RAS Techniques and Instruments

RASTAI **2,** 207–215 (2023)
Advance Access publication 2023 April 28

https://doi.org/10.1093/rasti/rzad010

Overcomplete tomography: a novel approach to imaging

Buse Turunçtur [1]★ Andrew Valentine [1,2] and Malcolm Sambridge [1]

[1]*Research School of Earth Sciences, The Australian National University, Canberra ACT 2601, Australia*
[2]*Department of Earth Sciences, Durham University, Durham, DH1 3LE, UK*

Accepted 2023 March 24. Received 2023 February 21; in original form 2022 November 6

ABSTRACT

Regularized least-squares tomography offers a straightforward and efficient imaging method and has seen extensive application across various fields. However, it has a few drawbacks, such as (i) the regularization imposed during the inversion tends to give a smooth solution, which will fail to reconstruct a multi-scale model well or detect sharp discontinuities, (ii) it requires finding optimum control parameters, and (iii) it does not produce a sparse solution. This paper introduces 'overcomplete tomography', a novel imaging framework that allows high-resolution recovery with relatively few data points. We express our image in terms of an overcomplete basis, allowing the representation of a wide range of features and characteristics. Following the insight of 'compressive sensing', we regularize our inversion by imposing a penalty on the L_1 norm of the recovered model, obtaining an image that is sparse relative to the overcomplete basis. We demonstrate our method with a synthetic and a real X-ray tomography example. Our experiments indicate that we can reconstruct a multi-scale model from only a few observations. The approach may also assist interpretation, allowing images to be decomposed into (for example) 'global' and 'local' structures. The framework presented here can find application across a wide range of fields, including engineering, medical and geophysical tomography.

Key words: algorithms – data methods – L_1-norm based inversion – model parametrization – multiscale tomography – overcomplete basis.

1 INTRODUCTION

In many imaging situations, it is not possible to observe the feature of interest directly. Instead, we must rely on whatever information can be extracted from secondary data sets. For example, our knowledge of the Earth's interior structure must be inferred from surface observables, such as measurements of seismic waves or variations in the gravity field. To achieve this, we first introduce a mathematical framework for describing the target feature – a model parametrization – and then develop techniques to simulate the corresponding observables. Finally, in a process often known as 'inversion', we can infer the set of model parameters that generate the best match to observations.

Usually, a system can be described using a wide variety of different model parametrizations. For example, models of Earth structure have been parametrized using (amongst others) spherical harmonics (e.g. Woodhouse & Dziewonski 1984; Ritsema et al. 2011), voxels or cells (Spakman 1991; Hilst et al. 1997), and wavelets (Simons et al. 2011). The choice of parametrization is sometimes regarded as a somewhat arbitrary choice (Valentine & Trampert 2016), but it can exert significant influence upon both the computational tractability of an inversion scheme, and upon the characteristics of the solution that is ultimately obtained – e.g. models expressed using spherical harmonics are necessarily smooth, whereas voxel models may exhibit sharp discontinuities.

To illustrate the consequences of this, Fig. 1 shows a simple synthetic X-ray imaging scenario. The true model, as shown in Fig. 1b, comprises some sharp, isolated features (expressed in a pixel basis; see below) superposed upon a smooth, long-wavelength background structure (expressed in a discrete cosine basis). Using synthetic attenuation data computed for the 300 ray paths shown in Fig. 1a, we compare results from two inversions performed using a Tikhonov-regularized least-squares algorithm: one seeking the best-fitting model expressed in (only) a pixel basis (Fig. 1c), and the other expressed in a discrete cosine basis (Fig. 1e). Unsurprisingly, neither succeeds in fully capturing the features of the true structure. While aspects of the isolated structures are visible in each, it is difficult to appreciate their true extent and character, and they cannot easily be distinguished from the background. In a real problem, this would likely be a barrier to analysis and exploitation of results, and interpretation might differ depending on the basis functions chosen to parameterize the inversion.

An obvious solution presents itself: why restrict the inversion to use only a single family of basis functions? In Fig. 1d, we again show results from a Tikhonov-regularized least-squares tomography, but with the model parametrized using both pixel and discrete cosine bases in combination, which we refer to as forming an 'overcomplete' basis, following Duffin & Schaeffer (1952). Unfortunately, this does not appear to yield a marked improvement in results: We see that most structure continues to be expressed using the pixel part of the model, and the difficulties of interpretation remain. However – as this paper describes – by using the overcomplete basis in conjuction with an alternative regularization scheme, we can achieve the results shown in Fig. 1f.

