Published in final edited form as:

Comput Methods Programs Biomed. 2016 February ; 124: 148–160. doi:10.1016/j.cmpb.2015.10.016.

Quadratic Blind Linear Unmixing: A Graphical User Interface for Tissue Characterization

O. Gutierrez-Navarro^a, D.U. Campos-Delgado^{a,**}, E. R. Arce-Santana^a, and Javier A. Jo^b

D.U. Campos-Delgado: ducd@fciencias.uaslp.mx

^aFacultad de Ciencias, Universidad Autonoma de San Luis Potosi, SLP, Mexico

^bDepartment of Biomedical Engineering, Texas A&M University, College Station, TX, USA

Abstract

Spectral unmixing is the process of breaking down data from a sample into its basic components and their abundances. Previous work has been focused on blind unmixing of multi-spectral fluorescence lifetime imaging microscopy (m-FLIM) datasets under a linear mixture model and quadratic approximations. This method provides a fast linear decomposition and can work without a limitation in the maximum number of components or end-members. Hence this work presents an interactive software which implements our blind end-member and abundance extraction (BEAE) and quadratic blind linear unmixing (QBLU) algorithms in Matlab. The options and capabilities of our proposed software are described in detail. When the number of components is known, our software can estimate the constitutive end-members and their abundances. When no prior knowledge is available, the software can provide a completely blind solution to estimate the number of components, the end-members and their abundances. The characterization of three case studies validates the performance of the new software: *ex-vivo* human coronary arteries, human breast cancer cell samples, and *in-vivo* hamster oral mucosa. The software is freely available in a hosted webpage by one of the developing institutions, and allows the user a quick, easy-to-use and efficient tool for multi/hyper-spectral data decomposition.

Keywords

linear spectral unmixing; endogenous fluorescence; chemometrics; graphical user interface; quadratic optimization

1. Introduction

Recent advances in optics and electronics have led to the development of new devices capable of multi-dimensional measurements. A particular application is the so-called optical biopsy technologies [1, 2]. Their goal is to characterize a sample by extracting their constituent components and contributions in a minimal invasive fashion. Fluorescence-based measurements can describe the biochemical environment [3] within a sample. They are

* Av. Salvador Nava s/n, Zona Universitaria, C.P. 78290, San Luis Potosi, Mexico, Tel. +52 (444) 8262491 ext 2903, Fax +52 (444) 8262384.

** Published by Computer Methods and Programs in Biomedicine (2015), doi:10.1016/j.cmpb.2015.10.016

minimal invasive techniques when no exogenous dyes are employed. However endogenous, intrinsic or auto-fluorescence emissions are characterized by a wide spectral response. Overlapping and high similarity among different endogenous fluorophores is a common problem and the reason why auto-fluorescence signals could be difficult to employ in quantitative characterizations. Multi-dimensional fluorescence measurements [4] provide information that facilitates the identification of auto-fluorescence information [5]. Particularly, fluorescence lifetime imaging microscopy (FLIM) data [6] can record the natural fluorescence decay and extract features for its identification and classification. Quantitative interpretation of the information obtained from FLIM measurements is an ongoing research problem. One approach relies on assuming a mono or multi-exponential model for the fluorescence impulse response. Such is the case of global analysis techniques [7] in frequency domain fluorescence measurements [8]. These methods determine the fractional contribution from the endogenous fluorophores without prior information, and with a fast response [9].

A different approach is to consider every measured fluorescence decay in the dataset as a linear mixture, produced by the contribution of different fluorophores within the sample. Several techniques can be employed to perform this decomposition. Multivariate curve resolution (MCR) methods aim to decompose input data into concentration or abundance profiles and spectra that satisfy multicomponent Beer's Law [10]. They can incorporate several constraints to enforce physical properties in the input data. In [11], the Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) software [12] is employed to separate the intrinsic and extrinsic fluorophores from an in-vivo sample. This method can also estimate the number of components based on the most significant singular values obtained from the input data [13]. Meanwhile, blind source separation methods are other techniques used to characterize fluorescence data. The Non-negative Matrix Factorization (NMF) [14] and its variants aim to decompose the input data into the product of two positive matrices. For example, the authors in [15] employ the NMF method to separate the auto-fluorescence contribution in the sample from extrinsic fluorophores. In fact, similar methods have been studied and applied since the 90's [16] in the decomposition of satellite imagery based on linear and non-linear mixture models [17, 18]. Although the latter can characterize complex relations, the former is usually preferred in practice by its efficiency. Finally, the estimation of the characteristic profiles, or end-members, has been also thoroughly studied specially in the remote sensing literature [19, 20]. Recently, many efforts have focused on the estimation of the number of components in the mixing model or the model order [21, 22, 23]. In fact, no single approach available gathers all the desirable characteristics, some methods require prior information, or they might be sensible to data with high similarity; others can be limited by a maximum number of components or rely on a fitting model. Furthermore, most of them are not tailored for multi/hyper-spectral fluorescence data.

Our previous work in [24, 25] has been dedicated to the development of linear unmixing methods for m-FLIM data based on quadratic optimization. These methods perform blind decomposition, without prior information, under a linear mixture model (LMM). They calculate the end-members, which correspond to the characteristic fluorescence intensity decays, and their abundances within the sample. Furthermore, the number of end-members in the dataset is estimated such that the measurements can be accurately reconstructed and

the abundances have significant spatial meaning. In this context, this paper presents the development of a graphical user interface (GUI) for our Quadratic Blind Linear Unmixing (QBLU) method [24, 25]. This program will be useful for the characterization of multi-spectral data given the input data is not highly mixed, i.e. it contains samples of the pure components. The program can perform decomposition based on a linear mixture model to estimate the correct number of constituents, their characteristic decays and proportional contributions. The software also provides basic tools to discard low SNR regions and border effects. The results can be employed to analyse different multi-spectral datasets according to the contribution of their mixed components.

