

# A Fast Wavelet-Based Reconstruction Method for Magnetic Resonance Imaging

M. Guerquin-Kern\*, M. Häberlin, K. P. Pruessmann, and M. Unser

**Abstract**—In this work, we exploit the fact that wavelets can represent magnetic resonance images well, with relatively few coefficients. We use this property to improve magnetic resonance imaging (MRI) reconstructions from undersampled data with arbitrary k-space trajectories. Reconstruction is posed as an optimization problem that could be solved with the iterative shrinkage/thresholding algorithm (ISTA) which, unfortunately, converges slowly. To make the approach more practical, we propose a variant that combines recent improvements in convex optimization and that can be tuned to a given specific k-space trajectory. We present a mathematical analysis that explains the performance of the algorithms. Using simulated and *in vivo* data, we show that our nonlinear method is fast, as it accelerates ISTA by almost two orders of magnitude. We also show that it remains competitive with TV regularization in terms of image quality.

**Index Terms**—Compressed sensing, fast iterative shrinkage/thresholding algorithm (FISTA), fast weighted iterative shrinkage/thresholding algorithm (FWISTA), iterative shrinkage/thresholding algorithm (ISTA), magnetic resonance imaging (MRI), non-Cartesian, nonlinear reconstruction, sparsity, thresholded Landweber, total variation, undersampled spiral, wavelets.

## I. INTRODUCTION

MAGNETIC RESONANCE IMAGING scanners provide data that are samples of the spatial Fourier transform (also known as *k-space*) of the object under investigation. The Shannon–Nyquist sampling theory in both spatial and k-space domains suggests that the sampling density should correspond to the field-of-view (FOV) and that the highest sampled frequency is related to the pixel width of the reconstructed images. However, constraints in the implementation of the k-space trajectory that controls the sampling pattern (e.g., acquisition duration, scheme, smoothness of gradients) may impose locally reduced sampling densities. Insufficient sampling results in reconstructed images with increased noise and artifacts, particularly when applying gridding methods.

Manuscript received January 05, 2011; revised March 22, 2011; accepted March 26, 2011. Date of publication April 07, 2011; date of current version August 31, 2011. This work was supported by the Swiss National Competence Center in Biomedical Imaging (NCCBI) and the Center for Biomedical Imaging of the Geneva-Lausanne Universities and EPFL. Asterisk indicates corresponding author.

\*M. Guerquin-Kern is with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.

M. Unser is with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.

M. Häberlin and K. P. Pruessmann are with the Institute for Biomedical Engineering, ETH Zürich, CH-8092 Zürich, Switzerland.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2011.2140121

The common and generic approach to alleviate the reconstruction problem is to treat the task as an inverse problem [1]. In this framework, ill-posedness due to a reduced sampling density is overcome by introducing proper regularization constraints. They assume and exploit additional knowledge about the object under investigation to robustify the reconstruction.

Earlier techniques used a quadratic regularization term, leading to solutions that exhibit a linear dependence upon the measurements. Unfortunately, in the case of severe undersampling (i.e., locally low sampling density) and depending on the strength of regularization, the reconstructed images still suffer from noise propagation, blurring, ringing, or aliasing errors. It is well known in signal processing that the blurring of edges can be reduced via the use of nonquadratic regularization. In particular,  $\ell_1$ -wavelet regularization has been found to outperform classical linear algorithms such as Wiener filtering in the deconvolution task [2].

Indicative of this trend as well is the recent advent of *Compressed Sensing* (CS) techniques in MRI [3], [4]. These let us draw two important conclusions.

- The introduction of randomness in the design of trajectories favors the attenuation of residual aliasing artifacts because they are spread incoherently over the entire image.
- Nonlinear reconstructions—more precisely,  $\ell_1$ -regularization—outperform linear ones because they impose constraints that are better matched to MRI images.

Many recent works in MRI have focused on nonlinear reconstruction via total variation (TV) regularization, choosing finite differences as a sparsifying transform [3], [5]–[7]. Nonquadratic wavelet regularization has also received some attention [3], [8]–[11], but we are not aware of a study that compares the performance of TV against  $\ell_1$ -wavelet regularization.

Various algorithms have been recently proposed for solving general linear inverse problems subject to  $\ell_1$ -regularization. Some of them deal with an approximate reformulation of the  $\ell_1$  regularization term. This approximation facilitates reconstruction sacrificing some accuracy and introducing extra degrees of freedom that make the tuning task laborious. Instead, the iterative shrinkage/thresholding algorithm [2], [12], [13] (ISTA) is an elegant and nonparametric method that is mathematically proven to converge. A potential difficulty that needs to be overcome is the slow convergence of the method when the forward model is poorly conditioned (e.g., low sampling density in MRI). This has prompted research in large-scale convex optimization on ways to accelerate ISTA. The efforts so far have followed two main directions.

- Generic multistep methods that exploit the result of past iterations to speed up convergence, among them: two-step

iterative shrinkage/thresholding [14] (TwIST), Nesterov schemes [15]–[17], fast ISTA [18] (FISTA), and monotonic FISTA [19] (MFISTA).

- Methods that optimize wavelet-subband-dependent parameters with respect to the reconstruction problem: multilevel thresholded Landweber (MLTL) [20], [21] and subband adaptive ISTA (SISTA) [22].

In this work, we exploit the possibility of combining and tailoring the two generic types of accelerating strategies to come up with a new algorithm that can speedup the convergence of the reconstruction and that can accommodate for every given k-space trajectory. Here, we first consider single-coil reconstructions that do not use sensitivity knowledge. In a second time, we confirm the results with SENSE reconstructions [1].

We propose a practical reconstruction method that turns out to sensibly outperform linear reconstruction methods in terms of reconstruction quality, without incurring the protracted reconstruction times associated with nonlinear methods. This is a crucial step in the practical development of nonlinear algorithms for undersampled MRI, as the problem of fixing the regularization parameter is still open. We also provide a mathematical analysis that justifies our algorithm and facilitates the tuning of the underlying parameters.

This paper is structured as follows. In Section II, we describe the basic data-formation model for MRI and derive the discrete forward model. The representation of the object by wavelet bases is considered in Section II-C2; in particular, we legitimize the use of a wavelet regularization term (which promotes sparsity) to distinguish the solution from other possible candidates. In Section III, we propose a fast algorithm for solving the nonlinear reconstruction problem and present theoretical arguments to explain its superior speed of convergence. Finally, we present in Section IV an experimental protocol to validate and compare our practical method with existing ones. We focus mainly on reconstruction time and signal-to-error ratio (SER) with respect to the reference image.

## II. MRI AS AN INVERSE PROBLEM

In this section, we present the MR acquisition model and the representation of the signal that is used to specify the reconstruction problem. The main acronyms and notations are summarized in Table 1. We motivate the sparsity assumptions in the wavelet domain and the variational approach with  $\ell_1$  regularization that is used to solve the inverse problem of imaging.

### A. Model of Data Formation

1) *Physics*: We consider MRI in two dimensions, in which case a 2D plane is excited. The time-varying magnetic gradient fields that are imposed define a trajectory in the (spatial) Fourier domain that is often referred to as k-space. We denote by  $\mathbf{k}$  the coordinates in that domain. The excited spins, which behave as radio-frequency emitters, have their precessing frequency and phase modified depending on their positions. The modulated part of the signal received by a homogeneous coil is given by

$$m(\mathbf{k}) = \hat{\rho}(\mathbf{k}) = \int_{\mathbb{R}^2} \rho(\mathbf{r}) e^{-j(\mathbf{k}, \mathbf{r})} d\mathbf{r}. \quad (1)$$

It corresponds to the Fourier transform of the spin density  $\rho$  that we refer to as *object*. The  $N$  measurements, concatenated in the vector  $\mathbf{m} = (m_1, \dots, m_N)$ , correspond to sampled values of this Fourier transform at the frequency locations  $\mathbf{k}_N$  along the k-space trajectory.

### 2) Model for the Original Data:

a) *Spatial discretization of the object*: From here on, we consider that the Fourier domain and, in particular, the sampling points  $\mathbf{k}_N$ , are scaled to make the Nyquist sampling interval unity. This can be done without any loss of generality if the space domain is scaled accordingly. Therefore, we model the object as a linear combination of pixel-domain basis functions  $\varphi_{\mathbf{p}}$  that are shifted replicates of some generating function  $\varphi$ , so that

$$\rho = \sum_{\mathbf{p} \in \mathbb{Z}^2} c[\mathbf{p}] \varphi_{\mathbf{p}} \quad (2)$$

with

$$\varphi_{\mathbf{p}}(\mathbf{r}) = \varphi(\mathbf{r} - \mathbf{p}). \quad (3)$$

In MRI, the implicit choice for  $\varphi$  is often Dirac's delta. Different discretizations have been proposed, for example by Sutton *et al.* [23] with  $\varphi$  as a boxcar function or later by Delattre *et al.* [24] with B-splines. But it has not been worked out in detail how to get back the image for general  $\varphi$  that are noninterpolating, which is the case, for instance for B-splines of order greater than 1. The image to be reconstructed [i.e., the sampled version of the object  $\rho(\mathbf{p})$ ] is obtained by filtering the coefficients  $c[\mathbf{p}]$  with the discrete filter

$$P(e^{j\omega}) = \sum_{\mathbf{h} \in \mathbb{Z}^2} \hat{\varphi}(\omega + 2\pi\mathbf{h}) \quad (4)$$

where  $\hat{\varphi}$  denotes the Fourier transform of  $\varphi$ .

b) *Wavelet discretization*: In the wavelet formalism, some constraints apply on  $\varphi$ . It must be a scaling function that satisfies the properties for a multiresolution [25]. In that case, the wavelets can be defined as linear combinations of the  $\varphi_{\mathbf{p}}$  and the object is equivalently characterized by its coefficients in the orthonormal wavelet basis. We refer to Mallat's book [26] for a full review on wavelets. There exists a discrete wavelet transform (DWT) that bijectively maps the coefficients  $\mathbf{c}$  to the wavelet coefficients  $\mathbf{w}$  that represent the same object  $\rho$  in a continuous wavelet basis. In the rest of the paper, we represent this DWT by the synthesis matrix  $\mathbf{W}$ . Note that the matrix multiplications  $\mathbf{c} = \mathbf{W}\mathbf{w}$  and  $\mathbf{w} = \mathbf{W}^{-1}\mathbf{c}$  have efficient filterbank implementations.

