



# Identification en temps fini et convergence linéaire locale de l'algorithme proximal implicite-explicite inertiel

Jingwei Liang, Jalal M. Fadili, Gabriel Peyré

## ► To cite this version:

Jingwei Liang, Jalal M. Fadili, Gabriel Peyré. Identification en temps fini et convergence linéaire locale de l'algorithme proximal implicite-explicite inertiel. GRETSI, 2015, Lyon, France. hal-02456431

**HAL Id: hal-02456431**

**<https://normandie-univ.hal.science/hal-02456431>**

Submitted on 27 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Identification en temps fini et convergence linéaire locale de l'algorithme proximal implicite-explicite inertiel

Jingwei LIANG<sup>1</sup>, Jalal M. FADILI<sup>1</sup>, Gabriel PEYRE<sup>2</sup>

<sup>1</sup>Normandie Université-ENSICAEN, GREYC UMR 6072, France

<sup>2</sup>CNRS et Ceremade UMR 7534, Université Paris-Dauphine, France

{Jingwei.Liang, Jalal.Fadili}@greyc.ensicaen.fr, Gabriel.Peyre@ceremade.dauphine.fr

**Résumé** – Nous considérons la classe des algorithmes proximaux implicites-explicites inertiels (iFB) pour minimiser la somme deux fonctions propres, semi-continues inférieurement et convexes, dont l'une est à gradient Lipschitz et l'autre est partiellement régulière relativement à une variété active  $\mathcal{M}$ . Cette classe inclut comme cas particuliers populaires les algorithmes FB et FISTA pour un choix approprié du paramètre inertiel dans ce dernier. Nous proposons une analyse unifiée et montrons que les itérées du schéma iFB, (i) identifient la variété active  $\mathcal{M}$  en temps fini, et ensuite (ii) entrent dans un régime de convergence linéaire locale dont nous caractérisons le taux de façon précise. Ces résultats donnent une justification théorique aux comportements typiques qui ont été observés numériquement pour plusieurs problèmes inclus dans notre cadre, comme par exemple le Lasso, le group-Lasso, la minimisation de la variation totale, ou celle de la norme nucléaire. Ces résultats peuvent avoir plusieurs applications en traitement du signal et des images ou en apprentissage statistique.

**Abstract** – We consider the class of inertial Forward-Backward (iFB) proximal splitting algorithms, to minimize the sum of two proper lower semi-continuous convex functions, one of which having a Lipschitz continuous gradient and the other being partly smooth relative to an active manifold  $\mathcal{M}$ . Special cases of this class include the FB and, for an appropriate choice of the inertial parameter, FISTA-like schemes. We propose a unified analysis, under which we show that iFB-type splitting, (i) correctly identifies the active manifold  $\mathcal{M}$  in a finite number of iterations, and then (ii) enters a local (linear) convergence regime, which is characterised precisely. This gives a grounded justification to the typical behaviour that has been observed numerically for many problems encompassed in our framework, including the Lasso, the group Lasso, total variation minimization and the nuclear norm regularization to name a few. These results may have numerous applications including in signal/image processing and machine learning.

## 1 Introduction

### 1.1 Formulation du problème

L'optimisation convexe est devenue omniprésente dans bon nombre de champs scientifiques. Notre objectif ici est la minimisation du problème composite structuré suivant

$$\min_{x \in \mathbb{R}^n} \{ \Phi(x) = F(x) + J(x) \}, \quad (\mathcal{P})$$

sous les hypothèses :

- (H.1)  $J \in \Gamma_0(\mathbb{R}^n)$ , la classe des fonctions propres, semi-continues inférieurement et convexes sur  $\mathbb{R}^n$  ;
- (H.2)  $F$  est convexe et  $C^{1,1}(\mathbb{R}^n)$  dont le gradient est  $\beta$ -Lipschitz continu ;
- (H.3)  $\text{Argmin } \Phi \neq \emptyset$ .

