Adaptive Sparsity Level and Dictionary Size Estimation for Image Reconstruction in Accelerated 2D Radial Cine MRI

Marie-Christine Pali[1], Tobias Schaeffter[2,3,4], Christoph Kolbitsch[2,3], Andreas Kofler[2]

[1]Department of Mathematics, University of Innsbruck, Innsbruck, Austria
[2]Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany
[3]Division of Imaging Sciences and Biomedical Engineering, Kings College London, London, UK
[4]Department of Medical Engineering, Technical University of Berlin, Berlin, Germany

Version typeset June 18, 2020

marie-christine.pali@uibk.ac.at

Abstract

Purpose: In the past, Dictionary Learning (DL) and Sparse Coding (SC) have been proposed for the regularization of image reconstruction problems. The regularization is given by a sparse approximation of all image-patches using a learned dictionary, i.e. an overcomplete set of basis functions learned from data. Despite its competitiveness, DL and SC require the tuning of two essential hyper-parameters: the sparsity level S - the number of basis functions of the dictionary, called atoms, which are used to approximate each patch, and K - the overall number of such atoms in the dictionary. These two hyper-parameters usually have to be chosen a-priori and are determined by repetitive and computationally expensive experiments. Further, the final reported values vary depending on the specific situation. As a result, the clinical application of the method is limited, as standardized reconstruction protocols have to be used.
Methods: In this work, we use adaptive DL and propose a novel adaptive sparse coding algorithm for 2D radial cine MR image reconstruction. Using adaptive DL and adaptive SC, the optimal dictionary size K as well as the optimal sparsity level S are chosen dependent on the considered data.
Results: Our three main results are the following: First, adaptive DL and adaptive SC deliver results which are comparable or better than the most widely used non-adaptive version of DL and SC. Second, the time needed for the regularization is accelerated due to the fact that the sparsity level S is never overestimated. Finally, the a-priori choice of S and K is no longer needed but is optimally chosen dependent on the data under consideration.
Conclusions: Adaptive DL and adaptive SC can highly facilitate the application of DL- and SC-based regularization methods. While in this work we focussed on 2D radial cine MR image reconstruction, we expect the method to be applicable to different imaging modalities as well.

Keywords: Adaptive Dictionary Learning, Adaptive Sparse Coding, Compressed Sensing, Radial Cine MRI, Unsupervised Learning, Parameter Estimation

Contents

I. Introduction

Magnetic Resonance Imaging (MRI) has become nowadays an indispensable imaging modality which is widely used in daily clinical routine to image the interior of a patient. For example, cardiac cine MRI can be used for the assessment of the cardiac function. For that, a slice of the patient's heart is scanned over multiple cardiac cycles and a sequence of 2D images showing the heart movement can be obtained. However, a major issue of MRI is the slow data-acquisition process due to physical limits imposed by the scanner. In particular, typical cardiac MR-scans are performed during a breathhold to avoid respiratory motion artefacts. Therefore the breathhold duration limits the spatial and temporal resolution of MR-scans, which represents a problem for ill patients with limited breathhold capabilities. The data-acquisition in MRI takes place in the so-called k-space, i.e. the Fourier space. Since the acquisition is often slow, undersampling in k-space is used to shorten scan times. This leads to undersampling artefacts due to the violation of the Nyquist sampling limit. Parallel imaging and regularized iterative reconstruction methods have been proposed to minimize undersampling artefacts, e.g.[1]. Regularization approaches using transforms learned from data, i.e. Dictionary Learning (DL), and sparse coding (SC) have been considered in the past[2],[3],[4],[5],[6],[7],[8],[9]. In DL-based regularization, the model assumption is patch-wise sparsity and therefore, the idea is to patch-wise impose the regularization on the image to be reconstructed.

The rationale behind the regularization based on learned dictionaries is that patches of an image have an inherently low-dimensional representation. DL aims to find building blocks (i.e. the basis functions) of such a representation in an unsupervised manner based on the given patches. SC then aims to find a sparse (low-dimensional) representation of a target patch with respect to this dictionary. The regularization of the solution is achieved by the fact that, given the incoherent undersampling scheme applied in k-space, the artefacts resulting from the direct reconstruction of an image are high-dimensional and thus suppressed by the low dimensional representation, which suffices to capture the important features.

In[2], for example, a pre-trained dictionary is used to regularize the images. Further, approaches in which the dictionary is learned from the current image estimate during the reconstruction have been proposed[3],[4] and successfully applied to cine MR image reconstruction[5],[6]. However, regardless of the excellent image quality which can be achieved by the latter mentioned methods, there still remain a few issues. First, the sparsity level S used for DL and SC as well as the number of atoms in the dictionary K need to be chosen a-priori and are typically determined by repeating the experiments for different choices of S and K. However, the parameters are clearly data-dependent and there is no guaranty on the achievable performance of the reconstruction algorithms on different datasets. Second, performing an S-sparse approximation of all image patches is computationally quite expensive, especially when S is chosen relatively high. These two issues make the method prohibitive for the application in the clinical routine where standardized reconstruction protocols have to be used.

In this work, to overcome the problem given by the computational complexity of the DL- and sparse coding (SC)-stage as well as the need for choosing the hyper-parameters S and K, we propose to use adaptive versions of DL and SC algorithms. While in other works the concept of adaptivity has been already introduced for the task of image recovery[10] and image reconstruction[11], these works only address the adaptive choice of the parameter which controls the contribution of the regularization using pre-defined sparsifying transforms, see for example[11,12]. In some well-known works, e.g.[4],[5] and[6], the authors refer to adaptive dictionary learning in the sense that the dictionary is learned during the reconstruction. In contrast, our notion of adaptivity refers to the adaptive choice of the sparsity level S and the number of atoms K in the dictionary based on the considered data. Adaptive sparse coding of signals has been previously considered mainly in the signal processing community, see e.g.[13],[14],[15],[16]. Our adaptive OMP is based on the selection of the atoms using thresholding which is similar to[13] and[15]. However, while[13] and[15] require the careful tuning of a hyper-parameter, in our case, the equivalent hyper-parameter is selected based on the dictionary size K. Therefore, to the best of our knowledge, this is the first work using adaptivity of the sparsity level S and the number of atoms K of the learned dictionary for the task of image reconstruction in MRI and therefore substantially differs from previous works.

The paper is structured as follows. In Section II.A. the reconstruction problem using the DL-based regularization technique is described and the general concepts of DL and SC are briefly revised and outlined. Section II.B. contains the main part of the work where we describe adaptive versions of ITKrM and OMP and their advantage over non-adaptive DL and SC algorithms. We conduct different experiments in Section II.C. which we discuss in Section IV. and then conclude the work with a summary in Section V..

II. Materials and Methods

II.A. Problem Formulation and Dictionary Learning-based Regularization Approaches

Mathematically, the process of undersampling can be formulated as applying a binary mask $\mathbf{S}_I$ to the measured Fourier data. Let $\mathbf{x} \in \mathbb{C}^N$ denote the vector representation of the unknown cine MR image with $N = N_x \cdot N_y \cdot N_t$, where $N_x \times N_y$ is the shape of a single 2D image and N_t corresponds to the number of cardiac phases. Let $\mathbf{F}$ denote the encoding operator and $I \subset J = \{1, \ldots, N\}$ the set of Fourier coefficients which are needed to properly reconstruct the image $\mathbf{x}$. The inverse problem one aims to solve is of the form

$$\mathbf{y}_I = \mathbf{F}_I \mathbf{x} + \mathbf{e}, \tag{1}$$

where $\mathbf{F}_I := \mathbf{S}_I \circ \mathbf{F}$ and $\mathbf{e}$ denotes random noise. Images directly reconstructed from undersampled k-space by applying the adjoint operator $\mathbf{F}_I^{\mathsf{H}}$ contain severe artefacts which limit the diagnostic quality. Since by discarding non-measured data problem (1) becomes underdetermined, there is an infinite number of solutions and, in order to constrain the space of solutions of interest, regularization techniques must be used. When DL and SC are used as a regularization method, possible formulations of the image reconstruction problem are the ones of joint minimization problems

$$\min_{\mathbf{x}, \{\boldsymbol{\gamma}_j\}_j} \|\mathbf{F}_I \mathbf{x} - \mathbf{y}_I\|_2^2 + \frac{\lambda}{2} \sum_j \left(\|\mathbf{R}_j \mathbf{x} - \boldsymbol{\Psi} \boldsymbol{\gamma}_j\|_2^2 + \|\boldsymbol{\gamma}_j\|_0 \right), \tag{P1}$$

