Sparse recovery in Magnetic Resonance Imaging with a Markov Random Field Prior

Marko Panić, Jan Aelterman, Vladimir Crnojević, and Aleksandra Pižurica

***Abstract*—Recent research in compressed sensing of magnetic resonance imaging (CS-MRI) emphasizes the importance of modelling structured sparsity, either in the acquisition or in the reconstruction stages. Subband coefficients of typical images show certain structural patterns, which can be viewed in terms of fixed groups (like wavelet trees) or statistically (certain configurations are more likely than others). Wavelet tree models have already demonstrated excellent performance in MRI recovery from partial data. However, much less attention has been given in CS-MRI to modelling statistically spatial clustering of subband data, although the potentials of such models have been indicated. In this work, we propose a practical CS-MRI reconstruction algorithm making use of a Markov Random Field prior model for spatial clustering of subband coefficients and an efficient optimization approach based on proximal splitting. The results demonstrate an improved reconstruction performance compared to both standard CS-MRI methods and recent related methods.**

***Index Terms*—MRI, compressed sensing, Markov Random Field, structured sparsity, alternating minimization.**

I. Introduction

MAGNETIC Resonance Imaging (MRI) with its inherently slow data acquisition process [1]–[3] calls for development of smart undersampling schemes [4], [5] and the corresponding reconstruction algorithms. Compressed Sensing (CS) [6], [7], demonstrated potential to improve the acquisition speed in MRI and since the seminal work of Lustig and collaborators [2], [3] on CS-MRI, a number of studies including [8]–[17], have addressed MRI recovery from partial data.

In a CS-MRI setup, the acquired k-space measurements $\mathbf{y} \in \mathbb{C}^M$ of an ideal image $\mathbf{x} \in \mathbb{C}^N$ are

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{1}$$

where $M \ll N$, $\mathbf{n} \in \mathbb{C}^M$ is white Gaussian noise, and $\mathbf{A} \in \mathbb{C}^{M\times N}$ denotes the undersampled Fourier operator [2], [3]. Estimation of $\mathbf{x}$ from measurements $\mathbf{y}$ is an ill-posed linear inverse problem, because the measurement matrix $\mathbf{A}$ is singular and/or very ill conditioned. Since there is no unique solution for the underdetermined system in (1), additional information about $\mathbf{x}$ is typically employed in the form of regularization to stabilize and guide the search towards relevant solutions. MRI images are naturally *compressible* in an appropriate transform domain (such as wavelet or related sparsyfing transform) [2], meaning that their sorted transform coefficients exhibit a power-law decay [18]. Let $\boldsymbol{\theta} = \mathbf{P}\mathbf{x} \in \mathbb{C}^D$ denote transform coefficients with $\mathbf{P} \in \mathbb{C}^{D\times N}$ the sparsyfing transform. A common analysis formulation of the estimator for $\mathbf{x}$, given (1), is [19]:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \phi(\mathbf{P}\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 \leq \epsilon \tag{2}$$

where $\phi : \mathbb{C}^N \mapsto \mathbb{R} \cup \{-\infty, +\infty\}$ is a regularization function and $\epsilon \geq 0$ a parameter related to the noise variance. Choosing ϕ as ℓ_1 norm: $\phi(\mathbf{x}) = \|\mathbf{x}\|_1$ leads to the basis pursuit denoising problem [20]. In CS-MRI, ϕ is typically the ℓ_1 norm, total variation (TV) norm in the image domain ($\mathbf{P} = \mathbf{I}$) or a linear combination of the two [2], [3], [9], [14]. Various reported methods focus on different aspects of this problem, such as improved iterative solvers [9], [11], the use of efficient sparsifying transforms such as shearlets and curvelets [12], [21]–[24] or trained dictionaries [25]–[27], and adaptive sampling schemes [5], [16], [28]. Recent work demonstrates benefits of encoding *structure* of the sparse, information bearing coefficients, either in the acquisition [28] or in the reconstruction [13], [17] stages. Subband coefficients of natural images, including MRI, obey certain structure, which can be viewed in terms of fixed groups (like wavelet trees) or statistically (certain clustering configuration are more likely than others). The wavelet-tree approach has already demonstrated an excellent performance in MRI reconstruction [17].

Much less attention has been devoted to modelling intra-scale coefficient dependencies, such as spatial clustering of subband data in CS-MRI. The so-called Lattice Matching Pursuit (LaMP) algorithm of [29], which models the support configurations with a Markov Random Field (MRF) prior, has demonstrated superior performances in background subtraction. LaMP was derived for images that are sparse in the canonical domain, and hence not directly applicable to most of the MRI images. By analogy with LaMP, some of us introduced earlier an algorithm called Lattice Split Bregman (LaSB) [30]. LaSB combines an MRF prior for the support of important subband coefficients with an augmented Lagrangian approach [31]. Although the presentation in [30] was drafted mainly as a proof of concept, without any elaborate analysis, the results demonstrated great potential for rapid MRI imaging which deserves to be studied more deeply. Motivated by these encouraging results, we develop and evaluate thoroughly an efficient MRF-based CS-MRI method.

The main contributions of this paper are: (1) We develop an efficient method for MRI reconstruction from partial Fourier data making use of a MRF prior for the support configurations of sparse coefficients. To our knowledge, this is the first

Marko Panić and Vladimir Crnojević are with the University of Novi Sad, BioSense Institute, e-mail: *mpanic@uns.ac.rs, crnjvc@gmail.com*

Jan Aelterman, and Aleksandra Pižurica are with Ghent University, Department Telecommunications and Information Processing, e-mail: *Jan.Aelterman@ugent.be, Aleksandra.Pizurica@ugent.be*

elaborate study on CS-MRI with MRF priors, although the potentials of such an approach were earlier demonstrated with a heuristic method LaSB in [30]. Compared to LaSB, the new method employs a different optimization technique, a more general MRF prior and achieves consistently better results; (2) We extend a constrained split augmented Lagrangian shrinkage algorithm (C-SALSA) of [19] with an MRF prior. In particular, we introduce a new regularization step, which admits support configurations favoured by the prior model. The resulting algorithm, coined LaSAL outperforms consistently C-SALSA too; (3) We develop a variant of the proposed method with compound regularization (MRF prior + TV norm), which further improves the reconstruction performance. A thorough evaluation is performed on MRI data sets acquired on Cartesian and radial grids, for which different undersampling strategies are simulated. For the radially acquired k-space data we perform undersampling based on golden ratio profile spacing. MRF-based CS-MRI methods demonstrate a clear improvement compared to alternative methods.

A preliminary version of parts of this work has been reported in a conference paper [32]. There is a substantial difference between this paper and the conference version. The presentation of the overall approach in the conference version is much less elaborate, the best performing algorithm with compound regularization was not presented at all, and the experimental evaluation was rather limited (to two MRI images only, without having access to original k-space data).

The paper is organized as follows: Section II reviews related work and introduces some concepts that we use later on. A general framework for MRF-based CS-MRI is in Section III, together with the concrete prior and conditional models and inference method that we employ. The proposed LaSAL algorithm and its variant with the compound prior are presented in Section IV. The experimental results are given in Section V and Section VI concludes the paper.

II. Related work

A. Reconstruction algorithms

Recent CS-MRI methods typically employ iterative reconstruction algorithms, both greedy and optimization-based. Well-known greedy methods include compressive sampling matching pursuit (CoSaMP) and subspace pursuit (SP) [33], [34], iterative hard thresholding (IHT) [10] and its extensions [35], [36], [15]. Methods employing convex non-smooth regularizers (TV and ℓ_1) typically consider, instead of the original problem in (2), the unconstrained problem [19]:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \phi(\mathbf{Px}) + \frac{\mu}{2}\|\mathbf{Ax}-\mathbf{y}\|_2^2 \tag{3}$$

with $\mu > 0$. Many state-of-the-art methods for solving this problem belong to the iterative soft-thresholding (IST) [37] algorithms and their variants TwIST [38], FISTA [39], and SpaRSA [40]. The solution of (3) is usually defined in terms of the Moreau proximal mapping of ϕ [41]

$$\mathbf{\Psi}_\phi(\mathbf{u};\mu) = \operatorname*{argmin}_{\mathbf{x}\in\mathbb{C}^N} \phi(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{x}-\mathbf{u}\|_2^2 \tag{4}$$

For $\phi(\mathbf{x}) = \|\mathbf{x}\|_1$, this operator is component-wise soft-thresholding $\mathbf{\Psi}_{\ell_1}(\mathbf{u};\mu) = \text{soft}(\mathbf{u}, 1/\mu)$, which replaces each component of $\mathbf{u}$ by $sign(u)\max\{|u| - 1/\mu, 0\}$. For the TV norm: $\|\mathbf{x}\|_{\text{TV}} = \sum_i \sum_j \sqrt{|\mathbf{x}_{i+1,j} - \mathbf{x}_{i,j}|^2 + |\mathbf{x}_{i,j+1} - \mathbf{x}_{i,j}|^2}$, $\mathbf{\Psi}_{\text{TV}}(\mathbf{u};\mu)$ is computed using Chambolle's algorithm [42].

Among the methods that employ extensions of wavelets as sparsifying transforms are [12], [22], [23]. A recent method method pFISTA [24] approximately solves the problem in (3) with $\phi(\mathbf{Px}) = \|\mathbf{Px}\|_1$ where $\mathbf{P}$ is a tight-frame. Other recent approaches employ dictionary learning [25], [27] or patch-based nonlocal operators (PANO) [26].

Although the formulation (3) is equivalent to (2) for the appropriate μ, and usually easier to solve, the formulation (2) has an important advantage: the parameter ϵ in (2) has a clear interpretation in terms of the noise level, while setting the correct μ in (3) is not evident in practice and requires a clever algorithm to adjust it properly. Motivated by this, the authors in [19] proposed an efficient algorithm for solving the constrained problem (2) directly. Their method, named constrained Split Augmented Lagrangian Shrinkage Algorithm (C-SALSA), has proved excellent performance in MRI-reconstruction. Therefore we decide to incorporate our MRF-based prior into this solver, as explained later, in Section IV.