Le problème  $(\mathcal{P})$  couvre plusieurs cas particuliers populaires que l'on rencontre en traitement du signal et des images ou encore en apprentissage statistique. Par exemple, en prenant  $F = \frac{1}{2\lambda} \|y - A \cdot\|^2$ , pour  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  un opérateur linéaire et  $\lambda > 0$ , on retrouve le Lasso lorsque  $J = \|\cdot\|_1$ , le Lasso par blocs pour  $J = \|\cdot\|_{1,2}$ , la régularisation par la variation totale lorsque  $J = \|D_{\text{DIF}}^* \cdot\|_1$  où  $D_{\text{DIF}}$  est un opérateur de différences finies, l'anti-parcimonie quand  $J = \|\cdot\|_\infty$ , et la régularisation par la norme nucléaire pour  $J = \|\cdot\|_*$ .

Dans cet article, nous considérons une forme généralisée inertielle de l'algorithme FB (iFB). L'iFB implémente la ré-

cursion suivante,

$$\begin{cases} y_a^k &= x^k + a_k(x^k - x^{k-1}), \\ y_b^k &= x^k + b_k(x^k - x^{k-1}) \\ x^{k+1} &= \text{prox}_{\gamma_k J}(y_a^k - \gamma_k \nabla F(y_b^k)); \end{cases} \quad (1.1)$$

où les paramètres inertiels  $(a_k, b_k) \in [0, \bar{a}] \times [0, \bar{b}]$ ,  $\bar{a}$  et  $\bar{b} \in [0, 1]$  et la suite des pas de descente est telle que  $0 < \underline{\gamma} \leq \gamma_k \leq \bar{\gamma} < 2/\beta$ . On rappelle l'opérateur proximal de  $J$ , pour  $\gamma > 0$ ,

$$\text{prox}_{\gamma J}(x) = \underset{z \in \mathbb{R}^n}{\text{argmin}} \frac{1}{2\gamma} \|z - x\|^2 + J(z).$$

Le schéma iFB (1.1) couvre plusieurs cas particuliers dans la littérature, dépendamment du choix de  $a_k$  et  $b_k$  : pour  $a_k = b_k \equiv 0$ , (1.1) n'est autre que FB (non-relaxé) dont la convergence est connue depuis [13] ;  $a_k \in [0, \bar{a}]$ ,  $\bar{a} < 1$ ,  $b_k = 0$  correspond à la méthode étudiée dans [15] ;  $a_k \in [0, \bar{a}]$ ,  $\bar{a} < 1$ ,  $b_k = a_k$  est celle dans [14]. Si dans ce dernier en plus,  $a_k \rightarrow 1$  et  $\gamma_k \in [\underline{\gamma}, 1/\beta]$ , alors (1.1) devient la méthode FISTA [1]. Toutefois, dans cette version originale de FISTA, seule la convergence de l'objective au taux  $O(1/k^2)$  est garantie. Ceci est resté un problème ouvert jusqu'au travail récent de [3] où la convergence des itérées est montrée mais pour le choix modifié

$$t_k = \frac{k+p-1}{p}, \quad a_k = b_k = \frac{t_k-1}{t_k}, \forall p > 2. \quad (1.2)$$

### 1.2 Contributions

Dans ce travail, nous établissons tout d'abord des conditions suffisantes garantissant la convergence globale de (1.1). En-

suite, lorsque  $J$  est aussi partiellement régulière relativement à une variété  $\mathcal{M}$  (voir Définition 3.1), nous prouvons que l'iFB permet d'identifier en temps fini  $\mathcal{M}$ , c.à.d. pour un nombre d'itérations  $K$  suffisamment grand, les itérées  $x^k \in \mathcal{M}$ ,  $\forall k \geq K$ . En exploitant cette propriété, nous montrons que  $x^k$  converge localement linéairement. Nous analysons précisément ce régime linéaire local et le taux associé notamment en fonction de la nature de la variété active  $\mathcal{M}$ . En général, la convergence locale est  $Q$ -linéaire, et elle devient  $R$ -linéaire dès lors que  $\mathcal{M}$  est affine (ou linéaire). Plusieurs résultats expérimentaux sur des problèmes populaires en traitement du signal et des images viennent conforter ces résultats. Toutes les preuves des résultats pourront être trouvées dans la version longue [12].

