



Subject Areas:

65K05, 90C26, 90C06, 90C30,
90C59

Keywords:

Compressed sensing, inverse
problems, non-linear constrained
optimisation, tomography

Author for correspondence:

T. Blumensath

e-mail:

thomas.blumensath@soton.ac.uk

Non-convexly constrained image reconstruction from non-linear tomographic x-ray measurements

Thomas Blumensath¹ Richard Boardman²

¹ISVR Signal Processing and Control Group,
University of Southampton, SO17 1BJ, UK

²Engineering Materials, University of Southampton,
SO17 1BJ, UK

The use of polychromatic x-ray sources in tomographic x-ray measurements leads to non-linear x-ray transmission effects. As these non-linearities are not normally taken into account in tomographic reconstruction, artefacts occur, which can be particularly severe when imaging objects with multiple materials of widely varying x-ray attenuation properties. In these settings, reconstruction algorithms based on a non-linear x-ray transmission model become valuable. We here study the use of one such model and develop algorithms that impose additional non-convex constraints on the reconstruction. This allows us to reconstruct volumetric data even when limited measurements are available. We propose a non-linear conjugate gradient iterative hard thresholding (NCGIHT) algorithm and show how many prior modelling assumptions can be imposed using a range of non-convex constraints.

1. Introduction

Tomographic x-ray imaging is used widely in areas as diverse as medical diagnostics and non-destructive material testing. Using sets of x-ray projection images taken from different directions, tomographic inverse algorithms reconstruct volumetric images of an object's internal structure. Most algorithms in use today rely on a linear x-ray attenuation model which assumes that the x-ray source produces mono energetic x-ray photons. However, most laboratory sources produce polychromatic x-ray beams and so, approximate pre-processing algorithms are often used to correct for some of the non-linear effects prior to volumetric reconstruction.

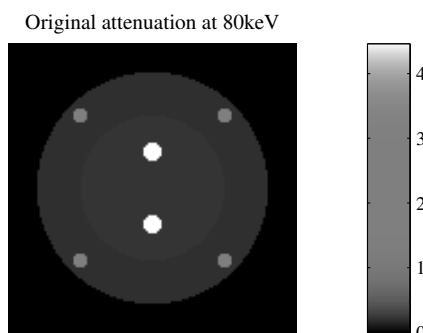


Figure 1. A simulated phantom made up of 4 materials which at 80 keV have mass attenuation coefficients of $0.2\text{cm}^2/\text{g}$, $0.23\text{cm}^2/\text{g}$, $2.77\text{cm}^2/\text{g}$ and $4.46\text{cm}^2/\text{g}$ respectively. To show the difference between the two low attenuating materials, the grey scale has been compressed (see colourbar).

Whilst these preprocessing techniques often work well when imaging objects in which x-ray attenuation does not vary to widely, in settings where there is a mix of highly attenuating regions and low attenuating regions, severe artefacts are introduced in the image reconstruction. Examples include metal implants imaged in human tissue or carbon composite components with metal inclusions.

As a motivating example, consider the simulated phantom shown in figure 1, which is made up of four materials, two materials with low x-ray attenuation factors that differ very little between materials (an aluminium outer ring with a silicon inner disk) and two highly attenuating materials (silver and gold inclusions). Figure 1 shows the mass attenuation for these materials at 80 keV. To differentiate the different materials, the greyscale has been compressed to better show the difference in the high and low attenuating regions (see the colourbar).

We simulated two different x-ray acquisitions, one with a monochromatic source at 80keV and one with a polychromatic source with a maximum energy of 130 keV. 720 fan beam projections were calculated for each system and reconstructed with the filtered backprojection algorithm (using matlab's image processing toolbox functions) and the Algebraic Reconstruction Technique (using the randomized Kaczmarz algorithm from the AIRTools toolbox [1]). The results are shown in figure 2. Whilst our compressed colour map shows reconstruction artefacts in the monochromatic reconstructing on the right, the left polychromatic reconstructions clearly show additional artefacts typical for polychromatic x-ray reconstructions. These artefacts can be seen as shadows between high attenuating regions (the darker area in the centre of the reconstruction) and brighter streaks that tangentially connect the boundaries of different high attenuating objects.

The realisation that polychromatic x-rays lead to a non-linear transmission models is typically known as beam-hardening [2]. Many different techniques for the correction of this effect have been proposed over the years, from physical filtering of the x-ray source spectrum to reduce the non-linear relationship between path length and x-ray absorption [2] to direct inversions of the full non-linear model (2.6).

For single material objects, intensity linearisation can be performed [2] based on phantom scans [3], initial uncorrected reconstructions [3,4] or simulation studies [5]. Measurements are interpolated [6–9] and this process is often iterated [6,7,10]. Intensity corrections can also be calculated for multiple material objects [3,8,11–16]. The initial reconstruction is segmented which again provides initial path length estimates. This process can then also be iterated. For objects made up of two materials with known attenuation coefficients, it is also possible to instead estimate the volume fraction for each spatial location [6,17,18].

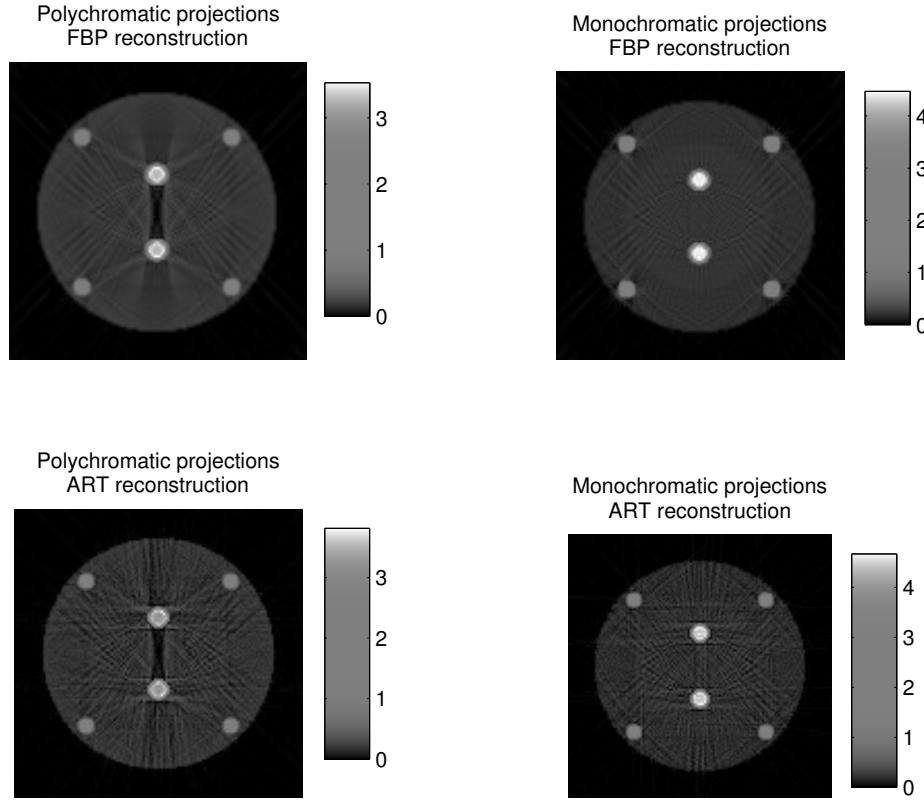


Figure 2. Reconstruction of the simulated phantom using two algorithms (the filtered backprojection FBP algorithm and the algebraic reconstruction technique (ART)). The reconstructions on the left use simulations of a polychromatic x-ray source whilst the reconstructions on the right use a monochromatic source. To show the difference between the two low attenuating materials, the tray scale has been compressed (see colourbar).

2. The non-linear x-ray observation model

We will here propose an advanced reconstruction algorithm that explicitly models non-linear x-ray attenuation. Furthermore, we will impose additional regularisation constraints, which will not only be able to overcome the general ill-conditioning of the non-linear inverse problem, but will also allow us to recover volumetric images when few x-ray measurements can be taken.

