Second Order Total Generalized Variation (TGV) for MRI

Florian Knoll,[1]* Kristian Bredies,[2] Thomas Pock,[3] and Rudolf Stollberger[1]

Total variation was recently introduced in many different magnetic resonance imaging applications. The assumption of total variation is that images consist of areas, which are piecewise constant. However, in many practical magnetic resonance imaging situations, this assumption is not valid due to the inhomogeneities of the exciting B1 field and the receive coils. This work introduces the new concept of total generalized variation for magnetic resonance imaging, a new mathematical framework, which is a generalization of the total variation theory and which eliminates these restrictions. Two important applications are considered in this article, image denoising and image reconstruction from undersampled radial data sets with multiple coils. Apart from simulations, experimental results from in vivo measurements are presented where total generalized variation yielded improved image quality over conventional total variation in all cases. Magn Reson Med 65:480–491, 2011. © 2010 Wiley-Liss, Inc.

Key words: total generalized variation; radial sampling; accelerated imaging; constrained reconstruction; denoising.

Total variation (TV) based strategies, which were originally designed for denoising of images (1) have recently gained wide interest for many MRI applications beyond denoising. Examples include regularization for parallel imaging (2,3), the elimination of truncation artifacts (4), inpainting of sensitivity maps (5), their use as a regularization method for undersampled imaging techniques within the compressed sensing framework (6) as well as in iterative reconstruction of undersampled radial data sets (7,8). TV models have the main benefit that they are very well suited to remove random noise, incoherent noise-like artifacts from random subsampling and streaking artifacts from undersampled radial sampling, while preserving the edges in the image. However, the assumption of TV is that the images consist of regions, which are piecewise constant. Due to inhomogeneities of the exciting B_1 field of high field systems with 3T and above and of the receive coils, this assumption is often violated in practical MRI examinations. Additionally, even in situations when this is no severe problem, due to the assumption of piecewise constancy, the use of TV often leads to staircasing artifacts and results in patchy, cartoon-like images which appear unnatural.

[1]Institute of Medical Engineering, Graz University of Technology, Kronesgasse 5, A-8010 Graz, Austria
[2]Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, A-8010 Graz, Austria
[3]Institute of Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16/II, A-8010 Graz, Austria

*Correspondence to: Florian Knoll, M.Sc., Graz University of Technology, Institute of Medical Engineering, Kronesgasse 5, A-8010 Graz, Austria. E-mail: florian.knoll@tugraz.at

This article introduces the new concept of total generalized variation (TGV) as a penalty term for MRI problems. This mathematical theory has recently been developed (9), and while it is equivalent to TV in terms of edge preservation and noise removal, it can also be applied in imaging situations where the assumption that the image is piecewise constant is not valid. As a result, the application of TGV in MR imaging is far less restrictive. It is shown in this work that TGV can be applied for image denoising and during iterative image reconstruction of undersampled radial data sets from phased array coils, and yields results that are superior to conventional TV.

THEORY

The Concept of Total Generalized Variation

The total generalized variation introduced in (9) is a functional which is capable to measure, in some sense, image characteristics up to a certain order of differentiation. In this section, we will restrict ourselves to give a short introduction which is not rigorous in the mathematical sense. The reader interested in the mathematical background may find more information in the Appendix.

First, recall the definition of the total variation, which is, for a given image u, usually expressed as

$$\text{TV}(u) = \int_\Omega |\nabla u| \, dx, \qquad [1]$$

The total generalized variation of second-order which we will use throughout the paper bases on this definition but is generalized to represent a minimization problem itself:

$$\text{TGV}^2_\alpha(u) = \min_v \ \alpha_1 \int_\Omega |\nabla u - v| \, dx + \alpha_0 \int_\Omega |\mathcal{E}(v)| \, dx \quad [2]$$

Here, the minimum is taken over all complex vector fields v on Ω and $\mathcal{E}(v) = \frac{1}{2}(\nabla v + \nabla v^{\text{T}})$ denotes the symmetrized derivative. Such a definition provides a way of balancing between the first and second derivative of a function (which is controlled by the ratio of positive weights α_0 and α_1). In contrast to that, TV only takes the first derivative into account. For compatibility with the usual TV semi-norm we let, $\alpha_1 = 1$ throughout the paper. Moreover, according to our experience, the default value $\alpha_0 = 2$ is suitable for most applications and does not need to be tuned, hence, one can say that no additional parameters are introduced with TGV.

Let us briefly mention some general properties of the functional. First, as TGV is the semi-norm of a Banach space, associated variational problems fit well into the well-developed mathematical theory of convex optimization problems, especially with respect to analysis and computational realization. Moreover, each function of bounded variation admits a finite TGV value, making the notion

FIG. 1. Denoising a ramp image with TV and TGV penalty. The noisy test image f (left), its total-variation regularization (middle) and its total-generalized-variation regularization (Eq. 3) (right). The absence of staircasing artifacts for TGV can nicely be observed here.
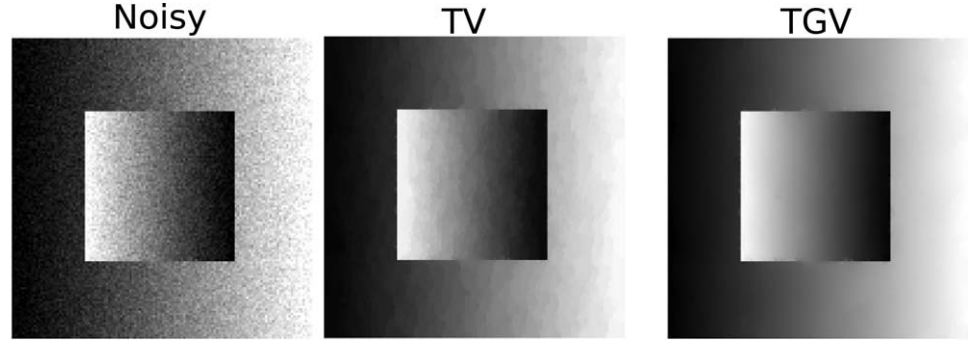
suitable for images. This means in particular that piecewise constant images can be captured with the TGV model which even extends the TV model. Finally, TGV is translation invariant as well as rotationally invariant, meaning that it is in conformance with the requirement that images are measured independent from the actual viewpoint.

It turns out that using TGV^2_α as a regularizer leads to an absence of the staircasing effect which is often observed in total variation regularization. In the following, we like to give an intuitive explanation for this. In smooth regions of a given image u, the second derivative $\nabla^2 u$ is locally "small". It is therefore beneficial for the minimization of Eq. 2 to choose locally $v = \nabla u$ in these regions, resulting locally in the penalization of $\nabla^2 u$. On the other hand, in the neighborhood of edges, $\nabla^2 u$ will be considerably "larger" than ∇u, so it is beneficial for the minimization of Eq. 2 to choose locally $v = 0$ there. This locally results in TGV^2_α measuring the first derivative which corresponds to the "jump," i.e., the difference of the values of u on one side of an edge compared to the other. Of course, this argumentation is only intuitively valid, the values of actual minimizers v may be located anywhere "between" 0 and ∇u. This balancing leads to edges measured in TGV^2_α not being more "expensive" than a smooth function, which is similar to TV, but to smooth regions being "cheaper" than staircases in terms of TGV^2_α, a feature which is absent in TV. Hence, total generalized variation used as a penalty functional prefers images which appear more natural and, as we will see, leads to a more faithful reconstruction of MR images.

Application of TGV in MRI

The effects suggested in the previous subsection can in fact be observed when applied to MRI problems. One example is TGV^2_α-based denoising which is briefly outlined in the following; for a detailed discussion of the noise suppression characteristics and a comparison with TV-based denoising, we refer to (9). Given a noisy image $f \in L^2(\Omega)$, and a regularization parameter $\lambda > 0$ the denoised version u is computed as the solution of the following problem.

Method 1 (TGV Image Denoising).

$$\min_u \ \frac{1}{2\lambda} \int_\Omega (u - f)^2 \, \mathrm{d}x + \mathrm{TGV}^2_\alpha(u). \qquad [3]$$

We propose Eq. 3 to remove noise from MR images which can include smooth signal modulations. Testing this method with a piecewise affine artificial image (see Fig. 1) clearly reveals its advantage over usual TV-based denoising. Here, the original image consists of regions with vanishing second derivative and since TGV^2_α takes into account the second derivative as well as the edges, the method is able to reconstruct almost perfectly.