see e.g.[2], or

$$\min_{\mathbf{x}, \boldsymbol{\Psi}, \{\boldsymbol{\gamma}_j\}_j} \|\mathbf{F}_I \mathbf{x} - \mathbf{y}_I\|_2^2 + \frac{\lambda}{2} \sum_j \left(\|\mathbf{R}_j \mathbf{x} - \boldsymbol{\Psi} \boldsymbol{\gamma}_j\|_2^2 + \|\boldsymbol{\gamma}_j\|_0 \right), \tag{P2}$$

see e.g.[5] and[6]. Here, $\mathbf{x}$ denotes the unknown solution, $\mathbf{y}_I$ the measured undersampled acquired k-space data, λ a regularization parameter, and $\|\boldsymbol{\gamma}_j\|_0$ counts the number of non-zero coefficients in $\boldsymbol{\gamma}_j$. The operator $\mathbf{R}_j$ extracts the j-th 3D patch from the image $\mathbf{x}$, $\boldsymbol{\Psi}$ denotes the dictionary, i.e. a set of K unit norm vectors also referred to as atoms, and $\boldsymbol{\gamma}_j$ the sparse representation of the patch $\mathbf{R}_j \mathbf{x}$ with respect to $\boldsymbol{\Psi}$. The difference between (P1) and (P2) is that in (P1), one assumes to have a pre-trained dictionary $\mathbf{\Psi}$, while in (P2), the dictionary $\mathbf{\Psi}$ is learned during the reconstruction based on the current image estimates. Note that in[6] and[5], a TV term is further added to the minimization problem (P2). However, since in this work we focus on the DL component of the reconstruction, we neglect the additional TV-regularization term. Problems (P1) and (P2) can be solved by the alternating direction method of multipliers (ADMM) which alternates between the minimization with respect to $\mathbf{x}$, the dictionary $\mathbf{\Psi}$ and the set of vectors $\{\boldsymbol{\gamma}_j\}_j$. Usually, the starting point for the iterative reconstruction algorithm is given by the direct reconstruction from the measured data, that is $\mathbf{x}_I = \mathbf{F}_I^{\mathsf{H}}\mathbf{y}_I$.

II.A.1. Dictionary and Sparse Code Update

Assuming a fixed $\mathbf{x}$, the minimization of (P1) and (P2) is achieved by solving the problems

$$\min_{\{\boldsymbol{\gamma}_j\}_j} \sum_j \left(\|\mathbf{R}_j\mathbf{x} - \mathbf{\Psi}\boldsymbol{\gamma}_j\|_2^2 + \|\boldsymbol{\gamma}_j\|_0 \right) \tag{2}$$

and

$$\min_{\mathbf{\Psi}, \{\boldsymbol{\gamma}_j\}_j} \sum_j \left(\|\mathbf{R}_j\mathbf{x} - \mathbf{\Psi}\boldsymbol{\gamma}_j\|_2^2 + \|\boldsymbol{\gamma}_j\|_0 \right), \tag{3}$$

respectively. Problem (2) is solved with any SC algorithm, while (3) is typically solved using an alternating minimization procedure, which alternates between DL to obtain $\mathbf{\Psi}$ and SC to obtain the set of vectors $\{\boldsymbol{\gamma}_j\}_j$. The choice of the algorithms used for training the dictionary $\mathbf{\Psi}$ and obtaining the sparse codes $\{\boldsymbol{\gamma}_j\}_j$ is the main focus of this work and will be discussed later.

II.A.2. Reconstruction Update

Assuming a fixed dictionary $\mathbf{\Psi}$ and a fixed set of sparse codes $\{\boldsymbol{\gamma}_j\}_j$, one can easily see that minimizing (P1) or (P2) with respect to $\mathbf{x}$ is equivalent to solving the system of linear equations

$$\mathbf{H}\mathbf{x} = \mathbf{b}, \tag{4}$$

where the operator $\mathbf{H}$ is given by

$$\mathbf{H} = \mathbf{F}_I^{\mathsf{H}}\mathbf{F}_I + \lambda \sum_j \mathbf{R}_j^{\mathsf{T}}\mathbf{R}_j, \tag{5}$$

and the right-hand-side $\mathbf{b}$ is given by a linear combination of the initial reconstruction $\mathbf{x}_I$ and an image which corresponds to a properly averaged combination of all patches $\mathbf{\Psi}\boldsymbol{\gamma}_j$, i.e.

$$\mathbf{b} = \mathbf{F}_I^{\mathsf{H}}\mathbf{y}_I + \lambda \sum_j \mathbf{R}_j^{\mathsf{T}}\mathbf{\Psi}\boldsymbol{\gamma}_j. \tag{6}$$

Since in general, the inversion of the operator $\mathbf{H}$ is computationally not feasible, problem (4) is solved using an iterative method. Given that $\mathbf{H}$ is symmetric, a common choice for the solver is the pre-conditioned conjugate gradient method[17].

II.A.3. Notation

In the following, for conciseness, we denote the vectorised patches extracted from an image by $y_n \in \mathbb{R}^d$ and call them signals. By I_n we denote the optimal sparse support of a signal y_n, i.e. the set of indices of the non-zero coefficients of the corresponding sparse vector, and by I_n^t the support obtained by thresholding. By $|I_n|$ we denote the cardinality of I_n, by $\mathbf{\Psi}_{I_n}$ the restriction of the dictionary $\mathbf{\Psi}$ to the atoms indexed by $i \in I_n$ and by $\mathbf{\Psi}_{I_n}^{\dagger}$ the pseudo inverse of $\mathbf{\Psi}_{I_n}$. By S we denote the sparsity of a signal y_n, i.e. the cardinality of its support I_n. The coherence of the dictionary $\mathbf{\Psi}$, i.e. the maximal absolute inner product between two different atoms, is denoted by $\mu(\mathbf{\Psi}) := \max_{i\neq j} |\langle \psi_i, \psi_j \rangle|$.

II.A.4. ITKrM Algorithm

One of the probably most popular and widely used DL algorithms is K-SVD (K-Singular Value Decomposition) introduced in[18]. While K-SVD yields meaningful representations in practice, a big drawback is its computational complexity. To circumvent this issue, the Iterative Thresholding and K-residual Means (ITKrM) algorithm was introduced in[19],[20], which like K-SVD, belongs to the class of alternating optimization algorithms. In contrast to K-SVD, ITKrM alternates between updating the sparse support using the much cheaper thresholding procedure instead of Orthogonal Matching Pursuit (OMP)[21] and updating the dictionary by calculating K residual means instead of calculating K singular value decompositions. In particular, in each iteration of ITKrM, for each signal y_n we calculate the thresholded support I_n^t and the residual $a_n = y_n - \mathbf{\Psi}_{I_n^t}\mathbf{\Psi}_{I_n^t}^{\dagger} y_n$, which captures the remaining signal energy and is used for the atom update,

$$\bar{\psi}_k = \sum_{n:k\in I_n^t} \big[a_n + \psi_k\langle\psi_k, y_n\rangle\big] \cdot \mathrm{sign}(\langle\psi_k, y_n\rangle). \tag{7}$$

Obviously, ITKrM exhibits a much lower computational complexity than K-SVD and, despite the fact that it is much simpler, was reported to yield similar results[19].

II.B. Proposed Adaptive Dictionary Learning and Adaptive Sparse Coding Algorithms

A difficulty which comes along with all popular DL algorithms is that the sparsity level S and dictionary size K (in terms of an initial dictionary) have to be chosen a-priori as input parameters. In applications such as image restoration, one typically chooses S and K empirically or experimentally. For instance, for d-dimensional signals, typical values are $d \leq K \leq 4d$ and $S = \sqrt{d}$, but depending on the situation they can highly vary and, as we will show later, they might have a significant impact on the reconstruction quality as well as the required computational time.

To circumvent this issue, a modification of ITKrM was introduced in[22], where S and K are adapted in each iteration. In the following, we briefly review the main ideas used to incorporate adaptivity into ITKrM, yielding its adaptive version aITKrM. For the interested reader, we refer to[22] for an extensive discussion of the introduced concepts, the algorithm and a `matlab`-toolbox. Inspired by some of the ideas used for adapting the sparsity level S, we further introduce adaptivity in the SC stage. In particular, we propose an adaptive version of OMP where not only the sparsity level S is chosen adaptively but which will also turn out to significantly accelerate the SC procedure.

II.B.1. Adaptive Dictionary Learning

In the following, we briefly describe the concept of adaptivity of the DL stage. A basic ingredient for the convergence of ITKrM is that the current estimate of the dictionary is not too coherent. Therefore, in order to avoid this, a replacement procedure and a strategy for finding good replacement candidates was introduced in[22], leading to a version of ITKrM where not only the learned dictionary exhibits good properties but also the dictionary size K and the sparsity level S are adapted in each iteration.

