Enhanced Compressed Sensing Recovery with Level Set Normals

Virginia Estellers*, Jean-Philippe Thiran*, Xavier Bresson$^\circ$

Abstract—We propose a compressive sensing algorithm that exploits geometric properties of images to recover images of high quality from few measurements. The image reconstruction is done by iterating the two following steps: 1) estimation of normal vectors of the image level curves and 2) reconstruction of an image fitting the normal vectors, the compressed sensing measurements and the sparsity constraint. The proposed technique can naturally extend to non local operators and graphs to exploit the repetitive nature of textured images in order to recover fine detail structures. In both cases, the problem is reduced to a series of convex minimization problems that can be efficiently solved with a combination of variable splitting and augmented Lagrangian methods, leading to fast and easy-to-code algorithms. Extended experiments show a clear improvement over related state-of-the-art algorithms in the quality of the reconstructed images and the robustness of the proposed method to noise, different kind of images and reduced measurements.

I. Formulation of the problem

COMPRESSED sensing (CS) is founded on the principle that, through optimization, the sparsity of a signal can be exploited to recover it from a reduced number of measurements. This simple and yet powerful idea is intriguing because it brings Shannon's sampling theorem into question. Compressed sensing is in fact the equivalent of Shannon's theorem from the point of view of sparsity: while Shannon states that to recover a band limited signal the sampling rate must be at least twice the maximum frequency present in the signal; CS relates the sparsity of a signal in certain basis with the number of measurements in another basis necessary to recover it from a minimization problem. A few definitions are necessary to understand the formulation of the CS problem.

We say that a signal $u \in \mathbb{R}^n$ is s-sparse in the basis or dictionary Ψ if it can be expressed by s non-zero coefficients in that basis, i.e. $\|\Psi u\|_0 = s$; while u is compressible if most of the energy in Ψu is contained in its largest s coefficients. Given Φ and Ψ two orthobasis or dictionaries of $\mathbb{R}^n$, the CS problem is formulated as the reconstruction of a signal $u \in \mathbb{R}^n$, sparse in basis Ψ, from $m < n$ linear measurements f in the sensing basis Φ. Ideally we should measure the n projections of u in basis Φ, that is Φu, but we only observe a small subset $f = Au$ of size $m < n$. The sampling matrix $A = R\Phi$ results from the combination of the sensing basis Φ and the matrix R that extracts the corresponding measurements

in f. Consequently, the system $f = Au$ is undetermined and the sparsity of the signal u must be exploited to "invert" the problem and obtain a correct reconstruction. The obvious strategy would be to recover the sparsest u agreeing with the measurements, that is, to solve the following non-convex problem

$$\min_u \|\Psi u\|_0 \text{ s.t } Au = f. \tag{1}$$

Problem (1) is NP-hard due to the ℓ_0 norm and only approximate solutions can be used in real applications. Relaxing the ℓ_0 norm to ℓ_1, problem (1) becomes the convex problem

$$\min_u \|\Psi u\|_1 \text{ s.t } Au = f. \tag{2}$$

Recent results in CS [1], [2] prove that (2) exactly recovers s-sparse signals with an overwhelming probability when the number of measurements is $\mathcal{O}(s \log n)$. In addition, if the sampling matrix A verifies certain restricted isometry condition, then (2) actually recovers the signal associated to the s largest coefficients of u in basis Ψ, i.e. exact recovery for s-sparse signals and recovery of the s-sparse ℓ_2 approximation for compressible signals.

When the measurements are contaminated with noise, the constraint $Au = f$ on the measurements is relaxed. In particular, under Gaussian noise the recovery is given by

$$\min_u \|\Psi u\|_1 \text{ s.t } \|Au - f\|_2 \leq \sigma, \tag{3}$$

where σ is related to the noise level. From optimization theory [3], we know that (3) is equivalent to

$$\min_u \|\Psi u\|_1 + \frac{\alpha}{2}\|Au - f\|_2^2 \tag{4}$$

in the sense that solving any of the two determines the parameter (σ, α) in the other and both have the same minimizer.

Designing the sparsifying basis depends on the signal at hand. For images a common choice are orthogonal wavelets or the discretized total variation (TV) semi-norm. TV assumes that the edges of an image are sparse and it is extensively used in inverse imaging problems as a regularizer. In its continuous formulation TV is a convex functional and its usual discretizations preserve that property. In CS $\|\Psi u\|_1$ is then substituted by the regularizing term $J(u) = \|u\|_{BV}$ in an abuse of notation.

Without loss of generality in this paper we adopt the Lagrangian formulation (4), use random Fourier samples as measurements [1] and choose total variation as sparsity criterion; but the proposed algorithm could be equally applied to

V. Estellers and J.P Thiran are with the Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne.

X. Bresson is with the Department of Computer Science, City University of Hong Kong.

[1]The proposed matrix A satisfy the restricted isometry condition with high probability and is therefore a common choice in MRI imaging [4].

other basis or dictionaries as investigated in [5]–[7]. The CS recovery problem that we consider is then

$$\min_u J(u) + \frac{\alpha}{2}\|Au - f\|_2^2. \tag{5}$$

With this formulation of CS recovery in imaging, we introduce an additional term in (5) inspired by image denoising techniques [8]. The resulting model exploits the geometry of the image to improve image recovery by aligning the normals associated to the levels sets of the image with the reconstructed signal. Our first contribution is therefore the introduction of a term for CS recovery based on geometric properties intrinsic to images. Our method can be beautifully extended to non local operators in order to recover textured images. In this case we exploit the geometry of the graph defined by the non local operators to recover finer details and structures of the images. This observation is a key contribution of our work because it can be easily adapted to improve existing non local denoising and deblurring methods, not only CS recovery. Finally, it is also important to mention that the proposed CS recovery model is based in the solution of two convex optimization problems and therefore can be efficiently solved with fast and easy-to-code algorithms.

The rest of the paper is organized as follows. After formulation the problem in this section, we present our method in Section II and explain its relation to similar techniques in Section III. Our method is then extended to non local operators in Section IV. Section V presents the associated minimization problems. Finally, experiments are presented in Section VI and conclusions drawn in Section VII.

II. CS WITH RECOVERED NORMALS

The main idea behind our method is that the recovered normals of an image can significantly improve the CS recovery results. This observation raises two main questions: how to recover normals robustly and accurately from CS measurements and how to introduce the estimated normals in the CS recovery. The answer that we propose is a two-step iterative method.

In the first step of each iteration, we estimate the normal vectors $\boldsymbol{n}$ associated to the level set curves of the image by solving a vectorial ROF [9] model that regularizes a first point-wise approximation of the level set normals $\hat{\boldsymbol{n}}$. In particular we adopt the extension of the weighted TV semi-norm for vector fields $J_w(\boldsymbol{n})$, where $w = g(|\nabla u_{k-1}|)$ is an edge detector designed to verify $w \approx 0$ near the edges and $w \approx 1$ on flat regions of u_{k-1}. Once the normals are estimated, we find an image that fits the measurements, the estimated normals and the sparsity criterion. The process is then iterated and can be summarized as

$$\begin{cases} \boldsymbol{n}_k = \arg\min_{|\boldsymbol{n}|\leq 1} J_w(\boldsymbol{n}) + \frac{\mu}{2}\|\boldsymbol{n} - \hat{\boldsymbol{n}}\|_2^2 \\ u_k = \arg\min_u J(u) + \gamma < \operatorname{div} \boldsymbol{n}_k, u > + \frac{\alpha}{2}\|Au - f\|_2^2 \end{cases} \tag{6}$$

On the following subsections we will detail each of these two steps, which both reduce to convex optimizations that can be efficiently solved. Combining the two stages into one would lead to a non convex model of higher order and the resulting minimization would be slower and suffer from local minima. A two step method is computationally more efficient in the

same way than splitting variables in Section V helps solving the minimization problems and leads to closed form solutions. The drawback of this two step procedure is the lack of rigorous theory and proof of convergence of the resulting algorithm. However, experimental results show that a single iteration of our method already improves the standard recovery (5), while the optimal performance (measured in terms of peak SNR) is attained after a few iterations. A similar limitation is present in the iterative edge-guided CS algorithm proposed in [10].

A. Estimation of level set normals

Each iteration, the normals of the image are estimated in two steps. We first obtain a noisy point-wise estimate $\hat{\boldsymbol{n}}$ from the previous solution u_{k-1} and we then regularize it to obtain $\boldsymbol{n}_k$.

Given an estimate of the underlying image u_{k-1}, the normal vectors associated to its level set curves are defined as

$$\hat{\boldsymbol{n}} = \begin{cases} \frac{\nabla u_{k-1}}{|\nabla u_{k-1}|} & \text{if } |\nabla u| > 0 \\ 0 & \text{if } |\nabla u| = 0. \end{cases} \tag{7}$$

Denoising of that first estimate of the normals $\hat{\boldsymbol{n}}$ is done with a combination of the vectorial ROF model [11] with the constraint $|\boldsymbol{n}| \leq 1$. In particular we define the vector field $\boldsymbol{n}_k = (n_x, n_y)_k$ as the solution of the following variational problem.

$$\min_{|\boldsymbol{n}|\leq 1} J_w(n_x, n_y) + \frac{\mu}{2}\|n_x - \hat{n}_x\|_2^2 + \frac{\mu}{2}\|n_y - \hat{n}_y\|_2^2 \tag{8}$$

where $J_w(n_x, n_y)$ is the extension of the weighted TV semi-norm to vector fields and $w = g(|\nabla u_{k-1}|)$ is an edge detector designed to verify $w \approx 0$ near the edges and $w \approx 1$ on flat regions of u_{k-1}.

By weighting the TV semi-norm with an edge detector $w = g(|\nabla u_{k-1}|)$, we encourage the edges of the regularized solution to coincide with the main edges of the noisy signal u_{k-1}. To be robust against false edges, we use the robust edge detector proposed by Black, Sapiro and Marimont in [12], where a statistical interpretation of the edge-stopping functions of anisotropic diffusion [13] is given. In this statistical interpretation, edges are considered outliers in the normal distribution of $|\nabla u|$ associated to noisy piece-wise constant regions and the edge-stopping functions $g(|\nabla u|)$ are derived from error norms robust to outliers. The edge detectors therefore have a parameter σ that acts as a soft-threshold in the detection of outliers and can be estimated a priori from the values of $|\nabla u|$ in the image. Based on the results of [12], we define

$$g(x) = \begin{cases} \frac{1}{2}\left(1 - \frac{x^2}{\sigma}\right)^2 & |x| \leq \sigma \\ 0 & |x| > \sigma \end{cases} \tag{9}$$

with $\sigma = 1.4826 \operatorname{median}(|\nabla u - \operatorname{median}(|\nabla u|)|)$. We refer the reader to the original publication [12] for the details of this edge detector and point out that other robust edge detection functions can be used in our formulation with similar results in the final CS reconstruction method.

The constraint $|\boldsymbol{n}| \leq 1$ in (8) corresponds to a relaxation of the condition $|\boldsymbol{n}| = 1$ inherent to the definition of normals. It

is numerically necessary in flat regions, where $\nabla u = \mathbf{0}$ and we cannot numerically normalize the gradient vector.

