Semi-supervised Learning of Sparse Linear Models in Mass Spectral Imaging

Fabian Ojeda[1⋆], Marco Signoretto[1], Raf Van de Plas[1,3], Etienne Waelkens[2,3], Bart De Moor[1,3], and Johan A.K. Suykens[1]

[1] ESAT-SCD-SISTA, Department of Electrical Engineering, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium.
[2] Laboratory for Phosphoproteomics, Katholieke Universiteit Leuven, O & N, Herestraat 49, B-3000 Leuven, Belgium.
[3] ProMeta, Interfaculty Centre for Proteomics and Metabolomics, Katholieke Universiteit Leuven, O & N 2, Herestraat 49, B-3000 Leuven, Belgium.
{fabian.ojeda,marco.signoretto,raf.vandeplas,bart.demoor,johan.suykens}@esat.kuleuven.be
{etienne.waelkens}@med.kuleuven.be
http://www.esat.kuleuven.be/sista

Abstract. We present an approach to learn predictive models and perform variable selection by incorporating structural information from Mass Spectral Imaging (MSI) data. We explore the use of a smooth quadratic penalty to model the natural ordering of the physical variables, that is the mass-to-charge (m/z) ratios. Thereby, estimated model parameters for nearby variables are enforced to smoothly vary. Similarly, to overcome the lack of labeled data we model the spatial proximity among spectra by means of a connectivity graph over the set of predicted labels. We explore the usefulness of this approach in a mouse brain MSI data set.

Keywords: MSI, sparsity, ordered variables, spatial information, smoothing penalty, graph Laplacian, convex optimization, regularization.

1 Introduction

Mass spectral imaging (MSI) is a developing technology that allows the detection of biomolecules such as proteins, peptides, and metabolites from organic tissue while retaining the spatial information intact [1]. Thus, MSI enables the study of the spatial tissue distribution for any detectable molecule that falls within a specified molecular mass range [2]. A typical MSI experiment consists of a grid of measurement locations or pixels covering the tissue section, with an individual mass spectrum attached to each pixel. The resulting data structure can be considered as three-dimensional array or tensor with two spatial dimensions (h and w) and one mass-over-charge (m/z) dimension as shown in Fig.1.

These characteristics pose challenges in the statistical analysis of MSI data. The high molecular specificity of MSI on one hand, delivers huge dimensional

⋆ Corresponding author.

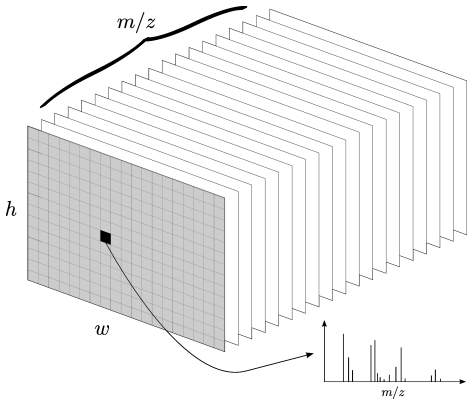


Fig. 1. A schematic representation of the MSI data structure. Individual mass spectra are collected from the tissue area of interest retaining their spatial relationships (h,w). The data is collected into a three-mode array where each *slide* corresponds to a particular (m/z) value and every point in the grid is attached to a spectrum.

data sets with thousands of measured variables that usually exceed the number of spectra (observations), often limited to a few hundreds. On the other hand, the spatial coordinates (h, w) associated to each spectrum define areas of interest and thus should be not neglected. In practice several methods have been already applied to MSI data set including but not limited to principal component analysis [3], clustering and multivariate analysis [4], and supervised classification [5],[6]. Besides the low number of observations, only a small fraction of them is labeled. This hinders many statistical methods and further limits the validation of the obtained results. Manual labeling requires dedicated expertise which can be time consuming, costly and in some cases inaccurate.

