RECOVERING NESTEROV ACCELERATED DYNAMICS FROM HEAVY BALL DYNAMICS VIA TIME RESCALING*

HEDY ATTOUCH†, RADU IOAN BOŢ‡, DAVID ALEXANDER HULETT§, AND DANG-KHOA NGUYEN¶

Abstract. In a real Hilbert space, we consider two classical problems: the global minimization of a smooth and convex function f (i.e., a convex optimization problem) and finding the zeros of a monotone and continuous operator M (i.e., a monotone equation). Attached to the optimization problem, first we study the asymptotic properties of a generalization of Polyak's Heavy Ball dynamics introduced in 1964; namely, we consider a positive function $b(\cdot)$ multiplying ∇f. We show small o convergence rates of the function values dependent on $b(\cdot)$ and weak convergence of trajectories towards minimizers of f. In 2015, Su, Boyd and Candés introduced a second-order system which could be seen as the continuous-time counterpart of Nesterov's accelerated gradient. As the first key point of this paper, we show that for a special choice for $b(t)$, these two seemingly unrelated dynamical systems are connected: namely, they are time reparametrizations of each other. Every statement regarding the continuous-time accelerated gradient system can be recovered from its Heavy Ball counterpart.

As the second key point of this paper, we observe that this connection extends beyond the optimization setting. Attached to the monotone equation involving the operator M, we again consider a Heavy Ball-like system suited for the monotone operator setting. We derive small o rates for the norm of the operator along the trajectories, and show the weak convergence of the trajectories towards zeros of M. For a particular case of this system, we establish a time reparametrization equivalence with the Fast OGDA dynamics introduced by Boţ, Csetnek and Nguyen in 2022, which can be seen as an analog of the continuous accelerated gradient dynamics, but for monotone operators. Every statement regarding the Fast OGDA system can be recovered from a Heavy Ball-like system.

Key words. Nesterov accelerated gradient method; Heavy Ball with friction; damped inertial dynamic; time scaling; monotone equations; monotone operator flow; convergence rates; convergence of trajectories

AMS subject classifications. 37N40, 47H05, 47J20, 90C25

1. Introduction.

1.1. Convex optimization. In a real Hilbert space $\mathcal{H}$, for a convex and continuously differentiable function $f : \mathcal{H} \to \mathbb{R}$, consider the minimization problem

$$\min_{x\in\mathcal{H}} f(x), \tag{1.1}$$

which we assume to have an optimal solution. Su, Boyd and Candès [25] observed that the second-order dynamical system

$$\ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + \nabla f(x(s)) = 0 \quad \text{for } s \geq s_0 > 0, \tag{1.2}$$

can be viewed as a continuous-time counterpart to Nesterov's accelerated gradient method [18, 19], designed to efficiently solve this problem. Attouch, Chbani, Peypouquet and Redont [6] and May [17] showed that when $\alpha > 3$, any solution $x(s)$ to this dynamical system satisfies $f(x(s)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{s^2}\right)$ and converges weakly to a global minimizer of f as $s \to +\infty$.

Also connected to (1.1), we have the Heavy Ball dynamics first introduced by Polyak [21, 22]

$$\ddot{y}(t) + \lambda\dot{y}(t) + \nabla f(y(t)) = 0, \tag{1.3}$$

where the name comes from the mechanical interpretation of this system: $y(t)$ describes the horizontal position of an object which moves alongside the graph of the function f in a medium with viscous friction coefficient λ. For more details about this derivation, we refer the reader to [8]. Álvarez [3] first showed that any solution $y(t)$ to the Heavy Ball dynamics satisfies $f(y(t)) \to \inf_{\mathcal{H}} f$ and converges weakly to a

*Submitted to the editors DATE.

Funding: RIB was partially supported by the Austrian Science Fund (FWF), 10.55776/P34922. DAH was supported by the Doctoral Programme *Vienna Graduate School on Computational Optimization (VGSCO)*, funded by the Austrian Science Fund (FWF), 10.55776/W1260. DKN's research was funded by the Postdoctoral Scholarship Programme of Vingroup Innovation Foundation (VINIF), code VINIF.2024.STS.37.

†IMAG, Univ. Montpellier, CNRS, Montpellier, France, hedy.attouch@umontpellier.fr.

‡Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, radu.bot@univie.ac.at.

§Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, david.alexander.hulett@univie.ac.at.

¶Faculty of Mathematics and Computer Science, University of Science, Ho Chi Minh City, Vietnam, AND Vietnam National University, Ho Chi Minh City, Vietnam, ndkhoa@hcmus.edu.vn.

global minimizer of f as $t \to +\infty$. Later, it was proved that the function values actually exhibit a rate $f(y(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t}\right)$ as $t \to +\infty$. Observe how the rates for the function values improve when instead of a constant friction coefficient λ, we have a so-called asymptotically vanishing damping $\frac{\alpha}{t}$ accompanying the velocity.

In [7], Attouch, Chbani and Riahi analyzed a more general version of (1.2) which reads

$$\ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + b(s)\nabla f(x(s)) = 0, \tag{1.4}$$

where $b : [s_0, +\infty[\to \mathbb{R}_{++}$ is a differentiable and nondecreasing function and $\mathbb{R}_{++}$ denotes the set of strictly positive real numbers. They explain how the presence of $b(\cdot)$ can be interpreted as the result of a time reparametrization of (1.2), which can yield faster convergence rates for the function values provided $b(\cdot)$ is correctly chosen. Following the same idea, in Section 2 we will study the asymptotic properties of

$$\ddot{y}(t) + \lambda\dot{y}(t) + b(t)\nabla f(y(t)) = 0 \quad \text{for } t \geq t_0 \geq 0. \tag{1.5}$$

We will show that if $b(\cdot)$ satisfies the growth condition $\sup_{t \geq t_0} \frac{\dot{b}(t)}{b(t)} < \lambda$, then any solution $y(t)$ to (1.5) satisfies $f(y(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{\int_{\frac{t_0+t}{2}}^{t} b(r)dr}\right)$, and it converges weakly to a global minimizer of f as $t \to +\infty$. Furthermore, and as one of the main points of this paper, in Section 3 we will show that for an appropriate choice for $b(\cdot)$, (1.5) and (1.2) are time reparametrizations of each other. Every statement regarding (1.2) can be recovered from properties of (1.5). To condense the previous discussion into a single assertion, we show that

Nesterov accelerated gradient dynamics can be recovered from the Heavy Ball dynamics through a time rescaling process.

1.2. Monotone equations. It turns out that this story goes beyond the optimization setting and has an analog for monotone equations. For a continuous and monotone operator $M : \mathcal{H} \to \mathcal{H}$, i.e. $\langle M(y) - M(x), y - x\rangle \geq 0$ for every $x, y \in \mathcal{H}$, consider the problem

$$\text{find } x \in \mathcal{H} \text{ such that } M(x) = 0, \tag{1.6}$$

which we assume to have at least one solution.

It is known that the monotone flow dynamical system applied to M,

$$\dot{x}(s) + M(x(s)) = 0, \tag{1.7}$$

fails in general to produce a solution which converges weakly to a zero of M, unless M is cocoercive, i.e., for some $\beta > 0$, it holds $\langle M(y) - M(x), y - x\rangle \geq \beta\|M(y) - M(x)\|^2$ for every $x, y \in \mathcal{H}$. Attouch, Boţ and Nguyen [4] proved that in the latter case any solution $x(s)$ to (1.7) satisfies $\|M(x(s))\| = o\left(\frac{1}{\sqrt{s}}\right)$ and converges weakly to a zero of M as $s \to +\infty$.

However, as shown by Attouch and Svaiter in [11], by adding a correction term which is the time derivative of the operator along the solution, i.e.,

$$\lambda(s)\dot{x}(s) + \frac{d}{ds}M(x(s)) + M(x(s)) = 0, \tag{1.8}$$

where $\lambda(\cdot)$ is a positive function, then even in case of a monotone operator M any solution $x(s)$ to this system satisfies $\|M(x(s))\| \to 0$ and weakly converges to a zero of M as $s \to +\infty$. For λ constant, the previous system can be rewritten equivalently as being of the form (1.7), but applied to the Moreau-Yosida regularization of M, which is always cocoercive. Combining second-order in time dynamics with an asymptotically vanishing damping of the form $\frac{\alpha}{s}$ has already been proven to produce fast convergence properties in the optimization setting; it is perhaps not too surprising that this effect can be transposed to the monotone inclusion setting. Indeed, for a general, possibly set-valued maximally monotone operator $M : \mathcal{H} \to 2^{\mathcal{H}}$, Attouch and Peypouquet studied in [9] the system

$$\ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + M_{\lambda(s)}(x(s)) = 0, \quad \text{where} \quad M_\lambda := \frac{1}{\lambda}\Big(\mathrm{Id} - (\mathrm{Id} + \lambda M)^{-1}\Big). \tag{1.9}$$

M_λ is the aforementioned Moreau-Yosida regularization of M of index $\lambda > 0$ of M. It is a known fact that M and M_λ share the same set of zeros. For the above dynamics, if $\lambda(s)$ grows as s^2, the authors show a rate of convergence of $o\left(\frac{1}{s^2}\right)$ for $\|M_{\lambda(s)}(x(s))\|$, as well as the weak convergence of $x(s)$ to a zero of M as $s \to +\infty$.

If M is single-valued and continuous, we would prefer a scheme that evaluates $M(x(s))$ directly, rather than through its Moreau-Yosida regularization, also to have a setting that allows explicit discretizations and therefore forward evaluations of M, not as for (1.9). Combining the ideas of having a correction term like in (1.8) together with second-order and asymptotically vanishing terms like in (1.9) and a time rescaling coefficient similar to that of (1.4) gives rise to the Fast OGDA dynamics

$$\ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + \beta(s)\frac{d}{ds}M(x(s)) + \frac{1}{2}\left(\dot{\beta}(s) + \frac{\alpha}{s}\beta(s)\right)M(x(s)) = 0 \quad \text{for } s \geq s_0 > 0, \tag{1.10}$$

introduced and studied by Boţ, Csetnek and Nguyen in [14]. Provided that $\beta(\cdot)$ fulfills a growth condition, the authors show a rate of $o\left(\frac{1}{s\beta(s)}\right)$ for $\|M(x(s))\|$ and the weak convergence of $x(s)$ towards a zero of M as $s \to +\infty$. For a more general system where a damping of the form $\frac{\alpha}{s^r}$ is considered, we refer the reader to [15]. When $\beta(\cdot) \equiv 1$, the dynamics (1.10) reads

$$\ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + \frac{d}{ds}M(x(s)) + \frac{\alpha}{2s}M(x(s)) = 0. \tag{1.11}$$

This is perhaps the most interesting case, since this system admits an explicit discretization which has identical convergence properties to its continuous-time counterpart, i.e., an algorithm which generates a sequence $(x_k)_{k\in\mathbb{N}}$, combining Nesterov momentum with operator correction terms and using only forward evaluations of M, and which fulfills $\|M(x_k)\| = o\left(\frac{1}{k}\right)$ and converges weakly towards a zero of M as $k \to +\infty$.

In Section 4, we add an inertial term $\ddot{y}(t)$ to (1.8) and we scale the terms $\frac{d}{dt}M(y(t))$ and $M(y(t))$ through positive functions $\mu(t)$ and $\gamma(t)$, which gives rise to the system

$$\ddot{y}(t) + \lambda\dot{y}(t) + \mu(t)\frac{d}{dt}M(y(t)) + \gamma(t)M(y(t)) = 0 \quad \text{for } t \geq t_0 \geq 0. \tag{1.12}$$

Since we have a constant viscous friction coefficient λ attached to $\dot{y}(t)$, this system can be seen as the Heavy Ball dynamics governed by a monotone and continuous operator M. Under a growth condition involving λ, $\mu(\cdot)$ and $\gamma(\cdot)$, we show that any solution $y(t)$ to this system fulfills $\|M(y(t))\| = o\left(\frac{1}{\mu(t)}\right)$ and converges weakly to a zero of M as $t \to +\infty$. Additionally, as the second main point of this paper, in Section 5 we will show that for an appropriate choice of $\mu(\cdot)$ and $\gamma(\cdot)$, (1.12) and (1.11) are time reparametrizations of each other. Every statement regarding (1.11) can be recovered from properties of (1.12). To put it succintly,

> The Fast OGDA dynamics can be recovered from the Heavy Ball dynamics for monotone equations through a time rescaling process.

2. Heavy Ball with friction dynamic from the time scaling perspective. As anticipated in the introduction, attached to the unconstrained minimization problem (1.1), we will study the asymptotic properties of the solutions to the Heavy Ball with Friction (HBF) dynamical system

$$\begin{cases} \ddot{y}(t) + \lambda\dot{y}(t) + b(t)\nabla f(y(t)) = 0 \quad \text{for } t \geq t_0 \geq 0, \\ y(t_0) = y_0, \quad \dot{y}(t_0) = y_1, \end{cases} \tag{HBF}$$

where $y_0, y_1 \in \mathcal{H}$. It will turn out that the friction parameter λ will play an important role when connecting (HBF) to other dynamics. First, let us collect the assumptions we will require in the optimization setting.

