

Chromatograms separation using Matrix decomposition

S.Anbumalar
Department of Instrumentation
and Control Engineering,
Sri Manakula Vinayagar
Engineering College,
Madagadipet, Puducherry,
India 605107.

P.Rameshbabu
Department of Electronics and
Instrumentation Engineering,
Pondicherry Engineering College
Puducherry, India 605014

R.Anandanatarajan
Department of Electronics and
Instrumentation Engineering,
Pondicherry Engineering College,
Puducherry, India 605014

ABSTRACT

Non – negative matrix factorization (NMF) was generally used to obtain representation of data using non – negativity constraints. It lead to parts – based (or) region based representation in the vector space because they allow only additive combinations of original data. NMF has been applied so far in image and text data analysis, audio signal separation, signal separation in bio-medical applications and spectral resolution. The original Lee and Seung 's NMF has to be modified for chemical analysis, based on the characteristics of that signal. In this paper, sparse NMF (sNMF) has been used for the deconvolution of overlapping chromatograms of chemical mixture. Before applying sNMF, the number of components in mixture was determined using Principal Component Analysis (PCA). The experimental overlapping chromatograms were obtained using Gas Chromatography –Flame Ionization Detector (GC-FID) for the chemical mixture of acetone and acrolein and they have been soundly resolved by sNMF algorithm. The proposed algorithm has also been tested with simulated two, three and four component chromatograms of severely overlapped and embedded peaks. Even though there are three or four components, the results are encouraging. The correlation coefficient is greater than 0.99 and signal to noise ratio is greater than 29dB always.

General Terms: Chemometrics; NMF.

Keywords

sparse Non-negative Matrix Factorization (sNMF); Principal Component Analysis (PCA); resolution; overlapping chromatograms; GC– FID; acetone and acrolein mixture.

1. INTRODUCTION

Many resolution techniques are available to extract information from multivariate data. Evolving Factor Analysis (EFA) [1-2], Fixed Size Moving Window Evolving Factor Analysis (FSMW-EFA) [3], SIMPLEx-to-use Interactive Self-modeling Mixture Analysis (SIMPLISMA) [4], Multivariate Curve Resolution (MCR) [5-7], Iterative Target Transformation Factor Analysis (ITTFA) [8-9] and Heuristic Evolving Latent Projections (HELP) [10-11] are some of the important techniques so far applied to many fields. All the above methods have their limitations in separation of overlapped chromatograms when they are severely overlapped [12]. Many methods have been developed further to improve the resolution. However, it is still difficult to resolve severely overlapping peaks accurately.

Lee and Seung Suggested Non-negative Matrix Factorization (NMF) in 1999 [13] and provided more algorithms in 2001 [14]. It has been generally applied to image and text data mining. Several works have been carried out in NMF for image analysis [15-18], audio signal separation [19], spectral resolution [20] and signal separation in bio-medical application [21]. A chemometric application of the NMF method is proposed by Li et al. [22] to detect chemical compounds from a chemical substance represented through Raman spectroscopy. Hong-Tao Gao et al., (2005) applied NMF for two components overlapped spectrum resolution [20] in which NMF was applied after imposing constraints based on unimodality, smoothness and sparseness of the data.

In this work, acetone and acrolein were mixed and analyzed using Gas Chromatography –Flame Ionization Detector (GC-FID) and it gives overlapping chromatograms due to their closer retention times (i.e., 1.127 min for acetone and 1.187 min for acrolein). Hence, if any product that contains acetone and acrolein is analysed by Gas Chromatography –Flame Ionization Detector (GC-FID), it will be difficult to identify these components due to non-separated chromatograms. Certain experimental separation procedure has been proposed already which involves costly chemicals and lengthier separation time [23]. Hence, sNMF algorithm has been used in this work for separation of overlapped chromatograms obtained. The sNMF algorithm has been applied on experimental chromatograms. The results of sNMF are encouraging. The algorithm is also tested with simulated overlapped and embedded chromatograms of complex 4 components mixtures which were not done in [20].

2. MATERIALS AND METHODS

2.1 Experiments

Initially the chromatograms for the acetone and acrolein standards were obtained using GC-FID by injecting the standards separately. Then the acetone and acrolein were mixed in 14 different concentrations. The mixture was vigorously stirred separately under ice cold condition for half an hour and 0.2 µL of these mixtures were injected in GC-FID.

2.1.1 Instrumental condition

Chemito 8610 make Gas chromatography with Flame Ionization Detector was used with the detector temperature of 150° C. HPFFAP Capillary Column (30 m × 0.25 mm ID, BP5) was used with a temperature of 100 ° C. The Temperature was programmed as 40 ° C for 2 min, 5° C/min, 68 ° C for 2min, 10 ° C/min, 100 ° C for 2min. Injector temperature was kept at 150 ° c. Nitrogen was used as a carrier gas at a flow rate of 90 ml/min. The sample of 0.2µL volume was injected into the column for analysis. All the graphs were obtained in the

software provided by Iris 32 chromatography provided by chemito technologies pvt ltd., running on PC with Intel (R) Pentium4 CPU 2.00GHz and 512MB RAM. The chromatographic data (i.e., retention time t and detector output) has been exported to an ASCII file through Iris 32 software and acquired through MATLAB R2008a software. For the above instrumental condition, the chromatograms obtained for seemed to be overlapping instead of separated peaks due to their closer retention times. It is difficult to identify the individual components in the overlapping chromatograms.