Besides image denoising, we propose to use TGV^2_α for recovering modulated images from incomplete Fourier measurements, which covers in particular the case of subsampled radial data from multiple coils. It bases on the approach analyzed in (10–12), was applied to radial MRI data in (7) and can be described as follows: Denoting the unknown true data by $f \in L^2(\Omega)$ with Ω being a bounded domain, and by $\Xi \subset \mathbf{R}^d$ a non-empty set representing the spatial frequencies (or k-space points) which are measured, we are looking for a solution of

$$\min_{u \in L^2(\Omega)} \mathrm{TV}(u) \quad \text{such that} \quad \hat{u}(\xi) = \hat{f}(\xi) \quad \text{for all } \xi \in \Xi \quad [4]$$

with $\hat{u}$ and $\hat{f}$ being the Fourier transforms of u and f, respectively. This approach is able to reduce artifacts from the incomplete measurements Ξ which usually appear in the minimum L^2-norm reconstruction. Moreover, surprisingly, for suitable sets Ξ and many piecewise constant f, the solution u of Eq. 4 coincides with f (10). However, the tendency of TV to produce staircasing artifacts in piecewise smooth images is again present. We therefore modify the problem to incorporate TGV^2_α.

$$\min_{u \in L^2(\Omega)} \mathrm{TGV}^2_\alpha(u) \quad \text{such that} \quad \hat{u}(\xi) = \hat{f}(\xi) \quad \text{for all } \xi \in \Xi.$$
$$[5]$$

This approach was tested with a noise-free numerical Shepp-Logan phantom that was modulated with a simulated coil sensitivity generated with the use of Biot-Savart law, to simulate the situation of undersampled radial multichannel acquisition. TGV yields again almost perfect reconstruction where conventional TV (Eq. 4) suffers from staircasing artifacts (see Fig. 2).

For reconstruction from multiple channel data $g_1, \ldots, g_P$, we follow the approach of (7) and assume that there are complex sensitivity estimates $\sigma_1, \ldots, \sigma_P$ such that $\widehat{(\sigma_i u)} \approx g_i$ on Ξ. Relaxing the constraint in Eq. 5 then leads to
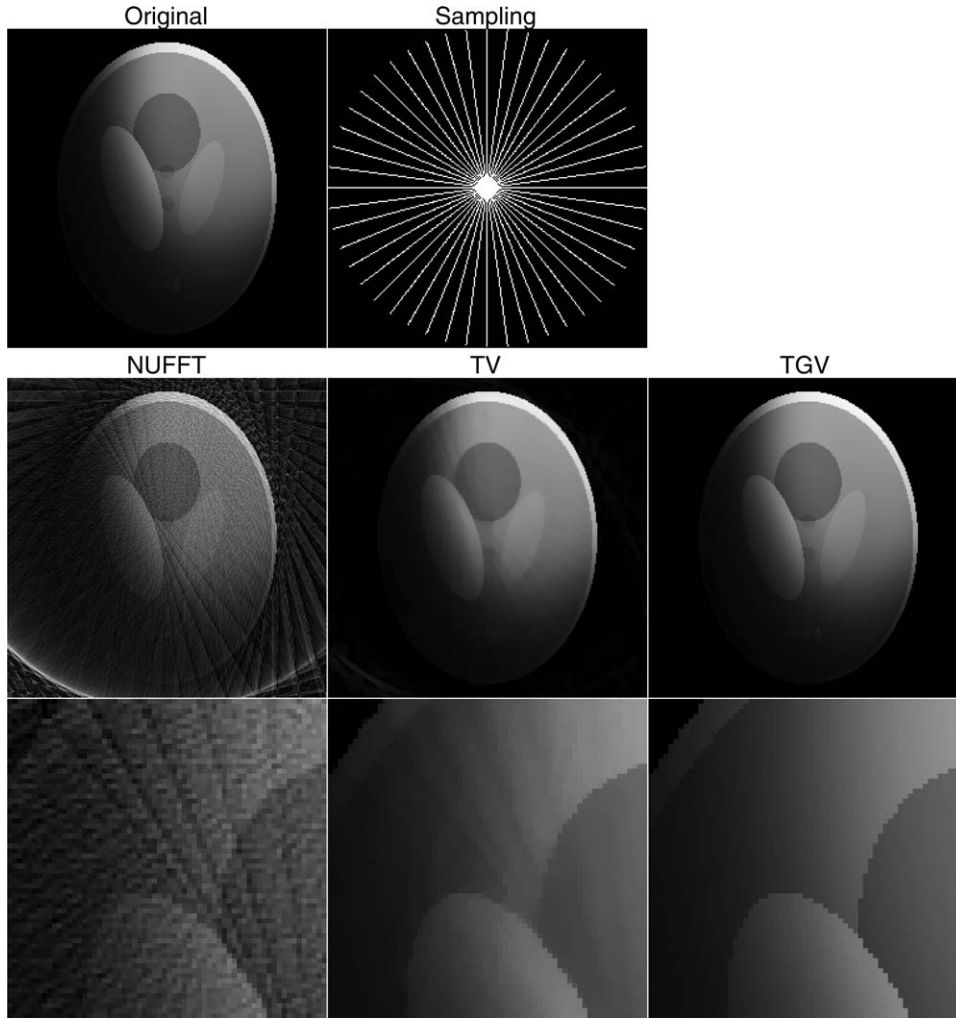
FIG. 2. Reconstructions of a Shepp-Logan phantom from undersampled radial *k*-space data, modulated with a simulated coil sensitivity generated with the use of Biot-Savart law. Top row: Original test image of size 256×256 (left) and the corresponding radial sampling pattern with 24 radial spokes (right). Middle row: conventional NUFFT reconstruction (left), minimum-TV-solution (Eq. 4) (middle) and minimum-TGV$_\alpha^2$-solution (Eq. 6) (right). Bottom row: Magnified views from the middle row.

Method 2 (TGV Undersampling Reconstruction).

$$\min_{u \in L^2(\Omega)} \sum_{i=1}^{P} \frac{\|\widehat{(\sigma_i u)} - g_i\|_2^2}{2\lambda} + \mathrm{TGV}_\alpha^2(u) \qquad [6]$$

where $\|\cdot\|_2$ is a suitable norm on $L^2(\Xi)$ and $\lambda > 0$ is a regularization parameter. We apply Eq. 6 for undersampled multi-channel radial MR data.

Numerical Algorithms

Both the denoising problem Eq. 3 and the undersampling reconstruction problem Eq. 6 constitute non-smooth convex optimization problems for which several numerical algorithms are available. To implement any numerical algorithm on a digital computer, we have to consider the discrete setting of the proposed reconstruction method [6]. For discretization, we use a two dimensional regular Cartesian grid of size $M \times N$ and mesh size $h > 0$. For more information on the discrete setting we refer to (9). From now on we will use the superscript h to indicate the discrete setting. The differential operators $\mathrm{div}^h, \mathcal{E}^h, \nabla^h$ are approximated using first order finite differences. These operators are chosen adjoint to each other, i.e.

$$\left(\mathrm{div}_2^h \right)^* = -\mathcal{E}^h, \qquad \left(\mathrm{div}_1^h \right)^* = -\nabla^h.$$

With $U^h = \mathbf{C}^{MN}$, the discrete version of the denoising problem Eq. 3 is given by

$$\min_{u^h \in U^h} \frac{1}{2\lambda} \|u^h - f^h\|_2^2 + \mathrm{TGV}_\alpha^2(u^h), \qquad [7]$$

where $f^h \in U^h$ is the noisy input data and TGV_α^2 the discrete total generalized variation according to Eq. 2 using the corresponding discrete differentiation operators. Discretized, problem Eq. 7 becomes

$$\min_{u^h \in U^h, \, v^h \in V^h} \frac{1}{2\lambda} \|u^h - f^h\|_2^2 + \alpha_1 \|\nabla^h u^h - v^h\|_1 + \alpha_0 \|\mathcal{E}^h(v^h)\|_1 \qquad [8]$$

where $V^h = \mathbf{C}^{2MN}$. Note that in this formulation, only first order differential operators are involved.

As recent trends in convex optimization (13,14) indicate and as they turn out to be suitable for our purposes, it is sensible to use first order algorithms since they are easy to implement, have a low memory print and can be efficiently parallelized. Furthermore, these methods come along with global worst case convergence rates. This allows to estimate

the required number of iterations which are needed to reach a certain accuracy. Here, we employ the first-order primal dual algorithm of (15,16). Details about how this method is applied can be found in the Appendix.

