Hybrid Regularization and Sparse Reconstruction of Imaging Mass Spectrometry Data

Andreas Bartels*, Dennis Trede*,†, Theodore Alexandrov*,†,‡ and Peter Maass*,†

*Center for Industrial Mathematics, University of Bremen, Bibliothekstr. 1, 28359 Bremen, bartels@math.uni-bremen.de.
†Steinbeis Innovation Center SCiLS Research, Richard-Dehmel-Str. 69, 28211 Bremen, Germany.
‡MALDI Imaging Lab, University of Bremen, Leobener Str., NW2, 28359 Bremen, Germany.

***Abstract*—Imaging mass spectrometry (IMS) is a technique to visualize the molecular distributions from biological samples without the need of chemical labels or antibodies. The underlying data is taken from a mass spectrometer that ionizes the sample on spots on a grid of a certain size. Mathematical postprocessing methods has been investigated twice for better visualization but also for reducing the huge amount of data. We propose a first model that applies compressed sensing to reduce the number of measurements needed in IMS. At the same time we apply peak picking in spectra using the ℓ_1-norm and denoising on the m/z-images via the TV-norm which are both general procedures of mass spectrometry data postprocessing, but always done separately and not simultaneous. This is realized by using a hybrid regularization approach for a sparse reconstruction of both the spectra and the images. We show reconstruction results for a given rat brain dataset in spectral and spatial domain.**

I. Introduction

A. Mass spectrometry

Mass spectrometry is a technique of analytical chemistry for the determination of the elemental 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.

As an example for a mass spectrometer we will now shortly describe the main principles of the so-called matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometer. In MALDI mass spectrometry the sample or compound to be analyzed is dissolved in a so-called matrix with crystallized molecules. Next, the ionization of the sample is triggered by intense laser pulses over a short duration. The ions are then accelerated by an electrostatic field. Since the velocity of the ions depends on the mass-to-charge ratio it is possible to measure the time-of-flight (TOF) to find the mass-to-charge ratio.

Most applications of mass spectrometry can be found in biology and medicine. But generally, mass spectrometry is not limited to the analysis of organic molecules. In principle any ionizable element can be analyzed with this technique.

B. Imaging mass spectrometry

The imaging mass spectrometry (IMS) is a technique used in mass spectrometry to visualize the spatial distribution of e.g. proteins or other chemical compounds. Given a thin sample, usually a tissue section, in MALDI-IMS mass spectra at discrete spatial points across the sample surface are acquired independently, providing a so-called datacube or hyperspectral image, with a mass spectrum measured at each pixel [1], see Figure 1. A mass spectrum represents the relative abundances of ionizable molecules with various mass-to-charge ratios (m/z), ranging for MALDI-IMS from several hundred up to a few tens of thousands m/z. A channel of a MALDI datacube corresponding to an m/z-value is called an m/z- or molecular image and expresses the relative spatial abundances of a molecular ion with this m/z-value. MALDI-IMS data

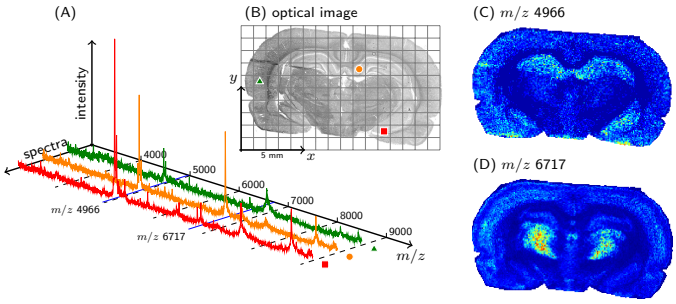

Fig. 1. Mass spectrometry data on an example of a rat brain tissue, taken from [2]. Each spot on the x, y-grid on the sample in (B) corresponds to one spectrum (A). Fixing an m/z-value yields to m/z-images representing the spatial distribution of the corresponding m/z-value, (C) and (D).

is large with a typical dataset comprising $10,000$ - $100,000$ spectra where an individual spectrum represents intensities measured at $10,000$ - $25,000$ m/z-bins.

