View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Apollo

arXiv:1807.01660v2 [math.NA] 4 Mar 2019

# Enhancing joint reconstruction and segmentation with non-convex Bregman iteration

**Veronica Corona[1], Martin Benning[2], Matthias J. Ehrhardt[3], Lynn F. Gladden[4], Richard Mair[5], Andi Reci[4], Andrew J. Sederman[4], Stefanie Reichelt[5], Carola-Bibiane Schönlieb[1]**

[1] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, United Kingdom
[2] School of Mathematical Sciences, Queen Mary University of London, United Kingdom
[3] Institute for Mathematical Innovation, University of Bath, United Kingdom
[4] Department of Chemical Engineering and Biotechnology, University of Cambridge, United Kingdom
[5]Cancer Research UK Cambridge Institute, University of Cambridge, United Kingdom

E-mail: vc324@cam.ac.uk

**Abstract.** All imaging modalities such as computed tomography (CT), emission tomography and magnetic resonance imaging (MRI) require a reconstruction approach to produce an image. A common image processing task for applications that utilise those modalities is image segmentation, typically performed posterior to the reconstruction. Recently, the idea of tackling both problems jointly has been proposed. We explore a new approach that combines reconstruction and segmentation in a unified framework. We derive a variational model that consists of a total variation regularised reconstruction from undersampled measurements and a Chan-Vese based segmentation. We extend the variational regularisation scheme to a Bregman iteration framework to improve the reconstruction and therefore the segmentation. We develop a novel alternating minimisation scheme that solves the non-convex optimisation problem with provable convergence guarantees. Our results for synthetic and real data show that both reconstruction and segmentation are improved compared to the classical sequential approach.

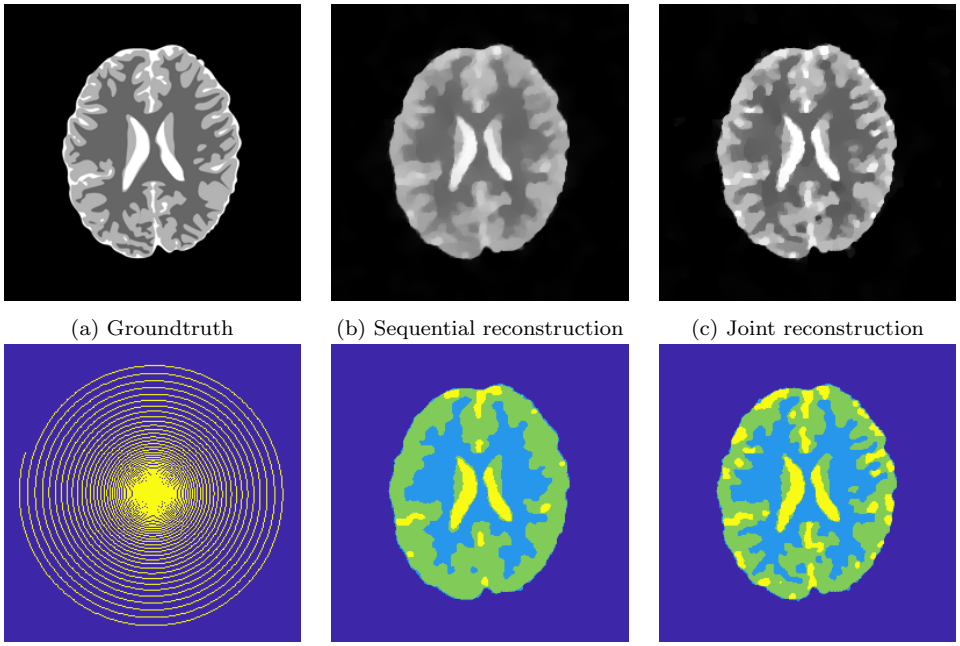
## 1. Introduction

Image reconstruction plays a central role in many imaging modalities for medical and non-medical applications. The majority of imaging techniques deal with incomplete data and noise, making the inverse problem of reconstruction severely ill-posed. Based on compressed sensing (CS) it is possible to tackle this problem by exploiting prior knowledge of the signal [1–3]. Nevertheless, reconstructions from very noisy and undersampled data will present some errors that will be propagated into further analysis,

e.g. image segmentation. Segmentation is an image processing task used to partition the image into meaningful regions. Its goal is to identify objects of interest, based on contours or similarities in the interior. Typically segmentation is performed after reconstruction, hence its result strongly depends on the quality of the reconstruction. Recently the idea of combining reconstruction and segmentation has become more popular. The main motivation is to avoid error propagations that occur in the sequential approach by estimating edges simultaneously from the data, ultimately improving the reconstruction. In this paper, we propose a new model for joint reconstruction and segmentation from undersampled MRI data. The underlying idea is to incorporate prior knowledge about the objects that we want to segment in the reconstruction step, thus introducing additional regularity in our solution. In this unified framework, we expect that the segmentation will also benefit from sharper reconstructions. We demonstrate that our joint approach improves the reconstruction quality and yields better segmentations compared to sequential approaches. In Figure 1, we consider a brain phantom from which we simulated the undersampled $k$-space data and added Gaussian noise. . Figure 1b and 1e present reconstructions and segmentations obtained with the sequential approaches, while Figure 1c and 1f show the results for our joint approach. The reconstruction using our method shows clearly more details and it is able to detect finer structures that are not recovered with the classical separate approach. As a consequence, the joint segmentation is also improved. In the following section we present the mathematical models that we used in our comparison. We investigated the performance of our model for two different applications: bubbly flow and cancer imaging. We show that both reconstruction and segmentation benefit from this method, compared to the traditional sequential approaches, suggesting that error propagation is reduced.

*Our contribution.* In our proposed joint method, we obtain an image reconstruction that preserves its intrinsic structures and edges, possibly enhancing them, thanks to the joint segmentation, and simultaneously we achieve an accurate segmentation. We consider the edge-preserving total variation regularisation for both the reconstruction and segmentation term using Bregman distances. In this unified Bregman iteration framework, we have the advantage of improving the reconstruction by reducing the contrast bias in the TV formulation, which leads to more accurate segmentation. In addition, the segmentation constitutes another prior for the reconstruction by enhancing edges of the regions of interest. Furthermore, we propose a non-convex alternating direction algorithm in a Bregman iteration scheme for which we prove global convergence.

The paper is organised as follows. In Section 2 we describe the problems of MRI reconstruction and region-based segmentation. We then introduce our joint reconstruction and segmentation approach in a Bregman iteration framework. This section also contains a detailed comparison of other joint models in the literature. In Section 3 we study the non-convex optimisation problem and present the convergence analysis for



(a) Groundtruth (b) Sequential reconstruction (c) Joint reconstruction

(d) Sampling matrix (e) Sequential segmentation (f) Joint segmentation

Figure 1: Sequential approach (left) versus unified approach (right). Combining reconstruction and segmentation in a single unified approach improves both the reconstructed image and its segmentation. See Figure 2 for more details.

this class of problems. Finally in Section 4 we present numerical results for MRI data for different applications. Here we investigate the robustness of our model by testing the undersampling rate up to its limit and by considering different noise levels.

## 2. MRI reconstruction and segmentation

In the following section we introduce the mathematical tools to perform image reconstruction and image segmentation. In this work, we focus on the specific MRI application; however, our proposed joint method can be applied to other imaging problems in which the measured data is connected to the image via a linear and bounded forward operator, cf. Subsection 2.1. Finally we present our model that combines the two tasks of reconstruction and segmentation in a unified framework.

### *2.1. Reconstruction*

In image reconstruction problems, we have the general setting

$$f = Au + \eta, \tag{1}$$

where $A : \mathbb{X} \to \mathbb{Y}$ is a bounded and linear operator mapping between two vector-spaces. The measured data $f \in \mathbb{Y}$ is usually corrupted by some noise $\eta$ and often only observed partially. In this formulation we are interested in recovering the image $u$ given the data $f$.

In this work, we focus on the application of MRI and we refer to the measurements $f$ as the $k$-space data. In standard MRI acquisitions, the Fourier coefficients are collected in the $k$-space by radio-frequency (RF) coils. Because the $k$-space data is acquired sequentially, the scanning time is constrained by physical limitations of the imaging system. One of the most common ways to perform fast imaging consists of undersampling the $k$-space; this, however, only yields satisfactory results if the dimension of the parameter space can implicitly be reduced, for example by exploiting sparsity in certain domains. In the reconstruction, this assumption is incorporated in the regularisation term. Let $\Omega := \{1, \dots, n_1\} \times \{1, \dots, n_2\}$ with $n_1, n_2 \in \mathbb{N}$ be a discrete image domain. Let $f = (f_i)_{i=1}^m \in \mathbb{C}^m$ with $m \ll n = n_1 n_2$ be our given undersampled $k$-space data, where $f_i \in \mathbb{C}$ are the measured Fourier coefficients that fulfil the relationship (1) with $A = SF$. The operator $A$ is now composed by $\mathcal{S} : \mathbb{C}^n \to \mathbb{C}^m$, which is a sampling operator that selects $m$ measurements from the $Fu$ data according to the locations provided by a binary sampling matrix (see e.g. Figure 1d), where $F$ is the discrete Fourier transform. In MRI, the noise $\eta$ is drawn from a complex-valued Gaussian distribution with zero mean and standard deviation $\sigma$ [4].

In problem (1) for MRI, the aim is to recover the image $u \in \mathbb{C}^n$ from the data. However, in this work we follow the standard assumption that in many applications we have negligible phase, i.e. we are working with real valued, non-negative images. Therefore, we are only interested in $u \in \mathbb{R}^n$; hence we consider the MRI forward operator as $A : \mathbb{R}^n \to \mathbb{C}^m$ and its adjoint $A^* : \mathbb{C}^m \to \mathbb{R}^n$ as modelled in [5]. Problem (1) is ill-posed due to noise and incomplete measurements. The easiest approach to approximate (1) is to compute the solution, for which the missing entries are replaced with zero

$$u_z = A^* f.$$

However, images reconstructed with this approach will suffer from aliasing artifacts because undersampling the $k$-space violates the Nyquist–Shannon sampling theorem. Therefore, we consider a mathematical model that incorporates prior knowledge by using a variational regularisation approach. A popular model is to find an approximate solution for $u$ as a minimiser of the Tikhonov-type regularisation approach

$$u^* \in \arg\min_u \left\{ \frac{1}{2} \|Au - f\|_2^2 + \alpha J(u) \right\}, \tag{2}$$

where the first term is the data fidelity that forces the reconstruction to be close to the measurements and the second term is the regularisation, which imposes some regularity on the solution. The parameter $\alpha > 0$ is a regularisation parameter that balances the two terms in the variational scheme. In this setting, different regularisation functionals $J$ can be chosen (see [6] for a survey of variational regularisation approaches).

Although problems of the form (2) are very effective, they also lead to a systematic loss of contrast [7–9]. This is typically observed for common choices of the regulariser $J$, i.e., convex functional. To overcome this problem, [10] proposed an iterative regularisation method based on the generalised Bregman distance [11,12]. The Bregman distance with

respect to $J$ is defined as

$$D_J^{p^k}(u, u^k) = J(u) - J(u^k) - \langle p^k, u - u^k \rangle \tag{3}$$

with $p^k \in \partial J(u^k)$, where $\partial J(u^k)$ is called sub-differential and it is a generalisation of the classical differential for convex functions. We replace problem (2) with a sequence of minimisation problems

$$u^{k+1} \in \arg\min_u \left\{ \frac{1}{2}\|Au - f\|_2^2 + \alpha D_J^{p^k}(u, u^k) \right\}. \tag{4}$$

The update on the subgradient can be conveniently computed by the optimality condition of (4)

$$p^{k+1} = p^k - \frac{1}{\alpha} A^*(Au^{k+1} - f). \tag{5}$$

In this work, we will focus on one particular choice for $J$, namely the *total variation*. The total variation (TV) regularisation is a well-known edge-preserving approach, first introduced by Rudin, Osher and Fatemi in [13] for image denoising. The TV regularisation, i.e., the 1-norm penalty on a discrete finite difference approximation of the two-dimensional gradient $\nabla : \mathbb{R}^n \to (\mathbb{R}^2)^n$, that is $\nabla u(i,j) = (\nabla_1 u(i,j), \nabla_2 u(i,j))^T$, is in the discrete setting

$$J(u) = \mathrm{TV}(u) = \|\nabla u\|_{2,1} = \sum_{(i,j)\in\Omega} \sqrt{|\nabla_1 u(i,j)|^2 + |\nabla_2 u(i,j)|^2}, \tag{6}$$

for the isotropic case.

We then consider the Bregman iteration scheme in (4) for $J(u) = \mathrm{TV}(u)$. This approach is usually carried on by initialising the regularisation parameter $\alpha$ with a large value, producing overregularised initial solutions. At every step $k$, finer details are added. A suitable criterion to stop iterations (4) and (5) (see [6]), is the Morozov's discrepancy principle [14]. The discrepancy principle suggests to choose the smallest $k \in \mathbb{N}$ such that $u^{k+1}$ satisfies

$$\|f - Au^{k+1}\|_2 \leq \sigma\sqrt{m} \tag{7}$$

where $m$ is the number of samples and $\sigma$ is the standard deviation of the noise in the data. Note that using Bregman iterations, the contrast is improved and in some cases even recovered exactly, compared to the variational regularisation model. In addition, it makes the regularisation parameter choice less challenging. Note that for different choices of $J$ in (2), e.g., the Mumford-Shah/Potts model [15–19], we do not have loss of contrast, but we deal with a non-convex NP hard problem, algorithmically more challenging.

