

On a Stochastic Differential Equation with Correction Term Governed by a Monotone and Lipschitz Continuous Operator

Radu Ioan Bot^{*} Chiara Schindler[†]

Abstract

In our pursuit of finding a zero for a monotone and Lipschitz continuous operator $M : \mathbb{R}^n \rightarrow \mathbb{R}^n$ amidst noisy evaluations, we explore an associated differential equation within a stochastic framework, incorporating a correction term. We present a result establishing the existence and uniqueness of solutions for the stochastic differential equations under examination. Additionally, assuming that the diffusion term is square-integrable, we demonstrate the almost sure convergence of the trajectory process $X(t)$ to a zero of M and of $\|M(X(t))\|$ to 0 as $t \rightarrow +\infty$. Furthermore, we provide ergodic upper bounds and ergodic convergence rates in expectation for $\|M(X(t))\|^2$ and $\langle M(X(t), X(t) - x^*) \rangle$, where x^* is an arbitrary zero of the monotone operator. Subsequently, we apply these findings to a minimax problem. Finally, we analyze two temporal discretizations of the continuous-time models, resulting in stochastic variants of the Optimistic Gradient Descent Ascent and Extragradient methods, respectively, and assess their convergence properties.

Key Words. monotone equation, stochastic differential equation, existence and uniqueness of solutions, ergodic upper bounds, ergodic convergence rates, stochastic algorithms for monotone equations

AMS subject classification. 34F05, 47H05, 60H10, 68W20

1 Introduction

For a monotone and L -Lipschitz continuous operator $M : \mathbb{R}^n \rightarrow \mathbb{R}^n$, we examine the equation

$$M(x) = 0. \tag{1}$$

We assume that the set of solutions of (1), denoted by $\text{zer}M := \{x \in \mathbb{R}^n : M(x) = 0\}$, is nonempty. The operator $M : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is said to be monotone if $\langle M(x) - M(y), x - y \rangle \geq 0$ for all $x, y \in \mathbb{R}^n$.

For $M := \nabla f$, where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex and differentiable function with a Lipschitz continuous gradient, solving equation (1) is equivalent to finding the global solutions of the optimization problem

$$\min_{x \in \mathbb{R}^n} f(x). \tag{2}$$

On the other hand, if we define

$$M(x, y) := (\nabla_x \Phi(x, y), -\nabla_y \Phi(x, y)),$$

^{*}Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, e-mail: radu.bot@univie.ac.at. Research partially supported by FWF (Austrian Science Fund), projects W 1260 and P 34922-N, and by a grant of the Romanian Ministry of Research, Innovation and Digitization, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2021-0138, within PNCDI III.

[†]Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, e-mail: chiara.schindler@univie.ac.at.

where $\Phi : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$ is a convex-concave function with Lipschitz continuous gradient, solving equation (1) is equivalent to finding the saddle points of the minimax problem

$$\min_{x \in \mathbb{R}^p} \max_{y \in \mathbb{R}^q} \Phi(x, y). \quad (3)$$

Inspired by the Newton and the Levenberg-Marquardt methods, Attouch and Svaiter introduced in [3] the following continuous time equation associated to (1) in the setting of a real Hilbert space

$$\begin{cases} \dot{x}(t) + \mu(t) \frac{d}{dt} M(x(t)) + \mu(t) M(x(t)) = 0 & \forall t > 0, \\ x(0) = x_0, \end{cases} \quad (4)$$

where $\mu : [0, +\infty) \rightarrow (0, +\infty)$ was assumed to be bounded, continuous and absolutely continuous on bounded intervals. Besides proving the existence and uniqueness of strong global solutions, they showed, under suitable assumptions on the parameter function, that $x(t)$ converges weakly to a zero of M as well as $M(x(t)) \rightarrow 0$ as $t \rightarrow +\infty$.

In many instances, the evaluation of the operator is subject to noise, which motivates us to transfer (4) to a stochastic setting. To achieve this, we consider the formal expression of a continuous time system

$$\begin{cases} dX(t) + \mu(t) dM(X(t)) = -\gamma(t) M(X(t)) dt + \sigma(t, X(t)) dW(t) & \forall t > 0, \\ X(0) = X_0, \end{cases} \quad (5)$$

which can consistently with Itô's chain rule be understood as

$$\begin{cases} d(X(t) + \mu(t) M(X(t))) = -(\gamma(t) - \dot{\mu}(t)) M(X(t)) dt + \sigma(t, X(t)) dW(t) & \forall t > 0, \\ X(0) = X_0, \end{cases} \quad (\text{SDE-M})$$

defined over a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ with the diffusion term $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ being matrix-valued and measurable, W a m -dimensional Brownian motion, and $X(\cdot)$ as well as $M(X(\cdot))$ are stochastic Itô processes with the same m -dimensional Brownian motion W . The parameter functions $\mu : [0, +\infty) \rightarrow (0, +\infty)$ and $\gamma : [0, +\infty) \rightarrow (0, +\infty)$ are assumed to be continuous differentiable and, respectively, integrable. For the diffusion term we assume

$$\begin{cases} \exists c_\sigma > 0 \text{ such that } \|\sigma(t, x') - \sigma(t, x)\|_F \leq c_\sigma \|x' - x\| & \forall t > 0, \\ \sigma_\infty(t) := \sup_{x \in \mathbb{R}^n} \|\sigma(t, x)\|_F \leq \sigma_* & \forall t > 0, \end{cases} \quad (6)$$

where $\|\cdot\|_F$ denotes the Frobenius norm on $\mathbb{R}^{n \times m}$.

Then, (SDE-M) can be rewritten as the system

$$\begin{cases} dM(X(t)) = Y(t) dt + \sigma_M(t) dW(t), \\ dX(t) = (-\mu(t) Y(t) - \gamma(t) M(X(t))) dt + \sigma_X(t) dW(t) & \forall t > 0, \\ X(0) = X_0, \end{cases}$$

where

$$\sigma(t, X(t)) = \sigma_X(t) + \mu(t) \sigma_M(t) \quad \forall t \geq 0$$

and $\sigma_X, \sigma_M : \mathbb{R}_+ \rightarrow \mathbb{R}^{n \times m}$ are measurable.

Our goal is to explore the existence and uniqueness of trajectory solutions for (SDE-M), along with their long-term behavior in terms of both almost sure convergence and convergence rates. Furthermore, we demonstrate how these properties transfer to convergence rates for stochastic numerical methods obtained through temporal discretizations of the stochastic differential equation.

1.1 Related works

The simplest continuous time approach associated to the optimization problem (2), for a convex and differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ with $L_{\nabla f}$ -Lipschitz continuous gradient, is the gradient flow

$$\begin{cases} \dot{x}(t) = -\nabla f(x(t)) & \forall t > 0, \\ x(0) = x_0. \end{cases}$$

Its trajectory solution $x(t)$ converges to a global minimizer of f , while $f(x(t))$ converges to the minimal function value $\min f$ of f with a convergence rate of $o(\frac{1}{t})$ as $t \rightarrow +\infty$.

Its discrete time counterpart, the gradient method,

$$x^{k+1} = x^k - \gamma \nabla f(x^k) \quad \forall k \geq 0,$$

where $x^0 \in \mathbb{R}^n$ and $\gamma \in \left(0, \frac{1}{L_{\nabla f}}\right]$, shares the same convergence properties. Specifically, the sequence $(x_k)_{k \geq 0}$ converges to a minimizer of f , and $f(x_k) - \min f = o(\frac{1}{k})$ as $k \rightarrow +\infty$.

In [25], to accommodate instances where the gradient input is subject to noise, Maulen-Soto, Fadili and Attouch proposed the following stochastic differential equation

$$\begin{cases} dX(t) = -\nabla f(X(t))dt + \sigma(t, X(t))dW(t) & \forall t > 0, \\ X(0) = X_0, \end{cases} \quad (7)$$

defined over a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ with $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ being a matrix-valued measurable diffusion term fulfilling (6), and W a m -dimensional Brownian motion. They derived for $f\left(\int_0^t X(s)ds\right) - \min f$ and $\frac{1}{t} \int_0^t f(X(s))ds - \min f$ upper bounds in expectation of $\mathcal{O}(\frac{1}{t}) + \mathcal{O}(\sigma_*^2)$. Additionally, assuming that σ_∞ is square-integrable, they demonstrated the almost sure convergence of the trajectory process to a minimizer of f and of $f(X(t))$ to $\min f$ as $t \rightarrow +\infty$, and that the two above quantities exhibit convergence rates of $\mathcal{O}(\frac{1}{t})$ as $t \rightarrow +\infty$.

There is a strong link between this stochastic differential equation and the stochastic gradient descent algorithm with constant stepsize ([9], [19], [22], [23], [29], [32])

$$x^{k+1} := x^k - \gamma \nabla f(x^k, \xi^k) \quad \forall k \geq 0,$$

where $(\mathcal{F}_k)_{k \geq 0}$ is an increasing family of σ -fields, $x^0 \in \mathbb{R}^n$ is \mathcal{F}_0 -measurable, $(\xi_k)_{k \geq 0}$ is such that $\nabla f(x, \xi^k)$ corresponds to a noisy observation of $\nabla f(x)$, with the properties that $\mathbb{E}(\nabla f(x, \xi^{k+1}) | \mathcal{F}_k) = \nabla f(x)$ for all $x \in \mathbb{R}^n$ and all $k \geq 0$, and $\mathbb{E}(\|\nabla f(x^*, \xi^{k+1})\|^2 | \mathcal{F}_k) \leq \sigma_{**}^2$ for a global minimizer x^* of f and all $k \geq 0$. Then (see, for instance, [4]) the upper bound in expectation of $f\left(\frac{1}{k} \sum_{i=0}^{k-1} x^i\right) - \min f$ is of $\mathcal{O}(\frac{1}{k}) + \mathcal{O}(\sigma_{**}^2)$.

Recently, in [26], (7) has been enhanced with a Tikhonov regularization term, resulting in a stochastic differential equation which, under the same upper bounds and convergence rates in expectation, in the case when σ_∞ is square-integrable, exhibits strong convergence of the trajectory process to the global minimizer of f of minimum norm.

Furthermore, in [27], the properties of first order stochastic differential equations have been transferred to second order ones using the time scaling and averaging methodology developed in [2]. The latter generalize the second order Langevin process, and demonstrate rapid convergence in expectation of the function values along the trajectory process to the minimal function value.

As for the monotone equation (1), one could replicate the gradient flow approach and associate with it the following continuous dynamics

$$\begin{cases} \dot{x}(t) = -M(x(t)) & \forall t > 0, \\ x(0) = x_0. \end{cases} \quad (8)$$

However, only the ergodic trajectory $\frac{1}{t} \int_0^t x(s) ds$ converges to a zero of M as $t \rightarrow +\infty$ ([5]), while the trajectory $x(t)$ does not converge in general, as can be easily seen for the case of the counterclockwise rotation operator by $\frac{\pi}{2}$ radians in \mathbb{R}^2 ([11]).

If the operator M is ρ -cocoercive with constant $\rho > 0$, meaning that

$$\langle M(x) - M(y), x - y \rangle \geq \rho \|M(x) - M(y)\|^2 \quad \forall x, y \in \mathbb{R}^n,$$

which obviously implies that M is monotone and $\frac{1}{\rho}$ -Lipschitz continuous, then the asymptotic behaviour of the trajectory of (8) improves substantially, even in the presence of small perturbations. Precisely, according to [2, Theorem 11], in this case, the trajectory solution of

$$\begin{cases} \dot{x}(t) = -M(x(t)) + g(t) & \forall t > 0, \\ x(0) = x_0, \end{cases} \quad (9)$$

where $g : [0, +\infty) \rightarrow \mathbb{R}^n$ is such that

$$\int_{t_0}^{+\infty} \|g(t)\| dt < +\infty \quad \text{and} \quad \int_{t_0}^{+\infty} t \|g(t)\|^2 dt < +\infty, \quad (10)$$

converges to a zero of M , and it holds $\|M(x(t))\| = o(\frac{1}{\sqrt{t}})$ as $t \rightarrow +\infty$.

In the case when M is a ρ -cocoercive operator with constant $\rho > 0$, the following stochastic counterpart to (8) has been proposed in [25]

$$\begin{cases} dX(t) = -M(X(t))dt + \sigma(t, X(t))dW(t) & \forall t > 0, \\ X(0) = X_0, \end{cases} \quad (11)$$

defined over a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$, where $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ fulfills (6), and W a m -dimensional Brownian motion. Alongside the existence and uniqueness of the trajectory process for (11), upper bounds in expectation of $\mathcal{O}(\frac{1}{t}) + \mathcal{O}(\sigma_*^2)$ for $\|\frac{1}{t} \int_0^t M(X(s)) ds\|^2$ and $\frac{1}{t} \int_0^t \|X(s)\|^2 ds$ have been derived. In case M is κ -strongly monotone with constant $\kappa > 0$, upper bounds in expectation of $\mathcal{O}(e^{-2\kappa t}) + \mathcal{O}(\sigma_*^2)$ for $\|X(t) - x^*\|^2$, where x^* is the unique zero of M , have been also provided. Additionally, assuming that σ_∞ is square-integrable, it has been proved that there is almost sure convergence of $X(t)$ towards a zero of M and of $\|M(X(t))\|$ towards 0 as $t \rightarrow +\infty$. Moreover, $\|\frac{1}{t} \int_0^t M(X(s)) ds\|^2$ and $\frac{1}{t} \int_0^t \|X(s)\|^2 ds$ exhibit convergence rates of $\mathcal{O}(\frac{1}{t})$ as $t \rightarrow +\infty$.

If M is only monotone and Lipschitz continuous, [3] suggests that incorporating a correction term into the formulation of the differential equation, such as the time derivative of $t \mapsto M(x(t))$ in (4), is crucial to improve the convergence properties of the dynamical system associated with (1). We have extended this idea to the stochastic domain, resulting in the formulation of (SDE-M), and assuming that $M(X(\cdot))$ is an stochastic Itô process.

The significance of the correction term in solving equations governed by monotone and L -Lipschitz continuous operators is well-reflected in the realm of numerical algorithms. On one hand, the forward method,

$$x^{k+1} := x^k - \gamma M(x^k) \quad \forall k \geq 0,$$

does not generally converge, as observed, for instance, with the counterclockwise rotation operator by $\frac{\pi}{2}$ radians in \mathbb{R}^2 .

On the other hand, for the Optimistic Gradient Descent Ascent method ([31], [6])

$$x^{k+1} := x^k - 2\gamma M(x^k) + \gamma M(x^{k-1}) = x^k - \gamma M(x^k) - \gamma(M(x^k) - M(x^{k-1})) \quad \forall k \geq 1,$$

which can be regarded as a natural discretization of (4), if $0 < \gamma < \frac{1}{2L}$, then the generated sequence of iterates converges to a zero of M . Additionally, if $0 < \gamma < \frac{1}{16L}$, then the method exhibits a best-iterate

convergence rate for the norm of the operator evaluated along the iterates of $\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$ as $k \rightarrow +\infty$ ([10], [14], [15]).

Furthermore, for the Extragradient method ([1], [21])

$$\begin{cases} y^k := x^k - \gamma M(x^k) \\ x^{k+1} := x^k - \gamma M(y^k) \end{cases} \quad \forall k \geq 0,$$

which also involves a correction term, if $0 < \gamma < \frac{1}{L}$, then the generated sequence of iterates converges to a zero of M and $\|M(x^k)\|$ converges to 0 with a convergence rate of $\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$ as $k \rightarrow +\infty$ ([17]).

