

# Partial Discreteness: A Novel Prior for Magnetic Resonance Image Reconstruction

Gabriel Ramos-Llordén\*, Member, IEEE, Arnold J. den Dekker, and Jan Sijbers

**Abstract**—An important factor influencing the quality of magnetic resonance (MR) images is the reconstruction method that is employed, and specifically, the type of prior knowledge that is exploited during reconstruction. In this work, we introduce a new type of prior knowledge, partial discreteness (PD), where a small number of regions in the image are assumed to be homogeneous and can be well represented by a constant magnitude. In particular, we mathematically formalize the partial discreteness property based on a Gaussian Mixture Model (GMM) and derive a partial discreteness image representation that characterizes the salient features of partially discrete images: a constant intensity in homogeneous areas and texture in heterogeneous areas. The partial discreteness representation is then used to construct a novel prior dedicated to the reconstruction of partially discrete MR images. The strength of the proposed prior is demonstrated on various simulated and real k-space data-based experiments with partially discrete images. Results demonstrate that the PD algorithm performs competitively with state-of-the-art reconstruction methods, being flexible and easy to implement.

**Index Terms**—MRI reconstruction, partial discreteness, Gaussian Mixture Model, sparsity, segmentation.

## I. INTRODUCTION

MAGNETIC Resonance Imaging (MRI) is a unique imaging method for medical diagnosis because of its excellent image quality, the absence of ionizing radiation and its versatility [1]. However, MRI is a relatively slow technique, which hinders its applicability in speed demanding modalities (e.g., time-resolved acquisitions or 3-D imaging [2]), limits the patient throughput [1], and increases the chances of subject motion.

Hardware improvements and dedicated sequence design have been proposed to speed up the imaging process [3], but physical constraints such as slew-rate or gradient amplitudes [4] limit the acquisition time of each sample. Hence,

Manuscript received October 15, 2016; revised December 13, 2016; accepted December 20, 2016. Date of publication December 23, 2016; date of current version April 30, 2017. This work was supported by the Research Foundation - Flanders (FWO, Belgium) through project funding G037813N, the Belgian Science Policy Office through the Inter-university Attraction Poles Program (P7/11), and by the University of Antwerp (TOP BOF project 26824). *Asterisk indicates corresponding author.*

\*G. Ramos-Llordén is with the IMEC-Vision Lab, Department of Physics, University of Antwerp, 2610 Antwerp, Belgium (e-mail: gabriel.ramos-llorden@uantwerpen.be).

A. J. den Dekker is with the IMEC-Vision Lab, Department of Physics, University of Antwerp, 2610 Antwerp, Belgium, and also with the Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands.

J. Sijbers is with the IMEC-Vision Lab, Department of Physics, University of Antwerp, 2610 Antwerp, Belgium.

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the authors.

Digital Object Identifier 10.1109/TMI.2016.2645122

the only way to further decrease the scan time is to acquire less data samples. Simply reducing the number of samples, however, leads to image reconstruction artefacts, hampering both visual and quantitative analysis [2].

In MRI, the acquired samples, a.k.a. the k-space data, are samples of the Fourier transform of the spatial magnetization distribution (the image) [5]. From sampling theory, it is well-known that if the k-space is sampled with a Cartesian scheme fulfilling the Nyquist condition, exact recovery of the finite support image is possible [2]. However, if the number of k-space data points is reduced, the Nyquist condition is violated, making the inverse problem of image reconstruction ill-posed without prior knowledge [2].

Fortunately, prior knowledge comes in different forms. It can be included by exploiting image properties, such as smoothness, both in the image and Fourier domain [6], [7], sparsity in a specific image representation [4] (e.g., Fourier [8], Wavelet [9], [10], Curvelet [11], Shearlet [12], or redundant dictionaries [1], [13]), number of discrete gray levels [14], minimal Total Variation (TV) [15], [16], [17], limited image support [18], [19], or spatial constraints [20]–[22]. Additionally, prior knowledge can come from anatomical information derived from a training dataset [23], [24] or can be extracted from reference images [25].

In this work, we introduce a novel type of prior knowledge for MR images that possess, apart from heterogeneous regions, a number of quasi-constant intensity regions. That is, the magnitude range is assumed to be partially discrete. Partial discrete tomography has recently successfully been introduced as a prior in X-ray tomography [26], [27] and electron tomography [28], [29], however, to the authors' knowledge not yet in MRI. There is a number of MRI applications where the *partial discreteness* assumption may be exploited. Implants MR imaging is a paradigmatic example, where the homogeneous composition of implants naturally leads to partially discrete images [30]. Contrast-enhanced MRI sequences also produce images that meet the *partial discreteness* assumption, e.g. contrast-enhanced MR Angiography (MRA) [4]. Finally, the use of specific pulse sequences, such as T2-weighted or short tau inversion recovery (STIR) sequences, may produce hyper-intense regions in brain cyst imaging [31].

Specifically, our contribution is the following. We mathematically formalize the *partial discreteness* property and propose a decomposition of every image into its *partial discreteness representation* and its residual form. The *partial discreteness representation* is constructed from 1) an auto-learned Gaussian Mixture Model (GMM) [32] specifically

designed to fulfill the *partial discreteness* assumption and 2) the fusion of the *a posteriori* probability maps with intensity information, both derived from the GMM and the image itself. Because partially discrete images admit an accurate *partial discreteness representation*, we enforce sparsity in the residual form to promote this type of solution in the reconstruction of under-sampled MR images. In this work, the *partial discreteness* prior is implemented in a phase-constrained MR reconstruction formulation [33]–[37] with the common assumption of a smoothly varying phase image [4], [18].

We illustrate the potential of the *partial discreteness* prior by showing examples of applications with under-sampled simulated and real k-space data. Thereby, the proposed partially discrete reconstruction method is compared to state-of-the-art MR reconstruction methods.

This paper is organized as follows. Section II describes the MR image reconstruction problem as an optimization problem. In Section III, we present the novel *partial discreteness* prior, which is incorporated in a constrained optimization method in Section IV. Section V illustrates the application of this method to a variety of under-sampling scenarios, in comparison with state-of-the-art methods. This section also summarizes a sensitivity analysis of the proposed method to various parameters and deviations from assumptions. Finally, conclusions are summarized in Section VI.

## II. IMAGE RECONSTRUCTION AS AN OPTIMIZATION PROBLEM

### A. Algebraic Linear Model

MR image reconstruction is often described as an algebraic reconstruction problem. The starting point is the linear forward model that results when the Fourier integral is discretized on a spatial grid [5]. Let  $\mathbf{y} \in \mathbb{C}^M$  denote the k-space data. Then, the image  $\mathbf{x} \in \mathbb{C}^N$  is related to  $\mathbf{y}$  through the following algebraic linear model [38]:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (1)$$

with matrix  $\mathbf{A} \in \mathbb{C}^{M \times N}$  the Fourier encoding matrix and  $\mathbf{n} \in \mathbb{C}^M$  complex valued white Gaussian noise [5]. In the absence of prior knowledge,  $\mathbf{x}$  is commonly estimated by minimizing  $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ , which is, given the Gaussian noise statistics, equivalent to the Maximum Likelihood (ML) estimator [39].

If prior knowledge is incorporated, the reconstruction problem can be recast as a constrained optimization problem [4], [40]–[41] in the form of

$$\min_{\mathbf{x}} \Phi(\mathbf{x}) \quad s.t. \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \epsilon, \quad (2)$$

where  $\epsilon$  in the so called data fidelity condition is usually set below the expected noise level [4].

The function  $\Phi(\cdot)$ , which is called the prior term or simply the prior, is defined such that it incorporates prior knowledge into the solution. The lower  $\Phi(\mathbf{x})$ , the more  $\mathbf{x}$  is in agreement with the prior knowledge.

### B. A Note on the Phase

*Partial discreteness* is applicable to the magnitude of the image. Many methods have been proposed that enforce prior knowledge on the magnitude image only [4], [15], though MR images are inherently complex valued [39]. By noting that  $\mathbf{x} = \Psi\mathbf{m}$  with  $\mathbf{m} \in \mathbb{R}_+^N$ ,  $\Psi = \text{diag}(e^{i\psi_{\mathbf{x}}})$  and  $\psi_{\mathbf{x}}$  the phase of  $\mathbf{x}$ , the optimization problem (2) restricted to magnitude images, also known as phase-constrained formulation [33]–[37], can be elegantly reformulated as:

$$\min_{\mathbf{m}} \Phi(\mathbf{m}) \quad s.t. \quad \|\mathbf{y} - \tilde{\mathbf{A}}\mathbf{m}\|_2^2 \leq \epsilon \quad (3)$$

with  $\mathbf{m} \in \mathbb{R}_+^N$  and  $\tilde{\mathbf{A}} = \mathbf{A}\Psi$ .

In our work, the required phase estimate,  $\hat{\psi}_{\mathbf{x}}$ , is obtained from a low resolution recovered image, as proposed in [4]. Details are provided in subsection IV-C. Hence, in what follows, we will assume that  $\mathbf{x} \in \mathbb{R}_+^N$ .

## III. PARTIAL DISCRETENESS

### A. A Bayesian Model for Partial Discreteness

Consider the following decomposition of an image into a piece-wise homogeneous part and a texture part:

$$\mathbf{x} = \underbrace{\sum_{k=1}^K \mathbf{x}_{\mathcal{A}_k}}_{\text{piece-wise homogeneous part}} + \underbrace{\mathbf{x}_{\bar{\mathcal{A}}}}_{\text{texture part}}, \quad (4)$$

where  $A = \cup_{k=1}^K \mathcal{A}_k$  represents the union of  $K$  disjoint homogeneous regions (i.e., pixel sets) of  $\mathbf{x}$ , and  $\bar{\mathcal{A}}$  is the texture region. For each pixel  $n = 1, \dots, N$ ,  $\mathbf{x}_{\bar{\mathcal{A}}} \in \mathbb{R}_+^N$  and  $\mathbf{x}_{\mathcal{A}_k} \in \mathbb{R}_+^N$  are defined as

$$(\mathbf{x}_{\bar{\mathcal{A}}})_n = \begin{cases} (\mathbf{x})_n, & \text{if } n \in \bar{\mathcal{A}} \\ 0, & \text{if } n \notin \bar{\mathcal{A}} \end{cases} \quad (5)$$

and

$$(\mathbf{x}_{\mathcal{A}_k})_n = \begin{cases} \eta_k, & \text{if } n \in \mathcal{A}_k \\ 0, & \text{if } n \notin \mathcal{A}_k, \end{cases} \quad (6)$$

respectively, with  $\eta_k$  the constant intensity of the homogeneous set  $\mathcal{A}_k$ . In this work, we assume that  $|\mathcal{A}_k| \gg 1, \forall k$ . While ideally suited to describe partially discrete images, this model is unpractical since the location, cardinality, and intensity of the sets  $\{\mathcal{A}_k\}_{k=1}^K$  are unknown in practice.