★ E-mail: buse.turunctur@anu.edu.au

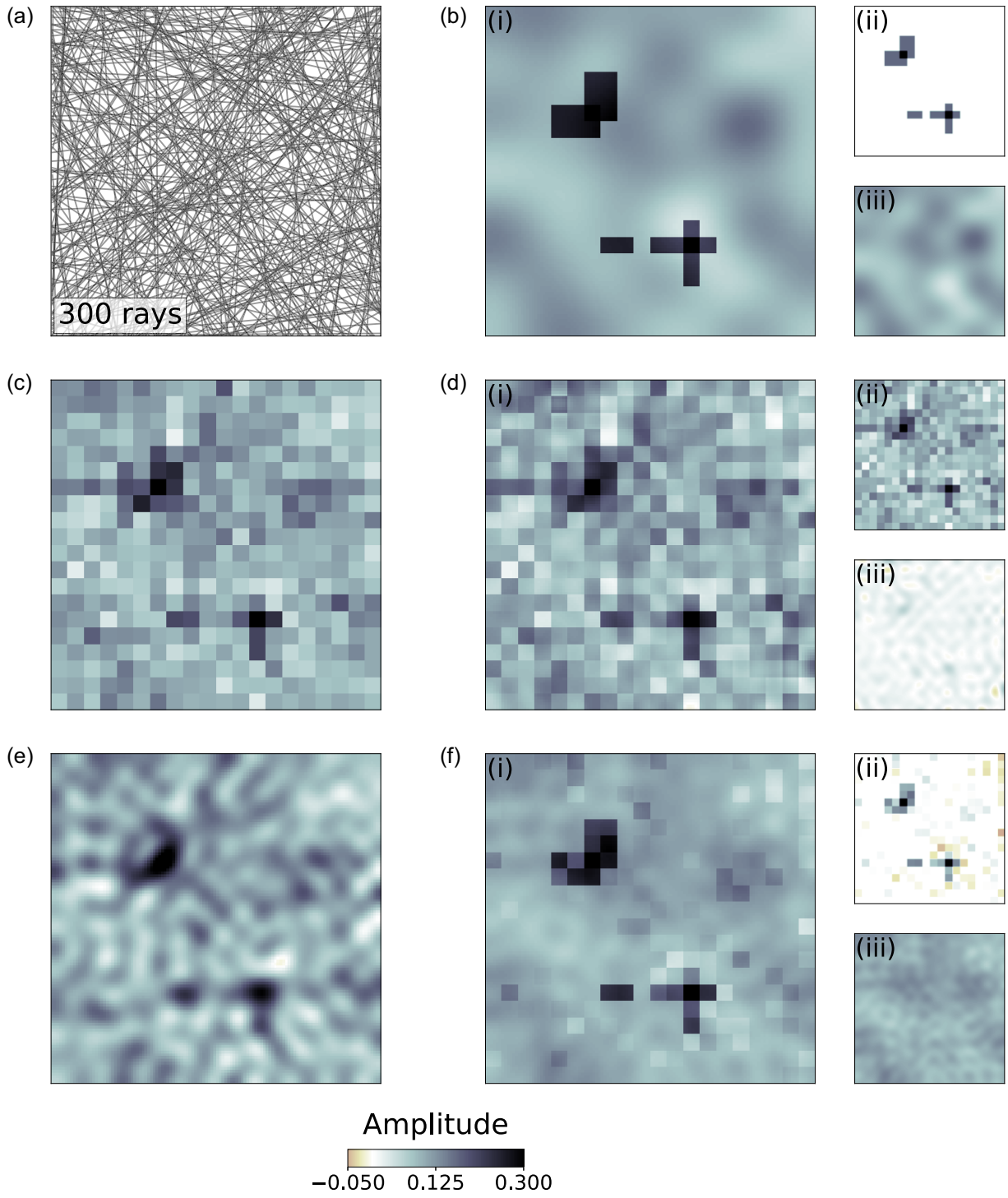

Figure 1 An illustration of tomographic solutions recovered with various model parametrizations and regularization approaches. (a) 300 Ray paths with randomly selected end points on a square boundary and (b) synthetic model used to generate line-integral X-ray attenuation data with equation (1). Recovered images after imposing Tikhonov regularization (c) using pixel basis; (d) overcomplete basis (Tikhonov-regularized overcomplete tomography; TROT); and (e) cosine basis functions. (f) Shows the recovered image using sparsity regularization and an overcomplete basis [referred to as sparsity-promoting overcomplete tomography (SPOT) in text]. For each overcomplete case, (i) shows combined model, (ii) pixel model, and (iii) the cosine model.

We refer to this as 'sparsity-promoting overcomplete tomography (SPOT)', and it enables excellent recovery of the target structures, revealing the distinct characteristics of background and anomalies.

To achieve this, we build on a significant body of work developed in support of the concept of 'compressive sensing'. Various studies, including those of Candès et al. (2006), Donoho (2006), and Candes & Tao (2006), have demonstrated that it is possible to recover signals from data sampled well below the Nyquist limit, provided that certain conditions are met. Underpinning this are two key ideas: First, that many systems of interest have 'sparse' representations, and can be

described using only a few parameters in an appropriately-chosen basis; and second, that minimizing the L_1 norm of a model can serve as a reasonably-accurate yet computationally-tractable proxy for minimization of its L_0 norm (Candès et al. 2006). Following this approach, we assume that real imaging targets have a sparse representation in the overcomplete basis, and we regularize the inversion based on the L_1 norm of the resulting model vector. As we show, this yields excellent practical results, particularly in situations where only a limited amount of data is available.

2 THEORETICAL BACKGROUND

In any tomographic imaging problem, our goal is to characterize a function $m(\mathbf{x})$, representing spatial variations in some physical property of interest. For example, in a geophysical setting m might represent the velocity of seismic waves within the Earth's interior; while in a medical setting it might represent the density of body tissue. However, rather than measuring m directly, we must infer it from its signature in observable data. In this paper, we consider only those problems where the data are assumed to be linearly dependent upon m, so that the ith datum, d_i, can be expressed:

$$d_i = \int_{\mathcal{X}} g_i(\mathbf{x}) m(\mathbf{x})\, \mathrm{d}^N\mathbf{x}\,. \tag{1}$$

Here, $g_i(\mathbf{x})$ is a kernel quantifying how the observation samples the model, and $\mathcal{X}$ represents the N-dimensional space within which the model is defined.

2.1 Least squares inversion

The conventional approach to inversion involves introducing some family of basis functions, $\phi_j(\mathbf{x})$, $(j = 1, \ldots, M)$. We assume that this family is sufficient to represent the system of interest, so that it forms a complete basis for representing m, and write

$$m(\mathbf{x}) = \sum_{j=1}^{M} m_j \phi_j(\mathbf{x}) \tag{2}$$

Within this representation, any model can be fully described by specifying the M parameters $m_{1\ldots M}$, which we collectively represent as the model vector $\mathbf{m}$. Substituting this into equation (1), we obtain

$$d_i = \sum_{j=1}^{M} m_j \int_{\mathcal{X}} g_i(\mathbf{x})\phi_j(\mathbf{x})\, \mathrm{d}^N\mathbf{x} \tag{3}$$

and this can be written in the form $\mathbf{d} = \mathbf{G}\mathbf{m}$, where the matrix $\mathbf{G}$ is defined to have elements $G_{ij} = \int_{\mathcal{X}} g_i(\mathbf{x})\phi_j(\mathbf{x})\, \mathrm{d}^N\mathbf{x}$.

In general, observational data will contain noise, and the system may not adhere precisely to the assumptions inherent to equation (1). We therefore do not expect to find a model that explains the data perfectly; instead, we seek the 'best-fitting' model and set out to minimize some objective function:

$$\chi(\mathbf{m}) = \frac{1}{2N}\,(\mathbf{d} - \mathbf{G}\mathbf{m})^{\mathsf{T}}\, \mathbf{C}_{\mathrm{d}}^{-1}\,(\mathbf{d} - \mathbf{G}\mathbf{m}) + \chi_{\mathrm{m}}(\mathbf{m}) \tag{4}$$

where $\mathbf{C}_{\mathrm{d}}$ is a covariance matrix representing any noise and/or correlations expected to be present within the data set, and where χ_{m} is a 'regularization function' that encodes any preferences we may wish to express regarding the model sought. One common choice is that of Tikhonov regularization, as already employed for Fig. 1, whereby

$$\chi_{\mathrm{m}}(\mathbf{m}) = \alpha\, \|\mathbf{m}\|_2^2 \tag{5}$$

with α ($\geq$0) chosen to balance the relative weights of data-fit and regularization terms.