### B. Matrix Representation of the Model

Since a FOV determines a finite number  $M$  of coefficients  $c[\mathbf{p}]$ , we handle them as a vector  $\mathbf{c}$ , keeping the discrete coordinates  $\mathbf{p}$  as implicit indexing. By simulating the imaging of the object (2), and by evaluating (1) for  $\mathbf{k} = \mathbf{k}_n$ , we find that the noise-free measurements are given by

$$\mathbf{m}_0 = \mathbf{E}\mathbf{c} \quad (5)$$

where  $\mathbf{E}$ , the MRI system matrix, is decomposed as

$$\mathbf{E} = \text{diag}(\hat{\varphi}(\mathbf{k}_n)) [\mathbf{s}_1 \dots \mathbf{s}_N]^T. \quad (6)$$

There,  $\mathbf{s}_n$  is a space-domain vector such that  $s_n[\mathbf{p}] = e^{-j\langle \omega_n, \mathbf{p} \rangle}$ .

A more realistic data-formation model is

$$\mathbf{m} = \mathbf{E}\mathbf{c} + \mathbf{b} \quad (7)$$

or

$$\mathbf{m} = \mathbf{M}\mathbf{w} + \mathbf{b} \quad (8)$$

with  $\mathbf{M} = \mathbf{E}\mathbf{W}$  and a residual vector  $\mathbf{b}$  representing the effect on measurements of noise and scanner imprecisions. The inverse problem of MRI is then to recover the  $M$  coefficients  $\mathbf{w}$  (or  $\mathbf{c}$ ) from the  $N$  corrupted measurements  $\mathbf{m}$ . Its degree of difficulty depends on the magnitude of the noise  $\mathbf{b}$  and the conditioning of the matrix  $\mathbf{M}$  (or  $\mathbf{E}$ ).

### C. Variational Formulation

1) *General Framework*: The solution  $\mathbf{w}^*$  is defined as the minimizer of a cost function that involves two terms: the data fidelity  $\mathcal{F}(\mathbf{b})$  and the regularization  $\mathcal{R}(\mathbf{w})$  that penalizes undesirable solutions. This is summarized as

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \mathcal{F}(\mathbf{m} - \mathbf{M}\mathbf{w}) + \lambda \mathcal{R}(\mathbf{w}) \quad (9)$$

where the regularization parameter  $\lambda \geq 0$  balances the two constraints. In MRI,  $\mathbf{b}$  is usually assumed to be a realization of a white Gaussian process, which justifies the choice  $\mathcal{F}(\mathbf{b}) = \|\mathbf{b}\|_{\ell_2}^2$  as a proper log-likelihood term. The ill-conditioning inherent to undersampled trajectories imposes the use of the suitable regularization term  $\mathcal{R}(\mathbf{w})$ .

Standard Tikhonov regularization corresponds to the quadratic term  $\mathcal{R}(\mathbf{w}) = \|\mathbf{R}\mathbf{w}\|_{\ell_2}^2$  and leads to the closed-form linear solution  $\mathbf{w}^* = (\mathbf{M}^H \mathbf{M} + \lambda \mathbf{R}^H \mathbf{R})^{-1} \mathbf{M}^H \mathbf{m}$  that is tractable both theoretically and numerically. When the reconstruction problem is sufficiently overdetermined to make noise propagation negligible, regularization is dispensable and the least-squares solution, which corresponds to Tikhonov with  $\lambda = 0$ , is adequate. The approach is statistically optimal if the object can also be considered as a realization of a Gaussian process. Unfortunately, this assumption is hardly justified for typical MR images. The quality of the solution obtained by this means quickly breaks down when undersampling increases.

TV reconstruction is related to the sum of the Euclidean norms of the gradient of the object. In practice, it is defined as  $\mathcal{R}(\mathbf{w}) = \|\nabla \mathbf{c}\|_{\ell_1}$ , where the operator  $\nabla$  returns pixelwise the  $\ell_2$ -norm of finite differences. The use of TV regularization is particularly appropriate for piecewise-constant objects such as the Shepp–Logan (SL) phantom.

2) *Sparsity-Promoting Regularization*: The main idea in this work is to exploit the fact that the object can be well represented by few nonzero coefficients (sparse representation) in an orthonormal basis of  $M$  functions  $\phi_{\mathbf{p}}$ . Formally, we write that

- $\exists S \subset \mathbb{R}^2$ ,  $|S| \ll M$  (*sparse support*)
- and
- $\exists \mathbf{a} : \|\rho - \sum_{\mathbf{p} \in S} a[\mathbf{p}] \phi_{\mathbf{p}}\| \ll \|\rho\|$  (*small error*).

It is well documented that typical MRI images admit sparse representation in bases such as wavelets or block DCT [3].

The  $\ell_1$ -norm is a good measure of sparsity with interesting mathematical properties (e.g., convexity). Thus, among the candidates that are consistent with the measurements, we favor a solution whose wavelet coefficients have a small  $\ell_1$ -norm. Specifically, the solution is formulated as

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \mathcal{C}(\mathbf{w}) \quad (10)$$

with

$$\mathcal{C}(\mathbf{w}) = \|\mathbf{m} - \mathbf{M}\mathbf{w}\|_{\ell_2}^2 + \lambda \|\mathbf{w}\|_{\ell_1}. \quad (11)$$

This is the general solution for wavelet-regularized inverse problems considered by [13] as well as many other authors.

### III. WAVELET REGULARIZATION ALGORITHMS

In this section, we present reconstruction algorithms that handle constraints expressed in the wavelet domain while solving the minimization problem (10). By introducing weighted norms instead of simple Lipschitz constants, we revisit the principle of the standard ISTA algorithm and simplify the derivation and analysis of this class of algorithms. We end up with a novel algorithm that combines different acceleration strategies and we provide a convergence analysis. Finally, we propose an adaptation of the fast algorithm to implement the random-shifting technique that is commonly used to improve results in image restoration.

#### A. Weighted Norms

Let us first define the weighted norm corresponding to a positive-definite symmetric weight matrix  $\mathbf{X}$  as  $\|\mathbf{w}\|_{\mathbf{X}}^2 = \mathbf{w}^H \mathbf{X} \mathbf{w}$ . The requirement that the eigenvalues of  $\mathbf{X}$ —denoted  $\lambda(\mathbf{X})$ —must be positive leads to the norm property

$$0 < \lambda_{\min}(\mathbf{X}) \|\mathbf{w}\|_2^2 \leq \|\mathbf{w}\|_{\mathbf{X}}^2 \leq \lambda_{\max}(\mathbf{X}) \|\mathbf{w}\|_2^2, \quad \forall \|\mathbf{w}\| \neq 0.$$

#### B. Principle of ISTA Revisited

An important observation to understand ISTA is to see that the nonlinear shrinkage operation, sometimes called soft-thresholding, solves a minimization problem [27], with

$$\begin{aligned} \mathcal{T}_{\lambda}(u) &= (|u| - \min(\lambda/2, |u|)) \cdot \text{sgn}(u) \\ &= \arg \min_{w \in \mathbb{C}} |u - w|^2 + \lambda |w|. \end{aligned} \quad (12)$$

By separability of norms, this applies component-wise to vectors of  $\mathbb{C}^N$ :  $\mathcal{T}_{\lambda}(\mathbf{u}) = \arg \min_{\mathbf{w}} \|\mathbf{u} - \mathbf{w}\|_{\ell_2}^2 + \lambda \|\mathbf{w}\|_{\ell_1}$ . This means that the  $\ell_1$ -regularized denoising problem (i.e., when  $\mathbf{M}$  in (11) is the identity matrix) is precisely solved by a shrinkage operation.

The iterative shrinkage/thresholding algorithm (ISTA) [2], [13], also known as thresholded Landweber (TL), generates a sequence of estimates  $\mathbf{w}_n$  that converges to the minimizer  $\mathbf{w}^*$  of (11) when it is unique. The idea is to define at each step a new functional  $\mathcal{C}'(\mathbf{w}, \mathbf{w}_n)$  whose minimizer  $\mathbf{w}_{n+1}$  will be the next estimate

$$\mathbf{w}_{n+1} = \arg \min_{\mathbf{w}} \mathcal{C}'(\mathbf{w}, \mathbf{w}_n). \quad (13)$$

Two constraints must be considered for the definition of  $C'$ .

- 1) It is sufficient for the convergence of the algorithm that  $C'(\mathbf{w}, \mathbf{w}_n)$  is an upper bound of  $C(\mathbf{w})$  and matches it at  $\mathbf{w} = \mathbf{w}_n$ ; this guarantees that the sequence  $\{C(\mathbf{w}_n)\}$  is monotonically decreasing.
- 2) The inner minimization (13) should be performed by a simple shrinkage operation to ensure the rapidity and accuracy of the algorithm.

In accordance with Constraint (1),  $C'$  can take the generic quadratically augmented form

$$C'(\mathbf{w}, \mathbf{w}_n) = C(\mathbf{w}) + \|\mathbf{w} - \mathbf{w}_n\|_{\Lambda - \mathbf{M}^H \mathbf{M}}^2 \quad (14)$$

with the constraint that  $(\Lambda - \mathbf{M}^H \mathbf{M})$  is positive definite, where the weighting matrix  $\Lambda$  plays the role of a tuning parameter.