Keeping Eq. (4) in mind, in this subsection we approximate Eq. (4) with a realizable model, effectively preserving the distinct characteristics of partially discrete images. To that end, a Bayesian framework is proposed which relies on 1) a Gaussian Mixture Model (GMM) that captures the particular intensity properties of Eq. (4) and 2) an unsupervised Bayesian probabilistic segmentation. The proposed methodology allows us to identify the different sets and estimate the probabilities that a pixel belongs to each of these sets (probabilistic segmentation) in an unsupervised manner. In what follows, we start with the construction of the GMM. The probabilistic segmentation is described afterwards. With these two main ingredients, we present the Bayesian model for partially discrete images in subsection III-B. With the *partial discreteness* property formalized, the *partial discreteness* prior is then defined

1) *The GMM Construction*: We consider the pixels' intensities of the image  $\mathbf{x}$ ,  $[x_1, \dots, x_N]^T$ , as a vector of realizations of a random variable,  $X : \Omega \mapsto \mathbb{R}^+$ . We assume that each pixel  $n = 1, \dots, N$  only belongs to one specific, unknown set  $\mathcal{A}_k$  or  $\bar{\mathcal{A}}$ . Assigning a probability to the event that a pixel belongs to a specific set, we define:

$$P(n \in \mathcal{A}_k) = \pi_k, \quad \text{with } k = 1, \dots, K \quad \text{and} \quad (7)$$

$$P(n \in \bar{\mathcal{A}}) = \pi_{\bar{\mathcal{A}}} = 1 - \sum_{k=1}^K \pi_k, \quad (8)$$

where it is assumed that the probabilities  $\pi_k$  and  $\pi_{\bar{\mathcal{A}}}$  are independent of the pixel  $n$ , and no a priori spatial information is incorporated. Hence, in what follows, we will denote  $P(n \in \mathcal{A})$  as  $P(\mathcal{A})$ . Each of the pixels' intensities,  $x_1, \dots, x_N$ , is assumed to have been generated by one (randomly selected and unknown) element of a set of  $K + 1$  random sources. The source-conditional distributions of the random variable  $X$  are characterized by the conditional probability density functions (PDFs)  $f_{X|\mathcal{A}_k}(x)$  and  $f_{X|\bar{\mathcal{A}}}(x)$  [42]. It is reasonable to assume that the individual conditional PDFs for the homogeneous sets all belong to the same location-scale family and can be well modeled by Gaussian distributions:

$$f_{X|\mathcal{A}_k}(x; \eta_k, \sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(x-\eta_k)^2}{2\sigma_k^2}}, \quad (9)$$

where the dispersion of  $X$  around the specific discrete value  $\eta_k$  is represented by the standard deviation  $\sigma_k$ . A small value of  $\sigma_k$  reflects the typical low dispersion of such homogeneous regions. Of course, that does not mean that no intensity variations are allowed, as these always occur in real scenarios.

The conditional PDF  $f_{X|\bar{\mathcal{A}}}(x; \theta_{\bar{\mathcal{A}}})$  models the random variable  $X$  in the texture part. Texture modeling through statistical distributions is beyond the scope of this paper. The interested reader is referred to [43]–[45]. We deem that the shape of  $f_{X|\bar{\mathcal{A}}}(x; \theta_{\bar{\mathcal{A}}})$  can be well modeled by a mixture of Gaussian PDFs, and therefore  $\theta_{\bar{\mathcal{A}}}$  are the Gaussian mixture model (GMM) parameters. The use of a GMM for  $f_{X|\bar{\mathcal{A}}}(x; \theta_{\bar{\mathcal{A}}})$  should be seen as a way to describe arbitrarily complex distributions [46] and not as an attempt to model quasi-discrete components, as in the case of  $f_{X|\mathcal{A}_k}(x; \eta_k, \sigma_k)$ .

Simple application of the law of total probability [42] with events  $\{n \in \mathcal{A}_k\}_{k=1}^K$  and  $\{n \in \bar{\mathcal{A}}\}$ , yields the final GMM:

$$f_X(x; \theta) = \sum_{k=1}^K \pi_k f_{X|\mathcal{A}_k}(x; \eta_k, \sigma_k) + \pi_{\bar{\mathcal{A}}} f_{X|\bar{\mathcal{A}}}(x; \theta_{\bar{\mathcal{A}}}) \quad (10)$$

with  $\theta = [\theta_{\mathcal{A}}, \theta_{\bar{\mathcal{A}}}]$  and where  $\theta_{\mathcal{A}} = \{\pi_k, \eta_k, \sigma_k\}_{k=1}^K$ .

2) *Bayesian Probabilistic Segmentation*: Bayes' theorem [42] now allows us to derive the *a posteriori* probabilities  $P(\mathcal{A}_k|x)$  and  $P(\bar{\mathcal{A}}|x)$  as

$$P(\mathcal{A}_k|x) = \frac{\pi_k f_{X|\mathcal{A}_k}(x; \eta_k, \sigma_k)}{f_X(x; \theta)} \quad \text{and} \quad (11)$$

$$P(\bar{\mathcal{A}}|x) = \frac{\pi_{\bar{\mathcal{A}}} f_{X|\bar{\mathcal{A}}}(x; \theta_{\bar{\mathcal{A}}})}{f_X(x; \theta)}, \quad (12)$$

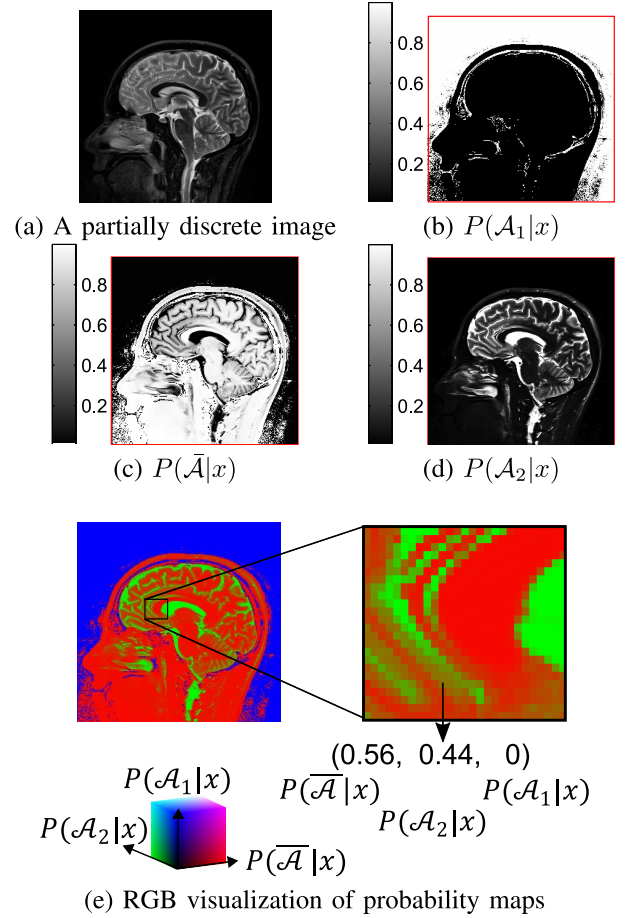


Fig. 1. Illustration of probability maps for a partially discrete image.

respectively. These *a posteriori* probabilities denote the probabilities of a pixel belonging to each of the  $K + 1$  sets given that its intensity is equal to  $x$ . If Eqs.(11)–(12) are evaluated for each pixel  $n = 1, \dots, N$ , we obtain what is dubbed throughout this work, probability maps, that is,  $\mathbf{p}_k \in [0, 1]^N$  and  $\mathbf{p}_{\bar{\mathcal{A}}} \in [0, 1]^N$  which are pixel-wise defined as

$$(\mathbf{p}_k)_n = \frac{\pi_k f_{X|\mathcal{A}_k}(x_n; \eta_k, \sigma_k)}{\sum_{k=1}^K \pi_k f_{X|\mathcal{A}_k}(x_n; \eta_k, \sigma_k) + \pi_{\bar{\mathcal{A}}} f_{X|\bar{\mathcal{A}}}(x_n; \theta_{\bar{\mathcal{A}}})} \quad (13)$$

and

$$(\mathbf{p}_{\bar{\mathcal{A}}})_n = \frac{\pi_{\bar{\mathcal{A}}} f_{X|\bar{\mathcal{A}}}(x_n; \theta_{\bar{\mathcal{A}}})}{\sum_{k=1}^K \pi_k f_{X|\mathcal{A}_k}(x_n; \eta_k, \sigma_k) + \pi_{\bar{\mathcal{A}}} f_{X|\bar{\mathcal{A}}}(x_n; \theta_{\bar{\mathcal{A}}})}. \quad (14)$$

Given a pixel, its *a posteriori* probability of belonging to a specific set can be determined from the corresponding probability map. This Bayesian framework frames a probabilistic segmentation scheme. Indeed, it is not possible to strictly assign pixels to specific sets but, instead, the probability of this assignment can be inferred. Another relevant point of the Bayesian segmentation is that not just local but *global* information is considered. This information is derived through  $f_X(x; \theta)$ . An illustration of the probabilistic segmentation associated to a partially discrete image is shown in Fig. 1.

These maps constitute the core of the *partial discreteness representation* that we propose in the next subsection III-B.



## B. The Partial Discreteness Prior

Given a partially discrete image  $\mathbf{x}$ , and its associated GMM  $f_X(x; \theta)$ , the *partial discreteness representation* of  $\mathbf{x}$ , denoted as  $\mathcal{P}(\mathbf{x}) \in \mathbb{R}_+^N$ , is defined as:

$$\mathcal{P}(\mathbf{x}) = \sum_{k=1}^K \eta_k \mathbf{p}_k + \mathbf{x}_\rho \circ \mathbf{p}_{\bar{A}} \quad (15)$$

where  $\circ$  denotes the Hadamard product and  $\mathbf{x}_\rho$  is a spatially filtered version of  $\mathbf{x}$  with a circularly symmetric Gaussian filter with standard deviation  $\rho$  in the image domain. For partially discrete images, the *partial discreteness representation* essentially behaves as Eq. (4).

### Behavior in homogeneous regions

In a homogeneous region  $\mathcal{A}_{k'}$ , the pixels' intensities minimally vary with respect to the mean  $\eta_{k'}$ . When the probability maps for the homogeneous regions are evaluated, all except the one associated to  $\mathcal{A}_{k'}$  approximate zero. This is due to the fact that their conditional PDFs  $\{f_{X|\mathcal{A}_k}(x; \eta_k, \sigma_k)\}_{k=1}^K$  are highly concentrated around their mean and when evaluated far from their mode they rapidly fall off to zero. As a consequence, each of the  $P(\mathcal{A}_k|x)$  with  $k \neq k'$  vanishes as well. Furthermore, the lack of intensity dispersion that characterizes the homogeneous regions is not captured by  $f_{X|\bar{A}}(x; \theta_{\bar{A}})$ . Therefore, exclusively the remaining probability map of  $\mathcal{A}_{k'}$ ,  $\mathbf{p}_{k'}$ , is approximately one. Hence, the *partial discreteness representation* becomes

$$\mathcal{P}(\mathbf{x}) \approx \eta_{k'} \mathbf{p}_{k'} \approx \mathbf{x}_{\mathcal{A}_{k'}}, \quad (16)$$

as desired.

### Behavior in texture regions

In texture regions, the characteristic intensity variability is solely represented by the conditional PDF  $f_{X|\bar{A}}(x; \theta_{\bar{A}})$ . As a result, we get

$$\mathcal{P}(\mathbf{x}) \approx \mathbf{x}_\rho \circ \mathbf{p}_{\bar{A}} \approx \mathbf{x}_\rho. \quad (17)$$

Note that  $\mathcal{P}(\mathbf{x})$  does not exactly approach  $\mathbf{x}$  but a Gaussian filtered version of  $\mathbf{x}$ . The use of the Gaussian filter should be seen as a way to make *partial discreteness* a stable representation under very high spatial frequency perturbations that do not correspond to the original texture  $\mathbf{x}$ . In our experiments, the value  $\rho = 2$  was consistently used. With this value, the corresponding cut-off frequency is high enough to preserve structural details. Note that in the limit case, which corresponds to  $\rho = 0$ ,  $\mathcal{P}(\mathbf{x}) \approx \mathbf{x}$ , since the Gaussian kernel degenerates into a delta function.

