Undersampling Raster Scans in Spectromicroscopy for reduced dose and faster measurements

OLIVER TOWNSEND,[1,2,*] SILVIA GAZZOLA,[1] SERGEY DOLGOV,[1] AND PAUL QUINN[2]

[1]*Department of Mathematics, 4W, University of Bath, Claverton Down, Bath, BA2 7AY, UK*
[2]*Diamond Light Source Ltd., Diamond House, Harwell Science and Innovation Campus, Didcot, OX11 0DE, UK*
[*]*ot280@bath.ac.uk*

Abstract: Combinations of spectroscopic analysis and microscopic techniques are used across many disciplines of scientific research, including material science, chemistry and biology. X-ray spectromicroscopy, in particular, is a powerful tool used for studying chemical state distributions at the micro and nano scales. With the beam fixed, a specimen is typically rastered through the probe with continuous motion and a range of multimodal data is collected at fixed time intervals.

The application of this technique is limited in some areas due to: long scanning times to collect the data, either because of the area/volume under study or the compositional properties of the specimen; and material degradation due to the dose absorbed during the measurement. In this work, we propose a novel approach for reducing the dose and scanning times by undersampling the raster data. This is achieved by skipping rows within scans and reconstructing the x-ray spectromicroscopic measurements using low-rank matrix completion. The new method is robust and allows for 5 to 6-fold reduction in sampling. Experimental results obtained on real data are illustrated.

1. Introduction

X-ray spectromicroscopy combines x-ray spectroscopy and x-ray microscopy to map changes in chemical state across a specimen on the micro and nano scales. These techniques have been broadly applied to problems across life and physical sciences, such as chemical engineering [1], material science [2], and biology [3]. A spectromicroscopy experiment involves measuring an $n_1 \times n_2$ grid at n_E energy levels across the absorption edge of an element of interest. The resulting $n_1 \times n_2$ spectra represents a big data challenge; to extract meaningful information they are typically analyzed using PCA and cluster analysis to reduce to a mapping of representative spectra, or to a low-rank representation of the data.

While the technique has been generally successful, the application of spectromicroscopy to in-situ studies and in areas of soft matter or biological materials is limited by two main factors. First, the total experiment time required to collect the data to a given statistical significance; second, the total radiation dose over the collection and any resulting damage to the object, or changes to the chemical state, that may occur as a result. The issue of damage due to dose and long collection times occurs across both x-ray and electron optical systems.

To alleviate this issue, a variety of approaches to reduce the number of samples has been proposed. In electron tomography, *compressed sensing* schemes have been investigated to solve the missing wedge problem, or to reduce the number of angles used [4–6]. Random sampling or jittered row sampling has also been used with in-painting to reduce dose and experiment time [7, 8]. In the x-ray regime, sparse studies are limited, but a low-rank matrix decomposition approach using PCA analysis of spectrotomography datasets has also recently been demonstrated,

which merges the angular and energy measurements to reduce overall measurement time [9].

Low-rank matrix completion is a well-known inverse problem that was widely used in sensor networks [10], computer vision [11] and medical imaging [12, 13]. See e.g. [14] for a review of those and further inverse problems. The classical formulation of matrix completion [15] aims to recover the missing elements of a low-rank matrix from an incomplete set of known entries, each of which is sampled *independently* at random (for example, uniformly).

However, there are important differences between x-ray and electron systems that make independent sampling unattractive for the reduction of the acquisition time in x-ray spectromicroscopy. An electron beam is rapidly moved across the specimen and can be blanked electrostatically at high rate to control dose, and recreate random patterns. In contrast, an x-ray beam is fixed, and the specimen is moved mechanically instead. Moreover, a mechanical chopping of the x-ray beam is needed to recreate sampling patterns. The mechanical operations in the x-ray regime limit the translation of some of these schemes from equivalent electron experiments.

The use of non-uniform sampling patterns in matrix completion has been widely studied, with the main goal of reducing the approximation error. Non-uniform sampling methods include adaptive cross interpolation [16], maximizing the volume of sampled submatrices [17], adaptive importance sampling using leverage scores [18] or application-oriented distributions [19], and supervised learning using a training dataset [13]. However, all of these methods still assume some arbitrary control over the sampling pattern rather than conventional acquisition patterns such as a raster scan.

In this paper we propose a novel approach for undersampling and reconstructing low-rank x-ray spectromicroscopy data that uses the *raster sampling pattern*. Our motivation with this work is to deliver a solution that can be deployed routinely at spectromicroscopy facilities by non-experts. This requires an approach where no intervention or tuning is needed to produce results. The method should also work with the standard raster acquisition approaches at these facilities, in particular our initial experimental focus was on optimizing fast low-dose in-situ experiments to study the evolution of battery materials over time. Scanning along a line can be carried out with a faster mechanical movement, hence the time per pixel can be reduced compared to independent or adaptive sampling. In addition, we can lower the x-ray dose on the specimen without compromising the recovery of missing entries: we develop a procedure to generate a robust raster pattern such that each row and column of the matrix has at least one sampled element.

Besides the sampling pattern, the performance of matrix completion depends on the cost function. In addition to the squared misfit of the sampled elements, the cost function may include regularization terms, promoting sparsity [13] or smoothness [20]. Alternatively, the entire completion problem can be turned into a Bayesian inference problem by introducing a prior distribution on the low-rank matrix factors, treated as random matrices, and a likelihood of the observed data samples [21, 22]. As a by-product, sparsity-promoting priors may provide an automatic selection of the rank as the number of nonzero posterior components [23]. However, mathematical study of spectromicroscopy is still in some infancy, and lacks well-recognised priors. The only regularization assumption that is generally valid is the low-rankness of the true absorption distribution. Therefore, we start with a simple Alternating Steepest Descent algorithm (ASD) [24] with the squared misfit cost only. For a reliable selection of the rank and number of samples, we propose an algorithm, LoopedASD, which successively increases the rank from 1 to some generous value (e.g. 20), taking a lower-rank result as the initial guess in each step. This provides a smoother convergence which allows the KNEEDLE [25] algorithm (already used in spectromicroscopy for the PCA analysis) to determine the final rank accurately.

Five datasets were used in this study. The first three (labelled *DS1, DS2, DS3*) are full datasets that can be undersampled numerically after the experiment; the final two (labelled *DS4, DS5*) are measured using both full and sparse raster patterns to verify the experimental implementation.

The sparse experimental measurements were conducted at a range of undersampling ratios (details on the different sampling approaches can be found in Section 3). The specimen were produced by mixing Fe_2O_3, Fe_3O_4 and FeO powders. The powders were ground in a ball mill and then drop cast onto a silicon nitride membrane. The dimensions of each data set can be seen in Table 1. The pairs (DS1, DS2) and (DS4, DS5) are data derived from the same specimen but at different spatial dimensions, i.e. DS2 is DS1 focused on a smaller area, as is DS4 of DS5. We aim to illustrate results using all 5 datasets, but where space is restrictive we only show the 3 independent datasets: DS1, DS3, DS5.

Table 1. Spectromicroscopy Meta Data

Dataset	Data Acquisition	n_E	n_1	n_2
DS1	Full	149	101	101
DS2	Full	150	92	79
DS3	Full	152	55	54
DS4	Full & Sparse	152	40	40
DS5	Full & Sparse	152	80	80

The paper is organised as follows. We begin in Section 2 with a description of the x-ray spectromicroscopy model, and derive its low rank nature. Next, we discuss general matrix completion and raster-aware sampling, which leads into our proposed matrix completion algorithm in Sections 3 & 4. Finally, in Sections 5 & 6, we discuss the results comparing the reconstructed sparse data against full data.

2. Low rank model

When the energy of a photon increases beyond the binding energy of a core electron, we see a sharp rise in a material's absorption - an *absorption edge*. The absorption coefficient, μ, will vary or be modulated by the local chemical environment. Measuring around the edge, x-ray near edge absorption spectroscopy (XANES) can be used as a fingerprinting tool to identify known standards or materials; investigation of the energy past the absorption edge, the Extended X-ray Absorption Fine Structure (EXAFS), can be used to extract information of nearest neighbour bonding and coordination. Further details regarding the different experimental setup and derivation of the following formula can be found in [26].

