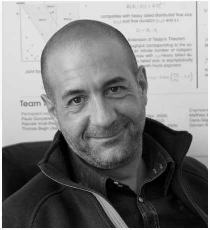
IML FISTA: A Multilevel Framework for Inexact and Inertial Forward-Backward. Application to Large Scale Image Restoration

Nelly Pustelnik

ENSL, CNRS, Laboratoire de physique, F-69342 Lyon, France

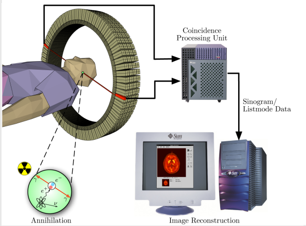
Collaboration



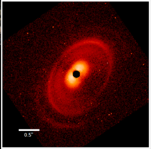
Guillaume LAUGA   Elisa RICCIETTI   Paulo GONCALVES

Inverse problems: various applications

Medical imaging



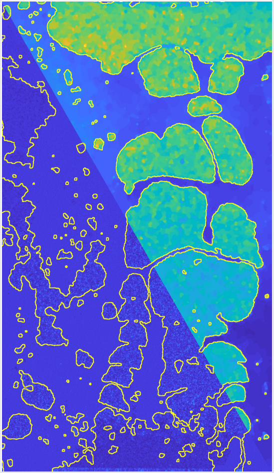
Astronomy



@ L. Denneulin

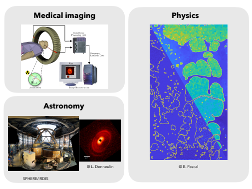
SPHERE/IRDIS

Physics



@ B. Pascal

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**

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

☛ **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 models $\widehat{x} = d_\Theta(z)$

➡ [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,...)

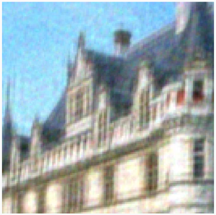
$$\widehat{x} \in \operatorname*{Argmin}_x \frac{1}{2}\|Ax - z\|_2^2 + \theta\|Lx\|_\star$$

Summary of inverse problems in imaging

Original

Degraded

SNR = 13.4 dB

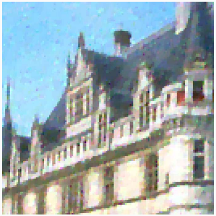


Tikhonov

SNR = 16.4 dB

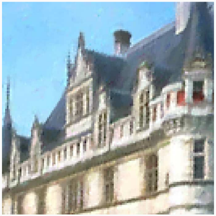
DTT

SNR = 16.6 dB



TV

SNR = 18.8 dB



NLTV

SNR = 19.4 dB

Une méthodologie

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

Interprétation bayésienne MAP (vraisemblance / a priori)

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ \psi(\mathrm{Ax};\mathrm{z}) + \varphi(\mathrm{Lx})$$

Iterative scheme

$$\mathrm{x}_{k+1} = \Phi(\mathrm{x}_k)$$

Iterative scheme

➡ **Minimization problem** :

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

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

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

$$(\forall k \in \mathbb{N}) \qquad \mathrm{x}_{k+1} = \mathbf{\Phi}(\mathrm{x}_k),$$

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

Iterative scheme

➡ **Minimization problem** :

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\; f(\mathrm{x}) + \varphi(\mathrm{Lx})$$

- **Require the computation of** $\mathrm{prox}_{\varphi(\mathrm{L}\cdot)}$. **Few closed form.**
- **Reformulation in the dual**: $\min_{\mathrm{w}\in\mathcal{G}} f^*(-\mathrm{L}^*\mathrm{w}) + \varphi^*(\mathrm{w})$,
- **Primal-dual algorithms**: $\min_{\mathrm{x}} f(\mathrm{x}) + \widetilde{f}(\mathrm{x}) + \varphi(\mathrm{Lx})$,

  [Condat,2013][Vũ,2013] [Chambolle-Pock,2011]

  $\rightarrow$ with f ν-gradient Lipschitz.

Hyperparameters setting: $\tau > 0$, $\gamma > 0$, such that $\frac{1}{\tau} - \gamma\|\mathrm{L}\|^2 > \frac{\nu}{2}$

For $k = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} \mathrm{w}^{[k+1]} = \mathrm{prox}_{\tau\widetilde{f}}\big(\mathrm{w}^{[k]} - \tau\nabla f(\mathrm{w}^{[k]}) - \tau\mathrm{L}^*\mathrm{x}^{[k]}\big) \\ \mathrm{x}^{[k+1]} = \mathrm{prox}_{\gamma\varphi^*}\big(\mathrm{x}^{[k]} + \gamma\mathrm{L}(2\mathrm{w}^{[k+1]} - \mathrm{w}^{[k]})\big) \end{array} \right.$$

Large scale

Main goal : provide acceleration for high dimensional problems

High dimensional problems $\rightarrow$ high computation time.

