}_*(\theta_t)\|^2] \,, \tag{5.126}$$

$$K(T) := U_1(T) + U_2(T) + C\epsilon_{\text{FA}} \,. \tag{5.127}$$

Using these definitions, we can rewrite Eq. (5.118) as follows:

$$F(T) \leq C\sqrt{F(T)}\sqrt{E(T)} + K(T) \,.$$

Solving this inequality yields:

$$F(T) = \mathcal{O}(E(T)) + \mathcal{O}(K(T)) \,. \tag{5.128}$$

We conclude the proof by remarking that items (1) to (3) above imply:

$$K(T) = \mathcal{O}(T^{\alpha-1}) + \mathcal{O}(\log^2(T)T^{-\alpha}) + \mathcal{O}(\epsilon_{\text{FA}}) \,. \tag{5.129}$$

Eqs. (5.128) and (5.129) combined can be explicitely written as follows:

$$\begin{aligned} \frac{1}{T - \tau_T + 1} \sum_{t=\tau_T}^{T} \mathbb{E}[\|\nabla J(\theta_t)\|^2] = \mathcal{O}(T^{\alpha-1}) + \mathcal{O}(\log^2(T)T^{-\alpha}) + \mathcal{O}(\epsilon_{\text{FA}}) \\ + \mathcal{O}\left(\frac{1}{T - \tau_T + 1} \sum_{t=\tau_T}^{T} \mathbb{E}[\|\omega_t - \bar{\omega}_*(\theta_t)\|^2]\right) \,. \end{aligned} \tag{5.130}$$

Then, we can write

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[\|\nabla J(\theta_t)\|^2] &= \frac{1}{T} \left(\sum_{t=1}^{\tau_T - 1} \mathbb{E}[\|\nabla J(\theta_t)\|^2] + \sum_{t=\tau_T}^{T} \mathbb{E}[\|\nabla J(\theta_t)\|^2] \right) \\ &\leq \frac{C \log T}{T} + \mathcal{O}\left(\frac{1}{T - \tau_T + 1} \sum_{t=\tau_T}^{T} \mathbb{E}[\|\nabla J(\theta_t)\|^2] \right) \,. \end{aligned}$$

This completes the proof given Eq. (5.130).

5.8.2.4 Proof of Cor. 5.7

The result is a consequence of combining Ths. 5.5 and 5.6 and simplifying the obtained rate using the fact that $0 < \beta < \xi < \alpha < 1$.

6

Conclusion and Perspectives

6.1 About non-convex stochastic optimization

The major part of this thesis was devoted to the study of momentum and adaptive gradient methods for non-convex stochastic optimization. Our analysis started with the popular Adam algorithm for which we considered two different stepsizes regimes. In the constant stepsize regime, we established a long run convergence result under a stability assumption. In the vanishing stepsizes regime, we proposed a novel variant of Adam for which we showed the almost sure convergence of the iterates towards the set of critical points of the objective function and a central limit theorem characterizing the fluctuations. In our study, the ODE method was pivotal to understand the dynamical behavior of Adam and derive convergence results in discrete time. In the second part of this thesis, beyond Adam and starting from a more general continuous-time dynamical system, we analyzed a general stochastic optimization algorithm in the decreasing stepsizes regime. This algorithm offers degrees of freedom and thereby encompasses several adaptive gradient methods including Adam and momentum methods such as S-NAG. In this part, similarly to our previous Adam analysis, we established stability, almost sure convergence and convergence rates results. A major issue we tackled in this thesis is that of avoidance of traps, showing that the stochastic algorithms we considered in this thesis cannot converge towards undesirable critical points such as local maxima and saddle points. The linchpin of our proof is a general avoidance of traps result extending the seminal works of Pemantle (1990); Brandière and Duflo (1996) to a non-autonomous setting, and which we believe is of independent interest. Finally, we proved some quantitative results complementing our asymptotic analysis. These results consists of bounds controlling the expected gradient norm of the objective function along the iterations in the same stochastic setting and function value gap convergence rates in the deterministic setting.

Concerning these last contributions, we bring to the attention of the reader that the literature has witnessed many developments after the publication on which Chapter 4 is based. In particular, the recent works of Défossez et al. (2020) and (Gadat and Gavra, 2020, Th. 2) provide some interesting quantitative results in the flavor of the results we have presented about bounding the expected gradient of the objective function along the iterates of the algorithm. We also mention that it would be interesting to extend our convergence rates based on the Kurdyka-Łojasiewicz property to the stochastic setting. Benaïm (2016) provides some results in this flavor for stochastic gradient algorithms.

Our analysis in the main part of this thesis opens the way for several directions of future research.

Constrained optimization. First, it can be interesting to explore constrained stochastic optimization problems and design proximal variants of momentum and adaptive gradient methods such as Adam.

Nonsmooth optimization. Second, this thesis was focused on the case where the objective function is differentiable. However, neural networks may involve points of non-differentiability because of the use of some popular activation functions such as ReLU (Rectified Linear Unit which is the positive part function). Recent results (Davis et al., 2020; Majewski et al., 2018) study the stochastic subgradient method for a class of nonsmooth non-convex functions via the differential inclusion approach (Benaïm et al., 2005). The mathematical artillery to tackle this nonsmoothness problem was recently set up in the literature. Bolte and Pauwels (2021) introduced a notion of generalized derivatives leading to a class of locally Lipschitz functions called path differentiable functions encompassing convex, Clarke regular and Whitney stratifiable functions (and many others). Authors then developed a generalized differential calculus which allows to analyze modern algorithms based on automatic differentiation in a nonsmooth context. In particular, following the differential inclusion approach of Benaïm et al. (2005), discrete stochastic algorithms used for training nonsmooth deep neural networks were analyzed within this framework to obtain almost sure subsequential convergence to steady states. This framework can be applied to the popular algorithms we have studied in this thesis. Another approach based on the use of closed measures from the calculus of variations and dynamical systems was proposed by Bolte et al. (2020) to study the subgradient method for Lipschitz path differentiable functions and could prove useful for more complex dynamics such as the ones induced by the algorithms considered in the present thesis.

Regarding avoidance of traps, more challenging problems arise in the nonsmooth setting. Very recent research efforts have been undertaken to address the case of the stochastic subgradient descent method in nonsmooth stochastic optimization (Bianchi et al., 2021; Davis et al., 2021; Schechtman, 2021). The question remains open for more complex algorithms.

Other applications. The success of Adam in deep learning and stochastic optimization inspired novel algorithms for other problems such as min-max optimization (see for instance Extra-Adam in Gidel et al. (2019) mixing Adam with the extragradient method of Korpelevich (1976)) for which no theoretical guarantees exist to the best of our knowledge. Algorithms based on Adam could also prove useful in estimating optimal transport distances as this problem was cast as a stochastic optimization problem by Genevay et al. (2016) and further explored by Bercu and Bigot (2021); Bercu et al. (2021).

6.2 About actor-critic methods with target networks

In the RL part of this thesis, we studied an actor-critic algorithm incorporating a target network inspired from several state of the art algorithms used in deep RL. Our algorithm uses three different timescales: one for the actor and two for the critic. Instead of using the single timescale TD learning algorithm as a critic, we use a two timescales target-based version of TD learning closely inspired from practical actor-critic algorithms implementing target networks. More precisely, we proved both asymptotic and non-asymptotic convergence results. First, we showed that the critic which uses a target variable tracks a slowly moving target corresponding to the well-known TD solution. Then, we proved that the actor parameter visits infinitely often a region of the parameter space where the norm of the policy gradient is dominated by a bias induced by linear function approximation. Second, we established a bound controlling the average expected

gradient (of the performance function) norm evaluated at the actor iterates generated by our target-based algorithm showing the quantitative impact of using a target network.

There are several interesting directions for future research regarding our actor-critic method.