B. Modelling structured sparsity

There are two principal approaches to modelling *structured sparsity* (structure of the sparse image coefficients): (1) in the acquisition stage, through an improved design of the sampling trajectories, and (2) in the recovery phase, through an improved regularization of the inverse problem. The first approach is advocated in [4], [16], [28] where efficient multilevel sampling schemes are constructed, showing a great potential over the standard sampling strategies. We focus on the second approach — modelling signal structure in the recovery phase.

Recent work has shown benefits of using *wavelet-tree* structure in the MRI recovery [13], [17]. This approach models the dependencies among wavelet coefficients on a quadtree structure through an additional group sparsity regularization term. Other related approaches employ Hidden Markov Tree (HMT) models [43], [44], [45]. Less attention has been devoted to modelling within-band (intrascale) dependencies in image recovery from compressive measurements. A representative of this approach is the LaMP (Lattice Matching Pursuit) algorithm [29], where a MRF prior models images that are canonically sparse (in applications such as background subtraction and moving object detection). A related algorithm [30] applied an MRF prior to subband data in CS-MRI recovery. Motivated by the encouraging results of [30], we build further on this approach and present a solid motivation, elaborate analysis and thorough evaluation, while previously only a proof of concept was given. Moreover using a different underlying optimization method and improved MRF modelling, we improve the performance over [30], and we also demonstrate, for the first time, potential benefits over the competing tree-structured approach.

III. MRF-based structure sparsity model

Let $\mathbf{P}^{D\times N}$ denote some sparsifying transform which yields coefficients $\boldsymbol{\theta} = \mathbf{Px} = \{\theta_1, ..., \theta_D\}$. The coefficient θ_i is

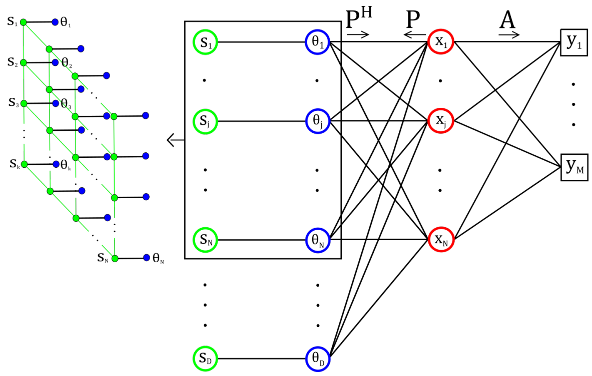

Fig. 1: A graphical representation of variables, operators and their connections in our model. **Left**: Hidden labels s_i, attached to each subband coefficient θ_i. Links among neighboring s_i indicate their statistical dependencies, encoded in a MRF. **Right**: A graphical model showing all the involved variables, measurements and operators in our problem.

significant if its magnitude is above a certain threshold. We assign a hidden label $s_i \in \{0, 1\}$ to θ_i to mark its significance: $s_i = 1$ if θ_i is significant and $s_i = 0$ otherwise. A particular configuration $\mathbf{s} = \{s_1, ..., s_D\}$ is assumed to be a realization of a Markov Random Field $\mathbf{S} = \{S_1, ..., S_D\}$. Fig. 1 illustrates this whole setup. Note that all measurements gathered in $\mathbf{y}$ are obtained as linear combinations of all N pixel intensities in $\mathbf{x}$ through the operator $\mathbf{A}$. The sparse coefficients $\boldsymbol{\theta}$ result from applying the analysis operator $\mathbf{P}$ to $\mathbf{x}$. Therefore, each coefficient θ_i is a linear combination of all pixel values, via $\mathbf{P}$. Conversely, each pixel value x_i is obtained as a linear combination of all coefficients θ_i through the synthesis operator $\mathbf{P}^H$.

A. Recovery problem with structured sparsity

Let us now instantiate a general recovery problem as (2), by replacing the arbitrary regularizer ϕ by our structured sparsity model. We use similar notation to [45]. Given the index set $\mathcal{N} = \{1, 2, 3, ..., D\}$, let $supp(\boldsymbol{\theta}) = \{i \in \mathcal{N} : \theta_i \neq 0\}$ denote the *support* of $\boldsymbol{\theta}$. Further on, for $\mathcal{S} \subseteq \mathcal{N}$, $\boldsymbol{\theta}[\mathcal{S}]$ denotes the elements of $\boldsymbol{\theta}$ indexed by $\mathcal{S}$, and $\bar{\mathcal{S}}$ is the complement of $\mathcal{S}$ with respect to $\mathcal{N}$. Denote the index set corresponding to the support $\mathbf{s}$ by $\Omega_{\mathbf{s}} = \{i \in \mathcal{N} : s_i = 1\}$ and define a model for $\boldsymbol{\theta}$ that conforms to the particular support configuration $\mathbf{s}$ as

$$\mathcal{M}_{\mathbf{s}} = \{\boldsymbol{\theta} \in \mathbb{C}^D : supp(\boldsymbol{\theta}) = \Omega_{\mathbf{s}}\}. \tag{5}$$

The objective of our approach is

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \quad \text{subject to} \quad \mathbf{P}\mathbf{x} \in \mathcal{M}_{\hat{\mathbf{s}}} \tag{6}$$

where $\hat{\mathbf{s}}$ is the estimate of the most likely spatial support of $\boldsymbol{\theta} = \mathbf{P}\mathbf{x}$. The constraint $\mathbf{P}\mathbf{x} \in \mathcal{M}_{\hat{\mathbf{s}}}$ can be equivalently replaced by $supp(\mathbf{P}\mathbf{x}) = \Omega_{\hat{\mathbf{s}}}$. In solving this problem, we shall involve a simpler one

$$\min_{\boldsymbol{\gamma} \in \mathbb{C}^D} \|\boldsymbol{\gamma} - \boldsymbol{\theta}\|_2^2 \quad \text{subject to} \quad \boldsymbol{\gamma} \in \mathcal{M}_{\hat{\mathbf{s}}} \tag{7}$$

for which the solution is $\hat{\boldsymbol{\gamma}}_H[\Omega_{\hat{\mathbf{s}}}] = \boldsymbol{\theta}[\Omega_{\hat{\mathbf{s}}}]$ and $\hat{\boldsymbol{\gamma}}_H[\bar{\Omega}_{\hat{\mathbf{s}}}] = 0$. Since $s_i \in \{0, 1\}$, this solution can be written as the *Hadamard product* $\hat{\boldsymbol{\gamma}}_H[\Omega_{\hat{\mathbf{s}}}] = \boldsymbol{\theta} \circ \hat{\mathbf{s}}$.

We search for the most likely support $\hat{\mathbf{s}}$ by applying the maximum a posteriori probability (MAP) criterion:

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmax}}\, P_{\mathbf{S}|\boldsymbol{\theta}}(\mathbf{s} \mid \boldsymbol{\theta}) = \underset{\mathbf{s}}{\operatorname{argmax}}\, p_{\boldsymbol{\theta}|\mathbf{S}}(\boldsymbol{\theta} \mid \mathbf{s}) P_{\mathbf{S}}(\mathbf{s}) \tag{8}$$

In practice, we shall re-estimate $\hat{\mathbf{s}}$ in each iteration of the complete recovery algorithm, starting from the current (temporary) estimate of the coefficient vector $\boldsymbol{\theta}$.

B. MRF prior

The global probability $P_{\mathbf{S}}(\mathbf{s})$ of a MRF is a Gibbs distribution [46], [47]

$$P_{\mathbf{S}}(\mathbf{s}) = \frac{1}{Z} e^{-H(\mathbf{s})/T} \tag{9}$$

where the energy $H(\mathbf{s})$ is a sum of clique potentials over all possible cliques: $H(\mathbf{s}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{s})$. The normalizing constant $Z = \sum_{\mathbf{s} \in \mathcal{L}} e^{-H(\mathbf{s})/T}$ is called the partition function and the temperature T controls the peaking in the probability density [46]. We use the Ising model as in [30], where

$$H(\mathbf{s}) = \sum_i V_1(s_i) + \sum_{\langle i,j \rangle \in \mathcal{C}} V_2(s_i, s_j) \tag{10}$$

with the single and pairwise potentials defined as

$$V_1(s) = \begin{cases} \alpha & s = 0 \\ -\alpha & s = 1 \end{cases}, \quad V_2(s,t) = \begin{cases} -\beta & s = t \\ \beta & s \neq t \end{cases} \tag{11}$$

Unlike in [30], we allow different a priori probabilities $\alpha \neq 0$, so that we can enforce the sparsity of the supports. The strength of the spatial clustering is controlled by the parameter $\beta > 0$.

C. Conditional model

We adopt the conditional model $p_{\boldsymbol{\Theta}|\mathbf{S}}(\boldsymbol{\theta}|\mathbf{s})$ of [30], [47]. With the common conditional independence assumption, we have $p_{\boldsymbol{\Theta}|\mathbf{S}}(\boldsymbol{\theta}|\mathbf{s}) = \prod_i p_{\Theta_i|S_i}(\theta_i|s_i)$. The observed coefficients are typically noisy versions of the ideal ones: $\theta = u + n$, where n denotes the noise component. We select the prior $p_U(u)$ as the generalized Laplacian and we estimate its parameters from the noisy coefficient histogram, knowing the noise standard deviation σ [47], [48]. In practice, σ is reliably estimated from the empty area on the borders of the MR image and rescaled appropriately in each subband. Let T_h denote the significance threshold for the ideal noise-free coefficients (u is significant if $|u| \geq T_h$). We relate this threshold to the noise level, but in a conservative manner, such that T_h is only a fraction of σ (in practice 10%). The conditional densities $p_{U|S}(u|0)$ and $p_{U|S}(u|1)$ are then obtained by rescaling the central part ($|u| < T_h$) and the tails ($|u| \geq T_h$) of $p_U(u)$, respectively, so that they both integrate to 1. The conditional densities of the noisy coefficients $p_{\Theta|S}(\theta|s)$ are obtained from the corresponding $p_{U|S}(u|s)$. For the additive noise model $\theta = u + n$ with $n \sim N(0, \sigma)$, $p_{\Theta|S}(\theta|s)$ is simply the convolution of $p_{U|S}(u|s)$ with $N(0, \sigma)$. Fig. 2 illustrates the adopted conditional model and the above described procedure.

