

COMPGV19 Project: Subsampled MRI and Image Reconstruction

Antonio Remiro Azócar, MSc Machine Learning

30 May 2017

Abstract: The use of two compressed sensing undersampling schemes, three reconstruction algorithms and different regularisation functionals is explored and evaluated to recover two biomedical images.

1 Introduction

Magnetic Resonance Imaging (MRI) is an essential modality in medical imaging with many applications including soft tissue imaging, brain imaging and angiography. The major issue with MRI techniques is that they suffer from very slow acquisition times, sometimes as long as one hour. Such lack of speed is intrinsic, primarily attributed to physical (e.g. magnetic field amplitudes limiting gradients) and physiological (e.g. peripheral nerve stimulation) constraints [1]. Undersampling techniques are often required to reduce acquisition times by sampling below the Nyquist rate. These result in a dataset no longer complete (therefore not suitable for standard inversion) and introduce artifacts which require computationally intensive image reconstruction to be removed.

Compressed sensing (CS) groups techniques seeking to exploit the (transform) sparsity of medical images to reconstruct the subsampled data without loss of information. Transform sparsity is inherent to MRI and is a fairly general concept, relating to the sparse representation of the image we wish to reconstruct in a transform domain. This is the fundamental philosophy of CS techniques, which fit well with MRI for the following reasons:

- **Medical images are intrinsically sparse**, either in the image domain e.g. angiograms (sparsity in the identity transform), or in other domains; e.g. the wavelet domain (brain images) or in a domain of spatial finite differences [2].
- **Medical images are naturally compressible:** They can be compressed with little loss of visual information via sparse coding (e.g. discrete cosine transform, wavelet transform) in an adequate domain.
- **MRI scanners acquire spatial-frequency encoded data;** instead of pixel samples in the image domain.

An MR image can be reconstructed via compressed sensing provided **three conditions** are satisfied [3]:

- The image **exhibits transform sparsity** and is therefore compressible via transform coding.
- **Aliasing artifacts** in the transform domain resulting from k -space subsampling are **incoherent** (noise-like).
- An adequate **non-linear recovery scheme** (one which preserves sparsity) is utilised to reconstruct the image.

Instead of acquiring a large volume of data for later compression, CS techniques sample the signal in a space of high information density and reconstruct it by means of an optimisation algorithm. Whereas classical compression algorithms must know a signal before it is compressed, in CS one “compresses” before physically acquiring the signal. The ultimate objective is to reduce the number of measurements made.

2 Mathematical Formulation and Regularisation Functionals

The concept of sparsity (the first condition in the above section) can be formalised as follows. A signal $x \in \mathbb{R}^N$ is sparse if there exists a basis $\Psi = (\psi_i)_i$, the sparsity transform, such that,

$$x = \sum_{i=0}^{N-1} \alpha_i \psi_i \quad \text{with} \quad \sum_{i=0}^{N-1} \|\alpha_i\|_{\ell^0} \ll N, \quad (1)$$

where $\sum_{i=0}^{N-1} \|\alpha_i\|_{\ell^0} = \#\{i | \alpha_i \neq 0\}$. As mentioned previously, one expects equation (1) to apply for medical images in a given transform domain. Provided that it does, and that basis Ψ is a priori known, the signal of interest $x \in \mathbb{R}^N$ can be reconstructed from a reduced number of samples z in a measurement basis $\Phi = (\phi_i)_i$ [4]. The bases Ψ and Φ must be incoherent i.e. noise-like (the second condition for MRI reconstruction via compressed sensing).

2.1 The ℓ^1 norm regularisation functional

The image reconstruction problem in question can be formulated as a convex optimisation problem:

$$\text{minimise}_x \quad \|\Psi x\|_{\ell^1} \quad \text{s.t. } \Phi x = z, \quad (2)$$

with Ψ as a sparsifying transform such that Ψx is sparse. Notice that the ℓ^0 norm has been replaced by the ℓ^1 norm ($\|\Psi x\|_{\ell^1} = \sum_{i=0}^{N-1} |[\Psi x]_i|$) in the sparsity constraint. The motivation for this change is that ℓ^0 minimisation, while often being correct, is intractable computationally, generally NP-hard and not a convex optimisation problem. On the other hand, ℓ^1 minimisation is convex and can be solved reasonably quickly [5]. In addition, it often results in a sparse solution, it is robust to noise and has been proved to guarantee near-optimal performance provided there is sufficient incoherence [5]. Least-squares estimation (ℓ^2 minimisation) is discarded, since while being simple to compute, it is often incorrect for underdetermined systems. It assigns greater penalties to larger coefficients, giving solutions with many small coefficients which are not sparse. In ℓ^1 minimisation, many small coefficients bear a larger penalty than only a few large coefficients. This promotes a sparse solution, restraining the presence of smaller coefficients. Equation (2) corresponds to **minimising the ℓ^1 norm with equality constraints**. This problem is also known as basis pursuit. If the entries of Φ , x and z are real valued, equation (2) can be reformulated as a linear problem (LP) [14].

The robustness to noise of ℓ_1 minimisation can be exploited as follows. Since for the noisy element of a signal, the sparsity constraint in equation (1) does not apply, one can remove such element from the image reconstruction procedure. This results in the following recovery scheme for image reconstruction [6,7], where the noise is filtered out by relaxing the constraint $\Phi x = z$, via:

$$\text{minimise}_x \quad \|\Psi x\|_{\ell^1} \quad \text{s.t. } \|\Phi x - z\|_{\ell^2} < \epsilon, \quad (3)$$

where z is the data measured by the scanner and ϵ is a threshold parameter enforcing consistency between the measurements and the reconstructed image. ϵ is usually tuned to a value roughly below the noise level. The constraint guarantees a solution which is compressible by Ψ and provides some sort of ‘denoising’. Equation (3) corresponds to **minimising the ℓ^1 norm with quadratic constraints**. If a sufficiently sparse signal of interest x_0 exists such that $z = \Phi x_0 + e$, where e represents an error term obeying $\|e\|_2 < \epsilon$, the solution to equation (3), x^* , lies close to x_0 . More rigorously, $\|x^* - x_0\|_2 < C \cdot \epsilon$, where C is a small constant [15]. Equation (3) can be reformulated as a second-order cone program (SOCP).

2.2 The Total-Variation regularisation functional

If the transform domain is sparse in terms of spatial finite differences, the objective function in (3) is the sum of absolute deviations in x . In this case, the objective function is denoted as the Total-Variation [8]. Provided that the gradients of a given 2D image are sparse, one can define the following operators:

$$\begin{aligned} D_{h;ij}x &= \left\{ \begin{array}{ll} x_{i+1,j} - x_{ij}, & i < n \\ 0 & i = n \end{array} \right\}, \\ D_{v;ij}x &= \left\{ \begin{array}{ll} x_{i,j+1} - x_{ij}, & j < n \\ 0 & j = n \end{array} \right\}, \\ D_{ij}x &= \begin{pmatrix} D_{h;ij}x \\ D_{v;ij}x \end{pmatrix}. \end{aligned}$$

In the above expressions, x_{ij} represents the pixel in the image domain located in row i and column j of an image of dimensions $n \times n$. The Total-Variation semi-norm of x is defined as:

$$\|x\|_{TV} = \sum_{ij} \sqrt{(D_{h;ij}x)^2 + (D_{v;ij}x)^2} = \sum_{ij} \|D_{ij}x\|_2.$$

The Total-Variation can be characterised as a sum of ‘discrete gradient’ magnitudes at every point of the image x ; the two-dimensional vector $D_{ij}x$ represents such ‘discrete gradients’.

If a medical 2D image is piecewise smooth, one can devise an optimisation problem replacing the minimisation of $\|\Psi x\|_{\ell^1}$ with that of $\|x\|_{TV}$ in equation (2) to give the problem,

$$\text{minimise}_x \quad \|x\|_{TV} \quad \text{s.t. } \Phi x = z. \quad (4)$$

This problem corresponds to **minimising the Total-Variation with equality constraints**. Equation (4) can reconstruct x exactly provided that a piecewise constant signal of interest x_0 exists such that only for a small number of indices, $D_{ij}x$ is non-zero [4].

Similarly, one can use Total-Variation minimisation to reconstruct an image from observations with noise [16]. The minimisation of $\|\Psi x\|_{\ell^1}$ can be replaced with that of $\|x\|_{TV}$ in equation (3) to give the problem,

$$\text{minimise}_x \quad \|x\|_{TV} \text{ s.t. } \|\Phi x - z\|_{\ell^2} < \epsilon, \quad (5)$$

This problem corresponds to **minimising the Total-Variation with quadratic constraints**. The TV semi-norm promotes the sparsity of Ψ and that of finite differences and is usually faster to compute than the ℓ^1 norm of Ψx . Both equations (4) and (5) can be reformulated as SOCPs.

This project will examine minimising problems (3),(5) and (2). i.e. minimising the ℓ^1 norm with quadratic constraints, minimising the Total-Variation with quadratic constraints and minimising the ℓ^1 norm with equality constraints respectively. The optimisation/reconstruction algorithms utilised to solve these problems are discussed in detail in Section 5. To solve problems (3) and (5), two variants of a standard log-barrier algorithm are implemented. To solve problems (2), a primal-dual algorithm for linear programming is implemented. Additionally, there is another objective function minimised in this project. This objective function is specific to the third reconstruction algorithm utilised, the NESTA algorithm.

