SECOND-ORDER DYNAMICS WITH HESSIAN-DRIVEN DAMPING FOR LINEARLY CONSTRAINED CONVEX MINIMIZATION*

SIMON K. NIEDERLÄNDER†

Abstract. In a real Hilbert space setting, we investigate the asymptotic properties of the solutions of a second-order differential system in view of linearly constrained convex minimization. The inertial dynamics are governed by a Hessian-driven damping term associated with the convex function to be minimized and potential effects induced by the linear constraints. We provide conditions on both the damping and the potential for which the solutions converge towards some feasible point of the convex minimization problem; this convergence is towards some minimizer provided that the solutions' initial data is specifically preselected. In addition, we present asymptotic estimates on the convergence rate of the solutions depending on the interaction between damping and potential effects. Our analysis is mainly based on energy-like arguments that capture the dissipative nature of the inertial dynamics by means of a Bregman distance. We complement our study with the fact that the second-order dynamics admit a first-order representation in terms of the Arrow–Hurwicz differential system. Numerical experiments further illustrate our theoretical findings.

Key words. dissipative inertial dynamics, Hessian-driven damping, Lyapunov analysis, asymptotic properties, Bregman distance, convex minimization

AMS subject classifications. 37N40, 46N10, 49M30, 65K05, 90C25

DOI. 10.1137/20M1323679

1. Introduction. Let X, Y be real Hilbert spaces endowed with inner products $\langle\,\cdot\,,\,\cdot\,\rangle_X$, $\langle\,\cdot\,,\,\cdot\,\rangle_Y$ and induced norms $\|\cdot\|_X$, $\|\cdot\|_Y$. Given a convex and twice continuously differentiable function $f : X \to \mathbb{R}$ and a continuous affine operator $h : X \to Y$, we investigate the second-order differential system

$$\text{(ID)} \qquad \ddot{x} + \nabla^2 f(x)\dot{x} + \nabla\|h(x)\|_Y^2/2 = 0_X,$$

where $\nabla\|h(\,\cdot\,)\|_Y^2/2$ denotes the gradient of $\|h(\,\cdot\,)\|_Y^2/2$ and $\nabla^2 f$ refers to the Hessian of f. The first and second derivatives of a solution $x : [0, +\infty) \to X$ of (ID) are denoted by $\dot{x}$ and $\ddot{x}$, respectively.

Although the Inertial Dynamics (ID) may appear in various physical contexts as Liénard-type equation [23], our motivation to study (ID) pertains to the dynamical approach of solving the linearly constrained convex minimization problem

$$\text{(P)} \qquad \inf\{f(x) \mid h(x) = 0_Y\}.$$

The constrained nature of (P) thereby necessitates specific structural properties of the dynamics to account for the minimization of f relative to h. As a decisive feature of (ID), the objective function f enters the dynamics through the Hessian-driven damping operator $\nabla^2 f(\,\cdot\,) : X \to X$ while the constraint function h further induces potential effects via the gradient mapping $\nabla\|h(\,\cdot\,)\|_Y^2/2 : X \to X$. The minimizing properties of (ID) with respect to the convex minimization problem (P) are originated by this structure: Under suitable assumptions, the potential effects tend to stabilize

*Received by the editors March 5, 2020; accepted for publication (in revised form) June 14, 2021; published electronically October 18, 2021.

https://doi.org/10.1137/20M1323679

Funding: This research was supported by the German Research Foundation (DFG).

†Institute for Systems Theory and Automatic Control, University of Stuttgart, Pfaffenwaldring 9, 70550 Stuttgart, Germany (simon.niederlaender@ist.uni-stuttgart.de).

the solutions of (ID) asymptotically towards feasible points of (P) whereas the geometric damping further allows one to asymptotically select for a minimizer of (P).

The structural properties of (ID) are in contrast to the various classical (inertial) approaches for unconstrained minimization, where the objective function typically induces the potential effects via its gradient; see, e.g., [1, 2, 6, 7, 8, 17]. While these methods in general allow for a fast minimization of the objective function, they require some arrangement to explicitly take constraints into account. Our study aims to highlight that the linearly constrained convex minimization problem (P) can be dynamically approached via the second-order differential system (ID) by interchanging the role of the damping and potential. Within this context, we aim at providing conditions on f and h for which the solutions of (ID) converge, at an explicitly characterized rate, towards some minimizer of (P).

Motivation and perspectives. To motivate our study, let us comment on the various perspectives of the inertial dynamics (ID) relative to the convex minimization problem (P).

"Generalized steepest descent". We first note that the inertial dynamics (ID) may be regarded as a differential version of the classical Arrow–Hurwicz evolution system

$$\text{(AH)} \qquad \begin{cases} \dot{x} + \nabla f(x) + h'(x)^* \lambda = 0_X \\ \dot{\lambda} - h(x) = 0_Y \end{cases}$$

with h' denoting the Fréchet derivative of h, and $h'(x)^*$ the adjoint operator of $h'(x)$. Indeed, differentiating the first component of the (AH) model yields

$$\ddot{x} + \nabla^2 f(x)\dot{x} + h'(x)^* \dot{\lambda} = 0_X.$$

The elimination of the dual variable by means of the second component of (AH) then leads to the inertial dynamics (ID). Conversely, an immediate integration shows that any solution $x : [0, +\infty) \to X$ of (ID) with corresponding initial data $(x_0, -\nabla f(x_0)) \in X \times X$ further obeys the integro-differential system

$$\dot{x} + \nabla f(x) + h'(x)^* \int h(x) = 0_X.$$

Introducing the auxiliary variable $\lambda : [0, +\infty) \to Y$ as the integral of $h(x)$ then gives rise to the (AH) differential system.

The (AH) dynamics were in essence originated by Arrow, Hurwicz, and Uzawa [4] and are known to be intimately related to the mini-maximization of the Lagrangian

$$\begin{aligned} L : X \times Y &\longrightarrow \mathbb{R} \\ (x, \lambda) &\longmapsto f(x) + \langle \lambda, h(x) \rangle_Y \end{aligned}$$

associated with the convex minimization problem (P). In fact, the (AH) differential system may be considered as a "generalized steepest descent method" acting on the convex-concave bifunction L which involves "steepest descent" in the first variable and "steepest ascent" in the second variable simultaneously. Moreover, it is well known that the zeros of the maximal monotone operator

$$(x, \lambda) \longmapsto (\nabla f(x) + h'(x)^* \lambda, -h(x)),$$

that is, the "generator" of the (AH) differential system, are precisely the saddle points of the Lagrangian L, cf. Rockafellar [28], which in turn are nothing but the solution pairs of (P) and its associated dual; we refer to Ekeland and Témam [19] for a general exposition on duality in convex optimization.

The previous discussion suggests that the inertial dynamics (ID) admit an inherent connection to the convex minimization problem (P) through the first-order differential system (AH). In this work, we explicitly characterize the minimizing properties of (ID) with respect to (P) without relying on the (AH) dynamics. As the differential nature of (ID) suggests, a particular preselection of its solutions' initial data will be essential to maintain the connection to the convex minimization problem (P). In this respect, it is interesting to note that the inertial dynamics (ID) neither involve any projection argument nor rely on any auxiliary variable to approach the linearly constrained convex minimization problem (P)—however, at the cost of second-order information on the objective function f.

"Heavy ball with friction" dynamics. The inertial dynamics (ID) also derive naturally from the "Heavy Ball with friction" differential system

$$\ddot{x} + \alpha\dot{x} + \nabla\|h(x)\|_Y^2/2 = 0_X \tag{HB}$$

associated with the convex potential function $\|h(\,\cdot\,)\|_Y^2/2$. The dynamics (HB) were first introduced, from a more general perspective, by Polyak [25, 26] and are known to inherit remarkable minimizing properties; cf. Alvarez [1] and Attouch, Goudou, and Redont [8] for an exposition involving general convex potential effects.

In the above (HB) model, the viscous damping operator $\alpha\,\mathrm{Id} : X \to X$, $\alpha > 0$, may render the system dissipative, but it acts isotropically on the velocity term $\dot{x}$ and neglects the geometry of the potential function $\|h(\,\cdot\,)\|_Y^2/2$. However, geometric damping effects may be crucial to diminish transversal oscillations, as observed by Alvarez et al. [2], or even to accelerate the convergence; see, e.g., Attouch et al. [7]. The importance of an anisotropic damping term in the (HB) model has already been noticed by Alvarez [1], who considered a generalized version of the differential system

$$\ddot{x} + \Gamma\dot{x} + \nabla\|h(x)\|_Y^2/2 = 0_X,$$

where $\Gamma : X \to X$ is an elliptic bounded self-adjoint linear operator; we also refer to Boţ and Csetnek [12] for a similar type of equation involving non-potential effects.

In this work, we naturally extend the (HB) model by introducing a general Hessian-driven damping term, thus leading to the inertial dynamics (ID). As a decisive feature of (ID), the damping operator $\nabla^2 f(\,\cdot\,) : X \to X$ is adapted to the objective function f of the convex minimization problem (P) rather than to the potential function $\|h(\,\cdot\,)\|_Y^2/2$ associated with the constraint function h. This particular and distinct feature allows us, as we shall see, to infer the minimizing properties of (ID) relative to the linearly constrained convex minimization problem (P).

Dynamics with variable metric. Finally, we remark that the inertial dynamics (ID) may also be seen as a second-order extension of the first-order differential system

$$\nabla^2 f(x)\dot{x} + \nabla\|h(x)\|_Y^2/2 = 0_X. \tag{SD}$$

It is customary to interpret (SD) as the Steepest Descent method associated with the convex potential function $\|h(\,\cdot\,)\|_Y^2/2$ and with variable metric induced by the Hessian of f. Indeed, assuming that $\nabla^2 f(\,\cdot\,)$ is non-degenerate, the (SD) model equivalently reads as

$$\dot{x} + \nabla_H\|h(x)\|_Y^2/2 = 0_X,$$

where $\nabla_H\|h(\,\cdot\,)\|_Y^2/2$ denotes the gradient of $\|h(\,\cdot\,)\|_Y^2/2$ relative to the inner product $\langle\nabla^2 f(\,\cdot\,)\,\cdot\,,\,\cdot\,\rangle_X$; we refer to Alvarez, Bolte, and Brahic [3] and Attouch et al. [5] for a general study of the steepest descent method with respect to Hessian–Riemannian metrics induced by Legendre-type convex functions.

In the (SD) model, the Hessian $\nabla^2 f(\,\cdot\,)$ may be degenerate, resulting in an ill-posed problem where (SD) is no longer defined as a differential system. Adding the second derivative $\ddot{x}$ to (SD) however causes a regularizing effect which renders the equation well-posed even in the case of a degenerate Hessian. In our work, we explicitly take the inertial effect induced by the acceleration term $\ddot{x}$ into account and further adhere to the notion of a variable metric relative to the Hessian $\nabla^2 f(\,\cdot\,)$ in (ID). Within this context, we highlight the role of the Bregman distance associated with f (cf. Bregman [13]) defined by

$$\begin{aligned} D_f : X \times X &\longrightarrow \mathbb{R} \\ (y, x) &\longmapsto f(y) - f(x) - \langle \nabla f(x), y - x \rangle_X, \end{aligned}$$

which implicitly captures the geometry of the objective function f of the convex minimization problem (P) and naturally relates it, as we shall see, to the Hessian $\nabla^2 f(\,\cdot\,)$ in (ID).

Presentation of the results. The analysis of the asymptotic behavior of the solutions of (ID) relies essentially on the decay property of the energy function

$$t \longmapsto \|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2.$$

In the case of a positive Hessian operator $\nabla^2 f(\,\cdot\,)$, we observe that energy may not be dissipated but conserved along the solutions of (ID); cf. Hale [20] and Haraux [21] for an exposition on dissipative systems. As a consequence, we infer that any non-stationary solution $x(t)$ of (ID) fails to converge as $t \to +\infty$, provided that

$$\dot{x}(t) \in \ker \nabla^2 f(x(t)), \quad \forall t \geq 0.$$

The above condition appears naturally as it reflects the vanishing damping effect of the Hessian $\nabla^2 f(\,\cdot\,)$ on the solutions of (ID). Despite the potential lack of damping, we show that any solution of (ID) converges in average towards a feasible point of (P), given that the potential function $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum with zero value.

Whenever the Hessian operator $\nabla^2 f(\,\cdot\,)$ is elliptic, we infer that energy is strictly dissipated as the solutions of (ID) evolve. In this case, it is proven that the energy function converges towards its infimal value. More precisely, we show that the convergence obeys, as $t \to +\infty$, the estimate

$$\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2 = o\Big(\frac{1}{t}\Big).$$

This result is in line with the worst-case estimate known for the "heavy ball with friction" differential system; see, e.g., Attouch and Cabot [6]. Under the ellipticity condition on $\nabla^2 f(\,\cdot\,)$, we also prove that any solution of (ID) converges towards a feasible point of (P), if one exists. In this case, the limit of a solution of (ID) is characterized as a "D_f-like projection" of its initial data onto the set of feasible points of (P); here, D_f refers to the Bregman distance associated with f. More precisely, given a solution $x : [0, +\infty) \to X$ of (ID) with initial data $(x_0, v_0) \in X \times X$, we show that $x(t)$ converges, as $t \to +\infty$, to the unique element $\bar{x} \in X$ satisfying

$$D_f(\bar{x}, x_0) - \langle \bar{x}, v_0 \rangle_X = \inf \{ D_f(x, x_0) - \langle x, v_0 \rangle_X \mid h(x) = 0_Y \}.$$

As an immediate consequence, we observe that whenever the solution is issued from $(x_0, -\nabla f(x_0)) \in X \times X$, the convergence is towards the unique element $\bar{x} \in X$ verifying

$$f(\bar{x}) = \inf \{f(x) \mid h(x) = 0_Y\},$$

that is, towards the unique minimizer of the linearly constrained convex minimization problem (P). Finally, we show that the solutions of (ID) decay asymptotically at an exponential rate provided that f satisfies

$$2D_f(y, x) - \langle \nabla^2 f(x)(x - y), x - y\rangle_X \geq 0, \quad \forall x, y \in X.$$

We note that the above condition essentially restricts f to be quadratic as it requires the graph of f to lie above all its second-order approximations.

Organization of the paper. We begin our discussion with a basic result on the existence and uniqueness of the solutions of (ID), and further characterize their asymptotic properties based on the decay information of the energy function. In section 3, we then investigate the asymptotic behavior of the solutions of (ID) under some more stringent assumptions on the problem data. In section 4, we establish the convergence of the solutions of (ID) and eventually characterize their minimizing properties with respect to the convex minimization problem (P). In section 5, we further derive asymptotic estimates on the convergence rate of the solutions of (ID). In section 6, we show that the inertial dynamics (ID) admit an equivalent first-order representation in terms of the Arrow–Hurwicz differential system (AH). Moreover, we introduce an augmented variant of the inertial dynamics (ID) to allow for a relaxation of the stringent assumptions imposed on the problem data. Finally, section 7 is devoted to numerical experiments.