In the present article we aim to address most of the aforementioned issues. In a first step, we start from regularized models that impose sparsity in the solution of coefficients. In the problem of interest variables admit a natural ordering due to their physical meaning. Therefore we enforce that the estimated coefficients of nearby variables should smoothly vary in terms of m/z. Unlike the so called *fused lasso* [7] where the absolute value of the differences is used, we employ a smooth quadratic penalty. Furthermore, to overcome the lack of labeled observations we exploit the prior assumption that nearby spectra are likely to have the same label. This is a meaningful assumption for many type of data: for instance a tumor is more likely to affect nearby cells than erratically affect disconnected regions of tissue. Our approach encodes the spatial proximity among spectra by means of a graph and hence can be seen as a semi-supervised method. The resulting proposed model is shown to be equivalent to a *lasso* formulation and therefore can be efficiently solved via the LARS (Least Angle Regression) [8] algorithm. Each component in our optimization problem clearly embodies the structural information of MSI data, whereas regularization parameters trade off

the complexity of the model in terms of sparsity, smoothness and unlabeled samples.

This paper is organized as follows. Section 2 introduces the notions about regularized linear models and notation with respect to MSI data. The general concept of encoding structural information via the graph Laplacian is presented in Section 3, while Section 3.1 deals in detail with the modeling of the ordering of the m/z variables and the resulting optimization problem. Section 3.2, elaborates on the encoding of spatial information using unlabeled samples and states the final proposed approach. Preliminary results on a mouse brain MSI data set are given in Section 4. Comparisons to related algorithms are reported along with visualization and interpretation of the obtained results.

2 Notation and preliminaries

The MSI data set can be represented by a collection of n observations (spectra) measured over p variables (mass-to-charge ratios). The set of labeled spectra is $\mathcal{D}_\ell = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, with $\boldsymbol{x}_i \in \mathbb{R}^p$, $y_i \in \mathbb{R}$, where y_i is the associated label to spectra $\boldsymbol{x}_i \in \mathbb{R}^p$. Denote by x_i^j the j-th component of $\boldsymbol{x}_i \in \mathbb{R}^p$, therefore $\boldsymbol{x}^j = (x_1^j, x_2^j \ldots, x_i^j, \ldots, x_n^j)^\top$ indicates the vector of measurements of a single variable. We deal with the problem of predicting the response y, from a corresponding observation $\boldsymbol{x}$. In this setting we consider the standard linear regression model

$$y_i = \sum_{j=1}^{p} \beta_j x_i^j + \varepsilon_i \ , \tag{1}$$

with errors ε_i. The variables are assumed to be standardized and the output to be centered. The vector of coefficients $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \ldots, \hat{\beta}_p)^\top \in \mathbb{R}^p$ is usually obtained by penalized empirical risk minimization:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 + \lambda P(\boldsymbol{\beta}) \ . \tag{2}$$

Common examples of penalized models are ridge regression with $P(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|^2$, or the *lasso* with $P(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$. The lasso penalty encourages sparse solutions while ridge regression keeps all the coefficients in the model. In general, a priori assumptions encoded via $P(\cdot)$ are needed to make the problem well-posed. In the following sections we aim to model the specific features of MSI data by translating them into useful structural information in the general optimization problem described in (2).

3 Structure Encoding via the Graph Laplacian

In order to incorporate structural information in our model fitting approach, we consider an undirected connectivity graph $\mathcal{G} = (V, E)$, where V is the set of nodes and E the set of edges. An edge between given nodes u and v ($u \sim v$) exists

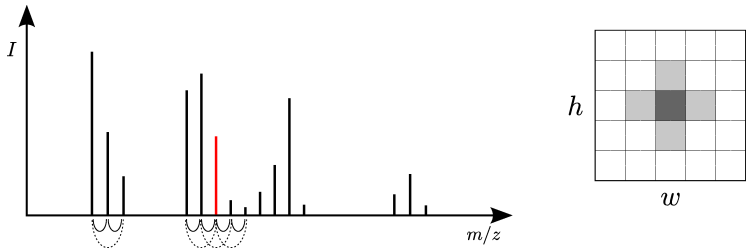


Fig. 2. *Left*: First order (solid) and second order (dashed) connectivity structure over the set of the variables. We consider first order connectivity to impose local smoothness on the coefficients. *Right*: *Cross-like* spatial neighborhood imposed over the set of spectra.

