An SDE Perspective on Stochastic Inertial Gradient Dynamics with Time-Dependent Viscosity and Hessian-Driven Damping

Rodrigo Maulen-Soto * Jalal Fadili† Hedy Attouch‡ Peter Ochs§

Abstract. Our approach is part of the close link between continuous dissipative dynamical systems and optimization algorithms. Motivated by solving stochastic convex optimization problems in real Hilbert spaces, we propose a class of stochastic inertial differential equations which are driven by the gradient of the objective function. This will provide a general mathematical framework for analyzing fast optimization algorithms with stochastic gradient input. Our goal is to develop convergence guarantees for second-order stochastic differential equations in time, incorporating a viscous time-dependent damping and a Hessian-driven damping. To develop this program, we rely on stochastic Lyapunov analysis. Assuming a square-integrability condition on the diffusion term times a function dependent on the viscous damping, and that the Hessian-driven damping is a positive constant, our first main result shows that almost surely, there is convergence of the objective values with a fast convergence rate in expectation. Besides, in the case where the Hessian-driven damping is zero, we get fast convergence of the values in expectation and in almost sure sense, and we also prove almost sure weak convergence of the trajectory. We provide a comprehensive complexity analysis by establishing several new convergence rates in expectation and in almost sure sense for the convex and strongly convex case.

Key words. Stochastic optimization, Inertial gradient system, Convex optimization, Stochastic Differential Equation, Time-dependent viscosity, Convergence rate, Asymptotic behavior.

AMS subject classifications. 37N40, 46N10, 49M99, 65B99, 65K05, 65K10, 90B50, 90C25, 60H10, 90C53, 60G12

1 Introduction

1.1 Problem Statement

In this paper, we consider the minimization problem

$$\min_{x\in\mathbb{H}} f(x), \tag{P}$$

where $\mathbb{H}$ is a real Hilbert space and the objective function $f : \mathbb{H} \to \mathbb{R}$ satisfies the following standing assumptions:

$$\begin{cases} f \text{ is convex and continuously twice differentiable with } L\text{-Lipschitz continuous gradient;} \\ \mathcal{S} \overset{\text{def}}{=} \operatorname{argmin}(f) \neq \emptyset. \end{cases} \tag{H_0}$$

First-order in time systems. To solve (P), a fundamental dynamic is the gradient flow system:

$$\begin{cases} \dot{x}(t) + \nabla f(x(t)) = 0, \quad t > t_0; \\ x(t_0) = x_0. \end{cases} \tag{GF}$$

The gradient system (GF) is a dissipative dynamical system, whose study dates back to Cauchy [1] in finite dimensions. It is key in optimization, as it converts minimizing f into analyzing the asymptotic behavior of the trajectories of (GF).

*Normandie Université, ENSICAEN, UNICAEN, CNRS, GREYC, France. E-mail: rodrigo.maulen@ensicaen.fr
†Normandie Université, ENSICAEN, UNICAEN, CNRS, GREYC, France. E-mail: Jalal.Fadili@ensicaen.fr
‡IMAG, CNRS, Université Montpellier, France. E-mail: hedy.attouch@umontpellier.fr
§Department of Mathematics and Computer Science, Saarland University, Germany, E-mail: ochs@math.uni-saarland.de

In fact, Brezis, Baillon, and Bruck showed in the 1970s that if $\operatorname{argmin}(f)$ in (P) is non-empty, then every solution trajectory of (GF) converges weakly to a point in $\operatorname{argmin}(f)$, with a convergence rate of $\mathcal{O}(t^{-1})$ (even $o(t^{-1})$) on the function values.

The Euler forward (or Euler-Maruyama) discretization of (GF), with stepsizes $h_k > 0$, yields the gradient descent scheme

$$x_{k+1} = x_k - h_k \nabla f(x_k). \tag{GD}$$

Under assumption (H_0) and for $(h_k)k \in \mathbb{N} \subset]0, 2/L[$, we have $f(x_k) - \min f = \mathcal{O}(1/k)$ (even $o(1/k)$) and the iterates $(x_k)_{k\in\mathbb{N}}$ converge weakly to a point in $\operatorname{argmin}(f)$. This rate can be improved under additional geometric conditions on f, such as error bounds or the Kurdyka-Łojasiewicz property in the convex case [2].

Second-order in time systems: Key role of inertia. Second-order inertial dynamical systems have been introduced to accelerate convergence compared to (GF). An abundant literature has been devoted to the study of inertial dynamics

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla f(x(t)) = 0, \quad t > t_0. \tag{IGS$_\gamma$}$$

Several authors (e.g., [3, 4]) highlighted that an asymptotically vanishing viscosity coefficient $\gamma(t)$ is crucial for acceleration. Most of the literature focuses on the case $\gamma(t) = \frac{\alpha}{t}$, stemming from the seminal work [5] which established an $\mathcal{O}(1/t^2)$ convergence rate on the values for $\alpha = 3$, linking it to Nesterov's accelerated gradient method [6]. Further research shows that $\alpha \geq 3$ is required for the $\mathcal{O}(1/t^2)$ rate [7], while $\alpha > 3$ yields an improved $o(1/t^2)$ rate and weak convergence of the trajectory [8, 9]. Another remarkable case of (IGS$_\gamma$) is the Heavy Ball with Friction (HBF) method, where $\gamma(t)$ is constant, as introduced by Polyak [10]. In the strongly convex setting, the trajectory converges exponentially with an optimal rate for a well-chosen constant γ [11].

Geometric Hessian-driven damping. Because of the inertial aspects, and the asymptotic vanishing viscous damping coefficient, (IGS$_\gamma$) may exhibit many small oscillations which are not desirable from an optimization point of view. To remedy this, a powerful tool consists in introducing a geometric damping driven by the Hessian of f into the dynamic. This yields the Inertial System with Explicit Hessian-driven Damping, first proposed in [12] (and further studied in [13, 14, 15]):

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \beta(t)\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0, \tag{ISEHD}$$

and the Inertial System with Implicit Hessian-driven Damping [16, 17]:

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla f(x(t) + \beta(t)\dot{x}(t)) = 0. \tag{ISIHD}$$

The rationale behind the use of the term "implicit" in (ISIHD) comes from a Taylor expansion of the gradient term (as $t \to +\infty$ we expect $\dot{x}(t) \to 0$). Following the physical interpretation of these ODEs, we call the non-negative parameters γ and β as the viscous and geometric damping parameters, respectively.

At first glance, the presence of the Hessian in (ISEHD) may seem to entail numerical difficulties. However, this is not the case as the term $\nabla^2 f(x(t))\dot{x}(t)$ is nothing but the time derivative of $t \mapsto \nabla f(x(t))$. This explains why the time discretization of this dynamic provides efficient first-order algorithms [14]. However, in our stochastic setting—the focus of this paper— the proposed approach applies only to the implicit form of the Hessian-driven damping. In this context, we do not have direct access to ∇f, and instead model errors with a continuous Itô martingale $M(t)$. The explicit form would involve the derivative of $\nabla f(X(t)) + M(t)$, which is meaningless since non-constant martingales are almost surely non-differentiable. Hence, we will only consider (ISIHD).

1.2 Motivations

In many practical situations, the gradient evaluation is subject to stochastic errors—either because each iteration is costly, requiring cheap random approximations, or due to other exogenous factors. The continuous-time approach through stochastic differential equations (SDE) is a powerful way to model these errors in a unified way, and stochastic algorithms can then be viewed as time- discretizations. In fact, several recent works have used the SDE perspective to model SGD-type algorithms motivated by various reasons; (see *e.g.* [18, 19, 20, 21, 22, 23, 24, 25, 26, 27]). The

continuous-time perspective offers a deep insight and unveils the key properties of the dynamic without being tied to a specific discretization.

In this context, a natural SDE is

$$\begin{cases} dX(t) &= -\nabla f(X(t)) + \sigma(t, X(t))dW(t), \quad t > t_0; \\ X(t_0) &= X_0, \end{cases} \tag{SGF}$$

defined over a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq t_0}, \mathbb{P})$, where the diffusion (volatility) term $\sigma : \mathbb{R}^+ \times \mathbb{H} \to \mathcal{L}_2(\mathbb{K}; \mathbb{H})$ is a measurable function, $\mathbb{H}, \mathbb{K}$ are real separable Hilbert spaces, W is a $\mathcal{F}_t$-adapted $\mathbb{K}$-valued cylindrical Brownian motion, and the initial data X_0 is an $\mathcal{F}_{t_0}-$measurable $\mathbb{H}$-valued random variable. We will elaborate more on these notations and concepts in Section 2. (SGF) is a stochastic counterpart of (GF) where the error is modeled as a stochastic integral with respect to the measure defined by a continuous Itô martingale.

SDE modeling of SGD. To simplify the discussion in this section, we restrict our attention to the finite-dimensional case ($\mathbb{H} = \mathbb{R}^d$, $\mathbb{K} = \mathbb{R}^m$). In various applications, specially in machine learning, the Stochastic Gradient Descent (SGD) is a powerful alternative to gradient descent, and consists in replacing the full gradient computation by a cheaper random version, serving as an unbiased estimator. The SGD updates the iterates according to

$$x_{k+1} = x_k - h(\nabla f(x_k) + e_k), \tag{SGD}$$

where $h > 0$ is the stepsize and e_k is the random noise term on the gradient at the k-th iteration. As such, (SGD) can be viewed as instance of the Robbins-Monro stochastic approximation algorithm [28]. When the objective takes the form $f(x) = \mathbb{E}[\hat{f}(x, \xi)]$, where the expectation is w.r.t. to the random variable ξ the single-batch version of SGD reads

$$x_{k+1} = x_k - h\nabla \hat{f}(x_k, \xi_k), \tag{SGD$_{\text{SB}}$}$$

where $(\xi_k)_{k\in\mathbb{N}}$ are i.i.d. random variables with the same distribution as ξ. Of course, (SGD$_{\text{SB}}$) is an instance of (SGD) with $e_k = \nabla \hat{f}(x_k, \xi_k) - \nabla f(x_k)$.

A natural continuous-time model for these stochastic algorithms is (SGF). In particular, when the noise e_k in (SGD) is Gaussian (i.e., $e_k \sim \mathcal{N}(0, \sigma_k I_d)$), it has been shown that (SGF) accurately approximates (SGD) [26, Proposition 2.1]. Similar approximations hold for (SGD$_{\text{SB}}$) under appropriate conditions, see *e.g.* [29, 18, 19, 20, 21, 22, 23, 30, 31, 32].

Overall, the continuous-time SDE framework provides a powerful tool for analyzing SGD-type algorithms. By modeling the dynamics with (SGF), one can exploit the rich theoretical foundations of SDEs, Itô calculus, and measure theory to uncover the essential properties of an algorithm. This perspective not only captures the underlying behavior of the dynamics in a way that is independent of a particular discretization, but it also enables the transfer of convergence results from the continuous model directly to the discrete setting. In essence, many stability and convergence properties that are proven for (SGF) can be used to predict and explain the behavior of the corresponding discrete algorithms, including their ability to escape saddle points in non-convex scenarios. For a more detailed discussion of these aspects, we refer the reader to [33].

Second-order SDE modeling of inertial SGD. Using a lifting argument to get an equivalent first-order reformulation, a natural reformulation of (ISIHD) yields the differential equation

$$\begin{cases} \dot{x}(t) = v(t), \quad t > t_0; \\ \dot{v}(t) = -[\gamma(t)v(t) + \nabla f(x(t) + \beta(t)v(t))], \quad t > t_0; \\ x(t_0) = x_0, \quad \dot{x}(t_0) = v_0. \end{cases} \tag{ISIHD$_{\text{R}}$}$$

In this setting, we can model the associated errors using a stochastic integral with respect to the measure defined by a continuous Itô martingale. This entails the following stochastic differential equation (SDE for short), which is the stochastic counterpart of (ISIHD$_{\text{R}}$):

$$\begin{cases} dX(t) &= V(t)dt, \\ dV(t) &= -\gamma(t)V(t)dt - \nabla f(X(t) + \beta(t)V(t))dt + \sigma(t, X(t) + \beta(t)V(t))dW(t), \\ X(t_0) &= X_0, \quad V(t_0) = V_0. \end{cases} \tag{S $-$ ISIHD}$$

Extending what we have discussed above for (SGF) as a good model of (SGD), it can be shown that (S − ISIHD) is a provably good model of the stochastic inertial algorithm

$$\begin{cases} X_{k+1} &= X_k + hV_k, \\ V_{k+1} &= (1-\gamma_k h)V_k - h\nabla f(X_k + \beta_k V_k) + \sqrt{h}\sigma_k G_k, \end{cases} \tag{1.1}$$

where $G_k \sim \mathcal{N}(0, I_d)$, $\gamma_k = \gamma(t_k)$, $\sigma_k = \sigma(t_k, X_k + \beta_k V_k)$ and $\beta_k = \beta(t_k)$. More precisely, this algorithm is obtained by simple Euler-Maruyama discretization of (S − ISIHD). It has been shown in [33, Appendix A] that (1.1) is a consistent discretization of (S − ISIHD) with an approximation error that vanishes at the rate $\mathcal{O}(h)$. This is much better than the approximation rate for (ISIHD) which is only $\mathcal{O}(\sqrt{h})$. As a consequence, this justifies that the continuous-time dynamics (S − ISIHD) is a better proxy of the stochastic inertial algorithm (1.1) and opens the door to new insights in the behavior of such algorithms. In turn, the convergence properties of such algorithms can be easily derived from those of (S − ISIHD) with minimal effort. For instance, when f is smooth with Lipschitz-continuous gradient, it is easy to see from the descent lemma that

$$\mathbb{E}[f(X_k) - \min f] = \mathbb{E}[f(X(kh)) - \min f] + \mathcal{O}(h),$$

where X_k are the iterates of (1.1) and $X(t)$ the solution to (S − ISIHD). This means that any rate proved on $\mathbb{E}[f(X(t)) - \min f]$ can be directly transferred to $\mathbb{E}[f(X_k) - \min f]$.

1.3 Objectives and Contributions

In this work, our goal is to provide a general mathematical framework for analyzing fast gradient-based optimization algorithms with stochastic gradient input. For this, we will study second-order stochastic differential equations in time featuring inertia that is crucial for acceleration, and whose drift term is the gradient of the function to be minimized. In this context, considering inertial dynamics combining a time-dependent viscosity coefficient and geometric damping is a key property to obtain fast convergent methods while reducing oscillations.

More precisely, we study the stochastic dynamics (S − ISIHD) and its long-time behavior in order to solve (P). We conduct a new analysis using specific and careful arguments that are much more involved than in the deterministic case. To get some intuition, keeping the discussion informal at this stage, let us first identify the assumptions needed to expect that the position state of (S − ISIHD) "approaches" $\operatorname{argmin}(f)$ in the long run. In the case where $\mathbb{H} = \mathbb{K}$, $\gamma(\cdot) \equiv \gamma > 0$, $\beta \equiv 0$, and $\sigma = \tilde{\sigma} I_{\mathbb{H}}$, where $\tilde{\sigma}$ is a positive real constant. Under mild assumptions one can show that (S − ISIHD) has a unique invariant distribution $\pi_{\tilde{\sigma}}$ in (x, v) with density proportional to $\exp\left(-\frac{2\gamma}{\tilde{\sigma}^2}\left(f(x) + \frac{\|v\|^2}{2}\right)\right)$, see e.g., [34, Proposition 6.1]. Clearly, as $\tilde{\sigma} \to 0^+$, $\pi_{\tilde{\sigma}}$ gets concentrated around $\operatorname{argmin}(f) \times \{0_{\mathbb{H}}\}$, with $\lim_{\tilde{\sigma}\to 0^+} \pi_{\tilde{\sigma}}(\operatorname{argmin}(f) \times \{0_{\mathbb{H}}\}) = 1$; see also Section 1.4 for further discussion. Motivated by these observations and the fact that we aim to exactly solve (P), our paper will then mainly focus on the case where $\sigma(\cdot, x)$ vanishes fast enough as $t \to +\infty$ uniformly in x.

Our main contributions are summarized as follows:

- We will develop a Lyapunov analysis to obtain convergence rates and integral estimates, in expectation and almost sure sense, in the general case of coefficients $\gamma(t)$ and $\beta(t)$.
- We will study two instances where the rates can be made even more explicit highlighting acceleration phenomena: when $\gamma(t) = \frac{\alpha}{t}, \beta(t) = \beta_0 + \frac{\gamma_0}{t}$, and when γ is decreasing and vanishing with vanishing derivative, and $\beta(t)$ is constant.
- In the case where the coefficient $\beta(t)$ is zero, we show that under some hypotheses, we have almost sure weak convergence of the trajectory, convergence rates, and integral estimates. As a special case, we focus on viscous damping coefficient $\gamma(t) = \frac{\alpha}{t^r}, r \in [0, 1], \alpha \geq 1 - r$.

