Magnetic Resonance Imaging 31 (2013) 1399–1411

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

journal homepage: www.mrijournal.com

Smoothly Clipped Absolute Deviation (SCAD) regularization for compressed sensing MRI Using an augmented Lagrangian scheme ☆

Abolfazl Mehranian [a], Hamidreza Saligheh Rad [b,c], Arman Rahmim [d], Mohammad Reza Ay [c,e], Habib Zaidi [a,f,g,*]

[a] Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland
[b] Quantitative MR Imaging and Spectroscopy Group, Research Center for Cellular and Molecular Imaging, Tehran University of Medical Sciences, Tehran P.O. Box 14185-615, Iran
[c] Department of Medical Physics and Biomedical Engineering, Tehran University of Medical Sciences, Tehran P.O. Box 14155-6447, Iran
[d] Department of Radiology, Johns Hopkins University, Baltimore, MD 21287, USA
[e] Medical Imaging Systems Group, Research Center for Molecular and Cellular Imaging, Tehran University of Medical Sciences, Tehran P.O. Box 14185-615, Iran
[f] Geneva Neuroscience Center, Geneva University, CH-1205 Geneva, Switzerland
[g] Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, Netherlands

ARTICLE INFO

Article history:
Received 15 February 2013
Revised 28 May 2013
Accepted 30 May 2013

Keywords:
Compressed sensing
Smoothly Clipped Absolute Deviation (SCAD)
Total variation
Augmented Lagrangian

ABSTRACT

Purpose: Compressed sensing (CS) provides a promising framework for MR image reconstruction from highly undersampled data, thus reducing data acquisition time. In this context, sparsity-promoting regularization techniques exploit the prior knowledge that MR images are sparse or compressible in a given transform domain. In this work, a new regularization technique was introduced by iterative linearization of the non-convex smoothly clipped absolute deviation (SCAD) norm with the aim of reducing the sampling rate even lower than it is required by the conventional l_1 norm while approaching an l_0 norm.

Materials and Methods: The CS-MR image reconstruction was formulated as an equality-constrained optimization problem using a variable splitting technique and solved using an augmented Lagrangian (AL) method developed to accelerate the optimization of constrained problems. The performance of the resulting SCAD-based algorithm was evaluated for discrete gradients and wavelet sparsifying transforms and compared with its l_1-based counterpart using phantom and clinical studies. The *k*-spaces of the datasets were retrospectively undersampled using different sampling trajectories. In the AL framework, the CS-MRI problem was decomposed into two simpler sub-problems, wherein the linearization of the SCAD norm resulted in an adaptively weighted soft thresholding rule with a sparsity enhancing effect.

Results: It was demonstrated that the proposed regularization technique adaptively assigns lower weights on the thresholding of gradient fields and wavelet coefficients, and as such, is more efficient in reducing aliasing artifacts arising from *k*-space undersampling, when compared to its l_1-based counterpart.

Conclusion: The SCAD regularization improves the performance of l_1-based regularization technique, especially at reduced sampling rates, and thus might be a good candidate for some applications in CS-MRI.

Published by Elsevier Inc.

1. Introduction

Magnetic resonance imaging (MRI) is one of the leading cross-sectional imaging modalities in clinical practice offering a great flexibility in representing the anato-functional characteristics of organs and soft tissues. However, MRI often suffers from long data acquisition time. Fast data acquisition is of particular importance for capturing temporal changes over whole organs in a short time. Beside ultra-fast imaging sequences [1], emerging trends focus on partial Fourier [2] and parallel MRI (pMRI) [3,4], which are based on the undersampling of *k*-space and estimation of missing data using the redundant information available in the acquired data or prior knowledge about the underlying image.

One of the issues of the data acquisition techniques employing *k*-space undersampling is reduced signal to noise ratio (SNR), since SNR is directly proportional to the number of phase-encoding steps or the amount of acquired data. Furthermore, at high undersampling rates or acceleration factors, the reconstructed images can exhibit residual aliasing artifacts that further degrade image quality. The

☆ Conflict of interest statement: The authors declare that they have no conflict of interest.

* Corresponding author. Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland. Tel.: +41 22 372 7258; fax: +41 22 372 7169.
E-mail address: habib.zaidi@hcuge.ch (H. Zaidi).

0730-725X/$ – see front matter. Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.mri.2013.05.010

reduced SNR and residual artifacts, in fact, arise from the ill-conditioning of the inverse problems encountered in this context [5]. Regularization and explicit incorporation of prior knowledge during reconstruction of MR images are efficient ways to improve the conditioning of the problem and thus to penalize unsatisfactory and noisy solutions. Several regularization schemes have been assessed in this context. Tikhonov regularization suppresses noise and artifacts by favoring smooth image solutions [6–8].The truncated singular value decomposition attempts to reduce noise by truncating small singular values on the assumption that noise amplification is associated with small singular values of solution [6,9]. Both regularizations are based on l_2 norm minimization and tend to blur the details and edges in the estimated image [6,10]. Recent developments in compressed sensing have introduced sparsity regularization techniques, which have garnered significant attention in MR reconstruction from highly undersampled *k*-spaces. In fact, CS-MRI reduces noise and aliasing artifacts by exploiting the prior knowledge that MR images are sparse or weakly sparse (compressible) in spatial and/or temporal domains [11,12], in a given transform domain such as wavelets, Fourier, discrete gradients [11,13] or in learned dictionaries [14,15]. By establishing a direct link between sampling and sparsity, CS theory provides an alternative sampling criterion to conventional Shannon–Nyquist theory [16,17]. According to this theory, it is possible to accurately recover the underlying signal or solution from the data acquired at sampling rates far below the Nyquist limit as long as i) it is sparse or has a sparse representation in a given transform domain and ii) the sampling pattern is random or such that the aliasing artifacts are incoherent (noisy-like) in that domain [17,18].

Sparsity regularization aims at finding a solution that has the sparsest representation in a given sparse transform domain. In this regard, the l_0 norm is an ideal regularizer (or prior), which counts the number of non-zero elements of the solution [19]. However, this non-convex prior results in an intractable and non-deterministic polynomial-time hard (NP-hard) optimization problem. For this reason, the l_1 norm has been widely used as a convex surrogate to the l_0 norm and has gained popularity in conjunction with wavelet [20,21] and discrete gradient transforms [22]. The latter is known as total variation (TV) regularization [23–26] and has been shown to outperform l_2-based regularizations in CS-(p)MRI [27,28]. The l_1-based regularizations; however, show a lower limit in the required sampling rate and hence in the maximum achievable acceleration rate [29]. In addition, the l_1 norm is known to be biased due to over-penalizing large sparse representation coefficients [30]. To further reduce the sampling rate and approach l_0 norm minimization, Candes et al. [30] proposed a reweighted l_1 norm minimization in which the sparsity induced by the l_1 norm is enhanced by the weighting factors that are derived from the current estimate of the underlying solution. This approach has been successfully applied in CS-(p)MRI [31–33]. Furthermore, non-convex priors homotopically approximating the l_0 norm have also been studied showing the improved performance of the resulting regularization techniques in the recovery of strictly sparse signals [19,34,35]. However, MR images are usually compressible rather than sparse, hence it is desirable to exploit the sparsity-promoting properties of both l_1 and l_0 norm minimizations [36]. To improve the properties of l_1 and pseudo l_0 norms in terms of unbiasedness, continuity and sparsity, Fan and Li [37] proposed a non-convex prior called smoothly clipped absolute deviation (SCAD) norm in the context of statistical variable selection. This norm has been designed to not excessively penalize large valued coefficients as in the l_1 norm and at the same time approaching an l_0 norm. Teixeira et al. [38] have previously studied the SCAD regularization for sparse signal recovery using a second-order cone optimization method. In this work, we employed, for the first time, the SCAD regularization with discrete gradients and wavelet transforms in the context of CS-MRI and solved the resulting problem using variable splitting and augmented Lagrangian (AL) methods. In the AL framework, the optimization problem is reduced to simpler sub-problems, leading to an improved convergence rate in comparison with state-of-the-art and general purpose optimization algorithms [39,40]. In this framework, the linearization of the SCAD norm resulted in a weighted soft thresholding rule that exploits the redundant information in image space to adaptively threshold the gradient fields and wavelet coefficients and to effectively reduce aliasing artifacts. In this study, we compared the performance of the proposed SCAD-based regularization with the conventional l_1-based approach using simulation and clinical studies, where *k*-spaces were retrospectively undersampled using different sampling patterns to demonstrate the potential application of the proposed method in CS-MR image reconstruction.

2. Materials and methods

2.1. Theory

For a single-coil CS-MRI, we formulate the following CS acquisition model:

$$\boldsymbol{y} = \Phi\mathcal{F}\boldsymbol{x} + \boldsymbol{n} \quad [1]$$

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

$$\hat{\boldsymbol{x}} = \mathrm{argim}_{\boldsymbol{x}}\frac{1}{2}\|\Phi\mathcal{F}\boldsymbol{x}-\boldsymbol{y}\|^2 + R(\boldsymbol{x}) \quad [2]$$

where the first term enforces data consistency and the second one, 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, thereby the resulting problem is amenable to optimization using convex programming. These regularizers are of the form $R(\boldsymbol{x}) = \lambda\|\boldsymbol{\Psi x}\|_1 = \lambda\sum_{i=1}^{N}|[\boldsymbol{\Psi x}]_i|$, where $\lambda > 0$ is a regularization parameter controlling the balance between regularization and data-consistency and $\boldsymbol{\Psi}$ is a sparsifying transform such as discrete wavelet, Fourier 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 $\boldsymbol{\Psi}$, i.e. most of the decomposition coefficients are zero, while few of them have a large magnitude, ii) the sensing matrix $\mathbf{A}$ should be sufficiently incoherent with the sparse transform $\boldsymbol{\Psi}$, thereby the aliasing artifacts arising from *k*-space undersampling would be incoherent (noise like) in the domain of $\boldsymbol{\Psi}$ [11,18].

2.2. Proposed method

The sparsity or compressibility of an image solution induced by l_1 based regularizers can be increased by introducing a non-convex potential function, ψ_λ, as follows:

$$R(\boldsymbol{x}) = \sum_{i=1}\psi_\lambda\left(|[\boldsymbol{\Psi x}]_i|\right) \quad [3]$$

where ψ_λ assigns a higher penalty on the coefficients of small magnitude, therefore, they are being shrunk towards zero. In this study, the non-convex SCAD potential function is applied for CS-MRI regularization. The SCAD function, which has been widely and successfully used for linear regression with variable selection [37], is defined as:

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

where $a > 2$. This potential function corresponds to a quadratic spline with knots at λ and $a\lambda$. Based on simulations and some Bayesian statistical arguments, Fan and Li [37] suggested $a = 3.7$.

