

On structured sparsity and selected applications in tomographic imaging

Aleksandra Pižurica^a, Jan Aelterman^a, F. Bai^a, Sam Vanlooche^b, Hiep Quang Luong^a, Bart Goossens^a and Wilfried Philips^a

^aDepartment of Telecommunications and Information Processing, TELIN-IPI-IBBT, Ghent University, Belgium;

^bDepartment of Information Technology, INTEC, Ghent University, Belgium

ABSTRACT

This work explores the potentials of *structure* encoding in sparse tomographic reconstructions. We are encoding *spatial* structure with Markov Random Field (MRF) models and employ it within Magnetic Resonance Imaging (MRI) and Quantitative Microwave Tomography. We illustrate thereby also different ways of MRF modelling: as a discrete, binary field imposed on hidden labels and as a continuous model imposed on the observable field. In case of MRI, the analyzed approach is a straightforward extension of sparse MRI methods and is related to the so-called LaMP (Lattice Matching Pursuit) algorithm, but with a number of differences. In case of Microwave Tomography, we give another interpretation of structured sparsity using much different, but also effective approach. Thorough experiments demonstrate clear advantages of MRF based structure encoding in both cases and motivate strongly further development.

Keywords: sparsity, structure, Markov Random Field, Magnetic Resonance Imaging, Microwave Tomography.

1. INTRODUCTION

We are witnessing rapid development of sparsity promoting reconstructions, among which the so-called compressed-sensing or compressive sampling (CS).¹⁻³ In Magnetic Resonance Imaging (MRI), CS approaches are gaining a lot of popularity after demonstrating great potentials for improving the imaging speed.^{4,5} Sparsity promoting reconstructions gained a lot of interest in other tomographic reconstructions too, including X-Ray Computerized Tomography (CT).⁶ Some recent works addressed the application of CS in Microwave Imaging,⁷ but to our knowledge only in the context of *qualitative* microwave tomography. The goal there is to locate the position of targets only and *not* to reconstruct exactly the complex permittivity profile, as we attempt to do in quantitative tomography.

The goal of this work is to explore the use of *structure* together with sparsity in tomographic reconstructions. We emphasize here *spatial* structure and encode it with Markov Random Field models. Our spatial modeling is applied in two types of tomographic reconstructions: *MRI* and *Quantitative Microwave Tomography*. With these two applications we also illustrate two types of MRF based structure encoding: (1) a discrete (binary) MRF imposed on the hidden labels (of transform coefficients, in MRI) and (2) a continuous MRF model imposed on the observable field (in quantitative microwave tomography). While the first approach relates closely to some of the recent ideas on structured sparsity, the second one offers an alternative formulation of structured sparsity.

Recently a number of works were reported on ideas of using data geometry, or structure, in sparse reconstructions. These include approaches like group sparsity^{8,9} or group Lasso,¹⁰ block sparsity,^{11,12} tree sparsity,¹³ more general graph sparsity,¹⁴ model based sparsity,¹⁵ etc. Despite a lot of theoretical results, few works were reported that demonstrate true benefits of using structured sparsity in practice. The existing related works in MRI¹⁶ concentrate on dynamic time sequences and not really on spatial structure that we wish to encode here. Our work is most closely related to LaMP (Lattice Matching Pursuit)¹⁷ where Markov Random Field prior

Further author information: (Send correspondence to Aleksandra Pižurica)
E-mail: sanja@telin.ugent.be, Telephone: +32 9 264 34 15

was used to encode structure in sparse reconstructions using Orthogonal Matching Pursuit (OMP). The LaMP estimator has demonstrated great potentials in surveillance videos and some limited evaluation in tomographic reconstructions (on synthetic Shepp-Logan phantom) was also demonstrated.^{17,18}

We develop here a related structured-sparsity approach, but with a number of differences (including different likelihood model and different optimization algorithm: Split-Bregman¹⁹ (versus Orthogonal Matching Pursuit in LaMP). To make a parallel with LaMP, we name this new method LaSB as an acronym of “Lattice Split-Bregman” and we evaluate it thoroughly in MRI. We start from our recent sparse MRI technique,²⁰ which employs Split-Bregman optimization algorithm and shearlet regularization. Essentially, we replace the soft shrinkage of the individual shearlet coefficients by structure-aware MRF-based estimation. This enables us to evaluate the gain that comes solely from MRF-based structure encoding and we show with a number of simulations that this gain is substantial in the case of Split-Bregman optimization too. The results support the findings of Cevher *et al*¹⁷ where their LaMP method (MRF-based structure encoding within the OMP) was found superior to other greedy reconstruction algorithms such as CoSaMP²¹ and fixed point continuation (FPC).²² We evaluate the gain from structure encoding within a more recent and better performing optimization algorithm, and also more thoroughly, on real MRI images and analyzing different parameter settings and showing the evolution over the iterations.

The second part of the paper considers a different type of tomographic reconstructions: microwave tomography. The problem is nonlinear and very challenging. We formulate a reconstruction algorithm where the energy of the Markov Random Field model plays a role of the structured sparsity prior. The approach itself relies on the known theory of MRF based regularizers but we think that its interpretation in terms of structured sparsity opens some interesting possibilities for further theoretical and practical developments. Our initial results are very encouraging and compare well with the state-of-the-art approaches in microwave tomography.

The paper is organized as follows. In Section 2, after reviewing briefly “sparse MRI” formulation and Split-Bregman optimization, we introduce our new LaSB algorithm and demonstrate its results. In Section 3, we formulate an alternative MRF-based structured sparsity approach and describe it in the setting of quantitative microwave tomography. Finally, we conclude the paper in Section 4.

2. STRUCTURED SPARSITY IN MRI USING HIDDEN MRF MODEL

The primary goal of the so-called “Sparse MRI” techniques^{4,5} is to optimize the acquisition time versus image quality. We consider the following problem

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} J(\mathbf{x}) \quad \text{s.t.} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 < \sigma^2 \quad (1)$$

where the complex vectors $\mathbf{y} \in \mathbb{C}^M$ and $\mathbf{x} \in \mathbb{C}^N$ ($M \leq N$) are the acquired k-space data and the ideal (or hypothetical desired) MRI image, respectively, in column stack ordering (i.e. the image columns stacked into one vector). The functional $J(\mathbf{x})$ is a given data complexity penalty. The measurement matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$ is an undersampled Fourier operator (or the non-uniform Fourier transform NUFT operator for non-Cartesian trajectories) and σ controls the fidelity of the reconstruction to the measured data.