In order to reduce the number of spectra required for reconstructing the hyperspectral IMS datacube we will make use of the compressed sensing (CS) idea: Instead of measuring spectra independently for each pixel we assume a setup that enables us to acquire some multiple sets of spectra at different points on the data which are then each summed up to a measurement-mean spectrum. Each measurement-mean spectrum then corresponds to one measurement and we would like to reconstruct the full dataset based on these measurements.

C. 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 CS. It turns

out that it is possible to represent or reconstruct given data with sampling rates much lower than the Nyquist rate [3], [4]. Mathematically spoken it means that given a signal, we do not need to acquire n periodic samples to get the discretized signal $x \in \mathbb{R}^n$. Instead, it suffices to take $m \ll n$ 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 $z_k \in \mathbb{R}$. In matrix notation this reads

$$y = \Phi x + z,$$

where Φ is called the measurement matrix whose rows are filled with the linear functionals φ_k. Using the a-priori information that the signal x is S-sparse in a basis Ψ, i.e. $x = \Psi\lambda$ with $\|\lambda\|_{\ell_0} := |\mathrm{supp}(\lambda)| \leqslant S \ll n$, we are then interested in recovering the data x from only few taken measurements y. This can, e.g., be done with the basis pursuit approach, i.e. by solving the following convex optimization problem

$$\operatorname*{argmin}_{\lambda \in \mathbb{R}^n} \|\lambda\|_1 \quad \text{s.t.} \quad \|y - \Phi\Psi\lambda\|_2 \leqslant \varepsilon. \tag{1}$$

One of the many applications of CS can be found in hyperspectral imaging. A hardware realization of CS in that situation applying the single-pixel camera [5] has been studied in, for example, [6]. From the theoretical point of view mathematical models has been studied for CS reconstruction under certain priors [7]–[9]. Suppose that we have 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 one aims to take $m \ll n$ measurements for each spectral channel $1 \leqslant j \leqslant c$ [8], [9] and to formulate a reconstruction strategy based on hyperspectral data priors. For example in [9] the authors assume the hyperspectral datacube to have low rank and piecewise constant channel images. Therefore the following convex optimization problem is presented

$$\operatorname*{argmin}_{\tilde{X} \in \mathbb{R}^{n \times c}} \|\tilde{X}\|_* + \lambda \sum_{j=1}^{c} \|\tilde{X}_j\|_{TV} \quad \text{s.t.} \quad \|Y - \Phi\tilde{X}\|_F \leqslant \varepsilon, \tag{2}$$

where $\|\cdot\|_*$ denotes the nuclear norm (i.e. the sum of the singular values), $\|\cdot\|_{TV}$ denotes the TV norm and the linear operator Φ is a measurement matrix as described above.

Another application of CS in hyperspectral imaging is the idea of calculating a compressed matrix factorization or a (blind) source separation of the data $X \in \mathbb{R}^{n \times c}$, i.e. $X = SH^T$, where $S \in \mathbb{R}^{n \times \rho}$ is a so called source matrix, $H \in \mathbb{R}^{c \times \rho}$ is a mixing matrix and ρ denotes the number of sources in the data which are supposed to be known. This model has been studied in the case of known mixing parameters H of the data X in [10] and with both matrices to be unknown in [7]. In case of the matrix H to be known and under the assumption that the columns of S are sparse or compressible in a basis Ψ, the problem being solved in [10] becomes

$$\operatorname*{argmin}_{\lambda \in \mathbb{R}^{\rho n}} \|\lambda\|_1 \quad \text{s.t.} \quad \|Y - \Phi\bar{H}\Psi\lambda\|_2 \leqslant \varepsilon, \tag{3}$$

where $\bar{H} = H \otimes I_n$ and $\otimes$ denotes the usual matrix Kronecker product and I_n the $n \times n$ identity matrix. The authors in [10] also studied the case where the ℓ_1-norm in (3) is replaced by the TV-norm with respect to the columns of S, i.e. $\sum_{j=1}^{\rho} \|S_j\|_{TV}$. However, as a result of (3) one has a decomposition of the data X as in two matrices S and H where the columns of S contain of the ρ most representative images of the hyperspectral datacube and those of H of the corresponding pseudo spectra.

