Rate of convergence of inertial gradient dynamics with time-dependent viscous damping coefficient
Hedy Attouch, Alexandre Cabot, Zaki Chbani, Hassan Riahi

To cite this version:
Hedy Attouch, Alexandre Cabot, Zaki Chbani, Hassan Riahi. Rate of convergence of inertial gradient dynamics with time-dependent viscous damping coefficient. Evolution Equations and Control Theory, American Institute of Mathematical Sciences (AIMS), 2018, 7 (3), pp.353-371. <10.3934/eect.2018018>. <hal-01660062v2>

HAL Id: hal-01660062
https://hal.archives-ouvertes.fr/hal-01660062v2
Submitted on 7 Apr 2018

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

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
RATE OF CONVERGENCE OF INERTIAL GRADIENT DYNAMICS  
WITH TIME-DEPENDENT VISCOUS DAMPING COEFFICIENT.  

HEDY ATTOUCH, ALEXANDRE CABOT, ZAKI CHBANI, AND HASSAN RIAHI

Abstract. In a Hilbert space $\mathcal{H}$, we study the convergence properties when $t \to +\infty$ of the trajectories of the second-order differential equation

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

where $\nabla \Phi$ is the gradient of a convex continuously differentiable function $\Phi : \mathcal{H} \to \mathbb{R}$, and $\gamma(t)$ is a time-dependent positive viscous damping coefficient. This study aims to offer a unifying vision on the subject, and to complement the article by Attouch and Cabot (J. Diff. Equations, 2017). We obtain convergence rates for the values $\Phi(x(t)) - \inf_{\mathcal{H}} \Phi$ and the velocities under general conditions involving only $\gamma(\cdot)$ and its derivatives. In particular, in the case $\gamma(t) = \frac{\alpha}{t}$, which is directly connected to the Nesterov accelerated gradient method, our approach allows us to cover all the positive values of $\alpha$, including the subcritical case $\alpha < 3$. Our approach is based on the introduction of a new class of Lyapunov functions.

1. Introduction

Throughout the paper, $\mathcal{H}$ is a real Hilbert space which is endowed with the scalar product $\langle \cdot, \cdot \rangle$, with $\|x\|^2 = \langle x, x \rangle$ for $x \in \mathcal{H}$. As a standing assumption, we suppose that $\Phi : \mathcal{H} \to \mathbb{R}$ is a convex differentiable function whose gradient $\nabla \Phi$ is Lipschitz continuous on bounded sets, and whose solution set $S = \text{argmin} \Phi$ is nonempty. For a fixed $t_0 \in \mathbb{R}$, we consider the second-order differential equation

\[(\text{IGS})_\gamma \quad \dot{x}(t) + \gamma(t)\dot{x}(t) + \nabla \Phi(x(t)) = 0, \quad t \geq t_0,\]

where $\gamma(\cdot)$ is a positive time-dependent damping parameter of class $C^2$. The appellation $(\text{IGS})_\gamma$ is for Inertial Gradient System with parameter $\gamma(\cdot)$. We will give conditions directly expressed on $\gamma(\cdot)$ which allow us to describe the asymptotic behavior of the trajectories when $t \to +\infty$. In particular, we will evaluate the convergence rate of the values $\Phi(x(t)) - \inf_{\mathcal{H}} \Phi$, and of the global energy.

1.1. Historical presentation. This study is part of the active research that has been devoted this last decade to the accelerated gradient methods for convex optimization, and their links with the continuous gradient-like inertial dynamics. For recent papers on inertial gradient systems with time-dependent friction, the reader is referred to [1, 2, 4, 5, 8, 9, 11, 12, 13, 14, 17, 19, 20, 21, 26]. Our paper comes as a natural complement to the recent article by Attouch and Cabot [2]. One of our goals is to broaden the scope of this study to cover the important case $\gamma(t) = \frac{\alpha}{t}$ for all positive values of the parameter $\alpha$. Let us briefly explain the importance of this case, and the recent progress regarding its study.

The tuning of the damping parameter $\gamma(t)$ plays a central role for obtaining fast gradient methods in numerical optimization. Recent developments show the close relationship between fast gradient methods for convex optimization and inertial gradient systems with vanishing damping coefficient, $\gamma(t) \to 0$ as $t \to +\infty$. As pointed out by Su-Boyd-Candes in [26], the $(\text{IGS})_\gamma$ system with $\gamma(t) = \frac{3}{t}$

\[(1) \quad \dot{x}(t) + \frac{3}{t}\dot{x}(t) + \nabla \Phi(x(t)) = 0\]

can be seen as a continuous version of the accelerated gradient method of Nesterov (see [22, 23]). This method has been developed to deal with large scale structured convex minimization problems, such as the FISTA algorithm of Beck-Teboulle [10]. These methods guarantee (in the worst case) the convergence rate $\Phi(x_k) - \inf_{\mathcal{H}} \Phi = \mathcal{O} \left( \frac{1}{k^2} \right)$, where $k$ is the number of iterations. Convergence of the trajectories generated by (1) and of the sequences

\textbf{Date:} March 29, 2018.  
1991 Mathematics Subject Classification. 37N40, 49M05, 65K05, 65K10, 90C25.  
Key words and phrases. Asymptotic stabilization; convex optimization; inertial gradient dynamics; fast gradient method; Lyapunov analysis; time-dependent viscosity.
Moreover, the energy function \( W(t) := (\Phi(x(t)) - \min_{\mathcal{H}} \Phi) + \frac{1}{2} \| \dot{x}(t) \|^2 \).

As a direct consequence of the equation (IGS)\( _\gamma \), differentiating \( W(\cdot) \) we get that
\[
\frac{d}{dt} W(t) = \langle \ddot{x}(t) + \nabla \Phi(x(t)), \dot{x}(t) \rangle = -\gamma(t) \| \dot{x}(t) \|^2.
\]

Hence, \( \frac{d}{dt} W(t) \leq 0 \), and \( W(\cdot) \) is a nonincreasing function.

Assume now that \( \Phi = 0 \). Then equation (IGS)\( _\gamma \) reduces to the linear differential equation
\[
\ddot{x}(t) + \gamma(t) \dot{x}(t) = 0.
\]

Let us multiply this equality by the integrating factor \( p(t) = e^{\int_{t_0}^t \gamma(\tau) \, d\tau} \) and integrate on \([t_0, t]\). We obtain
\[
p(t) \dot{x}(t) = \dot{x}(t_0)
\]
for every \( t \geq t_0 \). By integrating again, we find
\[
x(t) = x(t_0) + \left( \int_{t_0}^t \frac{ds}{p(s)} \right) \dot{x}(t_0).
\]

It ensues immediately that the trajectory \( x(\cdot) \) converges if and only if \( \dot{x}(t_0) = 0 \) or
\[
\int_{t_0}^{t_\infty} \frac{ds}{p(s)} < +\infty.
\]

Let us come back to the case of a general function \( \Phi \) and assume that the above condition is satisfied. We then define the function \( \Gamma : [t_0, +\infty] \rightarrow \mathbb{R}_+ \) by
\[
\Gamma(t) = \left( \int_t^{t_\infty} \frac{ds}{p(s)} \right) p(t).
\]

The theorem below gives the basic assumptions on the function \( \gamma \), that allow to obtain convergence of the trajectories of (IGS)\( _\gamma \).

**Theorem 1.1** (Attouch and Cabot [2]). Let \( \Phi : \mathcal{H} \rightarrow \mathbb{R} \) be a convex function of class \( \mathcal{C}^1 \) such that argmin\( \Phi \neq \emptyset \). Let us assume that \( \gamma : [t_0, +\infty] \rightarrow \mathbb{R}_+ \) is a continuous function satisfying:

(i) \( \int_{t_0}^{t_\infty} \frac{ds}{p(s)} < +\infty \);

(ii) There exist \( t_1 \geq t_0 \) and \( m < \frac{3}{2} \) such that \( \gamma(t) \Gamma(t) \leq m \) for every \( t \geq t_1 \);

(iii) \( \int_{t_0}^{t_\infty} \Gamma(s) \, ds = +\infty \).

