Tomographic image reconstruction via estimation of sparse unidirectional gradients

CrossMark

Adam G. Polak, Janusz Mroczka, Dariusz Wysoczański*

Faculty of Electronics, Wrocław University of Science and Technology, ul. B. Prusa 53/55, 50-317 Wrocław, Poland

A B S T R A C T

Since computed tomography (CT) was developed over 35 years ago, new mathematical ideas and computational algorithms have been continuously elaborated to improve the quality of reconstructed images. In recent years, a considerable effort can be noticed to apply the sparse solution of underdetermined system theory to the reconstruction of CT images from undersampled data. Its significance stems from the possibility of obtaining good quality CT images from low dose projections. Among diverse approaches, total variation (TV) minimizing 2D gradients of an image, seems to be the most popular method. In this paper, a new method for CT image reconstruction via sparse gradients estimation (SGE), is proposed. It consists in estimating 1D gradients specified in four directions using the iterative reweighting algorithm. To investigate its properties and to compare it with TV and other related methods, numerical simulations were performed according to the Monte Carlo scheme, using the Shepp-Logan and more realistic brain phantoms scanned at 9–60 directions in the range from 0 to 179°, with measurement data disturbed by additive Gaussians noise characterized by the relative level of 0.1%, 0.2%, 0.5%, 1%, 2% and 5%. The accuracy of image reconstruction was assessed in terms of the relative root-mean-square (RMS) error. The results show that the proposed SGE algorithm has returned more accurate images than TV for the cases fulfilling the sparsity conditions. Particularly, it preserves sharp edges of regions representing different tissues or organs and yields images of much better quality reconstructed from a small number of projections disturbed by relatively low measurement noise.

1. Introduction

Computed tomography (CT) has been one of the most powerful imaging technics used in medicine since its introduction at the turn of the 60 s and 70 s, enabling non-invasive imaging of object's cross-sections from projection data collected at different angles.

In hard-field transmission tomography (as X-ray, magnetic resonance or emission one) it is assumed that high-energy radiation beams propagate along straight lines through the object and their attenuation is caused by local energy absorption [1,2]. The relation of measured ray attenuation to tissue absorbing properties is described by the Radon Transform. According to the Central Slice Theorem, it is possible to determine attenuation information from these projections. One of the basic approaches for fast CT image reconstruction from well-sampled projections is to use the Fourier transform (the so-called *k*-space) with frequency-domain filtering and back-projection (FBP) [1]. In practice, a discrete image (consisting of pixels) is constructed from digital data. In this case, the problem can be written down as a set of large number of linear equations to be solved [1,2]. These pixels can be viewed as

unknown parameters, and the projection data – due to measurement noise – as random variables, so then the reconstruction takes a form of parameter estimation problem. Though increasing power of modern computers provides hope for using one-step procedures [3], typically iterative algorithms are used for solving these large systems of equations. Most of them are based on the method proposed by Kaczmarz [4] and further developed by Tanabe [5], and they are known as the algebraic reconstruction technique (ART) [6], with its later modifications [1,2]. Another widely used iterative approach is the expectation-maximization (EM) algorithm including the positivity constraint [7–9].

The limitation of an accumulated radiation dose can be ensured by applying a smaller number of projections or by decreasing beam energy. The former approach results in underdetermined systems of equations, while the latter one leads to a worsening of the signal-to-noise (SNR) ratio. Both the effects, particularly when being combined, introduce heavily geometric distortions and artifacts when using standard FBP reconstruction. In such cases, the improvement of image quality is possible by integrating prior knowledge and imposing

* Corresponding author.
  E-mail address: dariusz.wysoczanski@pwr.edu.pl (D. Wysoczański).

physical constraints into algebraic approaches. Among them there is a class of methods based on the minimization of a data fidelity term (commonly expressed as the least-squares), penalized by a term including prior knowledge (expressed by discrete operators imposed on pixel values), as Tikhonov regularization (TH), Twomey method or ridge regression (RR) [10]. Recently it has been shown that in the case of images with considerably large areas of dark background (pixels close to zero), the RR returns a slightly better solution than TH in terms the relative root-mean-square (RMS) reconstruction error [3].

In the above context, the method of total variation (TV) [11], were the norm of image 2D gradients is minimized (combined with the fidelity term in a constrained or penalized manner), appears to be particularly interesting. This method is well-poised to reconstruct "smooth" CT images consisting of almost homogeneous areas, representing organ or tissue cross-sections. TV was successfully explored in recent years, showing its potential to restore good quality images, even in the case of strongly underdetermined problems [12–22]. Other examples of using the concept of sparse gradients for image reconstruction include taking into account a local regional differential transformation [23] and approaches minimizing partial 1D gradients [24–26].

In this paper we also exploit the local consistency of CT medical images in terms of image gradients, however abandoning their global minimization in favor of focusing on nonzero gradients in chosen directions. Because of the nature of locally homogeneous CT images, in every direction only a limited number of gradients are noticeably nonzero, namely the ones at interfaces of different tissues. In this perspective, the image reconstruction problem comes down to searching the sparse solution (where only a very small number of entries are nonzero) of (usually) underdetermined system of linear equations – a method called thereinafter the *Sparse Gradients Estimation* (SGE). The aim of this paper is to present SGE and its main properties, comparing it with conceptually similar approaches.

The methodology of finding sparse solutions has been developed extensively for the last decades and successfully applied in a variety of areas, as spectral analysis, compressed sensing, image reconstruction or restoration, geophysics, and many others (e.g. [27–41]). In tomography, constrained TV was first proposed as a method for image reconstruction from sparse samples of its discrete Fourier transform [42]. Other sparsifying transformations were also investigated [43]. A survey on main algorithms for sparse solutions is given in [44]. Just recently Sakhaee et al. presented a pioneer work focusing on the estimation of image gradients instead of traditionally reconstructing the image pixels [45]. According to our best knowledge, other approaches to CT image reconstruction by direct gradients estimation besides proposed here SGE and the above method, have not been reported so far.

Specifically, the proposed SGE method consists in vectorizing the image in the main directions (vertical, horizontal and two diagonal ones), iterative finding the four interrelated sparse 1D gradients using the reweighting algorithm, and lastly – in their reversion to the averaged final image. This study shows that the SGE method outperforms other considered approaches to the reconstruction of CT images (including TV) in terms of the relative RMS error when the sparsity conditions are fulfilled. Additionally, it effectively preserves sharp edges between neighboring regions of different properties (as organs or tissue) – a feature especially valuable in medical images interpretation.

The reminder of the paper is organized as follows. In Section 2 the methods are described in more details, including SGE (image to gradients conversion, sparse problem to be solved, iterative reweighted algorithm, intermediate images averaging), analogous methods, as well as the procedure of performed Monte Carlo simulations with the Shepp-Logan and more realistic brain phantoms. Results of numerical analyses are presented in Section 3, and their discussion is covered by Section 4. Finally, Section 5 highlights the main conclusions.

2. Methods

2.1. Sparse gradients estimation (SGE) method

To simplify notations (without losing their generality), square 2D images $\mathbf{F}$ of the size n-by-n, consisting of $N = n^2$ pixels with values f representing radiation absorptions, will be considered hereinafter. Additionally, it is assumed that the considered images are normalized to pixel values from the range [0,1].

2.1.1. Gradients of the vectorized image

Each digital image $\mathbf{F}$ can be transformed into a column vector $\mathbf{f}_d$ consisting of subsequent pixels from a chosen direction d, and vice versa. There are four basic directions, natural for a digital 2D image, i.e. vertical, horizontal, main-diagonal and counter-diagonal one, thus the same image $\mathbf{F}$ can be represented simultaneously by four vectors $\mathbf{f}_v$, $\mathbf{f}_h$, $\mathbf{f}_m$, and $\mathbf{f}_c$, respectively.

Each vector $\mathbf{g}_d$ of a discrete approximation of gradients, resulting from a vectorized image $\mathbf{f}_d$, contains $N-1$ entries: $g_d(k) = f_d(k + 1) - f_d(k)$, and complementarily, the image vector can be reconstructed from its gradients, setting $f_d(1)=f_d(N)$ (this substitution concerns only two image pixels, i.e. $f(1,1)$ for v, h and c vectorization directions, and $f(n,1)$ in m vectorization). In the case of CT images with homogeneous areas (as tissues or organs), such gradients are *sparse*, i.e. only a small number $s \ll N$ of their entries are nonzero elements.

Since the four gradients considered in this paper ($\mathbf{g}_v$, $\mathbf{g}_h$, $\mathbf{g}_m$, and $\mathbf{g}_c$,) represent the same image $\mathbf{F}$, they are not independent – each modification of a given gradient entry changes the relevant original image and, as a consequence, alternates all the three remaining gradient vectors, so for every two gradient vectors ($\mathbf{g}_k$ and $\mathbf{g}_l$) it holds:

$$\mathbf{g}_k = \mathbf{H}_{kl}\mathbf{g}_l, \tag{1}$$

where $\mathbf{H}_{kl}$ is the transformation matrix (such that $\mathbf{H}_{kl} = \mathbf{H}_{lk}^{-1}$).

2.1.2. Sparse inverse problem

The original algebraic formulation of the hard-field transmission tomography for the image vectorized in the d-direction is as follows [1]:

$$\mathbf{p}_d = \mathbf{X}_d\mathbf{f}_d, \tag{2}$$

where $\mathbf{p}_d$ is the column vector of projections (with the arrangement of entries dependent on the direction of vectorization) and $\mathbf{X}_d$ is the system matrix of the size M-by-N with entries x_{ij} proportional to joint sections of the ith ray and the jth pixel area. CT image reconstruction consists then in finding $\mathbf{f}_d$ by solving the above system of linear equations, when $\mathbf{p}_d$ is measured and $\mathbf{X}_d$ is known from the system geometry.

This problem can be rewritten using the gradients $\mathbf{g}_d$ instead of $\mathbf{f}_d$ as unknowns:

$$\mathbf{p}_d = \mathbf{A}_d\mathbf{g}_d, \tag{3}$$

where it is further assumed that $f_d(1)=0$ and the entries of the M-by-$(N-1)$ integrated matrix $\mathbf{A}_d$ are given as [2]:

$$a_d(i, j) = \sum_{k=j}^{N-1} x(i, k + 1). \tag{4}$$

Calculating gradients $\mathbf{g}_d$ from (3) belongs now to the class of sparse inverse problems. Typically, they are formulated as a constrained optimization problem [44]:

$$\hat{\mathbf{g}}_d = \underset{\mathbf{g}_d}{\operatorname{argmin}}\|\mathbf{g}_d\|_0 \quad \text{subject to} \quad \mathbf{p}_d = \mathbf{A}_d\mathbf{g}_d, \tag{5}$$

where $\|\cdot\|_0$ denotes a pseudo-norm returning the number of nonzero entries of its argument. Because the recorded projections are contaminated by measurement errors, the data-fidelity constrain should be relaxed – typically to the minimum ℓ_2-norm, and then the sparse inverse problem can be presented in the equivalent penalized form as [44]

$$\hat{\mathbf{g}}_d = \underset{\mathbf{g}_d}{argmin}\left(\|\mathbf{p}_d - \mathbf{A}_d\mathbf{g}_d\|_2^2 + \lambda\|\mathbf{g}_d\|_0\right),$$

(6)

where λ is the regularization parameter. The data-fidelity term models here the statistics of measurements, and the regularization term reflects a priori information.

