Stochastic Differential Inclusions and Tikhonov Regularization for Stochastic Non-Smooth Convex Optimization in Hilbert spaces

Rodrigo Maulen-Soto
Normandie Université, ENSICAEN, UNICAEN, CNRS, GREYC
6 Bd. Marechal Juin, Caen, France
rodrigo.maulen@ensicaen.fr

Jalal Fadili
Normandie Université, ENSICAEN, UNICAEN, CNRS, GREYC
6 Bd. Marechal Juin, Caen, France
jalal.fadili@ensicaen.fr

Hedy Attouch
IMAG, CNRS, Université Montpellier
499-554 Rue du Truel 9, Montpellier, France
hedy.attouch@umontpellier.fr

Abstract

To solve convex optimization problems with a noisy gradient input, we analyze the global behavior of subgradient-like flows under stochastic errors. The objective function is composite, being equal to the sum of two convex functions, one being differentiable and the other potentially non-smooth. We then use stochastic differential inclusions where the drift term is minus the subgradient of the objective function, and the diffusion term is either bounded or square-integrable. In this context, under Lipschitz's continuity of the differentiable term and a growth condition of the non-smooth term, our first main result shows almost sure weak convergence of the trajectory process towards a minimizer of the objective function. Then, using Tikhonov regularization with a properly tuned vanishing parameter, we can obtain almost sure strong convergence of the trajectory towards the minimum norm solution. We find an explicit tuning of this parameter when our objective function satisfies a local error-bound inequality. We also provide a comprehensive complexity analysis by establishing several new pointwise and ergodic convergence rates in expectation for the convex, strongly convex, and Łojasiewicz case.

Keywords Stochastic optimization, inertial gradient system, Convex optimization, Non-smooth optimization, Stochastic Differential Equation, Stochastic Differential Inclusion, Tikhonov regularization, Error bound inequality, Łojasiewicz inequality, KL inequality, Convergence rate, Asymptotic behavior.

1 Introduction

1.1 Problem statement

We aim to solve convex minimization problems by means of stochastic differential inclusions (SDI), showing the existence, uniqueness, and properties of the solution. Then, we work with Tikhonov regularization, specifically when the drift term is the sum of the (sub-)gradient of the objective function and of a Tikhonov regularization term with a vanishing coefficient. This makes it possible to take into account a noisy (imprecise) gradient input and obtain convergence a.s. to the minimal norm solution.

Let us consider the minimization problem

$$\min_{x\in\mathbb{H}} F(x) \stackrel{\text{def}}{=} f(x)+g(x), \tag{P}$$

where $\mathbb{H}$ is a separable real Hilbert space, and the objective F satisfies the following standing assumptions:

$$\begin{cases} f:\mathbb{H}\to\mathbb{R} \text{ is continuously differentiable and convex with } L\text{-Lipschitz continuous gradient;} \\ g:\mathbb{H}\to\mathbb{R} \text{ is proper, lsc and convex;} \\ \mathcal{S}_F \stackrel{\text{def}}{=} \operatorname{argmin}(F)\neq\emptyset. \end{cases} \tag{H$_0$}$$

To solve (P), a fundamental dynamic to consider is the subgradient flow, which is the following differential inclusion (DI) starting in $t_0 \geq 0$ with initial condition $x_0 \in \mathbb{H}$:

$$\begin{cases} \dot{x}(t) \in -\partial F(x(t)), \quad t > t_0; \\ x(t_0) = x_0. \end{cases} \tag{DI}$$

It is well known since the founding articles of Brézis, Baillon, Bruck in the 1970s that, when the initial data x_0 is in the domain of F, (more generally when it is in its closure), there exists a unique strong global solution of (DI). Moreover, if the solution set $\mathrm{argmin}(F)$ of (P) is nonempty then each solution trajectory of (DI) converges weakly, and its limit belongs to $\mathrm{argmin}(F)$.

In many cases, the gradient input is subject to noise, for example, if the gradient cannot be evaluated directly, or due to some other exogenous factor. In such scenario, one 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 inclusion (SDI) as a stochastic counterpart of (DI)

$$\begin{cases} dX(t) \in -\partial F(X(t)) + \sigma(t, X(t))dW(t), \quad t \geq t_0; \\ X(t_0) = X_0, \end{cases} \tag{SDI}$$

where the diffusion (volatility) term $\sigma : [t_0, +\infty[\times\mathbb{H} \to \mathcal{L}_2(\mathbb{K}; \mathbb{H})$ (see notation in Section 2) is a measurable function, $\mathbb{K}$ a separable real Hilbert space, and W is a $\mathbb{K}$-valued cylindrical Brownian motion (see Section A.2.1 for a precise definition), and the initial data X_0 is a properly measurable $\mathbb{H}$-valued random variable. This dynamic can be viewed as a stochastic dissipative system that aims to minimize F if the diffusion term vanishes sufficiently fast. Also, it is the natural extension to the non-smooth setting of the work done in [30].

An important aspect of our work concerns the Tikhonov regularization of (DI) and (SDI). Given $t_0 > 0$, and a regularization parameter $\varepsilon : [t_0, +\infty[\to \mathbb{R}_+$, which is a measurable function that vanishes asymptotically in a controlled way, the Tikhonov regularization of (DI) is written:

$$\begin{cases} \dot{x}(t) \in -\partial F(x(t)) - \varepsilon(t)x(t), \quad t > t_0; \\ x(t_0) = x_0. \end{cases} \tag{DI-TA}$$

The stochastic counterpart of (DI-TA) (which is the Tikhonov regularization of (SDI)), is the following stochastic differential inclusion with initial data $X_0 \in \mathrm{L}^\nu(\Omega; \mathbb{H})$ (for some $\nu \geq 2$):

$$\begin{cases} dX(t) \in -\partial F(X(t)) - \varepsilon(t)X(t) + \sigma(t, X(t))dW(t), \quad t > t_0; \\ X(t_0) = X_0. \end{cases} \tag{SDI − TA}$$

It is well-known that in the deterministic case of (DI-TA), the Tikhonov regularization ensures that the trajectory generated by the system converges strongly to a particular minimizer of F: the one of minimum norm; see [3, 18] and references therein. The fact that the Tikhonov regularization parameter $\varepsilon(t)$ tends to zero not too fast as $t \to +\infty$ induces a hierarchical minimization property: the limit of any trajectory no longer depends on the initial data, it is precisely the minimum norm solution.

It is our aim in this paper to extend these results to the stochastic case (SDI − TA) based on the recent work of Maulen-Soto, Fadili, and Attouch [30]. More precisely, our objective is to study the dynamics (SDI) and (SDI − TA) and their long-time behavior in order to solve (P). If the diffusion term vanishes with time, one would expect to solve (P) with our dynamics and obtain for (SDI − TA) the hierarchical minimization property described above.

Motivated by this, our paper will primarily focus on the case where $\sigma(\cdot, x)$ vanishes sufficiently fast as $t \to +\infty$ uniformly in x. Additionally, we will provide some guarantees for uniformly bounded σ. Therefore, throughout the paper, we assume that σ satisfies:

$$\begin{cases} \sup_{t \geq t_0, x \in \mathbb{H}} \|\sigma(t, x)\|_{\mathrm{HS}} < +\infty, \\ \|\sigma(t, x') - \sigma(t, x)\|_{\mathrm{HS}} \leq L_0 \|x' - x\|, \end{cases} \tag{H}$$

for some $L_0 > 0$ and for all $t \geq t_0, x, x' \in \mathbb{H}$ (where the HS-norm is defined in Section 2). The Lipschitz continuity assumption is mild and required to ensure the well-posedness of (SDI) and (SDI − TA).

1.2 Contributions

This work goes well beyond that of [30] in three directions: we consider the non-smooth case, in infinite dimensional Hilbert spaces, and with Tikhonov regularization. The latter makes it possible to pass from weak convergence to strong convergence, and to a particular solution, that of minimal norm.

We first study the properties of the process $X(t)$ and $F(X(t))$ for the stochastic differential inclusion (SDI) on separable real Hilbert spaces from an optimization perspective, under the assumptions (H_0), (H) and (H_λ) (introduced in Section 3). When the diffusion term is uniformly bounded, we show convergence of $\mathbb{E}[F(X(t))-\min F]$ to a noise-dominated region both for the convex and strongly convex case. When the diffusion term is square-integrable, we show in Theorem 11 that $X(t)$ weakly converges almost surely to a solution of (P), which is a new result to the best of our knowledge. Moreover, in Theorem 12, we provide new ergodic and pointwise convergence rates of the objective in expectation, again, for both the convex and strongly convex case.

Next, we consider (SDI − TA), obtained by adding a Tikhonov regularization term to (SDI). We show in Theorem 14 that under certain conditions on the regularization term, $X(t)$ strongly converges almost surely to the minimum norm solution. Then, we show in Theorem 24 some practical situations where one can obtain an explicit form of the Tikhonov regularizer. Moreover, in Theorem 28, we show new convergence rates of the objective and the trajectory in expectation for the smooth case.

Table 1 summarizes the convergence rates obtained for $\mathbb{E}[F(X(t))-\min F]$. We use the following notation, $F=f+g$, $\sigma_*>0$ and $\sigma_\infty(\cdot)$ is defined as

$$\sigma_\infty(t) \stackrel{\text{def}}{=} \sup_{x\in\mathbb{H}} \|\sigma(t,x)\|_{\mathrm{HS}}, \quad \text{where} \quad \|\sigma(t,x)\|_{\mathrm{HS}}^2 \le \sigma_*^2, \quad \forall t\ge t_0, \forall x\in\mathbb{H}. \tag{1}$$

Property of F	**DI**	**SDI** ($\sup_{t\ge t_0}\sigma_\infty(t)\le\sigma_*$)	**SDI** ($\sigma_\infty\in\mathrm{L}^2([t_0,+\infty[)$)
Convex	t^{-1}	$t^{-1}+\sigma_*^2$	t^{-1}
μ-Strongly Convex	$e^{-2\mu t}$	$e^{-\mu t}+\sigma_*^2$	$\max\{e^{-\mu t},\sigma_\infty^2(t)\}$

Table 1 Summary of convergence rates obtained for $\mathbb{E}[F(X(t))-\min F]$.

We also denote $\mathrm{EB}^p(\mathcal{S})$ the local Error Bound Inequality defined in (30). In Table 2, we summarize the results obtained in the smooth case for the dynamics with Tikhonov regularization, *i.e.*, when $g\equiv 0$.

Property of f	**DI-TA** ($\varepsilon(t)=t^{-r}, r\in]0,1[$)	**SDI-TA** $\left(\varepsilon(t)=t^{-r}, r\in]\frac{2p}{2p+1},1[\right)$
Convex $\cap\,\mathrm{EB}^p(\mathcal{S})$	t^{-r}	t^{-r} whenever $\sigma_\infty^2(t)=\mathcal{O}(t^{-2r})$.

Table 2 Summary of convergence rates obtained for $\mathbb{E}[f(X(t))-\min f]$ for the dynamics with Tikhonov regularization when $\varepsilon(t)=t^{-r}$

1.3 Relation to prior work

The subgradient flow dynamic (DI), which is valid on a general real Hilbert space, is a dissipative dynamical system, whose study dates back to Cauchy [16]. It plays a fundamental role in optimization: it transforms the problem of minimizing F into the study of the asymptotic behavior of the trajectories of (DI). Its Euler forward discretization (with stepsize $\gamma_k>0$) is the subgradient method

$$x_{k+1}\in x_k-\gamma_k\partial F(x_k). \tag{Sub-G}$$

Or equivalently,

$$x_{k+1}=x_k-\gamma_k g_k, \tag{2}$$

where $g_k\in\partial F(x_k)$ for every $k\in\mathbb{N}$.

Let us focus on the finite-dimensional case ($\mathbb{H}=\mathbb{R}^d$). In [41, 1] they give conditions on the function and the stepsize to converge to within some range of the optimal value and to the optimal value. Despite (Sub-G) being a classical algorithm to solve the non-smooth convex minimization problem, it is not recommended for general use, as discussed in [28, 26]. Moreover, with the need to handle large-scale problems (such as in various areas of data science and machine learning), it has become necessary to find ways to get around the high computational cost per iteration that these problems entail. The Robbins-Monro stochastic approximation algorithm [37] is at the heart of Stochastic Gradient Descent methods, which, roughly speaking, consists in cheaply and randomly approximating the gradient at the price of obtaining a random noise in the solutions. In [23] they propose the natural generalization to the non-smooth setting, the stochastic subgradient method (S-Sub-G) that updates the iterates according to

$$x_{k+1} \in x_k - \gamma_k(\partial F(x_k) + \xi_k), \tag{S-Sub-G}$$

where ξ_k denotes the (random) noise term on the subgradient at the k-th iteration, and $\mathbb{E}[\xi_k] = 0$.

The SDI continuous-time approach is motivated by its relations to (S-Sub-G), where the latter can be viewed as an Euler forward time discretization, and the noise $\xi_k \sim \mathcal{N}(0, \sigma_k I_d)$ (hence not necessarily bounded). The advantage of the continuous-time perspective is that it offers a deep insight and unveils the key properties of the dynamic, without being tied to a specific discretization.

We extend the work of [30] to the case where the objective is "smooth+non-smooth", being able to show the almost sure weak convergence of the trajectory to the set of minimizers and new convergence rates for the objective in the convex and strongly convex case.

Besides, based on the work of [2], we add a Tikhonov term that let us obtain the almost sure strong convergence of the trajectory to the minimal norm solution. Moreover, we extend the convergence rates shown in [2, Theorem 5] to the stochastic case. In our way, we even prove new and useful results for the deterministic setting (e.g., Proposition 23 and Corollary 26).

While the use of Lyapunov analysis and vanishing Tikhonov regularization are known techniques in the deterministic case [2], their adaptation to the stochastic setting requires significant technical work and novel arguments. One has not only to handle carefully stochasticity through proper Itô's calculus, but also non-smoothness of the objective function.

1.4 Organization of the paper

Section 2 introduces notations and reviews some necessary material from convex and stochastic analysis. Section 3 states our main convergence results of (SDI) in the case of a convex objective function under (H_0) and with an extra assumption on the non-smooth term. We first show the almost sure weak convergence of the process towards the set of minimizers when the diffusion term is square-integrable, then we establish convergence rates for the values. Section 4 introduces an extra vanishing term called Tikhonov regularizer that let us obtain the almost sure strong convergence of (SDI − TA) to the minimal norm solution. Then we give some practical situations where we can obtain an explicit tuning of the Tikhonov regularizer. Finally in this section, we present convergence rates for the values and for the trajectory in the smooth case. Technical lemmas and theorems that are needed throughout the paper will be collected in the appendix A.

2 Notation and Preliminaries

We will use the following shorthand notations: Given $n \in \mathbb{N}$, $[n] \stackrel{\text{def}}{=} \{1, \dots, n\}$. 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 from $\mathbb{H}$ to $\mathbb{H}$. $\mathcal{L}(\mathbb{K}; \mathbb{H})$ 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

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

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 $M \in \mathcal{L}_2(\mathbb{K}; \mathbb{H})$,

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

is its Hilbert-Schmidt norm (in the finite-dimensional case is equivalent to the Frobenius norm). We denote by w-lim (resp. s-lim) the limit for the weak (resp. strong) 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 $[f \leq r] \stackrel{\text{def}}{=} \{x \in \mathbb{H} : f(x) \leq r\}$. 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 $p = +\infty$. 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 [38].

We denote by $\Gamma_0(\mathbb{H})$ the class of proper lsc and convex functions on $\mathbb{H}$ taking values in $\mathbb{R} \cup \{+\infty\}$. For $\mu > 0$, $\Gamma_\mu(\mathbb{H}) \subset \Gamma_0(\mathbb{H})$ is the class of μ-strongly convex functions, roughly speaking, this means that there exists a quadratic lower bound on the growth of these functions. 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 *subdifferential* of a function $f \in \Gamma_0(\mathbb{H})$ is the set-valued operator $\partial f : \mathbb{H} \rightrightarrows \mathbb{H}$ such that, for every x in $\mathbb{H}$,

$$\partial f(x) = \{u \in \mathbb{H} : f(y) \geq f(x) + \langle u, y - x \rangle \quad \forall y \in \mathbb{H}\}.$$

When f is continuous, $\partial f(x)$ is a non-empty convex and compact set for every $x \in \mathbb{H}$. If f is differentiable, then $\partial f(x) = \{\nabla f(x)\}$. For every $x \in \mathbb{H}$ such that $\partial f(x) \neq \emptyset$, the minimum norm selection of $\partial f(x)$ is the unique element $\{\partial^0 f(x)\} \stackrel{\text{def}}{=} \operatorname{argmin}_{u \in \partial f(x)} \|u\|$.

The projection of a point $x \in \mathbb{H}$ onto a closed convex set $C \subseteq \mathbb{H}$ is denoted by $\mathrm{P}_C(x)$.

2.1 Deterministic results on the subgradient flow with Tikhonov regularization

Let us first recall some basic facts about the deterministic case. To solve (P), a fundamental dynamic to consider is the subgradient flow of F, *i.e.*the following differential inclusion:

$$\dot{x}(t) \in -\partial F(x(t)). \tag{DI}$$

It is well known since the founding papers of Brézis, Baillon, and Bruck in the 1970s that, if the solution set $\operatorname{argmin}(F)$ of (P) is non-empty and F is convex, lower semicontinuous (lsc) and proper, then each solution trajectory of (DI) converges weakly, and its weak limit belongs to $\operatorname{argmin}(F)$.

In general, the limit solution depends on the initial data and is a priori difficult to specify when one has a set of solutions not reduced to only one element. To remedy this difficulty we consider the differential inclusion with vanishing Tikhonov regularization, $\varepsilon(t) \to 0$ (denoted (DI-TA)) which gives

$$\dot{x}(t) + \partial F(x(t)) + \varepsilon(t)x(t) \ni 0. \tag{DI − TA}$$

To analyze the convergence properties of this dynamic, let us recall basic facts concerning the Tikhonov approximation (1963). It consists in approximating the convex minimization problem (possibly ill-posed)

$$(\mathcal{P}) \quad \min\{F(x) : \ x \in \mathcal{H}\},$$

by the strongly convex minimization problem ($\varepsilon > 0$)

$$(\mathcal{P})_\varepsilon \quad \min\left\{F(x) + \frac{\varepsilon}{2}\|x\|^2 : \ x \in \mathcal{H}\right\}$$

whose unique solution is denoted by x_ε. The following result was first obtained by Browder in 1966 [13, 14].