To a first approximation, ignoring x-ray scattering, the fraction of photons with energy E emitted from the x-ray source that travel towards a given detector element $I_0(\mathbf{r}, E)$ and the number of photons with energy E counted at the detector $I(\mathbf{r}, E)$ are a function of the mass attenuation coefficient $\mu(\mathbf{z}, E)$ which describes the likelihood with which x-ray photons with energy E are absorbed or scattered at location \mathbf{z} . We here use \mathbf{r} to label the line between the x-ray source and the detector element. As most x-ray detectors cannot distinguish photons of different energies, the detector output integrates over all energy levels.

This leads to the following observation model used previously in [6,13,19–29] and [30]

$$I(\mathbf{r}) = \int I(\mathbf{r}, E) dE = \int I_0(E) \exp\left(-\int_{\mathbf{r}} \mu(\mathbf{z}, E) d\mathbf{z}\right) dE, \quad (2.1)$$

where we assume that $I_0(\mathbf{r}, E)$ is independent of \mathbf{r} so that we write $I_0(E)$ instead¹.

¹In practice, $I_0(\mathbf{r}, E)$ is a function of \mathbf{r} , however, the detector is normally calibrated so that photon counts are scaled to compensate for this non-uniformity in the source.

(a) The monochromatic model

For future reference, if we had a monochromatic source or if we were able to measure x-ray intensities $I(\mathbf{r}, E)$ in narrow energy bands, then we would have a simple linear relationship

$$-\ln \frac{I(\mathbf{r}, E)}{I_0(E)} = \int_{\mathbf{r}} \mu(\mathbf{z}, E) d\mathbf{z}. \quad (2.2)$$

This model is the basis for most traditional tomographic reconstruction algorithms such as filtered backprojection and algebraic reconstruction.

Algebraic methods start with a discretisation of (2.2). Using $\mathbf{x}(E) \in \mathbb{R}^N$ as the unknown coefficient vector that approximates $\mu(\mathbf{z}, E)$ in a spatial basis and using the vector $\mathbf{a}(\mathbf{r}_i)$ to describe the integrals of each individual basis function along the path \mathbf{r}_i . Let

$$\bar{\mathbf{y}} = - \left[\ln \frac{I(\mathbf{r}_i, E)}{I_0(E)} \right]_{i \in [1, \dots, M]} \quad (2.3)$$

be the vector containing all our normalised and log transformed projection measurements. If we furthermore stack all vectors $\mathbf{a}(\mathbf{r}_i)^T$ into a matrix \mathbf{A} , then we have the linear model

$$\bar{\mathbf{y}} \approx \mathbf{A}\mathbf{x}. \quad (2.4)$$

The bar over the vector $\bar{\mathbf{y}}$ here reminds us that this is the linear model for monochromatic x-ray interaction.

(b) The polychromatic model

To use the non-linear model in (2.1) for reconstruction, we also discretise the photon energies E [21]

$$I(\mathbf{r}_i) \approx \sum_{l \in [1, \dots, L]} I_0(E_l) \exp \left(-\mathbf{a}(\mathbf{r}_i)^T \mathbf{x}(E_l) \right), \quad (2.5)$$

where the sum is now over L different energy bands E_l .

We again collect the normalised and log transformed observations $I(\mathbf{r}_i)$ into a single vector

$$\mathbf{y} = -\log \left[\frac{I(\mathbf{r}_i)}{\sum_{\tilde{l}} (I_0(E_{\tilde{l}}))} \right]_{i \in [1, \dots, M]}. \quad (2.6)$$

Using the normalised x-ray intensity vector $\mathbf{I} = \left[\frac{I_0(E_l)}{\sum_{\tilde{l}} (I_0(E_{\tilde{l}}))} \right]_{l \in [1, \dots, L]}$, we thus have the measurement model

$$\mathbf{y} \approx -\log \left(\left[e^{-\mathbf{A}\mathbf{X}} \right] \mathbf{I} \right), \quad (2.7)$$

where the exponent is evaluated component wise and where we introduce the matrix notation $\mathbf{X} = [\mathbf{x}(E_1), \dots, \mathbf{x}(E_L)] \in \mathbb{R}^{N \times L}$.

It is important to note that the columns in \mathbf{X} are related to x-ray absorption at a certain x-ray energy levels and so are the different entries in \mathbf{I} . To use this model for a fixed set of energies, we would thus either need to know the source intensities at these energy levels, or estimate \mathbf{I} at the same time as \mathbf{X} from the measurements. If we did this, then we could incorporate additional knowledge about the x-ray absorption properties of materials known to be present in a sample. However, if we estimate \mathbf{X} without any prior assumption on the materials' absorption properties, then quantisation of the energy levels is to some extent arbitrary. Thus, instead of quantising the energy levels, we might as well quantise x-ray intensities \mathbf{I} , which in turn would correspond to an unknown quantisation of the energy levels. We can then estimate the average x-ray intensity and detector efficiency using a scan without the object or from a region of the detector not in the shadow of the object. This estimate $\tilde{I} \approx \sum_{\tilde{l}} (I_0(E_{\tilde{l}}))$ can then be used in the denominator of (2.6) to normalise the observed x-ray intensities before taking the logarithm.

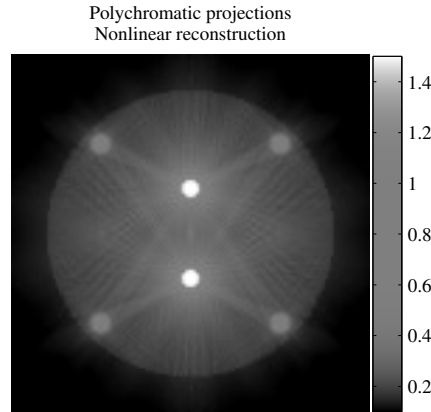


Figure 3. Reconstruction of the simulated phantom from polychromatic x-ray measurements by optimising equation (2.8).

(c) Measurement errors

So far, we have modelled the x-ray physics using the Beer-Lambert law. Whilst this model is motivated by the photon nature of x-rays, it does not directly model the quantised, probabilistic nature of x-ray generation and detection, which follow Poisson like distributions. In certain settings, especially those in which few x-rays are measured along each x-ray path, the influence of the x-ray statistics becomes more important and it is then advantageous to directly model the statistical distribution to formulate the inverse problem. In other settings, where the number of measured x-ray photons becomes large, the central limit theory allows us to make Gaussian noise assumptions.

Additional uncertainty in the measurement comes from system noise and from discretisation and quantisation effects. Spatial discretisation is further degraded as x-ray sources and detectors are not ideal points in space but cover extended areas whilst x-ray intensities are stored digitally in quantised form. Scattering [31] is another significant process that can degrade x-ray projection images and lead to reconstruction artefacts.

In this paper we will make Gaussian noise assumptions, though our approach is readily modified to use Poisson models when appropriate. We will start the development of our approach using a maximum likelihood estimate, though again, extension to maximum a posterior estimates follow similar lines. Assuming a fixed noise variance, the maximum likelihood estimate for a Gaussian model with observations described by (2.7) is then equivalent to the following minimisation problem

$$\min_{\mathbf{X}} \left\| \mathbf{y} - \log \left(\left[e^{-\mathbf{A}\mathbf{X}} \right] \mathbf{I} \right) \right\|^2. \quad (2.8)$$

To demonstrate the benefits of this non-linear optimisation problem over the linear model, we used the same simulated polychromatic x-ray measurements that were used on the left hand side of figure 2 and optimised (2.8) using 200 iterations of a gradient descend algorithm with fixed step size. We initialised the matrix \mathbf{X} as $\mathbf{x}(\phi + \theta)^T$ where \mathbf{x} is the ART reconstruction shown in figure 2 and where ϕ and θ are the photoelectric effect and Compton scattering basis functions proposed in [25].

The results are shown in figure 3, where we show the estimated mass attenuation averaged over energy. This average is calculated as

$$-\log \left(\left[e^{-\mathbf{X}} \right] \mathbf{I} \right), \quad (2.9)$$

where the exponent and logarithm are again evaluated component wise.