For completeness, we would like to mention that recently, in [7], a second-order dynamical system with momentum term and correction term associated with a monotone equation has been introduced. Alongside the convergence of the generated trajectory to a zero of the operator, the system exhibits convergence rates of $\mathcal{O}\left(\frac{1}{t}\right)$ for the norm of the operator and the gap function along the trajectory as $t \rightarrow +\infty$. Furthermore, an explicit numerical algorithm replicating the convergence properties of the continuous time dynamics has been obtained. These are the best-known convergence rate results for continuous time and explicit discrete time approaches for monotone equations.

1.2 Our contributions

In Section 2, we discuss the existence and uniqueness of a trajectory solution for the continuous time system (SDE-M). The presence of the corrector term requires quite involved analysis. Under the assumption that σ_∞ is square-integrable, we also prove the almost sure convergence of $X(t)$ to a zero of M and of $\|M(X(t))\|$ to 0 as $t \rightarrow +\infty$.

In Section 3, in the framework of assumption (6) for the diffusion term, we establish upper bounds in expectation for $\frac{1}{t} \int_0^t \|M(X(s))\|^2 ds$ and $\frac{1}{t} \int_0^t \langle X(s) - x^*, M(X(s)) \rangle ds$, where x^* is a zero of M . These bounds are of the form $\mathcal{O}\left(\frac{1}{t\mu(t)}\right) + \mathcal{O}(\sigma_*^2)$ and $\mathcal{O}\left(\frac{1}{t}\right) + \mathcal{O}(\sigma_*^2)$, respectively. Additionally, assuming that σ_∞ is square-integrable, we show that the squared norm of the operator $\|M(X(t))\|^2$ and the gap function $\langle X(t) - x^*, M(X(t)) \rangle$ exhibit ergodic convergence rates in expectation of $\mathcal{O}\left(\frac{1}{t\mu(t)}\right)$ and $\mathcal{O}\left(\frac{1}{t}\right)$, respectively, as $t \rightarrow +\infty$. In case M is κ -strongly monotone with constant $\kappa > 0$, we also derive upper bounds in expectation of $\mathcal{O}(e^{-2\kappa t}) + \mathcal{O}(\sigma_*^2)$ for $\|X(t) - x^*\|^2$, where x^* is the unique zero of M .

In Section 4, we demonstrate how these results transfer from continuous time to discrete time settings within the context of the stochastic Optimistic Gradient Descent Ascent (OGDA) method and of the stochastic Extragradient (EG) method, designed to solve (1), in the circumstances that the evaluation of M is subject to noise.

2 Continuous time system: existence, uniqueness and almost sure convergence

This section is dedicated to the study of (SDE-M) in the setting

$$\begin{cases} d(X(t) + \mu(t)M(X(t))) = -(\gamma(t) - \dot{\mu}(t))M(X(t))dt + \sigma(t, X(t))dW(t) \quad \forall t > 0, \\ X(0) = X_0. \end{cases}$$

Throughout this section, we will assume that the parameter functions satisfy the following conditions:

$$0 < \mu_{\text{low}} \leq \mu(t) \leq \mu_{\text{up}} := \mu(0) \text{ and } 0 < \gamma_{\text{low}} \leq \gamma(t) \leq \gamma_{\text{up}} \quad \forall t \geq 0, \text{ and } \mu \text{ is nonincreasing on } [0, +\infty). \quad (12)$$

Throughout the paper, $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ is a filtered probability space. The expectation of a random variable $\xi : \Omega \rightarrow \mathbb{R}^n$ is denoted by $\mathbb{E}(\xi) := \int_\Omega \xi(\omega) d\mathbb{P}(\omega)$. An event $E \in \mathcal{F}$ happens almost surely (a.s.) if $\mathbb{P}(E) = 1$.

An \mathbb{R}^n -valued stochastic process is a function $X : \Omega \times \mathbb{R}_+ \rightarrow \mathbb{R}^n$. It is said to be continuous if $X(\omega, \cdot)$ is continuous on \mathbb{R}_+ for almost all $\omega \in \Omega$. For simplicity, we will denote $X(t) := X(\cdot, t)$. Two stochastic processes $X, Y : [0, T] \rightarrow \mathbb{R}^n$, for $T > 0$, are said to be equivalent if $X(t) = Y(t)$ almost surely for all $t \in [0, T]$. This allows the definition of the equivalence relation \mathcal{R} which associates equivalent stochastic processes to the same class.

Next, we will introduce some notions that will characterize the space where the trajectory solution of (SDE-M) lies.

Definition 2.1. (i) A stochastic process $X : \Omega \times \mathbb{R}_+ \rightarrow \mathbb{R}^n$ is called progressively measurable if for every $t \geq 0$, the mapping

$$\Omega \times [0, t] \rightarrow \mathbb{R}^n, \quad (\omega, s) \mapsto X(\omega, s)$$

is $\mathcal{F}_t \otimes \mathcal{B}([0, t])$ -measurable, where \otimes denotes the product σ -algebra and \mathcal{B} is the Borel σ -algebra. Further, X is called \mathcal{F}_t -adapted if $X(\cdot, t)$ is \mathcal{F}_t -measurable for every $t \geq 0$.

For $T > 0$, we define the quotient space as

$$S_n^0[0, T] := \{X : \Omega \times [0, T] \rightarrow \mathbb{R}^n : X \text{ is a progressively measurable continuous stochastic process}\} / \mathcal{R},$$

and set $S_n^0 := \bigcap_{T \geq 0} S_n^0[0, T]$.

(ii) For $\nu > 0$ and $T > 0$, we define $S_n^\nu[0, T]$ as the following subset of stochastic processes in $S_n^0[0, T]$

$$S_n^\nu[0, T] := \left\{ X \in S_n^0[0, T] : \mathbb{E} \left(\sup_{t \in [0, T]} \|X(t)\|^\nu \right) < +\infty \right\}.$$

Finally, we set $S_n^\nu := \bigcap_{T \geq 0} S_n^\nu[0, T]$.

The following results establishes the existence and uniqueness of a solution of (SDE-M).

Theorem 2.2. Let the diffusion term $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ satisfy assumption (6). Then (SDE-M) has a unique solution $X \in S_n^\nu$, where $\nu \geq 2$.

Proof. Let $\nu \geq 2$. First, we establish the result on $[0, T]$, for $T > 0$ fixed. We rewrite the stochastic differential equations as

$$X(t) = (\text{Id} + \mu(t)M)^{-1} \underbrace{(X(t) + \mu(t)M(X(t)))}_{=: Z(t)} =: J_{\mu(t)M}(Z(t)) \quad \forall t \in [0, T],$$

where $J_{\mu(t)M} := (\text{Id} + \mu(t)M)^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ denotes the resolvent of M with parameter $\mu(t)$ and Id denotes the identity operator on \mathbb{R}^n . Denoting the Yosida approximation of M with parameter $\mu(t)$ by

$$M_{\mu(t)} := \frac{1}{\mu(t)} (\text{Id} - J_{\mu(t)M}),$$

a simple calculation yields that $M(X(t)) = M_{\mu(t)}(Z(t))$ for all $t \in [0, T]$. Hence,

$$dZ(t) = d(X(t) + \mu(t)M(X(t))) = (-\gamma(t) + \dot{\mu}(t))M_{\mu(t)}(Z(t))dt + \sigma(t, J_{\mu(t)M}(Z(t)))dW(t) \quad \forall t \in [0, T]. \quad (13)$$

Now setting

$$F(t, z) := (\dot{\mu}(t) - \gamma(t))M_{\mu(t)}(z), \quad G(t, z) := \sigma(t, J_{\mu(t)M}(z)) \quad \forall z \in \mathbb{R}^n \quad \forall t \in [0, T],$$

yields for all $x, y \in \mathbb{R}^n$ and all $t \in [0, T]$

$$\begin{aligned}
& \|F(t, x) - F(t, y)\| + \|G(t, x) - G(t, y)\|_F \\
&= |\gamma(t) - \dot{\mu}(t)| \|M_{\mu(t)}(x) - M_{\mu(t)}(y)\| + \|\sigma(t, J_{\mu(t)M}(x)) - \sigma(t, J_{\mu(t)M}(y))\|_F \\
&\leq (\gamma_{\text{up}} - \dot{\mu}_{\text{low}, T}) \frac{1}{\mu(t)} \|x - y\| + c_\sigma \|J_{\mu(t)M}(x) - J_{\mu(t)M}(y)\| \\
&\leq \left((\gamma_{\text{up}} - \dot{\mu}_{\text{low}, T}) \frac{1}{\mu_{\text{low}}} + c_\sigma \right) \|x - y\|,
\end{aligned}$$

where $\dot{\mu}_{\text{low}, T} := \inf\{\dot{\mu}(t) : t \in [0, T]\} \in \mathbb{R}$, since the continuous mapping $\dot{\mu}$ is bounded on the compact interval $[0, T]$.

This allows us to use Theorem A.1 to obtain the existence of a unique stochastic process $Z_T \in S_n^\mu[0, T]$ solving (13). Therefore, a reasonable candidate to solve (SDE-M), given by

$$X_T(t) := J_{\mu(t)M}(Z_T(t)) \quad \forall t \in [0, T],$$

is uniquely defined. It remains to show that it does indeed solve (SDE-M) and lies in $S_n^\mu[0, T]$.

We define

$$Z_T(t) := X_T(t) + \mu(t)M(X_T(t)) \quad \forall t \in [0, T].$$

Since now $M(X_T(t)) = M_{\mu(t)}(Z_T(t))$ and Z_T is a solution of (13), it follows

$$\begin{aligned}
d(X_T(t) + \mu(t)M(X_T(t))) &= (-\gamma(t) + \dot{\mu}(t))M_{\mu(t)}(Z_T(t))dt + \sigma(t, J_{\mu(t)M}(Z_T(t)))dW(t) \\
&= (-\gamma(t) + \dot{\mu}(t))M(X_T(t))dt + \sigma(t, X_T(t))dW(t) \quad \forall t \in [0, T],
\end{aligned}$$

and thus, X_T is indeed a solution of (SDE-M) on $[0, T]$.

Now, recall that $J_{\mu(t)M}$ is 1-Lipschitz continuous and observe that for all $x \in \mathbb{R}^n$ and all $\mu, \lambda > 0$ it holds

$$\begin{aligned}
\|J_{\lambda M}(x) - J_{\mu M}(x)\| &= \left\| J_{\mu M} \left(\frac{\mu}{\lambda}x + \left(1 - \frac{\mu}{\lambda}\right) J_{\lambda M}(x) \right) - J_{\mu M}(x) \right\| \\
&\leq \left\| \left(1 - \frac{\mu}{\lambda}\right) (x - J_{\lambda M}(x)) \right\| \\
&\leq |\lambda - \mu| \|M_\lambda x\|.
\end{aligned}$$

This will allow us to establish the continuity of X_T . Indeed, for all $t, s \in [0, T]$ it holds

$$\begin{aligned}
\|X_T(t) - X_T(s)\| &\leq \|J_{\mu(t)M}(Z_T(t)) - J_{\mu(s)M}(Z_T(t))\| + \|J_{\mu(s)M}(Z_T(t)) - J_{\mu(s)M}(Z_T(s))\| \\
&\leq |\mu(t) - \mu(s)| \|M_{\mu(t)}(Z_T(t))\| + \|Z_T(t) - Z_T(s)\|,
\end{aligned}$$

and, since Z_T is continuous by Theorem A.1 and μ is continuous by assumption, the continuity of X_T follows.

Next, consider

$$\begin{aligned}
\mathbb{E} \left(\sup_{t \in [0, T]} \|X_T(t)\|^\nu \right) &= \mathbb{E} \left(\sup_{t \in [0, T]} \|J_{\mu(t)M}(Z_T(t)) - 0\|^\nu \right) \\
&= \mathbb{E} \left(\sup_{t \in [0, T]} \|J_{\mu(t)M}(Z_T(t)) - J_{\mu(t)M}(\mu(t)M(0))\|^\nu \right) \\
&\leq \mathbb{E} \left(\sup_{t \in [0, T]} \|Z_T(t) - \mu(t)M(0)\|^\nu \right) \\
&\leq \mathbb{E} \left(2^\nu \sup_{t \in [0, T]} \left(\frac{1}{2} \|Z_T(t)\| + \frac{\mu_{\text{up}}}{2} \|M(0)\| \right)^\nu \right) \\
&\leq 2^{\nu-1} \left(\mathbb{E} \left(\sup_{t \in [0, T]} \|Z_T(t)\|^\nu \right) + \mu_{\text{up}}^\nu \|M(0)\|^\nu \right) < +\infty.
\end{aligned}$$

Finally, in order to prove the progressive measurability of X , we first notice that the mapping

$$\Omega \times [0, T] \rightarrow [0, T] \times \mathbb{R}^n, \quad (\omega, s) \mapsto (s, Z_T(\omega, s)) \quad (14)$$

is measurable. Indeed, the preimage of an element of the generating set $I \times O$, where $I \subseteq [0, T]$ is an interval and $O \subseteq \mathbb{R}^n$ an element of the generator of the Borel- σ -algebra on \mathbb{R}^n , is given by

$$Z_T^{-1}(O) \cap (\Omega \times I),$$

which is measurable, due to the measurability of Z_T and the fact that a finite intersection of measurable sets is again measurable. On the other hand, from the above estimate for the resolvent operator and its Lipschitz continuity, we see that the mapping

$$[0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad (s, z) \mapsto J_{\mu(s)M}(z),$$

is continuous and therefore measurable. This means that its composition with the mapping in (14) is also measurable. However, this composition is given by

$$\Omega \times [0, T] \rightarrow \mathbb{R}^n, \quad (\omega, s) \mapsto J_{\mu(s)M}(Z_T(\omega, s)) = X_T(\omega, s),$$

which proves that X_T is measurable, as claimed.

Since X_T is uniquely defined on any $[0, T]$, where $T > 0$, for $0 < T_1 < T_2$, we have $X_{T_2}|_{[0, T_1]} = X_{T_1}$. Thus, there exists a unique solution $X \in S_n^\nu$ of (SDE-M). \square

The following result establishes the almost sure convergence of $X(t)$ to a $\text{zer}M$ -valued random variable as $t \rightarrow +\infty$. The notion of a martingale will play an important role in its proof.

Definition 2.3. *Let X be a real-valued stochastic process such that $\mathbb{E}(|X(t)|) < +\infty$ for all $t \geq 0$.*

(i) *The σ -algebra generated by the random variables $X(s)$ for $0 \leq s \leq t$,*

$$\sigma(t) := \sigma(X(s) | 0 \leq s \leq t),$$

is called the history of the stochastic process X until (and including) time t .

(ii) *If*

$$X(s) = \mathbb{E}(X(t) | \sigma(s)) \quad \text{a.s.} \quad \forall t \geq s \geq 0,$$

then X is called a martingale.

