Smoothly Clipped Absolute Deviation (SCAD) Regularization for Compressed Sensing MRI Using an Augmented Lagrangian Scheme

A. Mehranian, H. Saligheh Rad, M. R. Ay, A. Rahmim, *Senior Member, IEEE*, and H. Zaidi[†], *Senior Member, IEEE*

***Abstract*– Compressed sensing (CS) in magnetic resonance imaging (MRI) enables the reconstruction of MR images from highly undersampled *k*-spaces, and thus substantial reduction of data acquisition time. In this context, edge-preserving and sparsity-promoting regularizers are used to exploit the prior knowledge that MR images are sparse or compressible in a given transform domain and thus to regulate the solution space. In this study, we introduce a new regularization scheme by iterative linearization of the non-convex clipped absolute deviation (SCAD) function in an augmented Lagrangian framework. The performance of the proposed regularization, which turned out to be an iteratively weighted total variation (TV) regularization, was evaluated using 2D phantom simulations and 3D retrospective undersampling of clinical MRI data by different sampling trajectories. It was demonstrated that the proposed regularization technique substantially outperforms conventional TV regularization, especially at reduced sampling rates.**

I. Introduction

MAGNETIC resonance imaging (MRI) is one of the leading cross-sectional imaging techniques used in clinical practice offering a great flexibility in terms of soft tissue contrast. MRI often suffers from slow data acquisition and long scanning time. Fast scanning is of particular importance in dynamic and cardiovascular imaging where whole-organ coverage within a short time is required. Beside ultra-fast imaging sequences [1], emerging trends focus on parallel MRI (pMRI) and partial Fourier measurements through *k*-space undersampling [2]. However, undersampling inevitably violates the Nyquist sampling criterion, whereby the reconstructed images exhibit aliasing artifacts and reduced signal-to-noise ratio (SNR). In fact, pMRI techniques, such as SMASH [3], GRAPPA [4] and SENSE [5], partially collect data from an array of receiver radiofrequency (RF) coils and hence reduce the number of Fourier encoding steps. In the reconstruction process, they exploit the redundant information contained in the spatial sensitivity of the coil array to avoid aliasing. SMASH and GRAPPA estimate the missing *k*-space data directly from coil sensitivity maps, while SENSE restores the original images by solving an underdetermined linear system of equations incorporating sensitivity maps. However, the inverse problems raised in pMRI are ill-posed and unstable due to *k*-space undersampling [5] and instability arising from the correlation of sensitivity maps [6].

Regularization and incorporation of some prior knowledge about the image solution in the reconstruction process is an efficient way to stabilize the problem and in fact to penalize unsatisfactory solutions. Many regularization methods have been studied in this context. The truncated singular value decomposition (SVD) technique attempts to stabilize the solution by truncating small singular values with the assumption that noise amplification is associated with small singular values of the solution [7, 8]. Tikhonov regularization suppresses noise based on the assumption that sharp transitions in image space are probably due to noise and hence favors smooth images [6, 7]. Both regularizations are based on L_2 norm minimization. Recent developments in compressed sensing (CS) have introduced sparsity regularization techniques, which have gained significant attention in MR image reconstruction from highly undersampled *k*-spaces. In fact, CS-MRI exploits the prior knowledge that MR images are sparse or weakly sparse (compressible) in spatial and/or temporal domains [9-11] or in a given transform domain such as wavelets, discrete gradients [10] and learned dictionaries [12, 13].

In this context, total variation (TV) regularization has been widely used to exploit and promote the sparsity of the solution [14-16]. TV regularization is based on the L_1 norm and makes use of discrete gradients as a sparsifying transform. It has been shown that this regularization outperforms the L_2 based regularization in CS-MRI [17]. In this study, we introduce the non-convex smoothly clipped absolute deviation (SCAD) regularization [18] for CS-MRI in order to improve the performance of TV regularizations. CS-MRI reconstruction is formulated as an optimization problem and solved using an augmented Lagrangian (AL) method, which is an efficient algorithm for solving large-scale problems encountered in CS-(p)MRI [19]. By linearization of the SCAD function in this framework, we then derive an iteratively weighted TV regularization scheme.

This work was supported by the Swiss National Science Foundation under grant SNSF 31003A-135576, the Indo-Swiss Joint Research Programme ISJRP 138866, and Geneva Cancer League.

A. Mehranian is with Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland (e-mail: mehrani1@etu.unige.ch).

H. Saligheh Rad and M. R. Ay are with the Department of Medical Physics and Biomedical Engineering, and Research Center for Molecular and Cellular Imaging, Tehran University of Medical Sciences, Tehran, Iran (e-mails: h-salighehrad@tums.ac.ir and mohammadreza_ay@sina.tums.ac.ir).

A. Rahmim is with the Department of Radiology, School of Medicine, and Department of Electrical & Computer Engineering, Johns Hopkins University, Baltimore, USA (e-mail: arahmim1@jhmi.edu).

[†]H. Zaidi is with the Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland, Geneva, Neuroscience Center, Geneva University, Geneva, Switzerland, and Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, Netherlands (e-mail: habib.zaidi@hcuge.ch).

II. Problem Formulation

For the standard CS-MRI, we formulate the following CS acquisition model:

$$y = \Phi\mathcal{F}x + n, \tag{1}$$

where $y \in \mathbb{C}^M$ is the undersampled k-space of the underlying MR image, $x \in \mathbb{R}^N$, contaminated with additive noise $n \in \mathbb{C}^M$. $\mathcal{F} \in \mathbb{C}^{N\times N}$ is a unitary Fourier basis through which x is being sensed and $\Phi \in \mathbb{R}^{M\times N}$ is a sampling matrix that compresses data to $M < N$ samples. The matrix $A = \Phi\mathcal{F}$ is often referred to as sensing or Fourier encoding matrix. The direct reconstruction of x from y (by zero-filling the missing data and then taking its inverse Fourier transform) results in aliasing artifacts attributed to the ill-conditioning nature of matrix A. As a result, regularization is required to regulate the solution space according to a priori knowledge. The solution is therefore obtained by the following optimization problem

$$\hat{x} = \text{argmin}_x \tfrac{1}{2}\|\Phi\mathcal{F}x - y\|^2 + R(x), \tag{2}$$

where the first term enforces data consistency and the second, known as regularizer, enforces data regularity. In the CS-MRI context, sparse L_1-based regularizers have been widely used because the L_1 norm is a convex and sparsity promoting norm. As such, the resulting problem is amenable to optimization using convex programming. These regularizers are of the form $R(x) = \lambda\|Dx\|_1 = \lambda\sum_{i=1}^{N}|[Dx]_i|$, where $\lambda > 0$ is a regularization parameter controlling the balance between regularization and data-consistency and D is a sparsifying transform such as discrete wavelet, cosine or gradient transform. The CS approach makes it possible to accurately reconstruct the image solution of problem (1), provided that i) the underlying image has a sparse representation in the domain of the transform D, i.e. most of the decomposition coefficients are zero, while few of them have a large magnitude, ii) the sensing matrix A should be sufficiently incoherent with the sparse transform D, thereby the aliasing artifacts arising from k-space undersampling would be incoherent (noise like) in the domain of D [10].

