Dictionary learning and time sparsity in dynamic MRI

Jose Caballero[1], Daniel Rueckert[1], and Joseph V. Hajnal[2]

[1] Department of Computing, Imperial College London, UK
[2] Institute of Clinical Sciences, Imperial College London and MRC Clinical Sciences Centre, Hammersmith Hospital, London, UK
{jose.caballero06, d.rueckert, jo.hajnal}@imperial.ac.uk

Abstract. Sparse representation methods have been shown to tackle adequately the inherent speed limits of magnetic resonance imaging (MRI) acquisition. Recently, learning-based techniques have been used to further accelerate the acquisition of 2D MRI. The extension of such algorithms to dynamic MRI (dMRI) requires careful examination of the signal sparsity distribution among the different dimensions of the data. Notably, the potential of temporal gradient (TG) sparsity in dMRI has not yet been explored. In this paper, a novel method for the acceleration of cardiac dMRI is presented which investigates the potential benefits of enforcing sparsity constraints on patch-based learned dictionaries and TG at the same time. We show that an algorithm exploiting sparsity on these two domains can outperform previous sparse reconstruction techniques.

1 Introduction

Medical research and diagnosis rely heavily on magnetic resonance imaging (MRI) for the visualisation of anatomical structures and physiological function. This technology is non-invasive, non-ionizing and offers an unmatched quality in soft tissues contrast. However, physical and physiological limits on scanning speed make this an inherently slow process. Sampling constraints are particularly challenging for dynamic MRI (dMRI). The use of compressed sensing (CS) theory [1, 2] has been shown repeatedly to be successful at reducing acquisition time. The philosophy behind CS is that if a signal is known a-priori to be sparse in some transform domain, much fewer samples are needed for its acquisition than those dictated by the Nyquist rate.

Spatio-temporal correlations were early on exploited for MRI acceleration by k-t BLAST/SENSE [3], but with the introduction of Lustig's developments [4] they could be linked to CS theory under a more general framework named k-t FOCUSS [5, 6]. Structural dMRI such as cardiac cine, has mainly exploited sparsity in the x-f space, which is obtained through a Fourier transform of the data along time [7, 8]. The wavelet domain has also been proposed for the sparse representation of spatial information [9]. However, fixed basis transforms such as the aforementioned have recently been outperformed by adaptive transforms

resulting from dictionary learning (DL) [10] for the case of 2D structural MRI reconstruction [11, 12]. Also, dMRI objects are known to vary through time at a few constrained locations, e.g. edges of a heart, and Fourier transforms along time are not able to capture information about the location of temporal changes.

In this paper, a novel algorithm is proposed for the acceleration of cardiac dMRI acquisition. The contribution is two-fold: First, DL is extended to the dynamic case for the first time through the use of 3D dictionaries as an attempt at finding a better sparsifying domain than those obtained with fixed basis. Secondly, sparsity is additionally enforced on the temporal gradient (TG) of the reconstruction. This simple transform has not received much attention in previous techniques although it can greatly improve the reconstruction performance by capturing very sparse constrained spatio-temporal changes.

The paper is organised as follows: Section 2 introduces the theory behind the compressed sensing MRI (CSMRI) problem and the sparsifying transforms exploited in the method proposed. Our algorithm is described in section 3 and its performance is evaluated and compared with the k-t FOCUSS algorithm [5] in section 4. We summarise the characteristics of the novel technique and future work in section 5.

