Multilevel FISTA for image restoration

Guillaume Lauga, Elisa Riccietti, Nelly Pustelnik, Paulo Gonçalves

MULTILEVEL FISTA FOR IMAGE RESTORATION

Guillaume Lauga[†], *Elisa Riccietti*[†], *Nelly Pustelnik*[⋆], *Paulo Gonçalves*[†]

[†]Univ Lyon, Inria, EnsL, UCBL, CNRS, LIP, UMR 5668, F-69342, Lyon Cedex 07, France
[⋆]Ens de Lyon, CNRS, Laboratoire de Physique, F-69342, Lyon, France

ABSTRACT

This paper presents a multilevel fast iterative soft thresholding algorithm (FISTA), based on the use of the Moreau envelope to incorporate correction from coarse models, which is easy to compute when the explicit form of the proximal operator for the considered functions is known. This approach is supported by strong theoretical guarantees: we prove both the rate of convergence and the convergence of the iterates to a minimum in the convex case, an important result for ill-posed problems. We evaluate our approach on image restoration problems and we show that it outperforms classical FISTA for large-scale images.

Index Terms— multilevel optimization, inertial methods, image restoration, proximal methods.

1. INTRODUCTION

Many problems in signal and image processing involve minimising a sum of a data fidelity term f and a regularization function g, formally:

$$\min_{x\in\mathcal{H}} F(x) := f(x) + g(x) \tag{1}$$

where $\mathcal{H}$ is a real Hilbert space ($\mathcal{H} = \mathbb{R}^N$ in the following), $f : \mathcal{H} \to (-\infty, +\infty]$ and $g : \mathcal{H} \to (-\infty, +\infty]$ belong to $\Gamma_0(\mathcal{H})$ the class of convex, lower semi-continuous, and proper functions. Moreover, f is assumed to be differentiable with gradient L_f-Lipschitz and F is supposed to be coercive.

In the context of restoration, we aim to recover a good quality image from a degraded image $z = \mathrm{A}\bar{x} + \epsilon$, where $\mathrm{A} \in \mathbb{R}^{N\times N}$ models a linear degradation operator and ϵ stands for the additive noise. To solve this ill-posed problem, we generally consider a regularized least squares formulation, where we denote g the regularization function allowing us to choose the properties that we wish to impose on the solution. A usual choice is to apply the l_1-norm on the coefficients raised by a linear transformation $\mathrm{W} \in \mathbb{R}^{K\times N}$ (wavelets, frames, dictionary, ...), thus promoting the sparsity of the solution [2]. Given a regularization parameter $\lambda > 0$, the associated minimization problem reads:

$$\widehat{x} \in \underset{x\in\mathbb{R}^N}{\mathrm{Argmin}} \frac{1}{2}\|\mathrm{A}x - z\|_2^2 + \lambda\|\mathrm{W}x\|_1. \tag{2}$$

Many algorithms have been proposed in the literature to estimate $\widehat{x}$ (cf. [3–6]). They suffer from a significant increase in computational time with the dimension. Preconditioning techniques can be investigated but generally require strong assumptions for the choice of the preconditioning matrix (e.g., diagonal matrix) leading to limited gains. For the solution of large-scale problems with smooth objective function, it is possible to take advantage of the local structure of the optimisation problem (cf. [7] or [8]).

In this paper, we focus on a different family of approaches, the multilevel schemes, which exploit different resolutions of the same problem. In such methods the objective function is approximated by a sequence of functions defined on reduced dimensional spaces (coarser spaces). The descent step is thus calculated at coarser levels with minimal cost and then projected to the fine levels.

These approaches have been mainly studied for the solution of partial differential equations (PDEs), in which f and g are supposed to be differentiable [9, 10], but recently this idea has also been exploited in [11–13] to define multilevel forward-backward proximal algorithms applicable to problem (1) in the case where g is non differentiable.