1.4 Relation to prior work

Kinetic diffusion dynamics for sampling Let us consider (S − ISIHD) in the case where $\mathbb{H} = \mathbb{K} = \mathbb{R}^d$, $\gamma(t) = \gamma > 0, \beta(t) = 0$ and $\sigma = \sqrt{2\gamma} I_d$. Then one recovers the kinetic Langevin diffusion (or second-order Langevin process). In this case, the continuous-time Markov process $(X(t), V(t))$ is positive recurrent and has a unique invariant distribution

which has the density proportional to $\exp\left(-f(x) - \frac{\|v\|^2}{2}\right)$ with respect to the Lebesgue measure on $\mathbb{R}^{2d}$. Time-discretized versions of this Langevin diffusion process have been studied in the literature to (approximately) sample from a distribution proportional to $\exp(-f(x))$ with asymptotic and non-asymptotic convergence guarantees in various topologies and under various conditions have been studied; see [35, 36, 37] and references therein.

Inexact inertial gradient systems There is an abundant literature regarding the dynamics (ISIHD) and (ISEHD), either in the exact case or with errors but only deterministic ones; see [16, 38, 39, 40, 41, 7, 42, 43, 44, 45, 46, 47]). We are not aware of any such work in the stochastic case. On a technical level, our Lyapunov analysis will be inspired by that in [38] and [48]. Only a few papers have been devoted to studying the second-order in-time inertial stochastic gradient systems with viscous damping, *i.e.* stochastic versions of (IGS$_\gamma$), either with vanishing damping $\gamma(t) = \alpha/t$ or constant damping $\gamma(t)$ (stochastic HBF); see *e.g.* [26, 49, 50]. For instance, [26] provide asymptotic $\mathcal{O}(1/t^2)$ convergence rate on the objective values in expectation under integrability conditions on the diffusion term as well as other rates under additional geometrical properties of of the objective.The corresponding stochastic algorithms for these two choices of γ, whose mathematical formulation and analysis is simpler, have been the subject of active research work; see *e.g.* [51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64].

Time scaling and averaging A first-order SDE to solve (P) has been thoroughly studied in [25]; see also [27] for the non-smooth setting. This SDE has the form

$$\begin{cases} dX(t) = -\nabla f(X(t))dt + \sigma(t, X(t))dW(t), \quad t \geq t_0, \\ X(t_0) = X_0. \end{cases} \tag{1.2}$$

Our work here is a natural extension of [25, 27] to second-order systems hence allowing for accelerated rates. By leveraging the time scaling and averaging trick pioneered in [65] to go from (GF) to (ISIHD), we were able in [33] to transfer the results in [25] for (1.2) to (S − ISIHD). The approach in [33] avoids, in particular, going through an intricate Lyapunov analysis for (S − ISIHD). However, this can only be done for a specific choice of β (related to the viscous damping function γ). On the other hand, getting full advantage of the Hessian-driven damping term requires a careful choice of β, that is possibly independent of γ. Thus handling flexible and general choices of β γ makes necessary to go through a dedicated Lyapunov analysis for (S − ISIHD).

2 Notation and Preliminaries

We will use the following shorthand notations: Given $n \in \mathbb{N}$, $[n] \overset{\text{def}}{=} \{1, \dots, n\}$. Consider $\mathbb{H}, \mathbb{K}$ real separable Hilbert spaces endowed with the inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ and $\langle \cdot, \cdot \rangle_{\mathbb{K}}$, respectively, and norm $\|\cdot\|_{\mathbb{H}} = \sqrt{\langle \cdot, \cdot \rangle_{\mathbb{H}}}$ and $\|\cdot\|_{\mathbb{K}} = \sqrt{\langle \cdot, \cdot \rangle_{\mathbb{K}}}$, respectively (we omit the subscripts $\mathbb{H}$ and $\mathbb{K}$ for the sake of clarity). $I_{\mathbb{H}}$ is the identity operator on $\mathbb{H}$. $\mathcal{L}(\mathbb{K}; \mathbb{H})$ is the space of bounded linear operators from $\mathbb{K}$ to $\mathbb{H}$, $\mathcal{L}_1(\mathbb{K})$ is the space of trace-class operators, and $\mathcal{L}_2(\mathbb{K}; \mathbb{H})$ is the space of bounded linear Hilbert-Schmidt operators from $\mathbb{K}$ to $\mathbb{H}$. For $M \in \mathcal{L}_1(\mathbb{K})$, is trace is defined by

$$\mathrm{tr}(M) \overset{\text{def}}{=} \sum_{i \in I} \langle M e_i, e_i \rangle < +\infty,$$

where $I \subseteq \mathbb{N}$ and $(e_i)_{i \in I}$ is an orthonormal basis of $\mathbb{K}$. Besides, for $M \in \mathcal{L}(\mathbb{K}; \mathbb{H})$, $M^\star \in \mathcal{L}(\mathbb{H}; \mathbb{K})$ is the adjoint operator of M, and for $M \in \mathcal{L}_2(\mathbb{K}; \mathbb{H})$,

$$\|M\|_{\mathrm{HS}} \overset{\text{def}}{=} \sqrt{\mathrm{tr}(MM^\star)} < +\infty$$

is its Hilbert-Schmidt norm (in the finite-dimensional case is equivalent to the Frobenius norm). We denote by w-lim the limit for the weak topology of $\mathbb{H}$. The notation $A : \mathbb{H} \rightrightarrows \mathbb{H}$ means that A is a set-valued operator from $\mathbb{H}$ to $\mathbb{H}$. Consider $f : \mathbb{H} \to \mathbb{R}$, the sublevel of f at height $r \in \mathbb{R}$ is denoted $[f \leq r] \overset{\text{def}}{=} \{x \in \mathbb{H} : f(x) \leq r\}$. For $1 \leq p \leq +\infty$, $\mathrm{L}^p([a, b])$ is the space of measurable functions $g : \mathbb{R} \to \mathbb{R}$ such that $\int_a^b |g(t)|^p dt < +\infty$, with the usual adaptation when $p = +\infty$. For two functions $f, g : \mathbb{R} \to \mathbb{R}$ we will denote $f \sim g$ as $t \to +\infty$, if $\lim_{t \to +\infty} \frac{f(t)}{g(t)} = 1$. On the

probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathrm{L}^p(\Omega; \mathbb{H})$ denotes the (Bochner) space of $\mathbb{H}$-valued random variables whose p-th moment (with respect to the measure $\mathbb{P}$) is finite. Other notations will be explained when they first appear.

Let us recall some important definitions and results from convex analysis; for a comprehensive coverage, we refer the reader to [66].

We denote by $C^s(\mathbb{H})$ the class of s-times continuously differentiable functions on $\mathbb{H}$. For $L \geq 0$, $C_L^{1,1}(\mathbb{H}) \subset C^1(\mathbb{H})$ is the set of functions on $\mathbb{H}$ whose gradient is L-Lipschitz continuous, and $C_L^2(\mathbb{H})$ is the subset of $C_L^{1,1}(\mathbb{H})$ whose functions are twice differentiable.

The class of $C_L^{1,1}(\mathbb{H})$ functions enjoys the well-known *descent lemma* which plays a central role in the analysis of optimization dynamics.

Lemma 2.1. *Let $f \in C_L^{1,1}(\mathbb{H})$, then*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2, \quad \forall x, y \in \mathbb{H}.$$

Corollary 2.2. *Let $f \in C_L^{1,1}(\mathbb{H})$ such that $\operatorname{argmin} f \neq \emptyset$, then*

$$\|\nabla f(x)\|^2 \leq 2L(f(x) - \min f), \quad \forall x \in \mathbb{H}.$$

On stochastic differential equations For the necessary notation and preliminaries on stochastic processes, see [27, Section A.2]. Moreover, the existence and uniqueness of a solution of (S − ISIHD) is discussed in Proposition A.9.

Let us now present Itô's formula which plays a central role in the theory of stochastic differential equations:

Proposition 2.3. *Consider (X, V) a solution of (S − ISIHD) and W a $\mathbb{K}-$valued cylindrical Brownian motion, let $\phi : [t_0, +\infty[\times \mathbb{H} \times \mathbb{H} \to \mathbb{R}$ be such that $\phi(\cdot, x, v) \in C^1([t_0, +\infty[)$ for every $x, v \in \mathbb{H}$, $\phi(t, \cdot, \cdot) \in C^2(\mathbb{H} \times \mathbb{H})$ for every $t \geq t_0$. Then the process*

$$Y(t) = \phi(t, X(t), V(t)),$$

is an Itô Process, such that for all $t \geq t_0$

$$\begin{aligned} Y(t) = Y(t_0) &+ \int_{t_0}^t \frac{\partial \phi}{\partial t}(s, X(s), V(s))ds + \int_{t_0}^t \langle \nabla_x \phi(s, X(s), V(s)), V(s) \rangle ds \\ &- \int_{t_0}^t \langle \nabla_v \phi(s, X(s), V(s)), \gamma(s)V(s) + \nabla f(X(s) + \beta(s)V(s)) \rangle ds \\ &+ \int_{t_0}^t \langle \sigma^\star(s, X(s) + \beta(s)V(s)) \nabla_v \phi(s, X(s), V(s)), dW(s) \rangle \\ &+ \frac{1}{2} \int_{t_0}^t \operatorname{tr}[\sigma(s, X(s) + \beta(s)V(s)) \sigma^\star(s, X(s) + \beta(s)V(s)) \nabla_v^2 \phi(s, X(s), V(s))] ds, \end{aligned} \tag{2.1}$$

where ∇_v^2 is the Hessian with respect to the double differentiation of v and $\sigma^\star$ is the adjoint operator of σ. Moreover, if for all $T > t_0$

$$\mathbb{E}\left(\int_{t_0}^T \|\sigma^\star(s, X(s) + \beta(s)V(s)) \nabla_v \phi(s, X(s), V(s))\|^2 ds \right) < +\infty,$$

then $\int_{t_0}^t \langle \sigma^\star(s, X(s) + \beta(s)V(s)) \nabla_v \phi(s, X(s), V(s)), dW(s) \rangle$ is a square-integrable continuous martingale and

$$\mathbb{E}\left(\int_{t_0}^t \langle \sigma^\star(s, X(s) + \beta(s)V(s)) \nabla_v \phi(s, X(s), V(s)), dW(s) \rangle \right) = 0 \tag{2.2}$$

Proof. Let $Z \in \mathbb{H} \times \mathbb{H}$, $\Psi : \mathbb{R}_+ \times \mathbb{H} \times \mathbb{H} \to \mathbb{H} \times H$, and $\Phi(t, Z) : \mathbb{R}_+ \times \mathbb{H} \times \mathbb{H} \to \mathcal{M}_{2\times 2}(\mathcal{L}_2(\mathbb{K}; \mathbb{H}))$ such that

$$Z = \begin{pmatrix} X \\ V \end{pmatrix}, \Psi(t, Z) = \begin{pmatrix} V \\ -\gamma(t)V - \nabla f(X + \beta(t)V) \end{pmatrix}, \Phi(t, Z) = \begin{pmatrix} 0 & 0 \\ 0 & \sigma(t, X + \beta(t)V) \end{pmatrix}.$$

Let also $W(t) = (W_1(t), W_2(t))$, where W_1, W_2 are two independent $\mathbb{K}-$valued cylindrical Brownian motions. Then, we can write (S − ISIHD) as

$$\begin{cases} Z(t) &= \int_{t_0}^t \Psi(t, Z(t))dt + \int_{t_0}^t \Phi(t, Z(t))dW(t). \\ Z(t_0) &= (X_0, V_0). \end{cases} \tag{2.3}$$

The hypotheses let us use [67, Section 2.3.2] to obtain

$$\begin{aligned} Y(t) = \phi(t, Z(t)) = Y(t_0) &+ \int_{t_0}^t \frac{\partial \phi}{\partial t}(s, Z(s))ds + \int_{t_0}^t \langle \nabla_z \phi(s, Z(s)), \Psi(s, Z(s)) \rangle_{\mathbb{H}\times\mathbb{H}} ds \\ &+ \int_{t_0}^t \langle \nabla_z \phi(s, Z(s)), \Phi(s, Z(s))dW(s) \rangle_{\mathbb{H}\times\mathbb{H}} \\ &+ \frac{1}{2} \int_{t_0}^t \mathrm{tr}[\nabla_z^2 \phi(s, Z(s))\Phi(s, Z(s))\Phi^\star(s, Z(s))]ds, \end{aligned}$$

where ∇_z and ∇_z^2 denote the gradient and Hessian matrix with respect to the variable z, corresponding to first- and second-order derivatives, respectively. We obtain the desired result by replacing the actual values of Ψ and Φ and using the usual inner product of $\mathbb{H} \times \mathbb{H}$:

$$\langle (x_1, x_2), (y_1, y_2) \rangle_{\mathbb{H}\times\mathbb{H}} \overset{\text{def}}{=} \langle x_1, y_1 \rangle_{\mathbb{H}} + \langle x_2, y_2 \rangle_{\mathbb{H}}.$$

□

3 General γ and β

In this section, we will develop a Lyapunov analysis based on [38] to study almost sure, and in expectation properties of the dynamic (S − ISIHD), when the parameters γ and β are general functions. This will allow to go much further and consider parameters not covered in [33] which exploits the relationship between first-order and second-order systems. We will also apply our results to two special cases: (i) γ is a differentiable, decreasing and vanishing function, with vanishing derivative, and β is a positive constant; and (ii) $\gamma(t) = \frac{\alpha}{t}$, and $\beta(t) = \gamma_0 + \frac{\beta}{t}$ (with $\gamma_0, \beta > 0$). These cases are again are not covered by results in [33].

Recall that our focus in this paper is on an optimization perspective, and as we argued in the introduction, we will study the long time behavior of (S − ISIHD) as the diffusion term vanishes when $t \to +\infty$. Therefore, throughout the paper, we assume that the diffusion (volatility) term σ satisfies:

$$\begin{cases} \sup_{t\geq t_0, x\in\mathbb{H}} \|\sigma(t, x)\|_{\mathrm{HS}} < +\infty, \\ \|\sigma(t, x') - \sigma(t, x)\|_{\mathrm{HS}} \leq l_0 \|x' - x\|, \end{cases} \tag{H_σ}$$

for some $l_0 > 0$ and for all $t \geq t_0, x, x' \in \mathbb{H}$. The Lipschitz continuity assumption is mild and required to ensure the well-posedness of (S − ISIHD).

Remark 3.1. Under the hypothesis (H_σ) we have that there exists $\sigma_*^2 > 0$ such that

$$\|\sigma(t, x)\|_{\mathrm{HS}}^2 = \mathrm{tr}[\Sigma(t, x)] \leq \sigma_*^2, \quad \forall t \geq t_0, \forall x \in \mathbb{H},$$

where $\Sigma \overset{\text{def}}{=} \sigma\sigma^\star$. Let us also define $\sigma_\infty : [t_0, +\infty[\to \mathbb{R}_+$ as: $\sigma_\infty(t) \overset{\text{def}}{=} \sup_{x\in\mathbb{H}} \|\sigma(t, x)\|_{\mathrm{HS}}$.

Now, we follow with the hypotheses we will require over the damping parameters.

For $t_0 > 0$, let $\gamma : [t_0, +\infty[\to \mathbb{R}_+$ be a viscous damping, we denote

$$p(t) \overset{\text{def}}{=} \exp\left(\int_{t_0}^{t} \gamma(s)ds\right). \tag{3.1}$$

Besides, if

$$\int_{t_0}^{+\infty} \frac{ds}{p(s)} < +\infty, \tag{H_γ}$$

we define $\Gamma : [t_0, +\infty[\to \mathbb{R}_+$ by

$$\Gamma(t) \overset{\text{def}}{=} p(t) \int_{t}^{+\infty} \frac{ds}{p(s)}. \tag{3.2}$$

Remark 3.2. Let us notice that Γ satisfies the relation

$$\Gamma' = \gamma\Gamma - 1.$$

For $t_0 > 0$, let $\beta : [t_0, +\infty[\to \mathbb{R}_+$ be a geometric damping that we will assume to be a differentiable function. We will occasionally need to impose the additional assumption that there exists $c_1, c_2 > 0$, and $t_1 > t_0$ such that for every $t \geq t_1$:

$$\begin{cases} \beta(t) & \leq c_1; \\ \left|\frac{\beta'(t) - \gamma(t)\beta(t) + 1}{\beta(t)}\right| & \leq c_2. \end{cases} \tag{H_β}$$

We recall also that $\mathcal{S} \overset{\text{def}}{=} \operatorname{argmin}(f)$.

3.1 Reformulation of (S − ISIHD)

The formulation of the dynamic (S − ISIHD) is known as the Hamiltonian formulation. However, it is not the only one. In the deterministic case, an alternative equivalent and more flexible first-order reformulation of (ISIHD) was proposed in [38]. The motivation there was that this equivalent reformulation can handle the case where f is non-smooth. Although we will not consider the non-smooth case here, we will still extend and use that equivalent reformulation to the stochastic case.