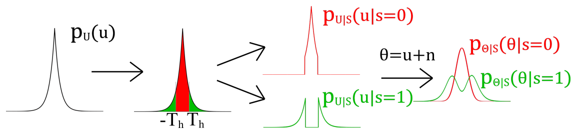

Fig. 2: The adopted conditional model from [30], [47]. Note that $p_U(u)$ is obtained from the noisy histogram. T_h is the only parameter.

D. Inference algorithm

Various inference algorithms can be employed to find the MAP estimate in (8), e.g., Iterative Conditional Modes (ICM) [49], Graph Cuts [50], loopy belief propabation (LBP) [51], and Markov Chain Monte Carlo (MCMC) samplers, such as Metropolis and Gibbs sampler [46]. We used the Metropolis sampler due to its flexibility and efficiency in this application. The Metropolis sampler starts from some initial configuration and in each step it switches a randomly chosen label s_i in the current mask $\mathbf{s}$ to produce the so-called "candidate" mask $\mathbf{s}^C$. The candidate gets accepted or not based on the change in the posterior probability $P_{\mathbf{S}|\mathbf{\Theta}}(\mathbf{s}^C|\boldsymbol{\theta})/P_{\mathbf{S}|\mathbf{\Theta}}(\mathbf{s}|\boldsymbol{\theta})$, which effectively reduces to

$$r = \left(\frac{p_{\theta_i|S_i}(\theta_i \mid 1)}{p_{\theta_i|S_i}(\theta_i \mid 0)}\right)^{\lambda} \exp\left\{2\alpha + 2\beta \sum_{j\in\mathcal{N}_i}(2s_j - 1)\right\} \quad (12)$$

when $s_i^C = 1$ and to $1/r$ when $s_i^C = 0$. Practically, the change is accepted if r exceeds a randomly generated number drawn from a uniform distribution on $[0,1]$. Parameter $\lambda > 0$ effectively simulates sampling at different temperatures; for details see [47]. This inference algorithm is in fact a step of the simulated annealing algorithm from [52] for a particular temperature — one could apply simulated annealing by changing gradually λ although we didn't do it in our experiments.

IV. CS-MRI ALGORITHM WITH MRF PRIORS

We now incorporate the spatial support estimation into practical CS-MRI recovery algorithms. The algorithms that we develop in this Section are inspired by and can be seen as extensions of the C-SALSA algorithm of [19].

A. LaSAL

Our optimization problem from (6) is equivalent to (2) under suitably defined regularization function ϕ. We follow the same steps for solving (2) as in [19], and we incorporate the particular ϕ that follows from our structured sparsity model described in the previous Section. To this end, let $E(\epsilon, \mathbf{A}, \mathbf{y}) = \{\mathbf{x} \in \mathbb{C}^N : \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \le \epsilon\}$ denote the feasible set for $\mathbf{x}$. By introducing an indicator function

$$\iota_{\mathcal{Q}}(\mathbf{q}) = \begin{cases} 0, & \mathbf{q} \in \mathcal{Q} \\ +\infty, & \text{otherwise} \end{cases} \quad (13)$$

the problem in (2) can be written as follows:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \phi(\mathbf{P}\mathbf{x}) + \boldsymbol{\iota}_{E(\epsilon,\mathbf{I},\mathbf{y})}(\mathbf{A}\mathbf{x}). \quad (14)$$

It has been shown in [19] that this problem is efficiently solved by a special type of the alternating direction method of multipliers (ADMM). The key step is variable splitting, which allows solving the composite problem as a sequence of minimizations over the separate components. In particular, for the problem in (14), two splitting variables are introduced $\mathbf{w} = \mathbf{x}$ and $\mathbf{v} = \mathbf{A}\mathbf{x}$, to split the original problem into separate mininimizations over each of the two terms. Together with a "binding" term that connects these two separate minimizations, we obtain the following three sub-problems:

$$\begin{aligned} \mathbf{x}^{\{k+1\}} &= \operatorname*{argmin}_{\mathbf{x}\in\mathbb{C}^N} \left\{ \|\mathbf{A}\mathbf{x} - \mathbf{u}'\|_2^2 + \mu\|\mathbf{x} - \mathbf{u}''\|_2^2 \right\} \\ \mathbf{v}^{\{k+1\}} &= \operatorname*{argmin}_{\mathbf{v}\in\mathbb{C}^M} \left\{ \frac{\boldsymbol{\iota}_{E(\epsilon,\mathbf{I},\mathbf{y})}(\mathbf{v})}{\mu} + \frac{1}{2}\|\mathbf{v}' - \mathbf{v}\|_2^2 \right\} \\ \mathbf{w}^{\{k+1\}} &= \operatorname*{argmin}_{\mathbf{w}\in\mathbb{C}^N} \left\{ \phi(\mathbf{P}\mathbf{w}) + \frac{\mu}{2}\|\mathbf{w}' - \mathbf{w}\|_2^2 \right\} \end{aligned} \quad (15)$$

where $\mathbf{v}' = \mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{b}^{\{k\}}$, $\mathbf{w}' = \mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}}$, $\mathbf{u}' = \mathbf{v}^{\{k\}} + \mathbf{b}^{\{k\}}$, $\mathbf{u}'' = \mathbf{w}^{\{k\}} + \mathbf{c}^{\{k\}}$, and $\mathbf{b}, \mathbf{c}$ are auxiliary variables.

The first sub-problem $\mathbf{x}^{\{k+1\}}$ is solved by the Gauss-Seidel method leading to a simple update equation. The second sub-problem $\mathbf{v}^{\{k+1\}}$ obviously does not depend on μ (because the indicator function defined in (13) takes only the values 0 or $+\infty$) and is simply the orthogonal projection of $\mathbf{v}$ on the closed ϵ-radius ball centered at $\mathbf{y}$ [19]:

$$\boldsymbol{\Psi}_{\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}}(\mathbf{v}) = \mathbf{y} + \begin{cases} \epsilon\frac{\mathbf{v}-\mathbf{y}}{\|\mathbf{v}-\mathbf{y}\|_2}, & \text{if } \|\mathbf{v}-\mathbf{y}\|_2 > \epsilon \\ \mathbf{v}-\mathbf{y}, & \text{if } \|\mathbf{v}-\mathbf{y}\|_2 \le \epsilon \end{cases} \quad (16)$$

The third sub-problem $\mathbf{w}^{\{k+1\}}$ has been typically solved by defining ϕ as the ℓ_1-norm. We define instead the regularization function $\phi(\boldsymbol{\theta})$ as a δ-loss function, prohibiting all realizations $\boldsymbol{\theta}$ that do not conform to the estimated support $\hat{\mathbf{s}}$. With the model $\mathcal{M}_{\hat{\mathbf{s}}}$ from (5), we define formally

$$\phi(\boldsymbol{\theta}) = \begin{cases} 0, & \text{if } \boldsymbol{\theta} \in \mathcal{M}_{\hat{\mathbf{s}}} \\ \infty, & \text{if } \boldsymbol{\theta} \notin \mathcal{M}_{\hat{\mathbf{s}}} \end{cases} \quad (17)$$

Substituting $\boldsymbol{\theta} = \mathbf{P}\mathbf{w}$, the third sub-problem in (15), following the transformation procedure given in [24], becomes

$$\boldsymbol{\theta}^{\{k+1\}} = \operatorname*{argmin}_{\boldsymbol{\theta}\in Range(\mathbf{P})} \left\{ \phi(\boldsymbol{\theta}) + \frac{\mu}{2}\|\boldsymbol{\theta}' - \boldsymbol{\theta}\|_2^2 \right\} \quad (18)$$

This problem has the same solution as its equivalent constrained formulation from (7), and thus $\boldsymbol{\theta}^{\{k+1\}} = \boldsymbol{\theta}' \circ \hat{\mathbf{s}}$ where

$$[\boldsymbol{\theta}' \circ \hat{\mathbf{s}}]_i = \begin{cases} \theta_i, & \text{if } \hat{s}_i = 1 \\ 0, & \text{if } \hat{s}_i = 0 \end{cases} \quad (19)$$

This completes the specification of our algorithm, named by analogy with the related methods as LaSAL, from Lattice Split Augmented Lagrangian. Its pseudo-code is listed in Algorithm 1. The step $\hat{\mathbf{s}} \leftarrow$ *MAP-support*$\{\boldsymbol{\theta}'\}$ denotes the support estimation using the MAP criterion in (8). The parameter $0 < \mu \le 1$, which controls the level of regularization, can be safely set to 1 as it was also done in [19], without a significant performance loss. We still decided to keep μ as a parameter in the algorithm, because we observed that allowing values of $\mu \le 1$ can yield a slightly higher peak signal to noise ratio (PSNR) in the reconstructions (up to 0.5 dB). Furthermore,

Algorithm 1 LaSAL

Input: $k = 0, \mu > 0, \mathbf{v}^{\{0\}}, \mathbf{w}^{\{0\}}, \mathbf{b}^{\{0\}}, \mathbf{c}^{\{0\}}$

