

Compressed Sensing in Imaging Mass Spectrometry

Andreas Bartels¹, Patrick Dülk¹, Dennis Trede^{1,2},
Theodore Alexandrov^{1,2,3} and Peter Maaß^{1,2}

¹Center for Industrial Mathematics, University of Bremen, Bibliothekstr. 1,
28359 Bremen, Germany.

²Steinbeis Innovation Center SCiLS Research, Fahrenheitstr. 1, 28359 Bremen,
Germany.

³MALDI Imaging Lab, University of Bremen, Leobener Str./NW2, 28359
Bremen, Germany.

E-mail: bartels@math.uni-bremen.de, pduelk@math.uni-bremen.de,
trede@math.uni-bremen.de, theodore@uni-bremen.de,
pmaass@math.uni-bremen.de

Abstract. Imaging mass spectrometry (IMS) is a technique of analytical chemistry for spatially-resolved, label-free and multipurpose analysis of biological samples, which is able to detect spatial distribution of hundreds of molecules in one experiment. The hyperspectral IMS data is typically generated by a mass spectrometer analyzing the surface of the sample. In this paper, we propose a compressed sensing approach to IMS which potentially allows for faster data acquisition by collecting only a part of pixels in the hyperspectral image and reconstructing the full image from this data. We present an integrative approach to perform both peak-picking spectra and denoising m/z -images simultaneously, whereas the state of the art data analysis methods solve these problems separately. We provide a proof of the robustness of the recovery of both spectra and individual channels of the hyperspectral image and propose an algorithm to solve our optimization problem which is based on proximal mappings. The paper concludes with numerical reconstruction results for a IMS dataset of a rat brain coronal section.

PACS numbers: 02.30.Zz, 82.80.Rt, 87.85.Ng, 07.05.Pj, 02.60.Cb

AMS classification scheme numbers: 47A52, 68U10, 94A12, 49N30, 49N45

1. Introduction

1.1. Mass spectrometry

Mass spectrometry is a widespread technique of analytical chemistry used to determine the molecular composition of a biological or chemical sample. The way this task is accomplished is through experimental measurement of the mass-to-charge ratio of gas-phase ions produced from molecules from the underlying analyte.

Apart, mass spectrometry is a technique of choice in various fields of biology and medicine. Among prominent applications are protein sequencing and discovery of novel biomarkers in urine, serum, or blood for such diseases as cancer, diabetes and neurodegenerative disorders.

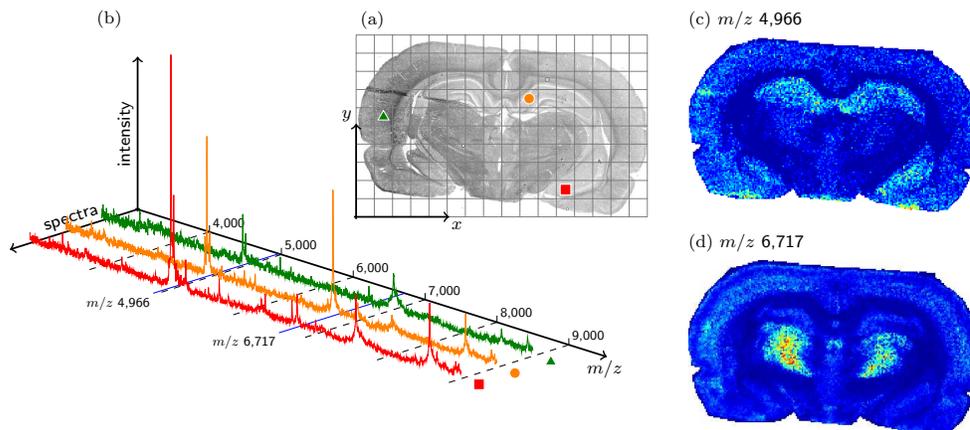


Figure 1. Imaging mass spectrometry data acquired from a rat brain tissue section, adapted from [3]. Each spot on the x, y -grid on the sample in (a) corresponds to one spectrum (b). An m/z -image corresponding to a m/z -value represents the spatial distribution of the ions with this m/z -value, (c) and (d).

1.2. Imaging mass spectrometry

Imaging mass spectrometry (IMS) is a mass spectrometry-based technique for spatially resolved chemical analysis. In this paper, we consider MALDI-IMS which uses the MALDI-TOF (time-of-flight) mass spectrometer. Given a tissue section, a MALDI imaging mass spectrometer acquires mass spectra at discrete spatial points across the sample surface, providing a so-called datacube or hyperspectral image with a mass spectrum acquired at a single pixel [1, 2], see figure 1. A mass spectrum represents relative abundances of ionizable molecules with various mass-to-charge ratios (m/z), ranging from several hundred up to tens of thousands m/z . A channel of a MALDI datacube corresponding to a particular m/z -value is called an m/z -image or a molecular image and expresses the relative spatial abundances of molecular ions with this m/z -value.

MALDI-IMS data is large, with a typical dataset containing 10,000-100,000 spectra across 10,000-100,000 m/z -values. In this paper, we propose a compressed sensing approach to MALDI-IMS which would allow for faster data acquisition by collecting only a part of a hyperspectral image and reconstructing the full image from this data. Instead of acquiring spectra independently for each pixel, we propose to perform a sequence of measurements which results in so called measurement-mean spectra. We then show how to reconstruct the full dataset from these spectra.

1.3. Compressed sensing and its applications to hyperspectral imaging

The combination of classical Shannon-Nyquist sampling and compression steps is one of the main ideas of compressed sensing (CS). It turns out that it is possible to represent or reconstruct data using sampling rates much lower than the Nyquist rate [4, 5, 6]. More formally, given a signal or data $x \in \mathbb{R}^n$, we do not need to acquire n periodic samples to return to the discretized signal x . Instead, it suffices to take only $k = 1, \dots, m \ll n$ linear measurements $y_k \in \mathbb{R}$ using linear test functions $\varphi_k \in \mathbb{R}^n$ (i.e. $y_k = \langle \varphi_k, x \rangle + z_k$), with some additive noise $z_k \in \mathbb{R}$ and noise level $\|z\| \leq \varepsilon$. In matrix

notation this reads

$$y = \Phi x + z, \quad (1.1)$$

where $\Phi \in \mathbb{R}^{m \times n}$ is called the measurement matrix and has rows filled with the functions φ_k . Using the a-priori information that the signal x is sparse or compressible in a basis $\Psi \in \mathbb{R}^{n \times n}$, we can then recover the signal x under suitable assumptions on Φ and Ψ from the measurements y_k with the basis pursuit approach, that is, by solving the following convex optimization problem

$$\operatorname{argmin}_{\lambda \in \mathbb{R}^n} \|\lambda\|_1 \quad \text{subject to} \quad \|y - \Phi\Psi\lambda\|_2 \leq \varepsilon. \quad (1.2)$$

One of the many applications of CS is in hyperspectral imaging. A hardware realization of CS in that hyperspectral situation applying the single-pixel camera [7] has been studied in, for example, [8]. From the theoretical point of view mathematical models have been studied for CS in hyperspectral image reconstruction under certain priors [9, 10, 11]. Suppose that we have a hyperspectral datacube $X \in \mathbb{R}^{n_x \times n_y \times c}$ whereas $n_x \times n_y$ denotes the spatial resolution of each image and c the number of channels. By concatenating each image as a vector we have $X \in \mathbb{R}^{n \times c}$ with $n := n_x \cdot n_y$. In the context of CS, we aim to take $m \ll n$ measurements for each spectral channel $1 \leq j \leq c$ [10, 11] and formulate a reconstruction strategy based on hyperspectral data priors. For example in [11] the authors assume the hyperspectral datacube to have low rank and piecewise constant channel images. Therefore the following convex optimization problem is presented

$$\operatorname{argmin}_{\tilde{X} \in \mathbb{R}^{n \times c}} \|\tilde{X}\|_* + \tau \sum_{j=1}^c \|\tilde{X}_j\|_{TV} \quad \text{subject to} \quad \|Y - \Phi\tilde{X}\|_F \leq \varepsilon, \quad (1.3)$$

where $\|\cdot\|_*$ and $\|\cdot\|_{TV}$ denote the nuclear norm (the sum of the singular values) and the TV semi-norm respectively. Furthermore the notation

$$\tilde{X}_j := \Omega_j \tilde{X} := (\Omega \circ C_j) \tilde{X}, \quad j = 1, \dots, c, \quad (1.4)$$

is used, where C_j maps from a hyperspectral data matrix to its j -th image in vectorized form and Ω concatenates it to an $n_x \times n_y$ image. τ is some positive balancing parameter, and the linear operator Φ is the measurement matrix as previously described. The reason for using the nuclear norm as one of the regularization terms arises from the fact that hyperspectral data often has high correlations in both the spatial and the spectral domains.

Another application of CS in hyperspectral imaging is in calculating a compressed matrix factorization or a (blind) source separation of the data $X \in \mathbb{R}^{n \times c}$, for example

$$X = SH^T, \quad (1.5)$$

where $S \in \mathbb{R}^{n \times p}$ is a so called source matrix, $H \in \mathbb{R}^{c \times p}$ is a mixing matrix and $p \ll \min\{n, c\}$ denotes the number of sources in the data (known a priori). This model has been recently studied in the case of known mixing parameters H of the data X in [12] and with both matrices unknown in [9]. If H is known and if the columns of S are sparse or compressible in a basis Ψ , the problem in [12] becomes

$$\operatorname{argmin}_{\lambda \in \mathbb{R}^n} \|\lambda\|_1 \quad \text{subject to} \quad \|Y - \Phi\bar{H}\Psi\lambda\|_2 \leq \varepsilon, \quad (1.6)$$

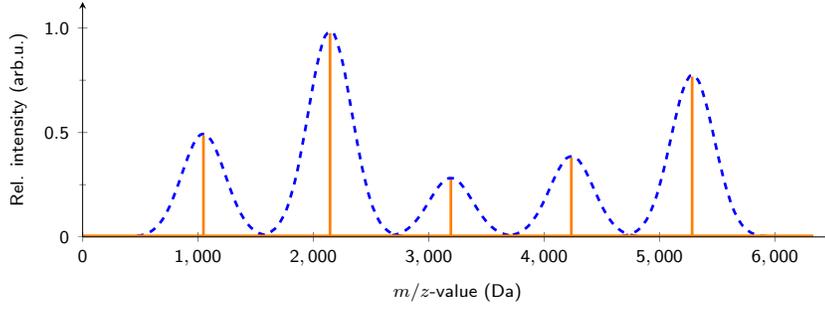


Figure 2. An illustration of a peak-picking approach in mass spectrometry, first published in the proceedings of SampTA 2013 published by EURASIP [17]. Instead of finding a reconstruction \tilde{X} via $\tilde{X}^T = \Psi\tilde{\Lambda}$, we aim to directly recover the features $\tilde{\Lambda}$. Dashed line (- -): Reconstruction of the i -th spectrum, i.e. $\tilde{X}_{(i,\cdot)}^T = (\Psi\tilde{\Lambda})_{(i,\cdot)}$. Solid line (—): Only the main features of the i -th spectrum $\tilde{\Lambda}_{(\cdot,i)}$, i.e. the main peaks, are extracted.

where $\bar{n} = p \cdot n$, $\bar{H} = H \otimes I_n$, with denoting \otimes the usual Kronecker product and I_n the $n \times n$ identity matrix. The authors in [12] also studied the case where the ℓ_1 -norm in (1.6) is replaced by the TV norm with respect to the columns of S , $\sum_{j=1}^p \|S_j\|_{TV}$, where S_j is defined as in (1.4) with proper dimensions. In this instance, solving (1.6) yields a decomposition as in (1.5), where the columns of S contain the p most representative images of the hyperspectral datacube and the rows of H contain the corresponding pseudo spectra.