In essence, the aim is to minimize data complexity penalty subject to data fidelity. The data complexity penalty (sparsity prior) is commonly the L1 norm of a convex function $\Phi(\mathbf{x})$: $J(\mathbf{x}) = |\Phi(\mathbf{x})|$, where $\Phi(\mathbf{x})$ is a sparsifying transform $\Phi(\mathbf{x}) = \mathbf{S}\mathbf{x}$ or a combination of such transforms. Usually, \mathbf{S} is a wavelet or “wavelet-like” (curvelet,²³ shearlet²⁴ or related) transform, or it is simply a discrete gradient²⁵ reducing in this case $J(\mathbf{x})$ to the so-called bounded variation, i.e., total variation²⁶ (TV norm). It is also common to use a combination of the TV norm and a wavelet-like (i.e. Besov) regularizer: $J(\mathbf{x}) = \|\mathbf{x}\|_{TV} + \|\mathbf{x}\|_{B_{1,1}} = |\nabla\mathbf{x}| + |\mathbf{S}\mathbf{x}|$. Such combined regularizers in sparse MRI were already discussed by other researchers^{27,28} and we will also incorporate a combination of TV and shearlet regularization in our experiments.

2.1 Split-Bregman optimization

There exist many competing algorithms^{21,22,29–34} to solve the constrained optimization problem in (1) with $J(\mathbf{x}) = |\Phi(\mathbf{x})|$ for convex $\Phi(\mathbf{x})$ and assuming mutually independent components of \mathbf{x} . In this paper we use the

Split-Bregman method.¹⁹ This method builds on the Bregman iteration³⁵ (or augmented Lagrangian) method, which can be expressed as:

$$\begin{cases} \mathbf{x}_{i+1} = \underset{\mathbf{x}}{\operatorname{argmin}} J(\mathbf{x}) + \frac{\lambda}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}_i\|^2 \\ \mathbf{y}_{i+1} = \mathbf{y}_i + \mathbf{y} - \mathbf{A}\mathbf{x}_{i+1} \end{cases} \quad (2)$$

The key to splitting method is to “de-couple” the L1 and L2 portions of the energy in this minimization.¹⁹ Denoting $J(\mathbf{x}) = |\Phi(\mathbf{x})|$ and introducing an auxiliary variable $\mathbf{d} = \Phi(\mathbf{x})$, the problem in the first line in (2) becomes restated as

$$\underset{\mathbf{x}, \mathbf{d}}{\operatorname{argmin}} |\mathbf{d}| + \frac{\lambda}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}_i\|^2 \quad \text{s.t.} \quad \mathbf{d} = \Phi(\mathbf{x}) \quad (3)$$

This is further transformed into a sequence of unconstrained problems using the Bregman formulation, so the problem in (2) becomes:

$$\begin{cases} (\mathbf{x}_{i+1}, \mathbf{d}_{i+1}) = \underset{\mathbf{x}, \mathbf{d}}{\operatorname{argmin}} |\mathbf{d}| + \frac{\lambda}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}_i\|^2 + \frac{\mu}{2} \|\mathbf{d} - \Phi(\mathbf{x}) - \mathbf{b}_i\|^2 \\ \mathbf{b}_{i+1} = \mathbf{b}_i + \Phi(\mathbf{x}_{i+1}) - \mathbf{d}_{i+1} \end{cases} \quad (4)$$

which is efficiently solved by iteratively minimizing over \mathbf{x} and \mathbf{d} separately, yielding the full algorithm as follows:

$$\begin{aligned} &\text{While } \|\mathbf{A}\mathbf{x}_i - \mathbf{y}\|^2 > \sigma^2 \\ &\quad \text{For } i = 1 \text{ to } N \\ &\quad \quad \mathbf{x}_{i+1} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{\lambda}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}_i\|^2 + \frac{\mu}{2} \|\mathbf{d}_i - \Phi(\mathbf{x}) - \mathbf{b}_i\|^2 \quad (\text{step } 1) \\ &\quad \quad \mathbf{d}_{i+1} = \underset{\mathbf{d}}{\operatorname{argmin}} |\mathbf{d}| + \frac{\mu}{2} \|\mathbf{d} - \Phi(\mathbf{x}_{i+1}) - \mathbf{b}_i\|^2 \quad (\text{step } 2) \\ &\quad \quad \mathbf{b}_{i+1} = \mathbf{b}_i + \Phi(\mathbf{x}_{i+1}) - \mathbf{d}_{i+1} \quad (\text{step } 3) \\ &\quad \text{End} \\ &\quad \mathbf{y}_{i+1} = \mathbf{y}_i + \mathbf{y} - \mathbf{A}\mathbf{x}_{i+1} \quad (\text{step } 4) \\ &\text{End} \end{aligned} \quad (5)$$

In our recent work,²⁰ (*step 1*) in (5) was solved by a preconditioned conjugate gradient method (due to the employed NUFT operator). The sparsity prior was L1 norm of the shearlet coefficients: $J(\mathbf{x}) = |\Phi(\mathbf{x})| = |\mathbf{S}\mathbf{x}|$ and *step 2* was solved by component-wise soft-thresholding:

$$[\mathbf{d}_{i+1}]_m = \Theta\left([\mathbf{S}\mathbf{x}_{i+1} - \mathbf{b}_i]_m, \frac{1}{\mu}\right) \quad (6)$$

where the soft thresholding function Θ is defined as

$$\Theta(u, \gamma) = \operatorname{sign}(u) \max(|u| - \gamma, 0) \quad (7)$$

and $[\mathbf{d}]_m$ denotes the m -th component of the vector d . Related approach was also discussed by the authors of Split-Bregman method¹⁹ in somewhat more general setting with $J(\mathbf{x})$ being a sum of TV norm and L1 norm of the wavelet coefficients. Other related methods include sparse MRI using the conventional Bregman iteration.²⁷

We will use in our experiments the same shearlet regularization as in our earlier work,²⁰ and we will also make some experiments with combined shearlet and TV regularization: $J(\mathbf{x}) = \lambda_{TV} \|\mathbf{x}\|_{TV} + \lambda_S |\mathbf{S}\mathbf{x}|$ with constants λ_{TV} and λ_S weighting the influence of the two components.

