

Imperial College of Science, Technology and Medicine  
Department of Computing

# **Dictionaries for fast and informative dynamic MRI acquisition**

Jose Caballero

A dissertation submitted in part fulfilment of the requirements for the degree of  
**Doctor of Philosophy**  
of  
**Imperial College London**

May 2015



## Abstract

Magnetic resonance (MR) imaging is an invaluable tool for medical research and diagnosis but suffers from inefficiencies. The speed of its acquisition mechanism, based on sequentially probing the interactions between nuclear atom spins and a changing magnetic field, is limited by atomic properties and scanner physics. Modern sampling techniques termed compressed sensing have nevertheless demonstrated how near perfect reconstructions are possible from undersampled, accelerated acquisitions, showing promise for more efficient MR acquisition paradigms. At the same time, information extraction from MR images through image analysis implies a considerable dimensionality reduction, in which an image is processed for the extraction of a few clinically useful parameters. This signals an inefficient handling of information in the separated treatment of acquisition and analysis that could be tackled by joining these two essential stages of the imaging pipeline.

In this thesis, we explore the use of adaptive sparse modelling for novel acquisition strategies of cardiac cine MR data. Conventional compressed sensing MR acquisition relies on fixed basis transforms for sparse modelling, which are only able to guarantee suboptimal sparse modelling. We introduce spatio-temporal dictionaries that are able to optimally adapt sparse modelling by absorbing salient features of cardiac cine data, and demonstrate how they can outperform sampling methods based on fixed basis transforms. Additionally, we extend the framework introduced to handle parallel data acquisition. Given the flexibility of the formulation, we show how it can be combined with a labelling model that provides a segmentation of the image as a by-product of the reconstruction, hence performing joint reconstruction and analysis.



## Acknowledgements

The past years have been a period of great change and growth in my life in which I have had the pleasure of working with extraordinary people. My graduate degree culminates in this thesis, but the experiences and friendships gathered along the way will shape me for the rest of my life.

I would like to first of all thank my supervisors Daniel Rueckert and Jo Hajnal, for giving me the opportunity to work alongside them. Daniel and Jo have been a constant source of support and inspiration throughout my degree, and our many conversations have defined my abilities to think about and conduct research. I could not have wished for a better supervision tandem, for their enthusiasm and curiosity about finding things out is exceptional.

The BioMedIA group at Imperial College London has been my second home. I am very thankful to everyone who made it an enjoyable and stimulating environment, especially to Ozan, Kevin, Salim and Jonathan for good memories, and Kanwal, Wenjia and Christian for their help and insightful feedback on my work.

I need to thank the people in the Division of Imaging Sciences and Biomedical Engineering Department team at King's College London for their contributions. In particular Anthony N. Price, for educating me on practical MRI scanning and for providing almost all the imaging data in this thesis.

I would also like to thank Sara for her loving support. Her stoic tolerance to my complaints when work was not going well has been a very pleasing personal discovery that is also the result of this thesis. For this and for her constant company, which has made my life so happy, I am extremely grateful.

Finally, I am thankful to my family, my mum, dad and sister, for their encouragements and their interest in my research, and for always having great advice whenever I need it.

*To my grandparents*

# Declaration of originality

I declare that the work presented in this thesis is my own, unless specifically acknowledged.

Jose Caballero

## **Copyright Declaration**

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

# Acronyms

**AWG** additive white Gaussian. 65, 83, 99, 100, 110

**CG** conjugate gradient. 104, 106, 107, 125, 143, 144

**CS** compressed sensing. 3, 4, 6, 9, 28, 30, 32–34, 36, 37, 42, 52, 56–58, 60, 61, 63, 65, 66, 81, 95–98, 100–102, 108, 117, 120, 126, 134, 140, 149, 150, 152, 161, 163, 166–169, 173

**DCT** discrete cosine transform. xx, xxi, 51, 52, 54–56, 75, 92–95, 150

**DFT** discrete Fourier transform. 20, 36, 56, 64, 75, 83

**DL** dictionary learning. 6, 8, 9, 43, 46–48, 54, 55, 59–61, 63–66, 84, 98, 102, 112, 166–168, 170, 173

**DLMRI** dictionary learning MRI. xxii, 74, 75, 78–85, 91

**DLTG** dictionary learning with temporal gradient. xxii, 73, 75, 78–84, 86, 87, 89, 91, 92, 170

**EM** expectation-maximisation. 137, 141, 142, 144, 145, 147, 148, 152, 157, 158, 162

**FOCUSS** focal underdetermined system solver. xxii, 42, 78–84, 90, 95

**FOV** field of view. xix, 21, 22, 27, 114, 120, 134

**GMM** Gaussian mixture model. 9, 135, 136, 138–142, 144, 146, 147, 149, 150, 156, 157, 159, 162, 169

**IID** independent and identically distributed. 94

**IRLS** iteratively re-weighted least squares. 41, 42

**LS** least-squares. 41, 47, 97

**ML** maximum likelihood. 47, 48, 136

**MOD** method of optimal directions. 47, 49

**MP** matching pursuit. 38, 39

**MR** magnetic resonance. i, xviii, xxi, 1–5, 7–10, 12, 22, 26, 28, 32, 53, 56, 59, 61–66, 95, 103, 108, 114, 133, 141, 147, 148, 150, 161, 163, 165–168, 172, 173

**MRF** Markov random field. 9, 147–149, 157, 158, 162, 163, 168

**MRI** magnetic resonance imaging. xviii, xix, 1, 3, 4, 6, 8–11, 15, 18, 20, 22–31, 36, 37, 56, 57, 59–65, 69, 77, 86, 94–98, 100–102, 126, 132–135, 162, 164–170, 173

**MSE** mean squared error. 51, 52, 155, 156

**MSSIM** mean structural similarity index. 76, 79, 80, 113

**NMR** nuclear magnetic resonance. xviii, 11–13, 15

**OMP** orthogonal matching pursuit. xx, 39, 48, 49, 51, 52, 54, 66, 70, 75, 89–91, 103, 109, 125, 141, 144, 161, 168

**PDF** probability density function. xxvi, 149

**PSNR** peak signal-to-noise ratio. xxii, 78–80, 83, 85, 89, 94, 113, 115, 167

**RF** radio-frequency. 12–16, 18, 22, 23

**RIP** restricted isometry property. 36, 37

**RMSE** root mean squared error. xxi, 54, 55

**ROI** region of interest. xxi, xxii, 76, 77, 79, 80, 153, 155, 167

**SNR** signal-to-noise ratio. xxiii, 100, 101, 108, 110–113, 123, 124

**SoS** sum of squares. 100, 112

**SVD** singular value decomposition. 48, 49, 51, 106

**TG** temporal gradient. 73, 74, 79, 80, 82, 91, 95

**TV** total variation. 8, 57, 59, 60, 63, 66, 67, 98

# Nomenclature

<b>D</b>	Dictionary of atoms.
<b>F</b>	2D DFT Fourier transform for single coil image.
$\tilde{\mathbf{F}}$	2D DFT Fourier transform for parallel coil data.
<b>G</b>	SPIRiT convolution kernel weights for parallel k-space recombination, or MRF matrix.
$\mathbf{\Gamma}$	Sparse coding matrix.
$\gamma_n$	Sparse coding vector, set as column $n$ in $\mathbf{\Gamma}$ .
$K$	Number of atoms in a dictionary.
$L$	Number of GMM components.
$M$	Dimension for undersampled data.
<b>M</b>	2D undersampling mask for single coil image.
$\tilde{\mathbf{M}}$	2D undersampling mask for parallel coil data.
$\{\mu, \sigma, \pi\}$	GMM parameters for mean, standard deviation and component mixing weights.
$N$	Dimension for fully sampled data.
$N_p$	Dimension for data patch.
$N_T$	Number of K-SVD training patches.

$R$	Acquisition acceleration factor.
$\mathbf{R}_n$	Patch extracting operator at space-time location $n$ .
$\mathbf{x}$	MR image reconstruction.
$\mathbf{x}_n$	MR image patch
$\hat{\mathbf{x}}$	Fourier transform of MR image.
$\mathbf{x}_f$	Fully sampled MR image.
$\mathbf{x}_D$	Reconstruction compliant with a patch-based dictionary sparse constraint.
$\mathbf{x}_{TG}$	Reconstruction compliant with a temporal gradient sparse constraint.
$\mathbf{x}_z$	Reconstruction with zero-filled k-space.



# Contents

<b>Abstract</b>	<b>i</b>
<b>Acronyms</b>	<b>vii</b>
<b>Nomenclature</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Challenges in MRI acquisition and analysis . . . . .	2
1.1.1 Acquisition speed . . . . .	3
1.1.2 Information extraction . . . . .	4
1.1.3 Reuse of past data . . . . .	4
1.2 Objectives and motivation . . . . .	6
1.3 Contributions . . . . .	8
1.4 Thesis overview . . . . .	9
<b>2 MRI principles</b>	<b>10</b>
2.1 Introduction . . . . .	10
2.2 Nuclear magnetic resonance . . . . .	11
2.2.1 Spin polarisation . . . . .	11
2.2.2 Excitation . . . . .	12
2.2.3 Relaxation . . . . .	13
2.2.4 Bloch equation . . . . .	14
2.3 Magnetic resonance imaging . . . . .	15
2.3.1 Encoding gradients . . . . .	15
2.3.2 NMR signal measurement . . . . .	18
2.3.3 K-space . . . . .	19
2.3.4 FOV and resolution . . . . .	20
2.4 Acquisition acceleration . . . . .	22
2.4.1 Explicit redundancy: Parallel imaging . . . . .	24

2.4.2	Implicit redundancy: Image correlation . . . . .	26
2.5	Conclusion . . . . .	28
<b>3</b>	<b>Adaptive compressed sensing</b>	<b>29</b>
3.1	Introduction . . . . .	29
3.2	The need for modern sampling . . . . .	30
3.3	Compressed sensing . . . . .	31
3.3.1	Signal sparsity . . . . .	32
3.3.2	Sampling incoherence . . . . .	35
3.3.3	Non-linear reconstruction . . . . .	38
3.4	Dictionary learning . . . . .	42
3.4.1	From linear transforms to adaptive dictionaries . . . . .	43
3.4.2	Dictionary learning problem . . . . .	46
3.4.3	Dictionary learning algorithms . . . . .	46
3.5	Patch-based dictionary sparse coding . . . . .	49
3.5.1	Overcompleteness . . . . .	50
3.5.2	Redundancy . . . . .	52
3.5.3	Adaptability . . . . .	53
3.6	CS and DL in dynamic MRI . . . . .	56
3.6.1	Non-adaptive sparsity . . . . .	57
3.6.2	Adaptive sparsity . . . . .	59
3.7	Conclusion . . . . .	61
<b>4</b>	<b>Dictionary learning for dynamic MRI</b>	<b>62</b>
4.1	Introduction . . . . .	63
4.2	Sparsifying transforms for CS dynamic MRI . . . . .	64
4.2.1	Adaptive spatio-temporal sparsity . . . . .	65
4.2.2	Temporal gradient sparsity . . . . .	66
4.3	Reconstruction with sparse model . . . . .	67
4.4	DLTG algorithm . . . . .	70
4.4.1	Dictionary training and sparse coding . . . . .	70
4.4.2	Temporal gradient filtering . . . . .	71
4.4.3	Acquisition data consistency . . . . .	71

4.4.4	Algorithm design . . . . .	73
4.4.5	Algorithm complexity . . . . .	75
4.5	Experiments and results . . . . .	76
4.5.1	Experimental setup . . . . .	76
4.5.2	Reconstruction of individual data sets . . . . .	78
4.5.3	Reconstruction with noise . . . . .	83
4.5.4	Spatial dictionaries . . . . .	85
4.5.5	Real and complex-valued dictionaries . . . . .	86
4.5.6	Parameter selection . . . . .	87
4.5.7	Algorithm convergence, speed and acceleration . . . . .	89
4.5.8	Influence of training data . . . . .	92
4.6	Conclusion . . . . .	94
<b>5</b>	<b>Parallel MRI reconstruction with dictionaries</b>	<b>96</b>
5.1	Introduction . . . . .	96
5.2	Compressed sensing parallel MRI . . . . .	98
5.2.1	SPIRiT . . . . .	100
5.2.2	$l_1$ -SPIRiT . . . . .	101
5.3	D-SPIRiT: Parallel MRI with dictionaries . . . . .	102
5.3.1	Problem formulation . . . . .	103
5.3.2	Solution with variable splitting and penalty method . . . . .	104
5.3.3	Implementation considerations . . . . .	106
5.4	Experiments and results . . . . .	107
5.4.1	Experimental setup . . . . .	108
5.4.2	Retrospective phantom tests . . . . .	111
5.4.3	Retrospective cardiac cine tests . . . . .	114
5.4.4	Prospective cardiac tests . . . . .	120
5.4.5	Parameter selection . . . . .	123
5.4.6	Algorithm speed and acceleration . . . . .	125
5.5	Conclusion . . . . .	126
<b>6</b>	<b>Application-driven MRI</b>	<b>131</b>
6.1	Introduction . . . . .	131

6.2	Joint reconstruction-segmentation . . . . .	134
6.3	Mixture model segmentation . . . . .	135
6.3.1	Gaussian mixture modelling . . . . .	135
6.3.2	EM for GMM fitting . . . . .	137
6.4	Reconstruction-segmentation for fast MRI . . . . .	140
6.4.1	Update of sparse model with OMP . . . . .	141
6.4.2	Update of GMM with EM . . . . .	141
6.4.3	Update of reconstruction with CG . . . . .	142
6.4.4	Avoiding EM trivial solution . . . . .	145
6.4.5	GMM extensions . . . . .	146
6.5	Experiments and results . . . . .	148
6.5.1	Impact of discriminative modelling on reconstruction . . . . .	150
6.5.2	Left ventricle cardiac cine segmentation . . . . .	151
6.5.3	Brain tissue segmentation . . . . .	156
6.5.4	Algorithm speed and acceleration . . . . .	161
6.6	Discussion . . . . .	161
6.7	Conclusion . . . . .	163
<b>7</b>	<b>Conclusion</b>	<b>165</b>
7.1	Achievements . . . . .	166
7.2	Limitations and future work . . . . .	168
7.3	Discussion . . . . .	170

# List of Tables

4.1	MSSIM comparison (mean $\pm$ std ) $\times 10^{-2}$ . . . . .	79
4.2	Training data influence on PSNR (dB) with $\epsilon = 0.01$ . . . . .	93
4.3	Training data influence on PSNR (dB) with $\epsilon = 0.007$ . . . . .	94
4.4	Influence of dictionary initialisation on PSNR (dB). . . . .	94
5.1	Subsampling mask design properties. . . . .	110
5.2	PSNR (dB) reconstruction performance mean $\pm$ standard deviation of retrospectively undersampled 32 coil data from 7 patients. . . . .	116
5.3	ISMRM challenge data specifications. . . . .	117
5.4	Prospective data with regular static undersampling acceleration specifications. . . . .	121
5.5	Prospective data with regular circularly shifted undersampling acceleration specifications. . . . .	122
6.1	Dice metrics for segmentation of a fully sampled data set. . . . .	158

# List of Figures

1.1	Examples of MR image modalities. Structural MR for brain (a) and knee (b) are static scans and functional brain MR (c) and cardiac cine (d) are dynamic modalities. . . . .	2
1.2	Conceptual differences in MR information extraction. Traditional MR analysis assumes a full image acquisition, which is later processed to obtain relevant clinical information (a). There is a significant dimensionality reduction between the fully sampled image and the extracted parameters that signals an inefficient handling of information. Modern approaches could ideally look for the low dimensional information of interest directly from incomplete raw data, minimising the discard of acquired information (b). Compressed sensing techniques would be a special case of this framework, where the signal of interest is the intensity image itself. . . . .	5
1.3	Number of imaging tests carried out by the NHS in England [101] (a). Magnetic resonance imaging (MRI) is the imaging modality with the highest cumulative increase rate in the last 20 years with over a 200% increase (b). . . . .	6
2.1	Different atomic nuclei phases in the nuclear magnetic resonance (NMR) experiment. Low-energy state nuclei are represented in light blue and high-energy state nuclei are dark blue. After spin polarisation with the $B_0$ field, the system of spins can be excited resulting in the transverse magnetisation $\mathbf{M}_{xy}$ . The bulk magnetisation relaxes at the end of the excitation, returning a state where it again aligns with the $B_0$ field. . . . .	13

2.2	Encoding gradients for spatial localisation. A slice selection gradient can create slabs of signal with different shifts from the Larmor frequency (a). This principle is applied three dimensions to fully encode the signal in space. . . . .	16
2.3	Spatial encoding in a 2D slice through phase and frequency encoding. Without gradient encoding, the signal received is a contribution of identically defined nuclei (a). The joint application of a phase (b) and a frequency (c) encoding allows to uniquely describe each location with different phases. . . . .	17
2.4	Simulated brain image [95] showing different contrast patterns. This is an example of the great contrast flexibility enabled by the physical principles underlying MRI. . . . .	18
2.5	Traversing k-space with the application of gradient fields. . . . .	20
2.6	Pulse sequence diagram representing the coordination of gradient pulses. Slice selection in the $z$ direction is applied with the RF excitation. Gradients to encode 2D spatial information are then applied sequentially, with phase encoding in the $y$ dimension and frequency encoding and read-out in the $x$ dimension. . . . .	21
2.7	K-space trajectories . . . . .	21
2.8	Field of view (FOV) and resolution relationships between k-space (a) and image domain (b). . . . .	22
2.9	Undersampling strategies in k-space (top) and the aliasing they produce in image domain (bottom). . . . .	24
2.10	Parallel MRI acquisition using 4 coils. The data acquired by each coil is filtered through its own sensitivity pattern. . . . .	25
3.1	Estimation of the amount of data produced and stored [132]. In 2007, the amount of information created surpassed the storage capabilities. . . . .	29
3.2	Timeline of signal modelling trends as summarised in [48] (left), and physical interpretation of the sparse synthesis model [105] (right). Natural signals are assumed to be a sparse linear combination of transform functions grouped in a basis or frame $\Psi$ . . . . .	33

3.3	Optimisation result using different norms. The intersection of the solution set (in green) and a norm ball represent a solution. The $l_p$ ball, $0 \leq p < 1$ , grows along the axes, hence finding a sparse solution, but is not a convex set. The $l_1$ norm is a convex set that can approximate the sparse result. . . . .	41
3.4	K-SVD dictionary update step. The left hand side of the figure shows the decomposition of the approximation error in $\mathbf{E}^k$ and the contribution from $\mathbf{d}_k$ and $\boldsymbol{\gamma}_t^k$ through a rank-1 matrix. Updating $\mathbf{d}_k$ and $\boldsymbol{\gamma}_t^k$ directly with an SVD decomposition of $\mathbf{E}^k$ does not guarantee the maintenance of sparsity in $\boldsymbol{\gamma}_t^k$ . Instead, shrinking the matrices as shown in the right hand side of the figure solves this problem as only the active support of $\boldsymbol{\gamma}_r^k$ is updated. . . . .	50
3.5	Dictionary overcompleteness translates into increased representation sparsity. The plot in (c) shows the average representation error of $10^4$ patches of size $8 \times 8$ from image (a) using a discrete cosine transform (DCT) dictionary of different sizes and a sparsity index $s = 6$ . Examples of image patches are shown in (b). . . . .	51
3.6	OMP sparse recovery examples. Artificial sparse signals of size $N = 8 \times 8$ are synthesised from $s = 5$ randomly chosen dictionary atoms from a DCT dictionary of size $K = 225$ . Compared are the best (a) and worst (b) recoveries with respect to MSE, out of $10^4$ tries. . . . .	53
3.7	Average empirical orthogonal matching pursuit (OMP) recovery performance using an overcomplete DCT dictionary. The support recovery rate falls almost to zero for $s/N = 10/64 = 0.15$ sparsity (a). This is due to the high degree of redundancy in the dictionary, which makes OMP choose sparse coding configurations that are not the ones used to generate the original patches. Nevertheless, OMP is able to maintain the predefined data consistency tolerance (b) by using a few additional atoms relative to the original signal (c). . . . .	54
3.8	Effect of K-SVD training on dictionary and on training dataset. An initial overcomplete DCT dictionary (a), is trained to yield an adapted dictionary (b). The representation accuracy of training signals improves over training iterations (c). . . . .	54

3.9	Sparse coding reconstruction root mean squared error (RMSE) of brain MR image. Adapting a dictionary to the brain image reduces the representation error with respect to it, while increasing the error with respect to data of different nature, such as a Gaussian noise image. . . . .	55
3.10	Sparse coded approximations of a brain MR image using a DCT dictionary (a, b) and a trained dictionary (c, d) with a sparsity index $s = 3$ . Error maps show absolute value differences with respect to the original image. . . . .	56
3.11	Incoherent subsampling of k-space. A variable density distribution is chosen to prioritise the acquisition of low-frequency content (a, b) from which a mask can be obtained (c). The aliasing produced by zero-filling non-acquired samples is shown in (d). . . . .	57
3.12	A cardiac cine dataset (a) and some examples of non-adaptive sparsity (b, c, d).	58
4.1	Comparison of the decay in transform coefficients of transforms $\nabla_x$ , $\nabla_y$ and $\nabla_t$ . The latter shows the fastest decay, implying that the signal energy concentrates in fewer coefficients than in the other gradient dimensions. . . . .	67
4.2	Visualisation of significant non-zero coefficients of transforms $\nabla_x$ , $\nabla_y$ and $\nabla_t$ . Important coefficients in $\nabla_t$ are concentrated around the dynamic region of the beating heart, whereas changes in other locations are kept below 0.1 and could be attributed to noise. . . . .	67
4.3	Algorithm flowchart for the DLTG (solid arrows) and the DLMRI (dashed arrows) algorithms. . . . .	74
4.4	Example of 2D temporal frame of one of the data sets used for testing. A region of interest (ROI) and a profile line are defined that will be used for evaluations.	77
4.5	Example of a magnitude temporal frame from one of the data sets analysed (a). The undersampling mask (b) applied in k-space reduces acquisition time but introduces aliasing in image space (c). All figures show a 2D frame on the left hand side and the temporal profile across the dashed line on the right hand side.	78

4.6	Mean and standard deviation peak signal-to-noise ratio (PSNR) performance of reconstructions from 10 scans retrospectively undersampled without added noise. Results are shown for the whole data sets (a) and within a ROI around the heart (b). . . . .	79
4.7	Visual comparison of a fully sampled magnitude frame (a), its undersampled by 8 zero-filled version (e), and reconstructions using k-t focal underdetermined system solver (FOCUSS) (b), dictionary learning MRI (DLMRI) (c) and dictionary learning with temporal gradient (DLTG) (d) with their respective errors multiplied by 6 (f, g, h). . . . .	80
4.8	Temporal profile of line shown in figure 4.4 in the fully sampled data set (a), and reconstructions from a 8 fold acceleration using k-t FOCUSS (b), DLMRI (d) and DLTG (f) with their respective errors amplified by 6 (c, e, g). . . . .	81
4.9	Visual comparison of a fully sampled magnitude frame (a), its undersampled by 15 zero-filled version (e), and reconstructions using k-t FOCUSS (b), DLMRI (c) and DLTG (d) with their respective errors multiplied by 6 (f, g, h). . . . .	82
4.10	Temporal profile of line shown in figure 4.4 in the fully sampled data set (a), and reconstructions from a 12 fold acceleration using k-t FOCUSS (b), DLMRI (d) and DLTG (f) with their respective errors amplified by 6 (c, e, g). . . . .	83
4.11	Phase reconstruction comparison in a scan accelerated by 8 of the fully sampled data (a) and the reconstructions provided by k-t FOCUSS, DLMRI and DLTG (b, c, d). . . . .	83
4.12	Algorithm evaluation with added noise. Reconstruction sensitivity to noise power and parameter $q$ for a scan accelerated by 4 (left) using the DLTG (solid) and DLMRI (dashed) algorithms, and their robustness to input noise (right). . . . .	84
4.13	Visual comparison of the reconstructions from a scan accelerated by 4 that has been contaminated by complex noise with input $\text{PSNR}_f = 25.8$ dB (a) using k-t FOCUSS (b), DLMRI (e) and DLTG (d) with their respective errors multiplied by 6 (f, g, h). The noise added to the fully sampled data set is shown in (e) amplified by 6. . . . .	85

4.14	Comparison of 2D spatial and 3D spatio-temporal reconstruction. Shown are a fully sampled frame (a), its undersampled by 6 zero-filled version (b), and the spatial only (c) and spatio-temporal (d) reconstructions. All figures are shown with their respective temporal profiles on the right hand side. . . . .	86
4.15	Average error per patch produced by the assumption of $s$ sparsity on a fully sampled scan using different dictionaries. . . . .	88
4.16	Influence of dictionary size (a) and patch size (b) on the reconstruction of a scan accelerated by 6. . . . .	88
4.17	MSE convergence of approximate results $\mathbf{x}_D$ and $\mathbf{x}_{TG}$ relative to the ground truth (a), and the intermediate sparsity level of the real, imaginary and training patches coding (b). . . . .	89
4.18	Example of a reconstruction's dependence to $\epsilon$ (a) and the Batch-OMP runtime for one coding of real and imaginary parts of a scan (b). . . . .	90
4.19	Convergence rate of DLMRI and DLTG for two different data fitting terms setups.	92
4.20	Examples of patches learnt in the last iteration of the 8 fold accelerated DLTG reconstruction shown in figure 4.7. . . . .	93
5.1	SPIRiT parallel reconstruction of a phantom image simulating 8 coils. Acceleration rates $R > 4$ present a much lower signal-to-noise ratio (SNR) than for $R \leq 4$ . . . . .	101
5.2	$l_1$ -SPIRiT parallel reconstruction of a phantom image simulating 8 coils. The regularisation with a sparsity term can reduce noise amplification stemming from parallel recombination instability. . . . .	101
5.3	Mask designs used for testing: 1D regular (a), 1D regular circularly shifted (b), 1D uniform Poisson (c), 1D uniform random (d), 2D variable density Poisson (e), 2D uniform Poisson (f), and 2D variable density random (g). . . . .	109
5.4	Simulated parallel coil sensitivity patterns. We assume the generation of 8 sensitivity patterns, where the 2D Gaussian function in (a) changes its center for different coils, and the phase image in (b) is rotated. The coil phase is a cropped section of the phase image pattern shown in (b). . . . .	111

5.5	Data used for phantom experiments. Using 8 coils, we show one coil example ground truth (a), the image with added noise equivalent for a received k-space SNR of 30 dB (b), the SoS recombination ground truth (c) and the same recombination from noisy coil images (d). . . . .	112
5.6	Reconstruction of a 6 fold accelerated 8 coil simulation of a phantom image with 30 dB k-space SNR. The SPIRiT reconstruction (a) suffers from noise amplification, which can be controlled by $l_1$ -SPIRiT regularisation (b), but especially by an overcomplete (c) and adaptive (d) dictionary regularisation. . . . .	113
5.7	Performance evaluation of SPIRiT and its sparsity regularised versions for 40 dB k-space SNR at different undersampling ratios. . . . .	114
5.8	Performance evaluation of SPIRiT and its sparsity regularised versions for 6 fold acceleration at different k-space SNR. . . . .	114
5.9	Reconstruction from 6.25 fold 1D regular circularly shifted undersampling. . . .	115
5.10	Reconstruction from 7.69 fold 1D regular circularly shifted undersampling. . . .	116
5.11	ISMRM cardiac cine reconstruction results. Coloured bars show reconstruction performance of competing methods, which are undisclosed, and the proposed technique D-SPIRiT is shown in black. . . . .	118
5.12	Examples of long-axis reconstructions from the ISMRM challenge dataset. We show the mask and DCT-SPIRiT reconstructions for cases 4 (a, b) and 5 (c, d). . . .	119
5.13	Examples of short-axis reconstructions from the ISMRM challenge dataset. We show the mask and DCT-SPIRiT reconstructions for cases 7 (a, b) and 8 (c, d). . . .	119
5.14	Single frames from original fully-sampled datasets used in prospective undersampled tests. We examine results using regular static undersampling on dataset (a) and regular circularly shifted undersampling on dataset (b). . . . .	122
5.15	Exploration of tuning parameter $\lambda$ in the SPIRiT initialisation for the D-SPIRiT algorithm. The tuning is dependent on the added k-space SNR, but not on the nature, dimension of the data, or undersampling rate, as shown using a cardiac (solid) and the Shepp-Logan phantom (dashed). . . . .	124

5.16	Look-up tables for tuning parameters derived from manual explorations of different noise levels. . . . .	124
5.17	Prospective undersampling reconstructions using regular mask. The comparison is between SENSE (left column), SPIRiT (middle column) and DCT-SPIRiT (right column), and acceleration factors are 2, 4, 6 and 8, from top to bottom rows. . . . .	127
5.18	Zoomed ROI of prospective undersampling reconstructions using regular mask. The comparison is between SENSE (left column), SPIRiT (middle column) and DCT-SPIRiT (right column), and acceleration factors are 2, 4, 6 and 8, from top to bottom rows. . . . .	128
5.19	Prospective undersampling reconstructions using circularly shifting regular mask. The comparison is between k-t SENSE (left column), SPIRiT (middle column) and DL-SPIRiT (right column), and acceleration factors are 2, 4, 6 and 8, from top to bottom rows. . . . .	129
5.20	Zoomed ROI of prospective undersampling reconstructions using circularly shifting regular mask. The comparison is between k-t SENSE (left column), SPIRiT (middle column) and DL-SPIRiT (right column), and acceleration factors are 2, 4, 6 and 8, from top to bottom rows. . . . .	130
6.1	Traditional imaging pipeline. . . . .	132
6.2	Toy image segmentation with GMM. A $L = 3$ component mixture model (b) is fit to the histogram of intensity image (a) producing segmentation (c). . . . .	137
6.3	Phantom brain MR image [95]. . . . .	139
6.4	GMM estimation and segmentation with EM algorithm. At the top of each image is shown the histogram of intensity values and the model fit, with axes probability against image intensity. . . . .	139

- 6.5 Joint reconstruction-segmentation trivial solution and EM instability. Each image shows the intensity reconstruction (top), the intensity histogram and GMM model fit, where axes are probability against intensity (middle), and the segmentation derived from the model fit (bottom). There exists a trivial solution to the global problem posed for which  $\sigma_l = 0$ . The figures show how the left-most Gaussian in the model collapses after iteration 8, and this destabilises the joint reconstruction-segmentation as the model needs to readjust itself and compensate for the missing cluster. . . . . 145
- 6.6 Undersampling mask simulation. A 2D Gaussian probability density function (PDF) is generated (a) from which a mask can be obtained as the result of a Bernoulli process (c). In (b) we show a single column of the pdf in (a). The variance of the PDF be modified to obtain different undersampling rates. . . . . 149
- 6.7 Visual appearance of reconstruction with different values of  $\beta$ . Each image shows the intensity histogram, where axes are probability against intensity (top), and the intensity image reconstruction (bottom). A joint segmentation is obtained for  $\beta > 0$ . As  $\beta$  increases pixel intensities are encouraged to cluster around Gaussian means, which forces a separation between different clusters in the intensity histograms shown above. This has an edge enhancement effect that is visually apparent in the images with highest  $\beta$  values. . . . . 151
- 6.8 Cardiac cine data (a) is segmented with a GMM of  $K = 3$  components (b) resulting in a segmentation (c) that uses a single component to label blood pool regions. . . . . 152
- 6.9 Cardiac cine reconstruction of a fully sampled data set (a) accelerated by 10 and reconstructed using the separate (b) and joint (c) methods. . . . . 153

6.10	Left ventricle segmentation from data undersampled at rate 24 seen on a temporal slide (a) and a temporal profile (b) corresponding to the horizontal line in (a). The segmentation of the fully sampled data (purple) is displayed along with deviations corresponding to errors in the segmentation of accelerated data that are exclusive to the separate method (red), the joint method (blue), or are shared by both methods (white). . . . .	154
6.11	Pixel misclassification rate for left ventricle segmentation of undersampled cardiac cine data. . . . .	156
6.12	Reconstruction and segmentation errors (a, b) produced at different acceleration rates by the joint (green) and separate (blue) approaches. The rate of improvement as a percentage is less pronounced for the reconstruction (c) than for the segmentation (d). . . . .	156
6.13	Comparison of brain tissue segmentation using the EM algorithm. The baseline EM method (b) is unable to provide an accurate result relative to a manual segmentation (a). The introduction of spatial probabilistic priors allows for a more precise result (c) and the use of MRF regularisation can additionally discard improbable configurations (d). . . . .	158
6.14	Brain tissue segmentation from data undersampled by 8. The joint method (c) is able at some locations, such as those highlighted by arrows, to better approximate the manual segmentation (a) than the separate approach (b). . . .	159
6.15	Average tissue segmentation Dice from undersampled data (excluding CSF). . . .	160
7.1	Detail loss at high undersampling factors shown for a single coil simulation. The case of 4 fold undersampling can recover fine details, but at high factors such as 10 fold fine details are lost by the patch-based regularisation method. . . . .	171



# Chapter 1

## Introduction

Magnetic resonance imaging (MRI) is today an indispensable tool for medical diagnosis and research. Its success can largely be explained by the detailed soft tissue contrast it provides and by the great flexibility enabled by its acquisition mechanism [19, 145]. Moreover, and contrary to other established imaging modalities, MRI is non-ionising and non-invasive. However, its acquisition mechanism and image analysis pipeline pose a number of efficiency challenges that once tackled could make acquisitions with this imaging modality shorter, cheaper and more comfortable. The sampling enabled by a magnetic resonance (MR) scanner is sequential, and the rate at which data can be collected is limited by physical and physiological constraints. The budget for data sampling is particularly tight for dynamic scans such as in cardiac cine image acquisition, where spatial and temporal sampling requirements present a fundamental compromise.

In this thesis, we explore the use of adaptive sparse modelling for novel acquisition strategies of cardiac cine MR data. Building upon methods that propose sparse modelling for the reconstruction of data sampled at sub-Nyquist rates, we experiment with patch-based approaches which are able to optimally adapt sparse modelling for a particular dataset. Additionally, we extend the framework introduced to allow for parallel data acquisition and to provide an analysis of the image as a by-product of the reconstruction. In the remainder of this chapter,

we expose the challenges addressed by this thesis, as well as the objectives and contributions expected.

## 1.1 Challenges in MRI acquisition and analysis

The flexibility of the MR acquisition process has allowed producing images of different body regions, and to tune acquisition parameters for the enhancement of different tissue attributes. A brief illustration of the diversity of MR images is shown in figure 1.1, ranging from structural, to functional and dynamic imaging.

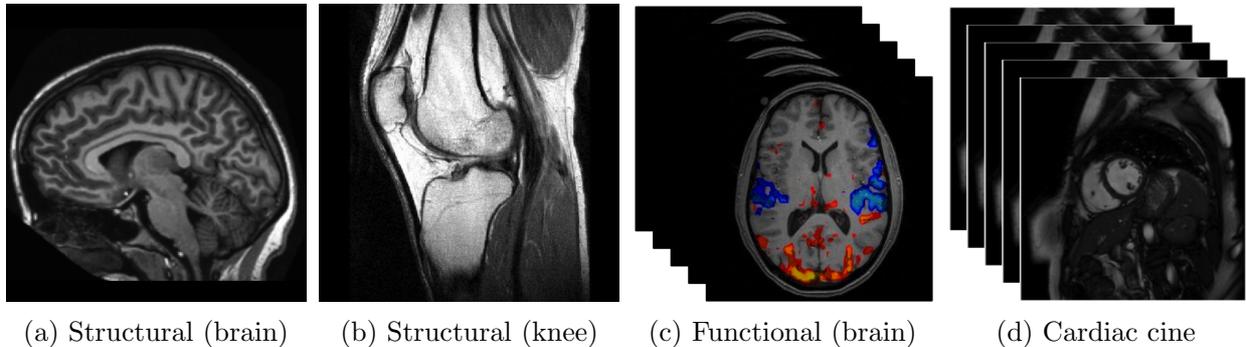


Figure 1.1: Examples of MR image modalities. Structural MR for brain (a) and knee (b) are static scans and functional brain MR (c) and cardiac cine (d) are dynamic modalities.

However, the current state of scanners and the established approaches to image processing make the acquisition and analysis of the examples shown inefficient. First, the sequential nature of MR acquisition limits the speed at which data can be collected. Furthermore, the images illustrated are often not an end in the imaging pipeline, but rather a means to obtain clinically useful information through post-processing stages. Also, the processing of medical images without the use of knowledge from past example cases implies a waste of resources given the amount of MR information produced every day. We detail below these three inefficiencies of current MR procedures, which we identify as challenges to be addressed in this thesis.

### 1.1.1 Acquisition speed

The production of MR signals relies on a physical mechanism imposing that the targeted body region is scanned along a trajectory in a sequential fashion. This is an inefficient strategy for data sampling which translates into inherently slow scanning procedures. It is particularly problematic in dynamic scans where the signal is acquired both in space and time and the resolution budget to be shared among both domains is rigidly constrained. One example of interest is cardiac cine imaging, which allows the imaging of the beating heart and is crucial for the prevention, diagnosis and treatment of cardiovascular diseases.

Despite the notable advances in MR hardware design with parallel coils, reducing the acquisition time of cardiac cine is still a challenging task, but a vast number of post-processing techniques based on modern sampling concepts of compressed sensing (CS) have recently shown promising potential to tackle this problem [86]. State-of-the-art developments, notably those combining compressed sensing (CS) with parallel MRI, provide satisfactory results at acceleration factors close to 8-fold [69, 85]. The approach considers the acquisition of only a fraction of the data normally necessary for a scan, hence linearly reducing acquisition time, and imposes a sparsity model onto the reconstruction to compensate for missing data.

An important drawback of this solution is the use of *a priori* assumptions which can sometimes try to fit the reconstruction to inaccurate models. Fixed-basis transforms designed to provide sparse representations for structured signals are popular for sparsifying spatial information [84, 97, 1] and, for the special case of dynamic MRI, a Fourier transform along time has been the preferred choice [86, 55, 69, 138]. However, it is generally agreed that not one single sparsity model can optimally fit all signals, and despite the clear indications that the search for suitable sparsifying bases is necessary, the focus has been on trying to develop better algorithms for sub-optimal models rather than trying to redefine the model itself.

### 1.1.2 Information extraction

MR images contain a large amount of data. In many cases, the utility of the image itself is to support a diagnosis or a differential diagnosis based on the extraction of clinical parameters from routine post-processing techniques. From an information theoretic point of view, the handling of MR information is extremely inefficient, given that to arrive at a few clinical parameters we require the acquisition and processing of images that contain thousands to millions of pixels or voxels. A common example is the processing of cardiac cine images, which consists of high dimensional spatio-temporal data very usually reduced to a small number of clinical measurements such as ejection fraction or ventricular mass.

It would be desirable that the development of future acquisition mechanisms of medical images accounts for this important dimensionality reduction that is performed to get from raw data to clinically useful information. Tackling this inefficiency seems to have much in common with the intuition of modern sampling methods, which attempt to reduce sampling constraints exploiting the assumption that the signal of interest lies in a subspace of smaller dimensionality than the image itself. In CS for MRI, the signal of interest is always presumed to be the magnetisation image, but if we modify our signal of interest to be the segmentation of the image or its registration to a different dataset, which are also of smaller dimensionality than the original image, we might be able to widen the spectrum of possibilities in MR acquisition and fuse reconstruction with analysis.

### 1.1.3 Reuse of past data

The past two decades have seen a rapid ascent of technology in modern societies and referring to some electronic devices as intelligent has become commonplace. This is so because the advance of hardware and software is at a point where they can be designed to learn patterns or rules of behaviour to direct actions autonomously, taking into consideration amounts of information in a way that is irreplaceable by human efforts. These developments, often referred to as machine

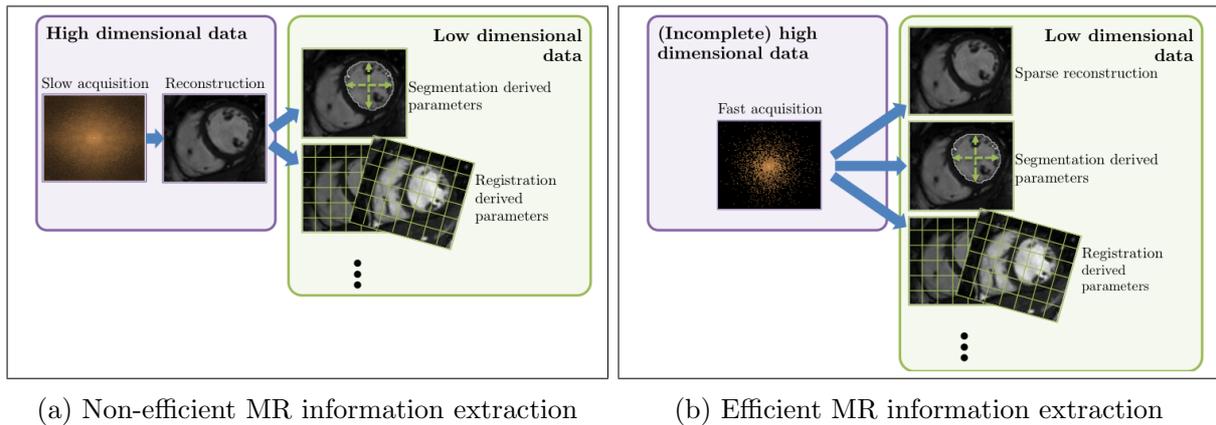


Figure 1.2: Conceptual differences in MR information extraction. Traditional MR analysis assumes a full image acquisition, which is later processed to obtain relevant clinical information (a). There is a significant dimensionality reduction between the fully sampled image and the extracted parameters that signals an inefficient handling of information. Modern approaches could ideally look for the low dimensional information of interest directly from incomplete raw data, minimising the discard of acquired information (b). Compressed sensing techniques would be a special case of this framework, where the signal of interest is the intensity image itself.

learning or artificial intelligence, have sprouted in numerous industries with noticeable benefits in augmenting their possibilities, but much of the potential in the medical imaging sector is still untapped.

Every day, thousands of medical images are produced and stored, capturing the state and progress of the human body and diseases across the entire population spectrum (see figure 1.3). This enormous load of information amasses studies and stories from the past which are valuable for the analysis of similar information yet to come. The convergence of machine learning and medical imaging is bound to bring great advances, and in this thesis we explore some of them. One of the objectives is to develop learning techniques at the service of medical imaging and analyse the benefit of training from past medical data. The outcome of the thesis will therefore not only be new methods directly applicable to current medical imaging, but also an illustration of how the community can progress towards the goals of intelligent imaging, where a present imaging experiment can draw information from images acquired in the past.

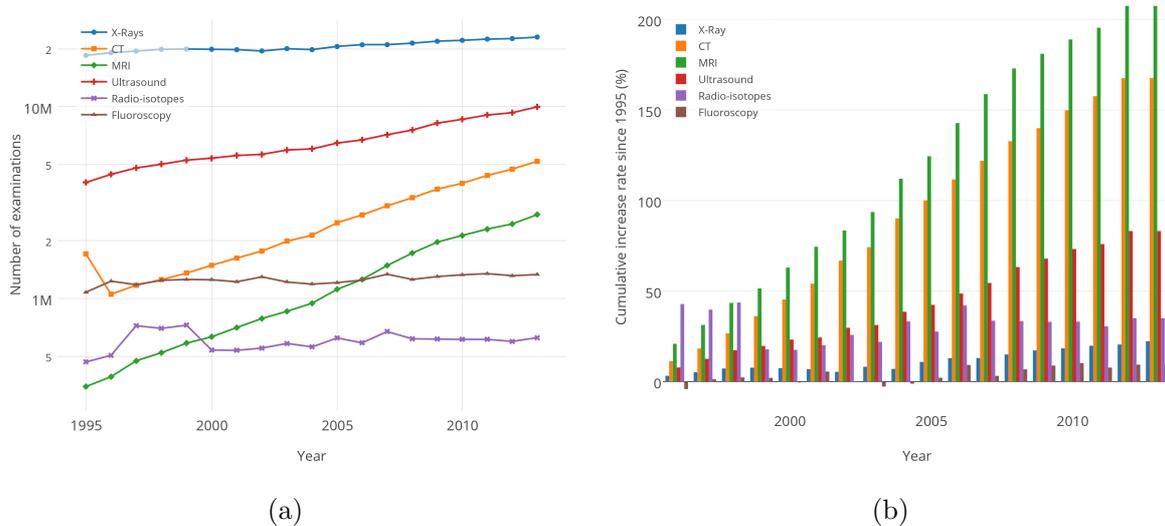


Figure 1.3: Number of imaging tests carried out by the NHS in England [101] (a). MRI is the imaging modality with the highest cumulative increase rate in the last 20 years with over a 200% increase (b).

## 1.2 Objectives and motivation

The objective of this thesis is to study the application of recent developments in dictionary learning (DL), a technology for training signal models, as a solution to the challenges introduced above. The term dictionary is used to denote a collection of elementary building blocks of data that are assumed to be capable of reconstructing natural signals through sparse linear combinations. One of their most attractive properties is their ability to adapt to a training data set, absorbing salient features to provide a model that is exclusively suitable for a specific type of data.

The use of DL techniques in MRI reconstruction and analysis has been growing in recent years. Adaptive modelling outperforms structured models by providing more accurate sparse representations of natural signals. This advantage can be exploited for the acceleration of MRI acquisition in the same setup as suggested by CS techniques and can be combined with other forms of modelling in order to perform reconstruction and analysis jointly from raw data. Furthermore, as long as the training set is generalisable, dictionaries provides a means to reuse imaging data by training on past medical images for an improved modelling of future scans.

The motivation for tackling these particular challenges can be divided into scientific, economic, and societal.

**Scientific** The reduction in MR acquisition time has been historically accompanied by the emergence of new imaging modalities. Accelerating dynamic MR acquisition would push boundaries in cardiac and fetal imaging, where currently spatial and temporal resolution need to be traded off against each other. Also, demonstrating the ability to extract analysis information directly from undersampled MR data would be a step forward in defining a new imaging paradigm, in which the information clinically relevant is directly targeted without the need to reconstruct high dimensional intensity images.

**Economic** The economical advantages in accelerating MR acquisition and reducing the processing necessary from acquisition to analysis are difficult to measure, but the current high cost of MR scans and the lengthening of patient waiting lists indicate that there could be considerable benefits. The increased demand for MR scans has been met in the past by a higher expenditure on machines [17, 98, 36]. Bringing into the equation the possibility of scanning patients in a fraction of the time currently necessary, would leverage much of the budgetary stress that MR scanning puts onto the healthcare sector.

**Societal** Current MR scanning procedures, and particularly for dynamic modalities, are long and uncomfortable. Producing high quality dynamic images of the heart usually requires that the patient performs repeated breath-holds to minimise breathing motion, and this might sometimes not even be possible depending on the condition of the patient. Moreover, long waiting lists degrade the quality of healthcare services and impact the quality of life for patients. Improving the overall functioning of healthcare services would be a valuable contribution of more intelligent MR acquisition techniques.

## 1.3 Contributions

The contributions of this thesis in the application of dictionary modelling for MRI are listed in the following subsections.

### **Dictionary learning for dynamic MR data reconstruction**

We introduce the use of spatio-temporal dictionaries for the acceleration of cardiac cine acquisition. This work is based on the foundations laid by Ravishankar *et al.* on the application of DL to 2D brain structural MRI in [117]. We explore its extension to handle dynamic data with the use of a spatio-temporal dictionary, jointly taking advantage of the high redundancy in time while ensuring spatial homogeneity. Additionally, we modify the modelling to include a temporal total variation (TV) penalty that can improve results in high acceleration scenarios. The results shown are simulated experiments using raw MR data and are therefore not directly applicable in practice, but demonstrate the capabilities of dictionary modelling for acquisition acceleration relative to a competing method that employs a non-adaptive sparsity model.

### **Parallel MRI reconstruction with overcomplete and adaptive dictionaries**

We demonstrate how the framework for MR acquisition acceleration using DL can be extended to handle reconstruction from parallel data. Parallel MR is today widely used as a hardware solution to accelerate acquisition and improve MR image quality. Innovative acquisition and reconstruction methods such as the DL framework introduced in this thesis must be accommodated to parallel coil technology to take advantage of their benefits and to be readily applicable to today's standard clinical procedures. We present a DL regularisation of an established parallel reconstruction method, which enables to control noise amplification and reduce aliasing artefacts. Results are demonstrated with retrospective undersampling of raw parallel data and examples of prospectively undersampled data.

## Application-driven MRI

The last contribution of the thesis targets efficient information extraction from MR images. The proposed concept of application-driven MRI encompasses techniques that can provide clinically useful information directly from raw data in a way that outperforms the traditional serial handling of information acquisition and analysis. Building on the fast DL-based reconstruction proposed, the model is extended such that MR images are jointly represented by a simpler labelling model, hence achieving image segmentation as a by-product of reconstruction. This model is chosen to be a Gaussian mixture model (GMM), which can be very effective at segmenting cardiac and brain images and flexibly incorporates additional constraints such as probabilistic priors in the form of atlases and Markov random field (MRF) regularisation.

## 1.4 Thesis overview

The thesis is structured in two main parts. The first part includes background information regarding the problems tackled and the state-of-the-art that forms the foundation for this thesis. The physical principles of MRI, the reasons for its speed limitations and previous attempts at tackling them are laid out in chapter 2. In chapter 3 we provide background information on the theory of CS and DL, and discuss some of the proposed methods which have successfully applied these novel sampling and modelling techniques to MRI acquisition, with a particular focus on cardiac cine. Chapter 4 introduces the DL method proposed for the acceleration of cardiac cine acquisition. Results are presented in a synthetic single coil scenario with retrospectively undersampled data, and compared against a method relying on a fixed-basis sparsity transform. In chapter 5 this framework is extended to the parallel coil scenario, showing results on raw data retrospectively and prospectively undersampled. The concept of application-driven MRI is discussed in chapter 6 and is demonstrated with results of joint reconstruction-segmentation on cardiac cine and brain MR images. To conclude, the thesis and its results are discussed and analysed in chapter 7, with a mention on potential future work avenues.

# Chapter 2

## MRI principles

### 2.1 Introduction

MRI is a non-invasive and non-ionising medical imaging modality largely popularised by an unmatched soft tissue contrast. Images obtained with MR are a reflection of the response of different tissues to a controlled stimulus. This is conceptually different from other imaging techniques that measure intrinsic material properties, such as computed tomography scans, which provide a measure of attenuation coefficients. MR images show a measure of the transverse magnetisation inside the body at a specific time. Transverse magnetisation is produced by applying a rotating magnetic field with rotational frequency matched to the resonant properties of certain atomic nuclei, and its behaviour across time is shaped by atomic properties that differ depending on tissues.

Generally, three stages can be clearly identified in the generation of MR images [19]: First, the magnetic pulses used to create a transverse magnetisation must be designed. Then, during the acquisition stage the signal emitted is sampled by a receiver coil in the scanner. The last stage is the reconstruction, where the signal captured is modelled and used to produce an image. The acquisition speed of MRI is limited by physical properties of the atoms manipulated

and hardware limitations, therefore restricting the use of this imaging tool due to the long scan times. This work focusses on the two latter stages, proposing alternative designs for the acquisition and reconstruction which allow faster and informative imaging.

In this chapter we introduce the nuclear magnetic resonance phenomenon that is at the core of MRI, and explain how it can be exploited to create images of the body. The problematic acquisition speed limits are also analysed and previous attempts at tackling them, which can be found in the literature, are summarised. The presentation of MRI given in this chapter follows the classical description found in [78].

## 2.2 Nuclear magnetic resonance

The essential physical phenomenon enabling MRI is nuclear magnetic resonance (NMR) [75], which refers to the exchange of energy between atoms and a magnetic field rotating at a resonant frequency. To introduce this concept we look at a sequence of three atomic phases composing the NMR experiment, which are spin polarisation, excitation and relaxation, and conclude with the Bloch equation, formally summarising these physical events.

### 2.2.1 Spin polarisation

The nuclei of atoms in our body with an odd number of protons or neutrons possess an angular momentum called spin with an associated microscopic magnetic field. One example of this kind of atom largely present in the body is hydrogen ( ${}^1H$ ). The backbone of an MR machine is a powerful superconducting magnet, capable of applying a high magnetic field which, by convention, is drawn along the  $z$  axis. Under the influence of this field, called  $B_0$  field, the magnetic moments of individual nuclei abandon a random orientation and adopt one out of a set of discrete values. For a spin  $\frac{1}{2}$  such as for  ${}^1H$  atoms, the orientation can be parallel or anti-parallel to the field, depending on whether they adopt a low or high energy state [75]. This

change of behaviour is represented in figures 2.1a and 2.1b.

The interaction of an ensemble of nuclei under a magnetic field results in a net effect of many individual spins acting to create a macroscopically detectable magnetisation vector. The magnetic vector is decomposed into a longitudinal component in the  $z$  axis, and a transverse component lying on the  $x$ - $y$  plane. The analysis of an ensemble of nuclei reveals that at room temperature a small majority of atoms aligns with the field, creating a net longitudinal magnetisation  $\mathbf{M} = M_z^0$  aligned with  $B_0$ . Individual transverse magnetisation of atoms however have random phases, and therefore do not create a net transverse magnetisation ( $\mathbf{M}_{xy} = 0$ ).

Nuclei have a resonant frequency proportional to the surrounding magnetic field. The  $B_0$  field causes similar nuclei to precess at the same frequency, called the Larmor frequency,

$$\nu = \frac{\gamma}{2\pi} B_0, \quad (2.1)$$

where  $\gamma$  is the gyromagnetic ratio and is nucleus dependent. The  $^1H$  atomic nuclei has  $\gamma = 2.68 \times 10^8$  rad/s/T, and so in a typical MR system of 1.5 T precesses at roughly 64 MHz [61]. This system state of a net longitudinal magnetisation with aligned spins is called equilibrium state, and sets the scene for the NMR experiment.

### 2.2.2 Excitation

The Larmor frequency is the resonance frequency at which energy from an external magnetic induction field can be absorbed by the nuclear spin system. In the MR literature, this field is produced by a radio-frequency (RF) pulse  $\mathbf{B}_1$  perpendicular to the static  $B_0$ . If enough energy is deposited in the system, the spins will come into phase and some low energy spins will jump to a high energy state, macroscopically tilting the net magnetisation onto the  $x$ - $y$  plane [78]. As a result, longitudinal magnetisation decreases and the net magnetisation now includes a net transverse component spinning at the Larmor frequency as depicted in figure 2.1c.

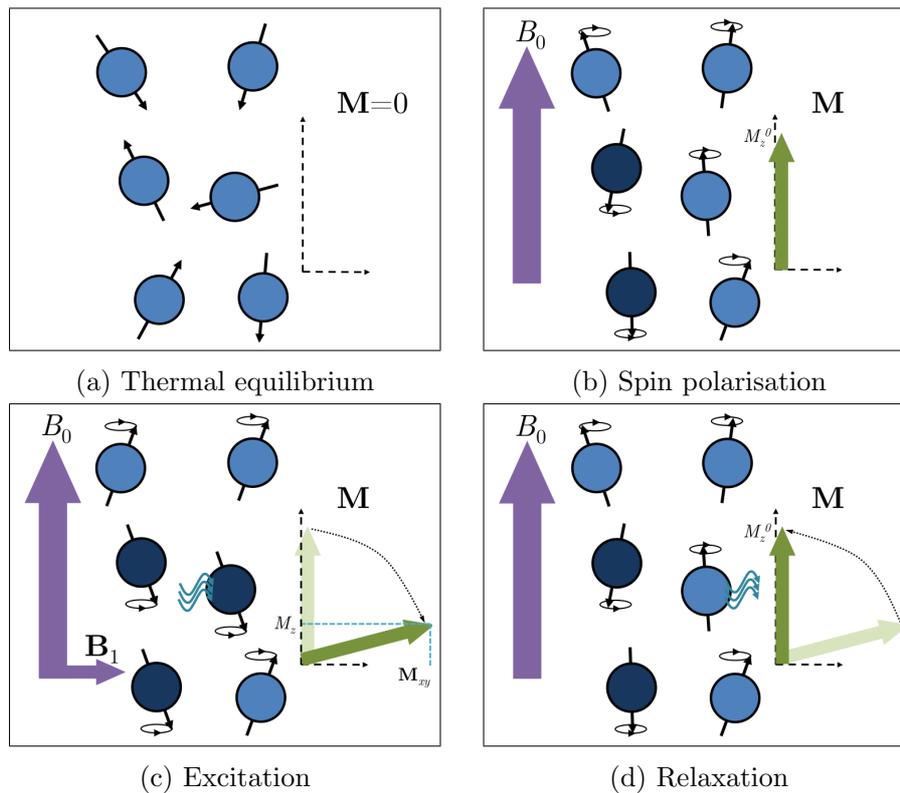


Figure 2.1: Different atomic nuclei phases in the NMR experiment. Low-energy state nuclei are represented in light blue and high-energy state nuclei are dark blue. After spin polarisation with the  $B_0$  field, the system of spins can be excited resulting in the transverse magnetisation  $\mathbf{M}_{xy}$ . The bulk magnetisation relaxes at the end of the excitation, returning a state where it again aligns with the  $B_0$  field.

### 2.2.3 Relaxation

After excitation, net magnetisation returns to the equilibrium state through a process called relaxation. In this time interval, RF energy absorbed during excitation by some spins to adopt a high energy state is liberated as they return to a low energy state, causing a gradual decay of transverse magnetisation and the recovery of the original longitudinal magnetisation as illustrated in figure 2.1d. The liberated energy constitutes the emitted signal, which can be sampled and used for imaging [75].

Two effects govern the transition from the excited state to equilibrium over a period of time  $\tau$  [78]. First, longitudinal relaxation is caused by energy exchange between the spins and the surrounding lattice, falling back to thermal equilibrium. It is described by an exponential curve

characterised by the spin-lattice relaxation time  $T_1$ :

$$M_z(t + \tau) = M_z^0 - (M_z^0 - M_z(t)) e^{-\frac{\tau}{T_1}}. \quad (2.2)$$

The term  $T_1$  is a tissue specific constant which is usually longer at higher field strengths.

Similarly, transverse relaxation results from spins dephasing in the absence of the RF signal:

$$\mathbf{M}_{xy}(t + \tau) = \mathbf{M}_{xy}(t) e^{-\frac{\tau}{T_2}}. \quad (2.3)$$

This decay is also governed by an exponential curve with the spin-spin relaxation time  $T_2$  which is also tissue dependent. It is crucial that  $T_1$  and  $T_2$  are tissue specific, given that this makes different tissues react differently to the same stimulus, becoming a mechanism not only to resolve them but to create images with different contrast patterns.

## 2.2.4 Bloch equation

The temporal behaviour of the net magnetisation of a nuclear spin system was formalised in 1946 by Felix Bloch [15] in what is known as the Bloch equation:

$$\frac{d\mathbf{M}(t)}{dt} = (\mathbf{M}(t) \times \gamma \mathbf{B}_1(t)) - \frac{M_z(t) - M_z^0}{T_1} \mathbf{e}_z - \frac{\mathbf{M}_{xy}(t)}{T_2}. \quad (2.4)$$

It synthesises the excitation reaction of a net magnetisation  $\mathbf{M}$  initially in equilibrium to an external electromagnetic radiation  $\mathbf{B}_1(t)$  and the subsequent relaxation phase governed by tissue specific constants  $T_1$  and  $T_2$ .

## 2.3 Magnetic resonance imaging

MRI refers to the technique exploiting the NMR signal produced by the transverse magnetisation of atomic nuclei to generate images. The emitted signal is a rotating entity dependent on space and time, and can therefore be described by the complex-valued variable  $m(\mathbf{r}, t) = m(\mathbf{r})e^{i\phi(\mathbf{r}, t)}$ , where  $\mathbf{r}$  refers to  $(x, y, z)$  space and  $m(\mathbf{r})$  is the signal of interest [78]. Crucial components for MRI are encoding gradients, which enable the spatial characterisation of the signal received, and the notion of k-space, relating the sampled signal with the magnetisation image.

### 2.3.1 Encoding gradients

The RF pulse causes all nuclei in the body precessing at the resonant frequency to emit a signal, regardless of their spatial location within the body. A MRI scanner is equipped with encoding gradient coils, which are capable of generating constant gradients that are superposed to the  $B_0$  field creating spatially linear varying precessing frequencies [145]. Following from equation (2.1), a gradient  $\mathbf{G}(t)$  will cause a frequency deviation from the Larmor frequency equivalent to

$$\Delta f(\mathbf{r}, t) = \frac{\gamma}{2\pi} \mathbf{G}(t) \cdot \mathbf{r}. \quad (2.5)$$

Encoding gradients therefore modify the frequency of rotating nuclei to be directly proportional to their location within the body as is illustrated in figure 2.2. There are generally two mechanisms for spatial localisation using encoding gradients: selective excitation and spatial encoding [78].