Theorem 1. *(Hierarchical minimization). Suppose that* $\mathcal{S}_F = \operatorname{argmin}(F) \neq \emptyset$. *Let* $x^\star = \mathrm{P}_{\mathcal{S}_F}(0)$. *Then,*

(i) $\|x_\varepsilon\| \leq \|x^\star\|$ *for all* $\varepsilon > 0$.

(ii) $\lim_{\varepsilon \to 0} \|x_\varepsilon - x^\star\| = 0$.

The system (DI − TA) is a special case of the general dynamic model

$$\dot{x}(t) + \partial F(x(t)) + \varepsilon(t)\nabla\Psi(x(t)) \ni 0 \tag{3}$$

which involves two functions F and Ψ intervening with different time scale. When $\varepsilon(\cdot)$ tends to zero moderately slowly, it was shown in [4] that the trajectories of (3) converge asymptotically to equilibria that are solutions of the following hierarchical problem: they minimize the function Ψ on the set of minimizers of F. The continuous and discrete-time versions of these systems have a natural connection to the best response dynamics for potential games, domain decomposition for PDEs, optimal transport, and coupled wave equations. In the case of the Tikhonov approximation, a natural choice is to take $\Psi(x) = \|x - x_d\|^2$ where x_d is a desired state (which is also the continuous model of the Halpern method [43]). By doing so, we obtain asymptotically the closest possible solution to x_d. By translation, we can immediately reduce ourselves to the case $x_d = 0$, as considered in our work.

The following theorem establishes the convergence of the trajectories of (DI − TA) towards the minimum norm solution under minimal assumptions on the parameter $\varepsilon(t)$. We assume that (DI − TA) admits a unique strong global solution $x : [0, +\infty[\to \mathbb{H}$, namely, x is absolutely continuous on each compact interval such that (DI − TA) holds for almost every $t \geq 0$. Sufficient conditions for this well-posedness may be found in [12].

Theorem 2. *Suppose that* $\varepsilon : [t_0, +\infty[\to \mathbb{R}_+$ *is a measurable function that satisfies:*

(i) $\varepsilon(t) \to 0$ *as* $t \to +\infty$*;*

(ii) $\displaystyle\int_{t_0}^{+\infty} \varepsilon(t)dt = +\infty$.

Let $x(\cdot)$ *be a solution trajectory of the continuous dynamic* (DI − TA)*. Then,* $\text{s-lim}_{t\to+\infty} x(t) = x^\star \stackrel{\text{def}}{=} \mathrm{P}_{\mathcal{S}_F}(0)$.

This result was established in [18, Theorem 2]. For the reader's convenience, we give a self-contained short proof in Appendix A.3.

2.2 Stochastic differential equations

As said before, in many cases, the drift term is subject to noise. In such a scenario, one can model these errors using a stochastic integral with respect to the measure defined by a continuous Itô martingale. In the smooth case without Tikhonov regularization, this approach has been well documented in Maulen-Soto, Fadili, Attouch [30]. This concerns the following stochastic differential equation as the stochastic counterpart of the gradient flow, let $t_0 \geq 0$ and initial data $X_0 \in \mathrm{L}^\nu(\Omega; \mathbb{H})$ (for some $\nu \geq 2$):

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

Let us make precise the ingredients of this stochastic differential equation. It is defined over a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathcal{P})$, where the diffusion (volatility) term $\sigma : [t_0, +\infty[\times \mathbb{H} \to \mathcal{L}_2(\mathbb{K}; \mathbb{H})$ is a measurable function, and W is a $\mathbb{K}$-valued cylindrical Brownian motion.

Throughout this article, the diffusion term σ is assumed to satisfy (H). In connection with this assumption, let us define $\sigma_* > 0$ and $\sigma_\infty(\cdot)$ by

$$\|\sigma(t, x)\|_{\mathrm{HS}}^2 \leq \sigma_*^2, \quad \forall t \geq 0, \forall x \in \mathbb{H}, \quad \sigma_\infty(t) \stackrel{\text{def}}{=} \sup_{x\in\mathbb{H}} \|\sigma(t, x)\|_{\mathrm{HS}}, \tag{4}$$

and $\sigma_\infty(\cdot)$ is a decreasing function.

Concerning the study of (SDI) and (SDI − TA), let us recall the following result of [30, Theorem 3.1] on which we will build our study. It establishes almost sure weak convergence of $X(t)$ to an $\mathcal{S}$-valued random variable as $t \to +\infty$.

Theorem 3. *Consider the dynamic* (SDE) *where* f *and* σ *satisfy the assumptions* (H_0) *and* (H)*. Let* $\nu \geq 2$*, and its initial data* $X_0 \in \mathrm{L}^\nu(\Omega; \mathbb{H})$*. Then, there exists a unique solution* $X \in S^\nu_{\mathbb{H}}[t_0]$ *of* (SDE)*. Additionally, if* $\sigma_\infty \in \mathrm{L}^2([t_0, +\infty[)$*, then:*

(i) $\sup_{t\geq 0} \mathbb{E}[\|X(t)\|^2] < +\infty$.

(ii) $\forall x^\star \in \mathcal{S}$, $\lim_{t\to+\infty} \|X(t) - x^\star\|$ *exists a.s. and* $\sup_{t\geq 0} \|X(t)\| < +\infty$ *a.s.*

(iii) $\lim_{t\to\infty} \|\nabla f(X(t))\| = 0$ *a.s. As a result,* $\lim_{t\to\infty} f(X(t)) = \min f$ *a.s.*

(iv) *There exists an* $\mathcal{S}$*-valued random variable* $X^\star$ *such that* $\text{w-lim}_{t\to+\infty} X(t) = X^\star$ *a.s.*

► Remark 4. To be precise, [30, Theorem 3.1] treats the finite-dimensional case, however in [42, Chapter 3] the general separable real Hilbertian case was considered.

3 Stochastic differential inclusions

In this section, we will work with stochastic differential inclusions. For the history of this concept, we refer the reader to [25, Preface]. We will start by showing a general version of the (SDI) dynamic, formally describing what it means to be a solution of that dynamic, and then we will move on to show the conditions under which we can have the existence and uniqueness of a solution. Existence is due to [34] and uniqueness is proven here. Then we will focus on (SDI) and study the conditions on the diffusion term in order to ensure the almost sure weak convergence of the trajectory towards the set of minimizers. Finally, we will show some convergence rates of the objective under convexity or strong convexity.

3.1 Existence and uniqueness of solution

For a set-valued operator $A : \mathbb{H} \rightrightarrows \mathbb{H}$, its domain is $\mathrm{dom}(A) = \{x \in \mathbb{H} : A(x) \neq \emptyset\}$. For $t_0 \geq 0$, let $b : [t_0, +\infty[\times\mathbb{H} \to \mathbb{H}$ and $\sigma : [t_0, +\infty[\times\mathbb{H} \to \mathcal{L}_2(\mathbb{K}; \mathbb{H})$, and consider the general stochastic differential inclusion:

$$\begin{cases} dX(t) \in b(t, X(t))dt - A(X(t))dt + \sigma(t, X(t))dW(t), \quad t > 0 \\ X(t_0) = X_0, \end{cases} \tag{SDI$_0$}$$

defined over a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq t_0}, \mathbb{P})$, where the diffusion (volatility) term $\sigma : [t_0, +\infty[\times\mathbb{H} \to \mathcal{L}_2(\mathbb{K}; \mathbb{H})$ is a measurable function; W is a $\mathcal{F}_t$-adapted $\mathbb{K}$-valued cylindrical Brownian motion; and the initial data X_0 is an $\mathcal{F}_0$-measurable $\mathbb{H}$-valued random variable.

Definition 5. *A solution of* (SDI$_0$) *is a couple* (X, η) *of* $\mathcal{F}_t$*-adapted processes such that almost surely:*

(i) *X is continuous with sample paths in the domain of A;*

(ii) *η is absolutely continuous, such that $\eta(t_0) = 0$, and $\forall T > t_0$, $\eta' \in \mathrm{L}^2([t_0, T]; \mathbb{H})$, $\eta'(t) \in A(X(t))$ for almost all $t \geq t_0$. These properties are collectively denoted by $\eta \in \mathcal{A}_{loc}([t_0, +\infty[; \mathbb{H})$;*

(iii) *For $t > t_0$,*

$$\begin{cases} X(t) &= X_0 + \int_{t_0}^t b(s, X(s))ds - \eta(t) + \int_{t_0}^t \sigma(s, X(s))dW(s), \\ X(t_0) &= X_0. \end{cases} \tag{5}$$

For the sake of brevity, we sometimes omit the process η and say that X is a solution of (SDI$_0$), meaning that, there exists a process η such that (X, η) satisfies the previous definition. The definition of uniqueness for the process X will be presented in Section A.2.1.

▶ Remark 6. There are different notions of solution to the SDI; see, for instance, [33, 17], where the process η is only assumed to be continuous, adapted, and of bounded variation. However, to make use of the tools developed in this work, we adopt the approach of [34], which considers an adapted and absolutely continuous process. In any case, we will show that, in our setting, existence and uniqueness of a solution hold under the notion we adopt.

Throughout the paper it will be assumed that:

$$\begin{cases} A \text{ is a maximal monotone operator with closed domain;} \\ \mathcal{S} \stackrel{\text{def}}{=} A^{-1}(0) \neq \emptyset. \end{cases} \tag{H$_0(A)$}$$

$$\begin{cases} \exists L > 0, \|b(t, x) - b(t, y)\| \vee \|\sigma(t, x) - \sigma(t, y)\|_{\mathrm{HS}} \leq L\|x - y\|, \forall t \geq t_0, \forall x, y \in \mathbb{H}; \\ \sup_{t\geq t_0}(\|b(t, 0)\| \vee \|\sigma(t, 0)\|_{\mathrm{HS}}) < +\infty. \end{cases} \tag{H$_0(b,\sigma)$}$$

The Lipschitz continuity assumption is mild and required to ensure the well-posedness of (SDI$_0$).

We are interested in ensuring the existence and uniqueness of a solution for (SDI$_0$). Although there are several works that deal with the subject of stochastic differential inclusions (see [25, 10, 8, 35, 34, 22]), those of [34, 22] are the closest to our setting and define a solution in the sense of Definition 5, thus generalizing the work of Brézis [12] in the deterministic case to the stochastic setting. In this paper, we consider the sequence of solutions $\{X_\lambda\}_{\lambda>0}$ of the stochastic differential equations

$$\begin{cases} dX_\lambda(t) = b(t, X(t))dt - A_\lambda(X_\lambda(t))dt + \sigma(t, X_\lambda(t))dW(t), \quad t > t_0 \\ X_\lambda(t_0) = X_0, \end{cases} \tag{SDE$_\lambda$}$$

where $A_\lambda = (I - (I + \lambda A)^{-1})/\lambda$ is the Yosida approximation of A with parameter $\lambda > 0$. Under $(\mathrm{H}_0(A))$ and $(\mathrm{H}_0(b, \sigma))$, as well as the integrability condition

$$\limsup_{\lambda \downarrow 0} \int_{t_0}^{T} \mathbb{E}(\|A_\lambda(X_\lambda(t))\|^2) dt < +\infty, \tag{H_λ}$$

it was shown in [34, Theorem 3.5] that there exists a couple (X, η) of stochastic processes such that for every $T > t_0$,

$$\lim_{\lambda \downarrow 0} \mathbb{E}\left(\sup_{t \in [t_0, T]} \|X_\lambda(t) - X(t)\|^2\right) = 0, \quad \lim_{\lambda \downarrow 0} \mathbb{E}\left(\sup_{t \in [t_0, T]} \|\eta_\lambda - \eta\|^2\right) = 0,$$

where $\eta_\lambda(t) = \int_{t_0}^{t} A_\lambda(X_\lambda(s)) ds$, and that (X, η) is a solution of (SDI_0) in the sense of Definition 5. Moreover, one can even have a.s. strong convergence of the process X_λ to X when the diffusion term is state-independent; see [34, Proposition 6.3].

▶ Remark 7. In view of [34, Proposition 3.1] and Fatou's lemma, for (H_λ) to hold, it is sufficient that the Yosida approximation A_λ obeys a linear growth condition of the form $\limsup_{\lambda \downarrow 0} \|A_\lambda(x)\| \leq C(1 + \|x\|)$ for all $x \in \mathbb{H}$. For instance, in light of [9, Proposition 23.43(i)], the last linear growth condition holds if $\mathrm{dom}(A) = \mathbb{H}$ and $\|A^0(x)\| \leq C(1 + \|x\|)$ for all $x \in \mathbb{H}$, where $A^0(x)$ is the minimal norm element of $A(x)$. Note that to control the linear growth of $A_\lambda(x)$ through that of $A^0(x)$, [9, Corollary 23.46(ii)] tells us that the domain condition $\mathrm{dom}(A) = \mathbb{H}$ cannot be removed since $\|A_\lambda(x)\| \uparrow +\infty$ as $\lambda \downarrow 0$ for $x \in \mathbb{H} \setminus \mathrm{dom}(A)$. A typical case of interest in the setting of non-smooth optimization, which is at the heart of this paper, is when A is the subdifferential of a globally Lipschitz continuous convex function. By classical properties of the Yosida approximation, we know that

$$(\partial g(x))_\lambda = \nabla g_\lambda(x) = \frac{1}{\lambda}(x - \mathrm{prox}_{\lambda g}(x)),$$

where g_λ is the Moreau envelope of g with parameter $\lambda > 0$. If g is convex, continuous, and its subgradient has linear growth, then (H_λ) is satisfied by ∂g. These examples cover a wealth of functions encountered in practice such as in machine learning and signal processing.

We insist, however, on the fact that these conditions are only sufficient but not necessary and (H_λ) can be verified beyond this case; see for instance the product structure of A and σ studied in [27] and specialized to normal cones in [34, Remark 3.3(ii)].

Although the existence of a solution to (SDI_0) was established in [34], uniqueness was not addressed. In the following theorem, we provide conditions under which uniqueness also holds for such SDI. The proof is given in Subsection A.4.

Theorem 8. *Consider* (SDI_0)*, where A and (b, σ) satisfy the assumption* $(\mathrm{H}_0(A))$ *and* $(\mathrm{H}_0(b, \sigma))$*, respectively. Additionally, suppose that A satisfy* (H_λ) *and let $\nu \geq 2$ such that $X_0 \in \mathrm{L}^\nu(\Omega; \mathbb{H})$ and is $\mathcal{F}_0$-measurable. Then, there exists a unique solution $(X, \eta) \in S_{\mathbb{H}}^{\nu}[t_0] \times \mathcal{A}_{loc}([t_0, +\infty[; \mathbb{H})$ of* (SDI_0).

Corollary 9. *Consider* (SDE_λ)*, where A and (b, σ) satisfy the assumption* $(\mathrm{H}_0(A))$ *and* $(\mathrm{H}_0(b, \sigma))$*, respectively. Additionally, let us consider that A satisfy* (H_λ) *and let $\nu \geq 2$ such that $X_0 \in \mathrm{L}^\nu(\Omega; \mathbb{H})$ and is $\mathcal{F}_0$-measurable. Then,*

$$\sup_{\lambda > 0} \mathbb{E}\left(\sup_{t \in [t_0, T]} \|X_\lambda(t)\|^\nu\right) < +\infty.$$

Proof. Since $A^{-1}(0) = A_\lambda^{-1}(0)$ and A_λ is monotone, we replace η' by $A_\lambda(X_\lambda)$ in the proof of Theorem 8, then we realize that the constant that bounds $\mathbb{E}\left(\sup_{t \in [t_0, T]} \|X_\lambda(t)\|^\nu\right)$ is independent from λ to conclude. ◀

Let us present our extension of Itô's formula for a multi-valued drift, which plays a central role in the study of SDIs.

Proposition 10. *Consider* (SDI_0) *under the assumptions of Theorem* (8)*. Let $(X, \eta) \in S_{\mathbb{H}}^{\nu}[t_0] \times \mathcal{A}_{loc}([t_0, +\infty[; \mathbb{H})$ be the unique solution of* (SDI_0)*, and let $\phi : [t_0, +\infty[\times \mathbb{H} \to \mathbb{R}$ be such that $\phi(\cdot, x) \in C^1([t_0, +\infty[)$ for every $x \in \mathbb{H}$ and $\phi(t, \cdot) \in C^2(\mathbb{H})$ for every $t \geq t_0$. Then the process $Y(t) = \phi(t, X(t))$, is an Itô Process such that for all $t \geq 0$*

$$Y(t) = Y(t_0) + \int_{t_0}^{t} \frac{\partial \phi}{\partial t}(s, X(s))ds + \int_{t_0}^{t} \langle \nabla\phi(s, X(s)), b(s, X(s)) - \eta'(s)\rangle\, ds \\ + \int_{t_0}^{t} \langle \sigma^\star(s, X(s))\nabla\phi(s, X(s)), dW(s)\rangle + \frac{1}{2}\int_{t_0}^{t} \mathrm{tr}[\sigma(s, X(s))\sigma^\star(s, X(s))\nabla^2\phi(s, X(s))]ds, \quad (6)$$

where $\eta'(t) \in A(X(t))$ a.s. for almost all $t \geq t_0$. Moreover, if $\mathbb{E}[Y(t_0)] < +\infty$, and if for all $T > t_0$

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

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

$$\mathbb{E}[Y(t)] = \mathbb{E}[Y(t_0)] + \mathbb{E}\left(\int_{t_0}^{t} \frac{\partial\phi}{\partial t}(s, X(s))ds\right) + \mathbb{E}\left(\int_{t_0}^{t} \langle \nabla\phi(s, X(s)), b(s, X(s)) - \eta'(s)\rangle\, ds\right) \\ + \frac{1}{2}\mathbb{E}\left(\int_{t_0}^{t} \mathrm{tr}[\sigma(s, X(s))\sigma^\star(s, X(s))\nabla^2\phi(s, X(s))]ds\right). \quad (7)$$

Proof. The unique solution $(X, \eta) \in S^\nu_{\mathbb{H}}[t_0] \times \mathcal{A}_{loc}([t_0, +\infty[; \mathbb{H})$ of (SDI_0) satisfies (by definition) the following equation:

$$\begin{cases} X(t) &= X_0 + \int_{t_0}^{t}[b(s, X(s)) - \eta'(s)]ds + \int_{t_0}^{t} \sigma(s, X(s))dW(s), \quad t > t_0, \\ X(t_0) &= X_0. \end{cases} \quad (8)$$

and $\eta'(s) \in A(X(s))$ for almost all $t \geq 0$ a.s.. Then, (8) is an Itô process with drift $s \mapsto b(s, X(s)) - \eta'(s)$ and diffusion $s \mapsto \sigma(s, X(s))$. Consequently, we can apply the classical Itô's formula (see [20, Section 2.3]) to obtain the desired. ◀

3.2 Almost sure weak convergence of the trajectory

We consider $f + g$ (called the potential) and study the dynamic (SDI) under the hypotheses (H_0) (*i.e.* $f \in C^{1,1}_L(\mathbb{H}) \cap \Gamma_0(\mathbb{H}), g \in \Gamma_0(\mathbb{H})$) and (H). Recall the definitions of σ_* and $\sigma_\infty(t)$ from (1). Throughout the rest of the paper, we use the notation:

$$\begin{aligned} F(x) &\stackrel{\text{def}}{=} f(x) + g(x), \\ \Sigma(t, x) &\stackrel{\text{def}}{=} \sigma(t, x)\sigma(t, x)^\star, \\ \mathcal{S}_F &\stackrel{\text{def}}{=} \mathrm{argmin}(F). \end{aligned}$$

Our first main result establishes almost sure weak convergence of $X(t)$ to a random variable taking values in $\mathcal{S}_F$. The proof relies on Lyapunov-type arguments and the use of Barbalat's Lemma and Opial's Lemma. While these tools are classical in the deterministic setting, applying them in a stochastic framework requires particular care. In our case, this involves working almost surely and leveraging the properties of Itô's formula and Robbins-Siegmund lemma (see Theorem 39). The overall strategy aligns with the approach used in [30, Theorem 3.1] (restated in Theorem 3), and extends that result to the non-smooth setting.

