

# Magnetic Resonance in Medicine

## Parallel Imaging with Nonlinear Reconstruction using Variational Penalties

Florian Knoll<sup>1\*</sup>, Christian Clason<sup>2</sup>, Kristian Bredies<sup>2</sup>, Martin Uecker<sup>3</sup>  
and Rudolf Stollberger<sup>1</sup>

<sup>1</sup>Institute of Medical Engineering, Graz University of Technology, Kronesgasse 5,  
A-8010 Graz, Austria

<sup>2</sup>Institute for Mathematics and Scientific Computing, University of Graz, Heinrich-  
strasse 36, A-8010 Graz, Austria

<sup>3</sup>Biomedizinische NMR Forschungs GmbH, Max-Planck-Institut für biophysikalische  
Chemie, Göttingen, Germany

\*Correspondence to: Florian Knoll,  
Graz University of Technology,  
Institute of Medical Engineering,  
Kronesgasse 5, A-8010 Graz, Austria.  
Phone: +43-316-873-5375 , Fax: +43-316-873-7890.  
E-mail: florian.knoll@tugraz.at

Grant sponsor: Austrian Science Fund: Grant number SFB F3209-18.

Running Title: Nonlinear Parallel Imaging with Variational Penalties

Number of Words (Abstract) = 84

Number of Words (Body) approximately 4000

Number of Figures: 7

Number of Citations: 47

**Magnetic Resonance in Medicine**

## **ABSTRACT**

A new approach based on nonlinear inversion for autocalibrated parallel imaging with arbitrary sampling patterns is presented. By extending the iteratively regularized Gauss–Newton method with variational penalties, the improved reconstruction quality obtained from joint estimation of image and coil sensitivities is combined with the superior noise suppression of total variation and total generalized variation regularization. In addition, the proposed approach can lead to enhanced removal of sampling artifacts arising from pseudorandom and radial sampling patterns. This is demonstrated for phantom and in-vivo measurements.

Key words: Parallel Imaging, Radial Sampling, Pseudorandom Sampling, Nonlinear Inversion, Total Variation, Total Generalized Variation.

## INTRODUCTION

It was shown recently (1–4) that nonlinear inversion can be applied successfully to image reconstruction of undersampled data from multiple coils, i.e., parallel imaging (5–7). The joint estimation of images and coil sensitivities, which can be achieved with an iteratively regularized Gauss–Newton (IRGN) method, leads to more accurate estimation of the coil sensitivities, and therefore yields results with improved image quality. It was also demonstrated that this method can be extended to non-Cartesian imaging (8,9). In particular, radial sampling has the advantage that the sampling pattern automatically leads to an oversampling of the central frequencies of  $k$ -space, which eliminates the need to acquire additional reference lines when performing auto-calibrated parallel imaging. Another important characteristic of radial sampling is the fact that aliasing artifacts, which are introduced by under-sampling, have a distinctively regular appearance which is usually different from the image content. For this reason, it is possible to remove these so-called streaking artifacts during reconstruction with the integration of suitable penalties on the reconstructed image, with total variation (TV) proving particularly effective in the radial case (10,11). Successful application of TV regularization to conventional Cartesian subsampling had been reported as well (12). This interest in total variation-type regularization has been stimulated by the success of compressed sensing (13,14) in MRI (15–17), which is often used in combination with pseudorandom sampling (i.e., randomly selecting sampling points from a regular Cartesian grid); see also (18) for a related idea. Recently, compressed sensing has also been used for parallel imaging with unknown coil sensitivities by formulating the reconstruction as a low-rank matrix completion problem (19).

The purpose of this work is to demonstrate that the formulation of autocalibrated parallel imaging as a nonlinear inverse problem yields a general framework that allows the joint estimation of coil sensitivities and image content in combination with arbitrary sampling patterns and variational penalties. This is illustrated via the integration of a TV penalty in the IRGN method. The immediate benefit is that TV regularization helps to suppress the noise amplification (20) that seriously limits parallel imaging with conventional methods (for linear reconstruction methods, the effect of noise suppression due to regularization is well known (21–24)). When applied to data sets obtained with radial or pseudorandom sampling patterns, TV also leads to enhanced removal of artifacts. The flexibility is further shown by replacing TV with a different convex regularization functional that is more suitable in cases when the assumption of piecewise constant images is not reasonable: total generalized variation (TGV) of second order (25,26). A similarly general approach

in terms of sampling patterns and variational penalties is followed in SPIRiT (27), which can also be formulated as a nonlinear minimization problem. In contrast to the approach considered here, calibration is carried out in a separate step, and the corresponding variant with non-quadratic regularization (L1 SPIRiT) is based on compressed sensing using a wavelet transform.

## THEORY

Mathematically, parallel MR imaging can be formulated as a nonlinear inverse problem where the sampling operator  $\mathcal{F}_S$  (defined by the  $k$ -space trajectory, e.g., Fourier transform followed by multiplication with a binary mask in the case of standard Cartesian subsampling) and the correspondingly acquired  $k$ -space data  $g = (g_1, \dots, g_N)^T$  from  $N$  receiver coils are given, and the spin density  $u$  and the unknown (or not perfectly known) set of coil sensitivities  $c = (c_1, \dots, c_N)^T$  have to be found such that

$$F(u, c) := (\mathcal{F}_S(u \cdot c_1), \dots, \mathcal{F}_S(u \cdot c_N))^T = g \quad [1]$$

holds. As was shown in (3, 8), this problem can be solved using the iteratively regularized Gauss–Newton (IRGN) method (28–31), i.e., computing in each step  $k$  for given  $x^k := (u^k, c^k)$  the solution  $\delta x := (\delta u, \delta c)$  of the minimization problem

$$\min_{\delta x} \frac{1}{2} \|F'(x^k)\delta x + F(x^k) - g\|^2 + \frac{\alpha_k}{2} \mathcal{W}(c^k + \delta c) + \beta_k \mathcal{R}(u^k + \delta u) \quad [2]$$

for given  $\alpha_k, \beta_k > 0$ , and then setting  $x^{k+1} := x^k + \delta x$ ,  $\alpha_{k+1} := q_\alpha \alpha_k$  and  $\beta_{k+1} := q_\beta \beta_k$  with  $0 < q_\alpha, q_\beta < 1$ . Here,  $F'(x^k)$  is the Fréchet derivative of  $F$  evaluated at  $x^k$ . The term  $\mathcal{W}(c) = \|Wc\|^2 = \|w \cdot \mathcal{F}c\|^2$  is a penalty on the high Fourier coefficients of the sensitivities and  $\mathcal{R}$  is a regularization term for the image. So far, the application of the IRGN method to parallel imaging has been formulated with a conventional  $L^2$  penalty (3, 8) (i.e.,  $\mathcal{R}(u) = \frac{1}{2}\|u\|^2$ ). As demonstrated in this work, the IRGN method can also be used with other regularization terms, which can be chosen dependent on the application. For example, the stability of the method with respect to noise can be improved: Since  $\alpha_k$  and  $\beta_k$  are decreasing during the iteration, the problem in Eq. 2 will become increasingly ill-conditioned. This leads to noise amplification, which can be counteracted by using a regularization term with stronger noise removal properties than the  $L^2$  penalty. A possible choice is the total variation (TV) of the image, i.e.,

$$\mathcal{R}(u) = \int |\nabla u|_2 dx, \quad [3]$$

---

**Algorithm 1** Solution of TV sub-problem (Eq. 6)
 

---

```

1: function TVSOLVE( $u, c, g, \alpha, \beta$ )
2:    $\delta u, \bar{\delta u}, \delta c, \bar{\delta c}, p \leftarrow 0$ , choose  $\sigma, \tau > 0$ 
3:   repeat
4:      $p \leftarrow \text{proj}_\beta(p + \tau \nabla(u + \bar{\delta u}))$ 
5:      $\delta u_{old} \leftarrow \delta u, \delta c_{old} \leftarrow \delta c$ 
6:      $\delta u \leftarrow \delta u - \sigma(\sum_{i=1}^N c_i^* \cdot \mathcal{F}_s^*(\mathcal{F}_s(u \cdot \bar{\delta c}_i + c_i \cdot \bar{\delta u}) + F(u, c) - g) - \text{div } p)$ 
7:      $\delta c \leftarrow \delta c - \sigma(u^* \cdot \mathcal{F}_s^*(\mathcal{F}_s(u \cdot \bar{\delta c}_i + c_i \cdot \bar{\delta u}) + F(u, c) - g) + \alpha W^*W(c_i + \bar{\delta c}_i))$ 
8:      $\bar{\delta u} \leftarrow 2\delta u - \delta u_{old}$ 
9:      $\bar{\delta c} \leftarrow 2\delta c - \delta c_{old}$ 
10:  until convergence
11:  return  $\delta u, \delta c$ 
12: end function
    
```