III. Proposed Approach

The sparsity of an image solution induced by L_1-based regularizers can be increased by introducing a non-convex potential function, ψ_λ, as follows:

$$R(x) = \sum_{i=1}^{N}\psi_\lambda(|[Dx]_i|), \tag{3}$$

where ψ_λ assigns a higher penalty on the coefficients of small magnitude so that they are being shrunk towards zero.

In this study, the non-convex SCAD potential function was applied for CS-MRI regularization. The most widely used SCAD function, which has been successfully applied for linear regression with variable selection [18], is defined as:

$$\psi_\lambda(|t|) = \begin{cases} \lambda|t| & |t| \le \lambda \\ (-|t|^2 + 2a\lambda|t| - \lambda^2)/2(a-1), & \lambda < |t| \le a\lambda \\ (1+a)\lambda^2/2 & |t| > a\lambda \end{cases} \tag{4}$$

where $a > 2$. This potential function corresponds to a quadratic spline with knots at λ and $a\lambda$. Fan and Li [18] suggested to use $a = 3.7$ based on simulations and Bayesian statistical arguments.

A 3D discrete gradient was employed as a sparsifying transform. It was defined as $D = [D^h, D^v, D^a] \in \mathbb{R}^{3N\times N}$, a derivative matrix composed of directional first-order finite difference matrices (horizontal, vertical and axial) with periodic boundary conditions. The magnitude of the gradient at voxel i is given by:

$$|[Dx]_i| = \sqrt{[D^h x]_i^2 + [D^v x]_i^2 + [D^a x]_i^2}\,, \quad [Dx]_i \in \mathbb{R}^3.$$

The summation over the magnitude of the gradient at all voxels defines the TV regularizer, which is known to be edge-preserving in image processing and sparsity-promoting in compressed sensing. However, it sometimes results in stair-casing artifacts, which are artificial flat regions in the image domain.

To solve the problem (2) with a SCAD-based regularizer, we follow the augmented Lagrangian (AL) method originally developed for constrained optimization problems [20]. The AL method, also known as the method of multipliers [21], not only allows for optimization of non-continuously differentiable regularizers such as L_1-based regularizer, but also through a variable splitting technique one can define auxiliary constraint variables and decompose the original optimization problem to simpler sub-problems [22]. Hence, we define the auxiliary variable $\theta = Dx$ and cast the problem (2), with the regularizer defined by equations (3) and (4), into the following constrained problem:

$$\min_{x,\theta}\left\{\Gamma(x,\theta) \triangleq \tfrac{1}{2}\|\Phi\mathcal{F}x - y\|^2 + \sum_{i=1}^{N}\psi_\lambda(|\theta_i|)\right\}, \quad \text{subject to } \theta = Dx \tag{5}$$

The augmented Lagrangian for this problem is defined as:

$$\mathcal{L}(x,\theta,\gamma) = \Gamma(x,\theta) - \gamma^T(\theta - Dx) + \tfrac{\rho}{2}\|\theta - Dx\|^2, \tag{6}$$

where $\gamma \in \mathbb{R}^{3N}$ and $\rho > 0$ are respectively the Lagrange multiplier and the penalty parameter associated with the equality constraint $\theta = Dx$. The AL method aims at finding a saddle point (x^*, θ^*) minimizing $\mathcal{L}(x,\theta,\gamma)$. The classical approach to solve Eq. (6) alternates between a joint-minimization and an update step as follows:

$$(x^{k+1}, \theta^{k+1}) = \text{argmin}_{x,\theta}\,\mathcal{L}(x,\theta,\gamma^k), \tag{7}$$

$$\gamma^{k+1} = \gamma^k - \rho(\theta^{k+1} - Dx^{k+1}), \tag{8}$$

As joint-minimization in Eq. (7) is not trivial, an alternating minimization with respect to a given variable while keeping the other constant can be followed. Using this approach, referred to as alternating direction method of multipliers (ADMM) [23], the optimization algorithm of Eq. (7) reads:

$$x^{k+1} = \text{argmin}_x\,\tfrac{1}{2}\|\Phi\mathcal{F}x - y\|^2 - \gamma_k^T(\theta^k - Dx) + \tfrac{\rho}{2}\|\theta^k - Dx\|^2\,, \tag{9}$$

$$\theta^{k+1} = \text{argmin}_\theta\,\textstyle\sum_{i=1}^{N}\psi_\lambda(|\theta_i|) - \gamma_k^T(\theta - Dx^{k+1}) + \frac{\rho}{2}\|\theta - Dx^{k+1}\|^2\,, \tag{10}$$

In the following, we solve the above sub-problems, individually.

A. Minimization with respect to x

The minimization in Eq. (9) is achieved by taking the derivative of the objective of the problem with respect to x and equating it to zero, thereby one arrives to the following normal equations:

$$(\mathcal{F}^H\Phi^T\Phi\mathcal{F}+\rho D^TD)x^{k+1}=\mathcal{F}^H\Phi^Ty+D^T(\rho\theta^k-\gamma^k), \quad (11)$$

where $(\cdot)^H$ denotes the Hermitian transpose and $\Phi^T\Phi\in\mathbb{R}^{N\times N}$ is a diagonal matrix with zeros and ones on the diagonal entries. To solve this problem, one need to invert the matrix $\mathcal{F}^H\Phi^T\Phi\mathcal{F}+\rho D^TD$. For periodic boundary conditions, matrix D has a block-circulant structure and its directional derivatives can be achieved by circular convolutions with two-element kernels. Therefore, D can be efficiently diagonalized using 3D discrete Fourier transform (DFT), i.e. $D=\mathcal{F}^H\Lambda\mathcal{F}$, where Λ is a diagonal complex matrix containing the DFT coefficients of the convolution kernels of D. Hence, one obtains $D^TD=\mathcal{F}^H|\Lambda|^2\mathcal{F}$, where $|\Lambda|^2\in\mathbb{R}^{N\times N}$ is the modulus of Λ, also the eigenvalue matrix of D. With this diagonalization, one can solve Eq. (9) as follows:

$$x^{k+1}=\mathcal{F}^H(\Phi^T\Phi+\rho|\Lambda|^2)^{-1}\mathcal{F}\left(\mathcal{F}^H\Phi^Ty+D^T(\rho\theta^k-\gamma^k)\right) \quad (12)$$

