Non-local Regularization of Inverse Problems

Gabriel Peyré, Sébastien Bougleux, Laurent D. Cohen

Non-local Regularization of Inverse Problems

Gabriel Peyré
Ceremade, Université Paris-Dauphine,
75775 Paris Cedex 16, France

Sébastien Bougleux
GREYC, Université de Caen,
14050 Caen Cedex, France

Laurent Cohen
Ceremade, Université Paris-Dauphine,
75775 Paris Cedex 16, France

Abstract

This article proposes a new framework to regularize imaging linear inverse problems using an adaptive non-local energy. A non-local graph is optimized to match the structures of the image to recover. This allows a better reconstruction of geometric edges and textures present in natural images. A fast algorithm computes iteratively both the solution of the regularization process and the non-local graph adapted to this solution. The graph adaptation is efficient to solve inverse problems with randomized measurements such as inpainting random pixels or compressive sensing recovery. Our non-local regularization gives state-of-the-art results for this class of inverse problems. On more challenging problems such as image super-resolution, our method gives results comparable to sparse regularization in a translation invariant wavelet frame.

1 Introduction

This paper studies the solution of linear ill-posed inverse problems in image processing. The goal is to recover a high resolution image $f_0 \in \mathbb{R}^n$ of n pixels from a set of $p \leqslant n$ noisy linear measurements

$$u = \Phi f_0 + \varepsilon \in \mathbb{R}^p.$$

where ε is an additive noise. The linear operator Φ typically accounts for some blurring, sub-sampling or missing pixels so that the measured data u only captures a small portion of the original image f one wishes to recover.

To solve this ill-posed problem, one needs some prior knowledge on the kind of typical images one expects to restore. This prior information should help to recover the missing information. Regularization methods assume that f_0 has some smoothness, for instance small derivatives (linear Sobolev regularization) or bounded variations (non-linear regularization). This paper derives a new prior model based on non-local comparison of patches.

A prior is a functional $J(f)$ that is small when f is close to the smoothness model. A regularized solution $f^\star$ to the inverse problem is written in variational

form as

$$f^\star \in \operatorname*{argmin}_{f\in\mathbb{R}^n} \frac{1}{2\lambda}\|u-\Phi f\|^2 + J(f), \tag{1}$$

where the minimum is not necessarily unique. The weight λ needs to be adapted to match the amplitude of the noise ε, which might be a non-trivial task in practical situations.

1.1 Previous Works

1.1.1 Classical smoothness priors.

The simplest prior model assumes an uniform smoothness of the image, and uses for instance a discretized Sobolev norm

$$J^{\mathrm{Sob}}(f) = \sum_x \|\nabla f(x)\|^2, \tag{2}$$

where $\nabla f(x)$ is a finite difference approximation of the gradient of f at pixel x. To enable the recovery of sharp features such as edges, Rudin, Osher and Fatemi [50] proposed to use the total variation norm for denoising purpose, when $\Phi = \mathrm{Id}$

$$J^{\mathrm{TV}}(f) = \sum_x \|\nabla f(x)\|. \tag{3}$$

Total variation regularization prior (3) has been used to solve super-resolution [38] and inpainting of small holes [13]. Inpainting of larger holes requires higher order regularizations, that take into account the curvature of the level lines [40, 5, 7] or a tensor diffusion [56].

Given a frame $\{\psi_m\}_m$ of $\mathbb{R}^n$, one defines an analysis sparsity enforcing prior in this frame using the ℓ^1 norm of the correlation with the frame atoms

$$J^{\mathrm{spars}}(f) = \sum_m |\langle f,\, \psi_m\rangle|. \tag{4}$$

This prior has been introduced by Donoho and Johnstone [21] with the orthogonal wavelet basis for denoising purpose, when $\Phi = \mathrm{Id}$. In this case, the solution $f^\star$ of (1) is obtained with a soft thresholding. Sparsity prior (4) has been used to solve general inverse problems, see for instance [19, 16] and the references therein. It can also be used in conjunction with redundant frames instead of orthogonal bases, see for instance [26]. For a redundant frame of $\mathbb{R}^n$, it is also possible to search for the coefficients of $f^\star$ in this frame. This corresponds to a synthesis sparsity prior, that differs from (4), see for instance [29].

It is possible to adapt the representation by learning the atoms $\{\psi_m\}_m$ used in the sparse regularization (4), see for instance [43]. This leads to redundant representations with state-of-the-art image denoising results [25]. Dictionaries can also be learned iteratively to perform image inpainting [37]and simultaneous image separation and inpainting [48].

1.1.2 Non-local diffusion.

In order to better respect edges in images than total variation and wavelet sparsity, several edge-aware filtering schemes have been proposed, among which

Yaroslavsky's filter [59], the bilateral filter [55], Susan filter [52] and Beltrami flow [53]. The non-local means filter [8] goes one step further by averaging pixels that can be arbitrary far away, using a similarity measure based on distance between patches.

As shown for instance in [45], for denoising $\Phi = \mathrm{Id}$, these edge adaptive filters are related to the minimization of (1) using a graph based regularization over the image

$$J_w^{\mathrm{graph}}(f) = \sum_{x,y} w_{x,y} |f(x) - f(y)|^\alpha \tag{5}$$

where $\alpha = 2$. The weights $w_{x,y}$ are computed from the input noisy image u using either the distance $|u(x) - u(y)|$ between the noisy pixel values [55, 59, 53] or the distance $\|p_x(f) - p_y(f)\|$ between the patches $(p_x(f), p_y(f))$ around x and y [8, 15, 54]. This variational denoising is related to sparsity in an adapted basis of eigenvectors of non-local diffusion operators [54, 45]

This graph based energy (5) is generalized using an arbitrary $\alpha \geqslant 1$ which, for $\alpha = 1$, defines a non-local total variation [32, 61, 27].

1.1.3 Non-local regularization of inverse problems.

For some class of inverse problems, the weights $w_{x,y}$ can be estimated from the observations u. This is for instance the case for inpainting small holes [34], deblurring [41, 35, 9], demosaicing [10] and segmentation [33].

These approaches share similarities with exemplar-based super-resolution, see for instance [18, 31, 23]. Although these methods operate using patches comparisons, they are different because they make use of pairs of low/high resolution exemplar patches.

For many other inverse problems, like inpainting of large or randomized holes, or compressive sensing, the observation u cannot be used directly to estimate the regularization graph w. One thus needs to iteratively estimate the graph while performing the inversion. For inpainting, computer graphics methods perform patch copy [17], that is closely related to an iterative estimation of a non-local graph. These methods are related to texture synthesis with patch recopy [24, 58], and can be re-casted as a non-convex variational problem [3].

We have presented in the conference paper [47] for the first time a general framework for the non-local regularization of inverse problems. This framework is used by Zhang et al. in [60], where Bregman iterations are used instead of the forward-backward splitting initially proposed in [47], which might result in a faster algorithm. A similar framework for image inpainting is developed by Facciolo et al. in [28], where a variational justification for our initial choice of weights [47] is given.

2 Contributions

This paper proposes a new framework to solve general inverse problems using a non-local and non-linear regularization on graphs. Our algorithm is able to efficiently solve for a minimizer of the proposed energy by iteratively computing an adapted graph and a solution of the inverse problem. We show applications to inpainting, super-resolution and compressive sampling where this new framework improves over wavelets and total variation regularizations.

This framework extends the initial proposal of our conference paper [47] as well as the Bregmanized version of [60] by considering a more general family of priors that compares patches and not pixel values. This allows us to prove the convergence of our scheme toward a stationary point of the non-local energy. It extends the work of [28] by treating arbitrary inverse problems beyond inpainting and by providing a convergence guarantee.

3 Non-local Regularization of Inverse Problems

3.1 Non-local Regularization

This section introduces a non-local graph-based regularization of the inverse problem $u = \Phi f_0 + \varepsilon$. This regularization is adaptive since the energy we consider

$$J(f) = J_w(f) + \gamma E(w) \tag{6}$$

is parameterized by a non-local graph w. This graph is a set of weights $w_{x,y} \geqslant 0$ that link pixels x and y over the image plane. The functional $J_w(f)$ regularizes the image and enforces a non-local regularity along the graph w, while the functional $E(w)$ constrains the graph itself.

