Robust Non-negative Tensor Factorization, Diffeomorphic Motion Correction, and Functional Statistics to Understand Fixation in Fluorescence Microscopy

Neel Dey[1], Jeffrey Messinger[2], R. Theodore Smith[3], Christine A. Curcio[2], and Guido Gerig[1]

[1] Department of Computer Science and Engineering, New York University, Brooklyn, NY 11201, USA
[2] Department of Ophthalmology and Visual Sciences, University of Alabama at Birmingham, Birmingham, AL 35294, USA
[3] Department of Ophthalmology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

Abstract. Fixation is essential for preserving cellular morphology in biomedical research. However, it may also affect spectra captured in multispectral fluorescence microscopy, impacting molecular interpretations. To investigate fixation effects on tissue, multispectral fluorescence microscopy images of pairs of samples with and without fixation are captured. Each pixel might exhibit overlapping spectra, creating a blind source separation problem approachable with linear unmixing. With multiple excitation wavelengths, unmixing is intuitively extended to tensor factorizations. Yet these approaches are limited by nonlinear effects like attenuation. Further, light exposure during image acquisition introduces subtle Brownian motion between image channels of non-fixed tissue. Finally, hypothesis testing for spectral differences due to fixation is non-trivial as retrieved spectra are paired sequential samples. To these ends, we present three contributions, (1) a novel *robust* non-negative tensor factorization using the β-divergence and $L_{2,1}$-norm, which decomposes the data into a low-rank multilinear and group-sparse non-multilinear tensor without making *any* explicit nonlinear modeling choices or assumptions on noise statistics; (2) a diffeomorphic atlas-based strategy for motion correction; (3) a non-parametric hypothesis testing framework for paired sequential data using functional principal component analysis. `PyTorch` code for robust non-negative tensor factorization is available at `https://github.com/neel-dey/robustNTF`.

1 Introduction

Imaging spectroscopy is a ubiquitous tool to investigate the chemical nature of biological samples. Each pixel contains an entire spectrum in a desired modality (e.g., reflectance) whose multivariate nature allows for rich applications in image analysis such as segmentation and unmixing. However, prior to any form of

imaging, fixation is performed to protect tissue from postmortem decay and improve mechanical strength. Doing so allows for the preparation of thin and storable tissue sections, a universal requirement. However, as fixation changes the chemistry of a sample to prevent decay, it may further alter its spectra and confound its interpretation. In image analysis applications where the spectra are of interest (e.g., detecting malignancy in surgical biopsies), it is imperative to verify whether fixation changes their shape characteristics [5].

Towards understanding this effect for applications in retinal biology and pathology, we perform a self-controlled experiment. Human eyes were obtained in pairs from deceased donors, and tissue sections from corresponding locations in both eyes were imaged. Sections from the right and left eyes were fixed and not fixed, respectively. Imaging was performed with multispectral fluorescence microscopy with multiple light sources (excitation wavelengths), allowing for the high-dimensional analysis of spectral differences in these paired samples. However, several image analysis challenges arise from this experimental design.

First, compound co-localization leads to pixels containing mixed spectra. To separate constituent spectra and retrieve their spatial distributions in the image, non-negative matrix and tensor factorizations are used when there are one or more light sources, respectively [2, 8]. Non-negative data analysis is required as physical spectra cannot be negative. However, these models are deficient when there are nonlinear effects present like absorption and scattering within tissue. A nonlinear matrix/tensor factorization was proposed in [6] which incorporated attenuation but required reference spectra, precluding several applications.

Second, without fixation, tissue heating during image acquisition causes subtle Brownian motion of organelles between image channels. Thus, the observed spectrum at a pixel may be erroneous and can be a source of noise to analysis. Third, our measurements are paired *sequential* observations. This precludes the use of standard paired hypothesis tests to test for differences caused by fixation as they assume normality and ignore the sequential nature of spectral curves.

Therefore, our image and statistical analysis contributions are threefold. First, we propose a novel robust non-negative tensor factorization (rNTF) that decomposes the tensor of multi-excitation multispectral images into a low-rank multilinear tensor and an additional group-sparse tensor which contains the nonlinearities. Existing methods for tensor factorization often make strong assumptions on noise statistics, whereas fluorescence microscopy is affected by a mixture of Poisson and Gaussian noise [12]. Therefore, we minimize the β-divergence, an information geometric metric which allows us to interpolate between noise statistics assumptions [3]. Group-sparsity is enforced on the resulting nonlinear tensor via $L_{2,1}$-norm regularization. We iteratively impute missing values common in fluorescence measurements via expectation maximization [11]. The algorithm has element-wise updates and is thus executed on GPUs for fast execution.