Concretely, for adapting the dictionary size, the replacement strategy, resulting from an analysis of the convergence behaviour of ITKrM, is separated into pruning of coherent and unused atoms and adding promising replacement candidates. This modification hence yields an improved dictionary and allows both increasing and decreasing the dictionary size. In particular, two atoms are considered too coherent if their inner product in absolute value is above a certain threshold $\mu_{\max}$. If this is the case, the less often used one is deleted or they are merged. To decide which atoms are useless one has to count how often an atom has been selected and additionally to check if its corresponding coefficient is larger than a certain threshold. Considering also the size of the coefficients prevents that we keep atoms representing noise, as coefficients corresponding to these atoms are small. If the number of times such an atom has been used is smaller than the minimal number of observations M, this atom is deleted. The same strategy is used to decide whether a well designed replacement candidate should be added or not, meaning one has to check if it is useful and if it is incoherent enough to all atoms which are already in the dictionary. Note that M only depends on the input data and can therefore be estimated[22]. Hence, the only parameter which has to be chosen is the maximal allowed coherence between two atoms $\mu_{\max}$. Compared to choosing S or K this is much simpler as it only determines how similar two atoms in our dictionary are allowed to be. For any dictionary $\mathbf{\Psi}$, we have $\mu(\mathbf{\Psi}) \in [0,1]$, where for example $\mu(\mathbf{\Psi}) = 0$ indicates that we haven an orthonormal basis and $\mu(\mathbf{\Psi}) = 1$ means that we have one atom twice.

The idea behind adaptively choosing S is to start learning the dictionary $\mathbf{\Psi}$ with estimated sparsity level $S_e = 1$ as each signal can be interpreted as being 1-sparse (with probably enormous error) in $\mathbf{\Psi}$. Yielding a reasonable first estimate of most dictionary atoms, one proceeds by carefully increasing, decreasing or keeping S_e the same, depending on the size of the estimated average sparsity level $\bar{S}$. More precisely, in each iteration of aITKrM, the sparsity level S_n of each signal y_n is estimated as the number of its squared coefficients $(|(\mathbf{\Psi}^\dagger_{I_n^t} y_n)(i)|^2)_{i \in I_n^t}$ and residual inner products with the dictionary $(|\langle \psi_i, a_n \rangle|^2)_{i \notin I_n^t}$ that are larger than some threshold θ^2 times the residual energy $\|a_n\|_2^2$. Note that this threshold is computed within the algorithm and has not be given as input parameter. If the average of these estimated sparsity levels $\bar{S} = \lfloor \frac{1}{N} \sum_n S_n \rceil$, is larger than S_e, this indicates that the current estimate S_e is too small and has to be increased by one, if $\bar{S} = S_e$, it is kept the same and if $S_e > \bar{S}$, S_e has to be decreased by one.

II.B.2. **Adaptive Sparse Coding**

As a next step, we introduce adaptivity into the SC procedure. In particular, we propose a version of OMP where the sparsity level S is no longer needed as input parameter but adaptively chosen for each signal. As we will demonstrate later, the sparsity level of an image can vary from position to position, meaning, each image-patch can have a different optimal sparsity level. More precisely, depending on the texture of each image patch, we have higher or lower S, hence, suggesting to introduce an adaptive choice of S in the SC step.

In order to incorporate adaptivity into OMP, we replace the condition of stopping after adding at most S atoms by a bound for the maximal inner product between any atom and the current residual. More precisely, in each iteration, we check if there exists an atom ψ_k for which the absolute value of the residual inner product $|\langle \psi_k, a_n \rangle|$ is larger than some threshold times the norm of the residual. The index corresponding to the atom yielding the largest inner product is then selected. Projecting the signal onto the span of already selected atoms and calculating the new residual, this procedure is repeated until the stopping condition is met. A suitable threshold is obtained using concentration of measure. More precisely, we want aOMP to stop if the residual consists only of noise. For that, assume our current residual is of the form $a_n = r$, where r denotes a Gaussian noise vector, and for the current support $|I_n| = S$. The expected number of remaining atoms for which the residual inner product is larger than $\tau \|r\|_2$ can be calculated as

$$\sum_{k \notin I_n} \mathbb{P}\big(|\langle \psi_k, r \rangle| > \tau \|r\|_2\big) < 2(K - S) \exp\left(-\frac{d\tau^2}{2}\right), \tag{8}$$

which for $\tau = \sqrt{2\log(4K)/d}$ is smaller than $\frac{1}{2}$. Inequality (8) is the main ingredient of the algorithm as it provides a proper threshold τ which is used as stopping condition in aOMP and prevents aOMP from overfitting the considered patch by discarding noise.

To further accelerate aOMP, we introduce a preliminary step where we select the 'strongest' part of the support. In particular, before always adding the next best fitting atom (one at a time) we will choose part of the support by thresholding with $\tau_1 = \sqrt{2\log(8K)/d}$, meaning we choose several atoms at a time, while having the previous expectation smaller than $\frac{1}{4}$. This partial support is subsequently refined/expanded by proceeding aOMP until one of the stopping conditions is met. A summary of the proposed algorithm can be found in Algorithm 1.

Algorithm 1: Proposed adaptive Orthogonal Matching Pursuit (aOMP)

Input: $\mathbf{\Psi}, \mathbf{Y}$ `//dictionary, signals`
Initialise: $\Gamma = 0$ `//` $d \times N$ `matrix`
$\tau_1 = \sqrt{2\log(8K)/d}$ `//thresholds`
$\tau_2 = \sqrt{2\log(4K)/d}$
foreach n **do**
  $I_n^t = \arg\text{where}\big(|\langle \psi_k, y_n\rangle| > \tau_1 \cdot \|y_n\|_2\big)$
  $a_n = y_n - P(\mathbf{\Psi}_{I_n^t})y_n$
  while $\max_k |\langle \psi_k, a_n\rangle| > \tau_2 \cdot \|a_n\|_2$ **do**
    $I_n^t = I_n^t \cup \arg\max_k |\langle \psi_k, a_n\rangle|$
    $a_n = y_n - P(\mathbf{\Psi}_{I_n^t})y_n$
  $\Gamma[I_n^t, n] = \mathbf{\Psi}_{I_n^t}^{\dagger} y_n$
Output: Γ `//sparse coefficient matrix`

Note that we suggest to replace OMP by thresholding only within the DL stage but to keep OMP for the SC stage. This choice is motivated by two reasons. First, thresholding is a computationally much cheaper procedure than OMP and hence, suitable for accelerating the regularization stage of the iterative reconstruction. Second, although OMP is known to yield better results than thresholding for sparse approximation, it is unstable under perturbations. More precisely, using an appropriate dictionary $\mathbf{\Psi}$ for the sparse approximation of a class of signals, OMP is known to yield much better results than simple thresholding. However, in the presence of perturbations of the dictionary, which is the case during the DL learning procedure, OMP performs worse. In particular, even in the presence of only small perturbations, the performance of OMP drastically decreases and hence, can be replaced by simple thresholding yielding similar results.

II.C. In-Vivo Experiments

Here, we conducted several experiments to study the behaviour of the proposed adaptive DL and SC algorithms used for the solution of (P1) and (P2). For that purpose, we ran experiments where we reconstructed 2D cine MR images from undersampled k-space data. In order to get an assessment of the quality of the obtained reconstructions for various combinations of DL and SC algorithms and to highlight some aspects of the adaptive DL and SC algorithms, we performed the following experiments.

1. *Adaptive Vs. Non-Adaptive DL and SC:* Here, we quantitatively compared the performance of the reconstruction algorithms used to solve problems (P1) and (P2) using three different combinations of DL and SC algorithms: K-SVD + OMP, ITKrM + OMP and aITKrM + aOMP. For these experiments, images obtained by kt-SENSE[23] were used as ground-truth images. From these images, the k-space data was retrospectively generated and corrupted by Gaussian noise in order to simulate an acceleration factor of 9. We repeated the experiments for different choices of the sparsity level S. More precisely, to demonstrate the impact of the choice of potentially too low/too high S, we used $S = 4$, $S = 8$ and $S = 16$ for the non-adaptive DL and SC algorithms.

2. *Convergence behaviour:* We investigated the convergence behaviour of the different combinations of DL and SC methods by tracking the average of the chosen image measures during the reconstruction.

