PAPER • OPEN ACCESS

SparseBeads data: benchmarking sparsity-regularized computed tomography

View the article online for updates and enhancements.

SparseBeads data: benchmarking sparsity-regularized computed tomography

Jakob S Jørgensen[1,4] , **Sophia B Coban**[2,5], **William R B Lionheart**[2],
Samuel A McDonald[3] **and Philip J Withers**[3]

[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark,
Richard Petersens Plads, Building 324, 2800 Kgs. Lyngby, Denmark
[2] School of Mathematics, University of Manchester, Manchester, M13 9PL, United Kingdom
[3] School of Materials, University of Manchester, Manchester, M13 9PL, United Kingdom

E-mail: jakj@dtu.dk and jakob.jorgensen@manchester.ac.uk

Abstract

Sparsity regularization (SR) such as total variation (TV) minimization allows accurate image reconstruction in x-ray computed tomography (CT) from fewer projections than analytical methods. Exactly how few projections suffice and how this number may depend on the image remain poorly understood. Compressive sensing connects the critical number of projections to the image sparsity, but does not cover CT, however empirical results suggest a similar connection. The present work establishes for real CT data a connection between gradient sparsity and the sufficient number of projections for accurate TV-regularized reconstruction. A collection of 48 x-ray CT datasets called SparseBeads was designed for benchmarking SR reconstruction algorithms. Beadpacks comprising glass beads of five different sizes as well as mixtures were scanned in a micro-CT scanner to provide structured datasets with variable image sparsity levels, number of projections and noise levels to allow the systematic assessment of parameters affecting performance of SR reconstruction algorithms[6]. Using the SparseBeads data, TV-regularized reconstruction quality was assessed as a function of numbers of projections and gradient sparsity. The critical number of projections for satisfactory TV-regularized reconstruction increased almost linearly with the gradient sparsity. This establishes a quantitative guideline from which one may predict how few projections to acquire based on expected sample sparsity level as an aid in planning of dose- or time-critical experiments. The results are expected to hold for samples of similar characteristics, i.e. consisting of few, distinct phases with relatively simple structure. Such cases are plentiful in porous media, composite materials, foams, as well as non-destructive testing and metrology. For samples of other characteristics the proposed methodology may be used to investigate similar relations.

Keywords: total variation, regularization, imaging, optimization, filtered backprojection, open-access data

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

---

[4] Present address: School of Mathematics, University of Manchester,
Manchester, M13 9PL, United Kingdom
[5] Present address: Centrum Wiskunde and Informatica, Science Park 123,
1098 XG, Amsterdam, Netherlands
[6] The SparseBeads dataset is available from https://zenodo.org/record/290117

1. Introduction

X-ray computed tomography (CT) is a noninvasive 3D imaging technique used extensively in medical imaging as well as a host of non-medical applications in areas such as materials science, industrial non-destructive testing and metrology. From a large number of x-ray radiograph images taken at a range of sample orientations a detailed 3D model can be reconstructed and subsequently segmented, visualized and quantified to understand the structure and processes occurring within. Substantial efforts are ongoing to further develop x-ray CT for a host of new demanding applications [1], and improve reliability, e.g. in terms of standardization [2], calibration [3] and bias and variance characterization [4]. In particular, the development of laboratory-based micro-CT imaging systems has enabled unprecedented micrometer- and even nanometer- resolution imaging being applied much more widely, previously only being possible in large-scale x-ray synchrotron facilities with restricted access. However, laboratory-based micro-CT remains slow with scan times on the order of hours, which presents a barrier to studying processes which occur on a faster time scale and to higher throughput in industrial imaging applications. An alternative for fast scans (seconds) is provided by real-time tomography (RTT) x-ray CT systems developed for airport baggage scanning [5] which have been explored for production inspection [6], however with resolution only on the millimeter scale. Naturally, accelerating high-resolution micro-CT is an active research venue with better sources and detectors coming to the market. A particularly promising direction for faster micro-CT occurs on the computational front, where new image reconstruction algorithms capable of accurate imaging from much fewer projections are being developed and approaching maturity for practical use.

X-ray CT still predominantly uses image reconstruction methods of the analytical type, i.e. filtered back-projection (FBP) for parallel-beam geometry and the Feldkamp–Davis–Kress (FDK) algorithm for cone-beam geometry [7]. In recent years novel reconstruction methods exploiting sparsity regularization (SR) based on an algebraic imaging model have been subject to extensive research. Unlike analytical methods which require a large number of projections independently of the image complexity, SR reconstruction can preserve or even improve image quality from substantially fewer projections [8–13] assuming the image or a suitable transform thereof is sparse, i.e. has few non-zero components. Reconstruction from few projections is of high interest due to its potential to reduce patient x-ray exposure in medical imaging and shorter acquisition times to capture rapid material changes or achieve fast throughput in non-destructive testing. The prototypical SR reconstruction method is total variation (TV) regularization [14], which has the ability to reduce noise and sub-sampling artifacts while preserving sharp edges in the reconstruction. TV-regularization can be interpreted as an SR method because it encourages a sparse gradient image of the reconstruction.

As mentioned, empirical results indicate that TV-regularization can produce high-quality reconstructions of gradient-sparse images from fewer projections than is possible using FBP. However, it remains poorly understood as to how few projections suffice and how sparse the images can be. One goal of the present work is exactly to clarify this point. As will be discussed further in section 3, we seek to establish empirically—using real x-ray CT data—the existence of a critical number of projections required for accurate TV-regularized reconstruction and quantify how this critical number depends on the image sparsity level.

This particular goal can be seen in a broader perspective. The current literature contains a multitude of proposed reconstruction methods and algorithms, many of which are demonstrated only on simple synthetic test problems or at most on a few selected real-data cases. Mathematical phantoms such as the ubiquitous Shepp–Logan phantom [15] and the FORBILD phantoms [16] serve an important purpose of testing ideal cases in controlled simulations. This is important to verify correctness of algorithms and implementations and to provide fundamental understanding of reconstruction behavior on known test images. On the other hand, more realistic and practically relevant tests naturally involve real datasets, but application-specific datasets can be very complex and are typically relevant only for determining if a new reconstruction method helps improve the specific clinical or industrial imaging task considered. From such studies it is difficult to deduce general behaviors and systematic performance dependencies on parameters such as image features (e.g. sparsity level), number of projections and noise level.

While many reconstruction algorithm developers have access to datasets through their facility, few of these are made openly available, which is unfortunate for developers without direct access to data or those wanting to make side-by-side comparisons. Furthermore, algorithm results are often published without supplying the dataset(s) used. This makes it difficult to assess the reproducibility of results and leads to a lack of clarity about how various proposed reconstruction methods compare. This motivates the need for openly available datasets specifically designed for benchmarking purposes.

Some open datasets for testing purposes are available in the literature including 'Tomographic x-ray data of a walnut' [17], 'Tomographic x-ray data of a lotus root filled with attenuating objects' [18] and 'The SophiaBeads Dataset Project' [19]. There are also databases with CT images, mainly from medical imaging, including the the visible human project[7] and radiopaedia.org[8]. From such images it is possible to simulate CT data for reconstruction testing purposes, which is advantageous relative to using simplistic standard phantoms such as the Shepp–Logan phantom, but critically actual measured raw data are not available. This limits their use as realistic tests of reconstruction algorithms. While the openly available datasets are immensely useful for testing on single problems, they are of limited use for large systematic tests of the central parameters for SR reconstruction, namely the the number of projections, the sparsity level as well as the noise level.

We argue that there is a need for something in between synthetic phantom data and isolated application-specific

---

[7] www.nlm.nih.gov/research/visible/

[8] http://radiopaedia.org/

datasets in order to assess advanced reconstruction methods more systematically and realistically. What is needed is a carefully designed set of real-data test problems with systematically varying properties across individual datasets that can serve as a benchmark for testing reconstruction methods. To this end we provide in this work a new, openly available, dataset that we refer to as SparseBeads. It consists of 48 individual 2D fan-beam x-ray CT datasets designed specifically to enable systematic comparisons of SR reconstruction algorithms. Glass beads of five different sizes were chosen as a suitably well-defined class of test objects to produce a range of sparsity levels. The smooth spherical shape makes it easy to verify reconstruction correctness, while the space between spheres offers more complex structure. This class of test objects is sufficiently simple to isolate the effect of sparsity, yet complex enough to be used as a practical model of, e.g. porous media. We do not claim that the SparseBeads dataset will offer complete answers but it may serve as one standardized way of assessing reconstruction algorithms.

Thus, this paper presents two contributions. First, in section 2, the new SparseBeads dataset is presented along with design considerations and suggestions of how this dataset may be of wide utility for testing SR reconstruction algorithms. The remainder of the paper describes a study seeking to establish a connection between the image sparsity level and the number of projections sufficient for TV-regularized reconstruction using a subset of the aforementioned SparseBeads dataset. Section 3 introduces the study and puts it in the context of previous work. Section 4 gives details of the computational setup we employ including TV-regularized reconstruction, optimization aspects, as well as image quality measures and determination of reference images and sparsity levels. Section 5 presents our results providing extensive empirical evidence that the required number of projections for accurate TV-regularized reconstruction depends approximately linearly on the gradient image sparsity. Section 6 offers a discussion before the work is concluded in section 7.

The novelty of the dataset lies in its design to allow systematic assessment of the dependence of SR reconstruction quality on factors such as image sparsity and number of projections. The subsequent study can be seen as an example hereof. We believe the dataset and the proposed methodology will be a significant step toward establishing a firm practical understanding of how and by how much SR methods can reduce the measurements required for accurate reconstruction. This will, for example, help accelerate micro-CT to enable imaging of fast microstructure evolution.

2. The SparseBeads dataset collection

2.1. Design considerations

The SparseBeads dataset collection is inspired by the SophiaBeads dataset project [19, 20], which is a collection of six cone-beam x-ray CT datasets designed specifically for testing reconstruction algorithms. In the SophiaBeads dataset the number of projections is varied while the total photon count (or the total exposure time) is kept constant. The sample

comprises glass beads and air, packed in a plastic container. The simplicity of having a few discrete materials and the fact that any cross section through the sample will be piecewise constant makes this dataset of interest for examining the capabilities of SR techniques, in particular TV-regularization. In [21] we carried out a preliminary study using the SophiaBeads dataset with the goal of assessing how the number of projections sufficient for accurate TV-regularized reconstruction depends on the image gradient sparsity. However, we encountered the following limitations:

(i) Limited sub-sampled datasets available, only 2048-, 1024-, 512-, 256-, 128-, and 64-projection datasets available providing only a few levels of sampling. It is possible to extract data subsets, e.g. from the 2048-projection dataset, but due to 2048 being a power of 2, only the already included smaller powers of 2 yield perfectly equiangular datasets.

(ii) Limited sparsity levels available—the gradient sparsity level depends on the bead size and only a single bead size is present in the SophiaBeads Dataset. In [21] an additional unpublished dataset with smaller-sized beads was included to yield different sparsity levels, but having only 1 or 2 sparsity levels severely limits the utility for assessing SR reconstruction methods.

(iii) Limited instances of test images available—only data for the same set of beads is provided. To assess reconstruction methods more thoroughly than on single test images, it is desirable to have multiple test images with similar characteristics, such as multiple different images of beads of the same size.

(iv) Difficulty in isolating the effect of decreasing number of projections due to variable noise levels between datasets. While the fixed total-dose design of the SophiaBeads Dataset is appropriate for the intended type of trade-off studies, datasets with fixed exposure time *per projection* would allow studies of the effect solely of the number of projections.

Nevertheless, the key aspects of the SophiaBeads dataset, namely the spherically-shaped beads ideal for TV-regularized reconstruction combined with a range of different numbers of projections, remain desirable for our purposes. Therefore, motivated by the SophiaBeads dataset, we designed and performed a new set of experiments, which we refer to as SparseBeads, specifically targeted at assessing SR reconstruction at reduced numbers of projections. In designing the new SparseBeads dataset collection we addressed the four limitations above.