In this paper we investigate a reconstruction model for hyperspectral data similar to (1.3) and (1.6), but with special motivation for IMS data. Let $X \in \mathbb{R}_+^{n \times c}$ be the hyperspectral IMS data and assume that there exists a sparse decomposition of the spectra $X_{(i,\cdot)} \in \mathbb{R}_+^c$ for $i = 1, \dots, n$ with respect to some basis $\Psi \in \mathbb{R}_+^{c \times c}$, i.e.

$$X^T = \Psi\Lambda \quad (1.7)$$

where $\Lambda \in \mathbb{R}_+^{c \times n}$. By applying compressed measurements via $\Phi \in \mathbb{R}^{m \times n}$ and (1.4), our minimization problem then becomes

$$\operatorname{argmin}_{\Lambda \in \mathbb{R}^{c \times n}} \|\Lambda\|_1 + \sum_{j=1}^c \|\Lambda_j\|_{TV} \quad \text{subject to} \quad \|Y - \Phi\Lambda^T\Psi^T\|_F \leq \varepsilon, \Lambda \geq 0. \quad (1.8)$$

Since we know a-priori that mass spectra in IMS are typically nearly sparse or compressible, we use the ℓ_1 -norm as one regularization term [13, 14]. The TV-term is used because the m/z -images have sparse image gradients [15]. A detailed derivation of the functional (1.8) is presented in sections 3.3 and 3.4.

1.4. Contributions and paper organization

This paper introduces a novel compressed sensing model able to reconstruct a full IMS dataset from only partial measurements. Moreover, with respect to (1.7) and the related work of Louis [16], we aim to directly reconstruct the main features Λ from the measured data Y without inverting the operator Ψ with a sparsity constraint as done in [13]. More precisely, we aim to directly reconstruct the features $\tilde{\Lambda}$ from $\tilde{X}^T = \Psi\tilde{\Lambda}$

from only $m \ll n_x \cdot n_y$ measurements, see figure 2. Usually in compressed sensing for hyperspectral imaging, the restriction on the number of measurements m is weaker since the number of channels is also taken into account, i.e. $\hat{m} = m \cdot c \ll n_x \cdot n_y \cdot c$, see e.g. [9, 12]. In our case the number of measurements does not scale with the number of channels since in MALDI-IMS we are restricted to measure a spectrum at each pixel.

While reconstructing the data, we extract its features in both the spectral and the spatial domains by peak-picking using the ℓ_1 -norm as well as image denoising with the TV semi-norm, both of which are common IMS postprocessing steps [15, 13]. We also prove, under certain assumptions, the robustness of the recovery of both the spectra and the m/z -images.

Since we would like to reconstruct the full dataset we do not need to know the number p of the mixing signatures of the underlying data, which makes this approach different from [9, 12]. Moreover, unlike in [9, 12] we assume our spectra rather than the images of the channels to be sparse or compressible in some known basis.

This paper is organized as follows: In section 2 we give the mathematical notation and background used in this paper. In section 3 we derive our mathematical model for compressed sensing in IMS in which peak-picking in the mass spectra as well as spatial denoising in the m/z -images is applied simultaneously. We also prove the robustness of the reconstruction of both the spectra and the m/z -images. Numerical results on an IMS test dataset are presented and discussed in section 5. Section 6 concludes with a general discussion on our proposed model as well as ideas for future work.

2. Preliminaries

For $p \geq 1$, we denote the matrix p -norm by $\|A\|_p = (\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^p)^{1/p}$ for a matrix $A \in \mathbb{R}^{m \times n}$, which is induced by the ℓ_p -vector norm $\|x\|_{\ell_p} := (\sum_i |x_i|^p)^{1/p}$ for some $x \in \mathbb{R}^n$. For $p = 0$ this is $\|x\|_0 := \|x\|_{\ell_0} := |\text{supp}(x)| := |\{x_j \mid x_j \neq 0\}|$, but it is neither a norm nor a semi-norm. However, we will refer to it as the ℓ_0 -“norm” [18]. The corresponding ℓ_0 -“norm” for matrices can be defined accordingly. In the case of $p = 2$, this is the Frobenius norm, denoted by $\|\cdot\|_F$. This norm is generated by the inner product

$$\langle A, B \rangle = \text{trace}(AB^T) = \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} A_{i,j} B_{i,j}.$$

where $A = (A_{i,j}) \in \mathbb{R}^{m \times n}$ and $B = (B_{i,j}) \in \mathbb{R}^{m \times n}$.

We introduce the notation $X := [X_1, \dots, X_c]$ for a stack of images $X_i \in \mathbb{R}^{n_x \times n_y}$, $i = 1, \dots, c$. A natural extension for the p -norm of these objects is then given as $\|X\|_p := (\sum_{i=1}^c \|X_i\|_p^p)^{1/p}$.

The discrete total variation (TV) semi-norm of A is defined by

$$\|A\|_{TV} = \|\nabla A\|_1,$$

In the paper we will use the anisotropic variant of the total variation norm which is given by

$$\|\nabla A\|_1 = \sum_{i,j} \|(\nabla A)_{i,j}\|_1,$$

where $(\nabla A)_{i,j}$ denotes the discretized gradient. More precisely we have the discrete directional derivatives

$$\begin{aligned} A_x : \mathbb{R}^{m \times n} &\rightarrow \mathbb{R}^{(m-1) \times n}, & (A_x)_{i,j} &= A_{i+1,j} - A_{i,j}, \\ A_y : \mathbb{R}^{m \times n} &\rightarrow \mathbb{R}^{m \times (n-1)}, & (A_y)_{i,j} &= A_{i,j+1} - A_{i,j}. \end{aligned}$$

The discrete gradient transform $\nabla : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n \times 2}$ is then defined component wise as follows

$$(\nabla A)_{i,j} = \begin{cases} ((A_x)_{i,j}, (A_y)_{i,j}), & 1 \leq i \leq m-1, 1 \leq j \leq n-1 \\ (0, (A_y)_{i,j}), & i = m, 1 \leq j \leq n-1 \\ ((A_x)_{i,j}, 0), & 1 \leq i \leq m-1, j = n \\ (0, 0), & i = m, j = n. \end{cases}$$

In the isotropic case one would have

$$\|\nabla A\|_1 = \sum_{i,j} \|(\nabla A)_{i,j}\|_2,$$

which is equivalent to the anisotropic case up to a factor $\sqrt{2}$. The results presented in the paper are therefore valid for both versions up to a factor including $\sqrt{2}$.

We say that $x \in \mathbb{R}^n$ is *s-sparse* when it has at most $s \leq n$ non-zero entries. We write $x \lesssim y$ to say that there exists some constant $C > 0$ such that $x \leq Cy$. Accordingly we define the notation $x \gtrsim y$. We also make use of the notation $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$.

For a Hilbert space \mathcal{H} , we denote $\Gamma_0(\mathcal{H})$ as the set of all proper lower semicontinuous convex functions from \mathcal{H} to $] -\infty, +\infty]$.

For a function $f \in \Gamma_0(\mathcal{H})$ and a point $x \in \mathcal{H}$ the *proximity operator* [19] is defined as the operator $\text{prox}_f : \mathcal{H} \rightarrow \mathcal{H}$ for which $\text{prox}_f(x)$ is the unique point in \mathcal{H} that satisfies

$$\text{prox}_f(x) = \underset{y \in \mathcal{H}}{\text{argmin}} \frac{1}{2} \|x - y\|^2 + f(y).$$

The existence of a minimizer of the function f is guaranteed because f is convex and lower semicontinuous. The uniqueness of $\text{prox}_f(x)$ follows from the additional quadratic data fidelity term which makes the underlying functional strictly convex. In the case of $f = \|\cdot\|_{\ell_1}$ and $\mathcal{H} = \mathbb{R}^n$ the proximity operator is a soft thresholding [20]

$$\text{prox}_{\gamma \|\cdot\|_{\ell_1}}(x) = \left(\max \left\{ 0, \left(1 - \frac{\gamma}{|x_k|} \right) \right\} x_k \right)_{1 \leq k \leq n}, \quad (2.1)$$

where γ is the threshold.

3. Compressed sensing model for imaging mass spectrometry

3.1. Imaging mass spectrometry data

Recall that IMS data is a hyperspectral datacube consisting of one mass spectrum for each pixel. The length of each spectrum depends on the number $c > 0$ of m/z -bins that have been selected before MS data acquisition. By fixing one specific m/z -value,

we have an m/z -image that represents the spatial distribution of the given mass in the biological sample, see figure 1. More formally, for $[1, \dots, n_x] \times [1, \dots, n_y] \subset \mathbb{Z}_+^2$ and $c \in \mathbb{N}_+$, the IMS datacube $X \in \mathbb{R}_+^{n_x \times n_y \times c}$ consists of m/z -images $X_{(\cdot, \cdot; k)} \in \mathbb{R}_+^{n_x \times n_y}$ for $k = 1, \dots, c$ of image resolution $n_x \times n_y$. Since in MALDI measurement process one counts the (relative) number of charged particles of a given mass that reaches the detector, it is natural to assume the data to be non-negative. By concatenating each image as a vector the hyperspectral data becomes

$$X \in \mathbb{R}_+^{n \times c} \quad (3.1)$$

where $n := n_x \cdot n_y$, so that each column in X corresponds to one m/z -image and each row corresponds to one spectrum.

3.2. The compressed sensing process

As described in section 1, part of the IMS measurement process consists of the ionization of the given sample. In MALDI-IMS, for instance, the tissue is ionized by a laser beam, which hits each of the n pixel of a predefined grid, producing n independently measured spectra. Our main goal is to use the theory of compressed sensing [21, 4, 5, 6, 22] to reduce the number of spectra required but still be able to reconstruct a full MALDI-IMS datacube X .

In the context of compressed sensing, each entry y_{ij} of the measurement vectors $y_i \in \mathbb{R}^c$ for $i = 1, \dots, m$ and $j = 1, \dots, c$ is the result of an inner product between the data $X \in \mathbb{R}_+^{n \times c}$ and a test function $\varphi_i \in \mathbb{R}^n$ with components φ_{ik} , i.e.

$$y_{ij} = \langle \varphi_i, X_{(\cdot, j)} \rangle. \quad (3.2)$$

From the IMS perspective these y_i for $i = 1, \dots, m$ are called the *measurement-mean spectra* since they are calculated by the mean intensities on each channel, see figure 3. This can be seen by rewriting (3.2) as

$$y_i^T = \varphi_i^T X = \sum_{k=1}^n \varphi_{ik} X_{(k, \cdot)}, \quad (3.3)$$

which directly shows that the measurement vectors y_i^T are linear combinations of the original spectra $X_{(k, \cdot)}$. We are looking for a reconstruction of the data X based on these m measurement-mean spectra, each measured by one linear function φ_i . In matrix form (3.2) or (3.3) becomes

$$Y = \Phi X \in \mathbb{R}^{m \times c}, \quad (3.4)$$

where $\Phi \in \mathbb{R}^{m \times n}$ is the measurement matrix. Clearly, by (3.3), each row in Y can be interpreted as a measurement-mean spectrum. By incorporating inherent noise $Z \in \mathbb{R}_+^{m \times c}$ that arises during the mass spectrometry measurement process, (3.4) becomes

$$Y = \Phi X + Z \in \mathbb{R}^{m \times c}, \quad (3.5)$$

with $\|Z\|_F \leq \varepsilon$. We explicitly assume this noise to be Gaussian for simplicity, but it should be noted that perhaps a Poisson noise framework might be more suitable to IMS data [15].

