Convergence of FISTA

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft : December 21, 2019
Last update : January 21, 2020

Problem set up

$$(\mathcal{P}) \ : \ \min_{\mathbf{x}} F(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$$

where

- $f : \mathbb{R}^n \to \mathbb{R}$ smooth convex
- $g : \mathbb{R}^n \to \mathbb{R}$ convex (possibly non-smooth)
- $(\mathcal{P})$ is assumed solvable : the solution set $\mathcal{X}^* = \arg\min_{\mathbf{x}} F(\mathbf{x}) \neq \varnothing$
- Notation : $\mathbf{x}^* \in \mathcal{X}^*$ and $F^* = F(\mathbf{x}^*)$

See here for details on solving this problem using ISTA.

This document : convergence of FISTA (an accelerated ISTA)
Note : the proofs are just purely algebraic tricks, nothing special and not inspiring

FISTA

FISTA solve ($\mathcal{P}$) by iterating the following 3 steps

$$\mathbf{x}_k = p_L(\mathbf{y}_k) \tag{1}$$

$$t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2} \tag{2}$$

$$\mathbf{y}_{k+1} = \mathbf{x}_k + \frac{t_k - 1}{t_{k+1}}(\mathbf{x}_k - \mathbf{x}_{k-1}) \tag{3}$$

Remarks

- Initialization : $\mathbf{x}_0 \in \mathbb{R}^n, \mathbf{y}_1 = \mathbf{x}_0, t_1 = 1, L \geq L_f$ (assume known)
- FISTA is a kind of Nesterov-type acceleration applied on ISTA
- FISTA belongs to the accelerated proximal gradient algorithm

Pre-requisite

- Local quadratic over-estimator

$$Q(\mathbf{x};\mathbf{y}) = f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle + \frac{L}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + g(\mathbf{x}),\ L \geq L_f$$

- Update operator

$$p_L(\mathbf{y}) = \arg\min_{\mathbf{x}} \left\{ \frac{L}{2} \left\| \mathbf{x} - \left(\mathbf{y} - \frac{1}{L}\nabla f(\mathbf{y}) \right) \right\|_2^2 + g(\mathbf{x}) \right\},\ L \geq L_f$$

- Lemma associated to the update operator
  For all $\mathbf{y}$, $L \geq L_f > 0$ s.t. $F(p_L(\mathbf{y})) \leq Q(p_L(\mathbf{y});\mathbf{y})$, then for all $\mathbf{x}$

$$F(\mathbf{x}) - F(p_L(\mathbf{y})) \geq \frac{L}{2}\|p_L(\mathbf{y}) - \mathbf{y}\|_2^2 + L\langle \mathbf{y}-\mathbf{x}, p_L(\mathbf{y}) - \mathbf{y}\rangle. \quad (4)$$

See here for details and proof.

Special Pythagoras Theorem

- A key theorem to be used in the analysis
- Scalar case

$$(b-a)^2 + 2(a-c)(b-a) = (b-c)^2 - (a-c)^2$$

- Vector case

$$\|\mathbf{b}-\mathbf{a}\|_2^2 + 2\langle \mathbf{a}-\mathbf{c}, \mathbf{b}-\mathbf{a}\rangle = \|\mathbf{b}-\mathbf{c}\|_2^2 - \|\mathbf{a}-\mathbf{c}\|_2^2 \tag{5}$$

Key lemma of FISTA

The sequence $\{\mathbf{x}_k, \mathbf{y}_k\}$ produced by FISTA with constant stepsize satisfies

$$\frac{2}{L}(t_k^2 v_k - t_{k+1}^2 v_{k+1}) \geq \|\mathbf{u}_{k+1}\|_2^2 - \|\mathbf{u}_k\|_2^2, \ \forall k$$

where $v_k := F(\mathbf{x}_k) - F^*$, $\mathbf{u}_k := t_k \mathbf{x}_k - (t_k - 1)\mathbf{x}_{k-1} - \mathbf{x}^*$

Proof. Put $\mathbf{x} = \mathbf{x}_k, \mathbf{y} = \mathbf{y}_{k+1}$ (and $p_L(\mathbf{y}_{k+1}) \overset{(1)}{=} \mathbf{x}_{k+1}$) in lemma (4) gives

$$F(\mathbf{x}_k) - F(\mathbf{x}_{k+1}) \geq \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{y}_{k+1}\|_2^2 + L\langle \mathbf{y}_{k+1} - \mathbf{x}_k, \mathbf{x}_{k+1} - \mathbf{y}_{k+1}\rangle. \tag{6}$$