Second, we propose a granular motion correction strategy in fluorescence microscopy using unbiased diffeomorphic atlas building [7], where all images in the stack are nonlinearly registered to an estimated template, minimizing subtle intra-stack motion without tearing the image due to the diffeomorphic

constraint. Third, inspired by the functional testing of fractional anisotropy along axonal tracts between groups in diffusion tensor imaging [9], we propose a framework for *paired* hypothesis testing of spectra using functional principal component analysis [10] and the Wilcoxon signed-rank test.

2 Methods

Preliminaries   We use the notation of $\mathbf{x}$ for a vector, $\mathbf{X}$ for a matrix, and $\mathcal{X}$ for a tensor. The tensor columns are mode-1 fibers, rows are mode-2 fibers and so on. Mode-i matricization refers to taking the mode-i fibers of a tensor and arranging them as columns of a matrix. The rank-P factorization of a tensor $\mathcal{L} \in \mathbb{R}^{I \times J \times K}$ can be formulated as the sum of P rank-one tensors, such that $\mathcal{L} \approx \sum_{i=1}^{P} \mathbf{a}_i \circ \mathbf{b}_i \circ \mathbf{c}_i$ where $\mathbf{a}_i \in \mathbb{R}^I$, $\mathbf{b}_i \in \mathbb{R}^J$, and $\mathbf{c}_i \in \mathbb{R}^K$. Factor matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are generated from $\mathcal{L}$ by concatenating vectors from the rank-one components, such that $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_P]$, $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_P]$, and $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_P]$.

Robust Non-negative Tensor Factorization   Robust tensor factorizations have a rich recent history in machine learning and computer vision [13]. Given a tensor $\mathcal{M}$ corrupted with gross outliers, it is possible to recover a low rank and sparse combination of tensors ($\mathcal{L}$ and $\mathcal{S}$, respectively) such that $\mathcal{M} \approx \mathcal{L} + \mathcal{S}$. Recently, there has been interest in replacing the squared Euclidean distance error term with other metrics and divergences which may accommodate other forms of data and noise statistics [4]. Further, there is interest in detecting structured outliers (common in medical imaging), motivating the replacement of L_1-norm regularization with the $L_{2,1}$-norm which induces group sparsity [13].

Consider a rank-P robust tensor factorization of $\mathcal{M} \approx \sum_{i=1}^{P} \mathbf{a}_i \circ \mathbf{b}_i \circ \mathbf{c}_i + \mathcal{S}$, where $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathcal{S} \geq 0$. Block coordinate descent using various tensor matricizations is the workhorse algorithm for calculating tensor factorizations and is adopted here by iteratively fixing three out of four quantities $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathcal{S}$ and solving for the remaining one. Using the formulation of factor matrices, the factorization can be written as, $\mathbf{M}_{(1)} \approx \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{S}_{(1)}$, $\mathbf{M}_{(2)} \approx \mathbf{B}(\mathbf{C} \odot \mathbf{A})^T + \mathbf{S}_{(2)}$ and $\mathbf{M}_{(3)} \approx \mathbf{C}(\mathbf{B} \odot \mathbf{A})^T + \mathbf{S}_{(3)}$ where $\mathbf{M}_{(i)}$ and $\mathbf{S}_{(i)}$ are the mode-i matricized representation of the tensors $\mathcal{M}$ and $\mathcal{S}$ respectively, and $\odot$ is the matrix Khatri-Rao product. Given the above considerations, we propose to solve the following model alternating between all matricizations, where k is the matricization mode,

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathcal{S}} \mathbf{E}(\mathbf{A},\mathbf{B},\mathbf{C},\mathcal{S}) = \mathbf{D}_\beta(\mathbf{M}_{(\mathbf{k})}, \mathbf{L}_{(\mathbf{k})} + \mathbf{S}_{(\mathbf{k})}) + \lambda\|\mathbf{S}_{(k)}\|_{2,1},$$

such that $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{S} \geq 0$, where $\mathbf{D}_\beta(\cdot, \cdot)$ is the beta divergence, and $\|\cdot\|_{2,1}$ is the $L_{2,1}$-norm, such that $\|\mathbf{S}_{(\mathbf{k})}\|_{2,1} = \sum_{i=1}^{G} \|\mathbf{s}_i\|_2$ where $\mathbf{S}_{(\mathbf{k})}$ has G columns. The β-divergence is an information-geometric measure of fit parameterized by a scalar β, which takes the squared Euclidean, Kullback-Leibler and Itakura-Saito divergences (corresponding to Gaussian, Poisson or Gamma noise assumptions) as limiting cases corresponding to $\beta = 2, 1, 0$ and all interpolating cases in between.

