

# COMPLEXITY REDUCTION BY CONVEX CONE DETECTION FOR UNMIXING HYPERSPECTRAL IMAGES OF BACTERIAL BIOSENSORS

Charles Soussen\*, Sebastian Miron\*, Fabrice Caland\*, David Brie\*, Patrick Billard†, Christian Mustin†

\*Centre de Recherche en Automatique de Nancy.  
Nancy-University, CNRS. Faculté des sciences,  
B.P. 70239, F-54506 Vandœuvre-lès-Nancy, France

†Laboratoire des Interactions Microorganismes-Minéraux-  
Matière Organique dans les Sols. CNRS, Nancy-University.  
B.P. 70239, F-54506 Vandœuvre-lès-Nancy, France

Tel: (+33)-3 83 68 44 71, fax: (+33)-3 83 68 44 62, email: Charles.Soussen@cran.uhp-nancy.fr

## ABSTRACT

We address the problem of complexity reduction in hyperspectral image unmixing. When the hyperspectral images are highly resolved, we propose to select a limited number of pixels, therefore reducing dramatically the size of the data. Then, the related mixtures are used as inputs to a positive source separation algorithm. Our pixel selection procedure is based on a convex cone analysis of the data mixtures; indeed, positive mixtures of sources are embedded in a convex cone whose boundary contains complete available information regarding the sources. We search for the least number of mixtures embedding the convex cone and then store the corresponding pixel indices as the selected pixels. We apply this method to the analysis of hyperspectral images of bacterial cells obtained on a confocal microscope. The bacterial cells, acting as whole-cell biosensors, display great potential as living transducers in sensing applications.

## 1. INTRODUCTION

Blind positive source separation (PSS) is known to be a difficult, ill-posed inverse problem, because the positivity of both source signals and mixture coefficients is not sufficient to guarantee a unique solution. In the last decade, intensive efforts were made to study the identifiability of the source separation problem under the positivity constraints, and efficient algorithms were proposed. The algorithms either use the positivity assumptions only [1, 2] or use additional constraints to force the solution to be unique [3, 4].

In imaging applications, the number of mixtures corresponds to the number of pixels. For highly resolved images, this number can typically reach  $512^2$ . Most positive source separation algorithms cannot be applied to such voluminous data sets because they require a considerable amount of time and/or memory. Therefore, it is necessary to reduce the data sets, e.g., by considering a limited number of pixels.

There exists a rich literature on pixel selection but we will only refer to the approaches similar to that proposed in this paper. One of the most popular algorithms for pixel selection is *Pixel Purity Index* (PPI) proposed by Boardman in [5]. PPI is based on the repeated projection of the mixtures onto random unit vectors, and the search for the extreme pixels in each projection. From a cumulative account recording the number of times each pixel is found to be extreme, PPI identifies the mixtures which are more likely to be the pure sources [5, 6]. In [7], the eigenvectors of the spectral correlation matrix are used to estimate the sources, in the convex cone analysis (CCA) framework. The main drawback

of these methods is that the selected pixels are directly considered as pure sources, which is not always a valid assumption in practice. In [8], the independent component analysis (ICA) was used to yield a rough classification of pixels, and then to design a heuristic selection procedure. This procedure avoids to select pixels belonging to artifact regions. However, the number of selected pixels is arbitrarily chosen and the positivity assumption is not taken into account.

The algorithm proposed in this paper aims at selecting effectively the relevant pixels, based on the CCA framework. The main idea in CCA is that each non-negative mixture of sources lays inside the convex cone spanned by the sources. The objective of CCA is to find the mixtures which fully identify the convex cone boundary. Contrary to [7], we do not set the sources as the “boundary mixtures”. Instead, these selected mixtures are used as the input of a positive source separation procedure.

In Section 2, we state the positive source separation problem and we discuss the complexity reduction issues. In Section 3, we introduce the CCA framework and we present our pixel selection algorithm. In Section 4, we apply this method to the analysis of hyperspectral images obtained on a confocal microscope, which characterize genetically engineered bacteria expressing optically active reporter molecules in response to chemical or environmental effectors [9].

