Patient-adaptive compressed sensing for MRI

Lee Gonzales (Vrije Universiteit Brussel), Koos Huijssen (VORTech), Rakesh Jha (Eindhoven University of Technology), Tristan van Leeuwen (Utrecht University), Alessandro Sbrizzi (UMC Utrecht), Willem van Valenberg (Utrecht University), Ian Zwaan (Eindhoven University of Technology), Johan van den Brink (Philips Healthcare), Elwin de Weerdt (Philips Healthcare)

Abstract

The theory of compressed sensing (CS) promises reconstruction of sparse signals with a sampling rate below the Nyquist criterion by taking randomized measurements under certain assumptions about the structure of the signal. Current compressed sensing techniques applied to magnetic resonance imaging (MRI) require measurements concentrated heavily around the center of the Fourier space in order to yield somewhat usable results. PHILIPS Healthcare is interested in the potential time-saving benefits of compressed sensing applied to MRI, but requires robust and accurate reconstruction results. During the 106th European Study Group for Mathematics and Industry, PHILIPS challenged us to improve upon existing Fourier space sampling patterns for compressed sensing. The patterns could, for example, include patient dependent prior information. We demonstrate (experimentally) that current CS-MRI techniques lack a sufficient amount of the property called *incoherence*. Incoherence is a measure of correlation between the measurement matrix and the basis in which the signal is sparse. A compressed sensing method without sufficient incoherence results in sub-optimal performance in terms of scan-time and/or image quality. By introducing the necessary incoherence, biased sampling in the Fourier space is no longer necessary. Increasing incoherence while keeping the similar sparsity level in a practical setting is not as straightforward. In this paper, we demonstrate an off-line approach based on prior patient information and evaluate the results with the structural similarity index (SSIM) as a measure of image reconstruction quality. The developed strategy provides an improved signal basis such that both scan-time reduction and good image reconstruction are attained.

1 Introduction

Magnetic resonance imaging (MRI) is the imaging modality of choice for diagnosing a vast range of diseases, including multiple sclerosis and cancer. Compared to X-ray based techniques, such as computerized tomography (CT), MRI provides superior soft-tissue contrast and does not expose the patient to ionizing radiation. Unfortunately, a typical MRI exam can take over 30 minutes (as compared to 5 minutes for a CT scan). One way to reduce the scan time is to collect fewer measurements. Current image reconstruction techniques, however, require a sampling strategy which obeys the Nyquist criterion, in order to reconstruct a usable image. Recent developments in the field of *compressed sensing* (CS) promise a dramatic reduction of the number of measurements needed to reconstruct an image, as long as the measurement process is *incoherent* and the image can be represented in terms of relatively few basis terms (*sparsity*).

PHILIPS Healthcare is one of the major MRI scanner manufacturers in the world and it is interested in optimization of the CS-MRI paradigm. During the SWI 2015, PHILIPS would like to investigate how the CS-MRI experiment can be optimally performed, in the sense that an accurate image is reconstructed from data obtained within the shortest possible scan-time. Some information about the patient is available prior to the scan thus it can be used to achieve this goal. As PHILIPS suggests, the approach should be *patient-dependent*, that is, the existing data has to be quickly processed to determine the scanner setup for the new scan.

In this report, we discuss how we combine CS and MRI in order to develop reconstruction algorithms that exploit previously acquired patient-specific information. The report is organized as follows. First, we review the mathematics behind basic MRI image reconstruction and compressed sensing and identify a possible bottleneck for the straightforward application of CS theory to MRI imaging. Then, we delineate a strategy to solve this problem and we show some preliminary results we have obtained during the SWI 2015.

2 Preliminaries

2.1 Magnetic resonance imaging

The simplest mathematical model for the MRI imaging process can be stated as follows. We are interested in reconstructing the transverse component of the so-called *spin magnetization vector*, which is tissue-dependent, and thus reveals internal structures. We represent the transverse magnetization of a 2D slice by a function $u(x, y)$ with $x, y \in [0, 1]$. In practice, u is a complex function but for simplicity, we assume $u \in \mathbb{R}$. The work presented in this report can be easily extended to the complex case.