3 Subsampling Schemes

As mentioned previously, MRI is well suited for the application of CS since data acquisition takes place in a transform domain instead of an image domain. This eases incoherent aliasing artifact generation. The transform domain (the domain of data acquisition) is the **k-space** trajectory of the MR pulse sequences. These are mapped to the image domain (the spatial domain) by a Fourier transformation [1]. In the context of Section 2, $x \in \mathbb{R}^n$, the image/signal of interest, lies in the spatial domain. z is the k -space data measured by the scanner and lies in the transform domain. The basis Ψ is an orthogonal sparsifying transform mapping the spatial domain to the new transform domain. The measurement basis Φ is the undersampled Fourier operator F_S ; that computed only at frequencies in a sampled subset S of k -space. The k -space data is therefore ‘partial’. One can devise a sampling pattern designing k -space trajectories for each image acquisition. In this case, Fourier basis vectors are used to design Φ with two sampling patterns in the transform domain: radial subsampling and random uniform subsampling (notice that we are reconstructing 2D MR images).

Radial subsampling is performed with **11 radial lines**, with **22 radial lines**, with **55 radial lines** (all star-shaped patterns). These structured sampling patterns consisting of radial lines in the transform domain have been inspired by [4]. Sampling along radial lines provides less sensitivity to motion artifacts than cartesian trajectories. Radial sampling schemes allow for undersampling [9], particularly when dealing with images of high contrast [10], but may be highly susceptible to system imperfections. To generate the radial k -space sampling patterns, the external function `LineMask.m` [11], developed by Justin Romberg, is utilised after pre-processing the images (see Section 4). The `LineMask` function samples k -space by applying a partial FFT to the images and adding Gaussian noise to the partial FFT results; additive Gaussian noise is introduced into both the real and the imaginary parts of Fourier coefficients. Such partial FFT generates the k -space samples in the Fourier domain in a star-shaped form. For each sampling pattern, the subsampling ratio is fixed for a given image. The subsampling ratio μ is given by:

$$\mu = \frac{\# \text{ number of subsampled measurements}}{\# \text{ number of full measurements}}. \quad (6)$$

Using 11 radial lines, for Image 1 (phantom), $\mu = 4.2\%$ and for Image 2 (MATLAB MRI), $\mu = 8.4\%$. Using 22 radial lines, for Image 1, $\mu = 8.4\%$ and for Image 2, $\mu = 16.3\%$. Using 55 radial lines, for Image 1, $\mu = 20.2\%$ and for Image 2, $\mu = 36.9\%$. See the next section for a detailed description of the utilised images/data. In this setting, the different subsampling ratios arise from sampling different numbers of lines. Figure 1 presents the different transform domain sampling positions arising from different numbers of lines for Image 1 (256×256). Figure 1 shows how the star-shaped patterns appear to characterise low frequencies more than high frequencies (the centre of k -space contains the low spatial frequency information whereas the periphery contains high spatial frequency information).

Random uniform subsampling is also performed, sampling 14%, 28% and 42% of the k -space. Random uniform patterns do not overrepresent low k -space frequencies with respect to high frequencies like star-shaped patterns. In random uniform sampling, the probability of sampling the coefficient corresponding to $z \in \mathbb{R}^2$ in the transform domain is uniform. Reference [4] encourages the use of this sampling pattern. Early CS research [3,4] sampled completely random k -space subsets. This approach guarantees extremely low coherence and mathematical simplicity even though it is impractical in problems of high dimensionality [6,7] (primarily because it does not result in reasonably smooth curves/lines). Other drawbacks of random point k -space sampling are that it lacks the robustness to withstand non-toy situations and that it does not account for the

non-uniform distribution of energy in MRI's frequency domain. Figure 2 presents the different k -space positions arising from different random uniform subsampling ratios for Image 1 (256×256).

Note: Points to consider for future work involve utilising random Gaussian sampling patterns. In random Gaussian sampling, the probability of sampling the coefficient corresponding to $z \in \mathbb{R}^2$ in the transform domain is given by $\exp(-(\frac{\|z\|_2}{\rho})^2)$, where ρ is modified depending on the specified sampling rate.

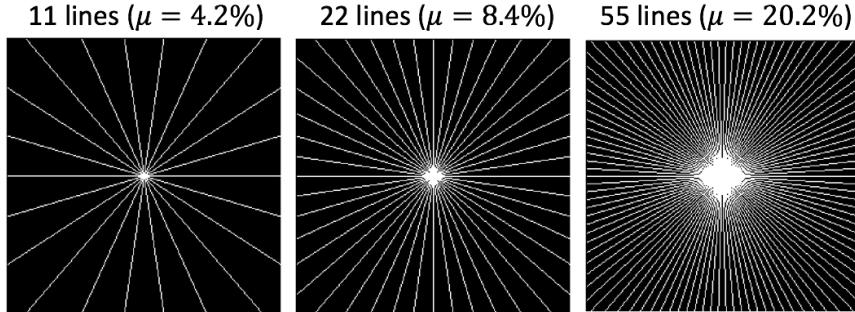


Figure 1: Transform domain sampling positions arising from 11, 22 and 55 line patterns for Image 1.

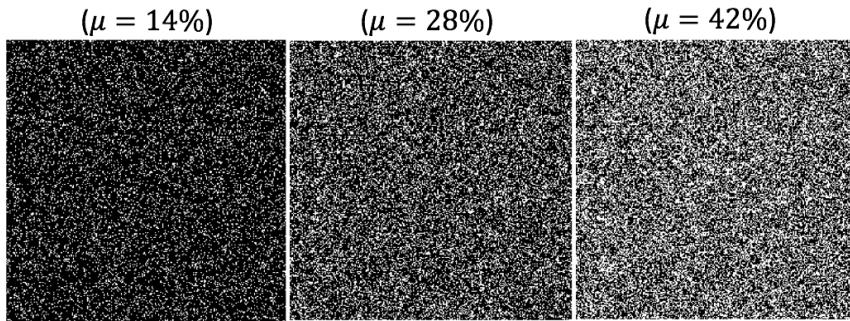


Figure 2: k -space domain positions arising from random uniform sampling with ratios $\mu = 14\%, 28\%, 42\%$.

4 Data

Reconstruction algorithms and subsampling schemes are evaluated on two images. The first image is a 256×256 variant of the Shepp-Logan phantom with improved contrast for refined visual perception (the MATLAB `Modified Shepp-Logan`) [12]. The Shepp-Logan phantom is a greyscale intensity image with one bigger ellipse (characterising the brain) including multiple smaller ellipses (characterising the brain's features). All the pixels of the phantom image have values between zero and one; each pixel value equals the sum of the additive intensity values of each ellipse that a pixel belongs to. The numerical accuracy of the implemented reconstruction algorithms is also tested in a 128×128 2D slice from an MR image from MATLAB's 3D `mri` dataset [13]. The slice extracted is a mid-horizontal slice (showing the view from the top of the head of the subject). The slice has been normalised using the MATLAB function `mat2gray`; hence all the pixels for the second image have values between zero and one.

5 Optimisation/Reconstruction Algorithms

Compressed sensing depends on solving efficiently the convex optimisation problems consisting of ℓ^1 norm and TV seminorm minimisation respectively. Solving these problems presents a challenge for the following reasons:

- The objective functions in the optimisation problems are non-smooth.
- The solution space's \mathbb{R}^N dimension N may be very large in a biomedical imaging context.

In this practical, three reconstruction algorithms are evaluated: a standard log-barrier method for SOCPs (two variants of such are implemented since the Newton step update differs for different objective functions), a primal-dual algorithm for LPs and the NESTA algorithm, introduced by Becker *et al.* [17]. The implementations of the log-barrier and the primal-dual method are general-purpose and follow closely the instructions presented in

Chapter 11 of [18] and in [19]. The motivation for implementing a primal-dual algorithm to solve (2) (minimising the ℓ^2 norm with equality constraints) has been the following. When using log-barrier methods and random uniform subsampling patterns, the reconstruction algorithms get stuck on cone iterations and keep returning previous iterates. Therefore, no reconstruction takes place and it is needed to resort to other methods to sample k -space with a random uniform distribution. Both the log-barrier and primal-dual are examples of interior point methods whose major computational bottleneck is computing the Newton step. The code provided for the interior point methods makes use of the conjugate gradient (CG) method. CG methods are iterative algorithms which are matrix-free, thereby being able to find approximate solutions to ‘large scale’ problems such as those presented by MRI. Notice than the log-barrier method has been utilised alongside radial subsampling schemes and the primal-dual algorithm has been used alongside random uniform subsampling schemes exclusively.

The NESTA algorithm is presented as a case study of the new heuristics introduced in the field. Notice that for this purpose, Becker *et al.*’s implementation [17] has been utilised and is therefore not included in the online submission. Whereas the aforementioned interior point methods are tractable computationally, computing a solution with them may require the inversion of a linear system of equations of at least size N . This makes the log-barrier method computationally expensive in problems of high dimensionality. In addition, interior point methods act as a black-box, in the sense that they do not take into account algebraic properties of Ψ and Φ . For example, that usually fast algorithms exist (such as fast Fourier/wavelet transform) to project a medical image. The NESTA algorithm is expected to capitalise more on the algebraic properties of Ψ and Φ , giving greater computational tractability to image reconstruction problems, and perhaps being more practical in a clinical context. The NESTA algorithm has been utilised alongside both random uniform and radial subsampling schemes.

