On the effect of perturbations, errors in first-order optimization methods with inertia and Hessian driven damping

Hedy Attouch, Jalal M. Fadili, Vyacheslav Kungurtsev

On the effect of perturbations, errors in first-order optimization methods with inertia and Hessian driven damping

Hedy Attouch* Jalal Fadili† Vyacheslav Kungurtsev‡

Abstract

Second-order continuous-time dissipative dynamical systems with viscous and Hessian driven damping have inspired effective first-order algorithms for solving convex optimization problems. While preserving the fast convergence properties of the Nesterov-type acceleration, the Hessian driven damping makes it possible to significantly attenuate the oscillations. To study the stability of these algorithms with respect to perturbations, errors, we analyze the behavior of the corresponding continuous systems when the gradient computation is subject to errors. We provide a quantitative analysis of the asymptotic behavior of two types of systems, those with implicit and explicit Hessian driven damping. We consider convex, strongly convex, and non-smooth objective functions defined on a real Hilbert space and show that, depending on the formulation, different integrability conditions on the perturbations are sufficient to maintain the convergence rates of the systems. We highlight the differences between the implicit and explicit Hessian damping, and in particular point out that the assumptions on the objective and perturbations needed in the implicit case are more stringent than in the explicit case.

Keywords: Hessian driven damping; damped inertial dynamics; accelerated convex optimization; convergence rates; Lyapunov analysis, perturbation; errors.
AMS subject classification 37N40, 46N10, 49M30, 65B99, 65K05, 65K10, 90B50, 90C25

1 Introduction

The continuous-time dynamic perspective of optimization algorithms, which can be viewed as temporal discretization schemes thereof, offers an insightful and powerful framework for the study of the behavior of these algorithms. In this paper, we study inertial systems involving both viscous and Hessian-driven damping, where the first-order gradient information is only accessible up to some *exogenous additive error*.

1.1 Problem statement

Throughout the paper, we make the following standing assumptions:

$$\boxed{f \text{ is a convex function on a real Hilbert space } \mathcal{H}, \text{ and } S := \operatorname{argmin}_{\mathcal{H}} f \neq \emptyset.}$$

We will study perturbed versions of two second-order ordinary differential equations (ODE). They differ from each other in that the Hessian driven damping appears explicitly in one and implicitly in the other.

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

†ormandie Univ, ENSICAEN, CNRS, GREYC, Caen, France. Jalal.Fadili@greyc.ensicaen.fr

‡Department of Computer Science and Engineering, Czech Technical University, Prague. vyacheslav.kungurtsev@fel.cvut.cz

1.1.1 Explicit Hessian

The first system we look at, which was proposed in [8] (see also [16]), takes the form

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \beta(t)\frac{d}{dt}(\nabla f(x(t))) + b(t)\nabla f(x(t)) = 0, \tag{ISEHD}$$

where $f \in \mathcal{C}^1$, $\gamma, \beta, b : [t_0, +\infty[\to \mathbb{R}_+$ are continuous functions, and $t_0 > 0$ is the initial time. The coefficients (γ, β, b) have a physical interpretation corresponding to natural phenomena:

- $\gamma(t)$ is the viscous damping coefficient,
- $\beta(t)$ is the Hessian-driven damping coefficient (which will be made clear),
- $b(t)$ is the time scaling coefficient (see [10]).

We term the above ODE an Inertial System with *Explicit* Hessian Damping (ISEHD for short), since

$$\frac{d}{dt}(\nabla f(x(t))) = \nabla^2 f(x(t))\dot{x}(t),$$

when f is of class $\mathcal{C}^2$. Throughout the paper, we consider (ISEHD) with the particular choice of parameters

$$\gamma(t) = \frac{\alpha}{t}, \quad \alpha \geq 0, \;\; \beta(t) \equiv \beta > 0 \text{ and } b(t) \equiv 1.$$

This choice of the viscous damping parameter $\gamma(t) = \frac{\alpha}{t}$ is justified by its direct link with the accelerated gradient method of Nesterov [27, 28], as shown in [6], [9], [18], [23], [33]. Related systems have been considered in [25] from the closed loop control perspective and in [33] by means of high-resolution of differential equations.

1.1.2 Implicit Hessian

The second system we consider, inspired by [3] (see also [26] for a related autonomous system in the case of a strongly convex function f), is

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f\Big(x(t) + \beta(t)\dot{x}(t)\Big) = 0, \tag{ISIHD}$$

where $\alpha \geq 3$ and $\beta(t) = \gamma + \frac{\beta}{t}, \gamma, \;\beta \geq 0$. We coin this ODE an Inertial System with *Implicit* Hessian Damping (ISIHD for short). The rationale justifying our use of the term "implicit" comes from the observation that by a Taylor expansion (as $t \to +\infty$ we have $\dot{x}(t) \to 0$ which justifies using Taylor expansion), one has

$$\nabla f\left(x(t) + \beta(t)\dot{x}(t)\right) \approx \nabla f(x(t)) + \beta(t)\nabla^2 f(x(t))\dot{x}(t),$$

hence making the Hessian damping appear indirectly in (ISIHD). This ODE was found to have a smoothing effect on the energy error and oscillations.

1.1.3 Exogenous additive error

We are interested in the situation where $\nabla f(x(t))$ is always evaluated with an exogenous additive error $e(t)$. With the choice of parameters made above, the perturbed dynamics (ISEHD) and (ISIHD) are written

$$\boxed{\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\frac{d}{dt}\Big(\nabla f(x(t)) + e(t)\Big) + \nabla f(x(t)) + e(t) = 0,} \tag{ISEHD-Pert}$$

$$\boxed{\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f\Big(x(t) + \beta(t)\dot{x}(t)\Big) + e(t) = 0.} \tag{ISIHD-Pert}$$

For system (ISEHD-PERT), the overall perturbation error affecting the system is $\beta \dot{e}(t) + e(t)$. Because the Hessian appears explicitly, both the error on the gradient and its derivative appear. It can then be anticipated that assumptions regarding both $e(t)$ and $\dot{e}(t)$, in particular their integrability, will be instrumental in deriving any convergence guarantees. On the other hand, in the system (ISIHD-PERT) with implicit Hessian damping, the error perturbation $e(t)$ appears without its time derivative. Naturally, we will see in this case that convergence results will be derived without any assumptions on the time derivative of the error. While this may be seen as an advantage at first glance, this comes at a price. Indeed, as we will also see, to maintain fast convergence guarantees, the integrability requirements on the error $e(t)$ will be more stringent for (ISIHD-PERT) than for (ISEHD-PERT), *i.e.* , higher-order moments of $e(t)$ will be required to be finite. We anticipate that when it comes to discrete algorithms, the assumptions on the objective and perturbations needed in the implicit case are more stringent than in the explicit case. We plan to study these questions in a future work. Note that similar questions arise when the perturbation is attached to a Tikhonov regularization term with an asymptotically vanishing coefficient [20].

1.2 Contributions

In [8] (resp. [3]), which studied the unperturbed system (ISEHD) (resp. (ISIHD)), fast convergence rates were obtained for the objective, velocities and gradients. Our main contribution in this paper is to analyze the robustness and stability of these systems, by quantifying their convergence properties in the presence of errors. We do this both in the general convex case and in the strongly convex case. We also study the case where the function f is non-smooth convex by proposing a first order formulation in time and space, with existence results and Lyapunov analysis. The main motivation for our work is to pave the way for the design and study of provably accelerated optimization algorithms that appropriately discretize the above dynamics while handling inexact evaluations of the gradient with deterministic and/or stochastic errors. The extension to the discrete setting of the results here will be the focus of a forthcoming paper.

1.3 Related Works

Due to the importance of the subject in optimization and control, several articles have been devoted to the study of perturbations in dissipative inertial systems and in the corresponding accelerated first order algorithms. The subject was first considered in the case of a fixed viscous damping, [11, 24]. Then it was studied within the framework of the accelerated gradient method of Nesterov, and of the corresponding inertial dynamics with vanishing viscous damping, see [7, 9, 17, 31, 34]. In the presence of the additional Hessian driven damping, first results have been obtained in [16] in the case of a smooth function. To the best of our knowledge, our work is the first to consider these questions in full generality, in view of recent progress concerning the associated algorithms [8, 32].

1.4 Contents

In Section 2, we prove that the two systems are well-posed both in the smooth and non-smooth cases. In Section 3, we study the convex case, and establish convergence rates for both systems under appropriate integrability assumptions on the error. In Section 4, we consider the strongly convex case. Section 5 is devoted to studying non-smooth f. In Section 6, we present some numerical illustrations of the results. In Section 7, we draw key conclusions and present some perspectives.

1.5 Main notations

$\mathcal{H}$ is a real Hilbert space, $\langle \cdot, \cdot \rangle$ is the scalar product on $\mathcal{H}$ and $\|\cdot\|$ is the corresponding norm. $\Gamma_0(\mathcal{H})$ is the class of proper, lower semicontinuous (lsc) and convex functions from $\mathcal{H}$ to $\mathbb{R} \cup \{+\infty\}$. For $g \in \Gamma_0(\mathcal{H})$, its domain is $\mathrm{dom}(g) := \{x \in \mathcal{H} : \ g(x) < +\infty\}$. ∂g denotes the (convex) subdifferential operator of g. When g is differentiable at $x \in \mathcal{H}$, then $\partial g(x) = \{\nabla g(x)\}$. We also denote $\mathrm{dom}(\partial g) := \{x \in \mathcal{H} : \ \partial g(x) \neq \emptyset\}$. We take $t_0 > 0$ as the origin of time. This is justified by the singularity of the viscous damping coefficient $\frac{\alpha}{t}$ at the origin. This is not restrictive since we are interested in asymptotic analysis.

2 Well-posedness

When $\beta > 0$, the presence of Hessian driven damping in the inertial dynamics makes it possible to reformulate the equations as first-order systems both in time and in space, without explicit evaluation of the Hessian. This will allow us to extend the existence of trajectories and the convergence results to the case $f \in \Gamma_0(\mathcal{H})$, by simply replacing the gradient of f with the subdifferential ∂f. This approach was initiated in [4] and used in [16] for the unperturbed case.

2.1 Explicit Hessian Damping

2.1.1 First order in time and space formulation

Let us start by establishing this equivalence in the case of a smooth function f.

Theorem 2.1. *Let $f : \mathcal{H} \to \mathbb{R}$ be a $\mathcal{C}^2$ function and $e : [t_0, +\infty[\to \mathcal{H}$ be $\mathcal{C}^1$. Suppose that $\alpha \geq 0$, $\beta > 0$. Let $(x_0, \dot{x}_0) \in \mathcal{H} \times \mathcal{H}$. The following statements are equivalent:*

1. *$x : [t_0, +\infty[\to \mathcal{H}$ is a solution trajectory of* (ISEHD-Pert) *with the initial conditions $x(t_0) = x_0$, $\dot{x}(t_0) = \dot{x}_0$.*

2. *$(x, y) : [t_0, +\infty[\to \mathcal{H} \times \mathcal{H}$ is a solution trajectory of the first-order system*

$$\begin{cases} \dot{x}(t) + \beta(\nabla f(x(t)) + e(t)) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right) x(t) + \frac{1}{\beta} y(t) & = 0 \\ \dot{y}(t) - \left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right) x(t) + \frac{1}{\beta} y(t) & = 0, \end{cases} \tag{1}$$

with initial conditions $x(t_0) = x_0$, $y(t_0) = -\beta(\dot{x}_0 + \beta \nabla f(x_0)) + (1 - \beta\alpha/t_0)x_0 - \beta^2 e(t_0)$.

Proof. *2.* $\Rightarrow$ *1.* Differentiating the first equation of (1) gives

$$\ddot{x}(t) + \beta\left(\nabla^2 f(x(t))\dot{x}(t) + \dot{e}(t)\right) - \frac{\alpha}{t^2}x(t) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)\dot{x}(t) + \frac{1}{\beta}\dot{y}(t) = 0. \tag{2}$$

Replacing $\dot{y}(t)$ by its expression as given by the second equation of (1) gives

$$\ddot{x}(t) + \beta\left(\nabla^2 f(x(t))\dot{x}(t) + \dot{e}(t)\right) - \frac{\alpha}{t^2}x(t) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)\dot{x}(t) + \frac{1}{\beta}\left(\left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right)x(t) - \frac{1}{\beta}y(t)\right) = 0. \tag{3}$$

Then replace $y(t)$ by its expression as given by the first equation of (1)

$$\ddot{x}(t) + \beta\left(\nabla^2 f(x(t))\dot{x}(t) + \dot{e}(t)\right) - \frac{\alpha}{t^2}x(t) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)\dot{x}(t)$$
$$+ \frac{1}{\beta}\left(\left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right)x(t) + \dot{x}(t) + \beta(\nabla f(x(t)) + e(t)) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)x(t)\right) = 0.$$

After simplification of the above expression, we obtain (ISEHD-Pert).

1. $\Rightarrow$ *2.* Define $y(t)$ by the first equation of (1). Differentiating $y(t)$ and using equation (ISEHD-Pert) allows one to eliminate $\ddot{x}(t)$, which finally gives the second equation of (1). □

2.1.2 Existence and uniqueness of a solution

Capitalizing on the result of Theorem 2.1, the following first order formulation assists in providing a meaning to our system when $f \in \Gamma_0(\mathcal{H})$. It is obtained by substituting the subdifferential ∂f for the gradient ∇f in the first-order formulation (1).

Definition 2.2. *Let $\alpha \geq 0$, $\beta > 0$ and $f \in \Gamma_0(\mathcal{H})$. Given $(x_0, y_0) \in \operatorname{dom}(f) \times \mathcal{H}$, the Cauchy problem for the perturbed inertial system with explicit generalized Hessian driven damping is defined by*

$$\begin{cases} \dot{x}(t) + \beta(\partial f(x(t)) + e(t)) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right) x(t) + \frac{1}{\beta} y(t) & \ni 0 \\ \dot{y}(t) - \left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right) x(t) + \frac{1}{\beta} y(t) & = 0 \\ x(t_0) = x_0, y(t_0) = y_0. \end{cases} \tag{4}$$

Let us formulate (4) in a condensed form as an evolution equation in the product space $\mathcal{H} \times \mathcal{H}$. Setting $Z(t) = (x(t), y(t)) \in \mathcal{H} \times \mathcal{H}$, (4) can be equivalently written

$$\dot{Z}(t) + \partial \mathcal{G}(Z(t)) + \mathcal{D}(t, Z(t)) \ni 0, \quad Z(t_0) = (x_0, y_0), \tag{5}$$

where $\mathcal{G} \in \Gamma_0(\mathcal{H} \times \mathcal{H})$ is the function defined by $\mathcal{G}(Z) = \beta f(x)$, and the time-dependent operator $\mathcal{D} : [t_0, +\infty[\times \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ is given by

$$\mathcal{D}(t, Z) = \left(\beta e(t) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right) x + \frac{1}{\beta} y, - \left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right) x + \frac{1}{\beta} y\right). \tag{6}$$

The differential inclusion (5) is governed by the sum of the maximal monotone operator $\partial\mathcal{G}$ (a convex subdifferential) and the time-dependent affine continuous operator $\mathcal{D}(t, \cdot)$. The existence and uniqueness of a global solution for the corresponding Cauchy problem is a consequence of the general theory of evolution equations governed by maximally monotone operators. In this setting, the notion of classical solution is replaced by the the notion of strong solution (see [21, Definition 3.1]), that we make precise now.

Definition 2.3. *Given $g \in \Gamma_0(\mathcal{H})$, and an operator $D : [t_0, +\infty[\times \mathcal{H} \to \mathcal{H}$, we say that $z : [t_0, T] \to \mathcal{H}$ is a strong solution trajectory on $[t_0, T]$ of the differential inclusion*

$$\dot{z}(t) + \partial g(z(t)) + D(t, z(t)) \ni 0, \tag{7}$$

if the following properties are satisfied:

1. *z is continuous on $[t_0, T]$ and absolutely continuous on any compact subset of $]t_0, T]$;*
2. *$z(t) \in \operatorname{dom}(\partial g)$ for almost every $t \in]t_0, T]$, and (7) is verified for almost every $t \in]t_0, T]$.*

$z : [t_0, +\infty[\to \mathcal{H}$ is a global strong solution of (7), if it is a strong solution on $[t_0, T]$ for all $T > t_0$.

The existence and uniqueness of a global strong solution of the Cauchy problem (4) is established in the following theorem.

Theorem 2.4. *Let $f \in \Gamma_0(\mathcal{H})$, $\alpha \geq 0$ and $\beta > 0$. Suppose that $e \in L^2(t_0, T; \mathcal{H})$ for every $T > t_0$. Then, for any Cauchy data $(x_0, y_0) \in \operatorname{dom}(f) \times \mathcal{H}$, there exists a unique global strong solution $(x, y) : [t_0, +\infty[\to \mathcal{H} \times \mathcal{H}$ of (4) satisfying the initial condition $x(t_0) = x_0$, $y(t_0) = y_0$. Moreover, this solution exhibits the following properties:*

(i) *y is of class $\mathcal{C}^1$ on $[t_0, +\infty[$, and $\dot{y}(t) - \left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right) x(t) + \frac{1}{\beta} y(t) = 0$, for $t \geq t_0$;*

(ii) *x is absolutely continuous on $[t_0, T]$ and $\dot{x} \in L^2(t_0, T; \mathcal{H})$ for all $T > t_0$;*

(iii) $x(t) \in \operatorname{dom}(\partial f)$ *for all* $t > t_0$*;*

(iv) x *is Lipschitz continuous on any compact subinterval of* $]t_0, +\infty[$*;*

(v) *the function* $t \mapsto f(x(t))$ *is absolutely continuous on* $[t_0, T]$ *for all* $T > t_0$*;*

(vi) *there exists a function* $\xi : [t_0, +\infty[\to \mathcal{H}$ *such that*

(a) $\xi(t) \in \partial f(x(t))$ *for all* $t > t_0$*;*

(b) $\dot{x}(t) + \beta\xi(t) + \beta e(t) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right) x(t) + \frac{1}{\beta} y(t) = 0$ *for almost every* $t > t_0$*;*

(c) $\xi \in L^2(t_0, T; \mathcal{H})$ *for all* $T > t_0$*;*

(d) $\dfrac{d}{dt} f(x(t)) = \langle \xi(t), \dot{x}(t) \rangle$ *for almost every* $t > t_0$*.*

Proof. It is sufficient to prove that (x, y) is a strong solution of (4) on $[t_0, T]$ and that the properties hold on $[t_0, T]$ for all $T > t_0$. So let us fix $T > t_0$. As we have already noticed, (4) can be written in the form (5) which is a Lipschitz perturbation of the differential inclusion governed by the subdifferential of a proper lsc convex function. A direct application of [21, Proposition 3.12] gives the existence and uniqueness of a strong global solution $Z = (x, y) : [t_0, T] \to \mathcal{H} \times \mathcal{H}$ to (5), or equivalently to (4), with initial condition $Z(t_0) = (x(t_0), y(t_0)) = (x_0, y_0)$. Verification of items (iii) to (vi) follows the same lines as the proof of [16, Theorem 4.4]. Of particular importance is the generalized derivation chain rule given in (vi)(d), which follows from [21, Lemma 3.3] after checking that the corresponding assumptions are met thanks to (ii), (vi)(a) and (vi)(c). $\square$

Combining Theorem 2.1 and Theorem 2.4 with $y(t_0) = -\beta(\dot{x}_0 + \beta \nabla f(x_0)) + (1 - \beta\alpha/t_0)x_0 - \beta^2 e(t_0)$, we obtain the existence and uniqueness result for (ISEHD-Pert).

Corollary 2.5. *Assume that* f *is a convex* $\mathcal{C}^2$ *function and* e *is* $\mathcal{C}^1$ *and* $e \in L^2(t_0, T; \mathcal{H})$ *for every* $T > t_0$*. For any* $t_0 > 0$*, and any Cauchy data* $(x_0, \dot{x}_0)$*, the system* (ISEHD-Pert) *with* $\alpha \geq 0$*,* $\beta(t) \equiv \beta > 0$ *and* $b(t) \equiv 1$ *admits a unique classical global solution* $x : [t_0, +\infty[\to \mathcal{H}$ *satisfying* $(x(t_0), \dot{x}(t_0)) = (x_0, \dot{x}_0)$*.*

2.2 Implicit Hessian Damping

2.2.1 First order in time and space formulation

Let us now turn to (ISIHD-Pert). We use the shorthand notation $\alpha(t) = \alpha/t$. Here and in the rest of the paper, we assume that $\beta(t)$ is $\mathcal{C}^1$ and $\inf_{t \in [t_0, +\infty[} \beta(t) > 0$.

Let us introduce the new function

$$y(t) := x(t) + \beta(t)\dot{x}(t), \tag{8}$$

whose time derivation gives

$$\dot{y}(t) = \dot{x}(t) + \beta(t)\ddot{x}(t) + \dot{\beta}(t)\dot{x}(t). \tag{9}$$

From (ISIHD-Pert) we know that

$$\ddot{x}(t) = -\alpha(t)\dot{x}(t) - \nabla f(y(t)) - e(t). \tag{10}$$

By combining (9) and (10) we obtain

$$\begin{aligned} \dot{y}(t) &= \dot{x}(t) + \beta(t)\left(-\alpha(t)\dot{x}(t) - \nabla f(y(t)) - e(t)\right) + \dot{\beta}(t)\dot{x}(t) \\ &= \left(1 - \alpha(t)\beta(t) + \dot{\beta}(t)\right)\dot{x}(t) - \beta(t)\left(\nabla f(y(t) + e(t)\right). \end{aligned} \tag{11}$$

From (8) and the fact that $\inf_{t\in[t_0,+\infty[} \beta(t) > 0$ we get $\dot{x}(t) = \frac{1}{\beta(t)}(y(t) - x(t))$. Replacing $\dot{x}(t)$ in (11) with this expression gives

$$\begin{aligned}
\dot{y}(t) &= \left(1 - \alpha(t)\beta(t) + \dot{\beta}(t)\right) \frac{1}{\beta(t)}(y(t) - x(t)) - \beta(t) \left(\nabla f(y(t)) + e(t)\right) \\
&= -\frac{1}{\beta(t)} \left(1 - \alpha(t)\beta(t) + \dot{\beta}(t)\right) x(t) + \frac{1}{\beta(t)} \left(1 - \alpha(t)\beta(t) + \dot{\beta}(t)\right) y(t) - \beta(t) \left(\nabla f(y(t)) + e(t)\right).
\end{aligned}$$

The reverse implication is obtained in a similar way. Let us summarize the results.

Theorem 2.6. *Let $f \in \mathcal{C}^1$. Suppose that $\alpha \geq 0$ and $\inf_{t\in[t_0,+\infty[} \beta(t) > 0$. The following statements are equivalent:*

1. *$x : [t_0, +\infty[\to \mathcal{H}$ is a solution trajectory of (ISIHD-Pert) with initial conditions $x(t_0) = x_0$, $\dot{x}(t_0) = \dot{x}_0$.*

2. *$(x, y) : [t_0, +\infty[\to \mathcal{H} \times \mathcal{H}$ is a solution trajectory of the first-order system*
$$\begin{cases} \dot{x}(t) + \frac{1}{\beta(t)}x(t) - \frac{1}{\beta(t)}y(t) = 0. \\ \dot{y}(t) + \beta(t) \left(\nabla f(y(t)) + e(t)\right) + \frac{1}{\beta(t)} \left(1 - \alpha(t)\beta(t) + \dot{\beta}(t)\right) (x(t) - y(t)) = 0 \end{cases}$$
with initial conditions $x(t_0) = x_0$, $y(t_0) = x_0 + \beta(t_0)\dot{x}_0$.