2.2 Overcomplete inversion

To extend this, we introduce K distinct families of basis functions, and use $\phi_j^{(k)}(\mathbf{x})$ to denote the jth basis function in the kth set. In analogy with equation (2), we then define

$$m(\mathbf{x}) = \sum_{k=1}^{K}\sum_{j=1}^{M_k} m_j^{(k)}\phi_j^{(k)}(\mathbf{x})\,, \tag{6}$$

with M_k denoting the number of basis functions comprising the kth set. The model function is therefore fully specified by $M = \sum_{k=1}^{K} M_k$ parameters, which can be regarded both as a single M-dimensional vector, $\mathbf{m}$, and as a collection of K subsidiary model vectors $\mathbf{m}^{(k)}$.

As before, we can substitute this into equation (1), leading to a system of equations of the form

$$\mathbf{d} = \sum_{k=1}^{K} \mathbf{G}^{(k)}\mathbf{m}^{(k)} \tag{7}$$

where each $\mathbf{G}^{(k)}$ has elements defined by $G_{ij}^{(k)} = \int_{\mathcal{X}} g_i(\mathbf{x})\phi_j^{(k)}(\mathbf{x})\, \mathrm{d}^N\mathbf{x}$. Defining

$$\mathbf{G} = \left(\mathbf{G}^{(1)} \ldots \mathbf{G}^{(K)}\right) \text{ and } \mathbf{m} = \begin{pmatrix} \mathbf{m}^{(1)} \\ \vdots \\ \mathbf{m}^{(K)} \end{pmatrix} \tag{8}$$

allows us to continue to pose the inverse problem in terms of the minimization of $\chi(\mathbf{m})$, as given in equation (4).

In principle, we could regularize this inversion using a Tikhonov-style regularization function, as in equation (5). However, experimental results prove poor (as we have already seen in Fig. 1). Instead, we adopt a regularization function of the form

$$\chi_{\mathrm{m}}(\mathbf{m}) = \alpha \sum_{k=1}^{K} \beta_k \left\|\mathbf{C}_{\mathrm{d}}^{-1/2}\mathbf{G}^{(k)}\right\|_2 \left\|\mathbf{m}^{(k)}\right\|_1 \quad \text{s.t.} \quad \sum_{k=1}^{K}\beta_k = 1\,. \tag{9}$$

Here, α continues to control the overall balance between data-fit and regularization terms, while the β_k (with $0 \leq \beta_k \leq 1$) allow for individual basis families to be up- or down-weighted relative to one another. The (scalar) factors of $\left\|\mathbf{C}_{\mathrm{d}}^{-1/2}\mathbf{G}^{(k)}\right\|_2$ are included to account for any differences in the 'units of measurement' associated with different basis function families, rescaling model norms into a common system; other choices may be possible here. We will refer to inversion based on minimizing this misfit as 'SPOT'; see below for justification of this name.

3 EXAMPLES

3.1 A synthetic test

To illustrate the behaviour and performance of this scheme, we first return to the simple example presented in Fig. 1. This is a simple two-dimensional imaging problem based on X-ray tomography, as discussed by Tarantola (2005, section 5.6). Our model function, $m(x, y)$ represents the 'density' of matter at location (x, y) within a domain $x \in [0, 1]$, $y \in [0, 1]$; we simulate X-ray transmission from sources of known intensity located on the boundary of this domain and compute the intensity received at distant receivers, also located on the boundary. This allows us to compute synthetic sinogram measurements:

$$d_i = \int_{\Gamma_i} m(x, y)\, \mathrm{d}s \tag{10}$$

where integration is along the straight-line path, Γ_i, between source and receiver.

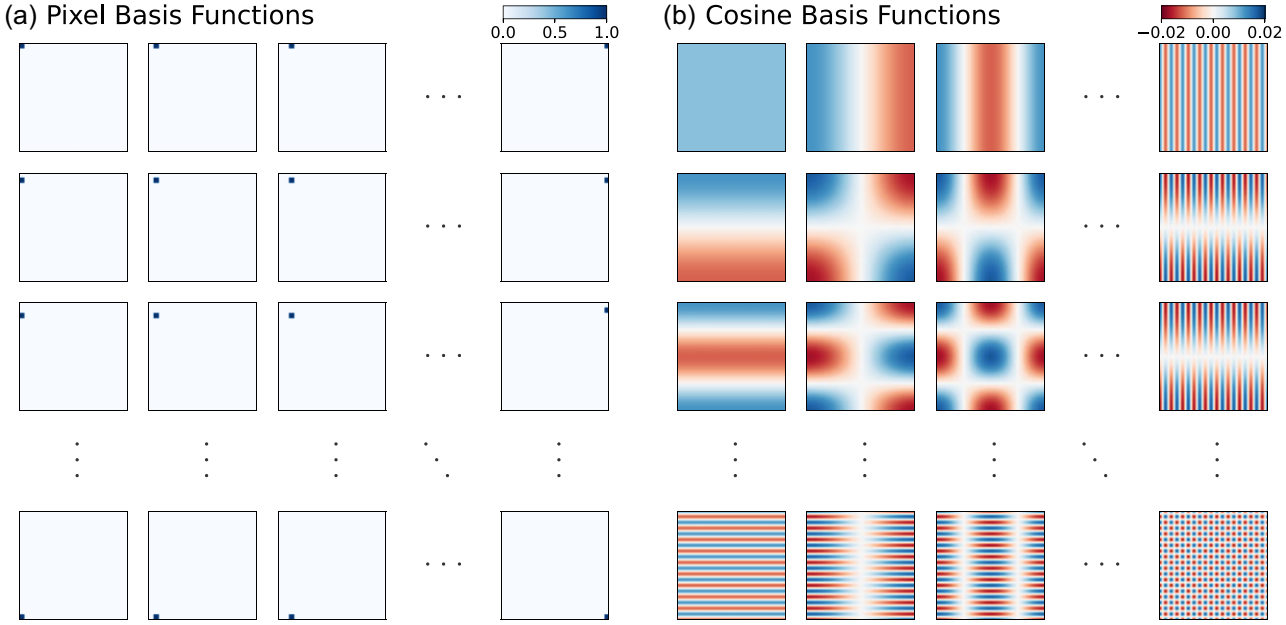

Figure 2 Visualization of the pixel and cosine basis functions. The first three rows/columns and the final row/column of the 20 × 20 set of basis functions are shown.

