



HAL
open science

Stochastic monotone inclusion with closed loop distributions

Hamza Ennaji, Jalal M. Fadili, Hedy Attouch

► **To cite this version:**

Hamza Ennaji, Jalal M. Fadili, Hedy Attouch. Stochastic monotone inclusion with closed loop distributions. 2024. hal-04652333v2

HAL Id: hal-04652333

<https://hal.science/hal-04652333v2>

Preprint submitted on 13 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **STOCHASTIC MONOTONE INCLUSION WITH CLOSED LOOP DISTRIBUTIONS**

2 HAMZA ENNAJI*, JALAL FADILI^{c,‡}, AND HEDY ATTOUCH[◊]

Dedicated to the memory of Hedy Attouch, outstanding mathematician and beloved collaborator.

ABSTRACT. In this paper, we study in a Hilbertian setting, first and second-order monotone inclusions related to stochastic optimization problems with decision dependent distributions. The studied dynamics are formulated as monotone inclusions governed by Lipschitz perturbations of maximally monotone operators where the concept of equilibrium plays a central role. We discuss the relationship between the W_1 -Wasserstein Lipschitz behavior of the distribution and the so-called coarse Ricci curvature. As an application, we consider the monotone inclusions associated with stochastic optimisation problems involving the sum of a smooth function with Lipschitz gradient, a proximable function and a composite term.

3 CONTENTS

4	1. Introduction	1
5	2. Notation and preliminaries	4
6	3. First-order monotone inclusions	6
7	4. Second-order monotone inclusions via Hessian damping	11
8	5. On coarse Ricci curvature	17
9	6. Application: Inertial primal-dual algorithm	20
10	7. Comments, extensions and future work	25
11	Acknowledgement	26
12	Appendices	26
13	Appendix A. Gronwall inequalities	26
14	Appendix B. Banach Fixed point theorem & Picard iterative method	27
15	References	27

16 1. INTRODUCTION

17 Recently, many problems in machine learning and risk management come in form of
18 stochastic optimisation problems. Such problems aim to learn a decision rule from a data

* UNIV. GRENOBLE ALPES, CNRS, GRENOBLE INP*, LJK, 38000 GRENOBLE, FRANCE.

^c CORRESPONDING AUTHOR.

[‡] ENSICAEN, NORMANDIE UNIVERSITÉ, CNRS, GREYC, FRANCE.

[◊] IMAG, UNIVERSITÉ MONTPELLIER, CNRS, FRANCE.

E-mail addresses: hamza.ennaji@univ-grenoble-alpes.fr,
hedy.attouch@umontpellier.fr.

jalal.fadili@ensicaen.fr,

2020 *Mathematics Subject Classification.* Primary 34G25, 37N40, 46N10, 47H05, 49M30, 60J20.

Key words and phrases. Monotone inclusions; Dynamical systems; Gradient flows; Inertial dynamics; Ollivier-Ricci curvature.

1 sample. This can be formulated in terms of optimization problems of the form

$$\min_x \mathbb{E}_{\xi \sim \mathfrak{m}} f(x, \xi) + g(x), \quad (1)$$

2 where \mathfrak{m} is a probability measure, $f(x, \xi)$ is a loss function of the decision x at data point ξ
 3 and g is a convex regularizer. In this work, we are interested in (1) in the case where the
 4 distribution \mathfrak{m} depends itself on the decision x , i.e., problems of the form

$$\min_x \mathbb{E}_{\xi \sim \mathfrak{m}_x} f(x, \xi) + g(x), \quad (2)$$

5 In this case, one tries to learn a decision rule from a decision-dependent data distribution
 6 \mathfrak{m}_x . Problems of the form (2) were addressed in the framework of performative prediction
 7 proposed in [43, 47] and discussed with further algorithmic aspects in [35]. A typical example
 8 concerns prediction of loan default risks, that is the chance that a borrower won't be able
 9 to repay their loan. More precisely, banks take into account several parameters, including
 10 the default risk, to decide whether to accept a consumer's loan application and, if so, what
 11 interest rate will apply. It is clear that a high default risk implies a high interest rate, but a
 12 high interest rate increases the consumer's default risk. Thus, the predictive performance
 13 of the bank's model is not calibrated with respect to future results obtained by acting on
 14 the model. Another example concerns navigation apps, such as Google Maps (see, e.g.,
 15 [33, 41]), which suggest routes with low travel time to users. This influences users' decisions
 16 to pick such routes and consequently, increases traffic on these routes, impacting travel times.
 17 Further applications and illustrations can be found in [47, Appendix A].

18 In general, problems of the form (2) are difficult to solve. However, a natural approach
 19 consists in performing a repeated minimization procedure, i.e., throughout iterations, one
 20 solves

$$x_{t+1} = \arg \min_x \mathbb{E}_{\xi \sim \mathfrak{m}_{x_t}} f(x, \xi) + g(x), \quad (3)$$

21 and then updates the distribution $\mathfrak{m}_{x_{t+1}}$. Under suitable assumptions that will be specified
 22 later, the sequence $(x_t)_t$ generated by the repeated minimization procedure (3) admits a
 23 fixed point \bar{x} . Such a point turns out to be an equilibrium with respect to the distribution
 24 $\mathfrak{m}_{(\cdot)}$ in the following sense:

$$\bar{x} = \arg \min_x \mathbb{E}_{\xi \sim \mathfrak{m}_{\bar{x}}} f(x, \xi) + g(x), \quad (4)$$

25 that is, \bar{x} solves (2) for the induced distribution $\mathfrak{m}_{\bar{x}}$. So instead of solving (2), we look at an
 26 equilibrium point in the sense of (4). Notice that in terms of operators, (4) can be written
 27 (formally, for instance) as a monotone inclusion:

$$A(\bar{x}) + G(\bar{x}) \ni 0 \quad (5)$$

28 with $A(x) = \partial g(x)$ and $G(x) = \mathbb{E}_{\xi \sim \mathfrak{m}_{\bar{x}}} \nabla_x f(x, \xi)$. That is \bar{x} is a zero of the sum of two
 29 monotone operators. One strategy to solve problems of the form (5) is to consider some
 30 continuous and discrete dynamical systems whose trajectories may converge, under suitable
 31 assumptions, to an element in $(A + B)^{-1}(0)$, the zero set of the sum $A + B$. For instance,
 32 when $B \equiv 0$ and $A = \nabla h$ where h is a given convex lower semicontinuous function on \mathbb{R}^n , it
 33 is well known since the works of Brézis, Baillon and Bucker [21, 29], that each trajectory of
 34 the gradient flow

$$\dot{x}(t) = -\nabla h(x) \text{ with } x(0) = x_0 \in \mathbb{R}^n, \quad (6)$$

35 converges to a minimizer of h , and thus a zero of ∇h , provided $\arg \min h \neq \emptyset$.

36 Designing algorithms and dynamical systems with rapid convergence properties to solve
 37 monotone inclusions is at the core of many fields in modern optimization, partial differential
 38 equation, game theory, etc. The literature is extensive, and to name only a few, the reader

1 is referred to [1, 2, 7, 9, 17, 25, 38, 40, 52] and the references therein.

2

3 In this work, we address monotone inclusions of the form

$$\dot{x}(t) + A(x(t)) + \mathbb{E}_{\xi \sim \mathbf{m}_{x(t)}}(B(x(t), \xi)) \ni 0, \quad (7)$$

4 where A is a maximally monotone operator, and B is a single valued mapping. The particu-
5 larity of such a dynamic is, of course, the presence of the random operator $\mathbb{E}_{\xi \sim \mathbf{m}_{(\cdot)}}(B(\cdot, \xi))$
6 where the random variable ξ has a trajectory-dependent distribution $\mathbf{m}_{x(t)}$. Thus, it is not
7 straightforward how to address (7) within the classical framework (see, e.g., [22, 27]). Yet,
8 a clever reformulation of (7), based on the notion of equilibrium, as a monotone inclusion
9 governed by a Lipschitz perturbation of a maximally monotone operator will allow us to
10 tackle this issue. Then we investigate inertial dynamics related to (7). Indeed, since the
11 work of Polyak [48], who considered a system of the form

$$\ddot{x}(t) + \gamma \dot{x}(t) + \nabla f(x(t)) = 0, \quad (\text{HBF})$$

12 where $\gamma > 0$ is called the viscous damping coefficient, the introduction of inertial dynamics to
13 accelerate optimization methods has gained a lot of attention and led to many developments
14 (see, e.g., [3, 10, 15, 18, 19, 52] and the references therein).

15 In this paper, we then consider second-order dynamics of the form

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla f_{\mathbf{m}_{\bar{x}}}(x(t)) + \omega \nabla^2 f_{\mathbf{m}_{\bar{x}}}(x(t))\dot{x}(t) + \mathbf{e}_{\bar{x}}(x(t)) = 0, \quad (8)$$

16 where $f_{\mathbf{m}_{\bar{x}}}(x) = \mathbb{E}_{\xi \sim \mathbf{m}_{\bar{x}}} f(x, \xi)$, $\mathbf{e}_{\bar{x}}$ is an perturbation function and ω is the so-called Hessian-
17 driven damping coefficient. When $\gamma(t) \equiv \gamma$, $f_{\mathbf{m}_{\bar{x}}} = f$ (i.e., without a stochastic structure)
18 and $\mathbf{e}_{\bar{x}} = 0$, systems of the form (8) were first studied in [3]. Later, this system was combined
19 with an asymptotic vanishing damping $\gamma(t) = \frac{\alpha}{t}$, for $\alpha > 0$ in [19]. Several recent studies
20 have been devoted to this topic (see, e.g., [9, 25, 38, 40, 51]).

21 **1.1. Statement of the problem.** Throughout, \mathcal{H} is a real Hilbert space and Ξ is a Polish
22 metric space, i.e., separable and completely metrizable. Given a maximal monotone operator
23 $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ such that $\text{int}(\text{dom}(A)) \neq \emptyset$, a single-valued mapping $B : \overline{\text{dom}(A)} \times \Xi \rightarrow \mathcal{H}$, and
24 consider the closed-loop inclusion

$$\begin{cases} \dot{x}(t) + A(x(t)) + \mathbb{E}_{\xi \sim \mathbf{m}_{x(t)}}(B(x(t), \xi)) \ni 0, & \text{a.e } t > t_0 > 0, \\ x(t_0) = x_0 \in \overline{\text{dom}(A)}, \end{cases} \quad (\text{SMI})$$

25 where \mathbf{m}_x is a family of probability distributions on Ξ indexed by $x \in \mathcal{H}$, such that $\mathbf{m}_x(C)$ is
26 a measurable function on \mathcal{H} for each fixed $C \in \mathcal{B}$, where \mathcal{B} is the Borel σ -algebra of Ξ .

27 We call the above dynamic (SMI), which stands for Stochastic Monotone Inclusion.

28 **► Example 1.1.** (1) Typically, (SMI) covers the case of stochastic optimization problems
29 with a state-dependent distribution studied in [35, 47] by taking $A = \partial g$ and $B = \nabla f$
30 for $g \in \Gamma_0(\mathcal{H})$ and $f(\cdot, \xi) \in C_{\beta}^{1,1}(\mathcal{H}) \cap \Gamma_0(\mathcal{H})$.

31 (2) Taking $A = N_C$ the normal cone of convex set C of admissible decisions, we recover
32 the framework of variational inequalities addressed recently in [32].

33 **Notation.** To simplify the presentation, we set, for any measure $\mathbf{m} \in \mathcal{P}(\Xi)$

$$B_{\mathbf{m}}(x) = \mathbb{E}_{\xi \sim \mathbf{m}}(B(x, \xi)) \text{ and } F_{\mathbf{m}}(x) = A(x) + B_{\mathbf{m}}(x), \quad (9)$$

34 where $\mathbb{E}_{\xi}(\cdot)$ is the expectation with respect to the random variable ξ .

35 Using this notation, the system (SMI) can be simply rewritten as

$$\begin{cases} \dot{x}(t) + F_{\mathbf{m}_{x(t)}}(x(t)) \ni 0, & \text{a.e } t > t_0, \\ x(t_0) = x_0 \in \overline{\text{dom}(A)}. \end{cases} \quad (\text{SMI})$$

1 In the sequel, we need one of the following assumptions.

2 **Assumption 1.** We suppose that $\xi \mapsto B(x, \xi)$ is measurable for any $x \in \mathcal{H}$ and that
3 $x \mapsto B(x, \xi)$ is β -Lipschitz continuous for all ξ .

4 The following assumption is essential for the convergence analysis. It describes the
5 sensitivity of the distribution to shifts in the index (here trajectory). It is widely used in the
6 literature (see, e.g., [35, 43, 47, 54]). We will discuss how it closely relates to the so-called
7 coarse Ricci curvature in Section 5.

8 **Assumption 2** (Lipschitz distributions). There exists $\tau > 0$ such that

$$\mathbb{W}_1(\mathbf{m}_x, \mathbf{m}_y) \leq \tau \|x - y\|, \text{ for all } x, y \in \mathcal{H}.$$

9 The last two assumptions are standard monotonicity assumptions essential to the well-
10 posedness of the dynamics and for the existence and uniqueness of equilibria.

11 **Assumption 3** (Strong monotonicity). $F_{\mathbf{m}_x}$ is μ -strongly monotone for all $x \in \mathcal{H}$, for $\mu > 0$.

12 Rather than using Assumption 3, one can use uniform monotonicity in the sense of
13 Definition 3 below.

14 **Assumption 4** (Uniform monotonicity). For all $x \in \mathcal{H}$, $F_{\mathbf{m}_x}$ is uniformly monotone with a
15 modulus ϕ such that

$$\phi(t) > \beta \tau t, \quad \forall t > 0.$$

16 Finally, define the following parameter $\rho := \frac{\beta \tau}{\mu}$. As we will see, (see also [35, 47]), the
17 parameter regime $\rho < 1$ will play a crucial role in the analysis of the convergence of the
18 trajectories.

19 **1.2. Organization of the paper.** The paper is organized as follows. In Section 3 we address
20 first-order monotone inclusions with closed-loop distributions. We prove the existence of
21 equilibria as well as the well-posedness of the dynamics and convergence properties of the
22 trajectories. In Section 4 we study asymptotic convergence properties of the trajectories of
23 second-order dynamics with closed-loop distributions via Hessian damping. This allows us in
24 particular to cover problems of the form (2). Section 5 contains a discussion concerning the
25 Lipschitz behavior of the family $(\mathbf{m}_x)_x$ with respect to \mathbb{W}_1 -distance and some consequences
26 in the framework of Markov chains on metric random walk spaces. In Section 6 we discuss
27 the inertial primal-dual algorithm as an application of our results. Finally, Section 7 contains
28 some conclusions and discusses some future works.

29 2. NOTATION AND PRELIMINARIES

30 In this section we fix some notation and present some notions and results that will be
31 used. Throughout this paper, \mathcal{H} is a real Hilbert space, $\langle \cdot, \cdot \rangle$ is the scalar product on \mathcal{H} and
32 $\|\cdot\|$ is the induced norm.

33 **2.1. Convex analysis.** We denote by $\Gamma_0(\mathcal{H})$ the class of proper, l.s.c and convex function
34 on \mathcal{H} with values in $\mathbb{R} \cup \{\infty\}$. Given $g \in \Gamma_0(\mathcal{H})$, its domain is defined by $\text{dom}(g) = \{x \in$
35 $\mathcal{H} : g(x) < \infty\}$. We say that g is α -strongly convex, for $\alpha > 0$, if $g - \frac{\alpha}{2} \|\cdot\|^2$ is convex.

36 The subdifferential of g is defined as

$$\partial g : x \in \mathcal{H} \mapsto \{v \in \mathcal{H} : g(y) \geq g(x) + \langle v, y - x \rangle\}.$$

37 We recall the following Fermat's optimality condition for $g \in \Gamma_0(\mathcal{H})$,

$$0 \in \partial g(x^*) \Leftrightarrow x^* \in \arg \min g(\mathcal{H}).$$

1 **Definition 1** (Differentiability). Let $g : \mathcal{H} \rightarrow \mathbb{R} \cup \{\infty\}$ and $x \in \text{int}(\text{dom}(g))$. We say that g is
 2 differentiable at x if there exists $v \in \mathcal{H}$ such that

$$\lim_{h \rightarrow 0} \frac{g(x+h) - g(x) - \langle v, h \rangle}{\|h\|} = 0.$$

3 The unique vector v satisfying this condition is the gradient of g at x and is denote by $\nabla g(x)$.

4 This being said, we recall the following

5 **Definition 2** (L -smoothness). Let $L \geq 0$ and $g : \mathcal{H} \rightarrow \mathbb{R} \cup \{\infty\}$. We say that g is L -smooth
 6 over $D \subset \mathcal{H}$ if it is differentiable over D and

$$\|\nabla g(x) - \nabla g(y)\| \leq L\|x - y\| \text{ for any } x, y \in D.$$

7 We denote by $C_L^{1,1}(D)$ the class of L -smooth functions over D .

8 Given a function $g : \mathcal{H} \rightarrow \mathbb{R} \cup \{\infty\}$, its proximal mapping is defined through

$$\text{prox}_f(x) = \arg \min_y \left\{ g(y) + \frac{1}{2}\|y - x\|^2 \right\} \text{ for any } x \in \mathcal{H}.$$

9 When $g = \iota_K$ the indicator function of a nonempty closed convex set $K \subset \mathcal{H}$, then
 10 $\text{prox}_f = \text{pr}_K$. For further details and notion, we refer the reader to [23].

