

Dictionary Learning and Time Sparsity for Dynamic MR Data Reconstruction

Jose Caballero*, Anthony N. Price, Daniel Rueckert, and Joseph V. Hajnal

Abstract—The reconstruction of dynamic magnetic resonance data from an undersampled k-space has been shown to have a huge potential in accelerating the acquisition process of this imaging modality. With the introduction of compressed sensing (CS) theory, solutions for undersampled data have arisen which reconstruct images consistent with the acquired samples and compliant with a sparsity model in some transform domain. Fixed basis transforms have been extensively used as sparsifying transforms in the past, but recent developments in dictionary learning (DL) have been shown to outperform them by training an over-complete basis that is optimal for a particular dataset. We present here an iterative algorithm that enables the application of DL for the reconstruction of cardiac cine data with Cartesian undersampling. This is achieved with local processing of spatio-temporal 3D patches and by independent treatment of the real and imaginary parts of the dataset. The enforcement of temporal gradients is also proposed as an additional constraint that can greatly accelerate the convergence rate and improve the reconstruction for high acceleration rates. The method is compared to and shown to systematically outperform k-t FOCUSS, a successful CS method that uses a fixed basis transform.

Index Terms—Compressed sensing (CS), dictionary learning, dynamic magnetic resonance imaging, image reconstruction, sparse coding.

I. INTRODUCTION

DYNAMIC magnetic resonance imaging (MRI) allows the visualization and analysis of anatomical and functional changes of internal body structures through time. This imaging modality has become a key tool in medical research and diagnosis mainly because it is noninvasive, nonionising and provides a unique quality of soft tissue contrast. The reconstruction of a sequence of images requires collecting a number of k-space samples at multiple time points, and the minimum number of samples needed for a predetermined spatio-temporal resolution is dictated in classical sampling theory by the Nyquist criterion [1]. However, the physics of MRI scanners impose a sequential sampling and the physiology and motion of body organs limit

Manuscript received October 22, 2013; revised January 13, 2014; accepted January 14, 2014. Date of publication January 17, 2014; date of current version March 31, 2014. Asterisk indicates corresponding author.

*J. Caballero is with the Department of Computing, Imperial College London, SW7 2AZ London, U.K. (e-mail: jose.caballero06@imperial.ac.uk).

A. N. Price and J. V. Hajnal are with the Division of Imaging Sciences and Biomedical Engineering Department, King's College London, St. Thomas' Hospital, SE1 7EH London, U.K. (e-mail: anthony.price@kcl.ac.uk; jo.hajnal@kcl.ac.uk).

D. Rueckert is with the Department of Computing, Imperial College London, SW7 2AZ London, U.K. (e-mail: d.rueckert@imperial.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>

Digital Object Identifier 10.1109/TMI.2014.2301271

the speed at which it can be performed, so meeting the Nyquist criterion often entails prohibitive scan durations or unfavorable balancing of spatial and temporal resolution. Acquiring fewer k-space samples than those imposed classically accelerates the process but introduces aliasing in image space. As a consequence, de-aliasing of signals from undersampled k-space acquisitions is an important stepping stone for the acceleration of MRI.

An advantage in cardiac cine is that the spatio-temporal redundancy can be predicted to some degree of accuracy and can be exploited in reconstruction. Methods such as UNFOLD [2] and Noquist [3] are based on the observation that only a fraction of the field of view requires the full temporal bandwidth. The option of data modelling has also been extensively explored for the optimization of dynamic magnetic resonance (MR) acquisition [4]–[10]. Methods like k-t BLAST [11] use *a priori* knowledge on the x-f support of the data obtained from a training process to prune a reconstruction that optimally reduces aliasing. Others like x-f choice [12] can learn the support region directly from the undersampled data.

Recently, the developing mathematical field of compressed sensing (CS) [13], [14], allowing theoretically error-free reconstructions of undersampled data using nonlinear processing, has been shown to be directly applicable to the problem of MRI acceleration [15]. Central to this approach is the assumption of sparsity of the reconstructed data under some transform domain, and the performance of the reconstruction relies heavily on the suitability of the sparsity model in terms of the approximation error it entails. Many of the methods proposed under the compressed sensing MRI (CSMRI) framework for dynamic imaging, seek for sparsity in the x-f support, which is the result of applying a Fourier transform of the image data along the temporal dimension. An example of successful methods enforcing x-f support sparsity is k-t FOCUSS [16]. The earlier method k-t SPARSE [17] additionally sparsifies data across space using a wavelet transform, which is known to provide approximately sparse representations for the vast majority of natural signals.

The potential of CSMRI methods is widely recognized, although exploration of the choice of sparsity model has been somewhat limited. Sparsity is often treated as synonymous with retrospective compressibility and this can lead to crude models that do not adhere well to the characteristics of the data. For high undersampling factors, the reconstruction relies strongly on the chosen *a priori* sparsity assumption, and models that are not representative will recover images with features that are unexpected in their fully sampled counterparts. Wavelet or total

variation (TV) based methods [17]–[19], for instance, are susceptible of generating blocky spatial artefacts, and very sparse x -f supports will miss out on rapid temporal changes [16], [20].

Adaptive techniques have been presented recently under the name of dictionary learning (DL) [21] that can learn characteristics from a signal and tailor an overcomplete set of basis functions providing higher levels of sparsity for that particular signal than those achievable by fixed bases transforms like wavelet or Fourier transforms. Their potential has been examined in image denoising [22], inpainting [23], and classification [24] among other general image processing tasks, but has also been demonstrated for reconstruction in accelerated MR within the CSMRI framework [25]–[28].

In this paper, a sparsity model combining patch based learnt and temporal gradient sparsity is introduced for the reconstruction of complex-valued cardiac cine from Cartesian MR undersampled data. A single spatio-temporal dictionary is trained for the encoding of the whole dataset. In addition, the temporal gradient transform is explored as an auxiliary sparsifying transform. This is mainly motivated by the observation from sample datasets that temporal gradients provide a sparser representation than spatial gradients. Additionally, the use of a penalty sparsity term that considers the entire temporal dimension of the data can enforce temporal cohesion globally in a way that patch based dictionaries cannot. This is particularly beneficial when using independent subsamples for each temporal frame, as this results in artefacts that flicker in time which tend to increase the temporal complexity of the target signal.

The aim of this study is to analyze the potential of the use of dictionaries in the context of cardiac cine. To carry out this analysis, synthetic experiments are arranged in which data acquired using parallel coils is retrospectively undersampled and reconstructed assuming a single coil setup. Conclusions drawn from these experiments should therefore not be taken as directly implementable in a real scan scenario, but will provide a simple paradigm for benchmarking the reconstruction capabilities of dictionaries. A practical implementation of DL for MRI will have to consider the extension of this technique to multi-coil MR technology.

This paper is organized as follows. Section II provides a formal introduction of the CSMRI problem when cardiac cine is considered. The two sparsifying transforms exploited in the algorithm proposed are described in Section III and the optimization problem they pose is stated in Section IV. In Section V we provide the details necessary for the implementation of the algorithm, which we term dictionary learning with temporal gradients (DLTG), as well as providing comments on its computational complexity. To conclude, the performance of the novel technique is analyzed in Section VI, where it is thoroughly compared with the k-t FOCUSS algorithm and where tuning parameters are examined.

II. COMPRESSED SENSING DYNAMIC MRI

The objective of CS is to devise a sampling strategy that will allow perfect reconstruction of a discrete signal from a minimum number of samples. The intuition behind it is that

any signal containing the redundancy necessary for its compression subsequent to a classical above-Nyquist sampling, can be acquired efficiently in an already compressed format from fewer samples. The downside is that reconstruction is nonlinear and hence also computationally demanding. Medical images are known to be compressible using the appropriate transforms [15] and are therefore suited to the CS acquisition paradigm.

The acquisition domain for MR data is k-space, which is equivalent to the Fourier domain. A sequence of images can be seen as 2D $P_x \times P_y$ k-space samples acquired at P_t different time instances that are stacked as a 3D volume. Let us refer to the fully sampled dataset as the one fulfilling the Nyquist sampling criterion. We will represent it in k-space as the column vector $\hat{\mathbf{x}}_f \in \mathbb{C}^P$, produced by the concatenation of columns in the 3D volume where $P = P_x P_y P_t$. The vector sequence in image space $\mathbf{x}_f \in \mathbb{C}^P$ is related to the k-space data by $\hat{\mathbf{x}}_f = \mathbf{F} \mathbf{x}_f + \mathbf{n}$, where \mathbf{F} performs a 2D discrete Fourier transform (DFT) on each temporal frame and $\mathbf{n} \in \mathbb{C}^P$ is additive white Gaussian (AWG) acquisition noise that is complex and circular. Now assume only a subset Ω of k-space is acquired, meaning that $\hat{\mathbf{x}}_u = \mathbf{M} \hat{\mathbf{x}}_f$ is the only data available for reconstruction. The undersampling mask $\mathbf{M} \in \mathbb{R}^{m \times P}$, $m \ll P$, contains the rows from the identity matrix that correspond to the samples of $\hat{\mathbf{x}}_f$ that are in Ω . Provided a sparsifying transform \mathbf{S} exists such that $\|\mathbf{S} \mathbf{x}_f\|_0 = L \ll m$, the CS dynamic MRI reconstruction is given by the solution to

$$\min_{\mathbf{x}} \|\mathbf{S} \mathbf{x}\|_0 \text{ s.t. } \|\mathbf{M} \mathbf{F} \mathbf{x} - \hat{\mathbf{x}}_u\|_2^2 < \epsilon \quad (1)$$

where ϵ is a small constant.

The l_p norm is defined as $\|\mathbf{x}\|_p = (\sum_i |x_i|^p)^{1/p}$, with x_i the scalar entry i of \mathbf{x} . The case when $p = 0$ is considered a pseudo-norm defined as a counter of nonzero elements in \mathbf{x} , hence $\|\mathbf{x}\|_0 = \{\#i : x_i \neq 0\}$. In many cases the $l0$ pseudo-norm is replaced by the $l1$ norm as a measure promoting sparsity in order to make problem (1) convex.

An important requirement for CSMRI is that the sparsifying transform \mathbf{S} is incoherent with the acquisition. This can be guaranteed if matrix $\mathbf{M} \mathbf{F} \mathbf{S}^{-1}$ satisfies the restricted isometry property [29]. Checking this property is an NP hard problem, but random partial Fourier measurements have been shown to provide sufficient incoherence with a number of transforms \mathbf{S} and are generally regarded as a good practical choice [15], [30].

III. SPARSIFYING TRANSFORMS FOR DYNAMIC MRI

There is a direct relationship between the sparsity level L provided by the transform \mathbf{S} and the minimum number of samples necessary for perfect reconstruction [29]. The sparser the model chosen, the higher the achievable acceleration rates will be. In this section, two sparsifying transforms that are well suited to the problem of cardiac cine are described. First, an adaptive patch based transform derived from DL theory is presented for the case of dynamic MR complex data. Then, the temporal gradient transform is proposed as a suitable global sparsity model for cardiac cine that can make reconstructions converge faster and improve performance at high undersampling rates.

A. Dictionary Learning for Dynamic MRI

DL refers to the process of adapting an initial set of basis functions to a specific signal through a training process such that it will provide a sparse representation of that particular signal. Denote training patches $\mathbf{x}_{T,i} \in \mathbb{R}^n$, $i = 1, \dots, M$, as column vectors of size n to be used for the training of a dictionary $\mathbf{D} \in \mathbb{R}^{n \times N}$ of N atoms arranged as column vectors. The training of a real-valued dictionary adapted to that training set can be stated as

$$\min_{\mathbf{D}, \boldsymbol{\Gamma}_T} \|\boldsymbol{\gamma}_{T,i}\|_0 \text{ s.t. } \|\mathbf{x}_{T,i} - \mathbf{D}\boldsymbol{\gamma}_{T,i}\|_2^2 < \epsilon \quad \forall i = 1, \dots, M \quad (2)$$

where $\boldsymbol{\Gamma}_T \in \mathbb{R}^{n \times M}$ is a matrix gathering the sparse representation of $\mathbf{x}_{T,i}$ as column vectors $\boldsymbol{\gamma}_{T,i}$. The subscript T specifies that these are variables for training. After this process, the dictionary can be used to find an approximation of a real-valued dataset as a sparse coding problem with a greedy l_0 minimization technique such as orthogonal matching pursuit (OMP) [31].

Recently, DL has been used for 2D structural MR image reconstruction largely outperforming competing techniques based on fixed basis transforms [25]. The extension to the case of MR sequences was introduced in [26] with the use of spatio-temporal 3D dictionaries, but only the reconstruction of synthetic real-valued sequences without a phase component was addressed. This is not feasible in practice since the observed k-space samples always relate to a complex image.

The training of complex-valued dictionaries that are suitable for MR data representation is possible defining $\boldsymbol{\Gamma}$ and \mathbf{D} as complex-valued variables as originally proposed in [25], [32]. Instead, we carry out this representation by using a single real-valued dictionary which is trained on real and imaginary parts of MR data for their independent coding. In Section VI-E we look at the differences between these two learning and coding strategies for the processing of MR data.