## 2. POSITIVE SOURCE SEPARATION AND COMPLEXITY REDUCTION ISSUES

### 2.1 Positive source separation

In source separation problems, the observations are grouped together in a matrix  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_M]^t$  of size  $M \times N$  representing a collection of  $M$  mixture signals  $\mathbf{x}_i \in \mathbb{R}^N$ . The linear instantaneous mixture model expresses each mixture  $\mathbf{x}_i$  as a linear combination of  $P$  source signals  $\mathbf{s}_1, \dots, \mathbf{s}_P$ :

$$\mathbf{x}_i = \sum_{j=1}^P a_{ij} \mathbf{s}_j. \quad (1)$$

In the matrix form, this model rereads:

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \quad (2)$$

where  $\mathbf{A}$  is the mixing matrix, of size  $M \times P$ , whose  $i$ -th row gathers the weight of all sources in the mixture  $\mathbf{x}_i$ , and  $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_P]^t$  is the source matrix, of size  $P \times N$ .

Given a set of observations  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_M]^t$ , the blind source separation problem consists in decomposing  $\mathbf{X}$  according to (2). In the following,  $P$  will be assumed to be the

rank of  $\mathbf{X}$ , and we will assume that  $\mathbf{A}$  and  $\mathbf{S}$  are full rank, *i.e.*,  $\text{rank}[\mathbf{A}] = \text{rank}[\mathbf{S}] = P$ .

For noisy data, model (2) is replaced by:

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N}, \quad (3)$$

where the noise matrix  $\mathbf{N}$  takes account of the modeling and the measurement errors. The blind source separation of  $\mathbf{X}$  consists in searching for the factorization  $\mathbf{X} \approx \mathbf{A}\mathbf{S}$  for which the residual between  $\mathbf{X}$  and  $\mathbf{A}\mathbf{S}$  is minimal. In this paper, we impose that  $\mathbf{A}$  and  $\mathbf{S}$  are non-negative matrices, *i.e.*, all elements  $a_{ij}$  and  $s_{ij}$  are non-negative.

## 2.2 Unicity issues and regularization

Let us start by considering the noiseless problem and the non-negativity constraint. This constraint is known to regularize the source separation problem, because it restricts the range of solutions for (2). However, it does not always guarantee the unicity of the decomposition [10, 11] (unicity is defined up to a permutation and scaling of the sources; if there is no other ambiguity, the decomposition is called “unique”). Typically, the non-negative matrix factorization algorithm [2] provides one exact solution among all the possible solutions satisfying  $\mathbf{A} \geq \mathbf{0}$  and  $\mathbf{S} \geq \mathbf{0}$ . Often, additional constraints must be imposed, *e.g.*, the sparsity of the source samples and/or of their weights in the mixtures in order to uniquely identify the sources [3, 4]. However, there exists a (rather technical) necessary and sufficient condition, based on the positive span of the mixtures  $\mathbf{x}_i$ , under which the identifiability of (2) is guaranteed with the only non-negativity constraint [10, 11]. Let us first define the notion of positive span.

**Definition.** The *positive span* of a family of vectors  $\{z_1, \dots, z_d\}$  is the set  $\text{vect}^+(z_1, \dots, z_d) = \{z = \sum_i \alpha_i z_i, \alpha_i \in \mathbb{R}_+\}$ , where  $\mathbb{R}_+$  denotes the set of non-negative reals.

In the following, we will denote by  $\text{vect}(\mathbf{X})$  and  $\text{vect}^+(\mathbf{X})$  the span and the positive span of the rows of  $\mathbf{X}$ , *i.e.*,  $\mathbf{x}_1, \dots, \mathbf{x}_M \in \mathbb{R}_+^N$ . Shortly speaking, the condition of [10, 11] states that  $\text{vect}^+(\mathbf{X})$  has to be “close enough” to  $\text{vect}(\mathbf{X}) \cap \mathbb{R}_+^N$  in order to guarantee the unicity of factorization (2).

In the noisy case, we do not search for an exact factorization of  $\mathbf{X}$  anymore, but rather for an approximate factorization in which  $\mathbf{A}$  and  $\mathbf{S}$  satisfy specific properties. The algorithms which jointly estimate  $\mathbf{A}$  and  $\mathbf{S}$  from noisy data often rely on prior assumptions on the sources and/or the mixture coefficients, *e.g.*, sparsity assumptions [3, 4].