2   Compressed sensing in dynamic MRI

2.1   The CSMRI problem

Assuming a desired cardiac sequence to be represented as a 3D volume of size $P = P_x \times P_y \times P_t$, time being the third dimension, the column vector $\mathbf{x}_d \in \mathbb{R}^P$ can be built as a concatenation of its columns. Denote $\hat{\mathbf{x}}_d$ as the Fourier transform of $\mathbf{x}_d$ such that $\mathcal{F}\{\mathbf{x}_d\} = \hat{\mathbf{x}}_d$. The acquisition of MR data takes place in the frequency domain, referred to as k-space. To accelerate the process, only $m << P$ k-space samples are acquired according to the sampling subset Ω, which are represented by $\hat{\mathbf{x}}_u \in \mathbb{C}^m$. If the sampling process is assumed to be corrupted by additive white Gaussian noise (AWGN) $\mathbf{n} \in \mathbb{C}^m$, then we can write $\hat{\mathbf{x}}_u = \mathcal{F}_u\{\mathbf{x}_d\} + \mathbf{n}$, where $\mathcal{F}_u\{.\}$ is the undersampling Fourier operator. Given that there exists a sparsifying transform $\mathcal{S}\{.\}$ such that $\|\mathcal{S}\{\mathbf{x}_d\}\|_0 << P$, the problem of recovering $\mathbf{x}_d$ from $\hat{\mathbf{x}}_u$ reduces to solving

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

where ϵ is a small constant accounting for sampling noise in k-space. In general both k-space data and the resulting image space are complex, although it is common to reconstruct magnitude images. In this paper fully sampled magnitude images have been used as gold standard input data, implying pure real objects.

Three requirements ensure that the underdetermined system of equations (1) has a unique and attainable solution [1]: (a) $\|\mathcal{S}\{\mathbf{x}_d\}\|_0$ must be sparse or compressible, (b) aliasing has to be incoherent in the transform domain and (c) there must be a non-linear reconstruction process able to enforce sparsity and data-consistency at the same time. The latter two conditions are easily

met in MRI with the use of random undersampling masks [4] and a variety of optimisation algorithms [13]. The success of a CSMRI method therefore relies heavily on the choice of the transform $\mathcal{S}\{.\}$, which should sparsify information as much as possible.

2.2 Sparsifying transforms in dynamic MRI

Local learned sparsity An alternative to fixed compression transforms are learned dictionaries. These are constructed through a training process based on example patches extracted from one or multiple signals sharing similar features with the target $\mathbf{x}_d$. Because they are adaptive and more flexible, they can provide sparser representations by discarding redundancies specific to a signal [10], therefore enhancing the potential of the CSMRI framework.

The learning of a dictionary is an optimisation problem in which a dictionary containing N patches of n pixels as columns in a matrix $\mathbf{D} \in \mathbb{R}^{n \times N}$ is tailored to sparsely represent $M > N$ training patches. Available algorithms such as the K-SVD [10] can perform this for 3D training patches and dictionary atoms, suitable for the coding of image sequences. Learning bases for natural video has been successfully tested [14], but the complexity involved can scale very fast if objects in the sequence are rapidly changing because at each time instance the dictionary must be relearned. This disadvantage, however, does not apply to dMRI where the signal's characteristics undergo minimal and locally constrained changes through time and a fixed dictionary is enough for the coding of an entire sequence.

Global temporal gradient sparsity Total variation (TV) sparsity has been exploited for dMRI reconstruction [15] and is generally included as an extra constraint in sparse reconstructions under the assumption that medical images are piecewise smooth. Nevertheless, the analysis of finite-difference sparsity along the three dimensions of a typical cardiac sequence reveals that temporal finite-differences, which approximate TG, are sparser than any other transform based on spatio-temporal finite-differences.

We empirically compared the sparsity of TV and finite-differences along the three dimensions from a cardiac sequence $\mathbf{x}_d$ of size $128 \times 128 \times 30$ that will be used in experiments of section 4. The l_1 norm, a usual sparsity metric, for these transforms was $\|TV\{\mathbf{x}_d\}\|_1 = 2.9 \times 10^4$, $\|\nabla_x\{\mathbf{x}_d\}\|_1 = 1.7 \times 10^4$, $\|\nabla_y\{\mathbf{x}_d\}\|_1 = 1.8 \times 10^4$ and $\|\nabla_t\{\mathbf{x}_d\}\|_1 = 6.2 \times 10^3$, where $\nabla_f\{.\}$ is the finite-difference operator along dimension f and $TV\{\mathbf{x}\} = \sqrt{\nabla_x\{\mathbf{x}\}^2 + \nabla_y\{\mathbf{x}\}^2 + \nabla_t\{\mathbf{x}\}^2}$.