### *2.2. Segmentation*

Image segmentation refers to the process of automatically dividing the image into meaningful regions. Mathematically, one is interested in finding a partition $\{\Omega_i\}_{i=1}^l$

of the image domain $\Omega$ subject to $\cup_{i=1}^{l}\Omega_i = \Omega$ and $\cap_{i=1}^{l}\Omega_i = \emptyset$. One way to do this is to use *region-based* segmentation models, which identify regions based on similarities of their pixels. The segmentation model we are considering was originally proposed by Chan and Vese in [20] and it is a particular case of the piecewise-constant Mumford-Shah model [15]. Given an image function $u : \Omega \to \mathbb{R}$, the goal is to divide the image domain $\Omega$ in two separated regions $\Omega_1$ and $\Omega_2 = \Omega \setminus \Omega_1$ by minimising the following energy function

$$\int_{\Omega_1} (u(x) - c_1)^2 \, dx + \int_{\Omega_2} (u(x) - c_2)^2 \, dx + \beta \cdot Length(C) \to \min_{c_1, c_2, C}$$

where $C$ is the desired contour separating $\Omega_1$ and $\Omega_2$, and the constants $c_1$ and $c_2$ represents the average intensity value of $u$ inside $C$ and outside $C$, respectively. The parameter $\beta$ penalises the length of the contour $C$, controlling the scale of the objects in the segmentation. From this formulation we can make two observations: first, the regions $\Omega_1$ and $\Omega \setminus \Omega_1$ can be represented by the characteristic function

$$v(x) = \begin{cases} 0, & \text{if } x \in \Omega_1 \cup C \\ 1, & \text{if } x \in \Omega_2, \end{cases}$$

second, the perimeter of the contour identified by the the characteristic function corresponds to its total variation, as shown by the Coarea formula [21]. This leads to the new formulation

$$\int_{\Omega} v(x)(u(x) - c_1)^2 dx + (1 - v(x))(u(x) - c_2)^2 \, dx + \beta \, \mathrm{TV}(v) \to \min_{c_1, c_2, v \in \{0,1\}} .$$

Even assuming fixed constants $c_1$, $c_2$ the problem is non-convex due to the binary constraint. In [22] the authors proposed to relax the constraint, allowing $v(x)$ to assume values in the interval $[0, 1]$. They showed that for fixed constants $c_1$, $c_2$, global minimisers can be obtained by minimising the following energy

$$\int_{\Omega} v(x)(u(x) - c_1)^2 dx + (1 - v(x))(u(x) - c_2)^2 \, dx + \beta \, \mathrm{TV}(v) \to \min_{v \in [0,1]} \tag{8}$$

followed by thresholding, setting $\Sigma = \{x : v(x) \geq \mu\}$ for a.e. $\mu \in [0, 1]$. As the problem is convex but not strictly convex, the global minimiser may not be unique. In practice we obtain solutions which are almost binary, hence the choice of $\mu$ is not crucial.
Setting

$$s(x) = (u(x) - c_1)^2 - (u(x) - c_2)^2$$

the energy (8) can be written in a more general form as

$$\int_{\Omega} v(x) s(x) \, dx + \beta \, \mathrm{TV}(v) \to \min_{v \in [0,1]} .$$

In this paper, we are interested in the extension of the two-class problem to the multi-class formulation [23]. Following the simplex-constrained vector function representation

for multiple regions and its convex relaxation proposed in [24], we obtain as a special case a convex relaxation of the Chan-Vese model for arbitrary number of regions, which reads

$$\int_\Omega \sum_{i=1}^{\ell} v_i(x)(c_i - u(x))^2 \, dx + \beta \, \mathrm{TV}(v) \to \min_{v \in \mathcal{C}}, \tag{9}$$

where $\mathcal{C} := \{v : \Omega \to \mathbb{R}^\ell \,|\, v(x) \geq 0, \sum_{i=1}^{\ell} v_i(x) = 1\}$ is a convex set which restricts $v(x)$ to lie in the standard probability simplex. As in the binary case, the constants $c_i$ describe the average intensity value inside region $i$. In this case we consider the vector-valued formulation of TV

$$\mathrm{TV}(v) = \int_\Omega \sqrt{\|\nabla v_1\|^2 + \cdots + \|\nabla v_\ell\|^2} \, dx.$$

*2.3. Joint reconstruction and segmentation*

MRI reconstructions from highly undersampled data are subject to errors, even when prior knowledge about the underlying object is incorporated in the mathematical model. It is often required to find a trade-off between filtering out the noise and retrieving the intrinsic structures while preserving the intensity configuration and small details. As a consequence, segmentations in the presence of artifacts are likely to fail.

In this paper, we propose to solve the two image processing tasks of reconstruction and segmentation in a unified framework. The underlying idea is to inform the reconstruction with prior knowledge of the regions of interest, and simultaneously update this belief according to the actual measurements. Mathematically, given the undersampled and noisy $k$-space data $f$, we want to recover the image $u \colon \Omega \to \mathbb{R}$ and compute its segmentation $v$ in $\ell$ disjoint regions, by solving the following problem

$$\begin{aligned}(u, v) = \arg\min_{u,v} \; & \underbrace{\frac{1}{2}\|Au - f\|_2^2 + \alpha \, \mathrm{TV}(u)}_{\text{reconstruction}} \\ & \underbrace{+ \delta \sum_{i=1}^{n} \sum_{j=1}^{\ell} v_{ij}(c_j - u_i)^2 + \beta \, \mathrm{TV}(v) + \imath_{\mathcal{C}}(v)}_{\text{segmentation}} .\end{aligned} \tag{10}$$

where $\imath_{\mathcal{C}}(v)$ is the characteristic function over $\mathcal{C} := \{v : \mathbb{R}^n \to \mathbb{R}^\ell \,|\, v_{ij} \geq 0, \sum_{j=1}^{\ell} v_{ij} = 1, \forall i \in \{1, \ldots, n\}\}$, and $\alpha$, $\beta$, $\delta > 0$ are some regularisation parameters. However, instead of solving (10), we consider the iterative regularisation procedure using Bregman distances. The main motivation is to exploit the contrast enhancement aspect for the reconstruction thanks to the Bregman iterative scheme. By improving the reconstruction, the segmentation is in turn refined. Therefore, we replace (10) with

the following sequence of minimisation problems for $k = 0, 1, 2, \ldots$

$$u^{k+1} = \arg\min_u \frac{1}{2}\|Au - f\|_2^2 + \alpha D_{\mathrm{TV}}^{p^k}(u, u^k) + \delta \sum_{i=1}^{n}\sum_{j=1}^{\ell} v_{ij}(c_j - u_i)^2 \tag{11a}$$

$$p^{k+1} = p^k - \frac{1}{\alpha}\left(A^*(Au^{k+1} - f) - 2\delta \sum_{j=1}^{\ell} v_j^k (u^{k+1} - c_j)\right) \tag{11b}$$

$$v^{k+1} \in \arg\min_v \delta \sum_{i=1}^{n}\sum_{j=1}^{\ell} v_{ij}(c_j - u_i^{k+1})^2 + \chi_{\mathcal{C}}(v) + \beta D_{\mathrm{TV}}^{q^k}(v, v^k) \tag{11c}$$

$$q^{k+1} = q^k - \frac{\delta}{\beta}(c_j - u^{k+1})^2. \tag{11d}$$

Note that (11) solves a problem different from (10). Assuming that a minimiser exists, the model (11) converges to a minimiser of

$$\frac{1}{2}\|Au - f\|_2^2 + \delta \sum_{i=1}^{n}\sum_{j=1}^{\ell} v_{ij}(c_j - u_i^{k+1})^2\,,$$

as we will show in Subsection 3.1. In case of noisy data $f$ this is not desirable, so that we combine the iteration with a stopping criterion in order to form a regularisation method.

This model combines the reconstruction approach described in (4) and the discretised multi-class segmentation in (9) with a variation in the regularisation term, which is now embedded in the Bregman iteration scheme. In [25] the authors used Bregman distances for the Chan-Vese formulation (8), combined with spectral analysis, to produce multiscale segmentations.

As described in the previous subsection, the parameters $\alpha$ and $\beta$ describe the scale of the details in $u$ and the scale of the segmented objects in $v$. By integrating the two regularisations into the same Bregman iteration framework, we obtain that these scales are now determined by the iteration $k+1$. At the first Bregman iteration $k = 0$, when $\alpha$ is very large, we obtain an over-smoothed $u^1$, and the value of $\beta$ is not very important. Intuitively, $u^1$ is almost piecewise constant with small total variation and a broad range of values of $\beta$ may lead to very similar segmentations $v^1$. However, at every iteration $k+1$, finer scales are added to the solution with the update $p^{k+1}$. Accordingly, with the update $q^{k+1}$, which is independent of $v^{k+1}$, the segmentation keeps up with the scale in the reconstructed image $u^{k+1}$.

The novelty of this approach is also represented by the role of the parameter $\delta > 0$. This parameter weighs the effect of the segmentation in the reconstruction, imposing regularity in $u$ in terms of sharp edges in the regions of interest. In Section 4 we show how different ranges of $\delta$ affects the reconstruction (see Figure 12). Intuitively, large values of $\delta$ force the solution $u$ to be close to the piecewise constant solution described by the constants $c_i$. This is beneficial in applications where MRI is a means to extract shapes and sizes of underlying objects, (e.g. bubbly flow in Subsection 4.1). On the

other hand, with very small $\delta$, the segmentation has little impact and the solutions for $u$ are close to the ones obtained by solving the individual problem (4). Instead, intermediate values of $\delta$ impose sharper boundaries in the reconstruction while preserving the texture.

Obviously, we need to stop the iteration before the residual brings back noise from the data $f$. As we cannot use Morozov discrepancy principle in this case (due to the fact that $\|Au^k - f\|_2$ will rather increase due to the effect of the coupling term controlled by the parameter $\delta$), we stop when two consecutive iterates in $v$ are smaller than a certain tolerance, $\|v^{k+1} - v^k\| < tol$, following the observation that the rate at which $u^{k+1}$ changes close to the optimal solution is low, in contrary to more abrupt changes at the beginning of the Bregman iteration and later on when it starts to add noise.

Clearly, problem (11) is non-convex in the joint argument $(u, v)$ due to the coupling term. However, it is convex in each individual variable. We propose to solve the joint problem by iteratively alternating the minimisation with respect to $u$ and to $v$ (see Section 3 for numerical optimisation and convergence analysis).

### *2.4. Comparison to other joint reconstruction and segmentation approaches*

In this section we will provide an overview of some existing simultaneous reconstruction and segmentation (SRS) approaches with respect to different imaging applications.

*CT/SPECT.* Ramlau and Ring [26] first proposed a simultaneous reconstruction and segmentation model for CT, that was later extended to SPECT in [27] and to limited data tomography [28]. In these work, the authors aim to simultaneously reconstruct and segment the data acquired from SPECT and CT. CT measures the *mass density distribution* $\mu$, that represents the attenuation of x-rays through the material; SPECT measures the *activity distribution* $f$ as the concentration of the radio tracer injected in the material. Given the two measurements $z^\delta$ and $y^\delta$, from CT and SPECT, they consider the following energy functional

$$E(f, \mu, \Gamma^f, \Gamma^\mu) = \|A(f, \mu) - y^\delta\|^2 + \beta\|R\mu - z^\delta\|^2 + \alpha(Length(\Gamma^f) + Length(\Gamma^\mu)).$$

They propose a joint model based on a Mumford-Shah like functional, in which the reconstructions of $\mu$ and $f$ and the given data are embedded in the data term in a least squares sense. The operators $A$ and $R$ are the attenuated Radon transform (SPECT operator) and the Radon transform (CT operator), respectively. The penalty term is considered to be a multiple of the lengths of the contours of $\mu$, $\Gamma^\mu$ and the contours of $f$, $\Gamma^f$. These boundaries are modelled using level set functions. In these segmented partitions of the domain, $\mu$ and $f$ are assumed to be piecewise constant. The optimisation problem is then solved alternatively with respect to the functional variables $f$ and $\mu$ with fixed geometric variables $\Gamma^\mu$ and $\Gamma^f$ and the other way around.

In [29] the simultaneous reconstruction and segmentation is applied to dynamic SPECT imaging, which solves a variational framework consisting of a Kullback-Leibler (KL) data fidelity and different regulariser terms to enforce sharp edges and sparsity for the segmentation and smoothness for the reconstruction. The cost function is

$$E(u,c) = KL\big(R(u \cdot c), g\big) + \alpha \sum_{k=1}^{K} \|\nabla u_k\| + \beta \sum_{k=1}^{K} \|u_k\|_1 + \frac{\delta}{2} \sum_{k=1}^{K} \|\frac{\partial}{\partial t} c_k\|_2^2.$$

Given the data $g$, they want to retreive the concentration curves $c_k(t)$ in time for K disjoint regions and their indication functions $u_k(x)$ in space. The optimisation is carried out alternating the minimisation over $u$ having $c$ fixed and then over $c$ having $u$ fixed.

In [30] they propose a variational approach for reconstruction and segmentation of CT images, with limited field of view and occluded geometry. The cost function

$$E(u,c,v) = \frac{1}{2}\|Ax - y\|^2 + \alpha\|\nabla u\| + \frac{\beta}{2}\left(\lambda \sum_{i}^{n} \sum_{k=1}^{K} v_{ik}(u_i - c_k)^2 + \frac{1}{2}\|Dv\|_2^2\right)$$

s.t. a box constraint on the image values $x$ and the simplex constraint on the labelling function $v$. The operator $A$ is the undersampled Radon transform modelling the occluded geometry and $y$ is the given data. The second term is the edge-preserving regularisation term for $u$, the third term is the segmentation term which aims at finding regions in $u$ that are close to the value $c_k$ in region $k$. The operator $D$ is the finite difference approximation of the gradient. The non-convex problem is solved by alternating minimisation between updates of $u, v, c$.

*PET and Transmission Tomography.* In [31], the authors propose a maximum likelihood reconstruction and doubly stochastic segmentation for emission and transmission tomography. In their model they use a Hidden Markov Measure Field Model (HMMFM) to estimate the different classes of objects from the given data $r$. They want to maximise the following cost function

$$E(u,p,\theta) = \log P(r|u) + \log P(u|p,\theta) + \log P(p).$$

The first term is the data likelihood which will be modelled differently for emission and transmission tomography. The second term is the conditional probability or class fitting term, for which they use HMMFM. The third term is the regularisation on the HMMFM. The optimisation is carried out in three steps, where first they solve for $u$ (image update) fixing $p, \theta$, then for $p$, holding $u, \theta$ (measure field update) and finally for $\theta$ (parameter update) having $u, p$ fixed.

A variant of this method has been presented in [32], in which they incorporate

prior information about the segmentation classes through a HMMFM. Here, the reconstruction is the minimisation over a constrained Bayesian formulation that involves a *data fidelity term* as a classical least squares fitting term, a *class fitting term* as a Gaussian mixture for each pixel given $K$ classes and dependent of the class probabilities defined by the HMMFM, and a *regulariser* also dependent of the class probabilities. The model to minimise is

$$E(u,\delta) = \lambda_{noise}\|Au-b\|_2^2 - \sum_{j=1}^{N}\log\left(\sum_{k=1}^{K}\frac{\delta_{jk}}{\sqrt{2\pi}\sigma_k}\exp\left(-\frac{(u_j-\mu_k)^2}{2\sigma_k^2}\right)\right) + \lambda_{class}\sum_{k=1}^{K}R(\delta_k)$$

$$\text{s.t.} \quad \sum_{k=1}^{K}\delta_{jk} = 1, \quad \delta_{jk} \geq 0, \quad j = 1,\dots,N, \quad k = 1,\dots,K.$$

The operator $A$ will be modelled as the Radon transform in case of CT and $b$ represents the measured data; $N$ is the number of pixel in the image; $\lambda_{noise}$ and $\lambda_{class}$ are the regularisation parameters; $\mu_k, \sigma_k$ are the class parameters. The cost function is non convex and they solve the problem in an alternating scheme where they either update the pixel values or the class probabilities for each pixel.

Storath and others [33] model the joint reconstruction and segmentation using the Potts Model with application to PET imaging and CT. They consider the variational formulation of the Potts model for the reconstruction. Since the solution is piecewise constant, this directly induces a partition of the image domain, thus a segmentation. Given the data $f$ and an operator $A$ (e.g. Radon transform), the energy functional is in the following form

$$E(u) = \lambda\|\nabla u\|_0 + \|Au-f\|_2^2$$

where the first term is the jump penalty enforcing piecewise constant solutions and the second term is the data fidelity. As the Potts model is NP hard, they propose a discretisation scheme that allows to split the Potts problem into subproblems that can be solved efficiently and exactly.

*MRI.* In [34], the authors proposed a joint model with application to MRI. Their reconstruction-segmentation model consists of a fitting term and a patch-based dictionary to sparsely represent the image, and a term that models the segmentation as a mixture of Gaussian distributions with mean, standard deviation and mixture weights $\mu$, $\sigma$, $\pi$. Their model is

$$E(u,\Gamma,\mu,\sigma,\pi) = \|Au-y\|^2 + \lambda\sum_{n=1}^{N}\|R_n u - D\gamma_n\|^2 - \beta \ln P(u|\mu,\sigma,\pi) \quad \text{s.t. } \|\gamma_n\|_0 \leq T \ \ \forall n$$

where $A$ is the undersampled Fourier transform, $y$ is the given data, $R_n$ is a patch extraction operator, $\lambda$ is a weighting parameter, $T$ is the sparsity threshold, and $\gamma_n$ is the sparse representation of patch $R_n u$ organised as column $n$ of the matrix $\Gamma$. The problem is highly non-convex and it is solved iteratively using conjugate gradient on $u$, orthogonal matching pursuit on $\Gamma$ and Expectation-Maximisation algorithm on $(\mu,\sigma,\pi)$.