11 **2.2. Operator theory.** We write $A : \mathcal{H} \rightrightarrows \mathcal{H}$ to denote that A is a set-valued operator on
 12 \mathcal{H} , its domain is defined as $\text{dom}(A) = \{x \in \mathcal{H} : A(x) \neq \emptyset\}$, its graph as $\text{gra}(A) = \{[x, u] \in$
 13 $\mathcal{H} \times \mathcal{H} : u \in Ax\}$ and its zeros set as: $\text{zer}(A) = \{x \in \mathcal{H} : 0 \in A(x)\} := A^{-1}(0)$. A selection
 14 of A is an operator $T : \text{dom } A \rightarrow \mathcal{H}$ such that, $Tx \in Ax$ for any $x \in \text{dom } A$. We write
 15 $A : \mathcal{H} \rightarrow \mathcal{H}$ to indicate that A is single-valued. In the following, we gather some main
 16 properties that are essential for the rest of the paper.

17 **Definition 3.** • We say that A is β -Lipschitz continuous if it is single-valued over $\text{dom } A$
 18 and

$$\|Ax - Ay\| \leq \beta\|x - y\| \quad \forall x, y \in \text{dom } A. \quad (10)$$

19 • We say that $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is monotone if

$$\langle x - y, u - v \rangle \geq 0 \quad \forall [x, u], [y, v] \in \text{gra } A. \quad (11)$$

20 • We say that A is maximal monotone if there exists no monotone operator B , i.e.,
 21 satisfying (11), such that $\text{gra } A \subset \text{gra } B$.

22 • We say that A is uniformly monotone with modulus $\phi : [0, \infty) \rightarrow [0, \infty)$ if ϕ is increasing,
 23 $\phi(0) = 0$, $\lim_{t \rightarrow \infty} \phi(t) = \infty$ and

$$\langle x - y, u - v \rangle \geq \|x - y\| \phi(\|x - y\|) \quad \forall [x, u], [y, v] \in \text{gra } A. \quad (12)$$

24 • We say that A is μ -strongly monotone, with $\mu > 0$, if

$$\langle x - y, u - v \rangle \geq \mu\|x - y\|^2 \quad \forall [x, u], [y, v] \in \text{gra } A. \quad (13)$$

25 **Remark 2.1.** • Note that if A is μ -strongly monotone is equivalent to saying that $A - \mu\text{Id}$
 26 is monotone.

27 • The definition of uniform monotonicity given by (12) is slightly different from the one
 28 in [23, Definition 22.1].

29 • If A is μ -strongly monotone, then it is uniformly monotone with modulus $\phi(t) = \mu t$.

30 **► Example 2.1.** • The typical example of a maximal monotone operator is the subdiffer-
 31 ential ∂g of a function $g \in \Gamma_0(\mathcal{H})$. We usually refer to such an operator as a subpotential
 32 maximal monotone operator.

1 **2.3. Monotone inclusions.** Let A be an operator on \mathcal{H} such that $\text{int}(\text{dom } A) \neq \emptyset$ and single-
 2 valued mapping $D : [0, \infty) \times \overline{\text{dom}(A)} \rightarrow \mathcal{H}$ and consider the following differential inclusion

$$\begin{cases} \dot{x}(t) + A(x(t)) + D(t, x(t)) \ni 0, & t \in [0, T] \\ x(0) = x_0 \in \text{dom}(A). \end{cases} \quad (14)$$

Definition 4. A strong solution to (14) is a function $x \in W^{1,1}((0, T]; \mathcal{H}) \cap C([0, T]; \mathcal{H})$ such that

$$-\dot{x}(t) \in A(x(t)) + D(t, x(t)), \quad \text{a.e. } t \in (0, T), x(0) = x_0.$$

3 Here, $W^{1,1}((0, T]; \mathcal{H}) = \{x \in L^1(0, T; \mathcal{H}); \dot{x} \in L^1(\epsilon, T; \mathcal{H}), \forall \epsilon \in (0, T)\}$.

4 A trajectory $x : [0, \infty) \rightarrow \mathcal{H}$ is a strong global solution of (14) if it is a strong solution on
 5 $[0, T]$ for any $T > 0$.

6 For further details, we refer the reader to the classical monograph [27] or [22].

7 **2.4. Transportation distance.** For $m_1, m_2 \in \mathcal{P}(\Xi)$, the W_1 -Wasserstein distance is defined by

$$W_1(m_1, m_2) = \sup_{u \in \text{Lip}_1} \mathbb{E}_{\xi \sim m_1} u(\xi) - \mathbb{E}_{\zeta \sim m_2} u(\zeta), \quad (15)$$

8 where Lip is the set of Lipschitz continuous function on Ξ .

9 3. FIRST-ORDER MONOTONE INCLUSIONS

10 In this section we perform the analysis of the first-order monotone inclusion (SMI). More
 11 precisely, we discuss the existence and uniqueness of solutions as well as the convergence of
 12 trajectories. Recall that the dynamic (SMI) is governed by the operator $F_{m_x} = A + B_{m_x}$.
 13 We first prove the existence of an equilibrium point \bar{x} which will allow us to reformulate
 14 (SMI) in a suitable form.

15 3.1. Existence and uniqueness of equilibria.

16 **Definition 5.** (Equilibrium point) We say that $\bar{x} \in \mathcal{H}$ is at equilibrium with respect to the
 17 distribution $m(\cdot)$ if

$$0 \in F_{m_{\bar{x}}}(\bar{x}). \quad (16)$$

18 In case $A = \partial g$ and $B = \nabla f$ where $g \in \Gamma_0(\mathcal{H})$ and $f(\cdot, \xi) \in C_{\beta}^{1,1}(\mathcal{H}) \cap \Gamma_0(\mathcal{H})$, this
 19 definition is to be compared to the one introduced in [47] (see also [35]). Indeed, (16) reduces
 20 to:

$$\bar{x} \in \arg \min_x \mathbb{E}_{\xi \sim m_{\bar{x}}} f(x, \xi) + g(x). \quad (17)$$

21 Solutions of (17) are exactly the fixed points of the repeated minimization procedure, that
 22 is, starting from some x_0 , we generate the following sequence for $t \geq 0$

$$x_{t+1} = S(x_t) := \arg \min_x \mathbb{E}_{\xi \sim m_{x_t}} f(x, \xi) + g(x). \quad (18)$$

23 In [47, Theorem 3.5] it is shown that if f is C^1 in both variables, $\xi \mapsto \nabla f(x, \xi)$ is
 24 β -Lipschitz and $\mathbb{E}_{\xi \sim m_x} f(\cdot, \xi)$ is μ -strongly convex for all $x \in \mathcal{H}$ with $\rho < 1$, then, under
 25 **Assumption 2**, the iterates of (18) converge to a unique stable point. Their proof is
 26 essentially based on a fixed point argument. In [47, Propostion 4.1], they show the existence
 27 of equilibrium points under weaker assumptions on the loss f . Specifically, they demonstrate
 28 that if f is convex and jointly continuous, then equilibrium points exist provided $\text{dom}(g)$ is
 29 compact. However, in this case the equilibrium is not necessarily unique. In the following
 30 lemma, we show the existence of equilibrium in the sense of (16).

1 **Theorem 1** (Existence and uniqueness of equilibrium point). Under [Assumption 4](#), the map

$$S : x \in \mathcal{H} \mapsto \text{zer}(F_{m_x}) = \{u \in \mathcal{H} : 0 \in F_{m_x}(u)\},$$

2 is a contraction. In particular, the equilibrium is unique. If moreover, [Assumption 3](#) holds
3 instead, i.e., F_{m_x} is μ -strongly monotone for $\mu > 0$, the mapping S is ρ -Lipschitz.

4 *Proof.* First, we see that S is well defined. Indeed, for any $x \in \mathcal{H}$, $\text{zer}(F_{m_x})$ is nonempty,
5 and is in fact a singleton due to [Assumption 4](#). To see this, we argue as in [[23](#), Proposition
6 22.11]. Fix $[y_0, u_0] \in \text{gra } F_{m_x}$. We have for any $[y, u] \in \text{gra } F_{m_x}$:

$$\begin{aligned} \|y - y_0\| \|u\| &\geq \langle y - y_0, u \rangle = \langle y - y_0, u - u_0 \rangle + \langle y - y_0, u_0 \rangle \\ &\geq \|y - y_0\| \phi(\|y - y_0\|) - \|y - y_0\| \|u_0\|. \end{aligned} \quad (19)$$

7 Since $\lim_{t \rightarrow \infty} \phi(t) = \infty$ we infer that $\inf_{u \in \text{gra } F_{m_x}(y)} \|u\| \rightarrow \infty$ as $\|y\| \rightarrow \infty$ and thus F_{m_x}
8 is surjective (see [[23](#), Corollary 21.25]). Moreover, in view of strict monotonicity of F_{m_x} ,
9 $\text{zer}(F_{m_x})$ is a singleton (see, e.g., [[23](#), Proposition 23.35]).

10 Now, pick $x, y \in \mathcal{H}$. We have that $0 \in F_{m_x}(S(x))$ and $0 \in F_{m_y}(S(y))$. In particular,
11 $-B_{m_y}(S(y)) \in A(S(y))$, which gives that $B_{m_x}(S(y)) - B_{m_y}(S(y)) \in F_{m_x}(S(y))$. We get,
12 thanks to [Assumption 4](#)

$$\|S(x) - S(y)\| \phi(\|S(x) - S(y)\|) \leq \langle u - v, S(x) - S(y) \rangle, \text{ for any } (u, v) \in F_{m_x}(S(y)) \times F_{m_x}(S(x)).$$

13 Then, taking $u = B_{m_x}(S(y)) - B_{m_y}(S(y))$ and $v = 0$, we get, using Cauchy-Schwarz inequality

$$\phi(\|S(x) - S(y)\|) \leq \|B_{m_x}(S(y)) - B_{m_y}(S(y))\|.$$

14 We get, using [Corollary 1](#) below

$$\phi(\|S(x) - S(y)\|) \leq \beta\tau \|x - y\|. \quad (20)$$

15 Since ϕ is strictly increasing, the last inequality gives thanks to [Assumption 4](#)

$$\|S(x) - S(y)\| \leq \phi^{-1}(\beta\tau \|x - y\|) < \|x - y\|,$$

16 and by Banach's fixed-point theorem (see [Theorem 6](#)), S has a unique fixed point \bar{x} .

17 Now if F_{m_x} is μ -strongly monotone, it is in particular uniformly monotone with modulus
18 $\phi(t) = \mu t$. Equation-([20](#)) gives

$$\|S(x) - S(y)\| \leq \rho \|x - y\|,$$

19 with $\frac{\beta\tau}{\mu} := \rho$. Again, we conclude using [Assumption 4](#) and [Theorem 6](#). ■

20 *Remark 3.1.* In the strongly monotone case, [Assumption 4](#) incorporates the parameter regime
21 $\rho < 1$ which appears in particular in [[47](#), Theorem 3.5].

22 *Remark 3.2.* Since $\text{zer}(F_{m_{\bar{x}}}) = \{\bar{x}\}$, we clearly see that $\bar{x} \in \text{zer}(F_{m_{\bar{x}}} + \mathbf{e}_{\bar{x}})$. Indeed, taking
23 into account ([23](#)), we have that $\mathbf{e}_{\bar{x}}(\bar{x}) = 0$ so that $0 \in (F_{m_{\bar{x}}} + \mathbf{e}_{\bar{x}})(\bar{x})$.

24 *Notation 1.* Before stating the main result of this section, let us fix the following notation.
25 Assume that [Assumption 4](#) holds, and denote by

$$\varphi(t) = \phi(t) - \beta\tau t, \quad (21)$$

26 where ϕ is the modulus of uniform monotonicity of $F_{m_{\bar{x}}}$.

27 We prove the following

1 **Lemma 1.** *Let $a > 0$ and define*

$$\theta(z) := \int_z^a \frac{ds}{\varphi(s)}. \quad (22)$$

2 *Then θ is nonincreasing and $\lim_{z \rightarrow 0^+} \theta(z) = \infty$.*

3 *Proof.* Indeed, we have $\dot{\theta}(z) = -1/\varphi(z) < 0$ since ϕ satisfies [Assumption 4](#). Moreover, since
4 $\phi(t) \geq \varphi(t)$, and $\lim_{z \rightarrow 0^+} \int_z^a \frac{ds}{\phi(s)} = \infty$, the result follows. ■

6 *Remark 3.3.* In the literature of ordinary differential equations, the above lemma is related
7 to the fact that φ is somehow an *Osgood modulus of continuity* (see, e.g., [20, Definitions
8 2.108 and 3.1]). If $\varphi(s) = s$, which corresponds to Lipschitz regularity, and $a = 1$ then
9 $\theta(z) = \log_+(z) = \max\{0, \log(1/z)\}$. If $a = 1/e$ and $\varphi(s) = s \log(1/s)$, which corresponds to
10 log-Lipschitz regularity, then $\theta(z) = \log \log_+(z)$. More generally, $\varphi(s) = s (\log(1/s))^r$ for
11 $r \leq 1$ are admissible choices.

12 **3.2. Well-posedness and convergence.** Now let us define the following gap

$$\mathbf{e}_{\bar{x}}(x) = B_{\mathbf{m}_x}(x) - B_{\mathbf{m}_{\bar{x}}}(x). \quad (23)$$

13 Using the notation in (9), we may rewrite (SMI) in the following form

$$\begin{cases} \dot{x}(t) + F_{\mathbf{m}_{\bar{x}}}(x(t)) + \mathbf{e}_{\bar{x}}(x(t)) \ni 0, & \text{a.e } t > t_0 \\ x(t_0) = x_0, \end{cases} \quad (\text{p-SMI})$$

14 One advantage of this formulation is that the mapping $x \mapsto \mathbf{e}_{\bar{x}}(x)$ exhibits Lipschitz behaviour
15 (see [Lemma 5](#)), and the operator $F_{\mathbf{m}}$ depends only on $\mathbf{m}_{\bar{x}}$ instead of $\mathbf{m}_{x(t)}$. This allows
16 us to treat (p-SMI) within the framework of evolution equations governed by Lipschitz
17 perturbations of maximal monotone operators (cf. [27, Chapter III]). This is behind the
18 notation (p-SMI), which stands for perturbed stochastic monotone inclusion. This being
19 said, our aim is to use [27, Proposition 3.13] and show the existence of a unique strong
20 solution (cf. [Definition 4](#)) to (p-SMI) (see also [27, Definition 3.1]). To this end we begin
21 with the following lemmas.

22 **Lemma 2.** *The operator $F_{\mathbf{m}_{\bar{x}}}$ is maximally monotone.*

23 *Proof.* Since A is maximal monotone, and $B(\cdot, \xi)$ is β -Lipschitz continuous for all ξ , the
24 result is a direct consequence of [27, Lemma 2.4]. ■

25 For completeness of the presentation, we also recall the following elementary results.

26 **Lemma 3.** *Under [Assumption 1](#), we have, for any $\mathbf{m}, \nu \in \mathcal{P}(\Xi)$*

$$\sup_{x \in \mathcal{H}} \|B_{\mathbf{m}}(x) - B_{\nu}(x)\| \leq \beta \mathbb{W}_1(\mathbf{m}, \nu).$$

27 *Proof.* By [Assumption 1](#), B is β -Lipschitz continuous. We then have from the definition of
28 the \mathbb{W}_1 -Wasserstein distance (15)

$$B_{\mathbf{m}}(x) - B_{\nu}(x) = \mathbb{E}_{\xi \sim \mathbf{m}} B(x, \xi) - \mathbb{E}_{\zeta \sim \nu} B(x, \zeta) \leq \beta \mathbb{W}_1(\mathbf{m}, \nu).$$

29 Taking the supremum over $x \in \mathcal{H}$ yields the result. ■

30 Combining [Lemma 3](#) and [Assumption 2](#) we get the following.

31 **Corollary 1.** *Under [Assumption 1](#) and [Assumption 2](#), for all $y, z \in \mathcal{H}$*

$$\sup_{x \in \mathcal{H}} \|B_{\mathbf{m}_y}(x) - B_{\mathbf{m}_z}(x)\| \leq \beta \tau \|y - z\|.$$

1 **Lemma 4.** Under [Assumption 1](#), we have for any $x, y \in \mathcal{H}$ and $\mathfrak{m} \in \mathcal{P}(\Xi)$

$$\|B_{\mathfrak{m}}(y) - B_{\mathfrak{m}}(x)\| \leq \beta \|x - y\|.$$

2 *Proof.* Let $\mathfrak{m} \in \mathcal{P}(\Xi)$ and $x, y \in \mathcal{H}$. Since $B(\cdot, \xi)$ is β -Lipschitz continuous, we have

$$B_{\mathfrak{m}}(x) - B_{\mathfrak{m}}(y) = \mathbb{E}_{\xi \sim \mathfrak{m}} (B(x, \xi) - B(y, \xi)) \leq \beta \|x - y\|.$$

3 The result follows immediately. \blacksquare

4 The key ingredient to use [[27](#), Proposition 3.13] is the Lipschitz continuity of the pertur-
5 bation $\mathbf{e}_{\bar{x}}(\cdot)$. This is the content of the following statement.

6 **Lemma 5.** Suppose that [Assumption 1](#) and [Assumption 2](#) hold, then $\mathbf{e}_{\bar{x}}(\cdot)$ is $\beta(2+\tau)$ -Lipschitz
7 continuous.