Consider the dynamic (S − ISIHD), and let us define the auxiliary variable

$$Y(t) = X(t) + \beta(t)V(t), \quad t > t_0.$$

We have that

$$\begin{aligned} dY(t) &= dX(t) + \beta'(t)V(t) + \beta(t)dV(t) \\ &= -\beta(t)\nabla f(Y(t))dt - (\beta'(t) - \gamma(t)\beta(t) + 1)\left(\frac{X(t) - Y(t)}{\beta(t)}\right)dt + \beta(t)\sigma(t, Y(t))dW(t). \end{aligned}$$

So we can reformulate (S − ISIHD) in terms of X, Y in the following way:

$$\begin{cases} dX(t) &= -\left(\frac{X(t)-Y(t)}{\beta(t)}\right)dt, \quad t > t_0, \\ dY(t) &= -\beta(t)\nabla f(Y(t))dt - (\beta'(t) - \gamma(t)\beta(t) + 1)\left(\frac{X(t)-Y(t)}{\beta(t)}\right)dt + \beta(t)\sigma(t, Y(t))dW(t), \quad t > t_0, \\ X(t_0) &= X_0, \quad Y(t_0) = X_0 + \beta(t_0)V_0, \end{cases} \tag{ISIHD − S_R}$$

where the subscript 'R' indicates that this is a reformulation. Moreover, we can reformulate (ISIHD $-$ $\mathrm{S_R}$) in the product space $\mathbb{H}\times\mathbb{H}$ by setting $Z(t)=(X(t),Y(t))\in\mathbb{H}\times\mathbb{H}$, and thus (ISIHD $-$ $\mathrm{S_R}$) can be equivalently written as

$$\begin{cases} dZ(t) &= -\beta(t)\nabla\mathcal{G}(Z(t))dt - \mathcal{D}(t,Z(t))dt + \hat{\sigma}(t,Z(t))dW(t), \quad t>t_0,\\ Z(t_0) &= (X_0, X_0+\beta(t_0)V_0), \end{cases} \tag{3.3}$$

where $\mathcal{G}:\mathbb{H}\times\mathbb{H}\to\mathbb{R}$ is the convex function defined as $\mathcal{G}(Z)=f(Y)$, and the time-dependent operator $\mathcal{D}:[t_0,+\infty[\times\mathbb{H}\times\mathbb{H}\to\mathbb{H}\times\mathbb{H}$ is given by

$$\mathcal{D}(t,Z)=\left(\frac{1}{\beta(t)}(X-Y),\frac{\beta'(t)-\gamma(t)\beta(t)+1}{\beta(t)}(X-Y)\right), \tag{3.4}$$

and the stochastic noise $\hat{\sigma}\in\mathcal{M}_{2\times2}(\mathcal{L}_2(\mathbb{K};\mathbb{H}))$ defined by $\hat{\sigma}(t,Z)=\begin{pmatrix}0 & 0\\ 0 & \beta(t)\sigma(t,Y)\end{pmatrix}$, and $W(t)=(W_1(t),W_2(t))$, where W_1,W_2 are two independent $\mathbb{K}-$valued cylindrical Brownian motions.

3.2 Fast convergence properties: convex case

To obtain properties in almost sure sense and in expectation of (S $-$ ISIHD), we are going to adapt the Lyapunov analysis shown on [38] for the dynamic (ISIHD).

To that purpose, let us consider $t_1>t_0$, $\gamma,\beta:[t_0,+\infty[\to\mathbb{R}_+$ be fixed functions and let $a,b,c,d:[t_0,+\infty[\to\mathbb{R}$ be differentiable functions (on $]t_0,+\infty[$) satisfying the following system for all $t>t_1$:

$$\begin{cases} a'(t)-b(t)c(t) &\leq 0\\ -a(t)\beta(t) &\leq 0\\ -a(t)\gamma(t)\beta(t)+a(t)\beta'(t)+a(t)-c(t)^2+b(t)c(t)\beta(t) &= 0\\ b'(t)b(t)+\frac{d'(t)}{2} &\leq 0\\ b'(t)c(t)+b(t)(b(t)+c'(t)-c(t)\gamma(t))+d(t) &= 0\\ c(t)(b(t)+c'(t)-c(t)\gamma(t)) &\leq 0. \end{cases} \tag{$\mathrm{S}_{a,b,c,d}$}$$

Given $x^\star\in\mathcal{S}$, we consider

$$\mathcal{E}(t,x,v)=a(t)(f(x+\beta(t)v)-\min(f))+\frac{1}{2}\|b(t)(x-x^\star)+c(t)v\|^2+\frac{d(t)}{2}\|x-x^\star\|^2. \tag{3.5}$$

Remark 3.3. It was shown in [16, Section 3.1] and [38, Lemma 1] that energy function $\mathcal{E}$ with a,b,c,d satisfying the system ($\mathrm{S}_{a,b,c,d}$) is a Lyapunov function for (ISIHD) when $\gamma(t)=\frac{\alpha}{t}$ (with $\alpha>3$) and $\beta(t)=\gamma_0+\frac{\beta}{t}$ (with $\gamma_0,\beta\geq0$), hence, useful to obtain convergence guarantees of that dynamic. We will see that the same system ($\mathrm{S}_{a,b,c,d}$) also covers the case of general coefficients γ and β, hence providing insights on the convergence properties of (S $-$ ISIHD) when one can find the corresponding functions a,b,c,d.

In the following proposition, we state an abstract integral bound, as well as almost sure and in expectation convergence properties for (S $-$ ISIHD).

Proposition 3.4. *Assume that f,σ satisfy (H_0) and (H_σ), respectively. Let $\nu\geq2$, and consider the dynamic (S $-$ ISIHD) with initial data $X_0,V_0\in\mathrm{L}^\nu(\Omega;\mathbb{H})$. Consider also γ,β from (S $-$ ISIHD) satisfying (H_γ) and (H_β). Then, there exists a unique solution $(X,V)\in S^\nu_{\mathbb{H}\times\mathbb{H}}[t_0]$ of (S $-$ ISIHD). Moreover, if there exist functions a,b,c,d satisfying ($\mathrm{S}_{a,b,c,d}$), and $t\mapsto m(t)\sigma^2_\infty(t)\in\mathrm{L}^1([t_0,+\infty[)$, where $m(t)\overset{\text{def}}{=}\max\{1,a(t),c^2(t)\}$, then the following statements hold:*

(i) If $b(t)c(t)-a'(t)=\mathcal{O}(c(t)(\gamma(t)c(t)-c'(t)-b(t)))$, then

$$\int_{t_0}^{+\infty}(b(s)c(s)-a'(s))(f(X(s))-\min f+\|V(s)\|^2)ds<+\infty\quad a.s..$$

(ii) If there exists $\eta > 0, \hat{t} > t_0$ *such that*

$$\eta \leq c(t)(\gamma(t)c(t) - c'(t) - b(t)), \quad \eta \leq a(t)\beta(t), \quad \gamma(t) \leq \eta, \quad \forall t > \hat{t},$$

then $\lim_{t\to+\infty} \|V(t)\| = 0$ *a.s.,* $\lim_{t\to+\infty} \|\nabla f(X(t) + \beta(t)V(t))\| = 0$ *a.s., and* $\lim_{t\to+\infty} \|\nabla f(X(t))\| = 0$ *a.s.*

(iii) If there exists $D > 0, \tilde{t} > t_0$ *such that* $d(t) \geq D$ *for* $t > \tilde{t}$*, then :*

$$\mathbb{E}\left(\|V(t)\|^2\right) = \mathcal{O}\left(\frac{1 + b^2(t)}{c^2(t)}\right),$$

and

$$\mathbb{E}\left(f(X(t)) - \min f\right) = \mathcal{O}\left(\max\left\{\frac{1}{a(t)}, \frac{\beta(t)\sqrt{1 + b^2(t)}}{\sqrt{a(t)}c(t)}, \frac{\beta^2(t)\left(1 + b^2(t)\right)}{c^2(t)}\right\}\right).$$

This is a compact version that extracts only the most important points from the more detailed and complete Propositions A.13, A.14 which are proved in the appendix.

The complete version of the previous proposition (*i.e.* Propositions A.13 and A.14) generalizes the results poved in [38] to the stochastic setting. However, they lack practical use if we cannot exhibit a, b, c, d functions that satisfy ($S_{a,b,c,d}$). Although we are not able to solve this system in general, in Corollaries 3.5 and 3.7 we will specify some particular cases for γ and β where such functions a, b, c, d can be exhibited to satisfy the system ($S_{a,b,c,d}$).

The following corollary provides a specific case where a solution to the system ($S_{a,b,c,d}$) can be exhibited, which was not discussed in [16, 38]. Moreover, we show the implications it has on the stochastic setting.

Corollary 3.5 (Decreasing and vanishing γ, with vanishing γ' and positive constant β). *Consider the context of Proposition 3.4 in the case where* $\beta(t) \equiv \beta > 0$*,* γ *satisfying* (H_γ)*, such that it is a differentiable, decreasing, and vanishing function, with* $\lim_{t\to+\infty} \gamma'(t) = 0$*, and satisfying that:*

$$\text{there exists } t_2 \geq t_0 \text{ and } m < \frac{3}{2} \text{ such that } \gamma(t)\Gamma(t) \leq m \text{ for every } t \geq t_2. \tag{H'_γ}$$

Let $b \in]2(m-1), 1[$*, then choosing*

$$\begin{aligned} a(t) &= \frac{\Gamma(t)(\Gamma(t) - \beta b)}{1 - \beta\gamma(t)}, \\ b(t) &= b, \\ c(t) &= \Gamma(t), \\ d(t) &= b(1-b), \end{aligned}$$

there exists $\hat{t} > t_0$ *such that the system* ($S_{a,b,c,d}$) *is satisfied for every* $t \geq \hat{t}$*.*

Given $x^\star \in \mathcal{S}$ *and* σ_∞ *be such that* $t \mapsto \Gamma(t)\sigma_\infty(t) \in \mathrm{L}^2([t_0, +\infty[)$*, then the following statements hold:*

(i) $\int_{t_0}^{+\infty} \Gamma(s)\left(f(X(s)) - \min f + \|V(s)\|^2\right) ds < +\infty$ *a.s..*

(ii) $\lim_{t\to+\infty} \|\nabla f(X(t))\| + \|V(t)\| = 0$ *a.s..*

(iii) $\mathbb{E}(f(X(t)) - \min f + \|V(t)\|^2) = \mathcal{O}\left(\frac{1}{\Gamma^2(t)}\right)$.

Remark 3.6. When $\gamma(t) = \frac{\alpha}{t}$ with $\alpha > 3$ and $t\sigma_\infty(t) \in \mathrm{L}^2([t_0, +\infty[)$, the previous corollary ensures fast convergence of the values, *i.e.*, $\mathcal{O}(t^{-2})$. Besides, by Corollary A.7, when $\gamma(t) = \frac{\alpha}{t^r}$ with $r \in]0, 1[, \alpha \geq 1 - r$, and $t^r\sigma_\infty(t) \in \mathrm{L}^2([t_0, +\infty[)$, the previous corollary ensures convergence of the objective at a rate $\mathcal{O}\left(t^{-2r}\right)$. The latter choice indicates that one can require a weaker integrability condition on the noise, compared to the case $\gamma(t) = \frac{\alpha}{t}$ ($\alpha > 3$), but at the price of a slower convergence rate.

Proof. We start by noticing that since γ is decreasing, by [48, Corollary 2.3] we have that $\Gamma(t)$ is increasing and $\gamma(t)\Gamma(t) \geq 1$, for every $t \geq t_0$. Also, it is direct that with a fixed $\beta > 0$ we satisfy (H_β).

Letting $b \in]2(m-1), 1[$ and $t_1 > t_0$ such that $\beta \leq \frac{1}{\gamma(t_1)}$, this t_1 exists since $t \mapsto \frac{1}{\gamma(t)}$ is an increasing function. We choose $c(t) = \Gamma(t)$, by the fifth equation of ($S_{a,b,c,d}$), we get that $d = b(1-b)$, and the fourth equation is trivial. The third equation implies that $a(t) = \frac{\Gamma(t)(\Gamma(t)-b\beta)}{1-\beta\gamma(t)}$ and the choice of β implies that the second equation is satisfied for $t \geq t_1$, since $\beta \leq \frac{1}{\gamma(t_1)} \leq \frac{1}{\gamma(t)} \leq \Gamma(t)$ for every $t > t_1$. By the definition of $c(t)$ and the fact that $b < 1$, we directly have that the sixth equation also holds. We just need to check the first equation, to do so, we can see that this equation is equivalent to

$$\frac{\Gamma'(t)(2\Gamma(t) - \beta b)(1-\beta\gamma(t)) + \beta\Gamma(t)(\Gamma(t) - \beta b)\gamma'(t)}{(1-\beta\gamma(t))^2} \leq b\Gamma(t),$$

by multiplying by $(1-\beta\gamma(t))^2$ at both sides and developing the terms, we get the previous inequality is equivalent to

$$\begin{aligned} 2\Gamma(t)\Gamma'(t) - \beta b\Gamma'(t) - 2\beta\gamma(t)\Gamma(t)\Gamma'(t) + b\beta^2\gamma(t)\Gamma'(t) + \beta\Gamma^2(t)\gamma'(t) - b\beta^2\Gamma(t)\gamma'(t) \\ \leq b\Gamma(t) - 2b\beta\gamma(t)\Gamma(t) + b\beta^2\gamma^2(t)\Gamma(t). \quad (3.6) \end{aligned}$$

By (H'_γ), there exists $m > \frac{3}{2}$ and $t_2 > t_0$ such that $1 \leq \gamma(t)\Gamma(t) \leq m$ and $0 \leq \Gamma'(t) \leq m-1$ for every $t \geq t_2$. Since the terms

$$-2\beta\gamma(t)\Gamma(t)\Gamma'(t), \quad -\beta b\Gamma'(t), \quad \beta\Gamma^2(t)\gamma'(t)$$

are negative, and $b\beta^2\gamma^2(t)\Gamma(t)$ is positive, thus if we could prove there exists $t_3 > \max\{t_0, t_1, t_2\}$ such that for every $t \geq t_3$

$$2\Gamma(t)\Gamma'(t) + b\beta^2\gamma(t)\Gamma'(t) - b\beta^2\Gamma(t)\gamma'(t) \leq b\Gamma(t) - 2b\beta\gamma(t)\Gamma(t), \quad (3.7)$$

we could conclude that (3.6) holds for every $t \geq t_3$. Rearranging the terms in (3.7), we get

$$b\beta^2\gamma(t)\Gamma'(t) + 2b\beta\gamma(t)\Gamma(t) \leq \Gamma(t)[b\beta^2\gamma'(t) + b - 2\Gamma'(t)], \quad (3.8)$$

and we note that the terms $b\beta^2\gamma(t)\Gamma'(t), 2b\beta\gamma(t)\Gamma(t)$ are upper bounded by a constant. Since $\lim_{t\to+\infty}\gamma(t) = 0$, we get that $\lim_{t\to+\infty}\Gamma(t) = +\infty$, therefore if we could prove that there exists $t_3 > \max\{t_0, t_1, t_2\}$ such that

$$-b\beta^2\gamma'(t) < \delta < b - 2(m-1) \leq b - 2\Gamma'(t)$$

for $t \geq t_3$, this would imply that there exists $\hat{t} \geq t_3$ such that (3.8) holds, which in turn would imply that (3.6) holds for every $t \geq \hat{t}$. In fact, we see that the previous inequality holds for t large enough (*i.e.* there exists such a t_3) since $\lim_{t\to+\infty} -\gamma'(t) = 0$ and the fact that $0 < b - 2(m-1)$ implies that there exists δ such that $0 < \delta < b - 2(m-1)$. Thus, we have checked that the proposed a, b, c, d satisfy the system ($S_{a,b,c,d}$) for $t > \hat{t}$.

The rest of the proof is direct by replacing the specified $a, b, c, d, \gamma, \beta$ functions in Proposition 3.4, and the fact that for t large enough, $b\Gamma(t) - a'(t) \geq (b - 2(m-1) - \delta)\Gamma(t) - C_b$ for some $C_b > 0$, that $\lim_{t\to+\infty}\Gamma(t) = +\infty$, and also that $a(t) = \mathcal{O}(\Gamma^2(t)), \Gamma^2(t) = \mathcal{O}(a(t))$. Since $\lim_{t\to+\infty}\Gamma(t) = +\infty, \lim_{t\to+\infty} a(t) = +\infty$, and $\lim_{t\to+\infty}\gamma(t) = 0$, the parameter $\eta > 0$ in Proposition 3.4 can be chosen arbitrarily.

□

The following result gives us another case in which we can satisfy the system ($S_{a,b,c,d}$). This generalizes to the stochastic setting the results presented in [16, Section 3.1] and [38, Lemma 1]. Besides, it ensures fast convergence of the values whenever $t \mapsto t\sigma_\infty(t) \in \mathrm{L}^2([t_0, +\infty[)$.