2.2.2 Existence and uniqueness of a solution

Existence and uniqueness of a global solution for the Cauchy problem associated with the unperturbed problem (ISIHD) was shown in [3] when ∇f is Lipschitz continuous using the Cauchy-Lipschitz theorem. This result can be easily extended to (ISIHD-Pert) if ∇f is assumed to be Lipschitz continuous. We take a different path here and proceed as in Section 2.1.2, so that we can extend the above formulation to the case where $f \in \Gamma_0(\mathcal{H})$, by replacing the gradient ∇f with the subdifferential ∂f.

Definition 2.7. *Let $\alpha(t) \geq 0$, $\beta(t) > 0$, $f \in \Gamma_0(\mathcal{H})$. Given $(x_0, y_0) \in \mathcal{H} \times \operatorname{dom}(f)$, the Cauchy problem associated with the perturbed inertial system with implicit generalized Hessian driven damping is defined by*

$$\begin{cases} \dot{x}(t) + \frac{1}{\beta(t)}x(t) - \frac{1}{\beta(t)}y(t) = 0 \\ \dot{y}(t) + \beta(t) \left(\partial f(y(t)) + e(t)\right) + \frac{1}{\beta(t)} \left(1 - \alpha(t)\beta(t) + \dot{\beta}(t)\right) (x(t) - y(t)) \ni 0 \\ x(t_0) = x_0, y(t_0) = y_0. \end{cases} \tag{12}$$

We reformulate (12) in the product space $\mathcal{H} \times \mathcal{H}$ by setting $Z(t) = (x(t), y(t)) \in \mathcal{H} \times \mathcal{H}$, and thus (12) can be equivalently written as

$$\dot{Z}(t) + \beta(t)\partial\mathcal{G}(Z(t)) + \mathcal{D}(t, Z(t)) \ni 0, \tag{13}$$

where $\mathcal{G} \in \Gamma_0(\mathcal{H} \times \mathcal{H})$ is the function defined as $\mathcal{G}(Z) = f(y)$, and the time dependent operator $\mathcal{D} : [t_0, +\infty[\times \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ is given by

$$\mathcal{D}(t, Z) = \left(\frac{1}{\beta(t)}(x - y), \beta(t)e(t) + \frac{1}{\beta(t)} \left(1 - \alpha(t)\beta(t) + \dot{\beta}(t)\right) (x - y)\right). \tag{14}$$

Constant β When β is independent of t, the differential inclusion (13) is governed by the sum of the convex subdifferential operator $\beta\partial\mathcal{G}$ and the time-dependent affine continuous operator $\mathcal{D}(t, \cdot)$. The existence and uniqueness of a global strong solution for the Cauchy problem associated to (12) follows exactly from the same arguments as those for Theorem 2.4. In turn, existence and uniqueness of a classical global solution to the Cauchy problem associated to (ISIHD-Pert) can be easily obtained thanks to the equivalence in Theorem 2.6.

Time-dependent β When β depends on time, one cannot invoke directly the results of [21]. Instead, one can appeal to the theory of evolution equations governed by general time-dependent subdifferentials as proposed in [12] for example. In fact, for a system in the simpler form (13), one can argue more easily, by making the change of time variable $t = \tau(s)$ with $\beta(\tau(s))\dot{\tau}(s) = 1$. Lemma A.4 then shows that (13) is equivalent to

$$\dot{W}(s) + \partial\mathcal{G}(W(s)) + \mathcal{F}(s, W(s)) \ni 0, \tag{15}$$

where $W(s) = Z(\tau(s))$, and $\mathcal{F}(s, W(s)) = \frac{1}{\beta(\tau(s))}\mathcal{D}(\tau(s), W(s))$ is affine continuous in its second argument. Provided that $\beta \notin L^1(t_0, +\infty; \mathbb{R})$, this defines a proper change of variable in time. With the formulation (15), we are brought back to the appropriate form to argue as before and invoke the results of [21]. We leave the details to the reader for the sake of brevity.

3 Smooth Convex Case

3.1 Explicit Hessian Damping

Consider first the explicit Hessian system (ISEHD-Pert), where we assume that $f \in \mathcal{C}^2(\mathcal{H})$, and recall the specific choices of $\gamma(t) = \frac{\alpha}{t}$, $\alpha > 0$, $\beta(t) \equiv \beta$ and $b(t) \equiv 1$. We will develop a Lyapunov analysis to study the dynamics of (ISEHD-Pert). Some of our arguments are inspired by the works of [9] and [8]. Throughout this section we use the shorthand notation

$$g(t) := e(t) + \beta\dot{e}(t) \tag{16}$$

for the overall contribution of the errors terms. We will first establish the minimization property which is valid by simply assuming the integrability of the error term and its derivative. Then, by reinforcing these hypotheses, we will obtain rapid convergence results, and the convergence of trajectories.

3.1.1 Minimizing properties

Define $u : [t_0, +\infty[\to \mathcal{H}$ by

$$u(t) := x(t) + \beta\int_{t_0}^{t} \nabla f(x(s))ds,$$

which will be instrumental in the proof of the following theorem. Note that, in the following statement, it is simply assumed that f is bounded from below, the set $\operatorname{argmin} f$ may be empty.

Theorem 3.1. *Let $f : \mathcal{H} \to \mathbb{R}$ be a $\mathcal{C}^2$ function which is bounded from below. Assume that $e : [t_0, +\infty[\to \mathcal{H}$ is a $\mathcal{C}^1$ function which satisfies the integrability properties $\int_{t_0}^{+\infty} \|e(t)\|\,dt < +\infty$ and $\int_{t_0}^{+\infty} \|\dot{e}(t)\|\,dt < +\infty$. Suppose that $\alpha, \beta > 0$. Then, for any solution trajectory $x : [t_0, +\infty[\to \mathcal{H}$ of (ISEHD-Pert), we have*

(i) $\sup_{t \geq t_0} \|\dot{u}(t)\| < +\infty$;

(ii) $\int_{t_0}^{+\infty} \frac{1}{t}\|\dot{x}(t)\|^2 dt < +\infty$, $\int_{t_0}^{+\infty} \|\nabla f(x(t))\|^2\, dt < +\infty$, $\int_{t_0}^{+\infty} \frac{1}{t}\|\dot{u}(t)\|^2 dt < +\infty$;

(iii) $\lim_{t\to+\infty} \|\dot{u}(t)\| = 0$; $\lim_{t\to+\infty} \|\dot{x}(t)\| = 0$; $\lim_{t\to+\infty} \|\nabla f(x(t))\| = 0$;

(iv) $\lim_{t\to+\infty} f(x(t)) = \inf_{\mathcal{H}} f$.

Proof. Denote $f^\star := \inf_{\mathcal{H}} f$. Since our analysis is asymptotic, there is no restriction in assuming that $t \geq t_1 := \max(t_0, 2\alpha\beta)$. We will then prove the statements in terms of t_1 and passing to t_0 is immediate thanks to the properties of the solution $x(t)$ in Theorem 2.4.

Claim (i) For $T \geq t \geq t_1$, define the function

$$W_T(t) := \frac{1}{2}\|\dot{u}(t)\|^2 + (f(x(t)) - f^\star) - \int_t^T \langle \dot{u}(\tau), g(\tau)\rangle\, d\tau.$$

Observe that W_T is well-defined under our assumptions. Thus, taking the derivative in time and using (ISEHD-Pert), we get,

$$\begin{aligned}
\dot{W}_T(t) &= \langle \dot{u}(t), \ddot{u}(t) + g(t)\rangle + \langle \dot{x}(t), \nabla f(x(t))\rangle \\
&= \big\langle \dot{x}(t) + \beta\nabla f(x(t)), \ddot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + g(t)\big\rangle + \langle \dot{x}(t), \nabla f(x(t))\rangle \\
&= \left\langle \dot{x}(t) + \beta\nabla f(x(t)), -\frac{\alpha}{t}\dot{x}(t) - \nabla f(x(t))\right\rangle + \langle \dot{x}(t), \nabla f(x(t))\rangle \\
&= -\frac{\alpha}{t}\|\dot{x}(t)\|^2 - \beta\|\nabla f(x(t))\|^2 - \frac{\alpha\beta}{t}\langle \dot{x}(t), \nabla f(x(t))\rangle \\
&\leq -\frac{\alpha}{2t}\|\dot{x}(t)\|^2 - \beta\left(1 - \frac{\alpha\beta}{2t}\right)\|\nabla f(x(t))\|^2 \\
&\leq -\frac{\alpha}{2t}\|\dot{x}(t)\|^2 - \frac{\beta}{2}\|\nabla f(x(t))\|^2,
\end{aligned} \tag{17}$$

where we used Young inequality and the fact that $t \geq t_1 > \alpha\beta$. This implies that W_T is non-increasing and in turn that $W_T(t) \leq W_T(t_1)$ for $t \in [t_1, T]$, *i.e.*

$$\frac{1}{2}\|\dot{u}(t)\|^2 + (f(x(t)) - f^\star) - \int_t^T \langle \dot{u}(\tau), g(\tau)\rangle\, d\tau \leq \frac{1}{2}\|\dot{u}(t_1)\|^2 + (f(x(t_1)) - f^\star) - \int_{t_1}^T \langle \dot{u}(\tau), g(\tau)\rangle\, d\tau.$$

Therefore,

$$\frac{1}{2}\|\dot{u}(t)\|^2 \leq \frac{1}{2}\|\dot{u}(t_1)\|^2 + (f(x(t_1)) - f^\star) + \int_{t_1}^t \|\dot{u}(\tau)\|\|g(\tau)\| d\tau.$$

Applying the Gronwall Lemma A.3, we get

$$\sup_{t\geq t_1}\|\dot{u}(t)\| \leq \big(\|\dot{u}(t_1)\|^2 + 2(f(x(t_1)) - f^\star)\big)^{1/2} + \int_{t_1}^{+\infty}\|e(\tau)\|d\tau + \beta\int_{t_1}^{+\infty}\|\dot{e}(\tau)\|d\tau < +\infty,$$

hence proving the first claim.

Claim (ii) Now define for all $t \geq t_1$

$$W(t) := \frac{1}{2}\|\dot{u}(t)\|^2 + (f(x(t)) - f^\star) - \int_t^{+\infty} \langle \dot{u}(\tau), g(\tau)\rangle\, d\tau.$$

This is again a well-defined function thanks to the first claim, and the integrability of g. Moreover, $W(t)$ is bounded from below,

$$\inf_{t\geq t_1} W(t) \geq -\left[\sup_{t\geq t_1}\|\dot{u}(t)\|\right]\left[\int_{t_1}^{+\infty}\|e(\tau)\|d\tau + \beta\int_{t_1}^{+\infty}\|\dot{e}(\tau)\|d\tau\right] > -\infty.$$

Observe that $\dot{W}(t) = \dot{W}_T(t)$. This together with (17) yields

$$\dot{W}(t) + \frac{\alpha}{2t}\|\dot{x}(t)\|^2 + \frac{\beta}{2}\|\nabla f(x(t))\|^2 \leq 0.$$

Integrating and using that W is bounded from below, we obtain the first two claims. From $\dot{u}(t) = \dot{x}(t) + \beta\nabla f(x(t))$, we deduce that $\|\dot{u}(t)\|^2 \leq 2(\|\dot{x}(t)\|^2 + \beta^2\|\nabla f(x(t))\|^2)$. After integration, we get the last claim

$$\begin{aligned}
\int_{t_1}^{+\infty}\frac{1}{t}\|\dot{u}(t)\|^2 dt &\leq 2\left(\int_{t_1}^{+\infty}\frac{1}{t}\|\dot{x}(t)\|^2 dt + \int_{t_1}^{+\infty}\frac{\beta^2}{t}\|\nabla f(x(t))\|^2 dt\right) \\
&\leq 2\left(\int_{t_1}^{+\infty}\frac{1}{t}\|\dot{x}(t)\|^2 dt + \frac{\beta^2}{t_1}\int_{t_1}^{+\infty}\|\nabla f(x(t))\|^2 dt\right) < +\infty.
\end{aligned}$$

Claim (iii) and (iv) Define $h : t \in [t_1, +\infty[\mapsto \frac{1}{2}\|u(t) - z\|^2$ for arbitrary $z \in \mathcal{H}$. We then have

$$\begin{aligned}
\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) &= \|\dot{u}(t)\|^2 + \left\langle u(t) - z,\, \ddot{u}(t) + \frac{\alpha}{t}\dot{u}(t)\right\rangle \\
&= \|\dot{u}(t)\|^2 + \left\langle u(t) - z,\, \ddot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \frac{\alpha\beta}{t}\nabla f(x(t))\right\rangle \\
&= \|\dot{u}(t)\|^2 - \left\langle u(t) - z,\, g(t) + \left(1 - \frac{\alpha\beta}{t}\right)\nabla f(x(t))\right\rangle \\
&= \|\dot{u}(t)\|^2 - \left(1 - \frac{\alpha\beta}{t}\right)\langle x(t) - z,\, \nabla f(x(t))\rangle - \langle u(t) - z,\, g(t)\rangle \\
&\quad - \beta\left(1 - \frac{\alpha\beta}{t}\right)\left\langle \int_{t_1}^{t} \nabla f(x(s))ds,\, \nabla f(x(t))\right\rangle \\
&= \|\dot{u}(t)\|^2 - \left(1 - \frac{\alpha\beta}{t}\right)\langle x(t) - z,\, \nabla f(x(t))\rangle - \langle u(t) - z,\, g(t)\rangle - \beta\left(1 - \frac{\alpha\beta}{t}\right)\dot{I}(t),
\end{aligned}$$

where $I(t) := \frac{1}{2}\left\|\int_{t_1}^{t} \nabla f(x(s))ds\right\|^2$. From the convexity of f and Cauchy-Schwarz inequality we get

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) + \left(1 - \frac{\alpha\beta}{t}\right)(f(x(t)) - f(z)) + \beta\left(1 - \frac{\alpha\beta}{t}\right)\dot{I}(t) \leq \|\dot{u}(t)\|^2 + \|u(t) - z\|\,\|g(t)\|\,.$$

Inserting $W(t)$ into this expression, we get,

$$\begin{aligned}
&\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) + \left(1 - \frac{\alpha\beta}{t}\right)(W(t) + f^* - f(z)) + \beta\left(1 - \frac{\alpha\beta}{t}\right)\dot{I}(t) \\
&\leq \left(\frac{3}{2} - \frac{\alpha\beta}{2t}\right)\|\dot{u}(t)\|^2 + \|u(t) - z\|\,\|g(t)\| - \left(1 - \frac{\alpha\beta}{t}\right)\int_{t}^{+\infty} \langle \dot{u}(\tau),\, g(\tau)\rangle\, d\tau. \qquad (18)
\end{aligned}$$

Since $W(t)$ is nonincreasing and bounded from below (see above), it converges to some $W_\infty \in \mathbb{R}$ as $t \to +\infty$. Since $\dot{u}$ is bounded, and g is integrable, we have that $\tau \to \langle \dot{u}(\tau),\, g(\tau)\rangle$ is integrable on $[t_0, +\infty[$. Therefore

$$\lim_{t \to +\infty} \int_{t}^{+\infty} \langle \dot{u}(\tau),\, g(\tau)\rangle\, d\tau = 0.$$

Consequently, as $t \to +\infty$

$$\frac{1}{2}\|\dot{u}(t)\|^2 + (f(x(t)) - f^\star) \to W_\infty.$$

If $W_\infty = 0$, since the two terms that enter the above expression (potential energy and kinetic energy) are nonnegative, we obtain that each of them tends to zero as $t \to +\infty$. This gives the claims (iii) and (iv). To prove that $W_\infty = 0$, we argue by contradiction, and show that assuming $W_\infty > 0$ leads to a contradiction. Since $W(t)$ is nonincreasing, we then have $W(t) \geq W_\infty > 0$. Take $z \in \mathcal{H}$ such that $f(z) < f^\star + \frac{1}{2}W_\infty$. Then

$$W(t) + f^* - f(z) > W_\infty - \frac{1}{2}W_\infty = \frac{1}{2}W_\infty.$$

Returning to (18) we deduce that, for $t \geq t_1$

$$\begin{aligned}
&\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) + \frac{1}{2}\left(1 - \frac{\alpha\beta}{t}\right)W_\infty + \beta\left(1 - \frac{\alpha\beta}{t}\right)\dot{I}(t) \\
&\leq \left(\frac{3}{2} - \frac{\alpha\beta}{2t}\right)\|\dot{u}(t)\|^2 + \|u(t) - z\|\,\|g(t)\| - \left(1 - \frac{\alpha\beta}{t}\right)\int_{t}^{+\infty} \langle \dot{u}(\tau),\, g(\tau)\rangle\, d\tau. \qquad (19)
\end{aligned}$$

Since $t < 2\alpha\beta$, we have $1 - \frac{\alpha\beta}{t} > \frac{1}{2}$. Therefore, after rearranging the terms, we obtain

$$\frac{1}{4}W_\infty \leq \frac{3}{2}\|\dot{u}(t)\|^2 + \|u(t)-z\|\,\|g(t)\| + \left(\sup_{t\geq t_1}\|\dot{u}(t)\|\right)\int_t^{+\infty}\|g(s)\|ds - \frac{1}{t^\alpha}\frac{d}{dt}(t^\alpha \dot{h}(t)) - \beta\left(1-\frac{\alpha\beta}{t}\right)\dot{I}(t).$$

Multiplying both sides by $\frac{1}{t}$, and integrating between t_1 and $\tau > t_1$,

$$\begin{aligned}\frac{1}{4}W_\infty \log\left(\frac{\tau}{t_1}\right) \quad \leq \quad & \frac{3}{2}\int_{t_1}^{\tau}\frac{1}{t}\|\dot{u}(t)\|^2 dt + \int_{t_1}^{\tau}\frac{\|g(t)\|\|u(t)-z\|}{t}dt + \left(\sup_{t\geq t_1}\|\dot{u}(t)\|\right)\int_{t_1}^{\tau}\left(\frac{1}{t}\int_t^{+\infty}\|g(s)\|ds\right)dt \\ & - \int_{t_1}^{\tau}\frac{1}{t^{\alpha+1}}\frac{d}{dt}(t^\alpha\dot{h}(t))dt - \beta\int_{t_1}^{\tau}\left(\frac{1}{t}-\frac{\alpha\beta}{t^2}\right)\dot{I}(t)dt. \end{aligned} \tag{20}$$

Throughout the rest of the proof, we will use the inequality

$$\|u(t)-z\| \leq \|u(t_1)-z\| + \int_{t_1}^{t}\|\dot{u}(s)\|\,ds \leq \|u(t_1)-z\| + t\sup_{s\geq t_1}\|\dot{u}(s)\|.$$

Let us examine successively the different terms which enter the second member of (20). The first term is bounded according to claim (ii). The second term is also bounded since

$$\int_{t_1}^{\tau}\frac{\|g(t)\|\|u(t)-z\|}{t}dt \leq \left(\frac{\|u(t_1)-z\|}{t_1} + \sup_{t\geq t_1}\|\dot{u}(t)\|\right)\int_{t_1}^{+\infty}(\|e(t)\| + \beta\|\dot{e}(t)\|)\,dt < +\infty.$$

The third term can be handled by integration by parts,

$$\int_{t_1}^{\tau}\left(\frac{1}{t}\int_t^{+\infty}\|g(s)\|ds\right)dt = \log\tau\int_\tau^\infty\|g(s)\|ds \; - \log t_1\int_{t_1}^{+\infty}\|g(s)\|ds + \int_{t_1}^{\tau}\|g(t)\|\log t\,dt.$$

For the fourth term, set $K(\tau) = -\int_{t_1}^{\tau}\frac{1}{t^{\alpha+1}}\frac{d}{dt}(t^\alpha\dot{h}(t))dt$ and integrate by parts twice to get,

$$\begin{aligned} K(\tau) &= -\left[\frac{1}{t}\dot{h}(t)\right]_{t_1}^{\tau} - (\alpha+1)\int_{t_1}^{\tau}\frac{1}{t^2}\dot{h}(t)dt = -\left[\frac{1}{t}\dot{h}(t)\right]_{t_1}^{\tau} - \frac{(1+\alpha)}{\tau^2}h(\tau) + \frac{(1+\alpha)}{t_1^2}h(t_1) - 2(1+\alpha)\int_{t_1}^{\tau}\frac{h(t)}{t^3}dt \\ &\leq -\left[\frac{1}{t}\dot{h}(t)\right]_{t_1}^{\tau} + \frac{(1+\alpha)}{t_1^2}h(t_1) \leq \frac{1}{t_1}|\langle\dot{u}(t_1),\,u(t_1)-z\rangle| + \frac{1}{\tau}|\langle\dot{u}(\tau),\,u(\tau)-z\rangle| + \frac{(1+\alpha)}{t_1^2}h(t_1) \\ &\leq C + \sup_{t\geq t_1}\|\dot{u}(t)\|\left(\frac{\|u(t_1)-z\|}{t_1} + \sup_{t\geq t_1}\|\dot{u}(t)\|\right) < +\infty. \end{aligned}$$

For the last term, we infer from Lemma A.2 that

$$\sup_{\tau > t_1} -\beta\int_{t_1}^{\tau}\left(\frac{1}{t}-\frac{\alpha\beta}{t^2}\right)\dot{I}(t)dt < +\infty.$$

Overall, we have shown that there exists a constant $C > 0$ such that (20) reads

$$\frac{1}{4}W_\infty\log\left(\frac{\tau}{t_1}\right) \leq C + \left(\sup_{t\geq t_1}\|\dot{u}(t)\|\right)\left(\log\tau\int_\tau^\infty\|g(s)\|ds + \int_{t_1}^{\tau}\|g(t)\|\log t dt\right).$$

Observe that $\lim_{\tau\to+\infty}\int_\tau^\infty\|g(s)\|ds = 0$ since g is integrable. Then, divide the last inequality by $\log\left(\frac{\tau}{t_1}\right)$ and let $\tau\to\infty$. According to Lemma A.5, we get that $W_\infty \leq 0$. Thus $W_\infty = 0$, hence the contradiction. The last statement, $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$, is obtained by following an argument similar to the one above, see also [16], which now uses the perturbed version of the classical energy function, namely

$$W_0(t) := \frac{1}{2}\|\dot{x}(t)\|^2 + (f(x(t)) - f^\star) - \int_t^{+\infty}\langle\dot{x}(\tau),\,g(\tau)\rangle\,d\tau.$$

We do not detail this proof for the sake of brevity. Then, according to $\nabla f(x(t)) = \frac{1}{\beta}(\dot{u}(t)-\dot{x}(t))$, we obtain the convergence of $\nabla f(x(t))$ to zero.

$\square$

3.1.2 Fast convergence rates

We now move on to showing fast convergence of the objective. For this, we will need to strengthen the integrability assumption on the errors. We will denote in this section the two functions

$$w(t) := 1 - \frac{\beta}{t} \text{ and } \delta(t) := t^2 w(t).$$

Theorem 3.2. *Let $f \in \mathcal{C}^2(\mathcal{H})$. Assume that e is $\mathcal{C}^1$ with $\int_{t_0}^{+\infty} t\|e(t)\|dt < +\infty$ and $\int_{t_0}^{+\infty} t\|\dot{e}(t)\|dt < +\infty$. Let x be a solution trajectory to (ISEHD-PERT) for $\alpha > 3$ and $\beta > 0$ such that $t_0 > \frac{\beta(\alpha-2)}{\alpha-3}$. Then the following holds:*

(i) $f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right)$ *as* $t \to +\infty$.

(ii) $\int_{t_0}^{+\infty} t^2\|\nabla f(x(t))\|^2 dt < +\infty$.

(iii) $\int_{t_0}^{+\infty} t\left(f(x(t)) - \inf_{\mathcal{H}} f\right) dt < +\infty$.