---

where  $|\cdot|_2$  denotes the Euclidean norm in  $\mathbb{R}^2$ . To calculate the solution of Eq. 2, we make use of the dual characterization of the TV semi-norm:

$$\beta \int |\nabla u|_2 dx = \sup_{p \in C_\beta} \langle u, -\text{div } p \rangle, \quad [4]$$

where  $p = (p_1, p_2)^T$ ,  $\text{div } p = \partial_x p_1 + \partial_y p_2$  with appropriate boundary conditions and

$$C_\beta = \{p \in L^2(\Omega; \mathbb{C}^2) : \text{div } p \in L^2(\Omega; \mathbb{C}), |p(x)|_2 \leq \beta \text{ for almost all } x \in \Omega\}. \quad [5]$$

The problem in Eq. 2 then becomes a non-smooth convex-concave saddle-point problem,

$$\min_{\delta u, \delta c} \max_{p \in C_{\beta_k}} \frac{1}{2} \|F'(x^k)\delta x + F(x^k) - g\|^2 + \frac{\alpha_k}{2} \mathcal{W}(c^k + \delta c) + \langle u^k + \delta u, -\text{div } p \rangle, \quad [6]$$

which can be solved efficiently using a projected primal-dual extra-gradient method (32,33), given as Algorithm 1. Since this requires only application of  $F'(x^k)$  and its adjoint  $F'(x^k)^*$ , the algorithm can be implemented efficiently on modern multi-core hardware such as graphics processing units (GPUs). Due to the bilinear structure of  $F$ , the action of  $F'(x^k)$  and  $F'(x^k)^*$  can be calculated explicitly in terms of the subsampling operator  $\mathcal{F}_S$  and its adjoint  $\mathcal{F}_S^*$ . The projection onto the convex set  $C_\beta$  can be calculated pointwise by setting for all  $x \in \Omega$

$$\text{proj}_\beta(q)(x) = \frac{q(x)}{\max(1, \beta^{-1}(|q(x)|_2)}. \quad [7]$$

Since TV regularization is known to introduce staircasing artifacts if the penalty parameter is large, we also consider second order total generalized variation (TGV),

---

**Algorithm 2** Solution of TGV sub-problem (Eq. 9)
 

---

```

1: function TGV SOLVE( $u, c, g, \alpha, \beta$ )
2:    $\delta u, \bar{\delta u}, \delta c, \bar{\delta c}, v, \bar{v}, p, q \leftarrow 0$ , choose  $\sigma, \tau > 0$ 
3:   repeat
4:      $p \leftarrow \text{proj}_\beta(p + \tau(\nabla(u + \bar{\delta u}) - v))$ 
5:      $q \leftarrow \text{proj}_\beta^2(q + \tau(\mathcal{E}v))$ 
6:      $\delta u_{old} \leftarrow \delta u, \delta c_{old} \leftarrow \delta c, v_{old} \leftarrow v$ 
7:      $\delta u \leftarrow \delta u - \sigma(\sum_{i=1}^N c_i^* \cdot \mathcal{F}_s^*(\mathcal{F}_s(u \cdot \bar{\delta c}_i + c_i \cdot \bar{\delta u}) + F(u, c) - g) - \text{div } p)$ 
8:      $\delta c \leftarrow \delta c - \sigma(u^* \cdot \mathcal{F}_s^*(\mathcal{F}_s(u \cdot \bar{\delta c}_i + c_i \cdot \bar{\delta u}) + F(u, c) - g) + \alpha W^* W(c_i + \bar{\delta c}_i))$ 
9:      $v \leftarrow v - \sigma(-p + \mathcal{E}^* q)$ 
10:     $\bar{\delta u} \leftarrow 2\delta u - \delta u_{old}$ 
11:     $\bar{\delta c} \leftarrow 2\delta c - \delta c_{old}$ 
12:     $\bar{v} \leftarrow 2v - v_{old}$ 
13:  until convergence
14:  return  $\delta u, \delta c$ 
15: end function
    
```

---

which is a generalization of TV that avoids the staircasing in regions of smooth signal change (25, 26). This amounts to setting

$$\beta \mathcal{R}(u) = \inf_v \beta \|\nabla u - v\| + 2\beta \|\mathcal{E}v\|, \quad [8]$$

where  $\mathcal{E}v = \frac{1}{2}(\nabla v + \nabla v^T)$  denotes the symmetrized gradient of the complex-valued vector field  $v \in \mathcal{C}^1(\Omega; \mathbb{C}^2)$ . We refer to (25, 26) for a detailed description of this functional and an explanation of its properties. Using again the dual representation of the norms, the Gauss–Newton step (Eq. 2) is equivalent to the saddle point problem

$$\min_{\delta u, \delta c, v} \max_{\substack{p \in C_{\beta_k} \\ q \in C_{\beta_k}^2}} \frac{1}{2} \|F'(x^k) \delta x + F(x^k) - g\|^2 + \frac{\alpha_k}{2} \mathcal{W}(c^k + \delta c) + \langle \nabla u^k + \delta u - v, p \rangle + \langle \mathcal{E}v, q \rangle, \quad [9]$$

where

$$C_\beta^2 = \left\{ q \in \mathcal{C}_c(\Omega, \mathcal{S}^{2 \times 2}) : (|q_{11}(x)|^2 + |q_{22}(x)|^2 + 2|q_{12}(x)|^2)^{1/2} \leq 2\beta \ \forall x \in \Omega \right\} \quad [10]$$

and  $\mathcal{S}^{2 \times 2}$  denotes the set of symmetric complex matrices. The corresponding extra-gradient method is given as Algorithm 2, where the projection  $\text{proj}_\beta^2$  onto  $C_\beta^2$  can again be computed pointwise.

## MATERIALS AND METHODS

### Data Acquisition

Experiments were performed for 3D pseudorandom as well as 2D radial sampling patterns. All measurements were performed on a clinical 3T system (Siemens Magnetom TIM Trio, Erlangen, Germany). Written informed consent was obtained from all volunteers prior to the examination.

Accelerated acquisition with pseudorandom sampling was tested with phantom experiments. A receive-only 12-channel head coil was used, and an SVD based coil compression (34) was applied to reduce the data to 8 virtual channels. Measurements were performed with a 3D FLASH sequence with the following sequence parameters: TR=20ms, TE=5ms, flip angle  $18^\circ$ , matrix size  $256 \times 256 \times 256$ . The pulse sequence was modified to include a binary 2D mask defining the subsampling of both phase-encoding directions. A resolution of  $1\text{mm} \times 1\text{mm} \times 5\text{mm}$  was used. Raw data was exported from the scanner, a 1D Fourier transform was performed along the readout direction, and partitions orthogonal to this axis were reconstructed.

Additionally, a fully sampled  $T_2$  weighted 2D turbo spin echo data set of the brain of a healthy volunteer was acquired with a 32-channel receive coil. The data was compressed to 12 virtual channels. Sequence parameters were TR=5000ms, TE=99ms, turbo factor 10, matrix size  $256 \times 256$ , slice thickness 4mm and an in-plane resolution of  $0.86\text{mm} \times 0.86\text{mm}$ . Raw data was exported from the scanner and then subsampled retrospectively with an adapted pseudorandom sampling pattern (16), where sampling points on a regular Cartesian grid are randomly selected according to a specified probability density function which is based on the energy distribution in  $k$ -space of medical images (35).

