Convergence of FISTA (Beck & Teboulle 2009)

Andersen Ang

ECS, Uni. Southampton, UK

andersen.ang@soton.ac.uk

Homepage angms.science

Version: July 17, 2023
First draft: Dece 21, 2019

Content

Reference: Amir Beck and Marc Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems." SIAM journal on imaging sciences, 2009.

Setup

Problem setup

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

- $f : \mathbb{R}^n \to \mathbb{R}$ smooth convex
  - f is smooth
    - f is continuously differentiable $\quad f \in \mathcal{C}^1$
      $\iff \nabla f(\boldsymbol{x})$ exists for all $\boldsymbol{x}$
    - ∇f is globally L_f-Lipschitz
      $L_f > 0, \; \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| \leq L_f \|\boldsymbol{x} - \boldsymbol{y}\|$
  - f is convex $\quad f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0) + \left\langle \nabla f(\boldsymbol{x}_0), \boldsymbol{x} - \boldsymbol{x}_0 \right\rangle$
- $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex and possibly non-smooth
- $(\mathcal{P})$ is assumed solvable $\mathcal{X}^* := \underset{\boldsymbol{x}}{\operatorname{argmin}} \; F(\boldsymbol{x}) \neq \varnothing$
- Notation: $\boldsymbol{x}^* \in \mathcal{X}^*$ and $F^* = F(\boldsymbol{x}^*)$
- See here for details on solving this problem using proximal gradient method.
- This document: convergence of FISTA (an accelerated ISTA)
- The proof is purely algebraic tricks, not special, not inspiring

Algorithm 1: FISTA

Initialize $\boldsymbol{y}_1 = \boldsymbol{x}_0 \in \mathbb{R}^n$
Initialize stepsize $t_1 = 1$
while *not converge* **do**

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

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

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

- Assume known $L \geq L_f$
- FISTA $=$ Nesterov's acceleration applied on ISTA
- FISTA $\in$ accelerated proximal gradient algorithms

Prerequisite

$$(\mathcal{P}) \ : \ \min_{\boldsymbol{x}} F(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x}), \ \ f \text{ is convex, } L_f\text{-smooth}, g \text{ is convex}$$

- **Local quadratic over-estimator**: f is L_f-smooth, so for any $L \geq L_f$,

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

- Let $L \geq L_f$, the **proximal gradient update operator on** g, denoted as $p_L : \mathbb{R}^n \to \mathbb{R}^n$

$$p_L(\boldsymbol{y}) \ = \ \operatorname*{argmin}_{\boldsymbol{x}} \left\{ \frac{L}{2}\left\| \boldsymbol{x} - \left(\boldsymbol{y} - \frac{1}{L}\nabla f(\boldsymbol{y}) \right) \right\|_2^2 + g(\boldsymbol{x}) \right\},$$

  - **A Lemma associated to the update operator** $\forall \boldsymbol{y}, L \geq L_f > 0 \,\text{s.t.}\, F(p_L(\boldsymbol{y})) \leq Q(p_L(\boldsymbol{y});\boldsymbol{y})$, then $\forall \boldsymbol{x}$

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

  See here for details and proof.

- **Special Pythagoras Theorem** A key theorem in the analysis

Scalar case $\quad (b-a)^2 + 2(a-c)(b-a) = (b-c)^2 - (a-c)^2$

Vector case $\quad \|\boldsymbol{b} - \boldsymbol{a}\|_2^2 + 2\langle \boldsymbol{a} - \boldsymbol{c}, \boldsymbol{b} - \boldsymbol{a} \rangle = \|\boldsymbol{b} - \boldsymbol{c}\|_2^2 - \|\boldsymbol{a} - \boldsymbol{c}\|_2^2 \quad (5)$

- Familiar with the proof of Nesterov's Accelerated Gradient Descent

A key lemma

$$\begin{array}{rcll} \boldsymbol{x}_k & = & p_L(\boldsymbol{y}_k) & (1) \\ F(\boldsymbol{x}) - F(p_L(\boldsymbol{y})) & \geq & \frac{L}{2}\|p_L(\boldsymbol{y}) - \boldsymbol{y}\|_2^2 + L\langle \boldsymbol{y} - \boldsymbol{x}, p_L(\boldsymbol{y}) - \boldsymbol{y}\rangle & (4) \end{array}$$

Lemma For the sequence $\{\boldsymbol{x}_k, \boldsymbol{y}_k\}_{k\in\mathbb{N}}$ produced by FISTA with constant stepsize $\frac{1}{L}$ with $L \geq L_f$, then

$$\frac{2}{L}(t_k^2\delta_k - t_{k+1}^2\delta_{k+1}) \quad \geq \quad \|\boldsymbol{u}_{k+1}\|_2^2 - \|\boldsymbol{u}_k\|_2^2, \ \forall k$$

