Joint image reconstruction and segmentation using the Potts model

Martin Storath[1], **Andreas Weinmann**[2,4]**, Jürgen Frikel**[3,4] **and Michael Unser**[1]

[1] Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Switzerland
[2] Department of Mathematics, Technische Universität München, Germany
[3] Department of Mathematics, Tufts University, USA
[4] Research Group Fast Algorithms for Biomedical Imaging, Helmholtz Zentrum München, Germany

E-mail: martin.storath@epfl.ch, andreas.weinmann@tum.de, juergen.frikel@helmholtz-muenchen.de and michael.unser@epfl.ch

CrossMark

Abstract

We propose a new algorithmic approach to the non-smooth and non-convex Potts problem (also called piecewise-constant Mumford–Shah problem) for inverse imaging problems. We derive a suitable splitting into specific sub-problems that can all be solved efficiently. Our method does not require *a priori* knowledge on the gray levels nor on the number of segments of the reconstruction. Further, it avoids anisotropic artifacts such as geometric staircasing. We demonstrate the suitability of our method for joint image reconstruction and segmentation. We focus on Radon data, where we in particular consider limited data situations. For instance, our method is able to recover all segments of the Shepp–Logan phantom from seven angular views only. We illustrate the practical applicability on a real positron emission tomography dataset. As further applications, we consider spherical Radon data as well as blurred data.

Keywords: Potts model, piecewise-constant Mumford–Shah model, image segmentation, Radon transform, spherical Radon transform, photoacoustic tomography

(Some figures may appear in colour only in the online journal)

1. Introduction

In this paper, we consider ill-posed imaging problems with incomplete data. Incomplete data are often due to technical restrictions or design issues of the imaging modality, like in the case of freehand SPECT [93] or limited-angle tomography [20, 66, 89]. Also health-related considerations lead to incomplete data. For example, sparse-angle setups are used to reduce radiation doses in x-ray tomography [6]. In addition to incompleteness, the data is usually corrupted by noise and may also suffer from blur [44, 55, 80, 97]. Altogether, this makes the reconstruction problems severely ill-posed which means that small perturbations in the data potentially lead to large errors in the reconstruction [33, 61, 62, 66]. It may even happen that certain features (singularities) are invisible from the incomplete data [42, 50, 76]. Consequently, the quality of reconstruction decreases significantly and one tends to lose fine details. Nevertheless, one can still try to reconstruct the object at a coarser scale. This is often of particular interest in medical imaging; for example, the locations of inner organs might be needed for surgery planning [77].

Classical reconstruction methods perform poorly in limited-data situations. Better approaches incorporate specific prior assumptions on the reconstruction. They are typically stated in terms of the minimization of some cost function. A popular representative is the convex total variation [27, 81]. A recent trend is to use non-convex regularizers [10, 29, 70, 77]. Although analytically and computationally more demanding, they give more freedom in the modeling and often yield better reconstructions [29, 30]. In their seminal work [65], Mumford and Shah introduced a cost functional on the piecewise-constant functions—now called the piecewise-constant Mumford–Shah functional—where the length of the discontinuity set is penalized. This functional has shown good performance, especially for the recovery of geometric macrostructures from poor data [54, 77, 78]. The piecewise-constant Mumford–Shah model also appears in statistics and image processing where it is often called *Potts model* [14–16, 74, 94]. The variational formulation of the Potts model is given by

$$\arg\ \min_u \gamma \, \| \nabla u \|_0 \, + \, \| Au - f \|_2^2. \tag{1}$$

Here, A is a linear operator (e.g., the Radon transform) and f is an element of the data space (e.g., a sinogram). A mathematically precise definition of the jump term $\| \nabla u \|_0$ is rather technical in a spatially continuous setting. However, if u is piecewise-constant and the discontinuity set of u is sufficiently regular, say, a union of C^1 curves, then $\| \nabla u \|_0$ is just the total arc length of this union. In general, the gradient ∇u is given in the distributional sense and the boundary length is expressed in terms of the $(d - 1)$-dimensional Hausdorff measure. When u is not piecewise-constant, the jump penalty is infinite [78]. The second term measures the fidelity of a solution u to the data f. The parameter $\gamma > 0$ controls the balance between data fidelity and jump penalty. The Potts model can be interpreted in two ways. On the one hand, if the imaged object is (approximately) piecewise-constant, then the solution is an (approximate) reconstruction of the imaged object. On the other hand, since a piecewise-constant solution directly induces a partitioning of the image domain, it can be seen as joint reconstruction and segmentation. Executing reconstruction and segmentation jointly typically leads to better results than performing the two steps successively [54, 77]; see figure 1.

The Potts problem is algorithmically challenging. For $A = \mathrm{id}$, it is NP-hard in dimension greater than one [88], and, for general linear operators A, it is even NP-hard for one-dimensional signals [87]. Thus, there is no hope to find a global minimizer in reasonable time. Nevertheless, due to its importance in image reconstruction and segmentation, several approximative strategies have been proposed. Bar *et al* [5] consider an Ambrosio–Tortorelli-type approximation. Kim *et al* [51] use a level-set based active contour method for
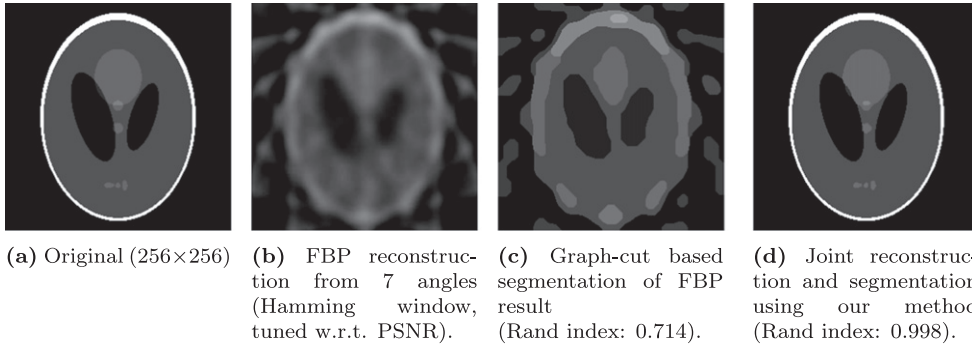
(a) Original (256×256)   **(b)** FBP reconstruction from 7 angles (Hamming window, tuned w.r.t. PSNR).   **(c)** Graph-cut based segmentation of FBP result (Rand index: 0.714).   **(d)** Joint reconstruction and segmentation using our method (Rand index: 0.998).

Figure 1. Segmentation from highly undersampled Radon data (seven projection angles). Reconstruction by a classical method and subsequent segmentation leads to unsatisfactory results. Our method produces a high-quality segmentation.

deconvolution. Ramlau and Ring [77] employ a related level-set approach for the joint reconstruction and segmentation of x-ray tomographic images; further applications are electron tomography [52] and SPECT [54]. The latter authors were the first to investigate the regularizing properties of such functionals [77, 78]. We elaborate further on Potts regularization in inverse problems and on existing algorithmic approaches in the sections 1.1 and 1.2, respectively.

In this paper, we first discretize the Potts problem as

$$u^* = \arg\min_{u\in\mathbb{R}^{m\times n}} \gamma \sum_{s=1}^{S} \omega_s \, \| \nabla_{p_s} u \|_0 \, + \, \| Au - f \|_2^2. \tag{2}$$

The symbol ∇_p denotes finite differences with respect to the displacement vector p so that $\nabla_p u = u(\,\cdot\,+ p) - u$, where $p \in \mathbb{Z}^2 \setminus \{0\}$. The symbol $\| \nabla_{p_s} u \|_0$ denotes the number of nonzero entries of $\nabla_{p_s} u$. The displacement vectors p belong to a neighborhood system $\mathcal{N} = \{p_1, ..., p_S\}$ and $\omega_1, ..., \omega_S$ are nonnegative weights. The simplest neighborhood system is made up of the two canonical basis vectors of $\mathbb{R}^2$ along with unit weights. Unfortunately, when refining the grid, this discretization converges to a limit that measures the boundary in terms of the ℓ^1 analogue of the Hausdorff measure [24]. The practical consequences are unwanted block artifacts in the reconstruction (geometric staircasing). The addition of diagonal or 'knight-move' finite differences (referring to the moves of a knight in chess) mildens such anisotropy effects [25]. We provide a general scheme for the proper choice of finite-difference systems and accompanying weights which allows for arbitrarily good approximations of the Euclidean length.

Based on (2), we propose a new minimization strategy for the Potts problem. Our key contribution is a particularly suitable splitting of the Potts problem (2) into specific sub-problems. The first subproblem is a classical Tikhonov-regularized problem with a solution that reduces to a linear system of equations. All the remaining subproblems can be solved efficiently by dynamic programming [24, 41, 65, 86]. We prove that our algorithm converges. A major advantage of our method is that neither the number of segments nor the gray-values of a solution have to be fixed *a priori*. Further, the method does not require any initial guess of the solution. Last but not least, it is highly parallelizable and easy to implement.

We demonstrate the suitability of our method for joint image reconstruction and segmentation in several setups. We consider Radon data, which appear in x-ray tomography (CT)

and in positron emission tomography (PET). In noise-free conditions, we achieve an almost perfect reconstruction/segmentation of the Shepp–Logan phantom from as few as seven projections (see figure 1). Also in the presence of noise, our method yields a high-quality segmentation from a small number of projections. On PET data of a physical phantom, we obtain a reliable segmentation of the anatomic structures. As further applications, we briefly discuss photoacoustic tomography, which is based on spherical Radon data, and the deconvolution of vector-valued data.

1.1. Potts regularization in inverse problems

In many imaging problems, the imaging operator A is not boundedly invertible. Examples are the Radon transform and the spherical Radon transform when viewed as operators on the corresponding L^2 spaces. Because of unboundedness, a direct inversion (if possible at all) may amplify small perturbations in the data. In order to attenuate such effects and produce more stable reconstructions, regularization is needed. A popular approach for regularization is by means of minimizing an energy functional of the form $\gamma J(u) + \|Au - f\|_2^2$. Here, the second term is the data-fidelity term while the first term—called the regularizing term—is a penalty that incorporates *a priori* knowledge on the solution. Classical regularizations are Besov or Sobolev seminorms which associate certain smoothness classes with the solution. A less classical choice is the TV seminorm which leads to solutions of bounded variation. In the context of sparsity regularization, non-convex functionals are also used as regularizing terms [18]. The Potts functional (1) fits into this framework by letting J be the jump penalty. It is non-convex and may be seen as a term that enforces a sparse gradient. We note that Mumford–Shah approaches (which include Potts functionals) also regularize the boundaries of the discontinuity set of the underlying signal [48].

The existence of minimizers of (1) is not guaranteed in a continuous domain setting [39, 40, 78, 87]. For example, if A is a compact convolution operator originating from a smooth function and if data is given by the impulse response $f = A\delta$, then the associated Potts functional does not have a minimizer [87]. In order to ensure the existence of minimizers, additional penalty terms such as an L^p ($1 < p < \infty$) term of the form $\|u\|_p^p$ [77, 78] or pointwise boundedness constraints [48] have been considered. We note that the existence of minimizers is guaranteed in the discrete domain setup [40, 87].

It is important to verify that the Potts model is a regularization method in the sense of inverse problems. The first work dealing with this task is [78]. Ramlau and Ring [78] assume that the solution space consists of non-degenerate piecewise-constant functions with at most k (arbitrary, but fixed) different values which are additionally bounded. Under relatively mild assumptions on the operator A, they show stability. In addition, they give a parameter choice rule and show that the solutions corresponding to the chosen parameters (which depend on the level of noise) converge to a noise-free solution as the noise vanishes. This means that the method is a regularizer in the sense of inverse problems. Related references are [53, 54] and the recent publication [48] which includes (non-piecewise-constant) Mumford–Shah functionals.

1.2. Existing algorithmic approaches to Potts and related problems

The name 'Potts model' for (1) has been retained in honor of Potts [75] who, as early as 1952, considered related jump penalties in his work in the field of statistical mechanics. The classical Potts model ($A = \mathrm{id}$) was first introduced in image processing by Geman and Geman [43] in a statistical framework. Their algorithmic approach is based on simulated annealing.

From a variational-calculus point of view, the problem was first considered by Mumford and Shah [65]. Ambrosio and Tortorelli [1] proposed an approximation by elliptic functionals. Currently, popular algorithmic approaches for the classical case $A =$ id are based on active contours [28, 32], graph cuts [14], convex relaxations [26, 74], and semi-global matching [47].