We employ two sets of basis functions for our experiments. The first comprises a 20 × 20 grid of square pixels (as shown in Fig. 2a), such that

$$\phi_i^{(p)}(x, y) = \begin{cases} 1 & \frac{m-1}{20} < x \leq \frac{m}{20} \text{ and } \frac{n-1}{20} < y \leq \frac{n}{20} \\ & \text{with } i = 20(m-1)+n \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

while the second is a 20 × 20 Fourier cosine basis (Fig. 2b):

$$\phi_i^{(c)}(x, y) = \cos(m\pi x)\cos(n\pi y) \quad \text{with } i = 20(m-1)+n\,. \tag{12}$$

The first of these is an example of a local basis, with each basis function being non-zero in only a small region of the model domain. Discontinuous structures can be represented at boundaries between pixels. In contrast, the second set of basis functions are all non-zero (other than at isolated points) throughout the domain: This is a global basis, and it can only be used to construct continuous functions. In this set-up, both sets of basis functions have a similar minimum scale length.

We construct a synthetic model:

$$m_s(x, y) = \sum_i m_i^{(p)}\phi_i^{(p)}(x, y) + \sum_i m_i^{(c)}\phi_i^{(c)}(x, y) \tag{13}$$

as shown in Fig. 1, with 20 randomly selected non-zero coefficients $m_i^{(c)}$ forming a smooth background, and 20 further non-zero coefficients $m_i^{(p)}$ chosen to simulate a target structure. We therefore have 40 non-zero coefficients within a parametrization that admits 800 model parameters (Fig. 3a). We compute synthetic sinograms according to equation (10) for 14 400 paths, derived from a randomly chosen distribution of sources and receivers, adding normally distributed random noise with zero mean and a standard deviation chosen to be 1.5 per cent of the absolute maximum sinogram amplitude.

Using this set of 14 400 simulated observables, we can select smaller subsets of data. Using a subset of 300 randomly selected ray paths, we conduct the experiments already presented in Fig. 1. As discussed above, we first pose and solve the inverse problem using (only) the 400 pixel basis functions (Fig. 1c) and, separately, the 400 cosine basis functions (Fig. 1e). In both cases, our objective function is as given in equation (4), with χ_m as in equation (5); we employ the solution algorithm of Valentine & Sambridge (2018) to automatically determine the regularization weight, α. The features in the resulting images can clearly be linked to those present in the 'true' structure (Fig. 1b), but most detail is lost, with no clear distinction between background and anomalies. Both images inherit characteristics from their basis set: The pixel image is blocky and discontinuous, while the cosine-derived image is smooth and blurry. Short-wavelength artefacts in the cosine-derived image are likely to be caused by Gibbs' phenomenon, with the continuous basis functions unable to represent the sharp discontinuities in our input structure.

Next, we perform inversion in the overcomplete basis, comprising a total of 800 basis functions (400 pixel; 400 cosine); we continue to use a Tikhonov-style regularization term, defined as

$$\chi_\mathrm{m}(\mathbf{m}) = \alpha \sum_{k=1}^{K} \beta_k \left\| \mathbf{C}_\mathrm{d}^{-1/2}\mathbf{G}^{(k)} \right\|_2 \left\| \mathbf{m}^{(k)} \right\|_2^2 \quad \text{s.t.} \quad \sum_{k=1}^{K} \beta_k = 1 \tag{14}$$

in analogy with equation (9); we will refer to this as 'Tikhonov-regularised overcomplete tomography (TROT)'. We assign equal weight to both basis subsets, choosing $\beta^{(p)} = \beta^{(c)} = \frac{1}{2}$, and adopt $\alpha = 10^{-4}$. We solve this by direct minimization of the objective function (equation 4), using the L-BFGS algorithm (Liu & Nocedal 1989; Virtanen et al. 2020), and obtain results as shown in Fig. 1d. These are rather similar to those obtained using just the pixel basis: It continues to be difficult to convincingly interpret the image produced.

Finally, we repeat the overcomplete tomography, but instead use the L_1-based regularization function of equation (9). Again, we use $\beta^{(p)} = \beta^{(c)} = \frac{1}{2}$, and $\alpha = 10^{-4}$. Results are markedly improved, both qualitatively and quantitatively: SPOT can achieve good separation of features across basis sets, and the image could readily be interpreted and analysed.

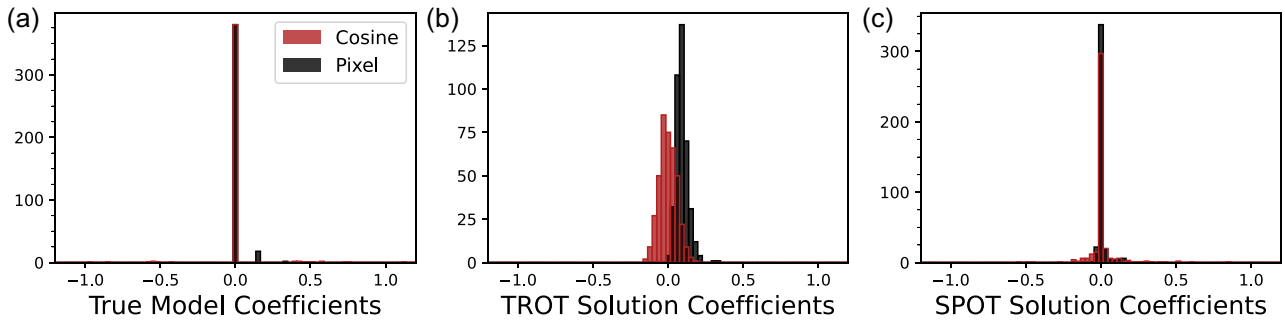

Figure 3 Histograms of model coefficients from the tomographic experiment in Fig. 1. (a) Shows the true model coefficients in Fig. 1b; (b) shows the TROT model coefficients in Fig. 1d; and (c) shows the SPOT model coefficients from Fig. 1f. Pixel coefficients are shown in black and cosine in red. Notice how SPOT produces few non-zero coefficients and is closer to the truth, whereas TROT has many non-zero coefficients. The constant background coefficient is omitted from each histogram.

