Multiscale analysis in image processing

Multilevel optimization

Barbara Pascal[†] **and Nelly Pustelnik**[‡]

bpascal-fr.github.io/talks

June 2025

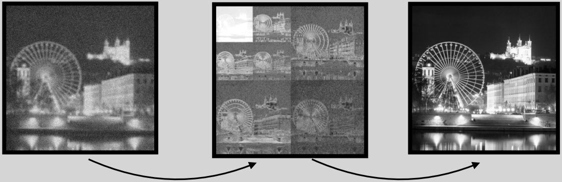
[†] Nantes Université, École Centrale Nantes, CNRS, LS2N, F-44000 Nantes, France

[‡] CNRS, ENSL, Laboratoire de physique, F-69342 Lyon, France

Multiresolution/multilevel

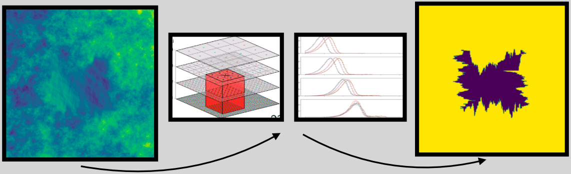
Multiresolution
to perform
image restoration

(~2000–2015)



Multiresolution
to perform
texture
segmentation

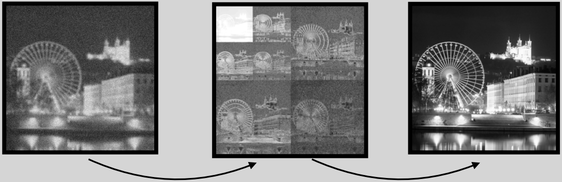
(~2014- now)



Multiresolution/multilevel

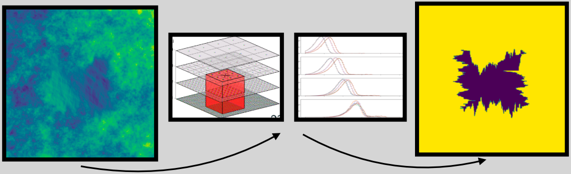
Multiresolution
to perform
image restoration

(~2000–2015)



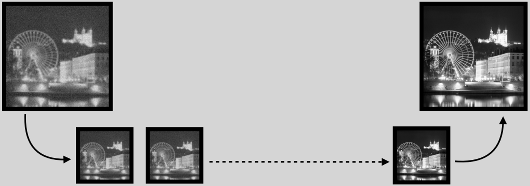
Multiresolution
to perform
texture
segmentation

(~2014- now)

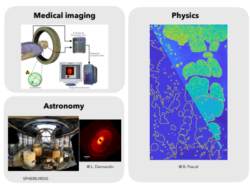


Multiresolution
to accelerate
algorithms

(~2016- now)



Inverse problems: variables and key equations



☛ Variables

- $\mathrm{z} \in \mathbb{R}^M$: data.
- $\overline{\mathrm{x}} \in \mathbb{R}^N$: unknown parameters.
- $\widehat{\mathrm{x}} \in \mathbb{R}^N$: estimated parameters.

☛ Forward model

$$\mathrm{z} = \mathcal{D}(\mathrm{A}\overline{\mathrm{x}})$$

Stochatic degradation **Linear** operator

☛ Inverse problem

$$\widehat{\mathrm{x}} = \mathrm{d}_\Theta(\mathrm{z})$$

☛ **Goal**: Estimate $\widehat{\mathrm{x}}$ close to $\overline{\mathrm{x}}$ from z, A, noise statistic $\mathcal{D}$, and prior information on the class of image to recover.

Inversion $\widehat{\mathrm{x}} = d_\Theta(\mathrm{z})$

- ➡ [1922] **Maximum likelihood** (Fisher).

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ \frac{1}{2}\|\mathrm{Ax} - \mathrm{z}\|_2^2 = (\mathrm{A}^*\mathrm{A})^{-1}\mathrm{A}^*\mathrm{z}$$

- ➡ [1963] **Regularization** (Tikhonov, Huber)

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ \frac{1}{2}\|\mathrm{Ax} - \mathrm{z}\|_2^2 + \theta\|\mathrm{Lx}\|_2^2 \qquad \text{avec} \quad \theta > 0$$

- ➡ [2000] **Sparsity** (Donoho, Daubechies-Defrise-DeMol,...)

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ \frac{1}{2}\|\mathrm{Ax} - \mathrm{z}\|_2^2 + \theta\|\mathrm{Lx}\|_\star$$

- ➡ [2010] **“End to end” neural networks**

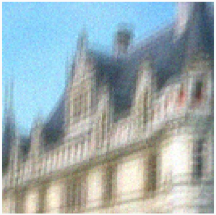
$$\widehat{\mathrm{x}} = \mathrm{NN}_\Theta(\mathrm{z})$$

- ➡ [2020] **Plug-and-Play**

$$0 \in \mathrm{A}^*(\mathrm{A}\widehat{\mathrm{x}} - \mathrm{z}) + \mathbf{B}(\widehat{\mathrm{x}})$$

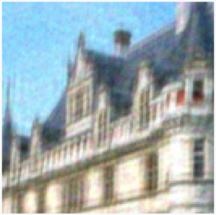
Summary of inverse problems in imaging

Original



Degraded

SNR = 13.4 dB

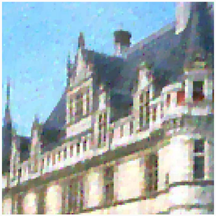


Tikhonov

SNR = 16.4 dB

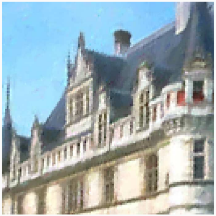
DTT

SNR = 16.6 dB



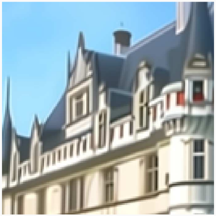
TV

SNR = 18.8 dB



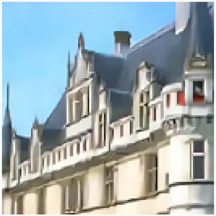
NLTV

SNR = 19.4 dB



PnP-DRUnet

SNR = 20.0 dB



PnP-ScCP

SNR = 20.2 dB

Focus in this presentation

- ➡ [1922] **Maximum likelihood** (Fisher).

$$\widehat{x} \in \operatorname*{Argmin}_{x} \frac{1}{2}\|Ax - z\|_2^2 = (A^*A)^{-1}A^*z$$

- ➡ [1963] **Regularization** (Tikhonov, Huber)

$$\widehat{x} \in \operatorname*{Argmin}_{x} \frac{1}{2}\|Ax - z\|_2^2 + \theta\|Lx\|_2^2 \qquad \text{avec} \quad \theta > 0$$

- ➡ [2000] **Sparsity** (Donoho, Daubechies-Defrise-DeMol,...)