2. General properties. Let X, Y be real Hilbert spaces endowed with inner products $\langle\cdot\,,\cdot\rangle_X$, $\langle\cdot\,,\cdot\rangle_Y$ and induced norms $\|\cdot\|_X$, $\|\cdot\|_Y$. Throughout the work, we presuppose that

(A1) $f : X \to \mathbb{R}$ is convex and twice continuously differentiable;
(A2) $\nabla^2 f : X \to \mathcal{B}(X)$ is Lipschitz continuous on bounded sets;
(A3) $A \in \mathcal{B}(X, Y)$, $b \in Y$ and $h(x) = Ax - b$, $x \in X$.

Given the above assumptions, consider the second-order differential system

$$\ddot{x} + \nabla^2 f(x)\dot{x} + \nabla\|h(x)\|_Y^2/2 = 0_X \tag{ID}$$

with initial data $(x_0, v_0) \in X \times X$.

Let us begin our discussion with a basic result on the existence and uniqueness of the solutions of (ID). Recall that $x : I \to X$ is a (classical) solution of (ID) on an interval $I \subset \mathbb{R}$, $0 \in I$, if $x \in \mathcal{C}^2(I; X)$ and x satisfies (ID) on I with $(x(0), \dot{x}(0)) = (x_0, v_0)$.

THEOREM 2.1. *For any $(x_0, v_0) \in X \times X$ there exists a unique solution $x : [0, +\infty) \to X$ of (ID). Moreover,*

(i) $t \mapsto \|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2$ is non-increasing on $[0, +\infty)$ and

$$\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 \leq \|v_0\|_X^2 + \|h(x_0)\|_Y^2, \quad \forall t \geq 0;$$

(ii) $\lim_{t\to+\infty}\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2$ exists;

(iii) it holds that

$$\int_0^\infty \langle \nabla^2 f(x(\tau))\dot{x}(\tau), \dot{x}(\tau)\rangle_X \,\mathrm{d}\tau < +\infty;$$

(iv) $\dot{x} \in \mathcal{L}^\infty([0,+\infty); X)$ *and* $h(x) \in \mathcal{L}^\infty([0,+\infty); Y)$.

Proof. By the Cauchy–Lipschitz theorem (see, e.g., [21]), for any $(x_0, v_0) \in X \times X$ there exists $\tilde{t} > 0$ such that (ID) admits a unique solution $x : [0, \tilde{t}] \to X$. Let

$$t_+ = \sup\{\tilde{t} > 0 \mid \exists! \text{ solution of (ID) on } [0, \tilde{t}]\}$$

and suppose, contrary to our claim, that $t_+ < +\infty$. For any $t \in [0, t_+)$, taking the inner product with $\dot{x}(t)$ in (ID) yields

$$\langle \ddot{x}(t) + \nabla \|h(x(t))\|_Y^2/2, \dot{x}(t)\rangle_X + \langle \nabla^2 f(x(t))\dot{x}(t), \dot{x}(t)\rangle_X = 0.$$

Using the chain rule, we obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2\big) + \langle \nabla^2 f(x(t))\dot{x}(t), \dot{x}(t)\rangle_X = 0. \tag{2.1}$$

Since $\nabla^2 f(\,\cdot\,)$ is positive (f being convex), it follows that $t \mapsto \|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2$ is non-increasing on $[0, t_+)$ and thus, for any $t \in [0, t_+)$,

$$\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 \le \|v_0\|_X^2 + \|h(x_0)\|_Y^2.$$

Clearly,

$$\sup_{t\in[0,t_+)} \|\dot{x}(t)\|_X < +\infty \quad \text{and} \quad \sup_{t\in[0,t_+)} \|h(x(t))\|_Y < +\infty.$$

Hence, for any $0 \le s \le t < t_+$, we deduce

$$\|x(t) - x(s)\|_X \le \int_s^t \|\dot{x}(\tau)\|_X \,\mathrm{d}\tau \le \sup_{\tau\in[s,t]} \|\dot{x}(\tau)\|_X (t-s) \le \sup_{\tau\in[0,t_+)} \|\dot{x}(\tau)\|_X (t-s),$$

so that $x : [0, t_+) \to X$ admits a unique continuous extension to $[0, t_+]$, that is, $\lim_{t\to t_+} x(t)$ exists. Consequently, x and $\dot{x}$ are bounded on $[0, t_+)$. Moreover, in view of the Lipschitz continuity of $\nabla^2 f$ and the boundedness of $\nabla\|h(x)\|_Y^2/2$ on $[0, t_+)$, we deduce from (ID) that $\ddot{x}$ remains bounded on $[0, t_+)$ as well. Hence, for any $0 \le s \le t < t_+$, we have

$$\|\dot{x}(t) - \dot{x}(s)\|_X \le \int_s^t \|\ddot{x}(\tau)\|_X \,\mathrm{d}\tau \le \sup_{\tau\in[0,t_+)} \|\ddot{x}(\tau)\|_X (t-s),$$

so that $\lim_{t\to t_+} \dot{x}(t)$ exists. Applying the Cauchy–Lipschitz theorem again with initial data $(\lim_{t\to t_+} x(t), \lim_{t\to t_+} \dot{x}(t))$ at $t_+ < +\infty$, we can extend the solution to an interval strictly larger than $[0, t_+)$, a contradiction. ∎

The previous theorem identifies the mapping

$$t \longmapsto \|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2$$

as a Lyapunov function whose decay property will be crucial for the analysis of the asymptotic behavior of the solutions of (ID). Assuming, moreover, the boundedness of the solutions of (ID), we have the following additional properties.

Corollary 2.2. *Let* $x : [0,+\infty) \to X$ *be a bounded solution of* (ID). *Then,*

(i) $\ddot{x}, \nabla^2 f(x)\dot{x}, \nabla\|h(x)\|_Y^2/2 \in \mathcal{L}^\infty([0,+\infty); X)$;

(ii) $\lim_{t\to+\infty} \langle \nabla^2 f(x(t))\dot{x}(t), \dot{x}(t)\rangle_X = 0$.

Proof. (i) By Theorem 2.1*(iv)*, we have both $\dot{x} \in \mathcal{L}^\infty([0,+\infty);X)$ as well as $h(x) \in \mathcal{L}^\infty([0,+\infty);Y)$. Consequently, $\nabla\|h(x)\|_Y^2/2$ belongs to $\mathcal{L}^\infty([0,+\infty);X)$. Since x is bounded on $[0,+\infty)$, it follows from the Lipschitz continuity of $\nabla^2 f$ and the boundedness of $\dot{x}$ on $[0,+\infty)$ that $\nabla^2 f(x)\dot{x}$ remains bounded on $[0,+\infty)$ as well; and so does $\ddot{x}$ in view of (ID).

(ii) By Theorem 2.1*(iii)*, we know that $\langle\nabla^2 f(x)\dot{x},\dot{x}\rangle_X \in \mathcal{L}^1([0,+\infty);\mathbb{R})$. Using $\ddot{x},\dot{x},x \in \mathcal{L}^\infty([0,+\infty);X)$ and the fact that $\nabla^2 f$ is Lipschitz continuous on bounded sets, we obtain $\langle\nabla^2 f(x)\dot{x},\dot{x}\rangle_X \in \mathrm{Lip}([0,+\infty);\mathbb{R})$. Observing that we have both

$$\langle\nabla^2 f(x)\dot{x},\dot{x}\rangle_X \in \mathcal{L}^1([0,+\infty);\mathbb{R}) \quad \text{and} \quad \langle\nabla^2 f(x)\dot{x},\dot{x}\rangle_X \in \mathrm{Lip}([0,+\infty);\mathbb{R}),$$

it follows from a classical result that $\lim_{t\to+\infty}\langle\nabla^2 f(x(t))\dot{x}(t),\dot{x}(t)\rangle_X = 0$. ∎

Remark 2.3. We note that the solutions of (ID) remain bounded, e.g., in the case when $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum at $\bar{x} \in X$, i.e.,

$$\exists\beta > 0\ \forall x \in X \quad \|h(x)\|_Y^2 \geq \inf\|h(\,\cdot\,)\|_Y^2 + \beta\|x-\bar{x}\|_X^2.$$

Indeed, it suffices to observe from Theorem 2.1*(i)* that for any $t \geq 0$,

$$\|h(x(t))\|_Y^2 \leq \|v_0\|_X^2 + \|h(x_0)\|_Y^2.$$

This majorization and the fact that $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum at $\bar{x} \in X$ clearly imply that x remains bounded on $[0,+\infty)$.

The following proposition characterizes the limit of a convergent solution of (ID) as a minimizer of $\|h(\,\cdot\,)\|_Y^2$. The technique we use to prove this fact is adapted from the corresponding literature on a second-order differential system with asymptotically small dissipation; cf. Cabot, Engler, and Gadat [17].

Proposition 2.4. *Let $x : [0,+\infty) \to X$ be a solution of* (ID) *and let $\bar{x} \in X$ be such that $x(t) \to \bar{x}$ strongly in X as $t \to +\infty$. Then,*

(i) it holds that

$$\lim_{t\to+\infty}\ddot{x}(t) = \lim_{t\to+\infty}\dot{x}(t) = 0_X \quad \textit{and} \quad \nabla\|h(\bar{x})\|_Y^2/2 = 0_X;$$

(ii) $\lim_{t\to+\infty}\|h(x(t))\|_Y^2 = \inf\|h(\,\cdot\,)\|_Y^2$.

Proof. (i) Let $\bar{x} \in X$ be such that $x(t) \to \bar{x}$ strongly in X as $t \to +\infty$. Consequently, x is bounded on $[0,+\infty)$ and the assertions of Corollary 2.2 hold. Hence, there exists $C \geq 0$ such that $\|\ddot{x}(t)\|_X \leq C$ for any $t \geq 0$. Applying Landau's inequality to $t \mapsto x(t) - \bar{x}$ yields for any $t \geq 0$,

$$\sup_{\tau\geq t}\|\dot{x}(\tau)\|_X \leq 2\sqrt{\sup_{\tau\geq t}\|x(\tau)-\bar{x}\|_X \sup_{\tau\geq t}\|\ddot{x}(\tau)\|_X} \leq 2\sqrt{C}\sqrt{\sup_{\tau\geq t}\|x(\tau)-\bar{x}\|_X}.$$

By taking the limit as $t \to +\infty$, it follows that $\lim_{t\to+\infty}\dot{x}(t) = 0_X$. Moreover, using the fact that $\nabla\|h(x(t))\|_Y^2/2 \to \nabla\|h(\bar{x})\|_Y^2/2$ strongly in X as $t \to +\infty$, we deduce that $\lim_{t\to+\infty}\ddot{x}(t) = -\nabla\|h(\bar{x})\|_Y^2/2$ in view of (ID) and the Lipschitz continuity of $\nabla^2 f$. Assuming now that $\nabla\|h(\bar{x})\|_Y^2/2 \neq 0_X$, an immediate integration shows that $\dot{x}(t)$ behaves equivalent to $-t\nabla\|h(\bar{x})\|_Y^2/2$ as $t \to +\infty$, a contradiction. Consequently, $\nabla\|h(\bar{x})\|_Y^2/2 = 0_X$ and thus, $\lim_{t\to+\infty}\ddot{x}(t) = 0_X$.

(ii) This is an immediate consequence of the convexity of $\|h(\,\cdot\,)\|_Y^2/2$. Indeed, for any $\eta \in X$, we have

$$\|h(\eta)\|_Y^2/2 \geq \|h(x(t))\|_Y^2/2 + \langle\nabla\|h(x(t))\|_Y^2/2, x(t)-\eta\rangle_X.$$

Since $x(t) \to \bar{x}$ and $\nabla \|h(x(t))\|_Y^2/2 \to 0_X$ strongly in X as $t \to +\infty$, it follows that

$$\|h(\eta)\|_Y^2 \geq \lim_{t\to+\infty} \|h(x(t))\|_Y^2 = \|h(\bar{x})\|_Y^2.$$

This inequality being true for any $\eta \in X$, we deduce

$$\lim_{t\to+\infty} \|h(x(t))\|_Y^2 = \inf \|h(\,\cdot\,)\|_Y^2 = \|h(\bar{x})\|_Y^2,$$

concluding the result. ∎

As a direct consequence of the previous result, we observe that any non-stationary solution of (ID) fails to converge under the condition

$$\dot{x}(t) \in \ker \nabla^2 f(x(t)), \quad \forall t \geq 0.$$

The assertion comes with no surprise, since under the above condition, the solutions of (ID) evolve solely in "regions of vanishing damping"; cf. the extreme case when $\nabla^2 f(\,\cdot\,)$ is identically zero. For simplicity of notation, we now set $S := \{x \in X \mid \nabla \|h(x)\|_Y^2/2 = 0_X\}$.

Corollary 2.5. *Let $x : [0, +\infty) \to X$ be a solution of (ID) with initial data $(x_0, v_0) \in X \times X$ and suppose that*

$$\dot{x}(t) \in \ker \nabla^2 f(x(t)), \quad \forall t \geq 0.$$

If $(x_0, v_0) \notin S \times \{0_X\}$, then $x(t)$ fails to converge as $t \to +\infty$.

Proof. Let $(x_0, v_0) \notin S \times \{0_X\}$ so that $\|v_0\|_X^2 + \|h(x_0)\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2 > 0$. For any $t \geq 0$, integration of (2.1) over $[0, t]$ yields

$$\begin{aligned}\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2 + 2\int_0^t \langle \nabla^2 f(x(\tau))\dot{x}(\tau), \dot{x}(\tau)\rangle_X \,\mathrm{d}\tau \\ = \|v_0\|_X^2 + \|h(x_0)\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2.\end{aligned}$$

Owing to the fact that, for any $t \geq 0$,

$$\dot{x}(t) \in \ker \nabla^2 f(x(t)),$$

by taking the limit in the above equality as $t \to +\infty$, it follows

$$\begin{aligned}\lim_{t\to+\infty} \|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2 \\ = \|v_0\|_X^2 + \|h(x_0)\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2 > 0.\end{aligned} \tag{2.2}$$

Suppose now, contrary to our claim, that there exists $\bar{x} \in X$ such that $x(t) \to \bar{x}$ strongly in X as $t \to +\infty$. By Proposition 2.4, we know that $\lim_{t\to+\infty} \dot{x}(t) = 0_X$ and $\lim_{t\to+\infty} \|h(x(t))\|_Y^2 = \inf \|h(\,\cdot\,)\|_Y^2$. Consequently,

$$\lim_{t\to+\infty} \|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2 = 0,$$

which contradicts (2.2). ∎

Let us now complement the previous discussion with a basic result on the ergodic convergence of the solutions of (ID). To this end, consider the Cesàro average $\sigma : (0, +\infty) \to X$ of a solution $x : [0, +\infty) \to X$ of (ID) defined by

$$\sigma(t) = \frac{1}{t}\int_0^t x(\tau)\,\mathrm{d}\tau.$$

The following result states that the Cesàro average $\sigma(t)$ of a solution of (ID) converges, as $t \to +\infty$, whenever $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum; cf. Remark 2.3. Clearly, the latter condition implies that S is non-empty and reduced to a single element.

