FAST CONVEX OPTIMIZATION VIA INERTIAL DYNAMICS
COMBINING VISCOUS AND HESSIAN-DRIVEN DAMPING WITH TIME RESCALING

Hedy Attouch
IMAG, Univ. Montpellier
CNRS, Montpellier, France

Aïcha Balhag, Zaki Chbani and Hassan Riahi
Cadi Ayyad University, Sémlika Faculty of Sciences
40000 Marrakech, Morocco

(Communicated by Alain Haraux)

Abstract. In a Hilbert setting, we develop fast methods for convex unconstrained optimization. We rely on the asymptotic behavior of an inertial system combining geometric damping with temporal scaling. The convex function to minimize enters the dynamic via its gradient. The dynamic includes three coefficients varying with time, one is a viscous damping coefficient, the second is attached to the Hessian-driven damping, the third is a time scaling coefficient. We study the convergence rate of the values under general conditions involving the damping and the time scale coefficients. The obtained results are based on a new Lyapunov analysis and they encompass known results on the subject. We pay particular attention to the case of an asymptotically vanishing viscous damping, which is directly related to the accelerated gradient method of Nesterov. The Hessian-driven damping significantly reduces the oscillatory aspects. We obtain an exponential rate of convergence of values without assuming the strong convexity of the objective function. The temporal discretization of these dynamics opens the gate to a large class of inertial optimization algorithms.

1. Introduction. Throughout the paper, $\mathcal{H}$ is a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$, and $f : \mathcal{H} \to \mathbb{R}$ is a convex and differentiable function. We aim at developing fast numerical methods for solving the optimization problem

$$\min_{x \in \mathcal{H}} f(x).$$

We denote by $\text{argmin}_{\mathcal{H}} f$ the set of minimizers of the optimization problem $(P)$, which is assumed to be non-empty. Our work is part of the active research stream that studies the close link between continuous dissipative dynamical systems and optimization algorithms. In general, the implicit temporal discretization of continuous gradient-based dynamics provides proximal algorithms that benefit from similar asymptotic convergence properties, see [29] for a systematic study in the case of first-order evolution systems, and [5, 6, 8, 12, 11, 19, 20, 21] for some recent developments.

2020 Mathematics Subject Classification. 37N40, 46N10, 49M15, 65K05, 65K10, 90C25, 93D05.

Key words and phrases. Damped inertial gradient dynamics, fast convex optimization, Hessian-driven damping, Nesterov accelerated gradient method, time rescaling.
results concerning second-order evolution equations. The main object of our study is the second-order in time differential equation

\[(\text{IGS})_{\gamma,\beta,b} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \beta(t)\nabla^2 f(x(t))\dot{x}(t) + b(t)\nabla f(x(t)) = 0,\]

where the coefficients \(\gamma, \beta : [t_0, +\infty[ \to \mathbb{R}_+\) take account of the viscous and Hessian-driven damping, respectively, and \(b : \mathbb{R}_+ \to \mathbb{R}_+\) is a time scale parameter. We take for granted the existence and uniqueness of the solution of the corresponding Cauchy problem with initial conditions \(x(t_0) = x_0 \in \mathcal{H}, \dot{x}(t_0) = v_0 \in \mathcal{H}\). We just draw the main lines of the proof. Indeed, this can be obtained by following an argument similar to that of [17, Theorem 4.4]: we first rewrite in an equivalent way \((\text{IGS})_{\gamma,\beta,b}\) as first order system in time and space. Assuming that \(\nabla f\) is Lipschitz continuous on the bounded sets, and that the coefficients are continuously differentiable, local existence follows from the non-autonomous version of the Cauchy-Lipschitz theorem, see [24, Prop. 6.2.1]. Then, the global existence follows from the energy estimates which will be established in the next section. Each of these damping and rescaling terms properly tuned, improves the rate of convergence of the associated dynamics and algorithms. An original aspect of our work is to combine them in the same dynamic. Let us recall some classical facts.

1.1. Damped inertial dynamics and optimization. The continuous-time perspective gives a mechanical intuition of the behavior of the trajectories, and a valuable tool to develop a Lyapunov analysis. A first important work in this perspective is the heavy ball with friction method of B. Polyak [30]

\[(\text{HBF}) \quad \ddot{x}(t) + \gamma \dot{x}(t) + \nabla f(x(t)) = 0.\]

It is a simplified model for a heavy ball (whose mass has been normalized to one) sliding on the graph of the function \(f\) to be minimized, and which asymptotically stops under the action of viscous friction, see [14] for further details. In this model, the viscous friction parameter \(\gamma\) is a fixed positive parameter. Due to too much friction (at least asymptotically) involved in this process, replacing the fixed viscous coefficient with a vanishing viscous coefficient (i.e. which tends to zero as \(t \to +\infty\)) gives Nesterov’s famous accelerated gradient method [27] [28]. The other two basic ingredients that we will use, namely time rescaling, and Hessian-driven damping have a natural interpretation (cinematic and geometric, respectively) in this context. We will come back to these points later. Precisely, we seek to develop fast first-order methods based on the temporal discretization of damped inertial dynamics. By fast we mean that, for a general convex function \(f\), and for each trajectory of the system, the convergence rate of the values \(f(x(t)) - \inf_{\mathcal{H}} f\) which is obtained is optimal (i.e. is achieved or nearly achieved in the worst case). The importance of simple first-order methods, and in particular gradient-based and proximal algorithms, comes from the applicability of these algorithms to a wide range of large-scale problems arising from machine learning and/or engineering.

1.1.1. The viscous damping parameter \(\gamma(t)\). A significant number of recent studies have focused on the case \(\gamma(t) = \frac{\alpha}{t}, \beta = 0\) (without Hessian-driven damping), and \(b = 1\) (without time rescaling), that is

\[(\text{AVD})_{\alpha} \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla f(x(t)) = 0.\]

This dynamic involves an Asymptotically Vanishing Damping coefficient (hence the terminology), a key property to obtain fast convergence for a general convex function
In [33], Su, Boyd and Candès showed that for \( \alpha = 3 \) the above system can be seen as a continuous version of the accelerated gradient method of Nesterov [27, 28] with \( f(x(t)) - \min_\mathcal{H} f = O(\frac{1}{t^2}) \) as \( t \to +\infty \). The importance of the parameter \( \alpha \) was put to the fore by Attouch, Chbani, Peypouquet and Redont [9] and May [26]. They showed that, for \( \alpha > 3 \), one can pass from capital \( O \) estimates to small \( o \). Moreover, when \( \alpha > 3 \), each trajectory converges weakly, and its limit belongs to \( \text{argmin } f \). Recent research considered the case of a general damping coefficient \( \gamma(\cdot) \) (see [4, 7]), thus providing a complete picture of the convergence rates for (AVD)_\alpha:

\[ f(x(t)) - \min_\mathcal{H} f = O(1/t^2) \]

when \( \alpha \geq 3 \), and

\[ f(x(t)) - \min_\mathcal{H} f = O \left( \frac{1}{t^{3/2}} \right) \]

when \( \alpha \leq 3 \), see [7, 13] and Apidopoulos, Aujol and Dossal [3].

1.1.2. The Hessian-driven damping parameter \( \beta(t) \). The inertial system

\[(DIN)_{\gamma,\beta} \quad \ddot{x}(t) + \gamma \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) = 0,\]

was introduced by Alvarez, Attouch, Bolte, and Redont in [2]. In line with (HBF), it contains a fixed positive friction coefficient \( \gamma \). As a main property, the introduction of the Hessian-driven damping makes it possible to neutralize the transversal oscillations likely to occur with (HBF), as observed in [2]. The need to take a geometric damping adapted to \( f \) had already been observed by Alvarez [1] who considered the inertial system

\[ \ddot{x}(t) + \mathcal{D} \dot{x}(t) + \nabla f(x(t)) = 0, \]

where \( \mathcal{D} : \mathcal{H} \to \mathcal{H} \) is a linear positive definite anisotropic operator. But still this damping operator is fixed. For a general convex function, the Hessian-driven damping in \((DIN)_{\gamma,\beta}\) performs a similar operation in a closed-loop adaptive way. (DIN) stands shortly for Dynamical Inertial Newton, and refers to the link with the Levenberg-Marquardt regularization of the continuous Newton method. Recent studies have been devoted to the study of the inertial dynamic

\[ \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) = 0, \]

which combines asymptotic vanishing damping with Hessian-driven damping [17].

1.1.3. The time rescaling parameter \( b(t) \). In the context of non-autonomous dissipative dynamic systems, reparameterization in time is a simple and universal means to accelerate the convergence of trajectories. This is where the coefficient \( b(t) \) comes in as a factor of \( \nabla f(x(t)) \). In [12] [11], in the case of general coefficients \( \gamma(\cdot) \) and \( b(\cdot) \) without the Hessian damping, the authors made in-depth study. In the case \( \gamma(t) = \frac{\alpha}{t} \), they proved that under appropriate conditions on \( \alpha \) and \( b(\cdot) \), \( f(x(t)) - \min_\mathcal{H} f = O(\frac{1}{t^{3/2}}) \). Hence a clear improvement of the convergence rate by taking \( b(t) \to +\infty \) as \( t \to +\infty \).

1.2. From damped inertial dynamics to proximal-gradient inertial algorithms. Let’s review some classical facts concerning the close link between continuous dissipative inertial dynamic systems and the corresponding algorithms obtained by temporal discretization. Let us insist on the fact that, when the temporal scaling \( b(t) \to +\infty \) as \( t \to +\infty \), the transposition of the results to the discrete case naturally leads to consider an implicit temporal discretization, i.e. inertial proximal algorithms. The reason is that, since \( b(t) \) is in front of the gradient, the application of the gradient descent lemma would require taking a step size that tends to

\[ f. \]
zero. On the other hand, the corresponding proximal algorithms involve a proximal coefficient which tends to infinity (large step proximal algorithms).

1.2.1. The case without the Hessian-driven damping. The implicit discretization of (IGS)$_{\alpha,b}$ gives the Inertial Proximal algorithm

$$
(IP)_{\alpha_k,\lambda_k} \quad \left\{ \begin{array}{l} y_k = x_k + \alpha_k (x_k - x_{k-1}) \\ x_{k+1} = \text{prox}_{\lambda_k f}(y_k) \end{array} \right.
$$

where $\alpha_k$ is non-negative and $\lambda_k$ is positive. Recall that for any $\lambda > 0$, the proximity operator $\text{prox}_{\lambda f} : \mathcal{H} \to \mathcal{H}$ is defined by the following formula: for every $x \in \mathcal{H}$

$$
\text{prox}_{\lambda f}(x) := \arg\min_{\xi \in \mathcal{H}} \left\{ f(\xi) + \frac{1}{2\lambda} \|x - \xi\|^2 \right\}.
$$

Equivalently, $\text{prox}_{\lambda f}$ is the resolvent of index $\lambda$ of the maximally monotone operator $\partial f$. When passing to the implicit discrete case, we can take $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ a convex lower semicontinuous and proper function. Let us list some of the main results concerning the convergence properties of the algorithm $(IP)_{\alpha_k,\lambda_k}$:

1. Case $\lambda_k \equiv \lambda > 0$ and $\alpha_k = 1 - \frac{\alpha}{\kappa}$. When $\alpha = 3$, the $(IP)_{1-3/k,\lambda}$ algorithm has a similar structure to the original Nesterov accelerated gradient algorithm\cite{27}, just replace the gradient step with a proximal step. Passing from the gradient to the proximal step was carried out by G"uler\cite[23, 22]{23}, then by Beck and Teboulle\cite{18} for structured optimization. A decisive step was taken by Attouch and Peynot in\cite{16} proving that, when $\alpha > 3$, $f(x_k) - \min_{\mathcal{H}} f = o\left(\frac{1}{k^\delta}\right)$. The subcritical case $\alpha < 3$ was examined by Apidopoulos, Aujol, and Dossal\cite{3} and Attouch, Chbani, and Riahi\cite{13} with the rate of convergence of values $f(x_k) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k^{\frac{3}{2}}}\right)$.