In this study, 3D discrete gradient and 2D wavelet transforms were employed as sparsifying transforms. For discrete gradient, we define $\boldsymbol{\Psi} = [\boldsymbol{\Psi}^h, \boldsymbol{\Psi}^v, \boldsymbol{\Psi}^a] \in \mathbb{R}^{3N\times N}$, which is composed of directional first-order finite difference matrices (horizontal, vertical and axial) with periodic boundary conditions. By convention, we define the magnitude of the gradient at voxel i by $[\boldsymbol{\Psi}, \boldsymbol{\Psi x}]_i = \sqrt{\left[\boldsymbol{\Psi}^h \boldsymbol{x}\right]_i^2 + \left[\boldsymbol{\Psi}^v \boldsymbol{x}\right]_i^2 + \left[\boldsymbol{\Psi}^a \boldsymbol{x}\right]_i^2}$, $[\boldsymbol{\Psi x}]_i \in \mathbb{R}^3$. The summation over the magnitude of the gradient at all voxels in Eq. (3) defines an isotropic TV regularizer, which is known to be edge-preserving in image processing and sparsity-promoting in compressed sensing. However, this is known to sometimes result in stair-casing artifacts, which are artificial flat regions in the image domain. For wavelet transforms, we make use of Daubechies 7/9 biorthogonal (D7/9) wavelets, with four resolution levels in a normalized tight (Parseval) frame of translation-invariant wavelets, implemented by undecimated discrete wavelet transforms (UDWT) and a lifting scheme. In UDWT, the decimation (downsampling) is eliminated in favor of invariance to the shifts of an input image, thus avoiding the star-like artifacts usually induced by the standard decimated wavelet transform. Note that in the case of a tight frame, $\boldsymbol{\Psi} \mathbb{R}^{D\times N}$, which is called decomposition or forward wavelet transform, satisfies $\boldsymbol{\Psi}^T \boldsymbol{\Psi} = \boldsymbol{I}$, where $\boldsymbol{\Psi}^T$ is reconstruction or inverse wavelet transform and $\mathbf{I}$ is the identity matrix.

To solve the problem defined in Eq. (2) using SCAD-based regularizer, we follow the augmented Lagrangian (AL) method, which has been originally developed for constrained optimization problems [41]. The AL method, also known as the method of multipliers [42], allows for the optimization of non-continuously differentiable regularizers through a variable splitting technique, in which auxiliary constraint variables are defined and the original optimization problem is decomposed to simpler sub-problems [39]. Hence, we define the auxiliary variable $\boldsymbol{\theta} = \boldsymbol{\Psi x}$ and cast the problem (2), with the regularizer defined by Eqs. (3) and (4), into the following constrained problem:

$$\min_{\boldsymbol{x},\boldsymbol{\theta}} \left\{ \Gamma(\boldsymbol{x},\boldsymbol{\theta}) \triangleq \frac{1}{2}\|\Phi\mathcal{F}\boldsymbol{x} - \boldsymbol{y}\|^2 + \sum_{i=1}^{N} \lambda(|\boldsymbol{\theta}_i|) \right\}, \text{subject to } \boldsymbol{\theta} = \boldsymbol{\Psi x}. \quad [5]$$

The augmented Lagrangian for this problem is defined as:

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{\gamma}) = \Gamma(\boldsymbol{x},\boldsymbol{\theta}) - \boldsymbol{\gamma}^T(\boldsymbol{\theta} - \boldsymbol{\Psi x}) + \frac{\rho}{2}\|\boldsymbol{\theta} - \boldsymbol{\Psi x}\|^2. \quad [6]$$

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

$$\left(\boldsymbol{x}^{k+1}, \boldsymbol{\theta}^{k+1}\right) = \text{argmin}_{\boldsymbol{x},\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\gamma}^k\right). \quad [7]$$

$$\boldsymbol{\gamma}^{k+1} = \boldsymbol{\gamma}^k - \rho\left(\boldsymbol{\theta}^{k+1} - \boldsymbol{\Psi x}^{k+1}\right). \quad [8]$$

As joint-minimization in Eq. (7) is not trivial, an alternating minimization with respect to a given variable while fixing the other one can be followed. Using this approach, referred to as alternating direction method of multipliers (ADMM) [43], the optimization algorithm of Eq. (7) reads:

$$\boldsymbol{x}^{k+1} = \text{argmin}_{\boldsymbol{x}} \left\{ \frac{1}{2}\|\Phi\mathcal{F}\boldsymbol{x} - \boldsymbol{y}\|^2 - \boldsymbol{\gamma}_k^T\left(\boldsymbol{\theta}^k - \boldsymbol{\Psi x}\right) + \frac{\rho}{2}\|\boldsymbol{\theta}^k - \boldsymbol{\Psi x}\|^2 \right\} \quad [9]$$

$$\theta^{k+1} = \text{argmin}_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{N} \psi_\lambda(|\theta_i|) - \boldsymbol{\gamma}_k^T\left(\boldsymbol{x} - \boldsymbol{\Psi}^{k+1}\right) + \frac{\rho}{2}\|\theta - \boldsymbol{\Psi x}^{k+1}\|2 \right\}. \quad [10]$$

Recently, Ramani et al. [40] studied the ADMM method for pMRI and demonstrated its outperformance over nonlinear conjugate gradient algorithms. In this work, we followed this method and derived solutions for the involved sub-problems as follows.

2.2.1. Minimization with respect to $\boldsymbol{x}$

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

$$\left(\mathcal{F}^H\Phi^T\Phi\mathcal{F} + \rho\Psi^T\Psi\right)\boldsymbol{x}^{k+1} = \mathcal{F}^H\Phi^T\boldsymbol{y} + \Psi^T\left(\rho\boldsymbol{\theta}^k - \boldsymbol{\gamma}^k\right). \quad [11]$$

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 needs to invert the matrix $\mathbf{G} \triangleq \mathcal{F}^H\Phi^T\Phi\mathcal{F} + \rho\boldsymbol{\Psi}^T\boldsymbol{\Psi}$.

In the case of discrete gradients with periodic boundary conditions, the matrix $\boldsymbol{\Psi} \in \mathbb{R}^{3N\times N}$ as a block-circulant structure and its directional derivatives can be achieved by circular convolutions with two-element kernels. Therefore, $\boldsymbol{\Psi}$ can be efficiently diagonalized using 3D discrete Fourier transform (DFT) [44], i.e. $\boldsymbol{\Psi} = \mathcal{F}^H\boldsymbol{\Lambda}\mathcal{F}$, where $\boldsymbol{\Lambda}$ is a diagonal complex matrix containing the DFT coefficients of the convolution kernels of $\boldsymbol{\Psi}$. Hence, one obtains $\boldsymbol{\Psi}^T\boldsymbol{\Psi} = \mathcal{F}^H|\boldsymbol{\Lambda}|^2\mathcal{F}$, where $|\boldsymbol{\Lambda}|^2 \in \mathbb{R}^{N\times N}$ is the modulus of $\boldsymbol{\Lambda}$, also the eigenvalue matrix of $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$. With this diagonalization, the solution of problem (9) is given by:

$$\boldsymbol{x}^{k+1} = \mathcal{F}^H\left(\Phi^T\Phi + \rho|\Lambda^2|\right)^{-1}\mathcal{F}\left(\mathcal{F}^H\Phi^T\boldsymbol{y} + \Psi^T\left(\rho\boldsymbol{\theta}^k - \boldsymbol{\gamma}^k\right)\right). \quad [12]$$

It should be noted that in the case of non-Cartesian MR data, the Fourier encoding matrix $\Phi\mathcal{F}$ is not diagonalizable using discrete Fourier transform. However, the solution can be obtained by regridding of data to Cartesian k-space or the use of iterative techniques such as the conjugate gradient algorithm for estimating the inverse of $\mathbf{G}$. Recently, Akcakaya et al. [45] proposed another approach to approximate the matrix $\Phi^T\Phi$, which is not diagonal in the case of non-Cartesian data acquisition, by a diagonal matrix which gives an approximate closed-form solution to this sub-problem.

In the case of wavelet frames, the inversion of matrix $\mathbf{G}$ can be obtained by Sherman–Morrison–Woodbury matrix inversion

formula and exploiting the fact that $\Psi^T\Psi = \Phi\Phi^T = \mathbf{I}$, as follows:

$$\boldsymbol{G}^{-1} = \frac{1}{\rho}\left(\mathbf{I} - \mathcal{F}^H\Phi^T\left(\Phi\Phi^T + \rho\mathrm{I}\right)^{-1}\Phi\mathcal{F}\right) = \frac{1}{\rho}\left(\mathbf{I} - \frac{1}{1+\rho}\mathcal{F}^H\Phi^T\Phi\right). \quad [13]$$

By doing some algebra and knowing that $\mathcal{F}\mathcal{F}^H = I$, one can show that the solution of Eq. (11) for wavelet transforms reads:

$$\boldsymbol{x}^{k+1} = \Psi^T\left(\boldsymbol{\theta}^k - \frac{1}{\rho}\boldsymbol{\gamma}^k\right) + \frac{1}{1+\rho}\mathcal{F}^H\Phi^T\left(\boldsymbol{y} - \Phi\mathcal{F}\Psi^T\left(\boldsymbol{\theta}^k - \frac{1}{\rho}\boldsymbol{\gamma}^k\right)\right). \quad [14]$$

2.2.2. Minimization with respect to $\boldsymbol{\theta}$

The SCAD potential function is non-convex; thereby the problem (10) might have multiple local minima. The minimization of non-convex problems often depends on the initial estimate and the choice of the optimization algorithm. Global search techniques, such as simulated annealing, can guarantee convergence to a global minimizer but they are impractical for routine use in image reconstruction. Hence, one can utilize an optimization transfer technique to iteratively surrogate the non-convex function by a convex function, which is amenable to optimization. Fan and Li [37] proposed a local quadratic approximation to the SCAD function near the point θ_i^k as follows:

$$Q\left(|\theta_i|, |\theta_i^k|\right) = \psi_\lambda\left(|\theta_i^k|\right) + \frac{1}{2}\frac{\psi_\lambda'\left(|\theta_i^k|\right)}{|\theta_i^k|\left(|\theta_i|^2 - |\theta_i^k|^2\right)} \quad [15]$$

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

$$\psi_\lambda'(|\theta_i|) = \begin{cases} \lambda & |\theta_i| \le \lambda \\ \max(0, a\lambda - |\theta_i|)/(a-1) & |\theta_i| > \lambda \end{cases} \quad [16]$$