Previous work has already investigated the use of dictionary learning for cardiac cine. In [32], complex spatio-temporal dictionaries are learnt and used as unique sparse model for patches of the reconstructed data. A key difference with the method proposed here apart from the training strategy is the fact that the dictionary is propagated and updated along the temporal dimension. This was proposed for dictionary learning applications in natural video processing, where the data structure is expected to rapidly change across time, and propagating and updating the dictionary adapts it to changing information [33]. However, the variability in cardiac cine data can be assumed to be limited and therefore a single training stage should be able to capture the structural information required, hence avoiding the additional computational load needed for updating the dictionary at each temporal frame.

Other variations of learning methods are possible. In [34]–[36], the trained frame consists of temporal functions that sparsely represent the temporal profile of each pixel in the dataset. Although there exist algorithmic differences in the reconstruction process, the concept of this previous work can be thought of as using independent patches that are only temporal and cover the entire dataset. Cardiac cine data is known to be redundant through space as well as through time (although it is usually the case that the temporal dimension is

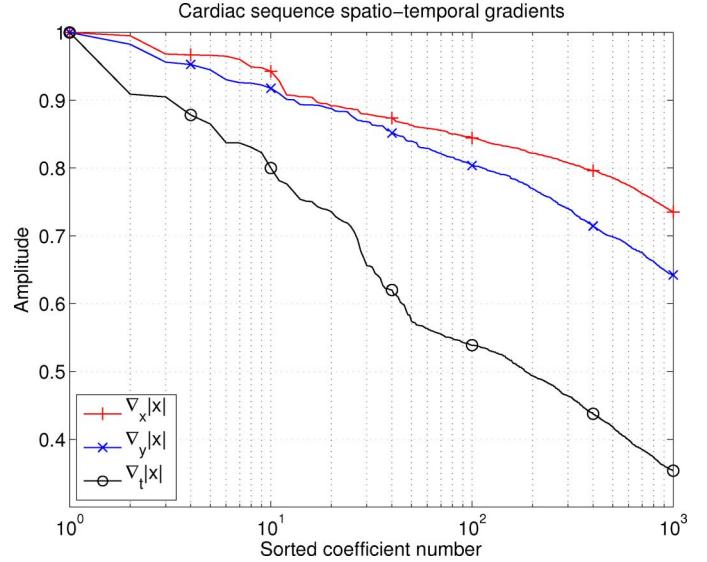


Fig. 1. Comparison of the decay in transform coefficients of transforms ∇_x , ∇_y , and ∇_t . The latter shows the fastest decay.

more redundant than the spatial dimension), which is why also exploiting spatial sparsity with a spatio-temporal dictionary could be advantageous.

B. Temporal Gradient Sparsity

An additional sparsity constraint can be imposed on the temporal finite differences (i.e., the first-order temporal gradient) of the dataset. Many authors have explored TV for imposing sparsity constraints on a CSMRI reconstruction because it provides sensible sparsity levels, but also because its optimization can be extremely efficient [15], [18], [37], [38]. In many cases, TV is not the main sparsifying transform but rather an auxiliary constraint that can stabilize and correct the solution provided by the main transform. This operation considers an equally weighted combination of the pixel-wise finite differences along space and time, but this is rarely a sensible assumption in cardiac cine because spatial and temporal gradients, which make up the individual dimensions of TV, will usually have different sparsity levels.

Consider a complex-valued single slice cardiac cine scan to be a volume \mathbf{X}^{3D} of entries $(\mathbf{X}^{3D})_{h,i,j} = X_{h,i,j}$ in the x , y , and t axes. An example magnitude temporal frame of this kind of datasets is shown in Fig. 3(a). We denote $(\nabla_x \mathbf{X}^{3D})_{h,i,j} = X_{h+1,i,j} - X_{h,i,j}$ the finite difference result of \mathbf{X}^{3D} along dimension x and similarly $(\nabla_y \mathbf{X}^{3D})_{h,i,j} = X_{h,i+1,j} - X_{h,i,j}$ and $(\nabla_t \mathbf{X}^{3D})_{h,i,j} = X_{h,i,j+1} - X_{h,i,j}$ along dimensions y and t . Expressing $(\nabla_x \mathbf{x})_{h,i,j}$, $(\nabla_y \mathbf{x})_{h,i,j}$, and $(\nabla_t \mathbf{x})_{h,i,j}$ as equivalent expressions when \mathbf{x} is a concatenated column vector version of \mathbf{X}^{3D} , and $|\mathbf{x}|$ referring to the element-wise absolute value of the complex vector \mathbf{x} , Fig. 1 shows the decay of magnitude coefficients of the three transforms, ordered and normalized.

The faster decay of the coefficients in the ∇_t domain confirms that the information in this example is sparser in first-order temporal gradients than in the spatial gradients. For this reason, we only use ∇_t as an auxiliary transform instead of the more

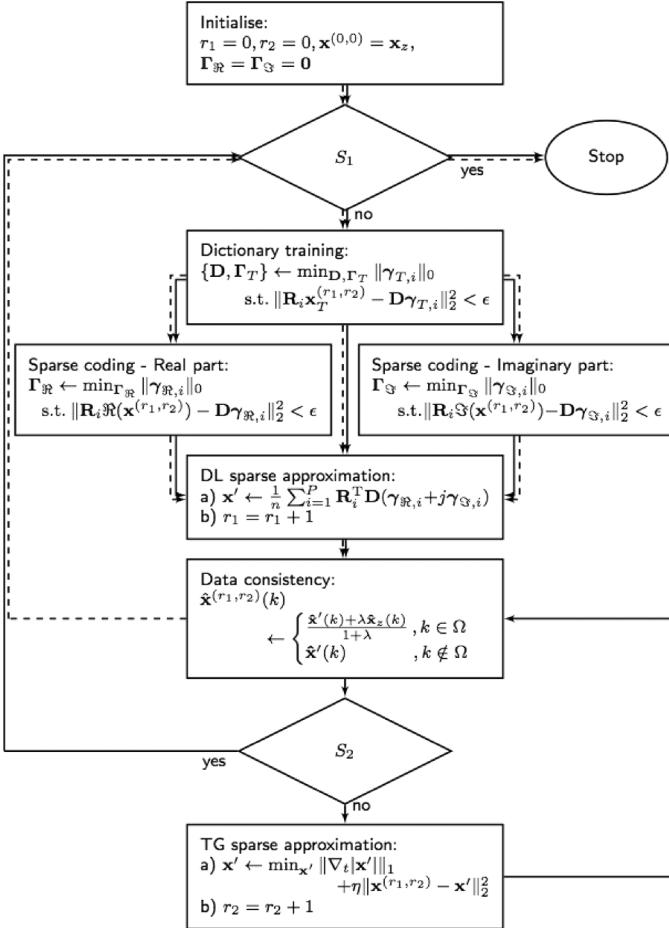


Fig. 2. Algorithm flowchart for DLTG (solid arrows) and the DLMRI (dashed arrows) algorithms.

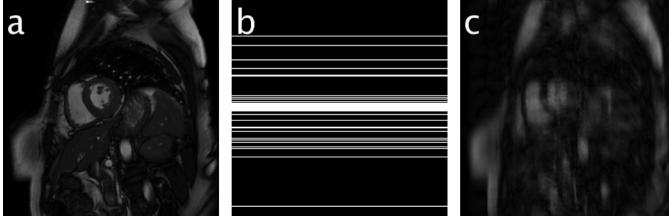


Fig. 3. Example of a magnitude temporal frame from one of the datasets analyzed (a). The undersampling mask (b) applied in k-space reduces acquisition time but introduces aliasing in image space (c).

extended TV. This decision is again in the spirit of choosing a sparsity model that is better suited to the particular case of cardiac cine. Higher sparsity enables more accurate CS reconstruction, so using ∇_t as an auxiliary sparsity transform instead of TV provides a better sparsity model on which to rely. This additional term on the model was shown in [26] to increase the performance of a similar algorithm designed for the reconstruction of purely real-valued sequences.

IV. PROBLEM STATEMENT

The problem posed is to simultaneously find a solution dataset $\mathbf{x} \in \mathbb{C}^P$ such that its real and imaginary parts $\Re(\mathbf{x})$ and $\Im(\mathbf{x})$ are sparsely represented with the same real-valued dictionary \mathbf{D} , which is to be learnt. Additionally, we penalize

solutions for which $\|\nabla_t \mathbf{x}\|_1$ is large and favour those that are overall consistent with the m k-space acquisitions $\hat{\mathbf{x}}_u$. This is a similar approach as that in [25], but instead of using 2D spatial dictionaries we use 3D spatio-temporal dictionaries and further exploit temporal redundancy in cardiac cine with an auxiliary temporal gradient constraint.

Let us define as $\mathbf{R}_i \in \mathbb{R}^{n \times P}$ an operator that extracts as a column vector the 3D patch in the dataset starting from pixel location i and wraps around volume boundaries. We use Γ_R and Γ_S to denote the sparse coding of $\Re(\mathbf{x})$ and $\Im(\mathbf{x})$ respectively, where each column $\gamma_{R,i}$ or $\gamma_{S,i}$ is the coding of patch $\mathbf{R}_i \Re(\mathbf{x})$ or $\mathbf{R}_i \Im(\mathbf{x})$. Throughout the description we assume that the step size for patch extraction is 1, meaning that $i = 1, \dots, P$ and that each pixel in \mathbf{x} is represented by n different patches.

The problem can be expressed as the following global statement:

$$\begin{aligned} \min_{\mathbf{D}, \Gamma_R, \Gamma_S} & \sum_{i=1}^P (\|\gamma_{R,i}\|_0 + \|\gamma_{S,i}\|_0) + \nu \|\mathbf{M}\mathbf{F}\mathbf{x} - \hat{\mathbf{x}}_u\|_2^2 \\ & + \mu \|\nabla_t \mathbf{x}\|_1 \\ \text{s.t. } & \begin{cases} \|\mathbf{R}_i \Re(\mathbf{x}) - \mathbf{D} \gamma_{R,i}\|_2^2 < \epsilon & \forall i \\ \|\mathbf{R}_i \Im(\mathbf{x}) - \mathbf{D} \gamma_{S,i}\|_2^2 < \epsilon & \forall i \end{cases}. \end{aligned} \quad (3)$$

This optimization problem is nonconvex, so we opt to split it into three simpler subproblems that are either convex or can efficiently be solved with greedy methods. Alternating the solution of these three subproblems iteratively will yield an approximation to problem (3). Without modification, we introduce an auxiliary variable \mathbf{x}_{TG} and write

$$\begin{aligned} \min_{\mathbf{D}, \Gamma_R, \Gamma_S} & \sum_{i=1}^P (\|\gamma_{R,i}\|_0 + \|\gamma_{S,i}\|_0) + \nu \|\mathbf{M}\mathbf{F}\mathbf{x} - \hat{\mathbf{x}}_u\|_2^2 \\ & + \mu \|\nabla_t \mathbf{x}_{TG}\|_1 \\ \text{s.t. } & \begin{cases} \|\mathbf{R}_i \Re(\mathbf{x}) - \mathbf{D} \gamma_{R,i}\|_2^2 < \epsilon & \forall i \\ \|\mathbf{R}_i \Im(\mathbf{x}) - \mathbf{D} \gamma_{S,i}\|_2^2 < \epsilon & \forall i \\ \mathbf{x} = \mathbf{x}_{TG} \end{cases}. \end{aligned} \quad (4)$$

Then, the new constraint is relaxed and included as a quadratic penalty term

$$\begin{aligned} \min_{\mathbf{D}, \Gamma_R, \Gamma_S} & \sum_{i=1}^P (\|\gamma_{R,i}\|_0 + \|\gamma_{S,i}\|_0) + \nu \|\mathbf{M}\mathbf{F}\mathbf{x} - \hat{\mathbf{x}}_u\|_2^2 \\ & + \mu \|\nabla_t \mathbf{x}_{TG}\|_1 + \beta \|\mathbf{x} - \mathbf{x}_{TG}\|_2^2 \\ \text{s.t. } & \begin{cases} \|\mathbf{R}_i \Re(\mathbf{x}) - \mathbf{D} \gamma_{R,i}\|_2^2 < \epsilon & \forall i \\ \|\mathbf{R}_i \Im(\mathbf{x}) - \mathbf{D} \gamma_{S,i}\|_2^2 < \epsilon & \forall i \end{cases}. \end{aligned} \quad (5)$$

There are four tuning parameters in (5): ν , μ , β , and ϵ . These control respectively the consistency with the acquired k-space samples, the temporal gradient sparsity of the dummy variable \mathbf{x}_{TG} , the distance of the result \mathbf{x} with respect to this dummy variable, and the representation accuracy of \mathbf{D} , Γ_R , and Γ_S . The last parameter is inversely related to the sparsity allowed in the dictionary representation.