Consider the matricization $\mathbf{M}_{(1)} \approx \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{S}_{(1)}$ to solve for $\mathbf{A}$ and $\mathbf{S}_{(1)}$. As derived in [3] for robust NMF, fixing $\mathbf{B}, \mathbf{C}$ and $\mathcal{S}$ allows us to multiplicatively update $\mathbf{A}$ such that $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathcal{S} \geq 0$ in a majorization-minimization framework. Using a convex-concave decomposition of the β-divergence, majorizing the convex and concave parts by the Jensen and Tangent inequalities respectively and minimizing in closed form w.r.t. to $\mathbf{A}$, we get,

$$\mathbf{A} \rightarrow \mathbf{A} * \Big(\frac{(\mathbf{M}_{(1)} * \hat{\mathbf{M}}_{(1)}^{\cdot(\beta-2)})(\mathbf{C} \odot \mathbf{B})}{\hat{\mathbf{M}}_{(1)}^{\cdot(\beta-1)}(\mathbf{C} \odot \mathbf{B})} \Big),$$

where the numerator and denominator undergo element-wise division, $\hat{\mathbf{M}}_{(1)}$ denotes the mode-1 matricization of the current low-rank approximation $\hat{\mathcal{M}}$, '$*$' denotes the element-wise (Hadamard) product, and the '.' operator in the exponents indicates element-wise power. To estimate $\mathbf{S}_{(1)}$, given fixed $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and an $L_{2,1}$ penalty term on $\mathbf{S}_{(1)}$, a similar optimization yields the following update,

$$\mathbf{S}_{(1)} \rightarrow \mathbf{S}_{(1)} * \Big(\frac{\mathbf{M}_{(1)} * \hat{\mathbf{M}}_{(1)}^{\cdot(\beta-2)}}{\hat{\mathbf{M}}_{(1)}^{\cdot(\beta-1)} + \lambda \mathbf{S}_{(1)} \mathrm{diag}(\|\mathbf{s}_1\|_2, \dots, \|\mathbf{s}_G\|_2)^{-1}} \Big),$$

where λ is the regularization weight on the $L_{2,1}$-norm and G is the number of columns in $\mathbf{S}_{(1)}$. The $\mathbf{S}_{(1)} \mathrm{diag}(\|\mathbf{s}_1\|_2, \dots, \|\mathbf{s}_G\|_2)^{-1}$ term in the denominator is the columnwise normalized matrix $\mathbf{S}_{(1)}$ which we compute by looping through the columns instead of direct evaluation for numerical stability. For brevity, we analogously update $\mathbf{B}, \mathbf{S}_{(2)}$ and $\mathbf{C}, \mathbf{S}_{(3)}$, as shown in the supplementary material.

If the input data tensor $\mathcal{M}$ has missing entries (as in our application), we can iteratively estimate the missing values by single imputation [11]. This involves generating an indicator tensor $\mathcal{W}$ with $w_{ijk} = 0$ if m_{ijk} is missing and vice-versa. During the iterations, we impute the missing entries of $\mathcal{M}$ via $\mathcal{M} \leftarrow \mathcal{W} * \mathcal{M} + (1 - \mathcal{W}) * \hat{\mathcal{M}}$ where $\hat{\mathcal{M}}$ is the current estimate of the reconstruction. As $\mathcal{W}$ is binary, the imputation reduces to expectation maximization [11].

Atlas-based Motion Correction   Subtle Brownian motion of organelles across image channels in a multispectral image must be corrected such that the organelles are stationary across the spectral sequence. Particle tracking methods can track individual particles across multiple images and obtain displacement fields, yet they are inapplicable for our registration-based correction as the fields need not be invertible (thus tearing the image) and assume constant intensity or require pre-segmentation. Further, a spectral channel must be arbitrarily chosen as the reference image towards registration, thus inducing user bias.

We propose to use a large deformation diffeomorphic metric mapping-based atlas building framework towards this correction. Unbiased atlas building [7] is used to generate a deformable template image by minimizing its distance to every channel in the spectral image. Symmetric diffeomorphic registrations and atlas estimation are performed using cross-correlation as a metric [1]. This approach has the following advantages: (1) the diffeomorphic constraint ensures
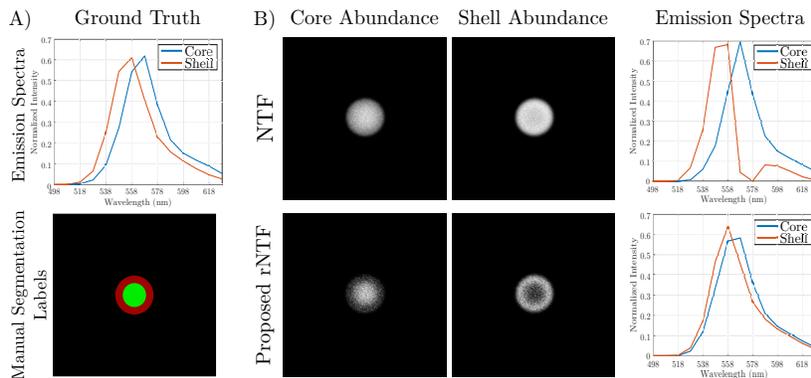
Fig. 1. rNTF validation. A bead with two concentric fluorescent chemicals with similar spectra imaged in a confocal microscope with two excitation wavelengths. Bovine hemoglobin was added to simulate attenuation in tissue imaging, thus creating non-linearities. As shown in **B**, the spectrum of the shell recovered by NTF is significantly distorted, and NTF is unable to spatially resolve the two substances due to the extrinsically added attenuation. rNTF recovers spectra which are undistorted, while also being able to spatially separate the two fluorophores on the bead as shown in the abundances images.