In MRI, we can measure the function u only in the Fourier domain, also called k-space. The 2D Fourier series of u is denoted by

$$\widehat{u}_{kl} = \int_0^1 \int_0^1 \mathrm{d}x \, \mathrm{d}y \, u(x, y) e^{\imath 2\pi(kx+ly)},$$

for $k, l \in \mathbb{Z}$. In practice, only the coefficients up to some maximum bandwidth $-B < k, l < B$ are measured, allowing us to reconstruct an $N \times N$ discrete representation of the transverse magnetization $u_{ij} = u(i/(N-1), j/(N-1))$ with $0 \leq i, j < N$. Given N, the Nyquist criterion determines the value of B which makes a correct reconstruction of u from its frequency coefficients possible.

By organizing all the coefficients in vectors, we can state the MRI measurement process as follows

$$\widehat{\mathbf{u}} = F\mathbf{u},$$

where $\mathbf{u} \in \mathbb{R}^n$, $\widehat{\mathbf{u}} \in \mathbb{C}^n$, $F \in \mathbb{C}^{n \times n}$ represents the 2D discrete Fourier transform and $n = N^2$ denotes tot total number of pixels. A schematic depiction of the process is shown in figure 1.

Since all the Fourier samples need to be measured sequentially, the time needed to acquire them scales quadratically with the required resolution. In the next section we will discuss an alternative sampling paradigm that promises a more favorable scaling assuming that $\mathbf{u}$ exhibits some additional structure.

2.2 Compressed sensing

The basic idea behind compressed sensing is that we can uniquely solve an underdetermined system $A\mathbf{x} = \mathbf{b}$ given that the solution we seek is *sparse* (i.e., has only a few non-zero elements) and the matrix A satisfies the so-called *restricted isometry property* (RIP).

Definition 2.1. A vector $\mathbf{x} \in \mathbb{R}^n$ is $k-$sparse when it has at most k non-zero elements.

Definition 2.2. A matrix $A \in \mathbb{R}^{m\times n}$ satisfies the restricted isometry property RIP(k,δ_k) if for every k-sparse vector $\mathbf{x}$ there exists a constant δ_k such that

$$(1-\delta_k)\|\mathbf{x}\|_2 \leq \|A\mathbf{x}\|_2 \leq (1+\delta_k)\|\mathbf{x}\|_2.$$

When the measurement matrix A satisfies the RIP and the solution x is sparse, or well approximated by a sparse solution, then one can solve the so-called *basis pursuit denoise (BPDN)* problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|A\mathbf{x}-\mathbf{b}\|_2 \leq \sigma.$$

Here, $\sigma \geq 0$ is the noise level of the measurement $\mathbf{b}$.

With these definitions we can now state the following theorem by Candès [3] regarding the recoverability of a sparse signal from noisy measurements.

Theorem 2.1. *Let the matrix A satisfies RIP($2k$,δ_{2k}) with $\delta_{2k} < \sqrt{2}-1$, and $\mathbf{b} = A\mathbf{x}+\mathbf{n}$ for given signal $\mathbf{x}$ and $\|\mathbf{n}\|_2 \leq \epsilon$. Then, the error between the solution $\overline{\mathbf{x}}$ of the BPDN problem and the true signal $\mathbf{x}$ is bounded as follows:*

$$\|\overline{\mathbf{x}}-\mathbf{x}\|_2 \leq C_0\|\mathbf{x}-\mathbf{x}_k\|_1/\sqrt{k} + C_1\epsilon,$$

where C_0 and C_1 are positive constants and $\mathbf{x}_k$ is the best k-sparse approximation to $\mathbf{x}$. Thus, if the given signal $\mathbf{x}$ is k-sparse, we have $\|\overline{\mathbf{x}}-\mathbf{x}\|_2 \leq C_1\epsilon$

With overwhelming probability, certain types of random matrices (e.g., matrices whose elements are i.i.d. Gaussian) satisfy the required RIP property when

$$m \geq Ck\log(n),$$

where C is a problem-specific constant [3]. For our MRI problem, this would mean that the number of measurements is no longer driven by the resolution but by the sparsity of the signal. It is in practice not feasible to check whether a given matrix satisfies RIP. Instead, one typically considers the *coherence*.