Assumptions in the optimization setting

The standing assumptions throughout Sections 2 and 3 are that $f : \mathcal{H} \to \mathbb{R}$ is convex and continuously differentiable, (1.1) has at least one solution, $\lambda > 0$, and $b : [t_0, +\infty[\to \mathbb{R}_{++}$ is continuously differentiable. At different points in our analysis we will a require one or a combination of the following further conditions:

($\mathcal{F}_1$) ∇f is Lipschitz continuous on bounded subsets of $\mathcal{H}$;

$(\mathcal{F}_2)$ $b : [t_0, +\infty[\to \mathbb{R}_{++}$ is nondecreasing and satisfies the growth condition

$$\sup_{t \geq t_0} \frac{\dot{b}(t)}{b(t)} < \lambda.$$

We lay out this section as follows: in Subsection 2.1, we show the existence and uniqueness of solutions to (HBF), and in Subsection 2.2 we discuss their asymptotic properties. We now introduce the energy function that will help in our analysis, and show an initial bound that will be needed later. Assume that $x_* \in \operatorname{argmin} f$, the set of minimizers of f, and that $t \mapsto y(t)$ solves (HBF) for $t \geq t_0$. For $0 \leq \eta \leq \lambda$ we define on $[t_0, +\infty)$

$$\mathcal{E}_\eta(t) := b(t)\left(f(y(t)) - \inf_{\mathcal{H}} f\right) + \frac{1}{2}\left\|\eta\left(y(t) - x_*\right) + \dot{y}(t)\right\|^2 + \frac{1}{2}\eta\left(\lambda - \eta\right)\left\|y(t) - x_*\right\|^2.$$

The time derivative of $\mathcal{E}_\eta(\cdot)$ at $t \geq t_0$ gives

$$\begin{aligned}\frac{d}{dt}\mathcal{E}_\eta(t) &= \dot{b}(t)\left(f(y(t)) - \inf_{\mathcal{H}} f\right) + b(t)\left\langle \nabla f(y(t)), \dot{y}(t)\right\rangle \\ &\quad + \left\langle \eta\left(y(t) - x_*\right) + \dot{y}(t), \eta\dot{y}(t) + \ddot{y}(t)\right\rangle + \eta\left(\lambda - \eta\right)\left\langle y(t) - x_*, \dot{y}(t)\right\rangle. \end{aligned} \tag{2.1}$$

According to the distributive property of the inner product and the equation (HBF) we have

$$\begin{aligned} & \left\langle \eta\left(y(t) - x_*\right) + \dot{y}(t), \eta\dot{y}(t) + \ddot{y}(t)\right\rangle \\ = \; & \left\langle \eta\left(y(t) - x_*\right) + \dot{y}(t), \lambda\dot{y}(t) + \ddot{y}(t)\right\rangle + (\eta - \lambda)\left\langle \eta\left(y(t) - x_*\right) + \dot{y}(t), \dot{y}(t)\right\rangle \\ = \; & - b(t)\left\langle \eta\left(y(t) - x_*\right) + \dot{y}(t), \nabla f(y(t))\right\rangle + \eta\left(\eta - \lambda\right)\left\langle y(t) - x_*, \dot{y}(t)\right\rangle + (\eta - \lambda)\left\|\dot{y}(t)\right\|^2. \end{aligned}$$

By plugging this expression into (2.1), we deduce that for every $t \geq t_0$

$$\begin{aligned}\frac{d}{dt}\mathcal{E}_\eta(t) &= \dot{b}(t)\left(f(y(t)) - \inf_{\mathcal{H}} f\right) - \eta b(t)\left\langle y(t) - x_*, \nabla f(y(t))\right\rangle + (\eta - \lambda)\left\|\dot{y}(t)\right\|^2 \\ &\leq \left(\dot{b}(t) - \eta b(t)\right)\left(f(y(t)) - \inf_{\mathcal{H}} f\right) + (\eta - \lambda)\left\|\dot{y}(t)\right\|^2, \end{aligned} \tag{2.2}$$

where the last inequality comes from the convexity of f.

2.1. Existence and uniqueness of global solutions.

We will rewrite (HBF) as a first-order system in two variables. It is simple to check that

$$\begin{cases} \ddot{y}(t) + \lambda\dot{y}(t) + b(t)\nabla f(y(t)) = 0, \\ y(t_0) = y_0, \quad \dot{y}(t_0) = y_1 \end{cases} \quad \text{is equivalent to} \quad \begin{cases} \dot{y}(t) & = u(t) - \lambda y(t), \\ \dot{u}(t) & = -b(t)\nabla f(y(t)), \\ y(t_0) & = y_0, \\ u(t_0) & = \lambda y_0 + y_1, \end{cases}$$

where $u(t) := \lambda y(t) + \dot{y}(t)$. We can write the system above in a more compact way, namely,

$$\begin{cases} \left(\dot{y}(t), \dot{u}(t)\right) & = G(t, y(t), u(t)) \\ \left(y(t_0), u(t_0)\right) & = \left(y_0, \; \lambda y_0 + y_1\right), \end{cases} \tag{2.3}$$

where $G : [t_0, +\infty[\times \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ is given by

$$G(t, y, u) := \left(u - \lambda y, \; -b(t)\nabla f(y)\right).$$

Notice that (2.2) suggests we need the growth condition $(\mathcal{F}_2)$ to obtain the decreasing property for $\mathcal{E}_\eta$.

THEOREM 2.1. *Suppose that $(\mathcal{F}_1)$ and $(\mathcal{F}_2)$ hold. Then, the dynamical system (HBF) admits a unique global solution $y : [t_0, +\infty[\to \mathcal{H}$.*

Proof. Since $b(\cdot)$ is continuously differentiable, it is Lipschitz continuous on the bounded subsets of $[t_0, +\infty[$. This, together with our other assumptions ensures that G is Lipschitz continuous on the bounded subsets of $[t_0, +\infty[\times \mathcal{H} \times \mathcal{H}$. According to [23, Theorems 46.2 and 46.3], the ordinary differential equation (2.3) admits a unique continuously differentiable solution $t \mapsto (y(t), u(t))$ defined on a maximal interval

$[t_0, T_{\max}[$, where

$$T_{\max} = +\infty \quad \text{or} \quad \Big\{ T_{\max} < +\infty \quad \text{and} \quad \lim_{t \to T_{\max}} \|(y(t), u(t))\| = +\infty \Big\}.$$

According to $(\mathcal{F}_2)$, there exists $0 < \eta_0 < \lambda$ such that

$$\frac{\dot{b}(t)}{b(t)} \leq \eta_0 < \lambda \quad \forall t \geq t_0.$$

We will now define the energy functional $\mathcal{E}_{\eta_0}(\cdot)$ restricted to the interval $[t_0, T_{\max}[$. According to the previous inequality and (2.2), we have for every $t \in [t_0, T_{\max}[$

$$\frac{d}{dt}\mathcal{E}_{\eta_0}(t) \leq \left(\sup_{t \geq t_0} \frac{\dot{b}(t)}{b(t)} - \eta_0 \right) b(t) \Big(f(y(t)) - \inf_{\mathcal{H}} f \Big) + (\eta_0 - \lambda) \big\| \dot{y}(t) \big\|^2 \leq 0.$$

This means that $\mathcal{E}_{\eta_0}(\cdot)$ is nonincreasing on $[t_0, T_{\max}[$. In particular, we obtain that for every $t \in [t_0, T_{\max}[$

$$\frac{1}{2} \big\| \eta_0 (y(t) - x_*) + \dot{y}(t) \big\|^2 + \frac{1}{2} \eta_0 (\lambda - \eta_0) \| y(t) - x_* \|^2 \leq \mathcal{E}_{\eta_0}(t) \leq \mathcal{E}_{\eta_0}(t_0).$$

This immediately yields that $t \mapsto \|y(t) - x_*\|$ and thus $t \mapsto \|y(t)\|$ are bounded on $[t_0, T_{\max}[$. Since we also obtain that $t \mapsto \|\eta_0(y(t) - x_*) + \dot{y}(t)\|$ is bounded on $[t_0, T_{\max}[$, using the triangle inequality, we obtain that $t \mapsto \big\|\dot{y}(t)\big\|$ and thus $t \mapsto \|u(t)\|$ are bounded on $[t_0, T_{\max}[$. So it must be $T_{\max} = +\infty$, which completes the proof of this theorem. ∎

2.2. Convergence rates for the function values and weak convergence of trajectories. From now on, we will assume that the existence of global solutions to (HBF) is given, since the following results do not require the assumption $(\mathcal{F}_1)$.

PROPOSITION 2.2. *Let $y : [t_0, +\infty[\to \mathcal{H}$ be a solution to* (HBF)*, and suppose that $(\mathcal{F}_2)$ holds. Then, the following statements are true:*

(i) (integrability results) *It holds*

$$\int_{t_0}^{+\infty} b(t) \left(f(y(t)) - \inf_{\mathcal{H}} f \right) dt < +\infty \quad \textit{and} \quad \int_{t_0}^{+\infty} \|\dot{y}(t)\|^2 \, dt < +\infty.$$

(ii) (energy functions convergence) *The limit $\lim_{t \to +\infty} \mathcal{E}_\eta(t) \in \mathbb{R}$ exists for every η satisfying $0 \leq \eta \leq \lambda$.*

Proof. (i) According to $(\mathcal{F}_2)$, there exists $\eta_0 > 0$ such that

$$\sup_{t \geq t_0} \frac{\dot{b}(t)}{b(t)} < \eta_0 < \lambda.$$

It follows from (2.2) that for every $t \geq t_0$

$$\frac{d}{dt}\mathcal{E}_{\eta_0}(t) \leq \left(\sup_{t \geq t_0} \frac{\dot{b}(t)}{b(t)} - \eta_0 \right) b(t) \left(f(y(t)) - \inf_{\mathcal{H}} f \right) + (\eta_0 - \lambda) \|\dot{y}(t)\|^2 . \tag{2.4}$$

The statement follows upon integration of this inequality.

(ii) Let $0 \leq \eta \leq \lambda$. From (2.2) we derive that for every $t \geq t_0$

$$\frac{d}{dt}\mathcal{E}_\eta(t) \leq (\lambda - \eta) b(t) \left(f(y(t)) - \inf_{\mathcal{H}} f \right).$$

The first statement in (i) ensures that the right hand side of this estimate belongs to $\mathbb{L}^1([t_0, +\infty[; \mathbb{R})$. Hence, the conclusion follows from Lemma A.2. ∎

For the purpose of studying the convergence and rate of convergence of (HBF), we introduce the function $W : [t_0, +\infty[\to \mathbb{R}_+$ defined for every $t \geq t_0$ as

$$W(t) := \left(f(y(t)) - \inf_{\mathcal{H}} f \right) + \frac{1}{2b(t)} \|\dot{y}(t)\|^2 .$$

The function is nonincreasing, which plays a crucial role in establishing convergence rates (see [13]). Indeed, for every $t \geq t_0$ we have

$$\dot{W}(t) = \langle \nabla f(y(t)), \dot{y}(t) \rangle - \frac{\dot{b}(t)}{2b^2(t)} \|\dot{y}(t)\|^2 + \frac{1}{b(t)} \langle \dot{y}(t), \ddot{y}(t) \rangle = -\frac{1}{b(t)} \left(\frac{\dot{b}(t)}{2b(t)} + \lambda \right) \|\dot{y}(t)\|^2 \leq 0, \tag{2.5}$$

which holds due to the fact that $\dot{b}(t) \geq 0$ for every $t \geq t_0$. We are now ready for the main convergence results for (HBF).

THEOREM 2.3. *Let $y : [t_0, +\infty[\to \mathcal{H}$ be a solution to* (HBF)*, and suppose that* ($\mathcal{F}_2$) *holds. Then, the following statements are true:*

(i) *(convergence rates) It holds*

$$W(t) = o\left(\frac{1}{\int_{\frac{t+t_0}{2}}^{t} b(r)\, dr} \right) \qquad \text{as } t \to +\infty.$$

In particular,

$$f(y(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{\int_{\frac{t+t_0}{2}}^{t} b(r)\, dr} \right) \qquad \text{and} \qquad \|\dot{y}(t)\| = o\left(\sqrt{\frac{b(t)}{\int_{\frac{t+t_0}{2}}^{t} b(r)\, dr}} \right) \qquad \text{as } t \to +\infty.$$

(ii) *(solution convergence) The solution $y(t)$ converges weakly to an element of* $\operatorname{argmin} f$ *as $t \to +\infty$.*

Proof. (i) From Proposition 2.2, we infer that $\int_{t_0}^{+\infty} b(t)W(t)dt < +\infty$. In particular, for $\psi : [t_0, +\infty[\to \mathbb{R}_+$ defined as $\psi(t) := \int_{t_0}^{t} b(r)\, W(r)\, dr$, the limit $\lim_{t\to+\infty} \psi(t) \in \mathbb{R}$ exists. Hence, for every $t \geq t_0$ and $r \in \left[\frac{t_0+t}{2}, t\right]$ we have, according to the decreasing property of $W(\cdot)$, that $W(t) \leq W(r)$ and thus

$$0 \leq W(t) \int_{\frac{t+t_0}{2}}^{t} b(r)\, dr \leq \int_{\frac{t+t_0}{2}}^{t} b(r)\, W(r)\, dr = \psi(t) - \psi\left(\frac{t+t_0}{2}\right) \to 0 \qquad \text{as } t \to +\infty,$$

which gives the desired small o rate for $W(\cdot)$.