A discrete version of the reconstruction problem (Eq. 6) can be derived as follows: Let $R^h = \mathbf{C}^{QP}$ represent the vector space associated with Q k-space measurements for P coils and denote by $K^h : U^h \to R^h$ and $(K^h)^*$ the linear operators which realize the NUFFT and its adjoint on the discrete data, respectively. We assume that K^h is normalized such that $\|K^h\| \leq 1$. Furthermore, denote by g^h the measured discrete radial sampling data. The continuous problem (Eq. 6) then becomes

$$\min_{u^h \in U^h} \frac{1}{2\lambda} \|K^h u^h - g^h\|_2^2 + \mathrm{TGV}_\alpha^2(u^h). \qquad [9]$$

The main difference to the denoising problem (Eq. 7) is that the linear operator K^h is involved in the data fitting term. However, the method introduced in (16) can be easily adapted to this setting by introducing an additional dual variable, we refer again to the Appendix for details.

MATERIALS AND METHODS

Application of TGV for Denoising

Denoising was the first natural application for TGV. Two different data sets were acquired on a clinical 3T scanner (Siemens Magnetom TIM Trio, Erlangen, Germany). Written informed consent was obtained from all subjects before the examinations. The first measurement was a T_1 weighted 3D gradient echo scan of the prostate with injection of a contrast agent. Sequence parameters were $T_R = 3.3\,\mathrm{ms}$, $T_E = 1.1\,\mathrm{ms}$, flip angle $\alpha = 15°$, matrix size $(x, y) = (256, 256)$, 20 slices, slice thickness 4 mm, in plane resolution 0.85 mm × 0.85 mm. The measurement showed severe signal inhomogeneities due to the exciting B_1 field and the surface coils used for reception (17). Therefore, the assumption that the images consist of piecewise constant areas is violated.

The other test data set was a T_2 weighted turbo spin echo scan of the human brain, using a 32 channel receive coil. Sequence parameters were repetition time $T_R = 5000\,\mathrm{ms}$, echo time $T_E = 99\,\mathrm{ms}$, turbo factor 10, matrix size $(x, y) = (256, 256)$, 10 slices with a slice thickness of 4 mm and an in plane resolution of 0.86 mm × 0.86 mm. Complex Gaussian noise was then added to the original k-space data. The standard deviation of the added noise was chosen such that the ratio to the norm of the data was 1/15. With this approach, a high SNR gold standard was available which allowed objective comparison of the TV and TGV results. This measurement did not show pronounced signal inhomogeneities, and the goal of the experiments with TGV was to show that even in these cases, TV introduces staircase artifacts which do not represent the underlying signal from the tissue.

Both data sets were exported, and offline denoising was performed with conventional TV, as well as with a Matlab (R2009a, The MathWorks, Natick, MA) implementation of TGV (Eq. 3).

Image Reconstruction of Subsampled Multicoil Radial Data with a TGV Constraint

Image reconstruction with TV regularization for subsampled radial data sets with removal of the intensity profiles and the phase of the individual coil elements before application of TV was first introduced by Block et al. (7). Our TV reference implementation roughly follows this approach. Analogously to (7), it employs, before application of TV, for each coil, a reconstruction with a quadratic regularization on the derivative (H^1-regularization) followed by the convolution with a smoothing kernel. The sensitivities are then obtained by dividing by the sum of squares image for each coil element. All results were also compared to an iterative parallel imaging method of CG-SENSE type (18), using the same coil sensitivities as in the TV experiments.

The fully sampled 32 channel TSE brain data set from the denoising experiments was subsampled retrospectively, to simulate an accelerated acquisition. Data from 32, 24, 16, and 12 radial projections was obtained. As $\frac{\pi}{2}n$ projections (402 for $n = 256$ in our case) have to be acquired to obtain a fully sampled data set in line with the Nyquist criterion according to the literature (19), this corresponds to undersampling factors of ~12, 16, 25, and 33.

Finally, undersampled 2D radial spin echo measurements of the human brain were performed with a clinical 3T scanner (Siemens Magnetom TIM Trio, Erlangen, Germany) using a receive only 12 channel head coil. Sequence parameters were: $T_R = 2500\,\mathrm{ms}$, $T_E = 50\,\mathrm{ms}$, matrix size $(x, y) = (256, 256)$, slice thickness 2 mm, in plane resolution 0.78 mm × 0.78 mm. 96, 48, 32, and 24 projections were acquired, and the sampling direction of every second spoke was reversed to reduce artifacts from off-resonances (8). Raw data was exported from the scanner, and image reconstruction was performed offline using a Matlab proof-of-principle implementation of the non-uniform fast Fourier transform (NUFFT) (20) and of the proposed TGV method (Eq. 6).

Additionally, reconstruction was performed using an implementation which exploits the high parallelization potential of current graphic processing units (GPUs). It uses NVIDIA's compute unified device architecture (CUDA) framework (21) and leads to a significant acceleration in comparison to the Matlab implementation, which is discussed in the section "Computational requirements."

RESULTS

Application of TGV for Denoising

Figure 3 shows the results from TV and TGV for the T_1 weighted scan of the human prostate. Both methods remove noise while maintaining sharp edges efficiently. However, in regions where the assumption that the image consists of areas that are piecewise constant is violated, it is not surprising that TV results are significantly corrupted by staircasing artifacts. TGV on the other hand works equally well in these regions, as already pointed out in Fig. 1.

The original image with high SNR, the noisy image and results with TV and TGV for the T_2 weighted brain scan are displayed in Fig. 4. Even in this data set, where the assumption that the image consists of areas that are piecewise constant is valid because there are neither severe B_1
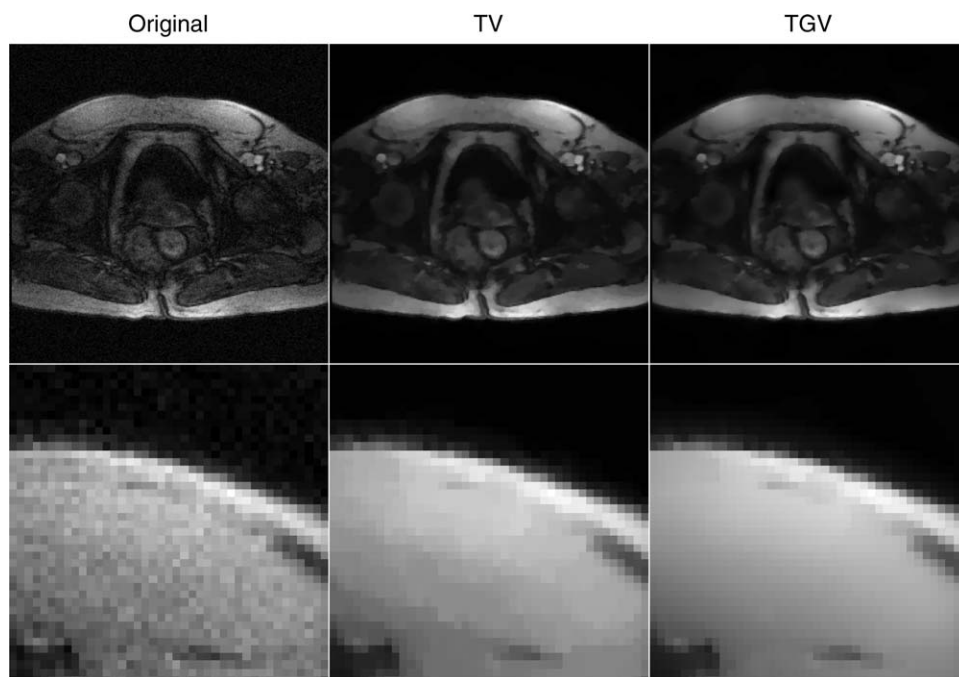
FIG. 3. TV (middle) and TGV (right) denoising results from the contrast enhanced T_1 weighted prostate scan. The images show smooth modulations due to inhomogeneities of the exciting B_1 field and the surface coils used for signal reception which lead to significant artifacts when conventional TV is applied.
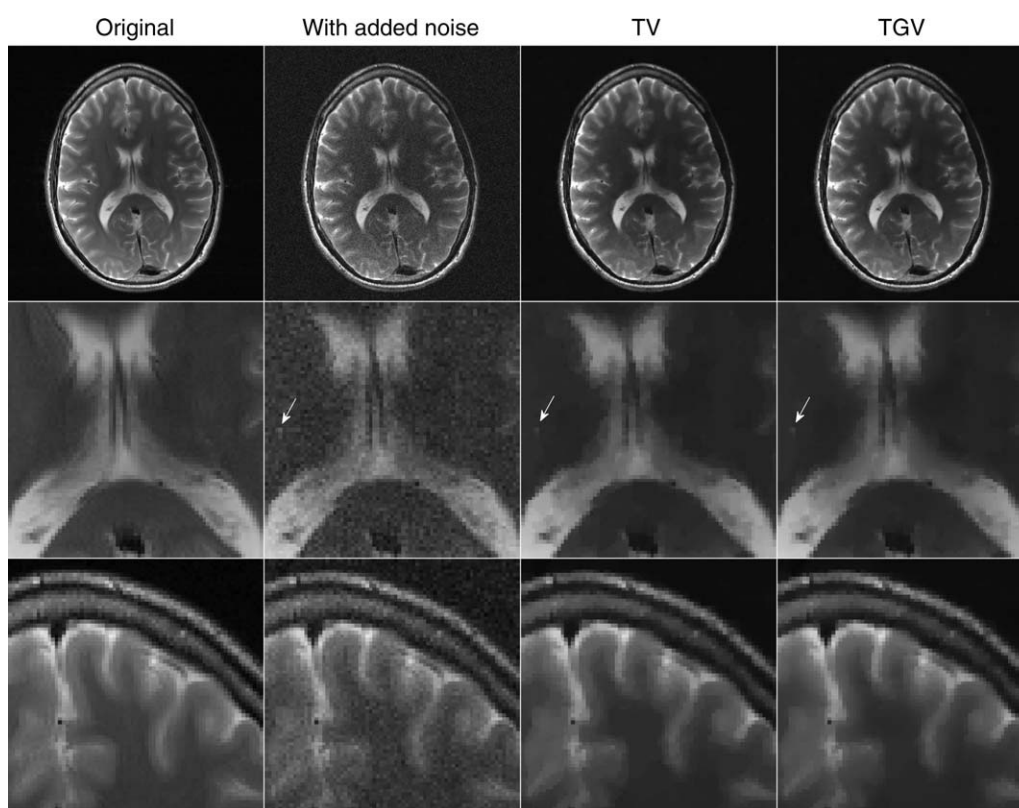

