

PAPER • OPEN ACCESS

# Directional sinogram inpainting for limited angle tomography

To cite this article: Robert Tovey *et al* 2019 *Inverse Problems* **35** 024004

View the [article online](#) for updates and enhancements.

## You may also like

- [deepCR: Cosmic Ray Rejection with Deep Learning](#)  
Keming Zhang, and Joshua S. Bloom
- [An image inpainting algorithm using exemplar matching and low-rank sparse prior](#)  
Qiangwei Peng and Wen Huang
- [A Review of PDE Based Local Inpainting Methods](#)  
Ahmed K. Al-Jaberi and Ehsan M. Hameed

# Directional sinogram inpainting for limited angle tomography

Robert Tovey<sup>1,6</sup>, Martin Benning<sup>2</sup>, Christoph Brune<sup>3</sup>,  
Marinus J Lagerwerf<sup>4</sup>, Sean M Collins<sup>5</sup>, Rowan K Leary<sup>5</sup>,  
Paul A Midgley<sup>5</sup> and Carola-Bibiane Schönlieb<sup>1</sup>

<sup>1</sup> Centre for Mathematical Sciences, University of Cambridge, Cambridge, United Kingdom

<sup>2</sup> School of Mathematical Sciences, Queen Mary University of London, London, United Kingdom

<sup>3</sup> University of Twente, Enschede, Netherlands

<sup>4</sup> Centrum Wiskunde & Informatica, Amsterdam, Netherlands

<sup>5</sup> Department of Materials Science and Metallurgy, University of Cambridge, Cambridge, United Kingdom

E-mail: [RobTovey@maths.cam.ac.uk](mailto:RobTovey@maths.cam.ac.uk)

Received 29 April 2018, revised 20 October 2018

Accepted for publication 22 November 2018

Published 2 January 2019



## Abstract

In this paper we propose a new joint model for the reconstruction of tomography data under limited angle sampling regimes. In many applications of tomography, e.g. electron microscopy and mammography, physical limitations on acquisition lead to regions of data which cannot be sampled. Depending on the severity of the restriction, reconstructions can contain severe, characteristic, artefacts. Our model aims to address these artefacts by inpainting the missing data simultaneously with the reconstruction. Numerically, this problem naturally evolves to require the minimisation of a non-convex and non-smooth functional so we review recent work in this topic and extend results to fit an alternating (block) descent framework. We perform numerical experiments on two synthetic datasets and one electron microscopy dataset. Our results show consistently that the joint inpainting and reconstruction framework can recover cleaner and more accurate structural information than the current state of the art methods.

Keywords: tomography, limited angle tomography, inpainting, anisotropy, variational, nonconvex, optimization

(Some figures may appear in colour only in the online journal)



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

<sup>6</sup> Author to whom any correspondence should be addressed.

## 1. Introduction

### 1.1. Problem formulation

Many applications in materials science and medical imaging rely on the x-ray transform as a mathematical model for performing 3D volume reconstructions of a sample from 2D data. We shall refer to any modality using the x-ray transform forward model as x-ray tomography. This encompasses a huge range of applications including positron emission tomography (PET) in front-line medical imaging [1], transmission electron microscopy (TEM) in materials or biological research [2–4], and x-ray computed tomography (CT) which enjoys success across many fields [5, 6].

The limited angle problem is common in x-ray tomography, for instance in TEM [7] and mammography [8], and is caused by a particular limited data scenario. Algebraically, we search for approximate solutions to the inverse problem

$$\text{Given data } b \text{ find optimal pair } (u, v) \text{ such that } Sv = b, \mathcal{R}u = v$$

where  $\mathcal{R}$  is the x-ray transform to be defined in (2.1),  $S$  represents the limited angle sub-sampling pattern described in figure 1. Typically, limited angle problems can occur due to having a large sample or because equipment does not allow the sample to be fully rotated. Mathematically, microlocal analysis can be used to categorise the limited angle problem and characterise artefacts that occur. Viewed through the Fourier slice theorem, it becomes clear that the Fourier coefficients of  $u$  are partitioned into those ‘visible’ in  $b$  and those contained in a ‘missing wedge’ [9]. These coefficients are referred to respectively as the visible and invisible singularities of  $u$ . The limited angle problem then is both a denoising and inpainting inverse problem, on the visible and invisible singularities respectively. The artefacts caused by the missing wedge can be explicitly characterised [10, 11] and examples of such streak artefacts and blurred boundaries can be seen in figures 2 and 3.

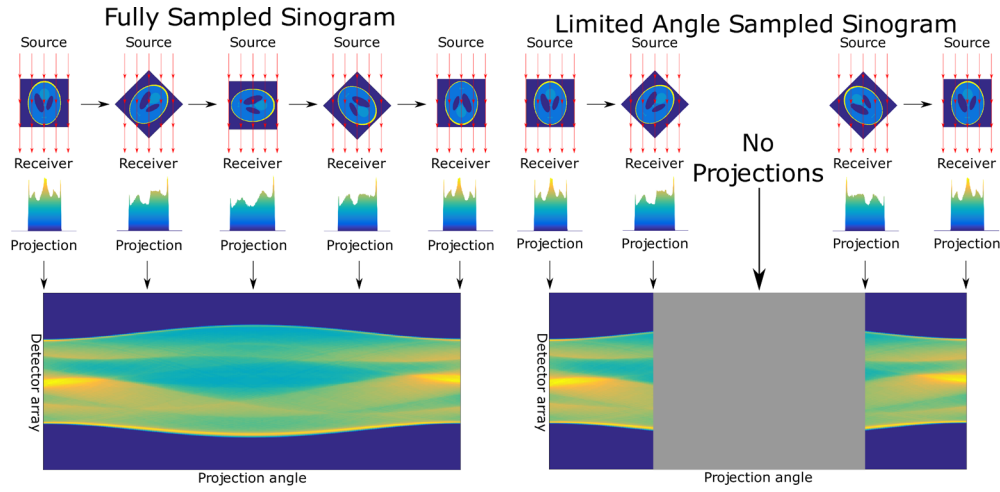
Whilst the techniques developed here can apply to any limited angle tomography problem, we focus on the application of TEM for specific examples.

### 1.2. Context and proposed model

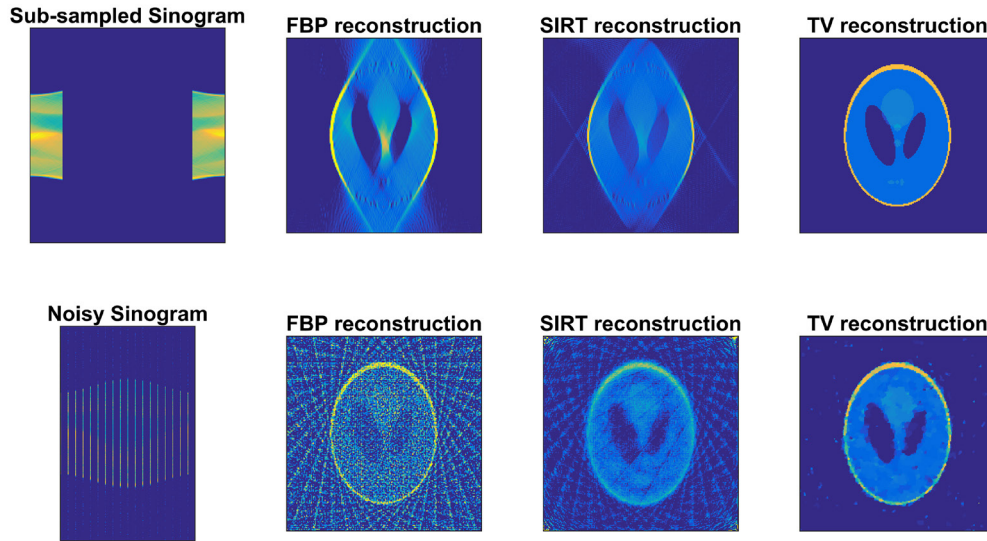
Traditional methods for x-ray tomography reconstruction find approximate solutions to  $S\mathcal{R}u = b$ , constraining  $\mathcal{R}u = v$  and only using prior knowledge of  $u$  or the sinogram,  $v$ . There are three main methods which fit into this category:

- Filtered back projection (FBP) is a linear inversion method with smoothing on the sinogram,  $v$ , to account for noise [12–14].
- The simultaneous iterative reconstruction technique (SIRT) can be thought of as a pre-conditioned gradient descent on the function  $\|S\mathcal{R}u - b\|_2^2$  [3, 15–17]. Regularisation is then typically implemented by an early-stopping technique.
- Variational methods where prior knowledge is encoded in regularisation functionals have now been applied in this field for nearly a decade. In particular, the current state of the art in electron tomography is total variation (TV) regularisation [2, 18, 19] where  $u$  is encouraged to have a piecewise constant intensity. This will be introduced formally in section 2.

FBP and SIRT are commonly used for their speed although variational methods like TV have quickly gained popularity as they enable prior physical knowledge to be explicitly incorporated in reconstructions. This added prior knowledge tends to stabilise the reconstruction

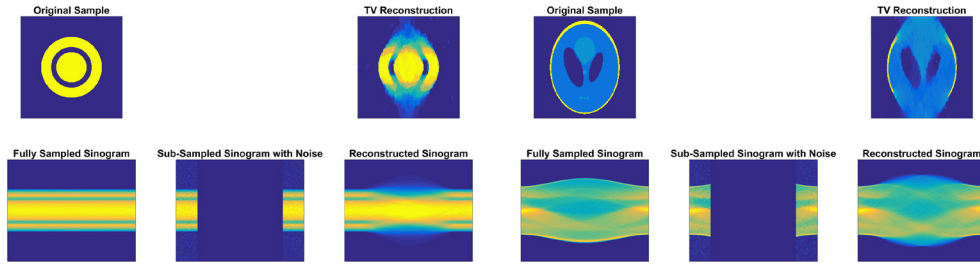


**Figure 1.** Diagrammatic representation of the acquisition of 2D x-ray transform data, the sinogram, in both full range and limited angle acquisition. Note that measurement at  $180^\circ$  is exactly a reflection of that at  $0^\circ$ . This symmetry allows us to consider a  $180^\circ$  range of the sinogram as a full sample. In the limited angle setting we can only rotate the sample a small amount clockwise and anti-clockwise which results in missing data in the middle of the sinogram.



**Figure 2.** Demonstration of TV reconstruction in comparison to FBP and SIRT. The top row shows reconstructions from noise-less limited angle data and the bottom shows reconstructions from noisy limited view data (far left images). Comparing the columns, we immediately see that FBP and SIRT are much more prone to angular artefacts than TV. In both cases we notice that the TV reconstructions better show the broad structure of the phantom.

process and figure 2 gives examples where TV can vastly outperform FBP and SIRT when either the noise level is large or the angular range small. However, figure 3 further shows the limitations of TV when high noise and limited angles are combined. The only difference



**Figure 3.** Examples when TV reconstructions cannot recover the global structures of samples. When there is a large missing wedge ( $\frac{2}{3}$  of data unseen) and noise on the projections then reconstructions exhibit characteristic blurring in the vertical direction. This can also be seen in the extrapolated region of the sinograms as a loss of structure.

between the Shepp–Logan phantom data shown in figures 2 and 3 is that the former is clean data, in the image of the forward operator, whilst the latter has Gaussian white noise added. We see that as soon as there is a combined denoising/inpainting reconstruction problem, the TV prior on  $u$  becomes insufficient to recover the structure of the sample.

Recently, these traditional methods have received a revival through machine learning methods, see for instance [20, 21]. In both of these examples the main artefact reduction is a learned denoising step which only enforces prior knowledge on  $u$ .

The most common method that has been used to reconstruct pairs  $(u, v)$  is to solve each inverse problem sequentially. Typically, we can express the pipeline for such methods as:

$$\begin{aligned} v &= \text{optimal inpainted sinogram given } b \\ u &= \text{optimal reconstruction given } v. \end{aligned}$$

This has seen much success in heavy metal artefact reduction [22, 23] where a regularisation functional for the inpainting problem may be constructed from dictionary learning [24], fractional order TV [23], and Euler’s Elastica [25]. Euler’s Elastica has also been used in the limited angle problem [26] along with more customised interpolation methods [27]. These approaches allow us to use prior knowledge on the sinogram to calculate  $v$  and then spatial prior knowledge to calculate  $u$  from  $v$ ; at no point is the choice of  $v$  influenced by our spatial prior. A full joint approach allows us to go beyond this and use all of our prior knowledge to inform the choice of both  $u$  and  $v$ . If our prior knowledge is consistent with the true data then this extra utilisation of our prior must have the potential to improve the recovery of both  $u$  and  $v$ . To build a model for this framework we shall encode our In this paper, therefore, we propose a full joint approach which allows us to use all of our prior knowledge at once. To realise this idea we encode prior knowledge and consistency terms into a single energy functional such that an optimal pair of reconstructions will minimise this joint functional, which we shall write as:

$$E(u, v) = \alpha_1 d_1(\mathcal{R}u, v) + \alpha_2 d_2(S\mathcal{R}u, b) + \alpha_3 d_3(Sv, b) + \alpha_4 r_1(u) + \alpha_5 r_2(v) \quad (1.1)$$

where  $\alpha_i \geq 0$  are weighting parameters,  $d_i$  are appropriate distance functionals and  $r_i$  are regularisation functionals which encode our prior knowledge. Note that choice of  $d_2$  and  $d_3$  are dictated by the data noise model. In what follows,  $r_1$  is chosen to be the total variation.