2.2 Method

This section describes the sNMF algorithm applied in this paper. The performance of the algorithm is discussed in the Results section.

2.2.1 sparse NMF algorithm

NMF is a technique of decomposing a non-negative matrix A into two non-negative matrices W and H as shown in equation 1.

$$A_{mn} \approx (WH)_{mn} = \sum_k W_{mk} H_{kn} \quad (1)$$

where $k=1$ to $r < \min(m,n)$.

Usually, r is the number of principal components. The decomposition is performed so that the product WH should compress and approximates the original data matrix A . W is called basis matrix whereas H is called encoding matrix, of which each column is in one-to-one correspondence with a column in A . Thus, the original data are represented as linear combinations of these basis vectors. Usually r is chosen to be smaller than n or m , so that W and H are smaller than the original matrix A . This results in a compressed version of the original data matrix. NMF allows only additive operations to provide meaningful decomposition. While PCA is used for decomposition, the two factorized matrices will contain positive and negative entries simultaneously, and these negative components make the result often unacceptable in chemical meanings. Instead, NMF does not allow negative entries in the factorized matrices W and H , permitting each column of basis matrix H to represent chromatogram in this work. NMF extensions also exist which includes, projective NMF [24], shifted NMF [25], incremental NMF [26] and sparse higher order NMF [27]. Although NMF is successful in Matrix Factorization, the NMF model does not impose the sparse constraints. Therefore, it can hardly yield a factorization, which reveals local sparse features in the data A . Related sparse coding is proposed in the work of [28] for matrix factorization. Inspired by the original NMF and sparse coding, the aim of this work is to propose sparse Non-negative Matrix Factorization (sNMF) which imposes the sparse and nonnegative constraint, for chromatograms resolution. Since NMF is an approximation factorization, we need to define the cost function to qualify this approximation. One natural way is to use the divergence function between A and WH . Hence, sNMF algorithm which is given in equation 2 has been applied in this paper to find the factors.

$$\sum_{i=1}^m \sum_{j=1}^n [A_{ij} \log((A + 1e^{-12})_{ij} / Y_{ij}) - A_{ij} + Y_{ij}] + \alpha \sum_{q=1}^k \sum_{j=1}^n H_{qj} \quad (2)$$

where $Y = WH$ and α obtained by experience was assumed a positive constant.

A sparse solution to the above constrained minimization can be found by the following update rules of W and H suggested by Lee and Seung (NIPS*2001) :

(i) Initialize matrices W and H randomly under nonnegative and non zero condition.

(ii) Calculate the new value of W using multiplicative update rule

$$W_{iq} = W_{iq} * (((A + 1e^{-12})_{ij} / (WH)_{ij}) * H^T)_{iq} / (B_{il} * (\sum_q H^T)_{iq})_{iq} \quad (3)$$

for $1 \leq i \leq m$ and $1 \leq q \leq k$.

(iii) Calculate the new value of H using multiplicative update rule

$$H_{qj} = (H_{qj} * (W^T * ((A + 1e^{-12})_{ij} / (WH)_{ij}))_{qj}) / (1 + \alpha) \quad (4)$$

for $1 \leq q \leq k$ and $1 \leq j \leq n$,

(iv) Column normalization of W (Liu, Zheng, and Lu added this normalization step) has been done during each iterations by

$$W_{iq} = W_{iq} / (B_{il} * \sum_q W_{iq})_{iq} \quad (5)$$

Where B_{il} is the unit matrix of size $m \times l$

(v) Calculate the objective function using new updated W and H .

Repeat from (ii) to (v) until convergence is achieved.

Element wise operation has been carried out in the iterative steps.

When NMF is introduced to resolve the chemical data matrix, it must be modified according to the characteristics of signal (i.e.,) smoothness, unimodality of chromatograms and sparseness of data. Based on that, constraints have to be imposed into the iterative procedure to get a reasonable and acceptable resolution [20].

In the present work, unimodality and selectivity constraints were integrated into the iterative procedure additionally. The chromatograms taken for analysis are smooth and hence the curve smoothing constraint was not imposed. The algorithm can handle sparse data and hence sparseness constraint was not imposed. The maximum number of iterations has to be specified. It takes only very few seconds to converge. To speed up the convergence, SVD-based initialization [29] was tested. But the results were not reliable. Then the random initialization was done for all cases.

2.2.2 Procedure

Step1: The chromatographic data were taken for acetone and acrolein mixtures of four different concentration ratios. The detector output (m) of each experiment has been taken as a column of matrix A and hence the size of A is $m \times 4$. It is proposed to use sNMF algorithm to perform the deconvolution of a data matrix A .

Step2: The size of matrix W is based on the number of components k in a mixture and it is determined by principal component analysis (PCA).

Step3: The W and H matrix, of size $m \times k$ and $k \times n$ respectively, have been initialized with random non negative and non zero values (using Matlab coding) and were updated with multiplicative step (Lee and Seung; NIPS*2001) as given in equation 3, 4 and 5.