Theorem 2.4. *Let $\text{zer}M \neq \emptyset$, the diffusion term $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ satisfy assumption (6), and X be the unique trajectory process of (SDE-M). If, in addition to (12), $\int_0^{+\infty} \sigma_\infty(s)^2 ds < +\infty$, then the following statements are true:*

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

(ii) *it holds $\sup_{t \geq 0} \|X(t)\| < +\infty$ a.s.*

If, in addition, $-\infty < \dot{\mu}_{up} \leq \dot{\mu}(t)$ for all $t \geq 0$ and $L\mu_{up} < 1$, then:

(iii) *almost surely $\lim_{t \rightarrow +\infty} \|M(X(t))\| = 0$ and $\lim_{t \rightarrow +\infty} \|X(t) - x^*\|$ exists and is finite for all $x^* \in \text{zer}M$;*

(iv) *there exists a $\text{zer}M$ -valued random variable x^* such that $\lim_{t \rightarrow +\infty} X(t) = x^*$ a.s.*

Proof. In order to perform the proof, we rewrite (SDE-M) as

$$\begin{cases} dM(X(t)) = Y(t)dt + \sigma_M(t)dW(t), \\ dX(t) = (-\mu(t)Y(t) - \gamma(t)M(X(t)))dt + \sigma_X(t)dW(t) \quad \forall t > 0, \\ X(0) = X_0, \end{cases}$$

where $\sigma(t, X(t)) = \sigma_X(t) + \mu(t)\sigma_M(t)$ for all $t \geq 0$ and $Y(t)$ is a stochastic process whose existence is guaranteed by $M(X(t))$ being an Itô process. This alternative form will be essential in the upcoming arguments, since we will be using Itô's chain rule on an anchor function which takes an additional argument to be filled by $M(X(t))$. Here, Itô's formula requires the process $Y(t)$ as an analogue to the derivative of $M(X(t))$ in the deterministic setting. Let $x^* \in \text{zer}M$, and define the anchor function $\phi : \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\phi(t, x, z) := \mu(t)\langle z, x - x^* \rangle + \frac{1}{2}\|x - x^*\|^2 + \frac{\mu(t)^2}{2}\|z\|^2 = \frac{1}{2}\|x + \mu(t)z - x^*\|^2.$$

The proposed anchor function takes an additional argument – this will be taken equal to $M(X(t))$ over the course of the proof. For all $(t, x, z) \in \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^n$ it holds

$$\begin{aligned} \frac{d}{dt}\phi(t, x, z) &= \dot{\mu}(t)\langle z, x - x^* \rangle + \mu(t)\dot{\mu}(t)\|z\|^2 \\ \nabla_x\phi(t, x, z) &= \mu(t)z + x - x^* \\ \nabla_z\phi(t, x, z) &= \mu(t)(x - x^*) + \mu(t)^2z \\ \nabla_x^2\phi(t, x, z) &= I \\ \nabla_x\nabla_z\phi(t, x, z) &= \nabla_z\nabla_x\phi(t, x, z) = \mu(t)I \\ \nabla_z^2\phi(t, x, z) &= \mu(t)^2I. \end{aligned}$$

Observe that the remaining second partial derivatives of ϕ are identically 0. Hence, the Itô formula yields for all $t \geq 0$

$$\begin{aligned} \phi(t, X(t), M(X(t))) &= \phi(0, X(0), M(X(0))) \\ &+ \int_0^t (\dot{\mu}(s)\langle M(X(s)), X(s) - x^* \rangle + \mu(s)\dot{\mu}(s)\|M(X(s))\|^2) ds \\ &+ \int_0^t \langle (\mu(s)M(X(s)) + X(s) - x^*), (-\mu(s)Y(s) - \gamma(s)M(X(s))) \rangle ds \\ &+ \int_0^t \langle \sigma_X^\top(s)(\mu(s)M(X(s)) + X(s) - x^*), dW(s) \rangle \\ &+ \int_0^t \langle (\mu(s)(X(s) - x^*) + \mu(s)^2M(X(s))), Y(s) \rangle ds \\ &+ \int_0^t \langle \sigma_M^\top(s)(\mu(s)(X(s) - x^*) + \mu(s)^2M(X(s))), dW(s) \rangle \\ &+ \frac{1}{2} \int_0^t \text{tr}(\sigma_X^\top(s)\sigma_X(s)) + 2\mu(s)\text{tr}(\sigma_M^\top(s)\sigma_X(s)) + \mu(s)^2\text{tr}(\sigma_M^\top(s)\sigma_M(s)) ds. \end{aligned}$$

We observe for all $s \geq 0$

$$\sigma_X^\top(s)\sigma_X(s) + 2\mu(s)\sigma_M^\top(s)\sigma_X(s) + \mu(s)^2(\sigma_M^\top(s)\sigma_M(s)) = \sigma^\top(s, X(s))\sigma(s, X(s)),$$

and denote this term by $\Sigma(s, X(s))$ for improved readability. Further, it holds for all $s \geq 0$

$$\begin{aligned} &\langle \sigma_X^\top(s)(\mu(s)M(X(s)) + X(s) - x^*) + \sigma_M^\top(s)(\mu(s)^2M(X(s)) + \mu(s)(X(s) - x^*)), dW(s) \rangle \\ &= \langle (\sigma_X(s) + \mu(s)\sigma_M(s))^\top(\mu(s)M(X(s)) + X(s) - x^*), dW(s) \rangle \\ &= \langle \sigma^\top(s, X(s))(\mu(s)M(X(s)) + X(s) - x^*), dW(s) \rangle. \end{aligned}$$

Then, performing some elementary computations, we get for all $t \geq 0$

$$\begin{aligned}
\phi(t, X(t), M(X(t))) &= \underbrace{\mu(0)\langle M(X_0), X_0 - x^* \rangle + \frac{1}{2}\|X_0 - x^*\|^2 + \frac{\mu(0)^2}{2}\|M(X_0)\|^2}_{=:\xi} \\
&\quad - \underbrace{\int_0^t (\gamma(s) - \dot{\mu}(s)) (\langle M(X(s)), X(s) - x^* \rangle + \mu(s)\|M(X(s))\|^2) ds}_{=:U(t)} \\
&\quad + \underbrace{\int_0^t \langle \sigma^\top(s, X(s))(\mu(s)M(X(s)) + X(s) - x^*), dW(s) \rangle}_{=:N(t)} \\
&\quad + \underbrace{\frac{1}{2} \int_0^t \text{tr}(\Sigma(s, X(s))) ds}_{=:A(t)},
\end{aligned} \tag{15}$$

in other words

$$\phi(t, X(t), M(X(t))) = \xi - U(t) + N(t) + A(t) \quad \forall t \geq 0.$$

(i) As $X \in S_n^2$, it holds for all $T > 0$ that

$$\begin{aligned}
&\mathbb{E} \left(\int_0^T \|\sigma^\top(s, X(s))(X(s) - x^* + \mu(s)M(X(s)))\|^2 ds \right) \\
&\leq \mathbb{E} \left(\int_0^T 2(\|\sigma^\top(s, X(s))(X(s) - x^*)\|^2 + \mu_{\text{up}}^2 \|\sigma^\top(s, X(s))(M(X(s)) - M(x^*))\|^2) ds \right) \\
&\leq \mathbb{E} \left(2 \sup_{t \in [0, T]} \|X(t) - x^*\|^2 \int_0^T \sigma_\infty^2(s) ds + \mu_{\text{up}}^2 \sup_{t \in [0, T]} \|M(X(t)) - M(x^*)\|^2 \int_0^T \sigma_\infty^2(s) ds \right) \\
&\leq 2(1 + \mu_{\text{up}}^2 L^2) \mathbb{E} \left(\sup_{t \in [0, T]} \|X(t) - x^*\|^2 \right) \int_0^T \sigma_\infty^2(s) ds < +\infty.
\end{aligned}$$

By Proposition A.2, this means that N is a square-integrable continuous martingale, and $\mathbb{E}(N(t)) = 0$ for all $t \geq 0$. Taking the expectation of (15) and using that

$$0 \leq \text{tr}(\Sigma(s, X(s))) \leq \sigma_\infty^2(s), \quad \langle M(X(s)), X(s) - x^* \rangle \geq 0, \quad \text{and} \quad \|M(X(s))\|^2 \geq 0 \quad \forall s \geq 0,$$

yields for all $t \geq 0$

$$\begin{aligned}
\mathbb{E} \left(\frac{1}{2} \|X(t) - x^*\|^2 \right) &\leq \mathbb{E}(\phi(t, X(t))) \\
&= \frac{1}{2} \|X_0 - x^*\|^2 + \mu_{\text{up}} \langle M(X_0), X_0 - x^* \rangle + \frac{\mu_{\text{up}}^2}{2} \|M(X_0)\|^2 \\
&\quad - \underbrace{\mathbb{E}(U(t))}_{\geq 0} + \underbrace{\mathbb{E}(N(t))}_{=0} + \mathbb{E}(A(t)) \\
&\leq \frac{1}{2} \|X_0 - x^*\|^2 + \mu(0) \langle M(X_0), X_0 - x^* \rangle + \frac{\mu(0)^2}{2} \|M(X_0)\|^2 + \frac{1}{2} \int_0^t \sigma_\infty^2(s) ds < +\infty.
\end{aligned}$$

Finally, by taking the supremum over $t \geq 0$, we establish the validity of the first claim.

(ii) A and U are continuous adapted increasing processes with $A(0) = U(0) = 0$ a.s., while $\lim_{t \rightarrow +\infty} A(t) < +\infty$ holds due to the integrability condition for the diffusion term. Thus, we can use Theorem A.3 to obtain almost surely that $\lim_{t \rightarrow +\infty} \phi(t, X(t), M(X(t)))$ exists and is finite, and

$$\int_0^{+\infty} (\gamma(s) - \dot{\mu}(s)) (\langle M(X(s)), X(s) - x^* \rangle + \mu(s) \|M(X(s))\|^2) ds < +\infty. \quad (16)$$

In other words, for all $x^* \in \text{zer}M$ there exists $\Omega_{x^*} \in \mathcal{F}$ such that $\mathbb{P}(\Omega_{x^*}) = 1$ and for all $\omega \in \Omega_{x^*}$ it holds that

$$\lim_{t \rightarrow +\infty} \mu(t) \langle M(X(\omega, t)), X(\omega, t) - x^* \rangle + \frac{1}{2} \|X(\omega, t) - x^*\|^2 + \frac{\mu(t)^2}{2} \|M(X(\omega, t))\|^2 \text{ exists and is finite.} \quad (17)$$

This implies that for all $x^* \in \text{zer}M$ there exists $\Omega_{x^*} \in \mathcal{F}$ such that $\mathbb{P}(\Omega_{x^*}) = 1$ and $\|X(\omega, \cdot) - x^*\|$ remains bounded for all $\omega \in \Omega_{x^*}$. This proves that $\sup_{t \geq 0} \|X(t)\| < +\infty$ a.s.

(iii) Using the monotonicity of M and the fact that μ and γ are bounded from below by a positive constant, by (16), it follows $\int_0^{+\infty} \|M(X(\omega, s))\|^2 ds < +\infty$ a.s. We denote by Ω_I the set of probability one for which it holds $\int_0^{+\infty} \|M(X(\omega, s))\|^2 ds < +\infty$ for all $\omega \in \Omega_I$.

Let $R(t) = \int_0^t \sigma(s, X(s)) dW(s)$. This is a continuous martingale with respect to \mathcal{F}_t , which, according to the Itô isometry property, satisfies

$$\mathbb{E}(\|R(t)\|^2) = \mathbb{E} \left(\int_0^t \|\sigma(s, X(s))\|_F^2 ds \right) \leq \mathbb{E} \left(\int_0^\infty \sigma_\infty^2(s) ds \right) < +\infty \quad \forall t \geq 0.$$

By Theorem A.5, there exists a \mathbb{R}^n -valued random variable R_∞ with respect to \mathcal{F}_∞ , the σ -algebra generated by $\cup_{t \geq 0} \mathcal{F}_t$, which satisfies $\mathbb{E}(\|R_\infty\|^2) < +\infty$, and

$$\lim_{t \rightarrow +\infty} R(\omega, t) = R_\infty(\omega) \quad \text{for every } \omega \in \Omega_R,$$

where $\Omega_R \in \mathcal{F}$ is such that $\mathbb{P}(\Omega_R) = 1$.

Then $\mathbb{P}(\Omega_I \cap \Omega_R) = 1$ and, as we will see below, $\lim_{t \rightarrow +\infty} \|M(X(\omega, t))\| = 0$ for all $\omega \in \Omega_I \cap \Omega_R$. Indeed, for all $\omega \in \Omega_I \cap \Omega_R$, since $\int_0^{+\infty} \|M(X(\omega, s))\|^2 ds < +\infty$, it holds $\liminf_{t \rightarrow +\infty} \|M(X(\omega, t))\| = 0$. In addition, for all $\omega \in \Omega_I \cap \Omega_R$, $\limsup_{t \rightarrow +\infty} \|M(X(\omega, t))\| = 0$.

Suppose by contradiction that there exists $\omega_0 \in \Omega_I \cap \Omega_R$ such that $\limsup_{t \rightarrow +\infty} \|M(X(\omega_0, t))\| > 0$. Then, by Lemma A.4, there exist $\delta > 0$ satisfying

$$0 = \liminf_{t \rightarrow +\infty} \|M(X(\omega_0, t))\| < \delta < \limsup_{t \rightarrow +\infty} \|M(X(\omega_0, t))\|,$$

and a sequence $(t_k)_{k \geq 0} \subseteq \mathbb{R}_+$ such that

$$\lim_{k \rightarrow \infty} t_k = +\infty, \quad \|M(X(\omega, t_k))\| > \delta, \quad \text{and} \quad t_{k+1} - t_k > 1 \quad \forall k \geq 0.$$

Let $\alpha > 0$ such that $1 - (1 + \alpha)L^2\mu_{\text{up}}^2 > 0$ and $\varepsilon > 0$ such that

$$\frac{\varepsilon}{2(|\dot{\mu}_{\text{low}}|^2 + \gamma_{\text{up}}^2)} < 1 \quad \text{and} \quad \left(3 + \frac{3}{\alpha}\right) \frac{\varepsilon L^2}{1 - (1 + \alpha)L^2\mu_{\text{up}}^2} < \frac{\delta^2}{2}.$$

Then $\left[t_k, t_k + \frac{\varepsilon}{2(|\dot{\mu}_{\text{low}}|^2 + \gamma_{\text{up}}^2)}\right]$ are disjoint intervals for any two distinct $k \geq 0$.