The variation of μ with energy depends on how an element is chemically bonded and when there are more than one chemical states present the absorption coefficients sum linearly. For a mixture of S materials, the measured x-ray absorption (or optical density), $D(E)$, can be modeled as,

$$D(E) = \sum_{s=1}^{S} \mu_s(E)\, t_s, \tag{1}$$

where $\mu_s(E)$ and t_s, $s = 1, ..., S$ are the distinct absorption coefficients (cm^{-1}), and thicknesses (cm) of the S materials, respectively.

With a focused x-ray beam, this experiment can be performed over $n_1 \times n_2$ positions, and at n_E distinct energies. By stacking each spatial scans, the distribution of x-ray absorption spectra can now be represented as a 3D tensor $D \in \mathbb{R}^{n_E \times n_1 \times n_2}$,

$$D_{ij_1j_2} = \sum_{s=1}^{S} \mu_s(E_i)(t_s)_{j_1j_2}\ , \tag{2}$$

where $(t_s)_{j_1 j_2}$ is the thickness of material s at pixel (j_1, j_2), and $\mu_s(E_i)$ is the absorption coefficient of material s at the $i-th$ energy level. This dataset can be flattened by considering a matrix $A \in \mathbb{R}^{n_E \times N}$, such that

$$A_{ij} = \sum_{s=1}^{S} \mu_s(E_i) t_{sj} + \eta_{ij} \ , \tag{3}$$

where $j = 1, ..., N,\ N = n_1 n_2$ indexes over all pixels, and $\eta \in \mathcal{N}(0, \delta^2 I)$ represents Gaussian noise with standard deviation $\delta \in \mathbb{R}$. With a slight abuse of notation, consider the matrices $\mu \in \mathbb{R}^{n_E \times S},\ \mu_{is} = \mu_s(E_i)$ and $t \in \mathbb{R}^{S \times N}$; the columns of μ represent the absorption coefficients of each material within the specimen, and the rows of t represent the thickness/presence of each material within each pixel. This allows us to write Eq. (3) as

$$A = \mu t + \eta. \tag{4}$$

This is significant, as it illustrates that for the majority of experiments, where the specimen is made up of only a few materials or chemical states (S is low), the corresponding spectromicroscopy data is inherently *low rank*, as $\text{rank}(\mu t) = S$. With the addition of noise, A is approximately low rank.

To analyse the spatial distribution of the absorption spectra, standard techniques are used to filter and decompose A back to a smaller set of representative spectra; see, for instance, [27]. First, PCA is applied to reduce the noise in the data by producing a rank-L approximation of the most significant components. We compute

$$A' = C' R', \tag{5}$$

with $A' \in \mathbb{R}^{n_E \times N},\ C' \in \mathbb{R}^{n_E \times L},\ R' \in \mathbb{R}^{L \times N}$. The variable L is chosen to capture as much variation in the data as possible with the smallest rank, and should approximate S: it is typically set to be the elbow point (point of maximum curvature) of the singular values of A, and can be selected automatically using algorithms like KNEEDLE [25].

The decomposition in Eq. (5) is similar to the model in Eq. (3), however the columns of C' are abstract spectra - linear combinations of the true spectra with no physical interpretation. Similarly, the rows of R' are the corresponding abstract thickness maps. Pixels are now clustered together based on the similarity of the normalised mixing factors of the PCA components. This is achieved by clustering the columns of R' using standard clustering algorithms such as kmeans [28] and lvq [29]. Taking the mean spectrum from the columns of A for each cluster increases the signal to noise ratio when compared to the measurements from an individual pixel, and produces accurate x-ray absorption spectra for the dominant material in each cluster.

In [27], it is noted that the first principal component from the PCA (the component with the greatest variation in the data) often describes the average x-ray absorption data across the whole specimen. In some cases the first principal component is discarded, or scaled down, and cluster analysis is only applied to the remaining components. This is done to emphasise the more subtle features and variations in the data and ensure the clustering results are determined by differences in materials not the thickness of the specimen. In this paper, we will refer to the process of discarding the first principal component as Reduction of Thickness Effect (RTE), which is achieved by simply removing the first row from R' in Eq. (5) before applying kmeans to its columns. Generally, the comparison between reconstructed and full data are worse when RTE is used, and they have been applied to several tests to provide the worst-case results. Any cluster results that have used RTE will be noted clearly. Further details on PCA, cluster analysis and RTE can be found in the supplemental document.

Figure 1 shows a schematic of the sparse spectromicroscopy process. In our proposed scheme, we measure only a small proportion of the data, and recover the missing entries later using low rank

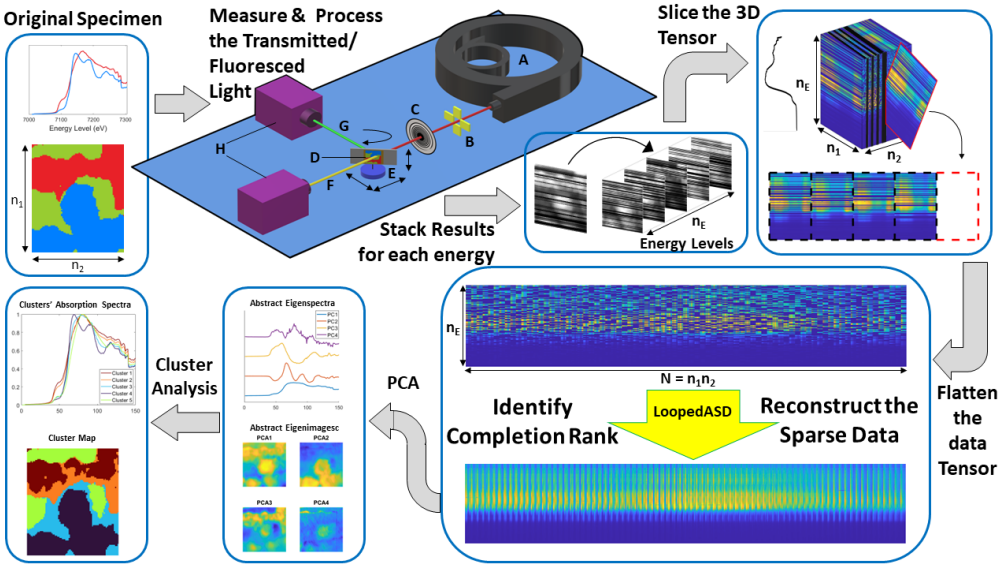


Fig. 1. Schematics of Sparse Spectromicroscopy using DS2. The original specimen contains a mixture of two materials: FeO in the blue region and Fe_2O_3 in the red region over the background (green region). The experimental setup is as follows: (A) a third/fourth generation synchrotron light source produces a beamline of energy E_i. (B) the intensity of the incident beam is measured. (C) the light is focused using an optic such as a zone plate or Kirkpatrick-Baez mirror to produce a micro or nanoscale beam. (D) The specimen is mounted on the stage, (E) which is able to move the specimen across in the XY plane to measure at different positions; the specimen can also be moved in the Z direction to bring the sample to the focus of the beam. Depending on the material and it's thickness, we measure *either* the transmitted flux, (F), *or* the fluoresced flux, (G), by counting photons using different detectors (H). The specimen is moved in a raster pattern; for sparse experiments we only sample pn_1 rows for each energy. The processed results for each energy are stacked as seen in the diagram. The tensor is flattened by concatenating slices of the sparse data. The missing entries are recovered using LoopedASD - our low rank matrix completion algorithm. PCA is applied to the completed data to produce the most significant components in the data - the eigenspectra and (reshaped) eigenimages seen in the schematic. The results of the cluster analysis with 5 clusters show similar results to the original specimen, with the locations of the two materials identified, and accurate absorption coefficients for each.

matrix completion. The remaining steps are consistent with a standard x-ray spectromicroscopy scheme. Precise details on the sampling are discussed in Section 3, and the completion methods are described in Section 4. The image illustrates the data acquisition, the formatting and flattening of the data tensor, the reconstruction of the sparse entries, and the PCA and cluster analysis.