The general case of A being a linear operator has been investigated to a lesser extent. Yet, strategies based on active contours emerge as an important class. The idea is to parameterize the jump set of u by a set of contours which evolve according to a deformation force. Active contours are used for A being a convolution operator in [5] and for A being the Radon transform in [77]. Both use level sets to parameterize the active contours. A similar method has been applied to joint segmentation of SPECT/CT data [53, 54]. Level-set methods have also been applied to stabilize sparse-angle tomography [56]. In [58], the authors use explicitly parameterized contours for the application to bioluminescence tomography. In general, active contours are quite flexible as the deformation force can be easily adjusted. Their main disadvantages are that they require a good guess on the initial contour and a good guess on the expected number of gray values of the solution.

Graph cuts [14] are a popular strategy to address the classical Potts problem with $A =$ id. In [87], the inverse problem for a general A is approached by iteratively using a graph-cut strategy on a classical Potts problem, followed by Tikhonov regularization. There, the authors apply their algorithm to deconvolution. A drawback of this approach is that graph cuts need an (*a priori*) discretization of the codomain of u. Thus, one needs either a good initial guess on the values that u may take, or a very fine (and expensive) discretization of the codomain of u.

To circumvent NP hardness [14], the jump penalty is often replaced by the total variation $\| \nabla u \|_1$; see [7, 17, 19, 22, 34, 98] and the references therein. TV minimization has been used for the reconstruction from sparse Radon data in [83–85]. TV minimization is theoretically and algorithmically easier to access since it is a convex problem. The theory of compressed sensing gives conditions when the minimizers of the relaxed problem coincide with those of the original problem [22, 68]. However, the conditions are relatively restrictive and fail to apply to some problems of practical importance. In limited-data situations, TV minimizers typically lack sharply localized boundaries [29, 30]. To sharpen the results of total variation minimization, various techniques such as iterative reweighting [23], simplex constraints [60], or iterative thresholding [21] have been proposed.

In order to come closer to the solution of the initial Potts problem, many authors propose the use of non-convex priors. Chartrand [29] uses priors based on the ℓ^p norm of the gradient, $0 < p < 1$, for the reconstruction of MRI-type images. The x-ray CT setup was considered in [31, 82]. Logarithmic priors are considered by Bostan *et al* [10] for the reconstruction of biomedical images. Nikolova *et al* [69, 70] propose a whole class of non-convex regularizers which are treated using a graduated non-convexity approach.

Another approach is to transform the Potts problem to an ℓ^0 problem [4, 30, 40]. The resulting problem is separable which allows for the application of iterative hard-thresholding-type algorithms [4, 8, 9, 40]. However, using this transformation comes with constraints in form of discrete Schwarz conditions [40] as well as a data term of the form $\| ABu - f \|_2^2$ with a full triangular matrix B. While the initial system matrix A is typically sparse, the modified matrix AB is not so in general.

1.3. Organization of the article

In section 2, we present our splitting approach to the Potts problem. We start by explaining the basic approach using an anisotropic discretization of (1). Next, we discuss strategies to get more isotropic discretizations, thus attenuating the unwanted geometric staircasing effect resulting from an anisotropic discretization. Then, we present our general algorithm. We briefly discuss more general data terms and prove the convergence of our algorithm. In section 3, we apply our method to ill-posed imaging problems. In particular, we consider Radon data as well as spherical Radon data. Furthermore, we apply our technique to real PET data. Eventually, we apply our method to deconvolution problems.

2. A splitting approach for the Potts problem

In this section, we present our splitting approach for the discrete-domain Potts functional (2). It seems instructive to first describe the basic idea of the splitting in the simplest case, which is the anisotropic discretization of the length term. This discretization turns out to be anisotropic as it measures the length of the discrete boundary in the Manhattan metric, which is the metric induced by the ℓ^1 norm on $\mathbb{R}^2$. This typically leads to block artifacts in the reconstruction. To avoid this we derive appropriate neighborhood systems and corresponding weights such that the discrete length term in (2) becomes more isotropic. Based on this discretization, we formulate our general splitting for the Potts problem. The problem reduces to smaller tractable subproblems that we briefly describe. We conclude the section with a convergence result.

2.1. Basic splitting algorithm for an anisotropic discretization

In the simplest case, the discretization of the regularizing term $\| \nabla u \|_0$ uses only finite differences with respect to the coordinate axes. Thus, $p_1 = (1, 0)$, $p_2 = (0, 1)$, and the weights ω_1, ω_2 are equal to 1. Then, the regularizing term reads

$$\| \nabla u \|_0 = \| \nabla_{p_1} u \|_0 + \| \nabla_{p_2} u \|_0 = \left| \left\{ (i, j) \colon u_{ij} \neq u_{i+1,j} \right\} \right| + \left| \left\{ (i, j) \colon u_{ij} \neq u_{i,j+1} \right\} \right|.$$

Plugging this discretization into (1), we rewrite the Potts problem as the constrained optimization problem

$$\begin{aligned}
\underset{u_1, u_2, v}{\text{minimize}} \quad & \gamma \left(\| \nabla_{p_1} u_1 \|_0 + \| \nabla_{p_2} u_2 \|_0 \right) + \| Av - f \|_2^2 \,, \\
\text{subject to} \quad & v - u_1 = 0, \quad v - u_2 = 0, \quad u_1 - u_2 = 0.
\end{aligned} \tag{3}$$

The augmented Lagrangian of this optimization problem reads

$$\begin{aligned}
L\left(u_1, u_2, v, \lambda_1, \lambda_2, \varrho \right) = \gamma & \left(\| \nabla_{p_1} u_1 \|_0 + \| \nabla_{p_2} u_2 \|_0 \right) + \| Av - f \|_2^2 \\
& + \langle \lambda_1, v - u_1 \rangle + \frac{\mu}{2} \| v - u_1 \|_2^2 + \langle \lambda_2, v - u_2 \rangle \\
& + \frac{\mu}{2} \| v - u_2 \|_2^2 + \langle \varrho, u_1 - u_2 \rangle + \frac{\nu}{2} \| u_1 - u_2 \|_2^2.
\end{aligned} \tag{4}$$

The constraints are now part of the (multivariate) target functional L. The parameter $\nu > 0$ controls how strong the split variables u_1, u_2 are tied to each other and $\mu > 0$ controls their coupling to v. The variables λ_1, λ_2, and ϱ are $(m \times n)$-dimensional arrays of Lagrange multipliers. The inner product is defined as $\langle x, y \rangle = \sum_{i,j} x_{ij} y_{ij}$. Completing the squares in (4),

we reformulate L in the convenient form

$$L\left(u_1, u_2, v, \lambda_1, \lambda_2, \varrho\right) = \gamma\left(\|\nabla_{p_1}u_1\|_0 + \|\nabla_{p_2}u_2\|_0\right) + \|Av - f\|_2^2$$

$$+ \frac{\mu}{2}\left\|v - u_1 + \frac{\lambda_1}{\mu}\right\|_2^2 - \frac{\mu}{2}\left\|\frac{\lambda_1}{\mu}\right\|_2^2 + \frac{\mu}{2}\left\|v - u_2 + \frac{\lambda_2}{\mu}\right\|_2^2$$

$$- \frac{\mu}{2}\left\|\frac{\lambda_2}{\mu}\right\|_2^2 + \frac{\nu}{2}\left\|u_1 - u_2 + \frac{\varrho}{\nu}\right\|_2^2 - \frac{\nu}{2}\left\|\frac{\varrho}{\nu}\right\|_2^2. \tag{5}$$

We now use the alternating direction method of multipliers (ADMM). The basic idea of ADMM is to minimize the augmented Lagrangian L with respect to $u_1, u_2,$ and v separately and to perform gradient ascent steps with respect to the Lagrange multipliers. (We refer to [11] for a detailed treatment on optimization strategies based on ADMM.) To simplify the expressions for arg $\min_{u_s} L$, $s = 1, 2$, and arg $\min_v L$, we will use the following lemma.

Lemma 2.1.   *For $a, b_1, ..., b_N \in \mathbb{R}$ and $x_1, ..., x_N > 0$, we have that*

$$\sum_i x_i\left(a - b_i\right)^2 = \left(\sum_i x_i\right)\left(a - \frac{\sum_i b_i x_i}{\sum_i x_i}\right)^2 + C,$$

where $C \in \mathbb{R}$ is a constant that does not depend on a.

Proof.   We calculate

$$\sum_i x_i\left(a - b_i\right)^2 = a^2\left(\sum_i x_i\right) - 2a\left(\sum_i b_i x_i\right) + \sum_i b_i^2 x_i$$

$$= \left(\sum_i x_i\right)\left(a^2 - 2a\frac{\sum_i b_i x_i}{\sum_i x_i} + \frac{\sum_i b_i^2 x_i}{\sum_i x_i}\right)$$

$$= \left(\sum_i x_i\right)\left(a^2 - 2a\frac{\sum_i b_i x_i}{\sum_i x_i} + \left(\frac{\sum_i b_i x_i}{\sum_i x_i}\right)^2 - \left(\frac{\sum_i b_i x_i}{\sum_i x_i}\right)^2 - \frac{\sum_i b_i^2 x_i}{\sum_i x_i}\right)$$

$$= \left(\sum_i x_i\right)\left(\left(a - \frac{\sum_i b_i x_i}{\sum_i x_i}\right)^2 - \left(\frac{\sum_i b_i x_i}{\sum_i x_i}\right)^2 - \frac{\sum_i b_i^2 x_i}{\sum_i x_i}\right).$$

The last two terms do not depend on a, which shows the assertion. $\square$

Using lemma 2.1, we rearrange the quadratic summands of arg $\min_{u_s} L$ for $s = 1, 2$ and those of arg $\min_v L$. Doing so, we get the iteration

$$
\begin{cases}
u_1^{k+1} \in \arg\ \min_{u_1} \dfrac{2\gamma}{\mu_k + \nu_k} \, \| \nabla_{p_1} u_1 \|_0 \\[2mm]
\qquad + \left\| u_1 - \dfrac{1}{\mu_k + \nu_k} \Big(\mu_k v^k + \nu_k u_2^k + \lambda_1^k - \varrho^k \Big) \right\|_2^2, \\[3mm]
u_2^{k+1} \in \arg\ \min_{u_2} \dfrac{2\gamma}{\mu_k + \nu_k} \, \| \nabla_{p_2} u_2 \|_0 \\[2mm]
\qquad + \left\| u_2 - \dfrac{1}{\mu_k + \nu_k} \Big(\mu_k v^k + \nu_k u_1^{k+1} + \lambda_2^k + \varrho^k \Big) \right\|_2^2, \\[3mm]
v^{k+1} = \arg\ \min_{v} \| A v - f \|_2^2 + \dfrac{\mu_k + \nu_k}{2} \\[2mm]
\qquad \left\| v - \dfrac{1}{2\mu_k} \Big(\mu_k u_1^{k+1} + \mu_k u_2^{k+1} - \lambda_1^k - \lambda_2^k \Big) \right\|_2^2, \\[3mm]
\lambda_1^{k+1} = \lambda_1^k + \mu_k \Big(v^{k+1} - u_1^{k+1} \Big), \\[1mm]
\lambda_2^{k+1} = \lambda_2^k + \mu_k \Big(v^{k+1} - u_2^{k+1} \Big), \\[1mm]
\varrho^{k+1} = \varrho^k + \nu_k \Big(u_1^{k+1} - u_2^{k+1} \Big).
\end{cases}
\qquad (6)
$$

As coupling parameter, we use an increasing sequence $(\mu_k)_{k \in \mathbb{N}}$. This is a slight refinement of the standard ADMM [71].