Main goal : provide acceleration for high dimensional problems

High dimensional problems $\rightarrow$ 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 and developed in particular for non-smooth optimization in [Parpas, 2017], our inspiration.

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)

$\rightarrow$ Successful attempts of accelerating minimization in imaging.

$\rightarrow$ 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: $\mathrm{x}_{k+1} = \Phi(\mathrm{x}_k) = \mathrm{x}_k - D_k$

• If f and g are differentiable: Gradient descent

$$D_k = \tau_k(\nabla f(\mathrm{x}_k) + \nabla g(\mathrm{x}_k))$$

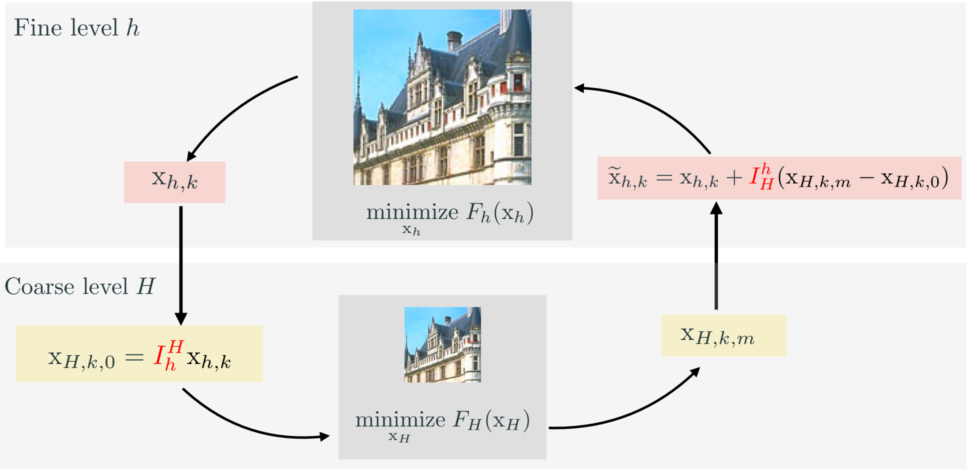
• If g is not differentiable: Proximal gradient descent

$$D_k = \mathrm{x}_k - \mathrm{prox}_{\tau_k g}\left(\mathrm{x}_k - \tau_k \nabla f(\mathrm{x}_k)\right)$$

Multilevel smooth optimization

Goal: Exploit hierarchy of approximations of the objective function.

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



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:

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

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 $y_h \in \mathbb{R}^{N_h}$ if the following equality holds:

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

☛ **Impact**: Coherence up to order one in the neighbourhood of the current iterates $y_h = y_{h,k}$.

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 $y_h \in \mathbb{R}^{N_h}$ as:

$$F_H = f_H + g_{H,\gamma_H} + \langle v_H, \cdot \rangle,$$

where

$$v_H = I_h^H \left(\nabla f_h(y_h) + \nabla g_{h,\gamma_h}(y_h) \right) - \left(\nabla f_H(I_h^H y_h) + \nabla g_{H,\gamma_H}(I_h^H y_h) \right).$$

Remarks:

- Adding the linear term $\langle v_H, \cdot \rangle$ 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

Coarse model F_H for non-smooth functions

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

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

where

$$v_H = I_h^H \left(\nabla f_h(y_h) + \nabla g_{h,\gamma_h}(y_h) \right) - \left(\nabla f_H(I_h^H y_h) + \nabla g_{H,\gamma_H}(I_h^H y_h) \right).$$

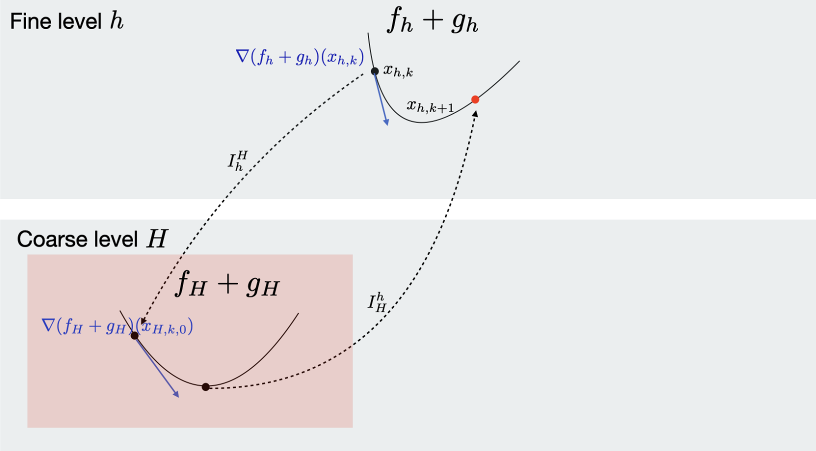
Lemma If F_H is given by definition (1), it necessarily verifies the first order coherence.

Proof.

