Biomed Eng Lett (2014) 4:80-92
DOI 10.1007/s13534-014-0126-2

ORIGINAL ARTICLE

Optimal Selection of Regularization Parameter in Total Variation Method for Reducing Noise in Magnetic Resonance Images of the Brain

Michael Osadebey, Nizar Bouguila and Douglas Arnold, for the ADNI[†]

Received: 15 February 2014 / Revised: 22 March 2014 / Accepted: 25 March 2014

Abstract
Purpose In the image processing community total variation (TV) is widely acknowledged as a popular and state-of-the-art technique for noise reduction because of its edge-preserving property. This attractive feature of TV is dependent on optimal selection of regularization parameter. Contributions in literature on TV focus on applications, properties and the different numerical solution methods. Few contributions which address the problem of regularization parameter selection are based on regression methods which pre-exist introduction of TV. They are generic and elegantly formulated, and their operation is in series with TV framework. For these reasons they render TV computationally inefficient and there is significant manual tuning when they are deployed in specific applications.
Methods This paper describes a non-regression approach for selection of regularization parameter. It is based on a new concept, the Variational-Bayesian (VB) cycle. Within the context of VB cycle we derive two important results. First, we confirm the notion held for a long time by researchers, within image processing and computer vision community, that variational and Bayesian techniques are equivalent. Second, the value of regularization parameter is equal to noise variance, and is determined, at no computational cost to TV denoising algorithm, from a mathematical model that describes relationship between Markov random field energy and noise level in magnetic resonance images (MRI) of brain. The second result is similar to one reported in [1] in which the authors, for special choice of regularization operator in different regression methods, derive value of regularization parameter as equal to noise variance.
Results Our proposal was evaluated on brain MRI images with different acquisition protocols from two clinical trials study management centers. It was based on visual quality, computation time, convergence and optimality.
Conclusions The result shows that our proposal is suitable in applications where high level of automation is demanded from image processing software.

Keywords Magnetic resonance imaging (MRI), Total variation (TV), Regularization parameter, Markov random field, Noise level

Michael Osadebey (✉)
Department of Electrical and Computer Engineering, Concordia University, 1515 St. Catherine Street West., Montreal, Quebec, H3G 2W1, Canada
Tel : +1-514-848-2424 ext 7135
E-mail : m_osadeb@encs.concordia.ca

Nizar Bouguila
Concordia Institute for Information Systems Engineering, Concordia University, 1515 St. Catherine Street West., Montreal, Quebec, H3G 2W1, Canada

Douglas Arnold
NeuroRx Research Inc, 3575 Parc Avenue, Suite # 5322, Montreal, QC, H2X 4B3 Canada

[†]Data used this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

INTRODUCTION

Brain MRI images in clinical trials

Magnetic resonance images (MRI) of human brain exhibit unique characteristics. They can be generally described as piecewise smooth and statistically simple [2]. Three structures namely white matter, gray matter and ventricular system dominate slices from brain MRI images of a single subject. There is also geometric similarity among different subjects across age, gender and race [3]. Brain MRI images are highly efficient for study and examination of brain anatomy as well as detection of signatures of neurodegenerative diseases such

as multiple sclerosis and Alzheimer [4, 5]. Daily, several thousands of brain MRI slice images are delivered from clinical trial sites around the globe to clinical research organizations that manage clinical trials of new drugs for pharmaceutical organizations. Efficient management of this large amount of data demands high level of automation from image processing and image analysis software at study management centers. The performance of image analysis system is strongly dependent on quality of MRI image. The quality of an MRI image reflects performance of noise reduction unit within image processing system. The noise reduction unit contribution to quality of acquired MRI image is preservation of sharp edges that defines boundaries of anatomical structures and disease signatures for efficient image analysis.

Problem statement

An MRI slice image $\mathbf{I_d}$, in its original form of acquisition, is a complex valued data having both real and imaginary components degraded by noise $\mathbf{n}$. The noise can be modeled by a Gaussian distribution of mean zero. The degradation processes in the two components of the complex plane are identical linear model:

$$\mathbf{I_d} = \mathbf{H}\mathbf{I_c} + \mathbf{n} \tag{1}$$

where $\mathbf{I_c}$ is the clean data and $\mathbf{H}$ is the point spread function of the imaging system. Combination of the real and imaginary components of the MRI data produces magnitude MRI image resulting in nonlinear transformation of the Gaussian distribution of the pixels in the complex plane into Rician distribution [6].

In 1992 Rudin, Osher and Fatemi (ROF) introduced total variation (TV) technique to the image processing community [7]. The trio observed that any image such as a MRI image $\mathbf{I_d}$ derived from degradation process in Eq. (1) is characterized by excessive details resulting in high total variation:

$$\sum_{i=1}^{M}\sum_{j=1}^{N}\left\|(\nabla \mathbf{I_d})_{(i,j)}\right\| \tag{2}$$

where $M \times N$ is the dimension of the image and $\nabla\mathbf{I_d}$ is the gradient of the image at pixel location (i,j). They reasoned that noise signatures expressed by excessive details can be suppressed if total variation of the image is subject to closeness in value to underlying image $\mathbf{I_c}$. They formulate problem of noise reduction as unconstrained minimization of total variation:

$$\hat{I}_c = \min_{\mathbf{I_c}} \sum_{i=1}^{M}\sum_{j=1}^{N}\left\|(\nabla \mathbf{I_d})_{(i,j)}\right\| + \lambda\left\|\mathbf{H}\mathbf{I_c} - \mathbf{I_d}\right\|_2^2 \tag{3}$$

where $\hat{I}_c$, $(\left\|\mathbf{H}\mathbf{I_c} - \mathbf{I_d}\right\|_2^2)$ and $\{\lambda : 0 \le \lambda \le \infty\}$ are the denoised image, fidelity term and regularization parameter respectively. The blurring function is an identity matrix $\mathbf{H} = \mathbf{I}$ and the fidelity term is a measure of closeness of observed image to underlying image. The regularization parameter weighs how the total variation and fidelity term is reflected in the denoised image. Values for λ that are too high, too low and equal to zero results in denoised images $\hat{I}_c$ with corresponding three different properties:

$$\lambda \begin{cases} \Rightarrow \textbf{Too High} & \hat{\mathbf{I}}_\mathbf{c} \Rightarrow \textbf{low noise \& oversmooth} \\ \Rightarrow \textbf{Too Low} & \hat{\mathbf{I}}_\mathbf{c} \Rightarrow \textbf{high noise \& piecewise smooth} \\ = 0 & \hat{\mathbf{I}}_\mathbf{c} \quad \textbf{same as observed image} \end{cases} \tag{4}$$

If the value of λ tends to be too high the algorithm is constrained to place more emphasize on noise removal but the denoised image is over-smooth because there is less emphasis on impacting piecewise smooth property of the underlying image on the denoised image. On the other hand, value of λ that tends to be too low results in denoised image that possesses the piecewise smooth property of the underlying image but retains noise. A special case is when $\lambda = 0$, in which the denoised image is same as the noisy image. Thus, optimal performance of TV technique is strongly dependent on optimal selection of λ [1]. Regularization parameter higher than optimal value will remove more noise, and this tends to destroy image details. On the other hand regularization parameter lower than optimal will preserve image detail but tends to allow noise to prevail.

Applications of total variation technique

Total variation technique is widely acknowledged as a popular and state-of-the-art technique. Its main attractive feature is ability to preserve edges. Today, the application of TV is beyond denoising. It is applied in super-resolution [8], computed tomography (CT) images analysis [9, 10], MRI images analysis [11], remote sensing [12], video restoration [13], blind deconvolution [14], and image inpainting [15, 16].

Detailed report on theory behind formulation of TV denoising, its operational characteristics and edge-preserving property can be found in [17]. Some researchers such as [18, 19] and [20] investigated the relationship between local image pixel intensities, image scale and regularization parameter and how global image restoration relates with image scale, frequency distribution of image pixel intensities and geometric features.