Our choice for  $r_2$ , the sinogram regularisation, is based on theoretical and heuristic observations. Thirion [28] has shown that discontinuities in  $u$  correspond to sharp edges in the sinogram. In figure 3 we also see that blurred reconstructions correspond to loss of structure in the

sinogram. Therefore,  $r_2$  will be chosen to detect sharp features in the given data and preserve these through the inpainting region. The exact form of  $r_2$  will be formalised in section 3.

A typical advantage of joint models is that they generalise previous ones. For instance, if we let  $\alpha_2, \alpha_4 \rightarrow \infty$  then we recover the TV reconstruction model. Alternatively, if we let  $\alpha_3, \alpha_5 \rightarrow \infty$  then we recover a method which performs the inpainting and then the reconstruction sequentially, as in [23, 25, 26]. Recent work [29] has shown that such a joint approach can be advantageous in similar situations but closest to our approach is that of [30] where  $r_1$  and  $r_2$  were chosen to encode wavelet sparsity in both  $u$  and  $v$ . We shall demonstrate, as seen in figure 4, that the flexibility of our joint model, (1.1), can allow for a better adaptive fitting to the data.

### 1.3. Overview and contributions

The main contribution of this work is to provide a framework for building models of the form described in (1.1) and provide new proofs for a numerical scheme for minimising these functionals. This numerical scheme is valid for a very large class of non-smooth and non-convex functionals  $r_i$  and thus could be used in many other applications.

Section 2 first outlines the necessary concepts and notation needed to formalise the statement of our specific joint model in section 3. It will become apparent that the main numerical requirements of this work will require minimising a functional which is neither convex or smooth. Section 4 will start by reviewing recent work from Ochs *et al* [31] and we then provide alternative concise and self-contained proofs. Our main contribution here is to extend the existing results to an alternating (block) descent scenario. Finally, we present numerical results including two synthetic phantoms and experimental electron microscopy data where the limited angle situation occurs naturally.

## 2. Preliminaries

### 2.1. The x-ray transform

The principal forward model for x-ray tomography is provided by the x-ray transform which can be defined for any bounded domain,  $\Omega \subset \mathbb{R}^n$ , by

$$\mathcal{R}: L^1(\Omega, \mathbb{R}) \rightarrow L^1(\mathcal{S}^{n-1} \times \mathbb{R}^n, \mathbb{R}) \text{ such that } \mathcal{R}u(\theta, y) = \int_{x=y+t\theta, t \in \mathbb{R}} u(x) dt \quad (2.1)$$

where  $\mathcal{S}^{n-1} = \{s \in \mathbb{R}^n \text{ s.t. } |s| = 1\}$ . In this work, for simplicity, we will only be using  $n = 2$  although the case  $n = 3$  is completely analogous.

### 2.2. Total variation regularisation

Total variation (TV) regularisation is extremely common across many fields of image processing [32–34]. The definition of the (isotropic) TV semi-norm on domains  $\Omega \subset \mathbb{R}^n$  is stated as

$$g \in L^1(\Omega, \mathbb{R}) \implies \text{TV}(g) = \sup \left\{ \int \langle g(x), \text{div}(\varphi)(x) \rangle dx \right. \\ \left. \text{s.t. } \varphi \in C_c^\infty(\Omega, \mathbb{R}^n), |\varphi(x)|_2 \leq 1 \text{ for all } x \right\}. \quad (2.2)$$

The intuition behind this is that when  $g \in W^{1,1}(\Omega, \mathbb{R})$  then we have exactly

$$\text{TV}(g) = \|\nabla g\|_{2,1} = \int |\nabla g(x)|_2 dx.$$

From this point forward we shall write the  $\|\cdot\|_{2,1}$  representation where  $|\nabla g|$  is to be understood as a measure if necessary. We shall denote the space of bounded variation as

$$\text{BV}(\Omega, \mathbb{R}) = \{g \in L^1 \text{ s.t. } \text{TV}(g) < \infty\}.$$

One property that we shall use about the space of bounded variation is  $\text{BV} \subset L^2 \subset L^1$  which holds whenever  $\Omega$  is compact by the Poincaré inequality:  $\|g - \int g / \int 1\|_2 \lesssim \text{TV}(g)$ .

The most common way to enforce this prior in reconstruction or inpainting problems is generalised Tikhonov regularisation, which gives us the basic reconstruction method [2, 18].

$$u = \underset{u \geq 0}{\text{argmin}} \frac{1}{2} \|SRu - b\|_2^2 + \lambda \text{TV}(u) \text{ for some } \lambda \geq 0. \quad (2.3)$$

The parameter  $\lambda$  is a regularisation parameter, which allows to enforce more or less regularity, depending on the quality of the data  $b$ .

### 2.3. Directional total variation regularisation

For our sinogram regularisation functional we shall use a directionally weighted TV penalty, motivated by the TV diffusion model developed by Joachim Weickert [35] for various imaging techniques including denoising, inpainting and compression. Such an approach has already shown great ability for enhancing edges in noisy or blurred images, and preserves line structures across inpainting regions [36–38]. The heuristic for our regularisation on the sinogram was described in figure 3 and we shall encode it in an anisotropic TV penalty which shall be written as

$$\text{DTV}(v) = \int |A(x) \nabla v(x)| dx = \|A \nabla v\|_{2,1} \text{ for some anisotropic } A: \mathbb{R}^2 \rightarrow \mathbb{R}^{2 \times 2}.$$

The power of such a weighted extension of TV is that once a line is detected, either known beforehand or detected adaptively, we can embed this in  $A$  and enhance or sharpen that line structure in the final result. The general form that we choose for  $A$  is

$$\begin{aligned} A(x) &= c_1(x) \mathbf{e}_1(x) \mathbf{e}_1(x)^T + c_2(x) \mathbf{e}_2(x) \mathbf{e}_2(x)^T \\ \text{such that } \mathbf{e}_i: \mathbb{R}^2 &\rightarrow \mathbb{R}^2, |\mathbf{e}_i(x)| = 1, \langle \mathbf{e}_1(x), \mathbf{e}_2(x) \rangle = 0 \end{aligned} \quad (2.4)$$

i.e.

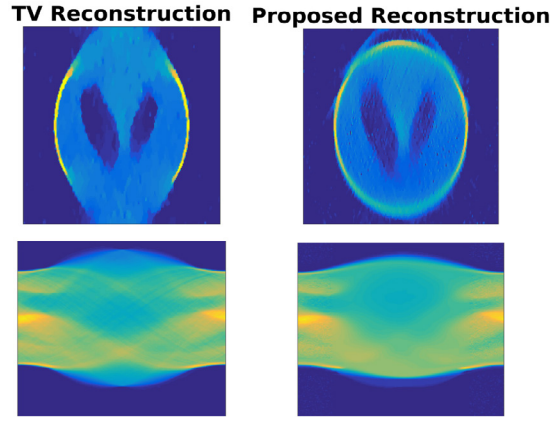
$$\text{DTV}(v) = \int \sqrt{c_1^2 |\langle \mathbf{e}_1, \nabla v \rangle|^2 + c_2^2 |\langle \mathbf{e}_2, \nabla v \rangle|^2} dx.$$

Examples of this are presented in figure 5. Note that the choice  $c_1 = c_2$  recovers the traditional TV regulariser but for  $|c_1| \ll c_2$  we allow for much larger (sparse) gradients in the direction of  $\mathbf{e}_1$ . This allows for large jumps in the direction of  $\mathbf{e}_1$  whilst maintaining flatness in the direction of  $\mathbf{e}_2$ . In order to generate these parameters we follow the construction of Weickert [35]. Given a noisy image,  $d$ , we can construct the structure tensor:

$$(\nabla d_\rho \nabla d_\rho^T)_\sigma(x) = \lambda_1(x) \mathbf{e}_1(x) \mathbf{e}_1(x)^T + \lambda_2(x) \mathbf{e}_2(x) \mathbf{e}_2(x)^T \text{ such that } \lambda_1(x) \geq \lambda_2(x) \geq 0$$

where





**Figure 4.** Demonstration of the improvement which can be achieved by using a model as in (1.1). Left hand images show state of the art reconstructions using total variational regularisation ( $\alpha_1 = \alpha_3 = \alpha_5 = 0$ ). This reconstruction clearly shows the characteristic blurring artefacts at the top and bottom. In our proposed joint reconstruction (right hand) we can minimise these effects.

$$d_\rho(x) = [d \star \exp(-|\cdot|^2/2\rho^2)](x) = \int \exp(-|y-x|^2/2\rho^2) d(y) \text{ etc.}$$

denotes convolution with the heat kernel. This eigenvalue decomposition is typically very informative in constructing  $A$ . If we define

$$\Delta(x) = \lambda_1(x) - \lambda_2(x) \text{ is coherence} \quad \Sigma(x) = \lambda_1(x) + \lambda_2(x) \text{ is energy}$$

then the eigenvectors give the alignment of edges and  $\Delta, \Sigma$  characterise the local behaviour, as in figure 6. In particular, we simplify the model to

$$A_d(x) := c_1(x|\Delta, \Sigma) \mathbf{e}_1(x) \mathbf{e}_1(x)^T + c_2(x|\Delta, \Sigma) \mathbf{e}_2(x) \mathbf{e}_2(x)^T \quad (2.5)$$

where the only parameters left to choose are  $c_i$ . Typical examples of include

$$c_1 = \frac{1}{\sqrt{1 + \Sigma^2}}, \quad c_2 = 1$$

$$c_1 = \varepsilon, \quad c_2 = \varepsilon + \exp(-1/\Delta^2) \text{ for some } \varepsilon > 0.$$

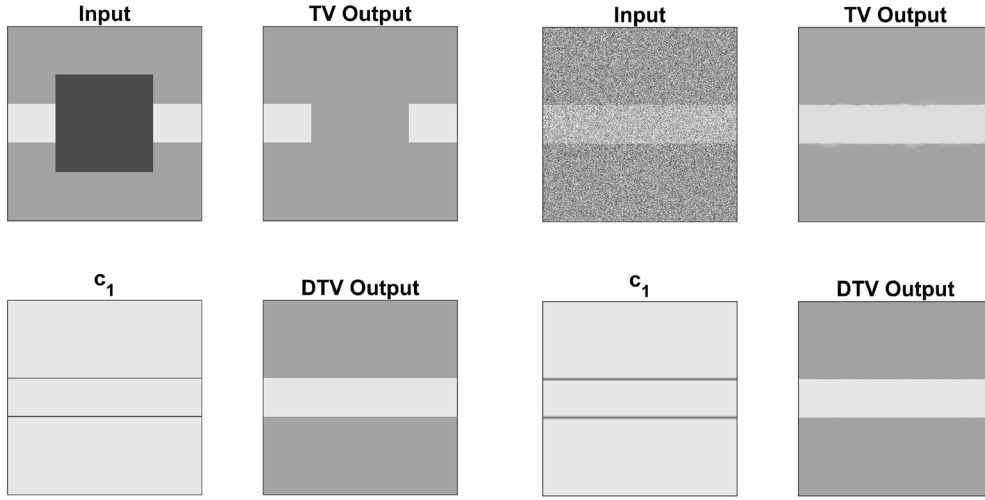
The key idea here is that  $c_1 \ll c_2$  near edges and  $c_1 = c_2$  on flat regions. In practice  $d$  will also be an optimisation parameter and so we shall require a regularity result on our choice of  $d \mapsto A_d$ , now characterised uniquely by our choice of  $c_i$ .

**Theorem 2.1.** *If*

- (i)  $c_i$  are  $2k$  times continuously differentiable in  $\Delta$  and  $\Sigma$ ,  $k \geq 1$
- (ii)  $c_1(x|0, \Sigma) = c_2(x|0, \Sigma)$  for all  $x$  and  $\Sigma \geq 0$
- (iii)  $\partial_\Delta^{2j-1} c_1(x|0, \Sigma) = \partial_\Delta^{2j-1} c_2(x|0, \Sigma) = 0$  for all  $x$  and  $\Sigma \geq 0, j = 1 \dots, k$

Then  $A_d$  is  $C^{2k-1}$  with respect to  $d$  for all  $\rho > 0, \sigma \geq 0$ .