Nonlinear function approximation. Our analysis addresses the linear function approximation setting. Many state of the art actor-critic algorithms in deep RL make use of target networks to stabilize the training (Lillicrap et al., 2016; Haarnoja et al., 2018; Fujimoto et al., 2018). A theoretical justification of the use of a target network in the nonlinear function approximation setting beyond linear function approximation is a challenging problem that merit further investigation. In particular, as practical algorithms in deep RL seem to indicate, it would be interesting to see if such a mechanism can be a theoretically grounded alternative to the failure of temporal difference learning with nonlinear function approximation as was noticed in (Tsitsiklis and Van Roy, 1997, Section 10). A significant gap still persists between theory and practice. Actor-critic methods are usually analyzed in the two timescales stochastic approximation framework where two different stepsizes decreasing at different rates are used for the actor and the critic. However, recent general two timescales stochastic approximation results established by (Karmakar and Bhatnagar, 2018) do not permit to show the convergence of actor-critic methods with target networks in the nonlinear function approximation setting as the assumptions made almost amount to consider linear function approximation and lead to suspect the existence of undesirable accumulation points (see Assumption (A6'), p. 7 and Remark 8, p. 12 in Karmakar and Bhatnagar (2018)). This challenging problem would probably need an approach tailored to actor-critic methods with target networks. Besides, we recall that the target network idea was originally proposed in Mnih et al. (2013, 2015) which introduced the DQN algorithm. Recently, some theoretical issues were pointed out in Avrachenkov et al. (2021) for DQN (with single timescale) and authors proposed a modification to fix these issues with some additional computation. This work gives interesting insights on this target network mechanism and prompts us to be cautious about using such an appealing scheme, at least when considering a single timescale (see also Wang and Ueda (2021)).

Experience replay. This consists in storing past transitions (i.e., (state, action, reward, next state) tuples) in memory and reusing them throughout learning instead of only using transitions as they occur during simulation or actual experience. This feature has proven useful in RL thanks to its numerous advantages such as variance reduction, robustness to anomalous transitions, preventing overfitting, allowing data re-use and handling delayed rewards as this had been reported and discussed for instance in (Avrachenkov et al., 2021, Section 2.3) which deals with the specific case of DQN. The insights provided in the aforementioned work may prove useful for the analysis of actor-critic algorithms with experience replay.

Nonsmoothness. As previously discussed, some nonsmoothness issues related to the use of nonsmooth activation functions (such as ReLU) also arise and need further investigation.

Off-policy learning. Another possible avenue of future work is to extend our proposed actor-critic algorithm to the off-policy setting which is pervasive in several successful deep RL actor-critic algorithms (Lillicrap et al., 2016; Haarnoja et al., 2018; Fujimoto et al., 2018).. In this setting, the policy followed to generate samples called the behavior policy differs from the so-called target policy which is the policy of interest to be

evaluated for example in policy evaluation. Off-policy learning is paramount to manage the celebrated exploration-exploitation trade-off and to learn efficiently in large-scale problems in general. An algorithm combining off-policy learning, function approximation and bootstrapping simultaneously is usually not guaranteed to be well behaved, leading to the famous deadly triad (Sutton and Barto, 2018, Chapter 11, Section 3). If the discounted reward setting has been solved more than ten years ago (Sutton et al., 2009b,a), provably convergent value-based algorithms for the average-reward case have only been recently introduced by Zhang et al. (2021a). Breaking the deadly triad is still a current research question (Zhang et al., 2021b) and designing off-policy RL algorithms remains a notoriously challenging task which is the subject of active research (Xu et al., 2021). Furthermore, very recent finite-time analysis of off-policy actor-critic RL algorithms with linear function approximation (Chen et al., 2021) can be an additional starting point to investigate the case of off-policy actor-critic algorithms with target networks.

6.3 Importing momentum and adaptive methods into RL

We would like to end this thesis by discussing some future directions we actively began to explore. These opportunities of future work would allow us to bridge together the seemingly independent parts of this thesis which we summarized in the previous sections of this conclusive chapter.

Momentum and adaptive gradient methods such as RMSProp and Adam are widely used (if not ubiquitous) as optimizers to train neural networks approximators in deep reinforcement learning (see Mnih et al. (2015); Lillicrap et al. (2016); Mnih et al. (2016); Schulman et al. (2017); Haarnoja et al. (2018) to name a few famous examples). During the last couple of years, a flurry of non-asymptotic theoretical analysis has been conducted around standard policy gradient, value-based and actor-critic methods to determine their sample complexity. However, in this recent growing line of research, only few works touch upon momentum and adaptive methods in RL and they are almost all restricted to the linear function approximation case excluding deep neural networks. We also highlight that theoretical guarantees for (two-layer) neural network approximators are very recent even for standard algorithms such as TD learning (see for e.g., Cayci et al. (2021)) and Q-learning (see for instance Xu and Gu (2020)). Overall, despite the popularity of momentum and adaptive methods in deep RL, theoretical understanding of their convergence behavior is still in its infancy. Moreover, it can be interesting to design novel algorithms by incorporating momentum and adaptive gradient methods ideas from optimization into the landmark principled classes of RL algorithms, namely, policy-based methods, value-based methods and actor-critic algorithms. Few recent preliminary works have followed this direction (Vieillard et al., 2020; Sun et al., 2020; Xiong et al., 2021; Romoff et al., 2021; Weng et al., 2021). To the best of our knowledge, all the works incorporating momentum or adaptive methods into RL have no guarantees in the nonlinear function approximation case. As a first step, a better theoretical understanding of adaptive optimizers in TD learning is still needed to discover optimizers achieving a significant improvement over vanilla TD learning in practice.

Momentum and eligibility traces. Although different, the concept of momentum incorporated to TD learning shares some similarity with the eligibility traces mechanism which is widely known in the RL research community. As explained in (Sutton and Barto, 2018, Chapter 12), eligibility traces can be seen as a tool to remedy the problem

of long-delayed rewards and can be suitable for the non-Markov setting, i.e., when the environment is not modeled as an MDP. Interestingly, the concept of momentum from optimization captures the idea of a memory of the gradients used all over the iterations. Therefore, transferring momentum to RL has the potential to address the issue of long-delayed rewards and could also be beneficial in non- Markov settings. The comparison of momentum to eligibility traces has not been conducted in the literature and the analysis of a momentum variant of TD learning is missing. The recent work of Bengio et al. (2021) is concerned with momentum in TD learning but does not address the questions we have raised here.

Policy gradient methods with momentum and adaptive stepsizes. To reduce the large variance of vanilla policy gradient estimates and improve the sample complexity, Huang et al. (2020) and Yuan et al. (2020) proposed to augment the update rules with an exponential moving average. Nevertheless, the proposed algorithms are not truly adaptive because the learning rates are forced to be decreasing unlike in Adam. Moreover, global convergence rates for policy gradient methods incorporating adaptive stepsizes are still missing. Since adaptive gradient methods are "geometry-aware" algorithms, these methods should have interesting connections with the recently introduced geometry-aware normalized policy gradient (GNPG) and its analysis combining the concept of non-uniform smoothness and the non-uniform Łojasiewicz inequality (Mei et al., 2020, 2021). It would also be interesting to compare the convergence rates of our method to the linear convergence of GNPG in the deterministic setting where exact gradients are supposed to be available.

Beyond the RL problem we tackle in this thesis, we can consider several other RL problems such as risk-sensitive RL or multi-agent RL.

Risk-sensitive RL setting. Beyond risk-neutral RL where the performance metric to be maximized by the agent is for instance the expectation of the expected total (possibly discounted) reward, the agent may also aim at minimizing at the same time a risk measure (e.g., variance or value at risk). Although risk-sensitive control dates back to the previous century (see for e.g., Whittle (1990)) and have witnessed many developments (see for e.g., Borkar (2001, 2005); Bhatnagar (2010) actor-critic algorithms in this setting, Borkar (2002) for Q-learning, and Borkar (2010); García and Fernández (2015) for surveys, and the references therein), risk-sensitive RL algorithms are less advanced in comparison to the empirical success of risk-neutral RL that we discussed so far. This gap calls for the need to design new algorithms for this setting which is motivated by real-world applications requirements. Some of the most recent efforts in this direction include the works of Bisi et al. (2020) and Whiteson et al. (2021) (see also Karmakar and Bhatnagar (2021) for error bounds). We could leverage our ideas in risk-neutral RL with advanced optimization methods for risk-sensitive RL.

Multi-agent RL setting. Finally, using momentum and adaptive gradient methods could also prove useful in the multi-agent RL setting. In the distributed RL problem, a network of agents seeks to maximize a global return cooperatively via communication with local neighbors. Consider for instance the setting where rewards are decentralized and each agent has access to the full state and action information. In the last few years, distributed versions of TD learning, Q-learning and actor-critic algorithms have been proposed in the literature (see Lee et al. (2020) for a recent survey). If the Adam algorithm has been recently extended to a distributed learning context by Chen et al. (2021) for supervised learning, an adaptation of the algorithm to the multi-agent RL setting is yet to be proposed.