*Summary.* Recently, the idea to solve the problems of reconstruction and segmentation simultaneously has become more popular. The majority of these joint methods have been proposed for CT, SPECT and PET data. Mainly they differ in the way they encode prior information in terms of regularisers and how they link the reconstruction and segmentation in the coupling term. Some imposes smoothness in the reconstruction [29], others sparsity in the gradient [26, 30, 33], other consider a patch-dictionary sparsifying approach [34]. In [33] they do not explicitly obtain a segmentation, but they force the reconstruction to be piecewise constant. Depending on the application, the coupling term is the data fitting term itself (e.g. SPECT), or the segmentation term. In [31,32,34] the authors model the segmentation as a mixture of Gaussian distribution, while [30] has a a region-based segmentation approach similar to what we propose. However, [30] penalises the squared 2-norm of segmentation, imposing spatial smoothness.
In our proposed joint approach, we perform reconstruction and segmentation in a unified Bregman iteration scheme, exploiting the advantage of improving the reconstruction, which results in a more accurate segmentation. Furthermore, the segmentation constitutes another prior imposing regularity in the reconstruction in terms of sharp edges in the regions of interest. We propose a novel numerical optimisation problem in a non-convex Bregman iteration framework for which we present a rigorous convergence result in the following section.

## 3. Optimisation

The cost function (11) is non-convex in the joint argument $(u, v)$, but it is convex in each individual variable. To solve this problem we derive a splitting approach where we solve the two minimisation problems in an alternating fashion with respect to $u$ and $v$. We present the general algorithm and its convergence analysis in the next subsection. First, we describe the solution of each subproblem.

*Problem in $u$.* The problem in $u$ reads

$$u^{k+1} = \arg\min_u \frac{1}{2}\|Au - f\|_2^2 + \alpha(\mathrm{TV}(u) - \langle p^k, u\rangle) + \delta \sum_{i=1}^{n}\sum_{j=1}^{\ell} v_{ij}^k (c_j - u_i)^2.$$

We solve the optimisation for $u$, fixing $v$, using the primal-dual algorithm proposed in [35–38]. We write $F(u) = \|u\|_1$, $K(u) = \nabla u$ and $G(u) = \frac{1}{2}\|Au - f\|_2^2 - \alpha\langle p^k, u\rangle + \delta \sum_{i=1}^{n}\sum_{j=1}^{\ell} v_{ij}^k (c_j - u_i)^2$ and obtain the following iterates for $\theta = 1$ and step sizes $\sigma = \tau = 0.99/\|\nabla\|$

$$\begin{aligned}
y^{n+1} &= \frac{y^n + \sigma\nabla\bar{u}^n}{\max(1, \|y^n + \sigma\nabla\bar{u}^n\|)} \\
u^{n+1} &= \frac{u^n + \tau\nabla\cdot y^{n+1} + 2\tau\delta\sum_{j=1}^{\ell} v_j^k c_j + \tau\alpha p^k + \tau A^* f}{1 + 2\tau\delta + \tau A^* A} \\
\bar{u}^{n+1} &= 2u^{n+1} - u^n.
\end{aligned}$$

After sufficiently many iterations we set $u^{k+1} = u^{n+1}$ and compute the update $p^{k+1}$ from the optimality condition of (3) as (11b).

*Problem in $v$.* The problem in $v$ reads

$$v^{k+1} = \arg\min_{v \in C} \langle v, \delta g - \beta q^k \rangle + \beta \, \mathrm{TV}(v)$$

with $g = \big((c_1 - u^{k+1})^2, \dots, (c_\ell - u^{k+1})^2\big)^T$. We now solve a variant of the primal-dual method [35] as suggested in [38, 39]. They consider the general problem including pointwise linear terms of the form

$$\min_{x \in C} \max_{y \in B} \langle Kx, y \rangle + \langle g, x \rangle - \langle h, y \rangle$$

where $C \subseteq X$, $B \subseteq Y$ are closed, convex sets.
Setting $K = \nabla$ and $h = 0$, $\theta = 1$ and step sizes $\sigma = \tau = 0.99/\|\nabla\|$, the updates are

$$\begin{aligned} w^{n+1} &= \Pi_B\big(w^n + \sigma(\nabla \bar{v}^n - h)\big) \\ v^{n+1} &= \Pi_C\big(v^n + \tau \nabla \cdot (v^{n+1} - \delta g + \beta q^k)\big) \\ \bar{v}^{n+1} &= 2v^{n+1} - v^n. \end{aligned}$$

At the end, we set $v^{k+1} = v^{n+1}$ and obtain the update $q^{k+1}$ as (11d).

### *3.1. Convergence analysis*

The proposed joint approach (11) is an optimisation problem of the form

$$\min_{u,v} E(u, v) + D^{p^k}_{J_1}(u, u^k) + D^{q^k}_{J_2}(v, v^k) \tag{12}$$

in the general Bregman distance framework for (nonconvex) functions $E : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$, for $k \in \{0, \dots, N\}$ and some positive parameters $\alpha$ and $\beta$. The functions $J_1 : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ and $J_2 : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ impose some regularity in the solution. In this work we consider a finite dimensional setting and we refer to the next section for the required definitons. To prove global convergence of (12), we consider functions that satisfy the Kurdika-Łojasiewicz property, defined below, and we make the following assumptions.

**Definition 1** *(Kurdyka-Łojasiewicz (KL) property).* Let $F : \mathbb{R}^d \to \mathbb{R}$ be a proper and lower semicontinuous function.

- Then the function $F$ is said to have the KL property at $\bar{u} \in \mathrm{dom}(\partial F) := \{u \in \mathbb{R}^d | \partial F \neq \emptyset\}$ if there exists a constant $\eta \in (0, \infty]$, a neighbourhood $N$ of $\bar{u}$ and a concave function $\varphi : [0, \eta) \to \mathbb{R}_{>0}$ that is continuous at 0 and satisfies $\varphi(0) = 0$, $\varphi \in C^1(]0, \eta[)$, and $\varphi'(s) > 0$ for all $s \in ]0, \eta[$, such that for all $u \in N \cap \{u \in \mathbb{R}^d | F(\bar{u}) < F(u) < F(\bar{u}) + \eta\}$ the inequality

$$\varphi'(F(u) - F(\bar{u}))\mathrm{dist}(0, \partial F(u)) \geq 1 \tag{KL}$$

  holds.

- If $F$ satisfies the KL property at each point of $\mathrm{dom}(\partial F)$, $F$ is called a KL function.

**Lemma 1** The function $E(u,v) = \frac{1}{2}\|Au - f\|_2^2 + \delta \sum_{i=1}^{n} \sum_{j=1}^{\ell} v_{ij}(c_j - u_i)^2$ in our joint problem (11) satisfies the KL property over $\mathbb{R}^n \times \mathbb{R}^m$.
*Proof.* It has been proved in [40] that real-analytic functions satisfy the KL property. The function $E(u,v)$ is polynomial and therefore it is a real-analytic function. ■

**Assumption 1**

(i) $E$ is a $C^1$ function
(ii) $E > -\infty$
(iii) $E$ is a KL function
(iv) $J_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2$, are proper, lower semi-continuous (l.s.c.) and strongly convex
(v) $J_i$, $i = 1, 2$, are KL function
(vi) for any fixed $v$, the function $u \to E(u,v)$ is convex. Likewise for any fixed $u$, the function $v \to E(u,v)$ is convex.
(vii) for any fixed $v$, the function $u \to E(u,v)$ is $C^1_{L_1(v)}$, hence the partial gradient is $L_1(v)$-Lipschitz continuous

$$\|\nabla_u E(u_1, v) - \nabla_u E(u_2, v)\| \leq L_1(v)\|u_1 - u_2\| \qquad \forall u_1, u_2 \in \mathbb{R}^n.$$

Likewise for any fixed $u$, the function $v \to E(u,v)$ is $C^1_{L_2(u)}$.

We want to study the convergence properties of the alternating scheme

$$u^{k+1} = \arg\min_u \left\{ E(u, v^k) + D_{J_1}^{p^k}(u, u^k) \right\} \tag{13a}$$

$$p^{k+1} = p^k - \nabla_u E(u^{k+1}, v^k) \tag{13b}$$

$$v^{k+1} = \arg\min_v \left\{ E(u^{k+1}, v) + D_{J_2}^{q^k}(v, v^k) \right\} \tag{13c}$$

$$q^{k+1} = q^k - \nabla_v E(u^{k+1}, v^{k+1}) \tag{13d}$$

for initial values $(u^0, v^0)$, $p^0 \in \partial J_1(u^0)$ and $q^0 \in \partial J_2(v^0)$.
We want to show that the whole sequence generated by (13) converges to a critical point of $E$.

In order for the updates (13a) and (13c) to exist, we want $J$ to be of the form $J = R + \varepsilon G$ (e.g. $R = \|\nabla u\|_1$ and $G = \|u\|_2^2$, see [41]) where $R$ and $G$ fulfil the following assumptions. In practice, we verify that $G$ does not significantly change the reconstruction and segmentation performance for the examples we consider in the next section, for sufficiently small parameter (e.g. $\varepsilon = 10^{-3}$). Therefore, in our model (11) and in the numerical results we omit it.

**Algorithm 1** Alternating splitting method with Bregman iterations for two blocks.

**Initialization:** $(u^0, v^0)$, $p^0 \in \partial J_1(u^0)$, $q^0 \in \partial J_2(v^0)$, $N \in \mathbb{N}$
**for** $k = 0, 1, \ldots, N$ **do**
$u^{k+1} = \arg\min_u \left\{E(u, v^k) + D_{J_1}^{p^k}(u, u^k)\right\}$
$p^{k+1} = p^k - \nabla_u E(u^{k+1}, v^k)$
$v^{k+1} = \arg\min_v \left\{E(u^{k+1}, v) + D_{J_2}^{q^k}(v, v^k)\right\}$
$q^{k+1} = q^k - \nabla_v E(u^{k+1}, v^{k+1})$
**end for**

**Assumption 2**

(i) The functions $G_1 : \mathbb{R}^n \to \mathbb{R}$ and $G_2 : \mathbb{R}^m \to \mathbb{R}$ are strongly convex with constants $\gamma_1$ and $\gamma_2$, respectively. They have Lipschitz continuous gradient $\nabla G_1$ and $\nabla G_2$ with Lipschitz constant $\delta_1$ and $\delta_2$, respectively.

(ii) The functions $R_1 : \mathbb{R}^n \to \mathbb{R}$ and $R_2 : \mathbb{R}^m \to \mathbb{R}$ are proper, l.s.c. and convex.

For $J_i = \alpha_i R_i + \varepsilon_i G_i$, $i \in \{1, 2\}$, we can write (13) as

$$u^{k+1} = \arg\min_u \left\{E(u, v^k) + \alpha_1 D_{R_1}^{p^k}(u, u^k) + \varepsilon_1 D_{G_1}(u, u^k)\right\} \tag{14a}$$

$$p^{k+1} = p^k - \frac{1}{\alpha_1}\Big(\nabla_u E(u^{k+1}, v^k) + \varepsilon_1\big(\nabla G_1(u^{k+1}) - \nabla G_1(u^k)\big)\Big) \tag{14b}$$

$$v^{k+1} = \arg\min_v \left\{E(u^{k+1}, v) + \alpha_2 D_{R_2}^{q^k}(v, v^k) + \varepsilon_2 D_{G_2}(v, v^k)\right\} \tag{14c}$$

$$q^{k+1} = q^k - \frac{1}{\alpha_2}\Big(\nabla_v E(u^{k+1}, v^{k+1}) + \varepsilon_2\big(\nabla G_2(v^{k+1}) - \nabla G_2(v^k)\big)\Big). \tag{14d}$$

**Theorem 1** *(Global convergence).* Suppose $E$ is a KL function for any $z^k = (u^k, v^k) \in \mathbb{R}^n \times \mathbb{R}^m$ and $r^k = (p^k, q^k)$ with $p^k \in \partial R_1(u^k)$, $q^k \in \partial R_2(v^k)$. Assume Assumptions 1 and 2 hold. Let $\{z^k\}_{k\in\mathbb{N}}$ and $\{r^k\}_{k\in\mathbb{N}}$ be sequences generated by (14), which are assumed to be bounded. Then

(i) The sequence $\{z^k\}_{k\in\mathbb{N}}$ has finite length, that is

$$\sum_{k=0}^{\infty} \|z^{k+1} - z^k\| < \infty. \tag{15}$$

(ii) The sequence $\{z^k\}_{k\in\mathbb{N}}$ converges to a critical point $\bar{z}$ of $E$.

*3.2. Proof of Theorem 1*

In the following we are going to show global convergence of this algorithm. The first step in our convergence analysis is to show a sufficient decrease property of a surrogate of the energy function (12) and a subgradient bound of the norm of the iterates gap. We first recall the following definitions.

**Definition 2** *(Convex Conjugate).* Let $G$ be a proper, l.s.c. and convex function. Then its convex conjugate $G^* : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is defined as

$$G^*(p) := \sup_{u \in \mathbb{R}^n} \{\langle u, p\rangle - G(u)\},$$

for all $p \in \mathbb{R}^n$.

**Lemma 2** Let $G$ be a proper, l.s.c. and convex function and $G^*$ its convex conjugate. Then for all arguments $u \in \mathbb{R}^n$ with corresponding subgradients $p \in \partial G(u)$ we know

- $\langle u, p\rangle = G(u) + G^*(p)$,
- $p \in \partial G(u)$ is equivalent to $u \in \partial G^*(p)$.

From Lemma 2 we can rewrite the Bregman distance in (3) as follows

$$D_J^{p^k}(u, u^k) = J(u) + J^*(p^k) - \langle u, p^k\rangle, \tag{16}$$

where we can see that now it does not depend on $u^k$ anymore, but it can be defined as a function of $u$ and $p^k$ only, $D_J(u, p^k)$.

**Definition 3** *(Strong convexity).* Let $G$ be a proper, l.s.c. and convex function. Then $G$ is said to be $\gamma$-strongly convex if there exists a constant $\gamma$ such that

$$D_G^p(u, v) \geq \frac{\gamma}{2}\|u - v\|^2$$

holds true for all $u, v \in \mathrm{dom}(G)$ and $q \in \partial G(v)$.

**Definition 4** *(Symmetric Bregman distance).* Let $G$ be a proper, l.s.c. and convex function. Then the symmetric generalised Bregman distance $D_G^{symm}(u, v)$ is defined as

$$D_G^{symm}(u, v) := D_G^p(u, v) + D_G^q(v, u) = \langle p - q, u - v\rangle$$

for $u, v \in \mathrm{dom}(G)$ with $p \in \partial G(u)$ and $q \in \partial G(v)$. We also observe that in case $G$ is $\gamma$-strongly convex we have

$$D_G^{symm}(u, v) \geq \gamma\|u - v\|^2.$$

**Definition 5** *(Lipschitz continuity).* A function $G : \mathbb{R}^n \to \mathbb{R}$ is (globally) Lipschitz-continuous if there exists a constant $L > 0$ such that

$$\|G(u) - G(v)\| \leq L\|u - v\|$$

is satisfied for all $u, v \in \mathbb{R}^n$.

Before we show global convergence, we first define the surrogate functions.