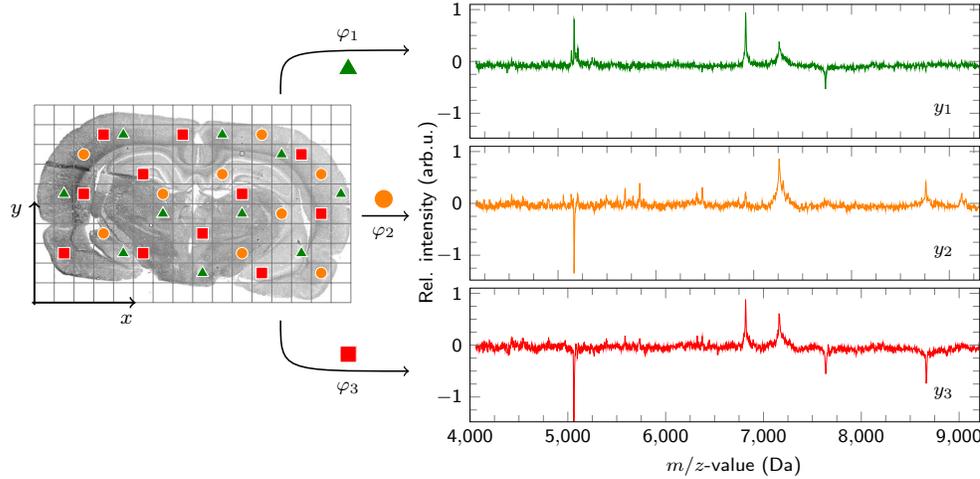


Figure 3. Compressed sensing measurements in imaging mass spectrometry. Each measurement φ_i on the sample (left) (triangle (Δ), circle (\circ), square (\square)) leads to a measurement y_i (right) called a measurement-mean spectrum.

Finding a reconstruction of the data X from the measurements Y in (3.4) is hopeless due to the ill-posed nature of the problem. Therefore, we need additional a-priori knowledge to find at least those reasonable solutions which also fulfill the given data properties. To remedy this, the next two subsections introduce two notions of sparsity that arise in imaging mass spectrometry.

3.3. First assumption: compressible spectra

For each pixel in the sample, we obtain a mass spectrum with positive real entries, i.e.

$$X_{(k,\cdot)} \in \mathbb{R}_+^c, \quad k = 1, \dots, n.$$

As motivated in figures 2 and 4, IMS spectra are compressible in spectral domain. We therefore assume that these spectra are sparsely presented by a suitable choice of functions $\psi_i \in \mathbb{R}_+^c$ for $i = 1, \dots, c$. More concretely, this means that there exists a matrix $\Psi \in \mathbb{R}_+^{c \times c}$ such that for each spectrum $X_{(k,\cdot)}$ we have a coefficient vector $\lambda_k \in \mathbb{R}_+^c$ with $\|\lambda_k\|_0 \ll c$, such that

$$X_{(k,\cdot)}^T = \Psi \lambda_k, \quad k = 1, \dots, n. \quad (3.6)$$

In this paper, we assume the basis functions to be shifted Gaussians [13, 23, 24, 15]

$$\psi_k(x) = \frac{1}{\pi^{1/4} \sigma^{1/2}} \exp\left(-\frac{(x-k)^2}{2\sigma^2}\right), \quad (3.7)$$

where the standard deviation σ needs to be set based on the data [15]. However, in matrix form, the sparsity property (3.6) can be written as

$$X^T = \Psi \Lambda, \quad (3.8)$$

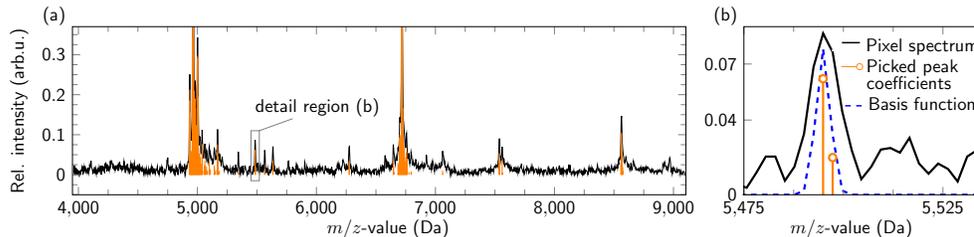


Figure 4. An example of a pixel spectrum from the rat brain sample dataset and its peak-picking result via ℓ_1 minimization with Gaussians as basis elements $\psi_k(\cdot)$ with $\sigma = 0.75$ (see (3.7)), (a). (b) presents a detailed view of the marked region in (a), in which a basis element as well as the coefficients of the detected peak are visualized.

where $\Lambda \in \mathbb{R}_+^{c \times n}$ is the coefficient matrix. The single-spectrum case from (3.6) can simply be found in (3.8): One column in X^T corresponds to one spectrum. The multiplication of Ψ with one column of Λ is exactly the same as in (3.6). However, in light of the compressible spectra, our aim should be to minimize each column $\Lambda_{(\cdot, i)}$ of Λ with respect to the l_0 -“norm”, since each represents the sparse peak-list information based on Ψ . Thus, for one spectrum we have $\|\Lambda_{(\cdot, i)}\|_0$ and for all spectra this reads

$$\|\Lambda\|_0. \quad (3.9)$$

Note that the notation from spectra and images (the order in the brackets in the index) changes for Λ due to the transposition in (3.8).

Putting (3.5) and (3.8) together leads to

$$Y = \Phi \Lambda^T \Psi^T + Z. \quad (3.10)$$

3.4. Second assumption: sparse image gradients

By fixing one m/z -value $i_0 \in \{1, \dots, c\}$ we get a vector $X_{(\cdot, i_0)} \in \mathbb{R}_+^n$ (one column of the dataset X), which by (1.4) is also an m/z -image $X_{i_0} \in \mathbb{R}_+^{n_x \times n_y}$ that represents the spatial distribution of the fixed mass m_0 in the measured biological sample. A priori, we know that these m/z -images are sparse with respect to their gradient. Additionally, we also note the large variance inside each individual m/z -image [15]. To handle both, we want to make use of the total variation (TV) model introduced by Rudin, Osher and Fatemi [25]. So we want each m/z -image to be minimized with respect to its TV semi-norm.

The matrix Ψ is columnwise filled with the shifted Gaussian kernels from (3.7) and it can therefore be interpreted as a convolution operator. With respect to (3.8), this means that the spectra $X_{(k, \cdot)}$, $k = 1, \dots, n$, are only sums of the shifted Gaussian kernels, see figure 2. From the spatial point of view this means that regions in an m/z -image are not separated sharply. In fact, they overlap each other, and this can also be observed in real measured data. However, the multiplication of Ψ in (3.8) only convolves or smoothes the boundaries in the m/z -images and does not effect the structure of each m/z -image X_i . Therefore, instead of minimizing

$$\|X_i\|_{TV} = \|(\Psi \Lambda)_i^T\|_{TV},$$

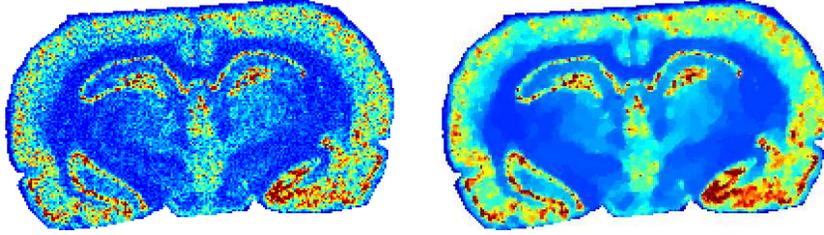


Figure 5. An example of a TV denoised m/z -image. Left: Noisy m/z -image from the rat brain dataset. Right: TV denoised image using the algorithm described in [26].

we conclude that it suffices to minimize the TV norm of the c images given through the coefficients Λ , i.e.

$$\sum_{i=1}^c \|\Lambda_i\|_{TV}. \quad (3.11)$$

3.5. The final model

In total, we are now able to formulate our model for CS in IMS. We seek a positive coefficient matrix $\tilde{\Lambda} \in \mathbb{R}_+^{c \times n}$ such that

1. the reconstructed datacube $\tilde{X}^T = \Psi \tilde{\Lambda}$ is consistent with the observed measurements Y up to a certain noise level ε , see (3.10)
2. the m/z -images X_i for $i = 1, \dots, c$ or, more precisely, the deconvoluted analogs Λ_i , have sparse image gradients, see (3.11)
3. each spectrum $X_{(i,\cdot)}$ can be represented by only a few peaks indicating sparse coefficient vectors $\Lambda_{(\cdot,i)}$

This leads us to the following optimization problem

$$\operatorname{argmin}_{\Lambda \in \mathbb{R}^{c \times n}} \|\Lambda\|_0 + \sum_{i=1}^c \|\Lambda_i\|_{TV} \quad \text{subject to} \quad \|Y - \Phi \Lambda^T \Psi^T\|_F \leq \varepsilon, \Lambda \geq 0. \quad (3.12)$$

It turns out that minimizing with respect to the ℓ_0 -“norm” is NP-hard [27]. Furthermore, this norm is not convex. To obviate this it is common to replace this norm with the ℓ_1 -norm [28, 29]. By introducing further the linear mapping

$$\mathcal{D}_{\Phi, \Psi} : \mathbb{R}_+^{c \times n} \rightarrow \mathbb{R}^{m \times c}, \Lambda \mapsto \Phi \Lambda^T \Psi^T, \quad (3.13)$$

(3.12) becomes

$$\operatorname{argmin}_{\Lambda \in \mathbb{R}^{c \times n}} \|\Lambda\|_1 + \sum_{i=1}^c \|\Lambda_i\|_{TV} \quad \text{subject to} \quad \|Y - \mathcal{D}_{\Phi, \Psi} \Lambda\|_F \leq \varepsilon, \Lambda \geq 0. \quad (3.14)$$

3.6. Robust recovery

We now show that the ℓ_1 reconstruction of the unknown matrix $\Lambda \in \mathbb{R}_+^{c \times n}$ in (3.14) is robust with respect to noise. In our case of compressed MALDI hyperspectral imaging, this means that the pixel spectra as well as the m/z -images are stably reconstructed. For this we need to generalize the results from [4, 30] and we will also assume, similar to [30], to have measurements on the image gradients.

One of the fundamental ideas in CS is the following restricted isometry property (RIP) whose definition is as follows.

Definition 3.1. *The linear operator $\mathcal{A} : \mathbb{R}^{n_x \times n_y} \rightarrow \mathbb{R}^{m \times p}$ has the restricted isometry property of order s and level $\delta \in (0, 1)$ if*

$$(1 - \delta)\|X\|_F^2 \leq \|\mathcal{A}(X)\|_F^2 \leq (1 + \delta)\|X\|_F^2 \quad \text{for all } s\text{-sparse } X \in \mathbb{R}^{n_x \times n_y}.$$

The smallest δ for which this holds is the restricted isometry constant for the operator \mathcal{A} and is denoted by δ_s .

We will make use of the following notation from [30]. For a matrix Φ , we denote Φ_0 and Φ^0 to be the matrices which arise from Φ by concatenating a row of zeros at the bottom or on top, respectively. The following lemma establishes a relation between measurements of directional gradients and these padded matrices.

Lemma 3.2. *Let $X \in \mathbb{R}^{n_x \times n_y}$, $\Phi \in \mathbb{R}^{(n_x-1) \times n_y}$ and $\Psi \in \mathbb{R}^{(n_y-1) \times n_x}$. Then*

$$\langle \Phi, X_x \rangle = \langle \Phi^0, X \rangle - \langle \Phi_0, X \rangle$$

and

$$\langle \Psi, X_y^T \rangle = \langle \Psi^0, X^T \rangle - \langle \Psi_0, X^T \rangle,$$

where X_x and X_y are defined as in the preliminaries.