3. Raster-aware sampling patterns

To formulate the sampling and reconstruction of x-ray spectromicroscopy data, we set out the following notation. Let $\Omega \subset \{1, ..., n_E\} \times \{1, ..., N\}$ be the set of known measurements, called the **sampling pattern**. For a matrix $X \in \mathbb{R}^{n_E \times N}$, we define the **sampling operator** $\mathcal{P}_\Omega$ as

$$\mathcal{P}_\Omega(X) = \begin{cases} X_{i,j}, & \text{if } (i,j) \in \Omega, \\ 0, & \text{if } (i,j) \notin \Omega. \end{cases} \tag{6}$$

Alternatively, the sampling pattern can be thought of as a binary matrix $\Omega \in \{0,1\}^{n_E \times N}$ with 1s in the locations of the known entries, and 0s everywhere else. It is then easy to compute $\mathcal{P}_\Omega(X) = \Omega \circ X$, where $\circ$ is the Hadamard product. The key parameter for matrix completion is the **undersampling ratio**, p, the proportion of known entries:

$$p = \frac{|\Omega|}{n_E N}. \tag{7}$$

The standard matrix completion problem involves computing a matrix of minimal rank such that the known entries, indexed by Ω, are equal to the given values. However, this problem is notoriously difficult, and many algorithms will instead seek to solve easier, but provably related, problems; see [30–33].

For this application, the presence of noise means the spectromicroscopy datasets are only approximately low rank - i.e. there is no exact low rank matrix that would perfectly match the subset of known entries $\mathcal{P}_\Omega(A)$. A more useful and efficient approach to reconstruction is to fix the rank, r, then solve an optimisation problem to find the best rank-r approximation to the known entries. Thus, to reconstruct the sparse set of measurements, $\mathcal{P}_\Omega(A)$, we seek to solve,

$$\min_{Z \in \mathbb{R}^{n_E \times N}} \frac{1}{2} ||\mathcal{P}_\Omega(A) - \mathcal{P}_\Omega(Z)||_F^2 \qquad \text{subject to} \qquad \text{rank}(Z) = r. \tag{8}$$

Here, we use the frobenius norm $||\cdot||_F$, defined for $X \in \mathbb{R}^{n_E \times N}$,

$$||X||_F^2 = \sum_{i,j} X_{ij}^2. \tag{9}$$

This approach is generally more effective (and far more computationally efficient) than attempting to solve the standard problem; we must, however, correctly input the completion rank r, since completion algorithms work best when r is close to the approximate rank of the full dataset. More details on accurately setting r can be found in the supplementary material.

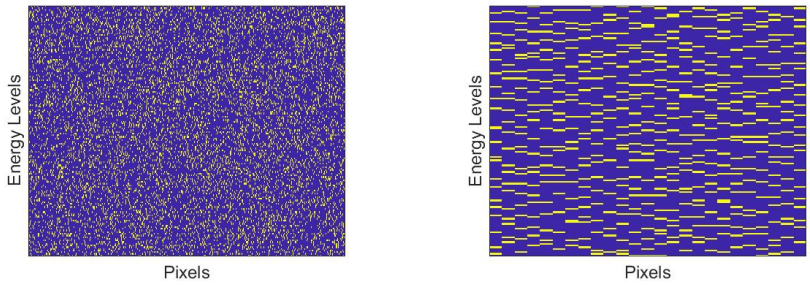
3.1. Setting the sampling pattern

In many applications of low rank matrix completion, it is impossible to predict which entries will be known. To model this, the known entries of the sampling pattern are typically set at random. Usually, *Bernoulli sampling* is used, where each entry is sampled i.i.d. (independent and identically distributed) with probability p.

Unlike other settings, in x-ray spectromicroscopy we have complete control over the data acquisition and can set the scanner to implement specific patterns. However, when formulating the sampling selection, we must consider the physical restrictions of the experiment. For x-ray spectromicroscopy, the specimen is on an XY stage and moves at constant velocity through the beam in a *raster pattern* - scanning across each row in turn before switching to the next energy level and repeating the spatial scan. To maximise the efficiency of the modelled sparse scans, we must ensure the known entries of the sampling pattern are collected together into spatial rows to be scanned. The scanner can then move between known entries, quickly passing over the empty rows.

To model the raster aligned patterns, we introduce *Raster sampling*, where spatial rows of the specimen are sampled i.i.d with probability p. Illustrations of both Bernoulli and Raster sampling patterns on flattened data sets can be found in Figure 2, in which the known entries are highlighted yellow. Notice that the raster sampling pattern groups the known entries into the spatial rows of the specimen, which appear as short segments after the data is flattened.

Data that has been sampled with a Bernoulli pattern generally allows completion at lower undersampling ratios, because the known entries are spread more evenly. In practice, undersampling



(a) Bernoulli Sampling (b) Raster Sampling

Fig. 2. Illustrations of Bernoulli and Raster sampling patterns. Patterns are shown on flattened datasets with $n_E = 150,\ n_1 = 25,\ n_2 = 25,\ p = 0.2$. Yellow indicates the point has been sampled, while purple indicates the entry is missing.

using Bernoulli sampling is possible, however it is only preferable under certain circumstances. Sampling patterns like these can be implemented for x-ray spectromicroscopy in two ways: either the specimen is moved through every position while rapidly blanking the beam, or the specimen is only positioned at the location of each known entry using stop-start motion. The former is difficult to implement and doesn't reduce the experiment time, however may be useful for dose reduction. Conversely, despite the potential use of lower undersampling ratios, stop-start motion is generally less efficient than the continuous motion used for raster sampling, resulting in longer experiments. Indeed, for typical dwell times of around $0.01s$, the time spent accelerating, decelerating and stabilising the scanner outweighs the time saved by scanning fewer entries. Despite this, Bernoulli sampling may be applicable for specimen that require longer dwell times (around $1s$), since reducing the number of known entries has a greater impact on the experiment. Developing methods for such specimen is left for further research.

One issue with Raster sampling is that the grouping of known entries increases the probability that certain rows are not scanned at all, especially for low undersampling ratios. Clearly, it is impossible to recover a row or column of the data set A (distinct from the spatial rows and columns of the specimen) that contains no known entries, so we must ensure that every spatial-row is measured at least once. By testing different sampling patterns, we have also seen that low rank completion is more reliable when the known entries are more evenly spread across the data.

To promote the spread of data, and ensure every row and column of A contains known entries, we have developed *Robust Raster sampling*. This is a variation of Raster sampling that ensures every row of the specimen is sampled exactly once before any row can be sampled a second time. Further details on Robust Raster Sampling can be found in the supplementary document.

4. Completion algorithm

To reconstruct undersampled data, we have developed a low rank matrix completion algorithm **LoopedASD**. This is a rank-incremental algorithm based on Alternating Steepest Descent (ASD) [24]; it uses the output of one ASD-completion as the input for the next higher-rank completion. After testing several different options, [31–33], it was found that this algorithm provided more accurate and reliable results on both simulated data and raster-sampled spectromicroscopy data.

ASD is an iterative method that fixes the rank, r, of its iterates by imposing the following decomposition:

$$Z = XY, \qquad \text{for} \quad Z \in \mathbb{R}^{n_E \times N},\ X \in \mathbb{R}^{n_E \times r},\ Y \in \mathbb{R}^{r \times N}. \tag{10}$$

Thus, the optimisation of the problem in Eq. (8) now becomes,

$$\min_{X,Y} f(X,Y), \quad \text{where} \quad f(X,Y) = \frac{1}{2}||\mathcal{P}_\Omega(A) - \mathcal{P}_\Omega(XY)||_F^2, \quad X \in \mathbb{R}^{n_E \times r}, \quad Y \in \mathbb{R}^{r \times N}. \tag{11}$$

We minimise the function f by alternately optimising the components X and Y using steepest descent with exact step sizes. Indeed, by fixing one component the gradient of f can be computed easily. It is then possible to analytically compute the exact step size required to minimise f along the gradient direction. Once the factor has been updated, we alternate the fixed component and repeat the process, iterating until suitable stopping conditions are satisfied.

