ORIGINAL PAPER

An algorithm for overlapped chromatogram separation

S. Anbumalar · R. Ananda Natarajan ·
P. Rameshbabu

Abstract Non-negative matrix factorization (NMF) is a recently developed method
for real time data analysis. In the past it has been used for facial recognition and
spectral data analysis. Most of the NMF algorithms do not converge to a stable limit
point and uniqueness in results is also a problem in NMF. To improve the conver-
gence, a new NMF algorithm with modified multiplicative update (ML-NMFmse) has
been proposed in this work for strongly overlapped and embedded chromatograms
separation. To get same results for all the runs, instead of random initialization, three
different initialization methods have been used namely, ALS–NMF (robust initial-
ization), NNDSVD based initialization and EFA based initializations. The proposed
ML-NMFmse algorithm is applied on the simulated and experimental overlapped chro-
matograms obtained for acetone and acrolein mixture, using Gas Chromatography–
Flame Ionization Detector. Before applying NMF, Principal Component Analysis
(PCA) was applied to determine number of components in the mixture taken. The
result of proposed ML-NMFmse is compared with that of existing Multivariate Curve
Resolution-Alternating Least Squares method in optimal conditions for both the algo-
rithms. In the case of embedded chromatogram, the proposed ML-NMFmse with
Robust method (ALS-NMF) of initialization performs better than all other methods.
For a resolution of severely overlapped chromatograms, the proposed ML-NMFmse
with NNDSVD method of initialization outperforms all other methods.

S. Anbumalar (✉)
Department of Electrical and Electronics Engineering, Sri Manakula Vinayagar
Engineering College, Puducherry 605107, India
e-mail: saravanan.anbumalar@gmail.com

R. Ananda Natarajan · P. Rameshbabu
Department of Electronics and Instrumentation Engineering,
Pondicherry Engineering College, Puducherry 605014, India
e-mail: ananda_natarajan@pec.edu

1 Introduction

Lee and Seung [1,2] suggested NMF algorithms. Several works have been carried out
in NMF for image analysis [3–6], audio signal separation [7], spectral resolution [8]
and signal separation in bio-medical applications [9]. A chemo metric application of
the NMF method is proposed by Li et al. [10] to detect chemical compounds from a
chemical substance represented through Raman spectroscopy. Other NMF extensions
also exist which includes, projective NMF [11], shifted NMF [12], incremental NMF
[13] and sparse higher order NMF [14] etc. Lee and Seung's NMF [8] has been applied
on two components GC-MS spectra resolution in which smoothness, unimodality and
sparseness constraints have been imposed based on the characteristics of the signal.
This inspiring and interesting application of NMF has motivated to test NMFmse,
which was not used so far for the deconvolution of overlapping chromatograms.
The already existing multivariate resolution techniques like Evolving Factor Analysis
(EFA) [15,16], SIMPLe-to-use Interactive Self-modeling Mixture Analysis (SIM-
PLISMA) [17], Multivariate Curve Resolution–Alternating Least Squares (MCR–
ALS) [18,19], Iterative Target Transformation Factor Analysis (ITTFA) [20] and
Heuristic Evolving Latent Projections (HELP) were applied to many fields. How-
ever, in all the above methods, the accuracy of deconvolution is less if the overlap is
high.

Initially, NMF algorithms have been used by the authors [21–23] for the separation
of overlapped chromatogram of acetone and acrolein mixture by taking random initial
matrices. But, there is no convergence of the algorithm to a stable limit point [24–26].
The algorithm results with different resolved chromatograms (differs in shape or in
eluting time) in different runs and same results in some runs due to random initial
matrices. It makes the analysis difficult. Hence, instead of random initial matrices, fixed
matrices using some initialization techniques have been proposed in this work. Three
such initialization strategies namely ALS-NMF (robust initialization) [27], NNDSVD
based initialization [28] and EFA based initialization are used, compared and the best
initialization method for each cases taken is identified.

To improve the convergence of NMFmse algorithm [22], the multiplicative update
has been modified [24] and unimodality and selectivity constraints have been imposed
in this proposed modified NMFmse (ML-NMFmse) algorithm. Constraints have also
been imposed to handle any zero value in the data.

Initially, the proposed ML-NMFmse algorithm has been extended for the sep-
aration of simulated overlapped chromatograms of acetone and acrolein mixture.
The results are encouraging. Then the proposed ML-NMFmse algorithm has been
applied on the experimental overlapping chromatogram obtained for acetone and
acrolein mixture using Gas chromatography–Flame Ionization Detector (GC–FID).
The results of proposed ML-NMFmse are compared with the existing Multivariate
Curve Resolution–Alternating Least Square (MCR–ALS) method. The results show

that the proposed ML-NMFmse is suitable for severely overlapped and embedded chromatograms' resolution.

2 Materials and methods

2.1 Materials

2.1.1 Instrumental condition

Gas chromatography with Flame Ionization Detector was used with the detector temperature of 150 °C. Capillary Column (30 m ×0.25 mm ID, BP5) was used at a temperature of 100 ° C. The Temperature was programmed as 40 °C for 2 min, 5 °C/min, 68 °C for 2 min, 10 °C/min, 100 °C for 2 min. Injector temperature was kept as 150 °C. Nitrogen was used as a carrier gas at a flow rate of 90 ml/min. All the graphs are obtained in the software provided by Iris 32 chromatography, running on PC with Intel (R) Pentium4 CPU 2.00 GHz and 1 GB 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.

2.1.2 Experiment

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 4 different concentrations. The mixture was vigorously stirred separately under ice cold condition for half an hour and $0.2\,\mu$L of these mixtures were injected and analyzed using GC–FID.

2.2 Method

2.2.1 Non-negative matrix factorization

NMF is a technique for decomposing a non-negative matrix A into two non-negative matrices W and H as shown in Eq. (1).

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

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 be approximately equal to 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. The number of components r in an unknown mixture is not known. So, PCA is required to find the number of components r in the mixture. Resultant is a compressed version of the original data matrix. NMF allows only additive operations to provide

meaningful decomposition. When PCA is used for decomposition, two factorized matrices will contain positive and negative entries, 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 W to represent chromatogram in this work.

2.2.2 NMFmse algorithm

NMFmse algorithm was suggested by Lee and Seung [9] with the following objective function which has to be minimized, subject to the constraints W, H > 0.

$$\text{Objective function} \, (f) = \quad (1/2) \sum_{i=1}^{m} \sum_{j=1}^{n} \left(A_{ij} - (WH)_{ij} \right)^2 \qquad (2)$$

The multiplicative update rules were given in [23].