$$\widehat{x} \in \operatorname*{Argmin}_{x} \frac{1}{2}\|Ax - z\|_2^2 + \theta\|Lx\|_\star$$

- ➡ [2010] **"End to end" neural networks**

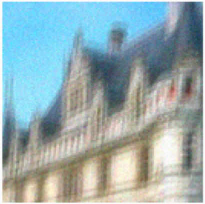
$$\widehat{x} = NN_\Theta(z)$$

- ➡ [2020] **Plug-and-Play**

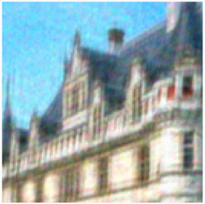
$$0 \in A^*(A\widehat{x} - z) + \mathbf{B}(\widehat{x})$$

Focus in this presentation


Original


Degraded

SNR = 13.4 dB


Tikhonov

SNR = 16.4 dB


DTT

SNR = 16.6 dB


TV

SNR = 18.8 dB


NLTV

SNR = 19.4 dB


PnP-DRUnet

SNR = 20.0 dB


PnP-ScCP

SNR = 20.2 dB[4]

Iterative scheme

➡ **Minimization problem** :

$$\boxed{\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\; f(\mathrm{x}) + g(\mathrm{x})}$$

with f and g either diff. with Lipschitz gradient or proximable.

➡ **Design of a recursive sequence of the form**

$$(\forall k \in \mathbb{N}) \qquad \mathrm{x}^{[k+1]} = \mathbf{T}(\mathrm{x}^{[k]}),$$

Gradient descent	$\mathbf{T} = \mathrm{Id} - \tau(\nabla f + \nabla g)$
Proximal point algorithm	$\mathbf{T} = \mathrm{prox}_{\tau(f+g)}$
Forward-Backward	$\mathbf{T} = \mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f)$
Peaceman-Rachford	$\mathbf{T} = (2\,\mathrm{prox}_{\tau g} - \mathrm{Id}) \circ (2\,\mathrm{prox}_{\tau f} - \mathrm{Id})$
Douglas-Rachford	$\mathbf{T} = \mathrm{prox}_{\tau g}(2\,\mathrm{prox}_{\tau f} - \mathrm{Id}) + \mathrm{Id} - \mathrm{prox}_{\tau f}$

Iterative scheme

➡ **Minimization problem** :

$$\boxed{\widehat{\mathrm{x}} \in \operatorname*{Argmin}_{\mathrm{x}} f(\mathrm{x}) + g(\mathrm{x})}$$

with f and g either diff. with Lipschitz gradient or proximable.

➡ **Design of a recursive sequence of the form**

$$(\forall k \in \mathbb{N}) \qquad \mathrm{x}^{[k+1]} = \mathbf{T}(\mathrm{x}^{[k]}),$$

Gradient descent	$\mathbf{T} = \mathrm{Id} - \tau(\nabla f + \nabla g)$
Proximal point algorithm	$\mathbf{T} = \mathrm{prox}_{\tau(f+g)}$
Forward-Backward	$\mathbf{T} = \mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f)$
Peaceman-Rachford	$\mathbf{T} = (2\,\mathrm{prox}_{\tau g} - \mathrm{Id}) \circ (2\,\mathrm{prox}_{\tau f} - \mathrm{Id})$
Douglas-Rachford	$\mathbf{T} = \mathrm{prox}_{\tau g}(2\,\mathrm{prox}_{\tau f} - \mathrm{Id}) + \mathrm{Id} - \mathrm{prox}_{\tau f}$

Main goal : provide acceleration for high dimensional problems

High dimensional problems → high computation time.

Alternatives :

- **FISTA** [Beck & Teboulle, 2009] [Chambolle & Dossal, 2015],
- **Preconditionning** [Donatelli, 2019][Repetti et al., 2014],
- **Blocks methods** [Liu, 1996] [Chouzenoux et al., 2016] [Salzo, Villa 2022],
- **multiresolution strategy**
  ☛ Idea that comes from the PDE field [Nash, 2000].
  ☛ preliminary results for non-smooth optimization in [Parpas, 2017].

Common aim of these methods:

☛ improve the gradient/proximal gradient steps with well chosen rules.

Multilevel algorithm for smooth optimization

Some references:

- A. Javaherian and S. Holman, **A Multi-Grid Iterative Method for Photoacoustic Tomography**, IEEE Transactions on Medical Imaging, (2017)
- S. W. Fung and Z. Wendy, **Multigrid Optimization for Large-Scale Ptychographic Phase Retrieval**, SIAM Journal on Imaging Sciences, 13 (2020)
- J. Plier, F. Savarino, M. Kočvara, and S. Petra, **First-Order Geometric Multilevel Optimization for Discrete Tomography**, in Scale Space and Variational Methods in Computer Vision, A. Elmoataz, J. Fadili, Y. Quéau, J. Rabin, and L. Simon, eds., vol. 12679, Springer International Publishing, Cham, (2021)

→ Successful attempts of accelerating minimization in imaging.

→ Restricted to smooth optimization.

Multilevel algorithms

First order descent methods

Goal:

$$\min_{\mathrm{x}\in\mathbb{R}^N} F(\mathrm{x}) := f(\mathrm{x}) + g(\mathrm{x})$$

f and g proper, lower semi-continuous, and convex.

f is assumed differentiable with Lipschitz gradient.

g is not necessarily differentiable.

Build a sequence: $\mathrm{x}^{[k+1]} = \Phi(\mathrm{x}^{[k]}) = \mathrm{x}^{[k]} - \mathrm{D}_k$

- If f and g are differentiable: Gradient descent

$$\mathrm{D}_k = \tau_k(\nabla f(\mathrm{x}^{[k]}) + \nabla g(\mathrm{x}^{[k]}))$$

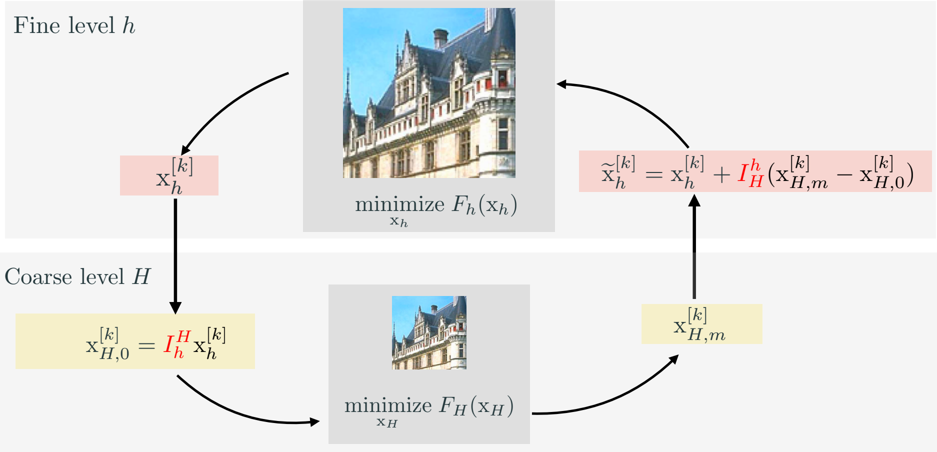
- If g is not differentiable: Proximal gradient descent

$$\mathrm{D}_k = \mathrm{x}^{[k]} - \mathrm{prox}_{\tau_k g}\left(\mathrm{x}^{[k]} - \tau_k \nabla f(\mathrm{x}^{[k]})\right)$$

Multilevel smooth optimization

Goal: Exploit hierarchy of approximations of the objective function.