**Definition 6** *(Surrogate objective).* Let $E, R_i, G_i, i \in \{1,2\}$ satisfy Assumption 1 and Assumption 2, respectively. For any $(u^k, v^k) \in \mathbb{R}^n \times \mathbb{R}^m$ and subgradients $p^k \in \partial R_1(u^k)$ and $q^k \in \partial R_2(v^k)$, we define the following *surrogate objectives* $F$, $F_1$ and $F_2$

$$\begin{aligned} F(u^{k+1}, v^{k+1}, p^k, q^k) = E(u^{k+1}, v^{k+1}) &+ \alpha_1 \underbrace{\big(R_1(u^{k+1}) + R_1^*(p^k) - \langle u^{k+1}, p^k\rangle\big)}_{= D_{R_1}^{p^k}(u^{k+1}, u^k)} \\ &+ \alpha_2 \underbrace{\big(R_2(v^{k+1}) + R_2^*(q^k) - \langle v^{k+1}, q^k\rangle\big)}_{= D_{R_2}^{q^k}(v^{k+1}, v^k)}, \end{aligned} \tag{17}$$

$$F_1(u^{k+1}, p^k) = E(u^{k+1}, v^{k+1}) + \alpha_1\big(R_1(u^{k+1}) + R_1^*(p^k) - \langle u^{k+1}, p^k\rangle\big), \tag{18}$$

$$F_2(v^{k+1}, q^k) = E(u^{k+1}, v^{k+1}) + \alpha_2\big(R_2(v^{k+1}) + R_2^*(q^k) - \langle v^{k+1}, q^k\rangle\big). \tag{19}$$

For convenience we will use the following notations

$$\begin{aligned} z^k :=& (u^k, v^k) \qquad \forall k \geq 0 \\ r^k :=& (p^k, q^k) \qquad p^k \in \partial R_1(u^k),\ q^k \in \partial R_2(v^k). \end{aligned}$$

The surrogate function $F$ will then read

$$F(z^{k+1}, r^k) = F(u^{k+1}, v^{k+1}, p^k, q^k).$$

We can now show the sufficient decrease property of (17) for subsequent iterates.

**Lemma 3** *(Sufficient decrease property).* The iterates generated by (14) satisfy the descent estimate

$$F(z^{k+1}, r^k) + \rho_2\|z^{k+1} - z^k\|^2 \leq F(z^k, r^{k-1}) \tag{20}$$

In addition we observe

$$\begin{aligned} \lim_{k\to\infty} D_{R_1}^{symm}(u^{k+1}, u^k) = 0 \qquad & \lim_{k\to\infty} D_{R_2}^{symm}(v^{k+1}, v^k) = 0 \\ \lim_{k\to\infty} D_{G_1}^{symm}(u^{k+1}, u^k) = 0 \qquad & \lim_{k\to\infty} D_{G_2}^{symm}(v^{k+1}, v^k) = 0. \end{aligned}$$

*Proof.* From (12) we consider the following step for $J_1 = \alpha_1 R_1 + \varepsilon_1 G_1$

$$\begin{aligned} u^{k+1} =& \arg\min_u \big\{E(u, v^k) + \alpha_1 D_{R_1}^{p^k}(u, u^k) + \varepsilon_1 D_{G_1}(u, u^k)\big\} \\ =& \arg\min_u \big\{E(u, v^k) + \alpha_1 R(u) + \varepsilon_1 G(u) - \langle \alpha_1 p^k + \varepsilon_1 \nabla G(u^k), u - u^k\rangle\big\}. \end{aligned}$$

Computing the optimality condition we obtain

$$\alpha_1(p^{k+1} - p^k) + \nabla_u E(u^{k+1}, v^k) + \varepsilon_1(\nabla G(u^{k+1}) - \nabla G(u^k)) = 0$$

Taking the dual product with $u^{k+1} - u^k$ yields

$$\alpha_1 \underbrace{\langle p^{k+1} - p^k, u^{k+1} - u^k \rangle}_{= D_{R_1}^{symm}(u^{k+1}, u^k)} + \underbrace{\langle \nabla_u E(u^{k+1}, v^k), u^{k+1} - u^k \rangle}_{\geq E(u^{k+1}, v^k) - E(u^k, v^k)}$$
$$+\varepsilon_1 \underbrace{\langle \nabla G_1(u^{k+1}) - \nabla G_1(u^k), u^{k+1} - u^k \rangle}_{= D_{G_1}^{symm}(u^{k+1}, u^k)} = 0.$$

Using the convexity estimate $E(u^{k+1}, v^k) - E(u^k, v^k) \leq -\langle \nabla_u E(u^{k+1}, v^k), u^{k+1} - u^k \rangle$ we obtain the inequality

$$\alpha_1 D_{R_1}^{symm}(u^{k+1}, u^k) + \varepsilon_1 D_{G_1}^{symm}(u^{k+1}, u^k) + E(u^{k+1}, v^k) - E(u^k, v^k) \leq 0$$

$$\alpha_1 \big( D_{R_1}^{p^k}(u^{k+1}, u^k) + D_{R_1}^{p^{k+1}}(u^k, u^{k+1}) \big) + \varepsilon_1 D_{G_1}^{symm}(u^{k+1}, u^k) + E(u^{k+1}, v^k)$$
$$\leq E(u^k, v^k).$$

Adding $\alpha_1 D_{R_1}^{p^{k-1}}(u^k, u^{k-1})$ to both sides, using the strong convexity of $G_1$ and the surrogate function notation, we get

$$F_1(u^{k+1}, p^k) + \alpha_1 \big( D_{R_1}^{p^{k+1}}(u^k, u^{k+1}) + D_{R_1}^{p^{k-1}}(u^k, u^{k-1}) \big) + \varepsilon_1 \gamma_1 \|u^{k+1} - u^k\|^2 \leq F_1(u^k, p^{k-1}).$$

Using the trivial estimate for the Bregman distances, we get the decrease property

$$F_1(u^{k+1}, p^k) + \varepsilon_1 \gamma_1 \|u^{k+1} - u^k\|^2 \leq F_1(u^k, p^{k-1}).$$

Similarly for $v$, we obtain

$$F_2(v^{k+1}, q^k) + \varepsilon_2 \gamma_2 \|v^{k+1} - v^k\|^2 \leq F_2(v^k, q^{k-1}).$$

Summing up these estimates, we verify the sufficient decrease property (20), with positive
$\rho_2 = \max\{\varepsilon_1 \gamma_1, \varepsilon_2 \gamma_2\}$. We also observe

$$0 \leq \Delta^k \leq E(z^k) - E(z^{k+1}).$$

with

$$\Delta^k := \alpha_1 D_{R_1}^{symm}(u^{k+1}, u^k) + \alpha_2 D_{R_2}^{symm}(v^{k+1}, v^k) + \varepsilon_1 D_{G_1}^{symm}(u^{k+1}, u^k) + \varepsilon_2 D_{G_2}^{symm}(v^{k+1}, v^k).$$

Summing over $k = 0, \ldots, N$

$$\sum_{k=0}^{N} \Delta^k \leq \sum_{k=0}^{N} E(z^k) - E(z^{k+1}) = E(z^0) - E(z^{N+1}) \leq E(z^0) - \inf_z E(z) < \infty.$$

Taking the limit $N \to \infty$ implies

$$\sum_{k=0}^{\infty} \Delta^k < \infty$$

thus $\lim_{k\to\infty} D_{R_1}^{symm}(u^{k+1}, u^k) = 0$ , $\lim_{k\to\infty} D_{G_1}^{symm} = 0$, $\lim_{k\to\infty} D_{R_2}^{symm}(v^{k+1}, v^k) = 0$, $\lim_{k\to\infty} D_{G_2}^{symm}(v^{k+1}, v^k) = 0$, due to $\alpha_1$, $\alpha_2$, $\varepsilon_1$, $\varepsilon_2 > 0$. ■

In order to show that the sequences generated by (14) approach the set of critical point we first estimate a bound for the subgradients of the surrogate functions and verify some properties of the limit point set. We first write the subdifferential of the surrogate function as

$$w^{k+1} := \begin{pmatrix} \nabla_u E(u^{k+1}, v^{k+1}) + \alpha_1(p^{k+1} - p^k) \\ \nabla_v E(u^{k+1}, v^{k+1}) + \alpha_2(q^{k+1} - q^k) \\ u^k - u^{k+1} \\ v^k - v^{k+1} \end{pmatrix} \in \partial F(z^{k+1}, r^k) \tag{21}$$

with $p^k \in \partial R_1(u^k)$ and $q^k \in \partial R_2(v^k)$ being equivalent to $u^k \in \partial R_1^*(p^k)$ and $v^k \in \partial R_2^*(q^k)$, respectively.

**Lemma 4** *(A subgradient lower bound for the iterates gap).* Suppose Assumptions 1 and 2 hold. Then the iterates (14) satisfy

$$\|w^{k+1}\| \le \rho_1 \|z^{k+1} - z^k\| \tag{22}$$

$w^{k+1} \in \partial F(z^{k+1}, r^k)$ as defined in (21) and $\rho_1 = \max\{1 + \varepsilon_1\delta_1, 1 + \varepsilon_2\delta_2 + L_2\}$.

*Proof.* From (21) we know

$$\begin{aligned} \|w^{k+1}\| \le\ & \|\nabla_u E(u^{k+1}, v^{k+1}) + \alpha_1(p^{k+1} - p^k)\| + \|\nabla_v E(u^{k+1}, v^{k+1}) + \alpha_2(q^{k+1} - q^k)\| \\ & + \|u^k - u^{k+1}\| + \|v^k - v^{k+1}\| \end{aligned}$$

From the optimality conditons of (14b) and (14d), we compute

$$\begin{aligned} \|w^{k+1}\| \le\ & \|\nabla_u E(u^{k+1}, v^{k+1}) + \alpha_1(p^{k+1} - p^k)\| + \|\nabla_v E(u^{k+1}, v^{k+1}) + \alpha_2(q^{k+1} - q^k)\| \\ & + \|u^k - u^{k+1}\| + \|v^k - v^{k+1}\| \\ =\ & \varepsilon_1 \underbrace{\|\nabla G_1(u^{k+1}) - \nabla G_1(u^k)\|}_{\le \delta_1\|u^{k+1} - u^k\|} + \underbrace{\|\nabla_u E(u^{k+1}, v^{k+1}) - \nabla_u E(u^{k+1}, v^k)\|}_{\le L_2\|v^{k+1} - v^k\|} \\ & + \varepsilon_2 \underbrace{\|\nabla G_2(v^{k+1}) - \nabla G_2(v^k)\|}_{\le \delta_2\|v^{k+1} - v^k\|} + \|u^{k+1} - u^k\| + \|v^{k+1} - v^k\| \\ \le\ & (1 + \varepsilon_1\delta_1)\|u^{k+1} - u^k\| + (1 + \varepsilon_2\delta_2 + L_2)\|v^{k+1} - v^k\| \\ \le\ & \rho_1 \|z^{k+1} - z^k\|. \end{aligned}$$

with $\rho_1 = \max\{1 + \varepsilon_1\delta_1, 1 + \varepsilon_2\delta_2 + L_2\}$. Here we used the Lipschitz-continuity of $\nabla G_i$ and $\nabla E$. ■

Following [41,42], we verify some properties of the limit point set. Let $\{z^k\}_{k\in\mathbb{N}}$ and $\{r^k\}_{k\in\mathbb{N}}$ be sequences generated by (14). The set of limit points is defined as

$$\omega(z^0, r^0) := \Big\{(\bar{z}, \bar{r}) \in \mathbb{R}^n \times \mathbb{R}^n : \exists \text{ an increasing sequence of integers } \{k_j\}_{j\in\mathbb{N}} \\ \text{such that } \lim_{j\to\infty} z^{k_j} = \bar{z} \text{ and } \lim_{j\to\infty} r^{k_j} = \bar{r}\Big\}.$$

As in [41, Definition 5.4, Proposition 5.5], we are going to assume that $R_i$, $i = 1, 2$ has locally bounded subgradients.

**Lemma 5** Suppose Assumptions 1 and 2 hold. Let $\{z^k\}_{k\in\mathbb{N}}$ be a sequence generated by (14) which is assumed to be bounded. Let $(\bar{z}, \bar{r}) \in \omega(z^0, r^0)$. Then the following assertion holds

$$\lim_{k\to\infty} F(z^{k+1}, r^k) = F(\bar{z}, \bar{r}) = E(\bar{z}). \tag{23}$$

*Proof.* Since $(\bar{z}, \bar{r})$ is a limit point of $\{(z^k, r^k)\}_{k\in\mathbb{N}}$, $\{(z^k, r^k)\}_{k\in\mathbb{N}}$, there exist subsequences $\{z^{k_j}\}_{j\in\mathbb{N}}$ and $\{r^{k_j}\}_{j\in\mathbb{N}}$ such that $\lim_{j\to\infty} z^{k_j} = \bar{z}$ and $\lim_{j\to\infty} r^{k_j} = \bar{r}$, respectively. We immediately obtain

$$\begin{aligned}\lim_{j\to\infty} F(z^{k_j}, r^{k_j-1}) &= \lim_{j\to\infty} \left\{E(z^{k_j}) + \alpha_1 D_{R_1}^{p^{k_j-1}}(u^{k_j}, u^{k_j-1}) + \alpha_2 D_{R_2}^{q^{k_j-1}}(v^{k_j}, v^{k_j-1})\right\} \\ &= E(\bar{z})\end{aligned}$$

due to the continuity of $E$ and $\lim_{j\to\infty} D_{R_1}^{p^{k_j-1}}(u^{k_j}, u^{k_j-1}) = 0$ and $\lim_{j\to\infty} D_{R_2}^{q^{k_j-1}}(v^{k_j}, v^{k_j-1}) = 0$. From the sufficient decrease property we conclude (23). ■

**Lemma 6** *(Properties of limit point set).* The limit point set $w(z^0)$ is a non empty, compact and connected set, the objective function $E$ is constant on $w(z^0)$ and we have $\lim_{k\to\infty} \operatorname{dist}(z^k, w(z^0)) = 0$. *Proof.* This follows steps as in [42, Lemma 5]. ■

To finally prove global convergence of (14), we will use the following Kurdyka-Łojasiewicz property defined and the result from [42]. Before recalling the definition, we introduce the notion of distance between any subset $S \subset \mathbb{R}^d$ and any point $x \in \mathbb{R}^d$ defined as

$$\operatorname{dist}(x, S) := \begin{cases} \inf\{\|y - x\| : y \in S\} & S \neq \emptyset \\ \infty & S = \emptyset \end{cases},$$

where $\|\cdot\|$ denotes the Euclidean norm.

**Lemma 7** *(Uniformised KL property).* Let $\Omega$ be a compact set and let $E : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ be a proper and l.s.c. function. Assume that $E$ is constant on $\Omega$ and satisfy the KL property at each point in $\Omega$. Then there exists $\varepsilon > 0$, $\eta > 0$ and $\varphi \in C^1((0, \eta))$ that satisfies the same conditions as in Definition KL, such that for all $\bar{u} \in \Omega$ and all $u$ in

$$\{u \in \mathbb{R}^n \,|\, \operatorname{dist}(u, \Omega) < \varepsilon\} \cap \{u \in \mathbb{R}^n \,|\, E(\bar{z}) < E(z) < E(z) + \eta\} \tag{24}$$

condition KL is satisfied. *Proof.* Follows from [42]. ■

With these results we can now show global convergence of (14).

*Proof of Theorem 1.* By the boundedness assumption on $\{(z^k, r^k)\}_{k\in\mathbb{N}}$, there exist converging subsequences $\{z^{k_j}\}_{j\in\mathbb{N}}$ and $\{r^{k_j}\}_{j\in\mathbb{N}}$ such that $\lim_{j\to\infty} z^{k_j} = \bar{z}$ and $\lim_{j\to\infty} r^{k_j} = \bar{r}$, respectively. We know from Lemma 5 that (23) is satisfied.