In this work we investigate a hybrid reconstruction model for hyperspectral data similar to (2) and (3), but with special motivation for imaging mass spectrometry data $X \in \mathbb{R}_+^{n \times c}$ and formulated as a Tikhonov functional:

$$\operatorname*{argmin}_{\tilde{X} \in \mathbb{R}_+^{n \times c}} \frac{1}{2}\|Y - \mathcal{D}_{\Phi,\Psi}(\tilde{X})\|_F + \alpha \sum_{j=1}^{c} \|\tilde{X}_j\|_{TV} + \beta\|\tilde{X}\|_1. \tag{4}$$

Furthermore, we are interested in reconstructing the full dataset $X \in \mathbb{R}_+^{n \times c}$ while extracting its main features.

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. The second, i.e. the TV-term, comes from the fact that the m/z-images are supposed to have sparse image gradients.

II. Compressed sensing model for imaging mass spectrometry

A. Imaging mass spectrometry data

IMS data is a hyperspectral datacube $X \in \mathbb{R}_+^{n_x \times n_y \times c}$ with c channels and m/z-images $X_{(\cdot,\cdot,k)} \in \mathbb{R}_+^{n_x \times n_y}$ for $k = 1, \ldots, c$. By concatenating each image as a vector, the hyperspectral data becomes a matrix $X \in \mathbb{R}_+^{n \times c}$ where $n := n_x \cdot n_y$.

B. The compressed sensing process

A part of the measurement process in IMS consists of the ionization of the given sample. In MALDI-MS, for instance, the tissue is ionized by a laser beam, which shoots on each of the n pixel of a predefined grid. This yields n independently measured spectra. In order to reduce the number of spectra needed for the reconstruction we make use of CS [4], [11].

In the context of compressed sensing, each entry y_{ij} of the measurement vectors $y_i \in \mathbb{R}^c$ for $i = 1, \ldots, m$ and $j = 1, \ldots, c$ 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.

$$y_{ij} = \langle \varphi_i, X_{(\cdot,j)} \rangle. \tag{5}$$

The results y_i for $i = 1, \ldots, m$ are in our IMS context so called *measurement-mean spectra* since they are calculated by the mean intensities on each channel. This can be seen by rewriting (5) as

$$y_i^T = \varphi_i^T X = \sum_{k=1}^{n} \varphi_{ik} X_{(k,\cdot)}, \tag{6}$$

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 m measurement-mean spectra, each measured by one linear function φ_i. In matrix form (5) becomes

$$Y = \Phi X \in \mathbb{R}_+^{m\times c}, \tag{7}$$

where $\Phi \in \mathbb{R}^{m\times n}$ is the measurement matrix. By incorporating noise $Z \in \mathbb{R}_+^{m\times c}$ that arises during the mass spectrometry measurement process, (7) becomes

$$Y = \Phi X + Z \in \mathbb{R}_+^{m\times c}, \tag{8}$$

at which $\|Z\|_F \leqslant \varepsilon$. By this we explicitly assume to have inherent Gaussian noise in the data and we will keep this for the rest of our analysis.