(ii) Let $0 < \eta_1 < \eta_2 < \lambda$ and $x_* \in \operatorname{argmin} f$. We have for every $t \geq t_0$

$$\begin{aligned} \mathcal{E}_{\eta_2}(t) - \mathcal{E}_{\eta_1}(t) &= \frac{1}{2}(\eta_2 - \eta_1)\lambda \|y(t) - x_*\|^2 + (\eta_2 - \eta_1)\langle y(t) - x_*, \dot{y}(t) \rangle \\ &= (\eta_2 - \eta_1)\left(\frac{1}{2}\lambda \|y(t) - x_*\|^2 + \frac{d}{dt}\left(\frac{1}{2}\|y(t) - x_*\|^2 \right) \right). \end{aligned}$$

Proposition 2.2 (ii) guarantees that $\lim_{t\to+\infty} (\mathcal{E}_{\eta_2}(t) - \mathcal{E}_{\eta_1}(t)) \in \mathbb{R}$ exists. If we set $q(t) := \frac{1}{2}\|y(t) - x_*\|^2$, Lemma A.4 ensures that $\lim_{t\to+\infty} \frac{1}{2}\|y(t) - x_*\|^2$ exists and is a real number. This shows that the first condition of the Opial Lemma (see Lemma A.1) is verified. Moreover, the second condition is verified due to the weak lower semicontinuity of f and the convergence of $f(y(t))$ to $\inf_{\mathcal{H}} f$ as $t \to +\infty$ from (i). ∎

Let us illustrate the preceding results with two specific choices. For $\kappa > 0$ and $\rho \geq 0$, consider

$$b(t) := \kappa \exp(\rho t) \quad \text{and} \quad b(t) := \kappa t^{\rho},$$

respectively. Note that in both cases the classical Heavy Ball Method with friction is recovered by setting $(\kappa, \rho) = (1, 0)$, yielding a convergence rate of $o(1/t)$ for the function values along the solution [5]. The $\mathcal{O}(1/k)$ convergence rate for the discrete counterpart can be found in [24]. For these choices of $b(\cdot)$ we respectively have, for $t \geq t_0$ sufficiently large,

$$\begin{aligned} \int_{\frac{t+t_0}{2}}^{t} b(r)dr &= \frac{\kappa}{\rho}\left[\exp(\rho t) - \exp\left(\frac{\rho(t+t_0)}{2} \right) \right] = \frac{\kappa}{\rho}\exp(\rho t)\left[1 - \exp\left(\frac{\rho(t_0 - t)}{2} \right) \right] \geq \frac{\kappa}{2\rho}\exp(\rho t), \\ \int_{\frac{t+t_0}{2}}^{t} b(r)dt &= \frac{\kappa}{\rho+1}\left[t^{\rho+1} - \left(\frac{t+t_0}{2} \right)^{\rho+1} \right] = \frac{\kappa}{\rho+1}t^{\rho+1}\left[1 - \left(\frac{t+t_0}{2t} \right)^{\rho+1} \right] \geq \frac{\kappa}{4(\rho+1)}t^{\rho+1}, \end{aligned}$$

respectively. To comply with the growth condition ($\mathcal{F}_2$) for these choices for $b(\cdot)$, we respectively ask that

$$\lambda > \rho > 0 \quad \text{and} \quad \lambda > \frac{\rho}{t_0} \geq 0.$$

The previous estimates lead to the following corollaries, respectively:

COROLLARY 2.4. *Let $y : [t_0, +\infty[\to \mathcal{H}$ be a solution to*

$$\ddot{y}(t) + \lambda \dot{y}(t) + \kappa \exp(\rho t) \nabla f(y(t)) = 0 \quad \text{for } t \geq t_0 \geq 0.$$

If $\lambda > \rho > 0$, then the following statements are true:

(i) *(convergence rates) It holds*

$$f(y(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{\exp(\rho t)}\right) \quad \text{and} \quad \|\dot{y}(t)\| \to 0 \quad \text{as } t \to +\infty.$$

(ii) *(solution convergence) The solution $y(t)$ converges weakly to an element of* argmin f *as $t \to +\infty$.*

COROLLARY 2.5. *Let $y : [t_0, +\infty[\to \mathcal{H}$ be a solution to*

$$\ddot{y}(t) + \lambda \dot{y}(t) + \kappa t^{\rho} \nabla f(y(t)) = 0 \quad \text{for } t \geq t_0 > 0.$$

If $\lambda > \dfrac{\rho}{t_0} \geq 0$, then the following statements are true:

(i) *(convergence rates) It holds*

$$f(y(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^{\rho+1}}\right) \quad \text{and} \quad \|\dot{y}(t)\| = o\left(\frac{1}{\sqrt{t}}\right) \quad \text{as } t \to +\infty.$$

(ii) *(solution convergence) The solution $y(t)$ converges weakly to an element of* argmin f *as $t \to +\infty$.*

Remark 2.6. Previous statements show that one can get faster rates from the Heavy Ball Method (even linearly, as in Collorary 2.4). Nevertheless, we emphasize that these dynamics are not only interesting in terms of their acceleration phenomenon but also because they have a connection with other dynamics in the literature, for a nonclassical choice of $b(\cdot)$. This was one of the main points made in the introduction, which we elaborate upon in the next section. Naturally, this framework is only truly meaningful in the continuous-time setting, since one cannot expect gradient-type algorithms with unbounded step sizes—arising as explicit discretizations of time-rescaled systems—to converge. In contrast, implicit discretizations, which give rise to proximal-type methods, inherit the convergence properties of the continuous-time systems.

3. Connection with the dynamical system with asymptotically vanishing damping.

For $\alpha > 3$, we will study the convergence behavior of the second-order dynamical system with an Asymptotically Vanishing Damping (AVD)

$$\text{(AVD)} \qquad \begin{cases} \ddot{x}(s) + \dfrac{\alpha}{s}\dot{x}(s) + \nabla f(x(s)) = 0 \quad \text{for } s \geq s_0 > 0, \\ x(s_0) = x_0, \quad \dot{x}(s_0) = x_1, \end{cases}$$

where $x_0, x_1 \in \mathcal{H}$. More specifically, we will show that (AVD) can be derived from (HBF) via a time rescaling argument. This connection allows us to transfer the convergence results established for (HBF) to (AVD).

3.1. Two equivalent dynamical systems through time rescaling.

We start with a solution $y : [t_0, +\infty[\to \mathcal{H}$ to

$$\ddot{y}(t) + \lambda \dot{y}(t) + b(t) \nabla f(y(t)) = 0, \tag{3.1}$$

and define $x(s) := y(\tau(s))$, where $\tau : [s_0, +\infty[\to [t_0, +\infty[$ is a continuously differentiable function such that $\dot{\tau}(s) > 0$ for every $s \geq s_0 > 0$ and $\lim_{s\to+\infty} \tau(s) = +\infty$. We have

$$\dot{x}(s) = \dot{\tau}(s)\dot{y}(\tau(s)) \quad \text{and} \quad \ddot{x}(s) = \ddot{\tau}(s)\dot{y}(\tau(s)) + \left(\dot{\tau}(s)\right)^2 \ddot{y}(\tau(s)).$$

These expressions lead to

$$\dot{y}(\tau(s)) = \frac{1}{\dot{\tau}(s)}\dot{x}(s) \quad \text{and} \quad \ddot{y}(\tau(s)) = \frac{1}{\left(\dot{\tau}(s)\right)^2}\Big[\ddot{x}(s) - \ddot{\tau}(s)\dot{y}(\tau(s))\Big] = \frac{1}{\left(\dot{\tau}(s)\right)^2}\Big[\ddot{x}(s) - \frac{\ddot{\tau}(s)}{\dot{\tau}(s)}\dot{x}(s)\Big].$$

Now, plugging $t = \tau(s)$ in (3.1) for $s \geq s_0$ gives

$$\frac{1}{\left(\dot{\tau}(s)\right)^2}\left[\ddot{x}(s) - \frac{\ddot{\tau}(s)}{\dot{\tau}(s)}\dot{x}(s)\right] + \frac{\lambda}{\dot{\tau}(s)}\dot{x}(s) + b(\tau(s))\nabla f(x(s)) = 0$$

or, equivalently,

$$\ddot{x}(s) + \left[\lambda\dot{\tau}(s) - \frac{\ddot{\tau}(s)}{\dot{\tau}(s)}\right]\dot{x}(s) + \big(\dot{\tau}(s)\big)^2 b(\tau(s))\nabla f(x(s)) = 0. \tag{3.2}$$

Recall that the Su-Boyd-Candès dynamics [25] are given by

$$\ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + \nabla f(x(s)) = 0.$$

Going back to (3.2), to match the asymptotically vanishing damping accompanying the velocity we need the following to hold

$$\begin{cases} \lambda\dot{\tau}(s) - \dfrac{\ddot{\tau}(s)}{\dot{\tau}(s)} & = \dfrac{\alpha}{s} \\ \tau(s_0) & = t_0. \end{cases}$$

It is straightforward to check that

$$\tau(s) := \frac{\alpha - 1}{\lambda}\ln\left(\frac{s}{s_0}\right) + t_0$$

satisfies the previous differential equation, since

$$\dot{\tau}(s) = \frac{\alpha-1}{\lambda s} \quad \text{and} \quad \ddot{\tau}(s) = -\frac{\alpha-1}{\lambda s^2}.$$

We need to assume that $\alpha > 1$. Furthermore, we want the coefficient attached to $\nabla f(x(s))$ to be 1, i.e.,

$$\frac{(\alpha-1)^2}{\lambda^2 s^2} b(\tau(s)) = \big(\dot{\tau}(s)\big)^2 b(\tau(s)) = 1 \quad \Leftrightarrow \quad b\left(\frac{\alpha-1}{\lambda}\ln\left(\frac{s}{s_0}\right) + t_0\right) = \frac{\lambda^2 s^2}{(\alpha-1)^2},$$

which is fulfilled if we choose

$$b(t) = \left(\frac{\lambda s_0}{\alpha - 1}\right)^2 \exp\left(\frac{2\lambda(t-t_0)}{\alpha-1}\right).$$

Furthermore, we need $b(\cdot)$ to satisfy $(\mathcal{F}_2)$. Indeed, we have

$$\frac{\dot{b}(t)}{b(t)} = \frac{\left(\frac{2\lambda}{\alpha-1}\right)\exp\left(\frac{2\lambda(t-t_0)}{\alpha-1}\right)}{\exp\left(\frac{2\lambda(t-t_0)}{\alpha-1}\right)} = \frac{2\lambda}{\alpha-1} < \lambda \quad \Leftrightarrow \quad \alpha > 3.$$

All in all, with these choices, the solution $t \mapsto x(t)$ fulfills

$$\ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + \nabla f(x(s)) = 0.$$

Conversely, if for $\alpha > 3$, $x : [s_0, +\infty[\to \mathcal{H}$ is a solution to the previous system and we define $y(t) = x(\sigma(t))$, where $\sigma : [t_0, +\infty[\to [s_0, +\infty[$ is a continuously differentiable function such that $\dot{\sigma}(t) > 0$ for $t \geq t_0 > 0$ and $\lim_{t\to+\infty}\sigma(t) = +\infty$, arguing in a similar fashion as it was done previously we arrive at

$$\ddot{y}(t) + \left[\alpha\frac{\dot{\sigma}}{\sigma(t)} - \frac{\ddot{\sigma}(t)}{\dot{\sigma}(t)}\right]\dot{y}(t) + \big(\dot{\sigma}(t)\big)^2\nabla f(y(t)) = 0.$$

We want the coefficient attached to $\dot{y}(t)$ to be λ, i.e., we want $\sigma(\cdot)$ to satisfy the differential equation

$$\begin{cases} \alpha\dfrac{\dot{\sigma}(t)}{\sigma(t)} - \dfrac{\ddot{\sigma}(t)}{\dot{\sigma}(t)} & = \lambda \\ \sigma(t_0) & = s_0, \end{cases} \quad \text{which is fulfilled by} \quad \sigma(t) := s_0\exp\left(\frac{\lambda(t-t_0)}{\alpha-1}\right).$$

With this choice for $\sigma(\cdot)$, the resulting system reads

$$\ddot{y}(t) + \lambda\dot{y}(t) + s_0^2\exp\left(\frac{2\lambda(t-t_0)}{\alpha-1}\right)\nabla f(y(t)) = 0.$$

We have thus showed the following statement.

Proposition 3.1. *Assume that $\alpha > 3$, $\lambda > 0$ and that $s_0 > 0, t_0 \geq 0$ are initial times. Consider the following second-order systems:*

$$\begin{cases} \ddot{y}(t) + \lambda\dot{y}(t) + s_0^2\exp\left(\frac{2\lambda(t-t_0)}{\alpha-1}\right)\nabla f(y(t)) = 0, \\ y(t_0) = y_0, \quad \dot{y}(t_0) = y_1, \end{cases} \tag{3.3}$$

and

$$\begin{cases} \ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + \nabla f(x(s)) = 0, \\ x(s_0) = x_0, \quad \dot{x}(s_0) = x_1. \end{cases} \tag{3.4}$$

Then, the following statements are true:

(i) *If $y : [t_0, +\infty) \to \mathcal{H}$ is a solution to (3.3) and the function $\tau : [s_0, +\infty) \to [t_0, +\infty)$ is given by*

$$\tau(s) := \frac{\alpha - 1}{\lambda} \ln\left(\frac{s}{s_0}\right) + t_0,$$

then the reparametrized solution $x : [s_0, +\infty) \to \mathcal{H}$ given by $x(s) := y(\tau(s))$ is a solution of (3.4) for initial conditions

$$x(s_0) = y_0 \quad \textit{and} \quad \dot{x}(s_0) = \frac{\alpha - 1}{\lambda s_0} y_1.$$

(ii) *If $x : [s_0, +\infty) \to \mathcal{H}$ is a solution to (3.4) and the function $\sigma : [t_0, +\infty) \to [s_0, +\infty)$ is given by*

$$\sigma(t) := s_0 \exp\left(\frac{\lambda(t - t_0)}{\alpha - 1}\right),$$

then the reparametrized solution $y : [t_0, +\infty) \to \mathcal{H}$ given by $y(t) := x(\sigma(t))$ is a solution of (3.3) for initial conditions

$$y(t_0) = x_0 \quad \textit{and} \quad \dot{y}(t_0) = \frac{\lambda s_0}{\alpha - 1} x_1.$$

3.2. Transferring the convergence results to the (AVD) framework. As a direct corollary of Theorem 2.3 and Proposition 3.1, we obtain the following theorem.

