Approximate Message Passing for Compressed Sensing Magnetic Resonance Imaging

Charles Millard
Exeter College
University of Oxford

A thesis submitted towards the degree of
Doctor of Philosophy

Michaelmas 2021

Disclaimer

The concepts and information presented in this thesis are based on research results that are not commercially available. Future availability cannot be guaranteed.

Acknowledgements

I am highly thankful for the guidance I have received from my supervisors Jared Tanner and Aaron Hess, who have been generous, patient and encouraging throughout my DPhil. I am also hugely grateful for the invaluable contributions from my industrial supervisor Boris Mailhé, without whom this thesis would not be possible.

I gratefully acknowledge EPSRC and Siemens for funding my studentship. I am particularly thankful to Mariappan Nadar and the Siemens reconstruction group in Princeton for welcoming me in to their team for two internships, which was highly enjoyable and enriching.

Thank you Peter Jezzard and the ONBI team for providing me with an enjoyable start to my time at Oxford and a solid foundation for my research career.

I have been fortunate to have been involved in two research groups simultaneously during my DPhil. Thank you to Nick Trefethen and my colleagues in Numerical Analysis, and to Damian Tyler and all the physicists in OCMR, for your many helpful insights.

Mark Chiew and Raphael Hauser gave me valuable feedback on my transfer and confirmation of status assessments, which I am very thankful for.

A special thanks to my office mates Michael Murray, Florentin Goyens, Pablo Martinez, Lindon Roberts, Ferenc Mozes and Liz Tunnicliffe, and to Thijs de Buck for kindly sharing his angiography data.

My Windmill Road friends Nicola Dinsdale, Veerle Brans, Charles Hill, Katie Banecki, Andy Tyler and Ryan Timms have been great housemates and provided a friendly home environment over the years, which I am so grateful for.

Thank you Jen for offering to proofread this thesis and for being an all-round lovely human being who I feel very lucky to have in my life. I am also very lucky to have a wonderful family. Thank you Mum, Dad, Granny, Emma and Joelle for your endless loving kindness and support.

Abstract

Magnetic Resonance Imaging (MRI) is a non-invasive, non-ionising imaging modality with unrivalled soft tissue contrast. A key consideration for MRI is data acquisition time, which is limited by inherent technological and physiological constraints. Compressed sensing is a relatively recent framework that can reduce the MRI acquisition time by undersampling randomly and exploiting presumed redundancies in the data.

The Approximate Message Passing (AMP) algorithm is an iterative compressed sensing method that efficiently reconstructs signals that have been sampled with i.i.d. sub-Gaussian sensing matrices. However, when Fourier coefficients of a signal with non-uniform spectral density are sampled, such as in MRI, AMP performs poorly in practice.

In response, this thesis proposes the Variable Density Approximate Message Passing (VDAMP) algorithm for undersampled MRI data. We present three versions of VDAMP: single-coil VDAMP, where receiver coil sensitivities are ignored, Parallel-VDAMP (P-VDAMP), which includes coil sensitivities, and Denoising-P-VDAMP (D-P-VDAMP), which incorporates the statistical modelling capabilities of neural networks. Central to VDAMP is a property that we term "coloured state evolution", where the difference between the intermediate image estimate at a given iteration and the ground truth is distributed according to a zero-mean Gaussian with known covariance. We demonstrate that coloured state evolution can be leveraged to yield an algorithm that converges rapidly, and to a competitive reconstruction quality, without the need to hand-tune model parameters.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Magnetic resonance imaging (MRI) is an imaging modality derived from radio frequency signals of atomic nuclei subject to a strong magnetic field. The underlying magnetic resonance physics of MRI is remarkably flexible and can be engineered to generate images that are sensitive to a number of contrast parameters [4]. Furthermore, unlike Computed Tomography (CT) and nuclear imaging techniques such as Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET), MRI does not use ionising radiation [5].

Acquisition time is a central consideration in many MRI applications, but is limited by inherent technological and physiological constraints. Measurements in MRI consist of smooth 1D curves through the 2D or 3D Fourier coefficients of the image, known as "k-space". The speed at which it is possible to traverse k-space is limited by the scanner's gradient specifications, in particular the maximum gradient amplitude and maximum slew-rate. Additionally, excessively high amplitude gradients and gradient switching can lead to peripheral nerve stimulation [6]. These intrinsic acquisition speed limitations have led researchers to investigate ways to accelerate MRI by undersampling.

Compressed sensing [7, 8] is a relatively recent framework that provides a way to

reconstruct accelerated MRI by exploiting presumed redundancies in the data [9]. Images of interest are well known to be compressible: for instance, an assumption of redundancy in the discrete cosine transform or wavelet transform are the bases of the JPEG and JPEG 2000 compression schemes respectively [10]. The core idea of these compression schemes is to transform the image in to a representation where most of the information is concentrated in a small number of coefficients, known as a "sparse representation". Given an assumption of sparsity, compressed sensing provides a toolbox of reconstruction algorithms that seek the sparsest solution that is consistent with the measurements. The primary success of compressed sensing theory is to establish that under certain conditions on the type of measurements, the number of measurements, and the degree of sparsity, it is possible to efficiently and robustly solve this problem [11].

The focus of this thesis is on the Approximate Message Passing (AMP) [12] algorithm, which efficiently reconstructs signals that have been sampled with large i.i.d. sub-Gaussian sensing matrices. Central to AMP is its "state evolution", which guarantees that the difference between the intermediate image estimate and ground truth (the "aliasing") at every iteration obeys a Gaussian distribution that can be fully characterised by a scalar. However, when Fourier coefficients of a signal with non-uniform spectral density are sampled, such as in MRI, AMP's scalar state evolution is no longer accurate and the algorithm encounters convergence problems.

This work presents a new AMP-based algorithm constructed specifically for variable density partial Fourier sensing matrices, which we term the Variable Density Approximate Message Passing (VDAMP) algorithm. We present empirical evidence that VDAMP obeys a "coloured state evolution", where the aliasing obeys a Gaussian distribution characterised by a vector. To our knowledge, VDAMP is the the first algorithm for MRI that exhibits a state evolution.

This thesis focuses on two practical advantages of state evolution for MRI re-

construction. Firstly, building on work on AMP in [13–15], Stein's Unbiased Risk Estimate (SURE) [16] is employed to tune model parameters to near optimality on-the-fly, relieving the need for manual parameter selection and leading to an algorithm that converges very quickly to a competitive reconstruction error. Secondly, the utility of state evolution in the context of deep Plug and Play (P&P) methods for MRI [17] is explored, where one step of the algorithm is replaced by a denoising neural network.

1.1 Thesis outline

This thesis has seven chapters. Chapter 2 reviews the mathematics of compressed sensing, placing emphasis on the AMP class of algorithms. Chapter 3 describes the MRI measurement model, motivates the application of compressed sensing, and reviews the literature on AMP for MRI. Chapter 4 onwards presents novel contributions. VDAMP for the single-coil MRI measurement model is presented in Chapter 4, and the extension to multiple coils, Parallel-VDAMP (P-VDAMP), is presented in Chapter 5. Chapter 6 presents the Denoising-P-VDAMP (D-P-VDAMP) algorithm, which incorporates a deep learning-based image model. Finally, Chapter 7 concludes and proposes a number of directions for future work.

1.2 Notation

In this thesis, boldface capital letters such as $\boldsymbol{A}$ are used for matrices and boldface lowercase letters such as $\boldsymbol{a}$ for vectors. For reference, a table of notations used in this thesis is shown in Table 1.1. Here, a_i denotes the ith element of a column vector $\boldsymbol{a} \in \mathbb{C}^N$ and A_{ij} denotes the entry of $\boldsymbol{A} \in \mathbb{C}^{M \times N}$ with ith row and jth column.

Notation	Definition	Description
$\boldsymbol{a}^T$ or $\boldsymbol{A}^T$	$[\boldsymbol{a}^T]_i = a_i$ or $[\boldsymbol{A}^T]_{ij} = A_{ji}$	Transpose of vector or matrix
$\boldsymbol{a}^H$ or $\boldsymbol{A}^H$	$[\boldsymbol{a}^H]_i = a_i^*$ or $[\boldsymbol{A}^H]_{ij} = A_{ji}^*$	Conjugate transpose of vector or matrix
$\|\boldsymbol{a}\|_p$	$\|\boldsymbol{a}\|_p = (\sum_i^N \|a_i\|^p)^{1/p}$	ℓ_p norm of a vector
$\mathrm{supp}(\boldsymbol{a})$	$\mathrm{supp}(\boldsymbol{a}) = \{i : a_i \neq 0\}$	Support of a vector
$\langle \boldsymbol{a} \rangle$	$\langle \boldsymbol{a} \rangle = \frac{1}{N} \sum_i^N a_i$	Empirical mean of a vector
$\boldsymbol{a} \odot \boldsymbol{b}$	$[\boldsymbol{a} \odot \boldsymbol{b}]_i = a_i \cdot b_i$	Entry-wise multiplication of vectors
$\boldsymbol{a} \oslash \boldsymbol{b}$	$[\boldsymbol{a} \oslash \boldsymbol{b}]_i = a_i / b_i$	Entry-wise division of vectors
$\mathrm{diag}(\boldsymbol{A})$	$[\mathrm{diag}(\boldsymbol{A})]_i = A_{ii}$	Vector formed from diagonal of matrix
$\mathrm{Diag}(\boldsymbol{a})$	$[\mathrm{Diag}(\boldsymbol{a})]_{ij} = a_i$ if $i = j$, 0 otherwise	Diagonal matrix formed from vector
$\mathbf{1}_N$	$[\mathbf{1}_N]_i = 1$	N-dimensional vector of ones
$\mathbf{0}_N$	$[\mathbf{0}_N]_i = 0$	N-dimensional vector of zeros
$\mathbb{1}_N$	$[\mathbb{1}_N]_{ij} = 1$ if $i = j$, 0 otherwise	$N \times N$ identity matrix
$\|\boldsymbol{a}\|$ or $\|\boldsymbol{A}\|$	$\|\boldsymbol{a}\|_i = \|a_i\|$ or $\|\boldsymbol{A}\|_{ij} = \|A_{ij}\|$	Entry-wise magnitude of a vector or matrix
$\boldsymbol{f}'(\boldsymbol{a})$	$\boldsymbol{f}'(\boldsymbol{a}) = \mathrm{diag}\left(\frac{\partial \boldsymbol{f}(\boldsymbol{a})}{\partial \boldsymbol{a}}\right)$	Diagonal of Jacobian of a function

Table 1.1: Table of notations used in this thesis.

Chapter 2

Compressed Sensing

This chapter reviews the mathematics of compressed sensing, placing an emphasis on the AMP algorithm [12]. The material covered in this chapter is preliminary to the MRI-specific compressed sensing review in Chapter 3.

2.1 Problem statement

Consider linear measurements $\boldsymbol{y} \in \mathbb{R}^n$ corrupted by additive white Gaussian noise,

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}_0 + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{n \times N}$, $\boldsymbol{x}_0 \in \mathbb{R}^N$, and $\boldsymbol{\varepsilon} \backsim \mathcal{N}(\mathbf{0}_n, \sigma_{\varepsilon}^2 \mathbb{1}_n)$, where $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^2)$ refers to the real Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}^2$. In the signal processing literature, the goal of recovering $\boldsymbol{x}_0$ from $\boldsymbol{y}$ is known as "solving a noisy linear inverse problem".

We are interested in the underdetermined case, where $n < N$. If $\boldsymbol{x}_0$ is known to be compressible, and the matrix $\boldsymbol{\Phi}$ satisfies certain conditions such as the Restriced Isometry Property (RIP) [18] discussed in Section 2.3, (2.1) is referred to as a compressed sensing problem [7, 8].

A vector $\boldsymbol{x}_0$ is defined to be compressible if there is a vector $\boldsymbol{w}_s \in \mathbb{R}^m$ in a known discrete representation $\boldsymbol{C} \in \mathbb{R}^{N \times m}$ such that $\|\boldsymbol{C}\boldsymbol{w}_s - \boldsymbol{x}_0\|_2 / \|\boldsymbol{x}_0\|_2$ is small, where $\boldsymbol{w}_s$ is "s-sparse". A vector is s-sparse if at most s of its coefficients are nonzero: $|\text{supp}(\boldsymbol{w}_s)| \leq s$, where $\text{supp}(\boldsymbol{w}_s) = \{j : w_{s,j} \neq 0\}$ is the support of $\boldsymbol{w}_s$. Then (2.1) can be written as

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{w}_s + \boldsymbol{\varepsilon}_s + \boldsymbol{\varepsilon} \tag{2.2}$$

where $\boldsymbol{\varepsilon}_s$ is the discrepancy due to using an s-sparse representation.

2.2 The (ℓ_0) problem

Given knowledge that $\boldsymbol{x}_0$ is compressible, it is natural to generate an estimate $\hat{\boldsymbol{x}}$ by seeking the sparsest solution in the compressed representation that is consistent with the data,

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \|\boldsymbol{w}\|_0 \text{ subject to } \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{w}\|_2 \leq \epsilon, \tag{2.3}$$

where $\epsilon \in \mathbb{R}^+$ is an estimate of the upper bound of the noise, and computing $\hat{\boldsymbol{x}} = \boldsymbol{C}\hat{\boldsymbol{w}}$. Eqn. (2.3) is known as the (ℓ_0) problem. To solve (ℓ_0) by combinatorially trying each $\binom{J}{N}$ solution with $\|\boldsymbol{w}\|_0 = J$ is NP-hard and computationally intractable for moderately sized problems.

A number of algorithms exist that are designed to approximate the solution of (2.3), including Orthogonal Matching Pursuit (OMP) [19–21], Iterative Hard Thresholding [22–24], Compressive Sampling Matching Pursuit (CoSaMP) [25], and Regularised OMP [26]. A popular approach in compressed sensing MRI and the focus of this thesis is on methods that employ convex relaxation.

2.3 Convex relaxation: the (ℓ_1) problem

One way to approach the (ℓ_0) problem is to replace the ℓ_0 norm by the convex ℓ_1 norm,

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \|\boldsymbol{w}\|_1 \text{ subject to } \|\boldsymbol{y} - \boldsymbol{\Phi C w}\|_2 \leq \epsilon, \tag{2.4}$$

which can be written in an unconstrained form as

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \frac{1}{2\sigma_\varepsilon^2}\|\boldsymbol{y} - \boldsymbol{\Phi C w}\|_2^2 + \lambda\|\boldsymbol{w}\|_1, \tag{2.5}$$

where λ is a sparse weighting. Equations (2.4) and (2.5) are referred to as the (ℓ_1) problem or "basis pursuit" [27].

A crucial theoretical question for compressed sensing is: under what conditions do the solutions of the (ℓ_0) and (ℓ_1) problems coincide? A common tool used to address this is the Restricted Isometry Property (RIP) [18]. A matrix $\boldsymbol{A}$ is said to satisfy the s-restricted isometry property if a constant $0 \leq \delta_s < 1$ exists such that

$$(1 - \delta_s)\|\boldsymbol{x}_s\|_2^2 \leq \|\boldsymbol{A x}_s\|_2^2 \leq (1 + \delta_s)\|\boldsymbol{x}_s\|_2^2$$

for all s-sparse $\boldsymbol{x}_s \in \mathbb{R}^N$. When the RIP holds, any $s \times s$ submatrix of $\boldsymbol{A}$ changes the ℓ_2 norm of a vector by only a small amount, so is nearly an isometry.

In the absence of noise, so that $\boldsymbol{\varepsilon}_s = \boldsymbol{0}_n$ and $\boldsymbol{\varepsilon} = \boldsymbol{0}_n$, the (ℓ_1) solution is guaranteed to be equivalent to the (ℓ_0) solution when $\boldsymbol{\Phi C}$ satisfies the RIP with $\delta_{2s} < \sqrt{2} - 1$ for s-sparse $\boldsymbol{w}_s$ [28]. In the presence of noise, the (ℓ_1) solution obeys

$$\|\hat{\boldsymbol{w}} - \boldsymbol{w}_0\|_2 \leq c_0 s^{-1/2}\|\boldsymbol{w}_0 - \boldsymbol{w}_s\|_1 + c_1\epsilon,$$

where $\boldsymbol{C w}_0 = \boldsymbol{x}_0$ and c_0 and c_1 are known, small constants [28]. In other words, when the RIP holds and a sufficiently sparse representation exists, basis pursuit recovers

$\boldsymbol{x}_0$ exactly in the noiseless case and approximately in the noisy case.

Unfortunately, to verify whether the RIP holds for general $\boldsymbol{A}$ is in itself an NP-hard problem [29]. However, for many classes of large *random* matrices, such as Gaussian, Bernoulli and partial Fourier, the RIP is known to hold with high probability [30–33]. It is not usual to compute the RIP explicitly; rather, as discussed further in Sections 3.3 and 3.4, the role of the RIP is to motivate sampling design choices.

2.3.1 Solving the (ℓ_1) problem

The noiseless (ℓ_1) problem, (2.4) with $\epsilon = 0$, can be solved with Linear Programming (LP) methods [18,34]. In [35–37], analytical predictions of the sparsity-undersampling regions of success and failure for LP methods were derived using combinatorial geometry. In the presence of noise, where $\epsilon > 0$, the ℓ_2 Quadratic Programming methods must be employed. The focus of this thesis is on proximal solution methods, which have low per-iteration computational cost and memory requirements, and are widely used in MRI. A number of such methods for the noisy (ℓ_1) problem exist, including but not limited to the Iterative Shrinkage-Thresholding Algorithm (ISTA) [38], Alternating Direction Method of Multipliers (ADMM) [39] and split Bregman method [40].

2.3.2 The ISTA algorithm

Due to its relevance to the AMP algorithm [12], which is central to this thesis, we focus on the ISTA algorithm here, stated in Algorithm 1. ISTA alternates between gradient descent in lines 3-4 and the action of the proximal operator of the ℓ_1 norm $\boldsymbol{\eta}(\boldsymbol{v}; t)$ in line 5, usually referred to as "soft thresholding", which acts component-wise as

$$\eta_j(v_j; t) = v_j \left(1 - \min\left\{\frac{t}{|v_j|}, 1\right\}\right), \tag{2.6}$$

Algorithm 1 Iterative Shrinkage-Thresholding Algorithm (ISTA) [38]

Require: Measurements $\boldsymbol{y} \in \mathbb{R}^n$, measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{n \times N}$, sparse representation $\boldsymbol{C} \in \mathbb{R}^{N \times m}$, sparse weighting $\lambda \in \mathbb{R}^+$, and number of iterations K_{it}.

1: Set $\hat{\boldsymbol{x}}_0 = \boldsymbol{0}_N$.
2: **for** $k = 0, 1, \ldots, K_{it} - 1$ **do**
3: $\quad \boldsymbol{z}_k = \boldsymbol{y} - \boldsymbol{\Phi}\hat{\boldsymbol{x}}_k$
4: $\quad \boldsymbol{r}_k = \hat{\boldsymbol{x}}_k + \boldsymbol{\Phi}^T \boldsymbol{z}_k$
5: $\quad \hat{\boldsymbol{x}}_{k+1} = \boldsymbol{C}\boldsymbol{\eta}(\boldsymbol{C}^T \boldsymbol{r}_k; \lambda)$
6: **end for**
7: **return** $\hat{\boldsymbol{x}}_{k+1}$

where $t \in \mathbb{R}$ is the threshold. For entries of $\boldsymbol{v}$ that have magnitude less that t, the soft thresholding operator sets $\eta(v_j; t) = 0$, while entries with a magntitude greater than t are scaled by $1 - t/|v_j|$.

The per-iteration computational cost of ISTA is dominated by the multiplication of $\boldsymbol{\Phi}$ and $\boldsymbol{C}$ and their transpose on a vector, which have complexity $O(nN)$ and $O(Nm)$ respectively in general. However, in many applications these multiplications have fast implementations, such as the $O(N \log N)$ complexity Fast Fourier Transform (FFT) or $O(N)$ complexity Discrete Wavelet Transform (DWT), yielding a far lower per-iteration cost.

2.3.3 The iterative denoising heuristic

This section describes the iterative denoising heuristic for ISTA, which is used to motivate the AMP algorithm introduced in Section 2.4. The discussion presented here is based on Section 1D of [12].

At the first iteration, line 4 of ISTA, Algorithm 1, gives $\boldsymbol{r}_0 = \boldsymbol{\Phi}^T \boldsymbol{y}$, which can be

written in terms of the unknown $\boldsymbol{x}_0$ as

$$\begin{aligned}\boldsymbol{r}_0 &= \boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{x}_0 + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{x}_0 + (\boldsymbol{\Phi}^T\boldsymbol{\Phi} - \mathbb{1}_N)\boldsymbol{x}_0 + \boldsymbol{\Phi}^T\boldsymbol{\varepsilon} \\ &= \boldsymbol{x}_0 + \boldsymbol{e}_0.\end{aligned}$$

where $\boldsymbol{e}_0 = (\boldsymbol{\Phi}^T\boldsymbol{\Phi} - \mathbb{1}_N)\boldsymbol{x}_0 + \boldsymbol{\Phi}^T\boldsymbol{\varepsilon}$. Therefore $\boldsymbol{r}_0$ can be thought of as the signal of interest $\boldsymbol{x}_0$ corrupted by a vector $\boldsymbol{e}_0$, where $\boldsymbol{e}_0$ has a contribution from the undersampling, $(\boldsymbol{\Phi}^T\boldsymbol{\Phi} - \mathbb{1}_N)\boldsymbol{x}_0$, and from the measurement noise $\boldsymbol{\Phi}^T\boldsymbol{\varepsilon}$.

Writing $\boldsymbol{r}_0$ in this way leads to an intuition for how ISTA works when $\boldsymbol{\Phi}$ is a random matrix. Consider the specific case where $\boldsymbol{\Phi}$ is a large matrix with zero-mean i.i.d. Gaussian entries with a variance that normalises the columns in expectation, so that $\mathbb{E}\{\boldsymbol{\Phi}^T\boldsymbol{\Phi}\} = \mathbb{1}_N$. In this case, the entries of $\boldsymbol{e}_0$ can accurately be modelled as drawn from an i.i.d. zero-mean Gaussian distribution, so that $\boldsymbol{r}_0 = \boldsymbol{x}_0 + \mathcal{N}(\boldsymbol{0}_N, \sigma_0^2\mathbb{1}_N)$ for some variance σ_0^2. In other words, when $\boldsymbol{\Phi}$ is an appropriately normalised i.d.d. Gaussian sampling matrix, ISTA's $\boldsymbol{r}_0$ behaves as the ground truth corrupted by i.i.d. Gaussian "noise". Line 5 of ISTA applies soft thresholding to the noisy vector $\boldsymbol{r}_0$, which is known to reduce the error when $\boldsymbol{x}_0$ is sufficiently sparse [41]. We therefore expect the error of the signal estimate to decreases after the first iteration.

It is instructive to question whether this denoising heuristic is accurate beyond $k = 0$. For general k the error vector is $\boldsymbol{e}_k = (\boldsymbol{\Phi}^T\boldsymbol{\Phi} - \mathbb{1}_N)(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_k) + \boldsymbol{\Phi}^T\boldsymbol{\varepsilon}$. Since $\boldsymbol{\Phi}^T\boldsymbol{\Phi} - \mathbb{1}_N$ correlates with $\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_k$ in general, the error vector $\boldsymbol{e}_k$ is no longer accurately modelled as a zero-mean i.d.d. Gaussian distribution, $\boldsymbol{r}_k \neq \boldsymbol{x}_0 + \mathcal{N}(\boldsymbol{0}_N, \sigma_k^2\mathbb{1}_N)$ for $k > 0$, so the iterative denoising heuristic breaks down.

This breakdown of the Gaussian denoising heuristic motivates the AMP algorithm [12]. AMP modifies ISTA in a way that ensures that the error vector $\boldsymbol{e}_k$ is guaranteed to be a zero-mean Gaussian for all iterations k, which, as described in the following,

Algorithm 2 Approximate Message Passing (AMP) [12]

Require: Measurements $\boldsymbol{y} \in \mathbb{R}^n$, measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{n \times N}$, denoiser $\boldsymbol{g}(\boldsymbol{r}; \tau)$ and number of iterations K_{it}.

1: Set $\boldsymbol{z}_{-1} = \boldsymbol{0}_n$ and $\hat{\boldsymbol{x}}_k = \boldsymbol{0}_N$.
2: **for** $k = 0, 1, \ldots, K_{it} - 1$ **do**
3: $\quad \boldsymbol{z}_k = \boldsymbol{y} - \boldsymbol{\Phi}\hat{\boldsymbol{x}}_k + \frac{N}{n}\boldsymbol{z}_{k-1} \left\langle \boldsymbol{g}'(\boldsymbol{r}_{k-1}; \tau_{k-1}) \right\rangle$
4: $\quad \boldsymbol{r}_k = \hat{\boldsymbol{x}}_k + \boldsymbol{\Phi}^{\boldsymbol{T}} \boldsymbol{z}_k$
5: $\quad$ Update τ_k e.g. $\tau_k = \|\boldsymbol{z}_k\|_2^2 / n$
6: $\quad \hat{\boldsymbol{x}}_{k+1} = \boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$
7: **end for**
8: **return** $\hat{\boldsymbol{x}}_k$

has a number of theoretical and practical advantages.

2.4 Approximate Message Passing (AMP)

The AMP algorithm [12], shown in Algorithm 2, is an iterative method that for certain sensing matrices and signal models obeys

$$\boldsymbol{r}_k = \boldsymbol{x}_0 + \mathcal{N}(\boldsymbol{0}_N, \sigma_k^2 \mathbb{1}_N), \tag{2.7}$$

for an iteration-dependent standard deviation σ_k. In other words, the denoising heuristic that is only valid at $k = 0$ for ISTA is accurate at all k for AMP. Eqn. (2.7) is referred to as AMP's "state evolution" because the "state" of the algorithm at iteration k can be fully characterised by the scalar σ_k^2.

AMP is stated in Algorithm 2. Lines 3-4 of AMP are identical to lines 3-4 of ISTA, Algorithm 1, except for the addition of a new term in line 3, $\frac{N}{n}\boldsymbol{z}_{k-1} \left\langle \boldsymbol{g}'(\boldsymbol{r}_{k-1}; \tau_{k-1}) \right\rangle$. This crucial term is referred to as AMP's "Onsager correction", and is responsible for removing the correlation that causes ISTA's denoising heuristic to break down for iterations $k > 0$. Here, $\boldsymbol{g}'(\boldsymbol{r}_k; \tau_k)$ is the diagonal of the Jacobian of $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$,

$$\boldsymbol{g}'(\boldsymbol{r}_k; \tau_k) = \operatorname{diag}\left(\frac{\partial \boldsymbol{g}(\boldsymbol{r}_k; \tau_k)}{\partial \boldsymbol{r}_k} \right),$$

and $\langle \boldsymbol{u} \rangle = \frac{1}{N} \sum_j^N u_j$ for $\boldsymbol{u} \in \mathbb{R}^N$ is the empirical averaging operator.

Line 5 of AMP estimates σ_k^2 from (2.7). It has been suggested that $\tau_k = \|\boldsymbol{z}_k\|_2^2/n$ is used, although other possibilities exist [42]. AMP's σ_k^2 estimate is not necessary for state evolution to occur; its role is as a parameter for the function $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$ in line 6. For instance, τ_k can be used to automatically tune parameters of $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$: see Section 2.6.1. Since, by (2.7), $\boldsymbol{r}_k$ behaves as a noisy version of the ground truth, $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$ is often referred to as a "denoiser" [44]. The particular choice of $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$ depends on the model of $\boldsymbol{x}_0$. One possibility for $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$ is the proximal operator associated with a penalty function $f(\boldsymbol{x})$, which is defined as

$$\boldsymbol{g}(\boldsymbol{r}_k; \tau_k) = \underset{\boldsymbol{x} \in \mathbb{C}^N}{\operatorname{argmin}} \frac{1}{2} \|\boldsymbol{r}_k - \boldsymbol{x}\|_2^2 + f(\boldsymbol{x}), \tag{2.8}$$

although, as discussed in Section 2.6.2, is not necessarily constrained to functions of this form. AMP that denoises with (2.8) has fixed point

$$\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + f(\boldsymbol{x}). \tag{2.9}$$

For instance, AMP was originally proposed with soft thresholding as the denoiser, $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k) = \boldsymbol{C}\boldsymbol{\eta}(\boldsymbol{C}^T \boldsymbol{r}_k; \lambda)$, to solve the (ℓ_1) problem (2.5).

Given $\boldsymbol{x}_0$, it is possible to recursively predict the σ_k from AMP's state evolution (2.7):

$$\sigma_{k+1}^2 = \sigma_\varepsilon^2 + \frac{1}{n} \mathbb{E}[\|\boldsymbol{g}(\boldsymbol{x}_0 + \mathcal{N}(\boldsymbol{0}_N; \sigma_k^2 \mathbb{1}_N); \tau_k) - \boldsymbol{x}_0\|_2^2]. \tag{2.10}$$

This is a valuable tool for analysis as it can be used to predict AMP's performance without actually running the algorithm. For example, for the noiseless (ℓ_1) problem, (2.4) with $\epsilon = 0$, it was shown in [12] using (2.10) that for a particular τ_k-dependent soft threshold, AMP coverges to the sparsest solutions in exactly the same cases as

predicted for LP methods from combinatorial geometry. To see how this works, one could imagine approximating σ_k^2 in the limit $k \to \infty$ by computing (2.10) for a large number of iterations and observing when $\sigma_k^2 \to 0$, which indicates exact reconstruction. Eqn. (2.10) has other analytic uses, such as to analyse AMP's sensitivity to measurement noise [45].

AMP was originally constructed for real, zero-mean, i.i.d. Gaussian measurements using a high-dimensional approximation of the "sum-product belief propagation equations" on a dense bipartite factor graph [46,47]. The sum-product belief propagation equations involve defining "messages" that are passed along the edges of the factor graph, hence the terminology "message passing". A discussion of its derivation would not inform any aspects of the novel work presented in this thesis, so is not included here; for a detailed discussion, see [48].

In this thesis we use the term "aliasing", which is common in the MRI literature, to refer to the difference between a given estimate and the ground truth. For instance, the aliasing of $\boldsymbol{r}_k$ is $\boldsymbol{r}_k - \boldsymbol{x}_0$. Also, when the covariance matrix is proportional to the identity, as in (2.7), the aliasing and state evolution are referred to as "white". AMP's white state evolution was proven for i.i.d. Gaussian sensing matrices in [49] and subsequently proven for i.i.d. sub-Gaussian measurements in [50]. It has also been shown empirically that white state evolution holds for uniformly undersampled Fourier measurements of an artificial i.i.d. signal [12]. However, for generic $\boldsymbol{\Phi}$, the behaviour of AMP is not well understood and it has been noted by a number of authors [14, 51–53] that it can encounter convergence problems, including for the type of sensing matrices relevent in MRI: see Section 3.7.

2.5 OAMP and VAMP

The following section introduces the recent Orthogonal AMP (OAMP) [54] and related Vector Approximate Message Passing (VAMP) [55], alternatives to AMP which obey a white state evolution for a broader class of measurement matrices $\boldsymbol{\Phi}$.

2.5.1 Description of OAMP and VAMP

Algorithm 3 states the OAMP [54] and VAMP [55] algorithms. Various choices of the $N \times n$ matrix $\boldsymbol{W}$ used in line 5 of Algorithm 3 are possible. For instance, "Matched-Filter" OAMP uses $\boldsymbol{W} = \boldsymbol{\Phi}^T$, while "Linear Minimum Mean-Squared Error" (LMMSE) OAMP uses

$$\boldsymbol{W} = \frac{N}{\mathrm{Tr}[\boldsymbol{\Phi}^T(u\mathbb{1}_n + v\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1}\boldsymbol{\Phi}]}\boldsymbol{\Phi}^T(u\mathbb{1}_n + v\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1} \tag{2.11}$$

for $u, v \in \mathbb{R}^+$. Lines 3-4 of Algorithm 3 with $\boldsymbol{W}$ set to (2.11) are equivalent to a Bayesian MMSE estimate where both the likelihood and prior are Gaussian [55].

VAMP [55] is a version of LMMSE-OAMP where $u = \sigma_\varepsilon^2$ and v takes the iteration-dependent value $v = \widetilde{\tau}_k$, where $\widetilde{\tau}_k$ is an estimate of $\|\widetilde{\boldsymbol{r}}_k - \boldsymbol{x}_0\|_2^2/N$. VAMP was proposed independently of OAMP, and was originally presented in a different form, but has since been shown to be equivalent to LMMSE-OAMP: see [56]. Although LMMSE OAMP and VAMP require the additional computation of a matrix inverse, the per-iteration complexity can be reduced to that of AMP after an initial Singular Value Decomposition (SVD) of $\boldsymbol{\Phi}$: see Algorithm 2 of [55]. It has also been suggested to use the Conjugate Gradient method to approximate $\boldsymbol{W}\boldsymbol{z}_k$ [57], which does not require the computation of an SVD or an explicit matrix inverse.

The c_k update in line 8 of Algorithm 3 is $c_k = 1/(1 - \alpha_k)$ for VAMP, while for OAMP it is recognised that state evolution holds or fails irrespective of the choice of c_k. For instance, [54] uses both $c_k = 1/(1 - \alpha_k)$ and three arbitrary chosen values

Algorithm 3 OAMP [54] and VAMP [55]

Require: Measurements $\boldsymbol{y} \in \mathbb{R}^n$, measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{n \times N}$, denoiser $\boldsymbol{g}(\boldsymbol{r}; \tau)$, choice of $\boldsymbol{W} \in \mathbb{R}^{N \times n}$ e.g. $\boldsymbol{W} = \boldsymbol{\Phi}^T$ and number of iterations K_{it}.

Set $\widetilde{\boldsymbol{r}}_0 = \boldsymbol{0}_N$
for $k = 0, 1, \ldots, K_{it} - 1$ **do**
  $\boldsymbol{z}_k = \boldsymbol{y} - \boldsymbol{\Phi}\widetilde{\boldsymbol{r}}_k$
  $\boldsymbol{r}_k = \widetilde{\boldsymbol{r}}_k + \boldsymbol{W}\boldsymbol{z}_k$
  Update τ_k, e.g. $\tau_k = \|\boldsymbol{z}_k\|_2^2 / n$
  $\hat{\boldsymbol{x}}_k = \boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$
  $\alpha_k = \langle \boldsymbol{g}'(\boldsymbol{r}_k; \tau_k) \rangle$
  Update c_k, e.g. $c_k = 1/(1 - \alpha_k)$
  $\widetilde{\boldsymbol{r}}_{k+1} = c_k \cdot (\hat{\boldsymbol{x}}_k - \alpha_k \boldsymbol{r}_k)$
end for
return $\hat{\boldsymbol{x}}_k$

$c_k = 1, 2, 3$, while in [58], Stein's Unbiased Risk Estimate (SURE) [16] was used to select a c_k update that approximately minimises the MSE of $\widetilde{\boldsymbol{r}}_k$. SURE is described in detail in Section 2.6.1.

2.5.2 Sensing matrix conditions for state evolution

Let the SVD of the sensing matrix be $\boldsymbol{\Phi} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$. It was proven in [55] that when $\boldsymbol{U}$ is an arbitrary real left-singular matrix, $\boldsymbol{S}$ is a real but otherwise arbitrary diagonal matrix of singular values, and $\boldsymbol{V}$ is a random matrix drawn from the group of orthogonal matrices, VAMP obeys a white state evolution. The term "right-orthogonally invariant" is used to decribe such a $\boldsymbol{\Phi}$, as the right-multiplication of an orthogonal matrix does not change its distribution. Shortly after the publication of [55], a similar result for the complex case, right-*unitarily* invariant sensing matrices, was proven in [56]. The class of right-unitarily invariant matrices is considerably broader than AMP's i.i.d. sub-Gaussian condition, which constrains $\boldsymbol{U}$ and $\boldsymbol{V}$ to random orthogonal matrices and a particular distribution on the singular values $\boldsymbol{S}$ [55].

2.5.3 Intuition for the form of OAMP

The relationship between Matched-Filter OAMP, Algorithm 3 with $\boldsymbol{W} = \boldsymbol{\Phi}^T$, and ISTA, Algorithm 1, can be seen by considering lines 6-9 of OAMP as a single function:

$$\tilde{\boldsymbol{g}}(\boldsymbol{r}_k; \tau_k) = c_k \cdot (\boldsymbol{g}(\boldsymbol{r}_k; \tau_k) - \alpha_k \boldsymbol{r}_k). \tag{2.12}$$

Then OAMP is equivalent to ISTA with $\tilde{\boldsymbol{g}}(\boldsymbol{r}_k; \tau_k)$ in place of the usual sparse-domain soft thresholding. Now recall that the difference between ISTA and AMP, Algorithm 2, is the Onsager correction, the final term in line 3 of AMP. An instructive question is the following: if OAMP were to have an AMP-style Onsager correction, $\frac{N}{n}\boldsymbol{z}_k \langle \widetilde{\boldsymbol{g}}'(\boldsymbol{r}_k; \tau_k)\rangle$, what would it be? Differentiating and averaging (2.12) and using $\alpha_k = \langle \boldsymbol{g}'(\boldsymbol{r}_k; \tau_k)\rangle$,

$$\begin{aligned}\langle \tilde{\boldsymbol{g}}'(\boldsymbol{r}_k; \tau_k)\rangle &= c_k \cdot \left\langle \boldsymbol{g}'(\boldsymbol{r}_k; \tau_k) - \langle \boldsymbol{g}'(\boldsymbol{r}_k; \tau_k)\rangle - \frac{1}{N}\sum_{j=1}^{N} \frac{\partial \boldsymbol{g}'(\boldsymbol{r}_k; \tau_k)}{\partial r_j} r_j \right\rangle \\ &= -\frac{c_k}{N} \boldsymbol{g}''(\boldsymbol{r}_k; \tau_k)^T \boldsymbol{r}_k.\end{aligned}$$

For the particular case where $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$ is the soft thresholding function, $\boldsymbol{g}''(\boldsymbol{r}_k; \tau_k) = \boldsymbol{0}_N$, and

$$\langle \tilde{\boldsymbol{g}}'(\boldsymbol{r}_k; \tau_k)\rangle = 0. \tag{2.13}$$

AMP's Onsager correction would therefore be zero. OAMP can therefore be understood as a version of AMP where the denoiser has been modified in such a way that the usual measurement-domain Onsager correction is zero. In other words, the AMP-style Onsager correction in the domain of $\boldsymbol{y}$, implemented in line 3 of Algorithm 2, is relocated in OAMP to the domain of $\boldsymbol{x}_0$, implemented in lines 7-9 of Algorithm 3.

The term "divergence-free" is used to describe a function that satisfies (2.13) [54]. Any divergence-free function can be employed in place of lines 6-9 of Algorithm 3; it

is not required to take the form of (2.12). For functions of the form of (2.12), any c_k is consistent with the divergence-free condition (2.13), so, as stated in Section 2.5.1, state evolution holds or fails irrespective of the c_k update in line 8 of Algorithm 3.

2.6 Leveraging state evolution

Section 2.4 described how state evolution can be used to choose thresholds for the noiseless (ℓ_1) problem. This section focuses on other practical advantages of state evolution in the $\sigma_\varepsilon \geq 0$ case, which play a central role in the novel contributions of this thesis.

2.6.1 AMP with Stein's Unbiased Risk Estimate

Selecting appropriate regularisation parameters is a notable challenge in real-world compressed sensing applications. This section decribes how AMP, OAMP or VAMP's state evolution can be leveraged to efficiently tune model parameters on-the-fly using Stein's Unbiased Risk Estimate (SURE) [13–16].

Consider a real vector $\boldsymbol{v}_0 \in \mathbb{R}^{N_v}$ corrupted by white Gaussian noise: $\boldsymbol{v} = \boldsymbol{v}_0 + \mathcal{N}(\mathbf{0}_{N_v}, \tau_v \mathbb{1}_{N_v})$. Let $\boldsymbol{d}(\boldsymbol{v}; \tau_k) = \boldsymbol{v} + \boldsymbol{h}(\boldsymbol{v}; \tau_k)$ be a weakly differentiable estimator of $\boldsymbol{v}_0$. SURE [16] is defined as

$$SURE(\boldsymbol{d}(\boldsymbol{v}; \tau_k)) = \|\boldsymbol{h}(\boldsymbol{v}; \tau_v)\|_2^2 + N_v \tau_v [2 \left\langle \boldsymbol{d}'(\boldsymbol{v}; \tau_k) \right\rangle - 1]. \tag{2.14}$$

SURE is of interest to denoising problems because it is an unbiased estimate of the risk [16],

$$\mathbb{E}\{SURE(\boldsymbol{d}(\boldsymbol{v}; \tau_k))\} = \mathbb{E}\{\|\boldsymbol{d}(\boldsymbol{v}; \tau_k) - \boldsymbol{v}_0\|_2^2\},$$

where $\mathbb{E}\{\cdot\}$ denotes the expectation over the random noise. Therefore, for large N_v, (2.14) can be used as a proxy for the true risk without requiring the ground truth.

For algorithms that obey state evolution, it is possible to tune parameters of $\boldsymbol{g}(\boldsymbol{r}_k;\tau_k)$ so that SURE is minimised, and $\boldsymbol{g}(\boldsymbol{r}_k;\tau_k)$ is approximately MSE-optimal [13, 14]. Concretely, a denoiser $\boldsymbol{g}(\boldsymbol{r}_k;\tau_k,\boldsymbol{\Theta})$, where $\boldsymbol{\Theta}\in\mathbb{R}^{N_p}$ is a parameter vector, is SURE-optimal and therefore approximately MSE-optimal when

$$\boldsymbol{\Theta}^* = \arg\min_{\boldsymbol{\Theta}} SURE(\boldsymbol{g}(\boldsymbol{r}_k;\tau_k,\boldsymbol{\Theta})) \approx \arg\min_{\boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{r}_k;\tau_k,\boldsymbol{\Theta}) - \boldsymbol{x}_0\|_2^2$$

is employed.

State evolution in conjunction with SURE yields an algorithm where all denoising parameters are automatically tuned on-the-fly, with approximately MSE-optimal denoising employed at every iteration. This is discussed in further detail in the context of the proposed novel algorithm in Section 4.2.3.

2.6.2 Denoising-AMP

Images have a rich structure beyond purely sparsity, and there has been considerable work on developing more sophisticated image models for reconstruction, such as block-sparsity [59], wavelet-tree sparsity [60] and non-local self-similarity [61, 62]. This section describes how the AMP framework offers a natural bridge between compressed sensing and the extensive literature on image denoising, where state-of-the-art image models have been developed.

Methods that replace one step of an iterative reconstruction method with a generic denoiser are referred to as Plug and Play (P&P) algorithms [17, 63–66]. Throughout this thesis, following [44], the prefix "Denoising" is used to refer to any such algorithm. For instance, AMP with a denoiser in place of soft thresholding, as proposed in [44], is referred to as Denoising-AMP (D-AMP).

The D-AMP, Denoising-OAMP (D-OAMP) [58] and Denoising-VAMP (D-VAMP) [67] algorithms exploit the observation that the goal of estimating $\boldsymbol{x}_0$ from a vector

$\boldsymbol{r}_k$ that obeys state evolution, (2.7), is exactly the white Gaussian image denoising problem [43]. These algorithms propose drawing from one of the wealth of sophisticated image denoisers already available, such as Block-Matching 3D (BM3D) [68,69], and using them as the denoiser $\boldsymbol{g}(\boldsymbol{r}_k;\tau_k)$. AMP with BM3D denoising was found to offer state-of-the-art reconstruction accuracy for i.i.d. Gaussian sampled images at the time [44].

Given sufficient representative training data, hand-designed denoisers such as BM3D can be outperformed by convolutional neural networks (CNNs) [2, 70, 71]. D-AMP with a denoising neural network based on the architecture from [2] was proposed in [72], which was found to outperform BM3D-AMP both in terms of reconstruction error and compute time. Deep learning is discussed in more detail in the context of MRI reconstruction in Section 3.6.

Since the denoiser in P&P methods is not necessarily a proximal operator, the usual convergence guarantees [39] may not apply, and their fixed points may not coincide with the minimum of a known cost function of the form stated in (2.9). In general it is unclear what such methods will converge to, if at all, and how this depends on the choice of denoiser. Some theoretical progress has been made on the convergence of P&P methods. For instance, in [64], it was shown that if the denoiser $\boldsymbol{g}(\boldsymbol{x})$ is non-expansive, $\|\boldsymbol{g}(\boldsymbol{x}_1)-\boldsymbol{g}(\boldsymbol{x}_2)\| \leq \|\boldsymbol{x}_1-\boldsymbol{x}_2\|$ for all $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and is a subgradient of some convex function, D-ADMM [63] will converge. It was shown in [73] that although these conditions are often not met in practice, a careful update of the algorithm parameters can ensure convergence for generic denoiser.

The D-AMP framework offers an instructive perspective on the value of state evolution: AMP essentially recasts compressed sensing reconstruction as a series of Gaussian denoising problems, where the Gaussian "effective noise" at every iteration has zero mean and approximately known variance. In other words, roughly speaking, how well one can reconstruct a signal depends on how well one can denoise a Gaussian-

corrupted version of it.

2.7 Summary

Compressed sensing provides a framework for reconstructing compressible signals from underdetermined measurements. The focus of this thesis is on the AMP algorithm, which efficiently reconstructs signals that have been sampled with large i.i.d. sub-Gaussian sensing matrices. AMP obeys a white state evolution, (2.7) and (2.10), which guarantees that the aliasing of the intermediate estimate $\boldsymbol{r}_k$ obeys a zero-mean Gaussian distribution that can be fully characterized by a scalar standard deviation σ_k.