invertible deformation fields and prevents tearing, (2) this does not require constant intensity due to the use of cross-correlation as a matching metric and (3) it removes user bias in picking a registration target.

Paired Hypothesis Testing for Spectra   The differences in retrieved spectra from each fixed/unfixed pair are difficult to interpret and necessitate a hypothesis testing framework. Here, our features are the spectral channels. Paired multivariate tests should not be directly used as they do not account for the sequential nature of these features and the number of features are comparable to the sample size. We start by noticing that spectra are discrete realizations of continuous curves (i.e. functional data). Inspired by [9], we use functional PCA to reduce dimensionality while accounting for the sequential nature of features and further extend their work to the case of paired samples.

Once we retrieve constituent spectra from all the tissue sections, corresponding length-m spectra from n tissue sections (i.e. the same spectral component in each donor identified by spatial localization) are stacked into a matrix of size $m \times n$. Applying functional PCA with rank-k, we get k functional eigenbases and their coefficients ξ. We wish to test whether the difference in distribution of these coefficients for paired samples (fixed/unfixed) are significantly different. We state our k null hypotheses as $H_0^k : [\xi_j^1 - \xi_j^2]_{j=1}^k$ comes from a symmetric distribution with zero median, where the coefficient superscript indicates group membership. Rejecting this null hypothesis implies significant differences between fixed and unfixed pairs. All k hypotheses are tested with the Wilcoxon signed-rank test
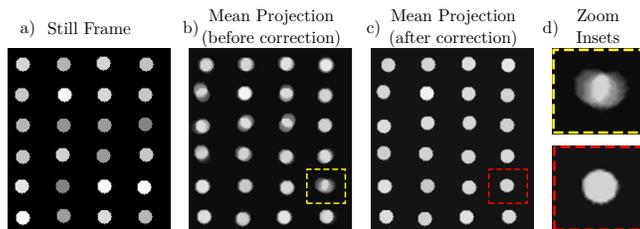
Fig. 2. Toy Example for Motion Correction. A grid of circles with random intensity changes and conditional deformations and are used to simulate the changing intensities and motion over image channels, respectively. With Brownian motion, the framework corrects the motion with some minor artefacts. **a)** A still frame from a sequence of images of circles on a randomly deforming grid. **b)** Mean projection of the raw image sequence. Note the blurry edges due to movement. **c)** Mean projection of the image sequence after correction. The circles are not returned to their initial positions due to our assumption of zero-mean motion. However, they are stationary which is sufficient for our application. **d)** Insets showing a sample circle with (bottom) and without (top) motion correction. Readers are encouraged to view supplementary material for videos of both synthetic and real examples.

which is a non-parametric test for paired samples. As there are k hypotheses, we apply the Bonferroni multiple comparisons correction. If we apply rank-r rNTF (i.e., r spectra from each section), we have r such matrices and repeat this for each matrix.

3   Experiments and Results

Experimental Dataset   24 pairs of tissue sections are used here. Tissue sections from corresponding locations in human donors from both eyes (one fixed, one not) were imaged with a multispectral fluorescence microscope, capturing multiple channels per image. Four excitation light sources were used, thus acquiring four multi-channel images per tissue section. After atlas-based motion correction, the images are preprocessed similarly to [2]. Each channel in an image is vectorized and treated as a row of a matrix. Repeating this across the four images and stacking the matrices, we get a 3D tensor (channels × pixels × light sources). This is illustrated in fig. **3A**, **B**, **C**, **D** and **E**. The multispectral images each have a different number of channels and therefore, when the images are stacked into a tensor, unavailable channels are treated as missing data.