2.1.3. Iterative reweighting algorithm

In low-dose CT, the number of projections is limited and the system of linear Eq. (3) becomes underdetermined ($M < N$). Fortunately, the sparse signals $\mathbf{g}_d$ can be still properly reconstructed, provided that the measurements satisfy the incoherence property. One of popular algorithms promoting the sparsity of solution is the FOCUSS (FOCal Underdetermined System Solver) [27], iteratively finding the weighted minimum norm solution (approximating the original $\|\mathbf{g}_d\|_0$ by the minimum-energy ℓ_2-norm $\|\mathbf{g}_d\|_2$). The solution of such a relaxed problem coincides with this of the original problem formulated in ℓ_0 when the conditions of compressed sensing are fulfilled [27,42].

The FOCUSS is widely used because of its straightforward computations. In this procedure, the unique minimum norm solution of the sparse problem

$$\hat{\mathbf{g}}_d = \mathbf{A}_d^+\mathbf{p}_d,$$

(7)

were $\mathbf{A}_d^+$ denotes the Moore-Penrose pseudoinverse of $\mathbf{A}_d$, is replaced by the weighed one with a diagonal matrix $\mathbf{W}_d$:

$$\hat{\mathbf{g}}_d = \mathbf{W}_d(\mathbf{A}_d\mathbf{W}_d)^+\mathbf{p}_d.$$

(8)

Introducing an auxiliary vector $\mathbf{q}_d = \mathbf{W}_d^+\mathbf{g}_d$ ($\mathbf{W}$ becomes singular when iterations proceed), the minimum norm solution

$$\hat{\mathbf{q}}_d = (\mathbf{A}_d\mathbf{W}_d)^+\mathbf{p}_d$$

(9)

finally minimizes $\|\mathbf{W}_d^+\mathbf{g}_d\|_2$. The weighting matrix $\mathbf{W}_d$ in the kth iteration is constructed from the previous solution $\hat{\mathbf{g}}_d^{(k-1)}$, and the three-steps iteration takes the following form:

$$\mathbf{W}_d^{(k)} = diag(\hat{\mathbf{g}}_d^{(k-1)}),$$
$$\hat{\mathbf{q}}_d^{(k)} = (\mathbf{A}_d\mathbf{W}_d^{(k)})^+\mathbf{p}_d,$$
$$\hat{\mathbf{g}}_d^{(k)} = \mathbf{W}_d^{(k)}\hat{\mathbf{q}}_d^{(k)}.$$

(10)

Large entries of $\mathbf{W}_d$ reduce the contribution of corresponding elements of $\hat{\mathbf{g}}_d$, attracting them to 0 (related entries of $|\hat{\mathbf{q}}_d|$ also approach 0, and the remaining ones become close to 1). In effect, the repeated weighting procedure (10) concentrates the solution in the minimal number of "active" elements of $\hat{\mathbf{g}}_d$, still reproducing the measurements sufficiently well. Additional advantage of using this procedure is the possibility to easily incorporate in it constraints or regularizing penalties [27].

2.1.4. Estimation of gradients and image reconstruction

As a core routine of SGE, we propose to simultaneously minimize the four data-fidelity terms in balance with four gradients energy, formulating the following convex functional V:

$$V = \|\mathbf{p}_v - \mathbf{A}_v\mathbf{g}_v\|_2^2 + \|\mathbf{p}_h - \mathbf{A}_h\mathbf{g}_h\|_2^2 + \|\mathbf{p}_m - \mathbf{A}_m\mathbf{g}_m\|_2^2 + \|\mathbf{p}_c - \mathbf{A}_c\mathbf{g}_c\|_2^2$$
$$+ \lambda\left(\|\mathbf{g}_v\|_2^2 + \|\mathbf{g}_h\|_2^2 + \|\mathbf{g}_m\|_2^2 + \|\mathbf{g}_c\|_2^2\right),$$

(11)

and using the above described iterative reweighting algorithm. The weighting matrix can be constructed only form one gradient vector, whereas there are four $\mathbf{g}_d$ in (11). To overcome this obstacle, each of gradient vectors is estimated separately, taking into account the relationships between them described by (1). Denoting indexes in a resulting permutation of the four gradients by d, j, k, and l, each of the four individual functionals to be actually minimized is given as:

$$V_d = \|\mathbf{p}_d - \mathbf{A}_d\mathbf{g}_d\|_2^2 + \|\mathbf{p}_j - \mathbf{A}_j\mathbf{H}_{jd}\mathbf{g}_d\|_2^2 + \|\mathbf{p}_k - \mathbf{A}_k\mathbf{H}_{kd}\mathbf{g}_d\|_2^2$$
$$+ \|\mathbf{p}_l - \mathbf{A}_l\mathbf{H}_{ld}\mathbf{g}_d\|_2^2 + \lambda\left(\|\mathbf{g}_d\|_2^2 + \|\mathbf{H}_{jd}\mathbf{g}_d\|_2^2 + \|\mathbf{H}_{kd}\mathbf{g}_d\|_2^2 + \|\mathbf{H}_{ld}\mathbf{g}_d\|_2^2\right),$$

(12)

where d stands for v, h, m, or c. Now the matrix $\mathbf{W}_d = diag(\hat{\mathbf{g}}_d)$ can be included in the equation ($\mathbf{W}_d\mathbf{W}_d^+ = \mathbf{I}$):

$$V_d = \|\mathbf{p}_d - \mathbf{A}_d\mathbf{W}_d\mathbf{W}_d^+\mathbf{g}_d\|_2^2 + \|\mathbf{p}_j - \mathbf{A}_j\mathbf{H}_{jd}\mathbf{W}_d\mathbf{W}_d^+\mathbf{g}_d\|_2^2$$
$$+ \|\mathbf{p}_k - \mathbf{A}_k\mathbf{H}_{kd}\mathbf{W}_d\mathbf{W}_d^+\mathbf{g}_d\|_2^2 + \|\mathbf{p}_l - \mathbf{A}_l\mathbf{H}_{ld}\mathbf{W}_d\mathbf{W}_d^+\mathbf{g}_d\|_2^2$$
$$+ \lambda\left(\|\mathbf{W}_d\mathbf{W}_d^+\mathbf{g}_d\|_2^2 + \|\mathbf{H}_{jd}\mathbf{W}_d\mathbf{W}_d^+\mathbf{g}_d\|_2^2 + \|\mathbf{H}_{kd}\mathbf{W}_d\mathbf{W}_d^+\mathbf{g}_d\|_2^2\right.$$
$$+ \left.\|\mathbf{H}_{ld}\mathbf{W}_d\mathbf{W}_d^+\mathbf{g}_d\|_2^2\right).$$

(13)

To simplify the notation, let us define the following matrices:

$$\mathbf{X}_d = \mathbf{A}_d\mathbf{W}_d, \quad \mathbf{X}_{jd} = \mathbf{A}_j\mathbf{H}_{jd}\mathbf{W}_d, \quad \mathbf{X}_{kd} = \mathbf{A}_k\mathbf{H}_{kd}\mathbf{W}_d, \quad \mathbf{X}_{ld} = \mathbf{A}_l\mathbf{H}_{ld}\mathbf{W}_d.$$

(14)

Then, recalling that the auxiliary vector $\mathbf{q}_d = \mathbf{W}_d^+\mathbf{g}_d$:

$$V_d(\mathbf{q}_d) = \|\mathbf{p}_d - \mathbf{X}_d\mathbf{q}_d\|_2^2 + \|\mathbf{p}_j - \mathbf{X}_{jd}\mathbf{q}_d\|_2^2 + \|\mathbf{p}_k - \mathbf{X}_{kd}\mathbf{q}_d\|_2^2 + \|\mathbf{p}_l - \mathbf{X}_{ld}\mathbf{q}_d\|_2^2$$
$$+ \lambda\left(\|\mathbf{W}_d\mathbf{q}_d\|_2^2 + \|\mathbf{H}_{jd}\mathbf{W}_d\mathbf{q}_d\|_2^2 + \|\mathbf{H}_{kd}\mathbf{W}_d\mathbf{q}_d\|_2^2 + \|\mathbf{H}_{ld}\mathbf{W}_d\mathbf{q}_d\|_2^2\right).$$

(15)

It is straightforward to find $\hat{\mathbf{q}}_d$ minimizing this quadratic functional by calculating its derivative $\partial V_d/\partial\mathbf{q}_d$ and equating it to $\mathbf{0}$:

$$\hat{\mathbf{q}}_d = [\mathbf{X}_d^T\mathbf{X}_d + \mathbf{X}_{jd}^T\mathbf{X}_{jd} + \mathbf{X}_{kd}^T\mathbf{X}_{kd} + \mathbf{X}_{ld}^T\mathbf{X}_{ld} + \lambda\mathbf{W}_d^T(\mathbf{I} + \mathbf{H}_{jd}^T\mathbf{H}_{jd}$$
$$+ \mathbf{H}_{kd}^T\mathbf{H}_{kd} + \mathbf{H}_{ld}^T\mathbf{H}_{ld})\mathbf{W}_d + \gamma\mathbf{I}]^{-1}$$
$$\cdot\left(\mathbf{X}_d^T\mathbf{p}_d + \mathbf{X}_{jd}^T\mathbf{p}_j + \mathbf{X}_{kd}^T\mathbf{p}_k + \mathbf{X}_{ld}^T\mathbf{p}_l\right),$$

(16)

where the term $\gamma\mathbf{I}$ has been additionally included to stabilize the solution at the expense of its bias – otherwise, the procedure would be numerically ill-conditioned due to the under-determination of the problem (causing that matrices $\mathbf{X}^T\mathbf{X}$ have determinants close to 0) and singularity of matrices $\mathbf{W}$. In fact, this term minimizes $\gamma\|\mathbf{q}_d\|_2^2$ and additionally promotes sparsity by attracting the entries of $\mathbf{q}_d$ to 0 instead to ± 1 [10], simultaneously suppressing the bias during iterations. Moreover, such formulation is suitable for both the over- and underdetermined systems [27,46]. Estimates $\hat{\mathbf{q}}_d$ are computed as solutions of the set of linear equations equivalent to (16) using the Cholesky factorization, instead of numerically inverting the matrix (which is symmetric and positive definite).

The next step in a given iteration is to retrieve the gradient vectors $\hat{\mathbf{g}}_d$ from $\hat{\mathbf{q}}_d$ and $\mathbf{W}_d$ (according to (10)), and then to reconstruct the four versions of the image: $\hat{\mathbf{F}}_v$, $\hat{\mathbf{F}}_h$, $\hat{\mathbf{F}}_m$, and $\hat{\mathbf{F}}_c$. The convergence can be improved by using all the assessments [25,47], so the final image estimate $\hat{\mathbf{F}}^{SGE}$ is calculated as their weighted mean, where the weight $\mu = 1/\sqrt{2}$ is assigned to the "diagonal" images $\hat{\mathbf{F}}_m$ and $\hat{\mathbf{F}}_c$[25], because they are reconstructed from geometrically more distant (diagonal) pixels:

$$\hat{\mathbf{F}}^{SGE} = \frac{\hat{\mathbf{F}}_v + \hat{\mathbf{F}}_h + \mu\hat{\mathbf{F}}_m + \mu\hat{\mathbf{F}}_c}{2 + 2\mu}.$$

(17)

At this stage, nonlinear constraints normalizing the estimate are imposed on pixel values clamped to [0,1] [18,21]. The resulting image, after vectorization and differentiation, is transformed into the four new gradient vectors $\hat{\mathbf{g}}_d$, so the weighting matrices $\mathbf{W}_d$ can be updated.