C. First assumption: compressible spectra

Within IMS data acquisition process for each pixel on the sample we gain a mass spectrum whose entries can be seen as positive real numbers, i.e. $X_{(k,\cdot)} \in \mathbb{R}_+^c, \quad k = 1,\dots,n$. As our first assumption we take into account that we know that these spectra are sparse or compressible in spectral domain. Therefore we assume that these spectra are well presented by a suitable choice of functions $\psi_i \in \mathbb{R}_+^c$ for $i = 1,\dots,c$. More precisely this means that there exists a matrix $\Psi \in \mathbb{R}_+^{c\times c}$ such that for each spectrum $X_{(k,\cdot)}$ there is a coefficient vector $\lambda_k \in \mathbb{R}_+^c$ with $\|\lambda_k\|_0 \ll c$, such that $X_{(k,\cdot)}^T = \Psi\lambda_k$. In matrix form this sparsity property can be written as

$$X^T = \Psi\Lambda, \tag{9}$$

where $\Lambda \in \mathbb{R}_+^{c\times n}$ is the coefficient matrix or *feature matrix*, see Figure 2. In light of the sparse spectra, our aim should be to minimize the columns $\Lambda_{(\cdot,i)}$ of Λ with respect to the l_0 "norm", i.e.

$$\sum_{i=1}^{n} \|\Lambda_{(\cdot,i)}\|_0. \tag{10}$$

Putting (8) and (9) together leads to

$$Y = \Phi\Lambda^T\Psi^T + Z. \tag{11}$$

D. Second assumption: sparse image gradients

By keeping one m/z-value $i_0 \in \{1,\dots,c\}$ fixed we get an m/z-image $X_{(\cdot,i_0)} \in \mathbb{R}_+^n$ (one column of the data X) that represents the spatial distribution of the fixed mass m_0 in the measured biological sample. Another a priori knowledge takes into account the sparsity of these m/z-images with respect to their gradient. Besides this, we are also aware of the large variance of noise variance in each m/z-image. To handle both, we want to make use of the total variation (TV) model introduced by Rudin, Osher and Fatemi [12]. So as a second statement, we want each m/z-image to be minimized with respect to its TV semi-norm. By taking into account the coefficient matrix Λ in (9), it arises to minimize its rows $\Lambda_{(i,\cdot)}$ for $i = 1,\dots,c$ since each of them corresponds to an m/z-image, i.e.

$$\sum_{i=1}^{c} \|\Lambda_{(i,\cdot)}\|_{TV}. \tag{12}$$

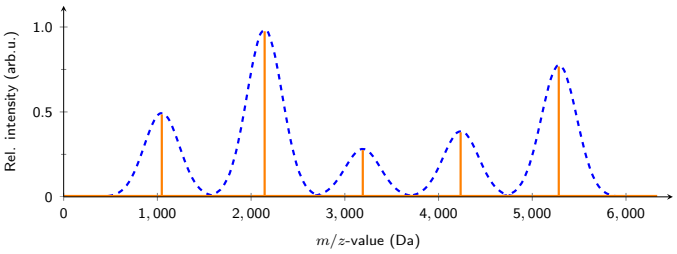

Fig. 2. Illustration of peak picking approach in mass spectrometry. Instead of finding a minimizer Λ and multiply it with a convolution operator Ψ, we aim to recover the features $\tilde{\Lambda}$. Dashed line (- - -): Reconstruction of the i-th spectrum, i.e. $\tilde{X}_{(\cdot,i)}^T = (\Psi\tilde{\Lambda})_{(\cdot,i)}$. Solid line (—): Only the main features of the i-th spectrum $\tilde{\Lambda}_{(\cdot,i)}$, i.e. the main peaks, are extracted.

As in the first assumption and also explained in Figure 2, we aim to extract the main features such as the main peaks in the data. For incorporating also the spatial domain information in each channel, we again only use the coefficient matrix in (12).

E. The final model

Putting it all together, we are now able to formulate our model for CS in IMS. Since minimizing with respect to the ℓ_0 "norm" is NP-hard, it is common to obviate this by replacing it with the ℓ_1-norm. Our minimization problem then finally becomes

$$\underset{\Lambda\in\mathbb{R}_+^{c\times n}}{\operatorname{argmin}} \frac{1}{2}\|Y - \Phi\Lambda^T\Psi^T\|_F^2 + \alpha\sum_{i=1}^{c}\|\Lambda_{(i,\cdot)}\|_{TV} + \beta\sum_{i=1}^{n}\|\Lambda_{(\cdot,i)}\|_1. \tag{13}$$

III. Numerical results

The algorithm we are using to solve (13) is based on the parallelizable primal-dual splitting algorithm presented in [13].