5.1 The standard log-barrier method

Problems (3) and (5) can be reformulated as:

$$\text{minimise}_y \quad \langle c_0, y \rangle \quad \text{s.t. } \Phi_0 y = z, f_i(y) \leq 0, \quad i = 1, \dots, m, \quad (7)$$

where $y \in \mathbb{R}^N$, $z \in \mathbb{R}^K$ and Φ_0 is a matrix of dimensions $K \times N$. f_i can represent either a linear or second order constraint indexed by $i = 1, \dots, m$. Under the log-barrier algorithm, problem (7) can be reformulated as a series of problems with linear constraints:

$$\text{minimise}_y \quad \langle c_0, y \rangle + \frac{1}{\tau^k} \sum_i -\log(-f_i(y)) \quad \text{s.t. } \Phi_0 y = z, \quad (8)$$

with $\tau^k > \tau^{k-1}$. The penalty introduced in (8), $\frac{1}{\tau^k} \sum_i -\log(-f_i(y))$, includes the inequality constraints of (7) into the program. Notice that this penalty term is smooth as long as the constraint f_i is not breached. When τ^k gets bigger, y^k , the solution to problem (8), gets close to the optimal solution y^* , the solution of (7). More formally, $\langle c_0, y^k \rangle - \langle c_0, y^* \rangle < \frac{m}{\tau^k}$. The term $\frac{m}{\tau^k}$ is called the **duality gap**; when the aforementioned inequality holds for a given iteration k , solution y^k is within $\frac{m}{\tau^k}$ from the optimal solution y^* . One can therefore solve the presented subproblems accurately by iterating Newton’s method a few times for each log-barrier iteration k . In particular, utilising y^k as the initial point for log-barrier iteration $k+1$. At each k , Newton’s method iteratively formulates a series of quadratic approximations to (8). Given $\Phi_0 y = z$, the change in y , Δy , that minimises $q(y + \Delta y)$, where $q(y + \Delta y) = y + \langle g_y, \Delta y \rangle + \frac{1}{2} \langle H_y \Delta y, \Delta y \rangle \approx f_0(y + \Delta y)$ corresponds to the quadratic approximation around a point y , solves the set of linear equations:

$$\tau \begin{pmatrix} H_y & \Phi_0^T \\ \Phi_0 & \mathbf{0} \end{pmatrix} \begin{pmatrix} \Delta y \\ v \end{pmatrix} = -\tau g_y, \quad (9)$$

where g_y is the gradient at point y , H_y the corresponding Hessian and v represents Lagrange multipliers from the constraints of the quadratic minimisation which are not directly utilised. In fact, when solving problems (3) and (5), $\Phi_0 = \mathbf{0}$ i.e. there are no equality constraints in the quadratic minimisation problem. Therefore, (9) is symmetric positive definite, and conjugate gradient methods can be utilised to solve system (9) large scale. One can obtain the Newton step direction since we know Δy . Now the specifics of the log-barrier implementation will be discussed for each functional. The notation follows that utilised in the implemented programs.

5.1.1 ℓ^1 minimisation with quadratic constraints

Problem (3) is reformulated as the SOCP:

$$\text{minimise}_{y,w} \quad \sum_i w_i \quad \text{s.t. } y - w \leq 0, -y - w \leq 0, \frac{1}{2} (\|\Phi y - z\|_2^2 - \epsilon^2) \leq 0$$

As in (9), the Newton step at point (y, w) can be cast as:

$$\begin{pmatrix} \Sigma_{11} - f_\epsilon \Phi^T \Phi + f_\epsilon^{-2} \Phi^T (\Phi y - z) (\Phi y - z)^T \Phi & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{11} \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta w \end{pmatrix} = \begin{pmatrix} f_{w_1}^{-1} - f_{w_2}^{-1} + f_\epsilon^{-1} \Phi^T (\Phi y - z) \\ \tau \mathbf{1} - f_{w_1}^{-1} - f_{w_2}^2 \end{pmatrix} := \begin{pmatrix} \chi_1 \\ \chi_2 \end{pmatrix},$$

for a given τ , with $f_{w_1} = y - w$, $f_{w_2} = -y - w$, $f_\epsilon = \frac{1}{2}(\|\Phi y - z\|_2^2 - \epsilon^2)$, $\Sigma_{11} = F_{w_1}^{-2} + F_{w_2}^{-2}$ and $\Sigma_{12} = -F_{w_1}^{-1} + F_{w_2}^{-2}$, where F are diagonal matrices s.t. $(F)_{ii} = f_i(y)$. Setting:

$$\Sigma_y = \Sigma_{11} - \Sigma_{12}^2 \Sigma_{11}^{-1},$$

and getting rid of Δw with:

$$\Delta w = \Sigma_{11}^{-1} (\chi_2 - \Sigma_{12} \Delta y),$$

yields:

$$(\Sigma_y - f_\epsilon^{-1} \Phi^T \Phi + f_\epsilon^{-2} \Phi^T r r^T \Phi) \Delta y = \chi_1 - \Sigma_{12} \Sigma_{11}^{-1} \chi_2,$$

where $r = \Phi y - z$. The above system is symmetric positive definite and so a solution can be found via the conjugate gradient method.

5.1.2 Total-Variation minimisation with quadratic constraints

Problem (5) is reformulated as the SOCP:

$$\begin{aligned} \text{minimise}_{y,u} \quad & \sum_{ij} u_{ij} \quad \text{s.t.} \quad \|D_{ij}y\|_2 \leq u_{ij}, \quad i,j = 1, \dots, n, \\ & \|\Phi y - z\|_2 \leq \epsilon. \end{aligned}$$

The following inequality functions are defined:

$$\begin{aligned} f_{u_{ij}} &= \frac{1}{2}(\|D_{ij}\|_2^2 - u_{ij}^2), \quad i,j = 1, \dots, n, \\ f_\epsilon &= \frac{1}{2}(\|\Phi y - z\|_2^2 - \epsilon^2), \end{aligned}$$

With,

$$\begin{aligned} \nabla f_\epsilon &= \begin{pmatrix} \Phi^T (\Phi y - z) \\ \mathbf{0} \end{pmatrix}, \quad \nabla f_\epsilon \nabla f_\epsilon^T = \begin{pmatrix} \Phi^T (\Phi y - z) (\Phi y - z)^T \Phi & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \nabla^2 f_\epsilon = \begin{pmatrix} \Phi^* \Phi & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \\ \nabla^2 f_{u_{ij}} &= \begin{pmatrix} D_{ij}^* D_{ij} & \mathbf{0} \\ \mathbf{0} & -\delta_{ij} \delta_{ij}^T \end{pmatrix}, \end{aligned}$$

where δ_{ij} is the Kronecker vector (consisting of zeros everywhere except in entry ij , which is one), one obtains the Newton system:

$$\begin{pmatrix} H_{11} & Z \Sigma_{12} \\ \Sigma_{12} Z^T & \Sigma_{22} \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta u \end{pmatrix} = \begin{pmatrix} D_h^T F_u^{-1} D_h y + D_v^T F_u^{-1} D_v y + f_\epsilon^{-1} \Phi^T (\Phi y - z) \\ -\tau \mathbf{1} - u f_u^{-1} \end{pmatrix} := \begin{pmatrix} \chi_1 \\ \chi_2 \end{pmatrix},$$

where $(u f_u^{-1})_{ij} = \frac{u_{ij}}{f_{u_{ij}}}$, F are diagonal matrices s.t. $(F)_{ii} = f_i(y)$, D_h has $D_{h;ij}$ as rows, D_v has $D_{v;ij}$ as rows and Γ is a matrix dependent on y such that:

$$Z = D_h^T \Sigma_{\partial h} + D_v^T \Sigma_{\partial v},$$

with $\Sigma_{\partial h} = \text{diag}(D_h y)$, $\Sigma_{\partial v} = \text{diag}(D_v y)$. Also:

$$\begin{aligned} H_{11} &= D_h^T (-F_u^{-1}) D_h + D_v^T (-F_u^{-1} D_v) + Z F_u^{-2} Z^T - f_\epsilon^{-1} \Phi^T \Phi + f_\epsilon^{-2} \Phi^T (\Phi y - z) (\Phi y - z)^T \Phi, \\ \Sigma_{12} &= -U F_u^{-2}, \quad \Sigma_{22} = F_u^{-1} + F_u^{-2} U^2, \end{aligned}$$

where $U = \text{diag}(u)$. Getting rid of Δu via:

$$\Delta u = \Sigma_{22}^{-1} (\chi_2 - \Sigma_{12} \Sigma_{\partial h} D_h \Delta y - \Sigma_{12} \Sigma_{\partial v} D_v \Delta y)$$

yields the reduced system:

$$H'_{11} \Delta y = \chi_1 - (D_h^T \Sigma_{\partial h} + D_v^T \Sigma_{\partial v}) \Sigma_{12} \Sigma_{22}^{-1} \chi_2,$$

where

$$\begin{aligned} H'_{11} &= H_{11} - Z \Sigma_{12}^2 \Sigma_{22}^{-1} Z^T \\ &= D_h^T (\Sigma_z \Sigma_{\partial h}^2 - F_u^{-1}) D_h + D_v^T (\Sigma_z \Sigma_{\partial v}^2 - F_u^{-1}) D_v + D_h^T (\Sigma_z \Sigma_{\partial h} \Sigma_{\partial v}) D_v + D_v^T (\Sigma_z \Sigma_{\partial h} \Sigma_{\partial v}) D_h - \\ &\quad f_\epsilon^{-1} \Phi^T \Phi + f_\epsilon^{-2} \Phi^T r r^T \Phi, \end{aligned}$$

with $\Sigma_z = F_t^{-2} - \Sigma_{12}^2 \Sigma_{22}^{-1}$ and $r = \Phi y - z$. The above reduced system is symmetric positive definite and so a solution can be found via the conjugate gradient method.

5.1.3 Implementation summary and parameter choices

The implementation of both variants of the log-barrier algorithm is summarised as follows:

1. The algorithm takes as inputs a value for the initial starting point y^0 , a tolerance value η , parameter ν and a starting value τ^1 , parametrising the duality gap. Set $k = 1$.
2. Subsequently, problem (8) can be solved by utilising Newton's method and later backtracking line search. If z^0 is the starting point, the solution of the iterative Newton procedure is z^1 (the solution of the Newton procedure with a starting point z^{k-1} is denoted z^k).
3. If the duality gap is lower than the specified tolerance value ($\frac{m}{\tau^k} < \eta$), the algorithm terminates and outputs the solution value z^k .
4. Else, update $\tau^{k+1} = \nu\tau^k$, $k = k + 1$ and return to Step 2.