For comparison, conventional TV filtering was applied to the magnitude images obtained from standard IRGN reconstruction of the  $T_2$  weighted data set of the brain. This experiment was repeated for an accelerated pseudorandom in-vivo measurement of the brain of a different volunteer. The same experimental setup was used as in the accelerated phantom experiments, except that 9 SVD compressed virtual channels were used and an isotropic spatial resolution of 1mm was achieved.

Radial sampling experiments were performed with an rf-spoiled radial FLASH sequence with sequence parameters TR=2.0ms, TE=1.3 ms, and a flip angle of  $8^\circ$ . Images of a water phantom and of the heart of a healthy volunteer were made. In-vivo data was acquired with a 32-channel body array coil, and data acquisition was performed with a protocol designed for real-time imaging (36–38) without cardiac gating and during free breathing (39). After the acquisition, the data was com-

pressed to 12 virtual channels for the in-vivo experiments and 8 virtual channels for the phantom experiments. An in-plane resolution of  $2\text{mm}\times 2\text{mm}$  and a slice thickness of 8mm was used in combination with  $128\times 128$  image matrices. Due to the two-fold oversampling, this resulted in 256 sample points for each radial spoke.

## Nonlinear reconstruction

All reconstructions were performed offline using a Matlab (R2010a, The MathWorks, Natick, MA, USA) implementation of the described nonlinear inversion method. For the reconstruction of radial data sets, Fessler and Sutton’s NUFFT (40) code was used. To facilitate comparison, the solution of Eq. 2 with  $\mathcal{R}(v) = \frac{1}{2}\|v\|^2$  was computed using the same extra-gradient scheme, which can be obtained from Algorithm 1 by removing step 4 and replacing the term  $-\text{div } p$  with  $+\beta u + \overline{\delta u}$  in step 6. In the following, we will refer to the Gauss–Newton reconstruction using an  $L^2$ -penalty simply as IRGN, while the reconstruction using TV and TGV penalties will be denoted by IRGN-TV and IRGN-TGV, respectively.

## Parameter Choice

The parameters in Algorithm 1 were chosen according to the convergence theory for the projected extra-gradient scheme. The step lengths  $\sigma$  and  $\tau$  were selected such that  $\sigma\tau L^2 < 1$  holds, where  $L$  is the Lipschitz constant of the gradient of the functional to be minimized. This constant depends on the subsampling strategy and the iterates  $u^k, c^k$ , but can be estimated using a few iterations of the power method to approximately compute the norms of the partial Fréchet derivatives of the linearized operator  $F'(x)$ . As the norm of the finite difference approximation of the divergence and gradient operators with mesh size 1 is  $\sqrt{8}$ , we set  $\tau = \sigma = 1/\sqrt{8 + 2\max(|\tilde{L}_u|, |\tilde{L}_c|)}$  in Algorithm 1, where  $\tilde{L}_u, \tilde{L}_c$  are the estimates from the power method. The step lengths in Algorithm 2 were set to  $\tau = \sigma = 1/\sqrt{12 + 2\max(|\tilde{L}_u|, |\tilde{L}_c|)}$  based on the norm of the linear operator involving the symmetrized derivative. The iteration was terminated after a fixed number of iterations, since the efficiency estimate for the extra-gradient method gives an upper bound on the required number of iterations to achieve a given accuracy. Since a high accuracy is not necessary during the initial Gauss–Newton iterations with large penalties, we started with  $N_0 = 20$  iterations and set  $N_k = 2N_{k+1}$ . These choices were stable and yielded good results for all data sets.

The parameters in the Gauss–Newton iteration were chosen according to a quasi-optimality criterion. The initial penalties  $\alpha_0, \beta_0$  were chosen such that the norm of

the residual  $\|F(u^1, c^1) - g\|$  after the first iteration was roughly 3/4 of the initial residual, and the reduction factors  $q_\alpha, q_\beta$  were set such that each further iteration roughly reduced the residual by a factor of 1/2. The iteration was terminated once the achieved reduction factor fell below 3/4. This led to the choice  $\alpha_0 = 1$ ,  $\beta_0 = 2$ ,  $q_\alpha = q_\beta = 1/10$ , and 5 Gauss–Newton iterations for the radial data set. For the pseudorandom data set,  $\beta_0 = 1$ ,  $q_\beta = 1/5$  and 6 Gauss–Newton iterations were used.

Since the TV regularization parameter is continually decreased during the Gauss–Newton iteration, the final reconstruction will typically not show strong signs of TV filtering such as a cartoon-like appearance. A more pronounced TV effect can be achieved if the decrease of the regularization parameter is stopped at the desired level. To illustrate this, we will also show reconstructions where we have set  $\beta_{k+1} = \max(\beta_{\min}, q_\beta \beta_k)$  with  $\beta_{\min} = 5 \cdot 10^{-3}$  for  $L^2$ , TV and TGV regularization (with otherwise unchanged parameters).

Due to the difference in functionals and normalization of raw data and reconstructed magnitude images, it is not sensible for comparison purposes to use the above values of  $\beta_{\min}$  as regularization parameters for TV filtering. Instead, based on visual inspection, an optimal parameter was chosen for each data set. For the subsampled data with  $R = 6$ , this was  $\beta_1^* = 5 \cdot 10^{-3}$ , while the accelerated in-vivo data with  $R = 4$  required  $\beta_2^* = 1.5 \cdot 10^{-2}$ .

## RESULTS

Figure 1 shows a partition of a water phantom reconstructed with IRGN and IRGN-TV from pseudorandomly subsampled 3D data using acceleration factors  $R = 4$  and  $R = 10$ . The reduced noise amplification and artifact removal characteristics of IRGN-TV are clearly visible for both acceleration factors. In the case of moderate acceleration with  $R = 4$ , the final TV regularization parameter  $\beta_{\min}$  was set to zero. For  $R = 10$ ,  $\beta_{\min} = 5 \cdot 10^{-3}$  was used to achieve a stronger TV regularization and a better removal of artifacts. It must be noted that the phantom only consists of regions that are piecewise constant, and therefore the underlying assumption of the TV penalty is fulfilled. However, in the case of  $R = 10$  with increased TV regularization, staircasing artifacts can be observed in regions where modulations caused by the inhomogeneity of the coil sensitivities affect the reconstructed image.

The above findings are confirmed for in-vivo conditions. IRGN and IRGN-TV reconstructions (with  $\beta_{\min} = 0$ ) of the retrospectively pseudorandomly subsampled T2 weighted data of the brain are displayed in Fig. 2, together with a sum-of-squares

image obtained from the fully sampled data. Shown are results for acceleration factors of  $R = 4$ ,  $R = 6$  and  $R = 8$  (defined as the ratio of total to acquired points on the underlying Cartesian sampling grid) as well as difference images to the fully sampled SOS image. With larger acceleration factors, an increasing amount of noise amplification and residual incoherent aliasing can be observed in the IRGN reconstructions, while this effect is reduced with IRGN-TV. The difference is especially noticeable in the magnified details shown in Fig. 3.

The comparison of IRGN-TV with IRGN followed by TV filtering of the reconstructed magnitude image is shown in Fig. 4. The conventional IRGN reconstruction of the T2 weighted brain data with  $R = 6$  (see Fig. 2) was postprocessed with a TV filter, where an optimal TV parameter was chosen based on visual inspection. To illustrate the robustness of IRGN-TV with respect to different data sets, this procedure was repeated for the accelerated brain scan obtained with the same FLASH sequence as in the phantom experiments (Fig. 1) with  $R = 4$ . Although postprocessing using an optimal TV parameter yields results that are comparable to IRGN-TV, it can be seen that this parameter value is specific to each data set: The optimal value  $\beta_1^* = 5 \cdot 10^{-3}$  for the subsampled data set (left column in Fig. 4) results in insufficient filtering for the accelerated in-vivo scan, while the optimal choice  $\beta_2^* = 1.5 \cdot 10^{-2}$  for the in-vivo scan (middle column in Fig. 4) already leads to over-regularization when applied to the first data set. In contrast, the results with IRGN-TV (right column) were obtained using the same parameter set – most notably  $\beta_{\min} = 0$  – in both cases.