PROPOSITION 2.6. *Let $\sigma : (0, +\infty) \to X$ be the Cesàro average of a solution of* (ID) *and suppose that $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum at $\bar{x} \in X$. Then, as $t \to +\infty$, it holds that*

$$\|\sigma(t) - \bar{x}\|_X = \mathcal{O}\Big(\frac{1}{t}\Big).$$

Consequently, $\sigma(t)$ converges strongly, as $t \to +\infty$, to the unique element $\bar{x} \in S$.

Proof. Since $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum at $\bar{x} \in S$, we have for any $t > 0$,

$$\beta\|\sigma(t) - \bar{x}\|_X^2 \leq \|h(\sigma(t))\|_Y^2 - \inf\|h(\,\cdot\,)\|_Y^2.$$

Moreover, using the fact that $\nabla\|h(\bar{x})\|_Y^2/2 = 0_X$ and the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}\|h(\sigma(t))\|_Y^2 - \inf\|h(\,\cdot\,)\|_Y^2 &= \langle \nabla\|h(\sigma(t))\|_Y^2/2, \sigma(t) - \bar{x}\rangle_X \\ &\leq \|\nabla\|h(\sigma(t))\|_Y^2/2\|_X\, \|\sigma(t) - \bar{x}\|_X.\end{aligned}$$

Combining the above inequalities yields

$$\beta\|\sigma(t) - \bar{x}\|_X \leq \|\nabla\|h(\sigma(t))\|_Y^2/2\|_X.$$

Equivalently,

$$t\|\sigma(t) - \bar{x}\|_X \leq \frac{1}{\beta}\Big\|\int_0^t \nabla\|h(x(\tau))\|_Y^2/2\,\mathrm{d}\tau\Big\|_X,$$

which, in view of (ID) and subsequent integration over $[0, t]$, reads

$$t\|\sigma(t) - \bar{x}\|_X \leq \frac{1}{\beta}\|\dot{x}(t) - v_0 + \nabla f(x(t)) - \nabla f(x_0)\|_X.$$

By Theorem 2.1*(iv)*, we know that $\dot{x}$ is bounded on $[0, +\infty)$. Since x is bounded on $[0, +\infty)$, cf. Remark 2.3, it follows from the Lipschitz continuity of $\nabla^2 f$ that $\nabla f(x)$ remains bounded on $[0, +\infty)$. Hence, there exists $C \geq 0$ such that for any $t > 0$,

$$t\|\sigma(t) - \bar{x}\|_X \leq C.$$

Passing to the upper limit as $t \to +\infty$ then yields

$$\limsup_{t\to+\infty} t\|\sigma(t) - \bar{x}\|_X < +\infty,$$

concluding the proof. ∎

As an immediate consequence of the previous result, we observe that whenever $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum at $\bar{x} \in X$ with zero value, i.e.,

$$\exists\beta > 0\ \forall x \in X \quad \|h(x)\|_Y^2 \geq \beta\|x - \bar{x}\|_X^2,$$

the Cesàro average $\sigma(t)$ of a solution of (ID) converges, as $t \to +\infty$, to the unique feasible (and thus optimal) point $\bar{x} \in S$ of the convex minimization problem (P).

3. The case of an elliptic Hessian. In this section, we investigate the asymptotic properties of the solutions of (ID) under some more stringent assumptions on the problem data. We take for granted the existence and uniqueness of the solutions of (ID) and suppose henceforth that

(A4) $\nabla^2 f(\,\cdot\,) : X \to X$ is α-elliptic, i.e.,

$$\exists \alpha > 0 \; \forall x, y \in X \quad \langle \nabla^2 f(x)y, y\rangle_X \geq \alpha \|y\|_X^2,$$

(equivalently, $f : X \to \mathbb{R}$ is α-strongly convex);

(A5) $S = \{x \in X \mid \nabla \|h(x)\|_Y^2/2 = 0_X\}$ is non-empty.

We remark that the latter assumption is verified in each of the following cases: (i) If $\operatorname{ran} A$ is closed, then $S = \xi + \ker A$ with $\xi \in X$ such that $A\xi = \operatorname{proj}_{\operatorname{ran} A} b$. (ii) If $b \in \operatorname{ran} A$, then $S = \xi + \ker A$ with $\xi \in X$ satisfying $A\xi = b$.

In our forthcoming study on (ID), we utilize the notion of the Bregman distance relative to the Legendre-type function f. Recall that the Bregman distance associated with f, denoted by $D_f : X \times X \to \mathbb{R}$, is defined by

$$D_f(y, x) = f(y) - f(x) - \langle \nabla f(x), y - x\rangle_X.$$

Since f is α-strongly convex ($\nabla^2 f(\,\cdot\,)$ being α-elliptic), it provides a natural measure of proximity in the sense that, for any $x, y \in X$, we have $D_f(y, x) \geq \alpha\|x - y\|_X^2/2$ and thus, $D_f(y, x) = 0$ if and only if $x = y$.

The following theorem characterizes the asymptotic properties of the solutions of (ID) and further provides an estimate on the decay of the mapping

$$t \longmapsto \|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2/2.$$

This decay property will be crucial for the subsequent convergence analysis on the solutions of (ID).

THEOREM 3.1. *Let $\nabla^2 f(\,\cdot\,)$ be α-elliptic, and let S be non-empty. Then, any solution $x : [0, +\infty) \to X$ of (ID) is bounded on $[0, +\infty)$. Moreover,*

(i) $\ddot{x}, \dot{x}, \nabla\|h(x)\|_Y^2/2 \in \mathcal{L}^2([0, +\infty); X)$*;*

(ii) it holds that

$$\lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \nabla\|h(x(t))\|_Y^2/2 = 0_X;$$

(iii) $\lim_{t\to+\infty} \|h(x(t))\|_Y^2 = \inf \|h(\,\cdot\,)\|_Y^2$*;*

(iv) it holds that

$$\int_0^\infty \|\dot{x}(\tau)\|_X^2 + \|h(x(\tau))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2 \,\mathrm{d}\tau < +\infty;$$

(v) $\lim_{t\to+\infty} t\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2\big) = 0$ *and thus, as* $t \to +\infty$,

$$\|\dot{x}(t)\|_X^2 = o\Big(\frac{1}{t}\Big) \quad \textit{and} \quad \|h(x(t))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2 = o\Big(\frac{1}{t}\Big).$$

Proof. Let $\xi \in S$ and define $\phi : [0, +\infty) \to \mathbb{R}$ by $\phi(t) = \|x(t) - \xi\|_X^2/2$ such that $\dot{\phi}(t) = \langle x(t) - \xi, \dot{x}(t)\rangle_X$ and $\ddot{\phi}(t) = \langle x(t) - \xi, \ddot{x}(t)\rangle_X + \|\dot{x}(t)\|_X^2$. Using (ID) and the fact that $\nabla\|h(\xi)\|_Y^2/2 = 0_X$, we have for any $t \geq 0$,

$$\ddot{\phi}(t) + \frac{\mathrm{d}}{\mathrm{d}t} D_f(\xi, x(t)) + \langle \nabla\|h(x(t))\|_Y^2/2 - \nabla\|h(\xi)\|_Y^2/2, x(t) - \xi\rangle_X = \|\dot{x}(t)\|_X^2.$$

Since $\nabla^2 f(\,\cdot\,)$ is α-elliptic, we observe from (2.1) that

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf\|h(\,\cdot\,)\|_Y^2\big) + \alpha\|\dot{x}(t)\|_X^2 \leq 0,$$

which, together with the above equation, gives

$$\begin{aligned}\ddot{\phi}(t) + \frac{\mathrm{d}}{\mathrm{d}t}D_f(\xi, x(t)) + \frac{1}{\alpha}\frac{\mathrm{d}}{\mathrm{d}t}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf\|h(\,\cdot\,)\|_Y^2\big)\\ + \|\dot{x}(t)\|_X^2 + \langle\nabla\|h(x(t))\|_Y^2/2 - \nabla\|h(\xi)\|_Y^2/2, x(t)-\xi\rangle_X \leq 0.\end{aligned} \tag{3.1}$$

Using this inequality and the monotonicity of $\nabla\|h(\,\cdot\,)\|_Y^2/2$, i.e.,

$$\langle\nabla\|h(x(t))\|_Y^2/2 - \nabla\|h(\xi)\|_Y^2/2, x(t)-\xi\rangle_X \geq 0,$$

we deduce that

$$t \longmapsto \dot{\phi}(t) + D_f(\xi, x(t)) + \frac{1}{\alpha}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf\|h(\,\cdot\,)\|_Y^2\big) \tag{3.2}$$

is non-increasing on $[0,+\infty)$. Hence, there exists $C \geq 0$ such that for any $t \geq 0$,

$$\dot{\phi}(t) + D_f(\xi, x(t)) + \frac{1}{\alpha}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf\|h(\,\cdot\,)\|_Y^2\big) \leq C.$$

Using successively $D_f(\xi, x(t)) \geq \alpha\phi(t)$ and the fact that

$$\alpha\phi(t) - \alpha\phi(0) = \alpha\int_0^t \dot{\phi}(\tau)\,\mathrm{d}\tau,$$

we obtain

$$\dot{\phi}(t) + \alpha\int_0^t \dot{\phi}(\tau)\,\mathrm{d}\tau \leq C - \alpha\phi(0).$$

Multiplying the previous inequality by $\mathrm{e}^{\alpha t}$ and subsequently integrating over $[0,t]$ shows that there exists $\tilde{C} > C$ such that

$$\phi(t) \leq \tilde{C}/\alpha,$$

which clearly implies that x remains bounded on $[0,+\infty)$.

(i) Recall the definition of h in terms of $A \in \mathcal{B}(X,Y)$ and $b \in Y$ (cf. assumption (A3)), and let $\|A\|$ denote the operator norm of A. Since $\nabla\|h(\,\cdot\,)\|_Y^2/2$ is $\|A\|^2$-Lipschitz continuous, it follows by the Baillon–Haddad theorem [9] that $\nabla\|h(\,\cdot\,)\|_Y^2/2$ is $1/\|A\|^2$-cocoercive, i.e.,

$$\langle\nabla\|h(x(t))\|_Y^2/2 - \nabla\|h(\xi)\|_Y^2/2, x(t)-\xi\rangle_X \geq \frac{1}{\|A\|^2}\|\nabla\|h(x(t))\|_Y^2/2 - \nabla\|h(\xi)\|_Y^2/2\|_X^2.$$

From (3.1) and the fact that $\nabla\|h(\xi)\|_Y^2/2 = 0_X$, we then obtain

$$\begin{aligned}\ddot{\phi}(t) + \frac{\mathrm{d}}{\mathrm{d}t}D_f(\xi, x(t)) + \frac{1}{\alpha}\frac{\mathrm{d}}{\mathrm{d}t}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf\|h(\,\cdot\,)\|_Y^2\big)\\ + \|\dot{x}(t)\|_X^2 + \frac{1}{\|A\|^2}\|\nabla\|h(x(t))\|_Y^2/2\|_X^2 \leq 0.\end{aligned}$$

Integration over $[0,t]$ yields

$$\begin{aligned}\dot\phi(t)+D_f(\xi,x(t))+\frac{1}{\alpha}\big(\|\dot x(t)\|_X^2+\|h(x(t))\|_Y^2-\inf\|h(\,\cdot\,)\|_Y^2\big)\\ +\int_0^t\|\dot x(\tau)\|_X^2\,\mathrm{d}\tau+\frac{1}{\|A\|^2}\int_0^t\|\nabla\|h(x(\tau))\|_Y^2/2\|_X^2\,\mathrm{d}\tau\le C.\end{aligned}$$

Using again $D_f(\xi,x(t))\ge\alpha\phi(t)$ and the fact that x and $\dot x$ are bounded on $[0,+\infty)$, we observe that $\dot\phi$ remains bounded on $[0,+\infty)$ as well. Hence,

$$\int_0^\infty\|\dot x(\tau)\|_X^2\,\mathrm{d}\tau<+\infty\quad\text{and}\quad\int_0^\infty\|\nabla\|h(x(\tau))\|_Y^2/2\|_X^2\,\mathrm{d}\tau<+\infty.$$

In view of (ID), the latter implies that $\ddot x+\nabla^2 f(x)\dot x\in\mathcal{L}^2([0,+\infty);X)$. Owing to the Lipschitz continuity of $\nabla^2 f$, it follows that $\ddot x=(\ddot x+\nabla^2 f(x)\dot x)-\nabla^2 f(x)\dot x$ belongs to $\mathcal{L}^2([0,+\infty);X)$ as well.

(ii) Since x is bounded on $[0,+\infty)$, the assertions of Corollary 2.2 hold. Hence, we have both

$$\dot x\in\mathcal{L}^2([0,+\infty);X)\quad\text{and}\quad\ddot x\in\mathcal{L}^\infty([0,+\infty);X),$$

which, according to a classical result, imply $\lim_{t\to+\infty}\dot x(t)=0_X$. Similarly, we have

$$\nabla\|h(x)\|_Y^2/2\in\mathcal{L}^2([0,+\infty);X)\quad\text{and}\quad\nabla\|h(x)\|_Y^2/2\in\mathrm{Lip}([0,+\infty);X),$$

so that $\lim_{t\to+\infty}\nabla\|h(x(t))\|_Y^2/2=0_X$. Using (ID) and the Lipschitz continuity of $\nabla^2 f$, we conclude that $\lim_{t\to+\infty}\ddot x(t)=0_X$.

(iii) This again follows immediately from the convexity of $\|h(\,\cdot\,)\|_Y^2/2$; cf. Proposition 2.4*(ii)*. Recall that for any $\eta\in X$, we have

$$\|h(\eta)\|_Y^2/2\ge\|h(x(t))\|_Y^2/2+\langle\nabla\|h(x(t))\|_Y^2/2,x(t)-\eta\rangle_X.$$

Since x is bounded on $[0,+\infty)$ and $\nabla\|h(x(t))\|_Y^2/2\to0_X$ strongly in X as $t\to+\infty$, it follows

$$\|h(\eta)\|_Y^2\ge\limsup_{t\to+\infty}\|h(x(t))\|_Y^2\ge\liminf_{t\to+\infty}\|h(x(t))\|_Y^2\ge\inf\|h(\,\cdot\,)\|_Y^2.$$

The above inequalities being true for any $\eta\in X$, we obtain the result.