The crucial observation is that we can solve all of the subproblems of (6) efficiently. The first line decomposes into n univariate Potts problems of the form

$$
\big(u_1^{k+1} \big)_{:,j} \in \arg\min_{g \in \mathbb{R}^m} \dfrac{2\gamma}{\mu_k + \nu_k} \, \| \nabla g \|_0
$$
$$
+ \left\| g - \dfrac{1}{\mu_k + \nu_k} \Big(\mu_k v_{:,j}^k + \nu_k \big(u_2^k \big)_{:,j} + \big(\lambda_1^k \big)_{:,j} - \varrho_{:,j}^k \Big) \right\|_2^2, \quad (7)
$$

where we use the subscript notation $x_{:,j}$ to denote the jth row of the $(m \times n)$-image x, that is, $x_{:,j} = (x_{ij})_{i=1,\dots,m}$. Analogously, we get a decomposition for the second line of (6) into the problems

$$
\big(u_2^{k+1} \big)_{i,:} \in \arg\min_{g \in \mathbb{R}^n} \dfrac{2\gamma}{\mu_k + \nu_k} \, \| \nabla g \|_0
$$
$$
+ \left\| g - \dfrac{1}{\mu_k + \nu_k} \Big(\mu_k v_{i,:}^k + \nu_k \big(u_1^{k+1} \big)_{i,:} + \big(\lambda_2^k \big)_{i,:} + \varrho_{i,:}^k \Big) \right\|_2^2. \quad (8)
$$

The third line of (6) is a classical L^2 Tikhonov regularization. The last three lines are simple gradient-ascent steps in the Lagrange multipliers. We briefly describe in section 2.4 the strategies to solve these subproblems.
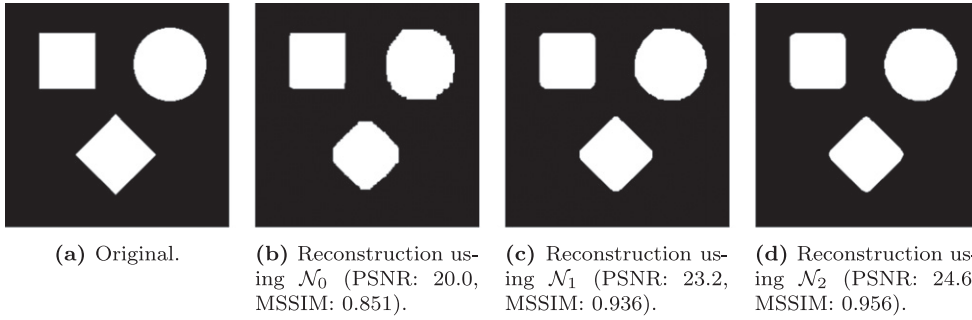
(a) Original.

(b) Reconstruction using $\mathcal{N}_0$ (PSNR: 20.0, MSSIM: 0.851).

(c) Reconstruction using $\mathcal{N}_1$ (PSNR: 23.2, MSSIM: 0.936).

(d) Reconstruction using $\mathcal{N}_2$ (PSNR: 24.6, MSSIM: 0.956).

Figure 2. Reconstruction of geometric shapes from Radon data with four angles and noise level 0.01 In case of the anisotropic discretization $\mathcal{N}_0$, jumps with respect to compass directions are significantly less penalized than jumps with respect to the diagonal directions (see figure 3). In consequence, horizontal and vertical edges are favored which results in a geometric staircasing effect. The neighborhood $\mathcal{N}_1$ improves the result significantly. The knight-move system $\mathcal{N}_2$ gives the most accurate reconstruction of the geometric shapes.

2.2. Design of isotropic discretizations

The anisotropic discretization of section 2.1 measures the length of the jump set in the anisotropic Manhattan metric [24]. This leads to geometric staircasing in the reconstructions illustrated in figure 2. The Euclidean length can be approximated better when complementing the neighborhood system with finite-difference vectors, for example, diagonal directions or 'knight-move' directions [25]. We now present a general scheme to construct appropriate neighborhood systems.

The starting point is the anisotropic neighborhood system

$$\mathcal{N}_0 = \{(1, 0), (0, 1)\}.$$

The vectors in this system have the (formal) slopes 0 and ∞. We add a new finite-difference vector $(x, y) \in \mathbb{Z}^2$ to the system only if its slope y/x is not yet contained in the system. For example, we can add the vector $(1, 1)$ with slope 1. For reasons of symmetry, we also add the orthogonal vector $(1, -1)$. Thus, we get the neighborhood system

$$\mathcal{N}_1 = \{(1, 0), (0, 1), (1, 1), (1, -1)\}. \tag{9}$$

The next vectors to include in the neighborhood system are the four knight move vectors $(\pm 2, 1), (1, \pm 2)$ which leads to the system

$$\mathcal{N}_2 = \{(1, 0), (0, 1), (1, 1), (1, -1), (-2, -1), (-2, 1), (2, 1), (2, -1)\}. \tag{10}$$

The general scheme of adding new vectors corresponds to the standard enumeration of the rational numbers.

Appropriate weights can be derived as follows. Let us assume that u is a binary $(n \times n)$ image with an edge along the direction $(x, y) \in \mathcal{N}$. We first look at lines with a slope y/x between (-1) and 1 going from the left to the right boundary of the image. (If the slope of (x, y) is not in the interval $[-1, 1]$ then we look at the $\pi/2$-rotated image and exchange the roles of x and y.) The Euclidean length of such a line is given by $n\sqrt{x^2 + y^2}/x$. Since we want that the total jump length of this image equals that Euclidean length, we get a condition on the weights that takes the form

$$\sum_{s=1}^{S} \omega_s \| \nabla_{p_s} u \|_0 = n \frac{\sqrt{x^2 + y^2}}{x}, \tag{11}$$

where $\| \nabla_{p_s} u \|_0$ is given by

$$\| \nabla_{p_s} u \|_0 = \left| \left\{ (i, j): u_{(i,j)+p_s} \neq u_{(i,j)} \right\} \right|.$$

It remains to evaluate the left-hand side of (11) for the binary image u. This can be done either manually for small neighborhood systems or with the help of a computer program for larger neighborhood systems. When counting the non-zero entries of $\nabla_{p_s} u$ we assume n to be large so that boundary effects are negligible. We end up with a system of S equations for the S unknowns. For the diagonal neighborhood system $\mathcal{N}_1$, (11) yields the conditions

$$\omega_1 + \omega_3 + \omega_4 = 1,$$
$$\omega_2 + \omega_3 + \omega_4 = 1,$$
$$\omega_1 + \omega_2 + 2\omega_3 = \sqrt{2},$$
$$\omega_1 + \omega_2 + 2\omega_4 = \sqrt{2}.$$

Solving this linear system, we get the weights

$$\omega_1 = \omega_2 = \sqrt{2} - 1 \quad \text{and} \quad \omega_3 = \omega_4 = 1 - \frac{\sqrt{2}}{2}.$$

For the knight-move neighborhood system $\mathcal{N}_2$, we get an analogous system of equations in $S = 8$ unknowns which gives us the weights

$$\omega_s = \begin{cases} \sqrt{5} - 2, & \text{for } s = 1, 2, \\ \sqrt{5} - \frac{3}{2}\sqrt{2}, & \text{for } s = 3, 4, \\ \frac{1}{2}\left(1 + \sqrt{2} - \sqrt{5}\right), & \text{for } s = 5,...,8. \end{cases}$$

We now turn to the question of how well we approximate the Euclidean length with the above discretizations. The neighborhood systems give rise to a norm $\| \cdot \|_{\mathcal{N}}$ defined for $p \in \mathbb{R}^2$ by

$$\| p \|_{\mathcal{N}} = \sum_{s=1}^{S} \omega_s \left| \langle p, p_s \rangle \right|.$$

By construction, the length $\| p_s \|_{\mathcal{N}}$ coincides with the Euclidean length $\| p_s \|_2$ for all vectors p_s in the neighborhood system as illustrated in figure 3. In [25], it is proposed to measure the isotropy of a finite-difference system by the ratio E between the longest and the shortest unit vector with respect to that length, compiled as

$$E = \max_{\|p\|_2=1} \| p \|_{\mathcal{N}} \Big/ \min_{\|p\|_2=1} \| p \|_{\mathcal{N}}.$$

The closer the quantity E is to one the higher is the isotropy. For the anisotropic system $\mathcal{N}_0$, we get the value $E_0 = \sqrt{2} \approx 1.41$. The introduction of diagonal directions reduces this value significantly to $E_1 \approx 1.08$. If we include also the knight-move differences ($\mathcal{N}_2$), then the ratio improves further to $E_2 \approx 1.03$. We note that the weights for the system $\mathcal{N}_1$ coincide with those proposed in [25], up to a normalization factor. For $\mathcal{N}_2$, our weights are more isotropic than the weights of [25].
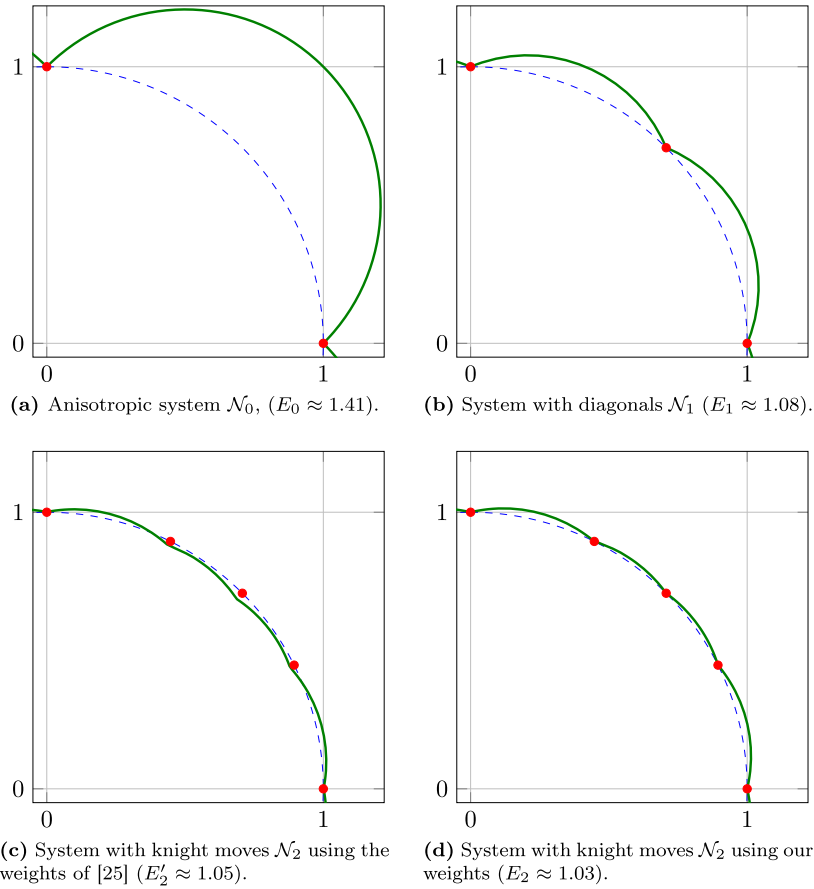
(a) Anisotropic system $\mathcal{N}_0$, ($E_0 \approx 1.41$).

(b) System with diagonals $\mathcal{N}_1$ ($E_1 \approx 1.08$).

(c) System with knight moves $\mathcal{N}_2$ using the weights of [25] ($E_2' \approx 1.05$).

(d) System with knight moves $\mathcal{N}_2$ using our weights ($E_2 \approx 1.03$).

Figure 3. The solid line represents the length of a Euclidean unit vector as measured in the finite-difference systems (as a function of the angle). The red dots identify the normalized vectors of the neighborhood system $p_s/\|p_s\|_2$ and the dashed line is the Euclidean unit circle. The isotropy increases significantly when passing from $\mathcal{N}_0$ to $\mathcal{N}_1$. The increase in isotropy is less substantial when passing from $\mathcal{N}_1$ to $\mathcal{N}_2$.