Numerical solutions to TV have attracted the attention of researchers. Partial differential equation technique similar to the original proposal by ROF was proposed by [21] and [22]. Primal-dual method proposed by [23] was an improvement in terms of speed and convergence over earlier work by [24]. There is the generalized accelerated proximal gradient

technique proposed by [25], which was reported to converge faster than the classical accelerated proximal gradient [26] and the alternating minimization technique [27], a variant of variable splitting technique [28].

Review of TV regularization parameter selection techniques

Fifteen years after the introduction of TV technique [29] and [30] were most likely the earliest contributions on selection of regularization parameter focused on TV image restoration. In [29] the image, regularization parameter and parameter associated with the prior image were jointly estimated using variational distribution within Bayesian framework. The generalized cross validation (GCV) method [31] was exploited to propose an automatic regularization parameter selection scheme in [30]. GCV is regarded as a weighted version of the classical cross validation method [32] for predicting the accuracy of a model in statistics. The observed data is partitioned into arbitrary number of independent sets. Each of the single observation is adopted, one at a time, as a test set and the remaining independent observations are regarded as training sets with which to predict the model. The reasoning here is that optimal choice is the observed regularization parameter that gives the best prediction.

The discrepancy rule-based method to automatically choose the regularization parameter was proposed in [33] and [34]. The discrepancy principle [35] recommends that, if there is a fair knowledge of the error $\varepsilon \approx \|\mathbf{H}\mathbf{I}_\mathbf{c} - \mathbf{I}_\mathbf{d}\|_2^2$ between observed and underlying image, the regularization parameter should be chosen when residual error of the regularized solution is less than or equal to $\beta\varepsilon$ for sone user defined constant $\beta > 1$.

Motivated by the performance of unbiased predictive risk estimator (UPRE) in Tikhonov regularization technique the authors of [36] propose UPRE to select regularization parameter in TV. Mallows [37] was the first to propose UPRE. The method relies on linear relationship between the regularized solution and data. Knowledge of noise variance is required to compute the prediction error P_λ, a function of the regularization parameter. P_λ is the difference between the computed solution and the ground truth solution. The optimal parameter is the minimizer of P_λ.

The authors in [38] and [39] reasoned that λ in the original formulation of ROF is a global parameter and does not satisfy local piecewise smoothness constraints in all regions within the image. They formulate a variant of TV model and propose multiple spatially dependent regularization parameters satisfying local constraints. The spatial dependency of λ was combined with the knowledge that images are comprised of multiple objects at different scales, and a spatially dependent multi scale total variation model was proposed by [40].

The use of Stein's Unbiased Risk Estimate (SURE) and GCV for optimal removal of speckle and Rician noise in synthetic aperture radar (SAR) and MRI images was reported in [41] and [42], respectively. Given a deterministic underlying image $\mathbf{I}_\mathbf{c}$ and its estimate $h(\mathbf{I}_\mathbf{d})$ computed in a denoising process, SURE, proposed by [43], is an unbiased estimator of variance-based expectation of the mean square error. For recent developments in numerical method solutions and applications of total variation we refer our readers to [44].

Limitations of current algorithms

The different techniques for selection of regularization parameter are based on classical parameter selection methods that were in existence before the introduction of TV technique. The algorithms are presented elegantly and their performances were evaluated using standard test images. In most cases their operation is a minimization process within another minimization process, the computationally intensive TV minimization algorithm. This results in increased computational cost of the entire denoising process [42]. In specific applications the need to understand the application environment for optimal performance calls for manual tuning of the parameter until the best peak signal-to-noise ratio (PSNR) of the image is attained [25, 27, 45]. The design and operation of the algorithms incorporate heuristics. For example, in generalized cross validation the minimum number of observations required for optimal performance has to be heuristically determined from experiment by the user. Noise will prevail in the regularized solution if a reasonable number of data points is not chosen as input into the algorithm. Moreover, the algorithm does not have universal application as there are reports of failure in some model parameter selection problems [46]. The performance of discrepancy principle, UPRE and SURE relies on a good estimate of the error level, otherwise there is risk of over-smoothing or retaining noise [1]. Most of current TV regularization parameter selection techniques will be computationally inefficient and useless in MRI-based clinical trials management centers where the daily routine include processing and analysis of several thousands of brain MRI images. Efficient operation in such application environment demands high level of automation from the image processing and image analysis software and little tolerance for manual task.

Our contribution

In this paper we introduce a new concept called the Variational-Bayesian (VB) cycle. Based on this concept we demonstrate that TV and Bayesian techniques are equivalent. Furthermore we describe how and why the noise variance of an image is the optimal regularization parameter. Our proposed method of computing regularization parameter is applied to MRI-based clinical trials where several thousands

of MRI brain images are processed daily. To satisfy the requirement of low computational cost the regularization parameter is computed by a third party algorithm before commencement of the TV denoising process. The equality of regularization parameter and noise variance which we derived from the VB concept was also derived in [47] and [1] using classical approach. The authors analyzed different classical methods for selection of regularization parameters. The methods are cross validation (CV), mean square error (MSE), predicted mean square error (PMSE), equivalent degrees of freedom (EDF), maximum likelihood (ML) and set theory (ST). The analysis shows that for a special choice of the regularization operator all the aforementioned methods yield the same value of λ equal to the noise variance σ^2:

$$\lambda_{\mathbf{MSE}} = \lambda_{\mathbf{PMSE}} = \lambda_{\mathbf{CV}} = \lambda_{\mathbf{EDF}} = \lambda_{\mathbf{ML}} = \lambda_{\mathbf{ST}} = \lambda = \sigma^2 \tag{5}$$

The special choice is expressing the regularization functional $\mathbf{Q}$ as periodogram-based estimate of autocorrelation function $\mathbf{R}_{\mathbf{I_d}}$ of the observed image $\mathbf{I_d}$:

$$\mathbf{Q}\mathbf{Q}^{-1} = \mathbf{R}_{\mathbf{I_d}} \tag{6}$$

The periodogram-based estimate is the equivalence of Fourier transform of spatial domain autocorrelation function of the observed image. For this special choice the denoised or estimate of the underlying image $\hat{I}_c$ is the linear minimum mean square error (LMMSE) solution of Wiener filter.

Outline of paper

This paper is organized as follows. The next section describes theory and methodology of our proposal. This is followed by experiments and display of experimental results. The results are discussed before conclusion of this paper.

THEORY AND METHOD

This section begins with the concept of VB cycle. It explains the equivalence of TV technique and the technique of Markov random field (MRF) model within Bayesian framework. Furthermore it describes how to *navigate* from TV technique to MRF-Bayesian technique and vice-versa. The concept was used to derive value of regularization parameter. Thereafter, we describe how the regularization parameter is computed.