Let $f(X,Y)$ be written as $f_Y(X)$ for fixed Y and $f_X(Y)$ for fixed X; the corresponding gradients are written $\nabla f_Y(X)$ and $\nabla f_X(Y)$, and the exact step sizes are written $\eta_X,\ \eta_Y$. Beginning with random matrices $X_0 \in \mathbb{R}^{n_E \times r},\ Y_0 \in \mathbb{R}^{r \times N}$, we implement ASD as follows,

$$\begin{cases} \text{Fix } Y_i, \text{ compute } \nabla f_{Y_i}(X_i) \ \& \ \eta_{X_i} \\ X_{i+1} = X_i - \eta_{X_i} \nabla f_{Y_i}(X_i) \\ \text{Fix } X_{i+1}, \text{ compute } \nabla f_{X_{i+1}}(Y_i) \ \& \ \eta_{Y_i} \\ Y_{i+1} = Y_i - \eta_{Y_i} \nabla f_{X_{i+1}}(Y_i). \end{cases} \tag{12}$$

The gradients and step sizes are given below. Full derivations can be found in the supplemental documents:

$$\nabla f_Y(X) = -(\mathcal{P}_\Omega(A) - \mathcal{P}_\Omega(XY^T))Y, \qquad \nabla f_X(Y) = -X^T(\mathcal{P}_\Omega(A) - \mathcal{P}_\Omega(XY^T)). \tag{13}$$

$$\eta_X = \frac{||\nabla f_Y(X)||_F^2}{||\mathcal{P}_\Omega(\nabla f_Y(X)Y^T)||_F^2}, \qquad \eta_Y = \frac{||\nabla f_X(Y)||_F^2}{||\mathcal{P}_\Omega(X[\nabla f_X(Y)]^T)||_F^2} \tag{14}$$

One advantage of ASD is that the residual $(\mathcal{P}_\Omega(A) - \mathcal{P}_\Omega(XY))$ can be easily updated between iterations, removing the need to compute the matrix product XY for each iteration. Thus, the per iteration cost of ASD has leading order $8|\Omega|r$ (see [24]).

A maximum number of iterations was set as a stopping condition, as well as a tolerance on the relative norm of the residual at each iteration,

$$\frac{||\mathcal{P}_\Omega(A) - \mathcal{P}_\Omega(X_iY_i)||_F^2}{||\mathcal{P}_\Omega(A)||_F^2}. \tag{15}$$

Once either stopping condition is satisfied, the two factors are output by the algorithm and are denoted X^* and Y^*.

To evaluate the success of the completion algorithms, we compute the *completion error*, often denoted e_c. This is simply the relative norm of the difference between the true low rank matrix A and the output of the ASD algorithm $A^* = X^*Y^*$, defined as,

$$e_c = \frac{||A - A^*||_F^2}{||A||_F^2} \tag{16}$$

A more detailed description of the algorithm can be found in the supplemental material.

4.1. Improving results with LoopedASD

Following rigorous testing of ASD for raster sampling, it was found that there exists an approximate optimal undersample ratio, p^*, that depends linearly on, and is correlated with, the data's rank r. For $p > p^*$, the probability of a successful completion was almost certain, and the completion

errors were consistently low. For $p < p^*$, the completion errors increased and the completion rate (the proportion of tests that successfully recover data to a certain accuracy - usually 10^{-4}) decreased rapidly to zero. Intuitively, the dependence of p^* on r makes sense: to reconstruct more complex datasets (with higher ranks), more known entries are required to successfully recover the missing ones.

LoopedASD was developed to take advantage of this heuristic: beginning with $r = 1$, we use the outputs of the $r = j$ completion as initial guesses for the $r = j + 1$ completion, iterating up to the completion rank set by the user.

Let $(X_0)^{(j)} \in \mathbb{R}^{n_E \times j}$, $(Y_0)^{(j)} \in \mathbb{R}^{j \times N}$ be the initial matrices for the j^{th} rank step, and $(X^*)^{(j)} \in \mathbb{R}^{n_E \times j}$, $(Y^*)^{(j)} \in \mathbb{R}^{j \times N}$ be the outputs of the j^{th} iteration of ASD with completion rank $r = j$. In order ensure the dimensions of the factors increase with each rank increment, we simply concatenate the j^{th} outputs $(X^*)^{(j)}$, $(Y^*)^{(j)}$ with a random column and row respectively to produce $(X_0)^{(j+1)}$, $(Y_0)^{(j+1)}$.

The aim of this variation is to allow the iterates to converge quickly to a rank-1 approximation of the sparse data, where fewer known entries are required. Once a rank-j approximation is known, the distance to the rank-$(j + 1)$ solution should be relatively low, and again fewer known entries are required to converge quickly to the next one. Ensuring iterates remain close to the minimum should also avoid convergence to a spurious local minima of f (which is non-convex).

The completion results of LoopedASD confirm the validity of this heuristic, as it yields good reconstructions more reliably at low undersampling ratios.

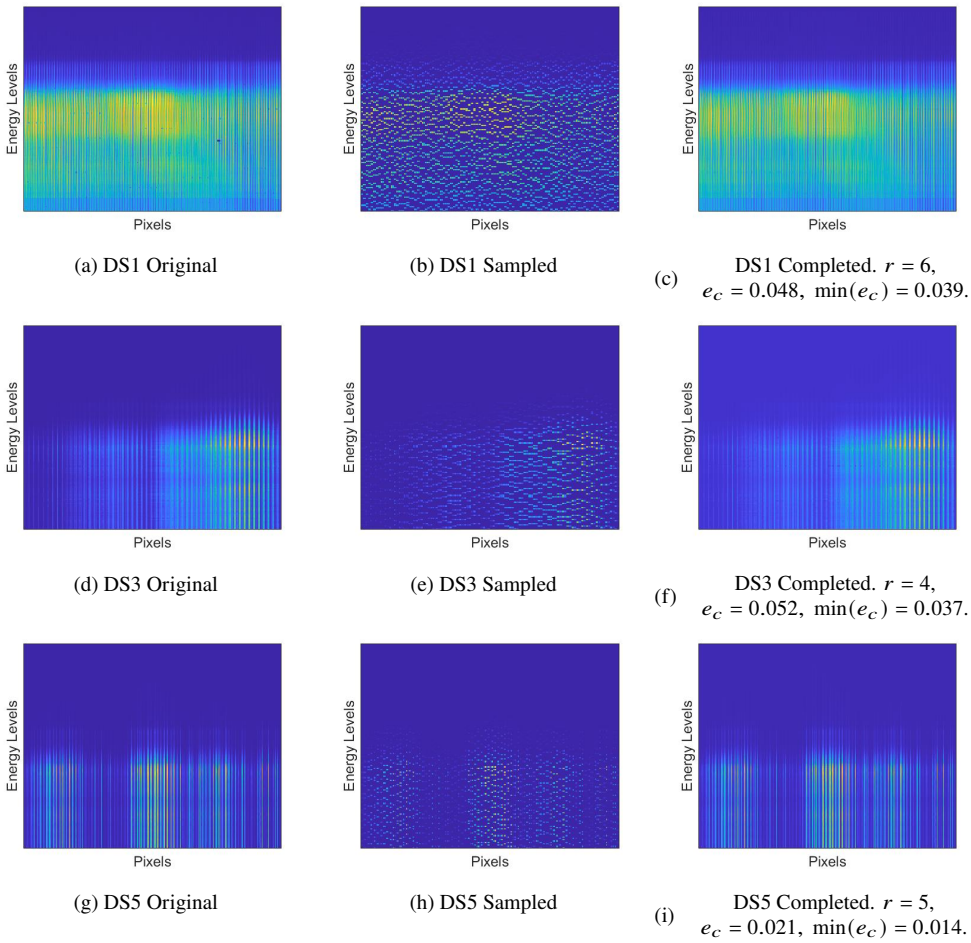
In addition to the usual stopping conditions, an early stopping procedure was required due to slow convergence rates when the iterates approached optimal solutions. In the case that the average change in the norm of the residual (Eq. (15)) over 50 iterations drops below a second tolerance (typically 10^{-5}), then the early stopping condition is satisfied and will halt the algorithm.

It has been noted that the completion rank must be set before implementing ASD and LoopedASD. This, however, is not an obvious choice since we cannot evaluate the accuracy of the different rank-k completions without prior knowledge of the results. To overcome this, we have developed a method that implements short completions over several ranks and evaluates the completion errors using cross validation so that we can efficiently identify the optimal completion rank to use. A full description and justification of the method can be found in the supplemental material.