Parameter/argument choices: For both ℓ^1 and Total-Variation minimisation, the algorithms (functions `logbar_TV.m` and `logbar_L1.m`) take 14 arguments. The first argument, `y_init` corresponds to the initial guess, which has been set to equal the minimum energy reconstruction (a $N \times 1$ vector). Such recovery consists of setting unobserved Fourier coefficients to zero after subsampling (the initial back-projection estimate). The arguments `phi` and `phi_t` are ‘implicit matrices’. These are created using two function handles, which allow us to efficiently solve the problem ‘large scale’. An $n \times n$ image is viewed as a vector of $N = n^2$ elements; such vector is subsampled at K frequencies, giving a $K \times N$ explicit measurement matrix. For example, for Image 1 (size 256×256), sampled at $K = 5481$ locations in the 2D transform domain using 22 radial lines, the explicit measurement matrix is size 5481×65536 . In this case, the system (9) becomes too big to compute or store explicitly. The function handles make the problem implicit and less computationally expensive by returning Φy (a vector of K elements) for each vector y (N elements); each Newton solver requires applying `phi` and `phi_t` only once. The argument `z` represents the k -space $K \times 1$ observation vector.

`nu` denotes ν , the factor by which to increase the barrier constant τ^k after each log-barrier iteration k . It has been set to 2 since for lower values ($1 \leq \nu < 2$), the number of Newton iterations required to reduce the duality gap ($\frac{m}{\tau^k}$) to a certain level increases dramatically. On the other hand, for higher values, the number of total iterations does not decrease significantly (it is actually not very sensitive to the value of ν for $\nu > 7$). We therefore opt to increase the barrier constant conservatively. The trade-off between the total number of Newton iterations required and the value of ν is explored in Section 6. The sixth argument, `eps`, corresponds to ϵ : the scalar constraint relaxation parameter. It has been set to 5×10^{-3} , tuned to a small value below the noise level and thereby guaranteeing a solution compressible by the basis Ψ . The argument `tol_logbar` denotes the tolerance of the log-barrier algorithm, which terminates when the duality gap is lesser than or equal to the specified tolerance. Recall that the duality gap quantifies how close solution y^k is from the optimal y^* for an iteration k . The log-barrier tolerance is set to 1×10^{-1} to ensure the duality gap is reduced to a sufficiently small value for accurate reconstruction. Notice that the total number of log-barrier iterations required is exclusively determined by `tol_logbar`. In fact, it is calculated in advance using:

$$\text{total log-barrier iterations} = \frac{\log m - \log \eta - \log \tau^1}{\log \nu},$$

where η is the log-barrier tolerance. The argument `tol_newton` represents the tolerance of the Newton subroutine, which terminates when the Newton decrement is lesser than or equal to the value specified (or when the number of Newton iterations reaches `newton_iterlimit`). It has been set to a value equal to that of `tol_logbar`, 1×10^{-1} , for the reasons before explained. `newton_iterlimit`, the maximum number of Newton iterations per log-barrier iteration, is set to 50; mostly as a precautionary measure since there are usually no more than 5 Newton iterations for each k . `tol_conjgrad` determines the desired precision of the conjugate gradients subroutine and has been set to 1×10^{-8} to maximise accuracy. The maximum number of conjugate gradients, `conjgrad_iterlimit`, has been set at 200. This means that a CG subroutine terminates when $\frac{\|Hz-g\|_2}{\|g\|_2} \leq 10^{-8}$ (for a symmetric positive definite system $Hx = g$) or when its number of iterations reaches 200.

Finally, the argument `newton_step` corresponds to the Newton step length for updating y . The step length $s \leq 1$ and must be picked such that $f_i(y + s\Delta y) < 0 \ \forall i = 1, \dots, m$ and the functional increases enough: $f_0(y + s\Delta y) < f_0(y) + \alpha s\Delta y(g_y, \Delta y)$ (minimum step size that stays in the interior), where α is another parameter to be defined (`alpha`). We set 0.9 initially for more aggressive updates and decrease its value by a multiple of β (`beta`) until the two aforementioned inequalities are satisfied. This is done via setting `alpha=0.01` and `beta=0.5`. The function does not take a specific initial τ^1 as an argument. This value is chosen prudently, computed within the program s.t. $\frac{m}{\tau^k}$ is roughly equal to $\langle c_0, y \rangle$ (the original norm) after log-barrier iteration 1 and is correspondingly adjusted after each log-barrier iteration.

The two following algorithms will not be explained in as much depth due to ‘word limit’ constraints.

5.2 The primal-dual algorithm

The primal-dual algorithm is used to solve problem (2), which can be reformulated as the LP:

$$\text{minimise}_y \quad \langle c_0, y \rangle \quad \text{s.t. } \Phi_0 y = z, f_i(y) \leq 0, \quad i = 1, \dots, m, \quad (10)$$

where $y \in \mathbb{R}^N$, $z \in \mathbb{R}^K$ and Φ_0 is a matrix of dimensions $K \times N$. f_i represents a linear functional $f_i(y) = \langle c_i, y \rangle + d_i$ indexed by $i = 1, \dots, m$. Let y^* be an optimal point. At y^* \exists also optimal dual vectors $\gamma^* \in \mathbb{R}^K$, $\lambda^* \in \mathbb{R}^m$, $\lambda^* \geq 0$ such that:

$$\begin{aligned} c_0 + \Phi_0^T \gamma^* + \sum_i \lambda_i^* c_i &= \mathbf{0}, \\ \lambda_i^* f_i(y^*) &= 0, \quad i = 1, \dots, m, \\ \Phi_0 y^* &= z, \\ f_i(y^*) &\leq 0, \quad i = 1, \dots, m. \end{aligned}$$

The above system of non-linear conditions corresponds to the Karush-Kuhn-Tucker (KKT) conditions and can be solved via the primal-dual algorithm to find the optimal y^* , γ^* , λ^* . However one must first modify $\lambda_i f_i = 0$ to $\lambda_i^k = f_i(y^k) = -\frac{1}{\tau^k}$, where k indexes an algorithm iteration, such for an interior point y^k, γ^k, λ^k (s.t. $f_i(y^k) < 0, \lambda^k > 0$), the point resulting from the next step $y^{k+1}, \lambda^{k+1}, \gamma^{k+1}$ stays in the interior. Prudently increasing τ^k in the aforementioned inequality shifts the system solution towards the interior.

Now, the goal is to find a step $(\Delta y, \Delta \gamma, \Delta \lambda)$ such that from a point (y, γ, λ) :

$$r_t(y + \Delta y, \gamma + \Delta \gamma, \lambda + \Delta \lambda) = 0, \quad (11)$$

where r_t represents a residual corresponding to the primal (p), dual (d) or central (c) problem, with $r_p = \Phi_0 y - z$, $r_d = c_0 + \Phi_0^T \gamma + \sum_i \lambda_i c_i$ and $r_c = -\lambda f = \frac{1}{\tau} \mathbf{1}$, $(\Lambda)_{ii} = \lambda_i$ (diagonal), $f = (f_1(y) \dots f_m(y))^T$. Using a Taylor expansion to turn (11) into a linear system, performing some algebraic expressions, and getting rid of Δw with:

$$\Delta \lambda = -\Lambda F^{-1} C \Delta y - \lambda - \frac{1}{\tau} f^{-1}, \quad (12)$$

where $(F)_{ii} = f_i(y)$ (diagonal) and C is a $m \times N$ matrix whose rows are c_i^T , yields the main system:

$$\begin{pmatrix} -C^T F^{-1} \Lambda C & \Phi_0^T \\ \Phi_0 & \mathbf{0} \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta \gamma \end{pmatrix} = \begin{pmatrix} -c_0 + \frac{1}{\tau} C^T f^{-1} - \Phi_0^T \gamma \\ z - \Phi_0 y \end{pmatrix} \quad (13)$$

This system is solved using the instructions provided in [18,19]; for this algorithm we do not explicitly derive the Newton step calculation in this report. The major computational bottle neck in the algorithm comes from solving (13). Since in ℓ^1 norm minimisation with equality constraints, matrix $-C^T F^{-1} \Lambda C$ is easily invertible, (13) can become a symmetric positive definite system. Therefore, conjugate gradient methods can be utilised to solve system (13) large scale. The term $\eta = -f^T \lambda$ denotes the surrogate duality gap. This term represents how near a point (y, γ, λ) is to optimality when r_d, r_p are small. The primal-dual method terminates when the number of iterations reaches `primal_dual_iterlimit` or when η is lesser than the specified `tol_primal_dual`.

$\frac{\|Hz-g\|_2}{\|g\|_2} \leq 10^{-8}$ (for a symmetric positive definite system $Hz = g$) or when its number of iterations reaches 200.

Parameter/argument choices: For ℓ^1 minimisation with equality constraints, the algorithm (function `primal_dual_L1`) takes 11 arguments: `y_init`, `phi`, `phi_tz`, `nu`, `eps`, `tol_primal_dual`, `primal_dual_iterlimit`, `tol_conjgrad`, `conjgrad_iterlimit`, `alpha` and `beta`. All argument names repeated from the log-barrier algorithm have the same interpretation and take the same values, except ν , which is set to 10 for faster convergence. Notice that `tol_logbar` has been replaced by `tol_primal_dual`, which also takes the value 1×10^{-1} and now an explicit iteration limit `primal_dual_iterlimit` has been introduced. This is set at 50.