In this paper we propose a variant of these methods, which we call MMFISTA for *Moreau Multilevel FISTA* providing a multilevel alternative to inertial strategies such as FISTA [14, 15]. Our framework relies on the Moreau envelope to define smooth coarse approximations of g, which can be easily constructed when the proximal operator of g is known in explicit form. Furthermore, we show under mild assumptions that the convergence guarantees of FISTA hold for MMFISTA, and in particular the convergence of the iterates, an important result for ill-posed problems and, to our knowledge, never established for multilevel inertial proximal methods.

The paper is organized as follows. In Section 2, we recall the main principles of FISTA. Then, we describe MMFISTA. In Section 3, we present its convergence guarantees. Finally,

The authors would like to thank the GdR ISIS for funding the MOMIGS project and the ANR-19-CE48-0009 Multisc'In project. We also gratefully acknowledge the support of the Centre Blaise Pascal's IT test platform at ENS de Lyon (Lyon, France) for the computing facilities. The platform operates the SIDUS [1] solution developed by Emmanuel Quemener. All experiments were performed on a machine with 8-cores Intel Core i7-10700, 4.8 GHz, 62 Go of RAM.

in Section 4, we present numerical results to confirm the good behaviour of MMFISTA in an image restoration context.

2. MULTILEVEL FISTA

FISTA – Among the numerous algorithms designed to solve a minimization problem of the form (1), the most standard strategy is FISTA [14], which relies on forward-backward iterations and extrapolation steps, such that, for every $k = 0, 1, \dots$

$$x_{k+1} = \text{prox}_{\tau g}(y_k - \tau_k \nabla f(y_k)) \tag{3}$$

$$y_{k+1} = x_{k+1} + \alpha_k (x_{k+1} - x_k) \tag{4}$$

where $x_0 = y_0$ and $\alpha_k = \frac{t_k - 1}{t_{k+1}}$ for all $k \geq 1$. Choosing $t_k = \left(\frac{k+a-1}{a}\right)$ where $a > 2$ [15, Definition 3.1] and $\tau \in (0, L_f^{-1})$ ensures various convergence guarantees (see [15, Theorem 3.5 and 4.1]). We will denote these conditions (AD) in the following.

Multilevel framework – The multilevel framework exploits a hierarchy of objective functions, which are representations of F at different resolutions and alternate minimization between these objective functions (following a V cycle procedure [9]). Without loss of generality and for the sake of clarity, we consider the two-level case: we index by h (resp. H) all quantities defined at the fine (resp. coarse) level. We thus define $F_h := F : \mathbb{R}^{N_h} \to (-\infty, +\infty]$ the objective function at the fine level where $N_h = N$, involving $f_h := f$ and $g_h := g$. Its approximation at the coarse level is denoted $F_H : \mathbb{R}^{N_H} \to (-\infty, +\infty]$ where $N_H < N_h$, which involves f_H and g_H. We also define transfer information operators: a linear operator $\mathrm{I}_h^H : \mathbb{R}^{N_h} \to \mathbb{R}^{N_H}$ that sends information from the fine level to the coarse level, and conversely $\mathrm{I}_H^h : \mathbb{R}^{N_H} \to \mathbb{R}^{N_h}$ that sends information from the coarse level back to the fine level. It is classical to choose $\mathrm{I}_H^h = \eta(\mathrm{I}_h^H)^T$, with $\eta > 0$.

In a multilevel scheme, we improve the intermediate iterate y_k^h by performing iterations at the coarse level: y_k^h is projected to the coarse level with I_h^H (5a), a sequence $(x_{k,\ell}^H)_{\ell \in \mathbb{N}}$ is defined (where k represents the current iteration at the fine level and ℓ indexes the iterations at the coarse level) such that: $x_{k,\ell+1}^H = \Phi_{k,\ell}^H(x_{k,\ell}^H)$, with $\Phi_{k,\ell}^H$ any operator such that $F_H(x_{k,m}^H) \leq F_H(x_{k,0}^H)$ for some $m > 0$. This yields after m iterations at the coarse level (5b) to a step being brought back at the fine level (5c).