The test data $X \in \mathbb{R}_+^{n\times c}$ is made of a well-studied rat brain coronal section [2] (see Figure 1) which consists of $c = 2,000$ channels with m/z-images of spatial resolution 121×202 and therefore $n = 24,442$ pixel. The spectra were normalized using total ion count (TIC) normalization which is nothing else than a division by the ℓ_1-norm of each spectrum. Furthermore, they were baseline-corrected using the TopHat algorithm with a minimal baseline with set to 10%. We assume the mass spectra to be sparse with respect to shifted Gaussians [14]

$$\psi_k(x) = \frac{1}{\pi^{1/4}\sigma^{1/2}}\exp\left(-\frac{(x-k)^2}{2\sigma^2}\right). \tag{14}$$

In (14), the variance should be set data dependent [15]. The measurement matrix Φ is randomly filled with numbers from an i.i.d. Gaussian distribution with zero mean and variance one. For our results we have further set the regularization parameters in (13) by hand as $\alpha = 1.6$ and $\beta = 1.4$ and applied 100 iterations.

First we present the mean spectrum, i.e. the sum over all pixel spectra $\Lambda_{(i,\cdot)}$ for $i = 1,\dots,n$ of the rat brain data as well as the mean spectrum of the reconstructed datacube, see Figure 4. The reconstruction is based on 40% taken measurements. We can see is that the main peaks from the

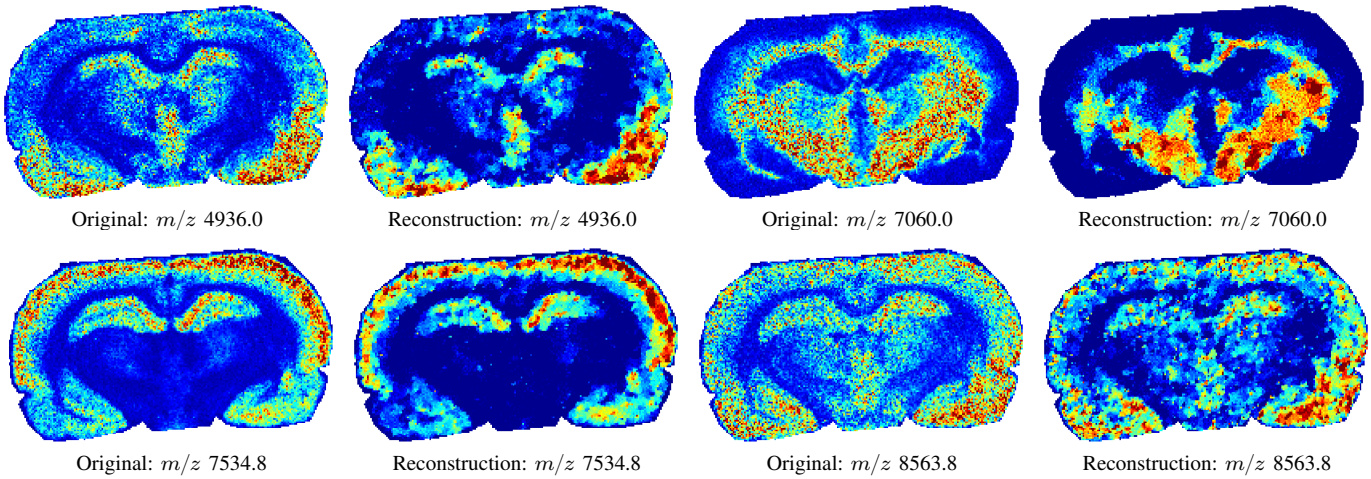

Fig. 3. Reconstruction results of four m/z-images based on 40% taken measurements of spectra.

