Blind Decomposition of Infrared Spectra Using Flexible Component Analysis

Ivica Kopriva,[*†] *Ivanka Jerić*[‡] *and Andrzej Cichocki*[§]

[†]Division of Laser and Atomic Research and Development

[‡]Division of Organic Chemistry and Biochemistry

Ruđer Bošković Institute, Bijenička cesta 54, HR-10000, Zagreb, Croatia

[§]Laboratory for Advanced Brain Signal Processing

Brain Science Institute, RIKEN 2-1, Hirosawa, Wako-shi, Saitama, 351-0198, Japan

[*]ikopriva@irb.hr; Tel.: +385-1-4571-286; Fax: +385-1-4680-104

Abstract

The paper presents flexible component analysis-based blind decomposition of the mixtures of Fourier transform of infrared spectral (FT-IR) data into pure components, wherein the number of mixtures is less than number of pure components. The novelty of the proposed approach to blind FT-IR spectra decomposition is in use of hierarchical or local alternating least square nonnegative matrix factorization (HALS NMF) method with smoothness and sparseness constraints simultaneously imposed on the pure components. In contrast to many existing blind decomposition methods no *a priori* information about the number of pure components is required. It is estimated from the mixtures using robust data clustering algorithm in the wavelet domain. The HALS NMF method is compared favorably against

three sparse component analysis algorithms on experimental data with the known pure component spectra. Proposed methodology can be implemented as a part of software packages used for the analysis of FT-IR spectra and identification of chemical compounds.

Keywords: Chemometrics, Blind source separation, Flexible component analysis, Sparse component analysis, Wavelet translation, FT-IR spectroscopy.

1. Introduction

Extraction of the pure component spectra from the mixtures of their linear combinations is of great interest in many applications. Classical approach to extraction of the spectra of pure components is to match the mixture's spectra with a library of reference compounds. This approach is ineffective with the accuracy strongly dependent on the library's content of the pure component spectra and can not reflect the variation of the spectral profile due to environmental changes. Alternatives to library matching approach are blind decomposition methods, wherein pure components' spectra are extracted using mixtures spectra only. Blind approaches to pure components spectra extraction have been reported in NMR spectroscopy [1], infrared (IR) [2-4] and near infrared (NIR) spectroscopy [4-6], EPR spectroscopy [7, 8], mass spectrometry [4, 9,1 0] Raman spectroscopy [11, 12] etc. In a majority of blind decomposition schemes independent component analysis (ICA) [13-15] is employed to solve related blind source separation (BSS) problem. ICA assumes that: (*i*) pure components are statistically independent, (*ii*) at most one is normally distributed and (*iii*) number of mixtures is greater than or equal to the unknown number of pure components. The two requirements: to have more linearly independent mixtures than pure components and to have statistically independent pure components seem to be most critical for the success of the BSS approach to blind decomposition of the mixtures spectra into pure components spectra [4, 5, 8, 10]. Statistical independence assumption is certainly not fulfilled in the case of IR spectra [2-6] because they are highly correlated i.e. overlapped. Raw data preprocessing technique by first or second order derivative has been used in FT-IR spectra analysis to reduce level of statistical dependence among pure components, [2-6]. This technique actually belongs to the

generalization of the ICA known as dependent component analysis (DCA), [14, 16-18]. An algorithm for blind decomposition of EPR spectra has been derived in [8] minimizing contrast function that exploits sparseness rather than statistical independence among the pure components. Unfortunately, sparseness criterion can not be used in the case of FT-IR spectra due to high degree of overlap between them, especially in wavelength or wavenumber domain. All discussed blind spectra decomposition methods require the number of mixtures spectra to be equal to or greater than the unknown number of pure components spectra. In a number of real world situations it is however not easy to acquire mixtures spectra with different concentrations of the pure components spectra. In this regard it is desirable property of blind decomposition methods to solve related BSS problem with as few mixtures as possible. Here, we demonstrate flexible component analysis (FCA) approach to blind decomposition of more than two pure components FT-IR spectra from two mixtures only. To solve related underdetermined BSS (uBSS) problem we use recently developed nonnegative matrix factorization (NMF) algorithm that is known as local or hierarchical alternating least squares (HALS) NMF algorithm [19, 20]. Its unique property is to estimate concentration or mixing matrix globally and pure components spectra locally, wherein smoothness and sparseness constraints are simultaneously imposed on the pure components spectra. Unlike majority of the BSS algorithms that assume the number of pure components to be known, proposed approach estimates it from the mixtures spectra in the wavelet domain by means of data clustering algorithm, [21]. Transformation of the mixtures spectra in wavelet domain yields representation that is significantly sparser than in original wavenumber domain. This enables more accurate estimation of the number of pure components spectra, especially due to the fact that used data clustering algorithm requires that pure components spectra are in average sparse in the chosen basis. Comparison of the HALS NMF approach against sparse component analysis (SCA) based approach [22-25] on experimental uBSS problem, which is presented in section 3, yields favorable results. Therefore, it is believed that proposed FCA-based approach to blind extraction of the FT-IR pure components spectra is practically important. The rest of the paper is organized as follows. We introduce data clustering algorithm, SCA and FCA concepts in section 2. Results and discussion of the

experimental comparative performance analysis of the FCA and SCA approaches on two mixtures of IR spectra containing three pure components are given in section 3. Conclusions are presented in section 4.

2. Computational methods

Like many decomposition methods proposed approach is based on static linear mixture model

$$\mathbf{X} = \mathbf{AS} \tag{1}$$

where $\mathbf{X} \in \mathrm{R}_{0+}^{N \times T}$ represents matrix of N measured mixtures spectra across T wavenumbers, $\mathbf{A} \in \mathrm{R}_{0+}^{N \times M}$ represents the matrix of concentration profiles also called the mixing matrix and matrix $\mathbf{S} \in \mathrm{R}_{0+}^{M \times T}$ contains M pure components spectra across T wavenumbers. Due to the nature of the problem all quantities in (1) are nonnegative. As already pointed out, the number of pure components M is in principle unknown although many BSS/ICA algorithms assume that it is either known in advance or can be easily estimated. This does not seem to be true in practice, especially when the BSS problem is underdetermined. Here, we shall treat M as unknown parameter that will be estimated by the clustering algorithm to be described in section 2.1. In addition to estimate the number of pure components used data clustering algorithm also estimates the concentration matrix. This is necessary for the SCA approach described in section 2.2., but is not necessary for HALS NMF approach described in section 2.3. In overall, the BSS problem related to blind FT-IR spectra decomposition consists of: (*i*) estimating the number of pure components spectra; (*ii*) estimating the matrix of the pure components spectra **S**; (*iii*) estimating the concentration matrix **A**. All three tasks are executed using matrix of mixtures spectra **X** only. In addition to that, we allow the number of pure components spectra M to be greater than the number of mixtures spectra N. Hence, blind FT-IR spectra decomposition problem becomes uBSS problem.

