Blind source separation techniques for the decomposition of multiply labeled fluorescence images

Richard A. Neher*, Mišo Mitkovski†‡, Frank Kirchhoff‡††, Erwin Neher‡§, Fabian J. Theis¶||, and André Zeug**††‡‡

* Kavli Institute for Theoretical Physics, University of California, address: Santa Barbara, CA, 93106
† Department of Molecular Biology of Neuronal Signals, Max Planck Institute of Experimental Medicine Göttingen, address: Hermann Rein Strasse 3, 37075 Göttingen, Germany
‡ Glial Physiology and Imaging, Neurogenetics, Max Planck Institute of Experimental Medicine Göttingen, address: Hermann Rein Strasse 3, 37075 Göttingen, Germany
§ Department Membranbiophysik, Max Planck Institute for Biophysical Chemistry Göttingen, address: Am Fassberg 11, 37077 Göttingen, Germany
¶ Institute for Bioinformatics and Systems Biology, Helmholtz Zentrum München, address: Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany
|| Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany, address: Bunsenstraße 10, 37073 Göttingen, Germany
** Department of Neurophysiology and Cellular Biophysics, University Göttingen, Germany
†† Deutsche Forschungsgemeinschaft-Research Center for the Molecular Physiology of the Brain (FZT 103), Göttingen, Germany
‡‡ present address: Department of Cellular Neurophysiology, Hannover Medical School, Germany, address: Carl-Neuberg-Str. 1, 30625 Hannover, Germany

Correspondence should be addressed to André Zeug (zeug.andre@mh-hannover.de).

Running title:

Blind unmixing of fluorescence images

Key words:

Spectral fingerprinting; image analysis, multispectral imaging,

NMF, PARAFAC, cytoskeleton, GFP
Abstract
Methods of blind source separation are used in many contexts to separate composite data sets into their sources. Multiply labeled fluorescence microscopy images represent such sets, in which the sources are the individual labels. Their distributions are the quantities of interest and have to be extracted from the images. This is often challenging since the recorded emission spectra of fluorescent dyes are environment and instrument specific.
We developed a non-negative matrix factorization (NMF) algorithm to detect and separate spectrally distinct components of multiply labeled fluorescence images. It operates on spectrally resolved images and delivers both the emission spectra of the identified components and images of their abundances. We tested the proposed method using biological samples labeled with up to 4 spectrally overlapping fluorescent labels. In most cases, NMF accurately decomposed the images into the contributions of individual dyes. However, the solutions are not unique, when spectra overlap strongly or else when images are diffuse in their structure. To arrive at satisfactory results in such cases, we extended NMF to incorporate preexisting qualitative knowledge about spectra and label distributions. We show how data acquired through excitations at two or three different wavelengths can be integrated and that multiple excitations greatly facilitate the decomposition.
By allowing reliable decomposition in cases, where the spectra of the individual labels are not or only inaccurately known, the proposed algorithms greatly extend the range of questions that can be addressed with quantitative microscopy.
Introduction

Multiple fluorescent labeling has become a key tool for the elucidation of signaling networks in cells and tissues (1,2). To understand a system’s properties it is essential to label and monitor simultaneously as many components as possible. Modern laser scanning and wide field microscopes allow rapid acquisition of spectrally resolved images, from which the separate contributions of simultaneously present labels can be obtained.

Traditionally, this task is solved by choosing narrow emission bands, where only one dye contributes significantly. This approach, however, discards the majority of the photons and severely limits the choice of available dyes to those with well-separated emissions. If the emission spectra of the dyes are known, these drawbacks can be overcome by linear unmixing or spectral fingerprinting (3). However, the relevant emission spectra of the fluorophores depend on the instrumentation and the chemical environment, and therefore their acquisition requires extensive calibration efforts.

One way to overcome these difficulties is to use fluorescence lifetime information or modulated excitation schemes (4-6 FLIM). On the other hand, several attempts have been made to use blind source separation (BSS) techniques that estimate spectra and concentrations simultaneously (7-10). However, the decompositions are often ambiguous. Furthermore, most algorithms do not account for the noise characteristics of fluorescence data and thereby put undue emphasis on some parts of the data. Here, we present an algorithm adapted to fluorescence microscopy and test it on a various samples to systematically investigate its reliability.

We discuss techniques to incorporate additional qualitative knowledge about spectra and spatial features of images for reducing the ambiguity in the decomposition. We conclude with a method to integrate data acquired at different excitation wavelengths, which further facilitates the decomposition (8,11). Our benchmark examples include the separation of the most commonly used fluorescent proteins (ECFP, EGFP and EYFP) and 3 or 4 subcellular structures labeled with common Alexa Fluor dyes, FITC and ethidium bromide (EtBr).

Non-negative matrix factorization for fluorescence microscopy

Modern laser scanning and wide field microscopes allow for rapid acquisition of the fluorescence emission in several spectral channels at each pixel of an image. In the absence of non-linear effects the recorded signal \( y_{ij} \) at pixel \( j \) in channel \( i \) is the sum of the contributions of the different labels. The contribution of label \( k \) is proportional to its concentration \( x_{kj} \) at pixel \( j \) and the fraction \( a_{ik} \) of its emission that falls into channel \( i \). This is summarized in the equation

\[
y_{ij} = \sum_{k=1}^{M} a_{ik} x_{kj},
\]

where the sum extends over all labels \( k=1..M \). Viewing \( y_{ij}, x_{kj}, \) and \( a_{ik} \) as matrices \( Y, X, \) and \( A \), Eq. 1 can be written, apart from noise, as \( Y=AX \). Typically, a researcher is interested in the concentration distributions \( X \) of the labels. If the spectra \( A \) are known, then \( X \) can be calculated from \( Y \) by ‘linear unmixing’(3,12). If \( A \) is not known, \( Y \) can be factorized into a pair of \( A^* \) and \( X^* \) in many different ways and additional assumptions have to be made to arrive at a unique solution. We will show that the trivial constraint, whereby all concentration and spectra have to be non-negative, suffices in many cases to achieve a reliable decomposition of the image. Such a factorization into non-negative \( A^* \) and \( X^* \) is efficiently achieved by an algorithm known as non-negative matrix factorization (NMF) (13,14).
decomposes \( \mathbf{Y} \) by an iterative minimization of a cost function, which reflects the deviation between the measured intensities and those predicted by the matrix product. The condition of non-negativity is imposed by choosing non-negative starting values and by choosing the update rules, such that no zero-crossings can occur. A detailed discussion of the estimation of matrices \( \mathbf{A}^* \) and \( \mathbf{X}^* \) from shot noise dominated microscopy data and a derivation of suitable update rules is presented in the appendix.