First, to provide a larger range of equiangular sub-sampled datasets the total number of projections was chosen as a *highly composite number*, i.e. a positive integer with more divisors than any smaller integer. In this way the large number of divisors will allow for many perfectly equiangular subsets by discarding all but every *d*th projection for any divisor *d*. In the present case, covering a full 360° rotation, the highly composite number 2520 was chosen, which by having 48 divisors allows for the same number of subsets, see table 1.

Table 1. Using the *highly composite number* 2520 (with more divisors than any smaller integer) as total number of projections admits a large number of perfectly equiangular sub-sampled datasets by discarding all but every *d*th projection for a sub-sampling factor of *d*. Shown here are all divisors of 2520, i.e. possible sub-sampling factors and the resulting number of projections.

Sub-sampling factor	1	2	3	4	5	6	7	8	9	10	12	14
No. projections	2520	1260	840	630	504	420	360	315	280	252	210	180
Sub-sampling factor	15	18	20	21	24	28	30	35	36	40	42	45
No. projections	168	140	126	120	105	90	84	72	70	63	60	56
Sub-sampling factor	56	60	63	70	72	84	90	105	120	126	140	168
No. projections	45	42	40	36	35	30	28	24	21	20	18	15
Sub-sampling factor	180	210	252	280	315	360	420	504	630	840	1260	2520
No. projections	14	12	10	9	8	7	6	5	4	3	2	1

Table 2. Dataset references of the 48 individual datasets contained in SparseBeads. Five different bead sizes were used giving five monodisperse beadpacks (B1–B5). The two smallest sizes given as 0.5 mm and 0.25 mm had sizes in the ranges 0.425–0.600 mm and 0.212–0.300 mm, respectively. Mixtures of two bead sizes are given as B12 for example of B1 and B2 mixture of equal parts. ML refers to a mixture of all five sizes. For each sample between one and five different vertical sample positions were used as given by L1–L5. Prefix of 'Low_' means the short exposure time of 134 ms per projection was used, while no prefix means 1000 ms.

Expos. (ms)	Size (mm)	Dataset reference (size B1–B45/ML; sample layer L1–L5)				
	5	B1L1	B1L2	B1L3		
	3	B2L1	B2L2	B2L3		
	1	B3L1	B3L2	B3L3		
	0.5	B4L1	B4L2	B4L3		
1000	0.25	B5L1	B5L2	B5L3		
	5 & 3	B12_L1				
	3 & 1	B23_L1				
	1 & 0.5	B34_L1				
	0.5 & 0.25	B45_L1				
	All	ML_L1	ML_L2	ML_L3	ML_L4	ML_L5
	5	Low_B1L1	Low_B1L2	Low_B1L3		
	3	Low_B2L1	Low_B2L2	Low_B2L3		
	1	Low_B3L1	Low_B3L2	Low_B3L3		
	0.5	Low_B4L1	Low_B4L2	Low_B4L3		
134	0.25	Low_B5L1	Low_B5L2	Low_B5L3		
	5 & 3	Low_B12_L1				
	3 & 1	Low_B23_L1				
	1 & 0.5	Low_B34_L1				
	0.5 & 0.25	Low_B45_L1				
	All	Low_ML_L1	Low_ML_L2	Low_ML_L3	Low_ML_L4	Low_ML_L5

Second, to achieve more sparsity levels we employed five different sizes of beads as specified in table 2. A wider range still was considered, but larger beads would not have packed properly in the container used, while smaller ones would have approached the detector resolution on the micro-CT instrument selected for the experiment. To achieve more intermediate sparsity levels several bead mixtures, see table 2, were used in addition to single-size beads only.

Third, to provide multiple similar images we scanned each sample at different vertical positions meaning that the central slice, which is the only part of the data of interest to us, samples different parts of the same beadpack, thus yielding different 2D images. This results in different images of same-sized beads, thus providing test images having approximately the same characteristics such as gradient sparsity level.

Finally, to isolate the effect of reducing the number of projections as well as enabling studies of the influence of the noise level in data, we recorded all datasets in both a high- and a low-exposure time versions. If further acquisition time had been available it would have been desirable to include more both longer and shorter exposure times to provide an even greater range of data quality.

2.2. Experimental setup

The SparseBeads datasets were collected over 4 days' usage of the 320/225 kV Nikon XTEK Bay micro-CT instrument, located at the Henry Moseley X-ray Imaging Facility (HMXIF), the University of Manchester. The apparatus consists of a cone-beam micro-focus tungsten-target x-ray source that projects polychromatic x-rays onto a 2000 × 2000-pixel, 16-bit, flat detector panel with pixel size of 0.2 mm × 0.2 mm. The source was operated at a voltage of 100 kV and current of 100 μA and the source-to-center and source-to-detector distances were 122 mm and 1400 mm, respectively. A total of
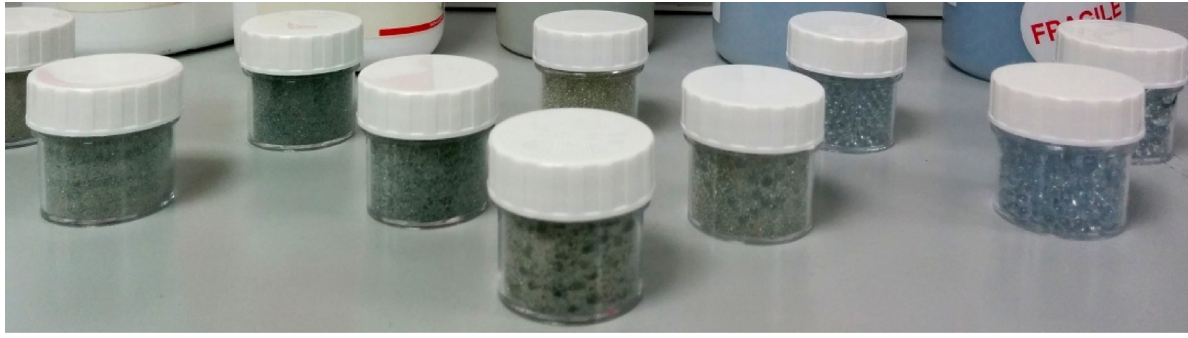
Figure 1. The 10 SparseBeads samples consisting of plastic containers packed with glass beads; five containers with beads only of a single size (back, increasing bead size left to right), four with mixtures of two sizes (middle, two nearest bead sizes mixed) and one with mixture of all five sizes (front).
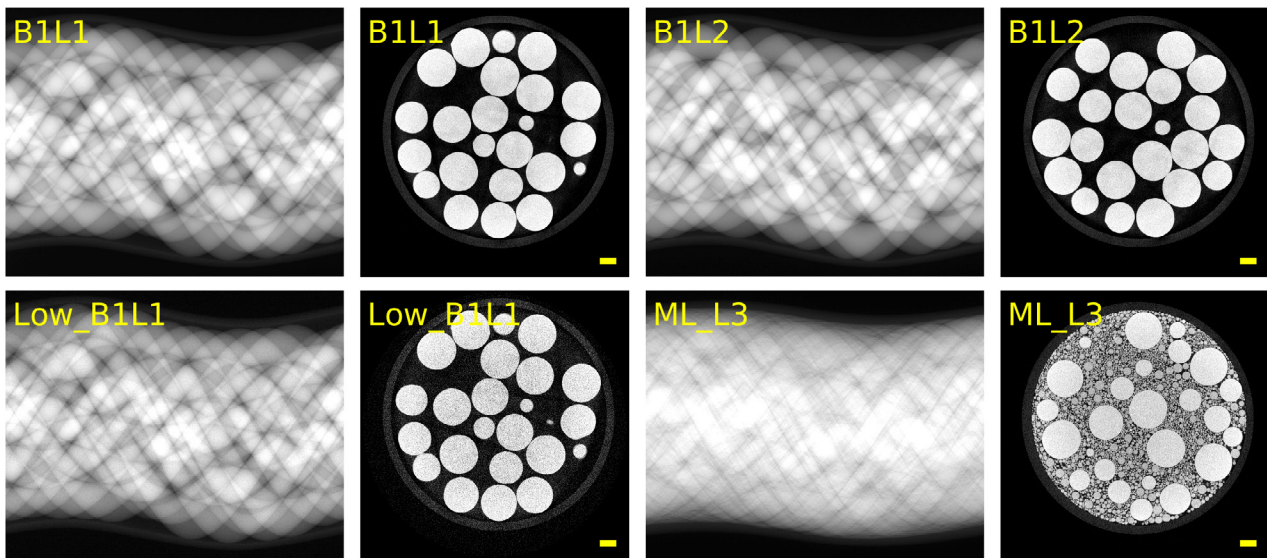

Figure 2. SparseBeads sinogram and vendor reconstruction examples. Top left, B1L1 sinogram and reconstruction. Top right, same for B1L2, i.e. same bead size, different layer. Bottom left, same for Low_B1L1, i.e. noisier sinogram and resulting lower-quality reconstruction. Bottom right, same for ML_L3, i.e. a mixture of all five bead sizes. Sinograms after normalization and negative log transform. Scalebar: 2 mm.

2520 equiangular projections over 360° were acquired with two choices of exposure time per projection, 1000 ms and 134 ms, to obtain low- and high-noise datasets. Projections were recorded using the so-called 'stop-start mode', in which the sample is stationary for acquisition between angular increments, to avoid any angular blurring resulting from operating in the faster 'continuous mode', where the sample is continuously rotated. To minimize possible beam-hardening artifacts, a 0.1 mm copper filter was used.

The samples were constructed using five sizes of uniform Soda-Lime glass (SiO_2-Na_2O) beads[9], see table 2 for sizes and mixtures used. The beads were packed in cylindrical 15 mL plastic (polystyrene) containers, 33 mm tall and with outer diameter 35 mm and closed with a screw cap, see figure 1. For bead-size mixtures equal volumes were measured and pre-mixed before packing. Samples were scanned at 1, 3 or 5 separate vertical positions.

2.3. Contents, availability and usage of the dataset collection

After data acquisition a central-slice reconstruction was produced using the Nikon XTEK CTPro software version 2.2 (presumably using the FDK algorithm, however not specified by the proprietary software). A center-of-rotation correction was applied as was a mild noise reduction (CTPro noise reduction level 2 of six possible levels—interpreted to mean the amount of filtering applied in the FDK algorithm) to preserve spatial resolution while reducing noise somewhat. The 2000×2000-pixel central-slice reconstruction (pixel size 17.4 μm) and the 2520×2000-pixel measured sinogram data corresponding to the central slice were exported as a 32-bit float binary file and 16-bit TIFF image, respectively, along with meta-data files describing the scan configuration to form a single complete dataset. This process was repeated resulting in the 48 datasets summarized in table 2. Example sinograms and reconstruction are shown in figure 2.

The SparseBeads dataset collection is available [22] from the Zenodo data sharing platform. Individual datasets can be downloaded independently, as can the full collection (in total

---

[9] Lead-free Soda-Lime glass beads distributed by BioSpec Products, Inc., PO Box 788, Bartlesville, OK 74005, USA, www.biospec.com

ca. 1.3 GB). The full cone-beam datasets and volume reconstructions are not provided as part of the SparseBeads dataset collection due to their size exceeding several TB. The data is exported in the same format as the SophiaBeads Dataset [19], which means that the SophiaBeads Dataset Project Codes [23] can be used to load the reconstruction and sinogram data files as well as carry out a conjugate-gradient least-squares (CGLS) reconstruction. This is demonstrated in a demo MATLAB script SparseBeads.m accompanying the SparseBeads dataset collection.