As can be seen, the beam hardening artefacts have been reduced though the performance is not as good as those achieved with monochromatic measurements. Part of this is due to the slow convergence of the gradient descent algorithm. This convergence is also the reason why we had to choose a sensible starting point for the algorithm, using a linear reconstruction to initialise the spatial distribution and a simple model of the energy dependent x-ray attenuation to model the energy distribution.

3. Regularisation with non convex constraints

When using model (2.6) additional assumptions are often imposed. Let us write the matrix \mathbf{X} as a factorisation $\mathbf{X} = \mathbf{D}\mathbf{S}$, where \mathbf{D} and \mathbf{S} are matrices. The matrices \mathbf{D} and \mathbf{S} can now be used to model different assumptions. For example, the i^{th} column in \mathbf{D} could contain the density of material i within each of the voxels and the i^{th} row in \mathbf{S} could contain the discretised attenuation spectrum for this material. Thus, if we assume that there are no more than J materials, then the matrix \mathbf{X} will have rank at most J . Alternatively, the formulation can be used to describe the model used in [23–26], where \mathbf{S} has two rows, one representing a known basis function to model the photoelectric effect and one basis function to model Compton scattering [25].

Different model assumptions also often lead to additional constraints on \mathbf{D} . For example, the assumption in [21] that each voxel contains a single material leads to the constraint that \mathbf{D} must have only a single entry per row. On the other hand, for the model where \mathbf{S} contains basis functions for the photoelectric effect and for Compton scattering, [24] proposed that the two coefficients in each row of \mathbf{D} could be modelled using a single parameter. A similar parameterisation of \mathbf{D} is also used in [27].

(a) Non convex regularisation

The constraints on \mathbf{X} discussed in the previous paragraph, namely the requirement that \mathbf{X} should be low rank or that \mathbf{D} should have one sparse columns can be enforced using non-convex regularisation constraints. Regularisation is a common approach used to stabilise tomographic reconstruction, preventing noise amplification and reducing artefacts due to non-ideal sampling. Traditional Tikhonov regularisation [32] as well as now standard Total Variation regularisation [33] typically lead to convex optimisation problems for which increasingly efficient algorithms are being developed [34]. Non-convex constraints, on the other hand, whilst offering increased flexibility in the way prior information can be modelled and exploited for regularisation, lead to non-convex and often combinatorial optimisation problems, which pose significantly more challenges. However, the advent of compressed sensing has, over the last ten years, shown, that many non-convex optimisation problems can nevertheless be solved efficiently using polynomial time algorithms [35,36]. Initially, compressed sensing was built around sparse, finite dimensional data models, though this has been extended to more general non-convex constraints and infinite dimensional models [37,38]. Another important extension has been the recent move away from the linear forward model of traditional compressed sensing to nonlinear forward models [39]. These recent innovations can now also be applied to x-ray tomographic reconstruction under non-linear observation models, such as that of equation (2.7). Before we look at the computational approaches that can be used to solve this inverse problem, we first introduce a range of possible non-convex constraints that might be useful in the tomographic reconstruction setting. As we will see, our algorithm will require the computation of projections of matrices or vectors onto the closest element in the non-convex constraint set so that we will pay particular attention to how this projection can be done efficiently for the different constraints we introduce.

(b) Sparsity

Sparsity is a powerful constraint. In its simplest form, sparsity constraints restrict model parameters to be predominantly zero so that the effective degrees of freedom of a model can be much smaller than the number of model parameters.

(i) Sparsity of a vector

The canonical sparse model is a vector model, where vector entries are restricted to contain few non-zero entries. More formally, consider the vector $\mathbf{x} \in \mathbb{R}^N$ in Euclidean space. We can denote the number of non-zero entries in this vector by $\|\mathbf{x}\|_0$. A sparsity constraint could then be written as

$$\|\mathbf{x}\|_0 \leq K, \quad (3.1)$$

for some $K < N$. This is a non-convex constraint. If we have two K sparse vectors \mathbf{x}_1 and \mathbf{x}_2 , then their convex combination $\lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$ is not guaranteed to be K -sparse. However, if both \mathbf{x}_1 and \mathbf{x}_2 have their non-zero elements at the same locations, then the linear combination will also be K -sparse. A K -sparse constraint thus restricts elements to lie on one of several subspaces. The set of all K -sparse vectors is thus a union of subspaces.

For an arbitrary vector \mathbf{x} , calculation of the K -sparse approximation $\hat{\mathbf{x}}_K$ with the smallest least squares error can be done extremely efficiently. All that is required is to sort the entries in \mathbf{x} by decreasing magnitude and set all but the K largest elements of \mathbf{x} to zero. We will call the calculation of the best K -sparse approximation of a vector a *projection* of the vector onto the union of subspaces.

The power of sparsity constraints comes from the fact that they are not restricted to the canonical basis, but that we can restrict coefficients in any basis. By carefully choosing the right basis, we are thus often able to model and constrain different characteristics of \mathbf{x} . For example, for our tomographic x-ray problem, wavelet bases can provide good sparse models. Let Φ be a basis matrix and define \mathbf{s} such that $\mathbf{x} = \Phi\mathbf{s}$. We then say that \mathbf{x} is sparse in the basis Φ if \mathbf{s} is sparse. If Φ is orthonormal, then it is again easy to commute the best K -sparse approximation of \mathbf{x} in the basis Φ . This is done by calculating $\mathbf{s} = \Phi^T \mathbf{x}$, which is then again thresholded by keeping only the K largest entries in magnitude of \mathbf{s} . However, if Φ is no longer orthonormal, or if Φ is over-complete, that is Φ has more than N columns, then the estimation of the best K -sparse approximation in this basis becomes a combinatorial problem in general and suboptimal algorithms have to be used.

(ii) Sparsity in matrices

We are interested in regularising the matrix valued optimisation problem 2.8 and for matrices, sparsity constraints are even more flexible. Mirroring the vector sparsity model, the simplest constraint could be a restriction on the total number of non-zero entries, however the matrix structure reflects underlying data properties and so it makes sense to treat columns separately. We might for example restrict matrix rows or columns to be sparse. A K -row-sparse matrix is then a matrix where all but K rows contain only zero entries. Again, for an arbitrary matrix \mathbf{X} , to compute the best approximation with a K -row-sparse matrix in terms of the Frobenius norm, we simply order the rows by the size of their Euclidean norm. We then set all but the largest K rows to zero, thus again projecting the matrix onto the closest K -row-sparse matrix.

Extensions of this principle again allow for the use of basis matrices to capture row or column structure. For example, a matrix \mathbf{X} is row sparse in a basis Φ if $\mathbf{X} = \Phi\mathbf{Z}$ with \mathbf{Z} being row sparse. Similarly, a matrix \mathbf{X} is column sparse in a basis Φ if $\mathbf{X} = \mathbf{Z}\Phi$ with \mathbf{Z} being column sparse. Again, for orthonormal basis Φ , calculating the best column or row sparse matrix in Φ can be done efficiently.

(iii) Block and tree sparse models

Often, additional structure can be imposed on a sparse model, further reducing the size of the constraint set. Block sparse vector models group elements of a vector into either non-overlapping

or overlapping blocks. A K sparse block model then restricts the number of blocks that can contain non-zero entries. For arbitrary vectors \mathbf{x} , computing the best block sparse approximation is easy if the model has non-overlapping blocks, but there are no efficient methods for overlapping blocks.

A special subset of bloc sparse models is a tree sparse model where a tree structure defines the blocks. A K -tree-sparse vector is then a rooted subtree with no more than K nodes. Optimal tree sparse models are difficult to compute, but fast approximate algorithms are available [40].

(c) Rank constraints

For matrix models, another powerful non-convex constraint is the rank constraint [41]. By restricting matrices to have a rank of at most K , we in effect constrain the sparsity of the singular values. The difference to sparse models is however that the left and right singular vectors can vary continuously. Whilst low rank matrices with the same left and right singular vectors again lie on the same subspace, there are infinitely many potential left and right singular vectors on so there are infinitely many of these subspaces. Low rank models thus restrict matrices to lie in one of infinitely many subspaces. For a general matrix \mathbf{X} , computing the best rank K approximation can be done using the singular value decomposition followed by a hard thresholding of the singular values based on their magnitude, which again computes the projection onto the union of subspaces model.