### 1.3 Relation à l'état de l'art

L'identification du support la convergence  $R$ -linéaire locale de FB a été établie dans [2, 5] pour le Lasso. [16] a montré la convergence  $Q$ -linéaire locale de FB pour résoudre  $(\mathcal{P})$  lorsque  $F$  vérifie une condition de régularité et de forte convexité restreinte, et  $J$  est une norme dite décomposable. Dans [7, 8, 6], les auteurs ont montré l'identification en temps fini de la variété active associée à des fonctions partiellement régulières pour certains algorithmes, comme la descente de (sous-)gradient projetée, Newton, ou l'algorithme du point proximal. La question du régime de convergence locale y est restée un problème ouvert. Les auteurs de [11] ont été les premiers à montrer l'identification en temps fini et la convergence linéaire locale du FB pour résoudre  $(\mathcal{P})$  dans le cadre très général où  $J$  est partiellement régulière. En finalisant cet article, nous avons pris connaissance des travaux très récents et indépendants de [17, 9], qui ont respectivement étudié les propriétés de convergence locale de FISTA et iFB<sup>1</sup> (avec  $a_k = b_k$  et  $\bar{a} < 1$ ) pour résoudre le Lasso. Nos contributions ici vont bien plus loin que le simple Lasso, et étendent les résultats de [11] au cas inertielle.

### 1.4 Notations

Pour un convexe non vide  $\mathcal{C} \subset \mathbb{R}^n$ ,  $\text{par}(\mathcal{C})$  est le sous-espace qui lui est parallèle, et  $\text{ri}(\mathcal{C})$  son intérieur relatif.  $P_{\mathcal{C}}$  est le projecteur orthogonal sur  $\mathcal{C}$ . Pour  $V$  un sous-espace de  $\mathbb{R}^n$  et une matrice  $A \in \mathbb{R}^{m \times n}$ ,  $A_V = AP_V$ .

Soit  $\mathcal{M} \subset \mathbb{R}^n$  une variété  $C^2$  autour de  $x \in \mathbb{R}^n$ , on dénote  $\mathcal{T}_{\mathcal{M}}(x)$  l'espace tangent de  $\mathcal{M}$  en  $x$ . On appellera le modèle tangent en  $x$   $T_x = \text{par}(\partial J(x))^\perp$ , où  $\partial J(x)$  est la sous-différentielle de  $J$  en  $x$ . Il est clair que  $P_{T_x}(\partial J(x))$  est un singleton. Par analogie avec la norme  $\ell_1$ , on définit alors le vecteur "signe" généralisé  $e_x = P_{T_x}(\partial J(x))$ .

## 2 Convergence Globale de l'iFB

Notre résultat de convergence globale est le suivant.

1. Strictement parlant, la dénomination FISTA utilisée dans [9] n'est pas appropriée.

**Théorème 2.1.** *Supposons que (H.1)-(H.3) sont vérifiées, que  $0 < \underline{\gamma} \leq \gamma_k \leq \bar{\gamma} < 2/\beta$ ,  $\bar{a} < 1$  et*

$$\sum_{k \in \mathbb{N}} \max\{a_k, b_k\} (\|x^{k+1} - x^k\|^2 + \|x^k - x^{k-1}\|^2) < +\infty. \quad (2.1)$$

*Alors le schéma iFB (1.1) converge vers un minimiseur de  $(\mathcal{P})$ .*

*Soit  $\tau > 0$  tel que l'une des deux conditions suivantes soient vraies,*

- (i)  $a_k \in ]0, \bar{a}]$ ,  $b_k \in ]0, \bar{b}]$ , pour  $b_k \leq a_k$  ou  $\frac{\gamma_k \beta}{2} b_k \leq a_k < b_k$ ,  
 $(1 - 3a_k) - \frac{\gamma_k}{2\beta}(1 - b_k)^2 > \tau$ ;
- (ii)  $a_k \in [0, \bar{a}]$ ,  $b_k \in ]0, \bar{b}]$ , et  $a_k < \frac{\gamma_k \beta}{2} b_k$ ,  
 $(1 + a_k) - \frac{\gamma_k}{2\beta}(1 + b_k)^2 > \tau$ .