B. Minimization with respect to θ

The SCAD potential function is non-convex; thereby the problem (10) is not tractable to minimization. Hence, one can utilize an optimization transfer technique to iteratively surrogate this function by a convex function. Fan and Li [18] proposed a local quadratic approximation to this function near the point θ^k as follows:

$$Q(|\theta_i|,|\theta_i^k|)=\psi_\lambda(|\theta_i^k|)+\frac{1}{2}\frac{\psi_\lambda'(|\theta_i^k|)}{|\theta_i^k|}\left(|\theta_i|^2-|\theta_i^k|^2\right), \quad (13)$$

where the first derivative of the SCAD function is given by:

$$\psi_\lambda'(|\theta_i|)=\begin{cases}\lambda & |\theta_i|\le\lambda\\ \max(0,a\lambda-|\theta_i|)/(a-1) & |\theta_i|>\lambda\end{cases}, \quad (14)$$

The quadratic surrogate in Eq. (13) is however undefined at points $\theta_i^k=0$. The denominator can be conditioned to $|\theta_i^k|+\epsilon$, where ϵ is a predefined perturbation parameter [24]. As erroneous ϵ_s potentially degrade the sparsity of the solution as well as the convergence rate of the optimization algorithm, Zou and Li [25] proposed the linear local approximation of the SCAD function near the point θ_i^k. As a result, the following convex surrogate is obtained:

$$L(|\theta_i|,|\theta_i^k|)=\psi_\lambda(|\theta_i^k|)+\psi_\lambda'(|\theta_i^k|)(|\theta_i|-|\theta_i^k|). \quad (15)$$

Fig. 1 compares the SCAD function with its quadratic and linear convex surrogates. Note that the linear surrogate is a non-smooth function and is similar to a scaled or weighted L_1-norm. Given the superiority of SCAD linearization, we adopted this convex surrogate and derived a closed-form solution to the problem (10). By dropping the terms independent of θ_i in Eq. (15), completing the square in Eq. (10) and defining the intermediate variable $\tilde{\theta}=Dx^{k+1}+\gamma^k/\rho$, we can rewrite the problem (10) as follows:

$$\theta^{k+1}=\text{argmin}_\theta\sum_{i=1}^N\psi_\lambda'(|\theta_i^k|)|\theta_i|+\frac{\rho}{2}\|\tilde{\theta}-\theta\|^2, \quad (16)$$

As the terms in the above optimization problem are separable, we obtain the solution by the following component-wise soft-thresholding estimator according to theorem (1) in [26]:

$$\theta_i^{k+1}=\begin{cases}0 & |\tilde{\theta}_i|\le w_i^k/\rho\\ \tilde{\theta}_i-w_i^k\,\text{sign}(\tilde{\theta}_i)/\rho, & |\tilde{\theta}_i|>w_i^k/\rho\end{cases}, \quad (17)$$

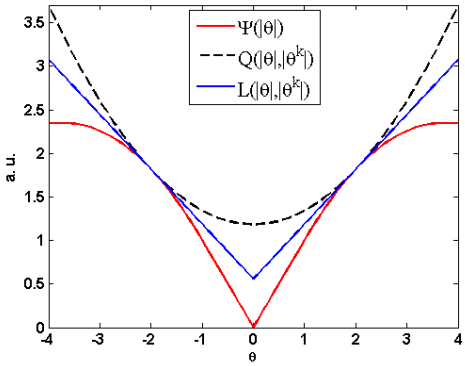

Fig. 1. The non-convex SCAD potential function (ψ) together with its convex quadratic (Q) and linear (L) surrogates ($\theta^k=2,\ a=3.7,\ \lambda=1$).

where $w_i^k=\psi_\lambda'(|\theta_i^k|)$ are iteratively derived weighting factors that promote or suppress the thresholding of the decomposition coefficients and $|\tilde{\theta}_i|=([\tilde{\theta}^h]_i^2+[\tilde{\theta}^v]_i^2+[\tilde{\theta}^a]_i^2)^{1/2}$. It worth noting that in the adopted framework, the problem (10) for quadratic approximation of SCAD has the following close-form solution:

$$\theta^{k+1}=\rho(W+\rho I)^{-1}\tilde{\theta},\ W=\text{diag}\left\{\frac{\psi_\lambda'(|\theta_1^k|)}{|\theta_1^k|+\epsilon},\dots,\frac{\psi_\lambda'(|\theta_N^k|)}{|\theta_N^k|+\epsilon}\right\},$$

where I is an identity matrix. Our preliminary results (not shown) showed that the linear approximation of SCAD outperforms its quadratic counterpart mainly due to i) its non-smooth nature and thus higher sparsity promotion and ii) the high dependence of quadratic approximation to its perturbation parameter, as noticed in [25].

As a result of SCAD function linearization, the regularizer $\sum_{i=1}^N\psi_\lambda'(|\theta_i^k|)|\theta_i|$ for $\theta_i=[Dx]_i$ in fact behaves as an iteratively-weighted TV regularizer with improved performance (see Results section). Note that by setting the weights $w_i^k=1$, the proposed regularization reduces to the conventional TV regularization. To this end, *Algorithm 1* summarizes the proposed SCAD-ADMM algorithm for CS-MRI. A global convergence is declared when the relative difference of x^{k+1} and x^k falls below a tolerance (η).

ALGORITHM 1: SCAD-ADMM

Choose ρ,λ,η, initialize $\theta^0,\gamma^0=0$ and pre-compute $|\Lambda|^2$.

While $(\|x^{k+1}-x^k\|/\|x^k\|)>\eta$ **do**

1. Compute x^{k+1} according to Eq. (12).
2. Define the intermediate variable $\tilde{\theta}=Dx^{k+1}+\gamma^k/\rho$.
3. Compute the weights $w^k=\psi_\lambda'(|\theta^k|)$ using Eq. (14).
4. Compute θ^{k+1} by weighted soft-thresholding of $\tilde{\theta}$ using Eq (17) and the weights w^k.
5. Update Lagrange multiplier according to Eq. (8).

Output: x^{k+1}.