The main challenge in this extension lies in justifying Itô's formula and handling the selection curve $\eta'(t) \in \partial g(X(t))$ a.s. The former is addressed through our specific notion of solution (see Definition 5 and Proposition 10), while the latter is of a more technical nature and is treated in detail in the proof. We now state the result precisely.

Theorem 11. *Consider $F = f + g$ and σ satisfying* (H_0) *and* (H) *respectively. Suppose further that ∂g verifies* (H_λ)*. Let $\nu \geq 2$, $t_0 \geq 0$, and consider the dynamic* (SDI) *with initial data $X_0 \in \mathrm{L}^\nu(\Omega; \mathbb{H})$, i.e.:*

$$\begin{cases} dX(t) &\in -\partial F(X(t))dt + \sigma(t, X(t))dW(t), \\ X(t_0) &= X_0, \end{cases} \quad (9)$$

where W is a $\mathbb{K}$-valued cylindrical Brownian motion. Then, there exists a unique solution (in the sense of Theorem 8) $(X, \eta) \in S^\nu_{\mathbb{H}}[t_0] \times \mathcal{A}_{loc}([t_0, +\infty[; \mathbb{H})$.

Moreover, if $\sigma_\infty \in \mathrm{L}^2([t_0, +\infty[)$, then the following holds:

(i) $\mathbb{E}[\sup_{t\geq t_0}\|X(t)\|^{\nu}] < +\infty$.

(ii) $\forall x^\star \in \mathcal{S}_F$, $\lim_{t\to+\infty}\|X(t)-x^\star\|$ *exists a.s. and* $\sup_{t\geq t_0}\|X(t)\| < +\infty$ *a.s..*

(iii) *If g is continuous, then $\forall x^\star \in \mathcal{S}_F$, $\nabla f(x^\star)$ is constant,* $\lim_{t\to\infty}\|\nabla f(X(t)) - \nabla f(x^\star)\| = 0$ *a.s. for any $x^\star \in \mathcal{S}_F$,and*

$$\int_{t_0}^{+\infty} F(X(t)) - \min F \, dt < +\infty.$$

(iv) *If (iii) holds, then there exists an $\mathcal{S}_F$-valued random variable $X^\star$ such that* $\text{w-}\lim_{t\to+\infty} X(t) = X^\star$.

Proof. (i) Directly from Theorem 8.

(ii) Since F is convex, we first notice that $\mathcal{S}_F = (\partial F)^{-1}(0)$.

Now let us consider $(X,\eta) \in S^{\nu}_{\mathbb{H}}[t_0] \times \mathcal{A}_{loc}([t_0,+\infty[;\mathbb{H})$ be the unique solution of (SDI$_0$) given by Theorem 8, and $\phi(x) = \frac{\|x-x^\star\|^2}{2}$, where $x^\star \in \mathcal{S}_F$. Then by Itô's formula

$$\begin{aligned}\phi(X(t)) = &\underbrace{\frac{\|X_0 - x^\star\|^2}{2}}_{\xi=\phi(X_0)} + \underbrace{\frac{1}{2}\int_{t_0}^{t} \mathrm{tr}[\Sigma(s,X(s))ds]}_{A_t} - \underbrace{\int_{t_0}^{t}\langle \eta'(s) + \nabla f(X(s)), X(s) - x^\star\rangle \, ds}_{U_t} \\ &+ \underbrace{\int_{t_0}^{t}\langle \sigma^\star(s,X(s))\,(X(s)-x^\star)\,, dW(s)\rangle}_{M_t}. \end{aligned} \tag{10}$$

Let us observe that, since $\nu \geq 2$, we have that $\mathbb{E}(\sup_{t\geq t_0}\|X(t)\|^2) < +\infty$. Moreover, since $\sigma_\infty \in \mathrm{L}^2([t_0,+\infty[)$ we have

$$\mathbb{E}\left(\int_{t_0}^{+\infty}\|\sigma^\star(s,X(s))\,(X(s)-x^\star)\|^2\,ds\right) \leq \mathbb{E}\left(\sup_{t\geq t_0}\|X(t)-x^\star\|^2\right)\int_{t_0}^{+\infty}\sigma_\infty^2(s)ds < +\infty.$$

Therefore M_t is a square-integrable continuous martingale. It is also a continuous local martingale (see [29, Theorem 1.3.3]), which implies that $\mathbb{E}(M_t) = 0$.

Moreover, since F is a convex function, then ∂F is a monotone operator. On the other hand $\eta'(t) \in \partial g(X(t))$ a.s. for almost all $t \geq t_0$, so

$$\langle \eta'(t) + \nabla f(X(t)), X(t) - x^\star\rangle \geq 0, a.s.\text{for almost all } t \geq t_0.$$

We have that A_t and U_t defined as in (10) are two continuously adapted increasing processes with $A_0 = U_0 = 0$ a.s.. Since $\phi(X(t))$ is nonnegative and $\sup_{x\in\mathbb{H}}\|\sigma(\cdot,x)\|_{\mathrm{HS}} \in \mathrm{L}^2([t_0,+\infty[)$, we deduce that $\lim_{t\to+\infty} A_t < +\infty$. Then, we can use Theorem 39 to conclude that

$$\int_{t_0}^{+\infty}\langle \eta'(t) + \nabla f(X(t)), X(t) - x^\star\rangle dt < +\infty \quad a.s. \tag{11}$$

and

$$\forall x^\star \in \mathcal{S}_F, \exists \Omega_{x^\star} \in \mathcal{F}, \text{such that } \mathbb{P}(\Omega_{x^\star}) = 1 \text{ and } \lim_{t\to+\infty}\|X(\omega,t) - x^\star\| \text{ exists } \forall \omega \in \Omega_{x^\star}. \tag{12}$$

Since $\mathbb{H}$ is separable, there exists a countable set $\mathcal{Z} \subseteq \mathcal{S}_F$, such that $\mathrm{cl}(\mathcal{Z}) = \mathcal{S}_F$ (where cl stands for the closure of the set). Let $\widetilde{\Omega} = \bigcap_{z\in\mathcal{Z}}\Omega_z$. Since $\mathcal{Z}$ is countable, a union bound shows

$$\mathbb{P}(\widetilde{\Omega}) = 1 - \mathbb{P}\left(\bigcup_{z\in\mathcal{Z}}\Omega_z^c\right) \geq 1 - \sum_{z\in\mathcal{Z}}\mathbb{P}(\Omega_z^c) = 1.$$

For arbitrary $x^\star \in \mathcal{S}_F$, there exists a sequence $(z_k)_{k\in\mathbb{N}} \subseteq \mathcal{Z}$ such that $\lim_{k\to\infty} z_k = x^\star$. In view of (12), for every $k \in \mathbb{N}$ there exists $\tau_k : \Omega_{z_k} \to \mathbb{R}_+$ such that

$$\lim_{t\to+\infty}\|X(\omega,t) - z_k\| = \tau_k(\omega), \quad \forall \omega \in \Omega_{z_k}. \tag{13}$$

Now, let $\omega \in \widetilde{\Omega}$. Since $\widetilde{\Omega} \subset \Omega_{z_k}$ for any $k \in \mathbb{N}$, and using the triangle inequality and (13), we obtain that

$$\tau_k(\omega) - \|z_k - x^\star\| \leq \liminf_{t\to+\infty} \|X(\omega,t) - x^\star\| \leq \limsup_{t\to+\infty} \|X(\omega,t) - x^\star\| \leq \tau_k(\omega) + \|z_k - x^\star\|.$$

Now, passing to $k \to +\infty$, we deduce

$$\limsup_{k\to+\infty} \tau_k(\omega) \leq \liminf_{t\to+\infty} \|X(\omega,t) - x^\star\| \leq \limsup_{t\to+\infty} \|X(\omega,t) - x^\star\| \leq \liminf_{k\to+\infty} \tau_k(\omega),$$

whence we deduce that $\lim_{k\to+\infty} \tau_k(\omega)$ exists on the set $\widetilde{\Omega}$ of probability 1, and in turn

$$\lim_{t\to+\infty} \|X(\omega,t) - x^\star\| = \lim_{k\to+\infty} \tau_k(\omega).$$

Let us recall that there exists $\Omega_{\text{cont}} \in \mathcal{F}$ such that $\mathbb{P}(\Omega_{\text{cont}}) = 1$ and $X(\omega,\cdot)$ is continuous for every $\omega \in \Omega_{\text{cont}}$. Now let $x^\star \in \mathcal{S}_F$ arbitrary, since the limit exists, for every $\omega \in \widetilde{\Omega} \cap \Omega_{\text{cont}}$ there exists $T(\omega)$ such that $\|X(\omega,t) - x^\star\| \leq 1 + \lim_{k\to+\infty} \tau_k(\omega)$ for every $t \geq T(\omega)$. Besides, since $X(\omega,\cdot)$ is continuous, by Bolzano's theorem

$$\sup_{t\in[0,T(\omega)]} \|X(\omega,t)\| = \max_{t\in[0,T(\omega)]} \|X(\omega,t)\| \stackrel{\text{def}}{=} h(\omega) < +\infty.$$

Therefore, $\sup_{t\geq t_0} \|X(\omega,t)\| \leq \max\{h(\omega), 1 + \lim_{k\to+\infty} \tau_k(\omega) + \|x^\star\|\} < +\infty$.

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

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

According to Theorem 38, 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

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

On the other hand, since $x^\star \in (\partial F)^{-1}(0) = (\nabla f + \partial g)^{-1}(0)$, then $-\nabla f(x^\star) \in \partial g(x^\star)$. Let $T > t_0$ such that $\eta'(t) \in \partial g(X(t))$ a.s., consequently,

$$\begin{aligned}\langle \eta'(t) + \nabla f(X(t)), X(t) - x^\star\rangle &= \underbrace{\langle \eta'(t) - (-\nabla f(x^\star)), X(t) - x^\star\rangle}_{\geq 0} \\ &\quad + \langle \nabla f(X(t)) - \nabla f(x^\star), X(t) - x^\star\rangle \\ &\geq \langle \nabla f(X(t)) - \nabla f(x^\star), X(t) - x^\star\rangle \\ &\geq \frac{1}{L}\|\nabla f(X(t)) - \nabla f(x^\star)\|^2,\end{aligned}$$

where $\langle \eta'(t) - (-\nabla f(x^\star)), X(t) - x^\star\rangle \geq 0$ by monotonicity of ∂g. Then by (11) we obtain

$$\int_{t_0}^{+\infty} \|\nabla f(X(t)) - \nabla f(x^\star)\|^2\, dt < +\infty \quad a.s.. \tag{14}$$

Let $\Omega_{\text{HS}} \in \mathcal{F}$ be the event where (11) (and consequently (14)) is satisfied ($\mathbb{P}(\Omega_{\text{HS}}) = 1$). Let $\Omega_\eta \in \mathcal{F}$ be the event where $\eta'(t) \in \partial g(X(t))$ for almost all $T > t_0$ ($\mathbb{P}(\Omega_\eta) = 1$). Finally, let $\Omega_{\text{conv}} \stackrel{\text{def}}{=} \widetilde{\Omega} \cap \Omega_{\text{cont}} \cap \Omega_{\text{HS}} \cap \Omega_N \cap \Omega_\eta$, hence $\mathbb{P}(\Omega_{\text{conv}}) = 1$. Let also $\omega \in \Omega_{\text{conv}} \subseteq \Omega_{\text{HS}}$ arbitrary, then

$$\liminf_{t\to+\infty} \|\nabla f(X(\omega,t)) - \nabla f(x^\star)\| = 0.$$

If also

$$\limsup_{t\to+\infty} \|\nabla f(X(\omega,t)) - \nabla f(x^\star)\| = 0,$$

then we conclude with the proof. Suppose by contradiction that there exists $\omega_0 \in \Omega_{\text{conv}}$ such that

$$\limsup_{t\to+\infty} \|\nabla f(X(\omega_0,t)) - \nabla f(x^\star)\| > 0.$$

Then, by Lemma 34, there exists $\delta(\omega_0) > 0$ satisfying

$$0 = \liminf_{t\to+\infty} \|\nabla f(X(\omega_0,t)) - \nabla f(x^\star)\| < \delta(\omega_0) < \limsup_{t\to+\infty} \|\nabla f(X(\omega_0,t)) - \nabla f(x^\star)\|,$$

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

$$\|\nabla f(X(\omega_0,t_k)) - \nabla f(x^\star)\| > \delta(\omega_0) \text{ and } t_{k+1} - t_k > 1, \quad \forall k \in \mathbb{N}.$$

Additionally, consider $\eta'(\omega_0,t) \in \partial g(X(\omega_0,t))$ for almost all $T > t_0$. Since $\sup_{t\geq t_0} \|X(\omega_0,t)\| < +\infty$, ∂g is full domain, and the fact that ∂g maps bounded sets onto bounded sets, we have that there exists $C_\eta(\omega_0) \geq 0$ such that $\|\eta'(\omega_0,t)\|^2 \leq C_\eta(\omega_0)$ for almost all $T > t_0$.

We allow ourselves the abuse of notation $X(t) \stackrel{\text{def}}{=} X(\omega_0,t), \eta'(t) \stackrel{\text{def}}{=} \eta'(\omega_0,t), C_\eta \stackrel{\text{def}}{=} C_\eta(\omega_0)$ and $\delta \stackrel{\text{def}}{=} \delta(\omega_0)$ during the rest of the proof from this point.

Let

- $C_0 \stackrel{\text{def}}{=} C_\eta + \|\nabla f(x^\star)\|^2$;
- $C_1 \stackrel{\text{def}}{=} \frac{(2C_0+1)^2-1}{C_0} > 0$;
- $\varepsilon \in \left]0, \min\{\frac{\delta^2}{4L^2}, C_1\}\right[$;
- and $C(\varepsilon) \stackrel{\text{def}}{=} \frac{\sqrt{C_0\varepsilon+1}-1}{4C_0} \in]0, \frac{1}{2}]$.

Note that this choice entails that the intervals $([t_k, t_k + C(\varepsilon)])_{k\in\mathbb{N}}$ are disjoint. On the other hand, according to the convergence property of N_t and the fact that $\|\nabla f(X(t)) - \nabla f(x^\star)\| \in \mathrm{L}^2([t_0,+\infty[)$, there exists $k' > 0$ such that for every $k \geq k'$

$$\sup_{t\geq t_k} \|N_t - N_{t_k}\|^2 < \frac{\varepsilon}{4} \text{ and } \int_{t_k}^{+\infty} \|\nabla f(X(t)) - \nabla f(x^\star)\|^2 dt < 1.$$

Also, we compute

$$\begin{aligned}\int_{t_k}^{t} \|\eta'(s) + \nabla f(X(s))\|^2 ds &\leq 2\int_{t_k}^{t} \|\nabla f(X(s)) - \nabla f(x^\star)\|^2 ds + 2\int_{t_k}^{t} \|\eta'(s) + \nabla f(x^\star)\|^2 ds \\ &\leq 2 + 4C_0(t - t_k).\end{aligned}$$

Furthermore, $C(\varepsilon)$ was chosen such that $C(\varepsilon) + 2C_0C(\varepsilon)^2 \leq \frac{\varepsilon}{8}$. Besides for every $k \geq k'$, $t \in [t_k, t_k + C(\varepsilon)]$,

$$\begin{aligned}\|X(t) - X(t_k)\|^2 &\leq 2(t - t_k)\int_{t_k}^{t} \|\eta'(s) + \nabla f(X(s))\|^2 ds + 2\|N_t - N_{t_k}\|^2 \\ &\leq 4(t - t_k) + 8C_0(t - t_k)^2 + \frac{\varepsilon}{2} \leq \varepsilon.\end{aligned}$$

Since ∇f is L-Lipschitz and $L^2\varepsilon \leq \left(\frac{\delta}{2}\right)^2$ by assumption on ε, we have that for every $k \geq k'$ and $t \in [t_k, t_k + C(\varepsilon)]$

$$\|\nabla f(X(t)) - \nabla f(X(t_k))\|^2 \leq L^2 \|X(t) - X(t_k)\|^2 \leq \left(\frac{\delta}{2}\right)^2.$$

Therefore, for every $k \geq k'$, $t \in [t_k, t_k + C(\varepsilon)]$

$$\|\nabla f(X(t)) - \nabla f(x^\star)\| \geq \|\nabla f(X(t_k)) - \nabla f(x^\star)\| - \underbrace{\|\nabla f(X(t)) - \nabla f(X(t_k))\|}_{\leq \frac{\delta}{2}} \geq \frac{\delta}{2}.$$

Finally,

$$\begin{aligned}\int_{t_0}^{+\infty} \|\nabla f(X(s)) - \nabla f(x^\star)\|^2 ds &\geq \sum_{k\geq k'} \int_{t_k}^{t_k+C(\varepsilon)} \|\nabla f(X(s)) - \nabla f(x^\star)\|^2 ds \\ &\geq \sum_{k\geq k'} \frac{\delta^2 C(\varepsilon)}{4} = +\infty,\end{aligned}$$

which contradicts $\|\nabla f(X(\cdot)) - \nabla f(x^\star)\| \in \mathrm{L}^2([t_0, +\infty[)$. So, for every $\omega \in \Omega_{\mathrm{conv}}$,

$$\begin{aligned}\limsup_{t\to+\infty} \|\nabla f(X(\omega,t)) - \nabla f(x^\star)\| &= \liminf_{t\to+\infty} \|\nabla f(X(\omega,t)) - \nabla f(x^\star)\| \\ &= \lim_{t\to+\infty} \|\nabla f(X(\omega,t)) - \nabla f(x^\star)\| = 0.\end{aligned}$$

On the other hand, since F is convex, by (11), we obtain

$$\int_{t_0}^{+\infty} F(X(t)) - \min F \, dt < +\infty, \quad a.s.. \tag{15}$$

Since $\sup_{t\geq t_0} \|X(t)\| < +\infty$ a.s., and (∂F) maps bounded sets onto bounded sets (since g is convex and continuous), we can show that there exists $\tilde{L} > 0$ such that

$$|F(X(t_1)) - F(X(t_2))| \leq \tilde{L}\|X(t_1) - X(t_2)\|, \quad \forall t_1, t_2 \geq t_0, a.s..$$

Using the same technique as before, we can conclude that $\lim_{t\to+\infty} F(X(t)) = \min F$ a.s..