Then, ISTA corresponds to the trivial choice  $\Lambda = (L/2)\mathbf{I}$ , with the value of  $L$  chosen to be greater or equal to the Lipschitz constant of the gradient of  $\|\mathbf{M}\mathbf{w}\|_{\ell_2}^2$ , so that  $L \geq 2\lambda_{\max}(\mathbf{M}^H \mathbf{M})$ .

Let us define  $\mathbf{a} = \mathbf{M}^H \mathbf{m}$ ,  $\mathbf{A} = \mathbf{M}^H \mathbf{M}$ , and

$$\mathbf{z}_n = \mathbf{w}_n + \frac{2}{L}(\mathbf{a} - \mathbf{A}\mathbf{w}_n). \quad (15)$$

Then, using standard linear algebra, we can write

$$\mathbf{w}_{n+1} = \arg \min_{\mathbf{w}} \|\mathbf{w} - \mathbf{z}_n\|_{\ell_2}^2 + \frac{2\lambda}{L} \|\mathbf{w}\|_{\ell_1} \quad (16)$$

$$= \mathcal{T}_{\frac{2\lambda}{L}}(\mathbf{z}_n). \quad (17)$$

This shows that Constraint (2) is automatically satisfied.

Note that both the intermediate variable  $\mathbf{z}_n$  in (15) and the threshold values will vary depending on  $L$ .

---



---

#### Algorithm 1: ISTA

---



---

**Repeat**  $\mathbf{w}_{n+1} \leftarrow \mathcal{T}_{2\lambda/L}(\mathbf{w}_n + (2/L)(\mathbf{a} - \mathbf{A}\mathbf{w}_n))$ ;

Beck and Teboulle [18, Thm. 3.1] showed that this algorithm decreases the cost function in direct proportion to the number of iterations  $n$ , so that  $C(\mathbf{w}_n) - C(\mathbf{w}^*) = \mathcal{O}(1/n)$ . Here, we present a slightly extended version of their original result which is valid for any reference point during iterations.

*Proposition 1:* Let  $\{\mathbf{w}_n\}$  be the sequence generated by Algorithm 1 with  $L \geq 2\lambda_{\max}(\mathbf{A})$ . Then, for any  $n > n_0 \in \mathbb{N}$ ,

$$C(\mathbf{w}_n) - C(\mathbf{w}^*) \leq \frac{L/2}{n - n_0} \|\mathbf{w}_{n_0} - \mathbf{w}^*\|_{\ell_2}^2. \quad (18)$$

*Proof:* [18, Thm. 3.1] gives the result for  $n_0 = 0$ . Consider a sequence  $\mathbf{w}'_n$  such that  $\mathbf{w}'_{n_0} = \mathbf{w}_0$ . As the iteration does not depend on  $n$ , we get  $\mathbf{w}'_n = \mathbf{w}_{n-n_0}$ . The result follows immediately. ■

Selecting  $L$  as small as possible will clearly favor the speed of convergence. It also raises the importance of a “warm” starting point.

Among the variants of ISTA, FISTA, proposed by Beck and Teboulle [18], ensures state-of-the-art convergence properties while preserving a comparable computational cost. Thanks to a controlled over-relaxation at each step, FISTA quadratically decreases the cost function, with  $C(\mathbf{w}_n) - C(\mathbf{w}^*) = \mathcal{O}(1/n^2)$ .

#### C. Subband Adaptive ISTA (SISTA)

SISTA is an extension of the ISTA that was introduced by Bayram and Selesnick [22]. Here, we propose an interpretation of SISTA as a particular case of (14) with a weighting matrix that replaces advantageously the step size  $2/L$  in (15). The idea is to use a diagonal weighting matrix  $\Lambda^{-1} = \text{diag}(\boldsymbol{\tau})$ —if it is not diagonal, Constraint (2) would not be fulfilled—with coefficients that are constant within a wavelet subband.

1) *SISTA:* In the same fashion as for ISTA, (15) and (17) can be adapted to subband-dependent steps and thresholds. Accordingly, SISTA is described in Algorithm 2.

---



---

#### Algorithm 2: SISTA

---



---

**Repeat**  $\mathbf{w}_{n+1} \leftarrow \mathcal{T}_{\lambda\boldsymbol{\tau}}(\mathbf{w}_n + \Lambda^{-1}(\mathbf{a} - \mathbf{A}\mathbf{w}_n))$ ;

2) *Convergence Analysis:* By considering the weighted scalar product  $\mathbf{x}^H \Lambda \mathbf{y}$  instead of  $(L/2)\mathbf{x}^H \mathbf{y}$ , we can adapt the convergence proof of ISTA by Beck and Teboulle (see Proposition 1). This result is new, to the best of our knowledge.

*Proposition 2:* Let  $\{\mathbf{w}_n\}$  be the sequence generated by Algorithm 2 with  $(\Lambda - \mathbf{A}) > 0$ . Then, for any  $n > n_0 \in \mathbb{N}$ ,

$$C(\mathbf{w}_n) - C(\mathbf{w}^*) \leq \frac{1}{n - n_0} \|\mathbf{w}_{n_0} - \mathbf{w}^*\|_{\Lambda}^2. \quad (19)$$

*Proof:* We rewrite the cost function (11) with the change of variable  $\mathbf{w}' = \Lambda^{1/2}\mathbf{w}$ . We then apply ISTA to solve that problem with  $\mathbf{a}' = \Lambda^{-1/2}\mathbf{a}$ ,  $\mathbf{A}' = \Lambda^{-1/2}\mathbf{A}\Lambda^{-1/2}$ , and  $L = 2$  (Note that  $\Lambda - \mathbf{A}$  is positive-definite if  $\mathbf{I} - \mathbf{A}'$  is positive-definite). The iteration  $\mathbf{w}'_{n+1} = \mathcal{T}_{\lambda}(\mathbf{w}'_n + \Lambda^{-1}(\mathbf{a}' - \mathbf{A}'\mathbf{w}'_n))$  can be rewritten, in terms of the original variable, as  $\mathbf{w}_{n+1} = \mathcal{T}_{\lambda\boldsymbol{\tau}}(\mathbf{w}_n + (\mathbf{a} - \mathbf{A}\mathbf{w}_n))$ . The latter is an iteration of SISTA (see Algorithm 2). According to Proposition 1, we have  $C(\Lambda^{-1/2}\mathbf{w}'_n) - C(\mathbf{w}^*) \leq \|\mathbf{w}'_{n_0} - \Lambda^{1/2}\mathbf{w}^*\|^2 / (n - n_0)$ , which translates directly into the proposed result. ■

Therefore, by comparing Propositions 1 and 2, an improved convergence is expected. The main point is that, for a “warm” starting point  $\mathbf{w}_0$  or after few iterations ( $n_0$ ), the weighted norm in (19) can yield significantly smaller values than the one weighted by  $L/2$  in (18).

3) *Selection of Weights:* Bayram and Selesnick [22] provide a method to select the values of  $\boldsymbol{\tau}$  for SISTA. To present this result, let us introduce some notations. We denote by  $s$  an index that scans all the  $S$  wavelet subbands, coarse scale included, by  $\tau_s$  the corresponding weight constant, and by  $\mathbf{M}_s$  the corresponding block of  $\mathbf{M}$ . We also define  $\gamma_{s_1, s_2} = \sqrt{\lambda_{\max}(\mathbf{M}_{s_2}^H \mathbf{M}_{s_1} \mathbf{M}_{s_1}^H \mathbf{M}_{s_2})}$ . The authors of [22] show that, for each subband, the condition

$$\frac{1}{\tau_s} > \sum_{s'=1}^S \gamma_{s, s'} \quad (20)$$

is sufficient to impose the positive definiteness of  $(\Lambda - \mathbf{M}^H \mathbf{M})$  that is required in (14). In the present context, we propose to compute the values  $\gamma_{s, s'}$  by using the power iteration method, once for a given wavelet family and k-space sampling strategy.



Fig. 1. Reference images from left to right: *in vivo* brain, SL reference, and wrist.

#### D. Best of Two Worlds: Fast Weighted ISTA (FWISTA)

Taking advantage of the ideas developed previously, we derive an algorithm that corresponds to the subband adaptive version of FISTA. In the light of the minimization problem (11), FWISTA generalizes the FISTA algorithm using a parametric weighted norm. We give its detailed description in Algorithm 3, where the modifications with respect to FISTA is the SISTA step in the loop.

---

#### Algorithm 3: FWISTA

---

**input:**  $\mathbf{A}$ ,  $\mathbf{a}$ ,  $\mathbf{w}_0$ , and  $\mathbf{\Lambda}^{-1} = \text{diag}(\boldsymbol{\tau})$ ;

Initialization:  $n = 0$ ,  $\mathbf{v}_0 = \mathbf{w}_0$ ,  $t_0 = 1$ ;

**repeat**

$$\mathbf{w}_{n+1} \leftarrow \mathcal{T}_{\lambda\tau}(\mathbf{v}_n + \mathbf{\Lambda}^{-1}(\mathbf{a} - \mathbf{A}\mathbf{v}_n)) \text{ (SISTA step);}$$

$$t_{n+1} \leftarrow (1 + \sqrt{1 + 4t_n^2})/2;$$

$$\mathbf{v}_{n+1} \leftarrow \mathbf{w}_{n+1} + (t_n - 1)/(\mathbf{w}_{n+1} - \mathbf{w}_n)t_{n+1};$$

$$n \leftarrow n + 1;$$

**until** *desired tolerance is reached*;

**return**  $\mathbf{w}_{n+1}$ ;

---

In the same fashion as for SISTA, we revisit the convergence results of FISTA [18, Thm. 4.4] for FWISTA.