Our software looks to ease the application of our unmixing methodology with data from different sources, as well as to take advantage of the graphical tools and numerical precision of Matlab. The software was tested for tissue or molecular characterization in three case studies to illustrate possible applications in different fields. The GUI as an application in Matlab and some m-FLIM datasets are freely available online, see Figure 1, at the following web-site: <http://galia.fc.uaslp.mx/~omargn/mflim.html>. The rest of the paper is organized as follows. The LMM employed for the processing of m-FLIM data is detailed in Section 2, as well as the physical constraints in end-members and abundances. Next, the QBLU algorithm is briefly overviewed in Section 3, where the quadratic approximations for the end-member and abundance estimations are highlighted. Section 4 describes some of the basic functionality of the GUI for preprocessing, visualization and output data storage, as well as its setting parameters. The case studies: ex-vivo human coronary arteries, breast tumor cell samples and in-vivo hamster oral mucosa are presented in Section 5. Finally, Section 6 discusses the results and provides final remarks.

2. Linear Mixture Model

The m-FLIM measurements $\mathbf{y}_k \in \mathbb{R}^L$ contain L -dimensional vectors of fluorescence decays recorded at positions $k \in [1, K]$ and at Λ different wavelengths. Due to its nature, the measured fluorescence decays are non-negative, i.e. $\mathbf{y}_k \succeq \mathbf{0}$ for all $k \in [1, K]$. These measurements are usually preprocessed to reduce $\mathbf{y}_k$ the effects of signal variability [26]. One common approach is to normalize the m-FLIM data so each vector sums to one

$$\mathbf{y}_k = \hat{\mathbf{y}}_k / (\mathbf{1}_L^\top \hat{\mathbf{y}}_k). \quad (1)$$

These K normalized measurements $\mathbf{y}_k \in \mathbb{R}^L$ are stored in a matrix $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_K] \in \mathbb{R}^{L \times K}$. The linear mixture model [3, 17] states that each fluorescence decay $\mathbf{y}_k$ is a linear combination of N end-members ($\mathbf{p}_1, \dots, \mathbf{p}_N$) and their abundances ($a_{k,1}, \dots, a_{k,N}$):

$$\mathbf{y}_k = \sum_{j=1}^N \alpha_{k,j} \mathbf{p}_j \quad (2)$$

where $\mathbf{p}_j \in \mathbb{R}^L$, and the abundances satisfy $a_{k,j} \geq 0$ and $\sum_j a_{k,j} = 1$. Our departing assumption is that the number of end-members N and their profiles ($\mathbf{p}_1, \dots, \mathbf{p}_N$) do not change in the m-

FLIM dataset, and the abundances reflect the fractional contribution of each end-member on each spatial location.

Given the number of end-members N , the matrix $\mathbf{P}^N = [\mathbf{p}_1, \dots, \mathbf{p}_N] \in \mathbb{R}^{L \times N}$ gathers the ideal end-members, and the vectors $\mathbf{a}_k = [a_{k,1}, \dots, a_{k,N}]^T \in \mathbb{R}^N$ contain the abundances at position k . The abundances of the N end-members in the dataset are grouped in the matrix $\mathcal{A}^N = [\mathbf{a}_1, \dots, \mathbf{a}_K] \in \mathbb{R}^{N \times K}$. Since we are looking for end-members with characteristics similar to the measurements $\mathbf{y}_k$, the estimation of $\mathbf{P}^N$ is also constrained to positivity and sum-to-one per column. Using these definitions, we can now state the linear mixture model in matrix notation as

$$\mathbf{Y} = \mathbf{P}^N \mathcal{A}^N \quad (3)$$

$$\mathbf{1}_N^T \mathcal{A}^N = \mathbf{1}_K^T \quad \& \quad \mathcal{A}^N \geq 0, \quad (4)$$

$$\mathbf{1}_L^T \mathbf{P}^N = \mathbf{1}_N^T \quad \& \quad \mathbf{P}^N \geq 0. \quad (5)$$

3. Methods

The developed software presented in this paper employs two methods reported in [24, 25] that provide a solution to the decomposition in (3) subject to constraints (4) and (5). The following subsections give a general overview of the formulation for QBLU for two conditions according to the knowledge on the number of end-members in the sample: (i) The value is fixed at N , (ii) The value is bounded in a range N to M ($N < M$).

3.1. Linear Unmixing of N end-members

The Blind End-member and Abundance Estimation (BEAE) algorithm in [25] is employed to decompose the input data $\mathbf{Y}$ into the end-members $\mathbf{P}^N$ and their correspondent abundances $\mathcal{A}^N$. This algorithm requires prior knowledge of the number of end-members N , as well as an initial set of them $(\mathbf{P}^N)^0$ which are selected from within $\mathbf{Y}$ or they could be provided by the user. In any case, the initial end-members need to be linearly independent, and to evaluate this property, the reciprocal condition estimator of the matrix $[(\mathbf{P}^N)^0]^T (\mathbf{P}^N)^0$ is computed, and this value has to be greater than a threshold parameter $\varphi > 0$. The abundances are calculated by using an active set method [27] to solve the quadratic optimization problem

$$\min_{\mathcal{A}^N \geq 0, \mathbf{1}_N^T \mathcal{A}^N = \mathbf{1}_K^T} \frac{1}{2} \|\mathbf{Y} - \mathbf{P}^N \mathcal{A}^N\|_F^2. \quad (6)$$