Example: Two levels case with fine (h) and coarse (H) levels.



Fine level h

$\mathrm{x}_h^{[k]}$

$\underset{\mathrm{x}_h}{\text{minimize}}\ F_h(\mathrm{x}_h)$

$\widetilde{\mathrm{x}}_h^{[k]} = \mathrm{x}_h^{[k]} + I_H^h(\mathrm{x}_{H,m}^{[k]} - \mathrm{x}_{H,0}^{[k]})$

Coarse level H

$\mathrm{x}_{H,0}^{[k]} = I_h^H \mathrm{x}_h^{[k]}$

$\underset{\mathrm{x}_H}{\text{minimize}}\ F_H(\mathrm{x}_H)$

$\mathrm{x}_{H,m}^{[k]}$

Design of I_H^h and I_h^H: Information transfer operators

Definition: *coherent information transfer (CIT)* operators

- $I_h^H : \mathbb{R}^{N_h} \to \mathbb{R}^{N_H}$ (transfer from fine to coarse scales)
- $I_H^h : \mathbb{R}^{N_H} \to \mathbb{R}^{N_h}$ (transfer from coarse to fine scales)

if there exists $\nu > 0$ such that:

$$I_H^h = \nu (I_h^H)^T.$$

- particular case of squared grids reads:

$$I_h^H = \frac{1}{16} \underbrace{\begin{pmatrix} 2 & 1 & 0 & \dots & & & 0 \\ 0 & 1 & 2 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & & & & 0 \\ 0 & \dots & & 0 & 1 & 2 & 1 \end{pmatrix}}_{\sqrt{N_h}/2 \times \sqrt{N_h}} \otimes \underbrace{\begin{pmatrix} 2 & 1 & 0 & \dots & & & 0 \\ 0 & 1 & 2 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & & & & 0 \\ 0 & \dots & & 0 & 1 & 2 & 1 \end{pmatrix}}_{\sqrt{N_h}/2 \times \sqrt{N_h}} \in \mathbb{R}^{N_H \times N_h}$$

Design of F_H: First order coherence

Smoothed convex function [Beck 2012, Definition 2.1]

Let g be a convex, l.s.c., and proper function on $\mathbb{R}^N$.

For every $\gamma > 0$, g_γ is a smoothed convex function if there exist scalars η_1, η_2 satisfying $\eta_1 + \eta_2 > 0$ such that the following holds:

$$(\forall \mathrm{y} \in \mathbb{R}^N) \qquad g(\mathrm{y}) - \eta_1\gamma \leqslant g_\gamma(\mathrm{y}) \leqslant g(\mathrm{y}) + \eta_2\gamma.$$

Design of F_H: First order coherence

First order coherence [Nash, 2000][Parpas et al. 1016, 2017]

The first order coherence between the smoothed version of the objective function F_h at the fine level and the coarse level objective function F_H is verified in a neighbourhood of $\mathrm{y}_h \in \mathbb{R}^{N_h}$ if the following equality holds:

$$\nabla F_H(I_h^H \mathrm{y}_h) = I_h^H \nabla \left(f_h + g_{h,\gamma_h}\right)(\mathrm{y}_h).$$

☛ **Impact**: Coherence up to order one in the neighbourhood of the current iterates $\mathrm{y}_h = \mathrm{y}_h^{[k]}$.

Design of F_H: First order coherence

Coarse model F_H for non-smooth functions

The coarse model F_H is defined for the point $\mathrm{y}_h \in \mathbb{R}^{N_h}$ as:

$$F_H = f_H + g_{H,\gamma_H} + \langle \mathrm{v}_h, \cdot \rangle, \tag{1}$$

where

$$\mathrm{v}_h = I_h^H \left(\nabla f_h(\mathrm{y}_h) + \nabla g_{h,\gamma_h}(\mathrm{y}_h) \right) - \left(\nabla f_H(I_h^H \mathrm{y}_h) + \nabla g_{H,\gamma_H}(I_h^H \mathrm{y}_h) \right).$$

Lemma If F_H is given by definition (1), it necessarily verifies the first order coherence.

Proof.
Considering the gradient of the coarse model F_H and combining it with the definition of v_h, yields

$$\begin{aligned} \nabla F_H(I_h^H \mathrm{y}_h) &= \nabla f_H(I_h^H \mathrm{y}_h) + \nabla g_{H,\gamma_H}(I_h^H \mathrm{y}_h) + \mathrm{v}_h, \\ &= I_h^H \left(\nabla f_h(\mathrm{y}_h) + \nabla g_{h,\gamma_h}(\mathrm{y}_h) \right). \end{aligned}$$

Design of F_H: First order coherence

Coarse model F_H for non-smooth functions

The coarse model F_H is defined for the point $\mathrm{y}_h \in \mathbb{R}^{N_h}$ as:

$$F_H = f_H + g_{H,\gamma_H} + \langle \mathrm{v}_h, \cdot \rangle,$$

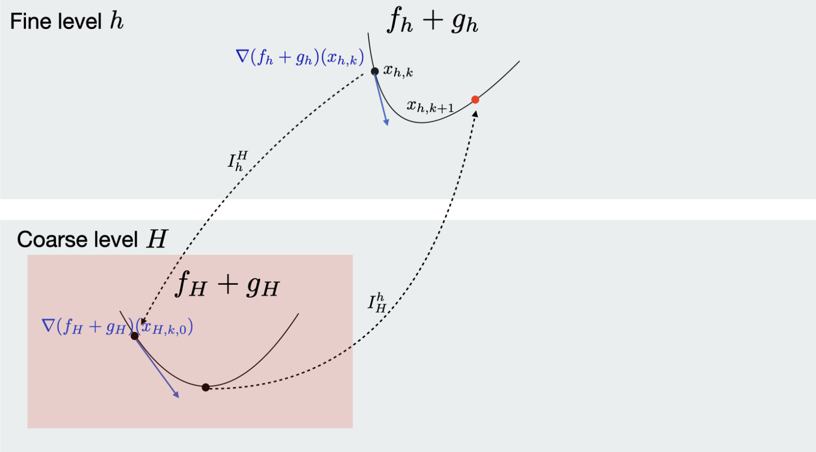
where

$$\mathrm{v}_h = I_h^H \left(\nabla f_h(\mathrm{y}_h) + \nabla g_{h,\gamma_h}(\mathrm{y}_h) \right) - \left(\nabla f_H(I_h^H \mathrm{y}_h) + \nabla g_{H,\gamma_H}(I_h^H \mathrm{y}_h) \right).$$