The variational-bayesian cycle

The focus of researchers on the computational speed of TV in different applications has led to generalizations and variations of the original TV denoising problem formulated by ROF [48]. In this paper we choose to adopt the original TV problem formulation expressed as a convex functional in [23] for our proposed method:

$$\hat{I}_c = \min_{I_c} \sum_{i=1}^{M}\sum_{j=1}^{N} \left\| (\nabla \mathbf{I_d})_{(i,j)} \right\| + \frac{1}{2\lambda} \left\| I_c - I_d \right\|^2 \tag{7}$$

The notations have same meaning as in Eq. (3). Natural logarithmic transformation on Eq. (7) maintains the equality of the right hand side (RHS) and left hand side (LHS) terms of the equation and also retains the convexity of the functional:

$$\ln[\exp-(\hat{I}_c)] = \min_{I_c} \ln\left[\exp-\left(\sum_{i=1}^{M}\sum_{j=1}^{N} \left\| (\nabla \mathbf{I_d})_{(i,j)} \right\| + \frac{1}{2\lambda}(I_c - I_d)^2 \right) \right] \tag{8}$$

Now we separate variables on RHS of Eq. (8) into product of two exponential functions:

$$\ln[\exp-(\hat{I}_c)] = \min_{I_c} \ln\left[\exp-\left(\sum_{i=1}^{M}\sum_{j=1}^{N} \left\| (\nabla \mathbf{I_d})_{(i,j)} \right\| \right) \exp-\left(\frac{1}{2\lambda}(I_c - I_d)^2 \right) \right] \tag{9}$$

With reference to our previous work [49] the first term on the RHS of Eq. (9) is the single layer Markov random field energy $U(f)$ expressed as a function of the image pixel configuration f:

$$U(f) = \sum_{i=1}^{M}\sum_{j=1}^{N} \left(\left\| (\nabla I(f))_{(i,j)} \right\| \right) \tag{10}$$

We insert this term into Eq. (9):

$$\ln[\exp-(\hat{I}_c)] = \min_{I_c} \ln\left[\exp-(U(f)) \exp-\left(\frac{1}{2\lambda}(I_c - I_d)^2 \right) \right] \tag{11}$$

Expressing RHS and LHS terms of Eq. (11) as strictly exponential functions maintains equality of both sides of the equation but the functional changes from a convex functional to a concave functional. Thus optimization criteria of the functional changes from minimization to maximization:

$$[\exp-(\hat{I}_c)] = \min_{I_c} \left[\exp-(U(f)) \exp-\left(\frac{1}{2\lambda}(I_c - I_d)^2 \right) \right] \tag{12}$$

In the classical Markov random field model [50] the Gibbs distribution $P(f)$, the probability distribution of each possible configuration of the prior I_c is proportional to the first exponential term on the RHS of Eq. (12):

$$P(f) = P(I_c) \propto [\exp-(U(f))] \tag{13}$$

The Gaussian distribution P_N, the probability distribution assumed for the observed image I_d given any realization of I_c is proportional to the second exponential term on the RHS of Eq. (12):

$$P_N = P(I_d|I_c) \propto \left[\exp - \left(\frac{(I_c - I(f))^2}{2\lambda}\right)\right] \tag{14}$$

Thus we have:

$$\exp - \left[\frac{(I_c - I(f))^2}{2\lambda}\right] \equiv \exp - \left[\frac{(\mu - \mathbf{X})^2}{2\sigma^2}\right] \tag{15}$$

where the mean μ of the Gaussian distribution is the underlying image I_c, the observed image I_d is the variable $\mathbf{X}$ of the Gaussian distribution. We conclude that the regularization parameter λ is equal to the noise variance σ^2 of the image:

$$\lambda = \sigma^2 \tag{16}$$

If we insert the probability notations on the RHS of Eq. (13) and Eq. (14) into RHS of Eq. (12) we obtain Bayes posterior probability $P(I_c|I_d)$ formula [51]:

$$P(I_c|I_d) \propto [P(I_d|I_c)P(I_c)] \tag{17}$$

Each possible configuration $f \in F$, where F is a discrete set of random variables, is independent and identically distributed. Thus Eq. (17) can be expressed as

$$P(I_c|I_d) \propto \left\{\left[\prod_{f\in F} \exp - (U(f))\right]\left[\prod_{f\in F} \exp - \left(\frac{1}{2\sigma^2}(I_c - I(f))^2\right)\right]\right\} \tag{18}$$

and we arrive at the Bayesian formulation of the classical Markov random field model for image restoration introduced by Geman and Geman [52]:

$$\begin{aligned} \hat{I}_c &= \arg\max_{I_c} \; P(I_c|I_d) \\ &= \arg\max_{I_c} \; [P(I_d|I_c)P(I_c)] \\ &= \arg\max_{I_c} \; \left\{\left[\prod_{f\in F} \exp - (U(f))\right]\left[\prod_{f\in F} \exp - \left(\frac{1}{2\sigma^2}(I_c - I(f))^2\right)\right]\right\} \end{aligned} \tag{19}$$

This formulation estimates the denoised image $\hat{I}_c$ from the image configuration which maximizes $P(I_c|I_d)$ in what is referred to as maximum a posteriori probability (MAP). The probability distributions are functions of the image pixel configurations. The configurations are determined by the level of noise and the level of noise is a function of the Markov random field energy.

MAP can be reformulated by adopting a two-step process. The first step is to regard MAP as minimization of the negative exponential terms in Eq. (19). The second step is natural logarithm transformation of the resulting exponential function:

$$\hat{I}_c = \arg\min_{I_c} \left\{U(f) + \frac{(I_c - I(f))^2}{2\sigma^2}\right\} \tag{20}$$

Replacing the image energy in the first term of Eq. (20) with the term on the RHS of Eq. (10) and substituting the expression for σ in Eq. (16) into the second term of Eq. (20):

$$\hat{I}_c = \arg\min_{I_c} \left[\sum_i \sum_j \left\|(\nabla I(f))_{(i,j)}\right\| + \frac{1}{2\lambda}\left\|I_c - I(f)\right\|^2\right] \tag{21}$$

which is a return back to variational mode of denoising expressed in Eq. (7), thus completing a full circular path which we refer to as the VB. Graphical expression of VB is shown in Fig. 1.

Method

Given a brain MRI image, its Markov random field energy expressed by total clique potential energy E is computed. The value of the energy is used to estimate variance of the image from the mathematical model that describes relationship between Markov random field energy and noise variance for brain MRI images [49]:

$$E = a\hat{\sigma}^b + c \tag{22}$$

where a, b, c are model parameters that have different values a_b, b_b, c_b and a_{fg}, b_{fg}, c_{fg} for MRI image having background

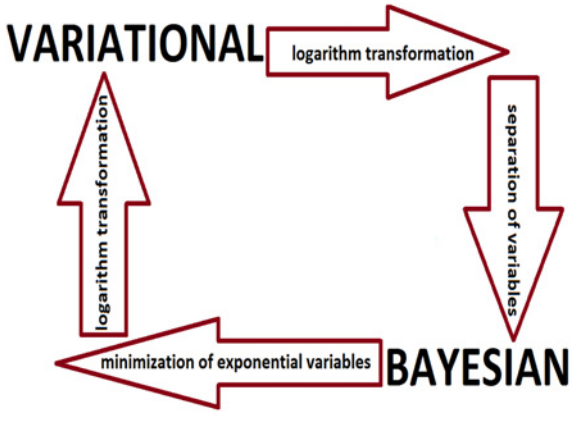

Fig. 1. The Variational-Bayesian (VB) cycle. Logarithmic transformation of the variational problem formulation followed by separation of exponential variables results in Bayesian problem formulation. A return to the variational problem formulation is by minimization of the exponential variables in Bayesian problem formulation followed by logarithmic transformation.

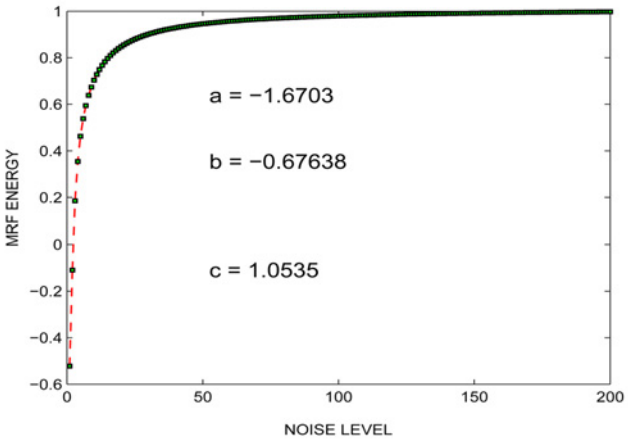

Fig. 2. Plot of mathematical model that describes relationship between Markov random field energy and noise level for magnetic resonance images of the brain with background pixels.

and foreground pixels respectively. For brain MRI images with background pixels $a_b = -1.67$, $b_b = -0.6764$, $c_b = 1.053$. Plot of the mathematical model is shown in Fig. 2. The variance of the image is further used to estimate, at no cost, the regularization parameter according to Eq. (16), before commencement of the TV denoising process.