**Materials and Methods**

**Implementation and data processing**

All algorithms were implemented as Matlab functions (Mathworks, Natick, MA, USA). The image data were pre-processed by subtracting the constant background signal, which was measured from a dark region of the image or with the laser shut off. Subsequently, all pixels below a background threshold (typically 100 counts) and above a saturation threshold of 4000 counts (4096 is the maximal range of the AD) were excluded from the analysis. The iterative algorithm was initialized by Gaussian spectra peaked at the wavelength of the reference spectra (Molecular Probes). The concentrations were initialized randomly and adjusted to the start spectra by 10 rounds of concentration updates only. The algorithm was run for 1000 iterations, which took about 2 min for a typical data set in the case of NMF. Eventually, the concentrations at the excluded pixels were calculated by non-negative linear unmixing using the estimated spectra and included into the image. A more complete account of our experience with different initial conditions, dependence on signal-to-noise ratio, and possible pitfalls is given in the supplementary material.

**Summary of sample preparation and Microscopy**

Reagents were obtained from Sigma-Aldrich unless stated otherwise. Stress-fiber formation was facilitated by growing HeLa SS6 (kindly provided by Prof. Lührmann, Max Planck Inst. for Biophysical Chemistry, Göttingen, Germany) and NIH-3T3 cells (DSMZ, Braunschweig, Germany) on fibronectin-coated coverslips. In the appropriate cases, cells were incubated with A555-conjugated transferrin (Invitrogen). The ensuing acrolein-paraformaldehyde fixation and Triton X-100 permeabilization allowed for simultaneous tubulin and F-actin (Sigma-Aldrich and Invitrogen) stains. Finally, cells were mounted in EtBr-containing medium after RNase treatment. Heterozygous, triple transgenic mice were generated by interbreeding the following mouse lines: TgN(Thy1-ECFP) (15), TgH(CX3CR1-EGFP) (16) and TgN(GFAP-EYFP) (15). These mice are characterized by selective expression of EGFP, EYFP and ECFP in microglia, subpopulations of neurons and astroglia, respectively. Bright fluorescence can be detected in microglia, Bergmann glia and mossy fiber terminals in the cerebellum. For imaging, transgenic mice (four-week-old) were anaesthetized and perfused transcardially with Hank’s Balanced Salt Solution (HBSS, Gibco) followed by perfusion with 4 % paraformaldehyde in phosphate-buffered saline for 15 min. The brain was removed and incubated in PFA over night at 4 °C. After washing twice in PBS, the cerebellum was dissected and 50 µm sagittal vibratome sections (Leica VT 1000S, Leica Instruments, Nussloch, Germany) were prepared and mounted with Immu-Mount (Shandon). Images were acquired with an Axiovert 200M equipped with LSM510-Meta confocal microscope (Carl Zeiss Jena, Germany) using a 63×/1.2 NA water-immersion objective. The 458, 477 and 488 nm line of a 40 mW argon laser was used at 100, 50 and 10 % power, respectively. The HFT458 and HFT488 dicroic mirrors were used for the mouse tissue sections and cultured cell samples, respectively. Channel settings
are different for the specific samples and are mentioned in the text and the appropriate figure captions, respectively. Reasonable detector gains were used between 550 and 650 a.u. in the Zeiss AIM software. The raw data provided by the microscope software was used in the NMF algorithm as described above.

Results
Single exposure measurements
We first tested the NMF algorithm on an image stack generated from adherent NIH-3T3 fibroblast cells. Herein, double-stranded nucleic acids (mainly ribosomal RNA and nuclear DNA) were labeled with EtBr (17,18); filamentous actin (F-actin) was stained with Alexa Fluor 532 (A532) conjugated phalloidin, while tubulin was labeled with an Alexa Fluor 488 (A488) linked antibody. The sample was imaged with a Zeiss LSM Meta 510 using 8 evenly spaced spectral emission channels from 508 to 657 nm (width 21.4nm) and excited with a 488 nm laser. Figure 1 shows the measured images in the 8 spectral channels on the left and the estimated concentrations \( \mathbf{X}^* \) on the right. The label distributions \( \mathbf{X}^* \) of the three dyes were consistent with the known morphology of the sample. Closer inspection, however, shows that a faint replica of the tubulin structure is superimposed onto the image of F-actin stain and similarly for the F-actin stain in the image of the DNA stain. This is a consequence of small deviations of the estimated spectra from the reference spectra, which we measured independently on singly stained samples (see Figure 1, lower left). The spectrum of A488, for example, is too narrow. To compensate for this deviation, the algorithm assigns about 10% of the photons originating from A488 to the F-actin image. The sum of 90% estimated A488 and 10% A532 yields precisely the true A488 spectrum.

Such problems in the decomposition are expected on theoretical grounds. They are due to the fact that any linear combination \( \mathbf{A}^* = \mathbf{A} \mathbf{B} \) of the true spectra \( \mathbf{A} \), such that \( \mathbf{A}^* \) and \( \mathbf{X}^* = \mathbf{B}^{-1} \mathbf{X} \) are non-negative, is a valid solution of the matrix Eq. 1 (\( \mathbf{B} \) is a full rank M-by-M matrix). The cost functions for \( \mathbf{X} \) and \( \mathbf{X}^* \) are identical and therefore the minimization of the cost function may result in any of such eligible linear combinations. This ambiguity is the basic problem of NMF. The main objective of our study is to explore its extent and to identify procedures, how to minimize its consequences. To illustrate this issue in more detail, we consider a graphical representation of the fitting results. The high-dimensional vectors \( \mathbf{y}_j \) representing the measured intensities at pixel \( j \) lie, up to fluctuations, in the three-dimensional subspace spanned by the three spectra of the dyes. In other words, \( \mathbf{y}_j \) can be described by three scalars corresponding to the abundances of the labels, rather than by the intensities in each spectral channel. This subspace can be further reduced to two dimensions if absolute values are not relevant, as is the case for the ambiguity discussion. We therefore normalized all data points and spectra. A projection of this 2D space is shown in Figure 1, bottom right (see Supplementary Materials for details). We will refer to this representation as the ‘simplex’ below. The circles corresponding to the spectra returned by NMF are shown as red dots connected by red lines forming a triangle (the ‘NMF triangle’). As argued above, the NMF-spectra are a linear combination of the reference spectra. The latter therefore can be represented in the same 2D subspace and are shown in Figure 1, bottom right, in blue (the ‘reference triangle’). The density of the data points is represented in the diagram by color and contour lines. A pixel containing contributions from two of the three spectra will lie on the side of the reference triangle connecting the two dyes, while a pixel containing
all three dyes will be located in the bulk of the triangle. Points outside the triangle correspond to negative contributions of one or several spectra. This observation highlights one central constraint to the possible decompositions: For all concentrations to be non-negative, the NMF triangle has to include the cloud of data in the two dimensional representations (apart from some scatter due to noise fluctuations). In the examples we studied, vertices of NMF triangles (representing the estimated spectra) often were located outside of the reference triangle. NMF favors such decompositions, since they result in non-negative coefficients even for many data points, which lie outside the reference triangle. Spectra outside the reference triangle represent linear combinations of reference spectra with some negative contributions. They are narrower than the correct spectra and have reduced spectral overlap. Such spectra are allowed in NMF as long as all their individual values remain non-negative. This restriction sets bounds to the decomposition errors. It also defines conditions under which the decomposition is unique. These are readily appreciated in the case of two overlapping dyes: Subtracting fractions of one spectrum from another is only allowed, if the spectrum to be subtracted vanishes at all wavelengths, at which the other dye does not emit. This condition, termed condition (i) below, protects dyes emitting predominantly at long wavelengths (and not at short ones) against distortions by short wavelength dyes.