3. *Computational Time:* We compared the different combinations K-SVD + OMP / ITKrM + OMP / aITKrM + aOMP in terms of computational time.

4. *Qualitative Comparison:* Here, we reconstructed images from the k-space data obtained from the scanner with the three different combinations of DL and SC.

For all experiments, we used the publicly available `Python`-implementations of K-SVD and OMP in the `scikit-learn` library[24] which are based on an efficient implementation of K-SVD using batch OMP[25]. Our `Python`-implementations of ITKrM, aITKrM and aOMP will be made available after peer-review. For our customized implementation of the forward and the adjoint operators $\mathbf{F}_I$ and $\mathbf{F}_I^{\mathsf{H}}$, we used the libraries `ODL`[26] and `PyNUFFT`[27]. The PCG method used to solve system (4) was provided by `ODL`.

II.C.1. Dataset

Our dataset consisted of $n = 15$ 2D cine MR images from patients as well as healthy volunteers and represents a subset of particularly interesting cases selected from[28]. Further, 10 different images were used to pre-train dictionaries used for solving (P1). The images were obtained using a bSSFP sequence on a 1.5 T MR scanner (Achieva, Philips Healthcare, Best, The Netherlands) within a single breathhold of 10 s (TR/TE = 3.0/1.5 ms, FA 60°). The images have a shape of $N_x \times N_y \times N_t = 320 \times 320 \times 30$, where $N_x \times N_y$ is the number of in-plane pixels and N_t is the number of cardiac phases which were acquired during the scan. The in-plane resolution of the images is 2 mm and the slice thickness is 8 mm. The acquired k-space data corresponds to the Fourier-data sampled along $N_\theta = 3400$ radial trajectories which were chosen according to[29]. From these images, we retrospectively generated the undersampled k-space data $\mathbf{y}_I$ by solely using $N_\theta = 1130$ radial spokes. Using only $N_\theta = 1130$ spokes corresponds to an undersampling factor of approximately ~ 9 and reduces the scan time to approximately 3 seconds. Further, the k-space data was corrupted by a normally distributed random noise vector $\mathbf{e}$ with a standard deviation of 0.05.

II.C.2. Experiment Set-Up

The patch-size used for all experiments was $4 \times 4 \times 4$, i.e. $d = 64$. As in[5], we approximated the real and imaginary part of the complex-valued images separately but using the same real-valued dictionary $\mathbf{\Psi}$. For the non-adaptive combinations of DL and SC algorithms, we fixed the number of atoms of the dictionary $\mathbf{\Psi}$ to be $K = 128$. Note that the empirical choice of K is typically in the range $d \leq K \leq 4d$ while using a sparsity level of $S = \sqrt{d}$, which, for a fixed size of patches $4 \times 4 \times 4$, results in $64 \leq K \leq 256$ and $S = 8$. In fact, in well-known literature, this choice is well-established. For example, in[6], the parameters are empirically set to $K = 256$ and $S = 15$. In[5], the number of atoms is set even higher, varying from $K = 300$ to $K = 600$, dependent on the experiments. However, due to the fact that in our work our forward model is given by a radial encoding operator using multiple coils, the artefacts contained in the NUFFT-reconstruction $\mathbf{x}_I$ are inherently different from the ones obtained by a zero-filled reconstruction as in[5] or[6]. Since the artefacts can be expected to have a more high-frequency type of texture, we decided to only use $K = 128$.

As already mentioned, the experiments were repeated for a relatively low choice of $S = 4$, a typical choice $S = \sqrt{d} = 8$ and a relatively high choice of $S = 16$. Since the k-space data $\mathbf{y}_I$ was contaminated by random noise, the regularization parameter λ was set to $\lambda = 1$ in order to achieve a relatively strong contribution of the regularization imposed by DL and SC and therefore being able to highlight the impact of the different DL and SC algorithms. The number of PCG iterations used to update the reconstruction by solving (4) and the number of overall iterations for ADMM were set to $n_{\mathrm{PCG}} = 4$ and $T = 12$, respectively.

For solving (P1), the dictionaries were pre-trained on patches extracted from the images of 10 different subjects. The dictionaries were initialized by $K = 128$ randomly selected patches and trained by randomly extracting 150 000 patches of the real and imaginary part of the images at each DL iteration. The maximal number of iterations for the respective DL algorithm was set to $n_{\mathrm{DL}} = 200$. The resulting size of the dictionary learned with aITKrM was $K = 151$. For solving (P2), the dictionaries were trained by randomly extracting $N = 10\,000$ patches of the real and imaginary part of the current image estimate $\mathbf{x}_k$ for each DL iteration. The maximal number of iterations of the respective dictionary algorithm within one ADMM iteration was set to $n_{\mathrm{DL}} = 20$. The dictionaries were initialized as for solving (P1) and continuously updated during the reconstruction. For each subsequent ADMM iteration, the dictionary Ψ was initialized with the one learned during the previous ADMM iteration. For the sparse approximation we used strides of 2 in N_x-, N_y- and N_t-direction, which reduces the number of patches to be sparsely approximated by a factor of 8. For the combination aITKrM + aOMP, we used $\mu_{\max} = 0.7$ and for the number of minimal observations we used $M = d$, which is suitable for this number of training signals. Note that we did not learn the constant atom since the patches were centred before training the dictionaries.

II.C.3. Quantitative Measures

For evaluating the performance of the different reconstruction algorithms, we report the peak signal-to-noise ratio (PSNR) and the normalized root mean squared error as error-based image metrics and the structural similarity index measure[30] (SSIM) and the Haar wavelet-based perceptual similarity index measure (HPSI)[31] as similarity-based image metrics. Note that the latter has been reported to exhibit a higher correlation with the human opinion tested on different benchmark datasets, see[31]. The hyper-parameters needed by SSIM and HPSI are the ones published in the respective works. In order to focus on the regions of the images with diagnostic content, the metrics were calculated on the images which were previously cropped to $N'_x \times N'_y = 220 \times 220$.

III. Results

III.A. Reconstruction Results

Here, we reconstructed all 15 cine MR images using the different combinations of DL and SC algorithms. Figure 1 shows an example of images reconstructed with the three different combinations of DL and SC algorithms for the different sparsity levels $S = 4$, $S = 8$ and $S = 16$. As can be seen from the point-wise error images, all non-adaptive and the adaptive DL and SC combinations led to visually comparable results. Table 1 lists the average PSNR, NRMSE, SSIM and HPSI for the different reconstructions. We see that for both non-adaptive combinations K-SVD + OMP and ITKrM + OMP, setting $S = 16$ yielded the worst results compared to $S = 8$ and $S = 4$. In particular, the gap between the both was larger for larger S, which can be attributed to issues during the dictionary learning and is a well known issue of ITKrM for overestimated sparsity levels[19]. The adaptive combination aITKrM + aOMP achieved similar reconstruction quality as K-SVD + OMP with the best reported choices of the sparsity level by further slightly improving SSIM.