3 Proposed method

The method suggested for undersampled dMRI reconstruction consists in an iterative refinement of the zero-filled sequenced $\mathbf{x}_{zf} \in \mathbb{R}^P$ until sparsity constraints under a learned dictionary and TG are simultaneously met. An additional step

ensures k-space samples of the solution in the sampling subset Ω are consistent with acquired samples in $\hat{\mathbf{x}}_{zf}$. The method, referred to as dictionary learning temporal gradient (DLTG), is broken down into three main blocks, in which an approximate generic solution $\mathbf{y} \in \mathbb{R}^P$ is refined towards a better solution $\mathbf{x} \in \mathbb{R}^P$. Algorithm 1 summarises the solution proposed.

– **Subproblem 1:** Local patch sparse coding

$$\min_{\boldsymbol{\Gamma}} \|\boldsymbol{\gamma}_i\|_0 \qquad s.t. \qquad \|\mathbf{R}_i\mathbf{y} - \mathbf{D}\boldsymbol{\gamma}_i\|_2^2 < \epsilon, \forall i \tag{2}$$

Assuming a dictionary $\mathbf{D}$ is trained on $\mathbf{y}$, this step will approximate $\mathbf{y}$ with a sparse representation under $\mathbf{D}$ up to a small error ϵ. A patch $\mathbf{R}_j\mathbf{y}$ is extracted with the matrix operator $\mathbf{R}_j$ and coded in $\boldsymbol{\gamma}_j$ for each pixel of $\mathbf{y}$. Patches wrap around sequence edges, meaning each pixel is contained by a total of n patches. Problem (2) can be efficiently solved through orthogonal matching pursuit [16].

Once a sparse coding is obtained for all pixels in $\mathbf{y}$, the output sequence is recovered through

$$\mathbf{x} = \frac{\sum_{j=1}^P \mathbf{R}_j^T \mathbf{D} \boldsymbol{\gamma}_j}{n}, \tag{3}$$

which numerically is interpreted as the superposition of all coded patches according to their spatial correspondence within $\mathbf{x}$, and the averaging of their contribution at each pixel position.

– **Subproblem 2:** Global TG sparse coding

$$\min_{\mathbf{x}} \lambda\|\nabla_t\{\mathbf{x}\}\|_1 + \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \tag{4}$$

The solution of (4) will be close to $\mathbf{y}$ in the least squares sense and will have sparse TG. This can be solved using an interior-point method with an empirical complexity of $O(P^{1.2})$ [17]. To solve equation (4) we operate on $\nabla_t\{\mathbf{x}\}$ directly and update $\mathbf{x}$ for the second term using forward cumulative sums of the TG. This requires the DC level to be set in an independent step. Further details are excluded here due to space restrictions.

– **Subproblem 3:** Data consistency in k-space

$$\hat{\mathbf{x}}(\mathbf{k}) = \begin{cases} \hat{\mathbf{y}}(\mathbf{k}) & , \mathbf{k} \notin \Omega \\ \frac{\hat{\mathbf{y}}(\mathbf{k})+\nu\hat{\mathbf{x}}_{zf}(\mathbf{k})}{1+\nu} & , \mathbf{k} \in \Omega \end{cases} \tag{5}$$

The entire process relies on acquired k-space samples $\mathbf{k} = (k_x, k_y, t) \in \Omega$, which are non-zero samples in $\hat{\mathbf{x}}_{zf}$. Hence, the k-space representation $\hat{\mathbf{x}}$ of the solution will have to be consistent with these. A weighting factor $\nu = q/\sigma$ determines how much these samples are trusted considering noise power σ in k-space.