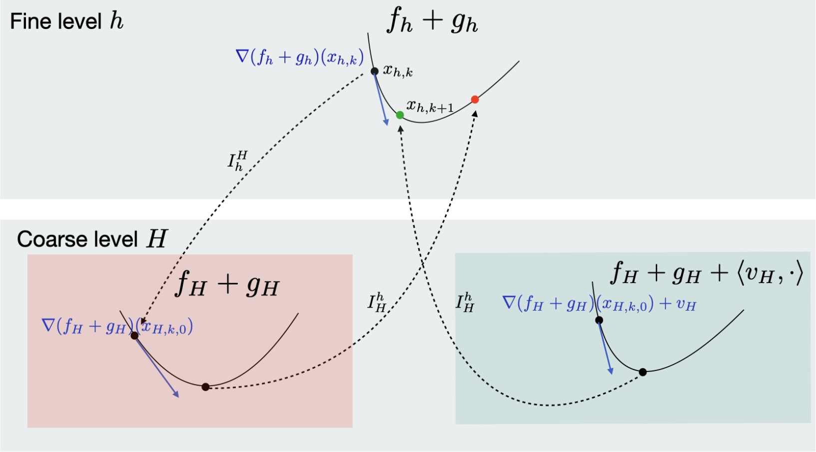
Considering the gradient of the coarse model F_H and combining it with the definition of v_H, yields

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

Design of F_H: First order coherence



Design of F_H: First order coherence



Design of Φ_H: Iterations at the coarse scale

Coarse model decrease

Let $(\Phi_{H,\ell})_{\ell\in\mathbb{N}}$ be a sequence of operators such that there exists an integer $m > 0$ that guarantees that if

$$\mathrm{x}_{H,m} = \Phi_{H,m-1} \circ \ldots \circ \Phi_{H,0}(\mathrm{x}_{H,0})$$

then

$$F_H(s_{H,m}) \leqslant F_H(\mathrm{x}_{H,0}).$$

Moreover, $\mathrm{x}_{H,m} - \mathrm{x}_{H,0}$ is bounded.

- ☛ $\Phi_{H,\bullet}$ aims to build a sequence producing a sufficient decrease of F_H after m iterations.
- ☛ Typical choices for $\Phi_{H,\ell}$: gradient descent step, inertial gradient descent step, forward-backward step or inertial forward-backward step (comparisons in [Lauga et al. 2022]).

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

Definition $I_h^H : \mathbb{R}^{N_h} \to \mathbb{R}^{N_H}$ (transfer from fine to coarse scales) and $I_H^h : \mathbb{R}^{N_H} \to \mathbb{R}^{N_h}$ (transfer from coarse to fine scales) are *coherent information transfer (CIT)* operators, if there exists $\nu > 0$ such that:

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

- most standard : dyadic decimated weighted operator [Briggs 2000].
- particular case of squared grids reads:

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

IML FB: Multilevel algorithm for nonsmooth optimization

Set $x_{h,0}, y_{h,0} \in \mathbb{R}^N$, $t_{h,0} = 1$
while Stopping criterion is not met **do**
  if Descent condition and $r < p$ **then**
    $r = r + 1$,
    $s_{H,k,0} = I_h^H x_{h,k}$ Projection
    $s_{H,k,m} = \Phi_{H,m-1} \circ .. \circ \Phi_{H,0}(s_{H,k,0})$ Coarse minimization
    Set $\bar{\tau}_{h,k} > 0$,
    $\bar{x}_{h,k} = x_{h,k} + \bar{\tau}_{h,k} I_H^h \left(s_{H,k,m} - s_{H,k,0}\right)$ Coarse step update
  else
    $\bar{x}_{h,k} = x_{h,k}$
  end if
  $x_{h,k+1} = \Phi_{h,k}(\bar{x}_{h,k})$ Forward-Backward step
end while

Convergence analysis

Lemma *(Descent direction for the fine level smoothed function)*.
Let us assume that I_h^H and I_H^h are CIT operators and that F_H satisfies Definition (1) and $\Phi_{H,\bullet}$ allows a decrease of the coarse model. Then, $I_H^h(s_{H,m} - s_{H,0})$ is a descent direction for $f_h + g_{h,\gamma_h}$.

Proof.
Set $\mathrm{x}_h \in \mathbb{R}^{N_h}$ and let us define $p_H := s_{H,m} - s_{H,0}$.
Recall that $s_{H,0} = I_h^H \mathrm{x}_h$.
From the definition of descent direction we have that:

$$\langle p_H, \nabla F_H(s_{H,0}) \rangle \leqslant 0.$$

By the first order coherence and imposing $I_h^H = \nu \left(I_H^h\right)^T$ we obtain

$$\begin{aligned} \langle p_H, \nabla F_H(s_{H,0}) \rangle &= \langle p_H, I_h^H \nabla (f_h + g_{h,\gamma_h})(\mathrm{x}_h) \rangle \\ &= \nu \langle I_H^h(p_H), \nabla (f_h + g_{h,\gamma_h})(\mathrm{x}_h) \rangle \leqslant 0. \end{aligned}$$