In practical applications, the signal of interest is not sparse itself, but admits a sparse representation in some orthonormal basis $\Psi \in \mathbb{C}^{n\times n}$. Modelling the measurement process as taking inner products of the signal with m rows of an orthonormal basis $\Phi \in \mathbb{C}^{n\times n}$, we can express the sensing matrix as $A = R\Phi\Psi$, where $R \in \mathbb{R}^{m\times n}$ selects m rows at random. The resulting matrix A is a suitable RIP when the *mutual coherence* between Ψ and Φ is low.

Definition 2.3. The mutual coherence of two orthonormal bases Ψ and Φ is defined as

$$\mu(\Psi, \Phi) = \max_{1\leq i,j\leq n} |(\Psi^T\Phi)_{ij}|,$$

Generally, the lower the coherence, the lower the RIP constant and the fewer measurements we expect to need in order to recover a given sparse signal. Note that for orthonormal bases we have $1 \leq \mu \leq \sqrt{n}$. In the remainder of the paper we will use the coherence as a heuristic to gauge how well a given pair (Ψ, Φ) is expected to perform.

2.3 Sparse recovery

There are a number of algorithms for solving the BPDN problem, most of which are based on one of two equivalent reformulations. The first is quadratic formulation of the problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 + \lambda\|A\mathbf{x} - \mathbf{x}\|_2,$$

and the second is the Lasso problem

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \leq \tau.$$

For a given σ, there exists a unique λ and τ such that the solutions of all three problems coincide [5, 2]. The relation between these parameters is given by the *Pareto curve*. This is illustrated in figure 2.

Finding these parameters λ or τ is not trivial, however, and typically relies on some sort of continuation method. A very elegant way of finding a τ corresponding to a given σ is described by [5]. Essentially, they develop a root-finding method to traverse the Pareto curve. The Lasso subproblems are solved via a projected gradient algorithm and is suitable for large-scale problems and complex data.

3 Approach and results

We have seen that the two important ingredients in CS are sparsity and coherence. The image needs to be as sparse as possible while the coherence needs to be as low as possible. The goal is to leverage these results to reduce the number of measurements needed to recover the image $\mathbf{u}$ from Fourier measurements

$$\mathbf{b} = RF\mathbf{u},$$

where $R \in \mathbb{R}^{m \times n}$ with $m < n$ is a restriction matrix that subsamples the full Fourier measurements. Since we are interested in taking as few measurements as possible, we want m to be as small as possible. Introducing the subsampling ratio $\rho = n/m$, the potential speedup of the measurement process is proportional to ρ.

A typical image is not sparse in the natural pixel basis. Instead, we need to find a basis in which the image can be sparsely represented. A common choice is wavelets, denoted here by a matrix W. The sparsity is illustrated in figure 3.

The recovery problem now is to find wavelet coefficients $\mathbf{z} = W\mathbf{x}$ such that $RFW^T\mathbf{z} \approx \mathbf{b}$. A problem with this approach is that the mutual coherence $\mu(F, W^T)$ is quite high, as illustrated in figure 4. A remedy to this is to insert a random ± 1 diagonal matrix in the measurement process and collect measurements as $\mathbf{b} = RFS\mathbf{u}$. Here, $S = \mathsf{diag}(\mathbf{s})$ where the $s_i \in \{+1, -1\}$ are i.i.d. Rademacher random variables [1, 4]. The mutual coherence $\mu(FS, W^T)$ is much lower, as illustrated in figure 4.

Reconstructions with and without S for homogeneous random sampling are presented in figure 5. We clearly see that incorporating the matrix S spreads the information more evenly over the Fourier space so that uniform random sampling makes more sense. Comparing the reconstruction quality in terms of the Structural Similarity Index Measure (SSIM) [6] for different subsampling ratios in figure 6 we see that incorporating S allows for a higher subsampling factor.