(iv) Let $\omega \in \Omega_{\mathrm{conv}}$ and $\widetilde{X}(\omega)$ be a weak sequential limit point of $X(\omega, t)$. 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

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

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 $\widetilde{X}(\omega) \in \mathcal{S}_F$. Finally, by Opial's Lemma (see [32]) we conclude that there exists $X^\star(\omega) \in \mathcal{S}_F$ such that

$$\operatorname*{w-lim}_{t\to+\infty} X(\omega, t) = X^\star(\omega).$$

In other words, since $\omega \in \Omega_{\mathrm{conv}}$ was arbitrary, there exists an $\mathcal{S}_F$-valued random variable $X^\star$ such that $\operatorname{w-lim}_{t\to+\infty} X(t) = X^\star$ a.s..

◀

3.3 Convergence rates of the objective

The following result, stated below, summarizes the global convergence rates in expectation satisfied by the trajectories of (SDI), and it is a natural extension of [30, Theorem 3.2] to the non-smooth setting.

Theorem 12. *Consider the dynamic* (SDI) *where* $F = f + g$ *and* σ *satisfy assumptions* (H_0) *and* (H). *Furthermore, assume that* ∂g *satisfies* (H_λ) *and that* $X_0 \in \mathrm{L}^2(\Omega; \mathbb{H})$ *and is* $\mathcal{F}_0$*-measurable. The following statements are satisfied by the unique solution trajectory* $X \in S^2_{\mathbb{H}}[t_0]$ *of* (SDI)*:*

(i) *Let* $\overline{F\circ X}(t) \stackrel{\text{def}}{=} t^{-1}\int_{t_0}^t F(X(s))ds$ *and* $\overline{X}(t) = t^{-1}\int_{t_0}^t X(s)ds$. *Then*

$$\mathbb{E}\left(F(\overline{X}(t)) - \min F\right) \leq \mathbb{E}\left(\overline{F\circ X}(t) - \min F\right) \leq \frac{\mathbb{E}\left(\mathrm{dist}(X_0, \mathcal{S}_F)^2\right)}{2t} + \frac{\sigma_*^2}{2}, \quad \forall t > t_0. \tag{16}$$

Besides, if σ_∞ *is* $\mathrm{L}^2([t_0, +\infty[)$, *then*

$$\mathbb{E}\left(F(\overline{X}(t)) - \min F\right) \leq \mathbb{E}\left(\overline{F\circ X}(t) - \min F\right) = \mathcal{O}\left(\frac{1}{t}\right). \tag{17}$$

(ii) *Moreover, if* $F \in \Gamma_\mu(\mathbb{H})$ *with* $\mu > 0$, *then* $\mathcal{S}_F = \{x^\star\}$ *and*

$$\mathbb{E}\left(\|X(t) - x^\star\|^2\right) \leq \mathbb{E}\left(\|X_0 - x^\star\|^2\right) e^{-\mu t} + \frac{\sigma_*^2}{\mu}, \quad \forall t > t_0. \tag{18}$$

Besides, if σ_∞ *is non-increasing and vanishes at infinity, then:*

$$\mathbb{E}\left(\|X(t) - x^\star\|^2\right) \leq \mathbb{E}\left(\|X_0 - x^\star\|^2\right) e^{-\mu t} + \frac{\sigma_*^2}{\mu} e^{\frac{\mu t_0}{2}} e^{-\frac{\mu t}{2}} + \sigma_\infty^2\left(\frac{t_0 + t}{2}\right), \quad \forall t > t_0. \tag{19}$$

Proof. By making use of the inequality

$$F(x) - \min F \leq \langle y, x - x^\star \rangle, \quad \text{for all } y \in \partial F(x) \text{ and } x^\star \in \mathcal{S}_F, \tag{20}$$

which is given by the convexity of F, and considering the anchor function defined by

$$\phi(x) = \frac{1}{2}\|x - x^\star\|^2, \quad \text{for } x^\star \in \mathcal{S}_F,$$

we apply Itô's formula to $\phi(X(t))$ and take expectation. Combining (20) with the result of Itô's formula leads to an integral equation governing the expected behavior of $\phi(X(t))$, which can subsequently be leveraged to establish convergence rates in expectation of the process. Since the arguments are essentially identical to those of [30, Theorem 3.2], we refrain from reproducing the details here. ◀

► Remark 13. In the deterministic setting, *i.e.*, (DI), the quantity $F(\overline{x}(t)) - \min F$ decreases monotonically and one directly obtains the standard $\mathcal{O}(1/t)$, besides when F is strongly convex, we have linear convergence rate of the distance to the unique minimizer, these convergence rates are obtained defining the same Lyapunov functions as in Theorem 12, *i.e.*, $\phi(x) = \frac{1}{2}\|x - x^\star\|^2$ for $x^\star \in \mathcal{S}_F$. By contrast, for the stochastic differential inclusion (SDI), the main challenge we must overcome is that the term $\sigma(t, X(t))dW(t)$ generates both a martingale difference noise and a quadratic variation term. We note that the martingale difference noise term vanishes after taking expectation and that the quadratic variation term is controlled by uniform bounds on σ, leading to the bias terms $\frac{\sigma_*^2}{2}$ (and $\frac{\sigma_*^2}{\mu}$ under strong convexity), moreover when σ_∞ is square integrable, the bias terms vanishes over time. These challenges, and the way they are addressed, are essential for extending the classical ODE-based analysis to the stochastic setting. Naturally, setting $\sigma \equiv 0$ in (SDI) eliminates the stochastic terms, and all the estimates in Theorem 12 reduce exactly to the classical bounds known for the deterministic subgradient flow.

4 Tikhonov regularization: Convergence properties for convex functions

It is important to provide insight into the technique of Tikhonov regularization. This allows us to pass from the almost sure weak convergence towards the set of minimizers of the trajectory generated by (SDI_0) to achieving almost sure strong convergence of the trajectory generated by $(\mathrm{SDI} - \mathrm{TA})$, not only towards the set of minimizers but to the minimal norm solution. The trade-off in order to achieve this is the proper tuning of the Tikhonov parameter that depends on a local constant that could be hard to compute, besides that, we obtain slower convergence rates of the objective, passing from $\mathcal{O}(t^{-1})$ to $\mathcal{O}(t^{-r} + R(t))$, where $r < 1$ and $R(t) \to 0$ (defined below in (37)).

4.1 Almost sure convergence of the trajectory to the minimal norm solution

Our second main result establish almost sure convergence of $X(t)$ to $x^\star = \mathrm{P}_{\mathcal{S}_F}(0)$ as $t \to +\infty$. It is based on a subtle tuning of the Tikhonov parameter $\varepsilon(t)$ formulated as conditions (T_1), (T_2), and (T_3) below. We know that $\|x^\star\|^2 - \|x_{\varepsilon(t)}\|^2$ tends to zero as $t \to +\infty$. By capitalizing on Proposition 23, we shall see in Theorem 24 that the conditions (T_1), (T_2), and (T_3) are compatible for functions verifying Hölderian-type error bounds, which is the case for Łojasiewicz functions (see Definition 16 and Proposition 18).

Theorem 14. *Consider the dynamic* $(\mathrm{SDI} - \mathrm{TA})$ *where* $F = f + g$ *and* σ *satisfy the assumptions* (H_0) *and* (H)*, respectively, furthermore assume that* ∂g *satisfy* (H_λ)*. Let* $\nu \geq 2$*, and its initial data* $X_0 \in \mathrm{L}^\nu(\Omega; \mathbb{H})$*. Then, there exists a unique solution* $X \in S_{\mathbb{H}}^\nu[t_0]$ *of* $(\mathrm{SDI} - \mathrm{TA})$*. Let* $x^\star \stackrel{\text{def}}{=} \mathrm{P}_{\mathcal{S}_F}(0)$ *be the minimum norm solution, and for* $\varepsilon > 0$ *let* x_ε *be the unique minimizer of* $F_\varepsilon(x) \stackrel{\text{def}}{=} F(x) + \frac{\varepsilon}{2}\|x\|^2$*. Suppose that* $\sigma_\infty \in \mathrm{L}^2([t_0, +\infty[)$*, and that* $\varepsilon : [t_0, +\infty[\to \mathbb{R}_+$ *satisfies the conditions:*

(T_1) $\varepsilon(t) \to 0$ *as* $t \to +\infty$*;*

(T_2) $\displaystyle\int_{t_0}^{+\infty} \varepsilon(t)dt = +\infty$*;*

(T_3) $\displaystyle\int_{t_0}^{+\infty} \varepsilon(t)\left(\|x^\star\|^2 - \|x_{\varepsilon(t)}\|^2\right)dt < +\infty.$

Then we have

(i) $\int_{t_0}^{+\infty} \varepsilon(t)\mathbb{E}[\|X(t)-x^\star\|^2]dt < +\infty.$

(ii) $\lim_{t\to+\infty}\|X(t)-x^\star\|$ *exists a.s. and* $\sup_{t\geq t_0}\|X(t)\| < +\infty$ *a.s.*.

(iii) $\int_{t_0}^{+\infty} \varepsilon(t)\|X(t)-x^\star\|^2 dt < +\infty$ *a.s.*.

(iv) $\text{s-}\lim_{t\to+\infty} X(t) = x^\star$ *a.s.*

▶ Remark 15. We note that assumptions (T_1) and (T_2) are the same as those required in the deterministic setting, namely in Theorem 2. However, assumption (T_3) is new and can be viewed as the price for moving to the stochastic case. This term appears in the proof of Theorem 2, specifically on the right-hand side of (40). While we could recover Theorem 2 by directly assuming (T_3), this is unnecessary in the deterministic case. In the latter, one deals with the differential inequality (40) to get the desired conclusion without relying on (T_3). In contrast, in the stochastic setting, we obtain an integral inequality that cannot be handled in the same way, as the stochasticity removes some of the monotonicity structure inherent to the deterministic problem.

Proof. The existence and uniqueness of a solution $X \in S^{\nu}_{\mathbb{H}}[t_0]$ follow directly from the fact that the conditions of Theorem 8 are satisfied under (H_0) and (H). The only subtlety to check is that $\sup_{t\geq t_0}|\varepsilon(t)| < +\infty$, but this can be assumed without loss of generality since $\varepsilon(t)\to 0$ as $t\to+\infty$ (it might be necessary a redefinition of t_0).

Our stochastic dynamic (SDI − TA) can be written equivalently as follows

$$\begin{cases} dX(t) \in -\partial F_{\varepsilon(t)}(X(t))dt + \sigma(t,X(t))dW(t), \quad t\geq t_0;\\ X(t_0) = X_0, \end{cases} \tag{SDIT}$$

(i) Let us define the anchor function $\phi(x) = \frac{\|x-x^\star\|^2}{2}$. Since ∂g satisfy (H_λ), there exists a stochastic process $\tilde{\eta} : \Omega\times[t_0,+\infty[\to \mathbb{H}$ such that $\tilde{\eta}(t) \in \partial F_{\varepsilon(t)}(X(t))$ a.s. for almost all $t \geq t_0$. Using Itô's formula we obtain

$$\begin{aligned}\phi(X(t)) = &\underbrace{\frac{\|X_0-x^\star\|^2}{2}}_{\xi} + \underbrace{\frac{1}{2}\int_{t_0}^t \mathrm{tr}[\Sigma(s,X(s))ds]}_{A_t} - \underbrace{\int_{t_0}^t \langle\tilde{\eta}(s), X(s)-x^\star\rangle\, ds}_{U_t}\\ &+ \underbrace{\int_{t_0}^t \langle\sigma^\star(s,X(s))\,(X(s)-x^\star)\,, dW(s)\rangle}_{M_t}.\end{aligned} \tag{21}$$

Since $X \in S^2_{\mathbb{H}}[t_0]$ by Proposition 10, we have for every $T > t_0$, that

$$\mathbb{E}\left(\int_{t_0}^T \|\sigma^\star(s,X(s))\,(X(s)-x^\star)\|^2\, ds\right) \leq \mathbb{E}\left(\sup_{t\in[t_0,T]}\|X(t)-x^\star\|^2\right)\int_{t_0}^{+\infty}\sigma^2_\infty(s)ds < +\infty.$$

Therefore M_t is a square-integrable continuous martingale. It is also a continuous local martingale, which implies that $\mathbb{E}(M_t) = 0$.

Let us now take the expectation of (21). Using that

$$0 \leq \mathrm{tr}[\Sigma(s,X(s))] \leq \sigma^2_\infty(s),$$

and (39) that we recall below

$$\langle y(t), X(t)-x^\star\rangle \geq \varepsilon(t)\phi(X(t)) + \frac{\varepsilon(t)}{2}\left(\|x_{\varepsilon(t)}\|^2 - \|x^\star\|^2\right), \tag{22}$$

where $y : \Omega\times[t_0,+\infty[\to\mathbb{H}$ is such that $y(t) \in \partial F_{\varepsilon(t)}(X(t))$ a.s.. We obtain that

$$\begin{aligned}&\mathbb{E}\left(\phi(X(t))\right) + \int_{t_0}^t \varepsilon(s)\mathbb{E}\left(\phi(X(s))\right)ds\\ &\leq \mathbb{E}\left(\frac{\|X_0-x^\star\|^2}{2}\right) + \frac{1}{2}\int_{t_0}^t \sigma^2_\infty(s)ds + \frac{1}{2}\int_{t_0}^t \varepsilon(s)\left(\|x^\star\|^2 - \|x_{\varepsilon(s)}\|^2\right)ds.\end{aligned}$$

According to our assumptions, we can write briefly the above relation as

$$\mathbb{E}\left(\phi(X(t))\right) + \int_{t_0}^{t} \varepsilon(s)\mathbb{E}\left(\phi(X(s))\right) ds \leq \Upsilon(t), \tag{23}$$

with Υ a nonnegative function defined by

$$\Upsilon(t) \stackrel{\text{def}}{=} \mathbb{E}\left(\frac{\|X_0 - x^\star\|^2}{2}\right) + \frac{1}{2}\int_{t_0}^{t} \sigma_\infty^2(s)ds + \frac{1}{2}\int_{t_0}^{t} \varepsilon(s)\left(\|x^\star\|^2 - \|x_{\varepsilon(s)}\|^2\right) ds$$

which satisfies $\lim_{t\to+\infty}\Upsilon(t) = \Upsilon_\infty < +\infty$ by the fact that $X_0 \in \mathrm{L}^2(\Omega;\mathbb{H})$, $\sigma_\infty \in \mathrm{L}^2([t_0,+\infty[)$ and (T_3).

Let us integrate the above relation (23). We set

$$\theta(t) \stackrel{\text{def}}{=} \int_{t_0}^{t} \mathbb{E}\left(\phi(X(s))\right) ds.$$

We have $\dot{\theta}(t) = \mathbb{E}\left(\phi(X(t))\right)$ and (23) is written equivalently as

$$\dot{\theta}(t) + \int_{t_0}^{t} \varepsilon(s)\dot{\theta}(s)ds \leq \Upsilon(t). \tag{24}$$

Equivalently

$$\frac{1}{\varepsilon(t)}\frac{d}{dt}\int_{t_0}^{t} \varepsilon(s)\dot{\theta}(s)ds + \int_{t_0}^{t} \varepsilon(s)\dot{\theta}(s)ds \leq \Upsilon(t), \tag{25}$$

that is

$$\frac{d}{dt}\int_{t_0}^{t} \varepsilon(s)\dot{\theta}(s)ds + \varepsilon(t)\int_{t_0}^{t} \varepsilon(s)\dot{\theta}(s)ds \leq \varepsilon(t)\Upsilon(t). \tag{26}$$

With $m(t) \stackrel{\text{def}}{=} \exp\int_{t_0}^{t}\varepsilon(s)ds$, we get

$$\frac{d}{dt}\left(m(t)\int_{t_0}^{t} \varepsilon(s)\dot{\theta}(s)ds\right) \leq \varepsilon(t)m(t)\Upsilon(t). \tag{27}$$

After integration we get

$$\int_{t_0}^{t} \varepsilon(s)\dot{\theta}(s)ds \leq \frac{1}{m(t)}\int_{t_0}^{t} m'(s)\Upsilon(s)ds. \tag{28}$$

Since Υ is bounded by assumption (T_3), we get

$$\sup_{t\geq t_0}\mathbb{E}\Big[\int_{t_0}^{t} \varepsilon(s)\|X(s) - x^\star\|^2\Big]ds < +\infty.$$

Equivalently

$$\int_{t_0}^{+\infty} \mathbb{E}\Big[\|X(t) - x^\star\|^2\Big]\varepsilon(t)dt < +\infty.$$

The assumption (T_2) guarantees that the above inequality forces $\mathbb{E}\Big[\|X(t) - x^\star\|^2\Big]$ to tend to zero.