(iv) *If, moreover, $t_0 \geq \frac{\beta(\alpha-2-\varepsilon)}{\alpha-3-\varepsilon}$ for some $\varepsilon \in]0, \alpha-3[$, then*

   (a) *for any $x^\star \in S$,* $\int_{t_0}^{+\infty} t\left\langle \nabla f(x(t)),\, x(t) - x^\star\right\rangle dt < +\infty$.

   (b) $\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty$ *and* $\sup_{t \geq t_0} \|x(t)\| < +\infty$.

   (c) $f(x(t)) - \inf_{\mathcal{H}} f = o\left(t^{-2}\right)$ *and* $\|\dot{x}(t)\| = o\left(t^{-1}\right)$ *as* $t \to +\infty$.

Proof. Denote $f^\star := \inf_{\mathcal{H}} f$. By assumption on the parameters, we have $w(t) \geq \frac{1}{\alpha-2} > 0$ for all $t \geq t_0$. Define,

$$v(t) = (\alpha-1)(x(t) - x^\star) + t(\dot{x}(t) + \beta\nabla f(x(t)))$$

and

$$\mathcal{E}(t) := \delta(t)(f(x(t)) - f(x^\star)) + \frac{1}{2}\|v(t)\|^2 - \int_t^T \tau\left\langle v(\tau),\, g(\tau)\right\rangle d\tau.$$

This is a well-defined differentiable function. Taking its derivative in time yields

$$\dot{\mathcal{E}}(t) = \dot{\delta}(t)(f(x(t)) - f(x^\star)) + \delta(t)\left\langle \nabla f(x(t)),\, \dot{x}(t)\right\rangle + \left\langle v(t),\, \dot{v}(t) + tg(t)\right\rangle.$$

From (ISEHD-PERT), we have,

$$\begin{aligned}
\dot{v}(t) &= \alpha\dot{x}(t) + \beta\nabla f(x(t)) + t\left(\ddot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t)\right) \\
&= \alpha\dot{x}(t) + \beta\nabla f(x(t)) + t\left(-\frac{\alpha}{t}\dot{x}(t) - \nabla f(x(t)) - g(t)\right) \\
&= -t\left(1 - \frac{\beta}{t}\right)\nabla f(x(t)) - tg(t).
\end{aligned}$$

Injecting this expression into the scalar product $\left\langle v(t),\, \dot{v}(t) + tg(t)\right\rangle$, developing and rearranging, we obtain

$$\dot{\mathcal{E}}(t) = \dot{\delta}(t)(f(x(t)) - f(x^\star)) - (\alpha-1)tw(t)\left\langle \nabla f(x(t)),\, x(t) - x^\star\right\rangle - \beta\delta(t)\|\nabla f(x(t))\|^2. \tag{21}$$

Convexity of f then yields

$$\dot{\mathcal{E}}(t) + \beta\delta(t)\|\nabla f(x(t))\|^2 + \left((\alpha-1)tw(t) - \dot{\delta}(t)\right)(f(x(t)) - f(x^\star)) \leq 0. \tag{22}$$

By assumption on the parameters, we have for any $t \geq t_0 > \frac{\beta(\alpha-2)}{\alpha-3}$,

$$(\alpha-1)tw(t) - \dot{\delta}(t) = t((\alpha-3)w(t) - t\dot{w}(t)) \geq ct$$

for some constant $c > 0$. Thus (22) implies that $\mathcal{E}$ is non-increasing. In turn,

$$\delta(t)f(x(t)) - f(x^\star) + \frac{1}{2}\|v(t)\|^2 \leq C - \int_{t_0}^{t} \tau \left\langle v(\tau),\, g(\tau)\right\rangle d\tau \tag{23}$$

with $C = \delta(t_0)(f(x_0) - f(x^\star)) + \frac{1}{2}\|v(t_0)\|^2$, from which we conclude that

$$\frac{1}{2}\|v(t)\|^2 \leq C + \int_{t_0}^{t} \|v(\tau)\|\left(\tau\|g(\tau)\|\right) d\tau.$$

Lemma A.3 then gives

$$\|v(t)\| \leq (2C)^{1/2} + \int_{t_0}^{t} \tau\left(\|e(\tau)\| + \beta\|\dot{e}(\tau)\|\right) d\tau, \tag{24}$$

and thus

$$\sup_{t\geq t_0} \|v(t)\| < +\infty. \tag{25}$$

Using this into (23) then shows that

$$\frac{t^2}{\alpha-2}\left(f(x(t)) - f(x^\star)\right) \leq \delta(t)\left(f(x(t)) - f(x^\star)\right) \leq C + \sup_{t\geq t_0}\|v(t)\| \int_{t_0}^{t} \left(\tau\|e(\tau)\| + \beta\tau\|\dot{e}(\tau)\|\right) d\tau < +\infty \tag{26}$$

hence proving statement (i). Finally, integrating (22) and using the fact that $\mathcal{E}(t)$ is bounded from below by (25) and the assumptions on the errors, we get

$$\beta \int_{t_0}^{+\infty} t^2 w(t)\|\nabla f(x(t))\|^2 dt \leq C,$$

and

$$c\int_{t_0}^{+\infty} t(f(x(t)) - f(x^\star)) \leq \int_{t_0}^{+\infty} \left((\alpha-1)tw(t) + \dot{\delta}(t)\right)(f(x(t)) - f(x^\star)) \leq C.$$

for some constant $C > 0$. This shows the integral estimates (ii) and (iii).

Let us turn to statement (iv)(a). We embark from (21) to write, for some $\rho \in]0,1[$ to be chosen shortly,

$$\begin{aligned}
\dot{\mathcal{E}}(t) &= \dot{\delta}(t)(f(x(t)) - f(x^\star)) - (1-\rho)(\alpha-1)tw(t)\left\langle \nabla f(x(t)),\, x(t) - x^\star\right\rangle \\
&\quad - \rho(\alpha-1)tw(t)\left\langle \nabla f(x(t)),\, x(t) - x^\star\right\rangle - \beta\delta(t)\|\nabla f(x(t))\|^2 \\
&\leq -\left((1-\rho)(\alpha-1)tw(t) - \dot{\delta}(t)\right)(f(x(t)) - f(x^\star)) - \rho(\alpha-1)tw(t)\left\langle \nabla f(x(t)),\, x(t) - x^\star\right\rangle - \beta\delta(t)\|\nabla f(x(t))\|^2.
\end{aligned}$$

To conclude, it remains to check that $\left((1-\rho)(\alpha-1)tw(t) - \dot{\delta}(t)\right)$ is non-negative. Taking $\rho = \varepsilon/(\alpha - 1) \in]0,1[$, in view of the assumption on the parameters, we have

$$\begin{aligned}
(1-\rho)(\alpha-1)tw(t) - \dot{\delta}(t) &= (\alpha-1-\varepsilon)tw(t) - \dot{\delta}(t) = t((\alpha-3-\varepsilon)w(t) - t\dot{w}(t)) \\
&= t\left((\alpha-3-\varepsilon) - \frac{(\alpha-2-\varepsilon)\beta}{t}\right) \geq t_0\left((\alpha-3-\varepsilon) - \frac{(\alpha-2-\varepsilon)\beta}{t_0}\right) \geq 0.
\end{aligned}$$

For claim (iv)(b), we multiply (ISEHD-Pert) by $t^2\dot{x}(t)$ to get

$$t^2\left\langle \ddot{x}(t),\, \dot{x}(t)\right\rangle + \alpha t\left\|\dot{x}(t)\right\|^2 + t^2\beta\left\langle \dot{x}(t),\, \nabla^2 f(x(t))\dot{x}(t)\right\rangle + t^2\left\langle \nabla f(x(t)),\, \dot{x}(t)\right\rangle + t^2\left\langle g(t),\, \dot{x}(t)\right\rangle = 0.$$

With the chain rule, Cauchy-Schwarz inequality and convexity of f, we obtain

$$\frac{1}{2}t^2\frac{d}{dt}\|\dot{x}(t)\|^2+\alpha t\|\dot{x}(t)\|^2+t^2\frac{d}{dt}(f(x(t))-f^\star)\leq\|tg(t)\|\|t\dot{x}(t)\|. \tag{27}$$

Integrating by parts on $[t_0,t]$ we get,

$$\frac{t^2}{2}\|\dot{x}(t)\|^2+(\alpha-1)\int_{t_0}^{t}s\|\dot{x}(s)\|^2ds\leq C_0+2\int_{t_0}^{t}s(f(x(s))-f^\star)ds+\int_{t_0}^{t}\|sg(s)\|\|s\dot{x}(s)\|ds \tag{28}$$

for some non-negative constant C_0, where we have used claim (i) of Theorem 3.2. Now by claim (iii) of Theorem 3.2 and ignoring the non-negative terms since $\alpha>1$, we obtain

$$\frac{1}{2}\|t\dot{x}(t)\|^2\leq C_1+\int_{t_0}^{t}\|sg(s)\|\|s\dot{x}(s)\|ds,$$

for another non-negative constant C_1. Lemma A.3 then yields

$$\sup_{t\geq t_0}t\|\dot{x}(t)\|<+\infty. \tag{29}$$

Using this in (28), we get also that

$$\int_{t_0}^{+\infty}t\|\dot{x}(t)\|^2dt<+\infty. \tag{30}$$

Moreover, (25), (29) and claim (ii) of Theorem 3.2 entail

$$\sup_{t\geq t_0}\|x(t)\|<+\infty.$$

We finally turn to statement (iv)(c). We embark from (27), use (29), and integrate on $[s,t]$ to see that

$$t^2\left(\frac{1}{2}\|\dot{x}(t)\|^2+(f(x(t))-f^\star)\right)-s^2\left(\frac{1}{2}\|\dot{x}(s)\|^2+(f(x(s))-f^\star)\right)$$
$$+(\alpha-1)\int_{s}^{t}\tau\|\dot{x}(\tau)\|^2d\tau-2\int_{s}^{t}\tau(f(x(\tau))-f^\star)d\tau-C\int_{s}^{t}\|\tau g(\tau)\|d\tau\leq 0,$$

where $C=\sup_{t\geq t_0}t\|\dot{x}(t)\|$. This means that the function

$$\mathcal{G}(t)=t^2\left(\frac{1}{2}\|\dot{x}(t)\|^2+(f(x(t))-f^\star)\right)+(\alpha-1)\int_{t_0}^{t}\tau\|\dot{x}(\tau)\|^2d\tau-2\int_{t_0}^{t}\tau(f(x(\tau))-f^\star)d\tau-C\int_{t_0}^{t}\|\tau g(\tau)\|d\tau$$

is non-increasing on $[t_0,+\infty[$. Since it is bounded from below by assumption on the errors and claim (iii), $\lim_{t\to+\infty}\mathcal{G}(t)$ exists. This together with assertions (iii) and (iv)(b) shows that the limit

$$0\leq L:=\lim_{t\to+\infty}t^2\left(\frac{1}{2}\|\dot{x}(t)\|^2+(f(x(t))-f^\star)\right)$$

exists. Suppose that $L>0$. Then, there exists $s\geq t_0$ such that

$$\int_{t_0}^{+\infty}\left(\frac{t}{2}\|\dot{x}(t)\|^2+t(f(x(t))-f^\star)\right)dt\geq\int_{s}^{+\infty}t^2\left(\frac{1}{2}\|\dot{x}(t)\|^2+(f(x(t))-f^\star)\right)t^{-1}dt\geq\int_{s}^{+\infty}\frac{L}{2t}dt=+\infty,$$

leading to a contradiction with claims (iii) and (iv)(b). $\square$

3.1.3 Convergence of the trajectories

We complete our analysis by showing weak convergence of the trajectories.

Theorem 3.3. *Assume that e is $\mathcal{C}^1$ with $\int_{t_0}^{+\infty}t\|e(t)\|dt<+\infty$ and $\int_{t_0}^{+\infty}t\|\dot{e}(t)\|dt<+\infty$. Let $x(t)$ be a solution trajectory to* (ISEHD-Pert) *for $\alpha>3$ and $\beta>0$ from initial time $t_0\geq\frac{\beta(\alpha-2-\varepsilon)}{\alpha-3-\varepsilon}$ for some $\varepsilon\in]0,\alpha-3[$. Then $x(t)$ converges weakly to a minimizer of f.*

Proof. Keeping in mind that the goal is to apply Opial's Lemma (see Lemma A.1), we will now show that $\lim_{t\to+\infty}\|x(t)-x^\star\|$ exists. Recall the Lyapunov function $\mathcal{E}$ from the proof of Theorem 3.2, and define its generalized version

$$\mathcal{E}_\varepsilon(t) := (\delta(t)+\varepsilon\beta t)\left(f(x(t))-f(x^\star)\right)+\frac{1}{2}\|v_\varepsilon(t)\|^2+\frac{\varepsilon(\alpha-1-\varepsilon)}{2}\|x(t)-x^\star\|^2-\int_t^T \tau\left\langle v_\varepsilon(\tau),\, g(\tau)\right\rangle d\tau, \tag{31}$$

where

$$v_\varepsilon(t)=(\alpha-1-\varepsilon)(x(t)-x^\star)+t(\dot{x}(t)+\beta\nabla f(x(t))).$$

One can check, arguing as for $\mathcal{E}$, that

$$\dot{\mathcal{E}}_\varepsilon(t)=\left(\dot{\delta}(t)+\varepsilon\beta\right)(f(x(t))-f(x^\star))-(\alpha-1-\varepsilon)tw(t)\left\langle\nabla f(x(t)),\, x(t)-x^\star\right\rangle-\beta\delta(t)\|\nabla f(x(t))\|^2-\varepsilon t\,\|\dot{x}(t)\|^2.$$

Convexity of f then entails

$$\dot{\mathcal{E}}_\varepsilon(t)+\left((\alpha-1-\varepsilon)tw(t)-\dot{\delta}(t)\right)(f(x(t))-f(x^\star))+\beta\delta(t)\|\nabla f(x(t))\|^2+\varepsilon t\,\|\dot{x}(t)\|^2\le 0.$$

The assumption on the parameters gives

$$(\alpha-1-\varepsilon)tw(t)-\dot{\delta}(t)=t((\alpha-3-\varepsilon)w(t)-t\dot{w}(t))\ge t_0\left((\alpha-3-\varepsilon)-\frac{(\alpha-2-\varepsilon)\beta}{t_0}\right)\ge 0.$$

Thus, ignoring the non-negative terms in this inequality entails that $\mathcal{E}_\varepsilon(t)$ is a decreasing function on $[t_0,\infty[$. As a consequence, the energy functions $\mathcal{E}(t)$ and $\mathcal{E}_\varepsilon(t)$ with $T=+\infty$ are well-defined on $[t_0,+\infty[$, and are then Lyapunov functions for the dynamical system (ISEHD-Pert). Both $\mathcal{E}(t)$ and $\mathcal{E}_\varepsilon(t)$ thus have limits as $t\to+\infty$, and so does their difference

$$\begin{aligned}\mathcal{E}_\varepsilon(t)-\mathcal{E}(t)=\varepsilon\beta t(f(x(t))-f(x^\star))-\frac{\varepsilon(\alpha-1)}{2}\|x(t)-x^\star\|^2-\varepsilon t\left\langle\dot{x}(t),\, x(t)-x^\star\right\rangle\\ -\varepsilon t\left\langle\nabla f(x(t)),\, x(t)-x^\star\right\rangle+\varepsilon\int_t^{+\infty}\tau\left\langle x(\tau)-x^\star,\, g(\tau)\right\rangle d\tau.\end{aligned}$$

By Theorem 3.2(i), the first term converges to 0 as $t\to+\infty$. By the integrability assumptions on the errors and boundedness of $x(t)$ (see Theorem 3.2), the last term also converges to 0 as $t\to+\infty$. We have then shown that the limit as t goes to infinity of

$$p(t):=\frac{\alpha-1}{2}\|x(t)-x^\star\|^2+t\left\langle\dot{x}(t),\, x(t)-x^\star\right\rangle+t\left\langle\nabla f(x(t)),\, x(t)-x^\star\right\rangle.$$

exists. Set

$$q(t):=\frac{\alpha-1}{2}\|x(t)-x^\star\|^2+(\alpha-1)\int_{t_0}^t\left\langle\nabla f(x(s)),\, x(s)-x^\star\right\rangle ds.$$

We obviously have

$$p(t)=q(t)+\frac{t}{\alpha-1}\dot{q}(t)-(\alpha-1)\int_{t_0}^t\left\langle\nabla f(x(s)),\, x(s)-x^\star\right\rangle ds.$$

By Theorem 3.2(iv)(a), and since $\left\langle\nabla f(x(s)),\, x(s)-x^\star\right\rangle$ is non-negative, we have that

$$\lim_{t\to+\infty}\int_{t_0}^t\left\langle\nabla f(x(s)),\, x(s)-x^\star\right\rangle ds \tag{32}$$

exists. Overall, we have shown that

$$\lim_{t\to+\infty}\left(q(t)+\frac{t}{\alpha-1}\dot{q}(t)\right)$$

exists. Since $\alpha>1$, it follows from [16, Lemma 7.2] that $\lim_{t\to+\infty}q(t)$ exists, and using again (32), we deduce that $\lim_{t\to+\infty}\|x(t)-x^\star\|$ exists for any $x^\star\in S$. From claim (i) of Theorem 3.2 (see also Lemma 3.1(iv)), it follows that for any sequence $(x(t_n))_{n\in\mathbb{N}}$ which converges weakly, to say $\bar{x}$, we have

$$f(\bar{x})\le\liminf_{n\to+\infty}f(x(t_n))=\lim_{t\to+\infty}f(x(t))=f^\star,$$

i.e. , $\bar{x}\in S$. Consequently, all the conditions of Lemma A.1 are satisfied, hence the weak convergence of the trajectories. □

3.2 Implicit Hessian Damping

We now turn to the second-order ODE (ISIHD-PERT) where $f \in \mathcal{C}^1(\mathcal{H})$ and $\beta(t) = \gamma + \frac{\beta}{t}$, $\gamma, \beta \geq 0$. Let us denote for brevity $f^\star := \inf_{\mathcal{H}} f$. Given $x^\star \in S$, we consider the function

$$\begin{aligned}&\mathcal{E}(t) = a(t)\left(f\left(x(t) + \beta(t)\dot{x}(t)\right) - f^\star\right) + \frac{1}{2}\|b(t)(x(t) - x^\star) + c(t)\dot{x}(t)\|^2 + \frac{d(t)}{2}\|x(t) - x^\star\|^2 \\ &- \int_t^{+\infty} c(\tau)\left\langle b(\tau)(x(\tau) - x^\star) + c(\tau)\dot{x}(\tau),\, e(\tau)\right\rangle d\tau - \int_t^{+\infty} a(\tau)\beta(\tau)\left\langle \nabla f\left(x(\tau) + \beta(\tau)\dot{x}(\tau)\right), e(\tau)\right\rangle d\tau\end{aligned} \tag{33}$$

parametrized by some functions $a(t)$, $b(t)$, $c(t)$ and $d(t)$ to be specified later.

3.2.1 Lyapunov function

We first show that for proper choices of $(a(t), b(t), c(t), d(t))$ as a function of the problem parameters (α, γ, β), $\mathcal{E}$ can serve as a Lyapunov function for (ISIHD-PERT). We will denote for short $\alpha(t) = \frac{\alpha}{t}$.

Lemma 3.4. *Assume that,*

$$\begin{cases} \dot{a}(t) - b(t)c(t) & \leq 0, \\ -a(t)\beta(t) & \leq 0, \\ -a(t)\alpha(t)\beta(t) + a(t)\dot{\beta}(t) + a(t) - c(t)^2 + b(t)c(t)\beta(t) & = 0, \\ \dot{b}(t)b(t) + \frac{\dot{d}(t)}{2} & \leq 0, \\ \dot{b}(t)c(t) + b(t)(b(t) + \dot{c}(t) - c(t)\alpha(t)) + d(t) & = 0, \\ c(t)(b(t) + \dot{c}(t) - c(t)\alpha(t)) & \leq 0. \end{cases} \tag{34}$$

Then

$$\begin{aligned}\dot{\mathcal{E}}(t) \;\leq\;& (\dot{a}(t) - b(t)c(t))(f(x(t) + \beta(t)\dot{x}(t)) - f^\star) - a(t)\beta(t)\|\nabla f(x(t) + \beta(t)\dot{x}(t))\|^2 \\ &+ \left(\dot{b}(t)b(t) + \frac{\dot{d}(t)}{2}\right)\|x(t) - x^\star\|^2 + c(t)(b(t) + \dot{c}(t) - c(t)\alpha(t))\,\|\dot{x}(t)\|^2 \leq 0. \end{aligned} \tag{35}$$

Proof. We proceed as in the proof of Theorems 3.1 and 3.2, and first consider the function $\mathcal{E}_T$ where the integrals involving the error terms are calculated on $[t, T]$, $T < +\infty$. This shows that $\mathcal{E}$ is well-posed under our assumptions. We can then compute the time derivative of $\mathcal{E}$ and use the chain rule to get

$$\begin{aligned}\dot{\mathcal{E}}(t) \;=\;& \dot{a}(t)\left(f\left(x(t) + \beta(t)\dot{x}(t)\right) - f^\star\right) + a(t)\left\langle \nabla f\left(x(t) + \beta(t)\dot{x}(t)\right), \dot{x}(t) + \dot{\beta}(t)\dot{x}(t) + \beta(t)\ddot{x}(t)\right\rangle \\ &+ \left\langle (b(t) + \dot{c}(t))\dot{x}(t) + c(t)\ddot{x}(t) + \dot{b}(t)(x(t) - x^\star),\, b(t)(x(t) - x^\star) + c(t)\dot{x}(t)\right\rangle \\ &+ \frac{\dot{d}(t)}{2}\|x(t) - x^\star\|^2 + d(t)\left\langle \dot{x}(t),\, x(t) - x^\star\right\rangle \\ &+ c(t)\left\langle b(t)(x(t) - x^\star) + c(t)\dot{x}(t),\, e(t)\right\rangle + a(t)\beta(t)\left\langle \nabla f\left(x(t) + \beta(t)\dot{x}(t)\right), e(t)\right\rangle \end{aligned} \tag{36}$$

Using (ISIHD-PERT) in the second term of (36), we get

$$\begin{aligned}&a(t)\left\langle \nabla f\left(x(t) + \beta(t)\dot{x}(t)\right), \dot{x}(t) + \dot{\beta}(t)\dot{x}(t) + \beta(t)\ddot{x}(t)\right\rangle \\ &= a(t)\left\langle \nabla f\left(x(t) + \beta(t)\dot{x}(t)\right), \left(1 + \dot{\beta}(t) - \alpha(t)\beta(t)\right)\dot{x}(t) - \beta(t)\nabla f\left(x(t) + \beta(t)\dot{x}(t)\right) - \beta(t)e(t)\right\rangle \\ &= -a(t)\beta(t)\left\|\nabla f\left(x(t) + \beta(t)\dot{x}(t)\right)\right\|^2 + \left(1 + \dot{\beta}(t) - \alpha(t)\beta(t)\right)a(t)\left\langle \nabla f\left(x(t) + \beta(t)\dot{x}(t)\right), \dot{x}(t)\right\rangle \\ &\qquad - \beta(t)a(t)\left\langle \nabla f\left(x(t) + \beta(t)\dot{x}(t)\right), e(t)\right\rangle. \end{aligned} \tag{37}$$