Even though phase information is relevant for applications such as flow measurement, in cardiac cine as in other structural MRI modalities, the information of most interest is normally

provided by the magnitude part of the complex signal formed. An appealing approach would therefore be to define the problem such that only a magnitude sequence $|x|$ is reconstructed discarding the reconstruction of its phase information. This is simpler than the problem posed in (5) because the number of unknowns is reduced by a half and the system of equations becomes better determined, which is why some solutions proposed have adopted this approach [26], [27]. However, this problem statement overlooks the nature of the observed data \hat{x}_u . The samples acquired in k-space always correspond in practice to a complex image, so trying to infer the k-space of a magnitude signal from an undersampled version of its complex representation is not a viable option. The reconstructed \mathbf{x} has to be complex-valued regardless of the clinical usefulness of its magnitude and phase information.

V. DLTG ALGORITHM

Problem (5) is broken down into three simpler subproblems. The DLTG algorithm iteratively refines a solution in three separate steps that each solve the global problem with some free variables fixed.

A. Dictionary Training and Sparse Coding

Begin by assuming \mathbf{x} and \mathbf{x}_{TG} fixed. The only free variables are \mathbf{D} , Γ_{\Re} and Γ_{\Im} . In words, we seek the dictionary and sparse codings that will represent both sparsely and accurately the real and imaginary parts of the dataset, or formally

$$\begin{aligned} \min_{\mathbf{D}, \Gamma_{\Re}, \Gamma_{\Im}} & \sum_{i=1}^P (\|\boldsymbol{\gamma}_{\Re, i}\|_0 + \|\boldsymbol{\gamma}_{\Im, i}\|_0) \\ \text{s.t. } & \begin{cases} \|\mathbf{R}_i \Re(\mathbf{x}) - \mathbf{D} \boldsymbol{\gamma}_{\Re, i}\|_2^2 < \epsilon & \forall i \\ \|\mathbf{R}_i \Im(\mathbf{x}) - \mathbf{D} \boldsymbol{\gamma}_{\Im, i}\|_2^2 < \epsilon & \forall i \end{cases} . \end{aligned} \quad (6)$$

This is the general dictionary learning problem, where the training data are patches $\mathbf{R}_i \Re(\mathbf{x})$ and $\mathbf{R}_i \Im(\mathbf{x})$. The K-SVD algorithm [39] can solve (6). This DL algorithm is capable of learning an overcomplete basis that can sparsely represent the underlying structure from corrupted measurements by iteratively solving for one of the two free variables at a time. It was originally proposed as a learning method for denoising [22] to provide an overcomplete frame that approximates the desired signal sparsely contrary to its added noise. For an undersampling mask providing sufficient aliasing incoherence, aliasing will have noise-like properties that the trained dictionary will not be able to approximate with a sparse representation either. Therefore, a sparse approximation of the corrupted sequence \mathbf{x} using the outcome dictionary from (6) will tend to reproduce anatomical structure and miss features derived from aliasing.

In practice, a reduced number of patches are used as training data for efficiency purposes. For the description of the algorithm, we will refer to the group of training patches as \mathbf{x}_T , which will be a collection of M patches chosen from a regular grid on $\Re(\mathbf{x})$ and $\Im(\mathbf{x})$. Their sparse representation will be denoted as \mathbf{G}_T . Once a dictionary \mathbf{D} has been learnt from the training set, it

can be used to code real and imaginary parts of the entire dataset independently with OMP

$$\min_{\Gamma_{\Re}} \sum_{i=1}^P \|\boldsymbol{\gamma}_{\Re, i}\|_0 \quad \text{s.t.} \quad \|\mathbf{R}_i \Re(\mathbf{x}) - \mathbf{D} \boldsymbol{\gamma}_{\Re, i}\|_2^2 < \epsilon \quad \forall i \quad (7)$$

$$\min_{\Gamma_{\Im}} \sum_{i=1}^P \|\boldsymbol{\gamma}_{\Im, i}\|_0 \quad \text{s.t.} \quad \|\mathbf{R}_i \Im(\mathbf{x}) - \mathbf{D} \boldsymbol{\gamma}_{\Im, i}\|_2^2 < \epsilon \quad \forall i. \quad (8)$$

Assuming patches are one pixel away from each other in the three dimensions and that they wrap around edges of the sequence, each pixel will be represented by n different coded patches.

B. Temporal Gradient Filtering

Let us fix now variables \mathbf{D} , Γ_{\Re} , Γ_{\Im} , and \mathbf{x} , and minimize the functional with respect to \mathbf{x}_{TG} . The problem becomes

$$\min_{\mathbf{x}_{TG}} \|\nabla_t \mathbf{x}_{TG}\|_1 + \eta \|\mathbf{x} - \mathbf{x}_{TG}\|_2^2 \quad (9)$$

with $\eta = \beta/\mu$. The magnitude of the dummy variable \mathbf{x}_{TG} is driven towards sparsity in its temporal gradients while minimizing the quadratic distance with respect to solution \mathbf{x} . The phase difference between \mathbf{x} and \mathbf{x}_{TG} will not alter the first term in (9), so we set $\angle \mathbf{x}_{TG} = \angle \mathbf{x}$ and solve

$$\min_{\mathbf{x}_{TG}} \|\nabla_t \mathbf{x}_{TG}\|_1 + \eta \|\mathbf{x} - \mathbf{x}_{TG}\|_2^2. \quad (10)$$

This can be very efficiently solved by the primal-dual method in [40] with an iterative clipping algorithm. Specifically, the following two computations are iterated:

$$|\mathbf{x}_{TG}|^{(r+1)} = |\mathbf{x}| - \nabla_t^T \mathbf{z}^{(r)} \quad (11)$$

$$\mathbf{z}^{(r+1)} = \text{clip} \left(\mathbf{z}^{(r)} + \frac{1}{\alpha} \nabla_t |\mathbf{x}_{TG}|^{(r+1)}, \frac{1}{2\eta} \right) \quad (12)$$

for $r \geq 0$, $\mathbf{z}^{(0)} = \mathbf{0}$ and $\alpha \geq \text{maxeig}(\nabla_t \nabla_t^T)$. For details on the derivation of this algorithm we refer to [40].

C. Data Consistency

The last subproblem looks at the case where \mathbf{x} is the only free variable. Equation (5) now becomes

$$\begin{aligned} \min_{\mathbf{x}} & \|\mathbf{M} \mathbf{F} \mathbf{x} - \hat{\mathbf{x}}_u\|_2^2 + \frac{\beta}{\nu} \|\mathbf{x} - \mathbf{x}_{TG}\|_2^2 \\ \text{s.t. } & \begin{cases} \|\mathbf{R}_i \Re(\mathbf{x}) - \mathbf{D} \boldsymbol{\gamma}_{\Re, i}\|_2^2 < \epsilon & \forall i \\ \|\mathbf{R}_i \Im(\mathbf{x}) - \mathbf{D} \boldsymbol{\gamma}_{\Im, i}\|_2^2 < \epsilon & \forall i \end{cases} . \end{aligned} \quad (13)$$

This can be rewritten as the following unconstrained optimization problem:

$$\begin{aligned} \min_{\mathbf{x}} & \sum_{i=1}^P (\|\mathbf{R}_i \Re(\mathbf{x}) - \mathbf{D} \boldsymbol{\gamma}_{\Re, i}\|_2^2 + \|\mathbf{R}_i \Im(\mathbf{x}) - \mathbf{D} \boldsymbol{\gamma}_{\Im, i}\|_2^2) \\ & + \nu' \|\mathbf{M} \mathbf{F} \mathbf{x} - \hat{\mathbf{x}}_u\|_2^2 + \beta' \|\mathbf{x} - \mathbf{x}_{TG}\|_2^2. \end{aligned} \quad (14)$$

Referring to this functional as f , the optimum solution is the one for which $\partial f / \partial \Re(\mathbf{x}) = \partial f / \partial \Im(\mathbf{x}) = 0$. Let us simplify the notation using $\mathbf{F}_u = \mathbf{MF}$ as the undersampled Fourier operator, and solve $\partial f / \partial \Re(\mathbf{x}) = 0$

$$\begin{aligned} & \left(\sum_{i=1}^P \mathbf{R}_i^T \mathbf{R}_i + \nu' \mathbf{F}_u^H \mathbf{F}_u + \beta' \right) \Re(\mathbf{x}) \\ &= \sum_{i=1}^P \mathbf{R}_i^T \mathbf{D} \gamma_{\Re,i} + \nu' \mathbf{F}_u^H \Re(\hat{\mathbf{x}}_u) + \beta' \Re(\mathbf{x}_{TG}). \end{aligned} \quad (15)$$

Taking the Fourier transform on both sides of (15), we have

$$\begin{aligned} & \left(\mathbf{F} \sum_{i=1}^P \mathbf{R}_i^T \mathbf{R}_i \mathbf{F}^H + \nu' \mathbf{F} \mathbf{F}_u^H \mathbf{F}_u \mathbf{F}^H + \beta' \mathbf{F} \mathbf{F}^H \right) \mathbf{F} \Re(\mathbf{x}) \\ &= \mathbf{F} \sum_{i=1}^P \mathbf{R}_i^T \mathbf{D} \gamma_{\Re,i} + \nu' \mathbf{F} \mathbf{F}_u^H \Re(\hat{\mathbf{x}}_u) + \beta' \mathbf{F} \Re(\mathbf{x}_{TG}) \end{aligned} \quad (16)$$

which yields the solution k-space of the real part of \mathbf{x}

$$\mathbf{F} \Re(\mathbf{x}) = \frac{\mathbf{F} \sum_{i=1}^P \mathbf{R}_i^T \mathbf{D} \gamma_{\Re,i} + \beta' \mathbf{F} \Re(\mathbf{x}_{TG}) + \nu' \mathbf{F} \mathbf{F}_u^H \Re(\hat{\mathbf{x}}_u)}{\mathbf{F} \sum_{i=1}^P \mathbf{R}_i^T \mathbf{R}_i \mathbf{F}^H + \beta' \mathbf{F} \mathbf{F}^H + \nu' \mathbf{F} \mathbf{F}_u^H \mathbf{F}_u \mathbf{F}^H}. \quad (17)$$

Replacing \Re by \Im in (17) we obtain the solution for the k-space of the imaginary part, hence we can write the k-space solution of \mathbf{x} (using notation $j = \sqrt{-1}$) as

$$\hat{\mathbf{x}} = \frac{\mathbf{F} \sum_{i=1}^P \mathbf{R}_i^T \mathbf{D} (\gamma_{\Re,i} + j \gamma_{\Im,i}) + \beta' \mathbf{F} \mathbf{x}_{TG} + \nu' \mathbf{F} \mathbf{F}_u^H \hat{\mathbf{x}}_u}{\mathbf{F} \sum_{i=1}^P \mathbf{R}_i^T \mathbf{R}_i \mathbf{F}^H + \beta' \mathbf{F} \mathbf{F}^H + \nu' \mathbf{F} \mathbf{F}_u^H \mathbf{F}_u \mathbf{F}^H}. \quad (18)$$

To better understand this expression we can use the following simplifications. The term $\mathbf{F} \mathbf{F}_u^H \hat{\mathbf{x}}_u$ is the zero-filled k-space acquisition which will be denoted as $\hat{\mathbf{x}}_z$. Assuming patches overlap and that the operator \mathbf{R}_i wraps around the boundaries such that each pixel is represented by n patches, the term $\mathbf{F} \sum_{i=1}^P \mathbf{R}_i^T \mathbf{R}_i \mathbf{F}^H$ is simply the weighted $P \times P$ identity matrix $n \mathbf{I}_P$. Moreover, $\mathbf{F} \mathbf{F}_u^H \mathbf{F}_u \mathbf{F}^H$ is equivalent to a $P \times P$ diagonal matrix containing a 1 in the diagonal whenever a k-space location was acquired or a 0 otherwise. The simplifications $\hat{\mathbf{x}}_{TG} = \mathbf{F} \mathbf{x}_{TG}$ and $\mathbf{I}_P = \mathbf{F} \mathbf{F}^H$ are trivial, and also notice that the expression $\mathbf{F} \sum_{i=1}^P \mathbf{R}_i^T \mathbf{D} \gamma_i / n$ represents the Fourier transform of the solution obtained by averaging the contribution of all coded patches relocated to their corresponding position within the dataset. For simplicity we refer to this solution as \mathbf{x}_{DL} , and $\hat{\mathbf{x}}_{DL}$ is its Fourier representation. The final expression for $\hat{\mathbf{x}}$ is therefore

$$\hat{\mathbf{x}}(k) = \begin{cases} \frac{\hat{\mathbf{x}}_{DL}(k) + \frac{\beta'}{n} \hat{\mathbf{x}}_{TG}(k) + \frac{\nu'}{n} \hat{\mathbf{x}}_z(k)}{1 + \frac{\beta'}{n} + \frac{\nu'}{n}}, & k \in \Omega \\ \frac{\hat{\mathbf{x}}_{DL}(k) + \frac{\beta'}{n} \hat{\mathbf{x}}_{TG}(k)}{1 + \frac{\beta'}{n}}, & k \notin \Omega. \end{cases} \quad (19)$$

The update of solution \mathbf{x} involves an average between the dictionary sparse solution $\hat{\mathbf{x}}_{DL}$, the temporal gradient sparse solution $\hat{\mathbf{x}}_{TG}$, and the original acquisitions $\hat{\mathbf{x}}_z$ for k-space locations that were acquired. This means that this step requires the tuning of parameters ν' and β' . To simplify this task, in the implementation of the algorithm we choose to update $\hat{\mathbf{x}}$ using only $\hat{\mathbf{x}}_z$ and either $\hat{\mathbf{x}}_{DL}$ or $\hat{\mathbf{x}}_{TG}$ with a single constant λ depending on which solution was updated last. Equation (19) then becomes

$$\hat{\mathbf{x}}(k) = \begin{cases} \frac{\hat{\mathbf{x}}'(k) + \lambda \hat{\mathbf{x}}_z(k)}{1 + \lambda}, & k \in \Omega \\ \hat{\mathbf{x}}'(k), & k \notin \Omega \end{cases} \quad (20)$$

with $\hat{\mathbf{x}}'$ alternating between $\hat{\mathbf{x}}_{DL}$ and $\hat{\mathbf{x}}_{TG}$. Noise standard deviation is taken into account by the regularization parameter $\lambda = q/\sigma$, where q is a constant that can be set empirically as is shown in Section VI-C.

