A Genetic Algorithm for Blind Source Separation of Spectral Data

WADY NAANAA* AND JEAN-MARC NUZILLARD**
*URIASIS
Faculty of Sciences
5000 Monastir
TUNISIA
naanaa.wady@planet.tn

**Pharmacognosy Laboratory
University of Reims
Moulin de la Housse
BP 1039, 51687 Reims Cedex 2
FRANCE
jm.nuzillard@univ-reims.fr

Abstract: - Blind source separation (BSS) consists in processing a set of observed mixed signals to separate them into a set of unobserved components. Various approaches have been employed to solve BSS problems assuming that the source signals are mutually uncorrelated (or orthogonal). However, in many real-life problems, signal orthogonality is not guaranteed.
This paper presents a new approach to BSS that can be applied to positive and partially correlated signals. The BSS problem is transformed into a combinatorial optimisation problem and solved by means of a Genetic Algorithm (GA). An algorithm implementing the approach has been proposed and successfully applied to an actual example from Nuclear Magnetic Resonance (NMR) spectroscopy.

Key-Words: - blind source separation, genetic algorithms, nuclear magnetic resonance spectroscopy

1 Introduction

Blind source separation (BSS) consists in retrieving n unknown source signals from m of their linear mixtures, despite the lack of information about the mixing process. Standard BSS problems may be defined by the following equation:

$$X = A.S + N \tag{1}$$

where X is an $m \times p$ matrix representing the detected signals, A is an $m \times n$ mixing matrix whose elements are the unknown mixing coefficients, S is an $n \times p$ matrix representing the unknown source signals, and N is the sensor noise. The mixtures, sources and noise are defined as sampled functions of an acquisition variable that may be time, frequency, position, wave numbers,... depending on the nature of the physical process under investigation. A full identification of A and S is impossible because the sources can be permuted and scaled, provided that the column of A are transformed accordingly. If P is a permutation matrix and Λ an invertible diagonal matrix then $A.S = (AP\Lambda).(\Lambda^{-1}P^{-1}S)$. The matrices S and $\Lambda^{-1}P^{-1}S$ are said to be equivalent in the sense of BSS.

There has been considerable interest in solving BSS problems because they occur in several areas such as speech recognition, data communication, medical science and analytical chemistry [1, 2, 8, 9]. The existing methods for solving BSS problems assume that the source signals are statistically independent. This primary constraint imposes source orthogonality. Consequently, these methods are not appropriate for non-orthogonal signals.

This paper presents a new approach to blind source separation that can be applied to positive

and partially correlated signals. The BSS problem is transformed into a combinatorial optimisation problem and solved by means of a procedure based on a genetic algorithm (GA). The use of GA to solve BSS problem has been made possible by introducing a new optimisation criterion: the *Normal Gram Determinant*, which is used to determining the coefficients of the mixing process.

The paper is organised as follows. Section 2 is devoted to the theoretical basis of the method. In section 3, the BSS problem is encoded in the terms of GAs. We then provide in section 4 the proposed BSS algorithm. Section 5 describes the experiment and an application of the method to an actual example from Nuclear Magnetic Resonance (NMR) spectroscopy. Section 6 is a brief conclusion. The following notations will be used throughout the paper. The notation A_i (resp. A^j) will be used to designate the i^{th} row (resp. the j^{th} column) of matrix A. An (i,j) entry of a matrix A will be denoted by a_{ij}.

2 The method

In order to establish the theoretical basis of the method we will consider the noiseless case:

$$X = A.S \tag{2}$$

It will be assumed without loss of generality that data are given in spectral form. It follows that each column index in S and X corresponds to a frequency value and each row contains data inherent to a particular spectrum. It is also assumed that the number of linearly independent mixtures is greater of equal to the number of sources and that the rank of the mixing matrix is equal to n, i.e. $\rho(A) = n$.