In the context of image denoising and inpainting [14], a combination of vectorial TV and ℓ_1 fidelity term has proven a better choice in the estimation of normal fields of images contaminated with salt-and-pepper noise. In our method, however, the first estimate of the normals $\hat{\boldsymbol{n}}$ is obtained from a CS recovery algorithm and we cannot assume any particular noise model but certain kinds of artifacts.Experimentally, we have observed that both the introduction of a weight in the vectorial TV semi-norm and the relaxation $|\boldsymbol{n}| \leq 1$ lead to more robust estimates of the normal field for their use in the second step of our reconstruction method. The use of a robust edge detector as a weighting function gives more weight to the data term in the estimation of the normals near the edges, where ∇u is large and $\hat{\boldsymbol{n}}$ is clearly defined. In regions with no clear orientation or where $\hat{\boldsymbol{n}}$ is affected by noise, the regularization tends to average neighbouring vectors and produce normal fields $\boldsymbol{n}_k$ close to zero, while the relaxation $|\boldsymbol{n}| \leq 1$ experimentally leads to small norm values $|\boldsymbol{n}_k|$ in these areas. Consequently, the value of $|\boldsymbol{n}_k|$ in non-flat areas can be considered an experimental measure of quality in the estimation of the normal field.

B. Matching normals and CS measurements

Once the normal field $\boldsymbol{n}_k$ is computed, we find an image that matches this field by including the term $- < \boldsymbol{n}_k, \nabla u >$ in the standard CS recovery model (5). This term tries to maximize the alignment of the estimated normals of the signal $\boldsymbol{n}_k$ with the normals of the reconstruction $\frac{\nabla u}{|\nabla u|}$. The resulting recovery model is

$$u_{k+1} = \arg\min_u J(u) - \gamma < \boldsymbol{n}_k, \nabla u > + \frac{\alpha}{2}\|Au - f\|_2^2 \quad . \tag{10}$$

Taking into account that the divergence div is the adjoint operator of the gradient ∇, the previous minimization can be rewritten as

$$u_{k+1} = \arg\min_u J(u) + \gamma < \operatorname{div} \boldsymbol{n}_k, u > + \frac{\alpha}{2}\|Au - f\|_2^2. \tag{11}$$

Our method then exploits the geometry of the image in the recovery process and obtains better regularization properties than standard TV. In particular the proposed model preserves edges like TV, by encouraging the gradients to be sparse with $J(u)$; but is also able to recover smooth regions by aligning the gradients of the reconstruction with the smoothed normals with the term $< \boldsymbol{n}_k, \nabla u >$. We make use of the adjoint properties of the divergence and gradient operators in order to overcome the following limitation of a direct implementation of our model in smoothly varying regions. In these regions, the minimization term associated to the direct alignment of normals $- \int_\Omega < \boldsymbol{n}_k, \nabla u >$ is negligible because the reconstructed gradient ∇u is small and the data and regularity terms dominate the minimization. In our implementation, however, this term is rewritten as $\int_\Omega \operatorname{div} \boldsymbol{n}_k u$, where the smoothness and orientation of the region are summarized by the divergence of the normal field estimated in the previous step $\operatorname{div} \boldsymbol{n}_k$. The divergence function accumulates the orientation of the

normals of the different neighbours around each pixel. It has constant non-zero values in smoothly varying regions, large magnitude close to the edges of the image and it is close to zero in flat areas contaminated with noise, where the estimated normal field has no clear orientation. The magnitude of the divergence, and the weight given to the alignment term in the minimization, is then proportional to the coherence in local orientation of the estimated normal field around each pixel.

In principle we could also use a smooth estimate of the gradients $\boldsymbol{v} = \nabla u_{k-1}$ instead of $\boldsymbol{n} = \frac{\nabla u_{k-1}}{|\nabla u_{k-1}|}$ to align the gradients of the reconstructed signal. However, discontinuities of the image would have a contribution to $\boldsymbol{v}$ proportional to their jump and the resulting term $< \boldsymbol{v}, \nabla u >$ would give different weights to discontinuities of different sizes. From a geometric perspective, if we want to recover the shapes of the image independently of their contrast we need to consider the normal vectors derived from its level sets. By the use of level sets, we treat all shapes equally and the term $\gamma < \boldsymbol{n}_{k+1}, \nabla u >$ only accounts for geometric quantities. Indeed, if we make use of the co-area formula, we can rewrite the alignment term as

$$\int_\Omega \boldsymbol{n}_k \cdot \nabla u \, \mathrm{d}\boldsymbol{x} = \int_\Omega \boldsymbol{n}_k \cdot \frac{\nabla u}{|\nabla u|}|\nabla u| \, \mathrm{d}\boldsymbol{x} = \int_{-\infty}^{\infty} \int_{u^{-1}(c)} |\boldsymbol{n}_k| \cos\theta(s) \, \mathrm{d}s \, \mathrm{d}c, \tag{12}$$

where $\theta(s)$ is the angle between the normal vector to the level curve $u^{-1}(c)$ of the reconstructed image and the regularized normal vector $\boldsymbol{n}_k$. The alignment term it thus is purely sensitive to angles and its contribution is weighted by $|\boldsymbol{n}_k|$. In flat regions, where $|\boldsymbol{n}_k| = 0$, the alignment term vanishes and our reconstruction method simplifies to TV. As we have previously explained, in smoothly varying regions and close to the edges the value of $|\boldsymbol{n}_k|$ is an experimental measure of quality in the estimation of the normal field $\boldsymbol{n}_k$. This results in an alignment term of Equation (12) weighted proportionally to the confidence we have on the estimated normal field $\boldsymbol{n}_k$, which reduces the sensitivity of our method to the correct estimation of angles in noisy regions.

III. RELATED APPROACHES IN CS

The method that we propose is inspired by image denoising and inpainting methods [8], [15] that align an estimate of the normals with the reconstructed image. In the context of image denoising, Lysaker, Osher and Tai in [8] first regularize the unit gradient of the noisy image and then improve reconstruction by fitting this gradient into the regularized vector. The resulting method outperforms the ROF model [9] and similar higher order PDE methods [16]. Dong et al. in [17] improve this model by regularizing the angles instead of the vectors and introducing an edge indicator as an extra weight. In image inpainting, an equivalent two-step method is proposed by Ballester et al. in [15], later improved with the divergence free constraint by Tai in [18], [19] and adapted to image decomposition and denoising in [14], [20]. In general, processing the normals to improve reconstruction has also been used in shape from shading [21] and mesh optimization [22]. In the context of edge integration, a similar term fitting the angle between the normal vector to the level curves of the image and the normal vector of the evolving curve was proposed in

[23]. However, this information has not been exploited before for CS image recovery.

In the CS field, several methods have been proposed to improve the quality of the ℓ_1 recovery. For general signals, greedy algorithms [24]–[26] and ℓ_p $0 < p < 1$ minimizations [27], [28] approximate the solution of the ℓ_0 problem (1) and improve its sparsity; but the resulting minimizations are not convex, the algorithms are slow and suffer from local minima. To improve the sparsity of ℓ_1 recovery (5) without increasing its complexity, Candès, Wakin and Boyd [29] proposed an iterative process solving a weighted ℓ_1 problem at each iteration. The weights are defined inversely proportional to the value of the recovered signal in the previous iterate, approximating the behaviour of the ℓ_0 norm and promoting sparser signal recovery. The resulting method efficiently solves a convex problem at each iteration, experimentally improves signal recovery and has been adopted for image processing with TV regularization in the edge-guided CS of Guo and Yin [10]. Edge-guided CS incorporates information about the magnitude of the gradient in the recovery process and it is therefore related to our method. However, we propose an additive method more robust to noise and exploit both the magnitude and directional information of the gradients.

CS recovery of images has also been improved modifying the data term $\|Au - f\|_2^2$ inspired by image denoising techniques. In particular, the Bregman iterations proposed by Osher et al. in [30] for image denoising and deblurring have been applied to CS in [31]. He et al. in [31] use Bregman iterations to improve CS image recovery for phantom MRI data, but fail in the recovery of real images due to the additional difficulties of reconstructing a signal from partial measurements compared to the original denoising problem. For the particular case of TV regularisation, the first Bregman iteration has a geometric interpretation similar to the second step of our recovery method. However, Bregman iterations do not include a regularization step for the normals and therefore fail for noisy and real MRI signals.

In the following, we summarize each of these to methods and clarify their relationship with our technique.

A. Edge-guided CS

Edge-guided CS [10] improves recovery of MRI images by exploiting edge information with an iterative process that weights TV with an edge detector associated to the image recovered in the previous iteration. The key idea is that edges correspond to locations where $|\nabla u|$ is large, TV corresponds to the ℓ_1 norm of the norm of the gradient and therefore an inverse edge detector can be used to re-weight TV and approximate the ℓ_0 norm in a similar fashion to the re-weighted ℓ_1 of Candès, Wakin and Boyd [29] for general signals. The method starts with the standard CS solution (5) to obtain a first estimate of the image u_1 and its edges. It then defines the weights $w_1 = g(|\nabla u_1|)$ inversely proportional to $|\nabla u_1|$) in order to recover an image with sparser edges at the second iteration by solving the re-weighted TV problem. The

process is iterated, leading to the following two step algorithm:

$$\begin{cases} u_{k+1} = \arg\min_u J_{w_k}(u) + \frac{\alpha}{2}\|Au - f\|_2^2 \\ w_{k+1} = g(|\nabla u_{k+1}|) \end{cases} \quad (13)$$

There is no stopping criterion or guarantee of convergence for this iterative process and usually, after a few iterations the reconstruction does not improve or even degrades. In fact, the multiplicative model of edge-guided CS is very sensitive to false edge detection. In particular, if an edge is detected in a wrong location, the weight associated to it on the next iteration will encourage an edge on this location and CS recovery will degrade with any new iterations. The iterative re-weighting process is designed to improve sparsity of the signal and recovery of piecewise constant functions, but it fails in the recovery of smooth regions in images. Compared to our method, edge-guided CS incorporates only information about the magnitude of ∇u, while we also use its directional information; it does not include a regularization step for the detected edges and it is specially designed for piecewise constant images.

B. Bregman methods

We also share similarities with Bregman methods, whose original idea was to restore normals and image intensity simultaneously. However, Bregman methods cannot recover normals as accurately and robustly as our method because they do not regularize the estimated normals. Our improvement is at the price of loosing global convexity.

Bregman iterations substitute the minimization problem (5) for a sequence of convex optimizations substituting $J(u)$ for its Bregman distance to the previous iterate. In particular, the first Bregman iteration has a geometric interpretation closely related to our method. Starting with $u = 0, v = 0$, the Bregman iterative process can be summarized as

$$\begin{cases} u_{k+1} = \arg\min_u J(u) + \frac{\alpha}{2}\|f + v_k - Au\|_2^2 \\ v_{k+1} = v_k + f - Au_{k+1} \end{cases} \quad (14)$$

While their first iteration corresponds to the standard CS model (5), their second iteration implicitly exploits the normals of the image recovered at iteration one to improve the recovery. For simplicity, here we show the connection to our method with the continuous formulation, where $A(\cdot)$ is the continuous functional operator of CS and A^* its adjoint. For the first iteration $u = 0, v = 0$ and the method solves

$$u_1 = \arg\min_{u \in \mathbb{R}^n} \int_\Omega |\nabla u| + \frac{\alpha}{2}\|f - A(u)\|_2^2, \quad (15)$$

The optimality condition associated to (15) derived from its the Euler-Lagrange equation is

$$\text{div}\, \frac{\nabla u_1}{|\nabla u_1|} = -\alpha A^*(u_1)\,(f - A(u_1)) \quad (16)$$

where $\boldsymbol{n}_1 = \frac{\nabla u_1}{|\nabla u_1|}$ correspond to the normals of u_1. At the next iteration we can introduce a term $< \boldsymbol{n}_1, \nabla u >$ aligning the normals of the reconstructed signal with the estimate of the normals from the previous iteration, that is

$$u_2 = \arg\min_u \int_\Omega |\nabla u| - < \boldsymbol{n}_1, \nabla u > + \frac{\alpha}{2}\|f - A(u)\|_2^2 \quad (17)$$

Integrating by parts and substituting div $\boldsymbol{n}_1$ in (16) we have

$$-<\boldsymbol{n}_1,\nabla u>=<\operatorname{div}\boldsymbol{n}_1,u>=-<\alpha A^*\left(f-A(u_1)\right),u>=$$
$$-\alpha<f-A(u_1),A(u)>=-\alpha<v_1,A(u)>$$
(18)

with $v_1=f-A(u_1)$ as defined in (14). If we substitute (18) in the minimization (17) and group together the terms with $A(u)$, we end up with the Bregman update rule

$$u_2=\int_\Omega|\nabla u|+\frac{\alpha}{2}\|f+v_1-A(u)\|_2^2. \qquad (19)$$

For the rest of iterations the geometric interpretation of the update is lost. Compared to Bregman iterations, our method explicitly uses the normals in the recovery model for all iterations, not only the second one, and it is not restricted to TV regularization. Indeed, this geometric interpretation is only possible for the TV term $J(u)$, while our method can be used with any sparsifying basis. We are also more robust to noise thanks to the regularization step and, unlike the Bregman iteration, experimentally improve the reconstruction model (5) for both phantom and real MRI data. In addition, our method extends to non local operators to exploit graph geometry and recover fine details in textured images.