(iv) From (3.1) and the fact that for any $t\ge0$,

$$\langle\nabla\|h(x(t))\|_Y^2/2-\nabla\|h(\xi)\|_Y^2/2,x(t)-\xi\rangle_X=\|h(x(t))\|_Y^2-\inf\|h(\,\cdot\,)\|_Y^2,$$

we obtain

$$\begin{aligned}\ddot\phi(t)+\frac{\mathrm{d}}{\mathrm{d}t}D_f(\xi,x(t))+\frac{1}{\alpha}\frac{\mathrm{d}}{\mathrm{d}t}\big(\|\dot x(t)\|_X^2+\|h(x(t))\|_Y^2-\inf\|h(\,\cdot\,)\|_Y^2\big)\\ +\|\dot x(t)\|_X^2+\|h(x(t))\|_Y^2-\inf\|h(\,\cdot\,)\|_Y^2\le0.\end{aligned}\tag{3.3}$$

Using a similar reasoning as above, there exists $C\ge0$ such that for any $t\ge0$,

$$\begin{aligned}\dot\phi(t)+\alpha\int_0^t\dot\phi(\tau)\,\mathrm{d}\tau+\frac{1}{\alpha}\big(\|\dot x(t)\|_X^2+\|h(x(t))\|_Y^2-\inf\|h(\,\cdot\,)\|_Y^2\big)\\ +\int_0^t\|\dot x(\tau)\|_X^2+\|h(x(\tau))\|_Y^2-\inf\|h(\,\cdot\,)\|_Y^2\,\mathrm{d}\tau\le C-\alpha\phi(0).\end{aligned}$$

Successively multiplying the above inequality by $\mathrm{e}^{\alpha t}$ and integrating over $[0,t]$ yields the existence of $\tilde{C} > C$ such that

$$\phi(t) + \frac{1}{\alpha}\int_0^t \|\dot{x}(\tau)\|_X^2 + \|h(x(\tau))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2 \,\mathrm{d}\tau \le \tilde{C}/\alpha,$$

which implies that $\|\dot{x}\|_X^2 + \|h(x)\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2 \in \mathcal{L}^1([0,+\infty);\mathbb{R})$.

(v) By Theorem 2.1*(i)*, the mapping $t \mapsto \|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2$ is non-increasing on $[0,+\infty)$ and thus, for any $t \ge 0$,

$$\int_{t/2}^t \|\dot{x}(\tau)\|_X^2 + \|h(x(\tau))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2 \,\mathrm{d}\tau \ge \frac{t}{2}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2\big).$$

Since $\|\dot{x}\|_X^2 + \|h(x)\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2$ belongs to $\mathcal{L}^1([0,+\infty);\mathbb{R})$, we observe that the above integral vanishes as $t \to +\infty$. Consequently,

$$\lim_{t\to+\infty} t\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2\big) = 0.$$

The remaining assertions now follow immediately. ∎

As an immediate consequence of Theorem 3.1*(ii)*, we recover the fact that the limit of a convergent solution of (ID) necessarily belongs to S; cf. Proposition 2.4*(i)*.

Corollary 3.2. *Under the hypotheses of Theorem 3.1, if there exists $\bar{x} \in X$ such that $x(t) \to \bar{x}$ strongly in X as $t \to +\infty$, then $\bar{x} \in S$.*

Proof. Since $\nabla\|h(\,\cdot\,)\|_Y^2/2$ is continuous and $\nabla\|h(x(t))\|_Y^2/2 \to 0_X$ strongly in X as $t \to +\infty$, the limit of $x(t)$ as $t \to +\infty$ clearly belongs to S. ∎

Let us now extend the previous discussion with a result on the convergence of the solutions of (ID) in the case when $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum. Recall that $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum at $\bar{x} \in X$ if

$$\exists \beta > 0 \;\forall x \in X \quad \|h(x)\|_Y^2 \ge \inf \|h(\,\cdot\,)\|_Y^2 + \beta\|x - \bar{x}\|_X^2.$$

In this case, the minimizing property in Theorem 3.1*(iii)* clearly implies that any solution $x(t)$ of (ID) converges, as $t \to +\infty$, to the unique element $\bar{x} \in S$. The following proposition further provides an asymptotic estimate on the decay of the solutions of (ID) based on the decay property stated in Theorem 3.1*(v)*.

Proposition 3.3. *Let $\nabla^2 f(\,\cdot\,)$ be α-elliptic and suppose that $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum at $\bar{x} \in X$. Let $x : [0,+\infty) \to X$ be a solution of* (ID)*. Then, as $t \to +\infty$, it holds that*

$$\|\dot{x}(t)\|_X^2 = o\Big(\frac{1}{t}\Big) \quad \textit{and} \quad \|x(t) - \bar{x}\|_X^2 = o\Big(\frac{1}{t}\Big).$$

Consequently, $x(t)$ converges strongly, as $t \to +\infty$, to the unique element $\bar{x} \in S$.

Proof. Since $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum at $\bar{x} \in S$, we have for any $t \ge 0$,

$$\|\dot{x}(t)\|_X^2 + \beta\|x(t) - \bar{x}\|_X^2 \le \|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2.$$

Multiplying the inequality by t and using that $\lim_{t\to+\infty} t\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf \|h(\,\cdot\,)\|_Y^2\big) = 0$, cf. Theorem 3.1*(v)*, we deduce

$$\lim_{t\to+\infty} t\big(\|\dot{x}(t)\|_X^2 + \beta\|x(t) - \bar{x}\|_X^2\big) = 0.$$

The remaining estimates are now readily deduced. ∎

Remark 3.4. We note that the above decay rate estimates remain valid even in the case when $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum with respect to the set S, i.e.,

$$\exists\beta > 0\ \forall x \in X \quad \|h(x)\|_Y^2 \geq \inf\|h(\,\cdot\,)\|_Y^2 + \beta \operatorname{dist}(x, S)^2.$$

Indeed, by following the above reasoning, we observe that any solution $x(t)$ of (ID) obeys, as $t \to +\infty$, the estimates

$$\|\dot{x}(t)\|_X^2 = o\Big(\frac{1}{t}\Big) \quad \text{and} \quad \operatorname{dist}(x(t), S)^2 = o\Big(\frac{1}{t}\Big).$$

In this case, however, we can not deduce the convergence of the solutions of (ID); we refer to section 4 for the respective results corresponding to the case when $\|h(\,\cdot\,)\|_Y^2$ admits multiple minima.

Finally, we observe from Proposition 3.3 that whenever $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum at $\bar{x} \in X$ with zero value, i.e.,

$$\exists\beta > 0\ \forall x \in X \quad \|h(x)\|_Y^2 \geq \beta\|x - \bar{x}\|_X^2,$$

any solution $x(t)$ of (ID) converges, as $t \to +\infty$, to the unique feasible (and thus optimal) point $\bar{x} \in S$ of the convex minimization problem (P).

4. Asymptotic analysis. In this section, we investigate the asymptotic properties of the solutions of (ID) in the case when $\|h(\,\cdot\,)\|_Y^2$ admits multiple minima. We focus our attention on the convergence of the solutions of (ID) in both the weak and the strong senses, and eventually characterize their minimizing properties with respect to the convex minimization problem (P).

4.1. Weak and strong convergence. Let us begin our discussion with a result on the weak convergence of the solutions of (ID). The argument we use to prove the following theorem relies essentially on Opial-like techniques; cf. Opial [24].

THEOREM 4.1. *Let $\nabla^2 f(\,\cdot\,)$ be α-elliptic, let S be non-empty and let $x : [0, +\infty) \to X$ be a solution of* (ID)*. Then, the following assertions hold:*
(i) $\forall \xi \in S$, $\lim_{t\to+\infty} D_f(\xi, x(t))$ *exists;*
(ii) $\forall t_n \to +\infty$ *with* $x(t_n) \rightharpoonup \bar{x}$ *weakly in* X*, it holds that* $\bar{x} \in S$.
Assuming moreover that ∇f is weakly sequentially continuous, then there exists $\bar{x} \in S$ such that $x(t) \rightharpoonup \bar{x}$ weakly in X as $t \to +\infty$.

Proof. *(i)* Let $\xi \in S$ and consider again the mapping $\phi : [0, +\infty) \to \mathbb{R}$ defined by $\phi(t) = \|x(t) - \xi\|_X^2/2$. Recall from (3.2) that

$$t \longmapsto \dot{\phi}(t) + D_f(\xi, x(t)) + \frac{1}{\alpha}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf\|h(\,\cdot\,)\|_Y^2\big)$$

is non-increasing on $[0, +\infty)$. Hence, it admits a limit as $t \to +\infty$. By Theorem 2.1*(ii)*, we know that $\lim_{t\to+\infty}\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2$ exists. Moreover, by Theorem 3.1, x is bounded on $[0, +\infty)$ and $\lim_{t\to+\infty}\dot{x}(t) = 0_X$. Consequently, $\lim_{t\to+\infty}\dot{\phi}(t) = 0$ and thus, $\lim_{t\to+\infty} D_f(\xi, x(t))$ exists.

(ii) Let $\bar{x} \in X$ and let $t_n \to +\infty$ be a sequence such that $x(t_n) \rightharpoonup \bar{x}$ weakly in X. By Theorem 3.1*(iii)*, we have $\lim_{n\to+\infty}\|h(x(t_n))\|_Y^2 = \inf\|h(\,\cdot\,)\|_Y^2$. Since $\|h(\,\cdot\,)\|_Y^2$ is weakly lower-semicontinuous ($\|h(\,\cdot\,)\|_Y^2$ being convex and continuous), it follows that

$$\inf\|h(\,\cdot\,)\|_Y^2 = \lim_{n\to+\infty}\|h(x(t_n))\|_Y^2 = \liminf_{n\to+\infty}\|h(x(t_n))\|_Y^2 \geq \|h(\bar{x})\|_Y^2,$$

which implies that $\bar{x} \in S$.

Let us assume now that ∇f is weakly sequentially continuous. To establish the weak convergence of $x(t)$ as $t \to +\infty$, it suffices to show that it admits a unique weak sequential cluster point. Let $x(t_n) \rightharpoonup \bar{y}$ and $x(s_n) \rightharpoonup \bar{z}$ weakly in X for some sequences $t_n, s_n \to +\infty$. Using (ii), we observe that $\bar{y}$ and $\bar{z}$ belong to S. Moreover, using (i), $\lim_{t\to+\infty} D_f(\bar{y}, x(t))$ and $\lim_{t\to+\infty} D_f(\bar{z}, x(t))$ exist. Hence,

$$\lim_{t\to+\infty} D_f(\bar{y}, x(t)) - D_f(\bar{z}, x(t)) \text{ exists.}$$

Replacing t successively by t_n and s_n gives

$$\lim_{n\to+\infty} D_f(\bar{y}, x(t_n)) - D_f(\bar{z}, x(t_n)) = \lim_{n\to+\infty} D_f(\bar{y}, x(s_n)) - D_f(\bar{z}, x(s_n)).$$

Since ∇f is weakly sequentially continuous, we deduce

$$\begin{aligned} &\lim_{n\to+\infty} D_f(\bar{y}, x(t_n)) - D_f(\bar{z}, x(t_n)) = -D_f(\bar{z}, \bar{y}), \text{ and} \\ &\lim_{n\to+\infty} D_f(\bar{y}, x(s_n)) - D_f(\bar{z}, x(s_n)) = D_f(\bar{y}, \bar{z}). \end{aligned}$$

Using that f is α-strongly convex, it follows

$$\alpha\|\bar{y} - \bar{z}\|_X^2/2 \le D_f(\bar{y}, \bar{z}) = -D_f(\bar{z}, \bar{y}) \le -\alpha\|\bar{y} - \bar{z}\|_X^2/2$$

and thus, $\bar{y} = \bar{z}$. ∎

The previous result establishes the weak convergence of the solutions of (ID) under the additional assumption that $\nabla f : X \to X$ is weakly sequentially continuous. This condition comes with no surprise since, in general, nonlinear sequentially continuous mappings are not weakly sequentially continuous; cf. Brézis [15]. We also remark that a similar result has been established by Bauschke, Borwein, and Combettes [10, Condition 4.3 and Theorem 4.11] based on the notion of Bregman monotonicity.

Remark 4.2. We note that $\nabla f : X \to X$ is weakly sequentially continuous, e.g., in the case when X is finite-dimensional or whenever ∇f is a continuous affine operator; cf. Bauschke and Combettes [11, Lemma 2.41].

Let us now turn our attention to the question on whether the solutions of (ID) are strongly convergent. The following theorem gives an affirmative answer to this question by exploiting the symmetry property of $\|h(\,\cdot\,)\|_Y^2$ relative to the set S, i.e., for any fixed $\xi \in S$, it holds that

$$\|h(x)\|_Y^2 = \|h(\xi - x + \xi)\|_Y^2, \quad \forall x \in X.$$

The above argument was in essence originated by Bruck [16] for the "steepest descent method"; see also Alvarez [1] for an extension to a second-order differential system.