In other respects, many physical analysis provide nonnegative signals. The signals we are concerned with, (NMR spectral signals), belongs to this category of signals. Furthermore, due to their physical origin, the mixing coefficients involved in the NMR experiment are nonnegative. The source orthogonality requirement is therefore replaced by the positivity constraint on A and S and by the following assumption: For each source, there is at least one value of the acquisition variable for which this source presents a non-null response, to the exclusion of all other sources. Such sources are said to be partially uncorrelated (or partially orthogonal). More formally, the source matrix S is assumed to satisfy the following condition:

Assumption 1 *for each* $i \in \{1,2,\ldots,n\}$ *there exists an* $j_i \in \{1,2,\ldots,p\}$ *such that* $s_{ij_i} > 0$ *and* $s_{kj_i} = 0, (k = 1,\ldots,i-1,i+1,\ldots,n)$.

Clearly, this assumption does not require orthogonal source signals. Indeed, orthogonality is required only for the indices $j_i, (i = 1,\ldots,n)$.
An other way to write Eq. 2 is the following:

$$X^j = \sum_{k=1}^{n} s_{kj} A^k, \quad (j = 1,\ldots,p) \tag{3}$$

For the particular subscripts $j_i, (i = 1,\ldots,n)$ described in assumption 1, Eq. 3 becomes

$$X^{j_i} = s_{ij_i} A^i, \quad (i = 1,\ldots,n) \tag{4}$$

since by assumption 1, s_{kj_i} is nonzero only if $k = i$. Equation 4 means that every column of A is collinear to at least one column of X (column X^{j_i}). Let $\hat{A}$ be the submatrix of X consisting of n columns each of which is collinear to a particular column of A, (i.e., the columns of X with superscripts $j_i, i = 1,\ldots,n$). Then, by Eq. 4, we have $\hat{A} = AP\Lambda$, where P is a $n \times n$ permutation matrix and Λ is an $n \times n$ diagonal real nonnegative matrix. It follows from Eq. 2 that:

$$\Lambda^{-1}P^{-1}S = \hat{A}^{\#}X$$

where the symbol $^{\#}$ denotes the Moore-Penrose pseudo-inverse [1]. For short, $\Lambda^{-1}P^{-1}S$ will be denoted by $\hat{S}$. $\hat{A}^{\#}X$ represents an acceptable solution to the separation problem since the intensity of the source signals as well as their arrangement in S are not relevant. Since X is known, in order to identify $\hat{S}$, we have to determine $\hat{A}$ which is a $m \times n$ submatrix of X. A generate and test algorithm for determining $\hat{A}$ must consider the different submatrices of X, until a submatrix such that $\hat{A}^{\#}X$ is nonnegative is found. Such an algorithm

[1] $A^{\#} = (A.A^t)^{-1}.A^t$

has two drawbacks. On the one hand, computing the matrix product $\hat{A}^{\#}X$ is time consuming because the number of columns in X, (which is equal to the number of frequency values), is often too large. On the other hand, the number of submatrices to be tested may reach $\binom{p}{n}$, since we have to consider all different combinations of n columns among the p columns of X.

To avoid computing the product $\hat{A}^{\#}X$, we remind the following definition first: the Gram determinant of a rectangular matrix M is defined by $G(M) = det({}^tM.M)$. In the following, the notation $\underline{M}$ will designate the matrix obtained from M by multiplying each column M^j by $\|M^j\|^{-1}$. The expression $\underline{G}(M)$ will designate $G(\underline{M})$ and will be called the *Normal Gram Determinant* of M.

To determine $\hat{A}$, we state the following proposition [2].

Proposition 2 *Let $Z \in M_{m,p}(\mathbb{R}^+)$ and $Y \in M_{n,p}(\mathbb{R}^+)$ such that $n \leq m$. Then, if there exists a $m \times n$ submatrix B of Z such that $Z = BY$ and $\rho(B) = n$ then, for any $m \times n$ submatrix B' of Z, we have $\underline{G}(B') \leq \underline{G}(B)$, with equality if and only if $B' = BP\Lambda$, where P is a permutation and Λ is diagonal.*

Proposition 2 may be applied in the context of BSS since the conditions which hangs over the matrices Z, B and Y are satisfied by $X, \hat{A}$ and $\hat{S}$ respectively. Indeed, $X, \hat{A}$ and $\hat{S}$ are nonnegative, they satisfy $X = \hat{A}\hat{S}$ and $\hat{A}$ is a submatrix of X. By proposition 2, $\hat{A}$ is necessary one of the submatrices of X that have a maximal Normal Gram Determinant value. This latter may therefore be viewed as a criterion to identify $\hat{A}$.