#### Selective excitation

Applying a gradient in a single direction, for instance in the  $z$  direction, will induce the spatially dependent precession  $f(z) = \frac{\gamma}{2\pi} (|B_0| + G_z z)$ , creating a linear variation of spin frequencies along

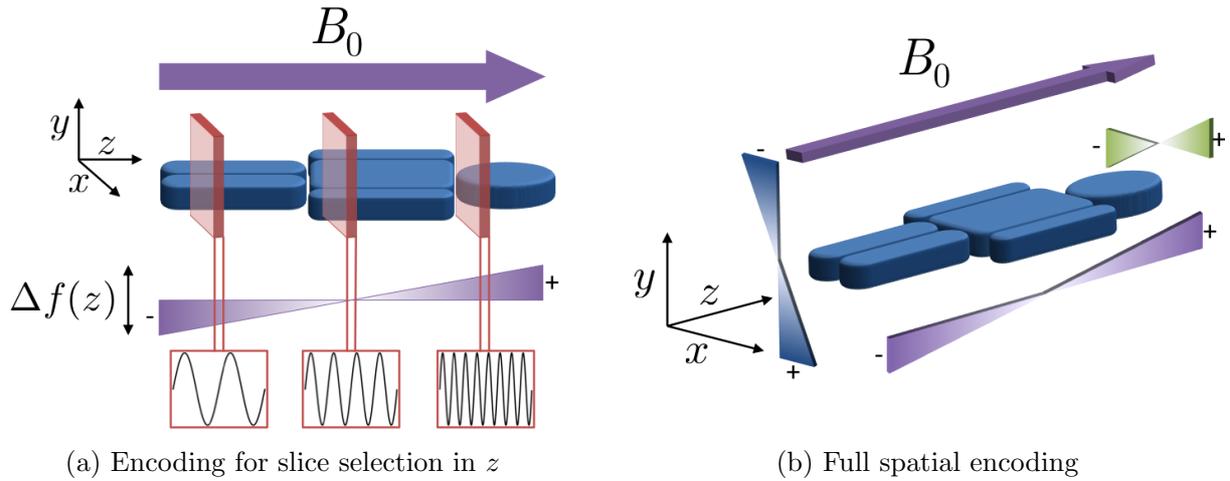


Figure 2.2: Encoding gradients for spatial localisation. A slice selection gradient can create slabs of signal with different shifts from the Larmor frequency (a). This principle is applied three dimensions to fully encode the signal in space.

the body. This set-up provides the ability to excite single  $x$ - $y$  slabs of nuclei sharing a similar frequency  $f(z)$ . The thickness of the slab is controlled by matching the properties of the RF pulse frequency content and a gradient of a specified amplitude to control the spatial extent for which the nuclei have Larmor frequencies that are matched by the pulse. This forms the first step for 2D imaging and is illustrated in figure 2.2a.

### Spatial encoding

Immediately after an RF pulse, the excited spin system in the  $x$ - $y$  plane is phase coherent and rotates at the Larmor frequency [78]. It is possible to determine the difference in spin phase caused by an additional gradient, for instance in the  $y$  direction, by integrating the frequency deviation over time, such that

$$\Delta\phi(y, \tau) = 2\pi \int_0^\tau \Delta f(y, t) dt = 2\pi \int_0^\tau \frac{\gamma}{2\pi} G_y(t) y dt = 2\pi k_y(\tau) y. \quad (2.6)$$

Here we have used  $k_y(\tau) = \int_0^\tau \frac{\gamma}{2\pi} G_y(t) dt$ . A simple encoding protocol will use a short gradient pulse duration  $\tau$  to create groups of atoms in the  $y$  direction sharing the same phase. This is the *phase encoding* step and is represented in figure 2.3b [78].

Moreover, it is clear how equation (2.5) can be used again in the remaining dimension  $x$  to create frequencies that are spatially varying in  $x$ . The use of this gradient to resolve the last encoding dimension is referred to as the *frequency encoding* step, and contributes in turn with the phase term

$$\Delta\phi(x, t) = 2\pi k_x(t)x, \quad (2.7)$$

where  $k_x(t) = \int_0^t \frac{\gamma}{2\pi} G_x(\tau) d\tau$  [78]. This stage is illustrated in figure 2.3c. At this point, each spatial location in the  $x$ - $y$  plane is uniquely described by its phase.

The phase component of the spatially encoded signal  $m(\mathbf{r}, t)$  can be included in the description of the emitted signal, which, ignoring the carrier Larmor frequency, becomes

$$m(\mathbf{r}, t) = m(\mathbf{r})e^{i\Delta\phi(\mathbf{r}, t)} = m(\mathbf{r})e^{i2\pi\mathbf{k}(t)\cdot\mathbf{r}}. \quad (2.8)$$

Here we have used  $\Delta\phi(\mathbf{r}, t) = 2\pi\mathbf{k}(t)\cdot\mathbf{r}$  to generalise encoding in the  $y$  and  $x$  dimensions of equations (2.6) and (2.7).

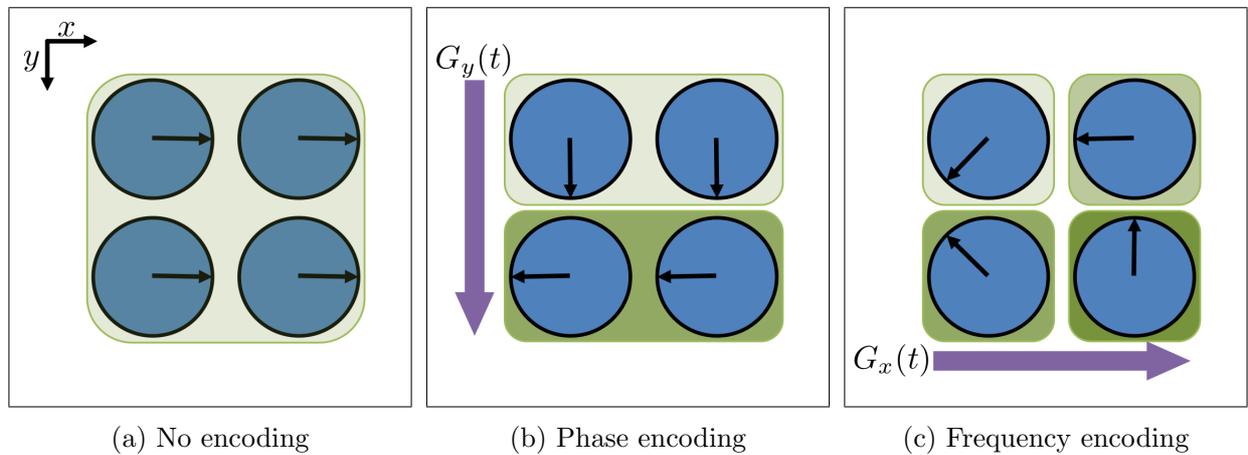


Figure 2.3: Spatial encoding in a 2D slice through phase and frequency encoding. Without gradient encoding, the signal received is a contribution of identically defined nuclei (a). The joint application of a phase (b) and a frequency (c) encoding allows to uniquely describe each location with different phases.

### 2.3.2 NMR signal measurement

The precession of  $m(\mathbf{r}, t)$  creates an oscillating magnetic flux that induces a voltage in a receiver coil perpendicular to the  $B_0$  field and tuned to the Larmor frequency. Samples of this voltage constitute the signal used for imaging. The transverse magnetisation can reveal different tissue properties depending on how the excitation pulse and the measurement are coordinated. Defining long time intervals between RF pulses and reading measurements immediately before relaxation will reveal the proton density of the tissue (PD weighting), whereas allowing for an interval of time between the end of the pulse and the sampling can reveal T2 relaxation properties of the tissue (T2 weighting) [19]. Examples of this flexibility in image contrast are shown in figure 2.4 on a simulated brain scan from [95].

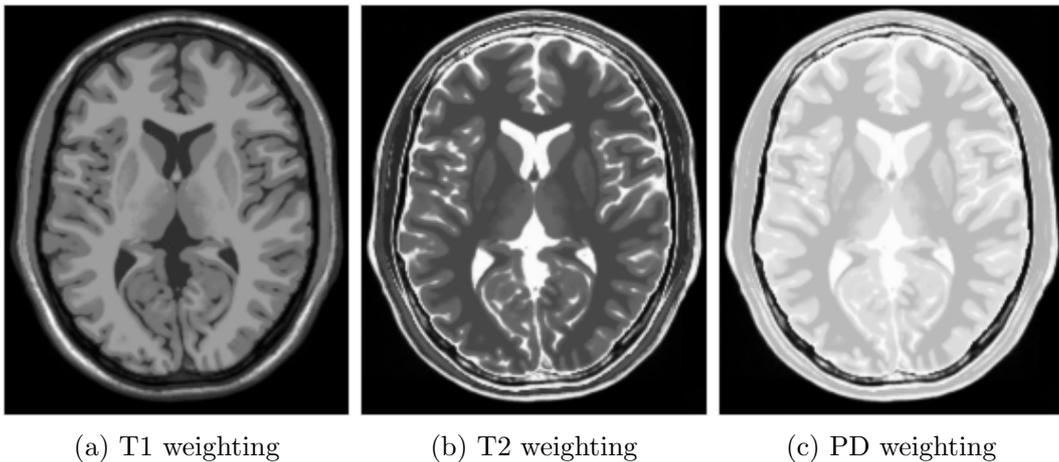


Figure 2.4: Simulated brain image [95] showing different contrast patterns. This is an example of the great contrast flexibility enabled by the physical principles underlying MRI.

The signal received is the aggregate response from nuclei at all spatial locations which have been subject to excitation. Usually, multiple receiver coils are used in a parallel arrangement, where the signal received is also dependent on the geometry and location of the coil [112, 59]. The magnetisation at coil  $c$  is therefore also multiplied by a sensitivity function  $S_c$ , yielding

$$\hat{m}_c(t) = \int_{-\infty}^{+\infty} S_c(\mathbf{r})m(\mathbf{r}, t)d\mathbf{r}. \quad (2.9)$$

We will assume for now the use of a single body coil to simplify the description before explaining

the use of parallel imaging in section 2.4.1. The received signal therefore simplifies to [78]

$$\hat{m}(t) = \int_{-\infty}^{+\infty} m(\mathbf{r}, t) d\mathbf{r} = \int_{-\infty}^{+\infty} m(\mathbf{r}) e^{-i2\pi\mathbf{k}(t)\cdot\mathbf{r}} d\mathbf{r}. \quad (2.10)$$

Note that this expression does not take into account factors such as the decay of transverse magnetisation and non-idealities like  $B_0$  inhomogeneity or changes of the object in time. Nevertheless, it provides a suitable model for the received signal and will be used throughout this work.

### 2.3.3 K-space

The direct mapping created between precessing frequencies and space through encoding gradients is identical to a Fourier relationship. Samples can therefore be interpreted as acquired in the spatial-frequency domain, also called k-space, where each sample is a sum of magnetisation from atoms in the entire volume weighted by the Fourier kernel  $e^{-i2\pi\mathbf{k}(t)\cdot\mathbf{r}}$  [78].

The coordinates of the sample in k-space are given by  $\mathbf{k}(t)$  and are directly controlled through the gradients  $\mathbf{G}(t)$  [19]. An example is given in figure 2.5. Applying a positive  $G_y(t)$  gradient for a short period of time  $\tau$  will produce a positive value  $k_y(\tau)$ , which is interpreted as a departure from the  $\mathbf{k}(t) = \mathbf{0}$  coordinate along the  $y$  axis. The use of both gradients  $G_y(t)$  and  $G_x(t)$  allows traversing the  $x$ - $y$  plane of k-space. Given that gradient pulse durations directly relate to k-space locations, the received signal can be rewritten as

$$\hat{m}(\mathbf{k}) = \int_{-\infty}^{+\infty} m(\mathbf{r}) e^{-i2\pi\mathbf{k}\cdot\mathbf{r}} d\mathbf{r}, \quad (2.11)$$

or if we consider the practical case of 2D discrete time samples we write

$$\hat{m}(k_x, k_y) = \sum_{x=-N_x/2}^{N_x/2} \sum_{y=-N_y/2}^{N_y/2} m(x, y) e^{-i2\pi(k_x x + k_y y)}. \quad (2.12)$$

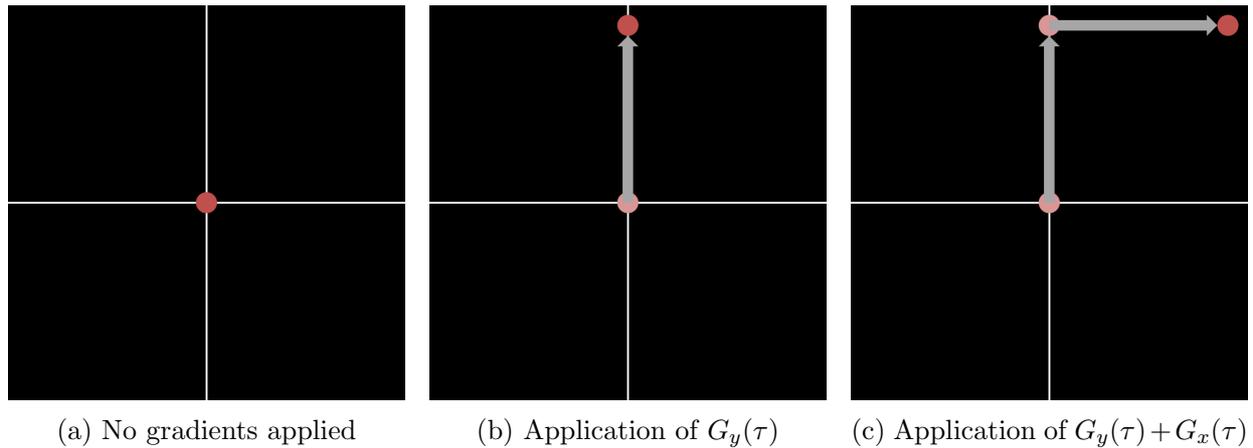


Figure 2.5: Traversing k-space with the application of gradient fields.

For the sake of simplicity we will adopt the following matrix notation using  $\mathbf{F}$  to denote a 2D discrete Fourier transform (DFT):

$$\hat{\mathbf{m}} = \mathbf{F}\mathbf{m}. \quad (2.13)$$

An appropriate coordination of gradients will lead to a scan protocol traversing locations of k-space and acquiring different samples. This protocol, usually called a trajectory, has an important impact on the appearance of the reconstructed image [19]. The design of a protocol is concisely represented in sequence diagrams such as the one shown in figure 2.6. The simplest example is the Cartesian trajectory, which traverses a rectangular k-space acquiring equidistant samples as shown in figure 2.7a. Cartesian sampling poses a very simple reconstruction stage, which is given by a DFT, but has the downside of requiring many excitation pulses to traverse a full k-space. Other possibilities exist, such as radial or spiral trajectories (see figures 2.7b and 2.7c), which require a gridding stage prior to DFT reconstruction.

### 2.3.4 FOV and resolution

From the classical description of the DFT relating time and frequency domains we know there is an inherent compromise between sampling frequency in time domain and spectrum replicas in frequency domain. Assuming Cartesian sampling in the MRI setup, this imposes the following

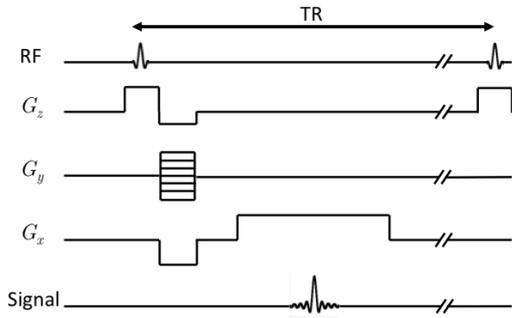


Figure 2.6: Pulse sequence diagram representing the coordination of gradient pulses. Slice selection in the  $z$  direction is applied with the RF excitation. Gradients to encode 2D spatial information are then applied sequentially, with phase encoding in the  $y$  dimension and frequency encoding and read-out in the  $x$  dimension.

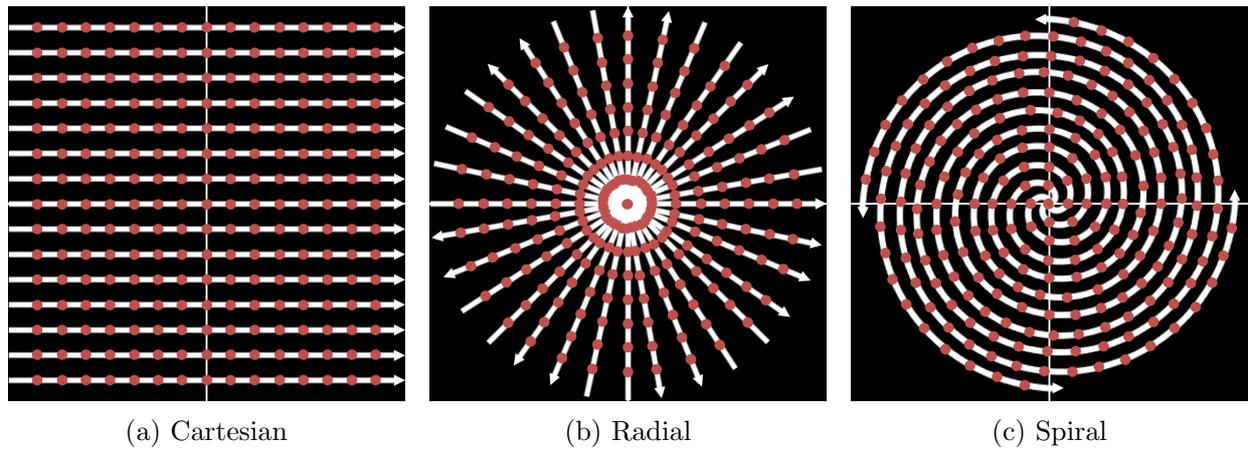


Figure 2.7: K-space trajectories

relationship between k-space sampling and image FOV [78]:

$$\text{FOV}_x = \frac{1}{\Delta k_x}, \quad \text{FOV}_y = \frac{1}{\Delta k_y}. \quad (2.14)$$

Choosing a sampling frequency smaller than the Nyquist rate  $f_{\text{Nyq}} = 2W$ , where  $W$  is the spatial support of the object imaged, will result in aliasing artefacts.

Similarly, finite sampling in Fourier domain implies a compromise between the sampled k-space support and resolution in image domain. This relationship is given as [78]

$$\Delta x = \frac{1}{\Delta k_x N_x}, \quad \Delta y = \frac{1}{\Delta k_y N_y}, \quad (2.15)$$

meaning the larger the k-space support sampled, the finer the resolution in image domain.

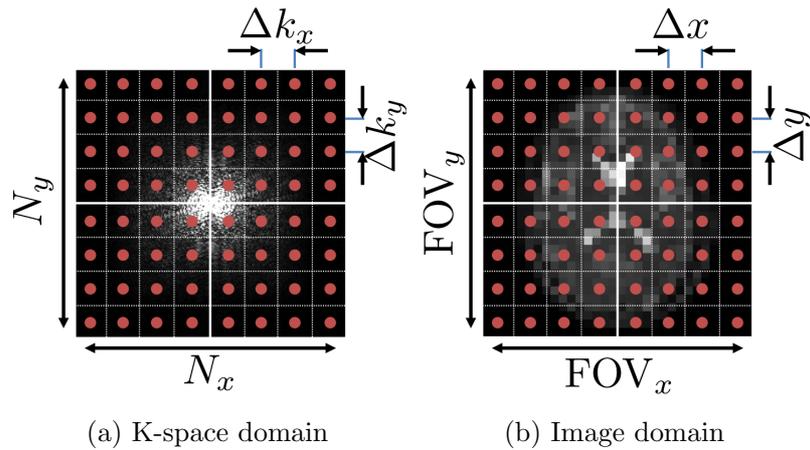


Figure 2.8: FOV and resolution relationships between k-space (a) and image domain (b).

## 2.4 Acquisition acceleration

The fact that k-space samples must be collected sequentially in MRI makes it a very time-consuming process. The acceleration of MRI acquisition is desirable for a number of reasons. The most direct consequence of faster scans is an increased comfort for patients. Ideally the patient needs to remain static for the duration of the scan, but patients who are severely ill or too young may not tolerate long scanning procedures. Indirectly, this translates for hospitals and clinics into a higher throughput and shorter waiting lists, therefore reducing the cost of MR scanning. Faster scans can also have an effect on the quality of images, especially for dynamic MR, as they can reduce the corruption introduced by motion artefacts.

In the most simple scan protocols the speed limit is dictated by the number of RF pulse excitations required and the time interval needed between pulses, known as repetition time TR. A simple 3D gradient echo Cartesian acquisition takes as long as

$$T_{\text{Acq}} = \text{TR} \times N_z N_y, \quad (2.16)$$

where  $N_z$  and  $N_y$  are the number of phase encodes and it is implied that readout is being performed along  $k_x$ . A full set of frequency encodes is acquired within the same RF pulse excitation so  $N_x$  can be ignored in calculating  $T_{\text{Acq}}$ . The arbitrary reduction of TR is not

viable given that it is set by image contrast requirements, readout properties, signal-to-noise ratio and specific absorption rate limits.

Many solutions have been proposed for MRI acquisition acceleration. One possibility is to look for pulse sequence designs that can traverse k-space with fewer RF excitations, such as spiral trajectories shown in figure 2.7c [3] or other existing examples [110, 102, 79]. Other trajectories such as radial sampling (see figure 2.7b) naturally favour denser sampling for low frequencies and sparser at higher frequencies, and have also been proposed for acceleration [114]. Moreover, carefully designing a pulse sequence can allow to collect multiple separately encoded readouts per RF pulse as is proposed by fast spin echo strategies. The extreme example is given by echo-planar imaging [93], which traverses a full k-space as a continuous trajectory. The quality of multi-echo sequences is nevertheless affected by relaxation effects and a variation of contrast given that phase encodes are acquired at different times after the excitation.

A different perspective on acquisition acceleration is taken by methods that assume a particular trajectory but undersample k-space by skipping the acquisition of some samples [86]. Given a binary mask  $\mathbf{M} \in \mathbb{R}^{N \times N}$ ,  $N = N_x N_y$ , with entries  $[\mathbf{M}]_{n,n} = 1$  if k-space sample  $n$  has been acquired, an undersampled acquisition is written as  $\hat{\mathbf{m}} = \mathbf{M}\mathbf{F}\mathbf{m}$ . The undersampling of a k-space intended to be sampled at the Nyquist rate provokes aliasing in image domain, which can take different forms as shown in figure 2.9. Sampling only low frequencies results in worse image resolution, regular undersampling creates coherent aliasing and random sampling leads to random aliasing. Physically this is interpreted as the energy of the image being folded onto itself in different locations.

Throughout this thesis we will refer to an aliasing-free acquisition satisfying the Nyquist criterion as a fully sampled acquisition. The problem posed for reconstruction is then an underdetermined system of linear equations, which is ill-posed unless sampling redundancy is exploited. Sampling redundancy refers to the fact that some samples may be dispensable without meaningful loss of information. These techniques will be of most interest in this work.

Early approaches proposed to create sampling redundancy through additional hardware. In

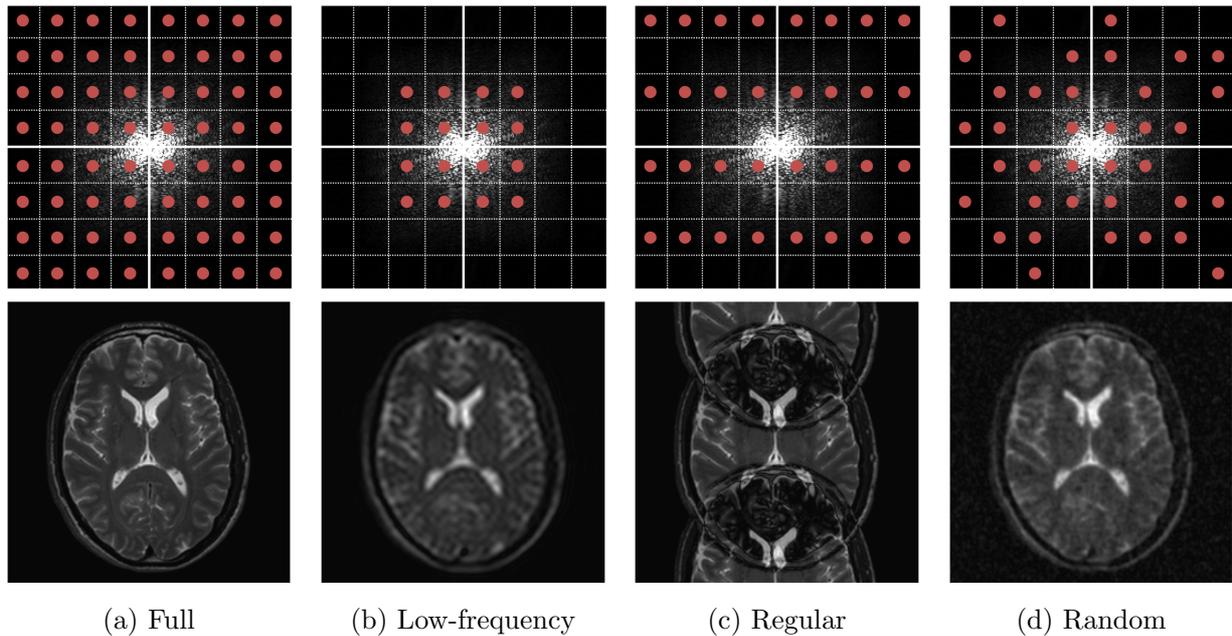


Figure 2.9: Undersampling strategies in k-space (top) and the aliasing they produce in image domain (bottom).

particular, the use of multiple receiver coils in a parallel imaging set-up linearly increases the number of samples for a particular image, which theoretically allows maintaining a fixed amount of information even after k-space undersampling [129, 112, 59, 136]. The last decade has seen the rise of methods that exploit inherent image redundancy, whereby some samples are presumed dispensable because they can be derived by assuming known properties of the image, such as compressibility [86, 55].

### 2.4.1 Explicit redundancy: Parallel imaging

Using multiple coils for the acquisition of MRI data is an effective method of reducing scan time. Each coil samples the signal emitted by the same object at the same time but filtered through different sensitivity patterns as depicted in figure 2.10. This was formally introduced in equation (2.9), which can be rewritten in discrete time and matrix form as

$$\hat{\mathbf{m}}_c = \mathbf{F}\mathbf{S}_c\mathbf{m}. \quad (2.17)$$

Assuming a total of  $N_c$  coils, this expression can be extended to account for all the data acquired by writing

$$\tilde{\mathbf{m}} = \tilde{\mathbf{F}}\tilde{\mathbf{S}}\mathbf{m}, \quad (2.18)$$

where we have used a tilde  $\tilde{\cdot}$  to refer to variables concatenating parallel data or operators. The term  $\tilde{\mathbf{m}} \in \mathbb{C}^{NN_c}$  is the vertical concatenation of k-space data from all coils,  $\tilde{\mathbf{F}}$  is a block diagonal concatenation of 2D Fourier transforms  $\mathbf{F}$ , and  $\tilde{\mathbf{S}}$  denotes a vertical concatenation of sensitivity matrices.

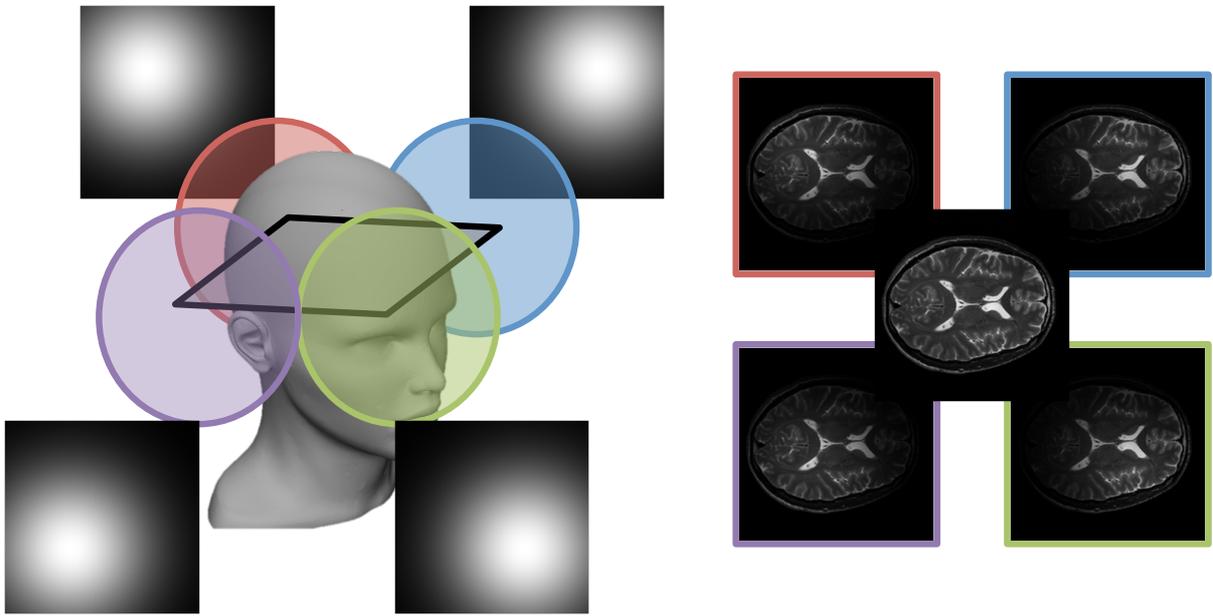


Figure 2.10: Parallel MRI acquisition using 4 coils. The data acquired by each coil is filtered through its own sensitivity pattern.

The expression in equation (2.18) is an overdetermined system of linear equations, meaning that in theory it is possible to design  $\tilde{\mathbf{S}}$  such that we can derive  $\mathbf{m}$  from as few as  $\frac{N}{N_c}$  samples of  $\tilde{\mathbf{m}}$ . For  $M$  samples acquired, we define an acceleration rate  $R = \frac{N}{M}$ . Using  $\tilde{\mathbf{M}}$  to denote the block diagonal concatenation of the undersampling mask  $\mathbf{M}$ , the accelerated scan is modelled as

$$\tilde{\mathbf{m}} = \tilde{\mathbf{M}}\tilde{\mathbf{F}}\tilde{\mathbf{S}}\mathbf{m}. \quad (2.19)$$

Different methods have been emerged for the reconstruction of  $\mathbf{m}$  given  $\tilde{\mathbf{m}}$ , which propose to

look for a solution in either the image or the k-space domain [129, 112, 59, 136]. As long as  $R \leq N_c$ , perfect reconstruction could in theory be achieved for orthogonal sensitivity patterns. However, the geometrical distribution of coils around a body prevents perfect decorrelation and in practice the acceleration rate is limited by numerical instabilities occurring when  $R$  approaches the theoretical maximum  $N_c$ . In chapter 5 we provide more details on parallel imaging limitations and solutions.

### 2.4.2 Implicit redundancy: Image correlation

An alternative route for acquisition acceleration in MRI is the use of inherent data redundancy. The more information about the object is known or assumed *a priori*, the fewer samples will be needed for its reconstruction. The diversity of methods which have been proposed incorporating prior knowledge into a model for the sampling process is very large, and their success depends on the validity and flexibility of the assumptions made, as well as on the accuracy with which these assumptions can be introduced into the reconstruction process.

A simple assumption about the magnetisation image, given that it is a reflection of spin density, is that it is expected to be a real-valued image, and therefore its Fourier representation should be conjugate symmetrical with respect to the k-space origin. The number of phase encodes can therefore potentially be reduced by almost half without loss of information. This is what partial Fourier sampling proposes [94, 35, 103]. MR images are nonetheless always complex-valued due to phase distortion effects from off-resonance, eddy currents and motion, amongst other causes [19]. It is possible to estimate and reproduce slowly varying phase patterns by fully acquiring additional phase encodes in central k-space, but even so the acceleration rates are generally limited to about  $R = 1.8$  [96].

Spatio-temporal correlations in dynamic MR images have also been the subject of much interest in trying to reduce the number of samples required. The changes between successive frames in a dynamic acquisition are usually confined to a small region of the image and are frequently slowly

varying. As we saw in section 2.3.4, FOV and sampling density are inversely proportional. The reduced FOV method [63] draws on this relationship to fully acquire a reference image that is later updated with a sparser sampling grid in order to capture changes occurring only in the (assumed) smaller spatial region. Such a strong assumption on the data degrades image quality when changes are not only present in the *a priori* defined reduced FOV, and therefore limits acceleration to  $R = 2$ .

More strategies have been proposed for the sequential update of a dynamic k-space. The keyhole method [142, 68] fully acquires a reference image and suggests the update of only the centre of k-space, which is where most of the signal's energy resides. A more elaborated variation is proposed by RIGR [77], where high frequency information is not directly obtained from the fully sampled reference scan, but instead it is fitted to a set of basis functions derived from this reference.

To try to cover the acquisition of full k-space while reducing scan time methods such as sliding window [37, 151] have emerged, where individual samples are shared by subsequent frames. A  $R = 2$  acceleration can be achieved by acquiring even and odd phase encodes separately for pairs of frames and then extrapolating missing samples from neighbouring frames. Reconstructions using these methods at higher acceleration rates are however prone to blurring effects.

Other approaches have attempted to target the suppression of aliasing, which is the direct consequence in image domain of k-space undersampling. The x-f space, which can be observed by taking a Fourier transform of the data along time, is generally a sparse representation of the data, accounting for the large temporal redundancy in dynamic MRI. Undersampling k-space with a regular lattice will create replicas of the spectrum at regular intervals. If the x-f space is sufficiently sparse, it may be possible to simply filter out unwanted aliasing. This method is known as UNFOLD [88]. A similar approach is taken by k-t BLAST and k-t SENSE [136], which includes an additional training stage where the support of the x-f spectrum is estimated and used for more accurate filtering.

A family of algorithms which has brought a lot of attention in the last decade has been meth-

ods based on CS. The fundamental observation used for CS reconstructions is that images or sequences of images can be sparsely represented in the image or a transform domain. This statement means that missing k-space data can be recovered with the *a priori* assumption that the reconstructed image is compressible. Its mechanism along with its application to MRI sampling is explained in more detail in the next chapter.

## 2.5 Conclusion

In this chapter we have reviewed the basic principles behind MRI acquisition and reconstruction. The MR experiment provides a means to interact with nuclei of the body using a superconducting magnet, and to generate a magnetisation signal dependent on nuclear magnetic properties that can be measured. It is then possible to produce images from this signal with the use of encoding gradients, which can resolve the magnetisation image through space by sequentially acquiring a number of phase and frequency encodes of a selected slice. The data resulting from this data is interpreted as the Fourier transform of the image of interest and can therefore be trivially inverted. This sequential collection of data is however inherently slow, and various techniques have been devised to reconstruct images from fewer samples exploiting sampling redundancy, which is created through parallel imaging or can be implicitly found in the image. In the following chapter we will introduce the mathematical foundations of CS, which allows the exploitation of inherent image redundancy, with a particular interest in novel adaptive methods, and discuss its application to MRI acquisition and reconstruction.

# Chapter 3

## Adaptive compressed sensing

### 3.1 Introduction

The quality of digital sensors is ever improving and is allowing us to capture larger amounts information each year. In 2007, the amount of digital information produced surpassed the storage capabilities in the world, producing a paradigm shift in which data is ephemeral, something to be immediately used or lost (see figure 3.1). In 2012 the estimated digital data in the cloud was 462 exabytes ( $40 \times 10^{18}$  bytes) and in 2020 the world's digital data is expected to grow to 40 zettabytes ( $40 \times 10^{21}$  bytes) [125]. MRI and medical imaging in general contributes to this overwhelming increase of digital information.

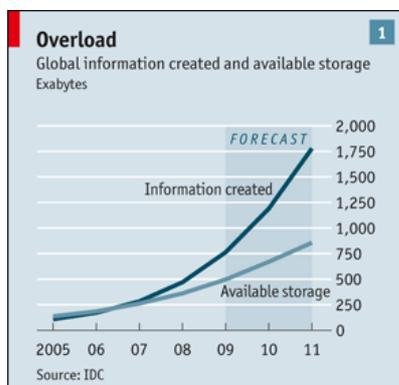


Figure 3.1: Estimation of the amount of data produced and stored [132]. In 2007, the amount of information created surpassed the storage capabilities.

The digitisation of natural information is produced through sampling. Analog-to-digital sam-

pling has been classically governed by the theorem proposed by Shannon in 1949 [123], which states that a continuous time signal of bandwidth  $B$  can be perfectly recovered after regular sampling provided the sampling rate is at least  $2B$ . This theorem, although not formally described, was implied by earlier results from Nyquist [104] that pointed to this critical sampling frequency, hence naming it the Nyquist rate. Reconstruction from samples obeying the Nyquist rate is attractive as it is linearly performed with sinc function interpolation, and although it is a sufficient condition for perfect reconstruction, it is not necessary.

Any signal reconstruction problem is an ill-posed problem, and its unique solution requires tailoring a given number of samples to a predefined signal model. The signal model is interpreted as prior knowledge about the signal and has a direct impact on the number of samples necessary for reconstruction, given that the more we know about the signal *a priori*, the less we need to discover about it from sampling. In Shannon's theorem the prior knowledge is a bandwidth limit, but in many situations it is possible to incorporate additional knowledge about the signal that reduce the sampling requirements imposed by Nyquist.

The term compressed sensing (CS) has been used to define a non-linear sampling theorem concentrating on the sampling of signals which are sparse or compressible [42, 28]. This condition encompasses many natural signals such as speech or images as has been demonstrated by data compression. Although CS is not the first to have proposed sub-Nyquist sampling strategies, it is one that is naturally suited for MRI given that the requirements for signal recovery can be satisfied without changing the physical scanner principles [84].

## 3.2 The need for modern sampling

There are many reasons why exploring newer forms of sampling is sensible. We summarise three of them in this section responding to the necessity for efficiency, cohesion and pragmatism in modern sampling.

**Efficiency** Most of the data acquired through digital sensors, such as speech or images, undergoes a compression stage after acquisition. This is a step needed to make the most out of the storage capabilities we have at hand and does not alter perceptually the quality of the data. From an efficiency perspective, it is not the best use of our resources, given that most of the information acquired is disposed and never used. It would therefore be interesting if we could estimate how compressible a signal is before acquisition and adapt the sampling stage such that information is captured in an already compressed format [28].

**Cohesion** The Shannon sampling theorem proposes a linear process both for sampling and reconstruction whereas compression is a highly non-linear process. Even though as separate stages sampling, compression and reconstruction can be thought of as optimal data processing modules, there is no reason to think that they are jointly optimal in providing compressed samples [7].

**Pragmatism** In certain cases the need for alternative forms of sampling is justified by the technical problems encountered in classical sampling. High-bandwidth signals impose demanding challenges on the acquisition, storage and processing hardware when sampled at the Nyquist rate [7, 133], and is a good example where traditional sampling is problematic. In the particular case of MRI, sampling time is proportional to the number of samples acquired, and therefore breaking the Nyquist rate by acquiring less samples enables faster scanning [84].

### 3.3 Compressed sensing

The Nyquist sampling of a 1D continuous time or space signal  $f(t)$  over a support of length  $NT$  is expressed as

$$x_i = f(nT), \quad n = 1, \dots, N \quad (3.1)$$

where  $1/T$  is the Nyquist rate. To analyse CS we replace samples with a set of linear measurements

$$\mathbf{y} = \Phi \mathbf{x}, \quad (3.2)$$

where  $\Phi \in \mathbb{C}^{M \times N}$  is a sensing operator containing sensing functions as rows. These functions are sometimes not up for design but are given by the physics of the acquisition mechanism [28], just as Fourier functions are sensing functions specific to MR sampling. Sub-Nyquist sampling is possible if  $\Phi$  can produce a set of measurements  $\mathbf{y} \in \mathbb{C}^M$  with  $M \ll N$  from which we can perfectly or approximately recover  $\mathbf{x}$ .

An immediate problem of sub-Nyquist sampling is that equation (3.2) produces an underdetermined system of linear equations, meaning that for a given  $\mathbf{y}$  there is an infinite number of solutions  $\mathbf{x}$ . CS shows that this problem is nevertheless solvable and tractable provided some characteristics of  $\Phi$  and  $\mathbf{x}$  are fulfilled. Specifically, there are three key topics to be discussed for a full description of CS:

- **Signal sparsity:** CS targets the recovery of signals that can be sparsely represented.
- **Sampling incoherence:** The measurement matrix  $\Phi$  must be a restricted isometry.
- **Non-linear reconstruction:** The CS problem imposes the use of a non-linear reconstruction.

### 3.3.1 Signal sparsity

CS requires that the signal of interest be sparse in its natural form or after transformation [42, 28]. This sparsity-based model is the prior knowledge which assumes redundancy in the signal and constrains the solution set. By definition, all signal models are wrong in the sense that they can not be able to fully explain a given signal [18]. Instead their goal is to encompass under the same umbrella signals sharing specific characteristics, and their usefulness can be evaluated based on their flexibility or specificity in accomplishing a particular task.

Sparsity-based models have gained prominence in the last decade as shown in the timeline of figure 3.2 [48], mainly due to their potential in compactly representing a wide range of natural signals. This has led to their successful application in general processing tasks such as denoising [91], super-resolution [153], classification [154] and segmentation [90]. Recently, one of the most successful models for a variety of applications has been the sparse synthesis model and, although CS methods are not restricted to this modelling, we will be using it for its theoretical description. The sparse synthesis model is characterised by two properties:

- **Synthesis:** Signals are assumed to be a linear combination of basis functions.
- **Sparsity:** The coding vector defining the synthesis of a signal from basis functions is sparse, *i.e.* uses a restricted number of non-zero coefficients.

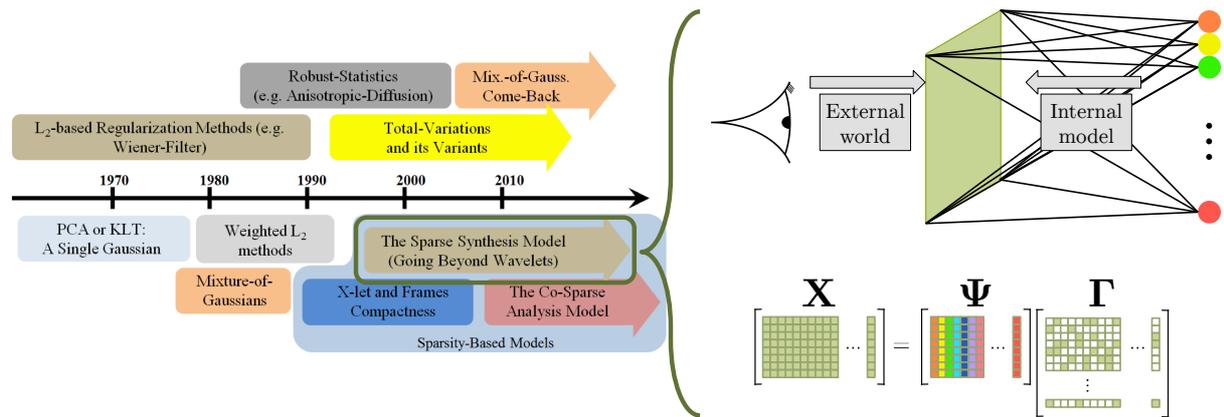


Figure 3.2: Timeline of signal modelling trends as summarised in [48] (left), and physical interpretation of the sparse synthesis model [105] (right). Natural signals are assumed to be a sparse linear combination of transform functions grouped in a basis or frame  $\Psi$ .

Given a signal  $\mathbf{x} \in \mathbb{C}^N$ , we can express it as a coding vector  $\boldsymbol{\gamma} \in \mathbb{C}^K$  in a basis or frame  $\Psi \in \mathbb{C}^{N \times K}$ , with normalised columns  $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \dots, \boldsymbol{\psi}_K$ , such that

$$\|\mathbf{x} - \Psi \boldsymbol{\gamma}\|_p = \epsilon, \quad (3.3)$$

ensures a small or null  $\epsilon$  for approximate or perfect representation respectively. The sparsity

property further imposes that the coding  $\boldsymbol{\gamma}$  is sparse, or equivalently that  $s = \|\boldsymbol{\gamma}\|_0 \equiv \#\{n|\gamma_n \neq 0\} \ll K$ . In approximation methods, norms typically used to measure model deviation are  $l_p$  norms with  $p = 1, 2$ , and  $\infty$ . Throughout this work we will concentrate on  $p = 2$  when measuring modelling errors.

Assuming a perfect approximation of the signal by its sparse representation, the system of linear equations modelling the sampling mechanism becomes

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\gamma}, \quad (3.4)$$

where we have used the composite matrix  $\mathbf{A} \in \mathbb{C}^{M \times K}$  to describe the combined effect of the sensing matrix  $\boldsymbol{\Phi}$  and the sparse synthesis matrix  $\boldsymbol{\Psi}$ . Recovering  $\mathbf{x}$  from samples  $\mathbf{y}$  is therefore equivalent to finding its representation  $\boldsymbol{\gamma}$ , from which we have the strong prior knowledge that it is a sparse vector.

Contrary to the forward transformation (finding  $\mathbf{y}$  from  $\boldsymbol{\gamma}$ ), the inverse transformation is highly non-linear, and consists on finding the  $s \ll K$  coefficients in  $\boldsymbol{\gamma}$  that will best approximate  $\mathbf{y}$ , or the sparsest code that will keep modelling error below a certain threshold:

$$\min_{\boldsymbol{\gamma}} \|\mathbf{y} - \mathbf{A}\boldsymbol{\gamma}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\gamma}\|_0 \leq s, \quad (\text{P0})$$

$$\min_{\boldsymbol{\gamma}} \|\boldsymbol{\gamma}\|_0 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\boldsymbol{\gamma}\|_2^2 \leq \epsilon. \quad (\text{P0-}\epsilon)$$

We will refer to these as the sparsity constrained and the error constrained  $l_0$  norm problems, and respectively refer to them as problems (P0) and (P0- $\epsilon$ ). Both problems are equivalent in that it is possible to select an error  $\epsilon$  that will provide the same solution as a maximum sparsity  $s$ . In practical situations,  $\epsilon$  is always non-zero because the sparse approximation is never perfect and any measurement mechanism inevitably introduces unwanted noise. The CS problem looks at the special case when  $M < K$  for this non-linear and NP-hard problem.

The analysis of CS has focussed on the case when  $\boldsymbol{\Psi}$  is an orthonormal basis, which is the case

for many wavelet transforms [91]. Contrary to the measurement matrix, the sparsity matrix can be fully designed and in some situations it is useful to exploit overcomplete more flexible transforms to find representations of a signal that are sparser than any orthonormal transform can provide [118]. In such cases it is common to refer to the transform as a dictionary  $\mathbf{D}$ , with more columns  $K$  than rows  $N$ , and where the columns of the transform can be highly correlated. This special case will be recurrent throughout this work and its implications will be explained in more detail in section 3.4.

### 3.3.2 Sampling incoherence

One of the most important questions is to determine under which circumstances equations (P0) and (P0- $\epsilon$ ) have a unique solution. It is simple to see that some designs of matrix  $\mathbf{A}$  will make finding a unique minimiser hopeless. For instance, if we are given the prior knowledge that signal  $\boldsymbol{\gamma}$  has sparsity 1 but all the columns of  $\mathbf{A}$  are identical, it will be impossible to locate the support of the non-zero coefficient in  $\boldsymbol{\gamma}$  from samples  $\mathbf{y}$  [56]. To avoid this scenario, we require that different  $\boldsymbol{\gamma}$  vectors give different sampling vectors  $\mathbf{y}$  after transformation  $\mathbf{A}$ .

This intuitive thought was formalised by Donoho and Elad [43, 44] with the notion of coherence of matrix  $\mathbf{A}$ , which measures the similarity between columns of  $\mathbf{A}$  and is given by

$$\mu(\mathbf{A}) = \max_{i \neq j} |\mathbf{a}_i^H \mathbf{a}_j|. \quad (3.5)$$

This is a concept that is recurrent in the analysis of the sparse recovery problem. A low coherence prevents two different sparse vectors  $\boldsymbol{\gamma}$  from collapsing onto the same measurement vector  $\mathbf{y}$  as in the trivial example above. It can be shown for instance that if a solution  $\boldsymbol{\gamma}^*$  is found for which

$$\|\boldsymbol{\gamma}^*\|_0 < \frac{1 + \frac{1}{\mu(\mathbf{A})}}{2}, \quad (3.6)$$

it is necessarily the sparsest solution of equation (P0- $\epsilon$ ) for  $\epsilon = 0$  [43].

In the particular case of CS, where the transform  $\mathbf{A}$  is a composite of a measurement matrix and a sparsity matrix, a similar concept is the mutual coherence [45] between the bases, given by

$$\mu(\Phi, \Psi) = \max_{i \leq M, j \leq K} |\phi_i^H \psi_j|. \quad (3.7)$$

Bounds similar to equation (3.6) dependent on mutual coherence can be found that ensure solution uniqueness. A low mutual coherence degree prevents a sparse combination of rows in  $\Phi$  to represent the columns of  $\Psi$  and vice-versa, meaning that a sparse vector  $\gamma$  will produce a dense signal  $\mathbf{y}$ . An example of bases producing minimal mutual coherence is the Fourier and the canonical (standard) basis.

A similar analysis was pioneered by Donoho and Elad [43] with the definition of the Spark of matrix  $\mathbf{A}$ , which is the minimum number of linearly dependent columns in  $\mathbf{A}$ . This deterministic analysis of the theory of sparse recovery however includes worst case scenarios and leads to very pessimistic uniqueness bounds. This motivated the probabilistic analysis led by Candès and Tao [27] with the idea of restricted isometry property (RIP). Matrix  $\mathbf{A}$  is said to satisfy the RIP with restricted isometry constant  $\delta_s$  if

$$(1 - \delta_s) \|\gamma\|_2^2 \leq \|\mathbf{A}\gamma\|_2^2 \leq (1 + \delta_s) \|\gamma\|_2^2. \quad (3.8)$$

The RIP effectively measures how much  $\mathbf{A}$  deviates from orthonormality, and is closely related to its Spark as defined by Donoho and Elad, as it geometrically ensures that two sparse vectors in the  $N$ -dimensional space maintain approximately their Euclidean distance in the reduced  $M$ -dimensional space.

Provided  $\delta_s$  is maintained small, it will be possible to recover  $\gamma$  from  $\mathbf{y}$ . It is an NP hard problem to determine whether a matrix is a restricted isometry, but it can be proven that some classes of matrices will exhibit a low restricted isometry constant with high probability [25]. This is the case for instance of matrices with random Gaussian or Bernoulli entries, and also applies to random subsets of the DFT [120]. In MRI we are restricted by the physical

acquisition mechanism to sample in the Fourier domain, meaning that a random subsampling of k-space is extremely likely to lead to a RIP system, hence enabling the natural application of CS to MRI without any hardware modifications.

The conditions discussed were derived for sparsity transforms which are orthonormal or tight frames, and are necessary if the requirement is to find the sparse code  $\gamma$ . Looking for a sparse representation of signal  $\mathbf{x}$  in an overcomplete dictionary  $\mathbf{D}$  would be a terrible choice according to this analysis because their columns are highly correlated, meaning that a single sparse vector  $\gamma$  would have many different representations in  $\mathbf{y} = \Phi\mathbf{x} = \Phi\mathbf{D}\gamma$ . However, if the focus is on recovering the signal  $\mathbf{x}$  instead of the original  $\gamma$ , dictionary incoherence is not a requirement as long as it is possible to recover one out of the many sparse representations that the signal  $\mathbf{x}$  has in dictionary  $\mathbf{D}$ .

Sparse recovery in overcomplete, highly correlated dictionaries has received less attention and recovery guarantees are very recent. Some results were presented in [56], where notions of  $D$ -Spark and  $D$ -RIP are introduced for theoretical analysis of recovery bounds. These are closely linked to the traditional Spark and RIP measures but focus the analysis on signal  $\mathbf{x}$  and the measurement matrix  $\Phi$ , so for instance, matrix  $\Phi$  satisfies the  $D$ -RIP condition with constant  $\delta_s^D$  if it is the smallest value for which

$$(1 - \delta_s^D)\|\mathbf{x}\|_2^2 \leq \|\Phi\mathbf{x}\|_2^2 \leq (1 + \delta_s^D)\|\mathbf{x}\|_2^2, \quad (3.9)$$

where  $\mathbf{x}$  is any vector which can be represented in  $\mathbf{D}$  by an  $s$ -sparse code  $\gamma$ . Notice that here there is no requirement on the sparse representation, which can take many forms and does not need to be unique. Sparse coding with overcomplete and redundant dictionaries will be studied in more detail in section 3.5.

### 3.3.3 Non-linear reconstruction

Ensuring a unique minimiser for equations (P0) and (P0- $\epsilon$ ) does not necessarily mean that the solution can be found. The use of the  $l_0$  norm as a measure of sparsity makes the problem non-convex and only a combinatorial search could guarantee finding the global minimiser. There have nevertheless been multiple attempts at solving this problem through heuristics with efficient algorithms. In this section we review a few successful solutions, some of which will be used later on in this work.

#### $l_0$ norm minimisation

Some greedy methods have been successful at approximating the combinatorial problem of  $l_0$  norm minimisation. A greedy algorithm is any method that proposes to iteratively seek the locally optimal choice in the hope of approximating the global minimum. This is a reasonable compromise for the sparse approximation problem given that the global solution to the NP hard problem is practically unreachable, but sequentially deciding the entries in  $\gamma$  that will minimise the approximation error is computationally very cheap.

This is the approach suggested by the matching pursuit (MP) algorithm family, initially proposed by Mallat and Zhang [92]. At each iteration, the support of the sparse vector  $\gamma$  is increased by one based on the transform function that best correlates with the residual left when the current approximation is subtracted from the signal. More specifically, starting from a residual equivalent to the measurements,  $R^{(0)} = \mathbf{y}$ , an empty sparse code  $\gamma = 0$ , and assuming normalised transform columns  $\|\mathbf{a}_k\|_2 = 1, \forall k$ , transform functions are iteratively chosen as

$$\omega^{(i)} = \arg \max_{1 \leq k \leq K} |\langle R^{(i-1)}, \mathbf{a}_k \rangle|, \quad (3.10)$$

providing an update for the sparse vector at the selected location

$$\gamma(\omega^{(i)}) = \langle R^{(i-1)}, \mathbf{a}_{\omega^{(i)}} \rangle, \quad (3.11)$$

and the update of the residual as

$$R^{(i)} = R^{(i-1)} - \gamma(\omega^{(i)})\mathbf{a}_{\omega^{(i)}}. \quad (3.12)$$

The approximation error  $\|\mathbf{y} - \mathbf{A}\boldsymbol{\gamma}\|_2^2$  decreases with each iteration until its energy falls below a target precision or a maximum number of chosen transform functions is reached. Although MP is extremely simple and its convergence is exponential, it requires an infinite number of iterations to perfectly represent a signal when  $\mathbf{A}$  is non-orthogonal. This is because MP updates the residual  $R^{(i)}$  such that it is orthogonal only to the current function  $\mathbf{a}_{\omega^{(i)}}$  at iteration  $i$ . If successive functions  $\mathbf{a}_k$  are selected which are not orthogonal, the algorithm will select the same functions multiple times.

The algorithm was extended to prevent the reuse of the same basis functions, accelerating the convergence rate. This can be achieved by introducing an additional operation at each iteration, involving the projection of the residual onto the span of the previously selected subset of functions. This is known as orthogonal matching pursuit (OMP) [109]. Although this step increases considerably the computational complexity due to the projection operator, it guarantees convergence in at most  $N$  steps. In OMP it is therefore necessary to keep track of the set  $\Omega$  of selected basis functions. The OMP algorithm is summarised in algorithm 1, where the orthogonalisation step is shown in line 4 with  $\mathbf{A}_{\Omega^{(i)}}^+$  denoting the pseudoinverse.

**Algorithm 1:** OMP algorithm for sparse coding

**Input:**  $\mathbf{y} \in \mathbb{C}^N$  - Signal to code  
 $\mathbf{A}$  - Sparse synthesis transformation from sparse representation to signal.  
 $s$  and/or  $\epsilon$  - Maximum number of sparse coefficients and/or maximum approximation tolerance  
**Output:**  $\boldsymbol{\gamma}$  - Sparse coding vector  
**Initialise:** Iteration count  $i = 0$ , residual  $R^{(0)} = \mathbf{y}$ , function set  $\Omega^{(0)} = \emptyset$ , sparse code  $\boldsymbol{\gamma}^{(0)} = 0$   
**repeat**  
    1.  $i \leftarrow i + 1$   
    2.  $\omega^{(i)} \leftarrow \arg \max_{1 \leq k \leq K} |\langle R^{(i-1)}, \mathbf{a}_k \rangle|$   
    3.  $\Omega^{(i)} \leftarrow \Omega^{(i-1)} \cup \mathbf{a}_{\omega^{(i)}}$   
    4.  $\boldsymbol{\gamma}^{(i)} \leftarrow \mathbf{A}_{\Omega^{(i)}}^+ \mathbf{x}$   
    5.  $R^{(i)} \leftarrow \mathbf{x} - \mathbf{A}_{\Omega^{(i)}} \boldsymbol{\gamma}^{(i)}$   
**until**  $i = s$  and/or  $\|\mathbf{y} - \mathbf{A}\boldsymbol{\gamma}^{(i)}\|_2^2 \leq \epsilon$ ;

Different extensions have been suggested, such as regularised orthogonal matching pursuit [100], stagewise orthogonal matching pursuit [47] and compressive sampling orthogonal matching

pursuit [99]. The objective is generally to accelerate convergence with alternative strategies to detect the correct support  $\Omega$  with as few iterations as possible.

### **$l_1$ norm minimisation**

A major inconvenience for solving equations (P0) and (P0- $\epsilon$ ) is that they are non-convex problems. An alternative is to look for the solution to the closest problem that is convex, for which the vast literature on convex optimisation would immediately apply. This was proposed by Chen, Donoho and Saunders in their seminal paper [32] with the relaxation of the  $l_0$  norm by the  $l_1$  norm, targeting the solutions to

$$\min_{\gamma} \|\mathbf{y} - \mathbf{A}\gamma\|_2^2 \quad \text{s.t.} \quad \|\gamma\|_1 \leq s, \quad (\text{P1})$$

$$\min_{\gamma} \|\gamma\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\gamma\|_2^2 \leq \epsilon. \quad (\text{P1-}\epsilon)$$

Finding solutions for these problems is generally referred to as the basis pursuit problem.

Under some circumstances the solutions to equations (P1) and (P1- $\epsilon$ ) coincide with the optimal minimiser of equations (P0) and (P0- $\epsilon$ ). A rigorous proof for this equivalence can be found in [27], but it can also be intuitively explained by considering the simple 2D problem shown in figure 3.3. Assume we look for a sparse signal  $\gamma \in \mathbb{R}^2$  of only two entries, for which we have a single measurement  $y = \Phi\Psi\gamma \in \mathbb{R}$  acquired through a sparsifying matrix  $\Psi \in \mathbb{R}^{2 \times 2}$  and a sensing matrix  $\Phi \in \mathbb{R}^{1 \times 2}$ . Letting the sparsifying matrix be represented by the Cartesian coordinate axes, the infinite solution set resulting from the measurement could be represented by a line. Solving the  $l_p$  norm optimisation is interpreted as inflating a ball of different shapes from the origin of the axes until it hits the solution set. Although the  $l_1$  norm diamond shape is very different from the spiky shape of the  $l_p$  norm,  $0 \leq p < 1$ , it also grows along the axes, hence favouring a solution where only one coefficient is non-zero. In higher dimensions the problem becomes abstract, but a similar sparsity promoting behaviour of the  $l_1$  norm can be expected.

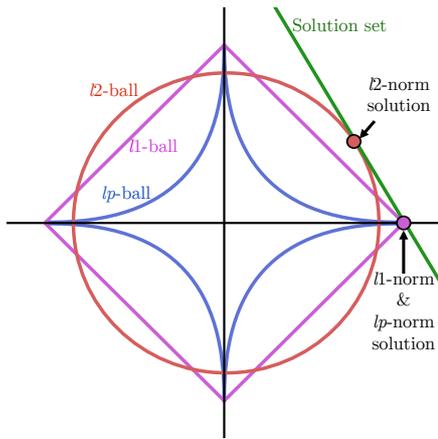


Figure 3.3: Optimisation result using different norms. The intersection of the solution set (in green) and a norm ball represent a solution. The  $l_p$  ball,  $0 \leq p < 1$ , grows along the axes, hence finding a sparse solution, but is not a convex set. The  $l_1$  norm is a convex set that can approximate the sparse result.

Lagrangian multipliers can be used to consider the unconstrained problem

$$\min_{\gamma} \|\mathbf{y} - \mathbf{A}\gamma\|_2^2 + \lambda\|\gamma\|_1. \quad (3.13)$$

The solution of this  $l_1$  norm regularised least-squares (LS) problem can be found with different approaches, such as homotopy methods [46], coordinate-wise descent methods [53], Bregman iterative methods [107] and iterative shrinkage methods [13]. Similarly, equation (3.13) can be recast as a quadratic program that can be solved with standard solvers relying for instance on interior-point methods [143, 72]. Although the latter solution has advantages such as problem tractability and recovery guarantees, it is computationally more demanding than greedy algorithms in general, especially for large scale problems.

### Weighted norm minimisation

An alternative to  $l_0$  and  $l_1$  norm minimisation are weighted norm minimisation methods, which exploit the simplicity of LS regularised minimisation. An analysis of this family of algorithms can be found in [38] under the name of iteratively re-weighted least squares (IRLS), finding the sparse code as  $\gamma = \mathbf{W}\alpha$ , where  $\mathbf{W}$  is a diagonal weighting matrix, and  $\alpha$  is the result to the LS problem

$$\min_{\alpha} \|\alpha\|_2^2 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{W}\alpha\|_2^2 \leq \epsilon. \quad (3.14)$$

Equivalently, we can consider the unconstrained optimisation problem

$$\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{A}\mathbf{W}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_2^2. \quad (3.15)$$

The algorithm to solve this problem is iterative and the solution at iteration  $i$  is given by

$$\boldsymbol{\gamma}^{(i)} = \mathbf{W}^{(i-1)}\boldsymbol{\alpha}^{(i-1)} = \boldsymbol{\Theta}^{(i-1)}\mathbf{A}^H \left( \mathbf{A}\boldsymbol{\Theta}^{(i-1)}\mathbf{A}^H + \lambda\mathbf{I} \right)^{-1} \mathbf{y}, \quad (3.16)$$

with  $\boldsymbol{\Theta} = \mathbf{W}\mathbf{W}^H$ .