THEOREM 3.2. *Let $\alpha > 3$ and $x : [s_0, +\infty[\to \mathcal{H}$ be a solution to*

$$\begin{cases} \ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + \nabla f(x(s)) = 0, \\ x(s_0) = x_0, \quad \dot{x}(s_0) = x_1. \end{cases}$$

Then, it holds

$$f(x(s)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{s^2}\right) \quad \textit{and} \quad \|\dot{x}(s)\| = o\left(\frac{1}{s}\right) \qquad \textit{as } s \to +\infty.$$

Furthermore, $x(s)$ converges weakly to an element of $\operatorname{argmin} f$ *as $s \to +\infty$.*

Proof. Let $\lambda > 0$ and $s_0 > 0$. As per Proposition 3.1, define $y(t) := x(\sigma(t))$. We know that $y : [t_0, +\infty) \to \mathcal{H}$ is a solution to (3.3). Taking into account the fact that $\sigma \circ \tau : [s_0, +\infty) \to [s_0, +\infty)$ is the identity function, we have $x(s) = y(\tau(s))$, and therefore $\frac{\lambda s}{\alpha - 1}\dot{x}(s) = \dot{y}(\tau(s))$. Thus, according to Theorem 2.3, we know that for $b(t) = s_0^2 \exp\left(\frac{2\lambda(t - t_0)}{\alpha - 1}\right)$, we have

$$f(x(s)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{\int_{\frac{t_0 + \tau(s)}{2}}^{\tau(s)} b(r)dr}\right) \quad \text{and} \quad \frac{\lambda s}{\alpha - 1}\|\dot{x}(s)\| = o\left(\sqrt{\frac{b(\tau(s))}{\int_{\frac{t_0 + \tau(s)}{2}}^{\tau(s)} b(r)dr}}\right) \quad \text{as} \quad s \to +\infty.$$

Notice that for every $s \geq s_0$ it holds

$$\begin{aligned} \int_{\frac{t_0 + \tau(s)}{2}}^{\tau(s)} b(r)dr &= s_0^2 \int_{\frac{t_0 + \tau(s)}{2}}^{\tau(s)} \exp\left(\frac{2\lambda(r - t_0)}{\alpha - 1}\right) dr \\ &= \frac{s_0^2(\alpha - 1)}{2\lambda}\left[\exp\left(\frac{2\lambda(\tau(s) - t_0)}{\alpha - 1}\right) - \exp\left(\frac{\lambda(\tau(s) - t_0)}{\alpha - 1}\right)\right] = \frac{s_0^2(\alpha - 1)}{2\lambda}\left[\left(\frac{s}{s_0}\right)^2 - \frac{s}{s_0}\right] \end{aligned}$$

and $b(\tau(s)) = s^2$. This gives $f(x(s)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{s^2}\right)$ and $\frac{\lambda s}{\alpha - 1}\|\dot{x}(s)\| \to 0$ as $s \to +\infty$, which verifies the rates we had claimed in the statement. Since $y(t)$ converges weakly to a global minimizer of f as $t \to +\infty$, so does $x(s) = y(\tau(s))$ as $s \to +\infty$. ∎

4. Heavy Ball dynamics governed by a maximally monotone and continuous operator. In this section, as anticipated in the introduction, we explore an analog to the Heavy Ball dynamics (HBF), but tailored to solve problem (1.6). We will study the asymptotic properties of the solutions to the Monotone Heavy Ball with Friction (M-HBF) dynamical system

$$\text{(M-HBF)} \qquad \begin{cases} \ddot{y}(t) + \lambda \dot{y}(t) + \mu(t) \dfrac{d}{dt} M(y(t)) + \gamma(t) M(y(t)) = 0 & \text{for } t \geq t_0 > 0, \\ y(t_0) = y_0, \quad \dot{y}(t_0) = y_1, \end{cases}$$

where $y_0, y_1 \in \mathcal{H}$. Let us collect the assumptions we will need in the monotone operator setting.

Assumptions in the monotone operator setting

The standing assumptions throughout Sections 4 and 5 are that $M : \mathcal{H} \to \mathcal{H}$ is monotone, single-valued and and continuous (thus, also maximally monotone), (1.6) has at least one solution, $\lambda > 0$, $\mu : [t_0, +\infty[\to \mathbb{R}_{++}$ is continuously differentiable and $\gamma : [t_0, +\infty[\to \mathbb{R}_{++}$ is continuous. At different points in our analysis we will require one or a combination of the following further conditions:

($\mathcal{M}_1$) M is L-Lipschitz continuous for some $L \geq 0$;

($\mathcal{M}_2$) $\mu(\cdot)$ is nondecreasing, and it holds

$$\lim_{t \to +\infty} \frac{\gamma(t)}{\mu(t)} =: L > 0, \quad \sup_{t \geq t_0} \frac{\dot{\mu}(t)}{\gamma(t)} < 1 \quad \text{and} \quad 2\lambda - 3L + \inf_{t \geq t_0} \frac{\dot{\mu}(t)}{\mu(t)} > 0.$$

When M is not cocoercive, which is the case for example when it does not arise from a convex potential, the presence of the "Hessian" term $\frac{d}{dt} M(y(t))$ in (M-HBF) is needed to ensure convergence rates for the function values and weak convergence results for the solution.

This section is organized as follows: in Subsection 4.1, we address the existence and uniqueness of strong global solutions to our system. In Subsection 4.2, we analyze the asymptotic properties of the global solutions to (M-HBF).

4.1. Existence and uniqueness of strong global solutions. We employ similar arguments to those used for the existence and uniqueness result for (HBF). First of all, we can rewrite (M-HBF) as a first-order system in two variables. Indeed, it is straightforward to check that

$$\begin{cases} \ddot{y}(t) + \lambda \dot{y}(t) + \mu(t) \frac{d}{dt} M(y(t)) + \gamma(t) M(y(t)) = 0, \\ y(t_0) = y_0, \quad \dot{y}(t_0) = y_1 \end{cases} \text{ is equivalent to } \begin{cases} \dot{y}(t) & = u(t) - \lambda y(t) - \mu(t) M(y(t)), \\ \dot{u}(t) & = \big(\dot{\mu}(t) - \gamma(t)\big) M(y(t)), \\ y(t_0) & = y_0, \\ u(t_0) & = \lambda y_0 + y_1 + \mu(t_0) M(y_0), \end{cases}$$

where $u(t) := \lambda y(t) + \dot{y}(t) + \mu(t) M(y(t))$, or

$$\begin{cases} \big(\dot{y}(t), \dot{u}(t)\big) & = G(t, y(t), u(t)), \\ (y(t_0), u(t_0)) & = \big(y_0,\ \lambda y_0 + y_1 + \mu(t_0) M(y_0)\big), \end{cases} \tag{4.1}$$

where $G : [t_0, +\infty) \times \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ is given by

$$G(t, y, u) := \Big(u - \lambda y - \mu(t) M(y),\ \big(\dot{\mu}(t) - \gamma(t)\big) M(y)\Big).$$

The existence and uniqueness of global solutions to (4.1) require strong assumptions, such as the Fréchet differentiability of M. It is more fitting to consider the existence and uniqueness of strong global solutions, which only require that M is Lipschitz continuous. We call $y : [t_0, +\infty) \to \mathcal{H}$ a strong global solution to (M-HBF) (see, for example, [12, Definition 3]) if the following statements hold:

(i) $y, \dot{y} : [t_0, +\infty) \to \mathcal{H}$ are locally absolutely continuous, that is, absolutely continuous on each interval $[t_0, t_1]$;

(ii) $\ddot{y}(t) + \lambda \dot{y}(t) + \mu(t) \frac{d}{dt} M(y(t)) + \gamma(t) M(y(t)) = 0$ for almost every $t \in [t_0, +\infty)$;

(iii) $y(t_0) = y_0$ and $\dot{y}(t_0) = y_1$.

THEOREM 4.1. *Suppose that ($\mathcal{M}_1$) holds. Then, the dynamical system (M-HBF) admits a unique strong global solution $y : [t_0, +\infty[\to \mathcal{H}$.*

Proof. We use the first-order reformulation (4.1), and show the existence of strong global solutions to this differential equation in the Hilbert space $\mathcal{H} \times \mathcal{H}$, which we endow with the inner product $\langle (y, u), (\overline{y}, \overline{u}) \rangle_{\mathcal{H} \times \mathcal{H}} :=$

$\langle y, \overline{y}\rangle + \langle u, \overline{u}\rangle$ and corresponding norm $\|(y,u)\|_{\mathcal{H}\times\mathcal{H}} := \sqrt{\|y\|^2 + \|u\|^2}$.

(a) First, we check the Lipschitz continuity of $G(t,\cdot,\cdot)$ for each $t \in [t_0, +\infty)$. For $(y,u), (\overline{y}, \overline{u}) \in \mathcal{H}\times\mathcal{H}$

$$\begin{aligned}
&\big\|G(t,y,u) - G(t,\overline{y},\overline{u})\big\|_{\mathcal{H}\times\mathcal{H}} \\
&\leq \Big\|(u-\overline{u}) - \lambda(y-\overline{y}) - \mu(t)(M(y) - M(\overline{y}))\Big\| + \Big\|(\dot{\mu}(t) - \gamma(t))(M(y) - M(\overline{y}))\Big\| \\
&\leq \|u-\overline{u}\| + \lambda\|y-\overline{y}\| + L|\mu(t)|\,\|y-\overline{y}\| + L\big|\dot{\mu}(t) - \gamma(t)\big|\|y-\overline{y}\| \\
&\leq \Big(1 + \lambda + L|\mu(t)| + L\big|\dot{\mu}(t) - \gamma(t)\big|\Big)\big\|(y,u) - (\overline{y},\overline{u})\big\|_{\mathcal{H}\times\mathcal{H}}.
\end{aligned}$$

By hypothesis, the (dependent on t) Lipschitz constant attached to $G(t,\cdot,\cdot)$ is locally integrable.

(b) Now, we prove the local integrability of $G(\cdot, y, u)$ for each $y, u \in \mathcal{H}$. We have

$$\begin{aligned}
\int_{t_0}^{t_1} \|G(s,y,u)\|_{\mathcal{H}\times\mathcal{H}} ds &\leq \int_{t_0}^{t_1} \Big(\Big\|u - \lambda y - \mu(s)M(y)\Big\| + \Big\|\big(\dot{\mu}(s) - \gamma(s)\big)M(y)\Big\|\Big) ds \\
&\leq \int_{t_0}^{t_1} \Big(\|u\| + \lambda\|y\| + |\mu(s)|\,\|M(y)\| + \big(\big|\dot{\mu}(s)\big| + \big|\gamma(s)\big|\big)\,\|M(y)\|\Big) ds,
\end{aligned}$$

and by hypothesis the right-hand side is finite.

With these two conditions verified, we invoke the Cauchy-Lipschitz-Picard theorem (see, for example, [16, Proposition 6.2.1]) to ensure the existence and uniqueness of a strong global solution to (4.1). The statement follows from the equivalence between the first- and second-order formulations. ∎

4.2. Convergence rates and weak convergence of the trajectories. In this subsection, we will assume that we have a global solution $y : [t_0, +\infty[\to \mathcal{H}$ of (M-HBF).