To overcome the second weak point picked out above, we propose to use a GA-based search so as to consider only the promising submatrices of X, i.e. those that own the highest $\underline{G}(\hat{A})$ values.

3 Applying GA to BSS

The genetic algorithm (GA) is a stochastic optimisation algorithm that was originally motivated by the mechanism of natural selection [3, 7]. Basically, a GA is a search procedure inspired by the survival of the fittest principle of natural evolution. The main elements of a population of individuals are represented by feature-encoding chromosomes. The chromosomes are composed of genes. In general, the GA acts on a population of individuals using three generic operators: selection, crossover and mutation. The process of selection copies chromosomes from one generation to the next, based on their fitness value. Crossover crosses pairs of randomly selected chromosomes in order to produce new offsprings. The reproduced chromosomes (children) contain materials from both of the original pairs (parents). Mutation acts on a simple chromosome by randomly changing some genes. During the evolution process, selection, crossover and mutation are repeatedly applied until a convergence criterion is met.

To apply a GA-based algorithm to BSS problem, we must consider the following points:

3.1 Encoding and fitness function

The GA-tool is devoted to achieving the main task of the separation process: determining $\hat{A}$. To this end, an individual must encode in some way an $m \times n$ matrix. If we take into account the fact that $\hat{A}$ is composed of n columns taken among the p columns X, then $\hat{A}$ may be fully identified by the n indices of its columns in X. Thus, an individual may be encoded by a vector consisting of n integers ranging in $\{1, \ldots, p\}$.

As it is suggested by proposition 2, the Normal Gram Determinant provides a mean to identify a submatrix $\hat{A}$ of X such that $\hat{A} = AP\Lambda$. The Normal Gram Determinant $\underline{G}(\hat{A})$ is therefore used as the fitness function. From the practical point of view, the computation of $\underline{G}(\hat{A})$ is not time consuming when compared to the computation of $A^{\#}X$ because $m, n \ll p$.

3.2 The initial population

Initialisation of a population to provide the genetic algorithm a starting point is usually done by generating random strings from the encoding alphabet. Nonetheless, for the problem treated

[2] The proof of proposition 2 cannot be included in this short paper.

herein, individuals involving more that one time a same column of X have zero fitness value, since the encoded matrix is singular in this case. In order to avoid such individuals, the initial population is generated as follows:

$$x_i = [i\%p, (i+1)\%p, \ldots, (i+n-1)\%p], \quad (i = 0, \ldots, p-1)$$

This population is composed of p individuals. In each individual, a same column cannot appear more than once since the integers $i\%p, \ldots, (i+n-1)\%p$ are all different when $1 < n \leq p$, (the symbol % denotes the modulo operator). Note that each column appears exactly n times in the whole population. Thus, at the beginning, all the columns of X are given the same chance to be present in the final solution.

3.3 Selection

The ranking selection scheme is used as the selection mechanism. Ranking methods assign a probability p_i to each individual i based on its fitness value. The probability p_i is defined as follows:

$$p_i = \frac{q(1-q)^{r-1}}{1-(1-q)^p}$$

where:

- q = the probability of selecting the best individual (this parameter is equal to 0.08)
- r = the rank of the individual, where 1 is the best rank
- p = the population size

A series of random numbers is generated and compared against the cumulative probability, $C_i = \sum_{k=1}^{i} p_k$. An individual i is selected and copied into the new population if $C_{i-1} < m < C_i$, where m is a random number in the range $[0, 1]$. The algorithm selects individuals in this manner until the entire population of next generation has been produced. The number of individuals surviving to selection remains constant, but in counterpart, the number of trials for the fittest chromosomes increases to the detriment of the number of trials for the less adapted chromosomes.