Algorithm 1 DLTG

Input: Undersampled zero-filled k-space data - $\hat{\mathbf{x}}_{zf}$
Output: Reconstructed MR sequence - $\mathbf{x}_{I_1,I_2}$
Initialise: $\mathbf{x}_{0,0} = \mathbf{x}_{zf} = \mathcal{F}^{-1}\{\hat{\mathbf{x}}_{zf}\}$
for $k = 0$ **to** $I_1 - 1$ **do**
    1. Train $\mathbf{D}$ using patches from $\mathbf{x}_{k,0}$
    2. Obtain local patch sparse coding $\mathbf{\Gamma}$ of $\mathbf{x}_{k,0}$ under $\mathbf{D}$ by solving (2)
    3. Update $\mathbf{x}_{k,0}$ based on $\mathbf{\Gamma}$ through (3)
    4. Enforce data consistency on $\mathbf{x}_{k,0}$ by (5)
    for $i = 0$ **to** $I_2 - 1$ **do**
        5. Enforce global TG sparsity on $\mathbf{x}_{k,i}$ solving (4)
        6. Enforce data consistency on $\mathbf{x}_{k,i}$ by (5) and store result in $\mathbf{x}_{k,i+1}$
    end
    7. Update reconstruction: $\mathbf{x}_{k+1,0} \leftarrow \mathbf{x}_{k,I_2}$
end

Many of the parameters do not require very precise settings and can be chosen from a reasonably large spectrum without excessively affecting the end result [11]. This is the case for patch sizes n, the number of dictionary atoms N, the number of training patches M and the scaling factor q. Stopping criteria are chosen as maximum number of iterations I_1 and I_2, although convergence to stable solutions is also a valid condition. Variables ϵ and λ set a compromise on the performance and speed of the algorithm. The smaller they are, the better the result but the slower the convergence rate. Ideally, a regularisation path would optimally initialise them with large values and decrease them at each iteration. For the production of results in section 4, they were sequentially updated such that at outer iteration k, $\epsilon_k = \sqrt{mse_{k,0} \times n}$, and at inner iteration i, $\lambda_{k,i} = \sqrt{mse_{k,i}}/2$, where $mse_{k,i} = \mathbb{E}\{\|\mathbf{x}_{k,i} - \mathbf{x}_d\|_2^2\}$ and $\mathbb{E}\{.\}$ is the expectation operator. The use of the gold standard $\mathbf{x}_d$ is exclusive for demonstration purposes. Fixed parameters ϵ and λ have been tested to provide the same results simply slower.

4   Experimental results

In the following experiments, a cardiac sequence of size $128 \times 128 \times 30$ with normalised intensity is artificially corrupted by multiplying its k-space representation with a binary undersampling mask and by adding complex AWGN of power σ. Assuming a Cartesian trajectory sampling, acquisitions of the mask are randomly drawn from a 1D Gaussian distribution prioritising the rejection of high frequency components and are independent realisations for each frame.

Results are all drawn using a dictionary of $N = 300$ cubic atoms of size $n = 125$ trained from $M = 5000$ patches. The noise regularisation factor is chosen to be $q = 150$ from empirical observation, but results were found to be reasonably stable (less than 0.5 PSNR variation) for a range between 60 and 200. The algorithm is run using $I_1 = 20$ and $I_2 = 10$ iterations, and its performance is evaluated in terms of its Peak Signal-to-Noise Ratio (PSNR). The result from
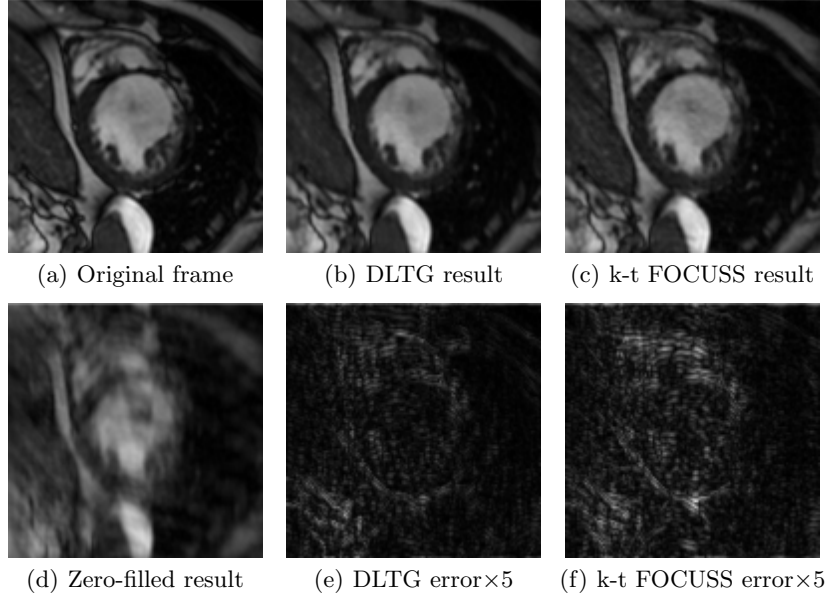
(a) Original frame   (b) DLTG result   (c) k-t FOCUSS result