To provide a clearer illustration of the 'sparsity-promoting' nature of our method, and justify this choice of name, Fig. 3 shows histograms of the model coefficient values for true model, and for the solutions obtained using the Tikhonov and sparsity-promoting methods. By construction, the target model is sparse, with only 5 per cent of coefficients non-zero. However, the Tikhonov-regularized solution is not sparse: As is well-known, regularization based on an L_2 norm smears the solution across all available model coefficients. In contrast, regularization using the L_1 norm results in only a small number of coefficients being significantly non-zero.

A key property of compressive sensing is the ability to accurately recover signals sampled at only a small number of points. To explore whether this result translates to overcomplete tomography, we conduct a series of experiments with varying numbers of rays. From our set of 14 400 simulated observations, we initially select 10 ray paths and perform two inversions, both in the overcomplete basis: one using TROT and the other using SPOT. This yields two models; we evaluate each by computing data misfit across all 14 400 original ray paths and plot these as points on Fig. 4. The measure of data misfit used is the objective function defined in equation (4), with the regularization function χ_{m} omitted. This provides a measure of the ability of the recovered model to give accurate predictions for even unseen paths. We then repeat the process, systematically adding more rays into the data set used for inversion.

Unsurprisingly, Fig. 4 shows that increasing data set size leads to better inversion results. However, it also reveals that results of SPOT have systematically better predictive power than those obtained using TROT. This is most marked for the smallest data set sizes, with sparsity-based results for a data set of 20 rays having predictive performance comparable to a TROT data set of around 120 rays. Plotting the recovered images for selected data sets, we see a similar story; even with comparatively small data sets, the sparsity-regularized tomography is able to produce informative, interpretable images, whereas the TROT results appear dominated by artefacts. With sufficiently large data sets, TROT also performs well, but SPOT is superior if data are limited.

3.2 Real data sets

To further illustrate this, we explore performance on two real X-ray tomography data sets: one based on imaging of a lotus-root (Bubba et al. 2016, containing 49 248 ray paths), and the other depicting a walnut (Hämäläinen et al. 2015, 19 680 ray paths). Together, these demonstrate a wide range of features common to real structures, including sharp contrasts and geometric complexity. Again, we select random subsets of the available data and perform sparsity-based tomography in an overcomplete basis. For the lotus root, our parametrization involves a total of 32 768 basis functions (16 384 pixel; 16 384 cosine); for the walnut, we have 53 792 basis functions (26 896 pixel; 26 896 cosine). We compare our SPOT results to a standard Tikhonov-regularized least squares in a pixel basis (only), using 16 384 basis functions for the lotus root, and 26 896 for the walnut.

The resulting images are shown in Fig. 5, and again we see that sparsity-regularized overcomplete tomography can recover useful interpretable images using a data set far smaller than that needed for robust standard Tikhonov-derived results. Even with the smallest data sets, SPOT is able to reveal the first-order features of the target objects. In particular, our approach is able to resolve the gross features of the lotus root, including some of the high-contrast features (which represent various objects inserted into the lotus root prior to imaging), using only 500 ray paths, as well as the overall shape of the walnut using only 400 ray paths. In contrast, no coherent object can be recognized within the standard least-squares results. Nevertheless, with sufficient data, both methods can obtain a reasonable image.

4 DISCUSSION

This paper is underpinned by two complementary ideas. First, we suggest that it is not necessary, nor necessarily desirable, to pose imaging problems in terms of a single family of basis functions. Second, we demonstrate that seeking sparse solutions can be an effective means of regularizing inversions in these 'overcomplete' settings, allowing good-quality results to be obtained from even small data sets. This is potentially powerful in settings such as geophysical imaging, where data collection is expensive and time-consuming, and where researchers may have limited control over acquisition geometries or source characteristics. It may also offer benefits in other fields, such as medical imaging, where exhaustive sampling of the target is currently routine: The ability to obtain comparable results using a smaller data set might offer tangible benefits such as faster scans and lower radiation exposure for patients.

4.1 Compressive sensing

Our approach is built on the intuition that physical systems tend towards simplicity – the principle of Occam's razor – but this simplicity may not conform to the rules of any one mathematical framework. In particular, real systems may arise from the interaction between multiple processes occurring at different scales and with