The key practical benefit of state evolution is that it can be used to guide the denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$. Using SURE, denoising parameters can be automatically tuned to near-optimality on-the-fly. Furthermore, state evolution offers a bridge between compressed sensing and the image denoising literature, allowing for richer signal modelling and a natural way to incorporate the statistical modelling capabilities of neural networks.

A widely acknowledged limitation of AMP is that when the sensing matrix deviates from its i.i.d. sub-Gaussian assumption, state evolution often fails and the algorithm does not perform well in practice. OAMP and VAMP are recent alternatives to AMP that preserve state evolution for a wider class of sensing matrices. However, as discussed in the following chapter, neither these algorithms nor prior attempts to modify them [74–77] successfully preserve state evolution for the type of sensing matrices seen in MRI. This motivates the primary goal of this thesis: to develop an algorithm for compressed sensing MRI that obeys state evolution.

Chapter 3

Compressed Sensing for Accelerated MRI

This chapter describes the data acquisition process in MRI, motivates the application of compressed sensing, and discusses prior work on AMP for MRI.

3.1 Data acquisition in MRI

The signal in MRI acquisitions is generated by the magnetisation of atomic nuclei, primarily protons from hydrogen atoms. Microscopically, nuclear magnetisation is best described by quantum mechanics [78]. However, for systems with many nuclei, most phoneomena can adequately be described with classical mechanics [79]. This thesis considers the classical description only.

A strong, static magnetic field B_0 causes the protons to have an net magnetic moment that tends to align with B_0. The protons precess at the Larmor frequency,

$$\omega_0 = \gamma B_0,$$

where $\gamma = 42.6$ MHzT^{-1} is the gyromagnetic ratio of a proton [80]. Precession causes

a changing magnetic flux which is detected by one or more receiver coils that are tuned to the Larmor frequency. To generate magnetisation in the direction orthogonal to B_0, known as "transverse magnetisation", radio waves tuned to ω_0 resonate with and excite the protons, tipping the magnetisation's direction. Over time, the magnetisation relaxes, returning in line with B_0. The evolution of the magnetisation over time depends on magnetic properties that are tissue-specific. This is the origin of contrast in MRI.

The magnetisation's evolution over time is mathematically governed by the Bloch equations [81], which depend on two decay constants: T_1, which models the time for the magnetisation to realign with B_0, and T_2, which models the transverse relaxation time. The transverse relaxation T_2 depends on the decoherence of spins, and is shorter than T_1 in general. Crucially for the clinical value of MRI, T_1 and T_2 are sensitive to a number of pathologies, as described in, for instance, [82–87]. The duration and timing of the radiofrequency pulses can be engineered so that the contrast is sensitive to T_1 or T_2, depending on the application.

The spatial origin of protons are encoded with additional magnetic fields that vary linearly in space, referred to as "gradient fields" [88], which manipulate the frequency of precession. Let the transverse magnetisation be denoted with $X_0(\boldsymbol{u})$, where $\boldsymbol{u}$ is a position vector. The precessing magnetisation has both magnitude and phase, so is represented as a complex number. The gradient field $\boldsymbol{q}(t)$ causes the Larmor frequency to vary in space and time,

$$\omega(\boldsymbol{u}, t) = \gamma(B_0 + \boldsymbol{q}(t)^T \boldsymbol{u}),$$

which implies that the phase at position $\boldsymbol{u}$, time t is [11]

$$\phi(\boldsymbol{u}, t) = 2\pi\gamma \int_0^t \boldsymbol{q}(t')^T \boldsymbol{u} dt',$$

where $t = 0$ is the time at excitation. The received signal is formed by an integration over the volume [89],

$$Y(t) = \int X_0(\boldsymbol{u}) \exp(-2\pi\gamma i \boldsymbol{k}(t)^T \boldsymbol{u}) d\boldsymbol{u} + \varepsilon(t), \tag{3.1}$$

where

$$\boldsymbol{k}(t) = \int_0^t \boldsymbol{q}(t')dt'$$

and $\varepsilon(t)$ is the measurement noise, which is primarily due to thermal effects in the subject and reciever coils [90]. The choice of gradient fields $\boldsymbol{q}(t)$ and the time-dependent resonanting radiofrequency pulses are collectively referred to as a "pulse sequence" [91].

The signal $Y(t)$ is a sample of the Fourier transform of $X_0(\boldsymbol{u})$ along a readout curve defined by the gradient field $\boldsymbol{q}(t)$. The data sampling process in MRI consists of traversing a series of 1D readout curves in the Fourier domain. The Fourier-domain representation of the image is known in the MRI literature as "k-space", and the 1D readout curves are often referred to as "sampling trajectories". Common choices for k-space sampling trajectories are discussed in Section 3.4. The speed at which it is possible to traverse k-space is limited by the scanner's gradient specifications, in particular the maximum gradient amplitude and maximum slew-rate. Additionally, excessively high amplitude gradients and gradient switching are limited for the patient's safety to prevent peripheral nerve stimulation [6].

The signal is acquired by an array of reciever coils that are placed around the region of anatomy of interest. The coils are more sensitive to regions that are physically closest to it, and its sensitivity gradually decays with increasing distance. This can be modelled by weighting (3.1) by the complex, spatially-dependent sensitivity

of the cth coil, $S_c(\boldsymbol{u})$:

$$Y_c(t) = \int S_c(\boldsymbol{u}) X_0(\boldsymbol{u}) \exp(-2\pi\gamma i \boldsymbol{k}(t)^T \boldsymbol{u}) d\boldsymbol{u} + \varepsilon_c(t), \qquad c = \{1, 2, \ldots, N_c\} \quad (3.2)$$

which corresponds to a sample of the Fourier transform of the coil-weighted image $S_c(\boldsymbol{u})X_0(\boldsymbol{u})$. Reciever coils are discussed further in Section 3.2.2.

3.2 Discrete measurement models

Discretisation in MRI consists of representing the transverse magnetisation $X_0(\boldsymbol{u})$ with a vector $\boldsymbol{x}_0 \in \mathbb{C}^N$, where each entry of $\boldsymbol{x}_0$ represents a voxel [92]. The indices of regions of k-space that have been sampled by a readout curve form a sampling set Ω, with $|\Omega| = n$ for $n \leq N$. The number of sampled voxels is $n = n_{read} \cdot n_{samp}$, where n_{read} is the number of read-outs and n_{samp} is the number of samples per readout, and $N = N_x \cdot N_y \cdot N_z$ is the total number of voxels, where N_x, N_y and N_z denote the number of voxels in each spatial dimension. This thesis considers two discrete measurement models: *single-coil*, based on (3.1), which is the subject of Chapter 4, and *multi-coil*, based on (3.2), which is the subject of Chapters 5 and 6. Both measurement models assume that k-space has been sampled on a Cartesian grid: see Section 3.4 for a discussion of non-Cartesian sampling.

3.2.1 Single-coil measurements

The discretised single-coil measurement model (3.1) is

$$\boldsymbol{y} = \boldsymbol{M}_\Omega(\boldsymbol{F}\boldsymbol{x}_0 + \boldsymbol{\varepsilon}), \quad (3.3)$$

where $\boldsymbol{F}$ is a multi-dimensional discrete Fourier transform and $\boldsymbol{M}_\Omega \in \mathbb{R}^{N\times N}$ is a diagonal undersampling mask with 1 on the jth diagonal entry if $j \in \Omega$ and 0 otherwise,

so that $y_j = 0$ for k-space that is not sampled. The measurement noise is typically assumed to follow a white complex Gaussian distribution, $\boldsymbol{\varepsilon} \backsim \mathcal{CN}(\mathbf{0}_N, \sigma_\varepsilon^2 \mathbb{1}_N)$, where $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^2)$ denotes the complex distribution with independent real and imaginary parts that are Gaussian distributed with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}^2/2$.

It is rarely accurate to model actual MRI receiver coils with $S_1(\boldsymbol{z}) = 1$ in practice. However, methods developed for solving (3.3) can usually be extended to non-trivial $S_1(\boldsymbol{z})$ by solving for the coil-weighted image. In other words, by seeking a solution for $\boldsymbol{x}_0^S = \boldsymbol{S}_1 \boldsymbol{x}_0$ in the forward model $\boldsymbol{y} = \boldsymbol{M}_\Omega(\boldsymbol{F}\boldsymbol{x}_0^S + \boldsymbol{\varepsilon})$, where $\boldsymbol{S}_1 \in \mathbb{C}^{N \times N}$ is a diagonal matrix that represents the sensitivity profile [93]. While the single-coil algorithm developed in Chapter 4 focuses on (3.3), we emphasise that it could also be applied in the context of a non-trivial $S_1(\boldsymbol{z})$.

3.2.2 Multi-coil measurements

Multi-coil measurements refer to measurements with multiple non-trivial coil sensitivity profiles modelled with (3.2). The discretised data on coil c is

$$\boldsymbol{y}_c = \boldsymbol{M}_\Omega(\boldsymbol{F}\boldsymbol{S}_c\boldsymbol{x}_0 + \boldsymbol{\varepsilon}_c), \qquad c = \{1, 2, \ldots, N_c\}, \tag{3.4}$$

where $\boldsymbol{S}_c \in \mathbb{C}^{N \times N}$ is the cth sensitivity profile. Alternatively, (3.4) can be expressed by concatenating the per-coil vectors, so that

$$\boldsymbol{y} = \boldsymbol{E}\boldsymbol{x}_0 + \boldsymbol{\varepsilon} \tag{3.5}$$

where

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_{N_c} \end{bmatrix}, \quad \boldsymbol{E} = \begin{bmatrix} \boldsymbol{M}_\Omega \boldsymbol{F} \boldsymbol{S}_1 \\ \boldsymbol{M}_\Omega \boldsymbol{F} \boldsymbol{S}_2 \\ \vdots \\ \boldsymbol{M}_\Omega \boldsymbol{F} \boldsymbol{S}_{N_c} \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{M}_\Omega \boldsymbol{\varepsilon}_1 \\ \boldsymbol{M}_\Omega \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{M}_\Omega \boldsymbol{\varepsilon}_{N_c} \end{bmatrix}.$$

The measurement noise $\boldsymbol{\varepsilon}$ correlates across coils, so is not i.i.d. in general. We denote the measurement noise covariance with $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^2$, so that $\boldsymbol{\varepsilon} \backsim \mathcal{CN}(\boldsymbol{0}_N, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^2)$. The covariance $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^2$ can be estimated empirically by time-averaging data acquired from an empty scanner [94].

The $\boldsymbol{S}_c$ depend on the coil geometry, coil positioning, and magnetic properties of the anatomy, and are unknown a priori. A common approach [1, 95, 96], and the approach used throughout this thesis, is to firstly estimate the coil sensitivities, and subsequently employ them as part of a separate reconstruction algorithm.

The total number of data acquired according to (3.4) is nN_c, so the measurements are only under-determined when $n/N < 1/N_c$. It therefore can be possible reconstruct $\boldsymbol{x}_0$ in the $n/N < 1$ realm, even without the tools of compressed sensing. However, since the coil sensitivities are not orthogonal in practice, parallel imaging inverse problems may be highly ill-conditioned. Approaches that reconstruct images from $n/N < 1$ multi-coil measurements are referred to in the MRI literature as "parallel imaging" methods [97, 98].

A number of algorithms have been proposed for parallel imaging, the most common of which include SENSitivity Encoding (SENSE) [95, 96], SiMultaneous Acquisition of Spatial Harmonics (SMASH), GeneRalized Autocalibrating Partially Parallel Acquisitions (GRAPPA) [99], and iterative Self-consistent Parallel Imaging ReconsTruction (SPIRiT) [100]. SENSE is among the oldest and most widely used, and involves $\boldsymbol{S}_c$ estimation via a series of low-resolution pre-scans acquired separately for

each coil. Reconstruction is then typically performed by solving the least squares problem [95, 101]

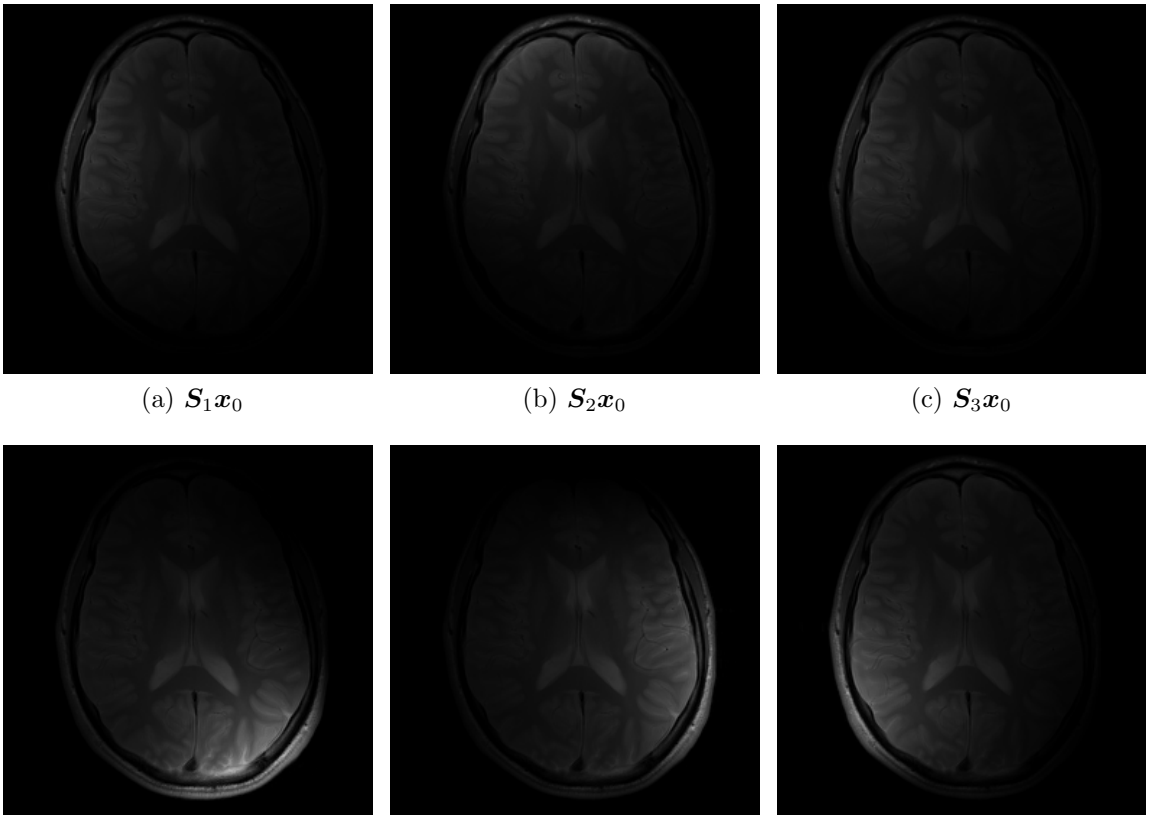
$$\hat{\boldsymbol{x}}_{SENSE} = \arg\min_{\boldsymbol{x}} \left\| \boldsymbol{\Sigma}_{\varepsilon}^{-1}(\boldsymbol{y} - \boldsymbol{E}\boldsymbol{x}) \right\|_2^2,$$

or possibly with a regularisation term [102–104].

The primary disadvantage of SENSE is that the pre-scans are reasonably time-consuming, so are vulnerable to motion artefacts causes by patient movement between the start of the pre-scans and the high resolution acquisition. The more recent Eigenvalue SPIRiT (ESPIRiT) [1] offers an alternative to coil sensitivity estimation that does not require a pre-scan. ESPIRiT estimates the $\boldsymbol{S}_c$ using a fully-sampled central region of k-space, known as the "autocalibration region", from the high-resolution acquisition. The need for an autocalibration region imposes a constraint on the sampling pattern: $j \in \Omega$ for all j in a central region of k-space, which is typically 24×24. Fig. 3.1 shows 6 of the $N_c = 16$ coil-weighted images using sensitivity maps estimated with ESPIRiT for a brain dataset via a 24×24 central autocalibration region calculated using the Berkeley Advanced Reconstruction Toolbox (BART)[1] [105]. Throughout this thesis we follow the ESPIRiT convention that the coil sensitivities are normalised as $\sum_c^{N_c} \boldsymbol{S}_c^H \boldsymbol{S}_c = \mathbb{1}_N$, computed via $\boldsymbol{S}_c = (\sum_{c'}^{N_c} \tilde{\boldsymbol{S}}_{c'}^H \tilde{\boldsymbol{S}}_{c'})^{-1/2} \tilde{\boldsymbol{S}}_c$, where $\tilde{\boldsymbol{S}}_c$ is the cth unnormalised coil.

Many iterative algorithms for parallel MRI apply the Fast Fourier Transform (FFT) once per coil, so have computational complexity $O(N_c N \log N)$. To speed up the algorithm, the poor conditioning of the parallel inverse problem can be exploited to perform dimensionality reduction, reducing the effective N_c. In the MRI literature, this procedure is referred to as "coil compression" [106–108], and the transformed coils are referred to as "virtual" coils. Throughout this thesis, coil compression is performed by Principle Component Analysis (PCA), as in [109], via the BART toolbox

[1] available from `https://mrirecon.github.io/bart/`



(a) $\boldsymbol{S}_1\boldsymbol{x}_0$ (b) $\boldsymbol{S}_2\boldsymbol{x}_0$ (c) $\boldsymbol{S}_3\boldsymbol{x}_0$

(d) $\boldsymbol{S}_4\boldsymbol{x}_0$ (e) $\boldsymbol{S}_5\boldsymbol{x}_0$ (f) $\boldsymbol{S}_6\boldsymbol{x}_0$

Figure 3.1: The magnitude of six of the $N_c = 16$ coil-weighted images from fully sampled brain k-space data, where the sensitivity maps are calculated by ESPIRiT [1] from a 24×24 central autocalibration region. Throughout this thesis, whenever a complex vector is visualised, the magnitude is shown.

[105]. For notational simplicity, whenever coil compression is performed, the new data and virtual coils are denoted in this work with the same notation as the raw data and coils ($\boldsymbol{y}_c$ and $\boldsymbol{S}_c$ respectively), and the number of original coils and virtual coils are stated explicitly in the text.

3.3 Accelerating MRI with compressed sensing

A major disadvantage of MRI is that the hardware and physiological limitations on the gradient fields $\boldsymbol{q}(t)$ imply that the data acquisition process is inherently slow, which can affect image quality. For instance, motion artefacts may arise due to respiration

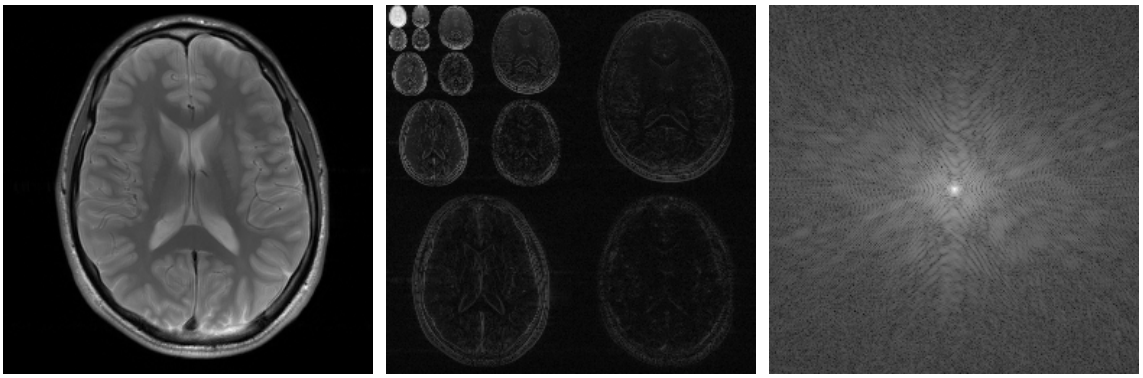
[110] or cardiac motion [111]. Further, long scan times limits throughput, increasing costs and reducing the numer of patients for which the modality is available [112].

Acquisition time can be reduced by using fewer readout curves, so that $n < N$. MRI is a prominent success of compressed sensing [9,113–116]. As discussed in Section 3.2.2, while single-coil measurements with $n/N < 1$ is an ill-*posed* problem, the multi-coil problem may only be ill-*conditioned*. Nonetheless, it is well known that multicoil reconstruction benefits from the random sampling and regularised optimisation developed for compressed sensing, accommodating further acceleration [117, 118]. The application of compressed sensing to MRI is well motivated for the following reasons:

- **Sampling can be randomised.** Compressed sensing recovery guarantees depend on certain deterministic properties of the sensing matrix, such as the RIP described in Section 2.3, which is known to hold for many classes of *random* matrices. This is well-suited to MRI: other than the constraint of smooth 1D trajectories, the choice of MRI sampling pattern can be freely engineered, so can be randomised. For MRI in practice the RIP *motivates* random sampling, without being explicitly calculated.

- **MR images are compressible.** Fig. 3.2b shows the wavelet coefficients of the brain from Fig. 3.2a for an order 4 Daubechies wavelet at 4 decomposition scales, a large proportion of which are near-zero. As well as transforms such as wavelets that compress most images of interest, there are often also further application-specific redundancies that can be exploited. For instance, in dynamic MRI, where acquisitions are taken at a series of time frames, many voxels remain the same, or change slowly, so a voxel-wise temporal Fourier transform increases sparsity [119, 120].

The problem size in MRI can be very large. For instance, the 3D multi-coil brain angiography datasets used in Chapter 5 are comprised of $N \backsim 10^9$ complex datapoints.



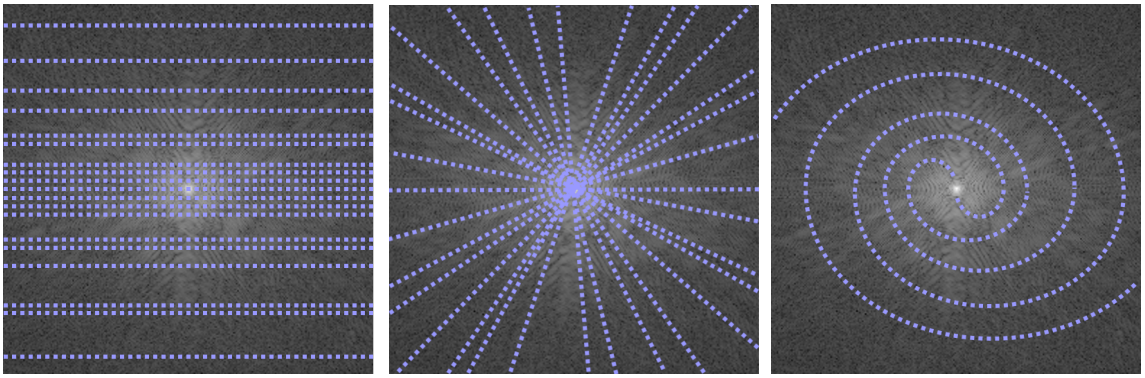
(a) Reference image, $\boldsymbol{x}_0$. (b) Wavelet domain, $\boldsymbol{\Psi x}_0$. (c) Single-coil k-space, $\boldsymbol{F x}_0$.

Figure 3.2: The sparse wavelet representation and non-uniform spectral density of a brain MR image, where $\boldsymbol{\Psi}$ is an order 4 Daubechies wavelet at 4 decomposition scales. Here, and throughout this thesis, all representations of k-space have low frequencies at the centre.

Computational efficiency and scalability is therefore a key consideration for algorithm design. The complexity of many iterative compressed sensing algorithms such as ISTA, Algorithm 1, is governed by the action of the sensing matrix and its transpose on a vector. For MRI in practice, the sensing matrix is never stored in memory, rather, the $O(N \log N)$ complexity Fast Fourier Transform (FFT) is employed. This is key to the computational feasibility of compressed sensing MRI reconstruction.

3.4 Sampling schemes

Images of interest typically have a highly non-uniform spectral density that is concentrated at low frequencies, as illustrated in Fig. 3.2c, which shows the Fourier transform of the brain from Fig. 3.2a. Accordingly, it is well-known that better image restoration is possible if the sampling set Ω is generated with variable density, so that there is a higher probability of sampling low frequencies. A variety of variable density sampling schemes for Fourier sampled images have been suggested [9, 121–125], including some with recovery guarantees [126, 127]. The focus of this thesis is on the reconstruction algorithm, not on the particular choice of sampling scheme. We



(a) Rows. (b) Radial. (c) Spiral.

Figure 3.3: Illustrations of 2D k-space sampling schemes.

therefore use sampling schemes that are known to work reasonably well in practice, without claiming that they are necessarily the best choice.

For 2D acquisitions, common sampling schemes include: (a) fully sampled rows or columns of k-space, (b) radial trajectories and (c) spiral trajectories. An example of each of these types is illustrated in Fig. 3.3. The sampling pattern is referred to as Cartesian when the sampling locations lie on a Cartesian grid. For non-Cartesian sampling strategies, the data is usually projected on to a Cartesian grid with an appropriate weighting, known as a "density compensation" factor [101, 128].

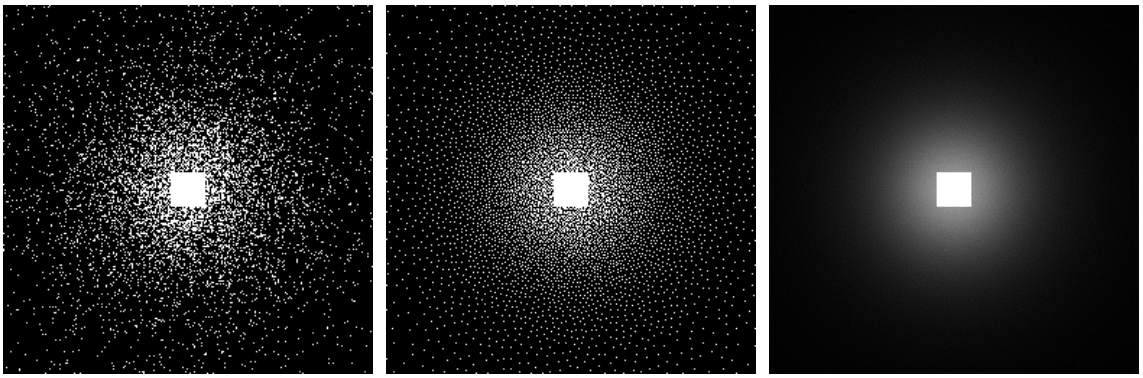
For 3D sampling, the additional spatial degree of freedom implies there is more flexibility in the sampling pattern design. The simplest possibility is to do a series of 2D acquisitions with slice-selective gradients, using, for instance, one of the sampling patterns from Fig. 3.3. Alternatively, it is possible to do fully 3D acquisition. One possible 3D acquisition would be to fully sample in one direction. An advantage of this approach is that a 1D inverse Fourier transform can be applied in the readout direction, so that the 3D k-space is spatially decomposed in to a series of 2D sub-problems with identical sampling sets Ω. This is commonly done to reduce the memory requirements of 3D reconstruction tasks [9, 125]. In this case, the Ω of each 2D sub-problem can be entirely arbitrary, and are not constrained to smooth

trajectories.

This thesis considers considers two types of unconstrained k-space sampling schemes. Firstly, we consider a sampling set Ω with elements drawn independently from a Bernoulli distribution with generic non-uniform probability, which we term "Bernoulli sampling". Secondly, for the multi-coil algorithms in Chapters 5 and 6, we also employ the widely used variable density "Poisson disc sampling" [125, 129], which constrains the sampling set to a minimum distance that varies as function of the position in k-space. Poisson disc sampling is motivated by the desire to improve the conditioning of the multi-coil sensing matrix, and is widely known to offer improved reconstruction quality compared with Bernoulli sampling [111, 125, 130–133]. An example of a variable density distribution Bernoulli sampling and Poisson disc sampling for a 256×256 k-space at $N/n = 10$ is shown in Fig. 3.4, which includes a fully sampled 24×24 central autocalibration region. Also shown is the expectation of the masks, $\boldsymbol{P} = \mathbb{E}_\Omega\{\boldsymbol{M}_\Omega\}$. All Poisson disc masks in this thesis were generated using code from the ESPIRiT package[2], which is based on the algorithm presented in [129].

Frequently in this thesis, we consider reconstruction problems with two spatial dimensions where the sampling set Ω is Bernoulli or Poisson disc, but emphasise that for actual, prospectively undersampled data this would be a single slice of a 3D problem with a fully sampled direction. This thesis does not consider data that have not been fully sampled in one direction, such as 3D radial [134] or 3D cones [135], nor to purely 2D acquisitions.

[2]downloaded from `https://people.eecs.berkeley.edu/~mlustig/Software.html`



(a) Bernoulli sampling $\boldsymbol{M}_\Omega$. (b) Poisson disc sampling $\boldsymbol{M}_\Omega$. (c) Variable density probability $\boldsymbol{P} = \mathbb{E}_\Omega\{\boldsymbol{M}_\Omega\}$.

Figure 3.4: Variable density Bernoulli and Poisson disc sampling schemes for $N/n = 10$ sampling of a 256×256 k-space, with a 24×24 fully sampled central autocalibration region. The sampling masks in figures 3.4a and 3.4b are represented with a white pixel when $j \in \Omega$ and black otherwise.

3.5 Image modelling

Given measurements $\boldsymbol{y}$ acquired with sampling set Ω, reconstruction is usually performed by seeking a solution of the optimisation problem

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{M}_\Omega \boldsymbol{F}\boldsymbol{x}\|_2^2 + f(\boldsymbol{x})$$

for single-coil measurements or

$$\min_{\boldsymbol{x}} \frac{1}{2}\sum_{c=1}^{N_c} \|\boldsymbol{y}_c - \boldsymbol{M}_\Omega \boldsymbol{F}\boldsymbol{S}_c\boldsymbol{x}\|_2^2 + f(\boldsymbol{x})$$

for multi-coil measurements, where $f(\boldsymbol{x})$ is the image model. Numerous models, many of which are application-specific, have been proposed. Motivated by the theory of compressed sensing, it is common to seek a sparse solution via an ℓ_1 penalty with a hand-selected sparse basis. For instance, a common penalty is $f(\boldsymbol{x}) = \lambda\|\boldsymbol{\Psi}\boldsymbol{x}\|_1$, where $\boldsymbol{\Psi}$ is a wavelet transform. A number of variations of sparse models exist, including patch-based sparsity [136] and convolutional sparse encoding [137, 138]. Compressed

sensing with dictionary learning [139] has been applied to MRI [140], as well a low rank [141, 142] and low rank plus sparse methods [143, 144].

3.6 Deep learning for MRI reconstruction

In recent years there has been significant research interest in MRI reconstruction algorithms that leverage the statistical modelling capabilities of neural networks [17, 145–155]. This section briefly introduces neural networks and reviews some of the ways they have been used to reconstruct MRI data.

3.6.1 Convolutional neural networks

A single layer of an artificial neural network acting on an input vector $\boldsymbol{v}$ consists of weights $\boldsymbol{\mathcal{W}}$, biases $\boldsymbol{b}$ and a nonlinear activation function $\boldsymbol{\sigma}(\cdot)$. The output of the lth layer is

$$\boldsymbol{d}_l(\boldsymbol{v}) = \boldsymbol{\mathcal{W}}_l \boldsymbol{\sigma}_l(\boldsymbol{v}) + \boldsymbol{b}_l.$$

A multi-layer neural network consists of a series of single layers,

$$\hat{\boldsymbol{u}}(\boldsymbol{v}) = \boldsymbol{d}_{n_l}(\boldsymbol{d}_{n_l-1}(\ldots \boldsymbol{d}_2(\boldsymbol{d}_1(\boldsymbol{v}))\ldots))$$

where $\hat{\boldsymbol{u}}(\boldsymbol{v})$ is the output of the neural network and n_l is the number of layers. Given training points $\boldsymbol{v}_i$ with associated labels $\boldsymbol{u}_i$, training consists of minimising a non-convex loss function $\sum_i L(\boldsymbol{u}_i, \hat{\boldsymbol{u}}(\boldsymbol{v}_i))$ with respect to the weights and biases (and activation functions, if parametrised). Network training is usually performed using Stochastic Gradient Descent (SGD) [156], or a variation of SGD such as AdaGrad [157] or Adam [158].

A convolutional neural network (CNN) is a subclass of neural networks where

the $\boldsymbol{\mathcal{W}}_l$ are primarily convolutions [159]. Since convolutional $\boldsymbol{\mathcal{W}}_l$s are defined by the kernel only, CNNs greatly reduce the computational burden of neural network computations, and are widely used for imaging tasks [160].

3.6.2 Approaches to MRI reconstruction with deep learning

A variety of approaches to MRI reconstruction with CNNs have been proposed [2, 161, 162]. They can roughly be divided in to:

1. **P&P algorithms**, as introduced in the context of AMP in Section 2.6.2, which replace one step of a compressed sensing algorithm with a denoiser [17, 145, 146], such as the denoising CNN described in Section 3.6.3.

2. **'Unrolled' algorithms**, which use a network architecture based on the structure of a compressed sensing algorithm [147–150].

3. **'Black box' mapping**, which learns the entire mapping between k-space and the image output, without being based on a compressed sensing algorithm [151–155].

The focus of Chapter 6 is on developing a P&P method for MRI with a CNN denoiser. To our knowledge, [145] and [17] are the only published works that apply a P&P algorithm with a CNN denoiser to MRI reconstruction. Both of these works used D-ADMM, originally proposed in [63]. Since the per-iteration aliasing of ADMM is not known, the network was trained to denoise images corrupted by white, complex Gaussian noise as a proxy for the true aliasing. This is a technical limitation of D-ADMM; there is a substantial mismatch between the aliasing the network is trained to denoise and the actual aliasing of the input to the denoiser. Nonetheless, D-ADMM was shown to offer performance improvements in comparison to a number of algorithms that do not employ deep learning, and similarly to a black box method

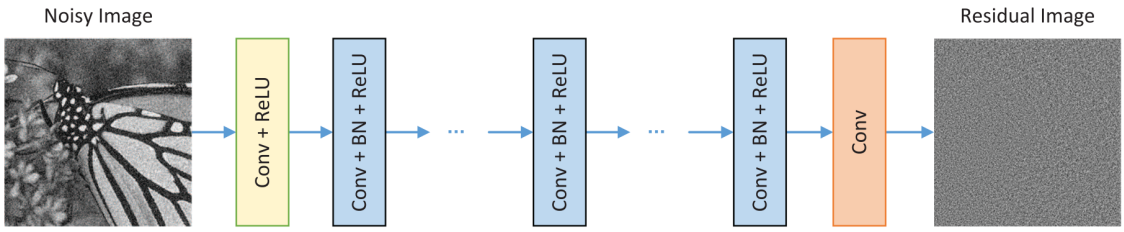


Figure 3.5: The network architecture of Denoising CNN (DnCNN). Figure from [2].

based on a U-Net architecture [163]. It was shown in [17] that P&P methods are more robust to variations in undersampling factors than the comparative U-Net. That they are not constructed for a specific k-space undersampling schemes is a key advantage of P&P methods over unrolled and black box methods.

3.6.3 Denoising CNN

The neural network architecture employed in many P&P methods, including [17, 72] and the method proposed in Chapter 6, is based on the Denoising CNN (DnCNN) from [2]. DnCNN, illustrated in Fig. 3.5, consists of three sections:

1. The first layer, illustrated with a yellow rectangle, consists of 64 filters of size $3 \times 3 \times c$, where c is the number of input channels, followed by the Rectified Linear Unit (ReLU) nonlinearity, which acts entry-wise as $\sigma(x_j) = \max(0, x_j)$.

2. The second type of layer, illustated with a blue rectangle, consists of 64 filters of size $3 \times 3 \times 64$, followed by batch normalisation [164] and ReLU. This layer is repeated $n_d - 2$ times, where n_d is the network depth.

3. The final layer, illustrated with an orange rectangle, consists of c filters of size $3 \times 3 \times 64$.

DnCNN uses "residual learning" [165], where the network learns the difference between the input and the ground truth. A considerable body of work exists ([166–168], for instance) that demonstrates the benefit of residual learning for deep networks.

3.7 Survey of AMP for MRI

As discussed in Section 2.4, AMP's state evolution is only guaranteed for i.i.d. sub-Gaussian measurement, while OAMP and VAMP's state evolution is guaranteed for right-unitarily invariant matrices. Sensing matrices comprised of random Fourier samples are not a member of this class, so there are no theoretical guarantees in the context of MRI. In [12], empirical evidence was presented that AMP's state evolution holds for uniformly random (i.e. non-variable density) Fourier samples of an artificial 1D sparse signal, without proof.

To our knowledge, there are three papers and one conference proceeding on AMP-based methods for variable density Fourier sampled images that were published prior to the development of the methods in this thesis, which are summarised here.

3.7.1 Location Constrained AMP

Location Constrained AMP (LCAMP) [74] is a version of AMP that uses presumed knowledge of the support set of the sparse representation of $\boldsymbol{x}_0$. They propose to replace the soft thresholding step with the application of a mask applied in the wavelet domain, so that the denoiser is $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k) = \boldsymbol{\Psi}^H \boldsymbol{M}_S \boldsymbol{\Psi} \boldsymbol{r}_k$, where $\boldsymbol{\Psi}$ is an orthogonal wavelet transform. Here, $S \subseteq \{1, \ldots, N\}$ is a presumed support set and $\boldsymbol{M}_S$ is a diagonal mask with nonzeroes indexed by S. A clear issue with this work is the selection of S. Even if S was known a priori, it would be possible to reconstruct the image directly via the sensing matrix pseudo-inverse; an iterative algorithm would not be required.

3.7.2 Dynamic Compressed Sensing AMP

In [75], a version of AMP modified for dynamic signals was proposed, with a signal model that assumes the support varies slowly in time and the non-zeros vary smoothly.

The algorithm was evaluated on a series of 10 acquisitions of the larynx, as in [169], and was found to perform competitively with the method proposed in [170].

3.7.3 BM3D-AMP-MRI

BM3D-AMP-MRI [76] is a version of D-AMP [44], as discussed in Section 2.6.2, that uses the Block-Matching 3D (BM3D) denoiser [68,171]. This algorithm is a modification of the BM3D-AMP algorithm previously proposed for i.i.d. Gaussian measurements in [69]. The authors compared BM3D-AMP-MRI with a number of state-of-the-art compressed sensing MRI algorithms, and demonstrate competitive performance. BM3D-AMP consistently outperforms the "BM3D-IT" algorithm, where the Onsager correction term is omitted, although state-evolution does not occur, so it is unclear why exactly the Onsager correction was beneficial for reconstruction quality.

3.7.4 VAMP for Image Recovery (VAMPire)

VAMP for Image Recovery (VAMPire) [77] is an adaption of VAMP, Algorithm 3, for variable density Fourier sampled images. VAMPire uses wavelets to decompose the effective noise in to frequency bands that are subsequently "whitened" by a pre-selected prediction of the per-subband energy, so that measurements are written as

$$\boldsymbol{y} = \boldsymbol{M}_\Omega(\boldsymbol{F}\boldsymbol{R}^{1/2}\boldsymbol{s} + \boldsymbol{\varepsilon})$$

where $\boldsymbol{R} \in \mathbb{R}^{N\times N}$ is an estimate of $\boldsymbol{x}_0\boldsymbol{x}_0^H$, which is approximated in the wavelet domain, and $\boldsymbol{s} \in \mathbb{R}^N$ is the whitened signal. It was found that adapting the algorithm in this way improves the convergence properties of VAMP, although it is not claimed that a state evolution occurred.

3.8 Summary

Data acquisition in MRI consists of traversing a series of 1D smooth trajectories through k-space, which is an inherently slow process. Since sampling can be randomised, and signals of interest are typically compressible, MRI is a natural application and prominent success of compressed sensing.

Despite performing well for certain abstract compressed sensing problems, the literature on AMP for MRI is limited. The key shortcoming of each of the AMP-based methods described in Section 3.7 is that *they do not obey a state evolution*, so do not benefit from the advantages discussed in Section 2.6.

The objective of this thesis is to develop an algorithm that obeys a state evolution for MRI measurements. Chapter 4 presents the Variable Density Approximate Message Passing (VDAMP) algorithm for the single-coil measurement model, (3.3). This framework is extended to the multi-coil measurement model (3.4) in Chapter 5, and in Chapter 6, we present a P&P version of VDAMP with a CNN denoiser.

Chapter 4

Single-coil VDAMP

This chapter presents VDAMP for the single-coil measurement model, (3.3). The key distinguishing feature of VDAMP compared with the methods discussed in Section 3.7 is that, to our knowledge, *VDAMP is the first algorithm for variable density Fourier sampled images that obeys a state evolution*[1]. In the context of wavelet-sparse models, we demonstrate how VDAMP's state evolution can be leveraged to automatically tune sparse weightings to near-optimality on-the-fly using SURE, alleviating the need to hand-tune model parameters, and converging very quickly to a reconstruction with competitive error. We find that VDAMP converges in around 15 times fewer iterations than Fast ISTA (FISTA) [38,179] on average, and typically to a lower or similar mean-squared-error, even when FISTA is optimally tuned: see Section 4.3.

Throughout this chapter we consider a sampling set Ω with elements drawn independently from a Bernoulli distribution with generic non-uniform probability, so that $\text{Prob}(j \in \Omega) = p_j \in [0, 1]$. This chapter is based on publications in the IEEE Open Journal of Signal Processing [172], and in the proceedings of the International Conference on Image Processing (ICIP) 2020 [173].

[1]After the publication of VDAMP [172,173], a number of AMP-based algorithms were proposed for MRI [174–178]. As for the methods described in Section 3.7, none of these proposals obey state evolution, except for [178], which is based on VDAMP.

4.1 Coloured aliasing

AMP, OAMP and VAMP assume that the sensing matrix is sufficiently random to ensure that the aliasing is white. However, when an image is sampled in the Fourier domain, the aliasing is innately coloured. To see this, consider the estimator

$$\widetilde{\boldsymbol{x}} = \boldsymbol{F}^H \boldsymbol{P}^{-1} \boldsymbol{y}, \tag{4.1}$$

where $\boldsymbol{P} = \mathbb{E}_\Omega\{\boldsymbol{M}_\Omega\}$ is a diagonal matrix formed from sampling probabilities p_j. Eqn. (4.1) is a natural initialisation for an AMP-based algorithm because $\widetilde{\boldsymbol{x}}$ is unbiased over the random mask and measurement noise,

$$\begin{aligned}\mathbb{E}_{\Omega,\varepsilon}\{\tilde{\boldsymbol{x}}\} &= \mathbb{E}_{\Omega,\varepsilon}\{\boldsymbol{F}^H \boldsymbol{P}^{-1} \boldsymbol{M}_\Omega (\boldsymbol{F}\boldsymbol{x}_0 + \boldsymbol{\varepsilon})\} \\ &= \boldsymbol{F}^H \boldsymbol{P}^{-1} \boldsymbol{P} \boldsymbol{F} \boldsymbol{x}_0 = \boldsymbol{x}_0,\end{aligned}$$

where $\mathbb{E}_\Omega\{\boldsymbol{M}_\Omega\} = \boldsymbol{P}$ and $\mathbb{E}_\varepsilon\{\boldsymbol{\varepsilon}\} = \boldsymbol{0}_N$ has been used. The following result derives the power spectrum of the aliasing of $\widetilde{\boldsymbol{x}}$, which is non-uniform in general.

Claim 4.1. The power spectrum of the aliasing of $\widetilde{\boldsymbol{x}}$ is

$$\mathbb{E}_{\Omega,\varepsilon}\{|\boldsymbol{y}_0 - \boldsymbol{P}^{-1}\boldsymbol{y}|^2\} = (\boldsymbol{P}^{-1} - \mathbb{1}_N)|\boldsymbol{y}_0|^2 + \sigma_\varepsilon^2 \boldsymbol{P}^{-1}\boldsymbol{1}_N, \tag{4.2}$$

where $\boldsymbol{y}_0 = \boldsymbol{F}\boldsymbol{x}_0$ and $|\cdot|$ is the entry-wise absolute value.