*Proposition 3:* Let  $\{\mathbf{w}_n\}$  be the sequence generated by Algorithm 3. Then, for any  $n \geq 1$

$$\mathcal{C}(\mathbf{w}_n) - \mathcal{C}(\mathbf{w}^*) \leq \left(\frac{2}{n+1}\right)^2 \|\mathbf{w}_0 - \mathbf{w}^*\|_{\mathbf{\Lambda}}^2. \quad (21)$$

*Proof:* In the spirit of the proof of Proposition 2, we consider the change of variable  $\mathbf{w}' = \mathbf{\Lambda}^{1/2}\mathbf{w}$  and apply FISTA to solve the new reconstruction problem. The step  $\mathbf{w}'_{n+1} = \mathcal{T}_{\lambda}(\mathbf{v}'_n + (\mathbf{a}' - \mathbf{A}'\mathbf{v}'_n))$  is equivalent to  $\mathbf{w}_{n+1} = \mathcal{T}_{\lambda\tau}(\mathbf{v}_n + \mathbf{\Lambda}^{-1}(\mathbf{a} - \mathbf{A}\mathbf{v}_n))$ . The convergence results of FISTA

[18, Thm. 4.4] applies on the sequence  $\{\mathbf{w}'_n\}$ , which leads to  $\mathcal{C}(\mathbf{\Lambda}^{-1/2}\mathbf{w}'_n) - \mathcal{C}(\mathbf{w}^*) \leq 4\|\mathbf{w}'_0 - \mathbf{\Lambda}^{1/2}\mathbf{w}^*\|^2/(n+1)^2$ . ■

This result shows the clear advantage of FWISTA compared to ISTA (Proposition 1) and SISTA (Proposition 2). Moreover, we note that FWISTA can be simply adapted in order to impose a monotonic decrease of the cost functional value, in the same fashion as MFISTA. The same convergence properties apply [19, Thm. 5.1].

#### E. Random Shifting

Wavelet bases perform well the compression of signals but can introduce artifacts that can be attributed to their relative lack of shift-invariance. In the case of regularization, this can be avoided by switching to a redundant dictionary. The downside, however, is a significant increase in computational cost. Alternatively, the practical technique referred to as random shifting (RS) [2] can be used. Applying random shifting is much simpler and computationally more efficient than considering redundant transforms and leads to sensibly improved reconstruction.

Here, we propose a variational interpretation that motivates our implementation of FWISTA with RS (see Algorithm 4). We consider the DWT  $[\mathbf{W}_1 \cdots \mathbf{W}_{N_s}]^H$ , with  $\mathbf{W}_i = \mathbf{S}_i\mathbf{W}$ , where  $\mathbf{S}_i$  represent the different shifting operations required to get a translation-invariant DWT. The desired reconstruction would be defined as the minimizer of

$$\mathcal{C}(\mathbf{c}) = \|\mathbf{m} - \mathbf{E}\mathbf{c}\|_{\ell_2}^2 + \frac{\lambda}{N_s} \left\| [\mathbf{W}_1 \cdots \mathbf{W}_{N_s}]^H \mathbf{c} \right\|_{\ell_1}. \quad (22)$$

In 1D, this formulation includes TV regularization, in other words a single-level undecimated Haar WT without coarse-scale thresholding.

Rewriting (22) in terms of wavelet coefficients, we get

$$N_s \mathcal{C}(\mathbf{c}) = \sum_i C_i(\mathbf{W}_i^{-1}\mathbf{c}) \quad (23)$$

with

$$C_i(\mathbf{w}_i) = \left\| \mathbf{m} - \mathbf{M}\mathbf{S}_i^{-1}\mathbf{w}_i \right\|_{\ell_2}^2 + \lambda \|\mathbf{w}_i\|_{\ell_1}. \quad (24)$$

For a current estimate, we select a transform  $\mathbf{W}_i$  and perform a step in the minimization of the cost with respect to  $\mathbf{w}_i$  while keeping  $\mathbf{w}_j$ , for  $j \neq i$  fixed. A SISTA step is used, as the minimization subproblem (24) takes the form (11). The minimizers of the functionals  $\mathcal{C}_i$  are expected to correspond to images close to each other and to the minimizer of (22). In the first iterations of the algorithm, the minimization steps with respect to any  $\mathbf{w}_i$  are functionally equivalent (i.e., the modification is mostly explained by the gradient step). This motivates the use of FWISTA iterations at first and the switch to ISTA steps afterwards.

As the scheme is intrinsically greedy, we do not have a theoretical guarantee of convergence. Yet, in practice, we have observed that the SER stabilizes at a much higher value than it does when using ISTA schemes with no RS (cf Fig. 4).

---



---

#### Algorithm 4: FWISTA with RS

---



---

**input:**  $\mathbf{A} = \mathbf{E}^H \mathbf{E}$ ,  $\mathbf{a} = \mathbf{E}^H \mathbf{m}$ ,  $\mathbf{c}_0$ ,  $\mathbf{\Lambda}^{-1} = \text{diag}(\boldsymbol{\tau})$ ,  $K$ ;

Consider the sequence of DWT with RS:  $\mathbf{W}_n$ ;

Initialization:  $n = 0$ ,  $k = 0$ ,  $\gamma = 1$ ,  $t_0 = 1$ ,  $C_0 = \mathcal{C}(\mathbf{c}_0)$ ,  $\mathbf{y}_0 = \mathbf{c}_0$ ;

**repeat**

$$\mathbf{w}_n \leftarrow \mathbf{W}_n^{-1} \mathbf{y}_n;$$

$$\mathbf{c}_{n+1} \leftarrow \mathbf{W}_n \mathcal{T}_{\lambda \boldsymbol{\tau}}(\mathbf{w}_n + \mathbf{\Lambda}^{-1} \mathbf{W}_n^{-1} (\mathbf{a} - \mathbf{A} \mathbf{y}_n));$$

$$C_{n+1} \leftarrow \mathcal{C}(\mathbf{c}_{n+1}) \text{ (cost-function evaluation);}$$

**if**  $C_{n+1} > C_n$  **then**

$$k \leftarrow k + 1;$$

**if**  $k = K$  **then**

$$\gamma \leftarrow 0;$$

$$\mathbf{\Lambda}^{-1} \leftarrow \max\{\boldsymbol{\tau}\} \mathbf{I};$$

$$\boldsymbol{\tau} \leftarrow \max\{\boldsymbol{\tau}\} \mathbf{1};$$

$$t_{n+1} \leftarrow (1 + \sqrt{1 + 4t_n^2})/2;$$

$$\mathbf{y}_{n+1} \leftarrow \mathbf{c}_{n+1} + \gamma(t_n - 1)(\mathbf{c}_{n+1} - \mathbf{c}_n)/t_{n+1};$$

$$n \leftarrow n + 1;$$

**until** *stopping condition is met*;

---

## IV. EXPERIMENTS

### A. Implementation Details

Our implementation uses Matlab 7.9 (Mathworks, Natick, MA). The reconstructions run on a 64-bit 8-core computer, clock rate 2.8 GHz, 8 GB RAM (DDR2 at 800 MHz), Mac OS X 10.6.5. For all iterative algorithms, a key point is that matrices are not stored in memory. They only represent operations that are performed on vectors (images). In particular,  $\mathbf{a} = \mathbf{M}^H \mathbf{m}$  is computed once per dataset. Matrix-to-vector multiplication with  $\mathbf{A} = \mathbf{M}^H \mathbf{M}$ , specifically,  $\mathbf{E}^H \mathbf{E}$ , have an efficient implementation thanks to the convolution structure of the problem [28], [29]. For these Fourier precomputations, we

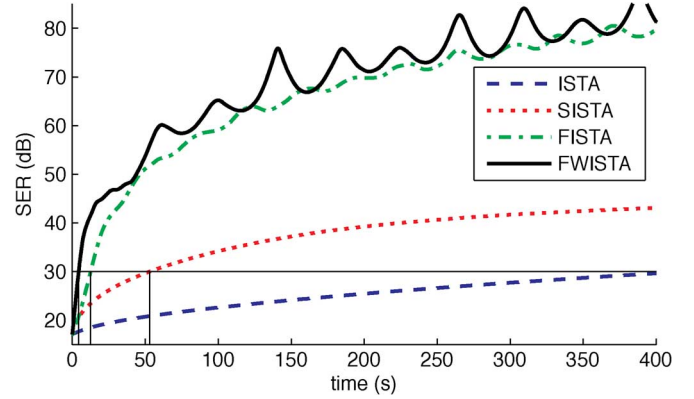


Fig. 2. Time evolution of the SER with respect to the minimizer for several ISTA algorithms. Times to reach 35 dB are delineated.

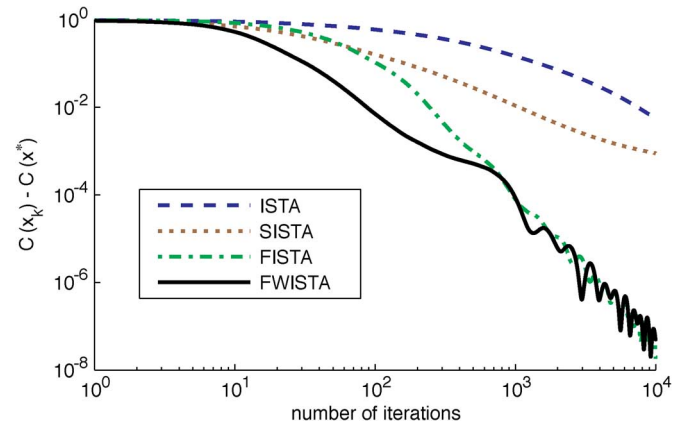


Fig. 3. Time evolution of the difference in cost function value with respect to the minimizer for several ISTA algorithms.

used the NUFFT algorithm [30] that is made available online.<sup>1</sup> For wavelet transforms, we used the code provided online<sup>2</sup> [21]. This Fourier-domain implementation proved to be faster than Matlab's when considering reconstructed images smaller than  $256 \times 256$  and the Haar wavelet. It must also be noted that the 2D-DFT were performed using the FFTW library which efficiently parallelizes computations.

