Non-local Regularization of Inverse Problems

Gabriel Peyré, Sébastien Bougleux, Laurent D. Cohen

Inverse Problems and Imaging
Volume X, No. 0X, 200X, X–XX

Web site: http://www.aimSciences.org

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

(Communicated by the associate editor name)

Abstract. This article proposes a new framework to regularize linear inverse problems using a total variation prior on an adapted non-local graph. The 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 particularly 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 translation invariant wavelet-based methods.

1. Introduction

1.1. Inverse Problems Regularization. This paper is focussed on the solution of linear ill-posed inverse problems in image processing. In this setting, one tries 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. Sections 1.1.1, 1.1.2 and 1.1.3 detail three examples of inverse problems we consider in our numerical experiments.

In order 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 theory assumes that f_0

2000 *Mathematics Subject Classification:* Primary: 68U10, 94A08; Secondary: 49N45.

Key words and phrases: Non-local regularization, inpainting, super-resolution, compressive sensing.

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

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 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 $J(f)$ is small when f is close to the smoothness model, and 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.

1.1.1. *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 [43, 27]. We consider here the problem of increasing the resolution of a single still image, which corresponds to the inversion of the operator

$$\forall\, f \in \mathbb{R}^n, \quad \Phi f = (f * h)\downarrow^k \tag{2}$$

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.

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

$$(\Phi f)(x) = \begin{cases} 0 & \text{if} \quad x \in \Omega, \\ f(x) & \text{if} \quad x \notin \Omega, \end{cases} \tag{3}$$

where $\Omega \subset \{0, \ldots, \sqrt{n}-1\}^2$ is the region where the input data has been damaged.

1.1.3. *Compressive Sensing.* Compressive sensing is a new sampling theory that uses a fixed set of linear measurements together with a non-linear reconstruction [9, 17]. The sensing operator computes the projection of the data on a set of $p \ll n$ vectors

$$\Phi f = \{\langle f, \varphi_i \rangle\}_{i=0}^{p-1} \in \mathbb{R}^p, \tag{4}$$

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.

1.2. SMOOTHNESS AND SPARSITY PRIOR MODELS. The simplest prior model assumes an uniform smoothness of the image, and use for instance a discretized Sobolev norm

$$J^{\mathrm{sob}}(f) = \sum_x \|\nabla_x f\|^2,$$

where $\nabla_x f$ 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 [49] proposed to use the total variation norm for denoising purpose, when $\Phi = \mathrm{Id}$

$$J^{\mathrm{tv}}(f) = \sum_x \|\nabla_x f\|. \tag{5}$$

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

$$J^{\mathrm{spars}}(f) = \sum_m |\langle f, \psi_m \rangle|. \tag{6}$$

This prior has been introduced by Donoho and Johnstone [18] 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.

1.2.1. *Smoothness and Sparsity for Inverse Problems.* Total variation regularization prior (5) has been used to solve super-resolution [35] and inpainting of small holes [11]. Inpainting of larger holes requires higher order regularizations, that take into account the curvature of the level lines [39, 3, 5] or a tensor diffusion [55].

Sparsity prior (6) has been used to solve general inverse problems, see for instance [16, 13] and the references therein. It can also be used in conjunction with redundant frames instead of orthogonal bases, see for instance [23]. 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 (6), see for instance [26].

1.2.2. *Sparsity for Compressive Sensing.* Compressive sampling theory gives hypotheses on both the input signal f_0 and the sensing vectors $\{\varphi_i\}_i$ for the non-uniform sampling process $u = \Phi f_0 + \varepsilon$ to be efficiently solved using the sparsity prior (6). In particular, the $\{\varphi_i\}_i$ must be incoherent with the orthogonal basis $\{\psi_m\}_m$ used for the sparsity prior, which is the case with high probability if they are drawn randomly from certain random distributions. Under the additional condition that f_0 is sparse in an orthogonal basis $\{\psi_m\}_m$

$$\#\{m \setminus \langle f_0, \psi_m \rangle \neq 0\} \leqslant s$$

then the optimization of (1) using the energy (6) leads to a recovery with a small error $\|f^\star - f_0\| \approx \|\varepsilon\|$ if $p = O(s \log(n/s))$. These results extend to approximately sparse signals, such as for instance signals that are compressible in an orthogonal basis.