Before proceeding, we introduce the energy function that will help us in our analysis and show some estimates that we will need later. Suppose x_* is a zero of M, and $t \mapsto y(t)$ solves (M-HBF) for $t \geq t_0$. For $0 < \eta < \lambda$, we define

$$\begin{aligned}
&(\mathcal{E}^1_\eta(t)) && \mathcal{E}_\eta(t) := \frac{1}{2}\Big\|2\eta(y(t) - x_*) + 2\dot{y}(t) + \mu(t)M(y(t))\Big\|^2 \\
&(\mathcal{E}^2_\eta(t)) && \qquad + 2\eta(\lambda - \eta)\|y(t) - x_*\|^2 \\
&(\mathcal{E}^3_\eta(t)) && \qquad + 2\eta\mu(t)\langle y(t) - x_*, M(y(t))\rangle \\
&(\mathcal{E}^4_\eta(t)) && \qquad + \frac{1}{2}\mu^2(t)\|M(y(t))\|^2.
\end{aligned}$$

We now compute the time derivative of $\mathcal{E}_\eta(\cdot)$ at a point $t \geq t_0$

$$\begin{aligned}
\frac{d}{dt}\mathcal{E}^1_\eta(t) &= \Big\langle 2\eta(y(t) - x_*) + 2\dot{y}(t) + \mu(t)M(y(t)),\ 2\eta\dot{y}(t) + 2\ddot{y}(t) + \dot{\mu}(t)M(y(t)) + \mu(t)\frac{d}{dt}M(y(t))\Big\rangle \\
&= \Big\langle 2\eta(y(t) - x_*) + 2\dot{y}(t) + \mu(t)M(y(t)), \\
&\qquad\qquad 2(\eta - \lambda)\dot{y}(t) + \Big[\dot{\mu}(t) - 2\gamma(t)\Big]M(y(t)) - \mu(t)\frac{d}{dt}M(y(t))\Big\rangle \\
&= 4\eta(\eta - \lambda)\langle y(t) - x_*, \dot{y}(t)\rangle + 2\eta\Big[\dot{\mu}(t) - 2\gamma(t)\Big]\langle y(t) - x_*, M(y(t))\rangle \\
&\quad - 2\eta\mu(t)\left\langle y(t) - x_*, \frac{d}{dt}M(y(t))\right\rangle + 4(\eta - \lambda)\big\|\dot{y}(t)\big\|^2 \\
&\quad + \Big\{2\Big[\dot{\mu}(t) - 2\gamma(t)\Big] + 2(\eta - \lambda)\mu(t)\Big\}\langle \dot{y}(t), M(y(t))\rangle - 2\mu(t)\left\langle \dot{y}(t), \frac{d}{dt}M(y(t))\right\rangle \\
&\quad + \mu(t)\Big[\dot{\mu}(t) - 2\gamma(t)\Big]\|M(y(t))\|^2 - \mu^2(t)\left\langle M(y(t)), \frac{d}{dt}M(y(t))\right\rangle, \\
\frac{d}{dt}\mathcal{E}^2_\eta(t) &= 4\eta(\lambda - \eta)\langle y(t) - x_*, \dot{y}(t)\rangle,
\end{aligned}$$

$$\frac{d}{dt}\mathcal{E}^3_\eta(t) = 2\eta\dot{\mu}(t)\langle y(t)-x_*, M(y(t))\rangle + 2\eta\mu(t)\langle \dot{y}(t), M(y(t))\rangle + 2\eta\mu(t)\left\langle y(t)-x_*, \frac{d}{dt}M(y(t))\right\rangle,$$
$$\frac{d}{dt}\mathcal{E}^4_\eta(t) = \mu(t)\dot{\mu}(t)\|M(y(t))\|^2 + \mu^2(t)\left\langle M(y(t)), \frac{d}{dt}M(y(t))\right\rangle.$$

Putting everything together yields

$$\begin{aligned}
\frac{d}{dt}\mathcal{E}_\eta(t) &= \frac{d}{dt}\mathcal{E}^1_\eta(t) + \frac{d}{dt}\mathcal{E}^2_\eta(t) + \frac{d}{dt}\mathcal{E}^3_\eta(t) + \frac{d}{dt}\mathcal{E}^4_\eta(t) \\
&= \Big\{2\eta\Big[\dot{\mu}(t) - 2\gamma(t)\Big] + 2\eta\dot{\mu}(t)\Big\}\langle y(t)-x_*, M(y(t))\rangle + 4(\eta-\lambda)\big\|\dot{y}(t)\big\|^2 \\
&\quad + \Big\{2\Big[\dot{\mu}(t) - 2\gamma(t)\Big] + 2(\eta-\lambda)\mu(t) + 2\eta\mu(t)\Big\}\langle \dot{y}(t), M(y(t))\rangle - 2\mu(t)\left\langle \dot{y}(t), \frac{d}{dt}M(y(t))\right\rangle \\
&\quad + \Big\{\mu(t)\Big[\dot{\mu}(t) - 2\gamma(t)\Big] + \mu(t)\dot{\mu}(t)\Big\}\|M(y(t))\|^2 \\
&= 4\eta\Big[\dot{\mu}(t) - \gamma(t)\Big]\langle y(t)-x_*, M(y(t))\rangle + 2\mu(t)\Big[\dot{\mu}(t) - \gamma(t)\Big]\|M(y(t))\|^2 \\
&\quad + 4(\eta-\lambda)\big\|\dot{y}(t)\big\|^2 - 2\mu(t)\left\langle \dot{y}(t), \frac{d}{dt}M(y(t))\right\rangle \\
&\quad + 2\Big\{\Big[2(\eta-\lambda)\mu(t) + \lambda\mu(t) - 2\gamma(t)\Big] + \dot{\mu}(t)\Big\}\langle \dot{y}(t), M(y(t))\rangle \\
&= 4\eta\Big[\dot{\mu}(t) - \gamma(t)\Big]\langle y(t)-x_*, M(y(t))\rangle + \frac{2}{3}\mu(t)\Big[\dot{\mu}(t) - \gamma(t)\Big]\|M(y(t))\|^2 \\
&\quad + (\eta-\lambda)\big\|\dot{y}(t)\big\|^2 - 2\mu(t)\left\langle \dot{y}(t), \frac{d}{dt}M(y(t))\right\rangle + \mathcal{R}(t),
\end{aligned} \tag{4.2}$$

where $\mathcal{R}(t)$ results from rearranging terms and will be made explicit momentarily. Our goal is to have $\frac{d}{dt}\mathcal{E}_\eta(t) \leq 0$ for t large enough, which will be ensured under the assumption $(\mathcal{M}_2)$. Indeed, the growth condition on $\mu(\cdot)$ and $\gamma(\cdot)$ and the positivity of $\mu(\cdot)$ ensure $\dot{\mu}(t) \leq \gamma(t)$, which makes the coefficients accompanying $\langle y(t)-x_*, M(y(t))\rangle$ and $\|M(y(t))\|^2$ nonpositive; the fact that $\eta < \lambda$ also makes the term corresponding to $\big\|\dot{y}(t)\big\|^2$ nonpositive. Furthermore, the monotonicity of M yields

$$\left\langle \dot{y}(t), \frac{d}{dt}M(y(t))\right\rangle = \lim_{s\to t}\frac{1}{(s-t)^2}\Big\langle y(s)-y(t), M(y(s)) - M(y(t))\Big\rangle \geq 0,$$

further multiplication by $-2\mu(t)$ makes this term nonpositive as well.

It only remains to check that $\mathcal{R}(t)$ is nonpositive as well, which requires some preliminary computations. Define, $\varepsilon := \lambda - \eta > 0$, so now $\mathcal{R}(t)$ reads

$$\mathcal{R}(t) = -3\varepsilon\big\|\dot{y}(t)\big\|^2 + 2\Big\{\Big[-2\varepsilon\mu(t) + \lambda\mu(t) - 2\gamma(t)\Big] + \dot{\mu}(t)\Big\}\langle \dot{y}(t), M(y(t))\rangle + \frac{4}{3}\mu(t)\Big[\dot{\mu}(t) - \gamma(t)\Big]\|M(y(t))\|^2. \tag{4.3}$$

To establish that $\mathcal{R}(t)$ is nonpositive for large enough t, we make use of Lemma A.3 and set $X := \dot{y}(t)$, $Y := M(y(t))$ and A, B, C as

$$A := -3\varepsilon, \quad B := \big(-2\varepsilon\mu(t) + \lambda\mu(t) - 2\gamma(t)\big) + \dot{\mu}(t), \quad C := \frac{4}{3}\mu(t)\Big[\dot{\mu}(t) - \gamma(t)\Big].$$

We have

$$\begin{aligned}
B^2 - AC &= \Big[\big(-2\varepsilon\mu(t) + \lambda\mu(t) - 2\gamma(t)\big) + \dot{\mu}(t)\Big]^2 + 3\cdot\frac{4}{3}\varepsilon\mu(t)\Big[\dot{\mu}(t) - \gamma(t)\Big] \\
&= \Big[-2\varepsilon\mu(t) + \lambda\mu(t) - 2\gamma(t)\Big]^2 - 4\varepsilon\mu(t)\dot{\mu}(t) + 2\big(\lambda\mu(t) - 2\gamma(t)\big)\dot{\mu}(t) + \big(\dot{\mu}(t)\big)^2 \\
&\quad + 4\varepsilon\mu(t)\dot{\mu}(t) - 4\varepsilon\mu(t)\gamma(t) \\
&= 4\varepsilon^2\mu^2(t) - 4\varepsilon\mu(t)\big(\lambda\mu(t) - 2\gamma(t)\big) + \big(\lambda\mu(t) - 2\gamma(t)\big)^2 + 2\big(\lambda\mu(t) - 2\gamma(t)\big)\dot{\mu}(t) \\
&\quad + \big(\dot{\mu}(t)\big)^2 - 4\varepsilon\mu(t)\gamma(t)
\end{aligned}$$

$$= \mu^2(t)\left\{4\varepsilon^2 - 4\left(\lambda - \frac{\gamma(t)}{\mu(t)}\right)\varepsilon + \left[\lambda - \frac{2\gamma(t)}{\mu(t)} + \frac{\dot{\mu}(t)}{\mu(t)}\right]^2\right\}. \tag{4.4}$$

Notice that $B^2 - AC$ is a quadratic expression in ε. Proposition 4.3, which is an intermediate result before the main theorem of this subsection, will show that there exists a suitable interval $I \subseteq]0, \lambda[$ with nonempty interior such that this quadratic expression eventually becomes negative for every $\varepsilon \in I$. It also shows integral statements which will be needed when showing weak convergence of trajectories and small o rates.

Remark 4.2. Assumption $(\mathcal{M}_2)$ implies in particular that $\lambda > L$, a fact that we will need later. Let us show this claim here. Since

$$2\lambda > 3L - \inf_{t\geq t_0} \frac{\dot{\mu}(t)}{\mu(t)}, \text{ there exists } \delta > 0 \text{ such that } 2\lambda - \delta \geq 3L - \inf_{t\geq t_0} \frac{\dot{\mu}(t)}{\mu(t)}.$$

Since $\lim_{t\to+\infty} \frac{\gamma(t)}{\mu(t)} = L$, for every given $0 < \varepsilon < \delta$ there exists $t_\varepsilon \geq t_0$ such that for every $t \geq t_\varepsilon$

$$L - \varepsilon \leq \frac{\gamma(t)}{\mu(t)} \leq L + \varepsilon, \text{ and in particular, } \frac{\dot{\mu}(t)}{\mu(t)} \leq (L+\varepsilon)\frac{\dot{\mu}(t)}{\gamma(t)} \leq L + \varepsilon,$$

where in the last estimate we used the assumption $\sup_{t\geq t_0} \frac{\dot{\mu}(t)}{\gamma(t)} < 1$ and the nondecreasing property of μ. Now, for every $t \geq t_\varepsilon$ we have

$$2\lambda - \delta \geq 3L - \inf_{t\geq t_0} \frac{\dot{\mu}(t)}{\mu(t)} \geq 3L - \frac{\dot{\mu}(t)}{\mu(t)} \geq 3L - L - \varepsilon = 2L - \varepsilon,$$

from which we deduce

$$2\lambda > 2L + \delta - \varepsilon > 2L \text{ and thus } \lambda > L.$$

PROPOSITION 4.3. *Let x_* be a zero of M and $y : [t_0, +\infty[\to \mathcal{H}$ be a solution to* (M-HBF)*, and suppose that $(\mathcal{M}_2)$ holds. Then, the following statements are true:*

(i) (energy functions convergence) *The limit $\lim_{t\to+\infty} \mathcal{E}_{\lambda-\varepsilon}(t) \in \mathbb{R}$ exists for every ε in a given interval $I \subseteq]0, \lambda[$.*

(ii) (integrability results) *It holds*

$$\int_{t_0}^{+\infty} \mu(t)\langle y(t) - x_*, M(y(t))\rangle dt < +\infty, \int_{t_0}^{+\infty} \left\|\dot{y}(t)\right\|^2 dt < +\infty, \int_{t_0}^{+\infty} \mu^2(t)\|M(y(t))\|^2 dt < +\infty.$$

Proof. (i) Just as we mentioned before the previous remark, we want to find an interval I with nonempty interior, independent of time, such that $I \subseteq]0, \lambda[$ and such that for every $\varepsilon \in I$ the quadratic expression (4.4) becomes negative for large enough t. We do this in two steps: first, we show that the discriminant, which depends on t, stays away from zero as $t \to +\infty$, and thus the (also dependent on t) roots are properly separated as $t \to +\infty$. Second, we show that the left and right roots stay respectively below and above certain thresholds as $t \to +\infty$. Combining these facts will yield the desired interval I.
To proceed, we first observe that, according to $(\mathcal{M}_2)$, and since $\lambda > L$ following Remark 4.2, we can choose δ_1, δ_2 such that

$$\max\left\{2 - \frac{\lambda}{L}, \sup_{t\geq t_0} \frac{\dot{\mu}(t)}{\gamma(t)}\right\} < \delta_1 < \delta_2 < 1. \tag{4.5}$$

After some algebraic manipulation, we deduce that

$$\lambda - \frac{2\gamma(t)}{\mu(t)} + \frac{\dot{\mu}(t)}{\mu(t)} < \lambda + (\delta_1 - 2)\frac{\gamma(t)}{\mu(t)} < \lambda + (\delta_2 - 2)\frac{\gamma(t)}{\mu(t)} < \lambda - \frac{\gamma(t)}{\mu(t)} \quad \forall t \geq t_0.$$

Moreover, we know that there exist $\tilde{\delta}_2 > \tilde{\delta}_1 > 0$ satisfying

$$0 < \lambda + (\delta_1 - 2)L < \tilde{\delta}_1 < \tilde{\delta}_2 < \lambda + (\delta_2 - 2)L. \tag{4.6}$$

Using $\lim_{t\to+\infty} \frac{\gamma(t)}{\mu(t)} = L > 0$, we know that there exist $t_1 \geq t_0$ such that

$$\lambda - \frac{2\gamma(t)}{\mu(t)} + \frac{\dot{\mu}(t)}{\mu(t)} < \lambda + (\delta_1 - 2)\frac{\gamma(t)}{\mu(t)} \leq \tilde{\delta}_1 < \tilde{\delta}_2 \leq \lambda + (\delta_2 - 2)\frac{\gamma(t)}{\mu(t)} < \lambda - \frac{\gamma(t)}{\mu(t)} \quad \forall t \geq t_1. \tag{4.7}$$

