The optimal decay for the solution of the Monotone Inclusion associated to FISTA for b

Jf Aujol, Charles H Dossal

The optimal decay for the solution of the Monotone Inclusion associated to FISTA for $b \leqslant 3$ is $\frac{2b}{3}$

JF Aujol *and Ch. Dossal †
IMB, Univ. Bordeaux

July 2017

Abstract

It was recently proved that the decay of the solution of the ODE associated to the Nesterov Fast Gradient Algorithm with a parameter $b \leqslant 3$ was $0(\frac{1}{t^{\frac{2b}{3}}})$. In this note we prove that this decay is achieved for the solution of the associated monotone inclusion for a specific function.

1 Setting

We study the following inclusion

$$\ddot{x}(t) + \frac{b}{t}\dot{x}(t) + \partial F(x(t)) \ni 0 \tag{1}$$

for $t \geqslant t_0 > 0$, where F is the absolute value defined by $F(x) = |x|$ and $b \in (0,3)$. This monotone inclusion for a general convex function F is associated to the optimization Algorithm FISTA of Beck et al. [4] and the Fast Gradient Descend Method of Nesterov [6] as it was remarked by Su et al in [7].
It was proved in [2] and [3] that for

$$\ddot{x}(t) + \frac{b}{t}\dot{x}(t) + \nabla F(x(t)) = 0 \tag{2}$$

*Jean-Francois.Aujol@math.u-bordeaux.fr
†Charles.Dossal@math.u-bordeaux.fr

where F is a differentiable convex function and $b \in (0,3)$ that if x^* is a minimizer of F,

$$F(x(t)) - F(x^*) = O\left(\frac{1}{t^{\frac{2b}{3}}}\right) \tag{3}$$

In this note, the main result is the following: we prove that this decay is actually optimal in the sense that it is achieved for (1) with $F(x) = |x|$. More precisely

Proposition 1. *Let $b \leqslant 3$, if $x(t_0) \neq 0$, for any solution x of* (1)

1. *it exists K_1 such that for any $t \geqslant t_0$,*

$$|x(t)| \leqslant \frac{K_1}{t^{\frac{2b}{3}}}, \tag{4}$$

2. *it exists $K_2 > 0$ such that for any $T > 0$, there exists $t > T$ such that*

$$|x(t)| \geqslant \frac{K_2}{t^{\frac{2b}{3}}} \tag{5}$$

Notice that for the case of the ODE (2), it was proved in [3] that the rate of (3) is asymptotically optimal by considering $F(x) = |x|^{(1+\beta)}$ with $\beta > 0$. The ideas presented in this note are an adaptation to the differential inclusion case of the results presented in [3]. In the first part, preliminaries results are given and in a second part the Propostion 1 is proved.

2 Preliminaries

We consider the inclusion

$$\ddot{x}(t) + \frac{b}{t}\dot{x}(t) + \partial F(x(t)) \ni 0 \tag{6}$$

with $F(x) = |x|$, and with initial condition $x(t_0) = x_0$, $\dot{x}(t_0) = v_0$. Notice that F admits a unique minimizer $x^* = 0$.

Let $I = [t_0, +\infty)$. In [1], it is proven that there exists a weak solution x of (1), with x in $W^{2,\infty}(I;\mathbb{R}) \cap C^{1,\lambda}(K;\mathbb{R})$, where $\lambda \in (0,1)$ and K is a compact subset of I, such that $x(t_0) = x_0$ and $\dot{x}(t_0) = v_0$. In particular, $\dot{x}$ is continuous and $\ddot{x}$ is bounded.

We recall that if F is a convex one homogeneous function, then we have $F(x) = zx$ for any z in $\partial F(x)$.

We also give the following lemma that will prove useful. If F is a proper lower semi continuous convex function, then there exists $z(t)$ in $\partial F(x(t))$ such that for almost every t (see Lemma 3.3 in [5]) for more details).

$$\frac{d}{dt}F(x(t)) = z(t)\dot{x}(t) \tag{7}$$

Moreover, $t \mapsto F(x(t))$ is an absolutely continuous function. We recall here that an absolutely continuous function is differentiable almost everywhere. Notice that in particular, if the derivative of an absolutely continuous function is non positive, then this function is non decreasing. See Lemma 4.1 in [1] for more details. These properties will be used intensively in the rest of this note.