Finally, the Newton step length $0 \leq s \leq 1$ for an update in step direction $(\Delta y, \Delta \gamma, \Delta \lambda)$, must be picked s.t. $\forall i \quad f_i(y + s\Delta y) < 0, \lambda_i > 0$ (i.e. $y + s\Delta y$ and $\lambda + s\Delta \lambda$ are in the interior), and the residuals' norm decreases enough: $r_\tau(y + s\Delta y, \gamma + s\Delta \gamma, \lambda + s\Delta \lambda) \|_2 \leq (1 - \alpha s) \cdot \|r_\tau(y, \gamma, \lambda)\|_2$. We set `alpha` to 0.01. Letting,

$$\mathcal{I}_f^+ = i : \langle c_i, \Delta y \rangle > 0, \quad \mathcal{I}_\lambda^- i : \Delta \lambda < 0,$$

and assigning,

$$s_{max} = 0.99 \cdot \min\{1, -\{f_i(y)/\langle c_i, \Delta y \rangle, i \in \mathcal{I}_f^+\}, \{-\frac{\lambda_i}{\Delta \lambda_i}, i \in \mathcal{I}_\lambda^-\}\},$$

we set the initial step length to $s = s_{max}$ i.e. the max. step length maintaining us in the interior is selected (since f_i is linear, this tackles $y + s\Delta y$ and $\lambda + s\Delta \lambda$ being in the interior). Then, we check if the residuals' norm has decreased enough, and if it has not, we set $s' = \beta \cdot s$ and attempt again the procedure (`beta` = 0.5).

5.3 NESTA

The NESTA algorithm has its origin in Nesterov's research involving the minimisation of non-smooth functions [20]. It can be used to minimise both the ℓ^1 norm and the Total-Variation seminorm. The algorithm iteratively produces a decreasing succession of iterates which eventually converge to a solution. The next guess for a given step is formulated as a linear combination of:

1. An expression inducing the iterate to move in a direction opposing that of the gradient of the objective function at the present point.
2. An expression that follows the previous directions of the gradient. This term enhances the algorithm's convergence attributes.

Both expressions can define solutions of other quadratic subroutines with analytical solutions. If $\Phi\Phi^*$ is the identity matrix (the rows of Φ make up an orthonormal family), where Φ is the measurement basis, a solution for the problem can be found efficiently. This holds since we can simply the inverse matrices in the analytical solutions to the subproblems. Each NESTA iteration has complexity $O(N + C_\Phi)$. C_Φ denotes the complexity of the application of Φ or Φ^* . Notwithstanding, note that $\Phi\Phi^*$ applies and $C_\Phi = O(N \log N)$ when Φ is a subset of the Fourier transform matrix rows. This makes NESTA an attractive option in this context and is used to both Image 1 and Image 2 with ℓ^1 minimisation and TV variation (radial line subsampling patterns only since NESTA requires the measurement matrix Φ to be a partial isometry i.e. $\Phi\Phi^T = \mathbf{I}$, which we did not account for when undersampling randomly).

Parameter/argument choices: The number of iterations in an inner loop `opts.maxiter` has been set to 5,000 and the number of continuation steps (`opts.maxintiter`) has been set to 5. `delta` (the ℓ^2 norm error bound) has been set to 0.001 to enforce that the variables fit the observations closely in $\|z - \Phi y\|_2 \leq \delta$. We have also tried setting $\delta = 0$ to minimise objective functions with equality constraints and $\delta = (\sqrt{m + 2\sqrt{(2m)}}) \cdot \sigma$. The tolerance for the stopping criteria (`opts.TolVar`) has been set to 1×10^{-5} to ensure an accurate reconstruction. `mu`, the desired value of μ at the last continuation step, is chosen to be small (0.0005) since a smaller value leads to higher accuracy.

6 Results and Discussion

6.1 Log-barrier algorithm (radial line subsampling)

Figure 3 shows visualisations of the original Image 1, its minimal-energy reconstruction, its reconstruction using Total-Variation minimisation with quadratic constraints and its reconstruction using ℓ^1 minimisation with quadratic constraints (left to right). Figure 4 shows visualisations of the original Image 2, its minimal-energy reconstruction (the initial back-projection estimate), its reconstruction using Total-Variation minimisation with quadratic constraints and its reconstruction using ℓ^1 minimisation with quadratic constraints. Results are displayed for radial sampling patterns with 11, 22 and 55 lines.

In Figures 3 and 4 we see the effects of not sampling k -space frequently enough. If this is the case, aliasing (wraparound) occurs in the image domain, as seen in the minimum Total-Variation and ℓ^1 recoveries for sampling patterns of 11 lines. For Image 1 (Phantom), Total-Variation minimisation appears to work better than ℓ^1 norm minimisation, requiring a lower value of μ for accurate reconstruction. Whereas precise recovery is attained with only 22 lines ($\mu = 8.4\%$) using TV, ℓ^1 appears to require 55 lines for a satisfactory reconstruction for Image 1 (for Image 2 55 radial lines are required in both cases). Note that more time was put into finding optimal starting parameters for TV than ℓ^1 . Perhaps different initial settings yield better results with less wraparound 'ghosts' for ℓ^1 recovery. Similar findings are obtained from Figure 4, where the lower quality of the reconstructions might be attributed to the lower resolution and most importantly, the lower sparsity of the original image. Table 1 presents a summary of reconstruction results for Image 1 using the log-barrier algorithm. Table 2 presents a summary of simulation results for Image 2.

Figure 5 presents plots of the residual error (the absolute difference in intensity) between the original images and those reconstructed through the log-barrier method. Both numerical and visual results confirm our original arguments; given the same image and the same number of radial lines (same value of μ), TV yields lower RMS values (greater accuracy) than ℓ^1 minimisation. On the other hand, ℓ^1 minimisation is less computationally expensive. For a given value of μ , it nearly always results in a greater number of Newton iterations. However, the time required per Newton iteration and the overall time of the reconstruction is always lower with ℓ^1 minimisation. This is expected since solving the TV Newton system requires the computation of additional terms involving 'discrete gradients' (D terms). Again, for a given number of radial lines and a given minimisation

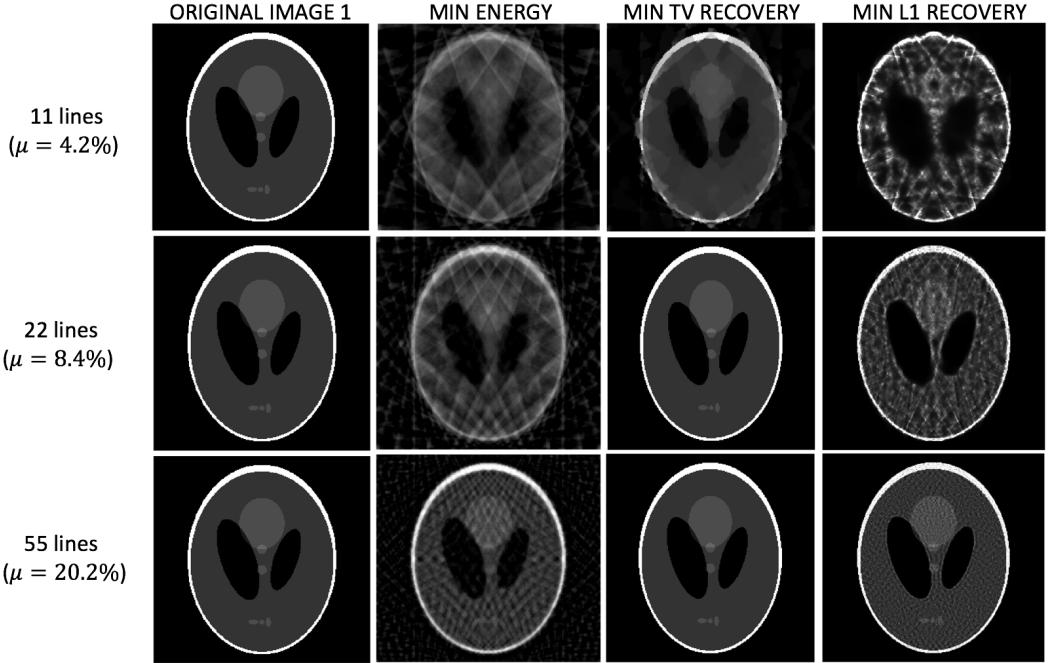


Figure 3: Original Image 1 (Phantom) and its reconstructions using different numbers of radial lines.

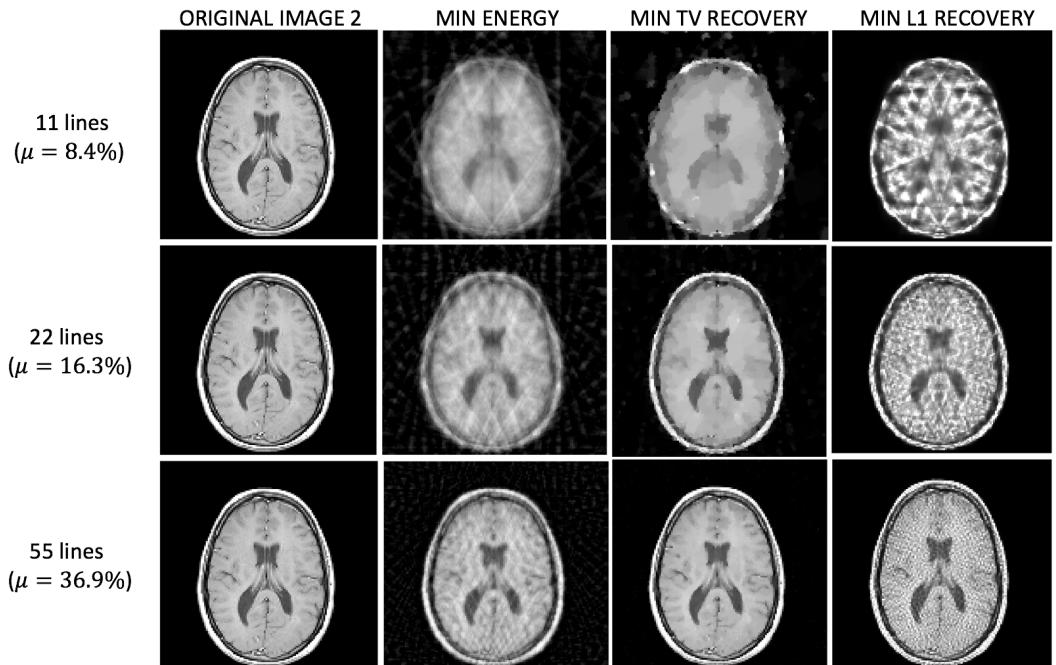


Figure 4: Original Image 2 (MRI MATLAB) and its reconstructions using different numbers of radial lines.