Figure 5 shows two slices of a water phantom acquired with subsampled radial measurements and reconstructed with IRGN and IRGN-TV. Here, 25 spokes were acquired to reconstruct a  $128 \times 128$  matrix, corresponding to an undersampling factor of approximately 8 in comparison to a fully sampled radial data set ( $128 \cdot \frac{\pi}{2} \approx 201$  spokes). Because of the high undersampling, an increased value of  $\beta_{\min} = 5 \cdot 10^{-3}$  was used in this example. Both slices show reduced noise and streaking artifacts when IRGN-TV is used. However, one of the images (top row in Fig. 5) again exhibits staircasing artifacts for the IRGN-TV solution.

Reconstructions of real-time images of the beating heart are displayed in Fig. 6 together with plots of the signal intensities across an indicated horizontal line. Results for 25 ( $R \approx 8.0$ ), 21 ( $R \approx 9.6$ ), and 19 ( $R \approx 10.6$ ) acquired spokes are shown, corresponding to image acquisition times of 50ms, 41ms, and 38ms. For all reconstructions,  $\beta_{\min} = 5 \cdot 10^{-3}$  was used. The image reconstructed from the 25 spokes data set does not show streaking artifacts for both IRGN and IRGN-TV. However, noise amplification is much stronger for IRGN. In contrast, reconstructions from 21

and 19 spokes show residual streaking artifacts due to increased subsampling, which are again reduced in the images reconstructed with IRGN-TV.

The effect of TGV regularization is demonstrated in Fig. 7. It shows highlighted regions of both phantom images affected by staircasing artifacts (pseudorandom and radial sampling, see Figs. 1 and 5) using TV as well as TGV regularization, both with  $\beta_{\min} = 5 \cdot 10^{-3}$ . The staircasing artifact is completely removed in the reconstruction with the IRGN-TGV method.

## DISCUSSION

The results from this work demonstrate that pronounced improvements in reconstruction quality of parallel imaging with nonlinear inversion can be achieved with TV and TGV based regularization instead of a simple  $L^2$  penalty. If moderate acceleration is used (e.g. Fig. 1, case of  $R = 4$ ), TV serves as a stabilization term against noise amplification, which otherwise limits the practical use of parallel imaging to low acceleration factors. In cases where acceleration is pushed to its limits (Fig. 1, case of  $R = 10$ ; Figs. 5 and 6), TV also leads to an additional removal of undersampling artifacts when combined with trajectories that produce incoherent aliasing. However, it must be noted that in this case, small image features with low contrast may also be removed during the reconstruction. This effect can be observed for some smaller structures, which are highlighted in Fig. 6. It can be seen in the top row that structures with low signal, which are still visible in conventional IRGN reconstructions, are removed when TV is applied. However, as indicated by the highlights in the middle row, objects with a slightly higher signal intensity, even though of same size, are preserved in both reconstructions. These effects are also represented in the cross-sectional plots. The elimination of structures like noise or residual streaking artifacts is visible as a reduced amount of high frequency oscillations in the IRGN-TV plots. Note that due to the nature of TV, blurring of sharp edges does not occur. This can be observed, e.g., at the sharp border of the ventricle, which is preserved equally well with IRGN and IRGN-TV.

As can be seen from Fig. 4, the results of IRGN-TV are comparable to those achievable by TV filtering, since the regularization term is the same in both cases. However, this requires (manual) parameter tuning in the case of TV filtering, even though the data sets were very similar in terms of anatomy, sampling pattern and normalization. In contrast, IRGN-TV performed similarly well on both data sets for the same choice of parameters, which can be attributed to the fact that the data

fitting term in IRGN-TV considers only the actually acquired  $k$ -space coefficients. This allows more accurate discrimination between image content and reconstruction artifacts. On the other hand, the data fitting term in TV filtering is based solely on image-space contrast. Furthermore, due to the iterative nature of the IRGN method, the noise suppression properties of strong TV regularization can take effect even for a small final regularization parameter value.

Our Matlab implementations reconstructed a single slice in a few minutes, where IRGN-TV took roughly 10% more time than IRGN (and similarly, IRGN-TGV was about 10% slower than IRGN-TV). It is possible to exploit the differentiability of the data consistency term to apply more efficient minimization algorithms such as the method of conjugate gradients in the case of IRGN or order optimal convex minimization methods such as those investigated in (33) for IRGN-TV and IRGN-TGV. Because this work focused on the effect of the different regularization techniques on image quality, the same primal-dual extra-gradient method was used as inner algorithm in all cases to allow a direct comparison for identical parameter choices. In this context, it should be noted that the parameters for the iteratively regularized Gauss–Newton method were independent of the chosen regularization term, and only depended on the trajectory type. Similarly, since the norm of the forward operator is estimated in the algorithm, the fixed parameters for the primal-dual extra-gradient methods were independent of the data set.

In this work, the flexibility to include different regularization terms was demonstrated on 2D examples, where each image (slice, i.e., 2D partition of a 3D data set, or frame of a time series) was reconstructed individually. While computationally more demanding, the extension of the penalties into a third space or a time dimension should further improve the image quality due to temporal redundancies in dynamic sequences – an effect that is well described in the literature (41–44). For example, earlier work has shown that residual streaking artifacts in radial imaging can be removed with the use of a median filter in the time dimension when using an interleaved  $k$ -space sampling scheme (45). Since the median filter can be interpreted as solving an  $L^1$  minimization problem, it is expected that the inclusion of a corresponding penalty – either in the form of an  $L^1$  penalty on the difference between the current and previous slice or frame, or of a higher-dimensional T(G)V penalty on the full data set – will yield even better results. The resulting convex minimization problems can be solved using the same primal-dual extra-gradient method as employed in this work.

Another possible extension of this work is the integration of additional information about the physical signal model into the functional in Eq. 2. An important

application is mapping of relaxation parameters in multi echo sequences (46, 47). Here, the framework of nonlinear inverse problems makes it straightforward to perform parameter identification during the reconstruction. Since the parameters are thus estimated directly from the raw data, instead of from reconstructed images, better quantification is possible.

## **CONCLUSIONS**

This work describes an approach to include additional variational penalties in parallel imaging with nonlinear inversion. The presented algorithms combine the advantages of nonlinear inversion, i.e., improved image quality through a better estimation of the coil sensitivities, with the advantageous properties of TV-based regularization terms. In addition to reducing the noise, the regularization is able to remove undersampling artifacts when combined with sampling strategies that produce incoherent aliasing, such as radial and pseudorandom sampling. The approach has the additional benefit that the inclusion of physiological parameters, either as additional penalties or as unknown parameters to be reconstructed, is straightforward.