Notations If $\lambda \in \mathbb{R}$ and x a solution of (1) we denote :

$$u(t) := t|x(t)| + \frac{1}{2t}|\lambda(x(t)) + t\dot{x}(t)|^2 \tag{8}$$

and

$$v(t) := \frac{1}{2t}|x(t)|^2 \tag{9}$$

The analysis is based on Lyapunov functions. Following Su, Boyd and Candès [7], for any $(\lambda, \xi) \in \mathbb{R}^2$ we define

$$\mathcal{E}_{\lambda,\xi}(t) = t^2|x(t)| + \frac{1}{2}|\lambda(x(t)) + t\dot{x}(t)|^2 + \frac{\xi}{2}|x(t)|^2. \tag{10}$$

One can observe that this energy can be defined using the functions u and v

$$\mathcal{E}(t) = tu(t) + \xi tv(t). \tag{11}$$

As usual, we need to compute the derivative $\mathcal{E}'_{\lambda,\xi}$. We will need two different expressions of the derivative.

Lemma 1.

$$\mathcal{E}'_{\lambda,\xi}(t) =(2-\lambda)t|x(t)| + t(\lambda+1-b)|\dot{x}(t)|^2 \tag{12}$$

$$+ (\lambda(\lambda+1-b)+\xi)\dot{x}(t)x(t) \tag{13}$$

for almost every t.

Proof. We differentiate and we use the recalls of the preliminary section to get the following computations that hold for almost every t :

$$\mathcal{E}'_{\lambda,\xi}(t) =2t|x(t)| + t^2 z(t)\dot{x}(t) \tag{14}$$

$$+ \lambda\dot{x}(t) + t\ddot{x}(t) + \dot{x}(t)\lambda(x(t)) + t\dot{x}(t) + \xi\dot{x}(t)x(t) \tag{15}$$

with $z(t) \in \partial F(x(t))$. We now make use of the differential inclusion:

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& 2t|x(t)| + t^2 z(t)\dot{x}(t) && (16)\\ &+ \lambda\dot{x}(t) + t\left(-z - \frac{b}{t}\dot{x}(t)\right) + \dot{x}(t)\lambda(x(t)) + t\dot{x}(t) + \xi\dot{x}(t)x(t) && (17)\end{aligned}$$

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& 2t|x(t)| + t^2 z(t)\dot{x}(t) && (18)\\ &+ (\lambda + 1 - b)\dot{x}(t) - tz\lambda(x(t)) + t\dot{x}(t) + \xi\dot{x}(t)x(t) && (19)\end{aligned}$$

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& 2t|x(t)| + x(t)(\lambda + 1 - b)\dot{x}(t) && (20)\\ &- \lambda x(t)z + (\lambda + 1 - b)t|\dot{x}(t)|^2 + \xi\dot{x}(t)x(t) && (21)\end{aligned}$$

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& (2 - \lambda)t|x(t)| + t(\lambda + 1 - b)|\dot{x}(t)|^2 && (22)\\ &+ (\lambda(\lambda + 1 - b) + \xi)\dot{x}(t)x(t) && (23)\\ & && (24)\end{aligned}$$

□

Lemma 2.

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& (2 - \lambda)t|x(t)| + (\xi - \lambda(\lambda + 1 - b))\dot{x}(t)x(t) && (25)\\ &- \frac{\lambda^2(\lambda + 1 - b)}{t}|x(t)|^2 && (26)\\ &+ \frac{\lambda + 1 - b}{t}|\lambda(x(t)) + t\dot{x}(t)|^2 && (27)\\ & && (28)\end{aligned}$$

Proof. We start from the result of Lemma 1. Observing that

$$\frac{1}{t}|\lambda(x(t)) + t\dot{x}(t)|^2 = t|\dot{x}(t)|^2 + 2\lambda\dot{x}(t)x(t) + \frac{\lambda^2}{t}|x(t)|^2 \quad (29)$$

we can write for almost every t

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& (2 - \lambda)t|x(t)| + (\lambda(\lambda + 1 - b) + \xi)\dot{x}(t)x(t) && (30)\\ &- 2\lambda(\lambda + 1 - b)\dot{x}(t)x(t) - (\lambda + 1 - b)\frac{\lambda^2}{t}|x(t)|^2 && (31)\\ &+ (\lambda + 1 - b)\frac{1}{t}|\lambda(x(t)) + t\dot{x}(t)|^2 && (32)\\ & && (33)\end{aligned}$$