1: **repeat**
2: $\quad \mathbf{r}^{\{k\}} = \mu(\mathbf{w}^{\{k\}} + \mathbf{c}^{\{k\}}) + \mathbf{A}^H(\mathbf{v}^{\{k\}} + \mathbf{b}^{\{k\}})$
3: $\quad \mathbf{x}^{\{k+1\}} = (\mu\mathbf{I} + \mathbf{A}^H\mathbf{A})^{-1}\mathbf{r}^{\{k\}}$
4: $\quad \mathbf{v}^{\{k+1\}} = \mathbf{\Psi}_{\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}}(\mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{b}^{\{k\}})$
5: $\quad \boldsymbol{\theta}' = \mathbf{P}(\mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}})$
6: $\quad \hat{\mathrm{s}} \leftarrow$ *MAP-support*$\{\boldsymbol{\theta}'\}$
7: $\quad \mathbf{w}^{\{k+1\}} = \mathbf{P}^H(\boldsymbol{\theta}' \circ \hat{\mathrm{s}})$
8: $\quad \mathbf{b}^{\{k+1\}} = \mathbf{b}^{\{k\}} - (\mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{v}^{\{k+1\}})$
9: $\quad \mathbf{c}^{\{k+1\}} = \mathbf{c}^{\{k\}} - (\mathbf{x}^{\{k+1\}} - \mathbf{w}^{\{k+1\}})$
10: $\quad k = k + 1$
11: **until** some stopping criterion is satisfied

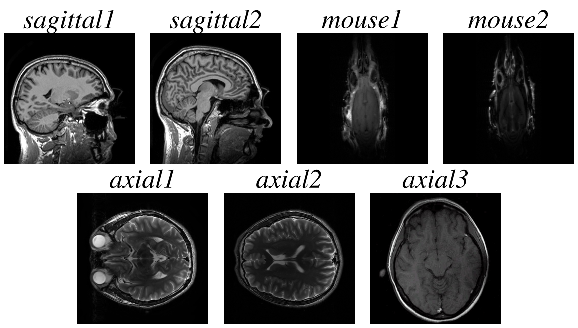

Fig. 3: Test images *sagittal1* and *sagittal2* are two slices from data set 1 comprising 248 images. The bottom three images are from [24], [26] and [25] resp. All images are 256×256, except *axial3*, which is 512×512.

observe that the update of the auxiliary variable $\mathbf{c}$ (line 9) is performed in the image-domain (while the equivalent step of the analysis-formulation of the related C-SALSA is applied in the transform domain, and with $\mathbf{P}^H$ applied to the first sum in $\mathbf{r}^{\{k\}}$). This is because we use $\mathbf{w} = \mathbf{x}$ in variable splitting instead of $\mathbf{w} = \mathbf{P}\mathbf{x}$ which leads to this type of update. Finally, as the stopping criterion we use in practice a fixed number of iterations (typically 50), because in all the simulations the differences in the resulting reconstruction error become practically negligible after this many iterations.

B. LaSAL2

We extend now the objective function in (14) with another regularization term: TV norm $\|\mathbf{x}\|_{\mathrm{TV}}$. The resulting objective function is an instance of the general form

$$\min_{\mathbf{x}\in\mathbb{C}^N} \sum_{j=1}^{J} g_j(\mathbf{H}^{(j)}\mathbf{x}) \tag{20}$$

with $J = 3$, $\mathbf{H}^{(1)} = \mathbf{P}$, $\mathbf{H}^{(2)} = \mathbf{I}$, $\mathbf{H}^{(3)} = \mathbf{A}$, $g_1(\mathbf{u}) = \phi(\mathbf{u})$, $g_2(\mathbf{u}) = \|\mathbf{u}\|_{\mathrm{TV}}$ and $g_3(\mathbf{u}) = \iota_{E(\epsilon,\mathbf{I},\mathbf{y})}(\mathbf{u})$. A detailed explanation and a compact pseudo-code for solving (20) is given in [19]. Here, we simply extend LaSAL from Algorithm 1 with an additional step that concerns with the TV regularization. This also requires introducing an additional auxiliary variable ($\mathbf{d}^{\{k\}}$, next to $\mathbf{b}^{\{k\}}$ and $\mathbf{c}^{\{k\}}$ in Algorithm 1). By applying

Algorithm 2 LaSAL2

Input: $k = 0, \mu_1, \mu_2 > 0, \mathbf{v}^{\{0\}}, \mathbf{w}^{\{0\}}, \mathbf{z}^{\{0\}}, \mathbf{b}^{\{0\}}, \mathbf{c}^{\{0\}}, \mathbf{d}^{\{0\}}$,

1: **repeat**
2: $\quad \mathbf{r}^{\{k\}} = \mu_1(\mathbf{z}^{\{k\}} + \mathbf{c}^{\{k\}}) + \mathbf{A}^H(\mathbf{v}^{\{k\}} + \mathbf{b}^{\{k\}})$
3: $\quad \mathbf{x}^{\{k+1\}} = (\mu_1\mathbf{I} + \mathbf{A}^H\mathbf{A})^{-1}\mathbf{r}^{\{k\}}$
4: $\quad \mathbf{v}^{\{k+1\}} = \mathbf{\Psi}_{\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}}(\mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{b}^{\{k\}})$
5: $\quad \mathbf{z}' = \frac{1}{(\mu_1+\mu_2)}\big(\mu_1(\mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}}) + \mu_2(\mathbf{w}^{\{k\}} + \mathbf{d}^{\{k\}})\big)$
6: $\quad \mathbf{z}^{\{k+1\}} = \mathbf{\Psi}_{\mathrm{TV}}(\mathbf{z}'; \mu_1 + \mu_2)$
7: $\quad \boldsymbol{\theta}' = \mathbf{P}(\mathbf{z}^{\{k+1\}} - \mathbf{d}^{\{k\}})$
8: $\quad \hat{\mathrm{s}} \leftarrow$ *MAP-support*$\{\boldsymbol{\theta}'\}$
9: $\quad \mathbf{w}^{\{k+1\}} = \mathbf{P}^H(\boldsymbol{\theta}' \circ \hat{\mathrm{s}})$
10: $\quad \mathbf{b}^{\{k+1\}} = \mathbf{b}^{\{k\}} - (\mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{v}^{\{k+1\}})$
11: $\quad \mathbf{d}^{\{k+1\}} = \mathbf{d}^{\{k\}} - (\mathbf{z}^{\{k+1\}} - \mathbf{w}^{\{k+1\}})$
12: $\quad \mathbf{c}^{\{k+1\}} = \mathbf{c}^{\{k\}} - (\mathbf{x}^{\{k+1\}} - \mathbf{z}^{\{k+1\}})$
13: $\quad k = k + 1$
14: **until** some stopping criterion is satisfied

variable splitting [53], the minimization sub-problem corresponding to the TV regularization can be written as

$$\begin{aligned} \mathbf{z}^{\{k+1\}} &= \operatorname*{argmin}_{\mathbf{g}\in\mathbb{C}^N} \left\{ \|\mathbf{z}\|_{\mathrm{TV}} + \frac{\mu_1 + \mu_2}{2}\|\mathbf{z}' - \mathbf{z}\|_2^2 \right\} \\ &= \mathbf{\Psi}_{\mathrm{TV}}(\mathbf{z}'; \mu_1 + \mu_2) \end{aligned} \tag{21}$$

where $\mathbf{z}' = \frac{1}{(\mu_1+\mu_2)}\big(\mu_1(\mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}}) + \mu_2(\mathbf{w}^{\{k\}} + \mathbf{d}^{\{k\}})\big)$ is a linear combination of the current solution $\mathbf{x}^{\{k+1\}}$ and the regularized solution $\mathbf{w}^{\{k\}}$ from the previous iteration, with parameters where $\mu_1, \mu_2 > 0$. For computing $\mathbf{\Psi}_{\mathrm{TV}}(\mathbf{u}; \mu)$ we used 5 iterations of Chambolle's algorithm [42] (more iterations only increased the computational cost with negligible improvement in PNSR). A pseudo-code of the resulting method that we named LaSAL2 is given in Algorithm 2. The source codes of both LaSAL and LaSAL2 algorithms are available at https://telin.ugent.be/~sanja/MRIreconstruction/LaSAL.

V. Experiments

A. Data sets and reference methods

We evaluate the proposed method on different MRI images, both using simulations starting from high-resolution MRI images, shown in Fig. 3, and using real data acquired in k-space. The first data set comprises 248 T1 MRI *brain slices* acquired on a Cartesian grid at Ghent University hospital (only two slices are shown in Fig. 3). The second set consists of two images of *mouse brain* with different modalities (T1 and T2) acquired on a Cartesian grid at Bio-Imaging Lab at the University of Antwerp. The remaining three images in Fig. 3 are from [24] (*axial1*), [26] (*axial2*) and [25] (*axial3*). All the test images have resolution 256×256, except *axial3*, which is 512×512. We simulate different sub-sampling trajectories: random lines, radial, Fibonacci spiral [54] and random sampling, illustrated in Fig. 4. All trajectories are defined as binary matrices on the Cartesian grid, which act as masks for selecting the corresponding Fourier coefficients. The data set that was acquired directly in k-space is an MRI scan of a pomelo (see Section V-E), acquired by radial acquisition, at the Bio-Imaging Lab in Antwerp.

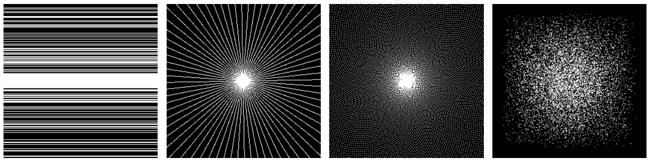

Fig. 4: Examples of sampling trajectories used in the experiments: random lines, radial, Fibonacci spiral and random.