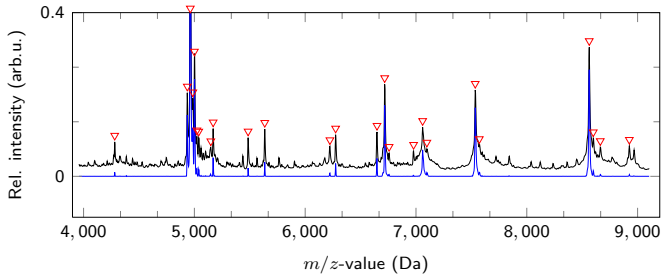

Fig. 4. Mean spectrum from the rat brain dataset (black) and its compressed reconstruction (blue), based on 40% taken measurements. As part of the reconstruction, also main peaks within the spectrum were detected (triangles).

original mean spectrum are recovered while the rest of the m/z-values are set to zero. At second we show four different m/z-images which all belong to a peak in the mean spectrum in Figure 4. We clearly see the influence of both the TV and the ℓ_1 penalty term. Where there are regions of high intensity pixels, total varations effects smoothing those ones while preserving the edges. In addition we see that due to ℓ_1 minimization other non-high intensity pixels were set to (nearly) zero.

IV. Conclusion

We have proposed a first CS model for imaging mass spectrometry. While reconstructing the data from fewer measurements we apply peak picking in mass spectra as well as TV-denoising on the m/z-images at the same time. Our results look promising and motivates for further research in this direction. Future work might be done by replacing the Gaussian noise model with a Poisson statistics approach which is supposed to be more suitable for MALDI-TOF spectrometry [15].

Acknowledgment

The authors would like to thank Michael Becker (Bruker Daltonik) for providing the rat brain dataset.

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*, vol. 23, no. 69, pp. 4751–4760, 1997.

[2] T. Alexandrov and J. H. Kobarg, "Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering," *Bioinformatics*, vol. 27, no. ISMB 2011, pp. i230–i238, 2011.

[3] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[4] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] D. Takhar, J. Laska, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. Kelly, and R. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," *Computational Imaging IV*, vol. 6065, pp. 43–52, 2006.

[6] T. Sun and K. F. Kelly, "Compressive sensing hyperspectral imager," *Comp. Optical Sensing and Imaging*, 2009.

[7] M. Golbabaee, S. Arberet, and P. Vandergheynst, "Distributed compressed sensing of hyperspectral images via blind source separation," *Presentation given at Asilomar conf. on signals, systems, and computers, Pacific Groove, CA, USA, November 7-10*, 2010.

[8] ——, "Multichannel compressed sensing via source separation for hyperspectral images," *Eusipco 2010, Aalborg, Denmark*, 2010.

[9] M. Golbabaee and P. Vandergheynst, "Joint trace/TV norm minimization: A new efficient approach for spectral compressive imaging," *IEEE Int. Conf. on Img. Proc (ICIP)*.

[10] M. Golbabaee, S. Arberet, and P. Vandergheynst, "Compressive source separation: Theory and methods for hyperspectral imaging," 2012, submitted.

[11] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[12] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, pp. 259–268, 1992.

[13] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued and Variational Analysis*, vol. 20, no. 2, pp. 307–330, 2012.

[14] L. Denis, D. A. Lorenz, and D. Trede, "Greedy solution of ill-posed problems: error bounds and exact inversion," *Inverse Problems*, vol. 25, no. 11, pp. 1–24, 2009.

[15] T. Alexandrov, M. Becker, S.-O. Deininger, G. Ernst, L. Wehder, M. Grasmair, F. v. Eggling, H. Thile, and P. Maass, "Spatial segmentation of imaging mass spectrometry data with edge-preserving image denoising and clustering," *J. Proteome Research*, vol. 9, no. 12, pp. 6535–6546, 2010.