In table 2 the provided datasets have been grouped according to size into six main groups: at each of the two exposure levels there are three groups, the first one being of single-size (monodisperse) beads labelled B1–B5; the second being mixtures of two bead sizes labelled B12, B23, B34 and B45; and the third being a mixture of all bead sizes labelled ML. For the monodisperse beadpacks three different realizations (sample scanned at different vertical positions or layers, thus L) labelled L1–L3, while the 2-bead size mixture have only a single repetition labelled L1, and the 5-bead size mixture has five labelled L1–L5. The long-exposure datasets in the top part are named by just the bead-size label and the layer label, e.g. B1L1, while the short-exposure datasets in the bottom half have an additional 'Low_' prefix.

As described previously the dataset collection is designed to allow a multitude of different experiments across systematically grouped datasets. In addition to the experiment described in the following sections we wish to briefly mention a few possible experiments enabled by the provided dataset collection:

(i) Very fine-grained studies of the effect of number of projections (preserving ideal equiangularity) can be undertaken due to the total number of projections chosen as the highly composite number 2520. Few-projection datasets can be used to study the performance of reconstruction algorithms as the number of projections is reduced.

(ii) Performance comparisons of reconstruction algorithms across several test images with same characteristics, for example approximately the same gradient sparsity level, using the various layers from the same bead size. This can yield some (limited) statistics and add more confidence in the results than running only on a single test image.

(iii) Comparisons of high- and low-exposure versions of the same datasets to assess performance of reconstruction algorithms at different noise levels on otherwise identical datasets.

(iv) Comparisons of reconstruction algorithms across multiple feature length scales using the 5-size beadpack. Some methods such as TV-regularization have a 'preferred' image feature scale below which details of finer scale tend to be treated as noise and smoothed away, with the scale governed by the choice of regularization parameter.

The remainder of the paper describes the experiment that motivated the acquisition of the SparseBeads dataset collection. The results establish for the first time using real CT data a relation between the image sparsity and the number of projections required for an accurate TV-regularized reconstruction. The specific questions addressed are given as **Q1** and **Q2**

in the following section. The experiment is interesting in its own right and also serves as an example of algorithm benchmarking using the SparseBeads dataset. Only a subset of the full dataset is used but the opportunity was taken to collect and provide, as open access, a larger collection of datasets for subsequent algorithm benchmarking by other researchers.

3. Critical sampling for TV-regularized reconstruction versus gradient sparsity

Research in SR reconstruction has in part been motivated by new mathematical results in the field of compressive sensing (CS). In a nutshell, CS establishes that accurate image reconstruction is possible from a substantially reduced number of measurements. This is tied to the sparsity of the image, under certain assumptions on the imaging process such as incoherence and the restricted isometry property [24, 25]. Unfortunately the recovery guarantees of CS do not extend to cover x-ray CT [26, 27] and the success of SR in CT therefore remains somewhat heuristic. In particular, little is known about conditions under which SR will succeed for tomographic data. In CS the sufficient number of measurements for accurate image reconstruction using SR methods depends directly on the the sparsity level of the image, i.e. the number of signal nonzero values or coefficients in some representation, see also [28, 29] for further details in relation to CT. It remains to be established whether a similar dependence holds in x-ray CT. In the absence of any useful theoretical guarantees in the CT setting thus far, an empirical approach may provide some initial insights.

Establishing even empirically a link between the image sparsity and a critical number of projections for accurate TV-regularized reconstruction may provide a quantitative guideline as to how few projections would suffice if an estimate of the sample gradient sparsity is known. This may lead to drastic reduction in dose and acquisition time thus enabling low-dose and ultra-fast scans without appreciable loss of image quality.

Previous work by the first author has addressed precisely this question in comprehensive studies with simulated data [28, 30–32]. These results indicate that under some circumstances an almost identical relation to what is seen in CS between the sparsity level and the critical number of measurements can be observed in x-ray CT. As explained in [28] a key CS result is the existence of phase-transition behavior for accurate reconstruction. Considering the imaging model $b = Ax$, where $b \in \mathbb{R}^M$ is the measured data, $A \in \mathbb{R}^{M \times N}$ is the measurement matrix and $x \in \mathbb{R}^N$ is the signal to be recovered which is assumed to have $s \leqslant N$ non-zeros. We refer to M as the sampling level and s as the sparsity level (where later in relation with TV-regularization the sparsity level is measured in the gradient-magnitude domain). The case of interest in the present work is the underdetermined case, i.e. where $M < N$, due to having few projections. For reconstruction L^1-norm minimization is used, that is, either

$$x^\star = \underset{x}{\arg\min} \|x\|_1 \quad \text{subject to} \quad Ax = b, \tag{1}$$
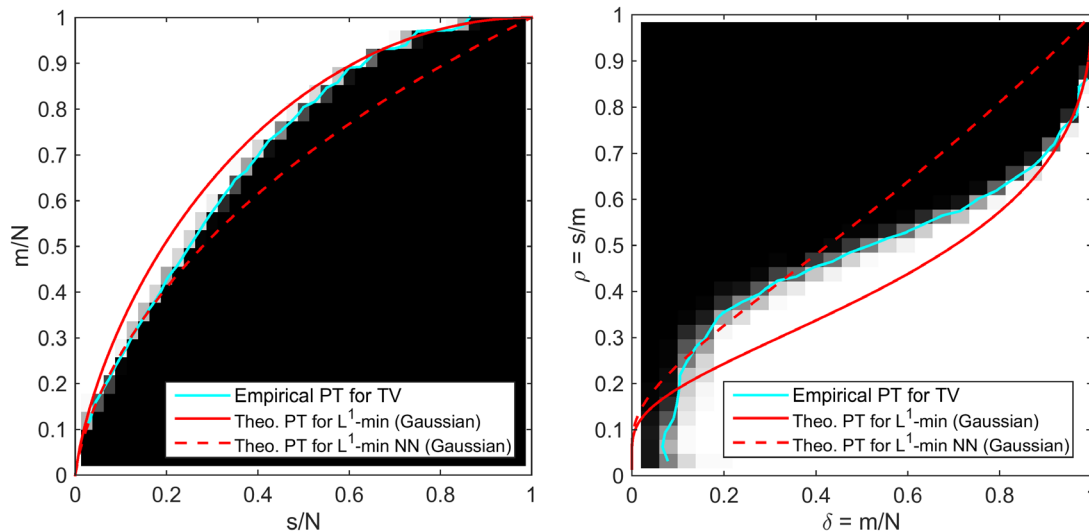
Figure 3. Empirical recovery phase diagrams with empirical and theoretical phase-transition (PT) curves overlaid (replotted from [28]). The gray-scale matrix plot shows empirical recovery rates for TV-minimization ranging from from 0% (black) to 100% (white) of all test instances as a function of sparsity and sampling levels for fan-beam CT measurements. Left is the Amelunxen–Lotz–McCoy–Tropp (ALMT) phase diagram and right is the Donoho–Tanner (DT) phase diagram. The full cyan curve is the empirically observed PT (50% of all instances recovered). The full/dashed red curve delineates the theoretical PT for Gaussian measurements and reconstruction by L^1-minimization without/with non-negativity (NN) constraint shown for reference. Note that full recovery occurs *above* the PT in the ALMT phase diagram and *below* the PT in the DT phase diagram.

or if entries of x are known to be non-negative

$$x^\star = \arg\min_x \|x\|_1 \quad \text{subject to} \quad Ax = b, \quad x \geqslant 0. \quad (2)$$

The question addressed is whether the true original image x_{orig}, from which the ideal data is generated $b = Ax_{\text{orig}}$, will be *perfectly recovered* through L^1-norm minimization. As shown by Donoho and Tanner, as summarized in [33], as well as by Amelunxen *et al* [34], if the entries of A are drawn from a standard normal (Gaussian) distribution (mean 0, standard deviation 1), then a certain phase-transition behavior will be encountered in the phase space of sparsity and sampling. Specifically, there exist combinations of sparsity and sampling levels such that in a specific well-defined sense almost all images will be recovered, while for other combinations almost all images will fail to be recovered. We refer to such regions as 'full-recovery' and 'no-recovery' regions and the two partition the entire sparsity-sampling phase space and are separated by a sharp phase transition. This is illustrated in figure 3, which is replotted and simplified from [28]. In the left Amelunxen–Lotz–McCoy–Tropp (ALMT) phase diagram the red curve is the theoretical phase transition above which almost all images are recovered, while below almost none are. The dashed red curve is for when non-negativity is known and enforced. In the right and equivalent, but differently parametrized, Donoho–Tanner (DT) phase diagram full recovery occurs below the curves. One of the results of [28] was the empirical observation that a similar sharp phase transition can be observed when changing to images sparse in the gradient domain reconstructed by TV-minimization instead of L^1-norm minimization and for CT-data instead of Gaussian samples. The empirically observed TV phase transitions are shown in cyan in figure 3 along with the raw percentage (0% = black, 100% = white) of image instances recovered at different sampling-sparsity positions.

The goal of the present work is to test whether similar observations can be made when considering real x-ray CT data. Specifically, we target two questions:

Q1 Can a relation between the critical number of projections for TV-regularized reconstruction and the image sparsity level be observed for real x-ray CT data?

Q2 If yes, can the critical number of projections be predicted from the image sparsity level using the empirical phase-transition established in previous simulation work [28]?

In the conference contribution [21] we attempted to answer **Q1** using the previously published SophiaBeads dataset [19]. While appropriate for its stated purpose of comparing reconstruction algorithms in a fixed-exposure trade-off study, and in principle also interesting for SR, we found the dataset to be of limited use for addressing **Q1** for the reasons explained in section 2.1. This prompted us to acquire the new and more extensive SparseBeads dataset for addressing the questions above, as explained in the following sections.

4. Computational setup

4.1. Image quality assessment

Meaningful quantitative image-quality assessment is a complex task. In principle, there is no general notion of image quality. As noted by Barrett [35] it is most meaningful to assess the quality of an image in relation to a particular task that it is intended to solve. Most work on task-based image-quality assessment has occurred in medical imaging (see [35] and references therein), where scans are generally performed with a specific purpose in mind, such as detection of certain features of interest for diagnosis, and the task is thus well-defined. In non-destructive testing and materials science typical imaging

tasks include detection (e.g. of cracks) and quantification (e.g. particle size and morphology).

The present work is not motivated by a specific practical imaging task, but from trying to establish theoretical connections observed in the recent work based on simulation studies [28, 30–32]. As an image-quality metric this previous work used the relative 2-norm error (closely related to the root-mean-square error) defined as

$$E_2(x) = \frac{\|x - x_{\text{ref}}\|_2}{\|x_{\text{ref}}\|_2}, \qquad (3)$$

which is a generic and widely used quality metric and in particular is appropriate for measuring whether an image exactly matches the reference image x_{ref}, which was the question of interest. While recognizing that exactly matching the reference image is not possible for real data, we employ this metric here in order to keep the connection to the previous work. Note that only pixels within the disc-shaped field of view are included in the computation.

The relative 2-norm error may however not always capture fully whether important features have been faithfully reconstructed. We therefore also employ additional quality measures which directly quantify whether certain geometric properties are preserved, as demonstrated in [36]. We do this by selecting a number of glass beads from the segmented reconstruction and do a quantification of the relevant properties using the software Avizo. We then compare with the same property measured on the same beads in the reference image and define the associated quality measure to be

$$E_Q = \frac{|M(x) - M(x_{\text{ref}})|}{M(x_{\text{ref}})}, \qquad (4)$$

where M here indicates the chosen geometry measure. In the present case of approximately spherical glass beads we will assess the following three geometric properties:

 (i) Aspect ratio: the ratio of the shortest to longest measured diameters of a bead.
 (ii) Shape: the ratio of the surface area of the object to the volume of a sphere. This can also be thought as the cubic-root of the sphericity measure defined in [37].
(iii) Volume: the bead volume determined from the number of pixels belonging to a bead.