1.3. Non-local Prior Models.

1.3.1. *Edge-aware 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 [57], the bilateral filter [54], Susan filter [51] and Beltrami flow [52]. The non-local means filter [6] goes one step further by averaging pixels that can be arbitrary far away, using a similarity measure based on distance between patches.

1.3.2. *Non-local regularization.* As shown for instance in [44], for denoising $\Phi = \mathrm{Id}$, these edge adaptive priors correspond to one step of gradient descent of (1) using a graph based regularization over the image

$$J_w^{\mathrm{graph}}(f) = \sum_{x,y} w_{x,y} |f(x) - f(y)|^\alpha \tag{7}$$

where $\alpha = 2$. The weights $w_{x,y}$ are computed from the input noisy image u using either a distance between the noisy pixel values $|u(x) - u(y)|$ [54, 57, 52] or a distance between the patches around x and y [6, 12, 53]. This variational denoising is related to sparsity in an adapted basis of eigenvectors of the non-local diffusion operator [53, 44]

This graph based energy (7) is generalized using an arbitrary $\alpha \geqslant 1$ which, for $\alpha = 1$, defines a non-local total variation [29, 59, 24].

1.3.3. *Non-local regularization of inverse problems.* For some class of inverse problems, the weights w can be estimated from the low dimensional observations u. This is for instance the case for inpainting small holes [31], deblurring [40, 32, 7], demosaicing [8] and segmentation [30].

We note that these approaches share similarities with exemplar-based super-resolution, see for instance see for instance [15, 28, 20]. 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 [14], that is closely related to an iterative estimation of a non-local graph. These methods are related to texture synthesis with patch recopy [21, 56], and can be re-casted as a non-convex variational problem [2].

In this paper, we extend the idea of iteratively estimating the graph w for general inverse problems. The quality of the graph thus improves as the algorithm progress toward the regularized solution of the inverse problem.

1.3.4. *Harmonic analysis with adaptive representations.* Non-local diffusion methods are also related to adaptive decompositions in dictionaries of orthogonal bases. The bandlets best basis decomposition [33, 38] re-transforms the wavelet coefficients of an image to better capture edges. The grouplet transform of Mallat [37] does a similar retransformation but makes use of an adaptive geometric flow that is well suited to capture oriented oscillating textures [46].

Another class of approaches adapts the representation by learning the set of atoms from a given set of examples, see for instance [42]. This leads to redundant representations that leads to state of the art image denoising results [22]. Dictionaries can also be learned iteratively to perform image inpainting [34]and simultaneous image separation and inpainting [48].

2. Recent Works and Contributions

2.1. Related works on non-local regularization of inverse problems. 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 [58], 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 [25], where a variational justification for our initial choice of weights [47] is given. This paper extends our initial method [47], and makes use of the weights regularization of [25] to obtain a self-contained non-local regularization framework.

2.2. 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.

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

$$J(f) = J_w(f) + \gamma E(w) \tag{8}$$

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

$$f^\star \in \operatorname*{argmin}_{f\in\mathbb{R}^n, w\in\mathcal{C}} \frac{1}{2\lambda}\|u - \Phi f\|^2 + J_w(f) + \gamma E(w), \tag{9}$$

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

3.2. Patch-based Local Signatures. 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

$$\forall\, t \in \{-(\tau-1)/2+1, \dots, (\tau-1)/2\}^2, \qquad \pi_x(f)(t) = f(x+t) \tag{10}$$

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

A local signature of dimension $q \leqslant \tau^2$ is obtained from $p_x(f)$ using an orthogonal projector $U \in \mathbb{R}^{q\times\tau^2}$ that reduces the dimensionality of the patch

$$p_x(f) = U\pi_x(f), \tag{11}$$

where U satisfies $UU^* = \mathrm{Id}$. 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 local signature framework extends to color images f of n pixels by considering patches of dimension $3\tau^2$.