Implementation Details   The four multispectral images are first affinely aligned. To build an image-specific atlas, 12 iterations of atlas building are computed with `ANTs` [1] for each channel in a stack with 80 iterations of diffeomorphic registration per iteration. Another atlas building step is done to create an atlas of atlases for each tissue section, to which each of the original images is mapped. Linear interpolation is used so as to not create values outside data range.
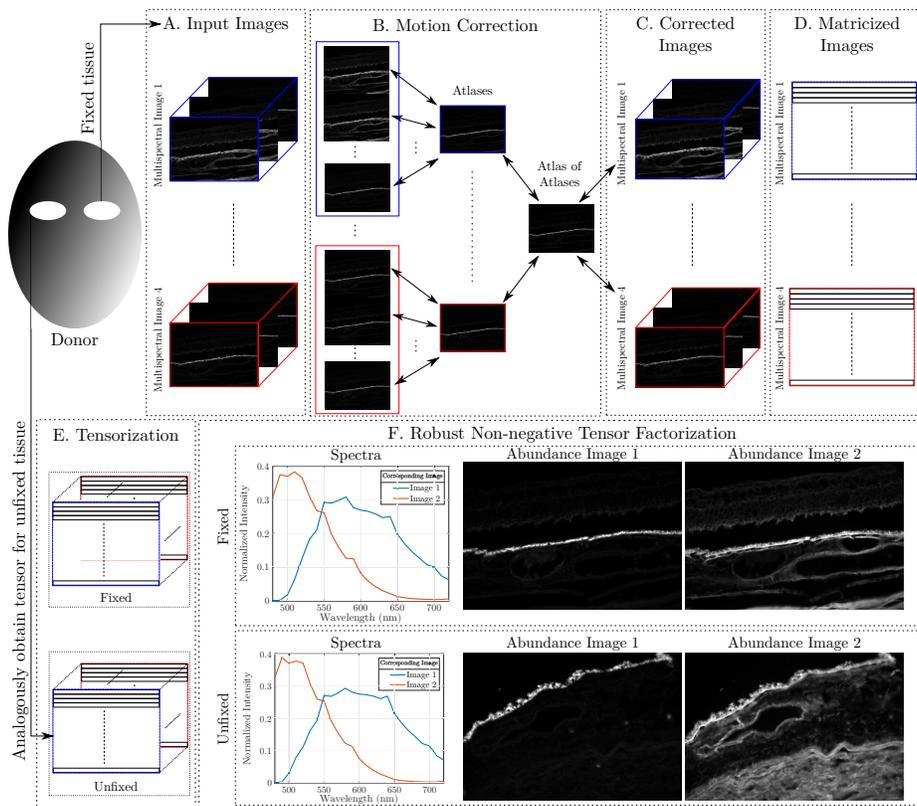
Fig. 3. An overview of the proposed pipeline and sample results. Panels **A**, **B**, **C** and **D** are only shown for the fixed tissue for brevity, with the unfixed tissue undergoing the same procedure. **Pipeline:** Fixed and unfixed images are obtained from the same donor. The images are motion corrected (**A**, **B** and **C** show the process for two out of four total images outlined in blue and red for each fixed tissue section). The images are then flattened into a matrix (**D**) and stacked into a tensor for each set (**E**). The two tensors undergo rNTF to reveal constituent spectra and spatial distributions (sample results shown in **F**). The spectra from all 24 pairs are used for functional statistics. See supplementary material for an expanded view of sample results.

We implement rNTF on GPUs for fast execution using `PyTorch` at fp64 precision. A rank-2 tensor factorization model is chosen using the CORCONDIA heuristic [11] and all factors were initialized uniformly at random. Parameters $\beta = 1.6$ (corresponding to mixed Poisson-Gaussian noise) and $\lambda = 2.5$ were chosen heuristically and a tolerance of 10^{-6} was used for all of the optimization. Sample tensor factorization results for a fixed/unfixed pair are shown in fig. **3F**.

Hypothesis Testing   Once constituent spectra are retrieved from all 24 pairs of tissues, a matrix is created with the spectrum from the blue curve in fig. **3F**

(lipofuscin, in retinal biology terminology) from each tissue section. All spectra are normalized to unit 2-norm to only consider shape changes. Functional PCA is then performed with model selection using the Bayesian Information Criterion. Two eigenbases are retrieved, with two sets of coefficients. The pairs of coefficients are given to the Wilcoxon signed-rank test for testing. After Bonferroni correction for two comparisons, we find no statistically significant differences between the pairs. The procedure is repeated for the red curves in fig. **3F** from each donor and finds no statistically significant difference between them, either.

4   Discussion

To our application, we provide image and statistical analysis methodologies and find no significant differences between the fluorescence spectra of fixed and unfixed tissue. This finding informs applications in multispectral retinal microscopy. Further validation is required as two limitations exist: (1) in this specific application, rNTF results are not guaranteed to be unique due to the additional need for data imputation; (2) the motion correction cannot retrieve the original positions and instead moves structures to the nonlinear mean of their movement.