## REFERENCES

1. Bauer F, Kannengiesser S. An alternative approach to the image reconstruction for parallel data acquisition in MRI. *Math Meth Appl Sci* 2007;30:1437–1451.
2. Ying L, Sheng J. Joint image reconstruction and sensitivity estimation in SENSE (JSENSE). *Magn Reson Med* 2007;57(6):1196–1202.
3. Uecker M, Hohage T, Block KT, Frahm J. Image reconstruction by regularized nonlinear inversion – joint estimation of coil sensitivities and image content. *Magn Reson Med* 2008;60(3):674–682.
4. Uecker M, Karaus A, Frahm J. Inverse reconstruction method for segmented multishot diffusion-weighted MRI with multiple coils. *Magn Reson Med* 2009; 62(5):1342–1348.
5. Sodickson DK, Manning WJ. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magn Reson Med* 1997; 38(4):591–603.
6. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med* 1999;42(5):952–962.
7. 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(6):1202–1210.
8. Knoll F, Clason C, Uecker M, Stollberger R. Improved reconstruction in non-cartesian parallel imaging by regularized nonlinear inversion. In *Proceedings of the 17th Scientific Meeting and Exhibition of ISMRM, Honolulu, HI. 2009*; .
9. Uecker M, Zhang S, Frahm J. Nonlinear inverse reconstruction for real-time MRI of the human heart using undersampled radial FLASH. *Magn Reson Med* 2010;63(6):1456–1462.
10. Chang T, He L, Fang T. MR image reconstruction from sparse radial samples using Bregman iteration. In *Proceedings of the 14th Scientific Meeting and Exhibition of ISMRM, Seattle. 2006*; 696.
11. Block KT, Uecker M, Frahm J. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magn Reson Med* 2007;57(6):1086–1098.
12. Liu B, King K, Steckner M, Xie J, Sheng J, Ying L. Regularized sensitivity encoding (SENSE) reconstruction using Bregman iterations. *Magn Reson Med* 2009;61(1):145–152.
13. Candes EJ, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory* 2006;52(2):489–509.

14. Donoho DL. Compressed sensing. *IEEE Transactions on Information Theory* 2006;52(4):1289–1306.
15. Lustig M, Lee J, Donoho D, Pauly J. Faster imaging with randomly perturbed, under-sampled spirals and L1 reconstruction. In *Proceedings of the 13th Scientific Meeting and Exhibition of ISMRM, Miami Beach, FL. 2005*; 685.
16. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson Med* 2007;58(6):1182–1195.
17. Gamper U, Boesiger P, Kozerke S. Compressed sensing in dynamic MRI. *Magnetic Resonance in Medicine* 2008;59(2):365–373.
18. Xiang QS. Accelerating MRI by skipped phase encoding and edge deghosting (SPEED). *Magn Reson Med* 2005;54(5):1112–1117.
19. Lustig M, Pauly JM. Calibrationless parallel imaging reconstruction by structured low-rank matrix completion. In *Proc. Intl. Soc. Mag. Reson. Med.*, 18. 2010; 2870.
20. Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Phys D* 1992;60(1-4):259–268.
21. Liang Z, Bammer R, Ji J, Pelc N, Glover G. Making better SENSE: wavelet denoising, Tikhonov regularization, and total least squares. In *Proceedings of the 10th Scientific Meeting and Exhibition of ISMRM, Honolulu, HI. 2002*; 2388.
22. Lin FH, Kwong KK, Belliveau JW, Wald LL. Parallel imaging reconstruction using automatic regularization. *Magn Reson Med* 2004;51(3):559–567.
23. Hoge WS, Brooks DH, Madore B, Kyriakos W. On the regularization of SENSE and space-RIP in parallel MR imaging. In *Proc. IEEE Int Biomedical Imaging: Nano to Macro Symp. 2004*; 241–244.
24. Hoge WS, Brooks DH, Madore B, Kyriakos WE. A tour of accelerated parallel MR imaging from a linear systems perspective. *Concepts in Magnetic Resonance Part A* 2005;27A:17–37.
25. Bredies K, Kunisch K, Pock T. Total generalized variation. *SIAM Journal on Imaging Sciences* 2010;3(3):492–526.
26. Knoll F, Bredies K, Pock T, Stollberger R. Second order total generalized variation (TGV) for MRI. *Magn Reson Med* 2010;65(2):480–491.
27. Lustig M, Pauly JM. SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space. *Magn Reson Med* 2010;64(2):457–471.
28. Bakushinsky AB, Kokurin MY. Iterative methods for approximate solution of inverse problems, volume 577 of *Mathematics and its Applications*. Dordrecht: Springer, 2004.
29. Engl HW, Hanke M, Neubauer A. Regularization of inverse problems, volume

- 375 of *Mathematics and its Applications*. Dordrecht: Kluwer Academic Publishers Group, 1996.
30. Blaschke B, Neubauer A, Scherzer O. On convergence rates for the iteratively regularized Gauss-Newton method. *IMA J Numer Anal* 1997;17(3):421–436.
  31. Hohage T. Logarithmic convergence rates of the iteratively regularized Gauss-Newton method for an inverse potential and an inverse scattering problem. *Inverse Problems* 1997;13(5):1279–1299.
  32. Pock T, Cremers D, Bischof H, Chambolle A. An algorithm for minimizing the Mumford-Shah functional. In *International Conference on Computer Vision (ICCV)*. 2009; 1133–1140.
  33. Chambolle A, Pock T. A first-order primal-dual algorithm for convex problems with applications to imaging. *J Math Imaging Vis* 2010;(online first).
  34. Buehrer M, Pruessmann KP, Boesiger P, Kozerke S. Array compression for MRI with large coil arrays. *Magn Reson Med* 2007;57(6):1131–1139.
  35. Knoll F, Clason C, Diwoky C, Stollberger R. Adapted random sampling patterns for accelerated MRI. *Magn Reson Mater Phy (MAGMA)* 2011;24(1):43–50.
  36. Frahm J, Haase A, Matthaei D. Rapid NMR imaging of dynamic processes using the FLASH technique. *Magn Reson Med* 1986;3(2):321–327.
  37. Riederer SJ, Tasciyan T, Farzaneh F, Lee JN, Wright RC, Herfkens RJ. MR fluoroscopy: technical feasibility. *Magn Reson Med* 1988;8(1):1–15.
  38. Wright RC, Riederer SJ, Farzaneh F, Rossman PJ, Liu Y. Real-time MR fluoroscopic data acquisition and image reconstruction. *Magn Reson Med* 1989; 12(3):407–415.
  39. Zhang S, Block KT, Frahm J. Magnetic resonance imaging in real time: advances using radial FLASH. *J Magn Reson Imaging* 2010;31(1):101–109.
  40. Fessler JA, Sutton BP. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Transactions on Signal Processing* 2003;51(2):560–574.
  41. Madore B, Glover GH, Pelc NJ. Unaliasing by fourier-encoding the overlaps using the temporal dimension (UNFOLD), applied to cardiac imaging and fMRI. *Magn Reson Med* 1999;42(5):813–828.
  42. Kellman P, Epstein FH, McVeigh ER. Adaptive sensitivity encoding incorporating temporal filtering (TSENSE). *Magn Reson Med* 2001;45(5):846–852.
  43. Tsao J, Boesiger P, Pruessmann KP. k-t BLAST and k-t SENSE: dynamic MRI with high frame rate exploiting spatiotemporal correlations. *Magn Reson Med* 2003;50(5):1031–1042.
  44. Xu D, King KF, Liang ZP. Improving k-t SENSE by adaptive regularization. *Magn Reson Med* 2007;57(5):918–930.

45. Uecker M, Zhang S, Voit D, Karaus A, Merboldt KD, Frahm J. Real-time MRI at a resolution of 20 ms. *NMR Biomed* 2010;23:986–994.
46. Block KT, Uecker M, Frahm J. Model-based iterative reconstruction for radial fast spin-echo MRI. *IEEE Trans Med Imaging* 2009;28(11):1759–1769.
47. Doneva M, Boernert P, Eggers H, Stehning C, Senegas J, Mertins A. Compressed sensing reconstruction for magnetic resonance parameter mapping. *Magn Reson Med* 2010;64(4):1114–1120.