Then, the generic iteration k of a multilevel method reads :

$$x_{k,0}^H = \mathrm{I}_h^H y_k^h \tag{5a}$$

$$x_{k,m}^H = \Phi_{k,m-1}^H \circ .. \circ \Phi_{k,0}^H(x_{k,0}^H) \tag{5b}$$

$$\bar{y}_k^h = y_k^h + \bar{\tau}_{h,k} \mathrm{I}_H^h \left(x_{k,m}^H - x_{k,0}^H\right) \tag{5c}$$

$$x_{k+1}^h = \text{prox}_{\tau_h g_h}(\bar{y}_k^h - \tau_h \nabla f_h(\bar{y}_k^h)) \tag{5d}$$

$$y_{k+1}^h = x_{k+1}^h + \alpha_{h,k}(x_{k+1}^h - x_k^h) \tag{5e}$$

By taking $x_{k,m}^H = x_{k,0}^H$ one recovers the standard FISTA iteration. To ensure that the correction term $x_{k,m}^H - x_{k,0}^H$, once projected from coarse level to fine level, induces a decrease of F_h, we need to appropriately choose :

- the coarse model F_H,
- the minimization scheme Φ_H.

Coarse model F_H – The coarse iterations are built using the Moreau envelope of g_h and of its coarse approximation g_H. The Moreau envelope provides a natural choice to extend ideas coming from the classical smooth case [10] to proximal gradient methods because of its smoothness and its expression involving the proximity operator. We first recall that for $\gamma > 0$ and g being a convex, lower semi-continuous, and proper function of $\mathcal{H}$ in $(-\infty, +\infty]$, its Moreau envelope, denoted ${}^{\gamma}g$, is the convex, continuous, real-valued function defined by

$${}^{\gamma}g = \inf_{y \in \mathcal{H}} g(y) + (1/2\gamma)\| \cdot - y\|^2, \tag{6}$$

which can be expressed explicitly with $\text{prox}_{\gamma g}$ [16, Remark 12.24]. The gradient of ${}^{\gamma}g$ is γ^{-1}-lipschitz and such that (Prop. 12.30 in [16])

$$\nabla({}^{\gamma}g) = \gamma^{-1}(\text{Id} - \text{prox}_{\gamma g}). \tag{7}$$

At iteration k, the coarse model F_H is defined as

$$F_H(x_H) = f_H(x_H) + g_H(x_H) + \langle v_{H,k}, x_H \rangle \tag{8}$$

where

$$\begin{aligned} v_{H,k} =& I_h^H \left(\nabla f_h(y_k^h) + \nabla({}^{\gamma_h} g_h)(y_k^h)\right) \\ & - (\nabla f_H(x_{k,0}^H) + \nabla({}^{\gamma_H} g_H)(x_{k,0}^H)). \end{aligned} \tag{9}$$

The third term in (8) is added to enforce the first order coherence between a smoothed coarse objective function

$$F_{H,\gamma_H}(x_H) = f_H(x_H) + {}^{\gamma_H} g_H(x_H) + \langle v_{H,k}, x_H \rangle \tag{10}$$

and a smoothed fine objective function F_{h,γ_h} [12] near $x_{k,0}^H$:

$$\nabla F_{H,\gamma_H}(x_{k,0}^H) = I_h^H \nabla F_{h,\gamma_h}(y_k^h). \tag{11}$$

The choice of the smoothing parameters γ_h and γ_H will be discussed in Section 4. This condition ensures that if $x_{k,m}^H - x_{k,0}^H$ is a descent direction for F_{H,γ_H} at $x_{k,0}^H$, then $I_H^h(x_{k,m}^H - x_{k,0}^H)$ is a descent direction for F_{h,γ_h} as well:

$$\langle I_H^h(x_{k,m}^H - x_{k,0}^H), \nabla F_{h,\gamma_h}(y_k^h) \rangle \leq 0. \tag{12}$$