2.1. Data clustering

In FT-IR spectra decomposition problem considered in this paper we shall assume that pure components spectra are in average $k=M-1$ sparse in wavelet domain. This implies that at each coordinate in wavelet domain in average only one pure component is active i.e. nonzero. This assumption allows to reduce number of mixtures to $N=2$, hence reducing the computational complexity of to be used data clustering algorithm [21] by reducing dimension of the concentration subspaces, that equals average number of active components, to 1. However, we are aware that it is not realistic to demand that pure components FT-IR spectra do not overlap in any representation domain including wavelet domain used here. That is why we expect that pure components spectra are only in average $k=M-1$ sparse in wavelet domain. Under such assumption the appropriately chosen function, see eq.(3), will effectively cluster data, wherein the number of clusters corresponds with the estimate of the number of pure components M. If the number of coordinates that violates $k=M-1$ sparseness assumption in wavelet domain is relatively large this will influence accuracy of the estimation of the concentration matrix due to the repositioning of the cluster centers. It will not however influence in the same amount the accuracy of the estimation of the number of clusters. Thus, performance of the SCA algorithms that require the estimate of the concentration matrix in order to proceed to the next phase and solve underdetermined system of linear equations will be affected significantly if FT-IR spectra are not sparse enough in the chosen basis. On the other hand proposed FCA approach will be significantly less sensitive to the level of sparseness of the FT-IR spectra because it only requires from the clustering algorithm the estimate of the number of pure components spectra.

Because solution of the BSS problem is generally characterized by scale indeterminacy we shall assume the unit norm constraint (in the sense of ℓ_2 norm) on the columns of the concentration matrix $\mathbf{A}$, i.e., $\left\{\left\|\mathbf{a}_m\right\|_2 = 1\right\}_{m=1}^{M}$. As already pointed out, in this paper we do assume the number of mixtures to be $N=2$. Thus, the normalized mixing vectors $\left\{\mathbf{a}_m\right\}_{m=1}^{M}$ lie in the first quadrant on the unit circle, i.e., they are parameterized as:

$$\mathbf{a}_m = [\cos(\varphi_m)\ \ \sin(\varphi_m)]^{\mathrm{T}} \qquad m = 1,...,M\ , \tag{2}$$

where φ_m represents mixing angle that is confined in the interval [0, π/2]. We do assume that mixtures are transformed into wavelet domain through wavelet transform

$$X_n(a,b) = \tfrac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x_n(t)\psi\left(\frac{t-b}{a}\right)dt \qquad n = 1,...,N$$

where the a and b represent respectively scale (resolution level) and time shift and $\Psi(t)$ represents wavelet function. After extensive experiments we have found out that symmlets with two to eight vanishing moments yield best results in terms of sparseness of $\mathbf{X}(a,b)$. Thus, the results reported in section 3 were obtained with the symmlets with the four vanishing moments. The fact that symmlets performed best is just experimental finding. We have also tried Daubechie's wavelet of different order, Haar wavelet, Morlet wavelet, Mexican hat wavelet, Coiflets and some biorthogonal wavelets. From the sparse representation point of view the key property of the wavelet is to match well the waveform of the particular signal of interest (in this case the FT-IR spectra). It is however very hard to find such a wavelet in case of FT-IR signals. Perhaps, the optimal solution would be to design new wavelet that will reflect better morphological properties of FT-IR data than standard wavelets do. Wavelet transform above can be used either as continuous or as discrete. In the results presented in section 3 we have used discrete shift invariant wavelet transform with the resolution levels corresponding to $a=2^1$ or $a=2^2$. By assuming *1*-dimensional concentration subspaces the clustering algorithm [21] is outlined by the following steps:

1) We remove all data points close to the origin for which applies: $\left\{\left\|\mathbf{x}(a,b_t)\right\|_2 \le \varepsilon\right\}_{t=1}^{T}$, where ε represents some predefined threshold. This corresponds with the case when pure components spectra are close to zero.

2) Normalize to unit ℓ_2 norm remaining data points $\mathbf{x}(a,b_t)$, i.e., $\left\{\mathbf{x}(a,b_t) \leftarrow \mathbf{x}(a,b_t) \big/ \left\|\mathbf{x}(a,b_t)\right\|_2\right\}_{t=1}^{\bar{T}}$, where $\bar{T} \le T$ denotes number of data points that remained after the elimination process in step 1.

3) Calculate function $f(\mathbf{a})$, where $\mathbf{a}$ is defined with (2):

$$f(\mathbf{a}) = \sum_{t=1}^{\bar{T}} \exp\left(-\frac{d^2\left(\mathbf{x}(a,b_t),\mathbf{a}\right)}{2\sigma^2}\right) \tag{3}$$

where $d\left(\mathbf{x}(a,b_t),\mathbf{a}\right) = \sqrt{1-\left(\mathbf{x}(a,b_t)\cdot\mathbf{a}\right)^2}$ and $\left(\mathbf{x}(a,b_t)\cdot\mathbf{a}\right)$ denotes inner product. Parameter σ in (3) is called dispersion. If set to sufficiently small value, in our experiments this turned out to be $\sigma \approx 0.035 \pm 0.007$, the value of the function $f(\mathbf{a})$ will approximately equal the number of data points close to $\mathbf{a}$. Thus by varying mixing angles $0 \le \varphi \le \pi/2$ we effectively cluster data.

4) Number of peaks of the function $f(\mathbf{a})$ corresponds with the estimate of the number of pure components spectra $\hat{M}$. Locations of the peaks correspond with the estimates of the mixing angles $\left\{(\hat{\varphi}_m)\right\}_{m=1}^{\hat{M}}$, i.e., mixing or concentration vectors $\left\{\hat{\mathbf{a}}_m\right\}_{m=1}^{\hat{M}}$, where $\hat{\mathbf{a}}_m$ is given with (2). The hat sign introduced here is used to denote estimate of the related quantity. Hence, at the end of data clustering phase estimates of the number of pure components M and concentration matrix $\mathbf{A}$ are obtained.