The methods presented are general and amenable to several applications. rNTF is suitable for applications in machine learning to handle grossly corrupted measurements (outliers and nonlinearities), making only mild assumptions on the outliers being sparse. The motion correction framework can be applied to any spectral image displaying nonlinear deformations between channels. Finally, when samples are paired functional observations, we develop a non-parametric hypothesis testing framework. E.g., this statistical framework can be applied to a longitudinal analysis of fiber tracts in diffusion MRI by registering fiber tracts from the same subjects across a baseline and followup visit to test for differences in fractional anisotropy.

References

1. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis **12**(1), 26–41 (2008)
2. Dey, N., Hong, S., Ach, T., Koutalos, Y., Curcio, C.A., Smith, R.T., Gerig, G.: Tensor decomposition of hyperspectral images to study autofluorescence in age-related macular degeneration. Medical Image Analysis **56**, 96–109 (2019)
3. Févotte, C., Dobigeon, N.: Nonlinear hyperspectral unmixing with robust non-negative matrix factorization. IEEE Transactions on Image Processing **24**(12), 4810–4819 (2015)
4. Hong, D., Kolda, T.G., Duersch, J.A.: Generalized canonical polyadic tensor decomposition. arXiv preprint arXiv:1808.07452 (2018)

5. Huang, Z., McWilliams, A., Lam, S., English, J., McLean, D.I., Lui, H., Zeng, H.: Effect of formalin fixation on the near-infrared raman spectroscopy of normal and cancerous human bronchial tissues. International journal of oncology **23**(3), 649–655 (2003)
6. Ikoma, H., Heshmat, B., Wetzstein, G., Raskar, R.: Attenuation-corrected fluorescence spectra unmixing for spectroscopy and microscopy. Optics express **22**(16), 19469–19483 (2014)
7. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage **23**, S151–S160 (2004)
8. Neher, R.A., Mitkovski, M., Kirchhoff, F., Neher, E., Theis, F.J., Zeug, A.: Blind source separation techniques for the decomposition of multiply labeled fluorescence images. Biophysical journal **96**(9), 3791–3800 (2009)
9. Pomann, G.M., Staicu, A.M., Ghosh, S.: A two-sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis. Journal of the Royal Statistical Society: Series C (Applied Statistics) **65**(3), 395–414 (2016)
10. Ramsay, J.O.: Functional data analysis. Springer, New York (2005)
11. Smilde, A., Bro, R., Geladi, P.: Multi-way analysis: applications in the chemical sciences. John Wiley & Sons (2005)
12. Waters, Jennifer C.: Accuracy and precision in quantitative fluorescence microscopy. Rockefeller University Press (2009)
13. Zhou, P., Feng, J.: Outlier-robust tensor PCA. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2263–2271 (2017)

A   Overall Source Separation Algorithm

Algorithm 1: Robust NTF with data imputation and rank estimation.

    Input: $\mathcal{M} \in \mathbb{R}^{I \times J \times K}, \beta, \lambda$.

    Output: $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathcal{S}$.

1 For imputation, set indicator tensor $\mathcal{W} \in \mathbb{R}^{I \times J \times K}$ with $w_{ijk} = 0$ if m_{ijk} is missing and $w_{ijk} = 1$ if m_{ijk} is available.

2 Initialize rNTF rank as $P = 1$.

3 **repeat**

4   Initialize factor matrices $\mathbf{A} \in \mathbb{R}^{I \times P}, \mathbf{B} \in \mathbb{R}^{J \times P}, \mathbf{C} \in \mathbb{R}^{K \times P}$, and outlier tensor $\mathcal{S} \in \mathbb{R}^{I \times J \times K}$, all uniformly at random.

5   **repeat**

6   Impute $\tilde{\mathcal{M}} \to \mathcal{W} * \mathcal{M} + (1 - \mathcal{W}) * \hat{\mathcal{M}}$, where $\hat{\mathcal{M}} = (\sum_{i=1}^{P} \mathbf{a}_i \circ \mathbf{b}_i \circ \mathbf{c}_i + \mathcal{S})$.
    `// Block 1: use mode-1 matricization.`

7   Fix $\mathbf{B}$, $\mathbf{C}$, and $\mathcal{S}$, and update $\mathbf{A} \to \mathbf{A} * \left(\frac{(\tilde{\mathbf{M}}_{(1)} * \hat{\mathbf{M}}_{(1)}^{\cdot(\beta-2)})(\mathbf{C} \odot \mathbf{B})}{\hat{\mathbf{M}}_{(1)}^{\cdot(\beta-1)}(\mathbf{C} \odot \mathbf{B})} \right)$.