### Behavior in the frontier between regions

In the frontier between regions, none of the *a posteriori* probabilities has a prevailing effect (see zoomed image of Fig.1(e)). Indeed,  $\mathcal{P}(\mathbf{x})$  is a mixture of intensities. Particularly, the intensity along a given profile which crosses two regions is a (convex) combination of two values. If the two regions are a homogeneous region  $\mathcal{A}_k$  and the texture region  $\bar{A}$ , such values are the mean  $\eta_k$  and  $\mathbf{x}_\rho$ , the filtered texture. The closer we are to  $\mathcal{A}_k$ , the higher  $\mathbf{p}_k = 1 - \mathbf{p}_{\bar{A}}$  is. Thus,  $\mathcal{P}(\mathbf{x})$  approaches  $\eta_k$ . The nearer we are to  $\bar{A}$ , the larger  $\mathbf{p}_{\bar{A}}$  is and  $\mathcal{P}(\mathbf{x})$  then approaches  $\mathbf{x}_\rho$ . If the interface divides two homogeneous regions, let's say  $\mathcal{A}_k$  and  $\mathcal{A}_{k'}$ , then  $\mathcal{P}(\mathbf{x})$  is a convex combination of the two corresponding mean values, that is,  $\eta_k$  and  $\eta_{k'}$ .

In summary,  $\mathcal{P}(\mathbf{x}) \approx \mathbf{x}$  for partially discrete images. Therefore, instead of the strict but unpractical model of Eq. (4), we can fairly justify the employment of the *partial discreteness representation* for the kind of images targeted in this work.

Based on this representation, the *partial discreteness* prior for the optimization problem (2) can be defined. We first note that a *partial discreteness representation*  $\mathcal{P}(\mathbf{x})$  can be assigned to every image  $\mathbf{x}$ . In practice, this requires that the number of homogeneous regions  $K$  is given and we have estimates of the GMM parameters,  $\hat{\theta} = [\hat{\theta}_{\mathcal{A}}, \hat{\theta}_{\bar{A}}]$ , (details about the GMM learning procedure are presented in subsection IV-B). Then, the estimated probability maps can be constructed pixel-wise from Eqs.(13-14) by replacing the given GMM parameters by the GMM estimates.

A *partial discreteness representation* is now obtained by substituting the estimated probability maps and  $\{\hat{\eta}_k\}_{k=1}^K$  in Eq. (15). The thus obtained *partial discreteness representation*  $\mathcal{P}(\mathbf{x})$  is unique for each  $\mathbf{x}$ . More interestingly, every image  $\mathbf{x}$  can be represented through its *partial discreteness representation*, including images that are not strictly partially discrete. Indeed, every image  $\mathbf{x} \in \mathbb{R}_+^N$  can be decomposed as

$$\mathbf{x} = \mathcal{P}(\mathbf{x}) + (\mathcal{I} - \mathcal{P})(\mathbf{x}), \quad (18)$$

where  $\mathcal{I}: \mathbb{R}^N \mapsto \mathbb{R}^N$  is the identity operator. The usefulness of this decomposition is that the second term, the residual form, serves as a measure of the partial discreteness error for all types of images. On the one hand, a partially discrete image is well represented by its partial discreteness representation, so its residual form,  $(\mathcal{I} - \mathcal{P})(\mathbf{x})$ , can be assumed to be small. On the other hand, images that do not obey the partial discreteness assumption possess a non-negligible residual form. Obviously, the prior  $\Phi(\cdot)$  should be defined in agreement with this reasoning.

For partially discrete images,  $(\mathcal{I} - \mathcal{P})(\mathbf{x})$  is almost zero except along edges. Since edges generally represent only a small fraction of the partially discrete image, we can justify that the residual form is sparse. Searching for the  $\mathbf{x}$  that has the sparsest residual form implies to make the  $l_0$  (pseudo) norm of the residual form minimal. As this combinatorial optimization problem is intractable, it is common practice to resort to other sparsity inducing norms, such as the  $l_1$  norm, the  $l_p$  ( $0 < p < 1$ ) semi-norms or other non-convex functionals such as log penalty functions [2]. To illustrate *partial discreteness*, we use the  $l_1$  norm in our experiments.

Hence, we define our *partial discreteness* prior  $\Phi(\cdot)$  as

$$\Phi(\mathbf{x}) = \|(\mathcal{I} - \mathcal{P})(\mathbf{x})\|_1. \quad (19)$$

## IV. METHOD

In this section, we propose a new reconstruction method that incorporates the *partial discreteness* (PD) prior term Eq.(19). The method will be denoted by the acronym PD.

### A. Split Bregman Reconstruction Algorithm

In most of the MRI reconstruction algorithms the inequality constrained problem (3) is transformed into an unconstrained

problem of the form

$$\min_{\mathbf{x} \in \mathbb{R}_+^N} J(\mathbf{x}, \lambda) \quad (20)$$

with

$$J(\mathbf{x}, \lambda) = \Phi(\mathbf{x}) + \frac{\lambda}{2} \|\tilde{\mathbf{A}}\mathbf{x} - \mathbf{y}\|_2^2. \quad (21)$$

Solving this new unconstrained problem is equivalent to solving (3) if and only if  $\lambda$  is selected according to the Karush-Kuhn-Tucker (KKT) conditions [2]. Otherwise, the unconstrained solution, that is, the solution of (20), may not meet the data fidelity condition. Within the KKT approach, the optimal  $\lambda$  is called the KKT multiplier [47]. Unfortunately, the analytic determination of the KKT multiplier is rather difficult or, in most of the cases, even impossible [2]. As an alternative to the KKT technique, several iterative optimization algorithms replace the original problem (3) by a sequence of unconstrained minimization problems (as (20)) where the cost function  $J(\cdot, \lambda)$  is augmented to account for the constraints [48]. These subproblems are iteratively solved in combination with an update of some of the parameters included now in the augmented cost function. This class of algorithms includes penalty-based methods [47], Augmented Lagrangian (AL) methods [49], [50], and Split Bregman methods [51]–[53]. Ideally, the sequence of solutions of each subproblem asymptotically approaches the original KKT solution, i.e., it solves problem (3). It should be noted that *partial discreteness* is not limited to a specific optimization algorithm. In this work, to illustrate its potential, the Split Bregman method is chosen for the following practical reasons: it is numerically more stable than penalty-based methods [53] and it is simpler than AL methods. With the Split Bregman method,  $J(\cdot, \lambda)$  is modified with the so-called Bregman distance [54]. We refer the reader to [53] for a more detailed explanation. After some algebra, it can be demonstrated [53] that this method adopts the following simplified recursive scheme:

$$\mathbf{x}^{(t+1)} = \arg \min_{\mathbf{x} \in \mathbb{R}_+^N} \Phi(\mathbf{x}) + \frac{\lambda}{2} \|\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b}^{(t)}\|_2^2 \quad \text{and} \quad (22)$$

$$\mathbf{b}^{(t+1)} = \mathbf{b}^{(t)} + \mathbf{y} - \tilde{\mathbf{A}}\mathbf{x}^{(t+1)}. \quad (23)$$

The parameters update is done in Eq. (23), through the modified data vector  $\mathbf{b}$  [53]. Each of the unconstrained minimization problems (Eq. (22)) is solved with a modified version of the Majorize-Minimize (MM)-based algorithm proposed by Muckley et al. [55], with the non-linear conjugate gradient method. Exact convergence properties for these subproblems can be demonstrated [55]. Further details on the MM-based algorithm, details on how to impose the real positiveness constraint, as well as the analytical derivation of the gradient of  $\Phi(\mathbf{x})$  are provided in the supplementary file which comes with this paper.

## B. GMM Learning

Estimating the GMM parameters,  $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\theta}}_{\mathcal{A}}, \hat{\boldsymbol{\theta}}_{\bar{\mathcal{A}}}]$ , is often called GMM training or learning. To learn the GMM, a set of samples drawn from the GMM, and a learning criterion are required.

**1) Training Data:** If a dataset of fully-sampled reconstructed partially discrete images from the same object is a priori available, it can be used to train the GMM. However, in this work, we focus on the automatic application of the algorithm, and the GMM is learned in situ from an image reconstructed from under-sampled k-space data. Specifically, in our experiments, the GMM is trained with the magnitude of a low-resolution image  $\mathbf{x}_{\text{LR}}$  calculated at the beginning of the algorithm. Such image is obtained as follows: an  $N_{\text{Hamm}}$ -point symmetric Hamming window is applied to the under-sampled k-space data  $\mathbf{y}$  and then, an inverse Fourier transform of the windowed k-space data is calculated. Additionally, the GMM can be retrained every  $T_{\text{GMM}}$  iterations in order to keep track of the variations in the GMM along the reconstruction process.

**2) GMM Learning Criterion:** To train the GMM described by Eq. (10), we use the algorithm proposed by Figueiredo and Jain [32]. In this algorithm, the Minimum Message Length criterion is implemented. It estimates the parameters of each component as well as the optimal number of components (i.e., the total number of classes) of the GMM. Another benefit of this learning algorithm is that a careful initialization is not required, as opposed to ML expectation-maximization based methods [32].

**3) The Selection of  $K$ :** Once the GMM described by Eq. (10) is learned, the number of homogeneous regions  $K$  should be selected and the corresponding conditional PDFs for the homogeneous regions,  $\{f_{X|\mathcal{A}_k}(x; \hat{\eta}_k, \hat{\sigma}_k)\}_{k=1}^K$ , should be detected. In our work, we advocate for a manual selection of  $K$ , prior to the application of the PD algorithm. In this approach,  $K$  is selected by visual inspection of the image obtained by basic zero-filled (ZF) reconstruction. For most images we investigated, at least two homogeneous regions can be discerned: a background and a hyper-intense region. The corresponding conditional PDFs are easily identified by selecting those with the lowest and highest estimated mean, respectively. The remaining conditional PDFs are arranged in ascending order with respect to their standard deviation. Selecting  $K \geq 3$  homogeneous regions may enhance the performance of the PD if the partially discrete image at hand is indeed constituted of more homogeneous regions than just the background and a hyper-intense region. This was illustrated in a simulation experiment that was set up to evaluate the sensitivity of the PD method to the selection of  $K$  and which is described in subsection II-A of the supplementary file. If  $K = 3$ , in addition to the conditional PDFs associated to the background and hyper-intense region, the conditional PDF with the lowest standard deviation is chosen. If  $K = 4$ , also the second conditional PDF is selected and so on. In the experiments that we performed to compare PD with state of the art reconstruction methods, we consistently set  $K$  equal to 2, which can be considered as a conservative choice.