8 *Proof.* Let $x, z \in \overline{\text{dom}(F_{\mathfrak{m}_{\bar{x}}})}$. We then have

$$\begin{aligned} \mathbf{e}_{\bar{x}}(x) - \mathbf{e}_{\bar{x}}(z) &= (B_{\mathfrak{m}_x}(x) - B_{\mathfrak{m}_{\bar{x}}}(x)) - (B_{\mathfrak{m}_z}(z) - B_{\mathfrak{m}_{\bar{x}}}(z)) \\ &= (B_{\mathfrak{m}_x}(x) - B_{\mathfrak{m}_x}(z)) + (B_{\mathfrak{m}_{\bar{x}}}(z) - B_{\mathfrak{m}_{\bar{x}}}(x)) + (B_{\mathfrak{m}_x}(z) - B_{\mathfrak{m}_z}(z)) \\ &\leq \beta \|x - z\| + \beta \|x - z\| + \beta \tau \|x - z\| \\ &= \beta(2 + \tau) \|x - z\|, \end{aligned} \quad (24)$$

9 where we have used [Corollary 1](#) and [Lemma 4](#) in the inequality. \blacksquare

10 **Proposition 1.** Assume [Assumption 1](#) and [Assumption 2](#) hold. Then, given $x_0 \in \overline{\text{dom}(F_{\mathfrak{m}_{\bar{x}}})}$,
11 the dynamic ([p-SMI](#)), and hence ([SMI](#)), admits a unique solution $x \in W^{1,1}(t_0, T; \mathcal{H})$.

12 *Proof.* Thanks to [Lemma 2](#) and [Lemma 5](#) $F_{\mathfrak{m}_{\bar{x}}}$ is maximal monotone and $\mathbf{e}_{\bar{x}}$ is Lipschitz
13 continuous. Since $t \mapsto \mathbf{e}_{\bar{x}}(x)$ is trivially in $L^\infty(t_0, T; \mathcal{H})$, we deduce, thanks to [[27](#), Proposition
14 3.13] the existence of a unique solution $x \in W^{1,1}(t_0, T; \mathcal{H})$ to ([SMI](#)). \blacksquare

15 *Remark 3.4.* Evolution problems of the form ([p-SMI](#)) were also studied in [[13](#)], where the
16 operator $\mathbf{e}_{\bar{x}}$ is potentially multivalued and dependent on both time and space variables.
17 While standard time-variable measurability is assumed, the authors' analysis relied on the
18 upper semicontinuity of $x \mapsto \mathbf{e}_{\bar{x}}(t, x)$ rather than Lipschitz continuity.

19 **Theorem 2.** Let x be the solution of ([SMI](#)), and suppose that [Assumption 4](#) holds. Then,
20 for all $t \geq t_0$, we have

$$\|x(t) - \bar{x}\| \leq \theta^{-1} (t - \hat{t}), \quad (25)$$

21 where θ is defined in [Lemma 1](#) and $\hat{t} = 2t_0 - \theta (\|x_0 - \bar{x}\|)$.

22 Furthermore, if [Assumption 3](#) holds instead, we have

$$\|x(t) - \bar{x}\| \leq C e^{-2\mu(1-\rho)t} \text{ for all } t \geq t_0, \quad (26)$$

23 with $C = \|x_0 - \bar{x}\| e^{2\mu(1-\rho)t_0}$ and $\rho := \frac{\beta\tau}{\mu}$.

24 *Proof.* Let $y \in \mathcal{H}$, we observe that

$$\frac{1}{2} \frac{d}{dt} \|x(t) - y\|^2 = \langle \dot{x}(t), x(t) - y \rangle.$$

25 We then have, for any selection $u(t)$ of $F_{\mathfrak{m}_{\bar{x}}}(x(t))$

$$\frac{1}{2} \frac{d}{dt} \|x(t) - y\|^2 = \langle \dot{x}(t), x(t) - y \rangle = -\langle u(t), x(t) - y \rangle - \langle \mathbf{e}_{\bar{x}}(x(t)), x(t) - y \rangle. \quad (27)$$

26 Taking $y \in \text{zer}(F_{\mathfrak{m}_{\bar{x}}}) = \{\bar{x}\}$, we get by [Assumption 3](#)

$$\langle u(t) - 0, x(t) - \bar{x} \rangle \geq \|x(t) - \bar{x}\| \phi(\|x(t) - \bar{x}\|).$$

1 Then, thanks to [Lemma 3](#) and [Assumption 2](#), equation (27) gives

$$\begin{aligned}
\frac{1}{2} \frac{d}{dt} \|x(t) - \bar{x}\|^2 &= -\langle u(t) - 0, x(t) - \bar{x} \rangle - \langle \mathbf{e}_{\bar{x}}(x(t)), x(t) - \bar{x} \rangle \\
&\leq -\|x(t) - \bar{x}\| \phi(\|x(t) - \bar{x}\|) + \|\mathbf{e}_{\bar{x}}(x(t))\| \|x(t) - \bar{x}\| \\
&\leq -\|x(t) - \bar{x}\| \phi(\|x(t) - \bar{x}\|) + \beta\tau \|x(t) - \bar{x}\|^2 \\
&= -\|x(t) - \bar{x}\| \varphi(\|x(t) - \bar{x}\|),
\end{aligned} \tag{28}$$

2 where $\varphi(t) = \phi(t) - \beta\tau t$ is as introduced in (21). Hence, dividing both sides of (28) by
3 $h(t) := \|x(t) - \bar{x}\| \neq 0$, we get

$$\dot{h}(t) \leq -2\varphi(h(t)). \tag{29}$$

4 From (28) and (29), we infer that

$$\frac{d}{dt} \theta(h(t)) = \frac{-1}{\varphi(h(t))} \dot{h}(t) \geq \left(\frac{1}{\varphi(h(t))} \right) (2\varphi(h(t))) = 2. \tag{30}$$

5 Integrating (30) between t_0 and $t > 0$, we get

$$\theta(h(t)) - \theta(h(t_0)) \geq 2(t - t_0).$$

6 Using [Lemma 1](#), we deduce that

$$h(t) \leq \theta^{-1}(2t - \hat{t}), \tag{31}$$

7 where $\hat{t} = 2t_0 - \theta(h(t_0))$. This proves (25), as desired.

8 To prove the second claim, we observe that if $F_{\mathbf{m}_{\bar{x}}}$ is μ -strongly monotone, $\phi(t) = \mu t$ so that
9 $\varphi(t) = (\mu - \beta\tau)t$. Assuming that $\mu \neq \beta\tau$, we get $\theta(s) = \frac{-\log(s)}{\mu - \beta\tau}$. So that $\theta^{-1}(s) = e^{-(\mu - \beta\tau)s}$.
10 Plugging this into (31), we obtain

$$h(t) \leq \|x_0 - \bar{x}\| e^{-2(\mu - \beta\tau)(t - t_0)}, \tag{32}$$

11 as desired. ■

12 *Remark 3.5.* Observe that one can obtain (26) from (28) using Gronwall's [Lemma 9](#). Indeed,
13 taking $\phi(t) = \beta\mu t$, we infer from (28)

$$\frac{1}{2} \frac{d}{dt} \|x(t) - \bar{x}\|^2 \leq -(\mu - \beta\tau) \|x(t) - \bar{x}\|^2.$$

14 By [Assumption 2](#) and [Theorem 2](#), we obtain a rate of convergence of the measure $\mathbf{m}_{x(t)}$
15 to $\mathbf{m}_{\bar{x}}$ in the \mathbb{W}_1 distance as $t \rightarrow \infty$.

16 **Corollary 2.** *Let $x : [t_0, \infty) \rightarrow \mathbb{R}$ be the solution of (SMI). Under [Assumption 4](#), we have*
17 *for all $t \geq t_0$*

$$\mathbb{W}_1(\mathbf{m}_{x(t)}, \mathbf{m}_{\bar{x}}) \leq \tau \theta^{-1}(t - \hat{t}),$$

18 *where θ is defined in [Lemma 1](#) and $\hat{t} = 2t_0 - \theta(\|x_0 - \bar{x}\|)$. If [Assumption 3](#) holds instead,*
19 *we have*

$$\mathbb{W}_1(\mathbf{m}_{x(t)}, \mathbf{m}_{\bar{x}}) \leq C e^{-2\mu(1-\rho)t},$$

20 *with $C = \|x_0 - \bar{x}\| e^{2\mu(1-\rho)t_0}$.*

4. SECOND-ORDER MONOTONE INCLUSIONS VIA HESSIAN DAMPING

This section is devoted to the analysis of the following second-order monotone inclusion

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + F_{m_{x(t)}}(x(t)) \ni 0, & \text{a.e } t \in [t_0, T] \\ x(t_0) = x_0 \in \overline{\text{dom}(F_{m_{x(t)}})}, \end{cases} \quad (\text{ISMI})$$

where $\gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is a continuous function, usually referred to as the viscous damping coefficient. We refer to this dynamic as an Inertial Stochastic Monotone Inclusion, (ISMI) for short. In a smooth deterministic setting, i.e., $F_{m_x} = \nabla f$ where f is a smooth, strongly convex function, systems of the form (ISMI) were first studied by Polyak in [48] with a fixed viscous damping coefficient. Later on, this kind of systems were studied by [15] and then [52] where the authors establish the link between the continuous dynamics with $\gamma(t) = \frac{3}{t}$ and the Nesterov's methods [44]. Further results and extensions followed (see, e.g., [5, 11, 16, 39] and the references therein).

Now, coming back to our setting, we reformulate (ISMI) using the notations of Section 3, as follows

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + F_{m_{\bar{x}}}(x(t)) + \mathbf{e}_{\bar{x}}(x(t)) \ni 0, & \text{a.e } t \in [t_0, T] \\ x(t_0) = x_0 \in \overline{\text{dom}(F_{m_{\bar{x}}})}. \end{cases} \quad (\text{p-ISMI})$$

One of the drawbacks of the perturbed stochastic inertial monotone inclusion (p-ISMI) is that it cannot be written as a dynamical system governed by Lipschitz perturbations of monotone operators. This being said, we shall consider (p-ISMI) with the presence of the so-called Hessian damping. To avoid technicalities, we restrict ourselves to the smooth case, i.e., $A = \nabla g$ for $g \in \Gamma_0(\mathcal{H}) \cap C^1(\mathcal{H})$ so that

$$F_{m_{\bar{x}}} = \nabla G_{m_{\bar{x}}} \text{ with } G_{m_{\bar{x}}} = g + \mathbb{E}_{\xi \sim m_{\bar{x}}} f(x, \xi). \quad (33)$$

In this case the perturbation defined in (23) is given by

$$\mathbf{e}_{\bar{x}} = \mathbb{E}_{\xi \sim m_x} \nabla f(x, \xi) - \mathbb{E}_{\xi \sim m_{\bar{x}}} \nabla f(x, \xi), \quad (34)$$

and it measure the gradient deviation. This being said, (p-ISMI) reduces to an inertial gradient system

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla G_{m_{\bar{x}}}(x(t)) + \mathbf{e}_{\bar{x}}(x(t)) = 0, \quad (35)$$

which, when $\gamma(t) \equiv \gamma$, is essentially a Polyak's heavy ball (HBF) system [48] with the presence of the perturbation $\mathbf{e}_{\bar{x}}$. The Hessian-driven damped system we shall consider reads

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla G_{m_{\bar{x}}}(x(t)) + \omega \nabla^2 G_{m_{\bar{x}}}(x(t))\dot{x}(t) + \mathbf{e}_{\bar{x}}(x(t)) = 0, \quad (\text{ISEHD}_{m, \gamma})$$

where $\omega : [0, \infty) \rightarrow \mathbb{R}^+$ is a continuous function usually referred to as the Hessian-driven damping coefficient, and which will be taken to be constant, i.e., $\omega(t) \equiv \omega > 0$. We call the above dynamic an Inertial System with Explicit Hessian Damping, (ISEHD_{m, γ, ω}) for short. The subscript m is here to emphasize the stochastic structure of the problem. Here, the Hessian damping is said to be explicit since, when $G_{m_{\bar{x}}}$ is of class C^2 , we have

$$\frac{d}{dt} (\nabla G_{m_{\bar{x}}}(x(t))) = \nabla^2 G_{m_{\bar{x}}}(x(t))\dot{x}(t).$$

One of the main advantages of considering dynamics with Hessian-driven damping is that it helps to attenuate the oscillation that might occur in (HBF)-like dynamics or more generally with inertial dynamics with viscous damping as observed in [3]. Variants and generalizations were studied by multiple authors (see, e.g., [9, 18, 19]). The study of the effect of perturbed Hessian-driven damping, i.e., problems of the form (ISEHD_{m, γ, ω}) was carried out in the recent work [14].

1 In particular, this will allow us to proceed as in [Section 3](#) to prove the well-posedness
 2 of the dynamics. As for convergence properties, this setting will allow us to exploit some
 3 techniques from [\[14\]](#).

4 4.1. Well-posedness.

5 4.1.1. Equivalent formulation.

6 **Proposition 2.** *Suppose that $\gamma(t) \geq 0, \omega > 0$. For any initial conditions $(x_0, v_0) \in \mathcal{H} \times \mathcal{H}$,
 7 the dynamics $(\text{ISEHD}_{\mathbf{m}, \gamma, \omega})$ admits an equivalent formulation of the form*

$$\begin{cases} \dot{x}(t) + \omega \left(\nabla G_{\mathbf{m}_{\bar{x}}} x(t) + \mathbf{e}_{\bar{x}}(x(t)) \right) - \left(\frac{1}{\omega} - \gamma(t) \right) x(t) + \frac{1}{\omega} y(t) & = 0 \\ \dot{y}(t) - \left(\frac{1}{\omega} - \gamma(t) - \dot{\gamma}(t)\omega \right) x(t) + \frac{1}{\omega} y(t) & = 0, \end{cases} \quad (36)$$

8 with initial conditions $x(t_0) = x_0, y(t_0) = -\omega(v_0 + \omega \nabla G_{\mathbf{m}_{\bar{x}}}(x_0)) + (1 - \omega\gamma(t_0))x_0 - \omega^2 \mathbf{e}_{\bar{x}}(x_0)$.

9 *Proof.* Let (x, y) be a solution of [\(36\)](#). By differentiation of the first equation in [\(36\)](#), we get

$$\ddot{x}(t) + \omega \nabla^2 G_{\mathbf{m}_{\bar{x}}}(x(t)) \dot{x}(t) + \dot{\gamma}(t)x(t) - \left(\frac{1}{\omega} - \gamma(t) \right) \dot{x}(t) + \frac{1}{\omega} \dot{y}(t) = 0.$$

Replacing \dot{y} by its expression from the second equation in [\(36\)](#), we obtain:

$$\begin{aligned} & \ddot{x}(t) + \omega \nabla^2 G_{\mathbf{m}_{\bar{x}}}(x(t)) \dot{x}(t) + \dot{\gamma}(t)x(t) - \left(\frac{1}{\omega} - \gamma(t) \right) \dot{x}(t) \\ & + \frac{1}{\omega} \left(\left(\frac{1}{\omega} - \gamma(t) - \dot{\gamma}(t)\omega \right) x(t) - \frac{1}{\omega} y(t) \right) = 0, \end{aligned} \quad (37)$$

and using again the first equation in [\(36\)](#) to eliminate $y(t)$, we get:

$$\begin{aligned} & \ddot{x}(t) + \omega \nabla^2 G_{\mathbf{m}_{\bar{x}}}(x(t)) \dot{x}(t) + \dot{\gamma}(t)x(t) - \left(\frac{1}{\omega} - \gamma(t) \right) \dot{x}(t) + \frac{1}{\omega} \left(\left(\frac{1}{\omega} - \gamma(t) \right) \right. \\ & \left. - \dot{\gamma}(t)\omega \right) x(t) + \dot{x}(t) + \omega \left(\nabla G_{\mathbf{m}_{\bar{x}}} x(t) + \mathbf{e}_{\bar{x}}(x(t)) \right) - \left(\frac{1}{\omega} - \gamma(t) \right) x(t) = 0, \end{aligned} \quad (38)$$

10 and after simplifications, we recover $(\text{ISEHD}_{\mathbf{m}, \gamma, \omega})$. Conversely, let x be a trajectory solution
 11 to $(\text{ISEHD}_{\mathbf{m}, \gamma, \omega})$ with initial conditions $(x_0, v_0) \in \mathcal{H} \times \mathcal{H}$ and define

$$y(t) = -\omega \left(\dot{x}(t) + \omega \left(\nabla G_{\mathbf{m}_{\bar{x}}} x(t) + \mathbf{e}_{\bar{x}}(x(t)) \right) - \left(\frac{1}{\omega} - \gamma(t) \right) x(t) \right).$$

12 By differentiating the previous formula and using $(\text{ISEHD}_{\mathbf{m}, \gamma, \omega})$, we recover the second
 13 equation of [\(36\)](#), as desired. ■

14 4.1.2. *Well-posedness.* Thanks to [Proposition 2](#), one can also consider [\(36\)](#) in the case where
 15 $G_{\mathbf{m}_{\bar{x}}} \in \Gamma_0(\mathcal{H})$. In fact, we have the following

16 **Definition 6.** Suppose that $\gamma(t) \geq 0, \omega > 0$. For any initial conditions $(x_0, v_0) \in \overline{\text{dom}(G_{\mathbf{m}_{\bar{x}}})} \times$
 17 \mathcal{H} , the dynamics $(\text{ISEHD}_{\mathbf{m}, \gamma, \omega})$ admits an equivalent formulation of the form

$$\begin{cases} \dot{x}(t) + \omega \left(\nabla G_{\mathbf{m}_{\bar{x}}} x(t) + \mathbf{e}_{\bar{x}}(x(t)) \right) - \left(\frac{1}{\omega} - \gamma(t) \right) x(t) + \frac{1}{\omega} y(t) & \ni 0 \\ \dot{y}(t) - \left(\frac{1}{\omega} - \gamma(t) - \dot{\gamma}(t)\omega \right) x(t) + \frac{1}{\omega} y(t) & = 0, \end{cases} \quad (39)$$

18 with initial conditions $x(t_0) = x_0, y(t_0) = v_0$.