The estimation of the end-members follows a common strategy in the remote sensing literature: search for a set of end-members which represent all the observed measurements [17, 28]. The BEAE method follows the strategy proposed in ICE [19], which penalizes the distance between the end-members instead of evaluating the volume of the simplex they

describe. In this way, BEAE estimate the end-members and abundances by iteratively minimizing the following quadratic cost function:

$$L(\mathbf{P}^N, \mathcal{A}^N) = \frac{1}{2} \|\mathbf{Y} - \mathbf{P}^N \mathcal{A}^N\|_F^2 + \rho \sum_{i=1}^{N-1} \sum_{j=i+1}^N \|\mathbf{p}_i - \mathbf{p}_j\|_2^2 \quad (7)$$

where $\rho > 0$ represents the regularization weight. The solution for the end-members is obtained from a quadratic optimization of (7) subject to (5). From [25], given N and the abundance matrix $\mathcal{A}^N$, the optimal end-member matrix is

$$\mathbf{P}^N = \left(\mathbf{I}_L - \frac{1}{L} \mathbf{1}_L \mathbf{1}_L^\top \right) \mathbf{Y} (\mathcal{A}^N)^\top \left(\mathcal{A}^N (\mathcal{A}^N)^\top + \rho \mathbf{O} \right)^{-1} + \frac{1}{L} \mathbf{1}_L \mathbf{1}_N^\top \quad (8)$$

where $\mathbf{O} = N \mathbf{I}_N - \mathbf{1}_N \mathbf{1}_N^\top$. The main steps of the BEAE algorithm are shown in Figure 2. The method employs two criteria to stop the optimization procedure:

- if $[L(\mathbf{P}^N, \mathcal{A}^N)^t - L(\mathbf{P}^N, \mathcal{A}^N)^{t-1}]^2 > \varepsilon$, where $\varepsilon > 0$ is a convergence threshold and the superscript t denotes the actual iteration.
- if $t > t_{\max}$, then the maximum number of iterations has been exceeded.

3.2. Linear Unmixing of N to M end-members

The QBLU is based on [24] and can perform blind linear decomposition while providing an estimation of the number of end-members. The estimation of this parameter is one of the most difficult tasks in blind unmixing [29]. A common approach is to determine the model order by using an information theoretic criterion [30] or performing a hypothesis test [31]. Meanwhile, QBLU relies on the BEAE method [25] which outputs are evaluated for different orders of the LMM under a certain range $[N, M]$. The method starts with an initial set of end-members $(\mathbf{P}^N)^0$ at iteration $t = 0$, and then BEAE is computed. Two Neyman-Pearson (NP) detectors [32] assess the temporal and spatial coherence of the current solution $\{\mathbf{P}^N, \mathcal{A}^N\}$. The temporal coherence is related to the linear independence of the end-member with respect to the rest, and the spatial coherence to have representative abundance in the sample.

A fake end-member is detected when the solution fails to comply both detectors by defining a probability of false alarm $P_{FA} > 0$. In the spatial coherence test, a histogram of B bins is constructed with the abundance of each end-member in the sample, and next an spatial coherence descriptor is constructed by defining $\theta > 0$ as the minimum number of occurrences for an abundance bin to be considered significant. Thus, based on the spatial descriptor, an end-member with low spatial coherence will have all its abundances distributed in the interval $[0, \omega_1]$. If at least one fake component has been detected among the solution set, the number of end-members has been exceeded and the correct value is lower $N-1$. On the other hand, when no fake candidate is identified, a new end-member candidate is selected from within the input data $\mathbf{Y}$ and the decomposition is re-evaluated for a higher model order $N+1$, or until the maximum number M is reached. The whole iterative process is depicted in Figure 3.

4. Graphical User Interface and Toolbox

The new software tool was implemented in Matlab and presents a GUI of dedicated functions to facilitate the application of the linear unmixing algorithms described in the previous sections. The software has been made freely available at the following address: <http://galia.fc.uaslp.mx/~omargn/mflim.html>. The GUI can be employed with m-FLIM or other multi/hyper-spectral dataset, as long as the LMM from equation (3), and the restrictions on the abundances (4) and end-members (5) apply. The GUI buttons and menus provide an intuitive and easy to operate interface.

The GUI is available in Matlab app format. The software is compatible with Matlab 2013a and most recent versions. The program requires the Signal Processing, and Statistics Toolboxes from Matlab. The hardware requirements are the same for Matlab 2013a, any Intel or AMD x 86 processor supporting SSE2 instruction set. The supported operative systems are Windows XP Service Pack 2, Ubuntu 12.04, Mac OS X 10.7.4 and more recent versions.

The main interface, shown in Figure 4, provides a basic environment to load multi-spectral data and decompose it employing two different methods: BEAE and QBLU (see Figures 2 and 3). The input dataset has to be arranged by the user in a three-dimensional array, where the first and second dimensions (spatial plane) correspond to the surface of the imaged sample, and the third one must contain the multi/hyper-spectral data. The GUI requires to keep this order in the input dataset, otherwise the decomposition will be infeasible. The user loads the dataset by the menu “File” and label “Load Dataset”, where a dialog box allows the user to locate a .mat file with the input data. Once the data is loaded, the program allows the decomposition by BEAE and QBLU.

The BEAE method decomposes the input data considering N end-members. This method can be selected with the “Decomposition into N ” button. The QBLU algorithm is available through the “Decomposition from N to M ” button. This second method estimates the decomposition from N initial end-members up to M candidates. This option also evaluates the number of components in the dataset, and the decomposition is halted if the program detects a fake end-member [24], where the program informs the user the number of components detected in the text-box “Information Dialog”. Next, the user can recompute the decomposition with the suggested number of end-members.

The status dialog notifies the user when the program is processing the dataset. In addition, the information dialog shows details about the decomposition process, such as the value of the cost function $L(\mathbf{P}^N, \mathcal{A}^N)$ and the results from the NP detectors. After BEAE or QBLU have finished the estimation process, the GUI illustrates the resulting end-members and their respective abundance maps (see Figure 4). A general block diagram of the GUI processing data flow is presented in Fig. 5.

4.1. Parameters

The user can modify some parameters which affect the performance of the BEAE algorithm (ε , $t_{\max}$, ρ , φ), and the NP detectors (P_{FA} , B , ω_1 , θ) in QBLU. These parameters can be

adjusted in the “*Parameters*” window available at the “*Edit*” menu. The default values employed for each parameter and its interpretation are summarized in Table 1, where these constants were selected based on our experience to obtain the best estimation results.