Proof. Using the definitions of the directional derivatives and the inner product from section 2, simple algebraic manipulations lead to

$$\begin{aligned} \langle \Phi, X_x \rangle &= \sum_{\substack{1 \leq i \leq n_x - 1 \\ 1 \leq j \leq n_y}} \Phi_{(i,j)}(X_x)_{i,j} \\ &= \sum_{\substack{1 \leq i \leq n_x - 1 \\ 1 \leq j \leq n_y}} \Phi_{(i,j)}(X_{i+1,j} - X_{i,j}) \\ &= \sum_{\substack{1 \leq i \leq n_x \\ 1 \leq j \leq n_y}} \left(\Phi_{(i,j)}^0 X_{i+1,j} - \Phi_{0,(i,j)} X_{i,j} \right) \\ &= \langle \Phi^0, X \rangle - \langle \Phi_0, X \rangle. \end{aligned}$$

The other equality follows similarly. \square

We we also make use of the *asymmetric isometry property* (A-RIP), the *restricted condition number* of a dictionary D as well as the *dictionary restricted isometry property* (D-RIP) as proposed in [31, 12, 32].

Definition 3.3. A matrix $D \in \mathbb{R}^{n_x \times n_x}$ satisfies the asymmetric restricted isometry property (A-RIP), if for all s -sparse $X \in \mathbb{R}^{n_x \times n_y}$ the following inequalities hold:

$$\mathcal{L}(D)\|X\|_F \leq \|DX\|_F \leq \mathcal{U}(D)\|X\|_F,$$

where $\mathcal{L}(D)$ and $\mathcal{U}(D)$ are the largest and the smallest constants for which the above inequalities hold. The restricted condition number of D is defined as

$$\xi(D) = \frac{\mathcal{U}}{\mathcal{L}}.$$

The D-RIP extends the notion of the standard RIP to matrices adapted to a dictionary.

Definition 3.4. A linear operator $\mathcal{A} : \mathbb{R}^{n_x \times n_y} \rightarrow \mathbb{R}^{m \times p}$ has the D-RIP of order s and level $\delta^* \in (0, 1)$, adapted to a dictionary D , if for all s -sparse $X \in \mathbb{R}^{n_x \times n_y}$ it holds

$$(1 - \delta^*)\|DX\|_F^2 \leq \|\mathcal{A}(DX)\|_F^2 \leq (1 + \delta^*)\|DX\|_F^2.$$

The robustness result that will be shown in theorem 3.6 rests mainly on the following proposition. They are generalizations of proposition 2 and theorem 4 in [30, 33]. The following proposition states, that if a family of noisy D-RIP-measurements fulfils generalized cone and tube constraints as introduced in [30], then robust recovery is possible.

Proposition 3.5. Fix the parameters $\varepsilon_i > 0$, $\sigma > 0$, $\delta_i^* < 1/3$, $C_i > 0$ as well as $k_i, n_x, n_y, t, p \in \mathbb{N}$. Suppose that for every $i = 1, \dots, m$ $\mathcal{A}_i : \mathbb{R}^{n_x^i \times n_y^i} \rightarrow \mathbb{R}^{t \times p}$ satisfies the D-RIP of order $5k_i\gamma^2$ and level δ_i^* , that each given dictionary $D_i \in \mathbb{R}^{n_x^i \times n_x^i}$ satisfies the A-RIP with constants $\mathcal{L}(D_i)$ and $\mathcal{U}(D_i)$ and suppose that each image $D_i X_i \in \mathbb{R}^{n_x^i \times n_y^i}$ satisfies a tube constraint

$$\|\mathcal{A}_i(D_i X_i)\|_F \leq C_i \varepsilon_i. \quad (3.15)$$

Set $C := \max C_i$, $\delta^* = \max \delta_i^*$, $\varepsilon := \max \varepsilon_i$, $k := \min k_i$, $K := \max k_i$, $\mathcal{U} := \max \mathcal{U}(D_i)$, $\mathcal{L} := \min \mathcal{L}(D_i)$, $\xi = \mathcal{U}/\mathcal{L}$, and suppose that $\gamma \geq \xi \sqrt{\frac{K}{k}}$. Further suppose that for each subset S_i of cardinality $|S_i| \leq k_i$ for $i = 1, \dots, m$ a cone constraint of the form

$$\sum_{i=1}^m \|X_{i_{S_i^C}}\|_1 \leq \sum_{i=1}^m \|X_{i_{S_i}}\|_1 + \sigma \quad (3.16)$$

is satisfied, where $X_{i_{S_i^C}}$ and $X_{i_{S_i}}$ denotes the matrix $X_i \in \mathbb{R}^{n_x^i \times n_y^i}$ restricted to the index set S_i^C and S_i , respectively. Then

$$\sum_{i=1}^m \|X_i\|_F \lesssim m\varepsilon + \frac{\sigma}{\sqrt{K}} \quad (3.17)$$

and

$$\sum_{i=1}^m \|X_i\|_1 \lesssim \sqrt{K}m\varepsilon + \sigma. \quad (3.18)$$

Proof. Let $s_i = k_i \gamma^2$ and let $S_i \subset [N_i]$, with $N_i = n_x^i \cdot n_y^i$, be the support set of an arbitrary s_i -term approximation. For each image X_i , $i = 1, \dots, m$, we will decompose its complement $S_i^C = [N_i] \setminus S_i$ as

$$X_{i_{S_i^C}} = X_{i_{S_i^1}} + X_{i_{S_i^2}} + \dots + X_{i_{S_i^{r_i}}} \quad \text{where} \quad r_i = \left\lfloor \frac{N}{4s_i} \right\rfloor.$$

Note that $X_{i_{S_i^1}}$ consists of the $4s_i$ largest magnitude components, i.e the four largest coefficients in 1-norm of X_i over S_i^C , $X_{i_{S_i^2}}$ then consists of the $4s_i$ largest magnitude components of X_i over $S_i^C \setminus S_i^1$ and so on. By definition, the average magnitude of the nonzero components of $X_{i_{S_i^{j-1}}}$ is larger than the magnitude of each of the nonzero components of $X_{i_{S_i^j}}$, thus

$$\|X_{i_{S_i^j}}\|_F \leq \frac{\|X_{i_{S_i^{j-1}}}\|_1}{2\sqrt{s_i}}, \quad j = 2, 3, \dots, r_i.$$

Together with the cone constraint (3.16), we obtain

$$\begin{aligned} \sum_{i=1}^m \sum_{j=2}^{r_i} \|X_{i_{S_i^j}}\|_F &\leq \sum_{i=1}^m \sum_{j=2}^{r_i} \frac{\|X_{i_{S_i^{j-1}}}\|_1}{2\sqrt{s_i}} \\ &= \sum_{i=1}^m \frac{\|X_{i_{S_i^C}}\|_1}{2\gamma\sqrt{k_i}} \\ &\leq \frac{1}{2\gamma\sqrt{k}} \sum_{i=1}^m \|X_{i_{S_i}}\|_1 + \frac{\sigma}{2\gamma\sqrt{k}} \\ &\leq \frac{\sqrt{K}}{2\gamma\sqrt{k}} \sum_{i=1}^m \|X_{i_{S_i}}\|_F + \frac{\sigma}{2\gamma\sqrt{k}}. \end{aligned}$$

In combination with the tube constraints (3.15), the D-RIP for each \mathcal{A}_i as well as the A-RIP for each D_i , we see

$$\begin{aligned} Cm\varepsilon &\geq \sum_{i=1}^m \|\mathcal{A}_i(D_i X_i)\|_F \\ &\geq \sum_{i=1}^m \|\mathcal{A}_i(D_i(X_{i_{S_i}} + X_{i_{S_i^1}}))\|_F - \sum_{i=1}^m \sum_{j=2}^{r_i} \|\mathcal{A}_i(D_i X_{i_{S_i^j}})\|_F \\ &\geq \sum_{i=1}^m \sqrt{1 - \delta_i^*} \|D_i(X_{i_{S_i}} + X_{i_{S_i^1}})\|_F - \sum_{i=1}^m \sum_{j=2}^{r_i} \sqrt{1 + \delta_i^*} \|D_i X_{i_{S_i^j}}\|_F \\ &\geq \mathcal{L} \sqrt{1 - \delta^*} \sum_{i=1}^m \|X_{i_{S_i}} + X_{i_{S_i^1}}\|_F - \mathcal{U} \sqrt{1 + \delta^*} \left(\frac{\sqrt{K}}{2\gamma\sqrt{k}} \sum_{i=1}^m \|X_{i_{S_i}}\|_F + \frac{\sigma}{2\gamma\sqrt{k}} \right) \\ &\geq \left(\mathcal{L} \sqrt{1 - \delta^*} - \mathcal{U} \frac{\sqrt{K}}{2\gamma\sqrt{k}} \sqrt{1 + \delta^*} \right) \sum_{i=1}^m \|X_{i_{S_i}} + X_{i_{S_i^1}}\|_F - \mathcal{U} \sqrt{1 + \delta^*} \frac{\sigma}{2\gamma\sqrt{k}}. \end{aligned}$$

Further calculations require that the bracket term is strictly positive, or

$$\sqrt{1 - \delta^*} - \xi \frac{\sqrt{K}}{2\gamma\sqrt{k}} \sqrt{1 + \delta^*} > 0.$$

With $\gamma \geq \xi\sqrt{K}/\sqrt{k}$ it is sufficient to have $\delta^* < 1/3$. For this it follows

$$\sum_{i=1}^m \|X_{i_{S_i}} + X_{i_{S_i^c}}\|_F \leq 5\frac{C}{\mathcal{L}}m\varepsilon + 3\xi\frac{\sigma}{\gamma\sqrt{k}}.$$

Because of the inequality

$$\begin{aligned} \sum_{i=1}^m \left\| \sum_{j=2}^{r_i} X_{i_{S_i^j}} \right\|_F &\leq \sum_{i=1}^m \sum_{j=2}^{r_i} \|X_{i_{S_i^j}}\|_F \\ &\leq \frac{\sqrt{K}}{2\gamma\sqrt{k}} \sum_{i=1}^m \|X_{i_{S_i}} + X_{i_{S_i^c}}\|_F + \frac{\sigma}{2\gamma\sqrt{k}} \\ &\leq \frac{1}{2\xi} \sum_{i=1}^m \|X_{i_{S_i}} + X_{i_{S_i^c}}\|_F + \frac{\sigma}{2\gamma\sqrt{k}} \end{aligned}$$

and the fact that $\xi \geq 1$, it follows that

$$\sum_{i=1}^m \|X_i\|_F \leq 8\frac{C}{\mathcal{L}}m\varepsilon + 5\xi\frac{\sigma}{\gamma\sqrt{k}} \leq 8\frac{C}{\mathcal{L}}m\varepsilon + 5\frac{\sigma}{\sqrt{K}}$$

which proves (3.17). Using the cone constraint (3.16), we conclude that

$$\begin{aligned} \sum_{i=1}^m \|X_i\|_1 &= \sum_{i=1}^m \|X_{i_S} + X_{i_{S^c}}\|_1 \\ &\leq \sum_{i=1}^m \|X_{i_{S_i}}\|_1 + \sum_{i=1}^m \|X_{i_{S_i^c}}\|_1 \\ &\leq 2 \sum_{i=1}^m \|X_{i_{S_i}}\|_1 + \sigma \\ &\leq 2\sqrt{K} \left(8\frac{C}{\mathcal{L}}m\varepsilon + 5\frac{\sigma}{\sqrt{K}} \right) + \sigma, \end{aligned}$$