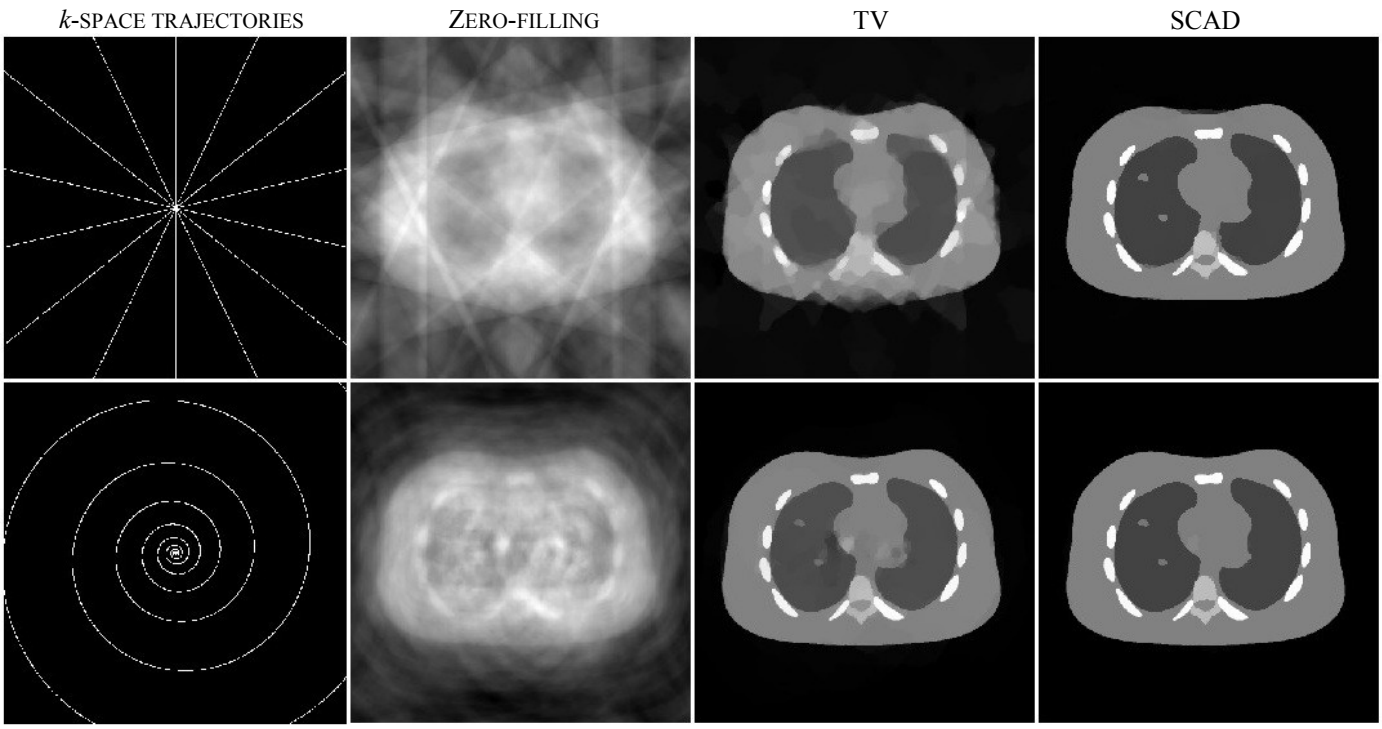

Fig. 2. Reconstruction of the XCAT phantom through zero-filling, TV-ADMM and SCAD-ADMM algorithms from *k*-spaces sampled using 7 equally spaced radial Fourier trajectories (Top) and a variable-density spiral trajectory (Bottom), respectively, corresponding to 98.37 and 98.48% *k*-space undersampling.

TABLE I

SNR (dB) PERFORMANCE OF THE STUDIED ALGORITHMS IN CS-MRI WITH RESPECT TO FULLY SAMPLED (REFERENCE) IMAGES.

DATASET (METHOD)	ZERO-FILLING	TV-ADMM	SCAD-ADMM
XCAT (RADIAL)	7.37	15.36	24.82
XCAT (SPIRAL)	7.19	21.58	30.31
BRAIN	12.67	18.41	20.19
MRA	11.08	16.62	18.12

IV. RESULTS

Several simulations and retrospective *k*-space undersampling using clinical datasets were performed to evaluate the performance of the proposed SCAD regularization with TV regularization. Radial and spiral Fourier trajectories were used for retrospective undersampling of the (fully sampled) Cartesian *k*-spaces of phantoms and clinical datasets.

The first experiment was performed on the XCAT phantom for the recovery of the underlying image (with a resolution of 512×512) from its undersampled *k*-space with 7 equally spaced radial trajectories and a single-shot variable-density spiral trajectory respectively, corresponding to 98.37 and 98.48% *k*-space undersampling, respectively). In this experiment, 20 dB complex noise was added to the undersampled *k*-spaces. Fig. 2 compares the images reconstructed by zero-filling and the best case performance of TV and SCAD-ADMM algorithms, i.e. heuristically optimized for involving parameters ρ, λ and a. For the radial sampling results, the optimal parameters were set as $\rho = 0.04,\ \lambda = 0.1,\ a = 3.7$ for SCAD and $\rho = 0.02,\ \lambda = 0.1$ for TV. Similarly, in the spiral sampling, the parameters were set as $\rho = 0.005, \lambda = 0.02,\ a = 3.7$ for SCAD and $\rho = 0.008,\ \lambda = 0.01$ for TV. As can be seen in Fig. 2, the proposed regularization technique has efficiently recovered the true image particularly in the case of radial sampling in which it outperforms the TV regularization in restoring the non-convex structures such as vertebra. The quantitative evaluation of the algorithms in terms of SNR is presented in Table I. The SNR results also confirm that the weighting scheme significantly improves the performance of TV regularization. It should be noted that at sufficiently high sampling rate the L_1-based TV regularization can restore the underlying image as faithfully as the SCAD regularization. However, we purposefully lowered the sampling rate to evaluate the ability of algorithms in CS-MRI from highly undersampled datasets.