In the more general case of several dyes the domain of all non-negative spectra can be defined algebraically (see Supplementary Material). A mapping of the case of 3 dyes into the 2D diagram is shown by the dashed red polygon in Figure 1, bottom right. While these constraints strongly limit the set of possible solutions, it is obvious from the diagram, that significant freedom remains and many solutions are equivalent from the NMF perspective. The outcome, therefore, can depend on initial conditions. The run in Figure 1 was initialized with broad, heavily overlapping Gaussians centered at emission peaks of the respective dyes. Runs with narrow or random initial spectra are presented in the supplementary material.

If the data points do not fill the reference triangle, NMF can also return spectra that lie inside the reference triangle. Such spectra are positive linear combinations of the reference spectra and therefore too broad. Deviations of this type are not possible, if a sufficient number of data points lies on the boundary of the reference triangle. These data points represent pixels at which one dye is absent, while others are present. This condition will be termed condition (ii) below. Note that this condition is much less restrictive than the requirement of singly labeled regions, which has to be fulfilled for the traditional acquisition of reference spectra.

Removing the ambiguity by applying constraints

The ambiguity of the NMF-decomposition can be removed by adding constraints or a bias, which favors certain solutions (19). This is readily achieved during the iterative optimization procedure by adding a bias term to the cost function, derived from additional knowledge about spectra or label distributions. If it is known, for example, that labels are sufficiently segregated, i.e. condition (ii) is fulfilled, the correct set of spectra is the one with the smallest possible triangle, which is equivalent to maximally overlapping spectra. To exploit this knowledge about the label distribution, we modified the NMF algorithm such that it returns the smallest possible triangle automatically. This can be achieved either by maximizing spectral overlap directly, or by favoring decompositions with segregated label distributions (20). For maximal label segregation, the data points have to be as close as possible to the boundary of the triangle, resulting in a bias towards small triangles. The latter strategy
proved most robust, and we implemented this ‘segregation bias’ by adding the ratio of the 1-norm and the 2-norm of the concentration vector at each pixel to the cost function. The implementation of these biases is detailed in the appendix (Eqs. A6-A9).

As a test sample with segregated labels, we imaged vibratome sections of the cerebellum obtained from triple transgenic mice with cell-type specific VFP expression (Bergmann glia: ECFP, microglia: EGFP and neuronal mossy fibers: EYFP). Without the segregation bias, NMF estimated too narrow spectra, i.e. the NMF-triangle was larger than the reference-triangle, and the label distributions exhibited some crosstalk (Figure 2 top row). When we increased the relative weight of the segregation bias, the spectra changed gradually and approached the actual spectra of the individual labels. At the same time the three labels expressed in different cell types became perfectly separated (Figure 2, middle row). One typically finds, that spectra and concentrations change very little after \( \lambda \) exceeds a certain value, until eventually upon further increase of \( \lambda \) - the additional term overwhelms the primary requirement to describe the data accurately. This is consistent with the interpretation that a variety of permissible solutions have almost identical cost functions and a small bias is sufficient to favor one over the others. A large bias is necessary to achieve spectra that require negative concentrations. Within this intermediate regime, the results are fairly independent of the choice of \( \lambda \). While the segregation bias works well in many cases, it is inapplicable if label distributions are not sufficiently segregated and the data leaves large parts of the reference triangle empty, violating criterion (ii). In this case the bias will result in too small triangles and spectra that exhibit secondary peaks (see supplementary). However, even images with strongly overlapping label distribution often show sufficient signal modulation and fulfill criterion (ii) approximately. The sample in Figure 1 represents such a case. When we subjected it to a segregation bias the tubulin pattern in the F-actin stain disappeared and the estimated spectra approached the true spectra. However, with such a bias EtBr can develop a secondary peak at small wavelengths (see below and supplementary).

Furthermore, if colocalization versus segregation is at the basis of the scientific question to be addressed, a segregation bias is certainly not an appropriate method even of criterion (ii) is fulfilled. In this case a bias targeting the spectral overlap directly should be used.

**Including prior knowledge about spectra**

We have shown above that the ambiguity of NMF can be reduced by biases, based on qualitative knowledge about the distribution of labels (segregation). An alternative way of invoking prior knowledge is to determine the spectra of some of the components separately, estimating only the remaining spectra. This is especially valuable for dyes, which are faint and therefore hard to estimate. To explore this possibility, we labeled F-actin with FITC-conjugated phalloidin, tubulin with Alexa Fluor 514 (A514), allowed cells to import Alexa Fluor 555 (A555) labeled transferrin and stained for nucleic acids with EtBr in adherent HeLa SS6 cells. The transferrin stain was comparatively weak and we fixed its spectrum to that provided by Molecular Probes (www.invitrogen.com). With a slight segregation bias as described above, NMF estimated the other three spectra with good accuracy and delivered satisfactory concentration maps for all four dyes (Figure 3). In fact, the precise shape
of the spectrum of a faint dye is not important as long as it captures the peak. The remaining part of the emission is then assigned to other dyes, which does not make a big difference, if such dyes are strong.

**Post-NMF Data Processing**

The strategies to reduce the ambiguity in NMF discussed so far involved the selection of a suitable bias, followed by an unsupervised run of the decomposition algorithm. For three dyes, there is an alternative strategy: The true spectra are (unknown) points in the 2D representation of the decomposition by NMF, and one can attempt to identify the appropriate spectra interactively. To this end, we created a software tool. Following an initial NMF-run, the tool presents the user the 2D representation of the data density and the NMF-triangle (compare simplex projections in Figure 1 and 2). It also displays the NMF-spectra and the label distributions. The user can now explore the set of possible spectra by moving the mouse cursor within the domain of non-negative spectra. The tool calculates and displays the spectrum that corresponds to the position of the mouse cursor in real time. Once a satisfactory spectrum is found, the user can drag the corresponding vertex of the NMF-triangle to the new location. The software then rapidly recalculates the label abundances. This way, secondary spectral peaks that may emerge as the consequence of a segregation bias (see above) can readily be removed. It is our experience, that for 2 to 3 dyes it is straightforward to arrive at a unique solution, which has neither unusual features in the spectra nor crosstalk between images in the form of shadows of characteristic structures (see supplemental material for an example). The successful application of this tool critically depends on a reasonable starting decomposition, such that each of the labels dominates one of the decomposed images. NMF almost always delivers appropriate starting values.