THEOREM 4.3. *Let $\nabla^2 f(\,\cdot\,)$ be α-elliptic, let S be non-empty and let $x : [0, +\infty) \to X$ be a solution of* (ID)*. Then, there exists $\bar{x} \in S$ such that $x(t) \to \bar{x}$ strongly in X as $t \to +\infty$.*

Proof. Let $\xi \in S$, let $\tilde{t} > 0$, and define $\psi : [0, \tilde{t}] \to \mathbb{R}$ by $\psi(t) = \|x(t) - \xi + x(\tilde{t}) - \xi\|_X^2/2$ such that $\dot{\psi}(t) = \langle x(t) - \xi + x(\tilde{t}) - \xi, \dot{x}(t)\rangle_X$ and $\ddot{\psi}(t) = \langle x(t) - \xi + x(\tilde{t}) - \xi, \ddot{x}(t)\rangle_X + \|\dot{x}(t)\|_X^2$. Using (ID), we have for any $t \in [0, \tilde{t}]$,

$$\begin{aligned} \ddot{\psi}(t) &+ \frac{\mathrm{d}}{\mathrm{d}t}\big(2D_f(\xi, x(t)) - 2D_f(\xi, x(\tilde{t})) - D_f(x(\tilde{t}), x(t))\big) \\ &+ \langle \nabla\|h(x(t))\|_Y^2/2, x(t) - \xi + x(\tilde{t}) - \xi\rangle_X = \|\dot{x}(t)\|_X^2. \end{aligned}$$

Since $\nabla^2 f(\,\cdot\,)$ is α-elliptic, we observe again from (2.1) that

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf\|h(\,\cdot\,)\|_Y^2\big) + \alpha\|\dot{x}(t)\|_X^2 \le 0.$$

This fact, together with the above equation, yields

$$\begin{aligned}\ddot{\psi}(t) &+ \frac{\mathrm{d}}{\mathrm{d}t}\big(2D_f(\xi,x(t)) - 2D_f(\xi,x(\tilde{t})) - D_f(x(\tilde{t}),x(t))\big)\\ &+ \frac{1}{\alpha}\frac{\mathrm{d}}{\mathrm{d}t}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf\|h(\,\cdot\,)\|_Y^2\big)\\ &\qquad + \|\dot{x}(t)\|_X^2 + \langle\nabla\|h(x(t))\|_Y^2/2, x(t)-\xi+x(\tilde{t})-\xi\rangle_X \le 0.\end{aligned}$$

Using this inequality and the convexity of $\|h(\,\cdot\,)\|_Y^2/2$, i.e.,

$$\|h(x(t))\|_Y^2/2 - \|h(\xi - x(\tilde{t})+\xi)\|_Y^2/2 \le \langle\nabla\|h(x(t))\|_Y^2/2, x(t)-\xi+x(\tilde{t})-\xi\rangle_X,$$

we obtain

$$\begin{aligned}\ddot{\psi}(t) &+ \frac{\mathrm{d}}{\mathrm{d}t}\big(2D_f(\xi,x(t)) - 2D_f(\xi,x(\tilde{t})) - D_f(x(\tilde{t}),x(t))\big)\\ &+ \frac{1}{\alpha}\frac{\mathrm{d}}{\mathrm{d}t}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf\|h(\,\cdot\,)\|_Y^2\big)\\ &\qquad + \|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2/2 - \|h(\xi - x(\tilde{t})+\xi)\|_Y^2/2 \le 0.\end{aligned}$$

Moreover, using successively that $t \mapsto \|\dot{x}(t)\|_X^2/2 + \|h(x(t))\|_Y^2/2$ is non-increasing on $[0,\tilde{t}]$, cf. Theorem 2.1*(i)*, and the fact that $\|h(\,\cdot\,)\|_Y^2/2$ satisfies the symmetry property relative to S, we have for any $t \in [0,\tilde{t}]$,

$$\begin{aligned}\|\dot{x}(t)\|_X^2/2 + \|h(x(t))\|_Y^2/2 &\ge \|\dot{x}(\tilde{t})\|_X^2/2 + \|h(x(\tilde{t}))\|_Y^2/2\\ &= \|\dot{x}(\tilde{t})\|_X^2/2 + \|h(\xi - x(\tilde{t})+\xi)\|_Y^2/2.\end{aligned}$$

Consequently,

$$\begin{aligned}\ddot{\psi}(t) &+ \frac{\mathrm{d}}{\mathrm{d}t}\big(2D_f(\xi,x(t)) - 2D_f(\xi,x(\tilde{t})) - D_f(x(\tilde{t}),x(t))\big)\\ &+ \frac{1}{\alpha}\frac{\mathrm{d}}{\mathrm{d}t}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2 - \inf\|h(\,\cdot\,)\|_Y^2\big) \le 0.\end{aligned}$$

Integration over $[t,\tilde{t}]$ yields

$$\begin{aligned}\dot{\psi}(\tilde{t}) - \dot{\psi}(t) &- \big(2D_f(\xi,x(t)) - 2D_f(\xi,x(\tilde{t})) - D_f(x(\tilde{t}),x(t))\big)\\ &+ \frac{1}{\alpha}\big(\|\dot{x}(\tilde{t})\|_X^2 + \|h(x(\tilde{t}))\|_Y^2\big) - \frac{1}{\alpha}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2\big) \le 0.\end{aligned}$$

Since f is α-strongly convex, we have $D_f(x(\tilde{t}),x(t)) \ge \alpha\|x(t)-x(\tilde{t})\|_X^2/2$ such that for any $t \in [0,\tilde{t}]$,

$$\begin{aligned}\frac{\alpha}{2}\|x(t)-x(\tilde{t})\|_X^2 &\le \dot{\psi}(t) - \dot{\psi}(\tilde{t}) + 2D_f(\xi,x(t)) - 2D_f(\xi,x(\tilde{t}))\\ &\qquad + \frac{1}{\alpha}\big(\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2\big) - \frac{1}{\alpha}\big(\|\dot{x}(\tilde{t})\|_X^2 + \|h(x(\tilde{t}))\|_Y^2\big).\end{aligned}$$

By Theorem 2.1*(ii)*, we know that $\lim_{t\to+\infty}\|\dot{x}(t)\|_X^2 + \|h(x(t))\|_Y^2$ exists. Moreover, Theorem 4.1*(i)* asserts that for any $\xi \in S$, $\lim_{t\to+\infty} D_f(\xi,x(t))$ exists. Since x is bounded on $[0,+\infty)$ and $\lim_{t\to+\infty}\dot{x}(t) = 0_X$ (cf. Theorem 3.1), we further have $\lim_{t\to+\infty}\dot{\psi}(t) = 0$. Hence, we deduce from the above inequality that

$$\lim_{t,\tilde{t}\to+\infty,\, t<\tilde{t}} \frac{\alpha}{2}\|x(t)-x(\tilde{t})\|_X^2 = 0,$$

so that the Cauchy criteria at infinity is satisfied. Consequently, $x(t)$ converges as $t\to+\infty$ and, by Corollary 3.2, the limit belongs to S. ∎

The above result complements our discussion in Remark 3.4 on the convergence of the solutions of (ID) corresponding to the case when $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum with respect to the set S, i.e.,

$$\exists\beta>0\ \forall x\in X\quad \|h(x)\|_Y^2 \geq \inf\|h(\,\cdot\,)\|_Y^2 + \beta\,\mathrm{dist}(x,S)^2.$$

In this case, however, it is not clear a priori to which point in S the solutions of (ID) converge and, in particular, whether the convergence is towards a minimizer of the convex minimization problem (P). This will be the subject of our investigation in the following subsection.

4.2. Localization of the limit. Let us now provide a localization result of the limit of the solutions of (ID). The following theorem characterizes the limit of a solution of (ID) in terms of a "D_f-like projection" of its initial data onto the closed affine subspace S.

THEOREM 4.4. *Let $\nabla^2 f(\,\cdot\,)$ be α-elliptic, let S be non-empty and let $x:[0,+\infty)\to X$ be a solution of* (ID) *with initial data $(x_0,v_0)\in X\times X$. Then, $x(t)$ converges strongly, as $t\to+\infty$, to the unique element $\bar{x}\in S$ satisfying*

$$D_f(\bar{x},x_0) - \langle\bar{x},v_0\rangle_X = \inf\nolimits_S D_f(\,\cdot\,,x_0) - \langle\,\cdot\,,v_0\rangle_X.$$

Proof. Let $\bar{x}\in S$ be such that $x(t)\to\bar{x}$ strongly in X as $t\to+\infty$ and let $\xi\in S$ be arbitrary. Using (ID) and the fact that $\nabla\|h(\bar{x})\|_Y^2/2 = \nabla\|h(\xi)\|_Y^2/2 = 0_X$, we have for any $t\geq 0$,

$$\begin{aligned}
\langle\ddot{x}(t)+\nabla^2 f(x(t))\dot{x}(t),\bar{x}-\xi\rangle_X &= -\langle\nabla\|h(x(t))\|_Y^2/2-\nabla\|h(\xi)\|_Y^2/2,\bar{x}-\xi\rangle_X\\
&= -\langle x(t)-\xi,\nabla\|h(\bar{x})\|_Y^2/2-\nabla\|h(\xi)\|_Y^2/2\rangle_X\\
&= 0.
\end{aligned}$$

Integration over $[0,t]$ yields

$$\langle\dot{x}(t)-v_0,\bar{x}-\xi\rangle_X + \langle\nabla f(x(t))-\nabla f(x_0),\bar{x}-\xi\rangle_X = 0.$$

Since $x(t)\to\bar{x}$ and $\dot{x}(t)\to 0_X$ strongly in X as $t\to+\infty$, we obtain

$$\langle\nabla f(\bar{x})-\nabla f(x_0),\bar{x}-\xi\rangle_X = \langle v_0,\bar{x}-\xi\rangle_X.$$

Using the Bregman three-points-identity [18, Lemma 3.1], it follows that

$$D_f(\xi,\bar{x}) + D_f(\bar{x},x_0) = D_f(\xi,x_0) + \langle v_0,\bar{x}-\xi\rangle_X, \tag{4.1}$$

and thus,

$$D_f(\bar{x},x_0) - \langle\bar{x},v_0\rangle_X \leq D_f(\xi,x_0) - \langle\xi,v_0\rangle_X.$$

This inequality being true for any $\xi\in S$, we deduce that

$$D_f(\bar{x},x_0) - \langle\bar{x},v_0\rangle_X = \inf\nolimits_S D_f(\,\cdot\,,x_0) - \langle\,\cdot\,,v_0\rangle_X.$$

Noticing that $D_f(\,\cdot\,,x_0) - \langle\,\cdot\,,v_0\rangle_X$ is α-strongly convex, we conclude the result. ∎

As a direct consequence of the previous result, we have the following localization estimate of the limit of a solution of (ID) given in terms of the "initial deflection" of the Bregman distance D_f with its quadratic lower bound.

COROLLARY 4.5. *Under the hypotheses of Theorem 4.4, let $\bar{x} \in S$ be such that $x(t) \to \bar{x}$ strongly in X as $t \to +\infty$. Then,*

$$\frac{\alpha}{2}\|\bar{x} - \mathrm{proj}_S(x_0 + \frac{1}{\alpha}v_0)\|_X^2 \le D_f(\mathrm{proj}_S(x_0 + \frac{1}{\alpha}v_0), x_0) - \frac{\alpha}{2}\|x_0 - \mathrm{proj}_S(x_0 + \frac{1}{\alpha}v_0)\|_X^2.$$

Proof. Let $\bar{x} \in S$ be such that $x(t) \to \bar{x}$ strongly in X as $t \to +\infty$ and let $\xi \in S$ be arbitrary. From (4.1) and the fact that f is α-strongly convex, we deduce that

$$\frac{\alpha}{2}\|\bar{x} - \xi\|_X^2 + \frac{\alpha}{2}\|x_0 - \bar{x}\|_X^2 - \alpha\langle \bar{x} - \xi, \frac{1}{\alpha}v_0\rangle_X \le D_f(\xi, x_0),$$

which, in view of a simple expansion, reads

$$\frac{\alpha}{2}\|\bar{x} - \xi\|_X^2 + \frac{\alpha}{2}\|\bar{x} - (x_0 + \frac{1}{\alpha}v_0)\|_X^2 - \frac{\alpha}{2}\|\xi - (x_0 + \frac{1}{\alpha}v_0)\|_X^2 \le D_f(\xi, x_0) - \frac{\alpha}{2}\|x_0 - \xi\|_X^2.$$

Choosing $\xi = \mathrm{proj}_S(x_0 + v_0/\alpha) \in S$ then gives the desired inequality. $\square$

Remark 4.6. In the particular case when $\nabla^2 f(\,\cdot\,) = \alpha\,\mathrm{Id}$, we infer that any solution $x(t)$ of (ID) converges, as $t \to +\infty$, to the unique element $\bar{x} \in S$ satisfying

$$\frac{\alpha}{2}\|\bar{x} - (x_0 + \frac{1}{\alpha}v_0)\|_X^2 = \inf_S \frac{\alpha}{2}\|\,\cdot\, - (x_0 + \frac{1}{\alpha}v_0)\|_X^2.$$

Indeed, it suffices to observe that the Bregman distance associated with f coincides with its quadratic lower bound, i.e.,

$$D_f(\mathrm{proj}_S(x_0 + \frac{1}{\alpha}v_0), x_0) - \frac{\alpha}{2}\|x_0 - \mathrm{proj}_S(x_0 + \frac{1}{\alpha}v_0)\|_X^2 = 0.$$

This fact, together with Corollary 4.5, clearly implies that $\bar{x} = \mathrm{proj}_S(x_0 + v_0/\alpha)$. We note that a similar characterization has yet been obtained by Alvarez [1, Proposition 2.5] in the study of "heavy ball with friction" differential system; see also Lemaire [22, Corollary 2.2] for the respective result on the "steepest descent method".

Finally, our next result characterizes the minimizing properties of the solutions of (ID) relative to a specific preselection of their initial data.

COROLLARY 4.7. *Under the hypotheses of Theorem 4.4, let $\bar{x} \in S$ be such that $x(t) \to \bar{x}$ strongly in X as $t \to +\infty$. Then, the following assertions hold:*

(i) If $v_0 = 0_X$, then $\bar{x} \in S$ is the unique element satisfying

$$D_f(\bar{x}, x_0) = \inf_S D_f(\,\cdot\,, x_0);$$

(ii) If $v_0 + \nabla f(x_0) = 0_X$, then $\bar{x} \in S$ is the unique element satisfying

$$f(\bar{x}) = \inf_S f(\,\cdot\,).$$

The proof is an immediate consequence of Theorem 4.4 and is left to the reader.

By virtue of Corollary 4.7*(ii)*, we observe that whenever $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum with respect to the set S and with zero value, i.e.,

$$\exists \beta > 0\ \forall x \in X \quad \|h(x)\|_Y^2 \ge \beta\,\mathrm{dist}(x, S)^2,$$

any solution $x(t)$ of (ID) with corresponding initial data $(x_0, -\nabla f(x_0)) \in X \times X$ converges, as $t \to +\infty$, to the unique minimizer of the convex minimization problem (P).

5. Exponential decay rate estimates. In this section, we further provide asymptotic estimates on the convergence rate of the solutions of (ID) in the restrictive but important case when f is quadratic. In particular, our analysis relies on the additional assumptions that

(A6) $f : X \to \mathbb{R}$ satisfies condition (C), i.e.,

$$2D_f(y,x) - \langle \nabla^2 f(x)(x-y), x-y\rangle_X \geq 0, \quad \forall x,y \in X;$$

(A7) $\nabla^2 f(\,\cdot\,) : X \to X$ is γ-bounded, i.e.,

$$\exists \gamma > 0 \;\forall x,y \in X \quad \langle \nabla^2 f(x)y, y\rangle_X \leq \gamma \|y\|_X^2.$$

We remark that condition (C) is verified whenever $f : X \to \mathbb{R}$ is minorized by its second-order Taylor approximations; cf. the case when f is a quadratic form. Since $\nabla^2 f(\,\cdot\,)$ is α-elliptic and γ-bounded, we further have for any $x, y \in X$,

$$\alpha\|x-y\|_X^2/2 \leq D_f(y,x) \leq \gamma\|x-y\|_X^2/2.$$

Given the above assumptions, the following theorem characterizes the exponential decay properties of the solutions of (ID) whenever $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum.