Unfortunately, it is not feasible in practice to incorporate S in the sampling as it would entail randomly perturbing the object prior to taking the measurement. Therefore, we will take a slightly different view on the problem.

3.1 Breaking the coherence

Instead of changing the measurement process, we will aim to find a basis $\widetilde{W}$ that reduces the mutual coherence. An obvious candidate is $\widetilde{W} = SW$, however, typical images are much less sparse in this new basis as illustrated in figure 7.

This leads us to two extreme cases: for $\mathbf{s} = \mathbf{1}$ we have a very sparse representation and a high coherence while for uniformly random $s_i \in \{-1, +1\}$ we have a low coherence and insufficient sparsity. This is shown in figure 8. The question is whether there exists a mask $\mathbf{s}$ that achieves an "optimal" trade-off between the two extremes. To investigate this we take the s_i to be correlated Rademacher variables or random checkerboard patterns and vary the scale to obtain a natural continuation from one extreme to the other. A few examples are shown in figure 9. For these matrices, we compute the coherence and the sparsity and plot them. Figure 10 shows that they indeed trace out a trade-off curve as argued earlier. Figure 11 shows the reconstruction quality for these various masks. Unfortunately, the partially coherent masks perform only marginally better for small subsampling ratios than the other two.

3.2 Experimental design

The goal of this section is to see if we can de better when we explicitly design S to minimize the coherence while maintaining the sparsity of a reference image $\mathbf{u}$. We could formulate this problem as

$$\min_S \mu(F, SW^T) \quad \text{s.t.} \quad \|WS^{-1}\mathbf{u}\|_1 \leq (1+\kappa)\|W\mathbf{u}\|_1,$$

with κ small. The inequality constraint ensures that the application of the new transform WS^{-1} to $\mathbf{u}$ gives a sparse representation.

If we take S to be a diagonal matrix $S = \mathbf{diag}(\mathbf{s})$ we can formulate this problem as

$$\min_{\mathbf{s}} \|A\mathbf{s}\|_\infty \quad \text{s.t.} \quad \|U(\mathbf{s})\|_1 \leq (1+\kappa), \tag{1}$$

where A is the matrix representation of the linear operation $F^T\mathsf{diag}(\mathbf{s})W^T$ (i.e., $A\mathbf{s} = \mathsf{vec}(F^T\mathsf{diag}(\mathbf{s})W^T)$) and $U(\mathbf{s}) = \frac{1}{\|W\mathbf{u}\|_1} W\mathbf{v}$ where $v_i \equiv 1/s_i$.

To asses the feasibility of this approach, we solve this optimization problem for a 16×16 reference image, shown in figure 12 and denoted by *True image*.

We use a black-box non-linear optimization routine in Matlab (`fmincon`) to solve the optimization problem. The starting value for $\mathbf{s}$ is the vector whose components follow a Rademacher distribution as introduced above, that is: $s_i \in \{1, -1\}$. This is shown in figure 12. The algorithm is manually halted after 40 iteration, when the convergence reaches a plateau. The convergence history is shown in figure 13.

The resulting solution is shown in figure 12 and it is denoted by *Optimized S*. Interestingly, the result has a similar structure as the reference image. This can be understood as follows. The reference image is multiplied point-wise with $\mathbf{s}$. If we take $s_i \approx u_i^{-1}$, the resulting normalized image will be almost constant, thus allowing for a very sparse approximation using only the coarsest scale wavelets.

We perform CS reconstructions of the test image with $N = 16$ and reduction factor 2 for the sparsity transforms W (standard approach) and WS^{-1}, respectively. To appreciate the improvement obtained by the experimental design algorithm, we also consider the reconstruction with the starting value for S, that is, the Rademacher distributed values.

The reconstructed images are shown in figure 12, bottom row. We consider the relative error given by $\|\mathbf{u} - \mathbf{u}^r\|_2 / \|\mathbf{u}\|_2 \times 100\%$ where the superscript r denotes the reconstructed image and we report the error value under the corresponding plots. Note the drastic reduction in the error when the optimized S is used.