Convergence analysis

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

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

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

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

□

Convergence analysis

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

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

- 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]:

$$x_{k+1} \approx_{\epsilon_k} \mathrm{prox}_{\tau\varphi\circ\mathrm{L}} \left(y_k - \tau\nabla f(y_k) + e_k\right)$$
$$y_{k+1} = x_{k+1} + \alpha_k(x_{k+1} - x_k)$$

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 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

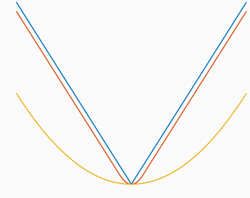
☛ **Moreau envelope of** g_H:

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

☛ **Properties of the Moreau envelope**:

- $\nabla^{\gamma} g_H = \gamma^{-1}(\mathrm{Id} - \mathrm{prox}_{\gamma g_H})$
- $\nabla^{\gamma} g_H$ 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

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

where

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

Minimization scheme at coarse level:

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

Multilevel algorithm for nonsmooth optimization

Set $x_{h,0}, y_{h,0} \in \mathbb{R}^N$, $t_{h,0} = 1$
while Stopping criterion is not met **do**
    if Descent condition and $r < p$ **then**
        $r = r + 1$,
        $s_{H,k,0} = I_h^H y_{h,k}$ Projection
        $s_{H,k,m} = \Phi_{H,m-1} \circ .. \circ \Phi_{H,0}(s_{H,k,0})$ Coarse minimization
        Set $\bar{\tau}_{h,k} > 0$,
        $\bar{y}_{h,k} = y_{h,k} +$
        $\bar{\tau}_{h,k} I_H^h \left(s_{H,k,m} - s_{H,k,0}\right)$ Coarse step update whose size is set by
    else
        $\bar{y}_{h,k} = y_{h,k}$
    end if
    $x_{h,k+1} = \Phi_i^{\epsilon_{h,k}}(\bar{y}_{h,k})$ Forward-backward step
    $t_{h,k+1} = \left(\frac{k+a}{a}\right)^d$, $\alpha_{h,k} = \frac{t_{h,k}-1}{t_{h,k+1}}$
    $y_{h,k+1} = x_{h,k+1} + \alpha_{h,k}(x_{h,k+1} - x_{h,k})$. Inertial step

Convergence analysis

Lemma *(Fine level decrease)*. Let assume that I_h^H and I_H^h are CIT operators and that F_H satisfies Definition (1) and $\Phi_{H,\bullet}$ allows a decrease of the coarse model. The iterations of IML FISTA ensure:

$$F_h(y_h + \bar{\tau} I_H^h(x_{H,m} - x_{H,0})) \leqslant F_h(y_h) + \eta\gamma_h$$

☛ Bound on one multilevel step but not enough for global convergence guarantees.

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:

$$x_{h,k+1} \simeq_{\epsilon_{h,k}} \mathrm{prox}_{\tau_h \varphi_h \circ \mathrm{L}_h} \left(y_{h,k} - \tau_h \nabla f_h \left(y_{h,k} \right) + e_{h,k} \right)$$
$$y_{h,k+1} = x_{h,k+1} + \alpha_{h,k}(x_{h,k+1} - x_{h,k})$$

☛ Rewriting coarse corrections:

$$e_{h,k} = \tau_h \Big(\nabla f_h(y_{h,k}) - \nabla f_h(\bar{y}_{h,k}) + \frac{\bar{\tau}_{h,k}}{\tau_h} I_H^h (s_{H,k,m} - s_{H,k,0}) \Big)$$

☛ Multilevel steps = bounded errors on the gradient

Convergence analysis

Lemma *(Coarse corrections are finite)*
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:

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

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 $z \in \text{dom } g$ 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 $\text{prox}_{\gamma g}(y)$ with precision ϵ, and we write $z \simeq_0 \text{prox}_{\gamma g}(y)$, if and only if $\|z - \text{prox}_{\gamma g}(y)\| \leqslant \sqrt{2\gamma\epsilon}$.

Type 1 approximation [Villa et al., 2013]

$z \in \mathbb{R}^N$ is a type 1 approximation of $\text{prox}_{\gamma g}(y)$ ith precision ϵ, and we write $z \simeq_1 \text{prox}_{\gamma g}(y)$, 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 $\text{prox}_{\gamma g}(y)$ with precision ϵ, and we write $z \simeq_2 \text{prox}_{\gamma g}(y)$, 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

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

with:

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

which can be solved iteratively with accuracy ϵ so that:

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

- Equivalent to:

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

Inexact proximity operator step

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

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

with:

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

which can be solved iteratively with accuracy ϵ so that:

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

- Equivalent to:

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

$\Rightarrow$ **Type 2 approximation**

Convergence analysis

Theorem

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 $x_{h,0} \in \mathbb{R}^{N_h}$ and choosing approximation of Type 0, the sequence $(x_{h,k})_{k\in\mathbb{N}}$ generated by IML FISTA converges to a minimizer of F_h.