Then every solution trajectory \( x : [t_0, +\infty] \rightarrow \mathcal{H} \) of (IGS)\( _\gamma \) converges weakly toward some \( x^* \in \text{argmin} \Phi \). Moreover, the energy function \( W \) satisfies the following rate of convergence:
\[
W(t) = o \left( \frac{1}{\int_{t_0}^t \Gamma(s) \, ds} \right) \quad \text{as} \; t \rightarrow +\infty.
\]
Taking $\gamma(t) = \alpha/t$ with $\alpha > 1$ gives after some elementary computation

$$\Gamma(t) = \left( \frac{t}{t_0} \right)^\alpha \left[ \int_t^{+\infty} \left( \frac{t_0}{\tau} \right)^\alpha d\tau \right] = t^\alpha \left[ \frac{\tau^{-\alpha} + 1}{-\alpha + 1} \right]_t^{+\infty} = \frac{t}{\alpha - 1}.$$ 

Then, the condition $\gamma(t) \Gamma(t) \leq m$ with $m < \frac{3}{2}$ is equivalent to $\alpha > 3$. As a consequence, for $\gamma(t) = \alpha/t$ and $\alpha > 3$, we recover the convergence of the trajectories of (IGS)$_\gamma$, together with the rate of convergence

$$W(t) = o \left( \frac{1}{t^2} \right) \quad \text{as} \ t \to +\infty.$$ 

The abovementioned theorem does not allow to handle the subcritical case, corresponding to $\alpha \leq 3$. This is the reason why we consider in this paper an alternative approach, based on a new Lyapunov function. We assume that $\gamma(\cdot)$ is of class $C^2$ and satisfies the following condition: There exists some positive real number $0 < r \leq \frac{1}{3}$ such that, for $t$ large enough,

$$(H_{\gamma,r}) \quad \dot{\gamma}(t) \geq 2r^2 \gamma(t)^3.$$ 

The value $r = \frac{1}{3}$ turns out to be critical (it corresponds to $\alpha = 3$ in the case $\gamma(t) = \frac{a}{t}$).

Under the assumption $(H_{\gamma,r})$ with $r \in [0, 1/3]$, we show in Section 2 (see Theorem 2.1) that

$$W(t) = O \left( e^{-2r \int_t^{t_0} \gamma(s) ds} \right) \quad \text{as} \ t \to +\infty.$$ 

Equivalently, $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = O \left( e^{-2r \int_t^{t_0} \gamma(s) ds} \right)$ and $\|\dot{x}(t)\| = O \left( e^{-r \int_t^{t_0} \gamma(s) ds} \right)$ as $t \to +\infty$. Section 3 is devoted to the case $r < \frac{1}{3}$. In this case, we obtain additional integral estimates that allow to prove in Theorem 3.1 the stronger result

$$W(t) = o \left( e^{-2r \int_t^{t_0} \gamma(s) ds} \right) \quad \text{as} \ t \to +\infty.$$ 

Equivalently $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o \left( e^{-2r \int_t^{t_0} \gamma(s) ds} \right)$ and $\|\dot{x}(t)\| = o \left( e^{-r \int_t^{t_0} \gamma(s) ds} \right)$ as $t \to +\infty$. In Section 4 we apply the above results to some particular situations. In addition to the important case $\gamma(t) = \frac{1}{(\ln t)^\beta}, \ 0 \leq \beta \leq 1$. Finally, in Section 5, we study the perturbed version of $(IGS)_\gamma$ obtained by introducing a second member $g$ in the dynamics. We show in Theorem 5.1 that the convergence results remain satisfied if the perturbation $g$ is not too large asymptotically. This reflects some structural stability of the dynamics $(IGS)_\gamma$.

2. Rate of convergence of the energy

Let us state our main result concerning the rate of convergence of the energy for $(IGS)_\gamma$.

**Theorem 2.1.** Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex differentiable function whose solution set $\text{argmin} \Phi$ is nonempty. Let us give $\alpha \in \mathbb{R}$ and $\gamma(\cdot) : [t_0, +\infty) \to \mathbb{R}_+$ a nonincreasing function that is twice continuously differentiable. Suppose that there exists some positive real number $0 < r \leq \frac{1}{3}$ such that, for $t$ large enough,

$$(H_{\gamma,r}) \quad \dot{\gamma}(t) \geq 2r^2 \gamma(t)^3.$$ 

Then, for each solution trajectory $x : [t_0, +\infty) \to \mathcal{H}$ of $(IGS)_\gamma$

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = O \left( e^{-2r \int_t^{t_0} \gamma(s) ds} \right) \quad \text{as} \ t \to +\infty.$$ 

Suppose moreover that $x(\cdot)$ is bounded. Then,

$$\|\dot{x}(t)\| = O \left( e^{-r \int_t^{t_0} \gamma(s) ds} \right) \quad \text{as} \ t \to +\infty.$$ 

**Proof.** Let us introduce the function $p : [t_0, +\infty) \to \mathbb{R}_+$ defined by

$$p(t) = e^{\int_t^{t_0} \gamma(s) ds}, \quad t \geq t_0.$$ 

It enters as a key ingredient of the energy-like function $\mathcal{E}_{r,\lambda,\xi} : [t_0, +\infty) \to \mathbb{R}_+$

$$\mathcal{E}_{r,\lambda,\xi}(t) = p(t)^2 \left[ \Phi(x(t)) - \min_{\mathcal{H}} \Phi \right] + \frac{1}{2} \lambda(t) \|x(t) - z\|^2 + p(t)^r \|\dot{x}(t)\|^2 + \xi(t)^2 \|\dot{x}(t)\|^2,$$ 

where $\lambda(t)$ and $\xi(t)$ are appropriate positive functions that will be chosen later. The right-hand side of (5) is of class $C^2$, and its time derivative is given by

$$\mathcal{L}(x(t), t) = \left[ \Phi(x(t)) - \min_{\mathcal{H}} \Phi \right] - \lambda(t) \|x(t) - z\|^2 + \xi(t)^2 \|\dot{x}(t)\|^2,$$ 

where $\mathcal{L}(x(t), t)$ is of class $C^2$.

Therefore, if $\mathcal{E}(x(t), t)$ is the energy-like function defined as in (5), then

$$\mathcal{L}(x(t), t) = \frac{d}{dt} \mathcal{E}(x(t), t) = \mathcal{L}(x(t), t).$$ 

Theorem 2.1 follows from the above functional properties of the energy-like function.

3. Rate of convergence of the perturbed system

Let us state our main result concerning the rate of convergence of the perturbed system for $(IGS)_\gamma$.

**Theorem 3.1.** Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex differentiable function whose solution set $\text{argmin} \Phi$ is nonempty. Let us give $\alpha \in \mathbb{R}$ and $\gamma(\cdot) : [t_0, +\infty) \to \mathbb{R}_+$ a nonincreasing function that is twice continuously differentiable. Suppose that there exists some positive real number $0 < r \leq \frac{1}{3}$ such that, for $t$ large enough,

$$(H_{\gamma,r}) \quad \dot{\gamma}(t) \geq 2r^2 \gamma(t)^3.$$ 

Then, for each solution trajectory $x : [t_0, +\infty) \to \mathcal{H}$ of $(IGS)_\gamma$

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = O \left( e^{-2r \int_t^{t_0} \gamma(s) ds} \right) \quad \text{as} \ t \to +\infty.$$ 

Equivalently $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o \left( e^{-2r \int_t^{t_0} \gamma(s) ds} \right)$ and $\|\dot{x}(t)\| = o \left( e^{-r \int_t^{t_0} \gamma(s) ds} \right)$ as $t \to +\infty$.
that will serve as a Lyapunov function. At this point $z \in \text{argmin} \Phi$ is fixed, while $r$ is a positive real number, $\lambda(\cdot)$, $\xi(\cdot)$ are positive functions we will use as parameters. The key of the proof is to wisely choose these parameters in order to make $E_{\lambda, \xi}(\cdot)$ a nonincreasing function. In short, without ambiguity, we will write $E$ instead of $E_{\lambda, \xi}$.

Thus, the core of the proof consists in proving that $\frac{d}{dt} E(t) \leq 0$ for $t$ large enough. In differentiating $E$, we will use systematically the following elementary relation

$$ \dot{p}(t) = \gamma(t)p(t), \quad t \geq t_0. $$

Let us first derivate each of the constitutive elements of $E$. Set $A(t) = p(t)^{2r} [\Phi(x(t)) - \text{min}_H \Phi]$, $B(t) = \frac{\xi(t)}{2} \|x(t) - z\|^2$, and $C(t) = \frac{1}{2} \|\lambda(t)(x(t) - z) + p(t)^r \dot{x}(t)\|^2$. So, $E(t) = A(t) + B(t) + C(t)$. By means of the classical derivation chain rule we obtain

$$ \frac{d}{dt} A(t) = 2r\gamma(t)p(t)^{2r}(\Phi(x(t)) - \text{min}_H \Phi) + p(t)^{2r} \langle \nabla \Phi(x(t)), \dot{x}(t) \rangle $$

$$ \frac{d}{dt} B(t) = \frac{\xi(t)}{2} \|x(t) - z\|^2 + \xi(t) \langle x(t) - z, \dot{x}(t) \rangle $$

$$ \frac{d}{dt} C(t) = \langle \lambda(t)(x(t) - z) + p(t)^r \dot{x}(t), \lambda(t) \dot{x}(t) + \dot{\lambda}(t)(x(t) - z) + r\gamma(t)p(t)^r \dot{x}(t) + p(t)^r \dot{x}(t) \rangle $$

$$ = \langle \lambda(t)(x(t) - z) + p(t)^r \dot{x}(t), (\lambda(t) - (1 - r)\gamma(t)p(t)^r) \dot{x}(t) + \dot{\lambda}(t)(x(t) - z) - p(t)^r \nabla \Phi(x(t)) \rangle $$

$$ \leq \langle \lambda(t)(x(t) - z) + p(t)^r \dot{x}(t), (\lambda(t) - (1 - r)\gamma(t)p(t)^r) \dot{x}(t) + \dot{\lambda}(t)(x(t) - z) \rangle $$

$$ - p(t)^{2r} \langle \nabla \Phi(x(t)), \dot{x}(t) \rangle - \lambda(t)p(t)^r \langle \Phi(x(t)) - \text{min}_H \Phi \rangle. $$