**Figure 5.** Examples comparing TV with directional TV for inpainting and denoising. Both examples have the same edge structure and so parameters in (2.4) are the same in both. DTV uses  $c_2 = 1$  and  $c_1$  as the indicator (0 or 1) shown in the bottom left plot, TV is the case  $c_1 = c_2 = 1$ . Left block: top left image is inpainting input where the dark square shows the inpainting region. The structure of  $c_1$  allows DTV (bottom right) to connect lines over arbitrary distances, whereas TV inpainting (top right) will fail to connect the lines if the horizontal distance is greater than the vertical separation of the edges. Right block: top left image is denoising input. DTV has two advantages. Firstly, the structure of  $c_1$  recovers a much straighter line than that in the TV reconstruction. Secondly, TV penalises jumps equally in each direction and so the contrast is reduced, DTV is able to penalise noise oscillations independently from edge discontinuities which allows us to maintain much higher contrast.

### Remark 2.2.

- Property (ii) is necessary for  $A_d$  to be well defined and continuous for all fixed  $d$
- If we can write  $c_i = c_i(\Delta^2, \Sigma)$  then property (iii) holds trivially

The proof of this theorem is contained in appendix A.

### 3. The joint model

Now that all of the notation and concepts have been defined we can formalise the statement of our particular joint model of the form in (1.1):

- The forward operator,  $\mathcal{R}$ , is the x-ray transform from (2.1)
- The desired reconstructed sample,  $u \in \text{BV}(\Omega, \mathbb{R})$  on some domain  $\Omega$
- The noisy sub-sampled data,  $b \in L^1(\Omega', \mathbb{R})$  on some  $\Omega' \subseteq \mathcal{S}^1 \times \mathbb{R}_{\geq 0}$ . We extend such that  $b|_{\Omega^c} = 0$  for notational convenience.
- The full reconstructed sinogram,  $v \in L^1(\mathcal{S}^1 \times \mathbb{R}_{\geq 0}, \mathbb{R})$

We also define  $S = S_{\Omega'}$  to be the indicator function on  $\Omega'$ . The joint model that we thus propose is

$$\begin{aligned}
(u, v) = \operatorname{argmin}_{u \geq 0} E(u, v) = \operatorname{argmin}_{u \geq 0} & \frac{1}{2} \|\mathcal{R}u - v\|_{\alpha_1}^2 + \frac{\alpha_2}{2} \|S\mathcal{R}u - b\|_2^2 \\
& + \frac{\alpha_3}{2} \|Sv - b\|_2^2 + \beta_1 \operatorname{TV}(u) + \beta_2 \operatorname{DTV}_{\mathcal{R}u}(v)
\end{aligned} \quad (3.1)$$

where

$$\operatorname{DTV}_{\mathcal{R}u}(v) = \|A_{\mathcal{R}u} \nabla v\|_{2,1}$$

and  $\alpha_i, \beta_i > 0$  are weighting parameters,  $A_{\mathcal{R}u}$  as defined in (2.5).  $\alpha_1$  is embedded in the norm as it is a spatially varying weight, taking different values inside and outside of  $\Omega'$ . We note that the norms involving  $b$  are determined by the noise model of the acquisition process, in this case Gaussian noise. The final metric pairing  $\mathcal{R}u$  and  $v$  was free to be chosen to promote any structural similarity between the two quantities. We have chosen the squared  $L^2$  norm for simplicity although if some structure is known to be important then there is a wide choice of specialised functions from which to choose (e.g. [1]).

The choice of regularisation functionals reflects prior assumptions on the expected type of sample; all of the examples shown later will follow these assumptions. The isotropic TV penalty is chosen as  $u$  is expected to be piece-wise constant. This will reduce oscillations from  $u$  and favour ‘stair-case’ like changes of intensity over smooth ones. The assumptions of our regularisation on  $v$  must also be derived from expected properties of  $u$ . What is known from [28], and can be seen in figure 3, is that discontinuities of  $u$  along curves in the spatial domain, say  $\gamma$ , generate a so called ‘dual curve’ in the sinogram.  $\mathcal{R}u$  will also have an edge, although possibly not a discontinuity, along this dual curve. Thus, perpendicular to the dual curve  $v$  should have sharp features and parallel to the dual curve intensity should vary slowly. The assumption of our regularisation is that if a dual curve is present in the visible component of the data then it should correspond to some  $\gamma$  in the spatial domain. The extrapolation of this dual curve must behave like the boundary of a level-set of  $u$ , i.e. preserve the sharp edge and slow varying intensities in  $v$ . The main influence of this regularisation is in the inpainting region and so any artefacts it introduces should also only effect edges corresponding to these invisible singularities, including streaking artefacts. Another bias that is present is an assumption that dual curves are themselves smooth. In the inpainting region, this will encourage dual curves with low curvature thus invisible singularities are likely to follow near-circular arcs in the spatial domain. Final parameter choices, such as  $\alpha_i, \beta_i$  and  $c_i$ , are not necessary at this point and will be chosen in section 5.1.

The immediate question to ask is whether this model is well posed. For a non-convex function we typically cannot expect numerically to find global minimisers but the following result shows we can expect some convergence to local minima. We present the following result which justifies looking for minima of this functional.

**Theorem 3.1.** *If*

- $c_i$  are bounded away from 0
- $\rho > 0$
- $A_d$  is differentiable in  $d$

*then sub-level sets of  $E$  are weakly compact in  $L^2(\Omega, \mathbb{R}) \times L^2(\mathbb{R}^2, \mathbb{R})$  and  $E$  is weakly lower semi-continuous. i.e. for all  $(u_n, v_n) \in L^2(\Omega, \mathbb{R}) \times L^2(\mathbb{R}^2, \mathbb{R})$ :*

$$E(u_n, v_n) \text{ uniformly bounded implies a subsequence converges weakly}$$

**Table 1.** Parameter choices for numerical experiments. Each algorithm was run for 300 iterations.

	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\beta_1$	$\beta_2$	$\beta_3$	$\rho$	$\sigma$
Concentric rings phantom	$\frac{1}{2^2} \mathbb{1}_{\Omega^c}$	1	$1 \times 10^{-1}$	$3 \times 10^{-5}$	$3 \times 10^3$	$10^{10}$	1	8
Shepp–Logan phantom	$\frac{1}{4^2} \mathbb{1}_{\Omega^c}$	1	$3 \times 10^{-1}$	$3 \times 10^{-5}$	$3 \times 10^2$	$10^{10}$	1	8
Experimental dataset (both sampling ratios)	$\frac{1}{2^2} \mathbb{1}_{\Omega^c}$	1	$3 \times 10^2$	$1 \times 10^{-5}$	$3 \times 10^1$	$10^6$	1	0

$$\liminf E(u_n, v_n) \geq E(u, v) \text{ whenever } u_n \rightharpoonup u, v_n \rightharpoonup v.$$

The proof of this theorem is contained in appendix B. This theorem justifies numerical minimisation of  $E$  because it tells us that all descending sequences  $(E(u_n, v_n) \leq E(u_{n-1}, v_{n-1}))$  have a convergent subsequence and any limit point must minimise  $E$  over the original sequence.

#### 4. Numerical solution

To address the issue of convergence we shall first generalise our functional and prove the result in the general setting. Functionals will be of the form  $E: X \times Y \rightarrow \mathbb{R}$  where  $X$  and  $Y$  are Banach spaces and  $E$  accepts the decomposition

$$E(x, y) = f(x, y) + g(J(x, y))$$

such that:

$$\text{Sub-level sets of } E \text{ are weakly compact} \quad (4.1)$$

$$f: X \times Y \rightarrow \mathbb{R} \text{ is jointly convex in } (x, y) \text{ and bounded below} \quad (4.2)$$

$$\begin{aligned} g: Z \rightarrow \mathbb{R} \text{ is a semi-norm on Banach space } Z, \text{ i.e. for all } t \in \mathbb{R}, z, z_1, z_2 \in Z \\ g(z) \geq 0, g(tz) = |t|g(z) \text{ and } g(z_1 + z_2) \leq g(z_1) + g(z_2) \end{aligned} \quad (4.3)$$

$$\begin{aligned} J: X \times Y \rightarrow Z \text{ is } C^1 \text{ and for all } K \subseteq X \times Y, \exists L_x, L_y < \infty \text{ such that } \forall (x, y) \in K \\ g(J(x + dx, y) - J(x, y) - \nabla_x J(x, y)dx) \leq L_x \|dx\|_X^2 \end{aligned} \quad (4.4)$$

$$g(J(x, y + dy) - J(x, y) - \nabla_y J(x, y)dy) \leq L_y \|dy\|_Y^2. \quad (4.5)$$

Note if  $g$  is a norm then  $L_x$  can be chosen to be the standard Lipschitz factor of  $\nabla_x J$ . If  $J$  is twice Fréchet-differentiable then these constants must be finite. In our case:

$$\begin{aligned} f(x, y) &= \frac{1}{2} \|Rx - y\|_{\alpha_1}^2 + \frac{\alpha_2}{2} \|\mathcal{S}Rx - b\|^2 \\ &\quad + \frac{\alpha_3}{2} \|Sy - b\|^2 + \beta_1 \text{TV}(x) + \begin{cases} 0 & x \geq 0 \\ \infty & \text{else} \end{cases} \\ g(z) &= \beta_2 \|z\|_{2,1} \\ J(x, y) = A_{Rx} \nabla y &\implies \tau_x \sim \beta_2 \|\nabla^2 A\| \|R\| \text{TV}(y), \tau_y = 0. \end{aligned}$$

Finiteness of  $\|\nabla^2 A\|$  and weak compactness of sub-level sets are given by theorems 2.1 and 3.1 respectively. The alternating descent algorithm is stated in algorithm 1. Classical alternating proximal descent would take  $x_{n+1} = \operatorname{argmin}_{x \in X} E(x, y_n) + \tau_x \|x - x_n\|_2^2$  but because of the complexity of  $A_{\mathcal{R}_d}$  each sub-problem would have the same complexity as the full problem, making it computationally infeasible. By linearising  $A_d$  we overcome this problem as both sub-problems are convex, for which there are many standard solvers such as [39, 40]. This second approach is similar to that of the ProxDescent algorithm [31, 41]. We extend this algorithm to cover alternating descent and achieve equivalent convergence guarantees. Using algorithm 1, our statement of convergence is the following theorem.

---

**Algorithm 1.**

---

**Input:** Any  $x_0 \in X$ ,  $\tau_x, \tau_y \geq 0$ .

$n \leftarrow 0$

**while** not converged **do**

$n \leftarrow n + 1$

$$x_n = \operatorname{argmin}_{x \in X} f(x, y_{n-1}) + \tau_x \|x - x_{n-1}\|_X^2 + g(J(x_{n-1}, y_{n-1}) + \nabla_x J(x_{n-1}, y_{n-1})(x - x_{n-1})) \quad (4.6)$$

$$y_n = \operatorname{argmin}_{y \in Y} f(x_n, y) + \tau_y \|y - y_{n-1}\|_Y^2 + g(J(x_n, y_{n-1}) + \nabla_y J(x_n, y_{n-1})(y - y_{n-1})) \quad (4.7)$$

**end while**

**Output:**  $(x_n, y_n)$

---

**Theorem 4.1.** *Convergence of alternating minimisation: if  $E$  satisfies (4.1)–(4.5) and  $(x_n, y_n)$  are a sequence generated by algorithm 1 then*

- $E(x_{n+1}, y_{n+1}) \leq E(x_n, y_n)$  for each  $n$ .
- A subsequence of  $(x_n, y_n)$  must converge weakly in  $X \times Y$ .
- If  $\{(x_n, y_n) \text{ s.t. } n = 1, \dots\}$  is contained in a finite dimensional space then every limit point of  $(x_n, y_n)$  must be a critical point (as will be defined in definition 4.4) of  $E$  in both the direction of  $x$  and  $y$ .

This result is the parallel of lemma 10, theorem 18 and corollary 21 in [31] without the alternating or block descent setting. There is some overlap in the analysis for the two settings although we present an independent proof which is more direct and we feel gives more intuition for our more restricted class of functionals. The rest of this section is now dedicated to the proof of this theorem.

For notational convenience we shall compress notation such that:

$$f_{n,m} = f(x_n, y_m), \quad J_{n,m} = J(x_n, y_m), \quad E_{n,m} = E(x_n, y_m) \text{ etc.}$$

#### 4.1. Sketch proof

The proof of theorem 4.1 will be a consequence of two lemmas.