2.3. Splitting scheme for general discretizations

We now derive a minimization strategy for the general discretization (2). Let us denote the neighborhood system by $\mathcal{N} = \{p_1, ..., p_S\}$ and let $\omega_1, ..., \omega_S > 0$ where $S \geqslant 2$. We first rewrite (2) as the constrained optimization problem

$$
\begin{aligned}
\underset{u_1, ..., u_S, v}{\text{minimize}} \quad & \gamma \sum_{s=1}^{S} \omega_s \parallel \nabla_{p_s} u_s \parallel_0 + \parallel Av - f \parallel_2^2 \\
\text{subject to} \quad & u_r - u_t = 0, \text{ for all } 1 \leqslant r < t \leqslant S, \\
& v - u_s = 0, \text{ for all } 1 \leqslant s \leqslant S.
\end{aligned}
\tag{12}
$$

The augmented Lagrangian of this optimization problem reads

$$L(u, \lambda, \varrho) = \gamma \sum_{s=1}^{S} \omega_s \left\| \nabla_{P_s} u_s \right\|_0 + \frac{\mu}{2} \left\| v - u_s + \frac{\lambda_s}{\mu} \right\|_2^2$$

$$+ \frac{\nu}{2} \sum_{1 \leqslant r < t \leqslant S} \left\| u_r - u_t + \frac{\varrho_{r,t}}{\nu} \right\|_2^2 + \left\| Av - f \right\|_2^2, \tag{13}$$

where L depends on the variables $\{u_s\}_{1 \leqslant s \leqslant S}$, $\{\lambda_s\}_{1 \leqslant s \leqslant S}$ and $\{\varrho_{r,t}\}_{1 \leqslant r < t \leqslant S}$. The parameter $\nu > 0$ controls how strong the split variables $u_1, \dots, u_S$ are tied to each other and $\mu > 0$ controls their coupling to v. The variables $\lambda_s, \varrho_{r,t} \in \mathbb{R}^{n \times m}$ are Lagrange multipliers. In the ADMM iteration, we minimize L sequentially with respect to $v, u_1, \dots, u_S$ followed by a gradient ascent step in the Lagrange multipliers. The minimization of L with respect to u_s reads

$$\arg\min_{u_s} L(u, \lambda, \varrho) = \arg\min_{u_s} \gamma \omega_s \left\| \nabla_{P_s} u_s \right\|_0 + \frac{\mu}{2} \left\| v - u_s + \frac{\lambda_s}{\mu} \right\|_2^2$$

$$+ \frac{\nu}{2} \sum_{1 \leqslant r < s} \left\| u_r - u_s + \frac{\varrho_{r,s}}{\nu} \right\|_2^2 + \frac{\nu}{2} \sum_{s < t \leqslant S} \left\| u_s - u_t + \frac{\varrho_{s,t}}{\nu} \right\|_2^2. \tag{14}$$

We modify this expression using lemma 2.1 to

$$\arg\min_{u_s} L(u, \lambda, \varrho) = \arg\min_{u_s} \frac{2\gamma\omega_s}{\mu + \nu(S-1)} \left\| \nabla_{P_s} u_s \right\|_0 + \left\| u_s - w_s \right\|_2^2$$

with

$$w_s = \frac{\mu v + \lambda_s + \sum_{1 \leqslant r < s} \left(\nu u_r + \varrho_{r,s} \right) + \sum_{s < t \leqslant S} \left(\nu u_t - \varrho_{s,t} \right)}{\mu + \nu(S-1)}.$$

In a similar way, we consider the minimizer with respect to v as

$$\arg\min_{v} L(u, \lambda, \varrho) = \arg\min_{v} \left\| Av - f \right\|_2^2 + \sum_{s=1}^{S} \frac{\mu}{2} \left\| v - u_s + \frac{\lambda_s}{\mu} \right\|_2^2$$

which we rewrite using lemma 2.1 as

$$\arg\min_{v} L(u, \lambda, \varrho) = \arg\min_{v} \left\| Av - f \right\|_2^2 + \frac{\mu S}{2} \left\| v - \frac{1}{S} \sum_{s=1}^{S} \left(u_s - \frac{\lambda_s}{\mu} \right) \right\|_2^2.$$

Having computed explicit expressions for the minimization with respect to each variable, we obtain the ADMM iteration

$$
\begin{cases}
u_1^{k+1} \in \arg\ \min_{u_1} \dfrac{2\gamma\omega_1}{\mu_k + \nu_k(S-1)}\ \|\nabla_{p_1}u_1\|_0 + \|u_1 - w_1^k\|_2^2, \\[2mm]
\qquad\qquad \vdots \\[2mm]
u_S^{k+1} \in \arg\ \min_{u_S} \dfrac{2\gamma\omega_S}{\mu_k + \nu_k(S-1)}\ \|\nabla_{p_S}u_S\|_0 + \|u_S - w_S^k\|_2^2, \\[2mm]
\ v^{k+1} = \arg\ \min_v\ \|Av - f\|_2^2 + \dfrac{\mu_k S}{2}\ \|v - z^k\|_2^2, \\[2mm]
\lambda_s^{k+1} = \lambda_s^k + \mu_k\big(v^{k+1} - u_s^{k+1}\big), \quad \text{for all}\quad 1 \leqslant s \leqslant S, \\[2mm]
\varrho_{r,t}^{k+1} = \varrho_{r,t}^k + \nu_k\big(u_r^{k+1} - u_t^{k+1}\big), \quad \text{for all}\quad 1 \leqslant r < t \leqslant S.
\end{cases}
\tag{15}
$$

Here, w_s^k is given by

$$
w_s^k = \frac{\mu_k v^k + \lambda_s^k + \sum_{1\leqslant r < s}\big(\nu_k u_r^{k+1} + \varrho_{r,s}^k\big) + \sum_{s < t \leqslant S}\big(\nu_k u_t^k - \varrho_{s,t}^k\big)}{\mu_k + \nu_k(S-1)}
$$

and z^k by

$$
z^k = \frac{1}{S}\sum_{s=1}^S \left(u_s^{k+1} - \frac{\lambda_s^k}{\mu_k}\right).
$$

The key observation is that every of the subproblem in the ADMM iteration can be solved efficiently. The minimization problems with respect to $u_1,...,u_S$ decompose into univariate Potts problems with respect to the paths induced by the finite difference vectors p_s. To fix ideas, consider the finite-difference vector $p_s = (1, 1)$. Then, the solution u_s^{k+1} is given by solving one univariate Potts problem for each diagonal path of the two-dimensional array w_s^k. The last subproblem, like in the anisotropic case, is a classical Tikhonov-type regularization.

We eventually remark that the anisotropic splitting (6) is a special case of the general form (15) if we choose the anisotropic finite-difference system $\mathcal{N}_0$.

2.4. Solution of the subproblems

Solution methods for the univariate Potts and classical Tikhonov problems are well studied. Since they are the important building blocks of our iteration (15), we briefly recall the idea of the algorithms.

The classical univariate Potts problem is given by

$$
P_\gamma(g) = \gamma\ \|\nabla g\|_0 + \|g - f\|_2^2 \to \min,
\tag{16}
$$

where $g, f \in \mathbb{R}^n$ and $\|\nabla g\|_0 = |\{i:\ g_i \neq g_{i+1}\}|$ denotes the number of jumps of g. This can be solved exactly by dynamic programming [24, 41, 64, 65, 86, 92, 95]. The basic idea is that a minimizer of the Potts functional for data $(f_1,...,f_r)$ can be computed in polynomial time provided that minimizers of the partial data (f_1), (f_1, f_2), $\cdots$, $(f_1,...,f_{r-1})$ are known. We denote the respective minimizers by g^1, g^2, ..., g^{r-1}. In order to compute a minimizer for data $(f_1,...,f_r)$, we first create a set of r minimizer candidates h^1, ..., h^r, each of length r. These minimizer candidates are given by

$$h^\ell = (g^{\ell-1}, \underbrace{\mu_{[\ell,r]}, ..., \mu_{[\ell,r]}}_{\text{Length } (r-\ell+1)}), \tag{17}$$

where g^0 is the empty vector and $\mu_{[\ell,r]}$ denotes the mean value of data $f_{[\ell,r]} = (f_\ell, ..., f_r)$. Among the candidates h^ℓ, one with the least Potts functional value is a minimizer for the data $f_{[1,r]}$.

In [41], Friedrich *et al* proposed the following $\mathcal{O}(n^2)$ time and $\mathcal{O}(n)$ space algorithm. They observed that the functional values of a minimizer $P_\gamma(g^r)$ for data $f_{[1,r]}$ can be computed directly from the functional values $P_\gamma(g^1), ..., P_\gamma(g^{r-1})$ and the squared mean deviations of the data $f_{[1,r]}, ..., f_{[r,r]}$. Indeed, using (17), the functional value of the minimizer g^r is given (setting $P_\gamma(g^0) = -\gamma$) by

$$P_\gamma(g^r) = \min_{\ell=1,...,r} \left\{ P_\gamma(g^{\ell-1}) + \gamma + d_{[\ell,r]} \right\}, \tag{18}$$

where $d_{[\ell,r]}$ denotes the squared deviation from the mean value

$$d_{[\ell,r]} = \min_{y \in \mathbb{R}} \| y - f_{[\ell,r]} \|_2^2 = \left\| \left(\mu_{[\ell,r]}, ..., \mu_{[\ell,r]} \right) - f_{[\ell,r]} \right\|_2^2.$$

The evaluation of (18) is $\mathcal{O}(n)$ if we precompute the first and second moments of data $f_{[\ell,r]}$. If ℓ^* denotes the minimizing argument in (18), then $(\ell^* - 1)$ indicates the rightmost jump location at step r, which is stored as $J(r)$. The jump locations of a solution g^r are thus $J(r)$, $J(J(r))$, $J(J(J(r))), ...$; the values of g^r between two consecutive jumps are given by the mean of f on this interval. Note that we only have to compute and store the jump locations $J(r)$ and the minimal Potts functional value $P_\gamma(g^r)$ in each iteration. The reconstruction of the minimizer from the jump locations only has to be done once for g^n at the end; it is thus not time-critical.

The algorithm for solving (16) consists of two nested loops for $r = 1, ..., n$ and $\ell = 1, ..., r$, which amounts to $n(n+1)/2$ iterations in total. Typically, a significant amount of configurations are unreachable and thus can be skipped [86 theorem 2]. The time complexity is still $\mathcal{O}(n^2)$, but the practical runtime is improved by a fourfold to fivefold factor. We refer the reader to [86] for the complete flow diagram of the accelerated algorithm. We remark that a minimizer of the univariate Potts problem need not be unique, which explains the '∈' in (6), (7), (8), and (15). However, those data which lead a non-unique minimizer form a negligible set [96].

Our second subproblem is the solution of a classical L^2 Tikhonov regularization

$$v^{k+1} = \arg \min_v \| Av - f \|_2^2 + \frac{\mu}{2} \| v - z \|_2^2 \tag{19}$$

with some $z \in \mathbb{R}^{m \times n}$. The unique minimizer of this problem is given by the solution of the normal equation

$$\left(A^* A + \frac{\mu}{2} I \right) v = A^* f + \frac{\mu}{2} z. \tag{20}$$

Here, A^* is the adjoint of A. This linear system can be solved using, for example, the conjugate-gradient method. In some cases we can exploit the structure of A for more efficient solution methods. This is the case when A is the Radon transform or a convolution operator (see section 3).

2.5. General data terms

Inspecting the ADMM iteration, we observe that the data term only appears in the first line. That line consists of a classical Tikhonov regularization with the L^2 data term $\|Av - f\|_2^2$. Minimizers of that problem can be computed efficiently for many other data terms $d(u, f)$, such as L^p data terms, $p \geqslant 1$, of the form

$$d(v, f) = \|Av - f\|_p^p$$

or a Huber data term which is a hybrid between L^1 and L^2 data terms [27, 91]. In general, our algorithm is applicable whenever the proximity operator $\mathrm{prox}_{d(\cdot,f)/\mu}$ of $d(\cdot, f)$, defined by

$$\mathrm{prox}_{d(\cdot,f)/\mu}(z) = \arg\min_v d(v, f) + \frac{\mu}{2}\|v - z\|_2^2,$$

can be evaluated efficiently. This is often the case when $d(\cdot, f)$ is a convex functional. We refer to [12] for an extensive overview on strategies for convex optimization.

2.6. Convergence

In this section, we show that algorithm (15) converges in the prototypical case $\nu_k = 0$ for all k (which implies that the $\varrho^k = 0$ for all k.) We leave a convergence proof for algorithm (15) with general ν_k as an open problem. For the proof, we use methods developed in [87].