Convergence analysis

Theorem

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 31 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 $(x_{h,k})_{k\in\mathbb{N}}$ the sequence generated by IML FISTA, then

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

Numerical experiments

Problem formulation:

- $N = 2048 \times 2048 \times 3$
- A: inpainting operator: 90% along all channels
- $\varphi \circ \mathrm{L}$: Non Local Total Variation penalization
- Gaussian noise $\sigma(\text{noise}) = 5e-2$
- $\theta = 3e-2$
- Parameters: $\tau = 1$, $\gamma = 1.1$, $l = 5$, $p = 2$, $m = 5$, $\theta_H = \lambda/4$

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

Image reconstruction with NLTV prior

Image reconstruction with NLTV prior

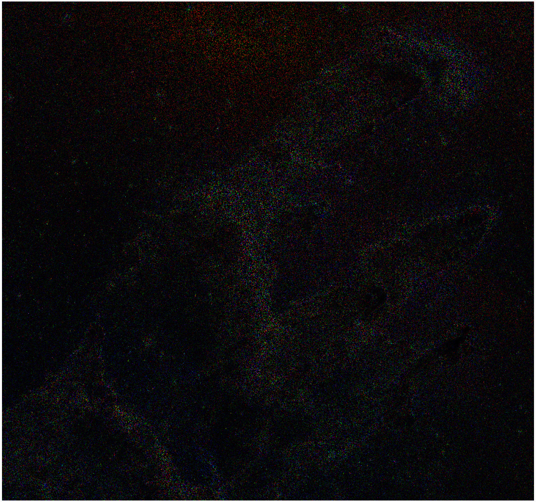
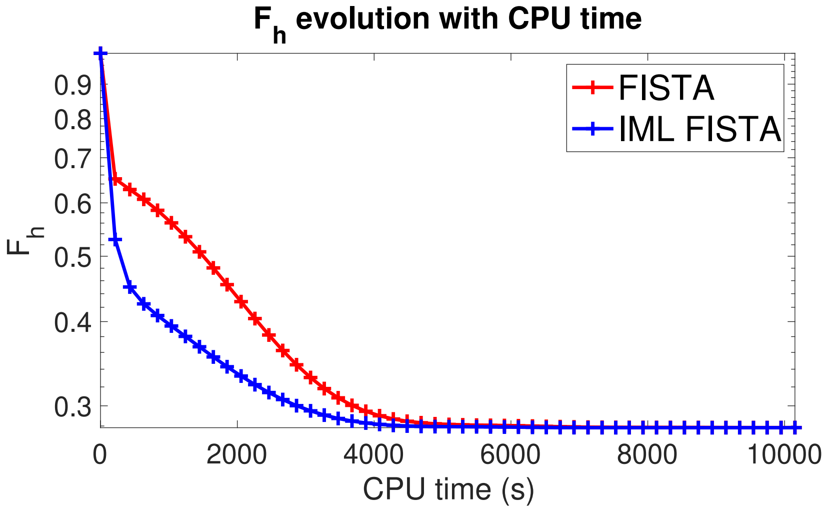
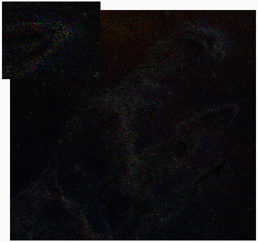


Image reconstruction with NLTV prior

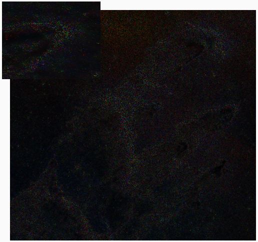
Evolution of F_h for a $N_h = 2048 \times 2048 \times 3$ image



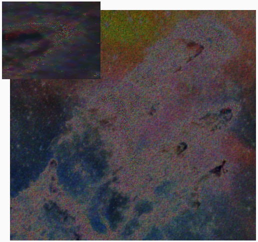
Reconstruction after 2 iterations with NLTV



z



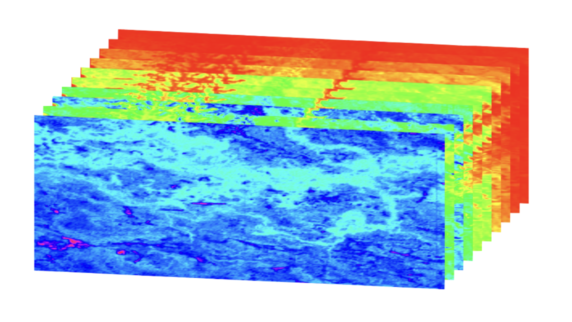
x_2 FISTA



x_2 IML FISTA

Numerical experiments on hyperspectral images

Hyperspectral images



- $x^{(i,b)} = x^{(i_1,i_2,b)}$
  pixel $i = (i_1, i_2) \in \{1, \ldots, N_r\} \times \{1, \ldots, N_c\}$
  of the band $b \in \{1, \ldots, L\}$ of the hypercube x.
- $N = N_r \times N_c \times L$