(d) Non-negative matrix factorisation constraint

An extension of a low rank constraint is the non-negative matrix factorisation constraint. For a non-negative matrix \mathbf{X} , a non-negative matrix factorisation is a decomposition of \mathbf{X} into two positive factors $\mathbf{D} \in \mathbb{R}^{M \times K}$ and $\mathbf{S} \in \mathbb{R}^{K \times N}$. For K smaller than M and N , this is again a low rank factorisation, however, the positivity constraint on \mathbf{D} and \mathbf{S} imposes additional constraints. Computing the best approximation of a positive matrix as a product of two positive factors is not a straightforward task, yet approximate solutions can be computed using non-negative matrix factorisation algorithms [42].

4. IHT and nonlinear IHT

There are two common approaches to deal with non-convex constraints such as those discussed above. The strategy advocated in the original compressed sensing literature is convex relaxation. Here a convex constraint set is specified that is as close as possible to the non-convex constraint. Surprisingly, in several sparse inverse problems, solving the relaxed convex problems is guaranteed to find good solutions to the original non-convex constraint. However, not all non-convex constraints lend themselves to efficient convex relaxation. Furthermore, the resulting convex algorithms are in many cases slower than alternative approaches.

The alternative to convex relaxation are so called greedy algorithms. These methods solve simpler optimisation problems in an iterative fashion. For sparse models, Greedy Pursuits are a fast family of methods that build up the sparse representation one (or a few) elements at a time. However, once chosen, a non-zero element is not normally un-selected so that the algorithm is unable to correct errors made in previous iterations. Furthermore, greedy pursuits are restricted to a small set of non-convex constraints. An alternative set of greedy methods are projection based methods. We have highlighted above the way in which we can compute the best approximation of an element with an element of our constraint set. There are now a range of algorithms that iteratively project elements back onto the non-convex constraint whilst also trying to reduce the error between the observed measurements and the model. For general non-convex constraints and linear models, the Compressed Sensing Matching Pursuit (CoSaMP) [38,43] and the Iterative Hard Thresholding algorithms (IHT) [37,44] are two common choices.

The work reported here is based on the Iterative Hard Thresholding (IHT) algorithm [44], its extension to non-linear problems suggested in [39] and the use of the conjugate gradient acceleration suggested in [45].

(a) IHT

The IHT algorithm was developed for the inversion of an underdetermined linear system under sparsity constraints. In particular, for given observations \mathbf{y} and matrix Φ , IHT tries to solve the non-convex optimisation problem

$$\min_{\mathbf{x}} \|\mathbf{y} - \Phi\mathbf{x}\|_2 \text{ such that } \|\mathbf{x}\|_0 \leq K. \quad (4.1)$$

This is done with the simple iteration:

$$\mathbf{x}^{x+1} = P(\mathbf{x}^n + \mu\Phi^T(\mathbf{y} - \Phi\mathbf{x}^n)), \quad (4.2)$$

where $P(\cdot)$ calculates the best K sparse approximation of a vector. The step size μ has to be chosen carefully to prevent instability. Under certain conditions on Φ and μ IHT is guaranteed to find near optimal solutions to the above optimisation problem [44].

The extension of the IHT algorithm to non-linear problems starts with the observation that $\Phi^T(\mathbf{y} - \Phi\mathbf{x}^n)$ in the above expression is proportional to the negative gradient of the unconstrained cost function $\|\mathbf{y} - \Phi\mathbf{x}\|_2$. Thus, if we were trying to optimise a more general non-linear cost function

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ such that } \|\mathbf{x}\|_0 \leq K, \quad (4.3)$$

then we simply replace the appropriate gradient into the algorithm and use the iteration

$$\mathbf{x}^{x+1} = P(\mathbf{x}^n - \mu\nabla f(\mathbf{x}^n)), \quad (4.4)$$

again, for a suitable choice of μ and under certain conditions on $f(\mathbf{x})$, this algorithm is guaranteed to solve the non-convex optimisation problem with a certain precision [39].

Whilst the previous discussion of the IHT algorithm concentrated on sparse models, other non-convex constraints can be enforced. This is achieved through a suitable replacement of the projection operator $P(\cdot)$. For example, if we want to estimate a matrix with rank less than K , then $P(\mathbf{X})$ is the operator that finds the best rank K approximation to \mathbf{X} . As discussed above, this can be computed using the singular value decomposition.

(b) Conjugate gradient based iterative hard thresholding for non-linear observation models

Several approaches have been suggested to increase convergence of IHT type algorithms [46–48]. For the non-linear x-ray model, we use the algorithm proposed in [39] for non-linear compressed sensing problems, however, to improve convergence, we implement an acceleration similar to that proposed in Blanchard et al. [45], who developed a conjugate gradient based variation of the linear IHT algorithm [44].

Our algorithm replaces the gradient descend step with a more general directional descend step. The update

$$\mathbf{x}^{x+1} = P(\mathbf{x}^n - \mu\nabla f(\mathbf{x}^n)), \quad (4.5)$$

is replaced by the more general update

$$\mathbf{x}^{x+1} = P(\mathbf{x}^n + \mu\mathbf{d}^n), \quad (4.6)$$

where \mathbf{d}^n is a descend direction. The computation of the descend direction is inspired by the conjugate gradient methods used for non-linear optimisation. In the first iteration or during a

conjugate gradient restart iteration, we use $\mathbf{d}^n = -\nabla f(\mathbf{x}^n)$. In other iterations, \mathbf{d} is a combination of the current gradient and the previous descend direction

$$\mathbf{d}^n = -0.5\nabla f(\mathbf{x}^n) + \beta\mathbf{d}^{n-1}, \quad (4.7)$$

where β is computed as $\beta = \frac{\|P_{\mathcal{U}}(\nabla f(\mathbf{x}^{[n]})\|^2}{\|(P_{\mathcal{U}}(\nabla f(\mathbf{x}^{[n-1]}))\|^2}$. We here use the projection operator $P_{\mathcal{U}}$ which projects the gradient at the current and the previous estimate onto the subspace \mathcal{U} , which is the subspace in the union of subspace model in which the current estimate \mathbf{x}^n lies. This is an important difference to the more standard conjugate gradient method. In effect, we assume that our optimisation problem is restricted to the subspace \mathcal{U} . Whenever we change subspaces (that is, whenever the projection $P(\cdot)$ projects onto a different subspace, then our optimisation problem changes so that we restart the conjugate gradient method by using $\beta = 0$ in the next iteration.

In this form, the algorithm can be used for non-convex constraint sets that are unions of linear subspaces. In this setting, the algorithm relies on four operations. The calculation of a cost function $f(\hat{\mathbf{x}})$, the evaluation of its gradient $g(\hat{\mathbf{x}})$, the non-linear projection operator $P(\cdot)$ and the linear projection operator $P_{\mathcal{U}}$. For our matrix valued problem with $\hat{\mathbf{x}} = \hat{\mathbf{X}}$ we here use

$$f(\hat{\mathbf{X}}) = \left\| \mathbf{y} + \log \left(\left[e^{-\mathbf{A}\hat{\mathbf{X}}} \right] \mathbf{I} \right) \right\|^2, \quad (4.8)$$

$$g(\hat{\mathbf{X}}) = \nabla f(\hat{\mathbf{X}}). \quad (4.9)$$

To better understand the projection operators $P(\cdot)$ and $P_{\mathcal{U}}$, assume we have a vector $\mathbf{x} = [-1 \ 0.2 \ 4 \ -2 \ 0.9]^T$ and we use a 2-sparse non-convex constraint set. The projection of \mathbf{x} onto this set is $P(\mathbf{x}) = [0 \ 0 \ 4 \ -2 \ 0]^T$. We can then also specify the subspace \mathcal{U} as all those vectors that have non-zero entries in the third and fourth position, that is, all vectors that can be written as

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}. \quad (4.10)$$

The projection $P_{\mathcal{U}}$ that projects onto this subspace simply sets the first, second and fifth entry in a matrix to zero. Alternatively, for matrix factorisation models, we use a similar approach. Assume $\mathbf{X} \approx \mathbf{D}\mathbf{S}$, where $\mathbf{D}\mathbf{S}$ is the best approximation of \mathbf{X} as a low rank (or a nonnegative low rank) decomposition. We then use the matrix \mathbf{D} to define the subspace onto which we can project any \mathbf{X} using $\mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{X}$.