Theorem 2.2. *Let the sequence $(\mu_k)_{k \in \mathbb{N}}$ be increasing and satisfy $\sum_k \mu_k^{-1/2} < \infty$. Further, let $\nu_k = 0$ for all k. Then, the iteration (15) converges in the sense that*

$$\left(u_1^k, \ldots, u_S^k, v^k\right) \to \left(u_1^*, \ldots, u_S^*, v^*\right) \quad \text{with} \quad u_1^* = \ldots = u_S^* = v^*,$$

$$\frac{\lambda_s^k}{\mu_k} \to 0 \quad \text{for all} \quad s \in \{1, \ldots, S\}. \tag{21}$$

Proof. We denote the S functionals appearing in the first S lines of (15) by F_s^k, i.e.

$$F_s^k(u_s) = \frac{2\gamma\omega_s}{\mu_k}\|\nabla_{p_s} u_s\|_0 + \left\|u_s - \left(v^k + \frac{\lambda_s^k}{\mu_k}\right)\right\|_2^2.$$

Using this notation, we rewrite the first S lines of (15) as $u_s^{k+1} \in \arg\min_u F_s^k(u_s)$ for all $s \in \{1, \ldots, S\}$. We first estimate the distance $\left\|u_s^{k+1} - (v^k + \frac{\lambda_s^k}{\mu_k})\right\|_2$. To that end, we note that $F_s^k(u_s^{k+1}) \leqslant F_s^k\left(v^k + \frac{\lambda_s^k}{\mu_k}\right)$ which holds true since u_s^{k+1} minimizes F_s^k. Applying the definition of F_s^k, we get

$$\gamma\omega_s \|\nabla_{p_s} u_s^{k+1}\|_0 + \frac{\mu_k}{2}\left\|u_s^{k+1} - \left(v^k + \frac{\lambda_s^k}{\mu_k}\right)\right\|_2^2 \leqslant \gamma\omega_s \left\|\nabla_{p_s}\left(v^k + \frac{\lambda_s^k}{\mu_k}\right)\right\|_0 \leqslant \gamma\omega_s L,$$

where $L = NM$ is the size of the considered $(N \times M)$ image. This is because $\|\nabla_{p_s} z\|_0 \leqslant MN$ for any p_s and any data z. Since the first summand on the left-hand side is nonnegative, we get that

$$\left\| u_s^{k+1} - \left(v^k + \frac{\lambda_s^k}{\mu_k} \right) \right\|_2^2 \leqslant \frac{\gamma \omega_s L}{\mu_k}. \tag{22}$$

In particular, for all $s \in \{1, \ldots, S\}$, we obtain that

$$\lim_{k \to \infty} u_s^{k+1} - \left(v^k + \frac{\lambda_s^k}{\mu_k} \right) = 0. \tag{23}$$

We now draw our attention to the $(S + 1)$th line of (15). We denote the corresponding functional by G^k, i.e.

$$G^k(v) = \| Av - f \|_2^2 + \frac{\mu_k S}{2} \left\| v - \frac{1}{S} \sum_{s=1}^{S} \left(u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} \right) \right\|_2^2.$$

The minimality of v^{k+1} implies the inequality

$$G^k\left(v^{k+1} \right) \leqslant G^k \left(\frac{1}{S} \sum_{s=1}^{S} \left(u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} \right) \right).$$

We now apply the definition of G^k to estimate

$$\| Av^{k+1} - f \|_2^2 + \frac{\mu_k S}{2} \left\| v^{k+1} - \left(\frac{1}{S} \sum_{s=1}^{S} \left(u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} \right) \right) \right\|_2$$

$$\leqslant \left\| A \left(\frac{1}{S} \sum_{s=1}^{S} \left(u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} \right) \right) - f \right\|_2^2$$

$$\leqslant \left\| A \left(\frac{1}{S} \sum_{s=1}^{S} \left(u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} - v^k \right) \right) + Av^k - f \right\|_2^2$$

$$\leqslant \left(\| A \| \left\| \frac{1}{S} \sum_{s=1}^{S} \left(u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} - v^k \right) \right\|_2 + \| Av^k - f \|_2 \right)^2. \tag{24}$$

Here, $\| A \|$ is the operator norm of A acting on ℓ^2. We now combine (24) and (22) to estimate the magnitude of the residuals $Av^{k+1} - f$. We get

$$\| Av^{k+1} - f \| \leqslant \frac{C}{\sqrt{\mu_k}} + \| Av^k - f \|,$$

where $C > 0$ is a constant that only depends on γ, ω_s L, and $\| A \|$. Solving this recursion yields

$$\| Av^{k+1} - f \| \leqslant C \sum_{j=1}^{k} \frac{1}{\sqrt{\mu_j}} + \| Av^0 - f \|,$$

which shows that the sequence of residuals $(Av^{k+1} - f)_{k \in \mathbb{N}}$ is bounded.

We consider the right-hand term in the first line of (24). Then we apply (24) to get

$$\frac{\mu_k S}{2} \| v^{k+1} - \frac{1}{S} \sum_{s=1}^{S} \left(u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} \right) \|^2$$

$$\leqslant \left(\|A\| \| \frac{1}{S} \sum_{s=1}^{S} \left(u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} - v^k \right) \| + \|Av^k - f\| \right)^2$$

$$\leqslant \left(\|A\| \frac{1}{S} \sum_{s=1}^{S} \| \left(u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} - v^k \right) \| + C' \right)^2 .$$

The last inequality is a consequence of the boundedness of the residuals where we denote the bound by the positive constant C' (which is independent of k). We now apply (23) to the first summand to conclude that the sequence (with respect to k)

$$\mu_k \left\| v^{k+1} - \frac{1}{S} \sum_{s=1}^{S} \left(u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} \right) \right\|_2^2 \quad \text{is bounded.} \tag{25}$$

We use this fact to establish the convergence of the sequence v^k by showing that it is a Cauchy sequence. We first apply the triangle inequality to get

$$\| v^{k+1} - v^k \| \leqslant \| v^{k+1} - \frac{1}{S} \sum_{s=1}^{S} \left(u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} \right) \| + \| \frac{1}{S} \sum_{s=1}^{S} \left(u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} - v^k \right) \| .$$

We now apply (25) to the first summand on the right-hand side as well as (22) to the second summand on the right-hand side to obtain

$$\| v^{k+1} - v^k \| \leqslant \frac{C''}{\sqrt{\mu_k}}$$

for some constant $C'' > 0$ which is again independent of k. The assumption on the sequence μ_k guarantees that v^k is a Cauchy sequence and hence that v^k converges to some v^*.

To establish the last statement in (21), we rewrite each of the last S lines in (15) to obtain the identity

$$\frac{\lambda_s^{k+1}}{\mu_k} = \left(\frac{\lambda_s^k}{\mu_k} + u_s^{k+1} - v^k \right) + \left(v^k - v^{k+1} \right). \tag{26}$$

By (23) and (25), each term in parenthesis converges to 0. Hence

$$\lim_{k \to \infty} \frac{\lambda_s^{k+1}}{\mu_k} = 0.$$

Since we assume that the sequence μ_k is nondecreasing, we have that $\mu_k / \mu_{k+1} \leqslant 1$ and thus, for all $s = 1, \ldots, S$

$$\lim_{k \to \infty} \frac{\lambda_s^k}{\mu_k} = 0.$$

This shows the last statement in (21). Finally, we rewrite the penultimate line of (15) as $v^{k+1} - u_s^{k+1} = (\lambda_s^{k+1} - \lambda_s^k) / \mu_k$ to obtain the inequality

$$\| u_s^{k+1} - v^{k+1} \| \leqslant \frac{\| \lambda_s^{k+1} \|}{\mu_k} + \frac{\| \lambda_s^k \|}{\mu_k} \to 0.$$

This means that $u_s^k - v^k \to 0$ for all $s = 1, \ldots, S$ and, since v^k converges, also each u_s^k converges and the corresponding limit u_s^* equals v^*, which completes the proof. $\qquad\square$

3. Application to Radon data

The result of our method is a joint reconstruction and segmentation of the imaged object. More precisely, we obtain a piecewise constant image which induces a partition on the image domain. We demonstrate the applicability to tomographic problems whose image acquisition process can be described in terms of the classical Radon transform [66]. We here consider x-ray computed tomography (CT) and PET. Recall that the Radon transform is defined by

$$\mathcal{R}u(\theta, s) = \int_{-\infty}^{\infty} u\left(s\theta + t\theta^{\perp} \right) \mathrm{d}t, \tag{27}$$

where $s \in \mathbb{R}$, $\theta \in S^1$, and $\theta^{\perp}$ is the unit vector $\pi/2$ radians counterclockwise from θ.

Measurement of the segmentation quality. We will focus on data from piecewise-constant images. This has the advantage that the ground truth, i.e. the desired partitioning, is induced directly by the original image. (For natural images, there is typically no ground truth available because segmentation is based on subjective impressions.) Having a ground truth at hand, we can objectively measure the quality of the segmentation using the Rand index (RI) [3, 79], which we briefly explain. Let $X = \{x_1, \ldots, x_N\}$ be a given set of points and let Y and Y' be two partitionings of the this set. (In our case, X is the set of pixels, and Y and Y' are the partitioning given by the ground truth and the result of our method, respectively.) The RI is defined by

$$\mathrm{RI}(Y, Y') = \binom{N}{2} \sum_{i<j}^{N} t_{ij},$$

where t_{ij} is equal to one if there exist k and k' such that both x_i and x_j are in both Y_k and $Y'_{k'}$, or if x_i is in both Y_k and $Y'_{k'}$ while x_j is in neither Y_k or $Y'_{k'}$. The RI is bounded from above by 1; a higher value means a better match. For the evaluation of the RI we used the implementation of K Wang available at the Matlab File Exchange.

Parameter choice for the algorithm. Unless stated otherwise, the setup for the numerical experiments is as follows: We use the ADMM iteration (15) with the coupling sequence $\mu_k = 10^{-7} \cdot k^{\tau}$ with $\tau = 2.01$, and we choose ν_k identically zero. This choice satisfies the hypothesis of theorem 2.2. For the experiments involving the Shepp–Logan phantom, we observed the best results using the neighborhood system $\mathcal{N}_1$ of (9) while the neighborhood $\mathcal{N}_2$ of (10) gave better result for the more realistic images (figures 8, 10, and 11); see also the comparison in figure 2. For the solution of the Tikhonov type problem (19), we use Matlab's conjugate-gradient method on the normal equation (20). We use a 'warm start', which means that we use the solution of the previous iteration v^k as initial guess for v^{k+1}. Then the conjugate-gradient iteration converges typically in few steps. The splitting variables v, u_s for $s \geqslant 1$, and the Lagrange multipliers are all initialized with 0. We stop the iteration when the relative deviation of u_1 and u_2, i.e. $\| u_1 - u_2 \|_2 / (\| u_1 \|_2 + \| u_2 \|_2)$, falls below some threshold. In our experiments, we have chosen the tolerance 10^{-3}; we did not observe an improvement of the results for lower thresholds. The distance of u_1 and u_2 is a natural choice because these variables appear in all the neighborhood systems including the anisotropic
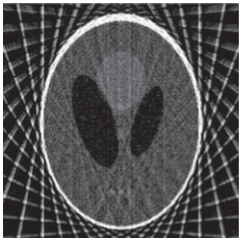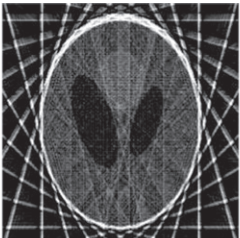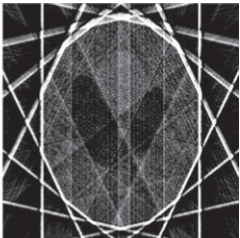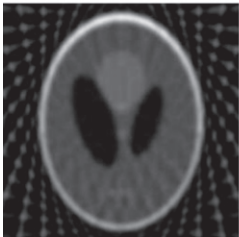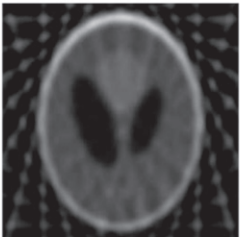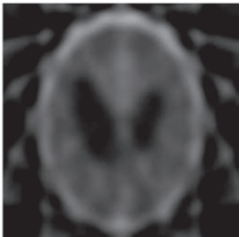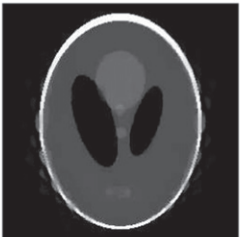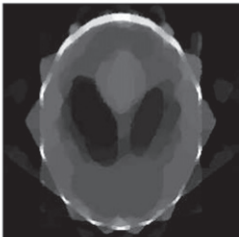
	17 Angles	12 Angles	7 Angles
FBP (Ram-Lak filter)			
	PSNR: 13.3 MSSIM: 0.180	PSNR: 11.1 MSSIM: 0.133	PSNR: 8.2 MSSIM: 0.090
FBP (Hamming window, optimized to PSNR)			
	PSNR: 17.7 MSSIM: 0.388	PSNR: 16.7 MSSIM: 0.355	PSNR: 15.6 MSSIM: 0.332
TV			
	$\alpha = 0.01$ PSNR: 39.3 MSSIM: 0.995	$\alpha = 0.01$ PSNR: 26.3 MSSIM: 0.898	$\alpha = 0.1$ PSNR: 17.2 MSSIM: 0.564
Our method			
	$\gamma = 0.03$ PSNR: 51.1 MSSIM: 0.999 RI: 0.999	$\gamma = 0.04$ PSNR: 46.7 MSSIM: 0.997 RI: 0.998	$\gamma = 0.03$ PSNR: 45.5 MSSIM: 0.997 RI: 0.998

Figure 4. Reconstruction of the Shepp–Logan phantom from highly undersampled Radon data. Filtered backprojection produces strong artifacts. Total variation regularization gives an almost perfect reconstruction up to about 17 projection angles but the quality decreases significantly for fewer angles. The proposed Potts based method yields a high quality reconstruction/segmentation from as few as seven projection angles.
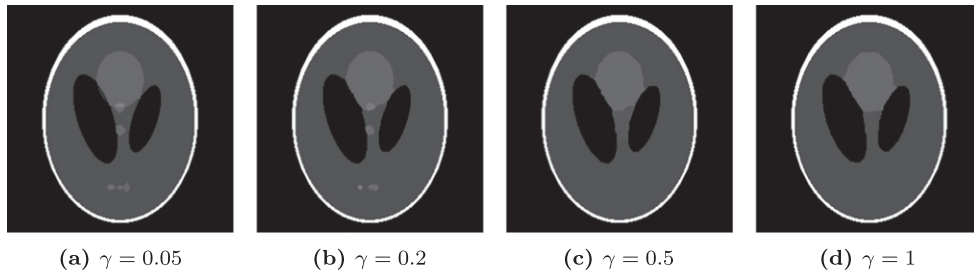
(a) $\gamma = 0.05$ **(b)** $\gamma = 0.2$ **(c)** $\gamma = 0.5$ **(d)** $\gamma = 1$

Figure 5. Influence of the model parameter γ on the result (12 noisefree projections of Radon data). For higher γ, the small segments vanish but the gross structures are recovered.