Corollary 3.7 ($\gamma(t) = \frac{\alpha}{t}$ **and** $\beta(t) = \gamma_0 + \frac{\beta}{t}$**).** *Consider the context of Proposition 3.4 in the case where* $\gamma(t) = \frac{\alpha}{t}$ *and* $\beta(t) = \gamma_0 + \frac{\beta}{t}$*, where* $\alpha > 3, \gamma_0 > 0, \beta \geq 0$*. Then choosing*

$$\begin{aligned} a(t) &= t^2\left(1 + \frac{(\alpha - b)\gamma_0 t - \beta(\alpha + 1 - b)}{t^2 - \alpha\gamma_0 t - \beta(\alpha+1)}\right), \\ b(t) &= b \in (2, \alpha - 1), \\ c(t) &= t, \\ d(t) &= b(\alpha - 1 - b), \end{aligned}$$

the system ($S_{a,b,c,d}$) *is satisfied.*

Given $x^\star \in \mathcal{S}$ *and* σ_∞ *be such that* $t \mapsto t\sigma_\infty(t) \in \mathrm{L}^2([t_0, +\infty[)$*, then the following statements hold:*

(i) $\int_{t_0}^{+\infty} s\left(f(X(s)) - \min f + \|V(s)\|^2\right) ds < +\infty$ *a.s..*

(ii) $\lim_{t\to+\infty} \|\nabla f(X(t))\| + \|V(t)\| = 0$ *a.s..*

(iii) $\mathbb{E}(f(X(t)) - \min f + \|V(t)\|^2) = \mathcal{O}\left(\frac{1}{t^2}\right)$.

Proof. Direct from replacing the specified $a, b, c, d, \gamma, \beta$ functions in Proposition 3.4, and the fact that for t large enough $bt - a'(t) \geq \frac{(\alpha-3)t}{2}$, and also $a(t) \geq t^2, 0 < \gamma_0 < \beta(t) \leq \gamma_0 + \frac{\beta}{t_0}$. □

Remark 3.8. We can use the choices for a, b, c, d presented in Corollaries 3.5 and 3.7 in Propositions A.13 and A.14 to obtain additional integral bounds, almost sure and in expectation properties of ($S - ISIHD$). We leave this to the reader.

3.3 Strongly convex case

In the following theorem, we consider the case where the objective function is strongly convex and we present a choice of parameters γ and β to obtain a fast linear convergence rate when the noise vanishes at a proper rate.

Theorem 3.9. *Assume that* $f : \mathbb{H} \to \mathbb{R}$ *satisfies* (H_0)*, and is* μ*-strongly convex,* $\mu > 0$*, and denote* $x^\star$ *its unique minimizer. Suppose also that* σ *obeys* (H_σ)*. Let* $\nu \geq 2$*, consider the dynamic* ($S - ISIHD$) *with initial data* $X_0 \in \mathrm{L}^\nu(\Omega; \mathbb{H})$*. Consider also* $\gamma \equiv 2\sqrt{\mu}$*, and a constant* β *such that* $0 \leq \beta \leq \frac{1}{2\sqrt{\mu}}$*. Moreover, suppose that* σ_∞ *is a non-increasing function such that* $\sigma_\infty \in \mathrm{L}^2([t_0, +\infty[)$*. Define the function* $\mathcal{E} : [t_0, +\infty[\times\mathbb{H}\times\mathbb{H} \to \mathbb{R}_+$ *as*

$$\mathcal{E}(t, x, v) \overset{\text{def}}{=} f(x + \beta v) - \min f + \frac{1}{2}\|\sqrt{\mu}(x - x^\star) + v\|^2.$$

Then, ($S - ISIHD$) *has a unique solution* $(X, V) \in S^\nu_{\mathbb{H}\times\mathbb{H}}[t_0]$*. In addition, there exists positive constants* M_1, M_2 *such that*

$$\mathbb{E}[\mathcal{E}(t, X(t), V(t))] \leq \mathcal{E}(t_0, X_0, V_0)e^{-\frac{\sqrt{\mu}}{2}(t-t_0)} + M_1 e^{-\frac{\sqrt{\mu}}{4}(t-t_0)} + M_2\sigma_\infty^2\left(\frac{t_0 + t}{2}\right), \quad \forall t > t_0.$$

Let $\Theta : [t_0, +\infty[\to \mathbb{R}_+$ *defined as* $\Theta(t) \overset{\text{def}}{=} \max\{e^{-\frac{\sqrt{\mu}}{4}(t-t_0)}, \sigma_\infty^2\left(\frac{t+t_0}{2}\right)\}$*. Consequently,*

$$\begin{aligned}
\mathbb{E}(f(X(t)) - \min f) &= \mathcal{O}(\Theta(t)),\\
\mathbb{E}(\|X(t)) - x^\star\|^2) &= \mathcal{O}(\Theta(t)),\\
\mathbb{E}(\|V(t)\|^2) &= \mathcal{O}(\Theta(t)),\\
\mathbb{E}(\|\nabla f(X(t))\|^2) &= \mathcal{O}(\Theta(t)).
\end{aligned}$$

Proof. Using Itô's formula with $\mathcal{E}$, taking expectation and denoting $E(t) \overset{\text{def}}{=} \mathbb{E}(\mathcal{E}(t, X(t), V(t)))$, we have

$$E(t) \leq E(t_0) - \int_{t_0}^t \frac{\sqrt{\mu}}{2}E(s)ds - \int_{t_0}^t C(s)ds + (L\beta^2 + 1)\int_{t_0}^t \sigma_\infty^2(s)ds,$$

where

$$\begin{aligned}
C(t) \overset{\text{def}}{=}\ & \beta\|\nabla f(X(t) + \beta V(t))\|^2 + \beta\sqrt{\mu}\langle\nabla f(X(t) + \beta V(t)), V(t)\rangle + \frac{\sqrt{\mu}}{2}(\beta^2\mu + 1)\|V(t)\|^2\\
&+ \beta\mu\sqrt{\mu}\langle X(t) - x^\star, V(t)\rangle + \frac{\mu\sqrt{\mu}}{4}\|X(t) - x^\star + \beta V(t)\|^2.
\end{aligned}$$

It was proved in [38, Theorem 4.2] that under the condition $0 \leq \beta \leq \frac{1}{2\sqrt{\mu}}$ we obtain that $C(t)$ is a non-negative function. Therefore, we can write the following

$$E(t) \leq E(t_0) - \int_{t_0}^t \frac{\sqrt{\mu}}{2}E(s)ds + (L\beta^2 + 1)\int_{t_0}^t \sigma_\infty^2(s)ds.$$

We continue by using [27, Lemma A.2], to do this, we need to solve the following Cauchy problem:

$$\begin{cases} Y'(t) & = -\frac{\sqrt{\mu}}{2}Y(t) + (L\beta^2+1)\sigma_\infty^2(t) \\ Y(t_0) & = \mathcal{E}(t_0, X_0, V_0). \end{cases}$$

Using the integrating factor method, we deduce that for all $t \geq t_0$:

$$\begin{aligned} Y(t) &= Y(t_0)e^{\frac{\sqrt{\mu}}{2}(t_0-t)} + (L\beta^2+1)e^{-\frac{\sqrt{\mu}}{2}t}\int_{t_0}^{t} e^{\frac{\sqrt{\mu}}{2}s}\sigma_\infty^2(s)ds \\ &\leq Y(t_0)e^{\frac{\sqrt{\mu}}{2}(t_0-t)} + (L\beta^2+1)e^{-\frac{\sqrt{\mu}}{2}t}\left(\int_{t_0}^{\frac{t_0+t}{2}} e^{\frac{\sqrt{\mu}}{2}s}\sigma_\infty^2(s)ds + \int_{\frac{t_0+t}{2}}^{t} e^{\frac{\sqrt{\mu}}{2}s}\sigma_\infty^2(s)ds\right) \\ &\leq Y(t_0)e^{\frac{\sqrt{\mu}}{2}(t_0-t)} + \frac{2(L\beta^2+1)}{\sqrt{\mu}}\sigma_\infty^2\left(\frac{t_0+t}{2}\right) + \frac{2(L\beta^2+1)}{\sqrt{\mu}}\sigma_\infty^2(t_0)e^{\frac{\sqrt{\mu}}{4}(t_0-t)} \\ &= \mathcal{O}(\Theta(t)). \end{aligned}$$

By [27, Lemma A.2], we conclude that $E(t) = \mathcal{O}(\Theta(t))$, immediately we observe that

$$\begin{aligned} \mathbb{E}(f(X(t)+\beta V(t)) - \min f) &= \mathcal{O}(\Theta(t)) \\ \mathbb{E}(\|\sqrt{\mu}(X(t)-x^\star)+V(t)\|^2) &= \mathcal{O}(\Theta(t)) \end{aligned}$$

By the strong convexity of f, we have that $\mathbb{E}(\|X(t)-x^\star+\beta V(t)\|^2) = \mathcal{O}(\Theta(t))$, since $\beta \neq \frac{1}{\sqrt{\mu}}$ ($\beta \leq \frac{1}{2\sqrt{\mu}}$), then $\mathbb{E}(\|X(t)-x^\star\|^2) = \mathbb{E}(\|V(t)\|^2) = \mathcal{O}(\Theta(t))$, on the other hand, using Lemma 2.1 and Lemma 2.2, $\mathbb{E}(f(X(t)) - f(X(t)+\beta V(t))) = \mathcal{O}(\Theta(t))$, thus,

$$\mathbb{E}(f(X(t)) - \min f) = \mathbb{E}(\|\nabla f(X(t))\|^2) = \mathcal{O}(\Theta(t)).$$

□

4 General γ and $\beta \equiv 0$

In this section we are going to study properties of the dynamic (S − ISIHD) in expectation and in almost sure sense, when the parameter γ is a general function and $\beta \equiv 0$. The noiseless case and under deterministic noise is well documented in [3].

Consider the dynamic (S − ISIHD) when $\beta \equiv 0$. This dynamic will be a stochastic version of the Hamiltonian formulation of (IGS$_\gamma$) and it will be described by:

$$\begin{cases} dX(t) & = V(t)dt, \ t > t_0, \\ dV(t) & = -\gamma(t)V(t)dt - \nabla f(X(t))dt + \sigma(t, X(t))dW(t), \ t > t_0, \\ X(t_0) & = X_0, \quad V(t_0) = V_0. \end{cases} \tag{IGS$_\gamma$ − S}$$

The main motivation for a separate analysis is that, in Section 3 we consider hypothesis (H$_\beta$) to establish the existence and uniqueness of a solution, from which, the rest of the results follow. This hypothesis is incompatible with the case $\beta \equiv 0$.

We will demonstrate almost sure convergence of the velocity to zero and of the objective to its minimum value, under assumptions that are satisfied for $\gamma(t) = \frac{\alpha}{t^r}$, with $r \in [0,1], \alpha \geq 1-r$. Additionally, we will show that for this particular choice of β, we can obtain almost sure (weak) convergence of the trajectory.

4.1 Minimization properties

Let us define for $c > 0$,

$$\lambda_c(t) = \frac{p(t)}{c + \int_{t_0}^t p(s)ds}.$$

We can deduce that $\lambda_c' + \lambda_c^2 - \gamma\lambda_c = 0$, besides, since $p \notin \mathrm{L}^1([t_0, +\infty[)$, then $\lambda_c \notin \mathrm{L}^1([t_0, +\infty[)$.

Theorem 4.1. *Assume that f and σ satisfy assumptions (H_0) and (H_σ), respectively. Let $\nu \geq 2$, and consider the dynamic ($\mathrm{IGS}_\gamma - \mathrm{S}$) with initial data $X_0, V_0 \in \mathrm{L}^\nu(\Omega; \mathbb{H})$. Then, there exists a unique solution $(X, V) \in S^\nu_{\mathbb{H}\times\mathbb{H}}[t_0]$ of ($\mathrm{IGS}_\gamma - \mathrm{S}$). Additionally, if $\sigma_\infty \in \mathrm{L}^2([t_0, +\infty[)$, then*

$$\int_{t_0}^{+\infty} \gamma(s)\|V(s)\|^2 ds < +\infty \quad a.s..$$

Moreover, suppose that

$$\text{there exists } \hat{t} \geq t_0, \text{ and } c > 0 \text{ such that } \gamma(t) \leq \lambda_c(t)\ \forall t \geq \hat{t}, \tag{H_a}$$

and

$$\int_{t_0}^{+\infty} \lambda_c(s)\|V(s)\|^2 ds < +\infty \quad a.s.. \tag{H_b}$$

Then the following properties are satisfied:

(i) $\int_{t_0}^{+\infty} \lambda_c(s)(f(X(s)) - \min f)ds < +\infty$ *a.s..*

(ii) $\lim_{t\to+\infty} \|V(t)\| = 0$ *a.s. and* $\lim_{t\to+\infty} f(X(t)) - \min f = 0$ *a.s..*

Proof. The existence and uniqueness of a solution of ($\mathrm{IGS}_\gamma - \mathrm{S}$) is a direct consequence of [27, Theorem 3.3] in the product space $\mathbb{H} \times \mathbb{H}$.

Let $x^\star \in \mathcal{S}$ and $\phi_0 : (x, v) \mapsto \mathbb{R}$ defined by $\phi_0(x, v) = f(x) - \min f + \frac{\|v\|^2}{2}$, by Itô's formula and Theorem A.12 we obtain that $\int_{t_0}^{+\infty} \gamma(s)\|V(s)\|^2 ds < +\infty$ a.s.. Moreover, if we assume the hypotheses (H_a) and (H_b), then:

(i) Let $x^\star \in \mathcal{S}$ and $\phi : (t, x, v) \mapsto \mathbb{R}$ defined by $\phi(t, x, v) = \frac{\|\lambda_c(t)(x-x^\star)+v\|^2}{2} + (f(x) - \min f)$. Let $\hat{t}$ defined in the statement, by Itô's formula from $\hat{t}$ to t, we have

$$\begin{aligned}
&f(X(t)) - \min f + \frac{\|\lambda_c(t)(X(t) - x^\star) + V(t)\|^2}{2} = f(X(\hat{t})) - \min f + \frac{\|\lambda_c(\hat{t})(X(\hat{t}) - x^\star) + V(\hat{t})\|^2}{2} \\
&+ \int_{\hat{t}}^t \lambda_c(s)\lambda_c'(s)\|X(s)) - x^\star\|^2 - \gamma(t)\|V(s)\|^2 - \lambda_c(t)\langle\nabla f(X(s)), X(s) - x^\star\rangle]ds \\
&+ \int_{\hat{t}}^t \lambda_c(t)\|V(s)\|^2 + \mathrm{tr}[\Sigma(s, X(s))]ds + \underbrace{\int_{\hat{t}}^t \langle[\lambda_c(s)(X(s) - x^\star) + V(s)]\sigma^\star(s, X(s)), dW(s)\rangle}_{M_t}.
\end{aligned}$$

By the hypotheses, we have that

$$\int_{\hat{t}}^{+\infty} \left(\lambda_c(s)\|V(s)\|^2 + \mathrm{tr}[\Sigma(s, X(s))]\right) ds \leq \int_{\hat{t}}^{+\infty} \left(\lambda_c(s)\|V(s)\|^2 + \sigma_\infty^2(s)\right) ds < +\infty \quad a.s..$$

Besides $(M_t)_{t\geq\hat{t}}$ is a continuous martingale. Moreover, by convexity of f and the fact that $\lambda_c'(t) \leq 0\ \forall t \geq \hat{t}$,

$$\begin{aligned}
&\int_{\hat{t}}^t \lambda_c(s)\lambda_c'(s)\|X(s)) - x^\star\|^2 - \gamma(t)\|V(s)\|^2 - \lambda_c(t)\langle\nabla f(X(s)), X(s) - x^\star\rangle]ds \\
&\leq -\int_{\hat{t}}^t \lambda_c(s)(f(X(s)) - \min f)ds.
\end{aligned}$$

Then, by Theorem A.12,

$$\int_{\hat{t}}^{+\infty} \lambda_c(s)(f(X(s)) - \min f)ds < +\infty \quad a.s., \tag{4.1}$$

and $\lim_{t\to+\infty} f(X(t)) - \min f + \frac{\|\lambda_c(t)(X(t)-x^\star)+V(t)\|^2}{2}$ exists a.s..

(ii) By Lemma A.2 and (4.1), we conclude that $\lim_{t\to+\infty} \frac{\|V(t)\|^2}{2} + f(X(t)) - \min f = 0$ a.s..

□

Corollary 4.2. *Consider the context of Theorem 4.1 with $\gamma(t) = \frac{\alpha}{t^r}$, where $r \in [0,1]$ and $\alpha > 1 - r$. Then (H_a) and (H_b) are satisfied and thus the conclusions of Theorem 4.1 hold.*

Proof.