□

Notations Now, we choose $\lambda = \frac{2b}{3}$ and $\xi = \lambda(\lambda+1-b)$ and we denote $\mathcal{E}_{\lambda,\xi} = \mathcal{E}$. For these parameters one can observe that $\lambda + 1 - b = 1 - \frac{b}{3} \geqslant 0$. We can define

$$c := 2 - \frac{2b}{3} \tag{34}$$

Notice that $c > 0$. We introduce the new Lyapunov energy:

$$\mathcal{H}(t) := t^{-c}\mathcal{E}(t). \tag{35}$$

Lemma 3. *We have for almost every t:*

$$\mathcal{E}'(t) = cu(t) - dv(t). \tag{36}$$

where $d > 0$ is defined as:

$$d := \left(\frac{2b}{3}\right)^2 \left(1 - \frac{b}{3}\right) \tag{37}$$

Proof.

$$\mathcal{E}'(t) =(2 - \frac{2b}{3})t|x(t)| + (1 - \frac{b}{3})\frac{1}{t}|\frac{2b}{3}(x(t)) + t\dot{x}(t)|^2 \tag{38}$$

$$- \left(\frac{2b}{3}\right)^2 (1 - \frac{b}{3})\frac{1}{t}|x(t)|^2 \tag{39}$$

Using

$$u(t) = t|x(t)| + \frac{1}{2t}|\frac{2b}{3}(x(t) - x^*) + t\dot{x}(t)|^2, \tag{40}$$

the definition of $v(t)$, c and d, we get the result. □

Proposition 2. *If $b \in (0, 3)$, if $\lambda = \frac{2b}{3}$ and $\xi = \frac{2b}{9}(3 - b)$, then for almost every t it holds*

$$\mathcal{H}'(t) = -2\xi t^{-c}v(t) \leqslant 0. \tag{41}$$

and $\mathcal{H}$ is an absolutely continuous function.

We can observe that for $b = 3$, we have, $c = 0$, $\xi = 0$, $\mathcal{E} = \mathcal{H}$ and $\mathcal{H}$ is constant.

Proof. $\mathcal{H}$ is an absolutely continuous function as a sum of absolutely continuous functions.

We have for almost every t

$$\mathcal{H}'(t) = t^{-c-1}(t\mathcal{E}'(t) - c\mathcal{E}(t)). \tag{42}$$

Using Lemma 3, we deduce that for almost every t:

$$\mathcal{H}'(t) = -2\xi t^{-c}v(t) \leqslant 0. \tag{43}$$

□

Lemma 4. *If* $x(t_0) > 0$ *then* $\lim_{t\to+\infty} \mathcal{H}(t) = \ell > 0$.

Proof. The function $\mathcal{H}$ is non-negative and from Proposition 2, $\mathcal{H}$ is non-increasing, thus $\mathcal{H}(t)$ converges to a limit $\ell \geqslant 0$.
Moreover for all $t \geqslant t_0$ we have

$$t^{-c}\mathcal{E}(t) = \mathcal{H}(t) \leqslant \mathcal{H}(t_0) \tag{44}$$

which implies that

$$|x(t)| \leqslant \frac{\mathcal{H}(t_0)}{t^{2-c}} = \frac{\mathcal{H}(t_0)}{t^{\frac{2b}{3}}} \tag{45}$$

We recall that $c := 2 - \frac{2b}{3}$ and then $c \in [0, 2)$ when $b \in (0, 3]$. From (43)

$$\mathcal{H}'(t) = -\xi t^{-c-1}|x(t)|^2 = -\xi t^{c-5} t^{2-c}|x(t)| t^{2-c}|x(t)|. \tag{46}$$

and from the definition of $\mathcal{H}$ we have $\mathcal{H}(t) \geqslant t^{2-c}|x(t)|$ which implies that

$$\mathcal{H}'(t) \geqslant -\xi t^{c-5} t^{2-c}|x(t)|\mathcal{H}(t) \tag{47}$$

Unsing the decay of $\mathcal{H}$ we have $\mathcal{H}(t_0) \geqslant t^{2-c}|x(t)|$ and thus

$$\mathcal{H}'(t) \geqslant -\xi t^{c-5}\mathcal{H}(t_0)\mathcal{H}(t). \tag{48}$$