4.2. Preprocessing

For some datasets, there might be regions where the SNR is low due to a larger FOV compared to the size of the sample or an uneven distribution, even the dataset could be too large to process. To deal with these scenarios, the GUI incorporates some preprocessing routines. These tools can be employed by the user after the dataset is loaded (see Figure 5) and they are: (i) downsampling and decimation of the input dataset, (ii) spatial average filtering to improve the SNR, (iii) masking of pixels with low SNR. Once the dataset has been loaded and after this initial preprocessing, the GUI enforces the positivity and sum-to-one constraints for every measurement available. Next, we describe the three preprocessing tools of the input dataset.

4.2.1. Downsampling and Decimation—In the current version, the toolbox make no use of parallel programming or any other software optimization technique. Therefore, the decomposition of large datasets could be a slow process. For a general reference, throughout this work a multi-spectral array of dimensions $60 \times 60 \times 720$ is considered as a small dataset. Meanwhile, any dataset bigger than $400 \times 400 \times 1200$ is regarded as a large one. When dealing with large datasets, the user can reduce the size of the input information, either by discarding some measurements in the spatial plane, which is equivalent to downsampling the spatial coordinates, or by performing decimation of the multi/hyper-spectral data, see Fig. 6. Decimation of the multi/hyper-spectral measurements is performed by the Matlab function “*resample*”; this function adjusts the sampling rate of each measurement by a factor $1/q$, where q is the value selected by the user. The function also applies a low-pass filter to avoid aliasing errors due to the re-sampling procedure. The user can also eliminate measurements at the border region of the spatial plane (first two dimensions of the input dataset array) of the dataset.

4.2.2. Average Spatial Filter—Datasets can also be averaged along the spatial dimensions instead of just discarding measurements. A window kernel of odd dimension $W = \{w_{i,j}\}_{i,j=1}^S \in \mathbb{R}^{S \times S}$ has to be specified by the user, see Figure 7. The values $w_{i,j}$, where $0 \leq w_{i,j} \leq 1$, define the weight given to each measurement in the sampling window for the averaging process along the multi/hyper-spectral dimension. As a result, the total size of the measurements is reduced in the first two dimensions by a factor of $S > 1$, since only one multi/hyper-spectral sample is calculated per spatial window. This process is performed on a copy of the original input dataset.

4.2.3. Mask Filter—Some datasets could present measurements with low SNR or no signal at all. Those measurements can affect the results of the BEAE and QBLU algorithms, specially if we take into account that the normalization process from equation (1) could amplify noisy measurements. For this reason, the GUI incorporates a mask filter, which calculates the average energy of the measurements using the ℓ_2 -norm [33]. The GUI generates an average energy map which can be used as a reference to select a threshold.

Measurements with an average energy below the chosen threshold value will be discarded by the GUI and displayed as white pixels in the output images.

4.3. Selection of Initial End-members

The GUI requires initial end-members $(\mathbf{P}^N)^0$ to start the BEAE and QBLU decomposition algorithms. These candidates can be provided by the user (“*File*” button) or selected from within the input dataset. When no prior information is available, the user is given the option to select them manually “*Manual*”, or let the software choose with the “*Automatic*” option. Manual selection of each candidate end-member is done by the user through the mouse input from the ℓ_2 -norm map of the processed input data. The GUI validates the selected positions, they should be in the unmasked portion of the energy image, and they should be linearly independent (according to the reciprocal condition estimator of $(\mathbf{P}^N)^0$). If they do not comply both conditions, the GUI asks the user to select a new set of candidates. Meanwhile, the automatic selection is based on a pure pixel methodology [21]. This condition tell us that for every j -th end-member, there exist at least one measurement for which $\alpha_{k,j} \approx 1$, while the rest satisfy $\alpha_{k,i} \approx 0$. The proposed automatic selection methodology is described next.

Based on the LMM in (3), for any measurement, its norm is bounded

$$\|\mathbf{y}_k\|_2 \leq \left\| \sum_{j=1}^N \mathbf{p}_j \alpha_{k,j} \right\|_2 \leq \sum_{j=1}^N \|\mathbf{p}_j\|_2 \alpha_{k,j}. \quad (9)$$

When a pure pixel is found, the equality from the previous expression is satisfied. As a result, the first pure candidate is selected by

$$(\mathbf{p}_1)^0 = \arg \max_{\mathbf{y} \in \mathcal{Y}} \|\mathbf{y}\|_2 \quad (10)$$

where $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_K\}$ denotes the set of measurements. The rest of the pure end-members are found in an iterative fashion by searching for the most different measurements. Therefore, given N initial end-members $(\mathbf{P}^N)^0$, the next candidate $(\mathbf{p}_{N+1})^0$ is defined as

$$(\mathbf{p}_{N+1})^0 = \arg \max_{\mathbf{y} \in \mathcal{Y} \setminus \{(\mathbf{p}_1)^0, \dots, (\mathbf{p}_N)^0\}} \sum_{n=1}^N \|\mathbf{y} - (\mathbf{p}_n)^0\|_2, \quad (11)$$

where $\setminus$ represents the difference operation between sets.

4.4. Output Data

The user can save the output data after the BEAE or QBLU methods have been applied. There are three possible outputs. First, a .jpg or .eps file of the resulting end-members plot can be save using the button “*Save End-members Figure*”. The abundance maps can also be saved in a .jpg or .eps file by the option “*Save Abundances Figure*”. Furthermore, a Matlab .mat file containing both outputs can be stored using the button “*Save Output Data*”. This output contains the estimated end-member matrix mv_P (the rows represent the multi-

spectral dimension and the columns the number of end-member) and the abundance maps *mv_Abundances* (three-dimensional array where the first two dimensions represent the spatial plane and the third one, the number of end-member).