The iterations are repeated until the convergence criterion is reached or their number exceeds the maximal assumed threshold. In the proposed algorithm, the convergence condition is fulfilled when the

largest change of pixel values between succeeding iterations is smaller than 0.001, and maximally 300 iterations are allowed.

2.2. Relevant approaches

2.2.1. Total variation (TV) minimization

The primary approach for CT image reconstruction by minimizing image gradients, developed and extensively used for recent years, is TV, and this method provides a natural benchmark for SGE evaluation.

Appling TV, the discrete gradient magnitude of a 2D image at the location specified by a pixel $f(k,l)$ is defined as [11,12,42]

$$|g_{kl}| = \sqrt{(f(k+1, l) - f(k, l))^2 + (f(k, l+1) - f(k, l))^2},$$ (18a)

being the ℓ_2-norm of the 2D image gradient [17]. Using the already introduced vectorized gradient notation, for the lth column and kth row one can also write:

$$|g_{kl}| = \sqrt{g_v^2(k) + g_h^2(l)}.$$ (18b)

The functional V_{TV} to be minimized is the ℓ_1-norm of all the gradient magnitudes subject to the data-fidelity constraint, or using the penalty term $\|\mathbf{f}\|_{TV}$ [15,16,18,21,44]:

$$V_{TV}(\mathbf{f}_d) = \frac{1}{2}\|\mathbf{p}_d - \mathbf{X}_d\mathbf{f}_d\|_2^2 + \lambda\|\mathbf{f}\|_{TV} = \frac{1}{2}\|\mathbf{p}_d - \mathbf{X}_d\mathbf{f}_d\|_2^2 + \lambda\sum_{k,l}|g_{kl}|,$$ (19)

where commonly the vertical direction d for image vectorization is chosen. In this study, we applied the *Unknown Parameter Nesterov* (UPN) variant of the TV minimization algorithm, characterized by fast convergence and automatic updating its internal parameters during iterations [16].

2.2.2. Classical regularization methods

The Tikhonov regularization (TH), introduced to smoothen reconstructed signals by minimizing their derivatives (i.e. differences between adjacent samples in the discrete approach) [48], is based on a similar idea as TV. The penalty term $\|\mathbf{f}\|_{TH}$ in TH may take many forms during 2D image reconstruction (e.g. [3,49,50]). In this work, the first-order smoothing, consisting in the minimization of differences between subsequent pixels, was considered with the differentiation matrix $\mathbf{D}$ applied to the vertical and horizontal directions:

$$V_{TH}(\mathbf{f}_d) = \|\mathbf{p}_d - \mathbf{X}_d\mathbf{f}_d\|_2^2 + \lambda(\|\mathbf{D}\mathbf{f}_v\|_2^2 + \|\mathbf{D}\mathbf{f}_h\|_2^2) = \|\mathbf{p}_d - \mathbf{X}_d\mathbf{f}_d\|_2^2$$
$$+ \lambda\left(\|\mathbf{g}_v\|_2^2 + \|\mathbf{g}_h\|_2^2\right).$$ (20)

Using the bijective transformation matrix $\mathbf{T}_{hv}$, such that $\mathbf{f}_h = \mathbf{T}_{hv}\mathbf{f}_v$ and traditionally choosing vertical vectorization leads to the following image estimator:

$$\hat{\mathbf{f}}_v^{TH} = [\mathbf{X}_v^T\mathbf{X}_v + \lambda(\mathbf{D}^T\mathbf{D} + \mathbf{T}_{hv}^T\mathbf{D}^T\mathbf{D}\mathbf{T}_{hv})]^{-1}\mathbf{X}_v^T\mathbf{p}_v.$$ (21)

Neglecting the gradients at bindings between succeeding vertical or horizontal vectors, $\|\mathbf{f}\|_{TH}$ can be well approximated as follows:

$$\|\mathbf{f}\|_{TH} \approx \sum_{k=1}^{n-1}\sum_{l=1}^{n-1}(g_v^2(k) + g_h^2(l)) = \sum_{k,l}|g_{kl}|^2,$$ (22)

so the similarity and diversity between TV and TH become clear: both of them minimize the magnitudes of discrete 2D gradients $|g_{kl}|$ all over the image, however in TH they are squared.

It has been shown that ridge regression (RR), minimizing the energy of reconstructed pixels (attracting estimates that are poorly defined by data towards zero) [10,51], returns comparable results in the case of images with dark homogeneous regions [3]. For vertical vectorization, the relevant functional takes the following form:

$$V_{RR}(\mathbf{f}_v) = \|\mathbf{p}_v - \mathbf{X}_v\mathbf{f}_v\|_2^2 + \lambda\|\mathbf{f}_v\|_2^2,$$ (23)

and finally the resulting image estimator is simply given as:

$$\hat{\mathbf{f}}_v^{RR} = (\mathbf{X}_v^T\mathbf{X}_v + \lambda\mathbf{I})^{-1}\mathbf{X}_v^T\mathbf{p}_v.$$ (24)

2.2.3. Recently proposed approaches

In recent years some new approaches, striving to overcome the problem with edge blurring by TV reconstruction, have focused on using 1D (partial) image gradients, so they are conceptually close to the proposed SGE.

The first idea is to replace the (vertically and horizontally isotropic) gradient magnitudes in TV by partial gradients with own weights ω – the approach called anisotropic TV (ATV):

$$\|\mathbf{f}\|_{ATV} = \sum_d \omega_d\|\mathbf{D}\mathbf{f}_d\|_1 = \sum_d \omega_d\|\mathbf{g}_d\|_1,$$ (25)

with directions d of gradients chosen in various ways, e.g. using the basic directions v and h with individual weights assigned to each pair of pixels by taking into account edges detected in the reconstructed image [26] or in arbitrary directions with the weights ω_d expressing links between a given gradient direction and those used during scanning [24].

A similar idea has been explored by Storath et al. [25]. First, these authors propose a strategy for more isotropic discretizations of an image and furthermore use the non-convex Mumford-Shah regularizer $\|\mathbf{f}\|_{MS}$ to solve the sparse problem, split into separate vectorization directions (in a notation following image discretization assumed in this paper):

$$\|\mathbf{f}\|_{MS} = \sum_d \omega_d\|\mathbf{D}\mathbf{f}_d\|_0 = \sum_d \omega_d\|\mathbf{g}_d\|_0,$$ (26)

where the weights ω_d are calculated in the way approximating the Euclidean length of $\|\mathbf{g}_d\|_0$. Splitting the problem enables to solve its components separately for each direction, and even for each path in a given direction. The optimization subproblems with ℓ_0-terms are solved using dynamic programming and those with ℓ_2-terms by solving linear systems of equations, all in the frame of the augmented Lagrangian (assuring consistency between every pair of $\hat{\mathbf{f}}_d$) minimized by the alternating direction method of multipliers (ADMM).

Just recently another pioneering work was presented, where the authors moved from reconstructing directly the image pixels $\mathbf{f}$ to the estimation of image gradients $\mathbf{g}$, used finally to recover the image. The approach proposed by Sakhaee et al. [45] is to estimate simultaneously vertical and horizontal gradients ($\mathbf{g}_v$ and $\mathbf{g}_h$) by minimizing their ℓ_1 norm, taking into account both data fidelity and the interdependence between these partial gradients by imposing a curl-free constraint:

$$[\mathbf{g}_v^T\mathbf{g}_h^T]^T = \underset{\mathbf{g}_v, \mathbf{g}_h}{\text{argmin}}\|\mathbf{r}_v - \mathbf{X}_v\mathbf{g}_v\|_2^2 + \|\mathbf{r}_h - \mathbf{X}_h\mathbf{g}_h\|_2^2 + \lambda\left(\|\mathbf{g}_v\|_1 + \|\mathbf{g}_h\|_1\right)$$
$$+ \gamma\|\mathbf{D}_h\mathbf{g}_v - \mathbf{D}_v\mathbf{g}_h\|_2^2,$$ (27)

where $\mathbf{X}_d$ are system matrices, $\mathbf{r}_d$ are partial derivatives of projections (such that their components related to projection angles θ are: $\mathbf{r}_v^\theta = \sin(\theta)\mathbf{D}_v\mathbf{p}^\theta$ and $\mathbf{r}_h^\theta = \cos(\theta)\mathbf{D}_h\mathbf{p}^\theta$), $\mathbf{D}_d$ are 1D differential operators (such that $\mathbf{g}_d = \mathbf{D}_d\mathbf{f}$), and the last term minimizes the curl of gradient vectors.

2.3. Numerical analyses

To present SGE algorithm effectiveness and to compare it to TV and other related methods, the Shepp-Logan phantom (SL) [52] and a more realistic brain phantom (RB), originally proposed for MRI [53], were chosen for numerical investigations. They represent examples of images possessing sparse gradients [33], determined by coherent areas with piecewise-constant cross-sections in any direction. Due to its simplicity and precisely defined properties, the SL phantom was often

used in studies testing image reconstruction algorithms (e.g. [12,14,19,30,31,46]), allowing comparisons of these methods with each other. On the other hand, the SL phantom is constructed with simple ellipses that cannot reproduce realistic anatomical features. Oppositely, the RB phantom includes regions parameterized by spline contours which are well suited to represent realistic anatomical regions.

The phantoms were used to generate synthetic measurement data in the form of a sinogram, i.e. the matrix $\mathbf{S}$ containing data yielded by the Radon transform, where the projection vectors recorded for consecutive angles were arranged as subsequent columns. It was assumed that the number of rays in a projection was equal to the dimension of a square image expressed in pixels (i.e. to n), so $\mathbf{S}$ were of the n-by-p size, where p denotes the amount of projections. This returned $n \cdot p$ measurements. To simulate low-dose CT, the projections were collected in the angle range from 0 to 179° with the step of 5, 10, 15 and 20° (resulting in p=36, 18, 12 and 9) for the SL phantom and 3, 5, 10 and 15° (resulting in p=60, 36, 18 and 12) for the RB phantom. Finally, to reproduce measurement errors, sinograms were disturbed by additive Gaussian noise of the relative level η=0.1, 0.2, 0.5, 1, 2 or 5%, giving the measurement matrix $\mathbf{P}= \mathbf{S} + \mathbf{E}$, with random samples put into the matrix $\mathbf{E}$, such that [16]:

$$\eta = \frac{\|\mathbf{e}_d\|_2}{\|\mathbf{s}_d\|_2} \cdot 100\% = \frac{\|\mathbf{E}\|_F}{\|\mathbf{S}\|_F} \cdot 100\%, \tag{28}$$

where $\mathbf{e}_d$ and $\mathbf{s}_d$ represent the matrices $\mathbf{E}$ and $\mathbf{S}$ vectorized in the direction d, and $\|\cdot\|_F$ denotes the Frobenius norm. Admittedly, the emission of photons by an X-ray source follows the Poisson process, and the radiation reaching detectors continues to be Poisson distributed. These current-integration detectors, however, are sources of Gaussian noise, so finally, also due to a polyenergetic nature of X-ray generation, system calibration and logarithmic transformation of measured data, the noise in Radon's projections is better approximated by Gaussian distribution [54]. The measurement matrices $\mathbf{P}$ prepared in such a way were vectorized in the four directions, yielding $\mathbf{p}_v$, $\mathbf{p}_h$, $\mathbf{p}_m$, and $\mathbf{p}_c$. The system matrices $\mathbf{X}_d$ of the size M-by-N, where $M=n \cdot p$ and $N=n^2$, were determined by solving the forward problem, as described in [3]. This implies that for $p < n$ the reconstruction problems were undetermined.