The non-adaptive regularization (1) is extended to this adaptive non-local setting by considering a minimization on both the image to recover and the graph

$$(f^\star, w^\star) \in \operatorname*{argmin}_{f \in \mathbb{R}^n, w \in \mathcal{C}} \mathcal{E}(f,w) = \frac{1}{2\lambda}\|u - \Phi f\|^2 + J_w(f) + \gamma E(w), \tag{7}$$

where $\mathcal{C}$ is an additional set of constraints on the graph.

3.2 Patch Extraction

A non-local regularization is obtained by comparing small patches that can be far away in the image plane. A patch of $\tau \times \tau$ pixels at location $x \in \{0, \dots, \sqrt{n}-1\}^2$ in the image is defined as

$$\forall\, t \in \{-(\tau-1)/2+1, \dots, (\tau-1)/2\}^2, \qquad \pi_x(f)(t) = f(x+t) \tag{8}$$

where τ is assumed to be an odd integer. A patch $\pi_x(f)$ is a vector of size τ^2.

To speed up the computation, the dimensionality of $\pi_x(f)$ is reduced from τ^2 to $q \leqslant \tau^2$ using an an orthogonal projector $U \in \mathbb{R}^{q \times \tau^2}$ that satisfies $UU^* = \mathrm{Id}$. The resulting patches of lower dimension are defined as

$$p_x(f) = U\pi_x(f) \in \mathbb{R}^q, \tag{9}$$

Using a small value for q speeds up the algorithms but might deteriorate the visual quality of the result. Section 5 describes how U is computed in practice for the numerical experiments.

We note that this framework extends to color images f of n pixels by considering patches of dimension $3\tau^2$. In this case, the projector U is useful to reduce the dimensionality by making use of the redundancy across color channels.

3.3 Graph-based Priors on Images

A non-local graph is a set of weights $w = \{w_{x,y}\}_{x,y}$ which assigns to each pair of pixels (x, y) a weight $w_{x,y} \geqslant 0$. We further impose that for each x, $\{w_{x,y}\}_y$ is a probability distribution and that the graph only connects pixels that are not too far away

$$\mathcal{C} = \left\{ w \setminus \sum_y w_{x,y} = 1, \quad \text{and} \quad \left\{ \begin{array}{l} \|x-y\| \leqslant \rho \;\Rightarrow\; w_{x,y} \geqslant w_{\min} \\ \|x-y\| > \rho \;\Rightarrow\; w_{x,y} = 0 \end{array} \right. \right\}. \tag{10}$$

Note that we impose a lower bound $w_{x,y} \geqslant w_{\min}$ on the weights, where $w_{\min} \geqslant 0$ is a small constant. Theorem 1 requires $w_{\min} > 0$ to ensure the convergence of our algorithm, although we observe numerically the convergence even if $w_{\min} = 0$.

The parameter ρ controls the degree of non-locality of the graph. For image containing periodic features, increasing the value of ρ might improve the numerical results but it also increases the complexity of the algorithms. For natural images, it might actually improve the results as well as the algorithmic complexity to impose a not so large value of ρ, which leads to a semi-non-local regularization.

This weighted graph is used to indicate which local patches should be compared in the image, and leads to the following non-local regularization functional

$$J_w(f) = \sum_{\|x-y\| \leqslant \rho} w_{x,y} \|p_x(f) - p_y(f)\|^\alpha, \tag{11}$$

where $\alpha \geqslant 1$ controls the shape of the functional. Such patch variations are considered by Peyré for texture synthesis [46] and for inpainting by Facciolo et al.[28].

For $\alpha = 2$, one obtains a quadratic energy that extends the Sobolev smoothness prior (2) from pixels to patches. For $\alpha = 1$, this energy extends the total variation (3) to patches. A major difference with previous works [32, 61, 27] is that J_w uses patch variations $\|p_x(f) - p_y(f)\|$ instead of pixel variations $|f(x) - f(y)|$ as this is the case in (5).

3.4 Maximum Entropy Prior on the Graph Weights

The constraint $w \in \mathcal{C}$ is not strong enough to select an efficient graph to process an image. Since $\sum_y w_{x,y} = 1$, the set of weights $\{w_{x,y}\}_y$ should be thought as being a probability distribution. To control the spread of this distribution around its modes, and following [28], we use a negative entropy to regularize the weights

$$E(w) = \left\{ \begin{array}{ll} \displaystyle\sum_{\|x-y\| \leqslant \rho} (w_{x,y} - w_{\min}) \log(w_{x,y} - w_{\min}) & \text{if} \quad w \in \mathcal{C} \\ +\infty & \text{otherwise.} \end{array} \right.$$

The parameter γ in (6) weights the influence of this entropy constraint on w and should be adapted to the geometry of the image to recover and to the noise level.

Decreasing the value of γ enforces the weight distribution $\{w_{x,y}\}_y$ to have a low entropy, and thus reduces the spread of the values, see Figure 1. When γ tends to 0, the distribution tends to a degenerate Dirac distribution. Section 4.3 shows more precisely the influence of the parameter γ on the computed weight.

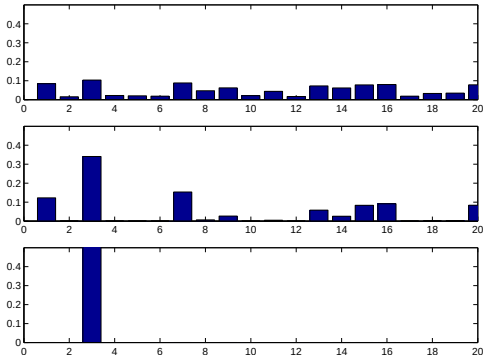


Figure 1: *Influence of the parameter γ on the weights computed according to (30) (only a small subset of the weights $\{w_{x,y}\}_y$ is shown). The values of $\gamma / \max_y \|p_x(f) - p_y(f)\|$ are $0.5, 0.1, 0.01$ from top to bottom.*

3.4.1 Non-local Patch Operators

Before describing in the next section our algorithm to solve (7), we introduce linear operators to re-write in a compact way graph-based priors.

The patch extraction process (8) defines a mapping from the pixel domain $\mathbb{R}^n$ to the patch domain $\mathcal{P}$

$$P : \left\{ \begin{array}{ccc} \mathbb{R}^n & \longrightarrow & \mathcal{P} \\ f & \longmapsto & \{p_x(f)\}_x \end{array} \right. .$$

A set of patches $\{p_x(f)\}_x \in \mathcal{P}$ is stored as a matrix of $q \times n$ elements. The adjoint mapping is defined as

$$P^* : \left\{ \begin{array}{ccc} \mathcal{P} & \longrightarrow & \mathbb{R}^n \\ \{p_x\}_x & \longmapsto & f \end{array} \right. \quad \text{where} \quad f(x) = \sum_y (U^* p_y)(x - y)$$

where the sum is restricted to pixels $y = (y_1, y_2)$ such that

$$-\frac{\tau - 1}{2} \leqslant x_1 - y_1 \leqslant \frac{\tau - 1}{2} \quad \text{and} \quad -\frac{\tau - 1}{2} \leqslant x_2 - y_2 \leqslant \frac{\tau - 1}{2}.$$

We note that a special care should be taken near boundaries of the image. In the numerical experiments, we use a symmetric extension of the image to avoid boundary artifacts.

The non-local energy defined in (11) is a vectorial ℓ^α norm

$$J_w(f) = \|G_w P f\|_\alpha^\alpha,$$

where the patch-valued gradient maps patches in $\mathcal{P}$ to patches differentials in $\mathcal{D}$

$$G_w : \left\{ \begin{array}{ccc} \mathcal{P} & \longrightarrow & \mathcal{D} \\ \{p_x\}_x & \longmapsto & \{d_{x,y}\}_{\|x-y\| \leqslant \rho} \end{array} \right. , \quad \text{where} \quad d_{x,y} = w_{x,y}^{\frac{1}{\alpha}} (p_x - p_y). \tag{12}$$

A patch differential $\{d_{x,y}\}_{\|x-y\| \leqslant \rho} \in \mathcal{D}$ is stored as an array of size $q \times n \times C_\rho$, where

$$C_\rho = |\{x \setminus \|x\| \leqslant \rho\}|, \tag{13}$$

is the number of pixels in an Euclidean ball of radius ρ. The adjoint of this patch-valued gradient is a patch-valued divergence

$$G_w^* : \left\{ \begin{array}{ccc} \mathcal{D} & \longrightarrow & \mathcal{P} \\ \{d_{x,y}\}_{\|x-y\|\leqslant\rho} & \longmapsto & \{p_x\}_x \end{array} \right. , \quad \text{where} \quad p_x = \sum_{\|x-y\|\leqslant\rho} w_{x,y}^{\frac{1}{\alpha}} d_{x,y} - w_{y,x}^{\frac{1}{\alpha}} d_{y,x}.$$

The ℓ^α norm of patch differentials $d = \{d_{x,y}\}_{\|x-y\|\leqslant\rho}$ is defined as

$$\forall\, \alpha < +\infty, \quad \|d\|_\alpha^\alpha = \sum_{\|x-y\|\leqslant\rho} \|d_{x,y}\|^\alpha, \quad \text{and} \quad \|d\|_\infty = \max_{\|x-y\|\leqslant\rho} \|d_{x,y}\|. \tag{14}$$

4 Non-local Regularization Algorithm

While the optimization problem (7) is separately convex with respect to f and to w, it is not jointly convex in (f, w). The minimization of (7) is thus difficult, and we propose to use a a coordinate descent algorithm that optimizes successively the graph w and then the image f to recover.