Bibliography

M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pages 265–283, 2016. page 3

P.-A. Absil, R. Mahony, and B. Andrews. Convergence of the iterates of descent methods for analytic cost functions. *SIAM Journal on Optimization*, 16(2):531–547, 2005. page 106

N. Agarwal, B. Bullins, X. Chen, E. Hazan, K. Singh, C. Zhang, and Y. Zhang. Efficient full-matrix adaptive regularization. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 102–110, Long Beach, California, USA, 09–15 Jun 2019. PMLR. URL http://proceedings.mlr.press/v97/agarwal19b.html. page 104

A. Alacaoglu, Y. Malitsky, and V. Cevher. Convergence of adaptive algorithms for weakly convex constrained optimization. *arXiv preprint arXiv:2006.06650*, 2020a. page 64

A. Alacaoglu, Y. Malitsky, P. Mertikopoulos, and V. Cevher. A new regret analysis for Adam-type algorithms. In H. D. III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 202–210, 13–18 Jul 2020b. URL http://proceedings.mlr.press/v119/alacaoglu20b.html. page 64

F. Alvarez. On the minimizing property of a second order dissipative system in Hilbert spaces. *SIAM Journal on Control and Optimization*, 38(4):1102–1119, 2000. page 69

M. Assran and M. Rabbat. On the convergence of Nesterov's accelerated gradient method in stochastic settings. In H. D. III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 410–420, 13–18 Jul 2020. URL http://proceedings.mlr.press/v119/assran20a.html. page 64

H. Attouch and J. Bolte. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Mathematical Programming*, 116(1-2):5–16, 2009. pages 23, 106, 111

H. Attouch, X. Goudou, and P. Redont. The heavy ball with friction method, i. the continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system. *Communications in Contemporary Mathematics*, 2(01):1–34, 2000. pages 28, 55

H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-łojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010. pages 111, 112

H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Mathematical Programming*, 168(1-2):123–175, 2018. page 57

J.-F. Aujol, C. Dossal, and A. Rondepierre. Optimal convergence rates for nesterov acceleration. *SIAM Journal on Optimization*, 29(4):3131–3153, 2019. doi: 10.1137/18M1186757. URL https://doi.org/10.1137/18M1186757. page 57

K. E. Avrachenkov, V. S. Borkar, H. P. Dolhare, and K. Patil. Full gradient dqn reinforcement learning: a provably convergent scheme. In *Modern Trends in Controlled Stochastic Processes:*, pages 192–220. Springer, 2021. pages 6, 171

F. Bach and E. Moulines. Non-strongly-convex smooth stochastic approximation with convergence rate o(1/n). In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper/2013/file/7fe1f8abaad094e0b5cb1b01d712f708-Paper.pdf. page 9

L. Balles and P. Hennig. Dissecting adam: The sign, magnitude and variance of stochastic gradients. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 404–413, 2018. pages 22, 28

E. Barnard. Temporal-difference methods and markov models. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(2):357–365, 1993. doi: 10.1109/21.229449. page 135

A. G. Barto, R. S. Sutton, and C. W. Anderson. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(5):834–846, 1983. doi: 10.1109/TSMC.1983.6313077. pages 7, 129

A. Basu, S. De, A. Mukherjee, and E. Ullah. Convergence guarantees for rmsprop and adam in non-convex optimization and their comparison to nesterov acceleration on autoencoders. *arXiv preprint arXiv:1807.06766*, 2018. pages 28, 64, 106, 115

A. Belotto da Silva and M. Gazeau. A general system of differential equations to model first-order adaptive algorithms. *Journal of Machine Learning Research*, 21(129):1–42, 2020. URL http://jmlr.org/papers/v21/18-808.html. pages 11, 12, 54, 55, 56, 57, 58, 69, 70

A. Belotto da Silva and M. Gazeau. A general system of differential equations to model first order adaptive algorithms. *arXiv preprint arXiv:1810.13108*, 31 Oct 2018. page 28

M. Benaïm. A dynamical system approach to stochastic approximations. *SIAM J. Control Optim.*, 34(2):437–472, 1996. ISSN 0363-0129. doi: 10.1137/S0363012993253534. URL https://doi.org/10.1137/S0363012993253534. pages 9, 143, 147

M. Benaïm. Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 1–68. Springer, Berlin, 1999. pages 9, 13, 20, 37, 47, 55, 68, 75, 76, 77, 79

M. Benaïm. On strict convergence of stochastic gradients. *arXiv preprint arXiv:1610.03278*, 2016. page 169

M. Benaïm and M. W. Hirsch. Asymptotic pseudotrajectories and chain recurrent flows, with applications. *J. Dynam. Differential Equations*, 8(1):141–176, 1996. ISSN 1040-7294. doi: 10.1007/BF02218617. URL http://dx.doi.org/10.1007/BF02218617. pages 37, 75

M. Benaïm and S. J. Schreiber. Ergodic properties of weak asymptotic pseudotrajectories for semiflows. *J. Dynam. Differential Equations*, 12(3):579–598, 2000. ISSN 1040-7294. page 20

M. Benaïm, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. *SIAM Journal on Control and Optimization*, 44(1):328–348, 2005. page 170

E. Bengio, J. Pineau, and D. Precup. Correcting momentum in temporal difference learning. *arXiv preprint arXiv:2106.03955*, 2021. page 173

A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1990. ISBN 3-540-52894-6. doi: 10.1007/978-3-642-75894-2. URL https://doi.org/10.1007/978-3-642-75894-2. Translated from the French by Stephen S. Wilson. pages 9, 130, 151

B. Bercu and J. Bigot. Asymptotic distribution and convergence rates of stochastic algorithms for entropic optimal transportation between probability measures. *The Annals of Statistics*, 49(2):968 – 987, 2021. doi: 10.1214/20-AOS1987. URL https://doi.org/10.1214/20-AOS1987. pages 9, 170

B. Bercu, J. Bigot, S. Gadat, and E. Siviero. A stochastic gauss-newton algorithm for regularized semi-discrete optimal transport. *arXiv preprint arXiv:2107.05291*, 2021. pages 9, 170

D. P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, 1st edition, 1996. ISBN 1886529108. pages 7, 129, 132, 135, 138, 145

D. P. Bertsekas and J. N. Tsitsiklis. Gradient convergence in gradient methods with errors. *SIAM Journal on Optimization*, 10(3):627–642, 2000. page 9

J. Bhandari, D. Russo, and R. Singal. A finite time analysis of temporal difference learning with linear function approximation. In S. Bubeck, V. Perchet, and P. Rigollet, editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 1691–1692. PMLR, 06–09 Jul 2018. URL http://proceedings.mlr.press/v75/bhandari18a.html. pages 131, 137, 139

S. Bhatnagar. An actor–critic algorithm with function approximation for discounted cost constrained markov decision processes. *Systems & Control Letters*, 59(12):760 – 766, 2010. ISSN 0167-6911. doi: https://doi.org/10.1016/j.sysconle.2010.08.013. URL http://www.sciencedirect.com/science/article/pii/S0167691110001246. page 173

S. Bhatnagar, R. S. Sutton, M. Ghavamzadeh, and M. Lee. Natural actor-critic algorithms. *Automatica*, 45(11):2471–2482, 2009. pages 8, 129, 130, 131, 136, 137

P. Bianchi, W. Hachem, and A. Salim. Constant step stochastic approximations involving differential inclusions: Stability, long-run convergence and applications. *Stochastics*, 91(2):288–320, 2019. page 42

P. Bianchi, W. Hachem, and S. Schechtman. Convergence of constant step stochastic gradient descent for non-smooth non-convex functions. *arXiv preprint arXiv:2005.08513*, 2020. page 2

P. Bianchi, W. Hachem, and S. Schechtman. Stochastic subgradient descent escapes active strict saddles. *arXiv preprint arXiv:2108.02072*, 2021. page 170

L. Bisi, L. Sabbioni, E. Vittori, M. Papini, and M. Restelli. Risk-averse trust region optimization for reward-volatility reduction. In C. Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4583–4589. International Joint Conferences on Artificial Intelligence Organization, 7 2020. Special Track on AI in FinTech. page 173