For Tikhonov regularizations, we implemented the classical conjugate gradient (CG) algorithm, with the identity as the regularization matrix. For TV regularizations, we considered the iteratively reweighted least-squares algorithm (IRLS), which corresponds to the additive form of half-quadratic minimization [31], [32]. We used 15 iterations of CG to solve the linear inner problems, always starting from the current estimate, which is crucial for efficiency. For the weights that permit the quadratic approximation of the TV term, we stabilized the inversion of very small values.

We implemented ISTA, SISTA, and FWISTA as described in Section III, with the additional possibility to use random shifting (see Section IV-C2). For the considered reconstructions using our method, described in Algorithm 4,  $K = 30$  was a reasonable choice. The Haar wavelet transform was used, with three decomposition levels when no other values are mentioned. As is

<sup>1</sup><http://www.eecs.umich.edu/~fessler/code/>

<sup>2</sup><http://bigwww.epfl.ch/algorithms/mltldconvolution>

TABLE I  
GLOSSARY

<b>Acronyms</b>	
MRI	Magnetic resonance imaging
ISTA	Iterative shrinkage/thresholding algorithm
(F)(W)(S)ISTA	(fast) (weighted) (subband adaptive) ISTA
TV	Total variation
FOV	Field of view
CS	Compressed sensing
DWT	Discrete wavelet transform
DCT	Discrete cosine transform
DFT	Discrete Fourier transform
CG	Conjugate gradient
IRLS	Iteratively reweighted least squares
ROI	Region of interest
MSE	Mean-squared error
S(E)(N)R	Signal-to-(error) (noise) ratio
RS	Random shifting
<b>Continuous Domain and Functions</b>	
$\mathbf{r}$	$\in \mathbb{R}^2$ spatial coordinates (XY plane)
$\mathbf{k}$	$\in \mathbb{R}^2$ k-space coordinates (XY plane)
$\rho(\mathbf{r})$	$\in \mathbb{R}^+$ object (proton density) in space
$m(\mathbf{k})$	$\in \mathbb{C}$ observation of the object in k-space
$\varphi(\mathbf{r})$	$\in \mathbb{R}$ generating function
$\hat{f}(\mathbf{k})$	$\in \mathbb{C}$ function $f$ in the k-space domain
$C$	cost function
$\mathcal{T}_\tau$	shrinkage operator with thresholds $\tau$
<b>Discrete Data and Linear Algebra</b>	
$\mathbf{p}$	$\in \mathbb{Z}^2$ discrete spatial coordinates
$M$	$\in \mathbb{N}$ number of pixels in the ROI
$N$	$\in \mathbb{N}$ number of k-space samples
$\mathbf{k}_n$	$\in \mathbb{R}^2$ $n$ th k-space sampling position
$m_n$	$\in \mathbb{C}$ $n$ th k-space observation
$\mathbf{m}$	$\in \mathbb{C}^N$ measurement vector
$\mathbf{b}$	$\in \mathbb{C}^N$ noise vector
$\mathbf{E}$	Fourier encoding matrix
$\mathbf{M}$	system matrix (wavelet domain to k-space)
$\mathbf{W}$	DWT matrix
$c[\mathbf{p}]$	$\in \mathbb{R}$ reconstructed spatial coefficient
$\mathbf{c}$	$\in \mathbb{R}^M$ vector of spatial coefficients
$\mathbf{w}$	$\in \mathbb{C}^M$ vector of wavelet coefficients
$\lambda_{\max}(\mathbf{X})$	largest eigenvalue of $\mathbf{X}$
$\mathbf{X}^H$	Hermitian transpose of the matrix $\mathbf{X}$
$\langle \mathbf{x}, \mathbf{y} \rangle$	$\in \mathbb{R}$ regular inner product
$\langle \mathbf{x}, \mathbf{y} \rangle_\Lambda$	$\in \mathbb{R}$ weighted inner product ( $\mathbf{x}^H \Lambda \mathbf{y}$ )
$\ \mathbf{x}\ _{\ell_2}$	$\in \mathbb{R}^+$ regular quadratic norm
$\ \mathbf{x}\ _\Lambda$	$\in \mathbb{R}^+$ weighted quadratic norm ( $\mathbf{x}^H \Lambda \mathbf{x}$ )
$\ \mathbf{x}\ _{\ell_1}$	$\in \mathbb{R}^+$ $\ell_1$ norm ( $\sum  x_i $ )

usual for wavelet-based reconstructions, the regularization was not applied to the coarse-level coefficients.

Reconstructions were limited to the pixels of the ROI for all algorithms. The regularization parameter  $\lambda$  was systematically adjusted such that the reconstruction mean-squared error (MSE) inside the ROI was minimal. For practical situations where the ground-truth reference is not available, it is possible to adjust  $\lambda$  by considering well-established techniques such as the discrepancy principle, generalized cross validation, or L-curve method [33].

**B. Spiral MRI Reconstruction**

In this section, we focused on the problem of reconstructing images of objects weighted by the receiving channel sensitivity, given undersampled measurements. This problem, which involves single-channel data and hence differs from SENSE, is challenging for classical linear reconstructions as it generates

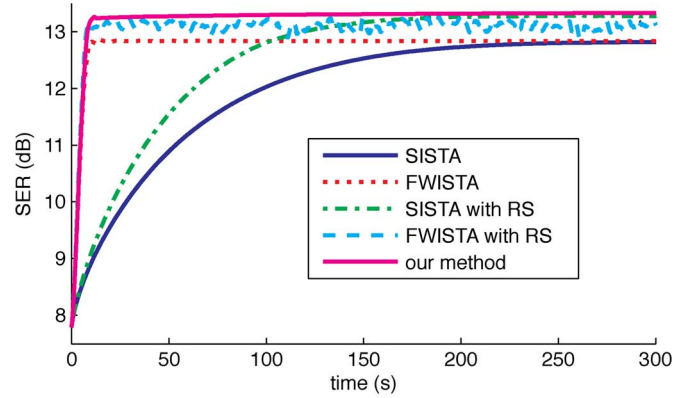


Fig. 4. Time evolution of the SER for several algorithms for the SL simulation using Haar wavelets.

artifacts and propagates noise. We considered spiral trajectories with 50 interleaves, with an interleave sampling density reduced by a factor  $R = 1.8$  compared to Nyquist for the highest frequencies and an oversampling factor 3.5 along the trajectory. Spiral acquisition schemes are attractive because of their versatility and the fact that they can be implemented with smooth gradient switching [34], [35].

We validate the results with the three sets of data that we present below. The corresponding reference images are shown in Fig. 1.

1) *MR Scanner Acquisitions*: The data were collected on a 3T Achieva system (Philips Medical Systems, Best, The Netherlands). A field camera with 12 probes was used to monitor the actual k-space trajectory [36]. An array of eight head coils provided the measurements. We acquired *in vivo* brain data from a healthy volunteer with parameters TR = 1000 ms and TE = 30 ms. The excitation slice thickness was 3 mm with a flip angle of 30°. The trajectory was designed for a FOV of 25 cm with a pixel size 1.5 mm. It was composed of 100 spiral interleaves. The interleaf distance for the highest sampled frequencies defined a fraction of the Nyquist sampling density ( $R = 0.9$ ).

The subset used for reconstruction corresponds to half of the 100 interleaves. The corresponding reduction factor, defined as the ratio of the distance of neighboring interleaves with the Nyquist distance, is  $R = 1.8$ .

2) *Analytical Simulation*: We used analytical simulations of the SL brain phantom with a similar coil sensitivity, following the method described in [37]. The values of these simulated data were scaled to have the same mean spatial value (i.e., the same central k-space peak) as the brain reference image. A realization of complex Gaussian noise was added to this synthetic k-space data, with a variance corresponding to 40 dB SNR. The  $176 \times 176$  rasterization of the analytical object provided a reliable reference for comparisons.

3) *Simulation of a Textured Object*: A second simulation was considered with an object that is more realistic than the SL phantom. We chose a  $512 \times 512$  MR image of a wrist that showed little noise and interesting textures. We simulated acquisitions with the same coil profile and spiral trajectory ( $176 \times 176$  reconstruction matrix), in presence of a 40 dB SNR



TABLE II  
VALUES OF THE OPTIMAL SER AND CORRESPONDING REGULARIZATION PARAMETERS ARE SHOWN FOR THE DIFFERENT WAVELET BASES

Experiment		SL simulation				Wrist simulation			
Wavelet basis		Haar	Spline 2	Spline 4	Spline 6	Haar	Spline 2	Spline 4	Spline 6
Without RS	SER (dB)	<b>12.65</b>	12.16	10.75	9.70	15.93	<b>17.33</b>	17.32	17.07
	$\lambda$	1 870	2 510	830	1 460	1 600	946	1 070	1 350
With RS	SER (dB)	<b>13.38</b>	12.53	11.58	10.38	<b>18.70</b>	18.24	18.05	17.87
	$\lambda$	5 650	3 900	7 770	1 370	1 490	850	1 190	1 260

TABLE III  
RESULTS OF THE PROPOSED WAVELET METHOD FOR DIFFERENT WAVELET DECOMPOSITION DEPTHS. VALUES OF THE REGULARIZATION PARAMETER, THE FINAL SER, THE RELATIVE MAXIMAL SPATIAL DOMAIN ERROR, AND THE TIME TO REACH  $-0.5$  dB OF THE FINAL SER

Experiment	SL simulation				Wrist simulation				Brain data			
	1	2	3	4	1	2	3	4	1	2	3	4
$\lambda$ opt.	5 330	5 680	5 110	5 200	1 500	1 520	1 570	1 700	9 650	10 800	11 400	11 400
SER (dB) opt.	13.25	13.34	<b>13.39</b>	13.35	18.54	18.66	<b>18.71</b>	18.70	18.78	<b>19.00</b>	18.99	18.96
$l_\infty$ error (%)	<b>50</b>	51	54	53	<b>15.5</b>	16.0	16.0	16.5	17.3	17.5	17.49	<b>17.0</b>
$t_{-0.5dB}$ (s)	6.56	<b>6.19</b>	6.93	7.59	5.88	<b>5.23</b>	5.33	4.25	8.99	7.03	<b>6.33</b>	7.97