Proof. Since the entries of $\boldsymbol{y}$ are independent and $\boldsymbol{P}^{-1}$ is diagonal, we can consider (4.2) as N distinct one-dimensional expressions. The ith entry of the left-hand-side

of (4.2) is

$$\begin{aligned}\mathbb{E}_{m_i,\varepsilon_i}\left\{\left|y_{0,i}-\frac{y_i}{p_i}\right|^2\right\}=&\mathbb{E}_{m_i,\varepsilon_i}\left\{\left|y_{0,i}-\frac{m_i}{p_i}(y_{0,i}+\varepsilon_i)\right|^2\right\}\\ =&\mathbb{E}_{m_i,\varepsilon_i}\left\{\left|\left(1-\frac{m_i}{p_i}\right)y_{0,i}-\frac{m_i}{p_i}\varepsilon_i\right|^2\right\},\end{aligned}\tag{4.3}$$

where m_i is the ith diagonal of $\boldsymbol{M}_\Omega$. By assumption, m_i is distributed according to a Bernoulli distribution with $\mathbb{E}_{m_i}\{m_i\}=p_i$, so the expectation over m_i can be found by resolving (4.3) at $m_i=1$ and $m_i=0$ and summing with weights p_i and $1-p_i$ respectively:

$$\mathbb{E}_{m_i,\varepsilon_i}\left\{\left|\left(1-\frac{m_i}{p_i}\right)y_{0,i}-\frac{m_i}{p_i}\varepsilon_i\right|^2\right\}=\mathbb{E}_{\varepsilon_i}\left\{p_i\left|\left(\frac{1-p_i}{p_i}\right)y_{0,i}-\frac{\varepsilon_i}{p_i}\right|^2+(1-p_i)|y_{0,i}|^2\right\}.$$

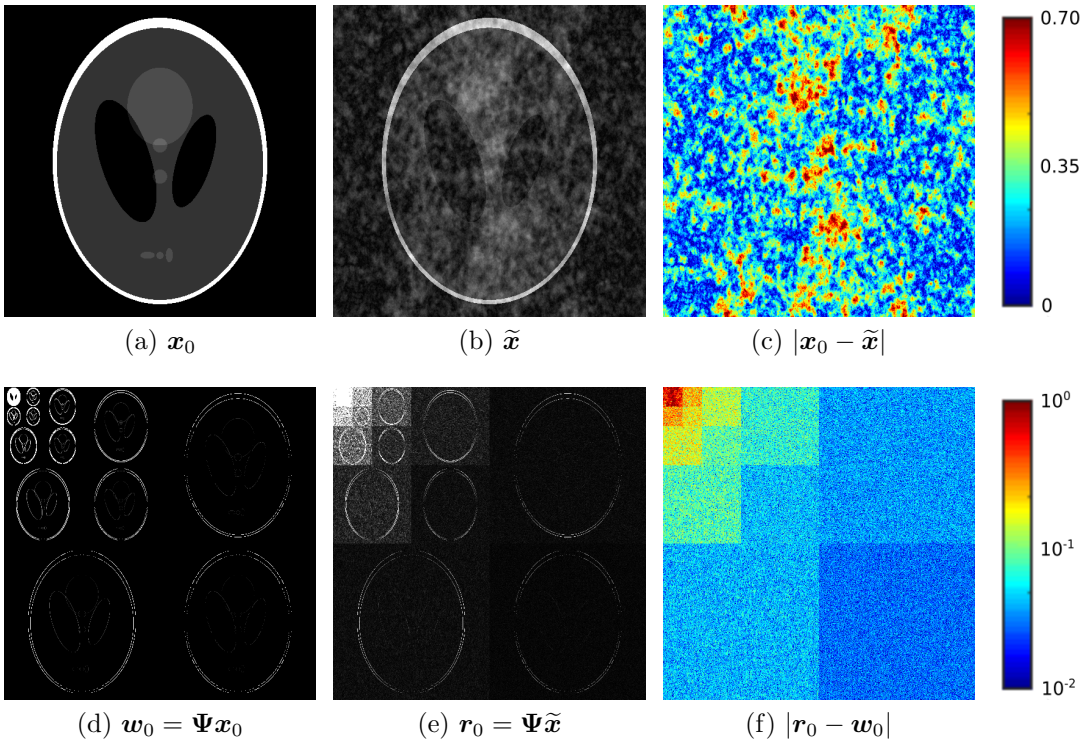
Expanding the first term, and noting that $\mathbb{E}_{\varepsilon_i}\{\varepsilon_i\}=0$ and $\mathbb{E}_{\varepsilon_i}\{|\varepsilon_i|^2\}=\sigma_\varepsilon^2$,

$$\begin{aligned}&\mathbb{E}_{\varepsilon_i}\left\{p_i\left|\left(\frac{1-p_i}{p_i}\right)y_{0,i}-\frac{\varepsilon_i}{p_i}\right|^2+(1-p_i)|y_{0,i}|^2\right\}\\ =&\mathbb{E}_{\varepsilon_i}\left\{p_i\left|\left(\frac{1-p_i}{p_i}\right)y_{0,i}\right|^2+p_i\left|\frac{\varepsilon_i}{p_i}\right|^2-\left(\frac{1-p_i}{p_i}\right)(y_{0,i}^*\varepsilon_i+y_{0,i}\varepsilon_i^*)+(1-p_i)|y_{0,i}|^2\right\}\\ =&\frac{(1-p_i)^2}{p_i}|y_{0,i}|^2+\frac{\sigma_\varepsilon^2}{p_i}+(1-p_i)|y_{0,i}|^2\\ =&\left(\frac{1-p_i}{p_i}\right)|y_{0,i}|^2+\frac{\sigma_\varepsilon^2}{p_i}\end{aligned}$$

which is the ith entry of the right-hand-side of (4.2). This completes the proof. □

Eqn. (4.2) depends on $\boldsymbol{P}$ and $|\boldsymbol{y}_0|^2$, which are non-uniform and anisotropic in general. Although the specific case $p_j=(\sigma_\varepsilon^2+|y_{0,j}|^2)/(\alpha+|y_{0,j}|^2)$ for constant α does lead to white aliasing, it requires knowledge of the ground truth spectral density $|\boldsymbol{y}_0|^2$ so is not a feasible sampling scheme in practice.

A visual example of coloured aliasing is shown in the top row of Fig. 4.1. Here, the unbiased estimate of a 512×512 synthetic Shepp-Logan phantom uniformly sampled



(a) $\boldsymbol{x}_0$ (b) $\widetilde{\boldsymbol{x}}$ (c) $|\boldsymbol{x}_0 - \widetilde{\boldsymbol{x}}|$

(d) $\boldsymbol{w}_0 = \boldsymbol{\Psi}\boldsymbol{x}_0$ (e) $\boldsymbol{r}_0 = \boldsymbol{\Psi}\widetilde{\boldsymbol{x}}$ (f) $|\boldsymbol{r}_0 - \boldsymbol{w}_0|$

Figure 4.1: The ground truth, unbiased estimate and entry-wise absolute error for a uniformly sampled Shepp-Logan with $p_j = 1/2$, where the colourbars are given as a proportion of the maximum of $\boldsymbol{x}_0$, as throughout this thesis. The top row shows the image domain and the bottom shows the wavelet domain. The coloured aliasing evident in (c) illustrates the infeasibility of white state evolution for Fourier sampling of signals with non-uniform spectral density. Taking the wavelet transform decomposes the aliasing so that it is approximately white subband.

with $p_j = 1/2$ for all j is shown, with $\sigma_\varepsilon = 0$. By (4.2), the power spectrum of the aliasing of $\tilde{\boldsymbol{x}}$ in this case is $\mathbb{E}_{\Omega,\varepsilon}\{|\boldsymbol{y}_0 - \boldsymbol{P}^{-1}\boldsymbol{y}|^2\} = |\boldsymbol{y}_0|^2$. Since $|\boldsymbol{y}_0|^2$ is tightly concentrated at low frequencies the aliasing has strong low-frequency components, which is manifest in Fig. 4.1c as local correlations. In this example, uniform sampling was chosen to exaggerate the coloured aliasing property. In practice, when variable density sampling is used, the sampling distribution partially compensates for the power spectrum of the signal, so the aliasing is still coloured, but less strongly.

The intrinsically coloured aliasing of variable density sampling from a non-uniform spectral density implies that the white state evolution of AMP, OAMP, and VAMP, (2.7), cannot be relied upon. A key component of this work is the use of the Discrete

Wavelet Transform (DWT) to compute a multiresolution decomposition of the power spectrum of the aliasing, as used for coloured noise analysis in [180–182]. In the wavelet domain, coloured aliasing has a structure that resembles a state evolution. To illustrate this, consider again the unbiased initialisation $\widetilde{\boldsymbol{x}}$. The corresponding estimator in the wavelet domain is

$$\boldsymbol{r}_0 = \boldsymbol{\Psi}\widetilde{\boldsymbol{x}}. \tag{4.4}$$

The bottom row of Fig. 4.1 shows $\boldsymbol{w}_0 = \boldsymbol{\Psi}\boldsymbol{x}_0$, $\boldsymbol{r}_0$ and the entry-wise absolute difference $|\boldsymbol{w}_0 - \boldsymbol{r}_0|$ for the same image and sampling set as the top row, where $\boldsymbol{\Psi}$ is a Haar DWT with 4 decomposition scales. Qualitatively, Fig. 4.1f suggests that the power spectrum given by (4.2) has a simple structure in the wavelet domain; in particular, it suggests that the per-subband aliasing within each subband is approximately uniform. In Section 4.3.5 we show that, in fact, *the aliasing within each subband is quantitatively consistent with a white Gaussian distribution.* Further, Section 4.3.5 presents evidence that the Gaussianity holds for a variety of image types and undersampling factors.

4.1.1 Coloured state evolution

Herein we present a new method for undersampled signal reconstruction that we term the Variable Density Approximate Message Passing (VDAMP) algorithm, see Algorithm 4. We present empirical evidence that VDAMP preserves the subband-dependent noise structure illustrated in Fig. 4.1f for all iterations. Explicitly, at iteration k the $\boldsymbol{r}_k$ of VDAMP behaves as

$$\boldsymbol{r}_k = \boldsymbol{w}_0 + \mathcal{CN}(\boldsymbol{0}_N, \boldsymbol{\Sigma}_k^2), \tag{4.5}$$

where the covariance matrix $\boldsymbol{\Sigma}_k^2$ is diagonal so that for a $\boldsymbol{\Psi}$ with s decomposition scales,

$$\boldsymbol{\Sigma}_k^2 = \begin{bmatrix} \sigma_{k,1}^2 \mathbb{1}_{N_1} & 0 & \dots & 0 \\ 0 & \sigma_{k,2}^2 \mathbb{1}_{N_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{k,1+3s}^2 \mathbb{1}_{N_{1+3s}} \end{bmatrix}, \tag{4.6}$$

where $\sigma_{k,b}^2$ and N_b refer to the variance and dimension of the bth subband respectively. We refer to (4.5) and (4.6) as the *coloured state evolution* of VDAMP and the aliasing of $\boldsymbol{r}_k$ as the *effective noise* of VDAMP.

The joint space-frequency localisation provided by the wavelet transform decomposes the colour of the effective noise while retaining incoherence. The algorithm presented in this chapter considers a sparse model on wavelet coefficients, however, we emphasise that the wavelet transform is primarily used as a tool for decomposing the aliasing, and is not necessarily constrained to model wavelet coefficients directly [44, 58, 178].

The VAMPire algorithm discussed in Section 3.7.4 also made use of the wavelet transform to decompose the aliasing. The key difference between VAMPire and the current work is that VAMPire's aliasing is approximately 'whitened' and its effective noise model is represented by a scalar τ_k. In contrast, we propose making the necessary algorithmic adaptions to *allow the aliasing to be coloured*, and to model the colour with a *vector* $\boldsymbol{\tau}_k$. To our knowledge, VDAMP is the first algorithm for variable density Fourier sampling of images where a state evolution has been observed.

AMP's state evolution was defined by (2.7), which states that its aliasing is white Gaussian distributed, and (2.10), which predicts the per-iteration variance σ_k^2 of the aliasing based solely on σ_{k-1}^2. Since (4.5) and (4.6) are analogous to (2.7) only, the coloured state evolution of VDAMP is weaker. Unlike AMP, the per-iteration aliasing $\boldsymbol{\Sigma}_k$ is *not* a function of $\boldsymbol{\Sigma}_{k-1}$; it is a function of the residual $\boldsymbol{z}_k$, as detailed

Algorithm 4 VDAMP

Require: Sampling set Ω, wavelet transform $\boldsymbol{\Psi}$, probability matrix $\boldsymbol{P}$, measurements $\boldsymbol{y}$, denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$, number of iterations K_{it}.

1: Set $\widetilde{\boldsymbol{r}}_0 = \boldsymbol{0}_N$ and compute $\boldsymbol{S} = |\boldsymbol{F}\boldsymbol{\Psi}^H|^2$
2: **for** $k = 0, 1, \ldots, K_{it} - 1$ **do**
3: $\quad \boldsymbol{z}_k = \boldsymbol{y} - \boldsymbol{M}_\Omega \boldsymbol{F}\boldsymbol{\Psi}^H \widetilde{\boldsymbol{r}}_k$
4: $\quad \boldsymbol{r}_k = \widetilde{\boldsymbol{r}}_k + \boldsymbol{\Psi}\boldsymbol{F}^H \boldsymbol{P}^{-1} \boldsymbol{z}_k$
5: $\quad \boldsymbol{\tau}_k = \boldsymbol{S}^T \boldsymbol{M}_\Omega \boldsymbol{P}^{-1} [(\boldsymbol{P}^{-1} - \mathbb{1}_N)|\boldsymbol{z}_k|^2 + \sigma_\varepsilon^2 \boldsymbol{1}_N]$
6: $\quad \hat{\boldsymbol{w}}_k = \boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$
7: $\quad \boldsymbol{\alpha}_k = \langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)) \rangle_{\text{sband}}$
8: $\quad$ Update $\boldsymbol{c}_k$ e.g. $\boldsymbol{c}_k = \boldsymbol{1}_N \oslash (\boldsymbol{1}_N - \boldsymbol{\alpha}_k)$
9: $\quad \widetilde{\boldsymbol{r}}_{k+1} = \boldsymbol{c}_k \odot (\hat{\boldsymbol{w}}_k - \boldsymbol{\alpha}_k \odot \boldsymbol{r}_k)$
10: **end for**
11: **return** $\hat{\boldsymbol{x}} = \boldsymbol{\Psi}^H \hat{\boldsymbol{w}}_k + \boldsymbol{F}^H (\boldsymbol{y} - \boldsymbol{M}_\Omega \boldsymbol{F}\boldsymbol{\Psi}^H \hat{\boldsymbol{w}}_k)$

in Section 4.2. VDAMP's coloured state evolution therefore does not have a self-contained prediction of the per-iteration variance analogous to (2.10), so cannot be used to analyse, for instance, the conditions under which VDAMP converges, without actually running the algorithm. Instead, the focus of this thesis is on the advantages of state evolution as a guide for denoising, as introduced in Section 2.6.

4.2 Description of VDAMP algorithm

The VDAMP algorithm, Algorithm 4, adapts OAMP, Algorithm 3, to a coloured aliasing model. We chose to modify Matched-Filter OAMP because because it lends itself more naturally to the per-subband aliasing model illustrated in Fig. 4.1 than AMP, VAMP, or other versions of OAMP, which implicitly employ a white aliasing model in the *measurement* domain. For AMP, the implicit measurement-domain white aliasing model is contained in the Onsager correction, the final term of line 3 of Algorithm 2, while for VAMP it is the $\widetilde{\tau}_k$ in the data consistency phase, shown in (2.11).

VDAMP's per-subband coloured aliasing model is updated in line 5, where the

scalar $\tau_k \in \mathbb{R}$ of OAMP is replaced by a vector $\boldsymbol{\tau}_k \in \mathbb{R}^N$ that models the diagonal of $\boldsymbol{\Sigma}_k^2$ from (4.6). In line 6, OAMP's denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$ is replaced by the denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$, which takes the vector $\boldsymbol{\tau}_k$ as its second input. Lines 7-9 of Algorithm 4 is the Onsager correction from lines 7-9 of Algorithm 3 adapted to a subband-wise aliasing model. Here, the notation $\boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k))$ in line 7 replaces OAMP's $\boldsymbol{g}'(\boldsymbol{r}_k; \tau_k)$, defined as the function with jth entry

$$\partial_j(\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)) = \frac{1}{2}\left(\frac{\partial \Re[g_j(\boldsymbol{r}_k; \boldsymbol{\tau}_k)]}{\partial \Re[r_j]} + \frac{\partial \Im[g_j(\boldsymbol{r}_k; \boldsymbol{\tau}_k)]}{\partial \Im[r_j]}\right), \tag{4.7}$$

where $\Re[\cdot]$ and $\Im[\cdot]$ are the real and imaginary parts respectively. Also in line 7, the notation $\langle\cdot\rangle_{\text{sband}}$ is an operator that empirically averages entries within subbands, so that $\boldsymbol{\alpha}_k$ has the structure

$$\boldsymbol{\alpha}_k = \begin{bmatrix} \alpha_{k,1}\mathbf{1}_{N_1} \\ \alpha_{k,2}\mathbf{1}_{N_2} \\ \vdots \\ \alpha_{k,1+3s}\mathbf{1}_{N_{1+3s}} \end{bmatrix}, \tag{4.8}$$

with

$$\alpha_{k,b} = \frac{1}{N_b}\sum_{j \in J_b} \partial_j(\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)),$$

where J_b is the set of indices associated with subband b and $N_b = |J_b|$. In line 8 of Algorithm 4, $\boldsymbol{c}_k$ is a vector with the piecewise-constant structure of (4.8). The notation $\oslash$ in line 8 refers to entry-wise division and $\odot$ in line 9 refers to entry-wise multiplication, so that $[\boldsymbol{a} \oslash \boldsymbol{b}]_i = a_i/b_i$ and $[\boldsymbol{a} \odot \boldsymbol{b}]_i = a_i \cdot b_i$ for vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. The remainder of this section describes Algorithm 4 in detail, line-by-line, with direct comparison to its basis algorithm OAMP, Algorithm 3.

4.2.1 Density compensated gradient descent, lines 3-4

To ensure that $\boldsymbol{r}_k$ is an unbiased estimate of $\boldsymbol{x}_0$, the sensing matrix must be correctly normalised. In VDAMP this is manifest in the gradient step of lines 3-4, which features a crucial weighting by $\boldsymbol{P}^{-1}$ that is absent in the previous applications of AMP to variable density sampling [74–77] discussed in Section 3.7. Note that VDAMP's $\boldsymbol{r}_0$ is the unbiased estimator from (4.4). The absent $\boldsymbol{P}^{-1}$ term in [74–77] implies that $\boldsymbol{r}_0$ is biased for these methods in general, so state evolution fails even at $k = 0$.

Such a rescaling is referred to as "density compensation" in the MRI literature, and was used in the original compressed sensing MRI paper with zero-filling to generate a unregularised, non-iterative baseline [9]. However, to our knowledge, VDAMP is the first *iterative* method that employs density compensated gradient descent. The density compensation factor can be understood as a left-multiplication of the forward model by $\boldsymbol{P}^{-1/2}$, so that the model is rescaled as

$$\boldsymbol{P}^{-1/2}\boldsymbol{y} = \boldsymbol{P}^{-1/2}\boldsymbol{M}_\Omega(\boldsymbol{F}\boldsymbol{x}_0 + \boldsymbol{\varepsilon}) \tag{4.9}$$

$$= \boldsymbol{P}^{-1/2}\boldsymbol{M}_\Omega(\boldsymbol{F}\boldsymbol{\Psi}^H\boldsymbol{w}_0 + \boldsymbol{\varepsilon}). \tag{4.10}$$

Lines 3-4 of VDAMP, Algorithm 4, are equivalent to lines 3-4 of OAMP, Algorithm 3, with $\boldsymbol{P}^{-1/2}\boldsymbol{M}_\Omega\boldsymbol{F}\boldsymbol{\Psi}^H$ and $\boldsymbol{P}^{-1/2}\boldsymbol{y}$ as the sensing matrix and measurements respectively[2]. Density compensation provides the correct normalisation of the density compensated sensing matrix in expectation over Ω: $\mathbb{E}_\Omega\{\boldsymbol{\Psi}\boldsymbol{F}^H\boldsymbol{P}^{-1}\boldsymbol{M}_\Omega\boldsymbol{F}\boldsymbol{\Psi}^H\} = \mathbb{1}_N$, which is a requirement of AMP, as discussed in Section 2.3.3.

Density compensation here differs from its usual use in MRI as an interpolation weighting when mapping non-Cartesian samples to a Cartesian grid [101, 128], as discussed in Section 3.4. If VDAMP were applied to non-Cartesian data, it would

[2]Multiplying a forward model by a matrix is common in numerical analysis for preconditioning [183]. However, we emphasise that the role of $\boldsymbol{P}^{-1/2}$ here is to ensure that $\boldsymbol{r}_k$ is unbiased, not to reduce the condition number of the sensing matrix, so we refrain from using the term "preconditioner" in this context.

be necessary to perform density compensation twice: firstly to map the samples to a Cartesian grid, and secondly to de-bias the estimator, which would have different density compensation weightings in general.

VDAMP's type of density compensation arises in recovery guarantees for variable density Fourier measurements in [126, 184], although it was considered an artefact of the proof and was not used in the numerical evaluations of these works. The connection of VDAMP to these theoretical results is beyond the scope of this thesis and is left as future work.

The measurement noise of the density compensated measurement model (4.10) is $\boldsymbol{P}^{-1/2}\boldsymbol{M}_\Omega\boldsymbol{\varepsilon}$. The variance of the measurement noise at frequencies sampled with low probability is therefore larger, and its inclusion in the gradient step will lead to a $\boldsymbol{r}_k$ with higher mean-squared error than an unweighted gradient step. However, as shown in Section 4.3, a careful choice of denoiser $\boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k)$ that leverages VDAMP's state evolution can cause lines 6-9 to be very effective, leading to faster overall convergence than competing methods.

The final step of VDAMP, line 11, is a gradient step without a $\boldsymbol{P}^{-1}$ weighting, which generates a biased image estimate $\hat{\boldsymbol{x}}$ with high data fidelity. In the experimental examples of Section 4.3, we found that this choice of output improved the mean-squared error of the reconstructed image compared with the natural alternative $\hat{\boldsymbol{x}} = \boldsymbol{\Psi}^H\hat{\boldsymbol{w}}_k$. However, it has been noted that when the measurement noise variance σ_ε^2 is large, $\boldsymbol{\Psi}^H\hat{\boldsymbol{w}}_k$ may be a better choice [178].

4.2.2 Coloured effective noise model, line 5

Line 5 of OAMP, Algorithm 3, computes a scalar that estimates the variance of $\boldsymbol{r}_k$. For VDAMP, this is modified to accomodate the colored aliasing of Fourier sampled images, as illustrated in Fig. 4.1. Line 5 of Algorithm 4 computes a *vector*, which models the diagonal of the wavelet-domain coloured effective noise covariance matrix

$\boldsymbol{\Sigma}_k^2$ from (4.5). Through a similar derivation to that for Claim 4.1, the power spectrum of the aliasing of $\boldsymbol{r}_k$ is

$$\mathbb{E}_{\Omega,\varepsilon}\{|\boldsymbol{F}\boldsymbol{\Psi}^H\boldsymbol{r}_k - \boldsymbol{y}_0|^2\} = (\boldsymbol{P}^{-1} - \mathbb{1}_N)|\boldsymbol{F}\boldsymbol{\Psi}^H\widetilde{\boldsymbol{r}}_k - \boldsymbol{y}_0|^2 + \sigma_\varepsilon^2\boldsymbol{P}^{-1}\mathbf{1}_N. \tag{4.11}$$

Eqn. (4.11) depends on the unknown ground truth $\boldsymbol{y}_0$, so is of limited practical use. The following claim demonstrates how (4.11) can be estimated in an unbiased fashion using the vector $\boldsymbol{z}_k = \boldsymbol{y} - \boldsymbol{M}_\Omega\boldsymbol{F}\boldsymbol{\Psi}^H\widetilde{\boldsymbol{r}}_k$ from line 2 of Algorithm 4 only, which is known.

Claim 4.2. The vector

$$\boldsymbol{\tau}_k^y = \boldsymbol{M}_\Omega\boldsymbol{P}^{-1}[(\boldsymbol{P}^{-1} - \mathbb{1}_N)|\boldsymbol{z}_k|^2 + \sigma_\varepsilon^2\mathbf{1}_N] \tag{4.12}$$

is an unbiased estimate of (4.11): $\mathbb{E}\{\boldsymbol{\tau}_k^y\} = \mathbb{E}_{\Omega,\varepsilon}\{|\boldsymbol{F}\boldsymbol{\Psi}^H\boldsymbol{r}_k - \boldsymbol{y}_0|^2\}$.

Proof. Define $\widetilde{\boldsymbol{y}}_k = \boldsymbol{F}\boldsymbol{\Psi}^H\widetilde{\boldsymbol{r}}_k$. The expectation of the ith entry of the right-hand-side of (4.12) in terms of $\boldsymbol{y}_0$ is

$$\mathbb{E}_{m_i,\varepsilon_i}\left\{\frac{m_i}{p_i}\left[\left(\frac{1-p_i}{p_i}\right)|y_i - m_i\widetilde{y}_{k,i}|^2 + \sigma_\varepsilon^2\right]\right\} = \mathbb{E}_{m_i,\varepsilon_i}\left\{\frac{m_i}{p_i}\left[\left(\frac{1-p_i}{p_i}\right)|m_i(y_{0,i}+\varepsilon_i) - m_i\widetilde{y}_{k,i}|^2 + \sigma_\varepsilon^2\right]\right\}.$$

Evaluating the expectation over m_i,

$$\mathbb{E}_{m_i,\varepsilon_i}\left\{\frac{m_i}{p_i}\left[\left(\frac{1-p_i}{p_i}\right)|m_i(y_{0,i}+\varepsilon_i) - m_i\widetilde{y}_{k,i}|^2 + \sigma_\varepsilon^2\right]\right\} = \mathbb{E}_{\varepsilon_i}\left\{\left(\frac{1-p_i}{p_i}\right)|y_{0,i} + \varepsilon_i - \widetilde{y}_{k,i}|^2 + \sigma_\varepsilon^2\right\}.$$

Since $\mathbb{E}_{\varepsilon_i}\{\varepsilon_i\} = 0$ and $\mathbb{E}_{\varepsilon_i}\{|\varepsilon_i|^2\} = \sigma_\varepsilon^2$,

$$\begin{aligned}\mathbb{E}_{\varepsilon_i}\left\{\left(\frac{1-p_i}{p_i}\right)|y_{0,i}+\varepsilon_i-\widetilde{y}_{k,i}|^2+\sigma_\varepsilon^2\right\} &= \left(\frac{1-p_i}{p_i}\right)(|y_{0,i}-\widetilde{y}_{k,i}|^2+\sigma_\varepsilon^2)+\sigma_\varepsilon^2 \\ &= \left(\frac{1-p_i}{p_i}\right)|y_{0,i}-\widetilde{y}_{k,i}|^2+\frac{\sigma_\varepsilon^2}{p_i}\end{aligned}$$

which is the ith entry of the right-hand-side of (4.11). □

The computation of the covariance $\boldsymbol{\tau}_k$ in line 5 of Algorithm 4 is a linear transform of (4.12) to the wavelet domain: $\boldsymbol{\tau}_k = |\boldsymbol{\Psi}\boldsymbol{F}^H|^2\boldsymbol{\tau}_k^y$.

Claim 4.3. The wavelet-domain estimator $\boldsymbol{\tau}_k = |\boldsymbol{\Psi}\boldsymbol{F}^H|^2\boldsymbol{\tau}_k^y$ is unbiased:

$$\mathbb{E}\{\boldsymbol{\tau}_k\} = \mathbb{E}\{|\boldsymbol{w}_0 - \boldsymbol{r}_k|^2\}.$$

Proof. Appendix A shows that $\mathbb{E}\{|\boldsymbol{A}\boldsymbol{u}|^2\} = |\boldsymbol{A}|^2\mathbb{E}\{|\boldsymbol{u}|^2\}$ for deterministic matrix $\boldsymbol{A}$ and random $\boldsymbol{u}$ with independent, zero-mean entries. Therefore

$$\begin{aligned}\mathbb{E}\{|\boldsymbol{w}_0 - \boldsymbol{r}_k|^2\} &= \mathbb{E}\{|\boldsymbol{\Psi}\boldsymbol{F}^H(\boldsymbol{y}_0 - \boldsymbol{F}\boldsymbol{\Psi}^H\boldsymbol{r}_k)|^2\} \\ &= |\boldsymbol{\Psi}\boldsymbol{F}^H|^2\mathbb{E}\{|\boldsymbol{y}_0 - \boldsymbol{F}\boldsymbol{\Psi}^H\boldsymbol{r}_k|^2\} \\ &= |\boldsymbol{\Psi}\boldsymbol{F}^H|^2\mathbb{E}\{\boldsymbol{\tau}_k^y\},\end{aligned}$$

where the final step is by Claim 4.2. □

The rows of $|\boldsymbol{\Psi}\boldsymbol{F}^H|^2$ are the power spectra of the rows of $\boldsymbol{\Psi}$. Since rows of $\boldsymbol{\Psi}$ that form the same subband are related by translation, which causes a phase shift in its Fourier transform, their power spectra are identical. The matrix $|\boldsymbol{\Psi}\boldsymbol{F}^H|^2$ therefore has $1+3s$ unique rows, one per subband, and the computation of $\boldsymbol{\tau}_k = |\boldsymbol{\Psi}\boldsymbol{F}^H|^2\boldsymbol{\tau}_k^y$ in line 5 of Algorithm 4 has complexity $O((1+3s)N)$. For fixed s VDAMP is therefore dominated computationally by the action of $\boldsymbol{F}$ and $\boldsymbol{F}^H$ in lines 3 and 4, which have complexity $O(N \log N)$.

4.2.3 Complex soft threshold tuning with SURE, line 6

Line 6 of VDAMP applies a denoising function $\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$, which replaces $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k)$ in line 6 of OAMP. In other words, OAMP's scalar alising model τ_k is replaced by VDAMP's vector aliasing model $\boldsymbol{\tau}_k$. The remainder of this section motivates and describes the specific denoiser employed in the chapter in detail.

A challenge for compressed sensing MRI frequently noted by researchers in the clinical community [185–187] is how to tune model parameters, which is typically done by hand [114]. The ideal parameter choice depends on a multitude of case-specific factors including the anatomy, underlying MR physics and sampling strategy, so hand-tuning generally yields sub-optimal reconstruction. We present an approach to parameter-free compressed sensing reconstruction that leverages VDAMP's state evolution by applying SURE [16], building on the work on AMP introduced in Section 2.6.1 [13–15].

A strength of automatic parameter tuning via SURE is that it is possible to have a richer regulariser than would usually be feasible for a hand-tuned denoiser, enabling higher order prior information such as anisotropy, variability across scales and structured sparsity, without the need to estimate the structure a priori [188]. In this chapter, the denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$ in line 6 of Algorithm 4 was the complex soft thresholding operator (2.6) with a subband-dependent threshold tuned automatically with SURE via the noise model $\boldsymbol{\tau}_k$. Concretely, the effective noise model $\boldsymbol{r}_k = \boldsymbol{w}_0 + \mathcal{CN}(\boldsymbol{0}_N, \mathrm{Diag}(\boldsymbol{\tau}_k))$ was used to compute

$$\boldsymbol{t}_{opt} \approx \underset{\boldsymbol{t} \in \mathbb{R}^N}{\mathrm{argmin}} \|\boldsymbol{\eta}(\boldsymbol{r}_k; \boldsymbol{t}) - \boldsymbol{w}_0\|_2, \tag{4.13}$$

where $\boldsymbol{t}$ is an entry-wise threshold that has the piecewise-constant structure of (4.8), and the denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k) = \boldsymbol{\eta}(\boldsymbol{r}_k; \boldsymbol{t}_{opt})$ was employed. The possibility of using scale-dependent thresholds is well known, such as in [38]. We emphasise that (4.13) is

subband-dependent rather than scale-dependent, enabling higher order, anisotropic signal modeling [189, 190].

Equation (4.13) was solved using a procedure related to SureShrink [191] but for Gaussian noise that is complex and coloured [182]. The following claim establishes the appropriate modification of real-valued SURE, (2.14), to complex variables.

Claim 4.4. Consider a vector $\boldsymbol{v}_0 \in \mathbb{C}^{N_v}$ corrupted by coloured complex Gaussian noise with diagonal covariance: $\boldsymbol{v} = \boldsymbol{v}_0 + \mathcal{CN}(\boldsymbol{0}_{N_v}, \mathrm{Diag}(\boldsymbol{\tau}_v))$. Let $\boldsymbol{d}(\boldsymbol{v}) = \boldsymbol{v} + \boldsymbol{h}(\boldsymbol{v})$ be an estimator of $\boldsymbol{v}_0$. Complex SURE,

$$cSURE(\boldsymbol{d}(\boldsymbol{v})) = \|\boldsymbol{h}(\boldsymbol{v})\|_2^2 + \boldsymbol{\tau}_v^T[2\boldsymbol{\partial}(\boldsymbol{d}(\boldsymbol{v})) - \boldsymbol{1}_{N_v}], \tag{4.14}$$

is an unbiased estimate of the risk: $cSURE(\boldsymbol{d}(\boldsymbol{v})) = \mathbb{E}\{\|\boldsymbol{d}(\boldsymbol{v}) - \boldsymbol{v}_0\|_2^2\}$.

Proof. See Appendix B.1. □

The optimal parameters of the denoiser $\boldsymbol{d}(\boldsymbol{v})$ can therefore be estimated by minimizing cSURE as a proxy for the true risk. The following claim evaluates (4.14) for when $\boldsymbol{d}(\boldsymbol{v})$ is the complex soft thresholding operator $\boldsymbol{\eta}(\boldsymbol{v}; t)$ with scalar threshold t.

Claim 4.5. When the denoiser $\boldsymbol{d}(\boldsymbol{v})$ is complex soft thresholding, cSURE (4.14) is

$$cSURE(\boldsymbol{\eta}(\boldsymbol{v}; t)) = \sum_{j:|v_j|\leq t}^{N_v} (|v_j|^2 - \tau_{v,j}) + \sum_{j:|v_j|>t}^{N_v} \left(t^2 + \tau_{v,j} - \frac{t\tau_{v,j}}{|v_j|}\right). \tag{4.15}$$

Proof. See Appendix B.2. □

The optimal threshold of VDAMP for the bth subband can be estimated with

$$t_{b,opt} = \underset{t}{\mathrm{argmin}}(cSURE(\boldsymbol{\eta}_b(\boldsymbol{r}_{k,b}; t))) \tag{4.16}$$

by evaluating (4.15) for trial thresholds $t = |v_1|, |v_2|, \ldots, |v_{N_v}|$ with $\boldsymbol{\tau}_v = \tau_b \boldsymbol{1}_{N_b}$. The near-optimal vector threshold $\boldsymbol{t}_{opt}$ from (4.13) can then be formed from the $t_{b,opt}$. For

large dimension N one would expect by the law of large numbers that cSURE is close to the true risk, and for the threshold to be almost optimal. Since a larger number of decomposition scales s give subbands with lower dimension, there is a trade-off between the size of s and the quality of threshold selection with cSURE.

SURE has previously been employed for parameter-free compressed sensing MRI in [192], where FISTA was used with SURE-minimising soft thresholds. This algorithm is herein referred to as SURE-IT, and is discussed in detail in Section 4.3.1.

4.2.4 Complex, coloured Onsager correction, lines 7-9

Lines 7-9 of VDAMP, Algorithm 4, is complex, per-subband Onsager correction. Intuitively, since lines 7-9 of OAMP, Algorithm 3, apply to white noise, and the effective noise of VDAMP is white within each subband, Onsager correction must be applied subband-by-subband. For soft thresholding, the use of $\langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k))\rangle_{\text{sband}}$ in place of OAMP's $\langle \boldsymbol{g}'(\boldsymbol{r}_k;\tau_k)\rangle$ leads to the function formed by merging lines 7-9,

$$\tilde{\boldsymbol{g}}(\boldsymbol{r}_k;\boldsymbol{\tau}_k) = \boldsymbol{c}_k \odot (\boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k) - \boldsymbol{\alpha}_k \odot \boldsymbol{r}_k), \tag{4.17}$$

to obey

$$\langle \boldsymbol{\partial}(\tilde{\boldsymbol{g}}(\boldsymbol{r}_k;\boldsymbol{\tau}_k))\rangle_{\text{sband}} = 0, \tag{4.18}$$

which is a coloured aliasing version of OAMP's divergence-free condition, (2.13), applied to both real and imaginary parts. The Onsager correction employed here is not the only choice that leads to a $\tilde{\boldsymbol{g}}(\boldsymbol{r}_k;\boldsymbol{\tau}_k)$ that satisfies (4.18) [193], however we have found that this particular choice performs well.

Like the scalar c_k updated on line 8 of Algorithm 3, the $\boldsymbol{c}_k$ updated in line 8 of Algorithm 4 is not constrained by (4.18), except to have the piecewise-constant structure of (4.8). In this chapter, two possibilities for the $\boldsymbol{c}_k$ update were considered.

First, we considered

$$\boldsymbol{c}_k^{\alpha} = \mathbf{1}_N \oslash (\mathbf{1}_N - \boldsymbol{\alpha}_k). \tag{4.19}$$

In this case, lines 6-9 of Algorithm 4 are a coloured version of the 'denoising' phase of VAMP when written in LMMSE form [55, Algorithm 3]. VDAMP with $\boldsymbol{c}_k$ updated with (4.19) is herein referred to as VDAMP-α.

Secondly, building on work from [58], we suggest using cSURE (4.14) for a second time to estimate the $\boldsymbol{c}_k$ that minimises the mean squared error of $\widetilde{\boldsymbol{r}}_k$:

$$\boldsymbol{c}_k^{SURE} \approx \arg\min_{\boldsymbol{c}} \|\tilde{\boldsymbol{g}}(\boldsymbol{r}_k; \boldsymbol{\tau}_k) - \boldsymbol{w}_0\|_2^2. \tag{4.20}$$

Since $\langle \boldsymbol{\partial}(\tilde{\boldsymbol{g}}(\boldsymbol{r}_k; \boldsymbol{\tau}_k))\rangle_{\text{sband}} = 0$, optimising (4.14) reduces to a series of ℓ_2 minimisation problems with a closed-form solution, so that for the bth subband

$$c_{k,b}^{SURE} = \arg\min_{c} \|c(\boldsymbol{g}_b(\boldsymbol{r}_k; \boldsymbol{\tau}_k) - \alpha_{k,b}\boldsymbol{r}_{k,b}) - \boldsymbol{r}_{k,b}\|_2^2 \tag{4.21a}$$

$$= \frac{\boldsymbol{r}_{k,b}^H(\boldsymbol{g}_b(\boldsymbol{r}_k; \boldsymbol{\tau}_k) - \alpha_{k,b}\boldsymbol{r}_{k,b})}{\|\boldsymbol{g}_b(\boldsymbol{r}_k; \boldsymbol{\tau}_k) - \alpha_{k,b}\boldsymbol{r}_{k,b}\|_2^2}. \tag{4.21b}$$

Vector $\boldsymbol{c}_k^{SURE}$ is formed from the $c_{k,b}^{SURE}$ so that it has the structure of (4.8). VDAMP with $\boldsymbol{c}_k$ updated with $\boldsymbol{c}_k^{SURE}$ is herein referred to as VDAMP-S.

As for OAMP [54], we do not claim that the either of the two $\boldsymbol{c}_k$ updates suggested here are necessarily the best choice. Instead, we show in Section 4.3 that both updates lead to aliasing consistent with state evolution, and empirically evaluate their performance.

4.3 Numerical experiments

This section evaluates the performance of VDAMP-α and VDAMP-S compared with the Fast Iterative Shrinking-Thresholding Algorithm (FISTA) algorithm [38, 179],

Algorithm 5 FISTA [179], S-FISTA [190] and SURE-IT [192] for Fourier sampled images

Require: Sampling set Ω, wavelet transform $\boldsymbol{\Psi}$, per-subband weighting $\boldsymbol{\Lambda}$ set to $\boldsymbol{\Lambda} = \mathbb{1}_N$ for FISTA and SURE-IT or calculated with (4.22) for S-FISTA, sparse weighting λ for FISTA and S-FISTA, measurements $\boldsymbol{y}$, number of iterations K_{it}.

1: Set $\widetilde{\boldsymbol{r}}_0 = \boldsymbol{0}_N$, $\hat{\boldsymbol{w}}_{-1} = \boldsymbol{0}_N$ and $h_{-1} = 1$
2: **for** $k = 0, 1, \ldots, K_{it} - 1$ **do**
3: $\quad \boldsymbol{z}_k = \boldsymbol{y} - \boldsymbol{F}\boldsymbol{\Psi}^H \widetilde{\boldsymbol{r}}_k$
4: $\quad \boldsymbol{r}_k = \widetilde{\boldsymbol{r}}_k + \boldsymbol{\Lambda}^{-1}\boldsymbol{\Psi}\boldsymbol{F}^H \boldsymbol{M}_\Omega \boldsymbol{z}_k$
5: $\quad$ Update τ_k
6: $\quad$ **if** SURE-IT **then**
7: $\quad\quad \hat{\boldsymbol{w}}_k = \boldsymbol{g}(\boldsymbol{r}_k; \tau_k \boldsymbol{1}_N)$
8: $\quad$ **else**
9: $\quad\quad \hat{\boldsymbol{w}}_k = \boldsymbol{\eta}(\boldsymbol{r}_k; \tau_k \lambda \boldsymbol{\Lambda}^{-1} \boldsymbol{1}_N)$
10: $\quad$ **end if**
11: $\quad h_k = (1 + \sqrt{1 + 4h_{k-1}^2})/2$
12: $\quad \widetilde{\boldsymbol{r}}_{k+1} = \hat{\boldsymbol{w}}_k + (h_{k-1} - 1)(\hat{\boldsymbol{w}}_k - \hat{\boldsymbol{w}}_{k-1})/h_k$
13: **end for**
14: **return** $\hat{\boldsymbol{x}} = \boldsymbol{\Psi}^H \hat{\boldsymbol{w}}_k + \boldsymbol{F}^H(\boldsymbol{y} - \boldsymbol{M}_\Omega \boldsymbol{F}\boldsymbol{\Psi}^H \hat{\boldsymbol{w}}_k)$

and two FISTA-based methods with subband-dependent thresholding [190, 192]. We also present empirical evidence for VDAMP's state evolution. All experiments were conducted in MATLAB 9.4 and can be reproduced with code online, available at `https://github.com/charlesmillard/VDAMP`.

4.3.1 Description of comparative FISTA-based algorithms

The three versions of FISTA [179, 190, 192] considered in this chapter are summarised in Algorithm 5. In line 7, $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k \boldsymbol{1}_N)$ refers to the subband-dependent thresholding function with thresholds tuned by cSURE, stated in (4.16), with a scalar aliasing model τ_k. In line 9, $\boldsymbol{\eta}(\boldsymbol{r}_k; \tau_k \lambda \boldsymbol{\Lambda}^{-1} \boldsymbol{1}_N)$ is the soft thresholding function from (2.6) with entry-wise threshold $\tau_k \lambda \boldsymbol{\Lambda}^{-1} \boldsymbol{1}_N$. The threshold employed here is weighted by the variance estimate τ_k, causing the threshold to decrease over iterations, which, as in [125], was found to significantly reduce the time to convergence. Line 14 is an

unweighted gradient descent step, as in line 11 of Algorithm 4, which outputs an estimate $\hat{\boldsymbol{x}}$ with high data fidelity. This was suggested in [125], and we found that this output generally had a lower reconstruction error than $\boldsymbol{\Psi}^H \hat{\boldsymbol{w}}_k$. The variations described in Algorithm 5, discussed in detail below, are referred to in this work as *FISTA*, *S-FISTA* [190] and *SURE-IT* [192].

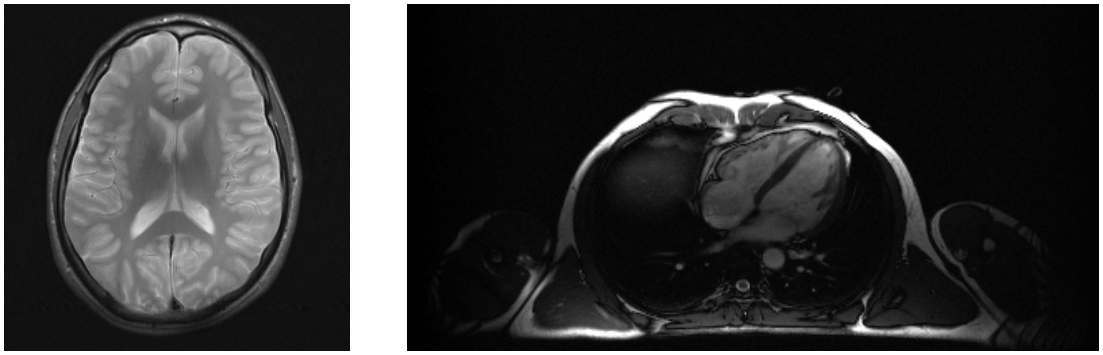
FISTA refers to Algorithm 5 with $\boldsymbol{\Lambda} = \mathbb{1}_N$, so that a global threshold $\lambda\tau_k$ is applied in line 9. Despite not discriminating between subbands, this version of FISTA was considered because it is widely-used in MRI. We found that it was an instructive baseline that performed well compared with the subband-dependent algorithms. FISTA with a hand-tuned subband-dependent λ was not considered as it is not feasible in practice.

S-FISTA refers to Algorithm 5 with a diagonal weight matrix $\boldsymbol{\Lambda}$ that has one unique entry per wavelet subband, so that $\mathrm{diag}(\boldsymbol{\Lambda})$ has the piecewise-constant structure of (4.8). In [190], a method was suggested for selecting the weight for the bth subband, which we denote as w_b. Let $\boldsymbol{\Phi}_b$ be the block of $\boldsymbol{M}_\Omega \boldsymbol{F} \boldsymbol{\Psi}^H$ corresponding to the bth subband. In [190], it is shown that a choice of w_b that satisfies

$$\frac{1}{w_b} > \sum_{b'=1}^{3s+1} \sqrt{\lambda_{\max}(\boldsymbol{\Phi}_{b'}^H \boldsymbol{\Phi}_b \boldsymbol{\Phi}_b^H \boldsymbol{\Phi}_{b'})}, \tag{4.22}$$

where $\lambda_{\max}(\cdot)$ is the largest eigenvalue, ensures that $\boldsymbol{\Lambda} - \boldsymbol{\Phi}^H \boldsymbol{\Phi}$ is a positive operator, and therefore guarantees that the algorithm converges: see [190] for details. Following [194], we computed the w_b by calculating $\lambda_{\max}$ once per Ω for all b and b' using the power iteration method. In contrast with VDAMP, S-FISTA's per-subband weighting is fixed across all iterations, and depends only on the sensing matrix and wavelet family, and not on the per-iteration signal estimate.

SURE-IT refers to Algorithm 5 with $\boldsymbol{\Lambda} = \mathbb{1}_N$ and a subband-dependent, automatically tuned denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \tau_k \mathbf{1})$, as in [192]. Although SURE-IT's thresholding



(a) 256×256 Brain (b) 208×416 Cardiac

Figure 4.2: MRI images used to evaluate the performance of VDAMP.