Also, by the convergence property of $R(\omega_0, \cdot)$ and the fact that $\int_0^{+\infty} \|M(X(\omega_0, s))\|^2 ds < +\infty$, there exists $k_0 \geq 0$ such that for every $k \geq k_0$,

$$\sup_{t \geq t_k} \|R(\omega_0, t) - R(\omega_0, t_k)\|^2 < \frac{\varepsilon}{2} \quad \text{and} \quad \int_{t_k}^{+\infty} \|M(X(\omega_0, s))\|^2 ds < 1.$$

For all $k \geq k_0$ and all $t \in \left[t_k, t_k + \frac{\varepsilon}{2(|\dot{\mu}_{\text{low}}|^2 + \gamma_{\text{up}}^2)} \right]$, it holds

$$\begin{aligned}
& \|X(\omega_0, t) - X(\omega_0, t_k)\|^2 \\
= & \|X(\omega_0, t) + \mu(t)M(X(\omega_0, t)) - X(t_k) - \mu(t_k)M(X(\omega_0, t_k)) + \mu(t_k)M(X(\omega_0, t_k)) - \mu(t)M(X(\omega_0, t))\|^2 \\
\leq & (1 + \alpha)\|\mu(t_k)(M(X(\omega_0, t_k)) - M(X(\omega_0, t)))\|^2 \\
& + \left(1 + \frac{1}{\alpha}\right) \left\| \int_{t_k}^t (\dot{\mu}(s) - \gamma(s))M(X(\omega_0, s))ds + \int_{t_k}^t \sigma(s, X(\omega_0, s))dW(s) + (\mu(t_k) - \mu(t))M(X(\omega_0, t)) \right\|^2 \\
\leq & (1 + \alpha)\mu_{\text{up}}^2 \|M(X(\omega_0, t_k)) - M(X(\omega_0, t))\|^2 \\
& + \left(3 + \frac{3}{\alpha}\right) \left(\left\| \int_{t_k}^t (\dot{\mu}(s) - \gamma(s))M(X(\omega_0, s))ds \right\|^2 + \left\| \int_{t_k}^t \sigma(s, X(\omega_0, s))dW(s) \right\|^2 \right. \\
& \quad \left. + |\mu(t_k) - \mu(t)|^2 \|M(X(\omega_0, t))\|^2 \right) \\
\leq & (1 + \alpha)\mu_{\text{up}}^2 \|M(X(\omega_0, t_k)) - M(X(\omega_0, t))\|^2 \\
& + \left(3 + \frac{3}{\alpha}\right) \left(2(|\dot{\mu}_{\text{low}}|^2 + |\gamma_{\text{up}}|^2)(t - t_k) \int_{t_k}^t \|M(X(\omega_0, s))\|^2 ds + \|R(\omega_0, t) - R(\omega_0, t_k)\|^2 + \right. \\
& \quad \left. + 4\mu_{\text{up}}^2 \|M(X(\omega_0, t))\|^2 \right) \\
\leq & (1 + \alpha)\mu_{\text{up}}^2 \|M(X(\omega_0, t_k)) - M(X(\omega_0, t))\|^2 \\
& + \left(3 + \frac{3}{\alpha}\right) \left(2(t - t_k)(|\dot{\mu}_{\text{low}}|^2 + \gamma_{\text{up}}^2) + \frac{\varepsilon}{2} + 4\mu_{\text{up}}^2 \|M(X(\omega_0, t))\|^2 \right) \\
\leq & (1 + \alpha)\mu_{\text{up}}^2 \|M(X(\omega_0, t_k)) - M(X(\omega_0, t))\|^2 + \left(3 + \frac{3}{\alpha}\right) (\varepsilon + 4\mu_{\text{up}}^2 \|M(X(\omega_0, t))\|^2),
\end{aligned}$$

where the first inequality is the one between the geometric and the arithmetic mean, the second inequality follows from the Cauchy-Schwarz inequality and the third inequality follows from the Hölder inequality for integrals.

By making use of the Lipschitz continuity of M , it yields

$$\begin{aligned}
& \|M(X(\omega_0, t_k)) - M(X(\omega_0, t))\|^2 \\
\leq & (1 + \alpha)L^2\mu_{\text{up}}^2 \|M(X(\omega_0, t_k)) - M(X(\omega_0, t))\|^2 + \left(3 + \frac{3}{\alpha}\right) L^2 (\varepsilon + 4\mu_{\text{up}}^2 \|M(X(\omega_0, t))\|^2),
\end{aligned}$$

consequently,

$$\|M(X(\omega_0, t_k)) - M(X(\omega_0, t))\|^2 \leq \left(3 + \frac{3}{\alpha}\right) \frac{L^2}{1 - (1 + \alpha)L^2\mu_{\text{up}}^2} (\varepsilon + 4\mu_{\text{up}}^2 \|M(X(\omega_0, t))\|^2).$$

Therefore, for all $k \geq k_0$ and all $t \in \left[t_k, t_k + \frac{\varepsilon}{2(|\dot{\mu}_{\text{low}}|^2 + \gamma_{\text{up}}^2)} \right]$, from

$$\|M(X(\omega_0, t))\|^2 \geq \frac{1}{2} \|M(X(\omega_0, t_k))\|^2 - \|M(X(\omega_0, t)) - M(X(\omega_0, t_k))\|^2,$$

we get

$$\left(1 + \left(3 + \frac{3}{\alpha}\right) \frac{4\mu_{\text{up}}^2 L^2}{1 - (1 + \alpha)L^2\mu_{\text{up}}^2}\right) \|M(X(\omega_0, t))\|^2 \geq \frac{\delta^2}{2} - \left(3 + \frac{3}{\alpha}\right) \frac{\varepsilon L^2}{1 - (1 + \alpha)L^2\mu_{\text{up}}^2}.$$

Finally,

$$\begin{aligned}
& \left(1 + \left(3 + \frac{3}{\alpha} \right) \frac{4\mu_{\text{up}}^2 L^2}{1 - (1 + \alpha)L^2 \mu_{\text{up}}^2} \right) \int_0^{+\infty} \|M(X(\omega_0, s))\|^2 ds \\
& \geq \sum_{k \geq k_0} \int_{t_k}^{t_k + \frac{\varepsilon}{2(|\mu_{\text{low}}|^2 + \gamma_{\text{up}}^2)}} \|M(X(\omega_0, s))\|^2 ds \\
& \geq \sum_{k \geq k_0} \frac{\varepsilon}{2(|\mu_{\text{low}}|^2 + \gamma_{\text{up}}^2)} \left(\frac{\delta^2}{2} - \left(3 + \frac{3}{\alpha} \right) \frac{\varepsilon L^2}{1 - (1 + \alpha)L^2 \mu_{\text{up}}^2} \right) = +\infty,
\end{aligned}$$

which contradicts $\int_0^{+\infty} \|M(X(\omega_0, s))\|^2 ds < +\infty$. This means that for all $\omega \in \Omega_I \cap \Omega_R$

$$0 = \liminf_{t \rightarrow +\infty} \|M(X(\omega, t))\| = \limsup_{t \rightarrow +\infty} \|M(X(\omega, t))\| = \lim_{t \rightarrow +\infty} \|M(X(\omega, t))\| = 0.$$

Then, for all $x^* \in \text{zer}M$ and all $\omega \in \Omega_{x^*} \cap \Omega_I \cap \Omega_R$, the Cauchy-Schwarz inequality gives

$$0 \leq \lim_{t \rightarrow +\infty} \langle M(X(\omega, t)), X(\omega, t) - x^* \rangle \leq \lim_{t \rightarrow +\infty} \underbrace{\|M(X(\omega, t))\|}_{\rightarrow 0} \underbrace{\|X(\omega, t) - x^*\|}_{\text{is bounded}} = 0.$$

Using (17) and that $\mu(t) \geq \mu_{\text{low}} > 0$ for all $t \geq 0$, it follows that $\lim_{t \rightarrow +\infty} \|X(\omega, t) - x^*\|$ exists and is finite.

Next, we will employ an argument from [25] to demonstrate the existence of a subset of Ω of probability one, which is independent of a previously chosen $x^* \in \text{zer}M$, such that for all ω in this subset and all $x^* \in \text{zer}M$ the limit $\lim_{t \rightarrow +\infty} \|X(\omega, t) - x^*\|$ exists and is finite.

Taking into account that $\text{zer}M$ is closed, by the separability of \mathbb{R}^n there exists a countable set $S \subseteq \mathbb{R}^n$ which is dense in $\text{zer}M$. Because S is countable,

$$\mathbb{P} \left(\bigcap_{s \in S} \Omega_s \right) = 1 - \mathbb{P} \left(\bigcup_{s \in S} \Omega_s^c \right) \geq 1 - \sum_{s \in S} \mathbb{P}(\Omega_s^c) = 1.$$

Then $\mathbb{P} \left(\bigcap_{s \in S} \Omega_s \cap \Omega_I \cap \Omega_R \right) = 1$. Let $\omega \in \bigcap_{s \in S} \Omega_s \cap \Omega_I \cap \Omega_R$ be fixed and choose $x^* \in \text{zer}M$. Then there exists a sequence $(s^k)_{k \geq 0} \subseteq S$ such that $s^k \rightarrow x^*$ as $k \rightarrow +\infty$. For every $k \geq 0$, since $\omega \in \Omega_{s^k}$,

$$\lim_{t \rightarrow +\infty} \|X(\omega, t) - s^k\| \text{ exists and is finite.}$$

Applying the triangle inequality, it follows

$$\left| \|X(\omega, t) - s^k\| - \|X(\omega, t) - x^*\| \right| \leq \|s^k - x^*\| \quad \forall k \geq 0 \quad \forall t \geq 0.$$

This gives for all $k \geq 0$

$$\begin{aligned}
-\|s^k - x^*\| + \lim_{t \rightarrow +\infty} \|X(\omega, t) - s^k\| & \leq \liminf_{t \rightarrow +\infty} \|X(\omega, t) - x^*\| \\
& \leq \limsup_{t \rightarrow +\infty} \|X(\omega, t) - x^*\| \leq \lim_{t \rightarrow +\infty} \|X(\omega, t) - s^k\| + \|s^k - x^*\|.
\end{aligned}$$

Letting $k \rightarrow +\infty$, we get

$$\lim_{t \rightarrow +\infty} \|X(\omega, t) - x^*\| = \lim_{k \rightarrow +\infty} \lim_{t \rightarrow +\infty} \|X(\omega, t) - s^k\| \in \mathbb{R}.$$

Finally, we 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 $\omega \in \Omega_{\text{cont}}$. Let $\omega \in \Omega_{\text{converge}} := \bigcap_{s \in S} \Omega_s \cap \Omega_I \cap \Omega_R \cap \Omega_{\text{cont}}$. It holds $\mathbb{P}(\Omega_{\text{converge}}) = 1$. For $x^* \in \text{zer}M$,

there exists $C(\omega) \in \mathbb{R}$ and $T(\omega) \leq 0$ such that $\|X(\omega, t) - x^*\| \leq C(\omega)$ for all $t \geq T(\omega)$. Because $X(\omega, \cdot)$ is continuous, it holds

$$\sup_{t \in [0, T(\omega)]} \|X(\omega, t)\| < +\infty.$$

Thus,

$$\sup_{t \geq 0} \|X(\omega, t)\| \leq \max \left(\sup_{t \in [0, T(\omega)]} \|X(\omega, t)\|, C(\omega) + \|x^*\| \right) < +\infty.$$

(iv) We will use Opial's Lemma to prove the convergence of the trajectory process. We fix $\omega \in \Omega_{\text{converge}}$ and recall that above we proved that for every $x^* \in \text{zer}M$ the limit $\lim_{t \rightarrow +\infty} \|X(\omega, t) - x^*\|$ exists. In addition, $X(\omega, \cdot)$ is bounded, therefore, its set of limit points is not empty. Let $\bar{x}(\omega)$ be such a limit point, which means that there exists a sequence $(t_k)_{k \geq 0} \subseteq \mathbb{R}_+$ such that

$$\lim_{k \rightarrow \infty} X(\omega, t_k) = \bar{x}(\omega).$$

From $\lim_{t \rightarrow +\infty} \|M(X(\omega, t))\| = 0$ and the continuity of M , we see that $\bar{x}(\omega) \in \text{zer}M$. Since both conditions in Opial's Lemma are satisfied, there exists $x^*(\omega) \in \text{zer}M$ such that $\lim_{t \rightarrow +\infty} X(\omega, t) = x^*(\omega)$.

Since $\omega \in \Omega_{\text{converge}}$ was arbitrarily chosen, there exists a $\text{zer}M$ -valued random variable x^* such that $\lim_{t \rightarrow +\infty} X(t) = x^*$ almost surely. \square

3 Continuous time system: convergence rates in expectation

In this section we will provide upper bounds and convergence rates in expectation for the ergodic squared norm of the operator and the ergodic gap function. In addition, in case M is κ -strongly monotone with constant $\kappa > 0$, we derive upper bounds in expectation for the squared distance of the trajectory process to the unique zero of M .

We recall that $M : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is called κ -strongly monotone with constant $\kappa > 0$ if $\langle Mx - My, x - y \rangle \geq \kappa \|x - y\|^2$ for all $x, y \in \mathbb{R}^n$.

Regarding the parameter functions, throughout this section, we will assume that:

$$0 < \mu(t) \leq \mu_{\text{up}} := \mu(0) \quad \forall t \geq 0 \text{ and } \mu \text{ is nonincreasing on } [0, +\infty). \quad (18)$$

Theorem 3.1. *Let X be a trajectory process of (SDE-M) and $x^* \in \text{zer}M$. Then the following statements are true:*

(i) *Assume that γ is nonincreasing on $[0, +\infty)$. For all $t > 0$ it holds*

$$\mathbb{E} \left(\frac{1}{t} \int_0^t \langle X(s) - x^*, M(X(s)) \rangle ds \right) \leq \frac{1}{\gamma(t)t} \left(\frac{\mu_{\text{up}}^2 L^2}{2} + \mu_{\text{up}} L + \frac{1}{2} \right) \text{dist}(X_0, \text{zer}M)^2 + \frac{\sigma_*^2}{2\gamma(t)\mu(t)}$$

and

$$\mathbb{E} \left(\frac{1}{t} \int_0^t \|M(X(s))\|^2 ds \right) \leq \frac{1}{\gamma(t)\mu(t)t} \left(\frac{\mu_{\text{up}}^2 L^2}{2} + \mu_{\text{up}} L + \frac{1}{2} \right) \text{dist}(X_0, \text{zer}M)^2 + \frac{\sigma_*^2}{2\gamma(t)\mu(t)}.$$

If, in addition, $\int_0^{+\infty} \sigma_\infty^2(s) ds < +\infty$, then

$$\mathbb{E} \left(\frac{1}{t} \int_0^t \langle X(s) - x^*, M(X(s)) \rangle ds \right) = \mathcal{O} \left(\frac{1}{\gamma(t)t} \right) \quad \text{as } t \rightarrow +\infty.$$

and

$$\mathbb{E} \left(\frac{1}{t} \int_0^t \|M(X(s))\|^2 ds \right) = \mathcal{O} \left(\frac{1}{\gamma(t)\mu(t)t} \right) \quad \text{as } t \rightarrow +\infty$$

(ii) Assume that $0 < \gamma_{low} \leq \gamma(t)$ for all $t \geq 0$. For all $t > 0$ it holds

$$\mathbb{E} \left(\frac{1}{t} \int_0^t \langle X(s) - x^*, M(X(s)) \rangle ds \right) \leq \frac{1}{\gamma_{low} t} \left(\frac{\mu_{up}^2 L^2}{2} + \mu_{up} L + \frac{1}{2} \right) \text{dist}(X_0, \text{zer}M)^2 + \frac{\sigma_*^2}{2\gamma_{low}\mu(t)}$$

and

$$\mathbb{E} \left(\frac{1}{t} \int_0^t \|M(X(s))\|^2 ds \right) \leq \frac{1}{\gamma_{low}\mu(t)t} \left(\frac{\mu_{up}^2 L^2}{2} + \mu_{up} L + \frac{1}{2} \right) \text{dist}(X_0, \text{zer}M)^2 + \frac{\sigma_*^2}{2\gamma_{low}\mu(t)}.$$

If, in addition, $\int_0^{+\infty} \sigma_\infty^2(s) ds < +\infty$, then

$$\mathbb{E} \left(\frac{1}{t} \int_0^t \langle X(s) - x^*, M(X(s)) \rangle ds \right) = \mathcal{O} \left(\frac{1}{t} \right) \quad \text{as } t \rightarrow +\infty.$$

and

$$\mathbb{E} \left(\frac{1}{t} \int_0^t \|M(X(s))\|^2 ds \right) = \mathcal{O} \left(\frac{1}{\mu(t)t} \right) \quad \text{as } t \rightarrow +\infty$$