## 2.3 Complexity reduction

For hyperspectral image applications,  $M$  is the number of pixels and the mixture  $\mathbf{x}_i$  is related to the  $i$ -th pixel. When the value of  $M$  is huge (typically,  $M$  can reach  $512^2$  or more), classical algorithms either yield poor results [1, 3] or cannot be used [4] because of their greediness in terms of memory storage and computation time. In all cases, it is of interest to reduce the dimension of the data  $\mathbf{X}$  by keeping a subset of the pixels  $\{1, \dots, M\}$ , for instance those for which only a limited number of sources are present. Formally, we want to extract a matrix  $\mathbf{X}_e$  from  $\mathbf{X}$  whose number of rows is limited, and then to unmix the new matrix  $\mathbf{X}_e$ :

$$\mathbf{X}_e \approx \mathbf{A}_e \mathbf{S}. \quad (4)$$

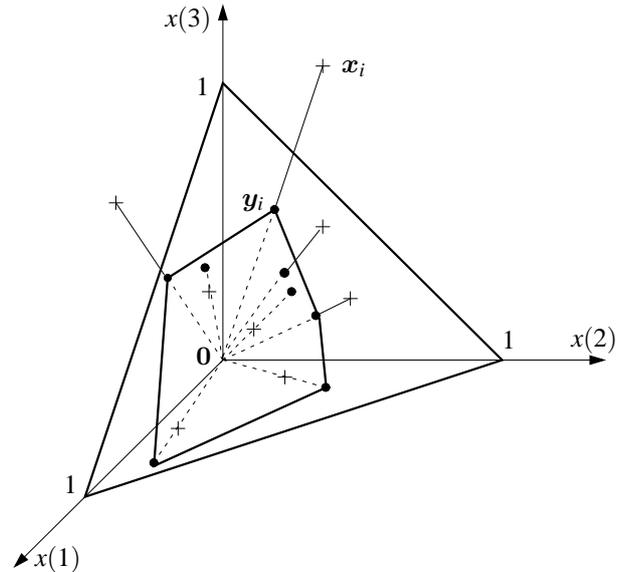


Figure 1: Pixel selection procedure: illustration for  $N = P = 3$ . A mixture signal  $\mathbf{x}_i$  is a vector of  $\mathbb{R}_+^3$  (+). Graphically, the affine projection  $\mathbf{y}_i$  of  $\mathbf{x}_i$  is defined as the intersection of line  $(\mathbf{0}, \mathbf{x}_i)$  with the plane of equation  $x(1) + x(2) + x(3) = 1$ . The projection  $\mathbf{y}_i$  is represented with a bullet (•). The pixel selection procedure consists in projecting all the data  $\mathbf{x}_i$  and then computing the  $(N - 1)$ -dimensional convex hull of the set of points  $\mathbf{y}_i$ . When  $N = 3$ , this convex hull is the 2D polygon represented in plain line.

In this equation,  $\mathbf{X}_e$  is of size  $m \times N$  with  $m \ll M$ ,  $\mathbf{A}_e$  is of size  $m \times P$  and  $\mathbf{S}$  remains of size  $P \times N$ .

The key issue is to keep the available information regarding the sources. This can be done by searching for the boundary of the convex cone  $\text{vect}^+(\mathbf{X})$  (also called the *conical hull*), which fully describes the “extreme” mixtures  $\mathbf{x}_i$ . Selecting the pixels  $i$  such that  $\mathbf{x}_i$  lays on the conical hull also facilitates the further source separation (4) because the size of the data is significantly reduced.

## 3. CONVEX CONE ANALYSIS FOR PIXEL SELECTION

In this section, we present our algorithm whose goal is to search for the conical hull, *i.e.*, the minimal family of mixtures whose positive span is equal to the convex cone  $\text{vect}^+(\mathbf{X})$ . The selected pixels are finally the corresponding indices. For the sake of simplicity and for illustration purpose, we start by the case where  $P = N$ .

### 3.1 Case where $P = N$

Given the mixtures  $\mathbf{x}_1, \dots, \mathbf{x}_M \in \mathbb{R}_+^N$  (see Fig. 1 for a representation of  $\mathbb{R}_+^3$ ), the method aims at researching the conical hull  $\text{vect}^+(\mathbf{X})$ , defined as the minimal subset  $\mathbf{X}_e = [\mathbf{x}_{\sigma_1}, \mathbf{x}_{\sigma_2}, \dots, \mathbf{x}_{\sigma_m}]^t$  of  $\mathbf{X}$  for which  $\text{vect}^+(\mathbf{X}) = \text{vect}^+(\mathbf{X}_e)$ , with  $\sigma_1, \dots, \sigma_m \in \{1, \dots, M\}$  and  $m \leq M$ .