which proves (3.18). \square

Theorem 3.6. Consider $n_x, n_y, c, k_0, k_i, m_1, m_2 \in \mathbb{N}$ and $K := \max\{\sum_{i=1}^c k_i, k_0\}$, $k := \min\{\sum_{i=1}^c k_i, k_0\}$ for $i = 1, \dots, c$. Let $\Lambda \in \mathbb{R}^{c \times n}$, where each row of Λ is the concatenation of an image of size $n_x \times n_y$. Let the dictionary Ψ fulfill the A-RIP of order $5k_0\gamma^2$ with constants \mathcal{L} and \mathcal{U} . Let $\gamma \geq \xi\sqrt{K}/\sqrt{k}$. Furthermore let $\mathcal{A} : \mathbb{R}^{(n_x-1) \times n_y} \rightarrow \mathbb{R}^{m_1}$ and $\mathcal{A}' : \mathbb{R}^{(n_y-1) \times n_x} \rightarrow \mathbb{R}^{m_1}$ be shifted measurements on the images $\Lambda_i, i = 1, \dots, c$, such that the operator $\mathcal{B} := [\mathcal{A} \ \mathcal{A}', \dots, \mathcal{A} \ \mathcal{A}']$, consisting of c concatenations of $[\mathcal{A} \ \mathcal{A}']$, has the RIP of order $5\sum_{i=1}^c k_i\gamma^2$ and level $\delta < 1/3$. Let the operator $\mathcal{D}_{\Phi, \Psi} : \mathbb{R}^{c \times n} \rightarrow \mathbb{R}^{m_2 \times c}$ possess the D-RIP of order $5k_0\gamma^2$ and level $\delta < 1/3$. Consider the linear operator $\mathcal{M}(\Lambda) : \mathbb{R}^{c \times n} \rightarrow \mathbb{R}^{4cm_1} \times \mathbb{R}^{m_2 \times c}$ with components

$$\begin{aligned} \mathcal{M}(\Lambda) = \left(\mathcal{A}^0 \Lambda_1, \mathcal{A}_0 \Lambda_1, \mathcal{A}'^0 \Lambda_1, \mathcal{A}'_0 \Lambda_1, \dots \right. \\ \left. \mathcal{A}^0 \Lambda_c, \mathcal{A}_0 \Lambda_c, \mathcal{A}'^0 \Lambda_c, \mathcal{A}'_0 \Lambda_c, \mathcal{D}_{\Phi, \Psi} \Lambda \right). \end{aligned} \quad (3.19)$$

If noisy measurements $Y = \mathcal{M}(\Lambda) + Z$ are observed with noise level $\|Z\|_F \leq \varepsilon$, then

$$\Lambda^\diamond = \operatorname{argmin}_{W \in \mathbb{R}^{c \times n}} \|W\|_1 + \sum_{i=1}^c \|W_i\|_{TV} \quad \text{s.t.} \quad \|\mathcal{M}(W) - Y\|_F \leq \varepsilon, \quad (3.20)$$

satisfies both

$$\begin{aligned} \|\Lambda - \Lambda^\diamond\|_F + \sum_{i=1}^c \|\nabla \Lambda_i - \nabla \Lambda_i^\diamond\|_F \\ \leq \frac{1}{\sqrt{K}} \left(\|\Lambda - \Lambda_{S_0}\|_1 + \sum_{i=1}^c \|\nabla \Lambda_i - (\nabla \Lambda_i)_{S_i}\|_1 \right) + \varepsilon, \end{aligned} \quad (3.21)$$

and

$$\begin{aligned} \|\Lambda - \Lambda^\diamond\|_1 + \sum_{i=1}^c \|\Lambda_i - \Lambda_i^\diamond\|_{TV} \\ \lesssim \|\Lambda - \Lambda_{S_0}\|_1 + \sum_{i=1}^c \|\nabla \Lambda_i - (\nabla \Lambda_i)_{S_i}\|_1 + \sqrt{K} \varepsilon. \end{aligned} \quad (3.22)$$

Proof. For $X \in \mathbb{R}^{c \times n}$ define $\tilde{X} := [\nabla X_1, \dots, \nabla X_c]^T$. To simplify the notation later on we define

$$\|\tilde{X}\|_{p,\Sigma} := \left(\sum_{i=1}^c \|\nabla X_i\|_p^p \right)^{1/p}.$$

With respect to proposition 3.5, since \mathcal{B} as well as $\mathcal{D}_{\Phi,\Psi}$ satisfy the RIP, it suffices to show that for $D = \Lambda - \Lambda^\diamond$, both D and \tilde{D} satisfy the tube and cone constraints. Write $L_i = [(D_i)_x, (D_i)_y^T]$ and let P denote the map which maps the indices of non-zero entries of ∇D_i to their corresponding indices in L_i . Let $L := [L_1, \dots, L_c]^T$ and extend the and $A_1, A_2, \dots, A_{m_1}, A'_1, A'_2, \dots, A'_{m_1}$ be such that for an image W

$$\mathcal{A}(W)_j = \langle A_j, W \rangle, \quad \text{and} \quad \mathcal{A}'(W)_j = \langle A'_j, W \rangle.$$

We now show that D as well as that \tilde{D} , satisfy the tube and cone constraints.

Cone constraint: Let S_0 be the support of the s_0 largest entries of Λ , and for $i = 1, \dots, c$, let S_i denote the support of the s_i largest entries of $\nabla \Lambda_i$ and S_i^c its complement and set $S := \bigcup S_i$. Using the minimality property of $\Lambda^\diamond = \Lambda - D$, it follows that

$$\begin{aligned} & \|\Lambda_{S_0}\|_1 - \|D_{S_0}\|_1 - \|\Lambda_{S_0^c}\|_1 + \|D_{S_0^c}\|_1 + \\ & \quad \|\tilde{\Lambda}_S\|_{1,\Sigma} - \|\tilde{D}_S\|_{1,\Sigma} - \|\tilde{\Lambda}_{S^c}\|_{1,\Sigma} + \|\tilde{D}_{S^c}\|_{1,\Sigma} \\ & \leq \|\Lambda_{S_0} - D_{S_0}\|_1 + \|\Lambda_{S_0^c} - D_{S_0^c}\|_1 + \|\tilde{\Lambda}_S - \tilde{D}_S\|_{1,\Sigma} + \|\tilde{\Lambda}_{S^c} - \tilde{D}_{S^c}\|_{1,\Sigma} \\ & = \|\Lambda^\diamond\|_1 + \|\tilde{\Lambda}^\diamond\|_{1,\Sigma} \\ & = \|\Lambda^\diamond\|_1 + \sum_{i=1}^c \|\nabla \Lambda_i^\diamond\|_1 \\ & \leq \|\Lambda\|_1 + \sum_{i=1}^c \|\nabla \Lambda_i\|_1 \\ & = \|\Lambda_{S_0}\|_1 + \|\tilde{\Lambda}_S\|_{1,\Sigma} + \|\Lambda_{S_0^c}\|_1 + \|\tilde{\Lambda}_{S^c}\|_{1,\Sigma}. \end{aligned}$$

Rewriting this equality leads to

$$\begin{aligned} \|D_{S_0^c}\|_1 + \|\tilde{D}_{S^c}\|_{1,\Sigma} &\leq \|D_{S_0}\|_1 + 2\|\Lambda_{S_0^c}\|_1 + \|\tilde{D}_S\|_{1,\Sigma} + 2\|\tilde{\Lambda}_{S^c}\|_{1,\Sigma} \\ &= \|D_{S_0}\|_1 + \|\tilde{D}_S\|_{1,\Sigma} + 2\|\Lambda - \Lambda_{S_0}\|_1 + 2\|\tilde{\Lambda} - \tilde{\Lambda}_S\|_{1,\Sigma}, \end{aligned}$$

Using the definition of $\|\cdot\|_{p,\Sigma}$ for $p = 1$ yields

$$\begin{aligned} &\|D_{S_0^c}\|_1 + \sum_{i=1}^c \|(\nabla D_i)_{S_i^c}\|_1 \\ &\leq \|D_{S_0}\|_1 + \sum_{i=1}^c \|(\nabla D_i)_{S_i}\|_1 + 2\|\Lambda - \Lambda_{S_0}\|_1 + 2\sum_{i=1}^c \|\nabla\Lambda_i - (\nabla\Lambda_i)_{S_i}\|_1. \end{aligned}$$

Now set $\sigma := 2\|\Lambda - \Lambda_{S_0}\|_1 + 2\sum_{i=1}^c \|\nabla\Lambda_i - (\nabla\Lambda_i)_{S_i}\|_1$. Using the projection P of the non-zero entries of ∇D_i on each L_i with $|P(S_i)| \leq |S_i|$, we have that D and L satisfy the cone constraint

$$\|D_{S_0^c}\|_1 + \|L_{P(S)^c}\|_1 \leq \|D_{S_0}\|_1 + \|L_{P(S)}\|_1 + \sigma.$$

Tube constraint: First, D immediately satisfies a tube constraint by feasibility since

$$\|\mathcal{M}(D)\|_F \leq \|\mathcal{M}(\Lambda) - Y\|_F + \|\mathcal{M}(\Lambda^\circ) - Y\|_F \leq 2\varepsilon.$$

Using lemma 3.2 for the j -th measurement of the derivative in both the x and y directions of each image D_i , $i = 1, \dots, c$, it follows that

$$\begin{aligned} |\langle A_j, (D_i)_x \rangle|^2 &= |\langle A_j^0, D_i \rangle - \langle A_{j,0}, D_i \rangle|^2 \\ &\leq 2|\langle A_j^0, D_i \rangle|^2 + 2|\langle A_{j,0}, D_i \rangle|^2 \end{aligned}$$

and

$$\begin{aligned} |\langle A'_j, ((D_i)_y)^T \rangle|^2 &= |\langle A'_j{}^0, (D_i)^T \rangle - \langle A'_{j,0}, (D_i)^T \rangle|^2 \\ &\leq 2|\langle A'_j{}^0, (D_i)^T \rangle|^2 + 2|\langle A'_{j,0}, (D_i)^T \rangle|^2. \end{aligned}$$

Thus, L satisfies a tube constraint

$$\|\mathcal{B}(L)\|_F^2 = \sum_{i=1}^c \sum_{j=1}^{m_1} |\langle A_j, (D_i)_x \rangle|^2 + |\langle A'_j, (D_i)_y \rangle|^2 \leq 2\|\mathcal{M}(D)\|_F^2 \leq 8\varepsilon^2.$$

To apply proposition 3.5, it remains to show that D also satisfies a tube constraint under the measurements $\mathcal{D}_{\Phi,\Psi}$. But this easily holds since

$$\|\mathcal{D}_{\Phi,\Psi} D\|_F \leq \|\mathcal{M}(D)\|_F \leq 2\varepsilon. \quad \square$$

Remark 3.7. A nonnegativity condition can be easily incorporated into theorem 3.6, since we only need feasibility of the true solution and the minimizer.