if the entities represented by u and v are linked. Denoting d_u as the degree of a node u, the normalized Laplacian matrix L associated to the graph $\mathcal{G}$ is given by [9]

$$L(u,v) = \begin{cases} 1, & \text{if} \quad u = v \quad \text{and} \quad d_u \neq 0 \ , \\ -1/\sqrt{d_u d_v}, & \text{if} \quad u \sim v \ , \\ 0, & \text{otherwise} \ . \end{cases} \tag{3}$$

The Laplacian is a symmetric semi-positive definite matrix which can be interpreted as an operator on functions of the type $\boldsymbol{f} : V \to \mathbb{R}$ namely vectors indexed by elements of V. It can be shown that [9]

$$\boldsymbol{f}^\top L \boldsymbol{f} = \sum_{u \sim v} \left(\frac{f_u}{\sqrt{d_u}} - \frac{f_v}{\sqrt{d_v}} \right)^2 \ , \tag{4}$$

and hence the quadratic term on the left-hand side of (4) can be used to define a penalty enforcing smooth variation over neighboring nodes. We use this fact to incorporate structural information of MSI into the learning framework.

3.1 Encoding ordered variables

In order to account for the natural ordering of the m/z measurements, we impose a graph $\mathcal{G}^p$ over the set of variables. The set of nodes are associated to the p input variables $\boldsymbol{x}^j, j = 1, ..., p$, thus modeling neighboring variables via the Laplacian matrix of the graph. The structure imposed can is visualized in the left panel in Fig.2, where every m/z variable $\boldsymbol{x}^j$ is connected to the preceding $\boldsymbol{x}^{j-1}$ and the subsequent $\boldsymbol{x}^{j+1}$. One might also consider second order relationships and so forth. By defining $L_\beta \in \mathbb{R}^{p \times p}$ as the Laplacian over the set of variables (cf. (3)) and considering the squared norm in (4) for $\boldsymbol{\beta}$, our regularized optimization problem takes then the form:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \boldsymbol{\beta}^\top L_\beta \boldsymbol{\beta} \ , \tag{5}$$

with $\lambda_1, \lambda_2 > 0$. While the second term enforces sparsity on the $\boldsymbol{\beta}$, the last term smooths the solution of $\boldsymbol{\beta}$ on the network. This is similar to the formulations in [10] and [11]. In the case that no structure is assumed in the network, that is taking $L_\beta = I$, the optimization problem resorts to the elastic net approach [12].

3.2 Encoding prior spatial information

Along the same lines of reasoning, we aim at imposing also a smooth structure on the predicted labels $\hat{y}_k,\ k = 1, ..., n_s$. The spatial distribution of the spectra in the square grid (see Fig.1) suggests that nearby spectra should correspond either to the same tissue area or, might represent connectivity tissues. In essence our goal is to extend the framework exposed in the previous section by incorporating additional information about the spatial structure of the MSI data. In order to get an empirical estimate of spectra distribution we make use of unlabeled examples [13]. Denoting by $\hat{\boldsymbol{y}} = (\hat{y}_1, \ldots, \hat{y}_{n_s})^\top$ the vector of predicted responses and, assigning each $\hat{y}_k$ to a node in a graph $\mathcal{G}^s$, we construct the corresponding Laplacian matrix $L_s \in \mathbb{R}^{n_s \times n_s}$ using (3). The entries of $L_s(h, w)$ are defined according to the *cross-like* neighborhood pattern shown on the right hand side of Fig.2. Likewise, we consider a similar quadratic form for the predicted responses as in (4), that is $\hat{\boldsymbol{y}}^\top L_s \hat{\boldsymbol{y}}$, which is bounded by an user specific parameter $\xi > 0$. Including this constraint into our optimization problem we have