The performance of the investigated reconstruction methods, implemented in the penalized forms, depends on the value of the regularization parameter λ, and additionally on γ in SGE. Taking advantage of iterative calculations, the optimal value of γ is found successively in the following way: at each step three images are reconstructed using the best γ from the last iteration and the values of $\gamma/10$ and $10 \cdot \gamma$. Then three corresponding values of the minimized functional V are computed using Eq. (11). The image and γ yielding the least V are accepted as the best (final) image and regularizing parameter at given iteration. Our intention was to compare the possibly best results yielded by the algorithms, so *a priori* optimal values of λ were first estimated. To do this, one has to remember that the same sinogram disturbed by various realizations of random measurement errors will yield different image reconstructions, characterized by distinct accuracy. In such a random problem, we can assess expected values of reconstruction errors for each setting (image, scanning, method, λ) using Monte Carlo methodology, by reconstructing images many times from a sinogram disturbed by new random errors drawn from the same distribution. Finally, for each image resolution, number of projections, noise level and reconstruction algorithm with λ alternated from 10^{-15} to 10 (in the logarithmic scale), 10 image estimates were calculated from the same sinogram disturbed by 10 realizations of noise (around 140 thousand reconstructions in total). Knowing the true image $\mathbf{F}$, the accuracy of reconstructed $\hat{\mathbf{F}}_i(\lambda)$ was assessed by calculating the ith relative root-mean-square error $\delta_{RMS}(\lambda,i)$ [3,15,17,18,19]:

$$\delta_{RMS}(\lambda,\, i) = \frac{\sqrt{\sum_{k,l} \left(\hat{f}_{kl}(\lambda,\, i) - f_{kl}\right)^2}}{\sqrt{\sum_{k,l} f_{kl}^2}} \cdot 100\% = \frac{\|\hat{\mathbf{F}}_i(\lambda) - \mathbf{F}\|_F}{\|\mathbf{F}\|_F} \cdot 100\%, \tag{29}$$

and then the mean relative RMS error estimated from r=10 evaluations for each λ:

$$\delta_{RMS}(\lambda) = \sqrt{\frac{1}{r} \sum_{i=1}^{r} \delta_{RMS}^2(\lambda,\, i)}. \tag{30}$$

Three of $\delta_{RMS}(\lambda)$ values, located around the minimum, were then selected and a quadratic function was fitted to them (in the logarithmic scale). The argument of the fitted function minimum was remembered as the optimal value of λ for a given case.

The optimal λ were then used in the essential part of investigations, also following the Monte Carlo scheme. For each case and optimal λ, 30 images were reconstructed from the same sinogram disturbed now by 30 subsequent realizations of noise (around 9 thousand of runs in total), and then the resulting δ_{RMS} were calculated using (31) with r=30, establishing the main indicator of algorithm effectiveness, used in further comparisons (a proper selection of λ in a practical application of the algorithms with the penalty term is a separate problem, not considered in this study). Additionally, images were also reconstructed by commonly applied FBP (with the Ram-Lak filter and spline interpolation), and the corresponding δ_{RMS} were assessed, too. Because the iterative algorithms of TV and SGE need an initial image to be supplied, the same FBP image was used to this end in the both cases.

The typical resolution of CT images ranges from 256×256 to 1024×1024 pixels. On the other hand, the performed comprehensive research on the methods dealing with random disturbances required the use of the Monte Carlo scheme, and thus hundreds of thousands of time-consuming reconstructions of CT images. Taking this into account, the phantoms resolutions were set to 25×25 and 51×51 pixels in Monte Carlo simulations. Using these low-resolution images shortened calculation time without losing the opportunity to compare the methods, and simultaneously enabled observation how the methods' properties change with image complexity (the higher resolution of the phantom, the relatively larger amount of non-zero gradients). For illustrative purposes, the final results are presented, however, with the image resolution of 101×101 pixels (optimal values of λ were extrapolated from the lower resolutions).

All numerical computations were performed with MATLAB (*The MathWorks, USA*), particularly using the procedures from the Image Processing Toolbox: modified Shepp-Logan phantom, Radon transform, filtered back-projection (FBP), Cholesky factorization and the pseudorandom generator of Gaussian noise.

3. Results

The results of numerical investigations on the accuracy of the TV and SGE methods used for CT image reconstruction for the SL phantom are presented in Fig. 1. Only simulations with the 25×25 pixels image scanned at 36 projections returned overdetermined systems of equations ($p > n$), which could be solved faultlessly from undisturbed measurement data (η=0) by all of the methods (Fig. 1a), including TH and RR. In this case also the 51×51 pixels image could be perfectly reconstructed using the TV and SGE algorithms for η=0. The estimation errors of these methods were always significantly lesser than the errors yielded by TH and RR (δ_{RMS} between 30% and 60%), and approximately positive-linearly dependent on η. From these two approaches, the SGE algorithm outperformed TV minimization in terms of reconstruction error in all the analyzed cases, returning roughly 1.5–2 times smaller δ_{RMS} for most of noise levels, and even better results for very small distributions. Furthermore, SGE – as the
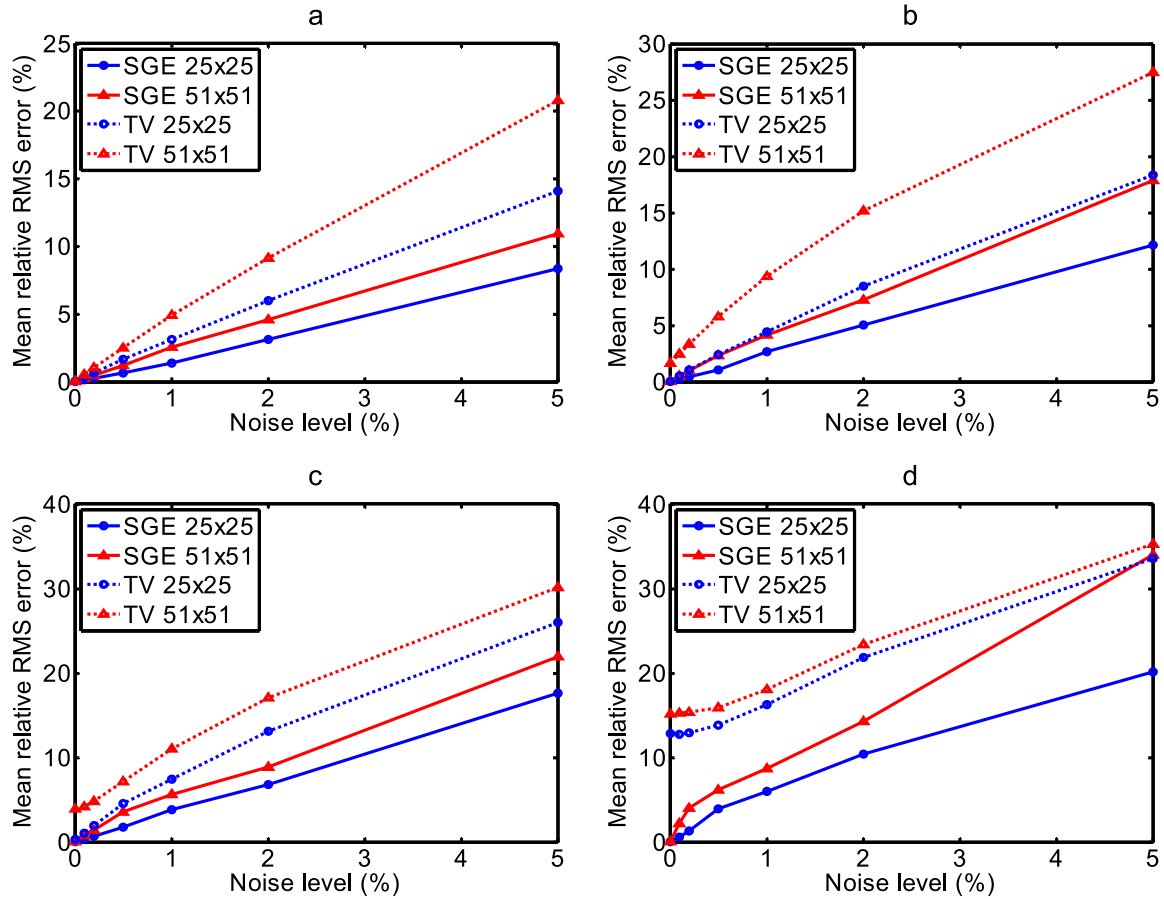
Fig. 1. Dependence of the mean relative RMS error δ_{RMS} on the noise level η for the SL phantoms (25×25 and 51×51 pixels) reconstruction by the TV and SGE algorithms, from 36 (a), 18 (b), 12 (c), and 9 (d) projections.

only method – appeared capable to perfectly reconstruct the SL image for $\eta=0$, regardless of its size and the number of investigated projections.

In the RB phantom investigations, the overdetermined systems of equations appeared for the cases of 25×25 and 51×51 pixels images scanned at 60 projections (Fig. 2a), and the 25×25 pixels image at 36 projections (Fig. 2b). Despite this, the TV algorithm was not able to perfectly reconstruct these images for $\eta=0$. Reversely, SGE returned ideal reconstructions in these instances, as well as for the under-determined cases (51×51 pixels image and $p=36$ – Fig. 2b, and 25×25 pixels image and $p=18$ – Fig. 2c). On the other hand, the dependence of the SGE reconstruction error on the noise level lost its quasi-linear character seen in Fig. 1 when SGE was applied to the more complex RB phantom. In effect, this method outperform TV only in the range of lower noise – approximately to 1–2%. Moreover, for images with higher resolution and lesser numbers of projections (Fig. 2c, d), the reconstruction error of SGE exceeded that of TV.

As one could expected, the above results show that while decreasing the number of projections, the error of image estimation is growing for both 25×25 and 51×51 pixels phantoms. Similarly, the higher resolution of an image, the larger δ_{RMS} is observed for given values of other parameters. Both effects are associated with the deteriorating ratio between the number of measurement data and the number of estimated pixel values, as well as with the changing number of nonzero image gradients.

The illustrative results of reconstruction by the FBP, TV and SGE methods, for the 101×101 pixels SL phantom scanned at 12 projections at the presence of 0.2% noise, and RB phantom scanned at 36 projections at the presence of 0.1% noise, are shown together with

the true image **F** in Figs. 3 and 4, respectively. The worst results were achieved with FBP (Figs. 3b and 4b), demonstrating its high sensitivity to data undersampling – the well-known weakness of this approach. The best estimates of the true images were obtained using SGE. TV minimization exhibited somewhat inconsistent regions and blurred edges (Figs. 3c and 4c) – the effects not visible in the images reconstructed by SGE (Figs. 3d and 4d).

Summarizing, the experimental results demonstrated the effectiveness and robustness of the proposed SGE algorithm, particularly for images with simpler content and reconstructions from less disturbed measurements.

4. Discussion

The fundamental point for the proposed SGE method is Eq. (11), where each of the four first terms represents the same measurement errors (thus three of them seem to be redundant). However, after transforming the interrelated gradients into a given direction (12) and using the reweighting algorithm, the corresponding terms in (15) become complementary, because in the ith iteration only unprecise surrogates $\hat{\mathbf{q}}_d^{(i)}$ and $\mathbf{W}_d^{(i)}$ are accessible. Numerical simulations have confirmed that using the complete functional (15) with four vectorization directions improves the convergence and accuracy of the SGE algorithm.