## C. Parameters Selection

For convex reconstruction problems with noiseless under-sampled data  $\mathbf{y}$ , independent of the  $\lambda$  selection in Eq. (22), the iterative solutions of the Split Bregman method asymptotically satisfy the data fidelity condition ( $\tilde{\mathbf{A}}\mathbf{x} = \mathbf{y}$ ) and monotonically

```

1:  $\hat{\psi}_{\mathbf{x}} \leftarrow \text{Phase-estimation}\{\mathbf{x}_{\text{LR}}\}$ 
2:  $\hat{\Psi} = \text{diag}(e^{i\hat{\psi}_{\mathbf{x}}})$ 
3:  $\hat{\mathbf{A}} = \mathbf{A}\hat{\Psi}$ 
4:  $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\theta}}_{\mathcal{A}}, \hat{\boldsymbol{\theta}}_{\bar{\mathcal{A}}}] \leftarrow \text{GMM-learning}\{|\mathbf{x}_{\text{LR}}|\}$ 
5: Define  $\mathcal{P}(\cdot)$  with  $\hat{\boldsymbol{\theta}}$  as in subsection III-B
6: set  $t = 0$ ,  $\mathbf{x}^{(0)} = |\mathbf{x}_{\text{LR}}|$ ,  $\mathbf{b}^{(0)} = \mathbf{y}$ 
7: while  $t < T_{\text{max}}$  and  $\|\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\|_2 \geq \text{Tol}$  do
8:    $\mathbf{x}^{(t+1)} = \arg \min_{\mathbf{x} \in \mathbb{R}_+^N} \Phi(\mathbf{x}) + \frac{\lambda}{2} \|\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b}^{(t)}\|_2^2$ 
9:    $\mathbf{b}^{(t+1)} = \mathbf{b}^{(t)} + \mathbf{y} - \mathbf{A}\mathbf{x}^{(t+1)}$ 
10:  if  $t = vT_{\text{GMM}}$  for any  $v \in \mathbb{N}^+$  then
11:     $\boldsymbol{\theta} = [\boldsymbol{\theta}_{\mathcal{A}}, \boldsymbol{\theta}_{\bar{\mathcal{A}}}] \leftarrow \text{GMM-learning}\{\mathbf{x}^{(t+1)}\}$ 
12:    Define  $\mathcal{P}(\cdot)$  with  $\boldsymbol{\theta}$  as in subsection III-B
13:  end if
14:   $t \leftarrow t + 1$ 
15: end while
16:  $\hat{\mathbf{x}} = \mathbf{x}^{(t)}$ 

```

Fig. 2. Pseudo-code of the PD algorithm.

decrease the prior term [53]. For non-convex prior terms, such as *partial discreteness*, and with noisy data  $\mathbf{y}$ , as in problem (3), global convergence cannot be guaranteed. Fortunately, in this situation, the Split Bregman algorithm has been experimentally observed to converge, even though theoretical proof is still lacking [56]–[60].

To achieve a good performance, a careful initialization  $\mathbf{x}^{(0)}$  and an adequate selection of  $\lambda$  are of great importance. Our choice for  $\mathbf{x}^{(0)}$  is the magnitude of the low-resolution reconstructed image  $\mathbf{x}_{\text{LR}}$ . This image was used as well for training the GMM. Our magnitude image  $|\mathbf{x}_{\text{LR}}|$  lacks details and texture is hardly preserved, but it has important advantages for *partial discreteness*: 1) artefacts are not so strongly manifested as in other  $\mathbf{x}^{(0)}$  choices (e.g., Tikhonov regularization on  $\mathbf{x}$ ), and 2) the background area is easily discernible. Consequently, the initial *partial discreteness representation*,  $\mathcal{P}(|\mathbf{x}_{\text{LR}}|)$ , does not undesirably magnify artefacts, and more important, the background is already accurately represented by  $\mathcal{P}(|\mathbf{x}_{\text{LR}}|)$ . The low resolution image  $\mathbf{x}_{\text{LR}}$  also serves to estimate the phase  $\psi_{\mathbf{x}}$ . From this image, the principal complex argument is voxel-wise calculated [4]. With the estimated phase image,  $\hat{\psi}_{\mathbf{x}} \in (-\pi, \pi]^N$ ,  $\tilde{\mathbf{A}}$  is defined. The basis for defining the weighting  $\lambda$  parameter is the following simple rule: the better the type of image we are reconstructing adheres to a partially discrete image, the lower  $\lambda$  should be. Furthermore, the design of  $\lambda$  should take into account that the prior term  $\Phi(\cdot)$  is often several orders of magnitude larger than the data fidelity  $l_2$  norm. Empirically, we have corroborated that a satisfactory formula in our experiments was the following:

$$\lambda = 2(1 - r) \cdot 10^3, \quad (24)$$

with  $0 \leq r < 1$  a value which we term the *partial discreteness degree*. In practice, the more constant we expect the hyperintense regions to be, the closer  $r$  should be to 1. The closer  $r$  is to 0, the less relevant the *partial discreteness* prior becomes.

The pseudo code of the proposed PD algorithm is presented in Fig. 2.

#### D. Multi-Coil Extension

The proposed method can be extended to be applicable to multi-coil acquisitions. In that case, the algebraic linear model (Eq. (1)) should be extended. Let  $\mathbf{y}_r \in \mathbb{C}^M$  be the k-space data acquired by the  $r$ -th coil, with  $r = 1, \dots, R$ . The relation with the reconstructed image  $\mathbf{x}_r$  is again

$$\mathbf{y}_r = \mathbf{A}\mathbf{x}_r + \mathbf{n}_r. \quad (25)$$

Each of  $\mathbf{x}_r \in \mathbb{C}^N$  is related to the true magnitude partially discrete image  $\mathbf{x} \in \mathbb{R}_+^N$  through the coil sensitivities,  $\mathbf{c}_r \in \mathbb{C}^N$ , as  $\mathbf{x}_r = \mathbf{C}_r \Psi \mathbf{x}$  with  $\mathbf{C}_r = \text{diag}(\mathbf{c}_r)$ . If we call  $\mathbf{y}^T = [\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_R^T]$  and  $\tilde{\mathbf{A}} = \mathbf{A}_{\text{Block}} \Psi$  with

$$\mathbf{A}_{\text{Block}} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \vdots \\ \mathbf{C}_R \end{bmatrix}, \quad (26)$$

*partial discreteness* can be applied as in subsection IV-A.

#### E. Relation to Non-Local Total Variation (NLTV)

Because partially discrete images are composed of several homogeneous regions with very low intensity variation, it may be tempting to reconstruct them with spatially-adapted versions of TV-based methods, such as Non-Local (NL) TV algorithms [61], [62], [16]. Here, we provide an insightful comparison between NLTV and our *partial discreteness* prior. While NLTV accounts for the low intensity variation using concepts rooted in calculus of variations and measure theory [61], *partial discreteness* uses tools from Bayesian inference and unsupervised learning and clustering.

Aside from being different in nature, both approaches handle prior information very differently. On one hand, *partial discreteness* promotes quasi-constant images in an explicit fashion, through the *partial discreteness representation*, i.e., an image. On the other hand, NLTV does it indirectly, through the non-local gradient [61], [63]. Indeed, in NLTV algorithms, the spatial gradient of  $\mathbf{x}$ , included in the original TV measure, is replaced by the non-local gradient of  $\mathbf{x}$ : a vector  $\nabla_m^w \mathbf{x} \in \mathbb{R}^N$  which at pixel  $m$  is defined as [16]

$$(\nabla_m^w \mathbf{x})_n = \sqrt{w_{mn}}(x_m - x_n) \text{ with } n = 1, \dots, N, \quad (27)$$

where  $\mathbf{W} \triangleq \{w_{mn}\} \in \mathbb{R}_+^{N \times N}$  is the graph matrix. In order to define the NLTV term, first the  $l_2$  norm of each  $\nabla_m^w \mathbf{x}$  with  $m = 1, \dots, N$  is taken and then, the  $l_1$  norm of the resulting vector is calculated [16]. The graph matrix  $\mathbf{W}$  plays the role of adaptive mechanism and weights the intensity differences according to the image spatial content. A large weight  $w_{mn}$  is assigned to two similar pixels, therefore penalizing deviation in intensity. Intensity deviations of pairs of dissimilar pixels must not contribute to the NLTV term, hence, zero values are selected in this case. The interested reader is referred to [16] for a more specific interpretation. At this point, a pertinent observation can be made. NLTV still needs a learning mechanism to account for the image structure and thus to define  $\mathbf{W}$ . This means that the performance of NLTV can not be separately assessed from the learning process. In fact, since the graph matrix  $\mathbf{W}$  is the tool which most leverages NLTV



performance, it should be very carefully designed. Yet, the selection of an appropriate similarity metric for weighting the pixels is still arguable and application-dependent [61], [63], not to mention the computational complexity of exhaustive searches in patch-based similarity algorithms [63]. Fortunately, the problem with the similarity metrics is circumvented with *partial discreteness*. The learning mechanism is naturally embedded into the GMM, which carries its own sophisticated learning/clustering procedure, based on Bayesian inference. The special image features of partially discrete images are condensed in the *partial discreteness representation*, derived from the unified framework presented in subsection III-A.

We emphasize the importance of the GMM in the construction of the *partial discreteness* prior. Its employment in MRI is not new. For instance, it has been used for unsupervised segmentation [64] and bias field correction [65], [66]. On top of that, it has been exploited as prior knowledge in model-based image restoration problems, where it has shown excellent results [67]. Certainly, in the Compressed Sensing (CS) field, it has been recently proved [68] that exact signal reconstruction drawn from a GMM is achievable with a substantially less number of measurements than commonly required with other sparse recovery methods.

## V. EXPERIMENTS

In this section, we describe the experiments that were carried out to evaluate the performance of the proposed PD algorithm. First, PD was compared against state-of-the-art reconstruction methods, conducting experiments on simulated as well as real k-space data. Next, dedicated simulation experiments on a digital phantom were performed to test the sensitivity of the algorithm to (i) the pre-selected number of homogeneous regions,  $K$ , (ii) phase profiles that do not satisfy the assumption of a slowly varying phase, and (iii) intensity inhomogeneity or bias field [69], [70]. In all experiments, the starting point was a fully sampled data set which was retrospectively under-sampled. Throughout, validation was not only based on visual assessment but also on quantitative results. We employed the following measures to evaluate the reconstruction quality: the Peak Signal-to-Noise Ratio (PSNR) [1], the Quantitative Index based on Local Variance (QILV) [71], the High Frequency Error Norm (HFEN) [1] and the Feature Similarity Index Metric (FSIM) [72].

### A. Comparison of PD With State-of-the-Art Methods

Comparative experiments were conducted on simulated as well as real k-space data. Different types of under-sampling scenarios were considered, namely, structural and random patterns. In this way, we show that PD does not require any specific assumption on the type of sampling, in contrast to common CS-based reconstruction methods [73]. PD was compared against the basic ZF reconstruction as well as three state-of-the-art reconstruction methods dedicated to the reconstruction of under-sampled data, namely:

- 1) SparseMRI, proposed by Lustig [4], which implements the concept of CS including, as prior terms, TV and  $l_1$ -wavelet sparsity. The Matlab code is publicly available [74].

- 2) CS+NLTV, proposed by Lian [16], which incorporates NLTV in the CS framework.
- 3) LORAKS proposed by Haldar [18]. LORAKS assumes images to have small finite support compared to the Field Of View (FOV) and/or a slowly varying phase. The Matlab code is publicly available [75], [76].