(i) KL property holds for $E$ and therefore for $E^k$ and we write

$$\varphi'\big(F(z^k, r^{k-1}) - E(\bar{z})\big)\mathrm{dist}\big(0, \partial F(z^k, r^{k-1})\big) \geq 1.$$

From Lemma 4 we obtain

$$\varphi'\big(F(z^k, r^{k-1}) - E(\bar{z})\big) \geq \rho_1^{-1}\|z^k - z^{k-1}\|^{-1},$$

and from the concavity of $\varphi$ we know that

$$\begin{aligned}&\varphi\big(F(z^k, r^{k-1}) - E(\bar{z})\big) - \varphi\big((F(z^{k+1}, r^k) - E(\bar{z})\big)\\ \geq\ &\varphi'\big(F(z^k, r^{k-1}) - E(\bar{z})\big)\big(F(z^k, r^{k-1}) - F(z^{k+1}, r^k)\big).\end{aligned}$$

Thus, we obtain

$$\frac{\varphi\big(F(z^k, r^{k-1}) - E(\bar{z})\big) - \varphi\big(F(z^k, r^{k-1}) - E(\bar{z})\big)}{F(z^k, r^{k-1}) - F(z^k, r^{k-1})} \geq \rho_1^{-1}\|z^k - z^{k-1}\|^{-1}.$$

From (20) with Lemma 3 and using the abbreviation

$$\varphi^k := \varphi(F(z^k, r^{k-1}) - E(\bar{z})),$$

it follows

$$\frac{\|z^{k+1} - z^k\|^2}{\|z^k - z^{k-1}\|} \leq \frac{\rho_1}{\rho_2}(\varphi^k - \varphi^{k+1}).$$

Multiplying by $\|z^k - z^{k-1}\|$ and using Young's inequality ($2\sqrt{ab} \leq a + b$ )

$$2\|z^{k+1} - z^k\| \leq \frac{\rho_1}{\rho_2}(\varphi^k - \varphi^{k+1}) + \|z^k - z^{k-1}\|.$$

Summing up from $k = 1, \ldots, N$ we get

$$\begin{aligned}\sum_{k=1}^{N}\|z^{k+1} - z^k\| &\leq \frac{\rho_1}{\rho_2}(\varphi^1 - \varphi^{N+1}) + \|z^1 - z^0\| + \|z^{N+1} - z^N\|\\ &\leq \frac{\rho_1}{\rho_2}\varphi^1 + \|z^1 - z^0\| < \infty.\end{aligned}$$

In addition we observe that the finite length property implies that the sequence $\{z^k\}_{k\in\mathbb{N}}$ is a Cauchy sequence and hence is a convergent sequence. For each $z^r$ and $z^s$ with $s > r > l$ we have

$$\|z^r - z^s\| = \|\sum_{k=r}^{s-1} z^{k+1} - z^k\| \leq \sum_{k=r}^{s-1}\|z^{k+1} - z^k\|.$$

(ii) The proof follows in a similar fashion as in [41, Lemma 5.9]

■

**Remark 2** *(Extension to d blocks).* The analysis described above holds for the general setting of $d$ blocks

$$\min_{\{u_1,\dots,u_d\}} E(u_1,\dots,u_d) + \sum_{i=1}^{n} \alpha_i^k D_{J_i}^{p_i^k}(u_i, u_i^k). \tag{25}$$

The update for each of the d blocks then reads

$$\begin{aligned} u_i^{k+1} &= \arg\min_{u_i} \left\{ E(u_1^{k+1}, u_2^{k+1}, \dots, u_{i-1}^{k+1}, u_i^k, u_{i+1}^k, \dots, u_d^k) + \alpha_i D_{J_i}^{p_i^k}(u_i, u_i^k) \right\} \\ p_i^{k+1} &= p_i^k - \frac{1}{\alpha_i}\Big(\nabla_{u_i} E(u_1^{k+1}, u_2^{k+1}, \dots, u_{i-1}^{k+1}, u_i^{k+1}, u_{i+1}^k, \dots, u_d^k)\Big). \end{aligned}$$

## 4. Numerical results

In this section we present numerical results for our joint reconstruction and segmentation model described in (11). We demonstrate its advantages and limitations, as well as a discussion on the parameter choice. In the first part, we focus on bubbly flow segmentation for simulated data. In the second part, we show results for real data acquired at the Cancer Research UK, Cambridge Institute, for tumour segmentation.

*Quality measure.* To assess the performance of the reconstruction we will compare our solutions $u$ with respect to the groundtruth $u^{gt}$. As quality measure we use the relative reconstuction error (RRE) and the peak signal to noise ratio (PSNR) defined as

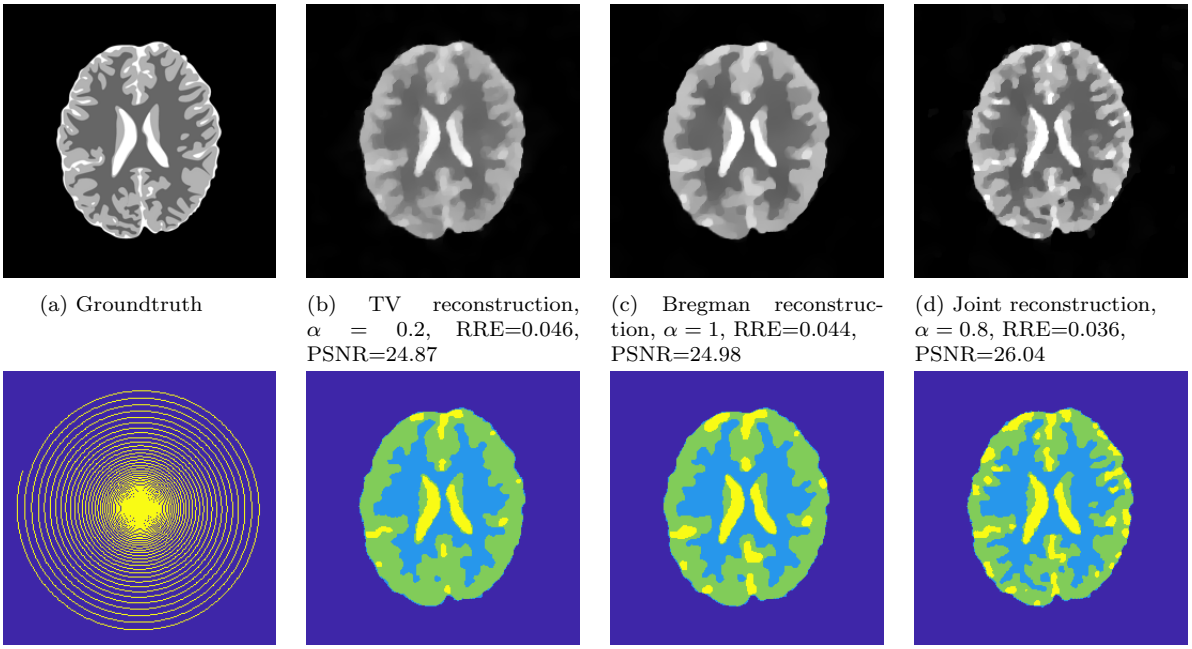
- $\mathrm{RRE}(u, u^{gt}) = \|u^{gt} - u\|_2 / \|u^{gt}\|_2$
- $\mathrm{PSNR}(u, u^{gt}) = 10 \log_{10}\left(\frac{\max(u)}{\|u^{gt}-u\|_2/N}\right)$

For the segmentation quality, we will use the relative segmentation error (RSE) to compare our segmentations $v$ with respect to the true segmentations $v^{gt}$

- $\mathrm{RSE}(v, v^{gt}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{v_i^{gt}, v_i}$

where $N$ is the number of pixels in the image, $\delta$ is the Kronecker delta function that will count the number of mis-classified pixels.

Before we present our two applications, we show a more detailed result of the phantom brain in Figure 1. In this example, we show the TV reconstruction 2b, where the parameter $\alpha$ has been optimised with respect to PSNR and its sequential segmentation 2f with optimal $\beta$ with respect to RSE. In 2c and 2g we present Bregman reconstruction and sequential segmentation where the Bregman iteration has been stopped according to the discrepancy principle Equation 7 and $\beta$ has been optimised with respect to RSE. These parameter choices for the sequential approaches will be used



(a) Groundtruth

(b) TV reconstruction, $\alpha = 0.2$, RRE=0.046, PSNR=24.87

(c) Bregman reconstruction, $\alpha = 1$, RRE=0.044, PSNR=24.98

(d) Joint reconstruction, $\alpha = 0.8$, RRE=0.036, PSNR=26.04

(e) Sampling matrix, 15%

(f) Segmentation, $\beta = 0.001$ RSE=0.061

(g) Bregman segmentation, $\beta = 0.001$, RSE=0.065

(h) Joint segmentation, $\beta = 0.001$, $\delta = 0.01$ RSE=0.057

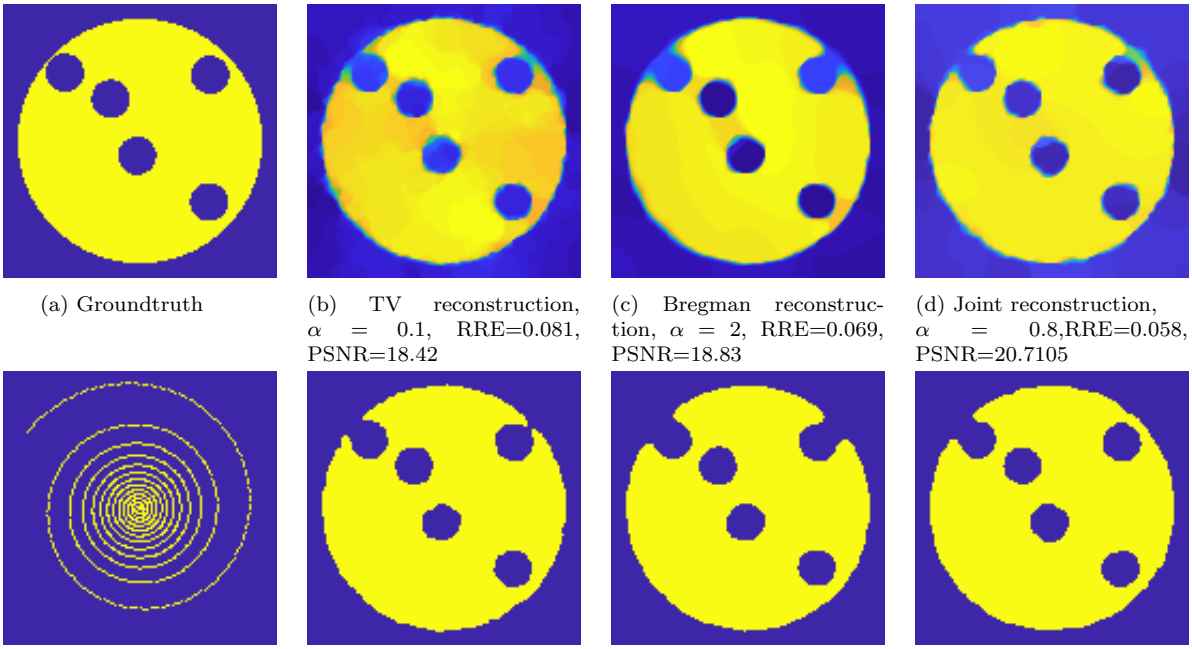
Figure 2: We consider 15% of the simulated $k$-space for the brain phantom, where Gaussian noise ($\sigma = 0.25$) was added. We compare results for the total variation reconstruction and total variation based Bregman iterative reconstruction and their segmentation in a sequential approach with our joint model. We show that both reconstruction and segmentation are improved.

in the whole paper.

In this first result, we clearly see that the joint approach performs much better compared to the separate steps in Figures 2b, 2f and 2c, 2g. Both reconstruction and segmentation are improved and more details are recovered. We refer to Appendix A for more simulated examples.

### *4.1. Bubbly flow*

The first application considered is the characterisation of bubbly flows using MRI. Bubbly flows are two-phase flow systems of liquid and gas trapped in bubbles, which are common in industrial applications such as bioreactors [43] and hydrocarbon processing units [44]. MRI has been successfully used to characterise the bubble size distribution [45, 46] and the liquid velocity field of bubbly flows [47, 48]; these properties govern the heat and mass transfer between the bubbles and the liquid which ultimately determine the efficiency of these industrial systems. However, when studying fast flowing systems, the acquisition time for fully sample $k$-space is too long to resolve the temporal changes; the most common method of breaking the temporal resolution barrier is through under-sampling. It is therefore critical to develop reconstruction techniques for highly under-sampled $k$-space data for the accurate reconstruction of the MRI images which would be subsequently used in calculating the bubble size distribution or in



(a) Groundtruth

(b) TV reconstruction, $\alpha = 0.1$, RRE=0.081, PSNR=18.42

(c) Bregman reconstruction, $\alpha = 2$, RRE=0.069, PSNR=18.83

(d) Joint reconstruction, $\alpha = 0.8$,RRE=0.058, PSNR=20.7105

(e) Sampling matrix, 8%

(f) Segmentation, $\beta = 0.001$, RSE=0.0093

(g) Bregman segmentation, $\beta = 0.001$, RSE=0.017

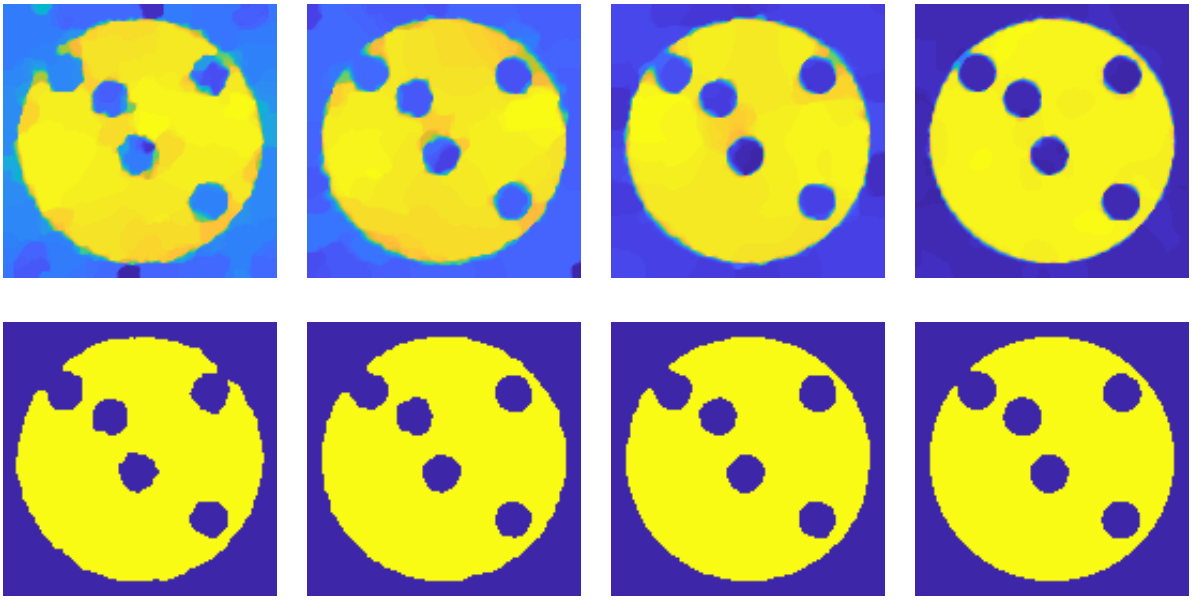
(h) Joint segmentation, $\beta = 0.001$, $\delta = 1$, RSE=0.0102

Figure 3: Results of the TV reconstruction and Bregman iterative reconstruction and their segmentation in the sequential approach are compared with our joint model. Both MSE and SSIM are improved in the joint approach. The data was corrupted with Gaussian noise with $\sigma = 0.35$.

studying the hydrodynamics of the system.