EXPERIMENTS AND RESULTS

Experiments design

Performance evaluation of our proposed algorithm was carried out using two sets of real magnitude MRI data with different MRI acquisition protocols. Sets labeled A and B were obtained from two international clinical trial study management centers, NeuroRx research and Alzheimer's Disease Neuroimaging Initiative (ADNI). Data from both sources were processed T2, T1, FLAIR and PD weighted images formatted as 16 bit unsigned integer DICOM file. The dimension of each slice is 256 × 256 and 256 × 228 for NeuroRx and ADNI respectively. For each subject MRI data only the useful slices were used as test images. The useless slices are the slices located towards the most superior and inferior sections of the head and are dominated by partial volume of nonbrain structures. We assume that each MRI slice is noise free, hence for evaluation, we induced artificial noise. For a m percent Rician noise level the standard deviation of the equivalent normal distribution is given by

$$\sigma \approx N\left(0, \frac{\tau m}{100}\right) \tag{23}$$

where τ is the maximum pixel intensity [53]. The algorithms were evaluated using MRI slices from all the four types of MRI acquisitions and noise levels in the range 0 percent ($\approx \sigma = 0$) to 30 percent ($\approx \sigma = 75$). However, to satisfy the requirement of page limit for this journal we display only T2 weighted images for Rician noise level of 10 percent ($\approx \sigma = 25$). The source code of the TV algorithm was the implementation reported in [54], and the algorithm runs through 100 iterations for each test image. The sources of MRI data are described below.

Experimental data - part A - NeuroRx Research Inc.

The first part of our experiments presents evaluation results using data from NeuroRx Research (http://www.neurorx.com/en/home.htm). NeuroRx was established in 2003 as a clinical research organization. NeuroRx is dedicated to working with the pharmaceutical industry to facilitate clinical trials of new drugs for multiple sclerosis (MS) and other neurological diseases. NeuroRx provides professional management of all MRI-related study activities and promptly delivers precise MRI outcome measurements that are performed in a regulatory compliant environment. The organization specializes in logistics of scan handling and tracking and can provide this service exclusively, if needed. NeuroRx uses advanced image analysis techniques to provide precise outcome data that maximize study power. Images are corrected for inhomogeneity and co-registered for perfect re-alignment and increased precision. Analyses are conducted in 3D, rather than on slices, so that information can be properly related to structures that span multiple slices. Customized automatic segmentation techniques are combined with expert supervision to maximize the precision of outcome measures related to both lesional and non-lesional pathology, as well as brain volume changes.

The CEO and president of NeuroRx is Douglas Arnold, MD. Douglas Arnold is currently Professor, Department of Neurology and Neurosurgery at McGill University and Director of the Magnetic Resonance Spectroscopy Unit in the Brain Imaging Center at the Montreal Neurological Institute.

Experimental data - part B - Alzheimer's Disease Neuroimaging Initiative

Data used in the preparation of the second part of this experiment were obtained from the ADNI database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI)

and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

Results

Fig. 3 displays images of a T2-weighted MRI slice from NeuroRx. The slice is indexed as slice number 32 in a single subject MRI data consisting of 45 slices. The original, degraded and denoised images are in Fig. 3a, Fig. 3b, and Fig. 3c, respectively. Fig. 4a and Fig. 4b are the convergence results, in terms of MSE, for the single slice of Fig. 3 and 31 MRI slices in the single subject MRI data, respectively.

The images in Fig. 5 are visual quality assessment to demonstrate optimality of the computed regularization parameter. In the figures are outputs of the TV algorithm and its MSE convergence plots for regularization parameter scaled to 10 percent, 100 percent and 190 percent of the computed value of λ. Graphical description of the optimality of the computed regularization parameter in terms of MSE convergence is shown in Fig. 6 for computed regularization parameter scaled to seven different values, from 0.1λ to 1.9λ

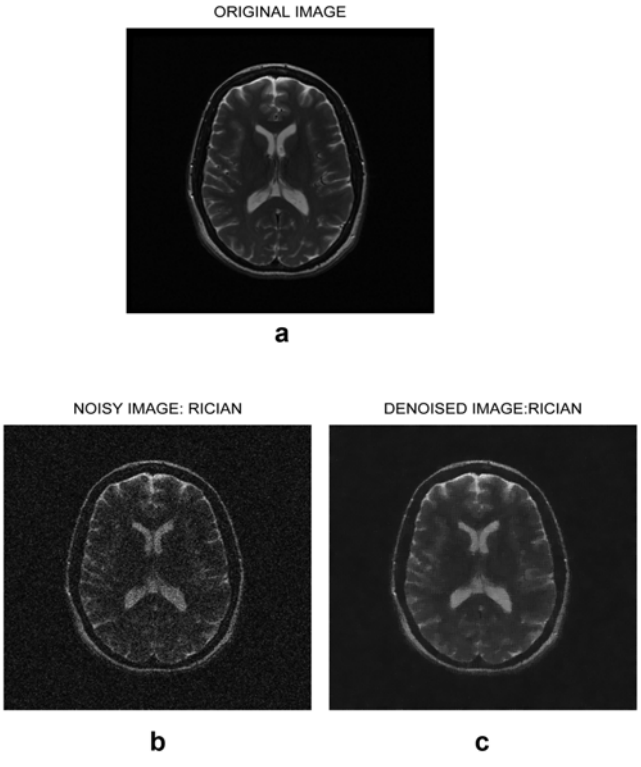

Fig. 3. A T2-weighted MRI slice image from NeuroRx Research in its (a) original state of acquisition (b) degraded state by noise level of $\sigma = 25$ (c) denoised version using TV algorithm with $\lambda = 25$.

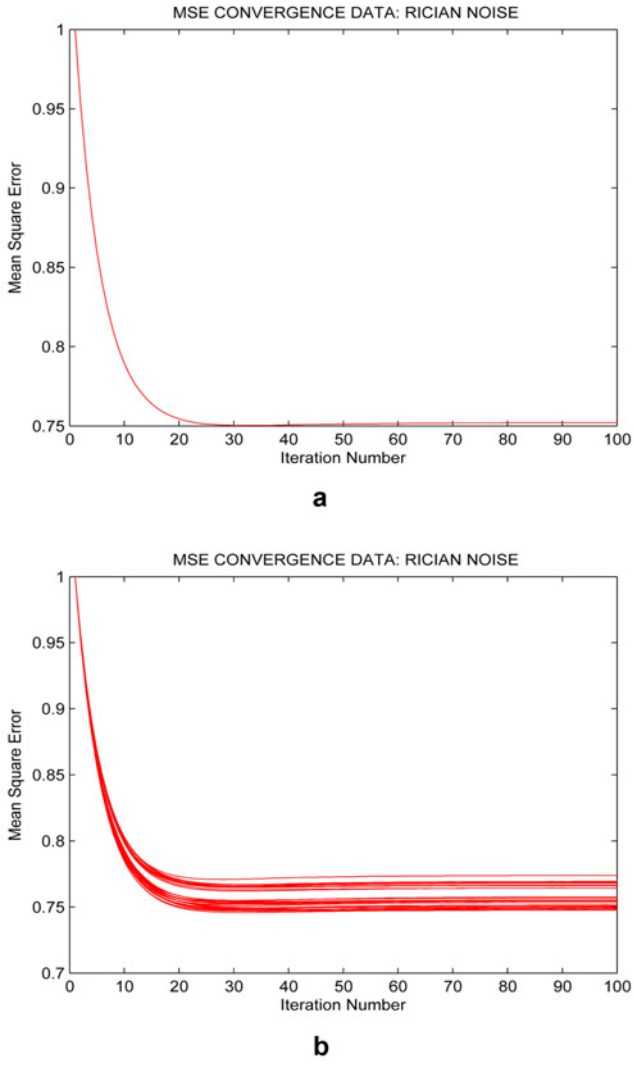

Fig. 4. Mean square error convergence test of TV algorithm for 100 iterations in the denoising ($\sigma = 25$) of (a) single MRI slice image shown in Fig. 3b (b) 31 MRI slice images of a single subject (including the image shown in Fig. 3b).