The resulting quality measures will be referred to as E_A, E_S and E_V. As a qualitative supplement to the quantitative measures, visual inspection will also be employed.

4.2. Reference image and gradient sparsity level determination

A reference image is required to quantitatively assess reconstruction errors. In this study the reference image is obtained as the 40-iteration CGLS reconstruction using all 2520 projections. The only preprocessing steps applied to the raw data before reconstruction are a center-of-rotation correction of 12 pixels as well as the negative logarithm of Lambert-Beer's law. The reconstructed image is filtered using a 5-by-5 median filter to reduce the noise in order to have a reference image

that is more piecewise constant, as expected from a sample consisting of glass beads, plastic and air. The filter size was chosen empirically by determining the best trade-off between noise reduction and preservation of spatial resolution across all bead samples and test filter sizes 3, 5, 7, 9, 11. While the larger beads allow large filters, the smaller ones are subject to extensive oversmoothing/loss of spatial resolution with filter sizes 7 and above, prompting the use of filter size 5×5 pixels. In figure 4 the selected reference images are shown along with gradient images.

Two methods are used for determining the gradient sparsity level of the reference images. Measure 1 computes the gradient magnitude at each pixel using forward differencing in each direction followed by squaring, summing directions and taking the square root. Since the images are obtained as reconstructions from real data some noise is present, which yields small nonzero values that should not be counted as part of the bead edges. For this reason only values larger than an empirically determined threshold are counted, and the threshold value was set to $2.0 \cdot 10^{-4}$. The thresholded gradient magnitude images capture exactly the edges between beads and air in all cases. The edge between the plastic container and air is below the threshold and thus corresponding pixels are not counted towards the sparsity level. This method may have a tendency to overestimate the number of gradient nonzeros, since several pixels across any given boundary will be counted due the transition from background to bead stretching over a few pixels. In principle the transition is sharp and to better capture this Measure 2 introduces a segmentation of the reference image before forward differencing. In this way a single band of pixels will have nonzero gradient and estimated sparsity values be lower and possibly more representative of the actual beads. The sparsity level values in table 3 are given as the percentage of nonzeros relative to the total number of pixels, 2000^2. It can be noted that, as expected, both sets of gradient sparsity levels increase when the bead diameter is decreased. Hence, different bead sizes allows to study the influence of gradient sparsity level on the number of measurements required for accurate reconstruction using TV-regularization.

4.3. Total variation regularization and algorithm

We denote the log-transformed, center-of-rotation-corrected projection data by b, the 2D fan-beam system matrix by A, a reconstructed image by x, and the number of projections by N_θ.

To determine a TV-regularized reconstruction (which can be seen as the maximum *a posteriori* estimate in a Bayesian formulation) of the discrete imaging model $Ax = b$ we solve an optimization problem as a compromise between fitting the data and enforcing regularity on the solution,

$$x_{\text{TV}} = \arg\min_x \frac{1}{2N_\theta} \|b - Ax\|_2^2 + \alpha T_\tau(x), \qquad (5)$$

where x_{TV} is the TV-regularized solution, α is the TV regularization parameter and $T_\tau(x)$ is the Huber-smoothed total variation, defined as
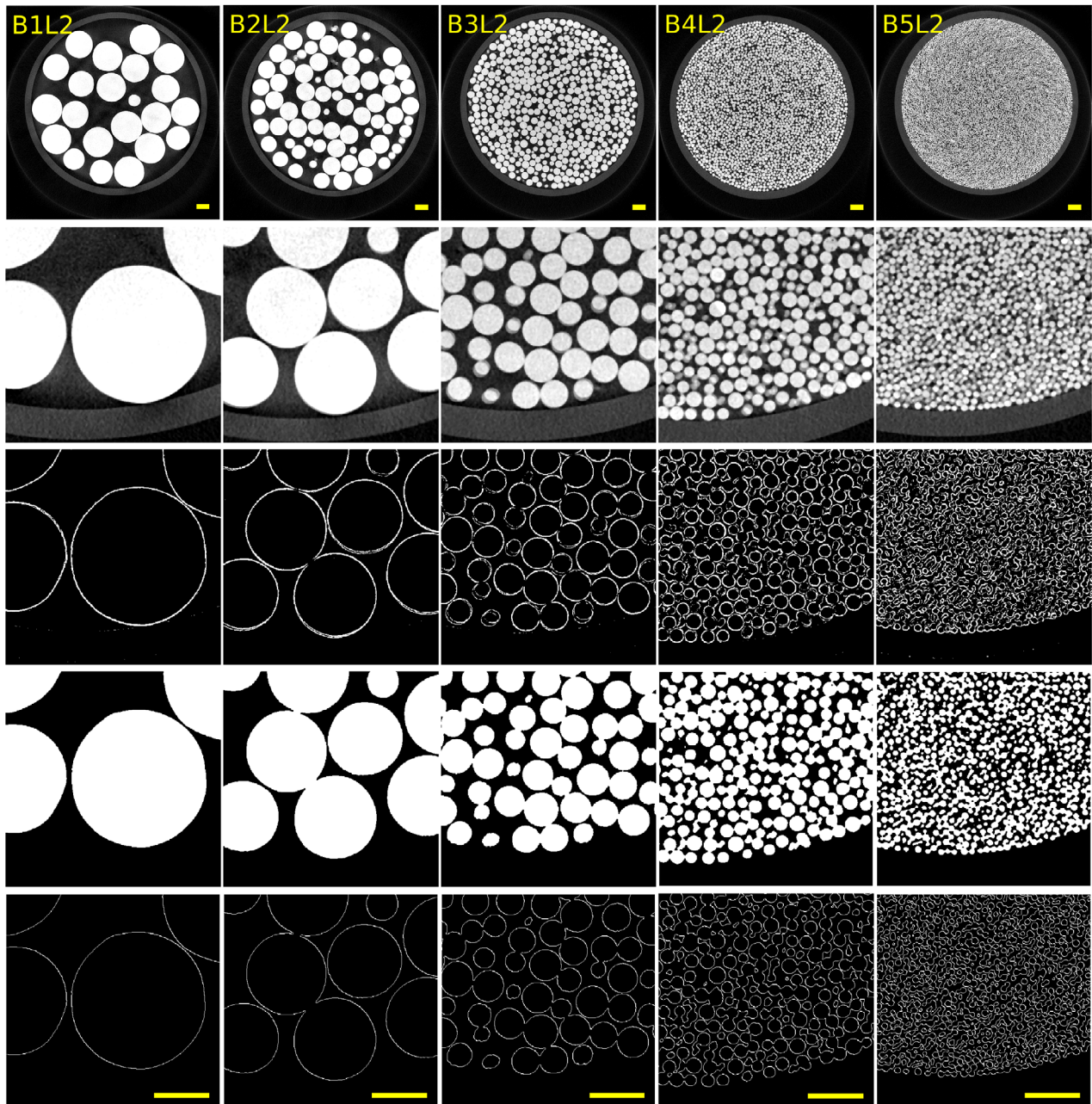
Figure 4. Regions of interest taken from reference images for B1L2–B5L2 obtained as 40 CGLS iterations followed by median filtering. 1st row: Full image, gray-scale window: $[0, 1.5 \cdot 10^{-3}]$. 2nd row: Zoom to 450×450 pixel region, same gray-scale window. 3rd row: Thresholded gradient magnitude images, threshold value $2 \cdot 10^{-4}$, basis for sparsity measure 1. 4th row: Segmented images. 5th row: Gradient magnitude of segmented images for sparsity measure 2. Scalebar: 2 mm.

$$T_\tau(x) = \sum_j \Phi_\tau(\|D_j x\|_2), \quad \text{where} \quad \Phi_\tau(u) = \begin{cases} |u| - \frac{1}{2}\tau, & \text{if } |u| \geqslant \tau, \\ \frac{1}{2\tau}u^2, & \text{otherwise.} \end{cases}$$

(6)

Here, τ is the Huber-smoothing parameter, D_j is the finite-difference approximation of the image gradient at pixel j, and $\|\cdot\|_2$ is the (Euclidean) 2-norm.

Smoothing is used to ensure a solution by smooth optimization techniques, which are generally faster than their non-smooth counterparts. However, depending on the choice of smoothing parameter, τ, this might modify the reconstruction. In our implementation, we use a sufficiently small value of τ relative to the image values that the smoothing

effects are negligible. This is discussed in more detail in the next section.

The normalization by N_θ helps to compare reconstructions obtained at different N_θ by compensating the magnitude of the first term. As a result, a fixed α value yields the same balance between the two terms at different N_θ, which reduces the range of relevant α values, making it more practical to find the optimal value.

A disc-shaped mask (covering the beads and the walls of the sample container) was used to ensure a reconstruction of the relevant field of view, and to reduce the number of unknown pixel values and thus the computation time. The pixels outside the field of view were set to zero since there

Table 3. Datasets used and determined sparsity levels (percentage of nonzero pixels in 2000×2000 reconstructed images) using sparsity measures 1 and 2, i.e. without and with a segmentation step.

	B1L2	B2L2	B3L2	B4L2	B5L2
Glass bead diameters (mm)	5	3	1	0.600–0.425	0.300–0.212
Average glass bead diameters (pixels)	287	172	57	29	15
Sparsity meas. 1 (%) (without segmentation)	1.29	2.21	4.58	8.34	11.60
Sparsity meas. 2 (%) (with segmentation)	0.48	0.85	2.25	4.56	7.90
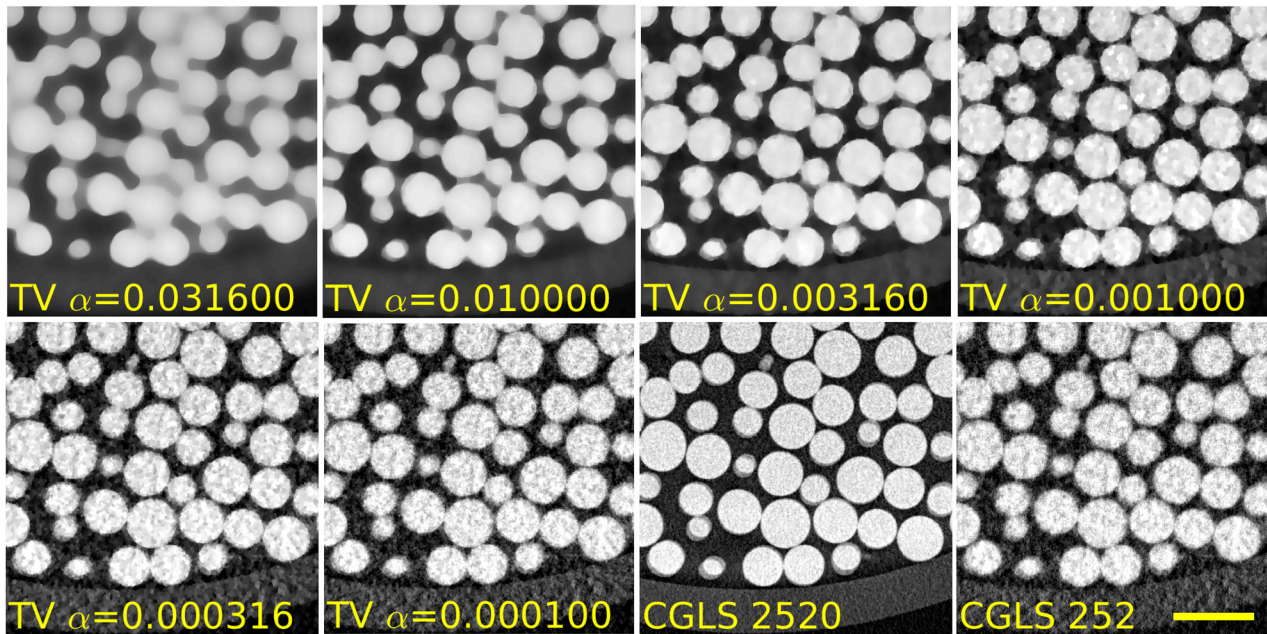
Figure 5. Effect of TV regularization parameter α for region of interest from B3L2 at 252 projections (downsampling factor 10). Top, left to right: $\alpha = 3.16 \cdot 10^{-2}, 1.0 \cdot 10^{-2}, 3.16 \cdot 10^{-3}, 1.0 \cdot 10^{-3}$. Bottom: $\alpha = 3.16 \cdot 10^{-4}, 1.0 \cdot 10^{-4}$, CGLS reconstructions at 2520 and 252 projections (downsampling factors 1 and 10) for comparison. Scalebar: 2 mm. Gray-scale window: $[0, 1.5 \cdot 10^{-3}]$.

are no other objects but air outside the sample holder. The masking was applied by determining the pixel indices within the sample holder, and passing only those pixels to the forward- and back-projector.