5. Case Studies

In this section, three case studies show the GUI and toolbox capabilities for the characterization of m-FLIM datasets. These examples illustrate possible applications of the GUI and toolbox for tissue/cellular quantification and identification. For validation purposes, the end-members obtained from the decomposition process need to be classified. One method to do so is to calculate their lifetimes in each wavelength band. In the case of time-resolved m-FLIM measurements, the fluorescence impulse responses were estimated [3] for each end-member and at each wavelength, and the Laguerre deconvolution method from [34] was employed for this purpose. For this strategy, a 12-th order model was selected with a shape parameter $\alpha = 0.88$ to provide a good fitting in all tested scenarios.

5.1. Ex-vivo Human Coronary Arteries

The first experiment consists in the study of four m-FLIM datasets from *ex-vivo* human atherosclerotic plaques. The m-FLIM data was obtained from [35]. The lumen of post-mortem human coronary arteries was imaged using a time-resolved m-FLIM device [36]. The multi-spectral emissions were recorded at the following wavelengths: 390 ± 40 nm, 450 ± 40 nm and 550 ± 80 nm. Each single dataset has a dimension of $60 \times 60 \times 768$. The imaged data has a field of view (FOV) of 2×2 mm² with a temporal resolution of 250 ps. All the single datasets are stacked into a large column matrix with a total dimension of $240 \times 60 \times 768$. The aim is to provide a detailed description by estimating the concentrations of each constituent fluorophore.

The QBLU method, button “*Decompose from N to M*”, was employed by using two initial end-members selected automatically. The parameters for this experiment are shown in Table 1. The decomposition stopped when the solution for four end-members was tested, see Figure 9. The fourth end-member was detected as fake by the spatial coherence detector. This is due to its low abundance, as can be seen in the rightmost picture in Figure 9 B). The GUI suggests the user to repeat the experiment with three end-members. The results with the correct number of end-members are shown in Figure 10. The lifetimes estimated for each end-member are shown in Table 2. In these datasets, three endogenous fluorophores are expected [37] in the sample: collagen, elastin and low-density lipoproteins (LDL). Thus, by the resulting lifetimes, the first end-member is identified as collagen, while the second and third ones are identified as LDL and elastin, respectively. According to the abundance maps in subplot B) of Figure 10, the most dominant component in the first two datasets (rows 1–60 and 61–120) is collagen; in the third dataset (rows 121–180), elastin has a more significant concentration with some contribution of collagen; and finally, in the last dataset (rows 181–240), LDL has a dominant abundance, where at the upper-right and lower-left edges, there is a larger concentration of collagen.

5.2. Human Breast Cancer Cells

The certain fluorophores can assess the diagnosis of pathologies, since they modify the cellular metabolism, such as cancer. In this experiment, m-FLIM data from human breast cancer cells was characterized. The aim is to estimate the abundances of two metabolic coenzymes, reduced NADH and FAD. The sample was obtained from [38] where the authors cultured the samples from a non-cancerous epithelium cell line, MCF10A. The data was acquired using a multiphoton fluorescence microscope (Prairie Technologies). This system employs two excitation sources, which were tuned at 750 nm and 890 nm for NADH and FAD, respectively. The device recorded the fluorescence emission with two bandpass filters: 440 ± 80 nm for NADH and 550 ± 100 nm for FAD. The m-FLIM dataset has a dimension of $256 \times 256 \times 380$.

The concentration levels of NADH and FAD were estimated by performing the linear decomposition. For this experiment, the number of components will be known a priori. Due to the presence of solution within the imaged region, several positions contain low SNR, plus there is a fluorescence emission from the cell nucleus. The mask filter from Section 4 was employed to discard low SNR samples, namely those measurements with an ℓ_2 -norm value lower than 22% of the peak value, see Figure 11. The BEAE algorithm was employed by button “*Decompose into N*” in the GUI. Two initial end-members were selected by using the automatic procedure, button “*Auto*”, and the parameters from Table 1 were used. The final end-members and their abundance maps are displayed in Figure 12. Once again, the lifetimes from the deconvolved intrinsic decays were calculated, see Table 3. Hence, as expected from the results in [38], each identified end-member has an alternated dominant contribution per channel, which is consistent with the spectral properties of FAD and NADH. From the abundance maps in subplot B) of Figure 12 and identified end-members in Table 3, the most significant component throughout the sample is NADH.

In order to illustrate that the number of end-members in the sample is exactly two, we performed a new evaluation by raising this value to three and running once more the BEAE by the button “*Decompose into N*” and automatically selecting the initial candidates. The resulting decomposition is shown in Fig. 13, where the algorithm detects that the third end-member is fake by its low spatial contribution in its abundance map. As a result, the original observation of a second order model in the LMM is confirmed for this dataset.

5.3. *In-vivo* Hamster Oral Mucosa

The third example involves the decomposition of *in-vivo* m-FLIM datasets from [39] that consider measurements from the cheek pouch of a golden Syrian hamster. The sample was treated with 7, 12-dimethylbenz[*a*]anthracene (DMBA) to induce oral carcinogenesis. The dataset was recorded using a time-resolved m-FLIM device [40] and has a dimension of $400 \times 400 \times 1200$. The imaged surface covers a FOV of 16×16 mm². The fluorescence intensity decays were recorded with a temporal resolution of 160 ps at three wavelength bands: 390 nm, 450 nm and >500 nm.

Before performing the linear decomposition, an average spatial filter of 3rd order was performed to improve the SNR, and a mask within a 5% threshold value was employed to

discard samples with a low signal level, see Figure 14. The QBLU algorithm was applied with 2 initial end-members and 5 components as maximum by using the parameters from Table 1. The initial decomposition stopped at 4 end-members, so the GUI detects that the last end-member could not satisfy the spatial coherence test. The resulting end-members for a 4-th order LMM are presented in Fig. 15, as well as their abundances, where the low spatial contribution of the 4th end-member is evident. Therefore, the GUI estimates that only 3 end-members are actually present in the sample, and advises the user to repeat the process considering 3 end-members as the maximum number allowed. The final results were estimated for three end-members, and are shown in subplot A) of Figure 16. The end-members were identified by using their average lifetimes, which are shown in Table 4. The first end-member corresponds to collagen, and the third one is a combination of NADH and FAD. The second end-member, which presents minimum emission in the 390 nm and 450 nm wavelength bands, was identified as porphyrin which is consistent with the results in [39]. Meanwhile, from the abundance maps in subplot B) of Figure 16 and the identified end-members in Table 4, collagen is the most significant component throughout the sample. However, in upper-right corner, now porphyrin and NADH/FAD become more dominant and there is no presence of collagen. This area in the sample is related to oral carcinogenesis [39].