The disparities between different IRLS methods mostly lie in the definition of the weighting matrix  $\mathbf{W}$ , which is the variable incorporating prior knowledge about the solution. A popular example of these algorithms is the FOCUSS algorithm introduced in [58], where  $\mathbf{W}$  is chosen to be  $\mathbf{W}^{(i)} = \text{diag}(\boldsymbol{\gamma}^{(i-1)})$  and can be proven to be equivalent to the minimisation of  $\sum_n \log |\boldsymbol{\gamma}(n)|$ . A slight variation was proposed in [57] to address the minimisation of sparsity promoting  $lp$  norms  $0 < p < 1$  with  $\mathbf{W}^{(i-1)} = \text{diag}(|\boldsymbol{\gamma}^{(i-1)}|^{1-p/2})$ .

The principle behind IRLS algorithms is that in order to minimise equation (3.15) where the penalty can be replaced by  $\|\mathbf{W}^{-1}\boldsymbol{\gamma}\|_2^2$ , the non-zero coefficients of the solution  $\boldsymbol{\gamma}$  must concentrate at locations  $n$  where the diagonal entry  $[\mathbf{W}^{-1}]_{n,n}$  is a small value. Starting close to a particular sparse solution, the algorithm is expected to provide a sequence  $\{\boldsymbol{\gamma}^{(i)}\}_{i=0}^{\infty}$  converging to it. A major drawback of this technique is that once a location has shrunk to zero it cannot be recovered as a non-zero coefficient, but some solutions exist that include a monotonically decreasing constant for the weighting matrix.

### 3.4 Dictionary learning

The assumption that the signal of interest can be sparsely represented in some domain is the foundation of CS, and it is therefore important to consider which form this domain should take. Throughout the past decades, different forms of representation have emerged, leading in

recent years to dictionary learning (DL). DL is the term given to the search for optimal sparse signal transforms which are obtained through a training stage. In what follows we present a brief history of signal transforms leading to DL and a summary of some of the DL algorithms available. Further details can be found in the excellent reviews [118, 134].

### 3.4.1 From linear transforms to adaptive dictionaries

The choice of a transform for signal representation is crucial and involves a number of compromises. The use of orthogonal or bi-orthogonal transforms has long been favoured because transform coefficients are given by a simple inner product between the signal and the transform or the transform inverse respectively. However, the use of complete bases has limitations in representation flexibility as some signals may not be well encompassed by their modelling. The desire for greater flexibility at the expense of mathematical complexity drove the switch from complete transform bases to overcomplete dictionaries, and from transform functions to dictionary atoms.

#### The convenience of orthogonal transforms

One of the most recurrent signal analysis tools is the Fourier transform, which was greatly popularised in the 1960s with the emergence of the fast Fourier transform proposed by Cooley and Tukey [34]. The decomposition of a signal into its global frequency content can sparsely represent uniformly smooth signals, but is very inefficient for capturing discontinuities given that their energy is spread among several frequency coefficients. Sharp discontinuities are rare in natural signals, but the periodic assumption of finite signals for the computation of its transform artificially creates them at the signal boundary. This naturally led to the use of the discrete cosine transform (DCT), which avoids this phenomenon by assuming odd periodicity and is the core ingredient of the JPEG image compression standard [146].

The following decades of the 1970s and 1980s centred the search of data simplicity on the

data itself. Statistical tools such as principal component analysis [67] and most notably the Karhunen-Loève transform [91] gained interest as they reduced the complexity of the signal on a low dimensional subspace with minimum  $l_2$  norm error. Using the  $s$  first eigenvectors of the eigenvalue decomposition of a signal's covariance matrix, it can be seen as a low dimensional Gaussian data fit. Although it is more powerful as a data sparsifier than the Fourier transform, it is considerably more complex given its data-driven nature.

During the 1980s, it became clear that the search for simpler, sparser representations required the departure from restrictive linear transforms leading to the design of non-linear transforms, where the support of non-zero coefficients is signal specific. Two major concepts are at the origin of wavelet design, emerging with the specific purpose of non-linear sparse coding for natural signals: localisation and multi-resolution.

The Fourier transform allows the identification of the different frequency content of a signal, but it does not reveal where in time or space this content can be found. This lack of localisation hinders compact signal representation, and results of this realisation were the short time Fourier transform and Gabor filters [54]. Multi-resolution analysis has been the consequence of noticing how natural signals exhibit fractal-like patterns, which repeat at different scales. Multi-scale wavelet analysis was introduced by Grossman and Morlet in [60] as the scaling and translation of a single function of finite support which could be designed to form an orthogonal basis. Mallat [91] later extended this concept for optimal 1D multi-resolution signal analysis, and most importantly, developed fast algorithms for wavelet decomposition enabling their practical use. Even though at higher dimensions wavelet analysis loses its optimality, these advances were adopted in the newer JPEG2000 image compression mechanism [127].

Wavelet analysis also suffers shortcomings, such as the lack of adaptability and geometric invariance. The orthogonality condition limits the range of temporal or spatial support of the functions which, if broken, allows for greater flexibility in representation. Wavelet packets [33] suggested the use of such an extension to wavelets which, given a signal, could be reduced to the optimal orthogonal subset, gaining adaptability but keeping the attractive properties of

orthogonal wavelets. Translation and rotation sensitivity are further drawbacks of standard wavelet transforms which were deemed unavoidable by orthogonal transforms in [126]. Thus began work on overcomplete transforms with early examples such as the stationary wavelet transform [9] seeking geometric invariance.

### **The flexibility of overcomplete dictionaries**

With the development in the 1990s of greedy algorithms for equation (P0) and the influential discovery that this problem could be approximated by the tractable equation (P1), the use of overcomplete frames adopting the name of dictionaries was popularised. Allowing multiple representations of the same signal in a dictionary of atoms opened new perspectives in coding design, which could now be driven by a cost function, and markedly separated the task of dictionary design from signal coding. Simple concatenations of bases could overcome what used to be fundamental limitations. For instance, a Fourier transform was unable to compactly represent discontinuities, but concatenating it with a Dirac basis could solve this problem.

Abandoning orthogonality paved the way for creative dictionary design. Two trends can currently be identified: the design of analytic dictionaries and data-driven adaptive dictionaries. The former approach relies on a mathematical model of the data to generate the dictionary and is usually characterised by efficient mechanisms to implicitly compute transform coefficients, as well as robust theoretical guarantees for signal approximation. Some examples of this category are curvelets [23], contourlets [40] and bandelets [74]. Data-driven dictionary design is more recent, and draws from example observations of a signal to obtain an optimal representation. Adaptive dictionaries are powerful as there is no reason to believe a single dictionary should be optimal for all kinds of signals, but come at the price of increased processing complexity and weaker theoretical guarantees. The search for optimal dictionaries for a specific set of training signal is known as the dictionary learning (DL) problem.

### 3.4.2 Dictionary learning problem

With the separation in the signal representation problem of dictionary design and coding stages, a question naturally emerges: what is the optimal dictionary for a particular task and signal? In this work DL will be applied on image patches, so we formulate the problem adopting the notation  $N_p$  to denote the dimension of a vectorised image patch. Given a set  $\mathbf{X}^T \in \mathbb{C}^{N_p \times T}$  of  $T$  observation signals of size  $N_p$ , a DL algorithm looks for the dictionary  $\mathbf{D} \in \mathbb{C}^{N_p \times K}$  of  $K$  atoms, that yields the sparse coding matrix  $\mathbf{\Gamma}^T \in \mathbb{C}^{K \times T}$  optimally solving

$$\min_{\mathbf{\Gamma}^T, \mathbf{D}} \|\mathbf{X}^T - \mathbf{D}\mathbf{\Gamma}^T\|_F^2 \quad \text{s.t.} \quad \|\boldsymbol{\gamma}_n^T\|_0 \leq s \quad \forall n, \quad (\text{DL})$$

or similarly

$$\min_{\mathbf{\Gamma}^T, \mathbf{D}} \|\boldsymbol{\gamma}_n^T\|_0 \quad \text{s.t.} \quad \|\mathbf{x}_n^T - \mathbf{D}\boldsymbol{\gamma}_n^T\|_2^2 \leq \epsilon \quad \forall n, \quad (\text{DL-}\epsilon)$$

where  $\mathbf{x}_n^T$  and  $\boldsymbol{\gamma}_n^T$  are columns of  $\mathbf{X}^T$  and  $\mathbf{\Gamma}^T$  respectively. The superscript  $.^T$  is chosen to highlight the fact that the observation dataset and the corresponding sparse codes are training signals, but it is dropped for the rest of the description for readability.

These problems are intricately related to sparse coding, but have the additional difficulty that, on top of finding a sparse code, the dictionary for sparse representation has to be simultaneously estimated. Similarly to the sparse coding problem, this is a non-convex optimisation statement for which only heuristics exist. Commonly, the problem is simplified by solving for the sparse code and the dictionary separately, and iteratively alternating their solutions until convergence.

### 3.4.3 Dictionary learning algorithms

Algorithms targeting equations (DL) and (DL- $\epsilon$ ) are relatively recent, but can already be distinguished based on the approach they take to provide an approximate solution. In the following section we highlight probabilistic methods and clustering methods, which have been among the most successful so far. We also note that the focus here is on learning general

dictionaries, and do not mention methods that try to enforce a specific structure and properties through a parametric training approach.

### Probabilistic methods

One of the pioneering works on DL was presented by Olshausen and Field in [105]. Instead of tackling equations (DL) and (DL- $\epsilon$ ) they proposed a maximum likelihood (ML) learning algorithm for natural images where the likelihood of the data is maximised as

$$\max_{\mathbf{D}} (\log P(\mathbf{x}_n | \mathbf{D})) = \max_{\mathbf{D}} \left( \log \int_{\gamma_n} P(\mathbf{x}_n | \gamma_n, \mathbf{D}) P(\gamma_n) d\gamma_n \right). \quad (3.17)$$

Two assumptions relate this expression to the general DL problem. First, assuming the priors  $P(\gamma_n)$  are Laplacian distributions enforces sparsity on the coding vector. Also, if the coding approximation error is modelled as Gaussian zero-mean, the problem becomes the energy optimisation

$$\min_{\mathbf{D}, \gamma_n} E(\mathbf{x}_n, \gamma_n | \mathbf{D}) = \min_{\mathbf{D}, \gamma_n} \|\mathbf{x}_n - \mathbf{D}\gamma_n\|_2^2 + \lambda \|\gamma_n\|_1. \quad (3.18)$$

As we have previously seen, the  $l_1$  norm can be seen as a convex relaxation of the  $l_0$  norm.

The alternate solution of equation (3.18) for  $\gamma_n$  and  $\mathbf{D}$  separately is trivial. The solution of the former can use a standard sparse coding technique whereas the latter requires solving a simple LS problem. In order to account for statistics of a set of images, the method minimises the average energy  $\mathbb{E}[E(\mathbf{x}_n, \gamma_n | \mathbf{D})]$  over multiple signals. Although the learnt dictionary is not guaranteed to be a global minimum, the result has deep implications unveiling an interesting relationship between sparse coding and the human visual cortex, which is known to be an efficient image encoder.

A similar two-step approach is taken by the method of optimal directions (MOD) [50]. Considering the entire example dataset, it looks for a solution to equation (DL). The MOD introduces two variations compared to the previous algorithm. The sparse coding stage is performed with

OMP instead of gradient descent and a closed-form solution is formulated for the update of the dictionary. Although both changes resulted in an accelerated convergence compared to the previous approach, this method is also prone to local minima.

One problem of the alternating approach to dictionary and sparse code update is that if the training set is extended with additional data the dictionary must be learnt from scratch again. Also, batch processing of the entire training data can be computationally expensive for very large training sets. Online dictionary learning [89] tries to solve these issues by optimising on a subset of the training set, which is then progressively increased to account for additional training data. Learnt dictionaries are successively reused as initialisations as the training set is augmented, leading to a more efficient training.

### Clustering methods

A different family of learning algorithms is based on clustering methods. Although these methods also share the alternating mechanism to iteratively update the sparse code and the dictionary, the dictionary is found as a clustering process rather than as a ML solution. One of the most recurrent clustering methods, K-means clustering, was exploited in [121] as a vector quantisation method for patch-based video coding. With the assumption that each patch is represented by a single atom, each patch can trivially be assigned the closest atom and the dictionary can subsequently be updated by individually modifying its atoms such that the distance to their assigned patches is minimised. Although efficient, sparse coding with a single atom is very restrictive.

The generalisation of this method was proposed with the K-SVD algorithm [2], by Aharon *et al.*. The name was chosen to emphasise the close link with the K-means algorithm, the only difference lying on the fact that the dictionary update requires a singular value decomposition (SVD) per dictionary atom instead of a mean operation. The general DL problem in equation (DL) is iteratively solved for  $\mathbf{\Gamma}$  and  $\mathbf{D}$  using a OMP sparse coding step and a dictionary update step.

This method is algorithmically very similar to the MOD algorithm but a fundamental difference lies in the dictionary update stage. For  $\mathbf{\Gamma}$  fixed, the K-SVD algorithm decomposes the penalty term by questioning one atom  $\mathbf{d}_k$  and its associated sparse codes given in row  $k$  of  $\mathbf{\Gamma}$ . Denoting this row vector as  $\boldsymbol{\gamma}_t^k$ , the penalty can be rewritten as

$$\begin{aligned} \|\mathbf{X} - \mathbf{D}\mathbf{\Gamma}\|_F^2 &= \left\| \mathbf{X} - \sum_{i=1}^K \mathbf{d}_i \boldsymbol{\gamma}_t^i \right\|_F^2 \\ &= \left\| \left( \mathbf{X} - \sum_{j \neq k} \mathbf{d}_j \boldsymbol{\gamma}_t^j \right) - \mathbf{d}_k \boldsymbol{\gamma}_t^k \right\|_F^2 \\ &= \|\mathbf{E}^k - \mathbf{d}_k \boldsymbol{\gamma}_t^k\|_F^2, \end{aligned} \tag{3.19}$$

where  $\mathbf{E}^k$  would be the approximation error if atom  $\mathbf{d}_k$  were to be removed from the dictionary.

With this separation, the approximation  $\mathbf{D}\mathbf{\Gamma}$  has been divided into  $K$  rank-1 matrices, only one of which is being questioned for update. Finding the rank-1 approximation of  $\mathbf{E}^k$  through SVD and using it for the update of  $\mathbf{d}_k$  and  $\mathbf{x}_t^k$  would be the optimal update step. However, this is likely to fill row  $\mathbf{x}_t^k$ , which we would like to keep as a sparse vector. A simple solution is to only consider the indices  $\omega_k$  of non-zero entries in  $\mathbf{x}_t^k$ , and define the shrunken vector  $\mathbf{x}_r^k$  and matrix  $\mathbf{E}_r^k$ . The rank-1 approximation of this new error matrix  $\mathbf{E}_r^k$  provides then the optimal update of  $\mathbf{d}_k$  and  $\mathbf{x}_r^k$  while the sparse coding support is either unchanged or reduced. This shrinkage operation is illustrated in figure 3.4, and the full K-SVD algorithm is summarised in algorithm 2.

### 3.5 Patch-based dictionary sparse coding

In this section we look at practical considerations of sparse recovery problems using dictionaries. Specifically, we present the implications for sparse coding brought by overcompleteness, redundancy and adaptability. The following results use the Batch-OMP implementation by Rubinstein described in [119].

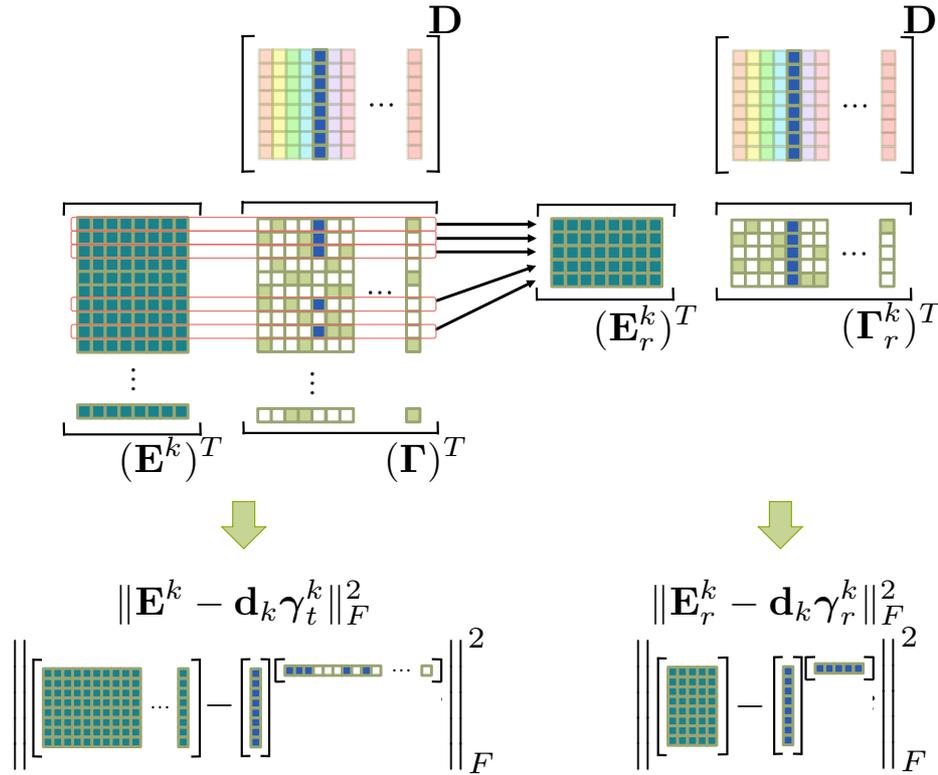


Figure 3.4: K-SVD dictionary update step. The left hand side of the figure shows the decomposition of the approximation error in  $\mathbf{E}^k$  and the contribution from  $\mathbf{d}_k$  and  $\gamma_t^k$  through a rank-1 matrix. Updating  $\mathbf{d}_k$  and  $\gamma_t^k$  directly with an SVD decomposition of  $\mathbf{E}^k$  does not guarantee the maintenance of sparsity in  $\gamma_t^k$ . Instead, shrinking the matrices as shown in the right hand side of the figure solves this problem as only the active support of  $\gamma_r^k$  is updated.

### 3.5.1 Overcompleteness

Consider the sparse recovery problem

$$\min_{\gamma} \|\mathbf{x} - \mathbf{D}\gamma\|_2^2 \quad \text{s.t.} \quad \|\gamma\|_0 \leq s, \quad (3.20)$$

where  $\mathbf{D} \in \mathbb{C}^{N \times K}$ ,  $N \leq K$ , is a dictionary. The solution to this problem is trivial in the complete, orthonormal case ( $N = K$ ,  $\mathbf{D}^H \mathbf{D} = \mathbf{I}$ ), given that the generalised Parseval's theorem holds between  $\mathbf{x}$  in the signal domain and  $\gamma$  in the sparsity domain. Therefore energy is preserved upon basis transformation implying that  $\|\mathbf{x} - \mathbf{D}\gamma\|_2^2 = \|\mathbf{D}^H \mathbf{x} - \gamma\|_2^2$ , and so the best  $s$  sparse representation is trivially given by the  $s$  largest coefficients of the transform  $\mathbf{D}^H \mathbf{x}$ . Energy preservation between domains is however violated as soon as  $N < K$ , which

**Algorithm 2:** K-SVD algorithm for dictionary learning

**Input:**  $\mathbf{X} \in \mathbb{C}^{N \times T}$  - Signal observation dataset  
 $s$  and/or  $\epsilon$  - Maximum number of sparse coefficients and/or maximum representation deviation from original signal  
 $I$  - Maximum number of training iterations

**Output:**  $\mathbf{D}$  - Trained dictionary

**Initialise:** Iteration count  $i = 0$ , dictionary  $\mathbf{D}^0$ , sparse code  $\mathbf{\Gamma}^0 = 0$

**repeat**

1.  $i \leftarrow i + 1$
2. Independent OMP sparse code update:  
 $\gamma_n^{(i)} \leftarrow \arg \min_{\gamma_n} \|\mathbf{x}_n - \mathbf{D}\gamma_n\|_2^2$  s.t.  $\|\gamma_n\|_0 \leq s \forall n$   
or  $\gamma_n^{(i)} \leftarrow \arg \min_{\gamma_n} \|\gamma_n\|_0$  s.t.  $\|\mathbf{x}_n - \mathbf{D}\gamma_n\|_2^2 \leq \epsilon \forall n$
3. Independent dictionary atom update with SVD decomposition:  
 $\omega_k \leftarrow \{n | 1 \leq n \leq N, \gamma_n^{(i)}(k) \neq 0\}$   
 $\mathbf{E}_r^k \leftarrow \mathbf{X} - \sum_{n \neq k} \mathbf{d}_n \gamma_n^{(i)}$   
Construct  $\mathbf{E}_r^k$  by choosing only columns assigned by  $\omega_k$   
SVD decomposition:  $\mathbf{E}_r^k = \mathbf{U}\mathbf{\Delta}\mathbf{V}^T$   
 $\mathbf{d}_k^{(i)} \leftarrow \mathbf{u}_1$   
 $\gamma_r^{k,(i)} \leftarrow \mathbf{\Delta}(1,1) \times \mathbf{v}_1$

**until**  $i = I$ ;

adds considerable flexibility to the sparse recovery problem given that it brings about a regime where one signal can have multiple dictionary representations.

To illustrate this, let us assume we extract a  $N = 8 \times 8$  patch from figure 3.5a. We then use OMP to solve equation (3.20) with a sparsity index  $s = 6$ . Lastly, we look at the mean squared error (MSE)  $\frac{1}{N} \|\mathbf{x} - \mathbf{D}\gamma\|_2^2$  of the result produced using discrete cosine transform (DCT) dictionaries of different sizes. This experiment is repeated with  $10^4$  different patches and average results are plotted in figure 3.5c. Examples of the test patches are displayed in figure 3.5b.

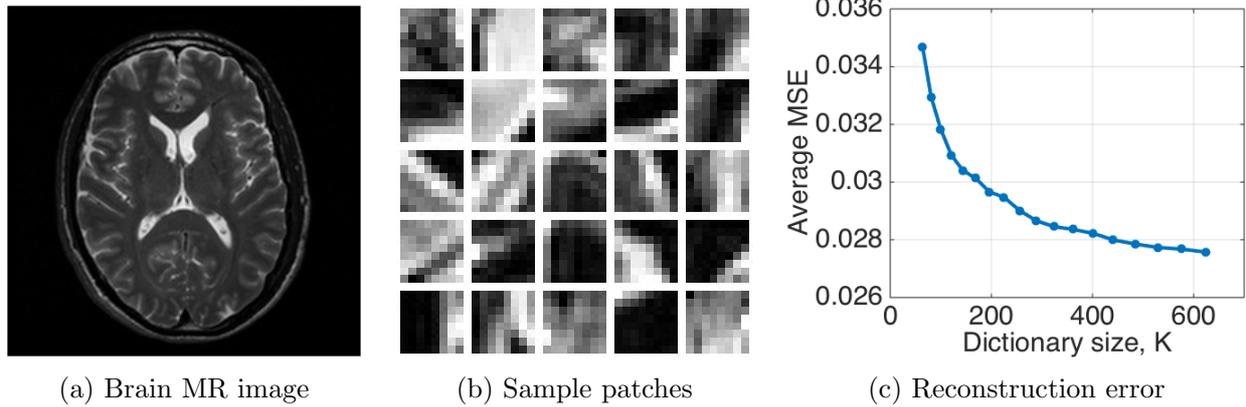


Figure 3.5: Dictionary overcompleteness translates into increased representation sparsity. The plot in (c) shows the average representation error of  $10^4$  patches of size  $8 \times 8$  from image (a) using a DCT dictionary of different sizes and a sparsity index  $s = 6$ . Examples of image patches are shown in (b).

Despite the advantage of domain transformation that an orthonormal dictionary provides, the

dictionary representation becomes more accurate with increasing overcompleteness. In practice, natural signals are only approximately sparse, *i.e.* compressible, but not strictly sparse. According to the CS literature, a penalty in signal reconstruction is paid which is proportional to the accuracy of our sparse approximation [26]. It is therefore important to seek for the data representations which are as accurate as possible for a given sparsity degree.

### 3.5.2 Redundancy

Redundant dictionaries are typically highly coherent, which immediately poses a problem for sparse recovery as we have seen transform incoherence is a necessary requirement. If the focus however is not on the recovery of the particular sparse code that generated the signal but on any sparse code that will approximate it, redundant dictionaries can be useful.

We analyse this statement by considering an overcomplete DCT dictionary with  $K = 225$  atoms of size  $N = 64$ . Given a sparsity degree  $s = 5$ , we synthesise an  $8 \times 8$  patch by linearly combining  $s$  randomly chosen atoms with random weights, and we then try to recover this sparse code with OMP solving

$$\min_{\gamma} \|\gamma\|_0 \quad \text{s.t.} \quad \|\mathbf{x} - \mathbf{D}\gamma\|_2 \leq \epsilon, \quad (3.21)$$

for  $\epsilon = 10^{-5}\sqrt{N}$ . We repeat this experiment  $10^4$  times and show the best and worst recoveries obtained in terms of signal MSE, alongside with the synthetic signal patches.

The greedy approach of OMP to sparse recovery is sometimes able to perfectly find the sequence of sparse coefficients that make up the signal as shown in figure 3.6a. In this case, signal recovery accuracy is down to precision error, well below the threshold set by  $\epsilon$ . However, the coherence of the dictionary can sometimes make OMP fail dramatically in finding the original support of the sparse code as shown in figure 3.6b. Nevertheless, the redundancy in the dictionary makes it possible to find an alternative sparse configuration that still achieves the data consistency level required in signal domain.

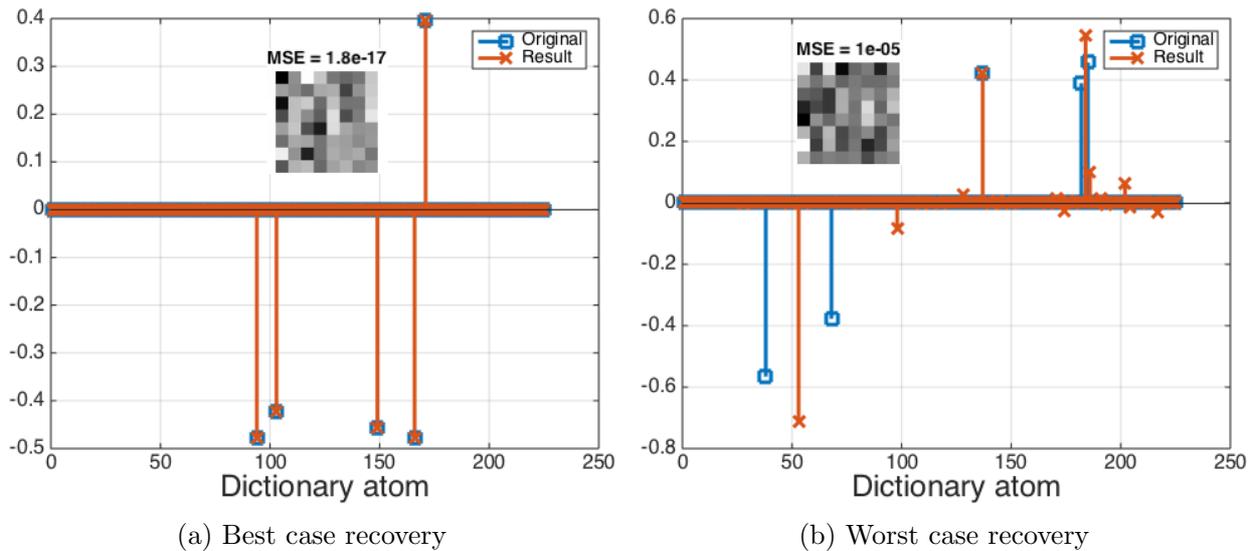


Figure 3.6: OMP sparse recovery examples. Artificial sparse signals of size  $N = 8 \times 8$  are synthesised from  $s = 5$  randomly chosen dictionary atoms from a DCT dictionary of size  $K = 225$ . Compared are the best (a) and worst (b) recoveries with respect to MSE, out of  $10^4$  tries.

This behaviour is analysed for different degrees of sparsity  $s$  in figure 3.7. The plots show average results for the same experiment with  $10^4$  different patches. Despite the correct support recovery decreasing quickly for  $s \geq 2$  (figure 3.7a), the signal domain reconstruction accuracy can be maintained below the predefined threshold (figure 3.7b) at the expense of a denser representation  $\gamma$  (figure 3.7c)<sup>1</sup>.

### 3.5.3 Adaptability

The main advantage of adaptive dictionaries over structured dictionaries is a sparser representation for a predefined set of signals. This comes at the cost of a computationally intensive training process and the loss of structure, meaning that implicit and efficient dictionary transforms are not available and theoretical guarantees of the dictionary are more difficult to derive.

In this section we compare the sparse representations of a structured dictionary with one that is trained using the K-SVD algorithm. For the comparison we use the magnitude brain MR

<sup>1</sup>The sparsity of the result was measured as the number of non-zero coefficients accounting for 99.9% of the energy of the sparse code  $\gamma$ .

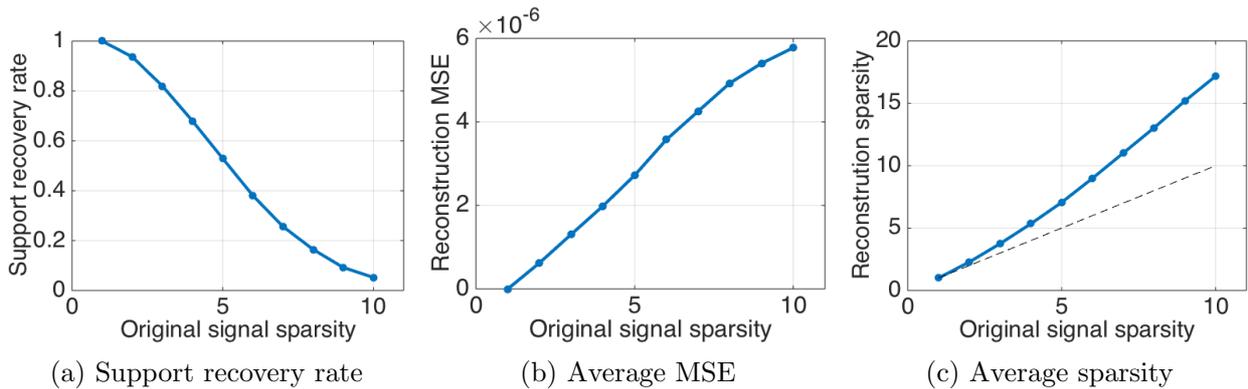


Figure 3.7: Average empirical OMP recovery performance using an overcomplete DCT dictionary. The support recovery rate falls almost to zero for  $s/N = 10/64 = 0.15$  sparsity (a). This is due to the high degree of redundancy in the dictionary, which makes OMP choose sparse coding configurations that are not the ones used to generate the original patches. Nevertheless, OMP is able to maintain the predefined data consistency tolerance (b) by using a few additional atoms relative to the original signal (c).

image shown in figure 3.5a, of size  $N = 256 \times 256$ . Breaking the image down into  $N_p = 8 \times 8$  overlapping patches and assuming patches wrap around the boundaries of the image, we have a total of  $N$  signals to be coded arranged as column vectors in  $\mathbf{X} \in \mathbb{R}^{N_p \times N}$ . Extracting a subset of  $T = 2 \times 10^4$  training patches from a regular grid on the image, we analyse the first 30 iterations of the K-SVD algorithm when we use it to solve equation (DL) with  $s = 5$  for the training of a  $K = 196$  atom dictionary. The initial dictionary is chosen to be a DCT dictionary.

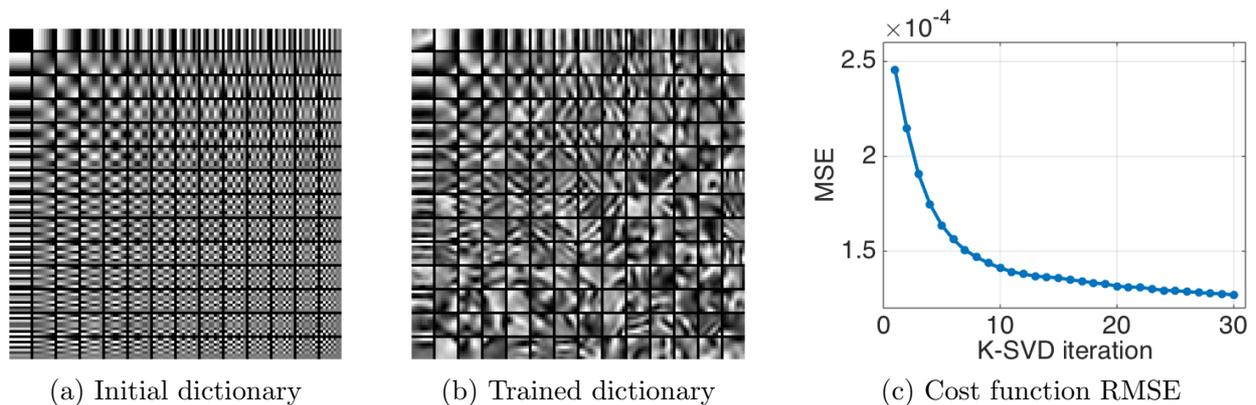


Figure 3.8: Effect of K-SVD training on dictionary and on training dataset. An initial overcomplete DCT dictionary (a), is trained to yield an adapted dictionary (b). The representation accuracy of training signals improves over training iterations (c).

The effects of the DL algorithm are patent in figure 3.8. It is clear how the RMSE of the

representation cost function decreases through the iterations of the K-SVD algorithm. This empirically confirms that the alternating strategy between a sparse coding stage and a dictionary update stage is effectively converging towards at least a local optimum of the problem. Furthermore, in order to achieve this, the dictionary is changing the shape of its atoms, and moves from the initially structured DCT dictionary towards one that incorporates new patterns. The edges and blobs seen in the trained patches are known to be good natural image sparsifiers.

We now focus on the implications that this adaptability has for the sparse approximation of the entire image. Assuming that the training set of patches, which is approximately a third of the full set, is a representative collection of the patches in the image, we should see the same improvement in the representation error when comparing the coding errors of both dictionaries. To recover a coded image from patches we average the contribution of overlapping patches and plot the RMSE obtained from DCT coding and K-SVD coding with 30 iterations. Results are plotted in figure 3.9 for a range of sparsity indices  $1 \leq s \leq 8$ . We also show the accuracy of recovering an image of only random noise with both dictionaries, to highlight how the lack of structure in random features can not be well captured by sparse coding.

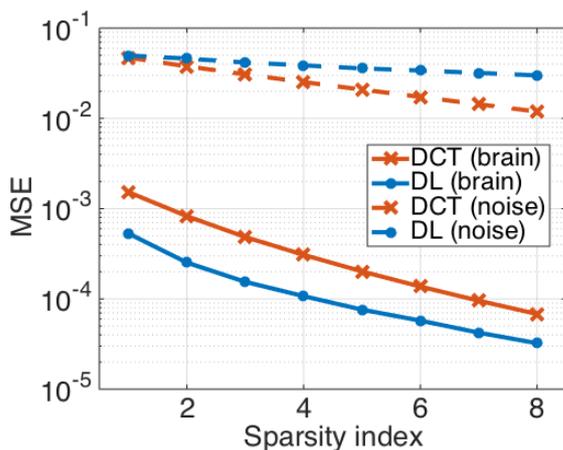


Figure 3.9: Sparse coding reconstruction RMSE of brain MR image. Adapting a dictionary to the brain image reduces the representation error with respect to it, while increasing the error with respect to data of different nature, such as a Gaussian noise image.

The gap in representation accuracy represents the gain that can be achieved through DL relative to the initial structured dictionary. This gap is also visible in the error maps of the approximated images for  $s = 3$ , shown in figure 3.10. Notice how most of the representation error concentrates on edges and fine details of the image. This is expected given that those features are precisely

the ones that will not abide to the sparsity criterion, and are therefore the first ones to be penalised with the assumption of sparsity.

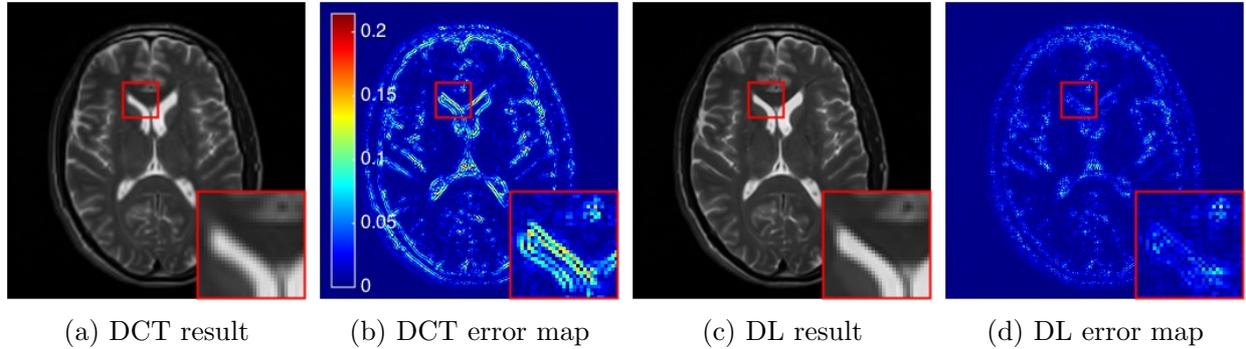


Figure 3.10: Sparse coded approximations of a brain MR image using a DCT dictionary (a, b) and a trained dictionary (c, d) with a sparsity index  $s = 3$ . Error maps show absolute value differences with respect to the original image.

### 3.6 CS and DL in dynamic MRI

MRI has been one of the most fruitful domains for the application of CS, mainly because it offers a solution to speed limitations but also because its acquisition mechanism lends itself to CS requirements without any hardware modifications. In [84], Lustig showed how CS is applicable to many different MR modalities, including structural 2D and dynamic MR. The general CS acquisition and reconstruction problem can be formulated as looking for the image  $\mathbf{x}$  solving the problem

$$\min_{\mathbf{x}} \|\mathbf{F}_u \mathbf{x} - \hat{\mathbf{x}}_u\|_2^2 \quad \text{s.t.} \quad \|\mathcal{S}(\mathbf{x})\|_0 \leq s, \quad (3.22)$$

where  $\mathbf{F}_u$  is an undersampled DFT operator,  $\hat{\mathbf{x}}_u$  is an undersampled k-space acquisition, and  $\mathcal{S}(\cdot)$  is a sparsifying operator.

It can be shown that a random undersampling of the DFT operator has high probability of ensuring incoherent sampling [120], which is a theoretical requirement for CS. Given that most of the signal energy concentrates around low frequencies, it is recommended to use variable density sampling which tries to mimic the expected energy distribution of the image [84]. This

results in a low resolution version of the image corrupted by an overlaid random aliasing pattern which can be treated as correlated noise. An example is shown in figure 3.11. For this reason, the removal of aliasing in a CS acquisition of MRI has a strong relationship to a denoising operation.

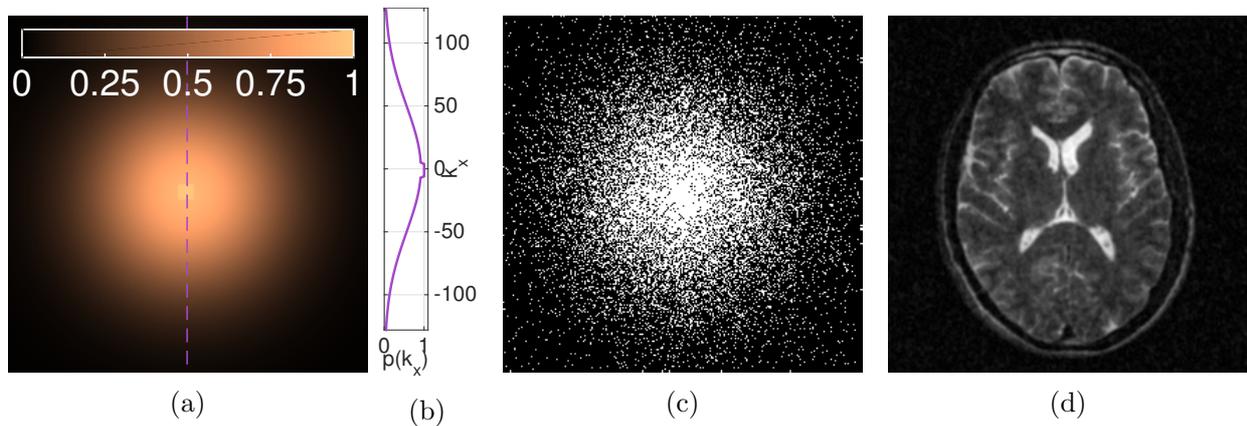


Figure 3.11: Incoherent subsampling of k-space. A variable density distribution is chosen to prioritise the acquisition of low-frequency content (a, b) from which a mask can be obtained (c). The aliasing produced by zero-filling non-acquired samples is shown in (d).

### 3.6.1 Non-adaptive sparsity

The sparsifying transform  $\mathcal{S}(\cdot)$  can take many different forms and generally non-adaptive transforms have initially been preferred for their orthogonality and fast operator properties. In the spatial domain, operators known to have good sparsification properties such as wavelets and TV have been suggested in pioneering methods [84]. The temporal dimension of cardiac cine data is a great advantage for CS acquisitions, given that generally temporal information is more redundant than spatial information. One of the most recurrent sparsifying transforms for the temporal dimension is the temporal Fourier transform, leading to signal representation in x-f domain. In [86], the k-t SPARSE method simultaneously ensures a sparse representation of the reconstruction under a Fourier transform along the temporal dimension and a 2D wavelet transform along space. Results at 7 fold acceleration factors were presented, albeit with some visible aliasing.

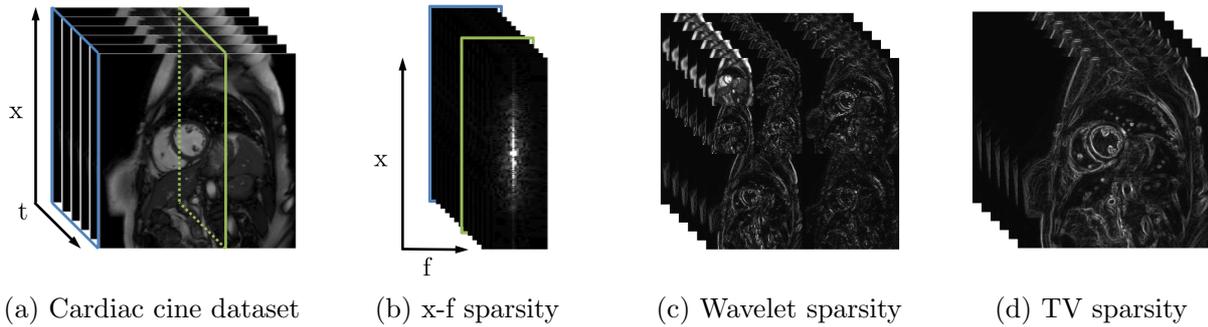


Figure 3.12: A cardiac cine dataset (a) and some examples of non-adaptive sparsity (b, c, d).

Not many researchers have explored changes in the sparsity model that was proposed early on by Lustig. Wavelet transforms are known to be suitable for natural image compression but are certainly not optimal for the sparsification of all signals. Moreover, the smooth temporal change of cardiac cine produces a compressible x-f support, but this transform does not contain locality information about time changes, meaning that individual approximation errors in x-f support spread among all time instances, which is an undesirable property. In [55], Gamper *et al.* developed a stochastic rearrangement of k-space undersampling that produces very low coherent aliasing. Results reported are all based on data sparsification in x-f space, and the achievable acceleration factors were similar to those presented by Lustig with reduced aliasing.

There is a vast list of algorithms that have been proposed for an efficient and more reliable solution of the optimisation problem proposed by Lustig. A good example is k-t FOCUSS [71, 69], where the sparsity model is on the residual of the x-f support relative to the temporal average image. Instead of using a conjugate gradient approach to solve the optimisation problem proposed by Lustig, the FOCUSS algorithm is used and the framework of k-t BLAST/SENSE [136] is linked to the CS problem. On top of being a faster approach, it provides a better approximation of the optimal sparse representation, hence producing better results. Nevertheless, using a temporal average image for the reconstruction can be unreliable and does not allow much flexibility for the undersampling pattern. A later improvement suggests relying on a reference fully sampled image [70], but this will not be discussed as it defies the purpose of CS undersampling. This technique is also adapted to the case of parallel MRI and results show

good quality reconstructions up to 8 fold acceleration factors using a 5 coil parallel acquisition.

One alternative to the x-f support sparsity model has been TV. The TV operator has been suggested in several occasions for the sparsification of medical 2D images with the assumption that they are piecewise constant [16, 124]. This idea was extended to dynamic MRI by Montefusco *et al.* in [97]. Although much of the aliasing can be removed at acceleration factors as high as 16 fold, this comes at the price of inevitable blocky artifacts that distort the reconstruction. This is because the assumption of piecewise constant sequences is very rigid and does not adhere to the characteristics of real sequences. It has nevertheless been reported many times that a TV sparsity constraint is a good auxiliary condition to guide the process towards a suitable reconstruction [84], but relying exclusively on it is restrictive. Also, a major drawback is to enforce the same level of sparsity in all three dimensions of the gradient, because it can be very easily seen empirically that the temporal gradient is often sparser than spatial gradients in cardiac cine images.

The majority of methods proposed assume a batch treatment of the dynamic data, in which the entire acquisition is available prior to reconstruction. Some online methods have also been suggested such as LS-CS [144], where a current observation frame is reconstructed from its undersampled acquisition and an estimation of the previous temporal frame. Such methods however cannot take advantage of the great redundancy offered by the temporal dimension, and clearly produce worse results than batch methods. This option would be desirable in real-time imaging where post-processing must be done online, but the current state of MR scans does not justify disregarding batch methods.

### 3.6.2 Adaptive sparsity

A considerable improvement in sparsity models has been brought by patch-based DL techniques [117]. By imposing patches of the signal in the image domain to be sparsely represented by a dictionary, the representation can benefit from the flexibility of overcompleteness and

redundancy, and the advantages brought by adaptability. Data patches are generally coded independently to save computational cost, and the full reconstruction handled as a two-step process of patch coding and image update. The full problem can be posed as

$$\min_{\bar{\gamma}, \bar{\mathbf{D}}} \|\mathbf{F}_u \bar{\mathbf{D}} \bar{\gamma} - \hat{\mathbf{x}}_u\|_2^2 \quad \text{s.t.} \quad \|\bar{\gamma}\|_0 \leq s, \quad (3.23)$$

where  $\bar{\mathbf{D}}$  and  $\bar{\gamma}$  are composite dictionary and sparse coding vector that concatenate patch contributions. It would be sufficient for  $\mathbf{F}_u$  to be a  $D$ -RIP matrix to satisfy sampling incoherence conditions. Although this has not been proven, results in [24, 115, 56] and empirical results from literature suggest it is a plausible assumption.

The use of DL to reduce sampling requirements in MRI has been proposed for different imaging techniques. In [41], the use of a dictionary is employed for an adaptive sparse representation of T1 and T2 parameter mapping of brain imaging. It has also been applied to directly code image intensity by Ravishankar *et al.* in [117], where an iterative process alternates between the patch-based sparsity in image domain and the consistency of acquired k-space locations. This method was shown to perform better than a fixed basis transform counterpart algorithm on 2D structural MRI.

### Adaptive spatio-temporal sparsity

In this work we explore the use of a single spatio-temporal dictionary to accurately code any time instance of the sequence and allow a CS reconstruction from undersampled data. Additionally, the data representation is regularised with a temporal TV constraint to further exploit the temporal redundancy within the data.

In the course of our investigation, alternative methods for DL applied to dynamic MRI have emerged. In [147], 3D spatio-temporal dictionaries are used in conjunction with a TV in a very similar manner as presented in this work. The authors propose to sequentially update the patch-based dictionary to individual temporal frames, which although desirable, it consid-

erably increases computational complexity. This is an idea that was proposed earlier on for natural video denoising [111]. The update of a dictionary for natural video is critical given that interframe information is expected to change dramatically, but in cardiac cine MR signal characteristics undergo minimal and locally constrained changes.

In [81] the authors consider training temporal signatures of MR data. This is a slightly different approach to dictionary training, but it could be related to the proposed method by assuming single pixel temporal patches of size  $1 \times 1 \times N_t$ , where  $N_t$  is the number of temporal frames in the scan. However, even though the temporal dimension is generally regarded as the one containing the largest redundancy in cardiac cine, spatial redundancy should also be exploited. Moreover, considering a joint spatio-temporal processing is expected to enhance reconstruction homogeneity compared to spatial only or temporal only processing.

## 3.7 Conclusion

In this chapter we have seen how implicit image redundancy, expressed in terms of compressibility, can be exploited to reduce sampling requirements below the traditional Nyquist rate. The non-linear sampling theory of CS imposes three conditions for its application: signal sparsity, sampling incoherence, and non-linear reconstruction. Additionally, we have discussed the transition in sparse representation from complete fixed basis transforms to overcomplete adaptive dictionaries, and the advantages in representation flexibility this entails. Some of the attempts at employing CS and DL for MRI acquisition acceleration have been described, with a particular focus on the case of dynamic cardiac MRI. In the following three chapters, we present our contribution to the application of DL for cardiac cine MRI reconstruction in single and multi-coil setups, and its combination with analysis processes.

# Chapter 4

## Dictionary learning for dynamic MRI

This chapter is based on the following publications:

- J. Caballero, A. N. Price, D. Rueckert, J. V. Hajnal, *Dictionary learning and time sparsity for dynamic MR data reconstruction*, IEEE Transactions on Medical Imaging, vol. 33(4), pp. 979-994, April 2014.
- J. Caballero, A. N. Price, D. Rueckert, J. V. Hajnal, *Dictionary-based reconstruction of dynamic complex MRI data*, Proceedings of the 21st International Society for Magnetic Resonance in Medicine (ISMRM) Annual Meeting and Exhibition, p. 3806, Salt Lake City, UT, USA, 20-26 April 2013.
- J. Caballero, D. Rueckert, J. V. Hajnal, *Dictionary learning and time sparsity for dynamic MRI*, Proceedings of the 15th International Conference on Medical Imaging Computing and Computer Assisted Interventions (MICCAI), vol. 1, pp. 256-263, Nice, France, 1-5 October 2012.

## 4.1 Introduction

As introduced in chapter 3, the developing mathematical field of CS [44, 26] for modern sampling has been shown to be naturally applicable to the problem of MRI acceleration [84]. Central to this approach is the assumption of sparsity of the reconstructed data under some transform domain, and the performance of the reconstruction relies heavily on the suitability of the sparsity model in terms of the approximation error it entails. The potential of CS methods is widely recognised, although exploration of the choice of sparsity model has been somewhat limited. Sparsity is often treated as synonymous with retrospective compressibility and this can lead to crude models that do not adhere well to the characteristics of the data. For high undersampling factors, the reconstruction relies strongly on the chosen a priori sparsity assumption, and models that are not representative will recover images with features that are unexpected in their fully sampled counterparts. Methods based on wavelets or TV [84, 97, 1], for instance, are susceptible of generating blocky spatial artefacts, and very sparse x-f supports will miss out on rapid temporal changes [71, 138]. Adaptive techniques exploiting the developments of DL have brought a step-change in performance of reconstruction from undersampled MRI data by tailoring an overcomplete set of basis functions to provide higher levels of sparsity than those achievable by fixed basis transforms.

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

target signal.

The aim of this study is to analyse the potential of the use of dictionaries in the context of cardiac cine images. To carry out this analysis, synthetic experiments are arranged in which data acquired using parallel coils is retrospectively undersampled and reconstructed assuming a single coil setup. Conclusions drawn from these experiments should therefore not be taken as directly implementable in a real scan scenario, but will provide a simple paradigm for benchmarking the reconstruction capabilities of dictionaries. A practical implementation of DL for MRI should consider the extension of this technique to multi-coil MR technology, and this will be discussed in chapter 5.

This chapter is organised as follows. The two sparsifying transforms exploited in the algorithm proposed are described in section 4.2 and the optimisation problem they pose is stated in section 4.3. In section 4.4 we provide the details necessary for the implementation of the algorithm, which we term dictionary learning with temporal gradient (DLTG), as well as providing comments on its computational complexity. To conclude, the performance of the novel technique is analysed in section 4.5, where it is thoroughly compared with the k-t FOCUSS algorithm, a method that also applies compressed sensing to dynamic cardiac MRI but uses a fixed basis, non-adaptive transform for sparse modelling. The influence of tuning parameters on results is also examined.

## 4.2 Sparsifying transforms for CS dynamic MRI

Throughout the chapter, we will assume a sequence of images can be seen as Fourier transforms of 2D  $N_x \times N_y$  k-space samples acquired at  $N_t$  different time instances that are stacked as a 3D volume. Let us refer to the k-space of a fully sampled data set compliant with the Nyquist sampling rate as the column vector  $\hat{\mathbf{x}}_f \in \mathbb{C}^N$ , produced by the concatenation of columns in the 3D volume where  $N = N_x \times N_y \times N_t$ . The vector sequence in image space  $\mathbf{x}_f \in \mathbb{C}^N$  is related to the k-space data by  $\hat{\mathbf{x}}_f = \mathbf{F}\mathbf{x}_f + \mathbf{n}$ , where  $\mathbf{F}$  performs a 2D DFT on each temporal frame and  $\mathbf{n}$

is additive white Gaussian (AWG) acquisition noise that is complex. Now assume only a subset  $\Omega$  of k-space is acquired, meaning that  $\hat{\mathbf{x}}_u = \mathbf{M}\hat{\mathbf{x}}_f$  is the only data available for reconstruction. The undersampling mask  $\mathbf{M} \in \mathbb{R}^{M \times N}$ ,  $M \ll N$ , contains the rows from the identity matrix that correspond to the samples of  $\hat{\mathbf{x}}_f$  that are in  $\Omega$ . The CS dynamic MRI reconstruction is given by the solution to

$$\min_{\mathbf{x}} \|S(\mathbf{x})\|_0 \quad \text{s.t.} \quad \|\mathbf{F}_u \mathbf{x} - \hat{\mathbf{x}}_u\|_2^2 < \epsilon, \quad (4.1)$$

where we have used  $\mathbf{M}\mathbf{F} = \mathbf{F}_u$  and  $\epsilon$  is a small constant.

There is a direct relationship between the sparsity degree provided by the transform  $\mathcal{S}(\cdot)$  and the minimum number of samples necessary for perfect reconstruction [27]. The sparser the model chosen, the higher the achievable acceleration rates will be. In this section, two sparsifying transforms that are well suited to the problem of dynamic cardiac imaging are described. First, an adaptive patch-based transform derived from DL theory is presented for the case of dynamic MR complex data. Then, the temporal gradient transform is proposed as a suitable global sparsity model for cardiac cine images that can make reconstructions converge faster and improve performance at high undersampling rates.

### 4.2.1 Adaptive spatio-temporal sparsity

DL was introduced in section 3.4 as a process to adapt an initial set of basis functions to a specific signal through a training stage such that it will provide a sparse representation of that signal. Denote training patches  $\mathbf{x}_n^T \in \mathbb{R}^{N_p}$ ,  $n = 1, \dots, T$ , as column vectors of size  $N_p$  to be used for the training of a dictionary  $\mathbf{D} \in \mathbb{R}^{N_p \times K}$  of  $K$  atoms arranged as column vectors. The training of a real-valued dictionary adapted to that training set can be stated as

$$\min_{\mathbf{\Gamma}^T, \mathbf{D}} \|\boldsymbol{\gamma}_n^T\|_0 \quad \text{s.t.} \quad \|\mathbf{x}_n^T - \mathbf{D}\boldsymbol{\gamma}_n^T\|_2^2 \leq \epsilon, \quad \forall n, \quad (\text{DL-}\epsilon)$$

where  $\mathbf{\Gamma}^T \in \mathbb{R}^{K \times T}$  is a matrix gathering the sparse representation of  $\mathbf{x}_n^T$  as column vectors  $\boldsymbol{\gamma}_n^T$ . The superscript  $\mathcal{T}$  specifies that these are variables for training. After this process, the

dictionary can be used to find an approximation of a real-valued data set as a sparse coding problem with a greedy  $l_0$  minimisation technique such as OMP.

Recently, DL has been used for 2D structural MR image reconstruction largely outperforming competing techniques based on complete bases and fixed transforms [117]. This framework is extended to the case of MR sequences in [21] with the use of spatio-temporal 3D dictionaries, but only the reconstruction of synthetic real-valued sequences without a phase component is addressed. This is not feasible in practice since the observed k-space samples always relate to a complex image.

The training of complex-valued dictionaries that are suitable for MR data representation is possible defining  $\mathbf{\Gamma}$  and  $\mathbf{D}$  as complex-valued variables as originally proposed in [117]. Instead, we carry out this representation by using a single real-valued dictionary which is trained on real and imaginary parts of MR data for their independent coding. In section 4.5.5 we look at the differences between these two learning and coding strategies for the processing of MR data.

## 4.2.2 Temporal gradient sparsity

An additional sparsity constraint can be imposed on the temporal finite differences (*i.e.* the first order temporal gradient) of the data set. Many authors have explored TV for imposing sparsity constraints on a CS reconstruction because it provides sensible sparsity levels, but also because its optimisation can be extremely efficient [84, 97, 87, 155]. In many cases, TV is not the main sparsifying transform but rather an auxiliary constraint that can stabilise and correct the solution provided by the main transform. This operation considers an equally weighted combination of the pixel-wise finite differences along space and time, but this is rarely a sensible assumption in cardiac cine images because spatial and temporal gradients, which make up the individual dimensions of TV, will usually have different sparsity levels.

Let us consider a complex-valued single slice cardiac sequence to be a volume  $\mathbf{X}^{3D}$  of entries  $(\mathbf{X}^{3D})_{g,h,i} = X_{g,h,i}$  in the  $x$ ,  $y$ , and  $t$  axes. We denote  $(\nabla_x \mathbf{X}^{3D})_{g,h,i} = X_{g+1,h,i} - X_{g,h,i}$  the finite

difference result of  $\mathbf{X}^{3D}$  along dimension  $x$  and similarly  $(\nabla_y \mathbf{X}^{3D})_{g,h,i} = X_{g,h+1,i} - X_{g,h,i}$  and  $(\nabla_t \mathbf{X}^{3D})_{g,h,i} = X_{g,h,i+1} - X_{g,h,i}$  along dimensions  $y$  and  $t$ . Expressing  $(\nabla_x \mathbf{x})_{g,h,i}$ ,  $(\nabla_y \mathbf{x})_{g,h,i}$  and  $(\nabla_t \mathbf{x})_{g,h,i}$  as equivalent expressions when  $\mathbf{x}$  is a concatenated column vector version of  $\mathbf{X}^{3D}$ , and  $|\mathbf{x}|$  referring to the element-wise absolute value of the complex vector  $\mathbf{x}$ , figure 4.1 shows the decay of magnitude coefficients of the three transforms, ordered and normalised. This implies the signal content is more compactly represented in the temporal gradient domain. Another visualisation of the difference in TV for the three dimensions considered is presented in figure 4.2, where significant coefficients are much more visibly present in spatial dimensions than in the temporal dimension.

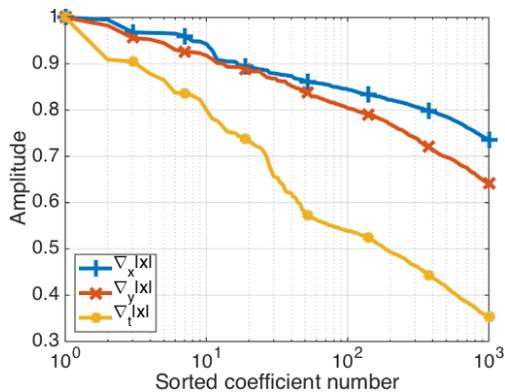


Figure 4.1: Comparison of the decay in transform coefficients of transforms  $\nabla_x$ ,  $\nabla_y$  and  $\nabla_t$ . The latter shows the fastest decay, implying that the signal energy concentrates in fewer coefficients than in the other gradient dimensions.

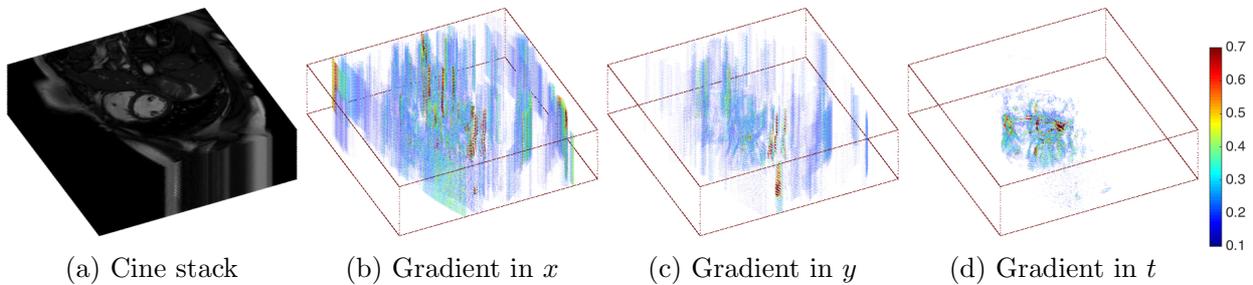


Figure 4.2: Visualisation of significant non-zero coefficients of transforms  $\nabla_x$ ,  $\nabla_y$  and  $\nabla_t$ . Important coefficients in  $\nabla_t$  are concentrated around the dynamic region of the beating heart, whereas changes in other locations are kept below 0.1 and could be attributed to noise.