1 One of the main advantages of (39) is that it can be easily recast as a monotone inclusion
 2 governed by Lipschitz perturbation of a maximal monotone operator on the product space
 3 $\mathcal{H} \times \mathcal{H}$. Indeed, setting $Z(t) = (x(t), y(t))$, $\mathcal{A}(x, y) = (\omega \partial G_{m_{\bar{x}}}(x), 0)$ and

$$\mathcal{E}(t, x, y) = \left(\omega \mathbf{e}_{\bar{x}}(x(t)) - \left(\frac{1}{\omega} - \gamma(t) \right) x(t) + \frac{1}{\omega} y(t), - \left(\frac{1}{\omega} - \gamma(t) - \dot{\gamma}(t) \omega \right) x(t) + \frac{1}{\omega} y(t) \right),$$

4 we immediately see that (39) can be written as

$$\dot{Z}(t) + \mathcal{A}(Z(t)) + \mathcal{E}(t, Z(t)) \ni 0_{\mathcal{H} \times \mathcal{H}}, \quad Z(0) = (x_0, v_0). \quad (40)$$

5 Then, (39) can be written as monotone inclusion in the product space $\mathcal{H} \times \mathcal{H}$

$$\begin{cases} \dot{Z}(t) + \mathcal{A}(Z(t)) + \mathcal{E}(t, Z(t)) \ni 0_{\mathcal{H} \times \mathcal{H}}, & \text{a.e } t \in [t_0, T] \\ Z(0) = (x_0, v_0) \in \mathcal{H} \times \mathcal{H}, \end{cases} \quad (\text{MIS})$$

6 which fits in the framework of Lipschitz perturbations of maximal monotone operators as in
 7 Section 3. Notice that this formulation is different from the classical Hamiltonian one.

8 Before stating the main result, let us recall that we endow the product space with the
 9 scalar product $\langle (u, v), (u^*, v^*) \rangle_{\mathcal{H} \times \mathcal{H}} = \langle u, u^* \rangle + \langle v, v^* \rangle$, and the induced norm $\|(u, v)\|_{\mathcal{H} \times \mathcal{H}} =$
 10 $\sqrt{\|u\|^2 + \|v\|^2}$. We have the following auxiliary results

11 **Lemma 6.** *Under Assumption 3, the operator \mathcal{E} is Lipschitz continuous.*

12 *Proof.* Let $u, v, u^*, v^* \in \mathcal{H}$ and set $p = \left(\frac{1}{\omega} - \gamma(t) \right), q = \left(\frac{1}{\omega} - \gamma(t) - \dot{\gamma}(t) \omega \right)$. We have, for
 13 $t \in [t_0, T]$

$$\begin{aligned} \|\mathcal{E}(t, u, v) - \mathcal{E}(t, u^*, v^*)\|_{\mathcal{H} \times \mathcal{H}} &= \left\| \left(\omega (\mathbf{e}_{\bar{x}}(u) - \mathbf{e}_{\bar{x}}(u^*)) + p(u^* - u) + \frac{1}{\omega} (v - v^*), \frac{1}{\omega} (v - v^*) + q(u^* - u) \right) \right\|_{\mathcal{H} \times \mathcal{H}} \\ &= \sqrt{\|(\omega (\mathbf{e}_{\bar{x}}(u) - \mathbf{e}_{\bar{x}}(u^*)) + p(u^* - u) + \frac{1}{\omega} (v - v^*))\|^2 + \|\frac{1}{\omega} (v - v^*) + q(u^* - u)\|^2} \\ &= \sqrt{2\omega^2 \|\mathbf{e}_{\bar{x}}(u) - \mathbf{e}_{\bar{x}}(u^*)\|^2 + (4p^2 + 2p^2) \|u - u^*\|^2 + \frac{6}{\omega^2} \|v - v^*\|^2} \\ &= \sqrt{(2\omega^2 \beta^2 (2 + \tau)^2 + 4p^2 + 2p^2) \|u - u^*\|^2 + \frac{6}{\omega^2} \|v - v^*\|^2} \\ &\leq \left(\sqrt{2} \omega \beta (2 + \tau) + 2|p| + \sqrt{2}|q| + \frac{\sqrt{6}}{\omega} \right) \|(u, v) - (u^*, v^*)\|_{\mathcal{H} \times \mathcal{H}} \\ &= K(\omega, \beta, \tau, \gamma) \|(u, v) - (u^*, v^*)\|_{\mathcal{H} \times \mathcal{H}} \end{aligned}$$

14 where we have used Young's inequality and Lemma 5 for the Lipschitz continuity of the
 15 operator $\mathbf{e}_{\bar{x}}$. ■

16 **Proposition 3.** *Assume that Assumptions 3 and 4 hold. Then, for any initial data $x_0 \in$
 17 $\overline{\text{dom}(G_{m_{\bar{x}}})}$ and $v_0 \in \mathcal{H}$, there exists a unique global solution $x \in C^2([t_0, \infty[, \mathcal{H})$ to (MIS)
 18 such that $x(t_0) = x_0$ and $\dot{x}(t_0) = v_0$. Moreover, the solution $Z = (x, y)$ satisfies the following
 19 properties*

- 20 (i) y is continuously differentiable on $[t_0, +\infty[$, and $\dot{y}(t) - \left(\frac{1}{\omega} - \gamma(t) - \omega \dot{\gamma}(t) \right) x(t) +$
 21 $\frac{1}{\omega} y(t) = 0$, for all $t \geq t_0$;
- 22 (ii) x is absolutely continuous on $[t_0, T]$ and $\dot{x} \in L^2(t_0, T; \mathcal{H})$ for all $T > t_0$;
- 23 (iii) $x(t) \in \text{dom}(\partial G_{m_{\bar{x}}})$ for all $t > t_0$;
- 24 (iv) x is Lipschitz continuous on any compact subinterval of $]t_0, +\infty[$;
- 25 (v) the function $t \in [t_0, +\infty[\mapsto G_{m_{\bar{x}}}(x(t))$ is absolutely continuous on $[t_0, T]$ for all
 26 $T > t_0$;

- 1 (vi) there exists a function $\xi : [t_0, +\infty[\rightarrow \mathcal{H}$ such that
 2 (a) $\xi(t) \in \partial G_{m_{\bar{x}}}(x(t))$ for all $t > t_0$;
 3 (b) $\dot{x}(t) + \omega \xi(t) - \left(\frac{1}{\omega} - \gamma(t)\right)x(t) + \frac{1}{\omega}y(t) = 0$ for almost every $t > t_0$;
 4 (c) $\xi \in L^2(t_0, T; \mathcal{H})$ for all $T > t_0$;
 5 (d) $\frac{d}{dt}\phi(x(t)) = \langle \xi(t), \dot{x}(t) \rangle$ for almost every $t > t_0$.

6 *Proof.* Thanks to [Assumption 4](#) and Moreau's theorem (see, e.g., [\[23, Theorem 20.25\]](#)) \mathcal{A}
 7 is maximal monotone and \mathcal{E} is Lipschitz continuous thanks to [Lemma 6](#). We conclude
 8 again the existence of a strong global solution $Z = (x, y) : [t_0, T] \rightarrow \mathcal{H} \times \mathcal{H}$ of [\(MIS\)](#) via
 9 [\[27, Proposition 3.12\]](#). The verification of items [\(i\)-\(vi\)](#) can be done by following the main
 10 arguments of [\[19, Theorem 4.4\]](#) \blacksquare

11 As a consequence, we have the following

12 **Corollary 3.** *Suppose that $G_{m_{\bar{x}}} : \mathcal{H} \rightarrow \mathbb{R}$ is a convex C^2 function. For any $t_0 > 0$, and any
 13 Cauchy data $(x_0, v_0) \in \mathcal{H} \times \mathcal{H}$, there exists a unique classical global solution $x : [t_0, +\infty[\rightarrow \mathcal{H}$
 14 to*

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla G_{m_{\bar{x}}}(x(t)) + \omega \nabla^2 G_{m_{\bar{x}}}(x(t))\dot{x}(t) + \mathbf{e}_{\bar{x}}(x(t)) = 0, \quad (\text{ISEHD}_{m, \gamma, \omega})$$

15 with $x(t_0) = x_0, \dot{x}(t_0) = v_0$.

16 *Proof.* We have thanks to [\(36\)](#) that

$$\dot{x}(t) + \omega \left(\nabla G_{m_{\bar{x}}}(x(t)) + \mathbf{e}_{\bar{x}}(x(t)) \right) - \left(\frac{1}{\omega} - \gamma(t) \right) x(t) + \frac{1}{\omega} y(t) = 0,$$

17 so thanks to [Proposition 3 Item \(i\)-Item \(ii\)](#), we have $\dot{x} \in C^1([t_0, \infty))$, and thus $x \in$
 18 $C^2([t_0, \infty))$. We conclude by combining [Propositions 2](#) and [3](#) with $y(t_0) = -\omega(v_0 + \omega \nabla G_{m_{\bar{x}}}(x_0)) +$
 19 $(1 - \omega\gamma(t_0))x_0 - \omega^2 \mathbf{e}_{\bar{x}}(x_0)$. \blacksquare

20 **4.2. Convergence properties.** Now let us examine the convergence properties of [\(ISEHD\)_{m, \gamma, \omega}](#).
 21 Recall that we restrict ourselves here to the smooth case, that is, we assume that $F_{m_{\bar{x}}} = \nabla G_{m_{\bar{x}}}$
 22 satisfies [Assumption 3](#), i.e., $G_{m_{\bar{x}}}$ is μ -strongly monotone for $\mu > 0$. Moreover, we tune the
 23 viscous damping function to the modulus of strong convexity μ of $G_{m_{\bar{x}}}$, by taking $\gamma(t) \equiv 2\sqrt{\mu}$.
 24 From now on, we focus on the following system

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla G_{m_{\bar{x}}}(x(t)) + \omega \nabla^2 G_{m_{\bar{x}}}(x(t))\dot{x}(t) + \mathbf{e}_{\bar{x}}(x(t)) = 0, \quad (\text{ISEHD}_{m, 2\sqrt{\mu}})$$

25 To perform Lyapunov analysis, let us define the following function $\mathcal{V} : [t_0, \infty[\rightarrow \mathbb{R}^+$ by

$$\mathcal{V}(t) := G_{m_{\bar{x}}}(x(t)) - G_{m_{\bar{x}}}^* + \frac{1}{2} \|v(t)\|^2, \quad \text{where } v(t) := \sqrt{\mu}(x(t) - \bar{x}) + \dot{x}(t) + \omega \nabla G_{m_{\bar{x}}}(x(t)).$$

26 Following the main ideas of [\[9, 14\]](#) we prove the following result.

27 **Theorem 3.** *Assume that $G_{m_{\bar{x}}}$ satisfies [Assumption 3](#) and let $x : [t_0, \infty[\rightarrow \mathbb{R}$ be the solution
 28 of [\(ISEHD\)_{m, 2\sqrt{\mu}}](#). Suppose that $0 \leq \omega \leq \frac{1}{2\sqrt{\mu}}$ and that ρ and the damping coefficient ω
 29 satisfy*

$$8\rho^2 + \omega < 1. \quad (41)$$

30 We then have:

- 31 • for all $t \geq t_0$

$$\mathcal{V}(t) \leq C e^{-\frac{\sqrt{\mu}}{4}t}, \quad \text{where } C = \mathcal{V}(t_0) e^{\frac{\sqrt{\mu}}{4}t_0}.$$

1 *In particular*

$$\frac{\mu}{2} \|x(t) - \bar{x}\|^2 \leq G_{\mathbf{m}_{\bar{x}}}(x(t)) - G_{\mathbf{m}_{\bar{x}}}^* \leq \mathcal{V}(t_0) e^{-\frac{\sqrt{\mu}}{4}(t-t_0)}.$$

2 • *There exists $C_1 > 0$ such that,*

$$e^{\sqrt{\mu}t} \int_{t_0}^t e^{\sqrt{\mu}s} \|\nabla G_{\mathbf{m}_{\bar{x}}}(x(s))\|^2 ds \leq C_1 e^{-\frac{\sqrt{\mu}}{4}t}, \quad \forall t \geq t_0.$$

3 *Proof.* We have

$$\begin{aligned} \dot{\mathcal{V}}(t) &= \langle \nabla G_{\mathbf{m}_{\bar{x}}}(x(t)), \dot{x}(t) \rangle + \langle v(t), \sqrt{\mu}\dot{x}(t) + \ddot{x}(t) + \omega \nabla G_{\mathbf{m}_{\bar{x}}}(x(t))\dot{x}(t) \rangle \\ &= \langle \nabla G_{\mathbf{m}_{\bar{x}}}(x(t)), \dot{x}(t) \rangle + \langle v(t), -\sqrt{\mu}\dot{x}(t) - \nabla G_{\mathbf{m}_{\bar{x}}}(x(t)) - \mathbf{e}_{\bar{x}}(x(t)) \rangle. \end{aligned} \quad (42)$$

We get from (ISEHD_{m,2√μ}) after some simplifications

$$\begin{aligned} \dot{\mathcal{V}}(t) + \mu \langle \dot{x}(t), x(t) - \bar{x} \rangle + \sqrt{\mu} \|\dot{x}(t)\|^2 + \sqrt{\mu} \langle \nabla G_{\mathbf{m}_{\bar{x}}}(x(t)), x(t) - \bar{x} \rangle \\ + \omega \sqrt{\mu} \langle \nabla G_{\mathbf{m}_{\bar{x}}}(x(t)), \dot{x}(t) \rangle + \omega \|\nabla G_{\mathbf{m}_{\bar{x}}}(x(t))\|^2 = -\langle v(t), \mathbf{e}_{\bar{x}}(x(t)) \rangle. \end{aligned} \quad (43)$$

4 Using μ -strong convexity of $G_{\mathbf{m}_{\bar{x}}}$, we have

$$\langle \nabla G_{\mathbf{m}_{\bar{x}}}(x(t)), x(t) - \bar{x} \rangle \geq G_{\mathbf{m}_{\bar{x}}}(x(t)) - G_{\mathbf{m}_{\bar{x}}}^* + \frac{\mu}{2} \|x(t) - \bar{x}\|^2, \quad (44)$$

5 and using this in (76), we get

$$\dot{\mathcal{V}}(t) + \sqrt{\mu} \Theta(t) \leq \|v(t)\| \|\mathbf{e}_{\bar{x}}(x(t))\|, \quad (45)$$

6 where

$$\Theta(t) := G_{\mathbf{m}_{\bar{x}}}(x(t)) - G_{\mathbf{m}_{\bar{x}}}^* + \frac{\mu}{2} \|x(t) - \bar{x}\|^2 + \sqrt{\mu} \langle \dot{x}(t), x(t) - \bar{x} \rangle + \|\dot{x}(t)\|^2 + \omega \langle \nabla G_{\mathbf{m}_{\bar{x}}}(x(t)), \dot{x}(t) \rangle + \frac{\omega}{\sqrt{\mu}} \|\nabla G_{\mathbf{m}_{\bar{x}}}(x(t))\|^2.$$

7 Using the definition of $\mathcal{V}(t)$, we may rewrite $\Theta(t)$ as

$$\Theta(t) = \mathcal{V}(t) + \frac{1}{2} \|\dot{x}(t)\|^2 - \omega \sqrt{\mu} \langle \nabla G_{\mathbf{m}_{\bar{x}}}(x(t)), x(t) - \bar{x} \rangle + \left(\frac{\omega}{\sqrt{\mu}} - \frac{\omega^2}{2} \right) \|\nabla G_{\mathbf{m}_{\bar{x}}}(x(t))\|^2.$$

8 Consequently, (45) becomes

$$\dot{\mathcal{V}}(t) + \sqrt{\mu} \mathcal{V}(t) + \frac{\sqrt{\mu}}{2} \|\dot{x}(t)\|^2 + \left(\left(\frac{\omega}{\sqrt{\mu}} - \frac{\omega^2}{2} \right) \|\nabla G_{\mathbf{m}_{\bar{x}}}(x(t))\|^2 - \omega \sqrt{\mu} \langle \nabla G_{\mathbf{m}_{\bar{x}}}(x(t)), x(t) - \bar{x} \rangle \right) \leq \|v(t)\| \|\mathbf{e}_{\bar{x}}(x(t))\|,$$

9 Using strong convexity we obtain again

$$\mathcal{V}(t) = \frac{1}{2} \mathcal{V}(t) + \frac{1}{2} \mathcal{V}(t) \geq \frac{1}{2} \mathcal{V}(t) + \frac{\mu}{4} \|x(t) - \bar{x}\|^2,$$

and observing that $\frac{\omega}{2\sqrt{\mu}} \leq \frac{\omega}{\sqrt{\mu}} - \frac{\omega^2}{2}$ for $0 \leq \omega \leq \frac{1}{\sqrt{\mu}}$. We end up with

$$\begin{aligned} \dot{\mathcal{V}}(t) + \frac{\sqrt{\mu}}{2} \mathcal{V}(t) + \frac{\sqrt{\mu}}{2} \|\dot{x}(t)\|^2 \\ + \left(\frac{\mu}{4} \|x(t) - \bar{x}\|^2 + \frac{\omega}{2\sqrt{\mu}} \|\nabla G_{\mathbf{m}_{\bar{x}}}(x(t))\|^2 - \omega \sqrt{\mu} \|\nabla G_{\mathbf{m}_{\bar{x}}}(x(t))\| \|x(t) - \bar{x}\| \right) \leq \|v(t)\| \|\mathbf{e}_{\bar{x}}(x(t))\|. \end{aligned} \quad (46)$$