2. For a general $\alpha_k$, the convergence properties of $(IP)_{\alpha_k,\lambda}$ were analyzed by Attouch and Cabot\cite{5}, then by Attouch, Cabot, Chbani, and Riahi\cite{6}, in the presence of perturbations. The convergence rates are then expressed using the sequence $(t_k)$ which is linked to $(\alpha_k)$ by the formula $t_k := 1 + \sum_{j=k}^{+\infty} \prod_{i=k}^{j-1} \alpha_j$. Under growth conditions on $t_k$, it is proved that $f(x_k) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t_k}\right)$. This last result covers the special case $\alpha_k = 1 - \frac{\alpha}{\kappa}$ when $\alpha \geq 3$.

3. For a general $\lambda_k$, Attouch, Chbani, and Riahi first considered in\cite{12} the case $\alpha_k = 1 - \frac{\alpha}{\kappa}$. They proved that under a growth condition on $\lambda_k$, we have the estimate $f(x_k) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{\lambda_k \alpha}\right)$. This result is an improvement of the one discussed previously in\cite{16}, because when $\lambda_k = k^\delta$ with $0 < \delta < \alpha - 3$, we pass from $\mathcal{O}\left(\frac{1}{t_k}\right)$ to $\mathcal{O}\left(\frac{1}{t_k^\delta}\right)$. Recently, in\cite{10} the authors analyzed the algorithm $(IP)_{\alpha_k,\lambda_k}$ for general $\alpha_k$ and $\lambda_k$. By including the expression of $t_k$ previously used in\cite{5, 6}, they proved that $f(x_k) - \min_{\mathcal{H}} f = \mathcal{O}\left(1/t_k^2 \lambda_{k-1}\right)$ under certain conditions on $\lambda_k$ and $\alpha_k$. They obtained $f(x_k) - \min_{\mathcal{H}} f = o\left(1/t_k^2 \lambda_k\right)$, which gives a global view of the convergence rate with small $o$, encompassing\cite{5, 10}.

1.2.2. The case with the Hessian-driven damping. Recent studies have been devoted to the inertial dynamic

$$
\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) = 0,
$$
which combines asymptotic vanishing viscous damping with Hessian-driven damping. The corresponding algorithms involve a correcting term in the Nesterov accelerated gradient method which reduces the oscillatory aspects, see Attouch-Peynot-Redont [17], Attouch-Chbani-Fadili-Riahi [8], Shi-Du-Jordan-Su [31]. The case of monotone inclusions has been considered by Attouch and László [15].

1.3. Contents. The paper is organized as follows. In section 2, we develop a new Lyapunov analysis for the continuous dynamic \((IGS)_{\gamma,\beta,b}\). In Theorem 2.1, we provide a system of conditions on the damping parameters \(\gamma(\cdot)\) and \(\beta(\cdot)\), and on the temporal scaling parameter \(b(\cdot)\) giving fast convergence of the values. Then, in sections 3 and 4, we present two different types of growth conditions for the damping and temporal scaling parameters, respectively based on the functions \(\Gamma_{\gamma}\) and \(p_\gamma\), and which satisfy the conditions of Theorem 2.1. In doing so, we encompass most existing results and provide new results, including linear convergence rates without assuming strong convexity. This will also allow us to explain the choice of certain coefficients in the associated algorithms, questions which have remained mysterious and only justified by the simplification of often complicated calculations. In section 5, we specialize our results to certain model situations and give numerical illustrations. Finally, we conclude the paper by highlighting its original aspects.

2. Convergence rate of the values. General abstract result. We will establish a general result concerning the convergence rate of the values verified by the solution trajectories \(x(\cdot)\) of the second-order evolution equation

\[
(IGS)_{\gamma,\beta,b} \quad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \beta(t)\nabla f(x(t))\dot{x}(t) + b(t)\nabla f(x(t)) = 0.
\]

The variable parameters \(\gamma(\cdot)\), \(\beta(\cdot)\) and \(b(\cdot)\) take into account the damping, and temporal rescaling effects. They are assumed to be continuously differentiable. To analyze the asymptotic behavior of the solutions trajectories of the evolution system \((IGS)_{\gamma,\beta,b}\), we will use Lyapunov’s analysis. It is a classic and powerful tool which consists in building an associated energy-like function which decreases along the trajectories, see [25] for a general presentation and recent advances. The determination of such a Lyapunov function is in general a delicate problem. Based on previous works, we know the global structure of such a Lyapunov function. It is a weighted sum of the potential, kinetic and anchor functions. We will introduce coefficients in this function that are a priori unknown, and which will be identified during the calculation to verify the property of decay. Our approach takes advantage of the technics recently developed in [4], [17], [11].

2.1. The general case. Let \(x(\cdot)\) be a solution trajectory of \((IGS)_{\gamma,\beta,b}\). Given \(z \in \text{argmin}_x f\), we introduce the Lyapunov function \(t \mapsto \mathcal{E}(t)\) defined by

\[
\mathcal{E}(t) := c(t)^2b(t)(f(x(t)) - f(z)) + \frac{\theta(t)\sigma(t)^2}{2}\|v(t)\|^2 + \frac{\xi(t)}{2}\|x(t) - z\|^2,
\]

where \(v(t) := x(t) - z + \frac{1}{\sigma(t)}(\dot{x}(t) + \beta(t)\nabla f(x(t)))\).
The four variable coefficients \( \beta(t), \theta(t), \sigma(t), \xi(t) \) will be adjusted during the calculation. According to the classical derivation chain rule, we obtain

\[
\frac{d}{dt} \mathcal{E}(t) = \frac{d}{dt} (c^2(t)b(t)(f(x(t)) - f(z))) + c(t)^2 b(t) \langle \nabla f(x(t)), \dot{x}(t) \rangle \\
+ \frac{1}{2} \frac{d}{dt} \theta(t)^2 \|v(t)\|^2 + \theta(t) a^2(t) \langle \dot{v}(t), v(t) \rangle \\
+ \frac{1}{2} \xi(t) \|x(t) - z\|^2 + \xi(t) \langle \dot{x}(t), x(t) - z \rangle.
\]

From now, without ambiguity, to shorten formulas, we omit the variable \( t \). According to the definition of \( v \), and the equation (IGS)_{\gamma, \beta, b}, we have

\[
\dot{v} = \dot{x} - \frac{\dot{x}}{\sigma^2} (\dot{x} + \beta \nabla f(x)) + \frac{1}{\sigma} \frac{d}{dt} (\dot{x} + \beta \nabla f(x)).
\]

Therefore,

\[
\langle \dot{v}, v \rangle = \left( 1 - \frac{\dot{x}}{\sigma^2} - \frac{\gamma}{\sigma} \right) \dot{x} + \left( \frac{\beta}{\sigma} - \frac{\dot{x}}{\sigma^2} - \frac{b}{\sigma} \right) \nabla f(x), x - z + \frac{1}{\sigma} (\dot{x} + \beta \nabla f(x)) \nabla f(x).
\]

According to the definition of \( v(t) \), after developing \( \|v(t)\|^2 \), we get

\[
\|v\|^2 = \|x - z\|^2 + \frac{1}{\sigma^2} (\|\dot{x}\|^2 + \beta^2 \|\nabla f(x)\|^2) + \frac{2}{\sigma} \langle \dot{x}, x - z \rangle \\
+ \frac{2\beta}{\sigma} (\nabla f(x), x - z) + \frac{2\beta}{\sigma^2} (\nabla f(x), \dot{x}).
\]

Collecting the above results, we obtain

\[
\frac{d}{dt} \mathcal{E}(t) = \frac{d}{dt} (c^2(t)b(t)(f(x) - f(z))) + c^2 b \langle \nabla f(x), \dot{x} \rangle + \frac{1}{2} \xi \|x - z\|^2 + \xi (\dot{x}, x - z) \\
+ \frac{1}{\sigma^2} \theta(t)^2 (\|x - z\|^2 + \frac{1}{\sigma^2} (\|\dot{x}\|^2 + \beta^2 \|\nabla f(x)\|^2)) + \frac{2}{\sigma} (\dot{x}, x - z) \\
+ \frac{1}{\sigma^2} \theta(t) (\frac{2\beta}{\sigma} (\nabla f(x), x - z) + \frac{2\beta}{\sigma^2} (\nabla f(x), \dot{x})) \\
+ \theta \sigma^2 \left( \left( 1 - \frac{\dot{x}}{\sigma^2} - \frac{\gamma}{\sigma} \right) \langle \dot{x}, x - z \rangle + \left( 1 - \frac{\dot{x}}{\sigma^2} - \frac{\gamma}{\sigma} \right) \frac{1}{\sigma} \|\dot{x}\|^2 \right) \\
+ \theta \sigma^2 \left( \left( 1 - \frac{\dot{x}}{\sigma^2} - \frac{\gamma}{\sigma} \right) \frac{\beta}{\sigma} + \left( \frac{\beta}{\sigma} - \frac{\dot{x}}{\sigma^2} - \frac{b}{\sigma} \right) \frac{1}{\sigma} \right) \langle \nabla f(x), \dot{x} \rangle.
\]


In the second member of the above formula, let us examine the terms that contain $d$. To get
\[
\langle \nabla f(x), x - z \rangle
\]
To majorize it, we use the convex subgradient inequality
\[
\left( \frac{\beta}{\sigma} - \frac{\hat{\sigma} \beta}{\sigma^2} - \frac{b}{\sigma} \right) \nabla f(x)\| f(x) \| ^2.
\]
In the second member of the above formula, let us examine the terms that contain $\langle \nabla f(x), x - z \rangle$. By grouping these terms, we obtain the following expression
\[
\left( \frac{\beta}{\sigma} d (\theta \sigma^2) + \theta \sigma^2 \left( \frac{\hat{\beta}}{\sigma} - \frac{\hat{\sigma} \beta}{\sigma^2} - \frac{b}{\sigma} \right) \right) \langle \nabla f(x), x - z \rangle.
\]
To majorize it, we use the convex subgradient inequality $\langle \nabla f(x), x - z \rangle \geq f(x) - f(z)$, and we make a first hypothesis $\frac{\beta}{\sigma} d (\theta \sigma^2) + \theta \sigma^2 \left( \frac{\hat{\beta}}{\sigma} - \frac{\hat{\sigma} \beta}{\sigma^2} - \frac{b}{\sigma} \right) \leq 0$. Therefore,
\[
\frac{d}{dt} \mathcal{E} \leq \left[ \frac{d}{dt} \left( c^2 b + \frac{\beta}{\sigma} \frac{d}{dt} (\theta \sigma^2) + \theta \sigma^2 \left( \frac{\hat{\beta}}{\sigma} - \frac{\hat{\sigma} \beta}{\sigma^2} - \frac{b}{\sigma} \right) \right) \right] (f(x) - f(z))
\]
\[
+ \left[ \frac{1}{\sigma} \frac{d}{dt} (\theta \sigma^2) + \theta \sigma^2 \left( 1 - \frac{\hat{\sigma}}{\sigma} - \frac{\gamma}{\sigma} \right) \right] \langle \dot{x}, x - z \rangle
\]
\[
+ \frac{1}{2} \left[ \frac{d}{dt} (\theta \sigma^2) + \dot{\xi} \right] \| x - z \|^2 + \left[ \frac{1}{2 \sigma^2} \frac{d}{dt} (\theta \sigma^2) + \theta \sigma \left( 1 - \frac{\hat{\sigma}}{\sigma} - \frac{\gamma}{\sigma} \right) \right] \| \dot{x} \|^2
\]
\[
+ \left[ \frac{\beta^2}{2 \sigma^2} \frac{d}{dt} (\theta \sigma^2) + \theta \sigma \beta \left( \frac{\hat{\beta}}{\sigma} - \frac{\hat{\sigma} \beta}{\sigma^2} - \frac{b}{\sigma} \right) \right] \| \nabla f(x) \|^2.
\]
To get $\frac{d}{dt} \mathcal{E}(t) \leq 0$, we are led to make the following assumptions:

(i) \[
\frac{\beta}{\sigma} \frac{d}{dt} (\theta \sigma^2) + \theta \sigma^2 \left( \frac{\hat{\beta}}{\sigma} - \frac{\hat{\sigma} \beta}{\sigma^2} - \frac{b}{\sigma} \right) \leq 0
\]

(ii) \[
\frac{d}{dt} \left( c^2 b + \frac{\beta}{\sigma} \frac{d}{dt} (\theta \sigma^2) + \theta \sigma^2 \left( \frac{\hat{\beta}}{\sigma} - \frac{\hat{\sigma} \beta}{\sigma^2} - \frac{b}{\sigma} \right) \right) \leq 0,
\]