We apply our joint reconstruction and segmentation approach to simulated bubbly flow imaging. In Figure 3 we present some results for synthetic data, where Figure 3a represents the groundtruth magnitude image, from which we simulate its $k$-space following the forward model described in (1). From the full $k$-space we collect 8% of the samples using the sampling matrix in Figure 3e and we corrupt the data with Gaussian noise of standard deviation $\sigma = 0.35$. In Figure 3b and 3f we show the results for the total variation regularised reconstruction and its segmentation performed sequentially. In the same sequential way, we show the results for the Bregman iterative regularization in Figure 3c and 3g. In the last column in Figure 3d and 3h, we finally show the results for our joint approach. Although the TV and the Bregman approaches are already quite good, we can see that both RRE and PSNR are improved using our model in the reconstruction and the segmentation. Smaller details, such as the top right bubble contour, are better detected when solving the joint problem. As the goal of the bubbly flow application is to detect bubble size distribution, this improvement is really advantageous.

We tested the robustness of our approach by corrupting the data with different signal to noise ratio (SNR) and by considering different amount of sampling. In Figure 5 we show in the top row the reconstructions obtained with the joint model for different SNR (which corresponds to different standard deviation $\sigma$) and in the bottom row the corresponding segmentation obtained by the joint approach. To complement

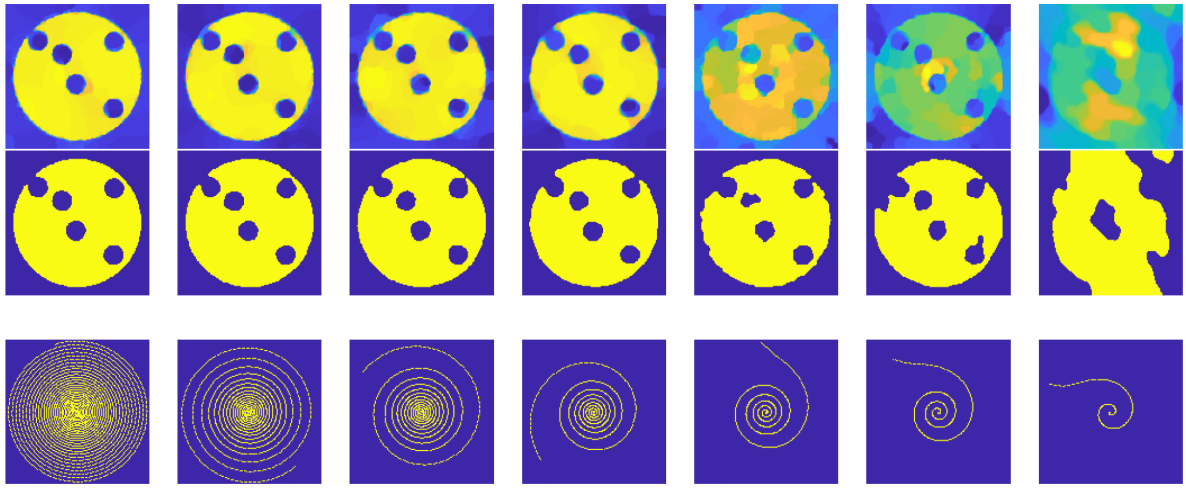


(a) SNR=10.56, $\sigma$=0.70 (b) SNR=12.69, $\sigma$=0.56 (c) SNR=16.68, $\sigma$=0.35 (d) SNR=32.83, $\sigma$=0.06

Figure 4: Top row: reconstructions obtained by the joint model with different SNR. Bottom row: corresponding segmentations.

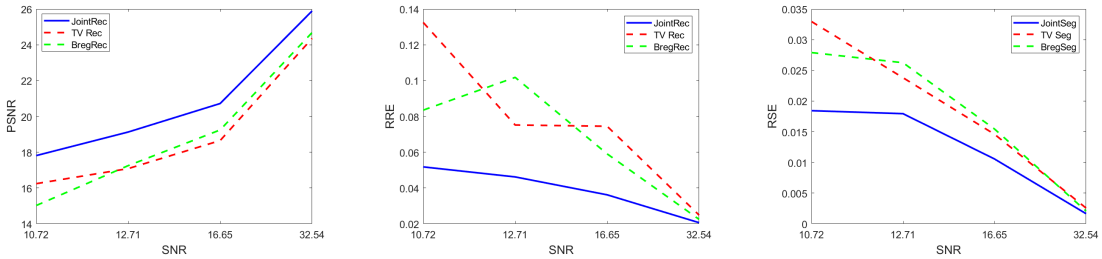
this information, we show in Figure 6 how the PSNR, RRE and RSE are affected, for the joint approach (blue lines) and for the separate approaches, TV (red dotted lines) and Bregman TV (green dotted lines). As expected, with the SNR increasing the error decreases. We can see that the joint approach performs better than the sequential approach for any SNR. The improvement is even more significant for very noisy data. As in practice we often observe high levels of noise, the joint approach is able to takle this problem better than the traditional sequential approaches.

It is also interesting to investigate how the joint approach performs with very low undersampling rates. In Figure 3e we show joint reconstructions (top row) and corresponding segmentations (bottom row) for decreasing sampling rates. We can see that up to 5% results are still very good. Using 3 and 2% of the samples the results are less clean but it is possible to identify the main structures. In contrast, 1% sampling is not enough to retrieve a good image reconstruction and consequently its segmentation. In Figure 7, we plot PSNR , RRE and RSE for different sampling rates. The blue lines represent the error for our joint approach, while the red and green dotted lines are for the sequential TV and sequential Bregman TV approaches. We can see that up to 5% sampling the error measures do not change significantly. However, for lower rates, the improvement is more significant. This is highly beneficial for the bubbly flow application as increasing the temporal resolution is really important to keep track of the gas flowing in the pipe.



(a) 25% (b) 12.5% (c) 8% (d) 5% (e) 3% (f) 2% (g) 1%

Figure 5: Top row: reconstructions obtained by the joint model with different sampling rates. Bottom row: corresponding segmentations. The joint reconstruction and segmentation is able to detect the main structures down to 5% of the samples. Up to 2% the results are less clean but still acceptable. Using only 1% of the data is not enough to produce the image and segmentation.



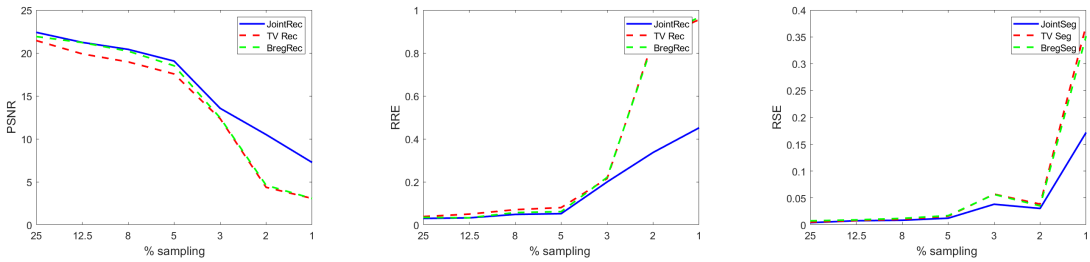

(a) PSNR (b) RRE (c) RSE

Figure 6: Error plots for different SNR. From left to right, we show the PSNR, RRE and RSE, respectively, for different levels of noise in the measurements. The blue lines represent the error for our joint approach, while the red and green dotted lines are for the sequential TV and sequential Bregman TV approaches. For each SNR, the joint model performs better than the separate methods. This improvement is even more significant for noisier data, which is highly advantageous as in practice we often observe lower SNR.

### *4.2. Cancer imaging*

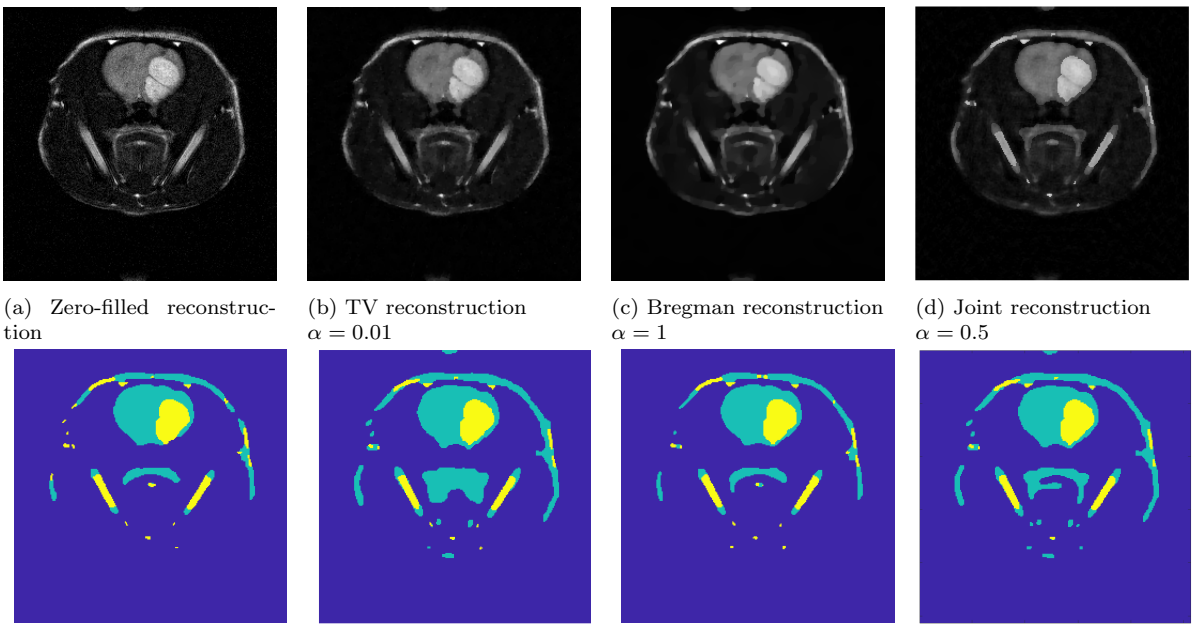
In this subsection, we illustrate the performance of the joint model for real cancer data. At the Cancer Research UK, Cambridge Institute, researchers acquire every day a huge amount of MRI scans to assess tumour progression and response to therapy [49]. For this reason, it is very convenient to have fast sampling through compressed sensing, and automatic segmentation methods. Furthermore, reconstructions with enhanced edges are advantageous to facilitate clinical analysis.

Here we show our results for MRI data of a rat bearing a glioblastoma. The MR image represents the rat head where the brain is the gray area in the top half of the image. Inside this gray region, a tumour is clearly visible appearing as a brighter area. For this




(a) PSNR (b) RRE (c) RSE

Figure 7: Error plots varying sampling rate. From left to right, we show the PSNR, RRE and RSE, respectively, for different levels of noise in the measurements. The blue lines represent the error for our joint approach, while the red and green dotted lines are for the sequential TV and sequential Bregman TV approaches. The joint appraoch performs better than the sequential cases. The gain is not very significant for higher sampling rates, but it becomes more important for lower rates, starting from 3%.

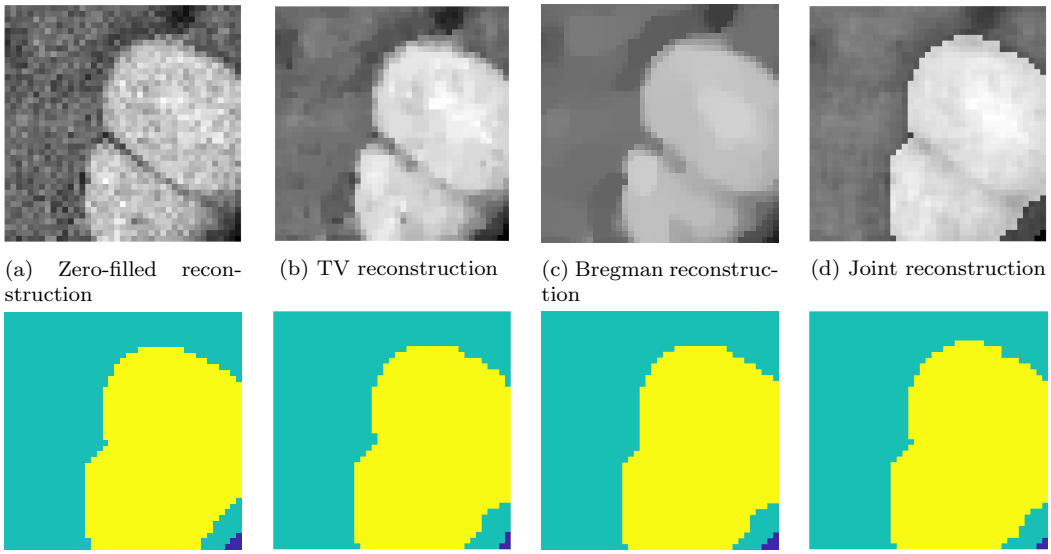


(a) Zero-filled reconstruction (b) TV reconstruction $\alpha = 0.01$ (c) Bregman reconstruction $\alpha = 1$ (d) Joint reconstruction $\alpha = 0.5$

(e) Segmentation (f) Segmentation $\beta = 0.07$ (g) Bregman segmentation $\beta = 0.07$ (h) Joint segmentation $\beta = 0.01$, $\delta = 0.01$

Figure 8: Reconstructions and segmentation for real MRI data. We select 15% of the samples using a spiral mask. The image show a rat brain bearing a tumour (brighter region). The zero-filled reconstruction 8a and the TV regularised reconstruction 8b are shown together with their sequential segmentation 8e and 8f respectively. In the last column 8d and 8h we show the results for our model. The parameter $\alpha$ for the TV reconstruction and for the joint reconstruction has been chosen such that it achieves visually optimal in the sense that it resolve all the details (e.g. the darker line cutting the tumour transversally).

experiment, we acquired the full $k$-space and present the zero-filling reconstruction in Figure 8a and the sequential segmentation in Figure 8e. As discussed already in the previous section, the zero-filled reconstruction presents noise and artefact which may complicate the segmentation. We want to show that the compressed sensing approach and in particular the joint model can improve this reconstruction. Given the full $k$-



(a) Zero-filled reconstruction (b) TV reconstruction (c) Bregman reconstruction (d) Joint reconstruction

(e) Segmentation (f) Segmentation (g) Bregman segmentation (h) Joint segmentation

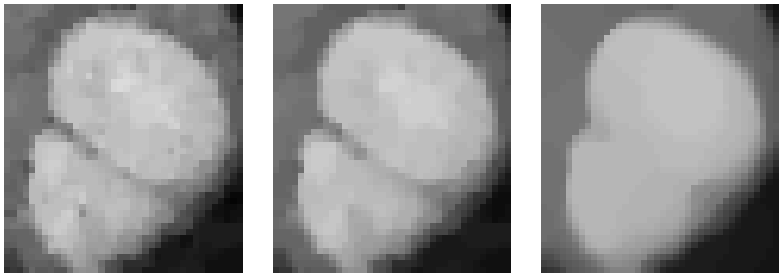
Figure 9: Zoomed section on the tumor for the different approaches.

space, we select 15% of the samples using a spiral mask. In Figure 8b, 8f and Figure 8c, 8g we show the results for the sequential approaches. In Figure 8d and 8h we show the joint reconstruction and the joint segmentation obtained for the same data. The regularised approaches perform better that the zero-filled reconstruction, producing less noisy results. However, our joint model is able to produce a cleaner reconstruction where the edges that defines the tumour and the brain are very well detected. In Figure 9, we show a zoomed section where it is easy to assess that the joint model tackle the noise and detect the region of interest. We can see that we are able to improve the reconstruction and automatically identify the tumour in the brain. The degree of enhancement of the edges in the reconstruction is controllable by the parameter $\delta$ in the model (11). In the next subsection we present a discussion on how to tune this parameter.