As reference methods, we use C-SALSA [19], the augmented Lagrangian method (Split-Bregman) SB [12] and LaSB [30], all implemented with the same shearlet transform. We also provide comparison with WaTMRI [13], [17], FCSA [14] and FCSANL [55] using the original implementations of the authors (http://ranger.uta.edu/~huang/index.html), and with pFISTA [24], dictionary learning approach DLMRI [25] and a patch-based method PANO [26], on images for which these methods were optimized. As an evaluation criterion we use the peak signal to noise ratio (PSNR) computed on the magnitude image. For the data acquired directly in k-space (no reference image available), we compute the structural similarity index (SSIM) [56] of the reconstructions from partial data relative to the reconstruction from all the available measurements.

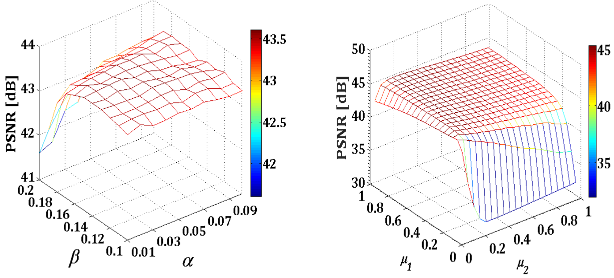

Fig. 5: Examples illustrating grid search results for the MRF parameters (α, β) (left) and regularization parameters (μ_1,μ_2) (right).

B. Parameter selection

As a sparsyfing transform, we use a nondecimated shearlet transform with 3 scales and by default 16, 8, and 4 orientations per scale, respectively, implemented as in [57].

The parameters (α, β) of the MRF model are optimized by grid search. Although the optimal values may slightly differ depending on the particular image, sampling rate and sampling trajectory, we observed a stable performance in a relatively wide range of the parameter values, as illustrated in the example from Fig. 5 (left). This diagram corresponds to LaSAL, applied on the test image *sagittal1*, with 48% of samples on a radial trajectory. Similar diagrams were obtained with other images and other sampling trajectories. We concluded that the same parameter values can be safely used for a wide range of sampling rates and for different images. The recommended values are α=0.01; β=0.16, with $\lambda = 0.2$ in (12). All the image reconstruction results reported in this paper were produced with these values. We did observe that somewhat more stable performance is in some cases reached with slightly different parameter values and at the price of slightly reduced PSNR, but this differences are not so significant in our experience.

We also optimize the parameter μ of LaSAL and (μ_1, μ_2) of LaSAL2 by grid search. It is important to note that all these parameters can be simply set to 1, without sacrificing significantly the reconstruction performance. This is evident from the grid search diagram in Fig. 5 (right) and agrees also with the general theory in [19]. Still, we observed that somewhat better reconstruction performance may be reached in practice by allowing other values of these parameters, so we opted to keep the possibility for their experimental optimization. In particular, we recommend $\mu = 0.04$ for LaSAL, and $\mu_1 = 0.11$, $\mu_2 = 0.01$ for LaSAL2 which are used for all experiments in the paper.

C. Benefit from the MRF model

We first explore how the incorporated MRF-based spatial context model influences the reconstruction performances. This can be directly observed by excluding the MRF-modelling part of LaSAL (lines 5–7 in Algorithm 1) and replacing it simply by soft-thresholding in the shearlet domain $\mathbf{w}^{\{k+1\}} = \mathbf{P}^H(\mathbf{\Psi}_{\ell_1}(\mathbf{P}(\mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}}; \mu)))$, which reduces our method to the corresponding version of C-SALSA [19].

The results in Fig. 6 demonstrate a clear improvement due to the MRF model. LaSB and SB share the same optimization algorithm, while LaSB is enriched with an MRF model. Similarly, C-SALSA and LaSAL share the same optimization method, extended with an MRF prior in LaSAL. We observe that in the same way as LaSB improves over SB, our new algorithm LaSAL improves over C-SALSA consistently for all sampling rates. Moreover, LaSAL yields a consistent improvement over LaSB, except at very low sampling rates. With the spiral trajectory, for sampling rates above 0.3, this improvement in PSNR is more than 1 dB and above 1.7 dB for the sampling rate around 0.5. Similar behaviour, with slightly smaller differences is observed in the case of radial trajectory.

Fig. 6 (bottom row) also shows the advantage of the compound prior: LaSAL2 indeed improves over LaSAL. The results also demonstrate improvement over the reference methods SB and LaSB implemented with compound priors, denoted for consistency as SB2 and LaSB2. The improvement of LaSAL2 over these methods is consistent at all sampling rates and for both sampling trajectories. The difference in PSNR relative to both LaSAL and LaSB2 ranges from 1 dB to more than 2 dB, while the improvement over SB2 is 3 to 5 dB.

D. Comparison with other methods

The reference methods FCSA [14], FCSANL [55] and WaTMRI [17] employ a compound regularization: TV and ℓ_1 (FCSA and WaTMRI) or non-local TV and ℓ_1 (FCSANL). WaTMRI employs next to it a tree-structured sparsity model. We adopt the experimental setup of [14], [17], [55] using random sampling matrices with variable density[1]. Seven sampling rates (14%, 20%, 25%, 32%, 38%, 42% and 50%) are used,

[1] http://ranger.uta.edu/~huang/index.html

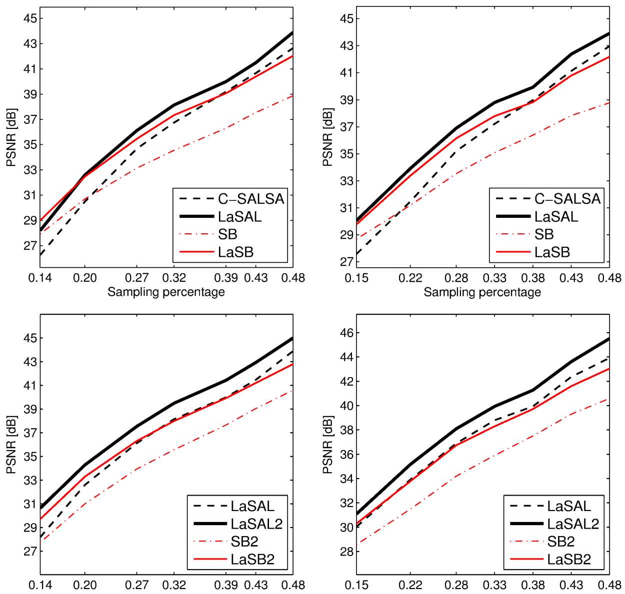

Fig. 6: Reconstruction results for *sagittal1* with different sampling rates using radial (left) and Fibonacci spiral (right) trajectories.

and for each of them ten sampling matrices are randomly generated and the average PSNR over the ten corresponding reconstructions is recorded.

Fig. 7 (top left) shows the result for *sagittal1* from Fig. 3. Obviously, the proposed LaSAL and LaSAL2 algorithms yield consistent improvement over all three reference methods FCSA, FCSANL and WaTMRI at all sampling rates. This improvement is in the range of 1.4 – 3 dB for LaSAL and in the range of 2.3 – 4.1 dB for LaSAL2. Similar conclusions hold for *sagittal2* (Fig. 7, top right): the improvement for LaSAL is now in the range of 0.5 – 2.6 dB and for LaSAL2 the same as on *sagittal1*. Two other diagrams in Fig. 7 show the PSNR results for the T1 and T2 *mouse brain* images. LaSAL2 yields again superior performance compared to all other tested methods. Among the three reference methods, FCSANL is now the best performing. LaSAL2 improves over this method at all sampling rates in the range of 0.6 dB to 2.7 dB for *mouse1* and in the range 2.2 – 3 dB for *mouse2*.

We also perform evaluation on the complete dataset of 248 MRI *brain slices*. Fig. 9 shows the mean PSNR per iteration across all the 248 images for different algorithms. LaSAL2 yields considerably higher PSNR than the reference methods: the improvement is more than 3.5 dB. Even LaSAL, which employs no TV regularization, improves over FCSANL and WaTMRI for about 2.4 dB, and reaches its highest PSNR in fewer iterations than the reference methods. In the same figure, we show the resulting distribution of the PSNR values per iteration. It can be seen that after 5 iterations LaSAL reaches a huge improvement in PSNR over all the reference methods, while LaSAL2 outperforms LaSAL after 30 iterations.

This huge improvement in PSNR comes at a price of an increased processing time. The computation times reported below were obtained on Intel© Core$^{\text{TM}}$ i7 processor (2.4 GHz, 8GB RAM). For LaSAL and LaSAL2, with a non-optimized Matlab implementation and (16, 8, 4) shearlet bands per scale, the processing time for a 256×256 image is 1.50 s per iteration, out of which 1.41 s goes on the support estimation, resulting in about 75 s for 50 iterations. For comparison, the fastest reference methods in 50 iterations require: FCSANL – 0.7 s, FCSA – 0.6 s, WaTMRI – 0.8 s, and C-SALSA – 10.9 s. There is much room for improving the computation time of our method by improving efficiency of the support configuration, e.g. by considering alternatives to Metropolis sampling, such as iterated conditional modes (ICM) [49] or belief propagation.

TABLE I: Comparison with PANO [26] and DLMRI [25]

axial2, random lines 40%			***axial3***, radial 14%		
Method	PSNR [dB]	Time [s]	Method	PSNR [dB]	Time [s]
LaSAL	**43**	111	LaSAL	36.6	507
LaSAL2	**45**	105	LaSAL2	**39.4**	**231**
PANO	41.3	**74**	DLMRI	37.5	763

For comparisons with pFISTA [24], we use their data – *axial1* in Fig. 3 and the original code provided by the authors. Fig. 10 shows the results for random and radial sampling trajectories with the sampling rate of 30%. Three variants of pFISTA from [24] are tested, using contourlets, shift-invariant discrete wavelet (SIDWT) and patch based directional wavelet (PBDW). We now used for LaSAL and LaSAL2 fewer shearlet bands (8, 4 and 2 per scale), resulting in comparable or smaller processing times with pFISTA, PANO and DLMRI. For the radial trajectory, the best performing variant of pFISTA gives a similar (slightly better) results than LaSAL, but LaSAL2 clearly yields the higher PSNR value. With random sampling, both LaSAL and LaSAL2 outperform clearly all the variants of pFISTA, and LaSAL2 is again the best performing method. Moreover, in all the cases LaSAL reached the maximum PSNR faster than pFISTA (see the caption of Fig. 3).