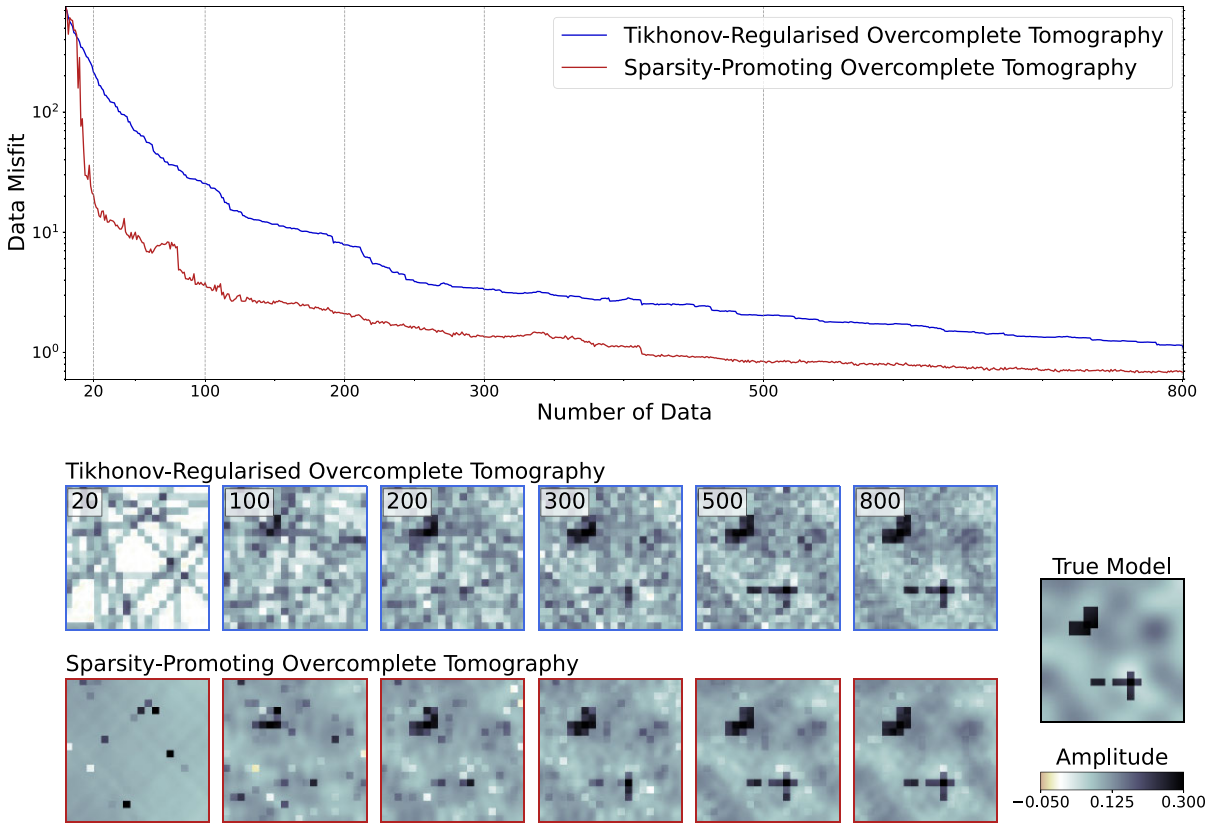

Figure 4 Data misfit as a function of the number of ray paths used in the inversion. SPOT misfit reduction is shown in red, and TROT in blue. In the lower panels, selected solutions for both cases are shown with increasing numbers of ray paths.

distinct characteristics. Each process may, individually, admit simple representation in some basis, but this need not be the same basis in each case. As a result, the combination of all processes may appear complex and difficult to represent relative to any single system. This motivates our use of an overcomplete basis, and we adopt the concept of 'sparsity' – only a few non-zero model parameters – as a proxy for 'simplicity'.

We therefore seek to construct the model that uses the least number of basis vectors necessary to explain the available data. Ideally, this would require us to regularize our inversion based on the L_0 norm of the model vector – that is, χ_m would simply count the number of non-zero model coefficients. However, the minimization of an objective with this form is known to be computationally intractable in realistic-scale problems. Instead, we follow Candès et al. (2006), who show that – with high probability – this would yield the same result as is obtained when we minimize an objective that depends upon the L_1 norm of the model vector. This is the same result that underpins compressive sensing; the key difference in our work is its use in combination with an overcomplete basis.

A characteristic feature seen in compressive sensing studies (e.g. Candes et al. 2008) is a sudden transition between 'poor' and 'excellent' signal recovery once the data set size passes a certain threshold. In Fig. 4, we saw a continual improvement in performance as the data set grows, but not the sharp 'cliff' that might be anticipated. However, if we perform an identical experiment using noise-free data, we obtain results as depicted in Fig. 6, with the predictive power of the sparsity-derived model abruptly improving by three orders of magnitude once the data set contains around 300 rays – a pattern that is also evident from the models themselves.

The absence of this cliff from Fig. 4 is probably attributable to the fact that we have not adhered strictly to the formal conditions necessary for compressive sensing to succeed. A central requirement in compressive sensing is that the signal be sampled in a manner that is incoherent to the basis in which it is sparse (Donoho & Elad 2003; Candes & Romberg 2007): In other words, each piece of data must carry information about many model components. However, while each of our rays is sensitive to the complete set of cosine model coefficients, each only samples a fraction of the pixel parameters. Given that our noise is not correlated between spatially adjacent measurements, it can easily be accommodated within the pixel part of our model – leading to a slight weakening of predictive performance. This could perhaps be mitigated through additional terms in the regularization function, although the details are likely to be problem-dependent; it may also suggest an alternative application of overcomplete inversion, for denoising of data sets or images.

4.2 Computational considerations

One major advantage of Tikhonov-style regularization is the resulting objective function can be minimized analytically, via a Gauss-Newton method. This is computationally efficient, allowing solutions to be obtained cheaply and rapidly. Unfortunately, the objective function underpinning our sparsity-promoting scheme (equation 9) cannot be minimized analytically, and instead some iterative optimization scheme must be used. For the examples presented here, we have employed the L-BFGS algorithm (Liu & Nocedal 1989; Virtanen et al. 2020), but a range of other approaches could also