IV. EXTENSION TO NON LOCAL METHODS

Total variation regularization is designed to recover images with sharp edges but, as other methods based on local gradients, it is not suited for textured images with fine structures. In this section we extend our method to textured images using both a non local TV regularization and a term aligning the estimated non local normals with the non local gradients of the reconstructed image.

A. Non local operators

Non local TV is a variational extension of the non local means filter proposed by Buades, Coll and Morel for image denoising [32]. Non local means exploits the repetition of patterns in natural and textured images to reconstruct sharp edges as well as fine meaningful structures. That principle is the basis of non local regularization methods in imaging, which outperform the classical methods by incorporating global information in the regularization process. In [33] Gilboa and Osher use graph theory to extended the classical TV to a non local functional. In the discrete setting, Zhou and Schölkopf [34] and Elmoataz et al. [35] use graph Laplacians to define similar non local regularization operators. The resulting non local methods have been applied to image denoising [33], segmentation [36], [37], inpainting [38], deconvolution and compressive sensing [39].

We adopt the discrete formulation of the continuous model presented in [33]. In this non local framework we consider the image domain as a graph $G=(\Omega,E)$; where Ω is the set of nodes of the graph, one for each pixel in the image, and E is the set of edges connecting the nodes. The edge connecting nodes i and j is weighted with a positive symmetric weighting function $w(i,j)$ that represents the distance between the two nodes in graph terms. Consequently, two pixels i and j spatially far away in the image can be considered neighbours in the graph and interact if $w(i,j)>0$ (we write then $i\sim j$). For that reason, the resulting approach is considered non local.

Given an image u defined on the graph, the non local gradient $\nabla_G u$ at node i is defined as the vector of all directional derivatives associated to the neighbours of i, that is

$$\nabla_G u\ (i,j)=(u(j)-u(i))\sqrt{w(i,j)}\ \forall j\in\Omega. \qquad (20)$$

In the graph, vectors $\boldsymbol{d}=d(i,j)$ are therefore functions defined in the domain $\Omega\times\Omega$.

In this setting we define the standard L_2 inner product between functions as

$$<u,v>_G=\sum_{i\in\Omega}u(i)v(i). \qquad (21)$$

For vectors, we define a dot product pixel-wise

$$(\boldsymbol{d}\cdot\boldsymbol{e})_G(i)=\sum_{j\sim i}d(i,j)e(i,j) \qquad (22)$$

and an inner product on the graph

$$<\boldsymbol{d},\boldsymbol{e}>_G=\sum_i(\boldsymbol{d}\cdot\boldsymbol{e})_G(i)=\sum_i\sum_{j\sim i}d(i,j)e(i,j). \qquad (23)$$

In order to have an equivalent to the TV semi-norm, we define a norm function on the graph $|\cdot|_G$. With the previous definitions, the magnitude of a vector at node i is given by

$$|\boldsymbol{d}|_G(i)=\sqrt{(\boldsymbol{d}\cdot\boldsymbol{d})_G(i)}=\sqrt{\sum_{j\sim i}d(i,j)^2} \qquad (24)$$

The standard TV is then naturally extended to a non local version as the ℓ_1 norm of the graph norm $|\cdot|_G$ associated to the non local gradient, that is,

$$TV_G\left(u\right)=J_G\left(u\right)=\sum_i|\nabla_G u|_G(i)=\parallel|\nabla_G u|_G\parallel_1. \qquad (25)$$

With the above inner products, the non local divergence of a vector $\boldsymbol{d}$ is defined as the adjoint of the non local gradient, that is

$$\operatorname{div}_G\boldsymbol{d}\ (i)=\sum_{j\sim i}(d(i,j)-d(j,i))\sqrt{w(i,j)}. \qquad (26)$$

With these definitions, if we consider only the immediate pixels as neighbours and fix their weights to $w(i,j)=1$, then the non local TV reduces to the standard TV definition. If we consider more general neighbours by defining a correct weighting function like in [32], the non local operators incorporate global information and the standard regularization process is improved. The weight function therefore has an important impact in the performance of the non local regularizers. Inspired by [32], [33], given a reference image u_0 we compute weighting function $w_0(i,j)$ measuring the difference of patches around each node as follows

$$w_0(i,j)=\exp^{-\frac{\|\mathcal{P}_0(i)-\mathcal{P}_0(j)\|^2}{2h^2}}, \qquad (27)$$

where h is a scaling factor and $\mathcal{P}_0(i)$ is a patch of u_0 centred at pixel i. This weighting function is designed to reduce Gaussian

noise while preserving the textures of the image. The reference image should be as close as possible to the true image in order to incorporate valid information related to image structures in the non local operators. For that reason, we initialize the weighting function in the non local methods with the standard CS solution (5) (on the following u_0) and iteratively solve the non local model and update the weights with the non local solution. The basic non local CS recovery is then

$$
\begin{cases}
\nabla_{G_k} \longleftarrow \text{ estimate non local operators from } u_{k-1} \\
u_k = \arg\min_u J_{G_k}(u) + \frac{\alpha}{2}\|Au - f\|_2^2.
\end{cases}
$$
(28)

B. Proposed non local method

Symmetrizing our local technique, we propose a two step iterative method for CS recovery. In the first step of each iteration, we estimate the non local normals $\boldsymbol{n}_G$ associated to the level set curves of the image in the graph. Once the non local normals are estimated, we find an image that fits the non local normals and the CS measurements and iterate the process.

In the local setting, the normal vectors associated to the level set curves of an image u are defined as $\boldsymbol{n} = \frac{\nabla u}{|\nabla u|}$. We extend that definition to our non local framework and exploit the geometry of the image in the graph to improve CS recovery. In particular, we derive the equivalent non local normals from the non local gradient $\nabla_G u$ by normalizing its components node-wise, i.e. all the components associated to node i are normalized by $|\nabla u|_G(i)$.

Given an estimate of the non local normals $\boldsymbol{n}_G$, we can include a term in the CS reconstruction (28) maximizing the alignment of the reconstructed signal with the normals. The resulting minimization is

$$
u = \arg\min_u J_G(u) - \gamma < \boldsymbol{n}_G, \nabla_G u >_G + \frac{\alpha}{2}\|Au - f\|_2^2
$$
(29)

Exploiting the adjoint relation of the non local divergence and gradient $< \boldsymbol{n}_G, \nabla_G u >_G = - < \text{div}_G \boldsymbol{n}_G, u >_G$, we have

$$
u = \arg\min_u J_G(u) + \gamma < \text{div}_G n, u >_G + \frac{\alpha}{2}\|Au - f\|_2^2.
$$
(30)

As before, the process can be iterated and we obtain the following analogue to the previous two step procedure:

$$
\begin{cases}
\nabla_{G_k} \longleftarrow \text{ estimate non local operators from } u_{k-1} \\
\text{div } G_k \boldsymbol{n}_{G_k} = \arg\min_v J_{G_k}(v) + \frac{\mu}{2}\|v - \hat{v}\|^2 \\
u_k = \arg\min_u J_{G_k}(u) + \gamma < \text{div } G_k \boldsymbol{n}_{G_k}, u >_G + \frac{\alpha}{2}\|Au - f\|_2^2
\end{cases}
$$

with $\hat{v} = (1 - g(|\nabla_G u_{k-1}|_G)) \text{div}_G \frac{\nabla_G u_{k-1}}{|\nabla_G u_{k-1}|}$.

The third step of our non local method is naturally derived from our local version and the geometric interpretation of the non local operators. However, the regularization step of the non local normals requires careful consideration, as we explain next.

C. Estimation of non local normals

The non local gradient operator, and consequently the non local normals, do not correspond to the discretization of standard vector fields in a grid. Indeed, $\nabla_G u$ has a different number of components for each pixel and the associated direction to $\nabla_G u(i, j)$ depend on the relative position of the node i and its neighbour j. Therefore, we cannot use standard techniques to regularize these vector fields and we prefer to regularize the term $\text{div}_G \boldsymbol{n}$ posteriorly used in the recovery algorithm. Compared to the regularization of the non local normals, we loose directional information, but the resulting method is simpler and faster.

Assume that we are given an estimate of the reconstructed signal u_{k-1}. We first compute a noisy estimate of the non local normals and their divergence pixel-wise and we then denoise it with standard denoising methods. In particular, we estimate the non local normals as

$$
\hat{\boldsymbol{n}}_{\boldsymbol{G}} = \frac{\nabla_G u_{k-1}}{|\nabla_G u_{k-1}|}
$$
(31)

and compute a rough estimate of the non local divergence as

$$
\hat{v} = (1 - g(|\nabla_G u_{k-1}|_G)) \text{div}_G \hat{\boldsymbol{n}}_{\boldsymbol{G}},
$$
(32)

where $g(x)$ is a function designed to verify $g \approx 0$ when x is large and $g \approx 1$ when x is small. In fact, $g(|\nabla_G u_{k-1}|_G)$ acts as the equivalent edge detector presented in Section II-A and is defined with the same expression (9). As in the local case, we adopt the statistical interpretation of the edge detector $g(|\nabla_G u|_G)$ presented in [12], where the edges are considered as outliers in the normal distribution of $|\nabla_G u|_G$ associated to homogeneous regions. Since the edge detector g is derived from error norms robust to outliers, weighting our estimate of the normals with the function $1 - g(|\nabla_G u_{k-1}|_G)$ in (32) is equivalent to soft-thresholding the non local normals when we suspect that the variations in u_{k-1} are due to noise inside homogeneous regions.

Finally, we regularize $\hat{v}$ to obtain a smoother estimate of the non local divergence, which will be used in the second step of our iterative method. There are two natural approaches for this regularization: we can ignore the non local nature of the divergence and gradient operators and use any local model to regularize $\hat{v}$, for instance the standard ROF [9] of equation (33); or use the non local neighbours to denoise $\hat{v}$ with (34), that is, use the natural distance and neighbouring relations inherent to de definition of $\hat{v}$ to denoise it.

$$
\text{div } G_k \boldsymbol{n}_{G_k} = \arg\min_v J(v) + \frac{\mu}{2}\|v - \hat{v}\|^2
$$
(33)

$$
\text{div } G_k \boldsymbol{n}_{G_k} = \arg\min_v J_G(v) + \frac{\mu}{2}\|v - \hat{v}\|^2
$$
(34)

In our experiments we obtained slightly better results with the first approach.