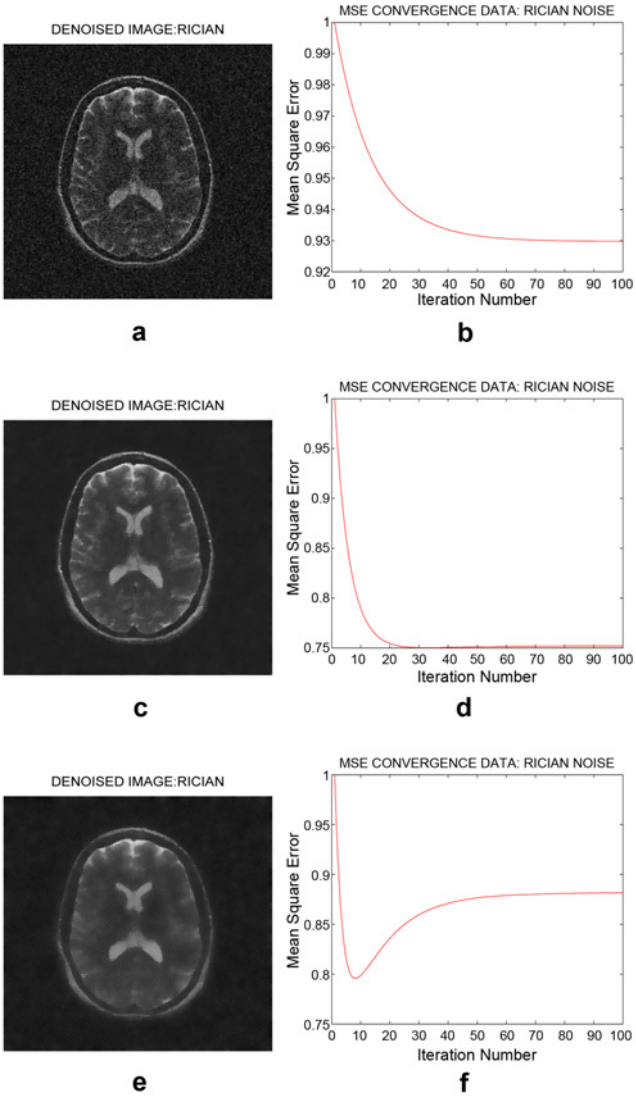

Fig. 5. Three NeuroRx images shown in (a), (c) and (e) were denoised using TV algorithm with regularization parameters of $\lambda = 0.1\lambda_{Op}$, $\lambda = \lambda_{Op}$ and $\lambda = 2\lambda_{Op}$ respectively where $\lambda_{Op} = 25$ is the computed optimal regularization parameter. Corresponding mean square error convergence test results are shown in (b), (d) and (f) respectively.

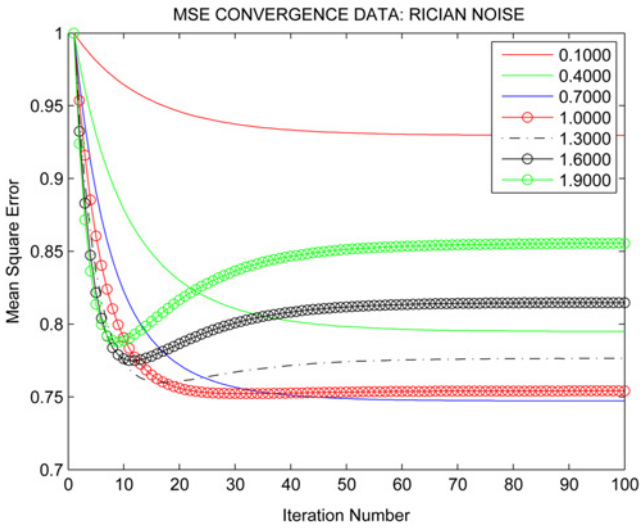

Fig. 6. Test of optimality of computed regularization parameter. The MRI slice image from NeuroRx shown in Fig. 3b was denoised using TV algorithm with the computed optimal regularization parameter scaled from 0.1 to 2 at interval of 0.3.

in steps of 0.3.

In Fig. 7 are images of a T2-weighted MRI slice from ADNI. The slice is indexed as slice number 28 in a single subject MRI data consisting of 42 slices. The original, degraded and denoised images are in Fig. 7a, Fig. 7b and Fig. 7c respectively. Fig. 8a and Fig. 8b are the convergence results in terms of MSE for the single slice of Fig. 7 and 26 MRI slices in the single subject MRI data respectively. Images for visual quality assessment of the optimality of the computed regularization parameter are displayed in Fig. 9. These figures display output of the TV algorithm and its

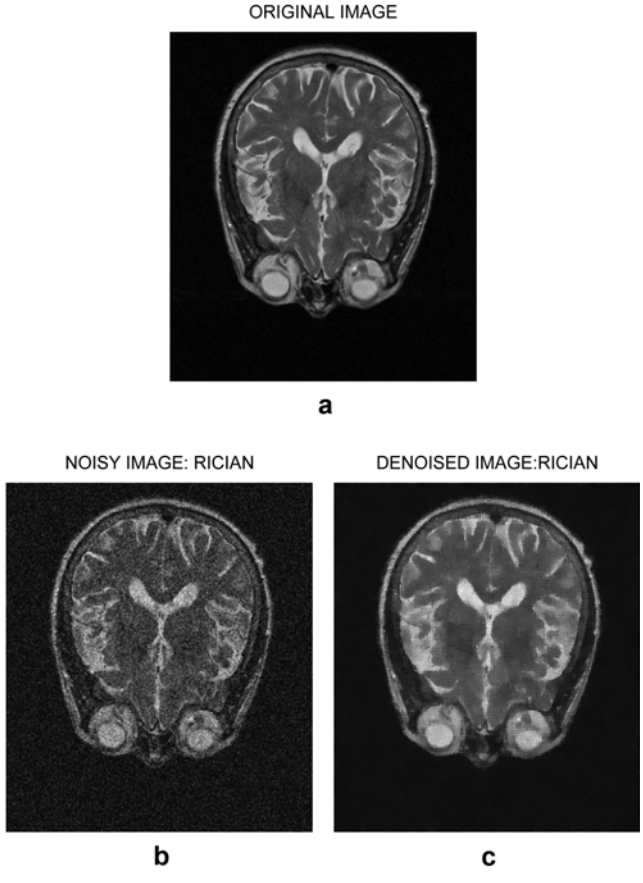

Fig. 7. A T2-weighted MRI slice image from ADNI in its (a) original state of acquisition (b) degraded state by noise level of $\sigma = 25$ (c) denoised version using TV algorithm with $\lambda = 25$.