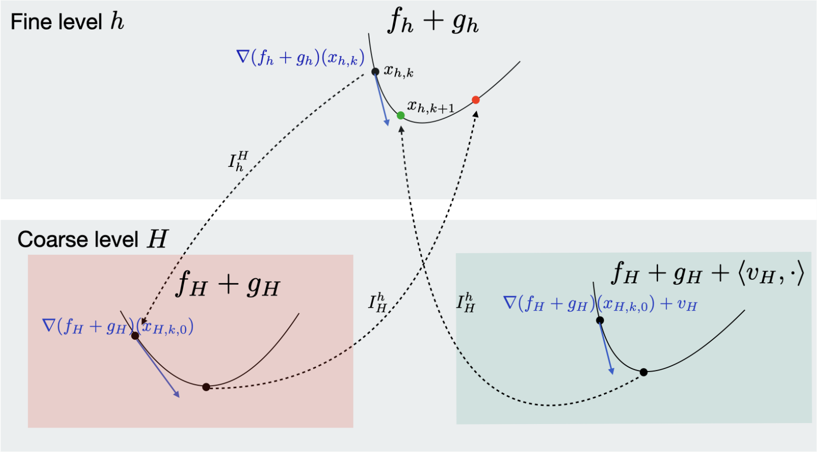
Remarks:

- ☛ Adding the linear term $\langle \mathrm{v}_h, \cdot \rangle$ to $f_H + g_{H,\gamma_H}$ allows to impose the so-called *first order coherence*.
- ☛ if g_h and g_H are smooth by design, one can simply replace g_{H,γ_H} and g_{h,γ_h} by g_H and g_h.

Design of F_H: First order coherence



Design of F_H: First order coherence



IML FB: Multilevel algorithm for nonsmooth optimization

[Lauga et al. 2024]

```
Set $\mathrm{x}_h^{[0]}, \mathrm{y}_h^{[0]} \in \mathbb{R}^N$, $t_{h,0} = 1$
while Stopping criterion is not met do
    if Descent condition then
        $\mathrm{s}_{H,0}^{[0]} = I_h^H \mathrm{x}_h^{[k]}$ Projection
        $\mathrm{s}_{H,m}^{[k]} = \mathbf{T}_{H,m-1} \circ .. \circ \mathbf{T}_{H,0}(\mathrm{s}_{H,0}^{[0]})$ Coarse minimization
        Set $\bar{\tau}_{h,k} > 0$,
        $\bar{\mathrm{x}}_h^{[k]} = \mathrm{x}_h^{[k]} + \bar{\tau}_{h,k} I_H^h \left(\mathrm{s}_{H,m}^{[k]} - \mathrm{s}_{H,0}^{[0]}\right)$ Coarse step update
    else
        $\bar{\mathrm{x}}_h^{[k]} = \mathrm{x}_h^{[k]}$
    end if
    $\mathrm{x}_h^{[k+1]} = \mathbf{T}_{h,k}(\bar{\mathrm{x}}_h^{[k]})$ Forward-Backward step
end while
```

Convergence analysis

Lemma *(Fine level decrease)*[Lauga et al. 2024] Let assume that I_h^H and I_H^h are CIT operators and that F_H satisfies Definition (1) and $\mathbf{T}_{H,\bullet}$ allows a decrease of the coarse model. The iterations of IML FB ensure:

$$F_h(\mathrm{x}_h + \bar{\tau} I_H^h(\mathrm{s}_{H,m} - \mathrm{s}_{H,0})) \leqslant F_h(\mathrm{x}_h) + (\eta_1 + \eta_2)\gamma_h.$$

Proof.
This directly comes from the definition of a smoothed convex function:

$$\begin{aligned} & F_h(\mathrm{x}_h + \bar{\tau}_h I_H^h(\mathrm{s}_{H,m} - \mathrm{s}_{H,0})) \\ & \quad \leqslant (L_h + R_{h,\gamma_h})(\mathrm{y}_h + \bar{\tau}_h I_H^h(\mathrm{s}_{H,m} - \mathrm{s}_{H,0})) + \eta_1\gamma_h \\ & \quad \leqslant (L_h + R_{h,\gamma_h})(\mathrm{x}_h) + \eta_1\gamma_h \\ & \quad \leqslant F_h(\mathrm{x}_h) + (\eta_1 + \eta_2)\gamma_h. \end{aligned}$$

Convergence analysis

Lemma *(Fine level decrease)*[Lauga et al. 2024] Let assume that I_h^H and I_H^h are CIT operators and that F_H satisfies Definition (1) and $\mathbf{T}_{H,\bullet}$ allows a decrease of the coarse model. The iterations of IML FB ensure:

$$F_h(\mathrm{x}_h + \bar{\tau} I_H^h(\mathrm{s}_{H,m} - \mathrm{s}_{H,0})) \leqslant F_h(\mathrm{x}_h) + (\eta_1 + \eta_2)\gamma_h.$$

- ☛ Coarse level minimization step, leads to a decrease of F_h, up to a constant $(\eta_1 + \eta_2)\gamma_h$ that can be made arbitrarily small by driving γ_h to zero.
- ☛ Commonly found in the literature of multilevel algorithms.
- ☛ Not sufficient to guarantee the convergence of the generated sequence.

What has been done:

☛ Remarks on multilevel framework to non-smooth optimization:

+ Handles non-smooth g.
+ Smoothing to define the first order coherence.
– Requires explicit form of $\mathrm{prox}_g = \mathrm{prox}_{\varphi \circ \mathrm{L}}$.
– No convergence guarantee to a minimizer.

☛ Some references:

- V. Hovhannisyan, P. Parpas, and S. Zafeiriou, **MAGMA: Multilevel Accelerated Gradient Mirror Descent Algorithm for Large-Scale Convex Composite Minimization**, SIAM J. Imaging Sciences (2016)
- P. Parpas, **A Multilevel Proximal Gradient Algorithm for a Class of Composite Optimization Problems**, SIAM J. Scient. Comp., 39 (2017)
- G. Lauga, E. Riccietti, N. Pustelnik, and P. Goncalves Multilevel FISTA for Image Restoration, IEEE ICASSP, 2023.

IML FISTA

Motivations and contribution

☛ **Goal**:

- **inexact proximal** steps to handle state-of-the-art regularization: Total Variation (TV) and Non-Local Total Variation (NLTV).
- obtain **state-of-the-art convergence guarantees**.