THEOREM 5.1. *Let $\nabla^2 f(\,\cdot\,)$ be α-elliptic and γ-bounded, and suppose that $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum at $\bar{x} \in X$ with constant β. Let f satisfy condition* (C) *and set*

$$\rho = \begin{cases} \alpha/2, & \text{if } \gamma^2 \leq 4\beta, \\ \min\{\alpha, \gamma - \sqrt{\gamma^2 - 4\beta}\}/2, & \text{if } \gamma^2 > 4\beta. \end{cases}$$

Let $x : [0,+\infty) \to X$ be a solution of (ID). *Then, the following assertions hold:*

(i) If $\rho^2 - \gamma\rho + \beta > 0$, then, as $t \to +\infty$, it holds that

$$\|\dot{x}(t)\|_X^2 = \mathcal{O}(\mathrm{e}^{-2\rho t}) \quad \text{and} \quad \|x(t)-\bar{x}\|_X^2 = \mathcal{O}(\mathrm{e}^{-2\rho t});$$

(ii) If $\rho^2 - \gamma\rho + \beta = 0$, then, as $t \to +\infty$, it holds that

$$\|\dot{x}(t)\|_X^2 = \mathcal{O}(t^2\mathrm{e}^{-2\rho t}) \quad \text{and} \quad \|x(t)-\bar{x}\|_X^2 = \mathcal{O}(t^2\mathrm{e}^{-2\rho t}).$$

Proof. Let $\bar{x} \in S$ and define $\vartheta : [0,+\infty) \to \mathbb{R}$ by $\vartheta(t) = \|\dot{x}(t) + \rho(x(t)-\bar{x})\|_X^2/2$ for some $\rho > 0$ (to be chosen) such that $\dot{\vartheta}(t) = \langle \dot{x}(t) + \rho(x(t)-\bar{x}), \ddot{x}(t) + \rho\dot{x}(t)\rangle_X$. Consider again $\phi : [0,+\infty) \to \mathbb{R}$ defined by $\phi(t) = \|x(t)-\bar{x}\|_X^2/2$. Using (ID) and the fact that $\nabla\|h(\bar{x})\|_Y^2/2 = 0_X$, we have for any $t \geq 0$,

$$\begin{aligned} &\dot{\vartheta}(t) + \rho^2\dot{\phi}(t) + \langle(\nabla^2 f(x(t)) - 2\rho\,\mathrm{Id})\dot{x}(t), \dot{x}(t) + \rho(x(t)-\bar{x})\rangle_X + \rho\|\dot{x}(t)\|_X^2 \\ &\quad + \langle \nabla\|h(x(t))\|_Y^2/2 - \nabla\|h(\bar{x})\|_Y^2/2, \dot{x}(t) + \rho(x(t)-\bar{x})\rangle_X = 0. \end{aligned}$$

Using successively the chain rule and the fact that

$$\langle \nabla\|h(x(t))\|_Y^2/2 - \nabla\|h(\bar{x})\|_Y^2/2, x(t)-\bar{x}\rangle_X = \|h(x(t))\|_Y^2 - \inf\|h(\,\cdot\,)\|_Y^2,$$

we obtain

$$\begin{aligned} &\dot{\vartheta}(t) + \rho^2\dot{\phi}(t) + \frac{\mathrm{d}}{\mathrm{d}t}\big(\|h(x(t))\|_Y^2/2 - \inf\|h(\,\cdot\,)\|_Y^2/2\big) \\ &\quad + 2\rho\big(\|\dot{x}(t)\|_X^2/2 + \|h(x(t))\|_Y^2/2 - \inf\|h(\,\cdot\,)\|_Y^2/2\big) \\ &\quad + \langle(\nabla^2 f(x(t)) - 2\rho\,\mathrm{Id})\dot{x}(t), \dot{x}(t) + \rho(x(t)-\bar{x})\rangle_X = 0. \end{aligned}$$

Developing the above expression yields

$$\begin{aligned}\dot\vartheta(t)+\rho^2\dot\phi(t)+\frac{\mathrm{d}}{\mathrm{d}t}\big(\|h(x(t))\|_Y^2/2-\inf\|h(\,\cdot\,)\|_Y^2/2-\rho D_f(\bar x,x(t))\big)\\ +2\rho\big(\vartheta(t)+\rho^2\phi(t)+\|h(x(t))\|_Y^2/2-\inf\|h(\,\cdot\,)\|_Y^2/2-\rho D_f(\bar x,x(t))\big)\\ +\langle(\nabla^2 f(x(t))-2\rho\,\mathrm{Id})(\dot x(t)+\rho(x(t)-\bar x)),\dot x(t)+\rho(x(t)-\bar x)\rangle_X\\ +\rho^2\big(2D_f(\bar x,x(t))-\langle\nabla^2 f(x(t))(x(t)-\bar x),x(t)-\bar x\rangle_X\big)=0.\end{aligned}$$

Since $\nabla^2 f(\,\cdot\,)$ is α-elliptic, we have

$$\langle(\nabla^2 f(x(t))-2\rho\,\mathrm{Id})(\dot x(t)+\rho(x(t)-\bar x)),\dot x(t)+\rho(x(t)-\bar x)\rangle_X\geq 2(\alpha-2\rho)\vartheta(t),$$

which, together with the above equation, gives

$$\begin{aligned}\dot\vartheta(t)+\rho^2\dot\phi(t)+\frac{\mathrm{d}}{\mathrm{d}t}\big(\|h(x(t))\|_Y^2/2-\inf\|h(\,\cdot\,)\|_Y^2/2-\rho D_f(\bar x,x(t))\big)\\ +2\rho\big(\vartheta(t)+\rho^2\phi(t)+\|h(x(t))\|_Y^2/2-\inf\|h(\,\cdot\,)\|_Y^2/2-\rho D_f(\bar x,x(t))\big)\\ +2(\alpha-2\rho)\vartheta(t)+\rho^2\big(2D_f(\bar x,x(t))-\langle\nabla^2 f(x(t))(x(t)-\bar x),x(t)-\bar x\rangle_X\big)\leq 0.\end{aligned}$$

An immediate integration over $[0,t]$ shows that there exists $C\geq 0$ such that

$$\begin{aligned}\vartheta(t)+\rho^2\phi(t)+\|h(x(t))\|_Y^2/2-\inf\|h(\,\cdot\,)\|_Y^2/2-\rho D_f(\bar x,x(t))\\ +2(\alpha-2\rho)\int_0^t \mathrm{e}^{-2\rho(t-\tau)}\vartheta(\tau)\,\mathrm{d}\tau+\rho^2\int_0^t \mathrm{e}^{-2\rho(t-\tau)}\kappa(\tau)\,\mathrm{d}\tau\leq C\mathrm{e}^{-2\rho t},\end{aligned}$$

where

$$\kappa(t)=2D_f(\bar x,x(t))-\langle\nabla^2 f(x(t))(x(t)-\bar x),x(t)-\bar x\rangle_X.$$

Using that $\nabla^2 f(\,\cdot\,)$ is γ-bounded and the fact that $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum at $\bar x\in S$ with constant β, we have $D_f(\bar x,x(t))\leq\gamma\phi(t)$ and $\|h(x(t))\|_Y^2/2-\inf\|h(\,\cdot\,)\|_Y^2/2\geq\beta\phi(t)$ such that for any $t\geq 0$,

$$\begin{aligned}\vartheta(t)+(\rho^2-\gamma\rho+\beta)\phi(t)+2(\alpha-2\rho)\int_0^t \mathrm{e}^{-2\rho(t-\tau)}\vartheta(\tau)\,\mathrm{d}\tau\\ +\rho^2\int_0^t \mathrm{e}^{-2\rho(t-\tau)}\kappa(\tau)\,\mathrm{d}\tau\leq C\mathrm{e}^{-2\rho t}.\end{aligned}\tag{5.1}$$

Let us now determine the largest value for $\rho\in(0,\alpha/2]$ such that $\rho^2-\gamma\rho+\beta\geq 0$. Clearly, if $\gamma^2\leq 4\beta$, then $\rho^2-\gamma\rho+\beta\geq 0$ holds for any $\rho>0$. On the other hand, if $\gamma^2>4\beta$, then $\rho^2-\gamma\rho+\beta\geq 0$ is attained whenever $\rho\leq\gamma/2-\sqrt{\gamma^2-4\beta}/2$. Consequently, we may take

$$\rho=\begin{cases}\alpha/2, & \text{if } \gamma^2\leq 4\beta,\\ \min\{\alpha,\gamma-\sqrt{\gamma^2-4\beta}\}/2, & \text{if } \gamma^2>4\beta.\end{cases}$$

We have either one of the following cases:

(i) Suppose that $\rho^2-\gamma\rho+\beta>0$. In this case, we deduce from (5.1) and the fact that f satisfies condition (C) that for any $t\geq 0$,

$$\mathrm{e}^{2\rho t}\vartheta(t)\leq C\quad\text{and}\quad \mathrm{e}^{2\rho t}\phi(t)\leq\frac{C}{\rho^2-\gamma\rho+\beta}.$$

Passing to the upper limit as $t \to +\infty$ yields

$$\limsup_{t\to+\infty} \mathrm{e}^{2\rho t}\vartheta(t) < +\infty \quad \text{and} \quad \limsup_{t\to+\infty} \mathrm{e}^{2\rho t}\phi(t) < +\infty.$$

Moreover, in view of the triangle inequality, we have

$$\mathrm{e}^{2\rho t}\|\dot{x}(t)\|_X^2 \le 4\mathrm{e}^{2\rho t}\vartheta(t) + 4\rho^2\mathrm{e}^{2\rho t}\phi(t)$$

and thus,

$$\limsup_{t\to+\infty} \mathrm{e}^{2\rho t}\|\dot{x}(t)\|_X^2 < +\infty.$$

(ii) Suppose now that $\rho^2 - \gamma\rho + \beta = 0$. In this case, we observe from (5.1) and the fact that f satisfies condition (C) that for any $t \ge 0$,

$$\mathrm{e}^{2\rho t}\vartheta(t) \le C.$$

Using this inequality together with the fact that

$$\sqrt{\mathrm{e}^{2\rho t}\phi(t)} \le \sqrt{\phi(0)} + \int_0^t \sqrt{\mathrm{e}^{2\rho\tau}\vartheta(\tau)}\,\mathrm{d}\tau,$$

we obtain

$$\sqrt{\mathrm{e}^{2\rho t}\phi(t)} \le \sqrt{C}t + \sqrt{\phi(0)}.$$

Taking the square and multiplying the resulting inequality by t^{-2} yields

$$t^{-2}\mathrm{e}^{2\rho t}\phi(t) \le C + 2\sqrt{C\phi(0)}t^{-1} + \phi(0)t^{-2}.$$

This majorization being valid for any $t > 0$, we deduce

$$\limsup_{t\to+\infty} \mathrm{e}^{2\rho t}\vartheta(t) < +\infty \quad \text{and} \quad \limsup_{t\to+\infty} t^{-2}\mathrm{e}^{2\rho t}\phi(t) < +\infty,$$

and thus,

$$\limsup_{t\to+\infty} t^{-2}\mathrm{e}^{2\rho t}\|\dot{x}(t)\|_X^2 < +\infty,$$

concluding the result. ∎

Remark 5.2. We note that the above decay rate estimates remain asymptotically correct whenever f satisfies condition (C) locally around $\bar{x} \in X$, i.e.,

$$\exists\delta > 0\ \forall x \in X \cap B_\delta(\bar{x}) \quad 2D_f(\bar{x}, x) - \langle\nabla^2 f(x)(x - \bar{x}), x - \bar{x}\rangle_X \ge 0.$$

However, to verify this assumption, one requires a priori knowledge of the unique minimizer $\bar{x} \in X$ of $\|h(\,\cdot\,)\|_Y^2$. We leave the details to the reader.

Remark 5.3. Let us further note that, in view of Theorem 5.1, whenever $\rho \in (0, \alpha/2]$ is chosen such that $\rho^2 - \gamma\rho + \beta \ge 0$, we have

$$\int_0^t \mathrm{e}^{-2\rho(t-\tau)}\kappa(\tau)\,\mathrm{d}\tau \le \frac{C}{\rho^2}\mathrm{e}^{-2\rho t}.$$

This estimate suggests that $\kappa(t)$ converges to zero in average, as $t \to +\infty$, at an exponential rate. In particular, it holds that

$$\int_0^\infty \mathrm{e}^{2\rho\tau}\kappa(\tau)\,\mathrm{d}\tau < +\infty.$$

The above result complements our discussion in Proposition 3.3 with exponential decay rate estimates on the solutions of (ID) in the case when f satisfies condition (C). Assuming moreover that $\nabla^2 f(\,\cdot\,) = \alpha\,\mathrm{Id}$, we have the following refined asymptotic decay properties.

Corollary 5.4. *Let $\nabla^2 f(\,\cdot\,) = \alpha\,\mathrm{Id}$ and suppose that $\|h(\,\cdot\,)\|_Y^2$ admits a strong minimum at $\bar{x} \in X$ with constant β. Let $x : [0, +\infty) \to X$ be a solution of (ID). Then, the following assertions hold:*

(i) If $\alpha^2 < 4\beta$, then, as $t \to +\infty$, it holds that

$$\|\dot{x}(t)\|_X^2 = \mathcal{O}(\mathrm{e}^{-\alpha t}) \quad \text{and} \quad \|x(t) - \bar{x}\|_X^2 = \mathcal{O}(\mathrm{e}^{-\alpha t});$$

(ii) If $\alpha^2 = 4\beta$, then, as $t \to +\infty$, it holds that

$$\|\dot{x}(t)\|_X^2 = \mathcal{O}(t^2\mathrm{e}^{-\alpha t}) \quad \text{and} \quad \|x(t) - \bar{x}\|_X^2 = \mathcal{O}(t^2\mathrm{e}^{-\alpha t});$$

(iii) If $\alpha^2 > 4\beta$, then, as $t \to +\infty$, it holds that

$$\|\dot{x}(t)\|_X^2 = \mathcal{O}(\mathrm{e}^{-(\alpha-\delta)t}) \quad \text{and} \quad \|x(t) - \bar{x}\|_X^2 = \mathcal{O}(\mathrm{e}^{-(\alpha-\delta)t}),$$

where $\delta = \sqrt{\alpha^2 - 4\beta}$.

Proof. (i)–(ii) This is an immediate consequence of Theorem 5.1*(i)–(ii)*.