- We will prove the case $r = 1$ first, since it is direct, in such case, letting $c = \frac{t_0}{\alpha+1}$ we have that $\lambda_c(t) = \frac{\alpha+1}{t}$, which satisfies ($H_a$), moreover, $\lambda(t) = \frac{\alpha+1}{\alpha}\gamma(t)$, so ($H_b$) is also satisfied.
- Let $r \in]0,1[$, $c = \frac{\int_0^{t_0} e^{\alpha s^{1-r}} ds}{e^{\alpha t_0^{1-r}}}$. Instead of proving $\gamma(t) \leq \lambda_c(t)$, we will prove the equivalent inequality $\frac{1}{t\lambda_c(t)} \leq \frac{1}{t\gamma(t)}$. In fact, by a change of variable we have that (see notation of I_p in Lemma A.8):

$$\frac{1}{\lambda_c(t)t} = \frac{(1-r)^{\frac{r}{1-r}}}{\alpha^{\frac{1}{1-r}}} I_{\frac{r}{1-r}}\left(\frac{\alpha}{1-r}t^{1-r}\right),$$

Moreover, by the first result of Lemma A.8 we have that

$$\frac{1}{t\lambda_c(t)} \leq \left(\frac{1-r}{\alpha}\right)^{\frac{1}{1-r}} \frac{t^{r-1}}{\alpha} \leq \frac{1}{t\gamma(t)}.$$

where the last inequality comes from the fact that $1 - r \leq \alpha$. Moreover, by the second result of Lemma A.8, we obtain that:

$$\left(\frac{\alpha}{1-r}\right)^{\frac{1}{1-r}} \frac{1}{t\lambda_c(t)} \sim \frac{1}{t\gamma(t)}, \quad \text{as } t \to +\infty.$$

This implies that for every $\varepsilon \in]0,1[$ there exists $\hat{t} > t_0, \Lambda_\varepsilon \geq 1$ such that $\lambda_c(t) \leq \Lambda_\varepsilon\gamma(t)$ for every $t > \hat{t}$ $\left(\Lambda_\varepsilon = \left(\frac{\alpha}{1-r}\right)^{\frac{1}{1-r}} \frac{1}{(1-\varepsilon)}\right)$, this implies ($H_b$).

□

Remark 4.3. Finding all (or at least a larger class of) continuous functions γ that satisfy (H_a) and for which one can prove (H_b) in general is an open problem.

4.2 Tighter convergence rates of the values

In order to illustrate the context of the following result, it is useful to mention that if $\gamma(t) = \frac{\alpha}{t}$, then Theorem 4.1 gives us minimization properties in the case $\alpha > 0$. However, as mentioned in the introduction, it is widely known in the continuous deterministic setting (IGS_γ), with $\gamma(t) = \frac{\alpha}{t}$ and $\alpha > 3$, then the values converge at the rate $o(1/t^2)$ (see [3, 8]). Based on [3], we will depict that effect for a general γ in the continuous stochastic setting.

We will rephrase assumption (H_0) on the objective f to:

$$\begin{cases} f \text{ is convex and continuously differentiable with } L\text{-Lipschitz continuous gradient;} \\ f \in C^2(\mathbb{H}) \text{ or } \mathbb{H} \text{ is finite-dimensional;} \\ \mathcal{S} \overset{\text{def}}{=} \operatorname{argmin}(f) \neq \emptyset. \end{cases} \tag{$H_0^\star$}$$

($H_0^\star$) coincides with (H_0) in the infinite-dimensional case, but is weaker than (H_0) when $\mathbb{H}$ is finite-dimensional.

Theorem 4.4. *Assume that f, σ and γ satisfy assumptions ($H_0^\star$), (H_σ) and (H_γ)-(H'_γ), respectively. Let $\nu \geq 2$, and consider the dynamic ($IGS_\gamma - S$) with initial data $X_0, V_0 \in L^\nu(\Omega; \mathbb{H})$. Then, there exists a unique solution $(X, V) \in S^\nu_{\mathbb{H}\times\mathbb{H}}[t_0]$ of ($IGS_\gamma - S$), for every $\nu \geq 2$. Additionally, if $\Gamma\sigma_\infty \in L^2([t_0, +\infty[)$, then:*

(i) $\int_{t_2}^{+\infty} \Gamma(t)(f(X(t)) - \min f + \|V(t)\|^2)dt < +\infty$ *a.s..*

(ii) $f(X(t)) - \min f + \|V(t)\|^2 = o\left(\frac{1}{\Gamma^2(t)}\right)$ *a.s..*

(iii) $\mathbb{E}(f(X(t)) - \min f + \|V(t)\|^2) = \mathcal{O}\left(\frac{1}{\Gamma^2(t)}\right)$.

Moreover, assume that $\Gamma \notin L^1([t_0, +\infty[)$, and let $\theta(t) \overset{\text{def}}{=} \int_{t_0}^t \Gamma(s)ds$. If also $\theta\sigma_\infty^2 \in L^1([t_0, +\infty[)$, then:

(iv) $f(X(t)) - \min f + \|V(t)\|^2 = o\left(\frac{1}{\theta(t)}\right)$ *a.s..*

(v) $\mathbb{E}(f(X(t)) - \min f + \|V(t)\|^2) = \mathcal{O}\left(\min\left\{\frac{\int_{t_0}^t \frac{ds}{\Gamma(s)}}{\theta(t)}, \frac{1}{\Gamma^2(t)}\right\}\right)$.

Remark 4.5. The claim (ii) is new even in the deterministic case (*i.e.* $\sigma(\cdot,\cdot) = 0_{\mathbb{K};\mathbb{H}}$). According to the first three items of the previous theorem, the conclusions of Remark 3.6 are also valid, this is that when $\gamma(t) = \frac{\alpha}{t}$ with $\alpha > 3$ and $t\sigma_\infty(t) \in L^2([t_0, +\infty[)$, the previous theorem ensures fast convergence of the values, *i.e.*, $\mathcal{O}(t^{-2})$ in expectation and $o(t^{-2})$ in almost sure sense. Besides, by Corollary A.7, when $\gamma(t) = \frac{\alpha}{t^r}$ with $r \in]0, 1[$, $\alpha \geq 1 - r$, and $t^r\sigma_\infty(t) \in L^2([t_0, +\infty[)$, the previous theorem ensures convergence of the objective at a rate $\mathcal{O}\left(t^{-2r}\right)$ in expectation and $o\left(t^{-2r}\right)$ in almost sure sense, moreover by ((iv)) the convergence rate is actually $o\left(t^{-(r+1)}\right)$ in almost sure sense, which is faster than $o\left(t^{-2r}\right)$. Regarding ((iv)), this can be seen as the extension of [3, Theorem 3.6] to the stochastic setting.

Proof. (i) Let $m < \frac{3}{2}$ and t_2 defined in (H'_γ), let also $b \in]2(m-1), 1[$ and $x^\star \in \mathcal{S}$. Based on ($S_{a,b,c,d}$) with $\beta \equiv 0$, we introduce $\phi_1 : (t, x, v) \mapsto \mathbb{R}$ defined by

$$\phi_1(t, x, v) = \Gamma^2(t)(f(x) - \min f) + \frac{\|b(x - x^\star) + \Gamma(t)v\|^2}{2} + \frac{b(1-b)}{2}\|x - x^\star\|^2.$$

Since $f \in C^2(\mathbb{H})$, we use Itô's formula from t_2 to t to get

$$\begin{aligned}\phi_1(t, X(t), V(t)) = \phi_1(t_2, X(t_2), V(t_2)) &+ \int_{t_2}^t \Gamma(s)[2\Gamma'(s)(f(X(s)) - \min f) - b\langle\nabla f(X(s)), X(s) - x^\star\rangle]ds \\ &+ (b-1)\int_{t_2}^t \Gamma(s)\|V(s)\|^2 ds + \int_{t_2}^t \Gamma^2(s)\mathrm{tr}[\Sigma(s, X(s))]ds \\ &+ \underbrace{\int_{t_2}^t \langle[\Gamma^2(s)V(s) + b\Gamma(s)(X(s) - x^\star)]\sigma^\star(s, X(s)), dW(s)\rangle}_{M_t}.\end{aligned} \tag{4.2}$$

When $\mathbb{H}$ is finite-dimensional but f is not $C^2(\mathbb{H})$, we can use mollifiers as in [24, Proposition C.2], and get (4.2) as an inequality in this case.

Besides, we have that

$$\int_{t_2}^{+\infty} \Gamma^2(s)\mathrm{tr}[\Sigma(s, X(s))]ds \leq \int_{t_2}^{+\infty} \Gamma^2(s)\sigma_\infty^2(s)ds < +\infty.$$

Besides $(M_t)_{t\geq t_2}$ is a continuous martingale. Moreover, by convexity of f, we have that

$$\begin{aligned}&\int_{t_2}^t \Gamma(s)[2\Gamma'(s)(f(X(s)) - \min f) - b\langle\nabla f(X(s)), X(s) - x^\star\rangle]ds \\ &\leq \int_{t_2}^t \Gamma(s)(2\Gamma'(s) - b)(f(X(s)) - \min f)ds.\end{aligned}$$

Since $b-1<0$, and

$$2\Gamma'(t)-b=2\gamma(t)\Gamma(t)-2-b\leq 2(m-1)-b<0,\quad \forall t>t_2.$$

By Theorem A.12,

$$\int_{t_2}^{+\infty}\Gamma(s)(f(X(s))-\min f+\|V(s)\|^2)ds<+\infty\quad a.s., \tag{4.3}$$

and

$$\lim_{t\to+\infty}\Gamma^2(t)(f(X(t))-\min f)+\frac{\|b(X(t)-x^\star)+\Gamma(t)V(t)\|^2}{2}+\frac{b(1-b)}{2}\|X(t)-x^\star\|^2 \text{ exists a.s..}$$

(ii) On the other hand, let $\phi_2:(t,x,v)\mapsto\mathbb{R}$ defined by $\phi_2(t,x,v)=\Gamma^2(t)\left(f(x)-\min f+\frac{\|v\|^2}{2}\right)$. Recalling the discussion for ϕ_1, we get that by Itô's formula from t_2 to t, we have

$$\begin{aligned}\phi_2(t,X(t),V(t))=\phi_2(t_2,X(t_2),V(t_2))&+\int_{t_2}^t 2\Gamma(s)\Gamma'(s)(f(X(s))-\min f)ds\\&-\int_{t_2}^t\Gamma(s)\|V(s)\|^2ds+\int_{t_2}^t\Gamma^2(s)\mathrm{tr}[\Sigma(s,X(s))]ds\\&+\underbrace{\int_{t_2}^t\Gamma^2(s)\langle V(s)\sigma^\star(s,X(s)),dW(s)\rangle}_{M_t}.\end{aligned} \tag{4.4}$$

And also, that

$$\begin{aligned}&\int_{t_2}^{+\infty}2\Gamma(s)\Gamma'(s)(f(X(s))-\min f)+\Gamma^2(s)\mathrm{tr}[\Sigma(s,X(s))]ds\\&\leq\int_{t_2}^{+\infty}\Gamma(s)(f(X(s))-\min f)+\Gamma^2(s)\sigma_\infty^2(s)ds<+\infty\quad a.s..\end{aligned}$$

Besides $(M_t)_{t\geq t_2}$ is a continuous martingale. By Theorem A.12, we get again that $\int_{t_2}^{+\infty}\Gamma(s)\|V(s)\|^2ds<+\infty$ a.s. and that

$$\lim_{t\to+\infty}\Gamma^2(t)\left(f(X(t))-\min f+\frac{\|V(t)\|^2}{2}\right)\text{ exists a.s.} \tag{4.5}$$

Let us recall that $\frac{1}{\Gamma}\notin\mathrm{L}^1([t_0,+\infty[)$ by Lemma A.3. Therefore, by (4.3) and (4.5), we can use Lemma A.2 to obtain that

$$\lim_{t\to+\infty}\Gamma^2(t)\left(f(X(t))-\min f+\frac{\|V(t)\|^2}{2}\right)=0\quad a.s..$$

(iii) Taking expectation on (4.2) and denoting

$$K_1\overset{\text{def}}{=}\Gamma^2(t_2)\mathbb{E}((f(X(t_2))-\min f))+\frac{1}{2}\mathbb{E}(\|b(X(t_2)-x^\star)+\Gamma(t_2)V(t_2)\|^2)+\frac{b(1-b)}{2}\|X(t_2)-x^\star\|^2,$$

$$K_\Gamma\overset{\text{def}}{=}\int_{t_2}^{+\infty}\Gamma^2(s)\sigma_\infty^2(s)ds,$$

we obtain directly that

$$\mathbb{E}\left(\Gamma^2(t)(f(X(t))-\min f)+\frac{\|b(X(t)-x^\star)+\Gamma(t)V(t)\|^2}{2}+\frac{b(1-b)}{2}\|X(t)-x^\star\|^2\right)\leq K_1+K_\Gamma.$$

From this, is direct that $\sup_{t\geq t_2}\mathbb{E}(\|X(t)-x^\star\|^2)<+\infty$, and this in turn imply

$$\mathbb{E}\left(f(X(t))-\min f+\frac{\|V(t)\|^2}{2}\right)=\mathcal{O}\left(\frac{1}{\Gamma^2(t)}\right).$$

(iv) Moreover, assume that $\Gamma \notin \mathrm{L}^1([t_0,+\infty[)$, and let $\theta(t) = \int_{t_0}^t \Gamma(s)ds$. If also $\theta\sigma_\infty^2 \in \mathrm{L}^1([t_0,+\infty[)$, then we consider $\phi_3(t,x,v) = \theta(t)\left(f(x) - \min f + \frac{\|v\|^2}{2}\right)$, by Itô's formula from t_2 to t, we get

$$\begin{aligned}\phi_3(t,X(t),V(t)) = \phi_3(t_2,X(t_2),V(t_2)) &+ \int_{t_2}^t \Gamma(s)\left(f(X(s)) - \min f + \frac{\|V(s)\|^2}{2}\right)ds \\ &- \int_{t_2}^t \gamma(s)\theta(s)\|V(s)\|^2 ds + \frac{1}{2}\int_{t_2}^t \theta(s)\mathrm{tr}[\Sigma(s,X(s))]ds \\ &+ \underbrace{\int_{t_2}^t \theta(s)\langle V(s)\sigma^\star(s,X(s)), dW(s)\rangle}_{M_t}.\end{aligned} \tag{4.6}$$

Also, by the first item and new hypothesis on the diffusion term, we get that

$$\begin{aligned}&\int_{t_2}^t \Gamma(s)\left(f(X(s)) - \min f + \frac{\|V(s)\|^2}{2}\right) + \theta(s)\mathrm{tr}[\Sigma(s,X(s))]ds \\ &\leq \int_{t_2}^{+\infty} \Gamma(s)\left(f(X(s)) - \min f + \frac{\|V(s)\|^2}{2}\right) + \theta(s)\sigma_\infty^2(s)ds < +\infty.\end{aligned} \tag{4.7}$$

Besides $(M_t)_{t\geq t_2}$ is a continuous martingale. By Theorem A.12, we get that $\int_{t_2}^{+\infty}\gamma(s)\theta(s)\|V(s)\|^2 ds < +\infty$ a.s. and that

$$\lim_{t\to+\infty}\theta(t)\left(f(X(t)) - \min f + \frac{\|V(t)\|^2}{2}\right) \text{ exists a.s.,} \tag{4.8}$$

Using Lemma A.4 with $q(t) = \theta(t)$, we get that $\frac{\Gamma}{\theta} \notin \mathrm{L}^1([t_2,+\infty[)$. Besides, recalling that

$$\int_{t_2}^{+\infty}\Gamma(s)\left(f(X(s)) - \min f + \frac{\|V(s)\|^2}{2}\right) < +\infty, \quad a.s.,$$

we invoke Lemma A.2 to conclude that $\lim_{t\to+\infty}\theta(t)\left(f(X(t)) - \min f + \frac{\|V(t)\|^2}{2}\right) = 0$ a.s..

(v) Taking expectation in (4.6) and upper bounding we get

$$\begin{aligned}\mathbb{E}(\phi_3(t,X(t),V(t))) \leq \mathbb{E}(\phi_3(t_2,X(t_2),V(t_2))) &+ \int_{t_2}^t \Gamma(s)\mathbb{E}\left(f(X(s)) - \min f + \frac{\|V(s)\|^2}{2}\right)ds \\ &+ \frac{1}{2}\int_{t_2}^{+\infty}\theta(s)\sigma_\infty^2(s)ds.\end{aligned} \tag{4.9}$$

By the third item, we have that $\mathbb{E}\left(f(X(s)) - \min f + \frac{\|V(s)\|^2}{2}\right) = \mathcal{O}\left(\frac{1}{\Gamma^2(s)}\right)$, so we conclude that

$$\mathbb{E}\left(\theta(t)\left(f(X(t)) - \min f + \frac{\|V(t)\|^2}{2}\right)\right) = \mathcal{O}\left(\int_{t_2}^t \frac{ds}{\Gamma(s)}\right). \tag{4.10}$$

Thus,

$$\mathbb{E}\left(f(X(t)) - \min f + \frac{\|V(t)\|^2}{2}\right) = \mathcal{O}\left(\min\left\{\frac{\int_{t_2}^t \frac{ds}{\Gamma(s)}}{\theta(t)}, \frac{1}{\Gamma^2(t)}\right\}\right).$$

□

4.3 Almost sure weak convergence of trajectories

In the deterministic setting with $\alpha > 3$, it is also well-known that one can obtain weak convergence of the trajectory. Our aim in this section is to show this claim for a general γ in the stochastic setting.

Theorem 4.6. *Consider the setting of Theorem 4.4. Then, if* $\Gamma\sigma_\infty \in \mathrm{L}^2([t_0, +\infty[)$ *we have that:*

(i) $\mathbb{E}\left[\sup_{t\geq t_2} \|X(t)\|^\nu\right] < +\infty$.