6. Discussion

We have presented a new software tool for the characterization of m-FLIM data that includes a GUI and a toolbox of dedicated functions. The GUI estimates the dimension of a multi/hyper-spectral dataset along with a quantitative description of the end-members and their abundances. The numerical methods are based on quadratic approximations for a LMM with robust results. The software was tested in three different scenarios to illustrate the characterization of laboratory data and potential applications. The blind decomposition for three components of a m-FLIM dataset of size $60 \times 60 \times 710$ is performed in around 40 seconds once the data is loaded. A bigger sample takes considerable more time. For example the m-FLIM dataset employed in Section 5.3, with dimensions $400 \times 400 \times 1200$, is decomposed for three components in around 8 minutes. The system employed for these tests was a Core i7-5500U CPU at 2.40 GHz with 8 GB in RAM. The GUI includes pre-processing tools to improve the SNR of the multi/hyper-spectral measurements or reduce the size of the dataset. Masking filters are also provided to deal with samples containing low or no signal within the imaged region. The numerical methods offer a quantitative analysis of m-FLIM data, which is not based on a predefined structure of the end-members or limited by a fixed number of them. Different input datasets can also be characterized using this software as long as they comply with the constraints of the LMM defined by equations (3), (4) and (5). The GUI is designed to facilitate the quantitative analysis of multi/hyper-spectral measurements for different chemometrical applications.

Future work will incorporate deconvolution tools for the estimation of the fluorescence impulse responses from the extracted end-members. This will be useful to quantify their average lifetimes and other features for the identification of the end-members. One current limitation for practical clinical applications is the simultaneous decomposition of several datasets. A solution for this problem is needed to characterize regions which cover FOVs in

the order of hundreds of square centimetres and will be addressed in our future work. Finally, we will consider to offer a standalone GUI interface for different operating systems to avoid the necessity of having installed Matlab in the host computer.

Acknowledgments

The authors acknowledge Dr. Melissa Skala and Alex Walsh, from the Department of Biomedical Engineering at Vanderbilt University, for providing the FLIM dataset of human breast cancer cells employed in Section 5.2. This research was supported by grants from CONACYT-TAMU (2012-034) and NIH (R01CA138653, R01HL11136). The work of O. Gutierrez-Navarro and D.U. Campos-Delgado was supported by CONACYT and COMEXUS Fulbright-Garcia Robles fellowships.

AcronymsNotation

BEAE	Blind End-member and Abundance Extraction
FLIM	Fluorescence Lifetime Imaging Microscopy
FOV	Field of View
GUI	Graphical User Interface
LDL	Low Density Lipo-proteins
LMM	Linear Mixture Model
MCR	Multivariate Curve Resolution
m-FLIM	multi-spectral FLIM
NMF	Non-negative Matrix Factorization
NP	Neyman Pearson
SNR	Signal-to-Noise Ratio
QBLU	Quadratic Blind Linear Unmixing
x	scalar element
$\mathbf{x}$	vector element
$\mathbf{X}$	matrix element
$\mathbb{R}^N$	N -dimensional real vectors
$\mathbb{R}^{N \times M}$	$N \times M$ real matrices
$(\bullet)^T$	transpose operator
$\succeq$	component-wise inequality for vectors or matrices
$\mathbf{I}_N$	identity matrix of dimension $N \times N$
$\mathbf{1}_N$	N -dimensional vector with unitary entries
$\ \bullet\ _2$	Euclidean vector norm
$\ \bullet\ _F$	Frobenious matrix norm