According to properties of the Moreau envelope and the principles developed in [17], if $x_{k,m}^H - x_{k,0}^H$ is a descent direction for F_{H,γ_H}, we obtain

$$F_h(y_k^h + \bar{\tau}_{h,k} I_H^h(x_{k,m}^H - x_{k,0}^H)) \leq F_h(y_k^h) + \beta\gamma_h \tag{13}$$

where $\bar{\tau}_{h,k}$ controls that the update is not too big and where $\beta > 0$ depends on g_h. This ensures that F_h is decreasing up to a constant $\beta\gamma_h$ (which can be made arbitrarily small) after a use of the coarse models. Now we show how to enforce the decrease of F_{H,γ_H}.

Minimization operator Φ_H **–** At the coarse level we can decide to consider either the non-smooth approximation (8) of the objective function or the smoothed version (10). Both cases lead to a decrease in F_{H,γ_H}: indeed, taking the Moreau envelope of g_H in $F_H(x_{k,m}^H) \leq F_H(x_{k,0}^H)$ yields $F_{H,\gamma_H}(x_{k,m}^H) \leq F_{H,\gamma_H}(x_{k,0}^H)$. The two cases are linked by the same choice of the correction term to ensure the coherence between the two levels (9). We consider here three different strategies :

1. Gradient steps on the smoothed F_{H,γ_H}: $\Phi_S^H = (\mathrm{Id} - \tau_H(\nabla(f_H + {}^{\gamma_H}g_H) + v_H))$
2. Proximal gradient steps on the non-smooth F_H: $\Phi_{FB}^H = \mathrm{prox}_{\tau_H g_H}\left(\mathrm{Id} - \tau_H(\nabla f_H + v_H)\right)$.
3. FISTA steps on the non-smooth F_H with the previous proximal gradient step and where $\alpha_{k,\ell}^H$ follows (AD) conditions. Noted Φ_{FISTA}^H in the following.

Practical considerations – Our algorithm is based on a simple construction of F_H and $v_{H,k}$, as long as the computation of the associated proximal operator has an explicit form, which is a rather reasonable assumption. Our method is sketched in Algorithm 1. The step length at both levels can be selected either by fixing a value below the threshold guaranteeing convergence, defined by the Lipschitz constants associated to the considered functions when they are known, or by a linear search to guarantee it. To ensure the convergence of the iterates, we impose at most p uses of the coarse models F_H (one use corresponds to a full V-scheme cycle), which is also recommended to significantly improve the computation time (cf. Section 4).

3. CONVERGENCE OF THE ITERATES

Provided that we use the coarse models a finite number of times, we can prove the convergence of the iterates to a minimizer of $F = F_h$ and that the rate of convergence remains $O(1/k^2)$. First, we consider the sequence of corrections from the coarse models.

Lemma 1. *Let $L_{f,h}$ and $L_{f,H}$ the Lipschitz constants of f_h and f_H, respectively. Let $\tau_h, \tau_H \in (0, +\infty)$ the step sizes taken at fine and coarse levels, respectively. Assume that $\tau_H < (L_{f,H})^{-1}$ and that $\tau_h < L_{f_h}^{-1}$ and denote $\widehat{\tau}_h = \sup_k \bar{\tau}_{h,k}$. The sequence $(c_k^h)_{k\in\mathbb{N}}$ in $\mathcal{H}$ generated by Algorithm 1 defined by :*

$$c_k^h = \nabla f_h(y_k^h) - \nabla f_h(\bar{y}_k^h) + (\tau_h)^{-1}\widehat{\tau}_h I_H^h(x_{k,m}^H - x_{k,0}^H)$$

if a coarse correction is used at iteration k and $c_k^h = 0$ otherwise, is such that $\sum_{k\in\mathbb{N}} k\|c_k^h\| < +\infty$.