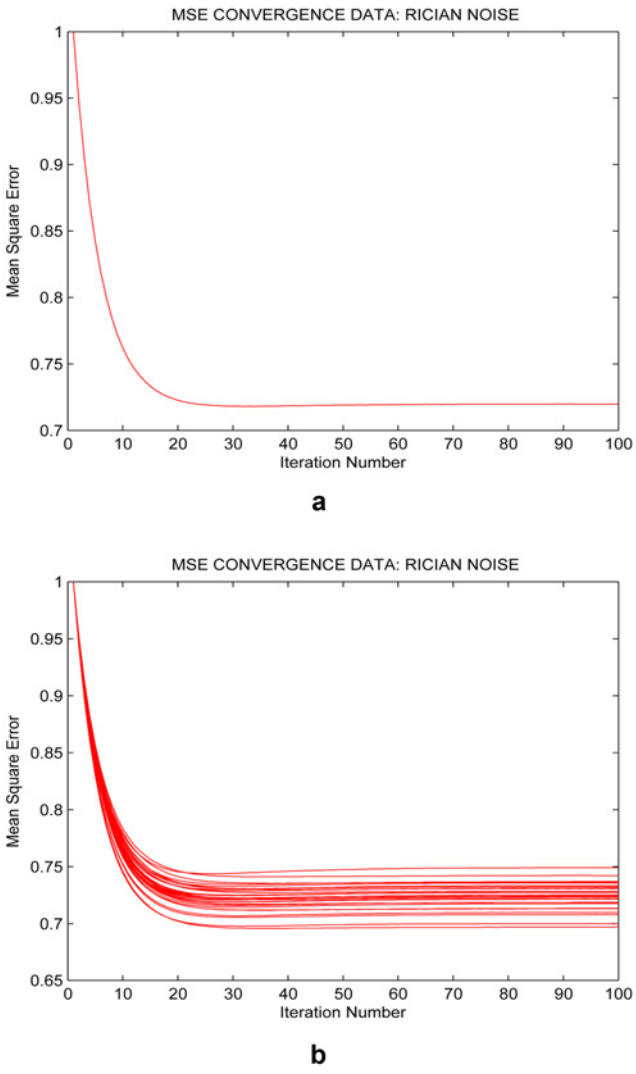

Fig. 8. Mean square error convergence test of TV algorithm for 100 iterations in the denoising ($\sigma = 25$) of (a) single MRI slice image shown in Fig. 7b (b) 26 MRI slice images of a single subject (including the image shown in Fig. 7b).

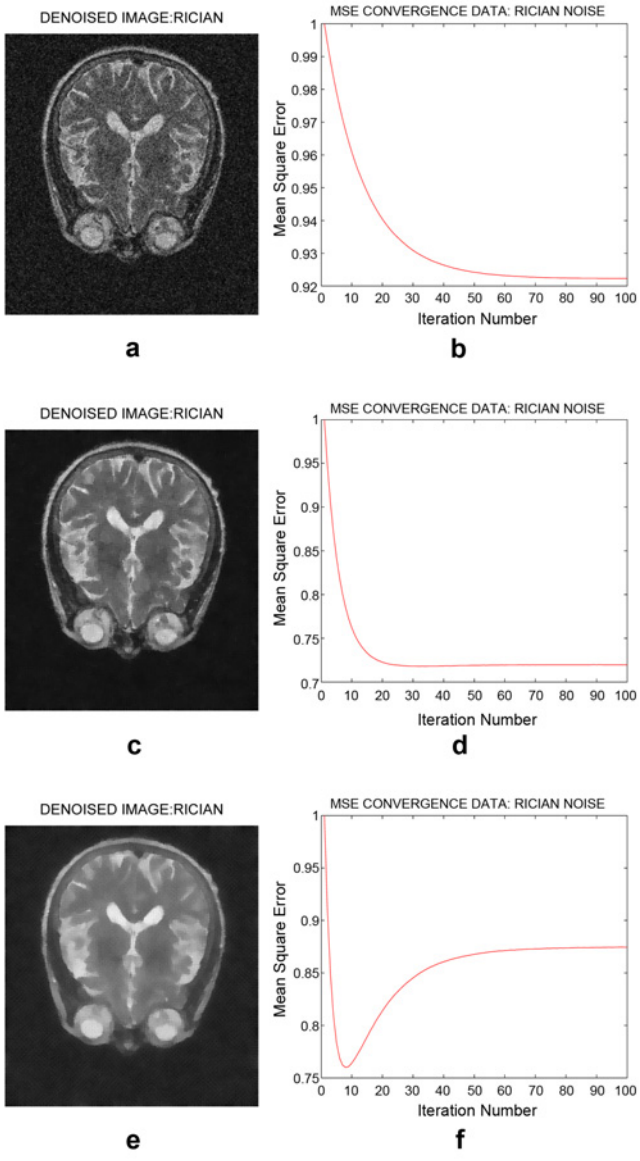

Fig. 9. Three ADNI images shown in (a), (c) and (e) were denoised using TV algorithm with regularization parameters of $\lambda = 0.1\lambda_{Op}$, $\lambda = \lambda_{Op}$ and $\lambda = 2\lambda_{Op}$ respectively where $\lambda_{Op} = 25$ is the computed optimal regularization parameter. Corresponding mean square error convergence test results are shown in (b), (d) and (f) respectively.

convergence in terms of MSE for regularization parameter scaled to 10 percent, 100 percent and 190 percent of the computed value of λ. Graphical description of the optimality of the computed regularization parameter in terms of MSE convergence is shown in Fig. 10 for regularization parameter scaled to seven different values from 0.1λ to 1.9λ in steps of 0.3. The computation time of the TV algorithm in denoising each slice image in each single subject MRI data from NeuroRx and ADNI is displayed in Table 1.

DISCUSSION

This section is in two parts. The first part is evaluation of our proposal. The second part is the limitations of total variation technique and future work.

Evaluation

The evaluation was based on four criteria: visual quality assessment, mean square error convergence, optimality of the computed regularization parameter and computation time.

Visual quality assessment

The grainy characteristics of Rician noise is absent in the

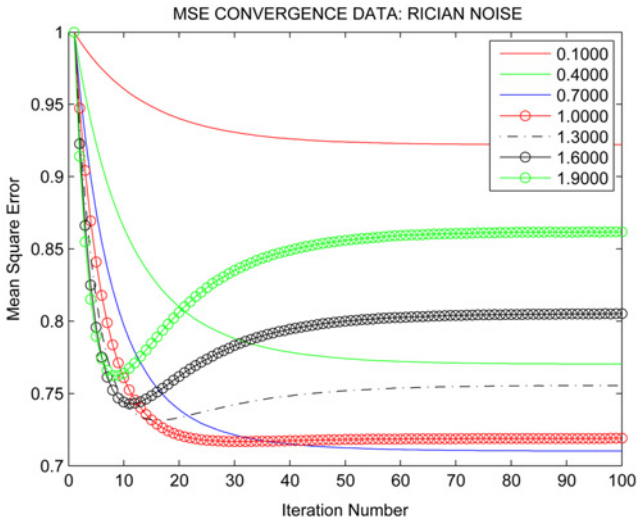

Fig. 10. Test of optimality of computed regularization parameter. The MRI slice image from ADNI shown in Fig. 7b was denoised using TV algorithm with the computed optimal regularization parameter scaled from 0.1 to 2 at interval of 0.3.

denoised images shown in Fig. 3c and Fig. 7c when compared to their noisy versions in Fig. 3b and Fig. 7b, respectively. Thus, we conclude that the TV algorithm significantly improves the visual quality of the degraded images.

Mean square error convergence test

The MSE data is computed relative to the MSE of the degraded image such that the degraded image has MSE value of 1. As shown in Fig. 4a the algorithm recorded about 25 percent improvement in MSE value, after 100 iterations, in the denoising of the single MRI slice shown in Fig. 3b. For the same number of iterations the algorithm improved the MSE of 31 MRI slices in a single subject MRI data by between 20 percent to 25 percent as shown in Fig. 4b.

For the corresponding ADNI data the algorithm recorded close to 30 percent improvement in MSE for a single MRI slice (see Fig. 8a) and between 25 percent to 30 percent for the 26 slices in the single subject MRI data (Fig. 8b).