References

1. Green B, Cobb AR, Brennan PA, Hopper C. Optical diagnostic techniques for use in lesions of the head and neck: review of the latest developments. *British Journal of Oral and Maxillofacial Surgery*. 2014; 52(8):675–680. [PubMed: 25015019]
2. Fuller C, Camilon R, Nguyen S, Jennings J, Day T, Gillespie MB. Adjunctive diagnostic techniques for oral lesions of unknown malignant potential: Systematic review with meta-analysis. *Head & Neck*. Forthcoming.
3. Lakowicz, JR. *Principles of Fluorescence Spectroscopy*. Springer; 2006.
4. Gadella, T. Elsevier. *FRET and FLIM Techniques*, Chapter 4: Multidimensional fluorescence imaging. 2008. p. 133-169.
5. Chorvat J, Chorvatova DA. Spectrally resolved time-correlated single photon counting: a novel approach for characterization of endogenous fluorescence in isolated cardiac myocytes. *European Biophysics Journal*. 2006; 36(1):73–83. [PubMed: 17033778]
6. Clegg, RM.; Periasamy, A. *FLIM Microscopy in Biology and Medicine*. Chapman and Hall/CRC; 2009.
7. Verveer PJ, Squire A, Bastiaenskremer PI. Global analysis of fluorescence lifetime imaging microscopy data. *Biophysical Journal*. 2000; 78:2127–2137. [PubMed: 10733990]
8. Kremers G-J, van Munster EB, Goedhart J, Gadella TWJJ. Quantitative lifetime unmixing of multiexponentially decaying fluorophores using single-frequency fluorescence lifetime imaging microscopy. *Biophysical Journal*. 95(1)
9. Warren SC, Margineanu A, Alibhai D, Kelly DJ, Talbot C, Alexandrov Y, Munro I, Katan M, Dunsby C, French PMW. Rapid global fitting of large fluorescence lifetime imaging microscopy datasets. *PLoS ONE*. 2013; 8(8):e70687. [PubMed: 23940626]
10. Ruckebusch C, Blanchet L. Multivariate curve resolution: a review of advanced and tailored applications and challenges. *Analytica chimica acta*. 2013; 765:28–36. [PubMed: 23410623]
11. Xu H, Rice BW. In-vivo fluorescence imaging with a multivariate curve resolution spectral unmixing technique. *Journal of Biomedical Optics*. 14:064011. [PubMed: 20059249]
12. Jaumot J, Gargallo R, de Juan A, Tauler R. A graphical user-friendly interface for mcr-als: a new tool for multivariate curve resolution in matlab. *Chemometrics and Intelligent Laboratory Systems*. 2005; 76(1):101–110.
13. Jaumot J, de Juan A, Tauler R. Mcr-als gui 2.0: New features and applications. *Chemometrics and Intelligent Laboratory Systems*. 2015; 140:1–12.
14. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature*. 1999; 401(6755):788–791. [PubMed: 10548103]
15. Montcuquet AS, Herve L, Navarro F, Dinten JM, Mars JI. Nonnegative matrix factorization: a blind spectra separation method for in vivo fluorescent optical imaging. *Journal of Biomedical Optics*. 2010; 15(5):056009. [PubMed: 21054103]
16. Craig M. Minimum-volume transforms for remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing*. 1994; 32(3):542–552.
17. Ma WK, Bioucas-Dias J, Chan TH, Gillis N, Gader P, Plaza A, Ambikapathi A, Chi CY. A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Processing Magazine*. 2014; 31(1):67–81.
18. Dobigeon N, Tournieret JY, Richard C, Bermudez JCM, McLaughlin S, Hero AO. Nonlinear unmixing of hyperspectral images: Models and algorithms. *IEEE Signal Processing Magazine*. 2014; 31(1):82–94.
19. Berman M, Kiiveri H, Lagerstrom R, Ernst A, Dunne R, Huntington J. ICE: a statistical approach to identifying endmembers in hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*. 2004; 42(10):2085–2095.
20. Miao L, Qi H. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*. 2007; 45(3):765–777.

21. Chan T-H, Chi C-Y, Keizer K. Hyperspectral data geometry-based estimation of number of endmembers using ℓ_1 -norm-based pure pixel identification algorithm. *IEEE Transactions on Geoscience and Remote Sensing*. 2013; 51(5):2753–2769.
22. Andreou C, Karathanassi V. Estimation of the number of endmembers using robust outlier detection method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2014; 7(1):247–256.
23. Eches O, Dobigeon N, Tourneret JY. Estimating the number of end-members in hyperspectral images using the normal compositional model and a hierarchical bayesian algorithm. *IEEE Journal of Selected Topics in Signal Processing*. 2010; 4(3):582–591.
24. Gutierrez-Navarro O, Campos-Delgado DU, Arce-Santana ER, Maitland KC, Cheng S, Jabbour J, Malik B, Cuenca R, Jo JA. Estimation of the number of fluorescent end-members for quantitative analysis of multispectral flim data. *Optics Express*. 2014; 22(10):12255–12272. [PubMed: 24921344]
25. Gutierrez-Navarro O, Campos Delgado D, Arce-Santana E, Mendez M, Jo J. Blind end-member and abundance extraction for multi-spectral fluorescence lifetime imaging microscopy data. *IEEE Journal of Biomedical and Health Informatics*. 2014; 18(2):606–617. [PubMed: 24608060]
26. Manolakis D, Shaw G. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Processing Magazine*. 2002; 19(1):29–43.
27. Gutierrez-Navarro O, Campos Delgado D, Arce-Santana E, Mendez M, Jo J. A fully constrained optimization method for time-resolved multi-spectral fluorescence lifetime imaging microscopy data unmixing. *IEEE Transactions on Biomedical Engineering*. 2013; 60(6):1711–1720. [PubMed: 23358941]
28. Bioucas-Dias J, Plaza A, Dobigeon N, Parente M, Du Q, Gader P, Chanussot J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2012; 5(2):354–379.
29. Ambikapathi A, Chan T-H, Chi C-Y, Keizer K. Hyperspectral data geometry-based estimation of number of endmembers using ℓ_p -norm-based pure pixel identification algorithm. *IEEE Transactions on Geoscience and Remote Sensing*. 2013; 51(5):2753–2769.
30. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974; 19(6):716–723.
31. Harsanyi J, Farrand W, Chang C-I. Determining the number and identity of spectral endmembers: an integrated approach using neyman-pearson eigen-thresholding and iterative constrained rms error minimization. *Proceedings of the Thematic Conference on Geologic Remote Sensing*. 1993; 1:395–395.
32. Levy, BC. *Principles of Signal Detection and Parameter Estimation*. 1. Springer Publishing Company, Incorporated; 2008.
33. Unser M. Texture classification and segmentation using wavelet frames. *IEEE Transactions on Image Processing*. 1995; 4(11):1549–1560. [PubMed: 18291987]
34. Pande P, Jo JA. Automated analysis of fluorescence lifetime imaging microscopy (flim) data based on the laguerre deconvolution method. *IEEE Transactions on Biomedical Engineering*. 2011; 58(1):172–181. [PubMed: 20934946]
35. Thomas P, Pande P, Clubb F, Adame J, Jo JA. Biochemical imaging of human atherosclerotic plaques with fluorescence lifetime angiography. *Photochemistry and Photobiology*. 2010; 86(3):727–731. [PubMed: 20331523]
36. Shrestha S, Applegate BE, Park J, Xiao X, Pande P, Jo JA. High-speed multispectral fluorescence lifetime imaging implementation for in vivo applications. *Optics Letters*. 2010; 35(15):2558–2560. [PubMed: 20680057]
37. Arakawa K, Isoda K, Ito T, Nakajima K, Shibuya T, Ohsuzu F. Fluorescence analysis of biochemical constituents identifies atherosclerotic plaque with a thin fibrous cap. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2002; 22(6):1002–1007.
38. Walsh AJ, Cook RS, Manning HC, Hicks DJ, Lafontant A, Arteaga CL, Skala MC. Optical metabolic imaging identifies glycolytic levels, subtypes, and early-treatment response in breast cancer. *Cancer research*. 2013; 73(20):6164–6174. [PubMed: 24130112]