The quadratic surrogate in Eq. (15) is, however, undefined at points $\theta_i^k = 0$. The denominator can be conditioned to $|\theta_i^k| + \varepsilon$, where ε is a predefined perturbation parameter [46]. Since an ε erroneous potentially degrades the sparsity of the solution as well as the convergence rate of the optimization algorithm, Zou and Li [47] proposed the linear local approximation of the SCAD function near the point θ_i^k. As a result, the following convex surrogate is obtained:

$$L\left(|\theta_i|, |\theta_i^k|\right) = \psi_\lambda\left(|\theta_i^k|\right) + \psi_\lambda'\left(|\theta_i^k|\right)\left(|\theta_i| - |\theta_i^k|\right). \quad [17]$$

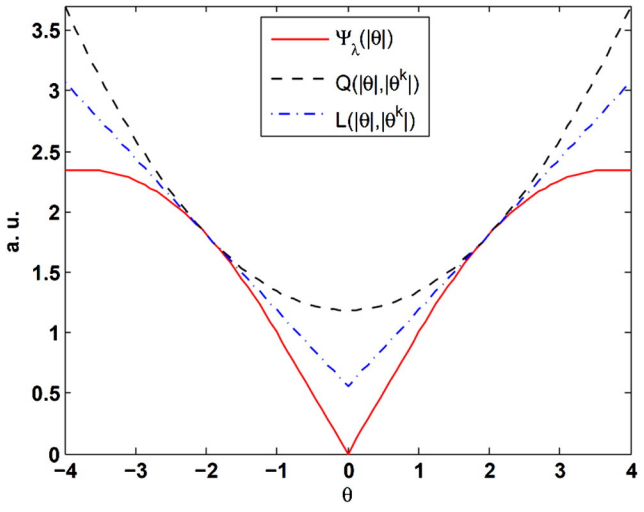

Fig. 1. The non-convex SCAD potential function (ψ) together with its convex quadratic (Q) and linear (L) surrogates ($\theta^k = 2$, $a = 3.7$, $\lambda = 1$).

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 linearization of SCAD, we adopted this convex surrogate and derived a closed-form solution to the problem (10). By dropping the terms independent of θ_i in (17), completing the square in (10) and defining the intermediate variable $\tilde{\boldsymbol{\theta}} = \Psi\boldsymbol{x}^{k+1} + \boldsymbol{\gamma}^k/\rho$, we can rewrite the problem (10) as follows:

$$\boldsymbol{\theta}^{k+1} = \mathrm{argmin}_{\boldsymbol{\theta}} \sum_{i=1}^{N} \psi_\lambda'\left(|\theta_i^k|\right)|\theta_i| + \frac{\rho}{2}||\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}||^2. \quad [18]$$

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 [48]:

$$\theta_i^{k+1} = \begin{cases} 0 & |\tilde{\theta}_i| \le w_i^k/\rho \\ \tilde{\theta}_i - w_i^k \mathrm{sign}\left(\tilde{\theta}_i\right)/\rho, & |\tilde{\theta}_i| > w_i^k/\rho. \end{cases} \quad [19]$$

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 in the case of discrete gradients, by convention, $|\tilde{\theta}_i| = \left(|\tilde{\theta}_i^h|^2 + |\tilde{\theta}_i^v|^2 + \left(|\tilde{\theta}_i^a|^2\right)\right)^{1/2}$, where $|c|^2 = c^*c$ is the modulus of the complex variable c. Note that in the case of discrete gradients, w_i^k, is isotropically used for the gradient fields at a voxel in three directions, that is, the weighting factors are concatenated to obtain $\mathbf{w}^k \in \mathbb{R}^{3N}$ and then are fed into Eq. (19).

As a result of linearization of the SCAD function, the regularizer $R(\boldsymbol{\theta}) = \sum_{i=1}^{N} \psi_\lambda'(|\theta_i^k|)|\theta_i|$ for $\theta_i = [\Psi\boldsymbol{x}]_i$ in fact behaves as an iteratively weighted l_1-based regularizer with improved performance (see Results section). Note that by setting the weights $w_i^k = \lambda$, the proposed regularization reduces to the conventional l_1-based regularization. To this end, Algorithm 1 summarizes the proposed SCAD-ADMM algorithm for CS-MRI. A global convergence is declared when the relative difference between $\boldsymbol{x}^{k+1}$ and $\mathbf{x}^k$ falls below a tolerance (η).

Algorithm 1
SCAD-ADMM.

Choose ρ, λ, η and initialize $\boldsymbol{\theta}^0$, $\boldsymbol{\gamma}^0 = 0$.
Pre-compute $|\boldsymbol{\Lambda}|^2$ if Ψ is discrete gradients.
While $(||\boldsymbol{x}^{k+1} - \boldsymbol{x}^k||/||\boldsymbol{x}^k||) > \eta$ **do**
1. Compute $\mathbf{x}^{k+1}$ according to Eq. (12) if Ψ is discrete gradients or Eq. (14) if Ψ is a wavelet transform.
2. Define the intermediate variable $\tilde{\theta} = \Psi\boldsymbol{x}^{k+1} + \gamma^k/\rho$.
3. Compute the weights $\boldsymbol{w}^k = \psi_\lambda'(|\boldsymbol{\theta}^k|)$ using Eq. (16).
4. Compute $\boldsymbol{\theta}^{k+1}$ by weighted soft-thresholding of $\tilde{\theta}$ using Eq. (19) and the weights $\mathbf{w}^k$.
5. Update Lagrange multipliers according to Eq. (8).

3. Experiments and evaluations

Several simulations and retrospective k-space undersampling in clinical datasets were performed to evaluate the performance of the proposed SCAD regularization with l_1-based regularization. The variable density random sampling and Cartesian approximations of the radial and spiral Fourier trajectories were used for retrospective undersampling of the (fully sampled) Cartesian k-spaces of

phantoms and clinical datasets. To demonstrate the performance of SCAD regularization for highly undersampled MR reconstruction, we first performed a set of simulated noisy data generated from the anthropomorphic XCAT phantom. In this experiment, the *k*-space of a 512 × 512 slice of the XCAT phantom was sampled by 8 equally spaced radial trajectories as well as a single-shot variable-density spiral trajectory, respectively corresponding to 98.33% and 98.30% undersampling, with 20 dB complex noise added to *k*-spaces. For the evaluation of the proposed SCAD-ADMM algorithm with 3D discrete gradients, an MR angiography (MRA) dataset in a patient with arterial bolus injection was obtained from Ref. [49]. The dataset has been synthesized from projection data collected for 3 frames per second for a total of 10 s (31 collected frames) and linearly interpolated into 200 temporal frames. In this study, 30 time frames of this dataset (with 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 78.7% undersampling. The performance of the algorithm was further evaluated with 2D translation-invariant wavelets using two brain datasets. A transverse slice of a 3D brain T1-weighted MRI dataset (of the size 181 × 217 × 181) was obtained from the BrainWeb database (McGill University, Montreal, QC, Canada) [50], which has been simulated for 1 mm slice thickness, 3% noise and 20% intensity non-uniformity. The image slice was zero-padded to 256 × 256 pixels and its *k*-space was retrospectively undersampled by a variable-density random sampling pattern with 85% undersampling. Finally in clinical patient study, a Dixon MRI dataset was acquired on a Philips Ingenuity TF PET–MRI scanner (Philips Healthcare, Cleveland, OH). The MRI subsystem of this dual-modality imaging system is equipped with the Achieva 3.0 T X-series MRI system. A whole body scan was acquired using a 16-channel receiver coil and a 3D multi-echo 2-Point FFE Dixon (mDixon) technique with parameters: TR = 5.7 ms, TE1/TE2 = 1.45/2.6, flip angle = 10º and slice thickness of 2 mm, matrix size of 480 × 480 × 880 and in plane resolution of 0.67 mm × 0.67 mm. From this sequence, in-phase, out-of-phase, fat and water (IP/OP/F/W) images are reconstructed. For the comparison of SCAD- and l_1-based wavelet regularizations, a representative image slice of OP image was zero-padded to 512 × 512 pixels and its *k*-space was retrospectively undersampled by the Cartesian approximation of radial trajectory with 83% undersampling. Note that the *k*-spaces of the studied datasets were obtained by forward Fourier transform of the image slices.

All of our CS-MR reconstructions were performed in MATLAB 2010a, running on a 12-core workstation with 2.40 GHz Intel Xeon processors and 32 GB memory. The improvement of image quality was objectively evaluated using peak signal to noise ratio (PSNR) and mean structural similarity (MSSIM) index, [51] between the ground truth fully sampled image, $\boldsymbol{x}^*$ and the images reconstructed by l_1- and SCAD-based regularizations, $\boldsymbol{x}$. The PSNR is defined as:

$$\mathrm{PSNR}(\boldsymbol{x},\boldsymbol{x}^*) = 20\log_{10}\left(\frac{\max_{1\le i\le N}|\boldsymbol{x}_i^*|}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}|\boldsymbol{x}_i^*-\boldsymbol{x}_i|^2}}\right)$$

MSSIM, which evaluates both nonstructural (e.g., intensity) and structural (e.g., residual streaking artifacts) deviations of an image from its reference image, is given by:

$$\mathrm{SSIM}(\boldsymbol{x},\boldsymbol{x}^*) = \frac{(2\mu_{\boldsymbol{x}}\mu_{\boldsymbol{x}}{}^* + C_1)(2\sigma_{\boldsymbol{x}\boldsymbol{x}}{}^* + C_2)}{(\mu_{\boldsymbol{x}}{}^2 + \mu_{\boldsymbol{x}}{}^{*2} + C_1)(\sigma_{\boldsymbol{x}}{}^2 + \sigma_{\boldsymbol{x}}{}^{*2} + C_2)}$$

$$\mathrm{MSSIM}(\boldsymbol{x},\boldsymbol{x}^*) = \frac{1}{L}\sum_{l=1}^{L}\mathrm{SSIM}(\boldsymbol{x}_l,\boldsymbol{x}_l^*)$$

where the mean intensity, μ and standard deviation σ, of $\boldsymbol{x}$ and $\boldsymbol{x}^*$ and their correlation coefficient $\sigma_{\boldsymbol{x}\boldsymbol{x}^*}$, are calculated over L local image patches. The constants $C_1 = (K_1D)^2$ and $C_2 = (K_2D)^2$ are introduced to avoid instability issues, where D is the dynamic range of pixel values. $K_1 = 0.01$ and $K_2 = 0.03$ according to Ref. [51]. Based on this metric, a perfect score, i.e. MSSIM ($\boldsymbol{x}$, $\boldsymbol{x}^*$) = 1 is achieved only when the image $\boldsymbol{x}$ is identical to the ground truth image $\boldsymbol{x}^*$. In addition, the reconstructed images were qualitatively compared through visual comparison and intensity profiles. In all reconstructions, a tolerance of $\eta = 1\times10^{-4}$ was used in Algorithm 1 to declare the convergence of the algorithms.