Several runs of the above said NMFmse algorithm with random initialization will give different answers in each run. Hence, it is important to have efficient and consistent initial matrices W and/or H because the solution and convergence provided by NMF algorithms highly depends on initial conditions.

The multiplicative update of NMFmse [22] do not assure convergence to a stationary point [24–26].There are difficulties which exist in the multiplicative algorithm due to zero values in the matrices [24].

2.2.3 ML-NMFmse algorithm

In this proposed ML-NMFmse algorithm, the update rules of NMFmse [22] have been replaced by the modified update to improve the convergence to a stable limit point. To speed up the convergence and to get same results in all runs, ALS-NMF (Robust initialization) [27], NNDSVD [28] and EFA based initialization strategies have been tested in this proposed algorithm for the separation of overlapped chromatograms.

The modified algorithm for the minimization of the objective function (2) is given as follows:

 (i) Assign $\varepsilon > 0$ and $\delta > 0$
 (ii) Initialize $W_{iq}^1 \geq 0$ and $H_{qj}^1 \geq 0$, $\forall i, q, j$, using Robust initialization or NNDSVD or EFA.
   (a) The following steps have to be used in Robust initialization to estimate initial matrices.

> Step: 1 Generate the initial matrices W and H based on the output from ALS-NMF algorithm (20 iterations) in which W and H are given as

$$H \leftarrow \left[(W^T W)^{-1} W A \right]_+ \quad W \leftarrow \left[A H^T (H H^T)^{-1} \right]_+$$

> Step: 2. Run the NMF algorithm for the initial matrices (10–20 iterations) and it will provide initial estimates of the matrices $W^{(r)}$ and $H^{(r)}$.
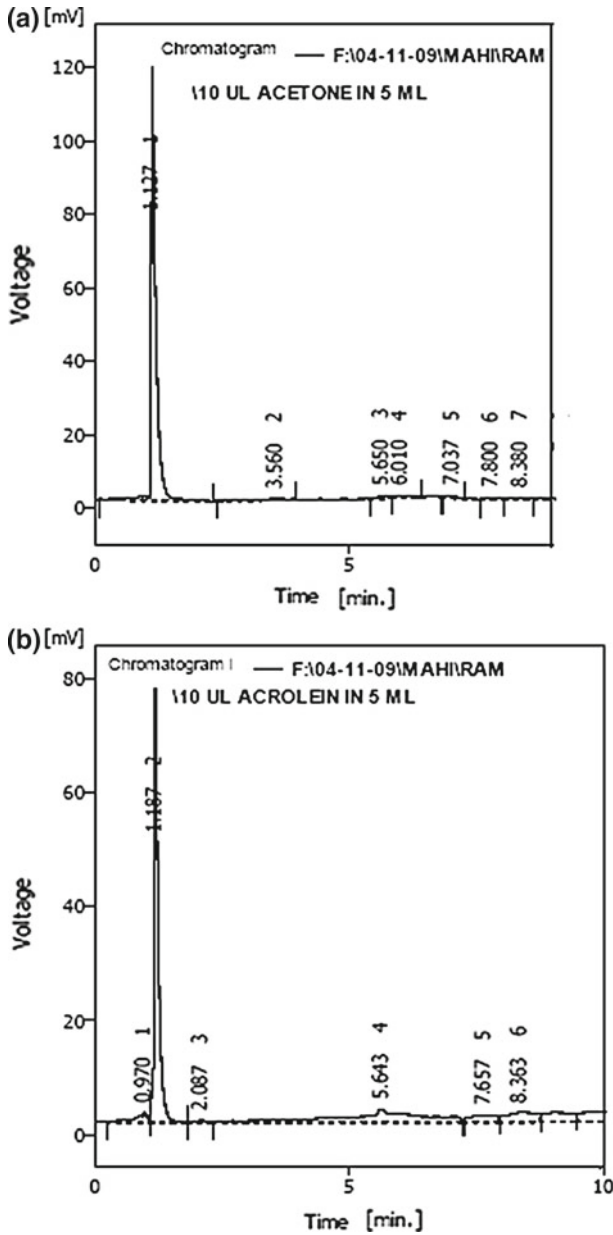
Fig. 1 **a** Real experimental chromatogram of acetone standard. Retention time of acetone 1.128 min.
b Real experimental chromatogram of acrolein standard. Retention time of acrolein 1.188 min