4.4. TV regularization parameter selection

Whereas CGLS requires only the number of iterations to be chosen and FDK only the filter, TV regularization needs a number of parameters to be set. These can be divided into two categories as pointed out by [38]: problem-specification parameters and algorithm parameters. The problem-specification parameters are part of the *a priori* assumptions along with the optimization problem itself and help specify the mathematical solution. In the present case the problem-specification parameters are the regularization parameter α and the Huber-smoothing parameter τ, as they occur in (5). The algorithm parameters specify the behavior of the optimization algorithm used to solve the mathematically defined problem. In our case, we use the GPBB (gradient projection with Barzilai–Borwein acceleration) method from the TVReg [39] MATLAB package and the noteworthy algorithm parameters are the maximum numbers of iterations to perform and the threshold value used in the convergence criterion; in addition a number of internal algorithm parameters, for example regarding step-size selection can also be chosen.

Considering the problem-specification parameters, the Huber-smoothing parameter is fixed at $\tau = 10^{-5}$ in all studies. A smaller value of τ means less smoothing and thus a better approximation to the ideal non-smoothed TV-regularization problem (which corresponds to $\tau = 0$). On the other hand, convergence speed of the algorithm worsens as τ is decreased [39], so in practice a trade-off must be used. To obtain a good approximation to the non-smoothed TV, it is sufficient that τ is small relative to the smallest difference between neighboring pixel values, which in the present study was found to be of the order of 10^{-4}–10^{-3} based on the reference images. Further, we empirically assessed the effect of τ by comparing reconstructions at $\tau = 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}$. Large differences (much smoother images) were evident for $\tau > 10^{-5}$, while for $\tau \leqslant 10^{-5}$ no noticeable differences were observed.

The regularization parameter α balances how strongly TV regularization is enforced compared to achieving the best least-squares fit to the data. Larger values of α yield a smoother image while smaller values yield images that exhibit no regularization, see figure 5. The optimal regularization parameter α, i.e. the value that yields in some sense the best reconstructed image, depends on many non-trivial factors including noise, number of projections, and scaling of the image pixel values. For each of the TV-reconstructions presented in section 5,

the regularization parameter α was optimized independently. The choice of regularization parameter is a trade-off between over- and under-smoothed images, where in some cases no good trade-off exists or the best trade-off can be subjective. In each case we tested a range of α parameters and chose the optimal parameter as the one yielding the smallest relative 2-norm error of the corresponding TV-reconstruction. This was to have an objective parameter selection method, which was also fully automated, in order to process the very large number of individual reconstructions computed in the results. Further fine-tuning the regularization parameter might slightly improve some reconstructions; however we consider it unlikely that this would affect any of our conclusions significantly.

Algorithm parameters were chosen to ensure an accurate solution to the TV-regularization problem, while preventing excessive computing times. In practice we observed negligible changes to the image beyond $10\,000$ iterations and therefore used this number as the maximum number of iterations to run. In addition we found it useful to limit unnecessary iterations to save time by terminating if the norm of the gradient of the objective function in (5) became sufficiently small. In practice we used the `epsb_rel` input of TVReg to 10^{-7}, as no noticeable changes were observed in tests beyond this value.

5. Results

5.1. Visual inspection of reconstructions at reduced numbers of projections

A key feature of the SparseBeads dataset is that many perfectly equiangular subsets can be extracted. The 48 possible downsampling factors and resulting numbers of projections are given in table 1. The datasets B1L2, B2L2, B3L2, B4L2 and B5L2 are TV-reconstructed from projection data at each of the 48 available sub-sampled datasets. Figure 6 shows a representative selection of a region of interest taken from the reconstructed images. Our previous simulation work showed that the undersampling factor, at which accurate TV-reconstruction is possible, grows when images become more sparse (have more zero pixels). Thus we expect larger beads (which have a more sparse gradient image) to allow accurate reconstruction at greater undersampling factors than smaller beads. Looking down columns of figure 6 this expectation appears to be confirmed visually: reconstructions of the largest beadpack B1L2 appear sharp and clean even at undersampling level 20–40. With decreasing bead size, we observe that only smaller undersampling factors yield reasonable reconstructions, roughly along the SW-NE diagonal, with the smallest beadpack, B5, only allowing reconstruction at undersampling factors between 1 and 5. Further, relative 2-norm errors E_2 are given in the lower-left corner of each reconstruction. The numbers are color-coded by reconstruction error according to satisfactory reconstruction thresholds to be defined in the following subsection: green ($E_2 < 0.10$), cyan ($0.10 \leqslant E_2 < 0.15$), yellow ($0.15 \leqslant E_2 < 0.20$), and magenta ($0.20 \leqslant E_2$). The error trends confirm the visual inspection in that transition from small to large errors occur around the SW–NW diagonal.

To further illustrate reconstruction quality as the number of projections is decreased, figure 7 shows horizontal line profiles through the region of interest of the TV-reconstructions in figure 6 for each bead size. In all cases contrast and spatial resolution are gradually lost with decreasing number of projections.

We note from figure 6 and figure 7 that the two larger bead sizes appear to have a larger intensity of around $1.6 \cdot 10^{-3}$ than the three smaller ones at approx. $1.4 \cdot 10^{-3}$. Due to the manufacturer's bead specifications we had expected all beads to consist of exactly the same material, however some manufacturing differences must be present. The important thing for the dataset and present study is however the large contrast to the background causing all images to be essentially binary, and as such we expect the slight intensity difference to be insignificant for the results.

5.2. Reconstruction error versus number of projections and image sparsity

To substantiate the visually observed connection between bead size and admitted undersampling level we compute the relative 2-norm errors for all 48 undersampling factors for all five bead sizes, see figure 8. To allow for a more detailed interpretation of the results we display the resulting error curves both in linear and log–log plots.

In the linear plot, different reconstruction error decrease rates are observed among the bead sizes. In agreement with the visual inspection, the reconstruction error of the largest B1L2 decays rapidly to an almost constant level after a couple of hundred projections. As the beads become smaller, the error decays become more gradual and error values larger. This confirms the expectation that smaller beads, which are less sparse in their gradient image, require more projections to achieve the same level of reconstruction accuracy as larger beads.

In the log–log plot the trends are even more pronounced: All reconstruction error curves have a similar shape and the number of projections to reach the same reconstruction error increases systematically with decreasing bead size. Again, this supports our hypothesis of a simple link between sparsity level and critical sampling level.