(iii) \[
c^2 b + \frac{\beta}{\sigma^2} \frac{d}{dt} (\theta \sigma^2) + \theta \sigma \left( 1 - \frac{\hat{\sigma}}{\sigma} - \frac{\gamma}{\sigma} \right) \beta + \left( \frac{\hat{\beta}}{\sigma} - \frac{\hat{\sigma} \beta}{\sigma^2} - \frac{b}{\sigma} \right) = 0,
\]

(iv) \[
\frac{1}{\sigma} \frac{d}{dt} (\theta \sigma^2) + \theta \sigma^2 \left( 1 - \frac{\hat{\sigma}}{\sigma} - \frac{\gamma}{\sigma} \right) + \dot{\xi} = 0,
\]

(v) \[
\frac{d}{dt} (\theta \sigma^2) + \dot{\xi} \leq 0,
\]

(vi) \[
\frac{1}{2 \sigma^2} \frac{d}{dt} (\theta \sigma^2) + \theta \sigma \left( 1 - \frac{\hat{\sigma}}{\sigma} - \frac{\gamma}{\sigma} \right) \leq 0,
\]

(vii) \[
\frac{\beta^2}{2 \sigma^2} \frac{d}{dt} (\theta \sigma^2) + \theta \sigma \beta \left( \frac{\hat{\beta}}{\sigma} - \frac{\hat{\sigma} \beta}{\sigma^2} - \frac{b}{\sigma} \right) \leq 0.
\]
After simplification, we get the following equivalent system of conditions:

**A: Lyapunov system of inequalities involving \( c(t), \theta(t), \sigma(t), \xi(t) \).**

\[
\begin{align*}
(i) & \quad \frac{d}{dt}(\beta \theta \sigma) - \theta b \sigma \leq 0, \\
(ii) & \quad \frac{d}{dt}(c^2 b + \beta \theta \sigma) - \theta b \sigma \leq 0, \\
(iii) & \quad b(c^2 - \theta) + \beta \theta (\sigma - \gamma) + \frac{d}{dt}(\beta \theta) = 0, \\
(iv) & \quad \frac{d}{dt}(\theta \sigma) + \theta \sigma (\sigma - \gamma) + \xi = 0, \\
(v) & \quad \frac{d}{dt}(\theta \sigma^2 + \xi) \leq 0, \\
(vi) & \quad \dot{\theta} + 2(\sigma - \gamma) \theta \leq 0, \\
(vii) & \quad \beta \left( \dot{\beta} \theta + 2 \left( \beta - b \right) \theta \right) \leq 0.
\end{align*}
\]

Let’s simplify this system by eliminating the variable \( \xi \). From \((iv)\) we get \( \xi = -\frac{d}{dt}(\theta \sigma) - \theta \sigma (\sigma - \gamma) \), that we replace in \((v)\), and recall that \( \xi \) is prescribed to be nonnegative. Now observe that the unknown function \( c \) can also be eliminated. Indeed, it enters the above system via the variable \( bc^2 \), which according to \((iii)\) verifies the equality \( bc^2 = b\theta - \beta \theta (\sigma - \gamma) - \frac{d}{dt}(\beta \theta) \). Replacing in \((ii)\), which is the only other equation involving \( bc^2 \), we obtain the equivalent system involving only the variables \( \theta(t), \sigma(t) \).

**B: Lyapunov system of inequalities involving the variables: \( \theta(t), \sigma(t) \)**

\[
\begin{align*}
(i) & \quad \frac{d}{dt}(\beta \theta \sigma) - \theta b \sigma \leq 0, \\
(ii) & \quad \frac{d}{dt}(b\theta + \beta \theta \gamma) - \frac{d^2}{dt^2}(\beta \theta) - \theta b \sigma \leq 0, \\
(iii) & \quad b\theta - \beta \theta (\sigma - \gamma) - \frac{d}{dt}(\beta \theta) \geq 0, \\
(iv) & \quad \frac{d}{dt}(\theta \sigma) + \theta \sigma (\sigma - \gamma) \leq 0, \\
(v) & \quad \frac{d}{dt} \left( -\frac{d}{dt}(\theta \sigma) + \theta \sigma \gamma \right) \leq 0, \\
(vi) & \quad \dot{\theta} + 2(\sigma - \gamma) \theta \leq 0, \\
(vii) & \quad \beta \left( \dot{\beta} \theta + 2 \left( \beta - b \right) \theta \right) \leq 0.
\end{align*}
\]

Then, the variables \( \xi \) and \( c \) are obtained by using the formulas

\[
\begin{align*}
\xi &= -\frac{d}{dt}(\theta \sigma) - \theta \sigma (\sigma - \gamma) \\
bc^2 &= b\theta - \beta \theta (\sigma - \gamma) - \frac{d}{dt}(\beta \theta).
\end{align*}
\]

Thus, under the above conditions, the function \( \mathcal{E}(\cdot) \) is nonnegative and nonincreasing. Therefore, for every \( t \geq t_0 \), \( \mathcal{E}(t) \leq \mathcal{E}(t_0) \), which implies that

\[
c^2(t)b(t)f(x(t)) - \min_{\mathcal{H}} f \leq \mathcal{E}(t_0).
\]
Therefore, as $t \to +\infty$
\[
f(x(t)) - \min_{\mathcal{H}} f = O\left(\frac{1}{c^2(t)b(t)}\right).
\]
Moreover, by integrating (3) we obtain the following integral estimates:

a) On the values:
\[
\int_{t_0}^{+\infty} \left( (\theta(t)b(t)\sigma(t) - \frac{d}{dt}(c^2(t)b(t) + \beta(t)\theta(t)\sigma(t))) \right) \left( f(x(t)) - \inf_{\mathcal{H}} f \right) dt < +\infty,
\]
where we use the equality:
\[-\left(\frac{d}{dt}(c^2b) + \frac{\sigma}{\sigma} \frac{d}{dt}(\theta\sigma^2) + \theta\sigma^2 \left(\frac{\sigma}{\sigma} - \frac{\partial^2}{\partial^2 - \frac{\partial}{\partial} \theta\sigma^2}\right)\right) = \theta b\sigma - \frac{d}{dt}(c^2b + \theta\sigma)\]
and the fact that, according to (ii), this quantity is nonnegative.

b) On the norm of the gradients:
\[
\int_{t_0}^{+\infty} q(t) \| \nabla f(x(t)) \|^2 dt < +\infty,
\]
where $q$ is the nonnegative weight function defined by
\[
q(t) := \theta(t)\beta(t) \left( \frac{\dot{\theta}(t)\beta(t)}{\sigma(t)} + b(t) - \dot{\beta}(t) \right) - \frac{\beta^2(t)}{2\sigma^2(t)} \frac{d}{dt}(\theta\sigma^2(t))
\]
\[
= b(t)\theta(t)\beta(t) - \frac{1}{2} \frac{d}{dt}(\theta\beta^2(t)).
\]
(4)

We can now state the following Theorem, which summarizes the above results.

**Theorem 2.1.** Let $f : \mathcal{H} \to \mathbb{R}$ be a convex differentiable function with $\text{argmin}_{\mathcal{H}} f \neq \emptyset$.

Let $x(\cdot)$ be a solution trajectory of
\[
\begin{align*}
(\text{IGS})_{\gamma,\beta,b} & \quad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \beta(t)\nabla^2 f(x(t))\dot{x}(t) + b(t)\nabla f(x(t)) = 0.
\end{align*}
\]

Suppose that $\gamma(\cdot)$, $\beta(\cdot)$, and $b(\cdot)$, are $\mathcal{C}^1$ functions on $[t_0, +\infty]$ such that there exists auxiliary functions $c(t), \theta(t), \sigma(t), \xi(t)$ that satisfy the conditions (i) – (vii) above.

Set
\[
\mathcal{E}(t) := c(t)^2b(t)(f(x(t)) - f(z)) + \frac{\theta(t)\sigma(t)^2}{2} \| v(t) \|^2 + \frac{\xi(t)}{2} \| x(t) - z \|^2, \quad (5)
\]
with $z \in \text{argmin}_{\mathcal{H}} f$ and $v(t) = x(t) - z + \frac{1}{\sigma(t)}(\dot{x}(t) + \beta(t)\nabla f(x(t)))$.

Then, $t \mapsto \mathcal{E}(t)$ is a nonincreasing function. As a consequence, for all $t \geq t_0$,

a) $f(x(t)) - \min_{\mathcal{H}} f \leq \frac{\mathcal{E}(t_0)}{c^2(t)b(t)}$; \quad (6)

b) $\int_{t_0}^{+\infty} \left( (\theta(t)b(t)\sigma(t) - \frac{d}{dt}(c^2b + \theta\sigma)) \right) \left( f(x(t)) - \inf_{\mathcal{H}} f \right) dt < +\infty$; \quad (7)

c) $\int_{t_0}^{+\infty} \left( b(t)\theta(t)\beta(t) - \frac{1}{2} \frac{d}{dt}(\theta\beta^2(t)) \right) \| \nabla f(x(t)) \|^2 dt < +\infty$. \quad (8)
2.2. **Solving system** (i) – (vii). The system of inequalities (i) – (vii) of Theorem 2.1 may seem complicated at first glance. Indeed, we will see that it simplifies notably in the classical situations. Moreover, it makes it possible to unify the existing results, and discover new interesting cases. We will present two different types of solutions to this system, respectively based on the following functions:

\[ p_\gamma(t) = \exp \left( \int_{t_0}^{t} \gamma(u) \, du \right), \quad (9) \]

and

\[ \Gamma_\gamma(t) = p_\gamma(t) \int_{t}^{+\infty} \frac{du}{p_\gamma(u)}. \quad (10) \]

The use of \( \Gamma_\gamma \) has been considered in a series of articles that we will retrieve as a special case of our approach, see [4], [5], [7], [11]. Using \( p_\gamma \) will lead to new results, see section 4.

3. **Results based on the function** \( \Gamma_\gamma \). In this section, we will systematically assume that condition \((H_0)\) is satisfied.

\[ \int_{t_0}^{+\infty} \frac{ds}{p_\gamma(s)} < +\infty. \]

Under \((H_0)\), the function \( \Gamma_\gamma(\cdot) \) is well defined. It can be equally defined as the solution of the linear non autonomous differential equation

\[ \dot{\Gamma}_\gamma(t) - \gamma(t) \Gamma_\gamma(t) + 1 = 0, \quad (11) \]

which satisfies the limit condition \( \lim_{t \to +\infty} \frac{\Gamma_\gamma(t)}{p_\gamma(t)} = 0. \)

3.1. **The case without the Hessian**, i.e. \( \beta \equiv 0 \). The dynamic writes

\[ (\text{IGS})_{\gamma, 0, b} \quad \ddot{x}(t) + \gamma(t) \dot{x}(t) + b(t) \nabla f(x(t)) = 0. \]

To solve the system (i) – (vii) of Theorem 2.1, we choose

\[ \xi \equiv 0, \quad c(t) = \Gamma_\gamma(t), \quad \sigma(t) = \frac{1}{\Gamma_\gamma(t)}, \quad \theta(t) = \Gamma_\gamma(t)^2. \]

According to (11), we can easily verify that conditions (i), (iii) – (vii) are satisfied, and (ii) becomes

\[ \frac{d}{dt} \left( \Gamma_\gamma(t)^2 b(t) \right) - \gamma(t) b(t) \leq 0. \]

After dividing by \( \Gamma_\gamma(t) \), and using (11), we obtain the condition

\[ 0 \geq \Gamma_\gamma(t) \dot{b}(t) - (3 - 2\gamma(t)\Gamma_\gamma(t))b(t). \]

This leads to the following result obtained by Attouch, Chbani and Riahi in [11].