In reconstructions with undisturbed sinograms, when using merely one vectorization direction, the convergence of SGE was achieved only for larger numbers of projections, i.e. when the number of nonzero gradients was much smaller than the number of measurements ($s \ll M$). This stems from forcing the local image constancy only in a given
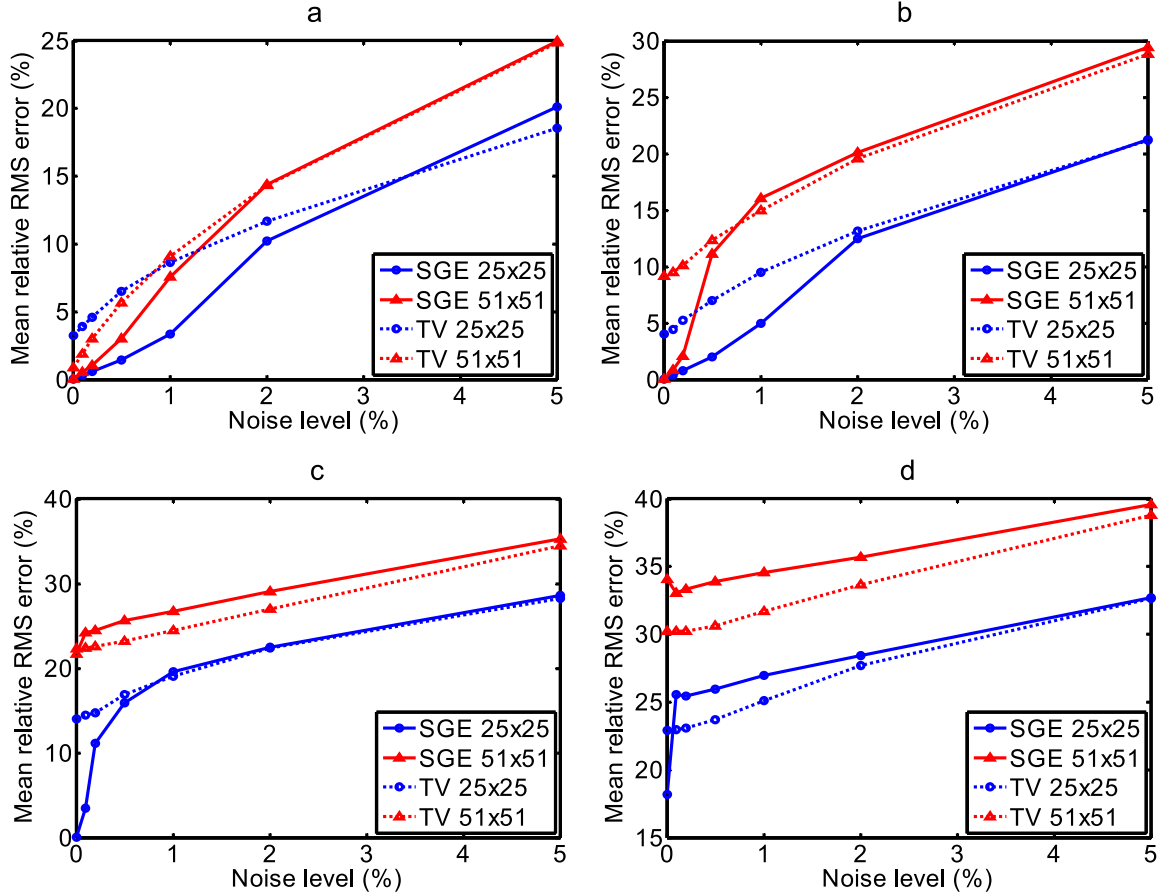
Fig. 2. Dependence of the mean relative RMS error δ_{RMS} on the noise level η for the RB phantoms (25×25 and 51×51 pixels) reconstruction by the TV and SGE algorithms, from 60 (a), 36 (b), 18 (c), and 12 (d) projections.

direction – the effect particularly well visible during first iterations (see Fig. 5). Perfect reconstructions from a smaller number of projections and undisturbed data can be achieved including just 2 or 3 directions (instead of 4) in SGE, however, at the expense of an increased number of iterations, which is essentially reduced when switching the algorithm from 2 to 3 directions (see Table 1). This observed tendency is in agreement with [25]. Generally, averaging these different versions of reconstructed image entails a regional smoothing of the image, however, the directional patterns disappear with successive iterations, and the finally reconstructed image preserves all sharp edges. Nevertheless, the image estimates in the main- and counter-diagonal directions may get stuck in the checkerboard pattern (mainly for non-convergent cases). This solution is specific because the ideal check-board image has zero both diagonal gradients. The above problem with convergence was successfully overcome by averaging the image esti-mates from all the four directions.

The proposed SGE method can be easy generalized from 4 to an arbitrary number of image vectorization directions. Including some or all of the directions that have been used when scanning an object ($K \leq p$ directions) seems to be a particularly attractive idea, because the procedure calculating the system matrices $\mathbf{X}_d$ from the problem's geometry returns also indexes of neighboring pixels in a given direction d [1]. This information can be used to compute directional gradients $\mathbf{g}_d$ in a straightforward way. Then, the generalized SGE functional V_G will take the following form:

$$V_G = \sum_{k=1}^{K}\left(\|\mathbf{p}_k - \mathbf{A}_k\mathbf{g}_k\|_2^2 + \lambda\|\mathbf{g}_k\|_2^2\right),$$ (31)

leading to K directional estimates:

$$\hat{\mathbf{q}}_d = \left[\mathbf{X}_d^T\mathbf{X}_d + \sum_{k\neq d}\mathbf{X}_{kd}^T\mathbf{X}_{kd} + \lambda\mathbf{W}_d^T\left(\mathbf{I} + \sum_{k\neq d}\mathbf{H}_{kd}^T\mathbf{H}_{kd}\right)\mathbf{W}_d + \gamma\mathbf{I}\right]^{-1}\left(\mathbf{X}_d^T\mathbf{p}_d + \sum_{k\neq d}\mathbf{X}_{kd}^T\mathbf{p}_k\right).$$ (32)

Such a formulation (and SGE with 4 directions in particular) ideally suits for parallel commutations, e.g. with the use of a GPU [13,21], because each directional estimate can be calculated independently and then used in averaging the resulting images. Thus, including additional directions will marginally affect the time of image reconstruction. It is worth to notice, however, that from the results shown in Table 1 and [25] it follows that additional directions included in the algorithm will only slightly improve the result.

The results of Monte Carlo simulations demonstrate that SGE outperforms TV in all analyzed cases for the simpler SL phantom (Fig. 1), as well as for larger numbers of projections and lower noise (up to 1–2%) in the case of more complex RB phantom (Fig. 2a, b). This effect comes from 2D gradient magnitude minimization by TV, in contrast to direct 1D gradient minimization by SGE. When minimizing the 2D gradient magnitude, 1D gradients in both perpendicular direction are to be reduced simultaneously (see (18)). As a result, e.g. in a region with a vertical contrast edge, this must lead to false image blurring in the horizontal direction, and vice-versa. This effect of edge distorting during CT image reconstruction by TV minimization was well presented by others [12–14,19], and can be also inspected in Fig. 6. On the contrary, when applying SGE operating with 1D gradients, it is possible to preserve sharp contrast edges in all of the four directions at the expense of small disturbances on flat sections (Fig. 6). Finally, the mean relative RMS error of image reconstruction is
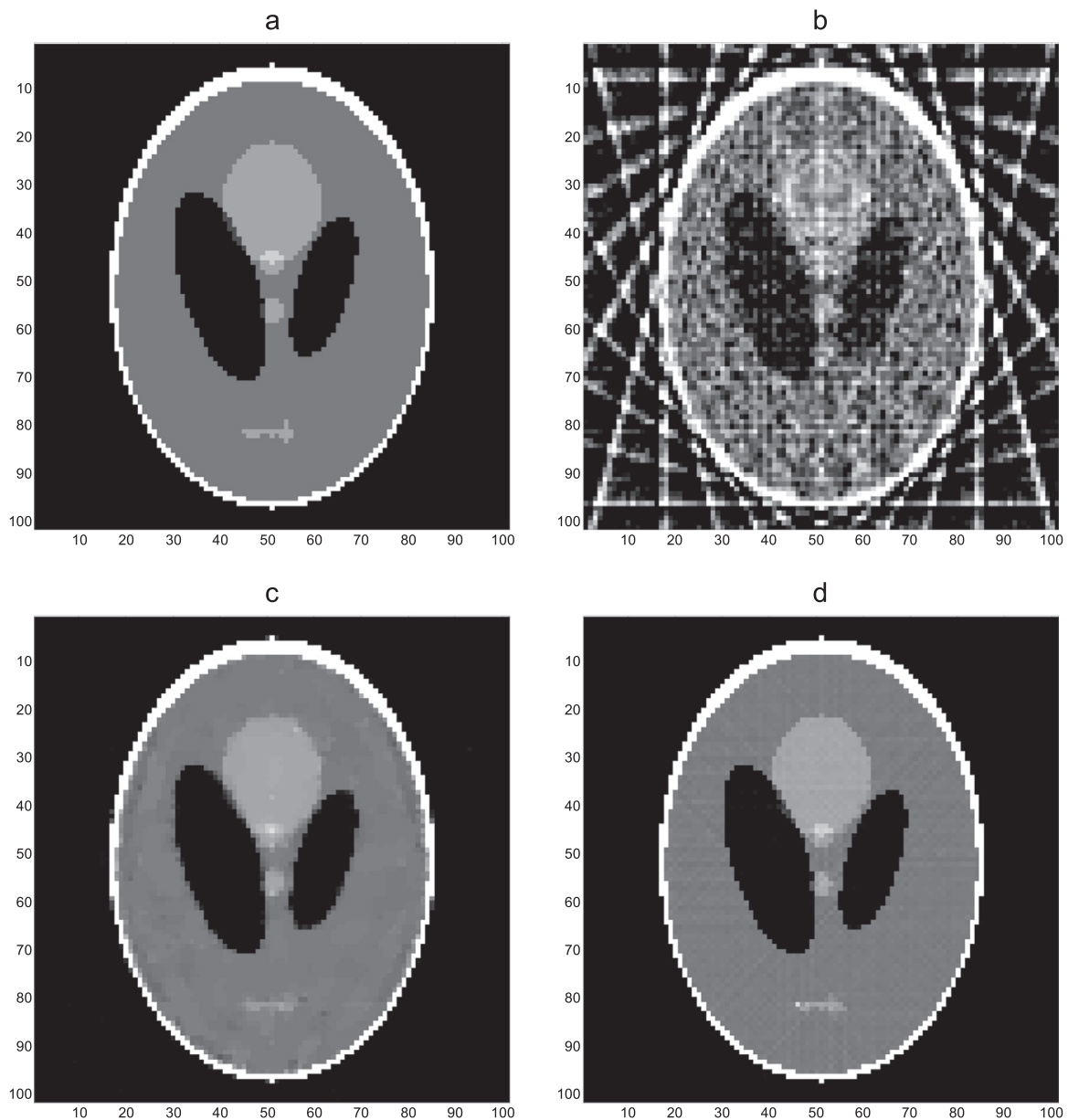
Fig. 3. Examples of 101×101 pixels SL phantom reconstruction: (a) true image, (b) FBP, (c) TV, (d) SGE (reconstructions from the same measurement data, for p=12, η=0.2 and optimal λ).

smaller when using SGE, compared to TV.

