Learning a Spatiotemporal Dictionary for Magnetic Resonance Fingerprinting with Compressed Sensing

Pedro A. Gómez[1,2,3(✉)], Cagdas Ulas[1], Jonathan I. Sperl[3], Tim Sprenger[1,2,3], Miguel Molina-Romero[1,2,3], Marion I. Menzel[3], and Bjoern H. Menze[1]

[1] Computer Science, Technische Universität München, Munich, Germany
`pedro.gomez@tum.de`
[2] Biomedical Engineering, Technische Universität München, Munich, Germany
[3] GE Global Research, Munich, Germany

Abstract. Magnetic resonance fingerprinting (MRF) is a novel technique that allows for the fast and simultaneous quantification of multiple tissue properties, progressing from qualitative images, such as T1- or T2-weighted images commonly used in clinical routines, to quantitative parametric maps. MRF consists of two main elements: accelerated pseudorandom acquisitions that create unique signal evolutions over time and the voxel-wise matching of these signals to a dictionary simulated using the Bloch equations. In this study, we propose to increase the performance of MRF by not only considering the simulated temporal signal, but a full spatiotemporal neighborhood for parameter reconstruction. We achieve this goal by first training a dictionary from a set of spatiotemporal image patches and subsequently coupling the trained dictionary with an iterative projection algorithm consistent with the theory of compressed sensing (CS). Using data from BrainWeb, we show that the proposed patch-based reconstruction can accurately recover T1 and T2 maps from highly undersampled k-space measurements, demonstrating the added benefit of using spatiotemporal dictionaries in MRF.

1 Introduction

Quantitative magnetic resonance imaging (qMRI) techniques measure relevant biological parameters, providing a profound characterization of the underlying tissue. In contrast to conventional weighted MRI, where the image signal is represented by intensity values and different tissues are described relative to each other, qMRI generates parametric maps of absolute measures that have a physical interpretation, leading to reduced bias and reproducible diagnostic information. On the other hand, obtaining quantitative maps is a time consuming task. It requires the repeated variation of typical MR acquisition parameters, such as flip angle (FA) or repetition time (TR), and the fitting of the measured signal to a model in order to estimate the parameters of interest, including the MR specific longitudinal (T1) and transversal (T2) relaxation times. Long acquisition times, together with high sensitivity to the imaging device

and system setup, are the main restrictions to clinical applications of qMRI techniques.

A recently proposed qMRI method, magnetic resonance fingerprinting (MRF), aims to overcome these limitations through accelerated pseudorandom acquisitions [6]. It is based on the idea that pseudorandom variations on acquisition parameters cause the signal response for different tissue types to be unique. This unique signal evolution can be matched to a precomputed dictionary created from known combinations of the parameters of interest (e.g. T1 and T2). Therefore, by matching the measured signal to one atom in the dictionary, all of the parameters used to simulate the corresponding atom can be simultaneously extracted. Furthermore, since the form of the signal evolution used for pattern matching is known *a priori*, MRF is less sensitive to measurement errors, facilitating accelerated acquisitions through the undersampling of the measurement space (k-space). It should be noted that, so far, all matching is done for one-dimensional temporal signals only.

The notion of reconstructing signals from undersampled measurements comes from the theory of compressed sensing (CS) [5]. CS has been successfully applied to accelerate parameter mapping [4] and recently Davies et al. [3] demonstrated a CS strategy for MRF that does not rely on pattern matching for error suppression and has exact recovery guarantees, resulting in increased performance for shorter pulse sequences. The authors further extend their CS model to exploit global spatial structure by enforcing sparsity in the wavelet domain of the estimated density maps, slightly improving the performance of their approach.

Spatial information can also be incorporated locally by using image patches. Patch-based dictionaries have the advantage of being able to efficiently represent complex local structure in a variety of image processing tasks. Furthermore, the use of overlapping patches allows for averaging, resulting in the removal of both noise and incoherent artefacts caused by undersampling. Patch-based dictionaries have been previously used for the task of MR image reconstruction [7], where the sparsifying dictionary was learnt directly from the measured data, resulting in accurate reconstructions for up to six fold undersampling.