☛ **Proposed scheme**: **IML FISTA** a *convergent multilevel inexact and inertial proximal gradient algorithm*:

- prox_g is explicit.
- $\mathrm{prox}_{\varphi \circ \mathrm{L}}$ is not known under closed form.

Inexact FISTA for solving $\min_{\mathbf{x}} f(\mathbf{x}) + \varphi(\mathrm{L}\mathbf{x})$

Inexact FISTA [Aujol, Dossal, 2015]:

$$\mathrm{x}^{[k+1]} \approx_{\epsilon_k} \mathrm{prox}_{\tau\varphi\circ\mathrm{L}}\left(\mathrm{y}^{[k]} - \tau\nabla f(\mathrm{y}^{[k]}) + e_k\right)$$

$$\mathrm{y}^{[k+1]} = \mathrm{x}^{[k+1]} + \alpha_k(\mathrm{x}^{[k+1]} - \mathrm{x}^{[k]})$$

where α_k is chosen with $t_{k+1} = \left(\frac{k+a}{a}\right)^d$, $\alpha_k = \frac{t_k - 1}{t_{k+1}}$.

Contribution: update $\mathrm{y}^{[k]}$ through a multilevel step.

- How to construct such multilevel update ?
- How to guarantee convergence ?

Smoothing of F_h and F_H with the Moreau envelope

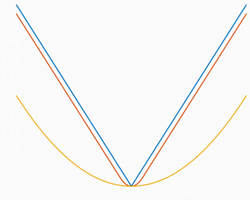
☛ **Moreau envelope of** g_H:

$$^{\gamma}g_H = \inf_{\mathrm{y}\in\mathcal{H}} g_H(\mathrm{y}) + \frac{1}{2\gamma}\|\cdot - \mathrm{y}\|^2$$

☛ **Properties of the Moreau envelope**:

- $\nabla {}^{\gamma}g_H = \gamma^{-1}(\mathsf{Id} - \mathrm{prox}_{\gamma g_H})$
- $\nabla {}^{\gamma}g_H$ γ^{-1} - Lipschitz
- $\nabla \left({}^{\gamma}\varphi_H \circ \mathrm{L}_H\right)(\cdot) = \gamma_H^{-1}\mathrm{L}_H^{*}\left(\mathrm{L}_H \cdot - \mathrm{prox}_{\gamma_H\varphi_H}(\mathrm{L}_H \cdot)\right)$

☛ **Illustration**: Moreau envelope of l_1-norm for $\gamma = 0.1$ and $\gamma = 1$



First order coherence for g non-smooth

Coarse model F_H for non-smooth functions

$$F_H = f_H + ({}^{\gamma_H}\varphi_H \circ \mathrm{L}_H) + \langle \mathrm{v}_h, \cdot \rangle$$

where

$$\begin{aligned}\mathrm{v}_h = I_h^H &\left(\nabla f_h(\mathrm{y}_h) + \nabla({}^{\gamma_h}\varphi_h \circ \mathrm{L}_h)(\mathrm{y}_h)\right) \\ &- \left(\nabla f_H(I_h^H \mathrm{y}_h) + \nabla({}^{\gamma_H}\varphi_H \circ \mathrm{L}_H)(I_h^H \mathrm{y}_h)\right)\end{aligned}$$

Minimization scheme at coarse level:

$$\mathbf{T}_H := \nabla f_H + \nabla({}^{\gamma_H} g_H \circ \mathrm{L}_H)$$

Multilevel algorithm for nonsmooth optimization

```
Set $\mathrm{x}_h^{[0]}, \mathrm{y}_h^{[0]} \in \mathbb{R}^N$, $t_{h,0} = 1$
while Stopping criterion is not met do
    if Descent condition then
        $\mathrm{s}_{H,0}^{[0]} = I_h^H \mathrm{y}_h^{[k]}$ Projection
        $\mathrm{s}_{H,m}^{[k]} = \mathsf{T}_{H,m-1} \circ .. \circ \mathsf{T}_{H,0}(\mathrm{s}_{H,0}^{[0]})$  Coarse minimization
        Set $\bar{\tau}_{h,k} > 0$,
        $\bar{\mathrm{y}}_h^{[k]} = \mathrm{y}_h^{[k]} + \bar{\tau}_{h,k} I_H^h \left(\mathrm{s}_{H,m}^{[k]} - \mathrm{s}_{H,0}^{[0]}\right)$  Coarse update
    else
        $\bar{\mathrm{y}}_h^{[k]} = \mathrm{y}_h^{[k]}$
    end if
    $\mathrm{x}_h^{[k+1]} = \mathsf{T}_i^{\epsilon_{h,k}}(\bar{\mathrm{y}}_h^{[k]})$ Forward-backward step
    $t_{h,k+1} = \left(\frac{k+a}{a}\right)^d$, $\alpha_{h,k} = \frac{t_{h,k}-1}{t_{h,k+1}}$
    $\mathrm{y}_h^{[k+1]} = \mathrm{x}_h^{[k+1]} + \alpha_{h,k}(\mathrm{x}_h^{[k+1]} - \mathrm{x}_h^{[k]})$. Inertial step
end while
```

Convergence of IML FISTA

Multilevel steps interpreted as gradient errors

- ☛ FISTA steps allow errors on the computation of the backward and on the forward steps:

$$\mathrm{x}_h^{[k+1]} \simeq_{\epsilon_{h,k}} \mathrm{prox}_{\tau_h \varphi_h \circ \mathrm{L}_h} \left(\mathrm{y}_h^{[k]} - \tau_h \nabla f_h \left(\mathrm{y}_h^{[k]} \right) + e_{h,k} \right)$$

$$\mathrm{y}_h^{[k+1]} = \mathrm{x}_h^{[k+1]} + \alpha_{h,k} (\mathrm{x}_h^{[k+1]} - \mathrm{x}_h^{[k]})$$

- ☛ Rewriting coarse corrections:

$$e_{h,k} = \tau_h \Big(\nabla f_h (\mathrm{y}_h^{[k]}) - \nabla f_h (\bar{\mathrm{y}}_h^{[k]} + \frac{\bar{\tau}_{h,k}}{\tau_h} I_H^h (\mathrm{s}_{H,m}^{[k]} - \mathrm{s}_{H,0}^{[0]}) \Big)$$

- ☛ Multilevel steps = bounded errors on the gradient

Convergence analysis

Lemma *(Coarse corrections are finite)*[Lauga et al, 2024]
Let β_h and β_H be the Lipschitz constants of f_h and f_H, respectively.
Assume that we compute at most p coarse corrections.
Let $\tau_h, \tau_H \in (0, +\infty)$ be the step sizes taken at fine and coarse levels, respectively.
Assume that $\tau_H < \beta_H^{-1}$ and that $\tau_h < \beta_h^{-1}$ and denote $\bar{\tau}_h = \sup_k \bar{\tau}_{h,k}$. Then the sequence $(e_{h,k})_{k\in\mathbb{N}}$ in $\mathbb{R}^{N_h}$ generated by IML FISTA is defined as:

$$e_{h,k} = \tau_h \left(\nabla f_h(\mathrm{y}_h^{[k]}) - \nabla f_h(\bar{\mathrm{y}}_h^{[k]} + (\tau_h)^{-1} \bar{\tau}_{h,k} I_H^h(\mathrm{s}_{H,m}^{[k]} - \mathrm{s}_{H,0}^{[0]}) \right),$$

if a coarse correction has been computed, and $e_{h,k} = 0$ otherwise.
This sequence is such that $\sum_{k\in\mathbb{N}} k\|e_{h,k}\| < +\infty$.