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 a weight $w_{x,y} \geqslant 0$. We further impose that these weights correspond to a probability distribution and that the graph only connects pixels that are not too far away

$$\mathcal{C} = \left\{ w \setminus w_{x,y} \geqslant 0,\ \sum_y w_{x,y} = 1, \quad \text{and} \quad \|x-y\| > \rho \ \Rightarrow\ w_{x,y} = 0 \right\}. \tag{12}$$

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 signatures should be compared in the image, and leads to the following non-local regularization functional

$$J_w(f) = \sum_{\|x-y\|\leqslant\rho} w_{x,y}\|p_x(f) - p_y(f)\|. \tag{13}$$

This energy is similar to the non-local total variation introduced by several authors [29, 59, 24] for denoising purpose. A major difference with these works is that J_w uses patch variations $\|p_x(f) - p_y(f)\|$ instead of pixel variations $|f(x) - f(y)|$. Such patch variations were already considered by Peyré for texture synthesis [45] and for inpainting by Facciolo et al. [25].

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. Following [25], we force the graph to have a large entropy by defining the following graph energy

$$E(w) = \sum_{x,y} w_{x,y} \log(w_{x,y}).$$

The parameter γ in (8) weights the influence of this entropy constraint on w and should be adapted to the geometry of the image to recover and on the noise level. In particular, choosing γ which tends to zero imposes a degenerate weight distribution that is a Dirac mass.

4. NON-LOCAL REGULARIZATION ALGORITHM

While the optimization problem (9) is separately convex with respect to f and to w, it is not jointly convex in (f, w). The minimization of (9) 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. NON-LOCAL PATCH OPERATORS. In order to derive the optimization procedure, we re-write the graph-based prior using operator notations.

The signature extraction process (10) defines a mapping from the pixel domain $\mathbb{R}^n$ to the signature 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 signatures $\{p_x(f)\}_x \in \mathcal{P}$ is stored as a matrix of $q \times n$ elements. The adjoint mapping is defined as

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

where the sum is restricted to pixels $y = (y_1, y_2)$ such that $-\tau \leqslant x_1 - y_1 \leqslant \tau$ and $-\tau \leqslant x_2 - y_2 \leqslant \tau$. 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 (13) is a vectorial ℓ^1 norm

$$J_w(f) = \|G_w P f\|_1,$$

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

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

A signature differential $\{d_{x,y}\}_{\|x-y\| \leqslant \rho} \in \mathcal{D}$ is stored as an array of size $q \times n \times (A\rho^2)$, where $A\rho^2$ is the number of pixels in the ball $\{x\}_{\|x\| \leqslant \rho}$ of radius ρ. The adjoint of

this signature-valued gradient is a signature-valued divergence

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

The ℓ^1 norm of signature differentials $d = \{d_{x,y}\}_{\|x-y\|\leqslant\rho}$ is defined as

$$\|d\|_1 = \sum_{\|x-y\|\leqslant\rho} \|d_{x,y}\|.$$

In the following we also make use of the ℓ^∞ norm,

$$\|d\|_\infty = \max_{\|x-y\|\leqslant\rho} \|d_{x,y}\|.$$

4.2. Optimization on the Graph. If f is fixed, optimizing (9) with respect to w defines an optimal graph $w(f)$, which is the solution of the following strictly convex optimization problem

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

As noticed in [25], the optimal graph corresponds to exponential weights

$$w(f)_{x,y} = \frac{\tilde{w}_{x,y}}{Z_x} \quad \text{where} \quad \tilde{w}_{x,y} = \left\{ \begin{array}{ll} e^{-\frac{\|p_x(f)-p_y(f)\|}{\gamma}} & \text{if } \|x-y\| \leqslant \rho, \\ 0 & \text{otherwise,} \end{array} \right. \tag{14}$$

where the normalizing constant is

$$Z_x = \sum_y \tilde{w}_{x,y}.$$

We note that these weights differ from those used in NL-means, that are Gaussian weights [6].