2.2 LaSB - Lattice Split Bregman Algorithm

We will refer to the algorithm described in (5)-(7) as *SB* algorithm. Our extension to it consists in solving *step 2* in (5) differently, i.e., replacing the soft-thresholding part in (6)-(7) by MRF-based estimation. This new algorithm will be denoted as *Lattice SB* or short *LaSB* to make a parallel with the related *LaMP*.¹⁷

We impose the MRF model on *hidden labels* attached to the (shearlet) coefficients. These hidden labels should tell whether the coefficient is important or not, based on its magnitude and based on the prior knowledge

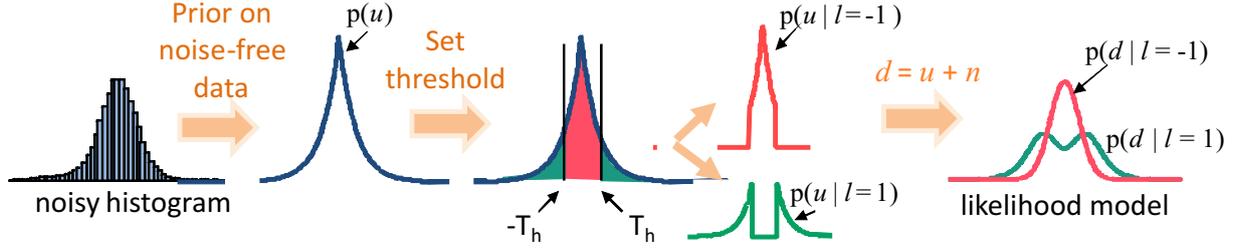


Figure 1. Estimating the likelihood model from the coefficients histogram (see text).

about spatial clustering of important coefficients (in typical, say, MRI images). Conforming to the notation above, let $[\mathbf{l}]_m \in \{-1, 1\}$ denote a *hidden label* attached to the m -th coefficient $[\mathbf{d}]_m$, where $[\mathbf{l}]_m = -1$ means that $[\mathbf{d}]_m$ is insignificant (i.e. “not of interest”) and, contrary, $[\mathbf{l}]_m = 1$ means that $[\mathbf{d}]_m$ is significant (i.e. “signal of interest”). We will denote by $\hat{\mathbf{l}}$ the Maximum a Posteriori (MAP) estimate of the unknown label field \mathbf{l} . Our LaSB algorithm replaces the soft-thresholding part in SB (6)-(7) by a *lattice index selector*

$$[\mathbf{d}_{i+1}]_m = \Lambda\left([\mathbf{S}\mathbf{x}_{i+1} - \mathbf{b}_i]_m, \hat{[\mathbf{l}]}_m\right) \quad (8)$$

with Λ defined as

$$\Lambda(u_i, l_i) = \begin{cases} u_i, & l_i = 1 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

In essence, we are estimating a “mask“ that denotes significant coefficients and then we select the surviving coefficients according to this mask, while zeroing all the others. We use the Metropolis sampler³⁶ as an inference engine to find the MAP estimate of the hidden label field, and we run it in practice in each subband of the sparsifying (here, shearlet) transform. Contrary to common perception about long computation time with this type of random samplers, the Metropolis sampler converges in this application very quickly. We use only few iterations (typically 5). Moreover, the Metropolis sampler leaves a lot of flexibility in terms of specifying the clique potentials and likelihood models, which is not present with all inference methods (like graph cuts). However, the use of one or the other inference method for the MAP mask estimation in our algorithm is an implementation detail that shouldn’t influence much the results.

2.3 Specifying prior and likelihood models

MRFs encode structure by assigning appropriate *potentials* (or “penalties”) to certain combinations of the labels in a group of interacting pixels, i.e., *cliques*. The joint probability of a MRF is a *Gibbs distribution*

$$P(\mathbf{L} = \mathbf{l}) = \frac{1}{Z} e^{-H(\mathbf{l})/T} \quad (10)$$

where the energy $H(\mathbf{l})$ is decomposed as a sum of clique potentials over all possible cliques: $H(\mathbf{l}) = \sum_{C \in \mathcal{C}} V_C(\mathbf{l})$. The normalizing constant Z , which sums (10) over all possible configurations \mathcal{L} : $Z = \sum_{\mathbf{l} \in \mathcal{L}} e^{-H(\mathbf{l})/T}$ is called the partition function and the “temperature” T controls the peaking in the probability density.³⁶ In particular we use the Ising model

$$H(\mathbf{l}) = \alpha \sum_i l_i + \beta \sum_{i,j \in \mathcal{C}} l_i l_j \quad (11)$$

with the second-order neighborhood. We give no preference to either “-1” or “1” labels, so we set $\alpha=0$. The parameter β controls the strength of the clustering. If $\beta < 0$ then equal neighboring labels contribute to the lower energy (i.e. favoring $l_i = l_j$), while the opposite is true for $\beta > 0$. For this isotropic model we can define the “neighborhood potential” as