(iii) Assume that $0 < \gamma_{low} \leq \gamma(t)$ for all $t \geq 0$. If M is κ -strongly monotone with constant $\kappa \geq \frac{1}{2\mu_{up}}$ and $X_0 \neq x^*$, then for all $t \geq 0$ it holds

$$\mathbb{E} \left(\frac{1}{2} \|X(t) - x^*\|^2 \right) \leq \left(\frac{1}{2} \|X_0 - x^*\|^2 + \mu_{up} \langle M(X_0), X_0 - x^* \rangle + \frac{\mu_{up}^2}{2} \|M(X_0)\|^2 \right) e^{-\frac{\gamma_{low}}{2\mu_{up}} t} + \frac{\sigma_*^2 \mu_{up}}{\gamma_{low}}.$$

If, in addition, σ_∞ is decreasing and vanishes at $+\infty$, then for all $\lambda \in (0, 1)$ and all $t > 0$ it holds

$$\begin{aligned} \mathbb{E} \left(\frac{1}{2} \|X(t) - x^*\|^2 \right) &\leq \left(\frac{1}{2} \|X_0 - x^*\|^2 + \mu_{up} \langle M(X_0), X_0 - x^* \rangle + \frac{\mu_{up}^2}{2} \|M(X_0)\|^2 \right) e^{-\frac{\gamma_{low}}{2\mu_{up}} t} \\ &\quad + \frac{\sigma_*^2 \mu_{up}}{\gamma_{low}} e^{-\frac{\gamma_{low}}{2\mu_{up}} (1-\lambda)t} + \frac{\mu_{up}}{\gamma_{low}} \sigma_\infty^2(\lambda t). \end{aligned}$$

Proof. (i) For all $t \geq 0$ we set (see (15))

$$\begin{aligned} \phi(t, X(t), M(X(t))) &= \mu(0) \langle M(X_0), X_0 - x^* \rangle + \frac{1}{2} \|X_0 - x^*\|^2 + \frac{\mu(0)^2}{2} \|M(X_0)\|^2 \\ &\quad - \int_0^t (-\dot{\mu}(s) + \gamma(s)) (\langle M(X(s)), X(s) - x^* \rangle + \mu(s) \|M(X(s))\|^2) ds \\ &\quad + \int_0^t \langle \sigma^\top(s, X(s)) (\mu(s) M(X(s)) + X(s) - x^*), dW(s) \rangle \\ &\quad + \frac{1}{2} \int_0^t \text{tr}(\Sigma(s, X(s))) ds \end{aligned}$$

and $G(t) := \mathbb{E}(\phi(t, X(t), M(X(t))))$. Then, using that the expectation of the stochastic integral is equal to zero and that $\dot{\mu}(t) \leq 0$, for all $t \geq 0$ it yields

$$\begin{aligned} G(t) - G(0) &= - \mathbb{E} \left(\int_0^t (-\dot{\mu}(s) + \gamma(s)) (\langle M(X(s)), X(s) - x^* \rangle + \mu(s) \|M(X(s))\|^2) ds \right) \\ &\quad + \frac{1}{2} \mathbb{E} \left(\int_0^t \text{tr}(\Sigma(s, X(s))) ds \right) \\ &\leq - \gamma(t) \mathbb{E} \left(\int_0^t \langle M(X(s)), X(s) - x^* \rangle ds \right) - \gamma(t) \mu(t) \mathbb{E} \left(\int_0^t \|M(X(s))\|^2 ds \right) + \frac{\sigma_*^2 t}{2}. \end{aligned} \tag{19}$$

Rearranging and using that $G(t) \geq 0$, we obtain for all $t > 0$

$$\begin{aligned} & \gamma(t)\mathbb{E} \left(\frac{1}{t} \int_0^t \langle M(X(s)), X(s) - x^* \rangle ds \right) + \gamma(t)\mu(t)\mathbb{E} \left(\frac{1}{t} \int_0^t \|M(X(s))\|^2 ds \right) \\ & \leq \frac{1}{t} \left(\mu(0)\langle M(X_0), X_0 - x^* \rangle + \frac{1}{2}\|X_0 - x^*\|^2 + \frac{\mu(0)^2}{2}\|M(X_0)\|^2 \right) + \frac{\sigma_*^2}{2} \\ & \leq \frac{1}{t} \left(\mu(0)L + \frac{1}{2} + \frac{\mu(0)^2 L^2}{2} \right) \|X_0 - x^*\|^2 + \frac{\sigma_*^2}{2}. \end{aligned}$$

Taking the infimum over all $x^* \in \text{zer}M$ yields the result.

If $\int_0^{+\infty} \sigma_\infty^2(s) ds < +\infty$, then we have for all $t > 0$

$$\begin{aligned} & G(t) - G(0) \\ & \leq -\gamma(t)\mathbb{E} \left(\int_0^t \langle M(X(s)), X(s) - x^* \rangle ds \right) - \gamma(t)\mu(t)\mathbb{E} \left(\int_0^t \|M(X(s))\|^2 ds \right) + \frac{1}{2}\mathbb{E} \left(\int_0^{+\infty} \sigma_\infty^2(s) ds \right), \end{aligned}$$

which, after rearrangement, gives

$$\begin{aligned} & \gamma(t)\mathbb{E} \left(\frac{1}{t} \int_0^t \langle M(X(s)), X(s) - x^* \rangle ds \right) + \gamma(t)\mu(t)\mathbb{E} \left(\frac{1}{t} \int_0^t \|M(X(s))\|^2 ds \right) \\ & \leq \frac{1}{t} \left(\mu(0)\langle M(X_0), X_0 - x^* \rangle + \frac{1}{2}\|X_0 - x^*\|^2 + \frac{\mu(0)^2}{2}\|M(X_0)\|^2 \right) + \frac{1}{2t}\mathbb{E} \left(\int_0^{+\infty} \sigma_\infty^2(s) ds \right), \end{aligned}$$

thus proving

$$\mathbb{E} \left(\frac{1}{t} \int_0^t \langle X(s) - x^*, M(X(s)) \rangle ds \right) = \mathcal{O} \left(\frac{1}{\gamma(t)t} \right) \text{ as } t \rightarrow +\infty$$

and

$$\mathbb{E} \left(\frac{1}{t} \int_0^t \|M(X(s))\|^2 ds \right) = \mathcal{O} \left(\frac{1}{\gamma(t)\mu(t)t} \right) \text{ as } t \rightarrow +\infty.$$

(ii) The proof follows in the same lines as that of statement (i).

(iii) By using the strong monotonicity of the operator, for all $t_2 > t_1 \geq 0$ it holds

$$\begin{aligned} & \mathbb{E} \left(\int_{t_1}^{t_2} (\dot{\mu}(s) - \gamma(s))(\langle M(X(s)), X(s) - x^* \rangle + \mu(s)\|M(X(s))\|^2) ds \right) \\ & \leq \mathbb{E} \left(\int_{t_1}^{t_2} (-\mu(s)\gamma(s)\|M(X(s))\|^2 - \gamma(s)\langle X(s) - x^*, M(X(s)) \rangle) ds \right) \\ & \leq \gamma_{\text{low}}\mathbb{E} \left(\int_0^t (-\mu(s)\|M(X(s))\|^2 - \langle X(s) - x^*, M(X(s)) \rangle) ds \right) \\ & \leq \frac{\gamma_{\text{low}}}{2\mu_{\text{up}}}\mathbb{E} \left(\int_{t_1}^{t_2} (-2\mu(s)\mu_{\text{up}}\|M(X(s))\|^2 - 2\mu_{\text{up}}\langle X(s) - x^*, M(X(s)) \rangle) ds \right) \\ & \leq \frac{\gamma_{\text{low}}}{2\mu_{\text{up}}}\mathbb{E} \left(\int_{t_1}^{t_2} (-2\mu(s)^2\|M(X(s))\|^2 - \mu(s)\langle X(s) - x^*, M(X(s)) \rangle - \mu_{\text{up}}\kappa\|X(s) - x^*\|^2) ds \right) \\ & \leq \frac{\gamma_{\text{low}}}{2\mu_{\text{up}}}\mathbb{E} \left(\int_{t_1}^{t_2} \left(-\frac{\mu(s)^2}{2}\|M(X(s))\|^2 - \mu(s)\langle X(s) - x^*, M(X(s)) \rangle - \frac{1}{2}\|X(s) - x^*\|^2 \right) ds \right) \\ & = -\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}}\int_{t_1}^{t_2} G(s) ds. \end{aligned}$$

Hence, according to (15), for all $t_2 > t_1 \geq 0$ it holds

$$G(t_2) \leq G(t_1) - \frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} \int_{t_1}^{t_2} G(s) ds + \frac{1}{2} \int_{t_1}^{t_2} \sigma_{\infty}^2(s) ds \leq G(t_1) - \frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} \int_{t_1}^{t_2} G(s) ds + \frac{1}{2} \int_{t_1}^{t_2} \sigma_*^2 ds.$$

The solution of the ordinary differential equation

$$\begin{cases} \varphi'(t) = -\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} \varphi(t) + \frac{\sigma_*^2}{2} & \forall t > 0 \\ \varphi(0) = G(0) \end{cases}$$

is

$$\varphi(t) = G(0)e^{-\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} t} + \frac{\sigma_*^2 \mu_{\text{up}}}{\gamma_{\text{low}}} (1 - e^{-\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} t}) \quad \forall t \geq 0,$$

hence, invoking Lemma A.6, it holds

$$\mathbb{E} \left(\frac{1}{2} \|X(t) - x^*\|^2 \right) \leq G(t) \leq G(0)e^{-\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} t} + \frac{\sigma_*^2 \mu_{\text{up}}}{\gamma_{\text{low}}} \quad \forall t \geq 0,$$

and the claim follows.

For σ_{∞} decreasing and vanishing at $+\infty$, we consider the ordinary differential equations

$$\begin{cases} \varphi'(t) = -\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} \varphi(t) + \frac{\sigma_{\infty}^2(t)}{2} & \forall t > 0 \\ \varphi(0) = G(0). \end{cases}$$

Employing a technique used in [25], we choose $\lambda \in (0, 1)$ and derive for its solution the following estimates, that hold for all $t \geq 0$

$$\begin{aligned} \varphi(t) &= G(0)e^{-\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} t} + e^{-\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} t} \int_0^t \frac{\sigma_{\infty}^2(s)}{2} e^{\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} s} ds \\ &\leq G(0)e^{-\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} t} + e^{-\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} t} \left(\int_0^{\lambda t} \frac{\sigma_{\infty}^2(s)}{2} e^{\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} s} ds + \int_{\lambda t}^t \frac{\sigma_{\infty}^2(s)}{2} e^{\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} s} ds \right) \\ &\leq G(0)e^{-\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} t} + e^{-\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} t} \left(\frac{\sigma_*^2}{2} \int_0^{\lambda t} e^{\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} s} ds + \frac{\sigma_{\infty}^2(\lambda t)}{2} \int_{\lambda t}^t e^{\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} s} ds \right) \\ &\leq G(0)e^{-\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} t} + e^{-\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} t} \left(\frac{\sigma_*^2 \mu_{\text{up}}}{\gamma_{\text{low}}} e^{\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} \lambda t} + \frac{2\mu_{\text{up}}}{\gamma_{\text{low}}} \frac{\sigma_{\infty}^2(\lambda t)}{2} e^{\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} t} \right). \end{aligned}$$

Hence, invoking again Lemma A.6, it holds

$$G(t) \leq G(0)e^{-\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} t} + \frac{\sigma_*^2 \mu_{\text{up}}}{\gamma_{\text{low}}} e^{-\frac{\gamma_{\text{low}}}{2\mu_{\text{up}}} (1-\lambda)t} + \frac{\mu_{\text{up}}}{\gamma_{\text{low}}} \sigma_{\infty}^2(\lambda t) \quad \forall t \geq 0,$$

proving the desired result. \square

Remark 3.2. While the boundedness of γ from above is essential in the proof of Theorem 2.4, the convergence rates in Theorem 3.1 hold also for γ not bounded from above.

Remark 3.3. Finding the saddle points of the minimax problem (3)

$$\min_{x \in \mathbb{R}^p} \max_{y \in \mathbb{R}^q} \Phi(x, y),$$

where $\Phi : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$ is a convex-concave function with L -Lipschitz continuous gradient, is equivalent to solving equation (1) for

$$M : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}, \quad M(x, y) := (\nabla_x \Phi(x, y), -\nabla_y \Phi(x, y)).$$

The corresponding stochastic differential equation (SDE-M) reads

$$\begin{cases} d(X(t) + \mu(t)\nabla_x \Phi(X(\cdot), Y(\cdot))) = -(\gamma(t) - \dot{\mu}(t))\nabla_x \Phi(X(\cdot), Y(\cdot))dt + \sigma_1(t, X(t), Y(t))dW(t), \\ d(Y(t) - \mu(t)\nabla_y \Phi(X(\cdot), Y(\cdot))) = (\gamma(t) - \dot{\mu}(t))\nabla_y \Phi(X(\cdot), Y(\cdot))dt + \sigma_2(t, X(t), Y(t))dW(t) \quad \forall t > 0, \\ X(0) = X_0, Y(0) = Y_0, \end{cases} \quad (\text{SDE-minimax})$$

and is defined over a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ with $\sigma_1 : \mathbb{R}_+ \times \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}^{p \times m}$ and $\sigma_2 : \mathbb{R}_+ \times \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}^{q \times m}$ measurable such that $\begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix} : \mathbb{R}_+ \times \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}^{(p+q) \times m}$ satisfies (6), W a m -dimensional Brownian motion, $X(\cdot), Y(\cdot), \nabla_x \Phi(X(\cdot), Y(\cdot))$, and $\nabla_y \Phi(X(\cdot), Y(\cdot))$ are stochastic Itô processes with the same m -dimensional Brownian motion W , and parameter functions $\mu : [0, +\infty) \rightarrow (0, +\infty)$ and $\gamma : [0, +\infty) \rightarrow (0, +\infty)$ assumed to be continuous differentiable and, respectively, integrable and to fulfill (12).