To quantify the observed connection between sparsity and critical sampling level we declare a reconstruction satisfactory if the relative error is below a given threshold. Clearly the choice of threshold is somewhat arbitrary and we thus compare three different threshold values ($E_2 = 0.20$, 0.15 and 0.10), as illustrated in figure 8. The smallest threshold is chosen at the point after which all error curves are settling to an almost constant level; the largest at the point where all curves have started their steep (log–log) decay; and the third threshold value as the average of the two. As can be seen from figure 6 these thresholds are meaningful to use as indicators of (increasingly) satisfactory reconstruction across all bead sizes, since as reconstruction errors decrease through the thresholds visual image quality changes from unsatisfactory to satisfactory. We determine the critical number of projections for each bead size (and each threshold) as the smallest number of projections that produce a satisfactory reconstruction. The critical
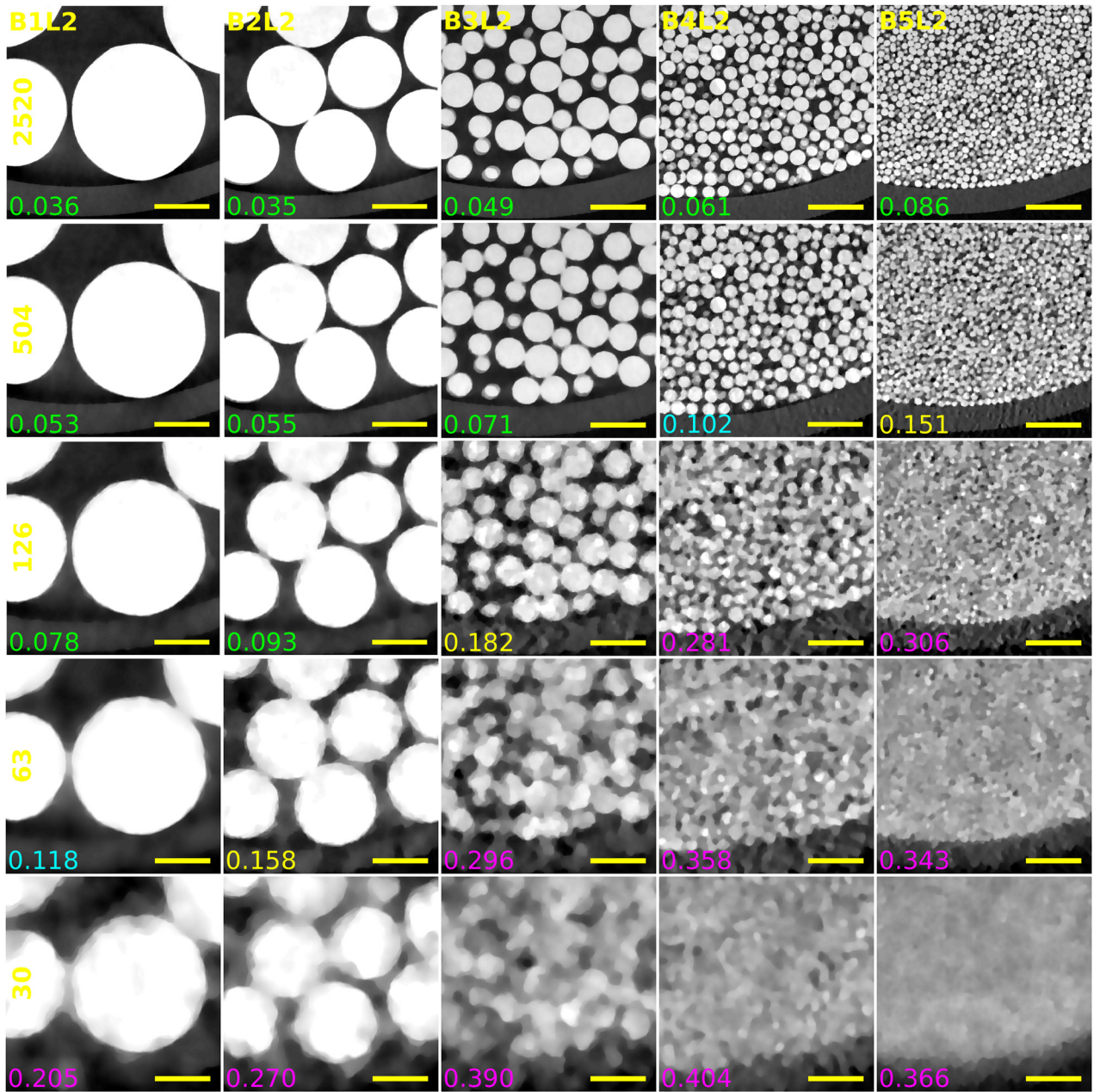
Figure 6. Regions of interest from TV reconstructions. Left to right: B1L2–B5L2. Top to bottom, number of projections (downsampling factors): 2520 (1), 504 (5), 126 (20), 63 (40), 30 (84). Bottom left corners: Relative 2-norm errors E_2, color-coded relative to satisfactory reconstruction thresholds from section 5.2: green: $E_2 < 0.10$, cyan: $0.10 \leqslant E_2 < 0.15$, yellow: $0.15 \leqslant E_2 < 0.20$, magenta: $0.20 \leqslant E_2$. Scalebar: 2 mm. Gray-scale window: $[0, 1.5 \cdot 10^{-3}]$.
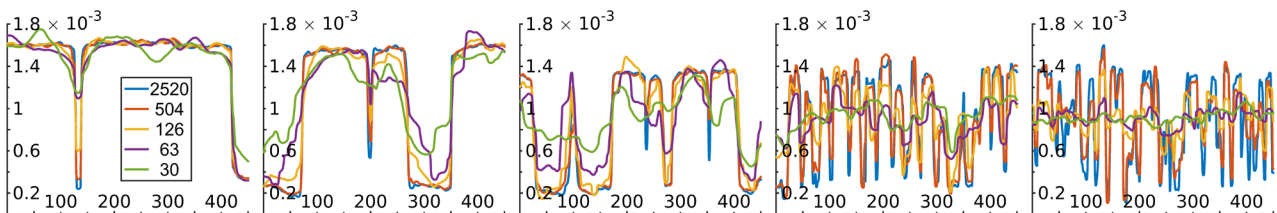
Figure 7. Central horizontal line profiles through 450×450-pixel region of interests of TV reconstructions shown in figure 6. Left to right: B1L2–B5L2. Horizontal axis: Pixel position. Vertical axis: Intensity. Legend (in left-most figure) shows the number of projections used.
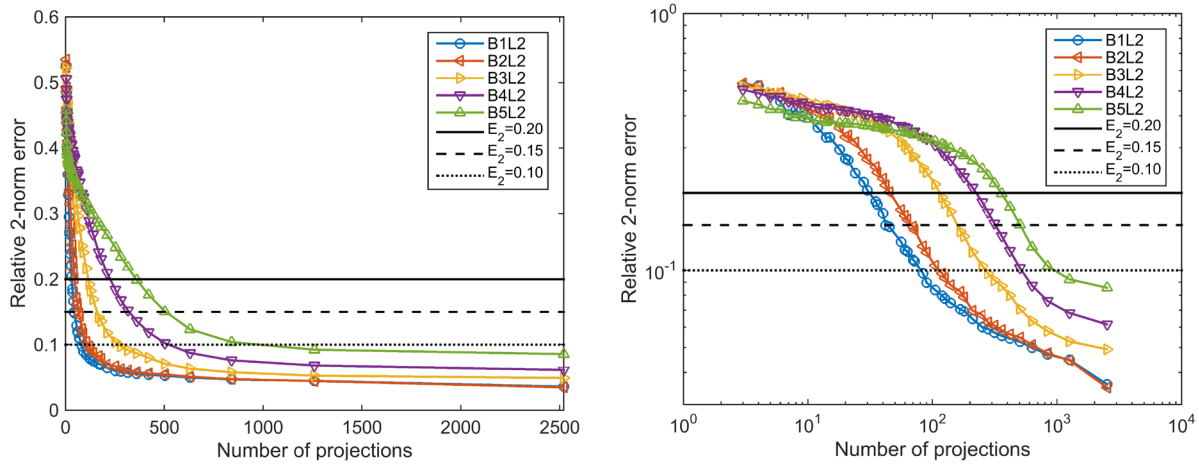
Figure 8. TV-reconstruction errors versus numbers of projections for single images (layer L2) at the five bead sizes (B1–B5). Left: Linear scale. Right: Log–log scale. Satisfactory reconstruction thresholds of 0.20, 0.15, and 0.10 are shown as horizontal lines.
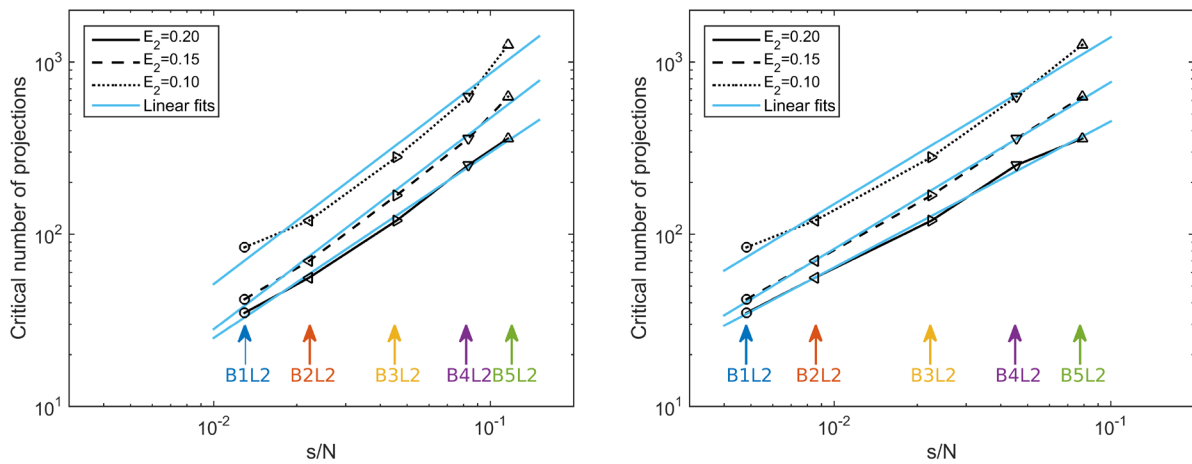

Figure 9. Relation between gradient sparsity level s/N, where s is the number of nonzero gradient values and N the total number of pixels, and the critical number of projections for a satisfactory TV-regularized reconstruction. Left: Sparsity levels determined determined using measure 1. Right: Using measure 2. The critical number of projections grows broadly linearly with the sparsity level.

numbers of projections are plotted against the gradient sparsity levels in figure 9. The log–log plots reveal in all cases an approximately linear dependency, which is supported by best linear (in the log–log scale) models, corresponding to power functions $N_{\text{crit}} \approx b \, (s/N)^a$, where best-fit a and b coefficients along with R^2 coefficients of determination are reported in table 4.

All R^2 values are close to 1, especially for thresholds 0.15 and 0.20, indicating that the linear model (power function after exponentiation) is appropriate. With the a-coefficient (exponent) being close to 1 in all cases it is reasonable to approximate the relation by a direct linear relationship

$$N_{\text{crit}} \approx b \, \frac{s}{N}. \tag{7}$$

In other words, from the data presented we observe an approximately linear relation between the gradient image sparsity and the critical number of projections using TV-regularized reconstruction. This establishes a positive answer to our targeted question **Q1**.

Note from table 4 that the linear coefficient b, which determines how quickly the critical number of projection grows

Table 4. Linear regression results for logarithm-transformed data yielding relation $N_{\text{crit}} \approx b \, (s/N)^a$, which due to all a coefficients being ≈ 1 can be approximated by $N_{\text{crit}} \approx b \, (s/N)$.

Sparsity meas.	Thresh.	a	$b/10^3$	R^2
1	0.20	1.0771	3.5623	0.9968
	0.15	1.2266	7.9794	0.9938
	0.10	1.2231	14.3665	0.9758
2	0.20	0.8487	3.1985	0.9970
	0.15	0.9686	7.1175	0.9985
	0.10	0.9693	12.9925	0.9874

with sparsity level, depends on the sparsity determination method. This is also observed in, and can be explained by, figure 9. The critical numbers of projections are the same in the left and right plots, only the sparsity values differ. Since sparsity measure 1 counts more nonzeros than sparsity measure 2, the curves are shifted to the right and are steeper, which corresponds to larger b-values. We conclude that both choices of sparsity measure lead to similar linear trends. Given that the determination of sparsity levels from reconstructed images is
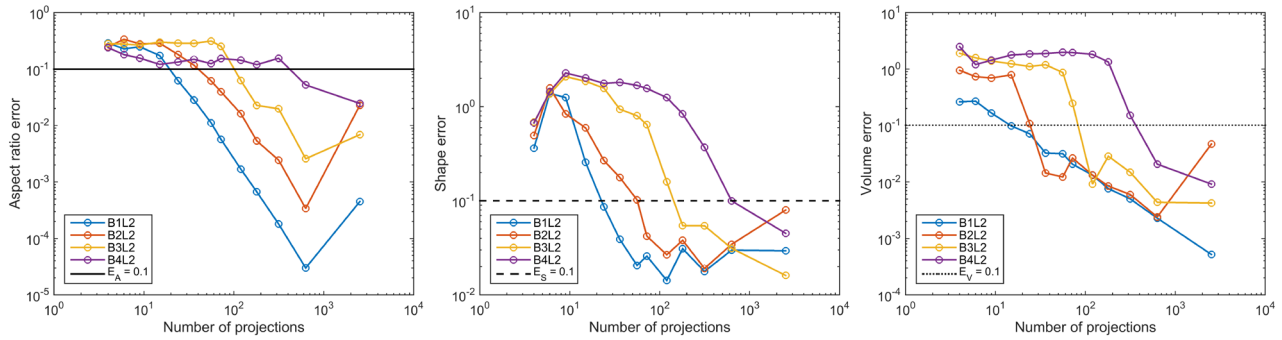
Figure 10. Aspect ratio, shape and volume segmented bead morphology quantification errors of TV-regularized reconstructions as function of numbers of projections. Satisfactory reconstruction thresholds of 0.10 are shown on each plot. Compare with figure 8.
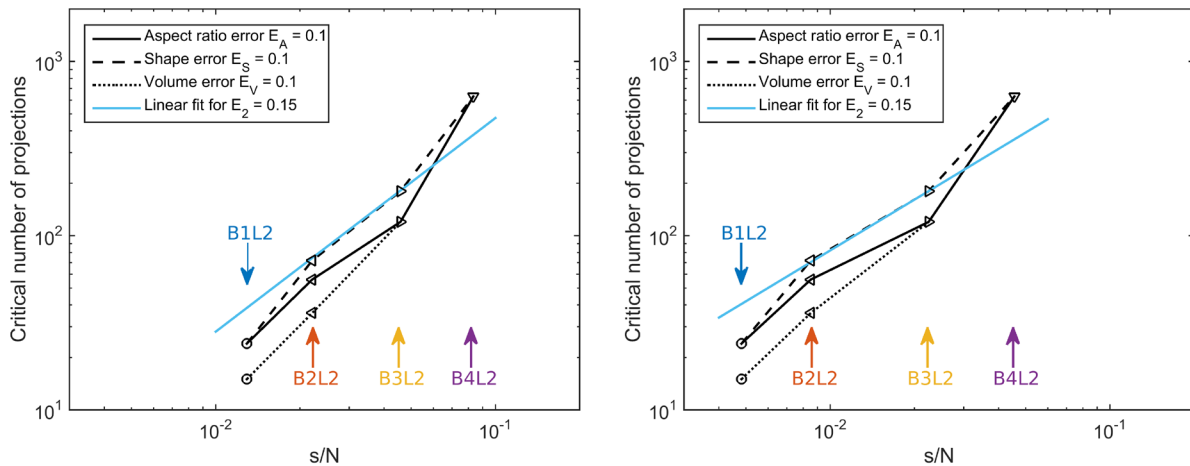

Figure 11. Critical number of projections for a satisfactory TV-regularized reconstruction as measured by the aspect ratio, shape and volume quantification errors. To help compare with the relative 2-norm error measure results in figure 9, the obtained linear models for threshold 0.15 are also included. Left: Sparsity levels determined determined using measure 1. Right: Using measure 2.

more ambiguous than from synthetic images, it is reassuring that the two choices employed here produce similar results.