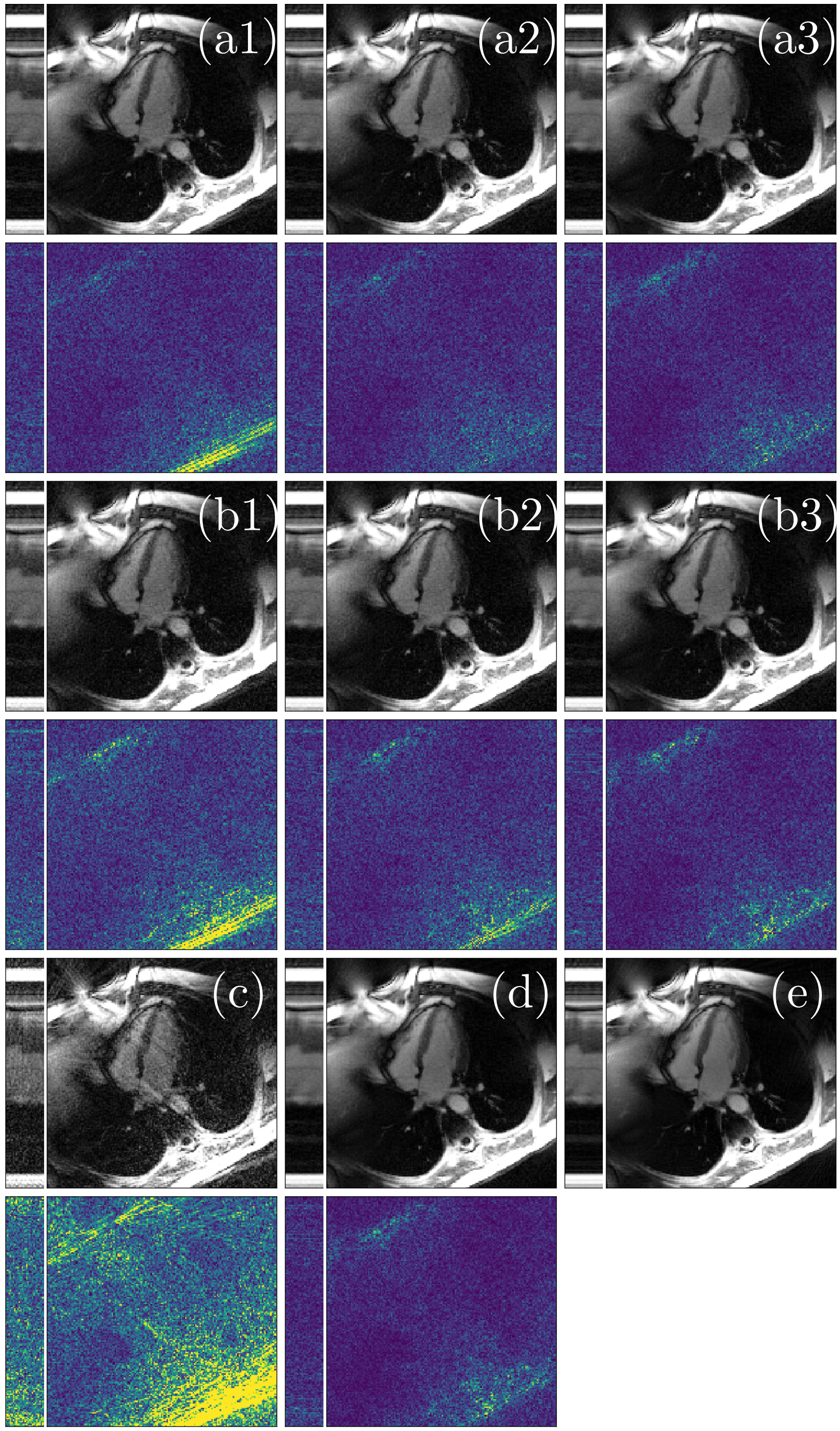

Figure 1: Results obtained by using different combinations of DL and SC algorithms. (a1)-(a3): K-SVD + OMP for $S = 16$ (a1), $S = 8$ (a2) and $S = 4$ (a3), (b1)-(b3): ITKrM + OMP for $S = 16$ (b1), $S = 8$ (b2) and $S = 4$ (b3), the initial NUFFT-reconstruction from $N_\theta = 1130$ radial spokes (c), aITKrM + aOMP (d) and the kt-SENSE reconstruction using $N_\theta = 3400$ radial spokes (e) which served as ground truth for the retrospective k-space data-generation.

Table 1: Comparison of the performance of different algorithms for DL and SC used in the reconstruction.

	Ψ Learned during Reconstruction						
	Non-Adaptive						**Adaptive**
DL	**K-SVD**			**ITKrM**			**aITKrM**
SC	**OMP**			**OMP**			**aOMP**
S	16	8	4	16	8	4	ad.
PSNR	43.870	44.538	44.354	40.825	43.017	43.628	44.491
NRMSE	0.068	0.062	0.064	0.096	0.074	0.069	0.063
SSIM	0.671	0.692	0.710	0.604	0.657	0.698	0.734
HPSI	0.989	0.992	0.992	0.982	0.989	0.991	0.992
	Pre-Trained Ψ						
	Non-Adaptive						**Adaptive**
DL	**K-SVD**			**ITKrM**			**aITKrM**
SC	**OMP**			**OMP**			**aOMP**
S	16	8	4	16	8	4	ad.
PSNR	44.856	45.205	44.594	43.117	44.483	44.009	45.314
NRMSE	0.06	0.058	0.062	0.073	0.063	0.066	0.057
SSIM	0.684	0.699	0.714	0.645	0.687	0.709	0.738
HPSI	0.992	0.992	0.992	0.988	0.991	0.99	0.993

III.B. Convergence Behaviour

For assessing the convergence speed of the reconstruction algorithms, we tracked the different measures used for the evaluation of the performance of the reconstruction algorithms during the iterative reconstruction. Figure 2 shows the mean PSNR, NRMSE, SSIM and HPSI averaged over the different images. Quite consistently, it can be observed that the reconstruction using the adaptive combinations aITKrM + aOMP surpassed the non-adaptive DL and SC combinations at early iterates with respect to all measures and tended to let the curves flatten out earlier than the non-adaptive counterparts. This could be particularly well observed for the case of NRMSE and PSNR and held true for all scenarios with different S. ITKrM + OMP with $S = 16$ revealed a semi-convergence type of behaviour which can be attributed to the fact that $S = 16$ is too high for ITKrM in the presence of noise in the

k-space. This also shows that the choice of S can have a high impact on the reconstruction.

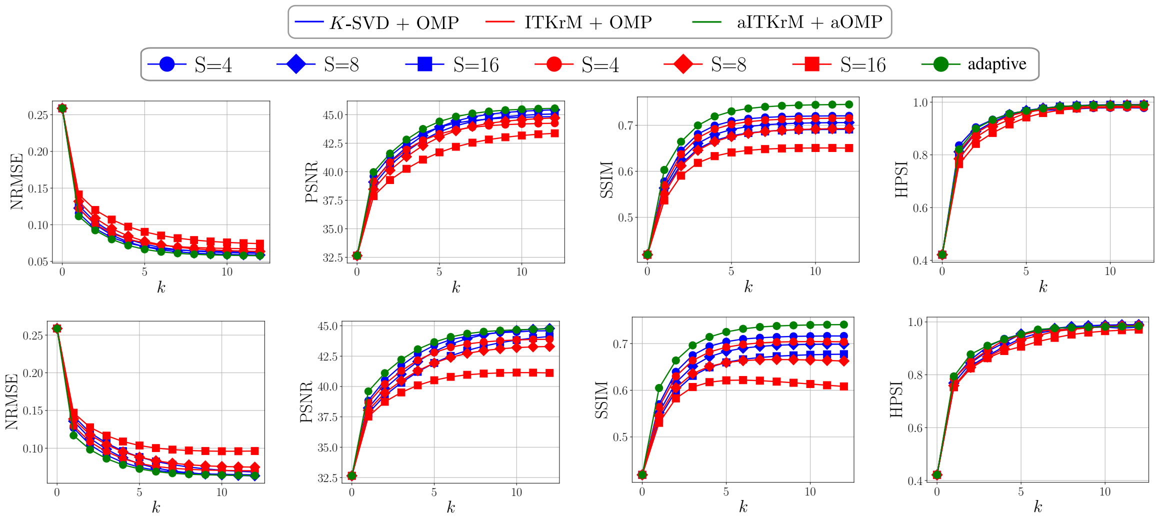

Figure 2: Convergence behaviour of the reconstruction scheme for solving (P1) (first row) and for solving (P2) (second row) using different combinations of DL and SC algorithms. The combination of aITKrM + aOMP yields better or equally good results compared to the non-adaptive combinations with respect to all measures, for solving (P1) and (P2). The images show the respective average measure over the iterations.

III.C. Reconstruction Times

Here, we report the times for the different components of the DL-based reconstruction algorithms. The components which significantly contributed to the relatively high reconstruction times were the DL and SC algorithms and the PCG method which is needed to obtain an approximate solution of (4). Obviously, the latter was constant for the three different combinations of DL and SC. Table 2 lists the average time needed for DL and SC for each ADMM iteration. Therefore, the overall time needed for a specific component can be obtained by multiplying the respective time by the number of ADMM iterations T.

Table 2: Comparison of DL and SC in terms of computational times in seconds for one ADMM iteration for solving problem (P2).