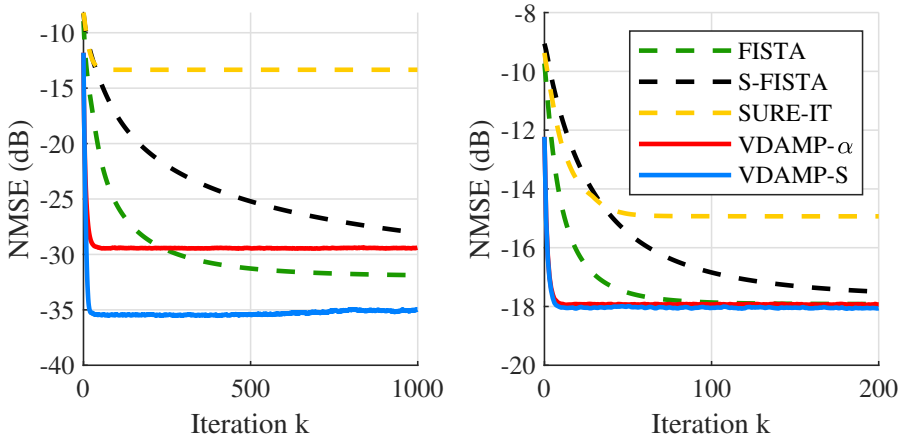
is iteration-dependent, the aliasing of its $\boldsymbol{r}_k$ is highly non-Gaussian, so deviates from a proper theoretical basis for using SURE and does not give near-optimal threshold selection. Further, unlike VDAMP, the aliasing of SURE-IT is modelled by a scalar τ_k.

In the following experiments, the scalar sparse weighting λ of FISTA and S-FISTA was tuned with an exhaustive search so that the mean-squared error was minimised at $k = 100$. Since the threshold was weighted by τ_k, we tuned λ separately for FISTA and S-FISTA. The variance estimate τ_k on line 5 of Algorithm 5 was updated using the ground truth: $\tau_k = \|\boldsymbol{r}_k - \boldsymbol{w}_0\|_2^2/N$.

4.3.2 Experimental method

We considered the reconstruction of 8 test images: the synthetic Shepp-Logan shown in Fig. 4.1a, a brain and a cardiac MR image, shown in Fig. 4.2, and 5 standard test images: Barbara, Boat, Cameraman, House and Peppers, downloaded from [195]. Unless otherwise stated, the undersampled data was artificially corrupted with complex Gaussian noise so that $\|\boldsymbol{x}_0\|_2^2/N\sigma_\varepsilon^2 = 40$dB. We assumed that σ_ε^2 was known a priori. For simplicity, $\boldsymbol{\Psi}$ was the Haar wavelet with s=4 scales.

In the experiments presented in this chapter, we generated a Bernoulli random Ω



(a) Shepp-Logan, $N/n = 10$ (b) Cardiac, $N/n = 4$

Figure 4.3: NMSE of FISTA, S-FISTA, SURE-IT, VDAMP-α and VDAMP-S of a Shepp-Logan undersampled at $N/n = 10$ and the cardiac MR image undersampled at $N/n = 4$. The NMSE at $k = 0$ differs between algorithms as the image estimate at $k = 0$ is defined to be after the first thresholding is applied. The cardiac example is shown up to $k = 200$, not $K_{it} = 500$, so that the behaviour of VDAMP can clearly be seen.

using the variable density sampling function from the Sparse MRI package[3]. We do not claim that this scheme is necessarily the best choice, only that it an instructive example because it is known to perform well in practice [9]. The Ω generated with this package for the $N/n = 4$ cardiac image is shown in Fig. 4.4f. We chose variable density schemes $\boldsymbol{P}$ with $\mathbb{E}_\Omega\{|\Omega|\} = n$ so that the acceleration factors $N/n = 4, 6, 8$ were used, except for the Shepp-Logan, where we used an increased acceleration of $N/n = 8, 10, 12$. All algorithms were initialised with a vector of zeros and run for $K_{it} = 500$ iterations, except for the Shepp-Logan, which was found to require more iterations, so was run until $K_{it} = 1000$.

4.3.3 Time to converge

Table 4.1 shows the time to converge for each algorithm, defined as the time taken until the NMSE is within 0.1dB of its value at K_{it}. In all test cases, both vari-

[3]available at https://people.eecs.berkeley.edu/∼mlustig/Software.html

Image	Shepp-Logan			Brain			Cardiac			Barbara		
N/n	8	10	12	4	6	8	4	6	8	4	6	8
FISTA	57.99	69.55	76.64	1.40	1.53	1.68	3.37	3.45	3.20	1.01	1.65	3.18
S-FISTA	79.91	80.62	81.13	4.22	4.93	4.79	10.41	8.53	9.66	6.00	6.78	10.10
SURE-IT	8.60	4.45	5.25	0.91	0.90	1.08	1.84	20.77	1.67	1.05	1.15	4.50
VDAMP-α	2.99	4.41	**0.96**	**0.16**	**0.13**	0.14	**0.39**	0.38	**0.24**	**0.18**	**0.19**	0.43
VDAMP-S	**2.21**	**2.37**	4.70	0.28	**0.13**	**0.11**	0.43	**0.26**	1.28	0.56	0.33	**0.31**

Image	Boat			Cameraman			House			Peppers		
N/n	4	6	8	4	6	8	4	6	8	4	6	8
FISTA	2.80	4.06	3.08	1.56	1.58	1.22	1.52	1.97	1.94	1.43	1.66	1.94
S-FISTA	10.44	11.71	8.13	5.15	4.94	4.20	4.44	5.08	4.44	5.24	5.29	5.57
SURE-IT	2.31	2.73	2.36	1.15	0.94	0.81	2.27	2.09	3.03	0.73	0.19	0.15
VDAMP-α	**0.44**	**0.42**	**0.31**	**0.19**	**0.13**	0.73	0.17	0.17	0.14	0.14	**0.09**	**0.09**
VDAMP-S	0.70	3.62	2.23	0.99	0.68	**0.06**	**0.14**	**0.09**	**0.09**	**0.10**	0.10	**0.09**

Table 4.1: Convergence time in seconds. The shortest time is highlighted in bold.

Image	Shepp-Logan			Brain			Cardiac			Barbara		
N/n	8	10	12	4	6	8	4	6	8	4	6	8
FISTA	-36.2	-31.9	-25.7	-20.4	-18.2	-17.1	-17.9	**-14.5**	**-13.0**	**-17.5**	**-16.3**	-15.5
S-FISTA	-33.2	-28.1	-23.5	-20.3	-18.1	-17.1	-17.7	-14.3	-12.7	**-17.5**	**-16.3**	-15.5
SURE-IT	-18.7	-13.3	-13.1	-18.9	-17.3	-16.6	-14.9	-12.9	-10.5	-17.0	-16.1	-15.5
VDAMP-α	-35.1	-29.4	-20.5	**-20.7**	**-18.5**	**-17.5**	-17.9	**-14.5**	-12.9	**-17.5**	**-16.3**	**-15.6**
VDAMP-S	**-38.1**	**-34.9**	**-30.8**	-20.6	-18.4	-17.4	**-18.0**	-14.3	-12.4	-16.8	-15.4	-14.1

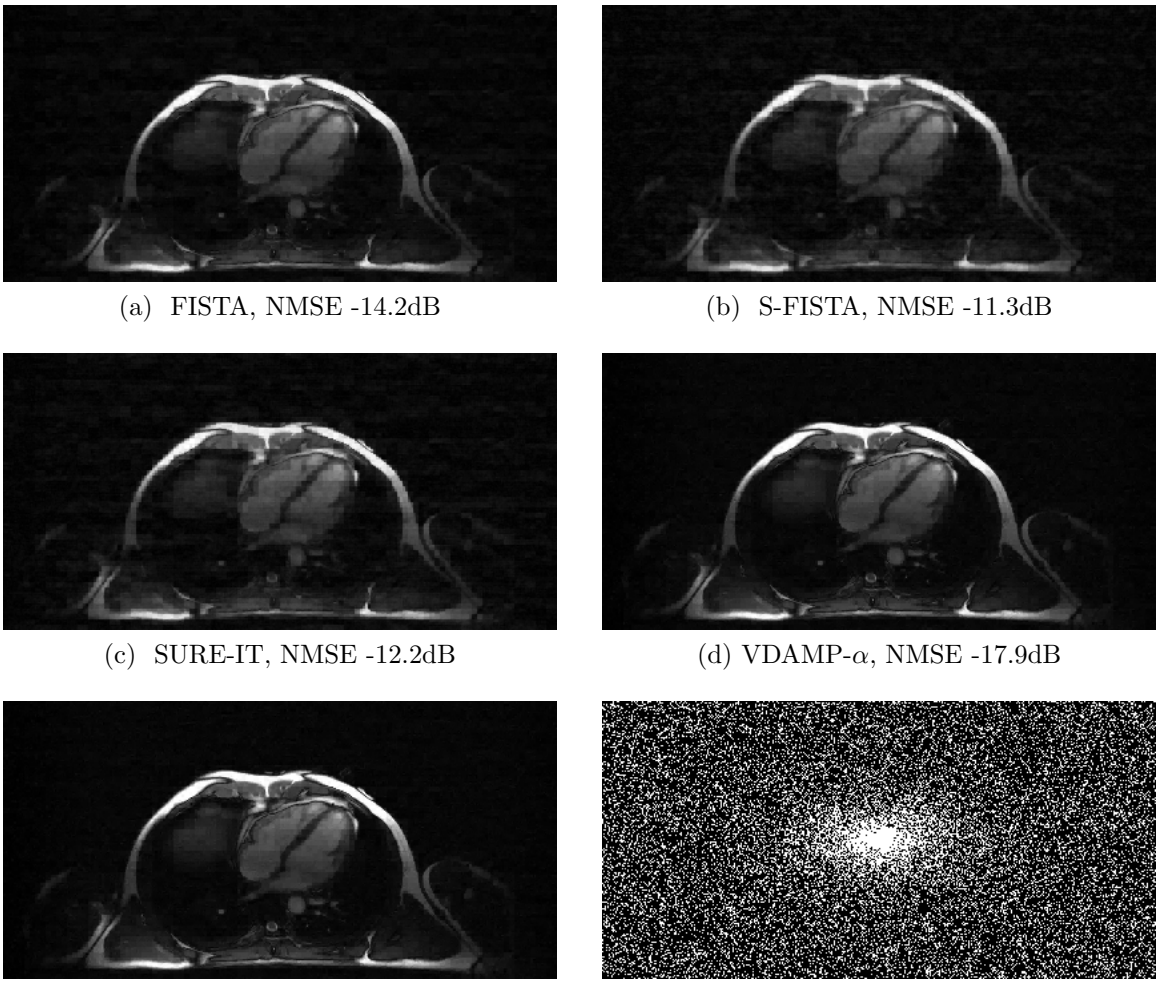
Image	Boat			Cameraman			House			Peppers		
N/n	4	6	8	4	6	8	4	6	8	4	6	8
FISTA	-21.4	-19.6	-18.6	**-20.9**	**-18.4**	**-17.1**	**-25.4**	-22.8	**-21.6**	**-20.0**	-17.5	-16.2
S-FISTA	-21.4	-19.5	-18.5	-20.7	-18.3	-17.0	-25.3	-22.7	-21.4	-19.9	-17.5	-16.1
SURE-IT	-20.3	-18.9	-18.3	-18.7	-16.7	-15.9	-23.3	-21.4	-20.6	-18.5	-16.0	-15.3
VDAMP-α	**-21.6**	**-19.9**	**-18.9**	-20.8	-18.2	-16.0	**-25.4**	**-22.9**	-21.5	-19.9	**-17.7**	**-16.7**
VDAMP-S	-19.5	-16.7	-15.6	-18.7	-16.1	-15.5	-21.5	-18.8	-17.8	-18.0	-15.7	-14.3

Table 4.2: Reconstruction NMSE in dB at K_{it} for all 8 test images at different undersampling factors. The lowest NMSE is highlighted in bold.

Image	Shepp-Logan			Brain			Cardiac			Barbara		
N/n	8	10	12	4	6	8	4	6	8	4	6	8
FISTA	27.99	32.57	38.01	2.86	4.43	5.76	3.94	6.35	9.55	2.10	2.75	3.09
S-FISTA	43.69	45.63	46.83	9.58	11.35	11.97	20.46	23.56	25.42	3.80	4.35	4.55
SURE-IT	39.27	49.64	51.08	4.57	7.32	8.76	6.82	13.92	21.99	2.38	2.88	3.21
VDAMP-α	**0.06**	**0.05**	0.13	-0.06	-0.02	**-0.03**	**-0.06**	**-0.05**	**-0.09**	0.04	**0.00**	**0.00**
VDAMP-S	0.10	0.07	**0.01**	**-0.03**	**-0.01**	0.06	0.20	0.07	0.11	**0.03**	0.10	0.04

Image	Boat			Cameraman			House			Peppers		
N/n	4	6	8	4	6	8	4	6	8	4	6	8
FISTA	2.17	3.35	4.20	3.70	5.33	7.25	3.01	5.23	6.96	2.90	5.13	7.19
S-FISTA	4.79	5.53	5.96	7.26	8.24	8.92	6.14	7.54	8.45	10.04	11.49	12.17
SURE-IT	3.30	4.74	5.27	5.68	8.09	8.92	5.26	7.55	8.45	5.21	10.01	11.30
VDAMP-α	0.02	0.09	0.05	**-0.06**	**-0.04**	**0.07**	**-0.05**	**-0.01**	**-0.01**	**0.00**	0.07	**0.03**
VDAMP-S	**0.01**	**0.02**	**0.01**	-0.08	0.13	-0.09	-0.06	-0.04	-0.07	-0.07	**0.00**	-0.06

Table 4.3: Test for Gaussianity using the mean excess kurtosis $\overline{\text{Kurt}}\{\Re[\boldsymbol{r}_k - \boldsymbol{w}_0]\}$ at K_{it}. An exact Gaussian has zero mean excess kurtosis. The smallest absolute values are highlighted in bold.



(a) FISTA, NMSE -14.2dB

(b) S-FISTA, NMSE -11.3dB

(c) SURE-IT, NMSE -12.2dB

(d) VDAMP-α, NMSE -17.9dB

(e) VDAMP-S, NMSE -18.0dB

(f) $N/n = 4$ sampling set Ω

Figure 4.4: Reconstructions of the cardiac image undersampled at $N/n = 4$ at $k = 10$, where both versions of VDAMP have converged, visualising the relative rapidity of VDAMP's convergence.

ations of VDAMP converged substantially faster than the competing FISTA-based methods. Across all experiments, convergence time compared with FISTA was 14.0 times shorter for VDAMP-α and 11.8 times shorter for VDAMP-S on average. In terms of the required number of iterations, there was 16.5 and 15.2 times reduction for VDAMP-α and VDAMP-S respectively. This corresponds to an increase in per-iteration compute time of 1.18 and 1.29 times; VDAMP-S's higher compute time is due to the additional inner products and norm calculations in (4.21). Of the three FISTA-based algorithms, SURE-IT had the shortest time to convergence, but still re-

quired 10.7 and 10.0 times more iterations on average than VDAMP-α and VDAMP-S respectively. The times listed in Table 4.1 do not include the time required to tune λ for FISTA and S-FISTA, nor the time to calculate S-FISTA's subband-weighting w_b.

The NMSE vs iteration for the Shepp-Logan undersampled at $N/n = 10$ and the cardiac image undersampled at $N/n = 4$ are shown up to K_{it} and $k = 200$ respectively in Fig. 4.3, which visualises VDAMP's comparative rapidity of convergence. In Fig. 4.4, the cardiac image is shown at $k = 10$, where VDAMP-α and VDAMP-S had converged, demonstrating a visible reduction in artefacts for VDAMP.

4.3.4 NMSE comparison

Table 4.2 shows the normalised mean-squared error (NMSE) $\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2^2 / \|\boldsymbol{x}_0\|_2^2$ of the reconstructed image for each algorithm. VDAMP-S performed well for the synthetic Shepp-Logan, but worse than VDAMP-α in all other cases except the $N/n = 4$ sampled cardiac image. The NMSE of FISTA, S-FISTA and VDAMP-α for the non-synthetic images are generally comparable. Given that VDAMP has 13 model parameters while FISTA has one, one might expect the NMSE of VDAMP would consistently be lower. However, for Cameraman, the cardiac image at $N/n = 8$, House at $N/n = 4$ and Peppers at $N/n = 4$, FISTA's NMSE was found to be lower. This is due to density compensation in the gradient step, line 4 of Algorithm 4, which leads to an effective measurement noise of $\boldsymbol{P}^{-1/2}\boldsymbol{M}_\Omega\boldsymbol{\varepsilon}$. When FISTA outperforms VDAMP, the gains from the richer model are insufficient to make up for the disadvantageous effective measurement noise. VDAMP's sensitivity to measurement noise is discussed further in Section 4.3.6. Note that FISTA's NMSE advantage in these instances may not necessarily arise in realistic, prospectively undersampled reconstruction tasks, as the ground truth cannot be used to hand-tune FISTA's sparse weighting λ to a near optimal value, as in the experiments here. Also note that VDAMP-S and VDAMP-α both perform comparatively well on the MRI images, which are of primary importance

for the algorithm's intended application.

Despite employing subband-dependent thesholding, S-FISTA's NMSE at K_{it} was often slightly higher than FISTA. Fig. 4.5 shows $k = 1, 2, 3$ of VDAMP and FISTA's wavelet-domain aliasing for the $N/n = 12$ Shepp-Logan, which demonstrates that the ratio of the aliasing between subbands may not be constant over iterations. For instance, at $k = 1$, VDAMP's coarse level has greater variance than the fine levels, but at $k = 3$ the coarse variance is visibly lower than the fine levels. This is poorly reflected by S-FISTA's threshold weighting, which is fixed over iterations and not dependent on the intermediate estimate. Further, [190] notes that while (4.22) is sufficient to ensure convergence, the inequality is not tight, so may lead to weights that are smaller than necessary, which slows convergence. In [194], which uses different sampling schemes to that employed here, it was found that S-FISTA performed slightly better than FISTA, suggesting that S-FISTA's relative performance may be particularly dependent on the sampling scheme employed. The comparatively poor performance of SURE-IT highlights the need for zero-mean Gaussian aliasing for effective automatic parameter tuning with SURE.

4.3.5 Empirical evidence of state evolution

This section presents empirical evidence that VDAMP obeys the coloured state evolution stated in (4.5) using kurtosis and quantile-quantile plots. The excess kurtosis of a real random variable X is defined as $\text{Kurt}\{X\} = \mu_4/\sigma^4 - 3$, where μ_4 is the fourth central moment and σ is the standard deviation. The Gaussianity of the aliasing of $\boldsymbol{r}_k$ was tested by calculating the mean of per-subband empirical kurtosis of the real part,

$$\overline{\text{Kurt}}\{\Re[\boldsymbol{r}_k - \boldsymbol{w}_0]\} = \frac{1}{3s+1}\sum_{b=1}^{3s+1} \text{Kurt}\{\Re[\boldsymbol{r}_{k,b} - \boldsymbol{w}_{0,b}]\},$$

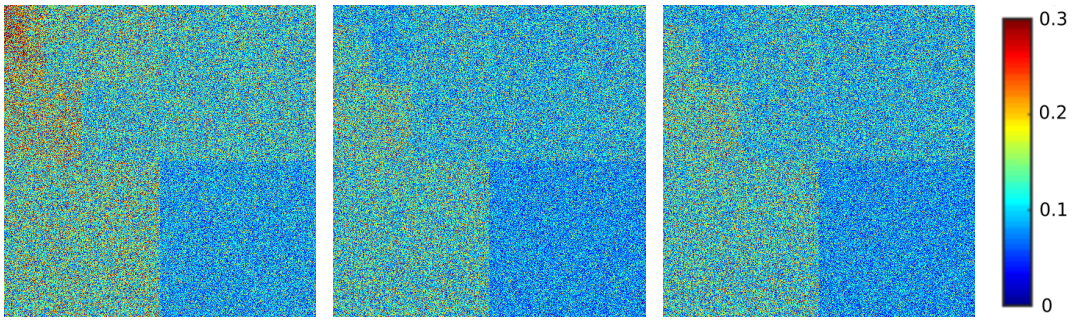
SNR	5	10	15	20	25	30	35	40	45	50
FISTA	**-9.89**	**-12.3**	**-14.5**	**-16.8**	-19.5	-22.2	-24.3	-25.7	-26.5	-26.8
VDAMP-α	-9.44	-11.4	-13.3	-15.6	-18.0	-20.2	-20.3	-20.5	-22.2	-22.7
VDAMP-S	-9.54	-11.4	-13.4	-15.5	**-19.6**	**-23.6**	**-27.2**	**-30.8**	**-35.0**	**-39.7**

Table 4.4: NMSE in dB of a $N/n = 12$ sampled Shepp-Logan for a range of measurement noise variances σ_ε^2.

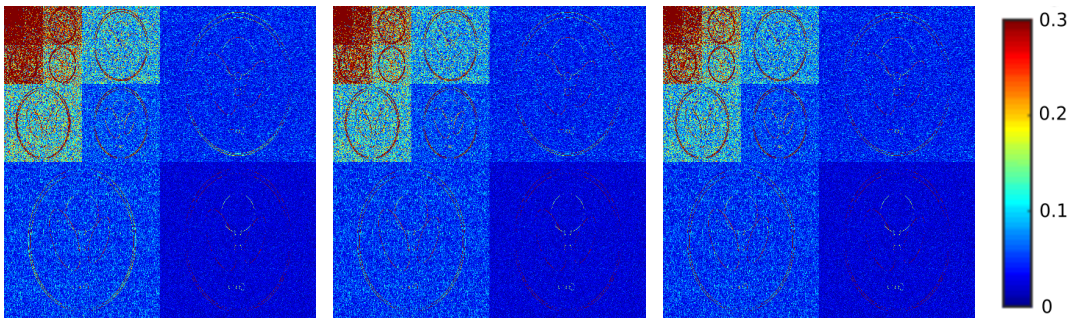
and comparing to zero, which is the kurtosis of a white Gaussian distribution. Table 4.3 shows the mean kurtosis for all images and sampling factors at $k = K_{it}$. The proximity to zero is consistent with a coloured state evolution for all image types and undersampling factors for both VDAMP-α and VDAMP-S, confirming that the difference in performance between the algorithms is not due to a breakdown in state evolution. The imaginary part, which is not included here for conciseness, was found to have a similarly small mean excess kurtosis.

Using the example of the Shepp-Logan undersampled with $N/n = 12$, Fig. 4.5 shows VDAMP-S and FISTA's $|\boldsymbol{r}_k - \boldsymbol{w}_0|$ for $k = 1, 2, 3$, visualising VDAMP-S's preservation of the unbiased subband-dependent aliasing structure shown for uniform sampling in Fig. 4.1, which contrasts with FISTA. For $k = 0, 5, 20$, Fig. 4.6 shows quantile-quantile plots against a Gaussian of the three illustrative subbands of $\boldsymbol{r}_k - \boldsymbol{w}_0$: the diagonal detail at scale 1, the horizontal detail at scale 2 and the vertical detail at scale 4, where scale 1 is the finest and scale 4 is the coarsest. The linearity of the blue points provides evidence that the per-subband effective noise is Gaussian.

The efficacy of automatic threshold selection with cSURE depends on how accurately the diagonal of the $\boldsymbol{\Sigma}_k^2$ from (4.5) is modelled by $\boldsymbol{\tau}_k$. For $k = 0, 1, \ldots, 20$, Fig. 4.7 shows the ground truth subband NMSE $\|\boldsymbol{r}_{k,b} - \boldsymbol{w}_{0,b}\|_2^2 / \|\boldsymbol{w}_{0,b}\|_2^2$ at all four scales and VDAMP's prediction. The ground-truth NMSE is closely tracked by $\boldsymbol{\tau}_k$ at all scales, which implies that parameter selection with cSURE is truly near-optimal.



(a) VDAMP's $|\boldsymbol{r}_1 - \boldsymbol{w}_0|$ (b) VDAMP's $|\boldsymbol{r}_2 - \boldsymbol{w}_0|$ (c) VDAMP's $|\boldsymbol{r}_3 - \boldsymbol{w}_0|$



(d) FISTA's $|\boldsymbol{r}_1 - \boldsymbol{w}_0|$ (e) FISTA's $|\boldsymbol{r}_2 - \boldsymbol{w}_0|$ (f) FISTA's $|\boldsymbol{r}_3 - \boldsymbol{w}_0|$

Figure 4.5: The magnitude of the aliasing of VDAMP-S and FISTA's $\boldsymbol{r}_k$ at iterations $k = 1, 2, 3$ for the $N/n = 12$ sampled Shepp-Logan, illustrating the different aliasing for the two algorithms.

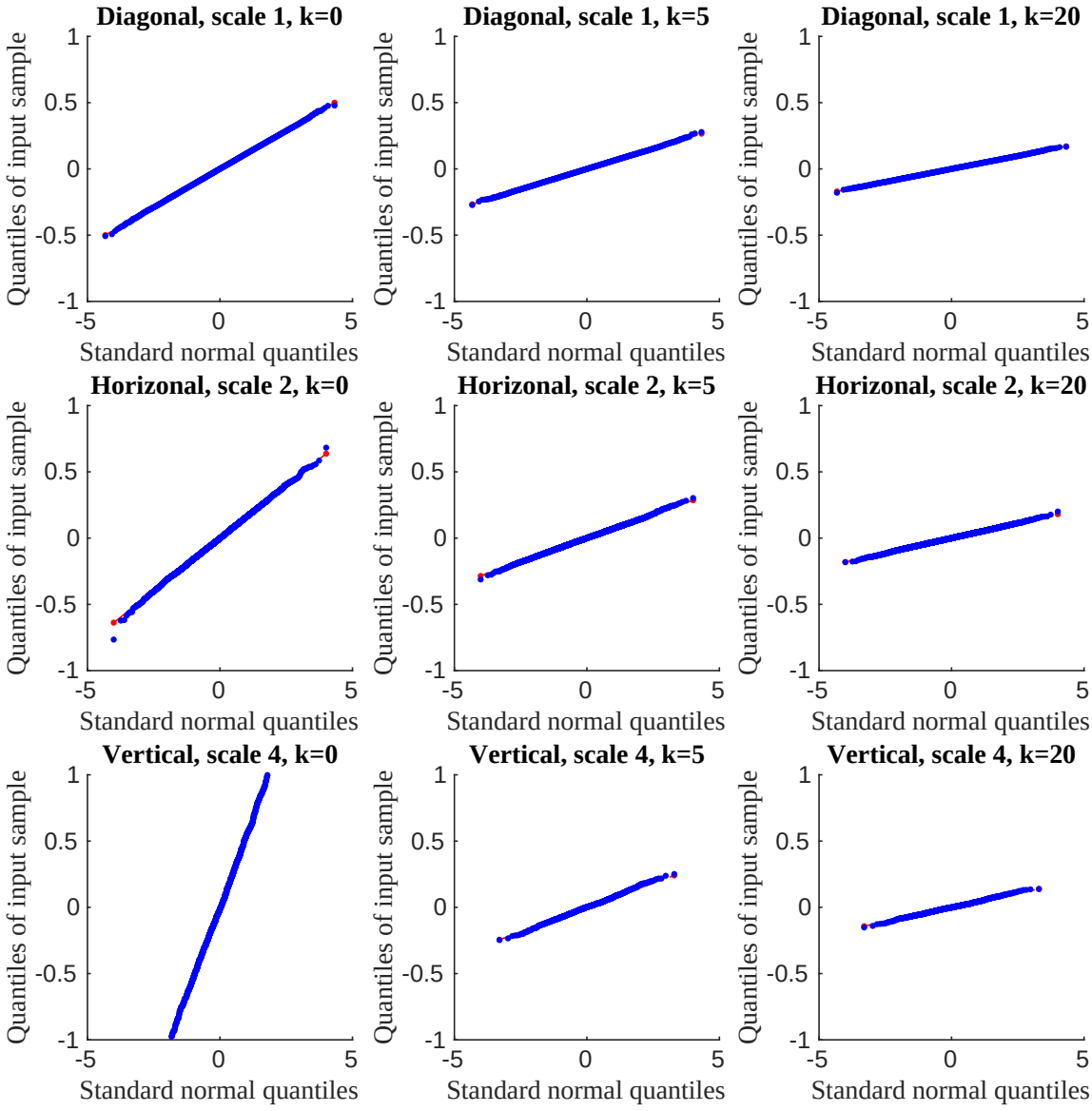


Figure 4.6: Normalised quantile-quantile plots against a Gaussian for three subbands of VDAMP-S's $\boldsymbol{r}_k - \boldsymbol{w}_0$ at $k = 0, 5, 20$ for the Shepp-Logan sampled with $N/n = 12$ in blue, and points along a straight line in red. The real part is plotted in the top and bottom rows and the imaginary is plotted in the middle row. Linearity of the blue points indicates that the data comes from a Gaussian distribution, and the decreasing gradient as k increases shows that the variance decreases with increasing k. Finite dimensional effects causing small deviations from an exact Gaussian are more apparent at coarse scales, where the dimension is smaller.

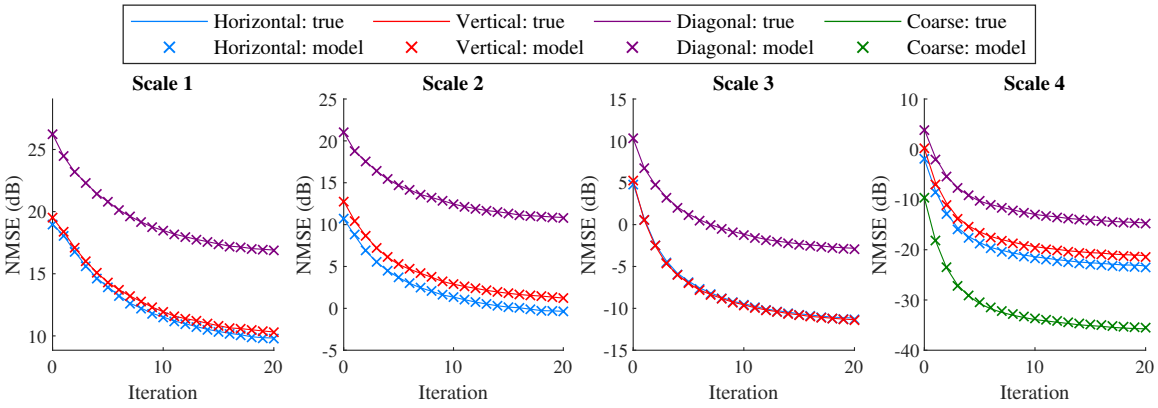


Figure 4.7: Per-subband NMSE $\|\boldsymbol{r}_{k,b} - \boldsymbol{w}_{0,b}\|_2^2 / \|\boldsymbol{w}_{0,b}\|_2^2$ versus iteration index k for a $N/n = 12$ undersampled Shepp-Logan reconstructed with VDAMP-S. Lines show the actual NMSE and crosses show the predictions from $\boldsymbol{\tau}_k$.

4.3.6 Sensitivity to measurement noise

This section evaluates the comparative performance of VDAMP for a range of measurement noise variances σ_ε^2. For this section we considered the $N/n = 12$ sampled Shepp-Logan only and varied the SNR from 5dB to 50dB in steps of 5dB. We compared with FISTA only since it was found in Section 4.3.4 that, out of the competing methods, it converged to the lowest NMSE. The NMSE at $K_{it} = 1000$ for each noise variance is shown in Table 4.4. While VDAMP-S substantially outperformed FISTA for high SNR, FISTA converged to a lower NMSE for an SNR of less than 25dB. As discussed in Section 4.2.1, this is a consequence of the density compensated measurement model (4.10), which has increased measurement noise $\boldsymbol{P}^{-1/2}\boldsymbol{M}_\Omega\boldsymbol{\varepsilon}$.

4.4 Conclusions

Based on the observation that Fourier sampling from a non-uniform spectral density leads to coloured aliasing, we proposed VDAMP, an OAMP-based algorithm that obeys a coloured state evolution (4.5). State evolution provided an effective way to automatically tune model parameters on-the-fly with cSURE, allowing more degrees of freedom in the model and implying that a single method could be used for any

image type without the need for manual adjustment. VDAMP typically converged to a similar NMSE to FISTA with an optimally-tuned scalar sparse weighting, but in around 15 times fewer iterations on average.

Chapter 5

Parallel-VDAMP

This chapter proposes the Parallel-VDAMP (P-VDAMP) algorithm, which reconstructs randomly sampled k-space data acquired across multiple receiver coils. As introduced in Section 3.2.2, and rewritten here for convenience, the data on coil c is modelled as

$$\boldsymbol{y}_c = \boldsymbol{M}_\Omega(\boldsymbol{F}\boldsymbol{S}_c\boldsymbol{x}_0 + \boldsymbol{\varepsilon}_c), \qquad c = \{1, 2, \ldots, N_c\} \tag{5.1}$$

where $\boldsymbol{S}_c \in \mathbb{C}^{N\times N}$ is a diagonal matrix known as the cth coil sensitivity and the $\boldsymbol{\varepsilon}_c$ are Gaussian but not necessarily i.i.d..

The crucial difference between P-VDAMP and single-coil VDAMP from Chapter 4 is the aliasing model update, $\boldsymbol{\tau}_k$. As demonstrated in Section 5.1, the sensitivity profiles of the receiver coils introduce intra-subband spatial modulation of the effective noise, which implies that white per-subband aliasing model employed in Chapter 4 is no longer valid, and a new formulation is required.

5.1 Multi-coil coloured aliasing

As for single-coil VDAMP, we construct P-VDAMP by qualitatively examining the aliasing of an unbiased estimate from the data, developing a method for modelling the aliasing, and constructing an algorithm based on the model. As in Chapter 4, the mathematical results in this section assume that the mask is Bernoulli random. Despite not being specifically constructed for Poisson disc sampled data, we find empirically in Section 5.3 that P-VDAMP applied to Poisson disc sampled data performs well without modification.

5.1.1 Unbiased estimator

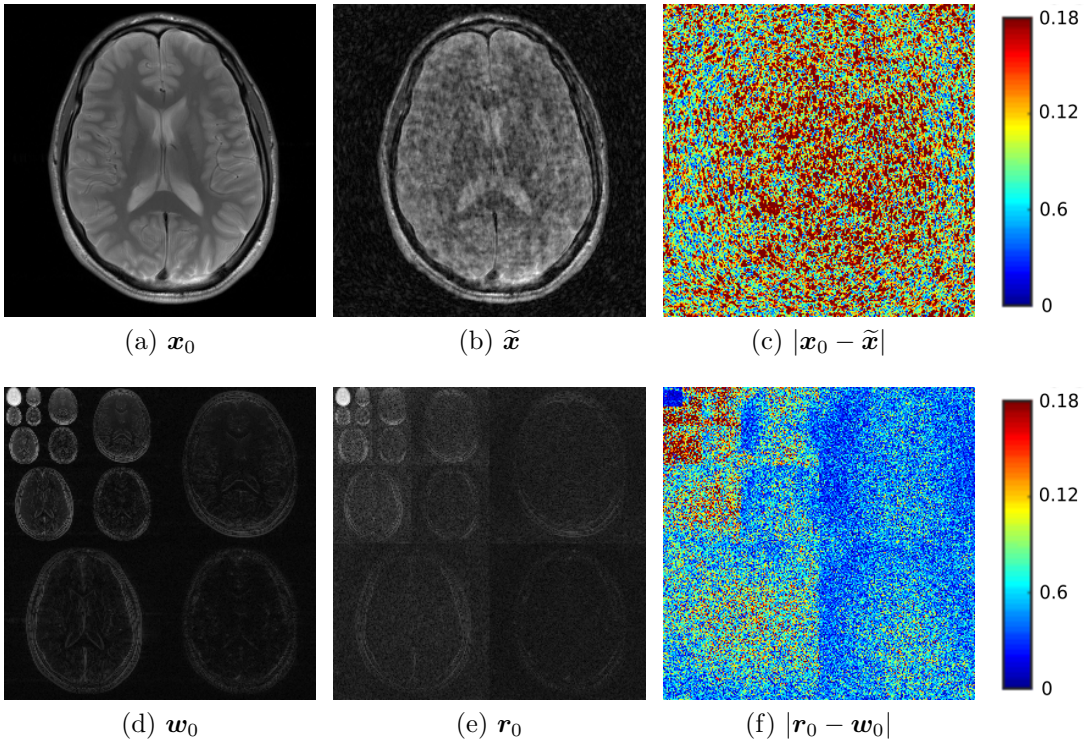
As for single-coil measurements, (4.1), the conjugate transpose of the sensing matrix applied to the density compensated data,

$$\tilde{\boldsymbol{x}} = \sum_c \boldsymbol{S}_c^H \boldsymbol{F}^H \boldsymbol{P}^{-1} \boldsymbol{y}_c, \tag{5.2}$$

is unbiased over the random mask and measurement noise:

$$\begin{aligned}
\mathbb{E}_{\Omega,\varepsilon}\{\tilde{\boldsymbol{x}}\} &= \mathbb{E}_{\Omega,\varepsilon}\{\sum_c \boldsymbol{S}_c^H \boldsymbol{F}^H \boldsymbol{P}^{-1} \boldsymbol{M}_\Omega (\boldsymbol{F}\boldsymbol{S}_c\boldsymbol{x}_0 + \boldsymbol{\varepsilon}_c)\} \\
&= \sum_c \boldsymbol{S}_c^H \boldsymbol{F}^H \boldsymbol{P}^{-1} \boldsymbol{P}\boldsymbol{F}\boldsymbol{S}_c\boldsymbol{x}_0 \\
&= \sum_c \boldsymbol{S}_c^H \boldsymbol{S}_c\boldsymbol{x}_0 = \boldsymbol{x}_0.
\end{aligned}$$

An example of $\tilde{\boldsymbol{x}}$ and $\boldsymbol{r}_0 = \boldsymbol{\Psi}\tilde{\boldsymbol{x}}$ for brain data acquired across a $N_c = 16$ channel receiver coil array is illustrated in Fig. 5.1. Here, k-space was retrospectively sampled at $N/n = 10$ using the Bernoulli sampling mask shown in Fig. 3.4a and $\boldsymbol{\Psi}$ is a order 4 Daubechies wavelet at 4 decomposition scales. It is instructive to compare Fig. 5.1 with the equivalent for single-coil measurements, Fig. 4.1. Fig. 5.1f illustrates



(a) $\boldsymbol{x}_0$ (b) $\widetilde{\boldsymbol{x}}$ (c) $|\boldsymbol{x}_0 - \widetilde{\boldsymbol{x}}|$

(d) $\boldsymbol{w}_0$ (e) $\boldsymbol{r}_0$ (f) $|\boldsymbol{r}_0 - \boldsymbol{w}_0|$

Figure 5.1: The ground truth, unbiased estimate and entry-wise absolute error of a variable density sampled 256×256 brain with $N/n = 10$, acquired on $N_c = 16$ coils. Compared with Fig. 4.1, it is clear that the white-per-subband aliasing model from the single-coil method is insufficient, and a new formulation is required.

that, unlike for single-coil measurements, the *multi-coil aliasing within each subband is not white*, which implies that a new type of aliasing model and state evolution formulation is required. The structure of the multi-coil aliasing can be understood as int*er*-subband colouring due to the non-uniformity of the signal's spectral density and int*ra*-subband modulation of the aliasing covariance due to the non-uniform coil sensitivities.

5.1.2 Multi-coil aliasing model

This section derives an approximate expression for the diagonal of the multi-coil wavelet-domain covariance of the aliasing of $\boldsymbol{r}_0 = \boldsymbol{\Psi}\widetilde{\boldsymbol{x}}$, the multi-coil equivalent of the single-coil $\boldsymbol{\tau}_k$ update described in Section 4.2.2. The $\boldsymbol{\tau}_k$ update rule derived in

this section is used in P-VDAMP once per iteration in place of line 5 of single-coil VDAMP, Algorithm 4: see Section 5.2.

Let $\boldsymbol{y}_{0,c} = \boldsymbol{F}\boldsymbol{S}_c\boldsymbol{x}_0$ be the signal of interest's coil-weighted Fourier coefficients. Using $\boldsymbol{x}_0 = \sum_c \boldsymbol{S}_c^H\boldsymbol{F}^H\boldsymbol{y}_{0,c}$, the wavelet-domain aliasing of $\boldsymbol{r}_0$ is

$$\begin{aligned}|\boldsymbol{r}_0 - \boldsymbol{w}_0|^2 =& |\boldsymbol{\Psi}(\tilde{\boldsymbol{x}} - \boldsymbol{x}_0)|^2 \\ =& |\boldsymbol{\Psi}\sum_c \boldsymbol{S}_c^H\boldsymbol{F}^H[\boldsymbol{P}^{-1}\boldsymbol{y}_c - \boldsymbol{y}_{0,c}]|^2 \\ =& |\sum_c \hat{\boldsymbol{\Psi}}^c[\boldsymbol{P}^{-1}\boldsymbol{y}_c - \boldsymbol{y}_{0,c}]|^2,\end{aligned}$$

where $\hat{\boldsymbol{\Psi}}^c = \boldsymbol{\Psi}\boldsymbol{S}_c^H\boldsymbol{F}^H$. Since $\boldsymbol{y}_c = \boldsymbol{M}_\Omega(\boldsymbol{y}_{0,c} + \boldsymbol{\varepsilon}_c)$,

$$\begin{aligned}|\sum_c \hat{\boldsymbol{\Psi}}^c[\boldsymbol{P}^{-1}\boldsymbol{y}_c - \boldsymbol{y}_{0,c}]|^2 =& |\sum_c \hat{\boldsymbol{\Psi}}^c[(\boldsymbol{P}^{-1}\boldsymbol{M}_\Omega - \mathbb{1})\boldsymbol{y}_{0,c} - \boldsymbol{P}^{-1}\boldsymbol{M}_\Omega\boldsymbol{\varepsilon}_c]|^2 \\ =& |\sum_c \hat{\boldsymbol{\Psi}}^c\boldsymbol{q}_c|^2 \qquad (5.3)\end{aligned}$$

where $\boldsymbol{q}_c = (\boldsymbol{P}^{-1}\boldsymbol{M}_\Omega - \mathbb{1})\boldsymbol{y}_{0,c} - \boldsymbol{P}^{-1}\boldsymbol{M}_\Omega\boldsymbol{\varepsilon}_c$.

Computing the expectation of (5.3) with a method analogous to the single-coil approach, described in Section 4.2.2, would involve computing and storing in memory the $N \times N$ matrices $|\hat{\boldsymbol{\Psi}}^c|^2$ for all N_c coils. Section 4.2.2 described how $|\boldsymbol{\Psi}\boldsymbol{F}^H|^2$ for the single-coil algorithm has rows that repeat for translated wavelet filters, so can be fully represented with only $3s+1$ N-dimensional vectors. The rows of the multi-coil $|\hat{\boldsymbol{\Psi}}^c|^2$ do not repeat, so the full $N \times N$ computation would be required. This also implies that multiplying $|\hat{\boldsymbol{\Psi}}^c|^2$ with $|\boldsymbol{q}_c|^2$ would be $O(N^2)$ rather than single-coil VDAMP's $O((3s+1)N)$. This method would therefore be highly demanding both in terms of memory requirements and compute time, and not practical for large N.

The remainder of this section develops an expression that approximates the expectation of (5.3) that is linear in N. To achieve this, the $\hat{\boldsymbol{\Psi}}^c$ are approximated in

terms of the unweighted spectra $\hat{\boldsymbol{\Psi}} = \boldsymbol{\Psi}\boldsymbol{F}^H$, so that the spectral density of the jth row of $\hat{\boldsymbol{\Psi}}^c$ is expressed as

$$\hat{\boldsymbol{\Psi}}_j^c \approx \xi_{c,j}\hat{\boldsymbol{\Psi}}_j, \tag{5.4}$$

where $\xi_{c,j} \in \mathbb{C}$ is a constant. Since $\hat{\boldsymbol{\Psi}}_j^c$ is the jth coil-weighted wavelet spectrum, (5.4) is interpretable as approximating the coil sensitivity as flat over the support of the jth wavelet filter, so that its spectrum is the unweighted spectrum multiplied by a filter and coil-dependent scalar. We suggest using the $\xi_{c,j}$ that minimises the Euclidean distance between the coil-weighted wavelet filter and its flat approximation,

$$\begin{aligned}
\xi_{c,j} &= \arg\min_{\xi} \|\hat{\boldsymbol{\Psi}}_j^c - \xi\hat{\boldsymbol{\Psi}}_j\|_2^2 \\
&= \arg\min_{\xi} \|\boldsymbol{\Psi}_j\boldsymbol{S}_c^H\boldsymbol{F}^H - \xi\boldsymbol{\Psi}_j\boldsymbol{F}^H\|_2^2 \\
&= \frac{\boldsymbol{\Psi}_j\boldsymbol{S}_c^H\boldsymbol{\Psi}_j^H}{\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^H} \\
&= \langle|\boldsymbol{\Psi}_j|^2, \mathrm{Diag}(\boldsymbol{S}_c^H)\rangle\,,
\end{aligned}$$

where $\boldsymbol{\Psi}_j$ is the jth row of $\boldsymbol{\Psi}$, and we have used the orthonormality of the rows of $\boldsymbol{\Psi}_j$ to set $\boldsymbol{\Psi}_j\boldsymbol{\Psi}_j^H = 1$.