Algorithm Design

Fig. 2 describes the ordering in which these three steps are performed in the DLTG algorithm. The data consistency step is interleaved between the other two steps and the TG module is iterated in an inner loop fashion to force a slow and smooth convergence towards a result jointly satisfying the sparse temporal gradient and the data consistency constraints. Additionally, a Dictionary Learning MRI (DLMRI) version of the method in which the TG term is ignored (equivalent to setting $\beta = \beta' = 0$) is also presented. Notice that this differs from the DLMRI method presented in [25] in that real and imaginary parts are coded independently with a real-valued dictionary.

The algorithm is initialized with the zero-filled sequence \mathbf{x}_z and the stopping criteria S_1 and S_2 can either be convergence to a stable solution or a maximum number of iterations I_1 and I_2 for outer and inner loops respectively. All the results below use small values for constants ϵ and η to ensure a slow convergence towards a result tightly adhering to the sparsity model imposed. We also use $I_2 = 10$ and a large number of outer iterations $I_1 > 100$ to guarantee the end result reaches a stable solution. The dictionary in the training stage is initialized at each iteration with an overcomplete DCT dictionary and coding matrices Γ_{\Re} and Γ_{\Im} are initially $\mathbf{0}$.

Algorithm Complexity

The complexity of the DLMRI algorithm, which constitutes the first half of the DLTG algorithm, is dominated by the dictionary learning and sparse coding steps. The training of the dictionary and sparse coding of patches are performed using the efficient implementations of the K-SVD and Batch-OMP algorithms made available by Rubinstein [41], [42]. In both cases, the computation is dominated by the cost of sparse coding a patch of size n with a dictionary of N atoms. The number of operations necessary for this can be described as a function of OMP iterations L . The stopping criterion described in Section IV for sparse coding is the data consistency level ϵ , and although the relationship between L and ϵ is nontrivial because of its nonlinearity, it is easily observed that a larger ϵ implies fewer OMP iterations for a given problem.

Batch-OMP precomputes the matrix $\mathbf{D}^T \mathbf{D}$ to save processing time when coding large sets of patches with the same dictionary. This first step requires nN^2 operations. The coding of a real-valued patch that requires L OMP iterations with Batch-OMP involves approximately $2nN + L^2N + L^3$ operations [41]. Given that we code a total of P patches twice for each dataset, we can provide the average number of operations per dataset coding as $nN^2 + 2P(2nN + \bar{L}^2N + \bar{L}^3)$, where \bar{L} is the average number of OMP iterations for coding real and imaginary parts. Also, a K-SVD training iteration will involve approximately $M(2nN + L^2N + L^3)$ operations for L OMP iterations [41], so on average we have $I_T M(2nN + \bar{L}^2N + \bar{L}^3)$ operations. The data consistency step only involves two fast Fourier transforms per temporal frame and averaging operations that are negligible compared to the sparse coding step. The extra computation for solving (9) in the DLTG case is also very small since only sparse matrix-vector multiplications and clipping operations are required. In Section VI-G we provide empirical results on the runtime of the algorithm.

VI. RESULTS

A. Experimental Setup

Fully sampled short-axis cardiac cine scans were acquired from 10 subjects for the analysis of the proposed method. All datasets contain 30 temporal frames of size 256×256 (i.e., $P = 256 \times 256 \times 30$) with a 320×320 mm field of view and 10 mm slice thickness, and were generated using an optimal combination of 32-channel data. For the 10 subjects the mean \pm standard deviation for heart rate was 62 ± 10.2 bpm, with the 30 frames giving a temporal rate of 33 ± 5.5 ms. The raw multi-coil data was reconstructed using SENSE [43] with no k-space undersampling and retrospective gating. Coil sensitivity maps were normalized to a body coil image to produce a single complex-valued image set that could then either be back-transformed to regenerate complex k-space samples or further processed to form final magnitude images.

These scans contain unavoidable k-space acquisition noise and this formally precludes quantitative comparison between a given reconstruction and the fully sampled dataset. Nonetheless, for the purposes of evaluation we will treat the fully sampled data as ground truth (i.e., treat them as if noiseless). To assess noisy scenarios we then artificially add noise to k-space. The quality of all reconstructions \mathbf{x} will be measured with $\text{PSNR}(\mathbf{x}) = 10 \log (1 / (\|\mathbf{x}_f - \mathbf{x}\|_2^2 / P))$, where \mathbf{x}_f denotes the fully sampled dataset. For reconstructions without artificial noise, visual inspection of magnitude and phase information as well as the mean structural similarity index (MSSIM) [44] will also be considered.

Throughout this section, fully sampled k-space data is artificially subsampled using 2D binary undersampling masks $\mathbf{M} \in \mathbb{R}^{m \times P}$. We only consider the case of Cartesian undersampling, which is the most common in practice, although more elaborate sampling strategies like radial or spiral [45], [46] could be equally possible with modifications on the data consistency step and using nonuniform Fourier transforms [47]. Even though greater aliasing incoherence can be achieved with 2D k-space undersampling [45], frequency encodes can be considered in-

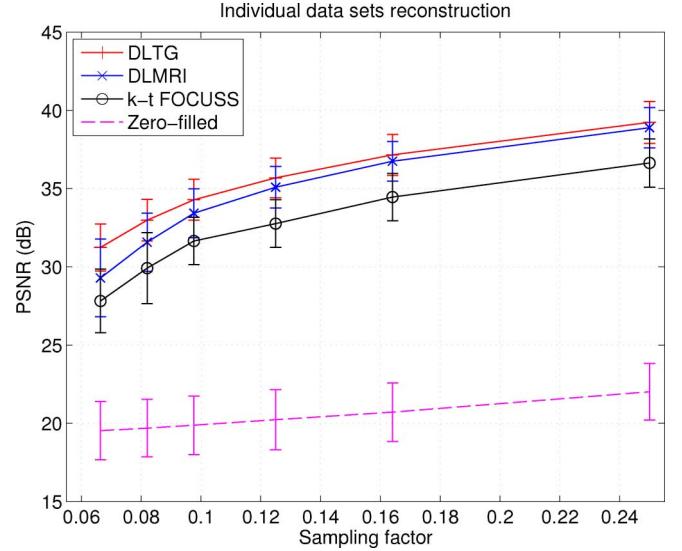


Fig. 4. Mean and standard deviation PSNR performance of reconstructions from 10 scans retrospectively undersampled without added noise.

stantaneous relative to phase encodes, so acceleration is only meaningful through phase encode undersampling.

It has been claimed in the past that drawing independent realizations of this random experiment for each temporal frame significantly increases eddy currents. In [48], this issue is remedied by pairing consecutive random phase encodes. At each frame, the eight lowest spatial frequencies are always acquired and other frequencies have a probability of being acquired determined by a Gaussian variable density function that is marginally offset, such that the probability of acquisition never reaches zero even at the highest frequencies. An implementation of this approach can be found in [49], and an example of a 2D mask and its effect on the magnitude of a temporal frame is shown in Fig. 3 for a six-fold acceleration.

The proposed method is compared to k-t FOCUSS [16] as a representative algorithm of the CS dynamic MRI methods using fixed bases sparsifying transforms. The implementation of this algorithm is publicly available in [49]. A single regularization parameter can be tuned to trade between a k-space data consistency term and an x-f domain sparsity term, and even though this parameter cannot be optimally determined *a priori*, we sweep across a large spectrum of values in all experiments and only show the best reconstruction in order to always compare to the best possible scenario.

In contrast, DLTG results are not optimized individually and, unless otherwise stated, its parameters are kept constant across all experiments. We use $M = 10^4$ patches of size $n = 4 \times 4 \times 4$ to train dictionaries of $N = 600$ atoms using $I_T = 10$ K-SVD iterations. These parameters were chosen based on empirical tradeoffs between performance and efficiency. The regularization parameters are $\epsilon = 0.007$ and $\eta = 4 \times 10^{-4}$, which allows to analyze the potential of the model by tightly adhering the result to it as will be shown in Section VI-G.

B. Reconstruction of Individual Datasets

The first experiment considers the simple case where datasets are undersampled and reconstructed with different acceleration

TABLE I
MSSIM COMPARISON (mean \pm std) $\times 10^{-2}$

Sampling factor	k-t FOCUSS	DLMRI	DLTG
0.25	95.8 \pm 2.5	97.3 \pm 0.8	97.4 \pm 1.0
0.16	93.9 \pm 3.4	96.2 \pm 0.9	96.5 \pm 1.1
0.12	91.4 \pm 6.0	95.0 \pm 1.1	95.5 \pm 1.1
0.10	89.9 \pm 5.4	93.3 \pm 1.6	94.4 \pm 1.3
0.08	86.2 \pm 6.3	91.2 \pm 1.8	93.1 \pm 1.3
0.06	80.5 \pm 7.4	87.7 \pm 3.0	91.0 \pm 1.4

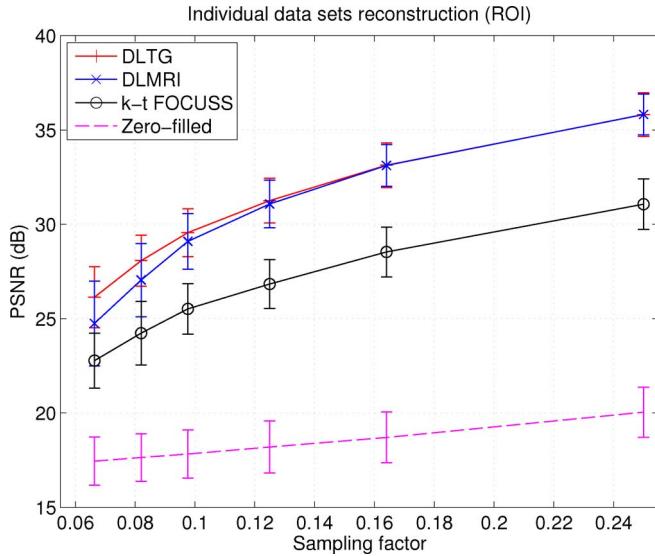


Fig. 5. PSNR performance of different reconstructions evaluated only within a ROI containing the heart.

factors without added noise. Fig. 4 plots the mean PSNR of the reconstructions and one standard deviation away from them against the acceleration factor for the 10 subjects considered. For this particular case, the noise regularization factor is set to $\lambda \rightarrow \infty$, such that in the DLMRI and DLTG algorithms all k-space samples that are acquired are constantly fed back to the reconstruction's k-space in the data consistency step without any weighting.

Even though it is ideally optimized, The k-t FOCUSS algorithm is clearly outperformed by the DLMRI and DLTG algorithms for all sampling ratios. A further comparison made in Table I using the MSSIM metric [44] supports this conclusion. MSSIM can quantify structural similarity between two images better than PSNR by comparing luminance and contrast in patches. For our results we consider the mean MSSIM of the 30 temporal frames in each dataset. Although the standard deviations of the PSNR and MSSIM results are relatively large, the rank order for both figures of merit for the three methods and for each individual reconstruction was DLTG > DLMRI > k-t FOCUSS.