In this work, we propose to use a dictionary with both temporal and local spatial information for parametric map estimation. We create a training set by using the Bloch equations to simulate the temporal signal response over a predefined spatial distribution obtained from anatomical images and train a spatiotemporal dictionary by clustering similar patches. The trained dictionary is incorporated into a patch-based iterative projection algorithm to estimate T1 and T2 parametric maps. We see two main benefits of our approach:

1. Incorporating spatial data increases the atom length, i.e. the amount of descriptive information available per voxel, requiring less temporal points for an accurate reconstruction.
2. Training improves the conditioning of the dictionary by creating atoms distinct to each other, leading to a better signal matching.

The rest of this paper is structured as follows. In Sect. 2 we describe the method, in particular the proposed patch-based algorithm for MRF. Section 3

depicts the experiments and demonstrates the application of recovering parametric maps from undersampled data, and in Sect. 4 we offer conclusions.

2   Methods

The goal of MRF is to obtain parametric maps $\boldsymbol{\theta} \in \mathbb{R}^{N \times Q}$ from a sequence of undersampled measurements $\mathbf{Y} \in \mathbb{C}^{M \times T}$, where Q is the number of tissue relaxation parameters (T1 and T2), T is the sequence length, every map $\boldsymbol{\theta}_q \in \mathbb{R}^N$ has a total of N voxels, every measurement $\mathbf{y}_t \in \mathbb{C}^M$ is sampled M times, and $M \ll N$. This is achieved in three steps: image reconstruction, template matching, and parameter extraction.

Image reconstruction is the task of obtaining the image sequence $\mathbf{X} \in \mathbb{C}^{N \times T}$ from the measurements $\mathbf{Y}$. This is generally formulated as a inverse problem: $\mathbf{Y} = \mathbf{E}\mathbf{X}$, where $\mathbf{E} \in \mathbb{C}^{M \times N}$ is the encoding operator. The reconstructed image is then matched to a precomputed dictionary $\mathbf{D} \in \mathbb{C}^{T \times L}$ of L atoms, to find the dictionary atom $\mathbf{d}_l \in \mathbb{C}^T$ that best describes it. This is done at every voxel location $\mathbf{x}_n \in \mathbb{C}^T$ by selecting the entry l_n that maximizes the modulus of the atom and the conjugate transpose of the signal:

$$\hat{l}_n = \arg\max_{\substack{l \\ l=1,\dots,L}} |\mathbf{x}_n^* \mathbf{d}_l| \tag{1}$$

where both, $\mathbf{d}_l$ and $\mathbf{x}_n$, were previously normalized to have unitary length. Finally, the T1 and T2 parameters used to construct the matching entry are assigned to the voxel n, creating $\boldsymbol{\theta}_n = \{T1_n, T2_n\}$. Thus, by repeating the matching over all voxels of the image, the parametric T1 and T2 maps are found.

Davies et al. [3] interpret the template matching as a projection of $\mathbf{x}_n$ onto the cone of the Bloch response manifold, and propose an iterative projection algorithm to accurately extract parametric maps. The algorithm, termed Bloch response recovery via iterated projection (BLIP), iteratively alternates between a gradient step, a projection step, and a shrinkage step to reconstruct the image sequence $\mathbf{X}$ and estimate the corresponding parameter maps $\boldsymbol{\theta}$.

2.1   Spatiotemporal Dictionary Design

Given a set of fully sampled 2D spatial parametric maps $\boldsymbol{\theta} \in \mathbb{R}^{N \times Q}$, where $N = N_i \times N_j$ and $Q = 2$, an image sequence $\mathbf{X} \in \mathbb{C}^{N \times T}$ of T temporal points can be created at each voxel using the Bloch equations to simulate the magnetization response of an inversion-recovery balanced steady state free-precession (IR-bSSFP) sequence with pseudorandomized acquisition parameters (see Fig. 1) [6]. $\mathbf{X}$ can be processed to create a spatiotemporal dictionary as follows.