J. Bolte and E. Pauwels. Conservative set valued fields, automatic differentiation, stochastic gradient methods and deep learning. *Mathematical Programming*, 188(1): 19–51, 2021. pages 2, 170

J. Bolte, A. Daniilidis, O. Ley, and L. Mazet. Characterizations of łojasiewicz inequalities: subgradient flows, talweg, convexity. *Transactions of the American Mathematical Society*, 362(6):3319–3363, 2010. page 110

J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, 2014. pages 23, 111, 112

J. Bolte, S. Sabach, M. Teboulle, and Y. Vaisbourd. First order methods beyond convexity and lipschitz gradient continuity with applications to quadratic inverse problems. *SIAM Journal on Optimization*, 28(3):2131–2151, 2018. pages 111, 112, 121, 122, 124

J. Bolte, E. Pauwels, and R. Rios-Zertuche. Long term dynamics of the subgradient method for lipschitz path differentiable functions. *arXiv preprint arXiv:2006.00098*, 2020. page 170

V. S. Borkar. A sensitivity formula for risk-sensitive cost and the actor–critic algorithm. *Systems & Control Letters*, 44(5):339–346, 2001. ISSN 0167-6911. doi: https://doi.org/10.1016/S0167-6911(01)00152-9. URL https://www.sciencedirect.com/science/article/pii/S0167691101001529. page 173

V. S. Borkar. Q-learning for risk-sensitive control. *Mathematics of operations research*, 27(2):294–311, 2002. page 173

V. S. Borkar. An actor-critic algorithm for constrained markov decision processes. *Systems & Control Letters*, 54(3):207–213, 2005. ISSN 0167-6911. doi: https://doi.org/10.1016/j.sysconle.2004.08.007. URL https://www.sciencedirect.com/science/article/pii/S0167691104001276. page 173

V. S. Borkar. *Stochastic approximation. A dynamical systems viewpoint*. Cambridge University Press, Cambridge; Hindustan Book Agency, New Delhi, 2008. ISBN 978-0-521-51592-4. pages 9, 13, 130, 136, 137, 141

V. S. Borkar. Learning algorithms for risk-sensitive control. In *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems–MTNS*, volume 5, 2010. page 173

V. S. Borkar and S. Chandak. Prospect-theoretic q-learning. *arXiv preprint arXiv:2104.05311*, 2021. page 6

V. S. Borkar and S. P. Meyn. The ode method for convergence of stochastic approximation and reinforcement learning. *SIAM Journal on Control and Optimization*, 38(2):447–469, 2000. page 136

L. Bottou, F. Curtis, and J. Nocedal. Optimization methods for large-scale machine learning. *SIAM Review*, 60(2):223–311, 2018. pages 1, 9

O. Brandière and M. Duflo. Les algorithmes stochastiques contournent-ils les pièges? *Ann. Inst. H. Poincaré Probab. Statist.*, 32(3):395–427, 1996. ISSN 0246-0203. URL http://www.numdam.org/item?id=AIHPB_1996__32_3_395_0. pages 13, 65, 68, 89, 94, 97, 169

A. Cabot, H. Engler, and S. Gadat. On the long time behavior of second order differential equations with asymptotically small dissipation. *Transactions of the American Mathematical Society*, 361(11):5983–6017, 2009. pages 28, 57, 58, 73

C. Castera, J. Bolte, C. Févotte, and E. Pauwels. An inertial newton algorithm for deep learning. *arXiv preprint arXiv:1905.12278*, 2019. page 110

S. Cayci, S. Satpathi, N. He, and R. Srikant. Sample complexity and overparameterization bounds for projection-free neural td learning. *arXiv preprint arXiv:2103.01391*, 2021. page 172

C. Chen, H. Wei, N. Xu, G. Zheng, M. Yang, Y. Xiong, K. Xu, and Z. Li. Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04): 3414–3421, Apr. 2020. doi: 10.1609/aaai.v34i04.5744. URL https://ojs.aaai.org/index.php/AAAI/article/view/5744. page 5

J. Chen, D. Zhou, Y. Tang, Z. Yang, and Q. Gu. Closing the generalization gap of adaptive gradient methods in training deep neural networks. *arXiv preprint arXiv:1806.06763*, 2018. pages 64, 107

T. Chen, Z. Guo, Y. Sun, and W. Yin. Cada: Communication-adaptive distributed adam. In *International Conference on Artificial Intelligence and Statistics*, pages 613–621. PMLR, 2021. pages 172, 173

X. Chen, S. Liu, R. Sun, and M. Hong. On the convergence of a class of adam-type algorithms for non-convex optimization. In *International Conference on Learning Representations*, 2019. pages 27, 28, 64, 106, 107, 115

G. Dalal, G. Thoppe, B. Szörényi, and S. Mannor. Finite sample analysis of two-timescale stochastic approximation with applications to reinforcement learning. In S. Bubeck, V. Perchet, and P. Rigollet, editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 1199–1233. PMLR, 06–09 Jul 2018. URL http://proceedings.mlr.press/v75/dalal18a.html. page 131

J. L. Daleckii and M. G. Krein. *Stability of solutions of differential equations in Banach space*. American Mathematical Society, Providence, R.I., 1974. Translated from the Russian by S. Smith, Translations of Mathematical Monographs, Vol. 43. pages 13, 54, 68, 89

D. Davis, D. Drusvyatskiy, S. Kakade, and J. Lee. Stochastic subgradient method converges on tame functions. *Foundations of Computational Mathematics*, 20(1): 119–154, 2020. pages 2, 28, 110, 111, 170

D. Davis, D. Drusvyatskiy, and L. Jiang. Subgradient methods near active manifolds: saddle point avoidance, local convergence, and asymptotic normality. *arXiv preprint arXiv:2108.11832*, 2021. page 170

A. Défossez, L. Bottou, F. Bach, and N. Usunier. A simple convergence proof of adam and adagrad. *arXiv preprint arXiv:2003.02395*, 2020. pages 64, 169

B. Delyon. General results on the convergence of stochastic algorithms. *IEEE Transactions on Automatic Control*, 41(9):1245–1255, 1996. page 9

B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the em algorithm. *Annals of statistics*, pages 94–128, 1999. pages 85, 86, 87

J. Diakonikolas and M. I. Jordan. Generalized momentum-based methods: A hamiltonian perspective. *arXiv preprint arXiv:1906.00436*, 2019. page 105

T. Dozat. Incorporating nesterov momentum into adam. 2016. page 107

S. S. Du, C. Jin, J. D. Lee, M. I. Jordan, A. Singh, and B. Poczos. Gradient descent can take exponential time to escape saddle points. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 1067–1077. Curran Associates, Inc., 2017. URL http://papers.nips.cc/paper/6707-gradient-descent-can-take-exponential-time-to-escape-saddle-points.pdf. page 68

J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159, 2011. pages 2, 18, 27, 61, 104

M. Duflo. *Random iterative models*, volume 34 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. ISBN 3-540-57100-0. page 9

J.-C. Fort and G. Pagès. Asymptotic behavior of a Markovian stochastic algorithm with constant step. *SIAM J. Control Optim.*, 37(5):1456–1482 (electronic), 1999. ISSN 0363-0129. page 45

S. Fujimoto, H. van Hoof, and D. Meger. Addressing function approximation error in actor-critic methods. In J. Dy and A. Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1587–1596. PMLR, 10–15 Jul 2018. URL http://proceedings.mlr.press/v80/fujimoto18a.html. pages 7, 129, 171

S. Gadat and I. Gavra. Asymptotic study of stochastic adaptive algorithm in non-convex landscape. *arXiv preprint arXiv:2012.05640*, 2020. pages 61, 64, 68, 169

S. Gadat, F. Panloup, and S. Saadane. Stochastic heavy ball. *Electron. J. Stat.*, 12(1):461–529, 2018. doi: 10.1214/18-EJS1395. URL https://doi.org/10.1214/18-EJS1395. pages 3, 11, 13, 27, 28, 54, 63, 64, 68, 73

J. García and F. Fernández. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16(1):1437–1480, 2015. page 173

A. Genevay, M. Cuturi, G. Peyré, and F. Bach. Stochastic optimization for large-scale optimal transport. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper/2016/file/2a27b8144ac02f67687f76782a3b5d8f-Paper.pdf. page 170