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \boldsymbol{\beta}^\top L_\beta \boldsymbol{\beta} \tag{6}$$

$$\text{s.t.} \quad \hat{\boldsymbol{y}}^\top L_s \hat{\boldsymbol{y}} \leq \xi \tag{7}$$

$$\hat{y}_k = \sum_{j=1}^{p} \beta_j x_k^j,\ k = 1, \ldots, n_s \ . \tag{8}$$

Expressing the vector of equality constrains in matrix form $\hat{\boldsymbol{y}}_s = X_s\boldsymbol{\beta}$, and replacing this term in the inequality constraint we get $\boldsymbol{\beta}^\top (X_s)^\top L_s (X_s)\boldsymbol{\beta} = \boldsymbol{\beta}^\top G_s \boldsymbol{\beta}$. By introducing a Lagrange multiplier $\lambda_3 > 0$ for the latter constraint, we write (6) as the following unconstrained optimization problem

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \boldsymbol{\beta}^\top L_\beta \boldsymbol{\beta} + \lambda_3 \boldsymbol{\beta}^\top G_s \boldsymbol{\beta} \ . \tag{9}$$

By grouping the quadratic terms of $\boldsymbol{\beta}$ and defining $H_{\lambda_3} = L_\beta + \frac{\lambda_3}{\lambda_2} G_s$, we can further cast the optimization problem into a *lasso* type formulation

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \arg\min_{\beta} \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 + | + \|\mathbf{0}_{p\times 1} - \sqrt{\lambda_2} H_{\lambda_3}^{1/2} \boldsymbol{\beta}\|_2^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| \ , \\ &= \left\| \begin{bmatrix} \boldsymbol{y} \\ \mathbf{0}_{p\times 1} \end{bmatrix} - \begin{bmatrix} X \\ \sqrt{\lambda_2} H_{\lambda_3}^{1/2} \end{bmatrix} \boldsymbol{\beta} \right\|_2^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| \ . \end{aligned} \tag{10}$$

This modified problem has dimensions $(n+p) \times p$, and can be solved using the LARS (Least Angle Regression) algorithm [8] up to fixing λ_2 and solving the regularization path for the constrained version using a bound $|\boldsymbol{\beta}|_1 \leq \tau$ instead of λ_1. This equation is our final proposed method to incorporate spatial information and to impose neighboring structure in the ordered variables.

By setting $\lambda_3 = 0$, we can relate our method to similar existing approaches. For instance, the Elastic net [12] is obtained by setting $L_\beta = I$. The network constrained-regularization (NET)[10] and the multiple NET [11] are not restricted to ordered variables and instead they impose prior groupings through graphs. The fused lasso algorithm in [7] penalized the absolute difference between adjacent coefficients whereas the the *group lasso* [14] assumes in advance groups of variables.

4 Experimental results

In this section we explore the usefulness and applicability of the proposed method to include the structural information of MSI data. While numerical validation of the obtained results is assessed via 10-fold cross-validation, visual interpretation and comparison appear more intuitive by translating the results into exploratory ion images. This visual aid compensates the limited availability of ground truth information.

4.1 Data set

The data set, acquired at University Hospital Leuven, comes from a sagittal section of mouse brain [15]. The spatial grid covering the tissue has 51×34 measurement locations (i.e. 1734 pixels). Each measurement spans a mass range from 2800 to 25000 Dalton in 6490 mass-over-charge (m/z) bins. Therefore, the data structure contains 1734 mass spectra measuring 6490 m/z variables per spectrum. Partial labeling information of 279 spectra is provided by a pathologist corresponding to four anatomical regions within the tissue. The labeled regions are the cerebellar cortex (cc), Ammon's horn in the hippocampus (ca), the caudaputamen (cp), and the lateral ventricle (vl) area. Figure 3(a) depicts the four partially labeled regions overlaid on a gray level microscopic image of the mouse brain section. The set of spectra is normalized with respect to the total ion current and is baseline corrected.