For theorem 3.6 it remains to validate that $\mathcal{D}_{\Phi,\Psi}$ satisfies the D-RIP, \mathcal{B} the RIP and Ψ the A-RIP. First note, that we can equivalently rewrite our problem as follows: Note that using the Kronecker product \otimes and the identity [34, lemma 4.3]

$$(CDE)_{\text{vec}} = (E^T \otimes C)D_{\text{vec}},$$

(3.12) or (3.14) can be formulated as in (1.8). The notation $(Z)_{\text{vec}}$ emphasizes the vectorized form of the matrix Z by stacking the columns of Z into a single column vector. With respect to the equations (3.4) and (3.8) we have

$$\begin{aligned} y &:= Y_{\text{vec}} = (\Phi \otimes I_{c \times c}) X_{\text{vec}} \\ &= (\Phi \otimes I_{c \times c}) (\Lambda^T \Psi^T)_{\text{vec}} \\ &= (\Phi \otimes I_{c \times c}) (I_{n \times n} \Lambda^T \Psi^T)_{\text{vec}} \\ &= \underbrace{(\Phi \otimes I_{c \times c})}_{:=\tilde{\Phi}} \underbrace{(\Psi \otimes I_{n \times n})}_{:=\tilde{\Psi}} \underbrace{(\Lambda^T)_{\text{vec}}}_{:=\tilde{\lambda}}. \end{aligned}$$

Then, the resulting matrix $\tilde{\Phi}$ is a $mc \times nc$ blockdiagonal matrix with entries Φ on the diagonal. If we now can show the D-RIP for $\tilde{\Phi}$ holds, it also follows that $\mathcal{D}_{\Phi, \Psi}$ fulfils the D-RIP, since

$$\|\Psi^T \Lambda^T\|_F^2 = \|\tilde{\Psi} \tilde{\lambda}\|_2^2 \quad \text{and} \quad \|\tilde{\Phi} \tilde{\Psi} \tilde{\lambda}\|_2^2 = \|\mathcal{D}_{\Phi, \Psi} \Lambda\|_F^2.$$

In [35] it has been shown that the RIP (or also the D-RIP) holds with overwhelming probability also for blockdiagonal matrices, if the elements of the matrix Φ are independently drawn at random from subgaussian distributions and

$$mc \gtrsim Kc \log^2(nc) \log^2(K). \quad (3.23)$$

As the authors state in [35], the results there may not be optimal and can be improved most likely. In fact, for our problem, the estimate for m in (3.23) is way to restrictive if we take a priori information about our problem into account and would always yield a measurement of the full data set.

Concerning the A-RIP, note that our dictionary is invertible and therefore its condition number κ provides an upper bound for all ξ . Therefore, Ψ fulfils the A-RIP and ξ from theorem 3.6 can be estimated from above by κ . In our case, the condition number is reasonably small if σ from (3.7) is small, e.g. $\sigma = 0.75$.

In theorem 3.6 we require $\gamma \geq \xi \sqrt{K}/\sqrt{k}$. If γ is too big, it would require m to be big as well according to the D-RIP of order $5K\gamma^2$. The value ξ has already been discussed before, so we need \sqrt{K}/\sqrt{k} to be small as well. Since Λ can be interpreted as the deconvoluted version of the datacube X (see figure 2 and (3.8)), it inherits the same physical structures. It is therefore reasonable to assume that the sparsity prior $\sum_{i=1}^c k_i \approx k_0$ holds for Λ as well and this implies $\sqrt{K} \approx \sqrt{k}$.

Regarding the RIP of \mathcal{B} , one could transform the operator into four blockdiagonal matrices with a structure similar to $\tilde{\Phi}$. Then, a discussion about the RIP can be done as before. These gradient measurements could theoretically be obtained by shifting the measurement mask. However, in the acquisition process of MALDI-TOF, an ionization of the biological sample is performed. Therefore, the sample is damaged at the ionized points and further measurements would make no sense. As the authors of [30], we believe that the additional $4m_1$ measurements $\mathcal{A}^0, \mathcal{A}_0, \mathcal{A}'^0, \mathcal{A}'_0$ in theorem 3.6 are not necessary. Indeed our numerical results in section 5 seem to confirm this.

4. Numerical implementation

In this paper we make use of the parallel proximal splitting algorithm (PPXA) [36] to solve the proposed optimization problem (3.14). To improve the ℓ_1 - as well as the

TV -minimization effects, we introduce additional regularization parameters $\alpha, \beta > 0$. Thus, the optimization problem becomes

$$\operatorname{argmin}_{\Lambda \in \mathbb{R}^{c \times n}} \alpha \|\Lambda\|_1 + \beta \sum_{i=1}^c \|\Lambda_i\|_{TV} \quad \text{subject to} \quad \|Y - \mathcal{D}_{\Phi, \Psi} \Lambda\|_F \leq \varepsilon, \Lambda \geq 0. \quad (4.1)$$

PPXA is an iterative method for minimizing a finite sum of lower semicontinuous convex functions. It is easy to implement and has the possibility to be parallelized. At each iteration of the algorithm one needs to calculate the proximity operator of each function and to average their results for updating the previous iterate.

To translate the optimization problem (4.1) in the PPXA context, we rewrite it to the sum of four lower semicontinuous convex functions

$$\operatorname{argmin}_{\Lambda \in \mathbb{R}^{c \times n}} f_1(\Lambda) + f_2(\Lambda) + f_3(\Lambda) + f_4(\Lambda), \quad (4.2)$$

where $f_1(\Lambda) = \alpha \|\Lambda\|_1$, $f_2(\Lambda) = \beta \sum_{i=1}^c \|\Lambda_i\|_{TV}$, $f_3(\Lambda) = \iota_{\mathcal{B}_2^\varepsilon}(\Lambda)$ and $f_4(\Lambda) = \iota_{\mathcal{B}_+}(\Lambda)$. Here, $\iota_{\mathcal{C}}$ is simply the indicator function which is defined as

$$\iota_{\mathcal{C}}(\Lambda) = \begin{cases} 0 & \text{if } \Lambda \in \mathcal{C} \\ +\infty & \text{otherwise} \end{cases}. \quad (4.3)$$

The indicator function is applied to the convex sets $\mathcal{B}_2^\varepsilon, \mathcal{B}_+ \subset \mathbb{R}^{c \times n}$, corresponding to the matrices that satisfy the fidelity constraint $\|Y - \mathcal{D}_{\Phi, \Psi} \Lambda\|_F \leq \varepsilon$ and to the ones lying in the positive orthant, respectively.

Algorithm 1: The Parallel Proximal Algorithm (PPXA) for solving (4.1)

Input: $Y, \Psi, \Phi, \alpha, \beta, \varepsilon, \gamma > 0$

Initializations: $k = 0; \Lambda_0 = \Gamma_{1,0} = \Gamma_{2,0} = \Gamma_{3,0} = \Gamma_{4,0} \in \mathbb{R}_+^{c \times n}$

repeat

for $j = 1 : 4$ **do**

$P_{j,k} = \operatorname{prox}_{\gamma f_j}(\Gamma_{j,k})$

end

$\Lambda_{k+1} = (P_{1,k} + P_{2,k} + P_{3,k} + P_{4,k})/4$

for $j = 1 : 4$ **do**

$\Gamma_{j,k+1} = \Gamma_{j,k} + 2\Theta_{k+1} - \Theta_k - P_{j,k}$

end

until *convergence*

The PPXA algorithm adapted to our problem is shown in algorithm 1. We will now shortly state the proximity operators of each of the function f_i , $i = 1, \dots, 4$, and refer the reader to [37, 36, 38] for further information.

The proximity operator of $f_1(\Lambda) = \alpha \|\Lambda\|_1$ is given via the well-known soft-thresholding operator as presented in (2.1) in the preliminaries. For the proximity operator of the sum of TV norm, namely $f_2(\Lambda) = \beta \sum_{i=1}^c \|\Lambda_i\|_{TV}$, we use an efficient implementation from [26]. Since the proximal operator of an indicator function $\iota_{\mathcal{C}}(\Lambda)$ is the orthogonal projection onto the convex set \mathcal{C} , the proximal operator of $\iota_{\mathcal{B}_+}(\Lambda)$ is simply given as

$$\operatorname{prox}_{\gamma \iota_{\mathcal{B}_+(\cdot)}}(Z) = \left(\max\{0, Z_{i,j}\} \right)_{\substack{1 \leq i \leq c \\ 1 \leq j \leq n}}.$$

For the projection onto the ℓ_2 ball we use a forward backward scheme as proposed in [39]. Note that all implementations are given within the UNLocBoX [40].

5. Numerical results

In this section we present reconstruction results for an example IMS dataset based on the proposed model. The well-studied dataset $X \in \mathbb{R}_+^{n \times c}$ was acquired from a rat brain coronal section (see figure 1) which consists of $c = 2,000$ data bins ranging from m/z 4,213 to m/z 9,104. The m/z -images have a spatial resolution of 121×202 . Therefore, we have $n = 24,442$ pixels. The spectra were normalized using total ion count (TIC) normalization, which is the normalization with respect to the ℓ_1 -norm [41]. Furthermore, they were baseline-corrected using the TopHat algorithm with a minimal baseline width set to 10%; for more details, see [15, 3].

In the following experiments, the mass spectra are assumed to be sparse or compressible with respect to shifted Gaussians as in (3.7), where we set the standard deviation $\sigma = 0.75$. By this, we still keep the idea of the peak picking as well as a low conditioning number $\xi \approx 8$, since the last is an important factor in the robustness theorem 3.6. The measurement matrix Φ is randomly filled with numbers from an i.i.d. standard normal Gaussian distribution. The initial guess Λ_0 for the desired solution Λ was set as a random matrix whose negative elements were set to zero. By experience, the noise level ε was set to $3.75 \cdot 10^3$ and we have applied 30 outer loop iterations in the PPXA 1. The regularization parameters in (4.1) were set for each amount of measurements by hand as follows: 20%: $\alpha = 0.15$, $\beta = 0.3$, 40%: $\alpha = 1.3$, $\beta = 1.6$, 60%: $\alpha = 2.0$, $\beta = 2.3$, 80%: $\alpha = 3.2$, $\beta = 3.5$ and 100%: $\alpha = 4.8$, $\beta = 5.1$.

Figure 6 presents the mean spectrum, i.e. the sum over all pixel spectra $X_{(i,\cdot)}$ for $i = 1, \dots, n$, of the rat brain data as well as the mean spectrum of the reconstruction, based on 20%, 40% and 60% measurements taken. The triangles in figure 6(a) show the peaks which are detected based the 20% level. 6(b) and 6(c) show which peaks are additionally extracted during reconstruction, visualized by additional squares and circles. We can clearly see the influence of more measurements on the feature extraction of the main peaks in the mean spectrum. As an example, the peak at m/z 7,060 is only slightly extracted in 6(a). More measurements not only lead to a higher intensity of this peak, but also in additional local information, see figure 6(b)-(c). Note that the described effect is only caused by the amount of samples and does only slightly alter with the regularization parameters.

The effect of this increasing peak intensities can be visualized, for example, by looking at the corresponding m/z -image at m/z 7,060, see figure 8. At the 20% level we extract the main spatial features of this image, but we miss details such as in the lower portion of the data. Increasing the number of measurements clearly leads to better reconstruction results. With 40 % of taken measurements we get almost all main features of the m/z -image.