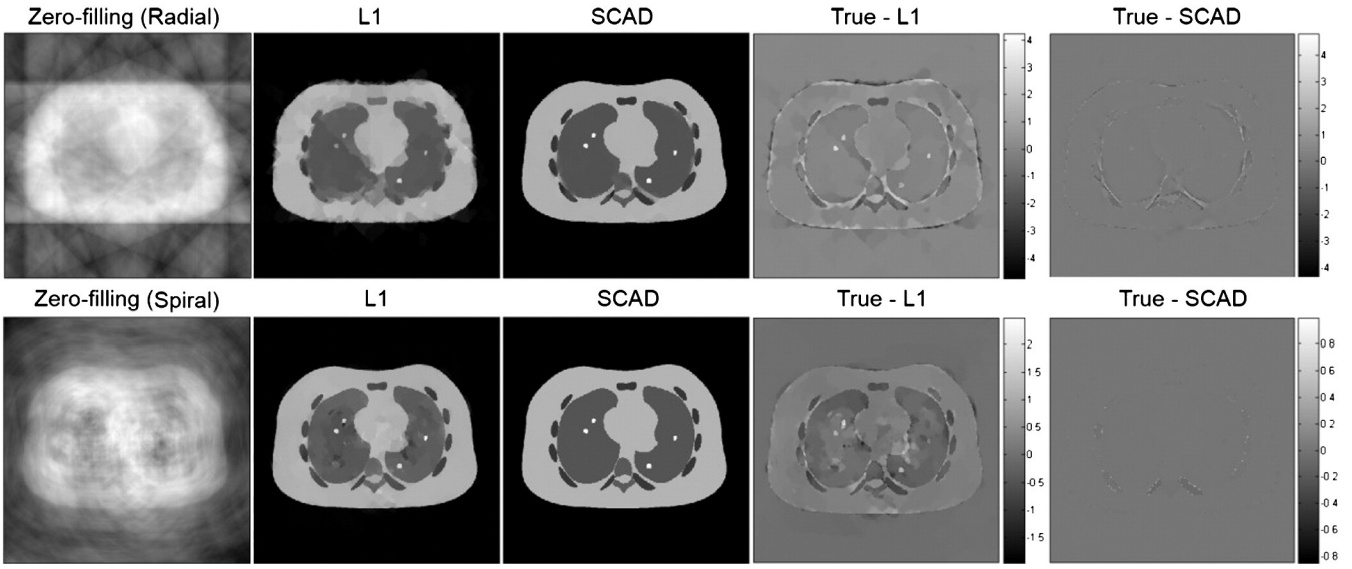

Fig. 2. Reconstruction of the XCAT phantom through zero-filling, gradient-based l_1 and SCAD ADMM algorithms from the *k*-spaces sampled by 8 equally spaced radial Fourier trajectories (top) and a variable-density spiral trajectory (bottom), respectively, corresponding to 98.33% and 98.30% undersampling. The difference images show the deviations of the reconstructed images from the true fully sampled image.

Table 1
Summary of the peak signal-to-noise ratio (PSNR) and mean structural similarity (MSSIM) performance of the studied algorithms in CS-MRI with respect to fully sampled (reference) images.

Dataset	PSNR (dB)			SSIM			Iterations | CPU time/Iter. (s)			
	Zero-Filling	L1	SCAD	Zero-Filling	L1	SCAD	L1		SCAD	
XCAT (Radial)	17.66	26.03	35.89	0.21	0.83	0.98	2130	0.06	2862	0.08
XCAT (Spiral)	14.21	31.41	52.66	0.16	0.84	0.99	2489	0.07	2745	0.08
MR Angiogram	24.65	28.32	28.60	0.64	0.69	0.70	94	0.83	138	0.97
BrainWeb	14.84	18.52	20.36	0.39	0.59	0.62	335	0.49	354	0.66
Brain mDixion	23.76	27.91	29.13	0.31	0.98	0.99	96	1.84	102	2.12

The number of iterations and the computation (CPU) time per iteration (in seconds) are also reported.

4. Results

4.1. CS-MR Image Reconstructions

Fig. 2 shows the results of image reconstruction in XCAT phantom for radial and single-shot variable density spiral trajectories, in first and second rows, respectively, and compares the images reconstructed by zero-filing, l_1- and SCAD-based ADMM algorithms with discrete gradient sparsifying transform. This figure also shows the difference images between true and reconstructed images. As can be seen, the proposed regularization technique has efficiently recovered the true image and outperformed its TV counterpart in both

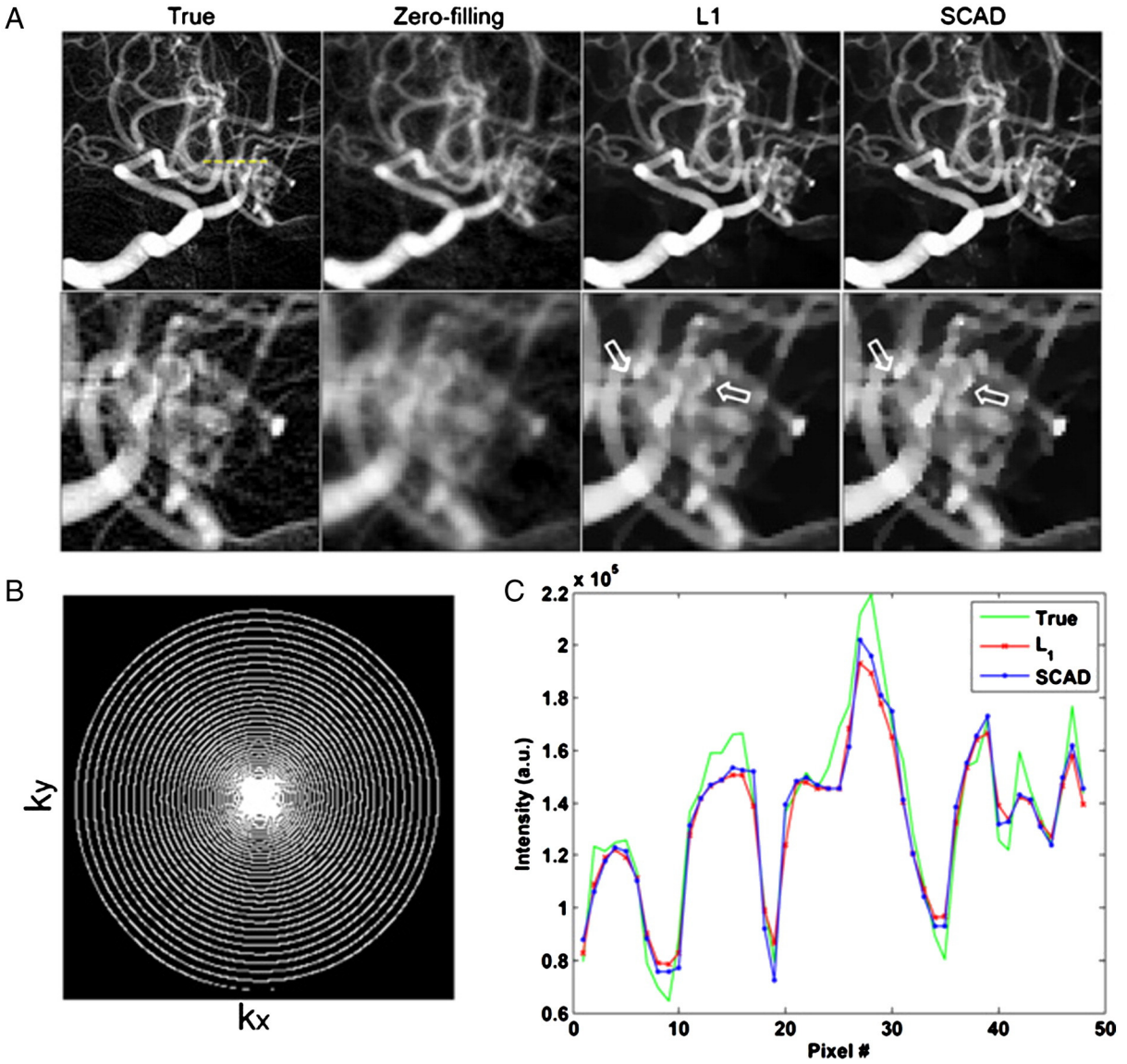

Fig. 3. (A) Reconstruction of the MR angiogram dataset through zero-filling, gradient-based l_1 and SCAD ADMM algorithms using a 3D stack of a single-shot variable-density spiral trajectory (78.7% undersampling). The L1 and SCAD images are shown with the same display window. (B) The illustration of *k*-space undersampling pattern. (C) The comparison of intensity profiles of reconstructed images along the dash line shown on the true image.