Let $\mathbf{R}_n \in \mathbb{C}^{P \times N}$ be the operator that extracts 2D image patches of size $P = P_i \times P_j$, so that the spatiotemporal image patch $\tilde{\mathbf{x}}_n \in \mathbb{C}^{P \times T}$ at a given spatial location n is given by

$$\tilde{\mathbf{x}}_n = \mathbf{R}_n \mathbf{X}. \tag{2}$$
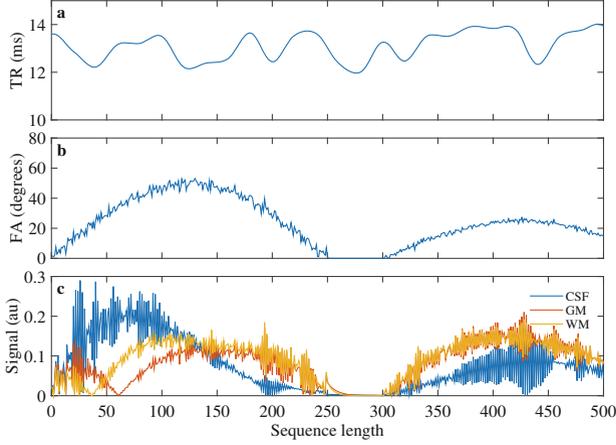
Fig. 1. Pseudorandom acquisition sequence and the corresponding signal response. **a**, TR values following a Perlin noise pattern. **b**, Flip angle series of repeating sinusoidal curves and added random values. **c**, Signal evolution for different tissue classes: white matter (WM), grey matter (GM), and cerebrospinal fluid (CSF).

It is then possible to create the patch-based image matrix $\tilde{\mathbf{X}} \in \mathbb{C}^{PT \times N}$ by concatenating the vector representation of every spatiotemporal patch of dimension $P_i \times P_j \times T$ for each spatial location in $\mathbf{X}$. Repeating the operation on $\boldsymbol{\theta}$ creates the patch-based multiparametric matrix $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^{PQ \times N}$. The spatiotemporal dictionary $\tilde{\mathbf{D}} \in \mathbb{C}^{PT \times K}$ is then constructed by using k-means to cluster atoms in $\tilde{\mathbf{X}}$ with similar signal values into K clusters, averaging the corresponding T1 and T2 values in $\tilde{\boldsymbol{\theta}}$ to create the clustered patch-based matrix $\boldsymbol{\Theta} \in \mathbb{C}^{PQ \times K}$, and simulating the signal evolution for each cluster. A new simulation of the signal evolution ensures that the atoms in $\tilde{\mathbf{D}}$ correspond exactly to the entries in $\boldsymbol{\Theta}$.

2.2 Patch-Based BLIP Reconstruction (P-BLIP)

The BLIP algorithm [3] reconstructs the image sequence $\mathbf{X}$ in an iterative fashion. Given an image sequence $\mathbf{X}^{(i)}$ at iteration i, the reconstructed sequence $\mathbf{X}^{(i+1)}$ in the next iteration is determined by

$$\mathbf{X}^{(i+1)} = \mathcal{P}_{\mathcal{A}}(\mathbf{X}^{(i)} + \mu \mathbf{E}^H(\mathbf{Y} - \mathbf{E}\mathbf{X}^{(i)})), \tag{3}$$

where $\mathcal{P}_{\mathcal{A}}$ represents the projection onto the signal model $\mathcal{A}$, $\mathbf{E}^H$ is the Hermitian adjoint of the encoding operator, and μ equals the step size. P-BLIP builds on this algorithm, incorporating the patch extraction operator in (2) and an update step to make (3) applicable to a spatiotemporal signal model.