Finally, figure 9 and figure 10 show images for six additional m/z -values and their corresponding reconstructions at the different measurement levels (20%, 40%, 60%, 80% and 100%). These six m/z -values correspond to six detected high intensity peaks in the mean spectrum as visualized in figure 6(a). Moreover, the m/z -images present main structures within the rat brain, as shown in the rat brain schematic in figure 7, adapted from [15]. As we can see, regions of high intensity pixels are mostly detected as such and were reconstructed well when using 40% measurements. In figures 9 and 10, we can also notice a slight loss of details when applying 40% measurements, as

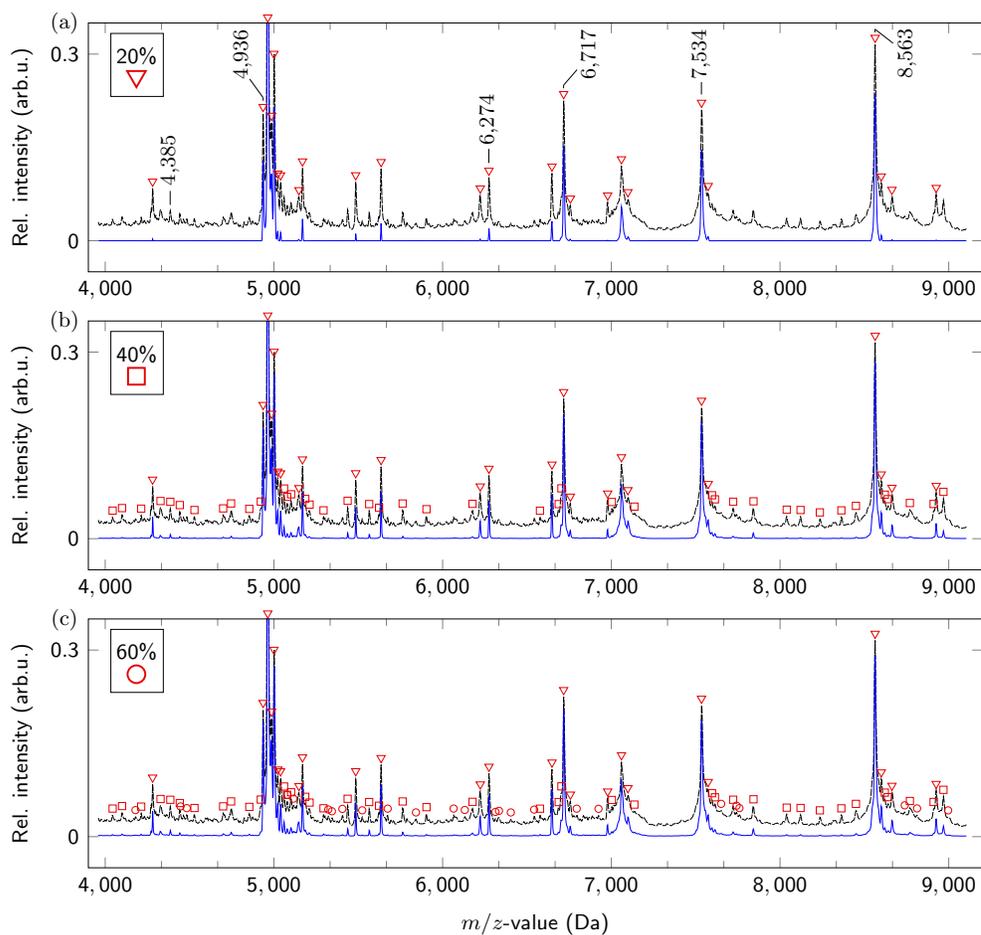


Figure 6. Original mean spectrum (black dashed line) and its reconstruction based on different number of taken measurements (blue line). The spectra are both normalized to $[0,1]$ and we leave out the upper half for better visualization. (a) Reconstruction based on 20% out of $n = 24,442$ taken compressed measurements. The triangles (∇) are set manually and visualize which peaks are mainly detected. (b) and (c) show reconstruction results for 40% and 60%, respectively. The squares (\square) and circles (\circ) show which peaks appear to be additionally detected.

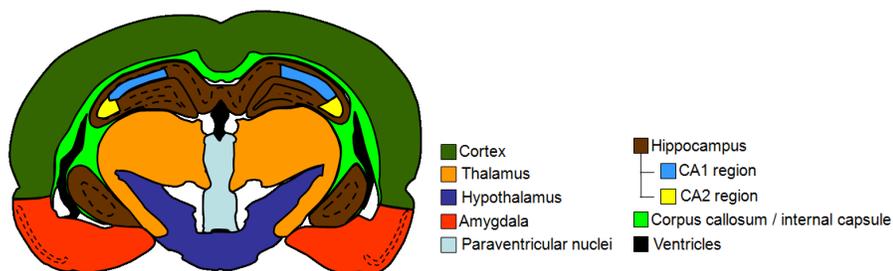


Figure 7. Schematic representation based on the rat brain atlas. Reprinted with permission from [15]. Copyright 2010 American Chemical Society.

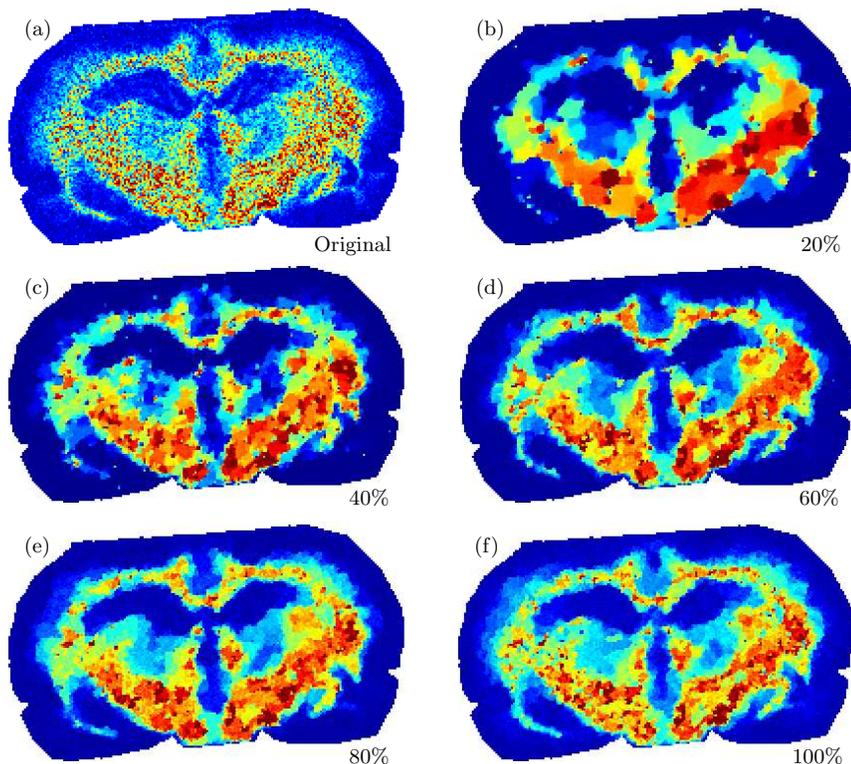


Figure 8. Influence of the number of taken measurements on the reconstruction of the full dataset; an m/z -image corresponding to m/z 7,060 is shown. (a) shows the original image. The images in (b), (c), (d), (e) and (f) show the reconstructions with respect to 20%, 40%, 60%, 80% and 100% taken measurements, respectively.

seen previously in figure 8. This loss clearly reduces with the amount of measurements taken. The image at m/z 4,385 illustrates the reconstruction results on a smaller peak, compared to the other selected. It is recognizable that 20% taken measurements lead to only an idea of where regions appear in the measured image slice, see also 6(a) in comparison with 6(b). In contrast, 40% taken measurements lead to reasonable reconstruction results. This behaviour can be observed on the other reconstructed m/z -images as well.

Note that we have not acquired any additional measurements on the gradients as they are required in theorem 3.6. As it is mentioned in section 3.6, we believe that they are not required in practice. Moreover, the actual theoretical bound in (3.23) on the number of measurements seems to be too pessimistic. Whereas the bound would lead to near full number of measurements, in our example, only few (around 40%) are needed for good reconstruction results.

6. Conclusions

In this paper, we proposed a compressed sensing model for imaging mass spectrometry. In reconstructing the data from less than the half of measurements than normally

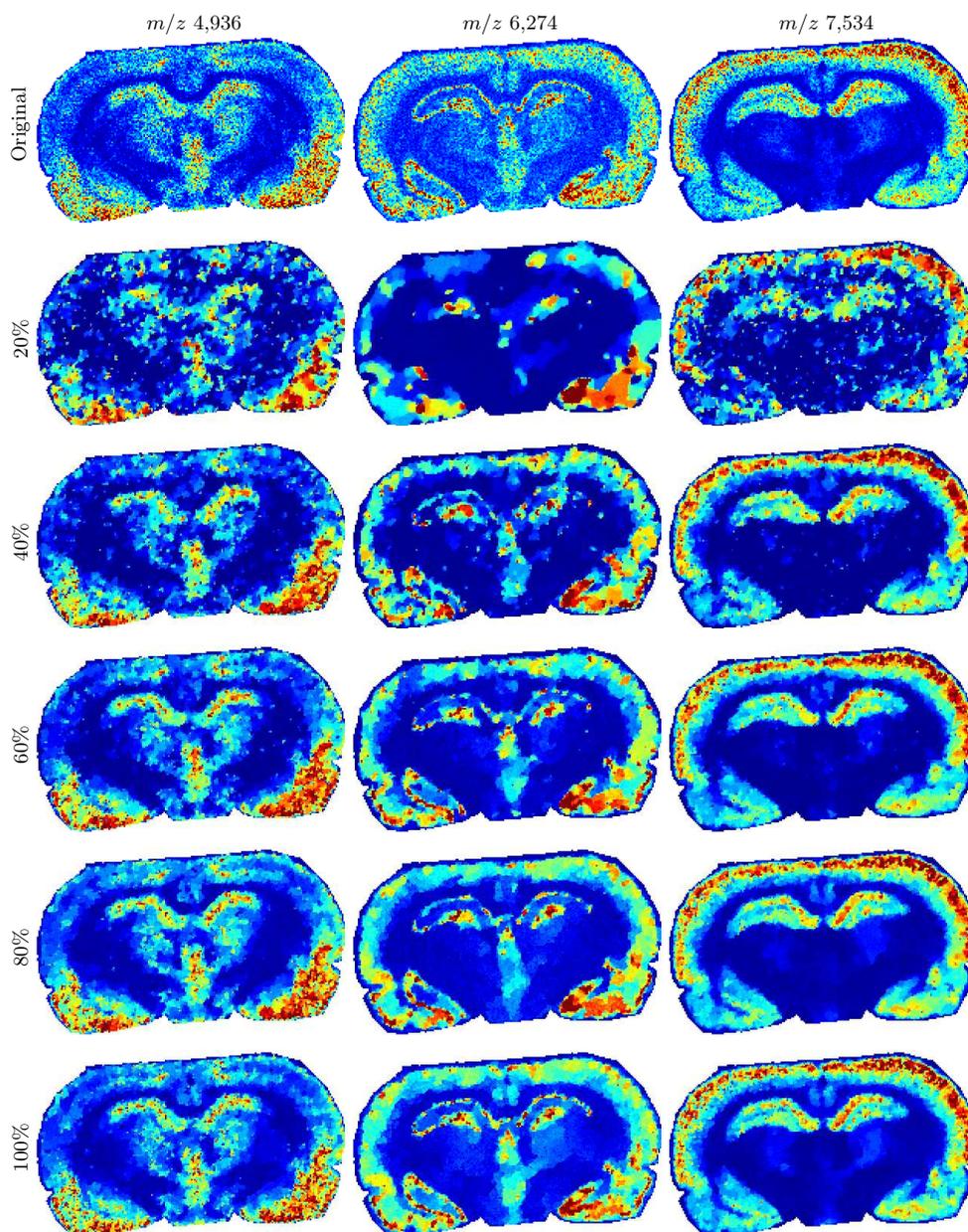


Figure 9. Reconstructions of three different m/z -images based on 20%, 40%, 60%, 80% and 100% of taken measurements. First row, m/z 4,936 with main structures in the middle and the lower part; second row, m/z 6,274 with structures at the boundaries and small regions of high intensity pixels in the middle and bottom part; third row, m/z 7,534 with high intensities at the boundary as well as in the middle and the bottom.