8   Update $\hat{\mathbf{M}}_{(1)} \to \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{S}_{(1)}$.

9   Fix $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, and update
$$\mathbf{S}_{(1)} \to \mathbf{S}_{(1)} * \left(\frac{\tilde{\mathbf{M}}_{(1)} * \hat{\mathbf{M}}_{(1)}^{\cdot(\beta-2)}}{\hat{\mathbf{M}}_{(1)}^{\cdot(\beta-1)} + \lambda \mathbf{S}_{(1)} \text{diag}(\|\mathbf{s}_1\|_2, ..., \|\mathbf{s}_{JK}\|_2)^{-1}} \right).$$

10   Update $\hat{\mathbf{M}}_{(1)} \to \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{S}_{(1)}$.
    `// Block 2: use mode-2 matricization.`

11   Fix $\mathbf{A}$, $\mathbf{C}$, and $\mathcal{S}$, and update $\mathbf{B} \to \mathbf{B} * \left(\frac{(\tilde{\mathbf{M}}_{(2)} * \hat{\mathbf{M}}_{(2)}^{\cdot(\beta-2)})(\mathbf{C} \odot \mathbf{A})}{\hat{\mathbf{M}}_{(2)}^{\cdot(\beta-1)}(\mathbf{C} \odot \mathbf{A})} \right)$.

12   Update $\hat{\mathbf{M}}_{(2)} \to \mathbf{B}(\mathbf{C} \odot \mathbf{A})^T + \mathbf{S}_{(2)}$.

13   Fix $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, and update
$$\mathbf{S}_{(2)} \to \mathbf{S}_{(2)} * \left(\frac{\tilde{\mathbf{M}}_{(2)} * \hat{\mathbf{M}}_{(2)}^{\cdot(\beta-2)}}{\hat{\mathbf{M}}_{(2)}^{\cdot(\beta-1)} + \lambda \mathbf{S}_{(2)} \text{diag}(\|\mathbf{s}_1\|_2, ..., \|\mathbf{s}_{KI}\|_2)^{-1}} \right).$$

14   Update $\hat{\mathbf{M}}_{(2)} \to \mathbf{B}(\mathbf{C} \odot \mathbf{A})^T + \mathbf{S}_{(2)}$.
    `// Block 3: use mode-3 matricization.`

15   Fix $\mathbf{A}$, $\mathbf{B}$, and $\mathcal{S}$, and update $\mathbf{C} \to \mathbf{C} * \left(\frac{(\tilde{\mathbf{M}}_{(3)} * \hat{\mathbf{M}}_{(3)}^{\cdot(\beta-2)})(\mathbf{B} \odot \mathbf{A})}{\hat{\mathbf{M}}_{(3)}^{\cdot(\beta-1)}(\mathbf{B} \odot \mathbf{A})} \right)$.

16   Update $\hat{\mathbf{M}}_{(3)} \to \mathbf{C}(\mathbf{B} \odot \mathbf{A})^T + \mathbf{S}_{(3)}$.

17   Fix $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, and update
$$\mathbf{S}_{(3)} \to \mathbf{S}_{(3)} * \left(\frac{\tilde{\mathbf{M}}_{(3)} * \hat{\mathbf{M}}_{(3)}^{\cdot(\beta-2)}}{\hat{\mathbf{M}}_{(3)}^{\cdot(\beta-1)} + \lambda \mathbf{S}_{(3)} \text{diag}(\|\mathbf{s}_1\|_2, ..., \|\mathbf{s}_{IJ}\|_2)^{-1}} \right).$$

18   Update $\hat{\mathbf{M}}_{(3)} \to \mathbf{C}(\mathbf{B} \odot \mathbf{A})^T + \mathbf{S}_{(3)}$.

19   **until** *convergence.*

20   **return** $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathcal{S}$.

21   Compute CORCONDIA between $\mathcal{M}$ and $(\sum_{i=1}^{P} \mathbf{a}_i \circ \mathbf{b}_i \circ \mathbf{c}_i + \mathcal{S})$.

22 $P = P + 1$.

23 **until** *CORCONDIA decrease exceeds user-defined threshold.*

24 **return** $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathcal{S}$.

B Scientific Packages and Parameters Used

Here, we briefly list heavily used packages and any important parameter values.

B.1 Robust Non-negative Tensor Factorization

Robust non-negative tensor factorization was computed with `PyTorch 1.0.0` in `python`. GPU-accelerated source code is available at `https://github.com/neel-dey/robustNTF`. $\beta = 1.6$ was chosen heuristically corresponding to a mixture of Poisson and Gaussian noise. $\lambda = 2.5$ was chosen to induce sufficient outlier-sparsity.

The comparison with standard non-negative tensor factorization with data imputation was done using the `N-way Toolbox` for `MATLAB`, available at `http://www.models.life.ku.dk/nwaytoolbox`.