The improvement with respect to k-t FOCUSS is also evident from the visual comparison shown in Fig. 6 of a frame from a dataset that has been accelerated by 8. It is nevertheless harder to visually identify the benefit of enforcing TG sparsity in DLTG relative to the DLMRI method. To better understand in what way the TG term modifies the solution of DLMRI, we

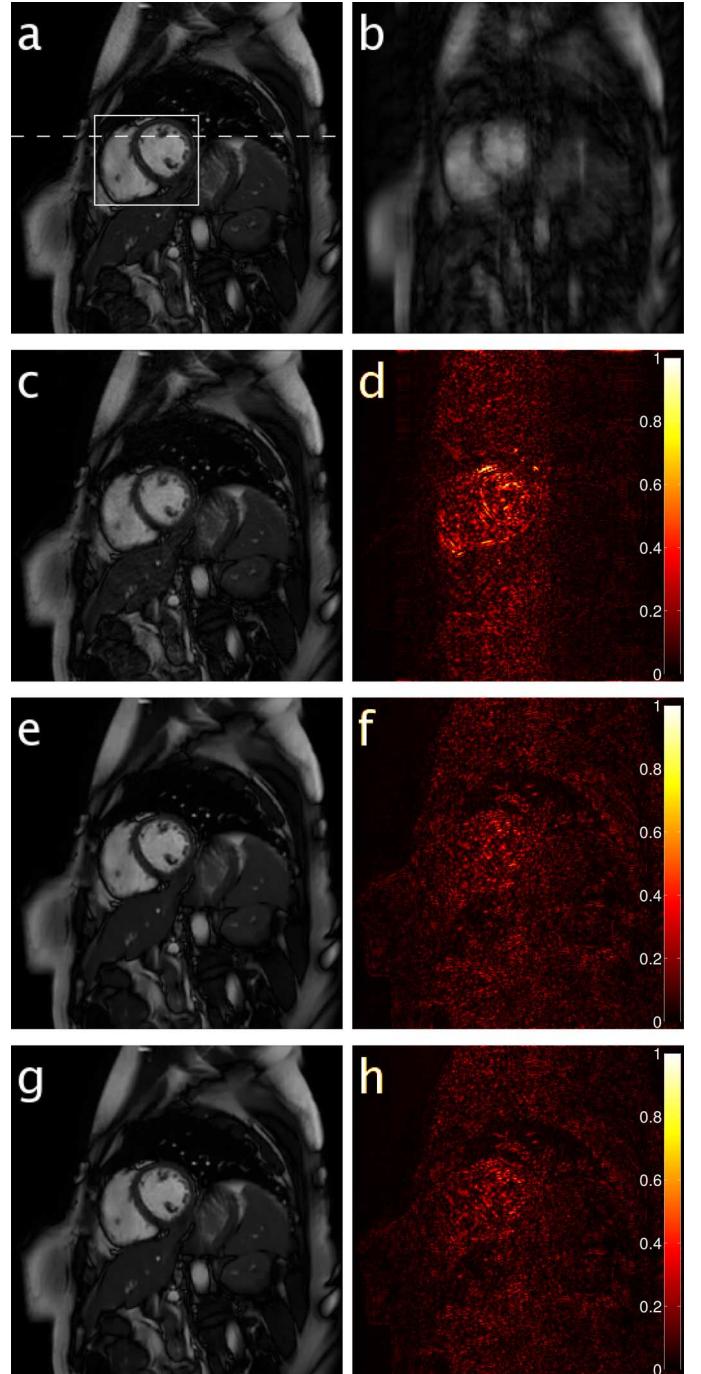


Fig. 6. Visual comparison of a fully sampled magnitude frame (a), its undersampled by eight zero-filled version (b), and reconstructions using k-t FOCUSS (c), DLMRI (e), and DLTG (g) with their respective errors multiplied by 6 (d, f, h).

compare in Fig. 5 the same measure when only a region of interest (ROI) around the heart is considered. An example of this ROI is shown in Fig. 6(a) for one of the scans considered. For this region, reconstruction performances of the DLMRI and the DLTG algorithms are almost the same except at very high acceleration rates.

A visual comparison of the dataset profile in Fig. 8 along the horizontal dashed line of Fig. 6(a) can help explain this

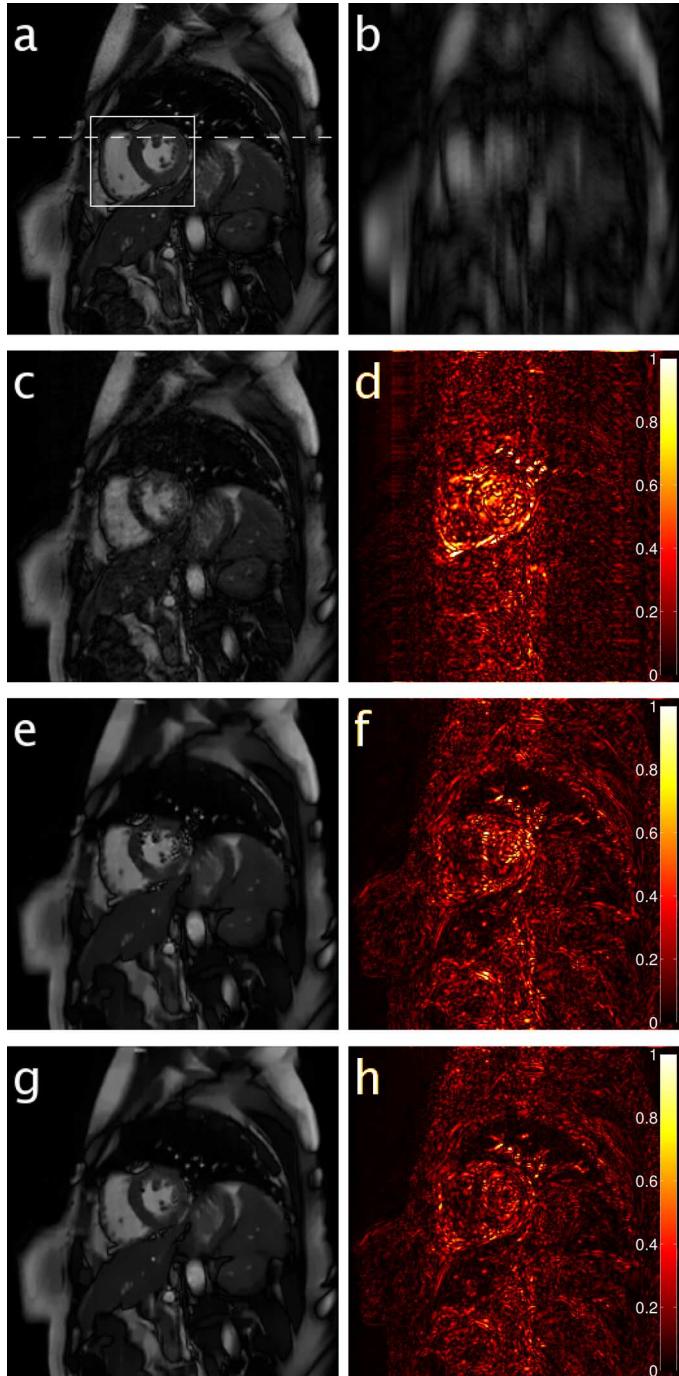


Fig. 7. Visual comparison of a fully sampled magnitude frame (a), its undersampled by 15 zero-filled version (b), and reconstructions using k-t FOCUSS (c), DLMRI (e), and DLTG (g) with their respective errors multiplied by 6 (d, f, h).

finding. The DLTG method slightly smooths the reconstruction out along time in regions of high motion as a result of the TG sparsity enforcement. A small improvement can be appreciated in the static regions outside the ROI (see arrow A), which is part of the reason for the improvement in the global PSNR and MSSIM metrics. We could argue that down to sampling ratios of 0.12, the DLMRI method is already able to recover a faithful representation of the signal and therefore imposing an extra sparsity penalty is liable to distort the solution undesirably

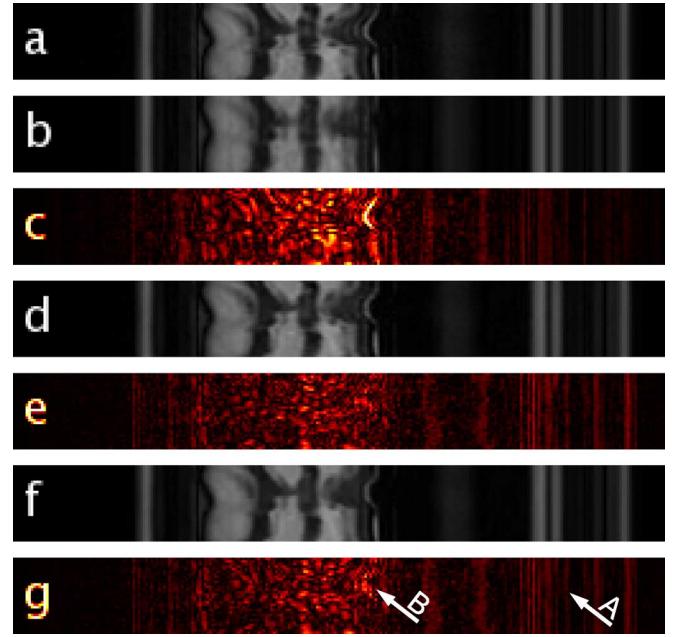


Fig. 8. Temporal profile of row 120 in the original dataset (a), its undersampled by eight zero-filled version (b), and reconstructions using k-t FOCUSS (c), DLMRI (e), and DLTG (g) with their respective errors amplified by 6 (d, f, h).

(see arrow B). However, this behavior is only occasionally perceived and at high acceleration factors the improvement in the reconstruction quality is evident inside and outside the ROI as is shown in Figs. 4 and 5.

Extremely low sampling factors inevitably degrade the results and are a difficult regime in which to carry out a CS based reconstruction, but they force the algorithm to rely heavily on the sparsity model it assumes from the data and can help analyze its suitability. Fig. 7 shows the reconstruction comparison with an acceleration rate of 15, which is the lowest sampling factor tested. Although fine details are lost with the three methods compared, the DLTG method overall provides the most satisfactory reconstruction. The k-t FOCUSS solution contains a lot of aliasing artefacts even though its optimal sparsity model was unrealistically optimized *a posteriori*. The DLMRI method presents a more blocky reconstruction and distorts the natural shape of the heart especially around the myocardium. The DLTG method also contains blocky artefacts, but it noticeably eliminates aliasing and preserves the coarse structure of the original frame.

Fig. 9 compares the temporal profile of the reconstruction at an acceleration rate of 12. The k-t FOCUSS method is not able to capture the dynamism well, and this is a direct consequence of the sparsity model it uses. Sparsity in k-t FOCUSS is imposed on the Fourier transform of pixels along the temporal dimension. If the reconstruction relies heavily on a reconstruction that is too sparse, it will only be able to capture very coarse movement, but fine temporal details will be missed as they are disregarded by the sparse model (see arrow A). The DLMRI result is able to better recover fine temporal changes, but contains a considerable amount of alias that is unresolved (see arrow B). Using the additional TG penalty removes the aliasing at the expense of slightly smoothing the reconstruction along time, as

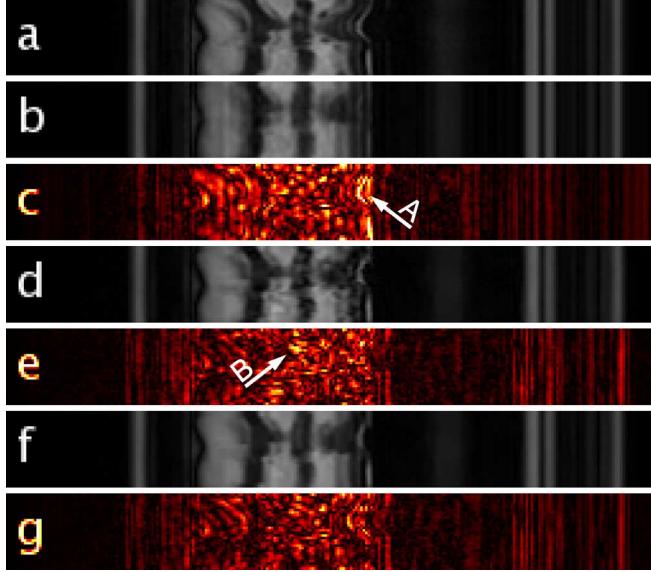


Fig. 9. Temporal profile of row 120 in the original dataset (a), its undersampled by 12 zero-filled version (b), and reconstructions using k-t FOCUSS (c), DLMRI (e), and DLTG (g) with their respective errors amplified by 6 (d, f, h).

shows the DLTG result. Despite this, it is structurally the most faithful reconstruction out of the comparison.

To conclude the analysis, we show in Fig. 10 the original phase and the reconstructed versions of a temporal frame from a dataset accelerated by 8. Except in locations where the amplitude is very low and hence phase is unstable (masked out in the fully sampled case), the three reconstructions are accurate. Quantitatively evaluating phase reconstruction only is difficult because of the many random values when magnitude is zero, but this is implicitly accounted for in Figs. 4 and 5.

C. Reconstruction With AWG Noise

This section analyses the impact of acquisition noise in reconstruction performance. The acquired undersampled data can be described as $\hat{\mathbf{x}}_u = \mathbf{M}(\mathbf{F}\mathbf{x}_f + \mathbf{n})$ where $\mathbf{n} \in \mathbb{C}^P$ is complex circular AWG noise of power σ^2 . The value of the noise regularization parameter $\lambda = q/\sigma$ in the data consistency step of the DLMRI and DLTG algorithms now plays an important role.