To assess whether the simple choice of the relative 2-norm error as the quality measure affects the obtained results we reanalyze the experiment using the bead morphology quantification measures of aspect ratio, shape and volume of the segmented beads. As the process of determining the measures is rather laborious we only provide values for 13 selected downsampling factors across the range of the 48 divisors of 2520. Figure 10 shows the error measures as function of numbers of projections. While the behavior is less defined than for the relative 2-norm error, the overall behavior is very similar for the four largest bead sizes: from approximately constant errors the curves for the different bead sizes experience a rapid decrease at different numbers of projections in the same order as before until a quite steady level is reached. The errors determined for the smallest bead size B5 have been omitted as it was not possible to segment and separate enough beads to provide a meaningful error measure. A single threshold for a satisfactory reconstruction is chosen as 0.1 for all three quality measures and the critical numbers of projections are computed and plotted as function of the two sparsity levels in figure 11. In both cases a linear trend is again observed but no linear regression is performed. Instead to allow comparison with the relative 2-norm error results in figure 9 we display the previously

obtained linear fits corresponding to the threshold of 0.15. In both cases the slope matches closely with the quantification results. We thus conclude that the quantification measures support the same positive answer to **Q1** as we obtained for the relative 2-norm error measure. This strengthens the conclusion as the quantification measures may be more representative of reconstruction quality, as they more physically reflect geometric properties of the glass beads.

5.3. Phase diagram analysis

Having established a positive answer to **Q1**, we now focus on **Q2**, which is to assess whether the found relation agrees with the empirical phase transition established in previous simulation work [28].

In [28] a sharp phase transition was observed empirically for TV-regularized reconstruction of a class of images with sparse gradient magnitude. This phase transition occurs in the $(\delta, \rho) \in [0, 1]^2$ phase space, where $\delta = m/N$ is the number of (single line integral) measurements relative to the total number of pixels and $\rho = s/m$ is the sparsity level relative to the number of measurements. A combination of an image of a given sparsity and a given number of measurements acquired determine a position in the phase diagram. Below the phase-transition curve essentially all combinations were
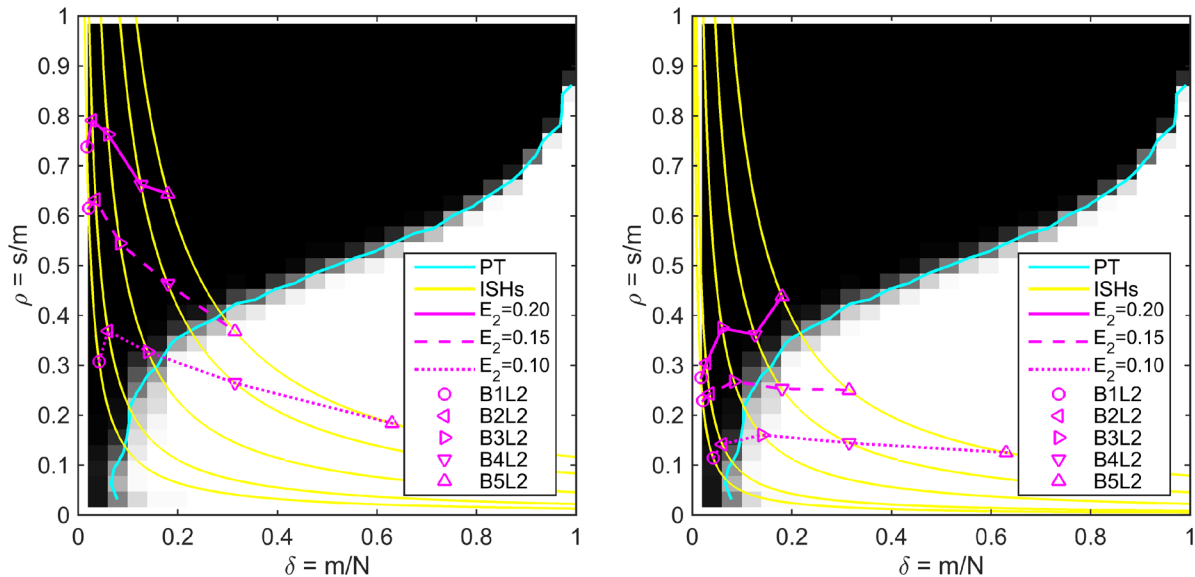
Figure 12. Empirical phase diagrams for sparsity measures 1 (left) and 2 (right) using E_2 image-quality measure. The empirical phase-transition (PT) curve sharply separates a full-recovery region (in white) from a no-recovery region (in black) and specifies the expected critical number of projections for a satisfactory reconstruction. Observed critical numbers of projections for each bead size are plotted as points based on the sparsity level and connected using lines with line styles indicating the threshold used. Auxiliary iso-sparsity hyperbolas (ISHs) shown trace out curves of constant sparsity levels of each bead size.

found to yield a satisfactory reconstruction using TV regularization, while above the reconstructions were unsatisfactory. The phase-transition curve specifies the expected critical sets of (δ, ρ) values separating the full-recovery and no-recovery regions of the (δ, ρ) phase space. The empirically determined phase-transition curve found in [28] is shown in figure 12.

If one knows the (gradient) sparsity s of an image with N pixels, then the phase transition curve specifies the critical number of measurements needed for TV-regularized reconstruction to produce a satisfactory reconstruction. By combining the expressions of δ and ρ we obtain

$$\rho = \left(\frac{s}{N}\right)\frac{1}{\delta}, \tag{8}$$

from which it is seen that an s-sparse image lies on a hyperbola in phase space, henceforth referred to as an iso-sparsity hyperbola (ISH). The point at which the ISH and the phase-transition curve intersect determines the critical number of measurements, m, that suggests a satisfactory reconstruction for the given sparsity level. In [28] this approach predicted that 69.3 projections would suffice to recover a digitized walnut image, which was almost perfectly correct, since the walnut was found empirically to be recovered at 68 projections.

In the present study instead of computing predicted critical numbers of projections we will compare the observed critical number of projections for all bead sizes directly with the phase-diagram predictions by plotting each one into the phase diagram. If the phase transition can predict the critical number of projections, each point should lie on or close to the phase-transition curve. This is done in figure 12 for both sparsity measures, the E_2 image-quality measure and all satisfactory-reconstruction thresholds. ISHs at each bead sparsity level are shown to illustrate the transition across the phase diagram.

We observe that only a few of the points are close to the phase-transition curves, and in particular for sparsity measure 1, most points are situated too high above the curve, which may again indicate that this method overestimates the true sparsity value. The points for sparsity measure 2 occur closer to the phase transition but do not appear to follow it at any of the chosen thresholds. From the data presented we must therefore conclude that the phase transition does not provide a useful method for predicting the critical number of projections from the (gradient) sparsity level. In other words, a negative answer to **Q2**.

Based on the almost-linear relationship between sparsity and critical sampling level (7), the lack of agreement with the phase transition should not be a surprise. We can convert (7) into the phase space variables to see what the relation corresponds to in the phase diagram. Given that each projection consists of $\sqrt{N} = 2000$ single measurements we find the critical number of measurements $m_{\text{crit}} = N_{\text{crit}}\sqrt{N}$ and the critical value of ρ to be

$$\rho_{\text{crit}} = \frac{s}{m_{\text{crit}}} = \frac{\sqrt{N}}{b}, \tag{9}$$

which is a constant independent of δ. For example for sparsity measure 2 and threshold value 0.15, this yields a constant of 0.2810. This matches quite well with the observed, almost-horizontal, dashed curve in figure 12, which lies at $\rho \approx 0.25$. In other words the observed almost-linear relationship providing a positive answer to **Q1** is incompatible with the non-linear phase transition and results in a negative answer to **Q2**.

For completeness we also present in figure 13 the phase diagram for the quantification measures. Similarly to the relative 2-norm results, the observed critical number of projections do not consistently occur close to the phase-transition
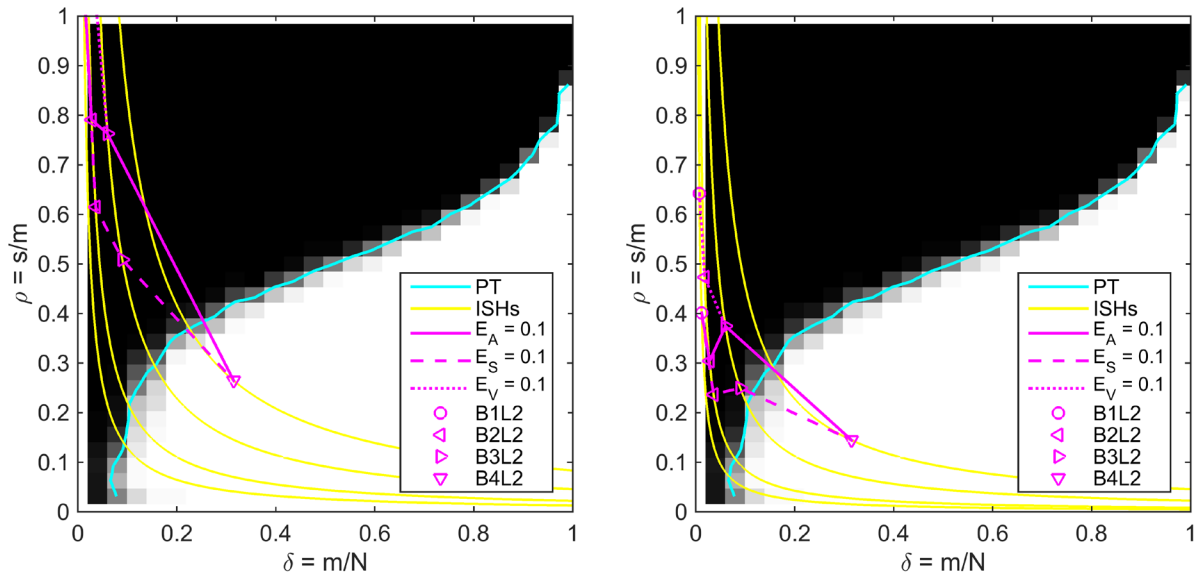
Figure 13. Comparison of observed critical number of projections by aspect ratio, shape and volume quantification errors with the phase diagram prediction of critical sampling level.

curve. We conclude again that the results of the conducted study are not consistent with a positive answer to **Q2**.

6. Discussion

Upon close inspection figure 6 shows some artifacts, mostly clearly seen for B3L2 and B4L2 and take the form of mainly smaller bead cross sections having a 'double appearance' with two almost identical copies slightly offset. The same artifacts are present in the CGLS reconstructions in figure 4, which means that it is not an artifact of TV regularization but rather inherent to the data. We believe this artifact may arise from slight bead movement during data acquisition. The movement artifacts may be removed by filling voids for example by epoxy to fix the beads, or by subjecting the packed beads to light sintering, which would consolidate the loose beads into a rigid structure, while preserving the spherical shape of the beads apart from at their contact points. A third option may be to 3D-print test phantoms of varying sparsity similar to the approach in [40] modeling metal foams.

Other possible sources of errors compared to a simpler ideal synthetic data scenario include computational aspects in the numerical solution of the TV-regularized reconstruction problem for large-scale data. We used the ASTRA tomography toolbox [41] for GPU-acceleration of the forward and back-projection operations, which are the most computationally expensive steps involved. For efficiency reasons this uses only single precision and involves an unmatched forward-backprojector pair implementation, i.e. the back-projection operation is not equivalent to applying the adjoint of the forward-projection operation represented by the matrix *A*. This may lead to inaccuracies and convergence issues [42], but the effect may be kept at bay by the additional regularity enforced by TV regularization. Furthermore a larger number of iterations may yield slightly different reconstructions, but as we observed essentially no change in the image after 10 000 iterations in a

range of inspected cases, we believe the reported observations will remain unchanged. Finally, any minor effect of using the Huber-smoothed TV may be removed by using an algorithm for non-smooth optimization such as the Chambolle–Pock algorithm [43, 44], but as explained in section 4.4 we believe to have already reduced the smoothing effect to be negligible. In the present work we chose to use the TVReg algorithm because we have found it to work very robustly when applied to real data and interfacing to the ASTRA Tomography Toolbox across a range of parameter regimes.