*Alors (2.1) est vraie.*

**Remarque 2.2.** Le Théorème 2.1 reste vrai dans tout espace de Hilbert réel. Par ailleurs, si  $\gamma_k \in [\underline{\gamma}, 1/\beta]$  et  $a_k = b_k$  est choisi selon la règle FISTA modifiée (1.2), [3] ont prouvé que la condition (2.1), est remplie. Il s'ensuit que les itérées de (1.1) convergent dans ce cas. Observons tout de même que ceci impose une borne supérieure sur  $\gamma_k$  deux fois plus petite.

## 3 Régularité Partielle

En plus de (H.1), notre hypothèse centrale dans la suite est que  $J$  est partiellement régulière. La régularité partielle des fonctions a été originellement définie dans [10]. Notre définition ci-après la spécialise pour les fonctions dans  $\Gamma_0(\mathbb{R}^n)$ .

**Définition 3.1.** Soit  $J \in \Gamma_0(\mathbb{R}^n)$ , et  $x \in \mathbb{R}^n$  tel que  $\partial J(x) \neq \emptyset$ .  $J$  est *partiellement régulière* en  $x$  relativement à l'ensemble  $\mathcal{M}$  contenant  $x$  si

- (1) (Régularité)  $\mathcal{M}$  est une variété  $C^2$  autour de  $x$  et la restriction de  $J$  à  $\mathcal{M}$  est  $C^2$  autour de  $x$ .
- (2) (Caractère piqué) L'espace tangent  $\mathcal{T}_{\mathcal{M}}(x)$  est  $T_x$ .
- (3) (Continuité) La sous-différentielle  $\partial J$  est continue en  $x$  relativement à  $\mathcal{M}$ .

On notera  $\text{PS}_x(\mathcal{M})$  la classe des fonctions partiellement régulières en  $x$  relativement à  $\mathcal{M}$ .

On peut montrer que la variété active  $\mathcal{M}$  est définie sans ambiguïté. De plus, sous hypothèses assez faibles de transversalité, l'ensemble des fonctions partiellement régulières est stable par addition et pré-composition par un opérateur linéaire [10]. Par ailleurs, les fonctions convexes, invariantes par permutation des valeurs singulières et partiellement régulières, c.à.d. fonctions spectrales, sont des fonctions convexes et partiellement régulières des matrices [4]. La classe des fonctions localement polyédrales est aussi partiellement régulière relativement à une variété affine (ou linéaire), pour laquelle le vecteur "signe" généralisé  $e_x$  est aussi localement constant [18]. On déduit de ces résultats que tous les exemples cités précédemment, comme les normes  $\ell_1$ ,  $\ell_{1,2}$ ,  $\ell_\infty$ , nucléaire, ou encore la semi-norme TV sont partiellement régulières.

## 4 Identification et Convergence Linéaire Locale

Dans cette section, nous énonçons notre principal résultat sur l'identification en temps fini et la convergence linéaire locale de l'iFB. Dans le reste de l'article, et pour des raisons de simplicité de l'exposé, nous nous concentrerons sur le cas  $a_k = b_k$ , bien que le cas général puisse être traité au prix d'arguments techniques supplémentaires. Nous posons  $d^{k+1} = \begin{pmatrix} x^{k+1} - x^* \\ x^k - x^* \end{pmatrix}$ .