4.3. Optimization on the Image. If w is fixed, optimizing (9) with respect to f defines an optimal image $f(w)$ which is a solution of the following convex optimization problem

$$f(w) \in \operatorname*{argmin}_{f\in\mathbb{R}^n} \frac{1}{2}\|u - \Phi f\|^2 + \lambda\|G_w P f\|_1. \tag{15}$$

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 [13, 4], Nesterov's algorithm [41, 1] and Bregman iterations [58]. In this paper, we use a 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

$$\tilde{f}^{(\ell)} = f^{(\ell)} + \nu\Phi^*(u - \Phi f^{(\ell)}) \tag{16}$$

and a proximal denoising correction

$$f^{(\ell+1)} = \operatorname{prox}_{\nu\lambda J_w}(\tilde{f}^{(\ell)}), \tag{17}$$

where the proximal operator is the solution of a denoising problem

$$\operatorname{prox}_{\omega J_w}(\tilde{f}) = \operatorname*{argmin}_{f\in\mathbb{R}^n} \frac{1}{2}\|f - \tilde{f}\|^2 + \omega\|G_w P f\|_1 \tag{18}$$

These forward backward iteration converge to a solution $f(w)$ of (15) as long as $0 < \nu < 2/\|\Phi^*\Phi\|$, see [13].

4.4. COMPUTATION OF THE PROXIMAL OPERATOR. The proximal operator (18) is the solution of a convex functional that generalizes the total variation denoising. Following Chambolle [10], we compute this operator through a dual optimization problem, that performs the optimization on a differential signature field $d \in \mathcal{D}$ rather than on the set of images.

Using standard duality analysis, the proximal operator is written as

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

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

$$d^\star \in \operatorname*{argmin}_{d \in \mathcal{D}, \|d\|_\infty \leqslant 1} \|\tilde{f}/\omega - P^* G_w^* d\|^2. \tag{19}$$

This constraint problem can be solved using several first order schemes, including foward-backward splitting [41] and Nesterov's algorithm [41].

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

$$\tilde{d}^{(k)} = d^{(k)} + \eta G_w P(\tilde{f}/\omega - P^* G_w^* d^{(k)}), \tag{20}$$

and a projection step on the ℓ^∞ constraint

$$d_{x,y}^{(k+1)} = \frac{\tilde{d}_{x,y}^{(k)}}{\max(1, \|\tilde{d}_{x,y}^{(k)}\|)}. \tag{21}$$

These projected gradient iterations converge to a solution $d^\star$ of (19) as long as $0 < \eta < 2/\|G_w P P^* G_w^*\|$, see [13]. The operator norm $\|G_w P P^* G_w^*\|$ is estimated numerically using a few power iterations to compute the largest singular value of $G_w P$.

4.5. THE NON-LOCAL REGULARIZATION ALGORITHM. Putting together the computation of the optimal weights $w(f)$ given in (14), the forward-backward iterations (16), (17) to compute $f(w)$ and the inner iterations (20), (21) of projected gradient to compute the proximal iteration leads to the algorithm 1 to minimize (9). This algorithm depends on two tolerance parameters tol_f and tol_d that control the outer and inner iteration loops.

Convergence of the forward-backward outer iteration on ℓ is guaranteed if the errors generated by the inner iterations on k are summable [13]. This requires to lower the tolerance $\mathrm{tol}_d = \mathrm{tol}_d^{(k)}$ as k is increasing. For the numerical experiments, we use a fixed tolerance tol_d, which in practice does not lead to convergence issue.

Since the energy to minimize is non-convex and non-smooth, little can be said about the convergence of this iterative scheme, beyond the fact that the energy (9) is decaying through the iterations. In the numerical experiments detailed in Section 5, we always observed convergence of the iterates. In fact, a small number of iterations (between 10 and 20 for natural images) are enough to obtain a stabilization of the graph parameter close to a limit value $w^\star$. Once this graph is fixed to $w = w^\star$, the regularization (9) corresponds to an ℓ^1 analysis prior, and forward backward splitting is known to converge to the solution [13].

Algorithm 1: Block coordinate descent algorithm to minimize (9).