In the second experiment, we performed CS-MR image reconstructions of a simulated T1-weighted 3D brain MRI dataset (30 slices with resolution of 256×256) , 1 mm slice thickness, 3% noise and 20% intensity non-uniformity [27]. A stack of 26 radial projections was used for 88.71% *k*-space undersampling. Fig. 3 compares the image reconstructed by zero-filling, TV and SCAD regularizations. In this case, both TV and SCAD regularizations have efficiently suppressed aliasing artifacts and thus recovered the underlying image. However, the TV regularization resulted in a global

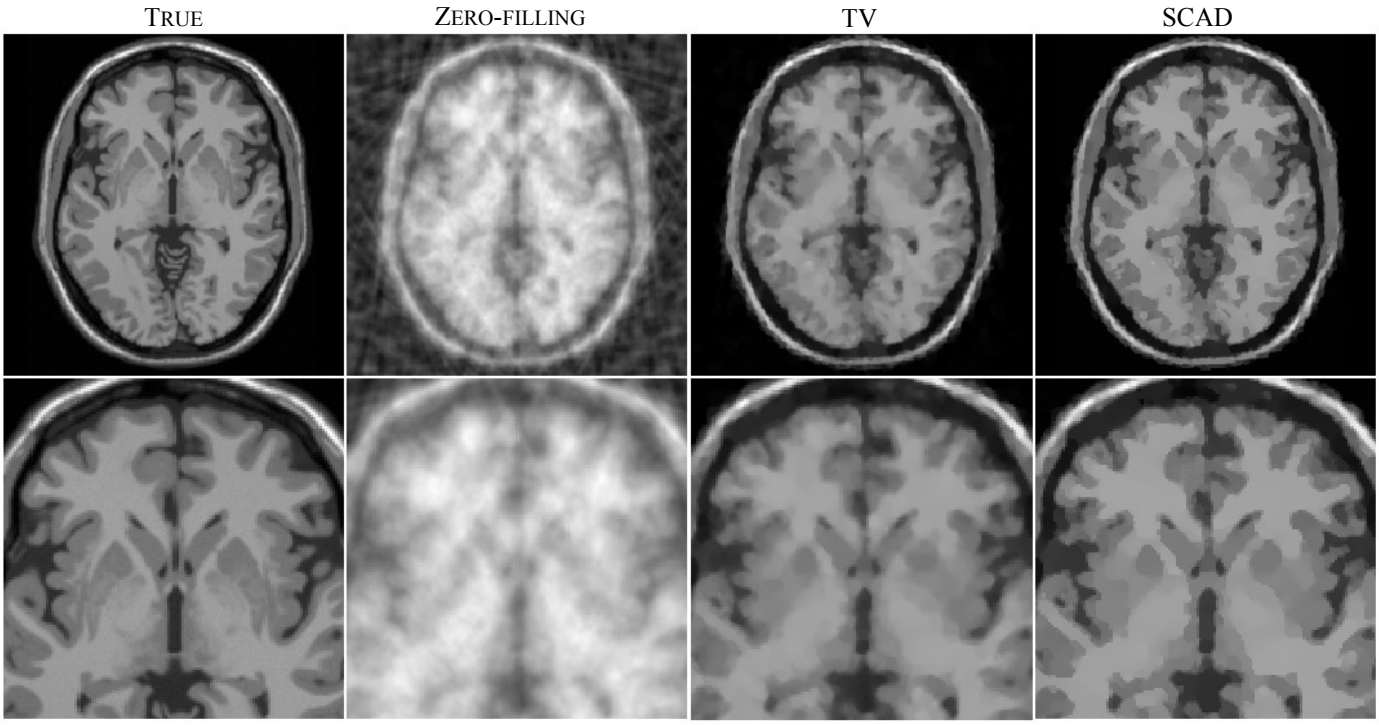

Fig. 3. Reconstruction of 3D Brain phantom through zero-filling, TV-ADMM and SCAD-ADMM algorithm using a 3D stack of 26 equally spaced radial Fourier trajectories (88.71% undersampling). The TV and SCAD images are shown with the same display window.

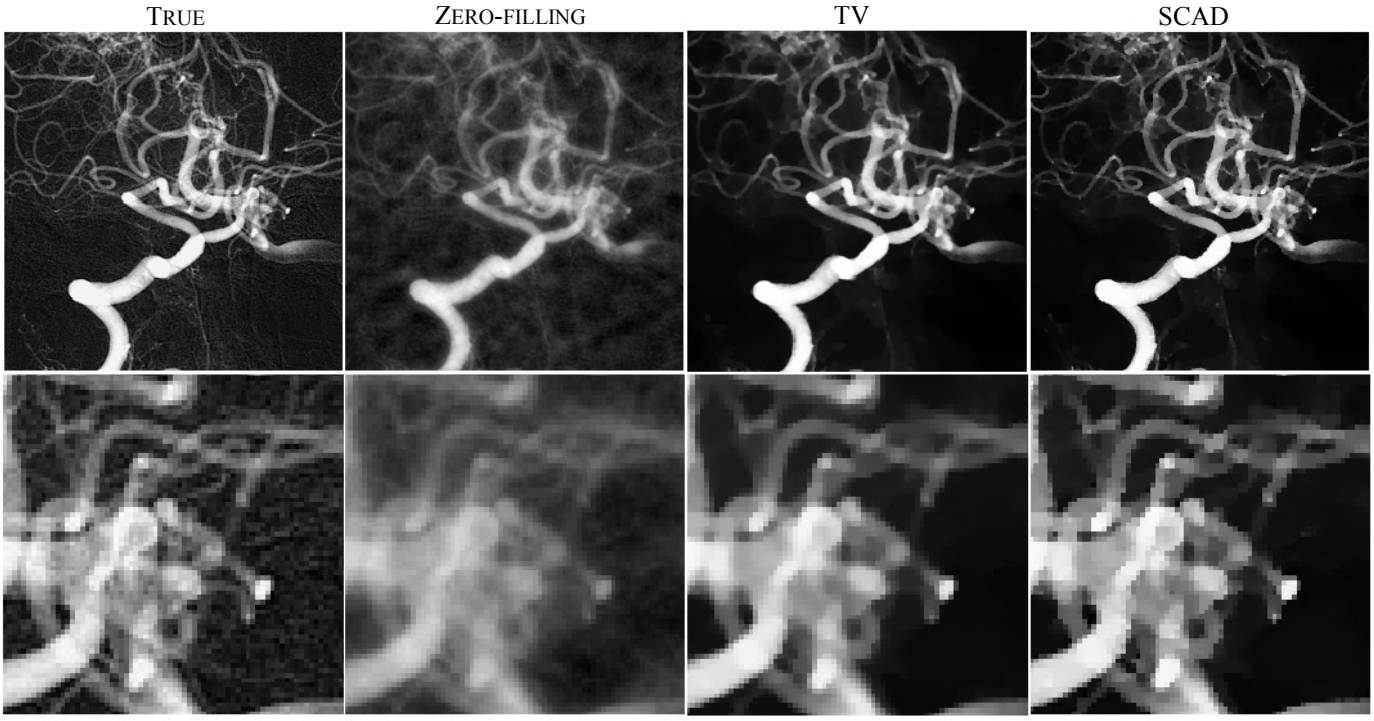

Fig. 4. Reconstruction of 3D brain MR angiogram through zero-filling, TV-ADMM and SCAD-ADMM algorithm using a 3D stack of a single-shot variable-density spiral trajectory (74.27% undersampling). The TV and SCAD images are shown with the same display window.