We expand the third term in (36) as

$$
\begin{aligned}
&\Big\langle (b(t)+\dot{c}(t))\dot{x}(t)+c(t)\ddot{x}(t)+\dot{b}(t)(x(t)-x^\star),\, b(t)(x(t)-x^\star)+c(t)\dot{x}(t)\Big\rangle\\
&=\langle (b(t)+\dot{c}(t)-c(t)\alpha(t))\dot{x}(t)-c(t)\nabla f\left(x(t)+\beta(t)\dot{x}(t)\right),\, b(t)(x(t)-x^\star)+c(t)\dot{x}(t)\rangle\\
&\quad+\Big\langle -c(t)e(t)+\dot{b}(t)(x(t)-x^\star),\, b(t)(x(t)-x^\star)+c(t)\dot{x}(t)\Big\rangle\\
&=-c(t)\left\langle \nabla f(x(t)+\beta(t)\dot{x}(t)),\, b(t)(x(t)-x^\star)+c(t)\dot{x}(t)\right\rangle\\
&\quad+\Big\langle \dot{b}(t)(x(t)-x^\star),\, b(t)(x(t)-x^\star)+c(t)\dot{x}(t)\Big\rangle\\
&\quad+c(t)(b(t)+\dot{c}(t)-c(t)\alpha(t))\left\|\dot{x}(t)\right\|^2-c(t)\left\langle e(t),\, b(t)(x(t)-x^\star)+c(t)\dot{x}(t)\right\rangle. \qquad (38)
\end{aligned}
$$

Plugging (37) and (38) into (36), we get,

$$
\begin{aligned}
\dot{\mathcal{E}}(t) &= \dot{a}(t)\left(f\left(x(t)+\beta(t)\dot{x}(t)\right)-f^\star\right)-a(t)\beta(t)\left\|\nabla f\left(x(t)+\beta(t)\dot{x}(t)\right)\right\|^2\\
&+ c(t)(b(t)+\dot{c}(t)-c(t)\alpha(t))\left\|\dot{x}(t)\right\|^2+\left(\dot{b}(t)b(t)+\frac{\dot{d}(t)}{2}\right)\left\|x(t)-x^\star\right\|^2\\
&+ \Big(b(t)^2+b(t)\dot{c}(t)+\dot{b}(t)c(t)-b(t)c(t)\alpha(t)+d(t)\Big)\left\langle \dot{x}(t),\, x(t)-x^\star\right\rangle\\
&+ \Big(-a(t)\alpha(t)\beta(t)+a(t)\dot{\beta}(t)+a(t)-c(t)^2\Big)\left\langle \nabla f\left(x(t)+\beta(t)\dot{x}(t)\right),\, \dot{x}(t)\right\rangle\\
&- b(t)c(t)\left\langle \nabla f\left(x(t)+\beta(t)\dot{x}(t)\right),\, x(t)-x^\star\right\rangle. \qquad (39)
\end{aligned}
$$

Since,

$\langle\nabla f(x(t)+\beta(t)\dot{x}(t)),\, x(t)-x^\star\rangle=\langle\nabla f(x(t)+\beta(t)\dot{x}(t)),\, x(t)+\beta(t)\dot{x}(t)-x^\star\rangle-\langle\nabla f(x(t)+\beta(t)\dot{x}(t)),\, \beta(t)\dot{x}(t)\rangle$
and using the convex (sub)differential inequality on f, we can write

$$
\begin{aligned}
&-b(t)c(t)\left\langle \nabla f(x(t)+\beta(t)\dot{x}(t),\, x(t)-x^\star\right\rangle\\
&\qquad\qquad\leq -b(t)c(t)(f(x(t)+\beta(t)\dot{x}(t))-f^\star)+b(t)c(t)\beta(t)\left\langle \nabla f(x(t)+\beta(t)\dot{x}(t)),\, \dot{x}(t)\right\rangle
\end{aligned}
$$

and we arrive at

$$
\begin{aligned}
&\dot{\mathcal{E}}(t)\leq (\dot{a}(t)-b(t)c(t))(f(x(t)+\beta(t)\dot{x}(t))-f^\star)-a(t)\beta(t)\|\nabla f(x(t)+\beta(t)\dot{x}(t))\|^2\\
&+(-a(t)\alpha(t)\beta(t)+a(t)\dot{\beta}(t)+a(t)-c(t)^2+b(t)c(t)\beta(t))\left\langle \nabla f(x(t)+\beta(t)\dot{x}(t)),\, \dot{x}(t)\right\rangle\\
&+\left(\dot{b}(t)b(t)+\frac{\dot{d}(t)}{2}\right)\left\|x(t)-x^\star\right\|^2\\
&+(b(t)^2+b(t)\dot{c}(t)+\dot{b}(t)c(t)-b(t)c(t)\alpha(t)+d(t))\left\langle \dot{x}(t),\, x(t)-x^\star\right\rangle+c(t)(b(t)+\dot{c}(t)-c(t)\alpha(t))\left\|\dot{x}(t)\right\|^2.
\end{aligned}
$$

Thus, conditions (34) guarantee that $\dot{\mathcal{E}}(t)\leq 0$, in particular they imply (35). $\square$

Following the discussion of [3, Remark 11], in the rest of the section, we take

$$
\begin{aligned}
\beta(t)=\gamma+\frac{\beta}{t},\quad \gamma,\beta\geq 0,\\
b(t)\equiv b\in]0,\alpha-1],\alpha>1,\qquad c(t)=t\ \text{ and }\ d(t)\equiv b(\alpha-1-b).
\end{aligned}
\qquad (40)
$$

Such a choice is reminescent of that in (31). The choices of $d(t)$ and $b(t)$ comply with the fourth, fifth and sixth conditions of (34). To satisfy the third condition, one has to take

$$
a(t)=t^2\left(1+\frac{(\alpha-b)\gamma t-\beta(\alpha+1-b)}{t^2-\alpha\gamma t-\beta(\alpha+1)}\right). \qquad (41)
$$

Clearly, for t large enough, one has $a(t)\geq t^2$ and $\beta(t)\geq\gamma/2$. Thus, the second condition is in force. One can also verify that the first inequality is satisfied for t large enough provided that $b>2$ (and thus $\alpha>3$) when $\gamma>0$, and $b=2$ (with $\alpha=3$) when $\gamma=0$.

3.2.2 Fast convergence rates

We start with the following boundedness properties.

Lemma 3.5. *Let*

$$E(t) = a(t)\left(f\left(x(t) + \beta(t)\dot{x}(t)\right) - f^\star\right) + \frac{1}{2}\|b(x(t)-x^\star) + t\dot{x}(t)\|^2 + \frac{b(\alpha-1-b)}{2}\|x(t)-x^\star\|^2.$$

Choose the parameters according to (40)-(41) with $\alpha > 3$, $\gamma > 0$. Define, for $t \geq t_0$,

$$m(t) := \max\left(t, L|a(t)\beta(t)|, L|a(t)|\beta(t)^2\right). \tag{42}$$

Assume that ∇f is L-Lipschitz continuous and $m(\cdot)e(\cdot) \in L^1(t_0, +\infty; \mathcal{H})$. Then, we have

$$\sup_{t\geq t_0} E(t) < +\infty,\ \sup_{t\geq t_0} t\,\|\dot{x}(t)\| < +\infty \text{ and } \sup_{t\geq t_0}\|x(t)-x^\star\| < +\infty.$$

Proof. Consider the function $\mathcal{E}(t)$ in (33) with the choices (40)-(41) for $c(t)$, $d(t)$, $b(t)$ and $a(t)$, with $b \in]2, \alpha-1[$. For such a choice, there exists $t_1 \geq t_0$ such that $a(t) > 0$, $\beta(t) > 0$ (and in turn, $m(t) > 0$), and all conditions of (34) are satisfied. Thus, $\mathcal{E}(t)$ is monotonically decreasing on $[t_1, +\infty[$ according to Lemma 3.4. We then have that, for $t > t_1$,

$$\begin{aligned}
&E(t) \leq E(t_1) + \int_{t_1}^t \left\langle \tau\left(b(x(\tau)-x^\star) + \tau\dot{x}(\tau)\right) + a(\tau)\beta(\tau)\nabla f\left(x(\tau)+\beta(\tau)\dot{x}(\tau)\right), e(\tau)\right\rangle d\tau \\
&= E(t_1) + \int_{t_1}^t \left\langle \tau\left(b(x(\tau)-x^\star) + \tau\dot{x}(\tau)\right) + a(\tau)\beta(\tau)\left(\nabla f\left(x(\tau)+\beta(\tau)\dot{x}(\tau)\right) - \nabla f(x^\star)\right), e(\tau)\right\rangle d\tau \\
&\leq E(t_1) + \int_{t_1}^t \Big(\tau\left\|b(x(\tau)-x^\star)+\tau\dot{x}(\tau)\right\| + a(\tau)\beta(\tau)\left\|\nabla f\left(x(\tau)+\beta(\tau)\dot{x}(\tau)\right) - \nabla f(x^\star)\right\|\Big)\|e(\tau)\|\,d\tau \\
&\leq E(t_1) + \int_{t_1}^t \Big(\tau\left\|b(x(\tau)-x^\star)+\tau\dot{x}(\tau)\right\| + a(\tau)\beta(\tau)L\left\|x(\tau)-x^\star+\beta(\tau)\dot{x}(\tau)\right\|\Big)\|e(\tau)\|\,d\tau \\
&\leq E(t_1) + \int_{t_1}^t \Big(\tau\left\|b(x(\tau)-x^\star)+\tau\dot{x}(\tau)\right\| + La(\tau)\beta(\tau)\left\|x(\tau)-x^\star\right\| + La(\tau)\beta(\tau)^2\left\|\dot{x}(\tau)\right\|\Big)\|e(\tau)\|\,d\tau.
\end{aligned} \tag{43}$$

Denote $d = b(\alpha-1-b)$. We have $d > 0$. Moreover, $a(t) > 0$ for $t \geq t_1$. One can then drop the first term in $E(t)$, and (43) becomes, for any $t \geq t_1$,

$$\begin{aligned}
&\frac{1}{2}\|b(x(t)-x^\star)+t\dot{x}(t)\|^2 + \frac{d}{2}\|x(t)-x^\star\|^2 \\
&\leq E(t_1) + \int_{t_1}^t \Big(\tau\left\|b(x(\tau)-x^\star)+\tau\dot{x}(\tau)\right\| + \sqrt{d}\frac{La(\tau)\beta(\tau)}{\sqrt{d}}\left\|x(\tau)-x^\star\right\| + \tau\frac{La(\tau)\beta(\tau)^2}{t_1}\left\|\dot{x}(\tau)\right\|\Big)\|e(\tau)\|\,d\tau \\
&\leq E(t_1) + \int_{t_1}^t \Big(\left\|b(x(\tau)-x^\star)+\tau\dot{x}(\tau)\right\| + \sqrt{d}\left\|x(\tau)-x^\star\right\| + \tau\left\|\dot{x}(\tau)\right\|\Big)\max\left(1, d^{-1/2}, t_1^{-1}\right)m(\tau)\,\|e(\tau)\|\,d\tau \\
&\leq E(t_1) + \int_{t_1}^t \Big(2\left\|b(x(\tau)-x^\star)+\tau\dot{x}(\tau)\right\| + (b+\sqrt{d})\left\|x(\tau)-x^\star\right\|\Big)\max\left(1, d^{-1/2}, t_1^{-1}\right)m(\tau)\,\|e(\tau)\|\,d\tau \\
&\leq E(t_1) + \int_{t_1}^t \Big(\left\|b(x(\tau)-x^\star)+\tau\dot{x}(\tau)\right\| + \sqrt{d}\left\|x(\tau)-x^\star\right\|\Big)Cm(\tau)\,\|e(\tau)\|\,d\tau,
\end{aligned}$$

for some constant $C \geq \max\left(1, d^{-1/2}, t_1^{-1}\right)\max\left(2, 1+\sqrt{\frac{b}{\alpha-1-b}}\right)$. Now, Jensen's inequality yields

$$\begin{aligned}
&\frac{1}{4}\left(\|b(x(t)-x^\star)+t\dot{x}(t)\| + \sqrt{d}\|x(t)-x^\star\|\right)^2 \leq \frac{1}{2}\|b(x(t)-x^\star)+t\dot{x}(t)\|^2 + \frac{d}{2}\|x(t)-x^\star\|^2 \\
&\leq E(t_1) + \int_{t_1}^t \Big(\left\|b(x(\tau)-x^\star)+\tau\dot{x}(\tau)\right\| + \sqrt{d}\left\|x(\tau)-x^\star\right\|\Big)C|m(\tau)|\,\|e(\tau)\|\,d\tau.
\end{aligned}$$

Using the Gronwall Lemma A.3, we conclude that, for all $t \geq t_1$

$$\|b(x(t) - x^\star) + t\dot{x}(t)\| + \sqrt{d}\|x(t) - x^\star\| \leq 2\sqrt{|E(t_0)|} + 2C \int_{t_0}^{+\infty} |m(\tau)|\|e(\tau)\| d\tau < +\infty, \tag{44}$$

whence we get boundedness of $\|x(t)-x^\star\|$ and $\|b(x(t)-x^\star)+t\dot{x}(t)\|$. The triangle inequality then shows that $t\,\|\dot{x}(t)\|$ is also bounded. Using this into (43) together with Cauchy-Schwarz inequality and our integrability assumption, we deduce boundedness of $E(t)$. $\square$

Remark 3.6. *Recall our discussion on the parameters in (40)-(41). Notice that we have $t^2 \leq a(t) \leq t^2 + \kappa_1$ and $\gamma/2 \leq \beta(t) \leq \beta_0$ for t large enough, where $\beta_0 = \gamma + |\beta|/t_0$ and κ is a non-negative constant. In turn, for t large enough, we have*

$$\max\left(1, \gamma/2\right) L\gamma/2t^2 \leq m(t) \leq \max\left(1, \beta_0\right) L\beta_0(t + \kappa_1)^2.$$

Clearly the condition $m(\cdot)e(\cdot) \in L^1(t_0, +\infty; \mathcal{H})$ is equivalent to $t^2 e(t) \in L^1(t_0, +\infty; \mathcal{H})$.

From Lemma 3.5, we obtain the following convergence rates and integral estimates.

Theorem 3.7. *Under the assumptions of Lemma 3.5, there exists $t_1 \geq t_0$ such that the following holds:*

(i) $f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right)$ *for all $t \geq t_1$;*

(ii) $\|\dot{x}(t)\| = \mathcal{O}\left(\frac{1}{t}\right)$ *for all $t \geq t_1$;*

(iii) $\int_{t_0}^{+\infty} t\left(f(x(t)) - \inf_{\mathcal{H}} f\right) dt < +\infty$*;*

(iv) $\int_{t_0}^{+\infty} t^2 \left\|\nabla f(x(t) + \beta(t)\dot{x}(t)\right\|^2 dt < +\infty$*;*

(v) $\int_{t_0}^{+\infty} t \left\|\dot{x}(t)\right\|^2 dt < +\infty$.

Proof. Claim (ii) follows from Lemma 3.5. Discarding the non-negative terms in $E(t)$, Lemma 3.5 together with the fact that $a(t) \geq t^2$ for t large enough, also gives

$$f(x(t) + \beta(t)\dot{x}(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right)$$

for t large enough. Let $\beta_0 = \gamma + |\beta|/t_0$. By the descent lemma,

$$\begin{aligned} f(x(t)) - f(x(t) + \beta(t)\dot{x}(t)) &\leq -\beta(t)\left\langle \nabla f(x(t) + \beta(t)\dot{x}(t)),\, \dot{x}(t)\right\rangle + \frac{L}{2}\beta(t)^2\left\|\dot{x}(t)\right\|^2 \\ &\leq \beta_0 \left\|\nabla f(x(t) + \beta(t)\dot{x}(t))\right\|\left\|\dot{x}(t)\right\| + \frac{L}{2}\beta_0^2\left\|\dot{x}(t)\right\|^2 \\ &= \mathcal{O}\left(\frac{1}{t}\right)\left\|\nabla f(x(t) + \beta(t)\dot{x}(t))\right\| + \mathcal{O}\left(\frac{1}{t^2}\right), \end{aligned} \tag{45}$$

where we used claim (ii) in the last inequality. According to (iv), to be shown shortly, we have that $\|\nabla f(x(t) + \beta(t)\dot{x}(t))\| = o\left(t^{-1}\right)$, whence claim (i) follows.

To show the remaining integral estimates, consider the function $\mathcal{E}(\cdot)$ in (33) with the choices (40)-(41) of $c(t)$, $d(t)$, $b(t)$ and $a(t)$, where $b \in]2, \alpha - 1[$. We first argue similarly to [3] to show that for t large enough, we have $\dot{a}(t) - bt \leq -\frac{(\alpha-3)t}{2}$, since $\alpha > 3$ and $b > 2$. In addition, for (a possibly different) t large enough, it is straightforward to see that $a(t)\,(\gamma + \beta/t) \geq t^2\gamma/2$. With these bounds, (35) reads, for t large enough,

$$\dot{\mathcal{E}}(t) \leq -\frac{(\alpha-3)t}{2}(f(x(t)+\beta(t)\dot{x}(t))-f^\star)-t^2\gamma/2\,\|\nabla f(x(t)+\beta(t)\dot{x}(t))\|^2 - t(\alpha-1-b)\,\|\dot{x}(t)\|^2\,. \quad (46)$$

Integrating (46), and using that $\mathcal{E}$ is bounded thanks to Lemma 3.5, we get statements (iv)-(v) and

$$\int_{t_0}^{+\infty} t\left(f(x(t)+\beta(t)\dot{x}(t)) - \inf_{\mathcal{H}} f\right) dt < +\infty. \quad (47)$$

From the descent lemma and Cauchy-Schwarz inequality, we have

$$\int_{t_0}^{+\infty} t\left(f(x(t)) - \inf_{\mathcal{H}} f\right) dt \leq \int_{t_0}^{+\infty} t\left(f(x(t)+\beta(t)\dot{x}(t)) - \min f(\mathcal{H})\right) dt$$

$$+\frac{\beta_0}{\sqrt{t_0}}\left(\int_{t_0}^{+\infty} t^2\,\|\nabla f(x(t)+\beta(t)\dot{x}(t))\|^2\,dt\right)^{1/2}\left(\int_{t_0}^{+\infty} t\,\|\dot{x}(t)\|^2\,dt\right)^{1/2} + \frac{L\beta_0^2}{2}\int_{t_0}^{+\infty} t\,\|\dot{x}(t)\|^2\,dt.$$

In view of (47) and claims (iv)-(v), statement (iii) follows. ☐

3.2.3 Convergence of the trajectories

We now turn to showing weak convergence of the trajectories to a minimizer.

Theorem 3.8. *Suppose that the assumptions of Lemma 3.5 hold. Then $x(t)$ converges weakly to a minimizer of f.*

Proof. As in the explicit case, we will invoke Opial's Lemma A.1. Recall that the trajectory $x(t)$ is bounded by Lemma 3.5. Thus, for any sequence $(x(t_n))_{n\in\mathbb{N}}$ which converges weakly to, say, $\bar{x}$, Theorem 3.7(i) entails that

$$f(\bar{x}) \leq \liminf_{n\to+\infty} f(x(t_n)) = \lim_{t\to+\infty} f(x(t)) = f^\star,$$

i.e. , each weak accumulation point of $\bar{x}$ belongs to S. To get weak convergence of the trajectory, it remains to show that $\lim_{t\to+\infty}\|x(t)-x^\star\|$ exists.

Let $h : t \in [t_0, +\infty[\ \mapsto \frac{1}{2}\|x(t)-x^\star\|^2$, and thus,

$$\dot{h}(t) = \langle \dot{x}(t),\, x(t)-x^\star\rangle \quad\text{and}\quad \ddot{h}(t) = \langle \ddot{x}(t),\, x(t)-x^\star\rangle + \|\dot{x}(t)\|^2\,.$$

From (ISIHD-Pert), we obtain

$$\begin{aligned}\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) &= \left\langle \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t),\, x(t)-x^\star\right\rangle + \|\dot{x}(t)\|^2\\ &= -\left\langle \nabla f(x(t)+\beta(t)\dot{x}(t)) + e(t),\, x(t)-x^\star\right\rangle + \|\dot{x}(t)\|^2\\ &= -\left\langle \nabla f(x(t)+\beta(t)\dot{x}(t)),\, x(t)+\beta(t)\dot{x}(t)-x^\star\right\rangle - \langle e(t),\, x(t)-x^\star\rangle + \|\dot{x}(t)\|^2\\ &\quad+\beta(t)\left\langle \nabla f(x(t)+\beta(t)\dot{x}(t)),\, \dot{x}(t)\right\rangle.\end{aligned}$$

Convexity of f implies,

$$-\left\langle \nabla f(x(t)+\beta(t)\dot{x}(t)),\, x(t)+\beta(t)\dot{x}(t)-x^\star\right\rangle \leq f^\star - f(x(t)+\beta(t)\dot{x}(t)) \leq 0,$$

and thus,

$$\begin{aligned}\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) &\leq \|x(t)-x^\star\|\,\|e(t)\| + \|\dot{x}(t)\|^2 + \beta_0\,\|\nabla f(x(t)+\beta(t)\dot{x}(t))\|\,\|\dot{x}(t)\|\\ &\leq C\,\|e(t)\| + \|\dot{x}(t)\|^2 + \beta_0\,\|\nabla f(x(t)+\beta(t)\dot{x}(t))\|\,\|\dot{x}(t)\|\end{aligned}$$

where $C = \sup_{t\geq t_0}\|x(t)-x^\star\| < +\infty$ thanks to Lemma 3.5, and we denoted $\beta_0 = 1+|\beta|/t_0$. Multiplying both sides by t, we arrive at

$$t\ddot{h}(t) + \alpha\dot{h}(t) \leq Ct\,\|e(t)\| + t\,\|\dot{x}(t)\|^2 + \frac{\beta_0}{\sqrt{t_0}}(t\,\|\nabla f(x(t)+\beta(t)\dot{x}(t))\|)(\sqrt{t}\,\|\dot{x}(t)\|).$$

The right-hand side of this inequality belongs to $L^1(t_0, +\infty; \mathbb{R})$ by assumption on the error, and using the Cauchy-Schwarz inequality and Theorem 3.7(iv)-(v) for the last two terms. It then follows from Lemma A.6 that $\lim_{t\to+\infty}\|x(t)-x^\star\|$ exists. We have now shown that all conditions of Lemma A.1 are satisfied, hence the weak convergence of the trajectories. ☐

3.3 Discussion

We now discuss the main differences between the two systems in terms of their stability to perturbations and the corresponding assumptions. Recall from Remark 3.6, the integrability assumption $m(\cdot)e(\cdot) \in L^1(t_0, +\infty; \mathcal{H})$ required to ensure stability for system (ISIHD-PERT) is equivalent to ensuring that the second-order moment of the error $e(\cdot)$ is finite. One may wonder whether this is more stringent than the integrability assumptions for the explicit Hessian system (ISEHD-PERT) involving the control of the first-order moments of the error and its derivative (see Section 3.1). The answer is clearly affirmative in the scalar case with a simple integration by parts argument. Indeed, supposing without loss of generaility that $e(t)$ is non-increasing, one has

$$\int_{t_0}^{+\infty} t|\dot{e}(t)|dt = \left|\int_{t_0}^{+\infty} t\dot{e}(t)dt\right| = \left|-t_0 e(t_0) - \int_{t_0}^{+\infty} e(t)dt\right| \leq t_0\left|e(t_0)\right| + t_0^{-2}\int_{t_0}^{+\infty} t^2|e(t)|dt.$$

Another intuitive way to understand this is to look at what happens if the system is discretized with finite differences. In this case, the integrability assumptions on the errors for system (ISEHD-PERT) boil down to controlling only the first-order moment of the (discretized) error. We conclude this discussion by noting that Lipschitz continuity of the gradient is not needed for the estimates and convergence analysis of (ISEHD-PERT) while it is used extensively to analyze (ISIHD-PERT). This is a distinctive avantage of (ISEHD-PERT) compared to (ISIHD-PERT). This will be even more notable when extending to the non-smooth case; see Section 5.