**Results – multiple exposures measurements**

Commonly used dyes differ not only in their emission spectra but also in their excitation spectra. If the same sample is imaged with different excitation wavelengths, the relative strengths of the dyes will vary from excitation to excitation, while the spatial distributions and the emission spectra remain unchanged. These differences in excitation efficiencies contain very valuable information for decomposing the image. Furthermore, it is much easier to collect a sufficiently large number of photons from each dye, since excitation wavelengths can be chosen such that each dye is strongly excited at least once. To handle such three-dimensional data (excitation wavelength, emission channel, and image pixels), NMF has to be generalized to what is known as non-negative tensor factorization (NTF) (21) or parallel factor analysis (PARAFAC) (8,11). We derived update rules for NTF that account for the Poisson distribution of photon counts in fluorescence microscopy (see Appendix, eq. A12).

We applied NTF to an image of the quadruply labeled cells of the previous section. The samples were imaged using the excitation wavelengths 458 nm, 477 nm and 488 nm. The emission was recorded in 16 channels from 502 to 663 nm (width 10.7 nm). NTF, initialized with Gaussian spectra, estimated all four spectra and label distributions correctly, although the excitation efficiencies of the dyes chosen do not differ greatly (Figure 4).

Another example with substantial variation in excitation efficiencies is provided by the brain slice expressing ECFP, EGFP and EYFP (see above). We used the same excitation wavelengths as above and recorded emission in eight spectral channels ranging from 470 to 545 nm. Channels of wavelengths shorter than the respective
excitation wavelengths, were excluded from the analysis (NTF seamlessly integrates overlapping spectral ranges in different excitations, see Appendix). The algorithm reliably separated the raw data into three components that corresponded to ECFP, EYFP and EGFP, without invoking any additional constraints as are needed for single shot measurements.

Care has to be taken that the sample or the apparatus does not drift between successive illuminations. NTF will fail in such cases, unless images are brought into register before processing. NTF is also prone to get stuck in local minima. The latter, however, is rarely a problem since good initial guesses for spectra are usually available from the literature.

Discussion
The conventional technique in fluorescence microscopy of separating fluorescent labels, using optical filter cubes, limits the choice of fluorophores to those with well-separated spectral bands. Newer methods to overcome this limitation include multi-epitope-ligand cartography (22), methods using multiple excitations or fluorescence lifetime information (4-6) and methods of spectral fingerprinting (3,23). The latter method can be used on laser scanning and wide field microscopes which provide spectrally resolved data. Data sets from such microscopes typically consist of image stacks of the emissions at up to 32 different wavelengths. We investigated the potential of blind source separation (BSS) techniques to decompose such data into the contributions of the individual labels, when emission spectra are not or only approximately known.

Different BSS algorithms use different criteria to determine the sources. Principal component analysis for example decomposes the data into eigenvectors of the covariance matrix of the data, yielding orthogonal sources. Independent component analysis tries to find a representation of the data, where different sources are as statistically independent as possible. However, typical spectra are not orthogonal, nor are the label distributions independent. On the other hand, both spectra and label concentrations are strictly non-negative. This is why we suggest NMF and NTF as the methods of choice. Non-negativity is a mild constraint and little prejudice is implicit in the algorithm. The flipside, however, is that the non-negativity constraint provides a unique decomposition only if the conditions (i) and (ii) formulated above are fulfilled. In that case only one non-negative solution is possible for geometrical reasons. Condition (i) states that the each label must not emit in at least one spectral channel where the other labels do, while condition (ii) states that the image has to contain pixels where one dye is absent and others are present in various concentration ratios. This ensures that the boundary of the simplex formed by the data is well defined. Both conditions are much less restrictive than those required for the conventional techniques, where singly labeled ROIs or spectral channels with emissions of only one dye are necessary. In other words, the condition are relaxed from ‘all absent but one’ to ‘one absent at a time’. If only one of the conditions is violated, we nevertheless can retrieve a unique (and correct) solution by biasing the algorithms towards well-segregated label distributions (if condition (ii) is fulfilled) or else towards spectra with minimal or maximal overlap (if conditions (i) or (ii) are fulfilled, respectively). Even when these conditions are only approximately fulfilled, the algorithm yields satisfactory results. However, it has to be stated clearly that the algorithm is not applicable to samples where both conditions are grossly violated, i.e. where label distributions are similar and spectra overlap strongly. As fluorescence
microscopy data is often noisy, the two conditions are somewhat soft and confidence intervals for the estimated spectra will depend on the degree to which the conditions are fulfilled. The problem of ambiguous solutions can be overcome by using multiple excitations and NTF.

We also created a tool that allows one to interactively correct for errors in the decomposition provided by the NMF algorithm. The best way to use this tool is to obtain an NMF run with a mild segregation bias. This usually provides decompositions, in which the strongly represented labels are estimated quite accurately. Weakly represented labels may be contaminated by ‘ghost images’ of the strong ones, while their spectra may show secondary peaks. The shape of the spectra and possible crosstalk between images is then readily corrected by eliminating such obvious artifacts.

Alternatively, it is straightforward to fix the spectra of a subset of labels to predetermined ones. This is indicated for weak labels with broad spectra and also for handling autofluorescence. Such constrained optimization can also be used to test whether the spectrum of a dye deviates in a given region from a known spectrum. To this end, one spectrum can be fixed to the known spectrum in an NMF run with one additional free spectrum. If the sample contains regions, where the spectrum deviates from the reference spectrum, NMF will yield a new spectrum localized to those regions, e.g. organelles. In this sense NMF can be used as an analytical tool.