TABLE IV  
RESULTS OF THE ALGORITHMS CG (LINEAR), IRLS (TV), AND OUR METHOD (WAVELETS) FOR DIFFERENT DEPTHS. VALUES OF THE REGULARIZATION PARAMETER, THE FINAL SER, THE RELATIVE MAXIMAL SPATIAL DOMAIN ERROR, AND THE TIME TO REACH  $-0.5$  dB OF THE FINAL SER

Experiment	SL simulation			Wrist simulation			Brain data		
	linear	TV	wavelets	linear	TV	wavelets	linear	TV	wavelets
$\lambda$ opt.	0.0247	4 090	6 380	0.436	760	1 620	0.471	6 050	16 800
SER (dB) opt.	8.46	<b>13.82</b>	13.17	16.14	18.41	<b>18.64</b>	15.81	18.88	<b>18.93</b>
$l_\infty$ error (%)	<b>48</b>	49	51	21	<b>16</b>	<b>16</b>	29	12	<b>11</b>
$t_{-0.5dB}$ (s)	<b>0.286</b>	18.1	5.40	<b>0.209</b>	10.5	4.64	<b>0.205</b>	15.2	6.13

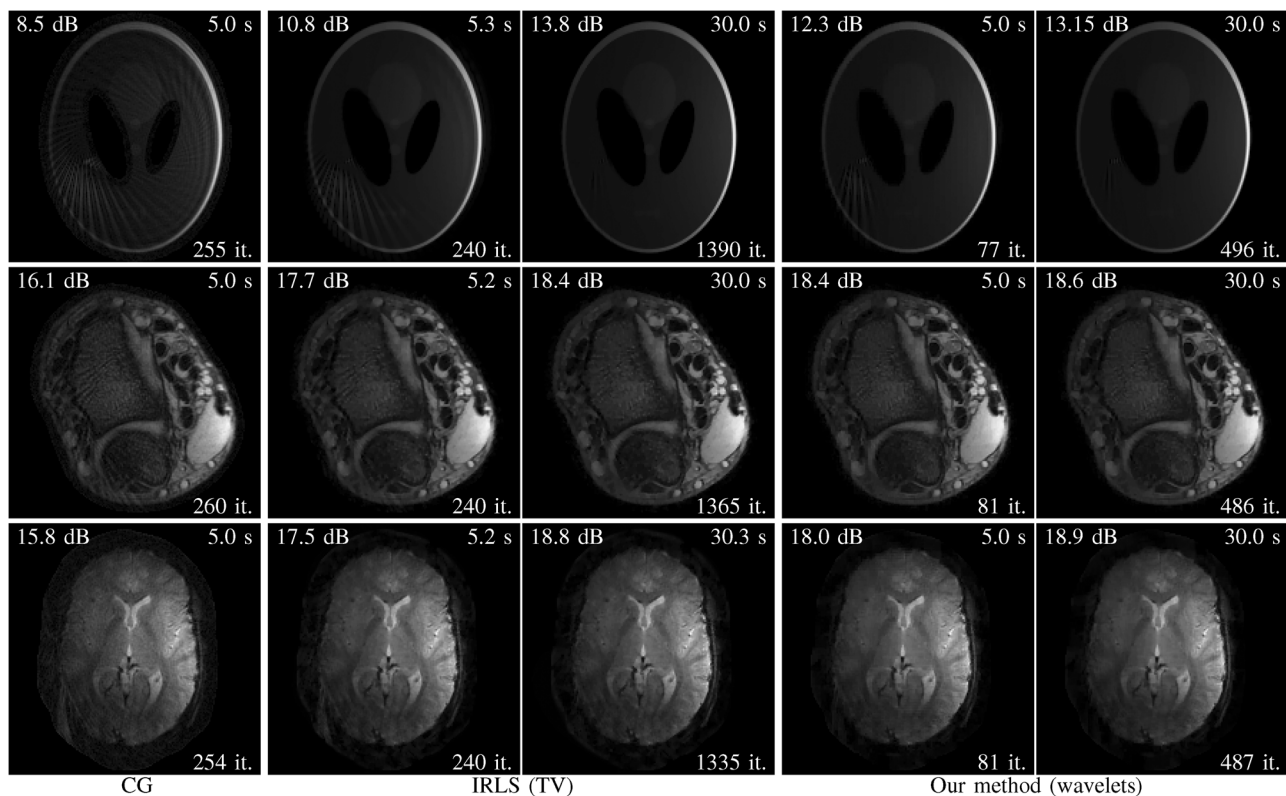


Fig. 5. Result of different reconstruction algorithms for the three experiments. For each reconstruction, the performance in SER with respect to the reference (top-left), the reconstruction time (top-right), and the number of iterations (bottom-right) are shown.

Gaussian complex noise. The height of the central peak was also adjusted to correspond to that of the brain data. The reference image was obtained by sinc-interpolation, by extracting the lowest frequencies in the DFT.

### C. Results

In this section, we present the different experiments we conducted. The two main reconstruction performance measures that we considered are as follows.



- Reconstruction duration, which excludes all aforementioned precomputations and the superfluous monitoring operations.
- Signal to error ratio with respect to a reference, defined as  $SER(\tilde{\rho}, \rho_{\text{ref}}) = \frac{\|\rho_{\text{ref}}\|_{\ell_2}}{\|\rho_{\text{ref}} - \tilde{\rho}\|_{\ell_2}}$  and  $SER_{\text{dB}}(\tilde{\rho}, \rho_{\text{ref}}) = 20 \log_{10}(SER(\tilde{\rho}, \rho_{\text{ref}}))$ . Practically, the references are either the ground-truth images or the minimizer of the cost functional. It is known that SER is not a foolproof measure of visual improvement but large SER values are encouraging and generally correlate with good image quality.

1) *Convergence Performance of IST-Algorithms*: In this first experiment, we compared the convergence properties of the different ISTA-type algorithms, as presented in Section III, with the Haar wavelet transform. The data we considered are those of the MR wrist image. The regularization parameter was adjusted to maximize the reconstruction SER with respect to the ground-truth data. The actual minimizer of the cost functional, which is the common fixed-point of this family of algorithms, was estimated by iterating FWISTA 100 000 times.

The convergence results are shown in Figs. 2 and 3 for the simulation of the MR wrist image. Similar graphs are obtained using the other sets of data.

For a fixed number of iterations, FISTA schemes (FISTA and FWISTA) require roughly 10% additional time compared to ISTA and SISTA. In spite of this fact, their asymptotic superiority appears clearly in both figures. The slope of the decrease of the cost functional in the log-log plot of Fig. 3 reflects the convergence properties in Propositions 1, 2, and 3. When considering the first iterations, which are of greatest practical interest, the algorithms with optimized parameters (SISTA and FWISTA) perform better than ISTA and FISTA (see Fig. 3). The times required by each algorithm to reach a 30 dB SER (considered as a threshold value to perceived changes) are 415 s (ISTA), 53 s (SISTA), 12.7 s (FISTA), and 4.4 s (FWISTA). With respect to this criterion, SISTA presents an eight-fold speedup over ISTA, while FWISTA presents a 12-fold speedup over SISTA and nearly a three-fold speedup over FISTA. It follows that FWISTA is practically close to two orders of magnitude faster than ISTA.

2) *Choice of the Wavelet Transform and Use of Random Shifting*: The algorithms presented in Section III apply for any orthogonal wavelet basis. For the considered application, we want to study the influence of the basis on performance. In this experiment we considered the Battle–Lemarié spline wavelets [26] with increasing degrees, taking into account the necessary postfilter mentioned in (4).

We compared the best results for several bases. They were obtained with FWISTA after practical convergence and are reported in Table II. Fig. 4 illustrates the time evolution of the SER using ISTA and FWISTA in the case of the SL reconstruction. Similar graphs are obtained with the other experiments.

It is known that the Haar wavelet basis efficiently approximates piecewise-constant objects like the SL phantom, which is consistent with our results. On the other hand, splines of higher degree, which have additional vanishing moments, perform better on the textured images (upper part of Table II).

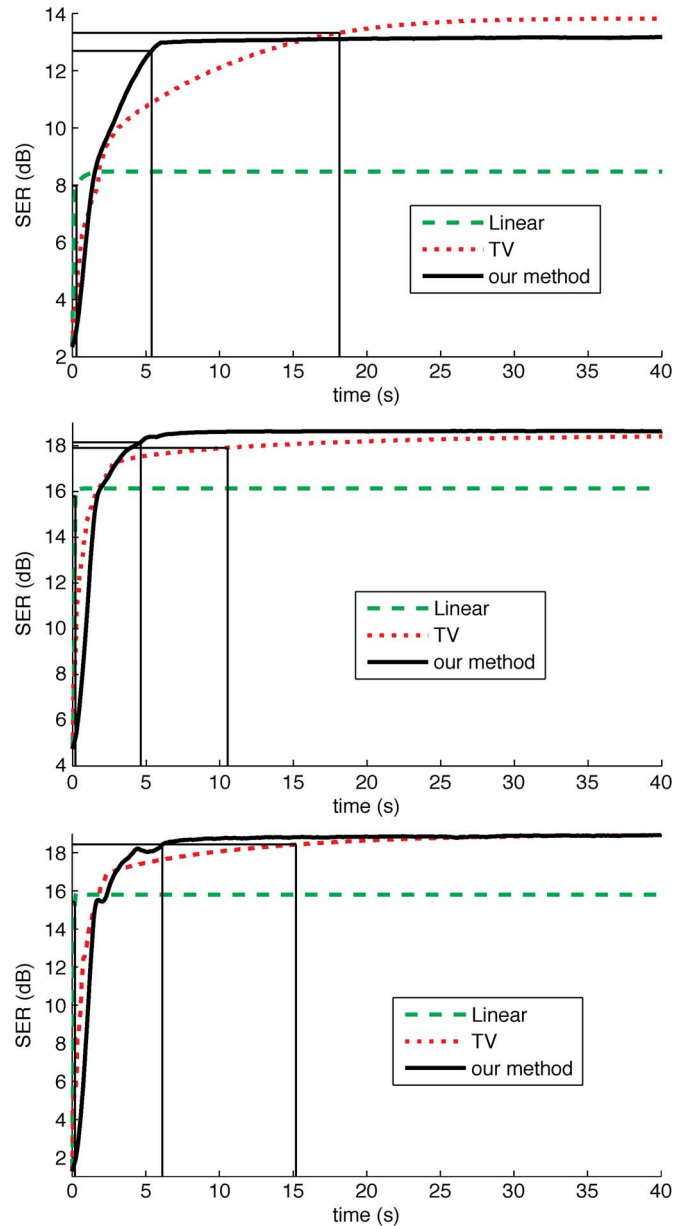


Fig. 6. Evolution of the performance of the algorithms. From top to bottom: SL simulation, wrist simulation, and brain data. Times required to reach  $-0.5$  dB of the asymptotic value are indicated.