smoothing, whereas, as seen in zoomed-in images (Fig. 3), the proposed regularization method results in the preservation of edges through its iterative weighting of the TV regularizer. This is particularly noticeable for the brain's fissures with sharper anatomical edges. The SNR performance of the algorithm (Table I,) also reveals the improved signal to

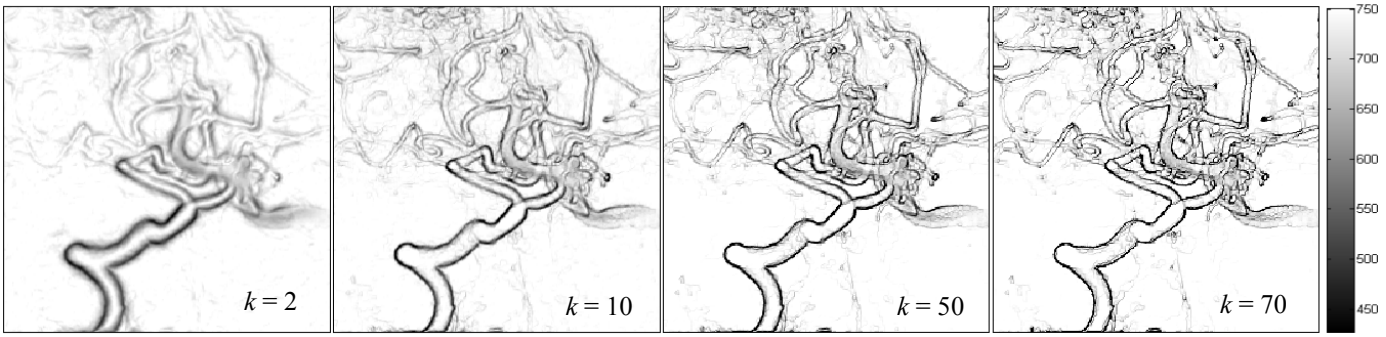

Fig. 5. Evolution of the weights (w^k) of the MRA frame shown in Fig. 4 with iteration number. The gray-color bar shows the dynamic range of the weights at iteration k = 70.

noise ratio and hence the reduction of aliasing artifacts achieved by the proposed regularization. In this reconstruction, the optimal parameters set were $\rho = 0.45, \lambda = 2.2, a = 10$ for SCAD and $\rho = 0.04, \lambda = 5$ for TV regularizations.

The performance of the algorithms was finally evaluated in the CS-MR reconstruction of an MR angiography (MRA) dataset in a patient with arterial bolus injection [28]. The dataset was synthetized from projection data collected for 3 frames per second, for a total of 10 seconds (31 collected frames) and the frames were linearly interpolated into 200 temporal frames. In this study, 30 time frames of this dataset (with a resolution of 256×256) were chosen and their 3D k-space was retrospectively undersampled using a stack of 2D single-shot variable-density spiral trajectories, yielding 74.27% undersampling. Fig. 4 shows reconstructed images using the studied algorithm with optimal parameters set as $\rho = 2, \lambda = 750, a = 50$ for SCAD and $\rho = 1.5, \lambda = 750$ for TV regularizations. The results show that the regularized reconstructions have substantially improved the quality of the images in comparison with zero-filling which is in fact an un-regularized reconstruction. At the same time, the regularized estimates are less noisy than the fully sampled image. On the other hand, close inspection of the zoomed-in images shows that the SCAD regularization enhances both image quality and contrast. The SNR results of the reconstruction algorithms are presented in Table I.

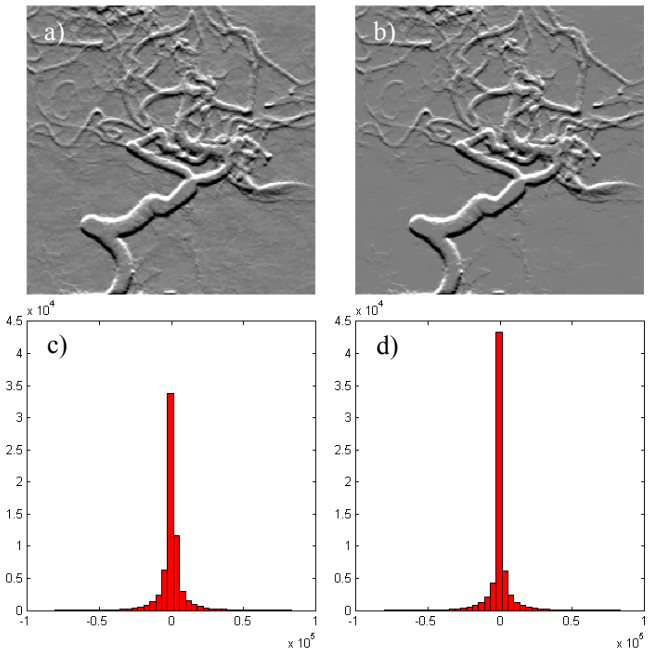

Fig. 6. The sparsity promotion of SCAD regularizer. (a-b) The vertical gradient field of the image frame shown in Fig. 4 (at iteration 10) thersholded by respectively soft and weighted-soft thresholding with the same regularization and penalty parameters. (c-d) The histograms of the images in shown in (a) and (b), respectively.

V. Discussion

A. Edge Preservation and Sparsity Promotion

The TV regularization and the linearized SCAD regularizations are in essence based on L_1 and reweighted L_1 norms. In general, the outperformance of SCAD regularization should be ascribed to the fact that the reweighted L_1 norm non-uniformly thresholds the gradient fields (decomposition coefficients) of the image estimate according to adaptively derived weighting factors (Step 4 in Algorithm 1). In fact, these weighting factors, on one hand, suppress the smoothing (thresholding) of edges and on the other hand, enforce the smoothing of the regions contaminated with noise and artifacts. Fig. 5 shows the weights associated with the image frame shown in Fig. 4 as a function of iteration number. It can be seen that as the iteration number is increased the emphasis on the preservation of edges (the wall of vessels) is increased by zero or close to zero weights and the suppression of in-between regions is continued by high-value weights. In fact, the end result of this procedure is the promotion of the sparsity of image estimates in the domain of the sparsifying transform. Figs. 6 (a–b) show the vertical gradient field (θ^v) of the MRA frame, shown in Fig. 4 at iteration number 10, thersholded respectively using soft-thresholding and weighted soft-thresholding with the same regularization and penalty parameters. The corresponding histograms of the images (20 bins) are shown in Figs. 6 (c–d). The results clearly show that the SCAD weighting scheme promotes the sparsity by zeroing or penalizing small value coefficients that appear as noise and incoherent (noisy-like) artifacts. This is also noticeable in the histograms where the frequency of the coefficients in close-to-zero bins has been reduced, while it has been increased in the zero-bin.