Step: 3. Select the estimates $W^{(rmin)}$ and $H^{(rmin)}$ corresponding to the lowest value of the cost function among the R runs as initial values for the final factorization.
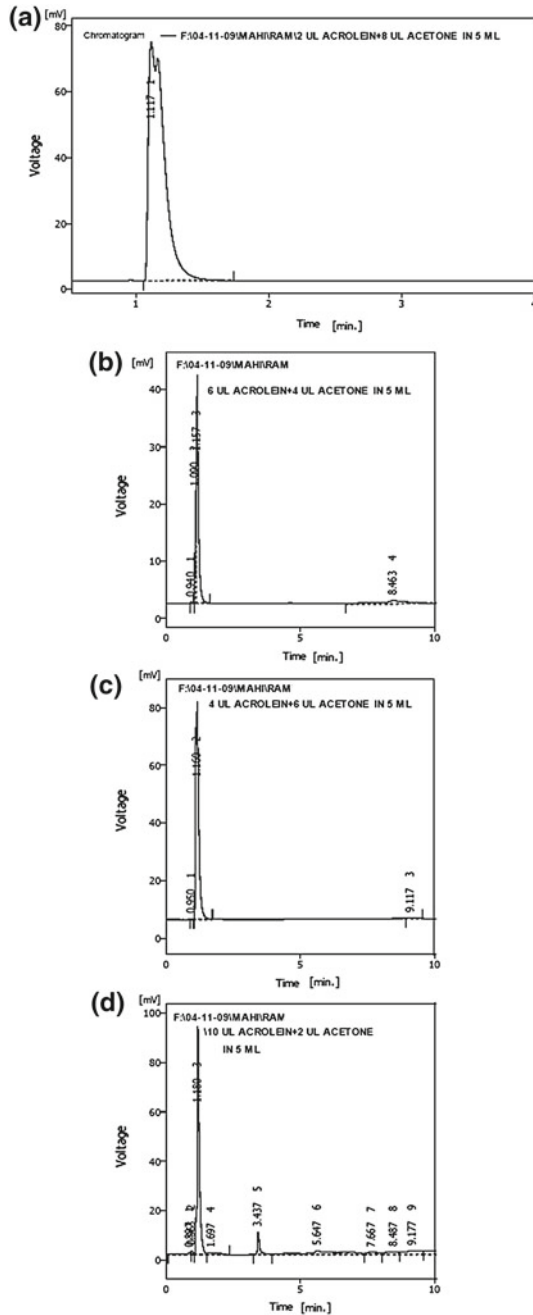
Fig. 2 a Real experimental severely overlapped chromatogram of acetone and acrolein mixture (2 μl acrolein and 8 μl acetone). **b** Real experimental severely overlapped chromatogram of acetone and acrolein mixture (6 μl acrolein and 4 μl acetone). **c**. Real experimental severely overlapped chromatogram of acetone and acrolein mixture (4 μl acrolein and 6 μl acetone). **d**. Real experimental severely overlapped chromatogram of acetone and acrolein mixture (10 μl acrolein and 2 μl acetone)
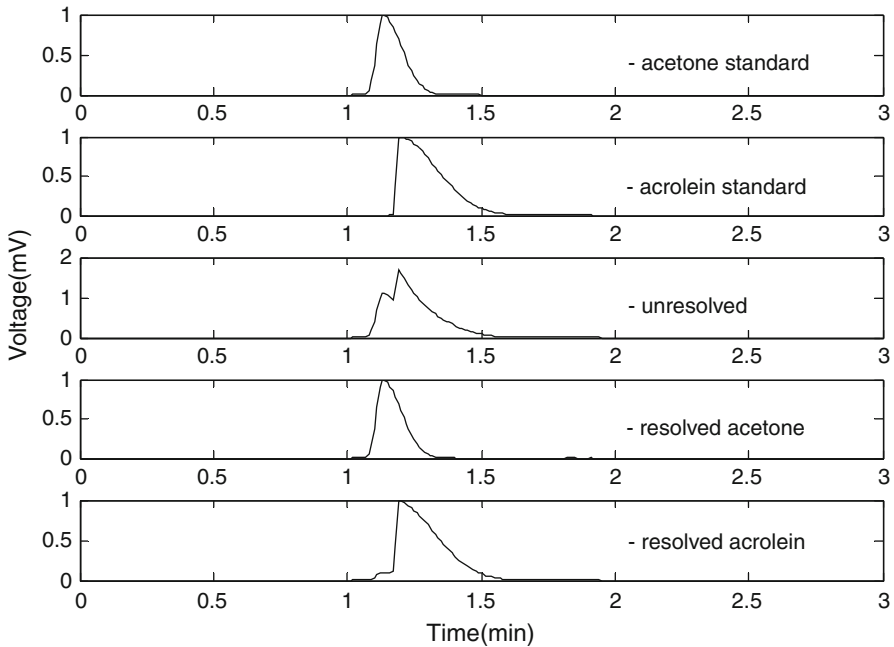
Fig. 3 Unresolved (partially overlapped) chromatogram of acetone and acrolein mixture and resolved chromatograms using proposed ML-NMFmse algorithm with EFA based initialization

Assume the estimate obtained in this initialization as an initial W and H matrices and follow the steps from (iii) to (v) for implementing proposed ML-NMFmse algorithm

b) Perform NNDSVD or EFA based initializations also to get the initial W and H matrices and use it separately for comparison.

(iii) For $k = 1, 2, \ldots$

a) If (W^k, H^k) is stationary, stop.
Else

$$H_{qj}^{k,n} = H_{qj}^k - \frac{\bar{H}_{qj}^k}{((W^k)^T W^k H^k)_{qj} + \varepsilon} \nabla_H f(W^k, H^k)_{qj}, \forall q, j \qquad (3)$$

$$W_{iq}^{k,n} = W_{iq}^k - \frac{\bar{W}_{iq}^k}{(\bar{W}^k H^{k,n}(H^{k,n})^T)_{iq} + \varepsilon} \nabla_W f(W^k, H^{k,n})_{qj}, \forall i, q \qquad (4)$$

where $\nabla_W f(W, H) = (WH - A)H^T$ and $\nabla_H f(W, H) = W^T(WH - A)$.

$$\frac{\bar{H}_{qj}^k}{((W^k)^T W^k \bar{H}^k)_{qj} + \varepsilon} \text{ is the step size}$$

Where

$$\bar{H}_{qj}^k \equiv \begin{cases} H_{qj}^k & if \ \nabla_H f(W^k, H^k)_{qj} \geq 0, \\ max(H_{qj}^k, \delta) & if \ \nabla_H f(W^k, H^k)_{qj} < 0 \end{cases}$$
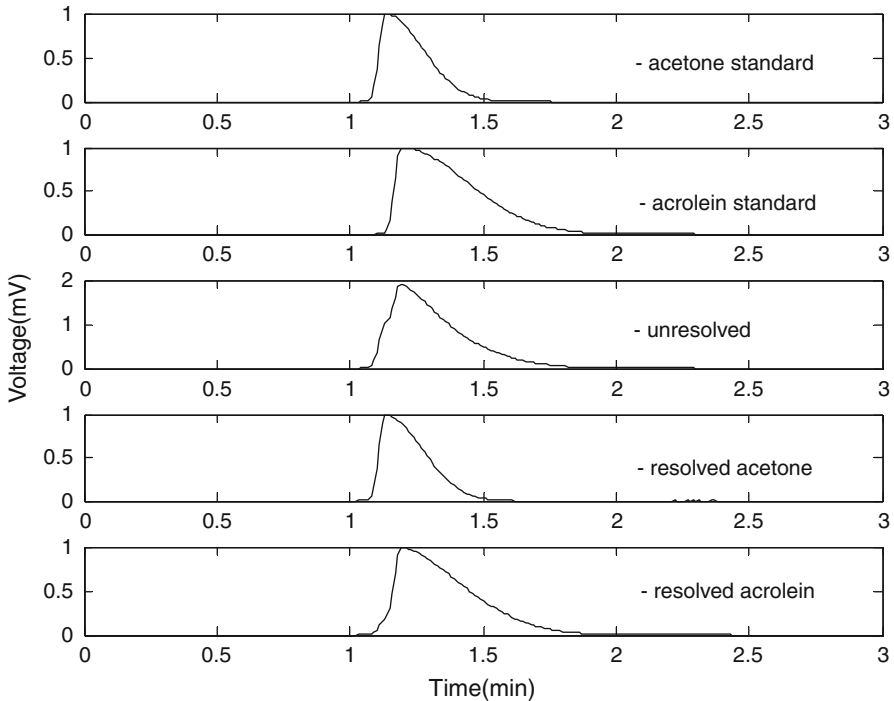
Fig. 4 Unresolved (severely overlapped) chromatogram of acetone and acrolein mixture and resolved chromatograms using proposed ML-NMFmse algorithm with EFA based initialization

   ε [29,30] and δ are the small positive numbers assumed to avoid division by zero.

b) Normalize $W^{k,n}$ to W^{k+1} so that W^{k+1}'s column sum is one.

c) Unimodality and selectivity constraints are additionally imposed into the algorithm so as to use the algorithm for chromatogram separation.

d) The iterative algorithm used has been stopped when the number of iterations achieves or exceeds a predefined maximum number of iteration.