V. MINIMIZATION PROBLEMS

In order to solve the minimization problems involved in each step of our method, we make use of recent advances in convex minimization [40], [41] and apply variable splitting and augmented Lagragians [42] to obtain efficient and easy-to-code algorithms. To simplify notation on this section we

remove the sub-indexes in u_k and n_k indicating the iterations of our two step procedure.

The minimizations associated to each of the local steps involve both a TV and a quadratic term similar to the ROF model [9]. Consequently, the resulting algorithms apply a similar strategy to overcome the non-linearity and non-differentiability of TV than the multitude of algorithms proposed for ROF. In the original ROF paper [9], the authors derive the Euler-Lagrange PDE of the model and propose a time marching method to solve it. The resulting method is slow due to the constraint on the time step associated to its stability conditions. In the last years more efficient methods have been proposed for the ROF model due to its extensive use in imaging. A popular class of methods is based on the dual formulations of the ROF model, e.g. Chambolle's projection method [43] or primal-dual approaches [44]–[46]. Other options are based on variable-splitting and equality constrained optimization; which is solved by quadratic-penalties [40], Bregman iterations [41], [47] or the equivalent augmented Lagrangian method [48]. In the case of CS, dual solvers are not usually adopted because they suffer from matrices A that are large-scale and dense. In particular for matrices corresponding to transforms with fast implementations (like the Fourier transform of this paper), splitting methods are a good choice because they can easily exploit fast transforms to compute Au and $A^T u$ [40], [41]. The algorithms that we propose fall in this last category. We rewrite the different problems as constraint minimizations and use augmented Lagrangians to solve them. The resulting Lagrangians are minimized with respect to each variable independently and the multipliers are then updated in a cyclic way. Since all the minimizations can be analytically solved, the resulting algorithms are extremely fast and easy to implement.

Similarly, the minimization algorithms that we propose for the non local method is closely related to the minimization of the non local ROF model proposed in [33], which was originally solved with a time consuming time marching algorithm. The non local CS problem has been solved with a combination of forward-backward splitting and Bregman iteration in [39], but for uniformity of the paper we use the same combination of splitting and augmented Lagrangians than in the local case to solve the non local problem (31).

A. Minimizations of local normal-guided CS

We discretize the image domain $\Omega \subset \mathbb{R}^2$ with a regular grid of size $n = n_x \times n_y$. In Ω we consider images as scalar functions with $u(i) \in \mathbb{R}$ and their gradients as vector-valued functions with $\nabla u(i) \in \mathbb{R}^2$. We use forward differences to compute the discrete gradients and backward differences for the divergence in order to preserve the adjoint relationship $\mathrm{div} = -\nabla^*$ in the discrete setting.

The discrete TV semi-norm is then given by

$$J(u) = \sum_i |\nabla u(i)| = \sum_i \sqrt{\nabla_x u(i)^2 + \nabla_y u(i)^2} \quad (35)$$

where we denote the pixel-wise norm of vectors as $|\boldsymbol{d}|(i) = \sqrt{d_x^2(i) + d_y^2(i)}$. Our discretized TV is then the ℓ_1 norm of the function computing the pixel-wise norm of the gradient,

i.e $J(u) = \| \, |\nabla u| \, \|_1$. For vector fields $\boldsymbol{d} = (d_x, d_y)$, we discretize the TV seminorm as follows

$$J(d_x, d_y) = \sum_i \sqrt{|\nabla d_x(i)|^2 + |\nabla d_y(i)|^2}. \quad (36)$$

In that case we observe that it corresponds to the ℓ_1 norm of the function computing the pixel-wise norm of the vector of combined gradients, i.e $J(d_x, d_y) = \| \, |(\nabla d_x, \nabla d_y)| \, \|_1$. With that observation it is easy then to extend it to a weighted TV norm as $J_w(d_x, d_y) = \| \, |W(\nabla d_x, \nabla d_y)| \, \|_1$, where W is the diagonal matrix of weights.

In the vector notation used in CS, we can efficiently compute the spatial derivatives multiplying the discrete functions arranged as a column vector with the sparse finite difference matrices $\nabla_x u = D_x u$, $\nabla_y u = D_y u$. Similarly, the discretization of the L_2 inner product in Ω corresponds to the usual dot product of vectors, i.e. $<v, u> = v^T u$.

1) Estimate u from CS measurements and normals: To reconstruct the image we need to solve the following convex minimization problem:

$$\min_u \| \, |\nabla u| \, \|_1 + \gamma v^T u + \frac{\alpha}{2} \|Au - f\|_2^2 \quad \text{with} \quad v = \mathrm{div}\, n. \quad (37)$$

We propose an iterative algorithm to solve (37) based on splitting and constraint minimization techniques. The main idea is to split the original problem into sub-optimization problems which are well-known and easy to solve, and combine them together using an augmented Lagrangian. The proposed algorithm is guaranteed to converge thanks to the convexity of (37).

Let us consider the following constrained minimization problem, which is equivalent to (11):

$$\min_{u, \boldsymbol{d}} = \| \, |\boldsymbol{d}| \, \|_1 + v^T u + \frac{\alpha}{2} \|Au - f\|_2^2 \quad \text{s.t.} \quad \boldsymbol{d} = \nabla u \quad (38)$$

The proposed splitting approach makes the original problem (11) easier to solve because (38) decouples the ℓ_1 norm and the gradient operator ∇.

Next, we reformulate this constrained minimization problem as an unconstrained optimization task. This can be done with an augmented Lagrangian approach, which translates the constraints into pairs of Lagrangian multiplier and penalty terms. Let us define the augmented Lagrangian energy associated to (38):

$$\mathcal{L}_1(u, \boldsymbol{d}, \boldsymbol{\lambda}) = \| \, |\boldsymbol{d}| \, \|_1 + v^T u + \frac{\alpha}{2} \|Au - f\|_2^2 + \lambda_x^T(d_x - D_x u)$$
$$+ \lambda_y^T(d_y - D_y u) + \frac{r}{2} \|d_x - D_x u\|_2^2 + \frac{r}{2} \|d_y - D_y u\|_2^2 \quad (39)$$

where $\boldsymbol{\lambda} = (\lambda_x, \lambda_y)$ are the Lagrange multipliers and r is a positive constant. The constraint minimization problem (38) reduces to finding the saddle-point of the augmented Lagrangian energy $\mathcal{L}_1$. The solution to the saddle point problem (39) can be approximated iteratively by the following algorithm: initialize the variables and Lagrange multipliers to zero; at each iteration find an approximate minimizer of $\mathcal{L}_1(u, \boldsymbol{d}, \boldsymbol{\lambda}^{k-1})$ with respect to the variables $u, \boldsymbol{d}$ and update the Lagrange multipliers with the residuals associated to the constraints; stop the process when u remains fix. As the Lagrangian $\mathcal{L}_1$ is convex with respect to $u, \boldsymbol{d}$, we can find

a minimizer by iteratively alternating the minimization with respect to each variable. The resulting method is equivalent to the alternating direction method of multipliers. The iterative method is summarized in Algorithm 1

Algorithm 1 Augmented Lagrangian method to solve (38), estimating u from CS measurements and normal matching

---

1: Initialize $u, \boldsymbol{d}, \boldsymbol{\lambda}$
2: For each iteration $l = 1, 2 \ldots$, find an approximate minimizer of $\mathcal{L}_1$ with respect to variables $(u, \boldsymbol{d})$ with fixed Lagrange multipliers $\boldsymbol{\lambda}^l$:

$$u^l = \arg\min_u \mathcal{L}_1(u, \boldsymbol{d}^{l-1}, \boldsymbol{\lambda}^l) \text{ solved in Fourier domain} \quad (40)$$

$$\boldsymbol{d}^l = \arg\min_{\boldsymbol{d}} \mathcal{L}_1(u^l, \boldsymbol{d}, \boldsymbol{\lambda}^l) \text{ solved by shrinkage} \quad (41)$$

3: Update Lagrange multipliers

$$\lambda_x^{l+1} = \lambda_x^l + r(d_x^l - D_x u^l)$$
$$\lambda_y^{l+1} = \lambda_y^l + r(d_y^l - D_y u^l)$$

4: Stop the iterative process when $\frac{\|u^l - u^{l-1}\|}{\|u^l\|} < \epsilon$.

The next step is to determine the solutions of the two sub-minimization problems (40),(41), which can be computed efficiently.

The sub-minimization problem (40) can be written as follows:

$$\min_u v^T u + \frac{\alpha}{2}\|Au - f\|_2^2 + \frac{r}{2}\|d_x + \frac{1}{r}\lambda_x - D_x u\|_2^2 + \frac{r}{2}\|d_y + \frac{1}{r}\lambda_y - D_y u - \|_2^2. \quad (42)$$

We see that it reduces to a quadratic minimization, with positive semi-definite Hessian $H = \alpha F^T R^T R F + r(D_x^T D_x + D_y^T D_y)$. The optimality conditions read

$$Hu = b \quad \text{with } b = \alpha F^T R f + D_x^T (r d_x + \lambda_x) + D_y^T (r d_y + \lambda_y).$$

Actually as R is a row selector, $R^T R$ is a sparse diagonal matrix with non-zero entries on the selected Fourier coefficients and we cannot assure the invertibility of H. We find an approximate solution defining the positive definite matrix $H_\epsilon = H + \epsilon I_n$ with small $\epsilon > 0$ and solving the approximate system

$$H_\epsilon u = b + \epsilon \hat{u}, \quad (43)$$

where we use the value of u from the previous augmented Lagrangian iteration to estimate $\hat{u} = u^{l-1}$. In the resulting system, H_ϵ is block circulant and we can use the Fourier transform to decompose it as $H_\epsilon = \mathcal{F}^T C \mathcal{F}$, with $C = \alpha R^T R + r \mathcal{F}(D_x^T D_x + D_y^T D_y)\mathcal{F}^T + \epsilon I_n$ a diagonal matrix. Consequently, the system (43) can easily be solved in the Fourier domain inverting the diagonal matrix C. In practice we use the FFT transform instead of doing the matrix multiplications with $\mathcal{F}$ and $\mathcal{F}^T$, which gives us a solution of complexity $\mathcal{O}(n \log n)$.

The minimization problem w.r.t. $\boldsymbol{d}$ corresponds to an ℓ_1 - ℓ_2 norm and can be solved by shrinkage. If we define $\boldsymbol{z} = \frac{1}{r}\boldsymbol{\lambda} - \nabla u$, Equation (41) is equivalent to

$$\min_{d_x, d_y} \sum_i |\boldsymbol{d}(i)| + \frac{r}{2}\sum_i |\boldsymbol{d}(i) - \boldsymbol{z}(i)|^2. \quad (44)$$

The minimization of (44) can be done pixel-wise and the solution is given by the shrinkage operator $\mathcal{S}(\boldsymbol{z}, 1/r)$.

$$\boldsymbol{d}(i) = \max\left\{|\boldsymbol{z}(i)| - \frac{1}{r}, 0\right\}\frac{\boldsymbol{z}(i)}{|\boldsymbol{z}(i)|} \quad i = 1, \ldots, n \quad (45)$$

2) Regularization of normals: To regularize the normals at each iteration we have to solve

$$\min_{|\boldsymbol{n}| \leq 1} \| |W(\nabla n_x, \nabla n_y)| \|_1 + \frac{\mu}{2}\|n_x - \hat{n}_x\|_2^2 + \frac{\mu}{2}\|n_y - \hat{n}_y\|_2^2,$$

where W is a diagonal matrix with weights associated to weighted TV seminorm. We use the same combination of splitting and augmented Lagrangian techniques than in Section V-A1. To avoid repetition, on the following we will simply write the form of the constraint minimization problem, the augmented Lagrangian and each of the subminimizations for a self-contained paper.