SparseMRI was applied with the built-in parameters, except for the TV weight and the number of iterations. Those values were extensively varied until no further artefacts reduction could be achieved. LORAKS was implemented with the built-in parameters. Experiments with different settings did not provide any remarkable difference. We followed the guidelines presented in the original work [16] to implement CS+NLTV. Parameter settings were chosen according to the recipe in the original work. Regarding PD, the length of the Hamming window  $N_{\text{Hammm}}$  was set in all experiments to 50. PD was stopped with the following parameters:  $T_{\text{max}} = 8$  and  $\text{Tol} = 10^{-4}$ . For simplicity and speed, we did not retrain the GMM. Regarding the weighting parameter  $\lambda$ , the corresponding  $r$  value is mentioned. The number of discrete classes were in all experiments set to  $K = 2$ .

1) *Simulated k-Space Data*: For the simulation experiment, a  $256 \times 256$  actual magnitude brain actual magnitude brain MR image (Fig. 3(a)) was used as ground-truth. The image was acquired with an Inversion Recovery (IR) pulse sequence on a 3T Siemens scanner with 32 coils. A smoothly varying phase was simulated by Legendre polynomials up to the second degree. Legendre polynomials were used because of their demonstrated suitability for simulating slow-varying intensity profiles [77]. The resulting complex image was polluted with additive complex-valued white Gaussian noise with uncorrelated real and imaginary parts of equal standard deviation  $\sigma$ . The value of  $\sigma$  was chosen such that the signal-to-noise ratio (SNR), defined as

$$\text{SNR} = \frac{\bar{x}}{\sigma} \quad (28)$$

with  $\bar{x}$  the spatial mean of the magnitude image  $\mathbf{x}$ , was equal to 10. From the noisy, complex valued image, k-space samples on a Cartesian grid were generated using the linear model of Eq. (1). Single-coil data were mimicked with structural under-sampling: a pseudo radial sampling mask (Fig. 3(b)) with 70% of non-acquired k-space points. This procedure was repeated  $N_{\text{Runs}} = 20$  times with different noise realizations. We chose a moderate value of the *partial discreteness degree*,  $r$ , i.e.,  $r = 0.1$ , which was observed to give reconstructed images with good details preservation.

The expected value of the magnitude of the reconstructed images, that is, the sample mean over the  $N_{\text{Runs}}$  realizations are shown in Fig. 3. Root-Mean-Squared Error (RMSE) maps are presented as well. Numerical results are provided in Table I.

From Fig. 3, it can be observed that SparseMRI and CS+NLTV, though successful in removing noise, do not recover a high resolution image. Their RMSE maps reveal substantial structural errors, mainly located at edges. This observation agrees with what was already pointed out in [18] for SparseMRI. As expected, edges are slightly better preserved with CS+NLTV [16]. All LORAKS versions and

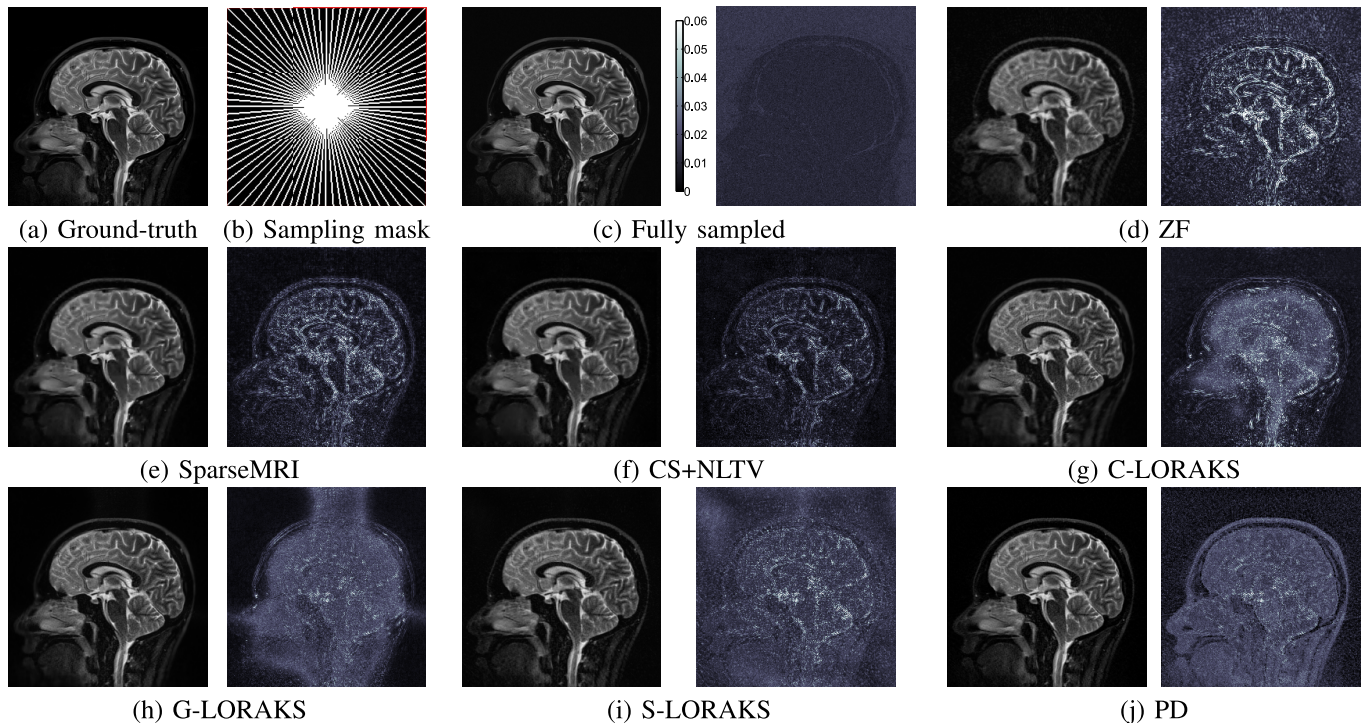


Fig. 3. Visual results for the experiment with simulated k-space data. The sample mean of the magnitude of the reconstructed images are shown in companion of the RMSE maps. To highlight small errors, the colorbar range of the RMSE maps was adapted to  $[0, T]$ , where  $T$  is 40% of the maximum value which was found in all RMSE maps (i.e., considering altogether).

TABLE I  
QUANTITATIVE RESULTS FOR THE EXPERIMENT  
WITH SIMULATED  $k$ -SPACE DATA

	PSNR [dB]	QILV	HFEN	FSIM
Fully sampled	35.3	0.998	0.070	0.990
ZF	28.7	0.747	0.235	0.903
SparseMRI	<b>34.6</b>	0.907	0.153	0.938
CS+NLTV	33.4	0.921	0.141	0.934
C-LORAKS	33.2	0.930	0.121	0.952
G-LORAKS	31.1	0.953	0.134	0.940
S-LORAKS	33.5	0.976	0.112	0.953
PD	34.5	<b>0.986</b>	<b>0.090</b>	<b>0.962</b>

specially PD restore images with substantially higher resolution than ZF, SparseMRI and CS+NLTV. Indeed, their RMSE maps exhibit a noisy pattern with a very moderate structural degradation effect. This is expected if reconstruction methods succeed in recovering high-resolution missing k-space data (see the fully sampled case).

Among all LORAKS versions, S-LORAKS seems to perform best. However, still some (small) structural details errors can be observed. These are considerably attenuated with PD though. Numerical results are in agreement with visual findings. PD obtains the best result (disregarding the fully sampled scenario), in terms of QILV, HFEN, and FSIM. PD is closely followed by S-LORAKS, while there is a significant difference compared to CS+NLTV and SparseMRI. The highest PSNR for SparseMRI might be attributed to its noise removal capability, having a relevant effect specially in the background. With an Intel Core i7-4770K 3.5 GHz (32 GB RAM) processor, the average time for LORAKS reconstruction was about 10 min. The Singular Value Decomposition (SVD) incorporated in the method is probably the reason of such a computational burden.

As already mentioned in subsection IV-E, the calculation of  $\mathbf{W}$  for the CS+NLTV method is computationally quite expensive. With the parameters chosen as those recommended in the original work [16], the average computation time for CS+NLTV was about 8 min. PD was able to reconstruct images within roughly 4 min. The main computational effort is in the GMM learning. Note that when the GMM parameters are known, each of the subproblems of the Split Bregman method can be rapidly implemented with the MM algorithm presented in the supplementary file. Clearly, the fastest algorithm is SparseMRI with a computation time sometimes below 1 min. It should be noted that all algorithms were implemented in Matlab. A C++ implementation would greatly speed up all algorithms. This holds especially for the graph matrix calculation of CS+NLTV and the GMM learning for PD.

*2) In Vivo Human Knee k-Space Data:* In this experiment, in vivo 3-D fully sampled k-space data of a human knee were employed. The k-space data, available at [78], were acquired with a 3-D Fast Spin Echo (FSE) pulse sequence on a 3T scanner with eight coils. From the 3-D k-space data, 2-D slice reconstruction was accomplished. We applied an inverse Fourier transform [18] along the third dimension, and then, one and the same mid-sagittal 2-D k-space slice was extracted per each coil. This approach is valid since the 3-D Fourier transform is a separable operator [79]. The corresponding k-space slices ( $320 \times 320$ ) were under-sampled with a random phase-encoding mask of 50% of missing k-space lines (Fig. 4(b)).

For ease of comparison, and because publicly available Matlab codes of SparseMRI and LORAKS only permit single-



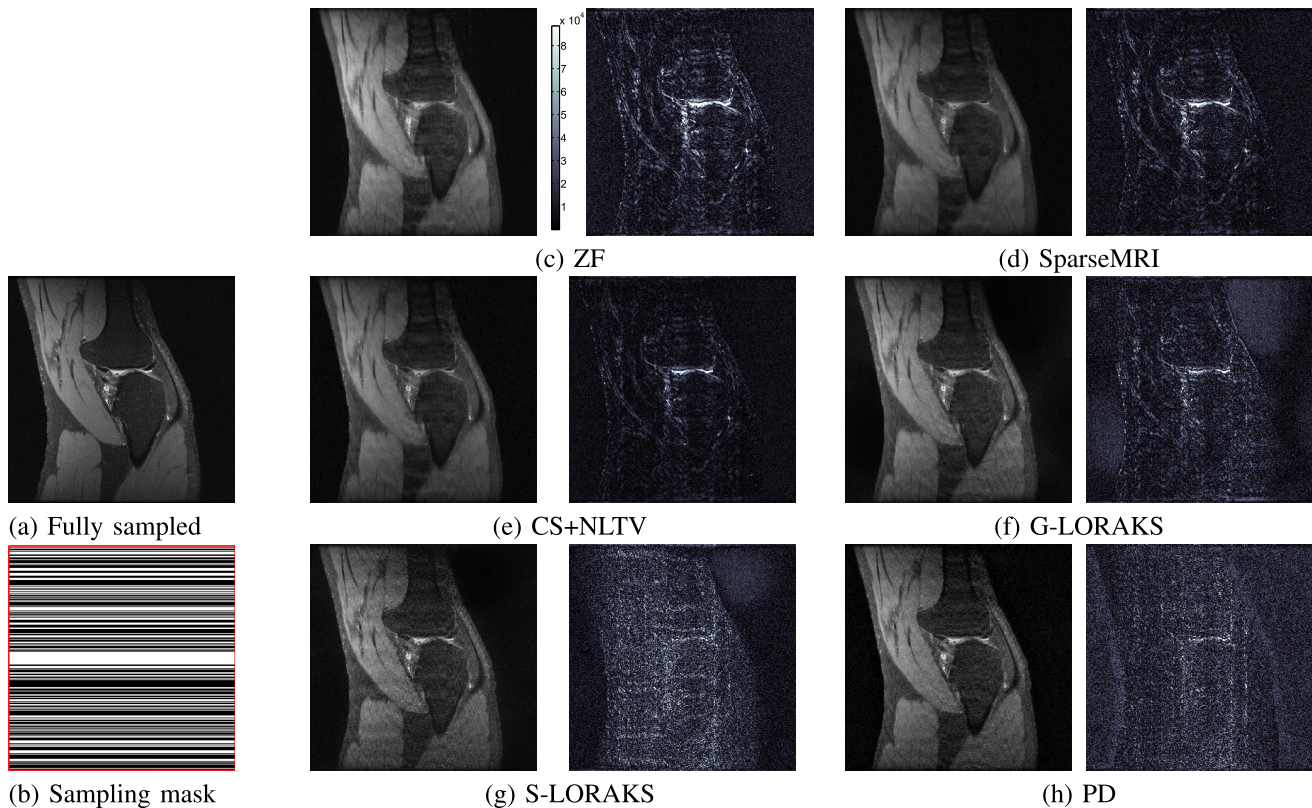


Fig. 4. Visual results for the experiment with real k-space data of the knee. The SoS reconstructed images are shown in companion of the absolute errors maps. To highlight small errors, the colorbar range of the absolute error maps was adapted to  $[0, T]$ , where  $T$  is 40% of the maximum value which was found in all absolute error maps (i.e., considering altogether).