FIG. 4. Denoising experiments for the T_2 weighted brain scan. In the top row, the original image with high SNR (left) which has been contaminated with additive Gaussian noise (middle left) is shown. TV (middle right) and TGV (right) denoising results are displayed as well as magnified views of the respective images (middle and bottom rows). An example of an unwanted image feature that is introduced during denoising is highlighted by arrows.
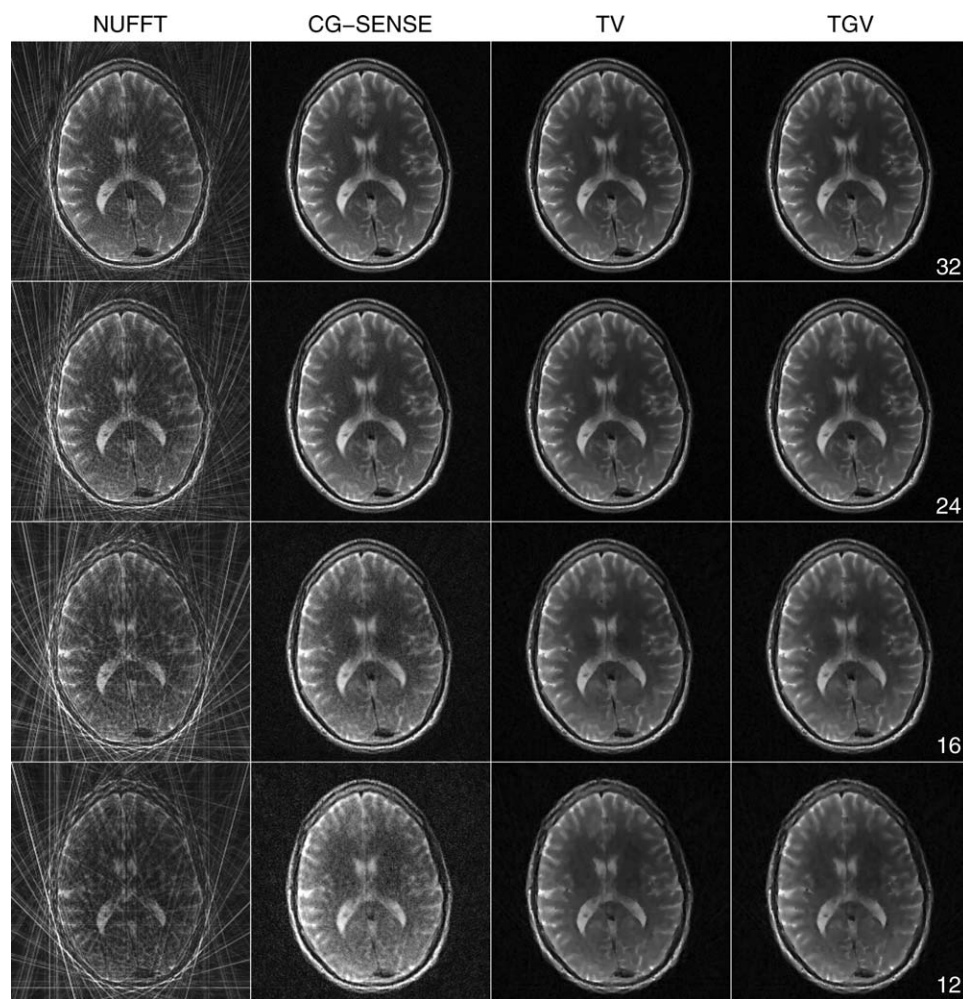
FIG. 5. Results from downsampling experiments of the human brain (256 × 256 matrix) using 32, 24, 16, and 12 radial projections. The columns depict conventional NUFFT reconstructions (left), reconstructions with CG-SENSE (middle left), with a TV constraint (middle right) and with a TGV constraint (right), respectively. The original, fully sampled image is displayed in Fig. 4.

signal inhomogeneities nor modulations from the receive coil, TV introduces additional staircases which cannot be found in the original image. This effect is reduced with TGV.

Image Reconstruction of Subsampled Multicoil Radial Data with a TGV Constraint

Figures 5 and 8 compare the reconstructions of the undersampled radial imaging experiments from retrospective subsampling and truly accelerated imaging. For comparison, the fully sampled image corresponding to the results of Fig. 5 can be found in Fig. 4. It is not surprising that conventional NUFFT reconstructions show streaking artifacts which get worse as the number of projections is reduced. Both TV and TGV reconstructions eliminate streaking artifacts efficiently. For moderate subsampling, results with CG-SENSE are of comparable quality. While the results of TV and TGV are of roughly the same quality for all amounts of subsampling in all experiments, magnified views (Fig. 6) again display staircasing artifacts in the TV results. This is also illustrated in Fig. 7, which displays surface plots of the

same magnified region as in Fig. 6 for the 32 projections data set. While this is still a level of subsampling where CG-SENSE achieves acceptable results, effects of noise amplification are already visible. The difference between TV and TGV is subtle, but the tendency of TV to introduce staircases which are not present in the fully sampled image is clearly visible, e.g., at the frontal interface of the ventricle.