S. Ghadimi and G. Lan. Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013. pages 109, 110

G. Gidel, H. Berard, G. Vignoud, P. Vincent, and S. Lacoste-Julien. A variational inequality perspective on generative adversarial networks. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=r1laEnA5Ym. page 170

R. M. Gower, N. Loizou, X. Qian, A. Sailanbayev, E. Shulgin, and P. Richtárik. Sgd: General analysis and improved rates. In *International Conference on Machine Learning*, pages 5200–5209. PMLR, 2019. page 9

S. Gu, E. Holly, T. Lillicrap, and S. Levine. Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 3389–3396. IEEE, 2017. page 5

H. Gupta, R. Srikant, and L. Ying. Finite-time performance bounds and adaptive learning rate selection for two time-scale reinforcement learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/e354fd90b2d5c777bfec87a352a18976-Paper.pdf. page 131

V. Gupta, T. Koren, and Y. Singer. A unified approach to adaptive regularization in online and stochastic optimization. *arXiv preprint arXiv:1706.06569*, 2017. page 104

T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In J. Dy and A. Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1861–1870. PMLR, 10–15 Jul 2018. URL http://proceedings.mlr.press/v80/haarnoja18b.html. pages 7, 129, 171, 172

A. Haraux. *Systèmes dynamiques dissipatifs et applications*, volume 17. Masson, 1991. pages 37, 38, 72, 75

A. Haraux and M. Jendoubi. *The convergence problem for dissipative autonomous systems*. Springer Briefs in Mathematics. Springer International Publishing, 2015. pages 23, 39, 40

P. Hartman. *Ordinary Differential Equations*. Society for Industrial and Applied Mathematics, second edition, 2002. doi: 10.1137/1.9780898719222. URL https://epubs.siam.org/doi/abs/10.1137/1.9780898719222. page 70

N. Heess, J. J. Hunt, T. P. Lillicrap, and D. Silver. Memory-based control with recurrent neural networks. *arXiv preprint arXiv:1512.04455*, 2015. pages 7, 129

M. Hong, H.-T. Wai, Z. Wang, and Z. Yang. A two-timescale framework for bilevel optimization: Complexity analysis and application to actor-critic. *arXiv preprint arXiv:2007.05170*, 2020. page 131

R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1994. ISBN 0-521-46713-6. Corrected reprint of the 1991 original. pages 90, 147

Y.-P. Hsieh, P. Mertikopoulos, and V. Cevher. The limits of min-max optimization algorithms: Convergence to spurious non-critical sets. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 4337–4348. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/hsieh21a.html. page 9

F. Huang, S. Gao, J. Pei, and H. Huang. Momentum-based policy gradient methods. In *International Conference on Machine Learning*, pages 4422–4433. PMLR, 2020. page 173

P. Jain, S. M. Kakade, R. Kidambi, P. Netrapalli, and A. Sidford. Accelerating stochastic gradient descent for least squares regression. In *Conference On Learning Theory*, pages 545–604. PMLR, 2018. page 3

C. Jin, R. Ge, P. Netrapalli, S. M. Kakade, and M. I. Jordan. How to escape saddle points efficiently. volume 70 of *Proceedings of Machine Learning Research*, pages 1724–1732. PMLR, 2017. URL http://proceedings.mlr.press/v70/jin17a.html. page 68

P. R. Johnstone and P. Moulin. Convergence rates of inertial splitting schemes for nonconvex composite optimization. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4716–4720. IEEE, 2017. pages 106, 111, 124

M. Kaledin, E. Moulines, A. Naumov, V. Tadic, and H. Wai. Finite time analysis of linear two-timescale stochastic approximation with markovian noise. In J. D. Abernethy and S. Agarwal, editors, *Conference on Learning Theory, COLT 2020, 9-12 July 2020, Virtual Event [Graz, Austria]*, volume 125 of *Proceedings of Machine Learning Research*, pages 2144–2203. PMLR, 2020. URL http://proceedings.mlr.press/v125/kaledin20a.html. page 131

I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*. Springer-Verlag, New York, second edition, 1991. pages 50, 84

H. Karimi, J. Nutini, and M. Schmidt. Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 795–811. Springer, 2016. page 112

P. Karmakar and S. Bhatnagar. Two time-scale stochastic approximation with controlled Markov noise and off-policy temporal-difference learning. *Math. Oper. Res.*, 43 (1):130–151, 2018. ISSN 0364-765X. doi: 10.1287/moor.2017.0855. URL https://doi.org/10.1287/moor.2017.0855. pages 136, 137, 141, 143, 171

P. Karmakar and S. Bhatnagar. On tight bounds for function approximation error in risk-sensitive reinforcement learning. *Systems & Control Letters*, 150:104899, 2021. ISSN 0167-6911. doi: https://doi.org/10.1016/j.sysconle.2021.104899. URL https://www.sciencedirect.com/science/article/pii/S0167691121000293. page 173

R. Kidambi, P. Netrapalli, P. Jain, and S. M. Kakade. On the insufficiency of existing momentum schemes for stochastic optimization. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJTutzbA-. page 3

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. pages 3, 10, 11, 17, 18, 20, 21, 27, 28, 53, 55, 103, 104, 110

B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. Al Sallab, S. Yogamani, and P. Pérez. Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 2021. page 5

P. E. Kloeden and M. Rasmussen. *Nonautonomous dynamical systems*, volume 176 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011. ISBN 978-0-8218-6871-3. doi: 10.1090/surv/176. URL https://doi.org/10.1090/surv/176. pages 13, 54, 68, 89, 91, 93, 94

V. R. Konda. *Actor-Critic Algorithms*. PhD thesis, USA, 2002. AAI0804543. pages 130, 133, 138, 152

V. R. Konda and V. S. Borkar. Actor-critic–type learning algorithms for markov decision processes. *SIAM Journal on control and Optimization*, 38(1):94–123, 1999. pages 7, 129, 136

V. R. Konda and J. N. Tsitsiklis. On actor-critic algorithms. *SIAM journal on Control and Optimization*, 42(4):1143–1166, 2003. pages 7, 129, 130, 131, 133, 137, 139, 149, 152

G. M. Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976. page 170

H. Kumar, A. Koppel, and A. Ribeiro. On the sample complexity of actor-critic method for reinforcement learning with function approximation. *arXiv preprint arXiv:1910.08412*, 2019. pages 131, 140

K. Kurdyka. On gradients of functions definable in o-minimal structures. In *Annales de l'institut Fourier*, volume 48, pages 769–783, 1998. pages 110, 111

H. J. Kushner and D. S. Clark. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer-Verlag, Berlin, Heidelberg, New York., 1978. page 9

H. J. Kushner and G. G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. ISBN 0-387-00894-2. Stochastic Modelling and Applied Probability. pages 9, 136

C. Lakshminarayanan and S. Bhatnagar. A stability criterion for two timescale stochastic approximation schemes. *Automatica*, 79:108–114, 2017. page 136

N. Lazic, T. Lu, C. Boutilier, M. Ryu, E. J. Wong, B. Roy, and G. Imwalle. Data center cooling using model-predictive control. In *Proceedings of the Thirty-second Conference on Neural Information Processing Systems (NeurIPS-18)*, pages 3818–3827, Montreal, QC, 2018. URL https://papers.nips.cc/paper/7638-data-center-cooling-using-model-predictive-control. page 5

D. Lee and N. He. Target-based temporal-difference learning. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 3713–3722. PMLR, 09–15 Jun 2019. URL http://proceedings.mlr.press/v97/lee19a.html. pages 15, 129, 131, 132, 134

D. Lee, N. He, P. Kamalaruban, and V. Cevher. Optimization for reinforcement learning: From a single agent to cooperative agents. *IEEE Signal Processing Magazine*, 37(3): 123–135, 2020. page 173

J. D. Lee, I. Panageas, G. Piliouras, M. Simchowitz, M. I. Jordan, and B. Recht. First-order methods almost always avoid strict saddle points. *Math. Program.*, 176(1-2, Ser. B):311–337, 2019. ISSN 0025-5610. doi: 10.1007/s10107-019-01374-3. URL https://doi.org/10.1007/s10107-019-01374-3. page 68

S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *The International Journal of Robotics Research*, 37(4-5):421–436, 2018. page 5