Algorithm 1: MMFISTA

```
Data: x_0^h, ε_h, γ, m, p > 0, t_0^h = 1, a > 2, k = 0, r = 0
while ||x_{k+1}^h − x_k^h|| > ε_h do
    if r < p then
        r = r + 1
        x_{k,0}^H = y_{k,0}^H = I_h^H y_k^h
        v_{H,k} = I_h^H ∇F_{h,γ_h}(y_k^h) − ∇F_{H,γ_H}(x_{k,0}^H)
        for ℓ = 0 ... m − 1 do
            y_{k,ℓ+1}^H = Φ_{k,ℓ}^H(x_{k,ℓ}^H)
            x_{k,ℓ+1}^H = y_{k,ℓ+1}^H + α_{k,ℓ}^H(y_{k,ℓ+1}^H − y_{k,ℓ}^H)
        end
        Set τ̄_{h,k} > 0, ȳ_k^h = y_k^h + τ̄_{h,k} I_H^h(x_{k,m}^H − x_{k,0}^H)
        x_{k+1}^h = prox_{τ_h g_h}(ȳ_k^h − τ_h ∇f_h(ȳ_k^h))
    else
        x_{k+1}^h = prox_{τ_h g_h}(y_k^h − τ_h ∇f_h(y_k^h))
    end
    t_k^h = ((k+a−1)/a), α_k^h = (t_k^h − 1)/t_{k+1}^h
    y_{k+1}^h = x_{k+1}^h + α_k^h(x_{k+1}^h − x_k^h)
end
```

The proof of this lemma is based on the fact that if the number of coarse corrections is finite, we only need to construct bounded sequences at coarse level so that $I_H^h(x_{k,m}^H - x_{k,0}^H)$ is also bounded. From this result we deduce the following theorem :

Theorem 1. *Consider Algorithm 1, suppose that for all $k \in \mathbb{N}^*$, t_k^h in Eqs.* (3) *and* (4) *satisfy (AD) conditions [15]. Suppose moreover that the assumptions of Lemma 1 hold. Then :*

- *The sequence $(k^2\left(F_h(x_k^h) - F_h(x^*)\right))_{k\in\mathbb{N}}$ belongs to $\ell_\infty(\mathbb{N})$.*
- *The sequence $(x_k^h)_{k\in\mathbb{N}}$ given by Algorithm 1 weakly converges to a minimizer of F_h.*

Proof. We combine [15, Theorem 3.5, 4.1, and Corollary 3.8] with Lemma 1 to prove the desired result. ☐

4. RESULTS

We numerically illustrate the performance of our algorithm in the context of image restoration.

Dataset and degradation – We consider large images of size 2048×2048, yielding $N = (2^J)^2 \simeq 4 \times 10^6$ with $J = 11$. The linear degradation operator A_h is constructed with HNO [18] as a Kroenecker product with Neumann boundary conditions and we add a Gaussian noise (see the legend of Fig.1 for details). In all tests, the regularization parameter λ_h was chosen by a grid search, in order to maximize the Signal-to-Noise-Ratio (SNR) of $\widehat{x}$ obtained with FISTA at convergence. Also, we initialise x_0 with the Wiener filtering of z.

Multilevel architecture – We use a 5-levels hierarchy: from 2048×2048 ($J = 11$) to 128×128 (indexed by $J - 4$). We choose I_h^H as the low scale projection on a symlet wavelet