(ii) $\forall x^\star \in \mathcal{S}$, $\lim_{t\to+\infty} \|X(t) - x^\star\|$ *exists a.s..*

(iii) If γ *is non-increasing, there exists an* $\mathcal{S}-$*valued random variable* $X^\star$ *such that* $\text{w-lim}_{t\to+\infty} X(t) = X^\star$ *a.s..*

Proof. (i) Analogous to the proof of the first point of [25, Theorem 3.1].

(ii) Recalling the proof of Theorem 4.4, we combine the fact that both

$$\lim_{t\to+\infty} \Gamma^2(t)(f(X(t)) - \min f) + \frac{\|b(X(t) - x^\star) + \Gamma(t)V(t)\|^2}{2} + \frac{b(1-b)}{2}\|X(t) - x^\star\|^2,$$

and

$$\lim_{t\to+\infty} \Gamma^2(t)\left(f(X(t)) - \min f + \frac{\|V(t)\|^2}{2}\right)$$

exist a.s.. We can substract both quantities to obtain that

$$\lim_{t\to+\infty} \frac{\|X(t) - x^\star\|^2}{2} + \Gamma(t)\langle X(t) - x^\star, V(t)\rangle \text{ exists a.s..}$$

Thus, for every $x^\star \in \mathcal{S}$ there exists $\Omega_{x^\star} \in \mathcal{F}$ with $\mathbb{P}(\Omega_{x^\star}) = 1$ and $\exists \ell : \Omega_{x^\star} \mapsto \mathbb{R}$ such that

$$\lim_{t\to+\infty} \frac{\|X(\omega,t) - x^\star\|^2}{2} + \Gamma(t)\langle V(\omega,t), X(\omega,t) - x^\star\rangle = \ell(\omega).$$

Let $Z(\omega,t) = \frac{\|X(\omega,t)-x^\star\|^2}{2} - \ell(\omega)$ and $\varepsilon > 0$ arbitrary. There exists $T(\omega) \geq t_0$ such that $\forall t \geq T(\omega)$

$$\left\|Z(\omega,t) + \Gamma(t)\langle V(\omega,t), X(\omega,t) - x^\star\rangle\right\| < \varepsilon.$$

Let $g(t) \overset{\text{def}}{=} \exp\left(\int_{t_2}^t \frac{ds}{\Gamma(s)}\right)$, multiplying the previous inequality by $\frac{g(t)}{\Gamma(t)}$, there exists $T(\omega) \geq t_0$ such that for every $t \geq T(\omega)$:

$$\left\|\frac{g(t)}{\Gamma(t)}Z(t) + g(t)\langle V(\omega,t), X(\omega,t) - x^\star\rangle\right\| < \frac{\varepsilon}{\Gamma(t)}g(t).$$

On the other hand, $dZ(t) = \langle V(t), X(t) - x^\star\rangle dt$ and

$$d\left(g(t)Z(t)\right) = \left(\frac{g(t)}{\Gamma(t)}Z(t) + g(t)\langle V(t), X(t) - x^\star\rangle\right)dt.$$

Thus,

$$\begin{aligned}\|g(t)Z(t) - g(T)Z(T)\| &= \left\|\int_T^t d(g(s)Z(s))\right\| = \left\|\int_T^t \left(\frac{g(s)}{\Gamma(s)}Z(s) + g(s)\langle V(s), X(s) - x^\star\rangle\right)ds\right\| \\ &\leq \varepsilon\int_T^t \frac{g(s)}{\Gamma(s)}ds = \varepsilon(g(t) - g(T)).\end{aligned}$$

So,

$$\|Z(t)\| \leq \frac{g(T)}{g(t)}\|Z(T)\| + \varepsilon.$$

By Lemma A.3, we obtain that $\lim_{t\to+\infty} g(t) = +\infty$. Hence, $\limsup_{t\to+\infty} \|Z(t)\| \leq \varepsilon$. And we conclude that for every $x^\star \in \mathcal{S}$, $\lim_{t\to+\infty} \frac{\|X(t)-x^\star\|}{2}$ exists a.s.. By a separability argument (see proof of [25, Theorem 3.1] or [27, Theorem 3.6]) there exists $\tilde{\Omega} \in \mathcal{F}$ (independent of $x^\star$) such that $\mathbb{P}(\tilde{\Omega}) = 1$ and $\lim_{t\to+\infty} \frac{\|X(\omega,t)-x^\star\|}{2}$ exists for every $\omega \in \tilde{\Omega}, x^\star \in \mathcal{S}$.

(iii) If γ is non-increasing, then Γ is non-decreasing (see [3, Corollary 2.3]). Then, by item (ii) of Theorem 4.4, we have that:

$$\lim_{t\to+\infty} f(X(t)) = \min f \quad a.s..$$

Let $\Omega_f \in \mathcal{F}$ be the set of events on which this limit is satisfied. Thus $\mathbb{P}(\Omega_f) = 1$. Set $\Omega_{\text{conv}} \overset{\text{def}}{=} \Omega_f \cap \tilde{\Omega}$. We have $\mathbb{P}(\Omega_{\text{conv}}) = 1$. Now, let $\omega \in \Omega_{\text{conv}}$ and $\widetilde{X}(\omega)$ be a weak sequential cluster point of $X(\omega, t)$ (which exists by boundedness on Ω_{conv}). Equivalently, there exists an increasing sequence $(t_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+$ such that $\lim_{k\to+\infty} t_k = +\infty$, and

$$\underset{k\to+\infty}{\text{w-lim}}\, X(\omega, t_k) = \widetilde{X}(\omega).$$

Since $\lim_{t\to+\infty} f(X(\omega, t)) = \min f$ and the fact that f is weakly lower semicontinuous (since it is convex and continuous), we obtain directly that $\widetilde{X}(\omega) \in \mathcal{S}$. Finally by Opial's Lemma (see [68]) we conclude that there exists $X^\star(\omega) \in \mathcal{S}$ such that $\text{w-lim}_{t\to+\infty} X(\omega, t) = X^\star(\omega)$. In other words, since $\omega \in \Omega_{\text{conv}}$ was arbitrary, there exists an $\mathcal{S}$-valued random variable $X^\star$ such that $\text{w-lim}_{t\to+\infty} X(t) = X^\star$ a.s..

□

A Auxiliary results

A.1 Deterministic results

Lemma A.1. *Let $a, b \in \mathbb{R}$ and $x, y \in \mathbb{H}$, then*

$$\|ax - by\| \leq \max\{|a|, |b|\}\|x - y\| + |a - b| \max\{\|x\|, \|y\|\}.$$

Lemma A.2. *Let $t_0 > 0$ and $a, b : [t_0, +\infty[\to \mathbb{R}_+$. If $\lim_{t\to\infty} a(t)$ exists, $b \notin \mathrm{L}^1([t_0, +\infty[)$ and $\int_{t_0}^{+\infty} a(s)b(s)ds < +\infty$, then $\lim_{t\to\infty} a(t) = 0$.*

Lemma A.3. *Under hypothesis (H_γ), then*

$$\int_{t_0}^{+\infty} \frac{ds}{\Gamma(s)} = +\infty.$$

Proof. Let $q(t) \overset{\text{def}}{=} \int_t^{+\infty} \frac{ds}{p(s)}$, since $\int_{t_0}^{+\infty} \frac{ds}{p(s)} < +\infty$, then $\lim_{t\to+\infty} q(t) = 0$ and $q'(t) = -\frac{1}{p(t)}$. On the other hand

$$\int_{t_0}^{+\infty} \frac{ds}{\Gamma(s)} = -\int_{t_0}^{+\infty} \frac{q'(s)}{q(s)} ds = \ln(q(t_0)) - \lim_{t\to+\infty} \ln(q(t)) = +\infty.$$

□

Lemma A.4. *Let $q : [t_0, +\infty[\to \mathbb{R}_+$ be a non-decreasing differentiable function, if $q \notin \mathrm{L}^1([t_0, +\infty[)$, then $\frac{q'}{q} \notin \mathrm{L}^1([t_0, +\infty[)$*

Proof. By definition,

$$\int_{t_0}^{+\infty} \frac{q'(s)}{q(s)} ds = \lim_{t\to+\infty} \ln(q(t)) - \ln(q(t_0)) = +\infty.$$

□

Lemma A.5. *For $a, x > 0$, let us define the upper incomplete Gamma function as:*

$$\Gamma_{inc}(a; x) = \int_x^{+\infty} s^{a-1} e^{-s} ds.$$

Then, the following holds:

(i) $x^{1-a} e^x \Gamma_{inc}(a; x) \leq 1$ *for* $0 < a \leq 1$.

(ii) $x^{1-a}e^x\Gamma_{inc}(a;x) \geq 1$ *for* $a \geq 1$.
(iii) $\lim_{x\to+\infty} x^{1-a}e^x\Gamma_{inc}(a;x) = 1$

Proof. See [69, Section 8]. □

Remark A.6. Do not confuse $\Gamma_{inc}(a;x)$ with $\Gamma(t)$ defined in (3.2).

Corollary A.7. *Let us consider the viscous damping function* $\gamma : [t_0, +\infty[\to \mathbb{R}_+$ *defined by* $\gamma(t) = \frac{\alpha}{t^r}$ *with* $r \in]0,1[$ *and* $\alpha \geq 1 - r$*, then:*
(i) γ *satisfies* (H_γ).
(ii) $\Gamma(t) = \mathcal{O}(t^r)$.
(iii) γ *satisfies* (H'_γ).

Proof. (i) Let $c \overset{\text{def}}{=} \frac{\alpha}{1-r} \geq 1$, we first notice that after the change of variable $u = cs^{1-r}$, we get

$$\int_0^{+\infty} \exp(-cs^{1-r})ds = \frac{1}{\alpha c^{\frac{r}{1-r}}} \int_0^{+\infty} u^{\frac{1}{1-r}-1}e^{-u}du < +\infty,$$

since the last integral is the classical Gamma function (see *e.g.* [69, Section 5]) evaluated at $\frac{1}{1-r}$, and this function is well defined for positive arguments, then (H_γ) is satisfied as $t_0 > 0$.

(ii) Besides, by definition $\Gamma(t) = \exp(ct^{1-r}) \int_t^{+\infty} \exp(-cs^{1-r})ds$. Using the same change of variable as before, we obtain that

$$\Gamma(t) = \frac{\exp(ct^{1-r})}{\alpha c^{\frac{r}{1-r}}} \Gamma_{inc}\left(\frac{1}{1-r}, ct^{1-r}\right). \tag{A.1}$$

By (iii) of Lemma A.5 with $a = \frac{1}{1-r} > 1$ and $x = ct^{1-r}$, for every $\varepsilon > 0$, there exists $t_1 > t_0$ such that for every $t > t_1$:

$$\Gamma(t) \leq \frac{1+\varepsilon}{\alpha c^{\frac{1}{1-r}}} t^r.$$

(iii) Moreover, if we restrict $\varepsilon \in]0, \frac{1}{2}[$, there exists $t_1 > t_0$ such that for every $t > t_1$:

$$\gamma(t)\Gamma(t) \leq \frac{1+\varepsilon}{c^{\frac{1}{1-r}}} \leq 1 + \varepsilon.$$

Defining m as $1 + \varepsilon$, we have that $m < \frac{3}{2}$, and we conclude.

□

Lemma A.8. *Let us define* $p > 0$ *and* $I_p(t) \overset{\text{def}}{=} \int_0^1 e^{-tu}(1-u)^p du$*. Then*
(i) $I_p(t) \leq t^{-1}$ *for every* $t > 0$.
(ii) $I_p(t) \sim t^{-1}$ *as* $t \to +\infty$.

Proof. The first result comes from bounding the term $(1-u)^p$ by 1 in the integral, then we can notice directly that $I_p(t) \leq t^{-1}$ for every $t > 0$. The second result is an application of Watson's Lemma (see [70]). □

A.2 Stochastic results

A.2.1 On stochastic processes

We refer to the notation and results discussed in [27, Section A.2].

Proposition A.9. *Consider* $\nu \geq 2$*,* $X_0, V_0 \in \mathrm{L}^\nu(\Omega;\mathbb{H})$*,* f *and* σ *satisfying* (H_0) *and* (H_σ)*, respectively. Consider also* γ *satisfying* (H_γ)*, and* β *satisfying* (H_β)*. Then* (S − ISIHD) *has a unique solution* $(X, V) \in S^\nu_{\mathbb{H}\times\mathbb{H}}[t_0]$.

Remark A.10. Hypothesis (H_β) does not allow us to consider the case $\beta \equiv 0$, nevertheless, this case is well studied in Section 4.

Proof. We rewrite ($S-ISIHD$) as in the reformulation ($ISIHD-S_R$), we recall (3.3)

$$\begin{cases} dZ(t) &= -\beta(t)\nabla\mathcal{G}(Z(t))dt - \mathcal{D}(t,Z(t))dt + \hat{\sigma}(t,Z(t))dW(t), \quad t > t_0, \\ Z(t_0) &= (X_0, X_0 + \beta(t_0)V_0), \end{cases}$$

Since $\beta(t) \leq c_1$, we have that $-\beta(t)\nabla\mathcal{G}(Z(t))$ is Lipschitz, besides, since $\left|\frac{\beta'(t)-\gamma(t)\beta(t)+1}{\beta(t)}\right| \leq c_2$, we have that $\mathcal{D}$ is a Lipschitz operator. Then, using the hypotheses on σ, we can use [27, Theorem 3.3] and conclude the existence and uniqueness of a process $Z \in S^{\nu}_{\mathbb{H}\times\mathbb{H}}[t_0]$, this, in turn, implies the existence and uniqueness of a solution $(X,V) \in S^{\nu}_{\mathbb{H}\times\mathbb{H}}[t_0]$ of ($S-ISIHD$). □

Theorem A.11. *Let $\mathbb{H}$ be a separable Hilbert space and $(M_t)_{t\geq 0} : \Omega \to \mathbb{H}$ be a continuous martingale such that $\sup_{t\geq 0}\mathbb{E}\left(\|M_t\|^2\right) < +\infty$. Then there exists a $\mathbb{H}-$valued random variable $M_\infty \in \mathrm{L}^2(\Omega;\mathbb{H})$ such that $\lim_{t\to\infty} M_t = M_\infty$ a.s..*

Proof. For the proof, we refer to [27, Theorem A.7]. □

Theorem A.12. *[71, Theorem 1.3.9] Let $\{A_t\}_{t\geq 0}$ and $\{U_t\}_{t\geq 0}$ be two continuous adapted increasing processes with $A_0 = U_0 = 0$ a.s.. Let $\{M_t\}_{t\geq 0}$ be a real-valued continuous local martingale with $M_0 = 0$ a.s.. Let ξ be a non-negative $\mathcal{F}_0$-measurable random variable. Define*

$$X_t = \xi + A_t - U_t + M_t \quad \text{for } t \geq 0.$$

If X_t is non-negative and $\lim_{t\to+\infty} A_t < \infty$, then $\lim_{t\to+\infty} X_t$ exists and is finite, and $\lim_{t\to+\infty} U_t < \infty$.

A.3 Abstract integral bounds, almost sure and in expectation properties of ($S-ISIHD$)

In the following proposition, we state different abstract integral bounds and almost sure properties for ($S-ISIHD$), finally concluding with the almost sure convergence of the gradient towards zero.

Proposition A.13. *Consider that f, σ satisfy (H_0) and (H_σ), respectively. Let $\nu \geq 2$, and consider the dynamic ($S-ISIHD$) with initial data $X_0, V_0 \in \mathrm{L}^\nu(\Omega;\mathbb{H})$. Consider also γ, β from ($S-ISIHD$) satisfying (H_γ) and (H_β), respectively, and suppose there exists a, b, c, d satisfying ($S_{a,b,c,d}$). Finally, we consider $\mathcal{E}$ the energy function defined in (3.5).*

Then, there exists a unique solution $(X,V) \in S^{\nu}_{\mathbb{H}\times\mathbb{H}}[t_0]$ of ($S-ISIHD$). Moreover, if $t \mapsto m(t)\sigma^2_\infty(t) \in \mathrm{L}^1([t_0,+\infty[)$, where $m(t) \overset{\text{def}}{=} \max\{1, a(t), c^2(t)\}$, then the following properties are satisfied:

(i)
$$\lim_{t\to+\infty} \mathcal{E}(t, X(t), V(t)) \quad \text{exists a.s..}$$

(ii) $\int_{t_0}^{+\infty}(b(s)c(s) - a'(s))(f(X(s) + \beta(s)V(s)) - \min f)ds < +\infty$, *a.s..*

(iii) $\int_{t_0}^{+\infty} a(s)\beta(s)\|\nabla f(X(s)) + \beta(s)V(s)\|^2 ds < +\infty$, *a.s..*

(iv) $\int_{t_0}^{+\infty}\left(b(s)b'(s) + \frac{d'(s)}{2}\right)\|X(s) - x^\star\|^2 ds < +\infty$, *a.s..*

(v) $\int_{t_0}^{+\infty} c(s)(\gamma(s)c(s) - c'(s) - b(s))\|V(s)\|^2 ds < +\infty$, *a.s..*

(vi) If $b(t)c(t) - a'(t) = \mathcal{O}(c(t)(\gamma(t)c(t) - c'(t) - b(t)))$, then

$$\int_{t_0}^{+\infty}(b(s)c(s) - a'(s))(f(X(s)) - \min f)ds < +\infty \quad a.s..$$

(vii) If there exists $\eta > 0, \hat{t} > t_0$ such that

$$\eta \leq c(t)(\gamma(t)c(t) - c'(t) - b(t)), \quad \eta \leq a(t)\beta(t), \quad \gamma(t) \leq \eta, \quad \forall t > \hat{t},$$

then $\lim_{t\to+\infty}\|V(t)\| = 0$ a.s., $\lim_{t\to+\infty}\|\nabla f(X(t) + \beta(t)V(t))\| = 0$ a.s., and $\lim_{t\to+\infty}\|\nabla f(X(t))\| = 0$ a.s.