In order to compute  $\text{vect}^+(\mathbf{X})$ , we use the affine projection whose center is  $\mathbf{0}$  onto the hyperplane of equation  $x(1) + \dots + x(N) = 1$ , where  $x(k)$  refers to the  $k$ -th coordi-

Table 1: Pixel selection procedure.

Input: $M$ mixtures stored as vectors $\mathbf{x}_1, \dots, \mathbf{x}_M$ of $\mathbb{R}_+^N$ .
Input: value of $P \approx \text{rank}[\mathbf{X}]$ .
<b>[Affine projection]</b>
For $i = 1, \dots, M$ ,
Compute the affine projection $\mathbf{y}_i$ of $\mathbf{x}_i$ using (5).
End For.
<b>[Dimensionality reduction from <math>N</math> to <math>P - 1</math>]</b>
Perform PCA of the set of $M$ points $\mathbf{y}_i \in \mathbb{R}_+^N$ .
Compute the new coordinates $\mathbf{c}_i \in \mathbb{R}^{P-1}$ of $\mathbf{y}_i$ with respect to the main $P - 1$ principal vectors.
<b>[Convex hull computation]</b>
Compute the convex hull of the points $\mathbf{c}_i, i = 1, \dots, M$ .
Store the output pixel indices as $\sigma_1, \dots, \sigma_m \in \{1, \dots, M\}$ .

nate of  $\mathbf{x}$ . The affine projection  $\mathbf{y}_i$  of a vector  $\mathbf{x}_i$  reads

$$\mathbf{y}_i = \frac{1}{\sum_{k=1}^N x_i(k)} \mathbf{x}_i. \quad (5)$$

It is easy to see that the conical hull  $\text{vect}^+(\mathbf{X})$  is supported by the origin  $\mathbf{0}$  and the *convex hull* of the set of points  $\{\mathbf{y}_i, i = 1, \dots, M\}$  which all lay in an  $(N - 1)$ -dimensional space (see figure 1). Their convex hull is thus an  $(N - 1)$ -dimensional polyhedron whose vertices are denoted by  $\mathbf{y}_{\sigma_1}, \dots, \mathbf{y}_{\sigma_m}$ . In other words, we have:

$$\text{vect}^+(\mathbf{X}) = \text{vect}^+(\mathbf{x}_{\sigma_1}, \dots, \mathbf{x}_{\sigma_m}) = \text{vect}^+(\mathbf{y}_{\sigma_1}, \dots, \mathbf{y}_{\sigma_m}). \quad (6)$$

The approach is illustrated on Fig. 1 in the case where  $N = 3$ : the mixtures  $\mathbf{x}_i$  are vectors of  $\mathbb{R}_+^3$  and their projections  $\mathbf{y}_i$  lay inside a common 2D affine plane.

In brief, the pixel selection procedure consists in computing the affine projection  $\mathbf{y}_i$  of each vector  $\mathbf{x}_i$ , and then the convex hull of the set of points  $\{\mathbf{y}_i, i = 1, \dots, M\}$ . In order to simplify the convex hull computation, and because the points  $\mathbf{y}_i$  all lay inside an affine hyperplane of dimension  $N - 1$ , we choose to describe them by  $(N - 1)$ -dimensional vectors. This can be done in a very simple manner by processing the Principal Component Analysis (PCA) of  $\{\mathbf{y}_1, \dots, \mathbf{y}_M\}$  and by keeping the  $N - 1$  main principal components.

### 3.2 Case where $P < N$

Let us extend the conical hull procedure described above. Because  $P$  is assumed to be the rank of  $\mathbf{X}$ , the mixtures  $\mathbf{x}_i$  are vectors laying inside a  $P$ -dimensional subspace of  $\mathbb{R}^N$ , thus their projections onto the affine plane of equation  $x(1) + \dots + x(N) = 1$  lay inside a  $(P - 1)$ -dimensional affine subspace. For this reason, the strategy described above remains valid, the only adaptation being to further reduce the dimension of the projected mixtures  $\mathbf{y}_i$  by keeping the main  $P - 1$  principal components. The procedure is finally summarized in Table 1.

