Local Linear Convergence of Inertial Forward-Backward Splitting for Low Complexity Regularization
Jingwei Liang, Jalal M. Fadili, Gabriel Peyré

To cite this version:
Jingwei Liang, Jalal M. Fadili, Gabriel Peyré. Local Linear Convergence of Inertial Forward-Backward Splitting for Low Complexity Regularization. SPARS, 2015, Cambridge, France. hal-02456434

HAL Id: hal-02456434
https://hal-normandie-univ.archives-ouvertes.fr/hal-02456434
Submitted on 27 Jan 2020
Local Linear Convergence of Inertial Forward–Backward Splitting for Low Complexity Regularization

Jingwei Liang and Jalal M. Fadili
CNRS, GREYC, ENSICAEN, Université de Caen
Email: {Jingwei.Liang, Jalal.Fadili}@ensicaen.fr

Gabriel Peyré
CNRS, Ceremade, Université Paris-Dauphine
Email: Gabriel.Peyre@ceremade.dauphine.fr

Abstract—In this abstract, we consider the inertial Forward-Backward (iFB) splitting method and its special cases (Forward-Backward/ISTA and FISTA). Under the assumption that the non-empty convex part of the objective is partly smooth relative to an active smooth manifold, we show that iFB-type methods (i) identify the active manifold in finite time, then (ii) enter a local linear convergence regime that we characterize precisely. This gives a grounded and unified explanation to the typical behaviour of the obtained results are illustrated by concrete examples.

I. INTRODUCTION

Consider the following structured optimization problem

$$\min_{x \in \mathbb{R}^n} \left\{ \Phi(x) \triangleq F(x) + J(x) \right\},$$

(P)

where $J \in \Gamma_0(\mathbb{R}^n)$, the set of proper, lower semi-continuous and convex functions, $F$ is convex, $C^{1,1}(\mathbb{R}^n)$ with $\nabla F$ being $\beta$-Lipschitz continuous. We assume that $\text{Argmin} \Phi \neq \emptyset$.

In this paper, we consider a generic form of inertial Forward–Backward for solving (P) which reads,

$$y_k = x_k^k + \alpha_k (x_k^k - x_{k-1}^k), \quad x_k^k = \arg \min_{y_k} \left\{ \gamma_k \nabla F(y_k) \right\},$$

where $\alpha_k \in [0, \bar{\alpha}]$ and $b_k \in [0, \bar{b}]$, $(\bar{\alpha}, \bar{b}) \in [0, 1]^2$, and the step-size $0 < \gamma_k \leq \gamma_k \leq \gamma < \min(2\alpha\beta^{-2}, 2\beta^{-1})$, then it converges to a minimizer $x^*$ of (P).

Theorem II.2 (Finite activity identification).

Condition (II.1) can be viewed as a geometric generalization of the strict complementarity of non-linear programming, and is almost necessary for the finite identification of $M_{x^*}$ [3].

III. LOCAL LINEAR CONVERGENCE

We now turn to the local linear convergence of the iFB-type methods with partly smooth functions. For space limitations, we mainly focus on the case where $a_k = b_k$, and denote $d_{k+1} = \left( x_{k+1}^k - x_k^k \right)$.

Theorem III.1. We assume the conditions of Theorem II.2 hold. If moreover $F$ is $C^2$ near $x^*$ and there exists $\alpha \geq 0$ such that $P_{T_{x^*}}\nabla^2 F(x^*)P_{T_{x^*}} > 0$, then for all $k$ large enough, we have 1) $Q$-linear rate: if $0 < \gamma \leq \gamma_k \leq \gamma < \min(2\alpha\beta^{-2}, 2\beta^{-1})$, then given any $\rho > 0$, then $1 > \rho \geq \tilde{\rho}_k$, the iterates satisfy

$$\|x_{k+1} - x^*\|_2^2 \leq \|x_k - x^*\|_2^2 - \eta(k), \quad k \in [0, 1],$$

where $\eta = \max \left\{ \eta_1(x), \eta_2(x) \right\}$, $\eta_1(x) = 1 - 2\alpha\gamma + \beta^2\gamma^2$, $\eta_2(x) = \min(2\alpha\beta^{-2}, 2\beta^{-1})$.

2) $R$-linear rate: if $M_{x^*}$ is affine/linear, then

$$\|x_{k+1} - x^*\|_2^2 \leq \|x_k - x^*\|_2^2 - \rho_k \|d_k\|_2^2,$$

where $\rho_k \in [0, 1]$.

$$\rho_k = \left\{ \begin{array}{ll}
\min(2\alpha\beta^{-2}, 2\beta^{-1}), & \text{if } \eta_k \in [-1, 0) \cup \left[ \frac{4\alpha\beta}{1+4\alpha\beta}, 1 \right], \\
\eta_k \in [0, \frac{4\alpha\beta}{1+4\alpha\beta}], & \text{if } \eta_k \in \left[ \frac{4\alpha\beta}{1+4\alpha\beta}, 1 \right].
\end{array} \right.$$
Fig. 1: Local linear convergence of iFB-type methods in terms of $\|x^k - x^*\|$. The forward model of the problem of interests reads $y = Ax_0 + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \delta^2)$. (a) $\ell_1$-norm, $(m, n) = (48, 128)$, $x_0$ is 8-sparse; (b) $\ell_{1,2}$-norm, $(m, n) = (60, 128)$, $x_0$ has 3 non-zero blocks with block-size 4; (c) 1D TV semi-norm, $(m, n) = (48, 128)$, $\nabla x_0$ is 8-sparse; (d) Nuclear norm, $(m, n) = (1425, 2500)$, $x_0 \in \mathbb{R}^{50 \times 50}$ and $\text{rank}(x_0) = 5$. The red, black and blue lines are respectively the results of FB, FISTA [5] and iFB (with $a_k = b_k \equiv \sqrt{5} - 0.01$). All algorithms were tested with $\gamma_k \equiv 1/\|A\|^2$. The solid lines are the practical observed profiles and the dashed ones the theoretical predictions. The beginning of the dashed lines are the points when $x^k$ identifies the manifold $\mathcal{M}_{x^*}$. As one can observe, FISTA has the fastest manifold identification, however, locally it is the slowest for all tested examples. Indeed, when the manifold is affine, it can be shown from Theorem III.1 that $\rho_k \in [\eta_k, \sqrt{\eta_k}]$ for $a_k > \eta_k$, i.e. FISTA is locally slower than FB.

ACKNOWLEDGMENT

This work has been partly supported by the European Research Council (ERC project SIGMA-Vision). JF was partly supported by Institut Universitaire de France.

REFERENCES