### *4.3. Parameter choice rule*

In the model proposed in (11), the parameters that we need to choose are $\alpha$, $\beta$ and $\delta$. In this section we discuss a rule to choose them depending on the desired results. Some examples will clarify these empirical choices.

- $\alpha$ balances the total variation regularization term in the reconstruction for the magnitude. The higher the $\alpha$, the more piecewise constant the reconstruction will be. See Figure 10.
- $\beta$ defines the scale of the objects that will be detected in the segmentation. Smaller values of $\beta$ will allow for smaller objects. See Figure 11.
- $\delta$ is the parameter linking the reconstruction and the segmentation. To better illustrate its role, let us consider a zero-filling like reconstruction and segmentation:



(a) $\alpha = 0.001$ (b) $\alpha = 0.01$ (c) $\alpha = 0.1$

Figure 10: The parameter $\alpha$ balances the data fidelity term and the total variation regularisation for the reconstruction. Smaller values of $\alpha$ produce a reconstruction closer to the data fitting term, hence less smooth as in 10a. As $\alpha$ increases in 10b the solution gets smoother and less noisy. Finally for large values it tends to become more piecewise constant as in 10c.

(a) $\beta = 0.1$ (b) $\beta = 1$ (c) $\beta = 3$

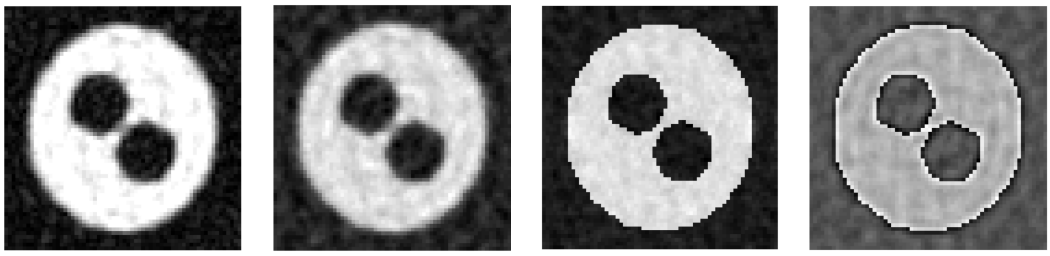
Figure 11: The parameter $\beta$ determines the scale of the objects that we are segmenting. Smaller values of $\beta$ can detect smaller objects 11a, which are lost for intermediate values 11b. Finally very large values only detect main structures 11c.

$$\imath_{\mathcal{SF}\cdot=f}(u) + \delta \sum_i \in \Omega \sum_{j=1}^{\ell} v_{ij}(c_j - u_i)^2 dx + \beta \|\nabla v\| \to \min_{u,v} \tag{26}$$

where $\imath(u) = \begin{cases} +\infty, & \text{if } \mathcal{SF}u \neq f \\ 0, & \text{if } \mathcal{SF}u = f \end{cases}$. This problem is solving the zero-filled reconstruction and segmentation jointly. For $\delta = 0$, the reconstruction is the zero-filling solution. In Figure 12 we can see the impact of the segmentation term on the reconstruction for increasing values of $\delta$. We can see that for very small$\delta$ the result is close to the zero-filling solution. For $\delta = 1$ the noise from the model is present as expected but in addition the boundaries are enhanced. For large $\delta$ the boundaries are still very pronounced and the noise is also amplified.

### *4.4. Comparison with another joint approach*

We present a comparison of our joint model with another non-convex method, namely the Potts model approach by [33], described in Subsection 2.4. The major advantage of the joint reconstruction and segmentation using the Potts model is that it does not require to select explicitely the number of regions to segment, although this depends on

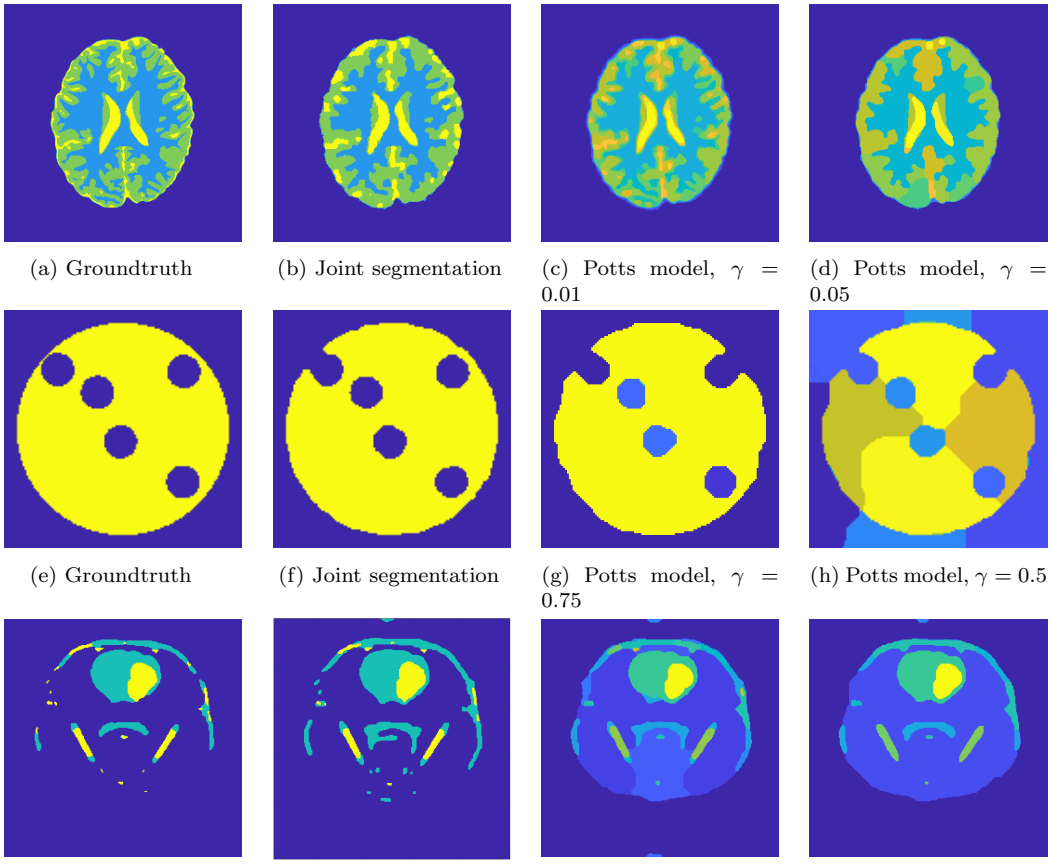


(a) $\delta = 0$ (b) $\delta = 0.1$ (c) $\delta = 1$ (d) $\delta = 2.5$

Figure 12: We show the reconstructions obtained solving (26) for different values of $\delta$. For $\delta = 0$ we get the zero-filling solution. For small $\delta$ we expect the solution to be similar to the zero-filling reconstruction. For $\delta = 1$ we see the effect of the joint term on the reconstruction. The solution presents the same noise artefacts but having in addition very sharp boundaries. Finally, for very large $\delta$ we still have enhanced boundaries but we also amplify the noise.

the choice of the regularisation parameter. However, by definition, it only produces a piecewise constant image, therefore a segmentation, and not a reconstruction. This is useful in some applications where one is only interested in the segmentation. In contrast, our model produces both reconstruction and segmentation. In Figure 13, we show the results for some examples. Note that because the results of the Potts model are in the range of the groundtruth image, while our segmentation are in label space, we can not directly use the RSE as before, or common metrics that compare actual intensities such as PSNR and structure similarity index measure (SSIM), for comparison. For example, for some tissue in class 1, to label it class 2 is as wrong as to label it class 3. However in this case, the SSIM and PSNR will favour the label class 2.

We therefore focus on a visual assessment and show the results of the Potts model for two different choices of the regularisation parameter $\gamma$. We recall that the proposed model requires to determine the number of classes in advance, while the model for comparison estimates the number of regions but this depends on the choice of the regularisation parameter. In the top row, we can see that the Potts model, although it retrieves the shape of the main structures for the brain phantom example, it overestimates the number of classes. By increasing the parameter $\gamma$, this issue is not resolved as it assigns different intensities to objects of the same category. In contrast, our approach is able to identify the desired classes as in the groundtruth. For the bubble case (middle row), we can see that our method works better and our segmentation is more accurate, while the Potts model fails to capture shape details (e.g., outer circle is distorted) and again overestimates the number of regions. We can also see that, when slightly decreasing $\gamma$, the Potts model is very sensitive to artefacts. For the real MR data (bottom row), we see that both methods identify the tumour quite well. Because we were only interested in identifying three classes as tumour, brain and background, we do not segment the outer region (rat's head), captured insted by the Potts model. However, the Potts model only produces the segmentation, while our method, as shown in Figure 8, also produces an enhanced reconstruction with sharp edges.



(a) Groundtruth (b) Joint segmentation (c) Potts model, $\gamma = 0.01$ (d) Potts model, $\gamma = 0.05$

(e) Groundtruth (f) Joint segmentation (g) Potts model, $\gamma = 0.75$ (h) Potts model, $\gamma = 0.5$

(i) Segmentation from zero-filled reconstruction (j) Joint segmentation (k) Potts model, $\gamma = 0.05$ (l) Potts model, $\gamma = 0.1$

Figure 13: Comparison of our joint approach with the Potts model. Noise level and undersampling rate are described in Figure 2, 3 and 8. The results are presented for three different examples and for two different choices of the regularisation parameter $\gamma$. We can see that the Potts model tends to overestimate the number of regions to segment.

## 5. Conclusion

In this paper, we have investigated a novel mathamatical approach to perform simultaneously reconstruction and segmentation from undersampled MRI data. Our motivation was to include in the reconstruction prior knowledge of the objects we are interested in. By interconnecting the reconstruction and the segmentation terms, we can achieve sharper reconstructions and more accurate segmentations. We derived a variational model based on Bregman iteration and we have verified its convergence properties. With our approach we show that by solving the more complicated joint model, we are able to improve both reconstruction and segmentation compared to the traditional sequential approach. This suggests that with the joint model it is possible to reduce error propagations that occur in sequential analysis, when the segmentation is separate and posterior to the reconstruction.

We have tested our method for two different application, which are bubbly flow

and cancer imaging. In both cases, the reconstructions are sharper and finer structures are detected. Additionally, the segmentations also benefit from the improvement in the reconstructions. We have found that the joint model outperforms the sequential approach by exploiting prior information on the objects that we want to segment. In addition, we also show that our method performs better than the well-known Potts model. We also presented a discussion on the parameter choice rule that offer some insight on how to tune the parameters according to the desired result. It is interesting to notice that, with our model, we are able to control the segmentation effect on the reconstruction. Furthermore, when the final analysis of the MR image is indeed the segmentation, it is possible to bias the reconstruction towards the piecewise constant solution, yet preserving finer details in the structure.

In our set-up, we have specified the intensity constants characteristic of the region of interests, which were known a priori for our applications. However, it is possible to also include the optimisation with respect to $c_j$ in our joint model, where the same convergence guarantees hold (see Remark 2). Nevertheless, one limitation of the model is the need to specify the number of regions to be segmented.

In our future research, we would like to study the extension of this model for the bubbly flow to the reconstruction of the magnitude image as well as the phase image. The goal is not only to extract the structure of the bubble, but also to estimate velocity information, which is encoded in the phase image. As the problem is non-convex in the joint argument but also non-convex with respect to the phase, we need to derive a different convergence analysis.

## Acknowledgements

VC acknowledges the financial support of the Cambridge Cancer Centre. MB acknowledges the Leverhulme Trust Early Career Fellowship "Learning from mistakes: a supervised feedback-loop for imaging applications", the Isaac Newton Trust and the Cantab Capital Institute for the Mathematics of Information. MJE and CBS acknowledge support from Leverhulme Trust project "Breaking the non-convexity barrier", EPSRC grant "EP/M00483X/1", EPSRC centre "EP/N014588/1", the Cantab Capital Institute for the Mathematics of Information, and from CHiPS and NoMADS (Horizon 2020 RISE project grant). Moreover, CBS is thankful for support by the Alan Turing Institute. We would like to thank Kevin Brindle and Alan Wright for the acquisition of the data in Figure 8.

## References

[1] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[2] David L. Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[3] Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic resonance in medicine*, 58(6):1182–1195, 2007.

[4] Albert Macovski. Noise in mri. *Magnetic Resonance in Medicine*, 36(3):494–497, 1996.

[5] Matthias J. Ehrhardt and Marta M. Betcke. Multi-contrast MRI reconstruction with structure-guided total variation. *SIAM Journal on Imaging Sciences*, 9(3):1084–1106, 2016.

[6] Martin Benning, Lynn Gladden, Daniel Holland, Carola-Bibiane Schönlieb, and Tuomo Valkonen. Phase reconstruction from velocity-encoded MRI measurements–a survey of sparsity-promoting variational approaches. *Journal of Magnetic Resonance*, 238:26–43, 2014.

[7] Yves Meyer. *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*, volume 22. American Mathematical Soc., 2001.

[8] David Strong and Tony Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse problems*, 19(6):S165, 2003.

[9] Martin Benning and Martin Burger. Ground states and singular vectors of convex variational regularization methods. *Methods and Applications of Analysis*, 2012.

[10] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[11] Lev M. Bregman. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex optimization. *USSR Computational Mathematics and Mathematical Physics*, 1967.

[12] Krzysztof C. Kiwiel. Proximal minimization methods with generalized bregman functions. *SIAM journal on control and optimization*, 35(4):1142–1168, 1997.

[13] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D, vol. 60, pp. 259-268*, 1992.

[14] Vladimir Alekseevich Morozov. On the solution of functional equations by the method of regularization. In *Soviet Math. Dokl*, volume 7, pages 414–417, 1966.

[15] David Mumford and Jayant Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5):577–685, 1989.

[16] Luigi Ambrosio and Vincenzo Maria Tortorelli. Approximation of functional depending on jumps by elliptic functional via t-convergence. *Communications on Pure and Applied Mathematics*, 43(8):999–1036, 1990.

[17] Antonin Chambolle. Image segmentation by variational methods: Mumford and shah functional and the discrete approximations. *SIAM Journal on Applied Mathematics*, 55(3):827–863, 1995.

[18] Leif Boysen, Angela Kempe, Volkmar Liebscher, Axel Munk, Olaf Wittich, et al. Consistencies and rates of convergence of jump-penalized least squares estimators. *The Annals of Statistics*, 37(1):157–183, 2009.

[19] Thomas Pock, Antonin Chambolle, Daniel Cremers, and Horst Bischof. A convex relaxation approach for computing minimal partitions. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 810–817, June 2009.

[20] Tony F. Chan and Luminita A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

[21] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*, volume 254. Clarendon Press Oxford, 2000.

[22] Tony F. Chan, Selim Esedoglu, and Mila Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM journal on applied mathematics*, 66(5):1632–1648, 2006.

[23] Jan Lellmann, Björn Lellmann, Florian Widmann, and Christoph Schnörr. Discrete and continuous models for partitioning problems. *International journal of computer vision*, 104(3):241–269, 2013.