Initialization: set $f = \Phi^* u$.
while *not converged* **do**
  Weight update: compute $w = w(f)$ using (14).
  Image update: initialize $f^{(0)} = f$, set $\ell = 0$.
  while $\|f^{(\ell+1)} - f^{(\ell)}\| > \mathrm{tol}_f$ **do**
    Gradient descent: compute $\tilde{f}^{(\ell)}$ using (16).
    Proximal correction: initialize $d^{(0)} = 0$, set $k = 0$.
    while $\|d^{(k)} - d^{(k+1)}\| > \mathrm{tol}_d$ **do**
      Gradient descent: compute $\tilde{d}^{(k)}$ using (20) with $\tilde{f} = \tilde{f}^{(\ell)}$.
      Projection: compute $d^{(k+1)}$ using (21).
    Set $f^{(\ell+1)} = \tilde{f}^{(\ell)} - \omega P^* G_w^* d^{(k+1)}$.
  Set $f = f^{(\ell+1)}$
Output: $f^\star = f$.

5. Numerical Experiments

In the numerical simulations, we consider three different regularizations:

- The total variation energy J^{tv}, defined in equation (5). An algorithm similar to algorithm 1 is used for this minimization, using the usual discrete gradient ∇_x instead of the graph gradient G_w.
- The sparsity energy J^{spars}, defined in equation (6), using a redundant tight frame of translation invariant 7-9 biorthogonal wavelets $\{\psi_m\}_m$, see [36]. 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 [26].
- The non-local total variation regularization J_w in an optimized graph, solved using algorithm 1. For this regularization, the size of the patch is set to $\tau = 5$ pixels. The parameter γ of equation (8) is set to $\tau\|f\|_\infty/10$, which gives satisfying results for the noise level we consider. The locality parameter ρ of equation (12) is fixed to $\rho = 15$ pixels. The size of the signature is set to $q = 14$, and the projector U used in (11) is obtained by considering the q leading PCA eigenvectors of random patches extracted from noise-free natural images.

Both total variation and non-local total variation require approximately the same number of iterations for a given tolerance. Non-local iterations are computationally more intensive since they require the computation of the non-local weights $\{w_{x,y}\}_y$ and the gradient operator is of much higher dimension. 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 $f^\star$

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

5.1. Inpainting. Inpainting corresponds to the making operator (3). In this case, $\Phi^* = \Phi$, and one can take a proximity step size $\nu = 1$ so that the proximal iteration (17) becomes

$$f^{(\ell+1)} = \mathrm{Prox}_{\lambda J}(\tilde{f}^{(\ell)}) \quad \text{where} \quad \tilde{f}^{(\ell)}(x) = \begin{cases} f^{(\ell)}(x) & \text{if} \quad x \in \Omega, \\ u(x) & \text{if} \quad x \notin \Omega. \end{cases}$$

Figure 1 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.