3.4 Crossover

Crossover operator takes two individuals which have survived to selection and produces two new individuals. We use one-point crossover mechanism. In one-point crossover, a crossover point is randomly selected. The portions of the two chromosomes beyond this cut-off point to the right are to be exchanged to form the new individuals. Precautions must be taken prior to applying crossover operator, because crossover may yield children that have zero fitness values, even though the fitness values of the parents are high. Indeed, a crossover involving parents that do not have duplicated columns may produce children with duplicated columns. To prevent this occurs, we proceed to a simple gene rearrangement before crossover in such a way that if a column is present in both of the parents which are going to reproduce then the position of this column should be the same. By so doing, we avoid children with duplicated columns.

3.5 Mutation

Mutation takes place immediately after crossover. The random alteration of genes during mutation may generate an individual with duplicated column. Therefore, the selected column for mutation must not be replaced by an column which is already present in the chromosome. A random integer in the range $\{1, 2, \ldots, p\}$ is repeatedly picked-up until getting an integer which is not already present in the chromosome. By so doing one is sure not to introduce a duplicated column during mutation.

4 The algorithm

The proposed algorithm for solving BSS problems is the following:

Algorithm 3 EvolBss*(X)*

1. *discard from X the columns X^j such that $\|X^j\| \leq \epsilon$. Note $\check{X}$ the resulting matrix.*
2. *form matrix $\hat{X}$ consisting of all the mutually non-collinear columns of $\check{X}$.*

3. $n = \rho(\hat{X})$

4. *apply algorithm* GA(n,$\hat{X}$) *and get the best individual* $\hat{A}$.

5. *replace each column in* $\hat{A}$ *by the average of all columns in* $\check{X}$ *that are collinear to it. Note* $\check{A}$ *the resulting matrix.*

6. *compute the Moore-Penrose pseudo-inverse* $\check{A}^{\#}$ *of* $\check{A}$ *and then the estimate of* S: $\check{S} = \check{A}^{\#}.X$

Although theoretically not compulsory, steps 1 and 2 are of great interest. Indeed, during these two steps, many columns of the input matrix X are discarded leading to a significant reduction of the search space. More precisely, step 1 aims to discarding from X the columns consisting of low-valued entries due to noise from X. The constant ϵ is a tolerance computed as the product of the standard deviation σ of the noise by an arbitrary coefficient k_1. In Step 2, we exploit the fact that a submatrix containing collinear columns has a zero fitness value. Then collinear columns are discarded in such a way as to keep as few collinear column as possible. Two column vectors are assumed to be collinear if their angle θ does not exceed a tolerance computed as the product of an other arbitrary constant k_2 by the standard deviation $\sigma(\theta)$ of the angles. The number of sources n is, in practice, determined by computing the rank of $\hat{X}$ (see step 3). In step 4, $\hat{A}$ is computed through the use of the genetic algorithm described in section 3. We used a population consisting of $\hat{p}$ individuals, where $\hat{p}$ is the number of columns in $\hat{X}$ (see algorithm 3). We used a adapted version of the GA implementation described in [4]. A possible way of reducing the influence of noise present in $\hat{A}$ is to replace each column by the average of the columns of $\check{X}$ that are collinear to it. This is done in step 5. An improvement is expected because the noise samples are uncorrelated. Finally, an estimate of the source matrix S is computed as $\check{S} = \check{A}^{\#}X$ (see step 6).

5 Experiment and results

NMR is a physical phenomenon that exploits the magnetic properties of atomic nuclei. It has found many applications in the fields of chemistry and medical diagnosis. The NMR spectrum of a mixture is ideally the linear combination of the spectra of the mixture components, designated by M_i. With only a single mixture available, the separation of the component spectra requires a way of modulating their intensity in a specific manner. This goal is achieved by the "Pulsed Gradient Spin Echo" (PGSE) method [10].

The mixing coefficients a_{ji} are all positive and represent the scaling factor of the signals from molecules M_i when submitted to gradient G_j. The spectra of the individual compounds that define the matrix S are also positive valued functions. The linear mixing model described by equation 1 is valid because all the data processing steps that produce the mixed spectra involve linear operations. This means that the spectrum of a mixture of systems is truly a linear combination of the spectra of the individual systems. Assumption 1 is met in most practical situations because there is very little chance that a compound does not display at least one characteristic spectral signal. Thus, the signals provided by the NMR PGSE experiment present adequate properties for their separation by the EvolBss algorithm, as described in section 4.