In Fig. 11(a), we plot the dependence of the reconstruction quality on parameter q at a sampling factor 0.25. The different input noise values displayed are represented as PSNR_f , which corresponds to the PSNR of the IFFT reconstruction of fully sampled data corrupted by noise. In terms of absolute noise power, $\text{PSNR}_f = \{25.8, 31.8, 35.8, 41.8\}$ dB is equivalent to $\sigma^2 = \{4 \times 10^{-8}, 10^{-8}, 4 \times 10^{-9}, 10^{-9}\}$. The constant q was set to 5×10^{-6} and 5×10^{-5} for the DLTG and DLMRI algorithms respectively based on these results. The same empirical value was found for q in other tests independent of the patch and dataset sizes. For high noise values ($\text{PSNR}_f \leq 31.8$ dB) a fine tuning of this constant is desirable as the reconstruction depends heavily on it, but for lower noise values results are not very sensitive to it.

We plot in Fig. 11(b) the reconstruction performance as a function of PSNR_f for a scan accelerated by a factor of 4. The same rank order between the three methods is preserved in this

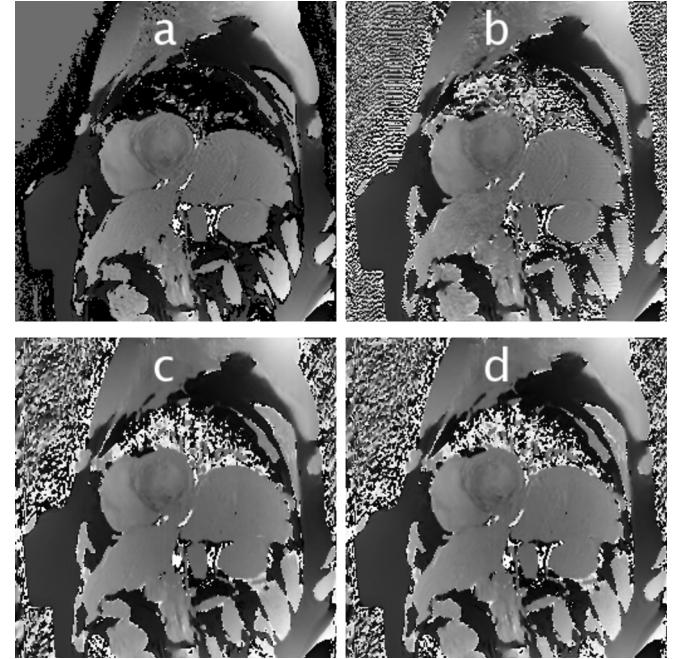


Fig. 10. Phase reconstruction comparison in a scan accelerated by 8 of the original data (a) and the reconstructions provided by k-t FOCUSS, DLMRI, and DLTG (b)–(d).

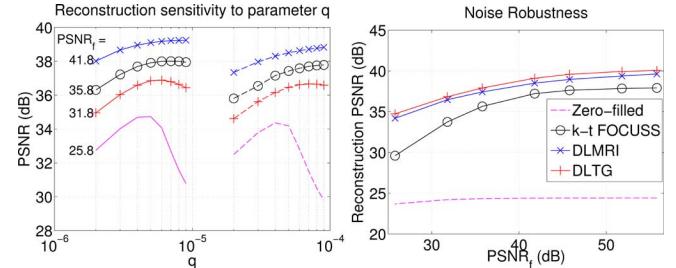


Fig. 11. Algorithm evaluation with added noise. Reconstruction sensitivity to noise power and parameter q for a scan accelerated by 4 (left) using the DLTG (solid) and DLMRI (dashed) algorithms, and their robustness to input noise (right).

test, and both methods using DL present a milder decay of performance at high noise values than k-t FOCUSS. The main reason for this improved robustness can be attributed to the denoising capabilities that have been demonstrated by the K-SVD algorithm. Sparse coding with a trained overcomplete dictionary and averaging overlapping patches is a powerful method to denoise natural images [22]. The visual comparison for $\text{PSNR}_f = 25.8$ dB in Fig. 12 shows how the k-t FOCUSS reconstruction is much more contaminated by noise than the dictionary based reconstructions. Notice that the performance decay at high noise values shown in [26] of DLTG compared to DLMRI is not experienced here. This is because in the present work parameter q is optimized independently for both methods.

D. Spatial Dictionaries

The main objective of this work is to analyze the potential of using spatio-temporal dictionaries for dynamic data instead of independently reconstructing temporal frames with a 2D (spatial) dictionary. In cardiac cine, the temporal dimension is

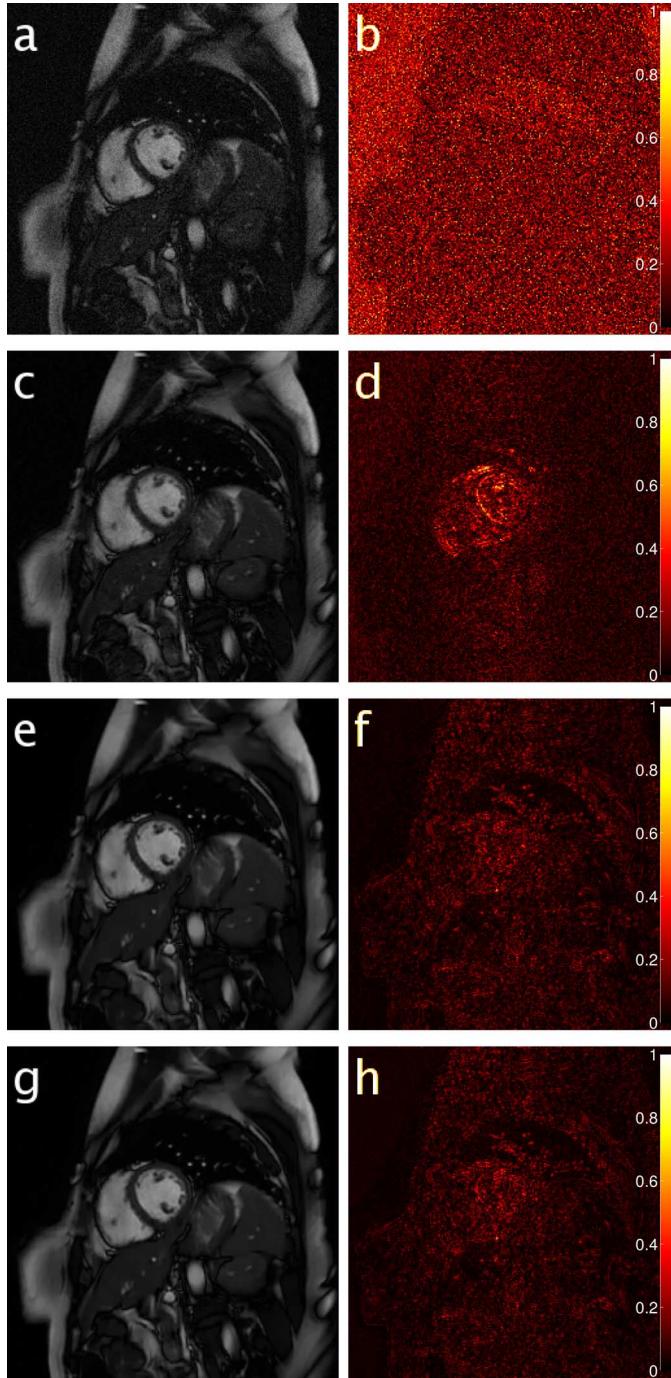


Fig. 12. Visual comparison of the reconstructions from a scan accelerated by 4 that has been contaminated by complex noise with input $\text{PSNR}_f = 25.8 \text{ dB}$ (a) using k-t FOCUSS (c), DLMRI (e), and DLTG (g) with their respective errors multiplied by 6 (d, f, h).

known to be highly redundant as changes through time are slow and confined to specific regions of the image. Hence, a sparsity model that exploits this temporal redundancy can be expected to perform better than one that disregards it.

Fig. 13 visually compares the reconstruction of a six-fold accelerated dataset using the DLMRI algorithm with a spatio-temporal and a spatial dictionary. The reconstruction using a spatio-temporal dictionary shows much better dealiasing properties whereas the reconstruction using the spatial dictionary

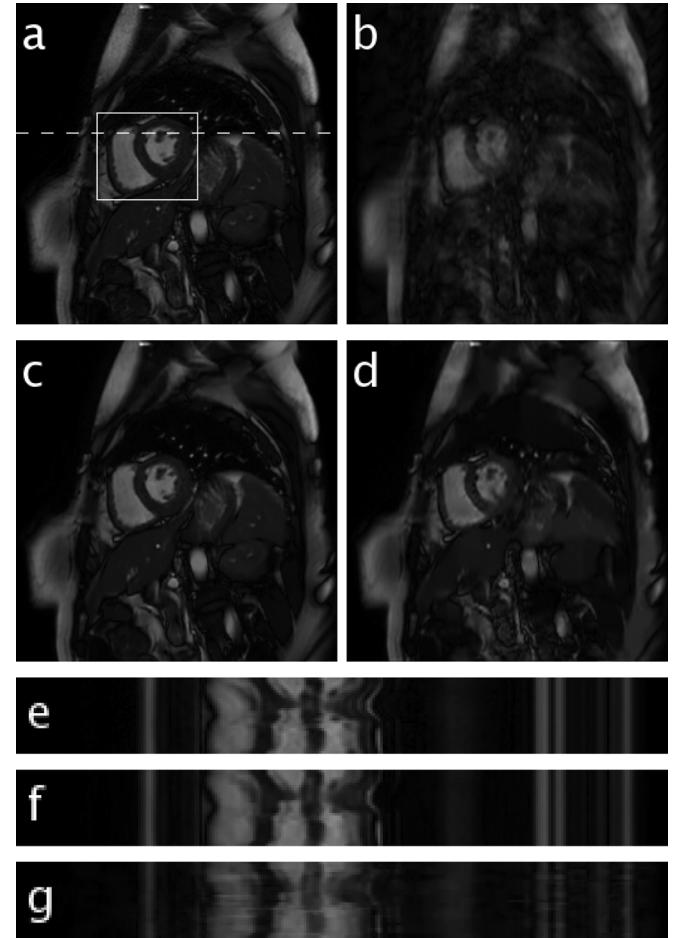


Fig. 13. Comparison of 2D spatial and 3D spatio-temporal reconstruction. Shown are a fully sampled frame (a), its undersampled by six zero-filled version (b), and the spatio-temporal (c) and spatial only (d) reconstructions. Images (e), (f), and (g) are the temporal profiles along the dashed line of datasets (a), (c), and (d), respectively.

has unresolved aliasing and an important loss of structure. The PSNR metric was 38 dB and 27 dB for the spatio-temporal and spatial dictionary reconstruction respectively. The temporal profiles also demonstrate a more accurate reconstruction with the spatio-temporal dictionary. Particularly, the spatial dictionary reconstruction shows a blocky structure that is less consistent across time in terms of both structure and contrast, which is a consequence of reconstructing temporal frames independently.

E. Real and Complex-Valued Dictionaries

The use of real-valued dictionaries is a main difference with respect to previous work on dictionary learning for MRI. In [25], [32], a complex-valued dictionary is trained which is then used to sparsely code complex-valued data. The DLTG algorithm instead trains a real-valued dictionary to sparsely code real and imaginary parts of the data independently. Although training a real-valued dictionary is not expected to outperform a complex-valued dictionary, it is worth understanding the differences between these two strategies.

We look at the error that is produced on a fully sampled dataset when these two different sparsity models are used to

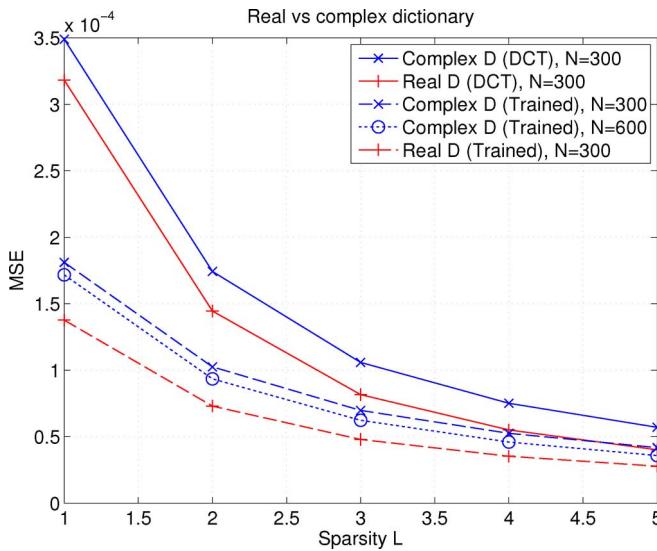


Fig. 14. Average error per patch produced by the assumption of L sparsity on a fully sampled scan using different dictionaries.

approximate it. In Fig. 14, we compare the average error per patch when a scan is coded using a maximum of L atoms from different dictionaries. It is difficult to portray a balanced comparison as the l_0 norm is different for a complex and a real-valued dictionary. The complex-valued dictionary allows for complex-valued sparse representations, meaning that the coding of a patch can use L atoms from the dictionary and their phase can be rotated. This allows for $2L$ degrees of freedom in the sparse representation. On the other hand, the coding using a real-valued dictionary uses L atoms for the real part of a patch and another independent L atoms for the imaginary part, hence matching the $2L$ degrees of freedom of the complex-valued dictionary strategy.