$$V_{N_i}(\mathbf{l}) = -\gamma \sum_{j \in N_i} l_i l_j \quad (12)$$

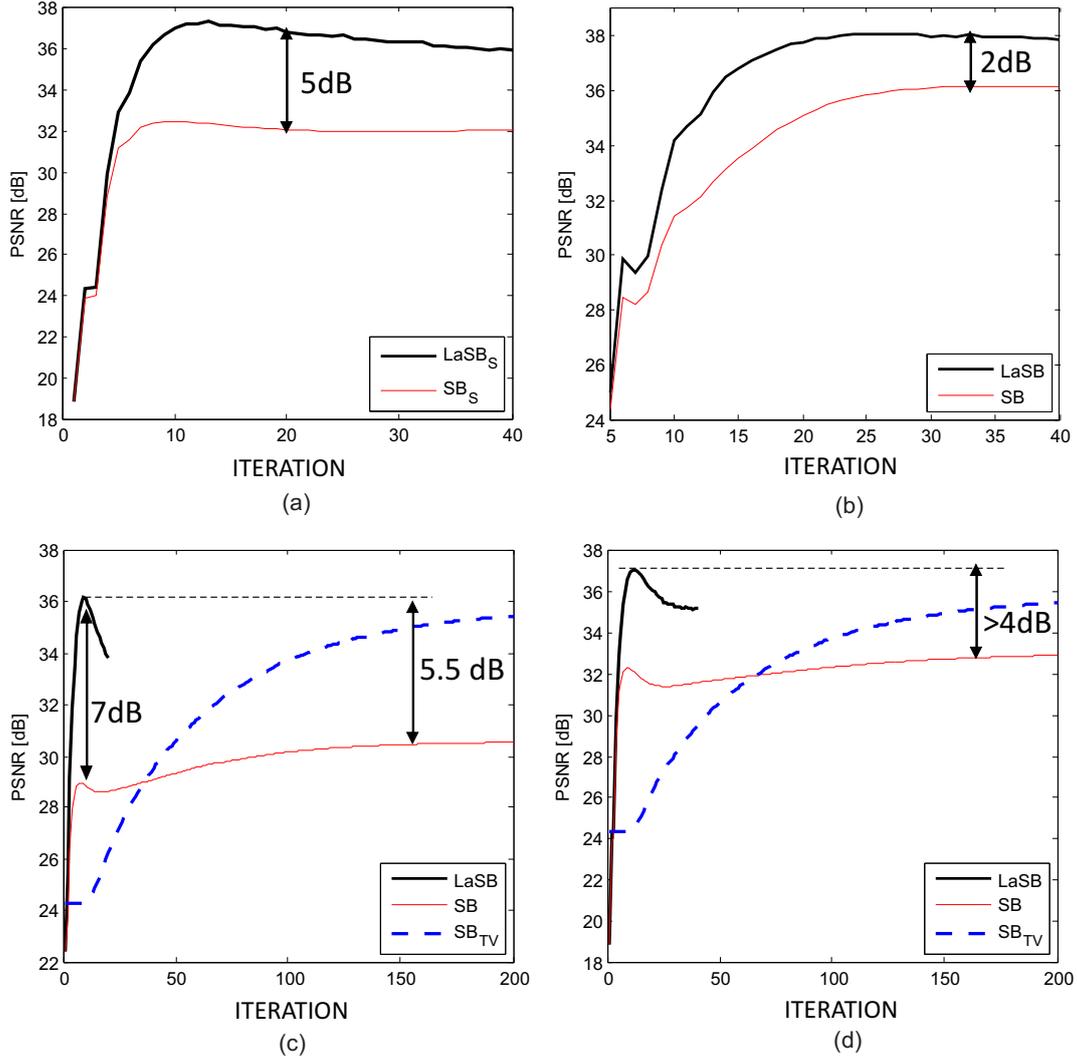


Figure 2. PSNR per iteration for LaSB versus SB, with different parameter settings and with random subsampling 50%. (a) Shearlet regularization only (SB_S and $LaSB_S$): $\lambda_{TV}=0$, $\lambda_S=0.4$; (b) $\lambda_{TV}=1$, $\lambda_S=1$; (c) $\lambda_{TV}=0.01$, $\lambda_S=0.15$; (d) $\lambda_{TV}=0.01$, $\lambda_S=0.4$. SB_{TV} denotes SB with TV regularization only ($\lambda_{TV}=0.01$, $\lambda_S=0$).

where N_i is the neighborhood of i , and $\gamma > 0$. In this way we can decompose the joint probability (10) into contributions summed over all indices (pixels) of our lattice:

$$P(\mathbf{L} = \mathbf{l}) = \frac{1}{Z} \exp\left(-\frac{1}{T} \sum_{i \in \mathcal{I}} V_{N_i}(\mathbf{l})\right) \quad (13)$$

Our likelihood model (conditional densities of the (noisy) coefficients d given the label l : $p_{d|l}(d|-1)$ and $p_{d|l}(d|1)$) is derived as in our previous works^{37,38} (see schematic representation in Fig. 1). We choose a prior model $p(u)$ for noise-free subband coefficients u (such as Laplacian, generalized Laplacian or other suitable, highly curtotic prior) and estimate the parameters of this prior from the noisy/degraded coefficient histogram and using knowledge about noise standard deviation (or more general, knowledge about the image degradation). In case of MRI we use the empty area on the borders of MRI image to estimate the noise standard deviation and scale it appropriately to obtain noise standard deviations σ in each subband. Once the parameters of the prior model are estimated, we need to specify a threshold T_h that identifies the notion of a “significant” signal component. This is the only parameter which is not estimated from the data. We relate this parameter to noise standard deviation,

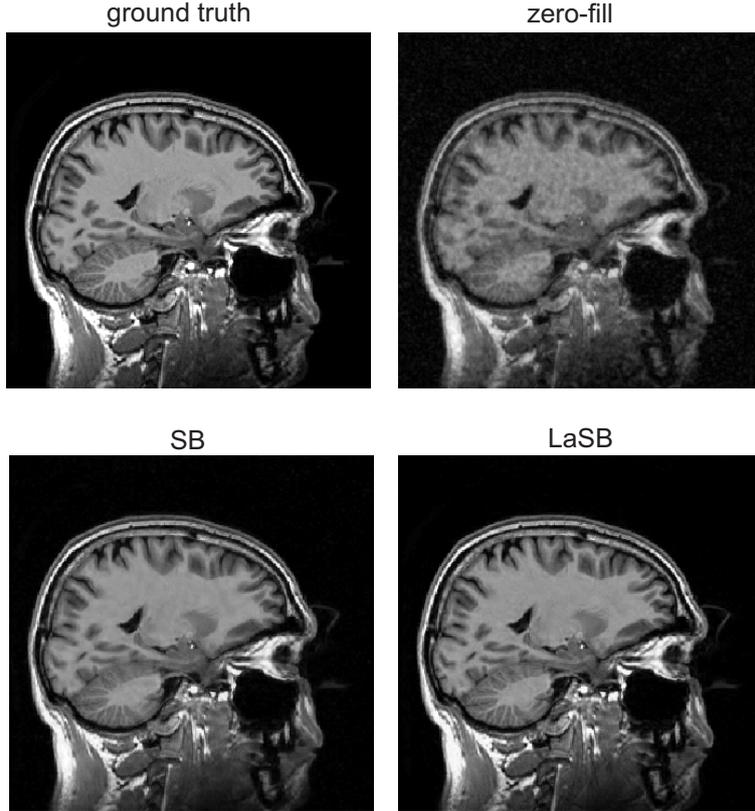


Figure 3. An illustration of the results in the 50% random subsampling experiment. Best results for both SB and LaSB are shown, with $\lambda_{TV}=0.01$ and $\lambda_S=0.4$. For SB: PSNR=32.9dB (reached after 200 iterations); For LaSB: PSNR=37dB (reached after 12 iterations).

similarly as in our earlier work,³⁸ but in a more conservative manner (T_h being in fact only a small fraction of σ in each subband, of the order of 10% or even less). Once the threshold T_h is specified, we obtain the conditional probability density functions of the noise-free coefficients $p_{u|l}(u|1)$ and $p_{u|l}(u|0)$ by rescaling the central part ($u \leq T_h$) and the tails ($u > T_h$) of the marginal prior $p(u)$ so that they integrate to 1. Finally, the corresponding conditional densities of the noisy coefficients, $p_{d|l}(d|1)$ and $p_{d|l}(d|0)$, are obtained from $p_{u|l}(u|1)$ and $p_{u|l}(u|0)$, respectively, for the given degradation model. If the degradation model in each subband is simplified by additive white Gaussian noise, then we just need to convolve the conditional densities of the noise-free data with a Gaussian function. In case of a more complex (and more realistic) degradation model, we might need to calculate the corresponding densities empirically, but essentially the main idea remains the same. For the results reported in this paper we have used the simplified white Gaussian noise model in each subband.