5. Validation of the algorithm with numerical undersampling

We test the method in two distinct ways. First, we numerically sample full datasets. To produce these datasets, 'unknown' entries are set to zero according to a randomly generated robust raster sampling pattern. Then, in Section 6, we implement sparse scans on the beamline to determine if any additional issues arise and to test in a new setting outside of our test set.

To properly evaluate the performance of the methods described in this paper, several measures of success must be considered. We can compare the original and reconstructed datasets directly by computing the average completion error. Perhaps more significantly, we evaluate the difference between the cluster maps and the absorption spectra themselves. Finally, we can plot and visually compare the clusters and absorption spectra. Due to specimen drifts (described in Section 6), the first two 'computational' approaches can only be achieved using numerically sampled data; visual comparisons must be used when evaluating reconstructions using the sparse scan data. During these tests, we use the optimal completion rank r^* determined by the rank selection algorithm described in the supplemental document to produce the completed matrix $(A^*)^{(r^*)} = (X^*)^{(r^*)}(Y^*)^{(r^*)}$. A^* is then used instead of the full data set A in the analytic processes (PCA and clustering).



(a) DS1 Original (b) DS1 Sampled (c) DS1 Completed. $r = 6$, $e_c = 0.048$, $\min(e_c) = 0.039$.

(d) DS3 Original (e) DS3 Sampled (f) DS3 Completed. $r = 4$, $e_c = 0.052$, $\min(e_c) = 0.037$.

(g) DS5 Original (h) DS5 Sampled (i) DS5 Completed. $r = 5$, $e_c = 0.021$, $\min(e_c) = 0.014$.

Fig. 3. Reconstructions of the three independent datasets, using LoopedASD as the completion algorithm. Each figure is a scaled colour image of the flattened dataset, with energy levels across the vertical axis, and all pixels across the horizontal axis. The left column shows the original datasets. The middle column shows the sampled data at $p = 0.20$. The right column shows the reconstructed data. Below each reconstruction is the completion rank, r, the completion error e_c (see Eq. (16)), and the minimum approximation error (see Eq. (17)) for that completion rank.

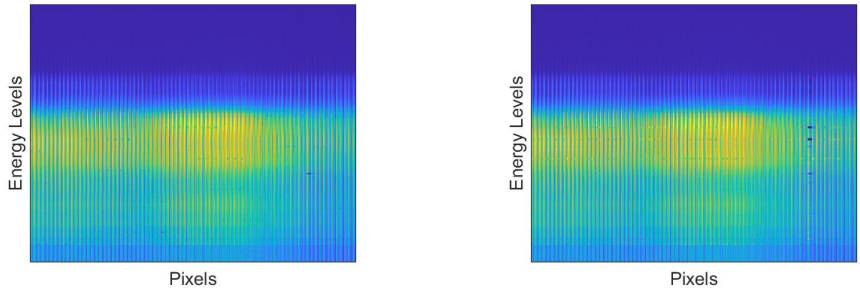
5.1. Completion errors of LoopedASD

We first examine the norm of the difference between data sets. When conducting experiments such as this, it is important to consider the constraints on the completion error, e_c (Eq. (16)), for a rank-r completion.

Consider a rank-r approximation of a matrix A, denoted $A^{(r)}$. We can compute the minimum approximation error of $A^{(r)}$ in the Frobenius norm using the Eckart-Young-Mirsky (EYM) Theorem [34]. We shall refer to this value as the minimal rank-r approximation error, and it is given by:

$$\min_{\text{rank}(A^{(r)})=r} ||A - A^{(r)}||_F = \sqrt{\sigma_{r+1}^2 + \sigma_{r+2}^2 + ... + \sigma_n^2}, \tag{17}$$

where $\sigma_1, \ ... \ , \ \sigma_n$ are the singular values (SVs) of A. Since any completion result with a



(a) DS2 Original

(b) DS2 Completed, $r = 5$, $e_c = 0.52,\ \min(e_c) = 0.026$

Fig. 4. Completion result following intentional sampling of the artifact in DS2 (dark spot towards the right). Note that the corrupted points are contained to the column of the initial artefact.

completion rank of r is just an example of a rank-r approximation, the corresponding minimum approximation errors are lower bounds for the corresponding completion errors. Thus the success of any completion should be evaluated by comparing back to the corresponding minimum error.

LoopedASD produces successful completions for all datasets, and for undersampling ratios as low as $p = 0.15$. By examining the mean completion errors, e_c, one can identify the optimal undersampling ratio, p^*. As before, for $p < p^*$ average completion errors rise quickly, and for $p > p*$ average completion errors decrease slowly, reducing the benefit of taking further measurements. It was found that LoopedASD produces lower completion errors more reliably than ASD, in particular for undersampling ratios around p^*.

To help with the visualization of the completion, we plot the flattened data sets themselves as colourmap images. Purple points indicate a value of zero, while bright yellow points indicate higher measured values. In Figure 3 we provide a visual representation of the sampling and reconstruction process at the optimal rank, r^*, and with a consistent undersampling ratio of $p = 0.20$. Note that completions often appear smoother than the original data - since it is low rank, it will have filtered out much of the noise.

One advantage of ASD and LoopedASD is that the impact of any artifacts that have been sampled remain contained in their rows/columns in the dataset. In Figure 4 we illustrate this by intentionally sampling a corrupted entry in DS2. We can see that the majority of the image remains accurate and it is just the row and column that contained the original artefact that are affected. Indeed, it is very easy to identify the corrupted regions, which can be cut from the dataset. The example in Figure 4 initially had a completion error of 0.052, compared to the mean 0.33 (the corresponding minimum approximation error is 0.026). Once the affected entries are removed, e_c is computed as 0.047 providing a much stronger result. In practice this property is very useful, since it implies the process is robust against artefacts found in the data.

5.2. Impact of completion on cluster analysis

We now compare the results of the cluster analysis from the full data and the sparse data. In some sense, this comparison is more relevant than simply comparing the completion errors because the success of the new methods will be determined by the quality of the cluster maps produced by the reconstructions, not necessarily the abstract difference between the two datasets.

We can evaluate the similarity of clusters using the Adjusted Rand Index (ARI) [35]. This is a symmetric score, with values between -1 and 1. An ARI score of 1 indicates clusters are a perfect match, while a score of 0 implies classifications have effectively been assigned at random.

Generally we use the inverted score of $(1 - \text{ARI})$ to evaluate cluster quality so that 0 indicates a perfect match, and smaller scores are generally better.

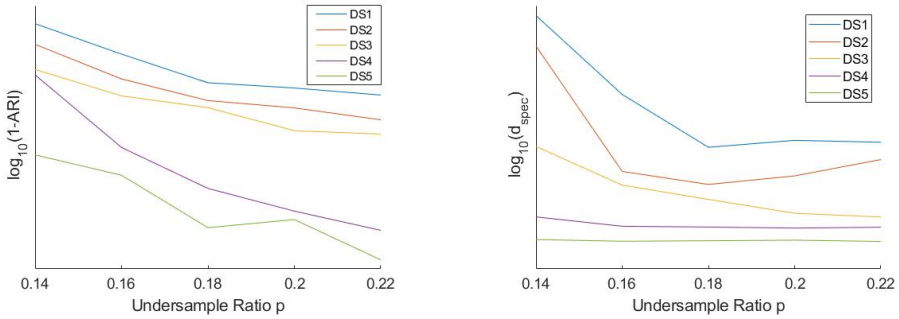
On the other hand, we can compare individual absorption coefficients using the Euclidean 2-norm, but we must ensure that we are taking the difference between coefficients for equivalent materials/clusters. Suppose we have performed cluster analysis on both the full data set and the completed data with N_{cluster} many cluster centres. Let $\mu_f \in \mathbb{R}^{n_E \times N_{\text{cluster}}}$ denote the absorption coefficients of the full data and $\mu_c \in \mathbb{R}^{n_E \times N_{\text{cluster}}}$ denote those from the completed data. In both cases, the i^{th} columns, $(\mu_f)_i$, $(\mu_c)_i$ respectively, represent the absorption coefficient corresponding to the i^{th} cluster. Note that the i^{th} clusters from the full data may not correspond to the i^{th} cluster from the completed data. We now align the clusters by taking each full-data cluster and finding the completed cluster that best fits it (minimises $(1 - \text{ARI})$); using this permutation index, we rearrange the columns of μ_c to get the aligned absorption coefficients, $\mu_a \in \mathbb{R}^{n_E \times N_{\text{cluster}}}$. We can now be sure the absorption coefficients represent the same area and the same material. We now wish to combine the relative differences of each cluster such that each cluster is weighted equally (not all absorption spectra are of the same magnitude) and is normalised for the number of clusters used. Thus, we compute the *spectral difference,* d_{spec}, as:

$$d_{\text{spec}} = \frac{1}{\sqrt{N_{\text{clusters}}}} \sqrt{\sum_{i=1}^{N_{\text{clusters}}} \left(\frac{||(\mu_f)_i - (\mu_a)_i||_2}{||(\mu_a)_i||_2} \right)^2}. \tag{18}$$