2.2. Sparse component analysis

SCA enables to find a possible good approximation of the true solution to an underdetermined system of linear equations subject to sparsity constraints. When in (1) $N<M$, the nullspace of $\mathbf{A}$ is nontrivial, and the inverse problem has many solutions. Therefore, additional constraint such as sparseness between the components of the column vectors $\left\{\mathbf{s}(t)\right\}_{t=1}^{T}$ is necessary. A sparse signal is a signal whose most samples are nearly zero, and just few percent take significant values. Signal that has at least $k \le M$

zero components is called *k*-sparse. The SCA is carried out using one of the two approaches. The first one employs NMF algorithms, wherein mixing matrix **A** and source matrix **S** are estimated simultaneously, usually through ALS methodology, [19, 26]. The second one, referred in [22-26] breaks down uBSS problem into two separate problems: estimation of the concentration matrix **A** and the number of pure component spectra using geometric concept known as data clustering [21-15] and estimation of the matrix of pure components spectra **S** (based on estimated **A**) by solving resulting underdetermined system of linear equations. The last step is carried out as linear programming, [22, 27, 28] ℓ_1-regularized least square problem [29, 30] or ℓ_2-regularized linear problem, [31]. Presuming that concentration matrix **A** and number of pure components spectra *M* are estimated through data clustering phase as well as that pure component spectra are in average *M*-1 sparse they can be estimated by means of linear programming in wavelet domain

$$\hat{\mathbf{z}}(a,b_t) = \arg\min_{\mathbf{z}(a,b_t)} \sum_{m=1}^{\hat{M}} z_m(a,b_t) \quad \text{subject to } \tilde{\mathbf{A}}\mathbf{z}(a,b_t) = \mathbf{x}(a,b_t) \quad \forall\, b_t = 1,...,T$$
$$\mathbf{z}(a,b_t) \geq \mathbf{0} \qquad (4)$$

where $\mathbf{z} = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}$. $\mathbf{u} \in \mathrm{R}_{0+}^{M}, \mathbf{v} \in \mathrm{R}_{0+}^{M}$ are nonnegative dummy vectors used to model source vector $\mathbf{s}(a,b_t)$ that can be both positive and negative, i.e. $\mathbf{s}(a,b_t)=\mathbf{u}-\mathbf{v}$. $\tilde{\mathbf{A}} = \left[\hat{\mathbf{A}} \;\; -\hat{\mathbf{A}}\right]$ is extended mixing matrix, while $\hat{\mathbf{A}}$ denotes estimate of the true mixing matrix **A** obtained by previously described data clustering algorithm. Pure component spectra in wavelet domain are obtained from the solution of linear program (4) as $\hat{\mathbf{s}}(a,b_t) = \mathbf{u} - \mathbf{v}$. If the noise is present in blind decomposition problem more robust sparse solution for $\{\mathbf{s}(a,b_t)\}_{t=1}^{T}$ is obtained by solving ℓ_1-regularized least square problem, [29, 30]:

$$\hat{\mathbf{s}}(a,b_t) = \arg\min_{\mathbf{s}(a,b_t)} \frac{1}{2}\left\|\hat{\mathbf{A}}\mathbf{s}(a,b_t) - \mathbf{x}(a,b_t)\right\|_2^2 + \lambda \left\|\mathbf{s}(a,b_t)\right\|_1 \quad \forall\, t = 1,...,T \qquad (5)$$

or ℓ_2-regularized linear problem [31]:

$$\hat{\mathbf{s}}(a,b_t) = \arg\min_{\mathbf{s}(a,b_t)} \left\|\mathbf{s}(a,b_t)\right\|_1 \text{ subject to } \left\|\hat{\mathbf{A}}\mathbf{s}(a,b_t) - \mathbf{x}(a,b_t)\right\|_2^2 \le \varepsilon \quad \forall\, t = 1,...,T \tag{6}$$

It can be shown that solution of (6) is minimizer of (5) for some λ>0, [34]. Also when ε=0 (6) reduces to linear program (5), [34, 29]. Note that all three pure component spectra in wave number domain are obtained by applying inverse wavelet transform to $\{\hat{s}_m(a,b_t)\}_{m=1}^{\hat{M}}$. In the experiments reported in section 3, in the SCA-based approach to estimate the pure components FT-IR spectra, we have tested linear programming method (4) and interior point [29, 37] and gradient projection [31, 38] methods to solve ℓ_1-regularized least square problem (5). As can be seen in section 3 all three algorithms yielded the same result, what implies that level of sparseness among pure component spectra was not high enough to yield good solution. That raised motivation to look for alternative solution of the blind underdetermined FT-IR spectra decomposition problem.

2.3. Flexible component analysis and HALS NMF

Majority of algorithms used for adaptive NMF are based on the alternating minimization of the squared Euclidean distance expressed by the Frobenius norm with respect to two sets of parameters $\{a_{nm}\}$ and $\{s_{mt}\}$ [19, 20, 26]:

$$D_F\left(\mathbf{X}\|\mathbf{AS}\right) = \frac{1}{2}\left\|\mathbf{X} - \mathbf{AS}\right\|_2^2 + \alpha_{\mathbf{S}} J_{\mathbf{S}}(\mathbf{S}) + \alpha_{\mathbf{A}} J_{\mathbf{A}}(\mathbf{A}) \tag{7}$$

where $J_{\mathbf{S}}(\mathbf{S})$ and $J_{\mathbf{A}}(\mathbf{A})$ represent constraints imposed on **S** and **A**, while $\alpha_{\mathbf{S}}$ and $\alpha_{\mathbf{A}}$ represent corresponding regularization constants. Since decomposition implied by SLMM (1) through minimization of the squared Euclidean distance only has many solutions, constraints are necessary in order to yield solutions for **A** and **S** that are meaningful. In a majority of cases sparseness constraints are imposed on **A** and **S** to obtain meaningful solutions. However, FT-IR spectra are not very sparse but they are reasonably smooth. Consequently, we shall simultaneously impose smoothness and sparseness

constraints on rows of **S**: $\left\{\underline{\mathbf{s}}_m\right\}_{m=1}^{M}$. In addition to that, to solve uBSS problem we shall employ minimization of the of the local cost functions [19,20]:

$$D_F^{(m)}\left(\mathbf{X}^{(m)}\middle\|\mathbf{a}_m\underline{\mathbf{s}}_m\right)=\frac{1}{2}\left\|\mathbf{X}^{(m)}-\mathbf{a}_m\underline{\mathbf{s}}_m\right\|_2^2+\alpha_{\mathbf{s}_p}^{(m)}\left\|\underline{\mathbf{s}}_m\right\|_1+\alpha_{\mathbf{s}_m}^{(m)}\left\|\varphi\left(\mathbf{L}\underline{\mathbf{s}}_m^{\mathrm{T}}\right)\right\|_1 \quad m=1,\ldots,M \tag{8}$$

with respect to $\{s_{mt}\}$ where

$$\mathbf{X}^{(m)}=\mathbf{X}-\sum_{j\neq m}\mathbf{a}_j\underline{\mathbf{s}}_j \tag{9}$$