To obtain the two last equations we have successively used $(IGS)_\gamma$ and the convexity of $\Phi$. Adding the above results, we obtain, after simplification,

$$ \frac{d}{dt} E(t) \leq p(t)^r \left[ 2r\gamma(t)p(t)^r - \lambda(t) \right] (\Phi(x(t)) - \text{min}_H \Phi) $$

$$ + \left[ \xi(t) + p(t)^r \dot{\lambda}(t) + \lambda(t) \{ \lambda(t) - (1 - r)\gamma(t)p(t)^r \} \right] \langle x(t) - z, \dot{x}(t) \rangle $$

$$ - p(t)^r \left[ (1 - r)\gamma(t)p(t)^r - \lambda(t) \right] \| \dot{x}(t) \|^2 + \left[ \lambda(t) \dot{\lambda}(t) + \frac{\xi(t)}{2} \right] \|x(t) - z\|^2. $$

Let us successively examine the different terms entering the second member of (7). Let us make the first two terms equal to zero by taking respectively

$$ \lambda(t) = 2r\gamma(t)p(t)^r $$

$$ \xi(t) = -p(t)^r \dot{\lambda}(t) - \lambda(t) \{ \lambda(t) - (1 - r)\gamma(t)p(t)^r \}. $$

Let us combine (8) with (9), and use that $\dot{\lambda}(t) = 2rp(t)^r (\gamma(t) + r\gamma(t)^2)$. We obtain

$$ \xi(t) = -\lambda(t)^2 + (1 - r)\lambda(t)\gamma(t)p(t)^r - \dot{\lambda}(t)p(t)^r $$

$$ = p(t)^{2r} \left[ -4r^2\gamma(t)^2 + 2r(1 - r)\gamma(t)^2 - 2r^2\gamma(t)^2 - 2r\gamma(t) \right] $$

$$ = p(t)^{2r} \left[ (-8r^2 + 2r)\gamma(t)^2 - 2r\gamma(t) \right] $$

$$ = -2rp(t)^{2r} \left[ (4r - 1)\gamma(t)^2 + \gamma(t) \right]. $$

Recall that we want $E(\cdot)$ to be a nonincreasing function. Therefore, we require the third term of the second member of (7) to satisfy

$$ (1 - r)\gamma(t)p(t)^r - \lambda(t) \geq 0. $$

Taking into account (8), that is $\lambda(t) = 2r\gamma(t)p(t)^r$, this is equivalent to assuming that

$$ (H_1): \quad r \leq \frac{1}{4}. $$

Let’s finally consider the last term of (7), and study under which conditions on the parameters, the inequality

$$ \lambda(t) \dot{\lambda}(t) + \frac{\xi(t)}{2} \leq 0, $$

is verified. It is equivalent to prove that $t \mapsto \beta(t)$ is nonincreasing, where $\beta(t) := \xi(t) + \lambda(t)^2$. We have

$$ \beta(t) = -2rp(t)^{2r} \left[ (4r - 1)\gamma(t)^2 + \gamma(t) \right] + 4r^2\gamma(t)^2p(t)^{2r} $$

$$ = -2rp(t)^{2r} \left[ (2r - 1)\gamma(t)^2 + \gamma(t) \right]. $$
By differentiating, we obtain,

\[ \dot{\beta}(t) = -4\gamma^2 \rho(t)^2 \gamma(t) \left[ (2\rho - 1)\gamma(t)^3 + \dot{\gamma}(t) \right] - 2\gamma \rho(t)^2 [2(2\rho - 1)\gamma(t)\dot{\gamma}(t) + \ddot{\gamma}(t)] \]

\[ = -2\gamma \rho(t)^2 \left[ 2(2\rho - 1)\gamma(t)^3 + 2\gamma(t)\dot{\gamma}(t) + 2(2\rho - 1)\gamma(t)\dot{\gamma}(t) + \ddot{\gamma}(t) \right] \]

\[ = -2\gamma \rho(t)^2 \left[ 2(2\rho - 1)\gamma(t)^3 + (6\rho - 2)\gamma(t)\dot{\gamma}(t) + \ddot{\gamma}(t) \right]. \]

As a consequence, inequality (11) is satisfied under the following additional assumption: for \( t \) large enough

\[ (H_2): \quad \dot{\gamma}(t) + (6\rho - 2)\gamma(t)\dot{\gamma}(t) + 2\gamma(2\rho - 1)\gamma(t)^3 \geq 0. \]

In summary, we have proved that under the assumptions \((H_1)\) and \((H_2)\), the energy function \( E(\cdot) \) is nonincreasing. Let us simplify this set of conditions. First observe that

\[ \dot{\gamma}(t) + (6\rho - 2)\gamma(t)\dot{\gamma}(t) + 2\gamma(2\rho - 1)\gamma(t)^3 = (\dot{\gamma}(t) - 2\gamma \rho(t)^2 \gamma(t)^3) + 2(3\rho - 1)\gamma(t)\dot{\gamma}(t) + (6\rho - 2)\gamma(t)^3 \]

\[ = (\dot{\gamma}(t) - 2\gamma \rho(t)^2 \gamma(t)^3) + 2(3\rho - 1)\gamma(t)^3 + \gamma(t)\dot{\gamma}(t) + r\gamma(t)^3 \]

Under condition \((H_1)\), we have \( 3\rho - 1 \leq 0 \). Therefore \((H_2)\) is satisfied under the two conditions

\[ (H_3): \quad \dot{\gamma}(t) \geq 2\gamma \rho(t)^2 \gamma(t)^3, \]

\[ (H_4): \quad \dot{\gamma}(t) \leq -r\gamma(t)^2. \]

So our set of assumptions amounts to \((H_1)-(H_3)-(H_4)\). Let’s reduce it by observing that \((H_4)\) implies \((H_4)\). This follows from an integration procedure. Recall that \( \gamma(\cdot) \) is assumed to be nonincreasing (indeed it is a necessary condition to get \((H_4)\)). Therefore \( \lim_{t \to +\infty} \gamma(t) = l \) exists with \( l \geq 0 \). Let’s show that necessarily \( l = 0 \). Otherwise, by \((H_4)\), we would have \( \dot{\gamma}(t) \geq r\gamma(t)^2 \) for \( t \) large enough, say \( t \geq T \). By integrating this inequality from \( T \) to \( t \), we would have \( \dot{\gamma}(t) \geq \dot{\gamma}(T) + r\gamma(t)^2 (t - T) \) for \( t \geq T \). This would imply \( \lim_{t \to +\infty} \gamma(t) = +\infty \), which in turn gives \( \lim_{t \to +\infty} \gamma(t) = +\infty \), a clear contradiction with \( \gamma(\cdot) \) nonincreasing. So \( \lim_{t \to +\infty} \gamma(t) = 0 \).

Let us multiply \((H_3)\) by \( \dot{\gamma}(t) \). Since this quantity is less than or equal to zero, we obtain

\[ \dot{\gamma}(t)\dot{\gamma}(t) \leq 2\gamma \rho(t)^2 \gamma(t)^3 \dot{\gamma}(t). \]

Equivalently

\[ \frac{d}{dt} \dot{\gamma}(t)^2 \leq 2r^2 \frac{d}{dt} \gamma(t)^4. \]

Let us integrate this differential inequality from \( t \) to \( T \geq t \). We obtain

\[ \dot{\gamma}(T)^2 - \dot{\gamma}(t)^2 \leq 2r^2 (\gamma(T)^4 - \gamma(t)^4), \]

which gives

\[ -\dot{\gamma}(t)^2 \leq 2r^2 (\gamma(T)^4 - \gamma(t)^4). \]

Letting \( T \to +\infty \), and using \( \lim_{T \to +\infty} \gamma(T) = 0 \), we obtain

\[ -\dot{\gamma}(t)^2 \leq -r^2 \gamma(t)^4. \]

Equivalently

\[ |\dot{\gamma}(t)| \geq r\gamma(t)^2. \]

Since \( \dot{\gamma}(t) \leq 0 \), this gives

\[ -\dot{\gamma}(t) \geq r\gamma(t)^2, \]

that’s \((H_4)\). To summarize, we proved that under the assumptions \((H_1)\) and \((H_3)\), \( \frac{d}{dt} E(t) \leq 0 \) for \( t \) large enough, and hence the function \( E(\cdot) \) is nonincreasing. Let us complete this result by showing that, under these assumptions, \( E(\cdot) \) is nonnegative. Returning to the definition (5) of \( E(\cdot) \), this will result from \( \xi(\cdot) \geq 0 \). By (10), and since \( -2\gamma \rho(t)^2 \leq 0 \), this property is equivalent to

\[ \dot{\gamma}(t) \leq (1 - 4\rho)\gamma(t)^2. \]

Let us now observe that under condition \((H_1)\), that is \( r \leq \frac{1}{3} \), we have \( -r \leq 1 - 4\rho \). Therefore, (13) is implied by assumption \((H_4)\), which itself is a consequence of \((H_3)\).