DL and SC	Sparsity Level	DL / SC Time
K-SVD + OMP	$S = 16$	71 / 849
	$S = 8$	69 / 415
	$S = 4$	69 / 206
ITKrM + OMP	$S = 16$	9 / 824
	$S = 8$	8 / 412
	$S = 4$	8 / 205
aITKrM + aOMP	adaptive	7 / 149

We see that K-SVD was the slowest DL algorithm and took approximately 69-71 seconds for one single ADMM iteration. ITKrM was considerably faster and took only between 8-9 seconds whereas its adaptive version was the fastest and took only around 7 seconds even though some additional time was needed to estimate the sparsity level and replacing coherent and unused atoms. For the SC, we see that for OMP, the chosen sparsity level obviously had an impact on the required computational time and took 824-849 seconds for $S = 16$, 412-415 seconds for $S = 8$ and 205-206 seconds for $S = 4$. Our adaptive version aOMP was even faster as OMP for the lowest choice of $S = 4$ and required about 149 seconds.
Figure 3 shows a diagram representing the overall time for the respective component of the reconstruction algorithm. From the bars we can see the time which each component took relative to the total reconstruction time. First, we see that for the non-adaptive experiments, the time needed for the SC of all patches constitutes the computational bottleneck of the method when S is chosen too high, i.e. $S = 16$. Second, we see that, as expected, ITKrM was able to substantially reduce the computational time compared to K-SVD. However, the gain in terms of acceleration was negligible when putting it in relation to the overall time because OMP still remains the computational overhead for $S = 16$. The last bar of the graph shows that first, by employing aITKrM, the time needed to learn the dictionary still amounted to approximately the same as for ITKrM, and second, in this configuration, the time needed for SC was clearly reduced and approximately corresponds to the one for OMP with $S = 4$.

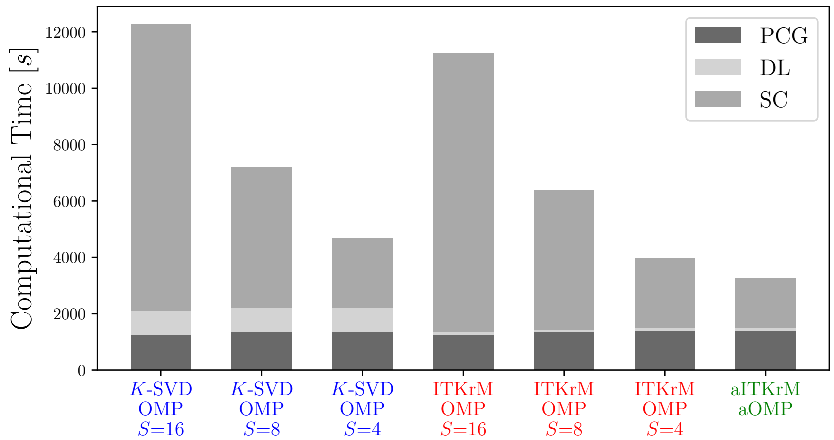

Figure 3: Reconstruction times grouped by components for different combinations of DL and SC algorithms for solving problem (P2). When solving (P1), the times needed for PCG and SC remain similar, while the time for learning the dictionary Ψ can be neglected since it is assumed to be given a-priori.

III.D. Experiments Using Real k-Space Data

In the following, we tested the reconstruction algorithm with the different combinations of DL and SC by using the real k-space data acquired along $N_\theta = 1130$ radial trajectories obtained from the scanner and compared it to kt-SENSE using $N_\theta = 3400$ radial trajectories. Note that sampling k-space along $N_\theta = 3400$ spokes already corresponds to an undersampling factor of ~ 3 which is needed to perform the scan in a single breathhold. Further, the kt-SENSE reconstruction algorithm itself imposes prior information to regularize the inverse problem and therefore, the kt-SENSE reconstructions obtained from the $N_\theta = 3400$ radial spokes cannot be considered as ground truth images for this experiment. Therefore, we abstain from reporting quantitative measures as well as point-wise error images. A rigorous quality assessment would need to be performed with respect to predefined clinical features and a clinical application. However, since this is beyond the scope of this work, we only show an example of the reconstruction for the sake of completeness and to demonstrate the applicability of aITKrM and aOMP for real k-space data. Figure 4 shows an example of images reconstructed with the three different combinations of DL and SC algorithms. Figure 4 (a1)-(a3) show the results obtained with K-SVD + OMP and (b1)-(b3) with ITKrM + OMP for different sparsity levels S, respectively. The initial NUFFT-reconstruction is visible in Figure 4 (c). Figure 4 (d) shows the result obtained with aITKrM + aOMP and (e) shows the kt-SENSE reconstruction using $N_\theta = 3400$ radial spokes. Visually, all methods performed similarly well, and K-SVD + OMP and ITKrM + OMP show a slightly higher noise level compared to aITKrM + aOMP, which is consistent with the results presented in Subsection III.A.. However, note again that the times needed to obtain the reconstructed images are substantially lower for aITKrM + aOMP and no a-priori choice of the hyper-parameters S and K was required.

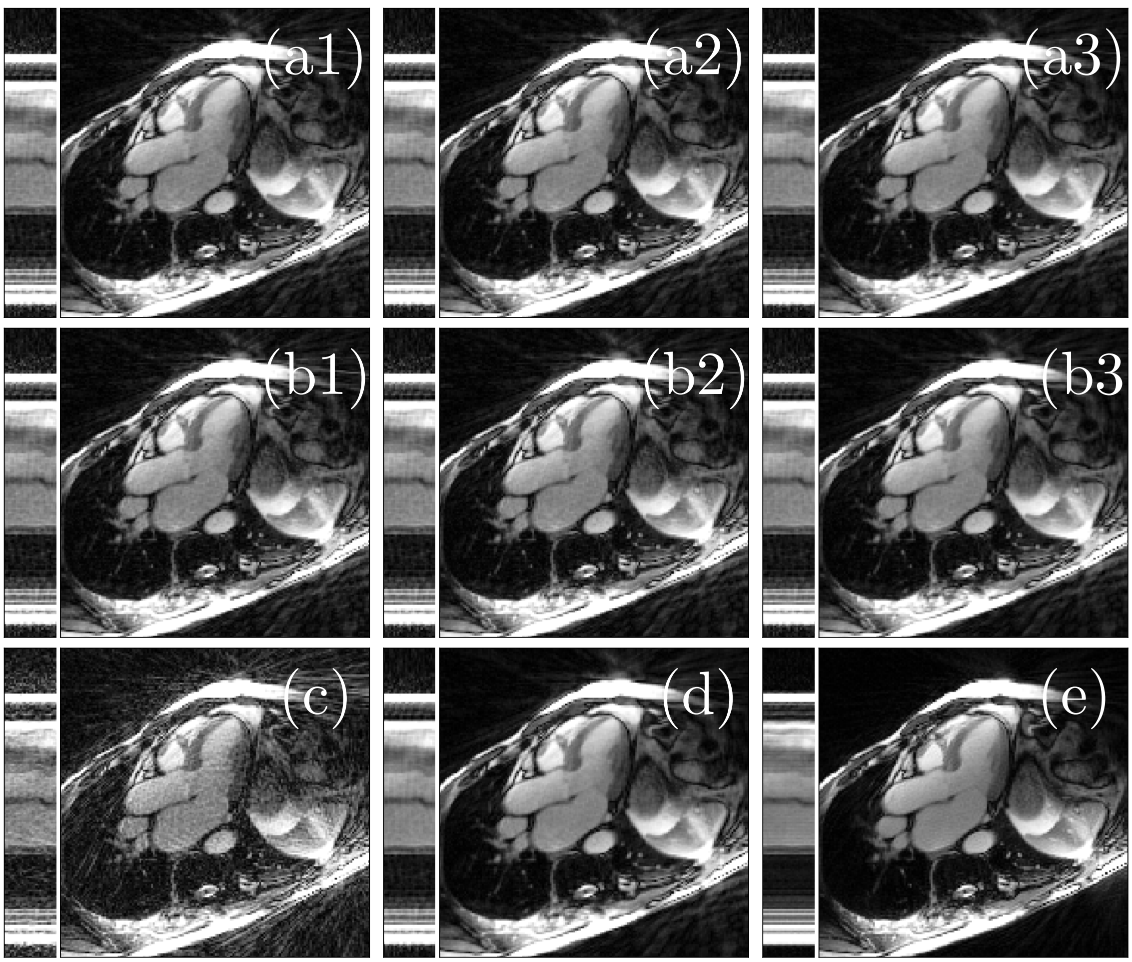

Figure 4: Results obtained from real k-space data obtained from the scanner measurements. K-SVD + OMP with $S = 4$ (a1), $S = 8$ (a2), $S = 16$ (a3), ITKrM + OMP with $S = 4$ (b1), $S = 8$ (b2), $S = 16$ (b3), NUFFT-reconstruction using $N_\theta = 1130$ radial lines (c), aITKrM + aOMP (d) and kt-SENSE using $N_\theta = 3400$ radial spokes (e).

IV. Discussion

The proposed adaptive versions of the DL and SC algorithm given by the adaptive Iterative Thresholding and K-residual Means (aITKrM) algorithm and adaptive Orthogonal Matching Pursuit (aOMP) provide valid alternatives to the well-established K-SVD algorithm and the non-adaptive SC algorithm OMP. As this work is of methodological nature, in the following we discuss advantages and limitations of the described algorithms in more detail.