Constant $\alpha_{\mathbf{s}_p}^{(m)}$ regulates level of sparseness while constant $\alpha_{\mathbf{s}_m}^{(m)}$ regulates level of smoothness of the pure component spectra $\left\{\underline{\mathbf{s}}_m\right\}_{m=1}^{M}$. Smoothness constraint employed here is of the form $\varphi(t)=|t|^2$ where **L** represents second order smoothing operator

$$\mathbf{L}=\begin{pmatrix}-2 & 2 & & & & & \\ 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & \cdots\cdots & & & & \\ & & & & 1 & -2 & 1 \\ & & & & & 2 & -2\end{pmatrix}=\begin{pmatrix}-2 & & & & & & \\ & -2 & & & & & \\ & & -2 & & & & \\ & & & \cdots\cdots & & & \\ & & & & -2 & & \\ & & & & & -2 & \end{pmatrix}+\begin{pmatrix}0 & 2 & & & & & \\ 1 & 0 & 1 & & & & \\ & 1 & 0 & 1 & & & \\ & & \cdots\cdots & & & & \\ & & & & 1 & 0 & 1 \\ & & & & & 2 & 0\end{pmatrix}=-2\mathbf{I}+\mathbf{P} \tag{10}$$

where **I** is identity matrix. Assuming that columns of **A** are normalized to ℓ_2 unit norm, minimization of (8) with respect to $\left\{\underline{\mathbf{s}}_m\right\}_{m=1}^{M}$ yields the following learning rules

$$\underline{\mathbf{s}}_m\leftarrow\frac{1}{1+\alpha_{\mathbf{s}_m}^{(m)}}\left[\mathbf{a}_m^{\mathrm{T}}\mathbf{X}^{(m)}-\alpha_{\mathbf{s}_p}^{(m)}\mathbf{1}_{1\times T}+\alpha_{\mathbf{s}_m}^{(m)}\underline{\mathbf{s}}_m\mathbf{P}^{\mathrm{T}}\right]_+ \tag{11}$$

As opposed to pure components spectra the concentration matrix is learned globally through minimization of (7) without any constraints imposed on it. This yields the following learning rule for **A**

$$\mathbf{A} \leftarrow \left[\mathbf{X}\mathbf{S}^{\mathrm{T}} \left(\mathbf{S}\mathbf{S}^{\mathrm{T}} + \lambda \mathbf{I}_M \right) \right]_+ \tag{12}$$

wherein after each iteration **A** is normalized to ℓ_2 unit column norm. In (12) $\mathbf{I}_M$ is an $M \times M$ identity matrix and $\mathbf{1}_{I \times T}$ is row vector with all entries equal to one. In (11) and (12) $[\xi]_+ = \max\{\varepsilon, \xi\}$ (e.g., $\varepsilon = 10^{-16}$) is used to prevent negative solutions for **A** and **S**. Regularization constant λ in (12) is used to improve ill-conditioning of the matrix $\mathbf{S}\mathbf{S}^{\mathrm{T}}$ and changes as a function of the iteration index k as $\lambda_k = \lambda_0 \exp(-k/\tau)$ (with $\lambda_0 = 100$ and $\tau = 0.02$ in the experiments). Additional improvement in the performance of the NMF algorithms can be obtained when they are applied in the multilayer mode [32, 35, 36], whereas sequential decomposition of the nonnegative matrices is performed as follows. In the first layer, the basic approximation decomposition is performed $\mathbf{X} \cong \mathbf{A}^{(1)}\mathbf{S}^{(1)} \in \mathrm{R}_{0+}^{N \times T}$. In the second layer result from the first layer is used to build up new input data matrix for the second layer $\mathbf{X} \leftarrow \mathbf{S}^{(1)} \in \mathrm{R}_{0+}^{M \times T}$ yielding $\mathbf{X}^{(1)} \cong \mathbf{A}^{(2)}\mathbf{S}^{(2)} \in \mathrm{R}_{0+}^{M \times T}$. After L layers the data decomposes as follows

$$\mathbf{X} \cong \mathbf{A}^{(1)}\mathbf{A}^{(2)} \cdots \mathbf{A}^{(L)}\mathbf{S}^{(L)} \tag{13}$$

Thus, learning rules (11) and (12) can be combined with multilayer mode of operation (13) constituting multilayer HALS NMF algorithm. Performance of the NMF algorithm critically depends on the strategy employed to select initial values for **A** and **S**. The reason is that cost functions (7) and (8) are convex with respect to **A** or **S** but not with respect to both of them. This increases chance, especially in a case of large scale problems, that NMF algorithm will be stuck in local minima yielding poor performance. Therefore, a multistart initialization procedure is proposed to alleviate these problems. It includes number of random guesses for **A** and **S**. For each random guess chosen const function is minimized. In

addition to squared Euclidean distance another cost function such as one based on generalized Kullback-Leibler divergence can be chosen for this purpose, [33]. We select as initial values for **A** and **S** the combination that yields minimum of the chosen cost function within predefined number of iterations. The multistart procedure is briefly outlined below:

Select: number of restart R*, number of alternating steps* K_i *and number of final alternating steps* K_f

for r=1 to R **do**

Initialize randomly $\mathbf{A}^{(0)}$ and $\mathbf{S}^{(0)}$

$\{\mathbf{A}^{(r)},\mathbf{S}^{(r)}\} \leftarrow$ nmf_algorithm($\mathbf{X},\mathbf{A}^{(0)},\mathbf{S}^{(0)},K_i$);

compute $d_r=D(\mathbf{X}||\mathbf{A}^{(r)}\mathbf{S}^{(r)})$;

end

$r_{min}=\text{argmin}_{1\le r\le R}\, d_r$;

$\{\mathbf{A},\mathbf{S}\} \leftarrow$nmf_algorithm($\mathbf{X},\mathbf{A}^{(r_{min})},\mathbf{S}^{(r_{min})}, K_f$);

3. Experimental

3.1. Software environment

Described approach for blind decomposition of FT-IR spectra that includes data clustering and HALS NMF algorithm was tested using custom scripts in MATLAB programming language (version 7.1.; The MathWorks, Natick, MA). The SCA algorithms have been implemented using `linprog` command from the Optimization toolbox for problem (4), using *interior point* method for problem (5) with a code provided at [37], and using *gradient projection* method with a code provided at [38]. All programs were executed on PC running under the Windows XP operating system using Intel Core 2 Quad Processor Q6600 operating with clock speed of 2.4 GHz and 4GB of RAM installed.