At every iteration the updated sequence $\mathbf{X}$ is transformed into the patch-based matrix $\tilde{\mathbf{X}}$ by (2). $\tilde{\mathbf{X}}$ is related to the trained dictionary $\tilde{\mathbf{D}}$ by

$$\tilde{\mathbf{X}} = \tilde{\mathbf{D}}\mathbf{W}, \tag{4}$$

where $\mathbf{W} \in \mathbb{R}^{K \times N}$ represents the weights. Equation 4 can be readily solved using greedy algorithms that find sparse solutions to linear systems of equations by adding a sparsity constraint to the ℓ_0-norm of each column vector $\mathbf{w}_n$:

$$\hat{\mathbf{W}} = \arg\min_{\mathbf{W}}\|\tilde{\mathbf{X}} - \tilde{\mathbf{D}}\mathbf{W}\|_2^2, \text{ s.t. } \|\mathbf{w}_n\|_0 \leq \gamma, \ n = 1, ..., N. \tag{5}$$

We set the sparsity constraint to $\gamma = 1$, equivalent to finding one dictionary atom, as done in the template matching used in [3,6].

After estimating the weights, the patch-based image matrix is projected onto the dictionary by $\tilde{\hat{\mathbf{X}}} = \tilde{\mathbf{D}}\hat{\mathbf{W}}$. At this point, each voxel is overrepresented a total of P times, requiring an update step to return to the original image sequence $\mathbf{X}$. This update is achieved by averaging the P temporal signals that contribute to a given voxel location. Finally, the parametric maps $\boldsymbol{\theta}$ are estimated by applying the weights and patch-wise updates on $\boldsymbol{\Theta}$.

3   Experiments and Results

Image Data. Experiments were performed using twenty digital brain phantoms from BrainWeb [2]. Of these, ten were used to train the spatiotemporal dictionary and ten to test the performance of three different reconstruction algorithms: the original MRF reconstruction [6], BLIP [3], and the proposed P-BLIP. Experiments were designed to evaluate the performance of each algorithm as a function of sequence length and acceleration factors, and, for the case of P-BLIP, also as a function of spatial patch size. Ground truth datasets were generated by selecting a slice of crisp datasets labeled with different tissue classes, and resampling them to a matrix size of 256×256 to accelerate computations. Quantitative maps were then obtained by replacing the tissue labels with their corresponding T1 and T2 values. The values for the three main tissue types grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) were equaled to those reported in [6], while the values for the rest of the classes (fat, bone, muscle, vessels, dura matter, and connective tissue) were obtained directly from [1].

Modeling the Signal Evolution. At every voxel, the ground truth quantitative maps served as a basis to simulate the temporal evolution of the signal based on the IR-bSSFP pulse sequence with acquisition parameters displayed in Fig. 1, where the TRs follow a Perlin noise pattern, FAs are a series of repeating sinusoidal curves with added random values, and the radio frequency phase alternates between 0° and 180° on consecutive pulses. Off-resonance frequencies were not taken into account. This pulse sequence was combined with all possible combinations of a given range of T1 and T2 values to create a temporal dictionary used in both MRF and BLIP. The selected range was reported in [3], where T1 spans from 100 ms to 6000 ms and T2 from 20 ms to 1000 ms, both sampled at varying step sizes. Additionally, the dictionary included the exact T1 and T2 combinations corresponding to the different tissue classes.