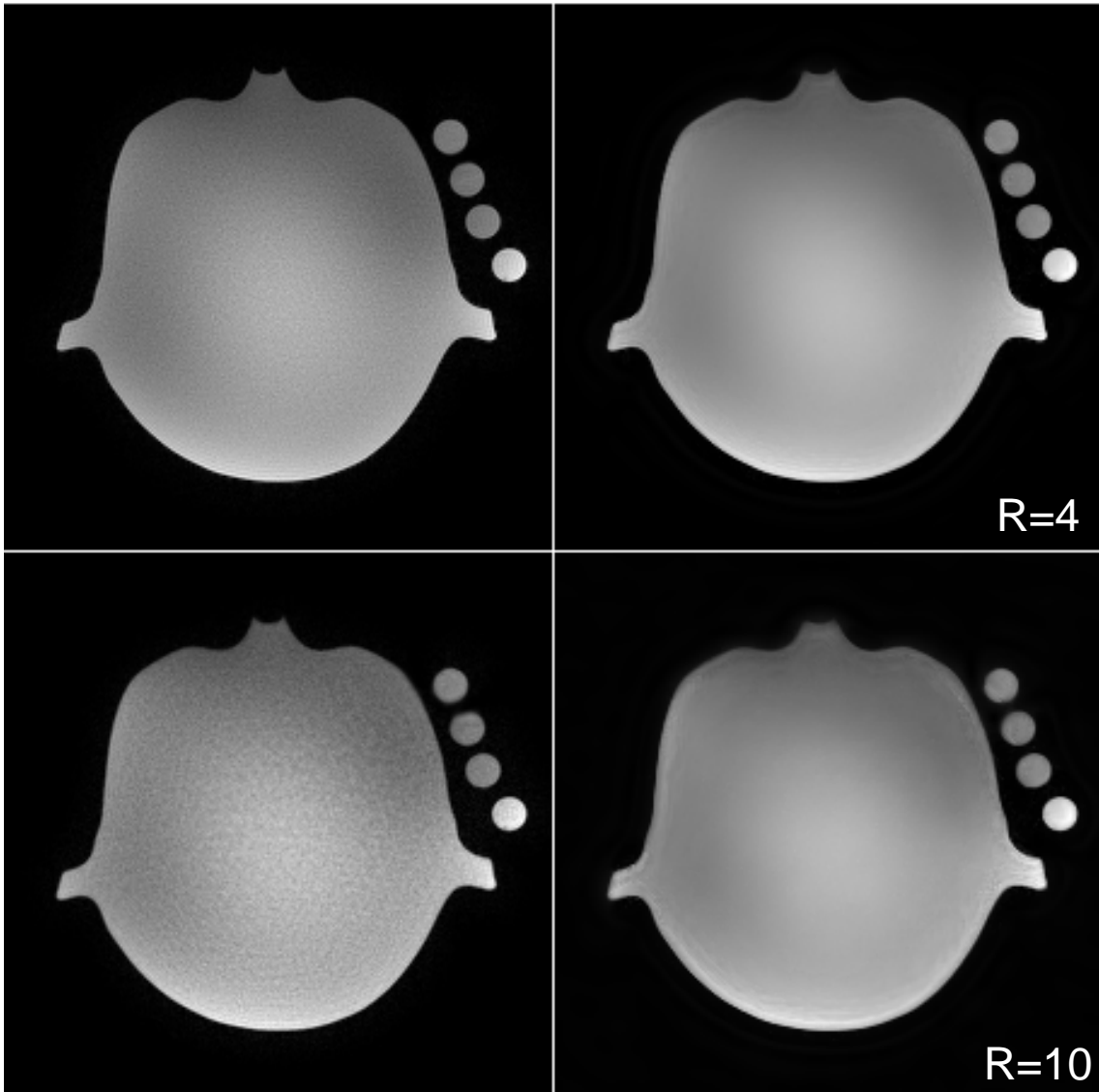
## LIST OF FIGURES

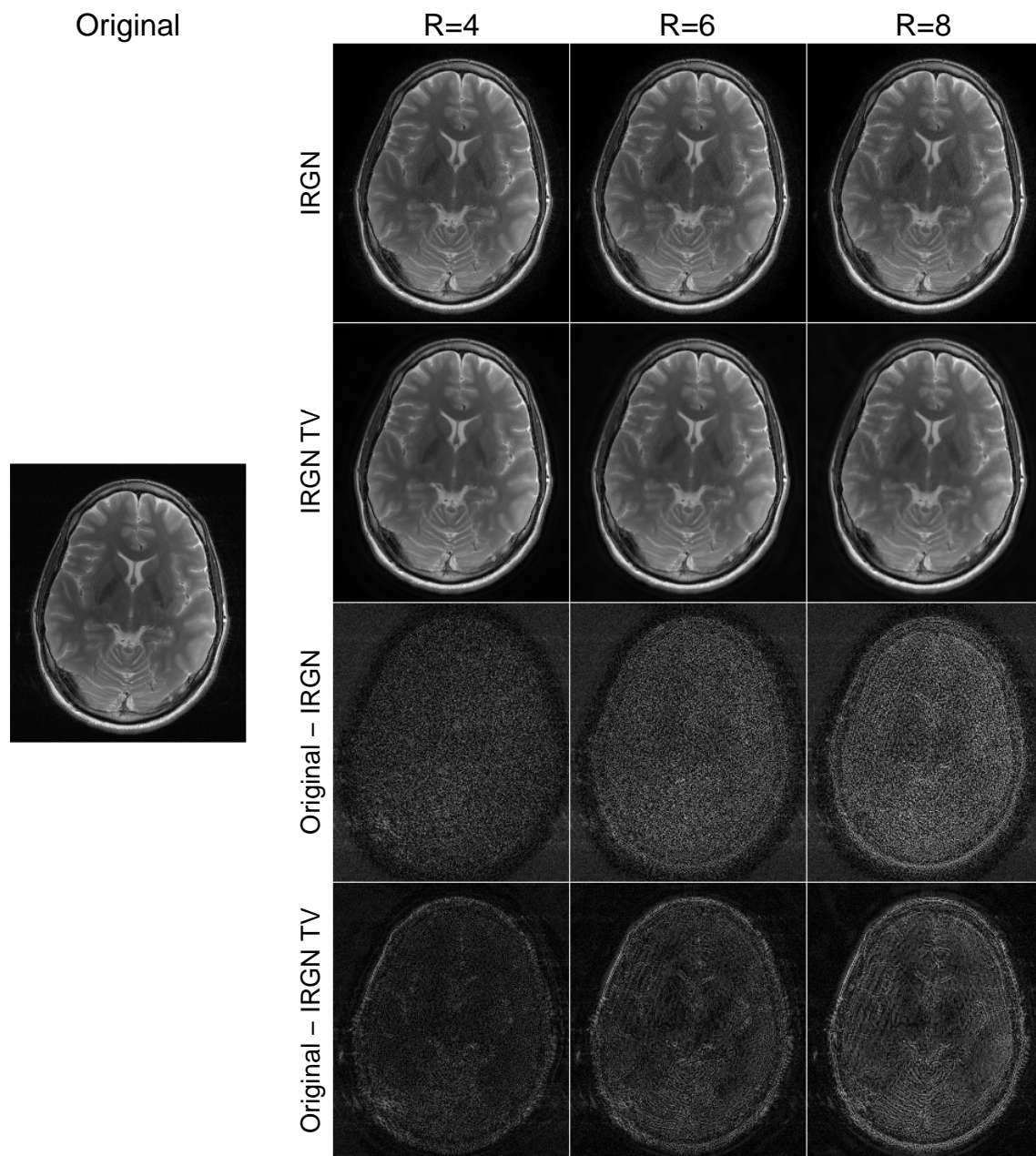
1. Comparison of IRGN (left) and IRGN-TV (right) for pseudorandom subsampling of a water phantom. Top: acceleration factor  $R = 4$ ,  $\beta_{\min} = 0$ ; bottom:  $R = 10$ ,  $\beta_{\min} = 5 \cdot 10^{-3}$ .
2. Comparison of IRGN (first row) and IRGN-TV (second row) for retrospective pseudorandom subsampling ( $\beta_{\min} = 0$ ). From left: fully sampled acquisition, acceleration factors  $R = 4$ ,  $R = 6$ ,  $R = 8$ . Difference images to the fully sampled SOS reconstruction are shown for IRGN (third row) and IRGN-TV (fourth row) and are rescaled individually for IRGN and IRGN-TV to allow better depiction of the pixel differences.
3. Magnified detail of Figure 2.
4. Comparison of IRGN-TV to TV filtering of conventional IRGN reconstructions for pseudorandom subsampling with  $R = 6$  (top, same data as Fig. 2) and accelerated in-vivo imaging with  $R = 4$  (bottom). For TV filtering, optimal regularization parameters were identified by visual inspection:  $\beta_1^* = 5 \cdot 10^{-3}$  (left, for T2 weighted TSE data) and  $\beta_2^* = 1.5 \cdot 10^{-2}$  (middle, for FLASH data). Both results for IRGN-TV (right) were obtained using the same parameter set, esp.  $\beta_{\min} = 0$ .
5. Comparison of IRGN (left) and IRGN-TV (right) for radial sampling of a phantom (25 spokes,  $\beta_{\min} = 5 \cdot 10^{-3}$ ). Shown are two slices.
6. Comparison of IRGN (left) and IRGN-TV (middle) for radial sampling of a human heart ( $\beta_{\min} = 5 \cdot 10^{-3}$ ). Top: 25 spokes. Highlighted are structures with little signal intensity that can be lost due to strong TV regularization. Middle: 21 spokes. Highlighted are structures of similar size but slightly higher signal intensity that are preserved even in case of TV regularization. Bottom: 19 spokes. The plots on the right show signal intensities across a horizontal line, indicated in the top row of the reconstruction results. The ability of IRGN-TV to preserve sharp edges is highlighted in the plot of the reconstruction from 19 spokes. The arrow marks the sharp boarder of the ventricle, which is depicted equally well with IRGN and IRGN-TV. The undesired loss of a small structure is highlighted in the plot of the reconstruction from 21 spokes. The plot crosses two adjacent vessels, which are both represented in the IRGN solution, but only the left one appears in the IRGN-TV reconstruction.