DISCUSSION

The results from this work clearly demonstrate the advantages of TGV over conventional TV for both denoising and for constrained image reconstruction of undersampled data. While this is no real surprise for the results in Fig. 3 because the TV assumption was violated in these cases, it is notable that even with quite homogeneous data sets like in Figs. 4–6 and 8, TV introduces additional staircasing artifacts which lead to a more blocky, less faithful result when compared to the original image because the underlying assumption of piecewise constancy does not follow the underlying signal from the tissue that
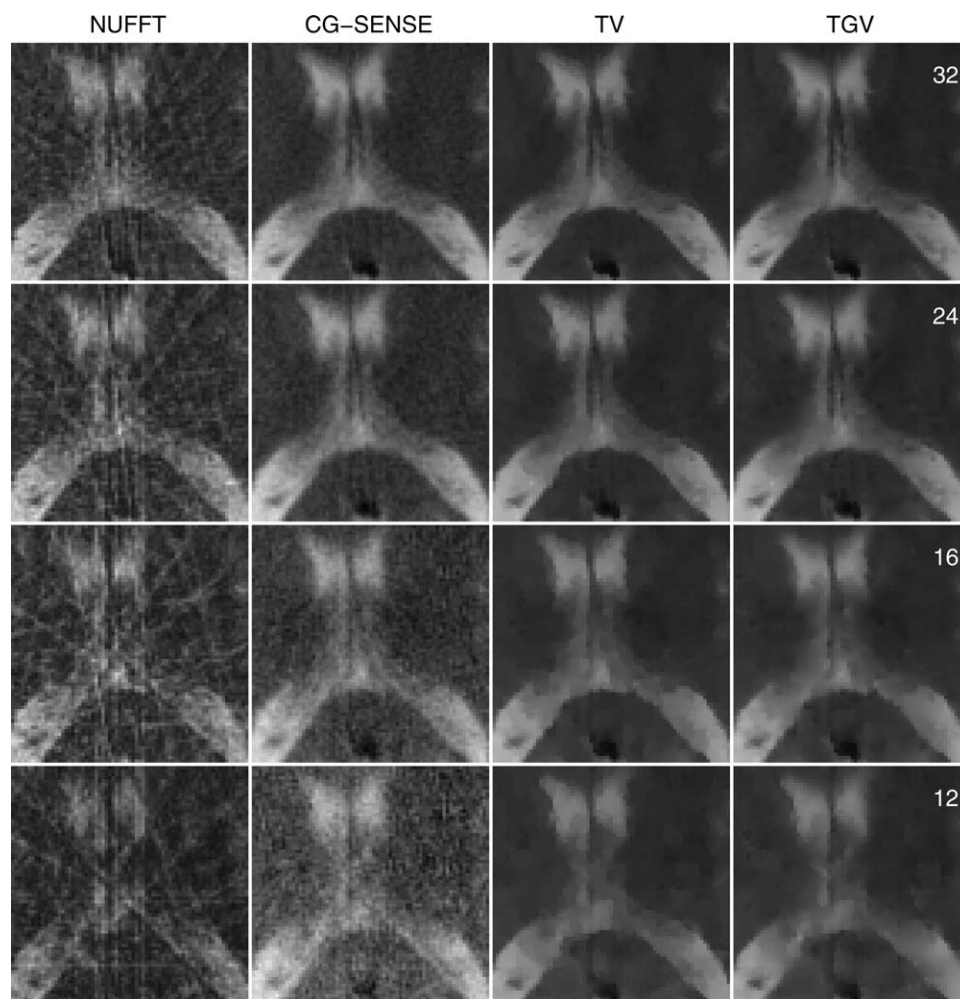
NUFFT   CG–SENSE   TV   TGV

FIG. 6.   Magnified views of the same data as shown in Fig. 5.

well. However, it must be noted that in some situations, the cartoon type, piecewise constant features of TV might actually be desirable. For example, the interfaces between grey and white matter in the magnified view of Fig. 4 (third row) are artificially enhanced with the use of TV. Although TGV is also edge-preserving (as can be seen by comparing the sharp edges of TV and TGV at the periphery of the brain), this effect gives TV results a slightly "sharper" visual appearance. In the case of image reconstruction this effect can be noticed by a direct comparison of Figs. 6 and 7. Again, the edges and piecewise constant regions introduced in the TV results are more emphasized in comparison to the TGV results and even the original image. In this particular case, the surface plots reveal that the TV result is not closer to the original image, but the described effect may still enhance the diagnostic value, comparable to the use of a postprocessing filter for edge enhancement.

As both TV and TGV introduce model based a-priori information, it is possible that this leads to losses in certain parts of the image, or the introduction of unwanted features. One example is highlighted by arrows in the second row of Fig. 4. In this particular case, noise corruption resulted in a group of bright pixels, which were

not eliminated during denoising. Therefore a bright spot remains in the denoised images, which might even be misinterpreted as a lesion. Of course this spot is also visible in the noisy image, but in this case a radiologist would be aware of the low SNR and be reluctant to base a diagnosis on such an image. The same holds true for fine details that are buried below the noise level. While it is possible to eliminate the noise, these features cannot be recovered. Therefore, if denoising is used, radiologists should always have access to the noisy data, too. In this way, it is possible to identify wrong image features and missing details more easily, and an unjustified overestimation of the diagnostic value of the images can be prevented.

It is obvious that both variational approaches (TV and TGV based) clearly outperform conventional L^2 based parallel imaging. Comparable image quality is only achieved for moderate undersampling (96 projections for the 12 channel data set (Fig. 8), 32 projections in the case of the 32 channel data (Figs. 5, 6). Of course, if the degree of undersampling is too high, some features may be buried without chance of recovery with either method. The cases where extremely high undersampling is performed (e.g., using only 12 projections) in Figs. 5 and 6 demonstrate these limits of the method.
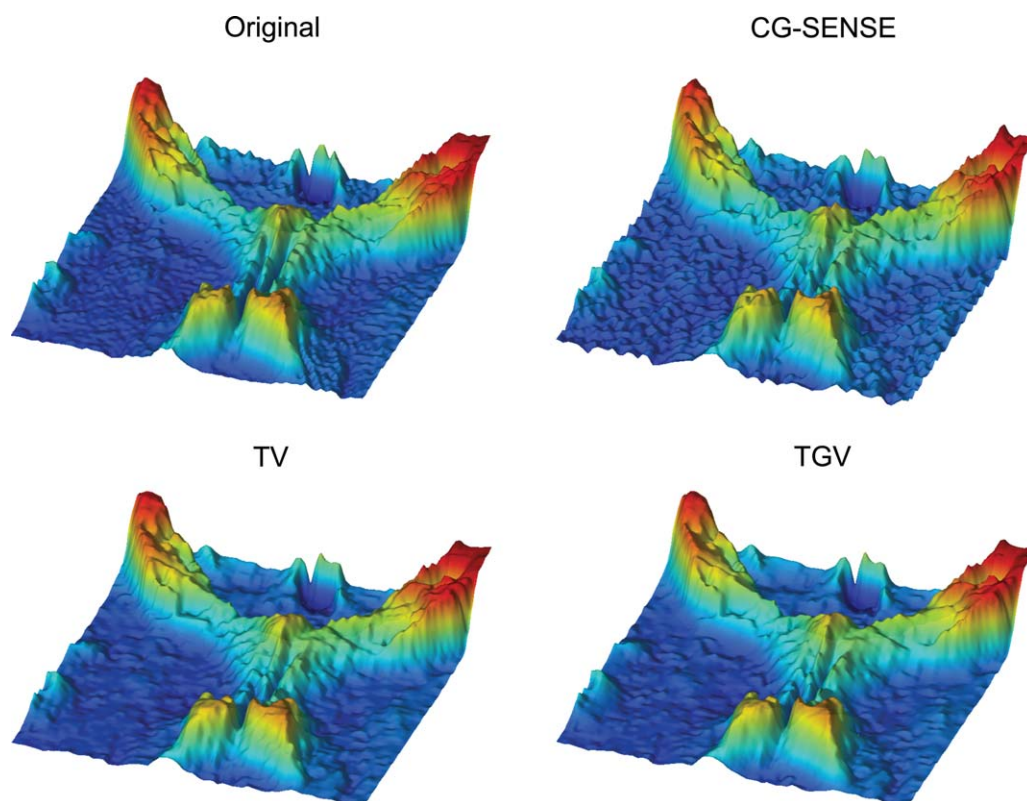
FIG. 7. Surface plots of the 32 projections data set of the same magnified region as in Fig. 6. The original fully sampled data, CGSENSE, TV and TGV reconstructions are displayed. The tendency of TV to introduce staircases is clearly visible, e.g., at the frontal interface of the ventricle.

Choice of the Regularization Parameter

TGV, like TV, needs a proper balancing between the regularization term and the data fidelity term. Overweighting of the TGV term removes image features while underweighting leads to residual noise or streaking artifacts in the image. In our current implementation, the right amount of regularization is determined in the following way. We start with very low regularization and then increase the parameter step by step until noise or artifacts are eliminated from the resulting image. An alternative would be the use of automatic strategies like L-curve (22) or discrepancy principles (23), which is the subject of future investigations. For the calculations of this work, the following weights were used: $\lambda = 0.02$ for prostate denoising (Fig. 3), $\lambda = 0.01$ for brain denoising (Fig. 4) and $\lambda = 8 \times 10^{-5}$ for all image reconstruction experiments, regardless of undersampling and the selected data set (Figs. 5, 6, and 8). Figure 9 demonstrates the influence of the regularization parameter on the reconstructed images for denoising and image reconstruction. Results with proper regularization, over-regularized ($\lambda_{high} = 5 \cdot \lambda_{proper}$) and under-regularized ($\lambda_{low} = 1/5 \cdot \lambda_{proper}$) images are shown. In the case of denoising, overweighting causes pronounced losses of object detail while residual noise remains in the image in the underregularized case. In contrast, image reconstruction is much more robust with respect to changes of the regularization parameter, which is an important condition when considering application in clinical practice. This also explains why excellent results were achieved with a "default" parameter for two different data sets and multiple levels of subsampling in the case of image reconstruction, while individual tuning had to be performed for denoising.