4.2 Numerical results

In order to set suitable values for the three regularization parameters, we first define a grid of values over the parameters λ_2 and λ_3. Secondly, for every pair of values we approximate the regularization path for the parameter τ (associated to λ_1) and pick the best combination via 10-fold cross-validation. In Table 1, we

report the results of the proposed approach among pairwise classes. Chosen values for the regularization parameters are reported along with the average 10-fold cross-validation accuracy and the number of non-zero β coefficients. Similarly, the performance of the lasso and elastic net algorithms are reported in Table 2.

Table 1. Multi-class one-vs-one results of the proposed approach. Regularization parameters associated to the quadratic penalties (λ_2, λ_3) are chosen from the grid $[10^{-3},\ 10^{-2},\ 10^{-1},\ 1,\ 10,\ 100]^2$. The regularization path for parameter τ (bound on the L_1 norm) associated to λ_1 is optimized via 10-fold crossvalidation on the labeled data.

Classes	τ^*	λ_2^*	λ_3^*	Non-zero β	10-fold mse	10-fold accuracy
cc vs *ca*	0.165	100	0.001	64	1.0529 (0.5482)	1 (0)
cc vs *cp*	0.213	100	0.001	106	1.3783 (0.3739)	0.9288 (0.1076)
cc vs *vl*	0.249	1	0.001	56	1.7588 (0.9709)	0.8938 (0.1719)
ca vs *cp*	0.162	1	0.001	54	2.3778 (0.9005)	0.9758 (0.0319)
ca vs *vl*	0.1640	100	0.01	102	3.2342 (1.2151)	0.9446 (0.0447)
cp vs *vl*	0.0460	10	0.1	14	5.9332 (1.6775)	0.9288 (0.0580)

Table 2. Multi-class one-vs-one results of the LASSO (lasso) and Elastic net (enet) algorithms.

	lasso			enet			
Classes	Non-zero β	τ^*	10-fold accuracy	Non-zero β	τ^*	λ_2^*	10-fold accuracy
cc vs *ca*	14	0.29	1.0000 (0.0)	17	0.211	0.001	1.0000 (0.0)
cc vs *cp*	10	0.16	0.9738 (0.0532)	15	0.136	0.01	0.9905 (0.0202)
cc vs *vl*	34	0.374	0.9250 (0.1208)	17	0.145	0.01	0.9333 (0.1097)
ca vs *cp*	31	0.212	0.9740 (0.0436)	18	0.1	0.001	0.9687 (0.0477)
ca vs *vl*	11	0.085	0.9143 (0.0732)	19	0.076	0.1	0.9330 (0.0549)
cp vs *vl*	5	0.031	0.9142 (0.0739)	6	0.033	0.001	0.9123 (0.0766)

Additionally, the performance for the combined *one-versus-one* predictions is presented in Table 3. All the three methods perform slightly similar with appreciable differences in the average number of coefficients. The proposed method tends to select more coefficients due to the effect of the two square penalties.

4.3 Visualization

By translating the predicted labels back to their position in the spatial domain, one can directly assess the performance of the algorithm via visual inspection.

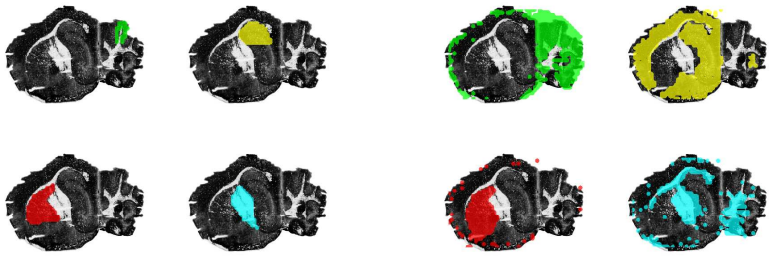
Table 3. Combined multi-class results on the MSI mouse brain rat data set.