 It is assumed that there is no need for peak shifting, base line correction and noise filtering.

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

(v) Repeat from (iii) to (iv) until convergence is achieved.

The modified multiplicative update overcomes the difficulties that exist in the convergence of NMFmse algorithm.

2.2.4 Procedure

Step 1: The detector output of each experiment has been taken as an individual column of matrix A. The experimental data were taken for mixtures of different concentration ratios to get a pseudo second order data. (Only one-way data can be taken with the help of detector available with us.) Hence, the shape and area
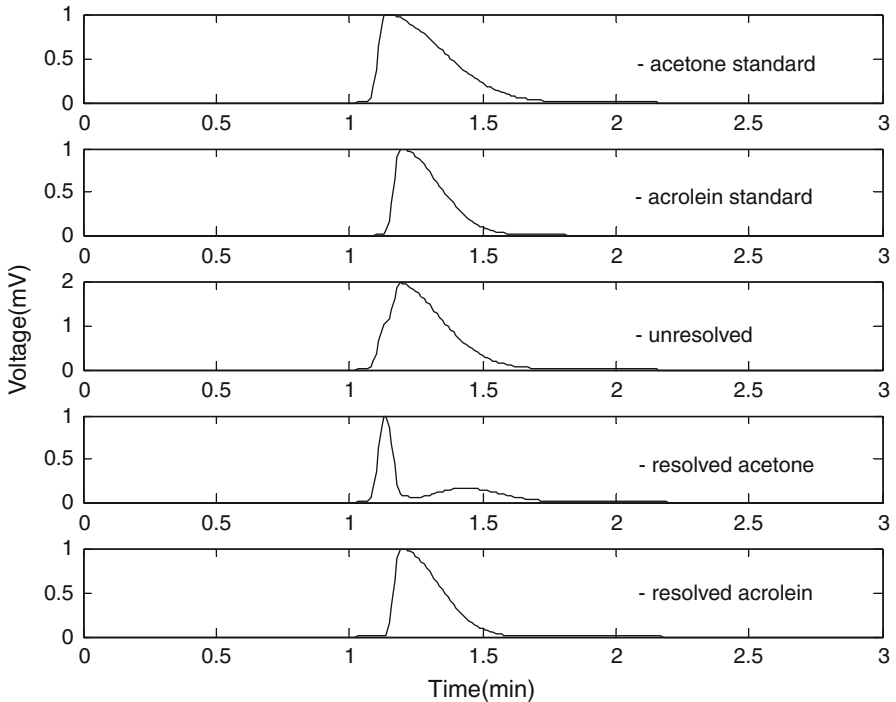
Fig. 5 Unresolved (embedded) chromatogram of acetone and acrolein mixture and resolved chromatograms using ML-NMFmse with EFA based initialization

Table 1 Performance measures of MCR–ALS and proposed ML-NMFmse algorithms on the separation of simulated partially overlapped chromatograms

Performance measures	ML-NMFmse with different initialization methods			MCR–ALS
	EFA	NNDSVD	Robust	
Correlation coefficient				
Resolved acetone	1.0000	1.0000	1.0000	**1.0000**
Resolved acrolein	0.9967	0.9967	0.9999	**1.0000**
Signal recovery (*dB*)				
Resolved acetone	52.1665	52.1992	49.7770	**49.5348**
Resolved acrolein	22.1431	22.1431	37.6111	**113.3708**

Bold values indicate that the algorithm gives better performance than the other algorithms given in table

of the overlapped chromatogram is based on its chemical concentration. Here, the shift in position or shape of the chromatogram couldn't be differentiated. Hence, preprocessing was not needed. The main focus of this work is resolving the overlapping components. It is proposed to use ML-NMFmse algorithm to perform the deconvolution of a data matrix A.

Table 2 Performance measures of MCR–ALS and proposed ML-NMFmse algorithms on the separation of simulated severely overlapped chromatograms

Performance measures	ML-NMFmse with different initialization methods			MCR–ALS
	EFA	NNDSVD	Robust	
Correlation coefficient				
Resolved acetone	**0.9999**	0.9999	0.9998	0.9999
Resolved acrolein	**0.9943**	0.9943	0.9943	0.9943
Signal recovery (dB)				
Resolved acetone	**39.7917**	36.2225	34.5066	35.8694
Resolved acrolein	**20.2171**	20.2163	20.2153	20.3776

Bold values indicate that the algorithm gives better performance than the other algorithms given in table

Table 3 Performance measures of MCR–ALS and proposed ML-NMFmse algorithms on the separation of simulated embedded chromatograms

Performance measures	ML-NMFmse with different initialization methods			MCR–ALS
	EFA	NNDSVD	Robust	
Correlation coefficient				
Resolved acetone	0.6915	0.5963	**0.9993**	0.5219
Resolved acrolein	0.9998	0.9526	**0.9563**	1.0000
Signal recovery (dB)				
Resolved acetone	2.6768	1.4826	**29.1994**	0.5967
Resolved acrolein	34.8281	10.5284	**10.8769**	52.1526

Bold values indicate that the algorithm gives better performance than the other algorithms given in table

Step 2: The size of matrix W is based on the number of components k in a mixture. The unknown number of components k, in a mixture taken for analysis, was determined by principal component analysis (PCA).

Step 3: The W and H matrices have been initialized with Robust initialization (or) NNDSVD based initialization (or) EFA based initialization. The zero elements, if any, in the matrices (A, W or H) are replaced by a small positive number (ε). Hence, the iteration never breaks and the algorithm can converge to minima.

Step 4: The proposed ML-NMFmse algorithm is used to decompose the matrix A into W and H matrices.