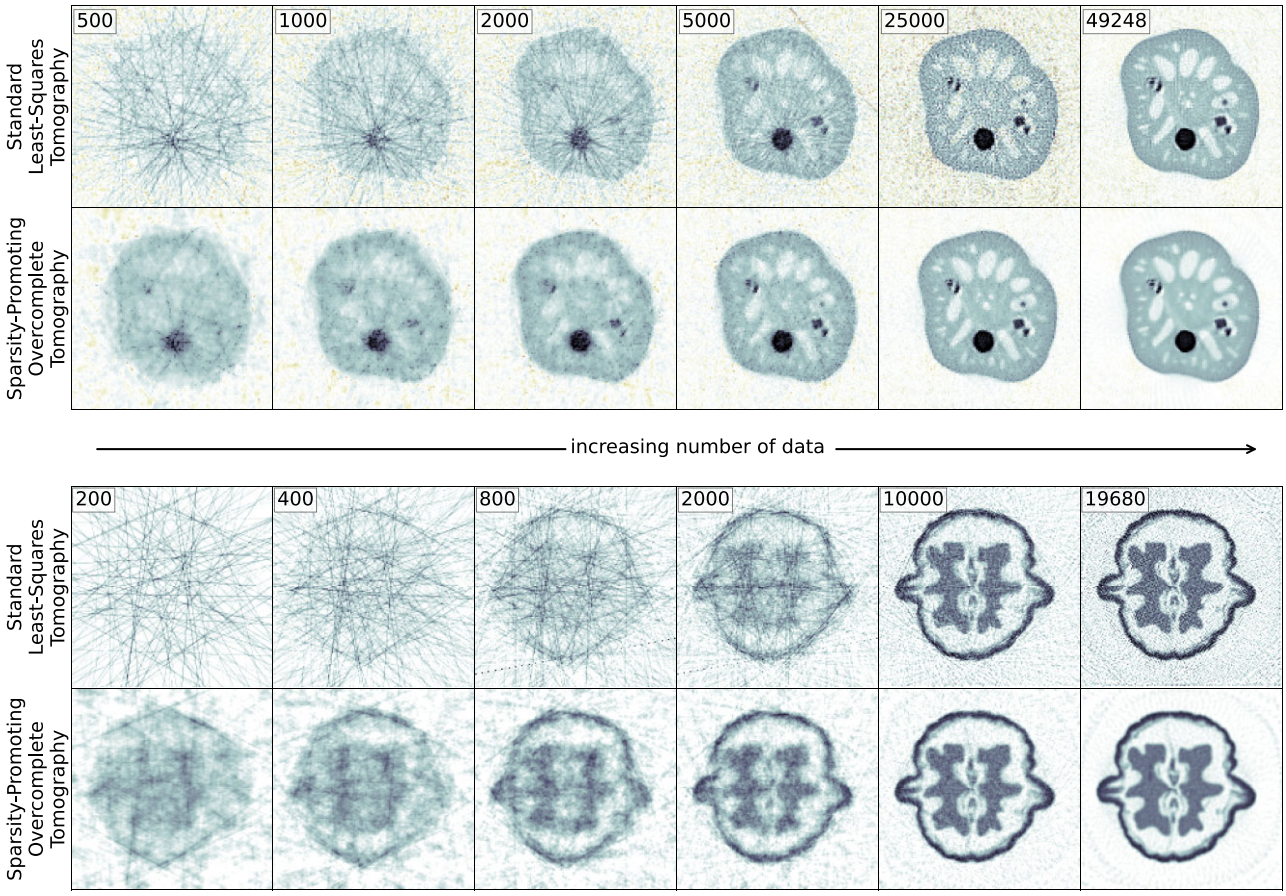

Figure 5 Reconstructed tomographic images as a function of the number of X-rays used in the real data experiment. The 1st and 3rd rows show the evolution of the recovered images from the TROT approach, while the second and fourth rows demonstrate the SPOT approach for the lotus and walnut cases, respectively. The number of data used in each case are displayed in the inset panel. These correspond to approximately 1 per cent, 2 per cent, 4 per cent, 10 per cent, 50 per cent, and 100 per cent of the available data set.

be explored. In particular, techniques such as the Lasso (Tibshirani 1996) and least angle regression (Efron et al. 2004) have been specifically designed to find sparse solutions to optimization problems, and quadratic programming approaches can also be employed (e.g. Solntsev et al. 2015). The suitability of these for use in SPOT remains to be fully explored. As an indication, solving the synthetic problem of Fig. 1 using our implementation of the SPOT approach takes 20.5 CPU-s, compared with 0.3 CPU-s for the TROT method and 0.2 CPU-s for a standard least-squares (Gauss-Newton) inversion.

As formulated in Section 2.2, implementation of SPOT requires the ability to compute $\mathbf{G}^{(k)}$ with respect to each of the basis sets that are to be employed. In some cases, this may present a practical barrier to adoption of the approach: Often, the numerical software available to compute these matrices will be designed to support only one specific basis. Similarly, any software that takes $m(\mathbf{x})$ as an input – perhaps to simulate observable phenomena – may expect the model to be expressed in a certain way. In the long term, it may be possible to develop codes that provide direct support for computations using overcomplete models. Alternatively, one can choose to express all basis functions $\phi_i^{(k)}$ relative to some reference basis $\{\psi_1, \psi_2, \ldots, \psi_{M_\psi}\}$:

$$\phi_i^{(k)}(\mathbf{x}) \approx \sum_{j=1}^{M_\psi} W_{ij}^{(k)} \psi_j(\mathbf{x}) \tag{15}$$

with, typically, $M_\psi \gg \max_k M_k$ to ensure that all $\phi_i^{(k)}$ can be approximated to a satisfactory level of accuracy. Then, it is straightforward to show that $G_{ij}^{(k)} \approx \sum_{l=1}^{M_\psi} W_{jl}^{(k)} G_{il}^{(\psi)}$, with $G_{il}^{(\psi)} = \int_{\mathcal{X}} g_i(\mathbf{x}) \psi_l(\mathbf{x}) \, \mathrm{d}^N\mathbf{x}$. Thus, quantities can be computed in the reference basis and then mapped into each of the overcomplete basis sets, allowing the SPOT approach to be employed.

4.3 Choosing basis families

For the examples presented in this paper, we have relied upon two families of basis functions: a pixel basis, and a discrete cosine basis. These were chosen for their conceptual simplicity, and for their clearly distinct properties: local versus global; discontinuous versus continuous. However, the method can be applied with any collection of basis vectors, and with more than two distinct families. The particular families used in any given situation are likely to be problem-dependent, informed by prior knowledge of the characteristics and features that should be present or are important to resolve. However, we believe that in many scenarios there will be particular power in mixing globally and locally supported basis functions: The former are able to bridge across gaps in data coverage, while the latter ensure that detailed features can be accurately reproduced. The trade-off parameters α and β_k govern the priority assigned to each basis set, and