**Théorème 4.1.** *Soit le schéma iFB (1.1) utilisé sous les hypothèses du Théorème 2.1 ou la Remarque 2.2, dont l'itérée  $x^k$  converge vers  $x^* \in \text{Argmin } \Phi$ , tel que  $J \in \text{PS}_{x^*}(\mathcal{M}_{x^*})$ ,  $F$  est  $C^2$  autour de  $x^*$  et*

$$-\nabla F(x^*) \in \text{ri}(\partial J(x^*)). \quad (4.1)$$

Alors les assertions suivantes sont vraies.

- (1) L'iFB (1.1) exhibe la propriété d'identification finie, i.e.  $\exists K \geq 0$ , tel que  $\forall k \geq K$ ,  $x^k \in \mathcal{M}_{x^*}$ . De plus, si  $\mathcal{M}_{x^*}$  est affine (ou linéaire), alors on a aussi  $y_a^k \in \mathcal{M}_{x^*}$  pour  $k \geq K$ .
- (2) Supposons par ailleurs  $\exists \alpha \geq 0$  tel que

$$P_T \nabla^2 F(x^*) P_T \succ \alpha \text{Id}, \quad (4.2)$$

où  $T := T_{x^*}$ . Alors,  $\forall k$  suffisamment grand,

- (i) Convergence  $Q$ -linéaire : si  $0 < \underline{\gamma} \leq \gamma_k \leq \bar{\gamma} < \min(2\alpha\beta^{-2}, 2\beta^{-1})$ , les itérées satisfont

$$\|x_{k+1} - x^*\|^2 \leq \|d^{k+1}\|^2 \leq \rho \|d^k\|^2,$$

pour tout  $\rho$  tel que  $1 > \rho > \tilde{\rho}_k$ , où

$$\tilde{\rho}_k = \begin{cases} \frac{(1+a_k)\eta + \sqrt{(1+a_k)^2\eta^2 - 4a_k\eta}}{2} & \eta \in [\frac{4a_k}{(1+a_k)^2}, 1], \\ \sqrt{a_k\eta} & \eta \in [0, \frac{4a_k}{(1+a_k)^2}], \end{cases}$$

$$\eta = \max\{q(\underline{\gamma}), q(\bar{\gamma})\} \in [0, 1], \text{ et } q(\gamma) = 1 - 2\alpha\gamma + \beta^2\gamma^2.$$

- (ii) Convergence  $R$ -linéaire : si  $\mathcal{M}_{x^*}$  est affine (ou linéaire),

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

où  $\rho_k \in [0, 1[$  et

$$\rho_k = \begin{cases} \frac{|(1+a_k)\eta_k| + \sqrt{(1+a_k)^2\eta_k^2 - 4a_k\eta_k}}{2}, & \eta_k \in ]-1, 0] \cup [\frac{4a_k}{(1+a_k)^2}, 1[, \\ \sqrt{a_k\eta_k}, & \eta_k \in [0, \frac{4a_k}{(1+a_k)^2}], \end{cases}$$

avec  $\eta_k \in ]-1, 1[$  une valeur propre de  $\text{Id} - \gamma_k P_T \int_0^1 \nabla^2 F(x^* + t(y_a^k - x^*)) dt P_T$ . Si  $\gamma_k \in [\underline{\gamma}, 1/\beta]$ , alors  $\eta_k \in [0, 1[$  est la plus grande valeur propre de cette matrice.

**Remarque 4.2.**(i) Le Théorème 4.1(i) garantit que  $x^k$  arrive sur la variété active en temps fini. Ainsi, le problème  $(\mathcal{P})$  est localement équivalent à sa version contrainte sur  $\mathcal{M}_{x^*}$ , qui par régularité partielle, devient un problème complètement lisse ( $C^2$ ). Ceci ouvre la voie à des méthodes d'accélération en utilisant l'information du second ordre. Par exemple, on peut se tourner vers des méthodes d'optimisation géométrique le long de la variété  $\mathcal{M}_{x^*}$  (e.g. Newton ou gradient conjugué non-linéaire), où une convergence bien plus rapide peut être garantie. Ces détails sont discutés dans la version longue [12].