(ii) Consider (21), we define

$$\tilde{A}_t \stackrel{\text{def}}{=} A_t + \int_{t_0}^{t} \frac{\varepsilon(s)}{2}(\|x^\star\|^2 - \|x_{\varepsilon(s)}\|^2)ds, \quad and \quad \tilde{U}_t \stackrel{\text{def}}{=} U_t + \int_{t_0}^{t} \frac{\varepsilon(s)}{2}(\|x^\star\|^2 - \|x_{\varepsilon(s)}\|^2)ds.$$

By (22) we have that $\tilde{U}_t \geq \int_{t_0}^{t}\varepsilon(s)\phi(X(s))ds \geq 0$. We can rewrite (21) as

$$\phi(X(t)) = \xi + \tilde{A}_t - \tilde{U}_t + M_t.$$

Since $\sigma_\infty \in \mathrm{L}^2([t_0,+\infty[)$ and (T_3), then $\lim_{t\to+\infty} \tilde{A}_t < +\infty$. Let us observe that, since $X \in S^2_{\mathbb{H}}[t_0]$ by Proposition 10, we have for every $T > t_0$ that

$$\mathbb{E}\left(\int_{t_0}^{T} \|\sigma^\star(s,X(s))\,(X(s)-x^\star)\|^2\,ds\right) \leq \mathbb{E}\left(\sup_{t\in[t_0,T]} \|X(t)-x^\star\|^2\right)\int_{t_0}^{+\infty} \sigma^2_\infty(s)ds < +\infty.$$

Therefore, M_t is a square-integrable continuous martingale. It is also a continuous local martingale (see [29, Theorem 1.3.3]), which implies that $\mathbb{E}(M_t) = 0$.

By Theorem 39, we get that $\lim_{t\to+\infty} \|X(t)-x^\star\|$ exists a.s. and that $\lim_{t\to+\infty} \tilde{U}_t < +\infty$ a.s..

(iii) Using the lower bound we had on $\tilde{U}_t$, we obtain

$$\int_{t_0}^{+\infty} \varepsilon(t)\,\|X(t)-x^\star\|^2\,dt < +\infty.$$

(iv) By the previous item, (T_2), and Lemma 32 we conclude that $\lim_{t\to+\infty} X(t) = x^\star$ a.s..

This completes the proof. ◀

4.2 Practical situations

We will consider situations where the three conditions (T_1), (T_2) and (T_3) are satisfied simultaneously. These are properties of the viscosity curve that we will now study. The difficulty comes from (T_2) and (T_3) which are a priori not compatible. Indeed, (T_2) requires the parameter $\varepsilon(t)$ to converge slowly towards zero for the Tikhonov regularization to be effective. On the other hand in (T_3) the parameter $\varepsilon(t)$ must converge sufficiently quickly towards zero so that the term $\left(\|x^\star\|^2 - \|x_{\varepsilon(t)}\|^2\right)$ converges to zero fairly quickly, and thus corrects the infinite value of the integral of $\varepsilon(t)$.

4.2.1 Łojasiewicz property

Our first objective is to evaluate the rate of convergence towards zero of $\left(\|x^\star\|^2 - \|x_\varepsilon\|^2\right)$ as $\varepsilon \to 0$. Using the differentiability properties of the viscosity curve is not a good idea, because the viscosity curve can be of infinite length in the case of a general differentiable convex function, see [44]. To overcome this difficulty, we assume that $F = f + g$ satisfies the Łojasiewicz property. This basic property has its roots in algebraic geometry, and it essentially captures a domination inequality between the objective value and its (sub)gradient.

Definition 16 (Łojasiewicz inequality)**.** *Let $f : \mathbb{H} \to \mathbb{R}$ be a convex function with $\mathcal{S} \neq \emptyset$ and $q \in [0,1[$. f satisfies the Łojasiewicz inequality on $\mathcal{S}$ with exponent q if there exists $r > \min f$ and $\mu > 0$ such that:*

$$\mu(f(x) - \min f)^q \leq \left\|\partial^0 f(x)\right\|, \quad \forall x \in [\min f < f < r], \tag{29}$$

where we recall $\|\partial^0 f(x)\| = \min_{u\in\partial f(x)} \|u\|$. We will write $f \in \text{Ł}^q(\mathcal{S})$.

Error bounds have also been successfully applied to various branches of optimization, and in particular to complexity analysis. Of particular interest in our setting is the Hölderian error bound.

Definition 17 (Hölderian error bound)**.** *Let $f : \mathbb{H} \to \mathbb{R}$ be a proper function such that $\mathcal{S} \neq \emptyset$. Then f satisfies a Hölderian (or power-type) error bound inequality on $\mathcal{S}$ with exponent $p \geq 1$, if there exists $\gamma > 0$ and $r > \min f$ such that:*

$$f(x) - \min f \geq \gamma \mathrm{dist}(x,\mathcal{S})^p, \quad \forall x \in [\min f \leq f \leq r], \tag{30}$$

and we will write $f \in \mathrm{EB}^p(\mathcal{S})$.

A deep result due to Łojasiewicz states that for arbitrary continuous semi-algebraic functions, the Hölderian error bound inequality holds on any compact set, and the Łojasiewicz inequality holds at each point. In fact, for convex functions, the Łojasiewicz property and Hölderian error bound are actually equivalent.

Proposition 18. *Assume that $f \in \Gamma_0(\mathbb{H})$ with $\mathcal{S} \neq \emptyset$. Let $q \in [0,1[$, $p \overset{\text{def}}{=} \frac{1}{1-q} \geq 1$ and $r > \min f$. Then f verifies the Łojasiewicz inequality with exponent q (see (29)) at $[\min f < f < r]$ if and only if the Hölderian error bound with exponent p (see (30)) holds on $[\min f < f < r]$.*

Proof. This is a consequence of [11, Theorem 5]. ◀

4.2.2 Quantitative stability of variational systems

Our first objective is to evaluate the rate of convergence towards zero of $\left(\|x^\star\|^2 - \|x_\varepsilon\|^2\right)$ as $\varepsilon \to 0$. Using the differentiability properties of the viscosity curve is not a good idea, because the viscosity curve can be of infinite length in the case of a general differentiable convex function, see [44]. To overcome this difficulty, we assume that $F = f + g$ satisfies the Łojasiewicz property (see (29)). This basic property has its roots in algebraic geometry, and it essentially describes a relationship between the objective value and its gradient (or subgradient). Once this is assumed, we will need tools from variational analysis to conclude.

We start by recalling the notion of bounded Hausdorff distance for functions introduced in [6] to study stability of minimization problems. All the results of this section until Theorem 24 do not need separability of $\mathbb{H}$.

For a set $C \subset \mathbb{H} \times \mathbb{R}$ and $\rho \geq 0$, we denote $C_\rho \stackrel{\text{def}}{=} C \cap \rho\mathbb{B}$, where $\mathbb{B}$ is the unit ball in the box norm on $\mathbb{H} \times \mathbb{R}$. For two sets $C, D \subset \mathbb{H} \times \mathbb{R}$, the excess function of C on D is defined as

$$e(C, D) \stackrel{\text{def}}{=} \sup_{x \in C} \operatorname{dist}(x, D).$$

For any $\rho \geq 0$, the ρ-Hausdorff distance between C and D is defined as

$$\operatorname{haus}_\rho(C, D) \stackrel{\text{def}}{=} \max(e(C_\rho, D), e(D_\rho, C)).$$

For $\rho = +\infty$, we recover the Hausdorff distance. A metrizable topology is naturally attached to the ρ-Hausdorff distance. When $\mathbb{H}$ is finite dimensional, the convergence with respect to the ρ-Hausdorff distances is nothing but the classical Painlevé-Kuratowski set-convergence.

Definition 19. *For $\rho \geq 0$, the ρ-Hausdorff (epi-)distance between two functions $f, g : \mathbb{H} \to \mathbb{R} \cup \{+\infty\}$ is*

$$\operatorname{haus}_\rho(f, g) \stackrel{\text{def}}{=} \operatorname{haus}_\rho(\operatorname{epi} f, \operatorname{epi} g).$$

This device was extended in [5] to set-valued operators by identifying them with their graphs.

Definition 20. *For $\rho \geq 0$, the ρ-Hausdorff distance between two operators $A, B : \mathbb{H} \rightrightarrows \mathbb{H}$ is*

$$\operatorname{haus}_\rho(A, B) \stackrel{\text{def}}{=} \operatorname{haus}_\rho(\operatorname{gph} A, \operatorname{gph} B),$$

where the unit ball is that of $\mathbb{H} \times \mathbb{H}$ equipped with the box norm.

We recall the following two results that have been obtained in [5, 6] and that will be important to prove our quantitative stability result. It is a particular case of [5, Proposition 1.2] with $\rho = 0, \lambda = 1$.

Proposition 21. *Let $A, B : \mathbb{H} \rightrightarrows \mathbb{H}$ be two maximal monotone operators, then*

$$\|J_A(0) - J_B(0)\| \leq 3\operatorname{haus}_{\|J_A(0)\|}(A, B),$$

where $J_A \stackrel{\text{def}}{=} (I + A)^{-1}, J_B \stackrel{\text{def}}{=} (I + B)^{-1}$ are the resolvent of the operators A and B, respectively.

The second abstract result is the equivalence of the uniform structure on the class of subdifferentials of convex lsc functions between the bounded Hausdorff distance and the uniform convergence on bounded sets of resolvents.

Proposition 22 ([6, Theorem 5.2])**.** *. Let f and $g \in \Gamma_0(\mathbb{H})$. To any $\rho > \max\left(\operatorname{dist}(0, \operatorname{epi}(f)), \operatorname{dist}(0, \operatorname{epi}(g))\right)$ there correspond some constants κ and ρ_0 (that depend on ρ) such that*

$$\operatorname{haus}_\rho(\partial f, \partial g) \leq \kappa \left(\operatorname{haus}_{\rho_0}(f, g)\right)^{\frac{1}{2}}.$$

The following proposition is new and is a consequence of the previous two results. Since this is not obvious, we are going to present the whole proof.

Proposition 23. *Let $f \in \Gamma_0(\mathbb{H})$ be a function such that $\mathcal{S} \stackrel{\text{def}}{=} \operatorname{argmin}_{\mathbb{H}}(f) \neq \emptyset$, and that $f \in \mathrm{EB}^p(\mathcal{S})$. Let also $x^\star = \mathrm{P}_{\mathcal{S}}(0)$ and for $\varepsilon > 0$, let x_ε be the unique minimizer of $f_\varepsilon(x) = f(x) + \frac{\varepsilon}{2}\|x\|^2$. Then there exists $C_0, \varepsilon^\star > 0$ such that*

$$\|x_\varepsilon - x^\star\| \leq C_0 \varepsilon^{\frac{1}{2p}}, \quad \forall \varepsilon \in]0, \varepsilon^\star]. \tag{31}$$

Consequently, there exists $C > 0$ such that

$$\|x^\star\|^2 - \|x_\varepsilon\|^2 \leq C \varepsilon^{\frac{1}{2p}}, \quad \forall \varepsilon \in]0, \varepsilon^\star]. \tag{32}$$

Proof. Let $\varphi_\varepsilon \stackrel{\text{def}}{=} \frac{1}{\varepsilon}(f - \min f)$. By optimality of x_ε, we have

$$x_\varepsilon = (I + \partial\varphi_\varepsilon)^{-1}(0) = J_{\partial\varphi_\varepsilon}(0).$$

We have that φ_ε increases to $\iota_{\mathcal{S}}$ as ε decreases to zero, and

$$x^\star = \mathrm{P}_{\mathcal{S}}(0) = (I + \partial\iota_{\mathcal{S}})^{-1}(0) = J_{\partial\iota_{\mathcal{S}}}(0),$$

where $\iota_{\mathcal{S}} : \mathbb{H} \to \{0, +\infty\}$ is the indicator function of $\mathcal{S}$, that takes 0 on $\mathcal{S}$ and $+\infty$ otherwise. Therefore

$$\|x_\varepsilon - x^\star\| = \|(I + \partial\varphi_\varepsilon)^{-1}(0) - (I + \partial\iota_{\mathcal{S}})^{-1}(0)\| = \|J_{\partial\varphi_\varepsilon}(0) - J_{\partial\iota_{\mathcal{S}}}(0)\|.$$

Applying Proposition 21 with $A = \partial\varphi_\varepsilon$, and $B = \partial\iota_{\mathcal{S}}$, we have that

$$\|x_\varepsilon - x^\star\| \le 3\mathrm{haus}_\rho(\partial\varphi_\varepsilon, \partial\iota_{\mathcal{S}}),$$

for $\rho > \|x^\star\|$. Now, since $\max(\mathrm{dist}(0, \mathrm{epi}(\varphi_\varepsilon)), \mathrm{dist}(0, \mathrm{epi}(\iota_{\mathcal{S}}))) \le \|x^\star\|$, we fix $\rho \ge \|x^\star\|$, and Proposition 22 entails that there exists constants $\kappa, \rho_0 > 0$ (depending on ρ) such that

$$\|x_\varepsilon - x^\star\| \le 3\kappa\left(\mathrm{haus}_{\rho_0}(\varphi_\varepsilon, \iota_{\mathcal{S}})\right)^{\frac{1}{2}}. \tag{33}$$

To complete our proof we just need to bound the right hand side of the last inequality. Observe first that since $\iota_{\mathcal{S}} \ge \varphi_\varepsilon$ we just need to compute $e\left((\mathrm{epi}\ \varphi_\varepsilon)_{\rho_0}, \mathrm{epi}\ \iota_{\mathcal{S}}\right) = e\left((\mathrm{epi}\ \varphi_\varepsilon)_{\rho_0}, \mathcal{S} \times \mathbb{R}_+\right)$. It then follows from Definition 19 that

$$\mathrm{haus}_{\rho_0}(\varphi_\varepsilon, \iota_{\mathcal{S}}) = \max_{(x_1,r_1)\in\mathrm{epi}(\varphi_\varepsilon)\cap\rho_0\mathbb{B}}\ \min_{(x_2,r_2)\in\mathcal{S}\times\mathbb{R}_+} \max\left(\|x_1 - x_2\|, |r_1 - r_2|\right),$$

where $\mathbb{B}$ is the unit ball of the max norm on $\mathbb{H} \times \mathbb{R}_+$. Besides, the inner minimization problem is bounded above by taking $r_2 = r_1$. Hence,

$$\begin{aligned}\mathrm{haus}_{\rho_0}(\varphi_\varepsilon, \iota_{\mathcal{S}}) &\le \max_{(x_1,r_1)\in\mathrm{epi}(\varphi_\varepsilon)\cap\rho_0\mathbb{B}} \mathrm{dist}(x_1, \mathcal{S}) = \max_{x_1\in[\min f\le f\le\varepsilon r_1+\min f],\|x_1\|\le\rho_0,r_1\le\rho_0} \mathrm{dist}(x_1, \mathcal{S}) \\ &\le \max_{x\in[\min f\le f\le\varepsilon\rho_0+\min f],\|x\|\le\rho_0} \mathrm{dist}(x, \mathcal{S}).\end{aligned}$$

We will now invoke the assumption that $f \in \mathrm{EB}^p(\mathcal{S})$. By the latter, there exists $\gamma > 0, r > \min f$ such that (30) holds. Now choose $\varepsilon_0 \stackrel{\text{def}}{=} \frac{r - \min f}{\rho_0} > 0$. We then have for any $\varepsilon \in]0, \varepsilon_0]$ that

$$\mathrm{haus}_{\rho_0}(\varphi_\varepsilon, \iota_{\mathcal{S}}) \le \max_{x\in[0\le f-\min f\le\varepsilon\rho_0]} \mathrm{dist}(x, \mathcal{S}) \le \max_{x\in[0\le f-\min f\le\varepsilon\rho_0]} \left(\frac{f(x) - \min f}{\gamma}\right)^{\frac{1}{p}} \le \left(\frac{\rho_0}{\gamma}\right)^{\frac{1}{p}} \varepsilon^{\frac{1}{p}}, \tag{34}$$

where we have used that $\rho_0\varepsilon \le \rho_0\varepsilon_0 \le r - \min f$ so that (30) applies. Combining (33) and (34) gives (31) where $C_0 = 3\kappa\left(\frac{\rho_0}{\gamma}\right)^{\frac{1}{2p}}$.

On the other hand, from Theorem 1(i) and the triangle inequality, we have

$$\|x^\star\|^2 - \|x_\varepsilon\|^2 \le 2\|x^\star\|\|x_\varepsilon - x^\star\|, \quad \forall\varepsilon \ge 0.$$

Taking $\varepsilon^\star = \varepsilon_0$ and $C = 2C_0\|x^\star\|$ and using (31), we get (32). ◀

The previous proposition was the key to deriving a proper tuning of the parameter $\varepsilon(t)$ that satisfies all the conditions presented in Theorem 14. In the following, we make this precise, recalling that the setting of Theorem 14 happened in the separable real Hilbert space $\mathbb{H}$, and that the previous proposition remains valid without separability.

Theorem 24. *Consider the setting of Theorem 14 and suppose that $F = f + g \in \mathrm{EB}^p(\mathcal{S}_F)$. Then taking the Tikhonov parameter $\varepsilon(t) = \frac{1}{t^r}$ with*

$$1 \ge r > \frac{2p}{2p+1},$$

then the three conditions (T_1), (T_2), and (T_3) of Theorem 14 are satisfied simultaneously. In particular, the solution $X \in S_{\mathbb{H}}^\nu[t_0]$ of (SDI − TA) *is unique and we get almost sure (strong) convergence of $X(t)$ to the minimal norm solution $x^\star = \mathrm{P}_{\mathcal{S}_F}(0)$.*

Proof. It is direct to check (T_1) and (T_2). In order to check (T_3), let $\varepsilon^\star > 0$ from Proposition 23 and $T^\star = \max\left(t_0, (\frac{1}{\varepsilon^\star})^{\frac{1}{r}}\right)$, then we have

$$\|x^\star\|^2 - \|x_{\varepsilon(t)}\|^2 \leq C\frac{1}{t^{\frac{r}{2p}}}, \quad \forall t \geq T^\star.$$

Therefore,

$$\int_{t_0}^{+\infty} \frac{\|x^\star\|^2 - \|x_{\varepsilon(t)}\|^2}{t^r} dt = \underbrace{\int_{t_0}^{T^\star} \frac{\|x^\star\|^2 - \|x_{\varepsilon(t)}\|^2}{t^r} dt}_{I_1} + \underbrace{\int_{T^\star}^{+\infty} \frac{\|x^\star\|^2 - \|x_{\varepsilon(t)}\|^2}{t^r} dt}_{I_2}.$$

Is clear that I_1 is bounded (by $T^\star t_0^{-r}\|x^\star\|^2$ for instance). Hence (T_3) holds under the condition that

$$\int_{T^\star}^{+\infty} \frac{1}{t^r}\frac{C}{t^{\frac{r}{2p}}} dt < +\infty,$$

which is true when $r + \frac{r}{2p} > 1$, whence we deduce our condition $1 \geq r > \frac{2p}{2p+1}$. ◀

4.3 Convergence rates of the objective in the smooth case

In this subsection, we turn our attention to the smooth case (*i.e.* $g \equiv 0$) and derive explicit global convergence rates in expectation. Recall that in the previous subsections, we established almost sure strong convergence with Tikhonov regularization, but did not obtain any convergence rate. To fill this gap, at least in the smooth setting, we first revisit the result of [2] which provides rates in the deterministic case. We then adapt their proof strategy–together with the stochastic analysis tools developed earlier– to control the additional variance term and obtain the decay of the objective value and distance to the minimal norm solution in expectation.