Spatiotemporal Dictionary. To train the spatiotemporal dictionary used in P-BLIP, a region of interest that accounted for the entire head area was defined.
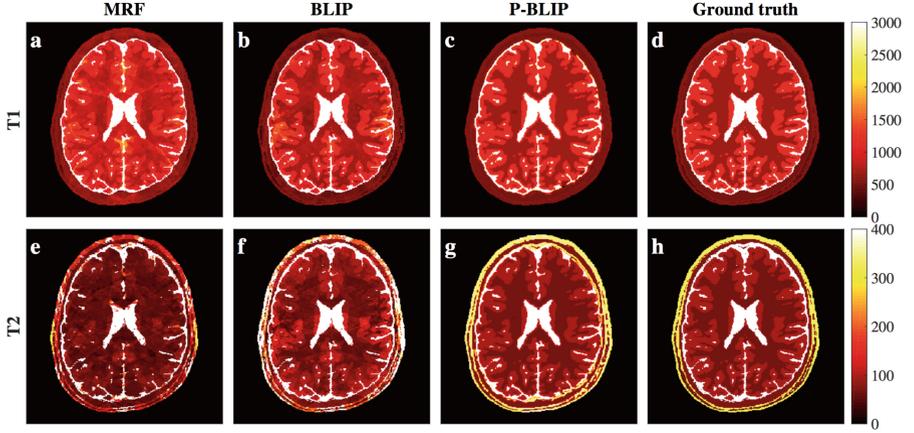
Fig. 2. Exemplary reconstruction results of one dataset with $T = 200$, $R = 10$, and $P = 3 \times 3$. The upper row shows T1 maps for all algorithms and the ground truth; and the bottom row the corresponding T2 maps. Most visible in T2 maps, subsampling artefacts can be effectively removed with P-BLIP.

The space covered by this region of interest was randomly and equally subsampled and each of the subsampled sets was assigned to a training subject. The selected parametric maps of each subject were then used as an input to train the dictionary as described in Sect. 2.1 with a total of $K = 200$ clusters.

Subsampling Strategy. We use a random EPI subsampling strategy for all experiments: the k-space is fully sampled in the read direction (k_x) and uniformly undersampled in the phase encoding direction (k_y) by an acceleration factor R. The sampling pattern is shifted by a random a number of k_y lines at every shot of the sequence.

Experimental Setup. An initial experiment was performed with spatiotemporal patches of size $3 \times 3 \times 200$ and an acceleration factor $R = 10$ to visually evaluate the reconstructed maps (see Fig. 2). Subsequently, three experiments assessed the reconstruction performance with respect to sequence length, acceleration factor and spatial patch size. The first experiment varied sequence lengths from 100 to 500 in step sizes of 100, the second experiment used acceleration factors of $R = \{2, 5, 10, 15, 20\}$, and the final experiment used spatial patches of sizes $P = \{1 \times 1, 3 \times 3, 5 \times 5, 7 \times 7\}$. The reconstruction error of the first two experiments was calculated using the signal-to-error ratio (SER) in decibels (dB), defined as $20 \log_{10} \frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2}$; and the third experiment with the SSIM values [8].

Results. Figure 2 displays the reconstructed parametric maps of an exemplary dataset. The MRF estimates show the characteristic ghosting artefacts caused by sub-Nyquist sampling. BLIP removes most of these artefacts from the T1 estimation, though they are still visible in the T2 maps. P-BLIP effectively removes these artefacts from both maps, resulting in reconstructions very close to the ground truth. These visual observations can be confirmed with quantitative
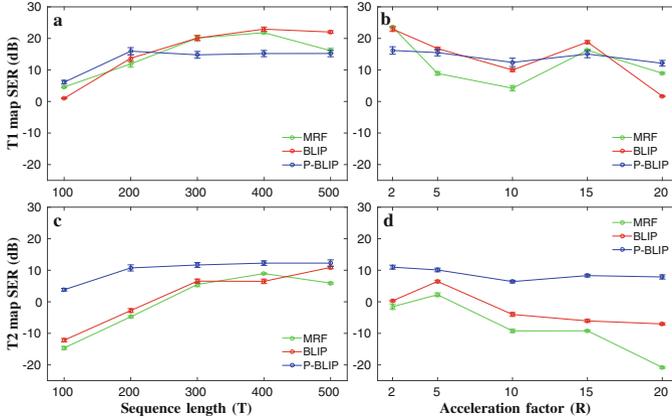
Fig. 3. a,c, Performance as a function of sequence length with $R = 10$, $P = 3 \times 3$; and **b,d**, as a function of acceleration factor with $T = 200$, $P = 3 \times 3$. P-BLIP is best in estimating T2 maps and shows better results for shorter sequences and higher acceleration for T1 maps.