Method	Avg. Non-zero β_j	Avg. $\|\beta\|_1$	10-fold accuracy
lasso	17	1.0303	0.9453 (0.0690)
enet	16	0.9608	0.9563 (0.0515)
proposed	66	1.6647	0.9502 (0.0608)

Figure 3 displays the combined *one-vs-one* predicted labels corresponding to the compared methods. All the three models effectively separate the lateral ventricle (vl) and cauda-putamen (cp) from the surrounding tissue. The classification for the ventricle area additionally draws in the elongated corpus callosum and cerebellar nucleus regions as well, as these regions share a panel of common molecules within the measured mass range. The cerebellar cortex (cc) label exceeds its intended boundaries due to the small number of labeled spectra (21 observations). The remaining hippocampus label (ca) extends to capture the complete hippocampus and most of the remaining unlabeled areas of the tissue. Furthermore, to visualize important selected variables, we look at the top three variables associated to the largest β coefficients. In particular we take those differentiating the lateral ventricle (vl) from the cauda-putamen (cc). In Fig.4, ion images highlight the presence of three of the top selected m/z in these two anatomical regions.

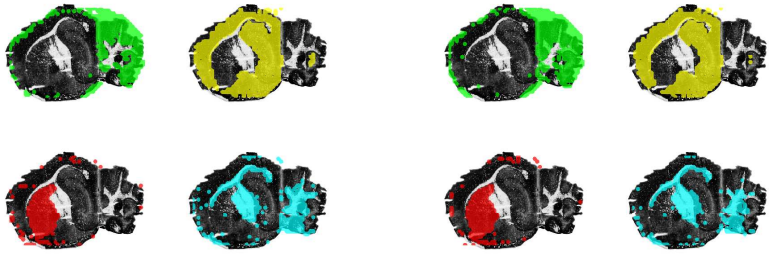
5 Conclusion

In this article we have presented a methodology to learn semi-supervised sparse linear models in MSI data. Starting from regularized learning models and structural information inherent to MSI data, we make use of the graph Laplacian to embed first, the natural ordering of the m/z variables and, secondly the spatial location of the spectra. Thereby, smooth quadratic penalties are imposed over neighboring *nodes* representing in the first case *variables* and in the second one *observations*. These penalties modify the standard learning algorithm resulting in an equivalent *lasso* formulation that can be solved efficiently. Moreover the lack of labeled data, typical of MSI experiments, is circumvented through modeling the predicted responses via the graph Laplacian. The applicability of the proposed approach is explored in a mouse brain MSI data set to distinguish amongst four anatomical regions, and it is compared to other learning models that do not, or partially, incorporate the structural information of MSI data. The presented case study shows that sparse linear models can already provide significant informative insight to assess tissue type, structure, and content. Additionally, our approach also holds value for more fundamental exploratory studies of tissue as it can highlight similarity in content between different tissue areas. Further work in this direction seems promising and should find applicability as more MSI data sets become available.



(a) cc (green), ca (yellow), cp (red), vl (cyan)

(b) lasso



(c) enet

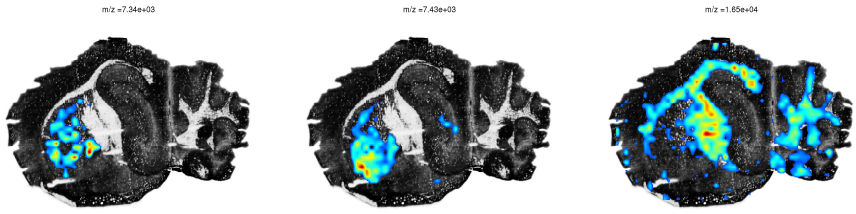
(d) proposed

Fig. 3. Labeled areas and corresponding predicted labels by the algorithms.