Going back to the definition of \( E(\cdot) \), and because \( E(\cdot) \geq 0 \), we obtain that for \( t \) large enough, say \( t \geq t_1 \)

\[ p(t)^2 (\Phi(x(t) - \min_{\nu} \Phi) \leq E(t) \leq E(t_1), \]
and the result follows.

Let us now suppose that the trajectory \( x(t) \) is bounded. From \( \mathcal{E}(t) \leq \mathcal{E}(t_1) \) for \( t \geq t_1 \), the definition of \( \mathcal{E}(\cdot) \) and \( \xi(t) \geq 0 \), we deduce that

\[
\frac{1}{2} \| \lambda(t)(x(t) - z) + p(t)\dot{x}(t) \|^2 \leq \mathcal{E}(t_1).
\]

According to the triangle inequality, it follows that

\[
p(t)^\gamma \| \dot{x}(t) \| \leq \sqrt{2\mathcal{E}(t_1)} + \lambda(t)\| x(t) - z \|.
\]

Recall that \( \lambda(t) = 2r\gamma(t)p(t)^r \), as defined in (8). Therefore, the above inequality gives

\[ (14) \quad \| \dot{x}(t) \| \leq \frac{\sqrt{2\mathcal{E}(t_1)}}{p(t)^r} + 2r\gamma(t)\| x(t) - z \|.
\]

Let us show that, under condition \((H_{\gamma,r})\), \( \gamma(t) \leq \frac{C}{p(t)^r} \). To that end, let us differentiate the function \( t \mapsto \gamma(t)p(t)^r \). Using the relation \( \dot{p}(t) = \gamma(t)p(t) \) (see (6)) we get

\[
\frac{d}{dt} \left( \gamma(t)p(t)^r \right) = p(t)^r \left( \dot{\gamma}(t) + r\gamma(t)^2 \right).
\]

As seen above, the condition \((H_4)\), that’s \( \dot{\gamma}(t) + r\gamma(t)^2 \leq 0 \) is implied by \((H_{\gamma,r})\). Hence the function \( t \mapsto \gamma(t)p(t)^r \) is nonincreasing on \([t_1, +\infty[\). It ensues that \( \gamma(t)p(t)^r \leq \gamma(t_1)p(t_1)^r \) for all \( t \geq t_1 \). In view of (14), we conclude that

\[
\| \dot{x}(t) \| \leq \frac{1}{p(t)^r} \left( \sqrt{2\mathcal{E}(t_1)} + 2r\gamma(t_1)p(t_1)^r \sup_{t \geq t_0} \| x(t) - z \| \right),
\]

which gives the claim. \( \square \)

**Remark 2.1.** It follows from \((H_{\gamma,r})\) and the proof of Theorem 2.1 that, under the assumptions of Theorem 2.1, \( \gamma(\cdot) \) is a nonnegative, nonincreasing, convex function that tends to zero as \( t \to +\infty \).

**Remark 2.2.** For any \( r, r' > 0 \) such that \( r \geq r' \), condition \((H_{\gamma,r})\) clearly implies \((H_{\gamma,r'})\). Therefore, if \((H_{\gamma,r})\) holds true for some \( r \geq 1/3 \), then the conclusions of Theorem 2.1 are satisfied with \( r' = 1/3 \) in place of \( r \).

**Remark 2.3.** In order to verify condition \((H_{\gamma,r})\) in Theorem 2.1, one just needs to consider the asymptotic equivalents of the functions \( \gamma(\cdot) \) and \( \dot{\gamma}(\cdot) \). This gives flexibility to this approach. Observe however that the coefficient \( r \) may change when considering an equivalent expression. If \( \gamma \) satisfies condition \((H_{\gamma,r})\) for some \( r > 0 \), an equivalent of \( \gamma \) may fulfill the condition only for \( r' < r \), and vice versa.

**Remark 2.4 (Case of equality in condition \((H_{\gamma,r})\)).** Assume that \( \gamma : [t_0, +\infty[ \to \mathbb{R}_+ \) is nonincreasing and satisfies \( \dot{\gamma}(t) = 2r^2\gamma(t)^3 \) for \( t \) large enough. We easily show that \( \lim_{t \to +\infty} \gamma(t) = \lim_{t \to +\infty} \dot{\gamma}(t) = 0 \), see the proof of Theorem 2.1. By using the integration procedure performed in the proof of Theorem 2.1, we obtain \( \gamma(t) = -r\gamma(t)^2 \) for \( t \) large enough. A second integration then yields \( \gamma(t) = 1/(rt + c) \), for some \( c \in \mathbb{R} \) and every \( t \) large enough.

**Remark 2.5.** The important role of the quantity \( \frac{\dot{\gamma}(t)}{\gamma(t)^2} \) in the asymptotic analysis of \((\text{IGS})\gamma \) was already underlined in [2]. In that article, it came in the form \( \frac{\dot{\gamma}(t)}{\gamma(t)^2} \geq -c \) with \( c \in ]0,1[ \). This shows the complementary aspect of our analysis, which is based on the opposite inequality.

**Remark 2.6.** Under the assumption

\[
\dot{\gamma}(t) + r\gamma(t)^2 \leq 0 \quad \text{for some } r \in ]0, 2/3[,\]

Cabot-Engler-Gadat [12] obtained the following estimate for the energy decay

\[
W(t) = \mathcal{O} \left( e^{-t \int_0^t \gamma(s)ds} \right) \quad \text{as } t \to +\infty,
\]

see [12, Proposition 3.5]. Theorem 2.1 improves the latter result by a factor 2 in the exponential.
Optimality of the rate of convergence results. Let us discuss the optimality of the convergence rates (3) and (4) obtained in Theorem 2.1. Optimality is understood in the worst case, which means that, for a given function $\gamma(\cdot)$, we are able to find potential functions $\Phi$ for which equality is reached (or approximated with arbitrary precision) in formulas (3) and (4). Such a discussion for an arbitrary damping function $\gamma(\cdot)$ is difficult because, in general, it is impossible to compute the general solution. In some cases, we can provide special solutions with a computable decay rate, allowing us to conclude to the optimality of these formulas. This is the case when $\gamma(t) = \frac{1}{t^2}$. The corresponding differential equation

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla \Phi(x(t)) = 0$$

plays an important role in optimization, because of its relation with the accelerated method of Nesterov, see [4, 26]. We will distinguish the two cases $\alpha \geq 3$ and $\alpha \leq 3$, which turn out to be two distinct regimes.

a) The case $\alpha \geq 3$. In this case we have $W(t) = \mathcal{O}\left(\frac{1}{t^2}\right)$ as a consequence of Theorem 2.1. See Section 4 for a detailed analysis. Let us show the optimality of this estimate. Following [4, Example 2.13], take $\mathcal{H} = \mathbb{R}$ and $\Phi(x) = c|x|^\delta$, where $c$ and $\delta$ are positive parameters. Let us look for positive solutions of

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + c\delta|x(t)|^{\delta-2}x(t) = 0$$

of the form $x(t) = \frac{1}{t^\theta}$, with $\theta > 0$. This means that the trajectory is not oscillating, which corresponds to over damping (also called heavy damping). We begin by determining the values of $c$, $\delta$ and $\theta$ that provide such solutions. On the one hand,

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) = \theta(\theta + 1 - \alpha) \frac{1}{t^{\theta+2}}.$$

On the other hand,

$$c\delta|x(t)|^{\delta-2}x(t) = c\delta \frac{1}{t^{\theta(\delta-1)}}.$$

Thus, $x(t) = \frac{1}{t^\theta}$ is a solution of (16) if, and only if,

i) $\theta + 2 = \theta(\delta - 1)$, which is equivalent to $\delta > 2$ and $\theta = \frac{2}{\delta - 2}$; and

ii) $c\delta = c(\alpha - \theta - 1)$, which is equivalent to $\alpha > \frac{\delta}{\delta - 2}$ and $c = \frac{2}{3(\delta - 2)}(\alpha - \frac{\delta}{\delta - 2})$.