(iii) Suppose that $\alpha^2 > 4\beta$ and let $\rho = (\alpha + \delta)/2$, where $\delta = \sqrt{\alpha^2 - 4\beta}$, so that $\rho^2 - \alpha\rho + \beta = 0$. From (5.1) and the fact that $\nabla^2 f(\,\cdot\,) = \alpha\,\mathrm{Id}$, we infer that there exists $C \geq 0$ such that for any $t \geq 0$,

$$\mathrm{e}^{(\alpha+\delta)t}\vartheta(t) \leq C + 2\delta \int_0^t \mathrm{e}^{(\alpha+\delta)\tau}\vartheta(\tau)\,\mathrm{d}\tau.$$

Applying Gronwall's inequality to $t \mapsto \mathrm{e}^{(\alpha+\delta)t}\vartheta(t)$ yields

$$\mathrm{e}^{(\alpha+\delta)t}\vartheta(t) \leq C\mathrm{e}^{2\delta t}.$$

Using again this inequality together with the fact that

$$\sqrt{\mathrm{e}^{(\alpha+\delta)t}\phi(t)} \leq \sqrt{\phi(0)} + \int_0^t \sqrt{\mathrm{e}^{(\alpha+\delta)\tau}\vartheta(\tau)}\,\mathrm{d}\tau,$$

we obtain

$$\sqrt{\mathrm{e}^{(\alpha+\delta)t}\phi(t)} \leq \frac{\sqrt{C}}{\delta}\mathrm{e}^{\delta t} + \sqrt{\phi(0)} - \frac{\sqrt{C}}{\delta}$$

and thus,

$$\mathrm{e}^{(\alpha-\delta)t}\phi(t) \leq \frac{C}{\delta^2} + 2\frac{\sqrt{C}}{\delta}\Big(\sqrt{\phi(0)} - \frac{\sqrt{C}}{\delta}\Big)\mathrm{e}^{-\delta t} + \Big(\sqrt{\phi(0)} - \frac{\sqrt{C}}{\delta}\Big)^2\mathrm{e}^{-2\delta t}.$$

Passing to the upper limit as $t \to +\infty$ yields the desired estimates. ∎

The previous result essentially recovers the optimal decay rate estimates known for the classical damped harmonic oscillator. In particular, we observe from Corollary 5.4 that the decay properties of the solutions of (ID) may be categorized into *(i)* the "underdamped case", *(ii)* the "critically damped case" and *(iii)* the "overdamped case". We refer to section 7 for a graphical illustration of the decay rate estimates.

6. Further extensions. In this section, we discuss further extensions on the inertial dynamics (ID) with the aim to weaken some of the previously stated assumptions. We first show that the inertial dynamics (ID) can be equivalently expressed as a first-order differential system with no occurrence of the Hessian. We then further introduce an augmented variant of (ID) which allows for a relaxation of the ellipticity condition imposed on the Hessian of f.

6.1. First-order representation. Let us show that the inertial dynamics (ID) admit an equivalent first-order representation in terms of the Arrow–Hurwicz differential system

$$\text{(AH)} \qquad \begin{cases} \dot{x} + \nabla f(x) + h'(x)^* \lambda = 0_X \\ \dot{\lambda} - h(x) = 0_Y \end{cases}$$

with initial data $(x_0, \lambda_0) \in X \times Y$. To this end, recall that $(x, \lambda) : [0, +\infty) \to X \times Y$ is a (classical) solution of (AH) if $(x, \lambda) \in \mathcal{C}^1([0, +\infty); X \times Y)$ and (x, λ) satisfies (AH) on $[0, +\infty)$ with $(x(0), \lambda(0)) = (x_0, \lambda_0)$.

Proposition 6.1. *Let $h'(\,\cdot\,) : X \to Y$ be surjective, and let $(x_0, v_0) \in X \times X$ and $\lambda_0 \in Y$ satisfy $v_0 + \nabla f(x_0) + h'(x_0)^* \lambda_0 = 0_X$. Then, the following assertions are equivalent:*

(i) $x : [0, +\infty) \to X$ is a solution of (ID) *with initial data (x_0, v_0);*

(ii) $\exists \lambda : [0, +\infty) \to Y$ such that $(x, \lambda) : [0, +\infty) \to X \times Y$ is a solution of (AH) *with initial data (x_0, λ_0).*

Proof. *(i) $\Longrightarrow$ (ii)* Let $x : [0, +\infty) \to X$ be a solution of (ID) with initial data $(x_0, v_0) \in X \times X$ and define $\lambda : [0, +\infty) \to Y$ such that for any $t \geq 0$,

$$\dot{x}(t) + \nabla f(x(t)) + h'(x(t))^* \lambda(t) = 0_X.$$

Observing that λ belongs to $\mathcal{C}^1([0, +\infty); Y)$, we deduce that

$$\ddot{x}(t) + \nabla^2 f(x(t)) \dot{x}(t) + h'(x(t))^* \dot{\lambda}(t) = 0_X,$$

which, in view of (ID), yields

$$h'(x(t))^* \dot{\lambda}(t) - h'(x(t))^* h(x(t)) = 0_X.$$

Since $h'(\,\cdot\,)$ is surjective, it follows that

$$\dot{\lambda}(t) - h(x(t)) = 0_Y.$$

Moreover, from $v_0 + \nabla f(x_0) + h'(x_0)^* \lambda_0 = 0_X$ and the fact that $v_0 + \nabla f(x_0) + h'(x_0)^* \lambda(0) = 0_X$, we obtain $\lambda(0) = \lambda_0$. Consequently, $(x, \lambda) : [0, +\infty) \to X \times Y$ is a solution of (AH) with initial data (x_0, λ_0).

(ii) $\Longrightarrow$ (i) Conversely, let $(x, \lambda) : [0, +\infty) \to X \times Y$ be a solution of (AH) with initial data $(x_0, \lambda_0) \in X \times Y$. Since f is twice continuously differentiable, we observe that $\dot{x} = -\nabla f(x) - h'(x)^* \lambda$ belongs to $\mathcal{C}^1([0, +\infty); X)$. Differentiation yields, for any $t \geq 0$,

$$\ddot{x}(t) + \nabla^2 f(x(t)) \dot{x}(t) + h'(x(t))^* \dot{\lambda}(t) = 0_X.$$

By taking (AH) into account, we obtain

$$\ddot{x}(t) + \nabla^2 f(x(t)) \dot{x}(t) + \nabla \|h(x(t))\|_Y^2 / 2 = 0_X.$$

Using again $v_0 + \nabla f(x_0) + h'(x_0)^* \lambda_0 = 0_X$ and the fact that $\dot{x}(0) + \nabla f(x_0) + h'(x_0)^* \lambda_0 = 0_X$, we conclude that $\dot{x}(0) = v_0$ and thus, $x : [0, +\infty) \to X$ is a solution of (ID) with initial data (x_0, v_0). ∎

The previous result provides conditions under which the inertial dynamics (ID) prove to be equivalent to the (AH) differential system. Despite being of first-order, the (AH) model is particularly favorable as it does not incorporate any second-order information on f. As such, it may still provide a meaning to (ID) even in the case when f is not twice continuously differentiable. Our next result asserts that the solutions of (AH) indeed enjoy the same basic properties as the solutions of (ID) given the weakened assumptions that

(A1)′ $f : X \to \mathbb{R}$ is convex and continuously differentiable;
(A2)′ $\nabla f : X \to X$ is Lipschitz continuous on bounded sets.

To ease the exposition, let $X \times Y$ be equipped with the Hilbertian product structure $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_X + \langle \cdot, \cdot \rangle_Y$ and induced norm $\| \cdot \|$.

THEOREM 6.2. *For any* $(x_0, \lambda_0) \in X \times Y$ *there exists a unique solution* $(x, \lambda) : [0, +\infty) \to X \times Y$ *of* (AH). *Moreover,*

(i) $t \mapsto \|(\dot{x}(t), \dot{\lambda}(t))\|$ *is non-increasing on* $[0, +\infty)$ *and*

$$\|(\dot{x}(t), \dot{\lambda}(t))\| \le \|(\nabla f(x_0) + h'(x_0)^* \lambda_0, -h(x_0))\|, \quad \forall t \ge 0;$$

(ii) $\lim_{t\to+\infty} \|(\dot{x}(t), \dot{\lambda}(t))\|$ *exists;*
(iii) $(\dot{x}, \dot{\lambda}) \in \mathcal{L}^\infty([0, +\infty); X \times Y)$.

The above result suggests that the conclusions of Theorem 2.1 on the inertial dynamics (ID) may directly be conveyed to the (AH) differential system although less regularity assumptions on f are imposed. In fact, the proof of Theorem 6.2 follows the same line of arguments used in the proof of Theorem 2.1 with the key difference that the decay property of the mapping $t \mapsto \|(\dot{x}(t), \dot{\lambda}(t))\|$ is enforced by the monotonicity of the operator

$$(x, \lambda) \longmapsto (\nabla f(x) + h'(x)^* \lambda, -h(x)),$$

that is, the "generator" of the (AH) differential system. We leave the details to the reader.

Remark 6.3. We note that the assertions of Theorem 6.2 remain valid even under the assumption that $f : X \to \mathbb{R} \cup \{+\infty\}$ is a proper convex lower-semicontinuous function. In this case, the (AH) dynamics generalize to the evolution system

$$\begin{cases} \dot{x} + \partial f(x) + h'(x)^* \lambda \ni 0_X \\ \qquad\qquad\quad \dot{\lambda} - h(x) = 0_Y \end{cases}$$

with ∂f denoting the convex subdifferential of f. The existence and uniqueness of the (strong) solutions of the above evolution system then follows from the general theory for semi-groups of contractions generated by maximal monotone operators. We shall not pursue this point here but instead refer to Brézis [14] for a detailed study of the subject.

6.2. Augmented inertial dynamics. Let us now introduce an augmented variant of the inertial dynamics (ID) aiming to weaken the ellipticity condition imposed on the Hessian of f.

Consider the second-order differential system

$$\text{(AD)} \qquad \ddot{x} + \big(\nabla^2 f(x) + \rho \nabla^2 \|h(x)\|_Y^2/2\big)\dot{x} + \nabla \|h(x)\|_Y^2/2 = 0_X$$

with initial data $(x_0, v_0) \in X \times X$ and constant $\rho > 0$. We take for granted the existence and uniqueness of the (classical) solutions of the Augmented Dynamics (AD) and assume henceforth that

(A3)′ $h'(\,\cdot\,) : X \to Y$ is surjective;
(A4)′ $\nabla^2 f(\,\cdot\,) : X \to X$ is α-elliptic on $\ker h'(\,\cdot\,)$, i.e.,

$$\exists\alpha > 0\ \forall x \in X \quad \langle \nabla^2 f(x)y, y\rangle_X \geq \alpha\|y\|_X^2$$

for all $y \in \ker h'(x)$;
(A5)′ $\nabla^2 f(\,\cdot\,) : X \to X$ is γ-bounded.

Given the above assumptions, the additional geometric damping in (AD) allows one to infer the ellipticity of $\nabla^2 f(\,\cdot\,) + \rho\nabla^2\|h(\,\cdot\,)\|_Y^2/2$ even though $\nabla^2 f(\,\cdot\,)$ is only assumed to be α-elliptic on $\ker h'(\,\cdot\,)$. The following result makes this precise. We omit the proof but refer to Polyak and Tret'yakov [27, Lemma 1] for a similar result on quadratic forms which are elliptic on subspaces.

LEMMA 6.4. *Let $h'(\,\cdot\,)$ be surjective, and let $\nabla^2 f(\,\cdot\,)$ be α-elliptic on $\ker h'(\,\cdot\,)$ and γ-bounded. Then, there exists $\bar\rho > 0$ such that*

$$\exists\bar\alpha > 0\ \forall x, y \in X \quad \langle(\nabla^2 f(x) + \rho\nabla^2\|h(x)\|_Y^2/2)y, y\rangle_X \geq \bar\alpha\|y\|_X^2$$

for all $\rho \geq \bar\rho$.

As an immediate consequence, we observe that $f(\,\cdot\,) + \rho\|h(\,\cdot\,)\|^2/2$ is $\bar\alpha$-strongly convex whenever $\rho > 0$ is chosen sufficiently large. In this case, we have for any $x, y \in X$,

$$D_f(y, x) + \rho D_{\|h(\,\cdot\,)\|_Y^2/2}(y, x) \geq \bar\alpha\|x - y\|_X^2/2,$$

where D_f and $D_{\|h(\,\cdot\,)\|_Y^2/2}$ denote, respectively, the Bregman distances associated with f and $\|h(\,\cdot\,)\|_Y^2/2$.

In view of the above discussion, we readily deduce that the conclusions of Theorems 3.1 and 4.3 on the inertial dynamics (ID) also remain valid for the augmented dynamics (AD) given that the constant $\rho > 0$ is chosen sufficiently large. Moreover, the minimizing properties of the solutions of (AD) with respect to the convex minimization problem (P) are maintained as the following variants of Theorem 4.4 and Corollary 4.7 suggest.

THEOREM 6.5. *Let $h'(\,\cdot\,)$ be surjective, and let $\nabla^2 f(\,\cdot\,)$ be α-elliptic on $\ker h'(\,\cdot\,)$ and γ-bounded. Let $\rho > 0$ be sufficiently large, and let $x : [0, +\infty) \to X$ be a solution of (AD) with initial data $(x_0, v_0) \in X \times X$. Then, $x(t)$ converges strongly, as $t \to +\infty$, to the unique element $\bar x \in S$ satisfying*

$$D_{f+\rho\|h(\,\cdot\,)\|_Y^2/2}(\bar x, x_0) - \langle\bar x, v_0\rangle_X = \inf\nolimits_S D_{f+\rho\|h(\,\cdot\,)\|_Y^2/2}(\,\cdot\,, x_0) - \langle\,\cdot\,, v_0\rangle_X.$$

COROLLARY 6.6. *Under the hypotheses of Theorem 6.5, let $\bar x \in S$ be such that $x(t) \to \bar x$ strongly in X as $t \to +\infty$. Then, the following assertions hold:*

(i) If $v_0 = 0_X$, then $\bar x \in S$ is the unique element satisfying

$$D_{f+\rho\|h(\,\cdot\,)\|_Y^2/2}(\bar x, x_0) = \inf\nolimits_S D_{f+\rho\|h(\,\cdot\,)\|_Y^2/2}(\,\cdot\,, x_0);$$

(ii) If $v_0 + \nabla f(x_0) + \rho\nabla\|h(x_0)\|_Y^2/2 = 0_X$, then $\bar x \in S$ is the unique element satisfying

$$f(\bar x) + \rho\|h(\bar x)\|_Y^2/2 = \inf\nolimits_S f(\,\cdot\,) + \rho\|h(\,\cdot\,)\|_Y^2/2.$$

By virtue of Corollary 6.6*(ii)*, we infer that any solution $x(t)$ of (AD) with corresponding initial data $(x_0, -\nabla f(x_0) - \rho\nabla\|h(x_0)\|_Y^2/2) \in X \times X$ converges, as $t \to +\infty$, to the unique minimizer of the augmented convex minimization problem

$$\inf\{f(x) + \rho\|h(x)\|_Y^2/2 \mid h(x) = 0_Y\}.$$

Noticing that the set of minimizers of the augmented problem coincides with the one of (P), we conclude that the convergence is towards the unique minimizer of the convex minimization problem (P).

7. Numerical experiments. In this section, we perform numerical experiments on the inertial dynamics (ID) and the Arrow–Hurwicz differential system (AH) to illustrate the minimizing properties of their solutions relative to the linearly constrained convex minimization problem (P). In our numerical tests, we consider three simple but representative examples on convex, respectively, strongly convex programming in two dimensions.