(a) Noise level 0.005,   **(b)** Noise level 0.01,   **(c)** Noise level 0.02, ($\gamma = 0.2$ RI: 0 990).   ($\gamma = 0.5$ RI: 0 957).   ($\gamma = 1.5$ RI: 0.909).
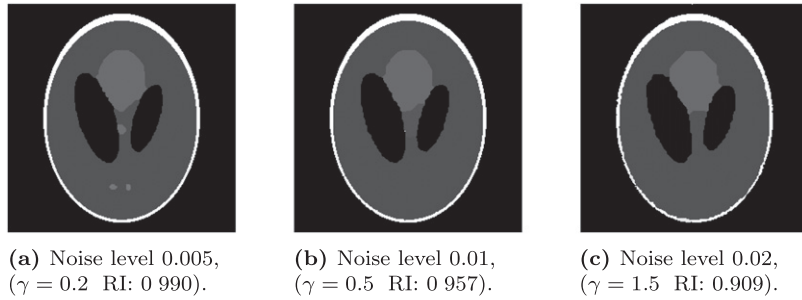
Figure 6. Effect of noise on the results (12 projection angles). The small structures vanish for higher noise levels whereas the large structures are segmented reliably.

system. For larger neighborhoods, other choices are possible, because all u_s converge to the same limit by theorem 2.2. Although the method is not independent from the ordering of the vectors in the neighborhood system, we observed no significant difference for other orderings. In our simulated experiments, the noise is Gaussian distributed with zero mean and standard deviation $\sigma = \text{noiselevel} \cdot \|f'\|_\infty$, where f' is the clean data.

3.1. Radon data with sparse angular sampling

We demonstrate the robustness of our method to drastic angular undersampling (figures 1, 4, 5, and 6). To set the results into context, we additionally show the results of filtered back-projection (FBP). FBP is the standard algorithm in many commercial CT scanners [72]. We use the Matlab function `iradon`, both with a standard Ram-Lak filter and with a Hamming window. The cutoff frequency for the Hamming window was tuned in steps of 0.01 with respect to PSNR. Recall that the PSNR is given by $\text{PSNR}(u) = 10 \log_{10}((mn \|g\|_\infty^2) / \|g - u\|_2^2)$ where $g \in \mathbb{R}^{m \times n}$ is the ground truth. For comparison, we further use the mean structural similarity index (MSSIM) [90]. We use Matlab's function `ssim` with standard parameters for the computation of the MSSIM. The MSSIM is bounded from above by 1. For both PSNR and MSSIM, higher values are better. We also compare with total variation regularization which uses the total variation $\alpha \| \nabla u \|_1$, $\alpha > 0$, as regularizing term. We follow the implementation of the Chambolle–Pock algorithm [27] provided by G. Peyre [73].

In figure 4, we observe that the classical reconstruction methods perform poorly when using only few projection angles. The standard FBP reconstruction produces streak artifacts which are typical for angular undersampling. The FBP reconstruction using optimized Hamming window smoothes out the edges. Total variation minimization achieves a high-quality reconstruction from 17 projections, but the quality decreases significantly for fewer angles. (Compare similar observations in [29] for MRI-type data.) In contrast, the proposed method achieves an almost perfect segmentation from as few as 7 projections.

In figure 1, we observe that the classical reconstruction combined with subsequent segmentation leads to poor results in a sparse angular setup. For the segmentation part in figure 1(c), we used the α-expansion graph-cut algorithm based on max-flow/min-cut of the library GCOptimization 3.0 of O Veksler and A Delong [13, 14, 57], which is a state-of-the-art image segmentation algorithm.

In figure 5, we illustrate the influence of the model parameter γ to the result. We observe that for higher γ smaller details vanish but the macrostructures are still recovered.

In figure 6, we add Gaussian noise to the observations. Due to the noise, the small structures vanish in the results. Nevertheless, the large scale structures are recovered reliably.

3.2. Radon data with dense angular sampling

The costs of evaluating $\mathcal{R}$ and $\mathcal{R}^*$ increases with the number of available projection angles. In total we need between 10^5 and 10^6 evaluations of $\mathcal{R}$ and $\mathcal{R}^*$. Therefore, using the conjugate gradients methods to solve the Tikhonov problem (15) can be time consuming for a dense angular sampling. In this setup, we can use the following efficient alternative implementation. Here, the minimizer of the Tikhonov problem is computed using a filtered-backprojection-type formula with a special filtering function which we describe next. Recall that the back-projection operator $\mathcal{R}^*$ is defined via

$$\mathcal{R}^* f(x) = \int_{S^1} f(\theta, x \cdot \theta) \mathrm{d}\theta.$$

Let $f = \mathcal{R}w$ for some $w \in L^2(\Omega)$ and $\alpha > 0$. Then, the solution v^* of the Tikhonov problem is given by

$$v^* = \arg\min_{v \in L^2(\Omega)} \| \mathcal{R}v - f \|_2^2 + \alpha \| v - z \|_2^2 = \mathcal{R}^* H_\alpha (f - \mathcal{R}z), \tag{28}$$

where the filtering operator is defined via

$$H_\alpha f = \mathcal{F}_s^{-1}\left(h_\alpha \mathcal{F}_s f \right)$$

with the filter function

$$h_\alpha(r) = \frac{|r|}{4\pi + \alpha |r|}.$$

Above, $\mathcal{F}_s$ (and $\mathcal{F}_s^{-1}$) denotes the one dimensional Fourier transform (and its inverse) of a function $f(\theta, s)$ with respect to the parameter s. Since we have not found the statement in this form in the literature, we provide a short proof in the appendix. We remark that this procedure can be applied only to densely sampled data. In a sparse angle setup, it produces unacceptably large errors.

In figure 7, we show the result of our method for the reconstruction of the Shepp–Logan phantom from Radon data with dense angular sampling (360 angles). For very high noise levels the small details vanish but the large geometric structures are still segmented reliably.
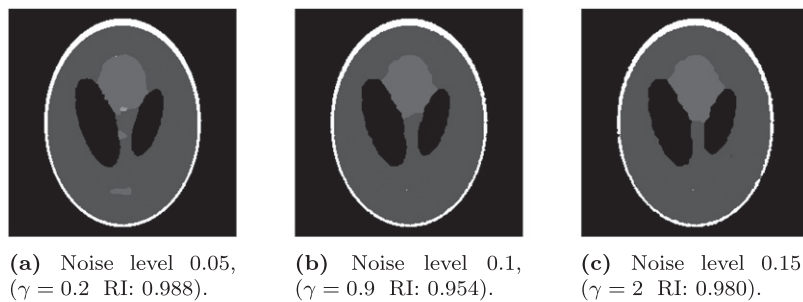
(a) Noise level 0.05, ($\gamma = 0.2$ RI: 0.988).

(b) Noise level 0.1, ($\gamma = 0.9$ RI: 0.954).

(c) Noise level 0.15, ($\gamma = 2$ RI: 0.980).

Figure 7. Reconstruction of the Shepp–Logan phantom. In the case of densely sampled Radon data (360 projection angles), our method is able to segment the large geometric structures for very high noise levels.
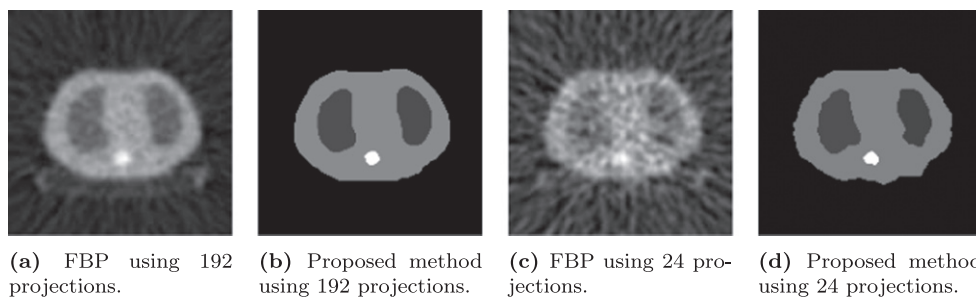

(a) FBP using 192 projections.

(b) Proposed method using 192 projections.

(c) FBP using 24 projections.

(d) Proposed method using 24 projections.

Figure 8. Joint reconstruction and segmentation from undersampled PET data of a physical thorax phantom [36]. Reproduced with permission from Fessler J *et al* 1997 *IEEE Trans. Med. Imaging* **16** 166–75. Copyright 1997 *IEEE*. The proposed method segments the anatomic structures (lung, spine, and thorax body). Filtered back-projection results are shown for comparison.

3.3. Real Radon data from a PET device

Next, we apply our method to PET data. The underlying PET model generates ideal data of the form $a \cdot \mathcal{R}g$ where $\mathcal{R}g$ is the Radon transform of the imaged object g and a is a known function depending on the attenuation. Eliminating the known function a, we are exactly in the setup of the classical Radon transform. In figure 8, we see the results for PET data of a physical thorax phantom [36]. Our method is able to jointly reconstruct and segment the anatomic structures (lung, spine, and thorax body) even from sparsely sampled data.

4. Further applications

We briefly discuss further applications. We consider spherical Radon data as well as blurred data.

4.1. Spherical Radon data

The reconstruction of a function $u: \mathbb{R}^2 \to \mathbb{R}$ from spherical averages plays an important role in the context of TAT/PAT. It has been intensively studied in recent years and still attracts much attention, cf [2, 20, 35, 38, 45, 59, 67, 89] to mention only a few. In the next example,
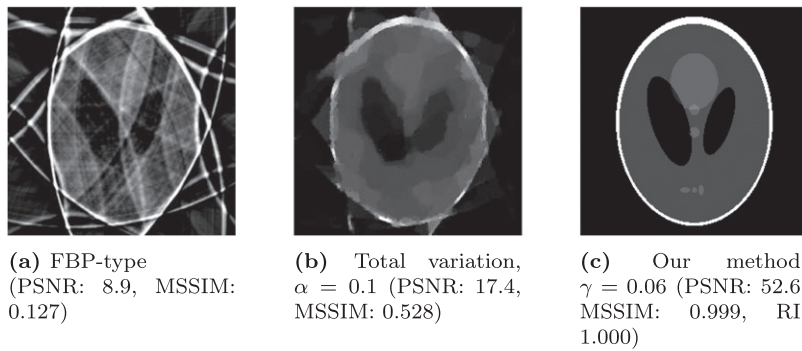
(a) FBP-type (PSNR: 8.9, MSSIM: 0.127)

(b) Total variation, $\alpha = 0.1$ (PSNR: 17.4, MSSIM: 0.528)

(c) Our method, $\gamma = 0.06$ (PSNR: 52.6, MSSIM: 0.999, RI: 1.000)

Figure 9. The proposed method reliably recovers all segments of the Shepp–Logan phantom from seven noisefree projections of spherical Radon data.

we apply our method to this problem. Here, we assume the data are given by the spherical mean Radon transform

$$\mathcal{M}u(\theta(\varphi), t) = \int_{S^1} u(\theta(\varphi) + t\zeta)\mathrm{d}\zeta, \tag{29}$$

for some angles $\varphi \in [0, 2\pi]$ and some radii $t \in (0, 2]$, where $\theta(\varphi) = (\cos\varphi, \sin\varphi)$. In our experiment, we computed the spherical means of the Shepp–Logan head phantom for seven equispaced angles and 512 equispaced radii. (See figure 1(a) for the original image.) The Tikhonov subproblem is solved using the standard conjugate gradient algorithm on the normal equation. For comparison, we show the result of FBP-type reconstruction algorithm as proposed in [2, 37] using R Seyfried's implementation of the algorithm. As can be observed in figure 9, the experimental results are similar to those of the classical Radon transform in section 3. The FBP-type reconstruction and the total variation reconstruction suffer from severe artifacts when only few data are available. In contrast, our method almost perfectly recovers the original image.

