

ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, IL 60439

**Non-negative matrix analysis in x-ray spectromicroscopy:
Choosing regularizers**

Rachel Mak, Stefan M. Wild, and Chris Jacobsen

Mathematics and Computer Science Division

Preprint ANL/MCS-P5328-0415

April 2015

Non-negative matrix analysis in x-ray spectromicroscopy: choosing regularizers

Rachel Mak^{*}, Stefan M. Wild[†] and Chris Jacobsen^{**,*}

^{*}*Dept. Physics & Astronomy, Northwestern University, Evanston, Illinois, USA*

[†]*Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois, USA*

^{**}*X-ray Science Division, Advanced Photon Source, Argonne National Laboratory, Argonne, Illinois, USA*

Abstract. In x-ray spectromicroscopy, a set of images can be acquired across an absorption edge to reveal chemical speciation. We previously described the use of non-negative matrix approximation methods for improved classification and analysis of these types of data. We present here an approach to find appropriate values of regularization parameters for this optimization approach.

Keywords: x-ray spectromicroscopy; x-ray microscopy; multivariate statistical analysis; non-negative matrix analysis; optimization

PACS: 68.37.Yz, 82.80.Ej

ANALYSIS OF SPECTROSCOPIC IMAGING DATA

Near-absorption edge resonances provide information on chemical speciation of the element in question. In biological and environmental systems, one might have a rich set of partially overlapping resonances, making interpretation difficult. In addition, chemical variation can occur on nanometer-length scales, so that spectroscopic imaging is required in order to see the regions with distinct chemical characteristics.

Several multivariate statistical analysis approaches have been used in spectroscopic microscopies. For cases where one does not have a set of well-defined reference spectra, we have described the use of cluster analysis [1] to find the dominant spectroscopic themes in a dataset of optical density $D(x, y, E) = -\ln[I(x, y, E)/I_0(E)] = [\mathbf{D}_{N \times P}]_{x,y}$, where P denotes the total count of all pixels, as stored in a 1D array, and N is the number of photon energies in the per-pixel spectra. Our goal is to describe the measured optical density in terms of a set of absorption spectra $\boldsymbol{\mu}_{N \times S}$ for S spectroscopically distinguishable chemical components in the sample, each associated with a thickness (or “weighting”) map $\mathbf{t}_{S \times P}$. That is, we wish to obtain the matrices $\boldsymbol{\mu}_{N \times S}$ and $\mathbf{t}_{S \times P}$, with S much smaller than both N and P , such that our data can be represented by

$$\mathbf{D}_{N \times P} \approx \boldsymbol{\mu}_{N \times S} \mathbf{t}_{S \times P}. \quad (1)$$