Notice that, out of the dictionaries of $N = 300$ atoms, using a real-valued dictionary entails a slightly smaller average error. This could be explained by the fact that the real-valued dictionary allows the independent representation of real and imaginary parts of a patch, whereas the complex-valued one reconstructs them jointly. Combinatorially, the representation capabilities are higher for the real-valued dictionary, meaning that using this strategy could be a more flexible solution. Its performance can only be matched by the complex-valued strategy if the overcompleteness of the dictionary is increased. The example using $N = 600$ supports this explanation by showing a smaller error with a larger complex-valued dictionary. This difference is nevertheless very small and does not seem to impact the reconstruction process noticeably.

F. Parameter Selection

The choice of algorithmic parameters is crucial for a suitable operation of the DLTG algorithm. The plots in Fig. 15 show the influence of the dictionary and patch size in the reconstruction performance of a scan accelerated by 6. In Fig. 15(a) the patch size was kept at $n = 4 \times 4 \times 4$ and in Fig. 15(b) the dictionary size was fixed at $N = 600$. The quality of the reconstruction is comparable in the broad range of values tested except for very large

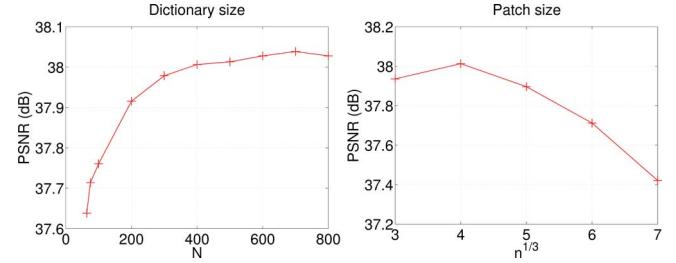


Fig. 15. Influence of dictionary size (left) and patch size (right) on the reconstruction of a scan accelerated by 6.

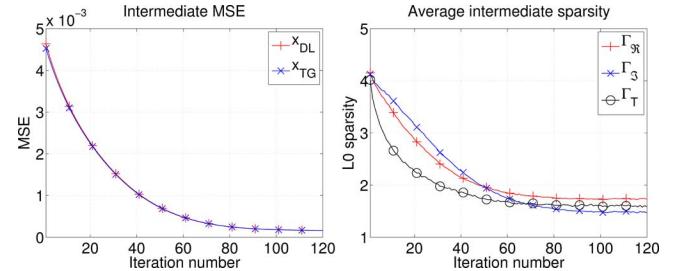


Fig. 16. MSE convergence of approximate results x_{DL} and x_{TG} relative to the ground truth (left), and the intermediate sparsity level of the real, imaginary, and training patches coding.

patch sizes. These tend to oversmooth the result missing out on fine details, which deteriorates the performance. It is also noticeable that above a certain dictionary overcompleteness, there is little improvement in using larger dictionaries.

Algorithm Convergence, Speed, and Acceleration

The intermediate convergence of terms in the global statement is summarized in Fig. 16 for an example reconstruction of a six-fold accelerated scan. Although the convergence of the DLTG algorithm is still to be proven, all the tests undergone showed convergence to a stable result.

The convergence rate, algorithm speed and reconstruction performance is largely dominated by the data consistency parameter ϵ . A large ϵ permits larger discrepancy between training and coding patches, which can be achieved with very few atoms of the dictionary using a very sparse representation. This means that few OMP iterations are needed both for training and coding and the computational load is small. However, the quality of a reconstruction with this parameter can be expected to be low as it will not be closely matched to the target dataset. On the other hand, a small ϵ requires a larger number of dictionary atoms to represent each training and coding patch, which slows down each OMP routine. Nonetheless, this is necessary to tightly adhere the reconstruction to the model and the observed dataset, which is what will ensure near optimal reconstruction.

This behavior is illustrated in Fig. 17(a), where the reconstruction performance of a six-fold accelerated dataset is evaluated in terms of PSNR and convergence rate for different values of ϵ . Fig. 17(b) shows the time needed for one coding step of the entire dataset (both real and imaginary parts), which is the most demanding step in the algorithm. MATLAB R2011b was used for this assessment of the runtime on an Intel Core i7-2600 CPU at 3.4 GHz and 8 GiB of memory. The C++ implementation of

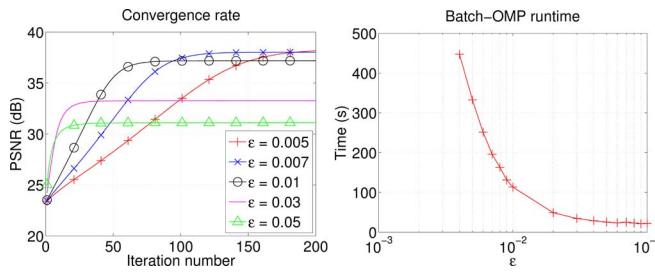


Fig. 17. Example of a reconstruction's dependence to ϵ (left) and the Batch-OMP runtime for one coding of real and imaginary parts of a scan (right).

K-SVD provided by Ron Rubinstein in [42] was used for the Batch-OMP stages.

Fig. 17 allows for an empirical assessment of the global runtime of the entire reconstruction. For instance, most of the results shown in this work were obtained with $\epsilon = 0.007$, which translates into a coding stage of 200 s, and the reconstruction in Fig. 17(a) took about 120 iterations to converge. The bulk of the computation therefore required $200 \times 120 / 3600 = 6.6$ h. Notice that there is nevertheless a lot of flexibility in the speed of the algorithm, as using $\epsilon = 0.01$ would reduce this time to $110 \times 75 / 3600 = 2.3$ h without much compromising the end result.

The use of temporal gradient sparsity in DLTG also has an important impact on the convergence rate compared to the DLMRI version. Fig. 18 shows the convergence of a ten-fold accelerated scan in terms of DLMRI iterations and DLTG outer iterations. Two settings are compared for each algorithm, a slow one using $\epsilon = 0.007$ and $\eta = 4 \times 10^{-4}$, which are the values used for all the results shown above, and a faster but less performant one using $\epsilon = 0.07$ with $\eta = 0.004$. The enforcement of temporal gradient sparsity in DLTG accelerates the process considerably. The reason for this is that the TG transform is able to reduce aliasing from a global perspective in a way that is unfeasible using only the patch based approach of the DLMRI algorithm. More specifically, the high temporal gradient complexity that independent masks produce in consecutive frames is highly penalized by this auxiliary sparsity term.

There is a large difference between these runtimes and the average 15 s required for a k-t FOCUSS reconstruction. Possible acceleration techniques are under review. From the behavior observed in Fig. 17(a) and Fig. 18, it seems natural to seek optimal trajectories for the data fidelity terms ϵ and η such that they start as large values benefiting from fast initial convergence, and are dynamically decreased as the algorithm converges to a stable solution to reach high reconstruction quality. This is a very simple addition that can reduce the total runtime considerably. Decreasing patch sizes and their overlap would also accelerate the algorithm although this could noticeably compromise the end result.

A very appealing modification would be to design a parallel implementation of the 2P sparse coding steps necessary in each iteration of the algorithm, given that they are completely independent from each other. This would dramatically decrease the runtime, because they are the most expensive processing blocks

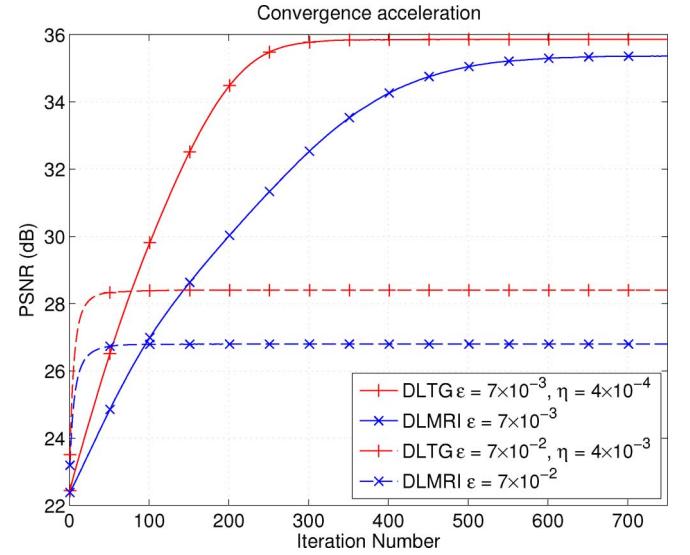


Fig. 18. Convergence rate of DLMRI and DLTG for two different data fitting terms setups.

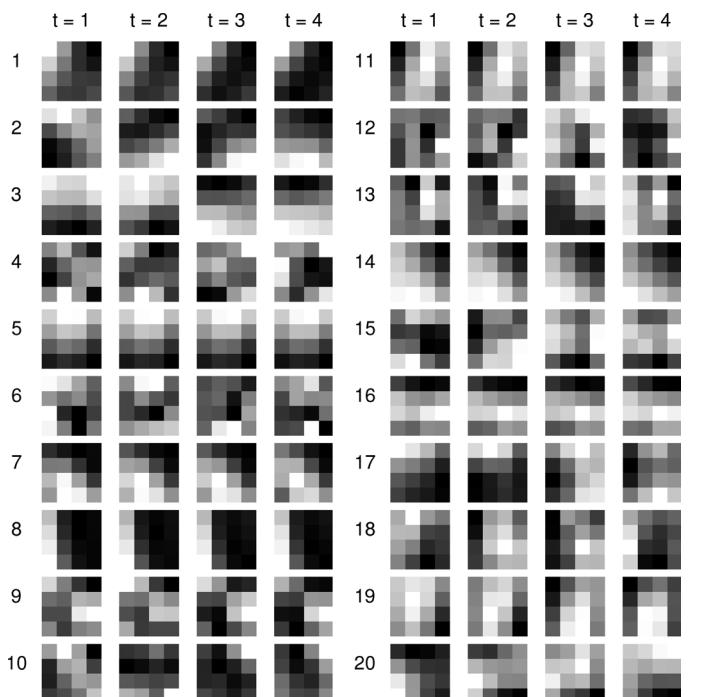


Fig. 19. Examples of spatio-temporal patches learnt in the last iteration of the eight-fold accelerated DLTG reconstruction shown in Fig. 6. Patches are $4 \times 4 \times 4$ patches and are displayed as sequences of spatial patterns for each time frame.

in terms of computation, and would not affect the quality of the reconstruction.

G. Influence of Training Data

The algorithm's behavior depends on the data chosen for dictionary training. The previous results were computed with online learning of a dictionary using patches from the target scan. Fig. 19 shows an example of 20 atoms of the spatio-temporal

TABLE II
TRAINING DATA INFLUENCE WITH $\epsilon = 0.01$ – PSNR (dB)

Training data	Sampling factor		
	0.25	0.16	0.125
DCT (no training)	38.86	36.78	35.22
True (online)	39.20	37.09	35.66
True (offline)	39.20	37.12	35.66
Different - 1 scan (offline)	39.20	37.09	35.65

TABLE III
TRAINING DATA INFLUENCE WITH $\epsilon = 0.007$ – PSNR (dB)

Training data	Sampling factor		
	0.25	0.16	0.125
DCT (no training)	39.8	37.3	36.2
True (online)	40.1	38.0	36.7
True (offline)	40.1	37.8	36.7
Different - 1 scan (offline)	39.9	37.6	36.5

dictionary trained in the last iteration of the DLTG reconstruction in Fig. 6. These look very different from the initial structured DCT dictionary. Some show coarse edges in different directions (atoms 5, 7, 8) whereas others do not have a significant structure and are necessary to capture image details (4, 10, 12). Furthermore, some contain temporal dynamism (1, 3, 20) while others are relatively static (5, 8, 11).

The strategy for dictionary training can also have an impact on the acceleration of the algorithm. Training can be performed offline taking advantage of the large amounts of available high quality medical data. This could be thought of as counterproductive since the aim is to tailor the dictionary to the dataset being reconstructed, and not to a prior scan that could come from a different patient. However, the K-SVD method generally learns coarse features of the data, so if different training datasets are used which are similar to the target dataset in a patch scale, they should provide a similar learning performance.