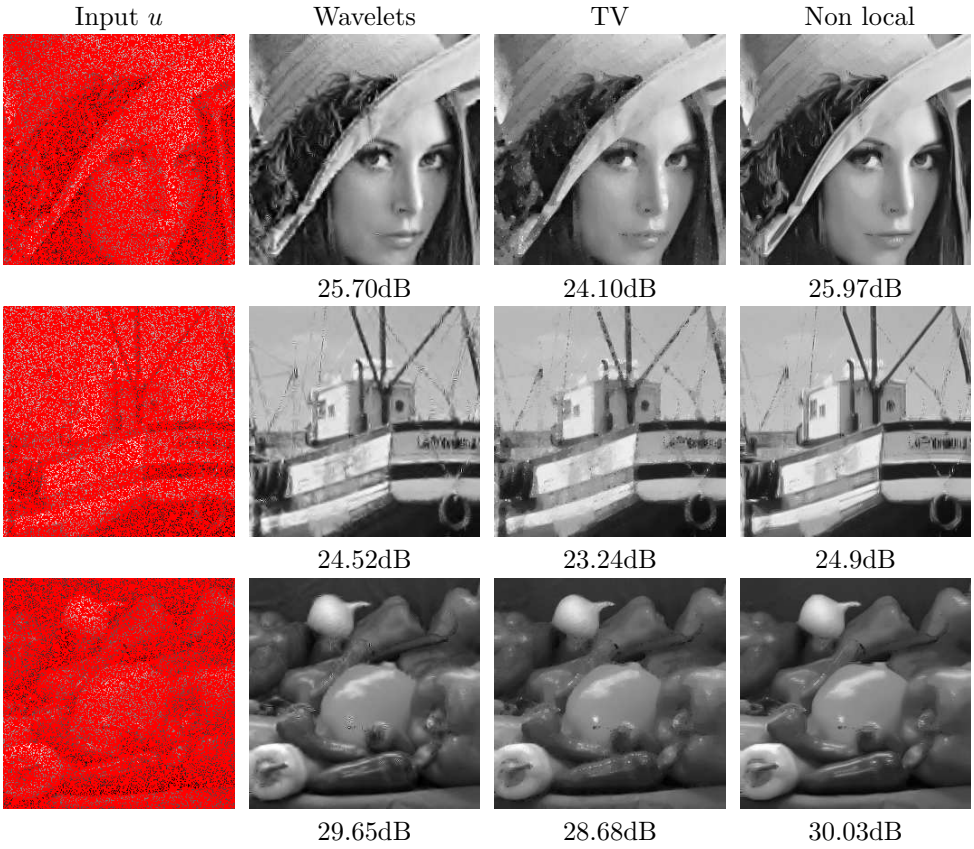

FIGURE 1. *Examples of inpainting where* Ω *occupates* 80% *of pixels. The original images* f_0 *are displayed on the left of figure 3.*

5.2. SUPER-RESOLUTION. Super-resolution inverse problem corresponds to the operator (2). 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 of standard deviation 6 pixels, and $k = 8$.

Figure 2 shows some graphical results of the three tested super-resolution methods. The results are comparable or slightly better than wavelet inpainting.

5.3. COMPRESSIVE-SAMPLING. Compressive sensing corresponds to an operator $\Phi \in \mathbb{R}^{p \times n}$ that is the realization of a random matrix distribution. In this paper, following for instance [19], we consider a fast sampling operator

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

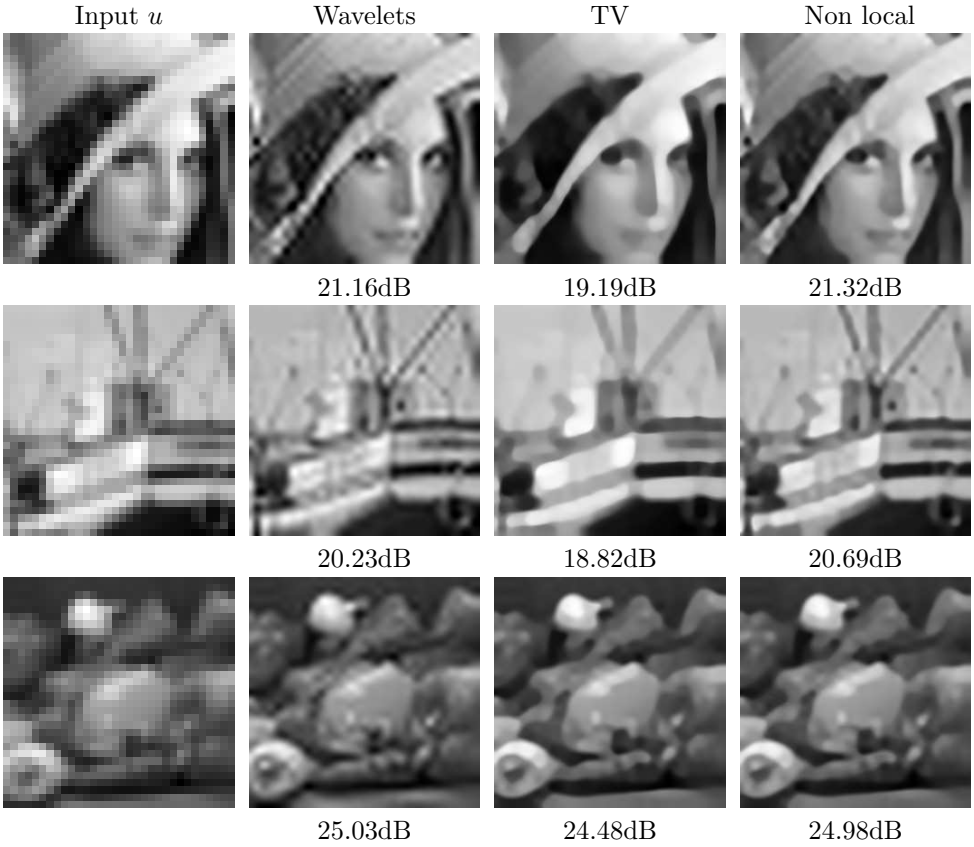

FIGURE 2. *Examples of image super-resolution with a down-sampling $k = 8$. The original images f_0 are displayed on the left of figure 3.*