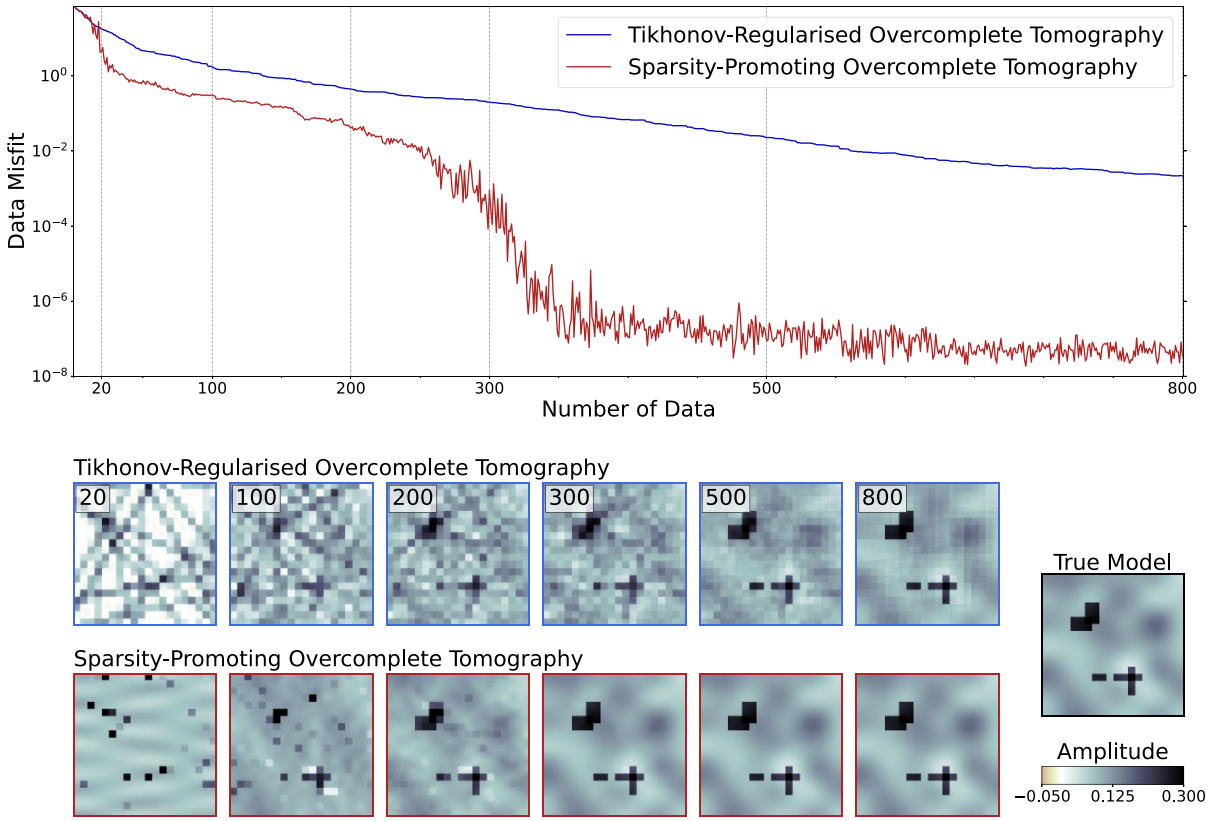

Figure 6 Results from a comparison of TROT and SPOT solutions for the synthetic experiment shown in Fig. 4 but for noiseless data. Notice that the SPOT solutions experience a three orders of magnitude reduction in misfit, as the number of ray paths increases beyond about 300. At the same location the true image (Fig. 1b) is recovered almost perfectly. This thresholding behaviour is consistent with that seen in experiments of compressed sensing.

their impact may need to be explored in the context of any particular problem.

5 CONCLUDING REMARKS

This paper has developed a new approach to model parametrization and regularization of inverse problems, which we refer to as 'overcomplete tomography'. This allows models to be expressed using a combination of basis functions from multiple distinct families, with regularization based on minimising the L_1 norm of the model coefficients. This enables the recovery of high-resolution images from comparatively small data sets, and allows us to avoid imposing arbitrary characteristics upon the solution.

We have demonstrated the efficacy of this approach using a number of examples. In particular, we have applied the method to a pair of real-world X-ray tomography data sets. We see that results are comparable to existing methods when the available data provides exhaustive sampling of the target, with our new method outperforming existing approaches when data are limited. We therefore foresee a range of applications in scenarios where data sets are intrinsically limited (e.g. geophysical imaging studies, where there may be little control over the distributions of sources and/or receivers), or where it is desirable to minimize the number of samples collected (e.g. to restrict acquisition time and cost, or to limit radiation exposure in medical settings). While our work has focussed exclusively on the case where data is linearly dependent on model parameters, we anticipate that the method can be employed iteratively for the solution of weakly non-linear problems.

As a result, we see overcomplete tomography as a powerful new methodology with applications in a range of settings such as next-generation seismic and electromagnetic imaging of subsurface anomalies beneath a volcano or in mineral resource provinces; medical CT imaging to improve the contrast of target features against background variations or other laboratory settings where tomographic approaches are used to detect anomalous structures within a target body. The basis functions must be chosen commensurately in all applications with the class of anomalies sought. With this new class of inference approach, we may soon see a significant improvement in resolving multiple structure classes within tomographic images, overcoming long-standing difficulties in this field.

ACKNOWLEDGEMENTS

We thank Matthias Scheiter for his insightful comments on this paper. This research was undertaken with the financial support by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Deep Earth Imaging Future Science Platform PhD scholarship, the Australian National University (ANU), and the Australian Research Council (ARC) via DP200100053. We thank the editor Zefeng Li and anonymous reviewers for their constructive comments and suggestions, which have greatly enhanced the quality of this paper.

DATA AVAILABILITY

The code associated with this article is available at https://github.com/buseet/OvercompleteTomo.

REFERENCES

Bubba T. A., Hauptmann A., Huotari S., Rimpeläinen J., Siltanen S., 2016, preprint (arXiv:1609.07299)
Candès E., Romberg J., 2007, Inverse Probl., 23, 969
Candès E. J., Tao T., 2006, IEEE Trans. Inf. Theor., 52, 5406
Candès E. J., Romberg J., Tao T., 2006, IEEE Trans. Inf. Theor., 52, 489
Candès E. J., Wakin M. B., Boyd S. P., 2008, J. Fourier Anal. Appl., 14, 877
Donoho D. L., 2006, IEEE Trans. Inf. Theor., 52, 1289
Donoho D. L., Elad M., 2003, Proc. Natl. Acad. Sci., 100, 2197
Duffin R. J., Schaeffer A. C., 1952, Trans. Am. Math. Soc., 72, 341
Efron B., Hastie T., Johnstone I., Tibshirani R., 2004, Ann. Stat., 32, 407
Hämäläinen K., Harhanen L., Kallonen A., Kujanpää A., Niemi E., Siltanen S., 2015, preprint (arXiv:1502.04064)
Liu D. C., Nocedal J., 1989, Math. Prog., 45, 503
Ritsema J., Deuss a. A., Van Heijst H., Woodhouse J., 2011, Geophys. J. Int., 184, 1223
Simons F. J. et al., 2011, Geophys. J. Int., 187, 969
Solntsev S., Nocedal J., Byrd R. H., 2015, Opt. Methods Softw., 30, 1213
Spakman W., 1991, Geophys. J. Int., 107, 309
Tarantola A., 2005, Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM, Philadelphia, PA
Tibshirani R., 1996, J. R. Stat. Soc. B, 58, 267
Valentine A. P., Sambridge M., 2018, Geophys. J. Int., 215, 1003
Valentine A. P., Trampert J., 2016, Geophys. J. Int., 204, 59
Virtanen P. et al., 2020, Nature Methods, 17, 261
Van der Hilst R. D., Widiyantoro S., Engdahl E., 1997, Nature, 386, 578
Woodhouse J. H., Dziewonski A. M., 1984, J. Geophys. Res.: Solid Earth, 89, 5953

This paper has been typeset from a TEX/LATEX file prepared by the author.