TABLE II  
QUANTITATIVE RESULTS FOR THE EXPERIMENT WITH  
*in vivo* HUMAN KNEE *k*-SPACE DATA

	PSNR [dB]	QILV	HFEN	FSIM
ZF	29.84	0.454	$2.96 \cdot 10^5$	0.887
SparseMRI	30.29	0.620	$2.72 \cdot 10^5$	0.860
CS+NLTV	<b>30.33</b>	0.710	$2.67 \cdot 10^5$	0.888
G-LORAKS	30.15	0.798	$2.48 \cdot 10^5$	<b>0.893</b>
S-LORAKS	27.78	0.847	$2.45 \cdot 10^5$	0.837
PD	27.92	<b>0.857</b>	<b><math>2.41 \cdot 10^5</math></b>	0.883

coil reconstruction, each of the eight 2-D k-space data sets were reconstructed independently. To create a final image per method, the eight reconstructed images were combined using the Sum of Squares (SoS) method [80].

Since the FSE sequence produces a highly bright and constant area in the cartilage region, it is reasonable to assume the *partial discreteness* holds even more than in the simulation experiment. Therefore, for PD, we increased the  $r$  value to 0.9.

Reconstructed SoS images as well as the absolute error maps (with the magnitude of the fully sampled image as ground truth), are shown in Fig. 4, while quantitative results are reported in Table II.

We can draw similar conclusions about the performance of PD in comparison to the rest of the methods. The artefacts manifested due to the random sampling are partially removed by SparseMRI and CS+NLTV, but both methods failed in recovering highly detailed clinical relevant areas, for example, the contours in the cartilage region. PD restored a higher

detailed image as can be seen as well by looking at the errors map. Concerning the LORAKS versions, we first notice that C-LORAKS failed to converge (results not shown). G-LORAKS was able to recover a better defined cartilage but still it is largely outperformed by S-LORAKS and PD. Reconstructed images with S-LORAKS are slightly less accurate than those obtained by PD, as is manifest in the metrics of Table II. Certainly, PD scores best for metrics which are specially conceived to assess small details preservation, that is, QILV and HFEN. The best FSIM case for G-LORAKS may be understood if we notice that G-LORAKS provides a good balance between artefacts suppression and structural details preservation. Finally, the highest PSNR for CS+NLTV could be largely based on its ability for artefacts removal.

The computational time of all methods, for each coil, were very similar to the times reported in the simulation experiment and hence they are not repeated here.

**3) *In Vivo* Human Brain *k*-Space Data:** To finish the experiments section, we validate PD in a truly multi-coil reconstruction scheme, where coil sensitivities estimation is required. *In vivo* 3-D fully sampled k-space data were used again, this time, from a human brain. K-space data were acquired with a 3T scanner with eight coils. To allow for 2-D slice reconstruction, we followed the same routine as in the knee experiment. Specifically, we applied an inverse Fourier transform along the second dimension and then, one and the same mid-axial 2-D k-space slice was extracted per each coil.

TABLE III  
QUANTITATIVE RESULTS FOR THE EXPERIMENT  
WITH *in Vivo* HUMAN BRAIN *k*-SPACE DATA

	PSNR [dB]	QILV	HFEN	FSIM
ZF	28.76	0.547	17.99	0.888
CS+NLTV	31.06	0.844	11.70	<b>0.930</b>
PD	<b>31.50</b>	<b>0.995</b>	<b>8.340</b>	0.926

The resulting *k*-space slices ( $230 \times 180$ ) were under-sampled with a 2-D variable density random mask of 75% of missing *k*-space points (Fig. 5(b)).

Coil sensitivities,  $\{\mathbf{c}_r\}_{r=1}^{R=8}$ , were estimated with the SoS method [80]. Next, the initial low-reconstructed image  $\mathbf{x}_{\text{LR}}$  was obtained as  $\mathbf{x}_{\text{LR}} = \mathbf{A}_{\text{Block}}^H \mathbf{y}_{N_{\text{Hammm}}}$  where  $\mathbf{A}_{\text{Block}}^H$  denotes the Hermitian transpose of  $\mathbf{A}_{\text{Block}}$  and  $\mathbf{y}_{N_{\text{Hammm}}}$  are the under-sampled *k*-space data filtered with a Hamming window ( $N_{\text{Hammm}} = 50$ ). The phase  $\psi_{\mathbf{x}}$  was estimated from  $\mathbf{x}_{\text{LR}}$  as well. From  $\hat{\psi}_{\mathbf{x}}$ ,  $\Psi$  was determined after which  $\tilde{\mathbf{A}}$  was redefined as proposed in subsection IV-D. Finally, PD was applied with the same Split Bregman method. As in the brain simulation experiment,  $r = 0.1$  was selected. We also compared PD with the ZF reconstruction, calculated as  $\mathbf{A}_{\text{Block}}^H \mathbf{y}$ , and with the CS+NLTV method. In both cases, the coil estimation procedure was done in exactly the same manner as for PD.

Undoubtedly, it can be seen from Fig. 5 that the magnitude of the reconstructed image with PD possesses higher resolution than the image restored with CS+NLTV and specially ZF. A closer look at zoomed images (Fig. 5(j-l)) reveals that the interfaces between white/gray matter are better preserved with PD compared to CS+NLTV. Absolute error maps also demonstrate that higher structural errors are more widely manifested with CS+NLTV than with PD.

Quantitative results in Table III further suggest the superiority of PD in detail preservation and resolution enhancement. PD ranked best for three of the four metrics.

Nevertheless, it is also clear that NLTV outperforms PD in noise suppression. This is not surprising since the NLTV prior term, incorporated in the CS+NLTV method, has proved to be a very effective denoising mechanism [62]. Indeed, the first application of the TV measure in the image processing field was noise removal [81].

It is interesting to remark that our *partial discreteness* prior does not attempt to remove noise but to recover the unknown partially discrete image using a priori information, which we carefully modeled from its special structure. Neither the GMM nor the Bayesian probabilistic segmentation were designed to tackle noise. However, showing its flexibility, our *partial discreteness* prior can easily accommodate a simple regularization term in companion of the *partial discreteness representation* to cope with noise while still exploiting all the potential of this novel image representation. For the reader's interest, we point out that the *a posteriori* probability maps derived from the GMM have a broad range of applications for MR reconstruction, tissue-selective filtering being one of them. A reduced list of these applications and some extensions of the GMM are given at the end of the conclusion section.

## B. Sensitivity Analysis of the PD Method

This subsection summarizes the main results of a sensitivity analysis of the proposed PD method to various parameters and deviations from assumptions. The analysis is based on dedicated simulation experiments, which are extensively discussed in the supplementary file accompanying this paper.

1) *Sensitivity to the Pre-Selected Number of Homogeneous Regions, K*: To test the sensitivity to the pre-selected number of homogeneous regions  $K$ , a dedicated simulation experiment was conducted on a digital phantom image. This simulation experiment is described in subsection II-A of the supplementary file. The results of the experiment show a clear gain in reconstruction quality if at least one homogeneous region (e.g., background) is chosen. A further substantial improvement was achieved by also incorporating the hyper-intense region ( $K = 2$ ). The optimal value of  $K$  for the phantom image was found to be equal to 4, which demonstrates that for particular partially discrete images, the performance of PD can be further improved by selecting  $K$  higher than 2. Obviously, the optimal value of  $K$  will depend on the image to be reconstructed. Note that in the experiments that we performed to compare PD with state-of-the-art reconstruction methods (see subsection V-A),  $K$  was consistently set equal to 2, which can be considered as a conservative choice.

2) *Sensitivity to Non-Slowly Varying Phase*: As described in subsection IV-B, the required estimate of the image phase is obtained from a low resolution image  $\mathbf{x}_{\text{LR}}$ . The rationale for this procedure is that real-life phase images are often slowly varying. Note that a slowly varying or smooth phase is a common assumption in MR image reconstruction methods [4]. Nevertheless, phase images may also have significantly higher spatial frequency content [82], especially when gradient echo instead of spin echo imaging sequences are used [83]. To study the sensitivity of the PD algorithm to non-slowly varying phase, we carried out a simulation experiment (cfr. subsection II-B of the supplementary file), in which the performance of PD for three different phase profiles was evaluated. As expected, the performance of PD degrades for highly-varying phase profiles, which indicates the importance of the smooth phase assumption. However, it was also found that PD is robust to moderate phase variations.