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

Proof

$$F(\boldsymbol{x}_k) - F\big(p_L(\boldsymbol{y}_{k+1})\big) \geq \frac{L}{2}\|p_L(\boldsymbol{y}_{k+1}) - \boldsymbol{y}_{k+1}\|_2^2 + L\big\langle \boldsymbol{y}_{k+1} - \boldsymbol{x}_k,\ p_L(\boldsymbol{y}_{k+1}) - \boldsymbol{y}_{k+1}\big\rangle \qquad \text{let } \boldsymbol{x} = \boldsymbol{x}_k, \boldsymbol{y} = \boldsymbol{y}_{k+1} \text{ in (4)}$$

$$F(\boldsymbol{x}_k) - F(\boldsymbol{x}_{k+1}) \overset{(1)}{\geq} \frac{L}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\|_2^2 + L\big\langle \boldsymbol{y}_{k+1} - \boldsymbol{x}_k,\ \boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\big\rangle$$

$$\delta_k - \delta_{k+1} \overset{\delta}{\geq} \frac{L}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\|_2^2 + L\big\langle \boldsymbol{y}_{k+1} - \boldsymbol{x}_k,\ \boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\big\rangle \qquad (6)$$

$$F^* - F\big(p_L(\boldsymbol{y}_{k+1})\big) \geq \frac{L}{2}\|p_L(\boldsymbol{y}_{k+1}) - \boldsymbol{y}_{k+1}\|_2^2 + L\big\langle \boldsymbol{y}_{k+1} - \boldsymbol{x}^*,\ p_L(\boldsymbol{y}_{k+1}) - \boldsymbol{y}_{k+1}\big\rangle \qquad \text{let } \boldsymbol{x} = \boldsymbol{x}^*, \boldsymbol{y} = \boldsymbol{y}_{k+1} \text{ in (4)}$$

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

$$-\delta_{k+1} \overset{\delta}{\geq} \frac{L}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\|_2^2 + L\langle \boldsymbol{y}_{k+1} - \boldsymbol{x}^*, \boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\rangle. \qquad (7)$$

$$\boxed{\begin{aligned} \delta_k - \delta_{k+1} &\geq \frac{L}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\|_2^2 + L\big\langle \boldsymbol{y}_{k+1} - \boldsymbol{x}_k ,\ \boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\big\rangle \quad (6) \\ -\delta_{k+1} &\geq \frac{L}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\|_2^2 + L\langle \boldsymbol{y}_{k+1} - \boldsymbol{x}^*, \boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\rangle \quad (7) \end{aligned}}$$

Re-arrange (6) and (7)

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

$$-\frac{2}{L}\delta_{k+1} \geq \|\boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\|_2^2 + 2\big\langle \boldsymbol{y}_{k+1} - \boldsymbol{x}^*, \boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\big\rangle. \tag{7}$$

Tricky step 1: $(t_{k+1} - 1)(6) + (7)$ gives

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

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

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

Tricky step 3: let $\boldsymbol{a} = t_{k+1}\boldsymbol{y}_{k+1}$, $\boldsymbol{b} = t_{k+1}\boldsymbol{x}_{k+1}$, $\boldsymbol{c} = (t_{k+1} - 1)\boldsymbol{x}_k + \boldsymbol{x}^*$

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

By (3) $\boldsymbol{a} = t_{k+1}\boldsymbol{y}_{k+1} \overset{(3)}{=} t_{k+1}\boldsymbol{x}_k + (t_k - 1)(\boldsymbol{x}_k - \boldsymbol{x}_{k-1})$, and by definition $\boldsymbol{u}_k = t_k\boldsymbol{x}_k - (t_k - 1)\boldsymbol{x}_{k-1} - \boldsymbol{x}^*$ we have

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

Theorem. Convergence of FISTA

The sequence $\{\boldsymbol{x}, \boldsymbol{y}_k\}_{k\in\mathbb{N}}$ generated by FISTA with constant stepsize $\frac{1}{L}$ with $L > L_f$ fulfills

$$F(\boldsymbol{x}_k) - F^* \leq \frac{2L\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2}{(k+1)^2} \quad \forall k.$$

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 δ_k, the theorem can be expressed as

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

and this give the "hint" that we should start from the key lemma.

- **Proof** By the key lemma:

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

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

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

by Lemma 1, assuming $a_1 + b_1 \leq c$ then
$a_k = \frac{2}{L}t_k^2\delta_k \leq R_0^2 = c$. Inequality $\frac{2}{L}t_k^2\delta_k \leq R_0^2$ gives

$$F(\boldsymbol{x}_k) - F^* =: \delta_k \leq \frac{LR_0^2}{2t_k^2}.$$

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

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

- The last thing to do is to show $a_1 + b_1 \leq c$

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

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

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

$$\iff$$

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

Hence $a_1 + b_1 \leq c$.

End of document