On the other hand, according to ($\mathcal{M}_2$), we know that

$$L - \lambda < \lambda - 2L + \inf_{t \geq t_0} \frac{\dot{\mu}(t)}{\mu(t)}.$$

Again recalling that $\lambda > L$, we can construct $\tilde{\delta}_4 > \tilde{\delta}_3 > 0$ and $t_2 \geq t_1$ such that

$$L - \lambda < -\tilde{\delta}_4 < -\tilde{\delta}_3 < \lambda - 2L + \inf_{t \geq t_0} \frac{\dot{\mu}(t)}{\mu(t)} \quad \text{and}$$

$$\frac{\gamma(t)}{\mu(t)} - \lambda \leq -\tilde{\delta}_4 < -\tilde{\delta}_3 \leq \lambda - \frac{2\gamma(t)}{\mu(t)} + \inf_{t \geq t_0} \frac{\dot{\mu}(t)}{\mu(t)} \leq \lambda - \frac{2\gamma(t)}{\mu(t)} + \frac{\dot{\mu}(t)}{\mu(t)} \quad \forall t \geq t_2. \tag{4.8}$$

Combining (4.7) and (4.8) yields

$$\frac{\gamma(t)}{\mu(t)} - \lambda \leq -\tilde{\delta}_4 < -\tilde{\delta}_3 \leq \lambda - \frac{2\gamma(t)}{\mu(t)} + \frac{\dot{\mu}(t)}{\mu(t)} \leq \tilde{\delta}_1 < \tilde{\delta}_2 \leq \lambda - \frac{\gamma(t)}{\mu(t)} \quad \forall t \geq t_2.$$

Without loss of generality, since both cases can be handed identically, we may assume that

$$\max\left\{\tilde{\delta}_1, \tilde{\delta}_3\right\} = \tilde{\delta}_1.$$

Therefore, the reduced discriminant of (4.4) satisfies

$$\Delta_t := 4\left[\left(\lambda - \frac{\gamma(t)}{\mu(t)}\right)^2 - \left(\lambda - \frac{2\gamma(t)}{\mu(t)} + \frac{\dot{\mu}(t)}{\mu(t)}\right)^2\right] \geq 4\left(\tilde{\delta}_2^2 - \tilde{\delta}_1^2\right) > 0 \quad \forall t \geq t_2, \tag{4.9}$$

and thus stays away from zero as $t \to +\infty$. Now, choose $\delta_3 > 0$ such that

$$\delta_3 < \min\left\{L,\ \lambda - L,\ \frac{1}{2}\sqrt{\tilde{\delta}_2^2 - \tilde{\delta}_1^2}\right\}.$$

Once again using the fact that $\lim_{t\to+\infty} \frac{\gamma(t)}{\mu(t)} = L$, there exists $t_3 \geq t_2$ such that

$$\lambda - L - \delta_3 \leq \lambda - \frac{\gamma(t)}{\mu(t)} \leq \lambda - L + \delta_3 \quad \forall t \geq t_3. \tag{4.10}$$

The roots of the quadratic expression (in ε) in equation (4.4) are

$$\underline{\varepsilon}_t := \frac{1}{2}\left(\lambda - \frac{\gamma(t)}{\mu(t)}\right) - \frac{1}{4}\sqrt{\Delta_t} \quad \text{and} \quad \overline{\varepsilon}_t := \frac{1}{2}\left(\lambda - \frac{\gamma(t)}{\mu(t)}\right) + \frac{1}{4}\sqrt{\Delta_t}.$$

Now, using (4.9) and (4.10) we can deduce

$$\delta_3 < \frac{1}{2}\sqrt{\tilde{\delta}_2^2 - \tilde{\delta}_1^2} \quad \Rightarrow \quad \lambda - L + \delta_3 - \sqrt{\tilde{\delta}_2^2 - \tilde{\delta}_1^2} < \lambda - L - \delta_3$$

$$\Rightarrow \quad \underline{\varepsilon}_t = \frac{1}{2}\left(\lambda - \frac{\gamma(t)}{\mu(t)}\right) - \frac{1}{4}\sqrt{\Delta_t} < \frac{1}{2}\left(\lambda - L + \delta_3 - \sqrt{\tilde{\delta}_2^2 - \tilde{\delta}_1^2}\right) < \frac{1}{2}\left(\lambda - L - \delta_3\right) \quad \forall t \geq t_3.$$

Similarly, we obtain

$$\frac{1}{2}\left(\lambda - L + \delta_3\right) < \overline{\varepsilon}_t \quad \forall t \geq t_3.$$

Thus, we define the interval I as

$$I := \left[\frac{1}{2}\left(\lambda - L - \delta_3\right), \frac{1}{2}\left(\lambda - L + \delta_3\right)\right] \subseteq \left]\underline{\varepsilon}_t, \overline{\varepsilon}_t\right[\quad \forall t \geq t_3.$$

The way we chose δ_3 also ensures that

$$I \subseteq \left]0, \lambda\right[.$$

Thus, for every $\varepsilon \in I$, the term (4.4) becomes negative, i.e.,

$$B^2 - AC < 0 \quad \text{for every } \varepsilon \in I \text{ and every } t \geq t_3,$$

and we recall that according to Lemma A.3, this in turn implies $\mathcal{R}(t) \leq 0$ for all $t \geq t_3$, which is what we wanted. Having shown (4.3), we go back to (4.2) and we drop the nonpositive terms corresponding to $\left\langle \dot{y}(t), \frac{d}{dt}M(y(t))\right\rangle$ and $\mathcal{R}(t)$. For every $\eta = \lambda - \varepsilon > 0$ with $\varepsilon \in I$, it holds

$$\frac{d}{dt}\mathcal{E}_\eta(t) \leq 4\eta\Big[\dot{\mu}(t) - \gamma(t)\Big]\langle y(t) - x_*, M(y(t))\rangle + (\eta - \lambda)\left\|\dot{y}(t)\right\|^2 + \frac{2}{3}\mu(t)\Big[\dot{\mu}(t) - \gamma(t)\Big]\left\|M(y(t))\right\|^2$$

$$\leq 0 \quad \forall t \geq t_3. \tag{4.11}$$

This means that for every $\eta = \lambda - \varepsilon$, where $\varepsilon \in I$, $\mathcal{E}_\eta(\cdot)$ is nonincreasing on $[t_3, +\infty[$, therefore

$$0 \leq \mathcal{E}_\eta(t) \leq \mathcal{E}_\eta(t_3) \quad \text{for } t \geq t_3 \tag{4.12}$$

and

$$\lim_{t\to+\infty} \mathcal{E}_\eta(t) \in \mathbb{R} \text{ exists.} \tag{4.13}$$

In particular, going back to ($\mathcal{E}_\eta^1(t)$)-($\mathcal{E}_\eta^4(t)$), we deduce that

$$\|M(y(t))\| \leq \frac{\sqrt{2\mathcal{E}_\eta(t_3)}}{\mu(t)} \quad \text{and} \quad \langle y(t) - x_*, M(y(t))\rangle \leq \frac{\mathcal{E}_\eta(t_3)}{2\eta\mu(t)} \quad \forall t \geq t_3. \tag{4.14}$$

(ii) From (4.5) and (4.10), for every $t \geq t_3$ we have

$$\frac{\dot\mu(t)}{\gamma(t)} \leq \delta_1 < 1 \quad \text{and} \quad L - \delta_3 \leq \frac{\gamma(t)}{\mu(t)} \leq L + \delta_3 \quad \Rightarrow \quad (1-\delta_1)\gamma(t) \leq \gamma(t) - \dot\mu(t)$$
$$\Rightarrow \quad (1-\delta_1)(L-\delta_3)\mu(t) \leq (1-\delta_1)\gamma(t) \leq \gamma(t) - \dot\mu(t).$$

Going back to (4.11), we integrate the inequality from t_3 to $t \geq t_3$ and obtain

$$\begin{aligned}
&4\eta(L-\delta_3)(1-\delta_1)\int_{t_3}^t \mu(s)\langle y(s) - x_*, M(y(s))\rangle ds + (\lambda - \eta)\int_{t_3}^t \left\|\dot y(s)\right\|^2 ds \\
&+ \frac{2}{3}(L-\delta_3)(1-\delta_1)\int_{t_3}^t \mu^2(s)\|M(y(s))\|^2 ds \\
\leq\ & 4\eta\int_{t_3}^t \Big[\gamma(s) - \dot\mu(s)\Big]\langle y(s) - x_*, M(y(s))\rangle ds + (\lambda-\eta)\int_{t_3}^t \left\|\dot y(s)\right\|^2 ds \\
&+ \frac{2}{3}\int_{t_3}^t \mu(s)\Big[\gamma(s) - \dot\mu(s)\Big]\|M(y(s))\|^2 ds \\
\leq\ & \mathcal{E}_\eta(t_3) - \mathcal{E}_\eta(t),
\end{aligned}$$

which produces the required integrability results. ◻

We are now in a position to state and prove the main theorem of this section.

THEOREM 4.4. *Let x_* be a zero of M and $y : [t_0, +\infty[\to \mathcal{H}$ be a solution to* (M-HBF)*, and suppose that* ($\mathcal{M}_2$) *holds. Then, the following statements are true:*

(i) *(convergence rates) It holds*

$$\|M(y(t))\| = o\left(\frac{1}{\mu(t)}\right), \quad \langle y(t) - x_*, M(y(t))\rangle = o\left(\frac{1}{\mu(t)}\right), \quad \left\|\dot y(t)\right\| \to 0 \quad \text{as } t \to +\infty. \tag{4.15}$$

(ii) *(solution convergence) The solution $y(t)$ converges weakly to a zero of M as $t \to +\infty$.*

Proof. (i) Let $I \subseteq]0, \lambda[$ be the interval provided by Proposition 4.3 (i). Choose $\varepsilon_1, \varepsilon_2 \in I$ such that $\varepsilon_1 \neq \varepsilon_2$. Set $\eta_i := \lambda - \varepsilon_i$, $i = 1, 2$. We have

$$\begin{aligned}
\mathcal{E}_{\eta_2}(t) - \mathcal{E}_{\eta_1}(t) =\ & \frac{1}{2}\Big\|2\eta_2(y(t) - x_*) + 2\dot y(t) + \mu(t)M(y(t))\Big\|^2 - \frac{1}{2}\Big\|2\eta_1(y(t) - x_*) + 2\dot y(t) + \mu(t)M(y(t))\Big\|^2 \\
&+ 2\Big[\eta_2(\lambda - \eta_2) - \eta_1(\lambda - \eta_1)\Big]\|y(t) - x_*\|^2 + 2(\eta_2 - \eta_1)\mu(t)\langle y(t) - x_*, M(y(t))\rangle \\
=\ & 2\big(\eta_2^2 - \eta_1^2\big)\|y(t) - x_*\|^2 + 2(\eta_2 - \eta_1)\Big\langle y(t) - x_*, 2\dot y(t) + \mu(t)M(y(t))\Big\rangle \\
&+ \Big[2\lambda(\eta_2 - \eta_1) - 2\big(\eta_2^2 - \eta_1^2\big)\Big]\|y(t) - x_*\|^2 + 2(\eta_2 - \eta_1)\mu(t)\langle y(t) - x_*, M(y(t))\rangle \\
=\ & 4(\eta_2 - \eta_1)\left[\frac{\lambda}{2}\|y(t) - x_*\|^2 + \Big\langle y(t) - x_*, \dot y(t) + \mu(t)M(y(t))\Big\rangle\right].
\end{aligned}$$

Define, for $t \geq t_0$,

$$p(t) := \frac{\lambda}{2}\|y(t) - x_*\|^2 + \Big\langle y(t) - x_*, \dot y(t) + \mu(t)M(y(t))\Big\rangle.$$

Because of (4.13), and since $\eta_2 - \eta_1 \neq 0$, we deduce that

$$\lim_{t\to+\infty} p(t) \in \mathbb{R} \text{ exists.}$$

With this at hand, we can now rewrite $\mathcal{E}_{\eta_1}(t)$ for every $t \geq t_0$ as

$$\begin{aligned}
\mathcal{E}_{\eta_1}(t) = & \frac{1}{2}\Big\|2\eta_1(y(t)-x_*) + 2\dot{y}(t) + \mu(t)M(y(t))\Big\|^2 + 2\eta_1(\lambda-\eta_1)\|y(t)-x_*\|^2 \\
& + 2\eta_1\mu(t)\langle y(t)-x_*, M(y(t))\rangle + \frac{1}{2}\mu^2(t)\|M(y(t))\|^2 \\
= & \frac{1}{2}\left[4\eta_1^2\|y(t)-x_*\|^2 + 4\eta_1\Big\langle y(t)-x_*, 2\dot{y}(t)+\mu(t)M(y(t))\Big\rangle + \Big\|2\dot{y}(t)+\mu(t)M(y(t))\Big\|^2\right] \\
& + 2\eta_1\lambda\|y(t)-x_*\|^2 - 2\eta_1^2\|y(t)-x_*\|^2 + 2\eta_1\mu(t)\langle y(t)-x_*, M(y(t))\rangle + \frac{1}{2}\mu^2(t)\|M(y(t))\|^2 \\
= & 4\eta_1 p(t) + \frac{1}{2}\Big\|2\dot{y}(t)+\mu(t)M(y(t))\Big\|^2 + \frac{1}{2}\mu^2(t)\|M(y(t))\|^2 \\
= & 4\eta_1 p(t) + \Big\|\dot{y}(t)+\mu(t)M(y(t))\Big\|^2 + \big\|\dot{y}(t)\big\|^2.
\end{aligned}$$