Equivalent constraint minimization is

$$\min_{\substack{\boldsymbol{n}=\boldsymbol{m}, |\boldsymbol{m}| \leq 1 \\ \boldsymbol{d}=\nabla n_x, \boldsymbol{e}=\nabla n_y}} \| |W(\boldsymbol{d}, \boldsymbol{e})| \|_1 + \frac{\mu}{2}\|n_x - \hat{n}_x\|_2^2 + \frac{\mu}{2}\|n_y - \hat{n}_y\|_2^2,$$

with associated augmented Lagrangian

$$\mathcal{L}_2(\boldsymbol{n}, \boldsymbol{m}, \boldsymbol{d}, \boldsymbol{e}, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\xi}) = \| |W(\boldsymbol{d}, \boldsymbol{e})| \|_1 + \frac{\mu}{2}\|\boldsymbol{n} - \hat{\boldsymbol{n}}\|_2^2 +$$
$$\lambda_x^T(d_x - D_x n_x) + \lambda_y^T(d_y - D_y n_x) + \frac{r_d}{2}\|d_x - D_x n_x\|_2^2 +$$
$$\frac{r_d}{2}\|d_y - D_y n_y\|_2^2 + \nu_x^T(e_x - D_x n_y) + \nu_y^T(e_y - D_y n_y)$$
$$+ \frac{r_e}{2}\|e_x - D_x n_y\|_2^2 + \frac{r_e}{2}\|e_y - D_y n_y\|_2^2 + \xi_x^T(n_x - m_x)$$
$$+ \frac{r_m}{2}\|n_x - m_x\|_2^2 + \xi_y^T(n_y - m_y) + \frac{r_m}{2}\|n_y - m_y\|_2^2.$$

The resulting minimization method is presented in Algorithm 2.

Algorithm 2 Augmented Lagrangian method to regularize normal vectors

---

1: Initialize $\boldsymbol{n}, \boldsymbol{m}, \boldsymbol{d}, \boldsymbol{e}, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\xi}$
2: For each iteration $l = 1, 2 \ldots$, find an approximate minimizer of $\mathcal{L}_2$ with respect to variables $(n_x, n_y, \boldsymbol{m}, \boldsymbol{d}, \boldsymbol{e})$ with fixed Lagrange multipliers $\boldsymbol{\lambda}^l, \boldsymbol{\nu}^l, \boldsymbol{\xi}^l$:

$$\boldsymbol{n}^l = \arg\min_{\boldsymbol{n}} \mathcal{L}_2(\boldsymbol{n}, \boldsymbol{m}^{l-1}, \boldsymbol{d}^{l-1}, \boldsymbol{e}^{l-1}, \boldsymbol{\lambda}^{l-1}, \boldsymbol{\nu}^{l-1}, \boldsymbol{\xi}^{l-1})$$
$$\boldsymbol{m}^l = \arg\min_{\boldsymbol{m}} \mathcal{L}(\boldsymbol{n}^l, \boldsymbol{m}, \boldsymbol{d}^{l-1}, \boldsymbol{e}^{l-1}, \boldsymbol{\lambda}^{l-1}, \boldsymbol{\nu}^{l-1}, \boldsymbol{\xi}^{l-1})$$
$$\boldsymbol{d}^l = \arg\min_{\boldsymbol{d}} \mathcal{L}(\boldsymbol{n}^l, \boldsymbol{m}^l, \boldsymbol{d}, \boldsymbol{e}^{l-1}, \boldsymbol{\lambda}^{l-1}, \boldsymbol{\nu}^{l-1}, \boldsymbol{\xi}^{l-1})$$
$$\boldsymbol{e}^l = \arg\min_{\boldsymbol{e}} \mathcal{L}(\boldsymbol{n}^l, \boldsymbol{m}^l, \boldsymbol{d}^l, \boldsymbol{e}, \boldsymbol{\lambda}^{l-1}, \boldsymbol{\nu}^{l-1}, \boldsymbol{\xi}^{l-1})$$

3: Update Lagrange multipliers

$$\lambda_x^l = \lambda_x^{l-1} + r_d(d_x^l - D_x n_x^l)$$
$$\lambda_y^l = \lambda_y^{l-1} + r_r(d_y^l - D_y n_x^l)$$
$$\nu_x^l = \nu_x^{l-1} + r_e(e_x^l - D_x n_y^l)$$
$$\nu_y^l = \nu_y^{l-1} + r_e(e_y^l - D_y n_y^l)$$
$$\xi^l = \xi^{l-1} + r_m(\boldsymbol{n} - \boldsymbol{m})$$

4: Stop the iterative process when $\frac{\|\boldsymbol{n}^l - \boldsymbol{n}^{l-1}\|}{\|\boldsymbol{n}^l\|} < \epsilon$.

The sub-minimization problem with respect to n_x can be written as follows:

$$\min_{n_x} \frac{\mu}{2}\|n_x - \hat{n}_x\|_2^2 + \xi_x^T(n_x - m_x) + \frac{r_m}{2}\|n_x - m_x\|_2^2$$
$$+\frac{r}{2}\|d_x + \frac{1}{r_d}\lambda_x - D_x n_x\|_2^2 + \frac{r_d}{2}\|d_y + \frac{1}{r_d}\lambda_y - D_y n_x - \|_2^2$$

We see that it reduces to a quadratic minimization, with positive definite Hessian $H = (\mu + r_m)I_n + r_d D_x^T D_x + r_d D_y^T D_y$. The optimality conditions read

$$Hn_x = \mu\hat{n}_x + r_m m_x + D_x^T(r_d d_x + \lambda_x) + D_y^T(r_d d_y + \lambda_y) - \xi_x.$$

As before, the resulting H is block circulant and we can use the Fourier transform to decompose it as $H = \mathcal{F}^T C \mathcal{F}$, with $C = (\mu + r_m)I_n + r_d\mathcal{F}(D_x^T D_x + D_y^T D_y)\mathcal{F}^T$ a diagonal matrix. We solve the linear system in the Fourier domain efficiently with the FFT transform. Observe that the minimization problem with respect to n_y has the same form and can be solved with the same technique.

The minimization problem w.r.t. d corresponds to the ℓ_1 - ℓ_2 problem

$$\min_{d_x, d_y} \sum_i |w(i)d(i)| + \frac{r_d}{2}\sum_i |d(i) - z(i)|^2, \qquad (46)$$

where $z = \frac{1}{r_d}\lambda - \nabla n_x$. As $w(i) > 0$, this minimization is equivalent to

$$\min_{d_x, d_y} \sum_i |d(i)| + \frac{r_d}{2w(i)}\sum_i |d(i) - z(i)|^2. \qquad (47)$$

A similar problem has already been solved in Section V-A1 with the shrinkage operator, which is now adapted to include the weights w. The solution is then

$$d(i) = \max\left\{|z(i)| - \frac{w(i)}{r}, 0\right\}\frac{z(i)}{|z(i)|} \quad i = 1, \ldots, n \quad (48)$$

Due to the symmetry of the problems, the same minimization technique is used for e.

Finally, the minimization problem w.r.t. m reads

$$\min_{|m(i)|=1} \frac{r_m}{2}\sum_i |m(i) - z(i)|^2, \text{ with } z = n + \frac{1}{r_m}\xi \quad (49)$$

and can be solved pixel-wise. For each pixel we have the following 2-D problem: given a point in space with coordinates $z(i) \in \mathbb{R}^2$ we want to find the point in the unit ball minimizing its distance to $z(i)$. It is clear that the solution corresponds to the projection of the unconstrained minimizer $z(i)$ into the unit ball, i.e

$$m(i) = \begin{cases} z(i) & |z(i)| \leq 1 \\ \frac{z(i)}{|z(i)|} & \text{otherwise} \end{cases}. \qquad (50)$$

B. Minimizations of non local normal-guided CS

In the discrete setting, the NL gradient is a linear operator. Arranging the image as a column vectors, it can be computed as a sparse matrix multiplication $\nabla_G u = Du$. The matrix $D \in \mathbb{R}^{N \times n}$ ($N = |E|$ indicates the number of nodes in the graph) is derived from the weights associated to the edges and is usually sparse. Consequently $d = Du \in \mathbb{R}^N$ is also a vector column, with as many components associated to node i as neighbours this node has. With the vector notation, the inner product of two vectors fields d, e defined in G is then computed as $< d, e >_G = d^T e$. As in the continuous setting, the NL divergence div_G is derived from its adjoint relation with the NL gradient $\nabla_G^* = -\text{div}_G$ and, consequently, in matrix notation it corresponds to $\text{div}_G d = -D^T d$.

Since the minimization associated to (33) has already been explained for the vectorial case, in the next paragraphs we focus on the minimizations associated to non local operators (31) and (34).

1) Minimization associated to CS reconstruction matching non local normals: With the previous notation, the minimization problem (31) reads

$$u = \arg\min_u \| |Du|_G \|_1 + \gamma v^T u + \frac{\alpha}{2}\|Au - f\|_2^2 \quad (51)$$

with $v = \text{div}_G \, n_G$. This minimization is also reformulated as a constraint minimization and solved efficiently with augmented Lagrangians. Compared to the local minimizations, in the splitting step we require an additional variable s to have efficient and analytic solutions for the posterior subminimization problems. The resulting constraint minimization formulation of (51) is

$$\min_{u,s,d} \| |d|_G \|_1 + v^T u + \frac{\alpha}{2}\|As - f\|_2^2 \text{ s.t. } \begin{cases} d = Du \\ s = u \end{cases} \quad (52)$$

The Lagrangian in that case reads

$$\mathcal{L}_3(u, s, d, \lambda_d, \lambda_u) = \| |d|_G \|_1 + v^T u + \frac{\alpha}{2}\|As - f\|^2$$
$$+\lambda_d^T(d - Du) + \frac{r_d}{2}\|d - Du\|^2 + \lambda_u^T(u - s) + \frac{r_u}{2}\|u - s\|^2.$$

The resulting minimization method is presented in Algorithm 3, where we have also hinted the solution to each of the subminimization problems.

Algorithm 3 Augmented Lagrangian method for CS reconstruction matching normals by (52)

1: Initialize $u, s, d, \lambda_d, \lambda_u$
2: For each iteration $l = 1, 2\ldots$, find an approximate minimizer of $\mathcal{L}_3$ with respect to variables (u, s, d) with fixed Lagrange multipliers λ_d^l, λ_u^l:

$u = \arg\min_u \mathcal{L}_3(u, s^{l-1}, d^{l-1}, \lambda_d^{l-1}, \lambda_u^{l-1})$ conjugate gradients

$s = \arg\min_s \mathcal{L}_3(u^l, s, d^{l-1}, \lambda_d^{l-1}, \lambda_u^{l-1})$ solved in Fourier domain

$d = \arg\min_d \mathcal{L}_3(u^l, s^l, d, \lambda_d^{l-1}, \lambda_u^{l-1})$ solved by non local shrinkage

3: Update Lagrange multipliers
$$\lambda_d^l = \lambda_d^{l-1} + r_d(d^l - Du^l)$$
$$\lambda_u^l = \lambda_u^{l-1} + r_u(u^l - s^l)$$

4: Stop the iterative process when $\frac{\|u^l - u^{l-1}\|}{\|u^l\|} < \epsilon$.