Step 5: After decomposition, the columns of W matrix are the individual chemical's chromatogram of the mixture taken. The resolved chromatograms are compared with standard's chromatogram and the correlation coefficients are obtained. The signal recovery is calculated using the formula given by

$$Pr_i(dB) = 10\log_{10} \frac{\sum_{t=1}^{p} [(S_i(t)/std(S_i))]^2}{\sum_{t=1}^{p} [(S_i(t)/std(S_i)) - (R_i(t)/std(R_i))]^2}$$
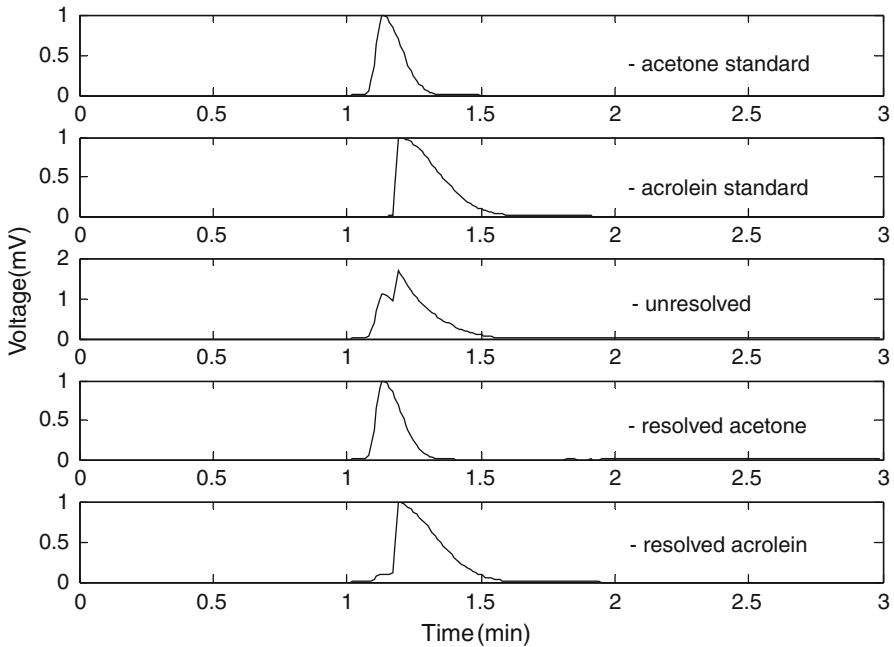
Fig. 6 Unresolved (partially overlapped) chromatogram of acetone and acrolein mixture and resolved chromatograms using proposed ML-NMFmse with NNDSVD based initialization

where

S_i chromatogram obtained for standard chemical (i.e., for acetone, acrolein in the experimental work; the chromatogram of individual component in the simulated mixtures)

R_i resolved signal;

Std standard deviation;

Pr_i (dB) closeness of ith resolved signal power with that of ith standard signal;

i 1,2…n; where 'n' is the number of components in the mixture.

t 1, 2…p; where 'p' is the end point of the chromatogram.

$S_i(1)$ detector output value at the starting point (t = 1) of the chromatogram for the ith standard chemical

$S_i(p)$ detector output value at the ending point (t = p) of the chromatogram for the ith standard chemical

$R_i(1)$ detector output value at the starting point (t = 1) of the chromatogram for the ith resolved chemical

$R_i(p)$ detector output value at the ending point (t = p) of the chromatogram for ith resolved chemical

The signal recovery gives information about how much the resolved chromatogram deviates from the standard chromatogram in retention time point of view (elution time) as well as in shape (quantification point of view) .The higher positive value says that the resolved signal matches with the standard chromatogram well. If one resolution
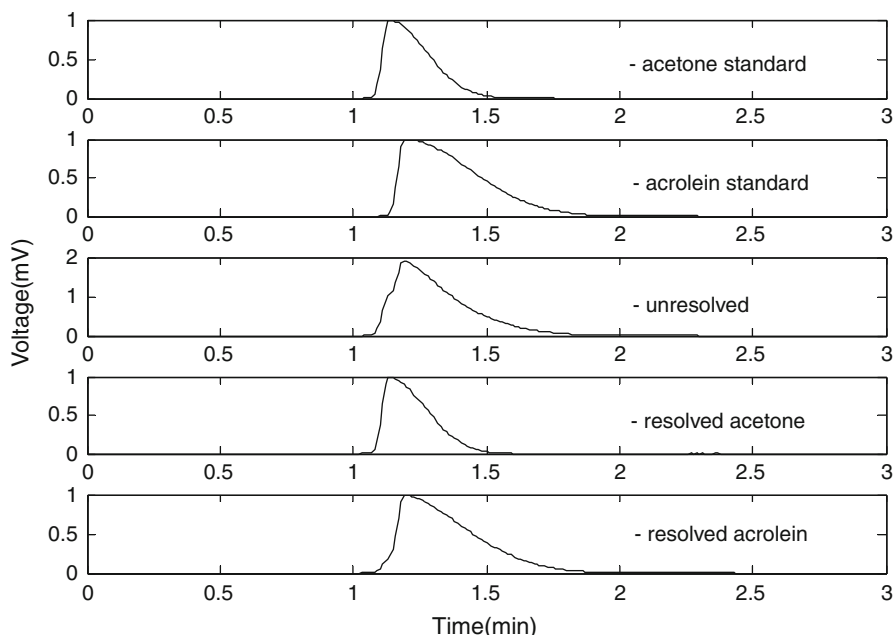
Fig. 7 Unresolved (severely overlapped) chromatogram of acetone and acrolein mixture and resolved chromatograms using proposed ML-NMFmse with NNDSVD based initialization

method produces 20 dB and another method produces 30 dB, second method is better than the first method. This is the performance measure used in signal processing to measure the efficiency of signal recovery. It has been first applied in our work as a performance measure in chromatogram separation.

All the above procedure was implemented using **Matlab** software.

Initially, the algorithm is tested on the simulated chromatograms of three cases, i.e., partially overlapped, severely overlapped and embedded peaks. One dimensional, GC–FID overlapped chromatogram containing tailing peaks (so that, it can be similar to the experimental chromatograms) of acetone and acrolein were simulated using cross product multiplication of Gaussian functions. The proposed ML-NMFmse algorithm has been performed to deconvolute the data matrix *A* of simulated overlapping and embedded chromatograms.

Finally, the proposed ML_NMFmse algorithm is applied on the experimental severely overlapping chromatograms of acetone and acrolein mixtures. All the results are compared with that of MCR–ALS method.