As an example, we consider a sample of a mixture containing two sugars: mannitol and beta-cyclodextrine at a concentration of 20 mmol/L and some partially deuteriated water HOD. The three spectra of figure 1a, b, c are obtained with gradient strength of 2.5, 15 and 25 G/cm. The spectra of pure mannitol, pure beta-cyclodextrine and deuteriated water are depicted in figure 1d, e, f. For the purposes of comparison, we also applied the Fixed Point algorithm [5]. The spectra of figures 2 are those computed by the EvolBss algorithm (a, b, c) and those obtained by the Fixed Point algorithm (d, e, f). The separation process took less that 1 minute by both of the two algorithms. The spectra obtained by the EvolBss algorithm are more akin to the source spectra (figure 1d, e, f).

The difference between the responses of the two algorithms can be observed notably in the frequency zone comprised between 2000 and 2500.

This result can be strengthened by means of the matrix distance defined in [2]. Let A and A' be two invertible matrices. The distance between A and A', $\epsilon(A, A')$ is built on the matrix $D = \underline{A}^{-1}\underline{A}'$ as:

$$\begin{aligned}\epsilon(A, A') &= \sum_i |\sum_j |d_{ij}| - 1|^2 + \\ &\sum_j |\sum_i |d_{ij}| - 1|^2 + \\ &\sum_i |\sum_j |d_{ij}|^2 - 1| + \\ &\sum_j |\sum_i |d_{ij}|^2 - 1|\end{aligned}$$

Comon proved that a value of $\epsilon(A, A')$ close to zero means that A and A' are nearly equivalent in the sense of BSS, i.e., $A' \approx AP\Lambda$. Let A_{eb} and A_{fp} be the mixing matrices, obtained by the EvolBss algorithm and Fixed Point algorithm respectively, and let A be the effective mixing matrix of the BSS problem instance described above. The values $\epsilon(A, A_{fp})$ and $\epsilon(A, A_{eb})$ are equal to 0.7718 and 0.3958 respectively. This means that A_{eb} is closer to the mixing matrix A than A_{fp}.

6 Conclusion

This paper presented a new blind source separation methods that can be applied to positive and partially correlated signals. The strong orthogonality constraint imposed by the existing methods is replaced by a weaker assumption that requires only partial orthogonality. Based on this assumption, the BSS problem is transformed into an optimisation problem which is solved by means of a genetic algorithm. An algorithm implementing the method is proposed and successfully applied to an actual example from NMR spectroscopy. Future work should include the application of the BSS method presented herein to image analysis as it is the case in [6].

References:

[1] Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., Moulines, E., A blind source separation technique using second-order statistics, *IEEE Transaction on signal processing*, Vol. 45, 1997, pp. 434-444.

[2] Comon, P., Independent component analysis, A new concept?, *Signal processing*, Vol. 36, 1994, pp. 287-314.

[3] Goldberg, D., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.

[4] Houck, C. R., Joines, J. A., Kay, M. G., A genetic algorithm for function optimization: A Matlab implementation, North Carolina State University-IE, Technical Report 95-09, 1995.

[5] Hyvärinen, A. 1999. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis, *IEEE Transactions on Neural Networks*, Vol. 10, No. 3, 1999, pp. 626-634.

[6] Kundur, D., Hatzinakos, D., A novel blind deconvolution scheme for image restoration using recursive filtering. *IEEE Transaction on Signal Processing*, Vol. 46, No. 2, 1998, pp. 375-390.

[7] Michalewics, Z., *Genetic algorithms + Data Structures = Evolution Programs*, AI series. Springer-Verlag, New York, 1994.

[8] Nuzillard, D. and Nuzillard, J.-M., Blind source separation to non-orthogonal signals, *1st Workshop on Independent Component Analysis*, Aussois, France, 1999, pp. 25-30.

[9] Pajunen, P., Karhunen, J., Least-squares methods for Blind source separation based on nonlinear PCA, *International Journal of Neural Systems*, Vol. 8, Nos. 5-6, 1997, pp. 601-612.

[10] Price, W. S., Hayamizu, K., Ide, H., Arata, Y., Strategies for Diagnosing and alleviating artifactual attenuation associated with large gradient pulses in PGSE NMR diffusion measurements, *Journal of Magnetic Resonance*, Vol. 139, 1999, pp. 205-212.