Inpainting problem for Hyperspectral Images

Degradation model: $z = \mathrm{A}\bar{x} + \epsilon$. Associated minimization problem:

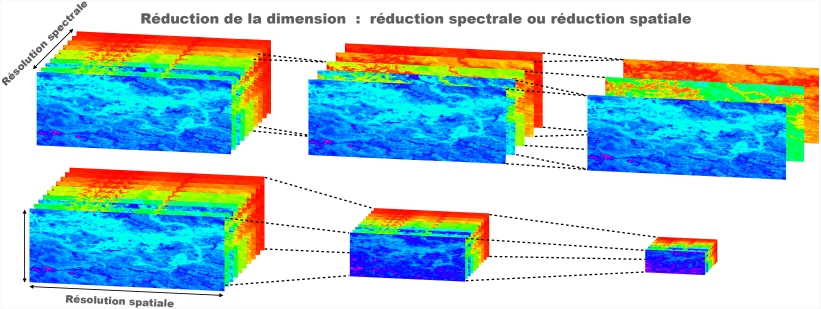
$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}\in\mathbb{R}^N}{\mathrm{Argmin}} \frac{1}{2}\|\mathrm{Ax} - \mathrm{z}\|_2^2 + \lambda g(\mathrm{Lx})$$

- $\mathrm{A} \in \mathbb{R}^{M\times N}$: inpainting operator which applies independant masks on each spectral band.
- $\mathrm{L} \in \mathbb{R}^{(N\times K\times L)\times N}$ such that:

$$(\mathrm{Lx})^{(i)} = \left[\omega^{(i,j)}(x^{(i,b)} - x^{(j,b)})\right]_{j\in\mathcal{N}_i, 1\leqslant b\leqslant L} \in \mathbb{R}^{\widetilde{K}\times L}$$

$$(\forall \mathrm{x} \in \mathbb{R}^N) \quad g(\mathrm{Lx}) = \sum_{i=1}^{N} \|(\mathrm{Lx})^i\|_*$$

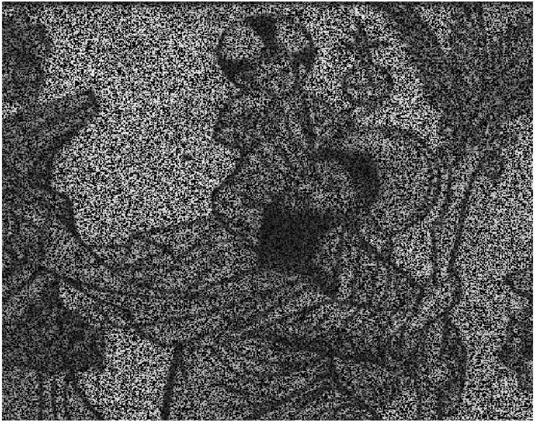
Dimension reduction



Results with Spectral IML FISTA

$\bar{x} \in \mathbb{R}^{512 \times 512 \times 33}$

Results with Spectral IML FISTA



z

Results with Spectral IML FISTA

$x_{\text{end,FISTA}}$

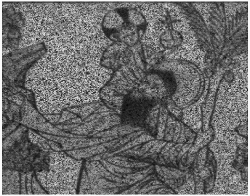
Results with Spectral IML FISTA

$x_{\text{end,IML FISTA}}$

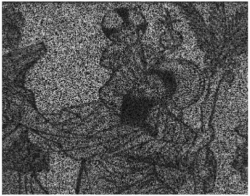
Results with Spectral IML FISTA

$\bar{x}$ (SNR)

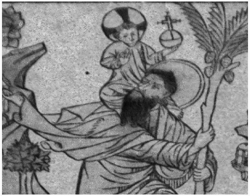
$x_{2,\text{FISTA}}$ (7 dB)



$x_{\text{end,FISTA}}$ (21 dB)



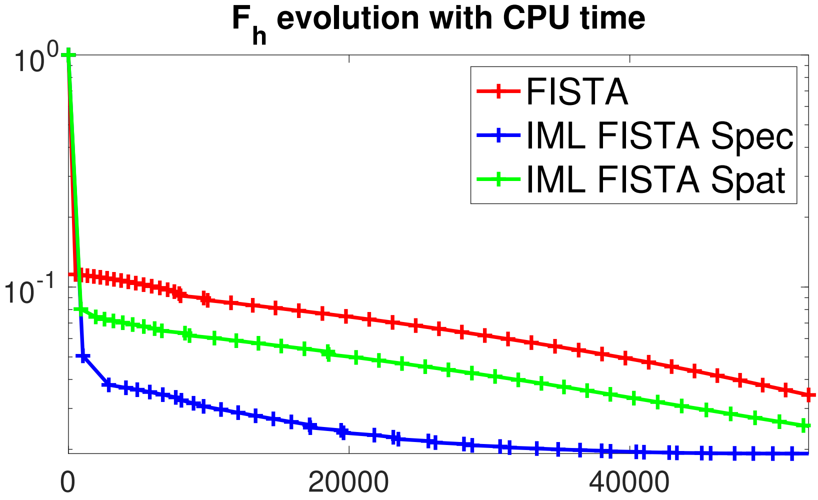
z (3 dB)