One complication is that cluster analysis is not deterministic, since there is some randomness in the starting vectors of the cluster centres. Because of this, we often find several common cluster maps resulting from the same dataset. This is true for both the full datasets and the reconstructed datasets, which can produce false ARI scores and spectral differences when opposing outputs are compared. To overcome this, we apply a clustering process to the clusters themselves - grouping similar maps together into 'Super Clusters'. Taking the mode within each super cluster produces a series of representative cluster maps that are produced by the original data. We can now compare the reconstructed cluster against each representative in turn, recording the highest score as the most appropriate comparison.

The spectromicroscopy datasets were sampled numerically, their optimal ranks were computed, and the data reconstructed using LoopedASD with early stopping. The completion ranks were set to be the optimal rank, r^*, using the method described in the supplementary document. PCA is used to decompose the data into its most significant components (we set the number of components used $L = r^*$, so the rank is consistent). Finally, we perform the cluster analysis using kmeans with 5 clusters and compute the associated absorption spectra. For DS1, DS2, DS3 we used RTE (dropping the first principal component before applying kmeans) to provide the worst-case results. Note that if RTE had not been used, both the ARI scores and the spectral difference would improve. For DS4 and DS5, we avoided using RTE due to the poorer quality of the clustering results for both the full and reconstructed data. When using RTE on DS4 and DS5, the clusters are not smooth, the images and spectra were noisy, and the outcomes were not representative of typical results.

Averaging over many iterations, we plot the clustering results for each dataset and for various undersampling ratios in Figure 5. In Figure 5a we plot $\log_{10}(1 - \text{ARI})$ against the undersampling ratio, and in Figure 5b we plot the log of the spectral difference, $log_{10}(d_{\text{spec}})$, against the undersampling ratio. The purpose of the plots is to illustrate the qualitative behaviour of the results, so have translated the individual plots vertically to improve clarity. The quantitative results can be found in Table 2. Here we see that, as we increase the undersampling ratio, the $(1 - \text{ARI})$ scores and spectral differences behave similarly to the completion error, e_c (Eq. (16)). Indeed, by plotting the logs of these scores we see the similarly shaped curves - a faster decline for lower p that flattens for higher p. Once again, we identify the optimal undersampling ratio,



(a) $\log_{10}(1 - ARI)$ against undersampling ratio, p (b) $\log_{10}(d_{\text{spec}})$ against undersample ratio, p

Fig. 5. Comparing $\log_{10}$ of the cluster results against undersampling ratio using the optimal completion rank. In each figure, *lower scores are better*. Note that the individual plots have been translated vertically to improve the clarity of the image.

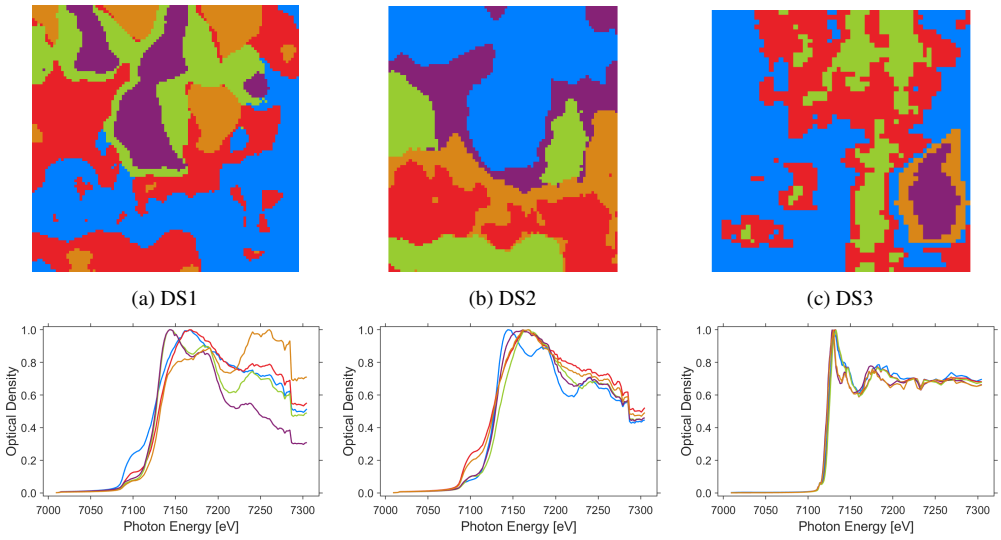
p^*, at the elbow of the plots - the points of maximum curvature. It is clear from the plot where the majority of the optimal undersampling ratios lie, however for DS2 and DS4 the elbows occur at different points across the two plots ($p = 0.16$ for Figure 5a and $p = 0.18$ for Figure 5b). In these cases, we take the higher value to ensure enough entries are known for better reliability. For DS4 and DS5, the pattern in Figure 5b is less clear and appears more as a constant value. This is simply because the variation in spectral difference over this range of undersample ratios is much smaller than for the other datasets, and by appropriately scaling these plots one can see the same curve as before.

In Table 2, we record the optimal completion rank, the corresponding optimal undersampling ratio (found at the elbow of the results in Figure 5), the completion errors, e_c, and the cluster results produced by these parameters.

Table 2. Optimal Clustering Results for numerically sampled Spectromicroscopy datasets. For the metrics used (e_c, 1 − ARI, d_{spec}), *lower scores are better*.

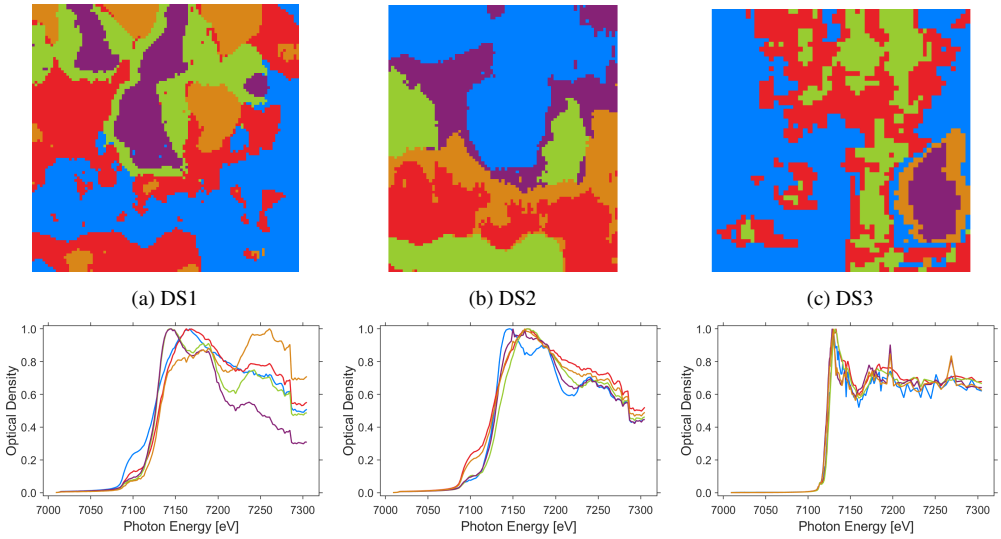
dataset	Optimal Completion Rank, r^*	Optimal undersampling ratio, p^*	Mean Completion Error (16)	RTE used	Mean (1-ARI)	Mean Spectral Difference (18)
DS1	6	0.18	0.0483	Yes	0.106	0.082
DS2	5	0.18	0.0326	Yes	0.108	0.115
DS3	4	0.20	0.0515	Yes	0.207	0.509
DS4	7	0.18	0.025	No	0.009	0.096
DS5	5	0.18	0.027	No	0.006	0.152

In Figures 6 & 7, we visually compare the cluster maps and absorption spectra of the full data and the reconstructed data for DS1, DS2 and DS3 (DS4 and DS5 are used in Section 6 to illustrate the sparse scans). The data sets are sampled numerically with $p = 0.20$, the optimal completion rank was used, and each of these results uses RTE. These images were produced by Mantis X-ray [36], an open source software package used for analysing x-ray spectromicroscopy data.