In many clinical and non-clinical cases the objects under study can be considerably more complex than the glass beads used here. With the present dataset we deliberately designed the samples with simplicity in mind and the different-sized bead edges being the dominant features present. This was motivated by our focus on establishing a connection between gradient sparsity level and the number of projections sufficient for accurate TV-regularized reconstruction. With a more complex set of test images, it would be substantially more involved to isolate this particular relation from influences from different contrast, feature scales, soft tissue variation, etc. We believe the results and dataset provided at present contribute an important step forward and are looking to generalize to more realistic cases in future work.

7. Conclusion

The data presented suggest that in the range of sparsity levels considered the critical number of projections for satisfactory reconstruction by TV regularization grows approximately linearly with the image gradient sparsity. To our knowledge this is first empirical evidence and quantification of a link between gradient image sparsity and the number of projections sufficient for accurate reconstruction by TV regularization. The closest related work we are aware of is [45], which as our work is motivated to assess the image quality of

TV-regularized reconstruction as function of the number of projections. It would be interesting to compare their results to ours, however their goal is slightly different in that they used a single test image without considering its sparsity level. The link between gradient sparsity and sufficient number of projections is the focus of our study, and as such a direct comparison is not possible.

We believe that under similar conditions the empirically established relation may be used as a guideline for determining the minimum number of projections to acquire for a satisfactory TV-regularized reconstruction based on the expected sparsity of the sample. For increasingly sparse samples this may lead to a large reduction in scanning effort (time, x-ray exposure, etc) of an order of magnitude or more, depending directly on the sparsity level as demonstrated by the present work, as opposed to dense angular sampling employed at present. From a practical perspective we can think of image sparsity as a proxy of 'sample simplicity'. Using this interpretation our results broadly say that complex samples (small features) do not permit a reduced number of projections, while as the sample simplicity increases (e.g. larger features) the number of projections can be reduced without loss of quality and at a roughly linear rate.

An important question for future work is how general the observed relation is when changing the scan conditions and the type of sample scanned. For example, we would expect the relation found to hold also for samples much larger or smaller with correspondingly larger or smaller pixel size used so that spherical features and sparsity levels would be comparable. If changing the detector resolution we would also expect the same relation to hold assuming the reconstruction resolution and number of projections are changed by the same factor. An interesting question is how the relation found depends on the shape and complexity of features. We expect the relation to hold for samples with similar characteristics of few, distinct phases with relatively simple structure. Many samples of interest to materials science including porous media, composites and foams as well as samples from non-destructive testing and metrology possess such characteristics. For more complex samples, the present results may still be used as a reference, while the proposed methodology could be used to investigate a possible similar relation.

Other SR methods including generalized forms of TV, wavelets and dictionary learning could also be explored using the proposed methodology. Quite likely more sophisticated methods than TV may perform better, and the SparseBeads dataset could help to quantify in a real-data setting how much better in terms for example of how few projections suffice. The purpose of the present work was not to claim that TV is the ideal SR method, but to systematically quantify the relation between image sparsity and number of projections. The proposed methodology may be used to establish similar relations for example for L^0- and L^1-norm regularization, where we based on our previous work [28] would expect slightly different performance. Other future work include to address limitations described in the previous section such as bead movement to design and acquire a more optimized dataset. This may further corroborate the link between sparsity and

critical sampling level and possibly reveal a closer relation with the phase-transition behavior observed for simulated data.

The SparseBeads dataset collection was designed and acquired to enable the study presented and has been made available [22] from the Zenodo data sharing platform. We provide 48 structured datasets offering a range of image sparsity levels, number of projections, noise levels and multiple similar-sparsity test images to enable systematic studies of key parameters for SR. It is the hope of the authors that it will become a useful benchmark for assessing and comparing reconstruction algorithms using real x-ray CT data.

Acknowledgments

ORCID iDs

Jakob S Jørgensen https://orcid.org/0000-0001-9114-754X

References

[1] Maire E and Withers P J 2014 Quantitative x-ray tomography *Int. Mater. Rev.* **59** 1–43
[2] Bartscher M, Sato O, Härtig F and Neuschaefer-Rube U 2014 Current state of standardization in the field of dimensional computed tomography *Meas. Sci. Technol.* **25** 064013
[3] Ferrucci M, Leach R K, Giusca C L, Carmignato S and Dewulf W 2015 Towards geometrical calibration of x-ray computed tomography systems—a review *Meas. Sci. Technol.* **26** 92003
[4] Bradley R S 2016 Estimation of bias and variance of measurements made from tomography scans *Meas. Sci. Technol.* **27** 095402
[5] Thompson W M, Lionheart W R B, Morton E J, Cunningham M and Luggar R D 2015 High speed imaging of dynamic processes with a switched source x-ray CT system *Meas. Sci. Technol.* **26** 055401
[6] Warnett J M, Titarenko V, Kiraci E, Attridge A, Lionheart W R B, Withers P J and Williams M A 2016 Towards in-process x-ray CT for dimensional metrology *Meas. Sci. Technol.* **27** 035401

[7] Pan X, Sidky E Y and Vannier M 2009 Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction? *Inverse Problems* **25** 123009

[8] Bian J, Siewerdsen J H, Han X, Sidky E Y, Prince J L, Pelizzari C A and Pan X 2010 Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT *Phys. Med. Biol.* **55** 6575–99

[9] Chen G-H, Tang J and Leng S 2008 Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets *Med. Phys.* **35** 660–3

[10] Ritschl L, Bergner F, Fleischmann C and Kachelrieß M 2011 Improved total variation-based CT image reconstruction applied to clinical data *Phys. Med. Biol.* **56** 1545–61

[11] Sidky E Y, Kao C-M and Pan X 2006 Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT *J. X-Ray Sci. Technol.* **14** 119–39

[12] Sidky E Y and Pan X 2008 Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization *Phys. Med. Biol.* **53** 4777–807

[13] Song J, Liu Q H, Johnson G A and Badea C T 2007 Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT *Med. Phys.* **34** 4476–83

[14] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68

[15] Shepp L A and Logan B F 1974 The Fourier reconstruction of a head section *IEEE Trans. Nucl. Sci.* **21** 21–43

[16] Bruder H *et al* FORBILD Phantoms, available online: www.imp.uni-erlangen.de/phantoms/ (Accessed: 14 March 2017)

[17] Hämäläinen K, Harhanen L, Kallonen A, Kujanpää A, Niemi E and Siltanen S 2015 Tomographic x-ray data of a walnut (arXiv:1502.04064)

[18] Bubba T A, Hauptmann A, Huotari S, Rimpeläinen J and Siltanen S 2016 Tomographic x-ray data of a lotus root filled with attenuating objects (arXiv:1609.07299)

[19] Coban S B and McDonald S A 2015 SophiaBeads Dataset Project (https://doi.org/10.5281/zenodo.16474)

[20] Coban S B, Withers P J, Lionheart W R B and McDonald S A 2015 When do iterative reconstruction methods become worth the effort? *Proc. 13th Int. Mtg. of Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine (Newport, RI, USA)* pp 244–7

[21] Jørgensen J S, Coban S B, Lionheart W R B and Withers P J 2016 Effect of sparsity and exposure on total variation regularized x-ray tomography from few projections *Proc. 4th Int. Mtg. on Image Formation in X-ray CT (Bamberg, Germany)* pp 279–82

[22] Jørgensen J S, Coban S B, Lionheart W R B, McDonald S A and Withers P J 2017 SparseBeads Dataset Collection (https://doi.org/10.5281/zenodo.290117)

[23] Coban S B 2017 SophiaBeads Dataset Project Codes available online http://sophilyplum.github.io/sophiabeads-datasets (Accessed: 3 May 2017)

[24] Candès E J, Romberg J and Tao T 2006 Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information *IEEE Trans. Inform. Theory* **52** 489–509

[25] Donoho D L 2006 Compressed sensing *IEEE Trans. Inf. Theory* **52** 1289–306

[26] Petra S and Schnörr C 2009 TomoPIV meets compressed sensing *Pure Math. Appl.* **20** 49–76

[27] Petra S and Schnörr C 2014 Average case recovery analysis of tomographic compressive sensing *Linear Algebra Appl.* **441** 168–98

[28] Jørgensen J S and Sidky E Y 2015 How little data is enough? Phase-diagram analysis of sparsity-regularized x-ray computed tomography *Phil. Trans. R. Soc.* A **373** 20140387

[29] Sidky E Y, Anastasio M A and Pan X 2010 Image reconstruction exploiting object sparsity in boundary-enhanced x-ray phase-contrast tomography *Opt. Express* **18** 10404–22

[30] Jørgensen J S, Kruschel C and Lorenz D A 2015 Testable uniqueness conditions for empirical assessment of undersampling levels in total variation-regularized x-ray CT *Inverse Problems Sci. Eng.* **23** 1283–305

[31] Jørgensen J S, Sidky E Y, Hansen P C and Pan X 2015 Empirical average-case relation between undersampling and sparsity in x-ray CT *Inverse Problems Imaging* **9** 431–46

[32] Jørgensen J S, Sidky E Y and Pan X 2013 Quantifying admissible undersampling for sparsity-exploiting iterative image reconstruction in x-ray CT *IEEE Trans. Med. Imaging* **32** 460–73

[33] Donoho D L and Tanner J 2010 Precise undersampling theorems *Proc. IEEE* **98** 913–24

[34] Amelunxen D, Lotz M, McCoy M B and Tropp J A 2014 Living on the edge: phase transitions in convex programs with random data *Inf. Inference* **3** 224–94

[35] Barrett H H and Myers K J 2004 *Foundations of Image Science* (Hoboken, NJ: Wiley)

[36] Coban S B 2015 Sophiabeads datasets project documentation S B and tutorials MIMS EPrint 2015-26 Manchester Institute for Mathematical Sciences, The University of Manchester UK April available online http://eprints.ma.man.ac.uk/2290/

[37] Wadell H 1935 Volume, shape, and roundness of quartz particles *J. Geol.* **43** 250–80

[38] Han X, Bian J, Ritman E L, Sidky E Y and Pan X 2012 Optimization-based reconstruction of sparse images from few-view projections *Phys. Med. Biol.* **57** 5245–73

[39] Jensen T L, Jørgensen J H, Hansen P C and Jensen S H 2012 Implementation of an optimal first-order method for strongly convex total variation regularization *BIT* **52** 329–56

[40] Hernández-Nava E, Smith C J, Derguti F, Tammas-Williams S, Léonard F, Withers P J, Todd I and Goodall R 2015 The effect of density and feature size on mechanical properties of isostructural metallic foams produced by additive manufacturing *Acta Mater.* **85** 387–95

[41] van Aarle W, Palenstijn W J, De Beenhouwer J, Altantzis T, Bals S, Batenburg K J and Sijbers J 2015 The ASTRA Toolbox: a platform for advanced algorithm development in electron tomography *Ultramicroscopy* **157** 35–47

[42] Zeng G L and Gullberg G T 2000 Unmatched projector/backprojector pairs in an iterative reconstruction algorithm *IEEE Trans. Med. Imaging* **19** 548–55

[43] 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

[44] Sidky E Y, Jørgensen J H and Pan X 2012 Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm *Phys. Med. Biol.* **57** 3065–91

[45] Tang J, Nett B E and Chen G-H 2009 Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms *Phys. Med. Biol.* **54** 5781–804