This leaves us with the problem of estimating a good step size μ . We here perform a line search in direction \mathbf{d} . This is done using a quadratic approximation to the line search cost function

$$\min_{\mu} f(\mathbf{x}^n + \mu P_{\mathcal{U}}(\mathbf{d})), \quad (4.11)$$

where we again only optimise within the subspace \mathcal{U} . The approximation of $f(\mathbf{x}^n + \mu P_{\mathcal{U}}(\mathbf{d}))$ fits a quadratic function to match the magnitude and gradient of $f(\mathbf{x}^n)$ evaluated at the current estimate \mathbf{x}^n as well as the value and gradient of $f(\mathbf{x}^n + a P_{\mathcal{U}}(\mathbf{d}))$ for a given test step-size a . Thus, the step size μ used is that which achieves the minimum of the quadratic function. In addition to the conjugate gradient restart due to the algorithm moving from one subspace to another, standard non-linear conjugate gradient restart conditions are also monitored (such as those based on a gradient orthogonality condition or on maximum iteration counts). Note, that for matrix factorisation based constraints, the subspace changes continuously between iterations so that a subspace based restart would mean that we simply use gradient descend. In this case, it makes sense to only use gradient orthogonality conditions or maximum iteration counts to define the conjugate-gradient restarts.

The algorithm can thus be summarised as follows:

- Input: $\mathbf{X}^{[0]}$, P , f , and g

- $\mathbf{p}^{[0]} = 0$
- if $\|X^{[0]}\| = 0$
 - $\mathcal{U} = \text{subspace}(P(g(\mathbf{X}^{[0]})))$
- else
 - $\mathcal{U} = \text{subspace}(P(\mathbf{X}^{[0]}))$
- $a = \|P_{\mathcal{U}}(g(\mathbf{X}^{[0]}))\|^{-1}$
- : Iterate ($n = 1, n + +$) until some stopping criterion is met:
 - if $n = 1$ or if the conjugate gradient restart condition is satisfied
 - $\beta = 0$
 - else
 - $\beta = \frac{\|P_{\mathcal{U}}(g(\mathbf{x}^{[n]}))\|^2}{\|(P_{\mathcal{U}}(g(\mathbf{x}^{[n-1]})))\|^2}$
 - $\mathbf{d} = -0.5g(\mathbf{X}^{[n]}) + \beta * \mathbf{p}^{[n-1]}$
 - $a = \text{linesearch}(f, g, \mathbf{d}, \mathcal{U}, a)$
 - $\mathbf{X}^{[n+1]} = P(\mathbf{X}^{[n]} + a\mathbf{d})$
 - $\mathcal{U} = \text{subspace}(\mathbf{X}^{[n+1]})$
 - $\mathbf{p}^{[n]} = \mathbf{d}$.

(c) Extension to mixed projections

The above algorithm assumes that P projects onto one of several linear subspaces. Whilst many of the constraints we have discussed above fall into this category, other constraints or combinations of constraints will lead to more general projections. For example, we know that the x-ray attenuation coefficients are positive. It thus would make sense to include positivity constraints into our non-convex constraints. This raises two issues, firstly, a model with a positivity constraint is no longer a subspace model and secondly, projecting onto the union of a non-convex subspace model and the positive orthant is far from trivial in most cases. We here use what we call mixed projections in which we first project onto the subspace model and then project onto the second constraint. The second constraint does not have to be convex (as the positivity constraint) but can be a more general constraint. For example, we might want to constrain the reconstruction matrix to be sparse in the wavelet domain as well as have a low-rank structure. We do this by first projecting onto the sparse wavelet model using a simple wavelet thresholding approach. This is then followed by a low-rank approximation.

We found that inclusion of this additional projection generally improves performance. In the above algorithm, we insert this projection step into the algorithm as the second to last line, that is, after the projection onto the non-convex constraint and the estimation of the subspace. Whilst this works well with a projection onto the positive orthant as well as the non-negative matrix decomposition, for the more general low-rank projection, we found that we were no longer optimising within a subspace. In this setting we thus used the conjugate gradient method to optimise in the full space, that is, we compute β as $\frac{\|g(\mathbf{x}^{[n]})\|^2}{\|(g(\mathbf{x}^{[n-1]}))\|^2}$, that is, using the unprojected direction vectors. Note however that we still conduct the line search within the subspace defined by the subspace projection.

5. Numerical results

(a) Artificial data

Our simulations are based on two phantoms. The disk phantom in figure 1 and the Shepp-Logan phantom. The disk phantom simulates a phantom with four materials, an aluminium ring with inner silicon disk, four silver inclusions in the aluminium ring and two gold inclusions in the silicon disk. For the Shepp-Logan phantom, we partitioned the original Shepp-Logan phantom into different areas, one for each of the different grey levels. Each of these areas was then assumed to be either a uniform mixture of three materials (Aluminium, Iron and Silver), or was assumed

to be one of these materials. For the first case, mixture coefficient are generated randomly, whilst in the second case, materials were assigned to regions randomly.

In all simulations, mass attenuation coefficients are taken from the NIST database. We model an x-ray source operating at 120keV using a tungsten target. The detector is assumed to be caesium iodine with a linear response. Energy levels were quantised in bins with centre energies of 30, 35, 40, 45, 50, 52.5, 55, 57.5, 60, 62.5, 65, 67.5, 70, 77.5, 85, 92.5, 100, 112.5 and 115 keV. Non-linear x-ray projections were generated on a different spatial grid to the grid used to generate the system matrix \mathbf{A} for the reconstruction algorithm. For the reconstruction, we used an energy quantisation of 10 levels with centre energies of 30, 40, 50, 55, 60, 65, 70, 85, 100 and 115 keV.

The disk phantom was measured using 40 cone-beam projections from equally spaced angles, covering 360 degrees. Each projection was measured with a linear x-ray detector array with 181 sensor elements. Both, the original phantom and the reconstruction were on a 128 y 128 spatial grid, with the difference that the simulated measurements used a grid rotated by 45 degrees. Measurements from the Shepp-Logan based phantom, were generated similarly, with the difference that we only used 32 projections. For the reconstruction, we assumed knowledge of I_0 but did not use knowledge of the mass attenuation coefficients for the three materials. Observations are then calculated as

$$\mathbf{y} = -\log \left(\left[e^{-\mathbf{A}\mathbf{D}\mathbf{S}} \right] I_0 \right). \quad (5.1)$$

When simulating the Shepp-Logan phantom in which each homogeneous region consists of a mixture of the three materials, then $\mathbf{D} \in \mathbb{R}^{N \times 3}$ is a matrix with repeated randomly generated rows (with entires drawn from a uniform distribution and normalised to sum to 1), with a different row for each grey level in the Shepp-Logan phantom. To simulate the Shepp-Logan phantom in which each homogeneous region consists of a single material, the matrix \mathbf{D} has a single non-zero entry per row, which is set to one. For each homogeneous region, the location of the non-zero entry is generated randomly. In both cases, $\mathbf{S} \in \mathbb{R}^{3 \times 19}$ is a matrix with the quantised mass attenuation coefficients for the three materials at the 19 different energy levels. I_0 is a vector that counts the average number of x-ray photons in a given energy band transmitted by the source and converted to visible light in the scintillator. We added Gaussian noise to the observations, normalised to give a signal to noise ration of 40dB.

In order to be able to compare linear reconstruction methods which estimate a single x-ray attenuation image and non-linear methods which estimate attenuation images for each energy level, we here compare the performance in terms of the algorithms ability to estimate the average attenuation calculated as

$$-\log \left(\left[e^{-\mathbf{X}} \right] \mathbf{I} \right). \quad (5.2)$$

The average attenuation can be compared directly to the estimates from the linear estimates, whilst for the non-linear estimates, we also average the estimated attenuation coefficients over energy levels before comparison.