3.2. FT-IR measurements

Amino acid derivatives Boc_2-Tyr-NH_2 (pure component **c1**), Boc-Phe-NH_2 (pure component **c2**) and Boc-Phe-NH-CH_2-C≡CH (pure component **c3**) and two mixtures, X_1 (**c1**:**c2**:**c3** = 3:3:1, w/w/w) and X_2 (**c1**:**c2**:**c3** = 2:5:3, w/w/w) were prepared. The powered sample was placed onto the ATR crystal and spectrum was recorded at a resolution of 4 cm^{-1} on an ABB Bomem MB102 spectrometer, equipped with CsI optics, DTGS detector, and a Specac 3000 Series high stability temperature controller with heating jacket.

4. Results and Discussion

To test the described approach to blind IR spectra decomposition, structurally similar amino acid derivatives Boc_2-Tyr-NH_2 (pure component **c1**), Boc-Phe-NH_2 (pure component **c2**) and Boc-Phe-NH-CH_2-C≡CH (pure component **c3**), see Figure 1, were chosen in this study. As clearly seen in Figure 1a to 1c, pure FT-IR spectra reflect high degree of similarity, thus providing sufficiently challenging experimental ground for mathematical algorithm. Number of pure components is estimated from two mixtures, shown in Figure 2, in wavelet domain with the clustering algorithm described in section 2.1. When dispersion factor in eq.(3) is set to σ=0.035 the number of the pure components is estimated as M=3. The corresponding clustering function given by eq.(3) is shown in Figure 3. The estimate of the number of pure components remains stable when dispersion factor is changed within σ=0.035±0.007. Thus, employed data clustering algorithms is quite robust. Clustering function shown in Figure 3 also reveals that one pure component, 3, was contained in low concentrations. However, it is quite realistic to expect that dispersion constant is set sub-optimally in which case too many clusters could be generated. Thus, we have to allow that some of them will not correspond with the true components but can be outliers caused by chemical noise or other types of imperfections that exist in experimental world. Hence, we propose information-theoretic criteria called negentropy, [13], to measure information content of to be estimated pure components and rank them according to estimated negentropy measure.

Negentropy is differential entropy defined relatively to the entropy of the Gaussian process. Approximation of negentropy for random process x is obtained as

$$J(x) \approx \frac{\left(C_3(x)\right)^2}{12} + \frac{\left(C_4(x)\right)^2}{48} \tag{14}$$

where $C_3(x)$ and $C_4(x)$ are third order and fourth order cumulants of the random process x, [41]. Because we shall calculate negentropy of the magnitude spectra of the estimated pure components in frequency domain we use the definition for the cumulants for the non-zero mean random process x

$$C_3(x) = E\left[x^3\right] - 3E\left[x\right]E\left[x^2\right] + 2E^3\left[x\right]$$

$$C_4(x) = E\left[x^4\right] - 4E\left[x\right]E\left[x^3\right] + 12E^2\left[x\right]E\left[x^2\right] - 6E^4\left[x\right] \tag{15}$$

where $E[x]$ in (15) denotes mathematical expectation of x. The Gaussian random process is the least informative among the random processes with unbounded support and has highest entropy. Hence, random processes that are informative are non-Gaussian with the non-zero negentropy measure. We intuitively expect that pure components are informative. Thus, the estimated pure components that correspond to the true pure components are expected to have significantly larger negentropy than the negentropy of the outliers. Figure 4 shows results obtained by HALS NMF algorithm, eq.(11) and (12), in single layer mode with regularization constants $\alpha_{\mathbf{s}_p}^{(m)} = 0.02$ and $\alpha_{\mathbf{s}_m}^{(m)} = 0.05$ after 2500 iterations where instead of M=3 pure components, as indicated by clustering results shown in Figure 3, we have assumed existence of M=4 pure components. As clearly seen in Figure 4d this extracted component represents an outlier. Estimated negentropies of the first three extracted pure components shown in Figures 4a to 4c were 1.086×10^6, 1.3×10^5 and 1.97×10^6, while estimated negentropies of the true pure components shown in Figures 1a to 1c were 3.437×10^7, 9.902×10^7 and 1.365×10^8. Estimated negentropy of the outlier shown in Figure 4d was 4.658×10^{-7}. Hence, estimated component shown in Figure 4d can be easily detected as outlier.

We have experimented with multilayer implementation, eq. (11), extensively but no significant improvement in the separation quality was obtained. We contribute this to the fact that level of sparseness between pure components FT-IR spectra was not high enough. As demonstrated in [32, 35] the multilayer implementation really helps to find solution that is sparser than one obtained in single layer mode. However, if true pure components are not sparse enough multilayer implementation could not help. That is why we have proposed HALS NMF algorithm that simultaneously imposes sparseness and smoothness constraints on the pure components. For the purpose of comparative performance analysis we have tested linear programming, eq.(4), based SCA approach to the same problem, as well as interior point [29, 37] and gradient projection [31, 38] methods to solve ℓ_1-regularized least square problem (5). The concentration matrix was estimated during data clustering phase. Figure 5, 6 and 7 show corresponding results. They are consistently similar what implies that lack of sparseness between pure components and not a noise is what affects the quality of the SCA-based solution. To quantify quality of the used blind decomposition schemes we have calculated normalized correlation coefficients between true pure components spectra shown in Figure 1, and pure component spectra estimated by HALS NMF algorithm, Figure 4, and SCA algorithm, Figure 5. Correlation coefficients between corresponding spectra in a case of HALS NMF algorithm were: 0.9101, 0.9804 and 0.9342. For linear programming-based SCA algorithm correlation coefficients in the same order as before were: 0.8468, 0.822 and 0.2125. For interior point method-based SCA algorithm correlation coefficients in the same order as before were: 0.8512, 0.8219 and 0.2100. For gradient projection method-based SCA algorithm correlation coefficients in the same order as before were: 0.8442, 0.8709 and 0.2396. Regularization constant in (5) was set to λ=0.1. Clearly, SCA algorithms failed to extract the pure component 3, while the other two pure components were extracted with significantly less accuracy than by HALS NMF algorithm.