We have also shown that combining data from multiple excitations at different wavelengths greatly facilitates the decomposition. We anticipate, that a large number of labels can be separated, when patching together measurements each exciting a subset of the dyes. The full potential of NTF is still to be explored.
Acknowledgement

We thank Prof. Dr. Mary Osborn and Prof. Dr. Reinhard Lührmann (both MPI-BPC, Göttingen) for helpful comments and discussions and for providing the HeLaSS6 cell line, as well as Sabine Klöppner (MPI-EM, Göttingen) for technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft through the Research Center Molecular Physiology of the Brain (FZT 103 and EXC 171). Richard A. Neher acknowledges financial support by the National Science Foundation under grant no. NSF PHY05-51164. Fabian J. Theis acknowledges financial support by the Helmholtz Alliance on Systems Biology (project CoReNe).
Appendix

The light $y_{ij}$ recorded in a particular channel $i$ at a given pixel $j$ is a sum of the contributions of the labels present at the pixel. The contribution of dye $k$ is proportional to its concentration $x_{ij}$ at this pixel and to the contribution of its emission $a_{ik}$ that falls into the spectral range of channel $i$. Hence, we have

$$y_{ij} = \sum_{k=1}^{M} a_{ik} x_{ij}, \quad (A1)$$

where the sum runs over all dyes $k=1,...,M$. This model is conveniently written as the matrix equation $Y=AX$, which describes all pixels simultaneously. This equation, however, is not quite correct since it equates the actual signal $y_{ij}$ with the expected signal $\sum_{k} a_{ik} x_{ij}$. This distinction is necessary, since light emission from fluorophores is not a deterministic process but the number of detected photons is distributed according to a Poisson distribution with mean $AX$, i.e. the recorded signal $Y$ will scatter around $AX$. Our aim here is to estimate $A$ and $X$ from a noisy spectrally resolved image $Y$. To this end we determine matrices $A^*$ and $X^*$ that maximize the probability of measuring $Y$, assuming a Poisson distribution of the data, which implies minimizing the negative log-likelihood function

$$C(A^*,X^*|Y) = \sum_{i,j} \sum_{k} a_{ik} x_{ij} - y_{ij} \ln \left( \sum_{k} a_{ik} x_{ij} \right) + \text{const.}, \quad (A2)$$

where $i$ is the index of spectral channels, $k$ the index of labels, $j$ the index of pixels of the image, and const. represents a constant background (to be determined separately). To distinguish inferred quantities from the actual one, we mark them with an asterix ($A^*$ and $X^*$ vs. $A$ and $X$; for the lower case quantities this is omitted for clarity). Since spectra and concentrations are non-negative quantities, the minimization has to be restricted to purely non-negative values. This kind of minimization problem, where one matrix $Y$ is approximated by a product of two non-negative matrices $A^*$ and $X^*$ is known as non-negative matrix factorization (NMF). Such a minimization is efficiently performed by iterative algorithms with multiplicative update rules that preserve the sign of the matrix entries (13,14). Following closely the derivation given in (11), we derive multiplicative update rules for the cost function (A2). One begins by considering an ordinary gradient descent with step size $\eta_{rs}$

$$a_{rs} \leftarrow a_{rs} - \eta_{rs} \frac{\partial C}{\partial a_{rs}} = a_{rs} - \eta_{rs} \sum_{j} x_{sj} - \frac{y_{rj}}{\sum_{k} a_{rk} x_{kj}} x_{sj}, \quad (A3)$$

as shown here for $a_{rs}$ with similar rules for $x_{rs}$. The step size can now be chosen to be $\eta_{rs} = \frac{a_{rs}}{\sum_{j} x_{sj}}$, in which case the update rule becomes multiplicative and preserves non-negativity. The update rule for concentrations can be derived analogously and when alternating the two update steps, we arrive at

$$a_{rs} \leftarrow \frac{a_{rs}}{\sum_{j} x_{sj}} \sum_{j} \frac{y_{rj}}{\sum_{k} a_{rk} x_{kj}} x_{sj}$$

$$x_{rs} \leftarrow \frac{x_{rs}}{\sum_{i} a_{ir}} \sum_{i} \frac{y_{ri}}{\sum_{k} a_{ir} x_{ks}} a_{ir}. \quad (A4)$$
It can be shown that these update rules converge to a local minimum of the cost function using similar arguments as in (14).

In the main text we discuss that the factorization of $Y$ into $A^*$ and $X^*$ is not unique in many cases. For any invertible matrix $B$, an equally valid decomposition of the data is given by

$$Y = A^* X^* = ABB^{-1}X = AX,$$

provided $A^* = AB$ and $X^* = B^{-1}X$ have non-negative entries only. The range of permissible matrices $B$ depends on the spectral and spatial overlap of the sources. To overcome this ambiguity, we suggest the use of several biases that favor some solution to others.

**Segregation bias:** When the label distributions are highly modulated, such that all possible combinations of label concentrations occur, the correct solution is the one with maximally overlapping spectra and segregated labels. To bias the NMF algorithm towards such solutions, we add the following term $E$ to the cost function (A2)

$$E = -\lambda \sum_j \frac{\sum |y_{ij}|}{\sqrt{\sum x_{ij}^2}}.$$

The first sum extends over all pixels, the fraction is the ratio of the 1-norm and 2-norm of the concentration vector at pixel $j$ and $\lambda$ is the weight of the additional term. The ratio of 1-norm and 2-norm is one, if only one label is present at the respective pixel, while it is equal to $\sqrt{M}$ if $M$ labels are present with equal amounts. Hence the term is the smaller the better segregated the labels are. The prefactor $\lambda$ is used to adjust the importance of the bias relative to the original cost function. This additional term changes the update rules for the concentrations to

$$x_{rs} \leftarrow \frac{x_{rs}}{\sum_i a_{ir}} \left[ \sum_i y_{ir} a_{ir} - \lambda \left( \frac{1}{\left( \sum_k x_{ks}^2 \right)^{0.5}} - \frac{x_{sr} \sum_k x_{ks}}{\left( \sum_k x_{ks}^2 \right)^{1.5}} \right) \right].$$

**Biasing spectral overlap:** To control the overlap of the spectra, we propose to maximize or minimize the overlap between pairs of spectra in certain circumstances. This can be achieved by adding the term

$$F = \sum_{v<w} m_{vw} \sum_i a_{iv} a_{iw}$$

to the cost function. The matrix elements $m_{vw}$ specify the weight of the bias for each pair of dyes $v,w$, while the second sum over $i$ is simply the scalar product between the spectra of the dyes $v$ and $w$. With this addition, the update rule for the spectra changes to

$$a_{rs} \leftarrow \frac{a_{rs}}{\sum_j x_{sj}} \left[ \sum_j y_{rj} x_{sj} - \sum_v m_{vs} a_{rv} \right].$$
This shows, that during one update, a small fraction of the spectrum of one dye is
subtracted or added (depending on the sign of $m_{vw}$) from another dye.
The update rules including biases can lead to negative values. However, for
reasonably small biases this is rarely the case. If some concentrations or spectra do
become negative during the update, they should be set to small non-negative values.