**Theorem 3.1.** [11, Theorem 2.1] **Suppose that for all** \( t \geq t_0 \)

\[ \Gamma_\gamma(t) \dot{b}(t) \leq b(t) (3 - 2\gamma(t)\Gamma_\gamma(t)), \quad (12) \]

**where** \( \Gamma_\gamma \) **is defined from** \( \gamma \) **by (10). Let** \( x : [t_0, +\infty[ \to \mathcal{H} \) **be a solution trajectory of** \( (\text{IGS})_{\gamma, 0, b} \). Given** \( z \in \text{argmin}_{\mathcal{H}} f \), **set**

\[ \mathcal{E}(t) := \Gamma_\gamma^2(t) b(t) (f(x(t)) - f(z)) + \frac{1}{2} \| x(t) - z + \Gamma_\gamma(t) \dot{x}(t) \|^2. \quad (13) \]
Then, $t \mapsto \mathcal{E}(t)$ is a nonincreasing function. As a consequence, as $t \to +\infty$

$$
 f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{\Gamma_{\gamma}(t)^2 b(t)}\right). 
$$

Precisely, for all $t \geq t_0$

$$
 f(x(t)) - \min_{\mathcal{H}} f \leq \frac{C}{\Gamma_{\gamma}(t)^2 b(t)},
$$

with $C = \Gamma_{\gamma}(t_0)^2 b(t_0) (f(x(t_0)) - \min_{\mathcal{H}} f) + d(x(t_0), \text{argmin } f)^2 + \Gamma_{\gamma}(t_0)^2 \|\dot{x}(t_0)\|^2$. Moreover,

$$
 \int_{t_0}^{+\infty} \Gamma_{\gamma}(t) \left( b(t) (3 - 2\gamma(t)\Gamma_{\gamma}(t)) - \Gamma_{\gamma}(t)b(t) \right) (f(x(t)) - \min_{\mathcal{H}} f) dt < +\infty.
$$

**Remark 1.** When $b \equiv 1$, condition (12) reduces to $\gamma(t)\Gamma_{\gamma}(t) \leq \frac{3}{2}$, introduced in [4].

### 3.2. Combining Nesterov acceleration with Hessian damping.

Let us specialize our results in the case $\beta(t) > 0$, and $\gamma(t) = \frac{\alpha}{t}$. We are in the case of a vanishing damping coefficient (i.e. $\gamma(t) \to 0$ as $t \to +\infty$). According to Su, Boyd and Candès [33], the case $\alpha = 3$ corresponds to a continuous version of the accelerated gradient method of Nesterov. Taking $\alpha > 3$ improves in many ways the convergence properties of this dynamic, see section 1.1.1. Here, it is combined with the Hessian-driven damping and temporal rescaling. This situation was first considered by Attouch, Chbani, Fadili and Riahi in [8]. Then the dynamic writes

$$(\text{IGS})_{\alpha/\Gamma, \beta, b} \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta(t)\nabla^2 f(x(t))\dot{x}(t) + b(t)\nabla f(x(t)) = 0.$$

Elementary calculus gives that $(H_0)$ is satisfied as soon as $\alpha > 1$. In this case,

$$
 \Gamma_{\gamma}(t) = \frac{t}{\alpha - 1}.
$$

After [8], let us introduce the following quantity which will simplify the formulas:

$$
 w(t) := b(t) - \beta(t) - \frac{\beta(t)}{t}.
$$

The following result will be obtained as a consequence of our general abstract Theorem 2.1. Precisely, we will show that under an appropriate choice of the functions $c(t), \theta(t), \sigma(t), \xi(t)$, the conditions (i) – (vii) of Theorem 2.1 are satisfied.

**Theorem 3.2.** [8, Theorem 1] Let $x: [t_0, +\infty[ \to \mathcal{H}$ be a solution trajectory of

$$(\text{IGS})_{\alpha/\Gamma, \beta, b} \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta(t)\nabla^2 f(x(t))\dot{x}(t) + b(t)\nabla f(x(t)) = 0.$$

Suppose that $\alpha > 1$, and that the following growth conditions are satisfied: for $t \geq t_0$

$$(G_2) \quad b(t) > \beta(t) + \frac{\beta(t)}{t};$$

$$(G_3) \quad tw(t) \leq (\alpha - 3)w(t).$$
Then, \( w(t) := b(t) - \dot{\beta}(t) - \frac{\beta(t)}{t} \) is positive and
\[
a) \quad f(x(t)) - \min f = \mathcal{O}\left(\frac{1}{t^2 w(t)}\right) \quad \text{as} \quad t \to +\infty; \\
b) \quad \int_{t_0}^{+\infty} t \left( (\alpha - 3) w(t) - t \dot{w}(t) \right) (f(x(t)) - \min f) dt < +\infty; \\
c) \quad \int_{t_0}^{+\infty} t^2 \beta(t) w(t) \|\nabla f(x(t))\|^2 dt < +\infty.
\]

**Proof.** Take \( \theta(t) = \Gamma \gamma(t)^2 \), \( \sigma(t) = \frac{1}{\Gamma \gamma(t)} \), \( \xi(t) \equiv 0 \), and
\[
c(t)^2 = \frac{1}{(\alpha - 1)^2} \frac{t}{b(t)} \left( t \beta(t) - \beta(t) - t \dot{\beta}(t) \right).
\]
(17)

This formula for \( c(t) \) will appear naturally during the calculation. Note that the condition \((G_2)\) ensures that the second member of the above expression is positive, which makes sense to think of it as a square. Let us verify that the conditions \((i)\) and \((iv), (v), (vi), (vii)\) are satisfied. This is a direct consequence of the formula (11) and the condition \((G_2)\):
\[
(i) \quad \frac{d}{dt}(\beta \theta \sigma) - \theta b \sigma = \frac{d}{dt} \beta \Gamma - \Gamma b = \frac{1}{\alpha - 1} \left( \frac{d}{dt}(t \beta) - tb \right) = \frac{1}{\alpha - 1} \left( \dot{\beta} + \frac{\beta}{t} - b \right) \leq 0.
\]
\[
(iv) \quad \frac{d}{dt}(\theta \sigma) + \theta \sigma (\sigma - \gamma) + \xi = \dot{\Gamma} + \Gamma (\frac{1}{t} - \gamma) = \dot{\Gamma} + 1 - \gamma \Gamma = 0.
\]
\[
(v) \quad Since \theta \sigma^2 \equiv 1 and \xi \equiv 1, we have \frac{d}{dt}(\theta \sigma^2 + \xi) = 0.
\]
\[
(vi) \quad \theta + 2(\sigma - \gamma)\theta = 2\dot{\Gamma} + 2(\Gamma - \gamma \Gamma^2) = 2\dot{\Gamma} + 1 - \gamma \Gamma = 0.
\]
\[
(vii) \quad \theta \beta + 2 \left( \dot{\beta} - b \right) \theta = 2\dot{\Gamma}(\beta \Gamma + (\dot{\beta} - b)\Gamma) = 2\dot{\Gamma}^2(\dot{\beta} + b + \frac{\beta}{t}) \leq 0.
\]
Let’s go to the conditions \((ii)\) and \((iii)\). The condition \((iii)\) gives the formula (17) for \( c(t) \). Then replacing \( c(t)^2 \) by this value in \((ii)\) gives the condition \((G_3)\). Note then that \( b(t)c(t)^2 = \frac{1}{(\alpha - 1)^2} t^2 \omega(t) \), which gives the convergence rate of the values
\[
f(x(t)) - \min f = \mathcal{O}\left(\frac{1}{t^2 w(t)}\right).
\]
Let us consider the integral estimate for the values. According to the definition (17) for \( c^2 b \) and the definition of \( w \), we have
\[
\theta b \sigma - \frac{d}{dt} \left( c^2 b + \beta \theta \sigma \right) = \frac{1}{\alpha - 1} tb - \frac{d}{dt} \left( \frac{1}{\alpha - 1} t^2 w + \frac{1}{\alpha - 1} t \beta \right)
= \frac{t}{(\alpha - 1)^2} \left( (\alpha - 1) b - 2 w - t \dot{w} - (\alpha - 1)(\dot{\beta} + \frac{\beta}{t}) \right)
= \frac{t}{(\alpha - 1)^2} \left( (\alpha - 3) w - t \dot{w} \right).
\]
According to Theorem 2.1 (ii)
\[
\int_{t_0}^{+\infty} t \left( (\alpha - 3) w(t) - t \dot{w}(t) \right) (f(x(t)) - \min f) dt < +\infty.
\]
Moreover, since \( \theta \sigma^2 = 1 \), the formula giving the weighting coefficient \( q(t) \) in the integral formula simplifies, and we get
\[
q(t) = \theta(t) \sigma(t) \beta(t) \left( \frac{\sigma(t) \beta(t)}{\sigma^2(t)} + \frac{b(t)}{\sigma(t)} - \frac{\dot{\beta}(t)}{\sigma(t)} \right)
\]
\[
= \beta(t) \Gamma_\gamma(t) \left( -\beta(t) \dot{\Gamma}_\gamma(t) + b(t) \Gamma_\gamma(t) - \dot{\beta}(t) \Gamma_\gamma(t) \right)
\]
\[
= \beta(t) \Gamma_\gamma(t)^2 \omega(t).
\]
According to Theorem 2.1 (iii)
\[
\int_{t_0}^{+\infty} t^2 \beta(t) w(t) \| \nabla f(x(t)) \|^2 \, dt < +\infty
\]
which gives the announced convergence rates. \( \square \)

**Remark 2.** Take \( \beta = 0 \). Then, according to the definition (16) of \( w \), we have \( w = b \), and the conditions of Theorem 3.2 reduce to
\[
t \dot{b}(t) - (3 - \alpha) b(t) \leq 0 \quad \text{for} \quad t \in [t_0, +\infty[.
\]
We recover the condition introduced in [11, Corollary 3.4]. Under this condition, each solution trajectory \( x \) of
\[
(\text{IGS})_{\alpha/\tau,0,b} \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + b(t) \nabla f(x(t)) = 0,
\]
satisfies
\[
f(x(t)) - \min_{\mathcal{H}} f = O \left( \frac{1}{t^2 b(t)} \right) \quad \text{as} \quad t \to +\infty.
\]

### 3.3. The case \( \gamma(t) = \frac{\alpha}{\tau} \), \( \beta \) constant

Due to its practical importance, consider the case \( \gamma(t) = \frac{\alpha}{\tau} \), \( \beta(t) \equiv \beta \) where \( \beta \) is a fixed positive constant. In this case, the dynamic \( (\text{IGS})_{\gamma,\beta,b} \) is written as follows
\[
\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + b(t) \nabla f(x(t)) = 0. \tag{18}
\]
The set of conditions \((G_2), (G_3)\) boils down to: for \( t \geq t_0 \)
\[
(G_2) \quad b(t) > \frac{\beta}{t},
\]
\[
(G_3) \quad t w(t) \leq (\alpha - 3) w(t),
\]
where \( w(t) = b(t) - \frac{\beta}{t} \). Therefore, \( b(\cdot) \) must satisfy the differential inequality
\[
t \frac{d}{dt} \left( b(t) - \frac{\beta}{t} \right) \leq (\alpha - 3) \left( b(t) - \frac{\beta}{t} \right).
\]
Equivalently
\[
t \frac{d}{dt} b(t) - (\alpha - 3) b(t) + \beta (\alpha - 2) \frac{1}{t} \leq 0.
\]
Let us integrate this linear differential equation. Set \( b(t) = k(t) t^{\alpha-3} \) where \( k(\cdot) \) is an auxiliary function to determine. We obtain
\[
\frac{d}{dt} \left( k(t) - \frac{\beta}{t^{\alpha-2}} \right) \leq 0,
\]
which gives \( k(t) = \frac{\beta}{t^{\alpha-2}} + d(t) \) with \( d(\cdot) \) nonincreasing. Finally, \( b(t) = \frac{\beta}{t} + d(t) t^{\alpha-3} \), with \( d(\cdot) \) a nonincreasing function to be chosen arbitrarily. In summary, we get the following result:
Proposition 1. Let \( x : [t_0, +\infty] \rightarrow \mathcal{H} \) be a solution trajectory of
\[
\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \frac{(\frac{\beta}{t} + d(t))}{t} \nabla f(x(t)) = 0
\] (19)
where \( d(\cdot) \) is a nonincreasing positive function. Then, the following properties are satisfied:

a) \( f(x(t)) - \min_{\mathcal{H}} f = O \left( \frac{1}{t^{\alpha-1}d(t)} \right) \) as \( t \to +\infty \);

b) \( \int_{t_0}^{+\infty} -\dot{d}(t) t^{\alpha-1} (f(x(t)) - \inf_{\mathcal{H}} f) dt < +\infty \);

c) \( \int_{t_0}^{+\infty} t^{\alpha-1} d(t) \| \nabla f(x(t)) \|^2 dt < +\infty \).