Recently, a regional (defined by fifth-order Markov connectivity) differential sparsifying transform (a special case of the nonlocal TV approach, better accounting for fine details), was also investigated, yielding reconstruction accuracy better than achieved with existing methods [23]. Other methods closely related to SGE in terms of using partial gradients are anisotropic TV (ATV) and those recently presented in [25,45]. Comparing to SGE, ATV is similar in including a chosen number of arbitrary vectorization directions [24] (vertical and horizontal in particular [26]) resulting in 1D gradients, each scaled by its own weight. This enables the reconstruction of sharp edges (with accuracy to square pixels). The simultaneous contribution of all partial gradients from all considered directions guarantees the consistency of a reconstructed image, however makes the problem algebraically much bigger in contrast to SGE, where consecutive systems of equations are solved only for gradients from one direction. Additionally, thus formulated ATV problem (with gradients from all directions) is less sparse than in SGE (gradients from one direction) in relation to the same number of sinogram data used in both algorithms, which can

result in worsened estimation by ATV when extravagating the conditions of compressed sensing. The advantage of ATS is, however, the implementation of ℓ_1-norm − a better approximation of the solved ℓ_0 problem than weighted ℓ_2-norm used in SGE.

An even better approach is the explicit use of $\|\mathbf{g}_d\|_0$ in the Mumford-Shah regularizer concerned by Storath and co-workers [25]. This method reveals several similarities to SGE. In each iteration both the algorithms include separate estimations in chosen directions that are independent from others, so these computations are easy parallelizable. In addition, the proposed strategy for deriving the weights returns the same relative weights for v, h, m and c directions as used in SGE (i.e. $1 - \sqrt{2}/2$ for v and h, and $-1/2 + \sqrt{2}/2$ for m and c). Also the interdependence between the directional estimates is taken into account. The main benefit offered by this method, stemming from arbitrary image discretizations, is the possibility of avoiding the staircase effect at the expense of algorithm complexity, exceeding the complexity of the SGE algorithm. Another difference is that it estimates directly an image $\mathbf{f}$ when minimizing the number of its gradients, whereas SGE estimates the gradients $\mathbf{g}$, used then to recover the image.
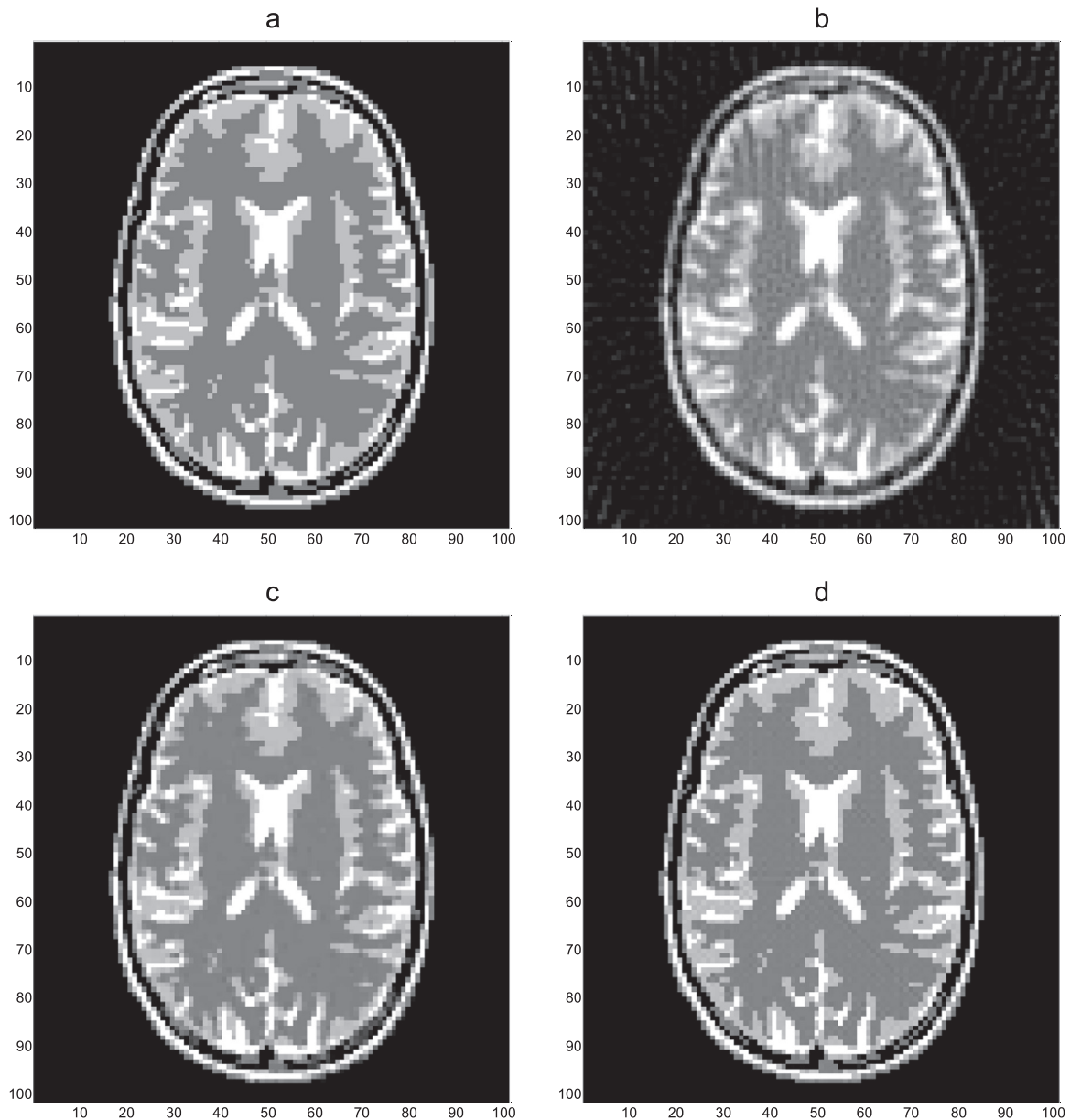
Fig. 4. Examples of 101×101 pixels RB phantom reconstruction: (a) true image, (b) FBP, (c) TV, (d) SGE (reconstructions from the same measurement data, for p=60, η=0.2 and optimal λ).

Focusing on the estimation of partial gradients **g** is a common feature of SGE and the method proposed recently by Sakhaee et al. [45]. Both the approaches exploit interdependences between directional gradients as penalty terms (based on the curl of gradient vectors or matrices **H**, respectively). A drawback of this method, in comparison to SGE, is its limitation to only two orthogonal directions (v and h), as well as simultaneous estimation of both gradient vectors, reducing problem sparsity. Additionally, the projection data **p** must be differentiated, increasing the relative content of measurement noise, i.e. reducing the SNR of data used during estimation. Instead, it assumes ℓ_1-norm minimization, while SGE uses the weighted ℓ_2-norm.

The sparse solution theory for underdetermined systems argues that the true solution can be retrieved from errorless data when it has less than $M/2$ nonzero terms [27], where M is the number of measurement data. M depends on the projections number p and the number n of rays in a projection (thus on the resolution of a reconstructed image): $M=p \cdot n$, whereas the amount of nonzero gradients s certainly hangs on the image size and its content [55]. To be able to find the exact image by the estimation of sparse gradients, the following condition must be therefore satisfied:

$$s < \frac{p_{\min} \cdot n}{2} \quad \Rightarrow \quad p_{\min} > \frac{2s}{n}, \tag{33}$$

where $p_{\min}$ denotes the minimal number of projections. Using the SL and RB phantoms, it is possible to check s for different image resolutions n and then to calculate $p_{\min}$ guaranteeing the errorless image reconstruction from undisturbed measurement data. Because the numbers of nonzero gradients are generally not the same in the four directions considered in this work (v, h, m, and c), $p_{\min}$ were calculated using the largest s for a given image resolution and put together into Table 2.

For $n \times n$ square images scanned in p projection with n rays, the ratio between $p \cdot n$ measurement data and n^2 unknowns equals to p/n. This means that for a given image size, the decreasing number of projection (graphs from (a) to (d) in Figs. 1 and 2) worsens that ratio, so the system of equations being solved becomes more underdetermined and finally, the error of reconstruction increases. Analogously, for a given number of projections, when increasing the image size (e.g.

a

b

c

d
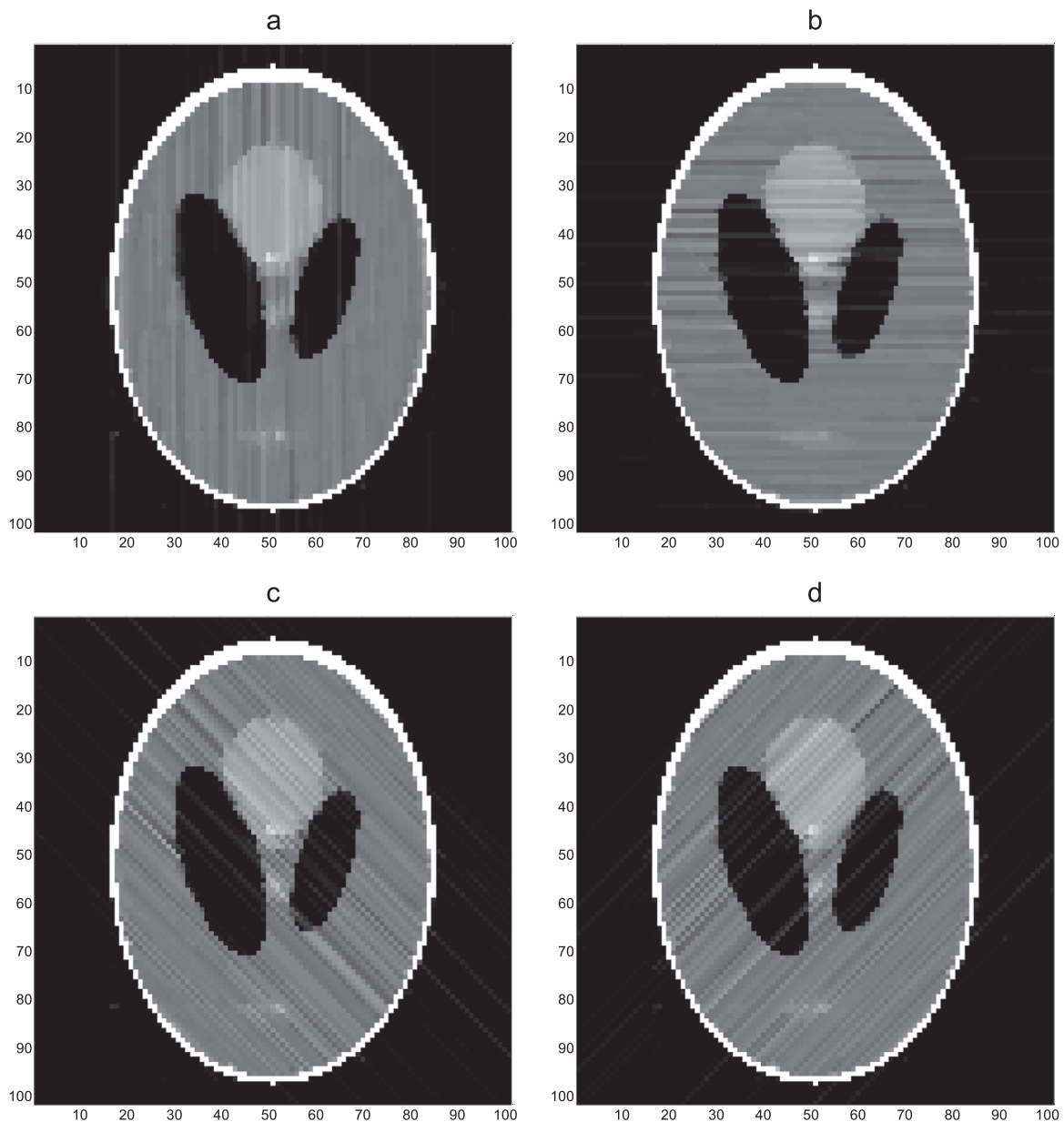
Fig. 5. Four constituent images reconstructed in the 7th iteration of SGE for the case shown in Fig. 3: (a) $\hat{\mathbf{F}}_v$, (b) $\hat{\mathbf{F}}_h$, (c) $\hat{\mathbf{F}}_m$, and (d) $\hat{\mathbf{F}}_c$.

from 25×25 to 51×51 pixels, as in Figs. 1 and 2), the error of reconstruction is also growing.