39. Jabbour JM, Cheng S, Malik BH, Cuenca R, Jo JA, Wright J, Cheng Y-SL, Maitland KC. Fluorescence lifetime imaging and reflectance confocal microscopy for multiscale imaging of oral precancer. *Journal of Biomedical Optics*. 2013; 18(4):046012–046012. [PubMed: 23595826]
40. Cheng S, Rico-Jimenez JJ, Jabbour J, Malik B, Maitland KC, Wright J, Cheng YSL, Jo JA. Flexible endoscope for continuous in vivo multispectral fluorescence lifetime imaging. *Optics Letters*. 2013; 38(9):1515–1517. [PubMed: 23632536]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Screenshot of the main website at <http://galia.fc.uaslp.mx/~omargn/mflim.html>.

Figure 2.

Block diagram of BEAE Algorithm. A) Input dataset Y and BEAE parameters are introduced. B) Initial end-members are chosen and verified. C) Abundances are estimated by using the method from [27]. D) End-members are refined by employing equation (8). E) The iteration is repeated until the cost function in (7) is lower than ε or the maximum number of iterations t_{max} is reached. F) The output data is the matrices containing the end-members P^N and abundances $\mathcal{A}^N$ for a N -th LMM.

Figure 3.

Block diagram for the QBLU procedure which also estimates the number of end-members.

A) Input dataset $\mathbf{Y}$ and QBLU parameters are introduced. B) Set the initial N and maximum number of end-members M . C) Initial end-members are chosen and verified. D) Estimate $\mathbf{P}^N$ and $\mathcal{A}^N$ using BEAE. E) The result is evaluated using the NP detectors. F) The procedure stops when one of the NPs are not satisfied. G) When no fake components are detected, a new initial candidate end-member $(\mathbf{P}^{N+1})^0$ is selected. H) Once the process is stopped, the GUI indicates the number of estimated end-members N and which NP detector was not

satisfied. The user is given the option to re-calculate the solution using the correct number of end-members.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4.
The main window of the GUI for QBLU.

Figure 5.
General block diagram of the GUI processing data flow.

Figure 6.
Interface for the downsampling and decimation tools in the GUI.

Figure 7.
Average filter window in the GUI.

Figure 8.
Initial end-members window in the GUI.

Figure 9.

Results obtained from the QBLU algorithm of the *ex-vivo* human coronary arteries sample considering four end-members. A) Plot of the extracted end-members B) Abundance maps obtained. The 4-th end-member (4-th column) triggered the NP spatial coherence detector due to its low abundance.

Figure 10.

Decomposition of four *ex-vivo* human coronary arteries samples considering three final end-members. A) Plot of the estimated end-members. B) Abundance maps where each column represents the estimated concentration for each end-member.

Figure 11.

Masking filter of the human breast cancer cell sample, where white pixels denote the samples discarded by the mask.

Figure 12.

Results obtained for two end-members from the human breast cancer cell sample. A) Extracted end-members correspond to FAD and NADH. B) Abundance maps highlight the abundance of FAD (left image) and NADH (right image), where white pixels were discarded by the mask filter.

Figure 13.

Results obtained for three end-members from the human breast cancer cell sample. A) Extracted end-members. B) Abundance maps where the third end-member has low spatial coherence.

Figure 14.
Mask filter applied to the *in-vivo* hamster oral mucosa data.

Figure 15.

Decomposition of the *in-vivo* hamster oral mucosa data with four end-members. A) Plot of the extracted end-members. B) Abundance maps obtained where each column represents the estimated concentration for each end-member. The fourth end-member was discarded by the spatial coherence detector.

Figure 16.

Final results obtained for the *in-vivo* hamster oral mucosa sample with three end-members.

A) Plot of the extracted end-members which were later identified as NADH/FAD, porphyrin and collagen. B) Abundance maps obtained where each column represents the estimated concentration for each end-member.

Table 1

Default parameters employed by BEAE and QBLU.

Description	Parameter	Value
Convergence Condition for BEAE	ε	5×10^{-6}
Maximum Number of Iterations in BEAE	$t_{\max}$	10
Regularization Weight for BEAE Cost Function	ρ	0.75
Similarity Condition for Initial End-members	φ	1×10^{-5}
False Alarm Probability	P_{FA}	0.001
Number of Bins in Abundance Histogram	B	30
Upper Limit of Uniform Distribution in Coherence Descriptor	ω_1	0.2
Threshold for Significant Abundances	θ	0.005

Table 2

Lifetimes estimated from the three estimated end-members in the *ex-vivo* human coronary arteries sample.

End-member	Channel Band			Id. Component
	$390 \pm 40nm$	$450 \pm 40nm$	$550 \pm 80nm$	
1	5.14 ns	5.15 ns	5.68 ns	Collagen
2	5.16 ns	7.01 ns	8.10 ns	LDL
3	3.18 ns	3.70 ns	3.10 ns	Elastin

Table 3

Lifetimes of the two end-members extracted from the human breast cancer cell sample

End-member	Channel Band		Id. Component
	440 ± 80 nm	550 ± 100 nm	
1	N/A	1.67 ns	FAD
2	1.86 ns	N/A	NADH

Table 4

Lifetimes of the three end-members obtained for the *in-vivo* hamster oral mucosa sample.

End-member	Channel band			Id. Component
	390 nm	450 nm	> 500 nm	
1	3.09 ns	2.59 ns	1.95 ns	NAHD/FAD
2	N/A	N/A	9.70 ns	Porphyrin
3	5.60 ns	6.06 ns	6.86 ns	Collagen