Proof. According to the definition of \( w(t) \) and \( b(t) \), we have the equalities
\( t^2 w(t) = t^2 \left( b(t) - \frac{\beta}{t} \right) = t^2 d(t) t^{\alpha-3} = t^{\alpha-1} d(t) \). Then apply Theorem 3.2. \( \square \)

3.4. Particular cases. According to Theorem 3.2 and Proposition 1, let us discuss the role and the importance of the scaling coefficient \( b(t) \) in front of the gradient term.

a) The first inertial dynamic system based on the Nesterov method, and which includes a damping term driven by the Hessian, was considered by Attouch, Pey-pouquet, and Redont in [17]. This corresponds to \( b(\cdot) \equiv 1 \), which gives:
\[
\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) = 0.
\]
In this case, we have \( w(t) = 1 - \frac{\beta}{t} \), and we immediately get that \((G_2), (G_3)\) are satisfied by taking \( \alpha > 3 \) and \( t > \beta \). This corresponds to take \( d(t) = \frac{1}{t^{\alpha-2}} - \frac{\beta}{t^{\alpha-3}} \), which is nonincreasing when \( t \geq \frac{\alpha}{\alpha-3} \).

Corollary 1. [17, Theorem 1.10, Proposition 1.11] Suppose that \( \alpha > 3 \) and \( \beta > 0 \). Let \( x : [t_0, +\infty] \rightarrow \mathcal{H} \) be a solution trajectory of
\[
\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) = 0.
\] (20)
Then,

a) \( f(x(t)) - \min_{\mathcal{H}} f = O \left( \frac{1}{t^2} \right) \) as \( t \to +\infty \);

b) \( \int_{t_0}^{+\infty} t(f(x(t)) - \inf_{\mathcal{H}} f) dt < +\infty \);

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

b) Another important situation is obtained by taking \( d(t) = \frac{1}{t^{\alpha-2}} \). This is the limiting case where the following two properties are satisfied: \( d(\cdot) \) is nonincreasing, and the coefficient of \( \nabla f(x(t)) \) is bounded. This offers the possibility of obtaining similar results for the explicit temporal discretized dynamics, that is to say the gradient algorithms. Precisely, we obtain the dynamic system considered by Shi, Du, Jordan, and Su in [31], and Attouch, Chbani, Fadili, and Riahi in [8].
Corollary 2. [8, Theorem 3], [31, Theorem 5]
Suppose that \( \alpha \geq 3 \). Let \( x : [t_0, +\infty) \to \mathcal{H} \) be a solution trajectory of

\[
\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \left(1 + \frac{\beta}{t}\right) \nabla f(x(t)) = 0. \tag{21}
\]

Then, the conclusions of Theorem 3.2 are satisfied:

a) \( f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right) \) as \( t \to +\infty \);

b) When \( \alpha > 3 \), \( \int_{t_0}^{+\infty} t(f(x(t)) - \inf_{\mathcal{H}} f)dt < +\infty \);

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

Note that (21) has a slight advantage over (20): the growth conditions are valid for \( t > 0 \), while for (20) one has to take \( t > \beta \). Accordingly, the estimates involve the quantity \( \frac{1}{t^2} \) instead of \( \frac{1}{t^{2(1-\frac{3}{2})}} \).

c) Take \( d(t) = \frac{1}{t^2} \) with \( s > 0 \). According to Proposition 1, for any solution trajectory \( x : [t_0, +\infty) \to \mathcal{H} \) of

\[
\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \left(\frac{\beta}{t} + t^{\alpha-3-s}\right) \nabla f(x(t)) = 0 \tag{22}
\]

we have:

a) \( f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^{2(\alpha-1-s)}}\right) \) as \( t \to +\infty \);

b) \( \int_{t_0}^{+\infty} t^{\alpha-s-2}(f(x(t)) - \inf_{\mathcal{H}} f)dt < +\infty, \int_{t_0}^{+\infty} t^{\alpha-s-1} \||\nabla f(x(t))||^2 dt < +\infty \).

4. Results based on the function \( p_\gamma \). In this section, we examine another set of growth conditions for the damping and rescaling parameters that guarantee the existence of solutions to the system \((i)-(vii)\) of Theorem 2.1. In the following theorems, the Lyapunov analysis and the convergence rates are formulated using the function \( p_\gamma : [t_0, +\infty) \to \mathbb{R}_+ \) defined by

\[
p_\gamma(t) := \exp\left(\int_{t_0}^{t} \gamma(s)ds \right).
\]

In Theorems 3.1 and 3.2, in line with the previous articles devoted to these questions (see [4], [7], [11]), the convergence rate of the values was formulated using the function \( \Gamma_\gamma(t) = p_\gamma(t) \int_{t_0}^{+\infty} \frac{1}{p_\gamma(s)}ds \). In fact, each of the two functions \( p_\gamma \) and \( \Gamma_\gamma \) captures the properties of the viscous damping coefficient \( \gamma(\cdot) \), but their growths are significantly different. To illustrate this, in the model case \( \gamma(t) = \alpha \frac{\alpha}{t}, \alpha > 1 \), we have \( p_\gamma(t) = \left(\frac{\alpha}{t}\right)^\alpha \), while \( \Gamma_\gamma(t) = \frac{t}{\alpha-1} \). Therefore, \( p_\gamma \) grows faster than \( \Gamma_\gamma \) as \( t \to +\infty \), and we can expect to get better convergence rates when formulating them using \( p_\gamma \). Moreover, \( p_\gamma \) makes sense and allows to analyze the case \( \alpha \leq 1 \), while \( \Gamma_\gamma \) does not. Thus, we will see that the approach based on \( p_\gamma \) provides results that cannot be captured by the approach based on \( \Gamma_\gamma \). To illustrate this, we start with a simple situation, then we consider the general case.
4.1. **A model situation.** Consider the system

\[
(\text{IGS})_{\gamma,0,b} \quad \ddot{x}(t) + \gamma(t)\dot{x}(t) + b(t)\nabla f(x(t)) = 0
\]

with \( \gamma(t) = \gamma_0(t) + \frac{1}{p_0(t)} \) and \( p_0(t) = \exp \left( \int_{t_0}^t \gamma_0(s) \, ds \right) \).

Choose

\[
\xi \equiv 0, \quad c(t) = p_0(t), \quad \sigma(t) = \frac{1}{p_0(t)}; \quad \theta(t) = p_0(t)^2.
\]

According to \( \dot{p}_0(t) = \gamma_0(t)p_0(t) \), we can easily verify that the conditions (i), (iii) – (vii) of Theorem 2.1 are satisfied, and (ii) becomes \( \frac{d}{dt} \left( p_0(t)^2b(t) \right) - p_0(t)b(t) \leq 0 \).

Then, a direct application of Theorem 2.1 gives the following result.

**Theorem 4.1.** Suppose that for all \( t \geq t_0 \)

\[
p_0(t)\dot{b}(t) + \left( 2\gamma_0(t)p_0(t) - 1 \right)b(t) \leq 0.
\]

Let \( x : [t_0, +\infty[ \to \mathcal{H} \) be a solution trajectory of \((\text{IGS})_{\gamma,0,b} \). Then, as \( t \to +\infty \)

\[
f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O} \left( \frac{1}{p_0(t)^2 b(t)} \right).
\]

Moreover,

\[
\int_{t_0}^{+\infty} p_0(t) \left( 1 - 2\gamma_0(t)p_0(t) - p_0(t)b(t) \right) (f(x(t)) - \min_{\mathcal{H}} f) \, dt < +\infty.
\]

**Remark 3.** Let us rewrite the linear differential inequality (23) as follows:

\[
\frac{\dot{b}(t)}{b(t)} \leq \frac{1}{p_0(t)} - 2\frac{\dot{p}_0(t)}{p_0(t)}.
\]

A solution corresponding to equality is \( b(t) = p_0(t)^{-2} \exp \left[ \int_{t_0}^t \left( \frac{1}{|p_0(s)|} \right) \, ds \right] \).

In the case \( \gamma_0(t) = \frac{a}{t}, \ 0 < a < 1, \ t_0 = 1, \) we have \( p_0(t) = t^\alpha, \) which gives

\[
b(t) = t^{-2\alpha} \exp \left[ \frac{t^{1-\alpha} - 1}{1-\alpha} \right].
\]

Therefore, for \( 0 < a < 1, \) and for this choice of \( b, \) (24) gives

\[
f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O} \left( \frac{1}{\exp \left[ \frac{t^{1-\alpha} - 1}{1-\alpha} \right]} \right).
\]

Thus, we obtain an exponential convergence rate in a situation that cannot be covered by the \( \Gamma_\gamma \) approach.

4.2. **The general case, with the Hessian-driven damping.**

**Theorem 4.2.** Let \( f : \mathcal{H} \to \mathbb{R} \) be a convex function of class \( C^1 \) such that \( \text{argmin}_{\mathcal{H}} f \neq \emptyset. \) Suppose that \( \gamma(\cdot), \beta(\cdot) \) are \( C^1 \) functions and \( b(\cdot) \) is a \( C^2 \) function which is nondecreasing. Suppose that \( r \) and \( m \) are positive parameters which satisfy \( 0 < r \leq \frac{1}{3} \) and \( 2r \leq m \leq 1 - r. \) Suppose that the following growth conditions are satisfied:

for \( t \geq t_0 \)

\[
(H_1) \quad \xi_0(t) \geq 0;
\]

\[
(H_2) \quad \xi_0(t)(2\sigma(t) - (m + r)\gamma(t)) - \frac{1}{2} \xi_0(t) + (m - (1 - r))\gamma(t)\sigma^2(t) \geq 0;
\]

\[
(H_3) \quad b(t) - \beta(t) + \beta(t)(\sigma(t) - \gamma(t)) \geq 0,
\]

\[
(H_4) \quad \frac{d}{dt} \left( \theta(w + \beta \sigma) \right)(t) - \theta(t)b(t)\sigma(t) \leq 0.
\]
Also, according to (34), we have
\[
\dot{x}(t) := (1 - 2(r + m))\gamma(t) + \sigma(t)\dot{\sigma}(t); \tag{26}
\]
\[
\sigma(t) := m\gamma(t) + \frac{1}{3}\frac{\dot{b}(t)}{b(t)}; \tag{27}
\]
\[
w(t) := b(t) - \dot{\beta}(t) + \beta(t)\sigma(t) + (1 - 2r - 2m)\gamma(t)\beta(t); \tag{28}
\]
\[
\theta(t) := p_\gamma(t)\gamma(t) - \frac{\dot{\sigma}(t)}{\dot{\beta}(t)}. \tag{29}
\]

Then, for each solution trajectory of \( x : [t_0, +\infty] \rightarrow \mathcal{H} \) of (IGS)\( \gamma, \beta, b \), we have,

\[
a) \quad f(x(t)) - \min_H f = O \left( \frac{1}{p_\gamma(t)^{3\gamma}w(t)b(t)^{\frac{3}{2}}} \right) \text{ as } t \rightarrow +\infty; \tag{30}
\]
\[
b) \quad \int_0^{+\infty} p_\gamma^2(t)\Upsilon(t) \left( f(x(t)) - \inf_H f \right) dt < +\infty; \tag{31}
\]
\[
c) \quad \int_0^{+\infty} \left( p_\gamma^2(t)b^{\frac{3}{2}}(t)\beta(t) - \frac{d}{dt}(p_\gamma^2 b^{-\frac{3}{2}}\beta^2)(t) \right) ||\nabla f(x(t))||^2 dt < +\infty. \tag{32}
\]

Here \( \Upsilon(t) := (3\sigma(t) - 2(r + m)\gamma(t))w(t) - \dot{w}(t) - 2(1 + r - m)\gamma(t). \)

Proof. According to Theorem 2.1, it suffices to show that, under the hypothesis (\( \mathcal{H}_1 \)) - (\( \mathcal{H}_4 \)), there exists \( c, \theta, \sigma, \xi \) which satisfy the conditions (i) - (vii) of Theorem 2.1. To perform the corresponding derivative calculation, let’s start by establishing some preliminary results.

- \( \ln p_\gamma(t) = \int_{t_0}^t \gamma(s)ds \), which by derivation gives \( \dot{p}_\gamma = \gamma \), that is to say \( p_\gamma = \gamma p_\gamma \).