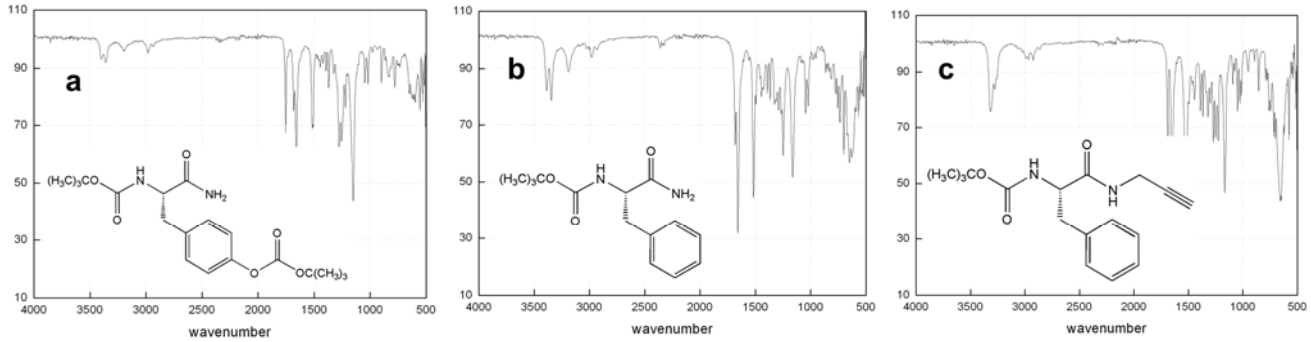

Figure 1. Pure components FT-IR spectra: a) pure component **c1** (Boc_2-Tyr-NH_2); b) pure component **c2** (Boc-Phe-NH_2); c) pure component **c3** (Boc-Phe-NH-CH_2-C≡CH).

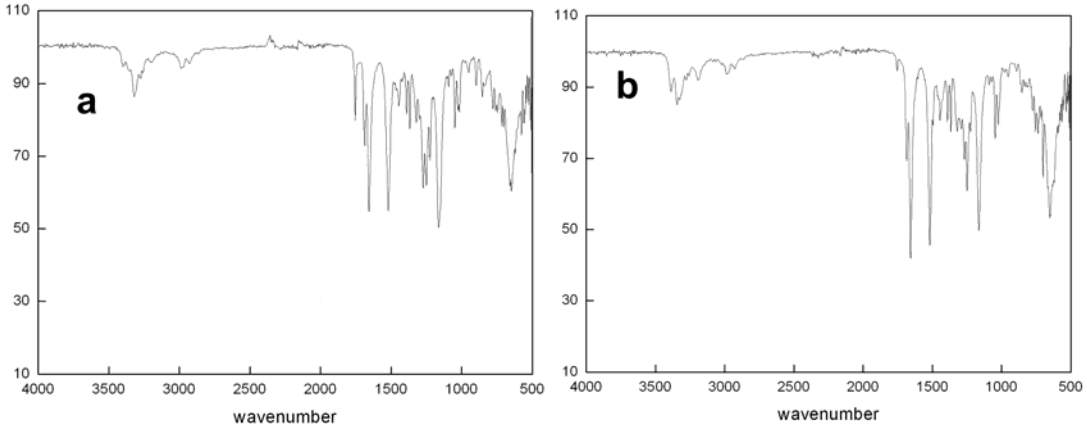

Figure 2. FT-IR spectra of two mixtures: a) X_1; b) X_2.

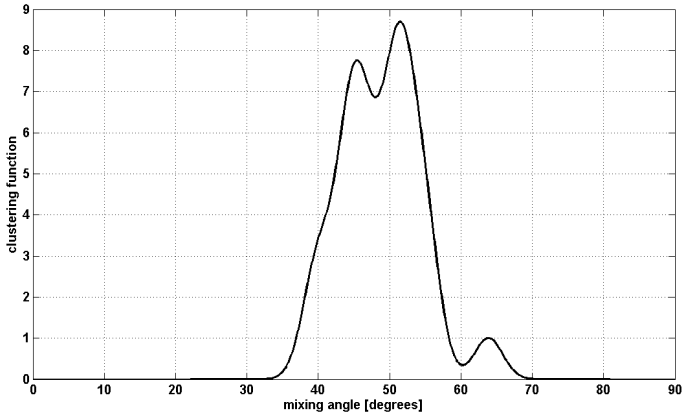

Figure 3. Clustering function, eq.(3), for two mixtures shown in Figure 2 transformed in wavelet domain. Dispersion factor was set to σ=0.035. Three peaks indicate existence of three pure components spectra in two mixtures.