3) *Sensitivity to Bias Fields*: Bias fields are undesired low frequency signals induced by inhomogeneities in the magnetic fields of the MRI system [84]. The presence of a bias field may challenge the main assumption underlying the PD method, namely that the image to be reconstructed contains regions of quasi-constant intensity. To evaluate the performance of the PD method in the presence of a bias field, dedicated simulation experiments were conducted on a digital phantom. These simulation experiments are described in subsection II-C of the supplementary file. Bias fields with different degrees of variation were considered. The results of the simulation experiments show that, in general, the performance of PD deteriorates when the degree of variation of the bias field increases. In addition, since not all homogeneous regions may be equally affected, a bias field may also influence the optimal



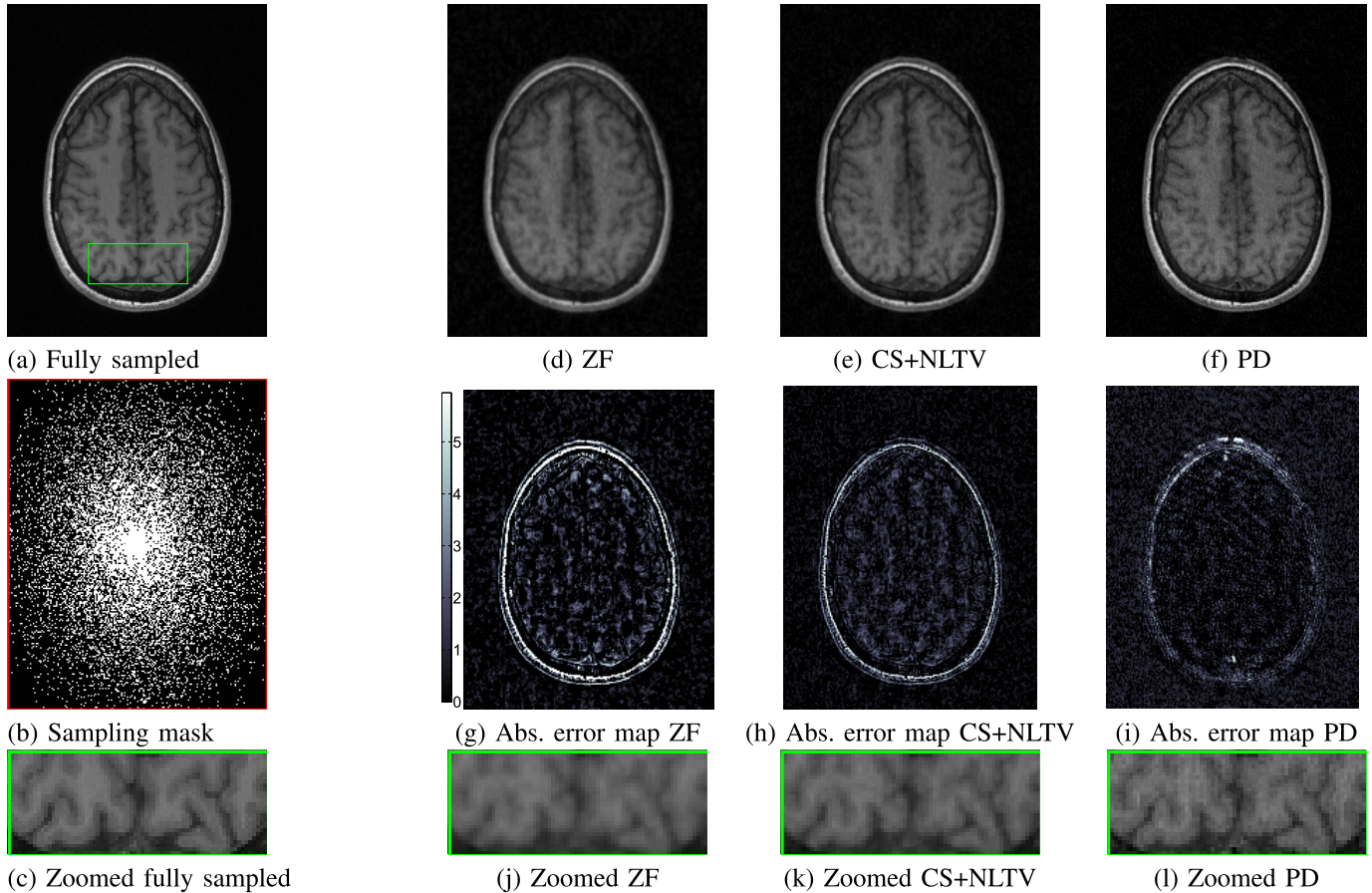


Fig. 5. Visual results for the experiment with real k-space data of the brain. The magnitude of the multi-coil reconstructed images are shown in companion of the absolute error maps. To highlight small errors, the colorbar range of the absolute error maps was adapted to  $[0, T]$ , where  $T$  is 40% of the maximum value which was found in all absolute error maps (i.e., considering altogether).

choice of  $K$ , suggesting that in the presence of a severe bias field, a conservative choice of  $K$  is advisable, as is further motivated in subsection II-C of the supplementary file. In this supplementary file, we also elaborate on the possibility of improving the robustness of the PD method by including a bias field correction technique.

## VI. CONCLUSIONS

In this work, we have presented a novel prior, *partial discreteness*, for the reconstruction of MR images with quasi-constant intensity regions as well as heterogeneous regions. We have shown that every image can be additively decomposed into its *partial discreteness representation* and its residual form. The *partial discreteness representation*, which is based on a GMM, embodies the basic features of partially discrete images: constant intensity in homogeneous regions and texture in heterogeneous regions. Exploiting this *partial discreteness representation* in MR image reconstruction, by enforcing sparsity on the residual form, we have been able to reconstruct highly detailed images from under-sampled data with structural and random under-sampling schemes, namely, pseudo-radial, random phase-encoding and pseudo-random variable density sampling. In this work, *partial discreteness*

has been implemented in a phase-constrained formulation where the phase map was estimated from a low-resolution image. Hence, we have implicitly made the common assumption of smoothly varying phase images [4], [18]. Furthermore, it seems that no special assumptions on the type of sampling pattern seem to be required for *partial discreteness*. Experiments performed on both simulated and real k-space data have shown that the newly proposed reconstruction method PD performs competitively with and often better than state-of-the-art reconstruction methods such as SparseMRI, LORAKS and CS+NLTV. The results suggest that PD allows better texture preservation than SparseMRI (CS with TV prior), avoiding the staircasing effect, and even CS+NLTV. This is because in the *partial discreteness representation*, which is the core of PD, edges are modeled not solely based on intensity or its gradient but merely based on Bayesian (*a posteriori*) probabilities of GMM classes. Enforcing sparsity on the residual form instead of the (non-local) gradient, as (non-local) TV promotes, is less restrictive. Furthermore, PD outperforms LORAKS in terms of computation time, while providing images as highly detailed as LORAKS does.

In future work, we will explore the extension of the univariate GMM within the PD method to a multivariate GMM. The use of a multivariate GMM allows the joint reconstruction



of a set of images, thereby exploiting the correlation between those images. Such an approach may be useful in, for example, accelerated dynamic MRI [85], diffusion-weighted MRI, 3D imaging or simultaneous multi-slice reconstruction [86].

Finally, we note that the probabilistic image presentation in PD inherently leads to tissue classification embedded in a Bayesian framework. *A posteriori* probability maps derived from the GMM may be used for tissue segmentation [64], texture analysis or tissue-selective filtering schemes [87], [88].

### ACKNOWLEDGMENTS

The authors would like to thank Dr. Michael Lustig and Dr. Shreyas Vasanawala for sharing in vivo human k-space data, Prof. Annie Cuyt for sharing her expert knowledge on numerical analysis and low-rank modeling, as well as the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

### REFERENCES

- [1] S. Ravishanker and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1028–1041, May 2011.
- [2] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic  $\ell_0$ -minimization," *IEEE Trans. Med. Imag.*, vol. 28, no. 1, pp. 106–121, Jan. 2009.
- [3] M. A. Bernstein, K. F. King, and X. J. Zhou, *Handbook of MRI Pulse Sequences*. Burlington, MA, USA: Elsevier, 2004.
- [4] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.
- [5] J. A. Fessler, "Model-based image reconstruction for MRI," *IEEE Signal Process. Mag.*, vol. 27, no. 4, pp. 81–89, Jul. 2010.
- [6] L. Ying, D. Xu, and Z.-P. Liang, "On Tikhonov regularization for image reconstruction in parallel MRI," in *Proc. 26th Conf. IEEE Eng. Med. Biol. Soc.*, vol. 1, Sep. 2004, pp. 1056–1059.
- [7] F. Knoll, C. Clason, K. Bredies, M. Uecker, and R. Stollberger, "Parallel imaging with nonlinear reconstruction using variational penalties," *Magn. Reson. Med.*, vol. 67, no. 1, pp. 34–41, Jan. 2012.
- [8] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1042–1054, May 2011.
- [9] C. Chen and J. Huang, "Compressive sensing MRI with wavelet tree sparsity," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1115–1123.
- [10] Z. Lai *et al.*, "Image reconstruction of compressed sensing MRI using graph-based redundant wavelet transform," *Med. Image Anal.*, vol. 27, pp. 93–104, Jan. 2016.
- [11] E. Candès, L. Demanet, D. Donoho, and X. Ying, "Fast discrete curvelet transforms," *Multiscale Model. Simul.*, vol. 5, no. 3, pp. 861–899, Sep. 2006.
- [12] J. Aelterman, H. Q. Luong, B. Goossens, A. Pižurica, and W. Philips, "Augmented Lagrangian based reconstruction of non-uniformly sub-Nyquist sampled MRI data," *Signal Process.*, vol. 91, no. 12, pp. 2731–2742, Dec. 2011.
- [13] J. Caballero, A. N. Price, D. Rueckert, and J. V. Hajnal, "Dictionary learning and time sparsity for dynamic MR data reconstruction," *IEEE Trans. Med. Imag.*, vol. 33, no. 4, pp. 979–994, Apr. 2014.
- [14] H. Segers, W. J. Palenstijn, K. J. Batenburg, and J. Sijbers, "Discrete tomography in MRI: A simulation study," *Fundam. Inform.*, vol. 125, nos. 3–4, pp. 223–237, 2013.
- [15] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint," *Magn. Reson. Med.*, vol. 57, pp. 1086–1098, Jun. 2007.
- [16] D. Liang, H. Wang, Y. Chang, and L. Ying, "Sensitivity encoding reconstruction with nonlocal total variation regularization," *Magn. Reson. Med.*, vol. 65, no. 5, pp. 1384–1392, May 2011.
- [17] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, "Second order total generalized variation (TGV) for MRI," *Magn. Reson. Med.*, vol. 65, no. 2, pp. 480–491, Feb. 2011.
- [18] J. P. Haldar, "Low-rank modeling of local k-space neighborhoods (LORAKS) for constrained MRI," *IEEE Trans. Med. Imag.*, vol. 33, no. 3, pp. 668–681, Mar. 2014.
- [19] J. P. Haldar and J. Zhuo, "P-LORAKS: Low-rank modeling of local k-space neighborhoods with parallel imaging data," *Magn. Reson. Med.*, vol. 75, no. 4, pp. 1499–1514, Apr. 2015.
- [20] B. M. Kelm, B. H. Menze, O. Nix, C. M. Zechmann, and F. A. Hamprecht, "Estimating kinetic parameter maps from dynamic contrast-enhanced MRI using spatial prior knowledge," *IEEE Trans. Med. Imag.*, vol. 28, no. 10, pp. 1534–1547, Oct. 2009.
- [21] J. V. Velikina and A. A. Samsonov, "Reconstruction of dynamic image series from undersampled MRI data using data-driven model consistency condition (MOCCO)," *Magn. Reson. Med.*, vol. 74, no. 5, pp. 1279–1290, Nov. 2014.
- [22] J. P. Haldar, D. Hernandez, S.-K. Song, and Z.-P. Liang, "Anatomically constrained reconstruction from noisy data," *Magn. Reson. Med.*, vol. 59, no. 4, pp. 810–818, Apr. 2008.
- [23] G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal, "Bayesian reconstruction of functional images using anatomical information as priors," *IEEE Trans. Med. Imag.*, vol. 12, no. 4, pp. 670–680, Dec. 1993.
- [24] Y. Cao and D. N. Levin, "Using an image database to constrain the acquisition and reconstruction of MR images of the human head," *IEEE Trans. Med. Imag.*, vol. 14, no. 2, pp. 350–361, Jun. 1995.
- [25] J. M. Hanson, Z.-P. Liang, E. C. Wiener, and P. C. Lauterbur, "Fast dynamic imaging using two reference images," *Magn. Reson. Med.*, vol. 36, no. 1, pp. 172–175, Jul. 1996.
- [26] K. J. Batenburg and J. Sijbers, "DART: A practical reconstruction algorithm for discrete tomography," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2542–2553, Sep. 2011.
- [27] A. Dabrovolski, K. J. Batenburg, and J. Sijbers, "A multiresolution approach to discrete tomography using DART," *PLoS ONE*, vol. 9, no. 9, p. e106090, Sep. 2014.
- [28] K. J. Batenburg *et al.*, "3D imaging of nanomaterials by discrete tomography," *Ultramicroscopy*, vol. 109, no. 6, pp. 730–740, May 2009.
- [29] T. Roelands, K. J. Batenburg, E. Biermans, C. Kübel, S. Bals, and J. Sijbers, "Accurate segmentation of dense nanoparticles by partially discrete electron tomography," *Ultramicroscopy*, vol. 114, pp. 96–105, Mar. 2012.
- [30] S. Juanpere, E. Perez, O. Huc, N. Motos, J. Pont, and S. Pedraza, "Imaging of breast implants—A pictorial review," *Insights Imag.*, vol. 2, no. 6, pp. 653–670, Dec. 2011.
- [31] M. T. Preece, A. G. Osborn, S. S. Chin, and J. G. Smirniotopoulos, "Intracranial neurenteric cysts: Imaging and pathology spectrum," *Amer. J. Neuroradiol.*, vol. 27, pp. 1211–1216, Jun./Jul. 2006.
- [32] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.
- [33] G. McGibney, M. R. Smith, S. T. Nichols, and A. Crawley, "Quantitative evaluation of several partial Fourier reconstruction algorithms used in MRI," *Magn. Reson. Med.*, vol. 30, no. 1, pp. 51–59, Jul. 1993.
- [34] A. A. Samsonov, E. G. Kholmovski, D. L. Parker, and C. R. Johnson, "POCSense: POCS-based reconstruction for sensitivity encoded magnetic resonance imaging," *Magn. Reson. Med.*, vol. 52, no. 6, pp. 1397–1406, Dec. 2004.
- [35] M. Bydder and M. D. Robson, "Partial Fourier partially parallel imaging," *Magn. Reson. Med.*, vol. 53, no. 6, pp. 1393–1401, Jun. 2005.
- [36] A. A. Samsonov, J. Velikina, Y. Jung, E. G. Kholmovski, C. R. Johnson, and W. F. Block, "POCS-enhanced correction of motion artifacts in parallel MRI," *Magn. Reson. Med.*, vol. 63, no. 4, pp. 1104–1110, Apr. 2010.
- [37] M. Blaimer, M. Heim, D. Neumann, P. M. Jakob, S. Kannengiesser, and F. A. Breuer, "Comparison of phase-constrained parallel MRI approaches: Analogies and differences," *Magn. Reson. Med.*, vol. 75, no. 3, pp. 1086–1099, Mar. 2016.
- [38] B. P. Sutton, D. C. Noll, and J. A. Fessler, "Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 178–188, Feb. 2003.
- [39] A. J. den Dekker and J. Sijbers, "Data distributions in magnetic resonance images: A review," *Phys. Med.*, vol. 30, no. 7, pp. 725–741, Nov. 2014.
- [40] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.
- [41] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008.