Step 4: The sNMF algorithm is used to decompose the matrix A into W and H matrices.

Step 5: After decomposition, each column of W matrix gives the chromatogram of individual chemical present in the mixture.

Step 6: The resolved components are compared with standard chromatograms and the correlation coefficients are calculated. The resolved signal power is calculated using the formula given in equation 6.

$$Pr(dB) = 10 \log_{10} \frac{\sum_i (S_i / std(S_i))^2}{\sum_i ((S_i / std(S_i)) - (R_i / std(R_i)))^2} \quad \text{----- (6)}$$

where S_i - standard signal,

R_i - resolved signal

Std - standard deviation

$Pr(dB)$ -- closeness of resolved signal power with that of standard signal

i -- 1,2...n ; where n is the end point of every chromatogram.

$r = 1,2,\dots$ number of components in the mixture.

The algorithm was tested on the simulated chromatograms of two cases, i.e., severely overlapped and embedded peaks. One dimensional, HPLC-FID data matrices of two, three and four component chromatograms were simulated using cross product multiplication of Gaussian functions. The components are strongly overlapping with almost equal retention times and widths. Due to that, their mixture appears as a single component chromatogram. The sNMF algorithm of Equation (2) has been performed to deconvolute the data matrix A of simulated overlapping and embedded chromatograms.

Finally, the sNMF algorithm was applied on the experimental severely overlapping chromatograms of acetone and acrolein mixtures. The experimental data were taken for mixtures of different concentration ratios. Due to that, the shape and area of the overlapped chromatograms are different. It makes it difficult to identify the shift in peak position and shape of the individual chromatograms. Hence, preprocessing was not applied. The efficiency of sNMF is evident from the good correlation coefficient of the resolved components with that of the standards. The main focus of this work is to resolve the overlapping components using sNMF. The results were unique when the matrix W was initialized with the help of evolving factor analysis or with the initial estimate obtained from the standard chromatograms. If the W matrix was randomly initialized, the results were unique for a two and three components chromatogram. But, for the complex chromatograms (i.e., four and more components chromatograms), the resolved results were not unique always but they were in a narrow feasible region that are reasonable and acceptable.

3. RESULTS AND DISCUSSION

3.1 Simulated HPLC-FID data matrix

One dimensional, HPLC-FID chromatograms of an overlapped and embedded peaks were simulated using cross product multiplication of Gaussian functions.

Two components system

One dimensional, HPLC-FID data matrices of a two component chromatograms of severely overlapped case were simulated. The sNMF algorithm was then applied on the overlapping peaks to resolve the individual components. The simulated severely overlapping peaks and resolved chromatograms are shown in Fig.1. The resolved components are compared with the standard chromatograms. The correlation coefficients are 0.9769 and 0.9541 for the first and second components respectively. The signal to noise ratio for the resolved components are 13.7608 dB and 10.7684 dB for the first and second components respectively.

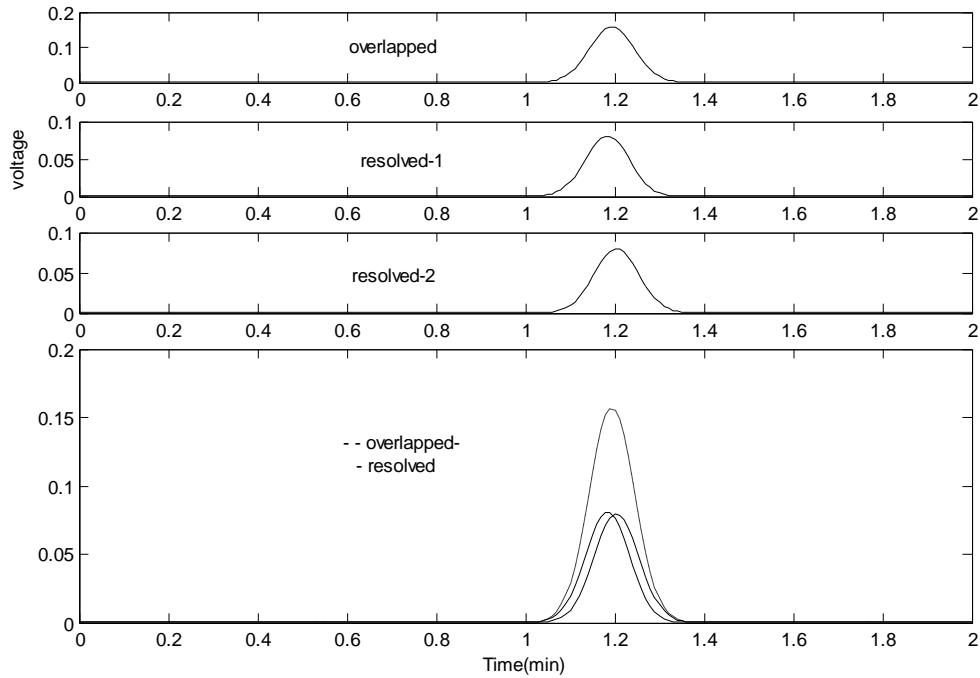


Fig 1: Simulated severely overlapped chromatogram of two components mixture and resolved chromatograms obtained using sNMF.