4.2. Blurred data

We finally demonstrate the applicability of our method to deblurring problems. Here, the operator A is a convolution operator. Hence, the normal equation (20) can be solved efficiently by fast Fourier transform techniques. In this experiment we particularly illustrate that our method can be applied to vector-valued data such as color images. We follow the splitting strategy as proposed in the present paper and extend the univariate Potts problem to vector-valued data according to [86]. It is worth mentioning that the computational effort grows only linearly in the dimension of the vectorial data. For example, the cost for processing a color image is about three times the cost of processing a gray-value image. In figure 10, we show the joint reconstruction and segmentation of an image blurred by a Gaussian kernel. (The image was taken from the Berkeley Segmentation Dataset [63]). In figure 11, we see the restoration of a traffic sign from simulated motion blur. Motion blur is modeled as a one-dimensional convolution along the direction $v \in \mathbb{R}^2$. Here, we use a moving average with respect to the horizontal direction. This experiment also illustrates that a positive coupling-parameter sequence ν_k can improve the result.

(a) Original (481 × 321). **(b)** Blurred and noisy data.

(c) Result using $\mathcal{N}_0$, **(d)** Result using $\mathcal{N}_1$. **(e)** Result using $\mathcal{N}_2$.
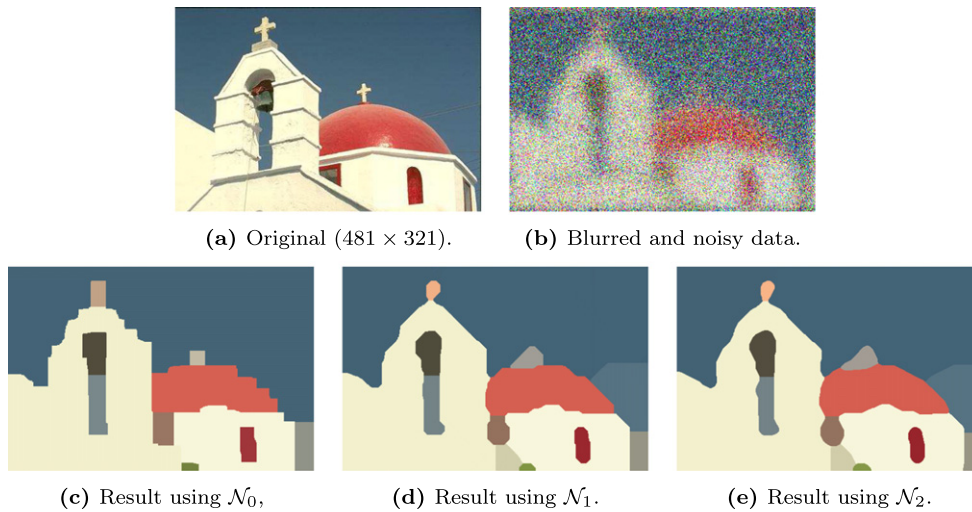
Figure 10. Joint reconstruction and segmentation of an image blurred by a Gaussian kernel of standard deviation ten and corrupted by extreme Gaussian noise of level 0.5. We used $\gamma = 0.5$. The higher the degree of isotropy becomes, the smoother are the segment boundaries. Reproduced with permission from Martin D *et al* 2001 *Proc. 8th Int. Conf. on Computer Vision* pp 416–23. Copyright 2001 Berkeley Segmentation Dataset.
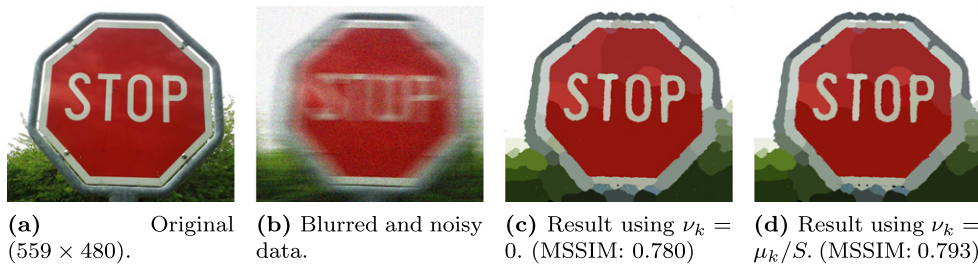

(a) Original (559 × 480). **(b)** Blurred and noisy data. **(c)** Result using $\nu_k = 0$. (MSSIM: 0.780) **(d)** Result using $\nu_k = \mu_k/S$. (MSSIM: 0.793)

Figure 11. Segmentation of an image from a simulated horizontal motion blur of 65 pixel length with Gaussian noise of level 0.1. The letters are recovered almost perfectly (c). The result improves if we use the algorithm of (15) with $\nu_k = \mu_k/S$ instead of $\nu_k = 0$ (d).

5. Conclusion

In this paper, we have developed a new splitting approach for the Potts model (or piecewise-constant Mumford–Shah model) for ill-posed inverse problems in imaging. We have presented a general discretization scheme which permits near-isotropic approximations of the length terms. This discretization allowed us to split the Potts problem into specific sub-problems that can be solved with efficient algorithms. We have demonstrated the capability of our method in various imaging applications. In particular, our algorithm has reconstructed all segments of the Shepp–Logan phantom from only seven projections of Radon and spherical Radon data, respectively. Further, we have obtained high-quality results from highly

incomplete and noisy data. Finally, we have demonstrated that it is applicable for joint reconstruction and segmentation of real tomographic data.

Appendix. Proof of equation (28)

Proof of equation (28). Setting $u = v - z$ and $g = f - \mathcal{R}z$ we first rewrite the cost functional as

$$\| \mathcal{R}v - f \|_2^2 + \alpha \| v - z \|_2^2 = \| \mathcal{R}u - g \|_2^2 + \alpha \| u \|_2^2. \tag{A.1}$$

In the following, we use the notation u_α to denote the minimizer of the right-hand side in (A.1). We first note that u_α satisfies the normal equation $\mathcal{R}^*\mathcal{R}u_\alpha + \alpha u_\alpha = \mathcal{R}^*g$ (see (20)). Further, we get from [66 theorem II.1.5] that $\mathcal{R}^*\mathcal{R}u_\alpha = 2(\| \cdot \|_2^{-1} * u_\alpha)$. Therefore, by taking the Fourier transform on both sides of the normal equation and applying the convolution theorem together with [46 Ch V, lemma 5.2], we obtain the following relation for the Fourier transform of u_α:

$$\hat{u}_\alpha(\xi) = \frac{\| \xi \|_2}{4\pi + \alpha \| \xi \|_2} \widehat{\mathcal{R}^*g}(\xi). \tag{A.2}$$

Formula (A.2) is a special case of a formula derived in [49 section 7]. Now, since $g = \mathcal{R}w'$ with $w' = w - z$, a similar argument shows that $\widehat{\mathcal{R}^*\mathcal{R}w'}(\xi) = 4\pi \| \xi \|_2^{-1} \widehat{w'}(\xi)$ and, therefore, that

$$\hat{u}_\alpha(\xi) = \frac{4\pi}{4\pi + \alpha \| \xi \|_2} \widehat{w'}(\xi). \tag{A.3}$$

Next, we use the Fourier-slice theorem [66 theorem II.1.1]. It states that $\widehat{\mathcal{R}_\theta w'}(r) = \sqrt{2\pi} \, \widehat{w'}(r\theta)$, where $\mathcal{R}_\theta w' = \mathcal{R}w'(\theta, \cdot)$. Using (A.3) we obtain

$$
\begin{aligned}
u_\alpha(x) &= \frac{1}{2\pi} \int_{\mathbb{R}^2} \widehat{u}_\alpha(\xi) e^{ix \cdot \xi} d\xi \\
&= \frac{1}{2\pi} \int_{\mathbb{R}^2} \frac{4\pi}{4\pi + \alpha \|\xi\|_2} \widehat{w'}(\xi) e^{ix \cdot \xi} d\xi \\
&= \int_{S^1} 2 \int_0^\infty \frac{1}{4\pi + \alpha |r|} \widehat{w'}(r\theta) e^{irx \cdot \theta} r \, dr \, d\theta \\
&= \int_{S^1} \int_{-\infty}^\infty \frac{|r|}{4\pi + \alpha |r|} \widehat{w'}(r\theta) e^{irx \cdot \theta} dr \, d\theta \\
&= \int_{S^1} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \frac{|r|}{4\pi + \alpha |r|} \widehat{\mathcal{R}_\theta w'}(r) e^{irx \cdot \theta} dr \, d\theta \\
&= \int_{S^1} H_\alpha \mathcal{R} w'(\theta, x \cdot \theta) \, d\theta \\
&= \mathcal{R}^* H_\alpha \mathcal{R} w'(x).
\end{aligned}
$$

Since $\mathcal{R}w' = \mathcal{R}w - \mathcal{R}z = f - \mathcal{R}z$, the assertion follows. $\qquad\square$

References

[1] Ambrosio L and Tortorelli V M 1990 Approximation of functional depending on jumps by elliptic functional via Γ-convergence *Commun. Pure Appl. Math.* **43** 999–1036

[2] Ansorg M, Filbir F, Madych W R and Seyfried R 2012 Summability kernels for circular and spherical mean data *Inverse Problems* **29** 015002

[3] Arbelaez P, Maire M, Fowlkes C and Malik J 2011 Contour detection and hierarchical image segmentation *IEEE Trans. Pattern Anal. Mach. Intell.* **33** 898–916

[4] Artina M, Fornasier M and Solombrino F 2013 Linearly constrained nonsmooth and nonconvex minimization *SIAM J. Optim.* **23** 1904–37

[5] Bar L, Sochen N and Kiryati N 2004 Variational pairing of image segmentation and blind restoration *ECCV 2004* (Berlin: Springer) pp 166–77

[6] Barkan O, Averbuch A, Dekel S and Tenzer Y 2014 A mathematical model for extremely low dose adaptive computed tomography acquisition *Mathematical Methods for Curves and Surfaces* (*Lecture Notes in Computer Science* vol 8177) ed T Lyche, M-L Mazure, K Morken and L Schumaker (Berlin: Springer) pp 13–33

[7] Beck A and Teboulle M 2009 A fast iterative shrinkage-thresholding algorithm for linear inverse problems *SIAM J. Imaging Sci.* **2** 183–202

[8] Blumensath T and Davies M 2008 Iterative thresholding for sparse approximations *J. Fourier Anal. Appl.* **14** 629–54

[9] Blumensath T and Davies M 2009 Iterative hard thresholding for compressed sensing *Appl. Comput. Harmon. Anal.* **27** 265–74

[10] Bostan E, Kamilov U, Nilchian M and Unser M 2013 Sparse stochastic processes and discretization of linear inverse problems *IEEE Trans. Image Process.* **22** 2699–710

[11] Boyd S, Parikh N, Chu E, Peleato B and Eckstein J 2011 Distributed optimization and statistical learning via the alternating direction method of multipliers *Found. Trends Mach. Learn.* **3** 1–122

[12] Boyd S and Vandenberghe L 2004 *Convex Optimization* (Cambridge: Cambridge University Press)

[13] Boykov Y and Kolmogorov V 2004 An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision *IEEE Trans. Pattern Anal. Mach. Intell.* **26** 1124–37

[14] Boykov Y, Veksler O and Zabih R 2001 Fast approximate energy minimization via graph cuts *IEEE Trans. Pattern Anal. Mach. Intell.* **23** 1222–39

[15] Boysen L, Bruns S and Munk A 2009 Jump estimation in inverse regression *Electron. J. Stat.* **3** 1322–59

[16] Boysen L, Kempe A, Liebscher V, Munk A and Wittich O 2009 Consistencies and rates of convergence of jump-penalized least squares estimators *Ann. Stat.* **37** 157–83