In the case of sparse problems, however, the accuracy of a solution depends also on the number s of nonzero elements. In investigations illustrated in Fig. 1, only the case (d) was characterized by too small p=9 (see Table 2), so the TV algorithm was able to find the errorless solutions for the 25×25 pixels LS phantom from undisturbed data (η=0) for p=12, 18 and 36. For larger images p_{min} is growing (Table 2), so the TV algorithm was not able to deliver the exact solutions for the 51×51 pixels image for p=9, 12 and 18. Fortunately, the proposed SGE algorithm returned the perfect solutions for the SL phantom in all the cases (inspect Fig. 1 for η=0).

In the case of images with more complex content, as the RB phantom, the number of nonzero gradients is much bigger, so is the minimal number p_{min} of projections necessary to perfectly reconstruct the image form noiseless data (see Table 2). This explains the results presented in Fig. 2, where only the cases shown in panels (a) and (b) fulfill the conditions of sparse sensing. Once again, this feature was effectively utilized only by the SGE algorithm, demonstrating its

supremacy over TV minimization, particularly when the level of measurement noise was small. An interesting observation is the ability of SGE to faultlessly reconstruct the both SL images for η=0 from 9 projections (Fig. 1d) despite p_{min} =11 and 13 in these cases, as well as the RB 25×25 pixels image from 18 projection (Fig. 2c), where p_{min} =19 (see Table 2). These outcomes result from the fact that p_{min} was calculated from the biggest directional s (the worst case), whereas some s in other directions are smaller, fulfilling the sparsity conditions and significantly improving the accuracy of the averaged image, updated inside the iterative loop. The better properties of SGE vs. TV stem also from the fact that there are more nonzero 2D gradients searched by TV compared to 1D gradients sought by SGE [30,31], so p_{min} is higher for TV than for SGE.

When the number s of nonzero gradients is too large for finding the exact sparse solution (for small p), one can suspect that the estimate may depend on the initial image, which is necessary in the TV and SGE iterative algorithms. Images reconstructed by FBP, used in the Monte Carlo simulations to this end, were highly inhomogeneous (see Fig. 3b), entering a large number of "false" nonzero initial gradients into TV and

Table 1
Numbers of iterations necessary to reconstruct the 51×51 pixels SL and RB phantoms by the SGE algorithm when using 2, 3 and 4 gradient directions.

Number of directions	Included directions	Number of iterations	
		SL	RB
2	*v h*	20	20
	v m	30	13
	v c	20	14
	h m	178	38
	h c	114	30
	m c	NC	16
3	*v h m*	12	12
	v h c	15	12
	v m c	14	10
	h m c	16	12
4	*v h m c*	11	10

SL phantom: $p=12$, $\lambda=10^{-6}$
RB phantom: $p=36$, $\lambda=10^{-5}$
NC – not converged

Table 2
Numbers s of nonzero 1D gradients for different resolutions n of the SL and RB phantoms together with the minimal number of projections p_{min} needed to faultlessly solve the sparse problem.

n		25	51	101	256	512	1024
p_{min}	SL	11	13	14	15	15	15
	RB	19	31	47	67	75	76

SGE. On the other hand, much smoother initial images (with a markedly reduced number of nonzero gradients) can be received from e.g. TH or RR with overvalued λ. Performed simulations demonstrated, however, that both TV and SGE are robust to the initial image.

The total dose in CT depends primarily on the number of projections p and ray energy. The vital question is, therefore, which of them should be preferably reduced to lower the dose, holding however a given accuracy of image reconstruction? Ray energy, assuming the approximate constancy of additive measurement noise, sets the relative noise level η. Both p and η influence the reconstruction error δ_{RMS}, as can be inspected in Figs. 1 and 2. It is possible to answer the question, converting η (given in %, see (28)) into the signal-to-noise ratio (SNR), defined in the linear scale as:

$$\text{SNR} = \frac{\|\mathbf{s}_d\|_2^2}{\|\mathbf{e}_d\|_2^2} = \left(\frac{1}{\eta \cdot 10^{-2}}\right)^2 = \frac{10^4}{\eta^2}, \tag{32}$$

and plotting these dependencies, as in Fig. 7. For example, the reconstruction error of 3% (dotted horizontal lines in Fig. 7) can be achieved for: $p=36$ and SNR$=7.5\cdot10^3$, $p=18$ and SNR$=2.4\cdot10^4$, $p=12$ and SNR$=6.3\cdot10^4$, or $p=9$ and SNR$=5.5\cdot10^5$ – for the SL phantom, and $p=60$ and SNR$=4.1\cdot10^4$ or $p=36$ and SNR$=2.1\cdot10^5$ – for the RB phantom (triangles in Fig. 7). Since a dose is proportional to $p\cdot$SNR, the resulting values are in relation as 1: 1.6: 2.8: 18.2 (SL phantom) and 1:3 (RB phantom). This shows a universal rule that it is better to record more projections using lower ray energy (resulting in lower SNR) in this case. Due to a nonlinear character of the presented

characteristics, however, the right decision depends also on the desirable reconstruction error, as well on the properties of a given CT scanner and image content.

The method for CT image reconstruction proposed in this paper includes two distinct topics: image reconstruction by estimation of four sparse gradients, and sparse gradient estimation form underdetermined data. The former one is an original idea, formulated as an alternative to popular and efficient TV minimization. The latter problem is successfully solved in SGE using the FOCUSS algorithm with ℓ_2-norm minimization. This algorithm has been effectively applied for image reconstruction, however to the sparse sampled k-space [46]. There is a group of iterative greedy algorithms for sparse solution finding in ℓ_1 space, as basic pursuit [56,57], marching pursuit [58], or orthogonal marching pursuit [38] – they need, however, a long computational time. Another interesting alternative is the lasso technique [59], Bregman iterations [39,60], the FIST algorithm [61], or iteratively reweighted ℓ_1-norm minimization [30,34]. Finally, other methods for finding sparse solutions, based on ℓ_p-norms, with $0 < p \leq 1$ [31,33] (including iterative reweighting algorithms [42,32,34]), being better approximations of ℓ_0 (which direct minimization is the NP-hard problem [44]), generally outperform the above approaches. For example, excellent results were achieved using them for image reconstruction from a highly-undersampled k-space [31,33]. The above remarks set a promising perspective for future improvements of SGE. Practical application of the method would additionally require implementing the mechanism for the (automatic) selection of optimal λ. Fortunately, λ can be modified in each iteration (as γ in SGE) using temporary estimates of an image and actual data. Potentially, one can select even different λs for each of the directions used (taking into account image anisotropy). The latter approach, however, would significantly increase algorithm's complexity, including its execution time.

5. Conclusion

In this paper, a new method for CT image reconstruction – *Sparse Gradients Estimation* (SGE) – is presented. It consists in estimating the image gradients in four directions: vertical, horizontal, main-
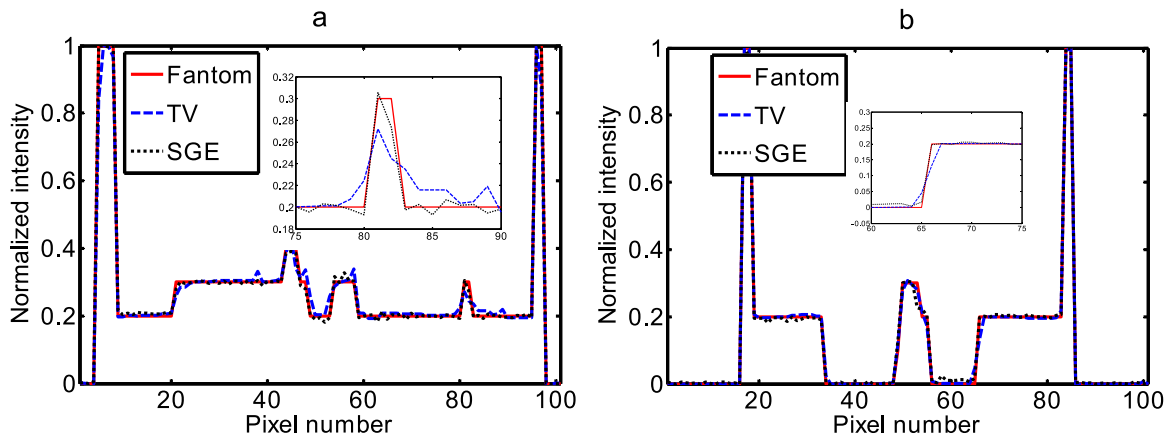
Fig. 6. Vertical (a) and horizontal (b) cross-sections of the SL phantom and images reconstructed by TV and SGE at lines 51 and 57, respectively, for the case shown in Fig. 3; inserts highlight general features of these methods.
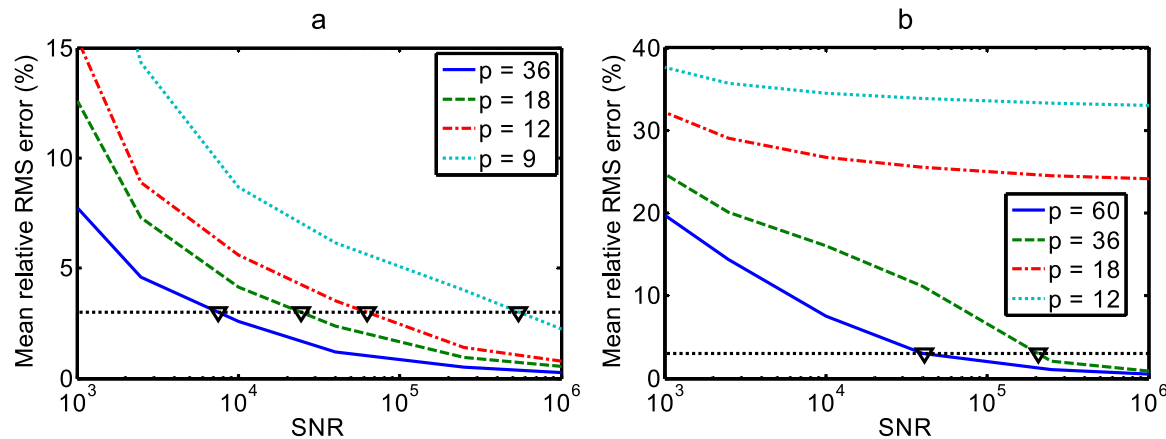
a

b

Fig. 7. Dependence of the mean relative RMS error δ_{RMS} on the linear SNR and the number of projections p for 51×51 pixels SL (a) and RB (b) phantom reconstructions by the SGE algorithm.

diagonal and counter-diagonal using FOCUSS – the iterative reweighting algorithm for finding sparse solutions of underdetermined problems, and then averaging the four estimated images in each iteration. This method allows to exactly reconstruct a CT image from errorless data when the conditions of compressed sensing are fulfilled. SGE is analogous to TV minimizations in terms of applying the sparse solution theory to find the nonzero gradients of a CT image. The fundamental difference between them is that SGE works with 1D gradients, while TV with 2D ones. In effect, the proposed SGE algorithm has returned more accurate (in terms of the mean relative RMS error) reconstructed images than TV for the cases fulfilling the sparsity conditions, particularly when the image content is not very complex. SGE preserves sharp edges of regions representing different tissues or organs, and yields images of high quality reconstructed from a small number of projections disturbed by low measurement noise. The still dynamically developed methods for finding sparse solutions of underdetermined problems set hope for future improvements of SGE.