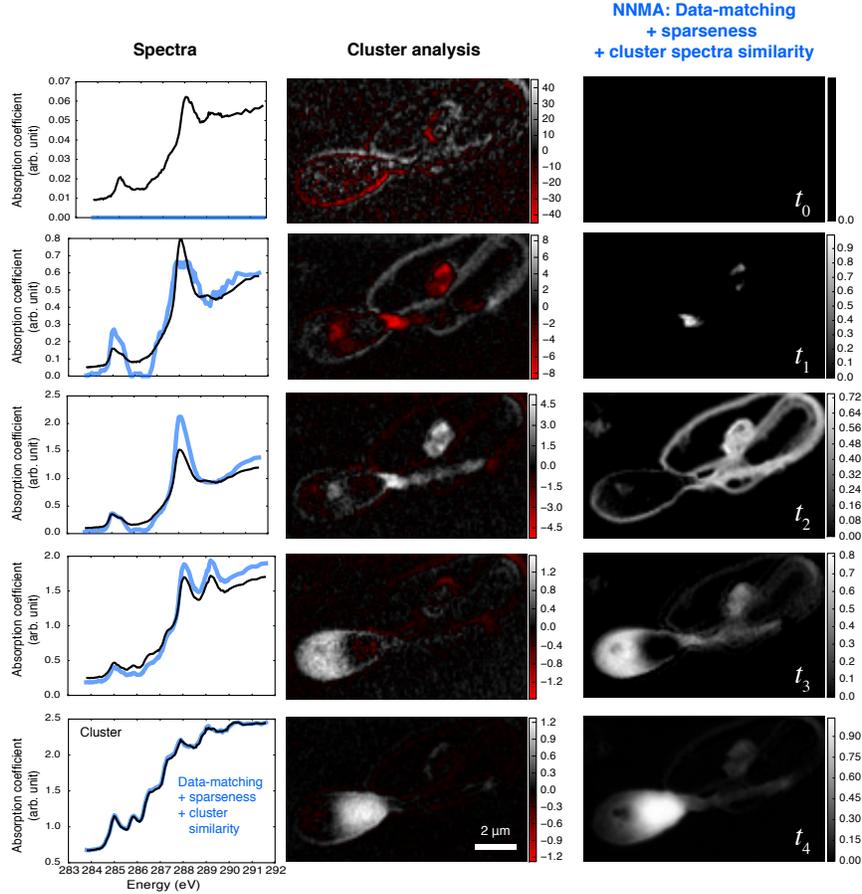
We have shown in [2] that clustering-based analysis approaches are not guaranteed to be free of aphysical characteristics such as the appearance of negative optical densities in the analysis results, while non-negative matrix approximation (NNMA, also sometimes called “non-negative matrix factorization” [3]) methods can yield improved interpretability. The results in [2], along with references therein, have led to improvements in spectromicroscopy analysis of carbon near-edge x-ray absorption data from human sperm; see Fig. 1.

OPTIMIZATION-BASED APPROACH USING REGULARIZERS

The approach we have described in [2] involves the constrained minimization of a global cost function F of the form

$$\begin{aligned} F(\boldsymbol{\mu}, \mathbf{t}) &= F_0(\boldsymbol{\mu}, \mathbf{t}) + \lambda_t F_1(\boldsymbol{\mu}, \mathbf{t}) + \lambda_\mu F_2(\boldsymbol{\mu}, \mathbf{t}) \\ &= \|\mathbf{D} - \boldsymbol{\mu} \mathbf{t}\|_F^2 + \lambda_t \|\mathbf{t}\|_1 + \lambda_\mu \|\boldsymbol{\mu} - \boldsymbol{\mu}_{\text{cluster}}\|_2^2, \end{aligned} \quad (2)$$

where $F_0(\boldsymbol{\mu}, \mathbf{t}) = \|\mathbf{D} - \boldsymbol{\mu} \mathbf{t}\|_F^2$ is a Frobenius norm “data-matching” cost based on Eq. 1; $F_1(\boldsymbol{\mu}, \mathbf{t}) = \|\mathbf{t}\|_1$ is a “sparsity” regularizer based on minimizing the one-norm of the thickness map \mathbf{t} (thus favoring solutions where each spectroscopically distinguishable component shows up in as few pixels as possible); and $F_2(\boldsymbol{\mu}, \mathbf{t}) = \|\boldsymbol{\mu} - \boldsymbol{\mu}_{\text{cluster}}\|_2^2$ is a regularizer seeking minimal adjustments to a target set of spectra $\boldsymbol{\mu}_{\text{cluster}}$ (in our case, found using cluster analysis, since these are