If we denote φ the function defined from $[t_0, +\infty]$ to $\mathbb{R}$ by

$$\varphi(t) := \xi \frac{\mathcal{H}(t_0)}{c-4} t^{c-4} \tag{49}$$

we have

$$\varphi'(t) = \xi \mathcal{H}(t_0) t^{c-5} \tag{50}$$

and defining $\mathcal{G}$ by

$$\mathcal{G}(t) := \mathcal{H}(t) e^{\varphi(t)} \tag{51}$$

which is an absolutely continuous function as a product of an absolutely continuous function by a continuous and bounded function, we have

$$\mathcal{G}'(t) := e^{\varphi(t)} \left(\mathcal{H}'(t) + \varphi'(t)\mathcal{H}(t)\right) \tag{52}$$

which is non negative from (48).
It follows that $\mathcal{G}$ is a non-decreasing function and thus that for all $t \geqslant t_0$

$$\mathcal{H}(t) \geqslant \mathcal{H}(t_0) e^{\varphi(t_0)-\varphi(t)} \geqslant \mathcal{H}(t_0) e^{\varphi(t_0)} > 0 \tag{53}$$

which ends the proof of the lemma. □

3 Proof of Proposition 1

The first point of the Proposition was proved in (45). To prove the second point of Proposition 1 we observe that

$$\mathcal{E}(t) = t^2|x(t)| + \frac{1}{2}|\lambda x(t) + t\dot{x}(t)|^2 + \frac{\xi}{2}|x(t)|^2. \tag{54}$$

Let $K_1 = \mathcal{H}(t_0)e^{\varphi(t_0)}$, from (53), $\forall t \geqslant t_0$, $\mathcal{E}(t) \geqslant K_1 t^{2-\frac{2b}{3}}$. We can notice here that $2 - \frac{2b}{3} \geqslant 0$. There are some cases where the conclusion holds directly.

1. If for $t_2 > T$, $\frac{1}{2}|\lambda x(t_2) + t\dot{x}(t_2)|^2 + \frac{\xi}{2}|x(t_2)|^2 \leqslant \frac{K_1}{2}t_0^{2-\frac{2b}{3}}$ we have $t_2^2|x(t_2)| \geqslant K_1 t_2^{2-\frac{2b}{3}} - \frac{K_1}{2}t_0^{2-\frac{2b}{3}} \geqslant \frac{K_1}{2}t_2^{2-\frac{2b}{3}}$ and thus we can conclude.
2. If there exists $t > T$ such that $\dot{x}(t) = 0$, using the fact that $\mathcal{E}(t) = t^2|x(t)| + \frac{\lambda^2+\xi}{2}|x(t)|^2$ and the fact that $\lim_{t\to\infty} \frac{\lambda^2+\xi}{2}|x(t)|^2 = 0$, we can conclude using the previous point.
3. If there exists $t > T$ such that $x(t) = 0$, since $\lim_{u\to\infty} |x(u)| = 0$, there exists $t_1 \geqslant t$ such that $\dot{x}(t_1) = 0$ and we can use the previous point.

We now suppose that $x(T) > 0$ and the sign of $\dot{x}$ is constant on $[T, +\infty)$. Since $\lim_{t\to\infty} x(t) = 0$ we deduce that $\forall t \in (T, +\infty)$, $\dot{x}(t) < 0$.
For any $t_1 > T$ we have

$$x(t_1) - x(T) = \int_{u=T}^{t_1} \dot{x}(t)dt \tag{55}$$

Since $x(t_1)$ converges to 0, we deduce that for any $\varepsilon' > 0$, there exists $t_2 > T$ such that $|t_2\dot{x}(t_2)| \leqslant \varepsilon'$. Hence for any $\varepsilon > 0$, there exists $t_2 > T$ such that $\frac{1}{2}|\lambda x(t_2) + t\dot{x}(t_2)|^2 + \frac{\xi}{2}|x(t_2)|^2 \leqslant \varepsilon$ and we can conclude since we are back to case 1. This concludes the proof of the Proposition. □

References

[1] V. Apidopoulos, J-F. Aujol, and C. Dossal. On a second order differential inclusion modeling the FISTA algorithm. *Hal Preprint*, (01517708), 2017.

[2] H. Attouch, Z. Chbani, and H. Riahi. Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$". *Archiv Preprint*, June 2017.

[3] J-F. Aujol and C. Dossal. Optimal rate of convergence of an ode associated to the fast gradient descent schemes for $b > 0$. *Hal Preprint*, June 2017.

[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] Haim Brezis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, volume 5. Elsevier, 1973.

[6] 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.

[7] 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.