Theorem 25. [2, Theorem 5] *Take* $\varepsilon(t) = \dfrac{1}{t^r}$ *and* $0 < r < 1$. *Let us consider* (DI − TA) *in the case where* $g \equiv 0$, *i.e.,*

$$\dot{x}(t) + \nabla f(x(t)) + \frac{1}{t^r}x(t) = 0. \tag{35}$$

Let $x : [t_0, +\infty[\to \mathbb{H}$ *be a solution trajectory of* (DI − TA). *For* $\varepsilon > 0$ *define* $f_\varepsilon(x) \overset{\text{def}}{=} f(x) + \frac{\varepsilon}{2}\|x\|^2$, *let* x_ε *be the unique minimizer of* f_ε, *and consider the Lyapunov function*

$$E(t) \overset{\text{def}}{=} f_{\varepsilon(t)}(x(t)) - f_{\varepsilon(t)}(x_{\varepsilon(t)}) + \frac{\varepsilon(t)}{2}\|x(t) - x_{\varepsilon(t)}\|^2.$$

Then, we have :

(i) $E(t) = \mathcal{O}\left(\dfrac{1}{t}\right)$ *as* $t \to +\infty$;

(ii) $f(x(t)) - \min(f) = \mathcal{O}\left(\dfrac{1}{t^r}\right)$ *as* $t \to +\infty$;

(iii) $\|x(t) - x_{\varepsilon(t)}\|^2 = \mathcal{O}\left(\dfrac{1}{t^{1-r}}\right)$ *as* $t \to +\infty$.

In light of Proposition 23, we can now characterize the rate at which the trajectory solution of (DI − TA) converges to the minimum norm solution. To the best of our knowledge, this convergence rate estimate is new.

Corollary 26. *Consider the setting of Theorem 25, then we have strong convergence of* $x(t)$ *to the minimum norm solution* $x^\star = \mathrm{P}_{\mathcal{S}}(0)$. *Moreover, if* $f \in \mathrm{EB}^p(\mathcal{S})$, *then*

$$\|x(t) - x^\star\|^2 = \begin{cases} \mathcal{O}\left(\dfrac{1}{t^{\frac{r}{p}}}\right), & \text{if } r \in \left]0, \frac{p}{p+1}\right[; \\ \mathcal{O}\left(\dfrac{1}{t^{1-r}}\right), & \text{if } r \in \left[\frac{p}{p+1}, 1\right[\end{cases} \quad \text{as } t \to +\infty. \tag{36}$$

Proof. Combine the third item of Theorem 25 and Proposition 23. ◀

▶ Remark 27. We observe that the convergence rate is governed by a piecewise function, attaining its optimum when $r = \frac{p}{p+1}$, in which case we obtain

$$\|x(t) - x^\star\|^2 = \mathcal{O}\left(\frac{1}{t^{\frac{1}{p+1}}}\right), \text{ as } t \to +\infty.$$

We also remark that this is strictly slower than the convergence rate of (deterministic) gradient flow when f satisfies a Hölderian error bound, in which case we have

$$\mathrm{dist}^2(x(t), \mathcal{S}) = \mathcal{O}\left(\frac{1}{t^{\frac{2}{p}}}\right), \text{ as } t \to +\infty.$$

This reflects the trade-off for ensuring strong convergence to the minimal norm solution with the Tikhonov regularization term $\varepsilon(t) = \frac{1}{t^{\frac{p}{p+1}}}, p \geq 1$.

We are ready now to state the main theorem of this subsection, which establishes global convergence rates in expectation for the trajectories of (SDI − TA) when $g \equiv 0$. Moreover, this result recovers the deterministic convergence rates of Theorem 25 and Corollary 26. Indeed, by setting $\sigma_\infty^2 = 0$, the stochastic term vanishes and we retrieve exactly the bounds of the deterministic case.

Theorem 28. *Consider* (SDI − TA) *with* $g \equiv 0$, $\varepsilon(t) = \frac{1}{t^r}$ *where* $0 < r < 1$, *i.e.*

$$\begin{cases} dX(t) = -\nabla f(X(t))dt - \dfrac{1}{t^r}X(t)dt + \sigma(t, X(t))dW(t), \quad t \geq t_0; \\ X(t_0) = X_0. \end{cases} \tag{SDE − TA}$$

where the initial data $X_0 \in \mathrm{L}^\nu(\Omega; \mathbb{H})$ *for* $\nu \geq 2$. *Assume that* $f \in \Gamma_0(\mathbb{H}) \cap C_L^2(\mathbb{H})$ *with* $\mathcal{S} \overset{\text{def}}{=} \mathrm{argmin}(f) \neq \emptyset$, *and* $f \in \mathrm{EB}^p(\mathcal{S})$. *Suppose also that* σ *satisfies* (H), *and that* $\sigma_\infty \in \mathrm{L}^2([t_0, +\infty[)$ *and is non-increasing. For* $\varepsilon > 0$, *let* $f_\varepsilon(x) \overset{\text{def}}{=} f(x) + \frac{\varepsilon}{2}\|x\|^2$, x_ε *be the unique minimizer of* f_ε. *Let* $x^\star \overset{\text{def}}{=} \mathrm{P}_{\mathcal{S}}(0)$. *Consider the energy function*

$$E(t, x) \overset{\text{def}}{=} f_{\varepsilon(t)}(x) - f_{\varepsilon(t)}(x_{\varepsilon(t)}) + \frac{\varepsilon(t)}{2}\|x - x_{\varepsilon(t)}\|^2,$$

and for $t_1 > t_0$,

$$R(t) \overset{\text{def}}{=} e^{-\frac{t^{1-r}}{1-r}} \int_{t_1}^t e^{\frac{s^{1-r}}{1-r}} \sigma_\infty^2(s) ds. \tag{37}$$

Then, the solution trajectory $X \in S_{\mathbb{H}}^\nu[t_0]$ *is unique, and the following holds:*

(i) R *converges to* 0 *at the rate,*

$$R(t) = \mathcal{O}\left(\exp(-t^r(1 - 2^{-r})) + t^r \sigma_\infty^2\left(\frac{t_1 + t}{2}\right)\right).$$

If, moreover, $\sigma_\infty^2(t) = \mathcal{O}(t^{-\alpha})$ *for* $\alpha > 1$, *then* $R(t) = \mathcal{O}(t^{r-\alpha})$.

(ii) $\mathbb{E}[E(t, X(t))] = \mathcal{O}\left(\frac{1}{t} + R(t)\right)$

(iii) $\mathbb{E}[f(X(t)) - \min(f)] = \mathcal{O}\left(\frac{1}{t^r} + R(t)\right)$. *In addition, if* $\sigma_\infty^2(t) = \mathcal{O}(t^{-\alpha})$ *for* $\alpha > 1$, *then*

$$\mathbb{E}[f(X(t)) - \min(f)] = \begin{cases} \mathcal{O}\left(\dfrac{1}{t^{\alpha-r}}\right), & \text{if } \alpha \in]1, 2r[; \\ \mathcal{O}\left(\dfrac{1}{t^r}\right), & \text{if } \alpha \geq 2r; \end{cases}$$

(iv) $\mathbb{E}[\|X(t) - x_{\varepsilon(t)}\|^2] = \mathcal{O}\left(\frac{1}{t^{1-r}} + t^r R(t)\right)$, *which goes to* 0 *as* $t \to +\infty$ *if* $r \in]0, \frac{1}{2}]$. *If, moreover,* $\sigma_\infty^2(t) = \mathcal{O}(t^{-\alpha})$ *for* $\alpha > \max\{2r, 1\}$, *then*

$$\mathbb{E}[\|X(t) - x_{\varepsilon(t)}\|^2] = \begin{cases} \mathcal{O}\left(\dfrac{1}{t^{\alpha-2r}}\right), & \text{if } \alpha \in]\max\{2r, 1\}, r + 1[; \\ \mathcal{O}\left(\dfrac{1}{t^{1-r}}\right), & \text{if } \alpha \geq r + 1. \end{cases}.$$

(v) $\mathbb{E}[\|X(t) - x^\star\|^2] = \mathcal{O}\left(\frac{1}{t^{1-r}} + \frac{1}{t^{\frac{r}{p}}} + t^r R(t)\right)$, *which goes to* 0 *as* $t \to +\infty$ *if* $r \in]0, \frac{1}{2}]$. *In addition, if* $\sigma_\infty^2(t) = \mathcal{O}(t^{-\alpha})$ *for* $\alpha > \max\{2r, 1\}$, *then*

$$\mathbb{E}[\|X(t) - x^\star\|^2] = \mathcal{O}\left(\frac{1}{t^{1-r}} + \frac{1}{t^{\frac{r}{p}}} + \frac{1}{t^{\alpha-2r}}\right).$$

In particular,

$$\mathbb{E}[\|X(t) - x^\star\|^2] = \begin{cases} \mathcal{O}\left(\frac{1}{t^{1-r}}\right), & \text{if } r \in \left]\frac{p}{p+1}, 1\right[, \alpha > r+1; \\ \mathcal{O}\left(\frac{1}{t^{\frac{r}{p}}}\right), & \text{if } r \in \left]0, \frac{p}{p+1}\right[, \alpha > \max\{1, \frac{r(2p+1)}{p}\}; \\ \mathcal{O}\left(\frac{1}{t^{\alpha-2r}}\right), & \text{if } r \in \left]\frac{p}{2p+1}, 1\right[, \alpha \in \left(\max\{2r, 1\}, \min\{r+1, \frac{r(2p+1)}{p}\}\right). \end{cases}$$

Proof. The existence and uniqueness of a solution was already stated in Theorem 14.

The first item is a direct consequence of Lemma 35, for the second one we recall that $\sigma_\infty \in \mathrm{L}^2([t_0, +\infty[)$ and is non-increasing, and we proceed as follows:

$$\begin{aligned} R(t) &= e^{-\frac{t^{1-r}}{1-r}} \int_{t_1}^{\frac{t_1+t}{2}} e^{\frac{s^{1-r}}{1-r}} \sigma_\infty^2(s) ds + e^{-\frac{t^{1-r}}{1-r}} \int_{\frac{t_1+t}{2}}^{t} e^{\frac{s^{1-r}}{1-r}} \sigma_\infty^2(s) ds \\ &\le e^{\left(\frac{t_0}{2}\right)^r} e^{-t^r(1-2^{-r})} \int_{t_1}^{+\infty} \sigma_\infty^2(s) ds + \sigma_\infty^2\left(\frac{t_1+t}{2}\right) D_{\frac{1}{1-r}, 1-r}(t), \end{aligned}$$

where

$$D_{a,b}(t) = e^{-at^b} \int_0^t e^{as^b} ds.$$

As a corollary of an upper bound of the Dawson integral shown in [31, Section 7.8], we have that

$$D_{a,b}(t) \le \frac{2}{ab} t^{1-b}, \quad 0 < b \le 2, a > 0, t > 0,$$

thus we obtain

$$R(t) = \mathcal{O}\left(\exp(-t^r(1-2^{-r})) + t^r \sigma_\infty^2\left(\frac{t_1+t}{2}\right)\right).$$

Since σ_∞^2 is non-increasing and $r < 1$, we have

$$0 \le t\sigma_\infty^2(t) \le 2\int_{\frac{t}{2}}^{t} \sigma_\infty^2(u) du,$$

and the right hand side goes to 0 as $t \to +\infty$ since $\sigma_\infty \in \mathrm{L}^2([t_0, +\infty[)$ by assumption. Thus we obtain that $\lim_{t\to\infty} t\sigma_\infty^2(t) = 0$ which proves claim (i).

The remainder of the proof follows by applying Itô's formula to (SDE − TA) with the function

$$\phi(t, x) \stackrel{\text{def}}{=} \Phi(t) E(t, x) \quad \text{where} \quad \Phi(t) \stackrel{\text{def}}{=} \exp\Big(\int_{t_1}^{t} s^{-r} ds\Big),$$

and then taking expectation. Following similar computations as in [2, Theorem 3], we obtain

$$\begin{aligned} \mathbb{E}[\phi(t, X(t))] &\le \mathbb{E}[\phi(t_0, X_0)] - \|x^\star\|^2 \int_{t_0}^{t} \dot\varepsilon(s)\Phi(s) ds + \int_{t_0}^{t} \mathbb{E}\left[\mathrm{tr}[\sigma(s, X(s))\sigma^\star(s, X(s))\nabla^2\phi(s, X(s))]\right] ds \\ &\le \mathbb{E}[\phi(t_0, X_0)] - \|x^\star\|^2 \int_{t_0}^{t} \dot\varepsilon(s)\Phi(s) ds + \left(L + 2t_0^{-r}\right) \int_{t_0}^{t} \Phi(s)\sigma_\infty^2(s) ds. \end{aligned}$$

Dividing by $\Phi(t)$, we have equivalently that

$$\begin{aligned} \mathbb{E}[E(t, X(t))] &\le \frac{\mathbb{E}[\phi(t_0, X_0)]}{\Phi(t)} - \frac{\|x^\star\|^2}{\Phi(t)} \int_{t_0}^{t} \dot\varepsilon(s)\Phi(s) ds + \frac{(L + 2t_0^{-r})}{\Phi(t)} \int_{t_0}^{t} \Phi(s)\sigma_\infty^2(s) ds \\ &\le \frac{\mathbb{E}[\phi(t_0, X_0)]}{\Phi(t)} + \frac{\|x^\star\|^2}{\rho t} + \frac{(L + 2t_0^{-r})}{\Phi(t)} \int_{t_0}^{t} \Phi(s)\sigma_\infty^2(s) ds, \end{aligned}$$

for $\rho < \frac{1}{r}$ (see the proof of [2, Theorem 5]). By definition $\frac{1}{\Phi(t)}\int_{t_0}^t \Phi(s)\sigma_\infty^2(s)ds = R(t)$, and since $\frac{1}{\Phi(t)}$ decays exponentially, we conclude with the claim of item (ii). Besides, by [2, Lemma 3], we have

$$f(x) - \min f \le E(t,x) + \frac{\varepsilon(t)}{2}\|x^\star\|^2 \quad \text{and} \quad \|x - x_{\varepsilon(t)}\|^2 \le \frac{E(t,x)}{\varepsilon(t)}.$$

Taking expectation and inserting the bound of (ii), we obtain claims (iii) and (iv). Finally, for item (v), we combine (iv) and Proposition 23. To conclude with the particular convergence rates, we plug in the derived rates for $R(t)$ obtained in (i) and the rate of $\sigma_\infty^2(t)$. ◀

▶ Remark 29. In the finite-dimensional case, *i.e.*, $\mathbb{H} = \mathbb{R}^d$ (not necessarily $\mathbb{K}$), we can weaken the assumption $f \in C_L^2(\mathbb{H})$ to $f \in C_L^{1,1}(\mathbb{H})$ thanks to [30, Proposition 2.2].

▶ Remark 30. Comparing Theorem 28 to its deterministic counterpart Theorem 25 (see also Corollary 26), one has the additional term $R(t)$ that appears in each rate. This necessarily slows down the convergence rate compared to the deterministic setting. But as expected, it is the price to be paid to account for stochastic noise while ensuring convergence.

▶ Remark 31. Our result in Theorem 14 ensures that with the Tikhonov regularization, the solution trajectory strongly converges in almost sure sense to the minimal norm solution, provided that the regularization coefficient $\varepsilon(t)$ is well chosen (verifies (T_1), (T_2) and (T_3)), and the diffusion term decays fast enough. While it is easy to choose $\varepsilon(t)$ so that (T_1)-(T_2) hold, fulfilling (T_3) required more involved arguments, and for instance that f verifies a Hölderian error bound. This also allowed to derive the (pointwise) convergence rates in expectation of Theorem 28. These quantitative estimates reveal that there is a trade-off between the decay of $\varepsilon(t)$ and that of the diffusion term $\sigma_\infty(t)$ in order to maintain convergence and have meaningful convergence rates. This is clearly reflected in the form of the function $R(t)$ in Theorem 28(i). For instance, mere square-integrability of $\sigma_\infty(t)$ is not sufficient as $\sigma_\infty(t)$ must decrease at least as $t^{-\alpha}$, $\alpha > 1$.

5 Conclusion, Perspectives

The purpose of this work was to study the convergence properties of trajectories of subgradient-like flows under stochastic errors in infinite dimensional separable real Hilbert spaces. The motivation stems from solving non-smooth convex optimization problems with noisy subgradient oracles with vanishing variance. We have shown important properties of these trajectories, such as the almost sure weak (resp. strong) convergence to a minimizer (resp. minimal norm one) without (resp. with) Tikhonov regularization. We have also investigated the convergence rates and highlighted the trade-off between the tuning of the Tikhonov regularization coefficient and the noise variance. This work leads us to important extensions, among which,we mention the following ones:

- Extend our results, with and without Tikhonov regularization, to the case of to the case of operators where ∇f and ∂g are replaced by, respectively, a co-coercive operator B and a maximal monotone operator A.
- Investigate the transition to second-order dynamics via time-scaling and averaging, and analyzing its corresponding convergence properties.
- Study second-order dynamics with inertia in view of understanding the behavior of accelerated dynamics in the presence of stochastic errors.

Some of these aspects are already the subject of ongoing research work.