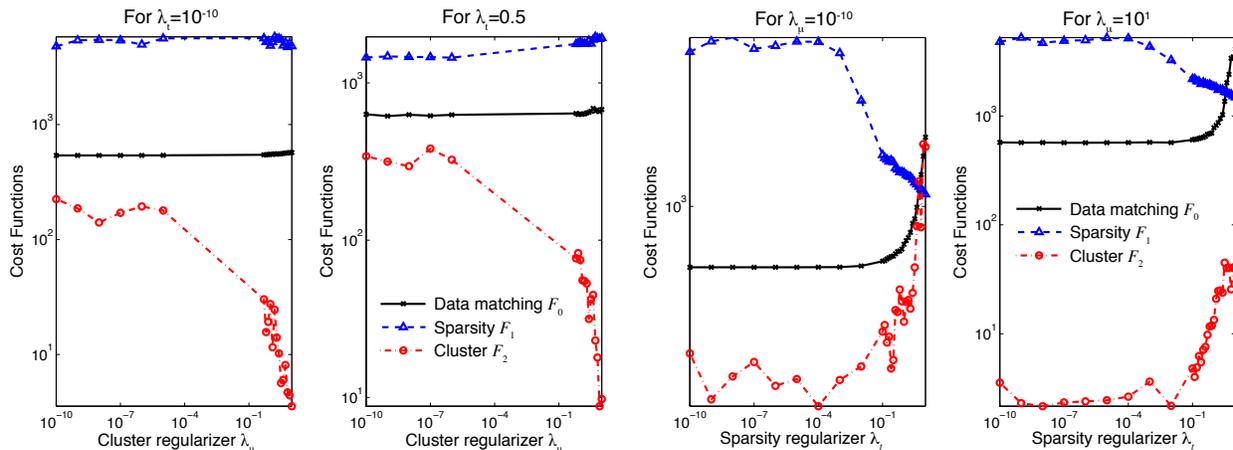
1: X-ray spectromicroscopy analysis results obtained by cluster analysis [1] and NNMA [2] including data matching, sparseness, and cluster similarity regularizers (Eq. 2, $\lambda_\mu = 10$ and $\lambda_\tau = 0.5$). Cluster analysis produces maps with negative weightings for some regions, which are not allowed by the physics of x-ray absorption, whereas the NNMA solution delivers an analysis result with recognizable x-ray absorption spectra and positive weightings or thickness maps. This NNMA result nicely illustrates the large-scale biochemical organization of sperm. Image t_3 highlights the acrosomal cap, flagellar motor, and mitochondrion; image t_4 highlights the nucleus where histones are bound to DNA; image t_2 highlights the lipid membrane and flagellum; and image t_1 isolates a high-density area in the flagellar motor with some combination of chemical sensitivity and experimental absorption saturation limits. Image t_0 shows residual errors in the cluster analysis map; no such errors are present in the t_0 map when using NNMA, so for NNMA the t_0 image is black to reflect this fact.

spectra obtained by averaging measurements from pixels with similar spectroscopic response). In [2], we described an algorithm for minimizing Eq. 2 subject to the constraints that $\boldsymbol{\mu}$ and \mathbf{t} be non-negative.

If one were to think of the total cost F as a financial one, the challenge would be that the component costs F_0 , F_1 , and F_2 are denominated in different currencies. What is the correct “exchange rate” between these costs? This is the role of the scalar regularization parameters $\lambda_\tau \geq 0$ and $\lambda_\mu \geq 0$. Clearly, we can obtain the lowest global cost $F(\boldsymbol{\mu}, \mathbf{t})$ if we set both regularization parameters to zero, but this would be done at the expense of losing the desired characteristics of sparseness or cluster spectra similarity. Formally, we can view our problem as one of multiobjective optimization [4], where we would like to trade off the three competing objectives F_0 , F_1 , and F_2 against one another. Such an approach would lead to (infinite) sets of approximations, with each approximation $(\mathbf{t}, \boldsymbol{\mu})$ being Pareto-optimal, meaning that no other $(\hat{\mathbf{t}}, \hat{\boldsymbol{\mu}})$ exists that is better in all three objectives simultaneously.

In this work, we focus on a more computationally tractable approach in studying the tradeoffs among the three objectives. In particular, minimizing the cost function in Eq. 2 for all possible values of $(\lambda_\tau, \lambda_\mu)$ corresponds to

generating the Pareto-optimal solutions in special cases. We study these cases by considering a wide range of regularization parameter values and examining the effects on the three component costs.