Optimality test

The plot in Fig. 5b shows that for $\lambda = 0.1\lambda_{Op}$ where λ_{Op} is the computed regularization parameter there is less than 10 percent decrease in the MSE of the degraded image in Fig. 3b compared to its denoised version in Fig. 5a. This slight decrease is reflected in the visual quality of the image as it is characterized by the grainy features seen in its degraded version. For $\lambda = 1.9\lambda_{Op}$ the image shown in Fig. 5e is significantly blurred. This visual quality is reflected in the MSE convergence plot in Fig. 5f where it can be seen that the high value of regularization parameter reduced the MSE by about 20 percent in less than 10 iterations and the process of degradation began immediately until the MSE value is reduced to less than 15 percent. The output of the TV algorithm for optimal regularization parameter $\lambda = \lambda_{Op}$ shown in Fig. 5c and its corresponding MSE convergence plot in Fig. 5d which indicates 25 percent reduction in MSE has the best visual quality when compared to the TV output for $\lambda = 0.1\lambda_{Op}$ and $\lambda = 1.9\lambda_{Op}$.

The image and the plot in Fig. 9a and Fig. 9b respectively shows that for $\lambda = 0.1\lambda_{Op}$ there is less than 10 percent reduction in MSE of the degraded image shown in Fig. 7b. The denoised image exhibit characteristics that are strongly similar to its degraded version. The output of the algorithm

Table 1. Computation time (in seconds) of total variation algorithm in the denoising of each slice in a single subject MRI data from NeuroRx (31 slices) and ADNI (26 slices).

Total variation algorithm computation time					
MRI slice number	NeuroRx	ADNI MRI	Slice number	NeuroRx	ADNI
1	12	12	17	10	12
2	13	12	18	10	12
3	11	12	19	10	12
4	10	12	20	11	12
5	15	12	21	10	10
6	10	12	22	10	11
7	10	11	23	10	12
8	12	11	24	10	11
9	11	11	25	10	11
10	11	11	26	10	11
11	12	11	27	12	NA
12	11	11	28	12	NA
13	11	11	29	12	NA
14	11	11	30	12	NA
15	11	11	31	12	NA
16	11	11	32	NA	NA

for $\lambda = 1.9\lambda_{Op}$ is blurry as shown in Fig. 9e. The profile of the MSE convergence plot in Fig. 9f exhibit instability similar to the plot in Fig. 5f. The output of the TV algorithm for optimal regularization parameter $\lambda = \lambda_{Op}$ shown in Fig. 9c and its corresponding MSE convergence plot in Fig. 9d indicates about 27 percent reduction in MSE in about 20 iterations, and this value of MSE is maintained up to 100 iterations. Thus, TV output corresponding to $\lambda = \lambda_{Op}$ has the best visual quality when compared to the TV output for $\lambda = 0.1\lambda_{Op}$ and $\lambda = 1.9\lambda_{Op}$.

Optimality test in terms of MSE using regularization parameters generated from the set $\{0.1\lambda_{Op} : 0.3 : 2\lambda_{Op}\}$ are the plots shown in Fig. 6 and Fig. 10. A cursory look at the plot indicates that three regularization parameter values $0.7\lambda_{Op}$ (blue colored solid line), λ_{Op} (red colored solid line with circle) and $1.3\lambda_{Op}$ (black colored dash line) are candidates for the optimal value. The plots identifying $1.3\lambda_{Op}$ in Fig. 6 and Fig. 10 shows 23 percent and 27 percent decrease respectively in MSE in about 10 iterations but lost its steam afterwards by reversing its earlier gains to 20 percent and 25 percent respectively after 100 iterations. The plots identifying the parameter $0.7\lambda_{Op}$ recorded 25 percent (see Fig. 6) and 30 percent (see Fig. 10) decrease in MSE after 40 iterations (see Fig. 6) and (Fig. 10). We conclude that the plot identifying λ_{Op} is the optimal value of λ because it reduced the MSE of the degraded image by 25 percent (see Fig. 6) and by about 30 percent (see Fig. 10) in 20 iterations which is half the number of iterations by the parameter $0.7\lambda_{Op}$ to attain same level of performance, and it maintained this same level of performance up to 100 iterations.

Computation time

The results shown in Table 1 indicates that the TV algorithm takes an average of 12 seconds to denoise an MRI slice image. This translates to denoising 1000 MRI slice images in less than three and a half hour.

Limitations of TV denoising technique

The main setback of TV technique is the transformation of smooth regions within the image into piecewise constant regions, a phenomenon referred to as staircasing effect [55]. This effect, which becomes increasingly significant with higher levels of noise in the degraded image, is consequence of the assumption that the underlying image is piecewise smooth. This assumption is not the ideal description for natural images such as brain MRI images. Cursory view of TV denoised images displayed in Fig. 3c, Fig. 5c, Fig. 7c and Fig. 9c shows that staircasing effect is much more reduced in the foreground regions which is useful in clinical diagnosis compared to the background regions where the staircasing effect is significant. The background regions contain no information, and are therefore irrelevant in clinical diagnosis. How to overcome staircasing effect in TV technique is outside the scope of this paper. However, detailed explanation on how and why staircasing effect is observed in TV denoised images, and how to address the problem can be found in [56-59].

CONCLUSION

On this paper we analyze the relationship between total variation and Bayesian problem formulations for noise reduction in images. Results of the analysis show that both techniques are equivalent, a notion that has been held for a long time within the image processing and computer vision community. Based on this equivalency we derive the value of TV regularization parameter as equal to the noise variance of the test image. We evaluated TV algorithm including our proposed method of computing the regularization parameter to noise reduction of images in two clinical trial study management centers. The performance evaluation result show that our proposed method for computation of regularization parameter makes TV algorithm computationally efficient and the quality of the denoised images are optimized by the computed regularization parameter. These feature makes our proposal suitable in application environment where there is little tolerance for manual task and high level of automation is demanded from image processing and image analysis software.

ACKNOWLEDGEMENTS

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is

providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for NeuroImaging at the University of California, Los Angeles.

CONFLICT OF INTEREST STATEMENTS

Osadebey M declares that he has no conflict of interest in relation to the work in this article. Bouguila N declares that he has no conflict of interest in relation to the work in this article. Arnold D declares that he has no conflict of interest in relation to the work in this article.