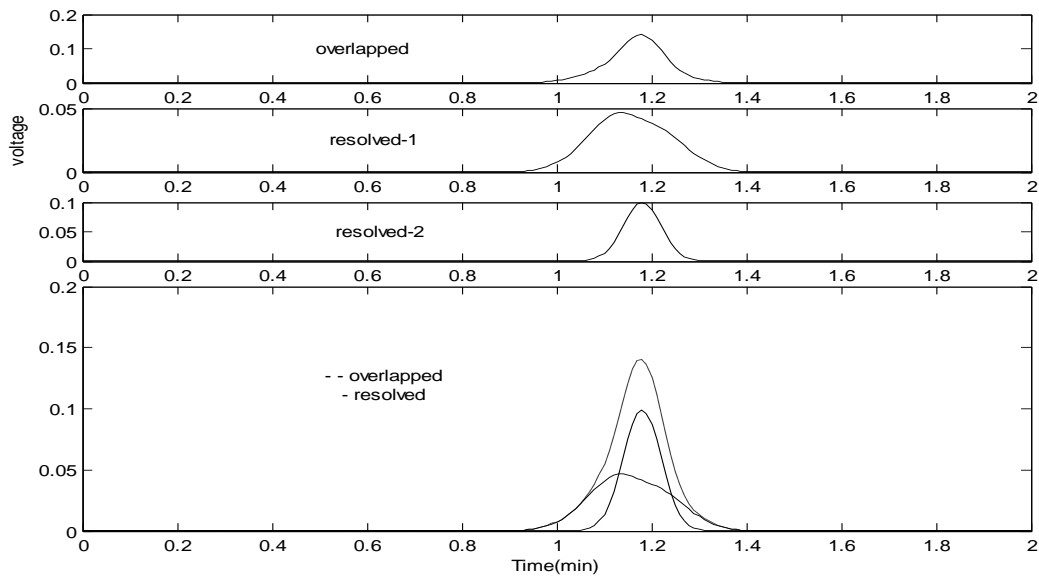


Fig 2: Simulated embedded chromatogram of two components mixture and resolved chromatograms obtained using sNMF .

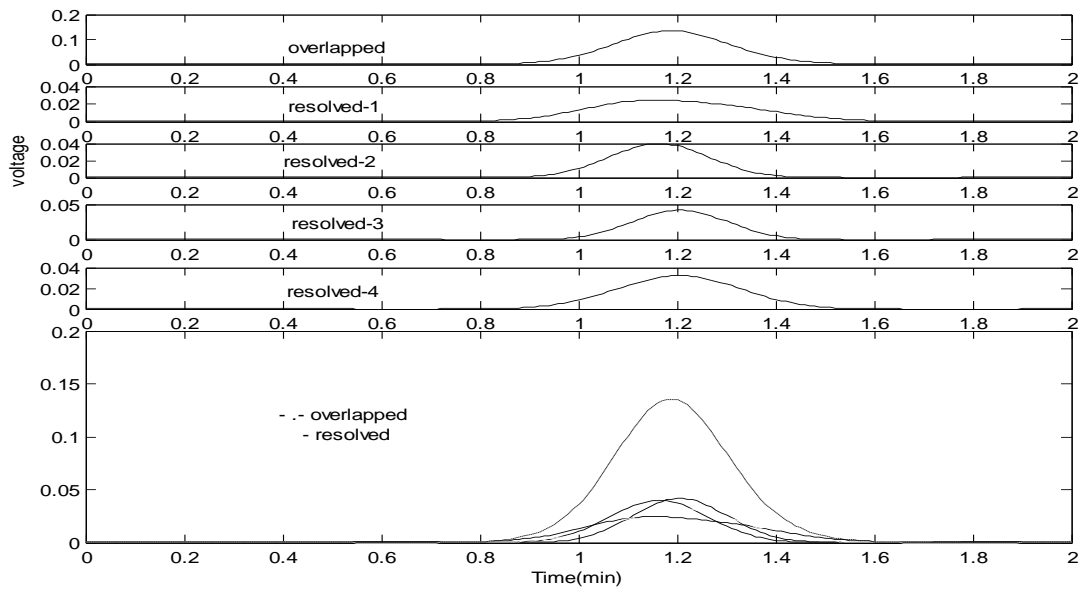


Fig 3: Simulated severely overlapping chromatogram of four components mixture and resolved chromatograms obtained using sNMF.