- [42] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. New York, NY, USA: McGraw-Hill, 2002.
- [43] S. C. Zhu, Y. Wu, and D. Mumford, "Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling," *Int. J. Comput. Vis.*, vol. 27, no. 2, pp. 107–126, Mar. 1998.
- [44] J. Huang and D. Mumford, "Statistics of natural images and models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 1999, pp. 541–547.
- [45] S. P. Awate, T. Tasdizen, N. Foster, and R. T. Whitaker, "Adaptive Markov modeling for mutual-information-based, unsupervised MRI brain-tissue classification," *Med. Image Anal.*, vol. 10, no. 5, pp. 726–739, Oct. 2006.
- [46] H. Permuter, J. Francos, and I. H. Jermyn, "Gaussian mixture models of texture and colour for image database retrieval," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 3, Apr. 2003, pp. III-569–III-572.
- [47] D. P. Bertsekas, "Necessary and sufficient conditions for a penalty method to be exact," *Math. Program.*, vol. 9, no. 1, pp. 87–99, Dec. 1975.
- [48] J. Nocedal and S. Wright, *Numerical Optimization*. New York, NJ, USA: Springer-Verlag, 2006.
- [49] C. Wu and X.-C. Tai, "Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 300–339, 2010.
- [50] E. J. G. Birgin and J. M. Martínez, "Augmented Lagrangian method with nonmonotone penalty parameters for constrained optimization," *Comput. Optim. Appl.*, vol. 51, no. 3, pp. 941–965, Apr. 2012.
- [51] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 460–489, Jul. 2005.
- [52] J.-F. Cai, S. Osher, and Z. Shen, "Split Bregman methods and frame based image restoration," *Multiscale Model. Simul.*, vol. 8, no. 2, pp. 337–369, Dec. 2009.
- [53] T. Goldstein and S. Osher, "The Split Bregman method for L1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, Apr. 2009.
- [54] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Comput. Math. Math. Phys.*, vol. 7, no. 3, pp. 200–217, 1967.
- [55] M. J. Muckley, D. C. Noll, and J. A. Fessler, "Fast parallel MR image reconstruction via B1-based, adaptive restart, iterative soft thresholding algorithms (BARISTA)," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 578–588, Feb. 2015.
- [56] Q. Liu, S. Wang, K. Yang, J. Luo, Y. Zhu, and D. Liang, "Highly undersampled magnetic resonance image reconstruction using two-level Bregman method with dictionary updating," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1290–1301, Jul. 2013.
- [57] R. Chartrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Jun./Jul. 2009, pp. 262–265.
- [58] F. Li, T. Zeng, and G. Zhang, "Lagrangian multipliers and split Bregman methods for minimization problems constrained on  $s^{n-1}$ ," *J. Vis. Commun. Image Represent.*, vol. 23, no. 7, pp. 1041–1050, Oct. 2012.
- [59] F. Wang, Z. Xu, and H.-K. Xu, (2014). "Convergence of Bregman alternating direction method with multipliers for nonconvex composite problems." [Online]. Available: <https://arxiv.org/abs/1410.8625>
- [60] V. Estellers, D. Zosso, R. Lai, S. Osher, J. P. Thiran, and X. Bresson, "Efficient algorithm for level set method preserving distance function," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4722–4734, Dec. 2012.
- [61] G. Gilboa and S. Osher, "Nonlocal linear image regularization and supervised segmentation," *Multiscale Model. Simul.*, vol. 6, no. 2, pp. 595–630, Jul. 2007.
- [62] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Model. Simul.*, vol. 7, no. 3, pp. 1005–1028, Nov. 2008.
- [63] A. Sawatzky, "(Nonlocal) total variation in medical imaging," Ph.D. dissertation, Dept. Math., Western Washington Univ., Münster, Germany, 2011.
- [64] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain," *IEEE Trans. Med. Imag.*, vol. 18, no. 10, pp. 897–908, Oct. 1999.
- [65] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based bias field correction of MR images of the brain," *IEEE Trans. Med. Imag.*, vol. 18, no. 10, pp. 885–896, Oct. 1999.
- [66] Z. Mai, R. Hanel, K. J. Batenburg, M. Verhoye, P. Scheunders, and J. Sijbers, "Bias field reduction by localized Lloyd-Max quantization," *Magn. Reson. Imag.*, vol. 29, no. 4, pp. 536–545, 2011.
- [67] J. Yang *et al.*, "Compressive sensing by learning a Gaussian mixture model from measurements," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 106–119, Jan. 2015.
- [68] F. Renna, R. Calderbank, L. Carin, and M. R. D. Rodrigues, "Reconstruction of signals drawn from a Gaussian mixture via noisy compressive measurements," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2265–2277, May 2014.
- [69] A. Fan *et al.*, "A unified variational approach to denoising and bias correction in MR," in *Information Processing in Medical Imaging*, vol. 2732. Springer, 2003, pp. 148–159.
- [70] U. Vovk, F. Pernus, and B. Likar, "A review of methods for correction of intensity inhomogeneity in MRI," *IEEE Trans. Med. Imag.*, vol. 26, no. 3, pp. 405–421, Mar. 2007.
- [71] S. Aja-Fernández, R. S. J. Estépar, C. Alberola-López, and C.-F. Westin, "Image quality assessment based on local variance," in *Proc. 28th Conf. IEEE Eng. Med. Biol. Soc.*, Aug./Sep. 2006, pp. 4815–4818.
- [72] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [73] J. P. Haldar, D. Hernando, and Z.-P. Liang, "Compressed-sensing MRI with random encoding," *IEEE Trans. Med. Imag.*, vol. 30, no. 4, pp. 893–903, Apr. 2011.
- [74] M. Lustig, *Sparse MRI*. [Online]. Available: <http://www.eecs.berkeley.edu/~mlustig/Software.html>
- [75] J. P. Haldar, *LORAKS: Implementation and Examples*. [Online]. Available: <http://mr.usc.edu/download/LORAKS>
- [76] J. P. Haldar, "Low-rank modeling of local k-space neighborhoods (LORAKS): Implementation and examples for reproducible research," Univ. Southern California, Los Angeles, CA, Tech. Rep. USC-SIPI-414, 2014.
- [77] M. Styner, C. Brechbühler, G. Szckely, and G. Gerig, "Parametric estimate of intensity inhomogeneities applied to MRI," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 153–165, Mar. 2000.
- [78] M. Lustig and S. Vasanawala, *MR Data for Compressed Sensing*. [Online]. Available: <http://mridata.org/>
- [79] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.
- [80] E. G. Larsson, D. Erdogmus, R. Yan, J. C. Principe, and J. R. Fitzsimmons, "SNR-optimality of sum-of-squares reconstruction for phased-array magnetic resonance imaging," *J. Magn. Reson.*, vol. 163, no. 1, pp. 121–123, Jul. 2003.
- [81] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.
- [82] F. Zhao, D. C. Noll, J. F. Nielsen, and J. A. Fessler, "Separate magnitude and phase regularization via compressed sensing," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1713–1723, Sep. 2012.
- [83] W. Feng, J. Neelavalli, and E. M. Haacke, "Catalytic multiecho phase unwrapping scheme (CAMPUS) in multiecho gradient echo imaging: Removing phase wraps on a voxel-by-voxel basis," *Magn. Reson. Med.*, vol. 70, no. 1, pp. 117–126, Jul. 2013.
- [84] J. Juntu, J. Sijbers, D. Van Dyck, and J. Gielen, *Bias Field Correction for MRI Images*. Berlin, Germany: Springer, 2005, pp. 543–551.
- [85] R. Otazo, E. Candès, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magn. Reson. Med.*, vol. 73, no. 3, pp. 1125–1136, Mar. 2015.
- [86] K. Setsompop, A. B. Gagoski, J. R. Polimeni, T. Witzel, V. J. Wedeen, and L. L. Wald, "Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty," *Magn. Reson. Med.*, vol. 67, no. 5, pp. 1210–1224, May 2012.
- [87] G. Vegas-Sánchez-Ferrero, S. Aja-Fernández, M. Martín-Fernández, A. F. Frangi, and C. Palencia, "Probabilistic-driven oriented speckle reducing anisotropic diffusion with application to cardiac ultrasonic images," in *Proc. 13th Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2010, pp. 518–525.
- [88] G. Ramos-Llordén, G. Vegas-Sánchez-Ferrero, M. Martín-Fernández, C. Alberola-López, and S. Aja-Fernández, "Anisotropic diffusion filter with memory based on speckle statistics for ultrasound images," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 345–358, Jan. 2015.