10 Now let us treat the right hand side of this inequality. Since $\mathcal{V}(t) \geq \frac{1}{2} \|v(t)\|^2$, we have, using
11 Young's inequality

$$\|v(t)\| \|\mathbf{e}_{\bar{x}}(x(t))\| \leq \frac{\sqrt{\mu}}{8} \|v(t)\|^2 + \frac{2}{\sqrt{\mu}} \|\mathbf{e}_{\bar{x}}(x(t))\|^2 \leq \frac{\sqrt{\mu}}{4} \mathcal{V}(t) + \frac{2\beta^2 \tau^2}{\sqrt{\mu}} \|x(t) - \bar{x}\|^2. \quad (47)$$

1 Using Lemma 5, we get after rearranging the terms

$$\dot{\mathcal{V}}(t) + \frac{\sqrt{\mu}}{4}\mathcal{V}(t) + \frac{\sqrt{\mu}}{2}\|\dot{x}(t)\|^2 + \sqrt{\mu}\Psi(t) \leq 0, \quad (48)$$

where

$$\Psi(t) = \left(\frac{\mu}{4} - \frac{2\beta^2\tau^2}{\mu}\right)\|x(t) - \bar{x}\|^2 + \frac{\omega}{2\sqrt{\mu}}\|\nabla G_{\mathbf{m}_{\bar{x}}}(x(t))\|^2 - \omega\sqrt{\mu}\|\nabla G_{\mathbf{m}_{\bar{x}}}(x(t))\|\|x(t) - \bar{x}\|.$$

2 Setting

$$\mathbf{a} = \frac{\mu}{4} - \frac{2\beta^2\tau^2}{\mu}, \mathbf{b} = \frac{\omega}{2\sqrt{\mu}}, \mathbf{c} = -\frac{\omega\sqrt{\mu}}{2}$$

3 and $X = \|x(t) - \bar{x}\|$ and $Y = \|\nabla f_{\mathbf{m}_{\bar{x}}}(x(t))\|$, we see that Ψ can be written as a quadratic
4 form $\mathcal{Q} : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$ with $\mathcal{Q}(X, Y) = \mathbf{a}\|X\|^2 + 2\mathbf{c}\langle X, Y \rangle + \mathbf{d}\|Y\|^2$. By assumption $\mathbf{a}, \mathbf{b} \geq 0$,
5 and the discriminant $\mathbf{c}^2 - \mathbf{a}\mathbf{b}$ of \mathcal{Q} is nonpositive. Indeed, since $0 \leq \omega \leq \frac{1}{2\sqrt{\mu}}$

$$\mathbf{c}^2 - \mathbf{a}\mathbf{b} = \frac{\omega^2\mu}{4} - \frac{\omega}{2\sqrt{\mu}}\left(\frac{\mu}{4} - \frac{2\beta^2\tau^2}{\mu}\right) \leq \frac{\omega^2\sqrt{\mu}}{8} - \frac{\omega\sqrt{\mu}}{2}\left(\frac{1}{4} - 2\rho^2\right) = \frac{\omega\sqrt{\mu}}{8}(\omega - 1 + 8\rho^2) \leq 0.$$

6 Hence, $\Psi(t) \geq 0$ and

$$\dot{\mathcal{V}}(t) + \frac{\sqrt{\mu}}{4}\mathcal{V}(t) + \frac{\sqrt{\mu}}{2}\|\dot{x}(t)\|^2 \leq 0,$$

7 which gives after integration

$$\mathcal{V}(t) \leq \mathcal{V}(t_0)e^{-\frac{\sqrt{\mu}}{4}(t-t_0)}. \quad (49)$$

8 Therefore, $\lim_{t \rightarrow \infty} \mathcal{V}(t) = 0$ and in particular

$$\lim_{t \rightarrow \infty} G_{\mathbf{m}_{\bar{x}}}(x(t)) - G_{\mathbf{m}_{\bar{x}}}^* = 0 \text{ and } \lim_{t \rightarrow \infty} \|v(t)\| = 0. \quad (50)$$

9 This implies, using strong convexity of $G_{\mathbf{m}_{\bar{x}}}$

$$\lim_{t \rightarrow \infty} \|x(t) - \bar{x}\| = 0,$$

10 which gives that

$$\lim_{t \rightarrow \infty} \|\nabla G_{\mathbf{m}_{\bar{x}}}(x(t))\| = 0.$$

11 We deduce from (50) that $\lim_{t \rightarrow \infty} \|\dot{x}(t)\| = 0$.

12 Coming back to (49), we have, by definition of \mathcal{V} , that

$$G_{\mathbf{m}_{\bar{x}}}(x(t)) - G_{\mathbf{m}_{\bar{x}}}^* \leq \mathcal{V}(t_0)e^{-\frac{\sqrt{\mu}}{4}(t-t_0)} \text{ and } \|v(t)\|^2 \leq 2\mathcal{V}(t_0)e^{-\frac{\sqrt{\mu}}{4}(t-t_0)}. \quad (51)$$

Developing in (51) we have

$$\begin{aligned} & \mu\|x(t) - \bar{x}\|^2 + \|\dot{x}(t)\|^2 + \omega^2\|\nabla G_{\mathbf{m}_{\bar{x}}}(x(t))\|^2 + 2\omega\sqrt{\mu}\langle \nabla G_{\mathbf{m}_{\bar{x}}}(x(t)), x(t) - \bar{x} \rangle \\ & + 2\omega\langle \nabla G_{\mathbf{m}_{\bar{x}}}(x(t)), \dot{x}(t) \rangle + 2\sqrt{\mu}\langle \dot{x}(t), x(t) - \bar{x} \rangle \leq Ce^{-\frac{\sqrt{\mu}}{4}t}, \end{aligned} \quad (52)$$

13 where $C = 2\mathcal{V}(t_0)e^{\frac{\sqrt{\mu}}{4}t_0}$.

14 Since $\langle \nabla G_{\mathbf{m}_{\bar{x}}}(x(t)), x(t) - \bar{x} \rangle \geq G_{\mathbf{m}_{\bar{x}}}(x(t)) - G_{\mathbf{m}_{\bar{x}}}^*$, we deduce from (52) that

$$\dot{U}(t) + \sqrt{\mu}U(t) + \omega^2\|\nabla G_{\mathbf{m}_{\bar{x}}}(x(t))\|^2 \leq Ce^{-\frac{\sqrt{\mu}}{4}t}, \quad (53)$$

15 where $U(t) := \sqrt{\mu}\|x(t) - \bar{x}\|^2 + 2\omega(G_{\mathbf{m}_{\bar{x}}}(x(t)) - G_{\mathbf{m}_{\bar{x}}}^*)$. Integrating (53), we obtain after
16 elementary computation

$$e^{\sqrt{\mu}t} \int_{t_0}^t e^{-\sqrt{\mu}s} \|\nabla G_{\mathbf{m}_{\bar{x}}}(x(s))\|^2 ds \leq C_1 e^{-\frac{\sqrt{\mu}}{4}t},$$

17 as desired. ■

1 *Remark 4.1.* Notice that contrary to [17], no assumption on the integrability of $\mathbf{e}_{\bar{x}}$ is made
 2 in [Theorem 3](#). In fact, thanks to [Lemma 4](#), the norm of the error term $\mathbf{e}_{\bar{x}}$ can be absorbed
 3 in the righthandside of (45).

4 *Remark 4.2.* Notice that in case $\omega = 0$, i.e., when the inertial dynamic is considered only
 5 with the viscous damping coefficient, the condition (41) reduces to $\rho < \frac{\sqrt{2}}{4}$. This contrasts
 6 with the convergence condition for first-order dynamics (cf. [Theorem 2](#)), where convergence
 7 is ensured in parameter regime $\rho < 1$. One possible explanation for this difference, is the
 8 potential occurrence of oscillations, which may necessitate a stricter compatibility condition
 9 between the parameters τ, β, μ .

10 We end this section with a similar result to [Corollary 2](#), which is a direct consequence of
 11 [Assumption 2](#) and [Theorem 3-3](#).

12 **Corollary 4.** *Let $x : [t_0, \infty) \rightarrow \mathbb{R}$ be the solution of (ISEHD $_{\mathbf{m}, 2\sqrt{\mu}}$), then $\forall t \geq t_0$*

$$\mathbb{W}_1(\mathbf{m}_{x(t)}, \mathbf{m}_{\bar{x}}) \leq C e^{-\frac{\sqrt{\mu}}{8}t},$$

13 *with $C = \tau \sqrt{\frac{2}{\mu} \mathcal{V}(t_0)} e^{\frac{\sqrt{\mu}}{8}t_0}$.*

14 5. ON COARSE RICCI CURVATURE

15 In this section we discuss some dynamical and geometrical properties of the family
 16 $(\mathbf{m}_x)_x$, particularly the notion of Ollivier-Ricci curvature and how it is tightly related to
 17 [Assumption 2](#). In fact, the family of probabilities $\mathbf{m} = (\mathbf{m}_x)_{x \in \mathcal{H}}$ and its Lipschitz behavior
 18 with respect to the \mathbb{W}_1 -Wasserstein distance, reveals that a natural setting to address
 19 monotone inclusions of the form (SMI), and thus stochastic optimization problems with
 20 decision-dependent distributions is the framework of *metric random walk spaces* (see, e.g.,
 21 [42, 45]). All definitions of this section can be found in [36, 42].

22 **5.1. Metric random walk spaces.** Before going further, let us recall the following definitions
 23 to introduce a couple of probabilistic notions.

24 **Definition 7** (Random walks [45]). Given a Polish space (X, d) . A family of probabilities
 25 $\mathbf{m} = (\mathbf{m}_x)_{x \in X}$ is a random walk on X if $\mathbf{m}_x \in \mathcal{P}(\Xi)$ for each $x \in X$ and

- 26 • \mathbf{m}_x depends measurably on $x \in X$,
- 27 • Each \mathbf{m}_x has finite first-order moment, i.e., for some $x^o \in X$, $\mathbb{E}_{y \sim \mathbf{m}_x} d(y, x^o) < \infty$.

28 Then (X, d) equipped with a random walk first-order \mathbf{m} is a metric random walk space
 29 (m.r.w.s for short), and we denote it by $[X, d, \mathbf{m}]$.

30 Let us recall the notion of invariant and ergodic measures.

31 **Definition 8** (Invariance). Let ν be a σ -finite measure on X and \mathbf{m} a random walk on (X, \mathcal{B}) .
 32 We say that ν is invariant with respect to \mathbf{m} if $\nu \star \mathbf{m} = \nu$, where $\nu \star \mathbf{m}$ is the convolution of ν
 33 by the random walk \mathbf{m} and is defined by

$$\nu \star \mathbf{m}(A) = \int_X \mathbf{m}_x(A) d\nu(x) \text{ for all } A \in \mathcal{B}.$$

34 As pointed out in [45], each measure \mathbf{m}_x can be seen as a replacement of a sphere around
 35 x . While in a probabilistic framework one thinks about a Markov chain whose transition
 36 kernel from x to y in n steps is defined by

$$d\mathbf{m}_x^{*n}(y) = \int_{z \in X} d\mathbf{m}_x^{*(n-1)}(z) d\mathbf{m}_z(y), \quad (54)$$

1 with $\mathfrak{m}_x^1 = \mathfrak{m}_x$ and $\mathfrak{m}_x^0 = \delta_x$.

2
3 In the sequel, we assume that (\mathcal{H}, d) is a separable real Hilbert space, and thus a Polish
4 space, where $d(x, y) = \langle x - y, x - y \rangle^{1/2}$.

5 **5.2. Feller Property.** Let us recall the following definition.

6 **Definition 9.** We say that $\mathfrak{m} := (\mathfrak{m}_x)_x$ has the weak-Feller property if and only if for every
7 sequence $x_n \rightarrow x^0 \in \mathcal{H}$ we have $\mathfrak{m}_{x_n} \rightarrow \mathfrak{m}_{x^0}$, i.e., $\int f d\mathfrak{m}_{x_n} \rightarrow \int f d\mathfrak{m}_{x^0}$ for any $f \in C_b(\mathcal{H})$.

8 It turns that [Assumption 2](#) implies directly that the family \mathfrak{m} is weak-Feller.

9 **Proposition 4.** Under [Assumption 2](#), \mathfrak{m} has the weak-Feller property. Moreover, for each
10 $x \in \mathcal{H}$, \mathfrak{m}_x has finite first-order moment.

11 *Proof.* Let $x^o \in \mathcal{H}$ and $(x_n)_n$ a sequence of \mathcal{H} such that $x_n \rightarrow x^o$ as $n \rightarrow \infty$. Then
12 [Assumption 2](#) gives

$$\mathbb{W}_1(\mathfrak{m}_{x_n}, \mathfrak{m}_{x^o}) \leq \tau \|x_n - x^o\|,$$

13 and thus $\lim_{n \rightarrow \infty} \mathbb{W}_1(\mathfrak{m}_{x_n}, \mathfrak{m}_{x^o}) = 0$. Thanks to [[4](#), Proposition 7.1.5], (\mathfrak{m}_{x_n}) has uniformly
14 integrable p -moments with $p \geq 1$ and narrowly converges towards \mathfrak{m}_{x^o} . In particular \mathfrak{m} is
15 weak-Feller and each \mathfrak{m}_x has finite first-order moments. ■

16 *Remark 5.1.* We already know that $\mathfrak{m}_x \in \mathcal{P}(\Xi)$ for each $x \in \mathcal{H}$ and that $x \mapsto \mathfrak{m}_x(C)$ is
17 measurable for each $C \in \mathcal{B}$. Moreover, thanks to [Proposition 4](#), we have finiteness of first-
18 order moments of each \mathfrak{m}_x , so that the family \mathfrak{m} satisfies the requirements of [Definition 7](#).
19 This shows that a natural setting to address dynamics of the form [\(SMI\)](#) is the metric
20 random walk space $[\mathcal{H}, d, \mathfrak{m}]$. Many diffusion and variational problems has been studies
21 within this framework, with allows in particular consider nonlocal continuum problems or
22 problems on weighted graphs (see, e.g., [[42](#)] and the references therein).

23 *Remark 5.2.* Let us point out that if ν is an invariant measure with respect to \mathfrak{m} then it is
24 also and invariant measure with respect to \mathfrak{m}^{*n} for every $n \in \mathbb{N}$, where \mathfrak{m}^{*n} is the n -step
25 transition probability function given by [\(54\)](#). It turns out that weak-Feller property implies
26 that every weak- $*$ limit ν of $(\mathfrak{m}^{*n})_n$ is an invariant measure of \mathfrak{m} cf. [[36](#), Proposition
27 7.2.2] (see also [[34](#), Proposition 12.3.4]). However, without assuming at first the existence
28 of an invariant measure with respect to \mathfrak{m} , the measure ν may be trivial. Without further
29 compactness assumptions on the metric space (see, e.g., [[36](#), Theorem 7.2.3]) one needs some
30 Lyapunov like condition to ensure the existence of an invariant measure ν of the weak-Feller
31 family \mathfrak{m} (see, e.g., [[36](#), Theorem 7.2.4] or [[34](#), Theorem 12.3.3]). As we will see in [Corollary 6](#),
32 another way to obtain the existence of invariant measures is having a positive lower bound
33 on the coarse Ricci curvature of $[\mathcal{H}, d, \mathfrak{m}]$.

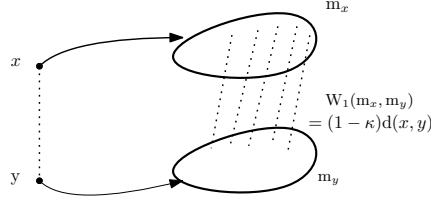
34 **5.3. Ollivier-Ricci curvature.** Let us discuss here the connexion between [Assumption 2](#) and
35 the so-called coarse or Ollivier-Ricci curvature (ORC for short). The results can be found in
36 [[45](#)] or [[46](#)]. A more recent presentation can be found in [[42](#)].

37 **Definition 10** (Ollivier-Ricci curvature [[45](#)]). Let $[\mathcal{H}, d, \mathfrak{m}]$ be a m.r.w.s. Then, for any distinct
38 points $x, y \in \mathcal{H}$, the ORC along (x, y) is defined as:

$$\kappa_{\mathfrak{m}}(x, y) = 1 - \frac{\mathbb{W}_1(\mathfrak{m}_x, \mathfrak{m}_y)}{d(x, y)}, \quad (55)$$

39 The ORC of $[\mathcal{H}, d, \mathfrak{m}]$ is defined as

$$\kappa_{\mathfrak{m}} := \inf_{x \neq y} \kappa_{\mathfrak{m}}(x, y). \quad (56)$$

**Figure 1:** Illustration of the ORC.

1 We clearly see from (55) that $\kappa_m(x, y) \leq 1$. Moreover, rearranging the terms, we have

$$\mathbb{W}_1(\mathbf{m}_x, \mathbf{m}_y) = (1 - \kappa_m(x, y)) d(x, y). \quad (57)$$

2 Consequently, having some lower bound $\kappa_m(x, y) \geq c \in \mathbb{R}$ for any $x, y \in \mathcal{H}$ gives

$$\mathbb{W}_1(\mathbf{m}_x, \mathbf{m}_y) \leq (1 - c)d(x, y), \quad (58)$$

3 which describes a Lipschitz behavior of the random walk \mathbf{m} . This has to be compared to
4 [Assumption 2](#). Indeed, we see from [Assumption 2](#) that, for $x \neq y$

$$1 - \tau \leq \kappa_m(x, y) \leq 1, \quad (59)$$

so according to the values of τ we have different regimes on the ORC κ_m (cf. [Table 1](#)).