4.1 Block Coordinate Descent Algorithm

Given some fixed graph $w \in \mathcal{C}$, an optimal image $f(w)$ solves

$$f(w) \in \operatorname*{argmin}_{f\in\mathbb{R}^n} \mathcal{E}(f, w) = \operatorname*{argmin}_{f\in\mathbb{R}^n} \frac{1}{2}\|u - \Phi f\|^2 + \lambda J_w(f). \tag{15}$$

where $\mathcal{E}$ is defined in (7). Section 4.2 details how to compute such an image. Given some fixed image $f \in \mathbb{R}^n$, the optimal graph is defined as

$$w(f) = \operatorname*{argmin}_{w\in\mathcal{C}} \mathcal{E}(f, w) = \operatorname*{argmin}_{w\in\mathcal{C}} J_w(f) + \gamma E(w). \tag{16}$$

Section 4.3 details how to compute such a graph.

The block coordinate descent algorithm starts from some initial graph w_0, for instance using constant weights. It then iterates

$$f_{k+1} = f(w_k) \quad \text{and} \quad w_{k+1} = w(f_{k+1}). \tag{17}$$

The following theorem studies the convergence of this iterative scheme in the case $\alpha > 1$. It makes use of the gradient $G_0 = G_{w^0}$ for constant weights $w_{x,y}^0 = 1/\tau^2$ for all $\|x - y\| \leqslant \rho$.

Theorem 1 *We suppose that the lower bound on the weights satisfies $w_{\min} > 0$, that $\alpha > 1$, and that*

$$\operatorname{Ker}(\Phi) \cap \operatorname{Ker}(G_0 P) = \{0\}. \tag{18}$$

The sequence $\{(f_k, w_k)\}_k$ is well defined and bounded. Every accumulation point $(f^\star, w^\star)$ of the sequence is a stationary point of $\mathcal{E}$ defined in (7), which means $\nabla\mathcal{E}(f^\star, w^\star) = 0$.

Proof: We check the hypothesis of Theorem 4.1 of Tseng [57], that proves our result. Following the notations of [57], the energy is rewritten as

$$\mathcal{E}(f,w) = \frac{1}{2\lambda}\|u-\Phi f\|^2 + J_w(f) + \mu E(w) + i_{\mathcal{C}}(w) = f_0(f,w) + f_1(w)$$

where $f_1(w) = \mu E(w) + i_{\mathcal{C}}(w)$. The indicator function is defined as $i_{\mathcal{C}}(w) = +\infty$ if $w \notin \mathcal{C}$ and $i_{\mathcal{C}}(w) = 0$ otherwise.

The function $\mathcal{E}$ is continuous on the level set

$$X_0 = \{(f,w) \setminus \mathcal{E}(f,w) \leqslant \mathcal{E}(f_0,w_0)\}.$$

Note that for any k, $(f_k,w_k) \in X_0$ because $\mathcal{E}$ is decaying with k. We then show that X_0 is bounded, which ensures the existence of convergent sub-sequences. For $w \in \mathcal{C}$, one has $w \geqslant w_{\min} > 0$ for all $\|x-y\| \leqslant \rho$, and hence

$$J_w(f) \geqslant w_{\min}\|G_0 P f\|_\alpha^\alpha \geqslant w_{\min} C_0^\alpha \|G_0 P f\|_2^\alpha$$

where we have used the fact that $\|\cdot\|_\alpha \geqslant C_0\|\cdot\|_2$, where $C_0 = 1$ if $1 < \alpha \leqslant 2$, and $C_0 = n^{1/\alpha-1/2}$ if $\alpha > 2$. We denote as Π the orthogonal projector on $\mathrm{Ker}(G_0 P)$. Denoting as $C_1 > 0$ is the smallest non zero singular value of $G_0 P$, one thus has

$$J_w(f) \geqslant w_{\min} C_0^\alpha \|G_0 P(f-\Pi f)\|_2^\alpha \geqslant w_{\min} C_0^\alpha C_1^\alpha \|f-\Pi(f)\|_2^\alpha. \tag{19}$$

We decompose $u = \Phi\bar{u} + r$ where $r \in \mathrm{Im}(\Phi)^\perp$, so that

$$\|u-\Phi f\|^2 \geqslant \|\Phi(\bar{u}-f)\|^2 \geqslant C_2^2 \|\Pi(\bar{u}-f)\|^2 \tag{20}$$

where $C_2 > 0$ the smallest non zero singular value of Φ, and where we have make use of the hypothesis that $\mathrm{Ker}(\Phi) \cap \mathrm{Ker}(G_0 P) = \{0\}$.

Putting (19) and (20) together shows that

$$\mathcal{E}(f,w) \geqslant \frac{C_2^2}{2\lambda}\|\Pi(\bar{u}) - \Pi(f))\|^2 + w_{\min} C_0^\alpha C_1^\alpha \|f-\Pi(f)\|_2^\alpha + \mu E(w).$$

Since under the constraint $w \in \mathcal{C}$, w is bounded, this shows that X_0 is also bounded.

The function $f_0(f,w)$ is of class C^1, because the mapping $(f,w) \mapsto J_w(f)$ is of class C^1 for $\alpha > 1$. The function $\mathcal{E}$ satisfies hypothesis (A1) of [57], and hence lemma 3.1 of [57] shows that $\mathcal{E}$ is regular at each point of X_0. □

If one does not perform any dimensionality reduction to compute the patches $p_x(f)$, meaning $U = \mathrm{Id}$, then condition (18) boils down to $1 \notin \mathrm{Ker}(\Phi)$. This is a classical non-degeneracy condition for variational regularizations involving derivatives, such as the Sobolev prior (2) and the total variation prior (3). For an arbitrary projector U, condition (18) is difficult to check analytically, and we verify numerically that (18) holds in all numerical examples.

Theorem 1 ensures convergence for $\alpha > 1$. Unfortunately, for $\alpha = 1$, the term $J_w(f)$ that mixes w and f is non smooth, so that one cannot guarantee the convergence in this case. Section 5 shows that this iterative scheme converges well in practical cases even for $\alpha = 1$.

4.2 Optimization on the Image

To compute $f(w)$ defined in (15) given some weights $w \in \mathcal{C}$, we distinguish between the cases $\alpha > 1$ and $\alpha = 1$.

4.2.1 Optimization on the Image, $\alpha > 1$

When $\alpha > 1$, computing $f(w)$ corresponds to the minimization of a smooth convex functional defined in (15). A simple scheme is a gradient descent, which reads

$$f^{(\ell+1)} = f^{(\ell)} - \nu_\ell \left(\Phi^*(\Phi f^{(\ell)} - u) + \lambda \alpha P^* G_w^* \tilde{d}^{(\ell)} \right) \tag{21}$$

where $\tilde{d}^{(\ell)} \in \mathcal{D}$ is computed as

$$d^{(\ell)} = G_w P f^{(\ell)} \quad \text{and} \quad \forall\, (x,y),\ \tilde{d}_{x,y}^{(\ell)} = d_{x,y}^{(\ell)} \| d_{x,y}^{(\ell)} \|^{\alpha-2}.$$

Since for $\alpha < 2$ the functional does not have a Lipshitz gradient, the step size ν_ℓ is computed using a 1-D line search procedure to ensure the following convergence result, see for instance [14].

Proposition 1 *One has $f^{(\ell)} \to f(w)$ when $\ell \to +\infty$, where $f(w)$ is a solution of (15).*

It is also possible to use more efficient second order schemes, such as quasi-Newton [4], that works well in practice.