Example 7.1 (Convex case). Let $X, Y = \mathbb{R}^2$ and consider the convex but not strongly convex function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x_1, x_2) = (x_1 + x_2)^2/2$. Moreover, let $h : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $h(x_1, x_2) = A(x_1, x_2) - b$, where $A(x_1, x_2) = (x_1, x_2)$ and $b = (1, 1)$. Clearly, $(x_1, x_2) \mapsto \|h(x_1, x_2)\|^2$ admits a strong minimum at $\bar{x} = (1, 1)$ and thus, $S = \{\bar{x}\}$. The evolution of the quantity $\|x(t)-\bar{x}\|$ along with the trajectories of the solutions $x(t) = (x_1(t), x_2(t))$ of (ID) and (AH) with corresponding initial data $x_0 = (0, 0)$, $v_0 = (-2, 0)$ and $\lambda_0 = (2, -1/2)$ is depicted in Figure 1.

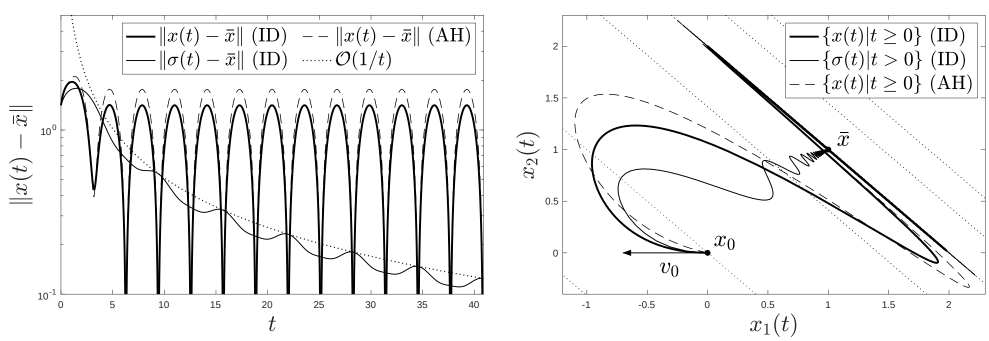

Figure 1: Graphical view on the evolution of $\|x(t) - \bar{x}\|$ and the trajectories of the solutions $x(t) = (x_1(t), x_2(t))$ of (ID) and (AH) in the convex case.

Analyzing Figure 1, we observe that the solutions $x(t)$ of (ID) and (AH) both fail to converge, as $t \to +\infty$, to the unique feasible (and thus optimal) point $\bar{x} \in S$ of the convex minimization problem (P). While the solutions of (ID) and (AH) remain bounded, cf. Remark 2.3, they both admit an oscillatory behavior corresponding to the "direction of vanishing damping" induced by the Hessian of f; cf. the sublevel sets of f depicted as dotted lines in the right panel of Figure 1. Despite the lack of damping, we observe that the Cesàro average $\sigma(t)$ of the solution of (ID) converges, as $t \to +\infty$, to the unique element $\bar{x} \in S$. Moreover, the convergence obeys, as $t \to +\infty$, the estimate $\|\sigma(t) - \bar{x}\| = \mathcal{O}(1/t)$; see Proposition 2.6. Finally, we remark that the solutions of (ID) and (AH) indeed share a similar behavior for the given set of initial conditions. In fact, both solutions are indistinguishable whenever their initial data is chosen according to Proposition 6.1.

Example 7.2 (Strongly convex case). Let $X = \mathbb{R}^2$, $Y = \mathbb{R}$, and consider the quadratic function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x_1, x_2) = (x_1^2 - x_1x_2 + x_2^2)/2$. Clearly, f is α-strongly convex with modulus $\alpha = 1/2$. Let further $h : \mathbb{R}^2 \to \mathbb{R}$ be defined by $h(x_1, x_2) = A(x_1, x_2) - b$, where $A(x_1, x_2) = x_1 + x_2$ and $b = 1$. In this case, we easily verify that $(x_1, x_2) \mapsto |h(x_1, x_2)|^2$ admits a strong minimum with respect to the set

$S = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 + x_2 = 1\}$. Moreover, $\operatorname{argmin}_S f = \bar{x}$ with $\bar{x} = (1/2, 1/2)$. Figure 2 displays the evolution of $\|x(t) - \bar{x}\|^2$ together with the trajectories of the solutions $x(t) = (x_1(t), x_2(t))$ of (ID) and (AH) with associated initial data $x_0 = (-1, 1)$, $v_0 = (3/2, -3/2)$ and $\lambda_0 = 1$.

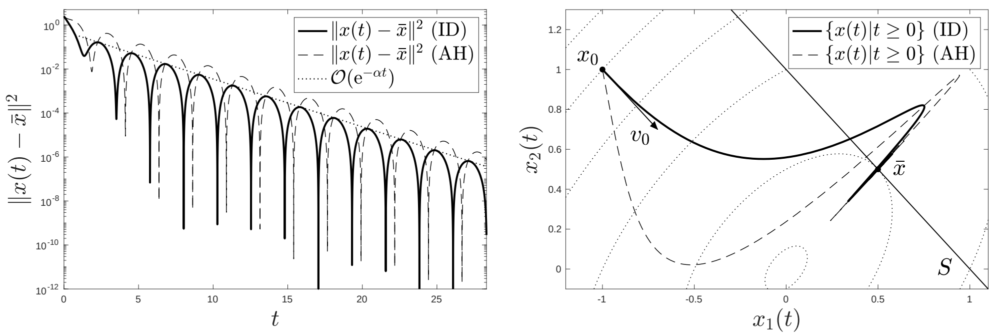

Figure 2: Graphical view on the evolution of $\|x(t) - \bar{x}\|^2$ and the trajectories of the solutions $x(t) = (x_1(t), x_2(t))$ of (ID) and (AH) in the strongly convex case.

According to Figure 2, the solutions of (ID) and (AH) tend to stabilize asymptotically towards an element of the set S; cf. Theorem 4.3. Moreover, as the initial velocity v_0 of the solution $x(t)$ of (ID) is taken along the negative direction of the gradient of f (cf. the right panel of Figure 2), we conclude its convergence, as $t \to +\infty$, towards the unique minimizer $\bar{x} \in S$ of the (strongly) convex minimization problem (P); see Corollary 4.7*(ii)*. Note that, however, for an arbitrary set of initial conditions, the limit of the solution of (ID) is characterized according to Theorem 4.4, following a hierarchical minimization principle involving the Bregman distance associated with f. It is also worth noting that the convergence appears to obey, as $t \to +\infty$, the estimate $\|x(t) - \bar{x}\|^2 = \mathcal{O}(\mathrm{e}^{-\alpha t})$ as predicted by Theorem 5.1*(i)*, even though the hypothesis on the strong minimum at $\bar{x} \in S$ is not verified. Geometrically, this is apparent as the trajectories of (ID) and (AH) evolve, for large values of t, solely in the direction where f admits its "lowest curvature" determined by the value of α.

Example 7.3 (Decay rate estimates). Finally, let $X, Y = \mathbb{R}^2$ and consider the parameterized quadratic function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x_1, x_2) = \alpha(x_1^2 + x_2^2)/2$, where $\alpha > 0$. In this case, the parameter α directly translates into a viscous damping coefficient of the inertial dynamics (ID). Let further $h : \mathbb{R}^2 \to \mathbb{R}^2$ be given in terms of $h(x_1, x_2) = A(x_1, x_2) - b$, where $A(x_1, x_2) = (x_1, x_2)$ and $b = (1, 1)$. Once again, $(x_1, x_2) \mapsto \|h(x_1, x_2)\|^2$ admits a strong minimum at $\bar{x} = (1, 1)$ with constant $\beta = 1$ and thus, $S = \{\bar{x}\}$. Figure 3 illustrates the decay properties of the squared error $\|x(t) - \bar{x}\|^2$ of the solutions $x(t) = (x_1(t), x_2(t))$ of (ID) and (AH) for the distinct values (i) $\alpha = 1$, (ii) $\alpha = 2$ and (iii) $\alpha = 3$. The initial data is set accordingly to $x_0 = (-1, 0)$, $v_0 = (\alpha, 0)$ and $\lambda_0 = (2, -1/2)$.

Figure 3 suggests that the solutions $x(t)$ of (ID) and (AH) converge, as $t \to +\infty$, at an exponential rate towards the unique feasible (and thus optimal) point $\bar{x} \in S$ of the (strongly) convex minimization problem (P). Indeed, the decay properties of the solutions of (ID) can be categorized as predicted by Corollary 5.4: In case (i), we have $\alpha^2 < 4\beta$, and thus, $\|x(t) - \bar{x}\|^2 = \mathcal{O}(\mathrm{e}^{-\alpha t})$ as $t \to +\infty$. We refer to this case as the "underdamped case" as the solutions of (ID) and (AH) both admit a significant oscillatory behavior. In case (ii), it holds that $\alpha^2 = 4\beta$, and thus,

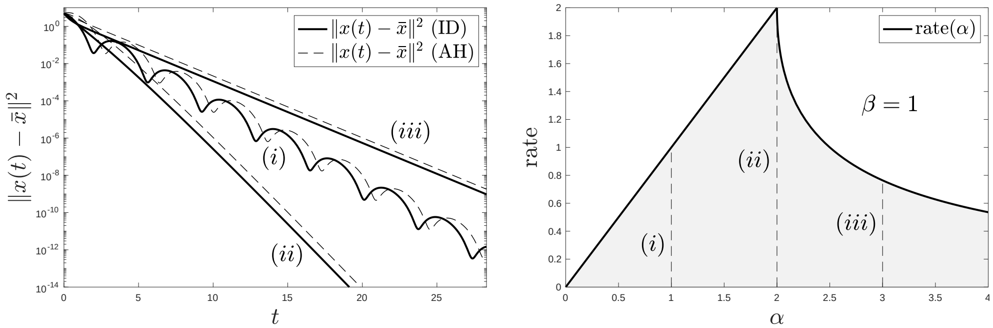

Figure 3: Decay properties of the squared error $\|x(t)-\bar{x}\|^2$ of the solutions $x(t) = (x_1(t), x_2(t))$ of (ID) and (AH) for distinct values of α.

$\|x(t)-\bar{x}\|^2 = \mathcal{O}(t^2 \mathrm{e}^{-\alpha t})$ as $t \to +\infty$. This case refers to the "critically damped case" for which we observe the fastest possible convergence of the solutions of (ID). Finally, in case (iii), we have $\alpha^2 > 4\beta$ such that $\|x(t)-\bar{x}\|^2 = \mathcal{O}(\mathrm{e}^{-(\alpha-\delta)t})$ as $t \to +\infty$, where $\delta = \sqrt{\alpha^2 - 4\beta}$. In this case, referred to as the "overdamped case", the decay of the solutions of (ID) and (AH) is considerably degraded; cf. the right panel of Figure 3 for a graphical illustration of the decay rate as a function of α. It is interesting to note that, in the first two cases, the decay rate estimates are fully characterized in terms of the damping coefficient α associated with f, whereas in the last case, the decay also depends on the constant β which implicitly captures the potential effects induced by h. Finally, we remark that the decay rate estimates on the solutions of (ID) appear to remain valid also for the solutions of the classical Arrow-Hurwicz differential system (AH).

Acknowledgments. The author expresses his gratitude to the two anonymous reviewers whose comments and suggestions led to a significant improvement of this manuscript.

REFERENCES

[1] F. Álvarez, *On the minimizing property of a second order dissipative system in Hilbert spaces*, SIAM J. Control Optim., 38 (2000), pp. 1102–1119.

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

[3] F. Álvarez, J. Bolte, and O. Brahic, *Hessian Riemannian gradient flows in convex programming*, SIAM J. Control Optim., 43 (2004), pp. 477–501.

[4] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in linear and non-linear programming*, Stanford University Press, Stanford, CA, 1958.

[5] H. Attouch, J. Bolte, P. Redont, and M. Teboulle, *Singular Riemannian barrier methods and gradient-projection dynamical systems for constrained optimization*, Optimization, 53 (2004), pp. 435–454.

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

[7] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi, *First-order optimization algorithms via inertial systems with Hessian driven damping*, Math. Program., 193 (2022), pp. 113–155.

[8] H. Attouch, X. Goudou, and P. Redont, *The heavy ball with friction method, I. The continuous dynamical system: Global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system*, Commun. Contemp. Math., 2 (2000), pp. 1–34.

[9] J.-B. Baillon and G. Haddad, *Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones*, Israel J. Math., 26 (1977), pp. 137–150.

[10] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, *Bregman monotone optimization algorithms*, SIAM J. Control Optim., 42 (2003), pp. 596–636.

[11] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, CMS Books in Mathematics, Springer, New York, 2017.

[12] R. I. Boţ and E. R. Csetnek, *Approaching monotone inclusion problems via second order dynamical systems with linear and anisotropic damping*, in New Trends in Differential Equations, Control Theory and Optimization, *Proceedings of the 8th Congress of Romanian Mathematicians*, World Scientific Publishing, (2016), pp. 53–72.

[13] L. M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, U.S.S.R. Comput. Math. and Math. Phys., 7 (1967), pp. 200–217.

[14] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, Mathematics Studies 5, North-Holland, Amsterdam, 1973.

[15] H. Brézis, *Functional analysis, Sobolev spaces and partial differential equations*, Springer, New York, 2011.

[16] R. E. Bruck, *Asymptotic convergence of nonlinear contraction semi-groups in Hilbert spaces*, J. Funct. Anal., 18 (1975), pp. 15–26.

[17] A. Cabot, H. Engler, and S. Gadat, *On the long time behavior of second order differential equations with asymptotically small dissipation*, Trans. Amer. Math. Soc., 361 (2009), pp. 5983–6017.

[18] G. Chen and M. Teboulle, *Convergence analysis of a proximal-like minimization algorithm using Bregman functions*, SIAM J. Optim., 3 (1993), pp. 538–543.

[19] I. Ekeland and R. Témam, *Convex analysis and variational problems*, Classics in applied mathematics, Society for Industrial and Applied Mathematics, Philadelphia, 1999.

[20] J. K. Hale, *Asymptotic behavior of dissipative systems*, Mathematical Surveys and Monographs 25, American Mathematical Society, Providence, Rhode Island, 1988.

[21] A. Haraux, *Systèmes dynamiques dissipatifs et applications*, Recherches en Mathématiques Appliquées 17, Masson, Paris, 1991.

[22] B. Lemaire, *An asymptotical variational principle associated with the steepest descent method for a convex function*, J. Convex Anal., 3 (1996), pp. 63–70.

[23] A. Liénard, *Étude des oscillations entretenues*, Rev. gén. d'électr., 23 (1928), pp. 901–912 and 946–954.

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

[25] B. T. Polyak, *Some methods of speeding up the convergence of iteration methods*, U.S.S.R. Comput. Math. and Math. Phys., 4 (1964), pp. 1–17.

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

[27] B. T. Polyak and N. V. Tret'yakov, *The method of penalty estimates for conditional extremum problems*, U.S.S.R. Comput. Math. and Math. Phys., 13 (1973), pp. 42–58.

[28] R. T. Rockafellar, *Monotone operators associated with saddle-functions and minimax problems*, In Nonlinear Functional Analysis, *Proceedings of Symposia in Pure Math.*, Amer. Math. Soc., (1969), pp. 241–250.