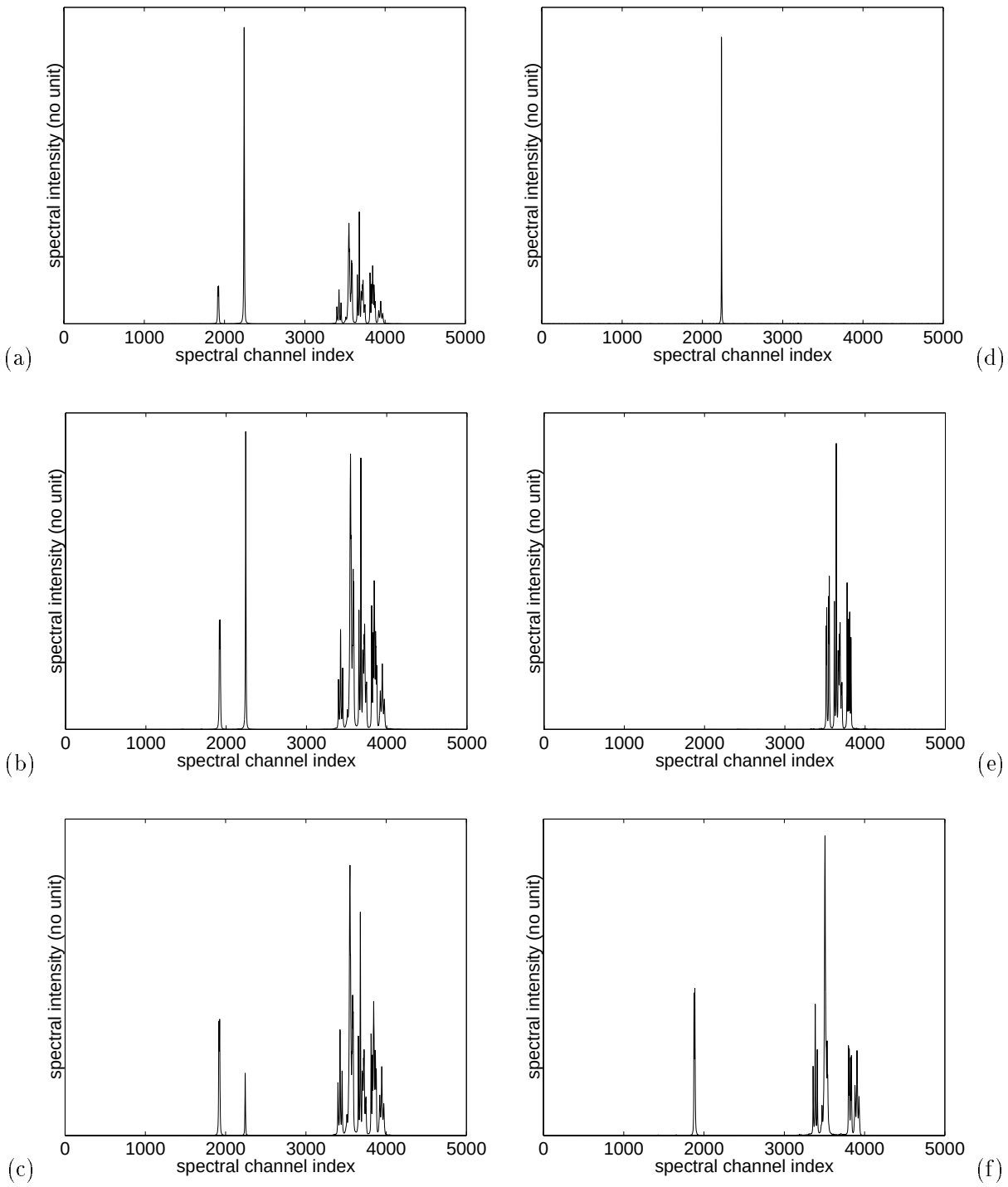

Fig. 1: Signals observed by submitting a sample containing a mixture of deuteriated water, mannitol, and beta-cyclodextrine to three gradient strength. (a-b-c), and effective source signals: Spectrum of deuteriated water, pure mannitol, and pure beta-cyclodextrine. (d-e-f)