The case $\alpha = 2$ deserves special attention, because it corresponds to a quadratic minimization, and a solution can be found by solving the following linear system

$$(\Phi^*\Phi + 2\lambda P^* G_w^* G_w P) f(w) = \Phi^* y. \tag{22}$$

A solution $f(w)$ can be computed efficiently using a conjugate gradient descent.

4.2.2 Optimization on the Image, $\alpha = 1$

For $\alpha = 1$, $f(w)$ is a solution of the following convex optimization problem

$$f(w) \in \operatorname*{argmin}_{f \in \mathbb{R}^n} \frac{1}{2} \|u - \Phi f\|^2 + \lambda \|G_w P f\|_1. \tag{23}$$

This corresponds to the minimization of the sum of a smooth quadratic functional $\|u - \Phi f\|^2$ and a non-smooth functional $\|G_w P f\|_1$. Several efficient first order schemes have been devised to perform such a minimization, among which forward-backward splitting [16, 6], Nesterov's algorithm [42, 2] and Bregman iterations [60]. We now detail the forward-backward splitting, that has the advantage of simplicity, although more efficient algorithms could be used as well.

Starting from an initial image $f^{(0)}$, forward-backward iterations alternate between a gradient descent step

$$\tilde{f}^{(\ell)} = f^{(\ell)} - \nu \Phi^*(\Phi f^{(\ell)} - u) \tag{24}$$

and a proximal denoising correction

$$f^{(\ell+1)} = \operatorname{prox}_{\nu \lambda J_w}(\tilde{f}^{(\ell)}), \tag{25}$$

where the proximal operator is the solution of a denoising problem

$$\operatorname{prox}_{\omega J_w}(\tilde{f}) = \operatorname*{argmin}_{f \in \mathbb{R}^n} \frac{1}{2} \|f - \tilde{f}\|^2 + \omega \|G_w P f\|_1. \tag{26}$$

Following for instance [16], one has the following convergence result.

Proposition 2 *If $0 < \nu < 2/\|\Phi^*\Phi\|$, one has $f^{(\ell)} \to f(w)$ when $\ell \to +\infty$, where $f(w)$ is a solution of (23).*

The proximal operator (26) is the solution of a convex functional that generalizes the total variation denoising. Following Chambolle [12], we compute this operator through a dual optimization problem, that performs the optimization on a patch differential $d \in \mathcal{D}$ rather than on the set of images.

Proposition 3 *One has*

$$\mathrm{prox}_{\omega J_w}(\tilde{f}) = \tilde{f} - \omega P^* G_w^* d^\star$$

where $d^\star$ is a solution of the following dual convex optimization problem

$$d^\star \in \operatorname*{argmin}_{d \in \mathcal{D}, \|d\|_\infty \leqslant 1} \|\tilde{f} - \omega P^* G_w^* d\|^2. \tag{27}$$

Proof: The proof of this duality relation follows closely the original work of Chambolle [12], by replacing the gradient operator ∇ by $G_w P$, and by replacing the duality between the ℓ^1/ℓ^∞ norms of 2-D vector fields by the duality among the ℓ^1/ℓ^∞ norms defined by (14). □

The constrained problem (27) can be solved using several first order schemes, including foward-backward splitting [42] and Nesterov's algorithm [42].

The foward-backward splitting corresponds to the usual projected gradient descent, that iterates between a gradient step

$$\tilde{d}^{[m]} = d^{[m]} + \eta G_w P(\tilde{f}/\omega - P^* G_w^* d^{[m]}), \tag{28}$$

and a projection step on the ℓ^∞ constraint

$$d_{x,y}^{[m+1]} = \frac{\tilde{d}_{x,y}^{[m]}}{\max(1, \|\tilde{d}_{x,y}^{[m]}\|)}. \tag{29}$$

Standard results for the convergence of a projected gradient descent, see for instance [16], shows the following convergence result.

Proposition 4 *If $0 < \eta < 2/\|G_w P P^* G_w^*\|$, one has $d^{[m]} \to d^\star$ when $m \to +\infty$, where $d^\star$ is a solution of (27).*

The operator norm $\|G_w P P^* G_w^*\|$ is estimated numerically using a few power iterations to compute the largest singular value of $G_w P$.

4.3 Optimization on the Graph

If f is fixed, optimizing (7) with respect to w defines an optimal graph $w(f)$, which is the solution of the following strictly convex optimization problem

$$\begin{aligned} w(f) &= \operatorname*{argmin}_{w \in \mathcal{C}} J_w(f) + \gamma E(w) \\ &= \operatorname*{argmin}_{w \in \mathcal{C}} \sum_{\|x-y\| \leqslant \rho} w_{x,y} \|p_x(f) - p_y(f)\|^\alpha + \gamma(w_{x,y} - w_{\min}) \log(w_{x,y} - w_{\min}). \end{aligned}$$

The optimal graph is computed as detailed in the following proposition, where

$$C_\rho = |\{x \setminus \|x\| \leqslant \rho\}|,$$

already introduced in (13), is the number of connexions between one patch and its neighboring patches.

Proposition 5 *For any $f \in \mathbb{R}^N$, if $w_{\min} \leqslant 1/C_\rho$, one has*

$$w(f)_{x,y} = \frac{\tilde{w}_{x,y}}{Z_x} \quad \textit{where} \quad \tilde{w}_{x,y} = \begin{cases} w_{\min} + e^{-\frac{\|p_x(f)-p_y(f)\|^\alpha}{\gamma}} & \textit{if } \|x-y\| \leqslant \rho, \\ 0 \quad \textit{otherwise}, \end{cases} \tag{30}$$

where the normalizing constant is

$$Z_x = \frac{\sum_y \tilde{w}_{x,y}}{1 - w_{\min} C_\rho}.$$

Proof: The proof follows closely the one in [28], that was made for $\alpha = 1$ and for $w_{\min} = 0$. For each x, one needs to find the minimizer with respect to $\{w_{x,y}\}_y$ of

$$\sum_y w_{x,y} \|p_x(f) - p_y(f)\|^\alpha + \gamma (w_{x,y} - w_{\min}) \log(w_{x,y} - w_{\min}) \tag{31}$$

under the constraints that $w_{x,y} \geqslant w_{\min}$ for $\|x-y\| \leqslant \rho$, and that $\sum_y w_{x,y} = 1$. If a minimum is reached at some w where for all y, $w_{x,y} > w_{\min}$, then the gradient with respect to $\{w_{x,y}\}_y$ of the objective (31) is proportional to the gradient of the constraint $\sum_y w_{x,y} = 1$, so that there exists some $C \in \mathbb{R}$ such that

$$\forall\, y, \quad \|p_x(f) - p_y(f)\|^\alpha + \gamma \left(\log(w_{x,y} - w_{\min}) + 1\right) = C.$$

Adapting C for $\sum_y w_{x,y} = 1$ to hold, one obtains the formula (30). If $w_{\min} \leqslant 1/C_\rho$, one has $w \in \mathcal{C}$ so that w is a valid minimizer of (31). Since the function to optimize is strictly convex and $\mathcal{C}$ is convex, it is the only one. □

The constraint $w_{\min} \leqslant 1/C_\rho$ imposed by the theorem is not an issue since for numerical applications, $w_{\min}$ is chosen very small. We note that for $\alpha = 2$, one recovers the Gaussian weights used in NL-means [8], but they differ if $\alpha \neq 2$.

4.4 The Non-local Regularization Algorithm

The overall block-wise coordinate descent algorithm is summerized in Table 1, both for the case $\alpha > 1$ (gradient descent) and $\alpha = 1$ (forward-backward). For the gradient descent, it makes use of two nested inner loops on (f_k, w_k) and $f^{(\ell)}$. For the forward-backward, it makes use of three nested inner loops, on (f_k, w_k), $f^{(\ell)}$ and $d^{[m]}$. The number of inner iterations is monitored by the precision parameters tol_f and tol_d.

Convergence of the forward-backward iterations when ℓ tends to $+\infty$ is guaranteed if the errors generated by the inner iterations on m are summable [16]. In theory, this requires to lower the tolerance $\mathrm{tol}_d = \mathrm{tol}_d^{(k)}$ as k is increasing. For the numerical experiments, we use fixed tolerances $\mathrm{tol}_f = \mathrm{tol}_d = 10^{-3}$, which in practice does not lead to convergence issue.

Algorithm 1: Block coordinate descent algorithm to minimize (7) for $\alpha = 1$.