4 Smooth Strongly Convex Case

4.1 Explicit Hessian Damping

In this section we consider the explicit Hessian system under the assumption of strong convexity of f.

Definition 4.1. *A function $f : \mathcal{H} \to \mathbb{R}$ is μ-strongly convex ($\mu > 0$) if $f - \frac{\mu}{2}\|\cdot\|^2$ is convex.*

Following Polyak's heavy ball system [30], consider the second-order perturbed system

$$\ddot{x}(t) + 2\sqrt{\mu}\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \beta\dot{e}(t) + \nabla f(x(t)) + e(t) = 0, \tag{48}$$

which has a fixed positive damping coefficient that is adjusted to the modulus of strong convexity of f. To study (48), we define the function $\mathcal{E} : [t_0, +\infty[\to \mathbb{R}_+$

$$t \mapsto \mathcal{E}(t) := f(x(t)) - \inf_{\mathcal{H}} f + \frac{1}{2}\|v(t)\|^2, \tag{49}$$

where

$$v(t) = \sqrt{\mu}(x(t) - x^\star) + \dot{x}(t) + \beta\nabla f(x(t)). \tag{50}$$

Theorem 4.2. *Suppose that $f : \mathcal{H} \to \mathbb{R}$ is μ-strongly convex for some $\mu > 0$, let $x^\star$ be the unique minimizer of f. Let $x(\cdot) : [t_0, +\infty[\to \mathcal{H}$ be a solution trajectory of* (48)*. Suppose that*

a) $0 \leq \beta \leq \dfrac{1}{2\sqrt{\mu}}$.

b) $\displaystyle\int_{t_0}^{+\infty} \|e(t)\|dt < +\infty$ *and* $\displaystyle\int_{t_0}^{+\infty} \|\dot{e}(t)\|dt < +\infty$.

(i) *Minimizing properties: for all $t \geq t_0$*

$$\mathcal{E}(t) \leq \mathcal{E}(t_0)e^{-\frac{\sqrt{\mu}}{2}(t-t_0)} + Me^{-\frac{\sqrt{\mu}}{2}t}\int_{t_0}^{t} e^{\frac{\sqrt{\mu}}{2}\tau}\|e(\tau) + \beta\dot{e}(\tau)\|d\tau,$$

where $M := \sqrt{2\mathcal{E}(t_0)} + \int_{t_0}^{+\infty} \|e(\tau) + \beta\dot{e}(\tau)\| d\tau$. *As a consequence,*

$$\lim_{t\to+\infty} \mathcal{E}(t) = 0; \lim_{t\to+\infty} f(x(t)) = \inf_{\mathcal{H}} f$$
$$\lim_{t\to+\infty} \|x(t) - x^\star\| = \lim_{t\to+\infty} \|\nabla f(x(t))\| = \lim_{t\to+\infty} \|\dot{x}(t)\| = 0.$$

(ii) *Convergence rates: suppose moreover that for some* $p > 0$, $\|e(t) + \beta\dot{e}(t)\| = \mathcal{O}\left(\frac{1}{t^p}\right)$, *as* $t \to +\infty$. *Then* $\mathcal{E}(t) = \mathcal{O}\left(\frac{1}{t^p}\right)$, *i.e.* $\mathcal{E}(t)$ *inherits the decay rate of the error terms. As a consequence, as* $t \to +\infty$

$$f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^p}\right);$$
$$\|x(t) - x^\star\|^2 = \mathcal{O}\left(\frac{1}{t^p}\right); \ \|\dot{x}(t)\|^2 = \mathcal{O}\left(\frac{1}{t^p}\right).$$

In addition, when $\beta > 0$

$$e^{-\sqrt{\mu}t} \int_{t_0}^{t} e^{\sqrt{\mu}s} \|\nabla f(x(s))\|^2 ds = \mathcal{O}\left(\frac{1}{t^p}\right).$$

Proof. Denote $f^\star = f(x^\star) = \inf_{\mathcal{H}} f$. Define $g(t) := e(t) + \beta\dot{e}(t)$, so that the constitutive equation is written in the compact form

$$\ddot{x}(t) + 2\sqrt{\mu}\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) + g(t) = 0. \tag{51}$$

Derivation of $\mathcal{E}(\cdot)$ gives

$$\begin{aligned}\dot{\mathcal{E}}(t) &= \langle \nabla f(x(t)), \dot{x}(t)\rangle + \langle v(t), \dot{v}(t)\rangle \\ &= \langle \nabla f(x(t)), \dot{x}(t)\rangle + \left\langle v(t), \sqrt{\mu}\dot{x}(t) + \ddot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t)\right\rangle.\end{aligned}$$

Using the definition of $v(t)$ and (51), we get

$$\dot{\mathcal{E}}(t) = \langle \nabla f(x(t)), \dot{x}(t)\rangle + \langle \sqrt{\mu}(x(t) - x^\star) + \dot{x}(t) + \beta\nabla f(x(t)), -\sqrt{\mu}\dot{x}(t) - \nabla f(x(t))\rangle - \langle v(t), g(t)\rangle.$$

After developing and simplifying, we obtain

$$\begin{aligned}&\dot{\mathcal{E}}(t) + \sqrt{\mu}\langle \nabla f(x(t)), x(t) - x^\star\rangle + \mu\langle x(t) - x^\star, \dot{x}(t)\rangle + \sqrt{\mu}\|\dot{x}(t)\|^2 \\ &\quad + \beta\sqrt{\mu}\langle \nabla f(x(t)), \dot{x}(t)\rangle + \beta\|\nabla f(x(t))\|^2 = -\langle v(t), g(t)\rangle.\end{aligned}$$

According to strong convexity of f, we have

$$\langle \nabla f(x(t)), x(t) - x^\star\rangle \geq f(x(t)) - f^\star + \frac{\mu}{2}\|x(t) - x^\star\|^2.$$

Thus, by combining the last two relations, and by the Cauchy-Schwarz inequality, we obtain

$$\dot{\mathcal{E}}(t) + \sqrt{\mu}A(t) \leq \|v(t)\|\|g(t)\|,$$

where

$$A(t) := f(x(t)) - f^\star + \frac{\mu}{2}\|x(t) - x^\star\|^2 + \sqrt{\mu}\langle x(t) - x^\star, \dot{x}(t)\rangle + \|\dot{x}(t)\|^2 + \beta\langle \nabla f(x(t)), \dot{x}(t)\rangle + \frac{\beta}{\sqrt{\mu}}\|\nabla f(x(t))\|^2.$$

Let us make appear $\mathcal{E}(t)$ in $A(t)$,

$$\begin{aligned}A(t) = \mathcal{E}(t) - \frac{1}{2}\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 - \sqrt{\mu}\langle x(t) - x^\star, \dot{x}(t) + \beta\nabla f(x(t))\rangle + \sqrt{\mu}\langle x(t) - x^\star, \dot{x}(t)\rangle \\ + \|\dot{x}(t)\|^2 + \beta\langle \nabla f(x(t)), \dot{x}(t)\rangle + \frac{\beta}{\sqrt{\mu}}\|\nabla f(x(t))\|^2.\end{aligned}$$

After developing and simplifying, we obtain

$$\dot{\mathcal{E}}(t)+\sqrt{\mu}\left(\mathcal{E}(t)+\frac{1}{2}\|\dot{x}(t)\|^2+\left(\frac{\beta}{\sqrt{\mu}}-\frac{\beta^2}{2}\right)\|\nabla f(x(t))\|^2-\beta\sqrt{\mu}\left\langle x(t)-x^\star,\, \nabla f(x(t))\right\rangle\right)\le\|v(t)\|\|g(t)\|.$$

Since $0\le\beta\le\frac{1}{\sqrt{\mu}}$, it holds that $\frac{\beta}{\sqrt{\mu}}-\frac{\beta^2}{2}\ge\frac{\beta}{2\sqrt{\mu}}$. Hence

$$\dot{\mathcal{E}}(t)+\sqrt{\mu}\left(\mathcal{E}(t)+\frac{1}{2}\|\dot{x}(t)\|^2+\frac{\beta}{2\sqrt{\mu}}\|\nabla f(x(t))\|^2-\beta\sqrt{\mu}\left\langle x(t)-x^\star,\, \nabla f(x(t))\right\rangle\right)\le\|v(t)\|\|g(t)\|.$$

Let us use again the strong convexity of f to write

$$\mathcal{E}(t)=\frac{1}{2}\mathcal{E}(t)+\frac{1}{2}\mathcal{E}(t)\ge\frac{1}{2}\mathcal{E}(t)+\frac{1}{2}\left(f(x(t))-f^\star\right)\ge\frac{1}{2}\mathcal{E}(t)+\frac{\mu}{4}\|x(t)-x^\star\|^2.$$

By combining the two inequalities above, we obtain

$$\dot{\mathcal{E}}(t)+\frac{\sqrt{\mu}}{2}\mathcal{E}(t)+\frac{\sqrt{\mu}}{2}\|\dot{x}(t)\|^2+\sqrt{\mu}B(t)\le\|v(t)\|\|g(t)\|,$$

where $B(t)=\frac{\mu}{4}\|x(t)-x^\star\|^2+\frac{\beta}{2\sqrt{\mu}}\|\nabla f(x(t))\|^2-\beta\sqrt{\mu}\|x(t)-x^\star\|\|\nabla f(x(t))\|$. Set $X=\|x-x^\star\|$, $Y=\|\nabla f(x)\|$. Elementary algebraic computation gives that, under the condition $0\le\beta\le\frac{1}{2\sqrt{\mu}}$

$$\frac{\mu}{4}X^2+\frac{\beta}{2\sqrt{\mu}}Y^2-\beta\sqrt{\mu}XY\ge 0.$$

Hence for $0\le\beta\le\frac{1}{2\sqrt{\mu}}$

$$\dot{\mathcal{E}}(t)+\frac{\sqrt{\mu}}{2}\mathcal{E}(t)+\frac{\sqrt{\mu}}{2}\|\dot{x}(t)\|^2\le\|v(t)\|\|g(t)\|. \tag{52}$$

(i) From (52), we first deduce that

$$\dot{\mathcal{E}}(t)\le\|v(t)\|\|g(t)\|,$$

which by integration gives

$$\mathcal{E}(t)\le\mathcal{E}(t_0)+\int_{t_0}^{t}\|v(\tau)\|\|g(\tau)\|d\tau.$$

By definition of $\mathcal{E}(t)$, we have $\mathcal{E}(t)\ge\frac{1}{2}\|v(t)\|^2$, which gives

$$\frac{1}{2}\|v(t)\|^2\le\mathcal{E}(t_0)+\int_{t_0}^{t}\|v(\tau)\|\|g(\tau)\|d\tau.$$

According to Lemma A.3, we obtain

$$\|v(t)\|\le\sqrt{2\mathcal{E}(t_0)}+\int_{t_0}^{t}\|g(\tau)\|d\tau.$$

Set $M:=\sqrt{2\mathcal{E}(t_0)}+\int_{t_0}^{+\infty}\|g(\tau)\|d\tau$. By assumption, $\displaystyle\int_{t_0}^{+\infty}\|g(\tau)\|d\tau<+\infty$, and thus $\sup_{t\ge t_0}\|v(t)\|\le M<+\infty$. Returning to (52) we deduce that

$$\dot{\mathcal{E}}(t)+\frac{\sqrt{\mu}}{2}\mathcal{E}(t)+\frac{\sqrt{\mu}}{2}\|\dot{x}(t)\|^2\le M\|g(t)\|. \tag{53}$$

Therefore

$$\dot{\mathcal{E}}(t)+\frac{\sqrt{\mu}}{2}\mathcal{E}(t)\le M\|g(t)\|. \tag{54}$$

By integrating the differential inequality above, we obtain

$$\mathcal{E}(t)\le\mathcal{E}(t_0)e^{-\frac{\sqrt{\mu}}{2}(t-t_0)}+Me^{-\frac{\sqrt{\mu}}{2}t}\int_{t_0}^{t}e^{\frac{\sqrt{\mu}}{2}\tau}\|g(\tau)\|d\tau. \tag{55}$$

We now use Lemma A.5, which is the continuous version of Kronecker's Theorem for series, with $f(t) = \|g(t)\|$ and $\varphi(t) = e^{\frac{\sqrt{\mu}}{2}t}$. By assumption we have $\int_{t_0}^{+\infty} \|g(\tau)\|d\tau < +\infty$. We deduce that

$$\lim_{t\to+\infty} \frac{1}{e^{\frac{\sqrt{\mu}}{2}t}} \int_{t_0}^{t} e^{\frac{\sqrt{\mu}}{2}\tau} \|g(\tau)\|d\tau = 0.$$

Therefore, from (55) we obtain

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

By definition of $\mathcal{E}(t)$ this implies

$$\lim_{t\to+\infty} f(x(t)) - \inf_{\mathcal{H}} f = 0 \tag{56}$$

$$\lim_{t\to+\infty} \|\sqrt{\mu}(x(t) - x^\star) + \dot{x}(t) + \beta\nabla f(x(t))\| = 0. \tag{57}$$

Acoording to (56) and the strong convexity of f we deduce that

$$\lim_{t\to+\infty} \|x(t) - x^\star\| = 0$$

By continuity of ∇f, and since $\nabla f(x^\star) = 0$, we deduce that

$$\lim_{t\to+\infty} \|\nabla f(x(t))\| = 0.$$

Combining the above results with (57), we deduce that

$$\lim_{t\to+\infty} \|\dot{x}(t)\| = 0.$$

(ii) Let us make precise the argument developed above, and assume that, as $t \to +\infty$

$$\|g(t)\| = \mathcal{O}\left(\frac{1}{t^p}\right),$$

where $p > 0$. Then, from (55) we get

$$\begin{aligned}
\mathcal{E}(t) &\leq \mathcal{E}(t_0)e^{-\frac{\sqrt{\mu}}{2}(t-t_0)} + Me^{-\frac{\sqrt{\mu}}{2}t}\Big(\int_{t_0}^{\frac{t}{2}} e^{\frac{\sqrt{\mu}}{2}\tau}\|g(\tau)\|d\tau + \int_{\frac{t}{2}}^{t} e^{\frac{\sqrt{\mu}}{2}\tau}\|g(\tau)\|d\tau\Big)\\
&\leq \mathcal{E}(t_0)e^{-\frac{\sqrt{\mu}}{2}(t-t_0)} + Me^{-\frac{\sqrt{\mu}}{2}t}\Big(C_1 e^{\frac{\sqrt{\mu}t}{4}} + \int_{\frac{t}{2}}^{t} e^{\frac{\sqrt{\mu}}{2}\tau}\frac{C_2}{\tau^p}d\tau\Big)\\
&\leq \mathcal{E}(t_0)e^{-\frac{\sqrt{\mu}}{2}(t-t_0)} + Me^{-\frac{\sqrt{\mu}}{2}t}\Big(C_1 e^{\frac{\sqrt{\mu}t}{4}} + \frac{C_2}{t^p}e^{\frac{\sqrt{\mu}}{2}t}\Big)\\
&\leq \mathcal{E}(t_0)e^{-\frac{\sqrt{\mu}}{2}(t-t_0)} + M\Big(C_1 e^{-\frac{\sqrt{\mu}t}{4}} + \frac{C_2}{t^p}\Big)\\
&= \mathcal{O}\left(\frac{1}{t^p}\right).
\end{aligned}$$

By definition of $\mathcal{E}(t)$ and strong convexity of f, we infer

$$\frac{\mu}{2}\|x(t) - x^\star\|^2 \leq f(x(t)) - \min f(\mathcal{H}) = \mathcal{O}\left(\frac{1}{t^p}\right) \text{ and } \|\sqrt{\mu}(x(t)-x^\star)+\dot{x}(t)+\beta\nabla f(x(t))\|^2 = \mathcal{O}\left(\frac{1}{t^p}\right).$$

Developing the left-hand side of the last expression, we obtain

$$\mu\|x(t) - x^\star\|^2 + \|\dot{x}(t)\|^2 + \beta^2\|\nabla f(x(t))\|^2 + 2\beta\sqrt{\mu}\,\langle x(t) - x^\star,\, \nabla f(x(t))\rangle$$
$$+ \langle \dot{x}(t),\, 2\beta\nabla f(x(t)) + 2\sqrt{\mu}(x(t) - x^\star)\rangle \leq \frac{C}{t^p}.$$

By convexity of f, we have $\langle x(t) - x^\star,\, \nabla f(x(t))\rangle \geq f(x(t)) - f^\star$. Moreover,

$$\langle \dot{x}(t),\, 2\beta\nabla f(x(t)) + 2\sqrt{\mu}(x(t) - x^\star)\rangle = \frac{d}{dt}\left(2\beta(f(x(t)) - f^\star) + \sqrt{\mu}\|x(t) - x^\star\|^2\right).$$

Combining the above results, we obtain

$$\sqrt{\mu}\left(2\beta(f(x(t))-f^\star)+\sqrt{\mu}\|x(t)-x^\star\|^2\right)+\beta^2\|\nabla f(x(t))\|^2+\frac{d}{dt}\left(2\beta(f(x(t))-f^\star)+\sqrt{\mu}\|x(t)-x^\star\|^2\right)\leq\frac{C}{t^p}.$$

Set $Z(t):=2\beta(f(x(t))-f^\star)+\sqrt{\mu}\|x(t)-x^\star\|^2$. We have

$$\frac{d}{dt}Z(t)+\sqrt{\mu}Z(t)+\beta^2\|\nabla f(x(t))\|^2\leq\frac{C}{t^p}.$$

By integrating this differential inequality, elementary computation gives

$$e^{-\sqrt{\mu}t}\int_{t_0}^t e^{\sqrt{\mu}s}\|\nabla f(x(s))\|^2ds\leq\frac{C}{t^p}.$$

This completes the proof.

□

4.2 Implicit Hessian Damping

We now turn to the implicit Hessian system, and take in the Polyak heavy ball system a fixed positive damping coefficient which is adjusted to the modulus of strong convexity of f. This gives the system

$$\ddot{x}(t)+2\sqrt{\mu}\dot{x}(t)+\nabla f\left(x(t)+\beta\dot{x}(t)\right)+e(t)=0. \tag{58}$$

To analyze (58), we define the function $\mathcal{E}:[t_0,+\infty[\to\mathbb{R}_+$

$$t\mapsto\mathcal{E}(t):=f\left(x(t)+\beta\dot{x}(t)\right))-\inf_{\mathcal{H}}f+\frac{1}{2}\|\sqrt{\mu}(x(t)-x^\star)+\dot{x}(t)\|^2. \tag{59}$$

Theorem 4.3. *Suppose that $f:\mathcal{H}\to\mathbb{R}$ is μ-strongly convex for some $\mu>0$, and let $x^\star$ be the unique minimizer of f. Let $x(\cdot):[t_0,+\infty[\to\mathcal{H}$ be a solution trajectory of (58). Suppose that*

a) $0\leq\beta\leq\dfrac{1}{2\sqrt{\mu}}$.

b) $\displaystyle\int_{t_0}^{+\infty}\|e(t)\|dt<+\infty$.

(i) *Minimizing properties: there exists a positive constant M such that for all $t\geq t_0$*

$$\mathcal{E}(t)\leq\mathcal{E}(t_0)e^{-\frac{\sqrt{\mu}}{2}(t-t_0)}+Me^{-\frac{\sqrt{\mu}}{2}t}\int_{t_0}^t e^{\frac{\sqrt{\mu}}{2}\tau}\|e(\tau)\|d\tau.$$

More precisely,

$$M:=\sqrt{\frac{\mathcal{E}(t_0)}{c}}+\frac{1}{2c}\int_{t_0}^{+\infty}\|e(\tau)\|d\tau\ \ \textit{with}\ \ c=\frac{\min\{\mu,1\}}{4\max\{\beta^2L^2,1\}}$$

and L is the Lipschitz constant of ∇f. Consequently,

$$\lim_{t\to+\infty}\mathcal{E}(t)=0;\ \lim_{t\to+\infty}f(x(t))=\inf_{\mathcal{H}}f$$
$$\lim_{t\to+\infty}\|x(t)-x^\star\|=\lim_{t\to+\infty}\|\nabla f(x(t))\|=\lim_{t\to+\infty}\|\dot{x}(t)\|=0.$$

(ii) *Convergence rates: suppose moreover that for some $p>0$, $\|e(t)\|=\mathcal{O}\left(\frac{1}{t^p}\right)$, as $t\to+\infty$. Then $\mathcal{E}(t)=\mathcal{O}\left(\frac{1}{t^p}\right)$, i.e. $\mathcal{E}(t)$ inherits the decay rate of the error terms. In turn, as $t\to+\infty$*

$$f\left(x(t)\right)-\inf_{\mathcal{H}}f=\mathcal{O}\left(\frac{1}{t^p}\right);$$
$$\|x(t)-x^\star\|^2=\mathcal{O}\left(\frac{1}{t^p}\right);\ \|\dot{x}(t)\|^2=\mathcal{O}\left(\frac{1}{t^p}\right);\ \|\nabla f(x(t))\|^2=\mathcal{O}\left(\frac{1}{t^p}\right).$$

Proof. Let us define

$$v(t) = \sqrt{\mu}(x(t) - x^\star) + \dot{x}(t). \tag{60}$$

and thus, $\mathcal{E}$ equivalently reads

$$\mathcal{E}(t) = f\left(x(t) + \beta\dot{x}(t)\right) - \inf_{\mathcal{H}} f + \frac{1}{2}\|v(t)\|^2. \tag{61}$$

Taking the derivative in time of $\mathcal{E}(\cdot)$ gives

$$\begin{aligned}
\dot{\mathcal{E}}(t) &= \langle \nabla f\left(x(t) + \beta\dot{x}(t)\right), \dot{x}(t) + \beta\ddot{x}(t)\rangle + \langle v(t), \dot{v}(t)\rangle \\
&= \langle \nabla f\left(x(t) + \beta\dot{x}(t)\right), \dot{x}(t) + \beta\ddot{x}(t)\rangle + \langle \sqrt{\mu}(x(t) - x^\star) + \dot{x}(t), \sqrt{\mu}\dot{x}(t) + \ddot{x}(t)\rangle .
\end{aligned}$$

Using the constitutive equation (51), we get

$$\begin{aligned}
\dot{\mathcal{E}}(t) = \langle \nabla f\left(x(t) + \beta\dot{x}(t)\right), (1 - 2\beta\sqrt{\mu})\dot{x}(t) - \beta\nabla f\left(x(t) + \beta\dot{x}(t)\right) - \beta e(t)\rangle \\
+ \langle \sqrt{\mu}(x(t) - x^\star) + \dot{x}(t), -\sqrt{\mu}\dot{x}(t) - \nabla f\left(x(t) + \beta\dot{x}(t)\right) - e(t)\rangle .
\end{aligned}$$

After developing and simplifying, we obtain

$$\begin{aligned}
\dot{\mathcal{E}}(t) + 2\beta\sqrt{\mu}\langle \nabla f\left(x(t) + \beta\dot{x}(t)\right), \dot{x}(t)\rangle + \sqrt{\mu}\langle \nabla f\left(x(t) + \beta\dot{x}(t)\right), x(t) - x^\star\rangle + \beta\|\nabla f\left(x(t) + \beta\dot{x}(t)\right)\|^2 \\
+ \sqrt{\mu}\|\dot{x}(t)\|^2 + \mu\langle x(t) - x^\star, \dot{x}(t)\rangle = -\langle \sqrt{\mu}(x(t) - x^\star) + \dot{x}(t) + \beta\nabla f\left(x(t) + \beta\dot{x}(t)\right), e(t)\rangle .
\end{aligned}$$