Put $\mathbf{x} = \mathbf{x}^*, \mathbf{y} = \mathbf{y}_{k+1}$ (and $p_L(\mathbf{y}_{k+1}) \overset{(1)}{=} \mathbf{x}_{k+1}$) in lemma (4) gives

$$F^* - F(\mathbf{x}_{k+1}) \geq \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{y}_{k+1}\|_2^2 + L\langle \mathbf{y}_{k+1} - \mathbf{x}^*, \mathbf{x}_{k+1} - \mathbf{y}_{k+1}\rangle. \tag{7}$$

Key lemma of FISTA

Re-arrange (6) and (7)

$$
\begin{aligned}
\frac{2}{L}\Big(F(\mathbf{x}_k) - F(\mathbf{x}_{k+1})\Big) &\geq \|\mathbf{x}_{k+1} - \mathbf{y}_{k+1}\|_2^2 + 2\langle \mathbf{y}_{k+1} - \mathbf{x}_k, \mathbf{x}_{k+1} - \mathbf{y}_{k+1}\rangle. \\
\frac{2}{L}\Big(F^* - F(\mathbf{x}_{k+1})\Big) &\geq \|\mathbf{x}_{k+1} - \mathbf{y}_{k+1}\|_2^2 + 2\langle \mathbf{y}_{k+1} - \mathbf{x}^*, \mathbf{x}_{k+1} - \mathbf{y}_{k+1}\rangle.
\end{aligned}
$$

By $v_k = F(\mathbf{x}_k) - F^*$

$$
\frac{2}{L}(v_k - v_{k+1}) \geq \|\mathbf{x}_{k+1} - \mathbf{y}_{k+1}\|_2^2 + 2\langle \mathbf{y}_{k+1} - \mathbf{x}_k, \mathbf{x}_{k+1} - \mathbf{y}_{k+1}\rangle. \quad (8)
$$

$$
-\frac{2}{L}v_{k+1} \geq \|\mathbf{x}_{k+1} - \mathbf{y}_{k+1}\|_2^2 + 2\langle \mathbf{y}_{k+1} - \mathbf{x}^*, \mathbf{x}_{k+1} - \mathbf{y}_{k+1}\rangle. \quad (9)
$$

Tricky step : $(t_{k+1} - 1)$(8) + (9) gives

$$
\frac{2}{L}\Big((t_{k+1} - 1)v_k - t_{k+1}v_{k+1}\Big) \geq t_{k+1}\|\mathbf{x}_{k+1} - \mathbf{y}_{k+1}\|_2^2 + 2\langle t_{k+1}\mathbf{y}_{k+1} - \mathbf{x}^* - (t_{k+1} - 1)\mathbf{x}_k, \mathbf{x}_{k+1} - \mathbf{y}_{k+1}\rangle
$$

Key lemma of FISTA

Tricky step 1 : multiply t_{k+1} to previous inequality, and apply $t_k^2 \overset{(2)}{=} t_{k+1}^2 - t_{k+1}$

$$\frac{2}{L}\Big(t_k^2 v_k - t_{k+1}^2 v_{k+1}\Big) \geq \|t_{k+1}\mathbf{x}_{k+1} - t_{k+1}\mathbf{y}_{k+1}\|_2^2 + 2\langle t_{k+1}\mathbf{y}_{k+1} - \mathbf{x}^* - (t_{k+1}-1)\mathbf{x}_k, t_{k+1}\mathbf{x}_{k+1} - t_{k+1}\mathbf{y}_{k+1}\rangle$$

Tricky step 2 : let $\mathbf{a} = t_{k+1}\mathbf{y}_{k+1}$, $\mathbf{b} = t_{k+1}\mathbf{x}_{k+1}$, $\mathbf{c} = (t_{k+1}-1)\mathbf{x}_k + \mathbf{x}^*$

$$\frac{2}{L}\Big(t_k^2 v_k - t_{k+1}^2 v_{k+1}\Big) \geq \|\mathbf{b}-\mathbf{a}\|_2^2 + 2\langle \mathbf{a}-\mathbf{c}, \mathbf{b}-\mathbf{a}\rangle \overset{(5)}{=} \|\mathbf{b}-\mathbf{c}\|_2^2 - \|\mathbf{a}-\mathbf{c}\|_2^2.$$