Initialization: set $f_0 = 0$, define w_0 using constant weights, set $k = 0$.
while *not converged* **do**
  Image update: initialize $f^{(0)} = f_k$, set $\ell = 0$.
  while $\|f^{(\ell+1)} - f^{(\ell)}\| > \mathrm{tol}_f$ **do**
    if $\alpha > 1$ **then**
      Compute $f^{(\ell+1)}$ using (21) with $w = w_k$.
    else
      Gradient descent: compute $\tilde{f}^{(\ell)}$ using (24).
      Proximal correction: initialize $d^{[0]} = 0$, set $m = 0$.
      while $\|d^{[m]} - d^{[m+1]}\| > \mathrm{tol}_d$ **do**
        Gradient descent: compute $\tilde{d}^{[m]}$ using (28) with $\tilde{f} = \tilde{f}^{(\ell)}$ and $w = w_k$.
        Projection: compute $d^{[m+1]}$ using (29).
      Set $f^{(\ell+1)} = \tilde{f}^{(\ell)} - \nu\lambda P^* G_w^* d^{[m+1]}$.
  Set $f_{k+1} = f^{(\ell+1)}$.
  Weight update: compute $w_{k+1} = w(f_{k+1})$ using (30).
Output: f_{k+1}.

5 Numerical Experiments

We have tested our algorithm on a wide range of images $f_0 \in \mathbb{R}^n$ of $n = 256 \times 256$ pixels, containing both edges and textures. In the numerical simulations, we consider three different regularizations:

- The total variation energy J^{TV}, defined in equation (3). An algorithm similar to algorithm 1 is used for this minimization, using the usual discrete gradient ∇ instead of the graph gradient G_w.
- The sparsity energy J^{spars}, defined in equation (4), using a redundant tight frame of translation invariant 7-9 biorthogonal wavelets $\{\psi_m\}_m$, see [39]. An algorithm similar to algorithm 1 is used for this minimization, excepted that the proximal projection is computed with a soft thresholding as detailed in [29].
- The regularization J_w in an optimized graph, solved using algorithm 1. Section 5.1compares the cases $\alpha = 1$ and $\alpha = 2$, and the remaining part of the section focus on $\alpha = 1$. For this regularization, the size of the patch is set to $\tau = 5$ pixels. The parameter γ of equation (6) is set to $\tau\|f\|_\infty/10$, which gives satisfying results for the noise level we consider. The locality parameter ρ of equation (10) is fixed to $\rho = 15$ pixels. The dimensionality reduction of the patches is set to $q = 14$, and the projector U used in (9) is obtained by considering the q leading PCA eigenvectors of random patches extracted from noise-free natural images.

In the three applications of Sections 5.1, 5.2 and 5.3, we use a Gaussian white noise ε of standard deviation $0.02\|u\|_\infty$. For all the proposed methods, the parameter λ is optimized in an oracle manner to maximize the PSNR of the recovered image $f^\star$

$$\mathrm{PSNR}(f^\star, f) = -20\log_2(\|f^\star - f\|/\|f\|_\infty).$$

Computational complexity. The computational complexity of the method depends on the dimensionality q of the patches, the non-locality factor ρ, and the

shape parameter α. For $\alpha = 1$, the overhead with respect to the total variation regularization is roughly $q\rho^2$ because of the higher dimensionality of the non-local gradient. The computation time is lower for $\alpha = 2$ because of the fast convergence of the conjugate gradient to solve (22). Note that several methods and data structures have been proposed to accelerate significantly non-local filtering, see for instance [36, 1]. These technics might also lead to a speed up of our regularization framework.

5.1 Inpainting

Inpainting aims at filling missing pixels from an image. It corresponds to the following masking operator

$$(\Phi f)(x) = \begin{cases} 0 & \text{if} \quad x \in \Omega, \\ f(x) & \text{if} \quad x \notin \Omega, \end{cases} \tag{32}$$

where $\Omega \subset \{0, \ldots, \sqrt{n}-1\}^2$ is the region where the input data has been damaged. In this case, $\Phi^* = \Phi$, and one can take a proximity step size $\nu = 1$ so that the gradient descent step (24) becomes a projection

$$\tilde{f}^{(\ell)} = \begin{cases} f^{(\ell)}(x) & \text{if} \quad x \in \Omega, \\ u(x) & \text{if} \quad x \notin \Omega. \end{cases} \tag{33}$$

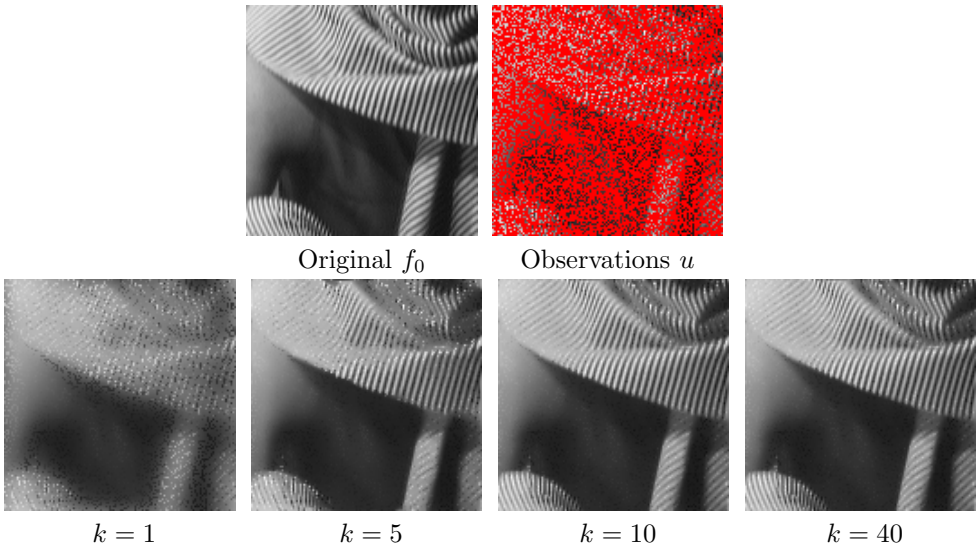
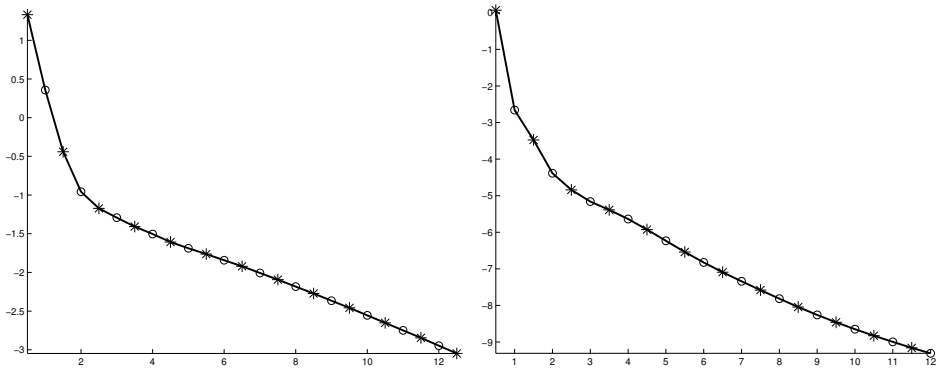


Figure 2: *Display of the iterates f_k of the algorithm for inpainting, in the case $\alpha = 1$.*

Figure 2 shows, for a textured image with $|\Omega|/n = 0.7$, the convergence of the iterates of the algorithm for $\alpha = 1$. One can see how each iteration of the algorithm progressively refines the geometry so that the texture is correctly reconstructed. This is possible because of the iterative refinement of the graph w_k. Although there

is no formal proof of convergence for the case $\alpha = 1$, we observe numerically a fast convergence.

This convergence is further confirmed on Figure 3 by the logarithmic plot of the decay of the energy $\mathcal{E}(f_k, w_k)$ toward $\mathcal{E}(f_\infty, w_\infty)$. We estimated numerically this final value using $(f_\infty, w_\infty) \approx (f_{100}, w_{100})$ and show the energy decay only for $k \leqslant 12$. The decay is fast (roughly geometrical) for both $\alpha = 1$ and $\alpha = 2$. We also display the intermediate energies $\mathcal{E}(f_{k+1}, w_k)$ to show that both the image and the graph update contribute to the decay of the energy.