For comparisons with PANO [26] and DLMRI from [25], we used the images from the corresponding papers (*axial2* and *axial3* from Fig. 3, resp.), the sampling trajectories that were used in the corresponding papers as indicated in Table. I, and the original publicly available codes. The resulting PSNR and processing times are listed in Table. I. LaSAL2 outperforms both PANO and DLMRI method on their respective test data. LaSAL2 was somewhat slower than PANO and significantly (more than three times) faster than DLMRI.

We also compared our approach to [22] and [23] on the test data from the original papers. Compared to the reported results in [22], LaSAL yields an improvement of nearly 2 dB, taking approximately the same time, and LaSAL2 an improvement of more than 3 dB. The source code of [23] was unavailable, but compared to the reported results from this work, LaSAL yields similar or slightly better signal to noise ratio and LaSAL2 yields an improvement of 2.5 dB.

E. Experiments on radially sampled data

Here we perform experiments on a data set acquired with radial sampling in the k-space — an MRI scan of a *pomelo*, consisting of 1608 radial lines, each with 1024 samples. We form under-sampled versions by leaving out some of the radial

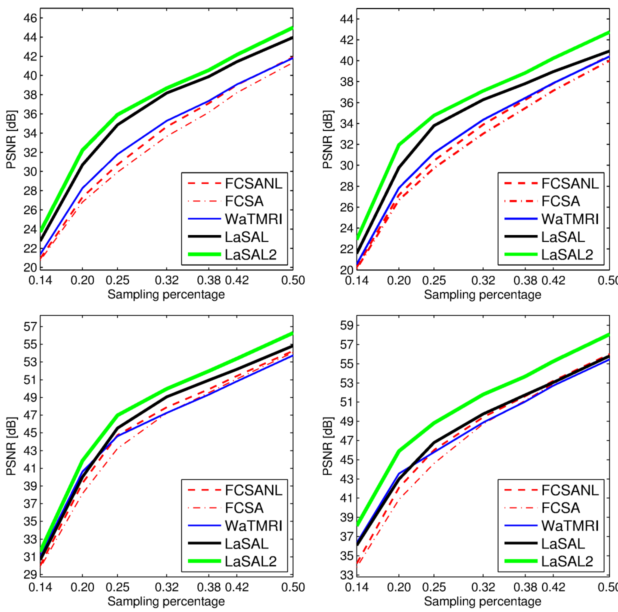

Fig. 7: Results with random sampling and different sampling rates on images *sagittal1* and *sagittal2* (top); *mouse1* and *mouse2* (bottom).

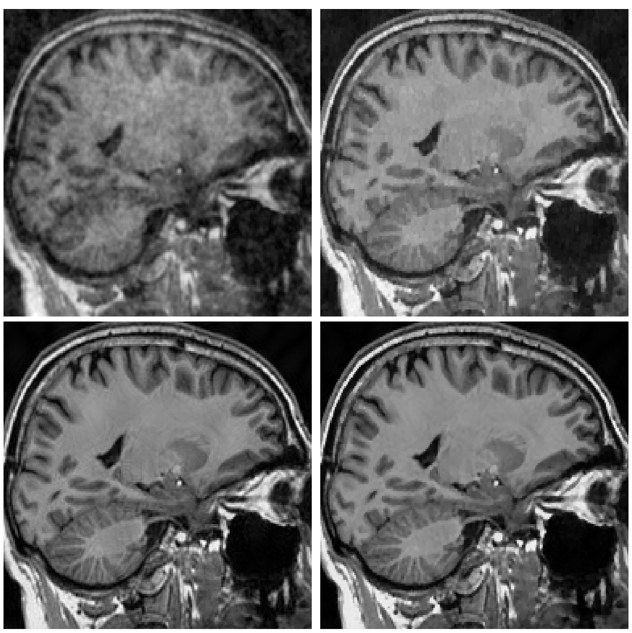

Fig. 8: Reconstructed *sagittal1* image from 20% of random measurements. **First row**: zero-fill (19.87 dB) and WaTMRI (28.78 dB), **Second row**: LaSAL (31.06 dB) and LaSAL2 (33.43 dB).

lines. In particular, we implement undersampling based on the golden ratio profile spacing [58], which guarantees a nearly uniform coverage of the space for an arbitrary number of the remaining radial lines. The procedure is as follows. Starting from an arbitrary selected radial line, each next line is chosen by skipping an azimuthal gap of $111.246°$. In practice we cannot always achieve this gap precisely (since we have a finite, although large, number of lines to start with). Therefore we choose the nearest available radial line relative to the position obtained after moving. Since we deal here with non-uniformly sampled k-space data, we need to employ the non-

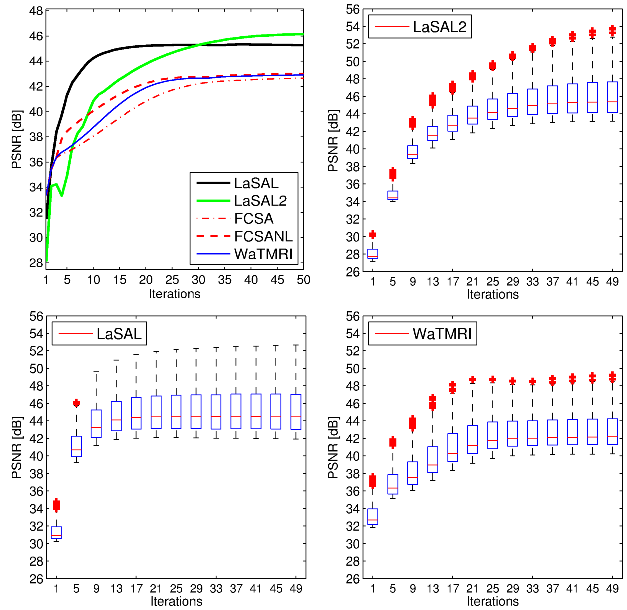

Fig. 9: PSNR values obtained from 248 MRI *brain slices* from the first data set, with random sampling. Mean PSNR (**top left**) and the PSNR distribution for LaSAL2 (**top right**), LaSAL (**bottom left**) and WaTMRI (**bottom right**). The results are presented as a box plot: the edges of the each box represents 25^{th} and 75^{th} percentile while the central mark (red line) in the box is median. The whiskers extend to the most extreme PSNR values which are not considered outliers while outliers are plotted separately with red crosses.

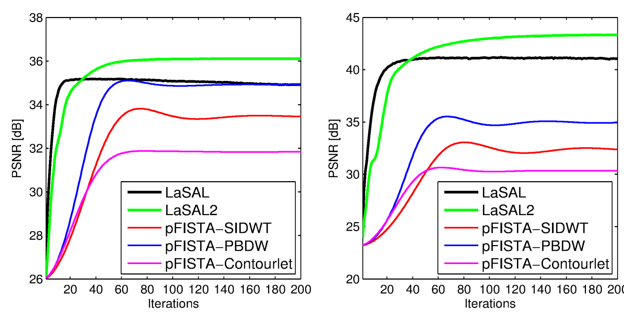

Fig. 10: Comparison with pFISTA [24] on the test image *axial1*. **Left**: radial sampling (maximum PSNR values reached in LaSAL – 27s; LaSAL2 – 51 s; pFISTA-PBDW – 47 s;). **Right**: random sampling (maximum PSNR values reached in LaSAL – 39 s; LaSAL2 – 78 s; FISTA-PBDW – 58 s).

uniform FFT procedures [58], which are commonly used in MRI reconstruction and readily available. The three reference methods (WaTMRI, FCSA and FCSANL) give similar results on this image, so we choose for comparison WaTMRI. Fig. 11 shows visual comparison and SSIM values for LaSAL2 and WaTMRI. At sampling rates up to 30%, LaSAL2 reaches the highest SSIM, while for higher sampling rates it yields the same SSIM scores as LaSAL. For all sampling rates, both LaSAL and LaSAL2 outperform WaTMRI.

F. Convergence

The optimization problem that our method solves is non-convex. For a similar non-convex problem with MRF regularization the authors in [44], argued that a local optimum can be

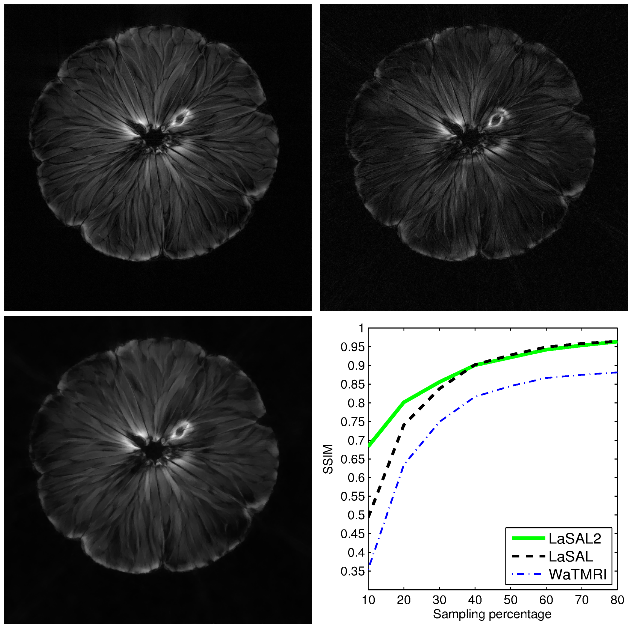

Fig. 11: Reconstructions of the radially sampled *pomelo*. **Top left**: reconstructed from all available data with the conjugate gradient algorithm (reference image). **Top right**: WaTMRI reconstruction from 20% samples, SSIM = 0.65. **Bottom left**: LaSAL2 reconstruction from the same 20% samples, SSIM = 0.80. **Bottom right**: SSIM values for different sampling rates.