- According to the definition of \( \sigma \),

\[
\frac{d}{dt} \left( p_\gamma^2 b^{-\frac{3}{2}} \right) = 2p_\gamma^2 b^{-\frac{3}{2}} \left( r\gamma - \frac{1}{3}\frac{\dot{b}}{b} \right) \tag{33}
\]

\[
= 2\theta ((r + m)\gamma - \sigma). \tag{34}
\]

Let us show that the following choice of the unknown parameters \( c, \theta, \sigma, \xi \) satisfies the conditions (i) - (vii) of Theorem 2.1:

\[
\theta := p_\gamma^2 b^{-\frac{3}{2}}, \quad \sigma := m\gamma + \frac{1}{3}\frac{\dot{b}}{b}, \quad \xi := \theta\xi_0,
\]

and

\[
c^2b := \theta w := \theta \left( b - \dot{\beta} + \beta\sigma + (1 - 2r - 2m)\gamma\beta \right). \tag{35}
\]

where \( \xi_0 \) has been defined in (26). We underline that under condition (\( \mathcal{H}_3 \)),

\[
c^2b = \theta \left( b - \dot{\beta} + \beta\sigma - \beta\gamma \beta + 2(1 - r - m)\gamma\beta \right) \geq 0.
\]

Also, according to (34), we have \( \dot{\theta} = 2\theta ((r + m)\gamma - \sigma). \)

\[
(i) \quad \frac{d}{dt}(\theta\beta) \sigma - \theta\sigma = \dot{\beta}\sigma + \beta(\dot{\beta}\sigma + \theta\dot{\sigma}) - b\theta\sigma
\]

\[
= \theta \left[ \dot{\beta}\sigma + 2(b\gamma - \beta\sigma)\beta\sigma + \beta\dot{\sigma} - b\sigma \right] \text{ because } \dot{\theta} = 2((r + m)\gamma - \sigma)
\]

\[
= \theta \left( -\beta\xi_0 - \sigma(b - \dot{\beta} + \beta\sigma - \gamma\beta) \right). \tag{36}
\]
According to the assumption \(m\) (see (18)), we have
\[
\frac{d}{dt} (\beta \sigma) - \theta b \sigma = \theta \left( -\beta \xi_0 - \sigma (b - \hat{\beta} + \beta \sigma - \gamma \beta) \right) \geq 0.
\]

(ii) According to the derivation chain rule and (H4), we conclude that
\[
\frac{d}{dt} (\xi) + \frac{d}{dt} (\beta \sigma) - \theta b \sigma + \frac{d}{dt} (\beta \sigma) - \theta b \sigma = \frac{d}{dt} (\theta w + \beta \sigma) - \theta b \sigma \leq 0.
\]

(iii) \(b(c^2 - \theta) + \beta \theta (\sigma - \gamma) + \frac{d}{dt} (\beta \theta) = 0\) results from (35).

(iv) According to the derivation chain rule, (33), and the definition of \(\sigma\)
\[
\frac{d}{dt} (\theta \sigma + \theta \sigma (\sigma - \gamma) + \xi) = \theta \sigma + \theta \sigma + \theta \sigma (\sigma - \gamma) + \xi
\]
\[
= 2 \theta \sigma ((r + m) \gamma - \sigma) + \theta \sigma + \theta \sigma (\sigma - \gamma) + \xi
\]
\[
= \theta \left( \sigma - \sigma ((1 - 2r - 2m) \gamma + \sigma) \right) + \xi.
\]

For this quantity to be equal to zero, we therefore take \(\xi = \theta \xi_0\), where \(\xi_0\) is defined in (27).

(v) According to our choice \(\xi = \theta \xi_0\), we have
\[
(v) \iff \frac{d}{dt} (\theta (\sigma^2 + \xi_0)) \leq 0.
\]

Let’s compute this quantity. According to the derivation chain rule and (34)
\[
\frac{d}{dt} (\theta (\sigma^2 + \xi_0)) = \left( \theta \dot{\xi}_0 + \theta (\sigma^2 + \xi_0) + 2 \theta \dot{\sigma} \right)
\]
\[
= 2 \theta \left( \frac{1}{2} \ddot{\xi}_0 + (\sigma^2 + \xi_0) ((r + m) \gamma - \sigma) + \dot{\sigma} \right)
\]
\[
= 2 \theta \left( \frac{1}{2} \ddot{\xi}_0 + \xi_0 ((r + m) \gamma - 2 \sigma) + \sigma (\xi_0 + \sigma ((r + m) \gamma - \sigma) + \dot{\sigma}) \right).
\]

By definition of \(\xi_0\), we have \(\xi_0 + \dot{\sigma} = \left( (1 - 2(r + m)) \gamma + \sigma \right) \sigma\). Therefore
\[
\frac{d}{dt} (\theta (\sigma^2 + \xi_0)) = 2 \theta \left( \frac{1}{2} \ddot{\xi}_0 + \xi_0 ((r + m) \gamma - 2 \sigma) + \sigma^2 (1 - (r + m)) \right).
\]

So, (v) is satisfied under the condition
\[
\frac{1}{2} \ddot{\xi}_0 + \xi_0 ((r + m) \gamma - 2 \sigma) + \sigma^2 (1 - (r + m)) \leq 0,
\]
which is precisely (H2).

(vi) Let’s compute
\[
\dot{\theta} + 2 (\sigma - \gamma) \theta = 2 \theta \left( r \gamma - \frac{1}{3} \frac{\dot{b}}{b} (m - 1) \gamma + \frac{1}{3} \frac{\dot{b}}{b} \right)
\]
\[
= 2 (r + m - 1) \theta \gamma.
\]

According to the assumption \(m \leq 1 - r\), this quantity is less or equal than zero.
We have (vii) ⇐⇒ \( \beta (\dot{\theta} + 2 (\dot{\beta} - b) \dot{\theta}) \leq 0 \).

According to condition \( H_3 \) and the assumption \( m \leq 1 - r \), we conclude
\[
\beta \dot{\theta} + 2 (\dot{\beta} - b) \dot{\theta} = 2 \theta (\beta (r + m) \gamma - \beta \sigma - b) \\
= 2 \theta \left[ - (b - \dot{\beta} + \beta \sigma - \gamma \sigma) - \beta \gamma \left( 1 - r - m \right) \right] \leq 0.
\]

So, (vii) is satisfied. According to Theorem 2.1, we obtain (30)-(31)-(32), which completes the proof.

4.3. The case without the Hessian. Let us specialize the previous results in the case \( \beta = 0 \), i.e. without the Hessian:

\( (IGS)_{\gamma,0,b} \)
\[\ddot{x}(t) + \gamma(t) \dot{x}(t) + b(t) \nabla f(x(t)) = 0.\]

**Theorem 4.3.** Suppose that the conditions \((H_1)\) and \((H_2)\) of Theorem 4.2 are satisfied. Then, for each solution trajectory \( x : [t_0, + \infty] \rightarrow \mathcal{H} \) of \((IGS)_{\gamma,0,b}\), we have, as \( t \rightarrow + \infty \)
\[f(x(t)) - \min_{\mathcal{H}} f = O \left( \frac{1}{p_r(t)^{2r} b(t)^{2r}} \right).\] (37)

Moreover, when \( m > 2r \)
\[\int_{t_0}^{+ \infty} p_r(t)^{2r} b(t)^{2r} \gamma(t) \left( f(x(t)) - \inf_{\mathcal{H}} f \right) dt < + \infty.\] (38)

**Proof.** Conditions \((H_1)\) and \((H_2)\) in Theorem 4.2 remain unchanged since they are independent of \( \beta \). We just need to verify \((H_4)\), because \((H_3)\) is written \( b(t) \geq 0 \) and becomes obvious. Since \( \beta = 0 \), we have \((H_4) \iff \frac{d}{dt} (\theta(t)b(t) - \theta(t)b(t) \sigma(t)) \leq 0 \).

According to
\[
\frac{d}{dt} (\theta(t)b(t) - \theta(t)b(t) \sigma(t)) = \theta(t)b(t) + \theta(t) \dot{b}(t) - \theta(t)b(t) \left( m \gamma + \frac{\dot{b}(t)}{3b(t)} \right)
\]
\[= b(t)^{1/3} \left[ \frac{d}{dt} \left( \theta(t)b(t)^{2/3} \right) - m \gamma(t) \left( \theta(t)b(t)^{2/3} \right) \right]
\]
\[= b(t)^{1/3} \left[ \frac{d}{dt} (p_r(t)^{2r}) - m \gamma(t) (p_r(t)^{2r}) \right]
\]
\[= (2r - m) \gamma(t) (b(t))^{2r} p_r(t)^{2r} \leq 0 \quad \text{since} \ 2r \leq m,
\]
we conclude that \((H_4)\) holds, which completes the proof.

Next, we show that the condition \((H_2)\) on the coefficients \( \gamma(\cdot) \) and \( b(\cdot) \) can be formulated in simpler form which is useful in practice.

**Theorem 4.4.** The conclusions of Theorem 4.3 remain true when we replace \((H_2)\) by
\((H_2^+)\): \( \sigma(t) \left( (r + m) \gamma(t) \right) \left( 2 \sigma(t) + (1 - 2(r + m)) \gamma(t) \right) + \frac{1}{2} \dot{\sigma}(t) \geq 0 \), and assume moreover that \( b(\cdot) \) is log-concave, i.e., \( \frac{d}{dt} (\ln(b(t))) \leq 0 \).

**Proof.** According to Theorem 4.3, it suffices to show that \((H_2)\) is satisfied under the hypothesis \((H_2^+)\). By definition of \( \sigma \), we have
\[(2 \sigma(t) - (m + r) \gamma(t)) = \left( (m - r) \gamma(t) + \frac{2 \dot{b}(t)}{3 b(t)} \right). \]
So \((H_2)\) can be written equivalently as \(A \geq 0\), where
\[
A := \xi_0(t) \left( (m - r)\gamma(t) + \frac{2 \dot{b}(t)}{3 \ddot{b}(t)} \right) - \frac{1}{2} \dot{\xi}_0(t) + \left( m + r - 1 \right) \gamma(t)\sigma^2(t). \tag{39}
\]
A calculation similar to the one above gives
\[
\xi_0(t) = (1 - 2r - m)\gamma(t) - m\gamma(t) + \sigma(t)\sigma(t) - \dot{\sigma}(t),
\]
\[
= (1 - 2r - m)\gamma(t) + \frac{1}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \sigma(t) - \dot{\sigma}(t). \tag{40}
\]
In (39), let’s replace \(\xi_0(\cdot)\) by its formulation (40), we obtain
\[
A = \frac{1}{2} \frac{d^2}{dt^2} \sigma(t) - \frac{1}{2} \frac{d}{dt} \left[ \sigma(t) \left( (1 - 2r - m)\gamma(t) + \frac{1}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right) \right]
- \dot{\sigma}(t) \left( (m - r)\gamma(t) + \frac{2}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right) + \left( m + r - 1 \right) \gamma(t)\sigma^2(t)
+ \left( (m - r)\gamma(t) + \frac{2}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right) \left( (1 - 2r - m)\gamma(t) + \frac{1}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right) \sigma(t).
\]
Set
\[
B := \left( m + r - 1 \right) \gamma(t)\sigma^2(t) + \left( (m - r)\gamma(t) + \frac{2}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right) \left( (1 - 2r - m)\gamma(t) + \frac{1}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right) \sigma(t),
\]
then we have (by omitting the variable \(t\) to shorten the formulas)
\[
B = \sigma \left[ (m + r - 1)\gamma + \left( m - r \right)\gamma(t) + \frac{2}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right] \left( (1 - 2r - m)\gamma + \frac{1}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right)
= \sigma \left[ (m + r - 1)\gamma + \frac{1}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right] \left( (1 - 2r)\gamma + \frac{2}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right)
= \sigma \left[ -r\gamma + \frac{1}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right] \left( (1 - 2r)\gamma + \frac{2}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right).
\]
Replacing \(B\) in \(A\), we obtain
\[
A = \sigma(t) \left( \sigma(t) - (m + r)\gamma(t) \right) \left( 2\sigma(t) + (1 - 2(m + r))(\gamma(t) + \frac{1}{2} \frac{d^2}{dt^2} \sigma(t) + C(t) \tag{41}\right)
\]
where
\[
C(t) := -\dot{\sigma}(t) \left( (m - r)\gamma(t) + \frac{2}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right) - \frac{1}{2} \frac{d}{dt} \left[ \sigma(t) \left( (1 - 2r - m)\gamma(t) + \frac{1}{3} \frac{\dot{b}(t)}{\ddot{b}(t)} \right) \right].
\]
Let us analyze the sign of the above quantity, which is the same as exponential growth. Take comparing the two approaches. Example, we will show that Theorem 4.4 better captures the case where growth conditions on the coefficients of the dynamic. Precisely, using the following

Here let us show that the condition \((H_2)\) of Theorem 3.1:

\[
\begin{align*}
(\gamma(t) - m + r)\gamma(t) & \geq \sigma(t) - (m + r)\gamma(t) \\
& \geq 2\sigma(t) + (1 - 2(m + r))\gamma(t) \\
& \geq \frac{1}{2}d^2\sigma(t).
\end{align*}
\]

The condition \((H_2^+)\) expresses that the second member of the above inequality is nonnegative. Therefore \((H_2^+)\) implies \((H_2)\), which gives the claim.