IV.A. Adaptive Estimation of S and K

Clearly, the major advantage of the presented adaptive DL and SC algorithms aITKrM and aOMP is to no longer need to choose the sparsity level S and the number of atoms K a-priori. This is not only important for making such algorithms more eligible for practical applications but also as a wrong choice of S and K can have a large impact on the computational time and the reconstruction quality. Intuitively speaking, a too small choice of S leads to too smooth results with probably missing details while a too high choice of S results in a preservation of undersampling artefacts which we are trying to remove. Also, the structure of an image varies from location to location and hence, also S should vary dependent on the considered image patch.

Moreover, the optimal number of atoms K is also data-dependent. In particular, for dictionaries learned on images containing more structure, a larger K is needed than for fairly smooth ones. Further, the optimal size of the dictionary was also shown to be dependent on the noise level of a corrupted image, i.e. the more noise, the smaller K should be chosen[22]. These observations suggest that a global choice of S and K cannot be optimal, disregarding from the fact that they are not known and can only be guessed.

Using aITKrM and aOMP, S and K are adaptively chosen based on the texture of the current image estimate. Intuitively, at early iterations of the iterative reconstruction, a stronger regularization of the image estimate is required in order to reduce the artefacts. At later iterations, where the current image estimate contains less noise and artefacts, a higher S is required to be able to represent fine anatomic details. In fact, this behaviour could be observed during the reconstruction and is illustrated in Figure 5. In Figure 5 (a1) and (b1), the real and imaginary part of the NUFFT-reconstruction $\mathbf{x}_I$ are displayed. In (c1) and (d1), we can see the corresponding patch-wise approximated images using aOMP and a dictionary learned by aITKrM. Figures 5 (e1) and (f1) show the estimated sparsity levels at various locations in the image. The second row of Figure 5 shows the same images at the penultimate iteration $T = 11$ of the reconstruction. As we can see in (e2) and (f2), the average estimated sparsity level S is significantly higher than for the NUFFT-reconstruction, especially in the regions of the image which contain the patient's anatomy. This indicates that the proposed adaptivity of aITKrM and aOMP is suitable for adapting to the texture of cine MR images. In particular, it achieves the desired property of strongly regularizing images with strong artefacts and noise at early iterates while still being able to well-represent image details at later iterations in the iterative reconstruction.

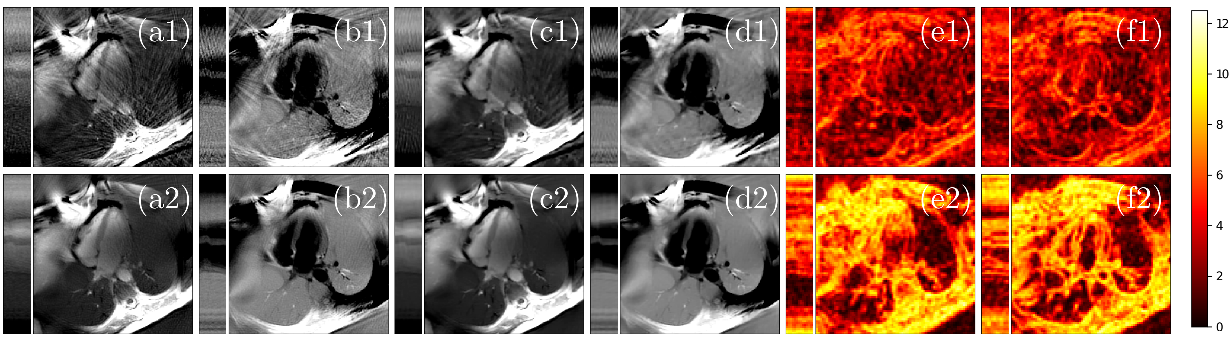

Figure 5: Estimated sparsity level at different stages during the iterative reconstruction for solving (P2). First row: (a1) and (b1) - real and imaginary part of the initial NUFFT-reconstruction $\mathbf{x}_I$, (c1) and (d1) - the correspondent patch-wise sparse approximations using aITKrM + aOMP, (e1) and (f1) - the estimated sparsity levels of image-patches at various locations. Second row: (a2) and (b2) - real and imaginary part of the twelfth iterate obtained by using aITKrM + aOMP, (c2) and (d2) - the correspondent patch-wise sparse approximations using aITKrM + aOMP, (e2) and (f2) - the estimated sparsity levels of image-patches at various locations. The average sparsity level S is therefore lower at early iterates in the reconstruction and higher at later iterates.

This demonstrates that for the specific task of iterative image reconstruction, the optimal sparsity level S of a patch first of all depends on the needed complexity to represent relevant features and second, might change during the reconstruction. Further, in Subsection III.C., we have observed that choosing a too high S clearly has significant impact on the computational time and at the same time does not necessarily increase the reconstruction quality.

In Figure 6 we see an example of eight atoms out of the dictionaries learned by the respective

DL algorithms. The atoms of the dictionaries shown in the figure were learned on a set of patches extracted from the initial NUFFT-reconstruction $\mathbf{x}_I$ (first row) and from the penultimate image estimate of the reconstruction (second row). We can see that the dictionaries learned by the non-adaptive DL algorithms with $S = 16$ tend to inherently contain quite a large portion of noise in the atoms which, on the other hand, is almost not present in the atoms learned by aITKrM. This observation is consistent with the theory discussed in[22] for the case where the sparsity level S or the dictionary size K are overestimated and suggests that $S = 16$ is a far to high choice of the sparsity level. The fact that the hyper-parameters S and K no longer need to be chosen a-priori could highly facilitate a possible application of the reconstruction algorithm in the clinical routine, where standardized acquisition and reconstruction protocols have to be used. Further, as we have seen in the examples shown in Subsection II.C., the S- and K-adaptivity achieves competitive results compared to K-SVD + OMP and additionally reduces the required reconstruction times.

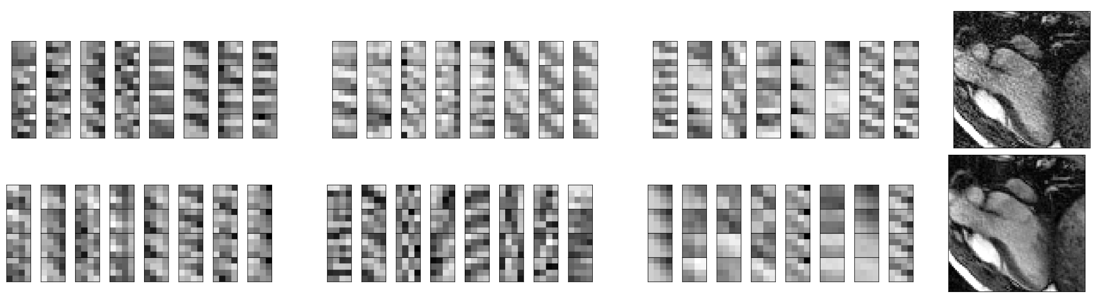

Figure 6: Examples of eight three-dimensional atoms (un-stacked along the time dimension) of the dictionaries learned by K-SVD (left), ITKRM (mid) and aITKrM (right). The dictionaries were learned on 3D patches extracted from the initial NUFFT-reconstruction $\mathbf{x}_I$ (first row) and the penultimate image estimate (second row). For K-SVD and ITKrM, the sparsity level was $S = 16$. Since $S = 16$ is relatively high, the atoms obtained by K-SVD and ITKrM contain quite some noise. Note that the constant atom is not shown in the images.

IV.B. Limitations

A possible limitation of the presented work is that the thresholds chosen for the algorithms underlie the theoretical consideration of Gaussian and sub-Gaussian noise which might not be true in general. However, sampling along radial trajectories is known to represent an incoherent sampling pattern with noise-like properties and similar or even better results could be probably obtained by using Compressed-Sensing Cartesian schemes[32].

As all iterative reconstruction methods which employ a-priori knowledge expressed as a penalty term, the DL-based regularization method requires to choose the regularization parameter λ. However, note that quite some work has been dedicated on how to adaptively choose the parameter λ as well, see e.g.[10,11], which might be incorporated in the reconstruction algorithm using aITKrM + aOMP.