Multiple excitations: The excitation efficiencies of most labels depend on the
wavelength of the excitation light. Hence, the different labels contribute with different
intensities, when the same sample is imaged at different wavelengths. This can be
incorporated into our data model by assigning an excitation efficiency $q_{kl}$ to dye $k$
at excitation wavelength $l$. The signal expected at the pixel $k$ in the emission channel $i$
and excitation wavelength $l$ is given by

$$y_{ijl} = \sum a_{ik} x_{kj} q_{kl}, \quad (A10)$$

where the sum extends over the labels in the sample. The cost function for a Poisson
distribution of light intensities is essentially unchanged and given by

$$C(A^*, X^*, Q^* | Y) = \sum_{i,j,l} \left[ \sum_k a_{ik} x_{kj} q_{kl} - y_{ijl} \ln \left( \sum_k a_{ik} x_{kj} q_{kl} \right) \right]. \quad (A11)$$

Methods to infer the three matrices $A^*$, $X^*$ and $Q^*$ from the three dimensional data $y_{ijl}$
are known as PARAFAC (11) or non-negative tensor factorization (21). While
PARAFAC often resorts to alternating least-square updates, NTF algorithms are a
direct generalization of NMF that naturally preserve positivity. For the above cost
function, we derived the following update rules:

$$a_{rs} \leftarrow \frac{a_{rs}}{\sum_{j,l} x_{sj} q_{sl}} \sum_{j,l} y_{rjl} x_{sj} q_{sl},$$

$$x_{rs} \leftarrow \frac{x_{rs}}{\sum_{i,l} a_{ir} q_{lr}} \sum_{i,l} y_{ijs} a_{ir} q_{lr}, \quad (A12)$$

$$q_{rs} \leftarrow \frac{q_{rs}}{\sum_{i,j} a_{ir} x_{sj}} \sum_{i,j} y_{ijs} a_{ir} x_{sj}.$$

The update rules can be derived in very much the same way as those described for
NMF above. If the spectral channels recorded differ for different excitations, the
summations on the right hand side of Eq. (A12) have to be restricted to the relevant
channels for each excitation.

It can be shown that the decomposition into $A^*$, $X^*$ and $Q^*$ is unique if the sources
differ sufficiently in their spectra $A$, their concentration distribution $X$ and their
excitation spectra $Q$. More specifically, the decomposition is unique if (11,24)

$$k(A) + k(X) + k(Q) \geq 2M + 2, \quad (A13)$$

where $k(A)$ is the k-rank of matrix $A$ and $M$ is the number dyes. The k-rank is the
maximal $k$, such that any combination of $k$ columns of $A$ has full rank.
Reference List


Figure captions

Figure 1: NMF estimation of spectra and label distributions of three spectrally and spatially overlapping labels; panel size 60x60 µm. **Left:** The emission in eight spectral channels from 508 to 657 nm of NIH-3T3 fibroblasts where tubulin is labeled with A488, F-actin with A532 and nucleic acids with EtBr. **Right:** The estimated concentrations maps for tubulin, F-actin and nucleic acids. Images are rescaled by the inverse maximum (relative scaling factors: 1 : 0.51 : 0.18 for A488 : A532 : EtBr). The F-actin image contains a faint replica of the tubulin structure, best visible in lower right nucleus region. Similarly, the nucleic acids stain was assigned a fraction of the emission of A532 (F-actin). The RGB false color representation of the individual concentration maps with blue: tubulin, green: F-actin, red: nucleic acids. **Bottom left:** Spectra plot from blue to red according to A488, A532 and EtBr. The estimated spectra (solid lines) are slightly narrower than the spectra measured in singly labeled specimen (dotted). **Bottom right:** The 8 dimensional data vectors can be projected into a simplex plane, see main text. The NMF run was initialized with Gaussian spectra with 524, 558 and 617 nm center position and a FWHM of 75 nm.

Figure 2: A segregation bias is necessary to find correct decomposition if spectra are overlapping strongly. A brain slice of a mouse expressing ECFP, EGFP and EYFP in different cell types is imaged in 8 spectral channels from 470 to 550 nm (z-stack image size 146x146x20 µm, z-projection presented). **Top row:** A bare NMF run returns concentration maps that exhibit some crosstalk between the images. This corresponds to too narrow spectra (dotted lined in spectra plot) and a NMF triangle much larger than the reference triangle (light red triangle in simplex projection). **Middle row:** When applying the segregation bias as described in the text, the NMF triangle gradually approaches the reference triangle (Appendix - Segregation bias). For λ = 0.3, the different labels are well separated. The RGB panels show an overlay of ECFP: blue, EGFP: green and EYFP: red. (all channels are over-saturated by 1.5 for better visibility), **Bottom row, left:** Schematic drawing of the different cell types; **center:** Spectra plot, bare NMF-spectra are dotted, NMF with segregation bias are solid and reference spectra from singly labeled specimen are dashed; **right:** The simplex projection with red triangles of resulting spectra with increasing segregation bias from light to dark red are shown. The spectra change very little for λ above 0.2 and below 0.01, such that the result is independent of the precise value of λ. In all cases the NMF run was initialized with Gaussian spectra with FWHM of 50 nm and centered between the half-maximum values of the literature spectra.

Figure 3: NMF is capable of separating four simultaneously present labels, when the spectrum of one label is known. A HeLa SS6 cell labeled with FITC phalloidin F-actin, A514 tubulin, A555 transferrin and EtBr DNA was excited at 488 nm and imaged over 16 spectral channels ranging from 503 to 663nm (width 10.7 nm, image size 49x49 µm). A555 transferrin is the weakest label and its spectrum was fixed to the literature spectrum. Running NMF with a slight segregation bias yielded the label distributions and the spectra from blue to red according FITC, A514, A555 and EtBr (bare NMF: dotted, NMF with segregation bias: solid and reference spectra from singly labeled specimen: dashed). Both, the labels distributions and the spectra are estimated to high accuracy. The RGB panels show the false color overlay of F-actin: blue, tubulin: green and nucleic acids: red (RGB₁₂₄), and of F-actin: blue, transferrin: green and nucleic acids: red (RGB₁₃₄). The NMF run was initialized with Gaussian
spectra with FWHM of 75 nm and 524, 558 and 617 nm centre position, which represents the FWHM centre of the literature spectra.