The system (SDE-minimax) has a unique trajectory solution $(X, Y) \in S_{p+q}^\nu$, for all $\nu \geq 2$. The trajectory (X, Y) satisfies all the statements of Theorem 2.4 under the corresponding hypotheses. Furthermore, for the gap function it holds almost surely $\lim_{t \rightarrow +\infty} \Phi(X(t), y^*) - \Phi(x^*, Y(t)) = 0$ for any saddle point (x^*, y^*) of (3). This is a straightforward consequence of the fact that for all $t \geq 0$

$$\begin{aligned} & \|(\nabla_x \Phi(X(t), Y(t)), -\nabla_y \Phi(X(t), Y(t)))\| \| (X(t), Y(t)) - (x^*, y^*) \| \\ & \geq \langle (\nabla_x \Phi(X(t), Y(t)), -\nabla_y \Phi(X(t), Y(t))), (X(t), Y(t)) - (x^*, y^*) \rangle \\ & = \langle \nabla_x \Phi(X(t), Y(t)), X(t) - x^* \rangle + \langle -\nabla_y \Phi(X(t), Y(t)), Y(t) - y^* \rangle \\ & \geq \Phi(X(t), Y(t)) - \Phi(x^*, Y(t)) - \Phi(X(t), Y(t)) + \Phi(X(t), y^*) \\ & = \Phi(X(t), y^*) - \Phi(x^*, Y(t)) \geq 0. \end{aligned} \quad (20)$$

The trajectory (X, Y) satisfies also all the statements of Theorem 3.1 under the corresponding hypotheses. Furthermore, for any saddle point (x^*, y^*) of (3) it holds

$$\begin{aligned} & \mathbb{E} \left(\Phi \left(\frac{1}{t} \int_0^t X(s) ds, y^* \right) - \Phi \left(x^*, \frac{1}{t} \int_0^t Y(s) ds \right) \right) \\ & \leq \frac{1}{\gamma(t)t} \left(\frac{\mu_{\text{up}}^2 L^2}{2} + \mu_{\text{up}} L + \frac{1}{2} \right) \text{dist}((X_0, Y_0), \text{zer}M)^2 + \frac{\sigma_*^2}{2\gamma(t)\mu(t)} \quad \forall t > 0, \end{aligned}$$

if γ is nonincreasing on $[0, +\infty)$, and

$$\begin{aligned} & \mathbb{E} \left(\Phi \left(\frac{1}{t} \int_0^t X(s) ds, y^* \right) - \Phi \left(x^*, \frac{1}{t} \int_0^t Y(s) ds \right) \right) \\ & \leq \frac{1}{\gamma_{\text{low}} t} \left(\frac{\mu_{\text{up}}^2 L^2}{2} + \mu_{\text{up}} L + \frac{1}{2} \right) \text{dist}((X_0, Y_0), \text{zer}M)^2 + \frac{\sigma_*^2}{2\gamma_{\text{low}}\mu(t)} \quad \forall t > 0, \end{aligned}$$

if $0 < \gamma_{\text{low}} \leq \gamma(t)$ for all $t \geq 0$. If, in addition, $\int_0^{+\infty} \sigma_\infty^2(s) ds < +\infty$, then

$$\mathbb{E} \left(\Phi \left(\frac{1}{t} \int_0^t X(s) ds, y^* \right) - \Phi \left(x^*, \frac{1}{t} \int_0^t Y(s) ds \right) \right) = \mathcal{O} \left(\frac{1}{\gamma(t)t} \right) \quad \text{as } t \rightarrow +\infty$$

and

$$\mathbb{E} \left(\Phi \left(\frac{1}{t} \int_0^t X(s) ds, y^* \right) - \Phi \left(x^*, \frac{1}{t} \int_0^t Y(s) ds \right) \right) = \mathcal{O} \left(\frac{1}{\gamma(t)t} \right) \quad \text{as } t \rightarrow +\infty,$$

respectively.

Indeed, invoking Jensen's inequality and (20), we have for all $t > 0$

$$\begin{aligned} 0 &\leq \mathbb{E} \left(\Phi \left(\frac{1}{t} \int_0^t X(s) ds, y^* \right) - \Phi \left(x^*, \frac{1}{t} \int_0^t Y(s) ds \right) \right) \leq \mathbb{E} \left(\frac{1}{t} \int_0^t (\Phi(X(s), y^*) - \Phi(x^*, Y(s))) ds \right) \\ &\leq \mathbb{E} \left(\frac{1}{t} \int_0^t \langle (\nabla_x \Phi(X(s), Y(s)), -\nabla_y \Phi(X(s), Y(s))), (X(s), Y(s)) - (x^*, y^*) \rangle ds \right). \end{aligned}$$

4 Discrete time considerations

The aim of this section is to show how the ergodic upper bound results in expectation for the squared norm of the operator and for the gap function transfer to the stochastic variants of the Optimistic Gradient Descent Ascent (OGDA) method and the Extragradient (EG) method with constant step sizes, which can be interpreted as temporal discretizations of (SDE-M).

We will consider only equations governed by operators which are monotone, but not necessarily strongly monotone. For a stochastic variant of the Optimistic Gradient Descent Ascent (OGDA) method designed to solve variational inequalities, ergodic upper bound results for the gap function in expectation have also been derived in [6]. For the stochastic Extragradient (EG) method, ergodic upper bound results in expectation for the squared norm of the operator have also been derived in [16] (see also [8, 20]), and for the gap function [6]. As seen in Remark 3.3, the upper bound results for the gap function can be straightforwardly transferred via Jensen's inequality and the gradient inequality for convex functions to the ergodic primal-dual gap when solving convex-concave minimax problems.

Throughout this section, we will assume that only a stochastic estimator $M(\cdot, \xi)$ is accessible instead of M itself. The stochastic estimator is assumed to be unbiased, meaning that $\mathbb{E}_\xi(M(x, \xi)) = M(x)$ for all $x \in \mathbb{R}^n$. Further, it is required to have bounded variance, i.e., $\mathbb{E}_\xi(\|M(x, \xi) - M(x)\|^2) < \sigma_*^2$ for all $x \in \mathbb{R}^n$.

4.1 Stochastic Optimistic Gradient Descent Ascent (OGDA) method

Recalling the stochastic differential equation (SDE-M)

$$d(X(t) + \mu(t)M(X(t))) = -(\gamma(t) - \dot{\mu}(t))M(X(t))dt + \sigma(t, X(t))dW(t),$$

and fixing the parameter functions $\mu(t) = \gamma(t) := \gamma > 0$ to be constant for all $t \geq 0$, we consider the temporal discretization

$$x^{k+1} - x^k + \gamma M(x^k, \xi_k) - \gamma M(x^{k-1}, \xi_{k-1}) = -\gamma M(x^k, \xi_k) \quad \forall k \geq 1.$$

Equivalently,

$$x^{k+1} := x^k - 2\gamma M(x^k, \xi_k) + \gamma M(x^{k-1}, \xi_{k-1}) \quad \forall k \geq 1, \quad (\text{Stochastic OGDA})$$

where $x^0, x^1 \in \mathbb{R}^n$ are given vectors and ξ_k encodes the stochasticity appearing at iteration k .

We consider the approximation errors

$$W_k := M(x^k, \xi_k) - M(x^k) \quad \forall k \geq 0,$$

and the sub-sigma algebras

$$\mathcal{F}_k := \sigma(x^0, x^1, \xi_0, \xi_1, \dots, \xi_{k-1}) \quad \forall k \geq 2.$$

Lemma 4.1. *Let $x^* \in \text{zer}M$ and $\gamma < \frac{1}{2L}$. For all $k \geq 2$ it holds*

$$\begin{aligned}
\|x^{k+1} + \gamma M(x^k, \xi_k) - x^*\| &\leq \|x^k + \gamma M(x^{k-1}, \xi_{k-1}) - x^*\|^2 \\
&\quad - 2\gamma \langle M(x^k, \xi_k), x^k - x^* \rangle + \frac{\gamma^2(16\gamma^2 L^2 - 1)}{2(1 - 4\gamma^2 L^2)} \|M(x^{k-1})\|^2 \\
&\quad + \frac{12\gamma^4 L^4}{1 - 4\gamma^2 L^2} (\|x^{k-1} - x^{k-2}\|^2 - \|x^k - x^{k-1}\|^2) \\
&\quad + 3\gamma^2 \|W_k\|^2 + \frac{\gamma^2(16\gamma^2 L^2 + 5)}{2(1 - 4\gamma^2 L^2)} \|W_{k-1}\|^2 + \frac{12\gamma^4 L^2}{1 - 4\gamma^2 L^2} \|W_{k-2}\|^2.
\end{aligned}$$

Proof. Let $k \geq 2$ be fixed. We observe that

$$\begin{aligned}
&\|x^{k+1} + \gamma M(x^k, \xi_k) - x^*\|^2 \\
&= \|x^k - \gamma M(x^k, \xi_k) + \gamma M(x^{k-1}, \xi_{k-1}) - x^*\|^2 \\
&= \|x^k - \gamma M(x^k, \xi_k) + 2\gamma M(x^{k-1}, \xi_{k-1}) - x^*\|^2 - \gamma^2 \|M(x^{k-1}, \xi_{k-1})\|^2 \\
&\quad - 2\gamma \langle x^k - \gamma M(x^k, \xi_k) + \gamma M(x^{k-1}, \xi_{k-1}) - x^*, M(x^{k-1}, \xi_{k-1}) \rangle \\
&= \|x^k + \gamma M(x^{k-1}, \xi_{k-1}) - x^*\|^2 + \|\gamma M(x^{k-1}, \xi_{k-1}) - \gamma M(x^k, \xi_k)\|^2 - \gamma^2 \|M(x^{k-1}, \xi_{k-1})\|^2 \\
&\quad + 2\gamma \langle x^k + \gamma M(x^{k-1}, \xi_{k-1}) - x^*, M(x^{k-1}, \xi_{k-1}) - M(x^k, \xi_k) \rangle \\
&\quad - 2\gamma \langle x^k - \gamma M(x^k, \xi_k) + \gamma M(x^{k-1}, \xi_{k-1}) - x^*, M(x^{k-1}, \xi_{k-1}) \rangle \\
&= \|x^k + \gamma M(x^{k-1}, \xi_{k-1}) - x^*\|^2 + \gamma^2 (\|M(x^{k-1}, \xi_{k-1}) - M(x^k, \xi_k)\|^2 - \|M(x^{k-1}, \xi_{k-1})\|^2) \\
&\quad - 2\langle \gamma M(x^k, \xi_k), x^k - x^* \rangle \\
&\leq \|x^k + \gamma M(x^{k-1}, \xi_{k-1}) - x^*\|^2 - 2\gamma \langle M(x^k, \xi_k), x^k - x^* \rangle - \gamma^2 \|M(x^{k-1}, \xi_{k-1})\|^2 \\
&\quad + \gamma^2 (\|W_{k-1} + M(x^{k-1}) - M(x^k)\| + \|W_k\|)^2 \\
&\leq \|x^k + \gamma M(x^{k-1}, \xi_{k-1}) - x^*\|^2 - 2\gamma \langle M(x^k, \xi_k), x^k - x^* \rangle - \gamma^2 \|M(x^{k-1}, \xi_{k-1})\|^2 \\
&\quad + 3\gamma^2 (\|W_{k-1}\|^2 + \|M(x^{k-1}) - M(x^k)\|^2 + \|W_k\|^2). \\
&\leq \|x^k + \gamma M(x^{k-1}, \xi_{k-1}) - x^*\|^2 - 2\gamma \langle M(x^k, \xi_k), x^k - x^* \rangle - \gamma^2 \|M(x^{k-1}, \xi_{k-1})\|^2 \\
&\quad + 3\gamma^2 (\|W_{k-1}\|^2 + L^2 \|x^{k-1} - x^k\|^2 + \|W_k\|^2). \tag{21}
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\|x^k - x^{k-1}\|^2 &= \gamma^2 \| -2M(x^{k-1}, \xi_{k-1}) + M(x^{k-2}, \xi_{k-2}) \|^2 \\
&= \gamma^2 \| -M(x^{k-1}, \xi_{k-1}) + W_{k-1} + M(x^{k-2}) - M(x^{k-1}) + W_{k-2} \|^2 \\
&\leq 4\gamma^2 (\|M(x^{k-1}, \xi_{k-1})\|^2 + \|W_{k-1}\|^2 + L^2 \|x^{k-1} - x^{k-2}\|^2 + \|W_{k-2}\|^2),
\end{aligned}$$

therefore,

$$\begin{aligned}
(1 - 4\gamma^2 L^2) \|x^k - x^{k-1}\|^2 &\leq 4\gamma^2 (\|M(x^{k-1}, \xi_{k-1})\|^2 + \|W_{k-1}\|^2 + \|W_{k-2}\|^2) \\
&\quad + 4\gamma^2 L^2 (\|x^{k-1} - x^{k-2}\|^2 - \|x^k - x^{k-1}\|^2). \tag{22}
\end{aligned}$$

Plugging (22) into (21), we obtain

$$\begin{aligned}
& \|x^{k+1} + \gamma M(x^k, \xi_k) - x^*\|^2 \\
& \leq \|x^k + \gamma M(x^{k-1}, \xi_{k-1}) - x^*\|^2 - 2\gamma \langle M(x^k, \xi_k), x^k - x^* \rangle - \gamma^2 \|M(x^{k-1}, \xi_{k-1})\|^2 \\
& \quad + \frac{12\gamma^4 L^2}{1-4\gamma^2 L^2} \|M(x^{k-1}, \xi_{k-1})\|^2 + 3\gamma^2 \|W_k\|^2 + \left(3\gamma^2 + \frac{12\gamma^4 L^2}{1-4\gamma^2 L^2}\right) \|W_{k-1}\|^2 + \frac{12\gamma^4 L^2}{1-4\gamma^2 L^2} \|W_{k-2}\|^2 \\
& \quad + \frac{12\gamma^4 L^4}{1-4\gamma^2 L^2} (\|x^{k-1} - x^{k-2}\|^2 - \|x^k - x^{k-1}\|^2) \\
& = \|x^k + \gamma M(x^{k-1}, \xi_{k-1}) - x^*\|^2 - 2\gamma \langle M(x^k, \xi_k), x^k - x^* \rangle + \frac{16\gamma^4 L^2 - \gamma^2}{1-4\gamma^2 L^2} \|M(x^{k-1}, \xi_{k-1})\|^2 \\
& \quad + 3\gamma^2 \|W_k\|^2 + \left(3\gamma^2 + \frac{12\gamma^4 L^2}{1-4\gamma^2 L^2}\right) \|W_{k-1}\|^2 + \frac{12\gamma^4 L^2}{1-4\gamma^2 L^2} \|W_{k-2}\|^2 \\
& \quad + \frac{12\gamma^4 L^4}{1-4\gamma^2 L^2} (\|x^{k-1} - x^{k-2}\|^2 - \|x^k - x^{k-1}\|^2) \\
& \leq \|x^k + \gamma M(x^{k-1}, \xi_{k-1}) - x^*\|^2 - 2\gamma \langle M(x^k, \xi_k), x^k - x^* \rangle + \frac{\gamma^2(16\gamma^2 L^2 - 1)}{2(1-4\gamma^2 L^2)} \|M(x^{k-1})\|^2 \\
& \quad + 3\gamma^2 \|W_k\|^2 + \frac{\gamma^2(16\gamma^2 L^2 + 5)}{2(1-4\gamma^2 L^2)} \|W_{k-1}\|^2 + \frac{12\gamma^4 L^2}{1-4\gamma^2 L^2} \|W_{k-2}\|^2 \\
& \quad + \frac{12\gamma^4 L^4}{1-4\gamma^2 L^2} (\|x^{k-1} - x^{k-2}\|^2 - \|x^k - x^{k-1}\|^2),
\end{aligned}$$

as claimed. \square

The following theorem provides ergodic upper bounds in expectation for the squared norm of the operator and the gap function for (Stochastic OGD).