To enhance sparsity, Candes *et al* [29] proposed a reweighted L_1 norm by iterative linearization of a quasi-

convex logarithmic potential function. They demonstrated that unlike the L_1 norm, the resulting reweighted L_1 norm (with the weights $w_i^k = \lambda(|\theta_i^k| + \epsilon)^{-1}$, in our notation) provides a more "democratic penalization" by assigning higher penalties on small non-zero coefficients while encouraging the preservation of larger coefficients. In this sense, a reweighted L_1 norm regularization resembles an L_0 norm, which is an ideal sparsity-promoting, but intractable norm. Recently, Trzasko *et al* [30] proposed a homotopic L_0 norm approximation by gradually reducing the perturbation parameter ϵ in quasi-convex norms (e.g. the logarithmic function) to zero. It has been shown that the solution of L_0 penalized least squares problems, such as the one presented in Eq. (16) with an L_0 regularizer, can be achieved with a hard thresholding rule [31, 32], which thresholds only the coefficients lower than a threshold $\sqrt{2\lambda}$. As observed by Trzasko *et al*, the hard thresholding rule associated with L_0 regularization increases sparsity and offers strong edge retention in comparison with the soft thresholding associated with L_1 regularization. In comparison with the weighting scheme of Candes *et al* and in connection with homotopic L_0 approximations, the linearized SCAD regularization invokes a weighted soft-thresholding rule that in limit approaches hard thresholding rule. Fig. 7(a) compares the standard hard and soft-thresholding rules with the weighted soft-thresholding rule obtained from the linearization of the SCAD function (according to Eq. 14), for different values of the parameter a and for $\lambda = 1$. Similarly, Fig. 7(b) compares those standard rules with the weighted soft-thresholding rule by Candes's weighting scheme for different values of the parameter ϵ and for $\lambda = 1$. It was observed that for small values of a, the SCAD weight soft rule resembles hard thresholding, while for the small values of ϵ the Candes weight soft rule is at best between the standard hard and soft thresholding. On the other hand, for large values of a and ϵ, the SCAD and Candes rules respectively approach soft thresholding and an identity rule (which can be thought as a soft-thresholding with zero threshold). In fact, Fan and Li [18] proposed the SCAD potential function to improve the properties of L_1 and hard thresholding penalty functions (those approximating L_0 norm) in terms of unbiasedness, continuity and sparsity. This function avoids the bias of soft-thresholding (L_1 norm) on over-penalizing large coefficients and the discontinuity and thus instability of hard-thresholding (quasi L_0 norm) and at the same time, promotes the sparsity of the solution [18].

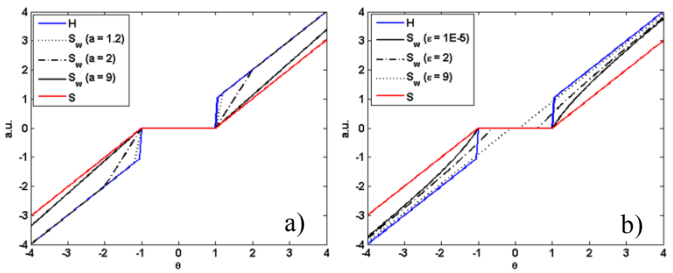

Fig. 7. Comparison of hard (H) and soft (S) thresholding rules with the soft-thresholding (S_w) rules weighted by (a) the linearization of SCAD function for different parameters a and (b) the Candes' approach for different parameters ϵ.

B. Computational Complexity and Parameter Selection

In this study, we solved the standard CS-MRI image reconstruction problem presented in Eq. (2) using an augmented Lagrangian method. Recently, Ramani *et al* [19] studied AL methods for pMRI and showed that this class of algorithms are computationally appealing in comparison with nonlinear conjugate gradient and monotone fast iterative soft-thresholding (MFISTA) algorithms. In the minimization of the AL function with respect to x, they solved Eq. (11), which also included sensitivity maps of array coils, using a few iterations of the conjugate-gradient algorithm. In contrast, we derived a closed-form solution for this equation in CS-MRI, which allowed speeding up the algorithm, particularly by the per-computation of the eigenvalue matrix of discrete gradient matrix D. The practical application of this analytic solution for pMRI remains to be addressed in future work.

All of our CS-MR experiments were implemented in MATLAB 2010a, running on a 12-core workstation with 2.40 GHz Intel Xeon processors and 32GB memory. In general, for retrospective reconstruction of a dataset 256×256×30, the SCAD and TV-ADMM algorithms required about 0.4 and 0.25 seconds per iteration in our implementation. A tolerance of $\eta = 5 \times 10^{-4}$ was used in Algorithm 1 to declare the convergence of the algorithms. For optimally tuned parameters, it was found that the TV-ADMM algorithm generally converges after a fewer number of iterations in comparison with the SCAD-ADMM. In the 3D Brain and MRA datasets, it converged after 32 and 43 iterations, respectively, while the SCAD converged after 93 and 72 iterations, respectively. This convergence behavior should be ascribed to the weighting scheme that SCAD regularization exploits. In fact, our results showed that the TV regularization has a comparatively better initial SNR improvement, however, after a sufficient number of iterations, it is not able to further improve the quality of the image estimate and hence it is stopped. Whereas, the SCAD algorithm showed a gradual improvement in SNR to a higher level, as summarized in Table I.

As mentioned, the algorithms were compared in their best case performance, that is, the involving parameters (the penalty parameter ρ, the regularization parameter λ and the SCAD's scale parameter a) were tuned to obtain the best qualitative and quantitative results. In general, it was observed that the parameter ρ has higher impact on the overall smoothing of image features compared with λ. In the SCAD regularization, the parameters λ and particularly a controlled the impact of weighting factors, and they showed a higher flexibility for selection than ρ. An alternative way to choose the pair parameters (a, λ) could be two-dimensional grids search using some criteria such as cross validation and L-curve methods [7], which calls for future investigations.

C. Future Prospects

For our experiments, we employed Cartesian approximations to radial and spiral k-space undersampling patterns to retrospectively perform a CS-MR acquisition model in Eq. (1). However, the extension of this model to true non-Cartesian data acquisitions is straightforward and the reconstruction can be done with regridding of data to a rectangular grid with

density compensation [14] or non-uniform fast Fourier transform (NUFFT) [33]. In the employed AL framework, the SCAD regularization can also be exploited for multi-scale sparsifying transforms such as orthogonal wavelets, a normalized tight frame of translation invariant wavelets or curvelet frames. Thereby, one can obtain a closed-form solution to the problem (11) using matrix inversion lemma (MIL) and the fact that $\Phi\Phi^T = I$ and for these transforms $DD^T = I$. Finally, in this framework, the linearization in Eq. (15) can be extended to derive a family of reweighted L_1 norm regularization from non-convex potential functions [34].