Having specified the prior model and the likelihood model, we can run the Metropolis sampler to yield the MAP estimate of the spatial support of significant coefficients $\hat{\mathbf{l}}$. The Metropolis sampler perturbs randomly labels producing from the current mask \mathbf{l} a next “candidate” mask \mathbf{l}^C , which gets accepted with probability³⁷ $p = p(d_i|l_i^C)/p(d_i|l_i) \exp(V_{N_i}(\mathbf{l}) - V_{N_i}(\mathbf{l}^C))$.

2.4 Results: LaSB versus SB

In this Section we compare the performance of our structured sparsity LaSB algorithm to SB. In both cases we use non-decimated shearlet transform (3 scales, 8 orientations). First we performed experiments with random subsampling on a Cartesian grid. The results in Fig. 2 - Fig. 4 correspond to random subsampling of k-space by 50%. For the sake of fair comparison between different methods and different parameter values, we created randomly subsampled matrix once and then saved it and used for all the presented results.

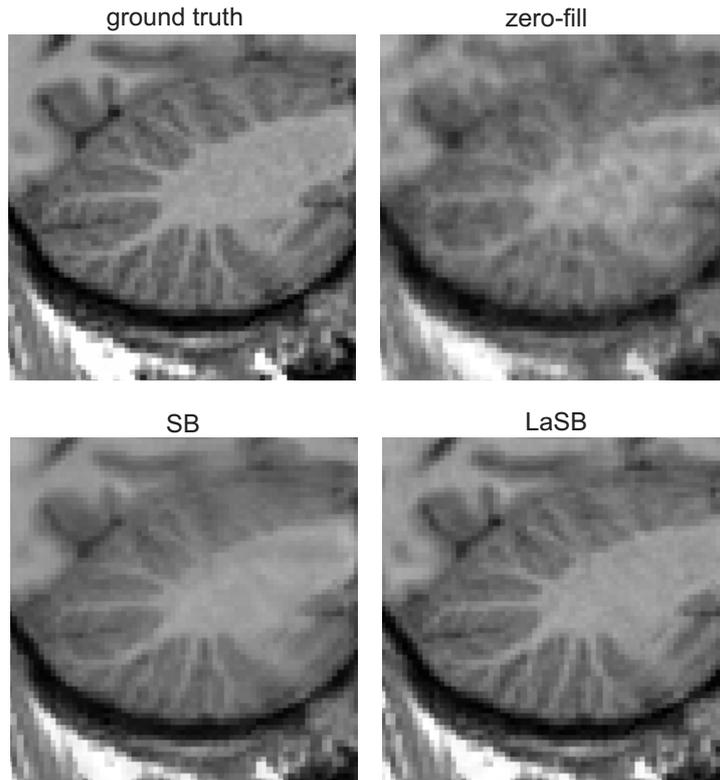


Figure 4. Details from Fig. 3.

The plots in Fig. 2 show Peak Signal to Noise Ratio (PSNR) as a function of the number of iterations. We repeated experiments with many different parameter settings, some of these are illustrated in Fig. 2 (a)-(d). In all cases, the relative comparison between LaSB and SB is similar: LaSB achieves higher PSNR and in less iterations. Visually, LaSB reconstructs details much better than SB, see Fig. 3 and Fig. 4.

Some problems still remain. With LaSB, the PSNR drops after certain number of iterations. While this in general can occur with here employed optimization algorithms, an additional complication could come from the non-convexity of the employed estimator. This aspect needs to be studied further.

We also applied our LaSB algorithm on MRI images with different sampling trajectories that are used in

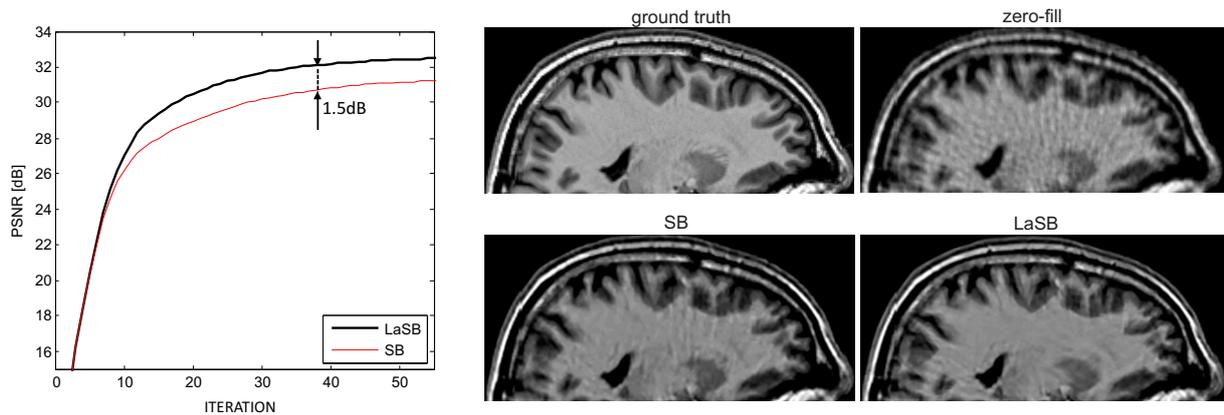


Figure 5. An illustration of the results in experiments with variable density spiral trajectories.

practice or that are more feasible than random sampling given the technological constraints. In particular, we experimented with different radial, Archimedean and variable density spirals. Remarkably, LaSB performance is much more stable with these trajectories than in the case of random subsampling (PSNR saturates after a given maximum). Comparing to SB, we achieve again higher PSNR and in fewer iterations. Figure 5 illustrates these results for a variable density spiral trajectory.

3. MICROWAVE TOMOGRAPHY USING STRUCTURED SPARSITY

Quantitative microwave tomography aims at estimating the permittivity profile of an unknown object through electromagnetic scattering experiments.^{39,40} The object is illuminated by transmitting antennas and the scattered field is measured by receiving antennas. The inversion problem (estimating the unknown permittivity given the scattered field and knowing the incident field) requires solving a system of nonlinear equations, which typically involves iterative procedures. This problem (also called “full wave”³⁹ microwave imaging) is both mathematically and computationally much more challenging than *qualitative* microwave imaging, where the goal is to determine only qualitatively the shape of the permittivity profile and not its exact values.

Estimation of the unknown permittivity profile iterates typically between the *forward* problem (simulating the scattered field based on the known or assumed permittivity profile) and the *inverse* problem (estimating the unknown permittivity from the scattered field). The permittivity profile is usually initialized to the known background permittivity. Depending on the error between the simulated and the measured scattered field, the process is stopped or the permittivity is updated. In this update process the regularization plays an important role, and this is where we will introduce a method based on discontinuity adaptive MRF.