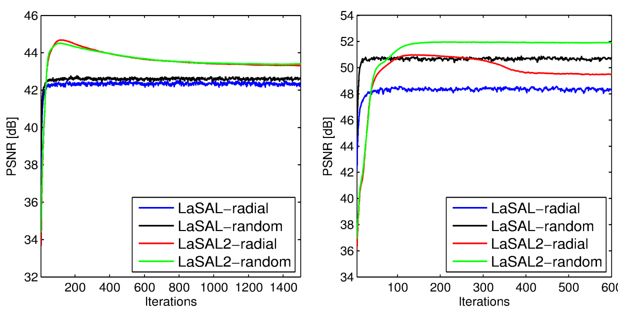

Fig. 12: Experimental evaluation of the stability of the proposed methods on different test images. **Left**: *sagittal1*, 50% sampling. **Right**: *axial2*, 48% sampling.

efficiently obtained by applying alternating minimization. The same argument holds for our method. Although we cannot provide a theoretical proof of convergence, we provide a solid empirical proof of convergence through simulation with different images and different trajectories. The experiments were conducted on various images using radial, spiral and random trajectories.

Fig. 12 shows results for two different images and different sampling trajectories. It can be observed that both LaSAL and LaSAL2 reach stable PSNR for all trajectories. In the case of LaSAL, only negligible oscillations persist around the converged value, while in the case of LaSAL2 no oscillations are observed. Changing the parameters of the MRF model can result in a higher maximum PSNR, at the cost of a less stable convergence.

We also investigated the effect of initializing the reconstruction differently: with a zero image, with white Gaussian noise image (zero mean, standard deviation 50) and with another MRI image as it is illustrated in Fig. 13. In the case of random noise initialization, we run 10 experiments and averaged results. The evolution of PSNR per iteration, after some initial iterations, practically does not depend on the initialization. We obtain similar results when initializing the reconstruction with an MRI image that is different from the one being reconstructed (see the diagram on the right of Fig. 13). In all our experiments, LaSAL and LaSAL2 reached their stable PSNR values that did not depend on the initial image.

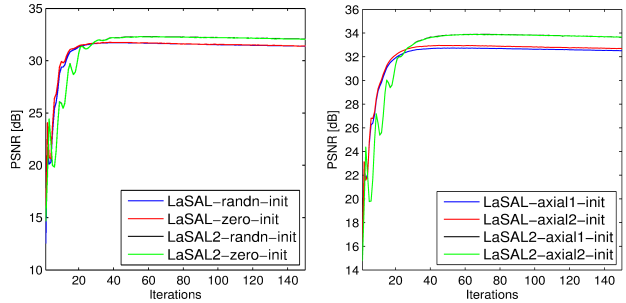

Fig. 13: Influence of the initialization on the reconstruction performance illustrated on reconstructions of *saggital1* from 20% of measurements. **Left**: initializations with zero-image and random noise; random trajectory. **Right**: initializations with different MRI images (*axial1* and *axial2*); radial trajectory.

Next, we analyze consistency of the proposed estimators. Estimation (reconstruction) of the original image from undersampled measurements is statistically consistent if the probability of reconstructing the true image converges to 1 as the number of measurements tends to infinity:

$$\lim_{n\to\infty} Pr(|T(\mathbf{y}_n) - \mathbf{x}| < \epsilon) = 1 \tag{22}$$

where T denotes estimator, n the number of samples in the measurements vector $\mathbf{y}$ and $\mathbf{x}$ the ground truth. Our proposed estimators LaSAL and LaSAL2 alternate between two minimization problems in an iterative procedure for image reconstruction. The first problem is minimization of an energy function composed of a data fitting term and a prior energy term, expressed as the energy of an Ising MRF model. This minimization results in an estimated support of the signal in a transformation domain. A detail analysis of Gibbs-Markov random fields models including statistical consistency of minimum contrast estimators employing these models is provided in [59]. The second problem is an objective function minimization that estimates the signal, constrained to the particular domain (signal space) imposed by the previously estimated signal support. Since this particular objective function is convex, its consistency is trivially proven.

LaSAL and LaSAL2 procedures alternate between the two estimators, inferring jointly the signal and its support in a transform domain. Proving the consistency of each estimator separately does not lead directly to the consistency proof of the joint estimator; the conditions are studied in literature [60], but such a rigorous analysis exceeds the scope of this work.

We provide a finite sample convergence analysis of the joint estimator using an experimental setup similar to the

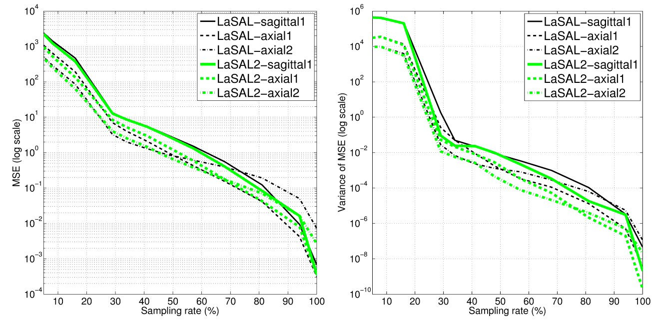

Fig. 14: Experimental evaluation of the finite sample convergence of the proposed methods. MSE (left) and its variance (right) in reconstruction trials with three images at different sampling rates with random trajectories. For each sampling rate 50 experiments were conducted with different realizations of random trajectories and averaged MSE and its variance are plotted.

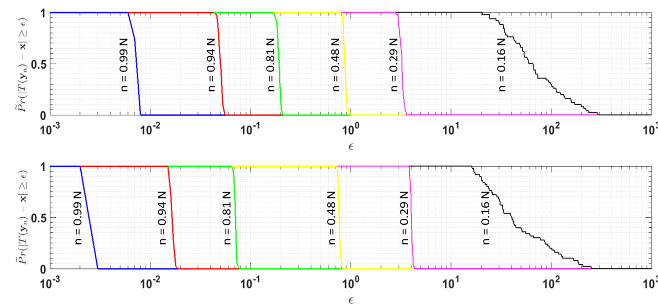

Fig. 15: Empirically estimated probability $\widetilde{Pr}(|T(\mathbf{y}_n) - \mathbf{x}| \geq \epsilon)$ for various values of ϵ and different number of measurements $n = SP \times N$, where SP denotes the sampling rate and N is the size of the ideal image $\mathbf{x}$. Top: reconstructions with LaSAL (on *sagittal1*) Bottom: the corresponding results for LaSAL2.

one in [61]. This experiment evaluates statistically image reconstruction quality as a function of an increasing number of measurements. At each of the considered sampling rates, we perform reconstructions over 50 realizations of randomly generated acquisition trajectories (where each reconstruction contains over 65000 pixels for a 256×256 image) and we record the averaged mean squared error (MSE) and its variance over all the realizations. The evaluation of MSE in this setting is commonly motivated in the statistical literature by Chebychev's inequality, from which it follows that:

$$Pr(|T(\mathbf{y}_n) - \mathbf{x}| \geq \epsilon) \leq \frac{E((T(\mathbf{y}_n) - \mathbf{x})^2)}{\epsilon^2} \qquad (23)$$

Fig. 14 shows the results obtained for three different input images. Relatively high values of variance at small sampling rates (less than 30%) can be attributed to the fact that random trajectories may miss (almost completely) or not the lowest frequency components which are essential for the quality of reconstruction. We conclude that as the number of measurements increases, the MSE and its variance decrease and tend to zero, as expected.

It is interesting to examine also the empirical estimates of the probabilities $Pr(|T(\mathbf{y}_n) - \mathbf{x}| \geq \epsilon)$, which can be obtained from the same experimental setup. For each n, we find an empirical estimate $\widetilde{Pr}(|T(\mathbf{y}_n) - \mathbf{x}| \geq \epsilon)$ as the fraction of the total number of experiments for which the absolute error of the reconstruction was exceeding ϵ. We illustrate these empirical probabilities for one of the test images in Fig. 15. The diagrams show that at very high sampling rates the empirical probability $\widetilde{Pr}(|T(\mathbf{y}_n) - \mathbf{x}| \geq \epsilon)$ indeed tends to zero for $\epsilon > 0.008$ (LaSAL) or $\epsilon > 0.003$ (LaSAL2) on grayscale images in the range [0,255]. These results indicate that the proposed algorithm reliably converges to solutions that lie within a standard deviation that can be ignored safely in any practical application.

VI. Conclusion

This work confirmed the potential of the MRF modelling framework for the recovery of compressively sampled MRI data, that was earlier hinted in [30]. Moreover, we now presented a more comprehensive study and developed a novel algorithm which incorporates the MRF modeling framework into a constrained split augmented Lagrangian method. The resulting algorithm improves upon the C-SALSA method in MRI reconstruction and it also outperforms the earlier method from [30]. The results also demonstrate superior performance of the proposed algorithm in comparison to state-of-the-art methods, both in terms of quantitative performance measures and visually. There is much room to optimize the computations in our method, especially regarding the inference procedure in the MRF-based support estimation. Belief propagation algorithms may be considered as well as various parallelization procedures to optimize the code. This aspect will be part of our future research.

Acknowledgements

This research has been supported by the FP7 project InnoSense. We are grateful to Prof. Dr. Karel Deblaere from the Radiology Department of UZ Gent hospital, Dr. Els Fieremans from the NYU school of Medicine and to Prof. Marleen Vehoye from the University of Antwerp for providing the data. We also thank the Associate Editor and the Reviewers for their constructive suggestions, which improved the quality of the manuscript.