- (ii) Pour  $a_k \equiv 0$ , on retrouve les résultats [11] pour FB. En particulier, on a  $\tilde{\rho}_k = \eta$  dans Théorème 4.1(2)(i), et  $\rho_k = \eta_k$  dans (ii).

- (iii) Bien que l'iFB puisse être plus globalement rapide que FB, le constat peut changer localement (voir Fig. 1). En effet, lorsque la variété est affine (ou linéaire), et avec le choix de  $a_k$  selon (1.2) et  $\gamma_k \in [\underline{\gamma}, 1/\beta]$ , on a  $\rho_k \in ]\eta_k, \sqrt{\eta_k}]$  dès lors que  $a_k > \eta_k$ . Ainsi, FISTA avec le pas (1.2) finira toujours par être plus lent que FB localement.

Lorsque  $J$  est localement polyédrale, on peut s'affranchir de l'hypothèse de forte convexité restreinte (4.2), mais au prix d'une estimation moins fine du taux de convergence linéaire.

**Théorème 4.3.** *Soit le schéma iFB (1.1) utilisé sous les hypothèses du Théorème 2.1 ou la Remarque 2.2, dont l'itérée  $x^k$  converge vers  $x^* \in \text{Argmin } \Phi$ , tel que  $J$  est localement polyédrale autour de  $x^*$ , (4.1) est vérifiée, et qu'il existe un sous-espace  $V$  tel que  $\text{Ker}(P_T \nabla^2 F(x) P_T) = V$  pour tout  $x$  autour de  $x^*$ . Alors il existe une constante  $\rho \in [0, 1[$  telle que pour tout  $k$  suffisamment grand*

$$\|x^k - x^*\| = O(\rho^k).$$

Un exemple typique de  $F$  où ce résultat s'applique est  $F = G \circ A$ , où  $G$  est localement fortement convexe, auquel cas  $V = \text{Ker}(A_T)$ .

## 5 Résultats expérimentaux

A titre d'illustration, on considère le problème suivant

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - Ax\|^2 + \lambda J(x) \quad (5.1)$$

où  $y \in \mathbb{R}^m$ ,  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , et  $\lambda$  est la paramètre de régularisation, et  $J$  est soit la norme  $\ell_1$ ,  $\ell_\infty$ ,  $\ell_{1,2}$ , la semi-norme TV ou la norme nucléaire.

**Exemple 5.1** ( $\ell_1$ ). La norme  $\ell_1$  est partiellement régulière en  $x$  relativement à  $\mathcal{M}_x = T_x = \{u \in \mathbb{R}^n : \text{supp}(u) \subseteq \text{supp}(x)\}$ , où  $\text{supp}(x) = \{i : x_i \neq 0\}$ . Son opérateur proximal est le seuillage doux.

**Exemple 5.2** ( $\ell_{1,2}$ ). La norme  $\ell_{1,2}$  n'est pas polyédrale, mais sa variété active en  $x$  est tout de même linéaire  $\mathcal{M}_x = T_x = \{u \in \mathbb{R}^n : \text{supp}_B(u) \subseteq \text{supp}_B(x)\}$ , où  $\text{supp}_B(x) = \bigcup \{b : x_b \neq 0\}$ , et  $\bigcup_{b \in B} b = \{1, \dots, n\}$ . Son opérateur proximal est un seuillage doux par bloc.

**Exemple 5.3** (TV anisotrope). La semi-norme TV  $\|x\|_{\text{TV}} = \|D_{\text{DIF}}^* x\|_1$  est partiellement régulièrement relativement à  $\mathcal{M}_x = T_x = \{u \in \mathbb{R}^n : \text{supp}(D_{\text{DIF}}^* u) \subseteq \text{supp}(D_{\text{DIF}}^* x)\}$ . L'opérateur proximal de TV 1D peut être obtenu en utilisant l'algorithme de la corde tendue, et celui de TV 2D anisotrope par coupure de graphe.