In view of strong convexity of f, we have

$$\begin{aligned}
\langle \nabla f\left(x(t) + \beta\dot{x}(t)\right), x(t) - x^\star\rangle &= \langle \nabla f\left(x(t) + \beta\dot{x}(t)\right), x(t) + \beta\dot{x}(t) - x^\star\rangle - \langle \nabla f\left(x(t) + \beta\dot{x}(t)\right), \beta\dot{x}(t)\rangle \\
&\geq f\left(x(t) + \beta\dot{x}(t)\right) - f^\star + \frac{\mu}{2}\|x(t) - x^\star + \beta\dot{x}(t)\|^2 - \langle \nabla f\left(x(t) + \beta\dot{x}(t)\right), \beta\dot{x}(t)\rangle .
\end{aligned}$$

Thus, by combining the last two relations, we obtain

$$\begin{aligned}
\dot{\mathcal{E}}(t) + \beta\sqrt{\mu}\langle \nabla f\left(x(t) + \beta\dot{x}(t)\right), \dot{x}(t)\rangle + \sqrt{\mu}\left(f\left(x(t) + \beta\dot{x}(t)\right) - f^\star + \frac{\mu}{2}\|x(t) - x^\star + \beta\dot{x}(t)\|^2\right) \\
+ \beta\|\nabla f\left(x(t) + \beta\dot{x}(t)\right)\|^2 + \sqrt{\mu}\|\dot{x}(t)\|^2 + \mu\langle x(t) - x^\star, \dot{x}(t)\rangle \leq \|w(t)\|\|e(t)\|, \quad (62)
\end{aligned}$$

where we have used Cauchy-Schwarz inequality, and we set

$$w(t) := \sqrt{\mu}(x(t) - x^\star) + \dot{x}(t) + \beta\nabla f\left(x(t) + \beta\dot{x}(t)\right).$$

Let us make $\mathcal{E}(t)$ appear on the left-hand side of (62). We get

$$\dot{\mathcal{E}}(t) + \sqrt{\mu}\mathcal{E}(t) + B(t) \leq \|w(t)\|\|e(t)\|$$

where

$$B(t) := \beta\|\nabla f\left(x(t) + \beta\dot{x}(t)\right)\|^2 + \frac{\sqrt{\mu}}{2}(\beta^2\mu + 1)\|\dot{x}(t)\|^2 + \beta\sqrt{\mu}\langle \nabla f\left(x(t) + \beta\dot{x}(t)\right), \dot{x}(t)\rangle + \beta\mu\sqrt{\mu}\langle x(t) - x^\star, \dot{x}(t)\rangle .$$

Let us use again the strong convexity of f to write

$$\mathcal{E}(t) = \frac{1}{2}\mathcal{E}(t) + \frac{1}{2}\mathcal{E}(t) \geq \frac{1}{2}\mathcal{E}(t) + \frac{1}{2}\left(f(x(t) + \beta\dot{x}(t)) - f^\star\right) \geq \frac{1}{2}\mathcal{E}(t) + \frac{\mu}{4}\|x(t) - x^\star + \beta\dot{x}(t)\|^2.$$

By combining the inequalities above, we obtain

$$\dot{\mathcal{E}}(t) + \frac{\sqrt{\mu}}{2}\mathcal{E}(t) + C(t) \leq \|w(t)\|\|e(t)\|,$$

where

$$\begin{aligned}
C(t) := \beta\|\nabla f(y(t)\|^2 + \beta\sqrt{\mu}\langle \nabla f(y(t)), \dot{x}(t)\rangle + \frac{\sqrt{\mu}}{2}(\beta^2\mu + 1)\|\dot{x}(t)\|^2 + \beta\mu\sqrt{\mu}\langle x(t) - x^\star, \dot{x}(t)\rangle \\
+ \frac{\mu\sqrt{\mu}}{4}\|x(t) - x^\star + \beta\dot{x}(t)\|^2,
\end{aligned}$$

and we set $y(t) := x(t) + \beta\dot{x}(t)$. Let us show that, for an adequate choice of the parameters, $C(t)$ is non-negative. Let us reformulate $C(t)$ as follows: Young's inequality gives the following minorization for the two first terms of $C(t)$

$$\beta\|\nabla f(y(t))\|^2 + \beta\sqrt{\mu}\langle\nabla f(y(t)), \dot{x}(t)\rangle \geq -\frac{1}{4}\beta\mu\|\dot{x}(t)\|^2.$$

By using this inequality in $C(t)$, and after simplification, we arrive at

$$\begin{aligned}
C(t) &\geq \left(\frac{\sqrt{\mu}}{2}(\beta^2\mu + 1) - \frac{1}{4}\beta\mu\right)\|\dot{x}(t)\|^2 + \beta\mu\sqrt{\mu}\left\langle x(t) - x^\star,\, \dot{x}(t)\right\rangle + \frac{\mu\sqrt{\mu}}{4}\|x(t) - x^\star + \beta\dot{x}(t)\|^2 \\
&= \frac{\mu\sqrt{\mu}}{4}\|x(t) - x^\star + \beta\dot{x}(t)\|^2 + \left(\frac{\sqrt{\mu}}{2}(\beta^2\mu + 1) - \frac{1}{4}\beta\mu - \beta^2\mu\sqrt{\mu}\right)\|\dot{x}(t)\|^2 + \beta\mu\sqrt{\mu}\left\langle x(t) - x^\star + \beta\dot{x}(t),\, \dot{x}(t)\right\rangle \\
&= \frac{\mu\sqrt{\mu}}{4}\|x(t) - x^\star + \beta\dot{x}(t)\|^2 + \sqrt{\mu}\left(-\frac{\beta^2\mu}{2} - \frac{1}{4}\beta\sqrt{\mu} + \frac{1}{2}\right)\|\dot{x}(t)\|^2 + \beta\mu\sqrt{\mu}\left\langle x(t) - x^\star + \beta\dot{x}(t),\, \dot{x}(t)\right\rangle.
\end{aligned}$$

Elementary algebra gives that $-\frac{\beta^2\mu}{2} - \frac{1}{4}\beta\sqrt{\mu} + \frac{1}{2} \geq 0$ if and only if $\beta\sqrt{\mu} \leq \frac{\sqrt{17}-1}{4}$. According to the classical rule for the sign of a quadratic function of a real variable, we get that $C(t) \geq 0$ under the condition

$$(\beta\mu\sqrt{\mu})^2 \leq \mu^2\left(-\frac{\beta^2\mu}{2} - \frac{1}{4}\beta\sqrt{\mu} + \frac{1}{2}\right).$$

Setting $Z = \beta\sqrt{\mu}$, the latter inequality is equivalent to ensuring

$$\frac{3}{2}Z^2 + \frac{1}{4}Z - \frac{1}{2} \leq 0.$$

which is satisfied for $0 \leq Z \leq \frac{1}{2}$, implying $\beta \leq \frac{1}{2\sqrt{\mu}}$. Since $\frac{1}{2} < \frac{\sqrt{17}-1}{4}$, we get as a final condition

$$\beta \leq \frac{1}{2\sqrt{\mu}}.$$

Thus under this condition we get

$$\dot{\mathcal{E}}(t) + \frac{\sqrt{\mu}}{2}\mathcal{E}(t) \leq \|w(t)\|\|e(t)\|. \tag{63}$$

From (63), we first deduce that

$$\dot{\mathcal{E}}(t) \leq \|w(t)\|\|e(t)\|,$$

which, after integration, gives

$$\mathcal{E}(t) \leq \mathcal{E}(t_0) + \int_{t_0}^{t}\|w(\tau)\|\|e(\tau)\|d\tau.$$

By definition of w we have

$$\begin{aligned}
\|w(t)\| &\leq \|v(t)\| + \beta\|\nabla f\left(x(t) + \beta\dot{x}(t)\right) - \nabla f(x^\star)\| \\
&\leq \|v(t)\| + \beta L\|x(t) - x^\star + \beta\dot{x}(t)\|,
\end{aligned}$$

where L is the Lipschitz constant of ∇f. On the other hand, strong convexity of f entails

$$\mathcal{E}(t) \geq \frac{\mu}{2}\|x(t) - x^\star + \beta\dot{x}(t)\|^2 + \frac{1}{2}\|v(t)\|^2.$$

Hence, there exists a positive constant c such that[1]

$$\mathcal{E}(t) \geq c\|w(t)\|^2.$$

This in turn gives

$$c\|w(t)\|^2 \leq \mathcal{E}(t_0) + \int_{t_0}^{t}\|w(\tau)\|\|e(\tau)\|d\tau.$$

[1] One can take $c = \frac{\min\{\mu,1\}}{4\max\{\beta^2L^2,1\}}$.

According to Lemma A.3, and $\int_{t_0}^{+\infty} \|e(\tau)\| d\tau < +\infty$, we deduce that

$$\sup_{t \geq t_0} \|w(t)\| \leq M := \sqrt{\frac{\mathcal{E}(t_0)}{c}} + \frac{1}{2c} \int_{t_0}^{+\infty} \|e(\tau)\| d\tau < +\infty.$$

Returning to (63) we deduce that

$$\dot{\mathcal{E}}(t) + \frac{\sqrt{\mu}}{2} \mathcal{E}(t) \leq M \|e(t)\|. \tag{64}$$

By integrating the differential inequality above, we obtain

$$\mathcal{E}(t) \leq \mathcal{E}(t_0) e^{-\frac{\sqrt{\mu}}{2}(t-t_0)} + M e^{-\frac{\sqrt{\mu}}{2} t} \int_{t_0}^{t} e^{\frac{\sqrt{\mu}}{2}\tau} \|e(\tau)\| d\tau. \tag{65}$$

(i) We first deduce from (65) that $\mathcal{E}(t)$ tends to zero as $t \to +\infty$. This implies that

$$\lim_{t \to +\infty} f(x(t) + \beta \dot{x}(t))) = \inf_{\mathcal{H}} f, \tag{66}$$

$$\lim_{t \to +\infty} \|\sqrt{\mu}(x(t) - x^\star) + \dot{x}(t)\| = 0. \tag{67}$$

From (66) and strong convexity of f we deduce that

$$\lim_{t \to +\infty} \|(x(t) - x^\star) + \beta \dot{x}(t)\| = 0. \tag{68}$$

From (67) and (68), and $\beta \neq \frac{1}{\sqrt{\mu}}$ (a consequence of the assumption $\beta \leq \frac{1}{2\sqrt{\mu}}$), elementary algebra gives

$$\lim_{t \to +\infty} \|x(t) - x^\star\| = \lim_{t \to +\infty} \|\dot{x}(t)\| = 0.$$

In turn, continuity of f and ∇f imply

$$\lim_{t \to +\infty} \|\nabla f(x(t))\| = 0 \text{ and } \lim_{t \to +\infty} f(x(t)) = \inf_{\mathcal{H}} f.$$

(ii) Let us now assume that, as $t \to +\infty$, we have $\|e(t)\| = \mathcal{O}\left(\frac{1}{t^p}\right)$, where $p > 0$. Based on (65), a similar argument as in the implicit case (see the proof of Theorem 4.2) gives $\mathcal{E}(t) = \mathcal{O}\left(\frac{1}{t^p}\right)$. By definition of $\mathcal{E}(t)$, we infer that

$$f(x(t) + \beta \dot{x}(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^p}\right) \tag{69}$$

and

$$\|\sqrt{\mu}(x(t) - x^\star) + \dot{x}(t)\|^2 = \mathcal{O}\left(\frac{1}{t^p}\right). \tag{70}$$

From (69) and strong convexity of f we deduce that

$$\|(x(t) - x^\star) + \beta \dot{x}(t)\|^2 = \mathcal{O}\left(\frac{1}{t^p}\right). \tag{71}$$

Combining (70) and (71), and recalling that $\beta\sqrt{\mu} \neq 1$ we immediately obtain

$$\|x(t) - x^\star\|^2 \leq \frac{C}{t^p} \text{ and } \|\dot{x}(t)\|^2 = \mathcal{O}\left(\frac{1}{t^p}\right). \tag{72}$$

According to the Lipschitz continuity of ∇f, and $\nabla f(x^\star) = 0$ we deduce that

$$\|\nabla f(x(t))\|^2 \leq L^2 \|x(t) - x^\star\|^2 = \mathcal{O}\left(\frac{1}{t^p}\right).$$

Now, combining the descent lemma with (69), (71) and (72) shows that

$$\begin{aligned}
f(x(t)) - \min f(\mathcal{H}) &\leq f(x(t) + \beta \dot{x}(t)) - \inf_{\mathcal{H}} f - \beta \left\langle \nabla f(x(t) + \beta \dot{x}(t)), \dot{x}(t) \right\rangle + \frac{L\beta^2}{2} \|\dot{x}(t)\|^2 \\
&\leq f(x(t) + \beta \dot{x}(t)) - \inf_{\mathcal{H}} f + L\beta \|x(t) - x^\star + \beta \dot{x}(t)\| \|\dot{x}(t)\| + \frac{L\beta^2}{2} \|\dot{x}(t)\|^2 \\
&= \mathcal{O}\left(\frac{1}{t^p}\right),
\end{aligned}$$

which completes the proof.

$\square$

5 The Non-smooth Case

5.1 Explicit Hessian Damping

In the sequel, we will show that most properties obtained in the smooth case still hold for the global strong solution of (4) (and in particular, all properties that do not require $x(t)$ to be twice differentiable).

5.1.1 Minimizing properties

From now on, we assume that, for all $T > t_0$, $e(\cdot) \in \mathcal{W}^{1,1}(t_0, T; \mathcal{H})$, which is the Sobolev space of functions $e(\cdot) \in L^1(t_0, T; \mathcal{H})$ with distributional derivative $\dot{e}(\cdot) \in L^1(t_0, T; \mathcal{H})$. Let $(x, y) : [t_0, +\infty[\to \mathcal{H} \times \mathcal{H}$ be the global strong solution to (4) with Cauchy data $(x(t_0), y(t_0)) = (x_0, y_0) \in \operatorname{dom}(f) \times \mathcal{H}$. For $t \geq t_0$ define

$$u(t) = \int_{t_0}^{t} \left(-\beta e(s) + \left(\frac{1}{\beta} - \frac{\alpha}{s} \right) x(s) - \frac{1}{\beta} y(s) \right) ds. \tag{73}$$

Thus u is continuously differentiable, with derivative satisfying

$$\begin{aligned} \dot{u}(t) &= -\beta e(t) + \left(\frac{1}{\beta} - \frac{\alpha}{t} \right) x(t) - \frac{1}{\beta} y(t), \quad \forall t \geq t_0, &(74)\\ &= \dot{x}(t) + \beta \xi(t), \qquad \text{for almost all } t > t_0, &(75) \end{aligned}$$

where $\xi(t) \in \partial f(x(t))$, and the last equality follows from Theorem 2.4(vi)-(b). Therefore, u can be also written equivalently as

$$u(t) = x(t) - x_0 + \beta \int_{t_0}^{t} \xi(s) ds.$$

With parts (i) and (ii) of Theorem 2.4, equality (74) shows that $\dot{u}$ is absolutely continuous on any compact subinterval of $[t_0, +\infty[$, hence differentiable almost everywhere on $[t_0, +\infty[$. Therefore,

$$\ddot{u}(t) = -\beta \dot{e}(t) + \frac{\alpha}{t^2} x(t) + \left(\frac{1}{\beta} - \frac{\alpha}{t} \right) \dot{x}(t) - \frac{1}{\beta} \dot{y}(t).$$

The equality above, combined with $\dot{y}(t) = \frac{\alpha\beta}{t^2} x(t) + \dot{x}(t) + \beta(\xi(t) + e(t))$ (which is obtained by taking the difference of the two equations in (4)), yields

$$\ddot{u}(t) = -\frac{\alpha}{t} \dot{x}(t) - \xi(t) - (e(t) + \beta \dot{e}(t)), \tag{76}$$

for almost all $t > t_0$. Using (75), we obtain

$$\ddot{u}(t) = \left(\frac{1}{\beta} - \frac{\alpha}{t} \right) \dot{x}(t) - \frac{1}{\beta} \dot{u}(t) - (e(t) + \beta \dot{e}(t)) \tag{77}$$

for almost all $t > t_0$. We will need the following energy function of the system, defined for all $T \geq t \geq t_0$ (recall (74) for the definition of $\dot{u}(t)$):

$$W_T(t) := \frac{1}{2} \|\dot{u}(t)\|^2 + f(x(t)) - \int_{t}^{T} \langle \dot{u}(\tau), e(\tau) + \beta \dot{e}(\tau) \rangle d\tau, \tag{78}$$

and when the following expression is well-defined (we will prove it later)

$$W(t) := \frac{1}{2} \|\dot{u}(t)\|^2 + f(x(t)) - \int_{t}^{+\infty} \langle \dot{u}(\tau), e(\tau) + \beta \dot{e}(\tau) \rangle d\tau. \tag{79}$$

Theorem 5.1. *Let $\alpha > 0$. Suppose that $\inf_{\mathcal{H}} f > -\infty$. Suppose that $e(\cdot) \in \mathcal{W}^{1,1}(t_0, T; \mathcal{H})$ for all $T > t_0$, with $\int_{t_0}^{+\infty} \|e(t)\| < +\infty$ and $\int_{t_0}^{+\infty} \|\dot{e}(t)\| < +\infty$. Then for any global strong solution of (4), $(x, y) : [t_0, +\infty[\to \mathcal{H} \times \mathcal{H}$*

(i) *W is well-defined and non-increasing on $[t_1, +\infty[$ for some $t_1 \geq t_0$.*

(ii) $\displaystyle\int_{t_0}^{+\infty} \frac{1}{t}\|\dot{x}(t)\|^2 dt < +\infty$, $\displaystyle\int_{t_0}^{+\infty} \frac{1}{t}\|\xi(t)\|^2 dt < +\infty$.

(iii) $\lim_{t\to+\infty} W(t) = \lim_{t\to+\infty} f(x(t)) = \inf_{\mathcal{H}} f \in \mathbb{R} \cup \{-\infty\}$, $\lim_{t\to+\infty} \|\dot{x}(t) + \beta\xi(t)\| = 0$.

(iv) *As $t \to +\infty$, every sequential weak cluster point of $x(t)$ belongs to S.*

(v) *If, moreover, the solution set $S \neq \emptyset$ and* $\displaystyle\int_{t_0}^{+\infty} \log t\, \|e(t)\| < +\infty$ *and* $\displaystyle\int_{t_0}^{+\infty} \log t\, \|\dot{e}(t)\| < +\infty$, *then*

  (a) $f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\dfrac{1}{\log t}\right)$ *and* $\|\dot{u}(t)\| = \mathcal{O}\left(\dfrac{1}{\sqrt{\log t}}\right)$ *as* $t \to +\infty$.

  (b) $\displaystyle\int_{t_0}^{+\infty} \frac{1}{t}(f(x(t)) - \inf_{\mathcal{H}} f) dt < +\infty$.

Proof. Since we are interested in asymptotic analysis, we can assume $t \geq t_1 = \max(t_0, 2\alpha\beta)$.

Claim (i) According to Theorem 2.4, W_T is absolutely continuous. Taking the derivative and using the chain rule we get

$$\dot{W}_T(t) = \langle \dot{u}(t), \ddot{u}(t)\rangle + \langle \xi(t), \dot{x}(t)\rangle + \langle \dot{u}(t), e(t) + \beta\dot{e}(t)\rangle,$$

for almost every $T > t > t_0$. Now use (75) and (77) to obtain

$$\begin{aligned}
\dot{W}_T(t) &= \left\langle \dot{u}(t),\, \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)\dot{x}(t) - \frac{1}{\beta}\dot{u}(t) - (e(t) + \beta\dot{e}(t))\right\rangle + \langle \xi(t),\, \dot{x}(t)\rangle + \langle \dot{u}(t),\, e(t) + \beta\dot{e}(t)\rangle \\
&= \left\langle \dot{u}(t),\, \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)\dot{x}(t) - \frac{1}{\beta}\dot{u}(t)\right\rangle + \langle \xi(t),\, \dot{x}(t)\rangle \\
&= -\frac{1}{\beta}\|\dot{u}(t)\|^2 + \left\langle \dot{x}(t),\, \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)\dot{u}(t) + \xi(t)\right\rangle \\
&= -\frac{1}{\beta}\|\dot{u}(t)\|^2 + \left\langle \dot{x}(t),\, \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)\dot{u}(t) + \frac{1}{\beta}(\dot{u}(t) - \dot{x}(t))\right\rangle \\
&= -\frac{1}{\beta}\|\dot{u}(t)\|^2 + \left\langle \dot{x}(t),\, \left(\frac{2}{\beta} - \frac{\alpha}{t}\right)\dot{u}(t) - \frac{1}{\beta}\dot{x}(t)\right\rangle \\
&= -\frac{1}{\beta}\|\dot{u}(t)\|^2 - \frac{1}{\beta}\|\dot{x}(t)\|^2 + \left(\frac{2}{\beta} - \frac{\alpha}{t}\right)\langle \dot{x}(t),\, \dot{u}(t)\rangle \\
&\leq -\frac{\alpha}{2t}\|\dot{x}(t)\|^2 - \frac{\alpha}{2t}\|\dot{u}(t)\|^2,
\end{aligned}$$

for almost every $t \geq t_1$. So W_T is non-increasing on $[t_1, +\infty[$, because it is absolutely continuous and its derivative is non-positive therein. Therefore $W_T(t) \leq W_T(t_1)$ for all $t \in [t_1, T]$. Equivalently

$$\frac{1}{2}\|\dot{u}(t)\|^2 + f(x(t)) - \int_t^T \langle \dot{u}(\tau), e(\tau) + \beta\dot{e}(\tau)\rangle d\tau \leq \frac{1}{2}\|\dot{u}(t_1)\|^2 + f(x(t_1)) - \int_{t_1}^T \langle \dot{u}(\tau), e(\tau) + \beta\dot{e}(\tau)\rangle d\tau.$$

After simplification, and setting $C = \frac{1}{2}\|\dot{u}(t_1)\|^2 + f(x(t_1)) - \inf f(\mathcal{H})$, we obtain

$$\frac{1}{2}\|\dot{u}(t)\|^2 \leq C - \int_{t_1}^t \langle \dot{u}(\tau), e(\tau) + \beta\dot{e}(\tau)\rangle d\tau.$$

By Cauchy-Schwarz inequality we get

$$\frac{1}{2}\|\dot{u}(t)\|^2 \leq C + \int_{t_0}^{t} \|\dot{u}(\tau)\|\|e(\tau) + \beta\dot{e}(\tau)\|d\tau.$$

According to Gronwall's Lemma A.3

$$\|\dot{u}(t)\| \leq \sqrt{2C} + \int_{t_0}^{t} \|e(\tau) + \beta\dot{e}(\tau)\|d\tau \leq M := \sqrt{2C} + \int_{t_0}^{+\infty} \|e(\tau) + \beta\dot{e}(\tau)\|d\tau. \tag{80}$$

So, $\|\dot{u}(t)\|$ is bounded on $[t_0, +\infty[$, which allows us to define

$$W(t) = \frac{1}{2}\|\dot{u}(t)\|^2 + f(x(t)) - \int_{t}^{+\infty} \langle \dot{u}(\tau), \, e(\tau) + \beta\dot{e}(\tau)\rangle \, d\tau. \tag{81}$$

Noticing that W and W_T have the same derivative we conclude that

$$\dot{W}(t) + \frac{\alpha}{2t}\|\dot{x}(t)\|^2 + \frac{\alpha}{2t}\|\dot{u}(t)\|^2 \leq 0, \tag{82}$$

and thus W is non-increasing on $[t_1, +\infty[$.

Claim (ii) Integrating (82), and using that f, and hence W, is bounded from below, we obtain,

$$\int_{t_0}^{+\infty} \frac{1}{t}\|\dot{x}(t)\|^2 dt < +\infty, \text{ and } \int_{t_0}^{+\infty} \frac{1}{t}\|\dot{u}(t)\|^2 dt < +\infty. \tag{83}$$

Using Jensen's inequality, we get the integrability claim on $\xi(t)$.