Table 1: Relation between the values of τ and κ_m .

Values of τ	0	< 1	≤ 1
Values of κ_m	1	$(0, 1]$	$[0, 1]$

5 Notice that [Assumption 2](#) excludes both the cases $\kappa_m \equiv 1$ and $\kappa_m < 0$. Typically, $\tau = 0$,
6 would give that $\kappa_m = 1$ in other words $\mathbb{W}_1(\mathbf{m}_x, \mathbf{m}_y) = 0$ for any $x, y \in \mathcal{H}$, i.e., the distribution
7 \mathbf{m} is constant.

8 Moreover, it turns that there is equivalence between the lower bound on κ_m in (59) and
9 the Lipschitz behavior (58). This is directly related to a \mathbb{W}_1 -contraction property cf. [[45](#),
10 [Proposition 20](#)].

11 **Proposition 5.** *Let \mathbf{m} be a random walk on (\mathcal{H}, d) and assume that \mathbf{m}_x has finite moment for
12 all $x \in \mathcal{H}$. Then*

$$\kappa_m(x, y) \geq c \in \mathbb{R}, \forall x \neq y \iff \mathbb{W}_1(\nu_1 \star \mathbf{m}, \nu_2 \star \mathbf{m}) \leq (1 - c)\mathbb{W}_1(\nu_1, \nu_2) \quad \forall \nu_1, \nu_2 \in \mathcal{P}_1(\Xi).$$

13 In view of [Proposition 5](#), taking $\nu_1 = \delta_x$ and $\nu_2 = \delta_y$ for $x \neq y$, and $c = 1 - \tau$, we get

$$\mathbb{W}_1(\mathbf{m}_x, \mathbf{m}_y) = \mathbb{W}_1(\delta_x \star \mathbf{m}, \delta_y \star \mathbf{m}) \leq \tau \mathbb{W}(\delta_x, \delta_y) = \tau d(x, y)$$

14 which is exactly [Assumption 2](#) since the above inequality is trivial for $x = y$.

15 In the case of positive curvature, this contraction result implies the existence of a unique
16 invariant measure for the random walk \mathbf{m} when the ORC is positive.

17 **Corollary 5** ([[42](#)]). *Assume that $\kappa_m(x, y) \geq c > 0$ for all $x \neq y$. Then, the random walk \mathbf{m}
18 has a unique invariant measure $\nu \in \mathcal{P}_1(\Xi)$. Moreover, for any $\nu \in \mathcal{P}_1(\Xi)$*

$$(1) \quad \mathbb{W}_1(\nu \star \mathbf{m}^{*n}, \nu) \leq (1 - c)^n \mathbb{W}_1(\nu, \nu), \quad \forall n \in \mathbb{N}.$$

$$(2) \quad \mathbb{W}_1(\mathbf{m}^{*n}, \nu) \leq \frac{(1 - c)^n}{c} \mathbb{W}_1(\delta_x, \nu), \quad \forall n \in \mathbb{N}, \forall x \in \mathcal{H}.$$

19 In our setting, the parameter regime $\tau < 1$ would define a positive ORC. Consequently,
20 we have the following

1 **Corollary 6.** *Assume that \mathfrak{m} satisfies [Assumption 2](#) with $\tau < 1$. Then, there exists a unique*
 2 *invariant measure $\nu \in \mathcal{P}_1(\Xi)$ with respect to \mathfrak{m} .*

3 6. APPLICATION: INERTIAL PRIMAL-DUAL ALGORITHM

4 This section is devoted to the application of the developed results in [Section 3](#) and
 5 [Section 4](#) to the following class of optimization problems

$$\min_x f_{\mathfrak{m}_x}(x) + g(x) + h(Kx). \quad (60)$$

6 In what follows, we make the following assumptions.

7 **Assumption 5.** (a) $f_{\mathfrak{m}_x}(x) = \mathbb{E}_{\xi \sim \mathfrak{m}_x} f(x, \xi)$, with $f(\cdot, \xi) \in C_{\beta_f}^{1,1}(\mathcal{H}) \cap \Gamma_0(\mathcal{H})$ and $f(\cdot, \xi)$ is
 8 μ -strongly convex for all $\xi \in \Xi$, and $\xi \in \Xi \mapsto f(x, \xi)$ is measurable for any $x \in \mathcal{H}$,
 9 (b) $h \in C_{\beta_h}^{1,1}(\mathcal{H}) \cap \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{H})$ and $K \in \mathcal{L}(\mathcal{H}, \mathcal{H})$,
 10 (c) $0 \in \text{sri}(K \text{ dom}(g) - \text{dom}(r))$,
 11 where μ, β_f, β_h are positive constants and $\text{sri}(K \text{ dom}(g) - \text{dom}(r))$ is the strong relative
 12 interior of $(K \text{ dom}(g) - \text{dom}(r))$ (see, e.g., [\[23, Definition 6.9\]](#)).

13 [Assumption 5-\(a\)](#) is to be compared to [Assumption 1](#). In particular, it ensure that for
 14 any measure $\nu \in \mathcal{P}(\Xi)$, f_ν is differentiable with $\nabla f_\nu(x) = \mathbb{E}_{\xi \sim \nu} \nabla f(x, \xi)$, moreover, strong
 15 convexity of $f(\cdot, \xi)$ for all $\xi \in \Xi$ implies strong convexity of $f_{\mathfrak{m}_x}$ for all $x \in \mathcal{H}$. While
 16 [Assumption 5\(b\)](#) and [\(c\)](#) are classical for strong duality as discussed below.
 17 Thanks to Fenchel-Rockafellar duality (see, e.g., [\[23, Chapter 15\]](#)) the dual problem of [\(60\)](#)
 18 reads

$$\min_y f_{\mathfrak{m}_x}^* \square g^*(-K^*y) + h^*(y), \quad (61)$$

19 where $f_{\mathfrak{m}_x}^* \square g^*$ is the infimal convolution of $f_{\mathfrak{m}_x}^*$ and g^* given by

$$f_{\mathfrak{m}_x}^* \square g^*(z) = (f_{\mathfrak{m}_x} + g)^*(z) = \inf_{y \in \mathcal{H}} f_{\mathfrak{m}_x}^*(y) + g^*(z - y),$$

20 and K^* is the adjoint operator of K . Notice that the dual problem [\(61\)](#) is difficult to tackle
 21 directly since $f_{\mathfrak{m}_x}^*$ cannot be computed explicitly due to the dependence of both the loss
 22 function f and the measure \mathfrak{m}_x on the state x .

23 Problems of the form [\(60\)](#) arise in many fields such as image and signal processing,
 24 machine learning and partial differential equations. In the deterministic case, i.e., f does
 25 not depend on the distribution \mathfrak{m} and $g \equiv 0$, such problems were studied in [\[30\]](#). Later
 26 on, extensions were addressed in several works (see, e.g., [\[31, 49, 53\]](#)). In [\[24\]](#), the authors
 27 studied a fully stochastic variant of [\(60\)](#), i.e., $f_{\mathfrak{m}_x}(x) = \mathbb{E}_{\xi \sim \mathfrak{m}} f(x, \xi)$, $g(x) = \mathbb{E}_{\xi \sim \mathfrak{m}} g(x, \xi)$
 28 and $K(x) = \mathbb{E}_{\xi \sim \mathfrak{m}} K(x)$ for some suitable functions f, g and K . Yet, the distribution \mathfrak{m} does
 29 not depend on the state x . To the best of our knowledge composite problems of the form
 30 [\(60\)](#) with state-dependent distributions have not been addressed in the literature.

31 **6.1. Formulation as a monotone inclusion.** As seen in [Section 3](#), the appropriate notion of
 32 solutions of [\(60\)](#) is that of equilibria. Thus, our aim is to find an equilibrium point \bar{x} , i.e., a
 33 solution of the static problem

$$\bar{x} \in \arg \min_x f_{\mathfrak{m}_x}(x) + g(x) + h(Kx). \quad (62)$$

34 Problems [\(62\)](#) and [\(61\)](#) can be written in an inf-sup form

$$\inf_{x \in \mathcal{H}} \sup_{y \in \mathcal{H}} f_{\mathfrak{m}_x}(x) + g(x) + \langle Kx, y \rangle - h^*(y), \quad (63)$$

1 and (\bar{x}, \bar{y}) is optimal for (63) if and only if the following optimality condition holds

$$\begin{cases} 0 \in K^*\bar{y} + \nabla f_{\mathbf{m}_{\bar{x}}}(\bar{x}) + \partial \mathbf{g}(\bar{x}) \\ 0 \in -K\bar{x} + \partial \mathbf{h}^*(\bar{y}). \end{cases} \quad (64)$$

2 Notice that the problem (63) is a saddle point problem

$$\inf_{x \in \mathcal{H}} \sup_{y \in \mathcal{H}} \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x, y), \quad (65)$$

3 of the Lagrangian functional: $\mathcal{L}_{\mathbf{m}_{\bar{x}}}: \text{dom}(\partial \mathbf{g}) \times \text{dom}(\partial \mathbf{h}^*) \rightarrow \mathbb{R}$ defined by

$$\mathcal{L}_{\mathbf{m}_{\bar{x}}}(x, y) = f_{\mathbf{m}_{\bar{x}}}(x) + \mathbf{g}(x) + \langle Kx, y \rangle - \mathbf{h}^*(y).$$

4 For any pair $(x_0, y_0) \in \text{dom}(\partial \mathbf{g}) \times \text{dom}(\partial \mathbf{h}^*)$ we define the following restrictions of the
5 Lagrangian functional

$$\mathcal{L}_{\mathbf{m}_{\bar{x}}}^{y_0}(x) = \begin{cases} \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x, y_0) & \text{if } x \in \text{dom}(\partial \mathbf{g}), \\ \infty & \text{otherwise} \end{cases}, \quad \mathcal{L}_{\mathbf{m}_{\bar{x}}}^{x_0}(x_0, y) = \begin{cases} \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x_0, y) & \text{if } y \in \text{dom}(\partial \mathbf{h}^*) \\ -\infty & \text{otherwise} \end{cases}.$$

6 Then, the optimality system (64) can be written as a monotone inclusion

$$0_{\mathcal{H} \times \mathcal{H}} \in \mathcal{T}_{\mathbf{m}_{\bar{x}}}(\bar{x}, \bar{y}), \quad (66)$$

7 where

$$\mathcal{T}_{\mathbf{m}_{\bar{x}}}(x, y) = (\partial \mathcal{L}_{\mathbf{m}_{\bar{x}}}^y(x), -\partial \mathcal{L}_{\mathbf{m}_{\bar{x}}}^x(y)). \quad (67)$$

8 Notice that $\mathcal{T}_{\mathbf{m}_{\bar{x}}}$ can be split into the sum of two operators: $\mathcal{T}_{\mathbf{m}_{\bar{x}}} = A + B_{\mathbf{m}_{\bar{x}}}$ where

$$A: (x, y) \mapsto (\partial \mathbf{g}(x) + K^*(y), \partial \mathbf{h}^*(y) - K(x)), \quad B_{\mathbf{m}_{\bar{x}}}: (x, y) \mapsto (\nabla f_{\mathbf{m}_{\bar{x}}}(x), 0). \quad (68)$$

9 *Remark 6.1.* Similar problems to (65) were addressed in [54] where the authors studied
10 problems of the form

$$\min_{x \in X} \max_{y \in Y} \mathbb{E}_{\xi \sim \mathbf{m}(\bar{x}, \bar{y})} \phi(x, y, \xi), \quad (69)$$

11 where X, Y are compact sets and ϕ is a convex-concave function that plays the role of
12 the Lagrangian in our case. Such problems fall into the scope of (65). Indeed, changing
13 the Lagrange functional in (65) to $\mathcal{L}_{\mathbf{m}(\bar{x}, \bar{y})}(x, y) + \delta_X(x) - \delta_Y(y)$, where δ_X and δ_Y are
14 the indicator functions of the sets X and Y respectively, and considering a distribution \mathbf{m}
15 depending on both the primal and dual variables (x, y) allows recovering (69).

16 First, let us recapitulate some facts in the following lemma.

17 **Lemma 7.** *Under Assumption 5(a)(b)(c), the following properties hold*

- 18 (i) *The operator A is maximally monotone,*
19 (ii) *The operator $B_{\mathbf{m}_{\bar{x}}}(\cdot, \cdot)$ is maximally monotone and $\beta_{\mathbf{f}}$ -Lipschitz continuous for any
20 $x \in \mathcal{H}$,*
21 (iii) *For any $y, z \in \mathcal{H}$*

$$\sup_{(x', y') \in \mathcal{H} \times \mathcal{H}} \|B_{\mathbf{m}_y}(x', y') - B_{\mathbf{m}_z}(x', y')\| \leq \beta_{\mathbf{f}} \tau \|y - z\|.$$

- 22 (iv) *The operator $\mathcal{T}_{\mathbf{m}_{\bar{x}}}$ is maximally monotone and $\tilde{\mu}$ -strongly monotone with $\tilde{\mu} =$
23 $\min(\mu, \beta_{\mathbf{h}}^{-1})$.*

24 *Proof.* First, we see that A is the sum of the maximally monotone operator $(x, y) \mapsto$
25 $(\partial \mathbf{g}(x), \partial \mathbf{h}^*(y))$ (see, e.g., [23, Theorem 21.2 and Proposition 20.23]) and a skew-symmetric
26 operator $(x, y) \mapsto (K^*(y), -K(x))$, which is maximally monotone (see, e.g., [23, Example

20.35]), this proves (i). The operator $B_{\mathbf{m}_{\bar{x}}}$ is the gradient of the function $\Phi(z) = \mathbf{f}_{\mathbf{m}_{\bar{x}}}(x)$ and thus is maximally monotone. Now take $(x, y), (x', y') \in \mathcal{H}$. We have

$$\begin{aligned} \|B_{\mathbf{m}_x}(x, y) - B_{\mathbf{m}_{x'}}(x', y')\|_{\mathcal{H} \times \mathcal{H}} &= \left\| \left(\nabla \mathbf{f}_{\mathbf{m}_x}(x) - \nabla \mathbf{f}_{\mathbf{m}_{x'}}(x'), 0 \right) \right\|_{\mathcal{H} \times \mathcal{H}} \\ &= \|\nabla \mathbf{f}_{\mathbf{m}_x}(x) - \nabla \mathbf{f}_{\mathbf{m}_{x'}}(x')\| \\ &\leq \beta \|x - x'\| \leq \beta \|(x, y) - (x', y')\|_{\mathcal{H} \times \mathcal{H}}, \end{aligned}$$

this proves (ii). The proof of (iii) can be directly deduced from Corollary 1. Indeed, let us fix $y, z \in \mathcal{H}$, we then have, for any $(x', y') \in \mathcal{H} \times \mathcal{H}$

$$\begin{aligned} \|B_{\mathbf{m}_y}(x', y') - B_{\mathbf{m}_z}(x', y')\|_{\mathcal{H} \times \mathcal{H}} &= \left\| \left(0, \nabla \mathbf{f}_{\mathbf{m}_y}(x') - \nabla \mathbf{f}_{\mathbf{m}_z}(x') \right) \right\|_{\mathcal{H} \times \mathcal{H}} \\ &= \|\nabla \mathbf{f}_{\mathbf{m}_y}(x') - \nabla \mathbf{f}_{\mathbf{m}_z}(x')\| \\ &\leq \beta_f \mathbf{W}(\mathbf{m}_y, \mathbf{m}_z) \leq \tau \beta_f \|y - z\|, \end{aligned}$$

where the last inequality follows from Assumption 2 and Assumption 5. As for (iv), we see that A is maximally monotone, and by (iii), $B_{\mathbf{m}_x}$ is Lipschitz continuous, we conclude again using [27, Lemma 2.4]. Now take $(x, y), (x', y') \in \mathcal{H} \times \mathcal{H}$ and $(u, v) \in \text{gra } \mathcal{T}_{\mathbf{m}_x}(x, y) \times \text{gra } \mathcal{T}_{\mathbf{m}_x}(x', y')$ with $u = (u_1, u_2), v = (v_1, v_2)$. Since by Assumption 5(a) $\nabla \mathbf{f}_{\mathbf{m}_x}$ is μ -strongly monotone and by Assumption 5(b) $\partial \mathbf{h}^*$ is β_h^{-1} -strongly monotone (see, e.g., [23, Theorem 18.15]), we get

$$\begin{aligned} \langle u - v, (x, y) - (x', y') \rangle &= \langle \nabla \mathbf{f}_{\mathbf{m}_x}(x) - \nabla \mathbf{f}_{\mathbf{m}_{x'}}(x'), x - x' \rangle + \langle u_1 - v_1, y - y' \rangle + \langle u_2 - v_2, x - x' \rangle \\ &\quad + \langle \mathbf{K}(x' - x), y - y' \rangle + \langle \mathbf{K}^*(y - y'), x - x' \rangle \\ &\geq \mu \|x - x'\|^2 + \beta_h^{-1} \|y - y'\|^2 \\ &\geq \min(\mu, \beta_h^{-1}) \|z - z'\|_{\mathcal{H} \times \mathcal{H}}^2, \end{aligned} \tag{70}$$

which proves $\min(\mu, \beta_h^{-1})$ -strong monotonicity of $\mathcal{T}_{\mathbf{m}_x}$. \blacksquare

6.2. Existence and uniqueness of equilibrium point. We are now in a position to prove the existence and uniqueness of an equilibrium to the problem (63)

Theorem 4 (Existence and uniqueness of equilibrium point). *Under Assumptions 2 and 5, the map*

$$S : (x, y) \in \mathcal{H} \mapsto \text{zer}(\mathcal{T}_{\mathbf{m}_x}) = \{(u, v) \in \mathcal{H} : (0, 0)_{\mathcal{H} \times \mathcal{H}} \in \mathcal{T}_{\mathbf{m}_x}(u, v)\}$$

is $\tilde{\rho}$ -Lipschitz with

$$\tilde{\rho} = \frac{\tau \beta_f}{\tilde{\mu}},$$

and $\tilde{\mu} = \min(\mu, \beta_h^{-1})$. In particular, if $\tilde{\rho} < 1$, the equilibrium (\bar{x}, \bar{y}) is unique.

