# Local Linear Convergence of FISTA for Sparse Optimization

Patrick R. Johnstone and Pierre Moulin
Department of ECE, UIUC, Urbana, Illinois 61820, USA.

## I. Introduction

### A. Problem Formulation

Sparse optimization is an important problem defined as

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ F(x) \triangleq f(x) + \rho\|x\|_1, \tag{1}$$

where $\rho \geq 0$, $\|x\|_1 = \sum |x_i|$, and $f$ is a smooth (twice differentiable in our analysis) convex function with $L$-Lipschitz continuous gradient. Problem (1) is of central importance in compressed sensing.

### B. Fast Iterative Soft Thresholding Algorithm

The Fast Iterative Soft Thresholding Algorithm (FISTA), introduced by Beck and Teboulle [1], is a well-known approach to solving problems where the objective function is the sum of a smooth and non-smooth term. Throughout we will refer to the following algorithm as FISTA:

$$\left| \begin{array}{ll} y^{k+1} = & x^k + \alpha_k(x^k - x^{k-1}) \\ x^{k+1} = & \text{prox}_{\lambda_k g}(y^{k+1} - \lambda_k \nabla f(y^{k+1})) \end{array} \right.$$

starting with arbitrary $x^0, x^1 \in \mathbb{R}^n$. This is the same as Beck and Teboulle's framework but we allow for arbitrary sequences for the parameters.

### C. Local Linear Convergence

FISTA is based on the well-known Iterative Soft Thresholding Algorithm (ISTA) for Problem (1). It has been observed that ISTA exhibits *local linear convergence* [2], which means that after some finite number of iterations ISTA identifies a manifold on which the solution lies, and thereafter convergence is linear. Unlike ISTA, it is not known whether FISTA obtains local linear convergence for Problem (1).

### D. Contributions

We show that FISTA obtains local linear convergence for Problem (1). Specifically we show that after a finite number of iterations, FISTA reduces to minimizing a local function on a reduced support subject to an orthant constraint. We provide explicit bounds on the number of iterations for this to occur thus generalizing the analysis of ISTA by Hale et al. [2].

## II. Optimality Conditions

Examination of the optimality conditions of Problem (1) reveals the following useful theorem. Let $X^*$ be the solution set.

**Theorem 1 ([2]).** *For problem (1), $x^* \in X^*$ if and only if $\nabla f(x^*) = h^*$ where for all $i$, $h^*$ satisfies*

$$\frac{h_i^*}{\rho} \begin{cases} = -1 \text{ if } \exists\, x \in X^* : x_i > 0 \\ = +1 \text{ if } \exists\, x \in X^* : x_i < 0 \\ \in [-1, 1] \quad \text{else.} \end{cases}$$

*Furthermore $\nabla f(x') = \nabla f(x^*) \triangleq h^*$ for all $x', x^* \in X^*$.*

The following two sets will be crucial to our analysis. Let $D \triangleq \{i : |h_i^*| < \rho\}$ and $E \triangleq \{i : |h_i^*| = \rho\}$. Note that $D \cap E = \emptyset$ and $D \cup E = \{1, 2, \ldots, n\}$. By Theorem 1, we can infer that $\text{supp}(x^*) \subseteq E$ for all $x^* \in X^*$. Finally, define $\omega$ to be the following useful quantity: $\omega \triangleq \min\{\rho - |h_i^*| : i \in D\} > 0$.

## III. Results

The following theorem proves finite convergence to 0 for the components in $D$ (i.e. in a finite number of iterations), and finite convergence to the correct sign for the components in $E$. The number of iterations of this "manifold identification period" can be explicitly bounded in terms of the salient parameters and variables of the problem. Let $\nu_k = \rho\lambda_k$. The full proof of Theorem 2 is available at [3].

**Theorem 2.** *Assume $\{\lambda_k\}$ is nondecreasing and satisfies $0 < \lambda_k \leq 1/L$, and there exist $\underline{\alpha}, \overline{\alpha} \in [0, 1)$ such that $\{\alpha_k\}$ satisfies $\underline{\alpha} \leq \alpha_k \leq \overline{\alpha}$ for all $k$. Then, there exist constants $K_D > 0$ and $K_E > 0$ such that, for all $k > K_E$ the iterates of FISTA applied to Problem (1) satisfy*

$$sgn\left(y_i^{k+1} - \lambda_k \nabla f(y^{k+1})_i\right) \quad = \quad -\frac{h_i^*}{\rho},\ \forall i \in E,$$

*and, for all $k > K_D$*

$$x_i^k = y_i^k = 0,\ \forall i \in D.$$

*Furthermore, for any $x^* \in X^*$, $K_E$ does not exceed*

$$\frac{1}{\nu_1^2}\left[\frac{2\overline{\alpha}(1+\overline{\alpha})(F(x^1) - F^*)}{\underline{\alpha}(1-\overline{\alpha})L^2} + \|x^1 - x^*\|^2 - \overline{\alpha}\|x^0 - x^*\|^2\right] + \frac{\underline{\alpha}}{1-\underline{\alpha}}$$

*and $K_D$ does not exceed*

$$\frac{1}{\omega^2\nu_1^2}\left[\frac{2\overline{\alpha}(1+\overline{\alpha})(F(x^1) - F^*)}{\underline{\alpha}(1-\overline{\alpha})L^2} + \|x^1 - x^*\|^2 - \overline{\alpha}\|x^0 - x^*\|^2\right] + \frac{\underline{\alpha}}{1-\underline{\alpha}} + 2.$$

**Remark** We can recover the result by Hale et al. for ISTA (Theorem 4.5 [2]) by setting $\overline{\alpha}$ and $\underline{\alpha}$ to 0. The theorem applies to Beck and Teboulle's parameter choice if one replaces $\alpha_k$ with $\min(\alpha_k, \overline{\alpha})$, with $\overline{\alpha}$ chosen in $[0, 1)$.

Further work in the spirit of [2] allows us to show that after the manifold identification period FISTA reduces to minimizing $f(x_E) + \rho\|x_E\|_1$, where $x_E$ is equal to the vector $x$ on $E$ but 0 elsewhere. This allows us to deduce a linear rate of convergence so long as either this local function is strongly convex or a strong complementarity condition holds.

## References

[1] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Img. Sci.*, vol. 2, no. 1, pp. 183–202, Mar. 2009.

[2] Elaine T. Hale, Wotao Yin, and Yin Zhang, "Fixed-Point Continuation for L-1 Minimization: Methodology and Convergence," *SIAM J. on Optimization*, vol. 19, no. 3, pp. 1107–1130, Oct. 2008.

[3] Patrick R Johnstone and Pierre Moulin, "Lyapunov Analysis of FISTA With Local Linear Convergence for Sparse Opimization," *Preprint, available at http://web.engr.illinois.edu/~prjohns2/pdfs/FistaLypv3.pdf*, January 2015.