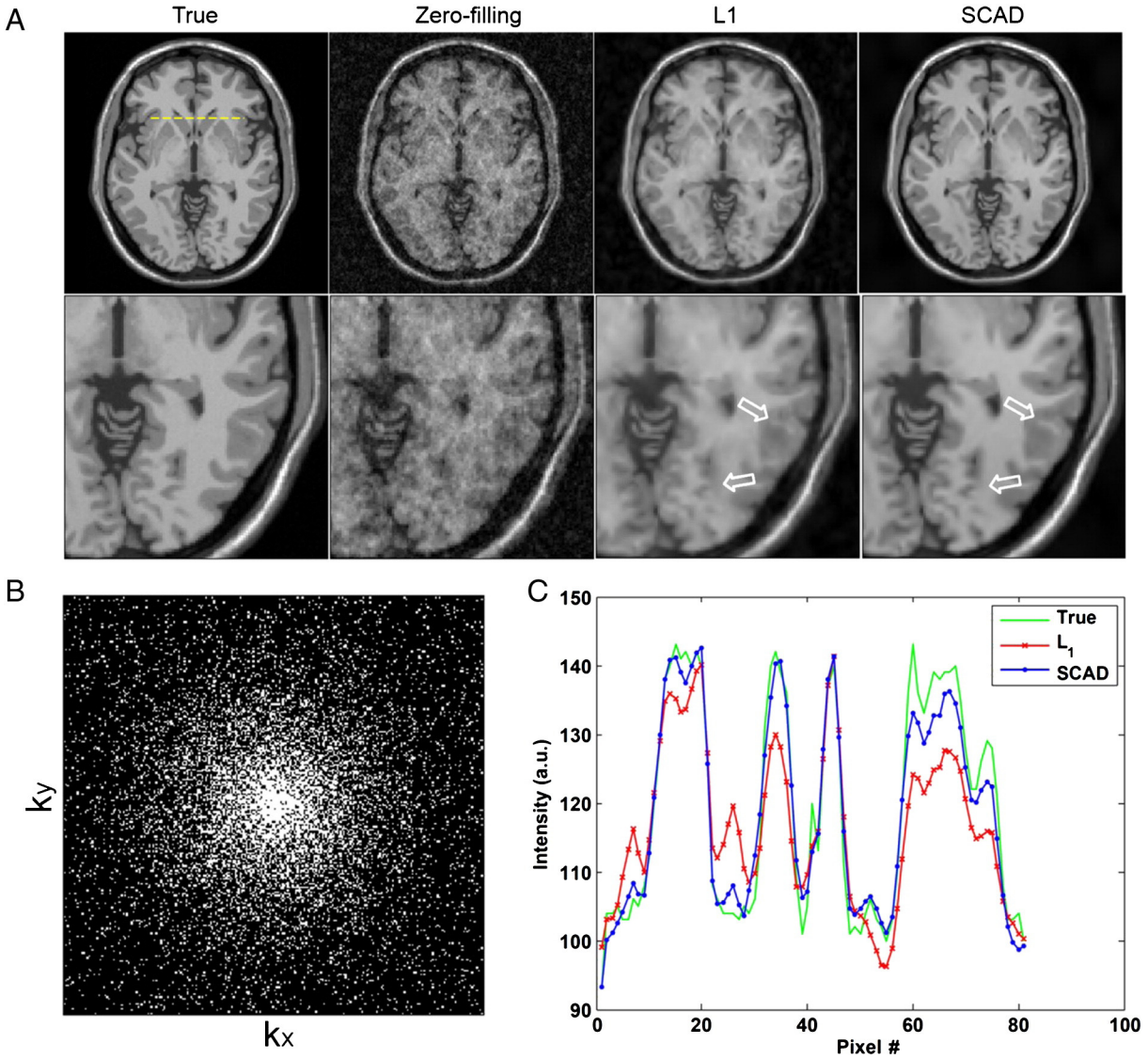

Fig. 4. (A) Reconstruction of the BrainWeb phantom through zero-filling, wavelet-based l_1 and SCAD ADMM algorithms using a variable density random sampling (85% undersampling). The L1 and SCAD images are shown with the same display window. (B–C) As in Fig. 3.

cases. In this simulation study with extremely high undersampling, the involved parameters, i.e. ρ, λ and a, were heuristically optimized to obtain the best case performance of the algorithms. For the radial sampling results, the optimized parameters were set to $\rho = 0.5$, $\lambda = 0.03$, $a = 3.7$ for SCAD and $\rho = 0.2$, $\lambda = 0.05$ for TV. Similarly, in the spiral sampling, the parameters were set to $\rho = 0.15$, $\lambda = 0.15$, $a = 3.7$ for SCAD and $\rho = 0.5$, $\lambda = 0.15$ for TV. The quantitative evaluation of the algorithms in terms of PSNR and SSIM index is presented in Table 1. The results show that the SCAD regularization significantly improves peak signal to noise ratio in the reconstructed images and also gives rise to a perfect similarity between the true and the reconstructed images. 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 Fig. 3 (A), a representative slice of the reconstructed MRA images is compared with the fully sampled ground truth. The visual comparison of the regularized reconstructions shows that both l_1-based TV and SCAD regularizations have noticeably suppressed noise and undersampling artifacts in comparison with zero-filling, which is in fact an un-regularized reconstruction. However, a close comparison of the images reveals that the SCAD regularization results in a higher image contrast (see arrows), since it exploits the weighting factors that suppress regularization across boundaries. Fig. 3 (B) also shows the k-space trajectory used for undersampling and the intensity profiles of the reconstructed images along the dashed line shown on the true image. The profiles also demonstrate that the SCAD regularization technique can improve the performance of its TV counterpart. The algorithms were also quantitatively evaluated based on PSNR and MSSIM metrics. The results summarized in Table 1 further demonstrate the outperformance of the proposed regularization technique. Note that during image reconstruction of this and the other two brain datasets, we first optimized the involved parameters of the ADMM algorithm, i.e. λ and ρ, for l_1-based regularizations to obtain the best case performance. Then, we optimized the performance of the SCAD regularization using the same values for the scale parameter a in Eq. (16). For this dataset, the optimal parameters were set to $\lambda = 1800$, $\rho = 1.2$ and $a = 100$.

Figs. 4 and 5 (A) show the image reconstruction results of the simulated (BrainWeb) and clinical brain Dixon datasets, respectively. As mentioned earlier, translation-invariant wavelets were employed as sparsifying transforms. As can be seen in both cases, the proposed regularization technique depicts improved performance in

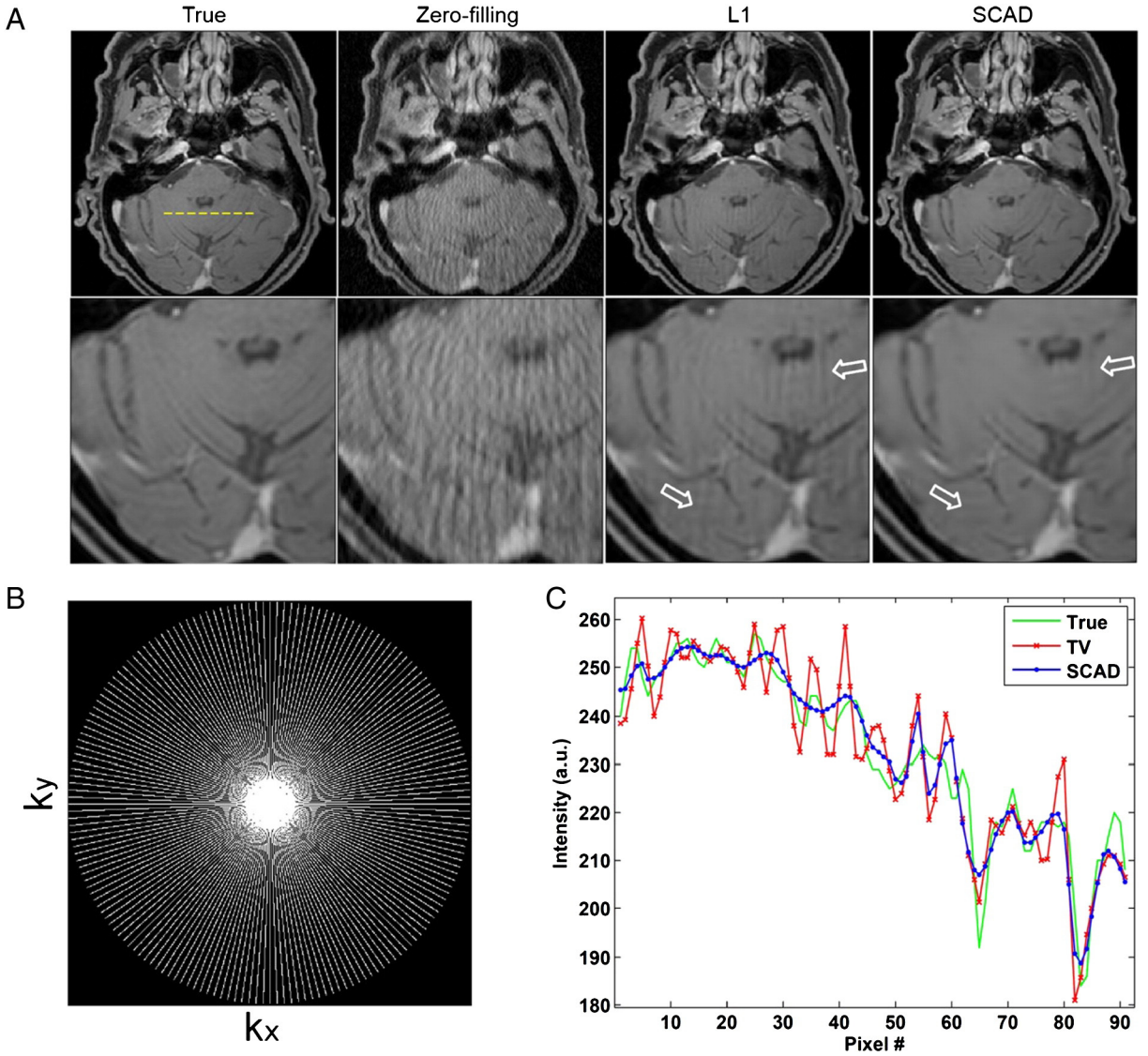

Fig. 5. (A) Reconstruction of the clinical brain Dixon (out-of-phase) dataset through zero-filling, wavelet-based l_1 and SCAD ADMM algorithms using a radial trajectory (87.54% undersampling). The L1 and SCAD images are shown with the same display window. (B–C) As in Fig. 3.

reducing the aliasing artifacts and restoring the details in comparison with l_1-based regularization (see arrows). In Figs. 4 and 5(B), *k*-space undersampling patterns as well as line profiles of the reconstructed images along the dash line in true images are shown. The line profiles demonstrate that the SCAD regularization can restore the true profiles more faithfully. As will be elaborated in the Discussion section, this regularizer exploits the redundant information in the image being reconstructed in order to suppress the thresholding of wavelet coefficients of image features and thereby to improve the accuracy of the reconstructed images. The quantitative evaluations of the l_1-based and SCAD regularizations presented in Table 1 demonstrate that the proposed regularizer achieves an improved SNR and structural similarity over its counterpart. The optimal parameters obtained for the simulated (BrainWeb) and clinical brain Dixon datasets were set to $\lambda = 3, \rho = 0.2$ and to $\lambda = 3, \rho = 0.5$ and $a = 10$, respectively.