A Auxiliary results

A.1 Deterministic results

Lemma 32. *Let $t_0 \ge 0$ 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 33 (Comparison Lemma)**.** *Let $t_0 \ge 0$ and $T > t_0$. Assume that $h : [t_0, +\infty[\to \mathbb{R}_+$ is measurable with $h \in \mathrm{L}^1([t_0, T])$, that $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and non-decreasing, $\varphi_0 > 0$ and the Cauchy problem*

$$\begin{cases} \varphi'(t) = -\psi(\varphi(t)) + h(t) & \text{for almost all } t \in [t_0, T] \\ \varphi(t_0) = \varphi_0 \end{cases}$$

has an absolutely continuous solution $\varphi : [t_0, T] \to \mathbb{R}_+$*. If a bounded from below lower semicontinuous function* $\omega : [t_0, T] \to \mathbb{R}_+$ *satisfies*

$$\omega(t) \leq \omega(s) - \int_s^t \psi(\omega(\tau))d\tau + \int_s^t h(\tau)d\tau$$

for $t_0 \leq s < t \leq T$ *and* $\omega(t_0) = \varphi_0$*, then*

$$\omega(t) \leq \varphi(t) \quad \text{for } t \in [t_0, T].$$

Lemma 34. *Let* $f : \mathbb{R}_+ \to \mathbb{R}$ *and* $\liminf_{t\to\infty} f(t) \neq \limsup_{t\to\infty} f(t)$*. Then there exists a constant* α*, satisfying* $\liminf_{t\to\infty} f(t) < \alpha < \limsup_{t\to\infty} f(t)$*, such that for every* $\beta > 0$*, we can define a sequence* $(t_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ *such that*

$$f(t_k) > \alpha, \quad t_{k+1} > t_k + \beta \quad \forall k \in \mathbb{N}.$$

Proof. See proof in [30, Lemma A.3]. ◀

Lemma 35. *Take* $t_0 > 0$*, and let* $f \in \mathrm{L}^1([t_0, +\infty[)$ *be continuous. Consider a non-decreasing function* $\varphi : [t_0, +\infty[\to \mathbb{R}_+$ *such that* $\lim_{t\to+\infty} \varphi(t) = +\infty$*. Then* $\lim_{t\to+\infty} \frac{1}{\varphi(t)} \int_{t_0}^t \varphi(s)f(s)ds = 0$*.*

Proof. See proof in [7, Lemma A.5] ◀

A.2 Stochastic results

A.2.1 On stochastic processes

Let us recall some elements of stochastic analysis. Throughout the paper, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\{\mathcal{F}_t | t \geq 0\}$ is a filtration of the $\sigma-$algebra $\mathcal{F}$. Given $\mathcal{C} \in \mathcal{P}(\Omega)$, we will denote $\sigma(\mathcal{C})$ the $\sigma-$algebra generated by $\mathcal{C}$. We denote $\mathcal{F}_\infty \stackrel{\text{def}}{=} \sigma\left(\bigcup_{t\geq 0} \mathcal{F}_t\right) \in \mathcal{F}$.

The expectation of a random variable $\xi : \Omega \to \mathbb{H}$ is denoted by

$$\mathbb{E}(\xi) \stackrel{\text{def}}{=} \int_\Omega \xi(\omega)d\mathbb{P}(\omega).$$

An event $E \in \mathcal{F}$ happens almost surely if $\mathbb{P}(E) = 1$, and it will be denoted as "E, $\mathbb{P}$-a.s." or simply "E, a.s.". The characteristic function of an event $E \in \mathcal{F}$ is denoted by

$$\mathbb{1}_E(\omega) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \omega \in E, \\ 0 & \text{otherwise.} \end{cases}$$

An $\mathbb{H}$-valued stochastic process starting at $t_0 \geq 0$ is a function $X : \Omega \times [t_0, +\infty[\to \mathbb{H}$. It is said to be continuous if $X(\omega, \cdot) \in C([t_0, +\infty[; \mathbb{H})$ for almost all $\omega \in \Omega$. We will denote $X(t) \stackrel{\text{def}}{=} X(\cdot, t)$. We are going to study SDEs, and in order to ensure the uniqueness of a solution, we introduce an equivalence relationship over stochastic processes. Two stochastic processes $X, Y : \Omega \times [t_0, T] \to \mathbb{H}$ are said to be equivalent if $X(t) = Y(t)$, $\forall t \in [t_0, T]$, $\mathbb{P}$-a.s. This leads us to define the equivalence relation $\mathcal{R}$, which associates the equivalent stochastic processes in the same class.

Furthermore, we will need some properties about the measurability of these processes. A stochastic process $X : \Omega \times [t_0, +\infty[\to \mathbb{H}$ is progressively measurable if for every $t \geq t_0$, the map $\Omega \times [t_0, t] \to \mathbb{H}$ defined by $(\omega, s) \to X(\omega, s)$ is $\mathcal{F}_t \otimes \mathcal{B}([t_0, t])$-measurable, where $\otimes$ is the product σ-algebra and $\mathcal{B}$ is the Borel σ-algebra. On the other hand, X is $\mathcal{F}_t$-adapted if $X(t)$ is $\mathcal{F}_t$-measurable for every $t \geq t_0$. It is a direct consequence of the definition that if X is progressively measurable, then X is $\mathcal{F}_t$-adapted.

Let us define the quotient space:

$$S^0_{\mathbb{H}}[t_0, T] \stackrel{\text{def}}{=} \{X : \Omega \times [t_0, T] \to \mathbb{H},\ X \text{ is a prog. measurable cont. stochastic process}\} \Big/ \mathcal{R}.$$

Set $S^0_{\mathbb{H}}[t_0] \stackrel{\text{def}}{=} \bigcap_{T\geq t_0} S^0_{\mathbb{H}}[t_0, T]$. For $\nu > 0$, we define $S^\nu_{\mathbb{H}}[t_0, T]$ as the subset of processes $X(t)$ in $S^0_{\mathbb{H}}[t_0, T]$ such that

$$S^\nu_{\mathbb{H}}[t_0, T] \stackrel{\text{def}}{=} \left\{ X \in S^0_{\mathbb{H}}[t_0, T] :\ \mathbb{E}\left(\sup_{t\in[t_0,T]} \|X_t\|^\nu \right) < +\infty \right\}.$$

We define $S^{\nu}_{\mathbb{H}}[t_0] \stackrel{\text{def}}{=} \bigcap_{T\geq t_0} S^{\nu}_{\mathbb{H}}[t_0,T]$.

Following the notation of [20, Section 2.1.2], we say that W_t is a $\mathbb{K}$-valued cylindrical Brownian motion defined on on the filtered space $(\Omega,\mathcal{F},\mathcal{F}_t,\mathbb{P})$ if:

(i) For an arbitrary $t\geq 0$, the mapping $W_t:\mathbb{K}\to \mathrm{L}^2(\Omega;\mathbb{R})$ is linear;

(ii) For an arbitrary $k\in\mathbb{K}$, $W_t(k)$ is an $\mathcal{F}_t$ Brownian motion;

(iii) For arbitrary $k,k'\in\mathbb{K}$ and $t\geq 0$, $\mathbb{E}[W_t(k)W_t(k')]=t\langle k,k'\rangle_{\mathbb{K}}$.

► Remark 36. There is no $\mathbb{K}$-valued process $\tilde{W}_t$ such that:

$$W_t(k)(\omega)=\langle \tilde{W}_t(\omega),k\rangle_{\mathbb{K}}.$$

However, if Q is a non-negative definite symmetric trace-class operator on $\mathbb{K}$, then a $\mathbb{K}-$valued $Q-$Brownian motion can be defined (see e.g. [20, Definition 2.6], [19, Section 4.1]). Moreover, if $\mathbb{K}=\mathbb{R}^m$, then $W_t(k)=\langle\tilde{W}_t,k\rangle_{\mathbb{K}}$, where $\tilde{W}_t$ denotes the standard m-dimensional Brownian motion. Thus, the $\mathbb{R}^m$-cylindrical Brownian motion coincides with the standard m-dimensional Brownian motion.

Besides, let $G:\Omega\times\mathbb{R}_+\to\mathcal{L}_2(\mathbb{K};\mathbb{H})$ be a measurable and $\mathcal{F}_t$-adapted process, then we can define $\int_0^t G(s)dW(s)$ which is the stochastic integral of G, and we have that $G\to\int_0^\cdot G(s)dW(s)$ is an isometry between the measurable and $\mathcal{F}_t-$adapted $\mathcal{L}_2(\mathbb{K};\mathbb{H})-$valued processes and the space of $\mathbb{H}$-valued continuous square-integrable martingales (see [20, Theorem 2.4]).

A.3 Proof of Theorem 2

Proof. Set $F_\varepsilon(x)\stackrel{\text{def}}{=}F(x)+\frac{\varepsilon}{2}\|x\|^2$. Then (DI − TA) can be written equivalently in a a compact form as

$$\dot{x}(t)+\partial F_{\varepsilon(t)}(x(t))\ni 0.$$

Set $h(t)\stackrel{\text{def}}{=}\frac{1}{2}\|x(t)-x^\star\|^2$ where $x^\star=\mathrm{P}_{\mathcal{S}_F}(0)$. Derivation of f and the constitutive equation (DI − TA) give

$$\dot{h}(t)+\langle-\dot{x}(t),x(t)-x^\star\rangle=0, \tag{38}$$

where $-\dot{x}(t)\in\partial F_{\varepsilon(t)}(x(t))$. By strong convexity of $F_{\varepsilon(t)}$, we get

$$F_{\varepsilon(t)}(x^\star)\geq F_{\varepsilon(t)}(x(t))+\langle y(t),x^\star-x(t)\rangle+\frac{\varepsilon(t)}{2}\|x(t)-x^\star\|^2,$$

for every $y(t)\in\partial F_{\varepsilon(t)}(x(t))$. Using that $F_{\varepsilon(t)}(x(t))\geq F_{\varepsilon(t)}(x_{\varepsilon(t)})$, we get

$$F(x^\star)+\frac{\varepsilon(t)}{2}\|x^\star\|^2\geq F(x_{\varepsilon(t)})+\frac{\varepsilon(t)}{2}\|x_{\varepsilon(t)}\|^2+\langle y(t),x^\star-x(t)\rangle+\frac{\varepsilon(t)}{2}\|x(t)-x^\star\|^2,$$

for every $y(t)\in\partial F_{\varepsilon(t)}(x(t))$. Besides, from $F(x^\star)\leq F(x_{\varepsilon(t)})$ we deduce

$$\langle y(t),x(t)-x^\star\rangle\geq\varepsilon(t)h(t)+\frac{\varepsilon(t)}{2}\left(\|x_{\varepsilon(t)}\|^2-\|x^\star\|^2\right), \tag{39}$$

for every $y(t)\in\partial F_{\varepsilon(t)}(x(t))$. Combining (38) with (39) we obtain

$$\dot{h}(t)+\varepsilon(t)h(t)\leq\frac{1}{2}\varepsilon(t)\left(\|x^\star\|^2-\|x_{\varepsilon(t)}\|^2\right). \tag{40}$$

Set $m(t)\stackrel{\text{def}}{=}\exp\int_{t_0}^t\varepsilon(s)ds$. Integrating (40) from t_0 to t, we get

$$h(t)\leq\frac{h(t_0)}{m(t)}+\frac{1}{2m(t)}\int_{t_0}^t m'(s)\left(\|x^\star\|^2-\|x_{\varepsilon(s)}\|^2\right)ds. \tag{41}$$

According to hypothesis (i) and the classical property of the Tikhonov approximation we have $x_{\varepsilon(t)}\to x^\star$, and hence $\|x^\star\|^2-\|x_{\varepsilon(s)}\|^2\to 0$. To pass to the limit on (41) we use hypothesis (ii) which tells us that $m(t)\to+\infty$. Let us complete the argument by using that convergence implies ergodic convergence. Precisely, given $\delta>0$, let $t_\delta>t_0$ such that $|\|x^\star\|^2-\|x_{\varepsilon(s)}\|^2|\leq\delta$ for $s\geq t_\delta$. Then split the integral as follows

$$\begin{aligned} h(t) &\leq \frac{h(t_0)}{m(t)}+\frac{1}{2m(t)}\int_{t_0}^{t_\delta}m'(s)\left(\|x^\star\|^2-\|x_{\varepsilon(s)}\|^2\right)ds+\delta\frac{1}{2m(t)}\int_{t_\delta}^t m'(s)ds && (42)\\ &\leq \frac{h(t_0)}{m(t)}+\frac{1}{2m(t)}\int_{t_0}^{t_\delta}m'(s)\left(\|x^\star\|^2-\|x_{\varepsilon(s)}\|^2\right)ds+\frac{\delta}{2}. && (43)\end{aligned}$$

Then let t tend to infinity, to get $\limsup_{t\to+\infty}h(t)\leq\frac{\delta}{2}$. This being true for any $\delta>0$ gives the result. ◄

A.4 Existence and uniqueness of the SDI

We now prove Theorem 8, which specifies conditions ensuring the existence and uniqueness of a solution to (SDI_0). The argument builds on prior results while addressing aspects not covered in [34].

Proof. The existence of a solution (X, η) in the sense of Definition 5 comes from [34, Theorem 3.5] (see [19, Section 7.1.1] for the SDE case). We now turn to uniqueness. Let (X_1, η_1) and (X_2, η_2) be two solutions of (SDI_0). By Itô's formula, we have

$$\begin{aligned}\|X_1(t) - X_2(t)\|^2 &= 2\int_{t_0}^{t} \langle b(s, X_1(s)) - b(s, X_2(s)), X_1(s) - X_2(s)\rangle ds \\ &- 2\int_{t_0}^{t} \langle \eta_1'(s) - \eta_2'(s), X_1(s) - X_2(s)\rangle ds + \int_{t_0}^{t} \|\sigma(s, X_1(s)) - \sigma(s, X_2(s))\|_{\mathrm{HS}}^2 ds \\ &+ \int_{t_0}^{t} \langle X_1(s) - X_2(s), [\sigma(s, X_1(s)) - \sigma(s, X_2(s))]dW(s)\rangle.\end{aligned}$$

Since for almost all $t \geq 0$, $\eta_i'(t) \in A(X_i(t))$, $i = \{1, 2\}$, by monotonicity of A, we have that for almost all $t \geq t_0$,

$$\langle \eta_1'(t) - \eta_2'(t), X_1(t) - X_2(t)\rangle \geq 0,$$

and thus the second term on the right-hand side is non-positive. Now, let $n \in \mathbb{N}$ arbitrary and consider the stopping time $\tau_n = \inf\{t \geq t_0 : \|X_1(t) - X_2(t)\| \geq n\}$ and evaluate the previous equation at $t \wedge \tau_n$, denoting $X_i^n(t) = X_i(t \wedge \tau_n)$ $(i = \{1, 2\})$, we have

$$\begin{aligned}\|X_1^n(t) - X_2^n(t)\|^2 &\leq 2\int_{t_0}^{t\wedge\tau_n} \langle b(s, X_1(s)) - b(s, X_2(s)), X_1(s) - X_2(s)\rangle ds \\ &+ \int_{t_0}^{t\wedge\tau_n} \|\sigma(s, X_1(s)) - \sigma(s, X_2(s))\|_{\mathrm{HS}}^2 ds \\ &+ \int_{t_0}^{t\wedge\tau_n} \langle X_1(s) - X_2(s), [\sigma(s, X_1(s)) - \sigma(s, X_2(s))]dW(s)\rangle \\ &\leq L(L+2)\int_{t_0}^{t\wedge\tau_n} \|X_1(s) - X_2(s)\|^2 ds \\ &+ \int_{t_0}^{t\wedge\tau_n} \langle X_1(s) - X_2(s), [\sigma(s, X_1(s)) - \sigma(s, X_2(s))]dW(s)\rangle \\ &\leq L(L+2)\int_{t_0}^{t} \|X_1^n(s) - X_2^n(s)\|^2 ds \\ &+ \int_{t_0}^{t\wedge\tau_n} \langle X_1(s) - X_2(s), [\sigma(s, X_1(s)) - \sigma(s, X_2(s))]dW(s)\rangle.\end{aligned}$$

Note that we have used Cauchy-Schwarz inequality and the Lipschitz assumption on (b, σ) in the second inequality. Taking expectation of both sides and using the properties of Itô's integral we obtain

$$\mathbb{E}(\|X_1^n(t) - X_2^n(t)\|^2) \leq L(L+2)\int_{t_0}^{t} \mathbb{E}(\|X_1^n(s) - X_2^n(s)\|^2) ds.$$

By Grönwall's inequality, we obtain that

$$\mathbb{E}(\|X_1^n(t) - X_2^n(t)\|^2) = 0, \forall t \geq t_0, \forall n \in \mathbb{N}.$$

On the other hand, we have that $\lim_{n\to+\infty} t \wedge \tau_n = t$. Therefore, taking $\liminf_{n\to+\infty}$ in the previous expression, using Fatou's Lemma and the fact that X_1, X_2 are a.s. continuous processes, we conclude that $\mathbb{E}(\|X_1(t) - X_2(t)\|^2) = 0$, consequently

$$\mathbb{P}(X_1(t) = X_2(t), \forall t \in [t_0, T]) = 1, \quad \text{for every } T > t_0.$$

Let $T > t_0$ arbitrary, let us prove that $\mathbb{E}\left(\sup_{t\in[t_0,T]}\|X(t)\|^\nu\right) < +\infty$. Using Itô's formula with the solution process X and the anchor function $\phi(x) = \|x - x^\star\|^2$ for $x^\star \in A^{-1}(0)$, we obtain for every $t \in [t_0, T]$:

$$\|X(t) - x^\star\|^2 = \|X_0 - x^\star\|^2 + 2\int_{t_0}^t \langle b(s, X(s)), X(s) - x^\star\rangle ds - 2\int_{t_0}^t \langle \eta'(s), X(s) - x^\star\rangle ds \\ + \int_{t_0}^t \|\sigma(s, X(s))\|_{\mathrm{HS}}^2 ds + 2\int_{t_0}^t \langle X(s) - x^\star, \sigma(s, X(s))dW(s)\rangle.$$

Since $\eta'(t) \in A(X(t))$ for almost all $t \geq 0$, and $0 \in A(x^\star)$, by monotonicity of A we have that for every $t \in [t_0, T]$,

$$\langle \eta'(t), X(t) - x^\star\rangle \geq 0, \quad \text{for almost all } t \geq 0.$$