We have $\min \Phi = 0$ and

$$\Phi(x(t)) = \frac{2}{\delta(\delta - 2)}(\alpha - \frac{\delta}{\delta - 2}) \frac{1}{t^{\frac{\delta}{\delta - 2}}}, \quad |\dot{x}(t)|^2 = \frac{4}{(\delta - 2)^2} \frac{1}{t^{\frac{2\delta}{\delta - 2}}}.$$

Hence the decay rate of the energy $W(t)$ is of order $\frac{1}{t^{\frac{\delta}{\delta - 2}}}$. As $\delta$ tends to infinity, the exponent $\frac{2\delta}{\delta - 2}$ tends to 2 from above, which shows in this case the optimality of the decay rate $W(t) = \mathcal{O}\left(\frac{1}{t^2}\right)$ given by Theorem 2.1. This limiting situation is obtained by taking a function $\Phi$ that becomes very flat around the set of its minimizers.

To achieve exactly the decay rate $W(t) \sim \frac{1}{t^2}$, we must take a function $\Phi$ extremely flat around its infimum, as given by the inverse of an exponential growth. Following [4, Example 2.11], take $\mathcal{H} = \mathbb{R}$ and $\Phi(x) = \frac{\alpha - 1}{2} e^{-2x}$ with $\alpha \geq 1$. Let us verify that $x(t) = \ln(t)$ is a solution of (15). On the one hand,

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) = \frac{\alpha - 1}{t^2}.$$

On the other hand, $\nabla \Phi(x) = -(\alpha - 1)e^{-2x}$ which gives

$$\nabla \Phi(x(t)) = -(\alpha - 1)e^{-2 \ln t} = -\frac{\alpha - 1}{t^2}.$$

Thus, $x(t) = \ln t$ is a solution of (15). Let us examine the decay rate. We have $\inf \Phi = 0$, and

$$\Phi(x(t)) = \frac{\alpha - 1}{2} e^{-2 \ln t} = \frac{\alpha - 1}{2t^2}, \quad |\dot{x}(t)|^2 = \frac{1}{t^2},$$

which gives\footnote{It would be interesting to develop a similar example in the case where $\arg \min \Phi \neq \emptyset$.} a decay rate of $W(t)$ of order $\frac{1}{t^2}$.

b) The case $\alpha \leq 3$. In this case, we have $W(t) = \mathcal{O}\left(\frac{1}{t^{\frac{3}{2}}}\right)$ as a consequence of Theorem 2.1. See Section 4 for a detailed analysis. Let us show the optimality of this estimate. The over-damped solution used in the...
case $\alpha \geq 3$ does no longer work. According to Aujol-Dossal [8], the idea is to take $\Phi(x) = |x|^\delta$, and let the exponent $\delta$ tend to 1. For $\delta > 1$, it is shown in [8, Proposition 6] that

$$W(t) \geq Kt^{-\frac{2\theta}{1-\theta}}.$$ 

The optimality of the decay rate is obtained by letting $\delta$ tend to one in this formula. To find a potential function $\Phi$ for which this rate is exactly obtained, we could take directly $\delta = 1$ as in Apidopoulos-Aujol-Dossal [1, Theorem 5.1], but in this case we have to consider a differential inclusion instead of a differential equation, see [3].

**Remark 2.7.** For clear numerical reasons, the multiplicative coefficient involved in the convergence rate also plays an important role. Its optimization is the subject of an active research trend, see Drori-Teboulle [16] and references therein.

**Some other examples.** Let us now give some other examples for which we are able to determine explicitly the rate of convergence of the energy. Taking $\Phi \equiv 0$, we obtain $\dot{x}(t) = \dot{x}(t_0)e^{-\int_0^t \gamma(s)\,ds}$, see the introduction. We then have for every $t \geq t_0$,

$$W(t) = \frac{1}{2} \|\dot{x}(t)\|^2 = \frac{1}{2} \|\dot{x}(t_0)\|^2 e^{-2\int_0^t \gamma(s)\,ds}.$$ 

The above rate of decay is the best that can be expected for the energy function $W(t)$, as shown by the next result.

**Proposition 2.1.** Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a function of class $C^1$ such that $\text{argmin}\Phi \neq \emptyset$. Let $\gamma : [t_0, +\infty) \to \mathbb{R}_+$ be a continuous function. Then any solution $x$ of $(\text{IGS})_\gamma$ satisfies for every $t \geq t_0$,

$$W(t) \geq W(t_0) e^{-2\int_0^t \gamma(s)\,ds}.$$ 

**Proof.** Taking into account the expression of $W$ and $\dot{W}$, we have

$$\dot{W}(t) + 2\gamma(t) W(t) = 2\gamma(t) (\Phi(x(t)) - \min_{\mathcal{H}} \Phi) \geq 0.$$ 

Let us multiply the above inequality by $e^{2\int_0^t \gamma(s)\,ds}$, we deduce that:

$$\frac{d}{dt} \left[ e^{2\int_0^t \gamma(s)\,ds} W(t) \right] \geq 0.$$ 

Formula (17) immediately follows. \qed

Assuming now that $\Phi = \frac{1}{2} \|\cdot\|^2$, the next proposition gives an equivalent of the energy $W(t)$ as $t \to +\infty$.

**Proposition 2.2.** Assume that $\Phi = \frac{1}{2} \|\cdot\|^2$ and that $\gamma : [t_0, +\infty) \to \mathbb{R}_+$ is a function of class $C^1$ such that $\lim_{t \to +\infty} \gamma(t) = 0$ and $\dot{\gamma} \in L^1(t_0, +\infty)$. Let $x$ be a solution of equation $(\text{IGS})_\gamma$. Then, either the solution $x$ is stationary and equal to zero, or there exists $C > 0$ such that

$$W(t) \sim C e^{-\int_0^t \gamma(s)\,ds} \quad \text{as} \quad t \to +\infty.$$ 

This result was obtained by Cabot-Frankel [14, Lemma 2.2] in the framework of linear hyperbolic evolution equations. For the sake of completeness, a self-contained proof is provided in the Appendix.

The above examples suggest that the speed of convergence of $W(t)$ as $t \to +\infty$ depends on the "index" of convexity of the function $\Phi$. This can be quantified as follows. Let $\theta > 0$ and $\Phi : \mathcal{H} \to \mathbb{R}$ be a function such that $\text{argmin}\Phi \neq \emptyset$. The function $\Phi$ is said to be $\theta$-power convex if $(\Phi - \min_{\mathcal{H}} \Phi)^\theta$ is convex. For $\theta = 1$, we recover the classical notion of convexity. Observe that $\Phi = \frac{1}{2} \|\cdot\|^2$ is $\frac{1}{2}$-power convex, while $\Phi = 0$ corresponds to $\theta \to 0$. For $\theta > 1$, the class of $\theta$-power convex functions goes beyond convexity. The class of $\theta$-power convex functions was handled by Cabot-Engler-Gadat [12, Section 3] in connection with the dynamical system $(\text{IGS})_\gamma$. More recently, the notion of $\theta$-power convexity was used by Aujol-Dossal in the case $\gamma(t) = \frac{t}{2}$. These authors quantified the rate of convergence of the energy $W(t)$ as a function of $\alpha$ and $\theta$, see [8, Theorem 1]. The latter suggests to extend Theorem 2.1 to the framework of $\theta$-power convex functions. This is out of the scope of the present paper, but certainly indicates matter for future investigation.
3. Rate of decay of the energy: case \(0 < r < \frac{1}{3}\)

Let us complete the preceding results by analyzing the decay rate of the global energy in the case \(0 < r < \frac{1}{3}\). We will get a better convergence rate, by passing from \(O\) estimates to \(o\) estimates. Recall that the global energy function \(t \mapsto W(t) \in \mathbb{R}_+\), which is the sum of the potential energy and the kinetic energy, is written

\[
W(t) = \left( \Phi(x(t)) - \min_{\mathcal{H}} \Phi \right) + \frac{1}{2} \| \dot{x}(t) \|^2.
\]

Our proof lies in establishing integral estimates for velocities and values. For this, we return to the basic estimate (7), which we recall below.

\[
\frac{d}{dt} \mathcal{E}(t) \leq p(t)^r \left[ 2r \gamma(t) p(t)^r - \lambda(t) \right] \left( \Phi(x(t)) - \min_{\mathcal{H}} \Phi \right) + \left[ \xi(t) + p(t)^r \lambda(t) + \lambda(t) \left( \lambda(t) - (1 - r) \gamma(t) p(t)^r \right) \right] \langle x(t) - z, \dot{x}(t) \rangle
+ p(t)^r \left[ (1 - r) \gamma(t) p(t)^r - \lambda(t) \right] \| \dot{x}(t) \|^2 + \left[ \lambda(t) \dot{\lambda}(t) + \frac{\dot{\xi}(t)}{2} \right] \| x(t) - z \|^2.
\]

Recall that this inequality is valid regardless of the choice of parameters. It simply uses the convex subdifferential inequality. Precisely, we will obtain these estimates by making different choices for the parameters \(\lambda(\cdot)\) and \(\xi(\cdot)\).