Claim (iii) Given $z \in \mathcal{H}$, let us define $h : [t_0, +\infty[\to \mathbb{R}_+$ by $h(t) = \frac{1}{2}\|u(t) - z\|^2$. The function h is continuously differentiable with

$$\dot{h}(t) = \langle u(t) - z, \, \dot{u}(t)\rangle,$$

and $\dot{h}$ is absolutely continuous on compact subintervals of $[t_0, +\infty[$ (since $\dot{u}$ is) and satisfies

$$\ddot{h}(t) = \langle u(t) - z, \, \ddot{u}(t)\rangle + \|\dot{u}(t)\|^2$$

for almost every $t > t_0$. Using (75) and (76) we get

$$\ddot{u}(t) + \frac{\alpha}{t}\dot{u}(t) = -\left(1 - \frac{\alpha\beta}{t}\right)\xi(t) - (e(t) + \beta\dot{e}(t)).$$

Therefore, for almost every $t > t_0$

$$\begin{aligned}
&\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) = \|\dot{u}(t)\|^2 - \left\langle u(t) - z, \, \left(1 - \frac{\alpha\beta}{t}\right)\xi(t)\right\rangle - \langle u(t) - z, \, e(t) + \beta\dot{e}(t)\rangle \\
&= \|\dot{u}(t)\|^2 - \left(1 - \frac{\alpha\beta}{t}\right)\left\langle x(t) - z - x_0 + \beta\int_{t_0}^{t}\xi(s)ds, \, \xi(t)\right\rangle - \langle u(t) - z, \, e(t) + \beta\dot{e}(t)\rangle \\
&\leq \|\dot{u}(t)\|^2 - \left(1 - \frac{\alpha\beta}{t}\right)\langle x(t) - z, \, \xi(t)\rangle - \left(1 - \frac{\alpha\beta}{t}\right)\left\langle -x_0 + \beta\int_{t_0}^{t}\xi(s)ds, \, \xi(t)\right\rangle + \|e(t) + \beta\dot{e}(t)\|\|u(t) - z\|
\end{aligned}$$

To interpret $\left\langle -x_0 + \beta\int_{t_0}^{t}\xi(s)ds \, , \, \xi(t)\right\rangle$ as a temporal derivative, let us introduce

$$I(t) = \tfrac{1}{2\beta}\left\|-x_0 + \beta\int_{t_0}^{t}\xi(s)ds\right\|^2.$$

Then $I(\cdot)$ is locally absolutely continuous and $\dot{I}(t) = \left\langle -x_0 + \int_{t_0}^{t} \xi(s)ds,\, \xi(t) \right\rangle$ almost everywhere, because $\xi \in L^2(t_0, T; \mathcal{H}) \subseteq L^1(t_0, T; \mathcal{H})$ for all $T > t_0$; see part (vi)-(c) of Theorem 2.4. So,

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) \leq \|\dot{u}(t)\|^2 - \left(1 - \frac{\alpha\beta}{t}\right)\langle x(t) - z,\, \xi(t)\rangle - \left(1 - \frac{\alpha\beta}{t}\right)\dot{I}(t) + \|e(t) + \beta\dot{e}(t)\|\|u(t) - z\|,$$

for almost every $t > t_0$. On the other hand, by convexity of f and $\xi(t) \in \partial f(x(t))$

$$\langle x(t) - z,\, \xi(t)\rangle \geq f(x(t) - f(z).$$

Therefore

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) + \left(1 - \frac{\alpha\beta}{t}\right)(f(x(t)) - f(z)) + \left(1 - \frac{\alpha\beta}{t}\right)\dot{I}(t) \leq \|\dot{u}(t)\|^2 + \|e(t) + \beta\dot{e}(t)\|\|u(t) - z\|.$$

Using the definition (81) of W, we get

$$\begin{aligned}\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) + \left(1 - \frac{\alpha\beta}{t}\right)(W(t) - f(z)) + \left(1 - \frac{\alpha\beta}{t}\right)\dot{I}(t) &\leq \left(\frac{3}{2} - \frac{\alpha\beta}{2t}\right)\|\dot{u}(t)\|^2 \\ &+ \|e(t) + \beta\dot{e}(t)\|\|u(t) - z\| - \left(1 - \frac{\alpha\beta}{t}\right)\int_t^{+\infty}\langle \dot{u}(\tau),\, e(\tau) + \beta\dot{e}(\tau)\rangle\, d\tau.\end{aligned}$$

According to (82), we have $\|\dot{u}(t)\|^2 \leq -\frac{2t}{\alpha}\dot{W}(t)$. Therefore,

$$\begin{aligned}\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) + \left(1 - \frac{\alpha\beta}{t}\right)(W(t) - f(z)) + \left(1 - \frac{\alpha\beta}{t}\right)\dot{I}(t) &\leq -\left(\frac{3t}{\alpha} - \beta\right)\dot{W}(t) \\ &+ \|e(t) + \beta\dot{e}(t)\|\|u(t) - z\| - \left(1 - \frac{\alpha\beta}{t}\right)\int_t^{+\infty}\langle \dot{u}(\tau),\, e(\tau) + \beta\dot{e}(\tau)\rangle\, d\tau.\end{aligned}$$

Dividing by t and rearranging the terms, we have with $g(t) := e(t) + \beta\dot{e}(t)$

$$\begin{aligned}\frac{1}{t}\ddot{h}(t) + \left(\frac{1}{t} - \frac{\alpha\beta}{t^2}\right)(W(t) - f(z)) &\leq -\left(\frac{3}{\alpha} - \frac{\beta}{t}\right)\dot{W}(t) - \left[\frac{\alpha}{t^2}\dot{h}(t) + \left(\frac{1}{t} - \frac{\alpha\beta}{t^2}\right)\dot{I}(t)\right] \\ &+ \frac{1}{t}\|g(t)\|\|u(t) - z\| - \left(\frac{1}{t} - \frac{\alpha\beta}{t^2}\right)\int_t^{+\infty}\langle \dot{u}(\tau),\, g(\tau)\rangle\, d\tau.\end{aligned}$$

After integration, and using Lemma A.2, we get

$$\frac{1}{t}\dot{h}(t) + \int_{t_1}^{t}\left(\frac{1}{s} - \frac{\alpha\beta}{s^2}\right)\big(W(s) - f(z)\big)\, ds \leq -\int_{t_1}^{t}\left(\frac{3}{\alpha} - \frac{\beta}{s}\right)\dot{W}(s)\, ds + C + K_1(t) + K_2(t), \tag{84}$$

where

$$K_1(t) = \int_{t_1}^{t}\frac{1}{s}\|g(s)\|\|u(s) - z\|ds \quad \text{and} \quad K_2(t) = \int_{t_1}^{t}\left(\frac{1}{s} - \frac{\alpha\theta}{s^2}\right)\int_s^{\infty}\|\dot{u}(\tau)\|\|g(\tau)\|d\tau ds.$$

Let us majorize $K_1(t)$ and $K_2(t)$. The relation

$$\|u(s) - z\| \leq \|u(t_1) - z\| + \int_{t_1}^{s}\|\dot{u}(\tau)\|\, d\tau,$$

and $\dot{u}(\cdot)$ bounded (see (80)) give

$$K_1(t) \leq \int_{t_1}^{t}\frac{1}{s}\|g(s)\|\|u(s) - z\|ds \leq \left(\frac{\|u(t_1) - z\|}{t_1} + \sup_{t\geq t_1}\|\dot{u}(\tau)\|\right)\int_{t_1}^{+\infty}\|g(s)\|ds \leq C < +\infty.$$

For $K_2(t)$, we use again $\dot{u}(\cdot)$ bounded (see (80)) and integration by parts to obtain

$$K_2(t) \leq C\int_{t_1}^{t}\left(\frac{1}{s}\int_s^{\infty}\|g(\tau)\|\, d\tau\right)ds \leq C\left(\log t\int_t^{\infty}\|g(\tau)\|\, d\tau + \int_{t_1}^{t}\|g(\tau)\|\log\tau\, d\tau + 1\right).$$

Let us examine the integral terms that enter (84). Since $W(\cdot)$ is non-increasing

$$\begin{aligned}\int_{t_1}^{t}\left(\frac{1}{s}-\frac{\alpha\beta}{s^2}\right)\big(W(s)-f(z)\big)\,ds &\geq (W(t)-f(z))\int_{t_1}^{t}\left(\frac{1}{s}-\frac{\alpha\beta}{s^2}\right)ds\\ &= (W(t)-f(z))\left(\log t-\log t_1+\frac{\alpha\beta}{t}-\frac{\alpha\beta}{t_1}\right).\end{aligned}\tag{85}$$

In turn, integration by parts gives

$$\begin{aligned}-\int_{t_1}^{t}\left(\frac{3}{\alpha}-\frac{\beta}{s}\right)\dot W(s)\,ds &= \left(\frac{3}{\alpha}-\frac{\beta}{t_1}\right)\big(W(t_1)-f(z)\big)-\left(\frac{3}{\alpha}-\frac{\beta}{t}\right)(W(t)-f(z))+\beta\int_{t_1}^{t}\frac{W(s)-f(z)}{s^2}\,ds\\ &\leq \left(\frac{3}{\alpha}-\frac{\beta}{t_1}\right)\big(W(t_1)-f(z)\big)-\left(\frac{3}{\alpha}-\frac{\beta}{t}\right)(W(t)-f(z))+\beta\big(W(t_1)-f(z)\big)\left(\frac{1}{t_1}-\frac{1}{t}\right),\\ &\leq \frac{3}{\alpha}\big|W(t_1)-f(z)\big|-\left(\frac{3}{\alpha}-\frac{\beta}{t}\right)(W(t)-f(z))\end{aligned}\tag{86}$$

since $t\mapsto W(t)-f(z)$ is non-increasing and $t\geq t_1\geq\alpha\beta$. Combining (84) with (85) and (86), we obtain

$$\frac{1}{t}\dot h(t)+(W(t)-f(z))\left(\log t+D+\frac{E}{t}\right)\leq C\left(\log t\int_{t}^{\infty}\|g(\tau)\|\,d\tau+\int_{t_1}^{t}\|g(\tau)\|\log\tau\,d\tau+1\right).$$

for appropriate constants $C,D,E\in\mathbb{R}$. Now, take $t_2\geq t_1$ such that $\log s+D+\frac{E}{s}\geq 0$ for all $s\geq t_2$. Integrate from t_2 to t and use again that W is non-increasing to obtain

$$\begin{aligned}&\frac{h(t)}{t}-\frac{h(t_2)}{t_2}+\int_{t_2}^{t}\frac{h(s)}{s^2}ds+(W(t)-f(z))\int_{t_2}^{t}\left(\log s+D+\frac{E}{s}\right)ds\\ &\qquad\leq C'\int_{t_2}^{t}\left(\log s\int_{s}^{\infty}\|g(\tau)\|\,d\tau+\int_{t_1}^{s}\|g(\tau)\|\log\tau\,d\tau+1\right)ds.\end{aligned}$$

Since h is non-negative, this implies

$$\begin{aligned}&(W(t)-f(z))\,(t\log t+(D-1)t+E\log t+F)\\ &\qquad\leq C'\left(t+t\log t\int_{t}^{\infty}\|g(\tau)\|\,d\tau+\int_{t_2}^{t}\|g(\tau)\|\,\tau\log\tau\,d\tau+t\int_{t_1}^{t}\|g(\tau)\|\log\tau d\tau\right)+G,\end{aligned}\tag{87}$$

for some appropriate constants $D,E,F,G\in\mathbb{R}$. Divide by $t\log t$, let $t\to+\infty$, and use Lemma A.5, to obtain $\lim_{t\to+\infty}W(t)\leq f(z)$. The integrability of g and $\dot u(\cdot)$ bounded (see (80)) yield $\lim\limits_{t\to+\infty}\int_{t}^{+\infty}\langle\dot u(\tau),\,g(\tau)\rangle\,d\tau=0$. As a consequence,

$$\lim_{t\to+\infty}\left(f(x(t))+\frac{1}{2}\|\dot x(t)+\beta\xi(t)\|^2\right)\leq f(z)$$

for each $z\in\mathcal{H}$. Thus

$$\inf_{\mathcal{H}}f\leq\liminf_{t\to+\infty}f(x(t))\leq\limsup_{t\to+\infty}f(x(t))\leq\lim_{t\to+\infty}\left(f(x(t))+\frac{1}{2}\|\dot x(t)+\beta\xi(t)\|^2\right)\leq\inf_{\mathcal{H}}f,$$

whence we get $\lim_{t\to+\infty}f(x(t))=\inf_{\mathcal{H}}f$, and thus $\lim_{t\to+\infty}\|\dot x(t)+\beta\xi(t)\|=0$.

Claim (iv) This follows from claim (iii) and lower semicontinuity of f.

Claim (v)-(a) Let $x^\star\in S$. We start from (87) with $z=x^\star$ and divide by t. To conclude, we note that

$$\log t\int_{t}^{\infty}\|g(\tau)\|\,d\tau\leq\int_{t}^{\infty}\log\tau\,\|g(\tau)\|\,d\tau<+\infty,$$

$$\int_{t_2}^{t}\|g(\tau)\|\,\frac{\tau}{t}\log\tau\,d\tau\leq\int_{t_2}^{t}\|g(\tau)\|\log\tau\,d\tau<+\infty\ \text{ and}$$

$$\log t\int_{t}^{+\infty}\langle\dot u(\tau),\,g(\tau)\rangle\,d\tau\leq C\int_{t}^{+\infty}\log\tau\,\|g(\tau)\|\,d\tau,$$

where $C=\sup_{t\geq t_0}\|\dot u(t)\|<+\infty$ (see (80)).

Claim (v)-(b) Putting together (84) and (86) with $z = x^\star \in S$, and using non-negativity of h, we infer that for some positive constant C

$$\left(1 - \frac{\alpha\beta}{t_1}\right) \int_{t_1}^{t} \frac{1}{s}\left(W(s) - f(z)\right) ds \leq C + K_2(t).$$

Arguing similarly as for proving part (v)-(a), we can show that $K_2(\cdot)$ is bounded. Thus

$$\int_{t_1}^{t} \frac{1}{s}\left(f(x(s)) - f(z) + \|\dot{u}(s)\|^2\right) ds \leq \int_{t_1}^{t} \frac{1}{s}\left(W(s) - f(z)\right) ds + \frac{1}{t_1}\int_{t_1}^{t} \|\dot{u}(s)\| \, \|g(s)\| \, ds < +\infty,$$

which completes the proof. □

5.1.2 Fast convergence rates

When $\alpha \geq 3$, under a reinforced integrability assumption on the perturbation term, we will show fast convergence results. The following theorem is the non-smooth counterpart of Theorem 3.2.

Theorem 5.2. *Suppose that $\alpha \geq 3$. Let $f \in \Gamma_0(\mathcal{H})$ such that $S \neq \emptyset$. Suppose that $e(\cdot) \in \mathcal{W}^{1,1}(t_0, T; \mathcal{H})$ for all $T > t_0$, with $\int_{t_0}^{+\infty} t\|e(t) + \beta\dot{e}(t)\|dt < +\infty$. Then, for any global strong solution (x, y) of (4)*

(i) $f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(t^{-2}\right)$.

(ii) $\displaystyle\int_{t_0}^{+\infty} t(f(x(t)) - \inf_{\mathcal{H}} f)dt < +\infty, \ \int_{t_0}^{+\infty} t^2\|\xi(t)\|^2 dt < +\infty, \ \int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty.$

(iii) $\|\dot{x}(t) + \beta\xi(t)\| = \mathcal{O}(t^{-1})$.

Proof. Let $(x, y) : [t_0, +\infty[\to \mathcal{H} \times \mathcal{H}$ be a global strong solution of (4). Take $\alpha \geq 3$ and $x^\star \in S$. Set $f^\star = \inf_{\mathcal{H}} f$ and $g(t) = e(t) + \beta\dot{e}(t)$. Our analysis relies on the non-smooth version of the Lyapunov function in (31), which is defined for $\lambda \in [2, \alpha - 1]$, as $\mathcal{E}_{\lambda,T} : [t_0, T] \to \mathbb{R}$ by

$$\mathcal{E}_{\lambda,T}(t) = t(t - \beta(\lambda + 2 - \alpha))(f(x(t)) - f^\star) + \frac{1}{2}\|v_\lambda(t)\|^2 + \lambda(\alpha - \lambda - 1)\frac{1}{2}\|x(t) - x^\star\|^2 - \int_t^T \tau \left\langle v_\lambda(\tau),\, g(\tau)\right\rangle d\tau, \tag{88}$$

where $v_\lambda(t) := \lambda(x(\tau) - x^\star) + \tau\dot{u}(t)$, and u is defined on $[t_0, +\infty[$ by (73) and $\dot{u}$ is given by (74).
$\mathcal{E}_{\lambda,T}(\cdot)$ is the sum of four terms, each of which is absolutely continuous on $[t_0, T]$ for all $T > t_0$. Hence $\mathcal{E}_{\lambda,T}$ is differentiable almost everywhere. We first differentiate each term of $\mathcal{E}_{\lambda,T}$:

$$\frac{d}{dt}\left[t(t - \beta(\lambda + 2 - \alpha))(f(x(t)) - f^\star)\right] = (2t - \beta(\lambda + 2 - \alpha))(f(x(t)) - f^\star) + t(t - \beta(\lambda + 2 - \alpha))\left\langle \xi(t),\, \dot{x}(t)\right\rangle.$$

Using (76), we have

$$\begin{aligned}
\frac{d}{dt}\frac{1}{2}\|v_\lambda(t)\|^2 &= \left\langle \lambda(x(t) - x^\star) + t\dot{u}(t),\, \lambda\dot{x}(t) + \dot{u}(t) + t\ddot{u}(t)\right\rangle \\
&= \left\langle \lambda(x(t) - x^\star) + t\dot{u}(t),\, (\lambda + 1 - \alpha)\dot{x}(t) - (t - \beta)\xi(t) - tg(t)\right\rangle \\
&= \lambda(\lambda + 1 - \alpha)\left\langle x(t) - x^\star,\, \dot{x}(t)\right\rangle - t(\alpha - \lambda - 1)\|\dot{x}(t)\|^2 - \beta t(t - \beta)\|\xi(t)\|^2 \\
&\quad - \lambda(t - \beta)\left\langle x(t) - x^\star,\, \xi(t)\right\rangle - t(t - \beta(\lambda + 2 - \alpha))\left\langle \xi(t),\, \dot{x}(t)\right\rangle - t\left\langle v_\lambda(t),\, g(t)\right\rangle, \\
\frac{d}{dt}\lambda(\alpha - \lambda - 1)\frac{1}{2}\|x(t) - x^\star\|^2 &= \lambda(\alpha - \lambda - 1)\left\langle x(t) - x^\star,\, \dot{x}(t)\right\rangle, \quad \text{and} \\
\frac{d}{dt}\left(-\int_t^T \tau\left\langle v_\lambda(\tau),\, g(\tau)\right\rangle d\tau\right) &= t\left\langle v_\lambda(t),\, g(t)\right\rangle.
\end{aligned}$$

By collecting these results, the perturbation terms cancel each other out. We get

$$\begin{aligned}\frac{d}{dt}\mathcal{E}_{\lambda,T}(t) &= (2t-\beta(\lambda+2-\alpha))(f(x(t))-f^\star)-\lambda(t-\beta)\left\langle x(t)-x^\star,\, \xi(t)\right\rangle \\ &\quad -t(\alpha-\lambda-1)\|\dot{x}(t)\|^2-\beta t(t-\beta)\|\xi(t)\|^2, \qquad (89)\end{aligned}$$

for almost all $t > t_0$. Since $\xi(t) \in \partial f(x(t))$ for all $t > t_0$, we have $\langle \xi(t),\, x(t)-x^\star\rangle \geq f(x(t)-f(x^\star)$, and we deduce from (89), that

$$\frac{d}{dt}\mathcal{E}_{\lambda,T}(t) \leq -((\lambda-2)t-\beta(\alpha-2))(f(x(t))-f^\star)-t(\alpha-\lambda-1)\|\dot{x}(t)\|^2-\beta t(t-\beta)\|\xi(t)\|^2, \quad (90)$$

for almost all $t \geq t_1 = \max(t_0, \beta)$. It follows that $\mathcal{E}_{\lambda,T}$ is non-increasing on $[t_1, T]$. In particular, $\mathcal{E}_{\lambda,T}(t) \leq \mathcal{E}_{\lambda,T}(t_1)$ for $t_1 \leq t \leq T$. This gives the existence of a constant C such that

$$\frac{1}{2}\|v_\lambda(t)\|^2 \leq C + \int_{t_0}^{t} \|v_\lambda(t)\|\|\tau g(\tau)\|d\tau. \qquad (91)$$

Applying Lemma A.3 to (91), and using the integrability of $t \mapsto tg(t)$, it follows that

$$\sup_{t\geq t_0}\|v_\lambda(t)\| \leq \sqrt{2C} + \int_{t_0}^{\infty} \|\tau g(\tau)\|d\tau < +\infty. \qquad (92)$$

As a consequence, we can define the energy function

$$\mathcal{E}_\lambda(t) := t(t-\beta(\lambda+2-\alpha))(f(x(t))-f^\star)+\frac{1}{2}\|v_\lambda(t)\|^2+\lambda(\alpha-\lambda-1)\frac{1}{2}\|x(t)-x^\star\|^2-\int_t^\infty \tau\left\langle v_\lambda(t),\, g(\tau)\right\rangle d\tau,$$

which has the same derivative as $\mathcal{E}_{\lambda,T}$. Hence $\mathcal{E}_\lambda(t) \leq \mathcal{E}_\lambda(t_0)$. Combined with (92), this gives

$$t(t-\beta(\lambda+2-\alpha))(f(x(t))-f^\star) \leq C + \sup_{t\geq t_0}\|v_\lambda(t)\| \int_{t_0}^{\infty} \|\tau g(\tau)\|d\tau < +\infty,$$

whence statement (i). Claim (iii) is obtained by letting $\lambda = 0$ in (92). Integration of (90) gives the integral estimates of (ii), which completes the proof. □

5.1.3 Convergence of the trajectories and faster asymptotic rates

Similar argument as in the smooth case (see Theorem 3.3), but now using the Lyapunov function (88), gives weak convergence of the trajectories of (4). Moreover, in the same vein as Theorem 3.2, $o(\cdot)$ rates can also be obtained. We leave the details to the readers for the sake of brevity.

Theorem 5.3. *Let $\alpha > 3$. Let $f \in \Gamma_0(\mathcal{H})$ and assume that $S = \operatorname{argmin} f \neq \emptyset$. Suppose that $e(\cdot) \in \mathcal{W}^{1,1}(t_0, T; \mathcal{H})$ for all $T > t_0$, with $\int_{t_0}^{+\infty} t\|e(t)+\beta\dot{e}(t)\|dt < +\infty$. Then, for any global strong solution (x, y) of (4)*

(i) *$x(t)$ converges weakly, as $t \to +\infty$ to a point in S;*

(ii) *$f(x(t)) - \inf_{\mathcal{H}} f = o\left(t^{-2}\right)$ and $\|\dot{x}(t)+\beta\xi(t)\| = o\left(t^{-1}\right)$ as $t \to +\infty$.*

5.2 Implicit Hessian Damping

As we have already discussed in the smooth case (see Section 3.3), the analysis of the convergence properties of the system with implicit Hessian driven damping heavily relies on Lipschitz continuity of the gradient. As shown above, such a property was not needed to analyze system (4). Therefore, the study of the convergence properties for the non-smooth system (12) (even without perturbations) is an open challenging topic.