**Exemple 5.4** (Norme nucléaire). La norme nucléaire est partiellement régulière relativement à la variété  $\mathcal{M}_x = \{u \in \mathbb{R}^{n_1 \times n_2} : \text{rang}(u) = \text{rang}(x)\}$ . Son opérateur proximal correspond au seuillage doux des valeurs singulières.

**Restauration à partir de mesures aléatoires** Dans ces exemples, le modèle d'observation est

$$y = Ax_0 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \delta^2), \quad (5.2)$$

où  $A \in \mathbb{R}^{m \times n}$  est une matrice Gaussienne dont les entrées sont i.i.d.  $\mathcal{N}(0, 1)$ . Le protocole expérimental est le suivant :

- $\ell_1$  :  $(m, n) = (48, 128)$ ,  $x_0$  est 8-parcimonieux ;
- TV :  $(m, n) = 48, 128$ ,  $(D_{\text{DIF}} x_0)$  est 8-parcimonieux ;
- $\ell_\infty$  :  $(m, n) = (123, 128)$ ,  $x_0$  a 10 entrées qui saturent ;
- $\ell_{1,2}$  :  $(m, n) = (48, 128)$ ,  $x_0$  a 2 blocs non-nuls de taille 4 ;
- Norme nucléaire :  $(m, n) = (1425, 2500)$ ,  $x_0 \in \mathbb{R}^{50 \times 50}$ ,  $\text{rang}(x_0) = 5$ .

Le nombre de mesures est choisi suffisamment grand,  $\delta$  suffisamment petit et  $\lambda$  de l'ordre de  $\delta$  tels que [19, Theorem 1] s'applique, ce qui implique que le minimiseur de (5.1) est unique et vérifie les conditions de non-dégénérescence et de forte convexité restreinte (4.1)-(4.2). Pour FB, iFB et FISTA, nous avons fixé  $\gamma_k \equiv 1/\beta$ . Pour iFB, nous avons utilisé  $a_k = b_k = \sqrt{5} - 2.01$  de façon à ce que Théorème 2.1(i) soit vrai. Pour FISTA, nous avons utilisé la suite  $a_k$  de (1.2) avec  $p = 2.001$ . Les profils de convergence de  $\|x^k - x^*\|$  sont portés en Fig. 1(a)-(e). Pour  $\ell_1$ , TV et  $\ell_\infty$ , les profils locaux prédits par Théorème 4.1(2)(ii) coïncident avec ceux observés. La raison est que les régularisateurs en question sont polyédraux et que l'attaque aux données est quadratique, rendant l'estimation du taux optimale. Pour la norme  $\ell_{1,2}$ , bien que la variété active soit encore un sous-espace, le vecteur "signe" généralisé n'est pas localement constant, et ainsi le taux prédit est légèrement sur-estimé, même si ça reste difficilement visible sur les figures. Enfin, pour la norme nucléaire, la variété n'étant plus linéaire, Théorème 4.1(2)(i) s'applique, mais comme attendu, l'estimation du taux est moins fine.

**Déconvolution TV 2D** Dans cet exemple 2D,  $y$  est une image générée selon (5.1), avec  $A$  un opérateur de convolution circulaire par un noyau Gaussien. La régularisation TV anisotrope est utilisée, qui est polyédrale. Les profils de convergence sont montrés en Fig. 1(f). Les conditions (4.1)-(4.2) ont été vérifiées a posteriori. Les taux théoriques sont encore une fois finement estimés.

## Références

- [1] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Img. Sci.*, 2(1) :183–202, 2009.
- [2] K. Bredies and D. A. Lorenz. Linear convergence of iterative soft-thresholding. *J. of Fourier Analysis and App.*, 14(5-6) :813–837, 2008.
- [3] A. Chambolle and D. Dossal. How to make sure the iterates of FISTA converge. *Preprint*, 2014.
- [4] A. Daniilidis, D. Drusvyatskiy, and A. S. Lewis. Orthogonal invariance and identifiability. *to appear in SIAM J. Matrix Anal. Appl.*, 2014.
- [5] E. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for  $\ell_1$ -minimization : Methodology and convergence. *SIAM J. on Optim.*, 19(3) :1107–1130, 2008.