Theorem 4.2. *Let $x^* \in \text{zer}M$ and $\gamma < \frac{1}{4L}$. For all $K \geq 0$ it holds*

$$\begin{aligned}
& \min_{k=1, \dots, K+1} \mathbb{E} (\|M(x^k)\|^2) \leq \mathbb{E} \left(\frac{1}{K+1} \sum_{k=1}^{K+1} \|M(x^k)\|^2 \right) \\
& \leq \frac{1}{K+1} \frac{2(1-4\gamma^2 L^2)}{\gamma^2(1-16\gamma^2 L^2)} \left(4\|x^1 - x^*\|^2 + \frac{4\gamma^2 L^2(1-\gamma^2 L^2)}{1-4\gamma^2 L^2} \|x^1 - x^0\|^2 + 8\gamma^2 \sigma_*^2 \right) + \frac{16\gamma^2 L^2 + 11}{1-16\gamma^2 L^2} \sigma_*^2
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E} \left(\frac{1}{K+1} \sum_{k=2}^{K+2} \langle M(x^k), x^k - x^* \rangle \right) & \leq \frac{1}{K+1} \left(\frac{2\|x^1 - x^*\|^2}{\gamma} + \frac{2\gamma L^2(1-\gamma^2 L^2)}{1-4\gamma^2 L^2} \|x^1 - x^0\|^2 + 4\gamma \sigma_*^2 \right) \\
& \quad + \frac{\gamma(16\gamma^2 L^2 + 11)}{4(1-4\gamma^2 L^2)} \sigma_*^2.
\end{aligned}$$

Proof. Let $K \geq 0$. According to Lemma 4.1, we have for all $k \geq 2$

$$\begin{aligned}
& \frac{\gamma^2(1-16\gamma^2 L^2)}{2(1-4\gamma^2 L^2)} \|M(x^{k-1})\|^2 + 2\gamma \langle M(x^k, \xi_k), x^k - x^* \rangle \\
& \leq \|x^k + \gamma M(x^{k-1}, \xi_{k-1}) - x^*\|^2 - \|x^{k+1} + \gamma M(x^k, \xi_k) - x^*\|^2 \\
& \quad + \frac{12\gamma^4 L^4}{1-4\gamma^2 L^2} (\|x^{k-1} - x^{k-2}\|^2 - \|x^k - x^{k-1}\|^2) \\
& \quad + \frac{12\gamma^4 L^2}{1-4\gamma^2 L^2} \|W_{k-2}\|^2 + \frac{\gamma^2(16\gamma^2 L^2 + 5)}{2(1-4\gamma^2 L^2)} \|W_{k-1}\|^2 + 3\gamma^2 \|W_k\|^2.
\end{aligned}$$

Summing the above inequality for k from 2 to $K + 2$ and multiplying by $\frac{1}{K+1}$ yields

$$\begin{aligned}
& \frac{\gamma^2(1-16\gamma^2L^2)}{2(1-4\gamma^2L^2)} \frac{1}{K+1} \sum_{k=1}^{K+1} \|M(x^k)\|^2 + \frac{2\gamma}{K+1} \sum_{k=2}^{K+2} \langle M(x^k, \xi_k), x^k - x^* \rangle \\
& \leq \frac{1}{K+1} (\|x^2 + \gamma M(x^1, \xi_1) - x^*\|^2 - \|x^{K+3} + \gamma M(x^{K+2}, \xi_{K+2}) - x^*\|^2) \\
& \quad + \frac{1}{K+1} \frac{12\gamma^4L^4}{1-4\gamma^2L^2} (\|x^1 - x^0\|^2 - \|x^{K+2} - x^{K+1}\|^2) \\
& \quad + \frac{1}{K+1} \sum_{k=2}^{K+2} \left(\frac{12\gamma^4L^2}{1-4\gamma^2L^2} \|W_{k-2}\|^2 + \frac{\gamma^2(16\gamma^2L^2+5)}{2(1-4\gamma^2L^2)} \|W_{k-1}\|^2 + 3\gamma^2 \|W_k\|^2 \right) \\
& \leq \frac{1}{K+1} \left(\|x^2 + \gamma M(x^1, \xi_1) - x^*\|^2 + \frac{12\gamma^4L^4}{1-4\gamma^2L^2} \|x^1 - x^0\|^2 \right) \\
& \quad + \frac{1}{K+1} \sum_{k=2}^{K+2} \left(\frac{12\gamma^4L^2}{1-4\gamma^2L^2} \|W_{k-2}\|^2 + \frac{\gamma^2(16\gamma^2L^2+5)}{2(1-4\gamma^2L^2)} \|W_{k-1}\|^2 + 3\gamma^2 \|W_k\|^2 \right).
\end{aligned}$$

For all $k = 2, \dots, K + 2$ we have that

$$\mathbb{E}(\langle M(x^k, \xi_k), x^k - x^* \rangle) = \mathbb{E}(\mathbb{E}(\langle M(x^k, \xi_k), x^k - x^* \rangle | \mathcal{F}_k)) = \mathbb{E}(\langle \mathbb{E}(M(x^k, \xi_k) | \mathcal{F}_k), x^k - x^* \rangle) = \langle M(x^k), x^k - x^* \rangle$$

and

$$\mathbb{E}(\|W_k\|^2) = \mathbb{E}(\mathbb{E}(\|W_k\|^2 | \mathcal{F}_k)) \leq \sigma_*^2.$$

This yields

$$\begin{aligned}
& \frac{\gamma^2(1-16\gamma^2L^2)}{2(1-4\gamma^2L^2)} \mathbb{E} \left(\frac{1}{K+1} \sum_{k=1}^{K+1} \|M(x^k)\|^2 \right) + 2\gamma \mathbb{E} \left(\frac{1}{K+1} \sum_{k=2}^{K+2} \langle M(x^k), x^k - x^* \rangle \right) \\
& \leq \frac{1}{K+1} \left(\mathbb{E}(\|x^2 + \gamma M(x^1, \xi_1) - x^*\|^2) + \frac{12\gamma^4L^4}{1-4\gamma^2L^2} \|x^1 - x^0\|^2 \right) \\
& \quad + \frac{1}{K+1} \sum_{k=2}^{K+2} \mathbb{E} \left(\frac{12\gamma^4L^2}{1-4\gamma^2L^2} \|W_{k-2}\|^2 + \frac{\gamma^2(16\gamma^2L^2+5)}{2(1-4\gamma^2L^2)} \|W_{k-1}\|^2 + 3\gamma^2 \|W_k\|^2 \right) \\
& \leq \frac{1}{K+1} \left(4\|x^1 - x^*\|^2 + 4\gamma^2L^2\|x^1 - x^0\|^2 + 8\gamma^2\sigma_*^2 + \frac{12\gamma^4L^4}{1-4\gamma^2L^2} \|x^1 - x^0\|^2 \right) + \frac{\gamma^2(16\gamma^2L^2+11)}{2(1-4\gamma^2L^2)} \sigma_*^2 \\
& = \frac{1}{K+1} \left(4\|x^1 - x^*\|^2 + \frac{4\gamma^2L^2(1-\gamma^2L^2)}{1-4\gamma^2L^2} \|x^1 - x^0\|^2 + 8\gamma^2\sigma_*^2 \right) + \frac{\gamma^2(16\gamma^2L^2+11)}{2(1-4\gamma^2L^2)} \sigma_*^2.
\end{aligned}$$

This concludes the proof. \square

4.2 Stochastic Extragradient (EG) method

The stochastic Extragradient method (EG) reads

$$\begin{cases} y^k := x^k - \gamma M(x^k, \xi_k) \\ x^{k+1} := x^k - \gamma M(y^k, \eta_k) \end{cases} \quad \forall k \geq 0, \quad (\text{Stochastic EG})$$

where $x^0 \in \mathbb{R}^n$ is a given vector and ξ_k and η_k encode the stochasticity appearing at iteration k when evaluating M at x^k and y^k , respectively.

We consider the approximation errors

$$W_k := M(x^k, \xi_k) - M(x^k) \quad \text{and} \quad Z_k := M(y^k, \eta_k) - M(y^k) \quad \forall k \geq 0$$

and the sub-sigma algebras

$$\mathcal{F}_0 := \sigma(x^0) \quad \text{and} \quad \mathcal{F}_k := \sigma(x^0, \xi_0, \xi_1, \dots, \xi_{k-1}, \eta_0, \dots, \eta_{k-1}) \quad \forall k \geq 1.$$

and

$$\widehat{\mathcal{F}}_0 := \sigma(x^0, \xi_0) \quad \text{and} \quad \widehat{\mathcal{F}}_k := \sigma(x^0, \xi_0, \xi_1, \dots, \xi_{k-1}, \xi_k, \eta_0, \dots, \eta_{k-1}) \quad \forall k \geq 1.$$

First, establish an inequality that will be essential to the proof of the convergence rate:

Lemma 4.3. *Let $x^* \in \text{zer}M$. For all $k \geq 0$ it holds*

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - 2\gamma \langle M(y^k, \eta_k), y^k - x^* \rangle + 2(3L^2\gamma^2 - 1)\|M(x^k)\|^2 \\ &\quad + \gamma^2(3\|Z_k\|^2 + (1 + 6L^2\gamma^2)\|W_k\|^2). \end{aligned}$$

Proof. Let $k \geq 0$. We have

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &= \|x^k - \gamma M(y^k, \eta_k) - x^*\|^2 \\ &= \|x^k - x^*\|^2 - 2\gamma \langle x^k - x^*, M(y^k, \eta_k) - M(x^*) \rangle + \gamma^2 \|M(y^k, \eta_k) - M(x^*)\|^2 \\ &= \|x^k - x^*\|^2 - 2\gamma \langle y^k + \gamma M(x^k, \xi_k) - \gamma M(x^*) - x^*, M(y^k, \eta_k) - M(x^*) \rangle \\ &\quad + \gamma^2 \|M(y^k, \eta_k) - M(x^*)\|^2 \\ &= \|x^k - x^*\|^2 - 2\gamma \langle y^k - x^*, M(y^k, \eta_k) - M(x^*) \rangle \\ &\quad - \gamma^2 (2\langle M(x^k, \xi_k) - M(x^*), M(y^k, \eta_k) - M(x^*) \rangle - \|M(y^k, \eta_k) - M(x^*)\|^2) \\ &= \|x^k - x^*\|^2 - 2\gamma \langle y^k - x^*, M(y^k, \eta_k) - M(x^*) \rangle - \gamma^2 \|M(x^k, \xi_k)\|^2 \\ &\quad + \gamma^2 (\|M(y^k, \eta_k) - M(x^k, \xi_k)\|^2) \\ &\leq \|x^k - x^*\|^2 - 2\gamma \langle y^k - x^*, M(y^k, \eta_k) - M(x^*) \rangle - \gamma^2 \|M(x^k, \xi_k)\|^2 \\ &\quad + \gamma^2 (3\|M(y^k, \eta_k) - M(y^k)\|^2 + 3\|M(y^k) - M(x^k)\|^2 + 3\|M(x^k) - M(x^k, \xi_k)\|^2) \\ &\leq \|x^k - x^*\|^2 - 2\gamma \langle y^k - x^*, M(y^k, \eta_k) - M(x^*) \rangle + \gamma^2 (3L^2 - 1)\|M(x^k, \xi_k)\|^2 \\ &\quad + \gamma^2 (3\|Z_k\|^2 + 3\|W_k\|^2) \\ &\leq \|x^k - x^*\|^2 - 2\gamma \langle y^k - x^*, M(y^k, \eta_k) - M(x^*) \rangle + 2(3L^2\gamma^2 - 1)\|M(x^k)\|^2 \\ &\quad + \gamma^2 (3\|Z_k\|^2 + (1 + 6L^2\gamma^2)\|W_k\|^2), \end{aligned}$$

as claimed. \square

The following theorem provides ergodic upper bounds in expectation for the squared norm of the operator in terms of the sequence $(x^k)_{k \geq 0}$ and the gap function in terms of the sequence $(y^k)_{k \geq 0}$.

Theorem 4.4. *Let $x^* \in \text{zer}M$ and $\gamma < \frac{1}{\sqrt{3}L}$. For all $K \geq 0$ it holds*

$$\min_{k=0, \dots, K} \mathbb{E} (\|M(x^k)\|^2) \leq \mathbb{E} \left(\frac{1}{K+1} \sum_{k=0}^K \|M(x^k)\|^2 \right) \leq \frac{1}{K+1} \frac{1}{2(1-3L^2\gamma^2)} \|x^0 - x^*\|^2 + \frac{\gamma^2(2+3L^2\gamma^2)}{1-3L^2\gamma^2} \sigma_*^2$$

and

$$\mathbb{E} \left(\frac{1}{K+1} \sum_{k=0}^K \langle M(y^k), y^k - x^* \rangle \right) \leq \frac{1}{K+1} \frac{1}{2\gamma} \|x^0 - x^*\|^2 + \gamma(2+3L^2\gamma^2)\sigma_*^2.$$

Proof. Let $K \geq 0$. According to Lemma 4.3, we have for all $k \geq 0$

$$\begin{aligned} 2(1-3L^2\gamma^2)\|M(x^k)\|^2 + 2\gamma \langle M(y^k, \eta_k), y^k - x^* \rangle &\leq \|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 \\ &\quad + \gamma^2(3\|Z_k\|^2 + (1+6L^2\gamma^2)\|W_k\|^2). \end{aligned}$$

Summing the above inequality for k from 0 to K and multiplying by $\frac{1}{K+1}$ it yields

$$\begin{aligned} & \frac{2(1-3L^2\gamma^2)}{K+1} \sum_{k=0}^K \|M(x^k)\|^2 + \frac{2\gamma}{K+1} \sum_{k=0}^K \langle M(y^k, \eta_k), y^k - x^* \rangle \\ & \leq \frac{1}{K+1} (\|x^0 - x^*\|^2 - \|x^{K+1} - x^*\|^2) + \frac{\gamma^2}{K+1} \left(\sum_{k=0}^K 3\|Z_k\|^2 + (1+6L^2\gamma^2)\|W_k\|^2 \right) \\ & \leq \frac{1}{K+1} \|x^0 - x^*\|^2 + \frac{\gamma^2}{K+1} \left(\sum_{k=0}^K 3\|Z_k\|^2 + (1+6L^2\gamma^2)\|W_k\|^2 \right). \end{aligned}$$

For all $k = 0, \dots, K$ we have that

$$\mathbb{E}(\langle M(y^k, \eta_k), y^k - x^* \rangle) = \mathbb{E}(\mathbb{E}(\langle M(y^k, \eta_k), y^k - x^* \rangle | \widehat{\mathcal{F}}_k)) = \mathbb{E}(\langle \mathbb{E}(M(y^k, \eta_k) | \widehat{\mathcal{F}}_k), y^k - x^* \rangle) = \langle M(y^k), y^k - x^* \rangle,$$

and

$$\mathbb{E}(\|W_k\|^2) = \mathbb{E}(\mathbb{E}(\|W_k\|^2 | \mathcal{F}_k)) \leq \sigma_*^2 \quad \text{and} \quad \mathbb{E}(\|Z_k\|^2) = \mathbb{E}(\mathbb{E}(\|Z_k\|^2 | \widehat{\mathcal{F}}_k)) \leq \sigma_*^2.$$

This yields

$$\begin{aligned} & 2(1-3L^2\gamma^2)\mathbb{E} \left(\frac{1}{K+1} \sum_{k=0}^K \|M(x^k)\|^2 \right) + 2\gamma\mathbb{E} \left(\frac{1}{K+1} \sum_{k=0}^K \langle M(y^k), y^k - x^* \rangle \right) \\ & \leq \frac{1}{K+1} \|x^0 - x^*\|^2 + 2\gamma^2(2+3L^2\gamma^2)\sigma_*^2. \end{aligned}$$

This concludes the proof. \square

4.3 Numerical experiments

In order to illustrate the convergence behaviour of the two stochastic algorithms, we consider the monotone equation associated to the following minimax problem (see also [7, 30])

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^n} \Phi(x, y),$$

where

$$\Phi(x, y) := \frac{1}{2} \langle x, Hx \rangle - \langle x, h \rangle - \langle y, Ax - b \rangle,$$

with

$$A := \frac{1}{4} \begin{pmatrix} & & & -1 & 1 \\ & & \ddots & \ddots & \\ & -1 & 1 & & \\ -1 & 1 & & & \\ 1 & & & & \end{pmatrix} \in \mathbb{R}^{n \times n}, \quad H := 2A^\top A, \quad b := \frac{1}{4} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix} \in \mathbb{R}^n \quad \text{and} \quad h := \frac{1}{4} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^n.$$

For $n = 10$, we performed the stochastic OGD and EG methods with 50000 iterations for a total of 100 times. In Figure (1) we plot the averaged squared norm of the operator

$$M(x, y) = (\nabla_x \Phi(x, y), -\nabla_y \Phi(x, y))$$

for the stochastic OGD and EG methods in loglog scale. The shaded areas represent the range between the best and worst-case instances, while the strongly colored lines depict the average over all 100 runs of

the algorithms. The error term in the k -th iteration was chosen as $\frac{a}{k\sqrt{n}}(1, \dots, 1)^T$, where a is normally distributed with mean 0 and standard deviation 10. While there does not seem to be much difference in performance between the stochastic OGDA and EG methods, at least in the example considered for the numerical experiments, the latter supports the $\mathcal{O}(\frac{1}{k})$ convergence rates as being the fastest one can expect from these methods.