Table II shows the PSNR quality of reconstructions at accelerations 4, 6, and 8 using different training datasets with $\epsilon = 0.01$. The comparison looks at the initial DCT dictionary without training, the dictionary trained online on the target scan, a dictionary trained offline on the fully sampled target scan, and an offline training on a fully sampled scan from a different subject. As expected, the initial DCT dictionary always provides the worst reconstruction quality and the best one is given by the fully sampled training data on the target scan. Notice however that training data coming from a different cardiac scan can be used to train a dictionary offline and closely approximate the performance of training on the true data.

Table III presents the same comparison using $\epsilon = 0.007$. In this regime, the dictionary is forced to very accurately represent the training data, so the tailoring to a particular scan during training is accentuated. Offline training on a different scan performs slightly worse than using the target scan, but still better than the initial DCT dictionary.

It is important to note that, although dictionary training enhances the reconstruction quality, this increase is relatively small and sometimes might not be noticeable visually. These results are in accordance with the comparisons shown in K-SVD denoising [22], [33], where dictionary training rarely

TABLE IV
DICTIONARY INITIALIZATION – PSNR (dB)

DCT	Data	IID Gaussian
38.0	37.6	36.2

improves PSNR by more than 0.5 dB. The additional computation necessary for training offline, and most importantly online, should be taken into account and questioned as being necessary or not depending on whether the improvement it can bring is clinically relevant.

To conclude, we show in Table IV the influence of the initial dictionary when reconstructing with online training. The test compares a DCT initial dictionary, with one that is initialized with random patches from the data and another that is filled in with independent and identically distributed (IID) Gaussian samples. The DCT initial dictionary provides the best reconstruction whereas the random IID Gaussian initialization performs noticeably worse. This is probably because, although the DCT dictionary is not tailored to a scan, it is still able to provide a sparse representation of it in the first iteration of each dictionary training step, facilitating the task of the K-SVD algorithm. The IID Gaussian initialization will most probably not find a sparse representation of the training data, and this deteriorates the behavior of K-SVD learning.

VII. CONCLUSION

We have presented an algorithm for the acceleration of dynamic MRI acquisition based on a sparsity model that can learn redundancy in the data and an auxiliary constraint on TG sparsity that helps speed up convergence and provides better results at high acceleration rates. The method was tested on data from 10 subjects and provided superior performance in every case. There were no failures and no gross errors. The reconstruction of complex-valued MR data was performed by splitting real and imaginary parts of the sequences and coding them with a single real-valued dictionary. The method proposed is shown to largely outperform the k-t FOCUSS algorithm, which supports a conclusion that a patch based reconstruction has greater potential than global sparsity transforms because they allow higher sparsity representations. Adapting the dictionary to training data can also enhance the reconstruction, although the increase in performance is usually modest.

The use of spatio-temporal dictionaries has proven to have a major advantage with respect to spatial dictionaries in reconstructing cardiac cine data. Imposing sparsity constraints on spatio-temporal patches has two beneficial properties. First, the temporal dimension is expected to be highly redundant in this kind of information, so if it is included in the sparsity model we can expect to create a better scenario for a compressed sensing reconstruction. Secondly, coding spatio-temporal patches implicitly enforces structural and intensity homogeneity across time, which can correct much of the aliasing that spatial dictionaries cannot handle.

Another important finding presented is the influence that the training has upon the reconstruction. It has been shown how dictionary training enhances the reconstruction with respect to an

initial DCT dictionary, but this increase in quality is very limited. This is the same conclusion that could be drawn from the original K-SVD results on denoising [22], [33]. Considering the computational overload that online dictionary training entails, it could be regarded as unnecessary if the reconstruction improvements are barely noticeable to the human eye. Nevertheless, we have also seen how offline training on a dataset from a different patient and with the same imaging setup could provide almost identical results to online training, and this is a viable option that would benefit from the increase in reconstruction quality while keeping the computation cost for training separate from the online runtime.

Future work will aim at extending the method presented such that its practical implementation is made efficient. On top of adding modifications that will accelerate the rate of convergence without degrading the reconstruction quality, the combination of a DL based sparsity model and a parallel MRI reconstruction might have a considerable potential. Currently, joint CS and parallel MRI reconstruction strategies are the ones that provide the best acquisition acceleration results, but only fixed bases sparsity transforms have so far been explored. One could hope that adaptive sparsity transforms would outperform them just like the method described here does with its CSMRI counterpart.

REFERENCES

- [1] C. E. Shannon, "Communication in the presence of noise," in *Proc. Int. Radio Eng.*, Jan. 1949, vol. 37, pp. 10–21, 1.
- [2] B. Madore, G. H. Glover, and N. J. Pelc, "Unaliasing by Fourier-encoding the overlaps using the temporal dimension (UNFOLD), applied to cardiac imaging and fMRI," *Magn. Reson. Med.*, vol. 42, no. 5, pp. 813–828, 1999.
- [3] M. E. Brummer, D. Moratal-Pérez, C.-Y. Hong, R. I. Pettigrew, J. Millet-Roig, and W. T. Dixon, "Noquist: Reduced field-of-view imaging by direct Fourier inversion," *Magn. Reson. Med.*, vol. 51, no. 2, pp. 331–342, 2004.
- [4] Q. Zhao, N. Aggarwal, and Y. Bresler, "Dynamic imaging of time-varying objects," in *Proc. 9th Annu. Meet. ISMRM*, Glasgow, UK, 2001, p. 1776.
- [5] N. Aggarwal, Q. Zhao, and Y. Bresler, "Spatio-temporal modeling and minimum redundancy adaptive acquisition in dynamic MRI," in *Proc. 1st IEEE Int. Symp. Biomed. Imag.*, Washington, DC, Jul. 2002, pp. 737–740.
- [6] Y. Bresler, "Fast acquisition and sampling in MRI: Introduction to time-sequential sampling of spatio-temporal signals," in *Proc. 1st IEEE Int. Symp. Biomed. Imag.*, Washington, DC, Jul. 2002, pp. 713–716.
- [7] N. Aggarwal, S. Bandyopadhyay, and Y. Bresler, "Spatio-temporal modeling and adaptive acquisition for cardiac MRI," in *Proc. 2nd IEEE Int. Symp. Biomed. Imag.*, Washington, DC, Apr. 2004, pp. 628–631.
- [8] N. Aggarwal, B. Sutton, and Y. Bresler, "Adaptive dynamic imaging in cardiac MR fluoroscopy: First in-vivo results," in *Proc. 14th Annu. Meeting ISMRM*, Seattle, WA, May 2006, p. 205.
- [9] N. Aggarwal and Y. Bresler, "Patient-adapted reconstruction and acquisition dynamic imaging method (PARADIGM) for MRI," *Inverse Problems*, vol. 24, no. 4, p. 045015 (29pp), 2008.
- [10] B. Sharif, J. A. Derbyshire, A. Z. Faranesh, and Y. Bresler, "Patient-adaptive reconstruction and acquisition in dynamic imaging with sensitivity encoding (PARADISE)," *Magn. Reson. Med.*, vol. 64, no. 2, pp. 501–513, Aug. 2010.
- [11] J. Tsao, P. Boesiger, and K. Pruessmann, "k-t BLAST and k-t SENSE: Dynamic MRI with high frame rate exploiting spatiotemporal correlations," *Magn. Reson. Med.*, vol. 50, no. 5, pp. 1031–1042, 2003.
- [12] S. J. Malik, S. Schmitz, D. O'Regan, D. J. Larkman, and J. V. Hajnal, "x-f Choice: Reconstruction of undersampled dynamic MRI by data-driven alias rejection applied to contrast-enhanced angiography," *Magn. Reson. Med.*, vol. 56, no. 4, pp. 811–823, Oct. 2006.
- [13] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [14] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [15] 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.
- [16] H. Jung, J. C. Ye, and E. Y. Kim, "Improved k-t BLAST and k-t SENSE using FOCUSS," *Phys. Med. Biol.*, vol. 52, no. 11, pp. 3201–3226, Jun. 2007.
- [17] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly, "k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proc. 14th Annu. Meet. ISMRM*, Seattle, WA, 2006, p. 2420.
- [18] L. B. Montefusco, D. Lazzaro, S. Papi, and C. Guerrini, "A fast compressed sensing approach to 3D MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1064–1075, May 2011.
- [19] G. Adluru, C. McGann, P. Speier, E. Khomovski, A. Shaaban, and E. Dibella, "Acquisition and reconstruction of undersampled radial data for myocardial perfusion magnetic resonance imaging," *J. Magn. Reson. Imag.*, vol. 29, no. 2, pp. 466–473, 2009.
- [20] M. Usman, C. Prieto, T. Schaeffter, and P. G. Batchelor, "k-t Group sparse: A method for accelerating dynamic MRI," *Magn. Reson. Med.*, vol. 66, no. 4, pp. 1163–1176, Oct. 2011.
- [21] I. Tasic and P. Frossard, "Dictionary learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, 2011.
- [22] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
- [23] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *Multiscale Model. Simulat.*, vol. 7, no. 1, pp. 214–241, 2008.
- [24] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. 23rd IEEE Conf. CVPR*, San Francisco, CA, Jun. 2010, pp. 3501–3508.
- [25] S. Ravishankar 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.
- [26] J. Caballero, D. Rueckert, and J. V. Hajnal, "Dictionary learning and time sparsity in dynamic MRI," in *Proc. 15th Int. Conf. MICCAI*, Nice, France, Oct. 2012, vol. 1, pp. 256–263.
- [27] S. P. Awate and E. V. R. Di Bella, "Spatiotemporal dictionary learning for undersampled dynamic MRI reconstruction via joint frame-based and dictionary-based sparsity," in *Proc. 9th IEEE Int. Symp. Biomed. Imag.*, Barcelona, Spain, 2012, pp. 318–321.
- [28] M. Doneva, P. Börnert, H. Eggers, C. Stehning, J. Sénégas, and A. Mertins, "Compressed sensing reconstruction for magnetic resonance parameter mapping," *Magn. Reson. Med.*, vol. 64, no. 4, pp. 1114–1120, 2010.
- [29] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [30] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Ap. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [31] J. A. Tropp, Anna, and C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [32] Y. Wang, L. Ying, and Y. Zhou, "Undersampled dynamic magnetic resonance imaging using patch-based spatiotemporal dictionaries," in *Proc. 10th IEEE Int. Symp. Biomed. Imag.*, San Francisco, CA, Apr. 2013, pp. 294–297.
- [33] M. Protter and M. Elad, "Image sequence denoising via sparse and redundant representations," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 27–35, Jan. 2009.
- [34] S. G. Lingala and M. Jacob, "A blind compressive sensing frame work for accelerated dynamic MRI," in *Proc. 9th IEEE Int. Symp. Biomed. Imag.*, Barcelona, Spain, May 2012, pp. 1060–1063.
- [35] S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic MRI," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1132–1145, Jun. 2013.
- [36] S. G. Lingala and M. Jacob, "Blind compressed sensing with sparse dictionaries for accelerated dynamic MRI," in *Proc. 10th IEEE Int. Symp. Biomed. Imag.*, San Francisco, CA, Apr. 2013, pp. 5–8.
- [37] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty, "An efficient algorithm for compressed MR imaging using total variation and wavelets," in *Proc. 21st IEEE Conf. CVPR*, Anchorage, AK, Jun. 2008, pp. 1–8.

- [38] J. Yang, Y. Zhang, and W. Yin, "A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 288–297, Apr. 2010.
- [39] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: Design of dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, pp. 4311–4322, Nov. 2006.
- [40] A. Chambolle, "Total variation minimization and a class of binary MRF models," in *Proc. 5th Int. Workshop EMMCVPR*, St. Augustine, FL, 2005, vol. 3757, pp. 136–152.
- [41] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit". Dept. Comput. Sci., Technion–Israel Inst. Technol., Tech. Rep, 2008.
- [42] R. Rubinstein, KSVD-Box v13 and OMP-Box v10 Dept. Comput. Sci., Technion–Israel Inst. Technol., 2012 [Online]. Available: <http://www.cs.technion.ac.il/~ronrubin/software.html>
- [43] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn. Reson. Med.*, vol. 42, no. 5, pp. 952–962, 1999.
- [44] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [45] M. Lustig, J. H. Lee, D. L. Donoho, and J. M. Pauly, "Faster imaging with randomly perturbed, undersampled spirals and L1 reconstruction," in *Proc. 13th Annu. Meet. ISMRM*, Miami Beach, FL, 2005, p. 685.
- [46] 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, no. 6, pp. 1086–1098, Jun. 2007.
- [47] J. A. Fessler, "On NUFFT-based gridding for non-Cartesian MRI," *J. Magn. Reson.*, vol. 188, no. 2, pp. 191–195, Oct. 2007.
- [48] M. Markl, K. Scheffler, and O. Bieri, "Analysis and compensation of eddy currents in balanced SSFP," *Magn. Reson. Med.*, vol. 54, no. 1, pp. 129–137, Jul. 2005.
- [49] Dynamic MRI KAIST, Bio Imaging Processing Lab, 2012 [Online]. Available: <http://bisip.kaist.ac.kr/ktFOCUSS.htm>