$\alpha = 1$ $\alpha = 2$

Figure 3: *Display of the decay of the energy* $\mathcal{E}$ *defined in* (7) *as a function of* k. *Star bullets display* $\log((\mathcal{E}(f_k, w_k) - \mathcal{E}(f_\infty, w_\infty))/\mathcal{E}(f_0, w_0))$ *while circle bullets display* $\log((\mathcal{E}(f_{k+1}, w_k) - \mathcal{E}(f_\infty, w_\infty))/\mathcal{E}(f_0, w_0))$.

Figure 4, left and center, compares the results obtained using $\alpha = 2$ and $\alpha = 1$. The non-local total variation prior ($\alpha = 1$) improves the inpainting results slightly. Figure 4, right, shows the result using the initial method proposed in [47], where the non-local prior J_w defined in (11)is replaced by pixel-wise comparison J_w^{graph} defined in (5). The proposed prior improves the inpainting results.

$\alpha = 2$ (23.43dB) $\alpha = 1$ (23.65dB) $\alpha = 1$, J_w^{graph} (23.05dB)

Figure 4: *Left and center: comparison of our reconstruction for* $\alpha = 2$ *and* $\alpha = 1$. *Right: reconstruction using the graph energy* J_w^{graph}.

Figure 5 shows some numerical examples of inpainting on images where 80% of

the pixels have been damaged, so that $|\Omega|/n = 0.8$. The wavelets method performs better than total variation in term of PSNR but tends to introduce some ringing artifacts. Non-local total variation performs better in term of PSNR and is visually more pleasing since edges are better reconstructed.

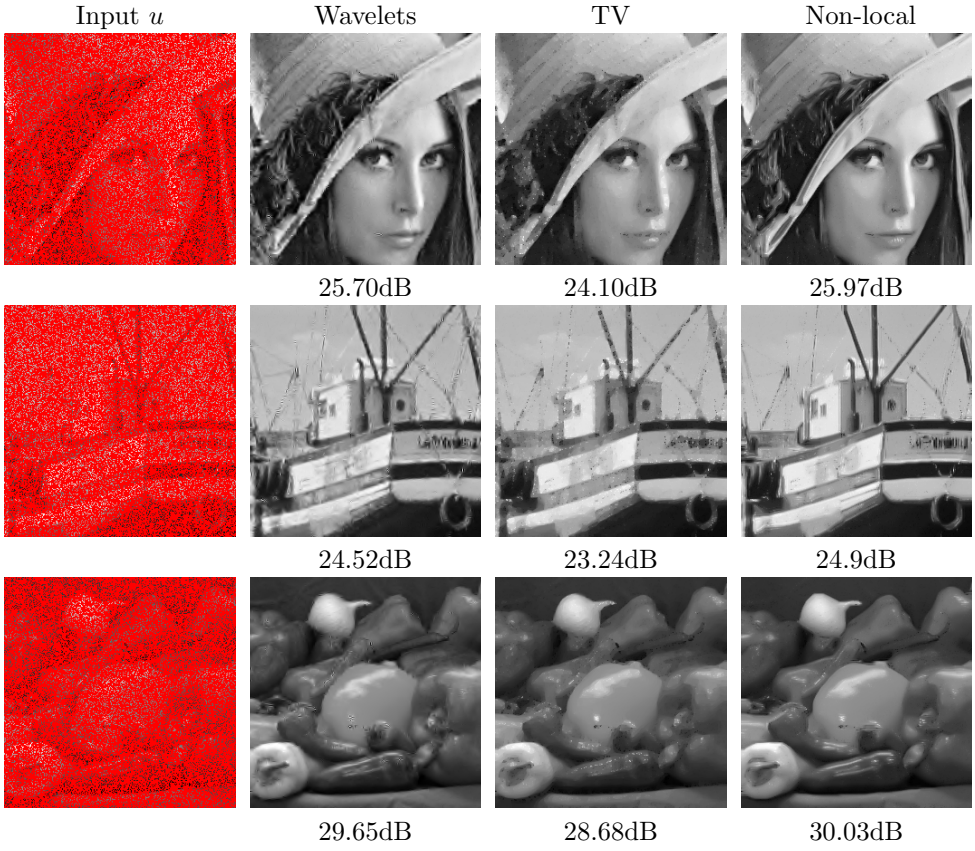


Figure 5: *Examples of inpainting where* Ω *occupates* 80% *of pixels. The original images* f_0 *are displayed on the left of figure 8.*

Similarly to other patch-regularizations [37], this method is not well adapted to inpaint large missing regions, where the hole is larger than the size of the patches. A similar issue appears with dictionary learning approaches. Note that both TV and wavelets also perform poorly (they produce a blurry inpainted region). For this setting, that corresponds to a texture synthesis problem, one should use patch recopy methods popular in computer graphics such as for instance [17]. Note also that for $\alpha = 1$, our algorithm is equivalent to the one proposed in [28].

In the following, we focus on the non-local total variation, and we thus present results obtained using $\alpha = 1$.

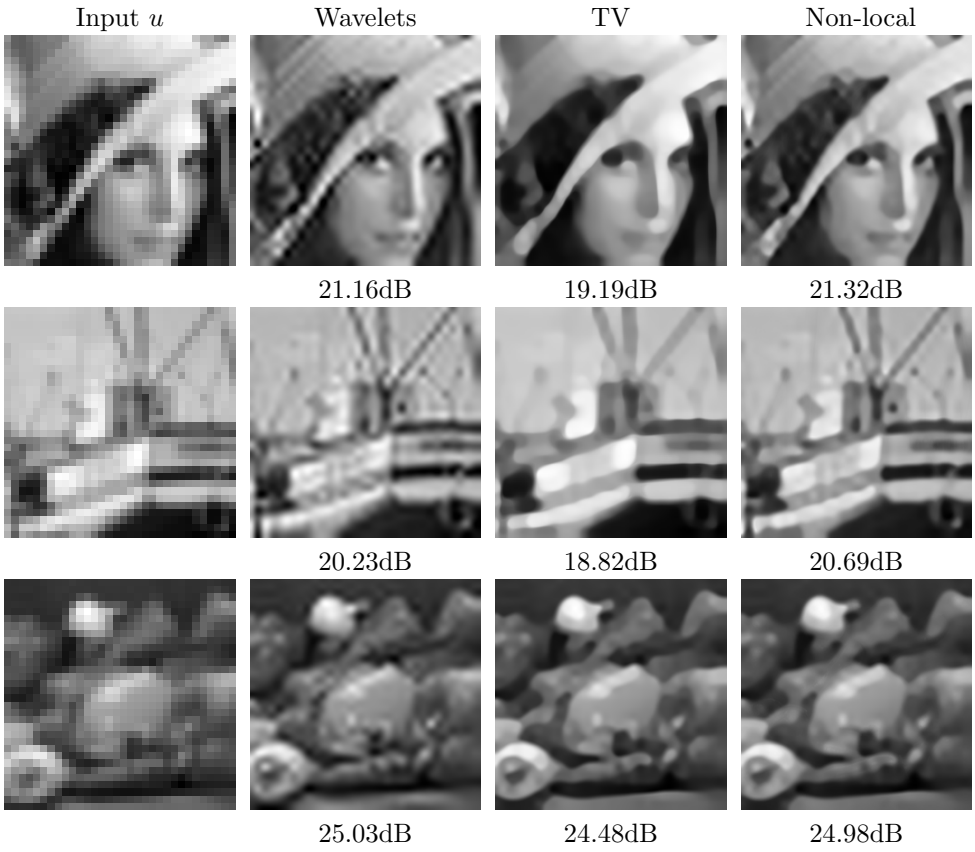


Figure 6: *Examples of image super-resolution with a down-sampling $k = 8$. The original images f_0 are displayed on the left of figure 8.*

5.2 Super-resolution

Super-resolution corresponds to the recovery of a high-definition image from a filtered and sub-sampled image. It is usually applied to a sequence of images in video, see the review papers [44, 30]. We consider here the problem of increasing the resolution of a single still image, which corresponds to the inversion of the operator

$$\forall\, f \in \mathbb{R}^n, \quad \Phi f = (f * h) \downarrow^k \tag{34}$$

where $*$ is the discrete convoluton, $p = n/k^2$, $h \in \mathbb{R}^n$ is a low-pass filter and $\downarrow^k : \mathbb{R}^n \to \mathbb{R}^p$ is the sub-sampling operator by a factor k along each axis.