The minimization w.r.t u corresponds to the following quadratic positive definite problem

$$\min_u v^T u + \lambda_d^T(d - Du) + \frac{r_d}{2}\|d - Du\|^2 + \lambda_u^T(u - s) + \frac{r_u}{2}\|u - s\|^2.$$

We find its minimizer solving its optimality conditions, which provide the following system of linear equations

$$K_u u = -\gamma v - \lambda_u + r_u s + D^T \left(\lambda_d + r_d d \right). \quad (53)$$

Matrix $K_u = r_u I + r_d D^T D$ is sparse, symmetric and positive definite and we have efficient algorithms to invert it. We choose an iterative method to invert the matrix, initializing it from the previous solution to the minimization problem u^{l-1}. In particular we use the conjugate gradient method to exploit the symmetry and positive definition of K, with preconditioning matrix given by its incomplete Cholesky factorization. The resulting method is very fast, converging to enough precision with $2-3$ iterations of the conjugate gradient method.

The minimization w.r.t s is also a quadratic problem which can be efficiently solved, in that case in the Fourier domain. The problem reads

$$\min_s \ \frac{\alpha}{2}\|As - f\|^2 + \lambda_u^T (u - s) + \frac{r_u}{2}\|u - s\|^2. \quad (54)$$

The optimality conditions in that case are

$$\left(\alpha A^T A + r_u I_n \right) s = \alpha A^T f + \lambda_u + r_u u. \quad (55)$$

As before, the matrix $K_s = \alpha A^T A + r_u I_n = F^T C F$ is block-circulant and the resulting system is diagonal in the Fourier domain with $C = R^T R + r_u I_n$. Therefore (55) can be efficiently solved with the FFT.

The introduction of the additional splitting variable $s = u$ allows us to split the inversion of the full matrix $\alpha A^T A + r_d D^T D$, resulting of the use of a single variable for s and u, into the inversion of two matrices K_u and K_s. The sparse matrix K_u can be efficiently solved with a sparse incomplete Cholesky factorization, while the full matrix K_s is easily inverted in the Fourier domain. The original matrix $\alpha A^T A + r_d D^T D$, on the other hand, does not present an evident sparsity pattern or a direct decomposition involving fast transforms.

The minimization with respect to d is equivalent to

$$\min_d \ = \ \| \, |d|_G \, \|_1 + \frac{r_d}{2}\|d - z\|^2 \ \text{ with } \ z = Du - \frac{\lambda_d}{r_d} \quad (56)$$

As in the local case, this minimization is decoupled for each pixel i as follows

$$\min_{d(i,j) \ j\sim i} \ = \ \sqrt{\sum_{j\sim i} d^2(i,j)} + \sum_{j\sim i}\frac{r_d}{2}\left(d(i,j) - z(i,j) \right)^2 \quad (57)$$

and can be solved by a straight-forward extension of the shrinkage operator to the graph. That is, for each node neighbour to i the solution is given by

$$d^\star(i,j) = \max \left\{ |z|_G(i) - \frac{1}{r_d}, 0 \right\} \frac{z(i,j)}{|z|_G(i)}. \quad (58)$$

2) Minimization associated to the regularization of normals: With the previous notation, the minimization problem (34) reads

$$v = \arg\min_v \ \| \, |Dv|_G \, \|_1 + \frac{\mu}{2}\|v - \hat{v}\|_2^2 \quad (59)$$

As in the local case, we decouple the ℓ_1 and ℓ_2 problems defining an additional variable $d = Du$ and rewrite (59) as the following constraint minimization problem

$$\min_{v,d} \ = \ \| \, |d|_G \, \|_1 + \frac{\mu}{2}\|v - \hat{v}\|_2^2 \ \text{ s.t. } \ d = Du \quad (60)$$

with associated augmented Lagrangian

$$\mathcal{L}_4\left(u, d, \lambda_d\right) = \ \| \, |d|_G \, \|_1 + \frac{\mu}{2}\|v - \hat{v}\|_2^2$$
$$+ \lambda_d^T \left(d - Dv \right) + \frac{r_d}{2}\|d - Dv\|_2^2. \quad (61)$$

To minimize the Lagrangian $\mathcal{L}_3$ with respect to u, d, we alternate the direction of minimization with respect to each variable and proceed as indicated by Algorithm 3, where we have also hinted the solution to each of the subminimization problems.

Algorithm 4 Augmented Lagrangian method to regularize non local divergence of normals from (60)

1: Initialize u, d, λ_d
2: For each iteration $l = 1, 2 \ldots$, find an approximate minimizer of $\mathcal{L}_4$ with respect to variables (u, d) with fixed Lagrange multipliers λ_d^l:

   $v = \arg\min\limits_v \mathcal{L}_4(v, d^{l-1}, \lambda_d^{l-1})$ solved with conjugate gradient

   $d = \arg\min\limits_d \mathcal{L}_4(v^l, d, \lambda_d^{l-1})$ solved by non local shrinkage

3: Update Lagrange multipliers

$$\lambda_d^l = \lambda_d^{l-1} + r_d(d^l - Du^l)$$

4: Stop the iterative process when $\frac{\|v^l - v^{l-1}\|}{\|v^l\|} < \epsilon$.

The minimization w.r.t v corresponds to the following quadratic positive definite problem

$$\min_v \ \frac{\mu}{2}\|v - \hat{v}\|_2^2 + \lambda_d^T \left(d - Dv \right) + \frac{r_d}{2}\|d - Dv\|_2^2. \quad (62)$$

We find its minimizer by solving its optimality conditions, which provide the following system of linear equations

$$\left(\mu I + r_d D^T D \right) v = \mu\hat{v} + D^T \left(\lambda_d + r_d d \right). \quad (63)$$

We find the same form of matrix $K = \mu I + r_d D^T D$ than in (53) and, therefore, we solve with linear system (53) with the same conjugate gradient method.

The minimization with respect to d is equivalent to (56) changing u for v, in particular we have

$$\min_d \ = \ \| \, |d|_G \, \|_1 + \frac{r_d}{2}\|d - z\|^2 \ \text{ with } \ z = Du - \frac{\lambda_d}{r_d} \quad (64)$$

and is solved with the same adaptation of the shrinkage operator to the graph. For each node neighbour to i, the solution is given by

$$d^\star(i,j) = \max \left\{ |z|_G(i) - \frac{1}{r_d}, 0 \right\} \frac{z(i,j)}{|z|_G(i)}. \quad (65)$$

VI. NUMERICAL RESULTS

In this section we present some of the numerical results obtained with our method and compare it to other techniques. We compare the local version of our method to standard CS recovery algorithm (5) and to the edge-guided CS proposed in [10]. The non local version of our model is compared to the non local CS recovery (28), which does not take into account the geometric information of the non local gradients into the recovery process. For the non local case, in our model we have regularized the divergence of the normal with the standard ROF model. A technical report with higher resolution images and code are available online in the authors webpage.

We use partial Fourier measurements for our reconstruction and perform radial sampling on R with different number of measurements in relation to the size of the signal (we specify it with the ratio $\frac{m}{n}$). For a fair comparison, we have used the same robust edge detector (9) for the edge-guided CS and our method and we have implemented the minimizations with the same splitting and augmented Lagrangian techniques for all the methods. The parameter α, which is related to the noise present in the CS measurements, has been manually tuned to obtain best reconstruction with the standard CS recovery models (5) and (28) and used with the other methods. The other parameters of our model γ, μ have also been chosen manually to obtain good CS recovery in terms of SNR. We have observed that γ (which controls the weight given to the alignment of the normals) takes similar values for the same kind of images (textured or brain IRM images) and remains stable for different sparsity and noise levels. On the other hand, the parameter μ controlling the smoothness of the estimated normals decreases when the number of measurements decreases or the noise level increases because the partial reconstructions and the estimated normals are noisier and require more regularization.

In a first set of experiments we test our method with MRI images, first with the Shepp-Logan phantom and then with a real MRI brain image.

Table I show the quantitative results of the different CS reconstruction methods for MRI images. Our method always outperforms the standard TV reconstruction and the edge-guided CS technique. In the experiments, both the edge-guided CS and our proposed method are initialized with the TV solution and, therefore, always improve its reconstruction. Comparing the gains of these two methods with respect to the TV reconstruction, we observe that our method more than doubles the gain of edge-guided CS. The Figures 1(e)-3 show qualitatively the improvement of our method over TV reconstruction. In the case of the phantom we are able to better reconstruct the phantom with fewer measurements both in the local and non local case, while with a real MRI image our reconstruction is able to capture better non-dominant edges of the white-grey matter interface. In Figure 3 we explicitly compare the normals associated to the TV solution and the regularized normals of our local reconstruction for the real brain MRI image. We observe that our method is able to better reconstruct the normals, and therefore the shapes, of the image. Performance improves with non local regularization, with our

method outperforming the non local CS reconstruction for all the experiments. As expected, the gain of our method compared to TV is lower than in the local approach because we loose part of the directional information of the normals by denoising their divergence instead of the vector fields. For each image we also added different levels of Gaussian

Image	$\frac{m}{n}$	local CS			non-local CS	
		TV	edge CS	normal CS	TV	normal CS
Phantom	8%	7.33 dB	7.37 dB	**12.78 dB**	28.28 dB	**33.13 dB**
Phantom	12%	38.60 dB	45.33 dB	**56.14 dB**	61.84 dB	**74.57 dB**
Brain	12%	17.14 dB	17.38 dB	**17.71 dB**	18.96 dB	**20.39 dB**
Brain	20%	22.16 dB	22.35 dB	**23.82 dB**	23.13 dB	**24.12 dB**

TABLE I: Comparison of CS reconstruction for MRI images. The first three columns show the results with the standard TV as regularization: TV stands for model of Equation (5), edge CS for (13) and normal CS for our method. The last two columns correspond to the definition of NL TV: NL-TV corresponds to standard non local CS recovery (28) and NL normal CS for the proposed non local method.

noise (σ_n) to the signal to investigate the robustness of our method to noise. Results are shown in table II. We observe that we are more robust to noise than edge-guided CS (which in fact does not improve the TV reconstruction for noise levels $\sigma_n = 15\%$, $\sigma_n = 10\%$) thanks to regularization step on the estimation of the normals. As before, non local regularization improves CS reconstruction, we observe that our non local method outperforms again the non local TV and is also robust to noise.

Image $\frac{m}{n} = 12\%$	noise σ_n	local CS			non-local CS	
		TV	edge CS	normal CS	TV	normal CS
Phantom	5%	11.90 dB	11.91 dB	**12.90 dB**	17.92 dB	**18.36 dB**
	10%	8.37 dB	8.38 dB	**9.44 dB**	12.15 dB	**13.03 dB**
	15%	6.59 dB	6.59 dB	**7.28 dB**	10.09 dB	**10.27 dB**
Brain	5%	13.37 dB	13.36 dB	**13.78 dB**	14.86 dB	**15.00 dB**
	10%	10.88 dB	10.88 dB	**11.57 dB**	12.31 dB	**12.50 dB**
	15%	9.89 dB	9.89 dB	**10.48 dB**	10.94 dB	**11.19 dB**

TABLE II: Comparison of CS reconstruction for noisy MRI images with 12% of samples and different levels σ_n of Gaussian noise. The first three columns show the results with standard TV as regularization: TV stands for the model of Equation (5), edge CS for (13) and normal CS for our method. The last two columns correspond to the definition of NL TV: NL-TV corresponds the standard non local CS recovery (28) and NL normal CS for the proposed non local method.