where P_1 and P_2 are realizations of a random permutation of the n entries of a vector in $\mathbb{R}^n$, H is a 2D orthogonal Hadamard transform, and $\downarrow_{[p]}$ selects the p first entries of a vector, see [50] for a definition of the Hadamard transform and its fast implementation. Such a random sensing operator is computed in $O(n \log(n))$ operations, which is important to process high dimensional data.

Figure 3 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. Numerical tests show how this method improves over some state of the art methods for inpainting, super-resolution and compressive sampling.

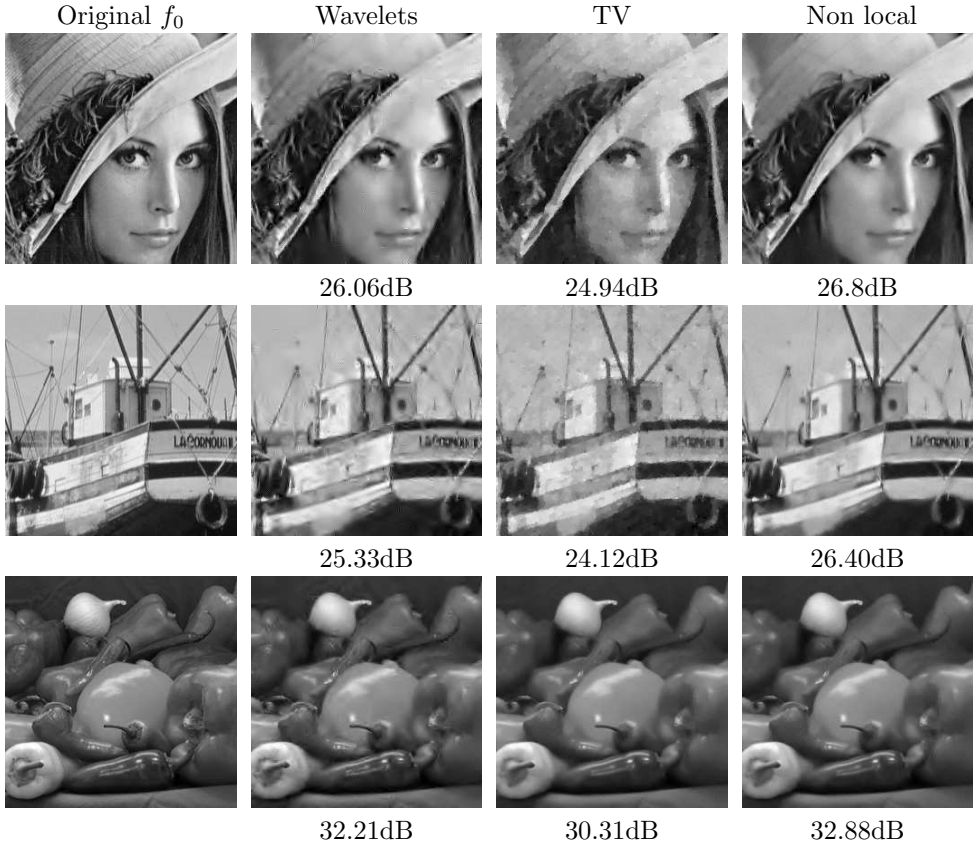

FIGURE 3. *Examples of compressive sensing reconstruction with* $p = n/8$.

REFERENCES

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

[2] J.-F. Aujol, S. Ladjal, and S. Masnou. Exemplar-based inpainting from a variational point of view. *CMLA Preprint 2008-42*, December 2008.

[3] 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.

[4] 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.