3.1. Integral estimate of the velocities. Following the proof of Theorem 2.1, let’s take in (18) the parameters \(\lambda(t)\) and \(\xi(t)\) respectively given by (8) and (9). So we have

\[
\lambda(t) = 2r \gamma(t) p(t)^r, \quad \xi(t) = -2r p(t)^{2r} \left[ (4r - 1) \gamma(t)^2 + \dot{\gamma}(t) \right].
\]

With this precise choice of the parameters, the two first terms of the right member of (18) are equal to zero. We showed that, under condition \((H_{\gamma,r})\) with \(r \in [0, 1/3]\), the last term of (18) is less than or equal to zero. So, (18) reduces to

\[
\frac{d}{dt} \mathcal{E}(t) + p(t)^r \left[ (1 - r) \gamma(t) p(t)^r - \lambda(t) \right] \| \dot{x}(t) \|^2 \leq 0.
\]

Let us observe that \(p(t)^r \left[ (1 - r) \gamma(t) p(t)^r - \lambda(t) \right] = (1 - 3r) \gamma(t) p(t)^{2r}\). Therefore the above inequality gives

\[
\frac{d}{dt} \mathcal{E}(t) + (1 - 3r) \gamma(t) p(t)^{2r} \| \dot{x}(t) \|^2 \leq 0.
\]

By integrating (19), and using that \(\mathcal{E}(\cdot)\) is nonnegative under the condition \((H_{\gamma,r})\) with \(r \in [0, 1/3]\), we obtain

\[
(1 - 3r) \int_{t_0}^{+\infty} \gamma(t) p(t)^{2r} \| \dot{x}(t) \|^2 dt < +\infty.
\]

This inequality brings information when \(0 < r < \frac{1}{3}\). As a direct consequence of (20) we obtain the following result.

**Proposition 3.1.** Suppose that \(\gamma(\cdot) : [t_0, +\infty] \to \mathbb{R}_+\) is a nonincreasing function that is twice continuously differentiable. Assume that \(\gamma(\cdot)\) satisfies \((H_{\gamma,r})\) for some \(r \in [0, 1/3]\). Then, for each solution trajectory \(x : [t_0, +\infty] \to \mathcal{H}\) of \((\text{IGS})_{\gamma}\)

\[
\int_{t_0}^{+\infty} \gamma(t) e^{2r \int_{t_0}^{t} \gamma(s) ds} \| \dot{x}(t) \|^2 dt < +\infty.
\]

3.2. Integral estimate of the values. Take now in (18) the parameters \(\lambda(t)\) and \(\xi(t)\) so as to make equal to zero the second and third terms of the right member. We immediately obtain

\[
\lambda(t) = (1 - r) \gamma(t) p(t)^r
\]

and

\[
\xi(t) = -p(t)^r \dot{\lambda}(t) = - (1 - r) p(t)^{2r} \left( \dot{\gamma}(t) + r \gamma(t)^2 \right).
\]

Let us verify that for these values of the parameters, and under condition \((H_{\gamma,r})\) with \(r \in [0, 1/3]\), the last term of (18), that is \(2 \lambda(t) \dot{\lambda}(t) + \xi(t)\) is less or equal than zero. We have \(2 \lambda(t) \dot{\lambda}(t) + \xi(t) = \frac{d}{dt} \dot{\beta}(t)\) where

\[
\dot{\beta}(t) = - (1 - r) p(t)^{2r} \left[ (2r - 1) \gamma(t)^2 + \dot{\gamma}(t) \right].
\]
Now $\tilde{\beta} = \frac{1}{2r^2} \beta(t)$, where $\beta(t)$ has been defined in (12). Under the condition $(H_{r,r})$ with $r \in [0, 1/3]$, it has been proved that $\beta(t)$ is nonincreasing. Therefore, this is also true for $\tilde{\beta}$, which gives $2\lambda(t)\lambda(t) + \xi(t) \leq 0$. Thus, under the condition $(H_{r,r})$ with $r \in [0, 1/3]$, and with the above choice of the parameters we have obtained
\[
\frac{d}{dt}\mathcal{E}(t) \leq p(t)^r [2r\gamma(t)p(t)^r - \lambda(t)] (\Phi(x(t)) - \min_{\mathcal{H}} \Phi).
\]
With the above choice of $\lambda(t)$ (see (21)), we have $p(t)^r [2r\gamma(t)p(t)^r - \lambda(t)] = -(1-3r)\gamma(t)p(t)^{2r}$, which gives
\[
\frac{d}{dt}\mathcal{E}(t) + (1-3r)\gamma(t)p(t)^{2r} \left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right) \leq 0.
\]
By integrating (23), and using that $\mathcal{E}(\cdot)$ is nonnegative under the condition $(H_{r,r})$ with $r \in [0, 1/3]$, we obtain
\[
(1-3r)\int_{t_0}^{+\infty} \gamma(t)p(t)^{2r} \left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right) dt < +\infty.
\]
This inequality brings information when $0 < r < \frac{1}{3}$. As a direct consequence of (24) we obtain the following result.

**Proposition 3.2.** Suppose that $\gamma(\cdot) : [t_0, +\infty] \to \mathbb{R}_+$ is a nonincreasing function that is twice continuously differentiable. Assume that $\gamma(\cdot)$ satisfies $(H_{r,r})$ for some $r \in [0, 1/3]$. Then, for each solution trajectory $x : [t_0, +\infty] \to \mathcal{H}$ of (IGS)\gamma
\[
\int_{t_0}^{+\infty} \gamma(t)e^{2r\int_{t_0}^{t} \gamma(s)ds} \left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right) dt < +\infty.
\]

### 3.3. Rate of decay of the global energy.

Putting together the integral estimates obtained in Propositions 3.1 and 3.2, we obtain
\[
\int_{t_0}^{+\infty} \gamma(t)p(t)^{2r}W(t) dt < +\infty.
\]
Assuming that $\int_{t_0}^{+\infty} \gamma(t)p(t)^{2r} dt = +\infty$, and since $W(\cdot)$ is a nonincreasing function, we can apply an argument developed by Attouch-Cabot in [2, Theorem 3.6], which gives
\[
W(t) = o\left(\frac{1}{\int_{t_0}^{+\infty} \gamma(s)p(s)^{2r} ds}\right) \text{ as } t \to +\infty.
\]
From $\int_{t_0}^{t} \gamma(s)p(s)^{2r} ds = \frac{1}{2r} (p(t)^{2r} - p(t_0)^{2r})$, we deduce that
\[
\int_{t_0}^{+\infty} \gamma(s)p(s)^{2r} ds = +\infty \iff \lim_{t \to +\infty} p(t) = +\infty \iff \int_{t_0}^{+\infty} \gamma(s) ds = +\infty.
\]
Collecting the above results we obtain the following statement.

**Theorem 3.1.** Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex differentiable function whose solution set argmin $\Phi$ is nonempty. Suppose that $\gamma(\cdot) : [t_0, +\infty] \to \mathbb{R}_+$ is a nonincreasing function that is twice continuously differentiable. Assume that $\gamma(\cdot)$ satisfies $(H_{r,r})$ for some $r \in [0, 1/3]$, together with $\int_{t_0}^{+\infty} \gamma(s) ds = +\infty$.

Then,
\[
W(t) = o\left( e^{-2r\int_{t_0}^{t} \gamma(s) ds} \right) \text{ as } t \to +\infty.
\]
Equivalently, $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left( e^{-2r\int_{t_0}^{t} \gamma(s) ds} \right)$ and $\|\dot{x}(t)\| = o\left( e^{-r\int_{t_0}^{t} \gamma(s) ds} \right)$ as $t \to +\infty$.

The following table gives a synthetic summary about the speed of convergence of the energy $W(t)$ as $t \to +\infty$.
Recall that $W(t) \to 0$ reflects both the rate of convergence to zero of the value $\Phi(x(t)) - \min_{\mathcal{H}} \Phi$, and of the velocity $\|\dot{x}(t)\|$. 
4. SOME PARTICULAR CASES.