We present in the lower part of Table II the performances observed when using ISTA with RS. We conclude that, in the case of realistic data, it is crucial to use RS as it improves results by at least 0.7 dB, whatever the wavelet basis is. The remarkable aspect there is that the Haar wavelet transform with RS consistently performs best. Two important things can be seen in Fig. 4: FWISTA is particularly efficient during the very first iterations, while SISTA with RS yields the best asymptotic results in terms of SER and stability. Our method combines both advantages.

In Table III, we present the results obtained using different depths of the wavelet decompositions. Our reconstruction method is used together with the Haar wavelet transform and RS. The performances are similar but there seems to be an advantage in using several decomposition levels both in terms of

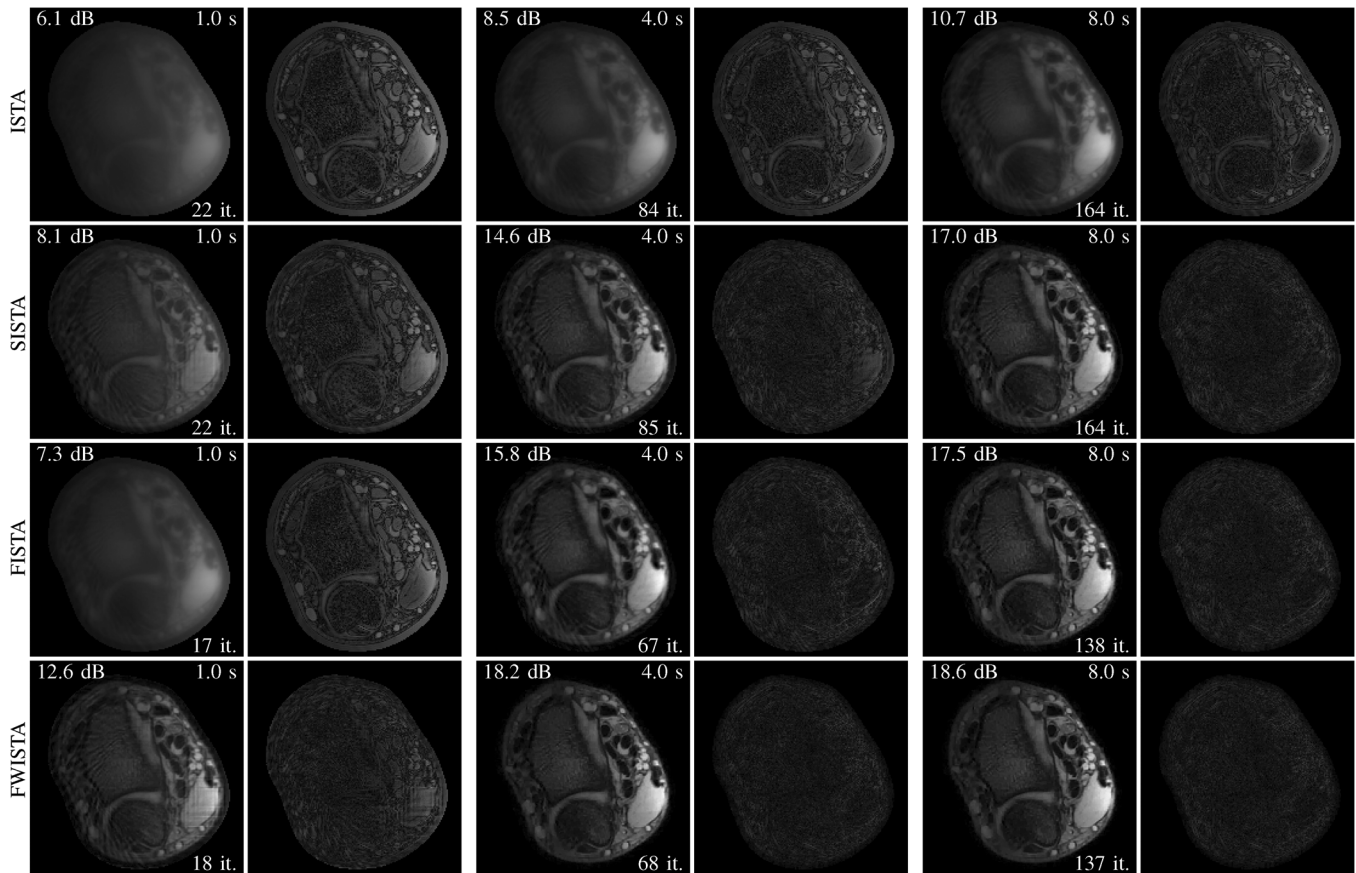


Fig. 7. Reconstructions and error maps of different IST-algorithms with RS for the wrist experiment. For each reconstruction, the performance in SER with respect to the reference (top-left), the reconstruction time (top-right), and the number of iterations (bottom-right) are shown.

SER and reconstruction speed. The FWISTA scheme seems to recoup the cost of the wavelet transform operations associated to an increase in the depth of decomposition.

3) *Practical Performance*: We report in Table IV the results obtained for different reconstruction experiments using state-of-the-art linear reconstruction, TV regularization, and our method. The images obtained when running the different algorithms after approximately 5 s, and after practical convergence as well, are shown in Fig. 5. We display in Fig. 6 the time evolution of the SER for the different experiments. In each case, we emphasize the time required to reach  $-0.5$  dB of the asymptotic value of SER. Finally, we present in Fig. 7 the reconstruction and error maps of the different IST-algorithms at different moments of reconstruction. This was done with the wrist simulated experiment using the Haar wavelet basis and RS.

Firstly, we observe that TV and our method achieve similar SER (Table IV) and image quality (Fig. 5). They both clearly outperform linear reconstruction, with a SER improvement from 1.5 to 3 dB, depending on the degree of texture in the original data. Moreover, the pointwise maximal reconstruction error appears to always be smaller with nonlinear reconstructions. Due to the challenging reconstruction task, which significantly undersamples of the  $k$ -space, residual artifacts remain in the linear reconstructions and at early stages of the nonlinear ones. Although the  $k$ -space trajectory is exactly the same in the three cases, artifacts are less perceived in the *in vivo* reconstructions, while they stand out for the synthetic experiments.

Secondly, it clearly appears that the linear reconstruction, implemented with CG, leads to the fastest convergence, unfortunately with suboptimal quality. For a reconstruction time one order of magnitude longer, our accelerated method provides better reconstructions. This is illustrated in Fig. 5 for reconstruction times of the order of five seconds (columns 1, 2, and 4).

Finally, we observe in Fig. 7 the superiority of the proposed FWISTA over the other types of algorithms. For the given reconstruction times, it consistently exhibits better image quality as can be seen in both reconstructions and error maps.

#### D. SENSE MRI Reconstruction

Our reconstruction method is applicable to linear MR imaging modalities. In this section, we report results obtained on a SENSE reconstruction problem. The data were acquired with the same scanner setup as in Section IV-B1. This time, the data from the eight receiving channels were used for reconstruction, as well as an estimation of the sensitivity maps. An *in vivo* gradient echo EPI sequence of a brain was performed with  $T_2^*$  contrast. The data was acquired with the following parameters: excitation slice thickness of 4 mm,  $TE = 35$  ms,  $TR = 900$  ms, flip angle of  $80^\circ$ , and trajectory composed of 13 interleaves, supporting a  $200 \times 200$  reconstruction matrix with pixel resolution  $1.18 \text{ mm} \times 1.18 \text{ mm}$ . The oversampling ratio along the readout direction was 1.62.

The reference image was obtained using the complete set of data and performing an unregularized CG-SENSE reconstruction.