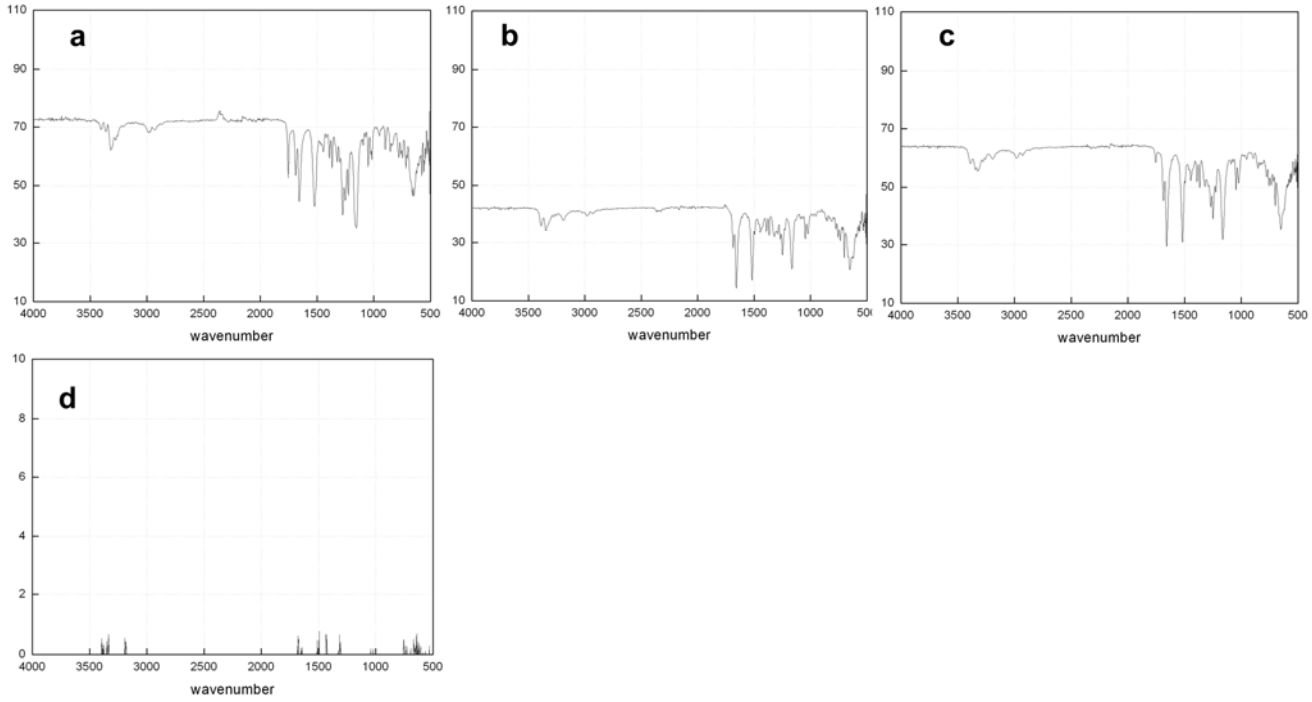

Figure 4. FT-IR spectra of three pure components extracted by HALS NMF algorithm (9) and (10): a) pure component **c1**; b) pure component **c2**; c) pure component **c3**; d) outlier.

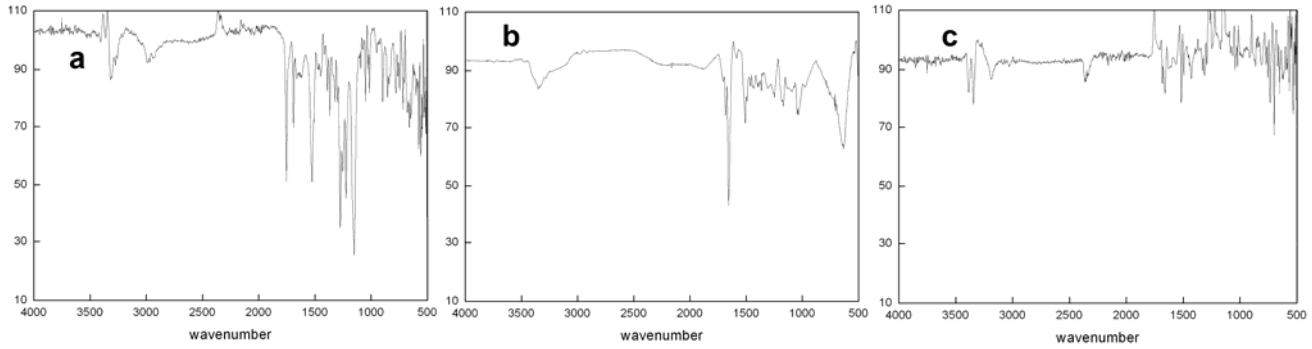

Figure 5. FT-IR spectra of three pure components extracted by linear programming-based SCA algorithm in wavelet domain: a) pure component **c1**; b) pure component **c2**; c) pure component **c3**.

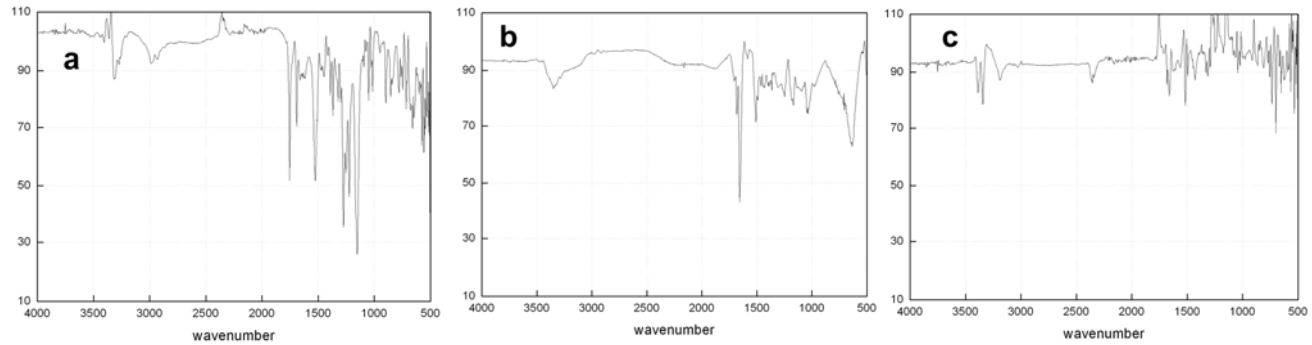

Figure 6. FT-IR spectra of three pure components extracted by interior point method-based SCA algorithm in wavelet domain: a) pure component **c1**; b) pure component **c2**; c) pure component **c3**.

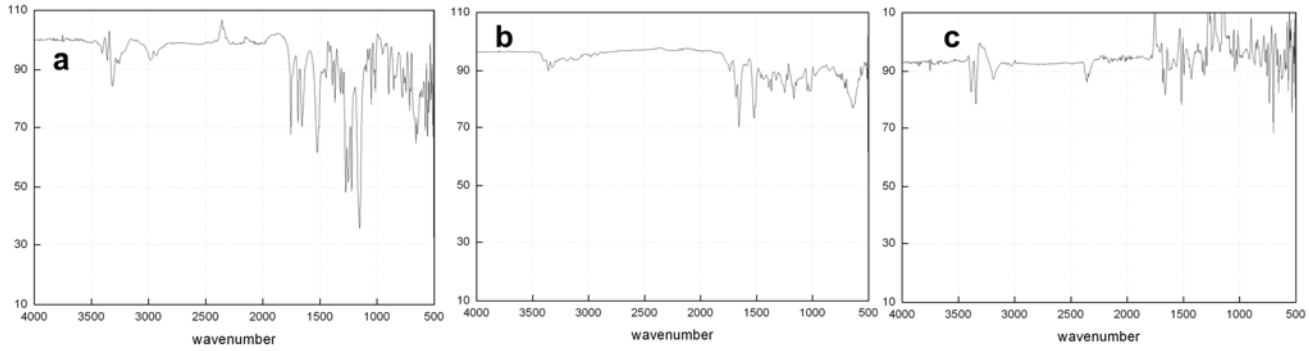

Figure 7. FT-IR spectra of three pure components extracted by gradient projection method-based SCA algorithm in wavelet domain: a) pure component **c1**; b) pure component **c2**; c) pure component **c3**.