Inexact proximal step

The ϵ-subdifferential of g at $z \in \operatorname{dom} g$ is defined as:

$$\partial_\epsilon g(z) = \{y \in \mathbb{R}^N \mid g(x) \geqslant g(z) + \langle x - z, y\rangle - \epsilon, \forall x \in \mathbb{R}^N\}.$$

Type 0 approximation [Combettes, Wajs, 2005]

$z \in \mathbb{R}^N$ is a type 0 approximation of $\operatorname{prox}_{\gamma g}(y)$ with precision ϵ, and we write $z \simeq_0 \operatorname{prox}_{\gamma g}(y)$, if and only if $\|z - \operatorname{prox}_{\gamma g}(y)\| \leqslant \sqrt{2\gamma\epsilon}$.

Type 1 approximation [Villa et al., 2013]

$z \in \mathbb{R}^N$ is a type 1 approximation of $\operatorname{prox}_{\gamma g}(y)$ ith precision ϵ, and we write $z \simeq_1 \operatorname{prox}_{\gamma g}(y)$, if and only if $0 \in \partial_\epsilon \left(g(z) + \frac{1}{2\gamma}\|z - y\|^2\right)$.

Type 2 approximation [Villa et al., 2013]

$z \in \mathbb{R}^N$ is a type 2 approximation of $\operatorname{prox}_{\gamma g}(y)$ with precision ϵ, and we write $z \simeq_2 \operatorname{prox}_{\gamma g}(y)$, if and only if $\frac{y-z}{\gamma} \in \partial_\epsilon g(z)$.

Inexact proximity operator step

- ☛ At each iteration of fine level minimization we need to compute

$$\mathrm{prox}_{\gamma\varphi_h\circ \mathrm{L}_h}(\mathrm{x}) = \mathrm{x} - \mathrm{L}_h^* \widehat{\mathrm{u}}$$

with:

$$\widehat{\mathrm{u}} \in \underset{\mathrm{u}\in\mathbb{R}^K}{\mathrm{argmin}}\ \frac{1}{2}\|\mathrm{L}_h^*\mathrm{u} - \mathrm{x}\|^2 + \gamma\varphi_h^*(\mathrm{u})$$

which can be solved iteratively with accuracy ϵ so that:

$$\mathrm{x} - \mathrm{L}_h^*\widehat{\mathrm{u}}_\epsilon \simeq_\epsilon \mathrm{prox}_{\gamma\varphi_h\circ \mathrm{L}_h}(\mathrm{x})$$

- ☛ Equivalent to:

$$\frac{\mathrm{L}_h^*\widehat{\mathrm{u}}_\epsilon}{\gamma} \in \partial_\epsilon \left(\varphi_h \circ \mathrm{L}_h\right)\left(\mathrm{x} - \mathrm{L}_h^*\widehat{\mathrm{u}}_\epsilon\right)$$

Inexact proximity operator step

- ☛ At each iteration of fine level minimization we need to compute

$$\mathrm{prox}_{\gamma\varphi_h \circ \mathrm{L}_h}(\mathrm{x}) = \mathrm{x} - \mathrm{L}_h^* \widehat{\mathrm{u}}$$

with:

$$\widehat{\mathrm{u}} \in \underset{\mathrm{u} \in \mathbb{R}^K}{\mathrm{argmin}} \ \frac{1}{2} \|\mathrm{L}_h^* \mathrm{u} - \mathrm{x}\|^2 + \gamma \varphi_h^*(\mathrm{u})$$

which can be solved iteratively with accuracy ϵ so that:

$$\mathrm{x} - \mathrm{L}_h^* \widehat{\mathrm{u}}_\epsilon \simeq_\epsilon \mathrm{prox}_{\gamma\varphi_h \circ \mathrm{L}_h}(\mathrm{x})$$

- ☛ Equivalent to:

$$\frac{\mathrm{L}_h^* \widehat{\mathrm{u}}_\epsilon}{\gamma} \in \partial_\epsilon \left(\varphi_h \circ \mathrm{L}_h\right) \left(\mathrm{x} - \mathrm{L}_h^* \widehat{\mathrm{u}}_\epsilon\right)$$

$\Rightarrow$ **Type 2 approximation**

Convergence analysis

Theorem [Lauga et al, 2024]

Considering $\forall k \in \mathbb{N}^*,\ \alpha_{h,k} = 0$ and the sequence $(\epsilon_{h,k})_{k\in\mathbb{N}}$ is such that $\sum_{k\in\mathbb{N}} \sqrt{\|\epsilon_{h,k}\|} < +\infty$. Set $\mathrm{x}_h^{[0]} \in \mathbb{R}^{N_h}$ and choosing approximation of Type 0, the sequence $(\mathrm{x}_h^{[k]})_{k\in\mathbb{N}}$ generated by IML FISTA converges to a minimizer of F_h.

Convergence analysis

Theorem [Lauga et al, 2024]

Let $\forall k \in \mathbb{N}^*,\ t_{h,k+1} = \left(\frac{k+a}{a}\right)^d$, with (a, d) satisfying the conditions in [Aujol, Dossal, 2015 – Definition 3.1], and that the assumptions of Lemma 28 hold. Moreover, if we assume that:

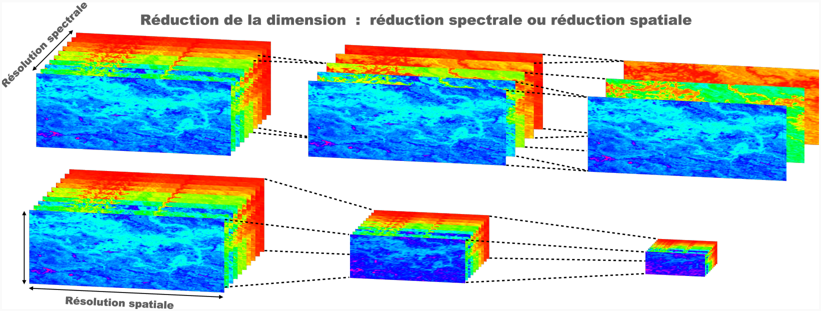
- $\sum_{k=1}^{+\infty} k^d \sqrt{\epsilon_{h,k}} < +\infty$ in the case of Type 1 approximation,
- $\sum_{k=1}^{+\infty} k^{2d} \epsilon_{h,k} < +\infty$ in the case of Type 2 approximation.