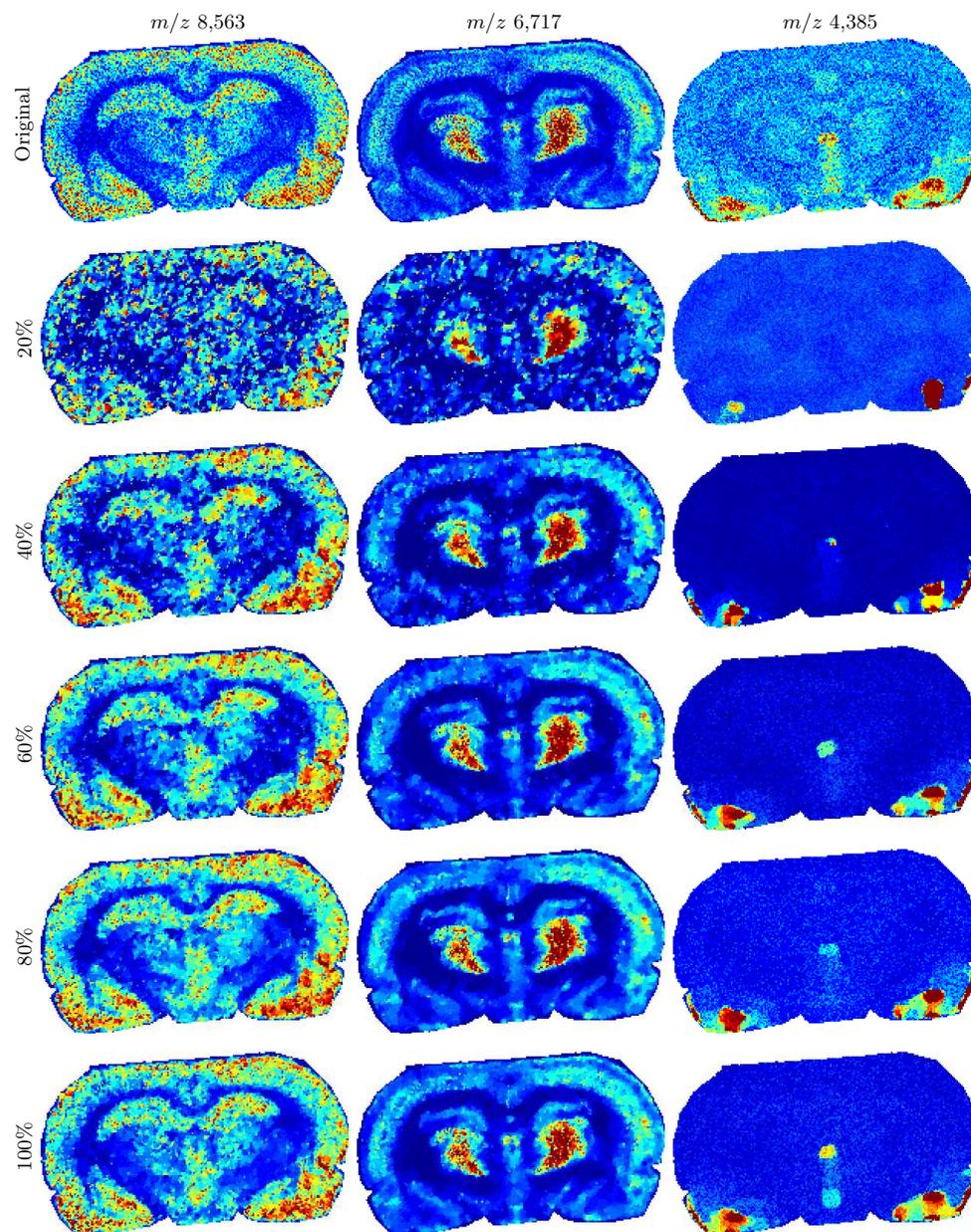


Figure 10. Reconstructions of three different m/z -images based on 20%, 40%, 60%, 80% and 100% of taken measurements. First row, m/z 8,563 with structures at the boundary as well as in the middle and the lower part; second row, m/z 6,717 with one main structure in the center and less intensive regions at the boundary; third row, m/z 4,385 with only small spots of high intensity pixels in the middle and the very bottom.

needed, we applied peak-picking in mass spectra and TV-denoising on the m/z -images at the same time. Both the reconstructed images as well as the spectra were shown to capture the features both in the spatial and the spectral domains. As visually judged, taking 40% to 60% of the typical number of measurements led to only a slight loss of spatial features even of small size.

Currently there are no mass spectrometers which allow for the acquisition of data in such manner. However, considering the recent developments of the single pixel camera [8, 7], one could theoretically implement such a mass spectrometry by splitting the laser into several beams analogously as it is done in the digital micromirror device used in the single pixel camera. Then, instead of analyzing each pixel separately, one could analyze several pixels simultaneously and accumulate a measurement-mean spectrum for such a measurement. Note that modern mass spectrometers indeed use complex optics to achieve non-flat structured laser profile as in Bruker Daltonics smartbeam mass spectrometers [42], although the current optics does not allow to change the profile during an experiment.

We have theoretically proven that both the reconstruction of the spectra and the reconstruction of the m/z -images are robust. Further research might investigate the analysis of how the additional measurements of the gradients in theorem 3.6 could be omitted. Also, the actual bound in (3.23) on the number of measurements to take for robust recovery could be improved. The numerical results presented in this paper suggest that it is too pessimistic.

We have used the parallel proximal algorithm [36] to solve our optimization problem. To improve the regularization effects, we have added regularization parameters α and β and set them by hand for each different amount of measurements. As it can be slightly seen in the results (e.g. 20% in figure 9), it is not feasible to set α by hand for all images. A future direction of investigation should therefore involve regularization terms with locally-dependent parameters $\alpha_i = \alpha(X_i)$ for $i = 1, \dots, c$ for the m/z -images as in [15] for locally-adaptive denoising, and $\beta_j = \beta(X_{(j,\cdot)})$ for $j = 1, \dots, N$ for the spectra.

Future work might also replace the Gaussian noise model with a Poisson statistics approach [43]. As it has been mentioned in [15], this model might be more suitable for MALDI-TOF spectrometry.

Acknowledgments

The authors thank the anonymous referees for their useful and constructive remarks leading to substantial improvements of the paper. We thank Thomas Page for his fruitful discussions on the manuscript and Luke Giberson for carefully proofreading the manuscript. We also thank Michael Becker (Bruker Daltonics) for providing the rat brain dataset, and Jan Hendrik Kobarg for his visualization of mass spectrometry data acquisition in figure 1. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n^o 305259 (3D-MASSOMICS). Andreas Bartels is supported by the BMBF project HYPERMATH under grant 05M13LBA.

References

- [1] R. M. Caprioli, T. B. Farmer, and J. Gile. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal. Chem.*, 23(69):4751–4760, 1997.

- [2] M. Stoeckli, P. Chaurand, D. E. Hallahan, and R. M. Caprioli. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nature Med.*, 7(4):493–496, 2001.
- [3] T. Alexandrov and J. H. Kobarg. Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. *Bioinformatics*, 27(ISMB 2011):i230–i238, 2011.
- [4] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.*, 59(8):1207–1223, 2006.
- [5] E. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.
- [6] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
- [7] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk. A new compressive imaging camera architecture using optical-domain compression. *Comput. Imaging IV*, 6065:43–52, 2006.
- [8] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Proc. Magazine*, 25(2):83–91, 2008.
- [9] M. Golbabaee, S. Arberet, and P. Vandergheynst. Distributed compressed sensing of hyperspectral images via blind source separation. In *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on*, pages 196–198, 2010.
- [10] M. Golbabaee, S. Arberet, and P. Vandergheynst. Multichannel compressed sensing via source separation for hyperspectral images. *Eusipco 2010, Aalborg, Denmark*, 2010.
- [11] M. Golbabaee and P. Vandergheynst. Joint trace/tv norm minimization: A new efficient approach for spectral compressive imaging. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 933–936, 2012.
- [12] M. Golbabaee, S. Arberet, and P. Vandergheynst. Compressive source separation: theory and methods for hyperspectral imaging. *IEEE Trans. Image Proc.*, 22(12):5096–5110, 2013.
- [13] L. Denis, D. A. Lorenz, and D. Trede. Greedy solution of ill-posed problems: error bounds and exact inversion. *Inverse Problems*, 25(11):1–24, 2009.
- [14] T. Alexandrov and A. Bartels. Testing for presence of known and unknown molecules in imaging mass spectrometry. *Bioinformatics*, 29(18):2335–2342, 2013.
- [15] T. Alexandrov, M. Becker, S.-O. Deininger, G. Ernst, L. Wehder, M. Grasmair, F. von Eggling, H. Thile, and P. Maass. Spatial segmentation of imaging mass spectrometry data with edge-preserving image denoising and clustering. *J. Proteome Research*, 9(12):6535–6546, 2010.
- [16] A. K. Louis. Feature reconstruction in inverse problems. *Inverse Problems*, 27(6):065010, 2011.
- [17] A. Bartels, D. Trede, T. Alexandrov, and P. Maass. Hybrid regularization and sparse reconstruction of imaging mass spectrometry data. In *10th international conference on Sampling Theory and Applications (SampTA 2013)*, pages 189–192, Bremen, Germany, 2013. EURASIP.
- [18] D. L. Donoho. Sparse components of images and optimal atomic decompositions. *Constr. Approx.*, 17(3):353–382, 2001.
- [19] J.-J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.
- [20] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011.
- [21] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
- [22] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007.
- [23] V. B. Di Marco and G. G. Bombi. Mathematical functions for the representation of chromatographic peaks. *J. Chromatography A.*, 931(1–2):1–30, 2001.
- [24] K. R. Coombes, J. M. Koomen, K. A. Baggerly, J. S. Morris, and R. Kobayashi. Understanding the characteristics of mass spectrometry data through the use of simulation. *Cancer Inform.*, 1(1):41–52, 2005.
- [25] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60:259–268, 1992.
- [26] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Proc.*, 18(11):2419–2434, 2009.
- [27] B. Natarajan. Sparse approximation solutions to linear systems. *SIAM J. Comput.*, 24(2):227–

- 234, 1995.
- [28] D. L. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM J. Appl. Math.*, 49(3):906–931, 1989.
 - [29] E. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.
 - [30] D. Needell and R. Ward. Stable image reconstruction using total variation minimization. *SIAM J. Imag. Sciences*, 6(2):1035–1058, 2013.
 - [31] Simon Foucart and M.-J. Lai. Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$. *Comput. Harmon. Anal.*, 26(3):395–407, 2009.
 - [32] E. Candès, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Appl. Comp. Harm. Anal.*, 31(1):59–73, 2011.
 - [33] D. Needell and R. Ward. Near-optimal compressed sensing guarantees for total variation minimization. *IEEE Trans. Im. Proc.*, 22(10):3941–3949, 2013.
 - [34] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, first edition, 1991.
 - [35] A. Eftekhari, H. L. Yap, M. B. Wakin, and C. Rozell. The restricted isometry property for block diagonal matrices. Submitted, 2012.
 - [36] P. L. Combettes and J.-C. Pesquet. A proximal decomposition method for solving convex variational inverse problems. *Inverse problems*, 24(6):27, 2008.
 - [37] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4(4):1168–1200, 2005.
 - [38] P. L. Combettes and J.-C. Pesquet. *Proximal splitting methods in signal processing*, volume 49 of *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer, 2011.
 - [39] M.-J. Fadili and J.-L. Starck. Monotone operator splitting for optimization problems in sparse recovery. In *Image Processing (ICIP), 16th IEEE International Conference on*, pages 1461–1464, 2009.
 - [40] P. Nathanaël, A. David, V. Pierre, and P. Gille. *UNLocBox: Short User Guide*. EPFL, Lausanne, July 2012.
 - [41] S.-O. Deinger, D. S. Cornett, R. Paape, M. Becker, C. Pineau, S. Rauser, A. Walch, and E. Wolski. Greedy solution of ill-posed problems: error bounds and exact inversion. *Anal. Bioanal. Chem.*, 401(1):167–181, 2011.
 - [42] A. Holle, A. Haase, M. Kyser, and J. Höndorf. Optimizing UV laser focus profiles for improved MALDI performance. *J. Mass Spec.*, 41(6):705–716, 2006.
 - [43] A. Sawatzky, C. Brune, J. Müller, and M. Burger. Total variation processing of images with Poisson statistics. In *CAIP 2009 - Lecture Notes in Computer Science*, volume 5702, pages 533–540. Springer Berlin Heidelberg, 2009.