Physically, (5.4) is motivated by the observation that the $\boldsymbol{S}_c$ vary smoothly in space, such as in Fig. 3.1. As well as coil smoothness, the accuracy of (5.4) depends on the choice of mother wavelet, whose support sizes vary, and the number of decomposition scales. For coarser scales, the wavelet filter has larger support, so (5.4) is expected to be a poorer assumption.

Claim 5.1 uses (5.4) to approximate the expectation of (5.3) in terms of the ground truth k-space, and Claim 5.2 shows how to estimate the aliasing in an unbiased fashion using the measurements only.

Claim 5.1. Let $\xi_{c,j} \in \mathbb{C}$ be a constant that approximates $\hat{\boldsymbol{\Psi}}_j^c \approx \xi_{c,j}\hat{\boldsymbol{\Psi}}_j$, for all j rows of $\hat{\boldsymbol{\Psi}}^c$. Then the expectation of (5.3) over the random mask and measurement noise

is

$$\mathbb{E}_{\Omega,\varepsilon}\{|r_{0,j} - w_{0,j}|^2\} \approx \boldsymbol{\xi}_j^H \left(\sum_i |\hat{\Psi}_{ji}|^2 \left[\left(\frac{1-p_i}{p_i} \right) \boldsymbol{y}_{0,i} \boldsymbol{y}_{0,i}^H + \frac{1}{p_i} \boldsymbol{\Sigma}_{\varepsilon,i}^2 \right] \right) \boldsymbol{\xi}_j, \tag{5.5}$$

where the $\boldsymbol{y}_{0,i} \in \mathbb{C}^{N_c}$ and $\boldsymbol{\xi}_{0,j} \in \mathbb{C}^{N_c}$ are comprised of the ith k-space location over all coils, having cth entry $[y_{0,i}]_c = [\boldsymbol{F}\boldsymbol{S}_c\boldsymbol{x}_0]_i$ and $[\boldsymbol{\xi}_{0,j}]_c = \xi_{c,j}$, and $\boldsymbol{\Sigma}_{\varepsilon,i}^2 = \mathbb{E}\{\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^H\}$ where $\boldsymbol{\varepsilon}_i \in \mathbb{C}^{N_c}$ similarly has cth entry $\varepsilon_{i,c} = [\boldsymbol{\varepsilon}_c]_i$.

Proof. Using (5.4), the jth entry of (5.3) is approximately

$$|r_{0,j} - w_{0,j}|^2 \approx |\sum_c \hat{\boldsymbol{\Psi}}_j \boldsymbol{q}_c \xi_{c,j}|^2 \tag{5.6a}$$

$$= |\sum_{c,i} \hat{\Psi}_{ji} q_{c,i} \xi_{c,j}|^2. \tag{5.6b}$$

Since the $q_{c,i}$ is comprised of independent entries with zero mean, by Appendix A the expectation of (5.6b) over the sampling mask and measurement noise is

$$\mathbb{E}\{|\sum_{c,i} \hat{\Psi}_{ji} q_{c,i} \xi_{c,j}|^2\} = \sum_i |\hat{\Psi}_{ji}|^2 \mathbb{E}\{|\sum_c \xi_{c,j} q_{c,i}|^2\} \tag{5.7a}$$

$$= \sum_i |\hat{\Psi}_{ji}|^2 \mathbb{E}\{|\boldsymbol{\xi}_j^H \boldsymbol{q}_i|^2\}, \tag{5.7b}$$

where $\boldsymbol{\xi}_j \in \mathbb{C}^{N_c}$ and $\boldsymbol{q}_i \in \mathbb{C}^{N_c}$. Resolving the expectation,

$$\mathbb{E}\{|\boldsymbol{\xi}_j^H \boldsymbol{q}_i|^2\} = \mathbb{E}\{\boldsymbol{\xi}_j^H \boldsymbol{q}_i \boldsymbol{q}_i^H \boldsymbol{\xi}_j\} \tag{5.8a}$$

$$= \mathbb{E}\left\{ \boldsymbol{\xi}_j^H \left[\left(\frac{m_i}{p_i} - 1 \right)^2 \boldsymbol{y}_{0,i} \boldsymbol{y}_{0,i}^H + \left(\frac{m_i}{p_i} \right)^2 \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^H \right] \boldsymbol{\xi}_j \right\}, \tag{5.8b}$$

where $\mathbb{E}\{\boldsymbol{\varepsilon_i}\} = \mathbf{0}_{N_c}$ has been applied to evaluate the cross-terms to zero. Using

$$\mathbb{E}\left\{ \left(\frac{m_i}{p_i} - 1 \right)^2 \right\} = \frac{1-p_i}{p_i}, \qquad \mathbb{E}\{\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^H\} = \boldsymbol{\Sigma}_{\varepsilon,i}^2,$$

(5.8b) is

$$\mathbb{E}\{|\boldsymbol{\xi}_j^H \boldsymbol{q}_i|^2\} = \boldsymbol{\xi}_j^H \left[\left(\frac{1-p_i}{p_i} \right) \boldsymbol{y}_{0,i} \boldsymbol{y}_{0,i}^H + \frac{1}{p_i} \boldsymbol{\Sigma}_{\varepsilon,i}^2 \right] \boldsymbol{\xi}_j. \tag{5.9}$$

Substituting (5.9) in to (5.7b) gives

$$\mathbb{E}_{\Omega,\varepsilon}\{|r_{0,j} - w_{0,j}|^2\} \approx \sum_i |\hat{\Psi}_{ji}|^2 \boldsymbol{\xi}_j^H \left[\left(\frac{1-p_i}{p_i} \right) \boldsymbol{y}_{0,i} \boldsymbol{y}_{0,i}^H + \frac{1}{p_i} \boldsymbol{\Sigma}_{\varepsilon,i}^2 \right] \boldsymbol{\xi}_j,$$

which is equal to (5.5). This completes the proof. □

In similar fashion to the single-coil method, as described in Section 4.2.2, we can estimate $\mathbb{E}_{\Omega,\varepsilon}\{|r_{0,j} - w_{0,j}|^2\}$ without requiring the ground truth. We use the update

$$\tau_j = \boldsymbol{\xi}_j^H \left(\sum_i |\hat{\Psi}_{ji}|^2 \left[\left(\frac{1-p_i}{p_i} \right) \boldsymbol{y}_i \boldsymbol{y}_i^H + \boldsymbol{\Sigma}_{\varepsilon,i}^2 \right] \frac{m_i}{p_i} \right) \boldsymbol{\xi}_j, \tag{5.10}$$

which the following proves is unbiased.

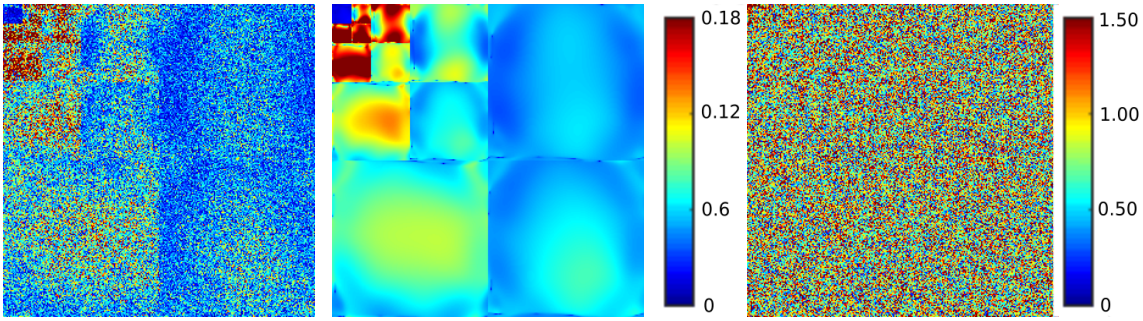
Claim 5.2. The estimate of the multi-coil wavelet-domain error stated in (5.10) is unbiased:

$$\mathbb{E}\{\tau_j\} = \boldsymbol{\xi}_j^H \left(\sum_i |\hat{\Psi}_{ji}|^2 \left[\left(\frac{1-p_i}{p_i} \right) \boldsymbol{y}_{0,i} \boldsymbol{y}_{0,i}^H + \frac{1}{p_i} \boldsymbol{\Sigma}_{\varepsilon,i}^2 \right] \right) \boldsymbol{\xi}_j. \tag{5.11}$$

Proof. Since $\boldsymbol{y}_i = m_i(\boldsymbol{y}_{0,i} + \boldsymbol{\varepsilon}_i)$, and using $\mathbb{E}\{\boldsymbol{\varepsilon}_i\} = \mathbf{0}_{N_c}$,

$$\begin{aligned}
&\mathbb{E}_{\Omega,\varepsilon} \left\{ \left[\left(\frac{1-p_i}{p_i} \right) \boldsymbol{y}_i \boldsymbol{y}_i^H + \boldsymbol{\Sigma}_{\varepsilon,i}^2 \right] \frac{m_i}{p_i} \right\} \\
=&\mathbb{E}_{\Omega,\varepsilon} \left\{ \left[m_i \left(\frac{1-p_i}{p_i} \right) (\boldsymbol{y}_{0,i} \boldsymbol{y}_{0,i}^H + \boldsymbol{\varepsilon}_i \boldsymbol{y}_{0,i}^H + \boldsymbol{y}_{0,i} \boldsymbol{\varepsilon}_i^H + \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^H) + \boldsymbol{\Sigma}_{\varepsilon,i}^2 \right] \frac{m_i}{p_i} \right\} \\
=& \left(\frac{1-p_i}{p_i} \right) \boldsymbol{y}_{0,i} \boldsymbol{y}_{0,i}^H + \frac{1}{p_i} \boldsymbol{\Sigma}_{\varepsilon,i}^2
\end{aligned} \tag{5.12}$$

Substituting (5.12) in (5.10) gives the right-hand-side of (5.11). □



(a) $|\boldsymbol{r}_0 - \boldsymbol{w}_0|$ (b) $\sqrt{\boldsymbol{\tau}_0}$ (c) $|\boldsymbol{r}_0 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_0}$

Figure 5.2: Error propagation with (5.10), using the same data and sampling as Fig. 5.1. The qualitative whiteness of (c) indicates that $\mathrm{Diag}(\boldsymbol{\tau}_0)$ is a reasonably accurate model of the true aliasing covariance.

To compute (5.10) is linear in N, so is scalable to large problem sizes. Further, it is only necessary to store in memory the $3s + 1$ N-dimensional unique rows of $|\hat{\boldsymbol{\Psi}}|^2$, as for single-coil VDAMP, and $\xi_{c,j}$, of which there are $N \cdot N_c$. Since it requires the computation of $\boldsymbol{y}_i \boldsymbol{y}_i^H$, (5.10) is quadratic in N_c, so especially benefits from coil compression.

Note that for the single-coil case with flat coil sensitivity, where $N_c = 1$, $\boldsymbol{S}_1 = \mathbb{1}_N$, $\boldsymbol{\Sigma}^2_{\varepsilon,i} = \sigma^2_\varepsilon$ and $\boldsymbol{\xi}_j = 1$, (5.4) is exact and (5.10) reduces to the single-coil aliasing model update, line 5 of Algorithm 4.

5.1.3 Parallel coloured state evolution

Fig. 5.2 shows $\boldsymbol{\tau}_0$ calculated according to (5.10) for the same brain data and sampling as Fig. 5.1. The NMSE estimate from error propagation, $\langle \boldsymbol{\tau} \rangle / \langle |\boldsymbol{x}_0|^2 \rangle = -11.67\text{dB}$, closely predicts the true NMSE, $\langle |\boldsymbol{x}_0 - \tilde{\boldsymbol{x}}|^2 \rangle / \langle |\boldsymbol{x}_0|^2 \rangle = -11.70\text{dB}$. Also shown is the aliasing of $\boldsymbol{r}_0$ divided entry wise by the error model, which resembles white Gaussian noise, qualitatively suggesting that $\boldsymbol{r}_0 \approx \boldsymbol{w}_0 + \mathcal{CN}(\boldsymbol{0}_N, \mathrm{Diag}(\boldsymbol{\tau}_0))$ is a reasonable model for the aliasing. For the coarsest scale, especially the approximation coefficients, there is a visible deviation from white noise, emphasising the poorer $\hat{\boldsymbol{\Psi}}^c_j \approx \xi_{c,j} \hat{\boldsymbol{\Psi}}_j$ approximation for filters with larger support. However, since the centre of k-space

is sampled with high density, the error of the coarsest scale is small, as evident in Fig. 5.2a. This implies that the algorithm's denoiser plays a less important role in reconstructing this region, so inaccuracy in the aliasing, which is used to guide the denoiser, is less crucial than for other subbands.

In the remainder of this chapter, we present P-VDAMP, Algorithm 6. In Section 5.3.7 we present empirical evidence that P-VDAMP retains zero-mean Gaussian aliasing with a wavelet-diagonal covariance for all iterations,

$$\boldsymbol{r}_k = \boldsymbol{w}_0 + \mathcal{CN}(\boldsymbol{0}_N, \boldsymbol{\Sigma}_k^2), \tag{5.13}$$

where $\boldsymbol{\Sigma}_k^2$ can be accurately estimated with $\mathrm{Diag}(\boldsymbol{\tau}_k)$, where $\boldsymbol{\tau}_k$ is updated according to (5.10). We term (5.13) P-VDAMP's *parallel coloured state evolution*.

5.2 Description of P-VDAMP algorithm

P-VDAMP is stated Algorithm 6. Lines 3-4 are density compensated gradient descent, the multi-coil analogue to lines 3-4 of single-coil VDAMP, Algorithm 4. In line 5, the error propagation formula (5.10) is applied to update $\boldsymbol{\tau}_k$ with the residual $\boldsymbol{z}_k$ in place of $\boldsymbol{y}$, as in line 5 of Algorithm 4.

Line 6 applies a denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$ to $\boldsymbol{r}_k$. As in Chapter 4, detailed in Section 4.2.3, we used the aliasing model $\boldsymbol{\tau}_k$ to employ SURE-optimal per-subband soft thresholding for $\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$ throughout this chapter. The possibility of using a CNN for the denoiser is the topic of Chapter 6.

Lines 6-9 is per-subband Onsager corrected denoising, identical to Algorithm 4. For single-coil VDAMP, per-subband Onsager correction arose naturally as a consequence of the white-per-subband aliasing model. For P-VDAMP, since the per-subband aliasing is not white, the use of per-subband Onsager correction is less intuitive. Nonetheless, we show empirically in Section 5.3.7 that such an Onsager

Algorithm 6 P-VDAMP

Require: Sampling set Ω, coil sensitivities $\boldsymbol{S}_c$, wavelet transform $\boldsymbol{\Psi}$, probability matrix $\boldsymbol{P}$, measurements $\boldsymbol{y}_c$, measurement noise covariance $\boldsymbol{\Sigma}_\varepsilon^2$, denoiser $\boldsymbol{g}(\boldsymbol{r};\boldsymbol{\tau})$, number of iterations K_{it}.

1: Compute $|\hat{\boldsymbol{\Psi}}|^2 = |\boldsymbol{\Psi}\boldsymbol{F}^H|^2$ and $\xi_{c,j} = \langle |\boldsymbol{\Psi}_j|^2, \mathrm{Diag}(\boldsymbol{S}_c^H)\rangle$ and set $\widetilde{\boldsymbol{r}}_0 = \boldsymbol{0}_N$
2: **for** $k = 0, 1, \ldots, K_{it} - 1$ **do**
3: $\quad \forall c, \boldsymbol{z}_{k,c} = \boldsymbol{y}_c - \boldsymbol{M}_\Omega \boldsymbol{F}\boldsymbol{S}_c\boldsymbol{\Psi}^H\widetilde{\boldsymbol{r}}_k$
4: $\quad \boldsymbol{r}_k = \widetilde{\boldsymbol{r}}_k + \boldsymbol{\Psi}\sum_c \boldsymbol{S}_c^H\boldsymbol{F}^H\boldsymbol{P}^{-1}\boldsymbol{z}_{k,c}$
5: $\quad \forall j, \tau_{k,j} = \boldsymbol{\xi}_j^H\left(\sum_i |\hat{\Psi}_{ji}|^2\left[\left(\frac{1-p_i}{p_i}\right)\boldsymbol{z}_{k,i}\boldsymbol{z}_{k,i}^H + \boldsymbol{\Sigma}_{\varepsilon,i}^2\right]\frac{m_i}{p_i}\right)\boldsymbol{\xi}_j$
6: $\quad \hat{\boldsymbol{w}}_k = \boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k)$
7: $\quad \boldsymbol{\alpha}_k = \langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k))\rangle_{\text{sband}}$
8: $\quad$ Update $\boldsymbol{c}_k$ e.g. $\boldsymbol{c}_k = \boldsymbol{1}_N \oslash (\boldsymbol{1}_N - \boldsymbol{\alpha}_k)$
9: $\quad \widetilde{\boldsymbol{r}}_{k+1} = \boldsymbol{c}_k \odot (\hat{\boldsymbol{w}}_k - \boldsymbol{\alpha}_k \odot \boldsymbol{r}_k)$
10: **end for**
11: **return** $\hat{\boldsymbol{x}} = \boldsymbol{\Psi}^H\hat{\boldsymbol{w}}_k + \sum_c \boldsymbol{S}_c^H\boldsymbol{F}^H(\boldsymbol{y}_c - \boldsymbol{M}_\Omega\boldsymbol{F}\boldsymbol{S}_c\boldsymbol{\Psi}^H\hat{\boldsymbol{w}}_k)$ or $\hat{\boldsymbol{x}} = \boldsymbol{\Psi}^H\boldsymbol{r}_k$

correction leads to parallel coloured state evolution. A thorough theoretical justification for the form of the Onsager correction is left as future work.

Line 11 of Algorithm 6 suggests two possible options for the algorithm output. The first is gradient descent without density compensation; the multi-coil analogue to line 11 of Algorithm 4. We also suggest $\boldsymbol{\Psi}^H\boldsymbol{r}_k$, which is unbiased when parallel coloured state evolution (5.13) holds. Section 5.3 compares these possibilities quantitatively and qualitatively.

5.2.1 Comments on P-VDAMP in practice

We have found that damping the denoiser by a constant factor $0 < \rho \leq 1$ aids convergence. For $k > 0$, we suggest replacing lines 6 and 7 of Algorithm 6 with

$$\begin{aligned} \hat{\boldsymbol{w}}_k &= \rho\boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k) + (1-\rho)\hat{\boldsymbol{w}}_{k-1} \\ \boldsymbol{\alpha}_k &= \rho\,\langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k))\rangle_{\text{sband}}\,. \end{aligned} \tag{5.14}$$

Note that $\rho = 1$ corresponds to zero damping. In all experiments in this chapter, the damping factor was set to $\rho = 0.75$ unless stated otherwise. The effect of damping

on P-VDAMP's state evolution is discussed in Section 5.3.7.

The single-coil VDAMP algorithm of Chapter 4 was run for a fixed number of iterations. Here, we use a more practical stopping criterion based on the aliasing model $\boldsymbol{\tau}_k$, following the suggestion in [178]. In the experiments in this chapter, VDAMP stops iterating when the mean-squared of the aliasing of $\boldsymbol{r}_k$ is predicted to have increased, $\langle \boldsymbol{\tau}_k \rangle > \langle \boldsymbol{\tau}_{k-1} \rangle$, or when $\langle \boldsymbol{\tau}_k \rangle$ changes by a small amount: $|\langle \boldsymbol{\tau}_k \rangle - \langle \boldsymbol{\tau}_{k-1} \rangle| / \langle \boldsymbol{\tau}_{k-1} \rangle < \epsilon$. We used $\epsilon = 10^{-3}$ throughout this chapter.

5.3 Numerical experiments

5.3.1 Description of data

We evaluated the performance of P-VDAMP on three types of MRI data:

- **Brain.** 256×256 acquisitions acquired on $N_c = 16$ coils, from an on-site 3T scanner.
- **Knee.** 368×640 acquisitions acquired on $N_c = 15$ coils, from the publicly available fastMRI dataset [196], which is comprised of a mixture of field strengths.
- **Brain angiography.** $640 \times 506 \times 56$ acquisitions acquired on $N_c = 32$ coils, from an on-site 7T scanner.

For each of the three categories, we had two fully sampled datasets, which are referred to as, for instance, *Brain A* and *Brain B*. The two knee datasets were selected randomly from the fastMRI dataset.

The aliasing model update in line 5 of Algorithm 6 assumes that k-space is Bernoulli random, and does not accurately model the aliasing for sampling patterns that include smooth trajectories. P-VDAMP is therefore only applicable to 3D data that has been decomposed in to a series of decoupled 2D subproblems, as described

in Section 3.4. The Bernoulli and Poisson disc sampling that we employ in this chapter for the 2D brain and knee data would therefore not be possible for actual, prospectively undersampled scans. Nonetheless, such a setting is instructive for evaluating algorithm performance as it is representative of a 2D subproblem that could be encountered as part of a 3D problem.

For the brain angiographies, we considered the entirely realistic case where a 1D inverse Fourier transform was applied in the readout direction and reconstructed slice-by-slice. The readout of the angiographies was in the first dimension, so that the 3D reconstruction consisted of 640 instances of 506×56 decoupled 2D reconstructions.

5.3.2 Description of experimental method

All datasets were retrospectively undersampled at factors $N/n = 5, 10$ with both Bernoulli and Poisson disc masks, as in the 256×256 examples of Fig. 3.4. Standard algorithms for generating variable density Poisson disc masks, such as [129], define a minimum distance between sampled points without explicitly being based on a variable density probability distribution. Therefore, to estimate $\boldsymbol{P} = \mathbb{E}\{\boldsymbol{M}_\Omega\}$, we computed the empirical mean of 10^5 instances of the Poisson disc mask. So that the Poisson disc and Bernoulli masks could be fairly compared, the Bernoulli sampled masks were generated using the same $\boldsymbol{P}$. All masks were fully sampled in a 24×24 central autocalibration region, with $p_j = 1$ for all j in that region, which was used to estimate the coil sensitivities $\boldsymbol{S}_c$ with ESPIRiT [1]. To reduce the per-iteration compute time, the coils were compressed to $N_c = 8$ virtual coils.

The reference image was computed with $\boldsymbol{x}_{\text{full}} = \sum_c \boldsymbol{S}_c^H \boldsymbol{F}^H \boldsymbol{y}_{\text{full},c}$, where $\boldsymbol{y}_{\text{full},c} = \boldsymbol{F}\boldsymbol{S}_c\boldsymbol{x}_0 + \boldsymbol{\varepsilon}_c$ is the noisy, fully sampled data on coil c. Note that since actual MRI data is used here, unlike the numerical experiments in Chapter 4, we do not have access to $\boldsymbol{y}_{0,c}$, the k-space data free of measurement noise, so cannot compute the ground truth $\boldsymbol{x}_0$. However, since the measurement noise variance is usually smaller

than the reconstruction MSE, the quantitative performance measures based on $\boldsymbol{x}_{\text{full}}$ are a reasonably accurate estimation of the truth.

Unfortunately, the data used here was not acquired with pre-scans for measurement noise estimation, so we were not able to predict $\boldsymbol{\Sigma}_{\varepsilon}^2$. We set P-VDAMP's measurement noise covariance estimate $\boldsymbol{\Sigma}_{\varepsilon}^2 = \mathbf{0}_{N\times N}$, where $\mathbf{0}_{N\times N}$ is the $N \times N$ matrix of zeroes, although we emphasise that for actual, prospectively undersampled scans, it would likely be beneficial to include an estimate of $\boldsymbol{\Sigma}_{\varepsilon}^2$.

5.3.3 Comparative algorithms

We evaluated the reconstruction quality of both output options in line 11 of Algorithm 6, and refer to the $\boldsymbol{\Psi}^H \boldsymbol{r}_k$ output as *Unbiased* P-VDAMP, and the alternative as simply P-VDAMP.

In Chapter 4, two $\boldsymbol{c}_k$ updates were considered: $\boldsymbol{c}_k^{\alpha}$ in (4.19) and $\boldsymbol{c}_k^{SURE}$ in (4.20). For single-coil measurements, $\boldsymbol{c}_k^{SURE}$ is SURE-optimal with respect to $\widetilde{\boldsymbol{r}}_k$ due to the white-per-subband aliasing model, which ensures that $\boldsymbol{\tau}_k^H \boldsymbol{\partial}(\widetilde{\boldsymbol{g}}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)) = 0$ and SURE optimisation reduces to an ℓ_2 minimisation problem with a closed form solution, as stated in (4.21). For multiple coils, the intra-subband modulation of the aliasing implies that $\boldsymbol{\tau}_k^H \boldsymbol{\partial}(\widetilde{\boldsymbol{g}}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)) \neq 0$ in general, and the computation of a SURE-optimal $\boldsymbol{c}_k$ update is less straight-forward. Therefore, for simplicity, we consider only the $\boldsymbol{c}_k^{\alpha} = \mathbf{1}_N \oslash (\mathbf{1}_N - \boldsymbol{\alpha}_k)$ version of the algorithm in this chapter. This choice is further motivated by the the empirical NMSE results from the single-coil method, Table 4.2, which found that VDAMP-α mostly outperformed VDAMP-S on data types other than the idealistic Shepp-Logan.

We compared P-VDAMP with three variations of multi-coil FISTA, Algorithm 5 with the gradient descent on lines 3 and 4 modified for multiple coils. Firstly, we implemented FISTA with a hand-selected λ. Unlike Chapter 4, where λ was tuned to near-optimality using the ground truth MSE on a case-by-case basis, here we tuned

λ less precisely to better reflect the tuning for actual, prospectively undersampled scans. The λ was tuned to be approximately MSE optimal for a specific case, and this was shared across a number of reconstructions, so that the reconstruction quality was suboptimal for reconstructions other than those for which λ was specifically tuned for. For the knee and brain data we used the λ that was approximately MSE-optimal for the Brain A data Bernoulli sampled at $N/n = 10$. For the angiographies, this tuning was performed on the central slice of Angiography A, also for Bernoulli sampling at $N/n = 10$.

Since it was consistently outperformed by FISTA in Section 4.3, we did not run S-FISTA [190] here. For comparative parameter-free methods, we ran SURE-IT, as detailed in Section 4.3, and a recently proposed parameter-free variant of FISTA [197], which automatically selects a per-scale sparse weighting heuristically, based on k-means clustering of the zero-filled estimate: see [197] for details. The latter is referred to as Automatic-FISTA (A-FISTA) throughout this chapter.

For FISTA and SURE-IT's MSE estimate in line 5 of Algorithm 5, we used the true error with respect to the reference, as in Chapter 4. The stopping criterion employed for FISTA and SURE-IT was analogous to P-VDAMP, stated in Section 5.2.1: when $\tau_k > \tau_{k-1}$, or when $|\tau_k - \tau_{k-1}|/\tau_{k-1} < 10^{-3}$. For A-FISTA, we found that such a stopping criterion typically terminated the algorithm prematurely, so instead used a fixed number of iterations. As in [197], we chose 200 iterations, which we found to be sufficient to ensure convergence.

For the wavelet transform $\mathbf{\Psi}$, we used the order 4 Daubechies mother wavelet with $s = 4$ decompositions scales for all algorithms, which is widely used in MRI reconstruction, such as in [9,197,198]. Since its mother wavelet is larger, P-VDAMP's aliasing model with order 4 Daubechies wavelet (5.10) is less accurate than it would be for Haar. However, we found that in practice the improved compressibility of the order 4 Daubechies wavelet coefficients was sufficient to improve the reconstruction

quality compared with Haar, despite its poorer aliasing model.

5.3.4 Performance metrics

The metric of utmost importance in MRI reconstruction is the diagnostic quality according to radiologists [199]. To reflect this, we extended the analysis employed in Chapter 4 to include simple post-reconstruction, pre-analysis processing and introduced additional metrics that are known to correlate with perceived quality.

To mitigate the effect of the background on the reconstruction quality metrics, we masked the reference $\boldsymbol{x}_{\text{full}}$ and image estimate $\hat{\boldsymbol{x}}$. For the structural brain and knee datasets, all pixels with absolute value less than 5% of the maximum absolute value of the reference were set to zero. For the angiographies we used a hand-selected mask that excluded both the background and the skull, a common procedure referred to in the MRI literature as "skull-stripping" [200]. All metrics for the angiographies were computed with respect to the maximum projection along the third dimension, which is a standard visualisation method [201, 202].

As in Chapter 4, we computed the NMSE and time to convergence, defined as the time to reach the $\boldsymbol{\tau}_k$-based stopping criterion described in Section 5.2.1 or, for A-FISTA, the time to 200 iterations. We also used the Structural Similarity Index Metric (SSIM) [203], a widely used metric that aims to predict the perceived change in structural information of an image, which returns a score between 0 and 1, where 1 is exact recovery. For the angiographies, we computed the SSIM after applying a vessel mask computed by thresholding with Otsu's method [204], which was shown in [199] to correlate well with radiologists' perceived reconstruction quality. The final metric we employed was the High-Frequency Error Norm (HFEN) [140], which quantifies the reconstruction quality of the image's fine features. Following [140], the HFEN was computed by applying a 15×15 Laplacian of Gaussian filter with standard deviation 1.5 to $\hat{\boldsymbol{x}} - \boldsymbol{x}_0$ and computing the ℓ_2 norm. We normalised the HFEN by dividing by

the ℓ_2 norm of the reference image.

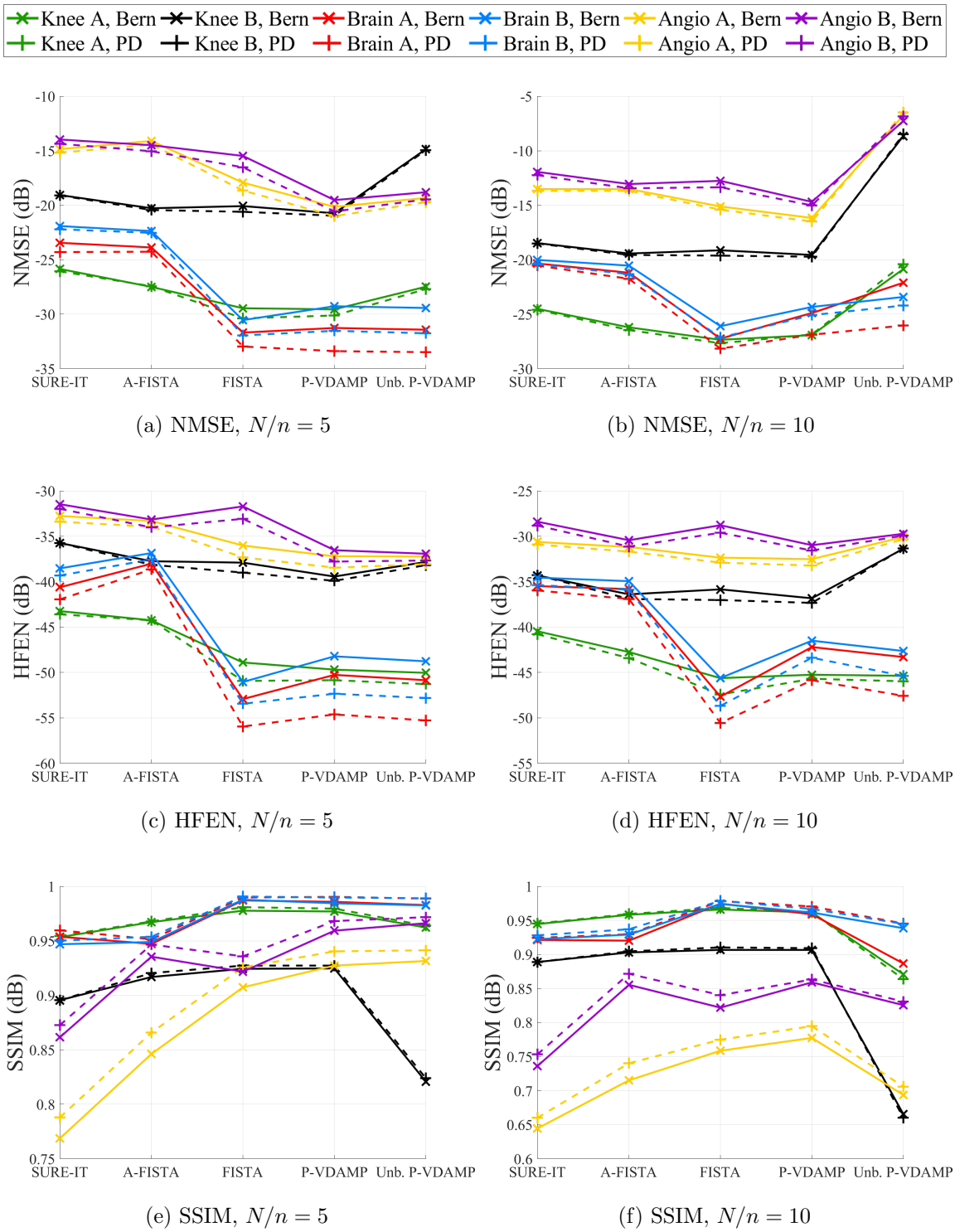
5.3.5 Quantitative discussion of reconstruction quality

The NMSE, HFEN and SSIM of each is algorithm is shown in Fig. 5.3. Despite P-VDAMP's aliasing model not being designed for Poisson disc sampled data, its comparative performance is similar to the Bernoulli sampled case. The aliasing model mismatch for Poisson disc sampled data is discussed further in Section 5.3.7.

P-VDAMP offered improved NMSE, SSIM and HFEN compared with SURE-IT for all reconstructions, highlighting the need for Gaussian aliasing for effective threshold selection with SURE, as for the single-coil method from Chapter 4. A-FISTA performs well in some instances, such as the $N/n = 10$ sampled Knee B, but is more often closer to the performance of SURE-IT than FISTA or P-VDAMP.

Compared with FISTA, the hierarchy appears to be case specific: FISTA generally outperforms P-VDAMP for the brain data, for the knee data it is more mixed, while for the angiographies P-VDAMP usually offers the best reconstruction. This does not necessarily imply that the comparative performance of P-VDAMP is anatomy-specific, as FISTA's performance depends on how well its sparse weighting is tuned. For instance, the $N/n = 10$ Bernoulli sampled Brain A, which was the case used to tune FISTA's sparse weighting, is an example where FISTA outperforms P-VDAMP by a greater margin. Since the reference MSE is not available for parameter tuning in actual, prospectively undersampled scans, FISTA's imperfectly tuned sparse weighting for reconstructions other than the $N/n = 10$ Bernoulli sampled Brain A is a reasonable reflection of reality.

Unbiased P-VDAMP often performs poorly in terms of NMSE and SSIM, but reasonably well in terms of HFEN. In the subsequent section, we show that even in the case of relatively poor NMSE and SSIM, Unbiased P-VDAMP performs well qualitatively.



(a) NMSE, $N/n = 5$

(b) NMSE, $N/n = 10$

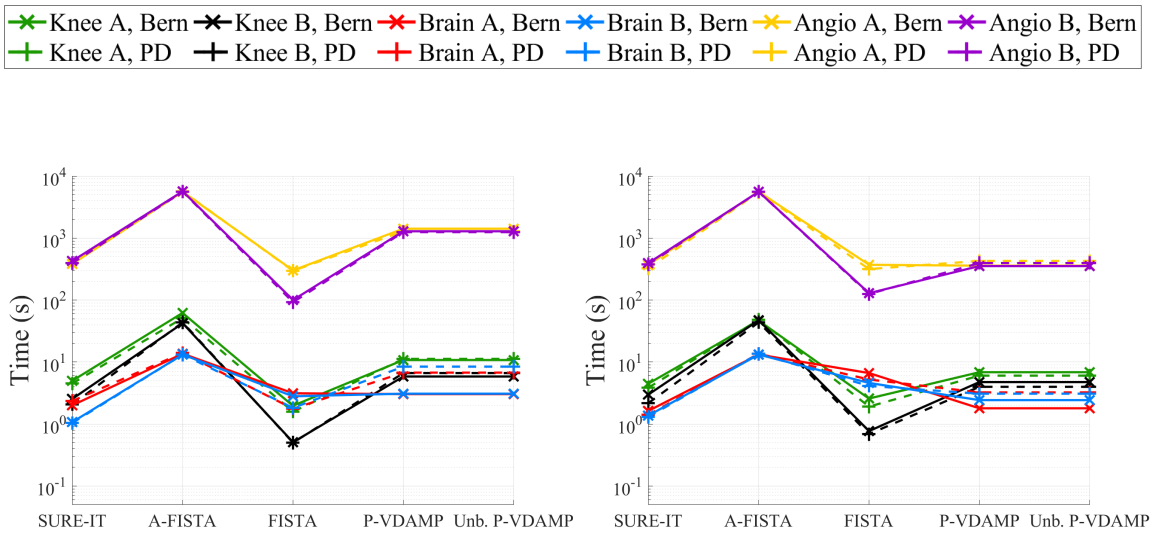
(c) HFEN, $N/n = 5$

(d) HFEN, $N/n = 10$

(e) SSIM, $N/n = 5$

(f) SSIM, $N/n = 10$

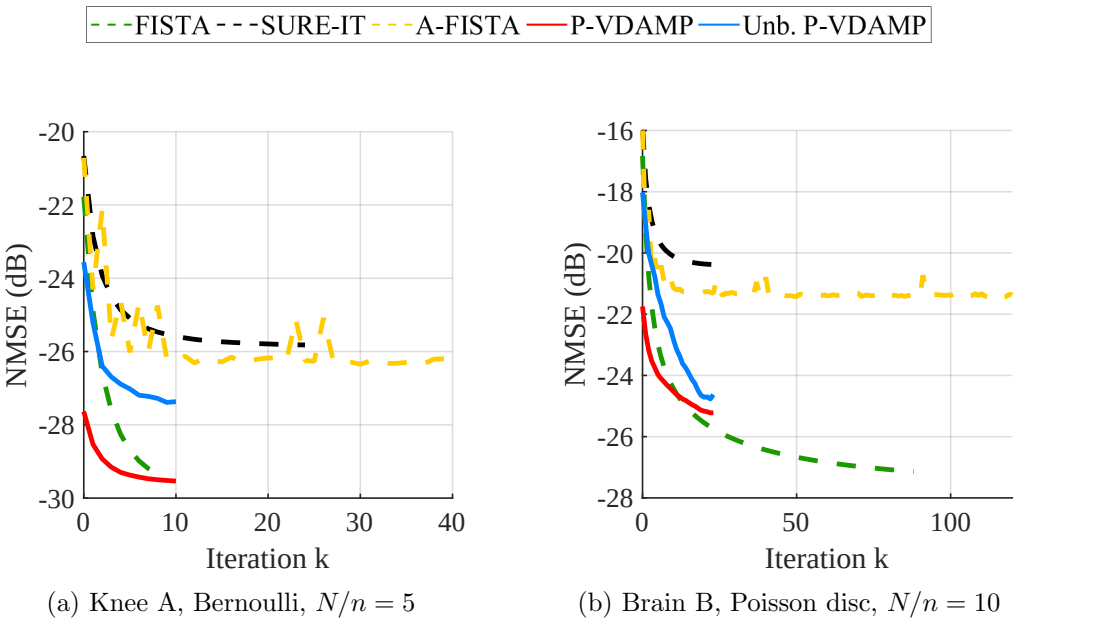
Figure 5.3: NMSE, HFEN and SSIM of all algorithms at $N/n = 5, 10$ for Bernoulli and Poisson disc sampling



(a) Time to convergence, $N/n = 5$

(b) Time to convergence, $N/n = 10$

Figure 5.4: Time to convergence, $N/n = 5, 10$



(a) Knee A, Bernoulli, $N/n = 5$

(b) Brain B, Poisson disc, $N/n = 10$

Figure 5.5: NMSE vs iteration for Bernoulli sampled Knee A and Poisson disc sampled Brain B, both at $N/n = 10$.

The time to convergence is shown in Fig. 5.4. Unlike single-coil VDAMP, P-VDAMP does not demonstrate a consistent reduction in the required compute time compared with FISTA. This is primarily due to the more computationally demanding $\boldsymbol{\tau}_k$ update, line 5 of Algorithm 6, which leads to around a 4 times longer per-iteration

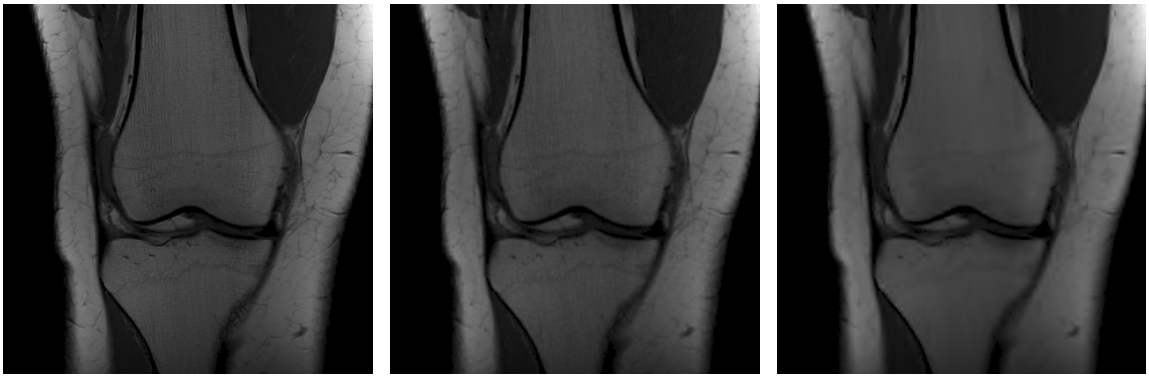
compute time than FISTA in our MATLAB implementation. For the $N/n = 5$ Bernoulli sampled Knee B and the $N/n = 10$ Poisson disc sampled Brain B, Fig. 5.5 shows the NMSE vs iteration. For the Knee A example, P-VDAMP converges in slightly more iterations than FISTA, while for the Brain B example, P-VDAMP converges in around 5 times fewer iterations than FISTA. Fig. 5.5 demonstrates why the $\boldsymbol{\tau}_k$-based stopping criterion stopped the algorithm prematurely; A-FISTA's NMSE does not decrease monotonically. Its longer reconstruction time is due to the lack of a better stopping criterion than running it for 200 iterations; A-FISTA has a similar per-iteration cost to FISTA.

5.3.6 Qualitative discussion of reconstruction quality

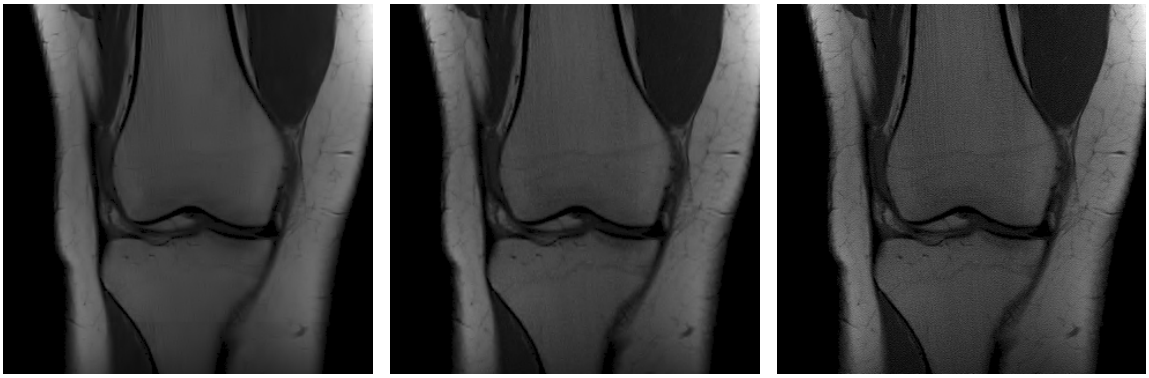
This section qualitatively discusses the reconstruction for one illustrative example from each data type. The first example, shown in Fig. 5.6, is the $N/n = 10$ Bernoulli sampled Knee A, where FISTA outperformed both P-VDAMP outputs by all reconstruction quality metrics. FISTA and both P-VDAMP estimates retain considerably more of the fine detail. Note that this is true for Unbiased P-VDAMP, despite having a far worse NMSE than all competing methods. For the competing parameter-free methods SURE-IT and A-FISTA, there is clear oversmoothing, leading to a loss of fine detail and indicating that A-FISTA's heuristic automatic tuning is sub-optimal.

The $N/n = 10$ Poisson disc sampled Brain A is shown in Fig. 5.7. As for the knee example, both parameter-free FISTA-based methods exhibit significant blurring, while FISTA and P-VDAMP retain considerably more anatomical detail. Unbiased P-VDAMP exhibits some high-frequency texturing in the aliasing not seen in P-VDAMP or FISTA.