3 Results and discussion

The chromatograms obtained for standards, acetone and acrolein, are shown in Fig. 1a, b respectively. The mixed solution of acetone and acrolein at four different concen-
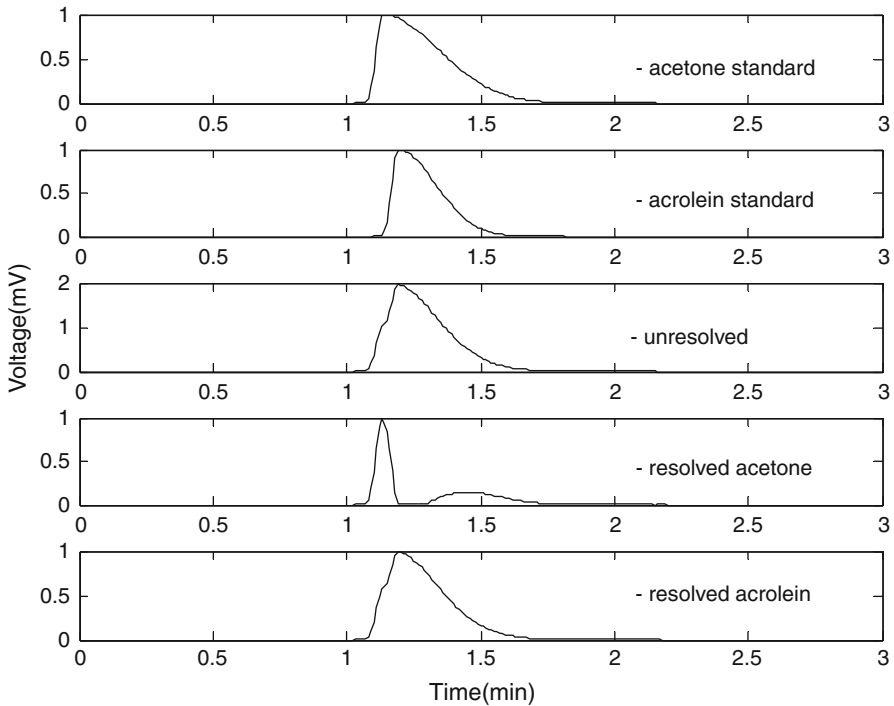
Fig. 8 Unresolved (embedded) chromatogram of acetone and acrolein mixture and resolved chromatograms using proposed ML-NMFmse with NNDSVD based initialization

trations were prepared. The mixer concentrations were (i) $2\,\mu L$ acrolein and $8\,\mu L$ acetone, (ii) $4\,\mu L$ a acrolein and $6\,\mu L$ acetone, (iii) $10\,\mu L$ acrolein and $2\,\mu L$ acetone, and (iv) $6\,\mu L$ acrolein and $4\,\mu L$ acetone respectively. The chromatograms obtained are as shown in Fig. 2a–d. It is found that the chromatograms are overlapped due to close retention time of acetone and acrolein. The chromatogram has some unknown peaks due to some impurities in the chemical mixture. But, focus is only on the overlapping region (from 1 to 2 min), which has to be resolved to separate severely overlapped acetone and acrolein, The remaining parts of the chromatogram are not considered in this work for analysis.

Three different strategies namely EFA, NNDSVD and Robust initialization are used to estimate the initial matrices of the proposed ML-NMFmse algorithm. Then the algorithm was used to resolve the overlapped chromatograms.

3.1 Simulation study using proposed ML-NMFmse algorithm with EFA based initialization

Initially, the EFA based initialization has been used to initialize the matrices of the proposed ML-NMFmse algorithm and used to resolve partially overlapped, severely overlapped and embedded chromatograms. The results are shown in Figs. 3, 4 and 5.
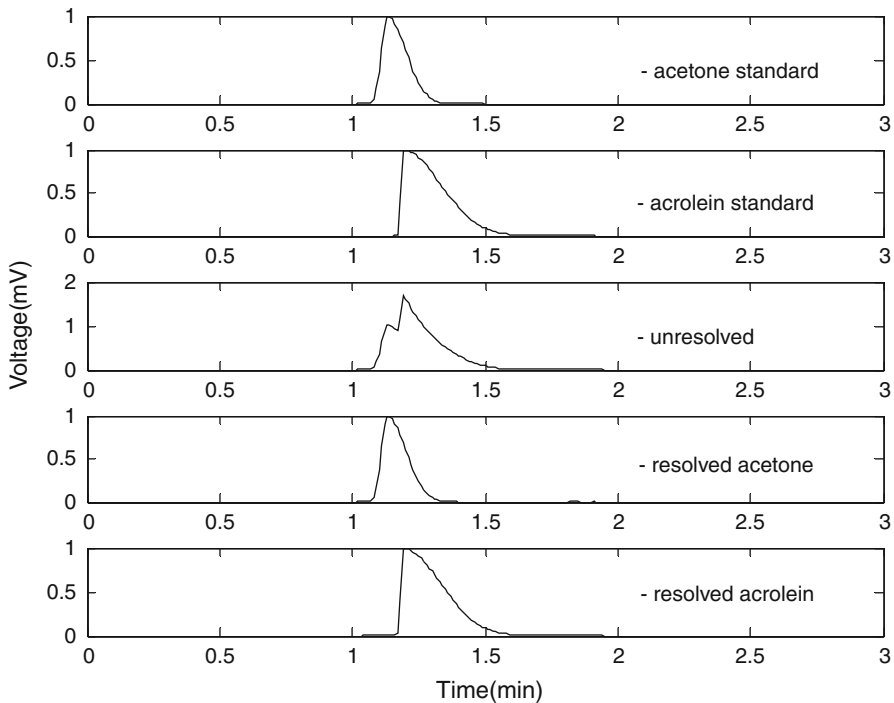
Fig. 9 Unresolved (partially overlapped) chromatogram of acetone and acrolein mixture and resolved chromatograms using proposed ML-NMFmse with Robust initialization

The resolved chromatograms are compared with standard's chromatogram and its correlation coefficient and signal recovery are given in Tables 1, 2 and 3.

3.1.1 Simulation study using proposed ML-NMFmse algorithm with NNDSVD based initialization

Secondly, the proposed ML-NMFmse algorithm has been initialized with NNDSVD method and used for overlapped chromatograms separation.

The proposed ML-NMFmse algorithm has been initialized with NNDSVD method and applied on partially overlapped, severely overlapped and embedded chromatograms for separation. The unresolved and the resolved chromatograms are shown in Figs. 6, 7 and 8. The resolved chromatograms are compared with standard's chromatogram and its correlation coefficient and signal recovery are given in Tables 1, 2 and 3.