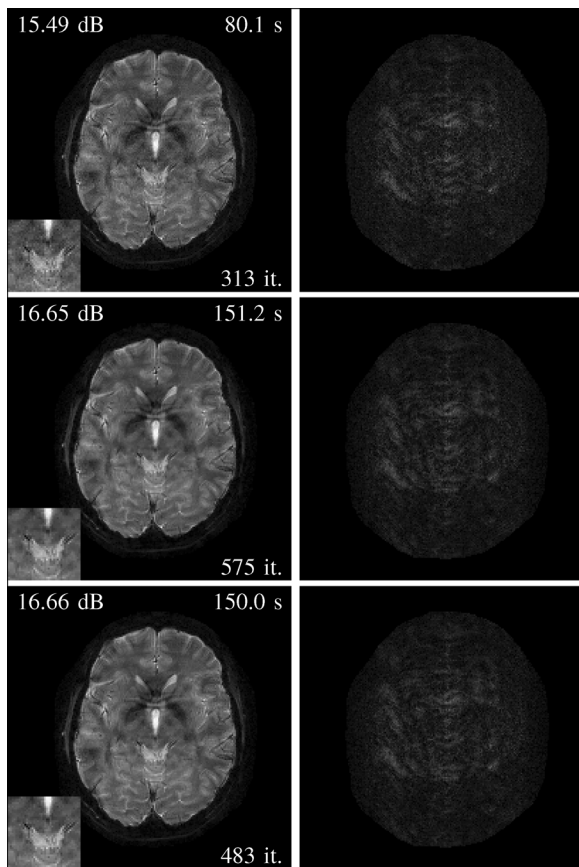


Fig. 8. Reconstructions (left column) and error maps (right column) for the SENSE EPI experiment using CG (first row), IRLS-TV (second row), and our method (third row). For each reconstruction, the performance in SER with respect to the reference (top-left corner), the reconstruction time (top-right corner), the number of iterations (bottom-right corner), and a magnification of the central part (bottom-left) are shown.

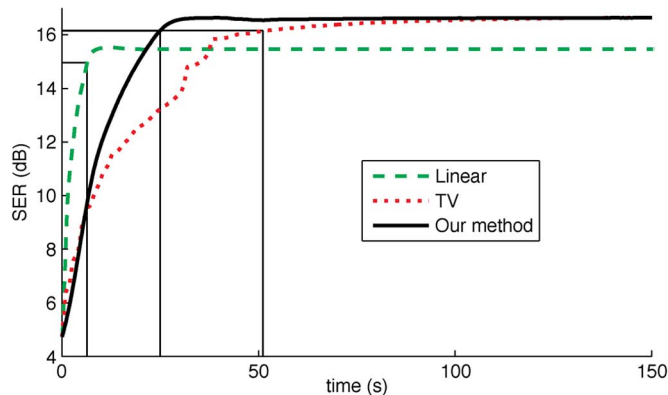


Fig. 9. Evolution of the performance of the algorithms. Times required to reach  $-0.5$  dB of the asymptotic value are indicated.

tion. The reconstruction involved 3 of the 13 interleaves, representing a significant undersampling ratio  $R = 4.33$ .

The images obtained using regularized linear reconstruction (CG), TV (IRLS), and our method are presented in Fig. 8. In Fig. 9, the SER evolution with respect to time is shown for the three methods. The times to reach  $-0.5$  dB of the asymptotic SER value are 5.9 s (CG), 49.1 s (IRLS), and 22.8 s (our method).

With this high undersampling, the errors maps show that reconstructions suffer from noise propagation mostly in the center of the image. It appears that TV and our method improve qualitatively and quantitatively image quality over linear reconstruction (cf. Fig. 8). As it was observed in Section IV-B with the spiral MRI reconstructions, this *in vivo* SENSE experiment confirms that our method is competitive with TV. In terms of reconstruction duration, our method proves to converge in a time that is of the same order of magnitude as CG (cf. Fig. 9).

## V. CONCLUSION

We proposed an accelerated algorithm for nonlinear wavelet-regularized reconstruction that is based on two complementary acceleration strategies: use of adaptive subband thresholds plus multistep update rule. We provided theoretical evidence that this algorithm leads to faster convergence than when using the accelerating techniques independently. In the context of MRI, the proposed strategy can accelerate the reference algorithm up to two orders of magnitude. Moreover, we demonstrated that, by using the Haar wavelet transform with random shifting, we are able to boost the performance of wavelet methods to make them competitive with TV regularization. Using different simulations and *in vivo* data, we compared the practical performance of our reconstruction method with other linear and nonlinear ones.

The proposed method is proved to be competitive with TV regularization in terms of image quality. It typically converges within five seconds for the single channel problems considered. This brings nonlinear reconstruction forward to an order of magnitude of the time required by the state-of-the-art linear reconstructions, while providing much better quality.

## ACKNOWLEDGMENT

The authors would like to thank J. Fessler and C. Vonesch to make their implementations of NUFFT and multidimensional wavelet transform available. We acknowledge C. Vonesch, I. Bayram, and D. Van De Ville for fruitful discussions. The authors would also like to thank P. Thévenaz for useful comments and careful proofreading.

## REFERENCES

- [1] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn. Reson. Med.*, vol. 42, no. 5, pp. 952–962, 1999.
- [2] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Signal Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.
- [3] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, pp. 1182–1195, 2007.
- [4] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magn. Reson. Med.*, vol. 59, no. 2, pp. 365–373, 2008.
- [5] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint," *Magn. Reson. Med.*, vol. 57, no. 6, pp. 1086–1098, 2007.
- [6] L. Ying, L. Bo, M. C. Steckner, W. Gaohong, W. Min, and L. Shi-Jiang, "A statistical approach to SENSE regularization with arbitrary k-space trajectories," *Magn. Reson. Med.*, vol. 60, pp. 414–421, 2008.
- [7] B. Liu, K. King, M. Steckner, J. Xie, J. Sheng, and L. Ying, "Regularized sensitivity encoding (SENSE) reconstruction using Bregman iterations," *Magn. Reson. Med.*, vol. 61, no. 1, pp. 145–152, 2009.
- [8] M. Lustig, J. H. Lee, D. L. Donoho, and J. M. Pauly, "Faster imaging with randomly perturbed, undersampled spirals and  $|L|_1$  reconstruction," in *Proc. ISMRM*, 2005.

- [9] B. Liu, E. Abdelsalam, J. Sheng, and L. Ying, "Improved spiral sense reconstruction using a multiscale wavelet model," in *Proc. ISBI*, 2008, pp. 1505–1508.
- [10] L. Chaâri, J.-C. Pesquet, A. Bebazza-Benyahia, and P. Ciucu, "Auto-calibrated regularized parallel MRI reconstruction in the wavelet domain," in *Proc. ISBI*, 2008, pp. 756–759.
- [11] M. Guerquin-Kern, D. Van De Ville, C. Vonesch, J.-C. Baritoux, K. P. Pruessmann, and M. Unser, "Wavelet-regularized reconstruction for rapid MRI," in *Proc. ISBI*, 2009, pp. 193–196.
- [12] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, "A  $\ell^1$ -unified variational framework for image restoration," *Lecture Notes in Computer Science*, vol. 3024, pp. 1–13, 2004.
- [13] I. Daubechies, M. DeFrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [14] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.
- [15] Y. E. Nestorov, Gradient methods for minimizing composite objective function CORE Rep., Tech. Rep., 2007.
- [16] P. Weiss, "Algorithmes rapides d'optimisation convexe. Application à la restauration d'images et à la détection de changements," Ph.D. dissertation, Univ. Nice, Nice, France, Dec. 2008.
- [17] S. Becker, J. Bobin, and E. J. Candès, NESTA: A fast and accurate first-order method for sparserecovery California Inst. Technol., Pasadena, 2009, Tech. Rep..
- [18] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
- [19] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.
- [20] C. Vonesch and M. Unser, "A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 539–549, Apr. 2008.
- [21] C. Vonesch and M. Unser, "A fast multilevel algorithm for wavelet-regularized image restoration," *IEEE Trans. Image Process.*, vol. 18, no. 3, pp. 509–523, Mar., 2009.
- [22] İ. Bayram and I. W. Selesnick, "A subband adaptive iterative shrinkage/thresholding algorithm," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1131–1143, Mar. 2010.
- [23] B. P. Sutton, D. C. Noll, and J. A. Fessler, "Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 178–188, Feb. 2003.
- [24] B. Delattre, J.-N. Hyacinthe, J.-P. Vallée, and D. Van De Ville, "Spline-based variational reconstruction of variable density spiral k-space data with automatic parameter adjustment," in *Proc. ISMRM*, 2009, p. 2066.
- [25] M. Unser and T. Blu, "Wavelet theory demystified," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 470–483, 2003.
- [26] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1999.
- [27] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier, "Non-linear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Signal Process.*, vol. 7, no. 3, pp. 319–335, Mar. 1998.
- [28] F. T. W. A. Wajer and K. P. Pruessmann, "Major speedup of reconstruction for sensitivity encoding with arbitrary trajectories," in *Proc. ISMRM*, 2001, p. 625.
- [29] J. A. Fessler, S. Lee, V. T. Olafsson, H. R. Shi, and D. C. Noll, "Toeplitz-based iterative image reconstruction for MRI with correction for magnetic field inhomogeneity," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3393–3402, Sep. 2005.
- [30] J. A. Fessler and B. P. Sutton, "Nonuniform fast Fourier transforms using min-max interpolation," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 560–574, 2003.
- [31] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.
- [32] F. Champagnat and J. Idier, "A connection between half-quadratic criteria and EM algorithms," *IEEE Signal Process. Lett.*, vol. 11, no. 9, pp. 709–712, Sep. 2004.
- [33] W. C. Karl, "Regularization in image restoration and reconstruction," in *Handbook of Image and Video Processing*, A. C. Bovik, Ed. New York: Elsevier, 2005, pp. 183–202.
- [34] G. H. Glover, "Simple analytic spiral k-space algorithm," *Magn. Reson. Med.*, vol. 42, no. 2, pp. 412–415, 1999.
- [35] D.-H. Kim, E. Adalsteinsson, and D. M. Spielman, "Simple analytic variable density spiral design," *Magn. Reson. Med.*, vol. 50, no. 1, pp. 214–219, 2003.
- [36] C. Barmet, N. De Zanche, B. J. Wilm, and K. P. Pruessmann, "A transmit/receive system for magnetic field monitoring of *in vivo* MRI," *Magn. Reson. Med.*, vol. 62, no. 1, pp. 269–276, 2009.
- [37] M. Guerquin-Kern, F. I. Karahanoğlu, D. Van De Ville, K. P. Pruessmann, and M. Unser, "Analytical form of Shepp-Logan phantom for parallel MRI," in *Proc. ISBI*, 2010, pp. 261–264.