4.1. Case $\gamma(t) = \frac{q}{t}$, $\alpha > 0$. In that case $p(t) = \left(\frac{1}{t^q}\right)^\alpha$. Let us examine the condition $(H_{\gamma,r})$.
We have $\dot{\gamma}(t) = -\frac{\alpha}{t^2}$, $\ddot{\gamma}(t) = \frac{2\alpha}{t^3}$. Hence

$$\ddot{\gamma}(t) \geq 2r^2\gamma(t)^3 \iff \frac{2\alpha}{t^3} \geq 2r^2\alpha^3 \iff 1 \geq \alpha^2 r^2,$$

which gives $r \leq \frac{1}{\alpha}$. Therefore the assumption $(H_{\gamma,r})$ is satisfied by taking $r = \frac{1}{\alpha}$. Let us now distinguish the cases $\alpha \leq 3$ and $\alpha > 3$.

(1) If $\alpha \leq 3$, then condition $(H_{\gamma,r})$ is fulfilled with $r = \frac{1}{\alpha}$ and we deduce from Theorem 2.1 that

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^{\frac{2}{3}}}\right) \quad \text{as } t \to +\infty.$$

Suppose moreover that $x(\cdot)$ is bounded. Then,

$$\|\dot{x}(t)\| = \mathcal{O}\left(\frac{1}{t^{\frac{2}{3}}}\right) \quad \text{as } t \to +\infty.$$

(2) If $\alpha > 3$, then condition $(H_{\gamma,r})$ is satisfied with $r = \frac{1}{\alpha} < \frac{1}{\alpha}$ and we infer from Theorem 3.1 that

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{t^{\frac{2}{3}}}\right) \quad \text{and} \quad \|\dot{x}(t)\| = o\left(\frac{1}{t^{\frac{1}{2}}}\right) \quad \text{as } t \to +\infty.$$

We can complete these results with the help of Propositions 3.1 and 3.2. In doing so, we derive the following statement, which covers the results obtained in [4] (case $\alpha \geq 3$), and in [1], [5], (case $\alpha \leq 3$).

**Corollary 4.1.** Given $\alpha > 0$, assume that $\gamma(t) = \frac{q}{t^2}$ for every $t \geq t_0 > 0$. Let $x(\cdot)$ be a solution of (IGS)$_{\gamma}$. Then, based on $\alpha$ lower or higher than 3, we have the following results:

a) Case $\alpha \leq 3$:

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^{\frac{2}{3}}}\right) \quad \text{as } t \to +\infty. \quad \text{If } x(\cdot) \text{ is bounded, then } \|\dot{x}(t)\| = \mathcal{O}\left(\frac{1}{t^{\frac{2}{3}}}\right) \quad \text{as } t \to +\infty.$$

b) Case $\alpha > 3$:

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{t^{\frac{2}{3}}}ight) \quad \text{and} \quad \|\dot{x}(t)\| = o\left(\frac{1}{t^{\frac{1}{2}}}ight) \quad \text{as } t \to +\infty.$$

Moreover, we have

$$\int_{t_0}^{+\infty} t (\Phi(x(t)) - \min_{\mathcal{H}} \Phi) \, dt < +\infty \quad \text{and} \quad \int_{t_0}^{+\infty} \|\dot{x}(t)\|^2 \, dt < +\infty.$$

Thus, we get a unified picture of the convergence properties for (IGS)$_{\gamma}$ in the case $\gamma(t) = \frac{q}{t^2}$, $\alpha > 0$. Cases $\alpha > 3$ and $\alpha \leq 3$ are clearly asymmetric. When $\alpha > 3$, the trajectories of (IGS)$_{\gamma}$ converge weakly and hence are bounded, see Theorem 1.1. On the other hand, when $\alpha \leq 3$, the convergence of the trajectories is not known, nor their boundedness. This is why the boundedness assumption is required in this case, see Corollary 4.1 a). We emphasize that historically the case $\alpha \geq 3$ was first solved. It is only recently that the case $0 < \alpha \leq 3$ has been well understood.
Remark 4.1. In [2] the authors give a complete picture in the case \( \gamma(t) \geq \frac{3}{t} \). In particular, this study covers the case \( \gamma(t) = \frac{\theta}{t^\alpha} \) for \( 0 < \theta < 1 \), which models a slow decrease to zero of the damping function. Our approach complements this study and covers the "subcritical case" \( \gamma(t) \leq \frac{3}{t} \). It might be tempting to have a unifying analysis including these two cases. Each of them relies on different Lyapunov functions. We believe that they rely on different regimes, and that a unifying theory would be too complicated technically, and therefore would not offer decisive progress.

4.2. Case \( \gamma(t) = \frac{1}{t(\ln t)^\beta} \), \( 0 \leq \beta \leq 1 \). Elementary computation gives

\[
\dot{\gamma}(t) = -\frac{\ln t + \beta}{t^2(\ln t)^{\beta+1}}
\]

and

\[
\ddot{\gamma}(t) = \frac{2(\ln t)^2 + 3\beta \ln t + \beta(\beta + 1)}{t^3(\ln t)^{\beta+2}}.
\]

Therefore, condition \((H_{\gamma,r})\) writes as follows: find \( r > 0 \) such that for \( t \) large enough

\[
\frac{2(\ln t)^2 + 3\beta \ln t + \beta(\beta + 1)}{t^3(\ln t)^{\beta+2}} \geq 2r^2 \frac{1}{t^3(\ln t)^{3\beta}}.
\]

Equivalently

\[
2(\ln t)^2 + 3\beta \ln t + \beta(\beta + 1) \geq 2r^2(\ln t)^{2(1-\beta)}.
\]

Since \( 0 \leq \beta \leq 1 \), this inequality is satisfied for any \( r \leq 1 \) and any \( t \geq e \). Let us compute \( p(t) \).

\[
p(t) = \exp\left(\int_0^t \frac{1}{s(\ln s)^\beta} ds\right) = \exp\left(\int_{\ln t_0}^{\ln t} \frac{1}{u^\beta} du\right).
\]

We have to distinguish two cases:

a) \( \beta = 1 \). In this case, \( p(t) = \frac{\ln t}{\ln t_0} \), and we get from Theorem 2.1 applied with \( r = \frac{1}{3} \)

\[
\Phi(x(t)) - \min_{\mathcal{H}} \Phi = O\left(\frac{1}{(\ln t)^{\frac{1}{2}}}\right) \quad \text{as } t \to +\infty.
\]

b) \( 0 \leq \beta < 1 \). Then, \( p(t) = \exp\left(\frac{1}{1-\beta} \left((\ln t)^{1-\beta} - (\ln t_0)^{1-\beta}\right)\right) \), and we get from Theorem 2.1 applied with \( r = \frac{1}{3} \)

\[
\Phi(x(t)) - \min_{\mathcal{H}} \Phi = O\left(\frac{1}{\exp\left(\frac{2}{\beta(1-\beta)} (\ln t)^{1-\beta}\right)}\right) \quad \text{as } t \to +\infty.
\]

Let us observe that when \( \beta \to 0 \) or \( \beta \to 1 \), we recover the previous formulae.

We can now complete the table giving a synthetic view about the speed of convergence in the different situations studied above.

<table>
<thead>
<tr>
<th>( \gamma(t) )</th>
<th>( \gamma(t) = \frac{1}{t(\ln t)^\beta}, \beta \in [0,1[ )</th>
<th>( \gamma(t) = \frac{2}{t}, \alpha \in [0,3] )</th>
<th>( \gamma(t) = \frac{2}{t}, \alpha &gt; 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( W(t) )</td>
<td>( O\left(\frac{1}{(\ln t)^\frac{1}{2}}\right))</td>
<td>( O\left(\frac{1}{\exp\left(\frac{2}{3(1-\beta)} (\ln t)^{1-\beta}\right)}\right))</td>
<td>( O\left(\frac{1}{t^{\frac{3}{2}}}\right))</td>
</tr>
</tbody>
</table>
5. The perturbed case

We consider the equation

\[(IGS)_{\gamma,g}\]

\[\dot{x}(t) + \gamma(t)\dot{x}(t) + \nabla \Phi(x(t)) = g(t)\]

with a second member \(g : [t_0, +\infty[ \to \mathcal{H}\). Depending on the situation, \(g\) is a forcing term, or comes from the approximation or computational errors in \((IGS)_{\gamma}\). To ensure existence and uniqueness for the corresponding Cauchy problem, we suppose that \(g\) is locally integrable. We will show that the results of the previous sections remain satisfied if the perturbation \(g\) is not too large asymptotically. This reflects some structural stability of the dynamics \((IGS)_{\gamma}\).