3.1 The inverse scattering problem

Before formalizing the electromagnetic inverse scattering problem, let us first set the notation and introduce some necessary background. We will consider an inhomogeneous dielectric object embedded in homogeneous (infinitely extending in all directions) background with permittivity ϵ_b and permeability μ_0 . The interaction of the object with time harmonic electromagnetic fields is entirely determined by its complex permittivity ϵ , which is a function of the 3D position vector \mathbf{r} and the angular frequency ω .⁴⁰

$$\epsilon(\mathbf{r}, \omega) = \epsilon_0 \epsilon_r(\mathbf{r}, \omega) - j \frac{\sigma(\mathbf{r}, \omega)}{\omega} \quad (14)$$

where ϵ_r is the (real) dielectric permittivity of the object and σ is the conductivity. We treat only single frequency scattering, hence ω -dependency of the permittivity and the $e^{j\omega t}$ dependency of the phasors will be omitted.

The goal of electromagnetic inverse scattering is to reconstruct $\epsilon(\mathbf{r})$ within a bounded investigation domain \mathcal{D} (which includes the unknown object) from the multi-view scattering data. The domain \mathcal{D} is therefore illuminated with a number of different incident fields \mathbf{E}_i^{inc} ($i=1, \dots, N^I$). For each such illumination i , the scattered field (i.e. the difference between the total field and the incident field) is measured in a number of measurement points $\mathbf{r}_{k,i}$ ($k=1, \dots, N^R$), which are possibly different for different i . The measurement procedure for the scattered fields is modelled by taking the component of \mathbf{E}_i^{scat} along the directions $\mathbf{v}_{k,i}$ in the points $\mathbf{r}_{k,i}$, i.e., $\mathbf{E}_i^{scat}(\mathbf{r}_{k,i}) \mathbf{v}_{k,i}$.

Let $\mathbf{e}^{scat}(\boldsymbol{\epsilon})$ be the N^D -dimensional vector that contains the simulated scattered field components $\mathbf{E}_i^{scat}(\mathbf{r}_{k,i}) \mathbf{v}_{k,i}$ for all combinations of N^I illuminations and N^R measurements and for a given permittivity vector $\boldsymbol{\epsilon}$. Furthermore, let the N^D -dimensional vector \mathbf{e}^{meas} contain the measured data. The least squares data fit cost function is then defined as⁴⁰

$$\mathcal{J}^{LS}(\boldsymbol{\epsilon}) = \frac{1}{N^{LS}} \|\mathbf{e}^{scat}(\boldsymbol{\epsilon}) - \mathbf{e}^{meas}\|^2 \quad (15)$$

where $N^{LS} = \|\mathbf{e}^{meas}\|^2$. Important to note is that \mathcal{J}^{LS} is *not* quadratic in $\boldsymbol{\epsilon}$, because the scattered field $\mathbf{E}_i^{inc}(\mathbf{r})$ is a *non-linear* function of the complex permittivity $\epsilon(\mathbf{r})$.^{39,40} This, together with high dimensionality, makes the problem even more challenging.

Recent approaches to solving the inverse scattering problem usually combine the data fit term from (15) with different regularizers, such as multiplicative smoothing (MS)^{39,41} and value picking (VP) regularization.⁴² VP achieves very good results on relatively simple piece-wise homogeneous profiles, but cannot cope well with more complex profiles (that would appear, e.g., in natural tissues). The MS smoothing is better suited for such profiles, but it tends to oversmooth the images.

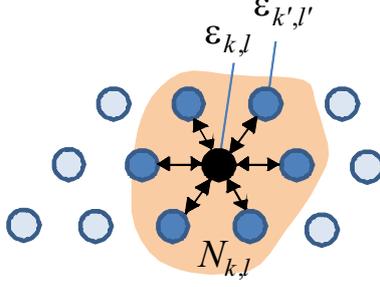


Figure 6. MRF model imposed on the field of permittivities.

3.2 MRF-based regularization and structured sparsity interpretation

Conforming to our notation of sparsity promoting reconstructions in Section 2 and without loss of generality, we can represent the electromagnetic inverse scattering problem as

$$\epsilon^* = \underset{\epsilon}{\operatorname{argmin}} J(\epsilon) + \frac{\lambda}{2} \|e^{\operatorname{scat}}(\epsilon) - e^{\operatorname{meas}}\|^2 \quad (16)$$

where $J(\epsilon)$ is our structured sparsity prior. This problem resembles obviously the formulation in (2), with the most important difference that the second term here is not quadratic in ϵ because of the nonlinearity of $e^{\operatorname{scat}}(\epsilon)$.

We will express $J(\epsilon)$ as the *energy* of the MRF field. In contrast to Section 2.2 where we used a binary MRF on the hidden labels, here we impose a *continuous* MRF model on the observable field of complex permittivities ϵ . Like in Section 2.3, the joint probability of the Markov Random Field model is the Gibbs distribution (10) with energy decomposed as a sum of clique potentials, which is here the structured sparsity prior:

$$J(\epsilon) = H(\epsilon) = \sum_{C \in \mathcal{C}} V_C(\epsilon). \quad (17)$$

Now the clique potentials $V_C(\epsilon)$ are in general continuous functions of the complex permittivity values in the pixels (voxels) that constitute the clique C . In practice, we will consider pairwise cliques. For simplicity of notation, consider 2-dimensional (2D) setting (extension to 3D is straightforward). The energy function of our MRF is hence

$$J(\epsilon) = \sum_{(i,j)} \sum_{(i',j') \in N_{i,j}} g(\epsilon_{i,j} - \epsilon_{i',j'}) \quad (18)$$

where $N_{i,j}$ denotes the neighborhood of the pixel (i,j) (in practice we use 8-neighborhood system), and $g(\eta)$ is a *discontinuity adaptive* clique potential, which should ideally penalize small differences in permittivity but without smoothing true transitions in the permittivity profile.