Image 1	TV (11)	L1 (11)	TV (22)	L1 (22)	TV (55)	L1 (55)
MSE	9.3×10^{-3}	2.3×10^{-2}	8.6×10^{-6}	8.5×10^{-3}	8.4×10^{-8}	2.6×10^{-3}
# iterations	31 (8)	39 (10)	48 (9)	40 (10)	33 (9)	34 (10)
Time (s)	30	27	48	27	46	25



Table 1: Simulation results for the log-barrier reconstruction of Image 1. MSE denotes the mean square error between the reconstruction and the original image. Results are presented for both Total-Variation (TV) and ℓ^1 (L1) minimisation, where TV(11), for example, denotes using Total-Variation minimisation with a subsampling pattern of 11 radial lines. # iterations denotes the number of Newton iterations required to complete the reconstruction; the numbers in parentheses indicate the log-barrier iterations required.

Image 2	TV (11)	L1 (11)	TV (22)	L1 (22)	TV (55)	L1 (55)
MSE	1.82×10^{-2}	3.41×10^{-2}	5.4×10^{-3}	1.1×10^{-2}	5.6×10^{-4}	5.4×10^{-3}
# iterations	26 (7)	32 (9)	27 (7)	32 (9)	29 (8)	32 (9)
Time (s)	10	7	9	7	8	7

Table 2: Simulation results for the log-barrier reconstruction of Image 2

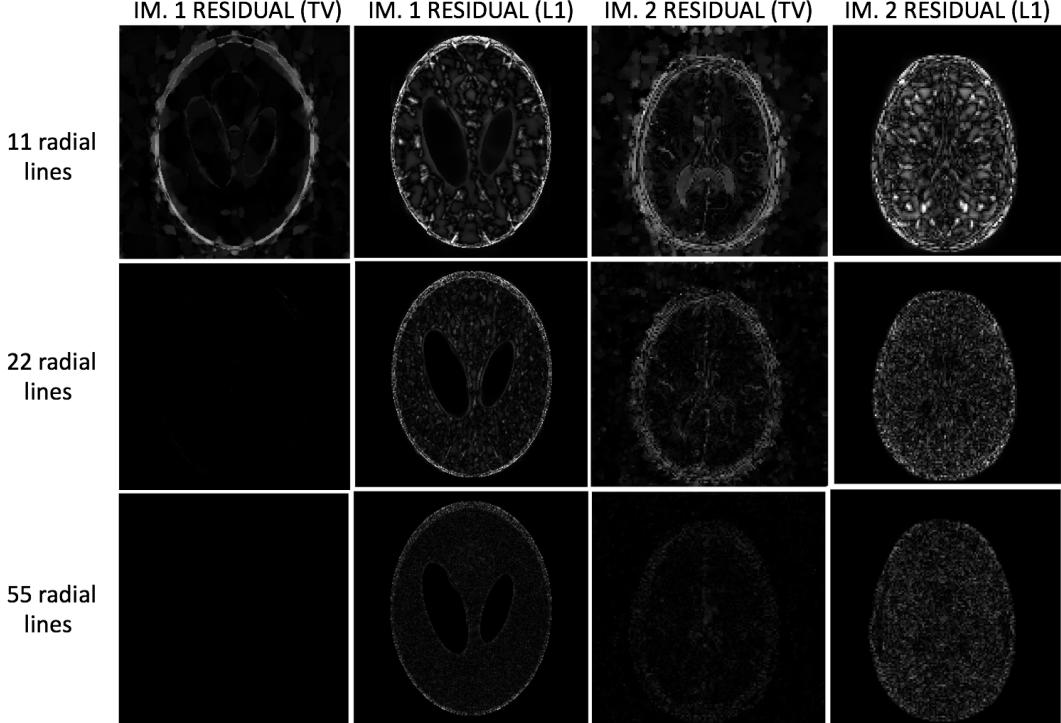


Figure 5: Residual images obtained for Image 1 (left) and Image 2 (right) through log-barrier reconstruction.

type, the reconstruction of Image 1 yields lower RMS values (greater accuracy) than that of Image 2. Note that computing the reconstruction for Image 2 is less computationally expensive than for Image 1 (in terms of Newton iterations and overall procedure time) because Image 2 is lower size.

The effect of sparsity: To understand why the reconstruction of Image 1 performs better than that of Image 2, we must consider the concept of sparsity. We are already familiar with sparsity in the image/pixel domain. In addition, recall that for any 2D real image y , if one thinks about each row as a signal, the horizontal ‘discrete’ gradient $D_{h;ijy}$ views the ‘difference’ between adjacent pixels (an edge denotes a sharp change). Similarly, if one thinks about each column as a signal, the vertical gradient $D_{v;ijy}$ views such difference. Also recall that, collectively, D_{ijy} denotes the magnitude (ℓ^2 norm) of both horizontal and vertical components. Figure 6 shows plots of $D_{h;ijy}$ (horizontal gradient), $D_{v;ijy}$ (vertical gradient) and D_{ijy} (gradient ℓ^2 norm) for the original Image 1 and Image 2. Table 3 presents the percentage of non-zero elements in the original image/pixel domain and in $D_{h;ijy}$, $D_{v;ijy}$ and D_{ijy} for each image. Both numerical and visual observations show that Image 1 is much more sparse (less dense in non-zero elements) than Image 2 in its image gradients. Recall that the TV of an image is simply the ℓ^1 norm of Dy (the sum of discrete gradients at every point). Therefore, TV minimisation exploits much more the sparse gradient property for Image 1 than for Image 2. This explains why TV minimisation improves the performance of ℓ^1 minimisation more significantly for Image 1 than for Image 2. Also, TV minimisation gives generally better performance for both images, because it not only exploits sparsity in the pixel domain, it also exploits sparsity in the gradient domains for both. Note that Image 2 has a lower percentage of non-zero elements than Image 1 in the pixel domain. However, most of the non-zero elements of Image 1 are in fact within 0.2 of zero (Figure 7) and in practice, it is more sparse than Image 2 (this is confirmed by visual assessments in Figures 3 and 4). Either this or simply more careful initial parameter optimisation for Image 1, may explain the better accuracy of ℓ^1 minimisation for the phantom. Figure 7 shows histograms of pixel intensities in the image domain for Image 1 and Image 2. The percentage of values lower than 0.2 for Image 1 is 91.2% (against 59.3% for Image 2), meaning that Image 1 has a greater proportion of smaller values and in this sense, is more sparse.

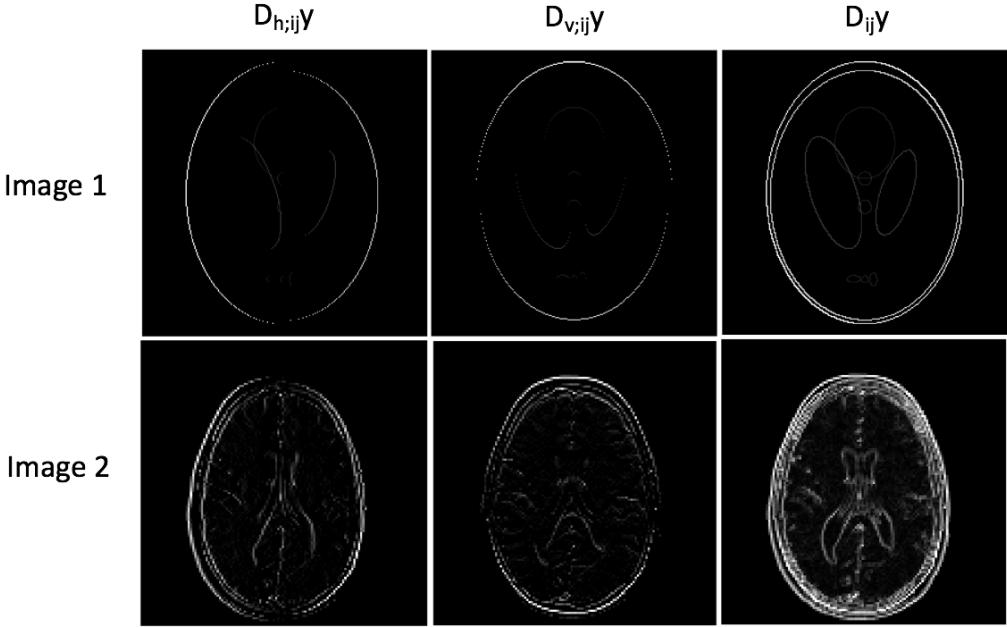


Figure 6: Horizontal, vertical and ℓ^2 norm gradient plots for Image 1 (top) and Image 2 (bottom)

Sparsity	Image Domain	$D_{h;ij}y$	$D_{v;ij}y$	$D_{ij}y$
Image 1	49.5%	2.26%	1.62%	3.33%
Image 2	43.2%	38.9%	38.1%	43.2%

Table 3: Percentage of non-zero elements (density) for Image 1 and Image 2 in different domains.

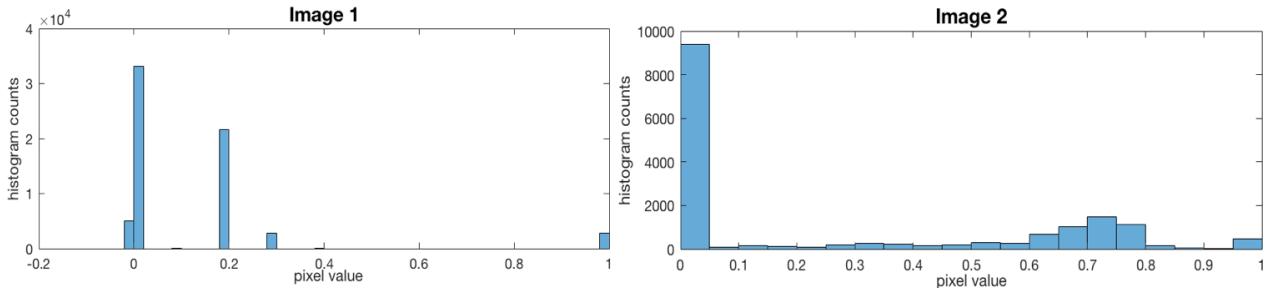


Figure 7: Histograms of pixel intensities for Image 1 (left) and Image 2 (right).