**Theorem 5.1.** Let \(\gamma(\cdot) : [t_0, +\infty[ \to \mathbb{R}_+\) be a nonincreasing function that is twice continuously differentiable, and satisfies \((H_{\gamma,r})\) for some \(r \in [0, 1/3]\). Suppose that \(g : [t_0, +\infty[ \to \mathcal{H}\) is locally integrable and satisfies

\[\int_{t_0}^{+\infty} e^{t - t_0} \gamma(s) ds \|g(t)\| dt < +\infty.\]

Then, for each solution trajectory \(x : [t_0, +\infty[ \to \mathcal{H}\) of \((IGS)_{\gamma,g}\)

\[\Phi(x(t)) = \min_{\mathcal{H}} \Phi = \mathcal{O}\left(e^{-2r \int_{t_0}^{t} \gamma(s) ds}\right) \text{ as } t \to +\infty.\]

**Proof.** The idea of proof is similar to that used in [2], [4], [5]. It consists of modifying the energy function \(E(\cdot)\) considered in the unperturbed case (see Section 2) by introducing an additional term taking account of the perturbation \(g\). Let \(t_1 \geq t_0\) be such that the inequality \(\hat{\gamma}(t) \geq 2r^2 \gamma(t)^3\) is satisfied for \(t \geq t_1\). Let us first fix some \(T > t_1\), and define for \(t \in [t_1, T]\) (without ambiguity we keep the same notations as in Section 2)

\[E(t) = \frac{1}{2} \|\lambda(t)(x(t) - z) + p(t)^T \dot{x}(t)\|^2 + \frac{\xi(t)}{2} \|x(t) - z\|^2 + \frac{\xi(t)}{2} \langle \lambda(t)(x(t) - z) + p(t)^T \dot{x}(t), \lambda(t)(x(t) - z) + p(t)^T \dot{x}(t) - g(t) \rangle.\]

By derivation of \(E(\cdot)\), we obtain

\[\frac{d}{dt} E(t) = 2r \gamma(t) p(t)^T (\Phi(x(t)) - \min_{\mathcal{H}} \Phi) + (\nabla \Phi(x(t)), \dot{x}(t)) + \frac{\xi(t)}{2} \|x(t) - z\|^2 + \xi(t) \langle x(t) - z, \dot{x}(t) - g(t) \rangle.\]

Note that the term \(\dot{x}(t) - g(t) = -\frac{d}{dt} \langle \Phi(x(t)) + \nabla \Phi(x(t)) \rangle\) gives exactly the same expression as in the unperturbed case. Therefore, by a similar argument, we obtain that, under the assumption \((H_{\gamma,r})\) with \(r \in [0, 1/3]\), \(\frac{d}{dt} E(t) \leq 0\) for \(t \geq t_1\). As a result, the energy function \(E(\cdot)\) is nonincreasing. In particular, for \(t \geq t_1\), we have \(E(t) \leq E(t_1)\), which gives, by definition of \(E(\cdot)\)

\[p(t)^2 \left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right) + \frac{1}{2} \|\lambda(t)(x(t) - z) + p(t)^T \dot{x}(t)\|^2 \leq C + \int_{t_1}^{t} \langle \lambda(s)(x(s) - z) + p(s)^T \dot{x}(s), p(s)^T g(s) \rangle ds,\]

where \(C\) is a constant that depends only on the values of the data at time \(t_1\). Applying Cauchy-Schwarz inequality, and since \(t_0 \leq t_1\), we deduce that

\[\frac{1}{2} \|\lambda(t)(x(t) - z) + p(t)^T \dot{x}(t)\|^2 \leq C + \int_{t_0}^{t} \|\lambda(s)(x(s) - z) + p(s)^T \dot{x}(s)\| \|p(s)^T g(s)\| ds.\]

Applying Gronwall inequality, see Lemma A.1 below, we obtain

\[\|\lambda(t)(x(t) - z) + p(t)^T \dot{x}(t)\| \leq \sqrt{2C} + \int_{t_0}^{t} p(s)^T \|g(s)\| ds.\]

By assumption \(\int_{t_0}^{+\infty} p(t)^T \|g(t)\| dt < +\infty\). Hence

\[\sup_{t \geq t_1} \|\lambda(t)(x(t) - z) + p(t)^T \dot{x}(t)\| \leq \sqrt{2C} + \int_{t_0}^{+\infty} p(s)^T \|g(s)\| ds < +\infty.\]

Injecting this estimate into (25), we obtain for \(t \geq t_1\)

\[p(t)^2 \left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right) \leq C + \left[\sqrt{2C} + \int_{t_0}^{+\infty} p(s)^T \|g(s)\| ds\right] \int_{t_0}^{+\infty} p(s)^T \|g(s)\| ds,\]

which immediately gives the claim. \(\square\)
Remark 5.1. Because of the numerical importance of the subject, several articles have been devoted to the study of the effect of perturbations on inertial gradient-like algorithms: to cite a few of them Aujol-Dossal [7], Schmidt-Le Roux-Bach [24], Solodov-Svaiter [25], Villa-Salzo-Baldassarres-Verri [27].

APPENDIX A

Proof of Proposition 2.2. Let us introduce the modified energy function \( \tilde{W} : [t_0, +\infty] \to \mathbb{R} \) given by

\[
\tilde{W}(t) = W(t) + \frac{\gamma(t)}{2} \langle \dot{x}(t), x(t) \rangle = \frac{1}{2} \| x(t) \|^2 + \frac{1}{2} \| \dot{x}(t) \|^2 + \frac{\gamma(t)}{2} \langle \dot{x}(t), x(t) \rangle.
\]

By differentiating the function \( \tilde{W} \), we find for every \( t \geq t_0 \)

\[
\dot{\tilde{W}}(t) = \dot{W}(t) + \frac{\gamma(t)}{2} \langle \dot{x}(t), x(t) \rangle + \frac{\gamma(t)}{2} \langle \dot{x}(t), x(t) \rangle + \frac{\gamma(t)}{2} \| \dot{x}(t) \|^2 = -\frac{\gamma(t)}{2} \| \dot{x}(t) \|^2 - \frac{\gamma(t)}{2} \| x(t) \|^2 + \left( \frac{\gamma(t)}{2} - \frac{\gamma(t)^2}{2} \right) \langle \dot{x}(t), x(t) \rangle.
\]

Therefore we have

\[
\dot{\tilde{W}}(t) + \gamma(t) \tilde{W}(t) = \frac{\gamma(t)}{2} \langle \dot{x}(t), x(t) \rangle.
\]

Observe that

\[
|\langle \dot{x}(t), x(t) \rangle| \leq \frac{1}{2} \| \dot{x}(t) \|^2 + \frac{1}{2} \| x(t) \|^2 = W(t).
\]

Recalling that \( \lim_{t \to +\infty} \gamma(t) = 0 \), the expression of \( \tilde{W} \) shows that

\[
\tilde{W}(t) \sim W(t) \quad \text{as} \quad t \to +\infty.
\]

We deduce from (26), (27) and (28) the existence of \( t_1 \geq t_0 \) such that for every \( t \geq t_1 \),

\[
\left| \dot{\tilde{W}}(t) + \gamma(t) \tilde{W}(t) \right| \leq |\gamma(t)| \tilde{W}(t).
\]

Let us multiply each member of this inequality by \( e^{\int_{t_1}^t \gamma(s)ds} \) and set \( F(t) = e^{\int_{t_1}^t \gamma(s)ds} \tilde{W}(t) \). We obtain

\[
|\dot{F}(t)| \leq |\gamma(t)| F(t).
\]

Observe that if \( F(t_2) = 0 \) for some \( t_2 \geq t_1 \), then we have \( \tilde{W}(t_2) = 0 \) and \( W(t_2) = 0 \). Since the map \( W \) is nonincreasing, we conclude that \( W(t) = 0 \) for every \( t \geq t_2 \), i.e. the solution \( x \) is stationary. Now assume that \( F(t) > 0 \) for every \( t \geq t_1 \) and divide each member of inequality (29) by \( F(t) \). Since \( \dot{\gamma} \in L^1(t_0, +\infty) \) by assumption, we deduce that

\[
\left| \frac{d}{dt} \ln F \right| (t) = \left| \frac{\dot{F}(t)}{F(t)} \right| \in L^1(t_0, +\infty).
\]

It ensues that \( \lim_{t \to +\infty} \ln F(t) \) exists in \( \mathbb{R} \). We deduce that \( \lim_{t \to +\infty} e^{\int_{t_1}^t \gamma(s)ds} \tilde{W}(t) = K > 0 \). The conclusion immediately follows from estimate (28). \( \square \)

Lemma A.1. Let \( m : [t_0, T] \to [0, +\infty) \) be integrable, and let \( c > 0 \). Suppose \( w : [t_0, T] \to \mathbb{R} \) is continuous and

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

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

\( ^2 \)The use of such an auxiliary function is classical, see for example [18, Lemma 3.2.6] in the case of an autonomous damping.
DAMPED INERTIAL GRADIENT DYNAMICS

References