B.2 Diffeomorphic Motion Correction

Initial inter-image stack affine registration was performed with the `SimpleITK` `python` library using normalized cross-correlation for image-matching. All registrations for intra-image stack motion correction were performed at fp32 with the `ANTs` toolbox, available at `http://stnava.github.io/ANTs/`. The greedy symmetric normalization algorithm is used with cross-correlation for image matching with default pixel radius. To avoid losing subtle motion, no smoothing or multi-scale strategies were used. 12 iterations of atlas building, with 80 inner registration iterations were used. Qualitatively, the algorithm was found to correct motion with fewer iterations (8 and 60, respectively). Another 12 iterations of atlas building were performed to construct the atlas of atlases for each tissue section.

B.3 Functional Statistics

Functional Principal Component Analysis was performed with the `PACE` toolbox for `MATLAB`, available at `http://www.stat.ucdavis.edu/PACE/`. Default toolbox settings were used for these experiments.
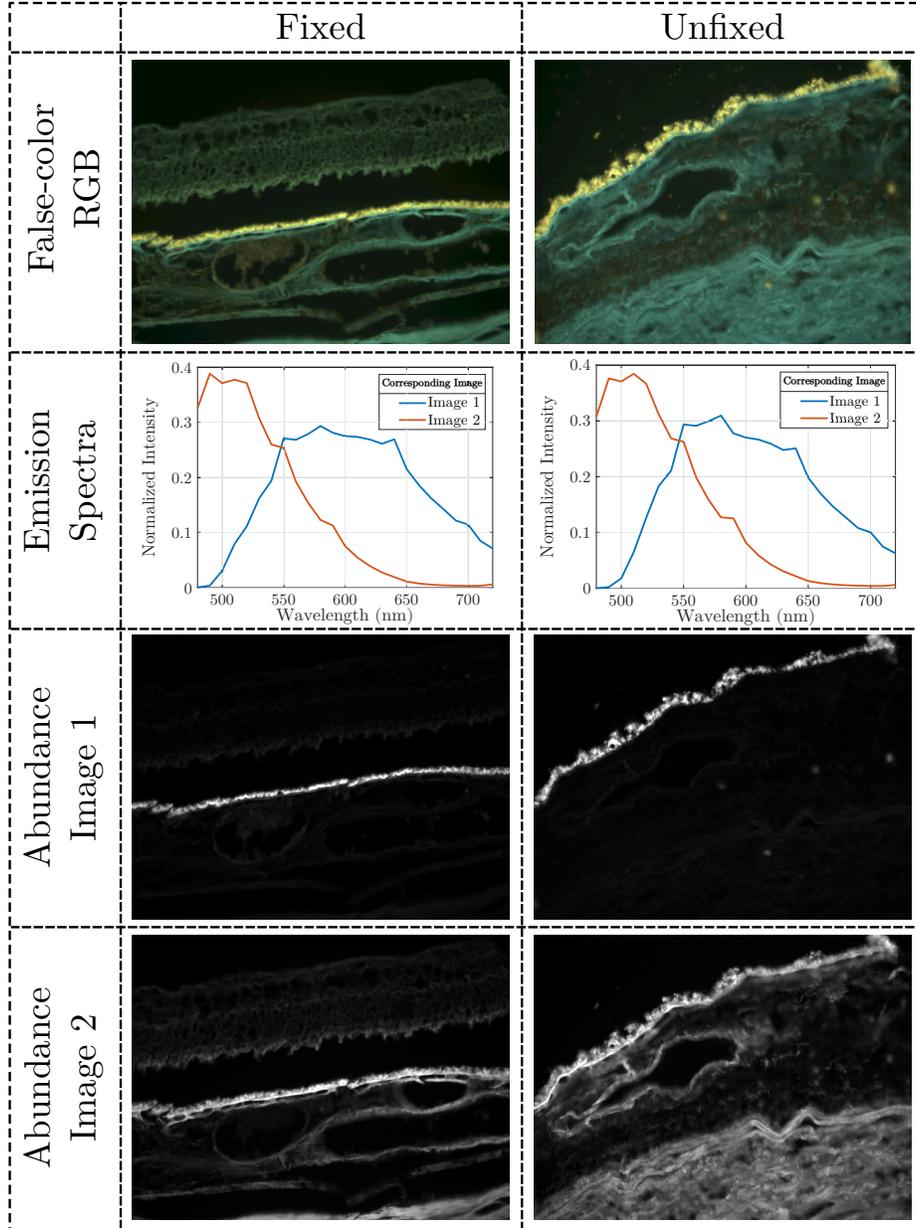
C Larger Results Panel

Fig. 4. A larger version of the sample results for a fixed and unfixed pair as shown in Figure 3 of the main text. **Row 1:** A false-color composite of the spectral images excited at 420 nm. **Rows 2, 3, and 4:** Emission spectra and abundance images retrieved by robust non-negative tensor factorization.