(a) DS1 (b) DS2 (c) DS3

Fig. 6. Cluster and spectral results from full datasets using RTE. Results computed using Mantis-xray [36].



(a) DS1 (b) DS2 (c) DS3

Fig. 7. Cluster and spectral results from reconstructions of sampled data, with $p = 0.20$, using RTE. Results computed using Mantis-xray [36]. Please note that for DS1 and DS2 there is a sharp downward going discontinuity at around $7270eV$ in both the full and completed results. This feature is present in the original data, and can be seen in the absorption spectra of many of the pixels with stronger signals. It is likely due to an error in the calibration of the energy stepping, which requires coordinated motion of an insertion device and monochromator, resulting in a strong variation in the intensity of the beam which was not normalised correctly.

We can see that both the cluster maps and absorption spectra are very similar between full and reconstructed data, with only some small (mostly insignificant) differences in each. The general outline of the reconstructions' clusters are very clear and are consistent with the full data results. The boundaries of the completion clusters lose some of their smoothness, but only a few isolated

Table 3. Time and efficiency (ratio of sparse and full scan times) of implementing sparse scans during experiments.

p	T $(hrs:min:s)$ - DS4	T $(hrs:min:s)$ - DS5	Efficiency - DS4	Efficiency - DS5
1.00	1:22:05.5	4:10:36.7	1.00	1.00
0.35	0:38:57.6	1:39:31.6	0.47	0.40
0.25	0:33:29.8	1:21:33.8	0.41	0.33
0.15	0:26:02.3	0:50:37.9	0.32	0.20

pixels have been misclassified with minimal impact to the overall image. Similarly the most important features of the absorption spectra have been preserved in the completions for DS1 and DS2, including the pre-edges, the peaks, and in particular the shifts in the absorption edges for each material. There are a few examples of noise affecting the data: DS3 is particular noisy, where for some clusters we see the sharp spikes indicative of noise. Despite this, the general outline is still consistent.

Overall, the similarity of the clusters ensures the sparse data would be interpreted in the same way as the original data, and the absorption spectra are sufficiently clear to be able to identify the materials within each cluster.

6. Practical sparse scanning experiments

We now bring all of the above material together to implement sparse scanning in practice. Instead of numerically sampling a full dataset, we physically implemented a sparse robust raster sampling pattern. This was done on the same specimen, with the same pixel sizes and dwell times for $p \in \{0.15,\ 0.25,\ 0.35,\ 1\}$.

Unfortunately, it is not possible to numerically compare the results of the data due to experimental variations between the known entries in each dataset. Small changes in the ambient temperature can cause the specimen to drift very small distances between each spatial scan during the experiment. Since measurements are being taken on the micro/nano-scopic scale, this effect can create pixel level change from start to end. This is usually compensated for during the stacking process, but sparse scans have a much shorter experimental time, and registration to correct drifts cannot be performed in the same way as full datasets. The intensity of the incident beam will also vary with time and, although this can be normalized, some variation can still occur. Because of these potential spatial discrepancies and changes in intensity and background noise, the same known entries across different scans will show different values. Thus, it is impossible to know whether the difference between the full and reconstructed data is because of differences in the measurements or due to the limitations of the completion algorithm.

Despite this, we can still *visually* compare the results. For each of our sparse scans (DS4 and DS5) and at each undersampling ratio, we use LoopedASD to reconstruct the data. We equip the algorithm with the optimal completion rank for the data set, described in the supplementary documents. We use Mantis X-ray to perform the cluster analysis, setting $L = r^*$ and using 5 clusters. Recall that RTE was **not** used for these results. In Figures 8 and 9, we plot these results.

Once again, despite a slight loss of smoothness around cluster boundaries and some individual miss-classified pixels, the overall structure of the cluster maps is near identical and clearly shows the same variations across the scans. Similarly, there are a few sharp discontinuities in the reconstructed absorption spectra, especially for lower undersample ratios, but each materials' XANES are still clearly identifiable.

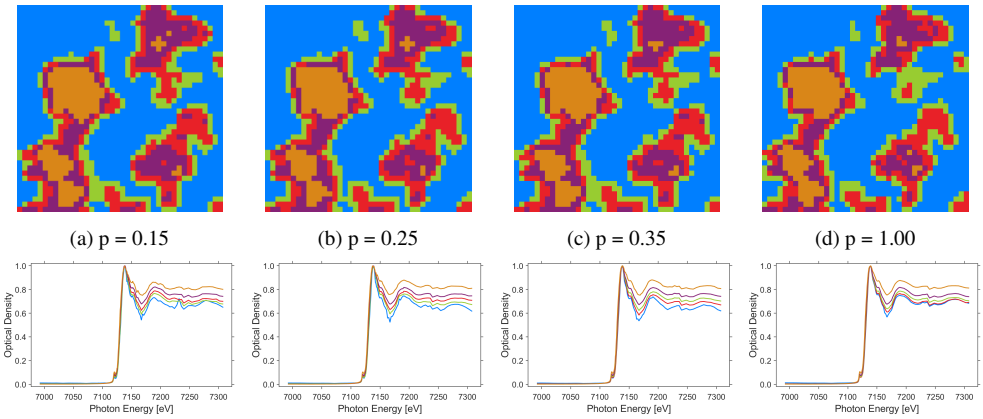


Fig. 8. Cluster and spectral results of sparse scanning for DS4. Measurements were taken at 15%, 25%, 35%, and 100% respectively. Cluster results computed using Mantis-xray [36]

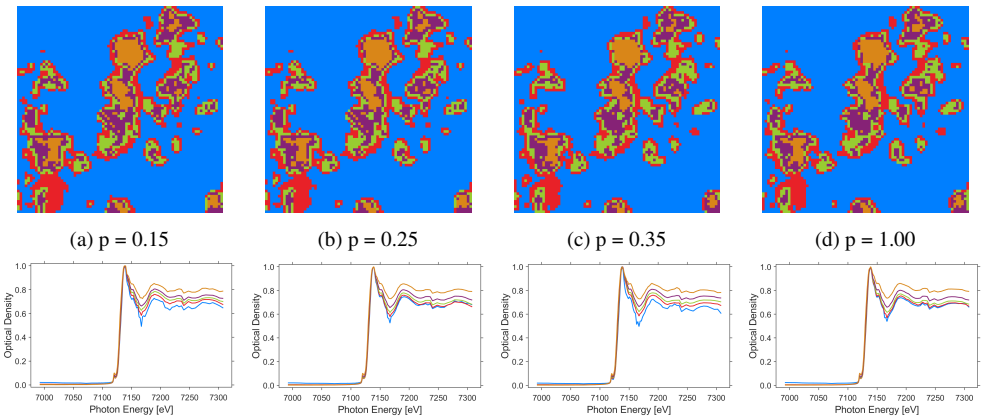


Fig. 9. Cluster and spectral results of sparse scanning for DS5. Measurements were taken at 15%, 25%, 35%, and 100% respectively. Cluster results computed by Mantis-xray [36].

After extracting the timestamps from the files' meta data, we can compare the run time for each sparse experiment, which have been summarised in Table 3.

The true time efficiency will never exactly match the undersampling ratio used due to 'dead time' as the scanner resets and for critical machine processes. Despite this, we see huge gains for the lowest undersample ratios, and a significant increase in experimental efficiency. In particular, we see greater improvements for the larger region of DS5, due to the higher measurement to dead time ratio. Another benefit for scanning larger regions is that the completion algorithms become more efficient. Because it is approximately low rank, the total number of entries grows much faster than the number of degrees of freedom - i.e. larger datasets are recoverable at lower undersampling ratios. Thus, it is in the interest of researchers using this approach to scan larger areas to improve both the scanning efficiency, and the quality of reconstruction.