References

[1] G. Wright, "Magnetic resonance imaging," *IEEE Signal Process. Mag.*, vol. 14, no. 1, pp. 56–66, 1997.

[2] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[3] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, 2008.

[4] B. Roman, A. Hansen, and B. Adcock, "On asymptotic structure in compressed sensing," *arXiv preprint arXiv:1406.4178*, 2014.

[5] A. Bastounis and A. C. Hansen, "On random and deterministic compressed sensing and the Restricted Isometry Property in levels," in *Sampling Theory and Applications (SampTA), 2015 International Conference on*. IEEE, 2015, pp. 297–301.

[6] E. J. Candès *et al.*, "Compressive sampling," in *Proceedings of the international congress of mathematicians*, vol. 3. Madrid, Spain, 2006, pp. 1433–1452.

[7] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[8] J. Starck, M. Elad, and D. L. Donoho, “Image decomposition via the combination of sparse representations and a variational approach,” *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1570–1582, 2005. [Online]. Available: http://dx.doi.org/10.1109/TIP.2005.852206

[9] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty, “An efficient algorithm for compressed MR imaging using total variation and wavelets,” in *IEEE Conf. on Computer Vision and Pattern Recognition, CVPR*, 2008, pp. 1–8.

[10] T. Blumensath and M. E. Davies, “Iterative hard thresholding for compressed sensing,” *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[11] J. Yang, Y. Zhang, and W. Yin, “A fast alternating direction method for TVℓ_1-ℓ_2 signal reconstruction from partial Fourier data,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 288–297, 2010.

[12] J. Aelterman, H. Q. Luong, B. Goossens, A. Pižurica, and W. Philips, “Augmented Lagrangian based reconstruction of non-uniformly subnyquist sampled MRI data,” *Signal Processing*, vol. 91, no. 12, pp. 2731–2742, 2011.

[13] C. Chen and J. Huang, “Compressive sensing MRI with wavelet tree sparsity,” in *Advances in neural information processing systems*, 2012, pp. 1115–1123.

[14] J. Huang, S. Zhang, and D. Metaxas, “Efficient MR image reconstruction for compressed MR imaging,” *Medical Image Analysis*, vol. 15, no. 5, pp. 670–679, 2011.

[15] S. R. Rajani and M. Ramasubba Reddy, “An iterative hard thresholding algorithm for CS MRI,” in *SPIE Medical Imaging*, 2012, pp. 83 143W1–83 143W7.

[16] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, “Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing,” *CoRR*, vol. abs/1302.0561, 2013. [Online]. Available: http://arxiv.org/abs/1302.0561

[17] C. Chen and J. Huang, “Exploiting the wavelet structure in compressed sensing MRI,” *Magnetic Resonance Imaging*, vol. 32, no. 10, pp. 1377–1389, 2014.

[18] V. Cevher, “Learning with compressible priors,” in *Advances in Neural Information Processing Systems*, 2009, pp. 261–269.

[19] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, “An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems,” *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2011.

[20] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[21] D. S. Smith, L. R. Arlinghaus, T. E. Yankeelov, and E. B. Welch, “Curvelets as a sparse basis for compressed sensing magnetic resonance imaging,” in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2013, pp. 866 929–866 929.

[22] X. Qu, W. Zhang, D. Guo, C. Cai, S. Cai, and Z. Chen, “Iterative thresholding compressed sensing MRI based on contourlet transform,” *Inverse Probl. Sci. Eng.*, vol. 18, no. 6, pp. 737–758, 2010.

[23] S. Pejoski, V. Kafedziski, and D. Gleich, “Compressed sensing MRI using discrete nonseparable shearlet transform and FISTA,” *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1566–1570, 2015.

[24] Y. Liu, Z. Zhan, J. Cai, D. Guo, Z. Chen, and X. Qu, “Projected iterative soft-thresholding algorithm for tight frames in compressed sensing magnetic resonance imaging,” *IEEE Trans. Med. Imag*, vol. 35, no. 9, pp. 2130–2140, 2016.

[25] S. Ravishankar and Y. Bresler, “MR image reconstruction from highly undersampled k-space data by dictionary learning,” *IEEE Trans. Med. Imag*, vol. 30, no. 5, pp. 1028–1041, 2011.

[26] X. Qu, Y. Hou, F. Lam, D. Guo, J. Zhong, and Z. Chen, “Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator,” *Medical image analysis*, vol. 18, no. 6, pp. 843–856, 2014.

[27] Y. Huang, J. Paisley, Q. Lin, X. Ding, X. Fu, and X.-P. Zhang, “Bayesian nonparametric dictionary learning for compressed sensing MRI,” *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5007–5019, 2014.

[28] B. Adcock, A. C. Hansen, and B. Roman, “The quest for optimal sampling: Computationally efficient, structure-exploiting measurements for compressed sensing,” *arXiv preprint arXiv:1403.6540*, 2014.

[29] V. Cevher, M. F. Duarte, C. Hegde, and R. Baraniuk, “Sparse signal recovery using Markov random fields,” in *Advances in Neural Information Processing Systems*, 2009, pp. 257–264.

[30] A. Pižurica, J. Aelterman, F. Bai, S. Vanloocke, Q. Luong, B. Goossens, and W. Philips, “On structured sparsity and selected applications in tomographic imaging,” in *SPIE Conference on Wavelets and Sparsity XIV*, vol. 8138, 2011, pp. 81 381D–1–12.

[31] T. Goldstein and S. Osher, “The split Bregman method for ℓ_1-regularized problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[32] M. Panić, J. Aelterman, V. Crnojević, and A. Pižurica, “Compressed sensing in MRI with a markov random field prior for spatial clustering of subband coefficients,” in *24th European Signal Processing Conference, EUSIPCO 2016, Budapest, Hungary, August 29 - Sept. 2, 2016*, in press.

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

[34] R. G. Baraniuk, V. Cevher, and M. B. Wakin, “Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 959–971, 2010.

[35] T. Blumensath and M. E. Davies, “Normalized iterative hard thresholding: Guaranteed stability and performance,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 298–309, 2010.

[36] T. Blumensath, “Accelerated iterative hard thresholding,” *Signal Processing*, vol. 92, no. 3, pp. 752–756, 2012.

[37] I. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Communications on pure and applied mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[38] J. M. Bioucas-Dias and M. A. Figueiredo, “A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration,” *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, 2007.

[39] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[40] S. J. Wright, R. D. Nowak, and M. A. Figueiredo, “Sparse reconstruction by separable approximation,” *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, 2009.

[41] P. L. Combettes and V. R. Wajs, “Signal recovery by proximal forward-backward splitting,” *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[42] A. Chambolle, “An algorithm for total variation minimization and applications,” *Journal of Mathematical imaging and vision*, vol. 20, no. 1-2, pp. 89–97, 2004.

[43] L. He and L. Carin, “Exploiting Structure in Wavelet-Based Bayesian Compressive Sensing,” *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3488–3497, 2009.

[44] V. Cevher, P. Indyk, L. Carin, and R. G. Baraniuk, “Sparse signal recovery and acquisition with graphical models,” *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 92–103, 2010.

[45] A. Kyrillidis, L. Baldassarre, M. El Halabi, Q. Tran-Dinh, and V. Cevher, “Structured sparsity: Discrete and convex approaches,” in *Compressed Sensing and its Applications*. Springer, 2015, pp. 341–387.

[46] S. Z. Li, *Markov random field modeling in image analysis*. Springer Science & Business Media, 2009.

[47] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, “A joint inter-and intrascale statistical model for Bayesian wavelet based image denoising,” *IEEE Trans. Image Process.*, vol. 11, no. 5, pp. 545–557, 2002.

[48] E. P. Simoncelli, “Statistical models for images: Compression, restoration and synthesis,” in *Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on*, vol. 1. IEEE, 1997, pp. 673–678.

[49] J. Besag, “On the statistical analysis of dirty pictures,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 259–302, 1986.

[50] V. Kolmogorov and R. Zabih, “What energy functions can be minimized via graph cuts ?” *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 26, pp. 65–81, 2004.

[51] K. P. Murphy, Y. Weiss, and M. I. Jordan, “Loopy belief propagation for approximate inference: An empirical study,” in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1999, pp. 467–475.

[52] S. Kirkpatrick, “Optimization by simulated annealing: Quantitative studies,” *J. Stat. Phys.*, vol. 34, no. 5-6, pp. 975–986, 1984.

[53] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, “Image restoration with compound regularization using a bregman iterative algorithm,” in *SPARS’09-Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[54] H. Cline and T. Anthony, “Magnetic resonance imaging with interleaved fibonacci spiral scanning,” Patent, 08 28, 2001. [Online]. Available: http://www.google.com/patents/US6281681

[55] J. Huang and F. Yang, "Compressed magnetic resonance imaging based on wavelet sparsity and nonlocal total variation," in *9th IEEE Internat. Symp. on Biomedical Imaging (ISBI)*, 2012, pp. 968–971.

[56] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[57] B. Goossens, J. Aelterman, H. Luong, A. Pižurica, and W. Philips, "Efficient design of a low redundant discrete shearlet transform," in *IEEE Internat. Workshop on Local and Non-Local Approximation in Image Processing, LNLA*, 2009, pp. 112–124.

[58] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers, and O. Doessel, "An optimal radial profile order based on the Golden Ratio for time-resolved MRI," *IEEE Trans. Med. Imag*, vol. 26, no. 1, pp. 68–76, 2007.

[59] C. Gaetan and X. Guyon, *Spatial statistics and modeling*. Springer, 2010, vol. 81.

[60] U. Niesen, D. Shah, and G. W. Wornell, "Adaptive alternating minimization algorithms," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 1423–1429, 2009.

[61] P.-L. Loh *et al.*, "Statistical consistency and asymptotic normality for high-dimensional robust M-estimators," *The Annals of Statistics*, vol. 45, no. 2, pp. 866–896, 2017.