**Figure 4:** Multiple excitations allow the separation of greater number of labels. HeLa SS6 cells labeled as the sample in Figure 3 were excited at 458, 477 and 488 nm while the emission was recorded in 16 channels from 503 to 663 nm (image size 73x73 µm). NTF delivered the label distributions (FITC F-actin, A514 tubulin, A555 transferrin and EtBr DNA); the RGB panel shows the false color overlay of F-actin: blue, transferrin: green and nucleic acids: red; the spectra (upper, right) are colored from blue to red according to FITC, A514, A555 and EtBr (NMF with segregation bias: solid, reference spectra from singly labeled specimen: dashed) and the normalized excitation efficiencies of the four labels (absolute values in table) without invoking auxiliary assumptions are shown in the lower right corner. Only the spectrum of FITC (blue) shows a significant deviation, which is due to cross talk between the FITC and A514. Their excitation efficiencies are exactly collinear, such that multiple excitations do not provide additional information. Hence we encounter too narrow spectra as already discussed for a single excitation. For the run shown, we used Gaussian waveforms with the width and peak position obtained from literature spectra.
Figure 1
Figure 2
Figure 3
Figure 4

**F-actin**

**tubulin**

**transferrin**

**DNA**

**RGB_{134}**

![Graph showing emission and excitation spectra](image)

- **Emission / a.u.**
  - 0.0
  - 0.1
  - 0.2
  - 0.3
  - 0.4
  - 0.5
  - 0.6
  - 0.7
  - 0.8
  - 0.9
  - 1.0

- **Excitation (normalized)**
  - 0.10
  - 0.15
  - 0.20
  - 0.25
  - 0.30
  - 0.35

- **Wavelength / nm**
  - 460
  - 480
  - 500
  - 520
  - 540
  - 560
  - 580
  - 600
  - 620
  - 640
  - 660
Supplementary material: Blind source separation techniques for the decomposition of multiply labeled fluorescence images

Richard A. Neher*, Mišo Mitkovski†,‡, Frank Kirchhoff‡,§, Erwin Neher§,‖, Fabian J. Theis¶,‖, and André Zeug**,††,‡‡

* Kavli Institute for Theoretical Physics, University of California, address: Santa Barbara, CA, 93106
† Department of Molecular Biology of Neuronal Signals, Max Planck Institute of Experimental Medicine Göttingen, address: Hermann Rein Strasse 3, 37075 Göttingen, Germany
‡ Glial Physiology and Imaging, Neurogenetics, Max Planck Institute of Experimental Medicine Göttingen, address: Hermann Rein Strasse 3, 37075 Göttingen, Germany
§ Department Membranbiophysik, Max-Planck Institute for Biophysical Chemistry Göttingen, address: Am Fassberg 11, 37077 Göttingen, Germany
‖ Institute for Bioinformatics and Systems Biology, Helmholtz Zentrum München, address: Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany
¶ Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany, address: Bunsenstraße 10, 37073 Göttingen, Germany
** Department of Neurophysiology and Cellular Biophysics, University Göttingen, Germany
†† Deutsche Forschungsgemeinschaft-Research Center for the Molecular Physiology of the Brain (FZT 103), Göttingen, Germany
‡‡ present address: Department of Cellular Neurophysiology, Hannover Medical School, Germany, address: Carl-Neuberg-Str. 1, 30625 Hannover, Germany

Correspondence should be addressed to André Zeug (zeug.andre@mh-hannover.de).

Table of contents:

1. Data visualization

   Projection into the plan p.2
   The positive domain p.3

2. NMF in Practice

   Pre-processing p.5
   Initial conditions p.6
   Post-processing of NMF results p.10
1. Data visualization

In a spectrally resolved image, each pixel corresponds to a vector $y$ describing the measurements in the different spectral channels. These vectors are a sum of contributions of the $m$ labels in the sample, each of which has a distinct emission spectrum. Since the variation of label concentrations are the main source of signal variation, the vectors $y$ lie to a good approximation in a $m$-dimensional plane. This plane is spanned by $m$ linearly independent spectra of the dyes and, as it contains the origin, is a subspace of the $n$-dimensional space of possible emission vectors. Variation perpendicular to the $m$-dimensional signal subspace is mainly due to photon shot noise. Since each vector $y$ is a combination of non-negative contributions of the $m$ labels, the data points fill only the cone spanned by the spectral vectors of the pure labels.

1.1 Data projection into the plane

While the projection described can be formulated for any number of dyes, we consider the case of $m=3$ dyes in the following. A three-dimensional space cannot be visualized in the plane. However, the absolute signal strength is not relevant for the following consideration. Hence we can normalize all vectors $y$ such that $\sum |y_i|=1$ and thereby reduce the relevant subspace to an affine two-dimensional space. To visualize the data in two dimensions, we have to project the data onto the affine 2-$d$ subspace and map the latter into the plane.

Assume we are given three spectra $s_1$, $s_2$ and $s_3$ with unit 1-norm that span the signal subspace. For each $n$-dimensional vector $y$ with unit 1-norm, we have to determine an orthogonal projection onto the affine 2-$d$ space. The projected vector can be written as

$$z = s_3 + \mu_1 t_1 + \mu_2 t_2,$$  \hspace{1cm} (S1)
where \( t_1 = s_1 - s_2 \) and \( t_2 = s_2 - s_3 \) are vectors within the affine space. From the condition that \( y - z \) is orthogonal to \( t_1 \) and \( t_2 \), we can calculate the coordinates within the affine subspace

\[
\begin{pmatrix}
\mu_1 \\
\mu_2
\end{pmatrix}
= G^{-1} \begin{pmatrix}
\langle y - s_3, t_1 \rangle \\
\langle y - s_3, t_2 \rangle
\end{pmatrix}
\text{ with } G = \begin{pmatrix}
\langle t_1, t_1 \rangle & \langle t_1, t_2 \rangle \\
\langle t_2, t_1 \rangle & \langle t_2, t_2 \rangle
\end{pmatrix},
\tag{S2}
\]

where \( \langle a, b \rangle \) is the scalar product between two vectors \( a \) and \( b \). The visualization of the data in the plane should preserve the angles and the relative length of points in the affine 2-d subspace. Both of these requirements are met, if the coordinates \( v_1 \) and \( v_2 \) in the plane are calculated from \( \mu_1 \) and \( \mu_2 \) as follows:

\[
\begin{pmatrix}
v_1 \\
v_2
\end{pmatrix}
= \begin{pmatrix}
\|t_1\| & \|t_2\| \cos \alpha \\
0 & \|t_1\| \sin \alpha
\end{pmatrix}
\begin{pmatrix}
\mu_1 \\
\mu_2
\end{pmatrix}
\text{ with } \alpha = \arccos \left( \frac{\langle t_1, t_2 \rangle}{\|t_1\| \|t_2\|} \right).
\tag{S3}
\]

Note that only relative locations and angles have a meaning, the absolute orientation and position of the representation of the data points in the plane are arbitrary; here for simplicity we mapped one edge onto the first unit vector.

The analogous projection for two-dimensional data maps the data onto a line segment. The end-points of the line segment correspond to pure dyes while the points in the interior correspond to mixtures of dyes. For more than three dyes, the data is mapped onto a \((m-1)\)-simplex (e.g. a tetrahedron for \(m=4\)).