### 4.3 Reconstruction with sparse model

The problem posed is to simultaneously find a solution data set  $\mathbf{x} \in \mathbb{C}^N$  such that its real and imaginary parts  $\Re(\mathbf{x})$  and  $\Im(\mathbf{x})$  are sparsely represented with a single dictionary  $\mathbf{D}$ , which is to

be learnt. Additionally, we penalise solutions for which  $\|\nabla_t|\mathbf{x}|\|_1$  is large and favour those that are overall consistent with the  $M$  k-space acquisitions  $\mathbf{F}_u\mathbf{x} = \hat{\mathbf{x}}_u$ . This is a similar approach as that in [117], but instead of using 2D spatial dictionaries we use 3D spatio-temporal dictionaries and further exploit temporal redundancy in the cardiac cine data with an auxiliary temporal gradient constraint.

Let us define as  $\mathbf{R}_n \in \mathbb{R}^{N_p \times N}$  an operator that extracts as a column vector the 3D patch in the data set starting from pixel location  $n$ . We use  $\mathbf{\Gamma} = \mathbf{\Gamma}_{\Re} + j\mathbf{\Gamma}_{\Im} \in \mathbb{C}^{N_p \times N}$  to denote the sparse coding of  $\mathbf{x}$ , where each column  $\gamma_n = \gamma_{\Re,n} + j\gamma_{\Im,n}$  in  $\mathbf{\Gamma}$  is the coding of patch  $\mathbf{R}_n\mathbf{x}$ . Throughout the description we assume that the step size for patch extraction is 1 and that patches wrap around edges, meaning that  $n = 1, \dots, N$ , and that each pixel in  $\mathbf{x}$  is covered by  $N_p$  different patches.

The problem can be expressed as the following global statement:

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{\Gamma}_{\Re}, \mathbf{\Gamma}_{\Im}, \mathbf{x}} \sum_{n=1}^N (\|\gamma_{\Re,n}\|_0 + \|\gamma_{\Im,n}\|_0) + \nu \|\mathbf{F}_u\mathbf{x} - \hat{\mathbf{x}}_u\|_2^2 + \mu \|\nabla_t|\mathbf{x}|\|_1 \\ \text{s.t.} \quad \begin{cases} \|\mathbf{R}_n\Re(\mathbf{x}) - \mathbf{D}\gamma_{\Re,n}\|_2^2 < \epsilon, \quad \forall n \\ \|\mathbf{R}_n\Im(\mathbf{x}) - \mathbf{D}\gamma_{\Im,n}\|_2^2 < \epsilon, \quad \forall n \end{cases} \end{aligned} \quad (4.2)$$

This optimisation problem is non-convex, so we opt to split it into three simpler subproblems that are either convex or can efficiently be solved with greedy methods. Alternating the solution of these three subproblems iteratively will yield an approximation to equation (4.2). Without modification, we introduce an auxiliary variable  $\mathbf{x}_{\text{TG}}$  and write

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{\Gamma}_{\Re}, \mathbf{\Gamma}_{\Im}, \mathbf{x}, \mathbf{x}_{\text{TG}}} \sum_{n=1}^N (\|\gamma_{\Re,n}\|_0 + \|\gamma_{\Im,n}\|_0) + \nu \|\mathbf{F}_u\mathbf{x} - \hat{\mathbf{x}}_u\|_2^2 + \mu \|\nabla_t|\mathbf{x}_{\text{TG}}|\|_1 \\ \text{s.t.} \quad \begin{cases} \|\mathbf{R}_n\Re(\mathbf{x}) - \mathbf{D}\gamma_{\Re,n}\|_2^2 < \epsilon, \quad \forall n \\ \|\mathbf{R}_n\Im(\mathbf{x}) - \mathbf{D}\gamma_{\Im,n}\|_2^2 < \epsilon, \quad \forall n \\ \mathbf{x} = \mathbf{x}_{\text{TG}} \end{cases} \end{aligned} \quad (4.3)$$

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

$$\begin{aligned} \min_{\substack{\mathbf{D}, \mathbf{\Gamma}_{\Re}, \mathbf{\Gamma}_{\Im} \\ \mathbf{x}, \mathbf{x}_{\text{TG}}} } \sum_{n=1}^N (\|\gamma_{\Re, n}\|_0 + \|\gamma_{\Im, n}\|_0) + \nu \|\mathbf{F}_u \mathbf{x} - \hat{\mathbf{x}}_u\|_2^2 + \mu \|\nabla_t |\mathbf{x}_{\text{TG}}|\|_1 + \beta \|\mathbf{x} - \mathbf{x}_{\text{TG}}\|_2^2 \\ \text{s.t.} \quad \begin{cases} \|\mathbf{R}_n \Re(\mathbf{x}) - \mathbf{D} \gamma_{\Re, n}\|_2^2 < \epsilon, \quad \forall n \\ \|\mathbf{R}_n \Im(\mathbf{x}) - \mathbf{D} \gamma_{\Im, n}\|_2^2 < \epsilon, \quad \forall n \end{cases} \end{aligned} \quad (4.4)$$

There are four tuning parameters in equation (4.4):  $\nu$ ,  $\mu$ ,  $\beta$  and  $\epsilon$ . These control respectively the consistency with the acquired k-space samples, the temporal gradient sparsity of the dummy variable  $\mathbf{x}_{\text{TG}}$ , the distance of the result  $\mathbf{x}$  with respect to this dummy variable, and the representation accuracy of  $\mathbf{D}$ ,  $\mathbf{\Gamma}_{\Re}$  and  $\mathbf{\Gamma}_{\Im}$ . The last parameter is inversely related to the sparsity allowed in the dictionary representation.

In dynamic cardiac imaging as in other structural MRI modalities, the information of most interest is provided by the magnitude part of the complex signal formed. An appealing approach would therefore be to define the problem such that only a magnitude sequence  $|\mathbf{x}|$  is reconstructed discarding the reconstruction of its phase information. This is simpler than the problem posed in equation (4.4) because the number of unknowns is reduced by a half and the system of equations becomes better determined, which is why some solutions proposed have adopted this approach [6, 22]. However, this problem statement overlooks the nature of the observed data  $\hat{\mathbf{x}}_u$ . The samples acquired in k-space always correspond in practice to a complex image, so trying to infer the k-space of a magnitude signal from an undersampled version of its complex representation is not a viable option. Alternatively, the reconstructed  $\mathbf{x}$  has to be complex if it is being inferred from its undersampled k-space, irrespective of the clinical usefulness of its magnitude and phase information.

## 4.4 DLTG algorithm

Equation (4.4) is broken down into three simpler subproblems. The dictionary learning with temporal gradient (DLTG) algorithm iteratively refines a solution in three separate steps that each solve the global problem with some free variables fixed.

### 4.4.1 Dictionary training and sparse coding

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

$$\min_{\mathbf{D}, \mathbf{\Gamma}_{\Re}, \mathbf{\Gamma}_{\Im}} \sum_{n=1}^N (\|\gamma_{\Re, n}\|_0 + \|\gamma_{\Im, n}\|_0) \quad \text{s.t.} \quad \begin{cases} \|\mathbf{R}_n \Re(\mathbf{x}) - \mathbf{D} \gamma_{\Re, n}\|_2^2 < \epsilon, \forall n \\ \|\mathbf{R}_n \Im(\mathbf{x}) - \mathbf{D} \gamma_{\Im, n}\|_2^2 < \epsilon, \forall n \end{cases}. \quad (4.5)$$

This is the general dictionary learning problem, where the training data are patches  $\mathbf{R}_n \Re(\mathbf{x})$  and  $\mathbf{R}_n \Im(\mathbf{x})$ . The K-SVD algorithm [2] can solve equation (4.5). For an undersampling mask providing sufficient aliasing incoherence, aliasing will have noise-like properties that the trained dictionary will not be able to approximate with a sparse representation. Therefore, a sparse approximation of the corrupted sequence  $\mathbf{x}$  using the outcome dictionary from equation (4.5) will tend to reproduce anatomical structure and miss features derived from aliasing.

In practice, a reduced number of patches are used as training data for efficiency purposes extracted from a regular grid on  $\Re(\mathbf{x})$  and  $\Im(\mathbf{x})$ . Once a dictionary  $\mathbf{D}$  has been learnt from the training set, it can be used to code real and imaginary parts of the entire data set independently with OMP:

$$\min_{\mathbf{\Gamma}_{\Re}} \sum_{n=1}^N \|\gamma_{\Re, n}\|_0 \quad \text{s.t.} \quad \|\mathbf{R}_n \Re(\mathbf{x}) - \mathbf{D} \gamma_{\Re, n}\|_2^2 < \epsilon, \forall n, \quad (4.6)$$

$$\min_{\mathbf{\Gamma}_{\Im}} \sum_{n=1}^N \|\gamma_{\Im, n}\|_0 \quad \text{s.t.} \quad \|\mathbf{R}_n \Im(\mathbf{x}) - \mathbf{D} \gamma_{\Im, n}\|_2^2 < \epsilon, \forall n. \quad (4.7)$$

### 4.4.2 Temporal gradient filtering

Let us now fix variables  $\mathbf{D}$ ,  $\mathbf{\Gamma}_{\Re}$ ,  $\mathbf{\Gamma}_{\Im}$  and  $\mathbf{x}$ , and minimise the functional with respect to  $\mathbf{x}_{\text{TG}}$ .

The problem becomes

$$\min_{\mathbf{x}_{\text{TG}}} \|\nabla_t |\mathbf{x}_{\text{TG}}|\|_1 + \eta \|\mathbf{x} - \mathbf{x}_{\text{TG}}\|_2^2, \quad (4.8)$$

with  $\eta = \beta/\mu$ . The magnitude of the dummy variable  $\mathbf{x}_{\text{TG}}$  is driven towards sparsity in its temporal gradient while minimising the quadratic distance with respect to solution  $\mathbf{x}$ . The phase difference between  $\mathbf{x}$  and  $\mathbf{x}_{\text{TG}}$  will not alter the first term in equation (4.8), so we fix the phase component  $\angle \mathbf{x}_{\text{TG}} = \angle \mathbf{x}$  and solve

$$\min_{|\mathbf{x}_{\text{TG}}|} \|\nabla_t |\mathbf{x}_{\text{TG}}|\|_1 + \eta \|\mathbf{x} - |\mathbf{x}_{\text{TG}}|\|_2^2. \quad (4.9)$$

This can be very efficiently solved by the primal-dual method in [29] with an iterative clipping algorithm. Specifically, the following two computations are iterated:

$$|\mathbf{x}_{\text{TG}}|^{(i+1)} = |\mathbf{x}| - \nabla_t^T \mathbf{z}^{(i)}, \quad (4.10)$$

$$\mathbf{z}^{(i+1)} = \text{clip} \left( \mathbf{z}^{(i)} + \frac{1}{\alpha} \nabla_t |\mathbf{x}_{\text{TG}}|^{(i+1)}, \frac{1}{2\eta} \right), \quad (4.11)$$

for  $i \geq 0$ ,  $\mathbf{z}^{(0)} = \mathbf{0}$  and  $\alpha \geq \text{maxeig}(\nabla_t \nabla_t^T)$ , where  $\text{maxeig}(\cdot)$  finds the maximum eigenvalue of a matrix. For details on the derivation of this algorithm we refer to [29].

### 4.4.3 Acquisition data consistency

The last subproblem looks at the case where  $\mathbf{x}$  is the only free variable. Equation (4.4) now becomes

$$\min_{\mathbf{x}} \|\mathbf{F}_u \mathbf{x} - \hat{\mathbf{x}}_u\|_2^2 + \frac{\beta}{\nu} \|\mathbf{x} - \mathbf{x}_{\text{TG}}\|_2^2 \quad \text{s.t.} \quad \begin{cases} \|\mathbf{R}_n \Re(\mathbf{x}) - \mathbf{D} \boldsymbol{\gamma}_{\Re, n}\|_2^2 < \epsilon & \forall n \\ \|\mathbf{R}_n \Im(\mathbf{x}) - \mathbf{D} \boldsymbol{\gamma}_{\Im, n}\|_2^2 < \epsilon & \forall n \end{cases}. \quad (4.12)$$

This can be rewritten as the following unconstrained optimisation problem:

$$\min_{\mathbf{x}} \sum_{n=1}^N (\|\mathbf{R}_n \Re(\mathbf{x}) - \mathbf{D}\gamma_{\Re,n}\|_2^2 + \|\mathbf{R}_n \Im(\mathbf{x}) - \mathbf{D}\gamma_{\Im,n}\|_2^2) + \nu' \|\mathbf{F}_u \mathbf{x} - \hat{\mathbf{x}}_u\|_2^2 + \beta' \|\mathbf{x} - \mathbf{x}_{\text{TG}}\|_2^2. \quad (4.13)$$

Referring to this functional as  $f$ , the optimum solution is the one for which  $\frac{\partial f}{\partial \Re(\mathbf{x})} = \frac{\partial f}{\partial \Im(\mathbf{x})} = 0$ .

Let us solve for  $\frac{\partial f}{\partial \Re(\mathbf{x})} = 0$ :

$$\left( \sum_{n=1}^N \mathbf{R}_n^T \mathbf{R}_n + \nu' \mathbf{F}_u^H \mathbf{F}_u + \beta' \right) \Re(\mathbf{x}) = \sum_{n=1}^N \mathbf{R}_n^T \mathbf{D}\gamma_{\Re,n} + \nu' \mathbf{F}_u^H \Re(\hat{\mathbf{x}}_u) + \beta' \Re(\mathbf{x}_{\text{TG}}). \quad (4.14)$$

Taking the Fourier transform on both sides of equation (4.14), we have

$$\begin{aligned} & \left( \mathbf{F} \sum_{n=1}^N \mathbf{R}_n^T \mathbf{R}_n \mathbf{F}^H + \nu' \mathbf{F} \mathbf{F}_u^H \mathbf{F}_u \mathbf{F}^H + \beta' \mathbf{F} \mathbf{F}^H \right) \mathbf{F} \Re(\mathbf{x}) \\ &= \mathbf{F} \sum_{n=1}^N \mathbf{R}_n^T \mathbf{D}\gamma_{\Re,n} + \nu' \mathbf{F} \mathbf{F}_u^H \Re(\hat{\mathbf{x}}_u) + \beta' \mathbf{F} \Re(\mathbf{x}_{\text{TG}}), \end{aligned} \quad (4.15)$$

Replacing  $\Re$  by  $\Im$  in equation (4.15) we obtain the solution for the k-space of the imaginary part, hence we can write the k-space solution of complex variable  $\mathbf{x}$  as

$$\begin{aligned} & \left( \mathbf{F} \sum_{n=1}^N \mathbf{R}_n^T \mathbf{R}_n \mathbf{F}^H + \beta' \mathbf{F} \mathbf{F}^H + \nu' \mathbf{F} \mathbf{F}_u^H \mathbf{F}_u \mathbf{F}^H \right) \hat{\mathbf{x}} \\ &= \mathbf{F} \sum_{n=1}^N \mathbf{R}_n^T \mathbf{D}(\gamma_{\Re,n} + j\gamma_{\Im,n}) + \beta' \mathbf{F} \mathbf{x}_{\text{TG}} + \nu' \mathbf{F} \mathbf{F}_u^H \hat{\mathbf{x}}_u. \end{aligned} \quad (4.16)$$

To better understand this expression we can use the following simplifications. The term  $\mathbf{F} \mathbf{F}_u^H \hat{\mathbf{x}}_u$  is the zero-filled k-space acquisition which will be denoted as  $\hat{\mathbf{x}}_z$ . Assuming patches overlap and that the operator  $\mathbf{R}_n$  wraps around the boundaries such that each pixel is represented by  $N_p$  patches, the term  $\mathbf{F} \sum_{n=1}^N \mathbf{R}_n^T \mathbf{R}_n \mathbf{F}^H$  is simply the weighted  $N \times N$  identity matrix  $N_p \mathbf{I}_N$ . Moreover,  $\mathbf{F} \mathbf{F}_u^H \mathbf{F}_u \mathbf{F}^H$  is equivalent to a  $N \times N$  diagonal matrix containing a 1 in the diagonal whenever a k-space location was acquired or a 0 otherwise. The simplifications  $\hat{\mathbf{x}}_{\text{TG}} = \mathbf{F} \mathbf{x}_{\text{TG}}$

and  $\mathbf{I}_N = \mathbf{F}\mathbf{F}^H$  are trivial, and also notice that the expression  $\mathbf{F} \frac{\sum_{n=1}^N \mathbf{R}_n^T \mathbf{D} \gamma_n}{N_p}$  represents the Fourier transform of the solution obtained by averaging the contribution of all coded patches relocated to their corresponding position within the data set. For simplicity we refer to this solution as  $\mathbf{x}_D$ , and  $\hat{\mathbf{x}}_D$  is its Fourier representation. The final expression for  $\hat{\mathbf{x}}$  is therefore

$$\hat{\mathbf{x}}(k) = \begin{cases} \frac{\hat{\mathbf{x}}_D(k) + \frac{\beta'}{N_p} \hat{\mathbf{x}}_{\text{TG}}(k) + \frac{\nu'}{N_p} \hat{\mathbf{x}}_z(k)}{1 + \frac{\beta'}{N_p} + \frac{\nu'}{N_p}}, & k \in \Omega, \\ \frac{\hat{\mathbf{x}}_D(k) + \frac{\beta'}{N_p} \hat{\mathbf{x}}_{\text{TG}}(k)}{1 + \frac{\beta'}{N_p}}, & k \notin \Omega. \end{cases} \quad (4.17)$$

The update of solution  $\mathbf{x}$  involves an average between the dictionary sparse solution  $\hat{\mathbf{x}}_D$ , the temporal gradient sparse solution  $\hat{\mathbf{x}}_{\text{TG}}$ , and the original acquisitions  $\hat{\mathbf{x}}_z$  for  $k$ -space locations that were acquired. This means that this step requires the tuning of parameters  $\nu'$  and  $\beta'$ . To simplify this task, in the implementation of the algorithm we choose to update  $\hat{\mathbf{x}}$  using only  $\hat{\mathbf{x}}_z$  and either  $\hat{\mathbf{x}}_D$  or  $\hat{\mathbf{x}}_{\text{TG}}$  with a single constant  $\lambda$  depending on which solution was updated last. Equation (4.17) then becomes

$$\hat{\mathbf{x}}(k) = \begin{cases} \frac{\hat{\mathbf{x}}'(k) + \lambda \hat{\mathbf{x}}_z(k)}{1 + \lambda}, & k \in \Omega, \\ \hat{\mathbf{x}}'(k), & k \notin \Omega, \end{cases} \quad (4.18)$$

with  $\hat{\mathbf{x}}'$  alternating between  $\hat{\mathbf{x}}_D$  and  $\hat{\mathbf{x}}_{\text{TG}}$ . Noise standard deviation is taken into account by the regularisation parameter  $\lambda = q/\sigma$ , where  $q$  is a constant that can be set empirically as is shown in section 4.5.3.

#### 4.4.4 Algorithm design

Figure 4.3 describes the ordering in which these three steps are performed in the DLTG algorithm. The data consistency step is interleaved between the other two steps and the temporal gradient (TG) module is iterated in an inner loop fashion to force a slow and smooth convergence towards a result jointly satisfying the sparse temporal gradient and the data consistency

constraints. Additionally, a DLMRI version of the method in which the TG term is ignored (equivalent to setting  $\beta = \beta' = 0$ ) is also presented. Notice that this differs from the DLMRI method presented in [117] in that real and imaginary parts are coded independently with a real-valued dictionary.

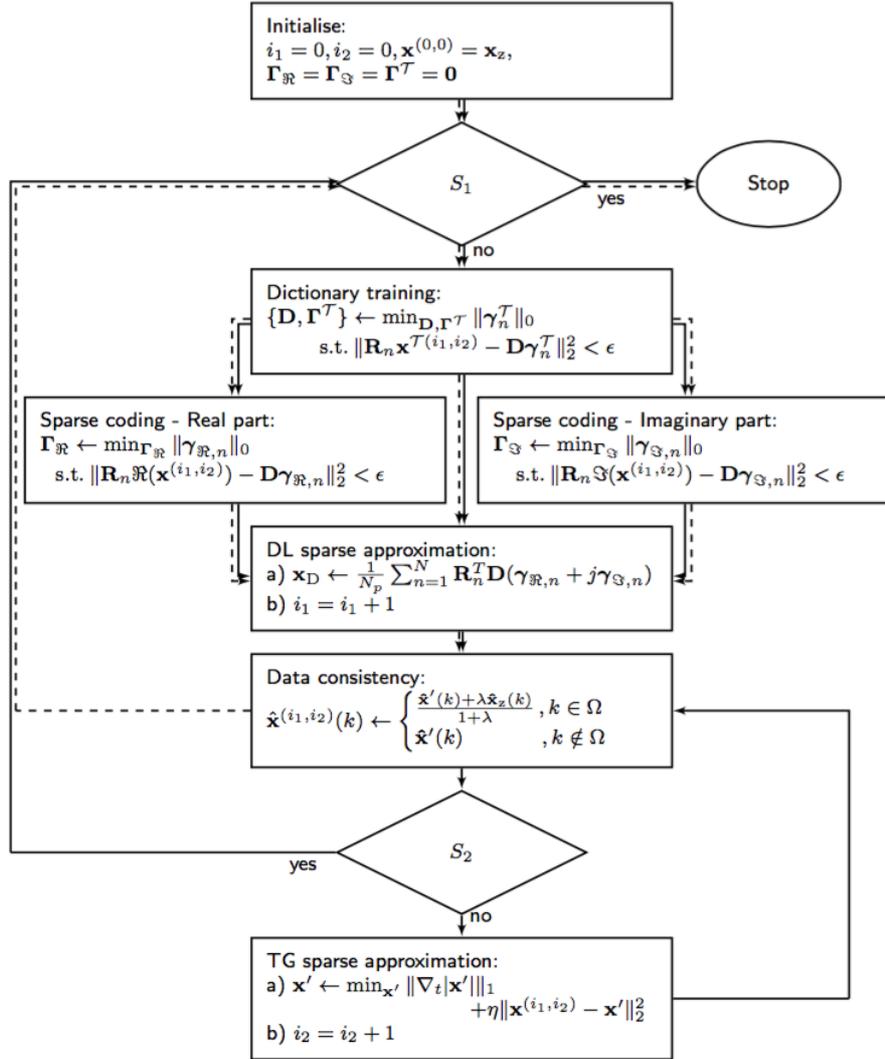


Figure 4.3: Algorithm flowchart for the DLTG (solid arrows) and the DLMRI (dashed arrows) algorithms.

The algorithm is initialised with the zero-filled sequence  $\mathbf{x}_z$  and the stopping criteria  $S_1$  and  $S_2$  can either be convergence to a stable solution or a maximum number of iterations  $\mathcal{I}_1$  and  $\mathcal{I}_2$  for outer and inner loops respectively. We use  $\mathcal{I}_2 = 10$  and a large number of outer iterations  $\mathcal{I}_1 > 100$  to guarantee the end result reaches a stable solution. All the results below use small

values for constants  $\epsilon$  and  $\eta$  to ensure a slow convergence towards a result tightly adhering to the sparsity model imposed. The dictionary in the training stage is initialised at each iteration with an overcomplete DCT dictionary and coding matrices  $\mathbf{\Gamma}_{\mathfrak{R}}$  and  $\mathbf{\Gamma}_{\mathfrak{S}}$  are initially  $\mathbf{0}$ .

#### 4.4.5 Algorithm complexity

The complexity of the DLMRI algorithm, which constitutes the first half of the DLTG algorithm, is dominated by the dictionary learning and sparse coding steps. The training of the dictionary and sparse coding of patches are performed using the efficient implementations of the K-SVD and Batch-OMP algorithms made available by R. Rubinstein [119, 113]. In both cases, the computation is dominated by the cost of sparse coding a patch of size  $N_p$  with a dictionary of  $K$  atoms. The number of operations necessary for this can be described as a function of  $s$  OMP iterations. The stopping criterion described in section 4.3 for sparse coding is the data consistency level  $\epsilon$ , and although the relationship between  $s$  and  $\epsilon$  is non-trivial because of its non-linearity, it is easily observed that a larger  $\epsilon$  implies fewer OMP iterations for a given problem.

Batch-OMP precomputes the matrix  $\mathbf{D}^T\mathbf{D}$  to save processing time when coding large sets of patches with the same dictionary. This first step requires  $N_p K^2$  operations. The coding of a real-valued patch that requires  $s$  OMP iterations with Batch-OMP involves approximately  $2N_p K + s^2 K + s^3$  operations [119]. Given that we code a total of  $N$  patches twice for each data set, we can provide the average number of operations per data set coding as  $N_p K^2 + 2N(2N_p K + \bar{s}^2 K + \bar{s}^3)$ , where  $\bar{s}$  is the average number of OMP iterations for coding real and imaginary parts. Also, a K-SVD training iteration for  $N_T$  training patches will involve approximately  $N_T(2N_p K + s^2 K + s^3)$  operations for  $s$  OMP iterations [119], so on average we have  $I_T N_T(2N_p K + \bar{s}^2 K + \bar{s}^3)$  operations. The data consistency step only involves two DFTs per temporal frame and averaging operations that are negligible compared to the sparse coding step. The extra computation for solving equation (4.8) in the DLTG case is also very small since only sparse matrix-vector multiplications and clipping operations are required. In section 4.5.7

we provide empirical results on the runtime of the algorithm.

## 4.5 Experiments and results

### 4.5.1 Experimental setup

Fully sampled short-axis cardiac cine scans were acquired from 10 subjects for the analysis of the proposed method. All data sets contain 30 temporal frames of size  $256 \times 256$  (*i.e.*  $N = 256 \times 256 \times 30$ ) with a  $320 \times 320$  mm<sup>2</sup> field of view and 10 mm slice thickness, and were generated using an optimal combination of 32-channel data. For the 10 subjects the mean  $\pm$  standard deviation for heart rate was  $62 \pm 10.2$  bpm, with the 30 frames giving a temporal rate of  $33 \pm 5.5$  ms. The raw multi-coil data was reconstructed using SENSE [112] with no k-space undersampling and retrospective gating. Coil sensitivity maps were normalised to a body coil image to produce a single complex-valued image set that could then either be back-transformed to regenerate complex k-space samples or further processed to form final magnitude images.

These scans contain unavoidable k-space acquisition noise and this formally precludes quantitative comparison between a given reconstruction and the fully sampled data set. Nonetheless, for the purposes of evaluation we will treat the fully sampled data as ground truth (*i.e.* treat them as if noiseless). To assess noisy scenarios we then artificially add noise to k-space. The quality of all reconstructions  $\mathbf{x}$  will be measured with  $\text{PSNR}(\mathbf{x}) = 10 \log \left( \frac{1}{\|\mathbf{x}_f - \mathbf{x}\|_2^2 / N} \right)$ , where  $\mathbf{x}_f$  denotes the fully sampled data set. An example of a data set frame is shown in figure 4.4, where we highlight a ROI and a vertical line, which will respectively serve for evaluations within a tight region around the heart and the temporal profile of the line. For reconstructions without artificial noise, visual inspection of magnitude and phase information as well as the mean structural similarity index (MSSIM) [148] will also be considered, which provides a measure of similarity that highlights structural information perception.

Throughout this section, fully sampled k-space data is artificially subsampled using 2D binary

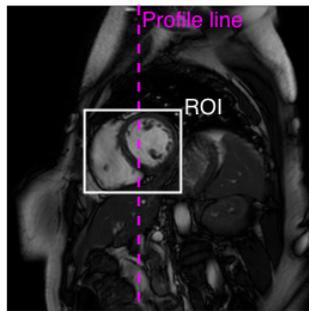


Figure 4.4: Example of 2D temporal frame of one of the data sets used for testing. A ROI and a profile line are defined that will be used for evaluations.

undersampling masks  $\mathbf{M}$ . We only consider the case of Cartesian undersampling, which is the most common in practice, although more elaborate sampling strategies like radial or spiral [84, 16] could equally be applicable with modifications on the data consistency step and using non-uniform Fourier transforms [52]. Even though greater aliasing incoherence can be achieved with 2D k-space undersampling [84] frequency encodes can be considered instantaneous relative to phase encodes, so acceleration is only meaningful through phase encode undersampling.

It has been claimed in the past that drawing independent realisations of this random experiment for each temporal frame significantly increases eddy currents. In [11], this issue is remedied by pairing consecutive random phase encodes. At each frame, the 8 lowest spatial frequencies are always acquired and other frequencies have a probability of being acquired determined by a Gaussian variable density function that is marginally offset with an added bias, such that the probability of acquisition never reaches zero even at the highest frequencies. An implementation of this approach can be found in [12], and an example of a 2D mask and its effect on the magnitude of a temporal frame is shown in figure 4.5 for a 6 fold acceleration.

The proposed method is compared to k-t FOCUSS [71] as a representative algorithm of the CS dynamic MRI methods using fixed basis sparsifying transforms. The implementation of this algorithm is publicly available in [12]. A single regularisation parameter can be tuned to trade between a k-space data consistency term and an x-f domain sparsity term, and even though this parameter cannot be optimally determined *a priori*, we sweep across a large spectrum of values in all experiments and only show the best reconstruction in order to always compare to the best possible scenario.

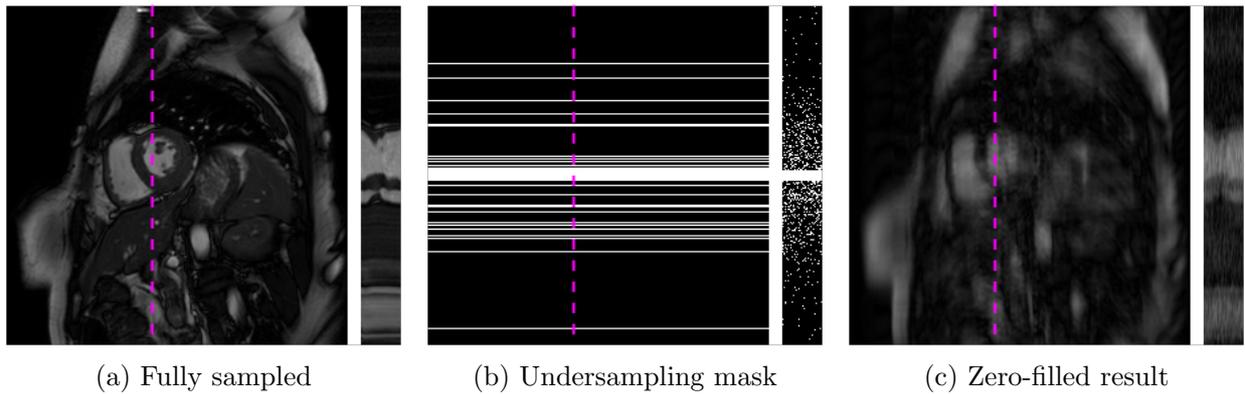


Figure 4.5: Example of a magnitude temporal frame from one of the data sets analysed (a). The undersampling mask (b) applied in k-space reduces acquisition time but introduces aliasing in image space (c). All figures show a 2D frame on the left hand side and the temporal profile across the dashed line on the right hand side.

In contrast, DLTG results are not optimised individually and, unless otherwise stated, its parameters are kept constant across all experiments. We use  $N_T = 10^4$  patches of size  $N_p = 4 \times 4 \times 4$  to train dictionaries of  $K = 600$  atoms. These parameters were chosen based on empirical trade-offs between performance and efficiency. The regularisation parameters for dictionary approximation accuracy and temporal gradient sparsity (see equations (4.5) and (4.8)) are  $\epsilon = 0.007$  and  $\eta = 4 \times 10^{-4}$ , which allows to analyse the potential of the model by tightly adhering the result to it as will be shown in 4.5.7.

## 4.5.2 Reconstruction of individual data sets

The first experiment considers the simple case where data sets are undersampled and reconstructed with different acceleration factors without added noise. Denote  $R$  the acceleration factor, figure 4.6a plots the mean PSNR of the reconstructions and one standard deviation away from them against the sampling ratio ( $1/R$ ) for the 10 subjects considered. For this particular case, the noise regularisation factor is set to  $\lambda \rightarrow \infty$ , such that in the DLMRI and DLTG algorithms all k-space samples that are acquired are constantly fed back to the reconstruction's k-space in the data consistency step without any weighting.

Even though it is ideally optimised, the k-t FOCUSS algorithm is outperformed by the DLMRI

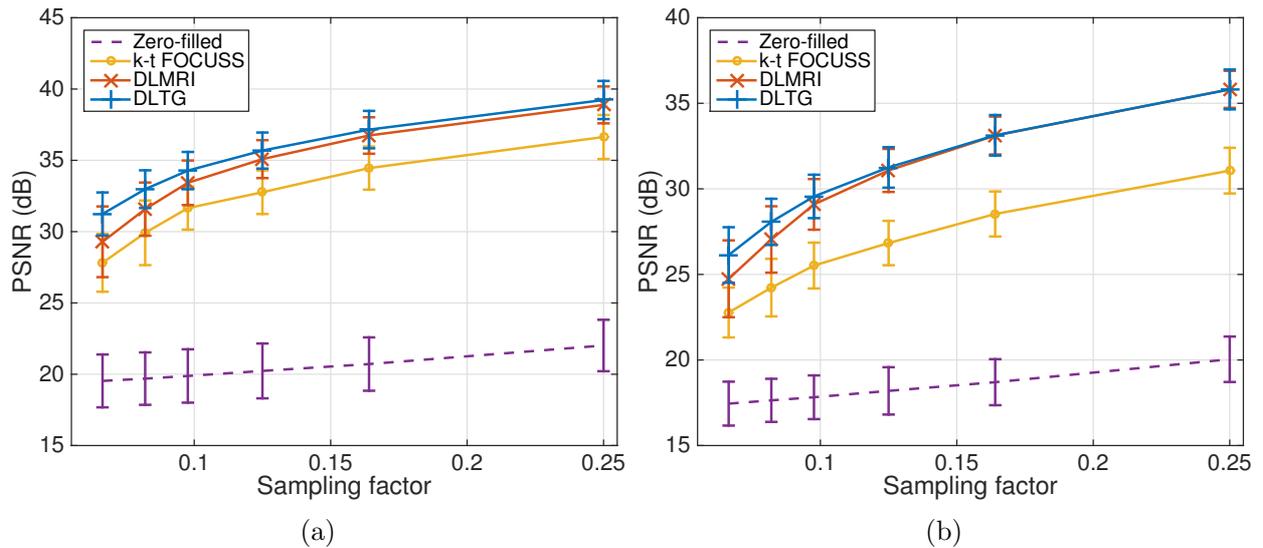


Figure 4.6: Mean and standard deviation PSNR performance of reconstructions from 10 scans retrospectively undersampled without added noise. Results are shown for the whole data sets (a) and within a ROI around the heart (b).

and DLTG algorithms in all experiments. A further comparison made in table 4.1 using the MSSIM metric supports this conclusion. MSSIM can quantify structural similarity between two images better than PSNR by comparing luminance and contrast in patches. For our results we consider the mean MSSIM of the 30 temporal frames in each data set. Although the standard deviations of the PSNR and MSSIM results are relatively large, the rank order for both figures of merit for the three methods and for each individual reconstruction was  $DLTG > DLMRI > k-t$  FOCUSS.

Table 4.1: MSSIM comparison (mean  $\pm$  std )  $\times 10^{-2}$ .

Sampling factor $1/R$	k-t FOCUSS	DLMRI	DLTG
0.25	$95.8 \pm 2.5$	$97.3 \pm 0.8$	$97.4 \pm 1.0$
0.16	$93.9 \pm 3.4$	$96.2 \pm 0.9$	$96.5 \pm 1.1$
0.12	$91.4 \pm 6.0$	$95.0 \pm 1.1$	$95.5 \pm 1.1$
0.10	$89.9 \pm 5.4$	$93.3 \pm 1.6$	$94.4 \pm 1.3$
0.08	$86.2 \pm 6.3$	$91.2 \pm 1.8$	$93.1 \pm 1.3$
0.06	$80.5 \pm 7.4$	$87.7 \pm 3.0$	$91.0 \pm 1.4$

The improvement with respect to k-t FOCUSS is also evident from the visual comparison shown in figure 4.7 of a frame from a data set that has been accelerated by 8. It is nevertheless harder to visually identify the benefit of enforcing TG sparsity in DLTG relative to the DLMRI

method. To better understand in what way the TG term modifies the solution of DLMRI, we compare in figure 4.6b the same measure when only the ROI around the heart shown in figure 4.4 is considered. For this region, reconstruction performances of the DLMRI and the DLTG algorithms are almost the same except at very high acceleration rates.

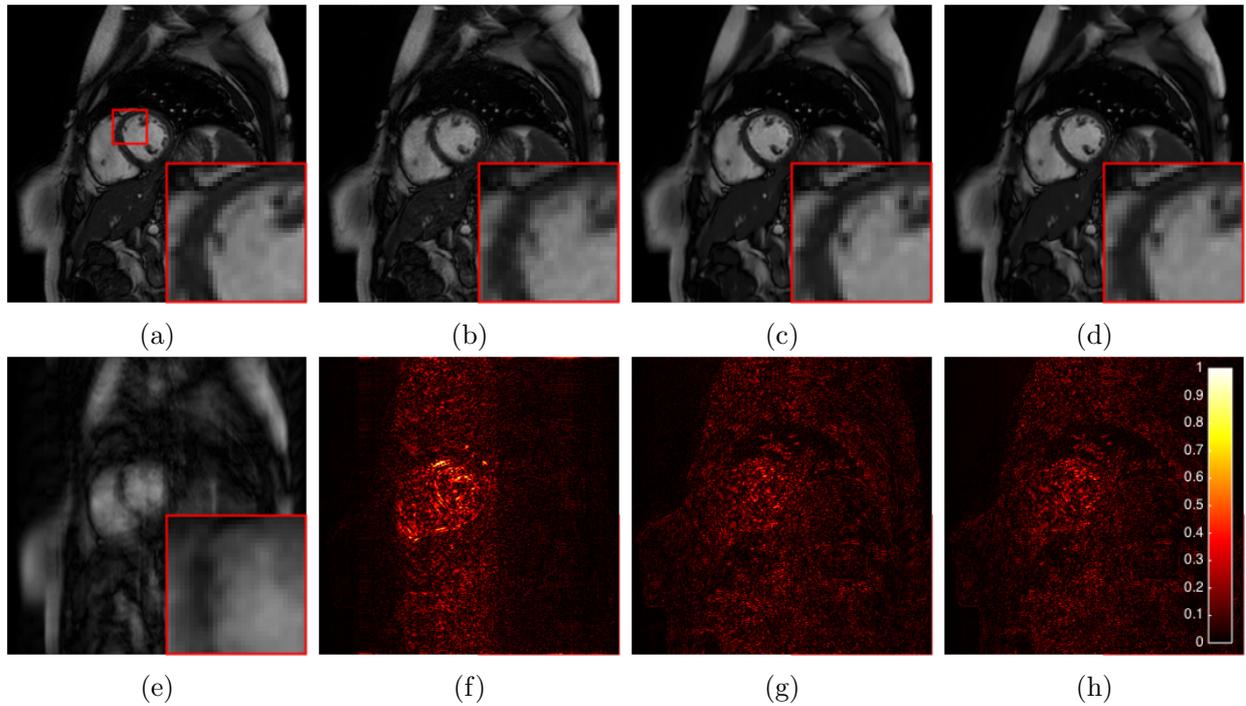


Figure 4.7: Visual comparison of a fully sampled magnitude frame (a), its undersampled by 8 zero-filled version (e), and reconstructions using k-t FOCUSS (b), DLMRI (c) and DLTG (d) with their respective errors multiplied by 6 (f, g, h).

A visual comparison of the data set profile in figure 4.8 along the vertical dashed line of figure 4.4 can help explain this finding. The DLTG method slightly smooths the reconstruction along time in regions of high motion as a result of the TG sparsity enforcement. A small improvement can be appreciated in the static regions outside the ROI (see arrow), which is part of the reason for the improvement in the global PSNR and MSSIM metrics. We could argue that down to sampling ratios of 0.12, the DLMRI method is already able to recover a faithful representation of the signal within the ROI and therefore imposing an extra sparsity penalty does not improve results considerably. However, at high acceleration factors the improvement in the reconstruction quality is evident inside and outside the ROI as is shown in figure 4.6.

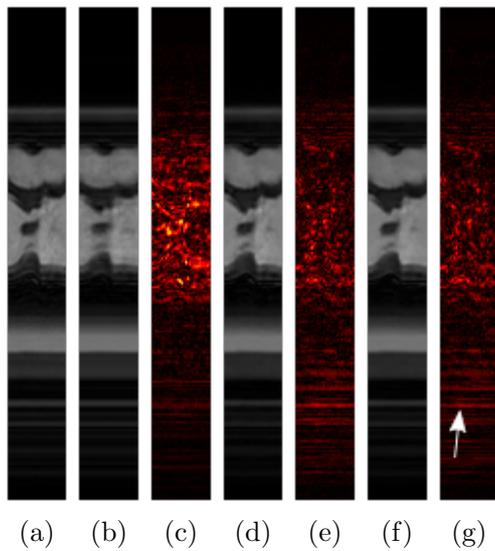


Figure 4.8: Temporal profile of line shown in figure 4.4 in the fully sampled data set (a), and reconstructions from a 8 fold acceleration using k-t FOCUSS (b), DLMRI (d) and DLTG (f) with their respective errors amplified by 6 (c, e, g).

Extremely low sampling factors inevitably degrade results and are a difficult regime in which to carry out a CS based reconstruction, but they force the algorithm to rely heavily on the sparsity model it assumes from the data and can help analyse its suitability. Figure 4.9 shows the reconstruction comparison with an acceleration rate of 15, which is the lowest sampling factor tested. Although fine details are lost with the three methods compared, the DLTG method overall provides the most satisfactory reconstruction. The k-t FOCUSS solution contains a lot of aliasing artefacts even though its optimal sparsity model was unrealistically optimised a posteriori. The DLMRI method presents a more blocky reconstruction and distorts the natural shape of the heart especially around the myocardium. The DLTG method also contains blocky artefacts, but it noticeably eliminates aliasing and preserves the coarse structure of the original frame.

Figure 4.10 compares the temporal profile of the reconstruction at an acceleration rate of 12. The k-t FOCUSS method is not able to capture the dynamism well, and this is a direct consequence of the sparsity model it uses. Sparsity in k-t FOCUSS is imposed on the Fourier transform of pixels along the temporal dimension. If the reconstruction relies heavily on a reconstruction that is too sparse, it will only be able to capture very coarse movement, but fine temporal details will be missed as they are disregarded by the sparse model (see arrow A). The DLMRI result is able to better recover fine temporal changes, but contains a considerable

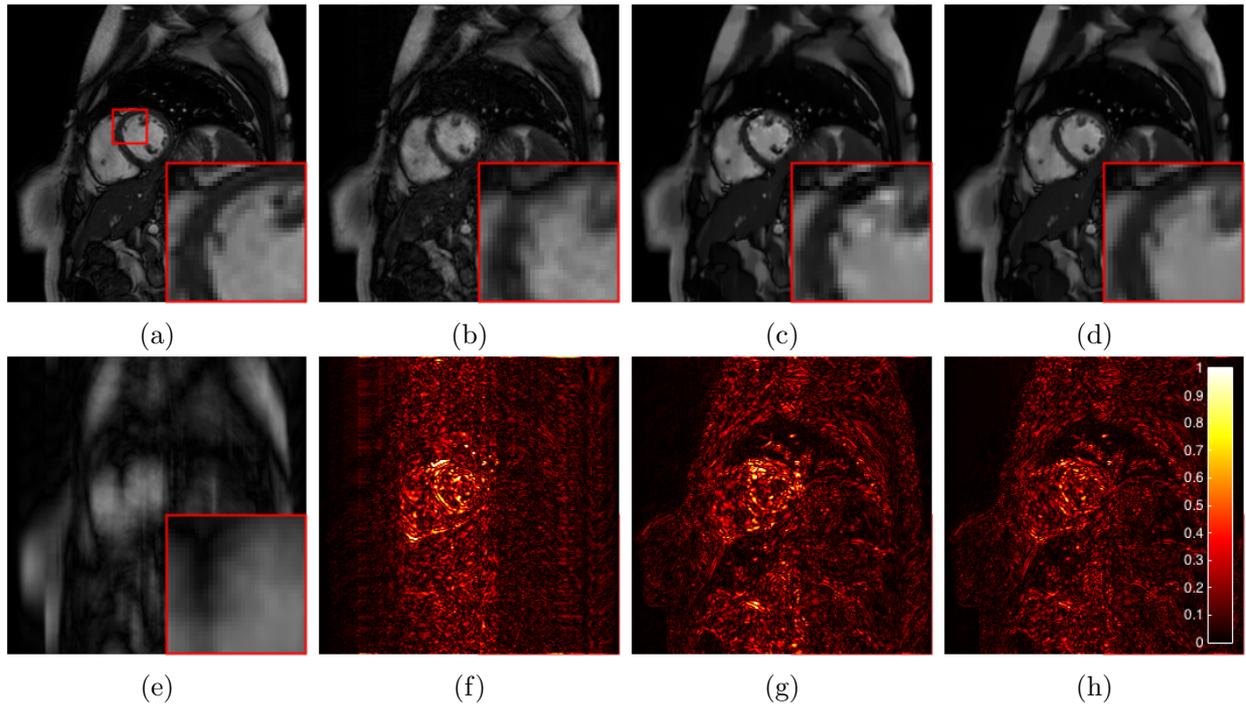


Figure 4.9: Visual comparison of a fully sampled magnitude frame (a), its undersampled by 15 zero-filled version (e), and reconstructions using k-t FOCUSS (b), DLMRI (c) and DLTG (d) with their respective errors multiplied by 6 (f, g, h).

amount of alias that is unresolved (see arrow B). Using the additional TG penalty removes the aliasing at the expense of slightly smoothing the reconstruction along time, as shows the DLTG result. Despite this, it is structurally the most faithful reconstruction out of the comparison.

To conclude the analysis, we show in figure 4.11 the original phase and the reconstructed versions of a temporal frame from a data set accelerated by 8. Except in locations where the amplitude is very low and hence phase is unstable (masked out in the fully sampled case), the three reconstructions are accurate. Quantitatively evaluating phase reconstruction only is difficult because of the many random values when magnitude is zero, but this is implicitly accounted for in figure 4.6.

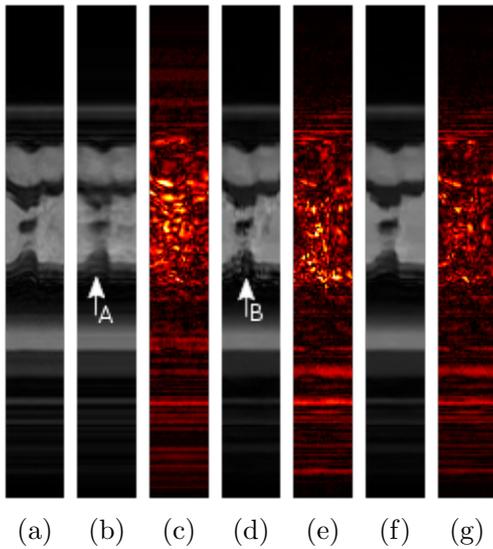


Figure 4.10: Temporal profile of line shown in figure 4.4 in the fully sampled data set (a), and reconstructions from a 12 fold acceleration using k-t FOCUSS (b), DLMRI (d) and DLTG (f) with their respective errors amplified by 6 (c, e, g).

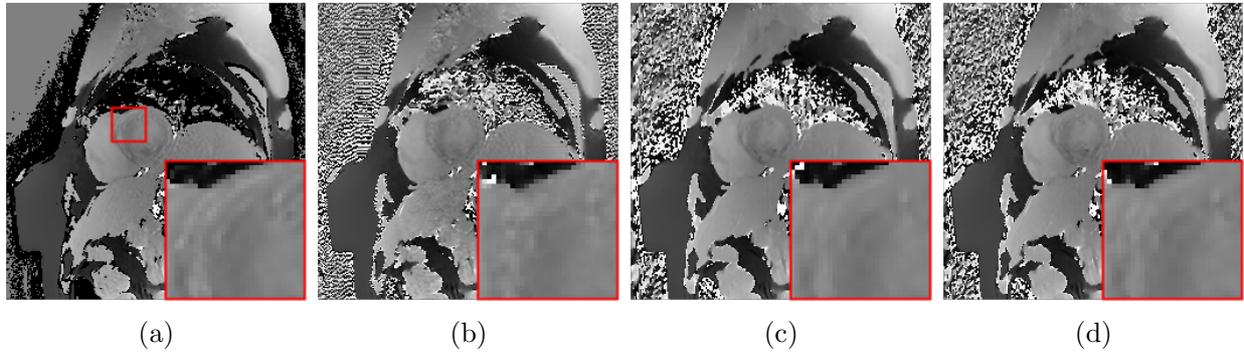


Figure 4.11: Phase reconstruction comparison in a scan accelerated by 8 of the fully sampled data (a) and the reconstructions provided by k-t FOCUSS, DLMRI and DLTG (b, c, d).

### 4.5.3 Reconstruction with noise

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

In figure 4.12a we plot the dependence of the reconstruction quality on parameter  $q$  at a sampling factor 0.25. The different input noise values displayed are represented as  $\text{PSNR}_f$ , which corresponds to the PSNR of the inverse DFT reconstruction of fully sampled data corrupted by noise. In terms of absolute noise power,  $\text{PSNR}_f = \{25.8, 31.8, 35.8, 41.8\}$  dB is equivalent to

$\sigma^2 = \{4 \times 10^{-8}, 10^{-8}, 4 \times 10^{-9}, 10^{-9}\}$ . The constant  $q$ , defining the consistency with respect to acquired data (see equation (4.18)), was set to  $5 \times 10^{-6}$  and  $5 \times 10^{-5}$  for the DLTG and DLMRI algorithms respectively based on these results. The same empirical value was found for  $q$  in other tests independent of the patch and data set sizes. For high noise values ( $\text{PSNR}_f \leq 31.8$  dB) a fine tuning of this constant is desirable as the reconstruction depends heavily on it, but for lower noise values results are not very sensitive to it.

We plot in figure 4.12b the reconstruction performance as a function of  $\text{PSNR}_f$  for a scan accelerated by a factor of 4. The same rank order between the three methods is preserved in this test, and both methods using DL present a milder decay of performance at high noise values than k-t FOCUSS. The main reason for this improved robustness can be attributed to the denoising capabilities that have been demonstrated by the K-SVD algorithm. Sparse coding with a trained overcomplete dictionary and averaging overlapping patches is a powerful method to denoise natural images [49]. The visual comparison for  $\text{PSNR}_f = 25.8$  dB in figure 4.13 shows how the k-t FOCUSS reconstruction is much more contaminated by noise than the dictionary based reconstructions.

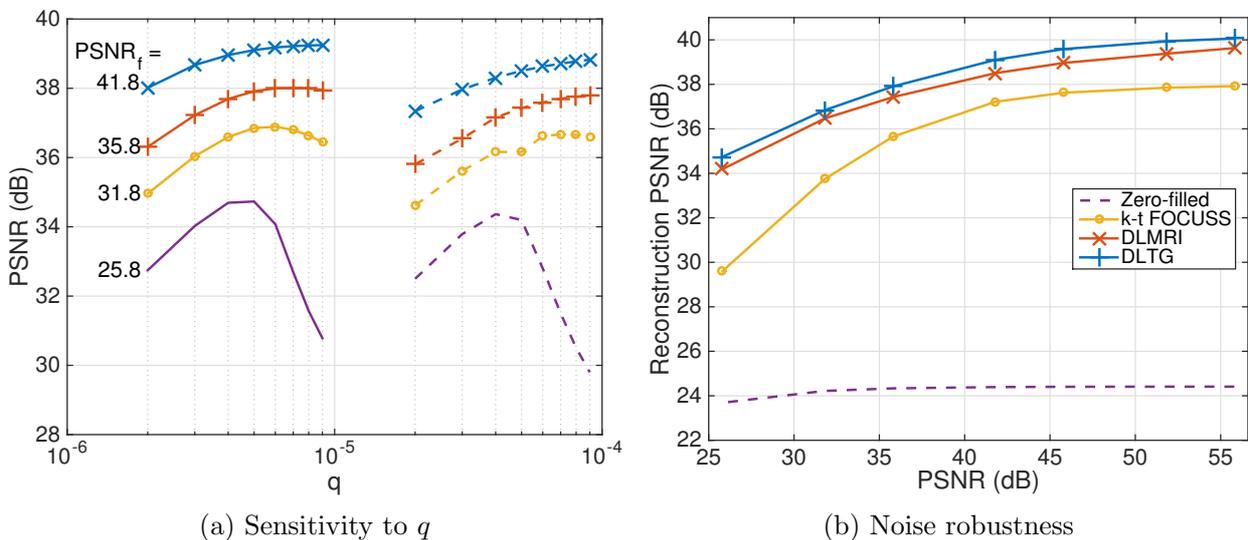


Figure 4.12: Algorithm evaluation with added noise. Reconstruction sensitivity to noise power and parameter  $q$  for a scan accelerated by 4 (left) using the DLTG (solid) and DLMRI (dashed) algorithms, and their robustness to input noise (right).

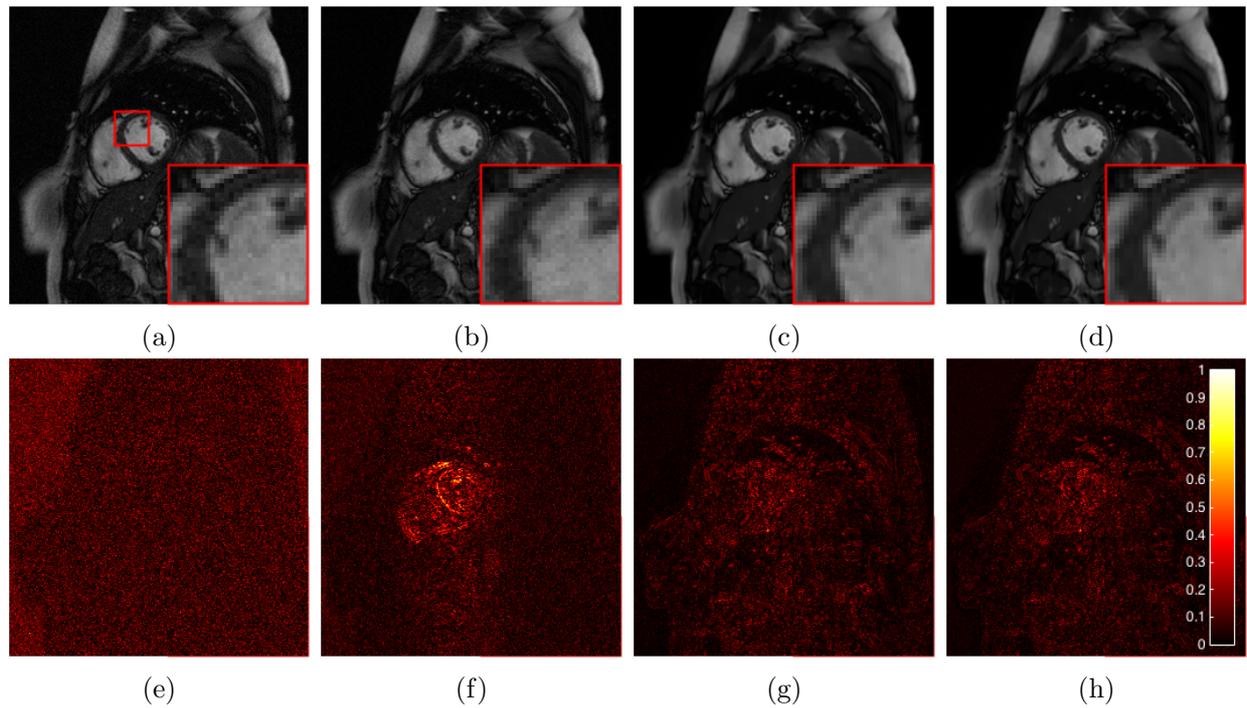


Figure 4.13: Visual comparison of the reconstructions from a scan accelerated by 4 that has been contaminated by complex noise with input  $\text{PSNR}_f = 25.8$  dB (a) using k-t FOCUSS (b), DLMRI (c) and DLTG (d) with their respective errors multiplied by 6 (f, g, h). The noise added to the fully sampled data set is shown in (e) amplified by 6.

#### 4.5.4 Spatial dictionaries

The main objective of this work is to analyse the potential of using spatio-temporal dictionaries for dynamic data instead of independently reconstructing temporal frames with a 2D (spatial) dictionary. In dynamic cardiac imaging, the temporal dimension is known to be highly redundant as changes through time are slow and confined to specific regions of the image. Hence, a sparsity model that specifically exploits this temporal redundancy can be expected to perform better than one that disregards it.

Figure 4.14 visually compares the reconstruction of a 6 fold accelerated data set using the DLMRI algorithm with a spatio-temporal and a spatial dictionary. The reconstruction using a spatio-temporal dictionary shows much better dealiasing properties whereas the reconstruction using the spatial dictionary has unresolved aliasing and shows an important loss of structure. The PSNR metric was 38 dB and 27.5 dB for the spatio-temporal and spatial dictionary recon-

struction respectively. The temporal profiles also demonstrate a more accurate reconstruction with the spatio-temporal dictionary. Particularly, the spatial dictionary reconstruction shows a blocky structure that is less consistent across time in terms of both structure and contrast, which is a consequence of reconstructing temporal frames independently.

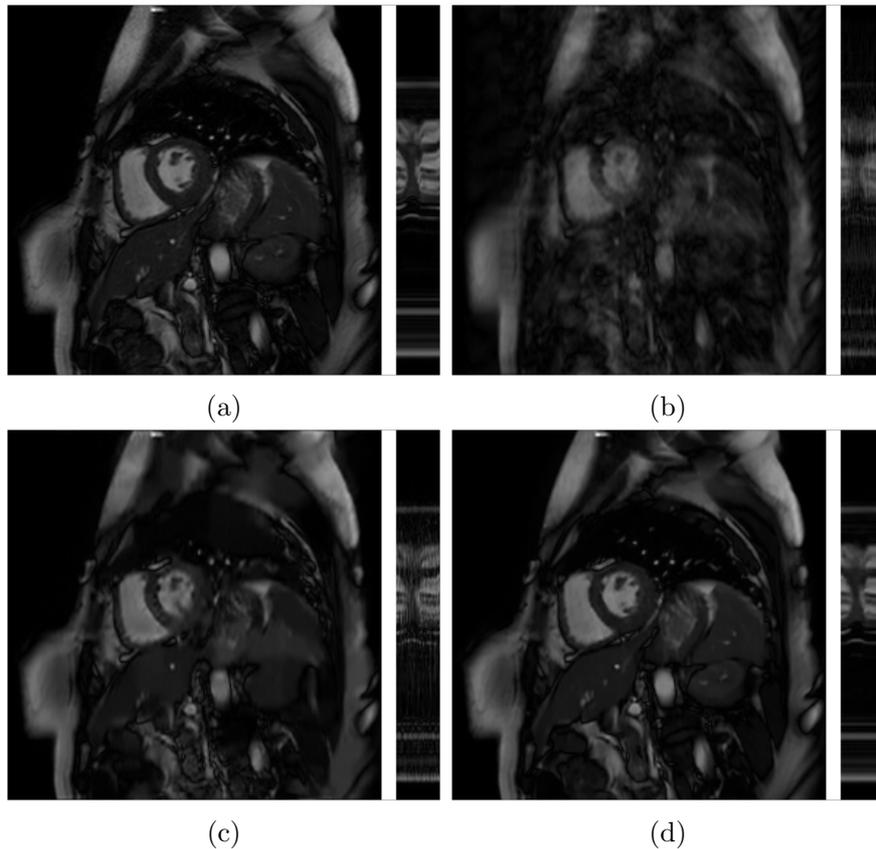


Figure 4.14: Comparison of 2D spatial and 3D spatio-temporal reconstruction. Shown are a fully sampled frame (a), its undersampled by 6 zero-filled version (b), and the spatial only (c) and spatio-temporal (d) reconstructions. All figures are shown with their respective temporal profiles on the right hand side.

### 4.5.5 Real and complex-valued dictionaries

The use of real-valued dictionaries is a main difference with respect to previous work on dictionary learning for MRI. In [117, 147], a complex-valued dictionary is trained which is then used to sparsely code complex-valued data. The DLTG algorithm instead trains a real-valued dictionary to sparsely code real and imaginary parts of the data independently. Although training

a real-valued dictionary is not expected to outperform a complex-valued dictionary, it is worth understanding the differences between these two strategies.