Fig. 5.8 shows the fully sampled Angiography B with its reconstructions for $N/n = 5$ Poisson disc sampled data. A 100×100 zoomed-in region is also shown, with contrast enhancement. The green arrow highlights a narrow region of a blood

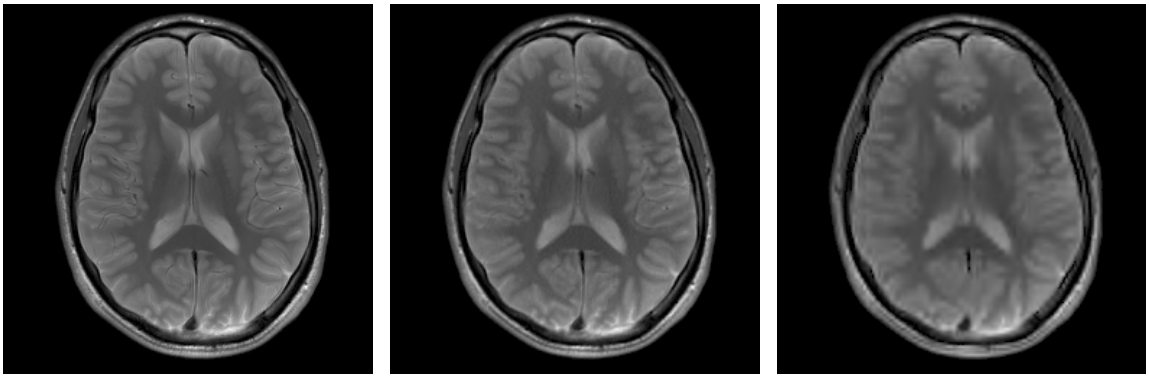


(a) Reference (b) FISTA (c) SURE-IT

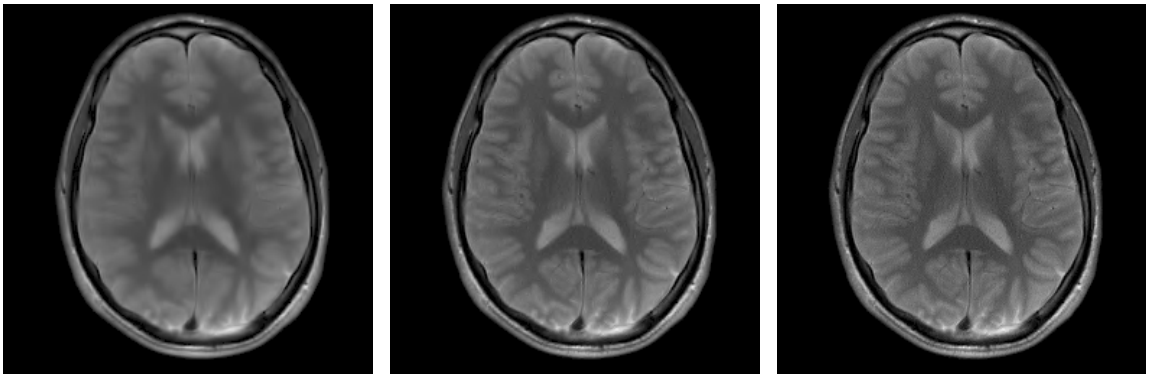


(d) A-FISTA (e) P-VDAMP (f) Unbiased P-VDAMP

Figure 5.6: Reconstructions of Knee A Bernoulli undersampled at $N/n = 10$.

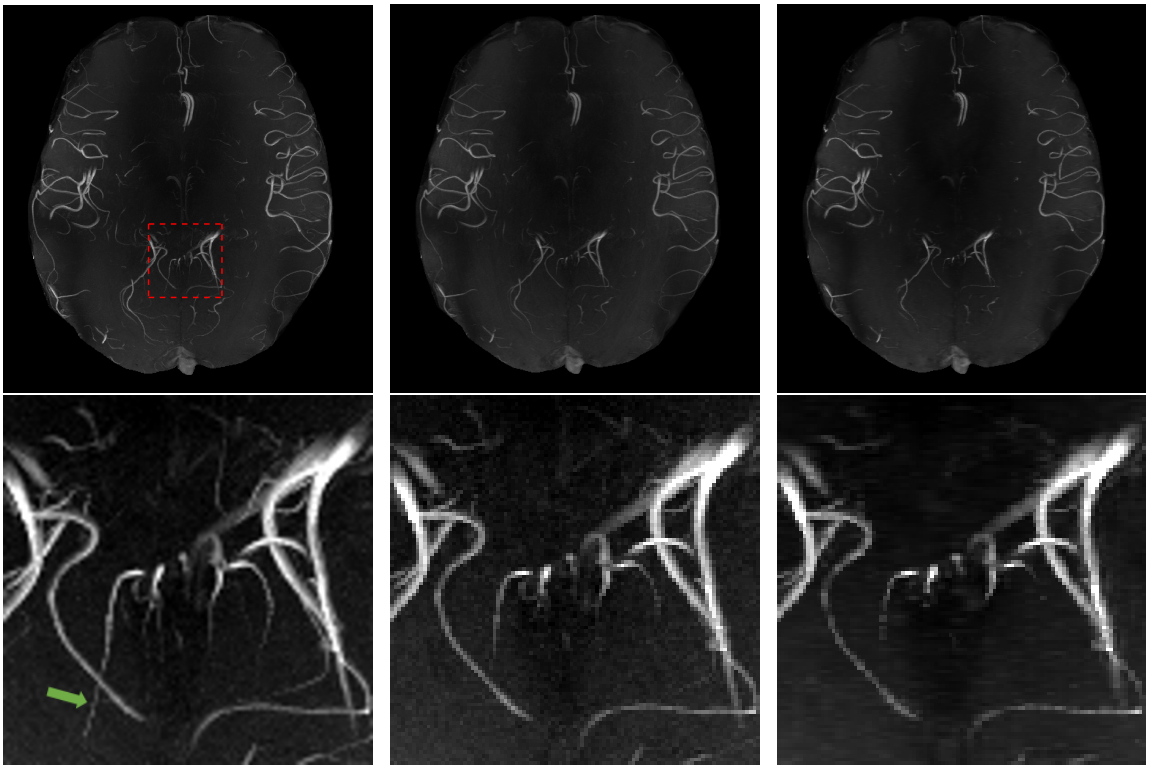


(a) Reference (b) FISTA (c) SURE-IT

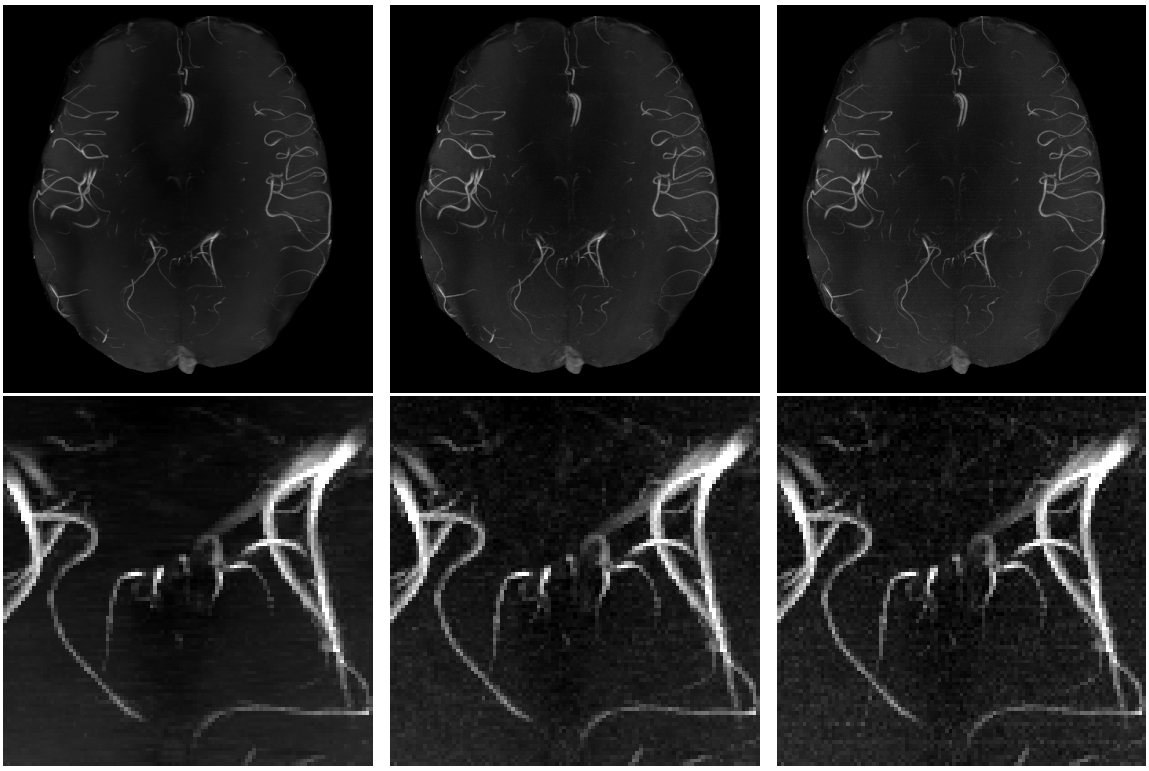


(d) A-FISTA (e) P-VDAMP (f) Unbiased P-VDAMP

Figure 5.7: Reconstructions of Brain A Poisson disc undersampled at $N/n = 10$.

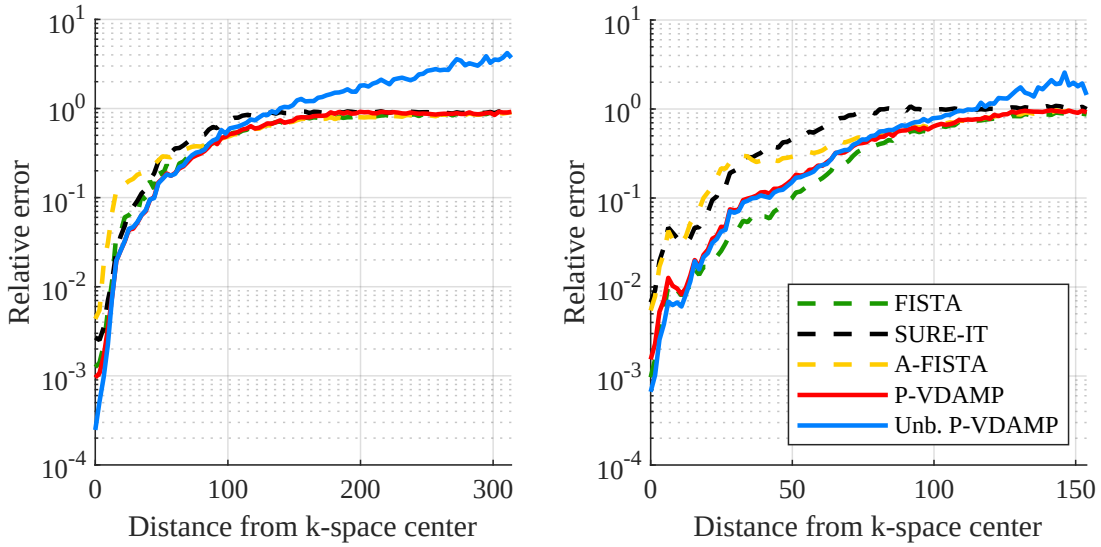


(a) Reference (b) FISTA (c) SURE-IT



(d) A-FISTA (e) P-VDAMP (f) Unbiased P-VDAMP

Figure 5.8: Angiography B Poisson disc undersampled at $N/n = 5$, with a contrast-enhanced 100×100 region. The green arrow shows part of a blood vessel which is retained only for P-VDAMP.



(a) Knee A, Bernoulli, $N/n = 10$

(b) Brain A, Poisson disc, $N/n = 10$

Figure 5.9: Error spectrum plot [3] for Bernoulli sampled Knee A and Poisson disc Sampled Brain A, both at $N/n = 10$. Unbiased P-VDAMP's error is characterised by a larger error in the high frequencies, but a reduction in low, which is manifest in figures 5.6 and 5.7 as a reduction in blurring but an increase in high-frequency texturing.

vessel, which is visible for Unbiased P-VDAMP and partially visible for P-VDAMP but mostly lost in the competing FISTA-based methods. For Unbiased P-VDAMP, there are horizontal streaking artefacts. For the more ambitious undersampling factor, $N/n = 10$, these artefacts are more prominent, as reflected in the poor quality metrics in Fig. 5.3c.

To analyse the structure of the reconstruction artefacts in more detail, we used the Error Spectrum Plot (ESP) proposed in [3], which characterises the error's frequency dependence. For the knee and brain examples of figures 5.6 and 5.7, Fig. 5.9 shows the ESP for each algorithm. These plots demonstrate that the error of Unbiased P-VDAMP is characterised by relatively accurate low frequencies, but less accurate high frequencies. This is qualitatively manifest in figures 5.6 and 5.7 as a reduction in blurring but an increase in high-frequency texturing.

Overall, P-VDAMP's reconstruction quality is comparable with FISTA, but offers

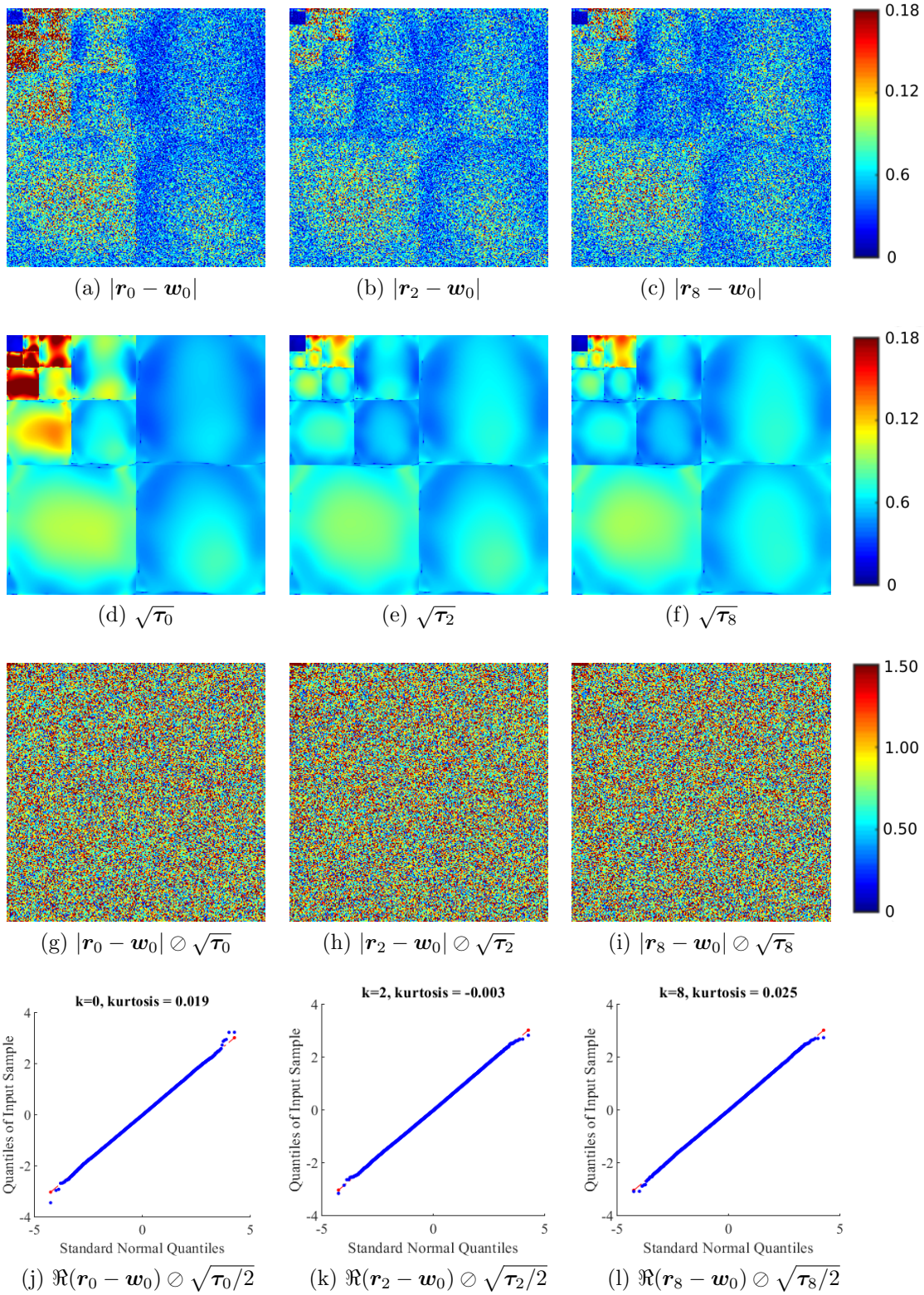
a clear improvement over the competing parameter-free methods, both quantitatively and qualitatively. The nature of Unbiased P-VDAMP's artefacts implies that the algorithm performs competitively qualitatively despite its comparatively poor quantitative metrics.

5.3.7 P-VDAMP's state evolution and damping

There are two aspects of state evolution of interest: (1) is the aliasing of $\boldsymbol{r}_k$ Gaussian? and (2) how accurately does $\boldsymbol{\tau}_k$ model its covariance? This section addresses these aspects using two representative examples.

In Chapter 4, we validated the existence of state evolution using kurtosis and quantile-quantile plots, comparing the effective noise of each wavelet subband with a white Gaussian distribution. The variances of the effective noise were estimated empirically by averaging over the subband, so the aliasing model $\boldsymbol{\tau}_k$ was not required. For the current, multi-coil method, since the per-subband aliasing is not white such a method is not possible and it is necessary to use $\boldsymbol{\tau}_k$ to normalise the effective noise to compare with a white Gaussian. This confounds the analysis of the two aforementioned aspects of state evolution. To evaluate aspect (1) of state evolution, all QQ plots and kurtosis calculations in this section were based *on the finest scale only*, where the wavelet filter is the least dilated and the aliasing model is the most accurate.

Fig. 5.10 shows the wavelet-domain effective noise of the the $N/n = 10$ Bernoulli sampled brain A at $k = 0, 2, 8$ when no damping is employed, so that $\rho = 1$, where ρ is defined in (5.14). The error model $\boldsymbol{\tau}_k$ is also shown, and the normalised effective noise, which qualitatively approximates a white Gaussian. The approximate Gaussianity of the real part of the finest scale is verified with QQ plots and kurtosis computations in the bottom row. Overall, Fig. 5.10 provides evidence that P-VDAMP with $\rho = 1$ obeys a parallel coloured state evolution. However, the reduction in aliasing is



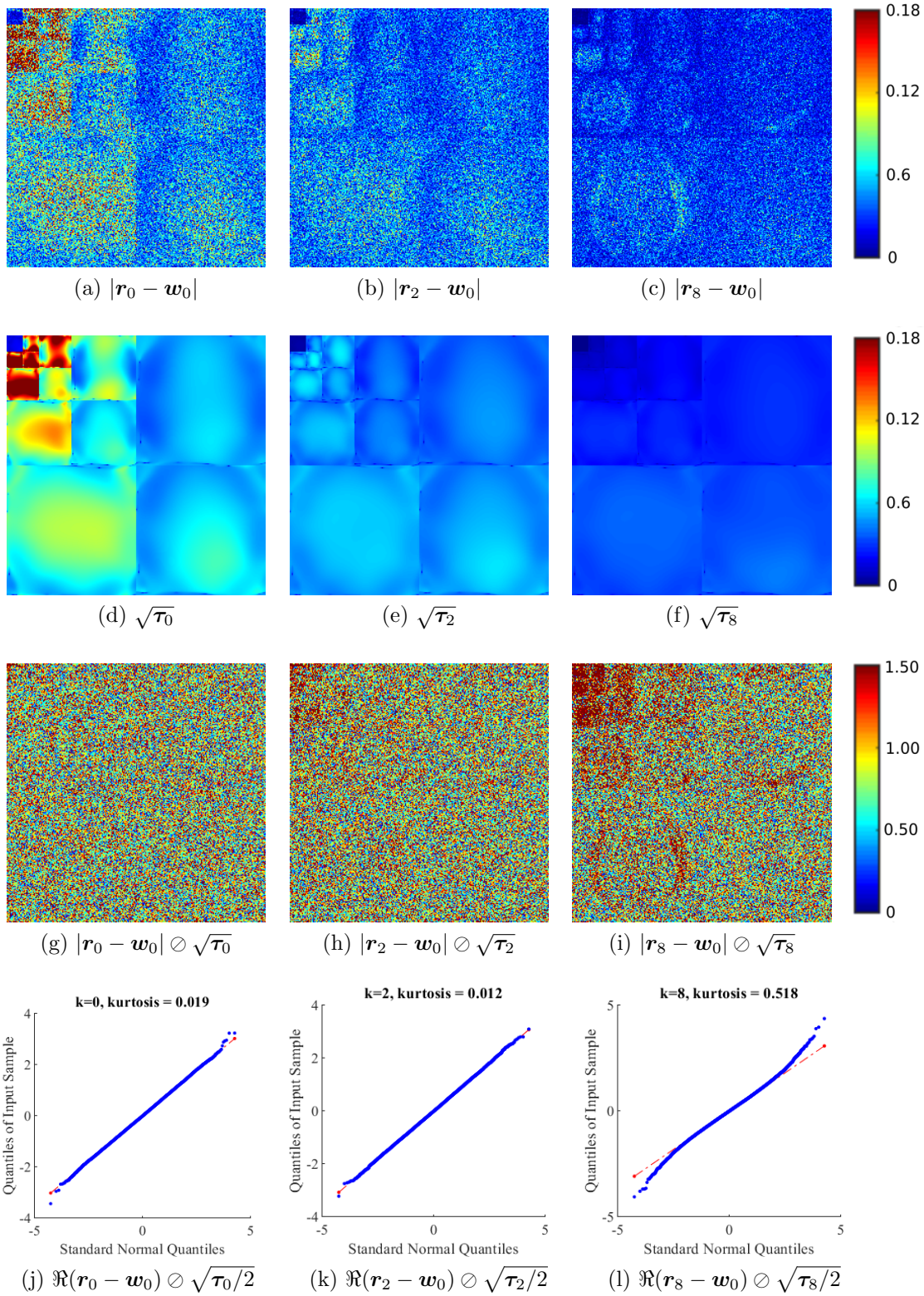
(a) $|\boldsymbol{r}_0 - \boldsymbol{w}_0|$ (b) $|\boldsymbol{r}_2 - \boldsymbol{w}_0|$ (c) $|\boldsymbol{r}_8 - \boldsymbol{w}_0|$

(d) $\sqrt{\boldsymbol{\tau}_0}$ (e) $\sqrt{\boldsymbol{\tau}_2}$ (f) $\sqrt{\boldsymbol{\tau}_8}$

(g) $|\boldsymbol{r}_0 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_0}$ (h) $|\boldsymbol{r}_2 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_2}$ (i) $|\boldsymbol{r}_8 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_8}$

(j) $\Re(\boldsymbol{r}_0 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_0/2}$ (k) $\Re(\boldsymbol{r}_2 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_2/2}$ (l) $\Re(\boldsymbol{r}_8 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_8/2}$

Figure 5.10: The wavelet-domain error and aliasing model of P-VDAMP with $\rho = 1$ at iterations $k = 0, 2, 8$ of Brain A, where k-space has been sampled with the $N/n = 10$ Bernoulli mask shown in Fig. 3.4a. The QQ plots are based on the real part at the finest scale.



(a) $|\boldsymbol{r}_0 - \boldsymbol{w}_0|$ (b) $|\boldsymbol{r}_2 - \boldsymbol{w}_0|$ (c) $|\boldsymbol{r}_8 - \boldsymbol{w}_0|$

(d) $\sqrt{\boldsymbol{\tau}_0}$ (e) $\sqrt{\boldsymbol{\tau}_2}$ (f) $\sqrt{\boldsymbol{\tau}_8}$

(g) $|\boldsymbol{r}_0 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_0}$ (h) $|\boldsymbol{r}_2 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_2}$ (i) $|\boldsymbol{r}_8 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_8}$

(j) $\Re(\boldsymbol{r}_0 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_0/2}$ (k) $\Re(\boldsymbol{r}_2 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_2/2}$ (l) $\Re(\boldsymbol{r}_8 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_8/2}$

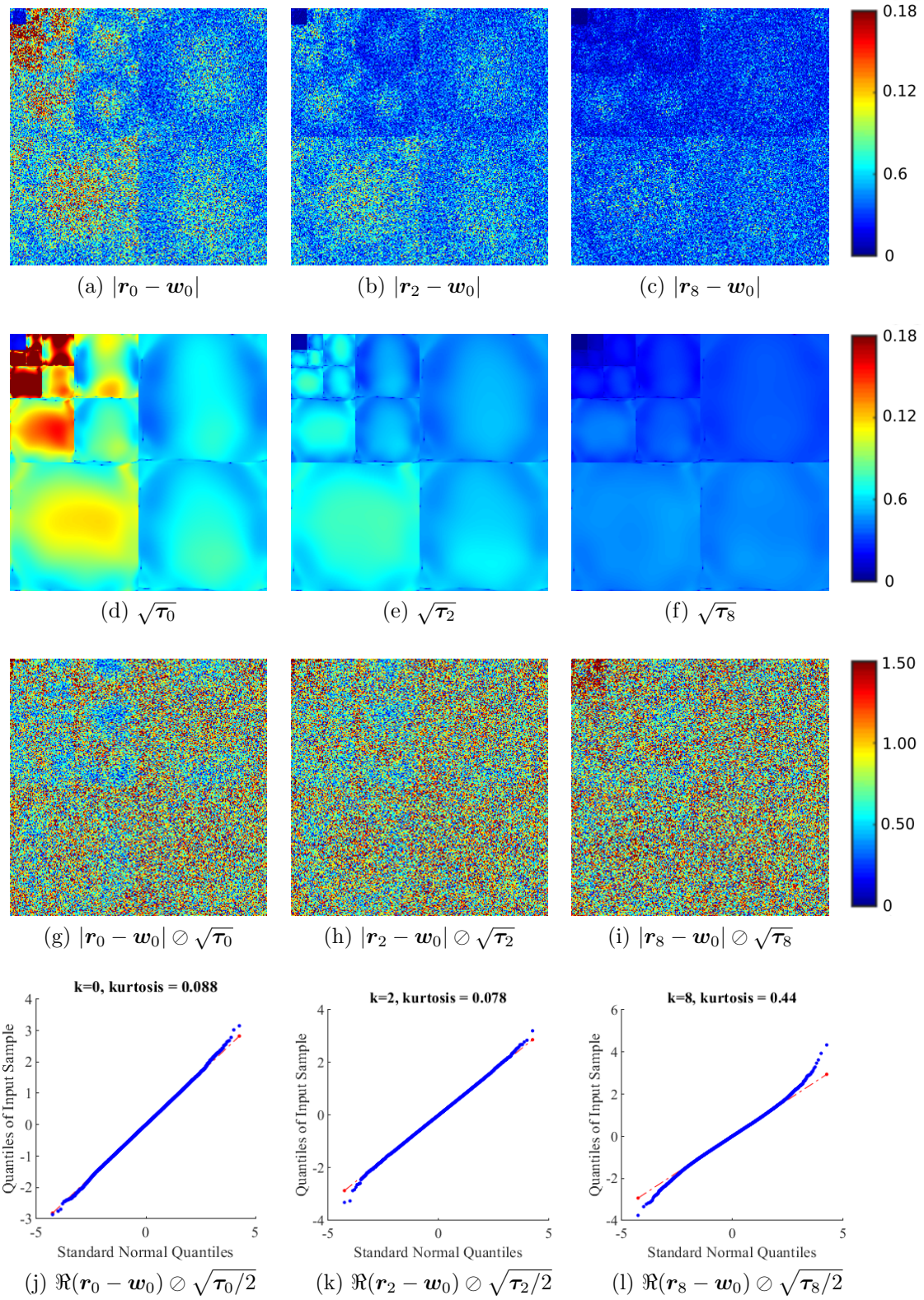
Figure 5.11: The wavelet-domain error and aliasing model of P-VDAMP with $\rho = 0.75$ at iterations $k = 0, 2, 8$ of $N/n = 10$ Bernoulli sampled Brain A. The aliasing decreases considerably more than in Fig. 5.10, despite a breakdown of P-VDAMP's state evolution for high k.

relatively small, and P-VDAMP's reconstruction is relatively poor, with an NMSE of -21.3dB when the stopping criterion was reached.

Fig. 5.11 shows a similar figure to 5.10 but with the damping constant set to $\rho = 0.75$, as used in the previous experimental sections in this chapter. With damping, there was a more substantial reduction in the aliasing, with an NMSE of -23.3dB when the stopping criterion was reached. However, the inexactness of Onsager correction in this case causes P-VDAMP's parallel coloured state evolution to fail as k increases, which is manifest at $k = 8$ in Fig. 5.11 as visibly biased aliasing and a large kurtosis. A breakdown in state evolution implies that parameter selection with SURE is no longer guaranteed to be near-optimal. Nonetheless, we found that the advantage of damping outweighed the disadvantage of poorer denoising for P-VDAMP's performance in general.

Fig. 5.12 shows similar plots with $\rho = 0.75$ for the same data and acceleration factor, but for Poisson disc sampling. In this case, the normalised effective noise visibly deviates from white, even at $k = 0$. Despite this, the QQ plots and kurtosis computations show that the normalised distribution is reasonably close to a Gaussian, demonstrating why P-VDAMP with SURE-optimal thresholding was found to perform reasonably well. For $k = 8$, the aliasing deviates more considerably from a Gaussian, as in Fig. 5.11.

To evaluate the accuracy of the aliasing model $\boldsymbol{\tau}_k$, aspect (2) of state evolution, Fig. 5.13 shows the true NMSE of $\boldsymbol{r}_k$ and its estimate, $\langle \boldsymbol{\tau}_k \rangle$, up to $k = 10$ for $N/n = 10$ Bernoulli and Poisson disc sampled Brain A and Knee A for both the damped and undamped versions of P-VDAMP. The only case where the aliasing is accurately tracked for all k is Bernoulli sampling without damping. For Bernoulli sampling with damping, the aliasing model is accurate for small k, but deviates as k increases, while for Poisson disc sampling, the aliasing model is inaccurate even for $k = 0$. This is consistent with the deviation from state evolution observed in figures



(a) $|\boldsymbol{r}_0 - \boldsymbol{w}_0|$ (b) $|\boldsymbol{r}_2 - \boldsymbol{w}_0|$ (c) $|\boldsymbol{r}_8 - \boldsymbol{w}_0|$

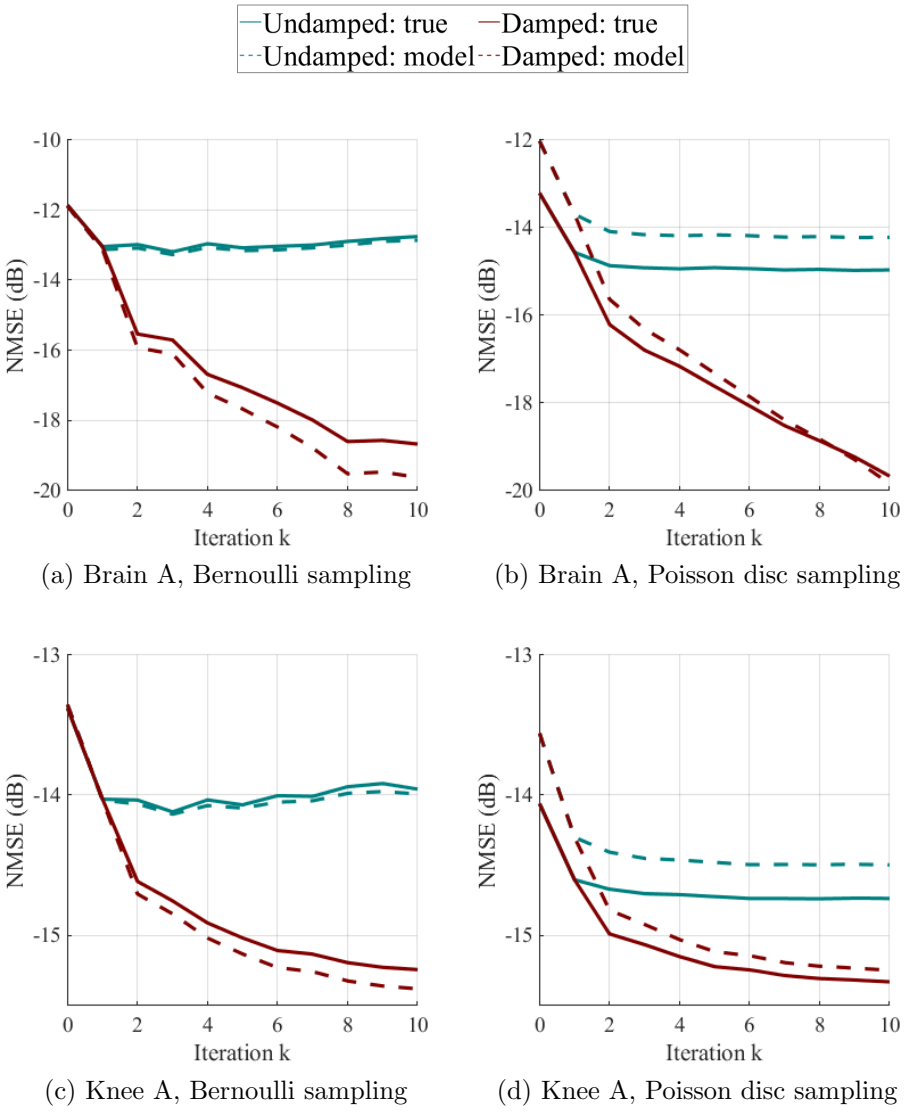
(d) $\sqrt{\boldsymbol{\tau}_0}$ (e) $\sqrt{\boldsymbol{\tau}_2}$ (f) $\sqrt{\boldsymbol{\tau}_8}$

(g) $|\boldsymbol{r}_0 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_0}$ (h) $|\boldsymbol{r}_2 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_2}$ (i) $|\boldsymbol{r}_8 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_8}$

(j) $\Re(\boldsymbol{r}_0 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_0/2}$ (k) $\Re(\boldsymbol{r}_2 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_2/2}$ (l) $\Re(\boldsymbol{r}_8 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_8/2}$

Figure 5.12: Equivalent of Fig. 5.10 for Poisson disc sampling with damping constant $\rho = 0.75$.

5.11 and 5.12.



(a) Brain A, Bernoulli sampling

(b) Brain A, Poisson disc sampling

(c) Knee A, Bernoulli sampling

(d) Knee A, Poisson disc sampling

Figure 5.13: True NMSE of $\boldsymbol{r}_k$ and its prediction $\langle \boldsymbol{\tau}_k \rangle$ for Bernoulli and Poisson disc sampled Brain A and Knee A at $N/n = 10$. Note that the NMSE values here were computed without the pre-analysis thresholding described in Section 5.3.4, so are not directly comparable with the results from figures 5.4a and 5.4b.

5.4 Conclusions

In this chapter, the single-coil VDAMP algorithm of Chapter 4 was extended to the multiple coil measurement model. The crucial modification was the $\boldsymbol{\tau}_k$ update, line 5

of Algorithm 6, which, to ensure scalability in N, approximates the coil sensitivities as flat over the wavelet filters' support. For Bernoulli sampling with no damping, P-VDAMP was found to obey a parallel coloured state evolution, (5.13). Although damping causes P-VDAMP's state evolution to break down for increasing k, we found that it nonetheless converged to a higher quality reconstruction. P-VDAMP performed well on Poisson disc sampled data without modification, despite a mismatch between the true aliasing and its model.

P-VDAMP with SURE-optimal wavelet thresholding offered significantly improved reconstruction quality over the competing tuning-free methods SURE-IT and A-FISTA. Compared with FISTA with a hand-selected threshold, P-VDAMP's reconstruction quality was more mixed, without a clear performance hierarchy. However, P-VDAMP has the clear advantage over FISTA that it alleviates the need to tune parameters by hand in a similar time to convergence.

Chapter 6

Denoising-Parallel-VDAMP

P-VDAMP's parallel coloured state evolution recasts the multi-coil compressed sensing MRI problem as a series of zero-mean, Gaussian denoising problems with approximately known covariance. In other words, how accurately one can reconstruct an image from randomly sampled k-space data depends simply on how well one can denoise a Gaussian-corrupted version of it.

Given sufficient representative training data, the current state-of-the-art in image denoising is with Convolutional Neural Networks (CNNs) [2,70,71]. This chapter proposes the Denoising-P-VDAMP (D-P-VDAMP) algorithm, where the SURE-optimal soft thresholding denoiser from Chapters 4 and 5 is replaced with a CNN.

CNNs are a recurring theme in MRI reconstruction research in recent years [17, 145–155]. Section 3.6.2 outlined three approaches to reconstruction with deep learning: Plug and Play (P&P), unrolled and black-box. D-P-VDAMP is a P&P method. Although for specific undersampling schemes unrolled and black-box methods can perform very well, P&P methods have the advantage that they are not constructed for specific k-space undersampling scheme, so are more robust to unseen sampling patterns, as shown in, for instance, [17]. The possibility of P-VDAMP as an unrolled method is discussed in Section 7.1.

6.1 Prior work: the D-VDAMP algorithm

Due to its relevance to the D-P-VDAMP method proposed in this chapter from Section 6.2 onwards, we discuss in detail work by other authors on the D-VDAMP algorithm [178], the VDAMP-based analogy of the D-AMP algorithm [44] discussed in Section 2.6.2. D-VDAMP was proposed after the publication of single-coil VDAMP [172] and before the development of the method proposed in this chapter. For a complete description of the method and results, see [178].

6.1.1 Summary of method and numerical results

D-VDAMP [178] is single-coil VDAMP, Algorithm 4, where the SURE-optimal soft thresholding $\boldsymbol{g}(\boldsymbol{r};\boldsymbol{\tau})$ is replaced by a generic denoiser. Four denoisers are proposed, each of which are applied in the image domain, so that $\boldsymbol{g}(\boldsymbol{r};\boldsymbol{\tau}) = \boldsymbol{\Psi}\boldsymbol{g}_{\text{im}}(\boldsymbol{\Psi}^H\boldsymbol{r};\boldsymbol{\tau})$ for some image-domain denoiser $\boldsymbol{g}_{\text{im}}$. The suggested image-domain denoisers are:

(A) The BM3D algorithm [68].

(B) A DnCNN [2], as described in Section 3.6.3, trained on images corrupted with white Gaussian noise.

(C) A DnCNN trained on images corrupted with coloured Gaussian noise.

(D) A DnCNN trained on images corrupted with coloured Gaussian noise, where each intermediate feature map is concatenated with a $N \times N \times (3s+1)$ tensor, where the bth slice consists of $N \times N$ copies of τ_b, the variance associated with subband b.

For each proposed architecture, four copies of the network were trained, specialised for particular effective noise standard deviation ranges. The denoising CNNs were

trained on around 127000 image patches from the Stanford 2D FSE dataset[1]. D-VDAMP's divergence term $\langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r};\boldsymbol{\tau}))\rangle_{\text{sband}}$ was approximated using an extension of the Monte-Carlo method developed in [205], described in detail in Section 6.2.2.

D-VDAMP was compared with the SURE-optimal wavelet thresholding VDAMP proposed in Chapter 4, TVAL3 [206], which is based on total variation, and two P&P algorithms: Regularisation by Denoising (RED) [207] and D-ADMM [208], with both denoisers (A) and (B) from above. The proposed method was found to substantially outperform all comparative methods, with denoiser (D) consistently achieving the lowest MSE. For instance, at acceleration factor 4, D-VDAMP with denoiser (D) outperformed its nearest (non-VDAMP-based) competitor RED by an MSE of 6dB. For a full set of results, see [178, Table 2].

The comparative improvement of D-VDAMP over the other P&P algorithms is a consequence of its state evolution. The absence of state evolution for RED and D-ADMM implies that there is a mismatch between the actual aliasing and the aliasing the network is trained to denoise, while for VDAMP, the aliasing is closely matched, leading to a far more effective denoising step.

6.1.2 Differences between D-VDAMP and the proposed D-P-VDAMP

The primary difference between D-VDAMP and the method proposed in the chapter is that we focus on modifying the *multi-coil* version of the algorithm, P-VDAMP from Chapter 5. The performance of D-VDAMP's denoiser (D) indicates that including the aliasing model $\boldsymbol{\tau}$ as an input to D-P-VDAMP's denoising CNN ought to improve its performance. However, since the multi-coil aliasing model includes intra-subband modulation, the specifics of denoiser (D) do not naturally generalise to P-VDAMP, as the variance of subband b cannot be represented with a scalar, so it is not possible

[1]available at `http://mridata.org/`

to concatenate $N \times N \times (3s+1)$ tensors in the same fashion. In response, we propose a novel *wavelet-domain* denoising architecture, described in Section 6.2.1, which uses a series of upsampling layers to gradually feed in the aliasing model and wavelet coefficients scale-by-scale.

6.2 Description of D-P-VDAMP

This section describes two developments required for P-VDAMP with a CNN $\boldsymbol{g}(\boldsymbol{r};\boldsymbol{\tau})$: 1) the development of an appropriate denoising CNN architecture, and 2) a method for estimating $\langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k))\rangle_{\text{sband}}$ for D-P-VAMP's Onsager correction.

6.2.1 Proposed wavelet-domain denoiser architecture

We propose a multiscale wavelet-domain denoiser based on the DnCNN described in Section 3.6.3. The proposed network architecture, which we term Wavelet-DnCNN (W-DnCNN), is represented in Fig. 6.1 for a 384×640 image decomposed at $s = 4$ scales. Unlike DnCNN, which takes only image-domain pixels as its input, W-DnCNN also takes wavelet coefficients. The wavelet coefficients are concatenated with their corresponding aliasing model and inputted to the network scale-by-scale, starting with the coarsest.

In Fig. 6.1, the black dashed boxes refer to network inputs. The $(12 \times 24 \times 40)$ input on the bottom left consists of a concatenation of the real part, imaginary part, and $\sqrt{\boldsymbol{\tau}}$ of the wavelet coefficients at the coarsest scale, including the approximation coefficients. The $(9 \times 48 \times 80)$, $(9 \times 96 \times 160)$ and $(9 \times 192 \times 320)$ inputs are the real part, imaginary part, and $\sqrt{\boldsymbol{\tau}}$ at the remaining three wavelet scales. These scales have 9 inputs channels because there are three subbands, unlike the coarsest scale which has 4 due to the additional subband of approximation coefficients. The $(2 \times 384 \times 640)$ input in the top right is the real and imaginary parts of the *image-domain* estimate.

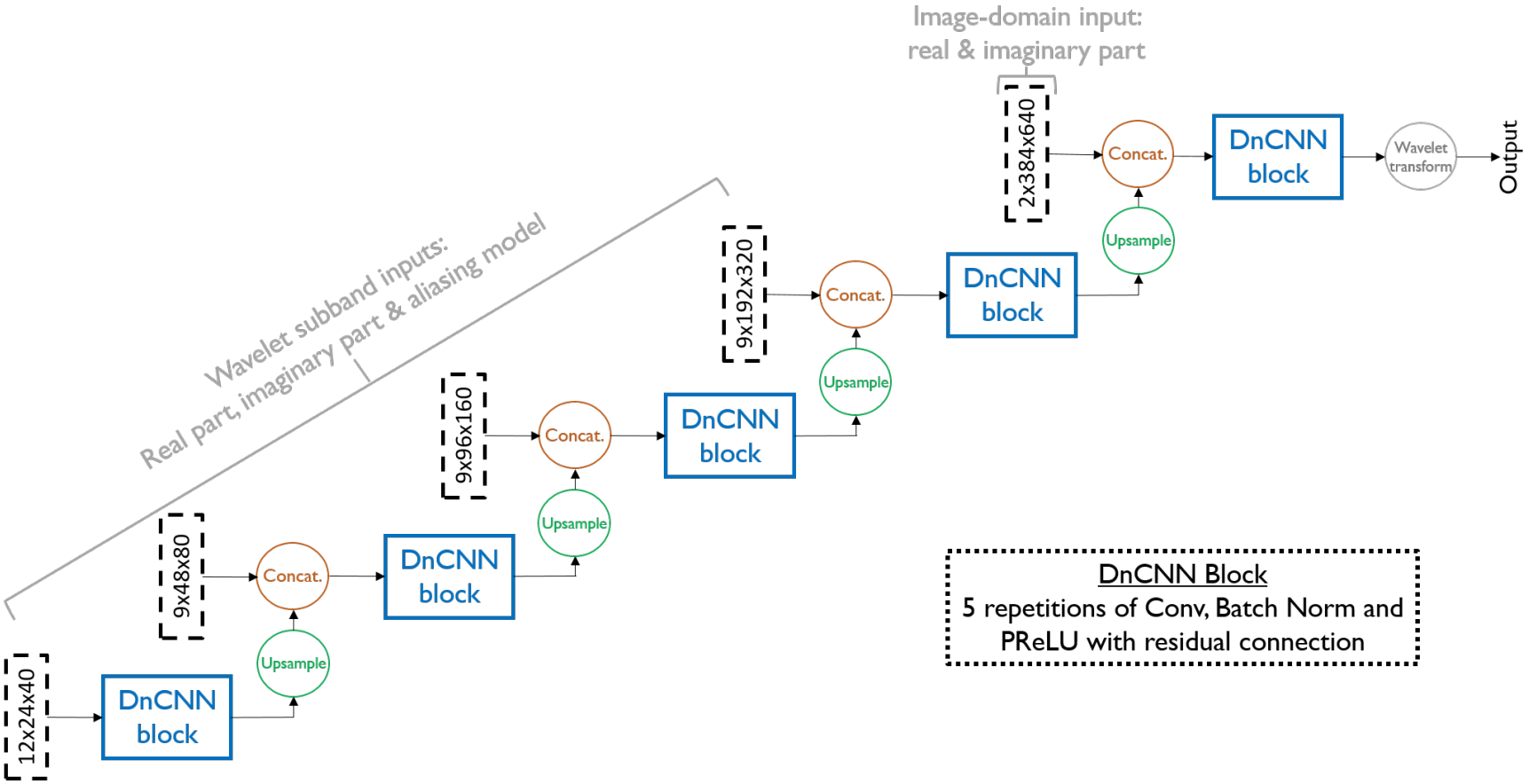


Figure 6.1: The proposed wavelet-based denoiser, W-DnCNN, for a 384×640 input with $s = 4$ scales.

The blue rectangles are "DnCNN Blocks", based on the DnCNN described in Section 3.6.3, which contain all the trainable parameters. The DnCNN block consists of 64 filters of size $(3 \times 3 \times c_{in})$, where c_{in} is the number of input channels to the block, followed by Batch Normalisation and Parametric ReLU, followed by 4 repetitions of 64 filters of size $(3 \times 3 \times 64)$, Batch Normalisation and Parametric ReLU. Parametric ReLU is defined as

$$\sigma_l(x_j) = \begin{cases} x_j, & \text{if } x > 0 \\ a_l x_j & \text{otherwise,} \end{cases}$$

where a_l is a trainable parameter. As in the original DnCNN, each DnCNN block in W-DnCNN has a residual connection [165].

