An ODE-based modeling of inertial Forward-Backward algorithms

Apidopoulos Vassilis
Université de Bordeaux
IMB, UMR 5251, F-33400 Talence, France
vasileios.apidopoulos@math.u-bordeaux.fr

Aujol Jean-François
Université de Bordeaux
IMB, UMR 5251, F-33400 Talence, France
jean-francois.aujol@math.u-bordeaux.fr

Dossal Charles
Université de Bordeaux
IMB, UMR 5251, F-33400 Talence, France
Charles.Dossal@math.u-bordeaux1.fr

***Abstract*—In this work we are interested in the asymptotic behaviour of the trajectory of solutions of a differential equation, driven by a discrete scheme, which corresponds to a particular inertial Forward-Backward (i-FB) algorithm considered in [2]. The interest of studying this ODE is its connexion with the fast minimization of a convex differentiable function F. More precisely, under some appropriate hypothesis, the convergence rates of the values of the functional F to its minimum and the norm of the velocity of the trajectory-solution are the same order to the ones obtained in [2] of their "discretized versions".**

I. Introduction

Let $\mathcal{H}$ be a separable Hilbert space endowed with the scalar product $\langle , \rangle$ and the norm $\|\cdot\|$. Let $t_0 \geq 1$ be a real number. We consider the function $F = f + g$, where F is coercive, $f, g : \mathcal{H} \to \mathbb{R}$ are lower semi-continuous convex functions and f is of class $\mathscr{C}^1(\mathcal{H})$ with ∇f L-Lipschitz. We denote by x^* a minimizer of F.

We split the paper into two parts. We first consider a classical ODE setting and then we consider the differential inclusion case.

A. The differential equation case

In this first setting we assume that $F \in \mathscr{C}^1(\mathcal{H})$ (that is $g \in \mathscr{C}^1(\mathcal{H})$) We propose to study the behaviour of trajectories of solutions of the following differential equation :

$$\ddot{x}(t) + \left(\frac{d}{t} + \frac{a^d}{t^d}\right)\dot{x}(t) + \nabla F(x(t)) = 0 \tag{E}$$

where $d \in (0, 1]$ and $a > 0$.

The motivation for the study of this differential equation comes from the fact that it models a specific inertial Forward-Backward algorithm which was introduced in [2]. In other words a discretization of (E) corresponds to this algorithm. We will show that under the hypothesis that (E) admits a solution in $[t_0 + \infty)$ with some initial conditions $(x(t_0), \dot{x}(t_0)))$ and under some supplementary hypothesis on the constant $a > 0$ as in [2], we can derive uniform bounds for $t^{2d}W(t) = t^{2d}(F(x(t)) - F(x^*))$ and $t^{2d}\|\dot{x}(t)\|^2$.

The analysis is similar to the one carried out in [1] and it is based on a Lyapunov energy function associated to (E) which was first considered in [8] and in [1], where the case $d = 1$ is treated (which corresponds to an ODE modeling the FISTA algorithm considered in [7], [4], [6], [1] and [9]). As a by-product the results allow us to deduce the weak convergence property of the trajectory $x(t)$ towards a minimizer x^* as already shown in [5].

Theorem I.1. *Let $x : [t_0, \infty) \longrightarrow \mathcal{H}$ be a solution of* (E). *There exist some positive constants C_1 and C_2, such that the following bounds hold for all $t \in (t_0, +\infty)$:*

- *If $a^d \geq 2d$ then :*

$$F(x(t)) - F(x^*) \leq \frac{C_1}{t^{2d}} \quad \textit{and } \|\dot{x}(t)\|^2 \leq \frac{C_2}{t^{2d}} \tag{1}$$

- *If $a^d > 2d$ then :*

$$\int_{t_0}^{+\infty} t^d(F(x(t)) - F(x^*))dt < +\infty \tag{2}$$

$$\textit{and} \quad \int_{t_0}^{+\infty} t^d\|\dot{x}(t)\|^2 dt < +\infty \tag{3}$$

B. The differential inclusion case

In a second time we turn our interest onto the following differential inclusion :

$$\ddot{x}(t) + \left(\frac{d}{t} + \frac{a^d}{t^d}\right)\dot{x}(t) + \partial F(x(t)) \ni 0 \tag{DI}$$

where $d \in (0, 1]$, $a > 0$. This framework corresponds in a more "direct" way to the i-FB algorithm studied in [2], since we do not make the supplementary hypothesis of $F \in \mathscr{C}^1(\mathcal{H})$ (or $g \in \mathscr{C}^1(\mathcal{H})$). In addition we suppose that $\mathcal{H}$ is of finite dimension. Apart of existence and uniqueness of a solution of (DI) in $[t_0, +\infty)$ issues, it turns out that the uniform bounds of the first point of Theorem I.1, still hold true, given that a classical solution to (DI) exists. In particular we have :

Theorem I.2. *Let $x : [t_0, +\infty) \longrightarrow \mathcal{H}$ be a (classical) solution of* (DI). *If $a \geq (2d)^{\frac{1}{d}}$, then there exist some positive constants C_1 and C_2, such that the following bounds hold for all $t \in (t_0, +\infty)$:*

$$W(t) \leq \frac{C_1}{t^{2d}} \quad \textit{and} \quad \|\dot{x}(t)\|^2 \leq \frac{C_2}{t^{2d}} \tag{4}$$

Future directions

As a future objectif, we aim at extending this study to a perturbed version of (E) :

$$\ddot{x}(t) + \left(\frac{d}{t} + \frac{a^d}{t^d}\right)\dot{x}(t) + \nabla F(x(t)) + p(t) = 0 \tag{EP}$$

which is similar to the one in [3]. The interest is to derive the same uniform bounds in $(t_0, +\infty)$, as in Theorem I.1 and to show that we can have a better trade-off between the different convergence rates and the impact of different perturbation levels for $d \in (0, 1]$ (it formally amounts to considering $\int_{t_0}^{+\infty} t^d\|p(t)\|dt < +\infty$ for $d \in (0, 1]$).

Acknowledgements

This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR GOTMI) (ANR-16-CE33-0010-01). Jean-François. Aujol is a member of Institut Universitaire de France (IUF).

REFERENCES

[1] Hedy Attouch, Zaki Chbani, Juan Peypouquet, and Patrick Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Mathematical Programming*, pages 1–53, 2016.

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

[3] Mounir Balti and Ramzi May. Asymptotic for the perturbed heavy ball system with vanishing damping term. *arXiv preprint arXiv:1609.00135*, 2016.

[4] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[5] Alexandre Cabot and Pierre Frankel. Asymptotics for some semilinear hyperbolic equations with non-autonomous damping. *Journal of Differential Equations*, 252(1):294–322, 2012.

[6] A Chambolle and Ch Dossal. On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". *Journal of Optimization Theory and Applications*, 166(3):968–982, 2015.

[7] Yurii Nesterov. A method of solving a convex programming problem with convergence rate o (1/k2). In *Soviet Mathematics Doklady*, volume 27, pages 372–376, 1983.

[8] Weijie Su, Stephen Boyd, and Emmanuel J Candes. A differential equation for modeling nesterov's accelerated gradient method: theory and insights. *Journal of Machine Learning Research*, 17(153):1–43, 2016.

[9] Andre Wibisono, Ashia C Wilson, and Michael I Jordan. A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, page 201614734, 2016.