Our next experiment is performed with two synthetic images in order to understand the properties of the local and non-local versions of our algorithm. The first step in our iterative method is designed to preserve the discontinuities in the level set normals of the reconstructed image, while the second step introduces this geometrical information in the CS reconstruction algorithm. This property is specially interesting for the recovery of images with geometrical structures like ridges or valleys, as shown in Figure 4. We observe that the proposed local method is able to recover ridges and valleys to certain extent and outperforms TV reconstruction in this kind of structures; but it is not able to capture the repetitive

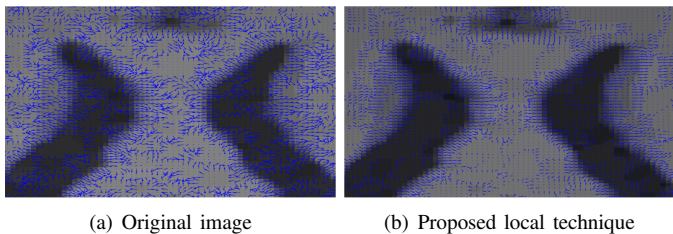(a) Original image   (b) Proposed local technique

Fig. 3: Zoom on reconstructed brain MRI image from 12% of measurements in Fourier domain. We superpose the reconstructed signals with the normals associated to their level sets for the standard TV solution (left) and for the local version of our method (right). Our method is able to better reconstruct the normals and shapes of the image.

nature of the image like the proposed non-local version of our method.

In a third set of experiments, we tested our method with natural images containing textures, where edge detection by itself is a difficult task and the images can not be considered piecewise constant. With these images, the local regularization looses all texture information, while the non-local approaches can recover repetitive patterns and better exploit the geometrical information of the image. Results with our method are presented in table III, with some of the reconstructed images shown in Figures 1-5 to qualitatively analyse the performance of our method.

A quantitative comparison of the different methods with textured images is presented in table III. We observe that the inclusion of an edge detector in edge-guided CS does not improve the TV reconstruction because the partially reconstructed images are not accurate enough to detect edges and the weighted TV term of edge-guided CS encourages edges in wrong positions. That effect is not observed in our method because it is additive and not multiplicative and it exploits the directional information of the regularized normals, which can partially capture texture information better than an edge detector. As a consequence, our local method always outperforms the TV reconstruction and edge-guided CS methods. For the non local regularizations our method outperforms non local TV, but the gain in some cases is negligible (fingerprint and baboon images for a ratio of measurements $\frac{m}{n} = 12\%$ or 20%). In fact, the non local methods require a good estimate of the reconstruction to initialize the non local gradient and divergence operators. Since our method requires both gradient and divergence to estimate the non local normals and align them with the reconstruction, we can only improve the non local TV reconstruction when the initialization (in our case we use the standard TV solution) has a minimum level of accuracy. The fact that more measurements are required for the fingerprint of baboon images is coherent with CS theory, as these images have finer details and are less sparse than Lena or Barbara in terms of total variation. In the reconstruction of Lena in Figures 2(a),2(b), 2(d),2(e),2(f) we can qualitatively observe the advantages of our method in comparison to local and non-local TV reconstruction for textured images. In the local case we avoid the staircase effect, which is clearly visible

in the TV reconstruction of Lena's cheek. In the non local case, we also capture better slowly varying textures changes, see for instance the different shadows in Lena's skin or hat. In both cases this improvement is due to the regularization of the level set normals of the image, which we exploit for CS recovery with our two step procedure.

$\frac{m}{n}$	image	local CS			non-local CS	
		TV	edge CS	normal CS	TV	normal CS
12%	Lena	14.53 dB	14.47 dB	**14.86 dB**	15.82 dB	**16.79 dB**
	Barbara	13.35 dB	13.31 dB	**13.59 dB**	15.00 dB	**15.52 dB**
	fingerprint	4.13 dB	4.11 dB	**4.13 dB**	5.97 dB	**5.98 dB**
	baboon	7.40 dB	7.25 dB	**7.40 dB**	7.65 dB	**7.65 dB**
20%	Lena	18.44 dB	18.36 dB	**19.27 dB**	19.95 dB	**21.09 dB**
	Barbara	16.71 dB	16.62 dB	**17.13 dB**	18.37 dB	**18.93 dB**
	fingerprint	5.70 dB	5.62 dB	**5.70 dB**	9.03 dB	**9.07 dB**
	baboon	9.13 dB	8.91 dB	**9.14 dB**	9.63 dB	**9.74 dB**
39%	Lena	25.39 dB	25.30 dB	**26.71 dB**	26.39 dB	**27.51 dB**
	Barbara	20.83 dB	20.68 dB	**21.36 dB**	24.68 dB	**25.33 dB**
	fingerprint	12.02 dB	11.84 dB	**12.03 dB**	14.52 dB	**14.56 dB**
	baboon	13.30 dB	13.14 dB	**13.41 dB**	13.44 dB	**13.82 dB**

TABLE III: Comparison of CS reconstruction for textured images. The first three columns show the results with the standard TV regularization: TV stands for the model of Equation (5), edge CS for (13) and normal CS for our method. The last two columns correspond to the definition of NL TV: NL-TV corresponds to the standard non local CS recovery (28) and NL normal CS for the proposed non local method.

VII. Conclusions

We propose a normal guided compressed sensing recovery method to recover images of higher qualities from fewer measurements. The normal vectors of image level curves have been exploited in denoising and inpainting algorithms, but in compressed sensing this information is embedded in the measurements and state-of-the-art recovery algorithms have just neglected it. To extract this geometric information we alternatively estimate the normals of the image level set curves and then improve the compressed sensing reconstruction matching the estimated normals, the compressed sensing measurements and the sparsity constraints. Although a we cannot provide a rigorous proof of convergence for this two-step procedure, experiments show a clear improvement over standard compressive sensing algorithms due to the introduction of the geometric information of level contours into the image recovery process. The proposed method is also extended to non local operators to recover textured images and could also be applied to improve existing non local denoising and deblurring methods. Our numerical experiments show that the proposed method improves image recovery in several ways: it is able to recover sharp edges as well as smoothly varying image regions, avoiding the staircase effect in the case of total variation regularization; it is robust to noise and the sparsity of the signal and relies on efficient minimization techniques to obtain a fast and easy-to-code algorithm.

REFERENCES

[1] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, Jul. 2011.

[2] D. Needell and R. Ward, "Stable image reconstruction using total variation minimization," *arXiv preprint 1202.6429*, pp. 1–23, 2012.

[3] S. P. Boyd and L. Vandenberghe, *Convex optimization theory*, 7th ed. New York: Cambridge University Press, Jun. 2004, vol. 25, no. 3.

[4] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and gaussian measurements," in *Conference on Information Sciences and Systems*, march 2006, pp. 207 –212.

[5] M. Lustig, J. Lee, D. Donoho, and J. Pauly, "Faster Imaging with Randomly Perturbed , Undersampled Spirals and L-1 Reconstruction," in *Proceedings of the Annual Meeting of International Society for Magnetic Resonance in Medicine*, 2005, p. 50.

[6] E. J. Candès and J. Romberg, "Practical signal recovery from random projections," *IEEE Transactions on Signal Processing*, vol. 2291, no. 626, 2005.

[7] S. Ma, Y. Wotao, Y. Zhang, and A. Chakraborty, "An efficient algorithm for compressed MR imaging using total variation and wavelets," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2008, pp. 1–8.

[8] M. Lysaker, S. Osher, and X.-C. Tai, "Noise Removal Using Smoothed Normals and Surface Fitting," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1345–1357, 2004.

[9] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.

[10] W. Guo and W. Yin, "EdgeCS: edge guided compressive sensing reconstruction," *Proceedings of Visual Communications and Image Processing*, no. 2, pp. 17–23, 2010.

[11] P. Blomgren and T. F. Chan, "Color TV: total variation methods for restoration of vector-valued images." *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 304–309, Jan. 1998.

[12] M. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 421–432, Jan. 1998.

[13] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, Jul. 1990.

[14] J. Hahn, C. Wu, and X.-C. Tai, "Augmented Lagrangian Method for Generalized TV-Stokes Model," *Journal of Scientific Computing*, vol. 50, no. 2, pp. 235–264, Apr. 2011.

[15] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera, "Filling-in by joint interpolation of vector fields and gray levels." *IEEE Transactions on Image Processing*, vol. 10, no. 8, pp. 1200–11, Jan. 2001.

[16] M. Lysaker, A. Lundervold, and X.-C. Tai, "Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time." *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1579–90, Jan. 2003.

[17] F. Dong, Z. Liu, D. Kong, and K. Liu, "An improved LOT model for image restoration," *Journal of Mathematical Imaging and Vision*, vol. 34, no. 1, pp. 89–97, May 2009.

[18] X. Tai, S. Borok, and J. Hahn, "Image Denoising Using TV-Stokes Equation with an Orientation-matching Minimization," in *Scale Space and Variational Methods in Computer Vision*, 2009, pp. 490–501.

[19] X. Tai, S. Osher, and R. Holm, "Image Inpainting Using a TV-Stokes Equation," in *Image Processing Based on Partial Differential Equations*, ser. Mathematics and Visualization, X. Tai, K.-A. Lie, T. Chan, and S. Osher, Eds. Springer, 2007, pp. 3–22.

[20] J. Hahn, X.-C. Tai, S. Borok, and A. M. Bruckstein, "Orientation-Matching Minimization for Image Denoising and Inpainting," *International Journal of Computer Vision*, vol. 92, no. 3, pp. 308–324, Jul. 2010.

[21] H. B. and B. M., "The variational approach to shape from shading," *Computer Vision, Graphics, and Image Processing*, vol. 33, no. 2, pp. 174–208, 1986.

[22] O. Y., B. A., and B. I., "Mesh regularization and adaptive smoothing," *Computer-Aided Design*, vol. 33, no. 11, pp. 789 – 800, 2001.

[23] R. Kimmel and A. M. Bruckstein, "Regularized Laplacian zero crossings as optimal edge integrators," *International Journal of Computer Vision*, vol. 53, no. 3, pp. 225–243, 2003.

[24] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655 –4666, dec. 2007.

[25] D. Needell and J. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301 – 321, 2009.

[26] B. T. and D. M.E., "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265 – 274, 2009.

[27] B. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Transactions on Signal Processing*, vol. 47, no. 1, pp. 187 –200, jan 1999.

[28] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707 –710, oct. 2007.

[29] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing Sparsity by Reweighted ℓ1 Minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, Oct. 2008.

[30] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An Iterative Regularization Method for Total Variation-Based Image Restoration," *Multiscale Modeling & Simulation*, vol. 4, no. 2, p. 460, 2005.

[31] L. He, T. Chang, S. Osher, T. Fang, and P. Speier, "MR image reconstruction by using the iterative refinement method and nonlinear inverse scale space methods," *UCLA CAM Report*, 2006.

[32] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *SIAM Journal on Multiscale Modelling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[33] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling and Simulation*, vol. 7, no. 3, pp. 1005–1028, 2008.

[34] D. Zhou and B. Schölkopf, "Regularization on Discrete Spaces," *Pattern Recognition*, vol. 80, pp. 361–368, 2005.

[35] A. Elmoataz, O. Lezoray, and S. Bougleux, "Nonlocal Discrete Regularization on Weighted Graphs: A Framework for Image and Manifold Processing," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1047–1060, 2008.

[36] G. G. and O. S., "Nonlocal linear image regularization and supervised segmentation," *Multiscale Modeling and Simulation*, vol. 6, no. 2, pp. 595–630, 2007.

[37] X. Bresson and T. F. Chan, "Non-local unsupervised variational image segmentation models," *UCLA cam report*, pp. 1–34, 2008.