### 3.3 Practical utilization of the pixel selection procedure

**Choice of  $P$  and implementation issues.** The value of  $P \approx \text{rank}[\mathbf{X}]$  can be chosen easily if some prior knowledge

on the sources is available, or else, by performing a singular value decomposition (SVD) of the raw data  $\mathbf{X}$ , and by searching for a significant gap between each pair of consecutive singular values.

When  $P$  is large, the computation burden of the pixel selection procedure dramatically increases, because a  $(P - 1)$ -dimensional convex hull computation is required. In this case, we propose an adaptation, which is approximate but significantly faster. It is based on the following remark: the convex hull of the orthogonal projections of a set of  $(P - 1)$ -dimensional vectors  $\mathbf{y}_i$  onto a given 2D space yields a number of indices  $\sigma_j$  such that for all  $j$ ,  $\mathbf{y}_{\sigma_j}$  necessarily belongs to the convex hull of  $\{\mathbf{y}_i, i = 1, \dots, M\}$ . Consequently, we propose to store the principal vectors  $\xi_k$  related to the PCA of the set of points  $\mathbf{y}_i$ , and, repeatedly, to consider each pair of vectors  $(\xi_k, \xi_l)$  (or the main vectors only), and then to project orthogonally the set of vectors  $\mathbf{y}_i$  onto the plane defined by  $\xi_k$  and  $\xi_l$ . The repeated 2D convex hull computations do not yield the exact conical hull  $\text{vect}^+(\mathbf{X})$  but a representative subset of pixels  $\mathbf{x}_i$  approximating  $\text{vect}^+(\mathbf{X})$ .

**Choice of the source separation algorithm.** Once the sequence of pixels  $\sigma_1, \dots, \sigma_m \in \{1, \dots, M\}$  is selected, we gather the  $m$  corresponding mixtures into matrix  $\mathbf{X}_e = [\mathbf{x}_{\sigma_1}, \mathbf{x}_{\sigma_2}, \dots, \mathbf{x}_{\sigma_m}]^t$ , and we solve (4) using a blind positive source separation algorithm, whose execution takes a reduced time since the number of mixtures is limited. In general, we cannot guarantee the identifiability of (4), except if the unicity conditions of [10, 11] are fulfilled. Since it is quite complex to compute this measure of unicity, we assume, by default, that the decomposition is not unique, and we choose the Bayesian Positive Source Separation (BPSS) algorithm of [4], which imposes sparsity of the source samples and/or of their weights in order to restrict the range of solutions to (4). This algorithm relies on Monte Carlo Markov Chain sampling of the sources and the mixture coefficients. Once the sources  $\mathbf{S}$  have been found, the remaining task is to estimate the mixture matrix  $\mathbf{A}$  (*i.e.*, the mixture coefficients for the *whole*  $M$  original pixels) from  $\mathbf{X}$  and  $\mathbf{S}$ . This task involves the minimization of the least-square error  $\|\mathbf{X} - \mathbf{A}\mathbf{S}\|^2$  with respect to  $\mathbf{A}$  under the constraint  $\mathbf{A} \geq \mathbf{0}$ .

## 4. APPLICATION TO REAL DATA

### 4.1 Imaging of bacterial sensors

The application context is the study of genetically engineered bacteria expressing optically active reporter molecules in response to chemical or environmental effectors. Bacterial cells have the capacity to act as whole-cell biosensors and display great potential as living transducers in sensing applications [9]. Most popular biosensors are produced by inserting into an appropriate bacterial host, *gfp* and *dsred* genes that code for a green fluorescent protein (GFP) and a Discosoma red fluorescent protein (DsRed), respectively. Because the expression of these fluorescent proteins can be assayed in individual cells by non-destructive means such as confocal laser scanning microscopy, engineered bacteria are attractive biosensors and are increasingly being employed for *in situ* studies in microbial ecology [12, 13].