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

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

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

[8] A. Buades, B. Coll, J-M. Morel, and C. Sbert. Non local demosaicing. *Preprint 2007-15*, 2007.

[9] 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.

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

[11] T. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math*, 62:1019–1043, 2002.
[12] 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*, 102:7426–7431, May 2005.
[13] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.
[14] 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.
[15] 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.
[16] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math*, 57:1413–1541, 2004.
[17] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.
[18] D. Donoho and I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, Dec 1994.
[19] D. Donoho, Y. Tsaig, I. Drori, and J-L. Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. *Preprint*, 2006.
[20] M. Ebrahimi and E.R. Vrscay. Solving the inverse problem of image zooming using 'self examples'. In *ICIAR07*, pages 117–130, 2007.
[21] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *Proc. of ICCV '99*, page 1033. IEEE Computer Society, 1999.
[22] 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.
[23] 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.
[24] 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.
[25] G. Facciolo, P. Arias, V. Caselles, and G. Sapiro. Exemplar-based interpolation of sparsely sampled images. *IMA Preprint Series # 2264*, 2009.
[26] M.J. Fadili, J.-L. Starck, and F. Murtagh. Inpainting and zooming using sparse representations. *The Computer Journal (to appear)*, 2006.
[27] 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.
[28] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, 2002.
[29] G. Gilboa, J. Darbon, S. Osher, and T.F. Chan. Nonlocal convex functionals for image regularization. *UCLA CAM Report 06-57*, 2006.
[30] G. Gilboa and S. Osher. Nonlocal linear image regularization and supervised segmentation. *SIAM Multiscale Modeling and Simulation*, 6(2):595–630, 2007.
[31] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *UCLA CAM Report 07-23*, 2007.
[32] 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.
[33] E. Le Pennec and S. Mallat. Bandelet Image Approximation and Compression. *SIAM Multiscale Modeling and Simulation*, 4(3):992–1039, 2005.
[34] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Trans. on Image Processing*, 17(1):53–69, 2008.
[35] 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.
[36] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 1998.
[37] S. Mallat. Geometrical grouplets. *to appear in Applied and Computational Harmonic Analysis*, 2008.
[38] S. Mallat and G. Peyré. Orthogonal bandlets bases for geometric image approximation. *Communications on Pure and Applied Mathematics*, 61(9):1173–1212, 2008.

[39] S. Masnou. Disocclusion: a variational approach using level lines. *IEEE Trans. Image Processing*, 11(2):68–76, February 2002.
[40] M. Mignotte. A non-local regularization strategy for image deconvolution. *Pattern Recognition Letters*, 29(16):2206–2212, December 2008.
[41] Y. Nesterov. Smooth minimization of non-smooth functions. *Math. Program.*, 103(1, Ser. A):127–152, 2005.
[42] 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.
[43] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20:21–36, 2003.
[44] G. Peyré. Image processing with non-local spectral bases. *SIAM Multiscale Modeling and Simulation*, 7(2):703–730, 2008.
[45] G. Peyré. Sparse modeling of textures. *J. Math. Imaging Vis.*, 34(1):17–31, 2009.
[46] G. Peyré. Texture synthesis with grouplets. *to appear in IEEE Transactions on Pattern Analysis and Machine Intelligence*, 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. *Preprint Hal-00415782*, 2009.
[49] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.
[50] J. Shanks. Computation of the fast walsh-fourier transform. *IEEE Transactions on Computers*, C(18):457–459, 1969.
[51] 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.
[52] 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.
[53] A. Szlam, M. Maggioni, and R. R. Coifman. A general framework for adaptive regularization based on diffusion processes on graphs. *to appear in Journ. Mach. Learn. Res.*, July 2007.
[54] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. of ICCV '98*, page 839, 1998.
[55] 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.
[56] 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.
[57] L. P. Yaroslavsky. *Digital Picture Processing — an Introduction*. Springer, Berlin, 1985.
[58] X. Zhang, M. Burger, X. Bresson, and S. Osher. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *CAM Report 09-03*, 2009.
[59] D. Zhou and B. Scholkopf. Regularization on discrete spaces. In *German Pattern Recognition Symposium*, page 361, 2005.