4 Discussion

We have seen that the coherence between the Fourier transform and wavelets leads to suboptimal performance of CS-type reconstructions and we have delineated a strategy to modify the Wavelet transform by means of the S matrix. The resulting transform maintains the sparsity and at the same time minimizes the coherence with the sampling operator RF. This trade-off solution gives excellent results in terms of CS reconstructions.

Note that the steps (experiment design and CS reconstructions) can be performed *off-line*, that is, after the end of the MRI exam. In this way, there is no surcharge of time for the clinical protocol, a major drawback for *on-line* design methods.

Our approach relies on the knowledge of a reference image. The question is, of course, how this approach will perform when the reference image is not the same as the true image. We expect minimal problems when the reference image is a previously acquired scan, a set of reference images, or other prior information. Alternatively, we could perform a first reconstruction by standard CS and use the resulting image as reference for designing S. This step could be repeated until no improvement is obtained with respect to the previously reconstructed image.

5 Conclusions

We have analyzed the pitfalls of CS applied to MRI and we have presented an innovative approach to improve the reconstructions. The large scale optimization problem can be performed off-line, making the way to the clinic potentially short.

References

[1] N. Ailon and B. Chazelle. The Fast Johnson-Lindenstrauss Transform and Approximate Nearest Neighbors. *SIAM Journal on Computing*, 39(1):302–322, 2009. doi: 10.1137/060673096. URL `http://dx.doi.org/10.1137/060673096`.

[2] A. Aravkin, J. Burke, and M. Friedlander. Variational properties of value functions. *SIAM Journal on optimization*, 23:1689–1717, Nov. 2013. ISSN 1052-6234. doi: 10.1137/120899157. URL `http://arxiv.org/abs/1211.3724http://epubs.siam.org/doi/abs/10.1137/120899157`.

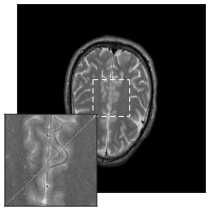
[3] E. Candes and M. Wakin. An Introduction To Compressive Sampling. *IEEE Signal Processing Magazine*, 25(March 2008):21–30, 2008. ISSN 1053-5888. doi: 10.1109/MSP.2007.914731.

[4] F. Krahmer and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property. *ArXiv e-prints*, Sept. 2010.

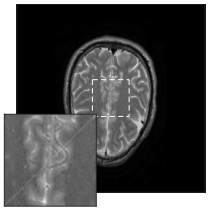
[5] E. Van Den Berg and M. Friedlander. PROBING THE PARETO FRONTIER FOR BASIS PURSUIT SOLUTIONS. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008. ISSN 1064-8275.

URL `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.161.9332\&rep=rep1\&type=pdf`.

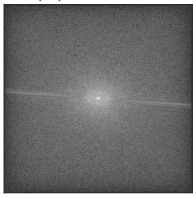
[6] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli. Wavelets for Image Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.



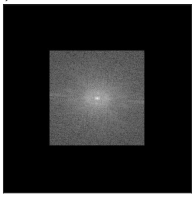
(1) Object



(4) Reconstruction



(2) full Fourier spectrum



(3) Bandlimited measurements

Figure 1: Schematic depiction of the MRI process. The ground object (1) is sampled in the Fourier domain (2), yielding a set of bandlimited measurements (3) from which we can reconstruct using a discrete inverse Fourier transform (4). Note the slight loss in resolution caused by the bandlimited nature of the measurements.

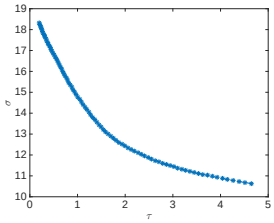
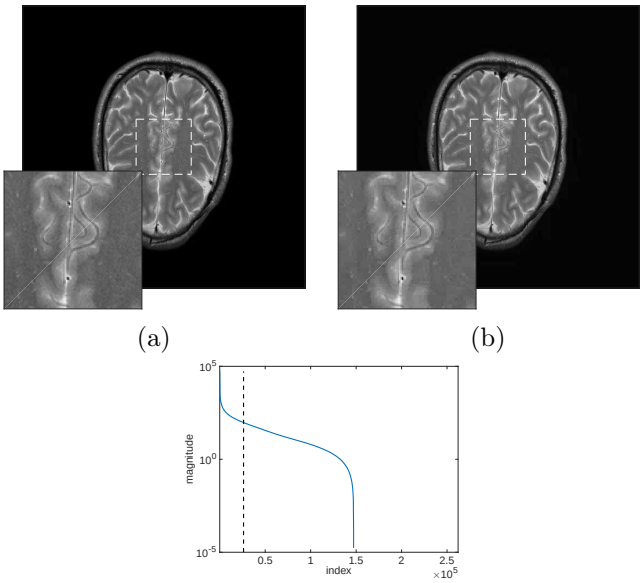