Thus the second integral is nonnegative, which implies

$$\|X(t) - x^\star\|^2 \leq \|X_0 - x^\star\|^2 + 2\int_{t_0}^t \langle b(s, X(s)), X(s) - x^\star\rangle ds + \int_{t_0}^t \|\sigma(s, X(s))\|_{\mathrm{HS}}^2 ds \\ + 2\int_{t_0}^t \langle X(s) - x^\star, \sigma(s, X(s))dW(s)\rangle. \tag{44}$$

Moreover, we have

$$2\langle b(t,x), x - x^\star\rangle + \|\sigma(t,x)\|_{\mathrm{HS}}^2 \leq 2\|b(t,x)\|\|x - x^\star\| + \|\sigma(t,x)\|_{\mathrm{HS}}^2 \leq C(1 + \|x - x^\star\|^2), \quad \forall t \geq t_0, \forall x \in \mathbb{H}.$$

We now proceed as in the proof of [24, Lemma 3.2] to conclude that $X \in S_{\mathbb{H}}^\nu[t_0]$. In fact, we take power $\frac{\nu}{2}$ at both sides of (44), then using that $(a+b+c)^{\frac{\nu}{2}} \leq 3^{\frac{\nu-2}{2}}(a^{\frac{\nu}{2}} + b^{\frac{\nu}{2}} + c^{\frac{\nu}{2}})$ we have

$$\|X(t) - x^\star\|^\nu \leq 3^{\frac{\nu-2}{2}}\left(\|X_0 - x^\star\|^\nu + C^{\frac{\nu}{2}}\left(\int_{t_0}^t 1 + \|X(s) - x^\star\|^2 ds\right)^{\frac{\nu}{2}}\right) \\ + 3^{\frac{\nu-2}{2}} 2^{\frac{\nu}{2}}\left(\int_{t_0}^t \langle X(s) - x^\star, \sigma(s, X(s))dW(s)\rangle\right)^{\frac{\nu}{2}}.$$

Now taking supremum $t \in [t_0, T]$ and then expectation at both sides, we have that there exists $K = K(\nu, T)$ such that:

$$\mathbb{E}\left(\sup_{t\in[t_0,T]}\|X(t) - x^\star\|^\nu\right) \leq K\left(1 + \mathbb{E}\left(\|X_0 - x^\star\|^\nu\right) + \int_{t_0}^T \mathbb{E}(\|X(s) - x^\star\|^\nu)ds\right) \\ + K\mathbb{E}\left(\sup_{t\in[t_0,T]}\left|\int_{t_0}^t \langle X(s) - x^\star, \sigma(s, X(s))dW(s)\rangle\right|^{\frac{\nu}{2}}\right).$$

By Proposition 37, we get that, for a redefined $K = K(\nu, T)$,

$$\mathbb{E}\left(\sup_{t\in[t_0,T]}\|X(t) - x^\star\|^\nu\right) \leq K\left(1 + \mathbb{E}\left(\|X_0 - x^\star\|^\nu\right) + \int_{t_0}^T \mathbb{E}(\|X(s) - x^\star\|^\nu)ds\right) \\ + K\mathbb{E}\left(\left|\int_{t_0}^T \|X(s) - x^\star\|^2\|\sigma(s, X(s))\|_{\mathrm{HS}}^2 ds\right|^{\frac{\nu}{4}}\right). \tag{45}$$

Note that by Cauchy-Schwarz and Young's inequality,

$$
\begin{aligned}
&\mathbb{E}\left(\left|\int_{t_0}^T \|X(s)-x^\star\|^2\|\sigma(s,X(s))\|_{\mathrm{HS}}^2 ds\right|^{\frac{\nu}{4}}\right)\\
&\leq \mathbb{E}\left(\sup_{t\in[t_0,T]}\|X(t)-x^\star\|^{\frac{\nu}{2}}\left(\int_{t_0}^T\|\sigma(s,X(s))\|_{\mathrm{HS}}^2\right)^{\frac{\nu}{4}}\right)\\
&\leq \frac{1}{2K}\mathbb{E}\left(\sup_{t\in[t_0,T]}\|X(t)-x^\star\|^{\nu}\right)+\frac{K}{2}\mathbb{E}\left[\left(\int_{t_0}^T\|\sigma(s,X(s))\|_{\mathrm{HS}}^2\right)^{\frac{\nu}{2}}\right]\\
&\leq \frac{1}{2K}\mathbb{E}\left(\sup_{t\in[t_0,T]}\|X(t)-x^\star\|^{\nu}\right)+\frac{KC^{\frac{\nu}{2}}}{2}\mathbb{E}\left[\left(\int_{t_0}^T 1+\|X(s)-x^\star\|^2 ds\right)^{\frac{\nu}{2}}\right]\\
&\leq \frac{1}{2K}\mathbb{E}\left(\sup_{t\in[t_0,T]}\|X(t)-x^\star\|^{\nu}\right)+\frac{KC^{\frac{\nu}{2}}}{2}T^{\frac{\nu-2}{2}}\mathbb{E}\left[\left(\int_{t_0}^T (1+\|X(s)-x^\star\|^2)^{\frac{\nu}{2}} ds\right)\right].
\end{aligned}
$$

Substituting this into (45), we have, for a possibly different $K=K(\nu,T)$,

$$
\mathbb{E}\left(\sup_{t\in[t_0,T]}\|X(t)-x^\star\|^{\nu}\right)\leq K\left(1+\mathbb{E}\left(\|X_0-x^\star\|^{\nu}\right)+\int_{t_0}^T\mathbb{E}\left(\sup_{t\in[0,s]}\|X(t)-x^\star\|^{\nu}\right)ds\right).
$$

By Grönwall's inequality, we obtain

$$
\mathbb{E}\left(\sup_{t\in[t_0,T]}\|X(t)-x^\star\|^{\nu}\right)\leq K\left(1+\mathbb{E}\left(\|X_0-x^\star\|^{\nu}\right)\right)e^{KT}<+\infty.
$$

Since $T>t_0$ is arbitrary, we conclude that $X\in S_{\mathbb{H}}^{\nu}[t_0]$. ◀

A.5 On martingales

Proposition 37. *(see* [15] *and* [36, Section 1.2]*) (Burkholder-Davis-Gundy Inequality) Let $p>0$, W be a $\mathbb{K}$-valued cylindrical Brownian motion defined over a filtered probability space $(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\geq 0},\mathbb{P})$ and $g:\Omega\times\mathbb{R}_+\to\mathbb{K}$ a progressively measurable process (with our usual notation $g(t)\stackrel{\text{def}}{=}g(\cdot,t)$) such that*

$$
\mathbb{E}\left[\left(\int_0^T\|g(s)\|^2ds\right)^{\frac{p}{2}}\right]<+\infty,\quad \forall T>0.
$$

Then, there exists $C_p>0$ (only depending on p) for every $T>0$ such that:

$$
\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t\langle g(s),dW(s)\rangle\right|^p\right]\leq C_p\mathbb{E}\left[\left(\int_0^T\|g(s)\|^2ds\right)^{\frac{p}{2}}\right].
$$

Theorem 38. *Let $\mathbb{H}$ be a real 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* s-$\lim_{t\to\infty}M_t=M_\infty$ *a.s..*

Proof. Consider $(M_k)_{k\in\mathbb{N}}$ to be the embedded discrete parameter martingale. Since $\sup_{k\in\mathbb{N}}\mathbb{E}\|M_k\|^2<+\infty$, then $(M_k)_{k\in\mathbb{N}}$ is uniformly integrable and by [40, Theorem 3], there exists a measurable $\mathbb{H}$-valued random variable $M_\infty\in\mathrm{L}^2(\Omega;\mathbb{H})$ such that $\lim_{k\to\infty}\|M_k-M_\infty\|=0$ a.s.. In turn, using the dominated convergence theorem (see [39, Theorem 1.34]), we also have

$$
\lim_{k\to\infty}\mathbb{E}(\|M_k-M_\infty\|^2)=0. \tag{46}
$$

The rest of the proof is inspired by the arguments in the proof of [21, Theorem 2.2].

We consider an arbitrary $k \in \mathbb{N}^*$ and $\delta > 0$. Since $(M_{t+k} - M_k)_{t\geq 0}$ is also a $\mathbb{H}-$valued martingale, we can use Doob's maximal inequalities for $\mathbb{H}-$valued martingales shown in [20, Theorem 2.2], which gives us

$$\delta^2 \mathbb{P}\left(\sup_{s\in[0,t]} \|M_{s+k} - M_k\| > \delta\right) \leq \mathbb{E}(\|M_{t+k} - M_k\|^2). \tag{47}$$

Let $n \in \mathbb{N}^*$ be arbitrary. We have

$$\mathbb{P}\left(\sup_{s\in\mathbb{Q}\cap[0,n]} \|M_{s+k} - M_\infty\| > \delta\right) \leq \mathbb{P}\left(\sup_{s\in\mathbb{Q}\cap[0,n]} \|M_{s+k} - M_k\| > \frac{\delta}{2}\right) + \mathbb{P}\left(\|M_k - M_\infty\| > \frac{\delta}{2}\right).$$

Using (47) and Markov's inequality, we get the bound

$$\begin{aligned}\delta^2 \mathbb{P}\left(\sup_{s\in\mathbb{Q}\cap[0,n]} \|M_{s+k} - M_\infty\| > \delta\right) &\leq 4\mathbb{E}(\|M_{n+k} - M_k\|^2) + 4\mathbb{E}(\|M_k - M_\infty\|^2) \\ &\leq 8\mathbb{E}(\|M_{n+k} - M_\infty\|^2) + 12\mathbb{E}(\|M_k - M_\infty\|^2).\end{aligned} \tag{48}$$

In turn, we get

$$\begin{aligned}\delta^2 \mathbb{P}\left(\sup_{s\in\mathbb{Q}, s\geq k} \|M_s - M_\infty\| > \delta\right) &\leq \delta^2 \mathbb{P}\left(\bigcup_{n\in\mathbb{N}^*} \left\{\sup_{s\in\mathbb{Q}\cap[0,n]} \|M_{s+k} - M_\infty\| > \delta\right\}\right) \\ &\leq \delta^2 \liminf_{n\to\infty} \mathbb{P}\left(\sup_{s\in\mathbb{Q}\cap[0,n]} \|M_{s+k} - M_\infty\| > \delta\right) \\ &\leq 12\mathbb{E}\left(\|M_k - M_\infty\|^2\right),\end{aligned}$$

where we have used (48) and in the last inequality, that $\lim_{n\to\infty} \mathbb{E}(\|M_{n+k} - M_\infty\|^2) = 0$ by (46). Taking $k \to \infty$, and using again (46), we conclude that for all $\delta > 0$

$$\lim_{k\to\infty} \mathbb{P}\left(\sup_{s\in\mathbb{Q}, s\geq k} \|M_s - M_\infty\| > \delta\right) = 0.$$

For $k \in \mathbb{N}^*$, we define $A_k \stackrel{\text{def}}{=} \{\omega \in \Omega : \sup_{s\in\mathbb{Q}, s\geq k} \|M_s(\omega) - M_\infty(\omega)\| > \delta\}$, since $(A_k)_{k\in\mathbb{N}^*}$ is a non-increasing sequence of sets:

$$0 = \lim_{k\to\infty} \mathbb{P}(A_k) = \mathbb{P}\left(\bigcap_{k\in\mathbb{N}^*} A_k\right).$$

Defining for $l \geq 0$, $D_l = \{\omega \in \Omega : \|M_l(\omega) - M_\infty(\omega)\| > \delta\}$, it is direct to check that $\bigcup_{l\geq k, l\in\mathbb{Q}} D_l \subseteq A_k$ for every $k \in \mathbb{N}^*$. Therefore, we obtain that

$$\mathbb{P}\left(\bigcap_{k\in\mathbb{N}^*} \bigcup_{l\geq k, l\in\mathbb{Q}} D_l\right) = 0,$$

which is equivalent to $\text{s-}\lim_{s\to\infty, s\in\mathbb{Q}} M_t = M_\infty$ a.s.. The result follows from classical arguments of continuity of the martingale. ◀

Theorem 39. [29, 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 nonnegative $\mathcal{F}_0$-measurable random variable. Define*

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

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

References

1 M. Akgul. Topics in relaxation and ellipsoidal methods. *Research Notes in mathematics*, 97, 1984.
2 H. Attouch, Z. Chbani, and H. Riahi. Accelerated gradient methods with strong convergence to the minimum norm minimizer: a dynamic approach combining time scaling, averaging, and Tikhonov regularization. *arXiv:2211.10140*, 2022.
3 H. Attouch and R. Cominetti. A dynamical approach to convex minimization coupling approximation with the steepest descent method. *Journal of Differential Equations*, 128(2):519–540, 1996.
4 H. Attouch and M.-O Czarnecki. Asymptotic behavior of coupled dynamical systems with multiscale aspects. *J. Differential Equations*, 248.6:1315–1344, 2010.
5 H. Attouch, A. Moudafi, and H. Riahi. Quantitative stability analysis for maximal monotone operators and semi-groups of contractions. *Nonlinear Analysis, TMA*, 21 (9):697–723, 1993.
6 H. Attouch and R. Wets. Quantitative stability of variational systems: I, the epigraphical distance. *Transactions of the American Mathematical Society*, 328 (2):695–729, 1991.
7 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.
8 J.P Aubin and G. Da Prato. The viability theorem for stochastic differential inclusions. *Stoch. Anal. Appl.*, 16:1–15, 1998.
9 H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer, 2nd edition, 2017.
10 G. Bocsan. On Wiener stochastic integrals of multifunctions. *Univ. Tim. FSN*, 87:1–7, 1987.
11 Jerome Bolte, Trong Phong Nguyen, Juan Peypouquet, and Bruce W. Suter. From error bounds to the complexity of first-order descent methods for convex functions. In Jon Lee and Sven Leyffer, editors, *Mathematical Programming*, volume 165, pages 471–507. Springer, 2016.
12 Haim Brezis. Operateurs maximaux monotones et semi-groupes de contractions dans les espace Hilbert. *North-Holland mathematics studies; 5.*, 1973.
13 F. E. Browder. Existence and approximation of solutions of nonlinear variational inequalities. *Proc. Nat. Acad. Sci. U.S.A*, 56:1080–1086, 1966.
14 F. E. Browder. Convergence of approximants to fixed points of nonexpansive nonlinear mappings in Banach spaces. *Arch. Rational Mech. Anal.*, 24:82–90, 1967.
15 D.L. Burkholder, B. Davis, and R.F. Gundy. Integral inequalities for convex functions of operators on martingales. *Proc. 6th Berkeley Symp. Math. Statistics and Probability*, 2:223–240, 1972.
16 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.
17 Emmanuel Cépa. équations différentielles stochastiques multivoques. *Séminaire de probabilités (Strasbourg)*, 29:86–107, 1995.
18 R. Cominetti, J. Peypouquet, and S. Sorin. Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization. *J. Differential Equations*, 245:3753–3763, 2008.
19 Giuseppe Da Prato and Jerzy Zabszyk. Stochastic equations in infinite dimensions. *Cambridge University Press*, 1992.
20 Leszek Gawarecki and Vidyadhar Mandrekar. Stochastic differential equations in infinite dimensions. *Springer*, 2011.
21 Joe Ghafari. A proof and an application of the continuous parameter martingale convergence theorem. *Research in Statistics*, 1 (1), 2023.
22 T.E. Govindan. Yosida approximations of stochastic differential equations in infinite dimensions and applications. *Springer: Probability Theory and Stochastic modelling*, 79, 2016.
23 J. Gwinner. Bibliography on nondifferentiable optimization and non-smooth analysis. *J. Comput. Appl. Math*, 7:277–285, 1981.
24 Desmond J. Higham, Xuerong Mao, and Andrew M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.*, 40(3):1041–1063, 2006.
25 Michal Kisielewicz. Stochastic differential inclusions and applications. *Springer Optimization and its applications*, 80, 2013.
26 Kiwiel and C Krzysztof. Lagrangian relaxation via ballstep subgradient methods. *Mathematics of Operations Research*, 32 (3):669–686, 2007.
27 Paul Krée. Diffusion equation for multivalued stochastic differential equations. *Journal of Functional Analysis*, 49(1):73–90, 1982.
28 Claude Lemarechal. Lagrangian relaxation. *Computational combinatorial optimization: Papers from the Spring School held in Schloss Dagstuhl*, 2241:112–156, 2001.
29 Xuerong Mao. Stochastic differential equations and applications. *Elsevier*, 2007.

30 Rodrigo Maulen-Soto, Jalal Fadili, and Hedy Attouch. An stochastic differential equation perspective on stochastic convex optimization. *Mathematics of Operations Research*, 2024. arXiv:2207.02750.

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

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

33 Etienne Pardoux and Aurel Rascanu. *Stochastic differential equations, backward SDE's, partial differential equations*. Springer, 2014.

34 Roger Pettersson. Yosida approximations for multivalued stochastic differential equations. *Stochastics and Stochastics reports*, 52:107–120, 1994.

35 Aurel Rascanu. Deterministic and stochastic differential equations in Hilbert spaces involving multivalued maximal monotone operators. *arXiv:1402.0748*, 2014.

36 Aurel Rascanu and Eduard Rotenstein. The Fitzpatrick function-a bridge betweem convex analysis and multivalued stochastic differential equations. *arXiv:0809.4447*, 2009.

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

38 R.T. Rockafellar. Convex analysis. *Princeton university press*, 28, 1997.

39 Walter Rudin. Real and complex analysis. *McGraw-Hill*, 1987.

40 Frank S. Scalora. Abstract martingale convergence theorems. *Pacific J. Math.*, 11 (1):347–374, 1961.

41 N.Z. Shor. Minimization methods for non-differentiable functions. *Springer Series in Computational Mathematics*, 1985.

42 Rodrigo Maulen Soto. *A dynamical system perspective ÎÂ¿n stÎÂ¿chastic and inertial methÎÂ¿ds fÎÂ¿r optimizatiÎÂ¿n.* PhD thesis, Normandie UniversitÃ©, 2024.

43 J. J. Suh, J. Park, and E. K. Ryu. Continuous-time analysis of anchor acceleration. In *Neural Information Processing Systems*, 2023.

44 D. Torralba. Convergence epigraphique et changements d'echelle en analyse variationelle et optimisation. *PhD Thesis, Université de Montpellier 2*, page 160, 1996.