5. Conclusions

FCA-based approach using HALS NMF algorithm has been proposed for blind extraction of *more than two* pure components spectra in FT-IR spectroscopy measuring *two mixtures only*. This is achieved by simultaneously imposing smoothness and sparseness constraints on the pure components FT-IR spectra. This appears to be the first time to report such result, because other blind decomposition methods require the number of mixtures to be equal to or greater than the unknown number of pure components. Unlike many existing BSS methods that assume the number of pure components to be known in advance, proposed FCA-based method estimates it by data clustering algorithm in wavelet domain. Proposed FCA-based approach can be used as a part of software packages for the analysis of FT-IR spectra and identification of the chemical compounds.

Acknowledgment

The work of I. Kopriva and I. Jerić was respectively supported by the Ministry of Science, Education and Sports, Republic of Croatia under Grants 098-0982903-2558 and 098-0982933-2936. We thank dr. Boris Zimmermann for performing FT-IR measurements.

References

[1] D. Nuzillard, S. Bourg, J.-M. Nuzillard, J. Magn. Reson., 133 (1998) 358-363.

[2] E. Visser, T.-W. Lee, Chemom. Intell. Lab. Syst., 70 (2004) 147-155.

[3] G. Wang, Q. Ding, Y. Sun, L. He, X. Sun, Spectrochim. Acta Part A., 70 (2008) 571-576.

[4] G. Wang, Q. Ding, Z. Hou, Trends in Anal. Chem., 27 (2008) 368-376.

[5] J. Chen, X. Z. Wang, J. Chem. Inf. Comput. Sci., 41 (2001) 992-1001.

[6] X. Shao, W. Wang, Z. Hou, W. Cai, Talanta, 69 (2006) 676-680.

[7] J. Y. Ren, C. Q. Chang, P. C. W. Fung, J. G. Shen, F. H. Y. Chan, J. Magn. Reson., 166 (2004) 82-91.

[8] Ch. Chang, J. Ren, P.C. Fung, Y. S. Hung, J. G. Shen, F. H. Y. Chan, J. Magn. Reson., 175 (2005) 242-255.

[9] X. Shao, G. Wang, S. Wang, Q. Su, Anal. Chem., 76 (2004) 5143-5148.

[10] G. Wang, W. Cai, X. Shao, Chemom. Intell. Lab. Syst., 82 (2006) 137-144.

[11] V. A. Shashilov, M. Xu, V.V. Ermolenkov, I. K. Lednev, J. Quant. Spect. & Rad. Trans., 102 (2006) 46-61.

[12] H. Li, T. Adali, W. Wang, D. Emge, A. Cichocki, J. VLSI Sig. Proc., 48 (2007) 83-97.

[13] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley Interscience, 2001.

[14] A. Cichocki, S. I. Amari, Adaptive Blind Signal and Image Processing, John Wiley, New York, 2002.

[15] P. Comon, Sig. Process., 36 (1994) 287-314.

[16] A. Hyvärinen, in Proc. of the International Conference on Artificial Neural Networks (ICANN'98), 1998, pp. 541-546.

[17] I. Kopriva, D. Seršić, Neurocomputing, 71 (2008) 1642-1655.

[18] A. Cichocki, P. Georgiev, IEICE Trans. on Fund. Elec. Comput. and Comput. Sci., E86-A (2003) 522-531.

[19] A. Cichocki, R. Zdunek, S. I. Amari, LNCS, 4666 (2007) 169-176.

[20] A. Cichocki, A. -H. Phan, R. Zdunek, L.-Q. Zhang, LNCS, 4984 (2008) 811-820.

[21] F. M. Naini, G. H. Mohimani, M. Babaie-Zadeh, Ch. Jutten, Neurocomputing, 71 (2008) 2330-2343.

[22] Y. Li, A. Cichocki, S. Amari, Neural Comput., 16 (2004) 1193-1234.

[23] Y. Li, S. Amari, A. Cichocki, D.W.C. Ho, S. Xie, IEEE Trans. on Sig. Proc., 54 (2006) 423-437.

[24] P. Georgiev, F. Theis, A. Cichocki, IEEE Trans. on Neural Networks, 16 (2005) 992-996.

[25] P. Bofill, M. Zibulevsky, Sig. Proc., 81 (2001) 2353-2362.

[26] A. Cichocki, R. Zdunek, S. Amari, IEEE Sig. Proc. Magazine, 25 (2008) 142-145.

[27] I. Takigawa, M. Kudo, J. Toyama, IEEE Tr. on Sig. Proc., 52 (2004) 582-591.

[28] D.L. Donoho, M. Elad, Proc. Nat. Acad. Sci., 100 (2003) 2197-2202.

[29] S. J. Kim, K. Koh, M. Lustig, S. Boyd, D. Gorinevsky, IEEE J. Selec. Top. in Sig. Proc., 1 (2007), 606-617.

[30] J. A. Tropp, A. C. Gilbert, IEEE Trans. on Inf. Theory., 53 (2007) 4655-4666.

[31] M. A. T. Figuiredo, R. D. Nowak, S. J. Wright, IEEE J. Selec. Top. in Sig. Proc., 1 (2007) 586-597.

[32] A. Cichocki, R. Zdunek, Elect. Lett., 42 (2006) 947-948.

[33] A. Cichocki, R. Zdunek, S. Amari, LNCS., 3889 (2006) 32-39.

[34] E. Van Den Berg, M. P. Friedlander, SIAM Journal of Scientific Computing, 31 (2008), 890-912.

[35] K. Stadlthanner, F. J. Theis, C. G. Puntonet, J. M. Górriz, A. M. Tomé, E. W. Lang, LNCS., 3745 (2005) 137-148.

[36] K. Stadlthanner, F. J. Theis, E. W. Lang, A. M. Tomé, C. G. Puntonet, J. M. Górriz, Neurocomputing, 71 (2008) 2356-2376.

[37] http://www.stanford.edu/~boyd/l1_ls/.

[38] www.lx.it.pt/~mtf/GPSR

[39] P. McCullagh, Tensor Methods in Statistics, Chapman & Hall, London, 1987.