G. Li and T. K. Pong. Calculus of the exponent of kurdyka–łojasiewicz inequality and its applications to linear convergence of first-order methods. *Foundations of computational mathematics*, 18(5):1199–1232, 2018. page 112

J. Li, W. Monroe, A. Ritter, D. Jurafsky, M. Galley, and J. Gao. Deep reinforcement learning for dialogue generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1192–1202, Austin, Texas, Nov. 2016. Association for Computational Linguistics. doi: 10.18653/v1/D16-1127. URL https://aclanthology.org/D16-1127. page 5

Q. Li, Y. Zhou, Y. Liang, and P. K. Varshney. Convergence analysis of proximal gradient with momentum for nonconvex optimization. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2111–2119. JMLR. org, 2017. pages 106, 111

X. Li and F. Orabona. On the convergence of stochastic gradient descent with adaptive stepsizes. In *Proceedings of Machine Learning Research*, volume 89, pages 983–992. PMLR, 16–18 Apr 2019. URL http://proceedings.mlr.press/v89/li19c.html. page 106

J. Liang, J. Fadili, and G. Peyré. A multi-step inertial forward-backward splitting method for non-convex optimization. In *Advances in Neural Information Processing Systems*, pages 4035–4043, 2016. page 106

T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. Continuous control with deep reinforcement learning. In *ICLR 2016*, 2016. URL http://arxiv.org/abs/1509.02971. pages 7, 129, 132, 135, 141, 171, 172

L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han. On the variance of the adaptive learning rate and beyond. *arXiv preprint arXiv:1908.03265*, 2019. page 107

S. Liu, K. C. See, K. Y. Ngiam, L. A. Celi, X. Sun, and M. Feng. Reinforcement learning for clinical decision support in critical care: comprehensive review. *Journal of medical Internet research*, 22(7):e18477, 2020. page 5

L. Ljung. Analysis of recursive stochastic algorithms. *IEEE transactions on automatic control*, 22(4):551–575, 1977. page 8

S. Łojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles*, 117:87–89, 1963. pages 22, 110

L. Luo, Y. Xiong, and Y. Liu. Adaptive gradient methods with dynamic bound of learning rate. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg3g2R9FX. pages 106, 107, 114, 115

J. Ma and D. Yarats. Quasi-hyperbolic momentum and adam for deep learning. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=S1fUpoR5FQ. page 107

V. V. Mai and M. Johansson. Convergence of a stochastic gradient method with momentum for nonsmooth nonconvex optimization. Proceedings of Machine Learning Research. PMLR, 2020. page 64

S. Majewski, B. Miasojedow, and E. Moulines. Analysis of nonsmooth stochastic approximation: the differential inclusion approach. *arXiv preprint arXiv:1805.01916*, 2018. pages 2, 170

P. Marbach and J. Tsitsiklis. Simulation-based optimization of markov reward processes. *IEEE Transactions on Automatic Control*, 46(2):191–209, 2001. doi: 10.1109/9.905687. pages 135, 144

H. B. McMahan and M. Streeter. Adaptive bound optimization for online convex optimization. pages 244–256, 2010. pages 2, 104

J. Mei, C. Xiao, C. Szepesvari, and D. Schuurmans. On the global convergence rates of softmax policy gradient methods. In H. D. III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 6820–6829. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/mei20b.html. page 173

J. Mei, Y. Gao, B. Dai, C. Szepesvari, and D. Schuurmans. Leveraging non-uniformity in first-order non-convex optimization. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 7555–7564. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/mei21a.html. page 173

P. Mertikopoulos, N. Hallak, A. Kavis, and V. Cevher. On the almost sure convergence of stochastic gradient descent in non-convex problems. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1117–1128. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/0cb5ebb1b34ec343dfe135db691e4a85-Paper.pdf. pages 9, 68

M. Métivier and P. Priouret. Applications of a kushner and clark lemma to general classes of stochastic algorithms. *IEEE Transactions on Information Theory*, 30(2): 140–151, 1984. page 9

M. Métivier and P. Priouret. Théorèmes de convergence presque sûre pour une classe d'algorithmes stochastiques à pas décroissant. *Probability Theory and Related Fields*, 74(3):403–428, Sep 1987. ISSN 1432-2064. doi: 10.1007/BF00699098. URL https://doi.org/10.1007/BF00699098. pages 9, 76

V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013. page 171

V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015. pages 5, 129, 132, 171, 172

V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In M. F. Balcan and K. Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1928–1937, New York, New York, USA, 20–22 Jun 2016. PMLR. URL http://proceedings.mlr.press/v48/mniha16.html. pages 129, 172

E. Moulines and F. Bach. Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24, pages 451–459. Curran Associates, Inc., 2011. URL https://proceedings.neurips.cc/paper/2011/file/40008b9a5380fcacce3976bf7c08af5b-Paper.pdf. page 9

Y. Nesterov. A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. In *Soviet Mathematics Doklady*, volume 27, pages 372–376, 1983. page 3

Y. Nesterov. *Introductory lectures on convex optimization: a basic course*. Springer: New York, NY, USA, 2004. pages 109, 121

P. Ochs. Local convergence of the heavy-ball method and ipiano for non-convex optimization. *Journal of Optimization Theory and Applications*, 177(1):153–180, 2018. pages 106, 111

P. Ochs, Y. Chen, T. Brox, and T. Pock. ipiano: Inertial proximal algorithm for nonconvex optimization. *SIAM Journal on Imaging Sciences*, 7(2):1388–1419, 2014. doi: 10.1137/130942954. URL https://doi.org/10.1137/130942954. pages 13, 105, 111, 120, 121

I. Panageas and G. Piliouras. Gradient descent only converges to minimizers: Non-isolated critical points and invariant regions. In *ITCS*, 2017. page 68

I. Panageas, G. Piliouras, and X. Wang. First-order methods almost always avoid saddle points: The case of vanishing step-sizes. In *Advances in Neural Information Processing Systems 32*, pages 6474–6483, 2019. page 68

R. Pascanu, T. Mikolov, and Y. Bengio. On the difficulty of training recurrent neural networks. In *International conference on machine learning*, pages 1310–1318, 2013. page 107

A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32:8026–8037, 2019. page 3

M. Pelletier. Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Annals of Applied Probability*, pages 10–44, 1998. pages 49, 50, 83, 84

R. Pemantle. Nonconvergence to unstable points in urn models and stochastic approximations. *Ann. Probab.*, 18(2):698–712, 1990. ISSN 0091-1798. URL http://links.jstor.org/sici?sici=0091-1798(199004)18:2<698:NTUPIU>2.0.CO;2-R&origin=MSN. pages 13, 65, 68, 89, 169

J. Peters and S. Schaal. Natural actor-critic. *Neurocomputing*, 71(7-9):1180–1190, 2008. page 129

B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964. pages 3, 105

C. Pötzsche and M. Rasmussen. Taylor approximation of integral manifolds. *J. Dynam. Differential Equations*, 18(2):427–460, 2006. ISSN 1040-7294. doi: 10.1007/s10884-006-9011-8. URL https://doi.org/10.1007/s10884-006-9011-8. page 91

L. A. Prashanth and S. Bhatnagar. Reinforcement learning with function approximation for traffic signal control. *IEEE Transactions on Intelligent Transportation Systems*, 12 (2):412–421, 2010. page 5

L. A. Prashanth and S. Bhatnagar. Reinforcement learning with average cost for adaptive control of traffic lights at intersections. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1640–1645, 2011. doi: 10.1109/ITSC.2011.6082823. page 5

M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014. pages 5, 129, 151

S. Qiu, Z. Yang, J. Ye, and Z. Wang. On the finite-time convergence of actor-critic algorithm. In *Optimization Foundations for Reinforcement Learning Workshop at Advances in Neural Information Processing Systems (NeurIPS)*, 2019. pages 131, 140

S. J. Reddi, S. Kale, and S. Kumar. On the convergence of adam and beyond. In *International Conference on Learning Representations*, 2018. pages 27, 28, 103, 106, 107, 108, 115

H. Robbins and S. Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951. pages 2, 8, 104

H. Robbins and D. Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing Methods in Statistics*, pages 233–257. Academic Press, New York, 1971. pages 48, 80

J. Romoff, P. Henderson, D. Kanaa, E. Bengio, A. Touati, P.-L. Bacon, and J. Pineau. Tdprop: Does adaptive optimization with jacobi preconditioning help temporal difference learning? In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1082–1090, 2021. page 172