- In lemma 4.3 we show for  $\tau_x, \tau_y$  (algorithm 1) sufficiently large, the sequence  $E_{n,n}$  is monotonically decreasing and sequences  $\|x_n - x_{n-1}\|_X, \|y_n - y_{n-1}\|_Y$  converge to 0. This follows by a relatively standard sufficient decrease argument as seen in [31, 42, 43].
- In lemma 4.6 we first define a critical point for functions which are non-convex and non-differentiable. If a subsequence of our iterates converges in norm then the limit must

be a critical point in each of the two axes. Note that it is very difficult to expect more than this in such a general setting, for instance example 4.2 shows a uniformly convex energy which shows this to be sharp. The common technique for overcoming this is assuming differentiability in the terms including both  $x$  and  $y$  [42, 44, 45]. These previous results and algorithms are not available to us as we allow non-convex terms which are also non-differentiable.

**Example 4.2.** Define  $E(x, y) = \max(x, y) + x^2 + y^2$  for  $x, y \in \mathbb{R}$ . This is clearly jointly convex in  $(x, y)$  and thus a simple case of functions considered in theorem 4.1. Suppose  $(x_0, y_0) = (0, 0)$  then

$$\begin{aligned} x_1 &= \operatorname{argmin} E(x, y_0) + \tau_x(x - x_0)^2 = 0 \\ y_1 &= \operatorname{argmin} E(x_1, y) + \tau_y(y - y_0)^2 = 0. \end{aligned}$$

Hence  $(0, 0)$  is a limit point of the algorithm but it is not a critical point,  $E$  is uniformly convex and so it has only one critical point,  $(-\frac{1}{2}, -\frac{1}{2})$ .

#### 4.2. Sufficient decrease lemma

In the following we prove the monotone decrease property of our energy functional between iterations.

**Lemma 4.3.** *If for each  $n$*

$$\tau_x \geq L_x + \tau_X, \quad \tau_y \geq L_y + \tau_Y$$

*for some  $\tau_X, \tau_Y \geq 0$  then*

$$\sum_{n=0}^{\infty} \tau_X \|x_n - x_{n-1}\|_X^2 + \tau_Y \|y_n - y_{n-1}\|_Y^2 \leq E(x_0, y_0)$$

*and*

$$E(x_{n+1}, y_{n+1}) \leq E(x_n, y_n) \text{ for all } n.$$

**Proof.** Note by equations (4.6) and (4.7) (definition of our sequence) we have

$$f_{n+1,n} + g(J_{n,n} + \nabla_x J_{n,n}(x_{n+1} - x_n)) + \tau_x \|x_{n+1} - x_n\|_X^2 \leq E_{n,n} \quad (4.8)$$

$$f_{n+1,n+1} + g(J_{n+1,n} + \nabla_y J_{n+1,n}(y_{n+1} - y_n)) + \tau_y \|y_{n+1} - y_n\|_Y^2 \leq E_{n+1,n}. \quad (4.9)$$

Further, by application of the triangle inequality for  $g$  and the mean value theorem we have

$$\begin{aligned} g(J_{n+1,n}) - g(J_{n,n} + \nabla_x J_{n,n}(x_{n+1} - x_n)) + \tau_x \|x_{n+1} - x_n\|_X^2 \\ \leq g(J_{n+1,n} - J_{n,n} - \nabla_x J_{n,n}(x_{n+1} - x_n)) + \tau_x \|x_{n+1} - x_n\|_X^2 \\ = g([\nabla_x J(\xi) - \nabla_x J_{n,n}](x_{n+1} - x_n)) + \tau_x \|x_{n+1} - x_n\|_X^2 \\ \leq \operatorname{Lip}_{X,g}(\nabla_x J(\cdot, y_n)) \|x_{n+1} - x_n\|_X^2 + \tau_x \|x_{n+1} - x_n\|_X^2 \\ \leq \tau_x \|x_{n+1} - x_n\|_X^2. \end{aligned} \quad (4.10)$$

By equivalent argument,

$$g(J_{n+1,n+1}) - g(J_{n,n+1} + \nabla_y J_{n+1,n}(y_{n+1} - y_n)) + \tau_Y \|y_{n+1} - y_n\|_Y^2 \leq \tau_Y \|y_{n+1} - y_n\|_Y^2. \quad (4.11)$$

Summing equations (4.8)–(4.11) gives

$$E_{n+1,n+1} + \tau_X \|x_{n+1} - x_n\|_X^2 + \tau_Y \|y_{n+1} - y_n\|_Y^2 \leq E_{n,n}.$$

This immediately gives the monotone decrease property of  $E_{n,n}$ . If we also sum this over  $n$  then we achieve the final statement of the theorem:

$$\sum_{n=1}^{\infty} \tau_X \|x_{n+1} - x_n\|_X^2 + \tau_Y \|y_{n+1} - y_n\|_Y^2 \leq E_{0,0} - \lim E_{n,n} \leq E_{0,0} \quad \square$$

#### 4.3. Convergence to critical points

First we follow the work by Drusvyatskiy *et al* [46] we shall define criticality in terms of the slope of a function.

**Definition 4.4.** We shall say that  $x_*$  is a critical point of  $F: X \rightarrow \mathbb{R}$  if

$$|\partial F(x_*)| = 0$$

where we define the slope of  $F$  at  $x_*$  to be

$$|\partial F(x_*)| = \limsup_{\|dx\| \rightarrow 0} \frac{\max(0, F(x_*) - F(x_* + dx))}{\|dx\|}.$$

The first point to note is that this definition generalises the concept of a first order critical point for both smooth functions and convex functions (in terms of the convex sub-differential). In particular

$$F \in C^1 \implies |\partial F(x_*)| = \max \left( 0, \sup_{\|dx\|=1} -\langle \nabla F(x_*), dx \rangle \right) = \|\nabla F(x_*)\|$$

$$\text{Hence } |\partial F(x_*)| = 0 \iff \|\nabla F(x_*)\| = 0 \iff \nabla F(x_*) = 0$$

$F$  convex, hence  $x^* \in \operatorname{argmin} F \iff F(x_*) \leq F(x_* + dx)$  for all  $dx$

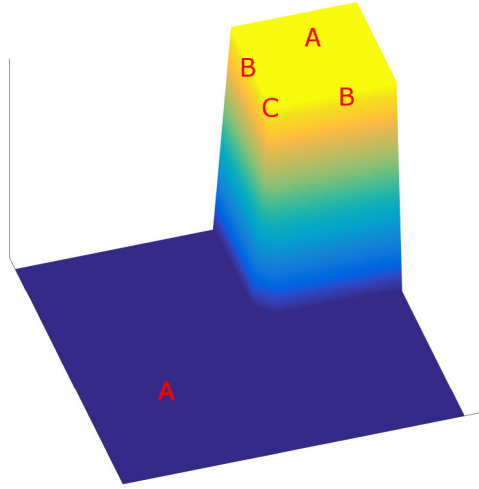
$$\text{Hence } |\partial F(x_*)| = 0 \iff \forall dx, 0 \geq \frac{F(x_*) - F(x_* + dx)}{\|dx\|} \iff x_* \in \operatorname{argmin} F.$$

For a differentiable function we cannot tell whether a critical point is a local minimum, maximum or saddle point. In general, this is also true for definition 4.4, however, at points of non-differentiability there is a bias towards local minima. This can be seen in the following example.

**Example 4.5.** Consider  $F(x) = -\|x\|$

$$|\partial F(0)| = \limsup_{\|dx\| \rightarrow 0} \max \left( 0, \frac{0 - \|0 + dx\|}{\|dx\|} \right) = 1$$

Hence, 0 is not a critical point of  $F$ . This bias is due to the  $\limsup$  in the definition which



**Figure 6.** Surface representing a characteristic image,  $d$ , to demonstrate the behaviour of  $\Sigma$  and  $\Delta$ . Away from edges (A) we have  $\Sigma \approx \Delta \approx 0$ . On simple edges (B) we have  $\Sigma \approx \Delta \gg 0$  and, finally, at corners (C) we have  $\Sigma \gg \Delta$ .

detects the strictly negative directional derivatives. This does not affect smooth functions as directional derivatives must vanish continuously to 0 about a critical point.

Some more examples are shown in figure 7. Now we shall show that our iterative sequence converges to a critical point in this sense.

**Lemma 4.6.** *If  $(x_n, y_n)$  are as given by algorithm 1 and  $X, Y$  are finite dimensional spaces then every limit point of  $(x_n, y_n)$ , e.g.  $(x_*, y_*)$ , is a critical point of  $E$  in each coordinate direction. i.e.*

$$|\partial_x E(x_*, y_*)| = |\partial_y E(x_*, y_*)| = 0.$$

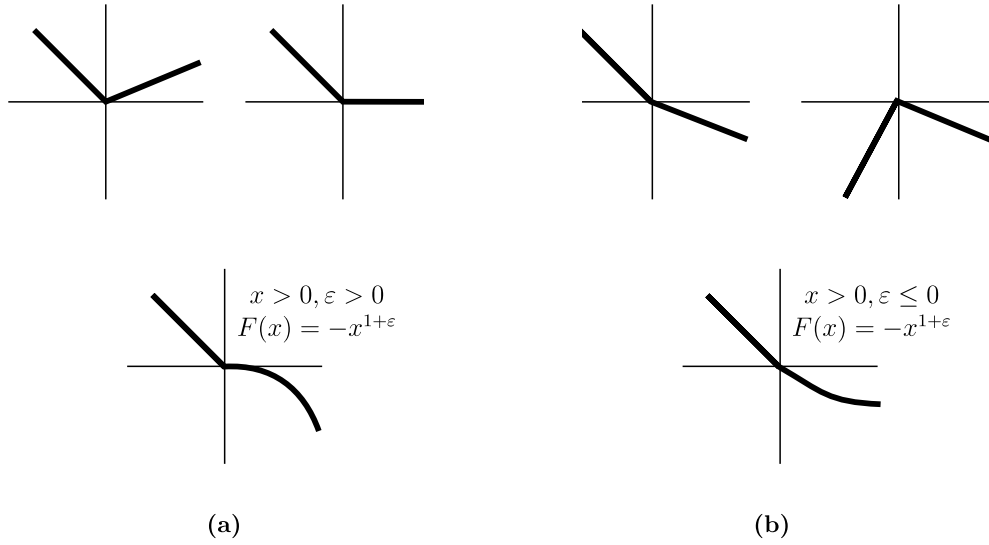
**Remark 4.7.**

- Finite dimensionality of  $X$  and  $Y$  accounts for what is referred to as ‘assumption 3’ in [31] and is some minimal condition which ensures that the limit is also a stationary point of our iteration (equations (4.6) and (4.7)).
- The condition for finite dimensionality, as we shall see, does not directly relate to the non-convexity. The difficulty of showing convergence to critical points in infinite dimensions is common across both convex [39] and non-convex [31, 42] optimisation.