Define

$$h(t) := \Big\|\dot{y}(t)+\mu(t)M(y(t))\Big\|^2 + \big\|\dot{y}(t)\big\|^2.$$

Since both $\lim_{t\to+\infty}\mathcal{E}_{\eta_1}(t)$ and $\lim_{t\to+\infty} p(t)$ exist, so does $\lim_{t\to+\infty} h(t)$. Notice that

$$\int_{t_0}^t h(s)ds \leq 3\int_{t_0}^t \big\|\dot{y}(s)\big\|^2 ds + 2\int_{t_0}^t \mu^2(s)\|M(y(s))\|^2 ds,$$

and the integrals on the right-hand side remain finite as $t \to +\infty$ according to Proposition 4.3 (i) (ii), i.e., $\int_{t_0}^{+\infty} h(t)dt < +\infty$. Combining this with the existence of $\lim_{t\to+\infty} h(t)$ allows us to deduce

$$\lim_{t\to+\infty} h(t) = 0.$$

This implies

$$\lim_{t\to+\infty} \|\dot{y}(t)\| = 0,$$

combining this with $\lim_{t\to+\infty}\big\|\dot{y}(t)+\mu(t)M(y(t))\big\| = 0$ gives

$$\lim_{t\to+\infty} \mu(t)\|M(y(t))\| = 0, \quad \text{or} \quad \|M(y(t))\| = o\left(\frac{1}{\mu(t)}\right) \quad \text{as} \quad t\to+\infty.$$

Using the boundedness of $t \mapsto \|y(t)-x_*\|$ produces

$$\langle y(t)-x_*, M(y(t))\rangle = o\left(\frac{1}{\mu(t)}\right) \quad \text{as} \quad t\to+\infty.$$

(ii) We will make use of Opial's Lemma. Define, for $t \geq t_0$,

$$q(t) := \frac{1}{2}\|y(t)-x_*\|^2 + \int_{t_0}^t \mu(s)\langle y(s)-x_*, M(y(s))\rangle ds.$$

Recalling the definition of $p(\cdot)$, we have

$$\begin{aligned}
\lambda q(t) + \dot{q}(t) &= \frac{\lambda}{2}\|y(t)-x_*\|^2 + \Big\langle y(t)-x_*, \dot{y}(t)+\mu(t)M(y(t))\Big\rangle + \lambda\int_{t_0}^t \mu(s)\langle y(s)-x_*, M(y(s))\rangle ds \\
&= p(t) + \lambda\int_{t_0}^t \mu(s)\langle y(s)-x_*, M(y(s))\rangle ds.
\end{aligned}$$

According to Proposition 4.3 (ii) , the integral in the previous sum converges as $t\to+\infty$. Since we already established that $\lim_{t\to+\infty} p(t)$ exists, we are lead to

$$\lim_{t\to+\infty}\big(\lambda q(t) + \dot{q}(t)\big) \in \mathbb{R} \text{ exists.}$$

Using Lemma A.4, we obtain the existence of $\lim_{t\to+\infty} q(t)$. Using again that $\int_{t_0}^{+\infty}\mu(s)\langle y(s)-x_*, M(y(s))\rangle ds$

$< +\infty$, we finally deduce that

$$\lim_{t\to+\infty} \|y(t) - x_*\| \text{ exists.}$$

Thus, the first condition of Opial's Lemma is met. For the second condition, let $\overline{y}$ be a sequential cluster point $y(t)$ as $t \to +\infty$, which means there exists a sequence $(t_n)_{n\in\mathbb{N}} \subseteq [t_0, +\infty[$ such that $t_n \to +\infty$ as $n \to +\infty$ and

$$y(t_n) \rightharpoonup \overline{y} \quad \text{as} \quad n \to +\infty.$$

Since we already know that $\|M(y(t))\| = o\left(\frac{1}{\mu(t)}\right)$ as $t \to +\infty$ and $\mu(\cdot)$ is nondecreasing, we have $M(y(t_n)) \to 0$ as $n \to +\infty$. Since M is maximally monotone, its graph is closed in $\mathcal{H}^{\text{weak}} \times \mathcal{H}^{\text{strong}}$, thus

$$M(\overline{y}) = 0.$$

Now both conditions of Opial's Lemma (see Lemma A.1) are fulfilled, from which we finally conclude the proof of this theorem. ∎

5. Connection with a system with asymptotically vanishing damping governed by a monotone and continuous operator. Here, as it was done in the optimization setting, we will study the convergence behavior of an analog to (AVD) which is suited for monotone equations. Namely, we will be studying the fast OGDA system (see [14]) for the case $\beta(s) \equiv 1$, which for $\alpha > 2$ reads

$$\text{(fOGDA)} \qquad \begin{cases} \ddot{x}(s) + \dfrac{\alpha}{s}\dot{x}(s) + \dfrac{d}{ds}M(x(s)) + \dfrac{\alpha}{2s}M(x(s)) = 0 \quad \text{for } s \geq s_0 > 0, \\ x(s_0) = x_0, \quad \dot{x}(s_0) = x_1, \end{cases}$$

where $x_0, x_1 \in \mathcal{H}$. More specifically, we will show that (fOGDA) can be derived from (M-HBF) via a time rescaling argument. Again, this connection allows us to transfer the convergence results established for (M-HBF) to (fOGDA).

5.1. Two equivalent dynamical systems through time rescaling. Similar to what was done in Section 3, we start with a solution $y : [t_0, +\infty[\to \mathcal{H}$ to

$$\ddot{y}(t) + \lambda\dot{y}(t) + \mu(t)\frac{d}{dt}M(y(t)) + \gamma(t)M(y(t)) = 0, \tag{5.1}$$

and define $x(s) := y(\tau(s))$, where $\tau : [s_0, +\infty[\to [t_0, +\infty[$ is a continuously differentiable function such that $\dot{\tau}(s) > 0$ for every $s \geq s_0 > 0$ and $\lim_{s\to+\infty} \tau(s) = +\infty$. We have

$$\dot{x}(s) = \dot{\tau}(s)\dot{y}(\tau(s)) \quad \text{and} \quad \ddot{x}(s) = \ddot{\tau}(s)\dot{y}(\tau(s)) + \big(\dot{\tau}(s)\big)^2\ddot{y}(\tau(s)).$$

These expressions lead to

$$\dot{y}(\tau(s)) = \frac{1}{\dot{\tau}(s)}\dot{x}(s) \quad \text{and} \quad \ddot{y}(\tau(s)) = \frac{1}{\big(\dot{\tau}(s)\big)^2}\Big[\ddot{x}(s) - \ddot{\tau}(s)\dot{y}(\tau(s))\Big] = \frac{1}{\big(\dot{\tau}(s)\big)^2}\left[\ddot{x}(s) - \frac{\ddot{\tau}(s)}{\dot{\tau}(s)}\dot{x}(s)\right].$$

Moreover

$$\frac{d}{ds}\Big(M(y(\cdot)) \circ \tau\Big)(s) = \dot{\tau}(s)\frac{d}{dt}M(y(t))\Big|_{t=\tau(s)} \;\Rightarrow\; \frac{d}{dt}M(y(t))\Big|_{t=\tau(s)} = \frac{1}{\dot{\tau}(s)}\frac{d}{ds}M(x(s)).$$

Now, plugging $t = \tau(s)$ in (5.1) gives

$$\frac{1}{\big(\dot{\tau}(s)\big)^2}\left[\ddot{x}(s) - \frac{\ddot{\tau}(s)}{\dot{\tau}(s)}\dot{x}(s)\right] + \frac{\lambda}{\dot{\tau}(s)}\dot{x}(s) + \mu(\tau(s))\frac{1}{\dot{\tau}(s)}\frac{d}{ds}M(x(s)) + \gamma(\tau(s))M(x(s)) = 0$$

or, equivalently,

$$\ddot{x}(s) + \left[\lambda\dot{\tau}(s) - \frac{\ddot{\tau}(s)}{\dot{\tau}(s)}\right]\dot{x}(s) + \dot{\tau}(s)\mu(\tau(s))\frac{d}{ds}M(x(s)) + \big(\dot{\tau}(s)\big)^2\gamma(\tau(s))M(x(s)) = 0. \tag{5.2}$$

We briefly recall that the Fast OGDA dynamical system [14], for constant $\beta(\cdot) \equiv 1$, is

$$\ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + \frac{d}{ds}M(x(s)) + \frac{\alpha}{2s}M(x(s)) = 0 \quad \text{for } s \geq s_0 > 0. \tag{5.3}$$

Going back to (5.2), if we want an asymptotically vanishing damping coefficient accompanying the velocity, we need to ask for

$$\begin{cases} \lambda\dot{\tau}(s) - \dfrac{\ddot{\tau}(s)}{\dot{\tau}(s)} & = \dfrac{\alpha}{s} \\ \tau(s_0) & = t_0. \end{cases}$$

Since it was done before in Section 3, we don't repeat the derivation for finding the solution here. We know that for $s \geq s_0 > 0$, the function

$$\tau(s) := \frac{\alpha-1}{\lambda}\ln(s) + \left(-\frac{\alpha-1}{\lambda}\ln(s_0) + t_0\right) = \frac{\alpha-1}{\lambda}\ln\left(\frac{s}{s_0}\right) + t_0 \tag{5.4}$$

satisfies this differential equation. We have

$$\dot{\tau}(s) = \frac{\alpha-1}{\lambda s} \quad \Rightarrow \quad \big(\dot{\tau}(s)\big)^2 = \frac{(\alpha-1)^2}{\lambda^2 s^2}.$$

We need, of course, to assume that $\alpha > 1$. Furthermore, we want the coefficient attached to $M(x(s))$ to be $\frac{\alpha}{2s}$, i.e.,

$$\big(\dot{\tau}(s)\big)^2\gamma(\tau(s)) = \frac{\alpha}{2s} \quad \Leftrightarrow \quad \gamma\left(\frac{\alpha-1}{\lambda}\ln\left(\frac{s}{s_0}\right) + t_0\right) = \frac{\alpha}{2}\cdot\frac{\lambda^2 s}{(\alpha-1)^2},$$

which is fulfilled if we choose

$$\gamma(t) := \frac{\alpha}{2}\cdot\frac{\lambda^2 s_0}{(\alpha-1)^2}\exp\left(\frac{\lambda(t-t_0)}{\alpha-1}\right).$$

Set $\mu(\cdot) = \gamma(\cdot)$. Notice that for every $t \geq t_0$

$$\frac{\dot{\mu}(t)}{\gamma(t)} = \frac{\dot{\mu}(t)}{\mu(t)} = \frac{\left(\frac{\lambda s_0}{\alpha-1}\right)^2\cdot\frac{\lambda}{\alpha-1}\exp\left(\frac{\lambda(t-t_0)}{\alpha-1}\right)}{\left(\frac{\lambda s_0}{\alpha-1}\right)^2\exp\left(\frac{\lambda(t-t_0)}{\alpha-1}\right)} = \frac{\lambda}{\alpha-1}.$$

Thus, the assumption $(\mathcal{M}_2)$ is satisfied if and only if

$$\frac{\lambda}{\alpha-1} < 1 \quad \text{and} \quad 2\lambda - 3 + \frac{\lambda}{\alpha-1} > 0.$$

After rearranging terms, the two previous inequalities are equivalent to

$$\frac{3(\alpha-1)}{2\alpha-1} = \frac{3}{2+\frac{1}{\alpha-1}} < \lambda < \alpha - 1. \tag{5.5}$$

In particular,

$$\frac{3(\alpha-1)}{2\alpha-1} < \alpha - 1 \quad \Leftrightarrow \quad 2 < \alpha.$$

We now turn our attention to the coefficient attached to $\frac{d}{ds}M(x(s))$ in (5.2). Since we want to reach (5.3), we need

$$\frac{\alpha}{2}\cdot\frac{\lambda}{\alpha-1} = \dot{\tau}(s)\mu(\tau(s)) = 1 \quad \Leftrightarrow \quad \lambda = \frac{2(\alpha-1)}{\alpha}.$$

We must verify that this choice of λ satisfies inequality (5.5). Indeed,

$$\frac{2(\alpha-1)}{\alpha} < \alpha - 1 \quad \Leftrightarrow \quad 2 < \alpha \quad \text{and}$$
$$\frac{3(\alpha-1)}{2\alpha-1} < \frac{2(\alpha-1)}{\alpha} \quad \Leftrightarrow \quad 3\alpha < 4\alpha - 2 \quad \Leftrightarrow \quad 2 < \alpha.$$

All in all, $s \mapsto x(s)$ fulfills

$$\ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + \frac{d}{ds}M(x(s)) + \frac{\alpha}{2s}M(x(s)) = 0.$$

Conversely, if for $\alpha > 2$, $x : [s_0, +\infty[\to \mathcal{H}$ is a solution of the previous system and we define $y(t) := x(\sigma(t))$, where $\sigma : [t_0, +\infty[\to [s_0, +\infty[$ is a continuously differentiable function such that $\dot{\sigma}(t) > 0$ for all $t \geq t_0 \geq 0$

and $\lim_{t\to+\infty} \sigma(t) = +\infty$, arguing in a similar fashion as it was done previously we arrive at

$$\ddot{y}(t) + \left[\alpha\frac{\dot{\sigma}(t)}{\sigma(t)} - \frac{\ddot{\sigma}(t)}{\dot{\sigma}(t)}\right]\dot{y}(t) + \dot{\sigma}(t)\frac{d}{dt}M(y(t)) + \frac{\alpha}{2}\cdot\frac{\left(\dot{\sigma}(t)\right)^2}{\sigma(t)}M(y(t)) = 0.$$

We want the coefficient attached to $\dot{y}(t)$ to be $\lambda = \frac{2(\alpha-1)}{\alpha}$. For this end, we need σ to satisfy the differential equation

$$\begin{cases} \alpha\dfrac{\dot{\sigma}(t)}{\sigma(t)} - \dfrac{\ddot{\sigma}(t)}{\dot{\sigma}(t)} & = \dfrac{2(\alpha-1)}{\alpha}, \\ \sigma(t_0) & = s_0, \end{cases} \quad \text{which is fulfilled by} \quad \sigma(t) := s_0 \exp\left(\frac{2(t-t_0)}{\alpha}\right).$$

With this choice for $\sigma(\cdot)$, the resulting system reads

$$\ddot{y}(t) + \frac{2(\alpha-1)}{\alpha}\dot{y}(t) + \frac{2s_0}{\alpha}\exp\left(\frac{2(t-t_0)}{\alpha}\right)\frac{d}{dt}M(y(t)) + \frac{2s_0}{\alpha}\exp\left(\frac{2(t-t_0)}{\alpha}\right)M(y(t)) = 0.$$

It is straightforward to check (we have actually done this already when going from the Heavy Ball system to the Fast OGDA dynamics) that

$$\lambda = \frac{2(\alpha-1)}{\alpha} \quad \text{and} \quad \mu(t) = \gamma(t) = \frac{2s_0}{\alpha}\exp\left(\frac{2(t-t_0)}{\alpha}\right) \quad \forall t \geq t_0$$

satisfy ($\mathcal{M}_2$), and thus all the results ensured by Theorem 4.4 hold.