REFERENCES

[1] Galatsanos NP, Katsaggelos AK. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. IEEE T Image Process. 1992; 1(3):322-36.
[2] Zhang Y, Brady M, Smith S. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. IEEE T Med Imaging. 2001; 20(1): 45-57.
[3] Osadebey M. Simulation of realistic head geometry using radial vector representation of magnetic resonance image data. Masters thesis; Tampere University of Technology; Finland; 2009.
[4] Gold R, Kappos L, Arnold DL, Bar-Or A, Giovannoni G, Selmaj K, Tornatore C, Sweetser MT, Yang M, Sheikh SI, Dawson KT. Placebo-controlled phase 3 study of oral bg-12 for relapsing multiple sclerosis. New Engl J Med. 2012; 367(12):1098-107.
[5] Thulborn KR, Uttecht SD. Volumetry and topography of the human brain by magnetic resonance. Int J Imag Syst Tech. 2000; 11(3):198-208.
[6] Wiest-Daessle N, Prima S, Coupe P, Morrissey S, Barillot C. Rician noise removal by non-Local Means filtering for low signal-to-noise ratio MRI: applications to DT-MRI. Med Image Comput Comput Assist Interv. 2008; 11(Pt 2):171-9.
[7] Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Physica D. 1992; 60:259-68.
[8] Ching WK, Ng MK, Sze KN, Yau AC. Superresolution image reconstruction from blurred observations by multisensors. Int J Imag Syst Tech. 2003; 13(3):153-60.
[9] Feng J, Zhang JZ. An adaptive dynamic combined energy minimization model for few-view computed tomography reconstruction. Int J Imag Syst Tech. 2013; 23(1):44-52.
[10] Zhang Y, Zhang WH, Chen H, Yang ML, Li TY, Zhou JL. Few-view image reconstruction combining total variation and a highorder norm. Int J Imag Syst Tech. 2013; 23(3):249-55.
[11] Zhu Y, Shi Y. A fast method for reconstruction of total-variation mr images with a periodic boundary condition. IEEE Signal Process Lett. 2013; 20(4):291-4.
[12] Vogel CR, Oman ME. Fast, robust total variation-based reconstruction of noisy, blurred images. IEEE T Image Process. 1998; 7(6):813-24.
[13] Chan SH, Khoshabeh R, Gibson KB, Gill PE, Nguyen TQ. An augmented lagrangian method for total variation video restoration. IEEE T Image Process. 2011; 20(11):3097-111.
[14] He L, Marquina A, Osher SJ. Blind deconvolution using tv regularization and bregman iteration. Int J Imag Syst Technol. 2005; 15(1):74-83.
[15] Chan TF, Yip AM, Park FE. Simultaneous total variation image inpainting and blind deconvolution. Int J Imag Syst Technol. 2005; 15(1):92-102.
[16] Guo W, Qiao LH. Inpainting based on total variation. Int Conf Wavelt Anal Pattern Recognit. 2007; 2:939-43.
[17] Strong D, Chan T. Edge-preserving and scale-dependent properties of total variation regularization. Inverse Probl. 2003; 19(6):S165.
[18] Bellettini G, Caselles V, Novaga M. The total variation flow in $\mathbb{R}^N$. J Differ Equations. 2002; 184(2):475-525.
[19] Chambolle A, Lions PL. Image recovery via total variation minimization and related problems. Numer Math. 1997; 76(2):167-88.
[20] Dobson DC, Santosa F. Recovery of blocky images from noisy and blurred data. SIAM J Appl Math. 1996; 56(4):1181-98.
[21] Vogel CR, Oman ME. Iterative methods for total variation denoising. SIAM J Sci Comput. 1996; 17(1):227-38.
[22] Lysaker M, Osher S, Tai XC. Noise removal using smoothed normals and surface fitting. IEEE T Image Process. 2004; 13(10):1345-57.
[23] Chambolle A. An algorithm for total variation minimization and applications. J Math Imaging Vision. 2004; 20(1-2):89-97.
[24] Chan TF, Golub GH, Mulet P. A nonlinear primal-dual method for total variation-based image restoration. SIAM J Sci Comput. 2006; 20(6):1964-77.
[25] Zuo W, Lin Z. A generalized accelerated proximal gradient approach for total-variation-based image restoration. IEEE T Image Process. 2011; 20(10):2748-59.
[26] Beck A, Teboulle M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE T Image Process. 2009; 18(11):2419-34.
[27] Wang Y, Yang J, Yin W, Zhang Y. A new alternating minimization algorithm for total variation image reconstruction. SIAM J Imaging Sci. 2008; 1(3):248-72.
[28] Afonso MV, Bioucas-Dias JM, Figueiredo MA. Fast image recovery using variable splitting and constrained optimization. IEEE T Image Process. 2010; 19(9):2345-56.
[29] Babacan SD, Molina R, Katsaggelos AK. Parameter estimation in tv image restoration using variational distribution approximation. IEEE T Image Process. 2008; 17(3):326-39.
[30] Liao H, Li F, Ng MK. Selection of regularization parameter in total variation image restoration. J Opt Soc Am A. 2009; 26(11):2311-20.
[31] Golub GH, Heath M, Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics. 1979; 21(2):215-23.
[32] Geisser S. Predictive inference : an introduction. Monographs on statistics and applied probability; New York, Chapman and Hall; 1993.
[33] Wen YW, Chan R. Parameter selection for total-variation-based image restoration using discrepancy principle. IEEE T Image Process. 2012; 21(4):1770-81.
[34] Chen A, Huo BM, Wen CW. Adaptive regularization for color image restoration using discrepancy principle. IEEE Int Conf Signal Process Commun Comput. 2013; 1-6.
[35] Engl HW, Hanke M, Neubauer A. Regularization of Inverse Problems. Mathematics and Its Applications; Springer; 1996.

[36] Lin Y, Wohlberg B, Guo H. UPRE method for total variation parameter selection. Signal Process. 2010; 90(8):2546-51.
[37] Mallows CL. Some comments on c_p. Technometrics. 1973; 15(4):661-75.
[38] Bertalmio M, Caselles V, Rouge B, Sole A. Tv based image restoration with local constraints. J Sci Comput. 2003; 19(1-3):95-122.
[39] Almansa A, Ballester C, Caselles V, Haro G. A tv based restoration model with local constraints. J Sci Comput. 2008; 34(3):209-36.
[40] Dong Y, Hintermuller M, Rincon-Camacho MM. Automated regularization parameter selection in multi-scale total variation models for image restoration. J Math Imaging Vis. 2011; 40(1):82-104.
[41] Palsson F, Sveinsson JR, Ulfarsson MO, Benediktsson JA. Sar image denoising using total variation based regularization with surebased optimization of the regularization parameter. IEEE Int Geosci Remote Sensing Symposium. 2012; 2160-3.
[42] Ramani S, Liu Z, Rosen J, Nielsen JF, Fessler JA. Regularization parameter selection for nonlinear iterative image restoration and mri reconstruction using gcv and sure-based methods. IEEE T Image Process. 2012; 21(8):3659-72.
[43] Stein CM. Estimation of the mean of a multivariate normal distribution. Ann Stat. 1981; 9(6):1135-51.
[44] Chan R, Chan T, Yip A. Numerical methods and applications in total variation image restoration. In: Scherzer O, editor. Handbook of Mathematical Methods in Imaging. Springer New York; 2011, p. 1059-94.
[45] Malgouyres F. Minimizing the total variation under a general convex constraint for image restoration. IEEE T Image Process. 2002; 11(12):1450-6.
[46] Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Stat Surv. 2010; 4:40-79.
[47] Galatsanos NP, Katsaggelos AK. Cross-validation and other criteria for estimating the regularizing parameter. Int Conf Acoust Speech Signal Process. 1991; 4:3021-4.
[48] Chan TF, Esedoglu S. Aspects of total variation regularized l1 function approximation. SIAM J Appl Math 2005; 65(5):1817-37.
[49] Osadebey M, Bouguila N, Arnold D, the ADNI. The clique potential of markov random field in a random experiment for estimation of noise levels in 2d brain mri. Int J Imag Syst Tech. 2013; 23(4):304-13.
[50] Li SZ. Markov Random Field Modeling in Image Analysis. Springer; 2009.
[51] Mackay DJC. Information Theory, Inference and learning algorithms. Cambridge: Cambridge University Press; 2003.
[52] Geman S, Geman D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE T Pattern Anal. 1984; 6(6):721-41.
[53] Coupe P, Manjon JV, Gedamu E, Arnold D, Robles M, Collins DL. Robust rician noise estimation for mr images. Med Image Anal. 2010; 14(4):483-93.
[54] Chopra A, Lian H. Total variation, adaptive total variation and nonconvex smoothly clipped absolute deviation penalty for denoising blocky images. Pattern Recogn. 2010; 43(8):2609-19.
[55] Marquina A, Osher S. Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal. SIAM J Sci Comput. 2000; 22(2):387-405.
[56] Chan T, Esedoglu S, Park F, Yip A. Total variation image restoration: Overview and recent developments. In: Paragios N, Chen Y, Faugeras O, editors. Handbook of Mathematical Models in Computer Vision. Springer US; 2006, p. 17-31.
[57] Zeng X, Li S. An efficient adaptive total variation regularization for image denoising. 2013 Seventh Int Conf Image Graph. 2013; 55-9.
[58] Chen Q, Montesinos P, Sun QS, Heng PA, Xia DS. Adaptive total variation denoising based on difference curvature. Image Vision Comput. 2010; 28(3):298-306.
[59] Blomgren P, Chan T, Mulet P, Wong CK. Total variation image restoration: numerical methods and extensions. Proc IEEE Int Conf Image Process. 1997; (3):384-7.