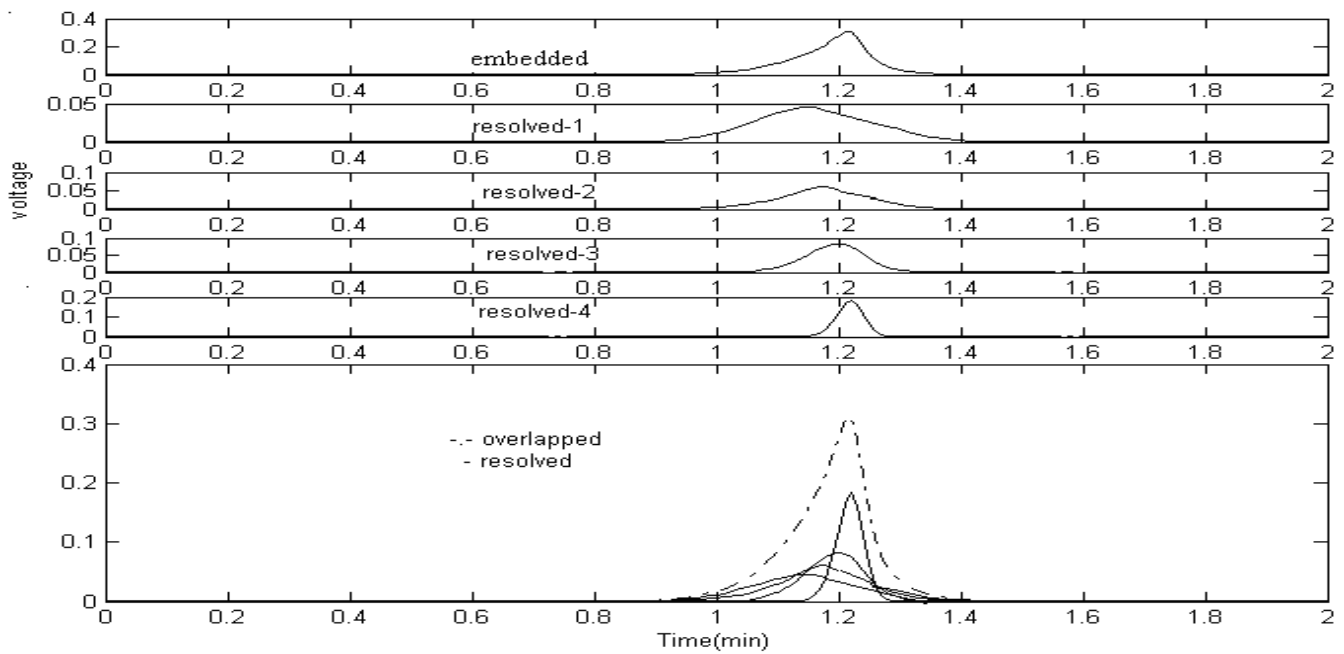


Fig 4: Simulated embedded chromatogram of four components mixture and resolved chromatograms obtained using sNMF.

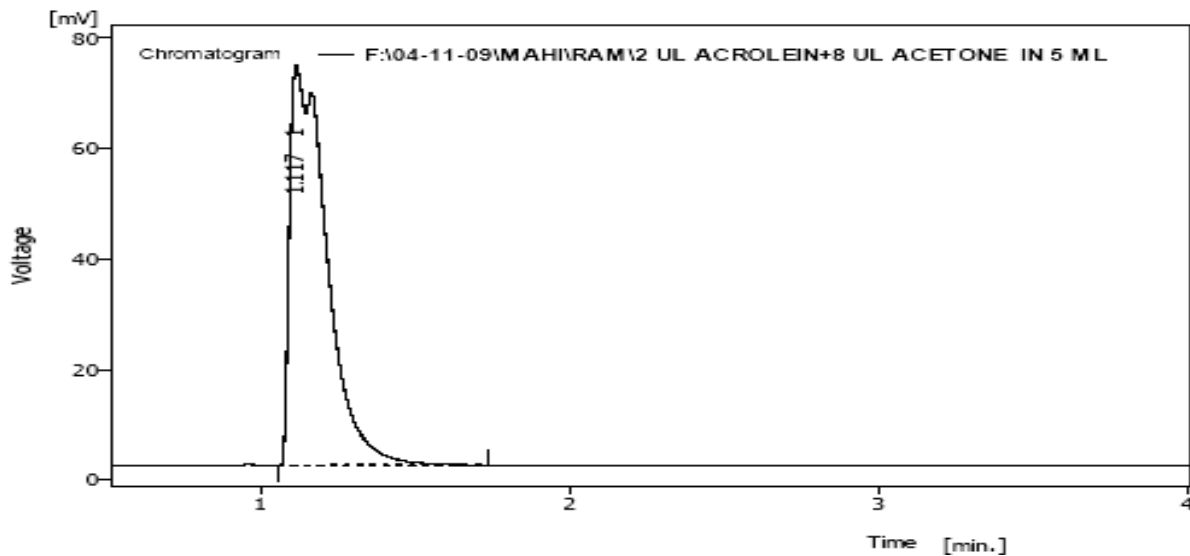


Fig 5: Real experimental severely overlapped chromatogram of acetone and acrolein mixture (2 μ l acrolein and 8 μ l acetone) . Experimental condition: Gas chromatography-Flame Ionization Detector; Carrier gas: Nitrogen (90ml/min); Detector temperature: 150° C; Retention time of acetone 1.127min; Retention time of acrolein 1.187 min.