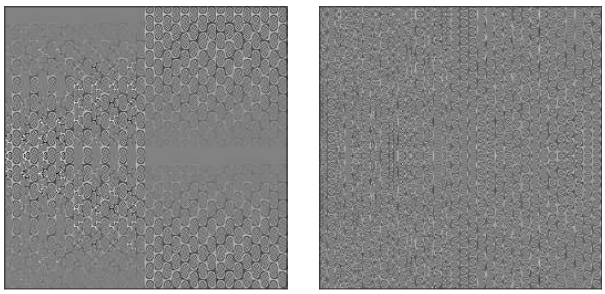
Figure 2: Schematic depiction of the Pareto curve, which relates the optimal solutions to the LASSO, BPDN and QP formulations of the sparse recovery problem. At a give (τ, σ), the derivative of the curve is proportional to λ.



(a) (b)

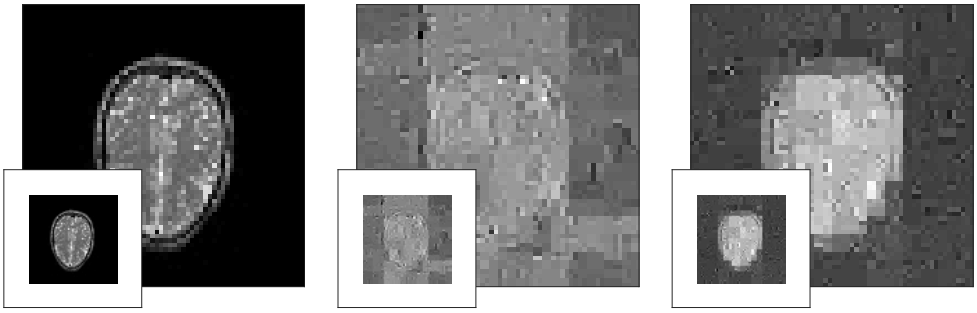
(c)

Figure 3: Sparsity in wavelets. (a) original image, (b) image using only 10% of the largest Wavelet coefficients, (c) magnitude of the Wavelet coefficients, the vertical line indicates the cut-off used to produce image (b).



(a) (b)

Figure 4: Coherence of (a) F and W^T and (b) FS and W^T. We see that the second matrix is much less coherent and hence is better suited for CS reconstruction.



(a) (b) (c)

Figure 5: (a) original image, (b) reconstruction with $\rho = 8$ without S, (c) reconstruction with $\rho = 8$ with S. The latter clearly gives a much better reconstruction, illustrating the importance of including the matrix S.

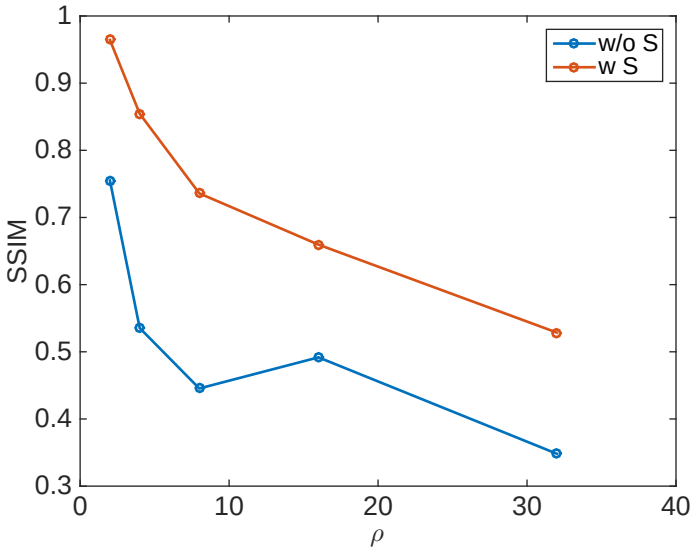


Figure 6: Reconstruction quality (in terms of the SSIM) for various subsampling ratios.