Proof. The existence and uniqueness of a solution is a direct consequence of Corollary A.9. Moreover, applying Proposition 2.3 with $\mathcal{E}$, we can obtain

$$\begin{aligned}\mathcal{E}(t,X(t),V(t)) \leq \mathcal{E}(t_0,X_0,V_0) &- \int_{t_0}^{t}(b(s)c(s)-a'(s))(f(X(s)+\beta(s)V(s))-\min f)ds \\ &- \int_{t_0}^{t} a(s)\beta(s)\|\nabla f(X(s)+\beta(s)V(s))\|^2 ds - \int_{t_0}^{t}\left(b(s)b'(s)+\frac{d'(s)}{2}\right)\|X(s)-x^\star\|^2 ds \\ &- \int_{t_0}^{t} c(s)(b(s)+c'(s)-c(s)\gamma(s))\|V(s)\|^2 ds + \int_{t_0}^{t}(La(s)\beta^2(s)+c^2(s))\sigma_\infty^2(s)ds + M_t,\end{aligned} \tag{A.2}$$

where $M_t = \int_{t_0}^{t}\langle \sigma^\star(s,X(s)+\beta(s)V(s))(a(s)\beta(s)\nabla f(X(s)+\beta(s)V(s))+c(s)[b(s)(X(s)-x^\star)+c(s)V(s)], dW(s)\rangle$. Since $\sup_{t\in[t_0,T]}\mathbb{E}(\|X(t)\|^2) < +\infty, \sup_{t\in[t_0,T]}\mathbb{E}(\|V(t)\|^2) < +\infty$ for every $T > t_0$, and a, b, c, β are continuous functions, we have that M_t is a continuous martingale, on the other hand, we have that

$$\int_{t_0}^{+\infty}(La(s)\beta^2(s)+c^2(s))\sigma_\infty^2(s)ds < +\infty.$$

Then, we can apply Theorem A.12 and conclude that $\lim_{t\to+\infty}\mathcal{E}(t,X(t),V(t))$ exists a.s. and

- $\int_{t_0}^{+\infty}(b(s)c(s)-a'(s))(f(X(s)+\beta(s)V(s))-\min f)ds < +\infty$ a.s..
- $\int_{t_0}^{+\infty} a(s)\beta(s)\|\nabla f(X(s))+\beta(s)V(s)\|^2 ds < +\infty$ a.s..
- $\int_{t_0}^{+\infty}\left(b(s)b'(s)+\frac{d'(s)}{2}\right)\|X(s)-x^\star\|^2 ds < +\infty$ a.s..
- $\int_{t_0}^{+\infty} c(s)(\gamma(s)c(s)-c'(s)-b(s))\|V(s)\|^2 ds < +\infty$ a.s..

This let us conclude with items (i) to (v).

Let $\tilde{b}(t) = b(t)c(t)-a'(t)$, and

$$\begin{aligned}I_f &\overset{\text{def}}{=} \int_{t_0}^{+\infty}\tilde{b}(s)(f(X(s))-\min f)ds \\ &\leq \int_{t_0}^{+\infty}\tilde{b}(s)(f(X(s))-f(X(s)+\beta(s)V(s)))ds + \int_{t_0}^{+\infty}\tilde{b}(s)(f(X(s)+\beta(s)V(s))-\min f)ds\end{aligned}$$

Using Descent Lemma, Cauchy Schwarz Inequality and Corollary 2.2:

$$\begin{aligned}I_f \leq \sqrt{2L}\beta_0\left(\int_{t_0}^{+\infty}\tilde{b}(s)(f(X(s)+\beta(s)V(s))-\min f)ds\right)^{\frac{1}{2}}&\left(\int_{t_0}^{+\infty}\tilde{b}(s)\|V(s)\|^2 ds\right)^{\frac{1}{2}} \\ + \frac{L\beta_0^2}{2}\int_{t_0}^{+\infty}\tilde{b}(s)\|V(s)\|^2 ds &+ \int_{t_0}^{+\infty}\tilde{b}(s)(f(X(s)+\beta(s)V(s))-\min f)ds.\end{aligned}$$

If $\tilde{b}(t) = b(t)c(t)-a'(t) = \mathcal{O}(c(t)(\gamma(t)c(t)-c'(t)-b(t)))$, we have that

$$\int_{t_0}^{+\infty}\tilde{b}(s)\|V(s)\|^2 ds < +\infty \quad a.s..$$

And we conclude with item (vi).

To prove (vii), in particular that $\lim_{t\to\infty}\|V(t)\| = 0$, we consider that if there exists $\eta > 0, \hat{t} > t_0$ such that $\eta \leq c(t)(\gamma(t)c(t)-c'(t)-b(t)), \forall t > \hat{t}$, then there exists $\Omega_v \in \mathcal{F}$ such that $\mathbb{P}(\Omega_v) = 1$ and

$$\int_{t_0}^{+\infty}\|V(\omega,s)\|^2 ds < +\infty, \quad \forall\omega\in\Omega_v.$$

Then, we have $\liminf_{t\to\infty}\|V(\omega,t)\| = 0, \forall\omega \in \Omega_v$. Let us suppose that $\limsup_{t\to\infty}\|V(\omega,t)\| > 0, \forall\omega \in \Omega_v$. Then, by [27, Lemma A.3], there exists $\delta > 0$ satisfying

$$0 = \liminf_{t\to+\infty}\|V(\omega,t)\| < \delta < \limsup_{t\to+\infty}\|V(\omega,t)\|, \quad \forall\omega \in \Omega_v.$$

And there exists $(t_k)_{k\in\mathbb{N}} \subset [t_0, +\infty[$ such that $\lim_{k\to+\infty} t_k = +\infty$,

$$\|V(\omega,t_k)\| > \delta, \quad \forall\omega \in \Omega_v \text{ and } t_{k+1} - t_k > 1, \quad \forall k \in \mathbb{N}.$$

Let $N_t \overset{\text{def}}{=} \int_{t_0}^t \sigma(s, X(s) + \beta(s)V(s))dW(s)$. This is a continuous martingale (w.r.t. the filtration $\mathcal{F}_t$), which verifies

$$\mathbb{E}(\|N_t\|^2) = \mathbb{E}\left(\int_{t_0}^t \|\sigma(s, X(s) + \beta(s)V(s))\|_{\text{HS}}^2 ds\right) \leq \mathbb{E}\left(\int_{t_0}^{+\infty} \sigma_\infty^2(s)ds\right) < +\infty, \forall t \geq t_0.$$

According to Theorem A.11, we deduce that there exists a $\mathbb{H}-$valued random variable N_∞ w.r.t. $\mathcal{F}_\infty$, and which verifies: $\mathbb{E}(\|N_\infty\|^2) < +\infty$, and there exists $\Omega_N \in \mathcal{F}$ such that $\mathbb{P}(\Omega_N) = 1$ and

$$\lim_{t\to+\infty} N_t(\omega) = N_\infty(\omega) \text{ for every } \omega \in \Omega_N.$$

Let $\omega_0 \in \Omega_{nv} \overset{\text{def}}{=} \Omega_N \cap \Omega_v$ ($\mathbb{P}(\Omega_{nv}) = 1$) and the notation $V(t) \overset{\text{def}}{=} V(\omega_0, t)$, $\varepsilon \in \left(0, \min\{1, \frac{\delta^2}{4}\}\right)$ arbitrary and recall that $\eta \leq a(t)\beta(t), \gamma(t) \leq \eta$ for every $t > \hat{t}$. Let $k' \in \mathbb{N}$ be such that $t_{k'} > \hat{t}$, $k > k'$ and $t \in [t_k, t_k + \varepsilon]$, then

$$\begin{aligned}
\|V(t) - V(t_k)\|^2 &\leq 3(t - t_k)\int_{t_k}^t \gamma^2(s)\|V(s)\|^2 ds + 3(t - t_k)\int_{t_k}^t \|\nabla f(X(s) + \beta(s)V(s))\|^2 ds \\
&\quad + 3\|N_t - N_{t_k}\|^2 \\
&\leq 3\eta^2(t - t_k)\int_{t_k}^t \|V(s)\|^2 ds + \frac{3}{\eta}(t - t_k)\int_{t_k}^t a(s)\beta(s)\|\nabla f(X(s) + \beta(s)V(s))\|^2 ds \\
&\quad + 3\|N_t - N_{t_k}\|^2 \\
&\leq 3\eta^2\varepsilon\int_{t_k}^t \|V(s)\|^2 ds + \frac{3}{\eta}\varepsilon\int_{t_k}^t a(s)\beta(s)\|\nabla f(X(s) + \beta(s)V(s))\|^2 ds \\
&\quad + 3\|N_t - N_{t_k}\|^2.
\end{aligned}$$

Now let $k'' \in \mathbb{N}$ be such that for every $k > k''$,

$$\int_{t_k}^{+\infty} \|V(s)\|^2 ds < \frac{1}{9\eta^2}, \int_{t_k}^{+\infty} a(s)\beta(s)\|\nabla f(X(s) + \beta(s)V(s))\|^2 ds < \frac{\eta}{9}, \sup_{t>t_k}\|N_t - N_{t_k}\|^2 < \frac{\varepsilon}{9}.$$

Then, we have that

$$\|V(t) - V(t_k)\|^2 \leq \varepsilon \leq \frac{\delta^2}{4}, \forall t \in [t_k, t_k + \varepsilon], k > \max\{k', k''\}.$$

For such t, we bound using the triangular inequality and obtain

$$\|V(t)\| \geq \|V(t_k)\| - \|V(t) - V(t_k)\| > \frac{\delta}{2}.$$

Now we consider

$$\int_{t_0}^{+\infty} \|V(s)\|^2 ds \geq \sum_{k>\max\{k',k''\}} \int_{t_k}^{t_k+\varepsilon} \|V(s)\|^2 ds \geq \sum_{k>\max\{k',k''\}} \frac{\varepsilon\delta^2}{4} = +\infty.$$

Which is a contradiction, then we conclude that $\liminf_{t\to+\infty}\|V(t)\| = \limsup_{t\to+\infty}\|V(t)\| = 0$, a.s..

To prove the second part of (vii), i.e. that $\lim_{t\to+\infty}\|\nabla f(X(t)+\beta(t)V(t))\| = 0$, we recall that there exists $\eta > 0, \hat{t} > t_0$ such that $\eta \leq a(t)\beta(t)$, then there exists $\Omega_y \in \mathcal{F}$ such that $\mathbb{P}(\Omega_y) = 1$ and

$$\int_{t_0}^{+\infty} \|\nabla f(X(\omega,s)+\beta(s)V(\omega,s))\|^2 ds < +\infty \quad \forall \omega \in \Omega_y$$

So we have that

$$\liminf_{t\to+\infty} \|\nabla f(X(\omega,t)+\beta(t)V(\omega,t))\| = 0, \quad \forall \omega \in \Omega_y.$$

Moreover, if we suppose that

$$\limsup_{t\to+\infty} \|\nabla f(X(\omega,t)+\beta(t)V(\omega,t))\| > 0, \forall \omega \in \Omega_y,$$

by [27, Lemma A.3], there exists $\delta > 0, (t_k)_{k\in\mathbb{N}} \subset [t_0, +\infty[$ such that $\lim_{k\to+\infty} t_k = +\infty$,

$$\|\nabla f(X(\omega,t_k)+\beta(t_k)V(\omega,t_k))\| > \delta \quad \forall \omega \in \Omega_y \text{ and } t_{k+1} - t_k > 1, \quad \forall k \in \mathbb{N}.$$

Recall that by (H_β), there exists β_0 such that $\beta(t) \leq \beta_0$. Let $\varepsilon \in \left(0, \min\{1, \frac{\delta^2}{4L^2}\}\right)$ arbitrary and consider $Y(t) = X(t) + \beta(t)V(t)$, let also $k \in \mathbb{N}$ arbitrary and $t \in [t_k, t_k + \varepsilon]$. Then, using Lemma A.1 and Jensen's inequality we can bound as follows:

$$\begin{aligned}
\|Y(t) - Y(t_k)\|^2 &\leq 2\|X(t) - X(t_k)\|^2 + 2\|\beta(t)V(t) - \beta(t_k)V(t_k)\|^2 \\
&\leq 2\Big\|\int_{t_k}^{t} V(s)ds\Big\|^2 + 2\left(\beta_0\|V(t) - V(t_k)\| + |\beta(t) - \beta(t_k)|\max\{\|V(t)\|, \|V(t_k)\|\}\right)^2 \\
&\leq 2(t - t_k)\int_{t_k}^{+\infty} \|V(s)\|^2 ds \\
&\quad + 2\left(\beta_0\|V(t) - V(t_k)\| + |\beta(t) - \beta(t_k)|\max\{\|V(t)\|, \|V(t_k)\|\}\right)^2 \\
&\leq 2(t - t_k)\int_{t_k}^{+\infty} \|V(s)\|^2 ds \\
&\quad + 4\beta_0^2\|V(t) - V(t_k)\|^2 + 4|\beta(t) - \beta(t_k)|^2 \max\{\|V(t)\|, \|V(t_k)\|\}^2.
\end{aligned}$$

By the previous point, we have that there exists $\Omega_v \in \mathcal{F}$ such that $\mathbb{P}(\Omega_v) = 1$ such that

$$\int_{t_0}^{+\infty} \|V(\omega,s)\|^2 ds < +\infty \quad \forall \omega \in \Omega_v,$$

and $k' \in \mathbb{N}$ such that for every $k > k'$, for all $t \in [t_k, t_k + \varepsilon]$:

$$\int_{t_k}^{+\infty} \|V(\omega,s)\|^2 ds < \frac{1}{6}, \quad \max\{\|V(\omega,t)\|, \|V(\omega,t_k)\|\} < 1, \quad \|V(\omega,t) - V(\omega,t_k)\|^2 \leq \frac{\varepsilon}{12\beta_0^2}.$$

We consider an arbitrary $\omega_0 \in \Omega_y \cap \Omega_v$ ($\mathbb{P}(\Omega_y \cap \Omega_v) = 1$), and we let us use the abuse of notation $X(t) = X(\omega_0, t), V(t) = V(\omega_0, t)$, and $Y(t) = Y(\omega_0, t)$ for the rest of this proof.
On the other hand, β is continuous, so there exists $\tilde{\delta} > 0$ such that, if $|t - t_k| < \tilde{\delta}$, then $|\beta(t) - \beta(t_k)| < \frac{\sqrt{\varepsilon}}{2\sqrt{3}}$.
Therefore, letting $\varepsilon' = \min\{\varepsilon, \tilde{\delta}\}$, we have that

$$\|\nabla f(Y(t)) - \nabla f(Y(t_k))\|^2 \leq L^2\|Y(t) - Y(t_k)\|^2 \leq L^2\varepsilon \leq \frac{\delta^2}{4}, \forall k > k', \forall t \in [t_k, t_k + \varepsilon'].$$

Then, we obtain

$$\|\nabla f(Y(t))\| \geq \|\nabla f(Y(t_k))\| - \|\nabla f(Y(t)) - \nabla f(Y(t_k))\| \geq \frac{\delta}{2}, \forall k > k', \forall t \in [t_k, t_k + \varepsilon'].$$

This implies that

$$\int_{t_0}^{+\infty} \|\nabla f(Y(s))\|^2 ds \geq \sum_{k>k'} \int_{t_k}^{t_k+\varepsilon'} \|\nabla f(Y(s))\|^2 ds \geq \sum_{k>k'} \int_{t_k}^{t_k+\varepsilon'} \frac{\delta^2}{4} = \sum_{k>k'} \frac{\varepsilon'\delta^2}{4} = +\infty.$$

Which is a contradiction, then we conclude that

$$\liminf_{t\to+\infty} \|\nabla f(X(t) + \beta(t)V(t))\| = \limsup_{t\to+\infty} \|\nabla f(X(t) + \beta(t)V(t))\| = 0, \quad a.s..$$

To prove the last part of (vii), we consider that $\beta(t) \leq \beta_0$, then

$$\begin{aligned}\|\nabla f(X(t))\| &\leq \|\nabla f(X(t) + \beta(t)V(t))\| + \|\nabla f(X(t)) - \nabla f(X(t) + \beta(t)V(t))\| \\ &\leq \|\nabla f(X(t) + \beta(t)V(t))\| + L\beta_0\|V(t)\|.\end{aligned}$$

With this bound, we can conclude that $\lim_{t\to+\infty} \|\nabla f(X(t))\| = 0$ a.s.. □

The following proposition states abstract bounds in expectation of (S − ISIHD).