P. Savarese. On the convergence of adabound and its connection to sgd. *arXiv preprint arXiv:1908.04457*, 2019. page 114

S. Schechtman. Stochastic subgradient descent on a generic definable function converges to a minimizer. *arXiv preprint arXiv:2109.02455*, 2021. page 170

J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. page 172

O. Sebbouh, R. M. Gower, and A. Defazio. Almost sure convergence rates for stochastic gradient descent and stochastic heavy ball. In M. Belkin and S. Kpotufe, editors, *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 3935–3971. PMLR, 15–19 Aug 2021. URL https://proceedings.mlr.press/v134/sebbouh21a.html. page 9

H. Shen, K. Zhang, M. Hong, and T. Chen. Asynchronous advantage actor critic: Non-asymptotic analysis and linear speedup. *arXiv preprint arXiv:2012.15511*, 2020. pages 131, 132, 134, 137, 139, 141, 152, 153, 154, 155, 162

R. Srikant and L. Ying. Finite-time error bounds for linear stochastic approximation and td learning. In A. Beygelzimer and D. Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 2803–2830, Phoenix, USA, 25–28 Jun 2019. PMLR. URL http://proceedings.mlr.press/v99/srikant19a.html. page 131

M. Staib, S. Reddi, S. Kale, S. Kumar, and S. Sra. Escaping saddle points with adaptive gradient methods. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 5956–5965, Long Beach, California, USA, 09–15 Jun 2019. PMLR. URL http://proceedings.mlr.press/v97/staib19a.html. page 104

W. Su, S. Boyd, and E. J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. *J. Mach. Learn. Res.*, 17:Paper No. 153, 43, 2016. ISSN 1532-4435. pages 57, 58, 73

T. Sun, H. Shen, T. Chen, and D. Li. Adaptive temporal difference learning with linear function approximation. *arXiv preprint arXiv:2002.08537*, 2020. page 172

I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147, 2013. pages 3, 105

R. S. Sutton. Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44, 1988. pages 14, 134

R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018. pages 4, 5, 7, 129, 132, 172

R. S. Sutton, D. Mcallester, S. Singh, and Y. Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Advances in Neural Information Processing Systems 12*, volume 99, pages 1057–1063. MIT Press, 1999. pages 6, 133, 139

R. S. Sutton, H. Maei, and C. Szepesvári. A convergent o(n) temporal-difference algorithm for off-policy learning with linear function approximation. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 21. Curran Associates, Inc., 2009a. URL https://proceedings.neurips.cc/paper/2008/file/e0c641195b27425bb056ac56f8953d24-Paper.pdf. page 172

R. S. Sutton, H. R. Maei, D. Precup, S. Bhatnagar, D. Silver, C. Szepesvári, and E. Wiewiora. Fast gradient-descent methods for temporal-difference learning with linear function approximation. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, page 993–1000, New York, NY, USA, 2009b. Association for Computing Machinery. ISBN 9781605585161. doi: 10.1145/1553374.1553501. URL https://doi.org/10.1145/1553374.1553501. pages 135, 172

C. Szepesvári. Algorithms for reinforcement learning. *Synthesis lectures on artificial intelligence and machine learning*, 4(1):1–103, 2010. pages 7, 132

T. Tieleman and G. Hinton. Lecture 6.e-rmsprop: Divide the gradient by a running average of its recent magnitude. *Coursera: Neural networks for machine learning*, pages 26–31, 2012. pages 3, 18, 27, 57, 61, 104

C. Traoré and E. Pauwels. Sequential convergence of adagrad algorithm for smooth convex optimization. *Operations Research Letters*, 49(4):452–458, 2021. ISSN 0167-6377. doi: https://doi.org/10.1016/j.orl.2021.04.011. URL https://www.sciencedirect.com/science/article/pii/S0167637721000651. page 27

J. N. Tsitsiklis and B. Van Roy. An analysis of temporal-difference learning with function approximation. *IEEE transactions on automatic control*, 42(5):674–690, 1997. pages 130, 131, 137, 138, 145, 171

N. Vieillard, B. Scherrer, O. Pietquin, and M. Geist. Momentum in reinforcement learning. In *International Conference on Artificial Intelligence and Statistics*, pages 2529–2538. PMLR, 2020. page 172

L. Wang, Q. Cai, Z. Yang, and Z. Wang. Neural policy gradient methods: Global optimality and rates of convergence. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=BJgQfkSYDS. page 131

Z. T. Wang and M. Ueda. A convergent and efficient deep q network algorithm. *arXiv preprint arXiv:2106.15419*, 2021. page 171

R. Ward, X. Wu, and L. Bottou. AdaGrad stepsizes: Sharp convergence over nonconvex landscapes. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 6677–6686, 2019a. page 27

R. Ward, X. Wu, and L. Bottou. Adagrad stepsizes: Sharp convergence over nonconvex landscapes. In *International Conference on Machine Learning*, pages 6677–6686, 2019b. page 106

C. J. C. H. Watkins. Learning from delayed rewards. 1989. page 6

B. Weng, H. Xiong, L. Zhao, Y. Liang, and W. Zhang. Finite-time theory for momentum q-learning. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2021. page 172

S. Whiteson, S. Zhang, and B. Liu. Mean- variance policy iteration for risk- averse reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*. Association for the Advancement of Artificial Intelligence, 2021. page 173

P. Whittle. *Risk-sensitive optimal control*, volume 2. Wiley, 1990. page 173

A. Wibisono, A. C. Wilson, and M. I. Jordan. A variational perspective on accelerated methods in optimization. *proceedings of the National Academy of Sciences*, 113(47): E7351–E7358, 2016. page 3

R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3):229–256, 1992. page 6

A. C. Wilson, B. Recht, and M. I. Jordan. A lyapunov analysis of accelerated methods in optimization. *Journal of Machine Learning Research*, 22(113):1–34, 2021. page 3

X. Wu, R. Ward, and L. Bottou. Wngrad: Learn the learning rate in gradient descent. *arXiv preprint arXiv:1803.02865*, 2018. page 106

Y. F. Wu, W. Zhang, P. Xu, and Q. Gu. A finite-time analysis of two time-scale actor-critic methods. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 17617–17628. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/cc9b3c69b56df284846bf2432f1cba90-Paper.pdf. pages 131, 132, 137, 139, 140, 141, 152, 154, 155, 163

Z. Wu and M. Li. General inertial proximal gradient method for a class of nonconvex nonsmooth optimization problems. *Computational Optimization and Applications*, 73 (1):129–158, 2019. page 105

Y. Xie, X. Wu, and R. Ward. Linear convergence of adaptive stochastic gradient descent. *arXiv preprint arXiv:1908.10525*, 2019. page 106

H. Xiong, T. Xu, Y. Liang, and W. Zhang. Non-asymptotic convergence of adam-type reinforcement learning algorithms under markovian sampling. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12):10460–10468, May 2021. URL https://ojs.aaai.org/index.php/AAAI/article/view/17252. page 172

P. Xu and Q. Gu. A finite-time analysis of q-learning with neural network function approximation. In *International Conference on Machine Learning*, pages 10555–10565. PMLR, 2020. page 172

T. Xu, S. Zou, and Y. Liang. Two time-scale off-policy td learning: Non-asymptotic analysis over markovian samples. In *Advances in Neural Information Processing Systems*, pages 10634–10644, 2019. page 131

T. Xu, Z. Wang, and Y. Liang. Improving sample complexity bounds for (natural) actor-critic algorithms. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 4358–4369. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/2e1b24a664f5e9c18f407b2f9c73e821-Paper.pdf. pages 131, 140

T. Xu, Z. Yang, Z. Wang, and Y. Liang. Doubly robust off-policy actor-critic: Convergence and optimality. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 11581–11591. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/xu21j.html. page 172

Y. Yan, T. Yang, Z. Li, Q. Lin, and Y. Yang. A unified analysis of stochastic momentum methods for deep learning. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 2955–2961, 2018. page 64

Z. Yang, K. Zhang, M. Hong, and T. Başar. A finite sample analysis of the actor-critic algorithm. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 2759–2764, 2018. doi: 10.1109/CDC.2018.8619440. page 131