Results for the disk phantom are shown in Figures 4 and 5 for the linear reconstruction and the non-linear reconstructions respectively. We here compared the filtered backprojection algorithm available through the Matlab image processing toolbox, the Algebraic Reconstruction Technique implemented in the AIRTools toolbox [1] (using the randomised Kaczmarz version of the algorithm) with our conjugate gradient algorithm that uses the linear model with two different constraints, wavelet sparsity and wavelet tree sparsity. We also used a total variation constrained inversion using the TVAL3 code (<http://www.caam.rice.edu/~optimization/L1/TVAL3/>). The algebraic method performs better than filtered back projections. Including additional constraints further improves performance, with the TV regularised results giving the highest signal to noise ratio here, though the beam hardening artefacts are still very strong. Also, none of the methods is able to allow us to distinguish between the aluminium and silicon.

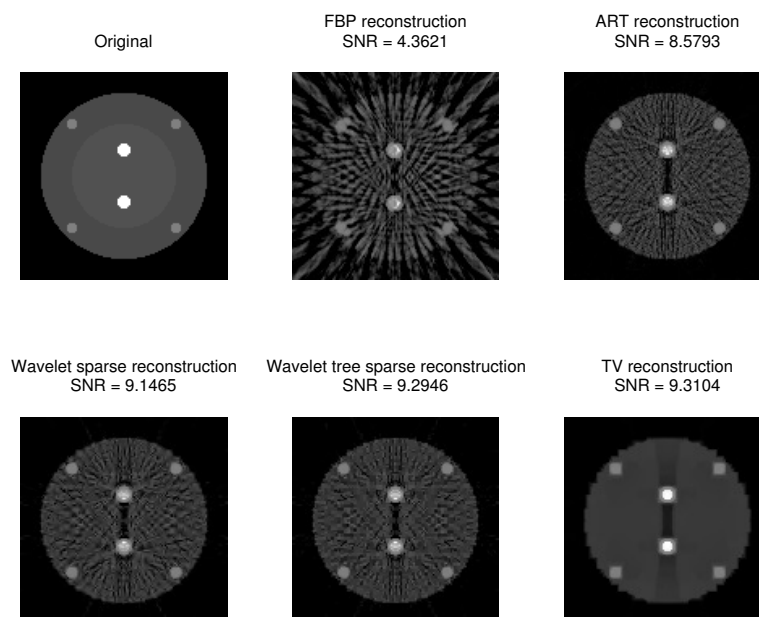


Figure 4. Comparison of linear reconstruction methods on a challenging phantom from 40 projections. The original is an object of aluminium, silicon, gold and silver. Filtered Backprojection gives a reconstruction with severe artefacts and of the iterative solvers, the TV regularised method performs best.

The reconstructions with the non-linear model are shown in figure 5. We here use our conjugate gradient method with a range of constraints, positivity, low rank, wavelet row-sparsity, wavelet-tree sparsity, a combination of wavelet row-sparsity and low-rank, a combination of wavelet row-sparsity and non-negative matrix factorisation and non-negative matrix factorisation on its own. Whilst some of the results have reduced beam-hardening artefacts, the artefacts are still visible here. However, the use of the matrix models has led to better SNR figures, especially for the model that combines wavelet row-sparsity with a non-negative matrix factorisation. For the same constraint, the non-linear model leads here to worse SNR results compared to the linear model. This is due to the fact that the non-linear model has significantly more parameters. But once we start to use some of the constraints that are only available for matrix models, such as the non-negative matrix decomposition constraint, do we see improvements over the linear model.

The results for the Shepp-Logan phantom are summarised in figures 6 and 7, where we show box plots of the SNR values achieved with the different methods for 10 different random instances of the two problem settings.

Again, whilst the non-convexly constrained linear approaches do in general work as well if not better than the approaches based on the non-linear cost using the same constraints, using more advanced constraints such as the non-negative matrix factorisation can perform much better.

We found that we consistently got the best results either with a combination of wavelet sparsity in the spatial domain combined with an additional projection onto a non-negative matrix factorisation or with the non-negative matrix factorisation constraint alone. Looking at these results in the light of the fact that we are unable to calculate exact projections onto the union of these two constraints and the fact that our algorithm is primarily designed for subspace models (which the non-negative factorisation is not) shows that the two constraints are capturing important complementary properties of the data. This demonstrates that there is a need to study these joint constraints in more detail in the future.

The only linear cost function based result that is not matched by the non-linear methods is the linear approach with the TV constraint. This suggests that the TV constraint is much more

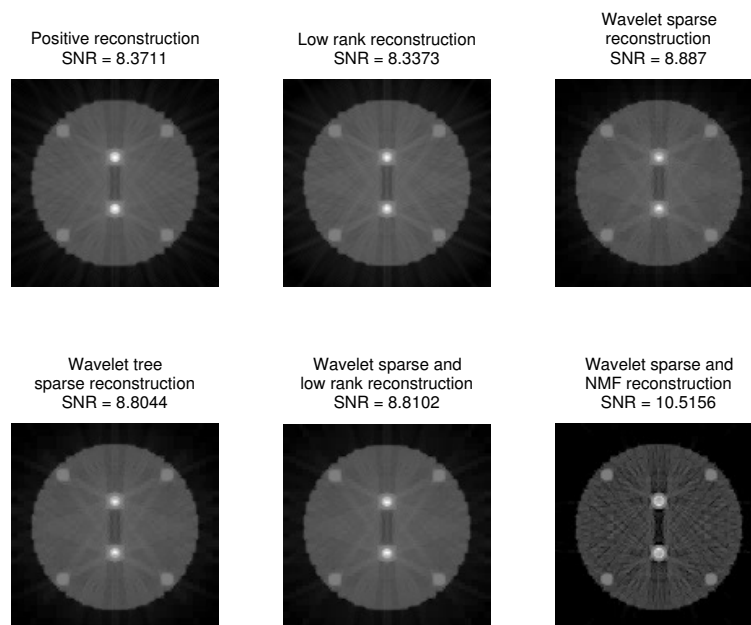


Figure 5. Comparison of non-linear reconstruction methods on a challenging phantom from 40 projections. The original is an object of aluminium, silicon, gold and silver. The non-linear reconstruction performs better under certain constraints than the linear methods. The best performing non-convex constraint is the combination of wavelet sparsity with a non-negative matrix factorisation. When using wavelet-sparsity here, we use a row sparse matrix model as described in subsection 3ii.

powerful for our phantoms than the related wavelet constraints we use here. This suggests the use of similar constraints also with our non-linear model, which can be easily done by adding a TV regularisation term to the non-linear cost function. Combining this with the non-negative matrix factorisation constraint is likely to offer additional benefits, though this is still under investigation.

(b) Real data

To demonstrate the applicability of the technique to real data, we applied it to two real data sets, comparing standard filtered back projection to the best performing approach in the previous experiment.

The first dataset was acquired at 360kV. We used 32 projections and a line array detector. The scanned object was a model of an engine block made of aluminium. The longest dimension of the object was about 15cm. Reconstruction of a two dimensional cross section using filtered backprojections is shown in figure 8. The reconstruction with a non-linear model and a sparsity and non-negative matrix factorisation constraint are shown in figure 10. Due to the extremely small number of projections, severe artefacts are visible in the filtered backprojection. Using our non-linear model and a wavelet sparsity constraint, the streak artefacts have been removed but the wavelet constraint has introduced some smoothing.