2: The three component costs (log scale) F_0 (“data”), F_1 (“sparsity”), and F_2 (“cluster”) obtained by minimizing Eq. 2 for particular (λ_t, λ_μ) values. A. (left) Costs are shown as a function of the cluster-spectra-matching regularizer λ_μ for two fixed values of the sparseness regularizer λ_t . B. (right) Costs are shown as a function of the sparseness regularizer λ_t for two fixed values of the cluster-spectra-matching regularizer λ_μ .

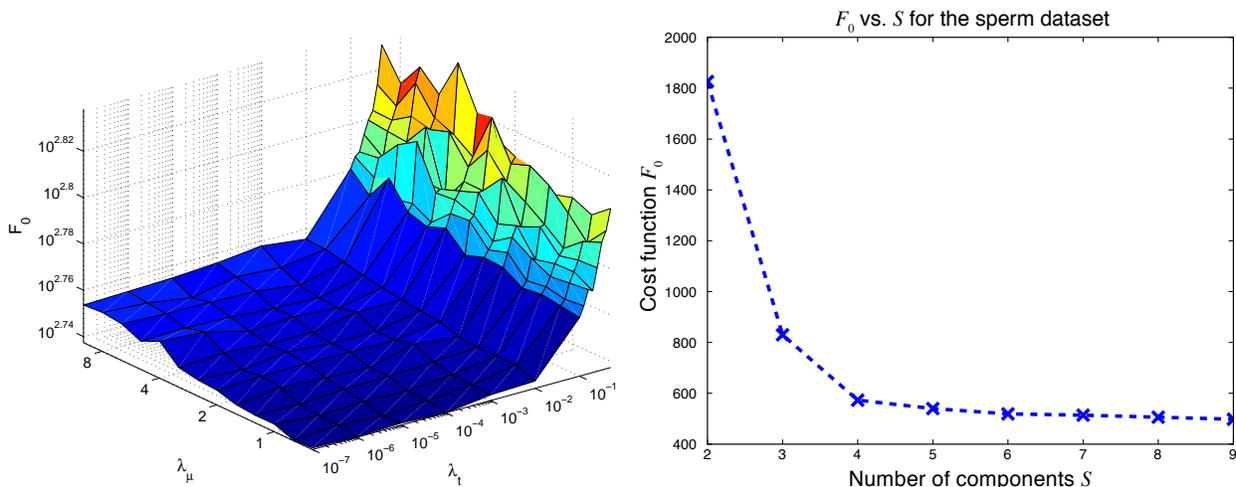
Figure 2 illustrates the resulting tradeoffs among the cost components for the sperm sample and $S = 5$ different components (see Fig. 1). Figure 2A shows that the sparsity and data-matching costs are relatively unaffected by demanding increasing levels of cluster similarity by increasing λ_μ . On the other hand, the behavior in Fig. 2B shows that demanding more sparsity (approximated here by F_1) can result in substantial degradation of both the data-matching and cluster-similarity component costs.

Figure 3A shows the behavior of the data-matching cost F_0 as a function of both λ_t and λ_μ . The resulting surface reiterates that this cost is considerably more sensitive to larger λ_t (sparsity demanding) values than it is to λ_μ (cluster similarity) and smaller λ_t values. We note that the F_0 plots in Fig. 2 correspond to four different slices through the 3D surface in Fig. 3A. Our final choice of the regularizers for sparsity λ_t and cluster similarity λ_μ is one which emphasizes very little rise in the most critical cost to minimize, which is the data-matching cost $F_0(\boldsymbol{\mu}, \mathbf{t})$.