In this work, a representative strain of the ubiquitous bacterial genus *Pseudomonas* (*P. putida*, KT2440) was genetically engineered to design a GFP-based biosensor (*gfp*) that

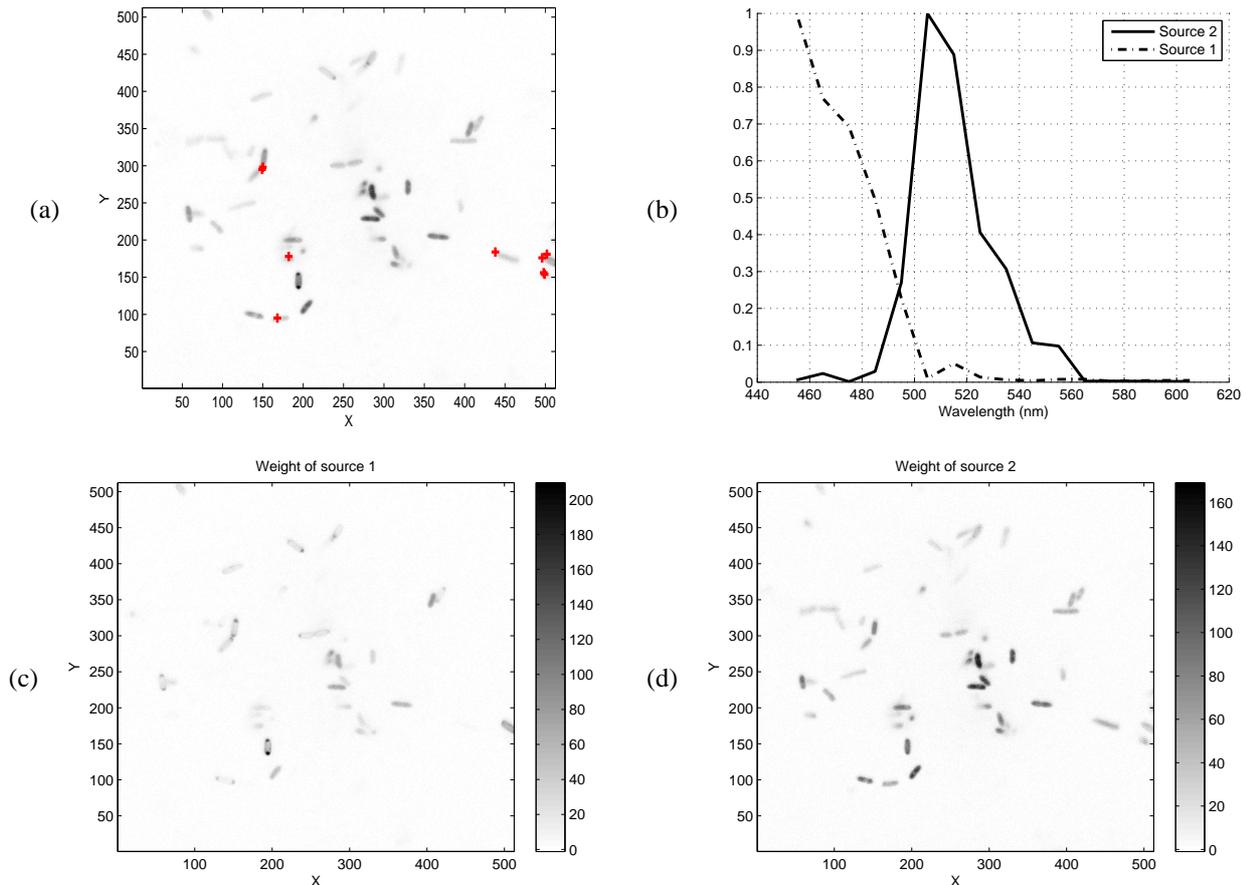


Figure 2: Bacterial cell imaging. The real data are 16 hyperspectral images of a set of genetically engineered bacteria (with *gfp* and *dsr* genes). The excitation beam was set to 405 nm while the emission signal ranging from 450 to 610 nm was recorded in steps of 10 nm. (a) Sum of the 16 hyperspectral images. The 9 pixels selected by setting the number of sources to  $P = 3$ , are superimposed with a (red) plus. (b) The two sources (wild-type strain *wt* alone and *gfp* gene) yielded by the BPSS algorithm ( $P = 2$ ) from the 9 data mixtures. (c,d) Weights of both sources in the pixels, displayed as two 2D images.

responds in a dose-dependent manner to toxic metal exposure. The biosensor strain was also engineered to constitutively express DsRed (*dsr*). Immobilized bacteria cells with *dsr* and *gfp* genes were examined using a Nikon inverted microscope (Eclipse TE2000-U) equipped with a Biorad confocal scan head (Radiance 2100 Rainbow). Fluorescence spectra were acquired pixel by pixel ( $512 \times 512$  pixels, of size  $100 \times 100 \text{ nm}^2$ ) and sequentially (16 wavelengths) in a focus plane (single cell layer). The excitation beam was provided by a blue laser diode (405 nm) while the emission signal ranging from 450 to 610 nm was recorded in steps of 10 nm.