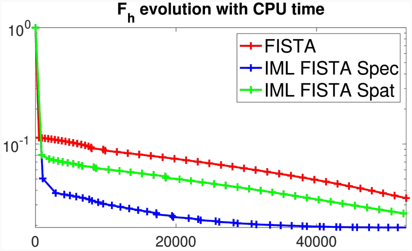
Let $(\mathrm{x}_h^{[k]})_{k\in\mathbb{N}}$ the sequence generated by IML FISTA, then

- The sequence $(k^{2d}\left(F_h(\mathrm{x}_h^{[k]}) - F_h(\mathrm{x}^*)\right))_{k\in\mathbb{N}}$ belongs to $\ell_\infty(\mathbb{N})$.
- The sequence $(\mathrm{x}_{h,k})_{k\in\mathbb{N}}$ converges to a minimizer of F_h.

Numerical experiments

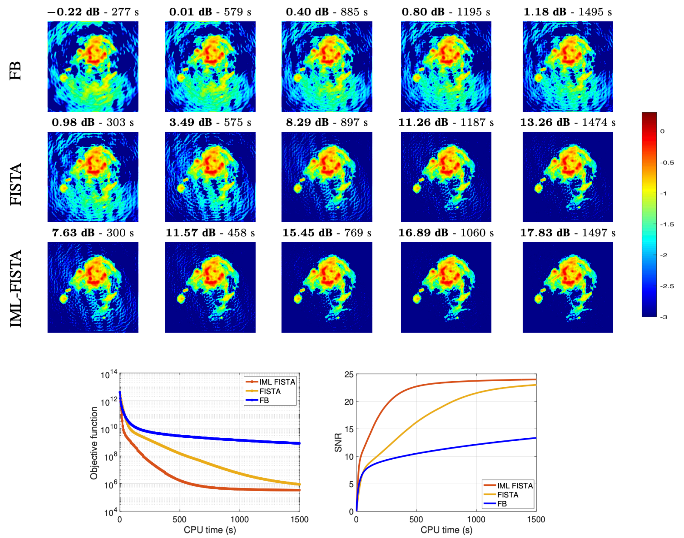
Restoration of blurred hyperspectral images with missing pixels





[G. Lauga et al., 2024, *SIAM J. Imaging Sci.*]

Numerical experiments in radio-interferometric imaging



[G. Lauga et al., 2024, *EUSIPCO*]

Partial conclusions

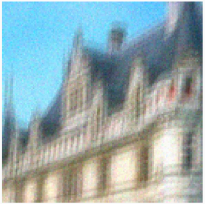
- Unifying and extended convergence guarantees for IML FB.
- Convergent IML FISTA.
- IML FISTA much faster than FISTA for large scale problems.

Future works:

- Deeper analysis of the design of I_h^H and I_H^h.
- Improve the rule to go from fine to a coarser step.
- What about multilevel PnP and unfolded networks ?

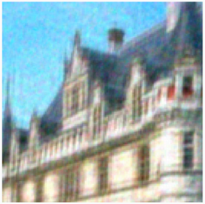
Perspective: Towards deep learning

Original



Degraded

SNR = 13.4 dB



Tikhonov

SNR = 16.4 dB

DTT

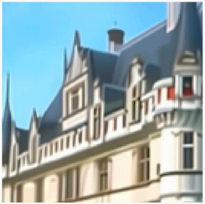
SNR = 16.6 dB

TV

SNR = 18.8 dB

NLTV

SNR = 19.4 dB



PnP-DRUnet

SNR = 20.0 dB

PnP-ScCP

SNR = 20.2 dB[4]

Perspective: Towards deep learning

- ➡ [1922] **Maximum likelihood** (Fisher).

$$\widehat{x} \in \underset{x}{\operatorname{Argmin}} \frac{1}{2}\|Ax - z\|_2^2 = (A^*A)^{-1}A^*z$$

- ➡ [1963] **Regularization** (Tikhonov, Huber)

$$\widehat{x} \in \underset{x}{\operatorname{Argmin}} \frac{1}{2}\|Ax - z\|_2^2 + \theta\|Lx\|_2^2 \qquad \text{avec} \quad \theta > 0$$

- ➡ [2000] **Sparsity** (Donoho, Daubechies-Defrise-DeMol,...)

$$\widehat{x} \in \underset{x}{\operatorname{Argmin}} \frac{1}{2}\|Ax - z\|_2^2 + \theta\|Lx\|_\star$$

- ➡ [2010] **"End to end" neural networks**

$$\widehat{x} = NN_\Theta(z)$$

- ➡ [2020] **Plug-and-Play**

$$0 \in A^*(A\widehat{x} - z) + \mathbf{B}(\widehat{x})$$

Make the algorithm robust and faster with multilevel strategy

Multilevel Plug-and-play

- **FB-PnP**:

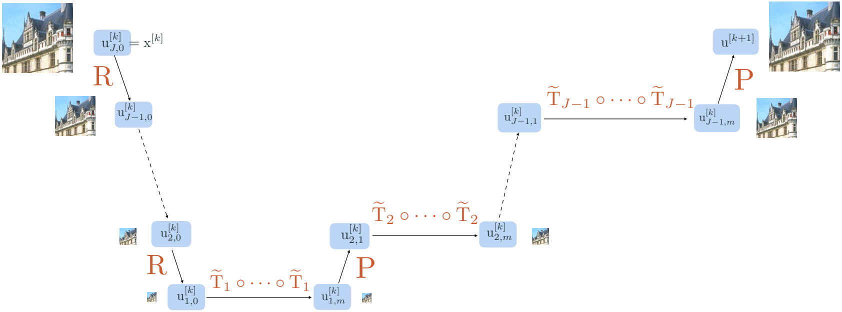
$$x^{[k+1]} = \boxed{d_\Theta}\left(x^{[k]} - \gamma A^\top (Ax^{[k]} - z) \right)$$

- **Multilevel FB-PnP**:

$$\begin{aligned} u^{[k]} &= \mathrm{ML}(x^{[k]}) \\ x^{[k+1]} &= \boxed{d_\Theta}\left(u^{[k]} - \gamma A^\top (Au^{[k]} - z) \right) \end{aligned}$$

- We denote: $\mathbf{T}(u^{[k]}) = d_\Theta(u^{[k]} - \gamma A^\top (Au^{[k]} - z)$
- Multilevel framework when $d_\Theta = \mathrm{prox}_f$: [Lauga, Riccietti, Pustelnik, Goncalves, 2024]