For a symmetric filter h, the dual operator is given by

$$\forall\, g \in \mathbb{R}^p, \quad \Phi^* g = (g \uparrow^k) * h$$

where $\uparrow^k : \mathbb{R}^p \to \mathbb{R}^n$ corresponds to the insertion of $k-1$ zeros along horizontal and vertical directions. In this experiment, we used a Gaussian kernel

$$h(x) = e^{-\frac{\|x\|^2}{2\sigma^2}}.$$

Figure 6 shows some graphical results of the three tested super-resolution methods, for $k = 8$ and $\sigma = 2$. The results are comparable or slightly better than wavelet inpainting.

When $k = 1$, the super-resolution problem is called deblurring, since it corresponds to removing the blur induced by the camera point spread function h. Figure 7 shows examples of deblurring with $k = 1$ and $\sigma = 4$. In this setting, non-local regularization improves with respect to both total variation and wavelets.

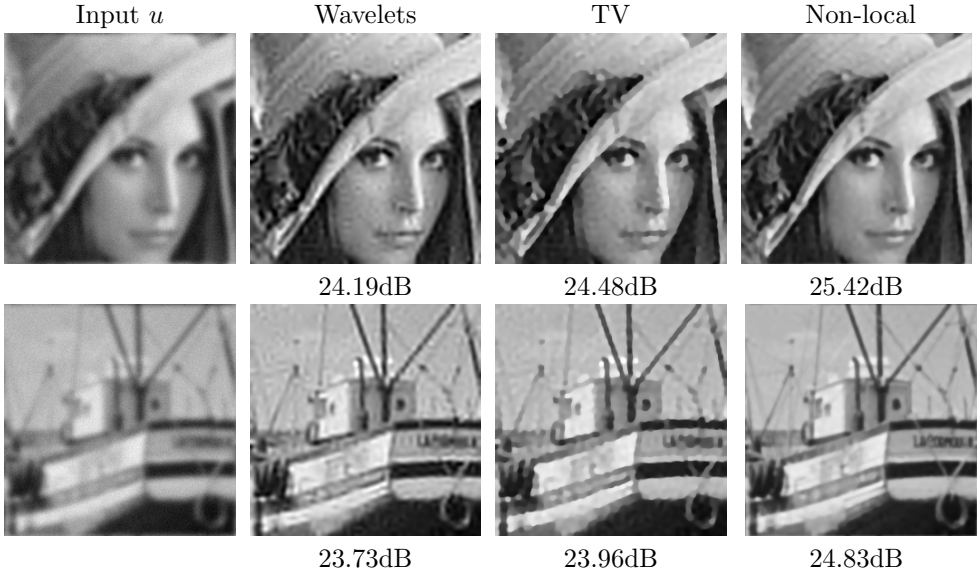


Figure 7: *Examples of image deblurring with a Gaussian kernel with $\sigma = 3$. The original images f_0 are displayed on the left of figure 8.*

5.3 Compressive Sensing

Compressive sensing is a new sampling theory that uses a fixed set of linear measurements together with a non-linear reconstruction [11, 20]. The sensing operator computes the projection of the data on a set of $p \ll n$ vectors

$$\Phi f = \{\langle f, \varphi_i \rangle\}_{i=0}^{p-1} \in \mathbb{R}^p, \tag{35}$$

where $\{\varphi_i\}_{i=0}^{p-1}$ are the rows of Φ. For the recovery of f_0 from partial measurements u to be efficient, compressive sensing theory makes use of operators Φ that are drawn from certain random matrix distributions.

Compressive sampling theory gives hypotheses on both the input signal f_0 and the sensing vectors $\{\varphi_i\}_i$ for the sampling process $u = \Phi f_0 + \varepsilon$ to be efficiently solved using the sparsity prior (4) when $\{\psi_m\}_m$ is an orthogonal basis. The initial theory was derived for φ_i being realizations of independent Gaussian vectors [11, 20]. In this setting, if the number of measurements is of the order of the sparsity s of the signal f_0,

$$p = O(s \log(n/s)) \quad \text{where} \quad s = |\{m \setminus \langle f_0, \psi_m \rangle \neq 0\}|,$$

then the recovery error satisfies $\|f^\star - f_0\| = O(\|\varepsilon\|)$ with high probability on the randomized Φ. These results extend to approximately sparse signals, such as for instance signals that are compressible in an orthogonal basis.

Gaussian matrices cannot be handled numerically for large imaging problems. Rudelson and Vershinin [49]show that compressed sensing results extend to the setting where Φf selects random entries from an orthogonal transform Hf of f, if $H \in \mathbb{R}^{n\times n}$ and $\{\psi_m\}_m$ enjoy some incoherence property. In our numerical experiments, we use H a 2-D orthogonal Hadamard transform, see [51] for a definition of the Hadamard transform and its fast implementation. To increase this incoherence when $\{\psi_m\}_m$ is a wavelet basis, following for instance [22], we consider the following sampling operator

$$\Phi f = (P_1 H P_2 f)\downarrow_{[p]},$$

where P_1 and P_2 are independent realizations of a random permutation of the n entries of a vector in $\mathbb{R}^n$, and $\downarrow_{[p]}$ selects the p first entries of a vector. Such a random sensing operator is computed in $O(n\log(n))$ operations, which is important to process high dimensional data.

The dual operator is given by

$$\forall\, g \in \mathbb{R}^p, \quad \Phi^* g = P_2^* H^* P_1^* (g \uparrow_{[p]})$$

where $\uparrow_{[p]}\colon \mathbb{R}^p \to \mathbb{R}^n$ happens $n-p$ zeros at the end of a vector, P_1^*, P_2^* are the reversed permutations, and H^* is the inverse Hadamard transform.

Figure 8 shows examples of compressive sampling reconstructions. The results are consistently better than both translation wavelets and total variation regularizations.

Conclusion

This paper proposed a new framework for the non-local resolution of linear inverse problems. The variational minimization computes iteratively an adaptive non-local graph that enhances the geometric features of the recovered image. We proved the convergence of the algorithm, up to a sub-sequence, to a stationary point of a non-convex energy. Numerical tests show that this method improves over some state-of-the-art methods for inpainting, super-resolution and compressive sampling.

Acknowledgments

This work was partially supported by ANR grants SURF-NT05-2_45825 and NatImages ANR-08-EMER-009.

References

[1] A. Adams, N. Gelfand, J. Dolson, and M. Levoy. Gaussian KD-trees for fast high-dimensional filtering. *ACM Transactions on Graphics*, 28(3), 2009.

[2] J.-F. Aujol. Some first-order algorithms for total variation based image restoration. *J. Math. Imaging Vis.*, 34-3:307–327, July 2009.

[3] J.-F. Aujol, S. Ladjal, and S. Masnou. Exemplar-based inpainting from a variational point of view. *SIAM Journal on Mathematical Analysis*, 42(3):1246–1285, 2010.

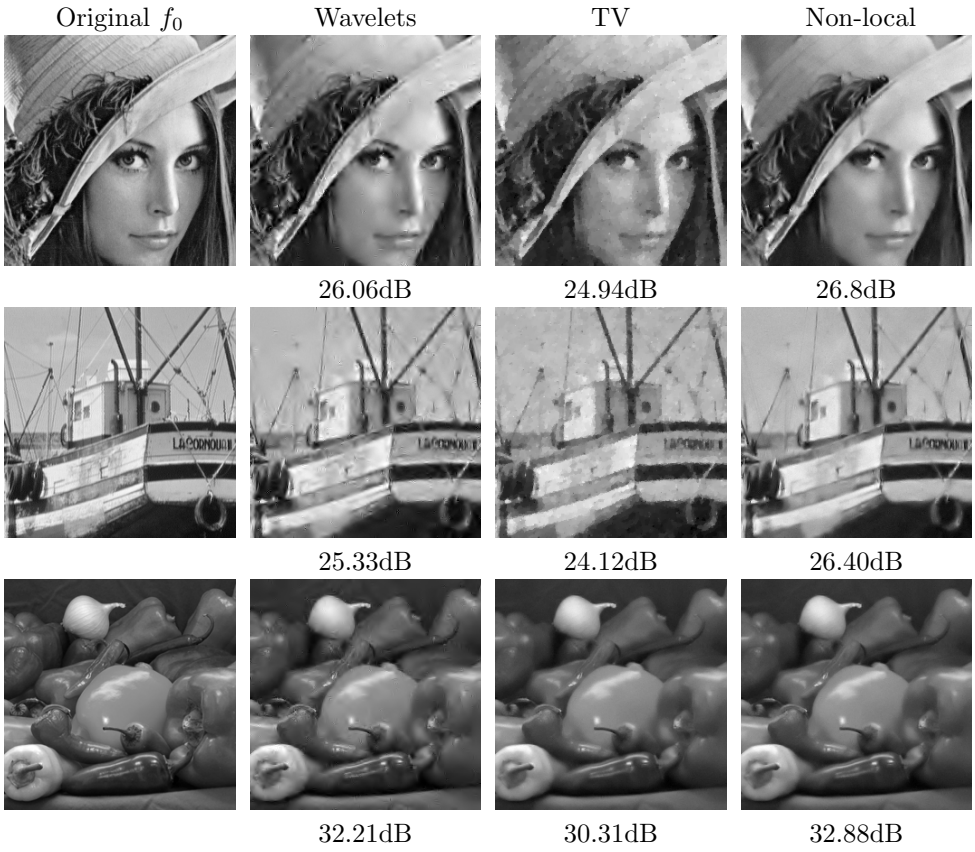


Figure 8: *Examples of compressive sensing reconstruction with $p = n/8$.*

[4] M. Avriel. *Nonlinear Programming: Analysis and Methods*. Dover Publishing, 2003.