VI. Conclusion

In this study, we proposed a new regularization technique for compressed sensing MRI through the linearization of the non-convex SCAD potential function in the framework of augmented Lagrangian (AL) methods. Using variable splitting technique, the CS-MRI problem was formulated as a constrained optimization problem and solved efficiently in the AL framework. We exploited discrete gradients as a sparsifying transform and demonstrated that the linearized SCAD regularization is an iteratively weighted TV regularization with improved edge-preserving and sparsity-promoting properties. The performance of the algorithm was evaluated in phantom simulations and retrospective CS-MRI in clinical datasets and it was found that the proposed regularization technique substantially outperforms the conventional TV regularization.

References

[1] S. Plein, *et al.*, "Steady-state free precession magnetic resonance imaging of the heart: Comparison with segmented k-space gradient-echo imaging," *J Magn Reson Imaging,* vol. 14, pp. 230-236, 2001.

[2] D. J. Larkman and R. G. Nunes, "Parallel magnetic resonance imaging," *Phys. Med. Biol.,* vol. 52, pp. R15-55, 2007.

[3] D. K. Sodickson and W. J. Manning, "Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays," *Magn. Reson. Med.* , vol. 38, pp. 591–603, 1997.

[4] M. A. Griswold, *et al.*, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magn. Reson. Med.,* vol. 47, pp. 1202–10, 2002.

[5] K. P. Pruessmann, *et al.*, "SENSE: Sensitivity encoding for fast MRI," *Magn Reson Med,* vol. 42, pp. 952-962, 1999.

[6] F. H. Lin, *et al.*, "Parallel Imaging Reconstruction Using Automatic Regularization," *Magn Reson Med,* vol. 51, pp. 559-567, 2004.

[7] W. C. Karl, "Regularization in Image Restoration and Reconstruction," in *Handbook of Image and Video Processing*, A. Bovik, Ed., ed San Diego: Academic Press, 2000.

[8] W. E. Kyriakos, *et al.*, "Sensitivity profiles from an array of coils for encoding and reconstruction in parallel (SPACE RIP)," *Magn Reson Med,* vol. 44, pp. 301-308, 2000.

[9] S. J. M. Lustig M, Donoho D L and Pauly J M a "k-t SPARSE: high frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proc. Intl. Soc. Mag. Reson. Med.*, Seattle, WA, 2006.

[10] M. Lustig, *et al.*, "Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging," *Magn. Reson. Med.,* vol. 58, pp. 1182-1195, 2007.

[11] H. Jung, *et al.*, "k-t FOCUSS: A General Compressed Sensing Framework for High Resolution Dynamic MRI," *Magn Reson Med,* vol. 61, pp. 103-116, 2009.

[12] M. Hong, *et al.*, "Compressed sensing MRI with singular value decomposition-based sparsity basis," *Phys. Med. Biol.,* vol. 56, pp. 6311-6325, 2011.

[13] S. Ravishankar and Y. Bresler, "MR Image Reconstruction From Highly Undersampled k-Space Data by Dictionary Learning," *IEEE Trans Image Process,* vol. 30, pp. 1028-1041, 2011.

[14] K. T. Block, *et al.*, "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint," *Magn Reson Med,* vol. 57, pp. 1086-1098, 2007.

[15] L. Chen, *et al.*, "Myocardial perfusion MRI with an undersampled 3D stack-of-stars sequence," *Med Phys,* vol. 39, pp. 5204-5211, 2012.

[16] G. Adluru, *et al.*, "Temporally Constrained Reconstruction of Dynamic Cardiac Perfusion MRI," *Magn Reson Med,* vol. 57, pp. 1027-1036, 2007.

[17] J. X. Ji, *et al.*, "Compressed sensing parallel Magnetic Resonance Imaging," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, 2008, pp. 1671-1674.

[18] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Statistical Assoc.,* vol. 96, pp. 1348-1360, 2001.

[19] S. Ramani and J. A. Fessler, "Parallel MR Image Reconstruction Using Augmented Lagrangian Methods," *Medical Imaging, IEEE Transactions on,* vol. 30, pp. 694-706, 2011.

[20] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2th ed.: Springer, 2006.

[21] M. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*, R. Fletcher, Ed., ed New York: Academic Press 1969, pp. 283-298.

[22] M. V. Afonso, *et al.*, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans Image Process,* vol. 19, pp. 2345-56, Sep 2010.

[23] J. Eckstein and D. Bertsekas, "On the DouglasRachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming,* vol. 5, pp. 293-318, 1992.

[24] D. Hunter and R. Li, "Variable selection using mm algorithms," *Ann. Statist.* , vol. 33, pp. 1617-1642, 2005.

[25] H. Zou and R. Li, "One-step sparse estimates in nonconcave penalized likelihood models," *Ann. Stat.,* vol. 36, pp. 1509-1533, 2008.

[26] A. Antoniadis and J. Fan, "Regularization of wavelets approximations," *J. Am. Stat. Assoc.,* vol. 96, pp. 939-963, 2001.

[27] *BRAINWEB, http://www.bic.mni.mcgill.ca/brainweb/.*

[28] *MRI UNBOUND, http://www.ismrm.org/mri_unbound/simulated.htm*.

[29] E. J. Candès, *et al.*, "Enhancing Sparsity by Reweighted L1 Minimization," *J Fourier Anal Appl,* vol. 14, pp. 877-905, 2008.

[30] J. Trzasko and A. Manduca, "Highly Undersampled Magnetic Resonance Image Reconstruction via Homotopic L0 Minimization," *Medical Imaging, IEEE Transactions on,* vol. 28, pp. 106-121, 2009.

[31] T. Blumensath, *et al.*, "Iterative Hard Thresholding and L0 Regularisation," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, 2007, pp. III-877-III-880.

[32] S. J. Wright, *et al.*, "Sparse Reconstruction by Separable Approximation," *Signal Processing, IEEE Transactions on,* vol. 57, pp. 2479-2493, 2009.

[33] J. A. Fessler and B. P. Sutton, "Nonuniform fast Fourier transforms using min-max interpolation," *Signal Processing, IEEE Transactions on,* vol. 51, pp. 560-574, 2003.

[34] Y.-B. Zhao and A. D. Li, "Reweighted L1-Minimization For Sparse Solutions To Underdetermined Linear Systems," *SIAM Journal on Optimization,* vol. 22, pp. 1065-1088, 2012.