Computational Requirements

The computational load of the proposed TGV models is essentially that of solving the respective minimization problems numerically. These constitute nonsmooth convex optimization problems in high-dimensional spaces with degrees of freedom determined by the number of pixels in the image. Because of the extensive mathematical study of this problem class, efficient and robust abstract algorithms are available.

In our Matlab implementation, which can be run on recent desktop PCs and notebooks, we decided to use rather simple general purpose methods to show that the adaptation to TGV does not require much effort in code development. This served as the basis for the CUDA implementation which was designed to run on every PC with a recent NVIDIA graphics processing unit.

While the algorithms for TGV and TV are essentially similar, one minor drawback of TGV in comparison to TV is the slightly higher computational load arising from the need to update more variables in each iteration step compared to the respective TV method. An extensive analysis of the computation times is beyond the scope of this article. However, our observations are that the NUFFT constitutes
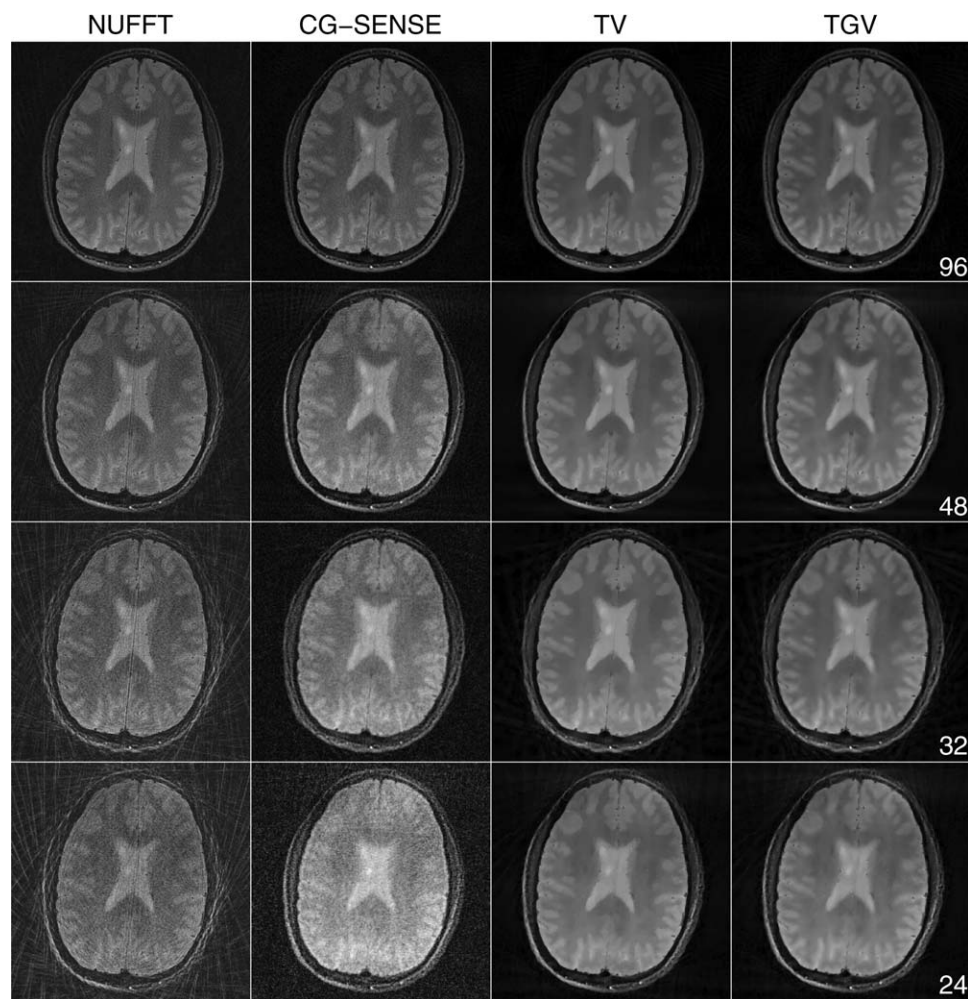
FIG. 8. Radial image reconstructions of the human brain (256 × 256 matrix) from 96, 48, 32, and 24 projections. The columns depict conventional NUFFT reconstructions (left), reconstructions with CG-SENSE (middle left), with a TV constraint (middle right) and with a TGV constraint (right), respectively.

the main part of the computational effort and hence, TGV reconstruction needs only slightly more computation time than TV reconstruction (roughly the factor 1.1 in our proof-of-principle implementation). While this means that TGV currently can only be used in applications where offline image reconstruction or postprocessing is acceptable, this is the case for many iterative TV or compressed sensing methods, too.

As it has also been observed for TV based artifact elimination (24), TGV offers great possibilities of acceleration through parallelization. Using our CUDA-based implementation, the TGV reconstruction of the 32 channel data presented in Fig. 5 was performed in roughly 60–100 seconds depending on amount of subsampling. Likewise, the 12 channel data depicted in Fig. 8 needed, due an increase of iteration steps, ∼80 to 110 seconds of reconstruction time on an NVIDIA GTX 280 GPU with 1GB of memory.

Extensions

Another advantage of TGV belonging to the well-examined class of convex regularization functionals is its versatile applicability to all kinds of variational problems. Roughly speaking, it can basically be used in all MRI applications in which the image is subject to optimization, in particular those where TV has already shown to work.

It is possible to use TGV as a regularizer that can be applied directly to array coil data of individual receiver channels. Thus, TGV can be used in combination with all parallel imaging reconstruction methods that first generate individual coil images, which are then combined to a single sum of squares image in the last step, like GRAPPA (25). This is not easily possible with conventional TV because before the sum of squares combination, the individual coil images are still modulated with the sensitivities of the receiver coils which will lead to considerable staircasing artifacts, as illustrated in Fig. 2.

Radial sampling was chosen as the primary application for image reconstruction in this paper, but of course it is also possible to use TGV as a penalty term for randomly subsampled data sets or for data that was not acquired with phased array coils, following the approach of compressed sensing (6).

Another promising application for TGV would be diffusion tensor imaging. Currently, TGV is only defined for scalar functions, but as symmetric tensor fields are inherent
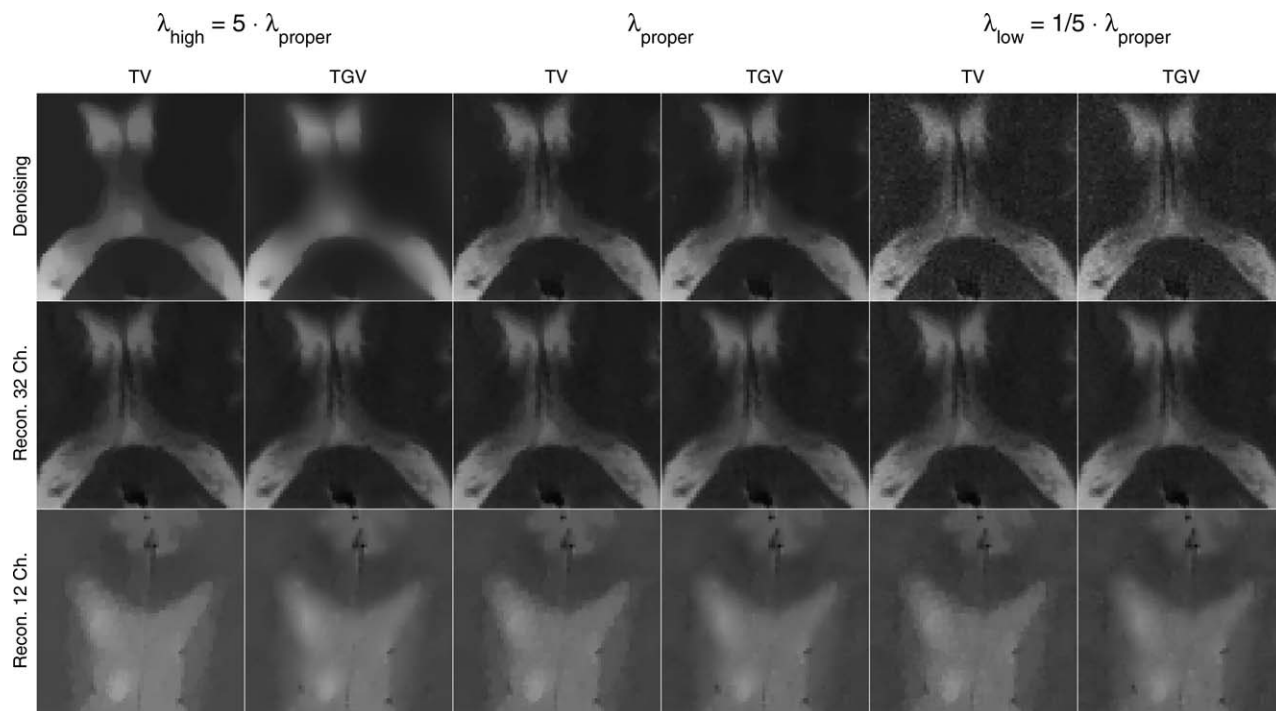
FIG. 9. Denoising (top row) and radial image reconstruction with 32 radial projections (256×256 matrix) from the downsampled (middle row), and the in vivo accelerated (bottom row) experiments. Results with a high regularization weight (left), an appropriate weight (middle) and a low weight (right) are shown.