Experimenting with a number of discontinuity adaptive potential functions, we found that best results are obtained with functions that resemble *Huber* loss function:

$$g_h = \begin{cases} \eta^2, & |\eta| \leq \tau \\ 2\tau|\eta| - \tau^2, & |\eta| > \tau \end{cases} \quad (19)$$

This is interesting from the point of view of the theory of sparse reconstructions and CS literature, where the use of Huber function as a substitute for L1 norm has been often advocated, especially from the point of view of robustness.^{43,44}

We have encouraging first results with this structured sparsity approach in microwave tomography. Fig. 7 shows the reconstruction results for 3 simple objects from the Fresnel database. We used a configuration with 8 transmitting and 36 receiving antennas, equally spaced on a circle within azimuthal plane. The figure shows only the real parts of the permittivity profile. The two reference methods are recent approaches from electromagnetics community: multiplicative smoothing (MS)⁴¹ and value picking (VP).⁴² The electromagnetic solvers used in the

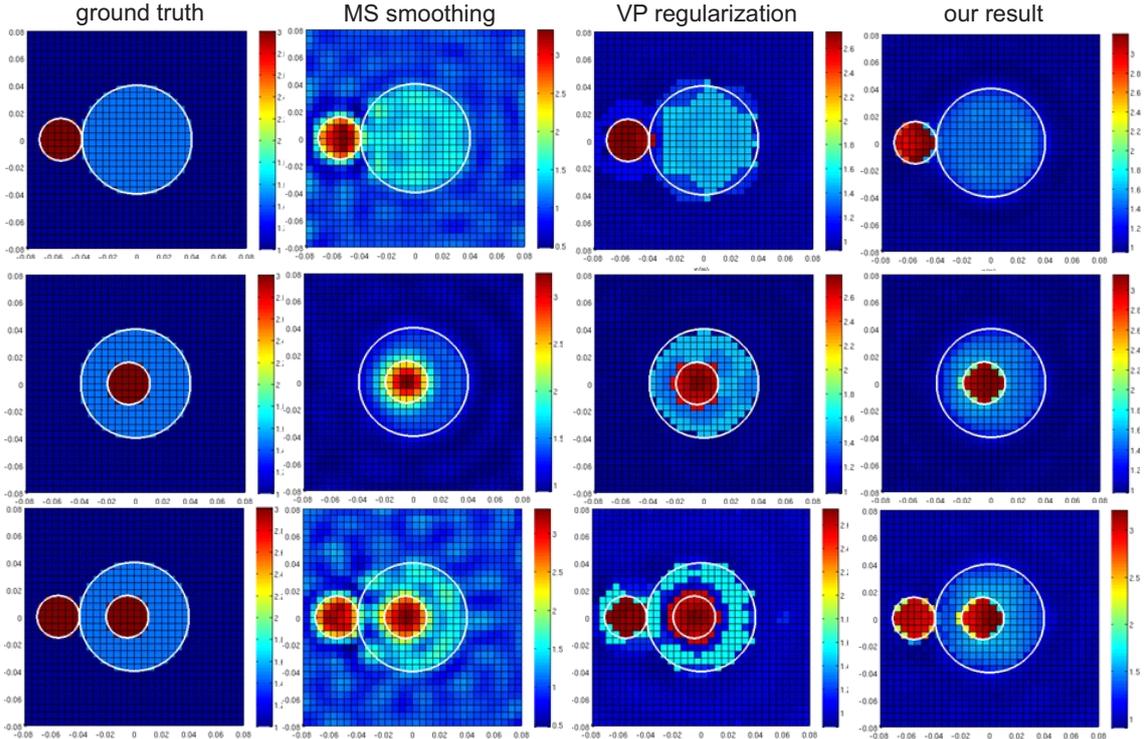


Figure 7. Real parts of the reconstructed permittivity profiles of three objects from the Fresnel database. Left to right: ground truth profiles and the results of MS,⁴¹ value picking (VP)⁴² and our structured sparsity approach.

forward and the inverse problems are detailed in these references. We used in our experiments the same software, provided by the authors, and for our method we replaced only the regularization part by the structured sparsity prior $J(\epsilon)$ in (18).

Our results in Fig. 7 are clearly better and sharper than those of MS. The VP method yields very sharp boundaries but delineation between different regions and the values of the permittivities are not always exact. Our approach does in this way already better, even though the transitions are somewhat less sharp than in the case of VP.

4. CONCLUSION

We presented here two types of structured sparsity approaches for tomographic reconstructions: imposing a hidden MRF model on the labels of the transform coefficients (in MRI) and imposing a continuous model on the observable field (in microwave tomography). We tried to use a unified notation making the parallels between the two approaches.

The hidden MRF formulation led to our new LaSB (Lattice Split Bregman) algorithm for MRI, motivated by the related LaMP.¹⁷ The results demonstrated clear advantages over the conventional approach which treats the coefficients of the sparsifying transform independently: much higher signal-to-noise ratio reached and in much less iterations. The present formulation is somewhat ad-hoc and further effort is needed to formalize it and study its properties. The presented results motivate strongly further efforts in this direction.

In the second considered problem - microwave tomography - we offered an alternative view on structured sparsity in tomographic reconstructions. The structured sparsity prior is formulated as energy of a continuous MRF model imposed on the observable field (of complex permittivities). The results demonstrate excellent performance of the Huber loss function as a discontinuity adaptive potential in this application. While this approach builds on the classical theory of MRF regularization, we think that its interpretation as structured sparsity can be of interest for further developments in sparsity and CS communities.

ACKNOWLEDGMENTS

The authors wish to thank Prof. Dr. Karel Deblaere from the Ghent University Hospital for providing MRI data sets and encouraging this work and to Prof. Ann Franchois and Prof. Daniel De Zutter from the Electromagnetics group at Ghent University for stimulating discussions and providing the reference methods (MS and VP regularized solvers) in microwave tomography. We are also grateful to Dr. Jurgen De Zaeytjij for sharing his electromagnetic solver code. This work is supported by FWO project “Advanced Quantitative Tomographic Image Reconstruction”.