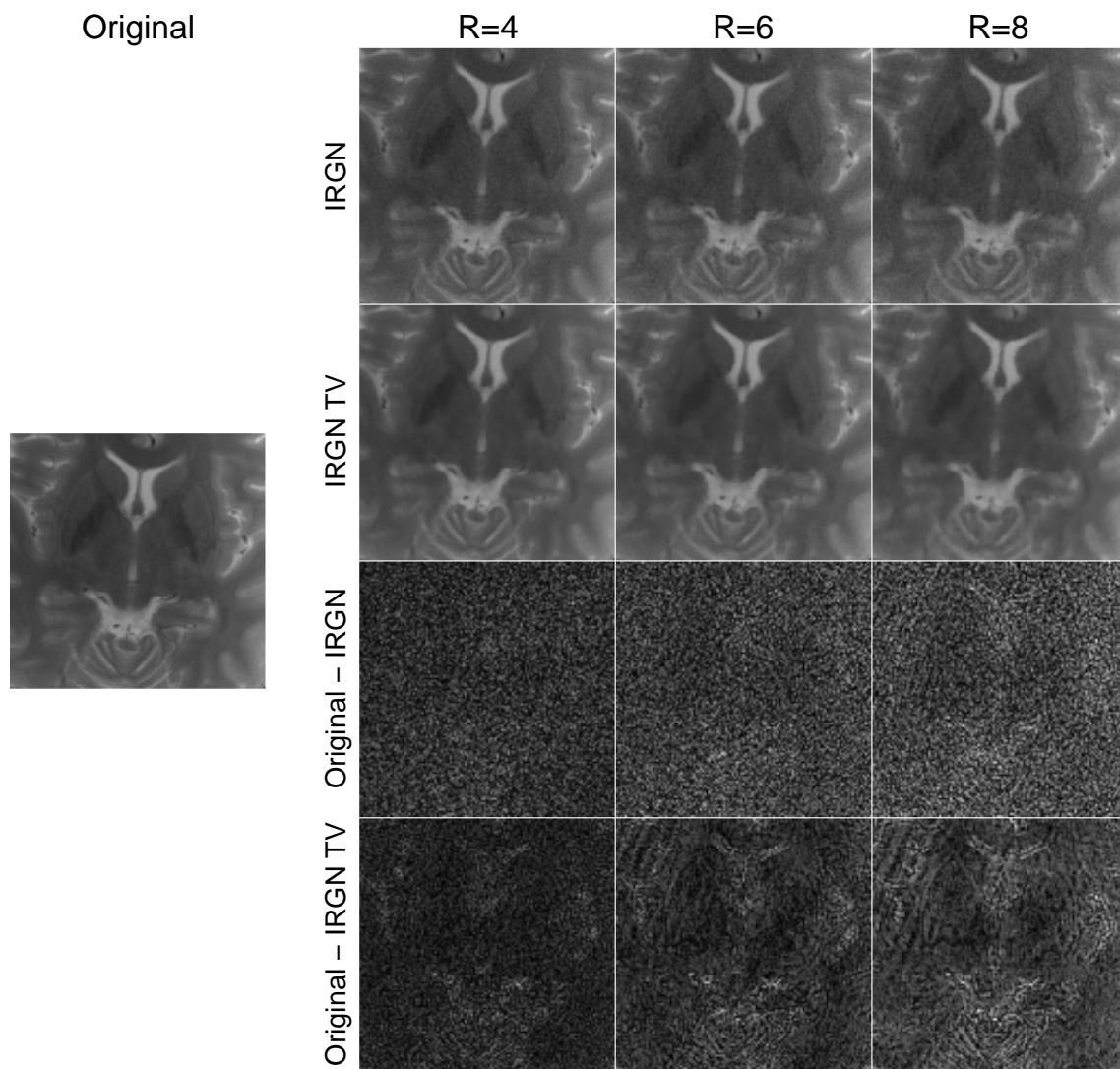
7. Comparison of IRGN-TV and IRGN-TGV for phantom data (top: pseudo-random sampling; bottom: radial sampling; both:  $\beta_{\min} = 5 \cdot 10^{-3}$ ). Left: IRGN-TV (magnified details from Figs. 1,  $R = 10$  and 5). Modulations from the coil sensitivities lead to pronounced staircasing artifacts from TV regularization. Right: IRGN-TGV. Staircasing artifacts are completely removed for TGV regularization.