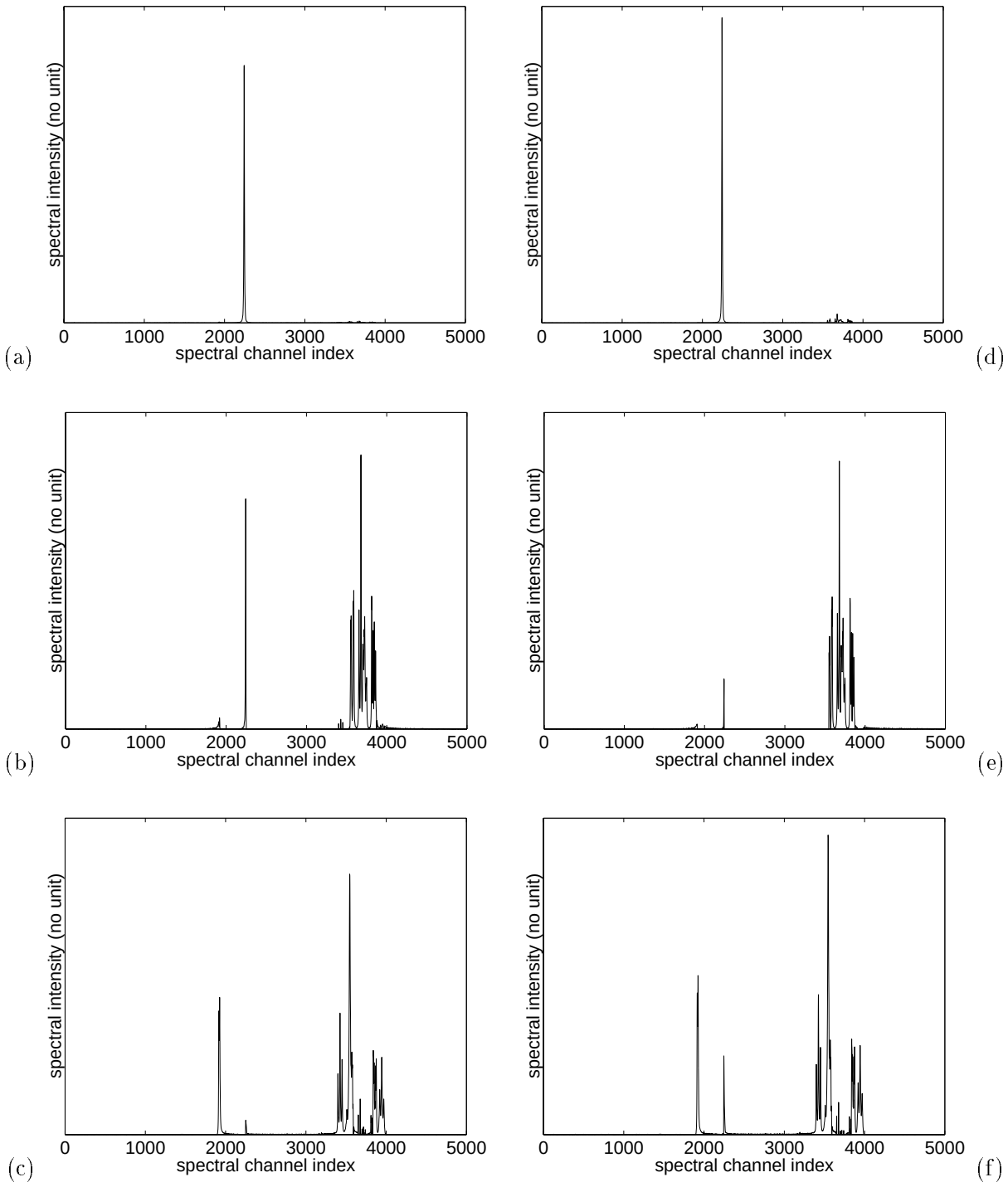

Fig. 2: Source signals computed by the EvolBss algorithm (a, b, c). Source signals computed by the Fixed Point algorithm (d, e, f).