4.2. Convergence rate and computation time

Table 1 summarizes the number of iterations and the computation (CPU) time per iteration for l_1 and SCAD-ADMM algorithms obtained for the studied datasets. Overall, for retrospective reconstruction of a dataset of size $256 \times 256 \times 30$, the TV and SCAD-ADMM algorithms require about 0.83 and 0.97 s per iteration in our MATLAB-based implementation. For the simulated BrainWeb and clinical brain Dixon datasets, which had matrix sizes of 256×256 and 512×512 and where wavelet transforms were used, the algorithms required an increased CPU time per iteration. This is due to the fact that wavelet transforms, particularly translation-invariant wavelets, require more arithmetic operations compared to finite differences (discrete gradients) and hence present with higher computational complexity. In practical settings, an iterative algorithm is said to be convergent if it reaches a solution where the image estimates do not change (or practically change within a certain tolerance determined by a stopping criterion) compared to the succeeding iteration [52]. As mentioned earlier, a tolerance of $\eta = 1 \times 10^{-4}$ was used in Algorithm 1 to declare the convergence of the algorithms. For the parameters optimally tuned, it was found that the l_1-based ADMM algorithm generally converges after a fewer number of iterations in comparison with the SCAD-ADMM. In the MRA, BrainWeb and brain Dixon datasets, it converged after 94, 335 and 94 iterations, respectively, while the SCAD converged after 138, 354 and 102 iterations, respectively. The same trend was also observed for the XCAT phantom with radial and spiral undersampling

patterns. This convergence behavior should be ascribed to the weighting scheme that SCAD regularization exploits. In fact, this weighting scheme attempts to iteratively recognize and preserve sharp edges. As a result, it takes the algorithm more iterations to identify true edges from those raising from aliasing and streaking artifacts. However, our results showed that the SCAD regularization achieves a higher PSNR at the same common iteration in comparison with the l_1-based regularization. Fig. 6 shows PSNR improvement with iteration number in CS-MR reconstruction of the MRA, BrainWeb and the brain Dixon datasets, respectively. As can be seen, the decreased convergence rate of (or the increased number iterations in) the proposed regularization technique is compensated with an overall increased PSNR.

5. Discussion

Fast MRI data acquisition is of particular importance in applications such as dynamic myocardium perfusion imaging and contrast-enhanced MR angiography. Compressed sensing provides a promising framework for MR image reconstruction from highly undersampled *k*-spaces and thus enables a substantial reduction of acquisition time [11,53]. In this study, we introduced a new regularization technique in compressed sensing MR image reconstruction based on the non-convex smoothly clipped absolute deviation norm with the aim of decreasing the sampling rate even lower than it is required by the conventional l_1 norm. The CS-MRI reconstruction was formulated as a constrained optimization problem and the SCAD norm was iteratively convexified by linear local approximations within an augmented Lagrangian framework. We employed finite differences and wavelet transforms as a sparsifying transform and compared the proposed regularizer with its l_1-based TV counterpart.

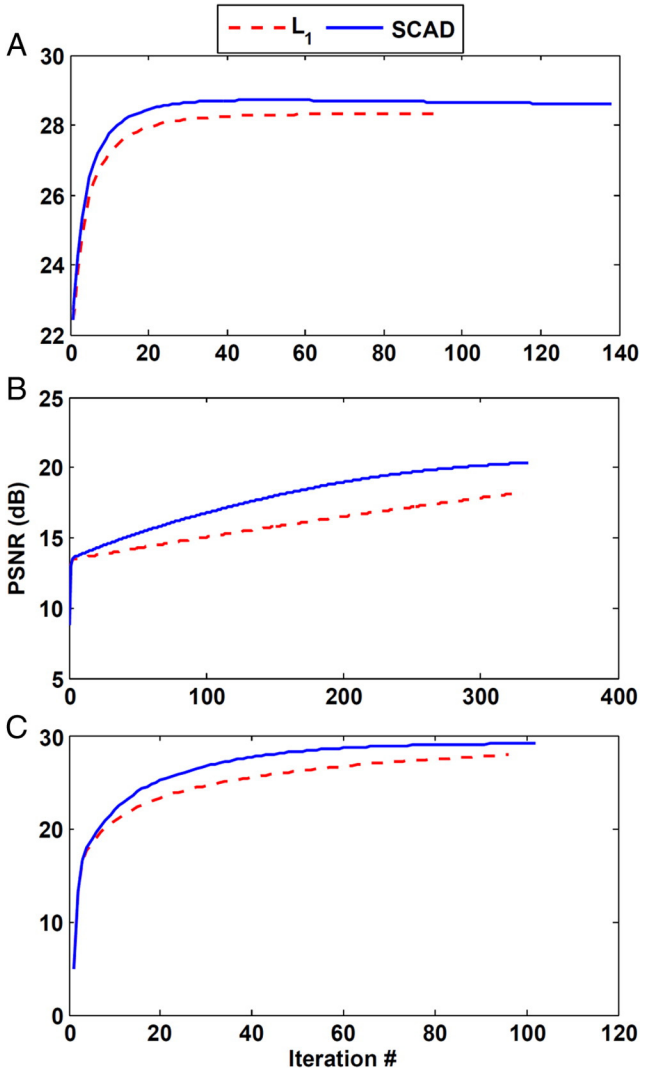

Fig. 6. PSNR improvement as a function of iteration number for the CS-MR reconstruction of: (A) the MRA, (B) BrainWeb and (C) the brain Dixon datasets using l_1 and SCAD-based regularizations.

5.1. Edge preservation and sparsity promotion

The linearization of the SCAD norm in effect gives rise to a reweighted l_1 norm. In general, our qualitative and quantitative results showed improved performance of the SCAD over the l_1 based regularization. This outperformance is due to the fact that the reweighted l_1 norm non-uniformly thresholds the gradient fields and wavelet 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 features and on the other hand, enforce the smoothing of regions contaminated by noise and artifacts. Fig. 7(A) and Fig. 7(B) show respectively the weighting factors associated with gradient fields of the MRA image (Fig. 3) and those with the wavelet coefficients of the brain Dixon image (Fig. 5) as a function of iteration number. Fig. 7(B) only shows the weights of the detail coefficients at resolution 4 in the diagonal direction. It is worth mentioning that the wavelet transform decomposes an image into one approximate subband and several (horizontal, vertical and diagonal) detail subbands. It can be seen that as iteration number increases: i) the true anatomical boundaries are being distinguished from false boundaries arising from artifacts, especially in the case of the brain dataset where the radial sampling pattern results in streaking artifacts and ii) the emphasis on edge preservation (the wall of vessels or the border of structures) increases by assigning zero or close to zero weights (dark intensities) to edges and the suppression of in-between regions is continued by high-value weights (bright intensities). Furthermore, the dynamic range of the weighting factors is continuous and varies based on the importance and sharpness of the edges, which demonstrates the adaptive nature of this weighting scheme. The end result of this procedure is in fact the promotion of the sparsity of image estimate in the domain of the sparsifying transform. Fig. 8(A) and Fig. 8(B) show the horizontal gradient field ($\boldsymbol{\theta}^h$) of the MRA frame shown in Fig. 3 at iteration number 5, thresholded respectively by soft-thresholding and weighted soft-thresholding with the same regularization and penalty parameters. The histograms of the images (20 bins) are shown in Fig. 8(C–D). The results 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 coefficients in close-to-zero bins has been reduced, while it has been increased in the zero-bin.

To enhance sparsity, Candes et al. [30] 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 $w_i^k = \lambda(|\theta_i^k| + \varepsilon)^{-1}$, in our notation in Eq. (19)] 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. [19] proposed a homotopic l_0 norm approximation by gradually

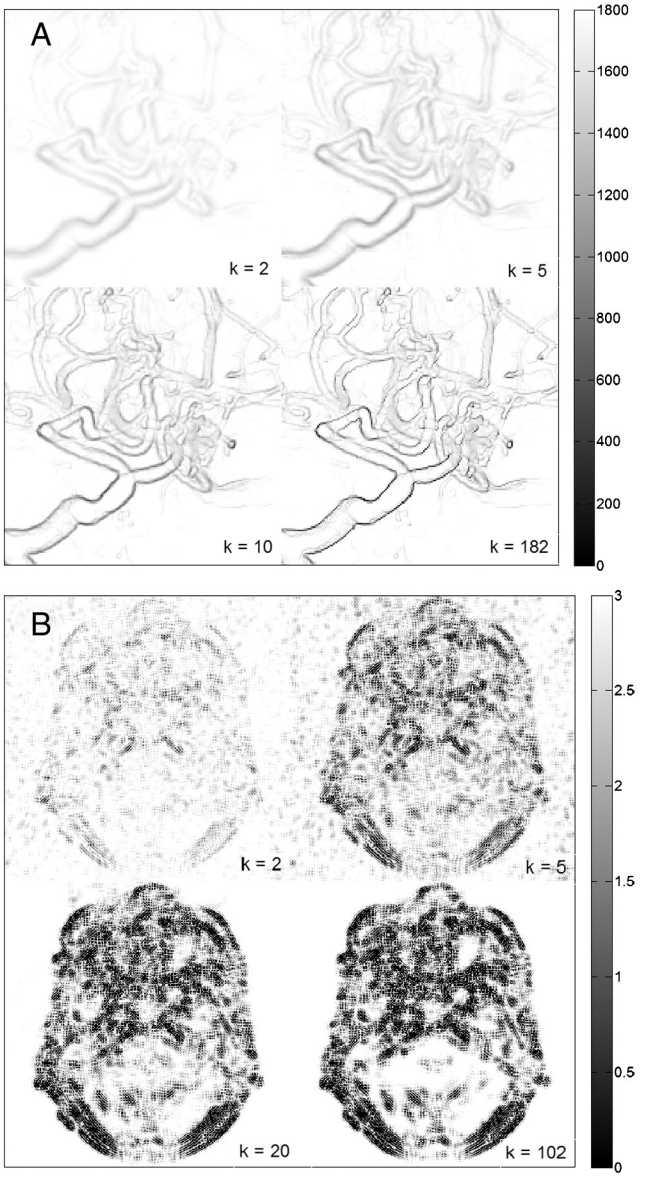

Fig. 7. Evolution of the weights ($\mathbf{w}^k$) of (A) the MRA frame and (B) brain Dixon image shown in Figs. 3 and 6 with iteration number. The gray-color bar shows the dynamic range of the weights.

reducing the perturbation parameter 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 in Eq. (18) with an l_0 regularizer, can be achieved with a hard thresholding rule [54,55], 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 soft thresholding, which is associated with l_1 regularization. In comparison with the weighting scheme of Candes et al. and in connection with the homotopic l_0 approximations, the linearized SCAD regularization invokes a weighted soft-thresholding rule that in limit approaches hard thresholding rule. Fig. 9(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. (17)], for different values of the scale parameter a and for $\lambda = 1$. Similarly, Fig. 9(B) compares those standard rules with the weighted soft-thresholding rule by Candes' weighting scheme for different values of the perturbation parameter ε and for $\lambda = 1$. As can be seen, for small values of a, the weighted soft-thresholding rule of SCAD resembles hard thresholding, while for small values of ε, the Candes' weighted 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 of as a soft-thresholding with zero threshold). In fact, Fan and Li [37] 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 tendency of soft-thresholding on over-penalizing large coefficients and the discontinuity and thus instability of hard-thresholding and at the same time, promotes the sparsity of the solution [37]. Generalized l_p Gaussian norms, for $0 < p < 1$, have also been successfully used in the context of CS-MRI reconstruction [19,56]. The sparsity induced by l_p norms is promoted as p approaches zero, since the resulting norm invokes a pseudo l_0 norm. In [57], Foucart and Lai showed that l_p norms give rise to a reweighted l_1 norm with weighting factors $w_i^k = \lambda(|\theta_i^k| + \varepsilon)^{p-1}$ (in our notation), which reduces to Candes' weights for $p = 0$. Chartrand [56] derived the so-called *p-shrinkage* rule for l_p norms, which is in fact a weighted soft thresholding rule with the weights $w_i^k = \lambda(|\theta_i^k|)^{p-1}$. Based on the asymptotic behavior of Candes' weighted soft thresholding (Fig. 9), one can conclude that l_p norms (with $0 < p < 1$) best approach a *semi-hard* thresholding rule for $p = 0$ and $\varepsilon = 0$, while the thresholding rule invoked by linearized SCAD can well approach a hard thresholding rule.