### 1.2 The positive domain

In addition to visualizing the data, we are interested in the range of eligible non-negative spectra. All eligible spectra have to lie in the same two-dimensional affine subspace and have non-negative entries in each of their \( n \) components vector. Hence, we are looking for the subset of the 2-d subspace that lies in the non-negative sector of the \( n \)-dimensional signal space. The boundary of this subset can be projected into
the plane just as the data points. To this end, one has to calculate the intersections of the coordinate hyperplanes with the affine 2D subspace. These intersections are lines, from which we have to extract the line segment that is non-negative. Its image in the plane can be calculated as above using Eqs. A15 and A16. For each coordinate hyperplane, we get one such line segment, and their union is the boundary of the domain of non-negative spectra.
2. NMF in Practice

2.1 Pre-processing

*Background:* Before an image can be successfully analyzed with NMF, any constant background should be removed. Preferably, the constant background should be determined by measuring the signal from a region in the field of view, which does not contain cells. Alternatively, one can also estimate the dark signal by taking the minimal pixel value of the intensity in each spectral channel. This dark signal has to be subtracted from the data.

*Autofluorescence:* In principle, autofluorescence could be included as an additional dye into the algorithm. Estimation of autofluorescence by NMF, however, is particularly difficult as its spectrum is typically broad, its signal is weak, and it is ubiquitously present. In many cases, autofluorescence is weak enough to be neglected entirely. If not, it is advisable to measure the spectrum of autofluorescence in an unlabeled sample and to include such a component during the NMF optimization as a fixed spectrum. Autofluorescence spectra may differ between different excitation wavelengths.

*Signal-to-noise ratio:* The quality of the decomposition increases with the signal-to-noise ratio of the data. Noisy pixels are detrimental, since the signal in these pixels may be dominated by contributions not captured by the model (e.g. autofluorescence). It is therefore advisable to restrict the analysis to pixels above a threshold. Similarly, saturated pixels have to be removed as their spectral signature is distorted.

*Large Data sets:* For single images a typical NMF run takes about a minute on a
personal computer. However, if large data sets such as z-stacks or time series are to be analyzed, it is advisable to perform NMF on a representative subset of the data. The obtained estimates for the spectra can either be used to calculate concentrations of the entire set by linear unmixing or as starting conditions for subsequent runs on a larger data set. This accelerates convergence, since existing estimates are refined much more quickly than the initial estimation from scratch. Another possibility is coarse-graining the images. Coarse-graining has the advantage of increasing signal-to-noise. On the other hand, it averages over spatial structures and thereby decreases signal modulation. Nevertheless, moderate coarse-graining is often advisable. Results from such runs can be used as starting conditions for higher resolution images.

2.2 Initial conditions

The non-negativity constraints often do not suffice to determine the decomposition uniquely, in which case the outcome of an NMF run can depend on the initial conditions. For three dyes, this ambiguity can be illustrated by projecting the density of data points and the domain of positive spectra into a plane, as described above and in the main text. Every set of non-negative spectra, the convex hull of which encloses the cloud of data points is an eligible solution and might be the outcome of an NMF. To impose initial conditions, we specify a set of spectra and solve for the corresponding concentrations, either by unmixing or by leaving spectra fixed during the initial 10 iterations. Concentrations are initialized with random numbers. To demonstrate the worst-case scenario, we ran NMF on the sample shown in Figure 1 of the main text with extremely narrow initial spectra. The results are shown in Figure S1 (see legend for a discussion of this run).

According to our experience best results are achieved when starting with spectra that
are broad and slightly red-shifted relative to the peak of the dye. In this case, the algorithm narrows the spectra in such a way that all data points can be described by non-negative concentrations and avoids the at times substantial ambiguity of too narrow spectra. The red-shift is indicated because most spectra are very asymmetric with long tails at long wavelengths.

A segregation bias as described in the main text forces spectra towards maximal overlap and thereby removes the ambiguity of NMF. A moderate segregation bias applied to the sample of Figure 1 of the main text significantly improves the decomposition. However, since the data points leave some part of the triangle empty, a strong segregation bias result in too broad spectra and too small triangles. Both of these cases are shown in Figure S2.

The two alternative biases discussed in the appendix of the main paper (Eqs. A7 and A9) have qualitatively similar effects, although they differ in their implementation. The segregation bias (A7) operates on the concentrations. This has the advantage that the weight of the bias is easy to tune, since the bias itself scales with the amount of data. A disadvantage is the high degree of non-linearity, which sometimes results in local minima in which the algorithm gets trapped. The bias (A9) targets the overlap of each pair of spectra directly, but is somewhat harder to tune.
Figure S1: The results of an NMF run with Gaussian initial spectra with a FWHM of 20 nm. The NMF-spectra of A532 F-actin and EtBr DNA are reasonably close to the true spectra, while the spectrum of A488 tubulin is far too narrow. This is a consequence of the vast space between the reference triangle (blue in simplex projection) and the boundary of positive spectra (broken red line). Indeed, the NMF triangle of this run is as large as possible and touches the boundary at each vertex.
Figure S2: Segregation bias applied to the sample of Figure 1 in the main text. A moderate segregation bias $\lambda = 0.005$ significantly improves the decomposition and causes the NMF triangle to enclose the data more tightly (top row). A strong segregation bias, however, results in a secondary peak in the EtBr spectrum, which now contains admixtures of A488. This is possible, since EtBr-DNA is always colocalized with A488-tubulin, as apparent in the absence of data points in the lower left of the reference triangle (blue) (bottom row). Segregation bias reduces the triangle to its smallest possible size, which corresponds to mixing A488 into the EtBr spectrum. The resulting secondary peak in the EtBr spectrum is readily removed by the post-processing tool (see below). Initial conditions for these runs are the same as in Figure 1 of the main text, i.e. Gaussian spectra with a FWHM of 75 nm.
2.3 Post-processing of NMF results
Typically, NMF in combination with a segregation bias provides a good initial decomposition. However, secondary peaks as apparent in the EtBr spectrum in Figure S2 can occur when dyes are strongly colocalized. These errors are often obvious and can be straightforwardly corrected using a software tool, which we developed. The use of the tool is illustrated in Figure S3.

Figure S3: The software tool uses the set of spectra returned by NMF and displays the concentration maps (upper, left), the 2D representation of the spectra and the data (upper right), the intensity distributions in the user interface (lower left) and the spectra simultaneously (lower right). When the mouse cursor is moved within the domain of non-negative spectra of the 2D simplex projection, the corresponding spectrum is displayed in real time in the spectra plot (dotted black line in spectra plot). When a better spectrum is found, the corresponding vertex of the NMF triangle (solid...
red triangle in simplex projection) can be dragged to that new location. The software then recalculates the concentration maps.