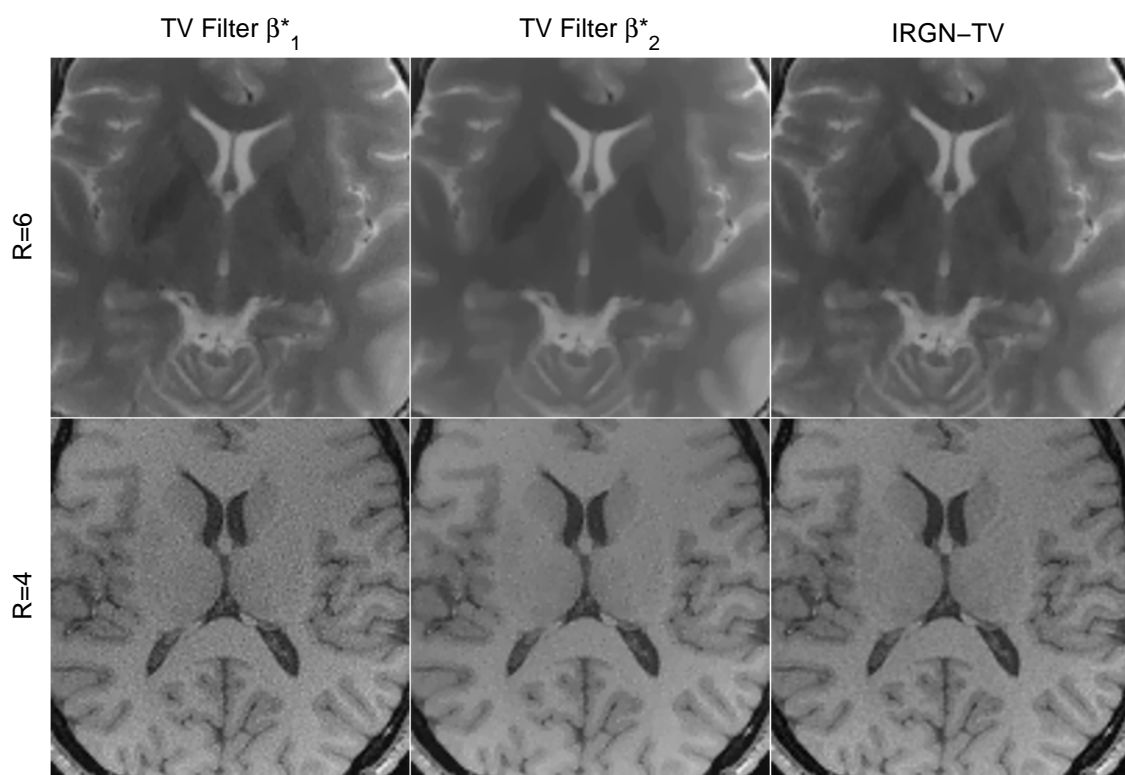
IRGN

IRGN TV



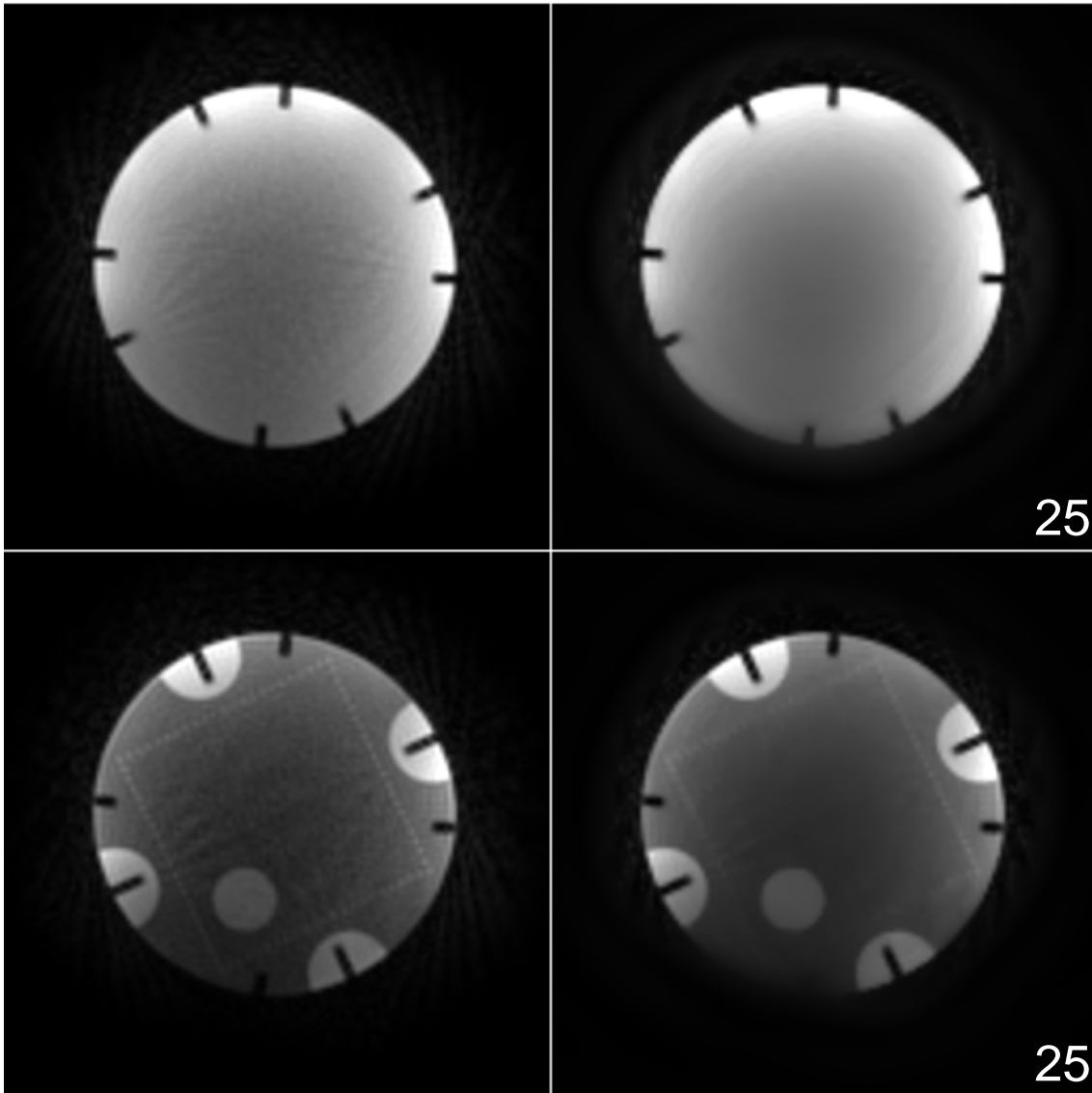


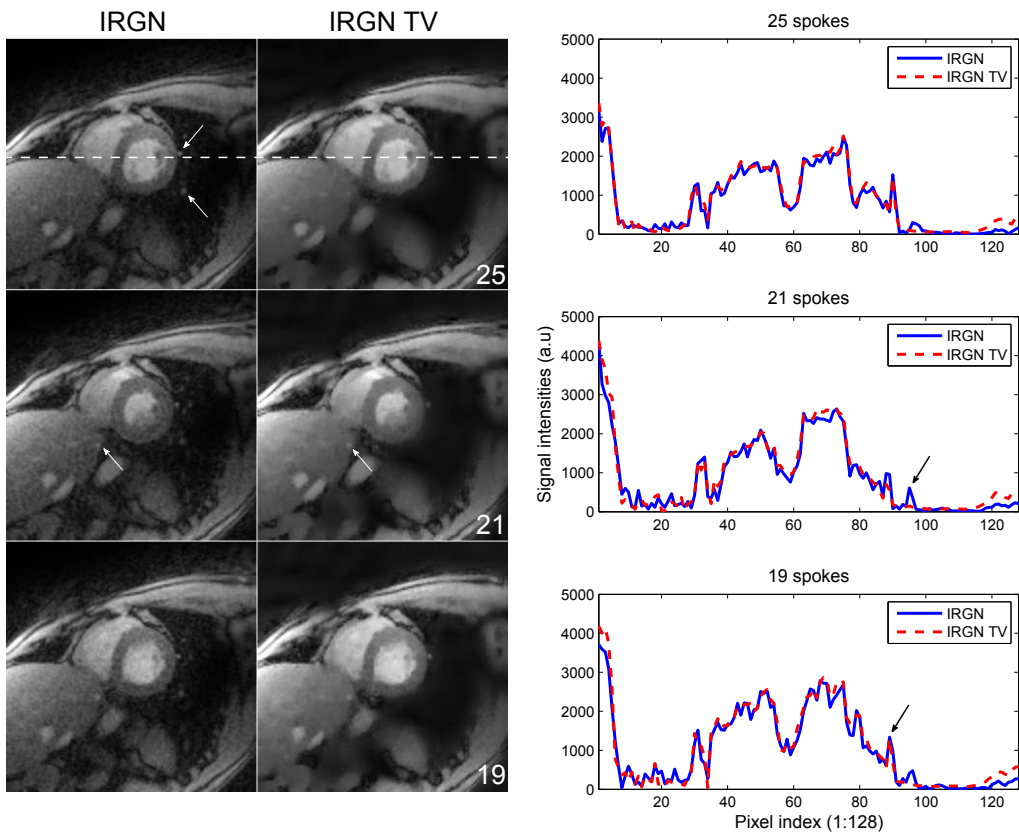




IRGN

IRGN TV





IRGN TV

IRGN TGV