7. Conclusion

We have demonstrated a new undersampling approach for spectromicroscopy data acquisition. By taking advantage of the inherent low rank structure of the datasets, we have produced algorithms that can accurately recover unknown entries from as little as 15% of the measurements when raster sampling. We have illustrated the robustness of these algorithms to machine artefacts, and how to derive parameters like the completion rank from the sparse data itself.

Finally, we showed the minimal impact reconstructing sparse data has on the cluster analysis that is currently used to interpret the spectromicroscopy data. By implementing these methods, we can conduct experiments at a much faster rate, over larger areas, and with lower dose on the sample: the method consistently produces near-identical results using 20% of the measurements, with potential improvements to 15% for larger samples. Alternate sampling schemes, compatible with continuous motion scans, may be able to produce further reductions. These improvements will help in the development of in-situ spectro-microscopy measurements and future developments will reduce times in areas such as XANES nano-tomography and nano-EXAFS, which are currently limited in their application by long acquisition times.

8. Backmatter

Funding. This research is partially supported by a University of Bath scholarship via the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa) under project EP/S022945/1.

This research is partially supported by Diamond Light Source Plc.

Content in the funding section will be generated entirely from details submitted to Prism. Authors may add placeholder text in the manuscript to assess length, but any text added to this section in the manuscript will be replaced during production and will display official funder names along with any grant numbers provided. If additional details about a funder are required, they may be added to the Acknowledgments, even if this duplicates information in the funding section. See the example below in Acknowledgements.

Acknowledgments. We acknowledge Diamond Light Source for time on Beamline/Lab I14 under Proposal MG31039.

The authors gratefully acknowledge the University of Bath's Research Computing Group (doi.org/10.15125/b6cd-s854) for their support in this work, see [37].

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

References

1. A. Braun, "Carbon speciation in airborne particulate matter with c (1s) nexafs spectroscopy." J. Environ. Monit. **7**, 1059–1065 (2005).
2. E. C. Miller, R. M. Kasse, K. N. Heath, B. R. Perdue, and M. F. Toney, "Operando spectromicroscopy of sulfur species in lithium-sulfur batteries." J. Electrochem. Soc. **165**, A6043–A6050 (2017).
3. A. P. Hitchcock, C. Morin, X. Zhang, T. Araki, J. Dynes, H. Stover, J. Brash, J. R. Lawrence, and G. G. Leppard, "Soft x-ray spectromicroscopy of biological and synthetic polymer systems," J. Electron Spectrosc. Relat. Phenom. **144-147**, 259–269 (2005).
4. R. Leary, Z. Saghi, P. A. Midgley, and D. J. Holland, "Compressed sensing electron tomography," Ultramicroscopy **131**, 70–91 (2013).
5. M. D. Guay, W. Czaja, M. A. Aronova, and R. D. Leapman, "Compressed sensing electron tomography for determining biological structure," Sci. Reports **6**, 1–14 (2016).
6. A. Cossa, V. Arluison, and S. Trépout, "Sparse cryo-STEM tomography for biological samples," Microsc. Microanal. **27**, 3028–3030 (2021).
7. L. Kovarik, A. Stevens, A. Liyu, and N. D. Browning, "Implementing an accurate and rapid sparse sampling approach for low-dose atomic resolution STEM imaging," Appl. Phys. Lett. **109**, 164102 (2016).
8. N. D. Browning, A. Stevens, L. Kovarik, A. Liyu, B. L. Mehdi, H. Yang, M. E. Gehm, B. Stanfill, S. Reehl, and L. Bramer, "Implementing Sparse Sub-Sampling Methods for Low-Dose/High Speed STEM," Microsc. Microanal. **24**, 1952–1953 (2018).

9. Z. Gao, M. Odstrcil, S. Böcklein, D. Palagin, M. Holler, D. F. Sanchez, F. Krumeich, A. Menzel, M. Stampanoni, G. Mestl, J. A. van Bokhoven, M. Guizar-Sicairos, and J. Ihli, "Sparse ab initio x-ray transmission spectrotomography for nanoscopic compositional analysis of functional materials," Sci. Adv. **7**, 1–12 (2021).
10. S. Oh, A. Montanari, and A. Karbasi, "Sensor network localization from local connectivity: Performance analysis for the mds-map algorithm," in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo),* (2010), pp. 1–5.
11. H. Ji, C. Liu, Z. Shen, and Y. Xu, "Robust video denoising using low rank matrix completion," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* (2010), pp. 1791–1798.
12. Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," in *2007 Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging,* (2007), pp. 181–182.
13. F. Sherry, M. Benning, J. C. De los Reyes, M. J. Graves, G. Maierhofer, G. Williams, C.-B. Schönlieb, and M. J. Ehrhardt, "Learning the sampling pattern for mri," IEEE Transactions on Med. Imaging **39**, 4310–4321 (2020).
14. P. C. Hansen, *Discrete Inverse Problems* (Society for Industrial and Applied Mathematics, 2010).
15. E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," Foundations Comput. Math. **9**, 717 (2009).
16. E. E. Tyrtyshnikov, "Incomplete cross approximation in the mosaic–skeleton method," Computing **64**, 367–380 (2000).
17. S. A. Goreinov and E. E. Tyrtyshnikov, "The maximal-volume concept in approximation by low-rank matrices," Contemp. Math. **280**, 47–51 (2001).
18. P. Drineas, M. Magdon-Ismail, M. W. Mahoney, and D. P. Woodruff, "Fast approximation of matrix coherence and statistical leverage," J. Mach. Learn. Res. **13**, 3475–3506 (2012).
19. S. Labouesse, S. C. Johnson, H. A. Bechtel, M. B. Raschke, and R. Piestun, "Smart scattering scanning near-field optical microscopy," ACS Photonics **7**, 3346–3352 (2020).
20. P. Kurschner, S. Dolgov, K. D. Harris, and P. Benner, "Greedy low-rank algorithm for spatial connectome regression," J. Math. Neurosc. **9**, 9 (2019).
21. S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse bayesian methods for low-rank matrix estimation," IEEE Transactions on Signal Process. **60**, 3964–3977 (2012).
22. C. Menzen, M. Kok, and K. Batselier, "Alternating linear scheme in a bayesian framework for low-rank tensor approximation," SIAM J. on Sci. Comput. **44**, A1116–A1144 (2022).
23. C. Hawkins, X. Liu, and Z. Zhang, "Towards compact neural networks via end-to-end training: A bayesian tensor approach with automatic rank determination," SIAM J. on Math. Data Sci. **4**, 46–71 (2022).
24. J. Tanner and K. Wei, "Low rank matrix completion by alternating steepest descent methods," Appl. Comput. Harmon. Analysis **40**, 417–429 (2016).
25. V. Satopaa, J. Albrecht, D. Irwin, and B. Raghavan, "Finding a "kneedle" in a haystack: Detecting knee points in system behavior," in *2011 31st International Conference on Distributed Computing Systems Workshops,* (2011), pp. 166–171.
26. M. Newville, "Fundamentals of xafs," Rev. Mineral. Geochem. **78**, 33–74 (2014).
27. M. Ierotic, C. Jacobsen, T. Schafer, and S. Vogt, "Cluster analysis of soft x-ray spectromicroscopy data," Ultramicroscopy **100.1**, 35–57 (2004).
28. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res. **12**, 2825–2830 (2011).
29. T. Kohonen, *Self-Organization and Associative Memory* (Springer Berlin, 1989).
30. E. J. Candès and B. Recht, "Exact matrix completion via convex optimisation," Foundations Comput. Math. **9**, 717–772 (2009).
31. J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," SIAM Soc. for Ind. Appl. Math. **20**, 1956 – 1982 (2010).
32. E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principle component analysis?" J. AMC **58**, 1 – 37 (2011).
33. R. Meka, P. Jain, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," (2009).
34. C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," Psychometrika **1**, 211 –218 (1936).
35. W. M. Rand, "Objective criteria for the evaluation of clustering methods." J. Am. Stat. Assoc. **66.336**, 846–850 (1971).
36. M. Lerotic, R. Mak, S. Wirick, F. Meirer, and C. Jacobsen, "Mantis xray," MANTiS: a program for the analysis of X-ray spectromicroscopy data. Created 2014, Sep. Accessed: 15/04/2021.
37. U. of Bath, "Research computing group," .