4.4. Comparing the two approaches. As we have already underlined, Theorems 3.1 and 4.4 are based on the Lyapunov analysis of the dynamic \((IGS)_{\gamma,0,b}\) using the functions \(\Gamma_\gamma\) and \(p_\gamma\), respectively. As such, they lead to significantly different growth conditions on the coefficients of the dynamic. Precisely, using the following example, we will show that Theorem 4.4 better captures the case where \(b\) has an exponential growth. Take

\(b(t) = e^{\mu ty}\) and \(\gamma(t) = \frac{\alpha}{t^{1-q}}\) with \(\alpha = \mu q > 0\), \(q \in ]0,1[\).

a) First, let us show that the condition \((H_2^+)\) of Theorem 4.4 is satisfied. We have

\[
\begin{align*}
\frac{1}{2}\sigma(t) + \sigma(t) & \geq 2\sigma(t) + (1 - 2(m + r))\gamma(t) \\
& \geq \frac{1}{2}d^2\sigma(t).
\end{align*}
\]

which is nonnegative because of the hypothesis \(r \leq \frac{1}{3}\) and \(q < 1\).

b) Let us now examine the growth condition used in Theorem 3.1:

\[
\Gamma(t)\dot{b}(t) \leq b(t)\left(3 - 2\gamma(t)\Gamma(t)\right) \quad \text{where } \Gamma(t) := p(t) \int_t^{+\infty} \frac{ds}{p(s)}.
\]

Here \(p(t) = e^{\mu(t^r - t_0)}\). Therefore \(\Gamma(t) = e^{\mu ty} \int_t^{+\infty} e^{-\mu s^y} ds\), which gives

\[
\Gamma(t)\dot{b}(t) - b(t)\left(3 - 2\gamma(t)\Gamma(t)\right) = 3e^{\mu ty} \left(\mu q t^{y-1} e^{\mu ty} \int_t^{+\infty} e^{-\mu s^y} ds - 1\right).
\]

Let us analyze the sign of the above quantity, which is the same as

\[
\mathcal{D}(t) := \mu q t^{y-1} e^{\mu ty} \int_t^{+\infty} e^{-\mu s^y} ds - 1
\]

\[
= -\mu q t^{y-1} e^{\mu ty} \int_t^{+\infty} \frac{ds}{ds} \left(e^{-\mu s^y}\right) \frac{1}{\mu q} s^{1-q} ds - 1
\]
After integration by parts, we get
\[ D(t) := \left( \frac{1}{q} - 1 \right) + \frac{1 - q}{q} t \left( q^{-1} e^{\mu t} \right) \int_t^{+\infty} e^{-\mu s} \frac{1}{s^q} ds > \left( \frac{1}{q} - 1 \right) > 0. \]
Therefore, the condition (42) is not satisfied.

5. Illustration of the results. Let us particularize our results in some important special cases, and compare them with the existing literature. We do not detail the proofs which result from the direct applications of the previous theorems and the classical differential calculus.

5.1. The case \( b(t) = p(t)^{3p_0} \). Recall that \( p(t) = \exp \left( \int_{t_0}^t \gamma(s) ds \right) \). We start with results in [7] concerning the rate of convergence of values in the case \( b(t) = c_0 p(t)^{3p_0} \) with \( p_0 \geq 0 \) and \( c_0 \geq 0 \). In this case, the system (IGS) becomes:

\[ \ddot{x}(t) + \gamma(t) \dot{x}(t) + c_0 \exp \left( 3p_0 \int_{t_0}^t \gamma(s) ds \right) \nabla f(x(t)) = 0. \] (43)

Observe that \( \dot{\gamma}(t) = p_0 \gamma(t) \) and \( \xi_0(t) = (m + p_0)(1 - 2r - (m + p_0)\gamma^2(t) - \gamma(t)) \).
Therefore, conditions (H1) and (H2) of Theorem 4.3 become after simplification:

\begin{align*}
(\text{H1}): \quad & [(p_0 - r) + (1 - r - m)]\gamma^2(t) - \gamma(t) \geq 0; \\
(\text{H2}): \quad & 2(p_0 - r)(1 + 2(p_0 - r))\gamma^3(t) - 2(1 + 3(p_0 - r)) \gamma(t)\gamma(t) + \gamma(t) \geq 0.
\end{align*}

Since \( m \leq 1 - r \), instead of (H1), it suffices to verify

\begin{align*}
(\text{H1}^+): \quad & (p_0 - r)\gamma^2(t) - \gamma(t) \geq 0.
\end{align*}

**Theorem 5.1.** Let \( \gamma : [t_0, +\infty[ \rightarrow \mathbb{R}_+ \) be a nonincreasing and twice continuously differentiable function. Suppose that there exists \( r \in \left[ 0, \frac{1}{3} \right] \) such that

\[ \dot{\gamma}(t) \geq 2\left[ \min(0, p_0 - r) \right]^2 \gamma^3(t) \quad \text{on} \quad [t_0, +\infty[. \] (44)

Then, for each solution trajectory \( x(\cdot) \) of (43), we have as \( t \to +\infty \)

\[ f(x(t)) - \min_{\mathcal{H}} f = O \left( \frac{1}{p(t)^{2r + p_0}} \right). \] (45)

**Proof.** To prove the claim, we use Theorem 4.2 and distinguish two cases:

* Suppose \( p - r \geq 0 \), then (44) implies \( \dot{\gamma}(t) \geq 0 \), and since \( \gamma \) is a nonincreasing, we also have \( \gamma(t) \leq 0 \); thus both conditions (H1) and (H2) are satisfied.

* Suppose \( p - r < 0 \), then (44) becomes

\[ \dot{\gamma}(t) \geq (2p - r)^2 \gamma^3(t) \quad \text{on} \quad [t_0, +\infty[. \] (46)

Since \( \gamma(\cdot) \) is a positive and nonincreasing, \( \lim_{t \to +\infty} \gamma(t) = \ell \) exists and is equal to zero. Otherwise, by integrating (46) on \([t_0, t]\) for \( t > t_0 \), we would have

\[ \dot{\gamma}(t) - 2 \int_{t_0}^t \gamma(s) ds \geq 2(p - r)^2 \ell^3(t - t_0). \]

This in turn gives \( \lim_{t \to +\infty} \dot{\gamma}(t) = +\infty \), which implies \( \lim_{t \to +\infty} \gamma(t) = +\infty \), that is a contradiction. Then, multiply (46) by \( \dot{\gamma}(t) \). Since \( \gamma(\cdot) \) is nonincreasing, we obtain

\[ \dot{\gamma}(t) \leq 2(p - r)^2 \gamma^3(t) \quad \text{and} \quad \frac{\dot{\gamma}(t)^2}{2} \leq 2(p - r)^2 \frac{d}{dt} \left( \gamma^4(t) \right). \]
By integrating this inequality from \( t \) to \( T > t \), we get
\[
\dot{\gamma}(T)^2 - \dot{\gamma}(t)^2 \leq (p - r)^2(\gamma^4(T) - \gamma^4(t)),
\]
Letting \( T \to +\infty \), and using \( \lim_{t \to +\infty} \gamma(T) = 0 \), we obtain \( \dot{\gamma}^2(t) \geq (p - r)^2\gamma^4(t) \), which is equivalent to \( |\dot{\gamma}(t)| \geq |p - r|\gamma^2(t) \). Since \( \dot{\gamma}(t) \leq 0 \) and \( p < r \), this gives \( -\dot{\gamma}(t) \geq (r - p)\gamma^2(t), \forall t > t_0 \), that is \( (H^+_1) \). We have
\[
(p - r)(1 - r - m)\gamma^2(t) - \dot{\gamma}(t)
\]
\[
= -2(p - r)^2\gamma^3(t) + \ddot{\gamma}(t) + 2(1 - 3r + 3p)\gamma(t)(p - r)\gamma^2(t) - \dot{\gamma}(t) \geq 0.
\]
Therefore, \( (H^+_1) \) and \( (H_2) \) are satisfied. Applying Theorem 4.2, we conclude. \( \square \)

As a particular case of Theorem 5.1, with \( p_0 = 0 \), we obtain the following result.

**Theorem 5.2.** [7, Theorem 2.1] Let \( \gamma(t) \) be a nonincreasing function of class \( C^2 \), and \( x(t) \) a solution trajectory of
\[
\ddot{x}(t) + \gamma(t)\dot{x}(t) + c_0\nabla f(x(t)) = 0.
\]
Suppose that
\[
(H_{r,\gamma}): \exists r > 0 \text{ such that } -2r^2\gamma^3(t) + \ddot{\gamma}(t) \geq 0 \text{ for } t \text{ large enough.}
\]
Then, \( f(x(t)) - \min_{\mathcal{H}} f = O\left(e^{-2\min(r, \frac{1}{2})\int_{t_0}^t \gamma(s)ds}\right) \) as \( t \to +\infty \).

**Remark 4.** The case \( \gamma(t) = \frac{1}{(\ln t)^\rho}, \) for \( 0 \leq \rho \leq 1 \), was developed in [7]. In that case condition \( (H_{3,\gamma}) \) writes as
\[
2(\ln t)^2 + 3\rho \ln t + \rho(\rho + 1) \geq 2r^2(\ln t)^2(1 - \rho),
\]
which is satisfied for any \( r \leq 1 \) and any \( \epsilon \neq 0 \).

\begin{itemize}
  \item If \( \rho = 1 \), then \( p(t) = \exp\left(\int_{t_0}^t \frac{1}{s(\ln s)^\rho}ds\right) = \exp\left(\int_{\ln t_0}^{\ln t} \frac{du}{x}\right) = \ln t \ln t_0^{-1} \),
  \end{itemize}
and for \( r = \frac{1}{3} \), we get \( f(x(t)) - \min_{\mathcal{H}} f = O\left(\frac{1}{(\ln t)^{\frac{5}{3}}}\right) \).

\begin{itemize}
  \item If \( 0 \leq \rho < 1 \), then \( p(t) = \exp\left(\int_{\ln t_0}^{\ln t} \frac{1}{w^\rho}dw\right) = \exp\left(\frac{1}{1 - \rho}\left((\ln t)^{1 - \rho} - (\ln t_0)^{1 - \rho}\right)\right)\)
  \end{itemize}
and, for \( r = \frac{1}{3} \), we also get \( f(x(t)) - \min_{\mathcal{H}} f = O\left(\exp\left(\frac{1}{2<(1-r)^{1-\rho}>}\ln t\right)\right) \).

5.2. **The case** \( b(t) = c_0t^\alpha \) and \( \gamma(t) = \frac{a}{t} \). When \( b(t) = c_0t^\alpha \) and \( \gamma(t) = \frac{a}{t} \), where \( \alpha > 0 \) and \( q \geq 0 \), we first observe that \( \gamma(t) = \exp\left(\int_{t_0}^t \frac{\gamma(s)ds}{t}\right) = \left(\frac{t}{t_0}\right)^{\alpha} \). The second-order continuous system becomes:
\[
\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + c_0\nabla f(x(t)) = 0.
\]
Applying Theorem 5.1, we obtain the following new result.

**Theorem 5.3.** Let \( x(t) \) be a solution trajectory of (48) with \( \alpha > 1 \) and \( q \geq 0 \). Suppose that \( 1 < \alpha \leq 3 + q \). Then,
\[
f(x(t)) - \min_{\mathcal{H}} f = O\left(\frac{1}{t^{(3+q)\alpha}}\right), \text{ as } t \to +\infty.
\]
**Remark 5.** Taking \( q = 0 \), a direct application of the above result covers the results obtained in [9, 33] (case \( \alpha \geq 3 \)), and in [3, 13], (case \( \alpha \leq 3 \)). It suffices to take \( \gamma(t) = \frac{\alpha}{2} \) and \( r = \frac{1}{2} \). More precisely, we get:

- if \( 0 < \alpha \leq 3 \) then \( f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}(t^{\frac{\alpha}{2}}) \),
- if \( \alpha > 3 \) then \( f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}(t^{\frac{1}{2}}) \).