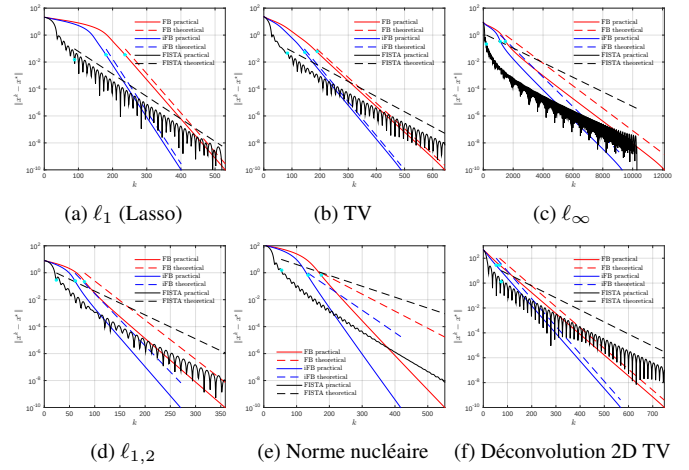


FIGURE 1 – Profils de convergence observés (ligne continue) et prédits (ligne discontinue) pour FB, iFB et FISTA en termes de  $\|x^k - x^*\|$  pour différentes régularisations. (a)  $\ell_1$  ; (b) TV ; (c)  $\ell_\infty$  ; (d)  $\ell_{1,2}$  ; (e) Norme nucléaire ; (f) Déconvolution par TV 2D anisotrope. Le point apparaissant sur chaque courbe montre l'itération à partir de laquelle  $\mathcal{M}_{x^*}$  a été identifié.

- [6] W. L. Hare. Identifying active manifolds in regularization problems. In H. H. Bauschke et al., editor, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer, 2011.
- [7] W. L. Hare and A. S. Lewis. Identifying active constraints via partial smoothness and prox-regularity. *J. of Conv. Ana.*, 11(2) :251–266, 2004.
- [8] W. L. Hare and A. S. Lewis. Identifying active manifolds. *Algorithmic Operations Research*, 2(2) :75–82, 2007.
- [9] P. R. Johnstone and P. Moulin. A lyapunov analysis of FISTA with local linear convergence for sparse optimization. Technical report, arXiv :1502.02281, Feb. 2015.
- [10] A. S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM J. on Optim.*, 13(3) :702–725, 2003.
- [11] J. Liang, M.J. Fadili, and G. Peyré. Local linear convergence of forward-backward under partial smoothness. *NIPS*, 27 :1970–1978, 2014.
- [12] J. Liang, M.J. Fadili, and G. Peyré. Activity identification and local linear convergence of forward-backward-type splitting. Technical report, 2015.
- [13] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. on Num. Analysis*, 16(6) :964–979, 1979.
- [14] D. A. Lorenz and T. Pock. An inertial forward-backward method for monotone inclusions. *J. of Math. Im. and Vis.*, 51(2) :311–325, 2015.
- [15] A. Moudafi and M. Oliny. Convergence of a splitting inertial proximal method for monotone operators. *J. of Comp. and App. Math.*, 155(2) :447–454, 2003.
- [16] S. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu. A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statistical Science*, 27(4) :538–557, December 2012.
- [17] S. Tao, D. Boley, and S. Zhang. Local linear of ISTA and FITSTA for the Lasso problem. Technical report, arXiv :1501.02888, Jan., 2015.
- [18] S. Vaïter, M. Golbabaee, J. M. Fadili, and G. Peyré. Model selection with low complexity priors. *Information and Inference : A Journal of the IMA (to appear)*, 2015.
- [19] S. Vaïter, G. Peyré, and J. M. Fadili. Partly smooth regularization of inverse problems. arXiv :1405.1004, 2014.