Noise \ Blur		(a) size(blur) = [40, 40], σ(blur) = 7.3						(b) size(blur) = [88, 88], σ(blur) = 16				
(1) $\sigma = 0.01$	FISTA CPU time	16	28	42	161	401		17	30	42	148	421
	$\Phi_{H,S}$	-20 $\color{red}\bullet$	-22 $\color{blue}\bullet$	$+1$ $\color{red}\bullet$	$+1$ $\color{red}\bullet$	-1 $\color{red}\bullet$		$\mathbf{-51}$ $\color{red}\bullet$	$\mathbf{-44}$ $\color{blue}\bullet$	-18 $\color{blue}\bullet$	$\mathbf{+4}$ $\color{red}\bullet$	$\mathbf{-1}$ $\color{red}\bullet$
	$\Phi_{H,FB}$	-19 $\color{red}\bullet$	-19 $\color{blue}\bullet$	$+5$ $\color{red}\bullet$	$\mathbf{+2}$ $\color{red}\bullet$	$\mathbf{+1}$ $\color{red}\bullet$		-50 $\color{red}\bullet$	-42 $\color{blue}\bullet$	-15 $\color{blue}\bullet$	$+6$ $\color{red}\bullet$	$+1$ $\color{red}\bullet$
	Φ^H_{FISTA}	$\mathbf{-51}$ $\color{red}\bullet$	$\mathbf{-32}$ $\color{blue}\bullet$	$\mathbf{-4}$ $\color{blue}\bullet$	$+2$ $\color{red}\bullet$	$+1$ $\color{red}\bullet$		-50 $\color{red}\bullet$	-42 $\color{blue}\bullet$	$\mathbf{-35}$ $\color{blue}\bullet$	$+8$ $\color{red}\bullet$	$+1$ $\color{red}\bullet$
(2) $\sigma = 0.04$	FISTA CPU time	14	22	34	108	220		15	25	34	122	315
	$\Phi_{H,S}$	$\mathbf{-22}$ $\color{red}\bullet$	-10 $\color{blue}\bullet$	-1 $\color{red}\bullet$	$\mathbf{-1}$ $\color{red}\bullet$	-1 $\color{red}\bullet$		-29 $\color{red}\bullet$	-25 $\color{blue}\bullet$	-18 $\color{blue}\bullet$	$\mathbf{+3}$ $\color{red}\bullet$	$\mathbf{+1}$ $\color{red}\bullet$
	$\Phi_{H,FB}$	-22 $\color{red}\bullet$	-10 $\color{blue}\bullet$	-1 $\color{red}\bullet$	$+1$ $\color{red}\bullet$	-1 $\color{red}\bullet$		-42 $\color{red}\bullet$	-31 $\color{blue}\bullet$	-16 $\color{blue}\bullet$	$+5$ $\color{red}\bullet$	$+2$ $\color{red}\bullet$
	Φ^H_{FISTA}	-21 $\color{red}\bullet$	$\mathbf{-12}$ $\color{blue}\bullet$	$\mathbf{-10}$ $\color{blue}\bullet$	-1 $\color{blue}\bullet$	$\mathbf{-2}$ $\color{blue}\bullet$		$\mathbf{-42}$ $\color{red}\bullet$	$\mathbf{-31}$ $\color{blue}\bullet$	$\mathbf{-22}$ $\color{blue}\bullet$	$+7$ $\color{red}\bullet$	$+2$ $\color{red}\bullet$

Table 1: For each degradation : the first line of each subtable represents the computation time (in sec) needed by FISTA to reach $5, 2, 1, 0.1$ and 0.01% of the distance $\|F_h(x_0^h) - F_h(x_*^h)\|$. Then for each type of minimization algorithm at coarse level, we display the CPU time relative to FISTA (14) (in %) for the best configuration with a colored bullet : $p = 1$ $\color{red}\bullet$ and $p = 2$ $\color{blue}\bullet$. In all cases : $m = 5$. SNR of z : (1a) 11.05 (1b) 9.64 (2a) 11.03 (2b) 9.63. SNR of $x_{h,300}$ computed by MMFISTA : (1a) 12.71 (1b) 11.02 (2a) 12 (2b) 10.6.

with 10 vanishing moments and $I_H^h = \frac{1}{4}(I_h^H)^T$. We then construct f_H with the blurring matrix $A_H = I_h^H A_h I_H^h$ (which is never used explicitly due to the properties of the Kroenecker product [12, 18]). Thus $f_h = \frac{1}{2}\|A_h x_h - z_h\|^2$ and $f_H = \frac{1}{2}\|A_H x_H - I_h^H z_h\|^2$. The penalty term $g_h = \|W_h x_h\|_1$ is defined using a full wavelet decomposition over J levels, we construct $g_H = \|W_H x_H\|_1$ with a decomposition over $J-1$ up to $J-4$ levels, with $\lambda_H = \lambda_h/4$. The Moreau envelope parameter associated with g_H is set to $\gamma_H = 1.1$ while γ_h is set to 1, but both values do not seem to be critical here.