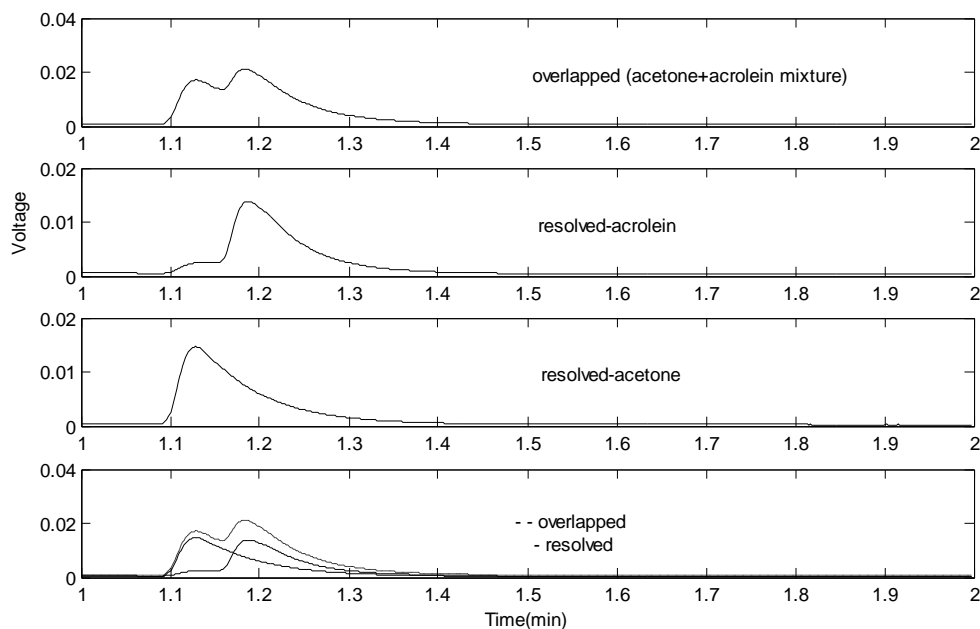


Fig 6: Experimental severely overlapped chromatogram of acetone and acrolein mixture and resolved chromatograms using sNMF (applied on 4 dataset). Experimental condition: Gas chromatography-Flame Ionization Detector; Carrier gas: Nitrogen (90ml/min); Detector temperature: 150° C; Retention time of acetone 1.127min; Retention time of acrolein 1.187 min.

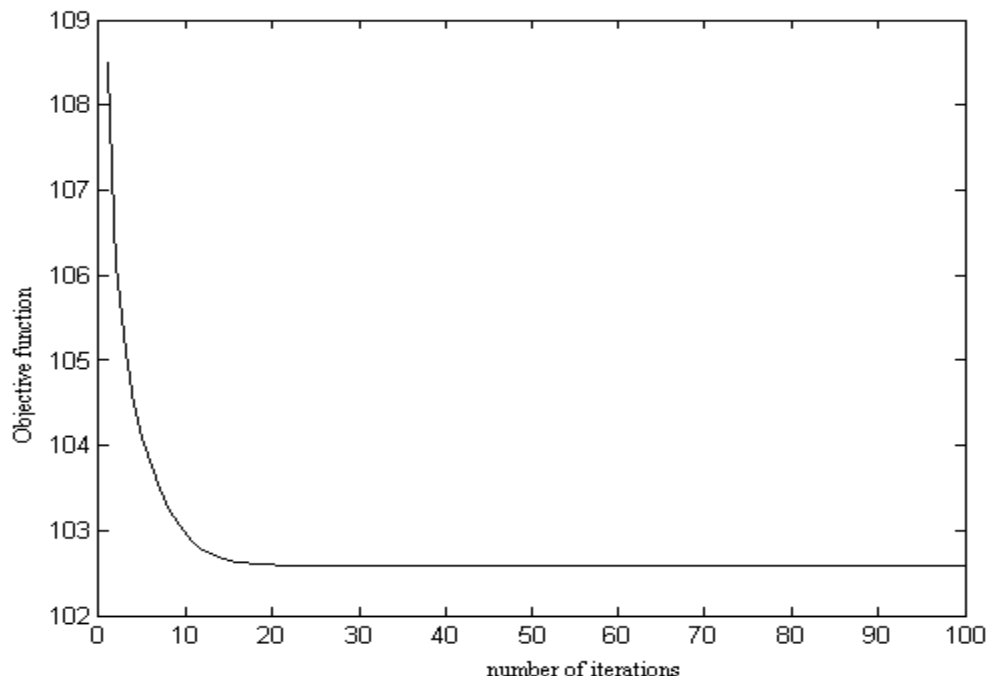


Fig 7: Objective function Vs number of iterations for sNMF implementation.

Two component chromatograms of embedded peaks were simulated. The sNMF algorithm was then applied on the chromatograms to resolve the individual components. The simulated embedded chromatogram and resolved chromatograms are shown in Fig.2. The resolved components are compared with the standard chromatograms. The correlation coefficients are 0.9950 and 1.0000 for the first and second components respectively. The signal to noise ratio for the resolved components are 20.5473 dB and 48.7652 dB for the first and second components respectively.

Three and Four component systems:

The simulated three components system has also been tested with sNMF algorithm. The results are encouraging. The correlation coefficient is greater than 0.99 and the signal to noise ratio is greater than 15dB always.