3.1.2 Simulation study using proposed ML-NMFmse algorithm with Robust initialization

In the third case, the proposed ML-NMFmse algorithm has been initialized with Robust initialization and used for the separation of overlapped chromatograms.
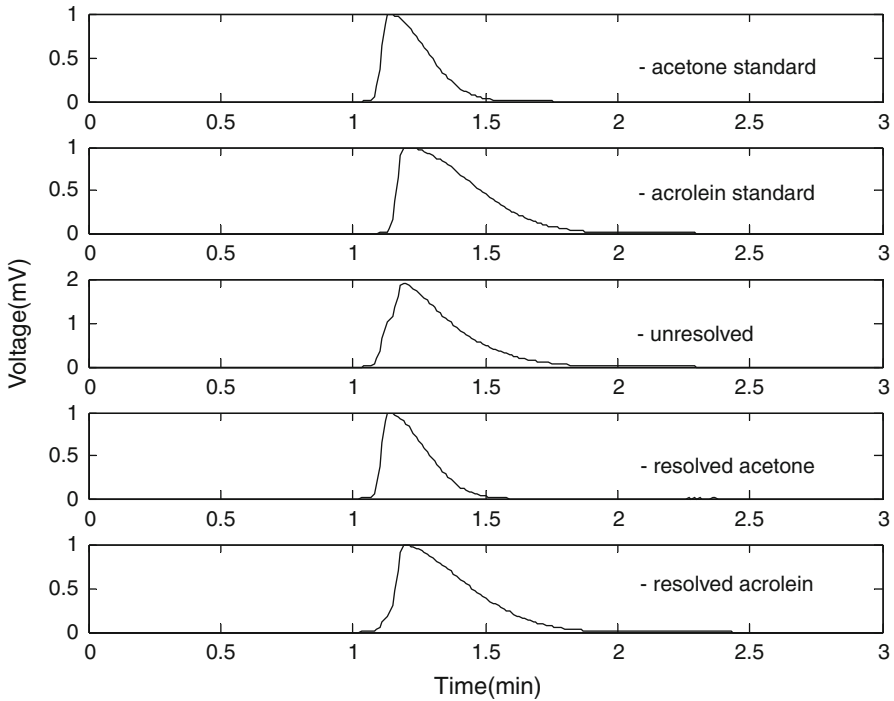
Fig. 10 Unresolved (severely overlapped) chromatogram of acetone and acrolein mixture and resolved chromatograms using proposed ML-NMFmse with Robust initialization

The proposed ML-NMFmse algorithm (Robust initialization) was used to separate partially overlapped, severely overlapped and embedded chromatograms. The results are shown in Figs. 9, 10 and 11. The resolved chromatograms are compared with the standard's chromatogram and the correlation coefficient and signal recovery of acetone and acrolein are listed in Table 1, 2 and 3.

The resolved components are indicating the retention time of chemicals only but the magnitude of the resolved chromatograms has scaling ambiguity**.** Hence the scale of the resolved chromatogram has been normalized and presented in all results. Even though the correlation coefficients are same for the resolved acetone and acrolein in some cases, their signal recoveries differ due to the shift in peak position and shape of the resolved chromatogram from its standard. Hence, the signal recovery seems to be a useful performance measure in chromatogram resolution. The performance of MCR–ALS method is better than the proposed ML-NMFmse algorithm for the partially overlapped chromatogram resolution which is evident from Table 1. The performance of proposed ML-NMFmse algorithm with EFA based initialization is better for severely overlapped case as shown in Table 2. The performance of proposed ML-NMFmse algorithm with Robust initialization is better for embedded case which is given in Table 3. Note that the resolution of chromatogram by MCR–ALS algorithm becomes poor as the severity of overlap increases. On the other hand,the proposed
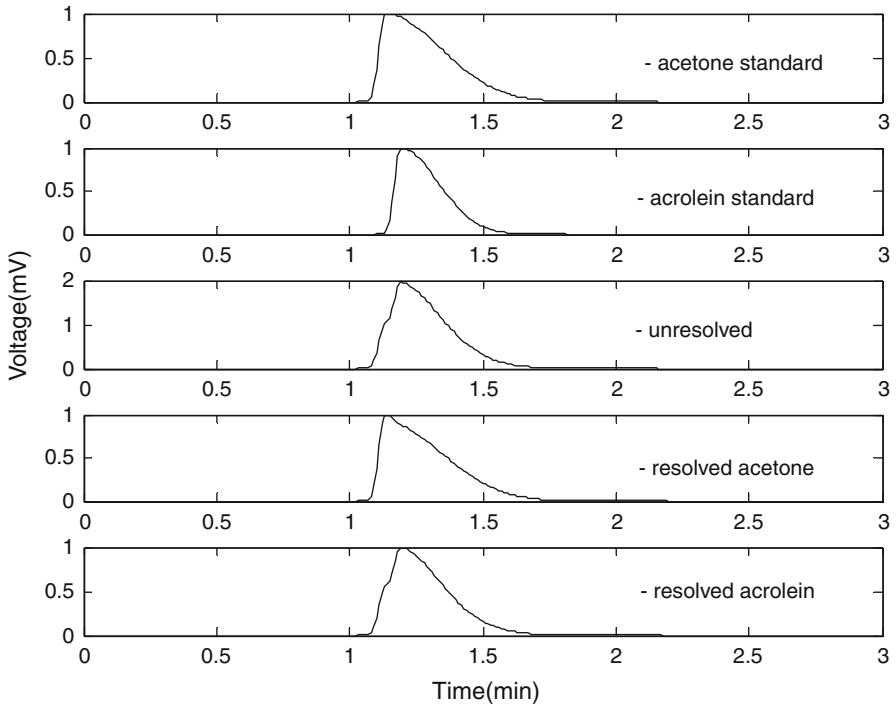
Fig. 11 Unresolved (embedded) chromatogram of acetone and acrolein mixture and resolved chromatograms using proposed ML-NMFmse with Robust initialization
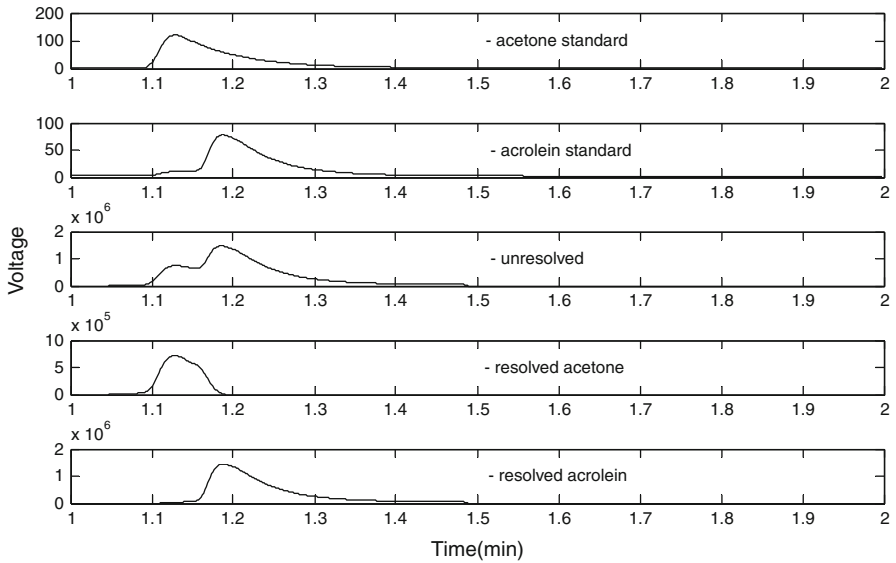