**Proof.** First we recall that, in finite dimensional spaces, convex functions are continuous on the relative interior of their domain [47]. Also note that by our choice of  $\tau_x$  in lemma 4.6, for all  $x, x', y$  we have

$$\begin{aligned} & |g(J(x, y) + \nabla_x J(x, y)(x' - x)) - g(J(x', y))| \\ & \leq |g([J(x, y) - J(x', y)] - \nabla_x J(x, y)(x' - x))| \\ & = |g(\nabla_x J(x, y)(x' - x) - \nabla_x J(x, y)(x' - x))| \\ & \leq \tau_x \|\xi - x\|_X \|x' - x\|_X \\ & \leq \tau_x \|x' - x\|_X^2 \end{aligned}$$





**Figure 7.** Examples of 1D functions where 0 is/is not a critical point by definition 4.4. If a function is piece-wise linear then 0 is a critical point iff each directional derivative is non-negative. If the function can be approximated on an interval of  $x > 0$  to first order terms by  $F(x) = cx^{1+\varepsilon}$  then criticality can be characterised sharply. If  $c \geq 0$  then 0 is always a critical point. If  $c < 0$  then 0 is a critical point iff  $\varepsilon > 0$ , however, 0 is never a local minimum. (a) Examples of critical points and examples of non-critical points.

where existence of such  $\xi$  is given by the mean value theorem. Hence, for all  $x$  we have

$$\begin{aligned}
 E(x_{n+1}, y_n) &= f_{n+1,n} + g(J_{n+1,n}) \\
 &\leq f_{n+1,n} + g(J_{n,n} + \nabla_x J_{n,n}(x_{n+1} - x_n)) + \tau_x \|x_{n+1} - x_n\|_X^2 \\
 &\leq f(x, y_n) + g(J_{n,n} + \nabla_x J_{n,n}(x - x_n)) + \tau_x \|x - x_n\|_X^2 \\
 &\leq f(x, y_n) + g(J(x, y_n)) + 2\tau_x \|x - x_n\|_X^2 \\
 &= E(x, y_n) + 2\tau_x \|x - x_n\|_X^2
 \end{aligned}$$

where the first and third inequality are due to the condition shown above and the second is due to the definition of  $x_{n+1}$  in (4.6). Finally, by continuity of  $f$ ,  $J$  and  $g$  we can take limits on both sides of this inequality:

$$\implies E(x_*, y_*) \leq E(x, y_*) + 2\tau_x \|x - x_*\|_X^2 \text{ for all } x. \quad (4.12)$$

This completes the proof for  $x_*$  as

$$|\partial_x E(x_*, y_*)| = \limsup_{x \rightarrow x_*} \max \left( 0, \frac{E(x_*, y_*) - E(x, y_*)}{\|x_* - x\|_X} \right) \leq \limsup 2\tau_x \|x - x_*\|_X = 0.$$

The proof for  $y_*$  follows by symmetry. □

**Remark 4.8.**

- The important line in this proof, and where we need finite dimensionality, is being able to pass to the limit for (4.12). In the general case we can only expect to have  $(x_n, y_n) \rightharpoonup (x_*, y_*)$ , typically guaranteed by choice of regularisation functionals as in our theorem 3.1. In this reduced setting the left hand limit of (4.12) still remains valid,

$$E(x_*, y_*) \leq \liminf E(x_{n+1}, y_n) \text{ by weak lower semi-continuity.}$$

However, on the right hand side we require:

$$\lim E(x, y_n) + 2\tau_x \|x - x_n\|_X^2 \leq E(x, y_*) + 2\tau_x \|x - x_*\|_X^2$$

In particular, we already require  $\|x - x_n\|_X$  to be weakly upper semi-continuous. Topologically, this is the statement that weak and norm convergence are equivalent which will not be the case in most practical (infinite dimensional) examples.

- The properties we derive for  $(x_*, y_*)$  are actually slightly stronger than that of definition 4.4 which only depends on an infinitesimal ball about  $(x_*, y_*)$ . However, (4.12) gives us a quantification for the more global optimality of this point. This is seen in figure 8.

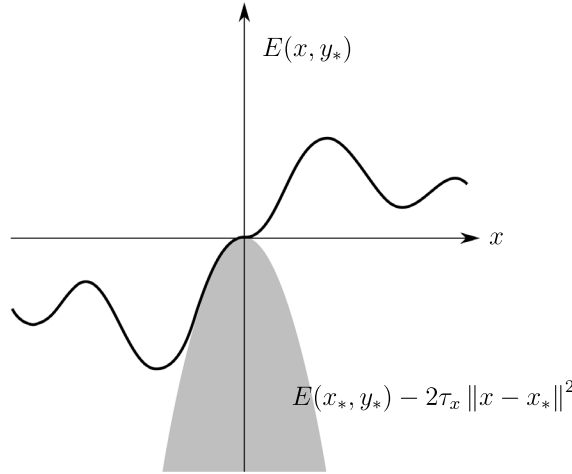
**4.4. Proof of theorem 4.1**

**Proof.** Fix arbitrary  $(x_0, y_0) \in X \times Y$  and  $\tau_X, \tau_Y \geq 0$ . Let  $(x_n, y_n)$  be defined as in algorithm 1. By lemma 4.3 we know that  $\{(x_n, y_n) \text{ s.t. } n \in \mathbb{N}\}$  is contained in a sub-level set of  $E$  which in turn must be weakly compact by (4.1). The assumption of lemma 4.6 is that we are in a finite dimensional space and so weak compactness is equivalent to norm compactness, i.e. some subsequence of  $(x_n, y_n)$  converges in norm. Also by lemma 4.6 we know that the limit point of this sequence must be an appropriate critical point.  $\square$

**5. Results**

For numerical results we present two synthetic examples and one experimental dataset from Transmission Electron Tomography. The two synthetic examples are discretised at a resolution of  $200 \times 200$  then simulated using the x-ray transform with a parallel beam geometry sampled at  $1^\circ$  intervals over a range of  $60^\circ$  resulting in a full sinogram of size  $287 \times 180$  and sub-sampled data at  $287 \times 60$ . Gaussian white noise (standard deviation 5% maximum signal) is then added to give the data. The experimental dataset was acquired with an annular dark field (parallel beam) Scanning TEM modality from which we have 46 projections spaced uniformly in  $3^\circ$  intervals over a range of  $135^\circ$ . Because of the geometry of the acquisition, we can treat the original 3D dataset as a stack of 2D and thus extract one of these slices as our example. This 2D dataset is then sub-sampled to 29 projections over  $87^\circ$ , reducing the size from  $173 \times 45$  to  $173 \times 29$ . This results in a reconstruction with  $u$  of size  $120 \times 120$  and  $v$  of size  $173 \times 180$ . A more detailed description of the acquisition and sample properties of the experimental dataset can be found in [48]. The code, and data, for all examples is available<sup>7</sup> under the Creative Commons Attribution (CC BY) license.

<sup>7</sup>[https://github.com/robtovey/2018\\_Directional\\_Inpainting\\_for\\_Limited\\_Angle](https://github.com/robtovey/2018_Directional_Inpainting_for_Limited_Angle)



**Figure 8.** Theorem 3.1 tells us that  $(x_*, y_*)$  is a local critical point but does not qualify the globality of the limit point. Equation (4.12) further allows us to quantify the idea that if a lower energy critical point exists then it must lie far from  $(x_*, y_*)$ . In particular, it must lie outside of the shaded cone given by the supporting quadratic.

### 5.1. Numerical details

All numerics are implemented in MATLAB 2016b. The sub-problem for  $u$  is solved with a PDHG algorithm [39] while the sub-problem for  $v$  is solved using the MOSEK solver via CVX [40, 49], the step size  $\tau$  is adaptively calculated. The initial point of our algorithm is always chosen to be a good TV reconstruction, i.e.

$$u_0 = \operatorname{argmin}_{u \geq 0} \frac{1}{2} \|S\mathcal{R}u - b\|_2^2 + \lambda \operatorname{TV}(u), \quad v_0 = \mathcal{R}u_0.$$

For clarity, we shall restate our full model with all of the parameters it includes. We seek to minimise the functional (3.1):

$$E(u, v) = \frac{1}{2} \|\mathcal{R}u - v\|_{\alpha_1}^2 + \frac{\alpha_2}{2} \|S\mathcal{R}u - b\|_2^2 + \frac{\alpha_3}{2} \|Sv - b\|_2^2 + \beta_1 \operatorname{TV}(u) + \beta_2 \|A_{Ru} \nabla v\|_{2,1}$$

$$A_d(x) = c_1(x|\lambda_1 - \lambda_2, \lambda_1 + \lambda_2) \mathbf{e}_1(x) \mathbf{e}_1(x)^T + c_2(x|\lambda_1 - \lambda_2, \lambda_1 + \lambda_2) \mathbf{e}_2(x) \mathbf{e}_2(x)^T$$

where  $(\nabla d_\rho \nabla d_\rho^T)_\sigma = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^T$  is a pointwise eigenvalue decomposition

$$c_1(x|\Delta, \Sigma) = 10^{-6} + \frac{\tanh(\Sigma(x))}{1 + \beta_3 \Delta(x)^2}, \quad c_2(x|\Delta, \Sigma) = 10^{-6} + \tanh(\Sigma(x)).$$

We chose these particular  $c_i$  according to two simple heuristics. If  $\Sigma$  is large (steep gradients) then it is likely a region with edges and so the regularisation should be largest but still bounded above. If  $\Delta = 0^+$  then there is a small or blurred ‘edge’ present and so we want to encourage it to become a sharp jump, i.e.  $\partial_\Delta c_1 < 0$ . Theorem 2.1 tells us that choosing  $c_i$  as functions of  $\Delta^2$  will guarantee accordance with our later convergence results; this leads to our natural choice above. The number of iterations for algorithm 1 was chosen to be 200 and 100 for the

synthetic and experimental datasets, respectively. To simplify the process of choosing values for the remaining hyper-parameters we made several observations:

- (i) The choice of  $\alpha_i$  and  $\beta_i$  appeared to be quite insensitive about the optimum. It is clear within 2–3 iterations whether values are of the correct order of magnitude. After this, values were only tuned coarsely. For example,  $\alpha_3$  and  $\beta_i$  are optimal within a factor of  $10^{\pm 1/2}$ .
- (ii) We can chose  $\alpha_2 = 1$  without any loss of generality. In which case, in general,  $\beta_1$  should be the same order of magnitude as when performing the TV reconstruction to get  $u_0, v_0$ .
- (iii)  $\alpha_2$  pairs  $u$  to the given data and  $\alpha_1$  pairs  $u$  to the inpainted data,  $v$ . As such,  $\alpha_1$  is spatially varying but should be something like a distance to the non-inpainting region. We chose the binary metric so that  $u$  is paired to  $v$  uniformly on the inpainting region and not at all outside.
- (iv) DTV specific parameters ( $\beta_2, \beta_3, \rho, \sigma$ ) can be chosen outside of the main reconstruction. These were chosen by solving the toy problem:

$$\operatorname{argmin} \frac{1}{2} \|v - v_0\|_2^2 + \beta_2 \|A_{v_0} \nabla v\|_{2,1}$$

which is a lot faster to solve.  $\rho > 0$  is required for the analysis and so this was fixed at 1.  $\sigma$  is a length-scale parameter which indicates the separation between distinct edges.  $\beta_3$  relies on the normalisation of the data. As can be seen in table 1, for the two synthetic examples, with same discretisation and scaling, these values are also consistent. The only value which changes is  $\beta_2$ , as expected, which weights how valid the DTV prior is for each dataset.

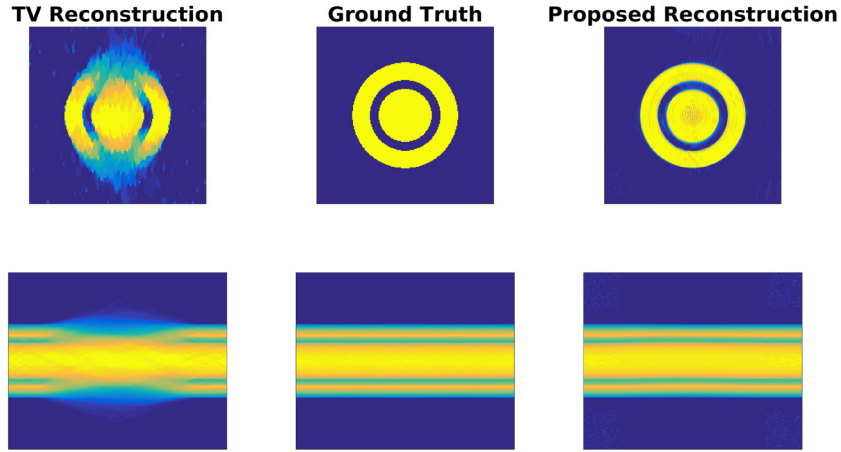
It is unclear whether a gridsearch may provide better results although, due to the number of parameters involved, this would definitely take a lot longer and mask some interpretability of the parameters. A further comparison of different choices of the main parameters can be found in the supplementary material.

### 5.2. Canonical synthetic dataset

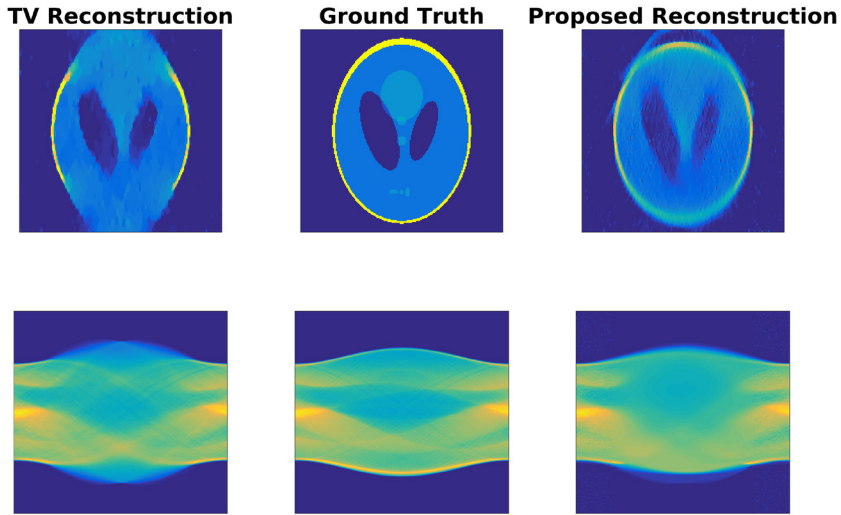
This example shows two concentric rings. This is the canonical example for our model because the exact sinogram is perfectly radially symmetric which should trivialise the directional inpainting procedure, even with noise present. As can clearly be seen in figure 9, the TV reconstruction is poor in the missing wedge direction which can be seen as a blurring out of the sinogram. By enforcing better structure in the sinogram, our proposed joint model is capable of extrapolating these local structures from the given data domain to recover the global structure and gives an accurate reconstruction.