Notes on the number of radial lines: The number of radial lines used (the subsampling ratio μ) does not make a difference in the computational time of the reconstruction procedure. This is evident; since one subsamples the k -space before initiating the optimisation algorithm. Nevertheless, in a clinical setting, increasing the subsampling ratio results in greater image acquisition times. As previously observed, not sampling k -space frequently enough results in aliasing (wraparound). However, if satisfactory accuracy is obtained with a small subsampling ratio (e.g. TV minimisation with 22 radial lines), there is no need to continue increasing μ (e.g. to 55 lines). In terms of data acquisition times at the scanner, this would be too costly for only a marginal improvement in accuracy. Notice that TV minimisation with 22 radial lines already achieves near-exact reconstruction of the phantom data with a small μ . These results are in line with those obtained by [19].

Notes on ℓ^2 minimisation: The minimum energy reconstruction (the initial back-projection estimate) in Figure 1 and Figure 2 is equivalent to ℓ^2 norm minimisation with equality constraints i.e. once the k -space is subsampled, minimising:

$$\|\Psi y\|_{\ell^2} = \left(\sum_i |\Psi y|_i^2 \right)^{\frac{1}{2}} \text{ s.t. } \Phi y = z.$$

Both through visual assessment (see Figure 3 and Figure 4 minimum energy reconstructions) and through

computing the MSE between the minimum energy reconstructions and the original images (see Table 4), we can confirm, that while being simple to compute ℓ^2 minimisation yields incorrect solutions than ℓ^1 and TV minimisation more often for these underdetermined systems. The higher MSE values for the minimum energy reconstructions result from ℓ^2 minimisation assigning greater penalties to larger coefficients, promoting less sparse reconstructions (as observed in Figures 3 and 4).

	MSE	ME(11)	ME (22)	ME (55)
Image 1	2.4×10^{-2}	1.8×10^{-2}	7.6×10^{-3}	
Image 2	2.2×10^{-2}	1.6×10^{-2}	6.5×10^{-3}	

Table 4: MSEs for minimum energy reconstructions of Image 1 and Image 2 with the radial line # in parentheses.

The duality gap: Recall that in the log-barrier algorithm, when for iteration k , τ^k gets bigger, y^k gets close to the optimal solution y^* . More formally, $\langle c_0, y^k \rangle - \langle c_0, y^* \rangle < \frac{m}{\tau^k}$. The term $\frac{m}{\tau^k}$ is called the **duality gap**; when the aforementioned inequality holds for a given iteration k , solution y^k is within $\frac{m}{\tau^k}$ from the optimal solution y^* . Therefore, the duality gap can be a useful metric to quantify how close we are from the optimal solution and how the reconstruction error is improving. Figure 8 presents a plot of the duality gap against the cumulative number of Newton iterations for a TV reconstruction of Image 2 using 22 radial lines for $\nu < 2$, $\nu < 5$ and $\nu < 10$. In this graph, we can observe that there is a trade off between ν and the total number of Newton iterations required to reduce the duality gap to a specific value (in this case, below 1×10^{-1} , the specified log-barrier tolerance).

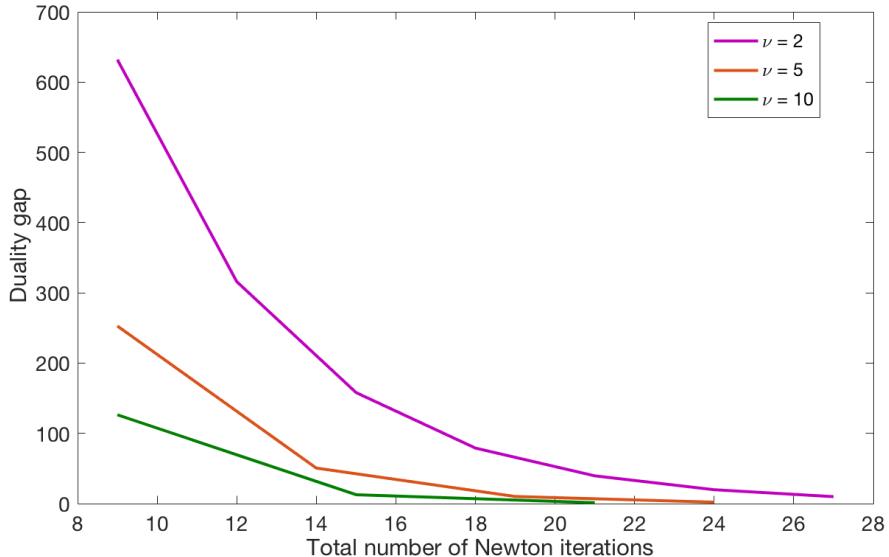


Figure 8: Plot of the duality gap against the cumulative number of Newton iterations for $\nu = 2$, $\nu = 5$ and $\nu = 10$ for Image 2 (22 radial lines), minimising the Total-Variation using the log-barrier algorithm.

6.2 Primal-dual algorithm (random uniform subsampling)



Figure 9 shows visualisations of the original Image 1, its minimal-energy reconstruction and its reconstruction using ℓ^1 minimisation with equality constraints (left to right). Figure 10 shows visualisations of the original Image 2, its minimal-energy reconstruction (the initial back-projection estimate) and its reconstruction using ℓ^1 minimisation with equality constraints. Results are displayed for random uniform sampling patterns with $\mu = 14\%$, $\mu = 28\%$ and $\mu = 42\%$.

Again, in Figures 9 and 10 we see the effects of not sampling k -space frequently enough (e.g. using only $\mu = 14\%$). In this case, aliasing occurs throughout the entire image domain; for 6.1 it was concentrated in specific areas of the image. This is due to the uniform random subsampling pattern. Table 5 presents a summary of reconstruction results for Image 1 using the primal-dual algorithm. Table 6 presents a summary of simulation results for Image 2. The primal-dual algorithm appears to be much less computationally expensive than the log-barrier method by looking at the overall procedure times (it is much less expensive to compute the Newton step for the primal-dual algorithm). However, it appears that with random uniform subsampling, reconstruction times scale with μ ; this was not obvious in the results for 6.1. It seems evident that in this case, a subsampling ratio of $\mu = 14\%$ is not enough for accurate reconstruction (radial patterns covering $\mu = 8.4\%$

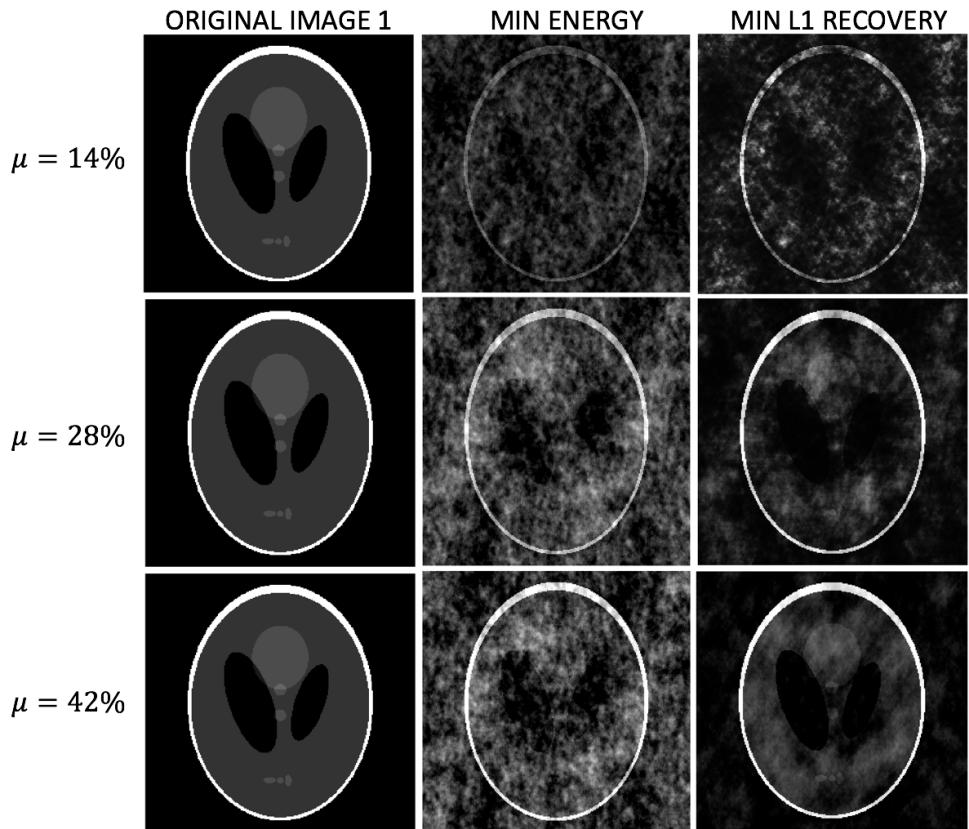


Figure 9: Original Image 1 (Phantom) and its reconstructions using different random uniform sampling ratios.