Z. Yang, Z. Fu, K. Zhang, and Z. Wang. Convergent reinforcement learning with function approximation: A bilevel optimization perspective. 2019. URL https://openreview.net/forum?id=ryfcCo0ctQ. page 131

H. Yuan, X. Lian, J. Liu, and Y. Zhou. Stochastic recursive momentum for policy gradient methods. *arXiv preprint arXiv:2003.04302*, 2020. page 173

M. Zaheer, S. J. Reddi, D. Sachan, S. Kale, and S. Kumar. Adaptive methods for nonconvex optimization. In *Advances in Neural Information Processing Systems*, pages 9793–9803, 2018. pages 28, 64, 107, 110, 115

J. Zeng, T. T. Lau, S. Lin, and Y. Yao. Global convergence of block coordinate descent in deep learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 7313–7323, Long Beach, California, USA, 09–15 Jun 2019. PMLR. URL http://proceedings.mlr.press/v97/zeng19a.html. page 111

J. Zhang, T. He, S. Sra, and A. Jadbabaie. Why gradient clipping accelerates training: A theoretical justification for adaptivity. In *International Conference on Learning Representations*, 2019. page 107

K. Zhang, A. Koppel, H. Zhu, and T. Başar. Global convergence of policy gradient methods to (almost) locally optimal policies. *SIAM J. Control Optim.*, 58(6):3586–3612, 2020a. ISSN 0363-0129. doi: 10.1137/19M1288012. URL https://doi.org/10.1137/19M1288012. pages 133, 150

S. Zhang, B. Liu, H. Yao, and S. Whiteson. Provably convergent two-timescale off-policy actor-critic with function approximation. In H. D. III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 11204–11213, Virtual, 13–18 Jul 2020b. PMLR. pages 138, 151

S. Zhang, Y. Wan, R. S. Sutton, and S. Whiteson. Average-reward off-policy policy evaluation with function approximation. *arXiv preprint arXiv:2101.02808*, 2021a. page 172

S. Zhang, H. Yao, and S. Whiteson. Breaking the deadly triad with a target network. *ICML 2021, arXiv preprint arXiv:2101.08862*, 2021b. pages 15, 130, 131, 132, 135, 144, 147, 151, 172

D. Zhou, Y. Tang, Z. Yang, Y. Cao, and Q. Gu. On the convergence of adaptive gradient methods for nonconvex optimization. *arXiv preprint arXiv:1808.05671*, 2018. pages 27, 64, 106, 107, 115

Z. Zhou, Q. Zhang, G. Lu, H. Wang, W. Zhang, and Y. Yu. Adashift: Decorrelation and convergence of adaptive learning rate methods. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=HkgTkhRcKQ. page 107

F. Zou, L. Shen, Z. Jie, W. Zhang, and W. Liu. A sufficient condition for convergences of adam and rmsprop. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11127–11135, 2019a. pages 64, 107, 115

S. Zou, T. Xu, and Y. Liang. Finite-sample analysis for sarsa with linear function approximation. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019b. URL https://proceedings.neurips.cc/paper/2019/file/9f9e8cba3700df6a947a8cf91035ab84-Paper.pdf. pages 139, 154

ECOLE
DOCTORALE

Titre : Contributions à l'optimisation stochastique non convexe et à l'apprentissage par renforcement

Mots clés : Optimisation, méthodes à gradient adaptatives avec momentum, évitement de pièges, approximation stochastique, systèmes dynamiques, apprentissage par renforcement, méthodes acteur-critique

Résumé : Cette thèse est centrée autour de l'analyse de convergence de certains algorithmes d'approximation stochastiques utilisés en machine learning appliqués à l'optimisation et à l'apprentissage par renforcement. La première partie de la thèse est dédiée à un célèbre algorithme en apprentissage profond appelé ADAM, utilisé pour entraîner des réseaux de neurones. Cette célèbre variante de la descente de gradient stochastique est plus généralement utilisée pour la recherche d'un minimiseur local d'une fonction. En supposant que la fonction objective est différentiable et non convexe, nous établissons la convergence des itérées au temps long vers l'ensemble des points critiques sous une hypothèse de stabilité dans le régime des pas constants. Ensuite, nous introduisons une nouvelle variante de l'algorithme ADAM à pas décroissants. Nous montrons alors sous certaines hypothèses réalistes que les itérées sont presque sûrement bornées et convergent presque sûrement vers des points critiques de la fonction objective. Enfin, nous analysons les fluctuations de l'algorithme par le truchement d'un théorème central limite conditionnel. Dans la deuxième partie de cette thèse, dans le régime des pas décroissants, nous généralisons nos résultats de convergence et de fluctuations à une procédure d'optimisation stochastique unifiant plusieurs variantes de descente de gradient stochastique comme la méthode de la boule pesante, l'algorithme stochastique de Nesterov accéléré ou encore le célèbre algorithme ADAM, parmi d'autres. Nous concluons cette partie par un résultat d'évitement de pièges qui établit la non convergence de l'algorithme général vers des points critiques indésirables comme les maxima locaux ou les points-selles. Ici, le principal ingrédient est un nouveau résultat indépendant d'évitement de pièges pour un contexte non-autonome. Enfin, la dernière partie de cette thèse qui est indépendante des deux premières parties est dédiée à l'analyse d'un algorithme d'approximation stochastique pour l'apprentissage par renforcement. Dans cette dernière partie, dans le cadre des processus décisionnels de Markov avec critère de récompense γ-pondéré, nous proposons une analyse d'un algorithme acteur-critique en ligne intégrant un réseau cible et avec approximation de fonction linéraire. Notre algorithme utilise trois échelles de temps distinctes: une échelle pour l'acteur et deux autres pour la critique. Au lieu d'utiliser l'algorithme de différence temporelle (TD) standard à une échelle de temps, nous utilisons une version de l'algorithme TD à deux échelles de temps intégrant un réseau cible inspiré des algorithmes acteur-critique utilisés en pratique. Tout d'abord, nous établissons des résultats de convergence pour la critique et l'acteur sous échantillonage Markovien. Ensuite, nous menons une analyse à temps fini montrant l'impact de l'utilisation d'un réseau cible sur les méthodes acteur-critique.

Title : Contributions to non-convex stochastic optimization and reinforcement learning

Keywords : Optimization, adaptive gradient methods with momentum, avoidance of traps, stochastic approximation, dynamical systems, reinforcement learning, actor-critic methods

Abstract : This thesis is focused on the convergence analysis of some popular stochastic approximation methods in use in the machine learning community with applications to optimization and reinforcement learning. The first part of the thesis is devoted to a popular algorithm in deep learning called ADAM used for training neural networks. This variant of stochastic gradient descent is more generally useful for finding a local minimizer of a function. Assuming that the objective function is differentiable and non-convex, we establish the convergence of the iterates in the long run to the set of critical points under a stability condition in the constant stepsize regime. Then, we introduce a novel decreasing stepsize version of ADAM. Under mild assumptions, it is shown that the iterates are almost surely bounded and converge almost surely to critical points of the objective function. Finally, we analyze the fluctuations of the algorithm by means of a conditional central limit theorem. In the second part of the thesis, in the vanishing stepsizes regime, we generalize our convergence and fluctuations results to a stochastic optimization procedure unifying several variants of the stochastic gradient descent such as, among others, the stochastic heavy ball method, the Stochastic Nesterov Accelerated Gradient algorithm, and the widely used ADAM algorithm. We conclude this second part by an avoidance of traps result establishing the non-convergence of the general algorithm to undesired critical points, such as local maxima or saddle points. Here, the main ingredient is a new avoidance of traps result for non-autonomous settings, which is of independent interest. Finally, the last part of this thesis which is independent from the two previous parts, is concerned with the analysis of a stochastic approximation algorithm for reinforcement learning. In this last part, we propose an analysis of an online target-based actor-critic algorithm with linear function approximation in the discounted reward setting. Our algorithm uses three different timescales: one for the actor and two for the critic. Instead of using the standard single timescale temporal difference (TD) learning algorithm as a critic, we use a two timescales target-based version of TD learning closely inspired from practical actor-critic algorithms implementing target networks. First, we establish asymptotic convergence results for both the critic and the actor under Markovian sampling. Then, we provide a finite-time analysis showing the impact of incorporating a target network into actor-critic methods.

Institut Polytechnique de Paris
91120 Palaiseau, France