### 5.3. Non-trivial synthetic dataset

This example shows the modified Shepp–Logan phantom which is built up as a sum of ellipses. This example has a much more complex geometry although the sinogram still has a clear geometry. In figure 10 we see that the largest scale feature, the shape of the largest ellipse, is recovered in our proposed reconstruction with minimal loss of contrast in the interior. One artefact we have not been able to remove is the two rays extending from the top of the reconstructed sample. Looking more closely we found that it was due to a small misalignment of the edge at the bottom of the sinogram as it crosses between the data to the inpainting region.

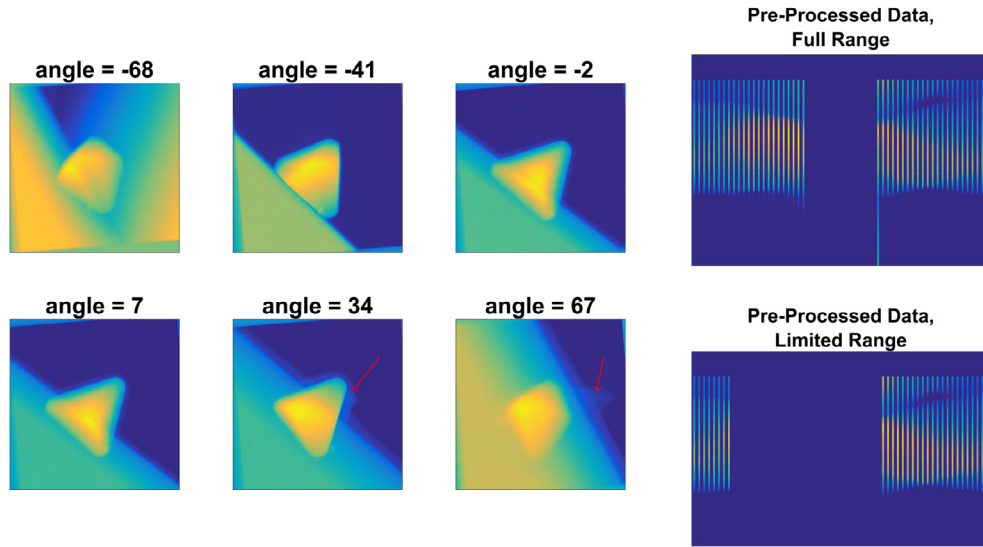


**Figure 9.** Canonical synthetic example. Top row shows the reconstructions,  $u$ , while the bottom row shows the reconstructed sinogram,  $v$ .



**Figure 10.** Non-trivial synthetic example of the modified Shepp–Logan phantom. Top row shows the reconstructions,  $u$ , while the bottom row shows the reconstructed sinogram,  $v$ . We regain the large-scale geometry of the shape without losing much of the interior features.

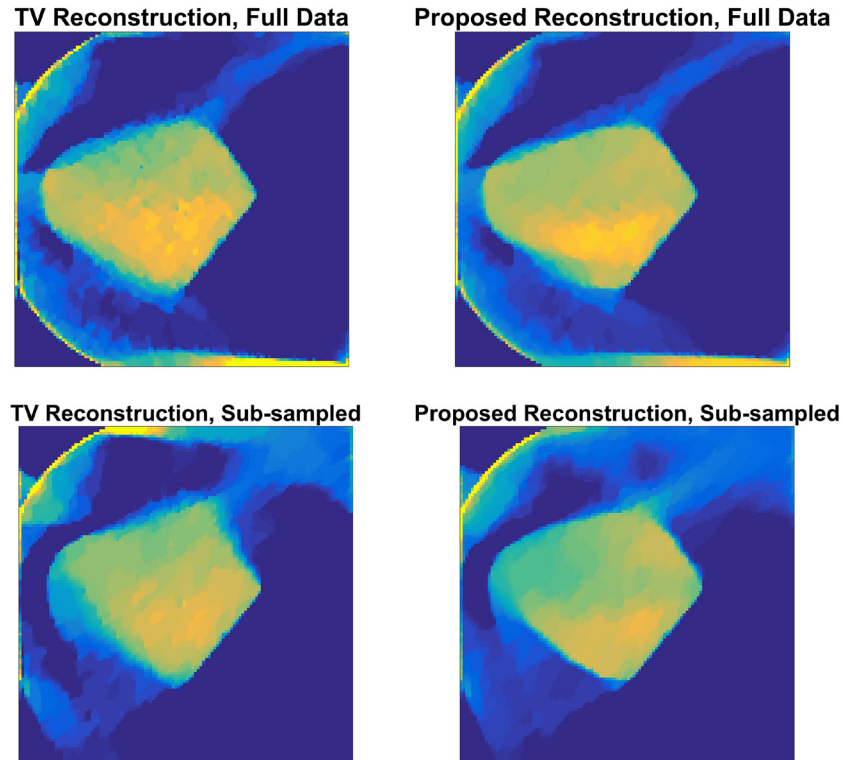
Numerically, this happens because of the convolutions which take place inside the directional TV regularisation functional. Having a non-zero blurring is essential for regularity of the regularisation (theorem 2.1) but the effect of this is that it does not heavily penalise misalignment on such a small scale. This means that at the interface between the fixed data-term there is a slight kink, the line is continuous but not  $C^1$ . The effect of this on the reconstruction is the two lines which extend from the sample at this point. Looking at quantitative measures, the PSNR value rises from 17.33 to 17.36 whereas the SSIM decreases from 0.76 to 0.62, from TV to the proposed reconstruction, respectively. These measures are inconclusive and the authors feel that they fail to balance the improvement to global geometry verses more local artefacts in the reconstructions.



**Figure 11.** Raw data for transmission electron microscopy example. Projections at large angles, e.g.  $-68^\circ$ , show the presence of the sample holder which violates the x-ray modelling assumption that outside of the region of interest is vacuum. If the violation is too extreme then this can cause strong artefacts in reconstructions and so the common action is to discard such data. The plane surface also violates this model but is relatively weak at low angles and so will cause weaker artefacts. A source of noise in this acquisition is that over time the surface becomes coated with carbon. This is first visible as a thin film at  $-2^\circ$  and progressively gets thicker through the remaining projections. At  $34^\circ$  we see a bump of carbon appear on the top right edge. After pre-processing, as shown top right artificially sub-sample to compare TV with our proposed reconstruction method.

#### 5.4. Experimental dataset

The sample is a silver bipyramidal crystal placed on a planar surface, and the challenges of this dataset are shown in figure 11. We immediately see that the wide angle projections have large artefacts which produces a very low signal to noise ratio. Another issue present is that there is mass seen in some of the projections which cannot be represented within the reconstruction volume. Both of these issues violate the simple x-ray model that is used. Exact modelling would require estimation of parameters which are not available *a priori* and so the preferred acquisition is one which automatically minimises these modelling errors. Another artefact is that over time each surface becomes coated with carbon. This is a necessary consequence of the sample preparation and this coating is known to occur during the microscopy. The result of modelling errors and time dependent noise is to prefer an acquisition with limited angular range and earliest acquired projections. Because of this, in numerical experiments we compare both TV and our proposed reconstruction using only  $\frac{3}{4}$  of the available data, 29 projections over an  $87^\circ$  interval, with a bias towards earlier projections. The artefacts due to the out-of-view mass are unavoidable but we can perform some further pre-processing to minimise the effect. In particular, if we shrink the field of view of the detector then the ‘heaviest’ part of the data will be the particle of interest and the model violations will be relatively small, increasing the signal to noise ratio. This can be seen as the sharp horizontal cut-off in the pre-processed sinograms seen on the right of figure 11. The effect of this on the reconstruction is going to be that there is a thin ring of mass placed at the edge of the (shrunk) detector

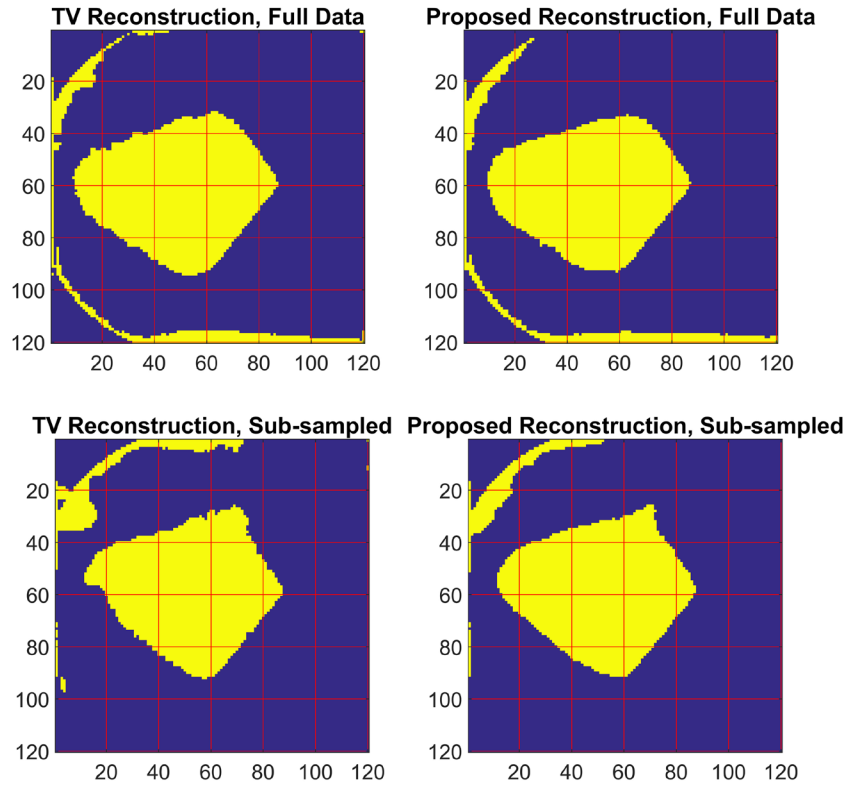


**Figure 12.** Reconstructions from a slice of the experimental data. We have chosen the slice half-way down through the projections shown in figure 11 to coincide with one of the rounded corners. The arc artefact was an anticipated consequence due to the out-of-view mass, the pre-processing has simply reduced the intensity. Proposed reconstructions consistently show better homogeneity inside the particle and sharper boundaries. The missing angles direction is the bottom-left to top-right diagonal where we see most error in each reconstruction, in particular, the blurring of the top right corner of the particle is a limited angle artefact.

view which will be clearly identifiable in the reconstruction. As a ground truth approximation we shall use a TV reconstruction on the full data for the location of the particle alongside prior knowledge of this sample for more precise geometrical features. We also note that the particle should be very homogeneous so this is another example where we expect the TV reconstruction to be very good.

The sample is a single crystal of silver and so we know it must have very uniform intensity and we are interested in locating the sharp facets which bound the crystal [48]. In figure 12 we immediately see that the combination of homogeneity and sharp edges is better reconstructed in our proposed reconstruction. Because we expect the reconstruction to be constant on the background and the particle, thresholding the reconstruction allows us to easily locate the boundaries and estimate interior angles of the particle. Figure 13 shows such images where the threshold is chosen to be the approximate midpoint of these two levels. We see that the proposed reconstruction consistently has less jagged edges and the left hand corner is better curved, as is consistent with our knowledge of the sample. Looking back at the full colour images we see that this is a result of lack of sharp decay at the boundary and homogeneity inside the sample. Looking for location error we see the biggest error in both TV and joint





**Figure 13.** Comparison between each reconstruction after thresholding. The geometrical properties of interest are that each edge should be linear, the left hand corner is rounded and the remaining corners are not. The particle of interest is homogeneous so thresholding the images should emphasise this in a way which is very unsympathetic to blurred edges. Again, the top right corner of each particle in the sub-sampled reconstructions coincides with the exacerbated missing wedge direction and so we expect each reconstruction to make some error here.

reconstruction is on the bottom-left edge where both reconstructions pull the line inwards. However, looking particularly at points  $(40, 80)$  and  $(20, 60)$ , we see that this was less severe in the proposed method. The other missing wedge artefact is in the top right corner which has been extended in both reconstructions although it is thinner in the proposed reconstruction. This indicates that it was better able to continue the straight edges either side of the corner and the blurring in the missing wedge direction is more localised than in the TV reconstruction. Overall, we see that the proposed reconstruction method is much more robust to an decrease in angular sampling range.