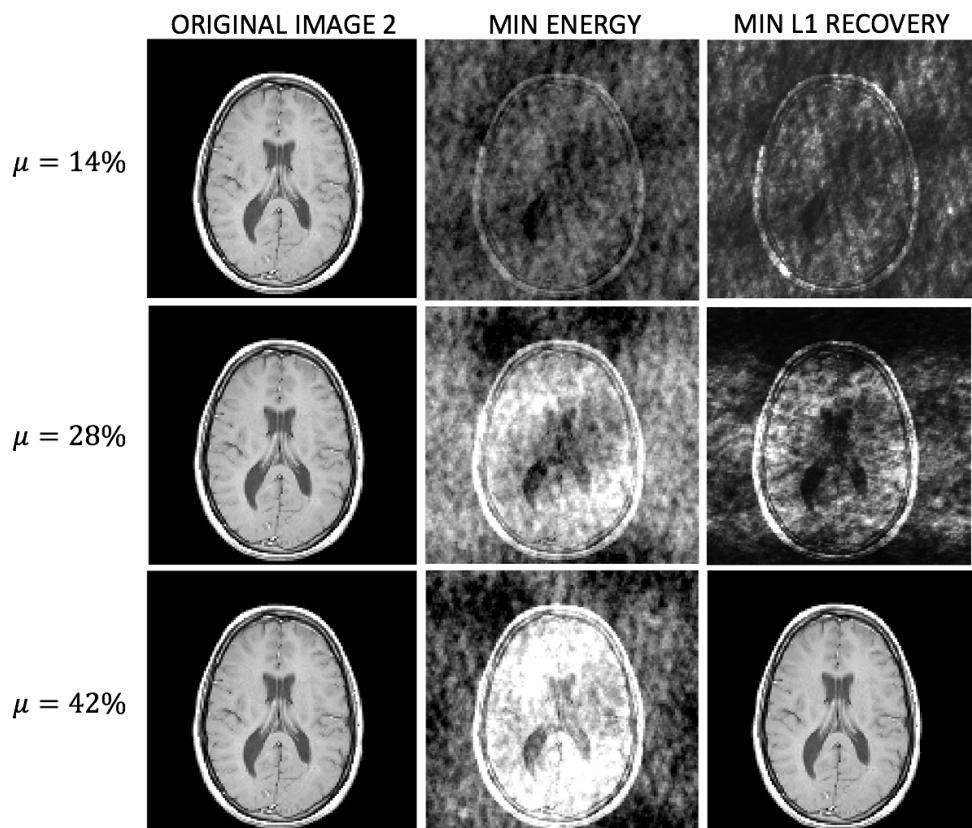


Figure 10: Original Image 2 (MRI) and its reconstructions using different random uniform sampling ratios.

were good enough in some instances in **6.1**; in a way random uniform sampling allows less for undersampling). Initially, it appeared that reconstructions for Image 1 were more accurate than for Image 2. However, an incredibly low RMS and accurate reconstruction is obtained using $\mu = 42\%$ for Image 2. Most importantly, we can observe how random uniform patterns do not overrepresent low k -space frequencies with respect to high k -space frequencies. For radial line subsampling, most of the residual error of the phantom is located in the “skull” (the white oval); there is barely no residual error in the central components because the low k -space frequencies are overrepresented. This does not occur for random uniform subsampling and may favour its use in Image 2, which is not centred in the vertical axis. A drawback of this subsampling method is that it does not account for the non-uniform distribution of energy in MRI’s frequency, and this can be assessed visually by looking at the intensity distribution for $\mu = 14\%, 28\%$ reconstructions.

Image 1	L1 (14%)	L1 (28%)	L1 (42%)
MSE	2.6×10^{-2}	9.6×10^{-3}	4.2×10^{-3}
# iterations	10	12	25
Time (s)	5	16	57

Table 5: Simulation results for the primal-dual reconstruction of Image 1. Results are presented for ℓ^1 (L1) minimisation with equality constraints, where L1(14%), for example, denotes using ℓ^1 minimisation with a random uniform subsampling pattern of $\mu = 14\%$. # iterations denotes the number of primal-dual iterations required to complete the reconstruction.

Image 2	L1 (14%)	L1 (28%)	L1 (42%)
MSE	1.0×10^{-1}	7.6×10^{-2}	1.4×10^{-8}
# iterations	8	9	13
Time (s)	1	3	8

Table 6: Simulation results for the primal-dual reconstruction of Image 2.

6.3 NESTA

Subsampling radial lines: Table 7 shows simulation results for the NESTA reconstruction of Image 1 (using 22 radial lines and both TV and ℓ^1 minimisation) and Image 2 (using 55 radial lines and both TV AND ℓ^1 minimisation). Figure 11 presents the aforementioned reconstructions visually. NESTA is roughly as time consuming as the primal-dual algorithm but achieves more computational tractability than the log-barrier algorithm. Additionally, it achieves more accurate reconstructions than both interior point methods using similar settings. The state-of-the-art NESTA algorithm capitalises more on the algebraic properties of Ψ and Φ , and is more practical in a clinical context. The NESTA computation times seem to be stable unlike those of the primal-dual algorithm.

	L1 (Im. 1, 22 radial lines)	TV (Im. 1, 22 lines)	L1 (Im. 2, 55 lines)	Im. 2 (55 lines)
MSE	4.4×10^{-3}	9.0×10^{-7}	2.0×10^{-3}	5.3×10^{-4}
# iterations	318	434	243	196
Time (s)	7	13	2	2

Table 7: Simulation results for the NESTA reconstructions of Image 1 and Image 2.

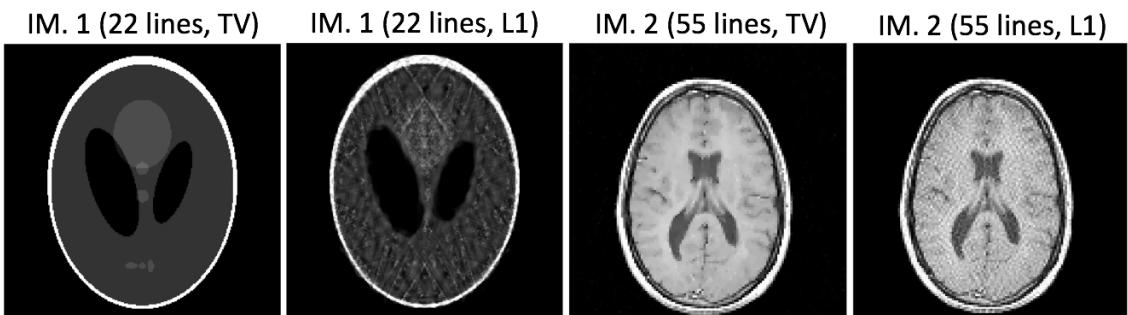


Figure 11: Reconstructions of Image 1 and Image 2 using different NESTA settings.

7 Conclusion

Simple reconstruction algorithms involving interior point methods can provide very accurate recoveries, even in comparison to state-of-the-art procedures like NESTA. Examples of such “successes” are:

- The reconstruction of Image 1 via TV minimisation with quadratic constraints (log-barrier algorithm) using radial line subsampling with 22 or 55 lines.
- The reconstruction of Image 1 via ℓ^1 minimisation with equality constraints (primal dual algorithm) using random uniform subsampling with $\mu = 42\%$.

In general, radial line subsampling gives better results than random uniform subsampling for a given value of μ . For a given subsampling level, specified patterns greatly influence the reconstruction quality. It is worth attempting to use a random Gaussian subsampling pattern.

References

- [1] G. Wright. Magnetic Resonance Imaging. *IEEE Signal Processing Mag*, 14(1), pp. 56-66, Jan. 1997.
- [2] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6): pp. 1182– 1195, 2007.
- [3] D. Donoho. Compressed sensing. *IEEE Trans Inf Theory*, 52: pp. 1289–1306, 2006.
- [4] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52: pp. 489–509, 2006.
- [5] T. Tao. Compressed sensing: or the equation $Ax = b$ revisited. [PowerPoint presentation]. *University of California, Los Angeles*. 2010. Available from: <http://www.math.hkbu.edu.hk/ttang/UsefulCollections/compressed-sensing1.pdf> [Accessed 18 March 2017].
- [6] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed Sensing MRI. *IEEE Signal Processing Magazine*, 25(2): pp. 72–82, March 2008.
- [7] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6): pp. 1182– 1195, 2007
- [8] L. Rudin, S. Osher, E. Fatemi. Non-linear total variation noise removal algorithm. *Phys. D*, 60: pp. 259–268, 1992.
- [9] K. Scheffler and J. Hennig. Reduced circular field-of-view imaging. *Magn. Reson. Med.*, 40(3): pp. 474–480, 1998.
- [10] D.C. Peters, F.R. Korosec, T.M. Grist, W.F. Block, J.E. Holden, K.K. Vigen, and C.A. Mistretta. Undersampled projection reconstruction applied to MR angiography. *Magn. Reson. Med.*, 43(1): pp. 91–101, 2000.
- [11] J.Romberg, LineMask.m. Available from: <https://uk.mathworks.com/matlabcentral/fileexchange/48428-cosbos--color-sensor-based-occupancy-sensing?focused=3846100&tab=function>.
- [12] Matlab Help: Create Modified Shepp-Logan Head Phantom Image.
- [13] Matlab Help: Exploring Slices from a 3-Dimensional MRI data set.
- [14] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20: pp. 33–61, 1999.
- [15] J. Nocedal and S. J. Wright. Numerical Optimization. Springer, New York, 1999.
- [16] E. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. Submitted to Communications on Pure and Applied Mathematics, March 2005.
- [17] S. Becker, J. Bobin, and E. Candes. NESTA: A fast and accurate first-order method for sparse recovery. Technical report, Caltech, 2009.
- [18] S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2004.
- [19] E. Candes and J.Romberg. 11-magic: Recovery of sparse signals via convex programming. *Stanford University*, 2005. Available from: <http://statweb.stanford.edu> [Accessed 5 May 2017]
- [20] Y. Nesterov. Gradient methods for minimizing composite objective function. Technical report, Université Catholique de Louvain, 2007.