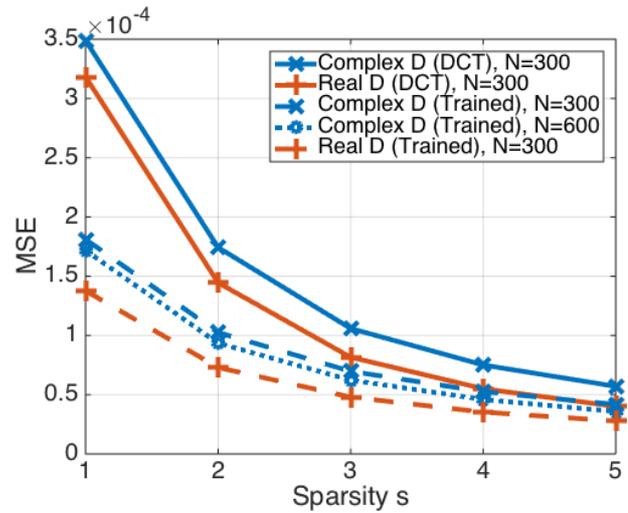
We look at the error that is produced on a fully sampled data set when these two different sparsity models are used to approximate it. In figure 4.15, we compare the average error per patch when a scan is coded using a maximum of  $s$  atoms from different dictionaries. It is difficult to portray a balanced comparison as the  $l_0$  norm is different for a complex and a real-valued dictionary. The complex-valued dictionary allows for complex-valued sparse representations, meaning that the coding of a patch can use  $s$  atoms from the dictionary and their phase can be rotated. This allows for  $2s$  degrees of freedom in the sparse representation. On the other hand, the coding using a real-valued dictionary uses  $s$  atoms for the real part of a patch and another independent  $s$  atoms for the imaginary part, hence matching the  $2s$  degrees of freedom of the complex-valued dictionary strategy.

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

### 4.5.6 Parameter selection

The choice of algorithmic parameters is crucial for a suitable operation of the DLTG algorithm. The plots in figure 4.16 show the influence of the dictionary and patch size in the reconstruction performance of a scan accelerated by 6. In figure 4.16a the patch size was kept at  $N_p = 4 \times 4 \times 4$

Figure 4.15: Average error per patch produced by the assumption of  $s$  sparsity on a fully sampled scan using different dictionaries.



and in figure 4.16b the dictionary size was fixed at  $K = 600$ . The quality of the reconstruction is comparable in the broad range of values tested except for very large patch sizes. These tend to oversmooth the result missing out on fine details, which deteriorates the performance. It is also noticeable that above a certain dictionary overcompleteness, there is little improvement in using larger dictionaries.

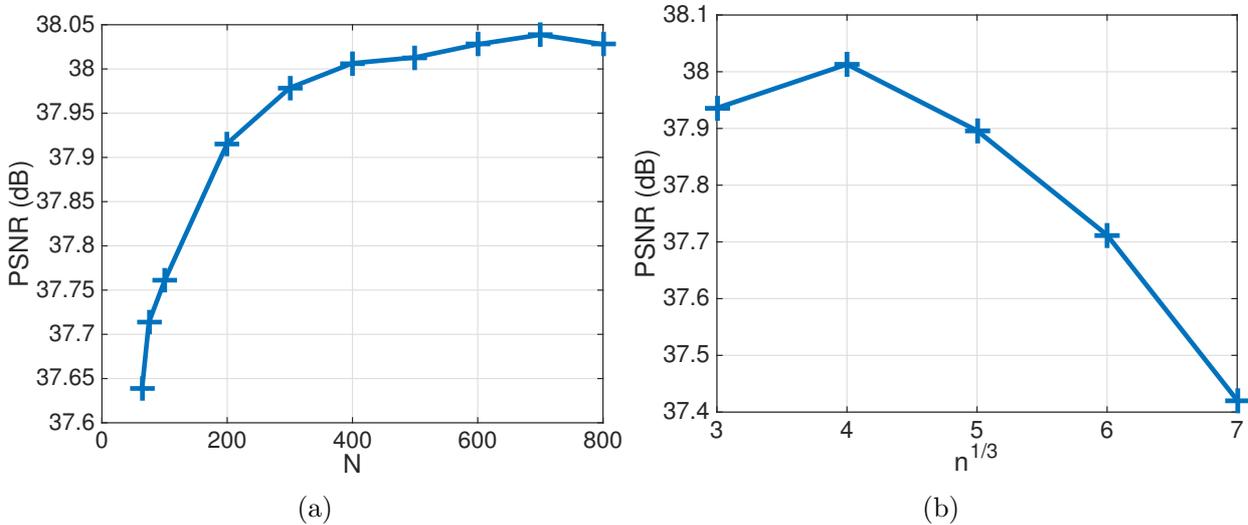


Figure 4.16: Influence of dictionary size (a) and patch size (b) on the reconstruction of a scan accelerated by 6.

### 4.5.7 Algorithm convergence, speed and acceleration

The intermediate convergence of terms in the global statement is summarised in figure 4.17 for an example reconstruction of a 6 fold accelerated scan. Although the convergence of the DLTG algorithm is still to be proven, all the tests undergone showed convergence to a stable result.

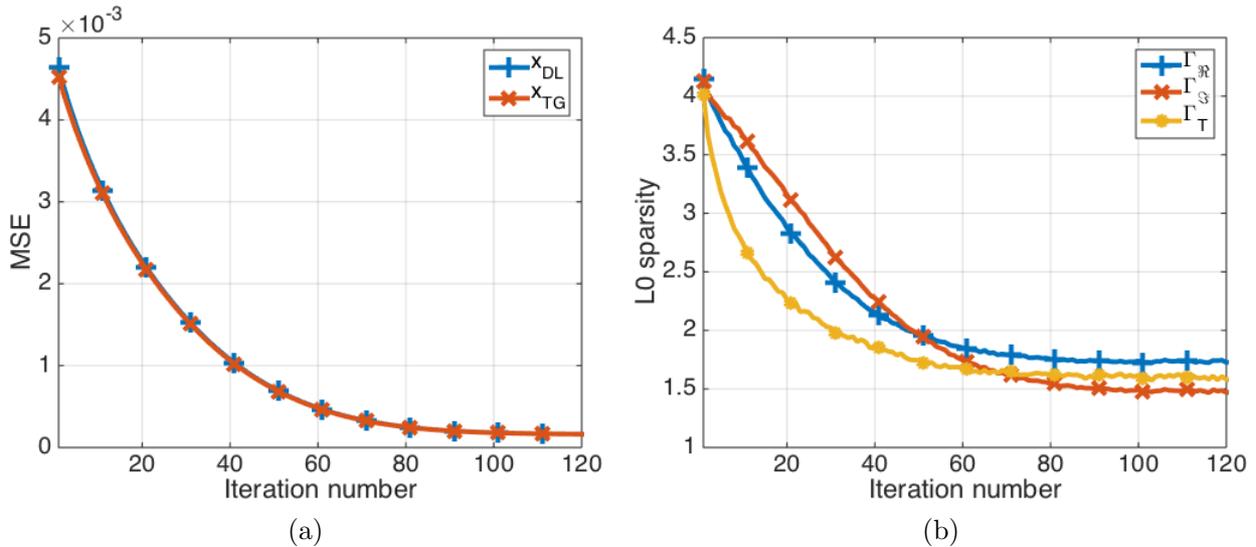


Figure 4.17: MSE convergence of approximate results  $\mathbf{x}_D$  and  $\mathbf{x}_{TG}$  relative to the ground truth (a), and the intermediate sparsity level of the real, imaginary and training patches coding (b).

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

This behaviour is illustrated in figure 4.18a, where the reconstruction performance of a 6 fold accelerated data set is evaluated in terms of PSNR and convergence rate for different values

of  $\epsilon$ . Figure 4.18b shows the time needed for one coding step of the entire data set (both real and imaginary parts), which is the most demanding step in the algorithm. Matlab R2011b was used for this assessment of the runtime on an Intel Core i7-2600 CPU at 3.4 GHz and 8 GiB of memory. The C++ implementation of K-SVD provided by Ron Rubinstein in [113] was used for the Batch-OMP stages.

Figure 4.18 allows for an empirical assessment of the global runtime of the entire reconstruction. For instance, most of the results shown in this work were obtained with  $\epsilon = 0.007$ , which translates into a coding stage of 200 s, and the reconstruction in figure 4.18a took about 120 iterations to converge. The bulk of the computation therefore required  $200 \times 120/3600 = 6.6$  h. Notice that there is nevertheless a lot of flexibility in the speed of the algorithm, as using  $\epsilon = 0.01$  would reduce this time to  $110 \times 75/3600 = 2.3$  h without much compromising the end result.

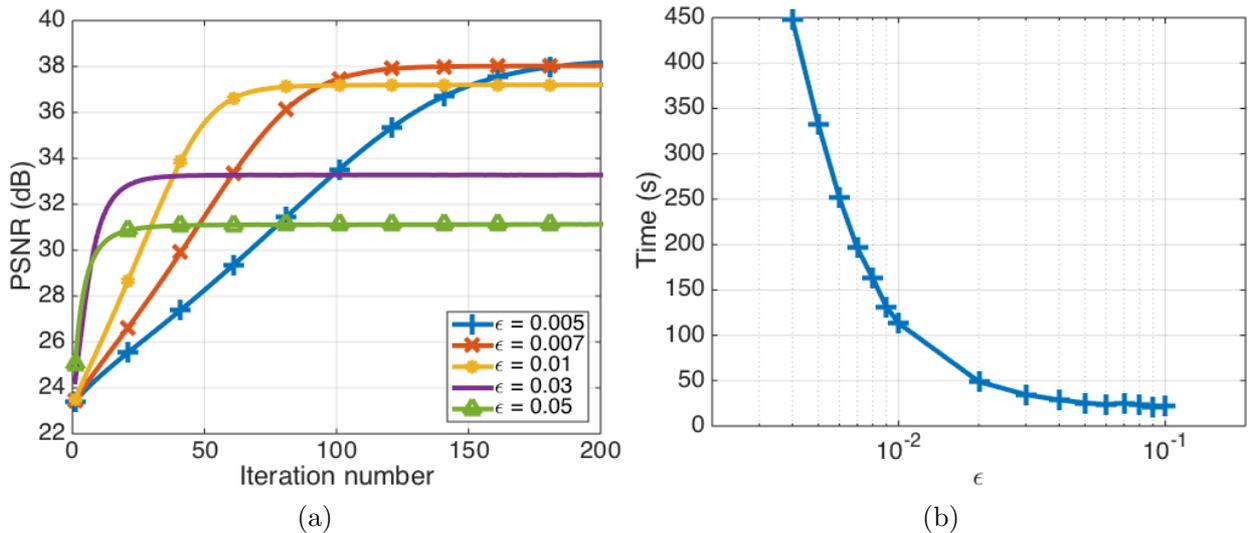


Figure 4.18: Example of a reconstruction’s dependence to  $\epsilon$  (a) and the Batch-OMP runtime for one coding of real and imaginary parts of a scan (b).

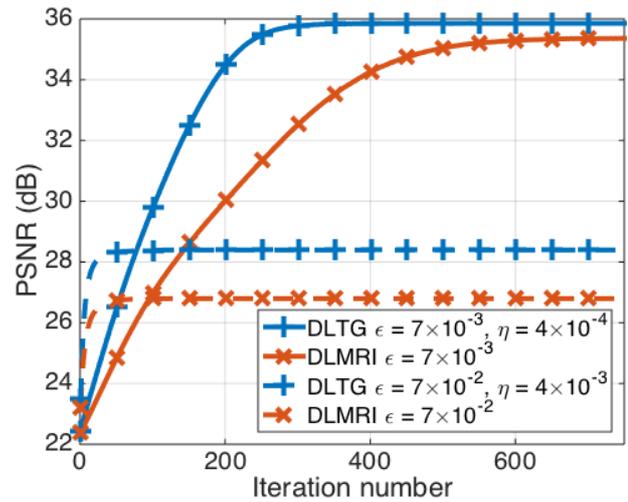
There is a large difference between these runtimes and the average 15 seconds required for a k-t FOCUSS reconstruction but there is scope for much improvement. Runtimes for our experiments were obtained with a sequential coding of individual time frames to save on memory requirements, but the coding can be performed on the full data set relaxing memory restrictions. Also, a very appealing modification is to process the  $2N$  sparse coding steps necessary in each

iteration of the algorithm in parallel, given that they are completely independent from each other. Additionally, from the behaviour observed in figure 4.18a, it seems natural to seek optimal trajectories for the data fidelity term  $\epsilon$  such that it starts as a large value benefiting from fast initial convergence, and is dynamically decreased as the algorithm converges to a stable solution to reach high reconstruction quality. These are very simple additions that can reduce the total runtime considerably. Decreasing patch sizes, their overlap stride and the number of dictionary atoms also accelerates the algorithm although this can compromise the end result.

These extensions have been tested on the same 6 fold experiment that was used to produce results in figure 4.18a. We run OMP on a full data set at once and parallelise coding among 8 CPU cores. Moreover, we choose  $\epsilon^{(r+1)} = \epsilon^{(r)}/\omega$  to be dynamically updated at each iteration  $r$ , setting  $\epsilon^{(0)} = 0.01$  and  $\omega = 1.2$  empirically, and we also reduce the number of dictionary atoms to  $K = 169$ . These settings allow to obtain a result with equivalent PSNR as the best performing example in figure 4.18a in only 6 minutes. Each OMP coding stage takes about 10 seconds and is still the most expensive operation. Further parallelisation of this step could be achieved with GPU processing.

The use of temporal gradient sparsity in DLTG also has an important impact on the convergence rate compared to the DLMRI version. Figure 4.19 shows the convergence of a 10 fold accelerated scan in terms of DLMRI iterations and DLTG outer iterations. Two settings are compared for each algorithm, a slow one using  $\epsilon = 0.007$  and  $\eta = 4 \times 10^{-4}$ , which are the values used for all the results shown above, and a faster but less performant one using  $\epsilon = 0.07$  with  $\eta = 0.004$ . The enforcement of temporal gradient sparsity in DLTG accelerates the process considerably. The reason for this is that the TG transform is able to reduce aliasing from a global perspective in a way that is infeasible using only the patch-based approach of the DLMRI algorithm. More specifically, the high temporal gradient complexity that independent masks produce in consecutive frames is highly penalised by this auxiliary sparsity term.

Figure 4.19: Convergence rate of DLMRI and DLTG for two different data fitting terms setups.



### 4.5.8 Influence of training data

The algorithm's behaviour depends on the data chosen for dictionary training. The previous results were computed with online learning of a dictionary using patches from the target scan. Figure 4.20 shows an example of 20 atoms of the spatio-temporal dictionary trained in the last iteration of the DLTG reconstruction in figure 4.7. These look very different from the initial structured DCT dictionary. Some show coarse edges in different directions (atoms 5, 15, 17) whereas others do not have a significant structure and are necessary to capture image details (4, 12, 14). Furthermore, some contain temporal dynamism (1, 3, 10), with changes across temporal instances  $t = 1, \dots, 4$ , while others are relatively static through time (5, 17, 18).

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

Table 4.2 shows the PSNR quality of reconstructions at accelerations 4, 6 and 8 using different training data sets with  $\epsilon = 0.01$ . The comparison looks at the initial DCT dictionary without

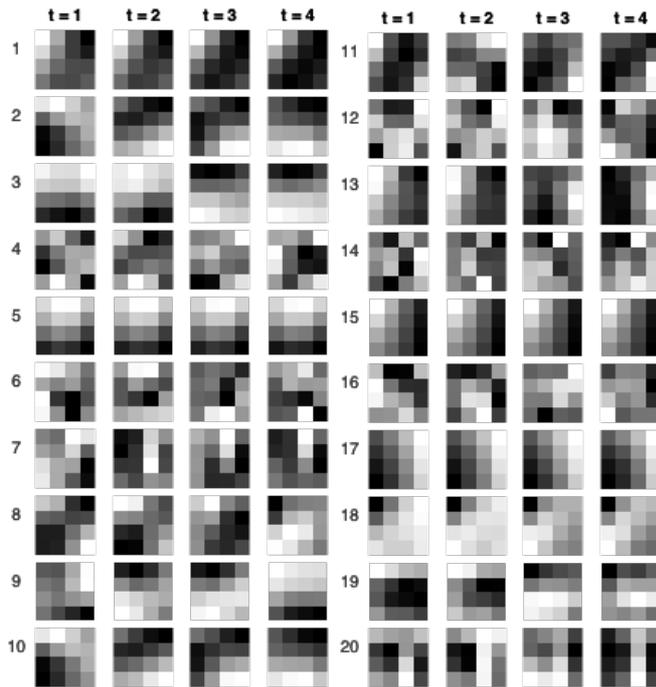


Figure 4.20: Examples of patches learnt in the last iteration of the 8 fold accelerated DLTG reconstruction shown in figure 4.7.

training, the dictionary trained online on the target scan, a dictionary trained offline on the fully sampled target scan, and an offline training on a fully sampled scan from a different subject. As expected, the initial DCT dictionary always provides the worst reconstruction quality and the best one is given by the fully sampled training data on the target scan. Notice however that training data coming from a different cardiac scan can be used to train a dictionary offline and closely approximate the performance of training on the true data.

Table 4.2: Training data influence on PSNR (dB) with  $\epsilon = 0.01$ .

Training data	$R = 4$	$R = 6$	$R = 8$
DCT (no training)	38.86	36.78	35.22
True (online)	<b>39.20</b>	37.09	<b>35.66</b>
True (offline)	<b>39.20</b>	<b>37.12</b>	<b>35.66</b>
Different - 1 scan (offline)	<b>39.20</b>	37.09	35.65

Table 4.3 presents the same comparison using  $\epsilon = 0.007$ . In this regime, the dictionary is forced to very accurately represent the training data, so the tailoring to a particular scan during training is accentuated. Offline training on a different scan performs slightly worse than using the target scan, but still better than the initial DCT dictionary.

Table 4.3: Training data influence on PSNR (dB) with  $\epsilon = 0.007$ .

Training data	$R = 4$	$R = 6$	$R = 8$
DCT (no training)	39.8	37.3	36.2
True (online)	<b>40.1</b>	<b>38.0</b>	<b>36.7</b>
True (offline)	<b>40.1</b>	37.8	<b>36.7</b>
Different - 1 scan (offline)	39.9	37.6	36.5

It is important to note that, although dictionary training enhances the reconstruction quality, this increase is relatively small and sometimes might not be noticeable visually. These results are in accordance with the comparisons shown in K-SVD denoising [49, 111], where dictionary training rarely improves PSNR by more than 0.5 dB. The additional computation necessary for training offline, and most importantly online, should be taken into account and questioned as being necessary or not depending on whether the improvement it can bring is clinically relevant.

To conclude, we show in table 4.4 the influence of the initial dictionary when reconstructing with online training. The test compares a DCT initial dictionary, with one that is initialised with random patches from the data and another that is filled in with independent and identically distributed (IID) Gaussian samples. The DCT initial dictionary provides the best reconstruction whereas the random IID Gaussian initialisation performs noticeably worse. This is probably because, although the DCT dictionary is not tailored to a scan, it is still able to provide a sparse representation of it in the first iteration of each dictionary training step, facilitating the task of the K-SVD algorithm. The IID Gaussian initialisation will most probably not find a sparse representation of the training data, and this deteriorates the behaviour of K-SVD learning.

Table 4.4: Influence of dictionary initialisation on PSNR (dB).

DCT	Data	IID Gaussian
<b>38.0</b>	37.6	36.2

## 4.6 Conclusion

We have presented a novel algorithm for the acceleration of dynamic MRI acquisition based on a sparsity model that can learn redundancy in the data and an auxiliary constraint on

TG sparsity that helps speed up convergence and provides better results at high acceleration rates. The method was tested on data from 10 subjects and provided superior performance in every case compared to the fixed basis transform CS MRI method k-t FOCUSS. There were no failures and no gross errors. The reconstruction of complex-valued MR data was performed by splitting real and imaginary parts of the sequences and coding them with a real-valued dictionary. The method proposed is shown to largely outperform the k-t FOCUSS algorithm, which supports a conclusion that patch-based reconstruction and adaptive sparsity bases have greater potential than fixed basis global sparsity transforms.

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

Another important finding presented is the influence that the training has upon the reconstruction. It has been shown how dictionary training enhances the reconstruction with respect to an initial DCT dictionary, but this increase in quality is limited. This is the same conclusion that could be drawn from the original K-SVD results on denoising [49, 111]. Considering the computational overhead that online dictionary training entails, it could be regarded as unnecessary if the reconstruction improvements are barely noticeable to the human eye. Nevertheless, we have also seen how offline training on a data set from a different patient could provide almost the same results as online training, and this is a viable option that would benefit from the increase in reconstruction quality while keeping the computation cost for training separate from the online runtime.

# Chapter 5

## Parallel MRI reconstruction with dictionaries

This chapter is based on the following publication:

- J. Caballero, A. N. Price, D. Rueckert, J. V. Hajnal, *Patch-based dictionaries for parallel MRI reconstruction*, Proceedings of the 22nd International Society for Magnetic Resonance in Medicine (ISMRM) Annual Meeting and Exhibition, p. 1560, Milan, Italy, 10-16 May 2014.

### 5.1 Introduction

Modern sampling techniques exploiting sparse properties of images have demonstrated potential in accelerating MRI acquisition. However, the speed constraints of this imaging modality have been tackled through other means, many of which are older than the ideas introduced by CS. Perfect recovery from incomplete data is theoretically only possible if the data to be reconstructed is redundant. Image sparsity implies an implicit data redundancy, but redundancy

can also be explicitly generated by sampling the same data multiple times, and this is the idea underlying parallel MRI.

Parallel MRI uses multiple coils that acquire the same data through different sensitivity patterns. In this respect, the burden of data collection is distributed among multiple coils, and produces an overcomplete set of linear equations that represent an oversampling of the image. The solution to this set of equations, which is the magnetisation image of interest, reduces to either a matrix inversion in image space [112] or a deconvolution in k-space [59]. Theoretically, the achievable acceleration rate can be as high as the number of parallel coils, but in practice it is bounded by coil geometry to typically 3 or 4 fold acceleration. Sensitivity patterns cannot be made perfectly orthogonal to each other, resulting in a rank deficient sampling process that when inverted causes numerical instability.

One possibility to deal with the instability of the parallel imaging process is to regularise the solution, discarding solutions of unlikely characteristics. A common choice is LS regularisation, also known as Tikhonov regularisation. This imposes a penalty on solutions of high energy, or on ones which deviate too much from a reference solution such as a low resolution acquisition scan of the image, which can be acquired fast [80]. This strategy can effectively control the noise amplification caused at high undersampling rates, but risks smoothing the reconstruction and losing detail.

With the introduction of CS reconstruction techniques, and previously with the analysis of sparse signals, newer and more versatile regularisation techniques have been proposed for parallel MRI regularisation. Sparsity promoting norms such as  $l_0$  or  $l_1$  norms can be incorporated into the parallel acquisition framework and can contribute to the reduction of noise and aliasing in the reconstruction as long as sampling incoherence constraints are ensured. The SPIRiT method [85] poses the parallel recombination as an energy minimisation of the reconstructed k-space, and is particularly well suited to incorporate regularisation penalty terms. A variation called  $l_1$ -SPIRiT can impose sparsity on the parallel reconstruction. The most investigated sparsity domains for MRI images are the image domain for angiography [84], wavelets and

TV for spatial data [84, 65, 31, 4], and the temporal Fourier transform for cardiac cine data [84, 70, 55] amongst others.

These transforms are all complete, and though they can be very effective and efficient to use at low acceleration rates, they may fail to provide sufficient sparsity at high accelerations. Moreover, they provide non-adaptive sparsity models, and when relying too much on them for the representation of missing data they can introduce unnatural looking artefacts. In the previous chapter we have seen how an overcomplete transforms could overcome some of these limitations, with added flexibility in sparsely representing images, and can additionally become adaptive to image features through a DL process [134, 49], better conditioning the CS problem and more faithfully reconstructing image features.

In this chapter we present an algorithm to combine DL and parallel imaging and analyse its performance against CS parallel MRI counterpart techniques that use non-adaptive transforms for sparse representations. Results are analysed with retrospective undersampling of phantom and raw cardiac cine data, and validated on prospectively undersampled raw cardiac data. The method proposed is based on SPIRiT, given that its formulation easily allows to impose constraints on the reconstruction, and the modification of the  $l_1$ -SPIRiT framework allows direct comparison with complete and non-adaptive based CS constraints.

## 5.2 Compressed sensing parallel MRI

Using multiple transceiver coils for the acquisition of MRI data is an effective method of reducing scan time. Each coil collects the same data at the same time but filtered by different sensitivity patterns. This effectively allows to induce redundancy through hardware in the sampling process, which is a necessary foundation for the recovery of undersampled data.

Assuming the full k-space of a 2D image of interest to be the signal  $\mathbf{x} \in \mathbb{C}^N$ ,  $N = N_x \times N_y$ , its parallel coil  $c$  image is  $\mathbf{x}_c = \mathbf{S}_c \mathbf{x}$ , where  $\mathbf{S}_c$  is a sensitivity pattern that is distinctive of coil

$c$  and is point-wise normalised by  $\mathbf{C} = \left( \sqrt{\sum_{c=1}^{N_c} \mathbf{S}_c^H \mathbf{S}_c} \right)^{-1}$  such that  $\mathbf{S}_c^H \mathbf{S}_c = \mathbf{I}_N$ . The k-space acquisition of coil  $c$  can be described as  $\hat{\mathbf{x}}_c = \mathbf{F} \mathbf{S}_c \mathbf{x}$ , where  $\mathbf{F}$  is a 2D discrete Fourier transform, and the undersampled, noisy acquisition of  $\hat{\mathbf{x}}_c$  can be represented as  $\hat{\mathbf{x}}_{u,c} = \mathbf{M} \hat{\mathbf{x}}_c + \mathbf{n}$ , where  $\mathbf{n}$  is AWG noise, complex and circular, with  $\mathbf{M} \in \mathbb{R}^{M \times N}$  a binary matrix with entries  $[\mathbf{M}]_{n,n} = 1$  if k-space sample  $n$  has been acquired. If we let  $\tilde{\mathbf{M}}$  and  $\tilde{\mathbf{F}}$  be the block diagonal concatenations of  $N_c$  replicas of  $\mathbf{M}$  and  $\mathbf{F}$  respectively, and  $\tilde{\mathbf{x}}$ ,  $\tilde{\mathbf{x}}$  and  $\tilde{\mathbf{S}}$  the vertical concatenations of  $\hat{\mathbf{x}}_c$ ,  $\mathbf{x}_c$  and  $\mathbf{S}_c, \forall c \in \{1, 2, \dots, N_c\}$ , the complete parallel acquisition can be compactly described as

$$\tilde{\mathbf{x}}_u = \tilde{\mathbf{M}} \tilde{\mathbf{F}} \tilde{\mathbf{S}} \mathbf{x} + \tilde{\mathbf{n}} = \tilde{\mathbf{M}} \tilde{\mathbf{x}} + \tilde{\mathbf{n}}, \quad (5.1)$$

where  $\tilde{\mathbf{n}} \in \mathbb{C}^{MN_c}$  is now noise received in all parallel coils. Noise will be assumed independent for each coil, although in practice noise correlation among channels exists.

Different methods have been proposed for the reconstruction of  $\mathbf{x}$  given  $\tilde{\mathbf{x}}_u$ , which can be categorised as image based or k-space based methods depending on the domain in which reconstruction is performed. SENSE [112] is an established image based parallel reconstruction technique that poses the problem as the unfolding of information in  $\tilde{\mathbf{x}}_u$  given a prior knowledge of the sensitivity patterns  $\tilde{\mathbf{S}}$  by a simple matrix pseudo-inverse. Although this method is simple and efficient, one drawback is that it requires prior estimation of coil sensitivities, but numerous techniques exist that can reliably estimate sensitivities.

The image space unfolding of SENSE can equivalently be performed in k-space by methods like GRAPPA [59]. Using the convolution property of the Fourier transform, the matrix inversion operation can be turned into a kernel deconvolution in k-space, where k-space samples are synthesised by a linear combination of their neighbouring samples. Given that coil sensitivities can be approximated as low-frequency bias fields, small k-space deconvolution kernels are sufficient for an accurate estimation of missing k-space samples. In a first stage, k-space convolution weights are calculated which relate a neighbourhood of acquired sampled and an additional auto-calibrating k-space line. These weights express the local relationships among

k-space samples, so they can be used to recover unknown samples. Furthermore and contrary to SENSE, GRAPPA does not require the explicit estimation of sensitivity patterns given that this information is implicit in the deconvolution weights used for reconstruction. The original GRAPPA method is however restricted to regular sampling patterns. There have been multiple attempts to combine the benefits of parallel MRI with CS, some examples extending SENSE and GRAPPA can be found in [76, 108, 149, 30].

### 5.2.1 SPIRiT

SPIRiT [85] is a reconstruction technique that generalises the GRAPPA methodology. It poses the reconstruction as the solution to the minimisation of a functional, where consistency with acquired k-space samples is weighted by the synthesis of k-space samples on the full solution  $\hat{\mathbf{x}}$ , which includes both acquired and non-acquired samples. This can be formally expressed as

$$\min_{\tilde{\mathbf{x}}} \|\tilde{\mathbf{x}}_{\mathbf{u}} - \tilde{\mathbf{M}}\tilde{\mathbf{x}}\|_2^2 + \lambda^2 \|(\mathbf{G} - \mathbf{I})\tilde{\mathbf{x}}\|_2^2, \quad (5.2)$$

with  $\mathbf{I}$  the identity matrix. The matrix  $\mathbf{G}$  is a linear operator containing weights of the convolution kernel which synthesise k-space samples from a local region of neighbour samples, and it can be computed prior to reconstruction from a small portion of fully sampled k-space. The k-space interpolation mechanism of SPIRiT recovers every coil image separately, which can then be recombined using a sum of squares (SoS) of all images.

As long as the acceleration rate  $R = \frac{N}{M}$  satisfies  $R \leq N_c$ , perfect reconstruction of image  $\mathbf{x}$  can in theory be achieved. However, in practice the acceleration rate is limited by numerical instabilities that occur when  $R$  approaches the theoretical maximum  $N_c$ , with a significant reduction in SNR. This phenomenon is demonstrated in figure 5.1, where the 8 coil reconstruction of a phantom with very little AWG noise is degraded as the acceleration rate increases.

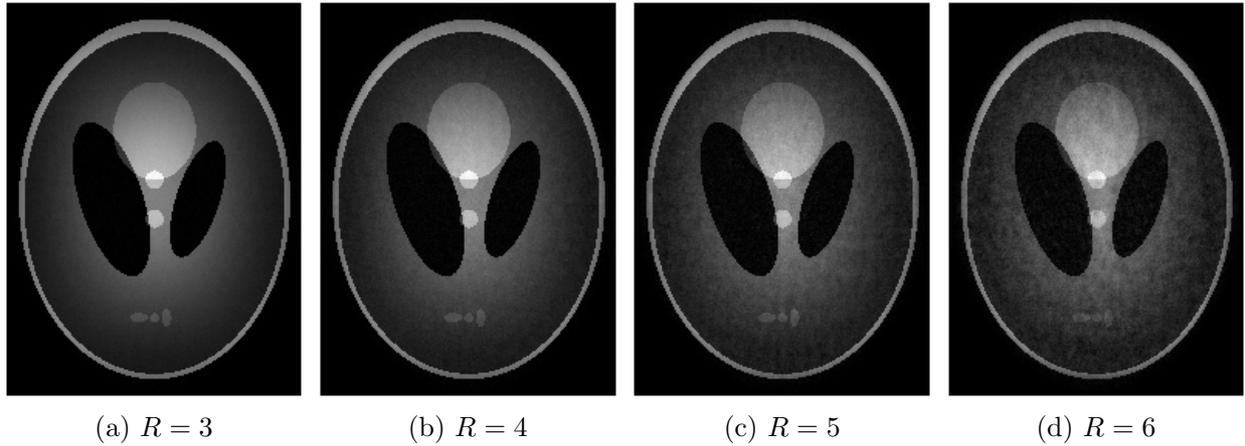


Figure 5.1: SPIRiT parallel reconstruction of a phantom image simulating 8 coils. Acceleration rates  $R > 4$  present a much lower SNR than for  $R \leq 4$ .

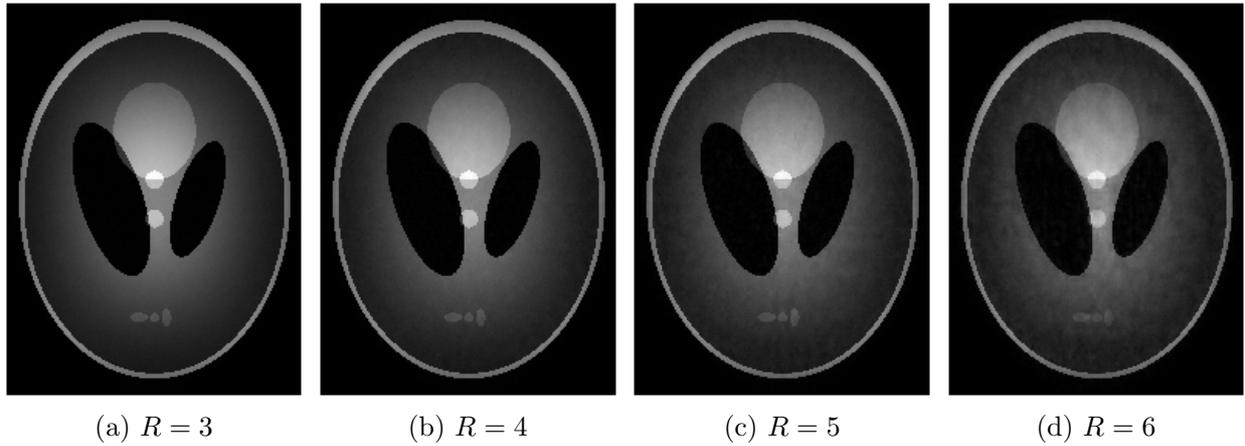


Figure 5.2:  $l_1$ -SPIRiT parallel reconstruction of a phantom image simulating 8 coils. The regularisation with a sparsity term can reduce noise amplification stemming from parallel recombination instability.

### 5.2.2 $l_1$ -SPIRiT

The formulation of the SPIRiT reconstruction as an optimisation task makes the addition of a constraint or penalty term to reduce the solution set straightforward. Regularisation terms that have demonstrated great potential when combined with parallel MRI are sparsity promoting norms  $\|\mathcal{S}(\mathbf{x})\|_p$ , where  $p = 0$  or  $p = 1$  and  $\mathcal{S}(\mathbf{x})$  is a sparsity transform. Building upon the description of SPIRiT, CS can be used to formulate an  $l_1$ -SPIRiT reconstruction [85]:

$$\min_{\tilde{\mathbf{x}}} \|\tilde{\mathbf{x}}_{\text{u}} - \tilde{\mathbf{M}}\tilde{\mathbf{x}}\|_2^2 + \lambda^2 \|(\mathbf{G} - \mathbf{I})\tilde{\mathbf{x}}\|_2^2 \quad \text{s.t.} \quad \|\mathcal{S}(\tilde{\mathbf{F}}^H\tilde{\mathbf{x}})\|_1 \leq s. \quad (5.3)$$

In figure 5.2 we show the same reconstructions as in figure 5.1 using  $l_1$ -SPIRiT, where the sparsity promoting transform is a Daubechies 4 wavelet transform.

How much an acquisition can be accelerated and perfectly recovered through CS is dependent on the sparsity degree of  $\mathcal{S}(\tilde{\mathbf{F}}^H \tilde{\mathbf{x}})$  and the incoherence between the sampling and the sparsity domains. The incoherence criterion can be generally satisfied in MRI with a random selection of k-space samples and the sparsity transform used will normally depend on the object being imaged. The sparsity degree available from complete and fixed transforms however can result to be too restrictive compared to overcomplete and adaptive transforms. Patch-based sparsity has been successfully used for the reconstruction of undersampled MRI data in the previous chapter, but only in a single coil, synthetic setup.

### 5.3 D-SPIRiT: Parallel MRI with dictionaries

In this section we present an extension of the SPIRiT method that forces the reconstruction to be sparsely represented by an overcomplete, and additionally adaptive, patch-based dictionary. Let the magnetisation image of interest  $\mathbf{x}$  of  $N$  pixels be decomposed into  $N$  overlapping patches of size  $\sqrt{N_p} \times \sqrt{N_p}$  assuming they wrap around image boundaries. Using  $\mathbf{R}_n, \forall n \in \{1, 2, \dots, N\}$ , to denote a linear operator extracting patch  $n$  from  $\mathbf{x}$ , then patches can be represented sparsely using  $\gamma_n$  as  $\mathbf{R}_n \mathbf{x} \approx \mathbf{D} \gamma_n$ , where  $\mathbf{D} \in \mathbb{C}^{N_p \times K}$  is an overcomplete dictionary of  $K$  atoms ( $N_p < K$ ). If  $\mathbf{\Gamma}$  denotes the matrix where column  $n$  is  $\gamma_n$ , the sparse coding of image  $\mathbf{x}$  by a dictionary  $\mathbf{D}$  can therefore be cast as

$$\min_{\mathbf{\Gamma}} \|\mathbf{R}_n \mathbf{x} - \mathbf{D} \gamma_n\|_2^2 \quad \text{s.t.} \quad \|\gamma_n\|_0 \leq s, \forall n. \quad (5.4)$$

The patch-based dictionary can optionally be defined as an optimisation variable to become adaptive to the reconstructed data through a DL process such as K-SVD.

We propose to impose the dictionary sparsity condition only on the recombined image, which

contains the object of interest, instead of forcing all coil images to satisfy sparsity criteria as suggested in  $l1$ -SPIRiT. There are two important reasons for this choice. First, the coding of images with a smooth intensity variation such as coil images is generally less sparse than that of an image with relatively homogeneous intensity like a recombined image. Also, the enforcement of sparsity on patches is computationally intensive and a typical coil number can be as high as 32, so sparse coding these many images would considerably increase computational efforts.

### 5.3.1 Problem formulation

The general problem, referred to as D-SPIRiT, can be described as the parallel coil solution  $\tilde{\mathbf{x}}$  to

$$\min_{\tilde{\mathbf{x}}, \mathbf{\Gamma}} \|\tilde{\mathbf{x}}_{\mathbf{u}} - \tilde{\mathbf{M}}\tilde{\mathbf{x}}\|_2^2 + \lambda^2 \|(\mathbf{G} - \mathbf{I})\tilde{\mathbf{x}}\|_2^2 + \frac{\beta^2}{N_p} \sum_{n=1}^N \|\mathbf{R}_n \mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_n\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\gamma}_n\|_0 \leq s, \forall n, \quad (5.5)$$

where the variable  $\mathbf{D}$  can be included as another optimisation variable if we want to adapt the dictionary online. For computational reasons however we will not explore this option and will only use dictionaries that are either non-adaptive or have been adapted offline prior to reconstruction.

Solving this problem directly poses a number of difficulties. Firstly, the  $l0$  norm in the constraint makes this problem non-convex and discards convex minimisation techniques. However, it can either be relaxed to an  $l1$  norm and solved with linear programming tools or solved approximately with greedy algorithms such as OMP [135]. A second problem is the fact that the first two penalty terms are defined in global k-space of all coil images whereas the third is defined in overlapping patches of the recombined image domain. A complex optimisation task with such asymmetric penalties is likely to contain many local minima that are difficult to avoid when solving directly. Motivated by these two major obstacles, we opt to split the problem into simpler subtasks and alternate their solutions. This splitting has already been proposed for similar problems imposing  $l1$  sparsity on parallel MR data [64].

### 5.3.2 Solution with variable splitting and penalty method

We introduce the variable  $\mathbf{x}'$ , and without loss of generality we write

$$\min_{\tilde{\mathbf{x}}, \Gamma, \tilde{\mathbf{x}}'} \|\tilde{\mathbf{x}}_u - \tilde{\mathbf{M}}\tilde{\mathbf{x}}\|_2^2 + \lambda^2 \|(\mathbf{G} - \mathbf{I})\tilde{\mathbf{x}}\|_2^2 + \frac{\beta^2}{N_p} \sum_{n=1}^N \|\mathbf{R}_n \mathbf{x}' - \mathbf{D}\gamma_n\|_2^2 \quad \text{s.t.} \quad \begin{cases} \|\gamma_n\|_0 \leq s, \forall n \\ \tilde{\mathbf{x}}' = \tilde{\mathbf{x}} \end{cases}, \quad (5.6)$$

where now the sparse coding is imposed on the auxiliary image  $\mathbf{x}' = \tilde{\mathbf{S}}^H \tilde{\mathbf{F}}^H \tilde{\mathbf{x}}'$ . The reconstruction with respect to  $\tilde{\mathbf{x}}$  and  $\tilde{\mathbf{x}}'$  is now split, and we can relax the additional constraint to obtain

$$\min_{\tilde{\mathbf{x}}, \Gamma, \tilde{\mathbf{x}}'} \|\tilde{\mathbf{x}}_u - \tilde{\mathbf{M}}\tilde{\mathbf{x}}\|_2^2 + \lambda^2 \|(\mathbf{G} - \mathbf{I})\tilde{\mathbf{x}}\|_2^2 + \frac{\beta^2}{N_p} \sum_{n=1}^N \|\mathbf{R}_n \mathbf{x}' - \mathbf{D}\gamma_n\|_2^2 + \mu^2 \|\tilde{\mathbf{x}}' - \tilde{\mathbf{x}}\|_2^2 \quad \text{s.t.} \quad \|\gamma_n\|_0 \leq s, \forall n. \quad (5.7)$$

The variable  $\tilde{\mathbf{x}}'$  can be treated as an intermediate solution that is forced to be sparsely represented by  $\mathbf{D}$ , and its contribution to the result of  $\tilde{\mathbf{x}}$  is controlled with parameter  $\mu$ . This is the fundamental concept behind the penalty method [128]. An alternated solution of problem (5.7) with respect to  $\tilde{\mathbf{x}}$  and  $\tilde{\mathbf{x}}'$  would be equivalent to (5.5) as  $\mu \rightarrow \infty$ . It is however observed empirically that a convergent solution can be found for  $\mu$  fixed. With this splitting, we can now solve the problem alternating the optimisation variables.

#### Conjugate gradient SPIRiT

Begin by assuming  $\Gamma$  and  $\tilde{\mathbf{x}}'$  to be fixed, the solution with respect to  $\tilde{\mathbf{x}}$  is given by

$$\min_{\tilde{\mathbf{x}}} \|\tilde{\mathbf{x}}_u - \tilde{\mathbf{M}}\tilde{\mathbf{x}}\|_2^2 + \lambda^2 \|(\mathbf{G} - \mathbf{I})\tilde{\mathbf{x}}\|_2^2 + \mu^2 \|\tilde{\mathbf{x}}' - \tilde{\mathbf{x}}\|_2^2, \quad (5.8)$$

which is simply a SPIRiT reconstruction regularised by the prior solution  $\tilde{\mathbf{x}}'$ , and can be solved fast using conjugate gradient (CG) descent [85].

### Orthogonal matching pursuit sparse coding

For  $\tilde{\mathbf{x}}$  and  $\tilde{\mathbf{x}}'$  fixed, the problem with respect to the sparse coding matrix  $\mathbf{\Gamma}$  becomes

$$\min_{\mathbf{\Gamma}} \|\mathbf{R}_n \mathbf{x}' - \mathbf{D} \boldsymbol{\gamma}_n\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\gamma}_n\|_0 \leq s, \forall n, \quad (5.9)$$

where the summation can be ignored because the coding of all patches is independent. Assuming the linear operator  $\mathbf{R}_n$  wraps around the boundaries of the image, this problem can be efficiently solved as  $N$  separate OMP steps, and is a highly parallelisable task.

### Least squares sparse prior update

To conclude, the solution of  $\tilde{\mathbf{x}}'$  for a given  $\tilde{\mathbf{x}}$  and  $\mathbf{\Gamma}$  is given by the solution to

$$\min_{\tilde{\mathbf{x}}'} \sum_{n=1}^N \|\mathbf{R}_n \mathbf{x}' - \mathbf{D} \boldsymbol{\gamma}_n\|_2^2 + \frac{\mu^2 N_p}{\beta^2} \|\tilde{\mathbf{x}}' - \tilde{\mathbf{x}}\|_2^2, \quad (5.10)$$

which is a LS problem with closed-form solution. To solve this problem we use an estimation of the sensitivity patterns  $\tilde{\mathbf{S}}$  and consider equivalently

$$\min_{\tilde{\mathbf{x}}'} \sum_{n=1}^N \|\mathbf{R}_n \tilde{\mathbf{S}}^H \tilde{\mathbf{F}}^H \tilde{\mathbf{x}}' - \mathbf{D} \boldsymbol{\gamma}_n\|_2^2 + \frac{\mu^2 N_p}{\beta^2} \|\tilde{\mathbf{x}}' - \tilde{\mathbf{x}}\|_2^2. \quad (5.11)$$

Denoting the functional to be minimised as  $f$ , it is a real-valued function of complex variables, for which the optimum solution for  $\tilde{\mathbf{x}}'$  is given by  $\frac{\partial f}{\partial \tilde{\mathbf{x}}'^H} = 0$ . Expanding equation (5.11), we have

$$\min_{\tilde{\mathbf{x}}'} \sum_{n=1}^N \left( \mathbf{R}_n \tilde{\mathbf{S}}^H \tilde{\mathbf{F}}^H \tilde{\mathbf{x}}' - \mathbf{D} \boldsymbol{\gamma}_n \right)^H \left( \mathbf{R}_n \tilde{\mathbf{S}}^H \tilde{\mathbf{F}}^H \tilde{\mathbf{x}}' - \mathbf{D} \boldsymbol{\gamma}_n \right) + \frac{\mu^2 N_p}{\beta^2} \left( \tilde{\mathbf{x}}' - \tilde{\mathbf{x}} \right)^H \left( \tilde{\mathbf{x}}' - \tilde{\mathbf{x}} \right), \quad (5.12)$$

and therefore, to satisfy  $\frac{\partial f}{\partial \tilde{\mathbf{x}}'^H} = 0$  it is required that

$$\left( \tilde{\mathbf{F}}\tilde{\mathbf{S}} \sum_{i=1}^N \mathbf{R}_n^T \mathbf{R}_n \tilde{\mathbf{S}}^H \tilde{\mathbf{F}}^H + \frac{\mu^2 N_p}{\beta^2} \right) \tilde{\mathbf{x}}' = \tilde{\mathbf{F}}\tilde{\mathbf{S}} \sum_{n=1}^N \mathbf{R}_n^T \mathbf{D} \gamma_n + \frac{\mu^2 N_p}{\beta^2} \tilde{\mathbf{x}}. \quad (5.13)$$

A number of observations can simplify the expression in (5.13). Assuming the wrap-around property of the patch extractor  $\mathbf{R}_n$ , the first term reduces to  $N_p \tilde{\mathbf{F}}\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H \tilde{\mathbf{F}}^H$ . Moreover, the matrix  $\sum_{n=1}^N \mathbf{R}_n^T \mathbf{D} \gamma_n$  is simply the added contribution of all coded patches relocated to their corresponding position within the dataset. Using the notation  $\tilde{\mathbf{x}}_D = \frac{1}{N_p} \tilde{\mathbf{F}}\tilde{\mathbf{S}} \sum_{n=1}^N \mathbf{R}_n^T \mathbf{D} \gamma_n$  to refer to the parallel k-space of the solution obtained when averaging the overlapping contributions of approximated patches  $\mathbf{D} \gamma_n, \forall n$ , dividing both sides by  $N_p$  gives

$$\left( \tilde{\mathbf{F}}\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H \tilde{\mathbf{F}}^H + \frac{\mu^2}{\beta^2} \right) \tilde{\mathbf{x}}' = \tilde{\mathbf{F}}\tilde{\mathbf{S}}\tilde{\mathbf{x}}_D + \frac{\mu^2}{\beta^2} \tilde{\mathbf{x}}. \quad (5.14)$$

This last expression can be solved via CG.

### 5.3.3 Implementation considerations

The three subproblems are iteratively alternated updating their respective optimisation variables, with the initialisation  $\tilde{\mathbf{x}} = \mathbf{0}$ ,  $\Gamma = \mathbf{0}$  and  $\tilde{\mathbf{x}}' = \text{SPIRiT}(\tilde{\mathbf{x}}_u)$ , where we use  $\text{SPIRiT}(\tilde{\mathbf{x}}_u)$  to denote the SPIRiT solution given undersampled data  $\tilde{\mathbf{x}}_u$ . To speed up the convergence of the CG steps, we use the zero-filled parallel k-space  $\tilde{\mathbf{M}}^T \tilde{\mathbf{x}}_u$  as initial solution in the first iteration and updated  $\tilde{\mathbf{x}}$  values in subsequent iterations. Convergence is assumed whenever the normalised consecutive change  $\frac{\|\tilde{\mathbf{x}}^{(i-1)} - \tilde{\mathbf{x}}^{(i)}\|_2}{\|\tilde{\mathbf{x}}^{(i)}\|_2}$  is smaller than  $10^{-4}$ , with superscript  $(i)$  denoting intermediate results at iteration  $i$ . For the estimation of the sensitivities needed we use the SVD method in [137] prior to reconstruction, although others could apply.

The tuning of three variables  $\lambda$ ,  $\beta$  and  $\mu$  in equation (5.7), which ideally has to provide optimal results and solution convergence, is problematic because it opens a large space for tuning

exploration. One way to reduce the number of tuning parameters is to let  $\beta \rightarrow \infty$  and force the variable  $\tilde{\mathbf{x}}' = \tilde{\mathbf{x}}$  instead of the weighted solution suggested by equation (5.14). With this modification, the variable  $\tilde{\mathbf{x}}'$  is treated as an intermediate solution that is close to  $\tilde{\mathbf{x}}$ , but which introduces the sparsity condition as a prior in the CG update of step 2. The final algorithm is compactly presented in algorithm 3. Empirical convergence is then easily observed, and the setting of remaining parameters  $\lambda$  and  $\mu$  can be defined from an estimation of noise power as will be seen in section 5.4.5. Only results using this simplified algorithm, referred to as D-SPIRiT, will be presented and discussed.

**Algorithm 3:** D-SPIRiT

**Input:** Undersampled k-space -  $\tilde{\mathbf{x}}_u$

**Output:** Reconstructed k-space -  $\tilde{\mathbf{x}}$

**Initialise:**  $\tilde{\mathbf{x}}^{(0)} = \tilde{\mathbf{M}}^T \tilde{\mathbf{x}}_u$ ,  $\mathbf{\Gamma}^{(0)} = \mathbf{0}$ ,  $\tilde{\mathbf{x}}'^{(0)} = \text{SPIRiT}(\tilde{\mathbf{x}}_u)$ , estimate  $\tilde{\mathbf{S}}$ .

**repeat**

1.  $i \leftarrow i + 1$

2. Update output k-space using CG with initial guess  $\tilde{\mathbf{x}}^{(i-1)}$ :

$$\tilde{\mathbf{x}}^{(i)} \leftarrow \min_{\tilde{\mathbf{x}}} \|\tilde{\mathbf{x}}_u - \tilde{\mathbf{M}}\tilde{\mathbf{x}}\|_2^2 + \lambda^2 \|(\mathbf{G} - \mathbf{I})\tilde{\mathbf{x}}\|_2^2 + \mu^2 \|\tilde{\mathbf{x}}'^{(i-1)} - \tilde{\mathbf{x}}\|_2^2$$

3. Update sparse coding with OMP:

$$\mathbf{\Gamma}^{(i)} \leftarrow \min_{\mathbf{\Gamma}} \|\mathbf{R}_n \mathbf{x}' - \mathbf{D}\mathbf{\gamma}_n\|_2^2 \quad \text{s.t.} \quad \|\mathbf{\gamma}_n\|_0 \leq s, \forall n$$

4. Recombine sparse coil images k-spaces:  $\tilde{\mathbf{x}}_D = \tilde{\mathbf{F}}\tilde{\mathbf{S}} \frac{1}{N_p} \sum_{n=1}^N \mathbf{R}_n^T \mathbf{D}\mathbf{\gamma}_n$

5. Update sparse prior:  $\tilde{\mathbf{x}}'^{(i)} \leftarrow \tilde{\mathbf{x}}_D$

**until**  $\frac{\|\tilde{\mathbf{x}}^{(i-1)} - \tilde{\mathbf{x}}^{(i)}\|_2^2}{\|\tilde{\mathbf{x}}^{(i)}\|_2^2} \leq 10^{-4}$ ;

## 5.4 Experiments and results

The objective of the following experiments is to provide an analysis of the proposed algorithm's performance, both quantitatively and qualitatively, and to study the advantages and drawbacks of using overcomplete and adaptive dictionaries for MRI reconstruction with respect to other sparsity transforms.

### 5.4.1 Experimental setup

We present result for both retrospective and prospective undersampling experiments. Results in sections 5.4.2 and 5.4.3 use retrospective undersampled data from a fully sampled simulated or real scan. The performance of reconstruction provided by the method presented is analysed at different sampling rates and under different SNR conditions, and is compared with the baseline SPIRiT algorithm and the regularised  $l_1$ -SPIRiT. In section 5.4.3 we also present results obtained while taking part in a reconstruction challenge relative to numerous undisclosed competing methods.

All undersampling masks used assume Cartesian k-space sampling. Different masks are used for each experiment, and benefits and drawbacks of each configuration are discussed. In figure 5.3 we provide as a reference an example of different 2D+t mask designs that will be used in the experiments below. For 2D phantom experiments, referring to one of these designs is equivalent to using a single 2D frame from one of the spatio-temporal masks shown. Undersampling in both phase and frequency encode directions can provide ideal sampling incoherence conditions, although acceleration is only effective when undersampling is performed in the phase encode direction.

In table 5.1 we compare distinctive characteristics of each mask design. We will see that, while it is important to maximise spatio-temporal sampling incoherence to enable CS reconstruction, k-space maximum sampling distance must also be taken into account. Given that SPIRiT recombines parallel MR data as a k-space deconvolution using a small kernel, allowing large gaps in k-space with no acquired data can be detrimental as the deconvolution process will have no information other than noise to synthesise missing samples. Also note that although a circularly shifted regular sampling is coherent in time, the periodic nature of the aliasing means that within a spatio-temporal patch the aliasing can vary very quickly through time.

Some D-SPIRiT reconstructions will be compared against  $l_1$ -SPIRiT. In order to have a direct comparison of the performance obtained with an overcomplete and adaptive transform instead

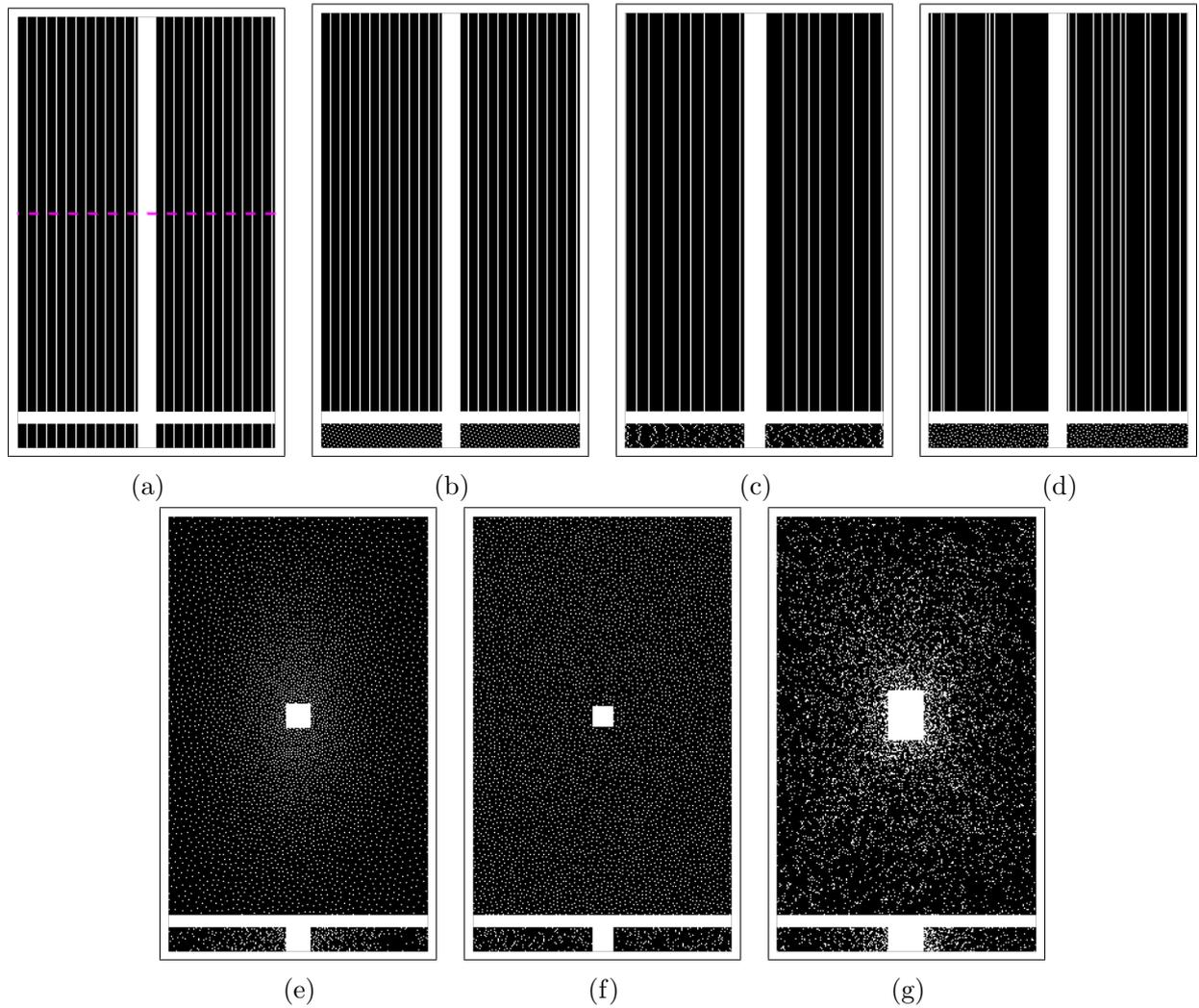


Figure 5.3: Mask designs used for testing: 1D regular (a), 1D regular circularly shifted (b), 1D uniform Poisson (c), 1D uniform random (d), 2D variable density Poisson (e), 2D uniform Poisson (f), and 2D variable density random (g).

of a complete and fixed transform, we rely on the D-SPIRiT algorithmic solution for the  $l_1$ -SPIRiT implementation. The only modifications between D-SPIRiT and  $l_1$ -SPIRiT are steps 3 and 4 in algorithm 3, where instead of using OMP to obtain an image that is sparse in a dictionary we enforce  $l_1$  sparsity solving

$$\min_{\tilde{\mathbf{x}}_{l_1}} \|\tilde{\mathbf{x}}' - \tilde{\mathbf{x}}_{l_1}\|_2^2 \quad \text{s.t.} \quad \|\mathcal{S}(\tilde{\mathbf{x}}_{l_1})\|_1 \leq s \quad (5.15)$$

This problem can be solved fast with unconstrained iterative shrinking methods such as [8].

Table 5.1: Subsampling mask design properties.

Mask design		Incoherence		Temporal aliasing	Sample distance
		Space	Time		
Regular	Constant	Low	Low	Constant Periodic	Minimal
	Circularly shifted				
Poisson disk	Uniform	Moderate	High	Random	Constrained
	Variable density				
Random	Uniform	High	High	Random	Unconstrained
	Variable density				

These methods require tuning for a weight on the sparse penalty, which was explored manually in all phantom tests individually to always compare against the best case scenario. For cardiac cine experiments, where fine tuning becomes time consuming, a unique value was set that was found to deliver optimal or close to optimal results in all cases. The sparse domain used for  $l_1$ -SPIRiT is the 2D Daubechies 4 transform domain for 2D data and x-f space 2D+t data, which is obtained with a Fourier transform along the temporal domain.

Some retrospective experiments include artificially added noise. Noise is assumed to be AWG, complex and circular, and independent for each coil. The metric we will use to refer to noise will be k-space SNR, which determines how much of the received signal power is attributed to noise. Formally, for the acquired signal  $\tilde{\mathbf{x}}_u = \tilde{\mathbf{M}}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}$ , we define k-space SNR as

$$k\text{SNR} = \frac{\mathcal{P}(\tilde{\mathbf{x}}_u)}{\mathcal{P}(\tilde{\mathbf{n}})} = \frac{\sum_{n=1}^{MNc} \|\tilde{x}_{u,n}\|_2^2}{\sum_{n=1}^{MNc} \|\tilde{n}_n\|_2^2}. \quad (5.16)$$

This metric allows to compare result at a given noise level irrespective of the undersampling rate.

The parameters for all D-SPIRiT reconstructions were kept constant, except for the tuning parameters  $\lambda$  and  $\mu$ , which are set based on the input noise power as is explained in section 5.4.5. Dictionaries of  $K = 196$  atoms were used to code patches of size  $N_p = 8 \times 8$  for images or  $N_p = 4 \times 4 \times 4$  for spatio-temporal data. The sparsity index was kept constant at  $s = 10$  for dictionary-based reconstructions, and the SPIRiT k-space convolution kernel size was  $5 \times 5$ .

Phantom experiments use simulated coil sensitivity patterns. These were generated by defining

coil centres to be distributed around the centre of the image, as shown in figure 5.4. A 2D Gaussian function was then used to simulate the magnitude of sensitivities, with variances adjusted in each dimension. At the centre of the coil the magnitude is 1 and far away it decays but does not reach 0. The phase of the coils was obtained from a 2D linear pattern that was rotated for each coil. The pattern was generated as shown in figure 5.4b, to traverse a full  $\pi$  rotation far away from the coil centre and a  $2\pi$  rotation close to the coil centre. The coil phase is a cropped section of this image pattern.