## 6. Conclusions and outlook

In this paper we have presented a novel method for tomographic reconstructions in a limited angle scenario along with a numerical algorithm with convergence guarantees. We have also tested our approach on synthetic and experimental data and shown consistent improvement over alternative reconstruction methods. Even when the x-ray transform model is noticeably

violated, as with our experimental data, we still better recover boundaries of the reconstructed sample.

There are three main directions which could be explored in future. Firstly, we think there is great potential to apply our framework to other applications, such as in tomographic imaging with occlusions and heavy metal artefacts where the inpainting region is much smaller [22, 23]. Secondly, we would like to find an alternative numerical algorithm with either faster practical convergence or one which is more capable of avoiding local minima in this non-convex landscape. Finally, we would like to explore the potential for an alternative regularisation functional on the sinogram which is better able to treat visible and invisible singularities, denoising and inpainting problems, independently. At the moment, the TV prior alone can reconstruct visible singularities well however, introducing a sinogram regulariser currently improves on the invisible region at the expense of damaging the visible. Overall, we feel that this presents the natural progression for the current work although it remains unclear how to regularise these invisible singularities.

## Acknowledgments

RT would like to thank Ozan Öktem for valuable discussions on the application of micro-local analysis in this setting. RT acknowledges support from the EPSRC grant EP/L016516/1 for the University of Cambridge Centre for Doctoral Training, the Cambridge Centre for Analysis. MB acknowledges support from the Leverhulme Trust Early Career Fellowship Learning from mistakes: ‘A supervised feedback-loop for imaging applications’, and the Isaac Newton Trust. CBS acknowledges support from the Leverhulme Trust project on ‘Breaking the non-convexity barrier’, EPSRC grant Nr. EP/M00483X/1, the EPSRC Centre Nr. EP/N014588/1, the RISE projects CHiPS and NoMADS, and the Alan Turing Institute. CBS, MB and RT also acknowledge the Cantab Capital Institute for the Mathematics of Information. CB acknowledges support from the EU RISE project NoMADS, the PIHC innovation fund of the Technical Medical Centre of UT and the Dutch 4TU programme Precision Medicine. MJL acknowledges financial support from the Netherlands Organization for Scientific Research (NWO), project 639.073.506. SMC acknowledges support from the Henslow Research Fellowship at Girton College, Cambridge. RKL acknowledges support from a Clare College Junior Research Fellowship. PAM acknowledges EPSRC grant EP/R008779/1.

## Appendix A. Theorem 2.1

**Theorem.** *If*

- (i)  $c_i$  are  $2k$  times continuously differentiable in  $\Delta$  and  $\Sigma$ ,  $k \geq 1$
- (ii)  $c_1(x|0, \Sigma) = c_2(x|0, \Sigma)$  for all  $x$  and  $\Sigma \geq 0$
- (iii)  $\partial_{\Delta}^{2j-1} c_1(x|0, \Sigma) = \partial_{\Delta}^{2j-1} c_2(x|0, \Sigma) = 0$  for all  $x$  and  $\Sigma \geq 0, j = 1 \dots, k$

Then  $A_d$  is  $C^{2k-1}$  with respect to  $d$  for all  $\rho > 0, \sigma \geq 0$ .

In this proof we will drop the  $x$  argument from  $c_i$  for conciseness of notation. Define

$$M_d = (\nabla d_{\rho} \nabla d_{\rho}^T)_{\sigma}.$$

Note that convolutions are bounded linear maps and  $\nabla d_{\rho} \in L^2$  by Young’s inequality hence  $M: L^1(\mathbb{R}^2, \mathbb{R}) \rightarrow L^{\infty}(\mathbb{R}^2, \text{Sym}_2)$  is well defined and differentiable w.r.t.  $d$ . Hence, it suffices to show that  $A$  is differentiable w.r.t.  $M$  where

$$M_d = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T, \lambda_1 \geq \lambda_2 \implies A = c_1(\Delta, \Sigma) e_1 e_1^T + c_2(\Delta, \Sigma) e_2 e_2^T$$

where  $\Delta = \lambda_1 - \lambda_2, \Sigma = \lambda_1 + \lambda_2$ . Note that this is not a trivial statement, we can envisage very simple cases in which the (ordered) eigenvalue decomposition is not even continuous. For instance

$$M(t) = \begin{pmatrix} 1-t & 0 \\ 0 & t \end{pmatrix} \implies \Sigma(t) = 1, \Delta(t) = |1-2t|, e_1 = \begin{cases} (1,0)^T & t < 1/2 \\ (0,1)^T & t > 1/2 \end{cases}.$$

Hence we can see that despite having  $M \in C^\infty$  we cannot even guarantee that the decomposition is continuous and so cannot employ any chain rule to say that  $A$  is smooth.

The structure of this proof breaks into four parts:

- (i) If  $c_1(0, \Sigma) = c_2(0, \Sigma)$  then  $A$  is well defined
- (ii) If  $c_i \in C^2$  for some open neighbourhood of point  $x$  such that  $\lambda_1(x) > \lambda_2(x)$  then  $A$  is differentiable w.r.t.  $M$  on an open neighbourhood of  $x$
- (iii) If  $\partial_\Delta c_1(0, \Sigma) = \partial_\Delta c_2(0, \Sigma) = 0$  at a point,  $x$ , where  $\lambda_1(x) = \lambda_2(x)$  then  $A$  is differentiable on an open neighbourhood of  $x$
- (iv) If  $\partial_\Delta^{2j-1} c_1(0, \Sigma) = \partial_\Delta^{2j-1} c_2(0, \Sigma) = 0$  at a point  $x$  where  $\lambda_1(x) = \lambda_2(x)$  and for all  $j = 1 \dots, k$  then  $A$  is  $C^{2k-1}$  on an open neighbourhood of  $x$

**Proof.** Proof of part i: note that when  $\lambda_1 = \lambda_2$  the choice of  $e_i$  is non-unique subject to  $e_1 e_1^T + e_2 e_2^T = \text{id}$  and so we get

$$A = c_1(0, \Sigma) \text{id} + (c_2(0, \Sigma) - c_1(0, \Sigma)) e_2 e_2^T.$$

Hence  $A$  is well defined if and only if  $c_1(0, \Sigma) = c_2(0, \Sigma)$  for all  $\Sigma \geq 0$ .

As we are decomposing  $2 \times 2$  matrices it will simplify the proof to write explicit forms for the values under consideration.

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{12} & M_{22} \end{pmatrix} \implies \lambda_i = \frac{M_{11} + M_{22} \pm \sqrt{(M_{11} - M_{22})^2 + 4M_{12}^2}}{2}$$

$$\Sigma = M_{11} + M_{22}, \quad \Delta = \sqrt{(M_{11} - M_{22})^2 + 4M_{12}^2}$$

$$\Delta \neq 0 \implies e_1 = \frac{(2M_{12}, \Delta - M_{11} + M_{22})^T}{\sqrt{(\Delta - M_{11} + M_{22})^2 + 4M_{12}^2}} = \frac{(\Delta + M_{11} - M_{22}, 2M_{12})^T}{\sqrt{(\Delta + M_{11} - M_{22})^2 + 4M_{12}^2}},$$

$$e_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} e_1.$$

We give two equations for  $e_1$  to account for the case when we get  $\frac{(0,0)^T}{0}$ .

Proof of part ii: note that  $\Sigma$  is always smooth and  $\Delta$  is smooth on the set  $\{\Delta > 0\}$

Case  $M_{12}(x) \neq 0$ : now both equations of  $e_1$  are valid (and equal) and the denominators non-zero on a neighbourhood. Hence, we can apply the standard chain rule and product rule to conclude.

Case  $M_{12}(x) = 0$ : in this case  $M(x)$  is diagonal but as  $\Delta = |M_{11} - M_{22}| > 0$  we know that one of our formulae for  $e_1$  must be valid with non-vanishing denominator in a neighbourhood and so we can conclude as in the first case.

Proof of part iii: given the Neumann condition on  $c_i$  we shall express  $c_i$  locally by Taylor's theorem.

$$c_i(\Delta, \Sigma) = c_i(0, \Sigma) + O(\Delta^2) = c_1(0, \Sigma) + O(\Delta^2).$$

Now we consider a perturbation:

$$\begin{aligned} M &= \begin{pmatrix} m & 0 \\ 0 & m \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{12} & \varepsilon_{22} \end{pmatrix} \\ \implies A(M + \varepsilon) - A(M) &= (c_1(0, 2m + \varepsilon_{11} + \varepsilon_{22}) - c_1(0, 2m))\text{id} + O(\Delta^2) \\ \Delta^2 &= (\varepsilon_{11} - \varepsilon_{22})^2 + 4\varepsilon_{12}^2 = O(\|\varepsilon\|^2) \implies O(\Delta^2) \leq O(\|\varepsilon\|^2) \\ \implies \frac{A(M + \varepsilon) - A(M)}{\|\varepsilon\|} &= \frac{\partial_{\Sigma} c_1(0, 2m) \text{tr}(\varepsilon)}{\|\varepsilon\|} + O(\|\varepsilon\|). \end{aligned}$$

In particular,  $A$  is differentiable w.r.t.  $M$  at  $x$ .

Proof of part iv: the proof of this follows exactly as the previous part,

$$c_i(\Delta, \Sigma) = \sum_0^{k-1} \frac{\Delta^{2j}}{j!} \partial_{\Delta}^{2j} c_i(0, \Sigma) + O(\Delta^{2k})$$

where the remainder term is sufficiently smooth. Hence  $c_i$  is at least  $2k - 1$  times differentiable w.r.t.  $M$ .  $\square$

## Appendix B. Theorem 3.1

**Theorem.** *If*

- $c_i$  are bounded away from 0
- $\rho > 0$
- $A_d$  is differentiable in  $d$

*then sub-level sets of  $E$  are weakly compact in  $L^2(\Omega, \mathbb{R}) \times L^2(\mathbb{R}^2, \mathbb{R})$  and  $E$  is weakly lower semi-continuous. i.e. for all  $(u_n, v_n) \in L^2(\Omega, \mathbb{R}) \times L^2(\mathbb{R}^2, \mathbb{R})$ :*

*$E(u_n, v_n)$  uniformly bounded implies a subsequence converges weakly*

$$\liminf E(u_n, v_n) \geq E(u, v) \text{ whenever } u_n \rightharpoonup u, v_n \rightharpoonup v$$

**Proof.** If  $c_i$  are bounded away from 0 then in particular we have  $A_{\mathcal{R}u_n} \gtrsim 1$  so  $\text{DTV}_{u_n}(v_n) = \|A_{\mathcal{R}u_n} \nabla v_n\| \gtrsim \|\nabla v_n\| = \text{TV}(v_n)$ . Hence,

$E(u_n, v_n)$  uniformly bounded  $\implies$

$$\begin{aligned} & \|S_{\Omega^c}(\mathcal{R}u_n - v_n)\|_2^2 + \|S_{\Omega'}\mathcal{R}u_n - b\|_2^2 + \|S_{\Omega'}v_n - b\|_2^2 \\ & \quad + \text{TV}(u_n) + \text{TV}(v_n) \text{ uniformly bounded} \\ \implies & \|A(u, v)^T - b\|_2^2 + \text{TV}((u, v)) \text{ uniformly bounded} \end{aligned}$$

for some linear  $A$  and constant  $b$ . Thus we are in a very classical setting where weak compactness can be shown in both the  $\|(u, v)\|_2$  norm and  $\|(u, v)\|_1 + \text{TV}((u, v))$  [50].

We now proceed to the second claim, that  $E$  is weakly lower semi-continuous. Note that all of the convex terms in our energy are lower semi-continuous by classical arguments so it remains to show that the non-convex term is too. i.e.

$$(u_n, v_n) \rightharpoonup (u, v) \stackrel{?}{\implies} \liminf \|A_{\mathcal{R}u_n} \nabla v_n\|_{2,1} \geq \|A_{\mathcal{R}u} \nabla v\|_{2,1}.$$

We shall first present a lemma from distribution theory.