Conflict of interest

None declared.

References

[1] A.C. Kak, M. Slaney, Principles of Computerized Tomographic Imaging, IEEE Press, New York, 1988.
[2] C.L. Byrne, Applied and Computational Linear Algebra: A First Course. Lecture Notes, 2013
[3] D. Wysoczański, J. Mroczka, A.G. Polak, Performance analysis of regularization algorithms used for image reconstruction in computed tomography, Bull. Pol. Acad. Sci.: Tech. 61 (2) (2013) 467–474.
[4] S. Kaczmarz, Angenäherte auflösung von systemen linearer gleichungen, Bull. Intern. Acad. Pol. Sci. Lett. A 35 (1937) 355–357.
[5] K. Tanabe, Projection method for solving a singular system of linear equations and its applications, Numer. Math. 17 (1971) 203–214.
[6] R. Gordon, R. Bender, G.T. Herman, Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography, J. Theor. Biol. 29 (1970) 471–481.
[7] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. B 39 (1) (1977) 1–38.
[8] Y. Teng, T. Zhang, Three penalized EM-type algorithms for PET image reconstruction, Comput. Biol. Med. 42 (6) (2012) 714–723.
[9] M.J. Rodríguez-Alvarez, A. Soriano, A. Iborra, F. Sánchez, A.J. González, P. Conde, L. Hernández, L. Moliner, A. Orero, L.F. Vidal, J.M. Benlloch, Expectation maximization (EM) algorithms using polar symmetries for computed tomography (CT) image reconstruction, Comput. Biol. Med. 43 (8) (2013) 1053–1061.
[10] A.G. Polak, An error-minimizing approach to regularization in indirect measurements, IEEE Trans. Instrum. Meas. 59 (2) (2010) 379–386.
[11] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D: Nonlinear Phenom. 60 (1) (1992) 259–268.
[12] E.Y. Sidky, C.M. Kao, X. Pan, Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT, J. X-ray Sci. Technol. 14 (2) (2006) 119–139.
[13] X. Jia, Y. Lou, R. Li, W.Y. Song, S.B. Jiang, GPU-based fast cone beam CT

reconstruction from undersampled and noisy projection data via total variation, Med. Phys. 37 (4) (2010) 1757–1760.
[14] X. Lu, Y. Sun, Y. Yuan, Optimization for limited angle tomography in medical image processing, Pattern Recogn. 44 (10) (2011) 2427–2435.
[15] M. Yan, L.A. Vese, Expectation maximization and total variation-based model for computed tomography reconstruction from undersampled data, Proc. SPIE 7961 (2011) 79612X.
[16] T.L. Jensen, J.H. Jørgensen, P.C. Hansen, S.H. Jensen, Implementation of an optimal first-order method for strongly convex total variation regularization, BIT, Numer. Math. 52 (2) (2012) 329–356.
[17] P.T. Lauzier, J. Tang, G.H. Chen, Prior image constrained compressed sensing: implementation and performance evaluation, Med. Phys. 39 (1) (2012) 66–80.
[18] T. Niu, L. Zhu, Accelerated barrier optimization compressed sensing (ABOCS) reconstruction for cone-beam CT: phantom studies, Med. Phys. 39 (7) (2012) 4588–4598.
[19] L. Liu, Z. Yin, X. Ma, Nonparametric optimization of constrained total variation for tomography reconstruction, Comput. Biol. Med. 43 (12) (2013) 2163–2176.
[20] Y. Liu, H. Shangguan, Q. Zhang, H. Zhu, H. Shu, Z. Gui, Median prior constrained TV algorithm for sparse view low-dose CT reconstruction, Comput. Biol. Med. 60 (2015) 117–131.
[21] D. Matenine, Y. Goussard, P. Després, GPU-accelerated regularized iterative reconstruction for few-view cone beam CT, Med. Phys. 42 (4) (2015) 1505–1517.
[22] Y. Xi, J. Zhao, J.R. Bennett, M.R. Stacy, A.J. Sinusas, G. Wang, Simultaneous CT-MRI reconstruction for constrained imaging geometries using structural coupling and compressive sensing, IEEE Trans. Biomed. Eng. 63 (6) (2016) 1301–1309.
[23] A. Wong, A. Mishra, P. Fieguth, D.A. Clausi, Sparse reconstruction of breast MRI using homotopic minimization in a regional sparsified domain, IEEE Trans. Biomed. Eng. 60 (3) (2013) 743–752.
[24] Z. Chen, X. Jin, L. Li, G. Wang, A limited-angle CT reconstruction method based on anisotropic TV minimization, Phys. Med. Biol. 58 (7) (2013) 2119–2141.
[25] M. Storath, A. Weinmann, J. Frikel, M. Unser, Joint image reconstruction and segmentation using the Potts model, Inverse Probl. 31 (2) (2015) 025003.
[26] J. Rong, W. Liu, P. Gao, Q. Liao, C. Jiao, J. Ma, H. Lu, CT reconstruction from few-views with anisotropic edge-guided total variance, Nucl. Instr. Meth. Phys. Res. A 820 (2016) 54–64.
[27] I.F. Gorodnitsky, B.D. Rao, Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm, IEEE Trans. Signal Process. 45 (3) (1997) 600–616.
[28] S.F. Cotter, B.D. Rao, K. Engan, K. Kreutz-Delgado, Sparse solutions to linear inverse problems with multiple measurement vectors, IEEE Trans. Signal Proces. 53 (7) (2005) 2477–2488.
[29] M.A.T. Figueiredo, R.D. Nowak, S.J. Wright, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems, IEEE J-STSP 1 (4) (2007) 586–597.
[30] E.J. Candès, M.B. Wakin, S.P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, J. Fourier Anal. Appl. 14 (5–6) (2008) 877–905.
[31] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Proces. Lett. 14 (10) (2007) 707–710.
[32] R. Chartrand, W. Yin, Iteratively reweighted algorithms for compressive sensing, IEEE Int. Conf. Acoust. Speech Signal Process. (2008) 3869–3872.
[33] J. Trzasko, A. Manduca, Highly undersampled magnetic resonance image reconstruction via homotopic-minimization, IEEE Trans. Med. Imag. 28 (1) (2009) 106–121.
[34] D. Giacobello, M.G. Christensen, M.N. Murthi, S.H. Jensen, M. Moonen, Enhancing sparsity in linear prediction of speech by iteratively reweighted 1-norm minimization, IEEE Int. Conf. Acoust. Speech Signal Process. (2010) 4650–4653.
[35] X. He, J. Liang, X. Wang, J. Yu, X. Qu, X. Wang, Y. Hou, D. Chen, F. Liu, J. Tian, Sparse reconstruction for quantitative bioluminescence tomography based on the incomplete variables truncated conjugate gradient method, Opt. Express 18 (24) (2010) 24825–24841.
[36] J. Aelterman, H.Q. Luong, B. Goossens, A. Pižurica, W. Philips, Augmented Lagrangian based reconstruction of non-uniformly sub-Nyquist sampled MRI data,

Signal Proces. 91 (12) (2011) 2731–2742.

[37] M.J. Lai, J. Wang, An unconstrained ℓ_q minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems, SIAM J. Optim. 21 (1) (2011) 82–101.

[38] D.L. Donoho, Y. Tsaig, I. Drori, J.L. Starck, Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit, IEEE Trans. Inform. Theory 58 (2) (2012) 1094–1121.

[39] Q. Liu, S. Wang, K. Yang, J. Luo, Y. Zhu, D. Liang, Highly undersampled magnetic resonance image reconstruction using two-level Bregman method with dictionary updating, IEEE Trans. Med. Imaging 32 (7) (2013) 1290–2301.

[40] Z. Zhang, B. Rao, Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation, IEEE Trans. Signal Proces. 61 (8) (2013) 2009–2015.

[41] E.A. Rashed, M. al-Shatouri, H. Kudo, Sparsity-constrained three-dimensional image reconstruction for C-arm angiography, Comput. Biol. Med. 62 (2015) 141–153.

[42] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory 52 (2) (2006) 489–509.

[43] M. Lustig, D. Donoho, J.M. Pauly, Sparse, MRI: the application of compressed sensing for rapid MR imaging, Magn. Reson. Med. 58 (6) (2007) 1182–1195.

[44] J.A. Tropp, S.J. Wright, Computational methods for sparse solution of linear inverse problems, Proc. IEEE 98 (6) (2010) 948–958.

[45] E. Sakhaee, M. Arreola, A. Entezari, Gradient-based sparse approximation for computed tomography, IEEE 12th Int. Symp. Biomed. Imag. (2015) 1608–1611.

[46] J.C. Ye, S. Tak, Y. Han, H.W. Park, Projection reconstruction MR imaging using FOCUSS, Magn. Reson. Med. 57 (4) (2007) 764–775.

[47] P. Sukovic, N.H. Clinthorne, Penalized weighted least-squares image reconstruction for dual energy X-ray transmission tomography, IEEE Trans. Med. Imag. 19 (11) (2000) 1075–1081.

[48] A.N. Tikhonov, Solution of incorrectly formulated problems and the regularization method, Sov. Math. Dokl. 4 (1963) 1035–1038.

[49] A. Lorent, R. Cierniak, Regularized image reconstruction from projections method, in: IEEE International Conference on Engineering and Telecommunication, 2014, pp. 82–86.

[50] L. Raczyński, P. Moskal, P. Kowalski, et al., Compressive sensing of signals generated in plastic scintillators in a novel J-PET instrument, Nucl. Instr. Method Phys. Res. A 786 (2015) 105–112.

[51] A.E. Hoerl, Application of ridge analysis to regression problems, Chem. Eng. Prog. 58 (1962) 54–59.

[52] L.A. Shepp, B.F. Logan, The Fourier reconstruction of a head section, IEEE Trans. Nucl. Sci. 21 (1974) 21–43.

[53] M. Guerquin-Kern, L. Lejeune, K.P. Pruessmann, M. Unser, Realistic analytical phantoms for parallel magnetic resonance imaging, IEEE Trans. Med. Imag. 31 (3) (2012) 626–636.

[54] J. Wang, H. Lu, Z. Liang, D. Eremina, G. Zhang, S. Wang, J. Chen, J. Manzione, An experimental study on the noise properties of X-ray CT sinogram data in Radon space, Phys. Med. Biol. 53 (12) (2008) 3327–3341.

[55] J. Tang, B.E. Nett, G.H. Chen, Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms, Phys. Med. Biol. 54 (19) (2009) 5781–5804.

[56] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.

[57] E. Van Den Berg, M.P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, SIAM J. Sci. Comput. 31 (2) (2008) 890–912.

[58] S.G. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Trans. Signal Proces. 41 (12) (1993) 3397–3415.

[59] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Statist. Soc. B 58 (1) (1996) 267–288.

[60] W. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing, SIAM J. Imag. Sci. 1 (1) (2008) 143–168.

[61] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imag. Sci. 2 (1) (2009) 183–202.