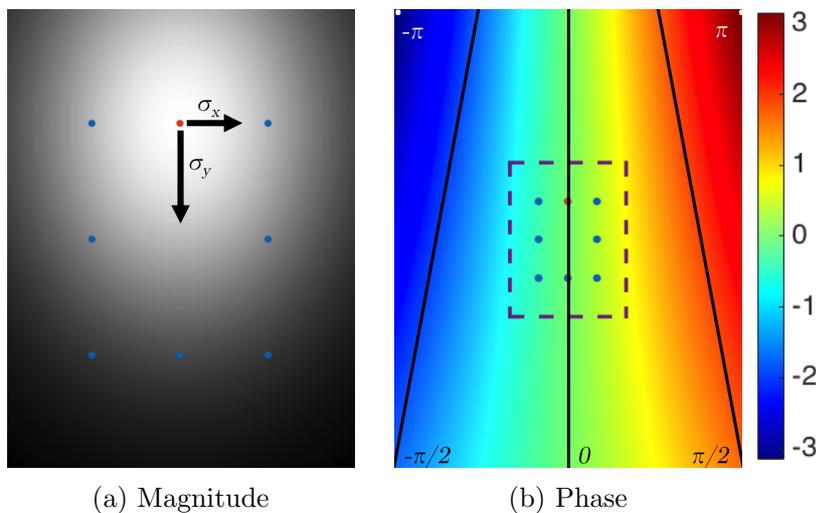


Figure 5.4: Simulated parallel coil sensitivity patterns. We assume the generation of 8 sensitivity patterns, where the 2D Gaussian function in (a) changes its center for different coils, and the phase image in (b) is rotated. The coil phase is a cropped section of the phase image pattern shown in (b).

### 5.4.2 Retrospective phantom tests

The first examination of the algorithm is performed with the Shepp-Logan phantom, which is an example image that can be very sparsely represented in many domains. Although exploiting redundancy in the temporal dimension is a great advantage of spatio-temporal dictionaries, a direct comparison with a complete, fixed-basis transform such as wavelets in a 2D setup is useful for the evaluation of the algorithm's behaviour.

A total of 8 parallel coils are assumed for this experiment. Figure 5.5 shows the magnitude ground truth image for one of those coil images simulated, alongside with the data after we add acquisition noise. Noise power was calculated such that the received k-space SNR was 30 dB. The visual comparison of the reconstruction results will refer to this setup, where each coil

image is undersampled using a 2D Poisson disk mask. Notice that with the chosen parameters, a SoS recombination of coil images produces an image with barely any noticeable noise as shown in figure 5.5d.

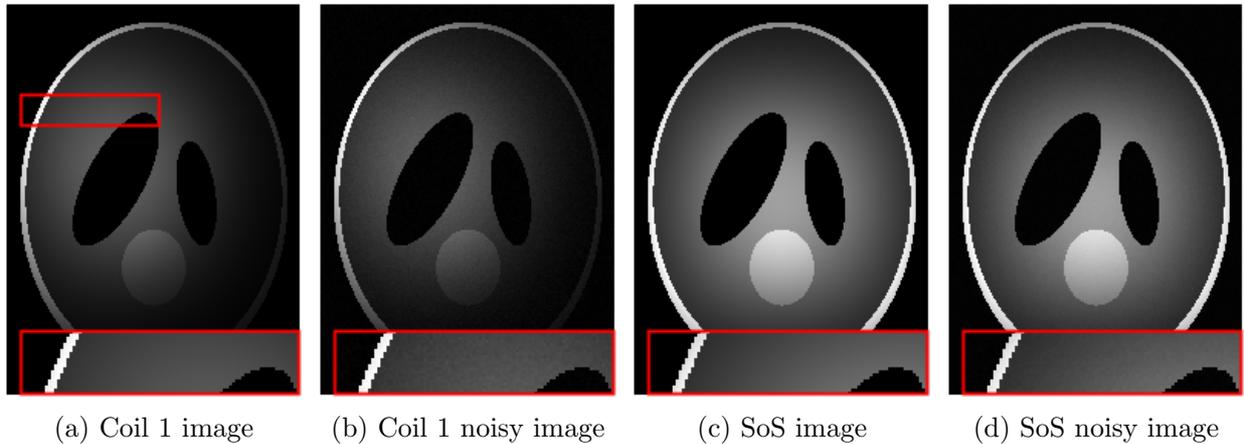


Figure 5.5: Data used for phantom experiments. Using 8 coils, we show one coil example ground truth (a), the image with added noise equivalent for a received k-space SNR of 30 dB (b), the SoS recombination ground truth (c) and the same recombination from noisy coil images (d).

Figure 5.6 compares reconstruction results for SPIRiT,  $l_1$ -SPIRiT, and D-SPIRiT with a DCT and a trained dictionary (DCT-SPIRiT and DL-SPIRiT respectively). For the training of the dictionary the ground truth image in figure 5.5c was used, hence simulating the scenario where ideal training data is available. Although the acceleration rate is below the theoretical maximum for perfect parallel reconstruction and the added noise is barely noticeable in the acquisitions, the numerical instability caused by coil geometry amplifies the noise in the SPIRiT reconstruction. A regularisation based on wavelets using  $l_1$ -SPIRiT can reduce some of this noise but presents some inaccuracies, particularly in high frequency regions. Using patch-based dictionary sparsity, as in figure 5.6c, we can control noise amplification better and this improvement should be attributed to the overcompleteness of the transform, which is much more flexible than a complete wavelet decomposition. A small advantage in image quality is gained by adapting the dictionary to this particular dataset through DL as shown in figure 5.6d.

In figure 5.7 we provide a quantitative evaluation of this comparison for a range of undersampling factors for 40 dB k-space SNR. The amplification of noise and unresolved aliasing

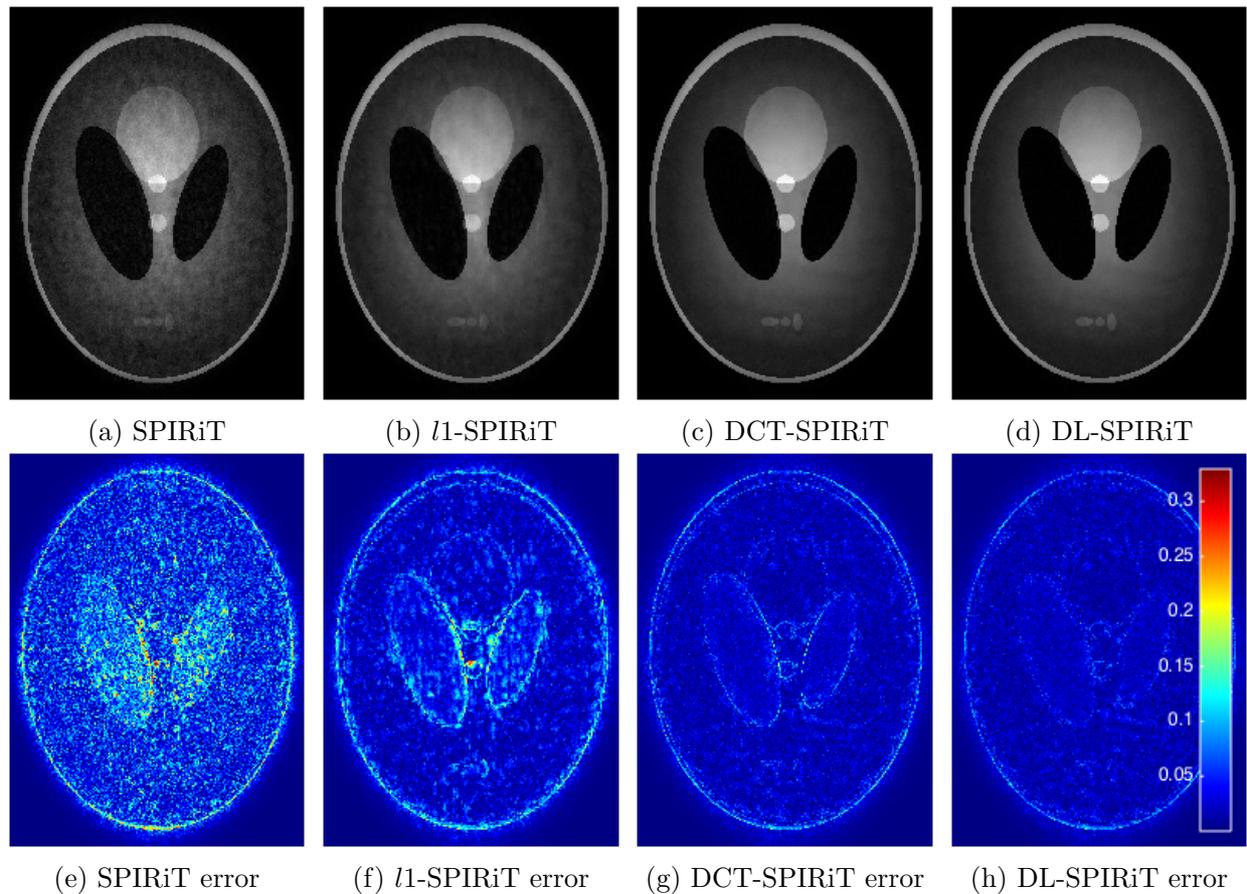


Figure 5.6: Reconstruction of a 6 fold accelerated 8 coil simulation of a phantom image with 30 dB k-space SNR. The SPIRiT reconstruction (a) suffers from noise amplification, which can be controlled by  $l_1$ -SPIRiT regularisation (b), but especially by an overcomplete (c) and adaptive (d) dictionary regularisation.

produces the decay of PSNR and MSSIM, but the rate of decay is milder for sparsity regularised reconstructions, particularly for the patch-based dictionary methods shown. The ability to preserve structural similarity at high accelerations by overcomplete sparsity transforms is notable. In figure 5.8 we show the same comparison for an acceleration rate of 6 and different k-space SNR values. In this comparison the decay of PSNR is similar for all methods, which maintain a roughly constant performance relative to each other, although again patch-based methods are able to preserve a high MSSIM metric even in very noisy regimes.

Figure 5.7: Performance evaluation of SPIRiT and its sparsity regularised versions for 40 dB k-space SNR at different undersampling ratios.

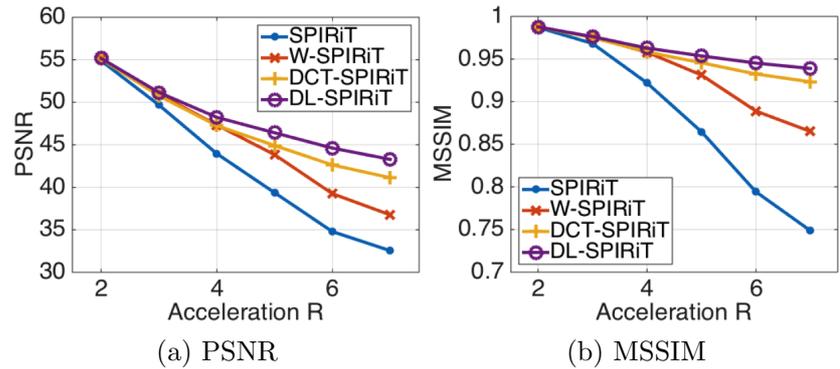
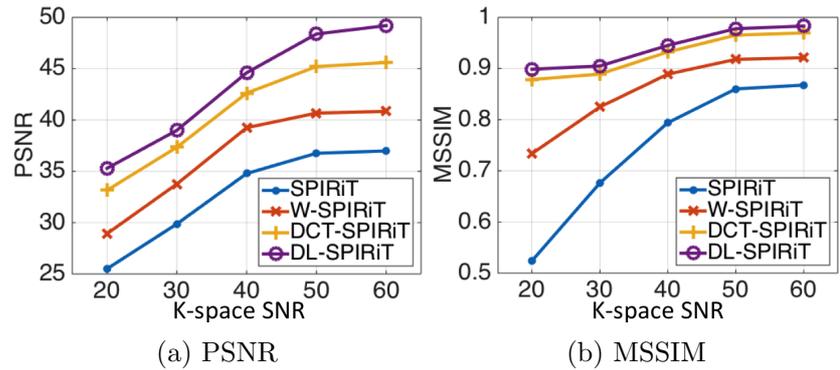


Figure 5.8: Performance evaluation of SPIRiT and its sparsity regularised versions for 6 fold acceleration at different k-space SNR.



### 5.4.3 Retrospective cardiac cine tests

In this section we present results with dynamic MR data. The undersampling is retrospective, hence allowing quantitative evaluation of the reconstructions. Two different sources of data are analysed, both acquired with a fully balanced steady-state free precession scan with breath-holding. Simulated noise was not added given that these datasets already contain acquisition noise.

#### Short-axis dataset tests

First, we present results on short-axis cardiac cine scans that were acquired using retrospective gating with 32 channels. The datasets contain 30 temporal frames of a single slice of thickness 4 mm. In-plane resolution was  $1.25 \times 1.25 \text{ mm}^2$  and the matrix size accounting for phase encode zero-padding was  $384 \times 192$ , corresponding to a  $640 \times 320 \text{ mm}^2$  FOV. For simulation purposes however the datasets were retrospectively cropped in image space to a matrix size of

$200 \times 189$ , getting rid of zero signal regions and allowing for faster reconstruction simulation. The undersampling masks applied to this data are 1D circularly shifted regular and 1D Poisson random undersampling.

In figures 5.9 and 5.10 we show one example of such reconstructions for acceleration rates of 6.25 and 7.69. In both cases, the noise amplification in the baseline SPIRiT reconstruction due to coil geometry restrictions is clear. The  $l_1$ -SPIRiT reconstruction, which exploits the x-f domain for sparse regularisation of the result also contains a considerable amount of aliasing and noise. This result can be modified according to how strongly the result is regularised, and although increasing sparsity regularisation can reduce noise it also smooths the reconstruction through time. This result is compared to the proposed DL-SPIRiT method, which uses a patch-based dictionary that has been trained *a priori* on a dataset from another patient. The reconstruction is cleaner than the  $l_1$ -SPIRiT method and at the same time can achieve good temporal resolution. The main drawback is the loss of some fine details, as shown in figure 5.10d, which can likely be attributed to the bad quality of the initial SPIRiT reconstruction and the lack of multiresolution in the patch-based dictionary.

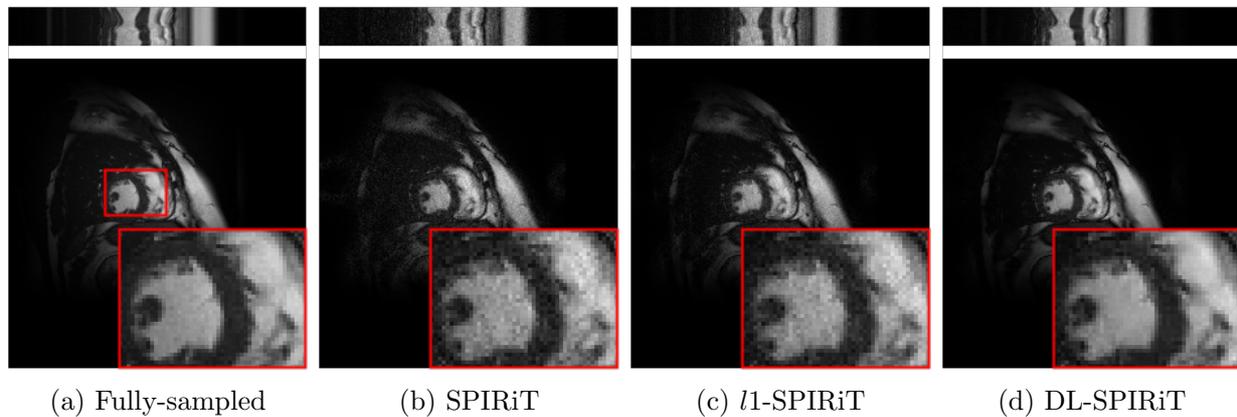


Figure 5.9: Reconstruction from 6.25 fold 1D regular circularly shifted undersampling.

In table 5.2 we provide a quantitative evaluation of this comparison for reconstructions from 7 different patients. The dictionary used for the DL-SPIRiT reconstructions was always the same, and was pre-trained on a single subject dataset. For all cases the patch-based method was shown to outperform the  $l_1$  regularised method, and the order of improvement in PSNR

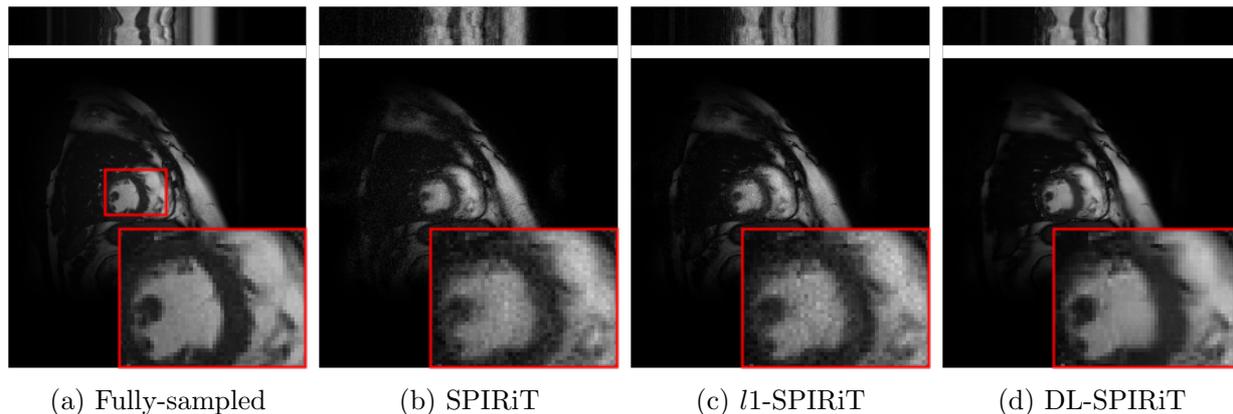


Figure 5.10: Reconstruction from 7.69 fold 1D regular circularly shifted undersampling.

of about 2 dB is similar to the one observed in the single channel simulations of chapter 4.

Table 5.2: PSNR (dB) reconstruction performance mean  $\pm$  standard deviation of retrospectively undersampled 32 coil data from 7 patients.

Sampling factor	SPIRiT	$l_1$ -SPIRiT	DL-SPIRiT
0.16	$33.64 \pm 1.88$	$36.02 \pm 1.98$	$38.31 \pm 2.01$
0.13	$32.14 \pm 1.86$	$34.71 \pm 1.97$	$36.09 \pm 1.91$

### ISMRM challenge dataset tests

The second dataset is available as part of the 2013 ISMRM sub-Nyquist cardiac cine reconstruction challenge [66]. This dataset contains 10 cases and is much more heterogeneous as it combines short-axis and long-axis scans, with undersampling patterns that are variable density random, Poisson disk random, and circularly shifted regular. However, the ground-truth fully-sampled data is not publicly available, so for evaluation we rely on visual inspection and on scores relative to other undisclosed algorithms.

The ISMRM challenge dataset allows testing the algorithm on a wider variety of examples, helping identify the strengths and weaknesses of the method. Specifications of each scan are given in table 5.3, showing the scan axis (SA: short axis, LA: long-axis), the undersampling rate and the mechanism used to generate the retrospective undersampling mask. A total of 22 teams entered the challenge, for which they had to upload reconstructions for the 10

undersampled cases. An automated system evaluated the performance of each reconstruction using the formula  $100 \times (1 - \text{RE})$ , where RE is reconstruction error relative to reference images as measured by a weighted Besov norm [39].

Table 5.3: ISMRM challenge data specifications.

Case	Scan axis	Undersampling rate	Mask type
1	LA	5	1D circularly shifted regular
2	LA	10	2D variable density random
3	LA	13	2D variable density Poisson disk
4	LA	6	1D circularly shifted regular
5	LA	5	1D uniform density random
6	SA	5	1D circularly shifted regular
7	SA	14	2D variable density Poisson disk
8	SA	5	1D uniform density random
9	SA	5	1D circularly shifted regular
10	SA	9	2D variable density random

Figure 5.11 shows the challenge results with our method highlighted as the black bottom row in each case. We applied the DCT-SPIRiT method in order to avoid dictionary training runtime. On average, the dictionary-based method was ranked 10<sup>th</sup>. One of the reasons for the modest average ranking was that for some cases the performance was relatively poor. This decrease in relative performance is particularly drastic for cases 5 and 8, which are the only examples for which a 1D uniform random sampling was used. This choice of mask, although it is beneficial from a CS perspective because it offers low sampling coherence in space and time, it is detrimental for k-space deconvolution methods such as SPIRiT. This is because the deconvolution kernel is smaller than some of the random gaps created in the acquisition, meaning that for those regions of k-space the only information available for reconstruction will be noise. The masks for these cases were provided as part of the data and are shown in figures 5.12c and 5.13c and reconstructions are shown in figures 5.12d and 5.13d, exhibiting unresolved aliasing artefacts. This corruption was already strong in the SPIRiT initialisation of the reconstructions due to inherent k-space deconvolution problems, and its regularisation is unable to resolve it.

Reconstructions are considerably better if a minimum distance between k-space samples is imposed, using for instance regular or Poisson disk random masks. For those cases, such as 4 and 7, the performance of the dictionary-based method was very close to the top performer.

These cases are shown in figures 5.12b and 5.13b. The SPIRiT reconstruction can provide a good result because the mask designs in figures 5.13a and 5.13c are particularly well-suited for parallel k-space deconvolution methods. However, it does introduce some noise and aliasing which is then cleaned by the dictionary sparsity regularisation.

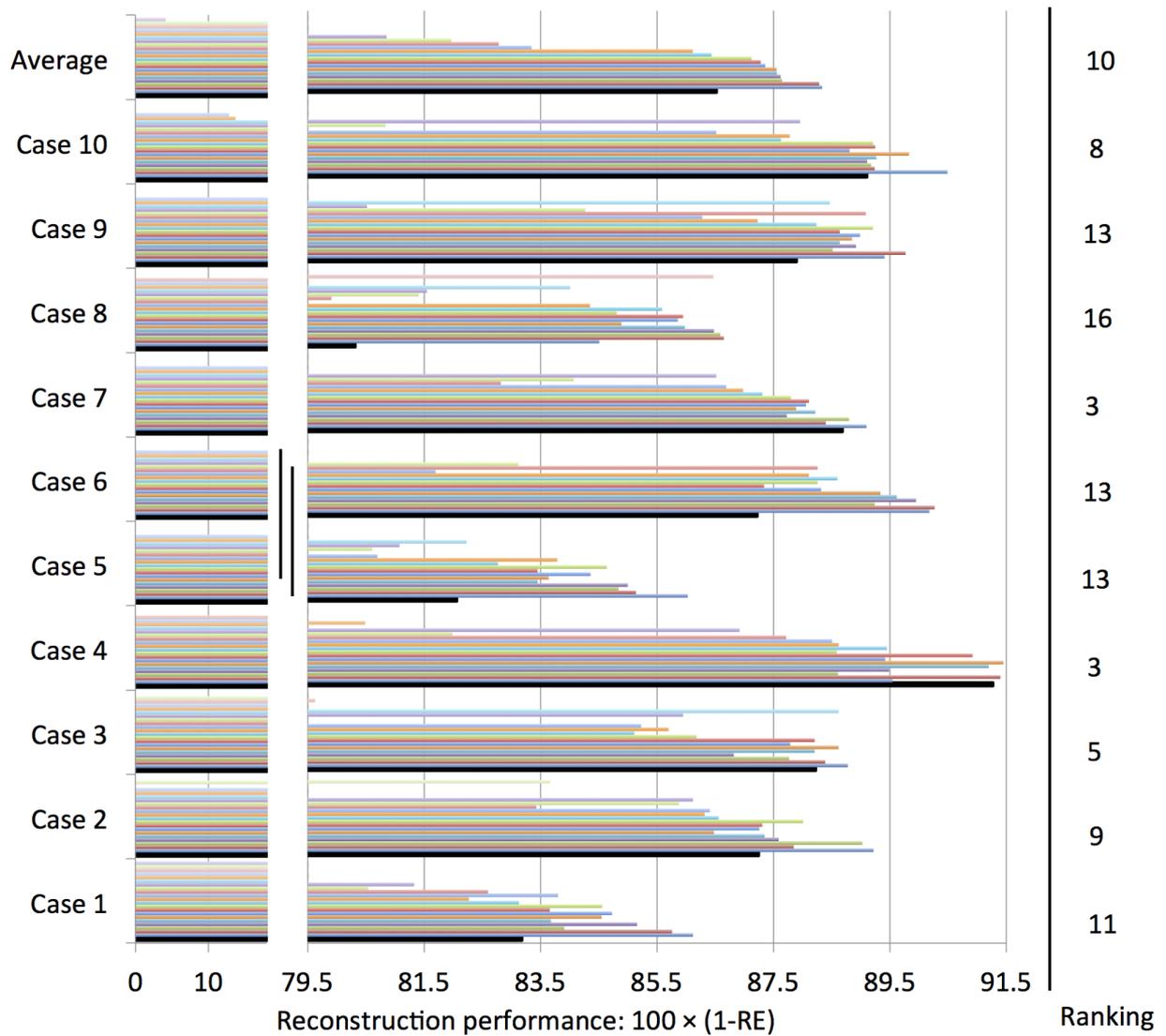


Figure 5.11: ISMRM cardiac cine reconstruction results. Coloured bars show reconstruction performance of competing methods, which are undisclosed, and the proposed technique D-SPIRiT is shown in black.

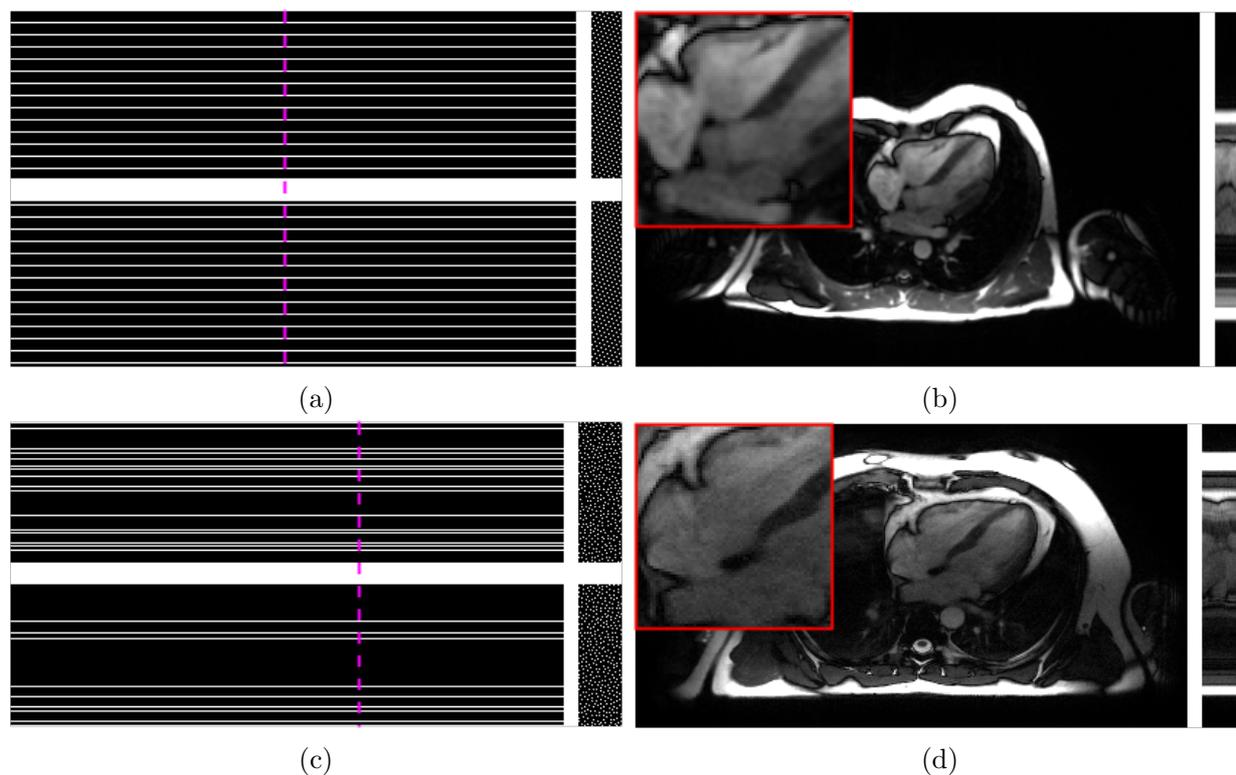


Figure 5.12: Examples of long-axis reconstructions from the ISMRM challenge dataset. We show the mask and DCT-SPIRiT reconstructions for cases 4 (a, b) and 5 (c, d).

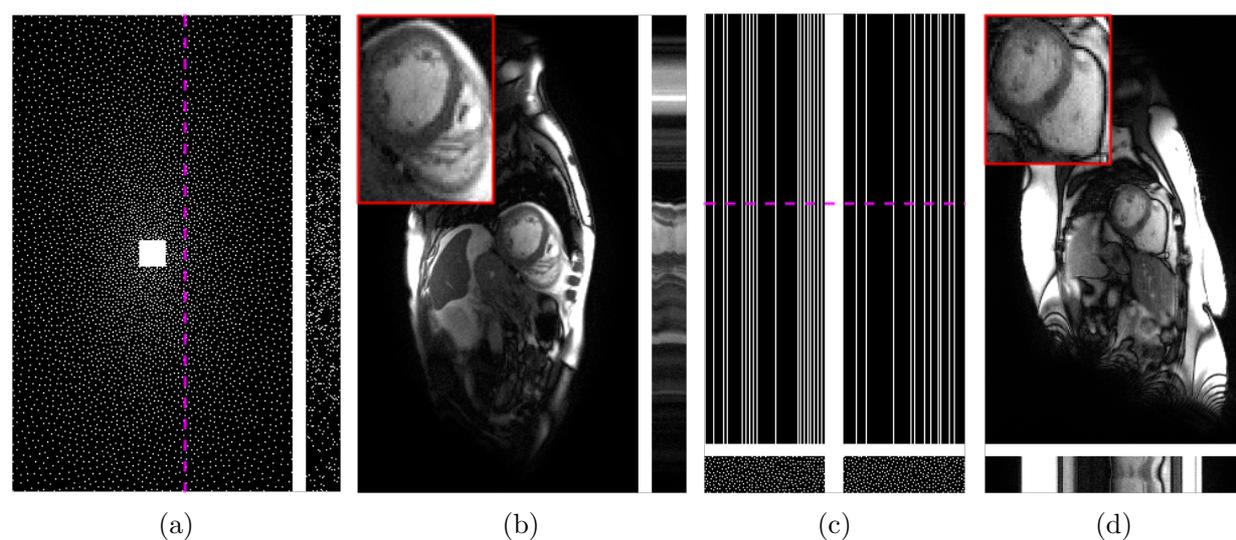


Figure 5.13: Examples of short-axis reconstructions from the ISMRM challenge dataset. We show the mask and DCT-SPIRiT reconstructions for cases 7 (a, b) and 8 (c, d).

#### 5.4.4 Prospective cardiac tests

We now turn to the application of the method in a practical and realistic setup. Commercial scanners do not readily allow random undersampling as is required for CS reconstruction, but we present results using typical undersampling patterns for SENSE and k-t SENSE, which are respectively regular static undersampling and regular circularly shifted undersampling. These experiments underline the importance of incoherent undersampling for the reconstruction method presented.

##### Regular undersampling

The data for these experiments was acquired in a 32 coil Philips Achieva scanner, using a balanced turbo field echo sequence. A total of 30 frames of size  $256 \times 256$  were produced with resolution  $1.25 \times 1.25 \text{ mm}^2$ , resulting in a  $320 \times 320 \text{ mm}^2$  FOV, and the slice thickness was 8 mm. A single patient was scanned at different undersampling rates with masks from the SENSE acquisition method provided by the scanner. The SENSE undersampling protocol uses a regular 1D undersampling mask that is invariant through time, hence does not comply with the incoherence requirements necessary for a CS reconstruction. This means that the reconstructions inevitably contain coherent aliasing, that is unresolvable through sparse regularisation, but these experiments provide nevertheless an assessment of the powerful noise reduction enabled by the method proposed. Furthermore, SENSE undersampling does not provide a fully sampled k-space region. We therefore learn the k-space deconvolution weights necessary for SPIRiT and estimate coil sensitivities from a fully sampled region in the centre of k-space of the fully sampled scan, hence simulating the undersampling mask in figure 5.3a.

A total of 12 short-axis slices were acquired as a stack scan for one patient, where raw data was first fully acquired, and then prospectively undersampled at ratios 2 to 8<sup>1</sup>. The acceleration

---

<sup>1</sup>Using standard scan parameters, a full slice could be reconstructed from a single heart beat with an undersampling of 4 fold. For this reason the protocol was changed from the standard 67% to 100%, meaning that the required 30 frames were reconstructed and not subdivided as is common practice. This resulted in a real scan acceleration from undersampling and avoided temporal blurring that can come from phase percentages

enabled by undersampling was translated by the number of slices that could be acquired per breath-hold or by the effective phase encode resolution. In table 5.4 we provide a short description of the acquisitions performed, where we see total acquisition time is effectively reduced up to a 6 fold acceleration, and the acceleration of the 8 fold case was used to improve the phase encode resolution relative to the 6 fold case.

Table 5.4: Prospective data with regular static undersampling acceleration specifications.

Case	Total scan time	Slices / BH	Scan time / BH	Scan time reduction	Scan percentage
x1	03:29.5	1	00:17.5	1	121%
x2	01:57.8	2	00:19.6	1.78	125%
x4	01:05.5	3	00:16.4	3.20	123%
x6	00:39.3	4	00:13.1	5.33	92%
x8	00:39.3	4	00:13.1	5.33	123%

We show one frame of this dataset in figure 5.14a. The SPIRiT and DCT-SPIRiT reconstructions are compared to the scanner’s baseline SENSE reconstruction in figures 5.17 and 5.18 for the reconstruction of a single slice. Given that aliasing artefacts are coherent, DCT-SPIRiT should not be expected to remove them. However, these results show how the patch-based regularisation of SPIRiT can enhance its reconstruction quality by reducing noise amplification. This is particularly noticeable in figure 5.18 for the 6 fold acceleration case, where the SENSE and SPIRiT reconstructions (figures 5.18g and 5.18h) contain a considerable amount of corruption that the DCT-SPIRiT method can alleviate (figure 5.18i). It can be expected, based on retrospective undersampling tests, that the remaining aliasing is resolvable using a circularly shifting or a Poisson sampling.

### Regular circularly shifted undersampling

An additional test was run where data was acquired using the k-t SENSE protocol, which uses a regular undersampling pattern that is circularly shifted through time. The data was again acquired on a 32 coil Philips Achieva scanner, using a balanced, prospectively gated and breath-held acquisition. The acquisition is divided into two stages: an undersampled acquisition lower than 100%.

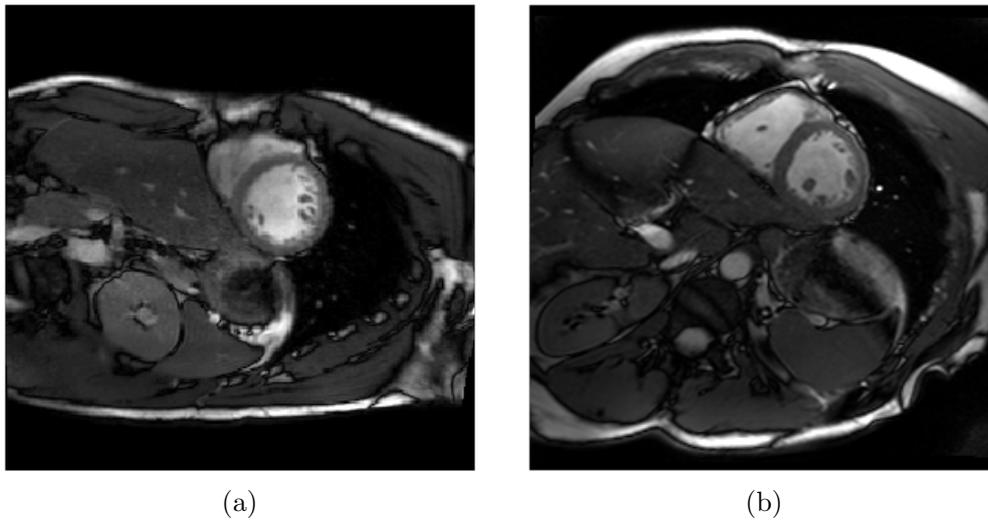


Figure 5.14: Single frames from original fully-sampled datasets used in prospective undersampled tests. We examine results using regular static undersampling on dataset (a) and regular circularly shifted undersampling on dataset (b).

stage to collect regularly undersampled data across the entire k-space over multiple cardiac cycles, and a training stage that collects 15 lines of fully-sampled central k-space data over two cardiac cycles. The two stages are separately acquired during different cardiac cycles, but are combined to make up an undersampling pattern similar to the one shown in figure 5.3b. The undersampling only affects the first stage, where data is allowed to be sparsely acquired, and translates into less number of cardiac cycles needed to collect the required k-space data. This datasets reconstruct 24 dynamic phases of the cardiac cycle. Details on the acquisitions analysed are provided in table 5.5.

Table 5.5: Prospective data with regular circularly shifted undersampling acceleration specifications.

Case	Total scan time	Cardiac cycles	Time / dyn. phase (s)
x1	00:20.3	20 + 2	0.845
x2	00:11.1	10 + 2	0.462
x4	00:06.4	5 + 2	0.226
x6	00:05.5	4 + 2	0.229
x8	00:04.6	3 + 2	0.192

The reconstructions are shown in figures 5.19 and 5.20, where in this case the dictionary was pre-trained on a dataset from a different patient. The comparison is made against the k-t SENSE reconstruction that is available directly from the scanner. In figure 5.19, we can see

how the k-t SENSE reconstruction provides good visual appearance for all accelerations up to case x8, whereas SPIRiT, shown in the middle column contains considerable noise and aliasing for x6 and x8, as was similarly seen in previous simulations. The patch-based regularisation is able to reduce the distortion caused by these corruption. An important difference is better appreciated in figure 5.20, where we also plot a temporal profile crossing the left ventricle. It can be seen that, although k-t SENSE produces a good in-plane result, the temporal resolution is dramatically affected by high acceleration. K-t SENSE regularises the reconstruction with an estimate of the x-f support from the training scan, but the acquired data is not enough to faithfully reconstruct the dynamism of the scan. In contrast, the DL-SPIRiT method proposed is able to much better reproduce the temporal pattern delineated by the ventricle even in the x8 case. Although remaining aliasing is not as destructive as with regular static undersampling, some of it is still unresolved, most likely because of the coherent undersampling strategy. Also, fine details are again lost, but just as with previous examples the coarse structure of the data and the dynamism is well preseved.

### 5.4.5 Parameter selection

The choice for  $\lambda$  and  $\mu$  was performed based on an estimation of the received k-space SNR. Given that this is a measure that is independent of the undersampling ratio and data dimensions, universal parameters could be chosen irrespective of the nature of the data. In figure 5.15 we show SPIRiT performance against varying choices of  $\lambda$  at three different k-space SNR values. The datasets used are one cardiac cine scan (solid) and the 2D phantom (dashed). The optimal tuning of  $\lambda$  is roughly dependent exclusively on input noise, irrespective of the data and the sampling ratio. A higher level of noise demands stronger regularisation of the data. Setting  $\lambda$  to the optimal choice based on these plots and running a similar test for  $\mu$  results in another set of optimal  $\mu$  choices, again independent on test conditions. These plots can therefore be used to generate look-up tables for  $\lambda$  and  $\mu$  that provide optimal choices based on noise estimation, as shown in figure 5.16

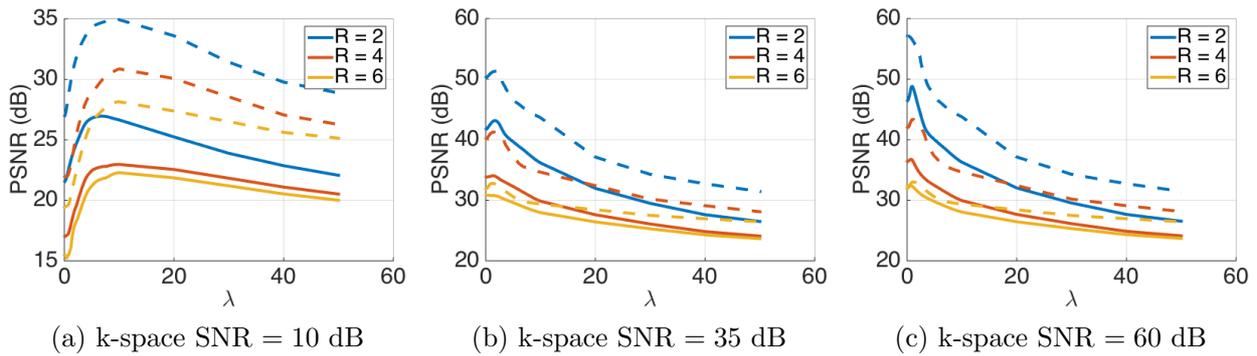


Figure 5.15: Exploration of tuning parameter  $\lambda$  in the SPIRiT initialisation for the D-SPIRiT algorithm. The tuning is dependent on the added k-space SNR, but not on the nature, dimension of the data, or undersampling rate, as shown using a cardiac (solid) and the Shepp-Logan phantom (dashed).

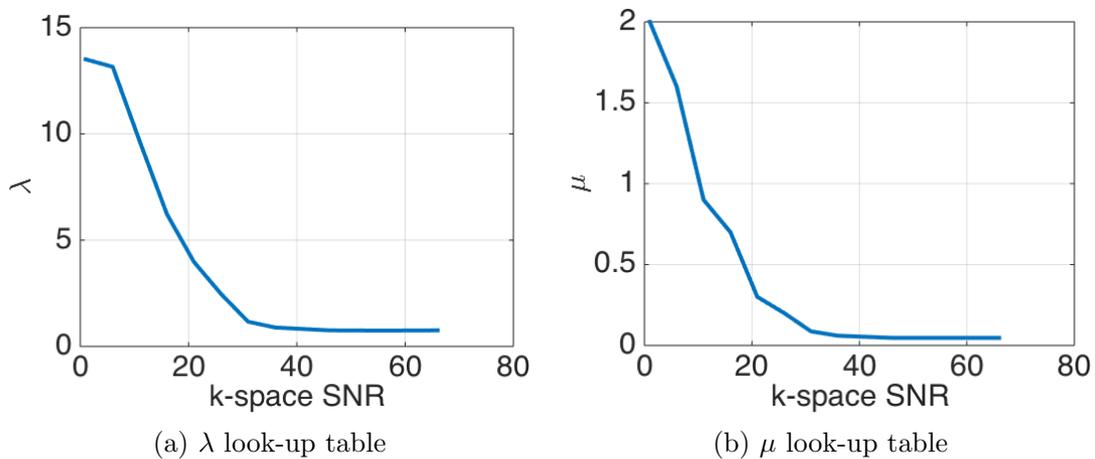


Figure 5.16: Look-up tables for tuning parameters derived from manual explorations of different noise levels.

Given that magnetisation signal is assumed to be very low at high frequencies, noise power was estimated for each experiment as the signal variance of data from the highest 10% frequencies. Knowing noise power and the total power received we could then derive k-space SNR. This results in an accurate estimation at low SNR, when noise in high frequencies is overwhelmingly larger than desired data, and less accurate at high SNR. However, in this latter regime optimal tuning parameters converge as shown in figure 5.16. In practice, a noise preparation scan is provided prior to imaging.

### 5.4.6 Algorithm speed and acceleration

The most intensive modules of the algorithm are the update of sparse coding with OMP and the update of the reconstruction in multi-coil k-spaces with CG. The former module is the same as the one discussed in the previous chapter in a single coil setup, and therefore the same ideas for acceleration presented in section 4.5.7 apply in this case. The update of the reconstruction with CG is derived from the SPIRiT methodology. In the baseline SPIRiT reconstruction proposed in [85], a single run of this descent algorithm is needed, but in our case it needs to be run iteratively to incorporate the sparse coding regularisation, and usually runs for about 50 iterations.

One main drawback of SPIRiT is that it looks for a solution in the multi-coil k-space domain, and hence its solution space grows with the number of coils. Using Matlab on a 8 GiB machine and 8 CPU cores, an 8 coil reconstruction of a 30 frame cardiac dataset can take about 5 minutes, but this algorithmic strategy lengthens reconstruction time by 6 times if the same data is reconstructed from 32 coils.

There are nevertheless possibilities for reducing runtime and making it independent from the number of coil channels used. One possibility would be to use a coil compression mechanism [160]. Given that complete coil orthogonality is not achievable, we can maximise the utility of the data acquired reducing the effective acquisition space to 8 or fewer virtual channels. This would reduce the solution space of SPIRiT and the algorithm could be used in the same way. Alternatively, we could try to use dictionaries to regularise a coil recombination method working in the space domain. The method ESPIRiT [137], for instance, has a strong relationship to SPIRiT but defines the solution space to be a single image or a combination of a few images in the space domain.

## 5.5 Conclusion

In this chapter we have introduced a method combining overcomplete and adaptive patch-based sparsity with parallel MRI reconstruction. The energy minimisation formulation of the SPIRiT reconstruction framework allows the straightforward addition of a dictionary sparsity term. We proposed a solution to the general problem formulation that splits it into three simpler subproblems and iterates their solution until convergence.

We demonstrated in retrospective undersampling tests using a 2D phantom image and 2D+t cardiac cine datasets that the algorithm proposed can outperform the baseline SPIRiT reconstruction, which is corrupted by noise amplification at high undersampling, and a CS regularised  $l_1$ -SPIRiT method, which uses a sparsity transform that is less flexible than overcomplete dictionaries. The method was validated in prospective undersampling cardiac cine tests, although only coherent undersampling from the SENSE and k-t SENSE protocol were analysed. Using regular static undersampling the method presented can enhance image quality compared to SENSE and SPIRiT by controlling noise amplification, but is theoretically unable to resolve coherent aliasing. When regular undersampling is circularly shifted, most of the aliasing corruption is attenuated but is again not fully resolvable, although the comparison against k-t SENSE and SPIRiT shows that DL-SPIRiT can faithfully recover coarse structure and dynamism. It is expected that more incoherent sampling mask designs such as Poisson random sampling could provide improved reconstructions in prospectively undersampled data, but this remains as future work.

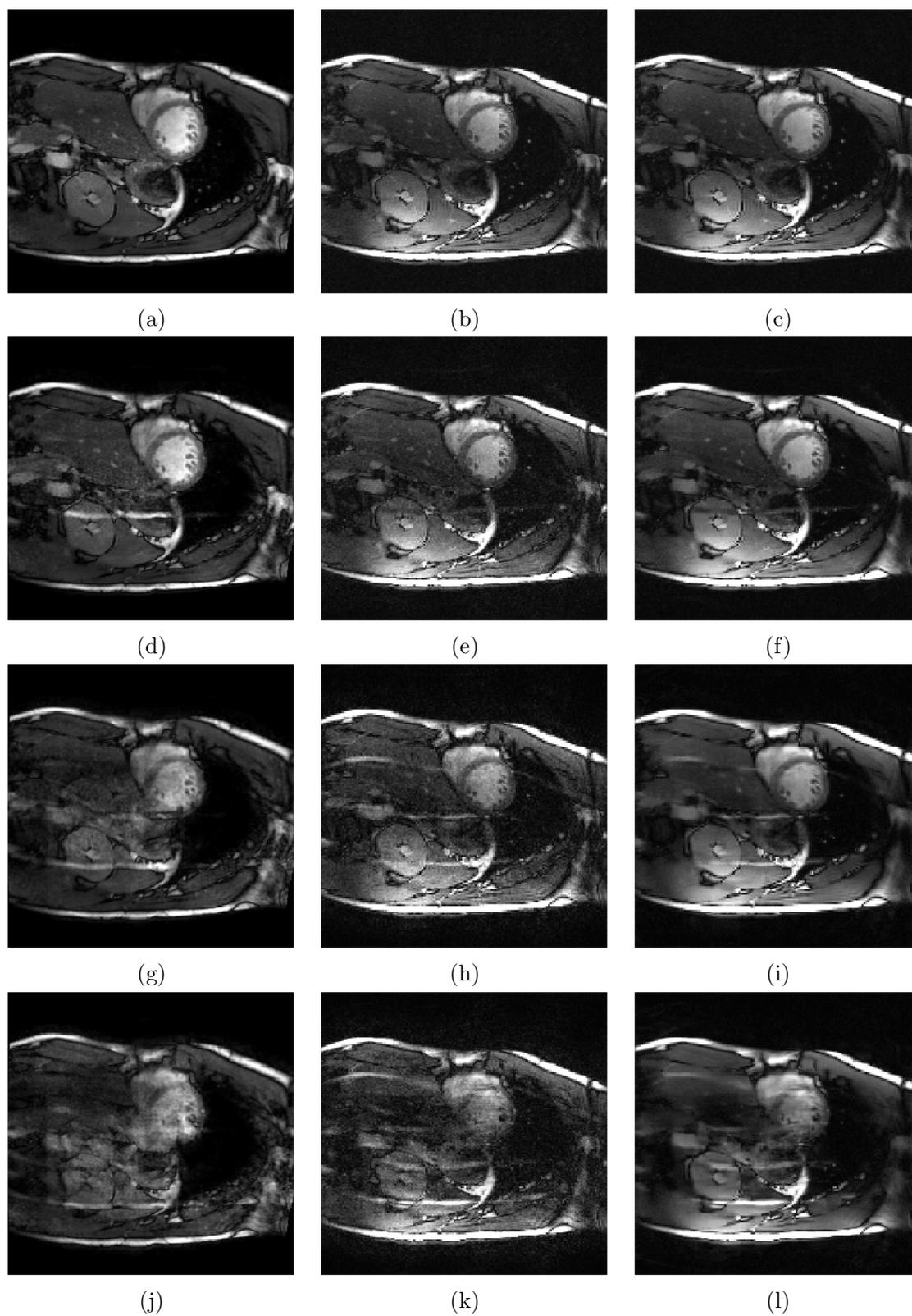


Figure 5.17: Prospective undersampling reconstructions using regular mask. The comparison is between SENSE (left column), SPIRiT (middle column) and DCT-SPIRiT (right column), and acceleration factors are 2, 4, 6 and 8, from top to bottom rows.

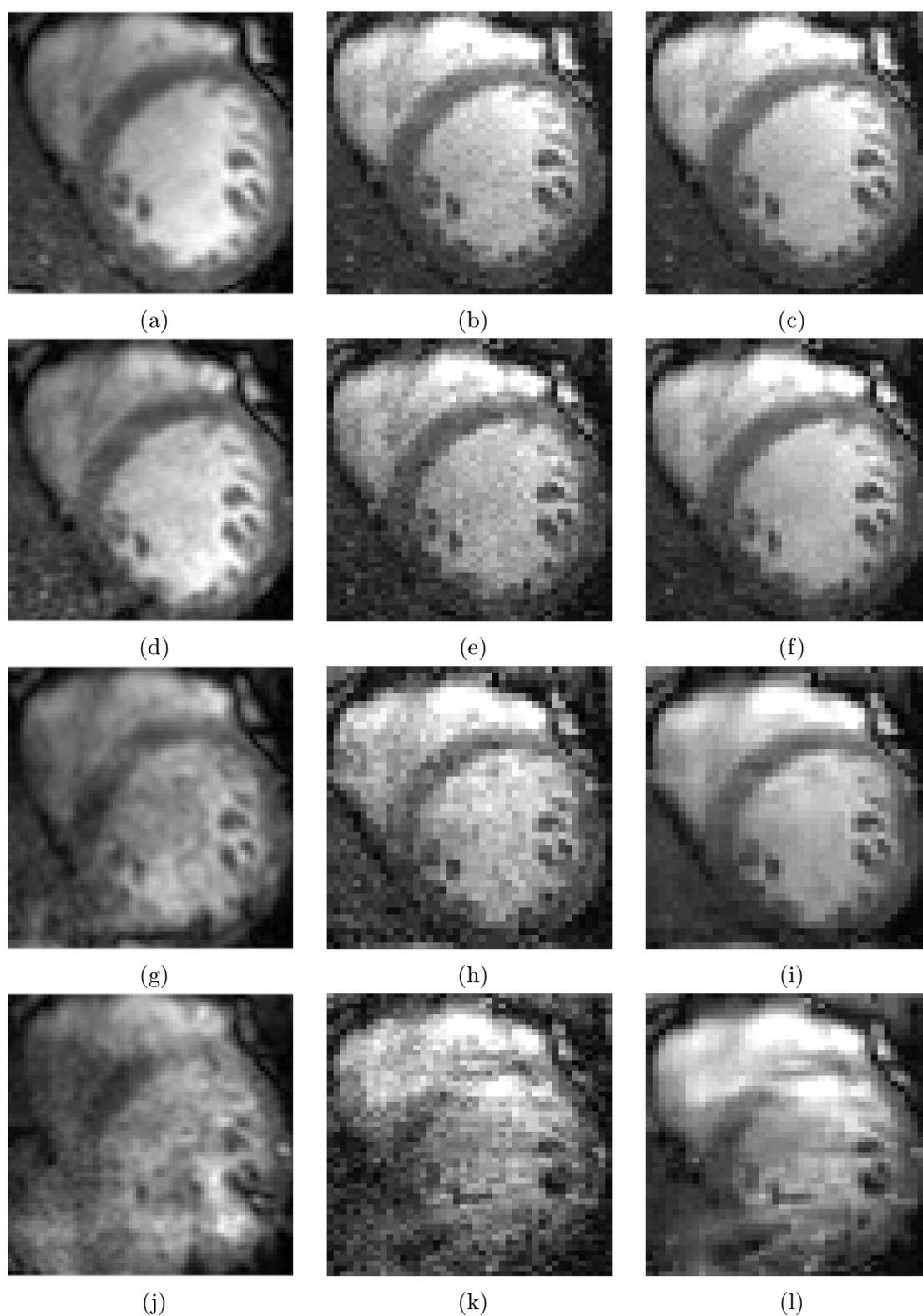


Figure 5.18: Zoomed ROI of prospective undersampling reconstructions using regular mask. The comparison is between SENSE (left column), SPIRiT (middle column) and DCT-SPIRiT (right column), and acceleration factors are 2, 4, 6 and 8, from top to bottom rows.

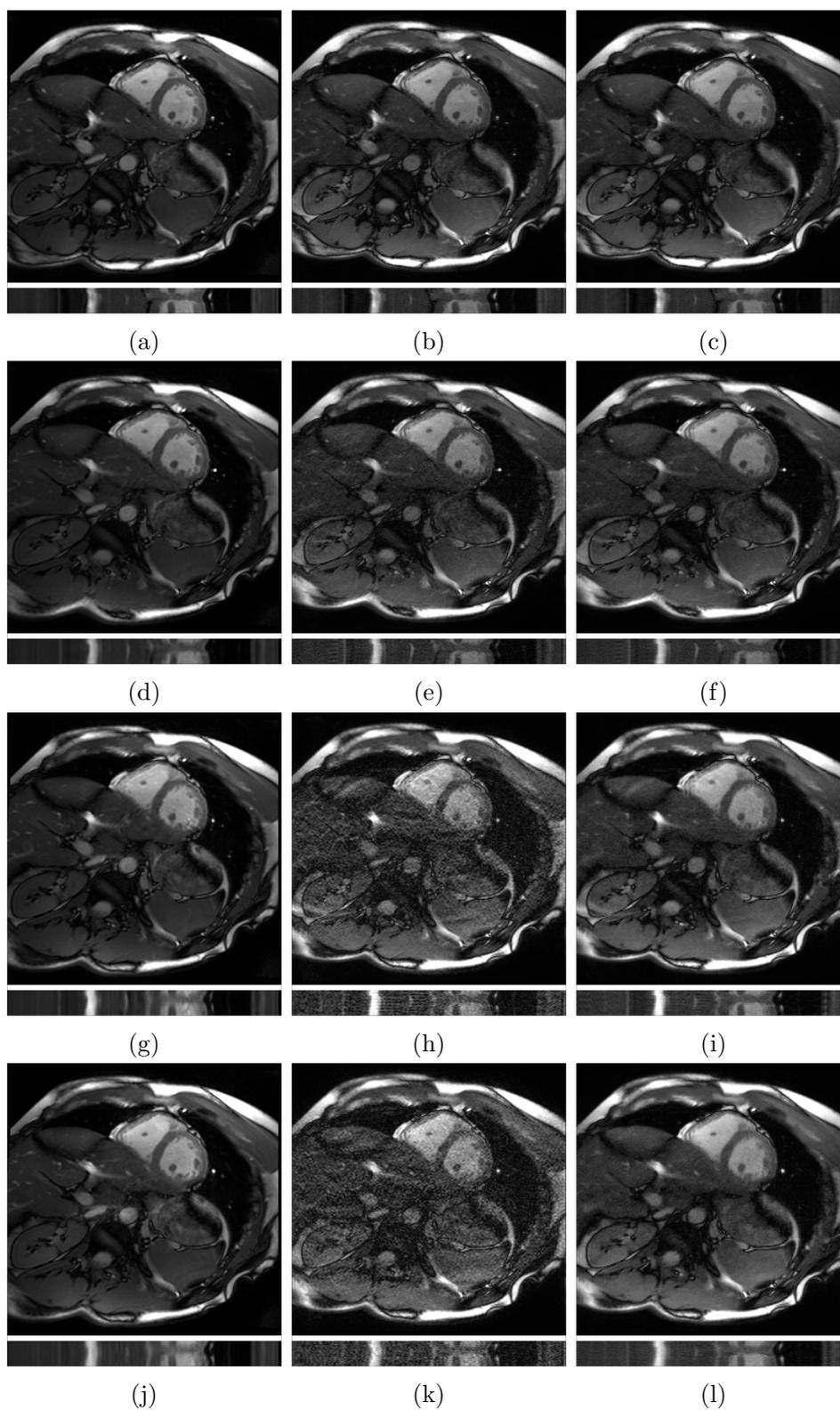


Figure 5.19: Prospective undersampling reconstructions using circularly shifting regular mask. The comparison is between k-t SENSE (left column), SPIRiT (middle column) and DL-SPIRiT (right column), and acceleration factors are 2, 4, 6 and 8, from top to bottom rows.

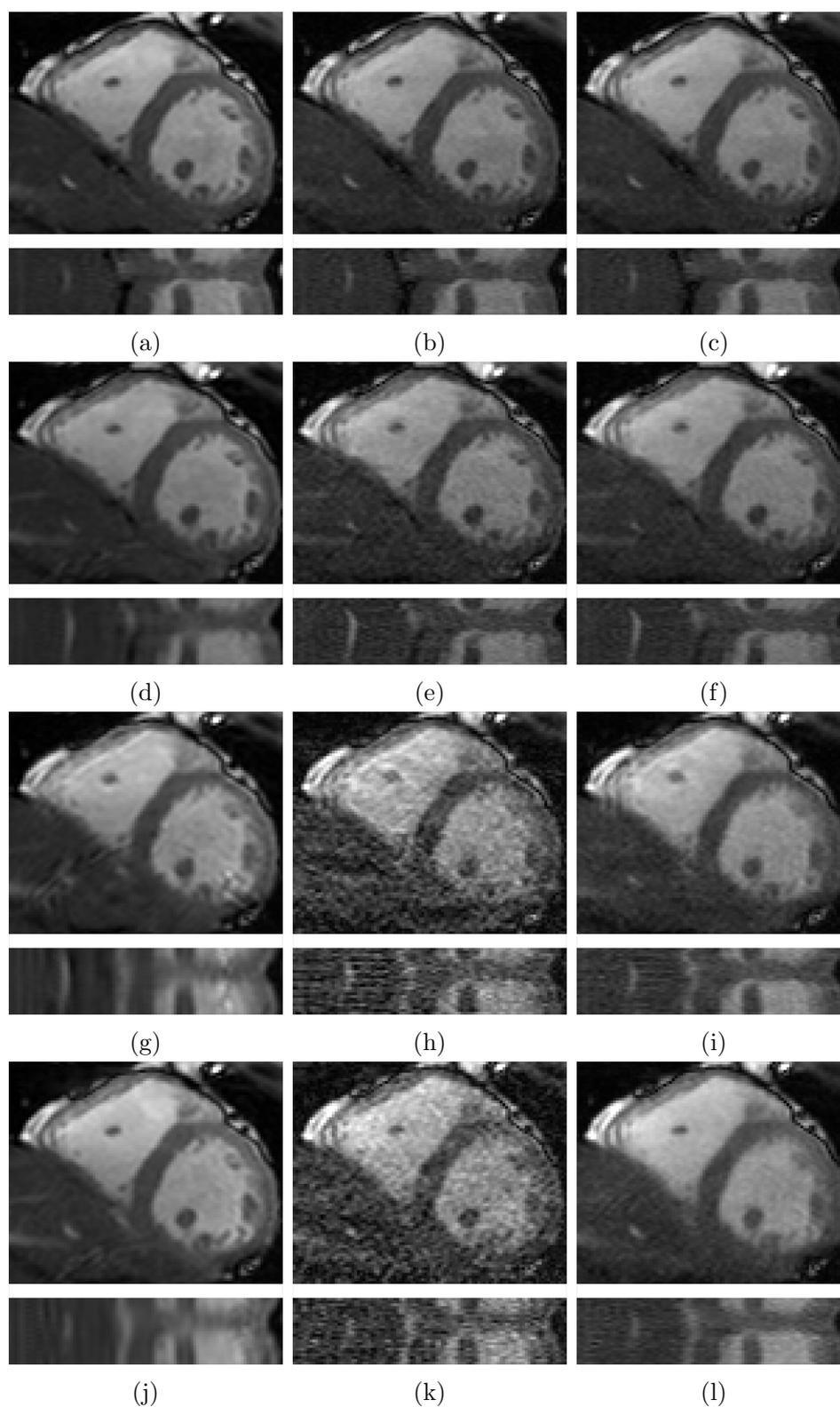


Figure 5.20: Zoomed ROI of prospective undersampling reconstructions using circularly shifting regular mask. The comparison is between k-t SENSE (left column), SPIRiT (middle column) and DL-SPIRiT (right column), and acceleration factors are 2, 4, 6 and 8, from top to bottom rows.