Acknowledgments. *Research Council KUL*: ProMeta, GOA Ambiorics, GOA MaNet, CoE EF/05/007 SymBioSys, START 1, Optimization in Engineering (OPTEC), IOF-SCORES4CHEM, several PhD/postdoc & fellow grants. *Flemish Government*: FWO: PhD/postdoc grants, G.0318.05 (subfunctionalization), G.0553.06 (VitamineD), G.0302.07 (SVM/Kernel), G0226.06 (cooperative systems and optimization), G0321.06 (Tensors), G.0302.07 (SVM/Kernel), G.0320.08 (convex MPC), G.0558.08 (Robust MHE), G.0557.08 (Glycemia2), G.0588.09 (Brain-machine) research communities (ICCoS, ANMMM, MLDM); G.0377.09 (Mechatronics MPC), G.0733.09 (3UTR), G.082409 (EGFR). IWT: PhD Grants, Eureka-Flite+, SBO LeCoPro, SBO Climaqs, SBO POM, O&O-Dsquare, Silicos, SBO-BioFrame, SBO-MoKa, TBM-IOTA3. FOD:Cancer plans. *Belgian Federal Science Policy Office*: IUAP P6/25 (BioMaGNet, Bioinformatics and Modeling: from Genomes to Networks, 2007-2011), IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007-2011). *EU-RTD: ERNSI*: European Research Network on System Identification; FP7-HEALTH CHeartED. FP7-HD-MPC (INFSO-ICT-223854), COST intelliCIS, FP7-EMBOCON (ICT-248940).

References

1. 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



(a) $m/z = 7.34 \times 10^3$ (b) $m/z = 7.43 \times 10^3$ (c) $m/z = 1.65 \times 10^4$

Fig. 4. Ion image visualization for the top three selected m/z variables discriminating the (cc) and (vl) tissue regions. The first two are common to the three compared methods, whereas the third variable at $m/z = 1.65e + 4$ Da that delineates the (vl) area only appears in the proposed model.

tissues. *Nature Medicine*, 7(4):493–496, April 2001.

2. L.A. McDonnell and R.M.A. Heeren. Imaging mass spectrometry. *Mass Spectrometry Reviews*, 26(4):606–643, 2007.
3. R. Van de Plas, F. Ojeda, M. Dewil, L. Van Den Bosch, B. De Moor, and E. Waelkens. Prospective exploration of biochemical tissue composition via imaging mass spectrometry guided by principal component analysis. In *Proceedings of the Pacific Symposium on Biocomputing 12*, pages 458–469, Maui, 2007.
4. G. McCombie, D. Staab, M. Stoeckli, and R. Knochenmuss. Spatial and spectral correlations in MALDI mass spectrometry images by clustering and multivariate analysis. *Analytical Chemistry*, (19):6118–6124, 2005.
5. M. Hanselmann, U. Köthe, M. Kirchner, B.Y. Renard, E.R. Amstalden, K. Glunde, R.M.A. Heeren, and F.A. Hamprecht. Toward digital staining using imaging mass spectrometry and random forests. *Journal of Proteome Research*, 8(7):3558–3567, 2009.
6. J. Luts, F. Ojeda, R. Van de Plas, B. De Moor, S. Van Huffel, and J.A.K. Suykens. A tutorial on support vector machine-based methods for classification problems in chemometrics. *Analytica Chimica Acta*, 665(2):129 – 145, 2010.
7. R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *Journal of The Royal Statistical Society Series B*, 67(1):91–108, 2005.
8. B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):407–451, 2004.
9. F.R.K. Chung. *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92)*. American Mathematical Society, February 1997.
10. C. Li and H. Li. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, 24(9):1175–1182, 2008.
11. M. Signoretto, A. Daemen, C. Savorgnan, and J.A.K. Suykens. Variable selection and grouping with multiple graph priors. In *2nd Neural Information Processing Systmes (NIPS) Workshop on Optimization for Machine Learning*, 2009.
12. H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320, 2005.

13. M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7:2399–2434, November 2006.
14. M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B*, 68:49–67, 2006.
15. R. Van de Plas, K. Pelckmans, B. De Moor, and E. Waelkens. Spatial querying of imaging mass spectrometry data: A nonnegative least squares approach. In *Neural Information Processing Systems Workshop on Machine Learning in Computational Biology*, 2007.