The green circles are fixed bilinear upsampling layers, which scale the dimension of the channels by 2, so that they match the dimension of the subbands at the next-coarsest scale. The channels are then concatenated with the next wavelet subbands and aliasing model or, at the largest scale, the image-domain input.

Since the output of the final DnCNN block is in the image-domain, and P-VDAMP's $\boldsymbol{g}(\boldsymbol{r}; \boldsymbol{\tau})$ is defined to be a wavelet-to-wavelet denoiser, a final wavelet transform is performed, represented in Fig. 6.1 with a grey circle.

6.2.2 Monte-Carlo divergence estimation

The soft thresholding operator employed for the denoiser $\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$ in Chapters 4 and 5 has an analytical expression for $\langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)) \rangle_{\text{sband}}$. For D-P-VDAMP, its CNN $\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)$ cannot be differentiated analytically. This section describes a method based on [205] that estimates $\langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k)) \rangle_{\text{sband}}$ using Monte-Carlo approximation, as applied in [178] but not described in detail. A similar approach was proposed to compute the denoiser divergence in the context of D-AMP [44] for real systems. The method

here describes the modifications required to compute the complex, subband-averaged $\langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k))\rangle_{\text{sband}}$.

Let $\boldsymbol{b} = \boldsymbol{b}_R + i\boldsymbol{b}_I \backsim \mathcal{CN}(\boldsymbol{0}_N, \mathbb{1}_N)$. By [205, Theorem 2], applied to the real and imaginary parts,

$$\langle \boldsymbol{\partial}(\boldsymbol{d}(\boldsymbol{v}))\rangle = \lim_{\eta\to 0} \mathbb{E}_{\boldsymbol{b}} \left\{ \boldsymbol{b}_R^T \left(\frac{\Re[\boldsymbol{d}(\boldsymbol{v}+\eta\boldsymbol{b}) - \boldsymbol{d}(\boldsymbol{v})]}{\eta} \right) + \boldsymbol{b}_I^T \left(\frac{\Im[\boldsymbol{d}(\boldsymbol{v}+\eta\boldsymbol{b}) - \boldsymbol{d}(\boldsymbol{v})]}{\eta} \right) \right\}. \tag{6.1}$$

To approximate the expectation over $\boldsymbol{b}$ using Monte-Carlo simulation, one generates a series of random $\boldsymbol{b}_t = \boldsymbol{b}_{t,R} + i\boldsymbol{b}_{t,I} \backsim \mathcal{CN}(\boldsymbol{0}_N, \mathbb{1}_N)$ for $t \in \{1, 2, \ldots, N_t\}$ and computes

$$\langle \boldsymbol{\partial}(\boldsymbol{d}(\boldsymbol{v}))\rangle \approx \frac{1}{N_t} \sum_t^{N_t} \left[\boldsymbol{b}_{t,R}^T \left(\frac{\Re[\boldsymbol{d}(\boldsymbol{v}+\eta\boldsymbol{b}_t) - \boldsymbol{d}(\boldsymbol{v})]}{\eta} \right) + \boldsymbol{b}_{t,I}^T \left(\frac{\Im[\boldsymbol{d}(\boldsymbol{v}+\eta\boldsymbol{b}_t) - \boldsymbol{d}(\boldsymbol{v})]}{\eta} \right) \right]$$

where $\eta \in \mathbb{R}$ is small. The value of η should be small enough to accurately approximate (6.1) but large enough to mitigate floating point errors in $\boldsymbol{d}(\boldsymbol{v} + \eta\boldsymbol{b}_t)$. For all numerics in this chapter we used $\eta = \|\boldsymbol{v}\|_\infty/10^6$. Since $\langle \boldsymbol{\partial}(\boldsymbol{d}(\boldsymbol{v}))\rangle$ is a scalar, the expectation can be well approximated with a single random $\boldsymbol{b}_t$, so that $N_t = 1$ [44], as long as the dimension of $\boldsymbol{v}$ is large. We used $N_t = 1$ for all results in this chapter. To approximate the jth subband of $\langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k))\rangle_{\text{sband}}$, we used

$$\begin{aligned} &\langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k))\rangle_{\text{sband},j} \\ &\quad\approx \boldsymbol{b}_{j,R}^T \left(\frac{\Re[\boldsymbol{g}(\boldsymbol{r}_k+\eta\boldsymbol{b};\boldsymbol{\tau}_k) - \boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k)]}{\eta} \right) + \boldsymbol{b}_{j,I}^T \left(\frac{\Im[\boldsymbol{g}(\boldsymbol{r}_k+\eta\boldsymbol{b};\boldsymbol{\tau}_k) - \boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k)]}{\eta} \right) \end{aligned} \tag{6.2}$$

where $\boldsymbol{b} \backsim \mathcal{CN}(\boldsymbol{0}_N, \mathbb{1}_N)$ and $\boldsymbol{b}_j$ is $\boldsymbol{b}$ where all subbands other than jth are zeroed.

To summarise, a Monte-Carlo estimate of $\langle \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k;\boldsymbol{\tau}_k))\rangle_{\text{sband}}$ can be computed with the following steps: (A) Compute $\eta = \|\boldsymbol{r}_k\|_\infty/10^6$; (B) Generate a random vector $\boldsymbol{b} \backsim \mathcal{CN}(\boldsymbol{0}_N, \mathbb{1}_N)$; (C) Compute $\boldsymbol{g}(\boldsymbol{r}_k + \eta\boldsymbol{b};\boldsymbol{\tau}_k)$, the denoised wavelet coefficients

from the perturbed wavelet-domain input, and (D) Evaluate (6.2) for all subbands $j \in \{1, 2, \ldots, 3s+1\}$. This procedure is required once per iteration, with its main computational burden being the additional implementation of the denoiser on the perturbed input. This increases the computational expense of D-P-VDAMP compared with other P&P methods such as D-FISTA and D-ADMM, where no divergence estimation is required. However, as shown in the following section, we find that D-P-VDAMP's reduced number of iterations to convergence leads to a reconstruction time comparable with competing P&P methods.

6.3 Numerical experiments

6.3.1 Description of data

We evaluated the performance of the proposed method on knee data from the publicly available fastMRI dataset [196][2], which was acquired on a mixture of 3T and 1.5T scanners on $N_c = 15$ channel knee coil arrays. The training and validation sets include both fully-sampled multi-coil k-space data and undersampled data. The suggested undersampling masks are 2D acquisitions in rows, as in Fig. 3.3a. Since P-VDAMP's aliasing model is not constructed for such sampling schemes, we ignored the suggested undersampling patterns and undersampled k-space ourselves with Bernoulli and Poisson disc masks. The test set does not include fully sampled k-space, only the suggested 2D row-undersampled k-space, so was excluded. The remainder of the dataset was comprised of 1153 fully sampled knee k-space volumes, each with around 30-40 slices. We composed our own training, validation and test sets from this remainder, randomly allocating 953 knees for training and 100 knees for both the validation and testing sets.

Unfortunately, since the fastMRI challenges are based on the suggested 2D row un-

[2]available at `https://fastmri.org/dataset/`

dersampling patterns only, we were not able to submit our results to the leaderboard or compare our results directly with previously proposed methods on the leaderboard, such as [209, 210].

6.3.2 Preparing k-space data for training

The W-DnCNN described in Section 6.2.1 takes aliased wavelet coefficients and the aliasing model $\boldsymbol{\tau}$ as its input. To prepare the raw k-space data for network training, the following was performed on each 2D slice in the dataset:

1. Multi-coil k-space was Bernoulli undersampled with an undersampling factor generated uniformly at random between $n/N = 0.05$ and $n/N = 0.5$ with a 24×24 autocalibration region.

2. The sensitivity maps $\boldsymbol{S}_c$ were estimated via ESPIRiT [1] via the autocalibration region and, if necessary, zero-padded or cropped to size 384×640.

3. To reduce the compute time of steps 4 and 5, coil compression was performed to 8 virtual coils.

4. The fully sampled wavelet coefficients were computed with $\boldsymbol{w}_0 = \boldsymbol{\Psi} \sum_c \boldsymbol{S}_c^H \boldsymbol{F}^H \boldsymbol{y}_{\text{full},c}$ and saved to disk.

5. The wavelet-domain covariance estimate $\boldsymbol{\tau}$ was computed via (5.10) with $\xi_{c,j} = \langle |\boldsymbol{\Psi}_j|^2, \text{Diag}(\boldsymbol{S}_c^H) \rangle$ and saved to disk.

Coil compression and coil sensitivity estimation with ESPIRiT in steps 2 and 3 were computed using the BART toolbox [105]. Across the whole dataset, data preparation took around 30 hours.

We emphasise that only $\boldsymbol{w}_0 \in \mathbb{C}^N$ and $\boldsymbol{\tau} \in \mathbb{R}^N$ are saved to disk, and the full k-space data is not required during training. This contrasts with unrolled and black-box CNN-based reconstruction, which require the full N-dimensional complex k-space

data over N_c coils. Since $\boldsymbol{\tau}$ is real, so has only half the memory requirements of $\boldsymbol{w}_0 \in \mathbb{C}^N$ at the same precision, this requires $\frac{3}{2}N_c$ times more storage than D-P-VDAMP's training data. For the fastMRI dataset, where $N_c = 15$, this is a 10 times reduction in storage requirements, which on our system allowed the training data to be stored internally rather than on an external hard-drive, substantially reducing the data load-in time and overall training time.

As discussed in Section 5.1.2, P-VDAMP's aliasing model is more accurate for wavelets filters with a smaller support set. In D-P-VDAMP, wavelets are used exclusively as a tool for decomposing the aliasing, and not to compress the image, so there is no advantage to using wavelets with a larger support. We therefore used the Haar mother wavelet for $\boldsymbol{\Psi}$, which has the smallest support set. As in Chapters 4 and 5, we used 4 decomposition scales.

6.3.3 Network training procedure

For D-P-VDAMP to perform well, the denoiser must be effective at removing the aliasing for a range of $\boldsymbol{\tau}$. It is therefore important to prioritise the generalisability of the CNN's performance with respect to $\boldsymbol{\tau}$ during training. This was addressed by simulating noisy wavelet coefficients with a range of $\boldsymbol{\tau}$, using the $\boldsymbol{\tau}$ generated during data preparation as a basis. We chose to simulate noisy wavelet coefficients rather than to use $\boldsymbol{r}_0 = \boldsymbol{\Psi} \sum_c \boldsymbol{S}_c^H \boldsymbol{F}^H \boldsymbol{P}^{-1} \boldsymbol{M}_\Omega \boldsymbol{y}_c$ because, as described in the following, it allows the denoiser to see a greater range of aliased inputs.

In this section, i and j index the 2D slices of the training points and aliasing models. We simulated the ith aliased training point $\boldsymbol{r}_i$ by corrupting the ground-truth wavelet coefficients $\boldsymbol{w}_{0,i}$ with complex Gaussian noise, so that

$$\boldsymbol{r}_i = \boldsymbol{w}_{0,i} + \mathcal{CN}(\boldsymbol{0}_N, \mathrm{Diag}(\boldsymbol{\tau}_j)), \tag{6.3}$$

where j is not necessarily equal to i. In other words, we shuffled the aliasing models with the ground-truth wavelet coefficients, so that the wavelet coefficients were not necessarily paired with the $\boldsymbol{\tau}$ generated during data preparation. These pairings were reshuffled at the start of every epoch.

Since the aliasing is expected to decrease for increasing k, D-P-VDAMP's denoiser is required to remove aliasing for both large and small $\langle \boldsymbol{\tau} \rangle$. In the D-VDAMP paper [178], this issue was addressed by using four denoising CNNs with identical architectures, which were specialised to a range of $\langle \boldsymbol{\tau} \rangle$. Here we take an alternative approach, using the same network for all aliasing models, but with a choice of loss function that encourages better performance at small $\langle \boldsymbol{\tau} \rangle$. We used the *root* mean-squared error,

$$L_{rmse} = \frac{1}{\sqrt{N_{train}}} \sum_{i=1}^{N_{train}} \|\boldsymbol{w}_{0,i} - \boldsymbol{g}(\boldsymbol{r}_i; \boldsymbol{\tau}_i)\|_2, \tag{6.4}$$

as the loss, where N_{train} is the number of training points. Easier denoising tasks, where $\langle \boldsymbol{\tau} \rangle$ is low, are more strongly penalised with the RMSE compared with the MSE, as used in [178]. This improved the denoising performance in these scenarios, which we found led to improved reconstruction performance overall.

We used a mini-batch size of 4 and trained for 50 epochs using the Adam optimiser [158] with a learning rate of 10^{-4}, and used the model with the lowest loss on the validation set. Training was completed in 76 hours on a single NVIDIA TITAN X GPU in PyTorch.

6.3.4 Comparative methods

The term P&P encompasses algorithms that use non-CNN denoisers, such as BM3D [68], which was applied to MRI in [76, 211]. This chapter does not consider non-learned denoisers like BM3D, as they are far slower to implement, and known to be

outperformed by CNNs given sufficient representative training data [2, 70, 71, 178].

For a comparative P&P method, we ran D-FISTA [212]. To avoid confounding the performance of the algorithm with the performance of the network architecture, we used the same W-DnCNN denoiser as D-P-VDAMP, described in Section 6.2.1, for D-FISTA. However, unlike D-P-VDAMP, we trained the network on *white* Gaussian noise, as in previous implementations of P&P for MRI [17, 145]. To train the white denoiser, the covariance of the simulated Gaussian noise was set to $\langle \boldsymbol{\tau}_j \rangle \mathbb{1}_N$. The training procedure was otherwise identical, using the same network depth and training hyper-parameters as for D-P-VDAMP, as described in Section 6.2.1.

In Chapters 4 and 5, FISTA's scalar MSE estimate τ_k, which scales the threshold, was computed using the fully sampled reference. Since D-FISTA's τ_k is used to guide the denoising CNN it plays a more crucial role, so in this chapter we consider the more realistic case where τ_k is estimated. For denoiser input $\boldsymbol{r}_k$, we used the mean-squared-error of the measured coefficients,

$$\tau_k = \sum_{c=1}^{N_c} \frac{1}{nN_c} \|\boldsymbol{y}_c - \boldsymbol{M}_\Omega \boldsymbol{F} \boldsymbol{S}_c \boldsymbol{\Psi}^H \boldsymbol{r}_k\|_2^2, \tag{6.5}$$

which would be reliable if the coil-weighted spectral density of the aliasing of $\boldsymbol{r}_k$ were approximately uniform.

In order to evaluate the effect of treating the noise as coloured, as well as D-P-VDAMP with the full, coloured aliasing model, we also ran the algorithm with the same white denoiser as D-FISTA. We whitened D-P-VDAMP's aliasing model by replacing $\boldsymbol{g}(\boldsymbol{r}; \boldsymbol{\tau})$ in Algorithm 6 with $\boldsymbol{g}(\boldsymbol{r}; \langle \boldsymbol{\tau} \rangle \mathbf{1}_N)$. As in Chapter 5, we considered the alpha variant of P-VDAMP only, so that line 8 of Algorithm 6 was $\boldsymbol{c}_k = \mathbf{1}_N \oslash (\mathbf{1}_N - \boldsymbol{\alpha}_k)$. Also as in Chapter 5, we evaluated the performance of both the standard and unbiased outputs of P-VDAMP. We used the damping factor $\rho = 0.75$ and $\boldsymbol{\tau}$-based stopping criterion detailed in Section 5.2.1.

For comparative non-CNN models, we ran the SURE-optimal soft thresholding P-VDAMP algorithm from Chapter 5 and multi-coil FISTA, Algorithm 5, with a threshold λ weighted by τ_k. We chose FISTA's λ by tuning it to be approximately MSE-optimal with respect to the reference on a randomly selected $N/n = 5$ Bernoulli undersampled slice from the test set.

In [17], it was shown that a U-Net [163] black-box method was consistently outperformed by D-ADMM with a DnCNN denoiser on the knee fastMRI dataset, even for a fixed sampling mask. Therefore, as in [178], we did not compare D-P-VDAMP with a baseline black box reconstruction method. The possibility of D-P-VDAMP as an unrolled neural network is discussed in Section 7.1.

6.3.5 Performance metrics

We ran all algorithms on the whole test set at $N/n = 5$ and $N/n = 10$ for both Poisson disc and Bernoulli sampling, using the same variable density distribution as in the numerical evaluation of P-VDAMP in Chapter 5, as illustrated in Figure 3.4. As discussed in Section 5.3.7, P-VDAMP's aliasing model does not accurately describe the true aliasing for Poisson disc sampling, see Fig. 5.12. However, as for the soft thresholding case, we found that D-P-VDAMP performed well without modification.

We used the same metrics as in Chapter 5 to evaluate the performance of the algorithms: NMSE, SSIM and HFEN, see Section 5.3.4 for details. These metrics were computed after the following two processing steps. As in Chapter 5, to mitigate contributions from the background, we masked all pixels with absolute values below 5% of the maximum of the reference. Also, as suggested in [196], the outputs of each algorithm were cropped to a central 320×320 region, which accounts for k-space oversampling that is standard in clinical practice.

N/n	5				10			
Quality metric	NMSE	SSIM	HFEN	Time	NMSE	SSIM	HFEN	Time
FISTA	-24.3	0.893	-22.7	4.86	-21.6	0.844	**-20.8**	5.67
D-FISTA (white)	-24.3	0.892	-22.2	**3.62**	-22.4	0.855	**-20.8**	4.63
P-VDAMP	-24.5	0.894	-22.5	23.71	-21.5	0.840	-20.1	7.75
D-P-VDAMP (white)	-25.0	0.900	-22.4	4.49	-22.9	**0.860**	**-20.8**	4.03
D-P-VDAMP (coloured)	**-25.8**	**0.906**	-22.8	4.81	**-23.1**	0.858	-20.7	**3.66**
Unbiased D-P-VDAMP (coloured)	-23.0	0.859	**-22.9**	4.81	-16.1	0.678	-20.3	**3.66**

Table 6.1: Average performance of algorithms on the test set for **Bernoulli sampled** data, where the NMSE and HFEN are in decibels and time is in seconds. The best performances are highlighted in bold.

N/n	5				10			
Quality metric	NMSE	SSIM	HFEN	Time	NMSE	SSIM	HFEN	Time
FISTA	-24.7	0.898	-23.2	4.44	-20.7	0.828	-19.9	**3.00**
D-FISTA (white)	-24.6	0.896	-22.7	2.77	-22.3	0.851	-20.5	3.51
P-VDAMP	-24.7	0.896	-22.6	5.03	-21.9	0.847	-20.5	8.08
D-P-VDAMP (white)	-25.8	0.905	-23.2	3.58	-23.4	**0.866**	**-21.4**	3.64
D-P-VDAMP (coloured)	**-26.2**	**0.908**	**-23.3**	**2.27**	**-23.5**	0.863	-21.1	3.39
Unbiased D-P-VDAMP (coloured)	-22.5	0.847	**-23.3**	**2.27**	-16.3	0.685	-20.9	3.39

Table 6.2: Average performance of algorithms on the test set for **Poisson Disc sampled** data.

6.4 Results

6.4.1 Quantitative Performance on test set

The average NMSE, SSIM and HFEN and time to convergence on the test set are shown for Bernoulli and Poisson disc sampling in Tables 6.1 and 6.2 respectively. Coloured D-P-VDAMP outperformed P-VDAMP according to every metric for all experiments, highlighting the advantage of using a richer, CNN-based model over soft thresholding. Coloured D-P-VDAMP's average NMSE was consistently lower than white D-P-VDAMP, demonstrating the benefits of using the full, coloured aliasing model, rather than the white aliasing model used in previous P&P MRI works [17, 145]. In some cases the SSIM and HFEN are marginally better for white D-P-VDAMP than coloured D-P-VDAMP. However, we note that the NMSE is the best reflection of the intended performance of the denoising CNN, as RMSE was used as the training loss, and that it would be expected that HFEN or SSIM improvements could be achieved by including appropriate terms in the loss function.

Except for HFEN at 10 times Bernoulli undersampling, coloured D-P-VDAMP's NMSE, SSIM and HFEN were consistently better than D-FISTA. Out of 3705 slices in the test set, coloured D-P-VDAMP outperformed D-FISTA in terms of NMSE in the overwhelming majority of cases: 97% and 85% of cases for 5 and 10 times undersampled Bernoulli sampling respectively and 95% and 98% of cases for Poisson disc.

White D-P-VDAMP's NMSE, SSIM and HFEN were consistently as good or better than D-FISTA, suggesting that any denoising CNN trained on white Gaussian noise, such as a simple image-domain DnCNN used in [17, 145], would benefit from using D-P-VDAMP as its basis algorithm.

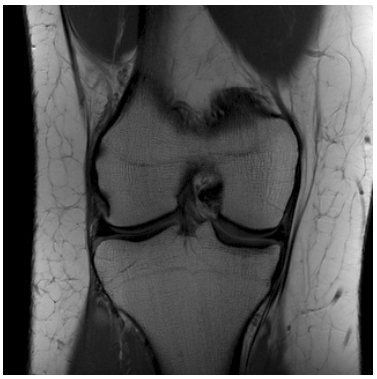
Unbiased, coloured D-P-VDAMP did not perform competitively in terms of NMSE and SSIM, but did perform reasonably well in terms of HFEN. As for the experimental results of Chapter 5, we found that Unbiased D-P-VDAMP nonetheless performed well qualitatively: see Section 6.4.2.

D-P-VDAMP's fairly expensive $\boldsymbol{\tau}$ update and the need to run two forward passes of the denoising CNN for Monte-Carlo divergence estimation via (6.2) means that the time per iteration was larger than D-FISTA. Averaged across all reconstructions, the time per iteration in our Python implementation was 1.53 and 0.58 seconds for coloured D-P-VDAMP and D-FISTA respectively. Despite being around 3 times slower per iteration, D-P-VDAMP's required number of iterations to convergence was substantially smaller, leading to a faster average time to convergence than D-FISTA in 3 out of 4 results from Tables 6.1 and 6.2.

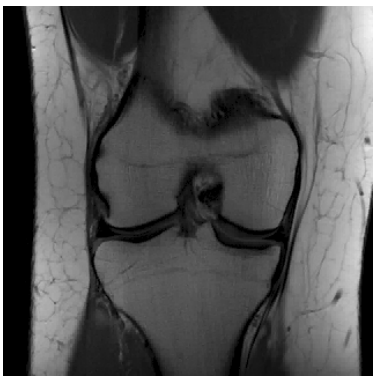
6.4.2 Qualitative analysis of reconstruction quality

Two sets of reconstructions for randomly selected slices from the test set are shown in figures 6.2 and 6.3 for Bernoulli at $N/n = 5$ and Poisson disc at $N/n = 10$ respectively.

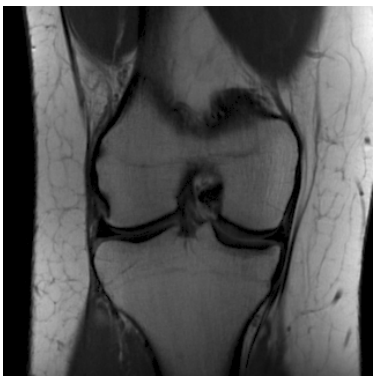
For FISTA and P-VDAMP there are blocking artefacts, especially at tissue bound-



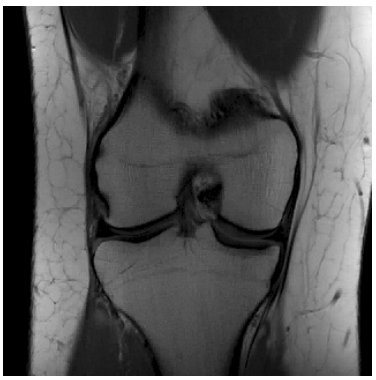
(a) Reference



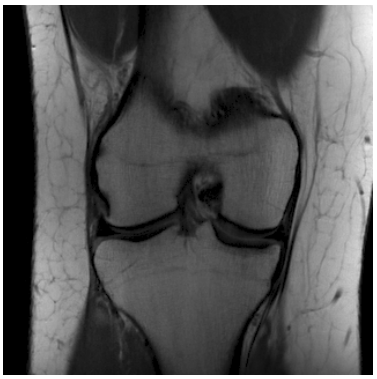
(b) FISTA



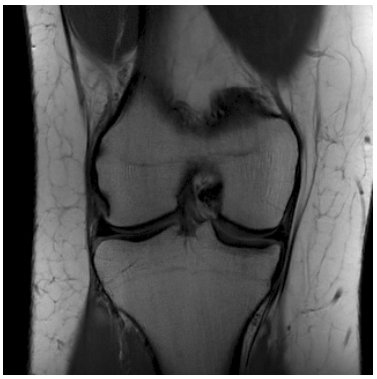
(c) D-FISTA



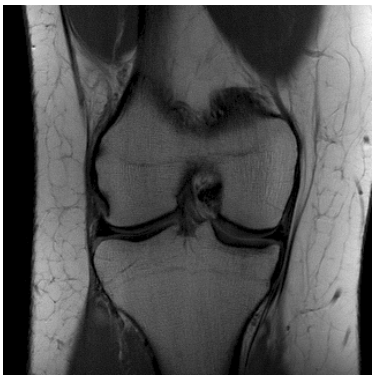
(e) P-VDAMP



(f) D-P-VDAMP (white)

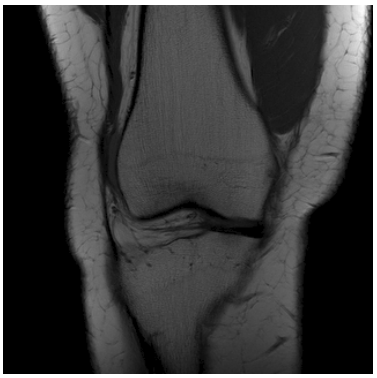


(g) D-P-VDAMP (coloured)

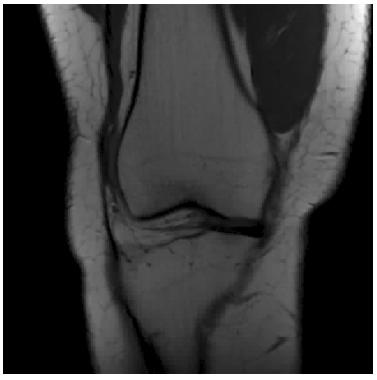


(h) Unbiased D-P-VDAMP (coloured)

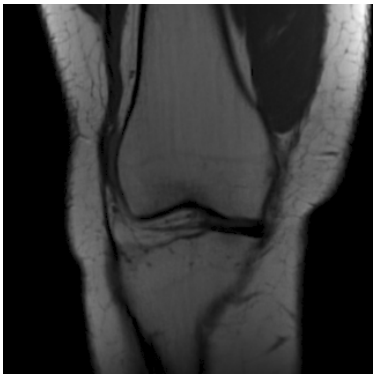
Figure 6.2: Reference and reconstructions of a slice from the test set Bernoulli undersampled at $N/n = 5$.



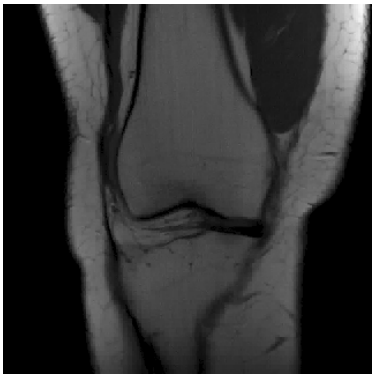
(a) Reference



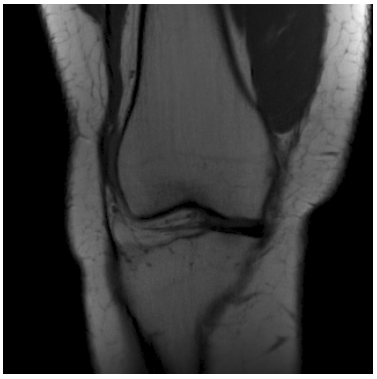
(b) FISTA



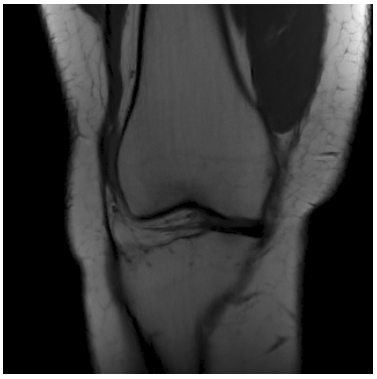
(c) D-FISTA



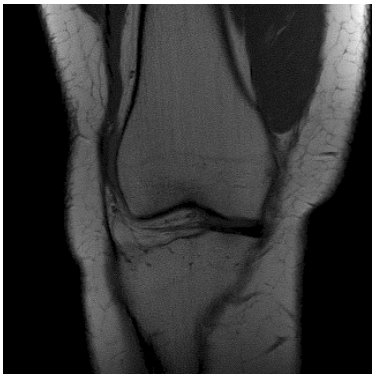
(e) P-VDAMP



(f) D-P-VDAMP (white)



(g) D-P-VDAMP (coloured)



(h) Unbiased D-P-VDAMP (coloured)

Figure 6.3: Reference and reconstructions of Poisson disc undersampled at $N/n = 10$.

aries, which arise due to wavelet thresholding. For D-FISTA and the two biased D-P-VDAMP algorithms there is no visible blockiness, but an increase in blurring at the edges and some loss of anatomical detail in the bulk tissue, which is especially prominent for the more challenging ten-fold undersampling of Fig 6.3.

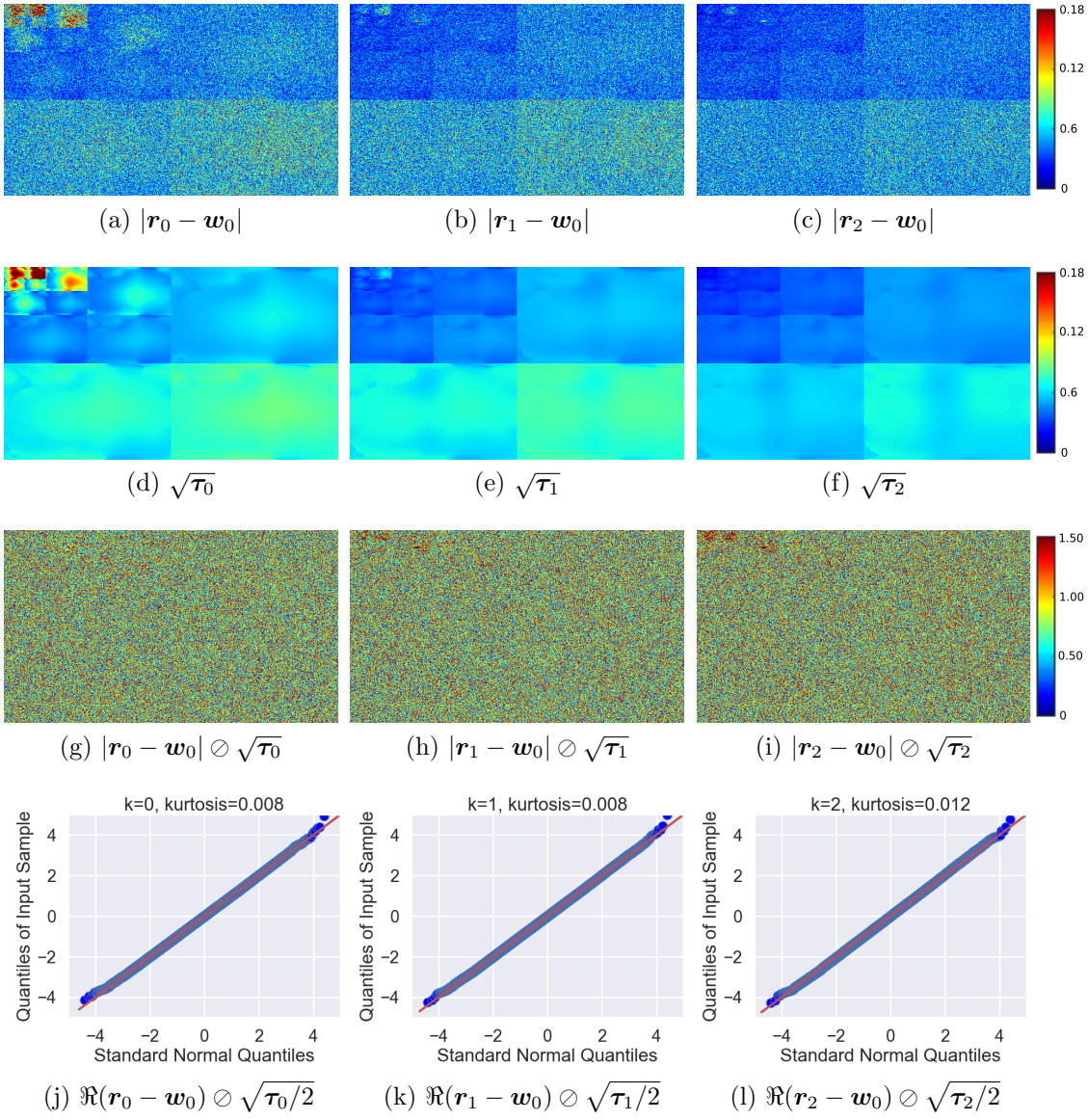
According to NMSE and SSIM, Unbiased D-P-VDAMP performs far worse in the examples of figures 6.2 and 6.3. For instance, its NMSE in Fig. 6.2 is -17.5dB, while the next worst, P-VDAMP, is -23.8dB. Nonetheless, Unbiased D-P-VDAMP qualitatively retains fine anatomical detail well. As discussed in Section 5.3.6, the unbiased estimate has a greater contribution to the aliasing from high-frequencies.

6.4.3 D-P-VDAMP's state evolution

Figures 6.4 and 6.5 show the aliasing of $\boldsymbol{r}_k$ at $k = 0, 1, 2$ and its model for D-P-VDAMP with a coloured denoiser, using the same examples as figures 6.2 and 6.3. These figures demonstrate that, as in the case of soft thresholding, the Bernoulli sampled reconstruction approximately obeys state evolution with some deviation due to damping, while the Poisson disc sampled example deviates even at $k = 0$.

6.5 Conclusions

Previous P&P methods for MRI such as D-FISTA have a mismatch between the noise of the training data and the actual per-iteration aliasing. D-P-VDAMP's state evolution addresses this issue, giving a way to train a denoising CNN on inputs with noise that closely matches the effective noise seen during reconstruction. D-P-VDAMP with a coloured denoiser outperformed D-FISTA's average NMSE and SSIM for all explored sampling schemes, and all average HFEN except one. Further, despite its fairly expensive aliasing update and need for two forward passes of the denoiser for the Monte-Carlo divergence estimate, D-P-VDAMP's low required number of iterations led to a competitive time to convergence.



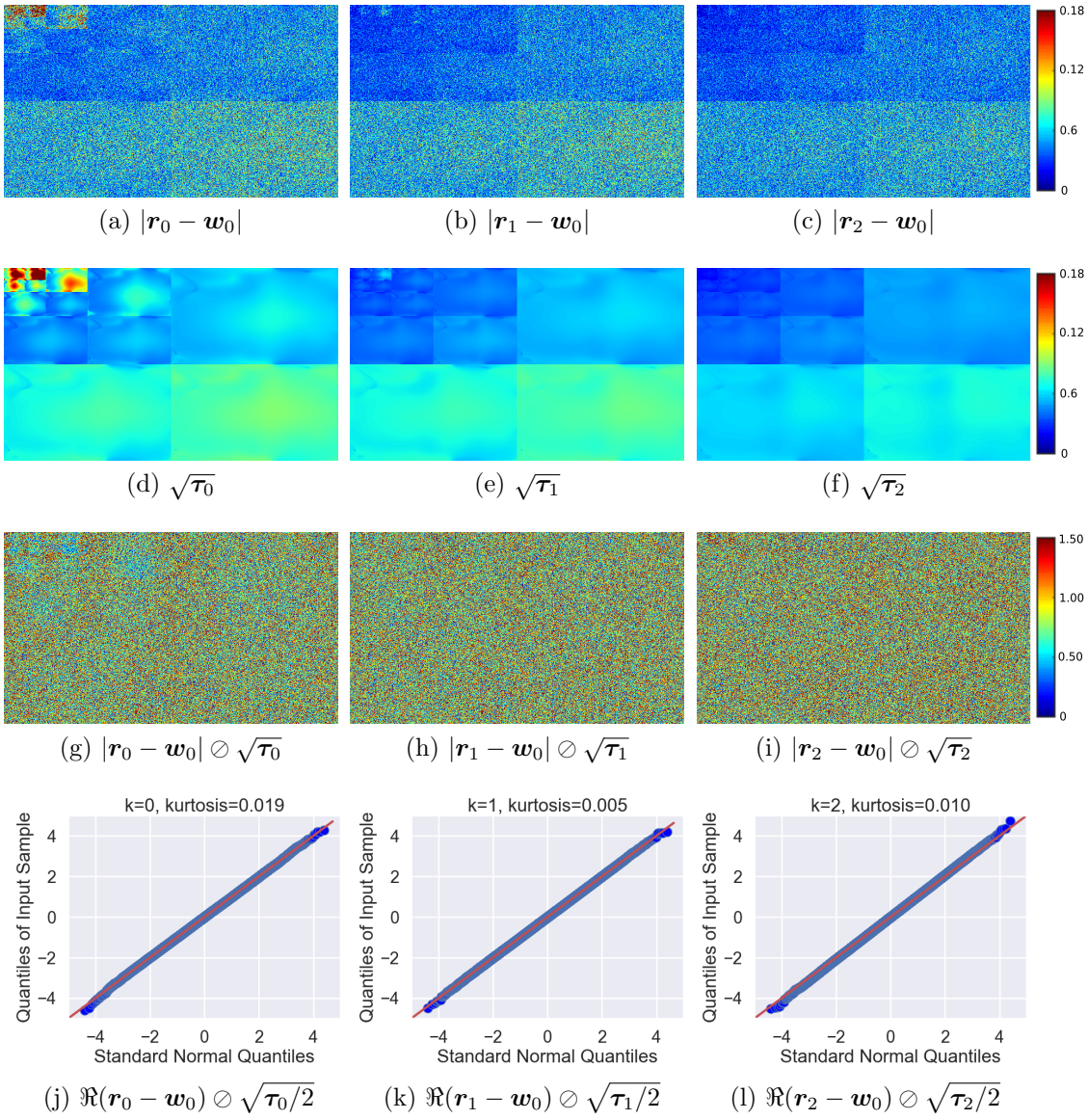
(a) $|\boldsymbol{r}_0 - \boldsymbol{w}_0|$ (b) $|\boldsymbol{r}_1 - \boldsymbol{w}_0|$ (c) $|\boldsymbol{r}_2 - \boldsymbol{w}_0|$

(d) $\sqrt{\boldsymbol{\tau}_0}$ (e) $\sqrt{\boldsymbol{\tau}_1}$ (f) $\sqrt{\boldsymbol{\tau}_2}$

(g) $|\boldsymbol{r}_0 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_0}$ (h) $|\boldsymbol{r}_1 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_1}$ (i) $|\boldsymbol{r}_2 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_2}$

(j) $\Re(\boldsymbol{r}_0 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_0/2}$ (k) $\Re(\boldsymbol{r}_1 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_1/2}$ (l) $\Re(\boldsymbol{r}_2 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_2/2}$

Figure 6.4: The wavelet-domain error and aliasing model at iterations $k = 0, 1, 2$ of the test set slice shown in Fig. 6.2, where k-space has been sampled with a $N/n = 5$ Bernoulli mask. As in Section 5.3.7, the Q-Q plots and kurtosis computations are based on the finest scale only.



(a) $|\boldsymbol{r}_0 - \boldsymbol{w}_0|$ (b) $|\boldsymbol{r}_1 - \boldsymbol{w}_0|$ (c) $|\boldsymbol{r}_2 - \boldsymbol{w}_0|$

(d) $\sqrt{\boldsymbol{\tau}_0}$ (e) $\sqrt{\boldsymbol{\tau}_1}$ (f) $\sqrt{\boldsymbol{\tau}_2}$

(g) $|\boldsymbol{r}_0 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_0}$ (h) $|\boldsymbol{r}_1 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_1}$ (i) $|\boldsymbol{r}_2 - \boldsymbol{w}_0| \oslash \sqrt{\boldsymbol{\tau}_2}$

(j) $\Re(\boldsymbol{r}_0 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_0/2}$ (k) $\Re(\boldsymbol{r}_1 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_1/2}$ (l) $\Re(\boldsymbol{r}_2 - \boldsymbol{w}_0) \oslash \sqrt{\boldsymbol{\tau}_2/2}$

Figure 6.5: The wavelet-domain error and aliasing model at iterations $k = 0, 1, 2$ of the test set slice shown in Fig. 6.3, where k-space has been sampled with a $N/n = 10$ Poisson disc mask.

Chapter 7

Conclusions

VDAMP and P-VDAMP are, to our knowledge, the first algorithms for single-coil and multi-coil MRI measurements that obey a state evolution. The central idea of VDAMP was to modify OAMP's standard white aliasing model to reflect the coloured effective noise that arises when an image is sampled in k-space. To model the aliasing we made use of the wavelet transform, which approximately diagonalises coloured noise.

This thesis explored two ways in which VDAMP's state evolution can be leveraged for MRI reconstruction. In Chapters 4 and 5, SURE was employed to tune parameters of the denoiser $\boldsymbol{g}(\boldsymbol{r};\boldsymbol{\tau})$ to near-optimality on-the-fly, alleviating the need to pre-tune parameters by hand and yielding an algorithm that converges rapidly to an image estimate with competitive reconstruction error. In Chapter 6, we showed how state evolution provides a natural way to train a CNN-based denoiser within the P&P framework, so that the actual per-iteration aliasing closely matches the aliasing the network is trained to denoise, leading to very effective denoising and state-of-the-art reconstruction quality.

7.1 Future work

The primary limitation of VDAMP is that a 3D acquisition must be acquired with a fully sampled readout dimension and decomposed into a series of 2D subproblems. Future work includes extending VDAMP to non-Cartesian sampling schemes such as radial or spiral, and to purely 2D acquistions. The key challenge for such sampling schemes would be to develop a suitable aliasing model, which would replace line 6 of Algorithms 4 and 6. Such developments would also allow VDAMP to be applied to other Fourier-based imaging applications of compressed sensing, such as astronomical imaging [213–215] and x-ray tomographic imaging [216–219].

A benefit of state evolution that could be further explored is as a tool for uncertainty quantification. After the publication of VDAMP, a SURE-based method was developed for D-AMP and (single-coil) D-VDAMP that generates pixel-wise uncertainty maps that were shown to match closely with the true error [220]. Future work includes extending this framework to D-P-VDAMP.

The theoretical avenue of utmost importance is the need for a proof of VDAMP's coloured state evolution, which was demonstrated with empirical evidence only in this thesis. OAMP's state evolution for unitarily invariant sensing matrices was proven in [56], which was based on a proof strategy originally proposed for AMP in [49]. However, the partial Fourier sensing matrices relevant in MRI are not unitarily invariant. To our knowledge, no analytical proof of state evolution for Fourier measurement currently exists, despite empirical evidence that it holds for AMP for Fourier samples of a signal with flat spectral density [12]. A proof of OAMP's state evolution for partial Fourier matrices would need to be appropriately modified in the variable density case, where VDAMP's coloured state evolution is required.

The remainder of this section suggests a number of future directions for deep-learning based P-VDAMP reconstruction.

The focus of Chapter 6 was on the P&P class of reconstruction methods. It

would be valuable to explore an unrolled version of P-VDAMP. For instance, a simple unrolled D-P-VDAMP would involve multiple iterations on the forward pass, with iteration-dependent denoising neural networks. This could offer reconstruction quality improvements, as the denoising neural network would be specialised for the iteration. A disadvantage of such a method is a substantial increase in memory requirements and training time due to the larger number of trainable parameters and the need to use the full multi-coil k-space data during training.

A number of recent works seek to train neural networks for deep learning reconstruction without ground truth data [146,174,221–225]. The D-P-VDAMP framework offers two potential methods for training without ground truth data. The first method would use a SURE-based loss as a proxy for the true MSE to train the denoiser, as in [226]. This would require an additional forward pass of the network for Monte-Carlo estimation of $\boldsymbol{\tau}_k^T \boldsymbol{\partial}(\boldsymbol{g}(\boldsymbol{r}_k; \boldsymbol{\tau}_k))$ in (4.21). The second method would compute another aliasing model $\boldsymbol{\tau}$ after the action of the denoiser. In other words, during training, lines 7-9 of Algorithm 6 would be computed, followed by line 3 and the aliasing model update in line 5. Then $\boldsymbol{\tau}$ could be used in the training loss, such as with the loss function $\sqrt{\langle \boldsymbol{\tau} \rangle}$. This would require computing the Onsager correction, loading in k-space data, as well as computing the $\boldsymbol{\tau}$ update, so would be considerably more computationally demanding than the method proposed in this chapter. Future work includes exploring the efficacy of these suggestions.

A computational disadvantage of D-P-VDAMP compared with D-FISTA is that it requires a second forward pass of the denoiser in order to compute the Monte-Carlo divergence estimate (6.2). To mitigate this, the denoiser could be trained to be approximately divergence-free by including (6.2) in the loss function during training. Then, when implementing D-P-VDAMP, there would be no need to perform Onsager correction; lines 7-9 of Algorithm 6 could be skipped. In other words, we would train the divergence free denoiser $\tilde{\boldsymbol{g}}(\boldsymbol{r}; \boldsymbol{\tau})$ directly. This would relocate the

computational burden of Onsager correction from implementation to training, and significantly reduce the per-iteration compute time of D-P-VDAMP.