By (3) and by definition $\mathbf{u}_k = t_k\mathbf{x}_k - (t_k-1)\mathbf{x}_{k-1} - \mathbf{x}^*$

$$\begin{aligned} \mathbf{a} = t_{k+1}\mathbf{y}_{k+1} &\overset{(3)}{=} t_{k+1}\mathbf{x}_k + (t_k-1)(\mathbf{x}_k - \mathbf{x}_{k-1}) \\ \mathbf{u}_k &= t_k\mathbf{x}_k - (t_k-1)\mathbf{x}_{k-1} - \mathbf{x}^* \end{aligned}$$

we have

$$\frac{2}{L}\Big(t_k^2 v_k - t_{k+1}^2 v_{k+1}\Big) \geq \underbrace{\|t_{k+1}\mathbf{x}_{k+1} - (t_{k+1}-1)\mathbf{x}_k - \mathbf{x}^*\|_2^2}_{\|\mathbf{u}_{k+1}\|_2^2} - \underbrace{\|t_{k+1}\mathbf{x}_k + (t_k-1)(\mathbf{x}_k - \mathbf{x}_{k-1}) - (t_{k+1}-1)\mathbf{x}_k - \mathbf{x}^*\|_2^2}_{\|\mathbf{u}_k\|_2^2} \qquad \square$$

Convergence rate of FISTA

Theorem. Convergence of FISTA

The sequence $\{\mathbf{x}, \mathbf{y}_k\}$ generated by FISTA with constant stepsize fulfill

$$F(\mathbf{x}_k) - F^* \leq \frac{2LR_0^2}{(k+1)^2} \ \forall k$$

where $R_0 = \|\mathbf{x}_0 - \mathbf{x}^*\|_2$

To prove this, we need the key lemma and

- Lemma 1. Let $\{a_k, b_k\}$ be positive sequence that $a_1 + b_1 \leq c, c > 0$ and $a_k - a_{k+1} \geq b_{k+1} - b_k \ \forall k$. Then $a_k \leq c$ for all k
- Lemma 2. t_k generated by FISTA with $t_1 = 1$ satisfies $t_k \geq \dfrac{k+1}{2}$ for all k.

Note that by the definition of v_k, the theorem can be expressed as

$$v_k \leq \frac{2LR_0^2}{(k+1)^2}$$

Convergence rate of FISTA

Proof. Key lemma :

$$\frac{2}{L}(t_k^2 v_k - t_{k+1}^2 v_{k+1}) \geq \|\mathbf{u}_{k+1}\|_2^2 - \|\mathbf{u}_k\|_2^2, \ \forall k.$$

Let $a_k := \frac{2}{L} t_k^2 v_k$, $b_k := \|\mathbf{u}_k\|_2^2, c := \|\mathbf{y}_1 - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 = R_0^2$, then

$$a_k - a_{k+1} \leq b_{k+1} - b_k,$$

by Lemma 1, assuming $a_1 + b_1 \leq c$ then $a_k = \frac{2}{L} t_k^2 v_k \leq R_0^2 = c$:

$$F(\mathbf{x}_k) - F^* = v_k \leq \frac{LR_0^2}{2t_k^2}$$

Apply lemma 2 $t_k \geq \frac{k+1}{2} \implies \frac{1}{t_k} \leq \frac{2}{k+1} \implies \frac{1}{t_k^2} \leq \frac{4}{(k+1)^2}$

$$F(\mathbf{x}_k) - F^* \leq \frac{2LR_0^2}{(k+1)^2}. \quad \square$$

The proof is really this short. However it is really not easy to come up with the lemmas and the structure of the proof.

To show $a_1 + b_1 \le c$

$$\begin{aligned}
a_1 &= \frac{2}{L} t_1^2 v_1 = \frac{2}{L} v_1, \\
b_1 &= \|\mathbf{u}_1\|_2^2 = \|\mathbf{x}_1 - \mathbf{x}^*\|_2^2, \\
c &= \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 = R_0^2
\end{aligned}$$

Apply Lemma (4) with $\mathbf{x} = \mathbf{x}^*, \mathbf{y} = \mathbf{y}_1, p_L(\mathbf{y}_1) = \mathbf{x}_1$

$$F^* - F(\mathbf{x}_1) \ge \frac{L}{2}\|\mathbf{x}_1 - \mathbf{y}_1\|_2^2 + L\langle \mathbf{y}_1 - \mathbf{x}^*, \mathbf{x}_1 - \mathbf{y}_1 \rangle$$

$\iff$

$$\begin{aligned}
\frac{2}{L} v_1 &\ge \|\mathbf{x}_1 - \mathbf{y}_1\|_2^2 + 2\langle \mathbf{y}_1 - \mathbf{x}^*, \mathbf{x}_1 - \mathbf{y}_1 \rangle \\
&\overset{(5)}{=} \|\mathbf{y}_1 - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_1 - \mathbf{x}^*\|_2^2 \\
&= R_0^2 - \|\mathbf{x}_1 - \mathbf{x}^*\|_2^2
\end{aligned}$$

Hence $a_1 + b_1 \le c$.