[38] G. Peyré, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," in *European Conference on Computer Vision*. Springer-Verlag, 2008, pp. 57–68.

[39] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized non-local regularization for deconvolution and sparse reconstruction," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 253–276, 2010.

[40] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging*, pp. 1–24, 2008.

[41] T. Goldstein and S. Osher, "The Split Bregman Method for L1-Regularized Problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, p. 323, 2009.

[42] R. Glowinski, J.-L. Lions, and R. Tremolieres, *Numerical Analysis of Variational Inequalities*. Elsevier, 1981.

[43] A. Chambolle, "An algorithm for Mean Curvature Motion," *Interfaces and Free Boundaries*, vol. 6, pp. 195–218, 2004.

[44] T. F. Chan, G. H. Golub, and P. Mulet, "A Nonlinear Primal-Dual Method for Total Variation-Based Image Restoration," *SIAM Journal on Scientific Computing*, vol. 20, no. 6, p. 1964, 1999.

[45] M. Zhu, S. J. Wright, and T. F. Chan, "Duality-based algorithms for total-variation-regularized image restoration," *Computational Optimization and Applications*, vol. 47, no. 3, pp. 377–400, Dec. 2010.

[46] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, pp. 120–145, 2011.

[47] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman Iterative Algorithms for l1 -Minimization with Applications to Compressed Sensing," *SIAM Journal on Imaging Sciences*, vol. 1, no. 1, pp. 143–168, 2008.

[48] C. Wu and X.-C. Tai, "Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 300–339, 2010.

(a) Shepp-Logan phantom

(b) TV reconstruction, 7.33 dB
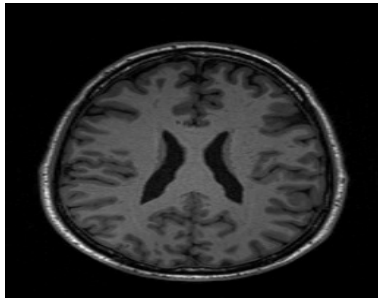
(c) edgeCS, 7.37 dB
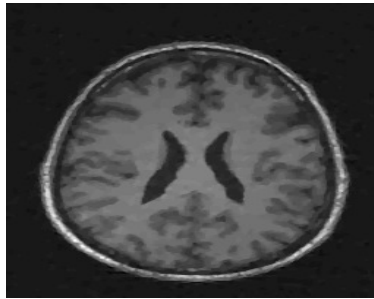
(d) non local TV reconstruction, 28.28 dB
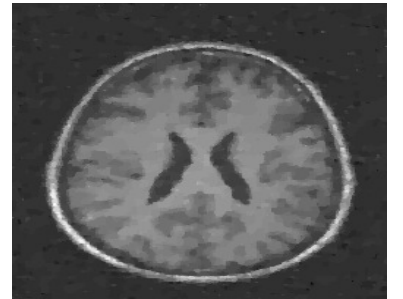
(e) proposed local method, 12.78 dB

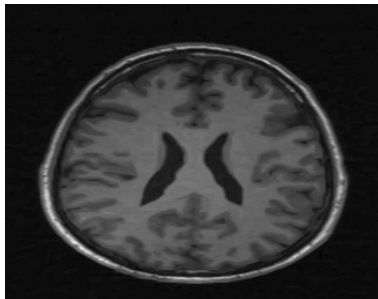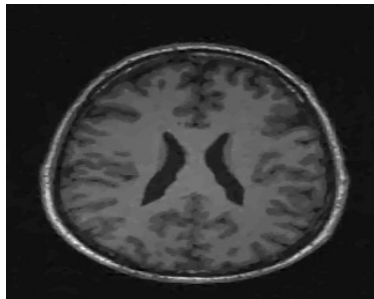(f) proposed non local method, 31.26 dB
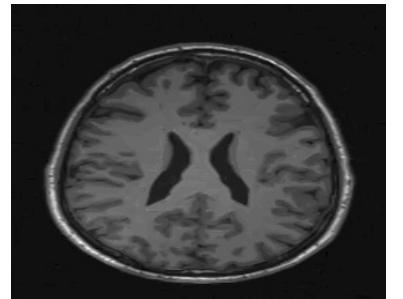
(g) Brain MRI

(h) TV reconstruction, 17.14 dB

(i) edgeCS reconstruction, 17.38 dB

(j) non local TV reconstruction, 18.96 dB

(k) proposed local method, 18.56 dB

(l) proposed non local method, 20.39 dB

Fig. 1: Columns 1-2: reconstruction of Shepp-Logan phantom from 8% of measurements in Fourier domain. Columns 3-4: reconstruction of MRI brain image from 12% of measurements in Fourier domain.

(a) Lena

(b) TV reconstruction, 18.44 dB

(c) edgeCS, 18.36 dB

(d) non local TV reconstruction, 19.45 dB

(e) proposed local method, 19.27 dB

(f) proposed non local method, 21.09 dB

Fig. 2: Reconstruction of Lena from 20% of measurements in Fourier domain.

(a) Original image

(b) Our local method, 16.73 dB

(c) Our non-local method, 30.47 dB

(d) Original image, from [14]

(e) Our local method, 5.90 dB

(f) Our non-local method, 14.11 dB
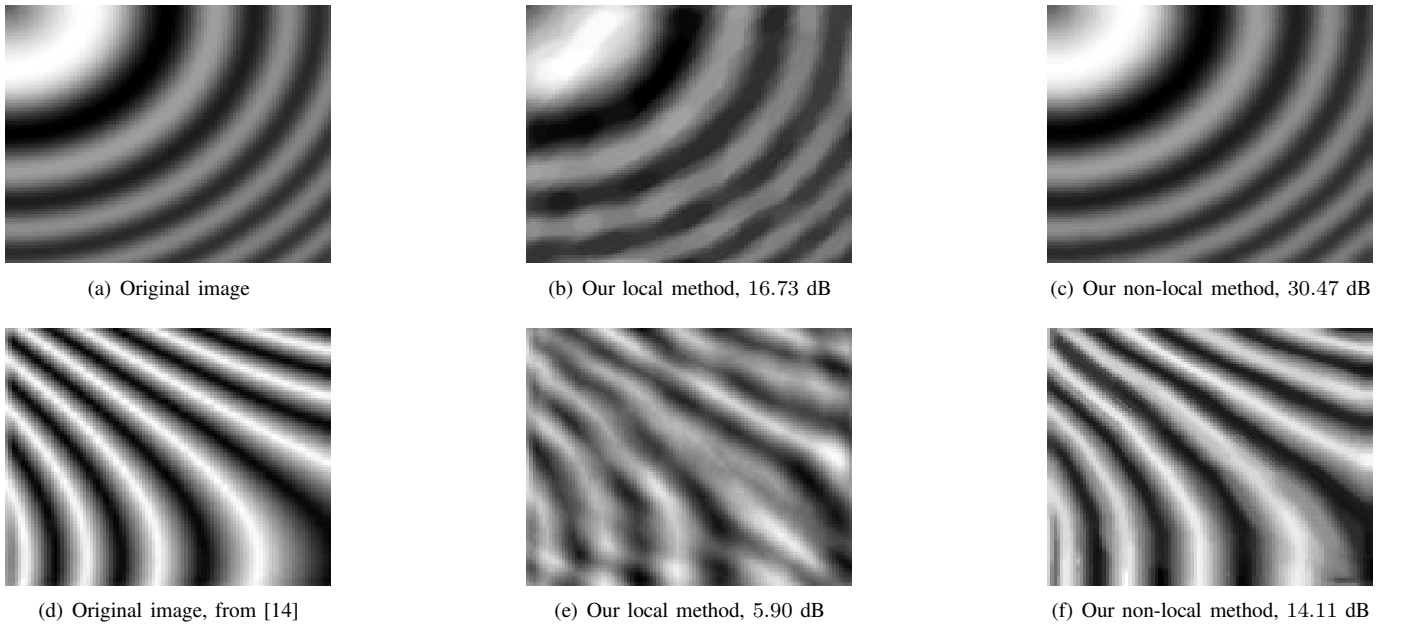
Fig. 4: Reconstruction of two synthetic images characterized by ridges from 10% and 14% of measurements in Fourier domain. Local and non-local TV results in SNR of 16.30 and 30.15 dB for the top image, and 5.86 and 12.99 dB for the bottom one.
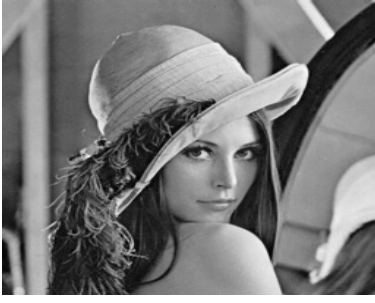
(a) Lena

(b) proposed local method, 14.86 dB

(c) proposed non local method, 16.79 dB

(d) Barbara

(e) proposed local method, 13.59 dB
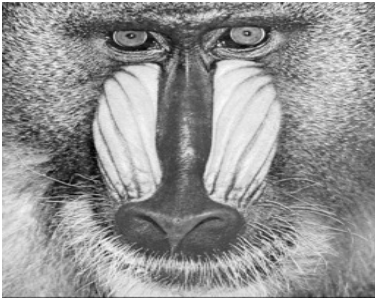
(f) proposed non local method, 15.52 dB

(g) Barbara

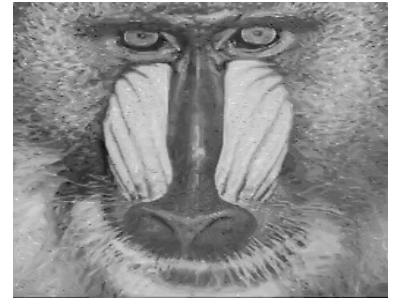(h) proposed local method, 17.13 dB
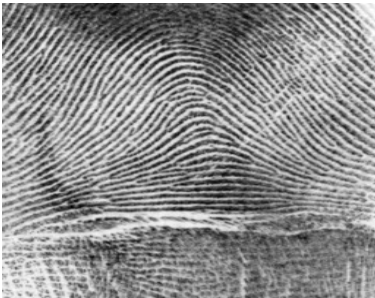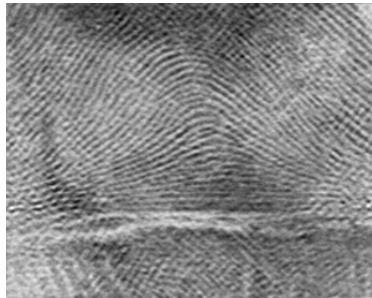
(i) proposed non local method, 18.92 dB

(j) Baboon

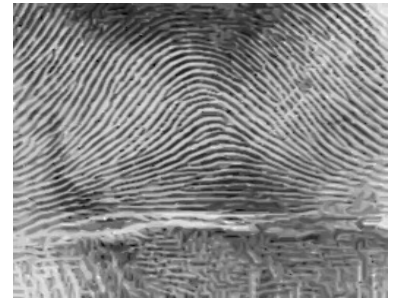(k) proposed local method, 9.14 dB

(l) proposed non local method, 9.74 dB

(m) Fingerprint

(n) proposed local method, 5.70 dB

(o) proposed non local method, 9.07 dB

Fig. 5: First row: reconstruction of Lena phantom from 12% of measurements in Fourier domain. Second row: reconstruction of Barbara from 12% of measurements in Fourier domain. Third row: reconstruction of Barbara from 20% of measurements in Fourier domain. Fourth row: reconstruction of baboon from 20% of measurements in Fourier domain. Fifth row: reconstruction of fingerprint from 20% of measurements in Fourier domain.