We have essentially shown the following proposition.

Proposition 5.1. *Assume that $\alpha > 2$ and that $s_0 > 0, t_0 \geq 0$ are initial times. Consider the following second-order systems:*

$$\begin{cases} \ddot{y}(t) + \frac{2(\alpha-1)}{\alpha}\dot{y}(t) + \frac{2s_0}{\alpha}\exp\left(\frac{2(t-t_0)}{\alpha}\right)\frac{d}{dt}M(y(t)) + \frac{2s_0}{\alpha}\exp\left(\frac{2(t-t_0)}{\alpha}\right)M(y(t)) = 0, \\ y(t_0) = y_0, \quad \dot{y}(t_0) = y_1, \end{cases} \tag{5.6}$$

and

$$\begin{cases} \ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + \frac{d}{ds}M(x(s)) + \frac{\alpha}{2s}M(x(s)) = 0, \\ x(s_0) = x_0, \quad \dot{x}(s_0) = x_1. \end{cases} \tag{5.7}$$

Then, the following statements are true:

(i) *If $y : [t_0, +\infty) \to \mathcal{H}$ is a solution to (5.6) and the function $\tau : [s_0, +\infty) \to [t_0, +\infty)$ is given by*

$$\tau(s) := \frac{\alpha}{2}\ln\left(\frac{s}{s_0}\right) + t_0,$$

then the reparametrized solution $x : [s_0, +\infty) \to \mathcal{H}$ given by $x(s) := y(\tau(s))$ is a solution to (5.7) for initial conditions

$$x(s_0) = y_0 \quad \text{and} \quad \dot{x}(s_0) = \frac{\alpha}{2s_0}y_1.$$

(ii) *If $x : [s_0, +\infty) \to \mathcal{H}$ is a solution to (5.7) and the function $\sigma : [t_0, +\infty) \to [s_0, +\infty)$ is given by*

$$\sigma(t) := s_0\exp\left(\frac{2(t-t_0)}{\alpha}\right),$$

then the reparametrized solution $y : [t_0, +\infty) \to \mathcal{H}$ given by $y(t) := x(\sigma(t))$ is a solution to (5.6) for initial conditions

$$y(t_0) = x_0 \quad \text{and} \quad \dot{y}(t_0) = \frac{2s_0}{\alpha}x_1.$$

5.2. Transferring the rates to Fast OGDA. A direct consequence of Theorem 4.4 and Proposition 5.1 yields the following theorem.

Theorem 5.2. *Let $\alpha > 2$, $s_0 > 0$, and $x : [s_0, +\infty[\to \mathcal{H}$ be a solution to*

$$\begin{cases} \ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + \frac{d}{ds}M(x(s)) + \frac{\alpha}{2s}M(x(s)) = 0, \\ x(s_0) = x_0, \quad \dot{x}(s_0) = x_1. \end{cases}$$

and let x_* be a zero of M. Then, it holds

$$\|M(x(s))\| = o\left(\frac{1}{s}\right), \quad \langle x(s) - x_*, M(x(s))\rangle = o\left(\frac{1}{s}\right), \quad \|\dot{x}(s)\| = o\left(\frac{1}{s}\right) \quad \text{as } s \to +\infty.$$

Furthermore, $x(s)$ converges weakly to a zero of M as $s \to +\infty$.

Proof. The proof is near identical to the one in the optimization setting. As per Proposition 5.1, define $y(t) := x(\sigma(t))$. We know that $y : [t_0, +\infty) \to \mathcal{H}$ is a solution to (5.6). Again, the function $\sigma \circ \tau : [s_0, +\infty) \to [s_0, +\infty)$ is the identity, which gives $x(s) = y(\tau(s))$ and $\frac{2s}{\alpha}\dot{x}(s) = \dot{y}(\tau(s))$ for every $s \geq s_0$. According to Theorem 4.4, if we set $\mu(t) = \gamma(t) = \frac{2s_0}{\alpha}\exp\left(\frac{2(t-t_0)}{\alpha}\right)$ for all $t \geq t_0$, then it holds, as $s \to +\infty$,

$$\|M(x(s))\| = o\left(\frac{1}{\mu(\tau(s))}\right), \quad \langle x(s) - x_*, M(x(s))\rangle = o\left(\frac{1}{\mu(\tau(s))}\right) \quad \text{and} \quad \frac{2s}{\alpha}\|\dot{x}(s)\| \to 0$$

We just need to compute $\mu(\tau(s))$. We readily see

$$\mu(\tau(s)) = \mu\left(\frac{\alpha}{2}\ln\left(\frac{s}{s_0}\right) + t_0\right) = \frac{2s_0}{\alpha}\exp\left(\frac{2}{\alpha}\cdot\frac{\alpha}{2}\ln\left(\frac{s}{s_0}\right)\right) = \frac{2s}{\alpha},$$

which immediately shows the rates claimed in the statement. Since $y(t)$ converges weakly to a zero of M as $t \to +\infty$, so does $x(s) = y(\tau(s))$ as $s \to +\infty$. $\square$

Appendix A. Auxiliary results. In what follows, we recall some results used throughout the paper. A key tool in proving the weak convergence of the solutions is Opial's Lemma (see [20])

LEMMA A.1. *(Opial) Let S be a nonempty subset of $\mathcal{H}$ and let $x : [t_0, +\infty[\to \mathcal{H}$. Assume that*

(i) *for every $x_* \in S$, $\lim_{t\to+\infty}\|x(t) - x_*\|$ exists;*
(ii) *every weak sequential limit point of $x(t)$, as $t \to +\infty$, belongs to S.*

Then $x(t)$ converges weakly as $t \to +\infty$, and its limit belongs to S.

The following result can be found in [1, Lemma 5.1].

LEMMA A.2. *Let $\delta > 0$. Suppose that $F : [\delta, +\infty) \to \mathbb{R}$ is locally absolutely continuous, bounded from below, and there exists $G \in \mathbb{L}^1([\delta, +\infty); \mathbb{R})$ such that for almost every $t \geq \delta$*

$$\frac{d}{dt}F(t) \leq G(t).$$

Then the limit $\lim\limits_{t\to+\infty} F(t) \in \mathbb{R}$ exists.

LEMMA A.3. *Let $A, B, C \in \mathbb{R}$ be such that $A \neq 0$ and $B^2 - AC \leq 0$. The following statements are true:*

(i) *if $A > 0$, then it holds*

$$A\|x\|^2 + 2B\langle x, y\rangle + C\|y\|^2 \geq 0 \quad \forall x, y \in \mathcal{H};$$

(ii) *if $A < 0$, then it holds*

$$A\|x\|^2 + 2B\langle x, y\rangle + C\|y\|^2 \leq 0 \quad \forall x, y \in \mathcal{H}.$$

The following lemma appears as Lemma A.3 in [15] and generalizes Lemma A.2 from [10].

LEMMA A.4. *Let $a > 0$, $r \in [0, 1]$ and $q : [t_0, +\infty) \to \mathbb{R}$ be a continuously differentiable function such that*

$$\lim_{t\to+\infty}\left(q(t) + \frac{t^r}{a}\dot{q}(t)\right) = \ell \in \mathbb{R}.$$

Then it holds $\lim_{t\to+\infty} q(t) = \ell$.

Acknowledgments. This work was started in the last year of Hedy Attouch's life, just months before his passing. R.I. Boţ, D. A. Hulett and D.-K. Nguyen would like to take this opportunity to pay tribute to a remarkable mathematician, a kind and generous soul, whose absence is deeply felt.

We also want to thank our colleague Ernö Robert Csetnek (University of Vienna) for his careful reading and comments that improved the manuscript. Last but not least, we thank the two anonymous referees for their comments and suggestions, which have improved the quality of the presentation.

REFERENCES

[1] B. Abbas, H. Attouch, B.F. Svaiter, *Newton-like dynamics and forward–backward methods for structured monotone inclusions in Hilbert spaces*, Journal of Optimization Theory and Applications **161** (2) (2014), 331–360

[2] C.D. Alecsa, S. László, T. Pinta, *An extension of the second-order dynamical system that models Nesterov's convex gradient method*, Applied Mathematics and Optimization **84** (2021), 1687–1716

[3] F. Álvarez, *On the minimizing property of a second-order dissipative system in Hilbert spaces*, SIAM Journal on Control and Optimization **38** (4) (2000), 1102–1119

[4] H. Attouch, R.I. Boţ, D.-K. Nguyen, *Fast convex optimization via time scale and averaging of the steepest descent*, Mathematics of Operations Research (2024), https://doi.org/10.1287/moor.2023.0186

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

[6] H. Attouch, Z. Chbani, J. Peypouquet, P. Redont, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity*, Mathematical Programming **168** (2018), 123–175

[7] H. Attouch, Z. Chbani, H. Riahi, *Fast proximal methods via time scaling of damped inertial dynamics*, SIAM Journal on Optimization **29** (3) (2019), 2227–2256

[8] H. Attouch, X. Goudou, P. Redont, *The heavy ball with friction method. The continuous dynamical system, global exploration of the local minima of a real-valued function by asymptotical analysis of a dissipative dynamical system*, Communications in Contemporary Mathematics **2** (1) (2000), 1–34

[9] H. Attouch, J. Peypouquet, *Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators*, Mathematical Programming **174** (1-2) (2019), 391–432

[10] H. Attouch, J. Peypouquet, P. Redont, *Fast convex minimization via inertial dynamics with Hessian driven damping*, Journal of Differential Equations, **261** (2016), 5734–5783

[11] H. Attouch, B.F. Svaiter *A continuous dynamical Newton-like approach to solving monotone inclusions*, SIAM Journal on Control and Optimizaion **49** (2) (2011), 574–598

[12] R.I. Boţ, E.R. Csetnek, *second-order forward-backward dynamical systems for monotone inclusion problems*, SIAM Journal on Control and Optimization **54** (3) (2016), 1423–1443

[13] R.I. Boţ, E.R. Csetnek, *A dynamical system associated with the fixed points set of a nonexpansive operator*, Journal of Dynamics and Differential Equations **29** (2017), 155-–168

[14] R.I. Boţ, E.R. Csetnek, D.-K. Nguyen, *Fast Optimistic Gradient Descent Ascent (OGDA) in continuous and discrete time*, Foundations of Computational Mathematics **25** (1) (2025), 162–222

[15] R.I. Boţ, D.A. Hulett, D.-K. Nguyen, *Fast second-order dynamics with slow vanishing damping approaching the zeros of a monotone and continuous operator*, Journal of Convex Analysis, to appear

[16] A. Haraux, *Systemes Dynamiques Dissipatifs et Applications*, Recherches en Mathématiques Appliqueés 17, Masson, Paris (1991)

[17] R. May, *Asymptotic for a second-order evolution equation with convex potential and vanishing damping term*, Turkish Journal of Mathematics, **41** (3) (2017), 681–685

[18] Y. Nesterov, *A method of solving a convex programming problem with convergence rate $O(1/k^2)$*, Soviet Mathematics Doklady, **27** (1983), 372–376

[19] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic Publishers, Boston (2004)

[20] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bulletin of the American Mathematical Society **73** (1967), 591–597

[21] B.T. Polyak, *Some methods of speeding up the convergence of iteration methods*, USSR Computational Mathematics and Mathematical Physics **4** (1964), 1–17

[22] B.T. Polyak, *Introduction to Optimization*, Optimization Software, New York (1987)

[23] G.R. Sell, Y. You, *Dynamics of Evolutionary Equations*, Springer, New York (2002)

[24] T. Sun, P. Yin, D. Li, C. Huang, L. Guan, H. Jiang, *Non-ergodic convergence analysis of heavy-ball algorithms.* In AAAI'19/IAAI'19/EAAI'19, AAAI Press, **33** Article 618 (2019), 5033–5040

[25] W.J. Su, S. Boyd, E.J. Candès, *A differential equation for modeling Nesterov's accelerated gradient method: theory and insights*. Neural Information Processing Systems, **27** (2014), 2510–2518