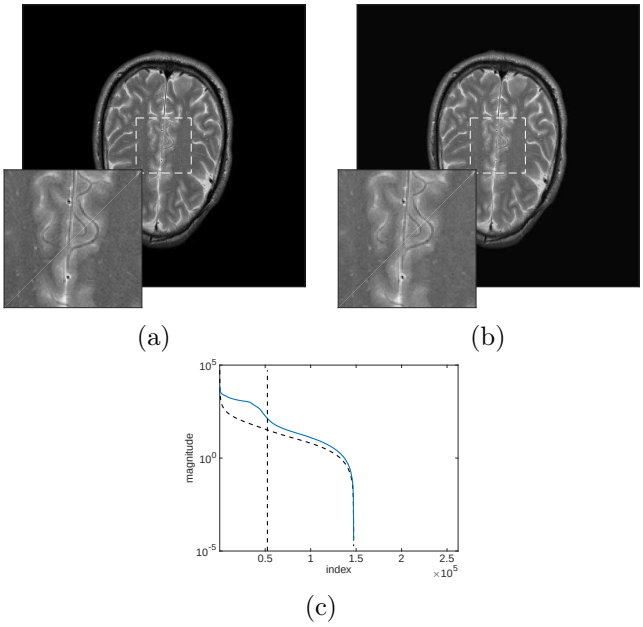


Figure 7: Sparsity in modified wavelets WS. (a) original image, (b) image using 20% of the largest Wavelet coefficients, (c) magnitude of the Wavelet coefficients. The dotted line indicates the magnitude of $W\mathbf{u}$ while the solid line indicates the magnitude of $WS\mathbf{u}$. The vertical line indicates the cut-off used to produce image (b). We see that the original image is less sparse in the modified wavelets.

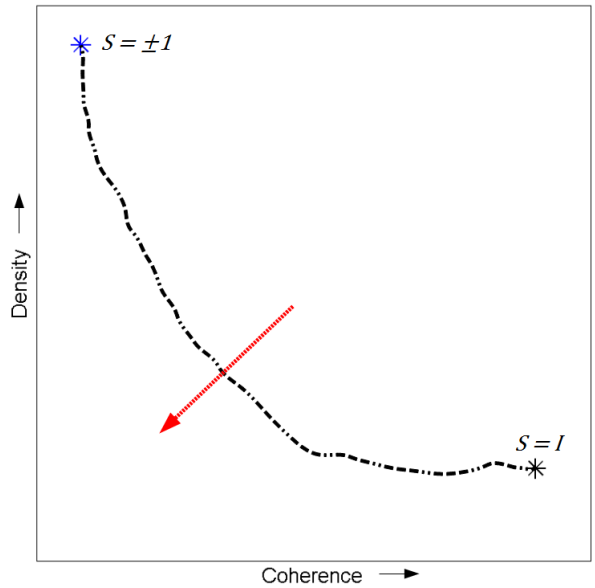


Figure 8: Schematic depiction of the tradeoff between sparsity and coherence.

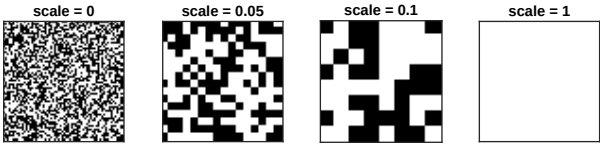


Figure 9: Examples of partiall coherent masks, ranging from completely incoherent (left) to completely coherent (right).