6 Numerical Experiments

To support our theoretical claims, we consider numerical examples in $\mathcal{H} = \mathbb{R}^2$ with two real-valued functions:

- The first one is given by $f(x_1, x_2)) = (x_1 - 1)^4 + (x_2 - 5)^2$. This function is obviously convex (but not strongly so) and smooth, and has a unique minimizer at $(1, 5)$. For this function, we consider the continuous time dynamical system (ISEHD-Pert) with parameters $(\alpha, \beta) = (3.1, 1)$, and (ISIHD-Pert) with parameters $(\alpha, \gamma, \beta) = (3.1, 1, 1)$.

- The second example we consider is with the convex non-smooth function $f(x_1, x_2) = (x_1 - 1)^4 + (x_2 - 5)^2 + 0.1(|x_1| + |x_2|)$. For this function, we use the continuous time non-smooth system (4) with parameters $(\alpha, \beta) = (3.1, 1)$. Although we have no theoretical guarantee for system (12), we do report the corresponding numerical results with parameters $(\alpha, \gamma, \beta) = (3.1, 1, 1)$.

For both examples, we take as an exogenous perturbation

$$e(t) = \frac{\cos(2\pi t)}{t^\delta} \text{ with } \delta \in \{0.1, 1.1, 3.1\}.$$

All systems are solved numerically with a Runge-Kutta adaptive method in MATLAB on the time interval $[1, 50]$ with initial data $(x_0, \dot{x}_0) = (-10, 20, 5, -5)$. The results are displayed in Figure 1 and Figure 2.

Let us first comment on the results for the smooth function. For $\delta = 3.1$, all required moment assumptions on the errors are fulfilled (for the explicit Hessian, the term $\dot{e}$ is dominated by e and can then be discarded). Hence the fast rates predicted by Theorem 3.2(i) and Theorem 3.7(i) as well as convergence of the trajectories (see Theorem 3.3 and Theorem 3.8) hold true. For the value $\delta = 0.1$, since the error is not even integrable, neither the convergence of the objective value nor that of the trajectories is ensured, with large oscillations appearing. The implicit Hessian damping seems also less stable as anticipated from our discussion in Section 3.3. For $\delta = 1.1$, though there is no convergence guarantee for the trajectory, the objective value for (ISEHD-Pert) decreases but at a rate which is dominated by the error decrease. This can be explained in light of the proof of Theorem 3.2(i), where a close inspection of (24) and (26) shows that the bound on the objective error decomposes as

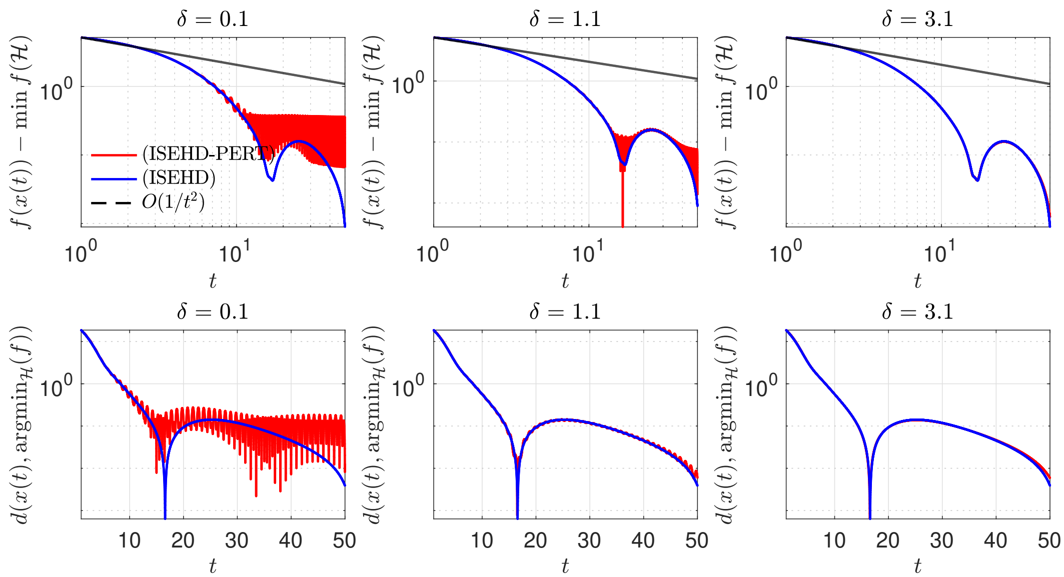
$$f(x(t)) - f^\star \leq \mathcal{O}\left(\tfrac{1}{t^2}\right) + \frac{C\left(\displaystyle\int_{t_0}^{t} \tau \|e(\tau)\| d\tau\right)^2}{t^2}.$$

For $\delta \in]1, 2]$, the second term indeed dominates the first one and decreases at the slower rate $t^{-2(\delta-1)}$. This confirms the known rule that there is a trade-off between fast convergence of the methods and their robustness to perturbations.

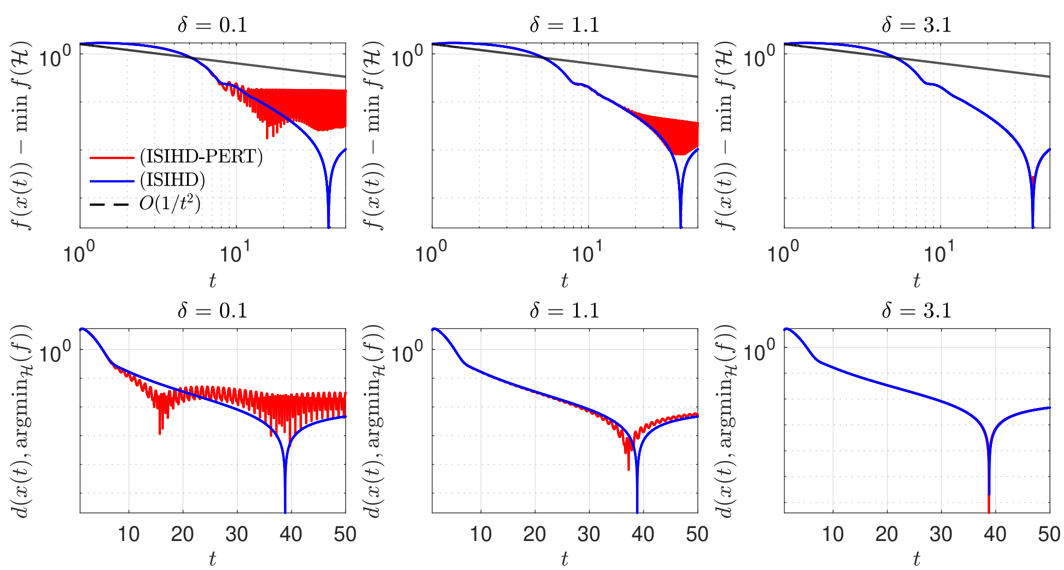
Similar observations remain true for the non-smooth function with system (4) where we now invoke Theorem 5.2 and Theorem 5.3. As for system (12), it seems that it has a behavior similar to what we observed in the smooth case for system (ISIHD-Pert). As we argued in Section 5.2, supplementing the numerical observations for system (12) with theoretical guarantees is an open problem that we leave to a future work.

7 Conclusion and Perspectives

The introduction of the correction term attached to the damping driven by the Hessian in first-order accelerated optimization algorithms makes it possible to considerably dampen the oscillations in the trajectory.

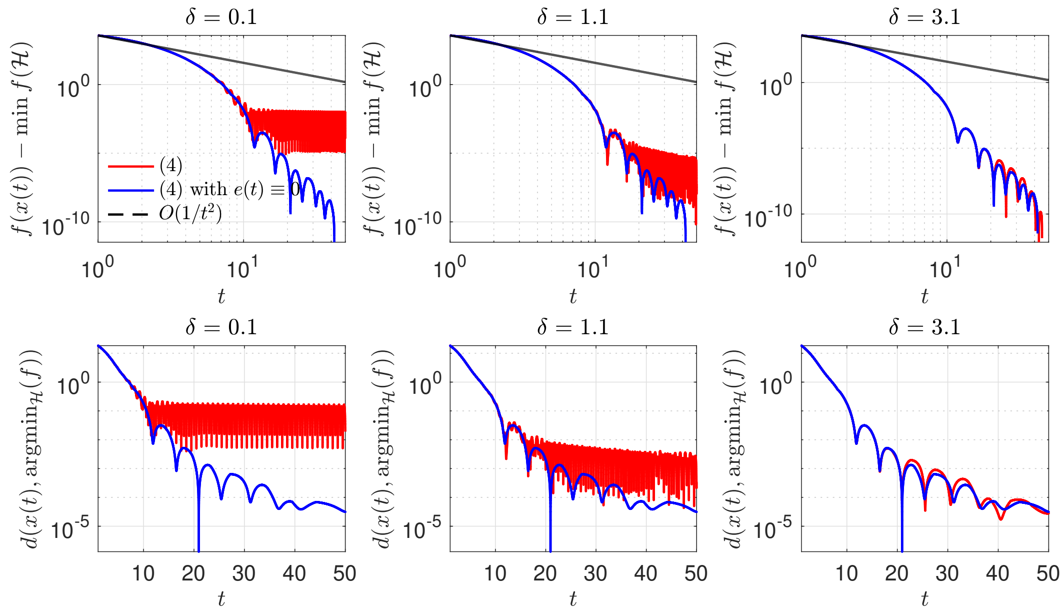


(a) Explicit Hessian damping: smooth function

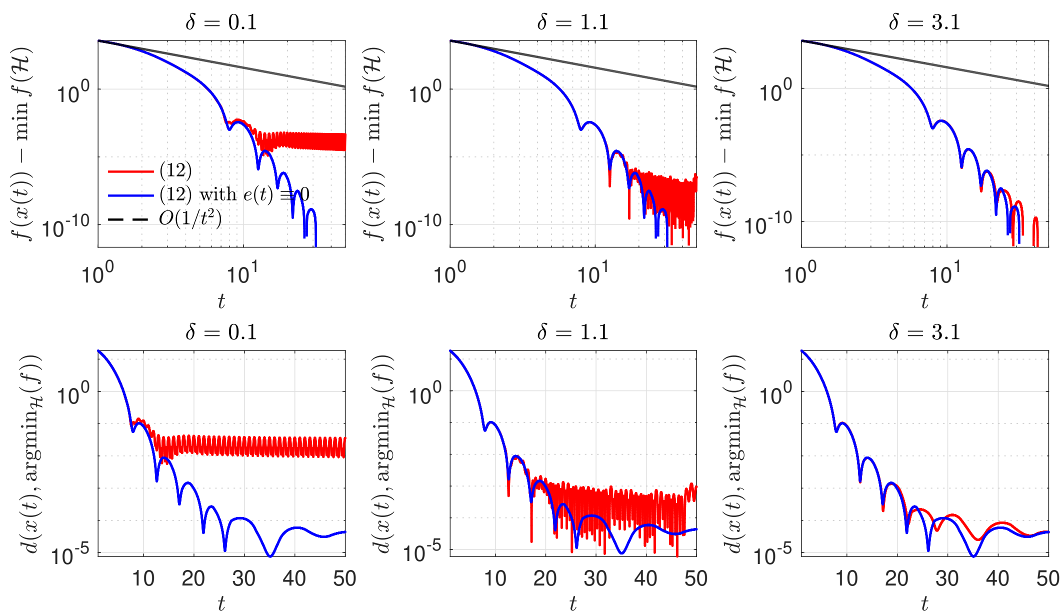


(b) Implicit Hessian damping: smooth function

Figure 1: Example on a smooth function: Evolution of the objective error and distance to the minimizer as a function of t for different error decay exponents.



(a) Explicit Hessian damping: non-smooth function



(b) Implicit Hessian damping: non-smooth function

Figure 2: Example on a non-smooth function: Evolution of the objective error and distance to the minimizer as a function of t for different error decay exponents.

The study of the robustness of these algorithms with respect to error perturbations is crucial for their further development in a stochastic framework. Our systematic study of these questions for the dynamics underlying these algorithms is a fundamental first step in this direction. We paid particular attention to the explicit and the implicit forms of the Hessian driven damping, showing several advantages of the explicit form. Our study concerns the dynamics with damping driven by the Hessian within the framework of the Nesterov acceleration gradient method. It shows that the convergence of the values still holds when the error terms satisfy an appropriate integrability condition, and fast convergence is satisfied when the (first or second-order) moment of the error is finite. Indeed, as a general rule, there is a balance between the rate of convergence of the methods and their robustness with respect to error disturbances. An interesting technique studied in [1, 2] is the introduction of a dry friction term. This makes it possible to have errors which do not necessarily go to zero, they must not exceed a certain threshold, but on the other hand we only obtain an approximate solution. Finding the right balance between the convergence rate and robustness is an important issue that should be the subject of further study. Another important aspect of our study is the fact that several results are valid in the case of a non-smooth function. This opens the door to the study of similar topics with respect to structured composite optimization problems involving a non-smooth term. These are some of the many facets of these flexible dynamics and algorithms which, in the unperturbed case, have been applied in various fields including PDE's and mechanical shocks [15], deep learning [22], non-convex optimization [5], monotone inclusions [13, 14] to mention a few important applications.

A Auxiliary results

Let us first recall the continuous form of the Opial's Lemma [29], a key ingredient to establish convergence of the trajectories.

Lemma A.1. *Let S be a nonempty subset of $\mathcal{H}$ and let $x : [0, +\infty[\to \mathcal{H}$. Assume that*

(i) *for every $z \in S$, $\lim_{t\to\infty} \|x(t) - z\|$ exists;*

(ii) *every weak sequential cluster point of $x(t)$, as $t \to \infty$, belongs to S.*

Then $x(t)$ converges weakly as $t \to \infty$ to a point in S.

Lemma A.2 ([16, Lemma 7.3])**.** *Let $\tau, p > 0$ and let $\psi :]\tau, +\infty[\to \mathbb{R}$ be $\mathcal{C}^2$ and bounded from below. Then,*

$$\inf_{t>\tau} \int_\tau^t \frac{\dot\psi(s)}{s^p}\, ds > -\infty \quad \textit{and} \quad \inf_{t>\tau} \int_\tau^t \frac{\ddot\psi(s)}{s^p}\, ds - \frac{\dot\psi(t)}{t^p} > -\infty.$$

Lemma A.3 ([21, Lemma A.5])**.** *Let $m : [t_0; T] \to [0, +\infty[$ be integrable. Suppose $w : [t_0, T] \to \mathbb{R}$ is continuous and*

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

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

$$|w(t)| \leq c + \int_{t_0}^t m(s) ds, \qquad t \in [t_0, T].$$

Lemma A.4. *Let $\beta : [t_0, +\infty[\to \mathbb{R}_+$ such that $\inf_{t\in[t_0,+\infty[} \beta(t) > 0$ and $\beta \notin L^1(t_0, +\infty; \mathbb{R})$. Then, the differential inclusion*

$$\dot z(t) + \beta(t)\partial\Phi(z(t)) + F(t, z(t)) \ni 0, \tag{93}$$

is equivalent to

$$\dot{w}(s) + \partial\Phi(w(s)) + G(s, w(s)) \ni 0, \tag{94}$$

with

$$G(s, w(s)) = \frac{1}{\beta(\tau(s))} F(\tau(s), w(s)), \quad t = \tau(s), \ \textit{and} \ \beta(\tau(s))\dot{\tau}(s) = 1.$$

Proof. Make the change of time variable $t = \tau(s)$ and $z(t) = z \circ \tau(s) = w(s)$. We then have

$$\frac{1}{\beta(\tau(s))\dot{\tau}(s)} \dot{w}(s) + \partial\Phi(w(s)) + \frac{1}{\beta(\tau(s))} F(\tau(s), w(s)) \ni 0,$$

Choose $\tau(\cdot)$ such that

$$\beta(\tau(s))\dot{\tau}(s) = 1. \tag{95}$$

Introduce a primitive of β, $p(t) = \int_{t_0}^{t} \beta(r)dr$ Therefore, (95) can be equivalently written

$$\tfrac{d}{ds} p(\tau(s)) = 1 \iff p(\tau(s)) = \textstyle\int_{t_0}^{\tau(s)} \beta(r)dr = s + C,$$

for any constant C. Thus, τ defines a change of variable if and only if $\int_{t_0}^{+\infty} \beta(r)dr = +\infty$, hence our assumption on β. $\square$

Lemma A.5. *Take $t_0 > 0$, and let $f \in L^1(t_0, +\infty; \mathbb{R})$ be non-negative and continuous. Consider a nondecreasing function $\varphi : [t_0, +\infty[\to \mathbb{R}_+$ such that $\lim\limits_{t \to +\infty} \varphi(t) = +\infty$. Then, $\lim_{t \to +\infty} \frac{1}{\varphi(t)} \int_{t_0}^{t} \varphi(s) f(s)ds = 0$.*

Proof. Given $\epsilon > 0$, fix t_ϵ so that $\int_{t_\epsilon}^{\infty} f(s)ds \leq \epsilon$. Then, for $t \geq t_\epsilon$, split the integral $\displaystyle\int_{t_0}^{t} \varphi(s) f(s)ds$ into two parts to obtain

$$\frac{1}{\varphi(t)} \int_{t_0}^{t} \varphi(s) f(s)ds = \frac{1}{\varphi(t)} \int_{t_0}^{t_\epsilon} \varphi(s) f(s)ds + \frac{1}{\varphi(t)} \int_{t_\epsilon}^{t} \varphi(s) f(s)ds \leq \ \tfrac{1}{\varphi(t)} \textstyle\int_{t_0}^{t_\epsilon} \varphi(s) f(s)ds + \int_{t_\epsilon}^{t} f(s)ds.$$

Let $t \to +\infty$ to deduce that $0 \leq \limsup_{t \to +\infty} \frac{1}{\varphi(t)} \int_{t_0}^{t} \varphi(s) f(s)ds \leq \epsilon$. Since this is true for any $\epsilon > 0$, the result follows. $\square$

Lemma A.6 ([9, Lemma A.4])**.** *Let $t_0 > 0$, and let $w : [t_0, +\infty[\to \mathbb{R}$ be a continuously differentiable function which is bounded from below. Assume that*

$$t\ddot{w}(t) + \alpha\dot{w}(t) \leq g(t),$$

for some $\alpha > 1$, almost every $t > t_0$, and some non-negative function $g \in L^1(t_0, +\infty, \mathbb{R})$. Then, the positive part $[w]_+$ of w belongs to $L^1(t_0, +\infty; \mathbb{R})$ and $\lim_{t \to +\infty} w(t)$ exists.

References

[1] S. Adly, H. Attouch, *Finite convergence of proximal-gradient inertial algorithms combining dry friction with Hessian-driven damping*, SIAM J. Optim., 30(3) (2020), pp. 2134–2162.

[2] S. Adly, H. Attouch, *Finite time stabilization of continuous inertial dynamics combining dry friction with Hessian-driven damping*, J. Conv. Analysis, 28 (2) (2021), pp. 281–310.

[3] C.D. Alecsa, S. László, T. Pinta, *An extension of the second order dynamical system that models Nesterov's convex gradient method*, Appl. Math. Optim., (2020), https://doi.org/10.1007/s00245-020-09692-1

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

[5] H. Attouch, R.I. Boţ, E.R. Csetnek, *Fast optimization via inertial dynamics with closed-loop damping*, Journal of the European Mathematical Society (JEMS), 2021, arXiv:2008.02261v2 [math.OC] Sep 2020.

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

[7] H. Attouch, A. Cabot, Z. Chbani, H. Riahi, *Accelerated forward-backward algorithms with perturbations. Application to Tikhonov regularization*, JOTA, 179 (1) (2018), pp. 1–36 .

[8] H. Attouch, Z. Chbani, J. Fadili, H. Riahi, *First order optimization algorithms via inertial systems with Hessian driven damping,* Math. Program. (2020), https://doi.org/10.1007/s10107-020-01591-1.

[9] H. Attouch, Z. Chbani, J. Peypouquet, P. Redont, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity, Mathematical Programming*, 168 (1-2) (2018), pp. 123–175.

[10] H. Attouch, Z. Chbani, H. Riahi, *Fast proximal methods via time scaling of damped inertial dynamics*, SIAM J. Optim., 29 (3) (2019), pp. 2227–2256.

[11] H. Attouch, M.-O. Czarnecki, *Asymptotic control and stabilization of nonlinear oscillators with non-isolated equilibria*, J. Differential Equations, 179 (1) (2002), pp. 278–310.

[12] H. Attouch, A. Damlamian, *Strong solutions for parabolic variational inequalities*, Nonlinear Analysis, TMA, 2(3) (1978), pp. 329-353.

[13] H. Attouch, S. C. László, *Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators*, SIAM J. Optim., 30(4) (2020), pp. 3252–3283.

[14] H. Attouch, S. C. László, *Continuous Newton-like Inertial Dynamics for Monotone Inclusions*, Set Valued and Variational Analysis, (2020), https://doi.org/10.1007/s11228-020-00564-y, hal-02577331.

[15] H. Attouch, P.E. Maingé, P. Redont, *A second-order differential system with Hessian-driven damping; Application to non-elastic shock law*s, Differential Equations and Applications, 4 (1) (2012), pp. 27–65.

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

[17] J.-F. Aujol and C. Dossal, *Stability of over-relaxations for the forward-backward algorithm, application to fista*, SIAM Journal on Optimization, 25 (4) (2015), pp. 2408–2433.

[18] J.-F. Aujol, C. Dossal, A. Rondepierre, *Optimal convergence rates for Nesterov acceleration*, SIAM Journal on Optimization, 29 (4) (2019), pp. 3131–3153.

[19] A. Beck, M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), No. 1, pp. 183–202.

[20] R. I. Bot, E. R. Csetnek, S.C. Laszlo, *Tikhonov regularization of a second order dynamical system with Hessian damping*, (2020), Math. Program., DOI:10.1007/s10107-020-01528-8.

[21] H. Brézis, *Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution*, Lecture Notes 5, North Holland, (1972).

[22] C. Castera, J. Bolte, C. Févotte, E. Pauwels, *An Inertial Newton Algorithm for Deep Learning*. 2019. HAL-02140748.

[23] A. Chambolle, Ch. Dossal, *On the convergence of the iterates of the Fast Iterative Shrinkage Thresholding Algorithm*, J. Opt. Theory Appl., 166 (2015), pp. 968–982.

[24] A. Haraux, M. A. Jendoubi, *On a second order dissipative ode in Hilbert space with an integrable source term*, Acta Math. Sci., 32 (2012), pp. 155–163.

[25] T. Lin, M. I. Jordan, *A Control-Theoretic Perspective on Optimal High-Order Optimization*, arXiv:1912.07168v1 [math.OC] Dec 2019.

[26] M. Muehlebach, M. I. Jordan, *A Dynamical Systems Perspective on Nesterov Acceleration*, (2019), arXiv:1905.07436

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

[28] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, volume 87 of Applied Optimization. Kluwer, 2004.

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

[30] B. T. Polyak, *Introduction to Optimization*, New York, Optimization Software, 1987.

[31] M. Schmidt, N. Le Roux, F. Bach, *Convergence rates of inexact proximal-gradient methods for convex optimization*, NIPS'11 - 25 th Annual Conference on Neural Information Processing Systems, Dec 2011, Grenada, Spain. (2011) HAL inria-00618152v3.

[32] B. Shi, S. S. Du, M. I. Jordan, W. J. Su, *Understanding the acceleration phenomenon via high-resolution differential equations*, arXiv:submit/2440124[cs.LG] 21 Oct 2018.

[33] W. Su, S. Boyd, E. J. Candès, *A Differential Equation for Modeling Nesterov's Accelerated Gradient Method*, Advances in Neural Information Processing Systems **27** (NIPS 2014).

[34] S. Villa, S. Salzo, L. Baldassarres, A. Verri, *Accelerated and inexact forward-backward*, SIAM J. Optim., 23 (3) (2013), pp. 1607–1633.