# Chapter 6

## Application-driven MRI

This chapter is based on the following publication:

- J. Caballero, W. Bai, A. N. Price, D. Rueckert, J. V. Hajnal, *Application-driven MRI: Joint reconstruction and segmentation from undersampled MRI data*, Proceedings of the 17th International Conference on Medical Imaging Computing and Computer Assisted Interventions (MICCAI), vol. 1, pp. 106-113, Boston, MA, USA, 14-18 September 2014.

### 6.1 Introduction

The journey and utility of a medical image can take many forms. A common pipeline can nonetheless be traced, where a patient or object is scanned, the data acquired is reconstructed into an image, this image can optionally be analysed, for instance to look for quantitative measures or to be registered and compared against another image, and the resulting information is interpreted by a physician for diagnosis or differential diagnosis. This traditional imaging pipeline is illustrated in figure 6.1.

Although this paradigm for diagnosis informed by imaging is accepted as the current norm, it can be improved at numerous fronts. An obvious limitation is error propagation, which is

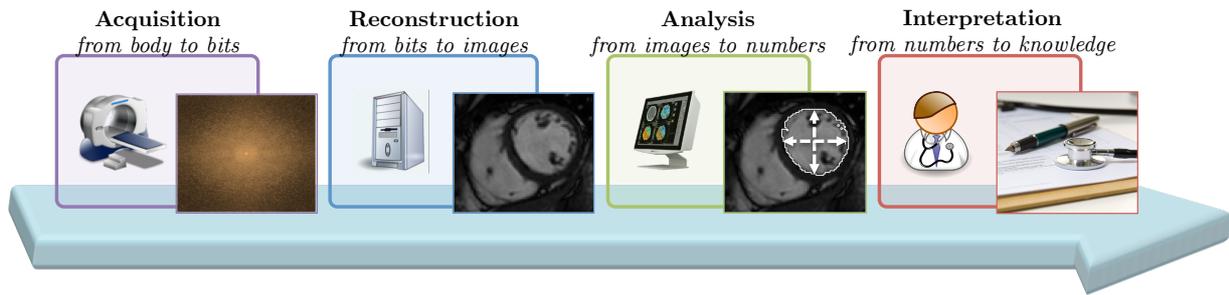


Figure 6.1: Traditional imaging pipeline.

inevitable in serial data manipulation. The imaging problem is ill-posed *per se*, and requires prior assumptions about the image carrying a degree of uncertainty about the aspect of the reconstructed image. This uncertainty is further aggravated by a myriad of factors, such as inherent systematic errors of the imaging method, image contrast limitations, or artefacts of motion, aliasing or scattering. Usually, it is impossible to include a measure of this uncertainty through stages hence leading to an accumulation of errors that is liable to contaminate interpretation stages.

Another point for discussion is the efficiency of the framework. Although the interpretation of imaging information is a very complex process, it is clear that an enormous dimensionality reduction exists when going from highly detailed images to clinical decisions based on a small number of quantitative measures. This would not be problematic if the data needed to reproduce an image was acquired at no cost, but we have already seen how, especially in the case of dynamic MRI, acquisition is a resource and time-consuming, expensive process.

This inefficiency has already been noted in the past, as highlights the following extract from a report on computational challenges of medical imaging from 2004 [130]:

The resulting images are now typically presented to experts (physicians) who use their extensive experience and professional judgment to extract knowledge from the pictures. Their interpretation is usually presented as a narrative qualitative appraisal, sometimes even only comprising a single bit of information (yes/no). We note for future reference that this analysis constitutes a very significant reduction in data volume, from a digital

image to a few succinct bits of pertinent extracted information.

It is important to note that imaging is often exploratory, and the capability for high-quality imaging of dense information is essential when the information required is unknown. However, in many situations such as in follow-up scans, specific quantitative measures informing interpretation are known to be of interest before data acquisition. There could therefore be scope for scanning procedures dedicated for a particular application, perhaps faster and more efficient than the current paradigm; an application-driven design.

The task of imaging data interpretation directly from raw MRI data is not simple. Diagnosis relies on countless factors ranging from the physicians expertise, to an analysis of physical conditions of the patient, or a subjective assessment of the situation relative to a history of medical records of the same or different patients. Automating diagnosis, although very desirable for the future, is an ambitious and delicate task. Furthermore, the Fourier relationship between acquisition and spatial domains in MRI being global, it is also non-trivial to design scan trajectories that are optimal for specific measures. Instead of considering the redefinition of the entire imaging process, we propose the integration of only reconstruction and analysis for the specific case of image segmentation.

This chapter is organised as follows. In section 6.2 we discuss how joint reconstruction and segmentation can be of interest in fast MRI scanning, and review similar problems already addressed in past literature. In section 6.3.1 we introduce the theory behind Gaussian mixture model (GMM) segmentation for images and apply it for undersampled MR data in section 6.4. We show results in section 6.5 of joint reconstruction and segmentation of accelerated 2D brain and cardiac cine MR data.

## 6.2 Joint reconstruction-segmentation

Image reconstruction and segmentation of MRI data have long been the subject of intense research activity. Image segmentation is usually treated as a distinct image processing stage rather than being integrated into the reconstruction stage even though its performance is fully conditioned by the quality and appearance of the latter. Similarly, the segmentation of an image often relies to some extent on identifying different intensity patterns, which is an information that is valuable to better condition the reconstruction problem.

We consider the problem of segmentation of undersampled MRI data. CS theory has been applied for fast MRI acquisition and reconstruction as discussed previously in this thesis. Even though this has great advantages as accelerated scanning can leverage constraints on patients, it devotes all resources into the approximation of an intensity valued image that does not explicitly reveal segmented information. Moreover, this approximation is never perfect, meaning that even if we carried out a segmentation of the non-linear reconstruction subsequently, adopting the traditional workflow, it is likely that reconstruction errors will propagate and influence segmentation. We investigate an alternative way of processing raw undersampled data, that could provide a segmentation both directly and accurately.

Simultaneous reconstruction and segmentation from undersampled measurements has already been studied in imaging methods where segmented data can easily be explained in a lower dimensional space. Hyperspectral imaging [158, 159], for instance, evaluates the response of a FOV across a spectrum of frequencies, hence requiring extensive sampling of the same image. Regions belonging to the same label will share a same signature across the frequency domain, and this generates a redundancy in the sampling process that can be exploited in a non-linear joint reconstruction-segmentation framework from undersampled data. Similarly, some tomographic images such as positron emission tomography images can be well approximated as piece-wise constant regions that directly relate to a label [62, 139].

We propose a non-linear joint reconstruction-segmentation of undersampled MRI using a com-

combination of a reconstructive and a discriminative term, which can trade intensity and segmentation accuracy. The reconstructive term of the algorithm employs a patch-based dictionary to impose sparsity on the reconstructed image. Simultaneously, the discriminative term promotes a representation where pixel intensities draw their values based on their labelling out of a few possible segmentation regions. For this we choose a GMM where pixel intensities can be explained by a linear combination of a few Gaussian distributions, and provides a simple segmentation mechanism where pixel labels are decided based on the probability that a pixel intensity was produced by a specific Gaussian. Discriminative representations obtained through methods different from a GMM could also apply within the general framework.

## 6.3 Mixture model segmentation

### 6.3.1 Gaussian mixture modelling

A GMM considers the data to be the realisation of an underlying probabilistic model that is a linear combination of Gaussian distributions. A recurrent assumption in image segmentation is that similar pixel intensities can be assumed to share the same label. This is often the case in brain MRI where individual tissues share similar intensity values [140, 141], or in cardiac cine where blood pool regions, often of most interest, reveal a characteristic high intensity relative to other anatomical regions such as the myocardium [122, 83].

Given observations  $\mathbf{x} = \{x_1, x_2, \dots, x_N\}$  to be pixel intensities, the assumption that each pixel intensity is the realisation of a single component from a mixture of  $L$  components can be formalised with the definition of a latent variable  $\mathbf{z} \in \mathbb{R}^{N \times L}$ . This variable is an indicator function relating one component from the mixture to data point  $x_n$ , such that for  $\mathbf{z}_n = \mathbf{e}_l$  data point  $n$  is assigned label  $l$ , where  $\mathbf{e}_l$  is 1 at entry  $l$  and 0 otherwise.

Let  $\mathcal{N}(x_n | \mu_l, \sigma_l)$  be a Gaussian distribution defined over  $x_n$  with mean  $\mu_l$  and standard deviation

$\sigma_l$ , compactly  $\theta_l = \{\mu_l, \sigma_l\}$ , we have

$$p(x_n | \mathbf{z}_n = \mathbf{e}_l, \theta_l) = \mathcal{N}(x_n | \theta_l), \quad (6.1)$$

and

$$p(x_n | \boldsymbol{\theta}) = \sum_{l=1}^L p(\mathbf{z}_n = \mathbf{e}_l) p(x_n | \mathbf{z}_n = \mathbf{e}_l, \theta_l) = \sum_{l=1}^L \pi_l \mathcal{N}(x_n | \theta_l), \quad (6.2)$$

where  $\pi_l = p(\mathbf{z}_n = \mathbf{e}_l)$  are mixture component weights and  $\boldsymbol{\theta} = \{\theta_l, \pi_l; l = 1, \dots, L\}$  is the complete set of model parameters. Notice that mixture weights are not dependent on location  $n$ , but can optionally become spatially dependent as will be shown in section 6.4.5. Assuming all pixels intensities to be generated independently, we have

$$p(\mathbf{x} | \boldsymbol{\theta}) = \prod_{n=1}^N p(x_n | \boldsymbol{\theta}). \quad (6.3)$$

The ML estimate of the model parameters  $\boldsymbol{\theta}$  is given by the marginal likelihood of the observation

$$\arg \max_{\boldsymbol{\theta}} \ln p(\mathbf{x} | \boldsymbol{\theta}) = \arg \max_{\boldsymbol{\theta}} \ln \left\{ \sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z} | \boldsymbol{\theta}) \right\}. \quad (6.4)$$

To illustrate the problem take the toy example given in figure 6.2, where the histogram of the synthetic image in figure 6.2a with intensities generated from Gaussian distributions is well approximated by a GMM in figure 6.2b. If model parameters  $\boldsymbol{\theta}$  are given, the optimal labelling of each pixel  $x_n$  in the absence of other prior knowledge (e.g. smoothness), is given by  $\arg \max_l \pi_l \mathcal{N}(x_n | \theta_l)$ , shown in figure 6.2c. The task is therefore to simultaneously find parameters  $\boldsymbol{\theta}$  from the observations and the hidden variable  $\mathbf{z}$  providing the segmentation as a by-product.

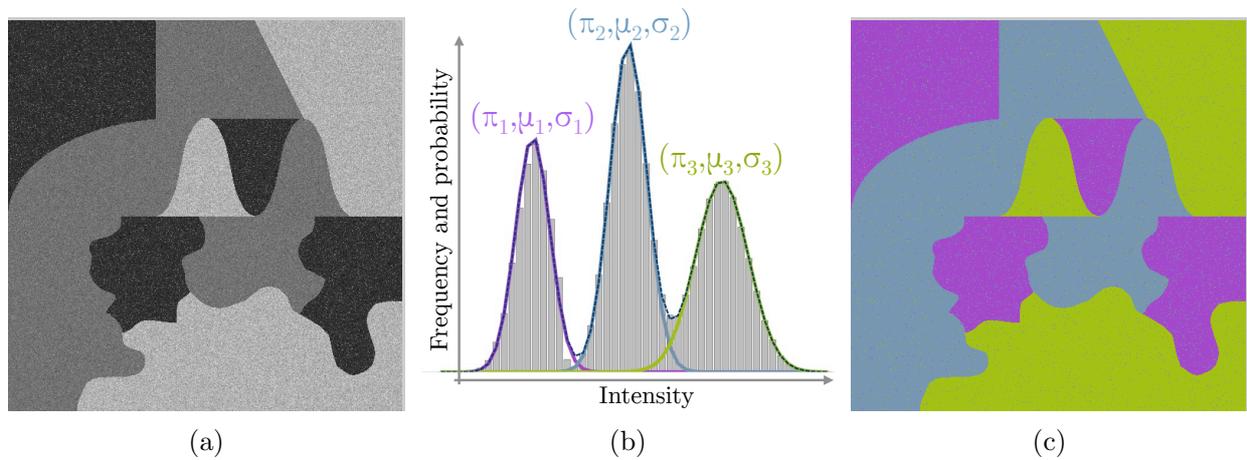


Figure 6.2: Toy image segmentation with GMM. A  $L = 3$  component mixture model (b) is fit to the histogram of intensity image (a) producing segmentation (c).

### 6.3.2 EM for GMM fitting

The solution of equation (6.4) is complicated by the sum inside the logarithm, which prevents a closed-form solution when setting derivatives to zero. If we were given the complete data set  $\{\mathbf{x}, \mathbf{z}\}$  the maximisation of the complete data log-likelihood  $\ln p(\mathbf{x}, \mathbf{z}|\boldsymbol{\theta})$  would be straightforward, but only the incomplete data set  $\mathbf{x}$  is at hand. Notice however that we can evaluate membership weights  $\mathbf{w} \equiv p(\mathbf{z}|\mathbf{x}, \boldsymbol{\theta})$  given the observations and a current estimate of the parameters using Bayes' rule. It is therefore possible to iteratively solve equation (6.4) by alternating the update of the complete data expectation under the posterior distribution of the latent variable, and the update of the resulting function maximisation.

This is formally known as the expectation-maximisation (EM) algorithm, iterating two steps:

- **Expectation step:** Compute the function

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \sum_{\mathbf{z}} p(\mathbf{z}|\mathbf{x}, \boldsymbol{\theta}^{(i)}) \ln p(\mathbf{x}, \mathbf{z}|\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{z}|\mathbf{x}, \boldsymbol{\theta}^{(i)}} \ln p(\mathbf{x}, \mathbf{z}|\boldsymbol{\theta}) \quad (6.5)$$

- **Maximisation step:** Evaluate new parameters as

$$\boldsymbol{\theta}^{(i+1)} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(i)}) \quad (6.6)$$

The log-likelihood  $\ln p(\mathbf{x} | \boldsymbol{\theta})$  is guaranteed to increase [152]. For the special case of GMMs, membership weights based on estimation  $\boldsymbol{\theta}^{(i)}$  are given in the expectation step by

$$w_{n,l}^{(i)} \equiv p(\mathbf{z}_n = \mathbf{e}_l | x_n, \boldsymbol{\theta}^{(i)}) = \frac{p(\mathbf{z}_n = \mathbf{e}_l) p(x_n | \mathbf{z}_n = \mathbf{e}_l, \theta_l^{(i)})}{\sum_{j=1}^L p(\mathbf{z}_n = \mathbf{e}_j) p(x_n | \mathbf{z}_n = \mathbf{e}_j, \theta_j^{(i)})} = \frac{\pi_l \mathcal{N}(x_n | \theta_l^{(i)})}{\sum_{j=1}^L \pi_j \mathcal{N}(x_n | \theta_j^{(i)})}. \quad (6.7)$$

Differentiating equation (6.3) with respect to  $\mu_l$ ,  $\theta_l$ , and  $\pi_l$ ,  $\forall l$ , and equating to zero provides an update of the parameter set  $\boldsymbol{\theta}^{(i+1)}$  as the maximisation step. Given  $N_l^{(i)} = \sum_{n=1}^N w_{n,l}^{(i)}$ , and  $\forall l$ ,

$$\mu_l^{(i+1)} = \frac{1}{N_l^{(i)}} \sum_{n=1}^N w_{n,l}^{(i)} x_n, \quad (6.8)$$

$$\sigma_l^{(i+1)} = \frac{1}{N_l^{(i)}} \sum_{n=1}^N w_{n,l}^{(i)} (x_n - \mu_l)^2, \quad (6.9)$$

$$\pi_l^{(i+1)} = \frac{N_l^{(i)}}{N}. \quad (6.10)$$

The algorithm is stopped when the rate of change of the log-likelihood falls below a predefined threshold or when a maximum number of iterations is reached. The full algorithm for fitting a GMM to a data set is given in algorithm 4, and figures 6.3 and 6.4 illustrate the algorithms behaviour on a simulated brain scan. For conciseness, throughout the chapter we show segmented images along with their corresponding intensity image histogram and the GMM that was used to fit it and produce the segmentation.

**Algorithm 4:** EM algorithm for GMM segmentation

**Input:**  $\mathbf{x} \in \mathbb{R}^N$  - Observation of pixel intensities  
 $K$  - Number of mixture components  
 $\mathcal{I}$  - Stopping criterion based on a maximum number of iterations or a rate of change threshold on the log-likelihood  
 $[\boldsymbol{\theta}^{(0)}]$  - Initial model [optional]

**Output:**  $\boldsymbol{\theta}$  - GMM model parameters  
 $w(\mathbf{z})$  - Membership weights, carrying soft segmentation

**Initialise:** Use  $\boldsymbol{\theta}^{(0)}$  or initialise model at random:  $\sigma_l^{(0)} = 0.5 \forall l$ ,  $\pi_l^{(0)} = 1/L \forall l$ ,  $\boldsymbol{\mu}_l^{(0)}$  randomly chosen from  $\mathbf{x}$

**repeat**

1.  $i \leftarrow i + 1$
2. E-step: Update  $w(\mathbf{z})^{(i)}$  with equation (6.7)
3. M-step: Update  $\boldsymbol{\theta}^{(i)}$  with equations (6.8) to (6.10)

**until** Condition  $\mathcal{I}$  is met;

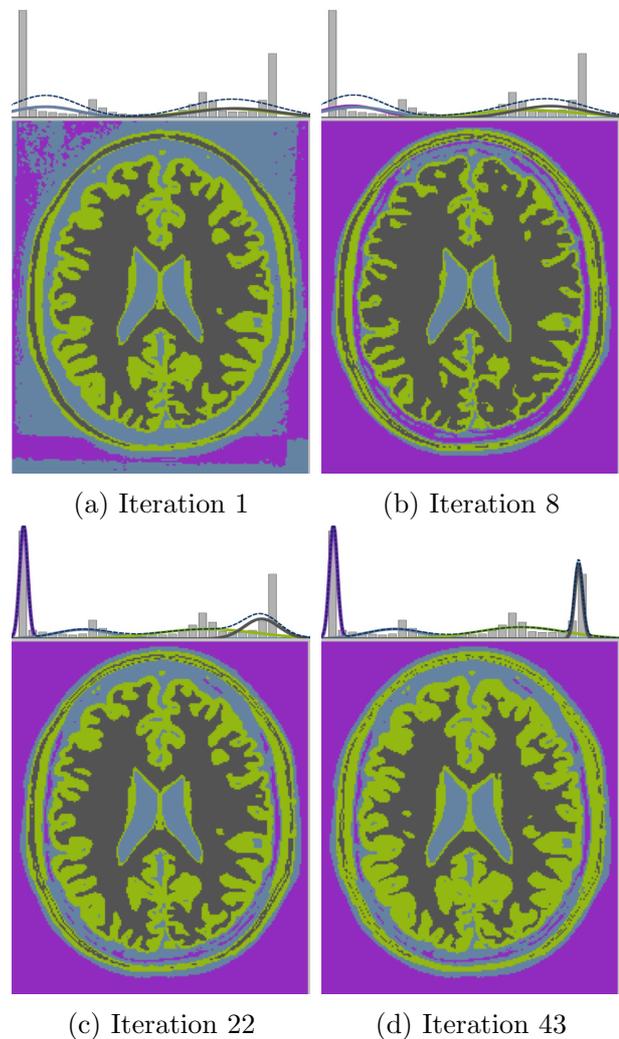
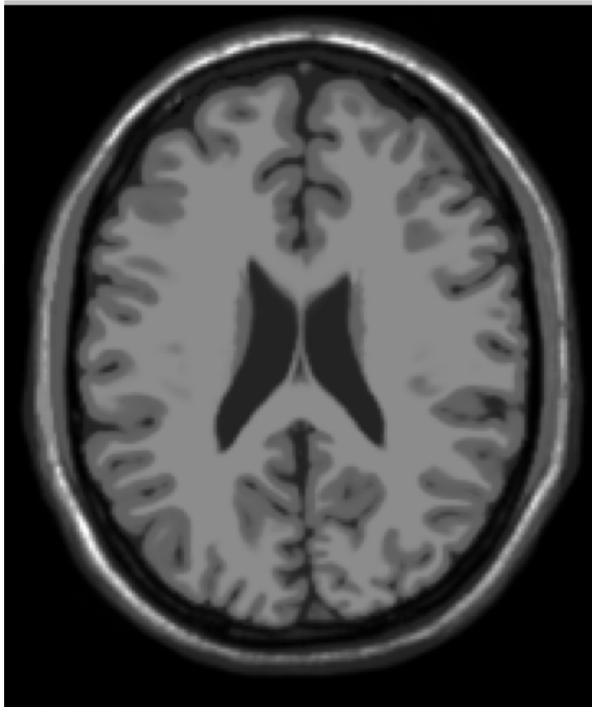


Figure 6.4: GMM estimation and segmentation with EM algorithm. At the top of each image is shown the histogram of intensity values and the model fit, with axes probability against image intensity.

Figure 6.3: Phantom brain MR image [95].

## 6.4 Joint reconstruction-segmentation of undersampled MRI data

Suppose now that we are interested in segmenting data  $\mathbf{x}$ , but that a CS acquisition protocol was used to accelerate its scan time with random k-space undersampling collecting data  $\mathbf{x}_u = \mathbf{M}\mathbf{F}\mathbf{x} = \mathbf{F}_u\mathbf{x} \in \mathbb{C}^M$ ,  $M < N$ . A reliable GMM segmentation of  $\mathbf{x}$  would require an accurate estimate of missing k-space data. CS reconstruction mechanisms such as the one presented in previous chapters provide approximate estimates of missing data. However, note too that having an accurate mixture model for the data, which decides labels purely based on intensity levels, would be useful information for the recovery of non-acquired k-space samples.

This intuitive mechanism for the joint reconstruction and segmentation of undersampled data is formalised in equation (6.11). This problem presents a non-linear reconstruction for an undersampled acquisition where the solution  $\mathbf{x}$  is forced to be consistent with the acquired data  $\mathbf{x}_u$  (first term), accurately and sparsely represented by a patch-based dictionary  $\mathbf{D}$  (second term and constraint), and consistent with a GMM term that is to be jointly estimated:

$$\min_{\mathbf{\Gamma}, \boldsymbol{\theta}, \mathbf{x}} \|\mathbf{F}_u\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\lambda}{N_p} \sum_{n=1}^N \|\mathbf{R}_n\mathbf{x} - \mathbf{D}\boldsymbol{\gamma}_n\|_2^2 - \beta \ln p(\mathbf{x}|\boldsymbol{\theta}) \quad \text{s.t.} \quad \|\boldsymbol{\gamma}_n\|_0 \leq s, \forall n. \quad (6.11)$$

This is a non-convex problem. We choose to approximate its solution by alternating the update of each optimisation variable  $\mathbf{\Gamma}$ ,  $\boldsymbol{\theta}$  and  $\mathbf{x}$  independently. Although this alternating projection approach is not guaranteed to converge, it decomposes the global problem into three subproblems that are either convex or can be solved with a greedy method. Sections 6.4.1 to 6.4.3 describe derivations for the three updates necessary, and the full algorithm is summarised in algorithm 5.

### 6.4.1 Update of sparse model with OMP

For  $\mathbf{x}$  and  $\boldsymbol{\theta}$  given, we require the update of the sparse model  $\boldsymbol{\Gamma}$  as

$$\min_{\boldsymbol{\Gamma}} \sum_{n=1}^N \|\mathbf{R}_n \mathbf{x} - \mathbf{D} \boldsymbol{\gamma}_n\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\gamma}_n\|_0 \leq s, \forall n, \quad (6.12)$$

which is a problem that has already been addressed by sparse coding algorithms such as OMP [109].

Additionally, this step could define the variable  $\mathbf{D}$  as an optimisation variable making the dictionary adaptive. Given that the GMM modelling defines intermediate segmentations of the data, an interesting extension would be to analyse the uses of intermediate labelled reconstructions to train class specific dictionaries for different regions of the image. For instance, dictionaries that are specifically trained from a given class, such as gray matter in brain MR tissue segmentation or myocardium in cardiac cine images, should reconstruct those areas better than a dictionary trained on data from all regions, because they have increased specificity. Similarly, dictionaries could be learnt from interfaces of pairs of regions. Some results have shown that such a classification of training data for independent dictionaries can improve reconstruction [156].

### 6.4.2 Update of GMM with EM

Given an estimation of the reconstruction  $\mathbf{x}$  and the sparse model  $\boldsymbol{\Gamma}$ , the minimisation problem in equation (6.11) reduces to

$$\min_{\boldsymbol{\theta}} -\ln p(\mathbf{x}|\boldsymbol{\theta}) = \max_{\boldsymbol{\theta}} \ln p(\mathbf{x}|\boldsymbol{\theta}), \quad (6.13)$$

which is the problem of fitting a GMM to a data set seen in equation (6.4) and solved through the EM algorithm.

When no estimate is available in the first iteration, the model is randomly initialised and the stopping criterion  $\mathcal{I}$  corresponds to 100 EM iterations. Subsequent updates of the model reuse the last estimation as initialisation and define a shorter stopping criterion of 30 iterations.

### 6.4.3 Update of reconstruction with CG

The last step is the update of the reconstruction  $\mathbf{x}$  given an estimation of  $\boldsymbol{\theta}$  and  $\boldsymbol{\Gamma}$ . Expanding the third term of equation (6.11) we have

$$\min_{\mathbf{x}} \|\mathbf{F}_u \mathbf{x} - \mathbf{y}\|_2^2 + \frac{\lambda}{N_p} \sum_{n=1}^N \|\mathbf{R}_n \mathbf{x} - \mathbf{D} \boldsymbol{\gamma}_n\|_2^2 - \beta \sum_{n=1}^N \ln \left( \sum_{l=1}^L \frac{\pi_l}{\sqrt{2\pi}\sigma_l} e^{-\frac{(\mathbf{P}_n \mathbf{x} - \mu_l)^2}{2\sigma_l^2}} \right) \quad (6.14)$$

using the notation  $x_n = \mathbf{P}_n \mathbf{x}$  with  $\mathbf{P}_n \in \mathbb{R}^{1 \times N}$  a row vector acting as pixel extraction operator to simplify analysis. Differentiating with respect to  $\mathbf{x}$  and equating to zero we find

$$2\mathbf{F}_u^H (\mathbf{F}_u \mathbf{x} - \mathbf{x}_u) + \frac{2\lambda}{N_p} \sum_{n=1}^N \mathbf{R}_n^T (\mathbf{R}_n \mathbf{x} - \mathbf{D} \boldsymbol{\gamma}_n) + \beta \sum_{n=1}^N \sum_{l=1}^L \frac{\frac{\pi_l}{\sqrt{2\pi}\sigma_l} e^{-\frac{(\mathbf{P}_n \mathbf{x} - \mu_l)^2}{2\sigma_l^2}}}{\sum_{j=1}^L \frac{\pi_j}{\sqrt{2\pi}\sigma_j} e^{-\frac{(\mathbf{P}_n \mathbf{x} - \mu_j)^2}{2\sigma_j^2}}} \left( \frac{\mathbf{P}_n^T (\mathbf{P}_n \mathbf{x} - \mu_l)}{\sigma_l^2} \right) = 0, \quad (6.15)$$

and using the compact definition of Gaussian distributions

$$2\mathbf{F}_u^H (\mathbf{F}_u \mathbf{x} - \mathbf{x}_u) + \frac{2\lambda}{N_p} \sum_{n=1}^N \mathbf{R}_n^T (\mathbf{R}_n \mathbf{x} - \mathbf{D} \boldsymbol{\gamma}_n) + \beta \sum_{n=1}^N \sum_{l=1}^L \frac{\pi_l \mathcal{N}(x_n | \theta_l)}{\sum_{j=1}^L \pi_j \mathcal{N}(x_n | \theta_j)} \left( \frac{\mathbf{P}_n^T (\mathbf{P}_n \mathbf{x} - \mu_l)}{\sigma_l^2} \right) = 0. \quad (6.16)$$

Notice that the multiplicative term where  $\mathbf{x}$  is non-linear is the same as the membership weights defined in equation (6.7), which is an output of the GMM update step. We can therefore replace this term by intermediate solutions of  $w_{n,l}$  and look for a fixed-point solution of  $\mathbf{x}$ . This results

in

$$2\mathbf{F}_u^H (\mathbf{F}_u \mathbf{x} - \mathbf{x}_u) + \frac{2\lambda}{N_p} \sum_{n=1}^N \mathbf{R}_n^T (\mathbf{R}_n \mathbf{x} - \mathbf{D} \boldsymbol{\gamma}_n) + \beta \sum_{n=1}^N \sum_{l=1}^L w_{n,l} \left( \frac{\mathbf{P}_n^T (\mathbf{P}_n \mathbf{x} - \boldsymbol{\mu}_l)}{\sigma_l^2} \right) = 0, \quad (6.17)$$

and isolating  $\mathbf{x}$  we find

$$\begin{aligned} & \left( \mathbf{F}_u^H \mathbf{F}_u + \frac{\lambda}{N_p} \sum_{n=1}^N \mathbf{R}_n^T \mathbf{R}_n + \frac{\beta}{2} \sum_{n=1}^N \sum_{l=1}^L \frac{w_{n,l}}{\sigma_l^2} \mathbf{P}_n^T \mathbf{P}_n \right) \mathbf{x} \\ & = \mathbf{F}_u \mathbf{x}_u + \lambda \sum_{n=1}^N \frac{\mathbf{R}_n^T \mathbf{D} \boldsymbol{\gamma}_n}{N_p} + \frac{\beta}{2} \sum_{n=1}^N \sum_{l=1}^L \mathbf{P}_n^T \frac{w_{n,l}}{\sigma_l^2} \boldsymbol{\mu}_l. \end{aligned} \quad (6.18)$$

To better understand this expression, we take the Fourier transform on both sides and analyse each term separately in equation (6.19) for the k-space update of  $\mathbf{x}$ :

$$\begin{aligned} & \left( \mathbf{F} \mathbf{F}_u^H \mathbf{F}_u \mathbf{F}^H + \frac{\lambda}{N_p} \mathbf{F} \sum_{n=1}^N \mathbf{R}_n^T \mathbf{R}_n \mathbf{F}^H + \frac{\beta}{2} \mathbf{F} \sum_{n=1}^N \sum_{l=1}^L \frac{w_{n,l}}{\sigma_l^2} \mathbf{P}_n^T \mathbf{P}_n \mathbf{F}^H \right) \mathbf{F} \mathbf{x} \\ & = \mathbf{F} \mathbf{F}_u^H \mathbf{x}_u + \lambda \mathbf{F} \sum_{n=1}^N \frac{\mathbf{R}_n^T \mathbf{D} \boldsymbol{\gamma}_n}{N_p} + \frac{\beta}{2} \mathbf{F} \sum_{n=1}^N \sum_{l=1}^L \mathbf{P}_n^T \frac{w_{n,l}}{\sigma_l^2} \boldsymbol{\mu}_l. \end{aligned} \quad (6.19)$$

The first term of the left-hand side reduces to an  $N \times N$  matrix with diagonal entry  $n$  set to 1 if k-space location  $n$  was acquired. The second term is the factor  $\frac{\lambda}{N_p}$  multiplied by the identity matrix scaled up by the dictionary atom size  $N_p$ , and hence reduces to  $\lambda \mathbf{I}$ . Similarly on the right-hand side, the first term is simply the zero-filled k-space acquisition and the second is the Fourier transform of the solution found by averaging the contribution of patches sparsely approximated with  $\mathbf{D}$ . The third term on the left-hand side, contrary to the first and second, is not a diagonal matrix. This prevents the update of  $\mathbf{x}$  to be performed pixel-wise independently, but the system of equations can nevertheless be easily solved with CG descent.

The third term on both sides is difficult to interpret, but it becomes intuitive if we let  $\beta \rightarrow \infty$ .

Consider then the update given by

$$\sum_{n=1}^N \sum_{l=1}^L \frac{w_{n,l}}{\sigma_l^2} \mathbf{P}_n^T \mathbf{P}_n \mathbf{x} = \sum_{n=1}^N \sum_{l=1}^L \mathbf{P}_n^T \frac{w_{n,l}}{\sigma_l^2} \mu_l. \quad (6.20)$$

where we have removed the left multiplication by  $\mathbf{F}$  on both sides. Given that the pixel extraction vector isolates the handling of individual pixels, the update of the reconstruction (now in image domain) is independent for all pixels and is defined as

$$x_n = \frac{\sum_{l=1}^L \frac{w_{n,l}}{\sigma_l^2} \mu_l}{\sum_{l=1}^L \frac{w_{n,l}}{\sigma_l^2}}. \quad (6.21)$$

The third terms on both sides of equation (6.19) therefore promote a solution where pixel values are determined as a weighted combination of GMM means. Individual contributions depend on membership weights and Gaussian variances. In a typical scenario, membership weights for most pixels are quickly attributed to only one Gaussian distribution  $j$  (*i.e.*  $w_{n,l} \approx 1$  for  $l = j$ , and  $w_{n,l} \approx 0$  for  $l \neq j$ ), making  $\mu_j$  dominate the contribution to the update of pixel intensity  $x_n$  by the GMM term.

**Algorithm 5:** Reconstruction-segmentation for fast MRI

**Input:**  $\mathbf{x}_u \in \mathbb{R}^M$  - Acquired k-space samples  
 $\mathbf{M} \in \mathbb{R}^{M \times N}$  - Undersampling mask  
 $K$  - Number of mixture components  
 $\mathcal{I}$  - Maximum number of iterations  
**Output:**  $\mathbf{x} \in \mathbb{R}^N$  - Reconstructed intensity image  
 $w(\mathbf{z})$  - Membership weights, carrying soft segmentation  
**Initialise:**  $\mathbf{x}^{(0)} = \mathbf{F}_u \mathbf{x}_u$ ,  $\mathbf{\Gamma}^{(0)} = \mathbf{0}$ ,  $\sigma_l^{(0)} = 0.5 \forall l$ ,  $\pi_l^{(0)} = 1/L \forall l$ ,  $\boldsymbol{\mu}_l^{(0)}$  randomly chosen from  $\mathbf{x}^{(0)}$   
**repeat**  
    1.  $i \leftarrow i + 1$   
    2. Update sparse coding  $\mathbf{\Gamma}^{(i)}$  with OMP using  $\mathbf{x}^{(i-1)}$  in equation (6.12)  
    3. Update model parameters  $\boldsymbol{\theta}^{(i)}$  with EM using  $\mathbf{x}^{(i-1)}$  in equation (6.13)  
    4. Update reconstruction  $\mathbf{x}^{(i)}$  with CG using  $\mathbf{\Gamma}^{(i)}$  and  $\boldsymbol{\theta}^{(i)}$  in equation (6.14)  
**until** Condition  $\mathcal{I}$  is met;

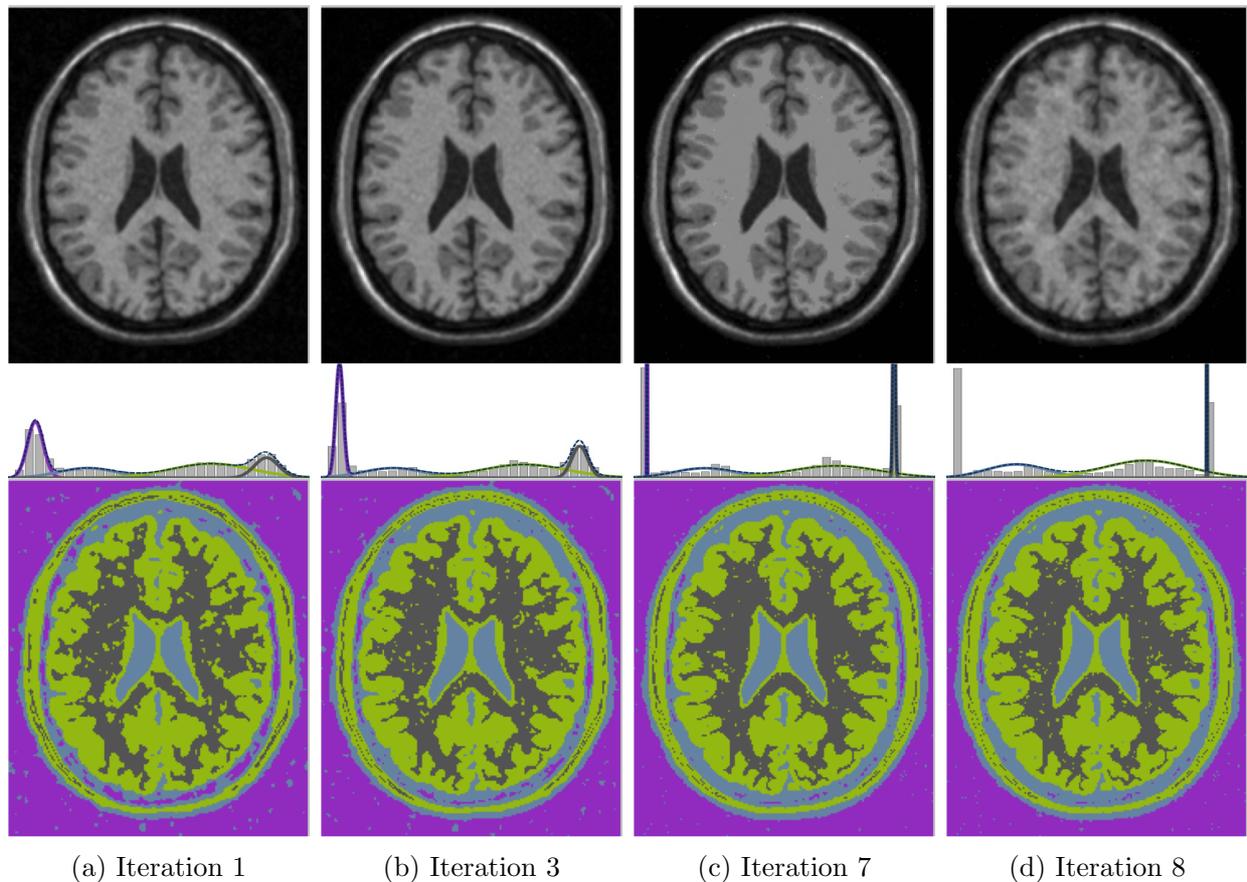


Figure 6.5: Joint reconstruction-segmentation trivial solution and EM instability. Each image shows the intensity reconstruction (top), the intensity histogram and GMM model fit, where axes are probability against intensity (middle), and the segmentation derived from the model fit (bottom). There exists a trivial solution to the global problem posed for which  $\sigma_l = 0$ . The figures show how the left-most Gaussian in the model collapses after iteration 8, and this destabilises the joint reconstruction-segmentation as the model needs to readjust itself and compensate for the missing cluster.

#### 6.4.4 Avoiding EM trivial solution

An important consideration for the cost function in equation (6.11) is the fact that the discriminative term  $\ln p(\mathbf{x}|\boldsymbol{\theta})$ , unlike the data consistency and the sparse model accuracy terms, is unbounded. We are seeking a joint solution for  $\mathbf{x}$ ,  $\boldsymbol{\Gamma}$  and  $\boldsymbol{\theta}$ , but as the log-likelihood can grow to infinity it will unavoidably preponderate over the first two terms in the minimisation of the functional.

The EM algorithm is known to present singularities. Consider the case when a data point  $x_n$

is exactly equal to the mean of mixture component  $l$ . This point will contribute a term to the log-likelihood of the form

$$\pi_l \mathcal{N}(x_n | \mathbf{z}_n = \mathbf{e}_l, \theta_l) = \frac{\pi_l}{\sqrt{2\pi\sigma_l}} e^{-\frac{(x_n - \mu_l)^2}{2\sigma_l^2}} = \frac{\pi_l}{\sqrt{2\pi\sigma_l}}, \quad (6.22)$$

hence the trivial solution that will maximise it is  $\sigma_l = 0$ . This causes component  $l$  to collapse upon data point  $x_n$  and reveals that the problem is not well-posed. Possible solutions exist to avoid this pathological behaviour consisting on heuristics, such as reinitialising the estimation if a component collapses, or considering a variational approach.

The problem we present in equation (6.11) is bound to find this trivial solution, given that we simultaneously look for  $\mathbf{x}$ . In practice, for  $\beta > 0$  pixel intensities are pulled towards the means of the Gaussians they have been assigned, and eventually pixel intensities reach the same value as the component mean triggering the singularity described. An example is given in figure 6.5, where after only 8 iterations of a joint reconstruction-segmentation of undersampled data from a brain phantom the GMM becomes unstable. One component from the initial  $L = 4$  component model collapses with  $\sigma = 0$ , hence reducing the model to only 3 components and disrupting the solution.

To prevent this from happening we propose the use of a minimum standard deviation that can be reached by Gaussians to constrain the valid search for parameters  $\theta$ . This is information that can easily be learnt from past similar images, but other options are available adding additional regularisers to equation (6.11) such as  $\|\sigma - \sigma_{\text{Ref}}\|_2^2$  using reference standard deviation values  $\sigma_{\text{Ref}}$ .

### 6.4.5 GMM extensions

The formulation of the segmentation as a by-product of a GMM allows for different natural extensions that have been studied in the past. We analyse the use of two additions: Markov random field (MRF) and spatial probabilistic priors.

### Markov random field

The model presented suggests a pixel-wise segmentation independent of space. MRFs are a regularisation technique allowing to incorporate spatial smoothness in the segmentation process, by considering the probability of pixel classification based on the labelling of neighbouring pixels. This is a simple way of incorporating prior knowledge about the labelling of the data being segmented. For instance, in the segmentation of cardiac cine it is expected that the blood pool regions are surrounded by myocardium tissue, or that cortical grey matter and white matter share boundaries in brain MR segmentation.

Formally, the latent variable  $\mathbf{z}$  is a random variable whose value now depends on a predefined set of weights  $\mathbf{G}$  setting the relationship between pairs of labels within a neighbourhood. The neighbourhood of pixel or voxel  $n$  is defined as the six first-degree neighbours for 2D+t and 3D data sets  $\mathcal{P}_n = \{n^n, n^s, n^e, n^w, n^t, n^b\}$ , corresponding to north, south, east, west, top and bottom.

This formally modifies the calculation of the GMM parameters through the EM procedure. As described in [141], we rely on an approximation from mean field theory [73, 157, 140], which translates into the replacement of equation (6.7) by

$$w_{n,l}^{(i)} = \frac{p(\mathbf{z}_n = \mathbf{e}_l | \mathbf{w}_{\mathcal{P}_n}^{(i)}, \mathbf{G}^{(i)}) \mathcal{N}(x_n | \theta_l^{(i)})}{\sum_{l=1}^L p(\mathbf{z}_n = \mathbf{e}_l | \mathbf{w}_{\mathcal{P}_n}^{(i)}, \mathbf{G}^{(i)}) \mathcal{N}(x_n | \theta_l^{(i)})}. \quad (6.23)$$

where the variable  $\pi_l = p(\mathbf{z}_n = \mathbf{e}_l)$  has been replaced by the neighbour dependent

$$p(\mathbf{z}_n = \mathbf{e}_l | \mathbf{w}_{\mathcal{P}_n}^{(i)}, \mathbf{G}^{(i)}) = \frac{e^{-U_{\text{MRF}}(\mathbf{e}_l | \mathbf{w}_{\mathcal{P}_n}^{(i)}, \mathbf{G}^{(i)})}}{\sum_{l=1}^L e^{-U_{\text{MRF}}(\mathbf{e}_l | \mathbf{w}_{\mathcal{P}_n}^{(i)}, \mathbf{G}^{(i)})}}. \quad (6.24)$$

The energy function used is defined as

$$U_{\text{MRF}}(\mathbf{e}_l | \mathbf{w}_{\mathcal{P}_n}^{(i)}, \mathbf{G}^{(i)}) = \mathbf{G}^{(i)} \mathbf{g}_n^{(i)}, \quad (6.25)$$

with  $\mathbf{g}_n \in \mathbb{R}^L, \forall n$ , a variable that accumulates class probabilities around the neighbourhood of  $x_n$ , where individual entries  $l = 1, 2, \dots, L$ , are defined as  $\sum_{j \in \mathcal{P}_n} w_{j,l}$ .

### Probabilistic spatial prior

It is often possible and desirable to exploit spatially dependent label priors from probabilistic atlases which constrain regions in the image where different classes are expected to belong. The benefits of probabilistic priors in image segmentation have been shown to contribute to intensity information to obtain a more robust and accurate segmentation through EM [141]. Priors can be derived from labelled atlases through registration and label propagation. In a joint reconstruction-segmentation setup, even if intermediate reconstructions contain aliasing, a coarse intensity reconstruction is often reliable enough for atlas registration.

The addition of probabilistic priors to the model regularised by a MRF further modifies the location independent variable  $\pi_l$ , which instead of equation (6.24) becomes

$$p(\mathbf{z}_n = \mathbf{e}_l | \mathbf{w}_{\mathcal{P}_n}^{(i)}, \mathbf{G}^{(i)}) = \frac{\pi_{n,l} e^{-U_{\text{MRF}}(\mathbf{e}_l | \mathbf{w}_{\mathcal{P}_n}^{(i)}, \mathbf{G}^{(i)})}}{\sum_{l=1}^L \pi_{n,l} e^{-U_{\text{MRF}}(\mathbf{e}_l | \mathbf{w}_{\mathcal{P}_n}^{(i)}, \mathbf{G}^{(i)})}}. \quad (6.26)$$

where the probabilistic prior is defined as  $\pi_{n,l} = p(\mathbf{z}_n = \mathbf{e}_l)$ .

## 6.5 Experiments and results

In all tests, a fully sampled MR data set  $\mathbf{x} \in \mathbb{R}^N$  corresponding to 2D, 3D or 2D+t data, is retrospectively undersampled in k-space with a mask  $\mathbf{M} \in \mathbb{R}^{M \times N}$  producing an undersampled observation  $\mathbf{x}_u = \mathbf{M}\mathbf{x} = \mathbf{F}_u\mathbf{x} \in \mathbb{R}^M$ . Note that the fully sampled acquisition simulates a single coil scenario and only the absolute value of the complex MR image is considered for k-space undersampling and reconstruction. Extending the algorithm to handle parallel k-space data and complex-valued data would be necessary for practical application. These cases could

be addressed with a CS regularisation of parallel reconstruction algorithms as proposed in the previous chapter, and with the use of a complex-valued GMM.

All undersampling masks simulate 2D variable density undersampling with a Gaussian PDF favouring the sampling of the center of k-space where most of the signal energy resides. The center  $11 \times 11$  square of k-space is always acquired. Modifying the variance of the 2D Gaussian allows for different undersampling rates. Figure 6.6a shows an example of this PDF, for which the mask in figure 6.6c with undersampling factor 5 is produced as the outcome of a point-wise Bernoulli process. In 3D or 2D+t data sets independent masks are drawn for successive slices or temporal frames.

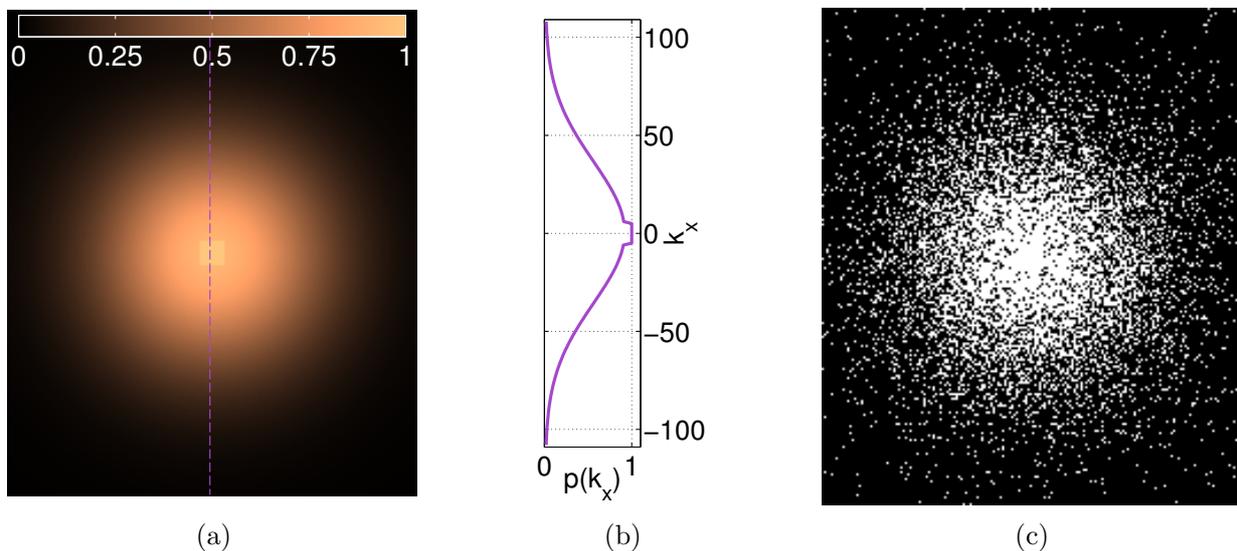


Figure 6.6: Undersampling mask simulation. A 2D Gaussian PDF is generated (a) from which a mask can be obtained as the result of a Bernoulli process (c). In (b) we show a single column of the pdf in (a). The variance of the PDF be modified to obtain different undersampling rates.

We first present a detailed analysis of the baseline algorithm without MRFs and probabilistic priors. A simulated 2D brain scan from [95] is used to provide a sense of how the algorithm reacts to parameter  $\beta$ . This parameter is of great importance because it can balance the output towards a purely reconstructive result or favour the search for the simpler modelling provided by the GMM. Cardiac cine data sets are then used to evaluate the performance of the baseline joint reconstruction-segmentation algorithm, whereas real 3D brain scans are used to demonstrate the capabilities of the algorithm extensions.

For all experiments a DCT dictionary with  $K = 200$  atoms of size  $N_p = 4 \times 4 \times 4$  (or  $8 \times 8$  for the 2D image test) is used to impose a sparsity of  $s = 8$  on all patches of the data set. Throughout, the choice  $\lambda = 10^{-3}$  was made to ensure high k-space consistency between the reconstruction and acquired samples. The joint reconstruction always undergoes 5 reconstructive iterations where  $\beta = 0$ , before starting the estimation of GMM parameters. This is a pragmatic choice as it was found that the high degree of undersampling aliasing in the first iterations of the algorithm is detrimental and a cleaner initialisation is beneficial. Parameter  $\beta$  was set to  $10^{-9}$  for cardiac experiments and  $6 \times 10^{-10}$  for brain data tests. The minimum standard deviation was set to 0.025 for both cardiac and brain data examples. This value was obtained from different fully sampled scans to be a representative minimum standard deviation achieved when a GMM was fitted to the fully sampled data.

### 6.5.1 Impact of discriminative modelling on reconstruction

It is interesting to visually understand the repercussion of the discriminative GMM model weighting on the intensity reconstruction. We use a simple simulation of a 2D brain MR image from BrainWeb [95] for demonstration purposes. The k-space of the original image is undersampled by 5 and then jointly reconstructed and segmented using with different weightings  $\beta$ .

With  $\beta = 0$ , the algorithm is purely reconstructive and corresponds to a simple version of dictionary CS reconstructions presented in [117], therefore no mixture model is jointly estimated. As  $\beta$  is set to a positive value, a mixture model is automatically estimated as shown in figure 6.7b. The minimum variance imposed on the model restricts the fitting of the histogram. Using a higher weighting in figure 6.7c sharpens edges between regions which have been identified with different labels. This is a direct consequence of forcing the reconstruction to be consistent with the GMM, because as data values approach their attributed Gaussian means the term  $\ln p(\mathbf{x}|\boldsymbol{\theta})$  increases. Therefore pixels deviating from their Gaussian means are increasingly penalised for higher  $\beta$ . For the highest weight shown in figure 6.7d, pixel intensities cluster even closer to

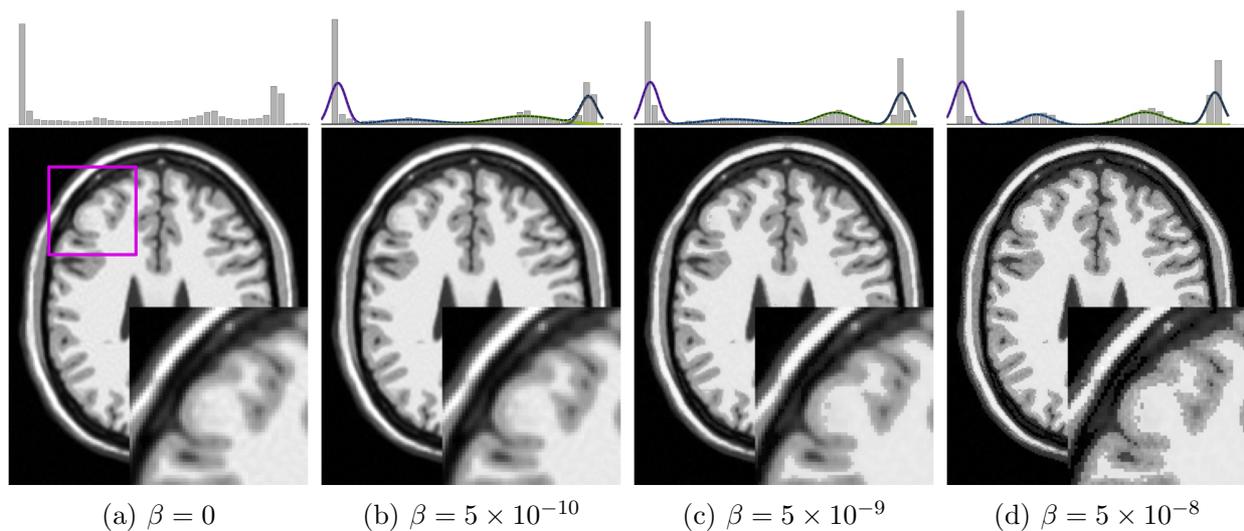


Figure 6.7: Visual appearance of reconstruction with different values of  $\beta$ . Each image shows the intensity histogram, where axes are probability against intensity (top), and the intensity image reconstruction (bottom). A joint segmentation is obtained for  $\beta > 0$ . As  $\beta$  increases pixel intensities are encouraged to cluster around Gaussian means, which forces a separation between different clusters in the intensity histograms shown above. This has an edge enhancement effect that is visually apparent in the images with highest  $\beta$  values.

Gaussian means and accentuate this effect. These changes in intensity appearance will lead to different segmentation results, so the tuning parameter  $\beta$  opens a range of different reconstruction and segmentation pairs and is a flexible mechanism to weight reconstruction fidelity and segmentation sharpness.

### 6.5.2 Left ventricle cardiac cine segmentation

We analyse the performance of the baseline algorithm in algorithm 5 for the specific problem of left ventricle segmentation in cardiac cine data. Fully sampled short-axis data was generated from 7 patients using an optimal combination of 32 channel data. All data sets contain 30 temporal frames of size  $256 \times 256$  (i.e.  $N = 256 \times 256 \times 30$ ) with a  $320 \times 320 \text{ mm}^2$  field of view and 10 mm slice thickness. Mean  $\pm$  standard deviation of heart rate was  $62 \pm 10.2$  bpm, giving a temporal rate of  $33 \pm 5.5$  ms for the 30 frames. The raw multi-coil data was reconstructed using SENSE [112] with no k-space undersampling and retrospective gating. Coil sensitivity maps were normalised to a body coil image to produce a single complex-valued image set that

was used to form final magnitude images.

The number of components in the mixture model was limited to  $L = 3$ . A very low  $L$  ensures that the blood pool regions that we are interested in are captured by a single component in the mixture model as shown in figure 6.8 on fully sampled data. Choosing a higher  $L$  would not prevent the algorithm from working but would complicate the assessment of left ventricle segmentation performance.

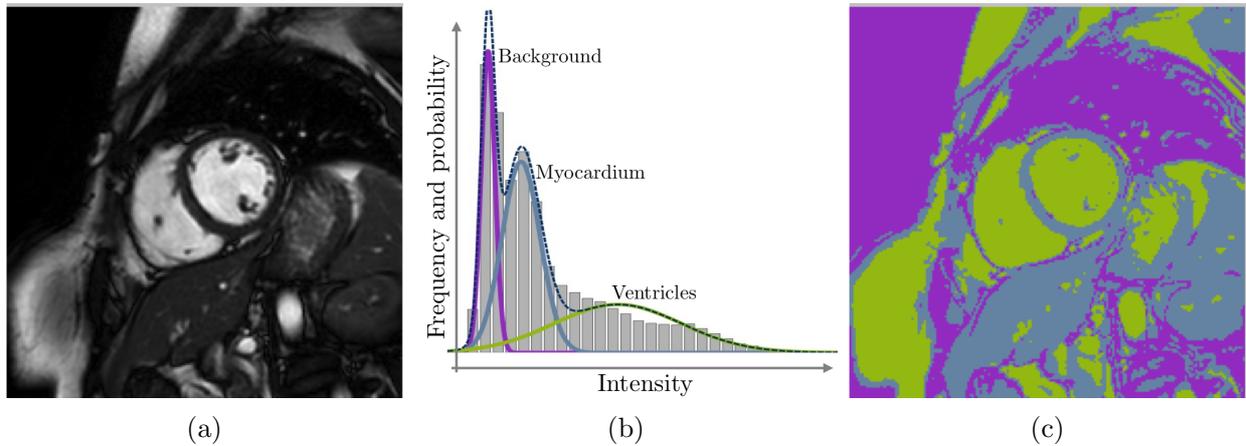


Figure 6.8: Cardiac cine data (a) is segmented with a GMM of  $K = 3$  components (b) resulting in a segmentation (c) that uses a single component to label blood pool regions.

The experiments for cardiac cine segmentation analyse three scenarios. Consider the original fully sampled acquisition of a data set. The intensity image reconstruction of this data is given by a Fourier transform and this intensity data can later be segmented using the EM algorithm. Assume now that the same k-space data is retrospectively undersampled. It can be handled in a similar fashion as the traditional imaging work flow, by first reconstructing with the patch-based CS method and then segmenting the result with EM. We will refer to this as the separate method. The alternative is to jointly reconstruct and segment the data using algorithm 5, which we will call the joint method.

The segmentations of the separate and joint methods can then be compared against the segmentation from the fully sampled data. Notice that the comparison is not against a ground truth manual segmentation, but rather tries to analyse which one of the separate or joint approaches

is less affected by undersampling compared to the fully sampled case.

### Qualitative analysis

Figure 6.9 compares a fully sampled temporal frame with the reconstructions of the separate and the joint methods with a 10 fold undersampling rate. A ROI around the heart is zoomed in. Notice how the separate method in figure 6.9c is unable to perfectly recover details, smooths texture for instance inside blood pool regions and blurs edges between ventricles and myocardium. The joint reconstruction result in figure 6.9c again shows enhanced edges between ventricles and myocardium because they have been identified as belonging to different Gaussians in the mixture model.

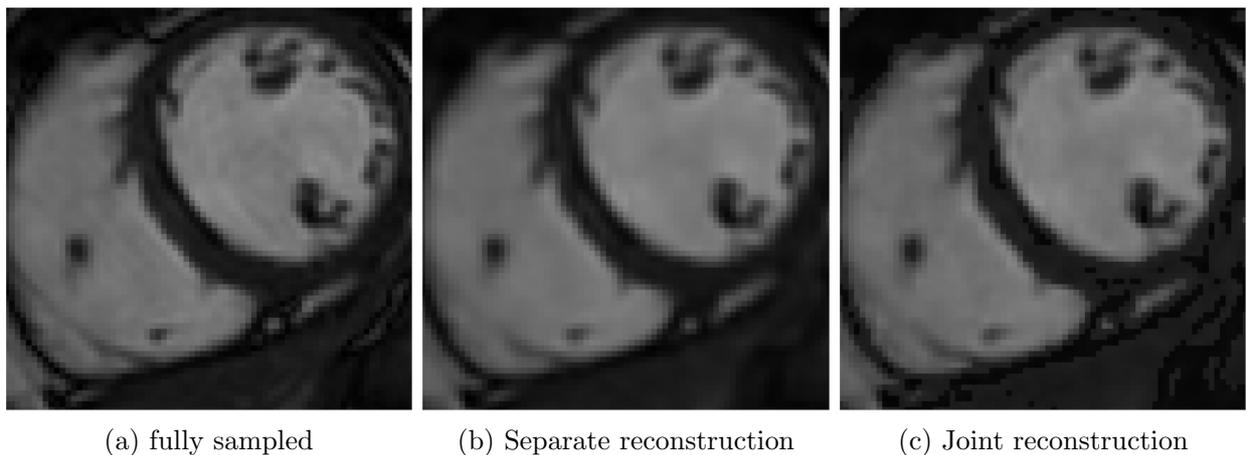
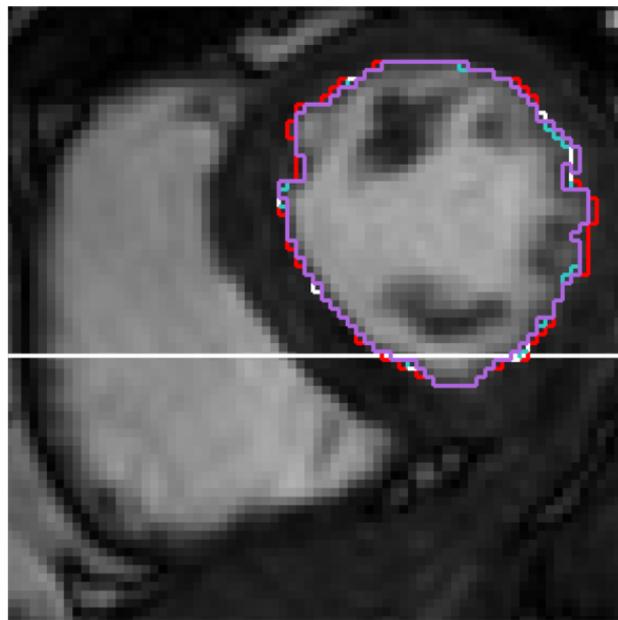


Figure 6.9: Cardiac cine reconstruction of a fully sampled data set (a) accelerated by 10 and reconstructed using the separate (b) and joint (c) methods.