The second set of x-ray projections was acquired with 200kV. The test object consisted of two concentric tubes. The outer tube was acrylic and the inner tube aluminium. Within the inner tube, aluminium wires of different diameters were attached. We used 200 projections acquired with a flat panel detector, taking only the central slice for reconstruction. The filtered backprojection reconstruction is shown on the left in figure 10 whilst our nonlinear reconstruction with a positivity constraint is shown on the right. The red oval highlights beam hardening artefacts

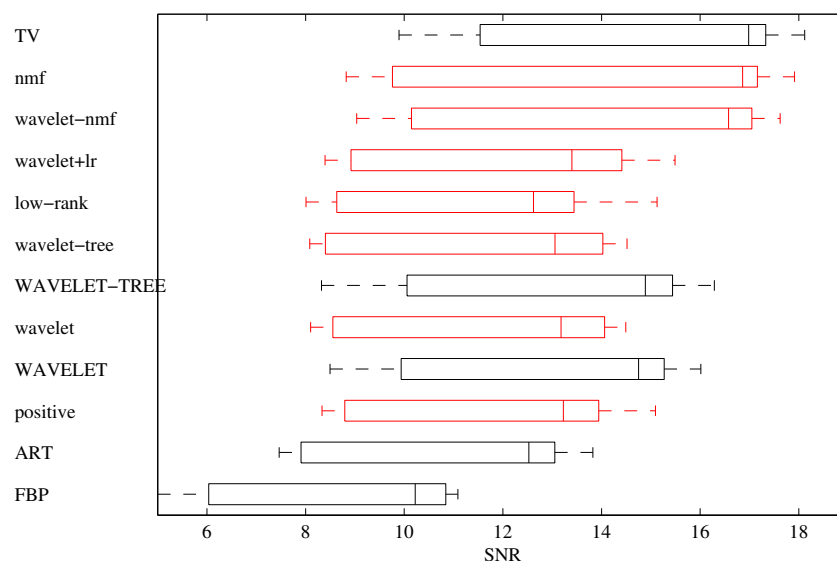


Figure 6. Distribution of results for different methods and for 10 different material assignments, where each area in the phantom is assigned a single material. The top seven methods (lower case labels) use the non-linear model in the reconstruction whilst the lower five methods (upper case labels) use the linear model. From top to bottom, the methods are constraint using 1) non-negative matrix factorisation projections, 2) mixed wavelet sparsity and non-negative matrix factorisation projections, 3) mixed wavelet sparsity and low rank matrix factorisation projections, 4) wavelet tree sparsity, 5) wavelet sparsity, 6) low-rank matrix projections, 7) positivity, 8) Total Variation regularisation, 9) wavelet tree sparsity and 10) wavelet sparsity. The second to last result are those achieved with Algebraic Reconstruction Technique [1] and the last results are those for Filtered Backprojection.

visible in the filtered back projection, which are not visible in the reconstruction achieved with our approach.

6. Conclusion

In this paper we have proposed a projected conjugate gradient algorithm that is able to solve nonlinear inverse problems under a range of non-convex constraints. We have applied the method to the inversion of the non-linear x-ray transmission model to solve x-ray computed tomography reconstruction. This not only allowed us to address the beam hardening problem but also allowed us to reconstruct images from few projections. Extensive simulation results on synthetic data have shown that advanced non-convex constraints, such as a combination of wavelet sparsity and non-negative matrix factorisation can have significant benefits over standard positivity constraints alone. Whilst these non-convex constraints do not guarantee globally optimal solutions, we could show that when we initialised our method with a good linear reconstruction, then the non-convexly constraint nonlinear reconstructions were often better. However, the improvement also comes at the cost of increased computational complexity. For real data, inverting the non-linear model often took hundreds of iterations to achieve the best performance.

The fact that total variation regularisation outperforms even our best non-linear model based estimate, shows the power to the TV constraint for objects with uniform material density as we simulated here. It is likely that the inclusion of such a TV regularisation term in our non-linear model will offer similar benefits, especially when combined with the non-negative matrix factorisation constraint. Using joint constraints has also demonstrated clear benefits. Whilst we enforced these here with consecutive projections, recent studies (not reported here) have shown that an approach that averages the projections onto the individual constraints tends to perform

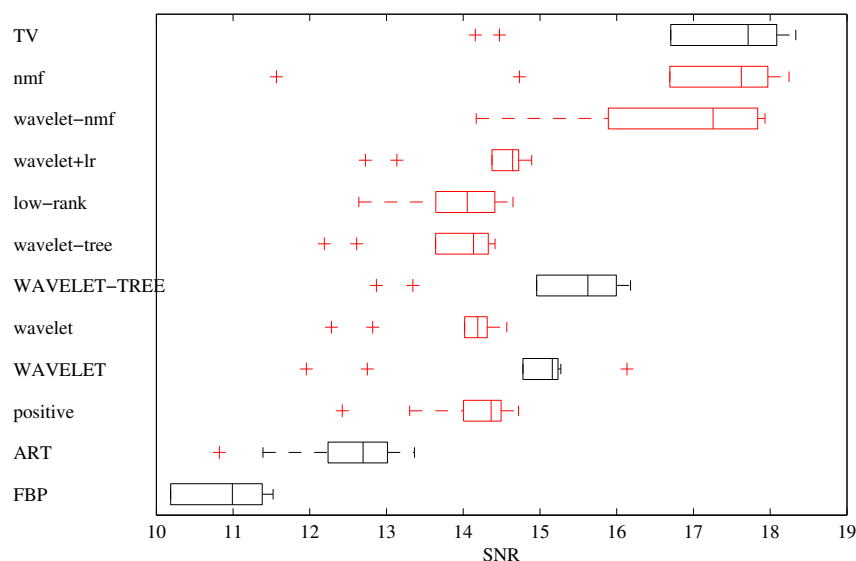


Figure 7. Distribution of results for different methods and for 10 different material assignments, where each area in the phantom is assigned a random mixture of all three materials. The top seven methods (lower case labels) use the non-linear model in the reconstruction whilst the lower five methods (upper case labels) use the linear model. From top to bottom, the methods are constraint using 1) non-negative matrix factorisation projections, 2) mixed wavelet sparsity and non-negative matrix factorisation projections, 3) mixed wavelet sparsity and low rank matrix factorisation projections, 4) wavelet tree sparsity, 5) wavelet sparsity, 6) low-rank matrix projections, 7) positivity, 8) Total Variation regularisation, 9) wavelet tree sparsity and 10) wavelet sparsity. The second to last result are those achieved with Algebraic Reconstruction Technique [1] and the last results are those for Filtered Backprojection.

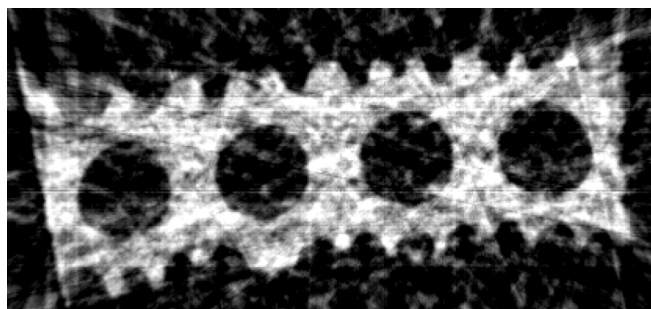


Figure 8. Filtered back-projection reconstruction from 32 projections of a slice through a model engine block made of aluminium.

better than the approach used here, though a detailed study how this works in the tomographic setting is still to be undertaken.

The main drawback of our approach currently is the slow convergence due to the complex nature of the non-convex cost function. This was here addressed with a sensible initialisation approach, though additional benefits are likely if better optimisation strategies are adopted. One possible improvement might be the use of a better line search strategy. The quadratic approximation does not always seem to give a good step size and other approaches are currently under investigation.

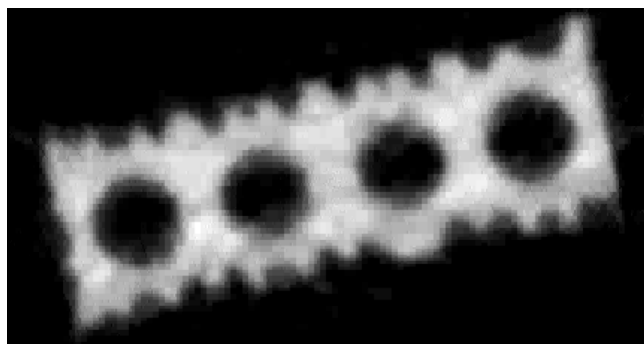


Figure 9. Reconstruction with the non-linear model and a sparsity and non-negative matrix factorisation constraint. Reconstruction from 32 projections of a slice through a model engine block made of aluminium.