#### 4.2 Hyperspectral unmixing

The hyperspectral signals  $x_i$  are thus of size  $N = 16$  (i.e., the number of wavelengths) and their number is  $M = 512^2$ . Each source is characterizing a “pure” component, i.e., the wild-type strain (*wt*) and the *gfp* and *dsr* genes, respectively. For this reason, we first set the number of sources to  $P = 3$ . In Fig. 2 (a), the image equal to the sum of the 16 hyperspectral images (in which the gray level of the  $i$ -th pixel is equal to  $\sum_{j=1}^N x_i(j)$ ) is shown. It provides a global view of all the

bacteria cells together. The pixel selection algorithm yields 9 mixtures among the  $512^2$  observations (see Fig. 2 (a)).

We first ran the BPSS algorithm with  $P = 3$  sources, but in output, two sources among the three were almost identical and shared similar weights. Thus, we concluded that only two sources are detectable from the data. Fig. 2 (b) shows the sources yielded by BPSS with  $P = 2$ , based on the same 9 mixtures. Here, it is not worth running the pixel selection algorithm with  $P = 2$  as it necessarily yields only 2 mixtures. We rather consider the 9 mixtures shown on Fig. 2 (a), since the BPSS algorithm remains fast with such a limited number of entries. Both sources can be interpreted as the wild-type (*wt*) and the *gfp* spectra, respectively, and the weight images of Fig. 2 (c,d) show their respective response with respect to the spatial dimensions  $X$  and  $Y$ .

#### 4.3 Interpretation of the results

Although the exact expected values of the sources are not known *a priori*, we can analyze and comment the source separation results by comparing them with the spectra measured by another modality of spectroscopy. We recorded average population spectra, referred to as *bulk fluorescence spec-*

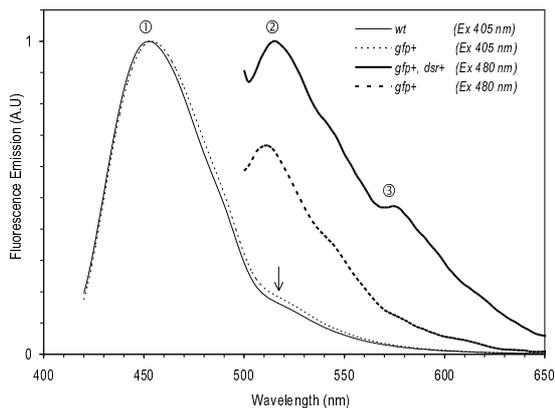


Figure 3: Bulk emission fluorescence spectra (excitation wavelengths:  $405 \pm 5$  nm and  $480 \pm 10$  nm) obtained from cell suspensions of wild-type (*wt*) and engineered strains containing reporter genes (*gfp+* and/or *dsr+*). Zone 1: broad autofluorescence band (440-470 nm); zone 2: principal *gfp* emission band (515 nm); zone 3: predominant *dsr* emission band (575 nm). Note that the *gfp* band is weakly detectable (arrow) in the *gfp+* spectrum acquired with the excitation beam set to 405 nm.

*tra*, using a SAFAS Xenius FLX spectrofluorometer with a xenon lamp excitation source. These average population spectra were obtained from suspensions of several millions of biosensor cells. We employed two excitation wavelengths at 405 and 480 nm while the emission wavelength varied from 420 to 650 nm in steps of 2 nm. The normalized emission spectra obtained for four cell suspensions of wild-type (*wt*) and engineered strains (*gfp+*, *dsr+*) are shown in Fig. 3.