Visual result – We display the restored image $\widehat{x}$ and the convergence curves as a function of the iterations and the CPU time for one case in Fig.1. For clarity, we only display the behaviour of the method with Φ^H_{FISTA}.

Performance assessment – We measure Time(MMFISTA), the CPU time needed to reach a threshold of $5, 2, 1, 0.1$ and 0.01% of the distance $\|F_h(x_0^h) - F_h(\widehat{x})\|$, with $\widehat{x}$ computed beforehand by FISTA, and we compare it to Time(FISTA). We tested the performance for several values of m, and among our numerous numerical experiments, $m = 5$ at the different coarse levels appears to be a good compromise whatever the noise and blur levels. We report in Tab.1, the quantity:

$$\frac{\text{Time(MMFISTA)} - \text{Time(FISTA)}}{\text{Time(FISTA)}} \times 100, \qquad (14)$$

for $m = 5$ at every coarse levels. In this table we evaluate:

– Impact of p. In our numerical experiments we only consider $p = 1$ ($\color{red}\bullet$) or $p = 2$ ($\color{blue}\bullet$) uses of the coarse models, performed at the beginning of the iterative process. They allow to quickly determine the low frequencies components of the solution at the fine level. The choice of p depends on the desired accuracy. For rough approximation (e.g. 5%), fixing $p = 1$ is a good choice, while $p = 2$ seems better for finer approximations (e.g. thresholds set to 2%). When very precise approximation is wanted (e.g. threshold below 1%), the use of a multilevel strategy does not pay off, even if it does not degrade performance.

– Impact of noise and blur level. For all methods acceleration increases significantly as the blur gets worse. Moreover, as the noise decreases, the improvement obtained with Φ^H_{FISTA} increases, compared to others Φ_H.

The main conclusion is that with a few coarse corrections, our method can significantly accelerate the achievement of coarse solution accuracies, while it stays competitive with FISTA for high precision approximations.

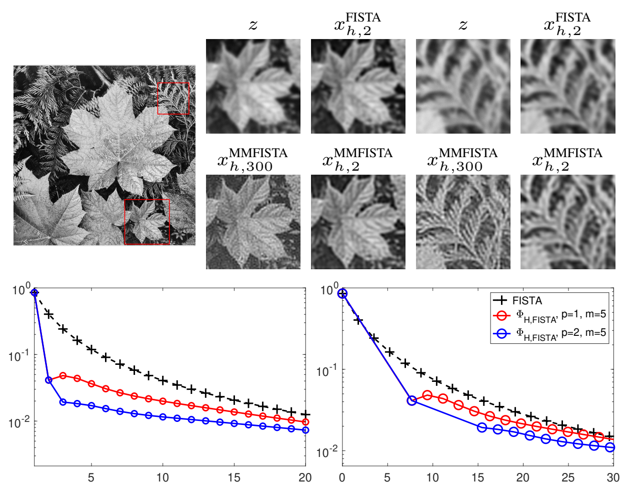

Fig. 1: *Top* : From left to right : Original 2048×2048 image[1] $\overline{x}$, (first row) zoom of the degraded image z for a noise with $\sigma = 0.01$ and a Gaussian blur of size 40×40 and 7.3 standard deviation and of x_2^h computed by FISTA. (second row) zoom of x_2^h and x_{300}^h computed by MMFISTA. *Bottom*: (left) Evolution of F_h versus iterations for MMFISTA with Φ^H_{FISTA} for $p = 1, 2$, $m = 5$; (right) Same for CPU time (in sec). $\lambda_h = 1.7 \times 10^{-4}$.