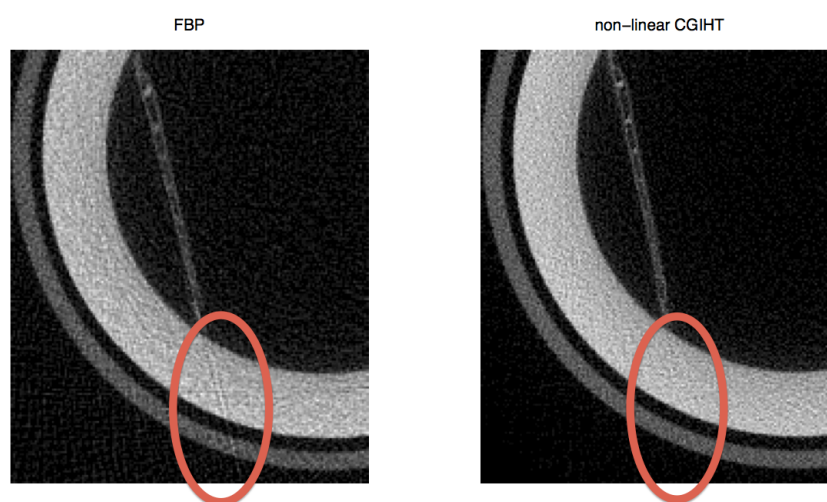


Figure 10. Detail of two reconstructions from 200 projections. Filtered back-projection (left) and non-linear conjugate gradient (right) reconstructions of a slice through a high density workpiece with internal structures. Beam hardening artefacts are visible in the Filtered Backprojection reconstruction which are not visible in the reconstruction achieved with our approach.

Acknowledgment

Thomas Blumensath acknowledges support from EPSRC grants EP/K037102/1, EP/J005444/1, the CCPi (funded through EPSRC grant EP/J010456/1) and the TSB (grant 101804) as well as a University of Southampton, Faculty of Engineering and the Environment "New Frontiers Fellowship".

References

1. PC Hansen M S 2012 *Journal of Computational and Applied Mathematics* **236** 2167–2178
2. Brooks R A and Di Chiro G 1976 *Phys. Med. Biol.* **21** 390–398
3. Van de Castele E, Van Dyck D, Sijbers J and Raman E 2004 *Journal of X-ray science and technology* **12** 53–57

4. So A, Hsieh J, Li J Y and Lee T Y 2009 *Phys. Med. Biol.* **54** 3031–3050 application of polynomial model
5. Hammersberg P and Mangard M 1998 *Journal of X-Ray Science and Technology* **8** 75–93
6. Herman G T 1979 *Phys. Med. Biol.* **24** 81–106
7. Herman G and Trivedi S 1983 *IEEE Transactions on Medical Imaging* **MI-2**
8. Hsieh J, Molthen R C, Dawson C A and Johnson R H 2000 *Med. Phys.* **27** 23–29 extends the single material polynomial model to multiple materials and estimates intensity ratios iteratively
9. Gao H, L Z, Chen Z, Xing Y and Li S 2006 *IEEE Transactions on Nuclear Science* **53**
10. Vedula V S V M and Munshi P 2008 *NDT & E International* **41** 25–31
11. ONalcioglu and Lou R Y 1979 *Phys. Med. Biol.* **24** 330–340
12. Rueggsegger P, Hangartner T, Keller H U and Hinderling T 1978 *J Comput Assist Tomogr* **2** 184–188 uses initial recon for classification and then estimates correction factor
13. Elbakri I A and Fessler J A 2002 *IEEE Transactions on Medical Imaging* **21** 89–99 uses initial material classification, single material per location model, and statistical reconstruction algorithm
14. Krumm M, Kasperl S and Franz M 2008 *NDT & E International* **41** 242–251
15. Krumm M, Kasperl S and Franz M 2008 Referenceless beam hardening correction in 3d computed tomography images of multi-material objects *World Conference on Nondestructive Testing*
16. Meagher J M, Mote Jr C D, and Skinner H B 1990 *IEEE Trans on Nuclear Science* **31**
17. Yan C H, Whalen R T, Beaupre G S, Yen S Y and Napel S 2000 *IEEE transactions on medical imaging* **19**
18. Chen C Y, Chuang K S, Wu J, Lin H R and Li M J 2001 *J. Digit. Imaging* **14** 54–61
19. McDavid W D, Waggenger R G, Payne W H and Dennis M J 1975 *Medical Physics* **2** 321 – 324
20. McDavid W D, Waggenger R G, Payne W H and Dennis M J 1977 *Medical Physics* **4** 54 – 58
21. Van Gompel G, Van Slambrouck K, Defrise M, Batenburg K J, de Mey J, Sijbers J and Nuyts J 2011 *Med Phys.* **38**
22. Kachelriess M and Kalender W A 2005 Improving pet/ct attenuation correction with iterative ct beam hardening correction *Proceedings of the IEEE Nuclear Science Symposium Conference*
23. Alvarez R E and Macovski A 1976 *Phys. Med. Biol.* **21** 733–744
24. Stonestrom J P, Alvarez R E and Macovski A 1981 *IEEE transactions on biomedical engineering* **BME-28**
25. De Man B, Nuyts J, Dupont P, Marchal G and Suetens P 2001 *IEEE Transactions on Medical Imaging* **20** 999–1007
26. Menvielle N, Goussard Y, Orban D and Soulez G 2005 Reduction of beam-hardening artefacts in x-ray ct *Proceedings of the IEEE Engineering in Medicine and Biology Annual Conference* (Shanghai, China)
27. Elbakri I A and Fessler J A 2003 *Phys. Med. Biol.* **48** 2453–2477
28. Olsen E A, Han K S and Pisano D J 1981 *IEEE Transactions on Nuclear Science* **NS-28**
29. Van de Castele E, Van Dyck D, Sijbers J and Raman E 2002 *Phys. Med. Biol.* **47** 4181–4190
30. Elbakri I A and Fessler J A 2001 Statistical x-ray-computed tomography image reconstruction with beam-hardening correction *Proc. Of SPIE. Medical Imaging: Image Processing*
31. Ruehrnschopf E P and Klingenberg K 2011 *Med. Phys.* **38** 4296–4311
32. Tikhonov A N and Arsenin V Y 1977 *Solution of Ill-posed Problems* (Washington: Winston & Sons)
33. Rudin L, Osher S and Fatemi E 1992 *Physica D: Nonlinear Phenomena* **60**(1-4) 259–268
34. Hämläinen K, Harhanen L, Hauptmann A, Kallonen A, Niemi E and Siltanen S 2014 *International Journal of Tomography and Simulation* **25**(1) 1–25
35. Candès E, Romberg J and Tao T 2006 *IEEE Trans. on Information Theory* **52**(2) 489 – 509
36. Donoho D 2006 *IEEE Trans. on Information Theory* **52**(4) 1289 – 1306
37. Blumensath T 2011 *IEEE Transactions on Information Theory* **57**(7) 4660–4671
38. Baraniuk R, Cevher V, Duarte M and Hegde C 2010 *IEEE Transactions on Information Theory* **56** 1982–2001
39. Blumensath T 2013 *IEEE Trans. on Information Theory* **59**(6) 3466–3474
40. Baraniuk R G 1999 *Wavelet Applications in Signal and Image Processing VII* **3813** 196–207
41. Candès E and Recht B 2009 *Foundations of Computational Mathematics* **9** 717–772
42. Paatero P and Tapper U 1994 *Environmetrics* **5** 111–126

43. Needell D and Tropp J A 2009 *Applied and Computational Harmonic Analysis* **26**(3) 301–321
44. Blumensath T and Davies M E 2009 *Applied and Computational Harmonic Analysis* **27**(3) 265–274
45. Blanchard J, Tanner J and Wei K 2014 CGIHT: Conjugate gradient iterative hard thresholding for compressed sensing and matrix completion Tech. rep. University of Oxford URL <http://eprints.maths.ox.ac.uk/1833/>
46. Blumensath T 2011 *Signal Processing* **92** 752–756
47. Foucart S 2011 *SIAM Journal on Numerical Analysis* **49**(6) 2543–2563
48. Cevher V 2011 An alps view of sparse recovery *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)* pp 5808–5811