The results presented in Figs. 3–5 show that the images reconstructed by l_1 based-ADMM algorithm suffer from an overall smoothing in comparison with SCAD and show to some extent reduced stair-casing artifacts that are sometimes associated with TV regularization using gradient and time-marching based optimization algorithms. These artifacts, which are false and small patchy regions within a uniform region, are often produced by first-order finite differences and can be mitigated by second-order differences [23] or a balanced combination of both [58]. As noted by Trzasko et al. [19] and implied from the above discussion, the soft thresholding (associated with l_1 norm) in the ADMM algorithm tends to uniformly threshold the decomposition coefficients, thereby it leads to a global smoothing of the image (similar to wavelet-based reconstructions) and thus the reduction of false patchy regions.

5.2. Computational complexity and parameter selection

In this study, we solved the standard CS-MRI image reconstruction defined in Eq. (2) using an augmented Lagrangian method. Recently, Ramani et al. [40] studied AL methods for pMRI and showed that this class of algorithms is 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 $\boldsymbol{x}$, 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 Cartesian CS-MRI, which allowed speeding up the algorithm, particularly by the per-computation of the eigenvalue matrix of the discrete gradient matrix $\boldsymbol{\Psi}$. As mentioned earlier, this analytic solution can also be assessed by regridding the non-Cartesian data or using the approximation method proposed in [45].

The convergence properties of the proposed SCAD-ADMM algorithm, similar to other optimization algorithms, depends on the choice of the involved parameters, i.e. the penalty parameter ρ, the regularization parameter λ and the SCAD's scale parameter a. The choice of these parameters in turn depends on a number of

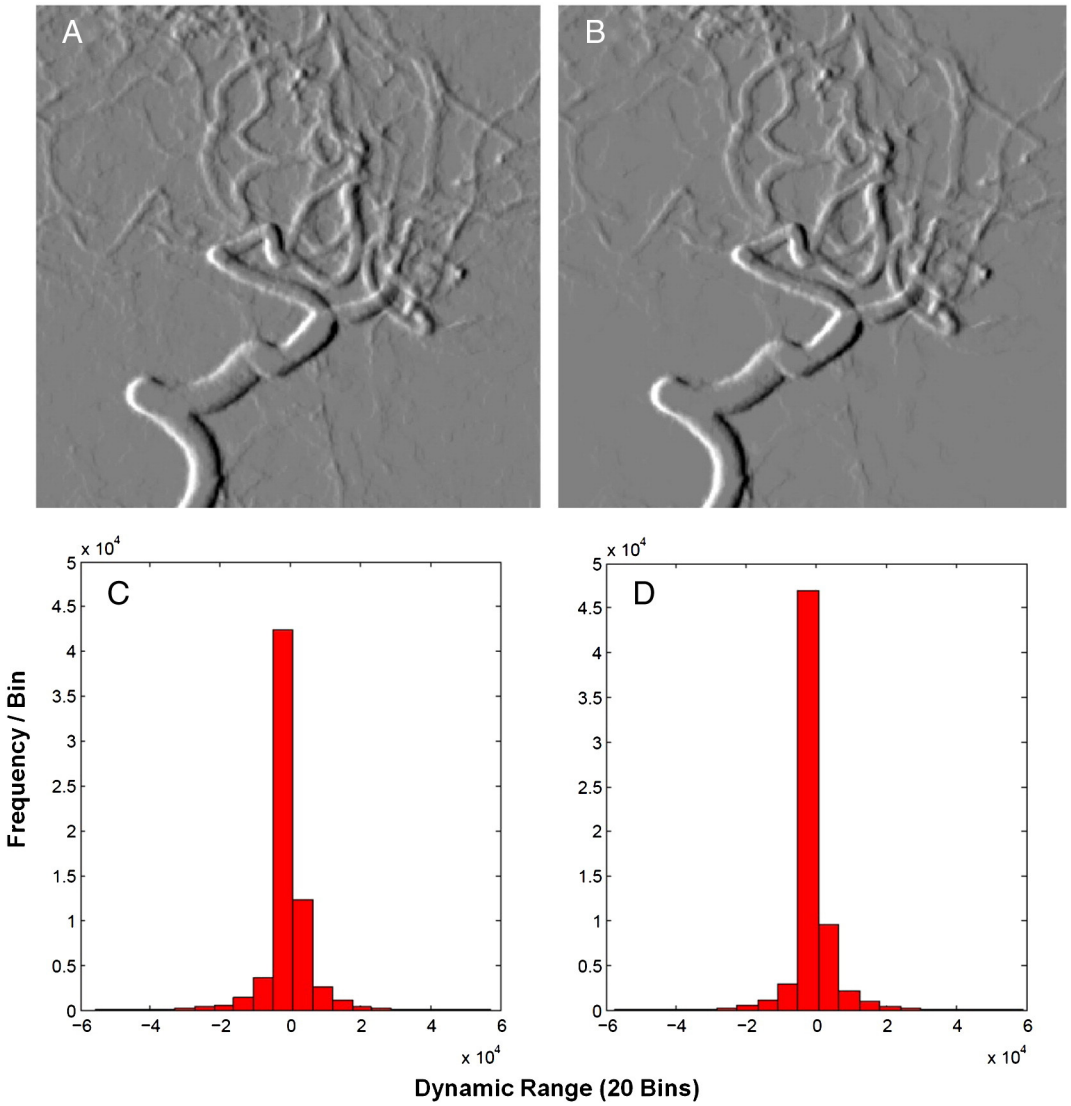

Fig. 8. The sparsity promotion of SCAD regularizer. (A–B) The horizontal gradient field of the image frame shown in Fig. 3 (at iteration 5) thresholded respectively by soft and weighted soft thresholding with the same regularization and penalty parameters. (C–D) The histograms of the images shown in (A) and (B), respectively (with 20 bins). Similar display window is used for the images shown in (A) and (B).

factors including the acquisition protocol, level of noise and the task at hand. As mentioned earlier in the clinical datasets, we optimized the parameter λ and ρ for l_1 regularization to obtain the best qualitative and quantitative results. Using the same parameters, the parameter a was optimized for SCAD regularization. In fact, we followed this parameter selection for an un-biased comparison of the l_1 and SCAD regularizations in order to demonstrate that the proposed technique can improve the performance of l_1 regularization by incorporating iteratively derived weighting factors. However, this parameter selection might not be optimal for SCAD and its convergence properties, since this regularizer tends to gradually and cautiously remove noisy-like artifacts. In particular, we found that higher values of λ improved the convergence rate of the l_1- and hence SCAD-based algorithms, but at the same time, resulted in overall smoothing of image features, which in the case of SCAD regularization could be compensated by decreasing the scale parameter a. The latter parameter, in fact, controls the impact of weighting factors in SCAD regularization. The lower values of this parameter increase edge-preservation because, as can be seen in Fig. 9(A), the resulting weighted soft thresholding approaches a hard thresholding rule, which is akin to l_0 regularization. While higher values of a reduce edge preservation, since the weighting factors become uniform and approach the parameter λ. In other words, as seen in Fig. 9(A), the weighted soft thresholding approaches the standard soft thresholding with a uniform threshold value of λ. In general, it was observed that the parameter a is less data-dependent and shows a higher flexibility for its selection in comparison with λ. An alternative way to choose the pair parameters (a, λ) for SCAD could be two-dimensional grids search using some criteria such as cross validation and L-curve methods [6], which calls for future investigations.

5.3. Future directions

As our results demonstrate, the SCAD regularization allows for more accurate reconstructions from highly undersampled data and hence reduced sampling rate for rapid data acquisitions. However, it is worth to mention the limitations and future outlook of the proposed SCAD-ADMM algorithm. In this study, we evaluated the proposed algorithm for Cartesian approximation of radial and spiral trajectories and compared its performance with l_1-based ADMM. These approximation and synthetization of the k-spaces from

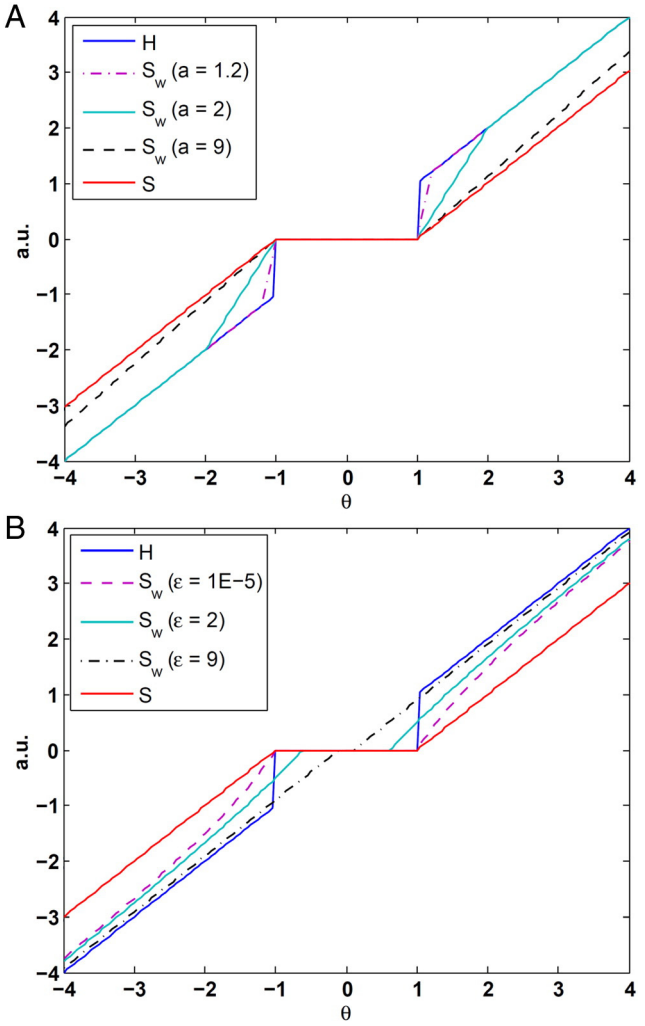

Fig. 9. Comparison of hard (H) and soft (S) thresholding rules with soft thresholding (S_w) rules weighted by: (A) the linearization of SCAD function for different scale parameters a and (B) Candes' approach for different perturbation parameters.

magnitude images cannot accurately represent the actual data acquisition process in MRI. Nevertheless, our comparative results showed that the proposed regularization can improve the performance of l_1-based regularization using adaptive identification and preservation of image features. Future work will focus on the evaluation of the proposed regularization for non-Cartesian datasets and also parallel MR image reconstruction using AL methods [40]. Performance assessment of the SCAD norm in comparison with a family of reweighted l_1 norms and l_0 homotopic norms within the presented AL framework is also underway.