The wild-type strain spectrum can be associated with our first source (Fig. 2 (b)). The predominant bands of wavelengths are indeed in good agreement (see Fig. 3, zone 1), although that of the bulk spectrum is wider due to the very large number of cells in the suspension, from which the average spectrum was obtained. Again, the predominant band (zone 2) of the *gfp+* spectra of Fig. 3 (*wt* and *gfp* together) is in good agreement with the second source of Fig. 2 representing the GFP protein alone. The accurate identification of the DsRed protein is more tricky, due to the very limited weight of *dsr* in the mixtures. Actually, this source is considered as noise when unmixing the hyperspectral images.

## 5. CONCLUSIONS

The proposed algorithm is efficient to select a limited number of relevant pixels in the hyperspectral mixtures. For the real data relative to genetically engineered bacterial cells, the further use of a positive source separation algorithm yields sources which are realistic with respect to our knowledge of the theoretical spectra of wild-type and engineered strains.

On the signal processing viewpoint, future work will be dedicated to the extension of the method by taking into account the neighborhood between pixels in the hyperspectral images. Indeed, the drawback of the pixel selection procedure is that the selected pixels are generally not neighbors on the image grid. This forbids the use of a source separation algorithm utilizing a spatial regularization on the weight coefficients (e.g., the weight of each source is constrained to

be a piecewise constant image). An extended pixel selection procedure would favor the joint selection of neighboring pixels in different zones of the space domain. In the application viewpoint, the proposed algorithm offers new insight in the analysis of living bacteria activities in complex environments, in which the background is unpredictable and the natural fluorescence of cells cannot be controlled. The key difficulties include the wide spectral overlapping between the biosensor emissions, the variability in the cell fitness and obviously, the impressive mass of data to compute [14].

## REFERENCES

- [1] R. Tauler, B. R. Kowalski, and S. Fleming, "Multivariate curve resolution applied to spectral data from multiple runs of an industrial process", *Analyt. Chem.*, vol. 65, no. 15, pp. 2040–2047, 1993.
- [2] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization", *Nature*, vol. 401, pp. 788–791, Oct. 1999.
- [3] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints", *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.
- [4] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, "Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling", *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4133–4145, Nov. 2006.
- [5] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data", in *Summaries of the Fifth Annual JPL Airborne Earth Science Workshop*, R. O. Green, Ed., Pasadena, CA, 1995, pp. 23–26.
- [6] C.-I. Chang and A. Plaza, "A fast iterative algorithm for implementation of pixel purity index", *IEEE Trans. Geosci. Remote Sensing*, vol. 3, no. 1, pp. 63–67, Jan. 2006.
- [7] A. Ifarraguerri and C.-I. Chang, "Multispectral and hyperspectral image analysis with convex cones", *IEEE Trans. Geosci. Remote Sensing*, vol. 37, no. 2, pp. 756–770, 1999.
- [8] S. Moussaoui, H. Hauksdóttir, F. Schmidt, C. Jutten, J. Chanussot, D. Brie, S. Douté, and J. A. Benediktsson, "On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation", *Neurocomputing*, vol. 71, no. 10–12, pp. 2194–2208, June 2008.
- [9] M.-A. Dollard and P. Billard, "Whole-cell bacterial sensors for the monitoring of phosphate bioavailability", *Journal of Microbiological Methods*, vol. 55, no. 1, pp. 221–229, 2003.
- [10] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?", *Advances in Neural Information Processing Systems*, vol. 16, pp. 1141–1148, 2004.
- [11] H. Laurberg, M. G. Christensen, M. D. Plumbley, L. K. Hansen, and S. H. Jensen, "Theorems on positive data: On the uniqueness of NMF", *Computational Intelligence and Neuroscience*, vol. 2008, pp. 1–9, Mar. 2008.
- [12] J. H. J. Leveau and S. E. Lindow, "Bioreporters in microbial ecology", *Current Opinion in Microbiology*, vol. 5, no. 3, pp. 259–265, June 2002.
- [13] E. Larrainzar, F. O'Gara, and J. P. Morrissey, "Applications of autofluorescent proteins for *in situ* studies in microbial ecology", *Annu. Rev. of Microbiol.*, vol. 59, pp. 257–277, 2005.
- [14] R. Tecon and J. R. Van der Meer, "Information from single-cell bacterial biosensors: what is it good for?", *Current Opinion in Microbiology*, vol. 17, no. 1, pp. 4–10, Feb. 2006.