**Lemma B.1.** If  $\Omega$  is compact,  $\varphi \in C^\infty(\Omega, \mathbb{R})$  and  $w_n \xrightarrow{L^p} w$  then

$$w_n \star \varphi \rightarrow w \star \varphi \text{ convergence in } C^k(\Omega, \mathbb{R}) \text{ for all } k < \infty.$$

*Proof.* Recall that  $w_n \rightharpoonup w \implies \|w_n\|_p \leq W$  for some  $W < \infty$ . By Hölder's inequality:

$$|w_n \star \varphi(x) - w \star \varphi(y)| \leq \int |w_n(z)| |\varphi(x-z) - \varphi(y-z)| \lesssim_{p,\Omega} |x-y| W \|\varphi'\|_\infty$$

$$|w_n \star \varphi(x)| \lesssim_{p,\Omega} W \|\varphi\|_\infty$$

i.e.  $\{w_n \text{ s.t. } n \in \mathbb{N}\}$  is uniformly bounded and uniformly (Lipschitz) continuous.

$$w_n \rightharpoonup w \implies w_n \star \varphi(x) - w \star \varphi(x) = \langle w_n - w, \varphi(x - \cdot) \rangle \rightarrow 0 \text{ for every } x.$$

Hence, we also know  $w_n \star \varphi$  converges point-wise to a unique continuous function. Suppose  $\|w_{n_k} \star \varphi - w \star \varphi\|_\infty \geq \varepsilon > 0$  for some  $\varepsilon$  and sub-sequence  $n_k \rightarrow \infty$ . By the Arzela–Ascoli theorem some further subsequence must converge uniformly to the point-wise limit,  $w \star \varphi$ , which gives us the required contradiction. Hence,  $w_n \star \varphi \rightarrow w \star \varphi$  in  $L^\infty = C^0$ . The general theorem follows by induction.  $\square$

This lemma gives us two direct results.

$$\rho > 0 \implies (Ru_n)_\rho \rightarrow (Ru)_\rho \text{ in } L^\infty$$

$$\{(Ru_n)_\rho\} \cup \{(Ru)_\rho\} \text{ compact, } A_d \in C^1(d) \implies A_{Ru_n} \rightarrow A_{Ru} \text{ in } \|\cdot\|_{2,\infty}.$$

Hence, whenever  $w \in W^{1,1}$  we have

$$\begin{aligned} \|A_{\mathcal{R}u_n} \nabla w\| & \geq \|A_{\mathcal{R}u} \nabla w\| - \|(A_{\mathcal{R}u_n} - A_{\mathcal{R}u}) \nabla w\| \\ & \geq \|A_{\mathcal{R}u} \nabla w\| - \|A_{\mathcal{R}u_n} - A_{\mathcal{R}u}\|_{2,\infty} \text{TV}(w). \end{aligned}$$

By density of  $W^{1,1}$  in the space of bounded variation, we know the same holds for  $w = v_n$ . We also know  $\text{TV}(v_n)$  is uniformly bounded thus

$$\liminf \|A_{\mathcal{R}u_n} \nabla v_n\| = \liminf \|A_{\mathcal{R}u} \nabla v_n\|.$$

Hence, for all  $\|\varphi\|_{2,\infty} \leq 1$  we have

$$\langle v, \nabla \cdot (A_{\mathcal{R}u} \varphi) \rangle = \liminf_n \langle v_n, \nabla \cdot (A_{\mathcal{R}u} \varphi) \rangle \leq \liminf \|A_{\mathcal{R}u} \nabla v_n\| \leq \liminf \|A_{\mathcal{R}u_n} \nabla v_n\|$$

as required.  $\square$

## ORCID iDs

Robert Tovey  <https://orcid.org/0000-0001-5411-2268>

Martin Benning  <https://orcid.org/0000-0002-6203-1350>

Christoph Brune  <https://orcid.org/0000-0003-0145-5069>

Marinus J Lagerwerf  <https://orcid.org/0000-0003-1916-4665>

Sean M Collins  <https://orcid.org/0000-0002-5151-6360>

Paul A Midgley  <https://orcid.org/0000-0002-6817-458X>

Carola-Bibiane Schönlieb  <https://orcid.org/0000-0003-0099-6306>

## References

- [1] Ehrhardt M J, Thielemans K, Pizarro L, Atkinson D, Ourselin S, Hutton B F and Arridge S R 2015 Joint reconstruction of PET-MRI by exploiting structural similarity *Inverse Problems* **31** 015001
- [2] Leary R K, Saghi Z, Midgley P A and Holland D J 2013 Compressed sensing electron tomography *Ultramicroscopy* **131** 70–91
- [3] Kübel C, Voigt A, Schoenmakers R, Otten M, Su D, Lee T-C, Carlsson A and Bradley J 2005 Recent advances in electron tomography: TEM and HAADF-STEM tomography for materials science and semiconductor applications *Microsc. Microanal.* **11** 378–400
- [4] Zhao G et al 2013 Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics *Nature* **497** 643–6
- [5] Kalender W A 2006 X-ray computed tomography *Phys. Med. Biol.* **51** 29–43
- [6] Cnudde V and Boone M N 2013 High-resolution x-ray computed tomography in geosciences: a review of the current technology and applications *Earth-Sci. Rev.* **123** 1–17
- [7] Kawase N, Kato M, Nishioka H and Jinnai H 2007 Transmission electron microtomography without the ‘missing wedge’ for quantitative structural analysis *Ultramicroscopy* **107** 8–15
- [8] Zhang Y, Chan H P, Sahiner B, Wei J, Goodsitt M M, Hadjiiski L M, Ge J and Zhou C 2006 A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis *Med. Phys.* **33** 3781–95
- [9] Quinto E T 1993 Singularities of the x-ray transform and limited data tomography in  $\{R^2\}$  and  $\{R^3\}$  *SIAM J. Math. Anal.* **24** 1215–25
- [10] Friel J and Quinto E T 2013 Characterization and reduction of artifacts in limited angle tomography *Inverse Problems* **29** 21
- [11] Katsevich A I 1997 Local tomography for the limited-angle problem *J. Math. Anal. Appl.* **213** 160–82
- [12] Feldkamp L A, Davis L C and Kress J W 1984 Practical cone-beam algorithm *J. Opt. Soc. Am. A* **1** 612
- [13] Flohr T, Stierstorfer K, Bruder H, Simon J, Polacin A and Schaller S 2003 Image reconstruction and image quality evaluation for a 16-slice CT scanner *Med. Phys.* **30** 832–45
- [14] Tang X, Hsieh J, Nilsen R A, Dutta S, Samsonov D and Hagiwara A 2006 A three-dimensional-weighted cone beam filtered backprojection (CB-FBP) algorithm for image reconstruction in volumetric CT-helical scanning *Phys. Med. Biol.* **51** 855–74
- [15] Gilbert P 1972 Iterative methods for the three-dimensional reconstruction of an object from projections *J. Theor. Biol.* **36** 105–17
- [16] Agulleiro J I, Garzón E M, García I and Fernández J J 2010 Vectorization with SIMD extensions speeds up reconstruction in electron tomography *J. Struct. Biol.* **170** 570–5

- [17] Spitzbarth M and Drescher M 2015 Simultaneous iterative reconstruction technique software for spectral-spatial EPR imaging *J. Magn. Reson.* **257** 79–88
- [18] Goris B, Van den Broek W, Batenburg K J, Heidari Mezerji H and Bals S 2012 Electron tomography based on a total variation minimization reconstruction technique *Ultramicroscopy* **113** 120–30
- [19] Chen Z, Jin X, Li L and Wang G 2013 A limited-angle CT reconstruction method based on anisotropic TV minimization *Phys. Med. Biol.* **58** 2119–41
- [20] Gu J and Ye J C 2017 Multi-scale wavelet domain residual learning for limited-angle CT reconstruction (arXiv submission)
- [21] Hammernik K, Würfl T, Pock T and Maier A 2017 A deep learning architecture for limited-angle computed tomography reconstruction *Bildverarbeitung für die Medizin* (Berlin: Springer) pp 92–7
- [22] Köstler H, Prümmer M, Rude U and Hornegger J 2006 Adaptive variational sinogram interpolation of sparsely sampled CT data *Proc. Int. Conf. on Pattern Recognition* vol 3 (IEEE) pp 778–81
- [23] Zhang Y, Pu Y F, Hu J R, Liu Y and Zhou J L 2011 A new CT metal artifacts reduction algorithm based on fractional-order sinogram inpainting *J. X-Ray Sci. Technol.* **19** 373–84
- [24] Li S, Cao Q, Chen Y, Hu Y, Luo L and Toumoulin C 2014 Dictionary learning based sinogram inpainting for CT sparse reconstruction *Optik* **125** 2862–7
- [25] Gu J, Zhang L, Yu G, Xing Y and Chen Z 2006 X-ray CT metal artifacts reduction through curvature based sinogram inpainting *J. X-Ray Sci. Technol.* **14** 73–82
- [26] Zhang H, Wang L, Duan Y, Li L, Hu G and Yan B 2017 Euler's elastica strategy for limited-angle computed tomography image reconstruction *IEEE Trans. Nucl. Sci.* **64** 2395–405
- [27] Kalke M and Siltanen S 2014 Sinogram interpolation method for sparse-angle tomography *Appl. Math.* **05** 423–41
- [28] Thirion J-P 1991 A geometric alternative to computed tomography *Engineering in Medicine and Biology Society* vol **13** p 34 (<https://doi.org/10.1109/IEMBS.1991.683828>)
- [29] Burger M, Müller J, Papoutsellis E and Schönlieb C-B 2014 Total variation regularisation in measurement and image space for PET reconstruction *Inverse Problems* **30** 105003
- [30] Dong B, Li J and Shen Z 2013 X-ray CT image reconstruction via wavelet frame based regularization and Radon domain inpainting *J. Sci. Comput.* **54** 333–49
- [31] Ochs P, Fadili J M and Brox T 2017 Non-smooth non-convex Bregman minimization: unification and new algorithms (arXiv submission)
- [32] Beck A and Teboulle M 2009 Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems *IEEE Trans. Image Process.* **18** 2419–34
- [33] Chan T F, Esedoglu S and Nikolova M 2006 Algorithms for finding global minimizers of image segmentation and denoising models *SIAM J. Appl. Math.* **66** 1632–48
- [34] Benning M, Möller M, Nossek R Z, Burger M, Cremers D, Gilboa G and Schönlieb C-B 2017 Nonlinear spectral image fusion *Scale Space and Variational Methods in Computer Vision* (Cham: Springer) vol 10302 pp 41–53
- [35] Weickert J 1998 *Anisotropic Diffusion in Image Processing* (Stuttgart: Teubner)
- [36] Berkels B, Burger M, Droske M, Nemitz O and Rumpf M 2006 Cartoon extraction based on anisotropic image classification *Vision, Modeling and Visualization* (Amsterdam: IOS Press) p 293
- [37] Estellers V, Soatto S and Bresson X 2015 Adaptive regularization with the structure tensor *IEEE Trans. Image Process.* **24** 1777–90
- [38] Bertalmio M, Sapiro G, Caselles V and Ballester C 2000 Image inpainting *Proc. of the 27th Annual Conf. on Computer Graphics and Interactive Techniques—SIGGRAPH* (New York: ACM) pp 417–24
- [39] Chambolle A and Pock T 2011 A first-order primal-dual algorithm for convex problems with applications to imaging *J. Math. Imaging Vis.* **40** 120–45
- [40] Mosek ApS 2010 The Mosek optimization software [www.mosek.com](http://www.mosek.com)
- [41] Drusvyatskiy D and Lewis A S 2018 Error bounds, quadratic growth, and linear convergence of proximal methods *Math. Oper. Res.* **43** 693–1050
- [42] Pock T and Sabach S 2016 Inertial proximal alternating linearized minimization (iPALM) for nonconvex and nonsmooth problems *SIAM J. Imaging Sci.* **9** 1756–87
- [43] Liang J, Fadili J M and Peyré G 2016 A multi-step inertial forward-backward splitting method for non-convex optimization *Adv. Neural Inf. Process. Syst.* **29** 4035–43



- [44] Ochs P, Chen Y, Brox T and Pock T 2014 iPiano: inertial proximal algorithm for non-convex optimization *SIAM J. Imaging Sci.* **7** 1388–419
- [45] Bolte J, Sabach S and Teboulle M 2014 Proximal alternating linearized minimization for nonconvex and nonsmooth problems *Math. Program.* **146** 459–94
- [46] Drusvyatskiy D, Ioffe A D and Lewis A S 2016 Nonsmooth optimization using Taylor-like models: error bounds, convergence, and termination criteria (arXiv submission)
- [47] Rockafellar R T and Wets R J B 1998 *Variational Analysis* (Berlin: Springer) vol 317 (<https://doi.org/10.1007/978-3-642-02431-3>)
- [48] Collins S M, Ringe E, Duchamp M, Saghi Z, Dunin-Borkowski R E and Midgley P A 2015 Eigenmode tomography of surface charge oscillations of plasmonic nanoparticles by electron energy loss spectroscopy *ACS Photonics* **2** 1628–35
- [49] Grant M and Boyd S CVX: Matlab Software for Disciplined Convex Programming, version 2.1 (<http://cvxr.com/cvx>) (Accessed: March 2014)
- [50] Chambolle A and Lions P-L 1997 Image recovery via total variation minimization and related problems *Numer. Math.* **76** 167–88