in the definition, it can be easily extended to higher order tensor fields and could therefore be used as a regularizer to reconstruct the diffusion tensor.

Let us finally mention that the TGV framework is independent of the space dimension and can thus as well be applied to the denoising/reconstruction of 3D image stacks taking full spatial information into account. The implementations can be extended without greater effort, however, an increase in computational complexity in comparison to sequential 2D processing has to be expected, especially in the case of image reconstruction, with 3D regridding being performed in each iteration step. For denoising applications, nevertheless, these kinds of problems do not occur.

CONCLUSIONS

This work presents TGV for MR imaging, a new mathematical framework which is a marked enhancement of the TV method. While it shares the existing desirable features of TV, it is not based on the limiting assumption of piecewise constant images. It is possible to use TGV in all applications where TV is used at the moment, and in the denoising and image reconstruction experiments of this work, it showed its important features to be able to follow the natural continuous signal changes of biological tissues and imaging methods. TGV is based on a solid mathematical theory within the framework of convex optimization, which ensures that numerical implementations are usually straightforward as a variety of numerical algorithms already exist for these types of problems. However,

while both TV and TGV are effective penalty functions for denoising and image reconstruction, the use of a-priori information always introduces other biases in the images. This can lead to image content that is missed, as well as to the introduction of false image features. The diagnostic value of TGV-regularized images in clinical practice has to be investigated in future work.

APPENDIX

Mathematical Background of TGV

One of the main reasons for the use of total variation for image processing problems is the fact that the underlying function space allows for images with discontinuities which usually occur on object boundaries. Even though discontinuities do usually not appear in "real world data," images are often perceived and idealized as having "jumps" at the edges. As functions are not differentiable at discontinuities, a definition of the total variation without derivatives of u is necessary. Such a requirement is realized by setting

$$\mathrm{TV}(u) = \sup \left\{ \int_\Omega u \operatorname{div} v \, \mathrm{d}x \; \middle| \; v \in \mathcal{C}_c^1(\Omega, \mathbf{C}^d), \; \|v\|_\infty \leq 1 \right\}$$

for a given complex-valued function u on a domain $\Omega \subset \mathbf{R}^d$. Note that this definition is informally linked to Eq. 1 by $v = -\nabla u / |\nabla u|$, the vector field for which the supremum is attained. The definition of the total generalized variation modifies the set of allowed test functions v as follows:

$$\mathrm{TGV}_\alpha^2(u) = \sup \left\{ \iint_\Omega u \,\mathrm{div}^2 v \,\mathrm{d}x \ \Big| \ v \in \mathcal{C}_c^2(\Omega, S^{d \times d}), \right.$$

$$\left. \|v\|_\infty \leq \alpha_0, \ \|\mathrm{div}\, v\|_\infty \leq \alpha_1 \right\}$$

letting $\alpha_0, \alpha_1 > 0$ and denoting by $S^{d \times d}$ the set of complex symmetric matrices. The respective definitions for the divergence and norms read as

$$(\mathrm{div}\, v)_i = \sum_{j=1}^d \frac{\partial v_{ij}}{\partial x_j}, \quad \mathrm{div}^2 v = \sum_{i=1}^d \frac{\partial^2 v_{ii}}{\partial x_i^2} + 2 \sum_{i<j} \frac{\partial v_{ij}}{\partial x_i \partial x_j}$$

and

$$\|v\|_\infty = \sup_{x \in \Omega} \left(\sum_{i=1}^d |v_{ii}(x)|^2 + 2 \sum_{i<j} |v_{ij}(x)|^2 \right)^{1/2},$$

$$\|\mathrm{div}\, v\|_\infty = \sup_{x \in \Omega} \left(\sum_{i=1}^d \left| \sum_{j=1}^d \frac{\partial v_{ij}}{\partial x_j}(x) \right|^2 \right)^{1/2}.$$

Again, u needs not to be differentiated. Also, the latter definition can be informally linked, via Fenchel duality theory, with Eq. 2 yielding the same values, see again (9).

ALGORITHM DESCRIPTION

TGV Denoising

The primal-dual algorithm we employed for solving the discretization of the TGV-constrained denoising problem (Eq. 7) bases on the formulation of Eq. 8 as a convex-concave saddle-point problem which is derived by duality principles:

$$\min_{u^h \in U^h, \ v^h \in V^h} \max_{p^h \in P^h, \ q^h \in Q^h} \frac{1}{2\lambda} \|u^h - f^h\|_2^2 + \left\langle \nabla^h u^h - v^h, p \right\rangle$$

$$+ \left\langle \mathcal{E}^h(v^h), q \right\rangle, \quad [\text{A1}]$$

where p^h, q^h are the dual variables. The sets associated with these variables are given by

$$P^h = \left\{ p^h \in \mathbf{C}^{2MN} \big| \|p^h\|_\infty \leq \alpha_1 \right\},$$

$$Q^h = \left\{ q^h \in \mathbf{C}^{3MN} \big| \|q^h\|_\infty \leq \alpha_0 \right\}.$$

We denote by $\mathrm{proj}_{P^h}(\tilde{p}^h)$ and $\mathrm{proj}_{Q^h}(\tilde{q}^h)$ the Euclidean projectors onto the convex sets P^h and Q^h. The projections can be easily computed by pointwise operations

$$\mathrm{proj}_{P^h}(\tilde{p}^h) = \frac{\tilde{p}^h}{\max\left(1, \frac{|\tilde{p}^h|}{\alpha_1}\right)}, \quad \mathrm{proj}_{Q^h}(\tilde{q}^h) = \frac{\tilde{q}^h}{\max\left(1, \frac{|\tilde{q}^h|}{\alpha_0}\right)}.$$

We further denote by $\mathrm{prox}_1^\tau(\tilde{u}^h)$ the proximal map

$$\mathrm{prox}_1^\tau(\tilde{u}^h) = \operatorname*{argmin}_{u^h \in U^h} \frac{\|u^h - \tilde{u}^h\|_2^2}{2\tau} + \frac{\|(u^h - \tau f^h)\|_2^2}{2\lambda}.$$

which can easily identified as the pointwise operation

$$\mathrm{prox}_1^\tau(\tilde{u}^h) = \frac{\lambda \tilde{u}^h + \tau f^h}{\lambda + \tau}.$$

The primal-dual algorithm for the denoising problem Eq. 7 reads as follows:

Algorithm 1 (Primal-dual method for TGV denoising).