5. CONCLUSION

We have proposed a convergent multilevel FISTA method for image restoration that reaches coarse approximations of the optimal solution in a much smaller CPU time than FISTA and that is well suited to large images. A future research perspective is to extend this approach to other proximal algorithmic schemes and to study/or improve the associated convergence rates. We also want to investigate the influence of the information transfer operators, which remains an open question.

[1] A close-up of leaves in Glacier National Park, Montana taken by Ansel Adams in the 1930s

6. REFERENCES

[1] E. Quemener and M. Corvellec, "SIDUS—the Solution for Extreme Deduplication of an Operating System," *Linux J.*, vol. 2013, no. 235, Nov. 2013.

[2] N. Pustelnik, A. Benazza-Benhayia, Y. Zheng, and J.-C. Pesquet, "Wavelet-based Image Deconvolution and Reconstruction," *Wiley Encyclopedia of EEE*, 2016.

[3] P. L. Combettes and V. R. Wajs, "Signal Recovery by Proximal Forward-Backward Splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, Jan. 2005.

[4] P. L. Combettes and J.-C. Pesquet, *Proximal Splitting Methods in Signal Processing*, pp. 185–212, Springer New York, New York, NY, 2011.

[5] N. Parikh and S. Boyd, "Proximal Algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2014.

[6] A. Chambolle and T. Pock, "An Introduction to Continuous Optimization for Imaging," *Acta Numerica*, vol. 25, pp. 161–319, 2016.

[7] E. Thiébaut, "Optimization Issues in Blind Deconvolution Algorithms," *Proceedings of SPIE - The International Society for Optical Engineering, Astronomical Data Analysis II*, vol. 4847, pp. 174–183, 12 2002.

[8] E. Chouzenoux, J.-C. Pesquet, and A. Florescu, "A Stochastic 3MG Algorithm with Application to 2D Filter Identification," in *2014 22nd European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, 11 2014, pp. 1587–1591.

[9] S. G. Nash, "A Multigrid Approach to Discretized Optimization Problems," *Optimization Methods and Software*, vol. 14, no. 1-2, pp. 99–116, 2000.

[10] H. Calandra, S. Gratton, E. Riccietti, and X. Vasseur, "On High-Order Multilevel Optimization Strategies," *SIAM Journal on Optimization*, vol. 31, no. 1, pp. 307–330, 2021.

[11] V. Hovhannisyan, P. Parpas, and S. Zafeiriou, "MAGMA: Multilevel Accelerated Gradient Mirror Descent Algorithm for Large-Scale Convex Composite Minimization," *SIAM Journal on Imaging Sciences*, vol. 9, no. 4, pp. 1829–1857, Jan. 2016.

[12] P. Parpas, "A Multilevel Proximal Gradient Algorithm for a Class of Composite Optimization Problems," *SIAM Journal on Scientific Computing*, vol. 39, no. 5, pp. S681–S701, 2017.

[13] G. Lauga, E. Riccietti, N. Pustelnik, and P. Gonçalves, "Méthodes proximales multi-niveaux pour la restauration d'images," Nancy, France, Sept. 2022.

[14] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, Jan. 2009.

[15] J.-F. Aujol and C. Dossal, "Stability of Over-Relaxations for the Forward-Backward Algorithm, Application to FISTA," *SIAM Journal on Optimization*, vol. 25, no. 4, pp. 2408–2433, Jan. 2015.

[16] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics. Springer International Publishing, New York, 2017.

[17] A. Beck and M. Teboulle, "Smoothing and First Order Methods: A Unified Framework," *SIAM Journal on Optimization*, vol. 22, no. 2, pp. 557–580, Jan. 2012.

[18] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images*, Society for Industrial and Applied Mathematics, 2006.