(d) Zero-filled result   (e) DLTG error×5   (f) k-t FOCUSS error×5

Fig. 1. One frame in a reconstruction of a sequence acquired at a 0.15 sampling factor.

k-t FOCUSS [5] is provided as a reference x-f based method and an algorithm exploiting only DL sparsity (ignoring steps 5 and 6 in algorithm 1) is assessed in order to understand the benefits of imposing TG sparsity on the solution.

4.1   Visual appearance

A frame from the reconstructed sequence is shown in figure 1. The sampling factor used in this example was $m/P = 0.15$ and a noise-free acquisition was assumed. Amplified error images are also shown. The DLTG method is able to recover the sequence with PSNR = 31.6dB from the original PSNR = 23.7dB of the zero-filled sequence compared to the PSNR = 29.7dB of k-t FOCUSS.

4.2   Noiseless case

Figure 2(a) shows the performance of the proposed method assuming $\sigma = 0$ for different sampling factors. For very high ones (above 0.5), the reconstruction power is completely attributed to the enforcement of sparsity under the learned dictionary. As the sampling is reduced, more aliasing is introduced in $\mathbf{x}_{zf}$ and the dictionary starts incorporating it as part of the sequence features to be learned, therefore lowering the performance of DL. The aliasing reproduced, however, prevents the reconstruction from meeting the sparsity requirement in the TG domain, which is why enforcing it is beneficial for the end result. The algorithm is shown to be superior to k-t FOCUSS for all sampling factors and the same was observed for 20 different short axis data as well as on long axis examples.
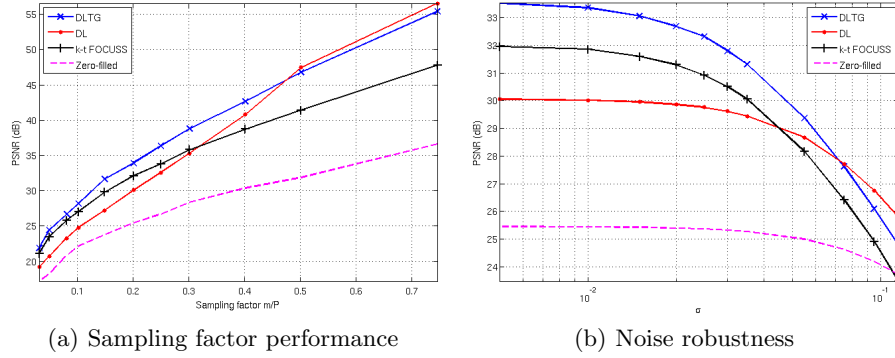
(a) Sampling factor performance   (b) Noise robustness

Fig. 2. Compared performance of the DLTG algorithm and k-t FOCUSS.

4.3   Noisy case

In figure 2(b), a sampling factor of 0.2 is imposed on the acquisition and the reconstruction performance is evaluated as a function of noise power σ. The spectrum covered is SNR = 34.5dB ($\sigma = 0.005$) up to SNR = 7.2dB ($\sigma = 0.115$) and the DLTG method is consistently superior to k-t FOCUSS. Also, it can be observed that the gain obtained from the TG constraint becomes smaller as noise is increased and its enforcement can be detrimental for high levels of noise.