SELECTING THE NUMBER OF COMPONENTS S

An important parameter in both NNMA and cluster analysis is the selection of the number of spectroscopically distinguishable components S to seek. If S is too small, we will arrive at a basis set that cannot reproduce all the important variations in the data; if S is too large, we may have simply analyzed variations due to noise from photon statistics or other sources. In principal component analysis, the ordering on a scree plot of the eigenvalues of the covariance matrix $\mathbf{Z} = \mathbf{D}\mathbf{D}^T$ can provide a good estimate of S . If the plot exhibits an “elbow,” beyond which eigenvalues are decreasing only slightly, this can indicate the factor order that marks a transition from variations in significant signals to only small variations due to noise factors [5]. In cluster analysis, we have used this technique to estimate the number of significant components \bar{S} in the sample [1], although in practice the precise number of factors to use may not be clear, and it has been found to be important to manually examine the analysis result. The number of clusters to seek would then be \bar{S} .

Since NNMA analysis involves the cost function $F_0(\boldsymbol{\mu}, \mathbf{t})$ that measures how well the solution $\boldsymbol{\mu}\mathbf{t}$ matches the data \mathbf{D} , we have a good basis for evaluating the effect of decreasing or increasing the number S of spectroscopically distinguishable components. By carrying out NNMA analysis with a range of values for S , we can see when the error $F_0(\boldsymbol{\mu}, \mathbf{t})$ no longer decreases as a function of S (Fig. 3B); we can similarly examine when decreases to S are insufficient to capture all the important spectroscopic variations in the sample. This topic will be explored further in future work, where we intend to examine how the component cost function tradeoffs change as a function of the number S of spectroscopically distinguishable components.



3: A. (left) Data-matching component costs (log scale) $F_0(\mathbf{t}^*(\lambda_t, \lambda_\mu), \boldsymbol{\mu}^*(\lambda_t, \lambda_\mu))$ obtained by minimizing Eq. 2 for different (λ_t, λ_μ) values. B. (right) By varying the number of spectroscopic components S and calculating its effect on the data-matching cost function F_0 , we can estimate the number of distinguishable components needed to account for the spectroscopic variation in the data. We advocate using 4 or 5 for S ; adding more components would not significantly reduce the cost function.

DISCUSSION

We have used a generalized cost function in NNMA analysis to factorize an optical density map into spectroscopic component factors and their associated weightings. Regularization parameters in the cost function control the balance between the need for the solution to closely match the data and the desired characteristics the solution should possess.

We have implemented a systematic and computationally tractable method for exploring the effects of a large range of regularization parameters on the cost function. An exploratory version of the NNMA analysis approach described here is implemented in a Python open-source code¹ called MANTiS [6], developed by 2nd Look Consulting; a more refined interface to NNMA analysis is planned for an upcoming release of MANTiS. For the data shown here, the combined cost function converged to a minimum over about 10,000 iterations, taking about 10 minutes on a laptop computer. These results show the potential of NNMA analysis on complicated data.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Energy (DOE) Office of Science under contract DE-AC02-06CH11357. Earlier exploration of NNMA as applied to sperm spectromicroscopy analysis was supported by the National Institutes of Health under grant R01 EB-000479; RM was supported in part by grant R01 GM-104530.

REFERENCES

1. M. Lerotic, C. Jacobsen, T. Schäfer, and S. Vogt, *Ultramicroscopy* **100**, 35–57 (2004).
2. R. Mak, M. Lerotic, H. Fleckenstein, S. Vogt, S. M. Wild, S. Leyffer, Y. Sheynkin, and C. Jacobsen, *Faraday Discussions* **171**, 357–371 (2014).
3. D. D. Lee, and H. S. Seung, *Nature* **401**, 788–791 (1999).
4. M. Ehrgott, *Multicriteria Optimization*, Springer-Verlag, 2005, 2nd edn.
5. E. R. Malinowski, *Factor Analysis in Chemistry*, John H. Wiley & Sons, New York, 1991, second edn.
6. M. Lerotic, R. Mak, S. Wirick, F. Meirer, and C. Jacobsen, *Journal of Synchrotron Radiation* **21**, 1206–1212 (2014).

¹ <http://bitbucket.org/mlerotic/spectromicroscopy>

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up, nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.