The sNMF algorithm was then applied on the simulated severely overlapping four components chromatograms. The simulated overlapped chromatogram and resolved chromatograms using sNMF are shown in Fig.3. The resolved components were compared with the simulated standard chromatograms and the correlation coefficients are 0.9269, 0.9912, 0.9984 and 0.9807 for the first, second, third and fourth components respectively. The signal to noise ratio for the resolved components are 8.4715 dB, 18.3927 dB, 25.4012dB and 14.5173 dB for the first, second, third and fourth components respectively.

The sNMF was then applied on the embedded peaks containing four components. The simulated embedded chromatogram and the resolved chromatograms using sNMF are shown in Fig.4. The resolved results were compared with the simulated standard chromatograms and its correlation coefficients are 0.9747, 0.9796, 0.9860 and 0.9974 for the first,

second, third and fourth components respectively. The signal to noise ratio for the resolved components are 13.7881 dB, 14.1252 dB, 15.6845 dB and 22.9471dB for the first, second, third and fourth components respectively.

Even though the degree of overlapping is very high, sNMF algorithm resolves the overlap efficiently with a correlation coefficient greater than 0.9 and signal to noise ratio greater than 8 db. The chromatographic situation of the cases taken will be difficult for an analyst to get a reasonable resolution. But the solutions of sNMF are reasonable and acceptable.

3.2 Real GC–FID experimental data

The overlapping chromatograms were obtained by GC-FID for acetone and acrolein mixture. It contains merging and tailing peaks. The experiments were conducted under condition mentioned in Section 2.1.1. The experimental chromatogram obtained for acetone and acrolein mixture is shown in fig.5. The individual components in the chromatogram can be qualitatively identified with the help of calibration standards. However, there were overlapping peaks due to acetone and acrolein mixture which have almost nearer retention times. (i.e., 1.127 min and 1.187 min respectively). If these overlapping peaks were not resolved, the chromatogram of mixtures measured could not get a good match with that of a pure component in the database.

Hence, sNMF algorithm was applied on the chromatograms obtained to resolve acetone and acrolein. Before applying sNMF, the number of components k in the data A was determined using PCA, which identified two components. Initially, sNMF was applied on A matrix formed from 4 datasets of acetone and acrolein mixture of 4 different concentrations.

The standard chromatograms of acetone and acrolein, overlapped chromatogram of acetone and acrolein mixture and

the resolved chromatograms are shown in Fig.6. The correlation coefficients were obtained by comparing the resolved results with the standard chromatograms and those coefficients are 1.0000 and 0.9993 for acetone and acrolein respectively. The signal to noise ratio for the resolved acetone and acrolein are 49.8145 dB and 29.4805dB respectively. The algorithm works well with random initial matrices of W and H . The resolving power of sNMF had been tested with different size of A matrix by varying n . Initially, the algorithm was tested only with 2 experiments. i.e., n was taken as 2. The result shows that there was reliable resolution even with 2 datasets. The algorithm was then tested by increasing n as three and then as four. The result shows that the signal recovery is higher when the number of datasets taken was increased.

The sNMF algorithm for chromatographic resolution is effective. The correlation coefficient is greater than 0.99 and signal to noise ratio is greater than 29 dB. When chromatographic profiles of two compounds mixture of similar peak-height and almost of the same retention time are analysed, it is quite a troublesome issue for analysts. But the solutions of sNMF are reasonable and acceptable. Fig.7 shows the minimization of objective function Vs number of iterations for sNMF. It shows that the objective function has come to minimum error in less than 100 iterations.

4. CONCLUSION

In this paper, sparse NMF algorithm was applied to resolve simulated data of two, three and four components mixtures of severely overlapped as well as embedded peaks. The algorithm was also tested on experimental overlapped chromatograms of acetone and acrolein mixture. The results are encouraging. The efficiency is checked from the correlation coefficient of the resolved components with standards which is always greater than 0.99 and the signal to noise ratio which is always greater than 29dB.