IV.C. Reconstruction Quality

The achieved image quality using aITKrM + aOMP is comparable with the one achieved using the standard combination K-SVD + OMP with the best reported choices of the sparsity level as can be seen in Figure 1 and Table 1. The performed experiments reveal that for K-SVD, choosing a too high S impairs image quality compared to a lower choice of S. This effect is even clearer for ITKrM, where a too high S is known to disturb atoms in the dictionary, especially in the presence of noise[22]. Moreover, in Figure 3 we have seen that overestimating S leads to a substantial increase of computational time. From these experiments we can further conclude that the choice of S is non-trivial. Also, relying on the choice of hyper-parameters suggested in the literature might not be optimal, as the reported parameters are always data- and problem dependent and usually adapted to a specific task. This observation makes the S- and K-adaptivity a particularly interesting feature of the combination aITKrM + aOMP from a practical point of view. First of all, it is potentially possible to reduce the computational time and further improve the reconstruction quality by properly estimating S and K. Second, the hyper-parameters are adaptively tuned to the considered data and no a-priori choice is needed.

IV.D. Reconstruction Times

Learning a dictionary with aITKrM instead of K-SVD leads to an acceleration factor of approximately 10 which is useful when the dictionary is learned during the reconstruction. The reason is that the computationally most expensive component of K-SVD is OMP, where aITKrM in contrast only requires the faster thresholding. More importantly, using aOMP has the potential of highly reducing the time needed for the sparse approximation of all patches since, instead of using a (as we have seen, potentially too high) global sparsity level

S, it is adaptively chosen according to the considered patch-example.

We point out that the used implementations of aITKrM as well as aOMP were not run on a GPU and could therefore be further accelerated and optimized. In particular, the underyling nature of ITKrM offers the possibility to transfer the calculations on a GPU and exploit parallelisation as it can process the patches sequentially. Also, note that for K-SVD we used an already optimized and efficient version based on batch OMP[33] and therefore, further improvements in terms of computational time could be expected from a more sophisticated implementation of ITKrM and aOMP.

V. Conclusion

In this work we have investigated the application of adaptive sparsity level and dictionary size estimation for the regularization of cine MR image reconstruction using dictionary learning (DL) and sparse coding (SC). We have used an adaptive version of ITKrM (Iterative Thresholding and K-residual Means) for DL and have presented a novel adaptive version of the Ortogonal Matching Pursuit (OMP) algorithm for SC. We have shown its competitiveness and advantages compared the wellestablished K-SVD and OMP algorithms. Most methods employing DL and SC for the regularization of image reconstruction in MR use a global sparsity level S for DL as well as for SC. Further, S and the number of atoms K to be used are usually determined by computationally expensive hyper-parameter searches. Using the adaptive methods aITKrM and aOMP, the a-priori choice of S and K is no longer needed. Instead, S and K are optimally determined for each patch within the iterative reconstruction and adapted to the texture of the currently considered image estimates. As we have seen, aOMP provides appropriate estimates of S for the sparse approximation of the patches and by this, a more efficient regularization is achieved. This also results in a significant acceleration of the regularization step, especially when compared to the case for standard a-priori choices of S and K.

Acknowledgements

A. Kofler acknowledges the support of the German Research Foundation (DFG), project number GRK 2260, BIOQIC. M. Pali is grateful for the support by the Austrian Science

Fund (FWF), Grant no. Y760. We thank K. Schnass for the support and critical discussions of the manuscript.

References

[1] B. J. Wintersperger, S. B. Reeder, K. Nikolaou, O. Dietrich, A. Huber, A. Greiser, T. Lanz, M. F. Reiser, and S. O. Schoenberg, Cardiac CINE MR imaging with a 32-channel cardiac coil and parallel imaging: impact of acceleration factors on image quality and volumetric accuracy, Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine **23**, 222–227 (2006).

[2] Y. Chen, X. Ye, and F. Huang, A novel method and fast algorithm for MR image reconstruction with significantly under-sampled data, Inverse Problems and Imaging **4**, 223–240 (2010).

[3] S. Ravishankar and Y. Bresler, MR image reconstruction from highly undersampled k-space data by dictionary learning, IEEE Trans. Med. Imag. **30**, 1028 (2011).

[4] 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 Transactions on Medical Imaging **32**, 1290–1301 (2013).

[5] J. Caballero, A. N. Price, D. Rueckert, and J. V. Hajnal, Dictionary learning and time sparsity for dynamic MR data reconstruction, IEEE Transactions on Medical Imaging (2014).

[6] Y. Wang and L. Ying, Compressed sensing dynamic cardiac cine MRI using learned spatiotemporal dictionary, IEEE Transactions on Biomedical Engineering (2014).

[7] N. G. Behl, C. Gnahm, P. Bachert, M. E. Ladd, and A. M. Nagel, Three-dimensional dictionary-learning reconstruction of 23Na MRI data, Magnetic Resonance in Medicine **75**, 1605–1616 (2016).

[8] Y. Wang, N. Cao, Z. Liu, and Y. Zhang, Real-time dynamic MRI using parallel dictionary learning and dynamic total variation, Neurocomputing **238**, 410–419 (2017).

9 P. Song, L. Weizman, J. F. Mota, Y. C. Eldar, and M. R. Rodrigues, Coupled dictionary learning for multi-contrast MRI reconstruction, IEEE Transactions on Medical Imaging (2019).

10 Y. Li, J. Zhang, G. Sun, and D. Lu, The Sparsity Adaptive Reconstruction Algorithm Based on Simulated Annealing for Compressed Sensing, Journal of Electrical and Computer Engineering **2019** (2019).

11 C. Chen, Y. Liu, P. Schniter, N. Jin, J. Craft, O. Simonetti, and R. Ahmad, Sparsity adaptive reconstruction for highly accelerated cardiac MRI, Magnetic resonance in medicine **81**, 3875–3887 (2019).

12 R. S. Mathew and J. S. Paul, Compressed Sensing Parallel MRI with Adaptive Shrinkage TV Regularization, arXiv preprint arXiv:1809.06665 (2018).

13 T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, Sparsity adaptive matching pursuit algorithm for practical compressed sensing, in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, pages 581–587, IEEE, 2008.

14 T. Blumensath and M. E. Davies, Stagewise weak gradient pursuits, IEEE Transactions on Signal Processing **57**, 4333–4346 (2009).

15 D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit, IEEE transactions on Information Theory **58**, 1094–1121 (2012).

16 H. Wu and S. Wang, Adaptive sparsity matching pursuit algorithm for sparse reconstruction, IEEE Signal Processing Letters **19**, 471–474 (2012).

17 M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, volume 49, NBS Washington, DC, 1952.

18 M. Aharon et al., K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. on signal processing **54**, 4311 (2006).

19 K. Schnass, Convergence radius and sample complexity of ITKM algorithms for dictionary learning, Applied and Computational Harmonic Analysis (2016).

[20] K. Schnass, Local Identification of Overcomplete Dictionaries, Journal of Machine Learning Research **16**, 12111242 (2015).

[21] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, in *Proceedings of 27th Asilomar conference on signals, systems and computers*, pages 40–44, IEEE, 1993.

[22] K. Schnass, Dictionary learning-from local towards global and adaptive, arXiv preprint arXiv:1804.07101 (2018).

[23] J. Tsao, P. Boesiger, and K. P. Pruessmann, k-t BLAST and k-t SENSE: Dynamic MRI With High Frame Rate Exploiting Spatiotemporal Correlations, Magnetic Resonance in Medicine (2003).

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research **12**, 2825–2830 (2011).

[25] R. Rubinstein, M. Zibulevsky, and M. Elad, Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit, Technical report, Computer Science Department, Technion, 2008.

[26] J. Adler, ODL - Operator Discretization Library, `https://github.com/odlgroup/odl`, 2013.

[27] J.-M. Lin, Python Non-Uniform Fast Fourier Transform (PyNUFFT): An Accelerated Non-Cartesian MRI Package on a Heterogeneous Platform (CPU/GPU), Journal of Imaging **4**, 51 (2018).

[28] A. Kofler, M. Dewey, T. Schaeffter, C. Wald, and C. Kolbitsch, Spatio-Temporal Deep Learning-Based Undersampling Artefact Reduction for 2D Radial Cine MRI with Limited Training Data, IEEE Transactions on Medical Imaging , 1–1 (2019).

[29] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers, and O. Doessel, An optimal radial profile order based on the Golden Ratio for time-resolved MRI, IEEE Transactions on Medical Imaging **26**, 68–76 (2006).

[30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Transactions on Image Processing (2004).

[31] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, A Haar wavelet-based perceptual similarity index for image quality assessment, Signal Processing: Image Communication (2018).

[32] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, Compressed sensing MRI, IEEE signal processing magazine **25**, 72–82 (2008).

[33] R. Rubinstein, A. M. Bruckstein, and M. Elad, Dictionaries for sparse representation modeling, Proceedings of the IEEE **98**, 1045–1057 (2010).