[17] Bredies K and Holler M 2014 Regularization of linear inverse problems with total generalized variation *J. Inverse Ill-posed Problems* at press

[18] Bredies K and Lorenz D A 2009 Regularization with non-convex separable constraints *Inverse Problems* **25** 085011

[19] Bronstein M, Bronstein A, Zibulevsky M and Azhari H 2002 Reconstruction in diffraction ultrasound tomography using nonuniform FFT *IEEE Trans. Med. Imaging* **21** 1395–401

[20] Buehler A, Rosenthal A, Jetzfellner T, Dima A, Razansky D and Ntziachristos V 2011 Model-based optoacoustic inversions with incomplete projection data *Med. Phys.* **38** 1694

[21] Cai X and Steidl G 2013 Multiclass segmentation by iterated ROF thresholding *Energy Minimization Methods in Computer Vision and Pattern Recognition* (Berlin: Springer) pp 237–50

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

[23] Candès E, Wakin M and Boyd S 2008 Enhancing sparsity by reweighted ℓ^1 minimization *J. Fourier Anal. Appl.* **14** 877–905

[24] Chambolle A 1995 Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations *SIAM J. Appl. Math.* **55** 827–63

[25] Chambolle A 1999 Finite-differences discretizations of the Mumford–Shah functional *ESAIM: Math. Modelling Numer. Anal.* **33** 261–88

[26] Chambolle A, Cremers D and Pock T 2012 A convex approach to minimal partitions *SIAM J. Imaging Sci.* **5** 1113–58

[27] Chambolle A and Pock T 2011 A first-order primal-dual algorithm for convex problems with applications to imaging *J. Math. Imaging Vis.* **40** 120–45

[28] Chan T and Vese L 2001 Active contours without edges *IEEE Trans. Image Process.* **10** 266–77

[29] Chartrand R 2007 Exact reconstruction of sparse signals via nonconvex minimization *IEEE Signal Process. Lett.* **14** 707–10

[30] Chartrand R 2009 Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data *IEEE Int. Symp. Biomedical Imaging: from Nano to Macro* pp 262–5

[31] Chartrand R, Sidky E and Pan X 2013 Nonconvex compressive sensing for x-ray CT: an algorithm comparison *IEEE Asilomar Conf. Signals, Systems, and Computers* pp 665–9

[32] Cremers D, Rousson M and Deriche R 2007 A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape *Int. J. Comput. Vis.* **72** 195–215

[33] Davison M E 1983 The ill-conditioned nature of the limited angle tomography problem *SIAM J. Appl. Math.* **43** 428–48

[34] Defrise M, Vanhove C and Liu X 2011 An algorithm for total variation regularization in high-dimensional linear problems *Inverse Problems* **27** 065002

[35] Elbau P, Scherzer O and Schulze R 2012 Reconstruction formulas for photoacoustic sectional imaging *Inverse Problems* **28** 045004

[36] Fessler J, Ficaro E, Clinthorne N and Lange K 1997 Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction *IEEE Trans. Med. Imaging* **16** 166–75

[37] Filbir F, Kunis S and Seyfried R 2014 Effective discretization of direct reconstruction schemes for photoacoustic imaging in spherical geometries

[38] Finch D, Patch S K and Rakesh 2004 Determining a function from its mean values over a family of spheres *SIAM J. Math. Anal.* **35** 1213

[39] Fornasier M, March R and Solombrino F 2013 Existence of minimizers of the Mumford–Shah functional with singular operators and unbounded data *Annali di Matematica Pura ed Applicata* **192** 361–91

[40] Fornasier M and Ward R 2010 Iterative thresholding meets free-discontinuity problems *Found. Comput. Math.* **10** 527–67

[41] Friedrich F, Kempe A, Liebscher V and Winkler G 2008 Complexity penalized M-estimation *J. Comput. Graph. Stat.* **17** 201–24

[42] Frikel J and Quinto E T 2013 Characterization and reduction of artifacts in limited angle tomography *Inverse Problems* **29** 125007

[43] Geman S and Geman D 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–41

[44] Hahn B 2014 Reconstruction of dynamic objects with affine deformations in computerized tomography *J. Inverse Ill-Posed Problems* **22** 323–39

[45] Haltmeier M 2014 Universal Inversion Formulas for Recovering a Function from Spherical Means *SIAM J. Math. Anal.* **46** 214–32

[46] Helgason S 1999 *The Radon Transform* (Boston: Birkhäuser)

[47] Hirschmüller H 2008 Stereo processing by semiglobal matching and mutual information *IEEE Trans. Pattern Anal. Mach. Intell.* **30** 328–41

[48] Jiang M, Maass P and Page T 2014 Regularizing properties of the Mumford–Shah functional for imaging applications *Inverse Problems* **30** 035007

[49] Jonas P and Louis A K 2001 A Sobolev space analysis of linear regularization methods for ill-posed problems *J. Inverse Ill-Posed Problems* **9** 59–74

[50] Katsevich A I 1997 Local tomography for the limited-angle problem *J. Math. Anal. Appl.* **213** 160–82

[51] Kim J, Tsai A, Cetin M and Willsky A S 2002 A curve evolution-based variational approach to simultaneous image restoration and segmentation *Proc. IEEE Int. Conf. Image Processing* **1** 1–109

[52] Klann E 2011 A Mumford–Shah-like method for limited data tomography with an application to electron tomography *SIAM J. Imaging Sci.* **4** 1029–48

[53] Klann E and Ramlau R 2013 Regularization properties of Mumford–Shah-type functionals with perimeter and norm constraints for linear ill-posed problems *SIAM J. Imaging Sci.* **6** 413–36

[54] Klann E, Ramlau R and Ring W 2011 A Mumford–Shah level-set approach for the inversion and segmentation of SPECT/CT data *Inverse Problems and Imaging* **5** 137–66

[55] Kohr H and Louis A K 2011 Fast and high-quality reconstruction in electron tomography based on an enhanced linear forward model *Inverse Problems* **27** 045008

[56] Kolehmainen V, Lassas M and Siltanen S 2008 Limited data x-ray tomography using nonlinear evolution equations *SIAM J. Sci. Comput.* **30** 1413–29

[57] Kolmogorov V and Zabih R 2004 What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Mach. Intell.* **26** 147–59

[58] Kreutzmann T and Rieder A 2014 Geometric reconstruction in bioluminescence tomography *Inverse Problems Imaging* **8** 173–97

[59] Kunyansky L A 2007 Explicit inversion formulae for the spherical mean Radon transform *Inverse Problems* **23** 373–83

[60] Lellmann J, Kappes J, Yuan J, Becker F and Schnörr C 2009 Convex multi-class image labeling by simplex-constrained total variation *Scale Space and Variational Methods in Computer Vision* (*Lect. Notes in Computer Science*) ed X-C Tai, K Mørken, M Lysaker and K-A Lie (Berlin: Springer) pp 150–62

[61] Louis A K 1984 The limited angle problem in computerized tomography *Technical Report* 5 (Kaiserslautern Technische Universität)

[62] Louis A K 1986 Incomplete data problems in x-ray computerized tomography. I. Singular value decomposition of the limited angle transform *Numer. Math.* **48** 251–62

[63] Martin D, Fowlkes C, Tal D and Malik J 2001 A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics *Proc. VIII Int. Conf. Computer Vision* pp 416–23

[64] Mumford D and Shah J 1985 Boundary detection by minimizing functionals *IEEE Conf. Comput. Vis. Pattern Recognit.* **17** 137–54

[65] Mumford D and Shah J 1989 Optimal approximations by piecewise smooth functions and associated variational problems *Commun. Pure Appl. Math.* **42** 577–685

[66] Natterer F 1986 *The Mathematics of Computerized Tomography* (Stuttgart: Teubner)

[67] Natterer F 2012 Photo-acoustic inversion in convex domains *Inverse Problems Imaging* **6** 315–20

[68] Needell D and Ward R 2013 Stable image reconstruction using total variation minimization *SIAM J. Imaging Sci.* **6** 1035–58

[69] Nikolova M, Ng M and Tam C-P 2010 Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction *IEEE Trans. Image Process.* **19** 3073–88

[70] Nikolova M, Ng M, Zhang S and Ching W 2008 Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization *SIAM J. Imaging Sci.* **1** 2–25

[71] Nocedal J and Wright S 2006 *Numerical Optimization* (New York: Springer)

[72] Pan X, Sidky E and Vannier M 2009 Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction? *Inverse Problems* **25** 123009

[73] Peyré G 2011 The numerical tours of signal processing-Advanced computational signal and image processing *IEEE Comput. Sci. Eng.* **13** 94–97

[74] Pock T, Chambolle A, Cremers D and Bischof H 2009 A convex relaxation approach for computing minimal partitions *IEEE Conf. Computer Vision and Pattern Recognition* pp 810–7

[75] Potts R 1952 Some generalized order-disorder transformations *Math. Proc. Camb. Phil. Soc.* **48** 106–9

[76] Quinto E T 1993 Singularities of the x-ray transform and limited data tomography in $\mathbb{R}^2$ and $\mathbb{R}^3$ *SIAM J. Math. Anal.* **24** 1215–25

[77] Ramlau R and Ring W 2007 A Mumford–Shah level-set approach for the inversion and segmentation of x-ray tomography data *J. Comput. Phys.* **221** 539–57

[78] Ramlau R and Ring W 2010 Regularization of ill-posed Mumford–Shah models with perimeter penalization *Inverse Problems* **26** 115001

[79] Rand W 1971 Objective criteria for the evaluation of clustering methods *J. Am. Stat. Assoc.* **66** 846–50

[80] Rosenthal A, Ntziachristos V and Razansky D 2011 Model-based optoacoustic inversion with arbitrary-shape detectors *Med. Phys.* **38** 4285

[81] Rudin L, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica D: Nonlinear Phenom.* **60** 259–68

[82] Sidky E, Chartrand R, Jorgensen J and Pan X 2013 Nonconvex optimization for improved exploitation of gradient sparsity in CT image reconstruction *12th Int. Meeting of Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine* (CA, USA: Lake Tahoe) pp 189–92

[83] Sidky E, Duchin Y, Pan X and Ullberg C 2011 A constrained, total-variation minimization algorithm for low-intensity x-ray CT *Med. Phys.* **38** 117–25

[84] Sidky E, Kao C-M and Pan X 2006 Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT *J. X-ray Sci. Technol.* **14** 119–39

[85] Sidky E and Pan X 2008 Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization *Phys. Med. Biol.* **53** 4777

[86] Storath M and Weinmann A 2014 Fast partitioning of vector-valued images *SIAM Journal on Imaging Sciences* **7** 1826–52

[87] Storath M, Weinmann A and Demaret L 2014 Jump-sparse and sparse recovery using Potts functionals *IEEE Trans. Signal Process.* **62** 3654–66

[88] Veksler O 1999 Efficient graph-based energy minimization methods in computer vision *PhD Thesis* Cornell University

[89] Wang L V 2009 *Photoacoustic Imaging and Spectroscopy (Optical Science and Engineering)* (Boca Raton, FL: Chemical Rubber Company)

[90] Wang Z, Bovik A, Sheikh H and Simoncelli E 2004 Image quality assessment: from error visibility to structural similarity *IEEE Trans. Image Process.* **13** 600–12

[91] Weinmann A, Demaret L and Storath M 2014 Total variation regularization for manifold-valued data *SIAM J. Imaging Sci.* at press (arXiv:1312.7710)

[92] Weinmann A, Storath M and Demaret L 2012 The L^1-Potts functional for robust jump-sparse reconstruction arXiv:1207.4642

[93] Wendler T *et al* 2010 First demonstration of 3D lymphatic mapping in breast cancer using freehand SPECT *Eur. Nucl. Med. Mol. Imaging* **37** 1452–61

[94] Winkler G 2003 *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction* (Berlin: Springer)

[95] Winkler G and Liebscher V 2002 Smoothers for discontinuous signals *J. Nonparametric Stat.* **14** 203–22

[96] Wittich O, Kempe A, Winkler G and Liebscher V 2008 Complexity penalized least squares estimators: Analytical results *Math. Nachr.* **281** 582–95

[97] Wolf P A, Jørgensen J S, Schmidt T G and Sidky E Y 2013 Few-view single photon emission computed tomography (SPECT) reconstruction based on a blurred piecewise constant object model *Phys. Med. Biol.* **58** 5629–52

[98] Yan M and Vese L A 2011 Expectation maximization and total variation-based model for computed tomography reconstruction from undersampled data *SPIE Medical Imaging* p 79612X