5. REFERENCES

- [1] Gampp, H., Maeder, M., Meyer, C.J., and Zuberbuhler, A.D. "Calculation of equilibrium constants from multiwavelength spectroscopic data. III. Model-free analysis of spectrophotometric and ESR titrations", *Talanta*, 32 (1985)1133-1139.
- [2] Maeder, M. "Evolving Factor Analysis for the resolution of overlapping chromatographic peaks", *Anal. Chem.*, 59 (1987) 527-530 .
- [3] Keller, H.R., and Massart, D.L. "Peak purity control in liquid chromatography with photodiode array detection by fixed size moving window evolving factor analysis", *Anal. Chim. Acta*, 246(1991) 379-390.
- [4] Windig, W., and Guilment, J. "Interactive self-modeling mixture analysis", *Anal. Chem.*, 63(1991) 1425-1432.
- [5] Tauler, R., Smilde, A.K., and Kowalski, B.R. "Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution", *J. Chemometr.* 9(1995) 31-58.
- [6] Tauler, R. "Multivariate curve resolution applied to second order data", *Chemom. Intell. Lab. Sys.*, 30(1995)133-146.
- [7] Tauler, R., Lacorte, S., and Barcelo, D. "Application of multivariate curve self-modeling curve resolution for the quantitation of trace levels of organophosphorous pesticides in natural waters from interlaboratory studies", *J. of Chromatogr. A*, 730(1996)177-183 .
- [8] Gemperline, P.J. "A priori estimates of the elution profiles of the pure components in overlapped liquid chromatography peaks using target factor analysis", *J. Chem. Inf. Comput. Sci.* 24 (1984) 206-212.
- [9] Vandeginste, B.G.M., Derks, W., and Kateman, G. "Multicomponent self-modelling curve resolution in high-performance liquid chromatography by iterative target transformation analysis", *Anal. Chim. Acta*, 173 (1985) 253-264.
- [10] Kvalheim, O.M., and Liang, Y.Z. "Heuristic evolving latent projections: resolving two-way multicomponent data. 1. Selectivity, latent-projective graph, datascope, local rank, and unique resolution", *Anal. Chem.*, 64 (1992) 936-946.
- [11]. Liang, Y.Z., Kvalheim, O.M., Keller, H.R., Massart, D.L., Kiechle, P., and Erni, F. "Heuristic evolving latent projections: resolving two-way multicomponent data. 2. Detection and resolution of minor constituents", *Anal. Chem.*, 64 (1992) 946-953.
- [12] Van Zomeren, P.V., Darwinkel, H., Coenegracht, P. M. J., and de Jong, G.J. "Comparison of several curve resolution methods for drug impurity profiling using high-performance liquid chromatography with diode array detection", *Anal. Chim. Acta*, 487(2003)155-170.
- [13] Lee, DD and Seung, H. "Learning the parts of objects by non-negative matrix factorization", *Nature*, 401(1999) 788-791.
- [14] Lee, DD. and Seung, H. "Algorithms for non-negative matrix factorization", *Adv. Neural Inf. Process. Syst.*, 13 (2001) 556-562.
- [15] Hoyer, P. O. "Non-negative Matrix Factorization with Sparseness Constraints", *Journal of Machine Learning Research*, 5 (2004) 1457-1469.
- [16] Li, S. Z., Hou, X. W., Zhang, H. J., and Cheng Q. S. "Learning Spatially Localized, Parts-based Representation", *International Conference on Computer Vision and Pattern Recognition*, (2001) 207- 212.
- [17] Zafeiriou, S., Tefas, A., Buciu, I., and Pitas, I. "Exploiting Discriminant Information in Nonnegative Matrix Factorization with Application to Frontal Face Verification", *IEEE Trans. on Neural Networks*, 17(3) (2006) 683-695.
- [18] Buciu, I., and Pitas, I. "A New Sparse Image Representation Algorithm Applied to Facial Expression Recognition", *IEEE Workshop on Machine Learning for Signal Processing*, (2004) 539- 548.
- [19] Schmidt, M. N. and Morup, M. "Nonnegative Matrix Factor 2- D deconvolution for blind single channel source separation, Independent Component Analysis and Blind

- Signal Separation” Lecture Notes in Computer Science, 3889/2006 (2006) 700-707.
- [20] Hong-Tao Gao, Tong-Hua Li, Kai Chen, Wei-Guang Li, and Xian Bi, ”Overlapping spectra resolution using non-negative matrix factorization”, *Talanta*, 66 (2005) 65– 73.
- [21] Liu Mingyu, Ji Hongbing, and Zhao Chunhong, Non negative Matrix Factorization and Its Application in EEG Signal Processing, *IEEE Xplore*, 978-1-4244-1748-3/08.
- [22].Li, H., Adali, T., Wang, W., Emge., D., and Cichocki, A. “Non-negative matrix factorization with orthogonality constraints and its application to Raman spectroscopy”, *J. of VLSI Signal Processing*, 48 (2007) 83–97.
- [23] Shin-Do Kim, Chang-Hwan Kim, Jin-Su Park, Jeong-Joo Lee, “A Study on the Peak Separation of Acetone and Acrolein based on High-Performance Liquid Chromatography (HPLC) Method”, *Bull. Korean Chem. Soc.* 30 (2009) 2011-2016.
- [24] Yuan, Z., and Oja, E “Projective nonnegative matrix factorization for image compression and feature extraction”, 14th Scandinavian Conference on Image Analysis, (2005) 333–342.
- [25] Morup, M., Madsen, K.H. and Hansen, L.K. “Shifted Non-negative Matrix Factorization”, *IEEE Workshop on Machine Learning for Signal Processing*, 2007.
- [26] Bucak, S.S., Günsel.B., and Gursoy, O.”Incremental non-negative matrix factorization for dynamic background modeling”, *ICEIS 8th International Workshop on Pattern Recognition in Information Systems*, 2007.
- [27] Morup, M., Hansen, L.K., and Arnfred, S.M. “Algorithms for sparse higher order non-negative matrix factorization(HONMF)”, *Technical Report*, 2006
- [28] Olshausen, B.A., and Field, D.J. “Emergence of simple-cell receptive field properties by learning a sparse code for natural images”, *Nature*, 381(1996)607-609.
- [29] Boutsidis. C., and Gallopoulos, E., “On SVD-based initialization for nonnegative Matrix factorization”, *Tech. Report, HPCLAB-SCG-6/08-05*, University of Patras, Patras, Greece, 2005.