Table 1. Average SSIM values for T1 and T2 map estimation with respect to different spatial patch sizes, $T = 200$ and $R = 10$.

Method	Baseline		Proposed: P-BLIP			
	MRF	BLIP	1×1	3×3	5×5	7×7
T1	0.761	0.814	0.848	**0.852**	0.691	0.625
T2	0.616	0.591	0.769	**0.857**	0.667	0.601

results. Figure 3c and d show how P-BLIP achieves better T2 estimates independently of the sequence length or acceleration factor. On the other hand, T1 maps for P-BLIP remain relatively constant for sequence lengths larger than 100 (Fig. 3a) and all acceleration factors (Fig. 3b), whilst the performance of MRF and BLIP increases with the sequence length and lower acceleration factors. The reason for these results is twofold. First, the IR-bSSFP sequence is mostly T1-weighted, favoring a better T1 matching over T2 matching for all methods. Second, a trained dictionary containing a longer sequence, but fixed K, is less flexible, and if the trained dictionary does not exactly contain the ground truth values, the quantitative error will be higher.

Table 1 indicates the performance of P-BLIP for different patch sizes in comparison to the performance of MRF and BLIP. A spatial patch size of $P = 1 \times 1$ implies that the training dataset was created from voxel-wise temporal evolutions and that the trained dictionary is a clustered version of the temporal dictionary. It can be seen that clustering a temporal dictionary alone improves the reconstruction with respect to MRF and BLIP, and that the spatiotemporal dictionary further improves these results for $P = 3 \times 3$. At larger spatial patch sizes the results begin to decline, indicating that the cluster size of $K = 200$ is not enough to capture the entire spatial variability of the parametric maps.

4 Conclusions

This work presents a novel patch-based reconstruction scheme for MRF consistent with the theory of CS. It is based on a spatiotemporal signal model and relies on the training of the corresponding dictionary from a set of examples. This patch-based scheme shows improved performance for shorter pulse sequences and at higher acceleration factors, leading to an increased efficiency of parameter mapping with MRF.

An important discussion point of our approach is the size of the dictionary in terms of space, time, and atoms. Larger spatial patches allow, in theory, for the acquisition of less temporal points, but the amount of atoms in the dictionary should in turn be large enough to account for large spatial variability. We have seen from our results that a dictionary size of $K = 200$ is not enough for spatial patch sizes larger than 3×3 for structures in the brain. A potencial solution to this shortcoming might be to make K dependant on the atom length or arbitrarily large at the cost of computational complexity. This point is currently under investigation and future work will focus on extending the method to incorporate 3D spatial patches and applying it to real datasets.

References

1. Aubert-Broche, B., Evans, A.C., Collins, L.: A new improved version of the realistic digital brain phantom. NeuroImage **32**, 138–145 (2006)
2. Aubert-Broche, B., Griffin, M., Pike, G.B., Evans, A.C., Collins, D.L.: Twenty new digital brain phantoms for creation of validation image data bases. IEEE Trans. Med. Imaging **25**(11), 1410–1416 (2006)
3. Davies, M., Puy, G., Vandergheynst, P., Wiaux, Y.: A compressed sensing framework for magnetic resonance fingerprinting. SIAM J. Imaging Sci. **7**(4), 2623–2656 (2014)
4. Doneva, M., Börnert, P., Eggers, H., Stehning, C., Sénégas, J., Mertins, A.: Compressed sensing reconstruction for magnetic resonance parameter mapping. Magn. Reson. Med. **64**, 1114–1120 (2010)
5. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theor. **52**, 1289–1306 (2006)
6. Ma, D., Gulani, V., Seiberlich, N., Liu, K., Sunshine, J.L., Duerk, J.L., Griswold, M.A.: Magnetic resonance fingerprinting. Nature **495**, 187–192 (2013)
7. Ravishankar, S., Bresler, Y.: MR image reconstruction from highly undersampled k-space data by dictionary learning. IEEE Trans. Med. Imaging **30**(5), 1028–1041 (2011)
8. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Proc. **13**, 600–612 (2004)