ML-step



$\mathrm{u}^{[k]}_{J,0} = \mathrm{x}^{[k]}$

R

$\mathrm{u}^{[k]}_{J-1,0}$

$\mathrm{u}^{[k]}_{2,0}$

R

$\mathrm{u}^{[k]}_{1,0}$

$\widetilde{\mathrm{T}}_1 \circ \cdots \circ \widetilde{\mathrm{T}}_1$

$\mathrm{u}^{[k]}_{1,m}$

P

$\mathrm{u}^{[k]}_{2,1}$

$\widetilde{\mathrm{T}}_2 \circ \cdots \circ \widetilde{\mathrm{T}}_2$

$\mathrm{u}^{[k]}_{2,m}$

$\mathrm{u}^{[k]}_{J-1,1}$

$\widetilde{\mathrm{T}}_{J-1} \circ \cdots \circ \widetilde{\mathrm{T}}_{J-1}$

$\mathrm{u}^{[k]}_{J-1,m}$

P

$\mathrm{u}^{[k+1]}$

Main ingredients

- R: restriction operator
- P: prolongation operator
- $\widetilde{\mathrm{T}}_j \circ \ldots \circ \widetilde{\mathrm{T}}_j$: coarser updates to insure first order coherence

Numerical experiments

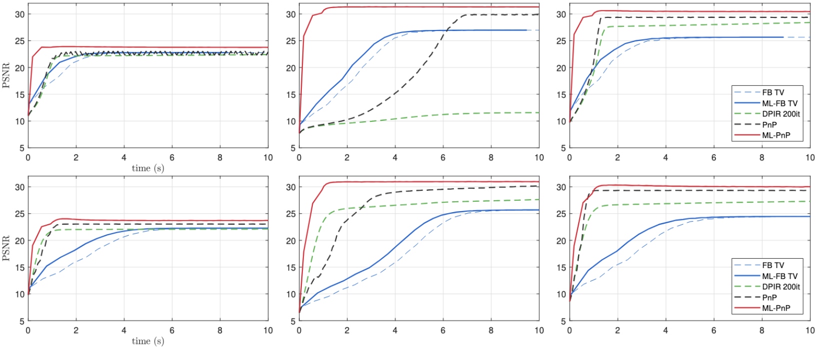
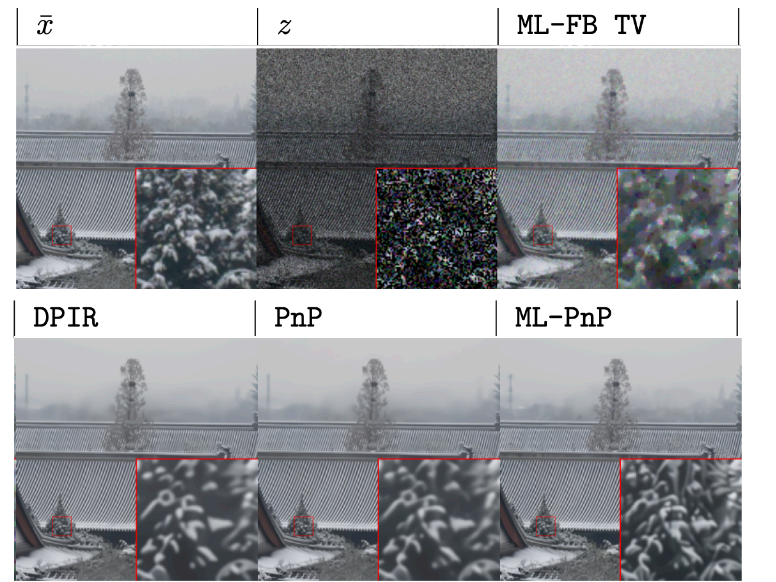
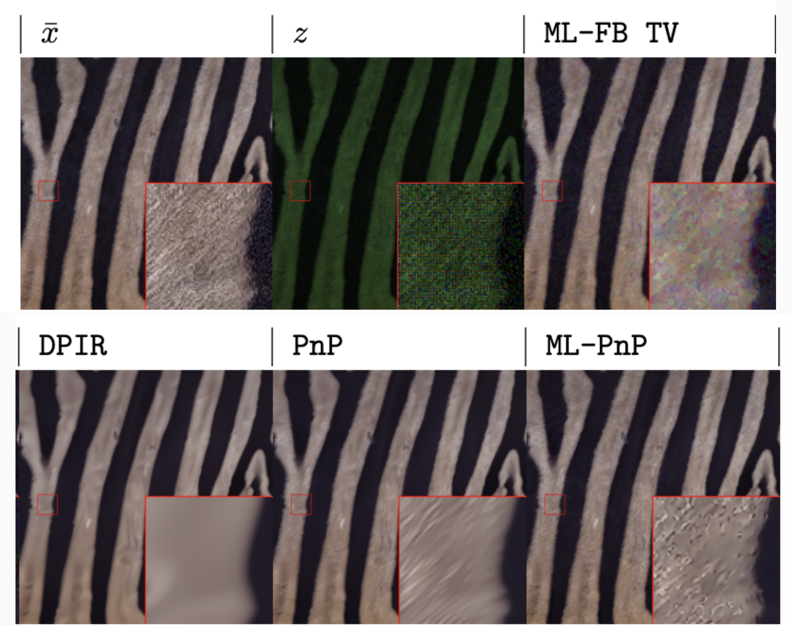


Figure 3: Reconstruction performance of multilevel algorithms and state of the art measured in PSNR with respect to time. First row presents results for the inpainting problem with 50% missing pixels, second row presents results for the demosaïcing problem. Each column is respectively associated with the images in the rows of Fig. 4 (inpainting) and Fig. 5 (demosaicing).

Numerical experiments



Numerical experiments



References

- G. Lauga, E. Riccietti, N. Pustelnik, and P. Goncalves.
  IML FISTA: A Multilevel Framework for Inexact and Inertial Forward-Backward. Application to Image Restoration , SIAM Journal on Imaging Sciences 17 (3), 1347-1376, 2024.

- G. Lauga, A. Repetti, E. Riccietti, N. Pustelnik, P. Gonçalves, Y.Wiaux,
  A multilevel framework for accelerating uSARA in radio-interferometric imaging, EUSIPCO, Lyon, 2024.

- G. Lauga, E. Riccietti, N. Pustelnik, P. Gonçalves,
  Méthodes multi-niveaux pour la restauration d'images hyperspectrales, Colloque GRETSI, September 2023.

- N. Laurent, E. Riccietti, J. Tachella, N. Pustelnik,
  Algorithme multiniveaux hybride pour la restauration d'images, Colloque GRETSI, August 2025.

- N. Laurent, J. Tachella, E. Riccietti, N. Pustelnik,
  Multilevel Plug-and-Play Image Restoration, submitted 2025.