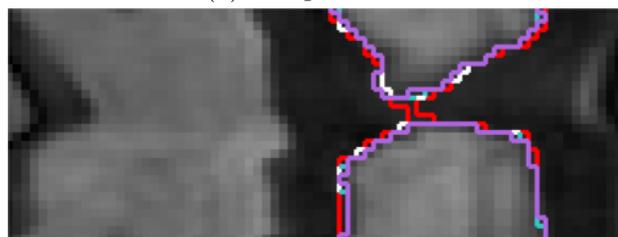
Changes in the reconstruction affect the segmentation of the data set. Figure 6.10 shows the segmented contour of the left ventricle in the three cases compared on a temporal frame and across a temporal profile for an undersampling rate of 24. The temporal profile corresponds to the white horizontal line in figure 6.10a. The segmentation contours are overlaid onto the fully sampled data and the fully sampled segmentation is shown in purple. Segmentations of the undersampled data are only shown when they deviate from the fully sampled segmentation, and are coded in red for the separate method, in blue for the joint method, and in white when both

methods coincide. Exclusive errors from the separate approach are predominant compared to the joint method in both the slide and profile.

In particular, figure 6.10b shows how the separate method classifies as left ventricle a connected region that is identified as myocardium between two left ventricle regions in the fully sampled segmentation. It is likely that this is a consequence of the blurred edges of the reconstruction produced by the separate method at high undersampling, but the sharpening of edges reconstructed by the joint method is able to improve segmentation accuracy.



(a) Temporal slide



(b) Temporal profile

Figure 6.10: Left ventricle segmentation from data undersampled at rate 24 seen on a temporal slide (a) and a temporal profile (b) corresponding to the horizontal line in (a). The segmentation of the fully sampled data (purple) is displayed along with deviations corresponding to errors in the segmentation of accelerated data that are exclusive to the separate method (red), the joint method (blue), or are shared by both methods (white).

### ROI pixel misclassification

In order to quantitatively assess the improvement of the joint reconstruction-segmentation over the separate approach we manually defined a ROI to tightly contain the left ventricle and myocardium boundary in the original data sets. For the 7 different patients, fully sampled k-spaces were undersampled at different acceleration rates and then reconstructed and segmented using both the separate and joint methods. Segmentation accuracy was evaluated in both cases with the percentage of misclassified pixels inside the ROI relative to the fully sampled segmentation.

Figure 6.11 shows box plots of pixel misclassification rates. Each dot at a given acceleration rate is the result from a different data set and the improvement in pixel misclassification of the joint method compared to the separate is shown by lines linking data sets for both approaches. For all data sets and all acceleration rates, it was shown that the segmentation of the joint approach was more accurate than the separate inside the ROI defined. A paired t-test was conducted for all the experiments in the plot analysing pixel misclassification difference between the separate and joint methods. The result was a mean misclassification improvement of  $\mu = 1.99\%$  and a p-value  $2.19 \times 10^{-18}$ , revealing that the improvement is statistically significant.

An interesting finding is illustrated in figure 6.12. We focus on a single data set and analyse reconstruction and segmentation errors at different undersampling rates with both methods. Reconstruction errors shown in figure 6.12a are evaluated as MSE, whereas segmentation errors in figure 6.12b look again at pixel misclassification rates.

The joint method outperforms the separate method in both in reconstruction and in segmentation accuracy. However, the change is very different as can be seen in figures 6.12c and 6.12d where we show the percentage improvement in reconstruction and segmentation accuracy obtained by the joint method. The reconstruction improvement is relatively small compared to the segmentation, which can reach above 30%. This finding suggests that the modification in modelling for reconstruction that is introduced by the mixture model might not necessarily

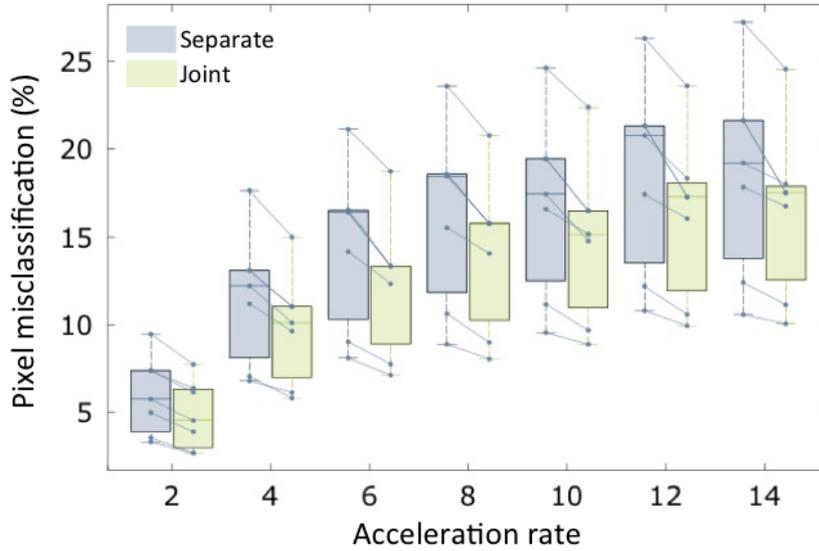


Figure 6.11: Pixel misclassification rate for left ventricle segmentation of undersampled cardiac cine data.

provide a much more accurate reconstruction in terms of MSE, but it does condition the data in a way that its segmentation is less affected by k-space undersampling.

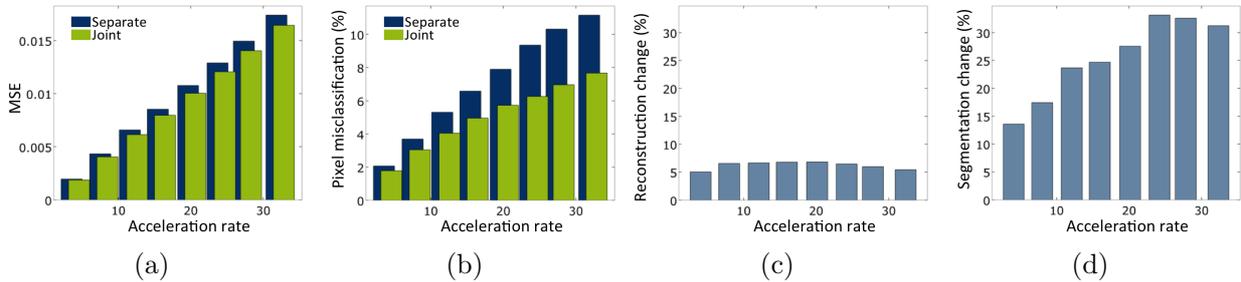


Figure 6.12: Reconstruction and segmentation errors (a, b) produced at different acceleration rates by the joint (green) and separate (blue) approaches. The rate of improvement as a percentage is less pronounced for the reconstruction (c) than for the segmentation (d).

### 6.5.3 Brain tissue segmentation

In the following experiments we investigate the applicability of the joint reconstruction segmentation method proposed to full brain tissue segmentation. In particular, we are interested in the segmentation of 5 tissues: cerebrospinal fluid (CSF), ventricles (Vent), cortical grey matter (cGM), deep grey matter (dGM), and white matter (WM). For this purpose we define a GMM

of  $L = 5$  components. It is in practice challenging to obtain accurate tissue segmentation relying only on intensity features. We therefore assess the benefits of incorporating MRFs and spatial priors to the model.

The data used for these experiments are T2-weighted 3D brain scans from 10 patients. Spatial priors are derived from the MNI atlas [95] by affine registering the intensity image to MNI space. This registration is performed whenever an update of the GMM is required. For fully sampled data or the separate method, only one registration is needed prior to the segmentation stage, whereas in the joint method a registration is performed in every iteration after the update of the reconstruction. The ground truth used for these experiments is a manual segmentation available for each data set.

### MRFs and spatial priors

To analyse changes induced by the introduction of spatial priors and MRFs we look at one data set with no undersampling and run a simple EM segmentation. We compare the segmentation given by the baseline EM algorithm (using only intensity information), the EM algorithm with spatial priors added but no MRF ( $\mathbf{G} = \mathbf{0}$ ), and with spatial priors and with MRF regularisation. The pairwise interactions defined between tissues for the MRF are rated as high ( $H = 0$ ), medium ( $M = 1$ ) or low ( $L = 1.5$ ), which respectively define to the probability that a tissue has to appear next to another. The weight matrix is set as

$$\mathbf{G} = \begin{bmatrix} H & L & L & M & M \\ L & H & L & M & M \\ L & L & H & M & H \\ M & M & M & H & H \\ M & M & H & H & H \end{bmatrix},$$

where columns and rows orderly relate to CSF, Vent, cGM, dGM and WM.

Figure 6.13 visually compares the three cases to the manual segmentation in a single coronal slice. The EM algorithm alone is insufficient for this data to correctly identify different tissues. The use of affine registered spatial priors provides a substantial improvement over the baseline EM algorithm, as it constrains the possible spatial locations where different tissues are expected to be found. Also, the additional use of MRF regularisation helps cleaning the segmentation by removing unlikely configurations, such as those highlighted with arrows in figure 6.13c where isolated Vent pixels in dGM are removed and cGM labelling is corrected.

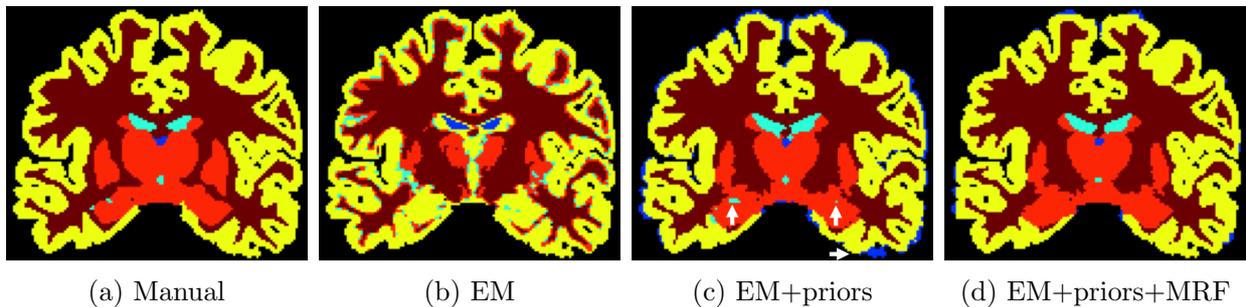


Figure 6.13: Comparison of brain tissue segmentation using the EM algorithm. The baseline EM method (b) is unable to provide an accurate result relative to a manual segmentation (a). The introduction of spatial probabilistic priors allows for a more precise result (c) and the use of MRF regularisation can additionally discard improbable configurations (d).

Table 6.1 displays Dice coefficients for the three cases compared: baseline EM, EM+priors and EM+priors+MRF. Similar effects are visible when data is undersampled and the joint reconstruction-segmentation method is used with these extensions. The Dice metric evaluates the overlap of different regions, where a Dice metric of 1 means perfect segmentation correspondence.

Table 6.1: Dice metrics for segmentation of a fully sampled data set.

Method	CSF	Vent	cGM	dGM	WM
EM	0.1312	0.0040	0.8033	0.1245	0.8955
EM+priors	0.0183	0.8132	0.9177	0.8651	0.9383
EM+priors+MRF	0.0611	0.8532	0.9561	0.8921	0.9406

### Segmentation from undersampled brain scans

The data from the 5 patients was retrospectively undersampled in k-space with different rates and reconstructed and segmented with spatial priors. Figure 6.14 compares tissue segmentations in a single slice of one data set when data is undersampled by 8. Both cases are very similar and in general fail in the same regions mainly because of the deterioration of the intensity reconstruction caused by the aggressive undersampling. There are however detailed locations where the joint method is able to better reproduce the true segmentation as shown in the arrows of figure 6.14c.

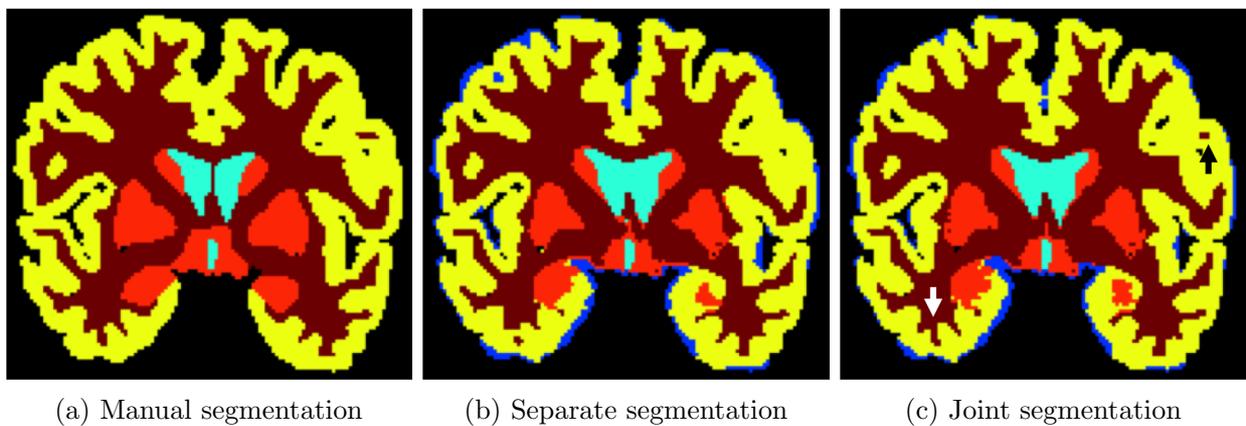


Figure 6.14: Brain tissue segmentation from data undersampled by 8. The joint method (c) is able at some locations, such as those highlighted by arrows, to better approximate the manual segmentation (a) than the separate approach (b).

In order to analyse global results on all cases, we consider segmentations for each data set at different acceleration rates and compute average tissue Dice metrics (the Dice for CSF is not taken into account for the averaging because of the poor quality in all examples). Results are shown in figure 6.15, where again linked dots correspond to the same data set. The plot reveals that the average Dice metric improves for almost all cases when using the joint method instead of the separate. A paired t-test conducted on all experiments to analyse the average Dice difference between the joint and the separate method statistically confirms this hypothesis with an average Dice improvement of  $\mu = 5.6 \times 10^{-3}$  with p-value  $2.10 \times 10^{-9}$ .

We can also use a GMM segmentation on the fully sampled data set as ground truth to compare

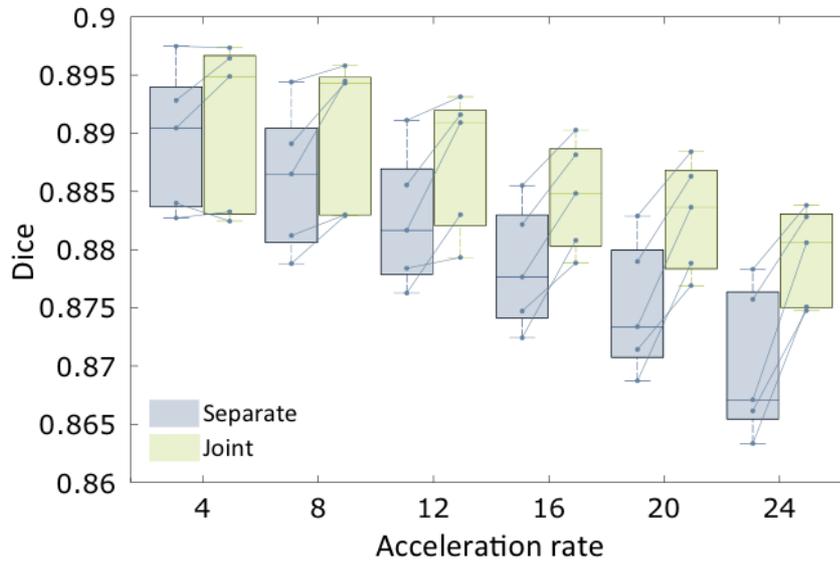


Figure 6.15: Average tissue segmentation Dice from undersampled data (excluding CSF).

against the resulting segmentations. This is the same test as presented in section 6.5.2 for cardiac examples where no manual segmentation was available, and reveals how much the segmentation in both methods is degraded by the undersampling. In this case, however, the joint method provides a segmentation that deviates more from the fully sampled segmentation than the separate method.

We speculate the cause for this, which contradicts what was found in section 6.5.2, is the contribution of spatial priors to the update of the reconstruction. Probabilistic priors are obtained from the registration and weighting of multiple manual segmentations performed by experts. These segmentations are not only based on intensity information, but also on anatomy and the experts's past experience. As a result, drawing intensity information from the labelling to inform the update of the reconstruction is likely to produce errors, given that pixels that are labelled based on a strong prior will be confidently pulled towards the class mean even though their true intensity value is not necessarily close to it.

### 6.5.4 Algorithm speed and acceleration

The core of the algorithm was implemented in Matlab, with the exception of the OMP sparse coding steps [119] and the affine registrations [131] which were performed in C++. As a reference for algorithm runtime, on a configuration Intel Core i72600 CPU at 3.4 GHz and 8 GiB of memory, and for a data set of size  $N$  and the parameters used throughout the experiments, the average runtime per joint reconstruction-segmentation iteration was  $N \times 3 \times 10^{-5}$  seconds. Therefore, the cardiac cine experiments took roughly  $256^2 \times 30$  (dimension)  $\times 35$  (iterations)  $\times 3 \times 10^{-5} \approx 35$  minutes to converge. Brain tissue segmentation experiments used larger data sets and approximately took  $256^3$  (dimension)  $\times 35$  (iterations)  $\times 3 \times 10^{-5} \approx 5$  hours.

The bottleneck of the computation is on the patch-based sparse coding stage with OMP. More information about the complexity of this coding can be found in [20]. It was shown in chapter 4 how this step could be accelerated with a parallel coding of independent patches without affecting the results presented, or by changing some parameters such as reducing the number of dictionary atoms  $K$  or the sparsity index of the coding  $s$ . Alternatively, the replacement of patch-based coding by a faster global sparsity transform such as proposed by other CS methods could also be investigated.

## 6.6 Discussion

The experiments presented try to answer a very specific question. Given that we have at hand accelerated MR data similar as the one delivered by a CS acquisition protocol, what is the best use that we can make of it provided we would like to obtain the best possible segmentation? Results suggest that a joint reconstruction-segmentation can directly provide segmented data, but also that the quality of this segmentation can outperform a traditional workflow of information, where an intensity image is first independently reconstructed and then a segmentation algorithm is used for classification.

The joint reconstruction-segmentation algorithm presented is able to balance through a single parameter reconstruction and segmentation accuracy. In doing so it produces segmentation results for undersampled data that are closer to those obtained on the image without undersampling than using the separate approach. Moreover, the formulation as an energy functional allows for very simple modifications as additional weighting terms, and relying on the EM segmentation framework enables the simple use of spatial priors and MRFs as shown in figure 6.13. To conclude, experiments on brain tissue segmentation demonstrate that the segmentations obtained with such extensions on undersampled data can be accurate with respect to a manually labelled reference.

The approach presented to join reconstruction and analysis achieves accurate segmentations for very fast acquired acquisitions while addressing potential issues in the imaging pipeline. A joint reconstruction-analysis is shielded against error propagation through stages, which could be the cause of the decreased performance in the separate approach. Similarly, it proposes a more efficient handling of information. Segmentation algorithms usually expect high detailed reconstructed images that are time consuming and expensive to produce in MRI. However, a segmentation is a much more concise and simple representation of the data in terms of information. With the introduction of the mixture model as part of the reconstruction we directly look for this simple representation. This modification in modelling does not necessarily improve the reconstruction objectively unless the true data follows a GMM, or improves it very subtly as shown in figure 6.12c, but it conditions the data towards a more accurate segmentation.

These results motivate the design of new methods to handle medical image data that break out of the traditional serial pipeline, and where the end application and utility of the image assumes a primary role. These new methods, which we refer to as application-driven, could also integrate acquisition and perhaps interpretation stages, providing dedicated scanning procedures that are more efficient and which directly and accurately reveal clinically relevant parameters such as ejection fraction or hippocampal volume changes in follow-up scans. It is also important to note that not all MRI scans could be application-driven because sometimes they are exploratory and

the relevant parameters looked for in the image are not known in advance.

Caution should nevertheless be taken when analysing the results presented in this work. In particular, it should be noted that the setup of the experiments is not identical to a practical situation because they simulate a single coil scenario and do not take in charge the complex-valued nature of MR images. Nonetheless as previously mentioned, we anticipate that the patch-based reconstruction method could be adapted to parallel complex-valued reconstructions as in the previous chapter, and the extension of the mixture model to represent complex data is straightforward. The analysis of parallel and complex-valued experiments however requires individual treatment and are left as future work. Additionally, undersampling trajectories as the one shown in figure 6.6c achieve high aliasing incoherence but are not in all cases feasible. Alternatives to 1D Cartesian sampling, such as radial or spiral sampling, could also be investigated.

## 6.7 Conclusion

We have presented a method for joint reconstruction and segmentation of undersampled MR data and have applied it for cardiac and brain tissue segmentation. The experiments compared the joint method with a separate approach, where data was first non-linearly reconstructed using CS and then subsequently segmented. The joint method proved to be more accurate than the separate for left ventricle segmentation in cardiac cine MR when compared against a segmentation of the fully sampled data. Moreover, extending the method with probabilistic priors and MRF regularisation, a brain tissue segmentation was achieved from highly under-sampled data comparable to a manual labelling of fully sampled scans. Again, the joint method outperformed the separate in terms of Dice metrics.

In scenarios where the end application of the imaging experiment is known *a priori*, it may be beneficial to purposely redesign the procedure for information extraction leading to a symbiotic joint processing such as the method presented. Future work could explore the direct extraction

of clinical parameters and investigate alternative designs for acquisition trajectories that are dedicated for a precise application. Such application-driven scans could tackle inefficiencies and error propagation of the traditional imaging workflow, ultimately opening the spectrum of scanning procedures to include cheaper and more efficient solutions for MRI.

# Chapter 7

## Conclusion

Medical imaging constitutes a fundamental component of medical research and diagnosis. The demand for MRI has risen particularly fast because it is non-invasive, non-ionising, and is a versatile tool to explore the structure and function of the human body. The technology underlying MR scanners is nevertheless far from ideal, and suffers from a number of inefficiencies that ballast its widespread use and availability, and the main objective of this thesis has been to provide solutions to these inefficiencies. In this chapter we summarise the contents of the thesis and we identify its achievements and limitations. We conclude with a discussion taking a broader outlook at the research undertaken.

We have identified and targeted three sources of inefficiency that are challenging the progress of MRI:

- Slow raw data acquisition,
- High information reduction from acquisition to analysis,
- Lack of reuse of past scans to improve future scans.

We have focussed on the case of dynamic cardiac imaging. Dynamic MR acquisition is particularly affected by speed constraints because it needs to trade-off spatial and temporal resolutions

against each other. This causes discomfort to patients, who are requested to perform a breath-hold for the duration of the scan, and prolongs imaging routines that need to resort to ECG gating. Fortunately, cardiac cine data has the advantage that it targets the inspection of relatively slow temporal changes, making it highly redundant in time and hence allowing to drastically reduce sampling requirements. Furthermore, the analysis of cardiac cine imaging in some cases reduces to the delineation of ventricle and myocardium boundaries, for an estimation of ejection fraction and cardiac mass. This is very succinct information to be derived from the highly complex raw data acquired, which is why more efficient paradigms for analysis information extraction would be plausible.

In this thesis, we have explored the applicability of recently proposed machine learning methods using adaptive dictionaries for fast and informative dynamic MRI acquisition. Sparse image penalties have shown promise in the acceleration of MRI acquisition drawing from CS theory, although this can be at the expense of some loss in image quality when undersampling becomes extreme. We have shown how the use of overcomplete and adaptive sparse transforms instead of common complete and fixed transforms can reduce the impact on image degradation that scan acceleration has. The additional ability to train dictionaries on past images also presents one way forward to learn information from past images that could be employed to improve the quality of future scans. Finally, the energy minimisation formulation of dictionary-based reconstructions is well suited for its combination with simpler models that can reveal analysis information about the image as a by-product of the reconstruction process.

## 7.1 Achievements

### Application of DL to cardiac MRI acceleration

The application of DL to cardiac cine data acquisition was proposed in chapter 4 in a simulated single MR coil scenario. The objective was to compare the reconstruction capabilities from an accelerated scan provided by the proposed method based on DL sparsity, and a competing

method based on a non-adaptive transform sparsity. The results obtained demonstrated that the flexibility of the transform derived from its overcompleteness can greatly improve reconstruction results in reducing the acceleration aliasing. It was shown to maintain reconstruction PSNR above 30 dB within a heart ROI for acceleration rates up to 8 fold, whereas the competing method could only guarantee this metric at a maximum of 4 fold acceleration. Irrespective of the acceleration rate, it was shown to achieve an increase of 5 dB in PSNR inside a heart ROI relative to the competing method, and was also validated to be more robust to noise.

### **Extension of the method to parallel MRI**

The concept of overcomplete and adaptive patch-based sparsity was successfully applied in the parallel coil setup in chapter 5, where the data is acquired by more than one coil hence generating redundancy through hardware. Parallel MRI developments were introduced earlier than CS, and have been established as the norm to accelerate MR scans by almost 4 fold. It is therefore important that the method is able to process parallel coil data, and this was achieved as an extension of the SPIRiT method. We demonstrated good performance in retrospective cardiac cine experiments compared to complete and non-adaptive transforms. Moreover, tests on prospectively acquired data showed an important quality enhancement for accelerated scans of up to a 8 fold acceleration compared to the scanner's reconstruction. The method took part in the ISMRM 2014 cardiac cine reconstruction challenge and ranked on average 10 out of 22 participants, with results comparable to the top performing method in cases using either Poisson or regular time-varying undersampling.

### **Joint reconstruction and segmentation from raw data**

In chapter 6 we went beyond MR acquisition speed limitations and investigated the joint reconstruction and analysis of MR data to suggest a more efficient handling of information from acquisition to clinical information extraction. Building upon the DL based reconstruction method for undersampled data, we proposed a joint reconstruction and segmentation for MR

images by forcing the reconstructed data to fit a discriminative model at the same time as it was being reconstructed. Traditionally, CS for MRI has focussed on providing the most accurate image reconstruction, even though many times in practice images are simply a means for clinical parameter measurement. Jointly reconstructing and segmenting MR data not only proved to be a more efficient mechanism to yield segmented images of cardiac cine data, but was also demonstrated to be more accurate than a disjoint handling of undersampled data reconstruction and segmentation, with around 2% less left ventricle pixel misclassification. This allowed to maintain left ventricle pixel misclassification below 18% for undersampling rates up to 14 fold in a single coil setup. The framework was also shown to be applicable to 3D brain tissue classification incorporating probabilistic priors and MRF regularisation, achieving over a 0.87 Dice coefficient relative to manual segmentations even for extreme 24 fold simulated undersampling rates. Most notably, the concept termed application-driven MRI could be extended to other applications in an attempt to efficiently grasp only useful information as early in the imaging pipeline as raw data acquisition.

## 7.2 Limitations and future work

### Computational cost

One of the major limitations of the application of DL to MRI acquisition acceleration is computational complexity, which became apparent in chapter 4. It was noted that optimal results for an acceleration rate of 6 fold required a computing time of 6.6 hours. However, relaxing memory constraints, using a simple dynamic choice of tuning parameters and parallelising computation over 8 CPU cores reduced runtime to 6 minutes without compromising image quality. The use of parallel OMP computations in a GPU architecture as proposed in [51, 14] is expected to dramatically cut runtime, and more clever reconstruction initialisations could also be explored as future work. A study submitted for publication [10] shows how an initialisation based on manifold learning that is very fast to obtain, can cut down the runtime of DL based

reconstruction by almost 6 times. These extensions could potentially speed up reconstruction time to match fast competing methods such as k-t FOCUSS, which converges in less than 1 minute.

### **Undersampling strategy for parallel MRI**

The extension of the method to parallel imaging in chapter 5 encountered other challenges. Although computational efforts in the parallel reconstruction method proposed are still considerable, the convergence of the method was increased by the initial reconstruction provided by the SPIRiT algorithm. The biggest limitation encountered were restrictions on data sampling. Participating in the ISMRM challenge helped reveal large differences in reconstruction quality depending on the undersampling mask design used. Random undersampling, allowing large gaps without samples in k-space, proved to be a bad choice for the method presented. This was the case for all participating methods, but appeared to be particularly harmful for our method because of the coil recombination mechanism of SPIRiT. However, CS methods theoretically thrive with random undersampling because it offers maximal sampling incoherence. It would be interesting to analyse the importance of incoherent sampling in hybrid parallel CS methods like the one presented here considering parallel imaging undersampling constraints. Also, a necessary extension to the results shown in the thesis would be to implement a CS compatible undersampling scheme to present prospective results under optimal conditions.

### **Feeding information from analysis to reconstruction domain**

A major limitation for the development of application-driven MRI as presented in chapter 6 has been designing a constructive interaction between raw data and the simpler discriminative model from which we can infer clinical parameters. In the segmentation application discussed in this work, we exploited the generative properties of a GMM to feed intensity information from a given labelling back to the image domain with a sense for uncertainty. This framework however brakes down if probabilistic prior information about the segmentation is incorporated into the

model, because they do not only carry intensity based information. Although comparisons in brain tissue segmentation from accelerated scans relative to manual segmentations were successful and convincing, they deviated from the fully sampled case more when data was jointly reconstructed and segmented than when it was processed disjointly. This is a limitation that is likely to be encountered if the concept is applied to a different application, and novel techniques to directly link clinical measures to raw data will have to be engineered.

### 7.3 Discussion

In this concluding section, we reflect upon the thesis's initial motivation, the discoveries made along the way and the outlook for future related research.

#### Thesis overview

The starting point for this thesis was to explore the potential use of sparse and adaptive dictionaries for cardiac MRI sequence reconstructions. Trained dictionaries are a very recent natural evolution of signal representation by frames, popularised amongst others by Elad [49], and had been proposed for undersampled MRI reconstruction by Ravishankar and Bresler [117] with considerable improvements in performance compared to the pioneering work of Lustig *et al.* in [86].

Cardiac cine acquisition was targeted given that it imposes particularly challenging sampling conditions and has the advantage of being highly redundant in time. These realisations were used as premises to design the reconstruction method DLTG, presented in chapter 4, where regularisation with a learnt dictionary combined with an additional temporal gradient constraint proved to be successful and robust. The method was extended in chapter 5 to provide the first solution using DL for parallel undersampled MRI reconstruction, and was validated on prospectively accelerated acquisitions.

However, in the course of our investigation the limitations of patch-based dictionary regularisation have also become apparent. Although quantitative comparisons prove satisfactory against complete and fixed regularisation methods, a recurring side effect of this regularisation when undersampling is high is a loss of fine details. This is noticeable both in the single coil simulations and in parallel coil tests. We reproduce in figure 7.1 the visual result of a single coil accelerated simulation of 10 fold. The structure and dynamism of the beating heart is nevertheless relatively well recovered. This recurring phenomenon led to new research questions: Can we afford the loss of fine detail? How much information are we really losing?

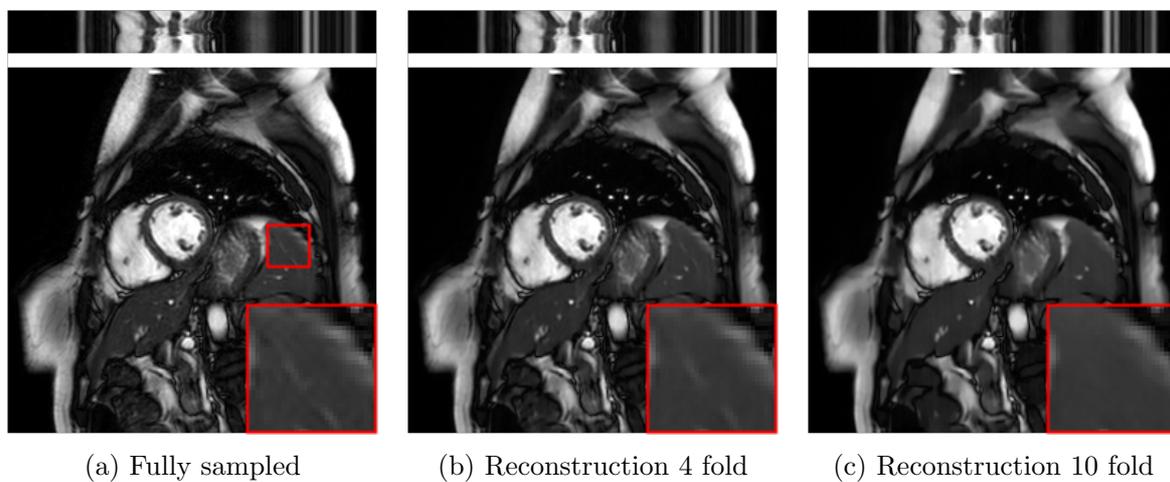


Figure 7.1: Detail loss at high undersampling factors shown for a single coil simulation. The case of 4 fold undersampling can recover fine details, but at high factors such as 10 fold fine details are lost by the patch-based regularisation method.

The answer to these questions is clearly dependent on the kind of knowledge we would like to extract from the image, leading to the application-driven paradigm presented in chapter 6, where the end goal of the imaging experiment must define the acquisition and reconstruction stages. This concept is broad, and the heterogeneity of clinically useful information makes it an abstract idea that could be materialised very differently depending on the application. We implemented a joint reconstruction and segmentation mechanism for extracting clinical information from cardiac cine scans, demonstrating the benefits of combining reconstruction and analysis.

Questioning the utility of a given image reconstruction broadened the scope of the thesis,

which could then focus on addressing inefficiencies in the entire MR imaging pipeline, where acquisition speed limitations are only one part of the problem. The advance of MR technology will steadily bring better, more precise scanning, which will enable new applications and improve the lives of patients, but a major breakthrough in imaging practice could come from smarter scanning.

### Thesis outlook

The shown reconstruction-segmentation method is one example of smarter imaging, but future research could try to take it further. An image segmentation is not clinical knowledge *per se* as it still requires some interpretation, so one possibility would be to try to look for clinical measures such as ejection fraction or cardiac mass directly from raw data as a by-product of reconstruction. Other routine applications could also be targeted. The assessment of Alzheimer's disease progress can for example be performed from an estimation of structural changes in the hippocampus, where multi-atlas registration and segmentation has proven to be very successful [5, 150]. A registration process does not normally exploit the entire raw data acquired, so there could be room for an imaging process that targets optimal registration against another image with reduced sampling requirements.

Proposing an imaging process where the image is not the main interest may seem contradictory, and could raise concerns for radiologists who are used to examine images that meet certain clinical requirements and translate them into knowledge. However, these innovative methods could be of great interest if the measurements they provide can be shown to be as accurate within a confidence interval as the ones obtained with traditional scanning, but with a much more efficient, faster and therefore cheaper acquisition.

We have mentioned how the use of overcomplete, adaptive dictionaries for accelerated MR imaging indirectly led to the definition of the broad thesis topic, but they are not indispensable. Patch-based dictionaries proved to be a good fit for a joint reconstruction and segmentation given that they preserve sharp edges and miss fine details, which can be beneficial for image

segmentation, but there have been extensions to this representation that could improve the work of this thesis. Some examples are the use of multi-resolution DL [106] or training features on domains sparser than the image domain [82]. Another interesting avenue could be to use not only the reconstructive power of patch-based dictionaries, but also employ other capabilities they are known for in the application-driven scenario, such as image segmentation [90], super-resolution [153] or inpainting [89].

Another front which has been very much unexplored in this thesis has been novel k-space undersampling. The sampling used for the collection of undersampled data was only ensured to contain some degree of randomness such as to comply with the incoherence requirements of CS reconstruction, but it is very likely that better solutions exist. It has already been shown that adapting the undersampling to the dataset can be beneficial for data recovery [116], and other designs could be explored in the context of application-driven MRI. The acquisition strategy of the joint reconstruction and segmentation method could for instance highlight sharp edges by missing low frequency components, which is the opposite of what was done in this work.

To conclude, in this thesis we have distinguished different sources of inefficiencies in acquisition and processing, and have adapted DL techniques for MRI to propose a possible solution, with its own advantages and limitations. This discussion highlights nonetheless the many different avenues that could have been taken and which open up possibilities to further explore the topics in this thesis. The pairing of machine learning with imaging process and the ever increasing amounts of MR data produced are bound to bring great advances to medical imaging and continue influencing the lives of patients.

# Bibliography

- [1] G. Adluru, C. McGann, P. Speier, E. G. Kholmovski, A. Shaaban, and E. V. R. Dibella. Acquisition and reconstruction of undersampled radial data for myocardial perfusion magnetic resonance imaging. *Journal of Magnetic Resonance Imaging*, 29(2):466–473, 2009. 3, 63
- [2] M. Aharon, M. Elad, and A. Bruckstein. K -SVD : An algorithm for designing over-complete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006. 48, 70
- [3] C. B. Ahn, J. H. Kim, and Z. H. Cho. High-speed spiral-scan echo planar NMR imaging-I. *IEEE transactions on medical imaging*, 5(1):2–7, 1986. 23
- [4] M. Akçakaya, S. Nam, P. Hu, M. Moghari, L. Ngo, V. Tarokh, W. Manning, and R. Nezafat. Compressed sensing with wavelet domain dependencies for coronary MRI: A retrospective study. *IEEE Transactions on Medical Imaging*, 30(5):1090–1099, 2011. 98
- [5] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage*, 46(3):726–738, 2009. 172
- [6] S. P. Awate and E. V. R. Dibella. Spatiotemporal dictionary learning for undersampled dynamic MRI reconstruction via joint frame-based and dictionary-based sparsity. In *Proceedings of the International Symposium on Biomedical Imaging*, pages 318–321, 2012. 69

- [7] R. G. Baraniuk. Compressive Sensing. *IEEE Signal Processing Magazine*, 24(4):118–124, 2007. 31
- [8] A. Beck and M. Teboulle. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. 109
- [9] G. Beylkin. On the Representation of Operators in Bases of Compactly Supported Wavelets. *SIAM Journal on Numerical Analysis*, 29(6):1716–1740, 1992. 45
- [10] K. K. Bhatia, J. Caballero, A. N. Price, Y. Sun, J. V. Hajnal, and D. Rueckert. KARMA: k-Space augmentation of random manifold projections for fast reconstruction of dynamic MRI. *Magnetic Resonance in Medicine*, Submitted, 2015. 168
- [11] O. Bieri, M. Markl, and K. Scheffler. Analysis and compensation of eddy currents in balanced SSFP. *Magnetic Resonance in Medicine*, 54(1):129–37, 2005. 77
- [12] K. Bio Imaging and Signal Processing Lab. k-t FOCUSS. 77
- [13] J. M. Bioucas-Dias and M. A. T. Figueiredo. An iterative algorithm for linear inverse problems with compound regularizers. *2008 15th IEEE International Conference on Image Processing*, 2:685–688, 2008. 41
- [14] J. D. Blanchard and J. Tanner. GPU accelerated greedy algorithms for compressed sensing. *Mathematical Programming Computation*, 5(3):267–304, 2013. 168
- [15] F. Bloch. Nuclear induction. *Physical Review*, 70(7-8):460–474, 1946. 14
- [16] K. T. Block, M. Uecker, and J. Frahm. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine*, 57(6):1086–1098, 2007. 59, 77
- [17] J. Boldt, A. W. Leber, K. Bonaventura, C. Sohns, M. Stula, A. Huppertz, W. Haverkamp, and M. Dorenkamp. Cost-effectiveness of cardiovascular magnetic resonance and single-photon emission computed tomography for diagnosis of coronary artery disease in Germany. *Journal of Cardiovascular Magnetic Resonance*, 15(1):30, 2013. 7

- [18] G. E. P. Box. Robustness in the Strategy of Scientific Model Building. In *Robustness in Statistics*, pages 201–236. 1979. 32
- [19] R. W. Brown, Y. C. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014. 1, 10, 18, 19, 20, 26
- [20] J. Caballero, W. Bai, A. N. Price, D. Rueckert, and J. V. Hajnal. Application-driven MRI: Joint reconstruction and segmentation from undersampled MRI data. In *Proceedings of the 17th International Conference on Medical Imaging Computing and Computer Assisted Interventions (MICCAI)*, pages 106–113, 2014. 161
- [21] J. Caballero, D. Rueckert, and J. V. Hajnal. Dictionary learning and time sparsity in dynamic MRI. In *Medical Image Computing and Computer Assisted Interventions (MICCAI)*, pages 256–263, 2012. 66
- [22] J. Caballero, J. A. Urigüen, S. R. Schultz, and P. L. Dragotti. Spike sorting at sub-Nyquist rates. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 585–588, 2012. 69
- [23] E. J. Candès and D. L. Donoho. Curvelets: A surprisingly effective nonadaptive representation for objects with edges. Technical report, 1999. 45
- [24] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73, 2011. 60
- [25] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006. 36

- [26] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006. 52, 63
- [27] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005. 36, 40, 65
- [28] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008. 30, 31, 32
- [29] A. Chambolle. Total variation minimization and a class of binary MRF models. *Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2005. 71
- [30] Y. Chang, K. F. King, D. Liang, Y. Wang, and L. Ying. A kernel approach to compressed sensing parallel MRI. *IEEE International Symposium on Biomedical Imaging*, pages 78–81, 2012. 100
- [31] C. Chen and J. Huang. The benefit of tree sparsity in accelerated MRI. *Medical Image Analysis*, 18(6):834–842, 2014. 98
- [32] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998. 40
- [33] R. R. Coifman, Y. Meyer, and V. Wickerhauser. Wavelet analysis and signal processing. *Wavelets and their Applications*, pages 153–178, 1992. 44
- [34] J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, 19(90):297–301, 1965. 43
- [35] J. Cuppen and A. van Est. Reducing MR imaging time by one-sided reconstruction. *Magnetic Resonance Imaging*, 5(6):526–527, 1987. 26
- [36] L. Curtis. Unit costs of health & social care 2011. *Personal Social Services Research Unit*, 2011. 7

- [37] J. A. D'Arcy, D. J. Collins, I. J. Rowland, A. R. Padhani, and M. O. Leach. Applications of sliding window reconstruction with cartesian sampling for dynamic contrast enhanced MRI. *NMR in biomedicine*, 15(2):174–183, 2002. 27
- [38] I. Daubechies, R. Devore, M. Fornasier, and C. S. Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63:1–38, 2010. 41
- [39] R. A. DeVore, B. Jawerth, and B. J. Lucier. Image compression through wavelet transform coding. *IEEE Transactions on Information Theory*, 38(2):719–746, 1992. 117
- [40] M. N. Do and M. Vetterli. Contourlets : A directional multiresolution image representation. *Image Processing*, 1:357–360, 2002. 45
- [41] M. Doneva, P. Börnert, H. Eggers, C. Stehning, J. Sénégas, and A. Mertins. Compressed sensing reconstruction for magnetic resonance parameter mapping. *Magnetic Resonance in Medicine*, 64(4):1114–1120, 2010. 60
- [42] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. 30, 32
- [43] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–202, Mar. 2003. 35, 36
- [44] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6–18, 2006. 35, 63
- [45] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001. 36

- [46] D. L. Donoho and Y. Tsaig. Fast solution of l1-norm minimization problems when the solution may be sparse. *IEEE Transactions on Information Theory*, 54(11):4789–4812, 2008. 41
- [47] 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(2):1094–1121, 2012. 39
- [48] M. Elad. Sparse and redundant representation modeling - What next ? *IEEE Signal Processing Letters*, 19(12):922–928, 2012. xix, 33
- [49] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. 84, 94, 95, 98, 170
- [50] K. Engan, S. O. Aase, and J. H. Husoy. Method of optimal directions for frame design. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 2443–2446, 1999. 47
- [51] Y. Fang, L. Chen, J. Wu, and B. Huang. GPU implementation of orthogonal matching pursuit for compressive sensing. In *IEEE 17th International Conference on Parallel and Distributed Systems*, pages 1044–1047, 2011. 168
- [52] J. A. Fessler. On NUFFT-based gridding for non-Cartesian MRI. *Journal of Magnetic Resonance*, 188(2):191–195, 2007. 77
- [53] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani. Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2):302–332, 2007. 41
- [54] D. Gabor. Theory of communication. *Journal of the Institution of Electrical Engineers*, 93(26):429–441, 1946. 44
- [55] U. Gamper, P. Boesiger, and S. Kozerke. Compressed sensing in dynamic MRI. *Magnetic Resonance in Medicine*, 59(2):365–73, 2008. 3, 24, 58, 98

- [56] R. Giryes and M. Elad. Can we allow linear dependencies in the dictionary in the sparse synthesis framework? In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5459–5463, 2013. 35, 37, 60
- [57] I. Gorodnitsky and B. Rao. Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, 45(3):600–616, 1997. 42
- [58] I. F. Gorodnitsky, J. S. George, and B. D. Rao. Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. *Electroencephalography and clinical neurophysiology*, 95(4):231–251, 1995. 42
- [59] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–10, 2002. 18, 24, 26, 97, 99
- [60] A. Grossmann and J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, 15(4):723–736, 1984. 44
- [61] R. H. Hashemi, W. G. Bradley, and C. J. Lisanti. *MRI: The basics*. 2012. 12
- [62] I. T. Hsiao, A. Rangarajan, and G. Gindi. Joint-MAP reconstruction/segmentation for transmission tomography using mixture-models as priors. In *IEEE Nuclear Science Symposium*, volume 3, pages 1689–1693, 1998. 134
- [63] X. Hu and T. Parrish. Reduction of field of view for dynamic imaging. *Magnetic Resonance in Medicine*, 31(6):691–694, 1994. 27
- [64] F. Huang, Y. Chen, W. Yin, W. Lin, X. Ye, W. Guo, and A. Reykowski. A rapid and robust numerical algorithm for sensitivity encoding with sparsity constraints: Self-feeding sparse SENSE. *Magnetic Resonance in Medicine*, 64(4):1078–1088, 2010. 103

- [65] J. Huang, S. Zhang, and D. Metaxas. Efficient MR image reconstruction for compressed MR imaging. *Medical Image Analysis*, 15(5):670–679, 2011. 98
- [66] ISMRM. ISMRM 2013 challenge on sub-Nyquist reconstruction of dynamic imaging. 116
- [67] I. Jolliffe. *Principal component analysis*. John Wiley & Sons, 2005. 44
- [68] R. A. Jones, O. Haraldseth, T. B. Müller, P. A. Rinck, and A. N. Oksendal. Kspace substitution: A novel dynamic imaging technique. *Magnetic resonance in medicine*, 29(6):830–834, 1992. 27
- [69] H. Jung, K. H. Sung, K. S. Nayak, and E. Y. Kim. K-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI. *Magnetic Resonance in Medicine*, 61(1):103–116, 2008. 3, 58
- [70] H. Jung and J. C. Ye. Motion estimated and compensated compressed sensing dynamic magnetic resonance imaging: What we can learn from video compression techniques. *International Journal of Imaging Systems and Technology*, 20(2):81–98, 2010. 58, 98
- [71] H. Jung, J. C. Ye, and E. Y. Kim. Improved k-t BLAST and k-t SENSE using FOCUSS. *Physics in Medicine and Biology*, 52(11):3201–3226, 2007. 58, 63, 77
- [72] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An interior-point method for large-scale  $l_1$ -regularized least squares. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):606–617, 2007. 41
- [73] D. A. Langan and K. J. Molnar. Use of the mean-field approximation in an EM-based approach to unsupervised stochastic model-based image segmentation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 57–60, 1992. 147
- [74] E. Le Pennec and S. Mallat. Sparse geometric image representations with bandelets. *IEEE Transactions on Image Processing*, 14(4):423–438, 2005. 45

- [75] M. H. Levitt. *Spin dynamics: Basics of nuclear magnetic resonance*. John Wiley & Sons, 2008. 11, 13
- [76] D. Liang, B. Liu, J. Wang, and L. Ying. Accelerating SENSE using compressed sensing. *Magnetic Resonance in Medicine*, 62(6):1574–84, Dec. 2009. 100
- [77] Z.-P. Liang and P. C. Lauterbur. An efficient method for dynamic magnetic resonance imaging. *Transactions on Medical Imaging*, 13(4):677–686, 1994. 27
- [78] Z.-P. Liang and P. C. Lauterbur. *Principles of magnetic resonance imaging: A signal processing perspective*. IEEE Press, Piscataway, NJ, 2000. 11, 12, 13, 15, 16, 17, 19, 21
- [79] R. Likes. Moving gradient zeugmatography, 1981. 23
- [80] Lin FaHsuan, K. K. Kwong, J. W. Belliveau, and L. L. Wald. Parallel imaging reconstruction using automatic regularization. *Magnetic Resonance in Medicine*, 51(3):559–567, 2004. 97
- [81] S. G. Lingala and M. Jacob. Blind compressed sensing with sparse dictionaries for accelerated dynamic MRI. In *IEEE International Symposium on Biomedical Imaging*, pages 5–8, 2013. 61
- [82] Q. Liu, S. Wang, L. Ying, X. Peng, Y. Zhu, and D. Liang. Adaptive dictionary learning in sparse gradient domain for image recovery. *IEEE Transactions on Image Processing*, 22(12):4652–4663, 2013. 173
- [83] M. Lorenzo-Valdés, G. I. Sanchez-Ortiz, A. G. Elkington, R. H. Mohiaddin, and D. Rueckert. Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. *Medical Image Analysis*, 8(3):255–265, 2004. 135
- [84] M. Lustig, D. L. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007. 3, 30, 31, 56, 57, 59, 63, 66, 77, 97, 98

- [85] M. Lustig and J. M. Pauly. SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space. *Magnetic Resonance in Medicine*, 64(2):457–471, 2010. 3, 97, 100, 101, 104, 125
- [86] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly. K-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity. In *ISMRM*, volume 50, page 2420, 2006. 3, 23, 24, 57, 170
- [87] S. Ma and W. Yin. An efficient algorithm for compressed MR imaging using total variation and wavelets. *Computer Vision and ...*, 2008. 66
- [88] B. Madore, G. H. Glover, and N. J. Pelc. Unaliasing by fourier-encoding the overlaps using the temporal dimension (UNFOLD), applied to cardiac imaging and fMRI. *Magnetic Resonance in Medicine*, 42(5):813–828, 1999. 27
- [89] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. In *Annual International Conference on Machine Learning*, pages 689–696, 2009. 48, 173
- [90] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. In *Computer Vision and Pattern Recognition*, pages 1–8, 2008. 33, 173
- [91] S. G. Mallat. *A wavelet tour of signal processing*. Academic press, 1999. 33, 35, 44
- [92] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993. 38
- [93] P. Mansfield. Multi-planar image formation using NMR spin echoes. *Journal of Physics C: Solid State Physics*, 10(3):L55–L58, 1977. 23
- [94] P. Margosian, F. Schmitt, and D. Purdy. Faster MR imaging: Imaging with half the data. *Health Care Instrumentation*, 1(6):195–197, 1986. 26
- [95] M. U. McConnell Brain Imaging Centre. BrainWeb: Simulated Brain Database. xix, xxv, 18, 139, 149, 150, 157

- [96] G. McGibney, M. R. Smith, S. T. Nichols, and A. Crawley. Quantitative evaluation of several partial Fourier reconstruction algorithms used in MRI. *Magnetic Resonance in Medicine*, 30(1):51–59, 1993. 26
- [97] L. B. Montefusco, D. Lazzaro, S. Papi, and C. Guerrini. A fast compressed sensing approach to 3D MR image reconstruction. *IEEE Transactions on Medical Imaging*, 30(5):1064–1075, 2011. 3, 59, 63, 66
- [98] National Audit Office. Department of Health: Managing high value capital equipment in the NHS in England. Technical report, 2011. 7
- [99] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009. 40
- [100] D. Needell and R. Vershynin. Signal recovery from inaccurate and incomplete measurements via regularized orthogonal matching pursuit. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):310–316, 2010. 39
- [101] NHS. Annual Imaging and Radiodiagnostics Data. Technical report, 2014. xviii, 6
- [102] D. C. Noll. Multishot rosette trajectories for spectrally selective MR imaging. *IEEE Transactions on Medical Imaging*, 16(4):372–377, 1997. 23
- [103] D. C. Noll, D. G. Nishimura, and A. Macovski. Homodyne detection in magnetic resonance imaging. *IEEE transactions on medical imaging*, 10(2):154–163, 1991. 26
- [104] H. Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928. 30
- [105] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37(23):3311–3325, 1997. xix, 33, 47
- [106] B. Ophir, M. Lustig, and M. Elad. Multi-scale dictionary learning using wavelets. *IEEE Journal of Selected Topics in Signal Processing*, 5(5):1014–1024, 2011. 173

- [107] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005. 41
- [108] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson. Combination of compressed sensing and parallel imaging for highly accelerated firstpass cardiac perfusion MRI. *Magnetic Resonance in Medicine*, 64(3):767–776, 2011. 100
- [109] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pages 40–44, 1993. 39, 141
- [110] J. P. Pipe. Motion correction with PROPELLER MRI: Application to head motion and free-breathing cardiac imaging. *Magnetic Resonance in Medicine*, 42(5):963–969, 1999. 23
- [111] M. Protter and M. Elad. Image sequence denoising via sparse and redundant representations. *IEEE Transactions on Image Processing*, 18(1):27–35, 2009. 61, 94, 95
- [112] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, 1999. 18, 24, 26, 76, 97, 99, 151
- [113] T. I. I. o. T. R. Rubinstein, Dept. of Computer Science. KSVD-Box v13 and OMP-Box v10. 75, 90
- [114] V. Rasche, R. W. De Boer, D. Holz, and R. Proksa. Continuous radial data acquisition for dynamic MRI. *Magnetic Resonance in Medicine*, 34(5):754–761, 1995. 23
- [115] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *IEEE Transactions on Information Theory*, 54(5):2210–2219, 2008. 60

- [116] S. Ravishankar and Y. Bresler. Adaptive sampling design for compressed sensing MRI. In *Annual International Conference of the IEEE on Engineering in Medicine and Biology Society*, pages 3751–3755. IEEE EMBS, 2011. 173
- [117] S. Ravishankar and Y. Bresler. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Transactions on Medical Imaging*, 30(5):1028–1041, 2011. 8, 59, 60, 66, 68, 74, 86, 150, 170
- [118] R. Rubinstein, A. M. Bruckstein, and M. Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, 2010. 35, 43
- [119] R. Rubinstein, M. Zibulevsky, and M. Elad. Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. *CS Technion*, 40(8):1–15, 2008. 49, 75, 161
- [120] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008. 36, 56
- [121] P. Schmid-Saugeon and A. Zakhor. Dictionary design for matching pursuit and application to motion-compensated video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(6):880–886, 2004. 48
- [122] J. Senegas, C. A. Cocosco, and T. Netsch. Model-based segmentation of cardiac MRI cine sequences: A Bayesian formulation. *Proceedings of SPIE*, 5370:432–443, 2004. 135
- [123] C. E. Shannon. Communication in the presence noise. *Proceedings of the IRE*, 37(1):10–21, 1949. 30
- [124] M. Shiqian, Y. Wotao, Z. Yin, and C. Amit. An efficient algorithm for compressed MR imaging using total variation and wavelets. In *Computer Vision and Pattern Recognition*, pages 1–8, 2008. 59
- [125] Siemens. From big data to smart data: Infographic smart data. 29

- [126] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, 38(2):587–607, 1992. 45
- [127] A. Skodras, C. Christopoulos, and T. Ebrahimi. The JPEG 2000 still image compression standard. *IEEE Signal Processing Magazine*, 18(5):36–58, 2001. 44
- [128] A. E. Smith, D. W. Coit, and Z. Michalewicz. Penalty functions. In *Handbook of Evolutionary Computation*, chapter Section C. 1997. 104
- [129] D. K. Sodickson and W. J. Manning. Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine*, 38(4):591–603, 1997. 24, 26
- [130] C. Stubbs, M. Brenner, A. Despain, R. Henderson, D. Long, W. Press, J. Tonry, and P. Weinberger. The computational challenges of medical imaging. Technical Report 3, JASON-MITRE, June 2004. 132
- [131] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999. 161
- [132] The Economist. Data, data everywhere. *Special Report: Managing Information*, 2010. xix, 29
- [133] Z. Tian and G. B. Giannakis. Compressed sensing for wideband cognitive radios. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 4, page 1357, 2007. 31
- [134] I. Tasic and P. Frossard. Dictionary learning. *IEEE Signal Processing Magazine*, 28(2):27–38, 2011. 43, 98
- [135] J. Tropp and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007. 103

- [136] 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*, 50(5):1031–1042, 2003. 24, 26, 27, 58
- [137] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, and M. Lustig. ESPIRiT - An eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, 71(3):990–1001, 2014. 106, 125
- [138] M. Usman, C. Prieto, T. Schaeffter, and P. G. Batchelor. K-t group sparse: A method for accelerating dynamic MRI. *Magnetic Resonance in Medicine*, 1176:1163–1176, 2011. 3, 63
- [139] D. Van de Sompel and M. Brady. Simultaneous reconstruction and segmentation algorithm for positron emission tomography and transmission tomography. In *IEEE International Symposium on Biomedical Imaging*, pages 1035–1038, 2008. 134
- [140] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based bias field correction of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):885–896, 1999. 135, 147
- [141] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908, 1999. 135, 147, 148
- [142] J. J. Van Vaals, M. E. Brummer, W. Thomas Dixon, H. H. Tuithof, H. Engels, R. C. Nelson, B. M. Gerety, J. L. Chezmar, and J. a. Den Boer. Keyhole method for accelerating imaging of contrast agent uptake. *Journal of Magnetic Resonance Imaging*, 3(4):671–675, 1993. 27
- [143] L. Vandenberghe and S. Boyd. A primal-dual potential reduction method for problems involving matrix inequalities. *Mathematical Programming*, 69(1-3):205–236, 1995. 41

- [144] N. Vaswani. LS-CS-residual (LS-CS): Compressive sensing on least squares residual. *IEEE Transactions on Signal Processing*, 58(8):4108–4120, 2010. 59
- [145] M. Vlaardingerbroek and J. den Boer. *Magnetic resonance imaging: Theory and practice*. Springer, 1999. 1, 15
- [146] G. Wallace. The JPEG still picture compression standard. *IEEE Transactions on Consumer Electronics*, 38(1):xviii–xxxiv, 1992. 43
- [147] Y. Wang and L. Ying. Compressed sensing dynamic cardiac cine MRI using learned spatiotemporal dictionary. *IEEE Transactions on Biomedical Engineering*, 61(4):1109–1120, 2014. 60, 86
- [148] 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*, 13(4):600–612, 2004. 76
- [149] D. S. Weller, J. R. Polimeni, L. Grady, L. L. Wald, E. Adalsteinsson, and V. K. Goyal. Combined compressed sensing and parallel MRI compared for uniform and random cartesian undersampling of K-space. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 553–556, 2011. 100
- [150] R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, and D. Rueckert. LEAP: Learning embeddings for atlas propagation. *NeuroImage*, 49(2):1316–1325, 2010. 172
- [151] R. C. Wright, S. J. Riederer, F. Farzaneh, P. J. Rossman, and Y. Liu. Real-time MR fluoroscopic data acquisition and image reconstruction. *Magnetic Resonance in Medicine*, 12(3):407–415, 1989. 27
- [152] C. F. J. Wu. On the convergence properties of the EM algorithm. *The Annals of statistics*, 11(1):95–103, 1983. 138
- [153] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, 2010. 33, 173

- [154] J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1794–1801, 2009. 33
- [155] J. Yang, Y. Zhang, and W. Yin. A fast alternating direction method for TVL1-L2. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):288–297, 2010. 66
- [156] Z. Zhan, J.-F. Cai, D. Guo, Y. Liu, Z. Chen, and X. Qu. Fast Multi-class dictionaries learning with geometrical directions in MRI reconstruction. *arXiv preprint arXiv:1503.02945*, 2015. 141
- [157] J. Zhang. The mean field theory in em procedures for Markov random fields. *IEEE Transactions on Signal Processing*, 40(10):2570–2583, 1992. 147
- [158] Q. Zhang, R. Plemmons, D. Kittle, D. Brady, and S. Prasad. Joint segmentation and reconstruction of hyperspectral data with compressed measurements. *Applied Optics*, 50(22):4417–4435, 2011. 134
- [159] Q. Zhang, R. Plemmons, D. Kittle, S. Prasad, and B. Sciences. Reconstructing and segmenting hyperspectral images from compressed measurements. In *3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pages 1–4, 2011. 134
- [160] T. Zhang, J. M. Pauly, S. S. Vasanawala, and M. Lustig. Coil compression for accelerated imaging with Cartesian sampling. *Magnetic Resonance in Medicine*, 69(2):571–582, 2013. 125