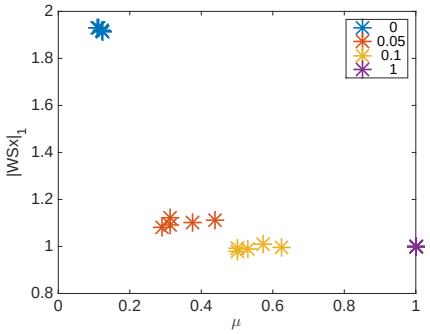


Figure 10: Tradeoff between sparsity and coherence for various image masks, ranging from completely incoherent (scale = 0) to completely coherent (scale = 1).

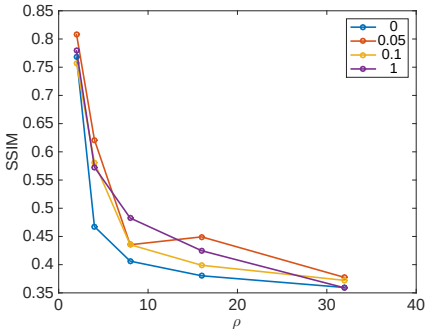


Figure 11: Reconstruction quality using modified wavelets with various image masks, ranging from completely incoherent to completely coherent. We see that the partially coherent masks with scales 0.05 and 0.1 perform slightly better than the other two.

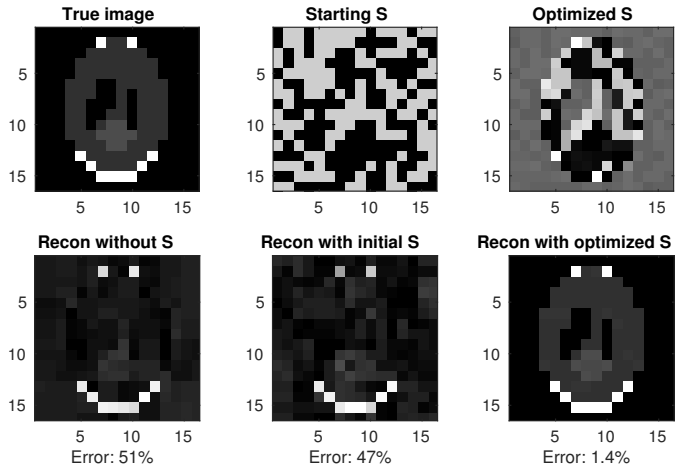


Figure 12: Experimental design. Top row: The ground truth image, the starting and optimized S, respectively. Bottom row: The three reconstructions, obtained without S, with the starting S and with the optimized S. Note the drastic improvementin the obtained image when the optimized S is employed.

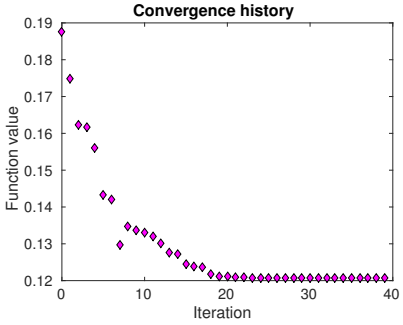


Figure 13: Experimental design. Convergence history for the design algorithm (Eq. 1).

A Matlab framework for rapid prototyping

To test the proposed algorithms, we used the `SPOT` toolbox, which allows us to define matrix-free linear operators. This toolbox allows us to use standard Matlab matrix-vector notation and manipulation while avoiding explicitly storing dense matrices. A 2D Fourier transform of a 2D signal `u`, for example, can be defined as follows.

```
F  = opDFT2(n,n);
ut = F*u;
```

Here, the Fourier operator `F` acts like a matrix, but upon multiplication it calls `fft2`. We can construct new operators by simply multiplying them together. The MRI measurement process, for example, is implemented as follows.

```
F = opDFT2(n,n);
I = randperm(n);
I = I(1:m);
R = opRestriction(n,I);
y = R*F*u;
```

For the sparse reconstruction we use `spgl1`. A complete reconstruction then, is done as follows.

```
F = opDFT2(n,n);
I = randperm(n);
I = I(1:m);
R = opRestriction(n,I);
y = R*F*u;
W = opWavelet2(n,n,'Haar');
z1 = spgl1(R*F*W',y,[],sigma);
u1 = W'*z;
```