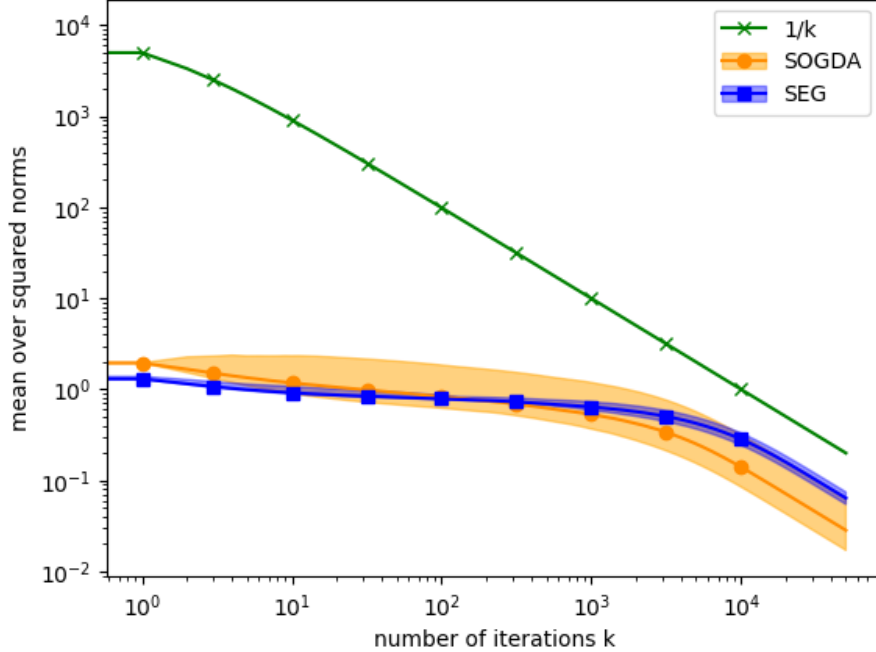


Figure 1: The convergence behaviour of the averaged squared norm of the operator for the stochastic OGDA and EG methods

Acknowledgements. The authors thank the anonymous reviewer for comments and suggestions that improved the quality of the paper.

A Appendix

In the appendix, we present several auxiliary results utilized in the analysis conducted in this paper. The following theorem will play a crucial role in establishing the existence and uniqueness of a solution to (SDE-M).

Theorem A.1. ([25, Theorem A.7], [28, Theorem 5.2.1]) *Let $F : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $G : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ be measurable functions satisfying, for every $T > 0$*

$$\|F(t, x) - F(t, y)\| + \|G(t, x) - G(t, y)\|_F \leq C_1 \|x - y\|, \quad \forall x, y \in \mathbb{R}^n \quad \forall t \in [0, T],$$

for some constant $C_1 \geq 0$. Then the stochastic differential equation

$$\begin{cases} dX(t) = F(t, X(t))dt + G(t, X(t))dW(t) & \forall t \in [0, T], \\ X(0) = X_0, \end{cases} \quad (\text{SDE-gen})$$

where W is an \mathcal{F}_t -adapted m -dimensional Brownian motion, has a unique solution $X \in S_n^\nu[0, T]$, for every $\nu \geq 2$.

Next, we introduce Itô's formula (see [13]).

Proposition A.2. *Let $X \in S_n^\nu$ be the solution of (SDE-gen)*

$$\begin{cases} dX(t) = F(t, X(t))dt + G(t, X(t))dW(t) & \forall t \geq 0, \\ X(0) = X_0, \end{cases}$$

where W is an \mathcal{F}_t -adapted m -dimensional Brownian motion, and, for all $T > 0$, $F = (F_1, \dots, F_n) : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ satisfies

$$\mathbb{E} \left(\int_0^T |F_i(t, X(t))| dt \right) < +\infty,$$

and $G = (G_{ij})_{i=\{1, \dots, n\}, j=\{1, \dots, m\}} : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ satisfies

$$\mathbb{E} \left(\int_0^T |G_{ij}(t, X(t))|^2 dt \right) < +\infty.$$

Further, let $\phi : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}$ be such that $\phi(\cdot, x) \in C^1(\mathbb{R}_+)$ for all $x \in \mathbb{R}^n$ and $\phi(t, \cdot) \in C^2(\mathbb{R}^n)$ for all $t \geq 0$. Then the process

$$\tilde{X}(t) = \phi(t, X(t))$$

is an Itô process such that

$$\begin{aligned} d\tilde{X}(t) &= \frac{d}{dt} \phi(t, X(t))dt + \sum_{i=1}^n \frac{d}{dx_i} \phi(t, X(t))dX^i(t) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^n \frac{d^2}{dx_i dx_j} \phi(t, X(t)) \sum_{\ell=1}^m G_{i\ell}(t, X(t))G_{j\ell}(t, X(t))dt \quad \forall t \geq 0. \end{aligned}$$

Further, if, for all $T > 0$,

$$\mathbb{E} \left(\int_0^T \|\sigma^\top(s, (X(s), Y(s))) \nabla_x \phi(s, X(s))\|^2 ds \right) < +\infty,$$

then $\int_0^t \langle \sigma^\top(s, (X(s), Y(s))) \nabla_x \phi(s, (X(s), Y(s))), dW(s) \rangle$ is for all $t \geq 0$ a square-integrable continuous martingale with expected value 0.

The proof of the convergence results in Theorem 2.4 relies on the following result.

Theorem A.3. ([24, Theorem 3.9], Theorem 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 $\{N(t)\}_{t \geq 0}$ be a real-valued continuous local martingale with $N(0) = 0$ a.s. Let ξ be a nonnegative \mathcal{F}_0 -measurable random variable. Define*

$$X(t) := \xi + A(t) - U(t) + N(t) \quad \text{for } \forall t \geq 0.$$

If $X(t)$ is nonnegative and $\lim_{t \rightarrow +\infty} A(t) < +\infty$ a.s., then a.s. $\lim_{t \rightarrow +\infty} X(t)$ exists and is finite, as well as $\lim_{t \rightarrow +\infty} U(t) < +\infty$.

In order to show that $\lim_{t \rightarrow +\infty} \|M(X(t))\| = 0$ a.s., we make use of the following lemma as well as of Doob's martingale convergence theorem presented below.

Lemma A.4. ([25, Lemma A.4]) Let $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ be such that $\liminf_{t \rightarrow +\infty} f(t) \neq \limsup_{t \rightarrow +\infty} f(t)$. Then, there exists a constant α such that $\liminf_{t \rightarrow +\infty} f(t) < \alpha < \limsup_{t \rightarrow +\infty} f(t)$, and for every $\beta > 0$ we can define a sequence $(t_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}$ with the properties that

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

Theorem A.5. ([12]) Let $\{M(t)\}_{t \geq 0} : \Omega \rightarrow \mathbb{R}$ be a continuous martingale such that $\sup_{t \geq 0} \mathbb{E}(|M(t)|^p) < +\infty$ for some $p > 1$. Then there exists a random variable M_∞ such that $\mathbb{E}(|M_\infty|^p) < +\infty$ and $\lim_{t \rightarrow +\infty} M(t) = M_\infty$ a.s.

The derivation of convergence rate in the strongly monotone case necessitates the following lemma.

Lemma A.6. ([25, Lemma A.2]) Let $t_0 \geq 0$ and $T > t_0$. Assume that $h : [t_0, +\infty) \rightarrow \mathbb{R}_+$ is measurable with $h \in L^1([t_0, T])$, that $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is continuous and nondecreasing, and the Cauchy problem

$$\begin{cases} \varphi'(t) = -\psi(\varphi(t)) + h(t) & \text{for almost all } t \in [t_0, T] \\ \varphi(t_0) = \varphi_0 > 0 \end{cases}$$

has an absolutely continuous solution $\varphi : [t_0, T] \rightarrow \mathbb{R}_+$. If a bounded lower semicontinuous function $\omega : [t_0, T] \rightarrow \mathbb{R}_+$ satisfies $\omega(t_0) = \varphi_0$ and

$$\omega(t_2) \leq \omega(t_1) - \int_{t_1}^{t_2} \psi(\omega(s)) ds + \int_{t_1}^{t_2} h(s) ds \quad \forall t_0 \leq t_1 < t_2 \leq T,$$

then

$$\omega(t) \leq \varphi(t) \quad \forall t \in [t_0, T].$$

References

- [1] A. S. Antipin. “On a method for convex programs using a symmetrical modification of the Lagrange function”. In: *Ekonomika i Matematicheskie Metody* 12 (1976), pp. 1164–1173.
- [2] H. Attouch, R. I. Boç, and D.-K. Nguyen. “Fast convex optimization via time scale and averaging of the steepest descent”. In: *Mathematics of Operations Research* (2024). DOI: 10.1287/moor.2023.0186.
- [3] H. Attouch and B. F. Svaiter. “A continuous dynamical Newton-like approach to solving monotone inclusions”. In: *SIAM Journal on Control and Optimization* 49(3) (2011), pp. 574–598.
- [4] F. Bach and E. Moulines. “Non-asymptotic analysis of stochastic approximation algorithms for machine learning”. In: *Advances in Neural Information Processing Systems 24 (NIPS 2011)* (2011).
- [5] J.-B. Baillon and H. Brézis. “Une remarque sur le comportement asymptotique des semi-groupes non linéaires”. In: *Houston Journal of Mathematics* 2 (1976), pp. 5–7.
- [6] A. Böhm, M. Sedlmayer, E. R. Csetnek, and R. I. Boç. “Two steps at a time – taking GAN training in stride with Tseng’s method”. In: *SIAM Journal on Mathematics of Data Science* 4(2) (2022), pp. 750–771.
- [7] R. I. Boç, E. R. Csetnek, and D.-K. Nguyen. “Fast Optimistic Gradient Descent Ascent (OGDA) method in continuous and discrete time”. In: *Foundations of Computational Mathematics* (2023). DOI: 10.1007/s10208-023-09636-5.
- [8] R. I. Boç, P. Mertikopoulos, M. Staudigl, and P. T. Vuong. “Minibatch Forward-Backward-Forward methods for solving Stochastic Variational inequalities”. In: *Stochastic Systems* 11(2) (2021), pp. 112–139.
- [9] P. Chaudhari and S. Sotito. “Stochastic gradient descent performs variational inference, converges to limit cycles for deep networks”. In: *Information Theory and Applications Workshop (ITA 2018)* (2018).

- [10] T. Chavdarova, M. I. Jordan, and M. Zampetakis. “Last-iterate convergence of saddle point optimizers via high-resolution differential equations”. In: *OPT2021: The 13th Annual Workshop on Optimization for Machine Learning* (2021).
- [11] R. Cominetti, J. Peypouquet, and S. Sorin. “Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization”. In: *Journal of Differential Equations* 245(12) (2008), pp. 3753–3763.
- [12] J. L. Doob. *Stochastic Processes*. Wiley, 1991.
- [13] L. C. Evans. *An Introduction to Stochastic Differential Equations*. Providence, Rhode Island: American Mathematical Society, 2013.
- [14] N. Golowich, S. Pattathil, and C. Daskalakis. “Tight last-iterate convergence rates for no-regret learning in multi-player games”. In: *34th Conference on Neural Information Processing Systems (NeurIPS 2020)* (2020).
- [15] N. Golowich, S. Pattathil, C. Daskalakis, and A. Ozdaglar. “Last iterate is slower than averaged iterate in smooth convex-concave saddle point problems”. In: *Conference on Learning Theory 2020* (2020), pp. 1758–1784.
- [16] E. Gorbunov, H. Berard, N. Loizou, and G. Gidel. “Stochastic extragradient: general analysis and improved rates”. In: *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS 2022)* (2022).
- [17] E. Gorbunov, N. Loizou, and G. Gidel. “Extragradient method: $\mathcal{O}(1/K)$ last-iterate convergence for monotone variational inequalities and connections with cocoercivity”. In: *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS 2022)* (2022).
- [18] Y.-G. Hsieh, F. Iutzeler, J. Malick, and P. Mertikopoulos. “On the convergence of single-call stochastic extra-gradient methods”. In: *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)* (2020).
- [19] W. Hu, C. J. Li, L. Li, and J.-G. Liu. “On the diffusion approximation of nonconvex stochastic gradient descent”. In: *Annals of Mathematical Sciences and Applications* 4(1) (2018), pp. 3–32.
- [20] A. N. Iusem, A. Jofré, R. I. Oliveira, and P. Thompson. “Extragradient method with variance reduction for stochastic variational inequalities”. In: *SIAM Journal on Optimization* 27(2) (2017), pp. 686–724.
- [21] G. M. Korpelevich. “An extragradient method for finding saddle points and for other problems”. In: *Ekonomika i Matematicheskie Metody* 12(4) (1976), pp. 747–756.
- [22] Q. Li, C. Tai, and W. E. “Stochastic modified equations and adaptive stochastic gradient algorithms”. In: *Proceedings of the 34th International Conference on Machine Learning* (2017).
- [23] Z. Li, S. Malladi, and S. Arora. “On the validity of modeling SGD with stochastic differential equations (SDEs)”. In: *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)* (2021).
- [24] X. Mao. *Stochastic Differential Equations and Applications*. Elsevier, 2007.
- [25] R. Maulen-Soto, J. Fadili, and H. Attouch. “An SDE perspective on stochastic convex optimization”. In: *arXiv: 2207.02750 [math.OA]* (2022).
- [26] R. Maulen-Soto, J. Fadili, and H. Attouch. “Tikhonov regularization for stochastic convex optimization in Hilbert spaces”. In: *arXiv: 2403.06708 [math.OA]* (2024).
- [27] R. Maulen-Soto, J. Fadili, H. Attouch, and P. Ochs. “Stochastic inertial dynamics via time scaling and averaging”. In: *arXiv: 2403.16775 [math.OA]* (2024).
- [28] B. Øksendal. *Stochastic Differential Equations*. Springer, 2003.
- [29] A. Orvieto and A. Lucchi. “Continuous-time models for stochastic optimization algorithms”. In: *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)* (2019).
- [30] Y. Ouyang and Y. Xu. “Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems”. In: *Mathematical Programming* 185 (2021), pp. 1–35.

- [31] L. D. Popov. “A modification of the Arrow–Hurwicz method for search of saddle points”. In: *Mathematical Notes of the Academy of Sciences of the USSR* 28 (5) (1980), pp. 845–848.
- [32] B. Shi, W. J. Su, and M. I. Jordan. “On learning rates and Schrödinger operators”. In: *Journal of Machine Learning Research* 24 (2023), pp. 1–53.