5.3. **The case** \( b(t) = e^{\mu t} \) **and** \( \gamma(t) = \frac{\alpha}{1-t} \). Suppose that \( \mu \geq 0 \), \( 0 \leq q \leq 1 \) and \( \alpha > 0 \). This will allow us to obtain the following exponential convergence rate of the values.

**Theorem 5.4.** Let \( x : [t_0, +\infty[ \rightarrow \mathcal{H} \) be a solution trajectory of

\[
\ddot{x}(t) + \frac{\alpha}{t-\mu} \dot{x}(t) + e^{\mu t} \nabla f(x(t)) = 0.\tag{50}
\]

Suppose that \( \alpha \leq \mu q \), then, as \( t \rightarrow +\infty \)

\[
f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(e^{-\frac{2q + \alpha}{1-t}} t^{\frac{1}{2}}\right).
\]

**Remark 6.** a) For \( q = \mu = 0 \), (50) reduces to the system initiated in [33], i.e.

\[
\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla f(x(t)) = 0.
\]

Just assuming \( \alpha > 0 \), we obtain \( \lim_{t \rightarrow +\infty} \left(f(x(t)) - \min_{\mathcal{H}} f\right) = 0 \).

b) For \( q = \frac{1}{2} \) we get

- if \( \alpha \leq \mu \), then \( f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(e^{-\frac{2q + \alpha}{1-t}} \sqrt{t}\right) \),
- if \( \alpha \geq \mu \), then \( f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(e^{-2\mu \sqrt{t}}\right) \).

c) For \( q = 1 \), direct application of Theorem 5.4 gives:

**Corollary 3 (Linear convergence).** Let \( x : [t_0, +\infty[ \rightarrow \mathcal{H} \) be a solution trajectory of

\[
\ddot{x}(t) + \alpha \dot{x}(t) + e^{\mu t} \nabla f(x(t)) = 0.\tag{51}
\]

If \( \alpha \leq \mu \), then \( f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(e^{-\frac{2q + \alpha}{1-t}}\right) \).

**Remark 7.** The above exponential convergence rate results from the time scale effect. Similar results are valid for the scaled steepest descent \( \dot{x}(t) + e^{\mu t} \nabla f(x(t)) = 0 \).

Let us illustrate these results. Take \( f(x_1, x_2) := \frac{1}{2} (x_1^2 + x_2^2) - \ln(x_1 x_2) \), which is a strongly convex function. Trajectories of

\[
\ddot{x}(t) + \alpha \dot{x}(t) + e^{\mu t} \nabla f(x(t)) + ce^{\nu t} \nabla^2 f(x(t)) \dot{x}(t) = 0,
\]

corresponding to different values of the parameters \( \alpha, \mu, \nu, \) and \( c \), are plotted in Figure 1. From Scilab version 6.1.0 [http://www.scilab.org](http://www.scilab.org) as an open source software.
Figure 1. $t \mapsto f(x(t)) - \min f$ for solutions of (51), (52), $f(x_1, x_2) = \frac{1}{2} (x_1^2 + x_2^2) - \ln(x_1 x_2)$.

| $\gamma(t)$ | $\beta(t)$ | $b(t)$ | $f(x(t)) - \min f$ | Reference |
|-------------|-------------|--------|---------------------|-----------|
| Cte         | 0           | 1      | $O(t^{-1})$         | (1964) [30] |
| Cte         | Cte         | 1      | $O(t^{-1})$         | (2002) [2]  |
| $\alpha/t$ | 0           | 1      | $O(t^{-2})$ if $0 < \alpha \leq 3$ | (2019) [13] |
|             |             |        | $O(t^{-2})$ if $\alpha \geq 3$ | (2014) [33] |
| $\alpha/t$ | Cte         | 1      | $O(t^{-2})$ if $\alpha \geq 3, \beta > 0$ | (2016) [17] |
| $\gamma(t)$ | 0           | $b(t)$ | $O\left(\left(\frac{p(t)}{t} \int_0^t \gamma(s) ds\right)^{-1}\right)$ | (2019) [13] |
| $\alpha/t$ | $\beta(t)$ | $b(t)$ | $O\left(\left(t^2 b(t) - \beta(t) - \frac{\beta(t)}{t}\right)^{-1}\right)$ | (2020) [8] |

Figure 2. Convergence rate of $f(x(t)) - \min f$ for instances of Theorem 2.1 and general $f$.

5.4.1. **Strongly convex case.** Suppose that $f$ is $s$-strongly convex. Following Polyak’s [30], the system

$$\ddot{x}(t) + 2\sqrt{s} \dot{x}(t) + \nabla f(x(t)) = 0$$

provides the linear convergence rate $f(x(t)) - \inf_{x \in \mathcal{H}} f \leq C e^{-\sqrt{s} t}$, see also [32, Theorem 2.2]. In the presence of an additional Hessian-driven damping term

$$\ddot{x}(t) + 2\sqrt{s} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) = 0 \quad (\beta \geq 0)$$

a related linear rate of convergence can be found in [8, Theorem 7]. Let us insist on the fact that, in Corollary 3, we obtain a linear convergence rate for a general convex differentiable function $f$. In Figure 1, for the strongly convex function $f(x_1, x_2) = \frac{1}{2} (x_1^2 + x_2^2) - \ln(x_1 x_2)$, we can observe that some values of $\mu$ give a...
better speed of convergence of $f(x(t)) - \min f$. We can also note that for $\mu$ correctly set, the system (51) provides a better linear convergence rate than the system (52).

5.4.2. Non-strongly convex case. We illustrate our results on the following simple example of a non strongly convex minimization problem, with non unique solutions.

$$\min_{\mathbb{R}^2} f(x_1, x_2) = \frac{1}{2}(x_1 + 10^3 x_2)^2.$$  

From Figure 3 we get the following properties:

a) The convergence rate of the values is in accordance with Figure 2.

b) The system (51) is best for its linear convergence of values.

c) The Hessian-driven damping reduces the oscillations of the trajectories.

$$\min_{\mathbb{R}^2} f(x_1, x_2) = \frac{1}{2}(x_1 + 1000 x_2)^2.$$  

6. Conclusion, perspectives. Our study is one of the first works to simultaneously consider the combination of three basic techniques for the design of fast converging inertial dynamics in convex optimization: general viscous damping (and especially asymptotic vanishing damping in relation to the Nesterov accelerated gradient method), Hessian-driven damping which has a spectacular effect on the reduction of the oscillatory aspects (especially for ill-conditioned minimization problems), and temporal rescaling. We have introduced a system of equations-inequations whose solutions provide the coefficients of a general Lyapunov function for these dynamics. We have been able to encompass most of the existing results and find new solutions for this system, thus providing new Lyapunov functions. Also, we have been able to explain the mysterious coefficients which have been used in recent algorithmic developments, and which were just justified until now by the simplification of complicated calculations. Finally, by playing on fast rescaling methods, we have obtained linear convergence results for general convex functions. This work provides a basis for the development of corresponding algorithmic results. Many questions remain to be studied, in particular the convergence of trajectories, the introduction of error terms, the study of the case of a non-smooth $f$ potential, to name just a few of the important questions for algorithmic applications.
REFERENCES

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

[2] F. Alvarez, 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.* (9), 81 (2002), 747–779.

[3] V. Apidopoulos, J.-F. Aujol and C. Dossal, Convergence rate of inertial forward-backward algorithm beyond Nesterov’s rule, *Math. Program.*, 180 (2020), 137–156.

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

[5] H. Attouch and A. Cabot, Convergence rates of inertial forward-backward algorithms, *SIAM J. Optim.*, 28 (2018), 849–874.

[6] H. Attouch, A. Cabot, Z. Chbani and H. Riahi, Inertial forward-backward algorithms with perturbations: Application to Tikhonov regularization, *J. Optim. Theory Appl.*, 179 (2018), 1–36.

[7] H. Attouch, A. Cabot, Z. Chbani and H. Riahi, Rate of convergence of inertial gradient dynamics with time-dependent viscous damping coefficient, *Evol. Equ. Control Theory*, to appear.

[8] H. Attouch, Z. Chbani, J. Fadili and H. Riahi, First-order optimization algorithms via inertial systems with Hessian driven damping, *Math. Program.*, (2020), 353–371.

[9] H. Attouch, Z. Chbani, J. Peynot and P. Redont, Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity, *Math. Program.*, 168 (2018), 123–175.

[10] H. Attouch, Z. Chbani and H. Riahi, Convergence rates of inertial proximal algorithms with general extrapolation and proximal coefficients, *Vietnam J. Math.*, 48 (2020), 247–276.

[11] H. Attouch, Z. Chbani and H. Riahi, Fast convex optimization via time scaling of damped inertial gradient dynamics, *Pure Appl. Functional Anal.*, to appear.

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

[13] H. Attouch, Z. Chbani and H. Riahi, Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$, *ESAIM Control Optim. Calc. Var.*, 25 (2019), 34pp.

[14] 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), 1–34.

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

[16] H. Attouch and J. Peynot, The rate of convergence of Nesterov’s accelerated forward-backward method is actually faster than $1/k^2$, *SIAM J. Optim.*, 26 (2016), 1824–1834.

[17] H. Attouch, J. Peynot and P. Redont, Fast convex optimization via inertial dynamics with Hessian driven damping, *J. Differential Equations*, 261 (2016), 5734–5783.

[18] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.*, 2 (2009), 183–202.

[19] R. I. Boţ and E. R. Csetnek, Second order forward-backward dynamical systems for monotone inclusion problems, *SIAM J. Control Optim.*, 54 (2016), 1423–1443.

[20] R. I. Boţ, E. R. Csetnek and S. C. László, Approaching nonsmooth nonconvex minimization through second-order proximal-gradient dynamical systems, *J. Evol. Equ.*, 18 (2018), 1291–1318.

[21] R. I. Boţ, E. R. Csetnek and S. C. László, Tikhonov regularization of a second order dynamical system with Hessian damping, *Math. Program.*, 2020.

[22] O. Güler, New proximal point algorithms for convex minimization, *SIAM J. Optim.*, 2 (1992), 649–664.

[23] O. Güler, On the convergence of the proximal point algorithm for convex minimization, *SIAM J. Control Optim.*, 29 (1991), 403–419.

[24] A. Haraux, *Systèmes Dynamiques Dissipatifs et Applications*, Recherches en Mathématiques Appliquées, 17, Masson, Paris, 1991.

[25] A. Haraux and M. A. Jendoubi, *The Convergence Problem for Dissipative Autonomous Systems. Classical Methods and Recent Advances*, SpringerBriefs in Mathematics, BCAM
R. May, Asymptotic for a second-order evolution equation with convex potential and vanishing damping term, *Turkish J. Math.*, 41 (2017), 681–685.

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

Y. Nesterov, *Introductory Lectures on Convex Optimization. A Basic Course*, Applied Optimization, 87, Kluwer Academic Publishers, Boston, MA, 2004.

J. Peypouquet and S. Sorin, Evolution equations for maximal monotone operators: Asymptotic analysis in continuous and discrete time, *J. Convex Anal.*, 17 (2010), 1113–1163.

B. T. Polyak, Some methods of speeding up the convergence of iteration methods, *USSR Comput. Math. Math. Phys.*, 4 (1964), 1–17.

B. Shi, S. S. Du, M. I. Jordan and W. J. Su, Understanding the acceleration phenomenon via high-resolution differential equations, preprint, arXiv:1810.08907.

J. W. Siegel, Accelerated first-order methods: Differential equations and Lyapunov functions, preprint, arXiv:1903.05671v1.

W. Su, S. Boyd and E. J. Candès, A differential equation for modeling Nesterov’s accelerated gradient method: Theory and insights, *J. Mach. Learn. Res.*, 17 (2016), 43pp.

Received for publication December 2020.

E-mail address: hedy.attouch@umontpellier.fr
E-mail address: aichabalhag@gmail.com
E-mail address: chbaniz@uca.ac.ma
E-mail address: h-riahi@uca.ac.ma