6. 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 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 and wavelet transforms as a sparsifying transform and demonstrated that the linearized SCAD regularization is an iteratively weighted l_1 regularization with improved edge-preserving and sparsity-promoting properties. The performance of the algorithm was evaluated using phantom simulations and retrospective CS-MRI of clinical studies. It was found that the proposed regularization technique outperforms the conventional l_1 regularization and can find applications in CS-MRI image reconstruction.

Acknowledgment

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

References

[1] Plein S, Bloomer TN, Ridgway JP, Jones TR, Bainbridge GJ, Sivananthan MU. Steady-state free precession magnetic resonance imaging of the heart: comparison with segmented k-space gradient-echo imaging. J Magn Reson Imaging 2001;14:230–6.
[2] McGibney G, Smith MR, Nichols ST, Crawley A. Quantitative evaluation of several partial fourier reconstruction algorithms used in MRI. Magn Reson Med 1993;30(1):51–9.
[3] Larkman DJ, Nunes RG. Parallel magnetic resonance imaging. Phys Med Biol 2007;52:R15–55.
[4] Blaimer M, Breuer F, Mueller M, Heidemann RM, Griswold MA, Jakob PM. SMASH, SENSE, PILS, GRAPPA: how to choose the optimal method. Top Magn Reson Imaging 2004;15:223–36.
[5] Pruessmann KP, Weigher M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. Magn Reson Med 1999;42:952–62.
[6] Karl WC. Regularization in image restoration and reconstruction. In: Bovik A, editor. Handbook of image and video processing. San Diego: Academic Press; 2000.
[7] Liang Z-P, Bammer R, Jim J, Pelc NJ, Glover GH. Improved image reconstruction from sensitivity-encoded data by wavelet denoising and Tikhonov regularization. Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI); 2002. p. 493–6.
[8] Lin FH, Wang FN, Ahlfors SP, Hamalainen MS, Belliveau JW. Parallel MRI reconstruction using variance partitioning regularization. Magn Reson Med 2007;58:735–44.
[9] Kyriakos WE, Panych LP, Kacher DF, et al. Sensitivity profiles from an array of coils for encoding and reconstruction in parallel (SPACE RIP). Magn Reson Med 2000;44:301–8.
[10] Ying L, Liu B, Steckner MC, Wu G, Wu M, Li S-J. A statistical approach to SENSE regularization with arbitrary k-space trajectories. Magn Reson Med 2008;60(2):414–21.
[11] Lustig M, Donoho DL, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn Reson Med 2007;58:1182–95.
[12] Jung H, Sung K, Nayak KS, Kim E, Yeop, Ye JC. k-t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI. Magn Reson Med 2009;61:103–16.
[13] Gamper U, Boesiger P, Kozerke S. Compressed sensing in dynamic MRI. Magn Reson Med 2008;59(2):365–73.
[14] Hong M, Yu Y, Wang H, Liu F, Crozier S. Compressed sensing MRI with singular value decomposition-based sparsity basis. Phys Med Biol 2011;56:6311–25.
[15] Ravishankar S, Bresler Y. MR image reconstruction from highly undersampled k-space data by dictionary learning. IEEE Trans Med Imaging 2011;30(5):1028–41.
[16] Candes E, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theory 2006;52:489–509.
[17] Donoho D. Compressed sensing. IEEE Trans Inf Theory 2006;52(4):1289–306.
[18] Candès E, Romberg J. Sparsity and incoherence in compressive sampling. Inverse Probl 2007;23(3):969.
[19] Trzasko J, Manduca A. highly undersampled magnetic resonance image reconstruction via homotopic L0 minimization. IEEE Trans Med Imaging 2009;28(1):106–21.
[20] Lustig M, Donoho DL, Pauly JM. k-t SPARSE: high frame rate dynamic MRI exploiting spatio-temporal sparsity. Proceedings of the annual meeting of ISMRM; 2006. p. 2420 Seattle, WA.
[21] Hu C, Qu X, Guo D, Bao L, Chen Z. Wavelet-based edge correlation incorporated iterative reconstruction for undersampled MRI. Magn Reson Imaging 2011;29(7):907–15.
[22] Lustig M, Donoho DL, Pauly JM. Rapid MR imaging with compressed sensing and randomly under-sampled 3DFT trajectories. Proceedings of the annual meeting of ISMRM; 2006. p. 695 Seattle, WA.
[23] Block KT, Uecker M, Frahm J. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. Magn Reson Med 2007;57(6):1086–98.
[24] Adluru G, McGann C, Speier P, Kholmovski EG, Shaaban A, DiBella EVR. Acquisition and reconstruction of undersampled radial data for myocardial perfusion magnetic resonance imaging. J Magn Reson Imaging 2009;29(2):466–73.
[25] Chen L, Adluru G, Schabel M, McGann C, Dibella E. Myocardial perfusion MRI with an undersampled 3D stack-of-stars sequence. Med Phys 2012;39(8):5204–11.

[26] Knoll F, Clason C, Bredies K, Uecker M, Stollberger R. Parallel imaging with nonlinear reconstruction using variational penalties. Magn Reson Med 2012;67(1):34–41.
[27] Ji JX, Chen Z, Tao L. Compressed sensing parallel magnetic resonance imaging. Proceedings of the IEEE EMBS Conference; 2008. p. 1671–4.
[28] Liu B, King K, Steckner M, Xie J, Sheng J, Ying L. Regularized sensitivity encoding (SENSE) reconstruction using bregman iterations. Magn Reson Med 2009;61(1):145–52.
[29] Otazo R, Kim D, Axel L, Sodickson DK. Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion mri. Magn Reson Med 2010;64:767–76.
[30] Candès EJ, Wakin MB, Boyd SP. Enhancing sparsity by reweighted L1 minimization. J Fourier Anal Appl 2008;14:877–905.
[31] Chang C-H, Ji J. Parallel compressed sensing mri using reweighted L1 minimization. Proceedings of the annual meeting of ISMRM 2011;19. p. 2866.
[32] Hu C, Qu X, Guo D, Bao L, Cai S, Chen Z. Undersampled MRI reconstruction using edge-weighted L1 norm minimization. Proceedings of the annual meeting of ISMRM, 19. 2011. p. 2832.
[33] Lustig M, Velikina J, Samsonov A, Mistretta C, Pauly JM, Elad M. Coarse-to-fine Iterative reweighted L1-norm compressed sensing for dynamic imaging. Proceedings of the annual meeting of ISMRM 2010;18. p. 4892.
[34] Weller DS, Polimeni JR, Grady L, Wald LL, Adalsteinsson E, Goyal VK. Evaluating sparsity penalty functions for combined compressed sensing and parallel MRI. Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI); 2011. p. 1589–92.
[35] Chartrand R. Exact reconstruction of sparse signals via nonconvex minimization. IEEE Signal Process Lett 2007;14(10):707–10.
[36] Yong W, Dong L, Yuchou C, Ying L. A hybrid total-variation minimization approach to compressed sensing. Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI); 2012. p. 74–7.
[37] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Amer Statistical Assoc 2001;96(456):1348–60.
[38] Teixeira FCA, Bergen SWA, Antoniou A. Signal recovery method for compressive sensing using relaxation and second-order cone programming. Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS); 2011. p. 2125–8.
[39] Afonso MV, Bioucas-Dias JM, Figueiredo MA. Fast image recovery using variable splitting and constrained optimization. IEEE Trans Image Process 2010;19(9):2345–56.
[40] Ramani S, Fessler JA. Parallel MR image reconstruction using augmented Lagrangian methods. IEEE Trans Med Imaging 2011;30(3):694–706.
[41] Nocedal J, Wright SJ. Numerical optimization. Springer; 2006.
[42] Powell M. A method for nonlinear constraints in minimization problems. In: Fletcher R, editor. Optimization. New York: Academic Press; 1969. p. 283–98.
[43] Eckstein J, Bertsekas D. On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math Program 1992;5:293–318.
[44] Wang Y, Yang J, Yin W, Zhang Y. A new alternating minimization algorithm for total variation image reconstruction. SIAM J Imaging Sciences 2008;3:248–72.
[45] Akcakaya M, Nam S, Basha T, Tarokh V, Manning WJ, Nezafat R. Iterative compressed sensing reconstruction for 3d non-cartesian trajectories without gridding & regridding at every iteration. Proceedings of the annual meeting of ISMRM 2011;19. p. 2550.
[46] Hunter D, Li R. Variable selection using mm algorithms. Ann Statist 2005;33:1617–42.
[47] Zou H, Li R. One-step sparse estimates in nonconcave penalized likelihood models. Ann Stat 2008;36:1509–33.
[48] Antoniadis A, Fan J. Regularization of wavelets approximations. J Am Stat Assoc 2001;96:939–67.
[49] MRI unbound. Available at http://wwwismrmorg/mri_unbound/simulatedhtm.
[50] BRAINWEB. Available at http://wwwbicmnimcgillca/brainweb/.
[51] Zhou W, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004;13(4):600–12.
[52] Lalush DS, Tsui BMW. A fast and stable maximum a posteriori conjugate gradient reconstruction algorithm. Med Phys 1995;22(8):1273–84.
[53] Lustig M, Donoho DL, Santos JM, Pauly JM. Compressed sensing MRI. IEEE Signal Process Mag 2008;25(2):72–82.
[54] Blumensath T, Yaghoobi M, Davies ME. Iterative hard thresholding and L0 regularisation. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2007;3. p. III-877–80.
[55] Wright SJ, Nowak RD, Figueiredo MAT. Sparse reconstruction by separable approximation. IEEE Trans Signal Process 2009;57(7):2479–93.
[56] Chartrand R. Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI); 2009. p. 262–5.
[57] Foucart S, Lai M. Sparsest solutions of undertermined linear systems via Lq-minimization for $0 < q < 1$. Appl Comput Harmon Anal 2009;26:395–407.
[58] Knoll F, Bredies K, Pock T, Stollberger R. Second order total generalized variation (TGV) for MRI. Magn Reson Med 2011;65(2):480–91.