Proof. Following the same lines as in Theorem 1 and using Lemma 7(i), we conclude. \blacksquare

Remark 6.2. Theorem 4 is to be compared to [54, Theorem 2.6]. Notice that the monotone inclusion (66) defining the equilibrium in [54] becomes a variational inequality due to the presence of constraints on x and y ([50, Example 12.50] for details).

Following the main steps of Sections-3-4, let us define the following gap function

$$\mathbf{E}_{\bar{x}}(x, y) = B_{\mathbf{m}_x}(x, y) - B_{\mathbf{m}_{\bar{x}}}(x, y). \tag{71}$$

Mimicking the proof of Lemma 5 and using Lemma 7, we prove a similar result

Lemma 8. *Under Assumptions 2 and 5, $\mathbf{E}_{\bar{x}}(\cdot)$ is $\beta(2 + \tau)$ -Lipschitz continuous.*

6.3. Related first and second-order dynamics.

1 6.3.1. *First ordre system.* To simplify the presentation, we set again $Z(t) = (x(t), y(t))$.
 2 Then, given an initial data $Z(t_0) = (x(t_0), y(t_0)) \in \text{dom}(\partial g) \times \text{dom}(\partial h^*)$, we consider the
 3 following first-order system associated to the monotone inclusion (66):

$$\dot{Z}(t) + \mathcal{T}_{m_x}(Z(t)) + \mathbf{E}_{\bar{x}}(Z(t)) \ni 0_{\mathcal{H} \times \mathcal{H}} \quad (72)$$

4 Or more explicitly

$$\begin{cases} \dot{x}(t) + K^*y(t) + \nabla f_{m_x}(x(t)) + \partial g(x(t)) + \mathbf{e}_{\bar{x}}(x(t)) \ni 0 \\ \dot{y}(t) - Kx(t) + \partial h^*(y(t)) \ni 0 \end{cases} \quad (\text{SPDS})$$

5 with $\mathbf{e}_{\bar{x}}(x(t)) = \nabla f_{m_x}(x(t)) - \nabla f_{m_x}(x(t))$. Arguing as in [Proposition 1](#) and [Theorem 2](#) we
 6 have the following result.

7 **Proposition 6.** *Assume that [Assumption 2](#) and [Assumption 5](#) hold. Then, for any initial*
 8 *data $Z(t_0) = (x(t_0), y(t_0)) \in \text{dom}(\partial g) \times \text{dom}(\partial h^*)$, (SPDS) admits a unique global solution*
 9 *$Z : t \in [t_0, \infty) \mapsto (x(t), y(t))$. Moreover,*

$$\|(x(t), y(t)) - (\bar{x}, \bar{y})\| \leq C e^{-2\tilde{\mu}(1-\tilde{\rho})t}, \quad \forall t \geq t_0,$$

10 with $C = \|(x(t_0), y(t_0)) - (\bar{x}, \bar{y})\| e^{2\tilde{\mu}(1-\tilde{\rho})t_0}$, $\tilde{\rho} = \frac{\tau\beta_t}{\tilde{\mu}}$ and $\tilde{\mu} = \min(\mu, \beta_h^{-1})$.

11 6.3.2. *second-order system.* To simplify the presentation, we assume that the following
 12 assumptions hold

- 13 **Assumption 5*.** (a) $f_{m_x}(x) = \mathbb{E}_{\xi \sim m_x} f(x, \xi)$, with $f(\cdot, \xi) \in C_{\beta_t}^{1,1}(\mathcal{H}) \cap \Gamma_0(\mathcal{H})$ and $f(\cdot, \xi)$
 14 is μ -strongly convex for all $\xi \in \Xi$, and $\xi \in \Xi \mapsto f(x, \xi)$ is measurable for any $x \in \mathcal{H}$,
 15 (b) $h \in C_{\beta_h}^{1,1}(\mathcal{H}) \cap \Gamma_0(\mathcal{H})$ is a strictly convex function, $g \in C^1(\mathcal{H})$ and $K \in \mathcal{L}(\mathcal{H}, \mathcal{H})$,
 16 (c) $0 \in \text{sri}(K \text{ dom}(g) - \text{dom}(r))$.

17 The main difference between [Assumption 5](#) and [Assumption 5*](#) is strict convexity of h
 18 which ensures differentiability of h^* (see, e.g., [23, Corollary 18.12] or [37, Theorem 4.1.1]).
 19 In this case we have

$$\mathcal{T}_{m_x}(x, y) = (\nabla_x \mathcal{L}_{m_x}(x, y), -\nabla_y \mathcal{L}_{m_x}(x, y)). \quad (73)$$

20 By analogy with ([ISEHD](#) _{$m, 2\sqrt{\tilde{\mu}}$}), we propose the following inertial system associated to ([66](#))

$$\begin{cases} \ddot{x}(t) + 2\sqrt{\tilde{\mu}}\dot{x}(t) + \nabla_x \mathcal{L}_m(x(t), y(t)) + \mathbf{e}_{\bar{x}}(x(t)) = 0 \\ \ddot{y}(t) + 2\sqrt{\tilde{\mu}}\dot{y}(t) - \nabla_y \mathcal{L}_m(x(t), y(t)) = 0 \end{cases} \quad (\text{ISPDS}_{m, 2\sqrt{\tilde{\mu}}})$$

21 Systems of the form ([ISPDS](#) _{$m, 2\sqrt{\tilde{\mu}}$}) can be found for instance in [8] or [26] in an unperturbed
 22 form, i.e., without the operator $\mathbf{e}_{\bar{x}}$. To lighten the presentation, we discuss ([ISPDS](#) _{$m, 2\sqrt{\tilde{\mu}}$})
 23 without a Hessian damping term. And for the convenience of the reader, we present the
 24 main ingredients of the Lyapunov analysis as in [Theorem 3](#). Notice that the existence of a
 25 global strong solution to ([ISPDS](#) _{$m, 2\sqrt{\tilde{\mu}}$}) can be obtained as in [Section 4](#). One can also see [8,
 26 Section 4].

27 Let us define the following function $\mathcal{V} : [t_0, \infty[\rightarrow \mathbb{R}^+$ by

$$\mathcal{V}(t) := \mathcal{L}_{m_x}(x(t), y(t)) - \mathcal{L}_{m_x}^* + \frac{1}{2} \|v(t)\|^2, \quad \text{where } v(t) := \sqrt{\tilde{\mu}} \left((x(t), y(t)) - (\bar{x}, \bar{y}) \right) + (\dot{x}(t), \dot{y}(t)),$$

28 where (\bar{x}, \bar{y}) is the equilibrium of ([63](#)) (cf. [Theorem 4](#)) and $\mathcal{L}_{m_x}^*(\bar{x}, \bar{y}) := \mathcal{L}_{m_x}^*$.

29 **Theorem 5.** *Assume that [Assumption 5*](#) holds and let $t \in [t_0, \infty[\mapsto (x(t), y(t))$ be the solution*
 30 *of ([ISPDS](#) _{$m, 2\sqrt{\tilde{\mu}}$}). Suppose that $\frac{\beta_t \tau}{\tilde{\mu}} < \frac{\sqrt{2}}{4}$. We then have:*

- 1 • for all $t \geq t_0$

$$\mathcal{V}(t) \leq C_1 e^{-\frac{\sqrt{\mu}}{4}t}, \text{ where } C_1 = \mathcal{V}(t_0) e^{-\frac{\sqrt{\mu}}{4}t_0}.$$

- 2 In particular

$$0 \leq \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x(t), y(t)) - \mathcal{L}_{\mathbf{m}_{\bar{x}}}^* \leq C e^{-\frac{\sqrt{\mu}}{4}t}. \quad (74)$$

- 3 • There exists $C_2 > 0$ such that,

$$\|(x(t) - \bar{x}, y(t) - \bar{y})\| \leq C_2 e^{-\frac{\sqrt{\mu}}{4}t}, \quad \forall t \geq t_0.$$

- 4 *Proof.* The proof is similar to the one of [Theorem 3](#). At first, we have, using [\(ISPDS_{m,2}√μ\)](#)

$$\dot{v}(t) = \left(\sqrt{\mu} \dot{x}(t) + \ddot{x}(t), \sqrt{\mu} \dot{y}(t) + \ddot{y}(t) \right) = \left(-\sqrt{\mu} \dot{x}(t) - \nabla_x \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x(t), y(t)) - \mathbf{e}_{\bar{x}}(x(t)), -\sqrt{\mu} \dot{y}(t) - \nabla_y \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x(t), y(t)) \right).$$

This gives

$$\begin{aligned} \dot{\mathcal{V}}(t) &= \langle \nabla_x \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x(t), y(t)), \dot{x}(t) \rangle + \langle \nabla_y \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x(t), y(t)), \dot{y}(t) \rangle \\ &\quad + \langle \sqrt{\mu}(x(t) - \bar{x}) + \dot{x}(t), -\sqrt{\mu} \dot{x}(t) - \nabla_x \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x(t), y(t)) - \mathbf{e}_{\bar{x}}(x(t)) \rangle \\ &\quad + \langle \sqrt{\mu}(y(t) - \bar{y}) + \dot{y}(t), -\sqrt{\mu} \dot{y}(t) - \nabla_y \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x(t), y(t)) \rangle \end{aligned} \quad (75)$$

We get after some simplifications

$$\begin{aligned} \dot{\mathcal{V}}(t) + \tilde{\mu} \langle \dot{x}(t), x(t) - \bar{x} \rangle + \tilde{\mu} \langle \dot{y}(t), y(t) - \bar{y} \rangle + \sqrt{\mu} \|\dot{x}(t)\|^2 \\ + \sqrt{\mu} \|\dot{y}(t)\|^2 + \sqrt{\mu} \langle \nabla_x \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x(t), y(t)), x(t) - \bar{x} \rangle + \sqrt{\mu} \langle \nabla_y \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x(t), y(t)), y(t) - \bar{y} \rangle = -\langle v(t), \mathbf{E}_{\bar{x}}(x(t)) \rangle. \end{aligned} \quad (76)$$

That is

$$\begin{aligned} \dot{\mathcal{V}}(t) + \tilde{\mu} \left\langle (\dot{x}(t), \dot{y}(t)), (x(t) - \bar{x}, y(t) - \bar{y}) \right\rangle + \sqrt{\mu} \|\dot{x}(t), \dot{y}(t)\|^2 \\ + \sqrt{\mu} \langle \mathcal{T}_{\mathbf{m}_{\bar{x}}}(x(t), y(t)), (x(t) - \bar{x}, y(t) - \bar{y}) \rangle = -\langle v(t), \mathbf{E}_{\bar{x}}(x(t)) \rangle. \end{aligned} \quad (77)$$

- 5 Using $\tilde{\mu}$ -strong monotonicity of $\mathcal{T}_{\mathbf{m}_{\bar{x}}}$, and the fact that $0_{\mathcal{H} \times \mathcal{H}} \in \mathcal{T}_{\mathbf{m}_{\bar{x}}}(\bar{x}, \bar{y})$ we have

$$\langle \mathcal{T}_{\mathbf{m}_{\bar{x}}}(x(t), y(t)), (x(t) - \bar{x}, y(t) - \bar{y}) \rangle \geq \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x(t), y(t)) - \mathcal{L}_{\mathbf{m}_{\bar{x}}}^* + \frac{\tilde{\mu}}{2} \|(x(t) - \bar{x}, y(t) - \bar{y})\|^2, \quad (78)$$

- 6 and using this in [\(77\)](#), we get

$$\dot{\mathcal{V}}(t) + \sqrt{\mu} \Theta(t) \leq \|v(t)\| \|\mathbf{E}_{\bar{x}}(x(t))\|, \quad (79)$$

- 7 where

$$\begin{aligned} \Theta(t) &:= \tilde{\mu} \left\langle (\dot{x}(t), \dot{y}(t)), (x(t) - \bar{x}, y(t) - \bar{y}) \right\rangle + \mathcal{L}_{\mathbf{m}_{\bar{x}}}(x(t), y(t)) - \mathcal{L}_{\mathbf{m}_{\bar{x}}}^* + \frac{\tilde{\mu}}{2} \|(x(t) - \bar{x}, y(t) - \bar{y})\|^2 \\ &= \mathcal{V}(t) + \sqrt{\mu} \|\dot{x}(t), \dot{y}(t)\|^2. \end{aligned} \quad (80)$$

- 8 Thus, [\(79\)](#) becomes

$$\dot{\mathcal{V}}(t) + \sqrt{\mu} \mathcal{V}(t) + \frac{\sqrt{\mu}}{2} \|\dot{x}(t), \dot{y}(t)\|^2 \leq \|v(t)\| \|\mathbf{E}_{\bar{x}}(x(t))\|, \quad (81)$$

- 9 Again, using $\tilde{\mu}$ -strong monotonicity of $\mathcal{T}_{\mathbf{m}_{\bar{x}}}$ we get

$$\dot{\mathcal{V}}(t) + \sqrt{\mu} \mathcal{V}(t) + \frac{\sqrt{\mu}}{2} \|\dot{x}(t), \dot{y}(t)\|^2 + \frac{\tilde{\mu}^{3/2}}{2} \|(x(t) - \bar{x}, y(t) - \bar{y})\|^2 \leq \|v(t)\| \|\mathbf{E}_{\bar{x}}(x(t))\|. \quad (82)$$

- 10 We have, using Young's inequality

$$\|v(t)\| \|\mathbf{E}_{\bar{x}}(x(t))\| \leq \frac{\sqrt{\mu}}{8} \|v(t)\|^2 + \frac{2}{\sqrt{\mu}} \|\mathbf{E}_{\bar{x}}(x(t))\|^2 \leq \frac{\sqrt{\mu}}{4} \mathcal{V}(t) + \frac{2\beta_f^2 \tau^2}{\sqrt{\mu}} \|(x(t) - \bar{x}, y(t) - \bar{y})\|^2.$$

1 Using Lemma 7-(iii), we get after rearranging the terms

$$\dot{\mathcal{V}}(t) + \frac{\sqrt{\tilde{\mu}}}{2}\mathcal{V}(t) + \frac{\sqrt{\tilde{\mu}}}{4}\|\dot{x}(t), \dot{y}(t)\|^2 + \sqrt{\tilde{\mu}}\left(\frac{\tilde{\mu}}{4} - \frac{2\beta_f^2\tau^2}{\tilde{\mu}}\right)\|(x(t) - \bar{x}, y(t) - \bar{y})\|^2 \leq 0.$$

2 Since $\frac{\beta_f\tau}{\tilde{\mu}} < \frac{\sqrt{2}}{4}$, we have

$$\dot{\mathcal{V}}(t) + \frac{\sqrt{\tilde{\mu}}}{2}\mathcal{V}(t) \leq 0,$$

3 which gives after integration

$$\mathcal{V}(t) \leq \mathcal{V}(t_0)e^{-\frac{\sqrt{\tilde{\mu}}}{4}(t-t_0)}, \quad \forall t \geq t_0. \quad (83)$$

4 By definition of \mathcal{V} , we have

$$\mathcal{L}_{m_{\bar{x}}}(x(t), y(t)) - \mathcal{L}_{m_{\bar{x}}}^* \leq \mathcal{V}(t_0)e^{-\frac{\sqrt{\tilde{\mu}}}{4}(t-t_0)}. \quad (84)$$

5 Coming back to (83), we have

$$\|v(t)\|^2 = \|\dot{x}(t), \dot{y}(t)\|^2 + 2\sqrt{\tilde{\mu}}\langle \dot{x}(t), \dot{y}(t), (x(t) - \bar{x}, y(t) - \bar{y}) \rangle + \tilde{\mu}\|(x(t) - \bar{x}, y(t) - \bar{y})\|^2 \leq C_1e^{-\frac{\sqrt{\tilde{\mu}}}{4}t}, \quad (85)$$

6 with $C_1 = 2e^{\frac{\sqrt{\tilde{\mu}}}{4}t_0}\mathcal{V}(t_0)$. Setting $U(t) = \sqrt{\tilde{\mu}}\|(x(t) - \bar{x}, y(t) - \bar{y})\|^2$, (85) gives

$$\dot{U}(t) + \sqrt{\tilde{\mu}}U(t) \leq C_1e^{-\frac{\sqrt{\tilde{\mu}}}{4}t},$$

7 which implies after integrating between t_0 and t that

$$\|(x(t) - \bar{x}, y(t) - \bar{y})\| \leq C_2e^{-\frac{\sqrt{\tilde{\mu}}}{4}t},$$

8 where the constant C_2 depends only on C_1 and $U(t_0) = \|(x(t_0) - \bar{x}, y(t_0) - \bar{y})\|$. This
9 finishes the proof. \blacksquare

10 7. COMMENTS, EXTENSIONS AND FUTURE WORK

11 In this paper, we adopted a dynamical system approach to study some stochastic opti-
12 mization problems with state-dependent distributions. We investigated the existence and
13 uniqueness of equilibrium points, well-posedness as well as convergence properties of the
14 trajectories, for both first and second-order dynamics. We highlighted some dynamical
15 and geometrical properties of the state-dependent distributions suggesting that the natural
16 framework to study problems of the form (5) is the one of *metric random walk spaces*. More
17 particularly, the notion of coarse Ricci curvature gives a new insight on the geometrical
18 hidden structure of this kind of problems. Finally, we discussed as an application the inertial
19 primal-dual algorithm. We present here some ongoing works, possible extensions as well as
20 some open problems.