REFERENCES

- [1] Candés, E., Romberg, J., and Tao, T., “Robust uncertainty principles: Exact frequency reconstruction from highly incomplete frequency information,” *IEEE Trans. Inform. Theory* **52**, 489–509 (2006).
- [2] Donoho, D., “Compressed sensing,” *IEEE Trans. Inform. Theory* **52**, 1289–1306 (2006).
- [3] Baraniuk, R., “Compressive sensing,” *IEEE Signal Processing Magazine* **24**, 118–121 (2007).
- [4] Lustig, M., Donoho, D., and Pauly, J., “Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging,” *Magnetic Resonance in Medicine* **58**, 1182–1195 (2007).
- [5] Lustig, M., Donoho, D., Santos, J., and Pauly, J., “Compressed sensing MRI,” *IEEE Signal Processing Magazine* **25**(2), 72–82 (2008).
- [6] Tang, J., Thériault Lauzier, P., and Chen, G.-H., “Dose reduction using prior image constrained compressed sensing (DR-PICCS),” in [*Proceedings of SPIE Medical Imaging*], **7961** (2011).
- [7] Nikolić, M., Tang, G., and Nehorai, A., “3D electromagnetic imaging using compressive sensing,” in [*IEEE Sensor Array and Signal Processing Workshop*], (2010).
- [8] Yuan, M. and Lin, Y., “Model selection and estimation in regression with grouped variables,” *Journal of The Royal Statistical Society Series B*, **68**(1), 49–67 (2006).
- [9] Huang, J. and Zhang, T., “The benefit of group sparsity,” *Annals of Statistics* **38**(4), 1978–2004 (2010).
- [10] Lv, X., Bi, G., and Wan, C., “The group Lasso for stable recovery of block-sparse signal representations,” *IEEE Trans. Signal Process.* **59**(4), 1371–1382 (2011).
- [11] Eldar, Y. and Mishali, M., “Robust recovery of signals from a structured union of subspaces,” *IEEE Trans. Information Theory*. **55**(11), 5302–5316 (2009).
- [12] Stojnic, M., Parvaresh, F., and Hassibi, B., “On the reconstruction of block-sparse signals with an optimal number of measurements,” *IEEE Trans. Signal Proc.* **57**(8), 3075–3085 (2009).
- [13] He, L. and Carin, L., “Exploiting structure in wavelet-based Bayesian compressive sensing,” *IEEE Trans. Signal Processing* **57**(9), 3488–3497 (2009).
- [14] Huang, J., Zhang, T., and Metaxas, D., “Learning with structured sparsity,” in [*Tech Report arXiv:0903.3002*], (2009).
- [15] Baraniuk, R., Cevher, V., Duarte, M., and Hegde, C., “Model-based compressive sensing,” *IEEE Trans. Information Theory* **56**, 1982–2001 (2010).
- [16] Trzasko, J. and Manduca, A., “Group Sparse Reconstruction of Vector-Valued Images,” in [*Proceedings ISMRM*], (2011).
- [17] Cevher, V., Duarte, M., Hegde, C., and Baraniuk, R., “Sparse signal recovery using Markov random fields,” in [*Proceedings of the Workshop on Neural Information Processing Systems (NIPS)*], (2008).
- [18] Cevher, V., Indyk, P., Carin, L., and Baraniuk, R., “Sparse signal recovery and acquisition with graphical models,” *IEEE Signal Processing Magazine* **27**(6), 92–103 (2010).
- [19] Goldstein, T. and Osher, S., “The split Bregman method for L1 regularized problems,” *SIAM Journal on Imaging Sciences* **2**(2), 323–343 (2009).
- [20] Aelterman, J., Quang Luong, H., Goossens, B., Pižurica, A., and Philips, W., “Augmented Lagrangian based reconstruction of non-uniformly sub-Nyquist sampled MRI data,” *Signal Processing* (2011). In press.
- [21] Needell, D. and Tropp, J. A., “Cosamp: Iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis* **26**(3), 301–321 (2009).
- [22] Hale, E., Yin, W., and Zhang, Y., “A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing,” in [*Technical Report TR07-07, Rice University, CAM Dept.*], (2007).

- [23] Candés, E., Demanet, L., Donoho, D., and Ying, L., “Fast Discrete Curvelet Transforms,” *Multiscale modeling and simulation* **5**(3), 861–899 (2006).
- [24] Easley, G., Labate, D., and Lim, W., “Sparse Directional Image Representations using the Discrete Shearlet Transform,” *Applied and Computational Harmonic Analysis* **25**, 25–46 (2008).
- [25] Chambolle, A., “An algorithm for total variation minimization and applications,” *J. of Math. Imaging and Vision* **20**, 89–97 (2000).
- [26] Rudin, L., Osher, S., and Fatemi, E., “Nonlinear total variation based noise removal algorithms,” *Physica D*, **60**, 259–268 (1992).
- [27] He, L., Chang, T.-C., Osher, S., Fang, T., and Speier, P., “MR image reconstruction from undersampled data by using the iterative refinement procedure,” *Proc. Appl. Math. and Mech.* **7**, 1011207–1011208, DOI 10.1002/pamm.200700776 (2007).
- [28] Huang, J., Zhang, T., Li, H., and Metaxas, D., “Composite splitting algorithms for convex optimization,” *Computer Vision and Image Understanding* (2011). In press.
- [29] Tropp, J. and Gilbert, A., “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Transactions on Information Theory* **53**, 4655–4666 (2007).
- [30] Yin, W., Osher, S., Goldfarb, D., and Darbon, J., “Bregman iterative algorithms for l_1 -minimization with applications to compressed sensing,” *SIAM Journal Imaging Sciences* **1**(1), 143–168 (2008).
- [31] Daubechies, I., Defrise, M., and De Mol, C., “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Communications on Pure and Applied Mathematics* **57**, 1413–1457 (2004).
- [32] Blumensath, T. and Davies, M., “Iterative thresholding for sparse approximation,” *The Journal of Fourier Analysis and Applications* **14**, 629–654 (2004).
- [33] Combettes, P. and Pesquet, J.-C., “A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery,” *IEEE Journal of Selected Topics in Signal Processing* **1**(4), 1–12 (2007).
- [34] Figueiredo, M., Nowak, R., and Wright, S., “Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems,” *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Convex Optimization Methods for Signal Processing* **1**, 586–598 (2007).
- [35] Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W., “An iterative regularization method for total variation-based image restoration,” *Multiscale Model. Simul.* **4**, 460–489 (2005).
- [36] Li, S., [*Markov Random Field Modeling in Computer Vision*], Springer-Verlag (1995).
- [37] Pižurica, A., Philips, W., Lemahieu, I., and Acheroy, M., “A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising,” *IEEE Trans. Image Process.* **11**(5), 545–557 (2002).
- [38] Pižurica, A. and Philips, W., “Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising,” *IEEE Trans. Image Process.* **15**(3), 654–665 (2006).
- [39] De Zaeytijd, J., Franchois, A., Eyraud, C., and Geffrin, J.-M., “Full-wave three-dimensional microwave imaging with a regularized Gauss-Newton method - Theory and experiment,” *IEEE Trans. Antennas and Propagation* **55**(11), 3279–3292 (2007).
- [40] De Zaeytijd, J., [*On the 3D Electromagnetic Quantitative Inverse Scattering Problem: Algorithms and Regularization*], PhD thesis, Ghent University (2009).
- [41] De Zaeytijd, J. and Franchois, A., “Three-dimensional quantitative microwave imaging from measured data with multiplicative smoothing and value picking regularization,” *Inverse Problems* **25**, doi:10.1088/0266-5611/25/2/024004 (2009).
- [42] De Zaeytijd, J., Franchois, A., and Geffrin, J.-M., “A new value picking regularization strategy - application to the 3-D electromagnetic inverse scattering problem,” *IEEE Trans. Antennas and Propagation* **57**(4), 1133–1149 (2005).
- [43] Li, B. and Yu, Q., “Robust and sparse bridge regression,” *Statistics and its Interface* **2**, 481–491 (2009).
- [44] Rosset, S. and Zhu, J., “Discussion of “Least Angle Regression” by Efron, Hastie, Johnstone and Tibshirani,” *Annals of Statistics* **32**, 469–475 (2004).