Proposition A.14. *Consider the setting of Proposition A.13, then we have that:*

(i) $\mathbb{E}(f(X(t) + \beta(t)V(t)) - \min f) = \mathcal{O}\left(\frac{1}{a(t)}\right)$.

Moreover, if there exists $D > 0, \tilde{t} > t_0$ such that $d(t) \geq D$ for $t > \tilde{t}$, then :

(ii) $\sup_{t>t_0} \mathbb{E}(\|X(t) - x^\star\|^2) < +\infty$.

(iii) $\mathbb{E}(\|V(t)\|^2) = \mathcal{O}\left(\frac{1 + b^2(t)}{c^2(t)}\right)$.

(iv) $\mathbb{E}\left(f(X(t)) - \min f\right) = \mathcal{O}\left(\max\left\{\frac{1}{a(t)}, \frac{\beta(t)\sqrt{1+b^2(t)}}{\sqrt{a(t)}c(t)}, \frac{\beta^2(t)\left(1+b^2(t)\right)}{c^2(t)}\right\}\right)$.

Proof. To prove this proposition we are going to take expectation in (A.2). First, we are going to bound the negative terms by 0, denoting $E_0 \overset{\text{def}}{=} \mathcal{E}(t_0) + \max\{1, L\}\int_{t_0}^{+\infty}(a(s)\beta^2(s) + c^2(s))\sigma_\infty^2(s)ds$, we obtain that

$$\mathbb{E}(\mathcal{E}(t, X(t), V(t))) \leq E_0.$$

This implies that

- $\mathbb{E}(f(X(t) + \beta(t)V(t)) - \min f) \leq \frac{E_0}{a(t)}$.
- $\mathbb{E}(\|b(t)(X(t) - x^\star) + c(t)V(t)\|^2) \leq 2E_0$.

If there exists $D > 0, \tilde{t} > t_0$ such that $d(t) \geq D$ for $t > \tilde{t}$, then for $t > \tilde{t}$:

- $\mathbb{E}(\|X(t) - x^\star\|^2) \leq \frac{2E_0}{D}$.
- And also,

$$\begin{aligned}\mathbb{E}(\|V(t)\|^2) &\leq \frac{2}{c^2(t)}[\mathbb{E}(\|b(t)(X(t) - x^\star) + c(t)V(t)\|^2) + b^2(t)\mathbb{E}(\|X(t) - x^\star\|^2)] \\ &\leq \frac{2}{c^2(t)}\left(2E_0 + \frac{2E_0 b^2(t)}{D}\right) = \frac{4E_0}{c^2(t)}\left(1 + \frac{b^2(t)}{D}\right).\end{aligned}$$

- We bound the following term using the Descent Lemma

$$\begin{aligned}\mathbb{E}(f(X(t)) - f(X(t) + \beta(t)V(t))) \leq \beta(t)&\sqrt{\mathbb{E}(\|\nabla f(X(t) + \beta(t)V(t))\|^2)}\sqrt{\mathbb{E}(\|V(t)\|^2)} \\ &+ \frac{L}{2}\beta^2(t)\mathbb{E}(\|V(t)\|^2).\end{aligned}$$

Using Corollary 2.2, we have

$$\begin{aligned}\mathbb{E}(f(X(t)) - f(X(t) + \beta(t)V(t))) &\leq \beta(t)\sqrt{2L\mathbb{E}(f(X(t) + \beta(t)V(t)) - \min f)}\sqrt{\mathbb{E}(\|V(t)\|^2)} \\ &\quad + \frac{L}{2}\beta^2(t)\mathbb{E}(\|V(t)\|^2) \\ &\leq 2E_0\sqrt{2L}\frac{\beta(t)}{\sqrt{a(t)}}\frac{\sqrt{1+\frac{b^2(t)}{D}}}{c(t)} + 2LE_0\frac{\beta^2(t)\left(1+\frac{b^2(t)}{D}\right)}{c^2(t)} \\ &= \mathcal{O}\left(\max\left\{\frac{\beta(t)}{\sqrt{a(t)}}\frac{\sqrt{1+b^2(t)}}{c(t)}, \frac{\beta^2(t)\left(1+b^2(t)\right)}{c^2(t)}\right\}\right)\end{aligned}$$

Then, we notice that

$$\begin{aligned}\mathbb{E}(f(X(t)) - \min f) &= \mathbb{E}[f(X(t)) - f(X(t) + \beta(t)V(t))] + \mathbb{E}[f(X(t) + \beta(t)V(t)) - \min f] \\ &= \mathcal{O}\left(\max\left\{\frac{\beta(t)}{\sqrt{a(t)}}\frac{\sqrt{1+b^2(t)}}{c(t)}, \frac{\beta^2(t)\left(1+b^2(t)\right)}{c^2(t)}, \frac{1}{a(t)}\right\}\right).\end{aligned}$$

□

References

[1] A. Cauchy. Méthode générale pour la résolution des systèmes d'équations simultanées. *Comptes Rendus de l'Académie des Sciences de Paris*, 25:536–538, 1847.

[2] J. Bolte, T.P. Nguyen, J. Peypouquet, and B. W. Suter. From error bounds to the complexity of first-order descent methods for convex functions. *Mathematical Programming*, 165:471–507, 2016.

[3] H. Attouch and A. Cabot. Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity. *Journal of Differential Equations*, 263-9:5412–5458, 2017.

[4] Alexandre Cabot, Hans Engler, and Gadat Sébastien. 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.

[5] Weijie Su, Stephen Boyd, and Emmanuel J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17:1–43, 2016.

[6] Y.E. Nesterov. A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. *Doklady Akademii Nauk SSSR*, 269(3):543–547, 1983.

[7] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Math. Program. Ser. B*, 168:123–175, 2018.

[8] Hedy Attouch and Juan Peypouquet. The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $\frac{1}{k^2}$. *SIAM Journal on Optimization*, 26(3):1824–1834, 2016.

[9] Ramzi May. Asymptotic for a second order evolution equation with convex potential and vanishing damping term. *Turkish Journal of Mathematics*, 41, 2017.

[10] Boris Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 1964.

[11] Hedy Attouch, Xavier Goudou, and Patrick Redont. The heavy ball with friction method. i- the continuous dynamical system. *Communications in Contemporary Mathematics*, 2 (1), 2011.

[12] F. Alvarez, H. Attouch, J. Bolte, and P. Redont. A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics. *J. Math. Pures Appl. (9)*, 81(8):747–779, 2002.

[13] Hedy Attouch, Juan Peypouquet, and Patrick Redont. Fast convex optimization via inertial dynamics with Hessian driven damping. *J. Differential Equations*, 261(10):5734–5783, 2016.

[14] Hedy Attouch, Zaki Chbani, Jalal Fadili, and Hassan Riahi. First-order optimization algorithms via inertial systems with Hessian driven damping. *Math. Program.*, 193(1, Ser. A):113–155, 2022.

[15] Hedy Attouch, Aïcha Balhag, Zaki Chbani, and Hassan Riahi. Fast convex optimization via inertial dynamics combining viscous and Hessian-driven damping with time rescaling. *Evol. Equ. Control Theory*, 11(2):487–514, 2022.

[16] C. Alecsa, S. László, and T. Pinta. An extension of the second order dynamical system that models Nesterov convex gradient method. *Appl. Math. Optim.*, 84:1687–1716, 2021.

[17] Michael Muehlebach and Michael I. Jordan. Optimization with momentum: dynamical, control-theoretic, and symplectic perspectives. *J. Mach. Learn. Res.*, 22:Paper No. 73, 50, 2021.

[18] Qianxiao Li, Cheng Tai, and Weinan E. Stochastic modified equations and adaptive stochastic gradient algorithms. *arXiv:1511.06251*, 2017.

[19] Antonio Orvieto and Aurelien Lucchi. Continuous-time models for stochastic optimization algorithms. *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, 2019.

[20] Wenqing Hu, Chris Junchi Li, Lei Li, and Jian-Guo Lui. On the diffusion approximation of nonconvex stochastic gradient descent. *arXiv:1705.07562v2*, 2018.

[21] Bin Shi, Weijie J. Su, and Michael I. Jordan. On learning rates and Schrödinger operators. *Journal of Machine Learning Research*, 24:1–53, 2023.

[22] Zhiyuan Li, Sadhika Malladi, and Sanjeev Arora. On the validity of modeling sgd with stochastic differential equations. *arXiv:2102.12470*, 2021.

[23] S. Soatto and P. Chaudhari. Stochastic gradient descent performs variational inference, converges to limit cycles for deep networks. *2018 Information Theory and Applications Workshop (ITA)*, pages 1–10, 2018.

[24] Panayotis Mertikopoulos and Mathias Staudigl. On the convergence of gradient-like flows with noisy gradient input. *SIAM Journal on Optimization*, 28(1):163–197, 2018.

[25] Rodrigo Maulen S., Jalal Fadili, and Hedy Attouch. An SDE perspective on stochastic convex optimization. *arXiv:2207.02750*, 2022.

[26] M. Dambrine, C. Dossal, B. Puig, and A. Rondepierre. Stochastic differential equations for modeling first order optimization methods. *Hal*, 2022.

[27] Rodrigo Maulen-Soto, Jalal Fadili, and Hedy Attouch. Tikhonov regularization for stochastic non-smooth convex optimization in Hilbert spaces. *arXiv:2403.06708v3*, 2024.

[28] Herbert Robins and Sutton Monro. A stochastic approximation method. *Ann. Math. Statist. 22*, pages 400–407, 1951.

[29] S. Mandt, M. Hoffman, and D. Blei. A variational analysis of stochastic gradient algorithms. In *International Conference on Machine Learning*, 2016.

[30] Z. Xie, I. Sato, and M. Sugiyama. A diffusion theory for deep learning dynamics: Stochastic gradient descent exponentially favors flat minima. In *International Conference on Learning Representations*, 2021.

[31] J. Latz. Analysis of stochastic gradient descent in continuous time. *Statistics and Computing*, pages 31–39, 2021.

[32] X. Li, Z. Shen, L. Zhang, and N. He. A Hessian-aware stochastic differential equation for modelling SGD. *arXiv:2405.18373*, 2024.

[33] Rodrigo Maulen-Soto, Jalal Fadili, Hedy Attouch, and Peter Ochs. Stochastic inertial dynamics via time scaling and averaging. *arXiv:2403.16775*, 2024.

[34] Grigorios A. Pavliotis. Stochastic processes and applications. *Springer*, 2014.

[35] Xiang Cheng, Niladri S. Chatterji, Peter L. Bartlett, and Michael I. Jordan. Underdamped Langevin MCMC: A non-asymptotic analysis. *arXiv:1707.03663*, 2017.

[36] Yi-An Ma, Niladri Chatterji, Xiang Cheng, Nicolas Flammarion, Peter Bartlett, and Michael I. Jordan. Is there an analog of Nesterov acceleration for MCMC? *Bernoulli*, 27 (3):1942–1992, 2021.

[37] Arnak S. Dalalyan, Lionel Riou-Durand, and Avetik G. Karagulyan. Bounding the error of discretized langevin algorithms for non-strongly log-concave targets. *J. Mach. Learn. Res.*, 23:235:1–235:38, 2019.

[38] Hedy Attouch, Jalal Fadili, and Vyacheslav Kungurtsev. On the effect of perturbations in first-order optimization methods with inertia and Hessian driven damping. *Evolution equations and Control*, 12(1):71–117, 2023.

[39] A. Haraux and Jendoubi M.A. On a second order dissipative ODE in Hilbert space with an integrable source term. *Acta Math. Sci.*, 32:155–163, 2012.

[40] B. Shi, S.S. Du, M.I. Jordan, and Su W.J. Understanding the acceleration phenomenon via high resolution differential equations. *Math. Program.*, 2021.

[41] H. Attouch, A. Cabot, Chbani Z., and H. Riahi. Accelerated forward-backward algorithms with perturbations: Application to Tikhonov regularization. *J. Optim. Theory Appl.*, 179:1–36, 2018.

[42] C. Dossal and J.F. Aujol. Stability of over-relaxations for the forward-backward algorithm, application to fista. *SIAM J. Optim.*, 25:2408–2433, 2015.

[43] M. Schmidt, N. Le Roux, and F. Bach. Convergence rates of inexact proximal-gradient methods for convex optimization. *NIPS'11*, 25th Annual Conference, 2011.

[44] S. Villa, S. Salzo, and Baldassarres L. Accelerated and inexact forward-backward. *SIAM J. Optim.*, 23:1607–1633, 2013.

[45] H. Attouch, J. Peypouquet, and P. Redont. Fast convex minimization via inertial dynamics with Hessian driven damping. *J. Differential Equations*, 261:5734–5783, 2016.

[46] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi. First order optimization algorithms via inertial systems with Hessian driven damping. *Math. Program.*, 2020.

[47] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi. Convergence of iterates for first-order optimization algorithms with inertia and Hessian driven damping. *Optimization*, 2021.

[48] Hedy Attouch and Alexandre Cabot. Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity. *J. Differential Equations*, 263(9):5412–5458, 2017.

[49] A. Orvieto, J. Kohler, and A. Lucchi. The role of memory in stochastic optimization. *Proceeding of Machine Learning Research*, 115:356–366, 2020.

[50] Sébastien Gadat and Fabien Panloup. Long time behaviour and stationary regime of memory gradient diffusions. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, 2014.

[51] Hongzhou Lin, Julien Mairal, and Zaid Harchaoui. Catalyst acceleration for first-order convex optimization: from theory to practice. *J. Mach. Learn. Res.*, 18:Paper No. 212, 54, 2017.

[52] R. Frostig, S. Kakade R. Ge, and A. Sidford. Un-regularizing: approximate proximal point and faster stochastic algorithms for empirical risk minimization. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pages 2540–2548, 2015.

[53] Prateek Jain, Praneeth Netrapalli, Sham M. Kakade, Rahul Kidambi, and Aaron Sidford. Parallelizing stochastic gradient descent for least squares regression: mini-batching, averaging, and model misspecification. *J. Mach. Learn. Res.*, 18:Paper No. 223, 42, 2017.

[54] M. Assran and M. Rabbat. On the convergence of nesterov's accelerated gradient method in stochastic settings. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 410–420, 2020.

[55] Zeyuan Allen-Zhu. Katyusha: the first direct acceleration of stochastic gradient methods. *J. Mach. Learn. Res.*, 18:Paper No. 221, 51, 2017.

[56] Bowei Yan. *Theoretical Analysis for Convex and Non-Convex Clustering Algorithms*. ProQuest LLC, Ann Arbor, MI, 2018. Thesis (Ph.D.)–The University of Texas at Austin.

[57] Sébastien Gadat, Fabien Panloup, and Sofiane Saadane. Stochastic heavy ball. *Electron. J. Stat.*, 12(1):461–529, 2018.

[58] Nicolas Loizou and Peter Richtárik. Momentum and stochastic momentum for stochastic gradient, Newton, proximal point and subspace descent methods. *Comput. Optim. Appl.*, 77(3):653–710, 2020.

[59] Maxime Laborde and Adam Oberman. A lyapunov analysis for accelerated gradient methods: from deterministic to stochastic case. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 602–612. PMLR, 26–28 Aug 2020.

[60] Guanghui Lan. *First-order and stochastic optimization methods for machine learning*. Springer Series in the Data Sciences. Springer, Cham, [2020] ©2020.

[61] Aaron Defazio and Samy Jelassi. Adaptivity without compromise: a momentumized, adaptive, dual averaged gradient method for stochastic optimization. *J. Mach. Learn. Res.*, 23:Paper No. [144], 34, 2022.

[62] Derek Driggs, Matthias J. Ehrhardt, and Carola-Bibiane Schönlieb. Accelerating variance-reduced stochastic gradient methods. *Math. Program.*, 191(2, Ser. A):671–715, 2022.

[63] Anis Hamadouche, Yun Wu, Andrew M. Wallace, and João F. C. Mota. Sharper bounds for proximal gradient algorithms with errors. *SIAM Journal on Optimization*, 34(1):278–305, 2024.

[64] Hedy Attouch, Jalal Fadili, and Vyacheslav Kungurtsev. The stochastic ravine accelerated gradient method with general extrapolation coefficients, 2024.

[65] Hedy Attouch, Radu Ioan Bot, and Dang-Khoa Nguyen. Fast convex optimization via time scale and averaging of the steepest descent. *arXiv:2208.08260*, 2022.

[66] R.T. Rockafellar. Convex analysis. *Princeton univerisity press*, 28, 1997.

[67] Leszek Gawarecki and Vidyadhar Mandrekar. Stochastic differential equations in infinite dimensions. *Springer*, 2011.

[68] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.*, 73:591–597, 1967.

[69] F.W.J. Olver, D.w. Lozier, R.F Boisvert, and Clarck C.W. Nist handbook of mathematical functions. *Cambridge University Press*, 2010.

[70] G.N. Watson. The harmonic functions associated with the parabolic cylinder. *Proceedings of the London Mathematical Society*, 2, no. 17:116–148, 1918.

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