$\mathrm{x}_{2,\text{IML FISTA}}$ (19 dB)

$\mathrm{x}_{\text{end,IML FISTA}}$ (35 dB)

Objective function evolution



Partial conclusions

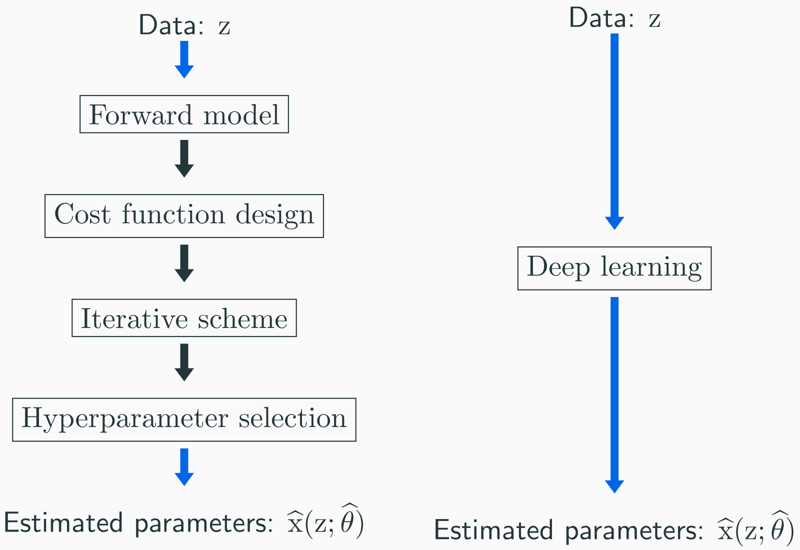
- Unifying and extended convergence guarantees for IML FB.
- Convergent IML FISTA.
- IML FISTA much faster than FISTA for large scale problems.

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 ?

Towards deep learning

Towards deep learning



➡ [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] **Sparisty** (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**

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

Towards deep learning: Plug and Play

Deep learning – *Framework*

- Database : $\mathcal{S} = \big\{ (\overline{\mathrm{x}}_i, \mathrm{z}_i) \in \mathbb{R}^N \times \mathbb{R}^M \;\big|\; i \in \{1, \ldots, \mathbb{I}\} \big\}$
- Prediction function : $\mathrm{NN}_\Theta(\mathrm{z}_i) = \eta^{[K]}\big(W^{[K]} \ldots \eta^{[1]}(W^{[1]}\mathrm{z}_i + b^{[1]}) \ldots + b^{[K]}\big)$

Variational approaches versus Plug and play

- **Forward-Backward**: $\mathrm{x}^{[k+1]} = \mathrm{prox}_{\gamma\theta\|\mathrm{L}\cdot\|_\star}\big(\mathrm{x}^{[k]} - \gamma \mathrm{A}^*(\mathrm{A}\mathrm{x}^{[k]} - \mathrm{z}) \big)$
- **Algo. PnP**: $\mathrm{x}^{[k+1]} = \mathrm{NN}_\Theta\big(\mathrm{x}^{[k]} - \gamma \mathrm{A}^*(\mathrm{A}\mathrm{x}^{[k]} - \mathrm{z}) \big)$

Toward deep learning: PnP and PNN

Deep learning – *Cadre général*

- Database : $\mathcal{S} = \big\{(\overline{\mathrm{x}}_i, \mathrm{z}_i) \in \mathbb{R}^N \times \mathbb{R}^M \;\big|\; i \in \{1, \ldots, \mathbb{I}\}\big\}$
- Prediction function : $\mathrm{NN}_\Theta(\mathrm{z}_i) = \eta^{[K]}\big(W^{[K]} \ldots \eta^{[1]}(W^{[1]}\mathrm{z}_i + b^{[1]}) \ldots + b^{[K]}\big)$

Unfolded schemes – *Informed neural network*

Analysis formulation : $\min_{\mathrm{x}} \frac{1}{2}\|\mathrm{Ax} - \mathrm{z}\|_2^2 + \theta\|\mathrm{Lx}\|_\star$

Toward deep learning: PnP and PNN

Deep learning – *Cadre général*

- Database : $\mathcal{S} = \left\{(\overline{\mathrm{x}}_i, \mathrm{z}_i) \in \mathbb{R}^N \times \mathbb{R}^M \;\middle|\; i \in \{1, \ldots, \mathbb{I}\}\right\}$
- Prediction function : $\mathrm{NN}_\Theta(\mathrm{z}_i) = \eta^{[K]}\big(W^{[K]} \ldots \eta^{[1]}(W^{[1]}\mathrm{z}_i + b^{[1]}) \ldots + b^{[K]}\big)$

Unfolded schemes – *Informed neural network*

Synthesis formulation : $\min\limits_{\mathrm{u}} \frac{1}{2}\|\mathrm{AL}^*\mathrm{u} - \mathrm{z}\|_2^2 + \theta\|\mathrm{u}\|_\star$