A key concern for MR reconstruction methods based on deep learning is how the network deals with deviations between training and test data [227–229]. For instance, an unusually shaped tumour could be at risk of being smoothed over by a network that has been trained on data without such tumours. Using the unbiased output of D-P-VDAMP could mitigate this risk, as it is unbiased even if the denoising CNN deals with unusual cases poorly; the role of deep learning in D-P-VDAMP is only to seek to reduce the per-iteration aliasing of the unbiased estimate, but does not change its unbiased Gaussian nature. To explore this conjecture, future work includes evaluating the performance of D-P-VDAMP, especially the unbiased output, on datasets that feature challenging cases in the test set.

Appendix A

Magnitude of the product of a matrix and a random vector

This appendix proves a simple result on the entry-wise magnitude of a matrix and a random, zero-mean vector, which is used in the proofs of claims 4.3 and 5.1.

Claim A.1. Consider a deterministic matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ and a random vector $\boldsymbol{u} \in \mathbb{C}^n$ comprised of independent entries with $\mathbb{E}\{\boldsymbol{u}\} = \boldsymbol{0}_n$. Then

$$\mathbb{E}\{|\boldsymbol{A}\boldsymbol{u}|^2\} = |\boldsymbol{A}|^2 \mathbb{E}\{|\boldsymbol{u}|^2\} \tag{A.1}$$

where $|\cdot|$ is the entry-wise magnitude of a vector or matrix.

Proof. The ith entry of $|\boldsymbol{A}\boldsymbol{u}|^2$ is

$$\begin{aligned} |\sum_j A_{ij} u_j|^2 &= \left(\sum_j A_{ij} u_j \right) \left(\sum_l A_{il}^* u_l^* \right) \\ &= \sum_j \left(A_{ij} A_{ij}^* u_j u_j^* + \sum_{l \neq j} A_{ij} A_{il}^* u_j u_l^* \right). \end{aligned}$$

Since $\boldsymbol{u}$ is zero-mean with independent entries, $\mathbb{E}\{u_j u_l^*\} = 0$ when $l \neq j$. Therefore

$$\begin{aligned}\mathbb{E}\{|\sum_j A_{ij}u_j|^2\} &= \mathbb{E}\{\sum_j A_{ij}A_{ij}^* u_j u_j^*\} \\ &= \sum_j |A_{ij}|^2 \mathbb{E}\{|u_j|^2\},\end{aligned}$$

which is the ith entry of the right-hand-side of (A.1). This completes the proof. □

Appendix B

SURE for complex variables

B.1 Proof that cSURE is unbiased

This appendix proves Claim 4.4, which is restated in Claim B.1 below for readability.

Claim B.1. Consider a vector $\boldsymbol{v}_0 \in \mathbb{C}^{N_v}$ corrupted by coloured complex Gaussian noise with diagonal covariance: $\boldsymbol{v} = \boldsymbol{v}_0 + \mathcal{CN}(\mathbf{0}_{N_v}, \mathrm{Diag}(\boldsymbol{\tau}_v))$. Let $\boldsymbol{d}(\boldsymbol{v}) = \boldsymbol{v} + \boldsymbol{h}(\boldsymbol{v})$ be an estimator of $\boldsymbol{v}_0$. Complex SURE,

$$cSURE(\boldsymbol{d}(\boldsymbol{v})) = \|\boldsymbol{h}(\boldsymbol{v})\|_2^2 + \boldsymbol{\tau}_v^T[2\boldsymbol{\partial}(\boldsymbol{d}(\boldsymbol{v})) - \mathbf{1}_{N_v}], \tag{B.1}$$

is an unbiased estimate of the risk:

$$\mathbb{E}\{cSURE(\boldsymbol{d}(\boldsymbol{v}))\} = \mathbb{E}\{\|\boldsymbol{d}(\boldsymbol{v}) - \boldsymbol{v}_0\|_2^2\}.$$

Proof. This proof is a coloured, complex noise variation on the standard proof that SURE is an unbiased estimate of the risk, as found in, for instance, [16, 230]. The Euclidean distance between the ground truth $\boldsymbol{v}_0$ and the denoised vector $\boldsymbol{d}(\boldsymbol{v})$ can be

expanded as

$$\mathbb{E}\{\|\boldsymbol{d}(\boldsymbol{v})-\boldsymbol{v}_0\|_2^2\}=\mathbb{E}\{\|\boldsymbol{h}(\boldsymbol{v})\|_2^2+\|\boldsymbol{v}-\boldsymbol{v}_0\|_2^2 \\ -2(\Re[\boldsymbol{h}(\boldsymbol{v})]^T\Re[\boldsymbol{v}-\boldsymbol{v}_0]+\Im[\boldsymbol{h}(\boldsymbol{v})]^T\Im[\boldsymbol{v}-\boldsymbol{v}_0])\} \quad \text{(B.2)}$$

By the noise model of $\boldsymbol{v}$, the second term on the right hand side is

$$\mathbb{E}\{\|\boldsymbol{v}-\boldsymbol{v}_0\|_2^2\}=\boldsymbol{\tau}_v^T\mathbf{1}_N$$

By Stein's lemma [16] (see also (6) of [230]), and recalling that $\mathcal{CN}(\mathbf{0}_{N_v},\mathrm{Diag}(\boldsymbol{\tau}_v))$ is defined such that the variance of the real and imaginary parts is $\boldsymbol{\tau}_v/2$, the final term of the right-hand-side of (B.2) is

$$\begin{aligned}\mathbb{E}\{\Re[\boldsymbol{h}(\boldsymbol{v})]^T\Re[\boldsymbol{v}-\boldsymbol{v}_0]\}&=\mathbb{E}\left\{\sum_j\frac{\tau_{v,j}}{2}\frac{\partial\Re[h_j(\boldsymbol{v})]}{\partial\Re[v_j]}\right\}\\&=\mathbb{E}\left\{\sum_j\frac{\tau_{v,j}}{2}\left(N_v-\frac{\partial\Re[d_j(\boldsymbol{v})]}{\partial\Re[v_j]}\right)\right\}\end{aligned}$$

and similarly for the imaginary part. Overall, (B.2) is therefore

$$\begin{aligned}&\mathbb{E}\{\|\boldsymbol{d}(\boldsymbol{v})-\boldsymbol{v}_0\|_2^2\}\\=&\mathbb{E}\left\{\|\boldsymbol{h}(\boldsymbol{v})\|_2^2-\boldsymbol{\tau}_v^T\mathbf{1}_N+\sum_j\tau_{v,j}\left(\frac{\partial\Re[d_j(\boldsymbol{v})]}{\partial\Re[v_j]}+\frac{\partial\Im[d_j(\boldsymbol{v})]}{\partial\Im[v_j]}\right)\right\}\\=&\mathbb{E}\{\|\boldsymbol{h}(\boldsymbol{v})\|_2^2+\boldsymbol{\tau}_v^T[2\boldsymbol{\partial}(\boldsymbol{d}(\boldsymbol{v}))-\mathbf{1}_N]\}\end{aligned}$$

which is the expectation of cSURE stated in (B.1). This completes the proof. □

B.2 Complex soft thresholding with cSURE

The appendix proves Claim 4.5, which is restated in Claim B.2.

Claim B.2. When complex soft thresholding is employed as the denoiser, cSURE is

$$cSURE(\boldsymbol{\eta}(\boldsymbol{v};t)) = \sum_{j:|v_j|\leq t}^{N_v} (|v_j|^2 - \tau_{v,j}) + \sum_{j:|v_j|>t}^{N_v} \left(t^2 + \tau_{v,j} - \frac{t\tau_{v,j}}{|v_j|}\right). \tag{B.3}$$

Proof. Since the soft thresholding operator acts entry-wise, we can split cSURE in to regions where $|v_j|$ is less than and greater than the threshold:

$$\begin{aligned} &cSURE(\boldsymbol{\eta}(\boldsymbol{v})) \\ &= \sum_{j:|v_j|<t} [|h_j(\boldsymbol{v})|^2 + 2\tau_{v,j}\partial_j(\boldsymbol{d}(\boldsymbol{v}))] + \sum_{j:|v_j|>t} [|h_j(\boldsymbol{v})|^2 + 2\tau_{v,j}\partial_j(\boldsymbol{d}(\boldsymbol{v}))] - \boldsymbol{\tau}_v^T \mathbf{1}_{N_v}. \end{aligned} \tag{B.4}$$

If $|v_j| < t$, the soft thresholding operator sets the jth entry to 0. Substituting $\boldsymbol{d}(\boldsymbol{v}) = \mathbf{0}_{N_v}$ into cSURE yields a contribution of

$$\sum_{j:|v_j|<t} [|h_j(v)|^2 + 2\tau_{v,j}\partial_j(\boldsymbol{d}(\boldsymbol{v}))] = \sum_{j:|v_j|<t} |v_j|^2. \tag{B.5}$$

The alternative is $|v_j| > t$, where the soft thresholding takes the form $v_j(1 - \frac{t}{|v_j|})$. In this case, $|h_j(v_j)|^2 = t^2$, and

$$\partial_j(\boldsymbol{d}(v_j)) = 1 - \frac{t}{2|v_j|}, \qquad \text{for } |v_j| > t$$

which implies that

$$\sum_{j:|v_j|>t} [|h_j(\boldsymbol{v})|^2 + 2\tau_{v,j}\partial_j(\boldsymbol{d}(\boldsymbol{v}))] = \sum_{j:|v_j|>t} t^2 + \tau_{v,j}\left(2 - \frac{t}{|v_j|}\right). \tag{B.6}$$

Substituting (B.5) and (B.6) into (B.4) and using

$$\boldsymbol{\tau}_v^T \mathbf{1}_{N_v} = \sum_{j:|v_j|<t} \tau_{v,j} + \sum_{j:|v_j|>t} \tau_{v,j}$$

completes the proof. □

Bibliography

[1] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, "ESPIRiT-an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA," *Magnetic Resonance in Medicine*, vol. 71, pp. 990–1001, mar 2014.

[2] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, pp. 3142–3155, jul 2017.

[3] T. H. Kim and J. P. Haldar, "The Fourier radial error spectrum plot: A more nuanced quantitative evaluation of image reconstruction quality," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 61–64, 2018.

[4] G. V. Martinez, "Introduction to MRI Physics," in *Methods in molecular biology*, vol. 1718, pp. 3–19, 2018.

[5] J. Peter, "Medical imaging modalities - an introduction," in *Advanced Imaging in Biology and Medicine*, pp. 225–254, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[6] B. A. Chronik and B. K. Rutt, "A comparison between human magnetostimulation thresholds in whole-body and head/neck gradient coils.," *Magnetic Resonance in Medicine*, vol. 46, pp. 386–94, aug 2001.

[7] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, apr 2006.

[8] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, feb 2006.

[9] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, pp. 1182–1195, dec 2007.

[10] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice*. Springer US, 2002.

[11] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Springer New York, 2013, illustrate ed., 2013.

[12] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, 2009.

[13] A. Mousavi, A. Maleki, and R. G. Baraniuk, "Parameterless optimal Approximate Message Passing," *arXiv preprint arXiv:1311.0035*, oct 2013.

[14] C. Guo and M. E. Davies, "Near optimal compressed sensing without priors: Parametric SURE Approximate Message Passing," *IEEE Transactions on Signal Processing*, vol. 63, pp. 2130–2141, apr 2015.

[15] M. Bayati, M. A. Erdogdu, and A. Montanari, "Estimating LASSO risk and noise level," *Advances in Neural Information Processing Systems*, pp. 1–9, 2013.

[16] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, vol. 9, pp. 1135–1151, nov 1981.

[17] R. Ahmad, C. A. Bouman, G. T. Buzzard, S. Chan, S. Liu, E. T. Reehorst, and P. Schniter, "Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery," *IEEE Signal Processing Magazine*, vol. 37, pp. 105–116, jan 2020.

[18] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, pp. 4203–4215, dec 2005.

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

[20] G. M. Davis, S. G. Mallat, and Z. Zhang, "Adaptive time-frequency decompositions," *Optical engineering*, vol. 33, no. 7, pp. 2183–2191, 1994.

[21] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[22] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier analysis and Applications*, vol. 14, no. 5-6, pp. 629–654, 2008.

[23] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and computational harmonic analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[24] T. Blumensath and M. E. Davies, "Normalized iterative hard thresholding: Guaranteed stability and performance," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 298–309, 2010.

[25] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and computational harmonic analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[26] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of computational mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[27] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, jan 1998.

[28] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes rendus mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[29] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM journal on computing*, vol. 24, no. 2, pp. 227–234, 1995.

[30] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[31] B. Bah and J. Tanner, "Improved bounds on restricted isometry constants for Gaussian matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 5, pp. 2882–2898, 2010.

[32] H. Rauhut, "Compressive sensing and structured random matrices," *Theoretical foundations and numerical methods for sparse recovery*, vol. 9, pp. 1–92, 2010.

[33] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[34] P. Bloomfield and W. L. Steiger, *Least absolute deviations: theory, applications, and algorithms*. Springer, 1983.

[35] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete & Computational Geometry*, vol. 35, no. 4, pp. 617–652, 2006.

[36] D. L. Donoho and J. Tanner, "Neighborliness of randomly projected simplices in high dimensions," *Proceedings of the National Academy of Sciences*, vol. 102, no. 27, pp. 9452–9457, 2005.

[37] D. L. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," *Discrete & computational geometry*, vol. 43, no. 3, pp. 522–541, 2010.

[38] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, pp. 1413–1457, nov 2004.

[39] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[40] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, pp. 323–343, jan 2009.

[41] D. L. Donoho and I. M. Johnstone, "Minimax risk overl p-balls forl p-error," *Probability Theory and Related Fields*, vol. 99, no. 2, pp. 277–303, 1994.

[42] A. Montanari, *Graphical models concepts in compressed sensing*, p. 394–438. Cambridge University Press, 2012.

[43] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale modeling & simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[44] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5117–5144, 2016.

[45] D. L. Donoho, A. Maleki, and A. Montanari, "The noise-sensitivity phase transition in compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6920–6941, 2011.

[46] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: an empirical study," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pp. 467–475, 1999.

[47] J. Pearl, *Reverend Bayes on inference engines: A distributed hierarchical approach*. Cognitive Systems Laboratory, School of Engineering and Applied Science ..., 1982.

[48] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *IEEE Information Theory Workshop 2010 (ITW 2010)*, pp. 1–5, IEEE, jan 2010.

[49] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, pp. 764–785, feb 2011.

[50] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and message passing algorithms," *The Annals of Applied Probability*, vol. 25, pp. 753–822, apr 2015.

[51] S. Rangan, P. Schniter, and A. Fletcher, "On the convergence of approximate message passing with arbitrary matrices," in *2014 IEEE International Symposium on Information Theory*, pp. 236–240, IEEE, jun 2014.

[52] F. Caltagirone, L. Zdeborova, and F. Krzakala, "On convergence of approximate message passing," in *2014 IEEE International Symposium on Information Theory*, pp. 1812–1816, IEEE, jun 2014.

[53] S. Rangan, P. Schniter, E. Riegler, A. K. Fletcher, and V. Cevher, "Fixed points of generalized approximate message passing with arbitrary matrices," *IEEE Transactions on Information Theory*, vol. 62, pp. 7464–7474, dec 2016.

[54] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.

[55] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector Approximate Message Passing," *IEEE Transactions on Information Theory*, pp. 1–1, 2019.

[56] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 501–505, IEEE, 2017.

[57] N. Skuratovs and M. Davies, "Compressed sensing with upscaled Vector Approximate Message Passing," *arXiv preprint arXiv:2011.01369*, 2020.

[58] Z. Xue, J. Ma, and X. Yuan, "D-OAMP: A denoising-based signal recovery algorithm for compressed sensing," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 267–271, IEEE, dec 2016.

[59] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.

[60] C. Hegde, P. Indyk, and L. Schmidt, "A fast approximation algorithm for tree-sparse recovery," in *2014 IEEE International Symposium on Information Theory*, pp. 1842–1846, 2014.

[61] G. Peyré, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," in *European Conference on Computer Vision*, pp. 57–68, Springer, 2008.

[62] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, "Compressive sensing via nonlocal low-rank regularization," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3618–3632, 2014.

[63] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, pp. 945–948, 2013.

[64] S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, G. T. Buzzard, L. F. Drummy, J. P. Simmons, and C. A. Bouman, "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Transactions on Computational Imaging*, vol. 2, no. 4, pp. 408–423, 2016.

[65] A. Brifman, Y. Romano, and M. Elad, "Turning a denoiser into a super-resolver using plug and play priors," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 1404–1408, 2016.

[66] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Image restoration and reconstruction using variable splitting and class-adapted image priors," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3518–3522, 2016.

[67] P. Schniter, S. Rangan, and A. Fletcher, "Denoising based vector approximate message passing," *arXiv preprint arXiv:1611.01376*, 2016.

[68] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 16, pp. 2080–95, aug 2007.

[69] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "BM3D-AMP: A new image recovery algorithm based on BM3D denoising," in *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 3116–3120, IEEE, sep 2015.

[70] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3929–3938, 2017.

[71] S. Lefkimmiatis, "Universal denoising networks: a novel CNN architecture for image denoising," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3204–3213, 2018.

[72] C. Metzler, A. Mousavi, and R. Baraniuk, "Learned D-AMP: Principled neural network based compressive image recovery," in *Advances in Neural Information Processing Systems 30* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 1772–1783, Curran Associates, Inc., 2017.

[73] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play admm for image restoration: Fixed-point convergence and applications," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2016.

[74] K. Sung, B. L. Daniel, and B. A. Hargreaves, "Location constrained approximate message passing for compressed sensing MRI," *Magnetic Resonance in Medicine*, vol. 70, pp. 370–381, 2013.

[75] J. Ziniel and P. Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5270–5284, 2013.

[76] E. M. Eksioglu and A. K. Tanc, "Denoising AMP for MRI reconstruction: BM3D-AMP-MRI," *SIAM Journal on Imaging Sciences*, vol. 11, pp. 2090–2109, jan 2018.

[77] P. Schniter, S. Rangan, and A. Fletcher, "Plug-and-play image recovery using Vector AMP," BASP Frontiers Workshop 2017, 2017.

[78] M. H. Levitt, *Spin dynamics: basics of nuclear magnetic resonance*. John Wiley & Sons, 2013.

[79] L. G. Hanson, "Is quantum mechanics necessary for understanding magnetic resonance?," *Concepts in Magnetic Resonance Part A: An Educational Journal*, vol. 32, no. 5, pp. 329–340, 2008.

[80] R. W. Brown, Y. C. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan, *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley, 2014.

[81] F. Bloch, "Nuclear induction," *Physical review*, vol. 70, no. 7-8, p. 460, 1946.

[82] D. R. Messroghli, T. Niendorf, J. Schulz-Menger, R. Dietz, and M. G. Friedrich, "T1 mapping in patients with acute myocardial infarction: myocardial infarction and scar," *Journal of Cardiovascular Magnetic Resonance*, vol. 5, no. 2, pp. 353–359, 2003.

[83] A. M. Maceira, J. Joshi, S. K. Prasad, J. C. Moon, E. Perugini, I. Harding, M. N. Sheppard, P. A. Poole-Wilson, P. N. Hawkins, and D. J. Pennell, "Cardiovascular magnetic resonance in cardiac amyloidosis," *Circulation*, vol. 111, no. 2, pp. 186–193, 2005.

[84] P. Williamson, D. Pelz, H. Merskey, S. Morrison, S. Karlik, D. Drost, T. Carr, and P. Conlon, "Frontal, temporal, and striatal proton relaxation times in schizophrenic patients and normal comparison subjects.," *The American journal of psychiatry*, 1992.

[85] A. Pitkanen, M. Laakso, R. Kalviainen, K. Partanen, P. Vainio, M. Lehtovirta, P. Riekkinen, and H. Soininen, "Severity of hippocampal atrophy correlates with the prolongation of MRI T2 relaxation time in temporal lobe epilepsy but not in Alzheimer's disease," *Neurology*, vol. 46, no. 6, pp. 1724–1730, 1996.

[86] H. Larsson, J. Frederiksen, J. Petersen, A. Nordenbo, I. Zeeberg, O. Henriksen, and J. Olesen, "Assessment of demyelination, edema, and gliosis by in vivo determination of T1 and T2 in the brain of patients with acute attack of multiple sclerosis," *Magnetic Resonance in Medicine*, vol. 11, no. 3, pp. 337–348, 1989.

[87] D. H. Hoffman, A. Ayoola, D. Nickel, F. Han, H. Chandarana, and K. P. Shanbhogue, "T1 mapping, T2 mapping and MR elastography of the liver for detection and staging of liver fibrosis," *Abdominal Radiology*, vol. 45, no. 3, pp. 692–700, 2020.

[88] D. W. McRobbie, *MRI from picture to proton*. Cambridge University Press, 2006.

[89] P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza, and O. M. Mueller, "The NMR phased array," *Magnetic Resonance in Medicine*, vol. 16, no. 2, pp. 192–225, 1990.

[90] D. Hoult and P. C. Lauterbur, "The sensitivity of the zeugmatographic experiment involving human samples," *Journal of Magnetic Resonance (1969)*, vol. 34, no. 2, pp. 425–433, 1979.

[91] M. A. Bernstein, K. F. King, and X. J. Zhou, *Handbook of MRI pulse sequences*. Elsevier, 2004.

[92] Z.-P. Liang and P. C. Lauterbur, *Principles of magnetic resonance imaging: a signal processing perspective*. SPIE Optical Engineering Press, 2000.

[93] E. R. McVeigh, M. J. Bronskill, and R. M. Henkelman, "Phase and sensitivity of receiver coils in magnetic resonance imaging," *Medical physics*, vol. 13, no. 6, pp. 806–814, 1986.

[94] D. J. Larkman and R. G. Nunes, "Parallel magnetic resonance imaging," *Physics in Medicine & Biology*, vol. 52, no. 7, p. R15, 2007.

[95] J. B. Ra and C. Y. Rim, "Fast imaging using subencoding data sets from multiple detectors," *Magnetic Resonance in Medicine*, vol. 30, no. 1, pp. 142–145, 1993.

[96] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: sensitivity encoding for fast MRI.," *Magnetic Resonance in Medicine*, vol. 42, pp. 952–62, nov 1999.

[97] Y. Wang, "Description of parallel imaging in MRI using multiple coils," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 44, no. 3, pp. 495–499, 2000.

[98] A. Deshmane, V. Gulani, M. A. Griswold, and N. Seiberlich, "Parallel MR imaging," *Journal of Magnetic Resonance Imaging*, vol. 36, no. 1, pp. 55–72, 2012.

[99] 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*, vol. 47, pp. 1202–1210, jun 2002.

[100] H. Pedersen, S. Kozerke, S. Ringgaard, K. Nehrke, and W. Y. Kim, "k-t PCA: Temporally constrained k-t BLAST reconstruction using principal component analysis," *Magnetic Resonance in Medicine*, vol. 62, pp. 706–716, sep 2009.

[101] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Magnetic Resonance in Medicine*, vol. 46, pp. 638–651, oct 2001.

[102] B. Liu, K. King, M. Steckner, J. Xie, J. Sheng, and L. Ying, "Regularized sensitivity encoding (SENSE) reconstruction using Bregman iterations," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 61, no. 1, pp. 145–152, 2009.

[103] H. Omer and R. Dickinson, "Regularization in parallel MR image reconstruction," *Concepts in Magnetic Resonance Part A*, vol. 38, no. 2, pp. 52–60, 2011.

[104] S. Ramani and J. A. Fessler, "Parallel MR image reconstruction using augmented Lagrangian methods," *IEEE Transactions on Medical Imaging*, vol. 30, no. 3, pp. 694–706, 2010.

[105] M. Uecker, J. I. Tamir, F. Ong, and M. Lustig, "The BART Toolbox for Computational Magnetic Resonance Imaging," in *ISMRM*, 2016.

[106] S. B. King, S. M. Varosi, and G. R. Duensing, "Optimum SNR data compression in hardware using an Eigencoil array," *Magnetic Resonance in Medicine*, vol. 63, no. 5, pp. 1346–1356, 2010.

[107] S. King, S. Varosi, F. Huang, and G. Duensing, "The MRI eigencoil: 2N-channel SNR with N-receivers," in *Proc. 11th Annual Meeting of the International Society for Magnetic Resonance in Medicine, Toronto*, vol. 712, 2003.

[108] M. Buehrer, K. P. Pruessmann, P. Boesiger, and S. Kozerke, "Array compression for MRI with large coil arrays," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 57, no. 6, pp. 1131–1139, 2007.

[109] F. Huang, S. Vijayakumar, Y. Li, S. Hertel, and G. R. Duensing, "A software channel compression technique for faster reconstruction with many channels," *Magnetic Resonance Imaging*, vol. 26, no. 1, pp. 133–141, 2008.

[110] S. S. Vasanawala, M. T. Alley, B. A. Hargreaves, R. A. Barth, J. M. Pauly, and M. Lustig, "Improved pediatric MR imaging with compressed sensing," *Radiology*, vol. 256, no. 2, pp. 607–616, 2010.

[111] A. Hsiao, M. Lustig, M. T. Alley, M. Murphy, F. P. Chan, R. J. Herfkens, and S. S. Vasanawala, "Rapid pediatric cardiac assessment of flow and ventricular volume with compressed sensing parallel imaging volumetric cine phase-contrast mri," *American Journal of Roentgenology*, vol. 198, no. 3, pp. W250–W259, 2012.

[112] M. S. Cohen and R. M. Weisskoff, "Ultra-fast imaging," *Magnetic resonance imaging*, vol. 9, no. 1, pp. 1–37, 1991.

[113] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson, "Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI," *Magnetic Resonance in Medicine*, vol. 64, pp. 767–776, sep 2010.

[114] O. N. Jaspan, R. Fleysher, and M. L. Lipton, "Compressed sensing MRI: a review of the clinical literature," *The British Journal of Radiology*, vol. 88, p. 20150487, dec 2015.

[115] J. C. Ye, "Compressed sensing MRI: a review from signal processing perspective," *BMC Biomedical Engineering*, vol. 1, p. 8, dec 2019.

[116] D. Donoho, "How high-dimensional geometry is transforming the MRI industry," 2017.

[117] M. Murphy, M. Alley, J. Demmel, K. Keutzer, S. Vasanawala, and M. Lustig, "Fast l1-SPIRiT compressed sensing parallel imaging MRI: scalable parallel implementation and clinically feasible runtime.," *IEEE transactions on medical imaging*, vol. 31, pp. 1250–62, jun 2012.

[118] D. Liang, B. Liu, J. Wang, and L. Ying, "Accelerating SENSE using compressed sensing," *Magnetic Resonance in Medicine*, vol. 62, pp. 1574–1584, dec 2009.

[119] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magnetic Resonance in Medicine*, vol. 59, pp. 365–373, feb 2008.

[120] M. Lustig, J. Santos, D. Donoho, and J. Pauly, "k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proceedings of the 13th annual meeting of ISMRM, Seattle*, 2006.

[121] C. M. Tsai and D. G. Nishimura, "Reduced aliasing artifacts using variable-density k-space sampling trajectories," *Magnetic Resonance in Medicine*, vol. 43, no. 3, pp. 452–458, 2000.

[122] Z. Wang and G. R. Arce, "Variable density compressed image sampling," *IEEE Transactions on Image Processing*, vol. 19, pp. 264–270, jan 2010.

[123] N. Chauffert, P. Ciuciu, and P. Weiss, "Variable density compressed sensing in MRI. Theoretical vs heuristic sampling strategies," *Proceedings - International Symposium on Biomedical Imaging*, pp. 298–301, 2013.

[124] C. Lazarus, P. Weiss, N. Chauffert, F. Mauconduit, L. El Gueddari, C. Destrieux, I. Zemmoura, A. Vignaud, and P. Ciuciu, "SPARKLING: variable-density k-space filling curves for accelerated T2*-weighted MRI," *Magnetic Resonance in Medicine*, vol. 81, no. 6, pp. 3643–3661, 2019.

[125] S. S. Vasanawala, M. J. Murphy, M. T. Alley, P. Lai, K. Keutzer, J. M. Pauly, and M. Lustig, "Practical parallel imaging compressed sensing MRI: Summary of two years of experience in accelerating body MRI of pediatric patients," *Proceedings - International Symposium on Biomedical Imaging*, pp. 1039–1043, 2011.

[126] F. Krahmer and R. Ward, "Stable and robust sampling strategies for compressive imaging," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 612–622, 2014.

[127] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: a new theory for compressed sensing," *Forum of Mathematics, Sigma*, vol. 5, p. e4, feb 2017.

[128] J. G. Pipe and P. Menon, "Sampling density compensation in MRI: Rationale and an iterative numerical solution," *Magnetic Resonance in Medicine*, vol. 41, pp. 179–186, jan 1999.

[129] R. Bridson, "Fast Poisson disk sampling in arbitrary dimensions.," *SIGGRAPH sketches*, vol. 10, p. 1, 2007.

[130] M. Lustig, M. Alley, S. Vasanawala, D. Donoho, and J. Pauly, "Autocalibrating parallel imaging compressed sensing using L1 SPIRiT with Poisson-Disc sampling and joint sparsity constraints," in *ISMRM Workshop on Data Sampling and Image Reconstruction*, pp. 25–28, 2009.

[131] A. Hsiao, M. Lustig, M. T. Alley, M. J. Murphy, and S. S. Vasanawala, "Evaluation of valvular insufficiency and shunts with parallel-imaging compressed-sensing 4D phase-contrast MR imaging with stereoscopic 3D velocity-fusion volume-rendered visualization," *Radiology*, vol. 265, no. 1, pp. 87–95, 2012.

[132] S. Rapacchi, F. Han, Y. Natsuaki, R. Kroeker, A. Plotnik, E. Lehrman, J. Sayre, G. Laub, J. P. Finn, and P. Hu, "High spatial and temporal resolution dynamic contrast-enhanced magnetic resonance angiography using compressed sensing with magnitude image subtraction," *Magnetic Resonance in Medicine*, vol. 71, no. 5, pp. 1771–1783, 2014.

[133] T. Zhang, S. Chowdhury, M. Lustig, R. A. Barth, M. T. Alley, T. Grafendorfer, P. D. Calderon, F. J. Robb, J. M. Pauly, and S. S. Vasanawala, "Clinical performance of contrast enhanced abdominal pediatric MRI with fast combined parallel imaging compressed sensing reconstruction," *Journal of Magnetic Resonance Imaging*, vol. 40, no. 1, pp. 13–25, 2014.

[134] K. M. Johnson, S. B. Fain, M. L. Schiebler, and S. Nagle, "Optimized 3D ultrashort echo time pulmonary MRI.," *Magnetic Resonance in Medicine*, vol. 70, pp. 1241–1250, nov 2013.

[135] P. T. Gurney, B. A. Hargreaves, and D. G. Nishimura, "Design and analysis of a practical 3D cones trajectory," *Magnetic Resonance in Medicine*, vol. 55, no. 3, pp. 575–582, 2006.

[136] S. Wang, S. Tan, Y. Gao, Q. Liu, L. Ying, T. Xiao, Y. Liu, X. Liu, H. Zheng, and D. Liang, "Learning joint-sparse codes for calibration-free parallel MR imaging," *IEEE transactions on medical imaging*, vol. 37, no. 1, pp. 251–261, 2017.

[137] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 301–315, 2015.

[138] T. Nguyen-Duc, T. M. Quan, and W.-K. Jeong, "Frequency-splitting dynamic MRI reconstruction using multi-scale 3D convolutional sparse coding and automatic parameter selection," *Medical image analysis*, vol. 53, pp. 179–196, 2019.

[139] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[140] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE transactions on medical imaging*, vol. 30, no. 5, pp. 1028–1041, 2010.

[141] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[142] M. Chiew, S. M. Smith, P. J. Koopmans, N. N. Graedel, T. Blumensath, and K. L. Miller, "k-t FASTER: acceleration of functional MRI data acquisition using low rank constraints," *Magnetic Resonance in Medicine*, vol. 74, no. 2, pp. 353–364, 2015.

[143] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011.

[144] R. Otazo, E. Candes, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magnetic Resonance in Medicine*, vol. 73, no. 3, pp. 1125–1136, 2015.

[145] A. P. Yazdanpanah, O. Afacan, and S. Warfield, "Deep plug-and-play prior for parallel MRI reconstruction," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 3952–3958, IEEE, 2019.

[146] J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov, "Rare: Image reconstruction using deep priors learned without groundtruth," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1088–1099, 2020.

[147] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Proceedings of the 30th international conference on neural information processing systems*, pp. 10–18, 2016.

[148] Y. Yang, J. Sun, H. Li, and Z. Xu, "ADMM-CSNet: A deep learning approach for image compressive sensing," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 3, pp. 521–538, 2018.

[149] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3055–3071, 2018.

[150] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1828–1837, 2018.

[151] K. Kwon, D. Kim, and H. Park, "A parallel MR imaging method using multilayer perceptron," *Medical physics*, vol. 44, no. 12, pp. 6209–6224, 2017.

[152] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang, "Accelerating magnetic resonance imaging via deep learning," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 514–517, IEEE, 2016.

[153] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, 2018.

[154] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1488–1497, 2018.

[155] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE transactions on medical imaging*, vol. 38, no. 1, pp. 167–179, 2018.

[156] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[157] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization.," *Journal of machine learning research*, vol. 12, no. 7, 2011.

[158] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[159] Y. LeCun, Y. Bengio, *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[160] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[161] D. Liang, J. Cheng, Z. Ke, and L. Ying, "Deep MRI reconstruction: Unrolled optimization algorithms meet neural networks," *arXiv preprint arXiv:1907.11711*, 2019.

[162] F. Knoll, K. Hammernik, C. Zhang, S. Moeller, T. Pock, D. K. Sodickson, and M. Akcakaya, "Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 128–140, 2020.

[163] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[164] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pp. 448–456, JMLR.org, 2015.

[165] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[166] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*, pp. 630–645, Springer, 2016.

[167] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1492–1500, 2017.

[168] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, 2017.

[169] W. Lu and N. Vaswani, "Modified compressive sensing for real-time dynamic MR imaging," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 3045–3048, IEEE, 2009.

[170] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4595–4607, 2010.

[171] M. Lebrun, "An analysis and implementation of the BM3D image denoising method," *Image Processing On Line*, vol. 2, pp. 175–213, aug 2012.

[172] C. Millard, A. T. Hess, B. Mailhe, and J. Tanner, "Approximate message passing with a colored aliasing model for variable density Fourier sampled images," *IEEE Open Journal of Signal Processing*, p. 1, 2020.

[173] C. Millard, A. T. Hess, B. Mailhe, and J. Tanner, "An approximate message passing algorithm for rapid parameter-free compressed sensing MRI," in *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 91–95, 2020.

[174] H. Kim, D. U. Kang, and S. Y. Chun, "Preliminary Studies On Training And Fine-Tuning Deep Denoiser Neural Networks In Learned D-Amp For Undersampled Real Mr Measurements," in *2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops)*, pp. 1–4, 2020.

[175] X. Qiao, J. Du, L. Wang, Z. He, and Y. Jia, "A model-based deep network for MRI reconstruction using approximate message passing algorithm," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1105–1109, 2020.

[176] S. Sarkar, R. Ahmad, and P. Schniter, "MRI image recovery using damped denoising vector AMP," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8108–8112, IEEE, 2021.

[177] S. Huang, J. J. Lah, J. W. Allen, and D. Qiu, "A Bayesian approach for T2* mapping and quantitative susceptibility mapping," *arXiv preprint arXiv:2103.05535*, 2021.

[178] C. A. Metzler and G. Wetzstein, "D-VDAMP: Denoising-based approximate message passing for compressive MRI," *arXiv:2010.13211*, 2020.

[179] A. Beck and M. Teboulle, "Fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, pp. 183–202, jan 2009.

[180] B. Goossens, J. Aelterman, H. Luong, A. Pizurica, and W. Philips, "Wavelet-based analysis and estimation of colored noise," in *Discrete Wavelet Transforms - Algorithms and Applications* (H. Olkkonen, ed.), ch. 15, pp. 255–280, Intech, 2011.

[181] T. Li, M. Wang, and W. Xiong, "A novel method for filtering of Gaussian colored noise in images with wavelet transform," *ICEIE 2010 - 2010 International Conference on Electronics and Information Engineering, Proceedings*, vol. 1, no. Iceie, pp. 184–189, 2010.

[182] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, vol. 59, no. 2, pp. 319–351, 1997.

[183] M. Benzi, "Preconditioning techniques for large linear systems: a survey," *Journal of computational Physics*, vol. 182, no. 2, pp. 418–477, 2002.

[184] G. Puy, P. Vandergheynst, and Y. Wiaux, "On variable density compressive sampling," *IEEE Signal Processing Letters*, vol. 18, pp. 595–598, oct 2011.

[185] A. Hsiao, M. Lustig, M. T. Alley, M. Murphy, F. P. Chan, R. J. Herfkens, and S. S. Vasanawala, "Rapid pediatric cardiac assessment of flow and ventricular volume with compressed sensing parallel imaging volumetric cine phase-contrast MRI," *American Journal of Roentgenology*, vol. 198, pp. W250–W259, mar 2012.

[186] K. G. Hollingsworth, D. M. Higgins, M. McCallum, L. Ward, A. Coombs, and V. Straub, "Investigating the quantitative fidelity of prospectively undersampled chemical shift imaging in muscular dystrophy with compressed sensing and parallel imaging reconstruction," *Magnetic Resonance in Medicine*, vol. 72, pp. 1610–1619, dec 2014.

[187] T. Akasaka, K. Fujimoto, T. Yamamoto, T. Okada, Y. Fushumi, A. Yamamoto, T. Tanaka, and K. Togashi, "Optimization of regularization parameters in compressed sensing of magnetic resonance angiography: Can statistical image metrics mimic radiologists' perception?," *PLOS ONE*, vol. 11, pp. 1–14, 01 2016.

[188] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-Based Compressive Sensing," *IEEE Transactions on Information Theory*, vol. 56, pp. 1982–2001, apr 2010.

[189] C. Vonesch and M. Unser, "A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 539–549, 2008.

[190] I. Bayram and I. W. Selesnick, "A subband adaptive iterative shrinkage/thresholding algorithm," *IEEE Transactions on Signal Processing*, vol. 58, no. 3 PART 1, pp. 1131–1143, 2010.

[191] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, p. 1200, dec 1995.

[192] K. Khare, C. J. Hardy, K. F. King, P. A. Turski, and L. Marinelli, "Accelerated MR imaging using compressive sensing with no free parameters," *Magnetic Resonance in Medicine*, vol. 68, pp. 1450–1457, nov 2012.

[193] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk, "Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP)," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4290–4308, 2013.

[194] M. Guerquin-Kern, M. Haberlin, K. P. Pruessmann, and M. Unser, "A fast wavelet-based reconstruction method for magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol. 30, no. 9, pp. 1649–1660, 2011.

[195] Y. Mäkinen, L. Azzari, E. Sánchez-Monge, M. Maggioni, A. Danielyan, K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image and video denoising by sparse 3D transform-domain collaborative filtering," 2020.

[196] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno, and Others, "fastMRI: An open dataset and benchmarks for accelerated MRI," *arXiv preprint arXiv:1811.08839*, 2018.

[197] G. Varela-Mattatall, C. A. Baron, and R. S. Menon, "Automatic determination of the regularization weighting for wavelet-based compressed sensing MRI reconstructions," *Magnetic Resonance in Medicine*, 2021.

[198] F. Ong, J. Y. Cheng, and M. Lustig, "General phase regularized reconstruction using phase cycling," *Magnetic Resonance in Medicine*, vol. 80, no. 1, pp. 112–125, 2018.

[199] T. Akasaka, K. Fujimoto, T. Yamamoto, T. Okada, Y. Fushumi, A. Yamamoto, T. Tanaka, and K. Togashi, "Optimization of regularization parameters in compressed sensing of magnetic resonance angiography: can statistical image metrics mimic radiologists' perception?," *PloS one*, vol. 11, no. 1, p. 146548, 2016.

[200] P. Kalavathi and V. S. Prasath, "Methods on skull stripping of MRI head scan images: a review," *Journal of digital imaging*, vol. 29, no. 3, pp. 365–379, 2016.

[201] B. Husson, G. Rodesch, P. Lasjaunias, M. Tardieu, and G. Sébire, "Magnetic resonance angiography in childhood arterial brain infarcts: a comparative study with contrast angiography," *Stroke*, vol. 33, no. 5, pp. 1280–1285, 2002.

[202] F. Nüssel, H. Wegmüller, and P. Huber, "Comparison of magnetic resonance angiography, magnetic resonance imaging and conventional angiography in cerebral arteriovenous malformation," *Neuroradiology*, vol. 33, no. 1, pp. 56–61, 1991.

[203] 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*, vol. 13, no. 4, pp. 600–612, 2004.

[204] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[205] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo Sure: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Transactions on Image Processing*, vol. 17, pp. 1540–1554, sep 2008.

[206] C. Li, W. Yin, and Y. Zhang, "User's guide for TVAL3: TV minimization by augmented lagrangian and alternating direction algorithms," *CAAM report*, vol. 20, no. 46-47, p. 4, 2009.

[207] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM Journal on Imaging Sciences*, vol. 10, pp. 1804–1844, jan 2017.

[208] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-Play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, pp. 945–948, IEEE, dec 2013.

[209] A. Sriram, J. Zbontar, T. Murrell, A. Defazio, C. L. Zitnick, N. Yakubova, F. Knoll, and P. Johnson, "End-to-end variational networks for accelerated MRI reconstruction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 64–73, Springer, 2020.

[210] Z. Ramzi, J. Starck, and P. Ciuciu, "XPDNet for MRI Reconstruction: An application to the 2020 fastMRI challenge," in *2021 ISMRM annual meeting*, vol. 275, 2021.

[211] E. M. Eksioglu, "Decoupled algorithm for MRI reconstruction using nonlocal block matching model: BM3D-MRI," *Journal of Mathematical Imaging and Vision*, vol. 56, no. 3, pp. 430–440, 2016.

[212] U. S. Kamilov, H. Mansour, and B. Wohlberg, "A plug-and-play priors approach for solving nonlinear imaging inverse problems," *IEEE Signal Processing Letters*, vol. 24, no. 12, pp. 1872–1876, 2017.

[213] J. Bobin, J. L. Starck, and R. Ottensamer, "Compressed sensing in astronomy," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 718–726, 2008.

[214] Y. Wiaux, L. Jacques, G. Puy, A. M. Scaife, and P. Vandergheynst, "Compressed sensing imaging techniques for radio interferometry," *Monthly Notices of the Royal Astronomical Society*, vol. 395, no. 3, pp. 1733–1742, 2009.

[215] J. L. Starck and J. Bobin, "Astronomical data analysis and sparsity: From wavelets to compressed sensing," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1021–1030, 2010.

[216] J. Song, Q. H. Liu, G. A. Johnson, and C. T. Badea, "Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT," *Medical physics*, vol. 34, no. 11, pp. 4476–4483, 2007.

[217] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine and Biology*, vol. 53, pp. 4777–4807, aug 2008.

[218] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Medical physics*, vol. 35, no. 2, pp. 660–663, 2008.

[219] A. Perelli, M. Lexa, A. Can, and M. E. Davies, "Compressive computed tomography reconstruction through denoising approximate message passing," *SIAM Journal on Imaging Sciences*, vol. 13, no. 4, pp. 1860–1897, 2020.

[220] R. Kitichotkul, C. A. Metzler, F. Ong, and G. Wetzstein, "SUREMap: Predicting uncertainty in CNN-based image reconstruction using Stein's Unbiased Risk Estimate," *arXiv preprint arXiv:2010.13214*, 2020.

[221] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.

[222] C. Eldeniz, W. Gan, S. Chen, T. J. Fraum, D. R. Ludwig, Y. Yan, J. Liu, T. Vahle, U. Krishnamurthy, U. S. Kamilov, *et al.*, "Phase2phase: Respiratory motion-resolved reconstruction of free-breathing magnetic resonance imaging using deep learning without a ground truth for improved liver imaging," *Investigative Radiology*, 2021.

[223] E. K. Cole, J. M. Pauly, S. S. Vasanawala, and F. Ong, "Unsupervised MRI reconstruction with generative adversarial networks," *arXiv preprint arXiv:2008.13065*, 2020.

[224] Z. Ke, J. Cheng, L. Ying, H. Zheng, Y. Zhu, and D. Liang, "An unsupervised deep learning method for multi-coil cine MRI," *Physics in Medicine & Biology*, vol. 65, no. 23, p. 235041, 2020.

[225] J. I. Tamir, S. X. Yu, and M. Lustig, "Unsupervised deep basis pursuit: Learning inverse problems without ground-truth data," *arXiv preprint arXiv:1910.13110*, 2019.

[226] S. Soltanayev and S. Y. Chun, "Training deep learning based denoisers without ground truth data," *arXiv preprint arXiv:1803.01314*, 2018.

[227] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen, "On instabilities of deep learning in image reconstruction and the potential costs of AI," *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30088–30095, 2020.

[228] D. Liang, J. Cheng, Z. Ke, and L. Ying, "Deep magnetic resonance image reconstruction: Inverse problems meet neural networks," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 141–151, 2020.

[229] F. Knoll, K. Hammernik, E. Kobler, T. Pock, M. P. Recht, and D. K. Sodickson, "Assessment of the generalization of learned image reconstruction and the potential for transfer learning," *Magnetic Resonance in Medicine*, vol. 81, no. 1, pp. 116–128, 2019.

[230] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, 2007.