[5] C. Ballester, M. Bertalmìo, V. Caselles, G. Sapiro, and J. Verdera. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans. Image Processing*, 10(8):1200–1211, August 2001.

[6] J. Bect, L. Blanc Féraud, G. Aubert, and A. Chambolle. A ℓ_1-unified variational framework for image restoration. In *Proc. of ECCV04*, pages Vol IV: 1–13. Springer-Verlag, 2004.

[7] M. Bertalmìo, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Siggraph 2000*, pages 417–424, 2000.

[8] A. Buades, B. Coll, and J. M. Morel. On image denoising methods. *SIAM Multiscale Modeling and Simulation*, 4(2):490–530, 2005.

[9] A. Buades, B. Coll, and J-M. Morel. Image enhancement by non-local reverse heat equation. *Preprint CMLA 2006-22*, 2006.

[10] A. Buades, B. Coll, J-M. Morel, and C. Sbert. Self similarity driven demosaicking. *IEEE Trans. Image Proc.*, 18(6):1192–1202, 2009.

[11] E. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.

[12] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1–2):89–97, 2004.

[13] T. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math*, 62(3):1019–1043, 2002.

[14] P. G. Ciarlet. *Introduction to Numerical Linear Algebra and Optimisation*. Cambridge University Press, Cambridge, 1989.

[15] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc. of the Nat. Ac. of Science*, 102(21):7426–7431, May 2005.

[16] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.

[17] A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13(9):1200–1212, 2004.

[18] D. Datsenko and M. Elad. Example-based single image super-resolution: A global map approach with outlier rejection. *Journal of Mult. System and Sig. Proc.*, 18(2–3):103–121, 2007.

[19] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math*, 57(11):1413–1541, 2004.

[20] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[21] D. Donoho and I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81(3):425–455, Dec 1994.

[22] D. Donoho, Y. Tsaig, I. Drori, and J-L. Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. *Preprint*, 2006.

[23] M. Ebrahimi and E.R. Vrscay. Solving the inverse problem of image zooming using 'self examples'. In *ICIAR07*, pages 117–130, 2007.

[24] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *Proc. of ICCV '99*, page 1033. IEEE Computer Society, 1999.

[25] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. on Image Processing*, 15(12):3736–3745, 2006.

[26] M. Elad, J.-L Starck, D. Donoho, and P. Querre. Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). *Journal on Applied and Computational Harmonic Analysis*, 19:340–358, 2005.

[27] A. Elmoataz, O. Lezoray, and S. Bougleux. Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing. *IEEE Tr. on Image Processing*, 17(7):1047–1060, 2008.

[28] G. Facciolo, P. Arias, V. Caselles, and G. Sapiro. Exemplar-based interpolation of sparsely sampled images. *IMA Preprint Series # 2264*, 2009.

[29] M.J. Fadili, J.-L. Starck, and F. Murtagh. Inpainting and zooming using sparse representations. *The Computer Journal*, 52(1):64–79, 2009.

[30] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Advances and challenges in super-resolution. *Int. Journal of Imaging Sys. and Tech.*, 14(2):47–57, 2004.

[31] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, 2002.

[32] G. Gilboa, J. Darbon, S. Osher, and T.F. Chan. Nonlocal convex functionals for image regularization. *UCLA CAM Report 06-57*, 2006.

[33] G. Gilboa and S. Osher. Nonlocal linear image regularization and supervised segmentation. *SIAM Multiscale Modeling and Simulation*, 6(2):595–630, 2007.

[34] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *SIAM Multiscale Modeling & Simulation*, 7(3):1005–1028, 2008.

[35] S. Kindermann, S. Osher, and P. W. Jones. Deblurring and denoising of images by nonlocal functionals. *SIAM Mult. Model. and Simul.*, 4(4):1091–1115, 2005.

[36] M. Mahmoudi and G. Sapiro. Fast image and video denoising via nonlocal means of similar neighborhoods. *IEEE Signal Processing Letters*, 12(12):839–842, December 2005.

[37] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Trans. Image Proc.*, 17(1):53–69, 2008.

[38] F. Malgouyres and F. Guichard. Edge direction preserving image zooming: A mathematical and numerical analysis. *SIAM Journal on Numer. An.*, 39(1):1–37, February 2001.

[39] S. Mallat. *A Wavelet Tour of Signal Processing, 3^{rd} edition.* Academic Press, San Diego, 2008.

[40] S. Masnou. Disocclusion: a variational approach using level lines. *IEEE Trans. Image Processing*, 11(2):68–76, February 2002.

[41] M. Mignotte. A non-local regularization strategy for image deconvolution. *Pattern Recognition Letters*, 29(16):2206–2212, December 2008.

[42] Y. Nesterov. Smooth minimization of non-smooth functions. *Math. Program.*, 103(1, Ser. A):127–152, 2005.

[43] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive-field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, June 1996.

[44] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20(3):21–36, 2003.

[45] G. Peyré. Image processing with non-local spectral bases. *SIAM Multiscale Modeling and Simulation*, 7(2):703–730, 2008.

[46] G. Peyré. Sparse modeling of textures. *J. Math. Imaging Vis.*, 34(1):17–31, 2009.

[47] G. Peyré, S. Bougleux, and L. D. Cohen. Non-local regularization of inverse problems. In D. A. Forsyth, P. H. S. Torr, and A. Zisserman, editors, *ECCV'08*, volume 5304 of *Lecture Notes in Computer Science*, pages 57–68. Springer, 2008.

[48] G. Peyré, J. Fadili, and J-L. Starck. Learning the morphological diversity. *SIAM Journal on Imaging Sciences*, to appear, 2010.

[49] M. Rudelson and R. Vershynin. On sparse reconstruction from fourier and gaussian measurements. *Commun. on Pure and Appl. Math.*, 61(8):1025–1045, 2008.

[50] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.

[51] J. Shanks. Computation of the fast walsh-fourier transform. *IEEE Transactions on Computers*, C(18):457–459, 1969.

[52] S. M. Smith and J. M. Brady. SUSAN - a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78, May 1997.

[53] A. Spira, R. Kimmel, and N. Sochen. A short time beltrami kernel for smoothing images and manifolds. *IEEE Trans. Image Processing*, 16(6):1628–1636, 2007.

[54] A. D. Szlam, M. Maggioni, and R. R. Coifman. Regularization on graphs with function-adapted diffusion processes. *Journal of Machine Learning Research*, 9:1711–1739, 2008.

[55] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. of ICCV '98*, pages 839–846, 1998.

[56] D. Tschumperlé and R. Deriche. Vector-valued image regularization with PDEs: Acommon framework for different applications. *IEEE Trans. Pattern Anal. Mach. Intell*, 27(4):506–517, 2005.

[57] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3):475–494, 2001.

[58] L-Y. Wei and M. Levoy. Fast texture synthesis using tree-structured vector quantization. In *Proc. of SIGGRAPH '00*, pages 479–488. ACM Press/Addison-Wesley Publishing Co., 2000.

[59] L. P. Yaroslavsky. *Digital Picture Processing — an Introduction*. Springer, Berlin, 1985.

[60] X. Zhang, M. Burger, X. Bresson, and S. Osher. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM Journal on Imaging Sciences*, 3(3):253–276, 2010.

[61] D. Zhou and B. Scholkopf. Regularization on discrete spaces. In W. G. Kropatsch, R. Sablatnig, and A. Hanbury, editors, *German Pattern Recognition Symposium*, volume 3663, pages 361–368. Springer, 2005.