1: **function** TGV_{DENOISE}(f^h)
2:     $u^h, \bar{u}^h \leftarrow f^h, \ v^h, \bar{v}^h \leftarrow 0, \ p^h \leftarrow 0, \ q^h \leftarrow 0$, choose $\tau, \sigma > 0$.
3:     **repeat**
4:        $p^h \leftarrow \mathrm{proj}_{P^h}\left(p^h + \sigma(\nabla^h \bar{u}^h - \bar{v}^h)\right)$
5:        $q^h \leftarrow \mathrm{proj}_{Q^h}(q^h + \sigma \mathcal{E}^h \bar{v}^h)$
6:        $u_{\mathrm{old}}^h \leftarrow u^h$
7:        $u^h \leftarrow \mathrm{prox}_1^\tau\left(u^h + \tau \,\mathrm{div}_1^h p^h\right)$
8:        $\bar{u}^h \leftarrow 2u^h - u_{\mathrm{old}}^h$
9:        $v_{\mathrm{old}}^h \leftarrow v^h$
10:       $v^h \leftarrow v^h + \tau\left(p^h + \mathrm{div}_2^h q^h\right)$
11:       $\bar{v}^h \leftarrow 2v^h - v_{\mathrm{old}}^h$
12:     **until** convergence of u^h
13:     **return** u^h
14: **end function**

Note that iterations are very simple and can therefore be efficiently implemented on parallel hardware such as graphics processing units. The algorithm is known to converge for step-sizes $\sigma\tau \leq 2h^2/(16 + h^2 + \sqrt{32h^2 + h^4})$, in particular for mesh-size $h = 1$, the choice $\sigma = \tau = 1/\sqrt{12}$ leads to convergence, see (16) for details.

TGV Undersampling Reconstruction

The discrete optimization problem (Eq. 9) can solved by applying the same primal-dual method. With a straightforward adaptation of Algorithm 1, however, one needs to compute the proximal map with respect to the data fitting term. Such an approach involves the solution of a linear system of equations which is quite costly. To circumvent this, we also dualize with respect to the data fitting term. The resulting convex-concave saddle-point problem is obtained as

$$\min_{u^h \in U^h, \ v^h \in V^h} \max_{p^h \in P^h, \ q^h \in Q^h, \ r^h \in R^h} \left\langle \nabla^h u^h - v^h, p^h \right\rangle + \left\langle \mathcal{E}^h(v^h), q^h \right\rangle$$

$$+ \left\langle K^h u^h - g^h, r^h \right\rangle - \frac{\lambda}{2}\|r^h\|_2^2, \quad [\text{A2}]$$

where $r^h \in R^h$ is the dual variable with respect to the data fitting term. Further, let $\mathrm{prox}_2^\sigma(\tilde{r}^h)$ be the proximal map

$$\mathrm{prox}_2^\sigma(\tilde{r}^h) = \operatorname*{argmin}_{r^h \in R^h} \frac{\|r^h - \tilde{r}^h\|_2^2}{2\sigma} + \frac{\lambda}{2}\|r^h\|_2^2$$

$$\Leftrightarrow \mathrm{prox}_2^\sigma(\tilde{r}^h) = \frac{\tilde{r}^h}{1 + \sigma\lambda}.$$

The primal-dual algorithm for the reconstruction problem Eq. 6 is as follows:

Algorithm 2 (PD method for TGV undersampling reconstruction).

1: **function** TGV_{RECONSTRUCT}(g^h)
2:     $u^h, \bar{u}^h \leftarrow 0, \ v^h, \bar{v}^h \leftarrow 0, \ p^h \leftarrow 0, \ q^h \leftarrow 0, \ r^h \leftarrow 0$, choose $\tau, \sigma > 0$.

3:    **repeat**

4:    $p^h \leftarrow \text{proj}_{P^h}\left(p^h + \sigma(\nabla^h \bar{u}^h - \bar{v}^h)\right)$

5:    $q^h \leftarrow \text{proj}_{Q^h}\left(q^h + \sigma \mathcal{E}^h \bar{v}^h\right)$

6:    $r^h \leftarrow \text{prox}_2^\sigma\left(r^h + \sigma(K^h \bar{u}^h - g^h)\right)$

7:    $u_{\text{old}}^h \leftarrow u^h$

8:    $u^h \leftarrow u^h + \tau\left(\text{div}_1^h p^h - (K^h)^* r^h\right)$

9:    $\bar{u}^h \leftarrow 2u^h - u_{\text{old}}^h$

10:    $v_{\text{old}}^h \leftarrow v^h$

11:    $v^h \leftarrow v^h + \tau\left(p^h + \text{div}_2^h q^h\right)$

12:    $\bar{v}^h \leftarrow 2v^h - v_{\text{old}}^h$

13:    **until** convergence of u^h

14:    **return** u^h

15: **end function**

Again, this algorithm is convergent provided that $\sigma\tau < h^2/(h^2 + \sqrt{8}h + 8)$, in particular for mesh-size $h = 1$, the choice $\sigma = \tau = 1/\sqrt{12}$ is sufficient.

ACKNOWLEDGMENTS

REFERENCES

1. Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Phys D 1992;60:259–268.
2. Liu B, King K, Steckner M, Xie J, Sheng J, Ying L. Regularized sensitivity encoding (SENSE) reconstruction using Bregman iterations. Magn Reson Med 2009;61:145–152.
3. Wald MJ, Adluru G, Song HK, Wehrli FW. Accelerated high-resolution imaging of trabecular bone using total variation constrained reconstruction. In: Proceedings of the 17th Scientific Meeting and Exhibition of ISMRM, Honolulu, HI. 2009.
4. Block KT, Uecker M, Frahm J. Suppression of MRI truncation artifacts using total variation constrained data extrapolation. In: Proceedings of the 16th Scientific Meeting and Exhibition of ISMRM, Toronto, Canada. 2008.
5. Huang F, Chen Y, Duensing GR, Akao J, Rubin A, Saylor C. Application of partial differential equation-based inpainting on sensitivity maps. Magn Reson Med 2005;53:388–397.
6. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn Reson Med 2007;58:1182–1195.
7. Block KT, Uecker M, Frahm J. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. Magn Reson Med 2007;57:1086–1098.
8. Block T. Advanced methods for radial data sampling in MRI. Ph.D. thesis, Georg-August-Universitaet Goettingen, 2008.
9. Bredies K, Kunisch K, Pock T. Total generalized variation. SIAM J Imaging Sci 2010;3:492–526.
10. Candes EJ, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inform Theory 2006;52:489–509.
11. Donoho DL. Compressed sensing. IEEE Trans Inform Theory 2006;52:1289–1306.
12. Lustig M, Donoho DL, Santos JM, Pauly JM. Compressed sensing MRI. IEEE Signal Process Magazine 2008;25:72–82.
13. Nesterov Y. Smooth minimization of nonsmooth functions. Math Programming Series A 2005;103:127–152.
14. Nemirovski, A. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM J Optim 2004;15:229–251.
15. Pock T, Cremers D, Bischof H, Chambolle A. An Algorithm for Minimizing the Mumford-Shah Functional. In: International Conference on Computer Vision (ICCV), Kyoto, Japan; 2009.
16. Chambolle A, Pock T. A first-order primal-dual algorithm for convex problems with applications to imaging. Technical report, Institute for Computer Graphics and Vision, Graz University of Technology, 2010.
17. Merwa R, Reishofer G, Feiweier T, Kapp K, Ebner F, Stollberger R. Improved quantification of pharmacokinetic parameters at 3 Tesla considering B1 inhomogeneities. In: Proceedings of the 17th Scientific Meeting and Exhibition of ISMRM, Honolulu, HI. 2009; p 4363.
18. Pruessmann KP, Weiger M, Börnert P, Boesiger P. Advances in sensitivity encoding with arbitrary k-space trajectories. Magn Reson Med 2001;46:638–651.
19. Bernstein MA, King KF, Zhou XJ. Handbook of MRI Pulse Sequences. Academic Press; 2004.
20. Fessler JA, Sutton BP. Nonuniform fast Fourier transforms using min-max interpolation. IEEE Trans Signal Process 2003;51:560–574.
21. NVIDIA. NVIDIA CUDA Programming Guide 2.0. NVIDIA Cooperation; 2008.
22. Engl HW, Hanke M, Neubauer A. Regularization of inverse problems, Vol. 375: *Mathematics and its applications*, Dordrecht: Kluwer Academic Publishers Group; 1996.
23. Morozov VA. On the solution of functional equations by the method of regularization. Soviet Math Dokl 1966;7:414–417.
24. Knoll F, Unger M, Diwoky C, Clason C, Pock T, Stollberger R. Fast reduction of undersampling artifacts in radial MR angiography with 3D total variation on graphics hardware. MAGMA 2010;23:103–114.
25. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 2002;47:1202–1210.