[24] Jan Lellmann, Jörg Kappes, Jing Yuan, Florian Becker, and Christoph Schnörr. Convex multi-class image labeling by simplex-constrained total variation. In *International conference on scale space and variational methods in computer vision*, pages 150–162. Springer, 2009.

[25] Leonie Zeune, Guus van Dalum, Leon W. M. M. Terstappen, Stephan A. van Gils, and Christoph Brune. Multiscale segmentation via bregman distances and nonlinear spectral analysis. *SIAM journal on imaging sciences*, 10(1):111–146, 2017.

[26] Ronny Ramlau and Wolfgang Ring. A Mumford-Shah level-set approach for the inversion and segmentation of X-ray tomography data. *Journal of Computational Physics, 221, 2, 539 - 557*, 2007.

[27] Ronny Ramlau Esther Klann and Wolfgang Ring. A Mumford-Shah level-set approach for the inversion and segmentation of SPECT/CT data. *Inverse Problems and Imaging, 5, 1, 137 - 166*, 2011.

[28] Esther Klann. A Mumford-Shah-like method for limited data tomography with an application to electron tomography. *SIAM J. Imaging Sci., 4(4), 10291048*, 2011.

[29] Martin Burger, Carolin Rossmanith, and Xiaoqun Zhang. Simultaneous reconstruction and segmentation for dynamic SPECT imaging. *Inverse Problems*, 32(10):104002, 2016.

[30] François Lauze, Yvain Quéau, and Esben Plenge. Simultaneous reconstruction and segmentation of CT scans with shadowed data. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 308–319. Springer, 2017.

[31] Dominique Van de Sompel and Michael Brady. Simultaneous reconstruction and segmentation algorithm for positron emission tomography and transmission tomography. *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 1035-1038*, 2008.

[32] Mikhail Romanov, Anders Bjorholm Dahl, Yiqiu Dong, and Per Christian Hansen. Simultaneous tomographic reconstruction and segmentation with class priors. *Inverse Problems in Science and Engineering, 24,8, 1432-1453*, 2015.

[33] Martin Storath, Andreas Weinmann, Jürgen Frikel, and Michael Unser. Joint image reconstruction and segmentation using the Potts model. *Inverse Problems 31 (2015) 025003 (29pp)*, 2015.

[34] Jose Caballero, Wenjia Bai, Anthony N. Price, Daniel Rueckert, , and Joseph V. Hajnal. Application-driven MRI: Joint reconstruction and segmentation from undersampled MRI data. *Med Image Comput Comput Assist Interv;17(Pt 1):106-13.*, 2014.

[35] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J Math Imaging Vis 40: 120. doi:10.1007/s10851-010-0251-1*, 2011.

[36] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.

[37] Ernie Esser, Xiaoqun Zhang, and Tony F. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.

[38] Thomas Pock, Daniel Cremers, Horst Bischof, and Antonin Chambolle. An algorithm for minimizing the Mumford-Shah functional. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1133–1140. IEEE, 2009.

[39] Antonin Chambolle, Daniel Cremers, and Thomas Pock. A convex approach to minimal partitions. *SIAM Journal on Imaging Sciences*, 5(4):1113–1158, 2012.

[40] Stanislaw Lojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles*, 117:87–89, 1963.

[41] Martin Benning, Marta M. Betcke, Matthias J. Ehrhardt, and Carola-Bibiane Schönlieb. Choose your path wisely: gradient descent in a Bregman distance framework. *arXiv preprint arXiv:1712.04045*, 2017.

[42] Jerome Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating minimization for nonconvex and nonsmooth problems. *Mathematical Programming, 146(1-2):459-494*, 2014.

[43] Jeffrey J. Chalmers. Cells and bubbles in sparged bioreactors. In *Cell Culture Engineering IV*,

pages 311–320. Springer, 1994.

[44] Wolf-Dieter Deckwer and Robert W. Field. *Bubble column reactors*, volume 200. Wiley New York, 1992.

[45] Daniel J. Holland, Andrew Blake, Alexander B. Tayler, Andrew J. Sederman, and Lynn F. Gladden. Bubble size measurement using bayesian magnetic resonance. *Chemical engineering science*, 84:735–745, 2012.

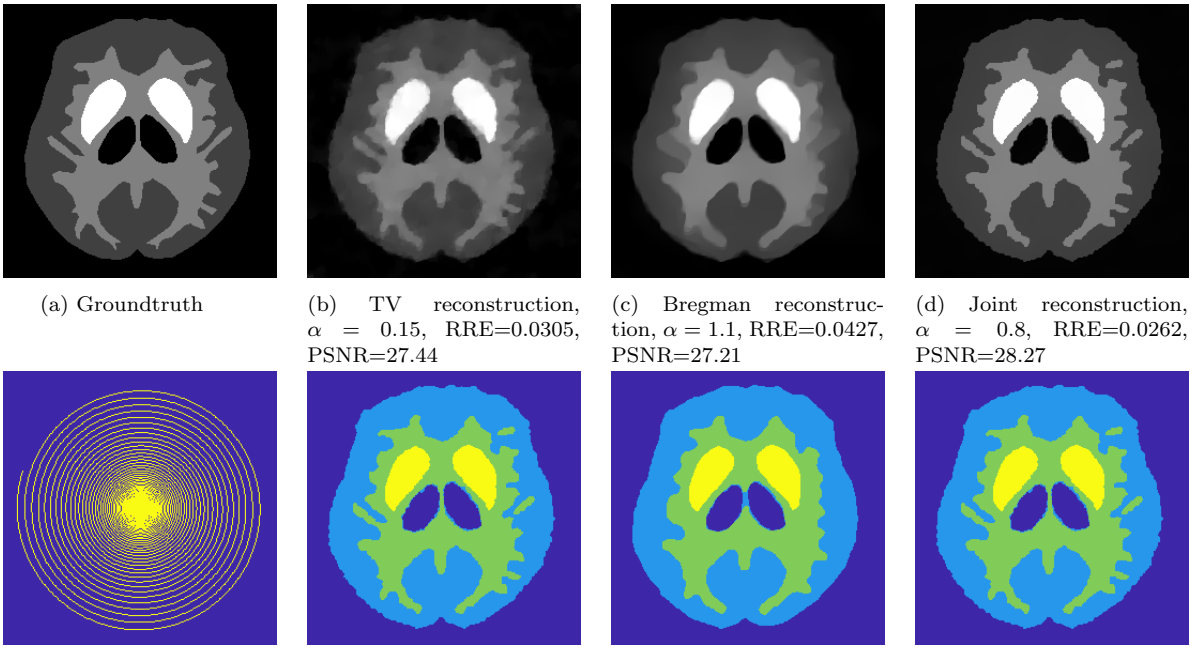
[46] Alexander B. Tayler, Daniel J. Holland, Andrew J. Sederman, and Lynn F. Gladden. Applications of ultra-fast MRI to high voidage bubbly flow: measurement of bubble size distributions, interfacial area and hydrodynamics. *Chemical engineering science*, 71:468–483, 2012.

[47] Daniel J. Holland, Dmitry M. Malioutov, Andrew Blake, Andrew J. Sederman, and L. F. Gladden. Reducing data acquisition times in phase-encoded velocity imaging using compressed sensing. *Journal of magnetic resonance*, 203(2):236–246, 2010.

[48] Alexander B. Tayler, Daniel J. Holland, Andrew J. Sederman, and Lynn F. Gladden. Exploring the origins of turbulence in multiphase flow using compressed sensing MRI. *Physical review letters*, 108(26):264505, 2012.

[49] Tiago B. Rodrigues, Eva M. Serrao, Brett W. C. Kennedy, De-En Hu, Mikko I. Kettunen, and Kevin M. Brindle. Magnetic resonance imaging of tumor glycolysis using hyperpolarized 13 c-labeled glucose. *Nature medicine*, 20(1):93, 2014.

## Appendix A. Numerical results on phantoms



(a) Groundtruth

(b) TV reconstruction, $\alpha = 0.15$, RRE=0.0305, PSNR=27.44

(c) Bregman reconstruction, $\alpha = 1.1$, RRE=0.0427, PSNR=27.21

(d) Joint reconstruction, $\alpha = 0.8$, RRE=0.0262, PSNR=28.27

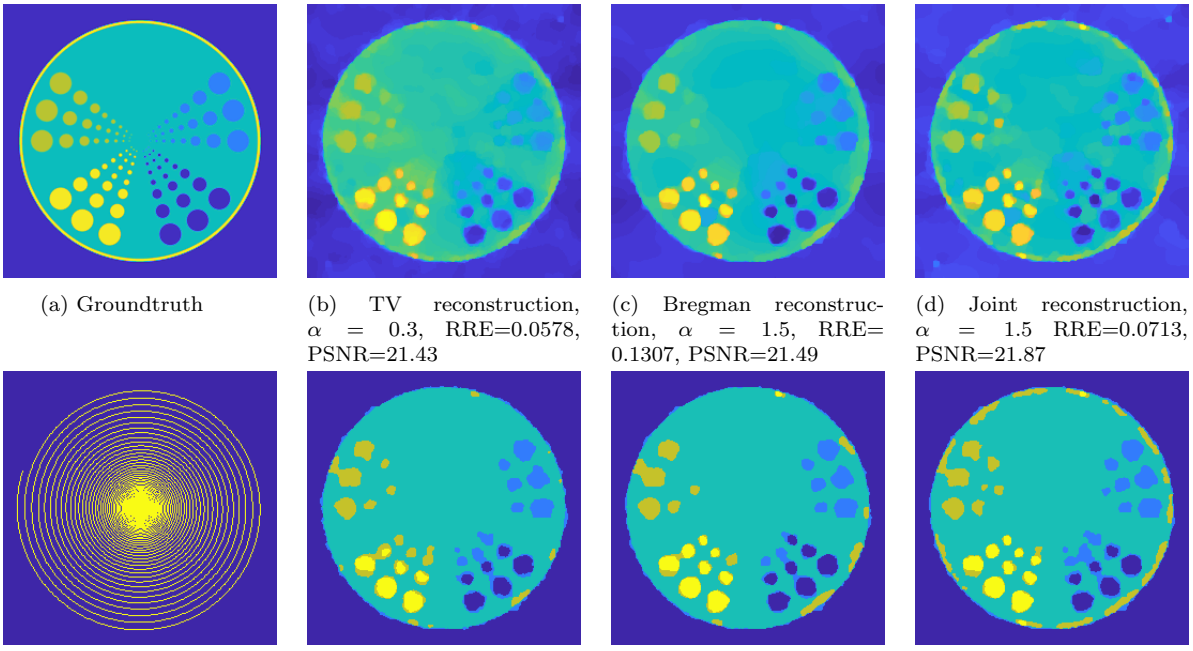
(e) Sampling matrix, 15%

(f) Segmentation, $\beta = 0.001$ RSE=0.0219

(g) Bregman segmentation, $\beta = 0.001$ RSE=0.0399

(h) Joint segmentation, $\beta = 0.001$, $\delta = 2$, RSE=0.0219

Figure A1: This example shows clearly the effect of the parameter $\delta$ in the joint model. The segmentation is easy to achieve and we do not see a significant improvement in joint segmentation compared to the TV sequential segmentation, but there is a small gain compared to the sequential Bregman segmantation. However, the joint reconstruction results improved thanks to the parameter $\delta$ which biases the reconstruction to be closer to the segmentation.



(a) Groundtruth

(b) TV reconstruction, $\alpha = 0.3$, RRE=0.0578, PSNR=21.43

(c) Bregman reconstruction, $\alpha = 1.5$, RRE= 0.1307, PSNR=21.49

(d) Joint reconstruction, $\alpha = 1.5$ RRE=0.0713, PSNR=21.87

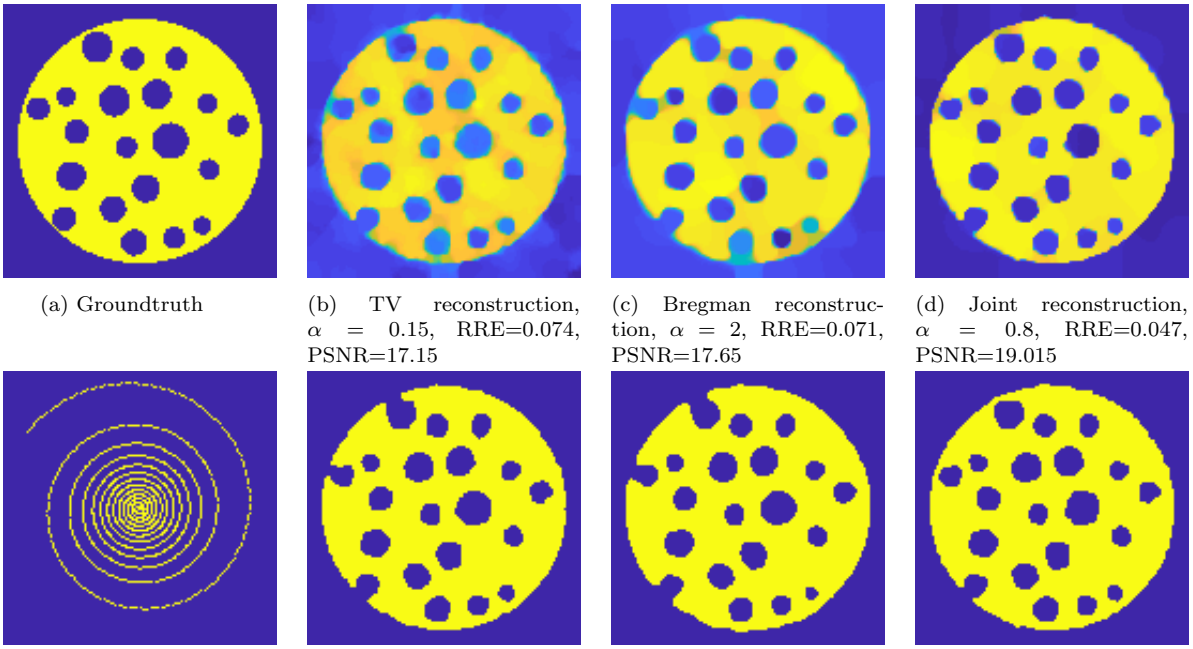
(e) Sampling matrix, 15%

(f) Segmentation, $\beta = 0.001$, RSE=0.096

(g) Bregman segmentation, $\beta = 0.001$, RSE=0.121

(h) Joint segmentation, $\beta = 0.001$, $\delta = 0.1$, RSE=0.091

Figure A2: In this example, we can see that the reconstructions are quite similar. However in the joint reconstruction, the outer yellow circle, which is completely ignored by the sequential reconstructions, is partially detected. This is also the case for the joint segmenation.



(a) Groundtruth

(b) TV reconstruction, $\alpha = 0.15$, RRE=0.074, PSNR=17.15

(c) Bregman reconstruction, $\alpha = 2$, RRE=0.071, PSNR=17.65

(d) Joint reconstruction, $\alpha = 0.8$, RRE=0.047, PSNR=19.015

(e) Sampling matrix, 8%

(f) Segmentation, $\beta = 0.01$, RSE=0.016

(g) Bregman segmentation, $\beta = 0.01$, RSE=0.022

(h) Joint segmentation, $\beta = 0.01$, $\delta = 1$, RSE=0.014

Figure A3: In this example for the bubbly flow, we can see clearly an improvement for both joint reconstruction and joint segmentation. The contrast in the joint reconstruction is better recovered and the segmentation is more accurate, especially for the bubbles close to the edge of the pipe. The joint method results particularly useful for the bubbly flow application.