21 **7.1. Inertial algorithms.** Relying on the discretization of the dynamics studied in Section 4
22 and Section 6, more specifically, $(\text{ISEHD}_{m,2\sqrt{\tilde{\mu}}})$ and $(\text{ISPDS}_{m,2\sqrt{\tilde{\mu}}})$, we obtain new inertial
23 algorithms with Hessian-driven damping for stochastic optimization problems with decision-
24 dependent distributions. These algorithms exhibit rapid convergence properties and can also
25 be adapted to the nonsmooth case. This is being addressed in ongoing work.

1 **7.2. Implicit Hessian damping.** We focused in [Section 4](#) on the explicit Hessian damping in
 2 the smooth case. Yet, it is possible to consider implicit damping as in [\[9, 14\]](#)

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla G_{m_{\bar{x}}}(x(t) + \omega\dot{x}(t)) + \mathbf{e}_{\bar{x}}(x(t)) = 0, \quad (\text{ISIHd}_{m,\gamma})$$

3 The dynamics $(\text{ISIHd}_{m,\gamma})$ is referred to as an Inertial System with Implicit Hessian damping,
 4 since one can observe, using Taylor expansion

$$\nabla G_{m_{\bar{x}}}(x(t) + \omega\dot{x}(t)) \approx \nabla G_{m_{\bar{x}}}x(t) + \nabla^2 G_{m_{\bar{x}}}(x(t))\dot{x}(t).$$

5 As it was observed in [\[14\]](#), higher-order moments of the perturbation $\mathbf{e}_{\bar{x}}$ are required to get
 6 fast convergence guarantees in the implicit case compared to the explicit one. Since in our
 7 analysis (see [Theorem 3](#)) no integrability assumption on $\mathbf{e}_{\bar{x}}$ is needed, it is interesting to
 8 investigate the effect of implicit Hessian damping, both in the smooth and nonsmooth cases.

9 **7.3. Tikhonov-regularization.** In [Section 4](#) we restricted ourselves, for sake of simplicity, the
 10 analysis to the case where the operator $F_{m_{\bar{x}}}$ is smooth. However, it is possible to consider
 11 second-order dynamics for general (and possibly nonpotential) operators, by considering, for
 12 $\lambda > 0$ the following dynamic

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + F_{m_{\bar{x}}^{\lambda(t)}}^{\lambda(t)}(x(t)) + \omega \frac{d}{dt} \left(F_{m_{\bar{x}}^{\lambda(t)}}^{\lambda(t)}(x(t)) \right) + \mathbf{e}_{\bar{x}}(x(t)) = 0. \quad (\text{ISEHD}_{m,\lambda,\gamma})$$

13 where $F_{m_{\bar{x}}}^{\lambda}$ is the so-called Yosida approximation of $F_{m_{\bar{x}}}$ defined by $F_{m_{\bar{x}}}^{\lambda} = \frac{1}{\lambda}(\text{id} - J_{\lambda F_{m_{\bar{x}}}})$
 14 and $J_{\lambda F_{m_{\bar{x}}}} = (\text{id} + JF_{m_{\bar{x}}})^{-1}$ is the resolvent of $F_{m_{\bar{x}}}$. This approach comes with several
 15 advantages. First, the Yosida approximation is single valued so that the monotone inclusion
 16 $(\mathbf{p}\text{-ISMI})$ reduces to the classical differential equation $(\text{ISEHD}_{m,\lambda,\gamma})$. In addition one can
 17 exploit the λ -cocoercivity of $F_{m_{\bar{x}}}^{\lambda}$ and the fact that $\text{zer } F_{m_{\bar{x}}} = \text{zer } F_{m_{\bar{x}}}^{\lambda}$. The approach was
 18 used in [\[17\]](#) for $\omega = 0$ and in the recent work [\[12\]](#) for Newton-like dynamics. We are exploring
 19 the adaption of this techniques to stochastic monotone inclusions with state dependent
 20 distribution in an ongoing work.

21 **7.4. Weaker Assumptions.** We have seen that one of the crucial assumptions in the analysis
 22 is [Assumption 2](#), which concerns the Lipschitz behavior of the distribution $(m_x)_x$. A natural
 23 question that arises is what happens under a weaker assumption. For example, when $x \mapsto m_x$
 24 is Hölder continuous. We are not aware of any existing results in this direction. We plan to
 25 investigate this in future work.

ACKNOWLEDGEMENT

26 The work of H.E was supported by the ANR grant, reference: ANR-20-CE92-0037.

A. GRONWALL INEQUALITIES

27 In this section we list several auxiliary results that we make use of in the paper.

28 **Lemma 9** (Gronwall's lemma: differential form). *Let u, v be two C^0 (resp. C^1) nonnegative
 29 function on $[0, T]$ and let w be a continuous function on $[0, T]$. We assume that*

$$\frac{1}{2} \frac{d}{dt} u^2(t) \leq w(t)u^2(t) + u(t)v(t) \text{ on } (0, T), \quad (86)$$

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

$$u(t) \leq u(0)e^{K(t)} + \int_0^t v(s)e^{K(t)-K(s)} ds, \quad (87)$$

31 where $K(t) = \int_0^t w(s) ds$.

1 **Lemma 10.** [27, Lemma A.5] Let $v \in L^1(t_0, T; \mathbb{R}^+)$ and $u \in C^0(t_0, T)$ such that

$$\frac{1}{2}u^2(t) \leq \frac{1}{2}c^2 + \int_{t_0}^t u(s)v(s)ds,$$

2 for some $c \geq 0$ for all $t \in [t_0, T]$. Then

$$|u(t)| \leq c + \int_{t_0}^t v(s)ds.$$

3 B. BANACH FIXED POINT THEOREM & PICARD ITERATIVE METHOD

4 **Theorem 6** (see, e.g., [6, 28]). Let (\mathcal{X}, d) be a complete metric space and $S : \mathcal{X} \rightarrow \mathcal{X}$ be a
5 strict contraction, i.e., there exists a constant $\rho < 1$ such that

$$d(S(x), S(y)) \leq \rho d(x, y), \forall x, y \in \mathcal{X}.$$

6 Then, there exists a unique $\bar{x} \in \mathcal{X}$ such that $S(\bar{x}) = \bar{x}$. Moreover, for any $x_0 \in \mathcal{X}$, the
7 sequence starting from x_0 with $x_{n+1} = S(x_n)$ for all $n \in \mathbb{N}$ converges to \bar{x} as n goes to ∞ .

8 REFERENCES

- 9 [1] F. Alvarez. On the minimizing property of a second order dissipative system in Hilbert spaces. *SIAM J.*
10 *Control Optim.*, 38(4):1102–1119, 2000.
- 11 [2] F. Alvarez and H. Attouch. An inertial proximal method for maximal monotone operators via discretiza-
12 tion of a nonlinear oscillator with damping. *Set-Valued Anal.*, 9(1-2):3–11, 2001.
- 13 [3] F. Alvarez, H. Attouch, J. Bolte, and P. Redont. A second-order gradient-like dissipative dynami-
14 cal system with hessian-driven damping.: Application to optimization and mechanics. *Journal de*
15 *mathématiques pures et appliquées*, 81(8):747–779, 2002.
- 16 [4] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability*
17 *measures*. Basel: Birkhäuser, 2nd ed. edition, 2008.
- 18 [5] V. Apidopoulos, J.-F. Aujol, and C. Dossal. Convergence rate of inertial forward–backward algorithm
19 beyond nesterov’s rule. *Mathematical Programming*, 180(1):137–156, 2020.
- 20 [6] H. Attouch, G. Buttazzo, and G. Michaille. *Variational analysis in Sobolev and BV spaces*, volume 17 of
21 *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadel-
22 phia, PA; Mathematical Optimization Society, Philadelphia, PA, second edition, 2014. Applications to
23 PDEs and optimization.
- 24 [7] H. Attouch, A. Cabot, and P. Redont. The dynamics of elastic shocks via epigraphical regularization of
25 a differential inclusion. Barrier and penalty approximations. *Adv. Math. Sci. Appl.*, 12(1):273–306, 2002.
- 26 [8] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi. Fast convergence of dynamical ADMM via time scaling
27 of damped inertial dynamics. *J. Optim. Theory Appl.*, 193(1-3):704–736, 2022.
- 28 [9] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi. First-order optimization algorithms via inertial systems
29 with Hessian driven damping. *Math. Program.*, 193(1, Ser. A):113–155, 2022.
- 30 [10] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont. Fast convergence of inertial dynamics and
31 algorithms with asymptotic vanishing viscosity. *Mathematical Programming*, 168:123–175, 2018.
- 32 [11] H. Attouch, Z. Chbani, and H. Riahi. Rate of convergence of the nesterov accelerated gradient method
33 in the subcritical case $\alpha \leq 3$. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:2, 2019.
- 34 [12] H. Attouch and S. Csaba László. Continuous Newton-like inertial dynamics for monotone inclusions.
35 *Set-Valued Var. Anal.*, 29(3):555–581, 2021.
- 36 [13] H. Attouch and A. Damlamian. On multivalued evolution equations in hilbert spaces. *Israel Journal of*
37 *Mathematics*, 12:373–390, 1972.
- 38 [14] H. Attouch, J. Fadili, and V. Kungurtsev. On the effect of perturbations in first-order optimization
39 methods with inertia and Hessian driven damping. *Evol. Equ. Control Theory*, 12(1):71–117, 2023.
- 40 [15] H. Attouch, X. Goudou, and P. Redont. The heavy ball with friction method, i. the continuous
41 dynamical system: global exploration of the local minima of a real-valued function by asymptotic
42 analysis of a dissipative dynamical system. *Communications in Contemporary Mathematics*, 2(01):1–34,
43 2000.

- 1 [16] H. Attouch and J. Peypouquet. The rate of convergence of nesterov’s accelerated forward-backward
2 method is actually faster than $1/k^2$. *SIAM Journal on Optimization*, 26(3):1824–1834, 2016.
- 3 [17] H. Attouch and J. Peypouquet. Convergence of inertial dynamics and proximal algorithms governed by
4 maximally monotone operators. *Math. Program.*, 174(1-2, Ser. B):391–432, 2019.
- 5 [18] H. Attouch, J. Peypouquet, and P. Redont. A dynamical approach to an inertial forward-backward
6 algorithm for convex minimization. *SIAM Journal on Optimization*, 24(1):232–256, 2014.
- 7 [19] H. Attouch, J. Peypouquet, and P. Redont. Fast convex optimization via inertial dynamics with hessian
8 driven damping. *Journal of Differential Equations*, 261(10):5734–5783, 2016.
- 9 [20] H. Bahouri, J.-Y. Chemin, and R. Danchin. *Fourier analysis and nonlinear partial differential equa-*
10 *tions*, volume 343 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of*
11 *Mathematical Sciences]*. Springer, Heidelberg, 2011.
- 12 [21] J. Baillon and H. Brézis. Une remarque sur le comportement asymptotique des semigroupes non linéaires.
13 *Houston J. Math.*, 2:5–7, 1976.
- 14 [22] V. Barbu. *Nonlinear differential equations of monotone types in Banach spaces*. Springer Monographs
15 in Mathematics. Springer, New York, 2010.
- 16 [23] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*.
17 CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2011. With
18 a foreword by Hedy Attouch.
- 19 [24] P. Bianchi, W. Hachem, and A. Salim. A fully stochastic primal-dual algorithm. *Optimization Letters*,
20 15(2):701–710, 2021.
- 21 [25] R. I. Boş, E. R. Csetnek, and S. C. László. Tikhonov regularization of a second order dynamical system
22 with Hessian driven damping. *Math. Program.*, 189(1-2 (B)):151–186, 2021.
- 23 [26] R. I. Boş, E. R. Csetnek, and D.-K. Nguyen. Fast augmented Lagrangian method in the convex regime
24 with convergence guarantees for the iterates. *Math. Program.*, 200(1 (A)):147–197, 2023.
- 25 [27] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de*
26 *Hilbert*. North-Holland Mathematics Studies, No. 5. North-Holland Publishing Co., Amsterdam-London;
27 American Elsevier Publishing Co., Inc., New York, 1973.
- 28 [28] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer,
29 New York, 2011.
- 30 [29] R. E. Bruck Jr. Asymptotic convergence of nonlinear contraction semigroups in hilbert space. *Journal*
31 *of Functional Analysis*, 18(1):15–26, 1975.
- 32 [30] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications
33 to imaging. *Journal of mathematical imaging and vision*, 40:120–145, 2011.
- 34 [31] L. Condat. A primal–dual splitting method for convex optimization involving lipschitzian, proximable
35 and linear composite terms. *Journal of optimization theory and applications*, 158(2):460–479, 2013.
- 36 [32] J. Cutler, M. Díaz, and D. Drusvyatskiy. Stochastic approximation with decision-dependent distributions:
37 asymptotic normality and optimality. *arXiv preprint arXiv:2207.04173*, 2022.
- 38 [33] A. Derrow-Pinion, J. She, D. Wong, O. Lange, T. Hester, L. Perez, M. Nunkesser, S. Lee, X. Guo,
39 B. Wiltshire, P. W. Battaglia, V. Gupta, A. Li, Z. Xu, A. Sanchez-Gonzalez, Y. Li, and P. Velickovic. Eta
40 prediction with graph neural networks in google maps. In *Proceedings of the 30th ACM International*
41 *Conference on Information & Knowledge Management, CIKM ’21*, page 3767–3776, New York, NY,
42 USA, 2021. Association for Computing Machinery.
- 43 [34] R. Douc, E. Moulines, P. Priouret, and P. Soulier. *Markov chains*. Springer Series in Operations
44 Research and Financial Engineering. Springer, Cham, 2018.
- 45 [35] D. Drusvyatskiy and L. Xiao. Stochastic optimization with decision-dependent distributions. *Mathematics*
46 *of Operations Research*, 2022.
- 47 [36] O. Hernández-Lerma and J. B. Lasserre. *Markov chains and invariant probabilities*, volume 211 of *Prog.*
48 *Math*. Basel: Birkhäuser, 2003.
- 49 [37] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. Grundlehren Text Edit.
50 Berlin: Springer, 2001.
- 51 [38] D. Kim. Accelerated proximal point method for maximally monotone operators. *Mathematical*
52 *Programming*, 190(1):57–87, 2021.
- 53 [39] S. C. László. Convergence rates for an inertial algorithm of gradient type associated to a smooth
54 non-convex minimization. *Mathematical Programming*, 190(1):285–329, 2021.

- 1 [40] T. Lin and M. I. Jordan. A control-theoretic perspective on optimal high-order optimization. *Mathematical Programming*, pages 1–47, 2022.
- 2
- 3 [41] J. Macfarlane. Your navigation app is making traffic unmanageable. *IEEE Spectrum*, 19, 2019.
- 4 [42] J. Mazón, M. Solera-Diana, and J. Toledo-Melero. *Variational and Diffusion Problems in Random Walk Spaces*. Progress in Nonlinear Differential Equations and Their Applications. Springer Nature Switzerland, 2023.
- 5
- 6
- 7 [43] C. Mendler-Dünner, J. C. Perdomo, T. Zrnic, and M. Hardt. Stochastic optimization for performative prediction. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS’20, Red Hook, NY, USA, 2020. Curran Associates Inc.
- 8
- 9
- 10 [44] Y. Nesterov. A method for solving the convex programming problem with convergence rate $o(1/k^2)$. *Proceedings of the USSR Academy of Sciences*, 269:543–547, 1983.
- 11
- 12 [45] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.
- 13 [46] Y. Ollivier. A survey of ricci curvature for metric spaces and markov chains. In *Probabilistic approach to geometry*, volume 57, pages 343–382. Mathematical Society of Japan, 2010.
- 14
- 15 [47] J. Perdomo, T. Zrnic, C. Mendler-Dünner, and M. Hardt. Performative prediction. In H. D. III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7599–7609. PMLR, 13–18 Jul 2020.
- 16
- 17
- 18 [48] B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *Ussr computational mathematics and mathematical physics*, 4(5):1–17, 1964.
- 19
- 20 [49] H. Raguét, J. Fadili, and G. Peyré. A generalized forward-backward splitting. *SIAM Journal on Imaging Sciences*, 6(3):1199–1226, 2013.
- 21
- 22 [50] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.
- 23
- 24 [51] B. Shi, S. S. Du, M. I. Jordan, and W. J. Su. Understanding the acceleration phenomenon via high-resolution differential equations. *Mathematical Programming*, pages 1–70, 2022.
- 25
- 26 [52] W. Su, S. Boyd, and E. J. Candès. A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights. *J. Mach. Learn. Res.*, 17:43, 2016. Id/No 153.
- 27
- 28 [53] B. C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Advances in Computational Mathematics*, 38(3):667–681, 2013.
- 29
- 30 [54] K. Wood and E. Dall’Anese. Stochastic saddle point problems with decision-dependent distributions. *SIAM Journal on Optimization*, 33(3):1943–1967, 2023.
- 31