5   Conclusion and future work

A novel reconstruction method for the acceleration of dMRI acquisition called DLTG has been proposed which, for the first time, adapts DL to dMRI data and introduces the TG transform as a suitable sparsifying transform. A dictionary learned from the acquired k-space data provides a sparse representation that can outperform methods based on fixed basis transforms. Sparsity enforced on TG is seen as the best performing one out of those exploiting spatio-temporal finite-differences and effectively reduces aliasing further whenever the DL constraint alone is not sufficient. On top of being reasonably robust to noise, results show that the method provides better results than the k-t FOCUSS method.

One important improvement of the DLTG method needs to explore the reconstruction of complex data. All the results were based on the assumption that the target sequence was real, but information such as blood flow in cardiac dMRI is encoded in complex format. Also, a more efficient way to combine DL and TG sparsity will be imperative to widen the algorithm's applicability. The DLTG algorithm is nevertheless an encouraging first attempt at using learned bases for dMRI and sets path for the extension of learned sparsity to other MRI modalities such as functional or parallel MRI, with the expectation of experiencing the same benefits previously observed in 2D MRI and confirmed here for structural dMRI.

References

1. Candes, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory **52**(2) (2006) 489–509
2. Donoho, D.: Compressed sensing. IEEE Transactions on Information Theory **52**(4) (2006) 1289–1306
3. Tsao, J., Boesiger, P., Pruessmann, K.P.: k-t BLAST and k-t SENSE: Dynamic MRI with high frame rate exploiting spatiotemporal correlations. Magnetic Resonance in Medicine **50**(5) (2003) 1031–1042
4. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: The application of compressed sensing for rapid MR imaging. Magnetic Resonance in Medicine **58**(6) (2007) 1182–1195
5. Jung, H., Ye, J.C., Kim, E.Y.: Improved k-t BLAST and k-t SENSE using FOCUSS. Physics in Medicine and Biology **52**(11) (2007) 3201–3226
6. Jung, H., Sung, K., Nayak, K.S., Kim, E.Y., Ye, J.C.: k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI. Magnetic Resonance in Medicine **61**(1) (2009) 103–116
7. Gamper, U., Boesiger, P., Kozerke, S.: Compressed sensing in dynamic MRI. Magnetic Resonance in Medicine **59**(2) (2008) 365 – 373
8. Usman, M., Prieto, C., Schaeffter, T., Batchelor, P.G.: k-t group sparse: A method for accelerating dynamic MRI. Magnetic Resonance in Medicine **66**(4) (2011) 1163–1176
9. Lustig, M., Santos, J.M., Donoho, D.L., Pauly, J.M.: k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity. In Proc. of the 13th Annual Meeting of ISMRM, Seattle **50**(5) (2006) 2420
10. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: Design of dictionaries for sparse representation. In Proc. of SPARS **5** (2005) 9–12
11. Ravishankar, S., Bresler, Y.: MR image reconstruction from highly undersampled k-space data by dictionary learning. IEEE TMI **30**(5) (2011) 1028–1041
12. Chen, Y., Ye, X., Huang, F.: A novel method and fast algorithm for MR image reconstruction with significantly under-sampled data. Inverse Problems and Imaging **4**(2) (2010) 223–240
13. Tropp, J.A., Wright, S.J.: Computational methods for sparse solution of linear inverse problems. In Proc. of the IEEE **98**(6) (2010) 948–958
14. Protter, M., Elad, M.: Image sequence denoising via sparse and redundant representations. IEEE Transactions on Image Processing **18**(1) (2009) 27–35
15. Montefusco, L.B., Lazzaro, D., Papi, S., Guerrini, C.: A fast compressed sensing approach to 3D MR image reconstruction. IEEE TMI **30**(5) (2011) 1067–1075
16. Pati, Y., Rezaiifar, R., Krishnaprasad, P.: Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In Proc. of 27th Asilomar Conference on Signals, Systems and Computers. **1** (2011) 40–44
17. Kim, S.J., Koh, K., Lustig, M., Boyd, S., Gorinevsky, D.: An interior-point method for large-scale l1-regularized least squares. IEEE Journal of Selected Topics in Signal Processing **1**(4) (2007) 606–617