Forward-backward :

$$\mathrm{u}^{[k+1]} = \mathrm{prox}_{\gamma\theta\|\cdot\|_\star}\left(\mathrm{u}^{[k]} - \gamma\mathrm{LA}^*(\mathrm{AL}^*\mathrm{u}^{[k]} - \mathrm{z})\right)$$

Toward deep learning: PnP and PNN

Deep learning – *Cadre général*

- Database : $\mathcal{S} = \big\{(\overline{\mathrm{x}}_i, \mathrm{z}_i) \in \mathbb{R}^N \times \mathbb{R}^M \mid i \in \{1, \ldots, \mathbb{I}\}\big\}$
- Prediction function : $\mathrm{NN}_\Theta(\mathrm{z}_i) = \eta^{[K]}\big(W^{[K]} \ldots \eta^{[1]}(W^{[1]}\mathrm{z}_i + b^{[1]}) \ldots + b^{[K]}\big)$

Unfolded schemes – *Informed neural network*

Synthesis formulation : $\min\limits_{\mathrm{u}} \frac{1}{2}\|\mathrm{AL}^*\mathrm{u} - \mathrm{z}\|_2^2 + \theta\|\mathrm{u}\|_\star$

Forward-backward :

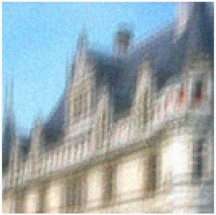
$$\mathrm{u}^{[k+1]} = \mathrm{prox}_{\gamma\theta\|\cdot\|_\star}\Big(\mathrm{u}^{[k]} - \gamma \mathrm{LA}^*(\mathrm{AL}^*\mathrm{u}^{[k]} - \mathrm{z})\Big)$$

PNN: Proximal Neural Network:

$$\mathrm{u}^{[k+1]} = \underbrace{\mathrm{prox}_{\gamma\theta\|\cdot\|_\star}}_{\eta^{[k]}}\Big(\underbrace{\mathrm{Id} - \gamma\mathrm{LA}^*\mathrm{AL}^*}_{\mathrm{W}^{[k]}}\ \mathrm{u}^{[k]} + \underbrace{\gamma\mathrm{LA}^*\mathrm{z}}_{\mathrm{b}^{[k]}}\Big)$$

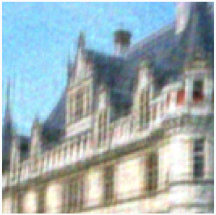
Summary of inverse problems in imaging

Original



Degraded

SNR = 13.4 dB

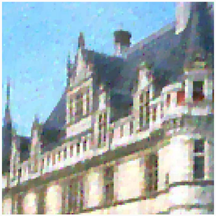


Tikhonov

SNR = 16.4 dB

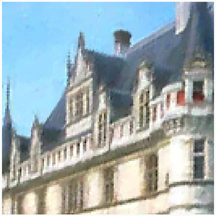
DTT

SNR = 16.6 dB



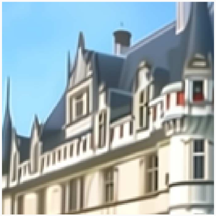
TV

SNR = 18.8 dB



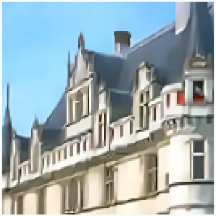
NLTV

SNR = 19.4 dB



PnP-DRUnet

SNR = 20.0 dB



PnP-ScCP

SNR = 20.2 dB

Conclusions

- Multilevel method combined with proximal algorithms appear to be a interesting way to accelerate standard schemes.
- Adapted to very large scale problems.
- Compatible with PnP and unfolded ?

References

- G. Lauga, E. Riccietti, N. Pustelnik, and P. Goncalves IML FISTA: Inexact MuLtilevel FISTA for Image Restoration, submitted, 2023.
- G. Lauga, E. Riccietti, N. Pustelnik, and P. Goncalves Multilevel FISTA for Image Restoration, IEEE ICASSP, 2023.
- G. Lauga, E. Riccietti, N. Pustelnik, P. Gonçalves, Méthodes multi-niveaux pour la restauration d'images hyperspectrales, Colloque GRETSI, September 2023.
- G. Lauga, E. Riccietti, N. Pustelnik, P. Gonçalves, Méthodes proximales multi-niveaux pour la restauration d'images, Colloque GRETSI, September 2022.

Journées SMAI MODE

société de mathématiques appliquées et industrielles

Journées SMAI MODE 2024

27–29 mars 2024
Lyon
Fuseau horaire Europe/Paris

Entrer le texte à rechercher

- Accueil
- Dates clés
- Soumission de communication
- Demande de bourse
- Inscription
- Tarifs
- Prix Dodu et prix poster
- Informations pratiques
- Comités
- Partenaires
- Photos
- Affiche
- Mention d'information

Accueil