Fig. 12 Experimental overlapped chromatogram of acetone and acrolein mixture and resolved chromatograms using MCR–ALS method
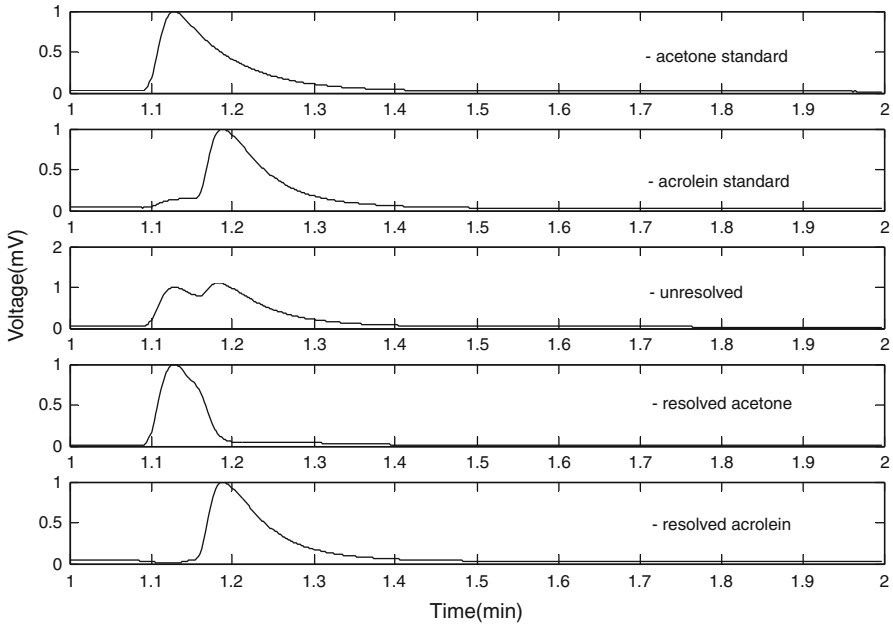
Fig. 13 Experimental overlapped chromatogram of acetone and acrolein mixture and resolved chromatograms using proposed ML-NMFmse (**Robust initialization**)

Fig. 14 Experimental overlapped chromatogram of acetone and acrolein mixture and resolved chromatograms using proposed ML-NMFmse (**NNDSVD based initialization**)
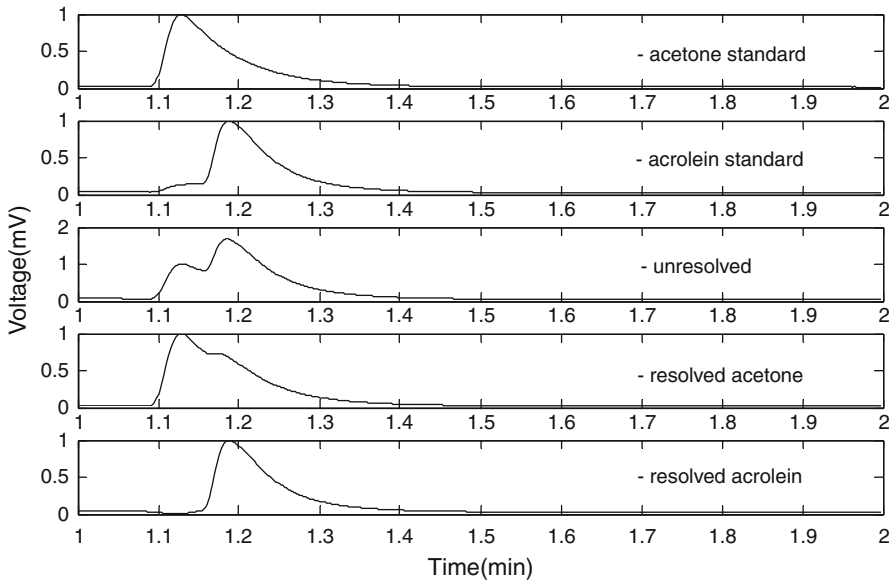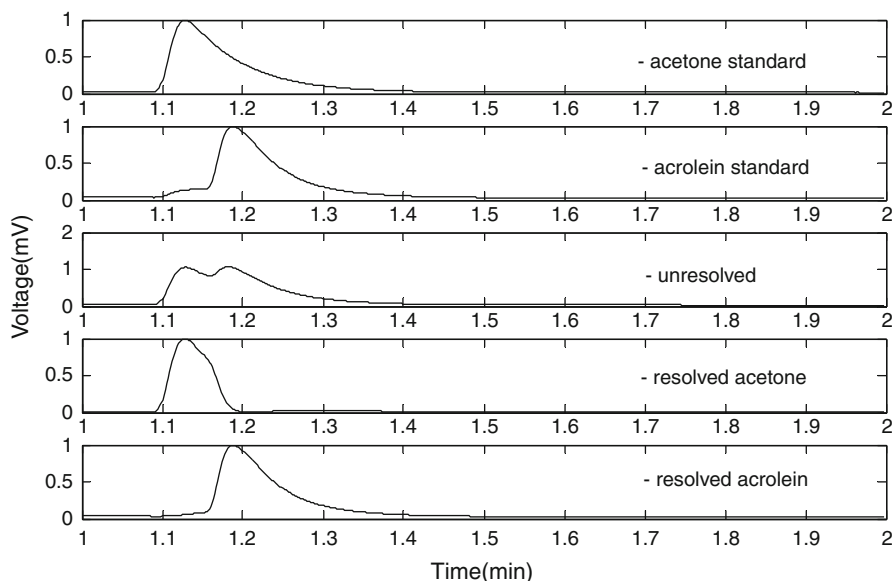
Fig. 15 Experimental overlapped chromatogram of acetone and acrolein mixture and resolved chromatograms using proposed ML-NMFmse (**EFA based initialization**)

method with Robust initialization gives better resolution as the severity of overlap increases.

Table 4 Performance measures of different initialization strategies applied in proposed ML-NMFmse algorithm for experimental chromatogram separation

Performance measures	ML-NMFmse with different initialization methods			MCR–ALS
	EFA	NNDSVD	Robust	
Correlation coefficient				
Resolved acetone	0.9087	**0.9875**	0.9276	0.9043
Resolved acrolein	0.9970	**0.9917**	0.9920	0.9948
Signal recovery (dB)				
Resolved acetone	7.6430	**16.7159**	8.6846	7.5689
Resolved acrolein	23.1455	**18.6236**	18.7715	20.5206

Bold values indicate that the algorithm gives better performance than the other algorithms given in table

3.2 Experimental study on ML-NMFmse

The overlapping chromatograms were obtained by GC–FID for acetone and acrolein mixtures. The experiments were conducted under condition mentioned in Sect. 2.1. Chromatograms obtained are shown in Fig. 2. It can be analyzed qualitatively with the
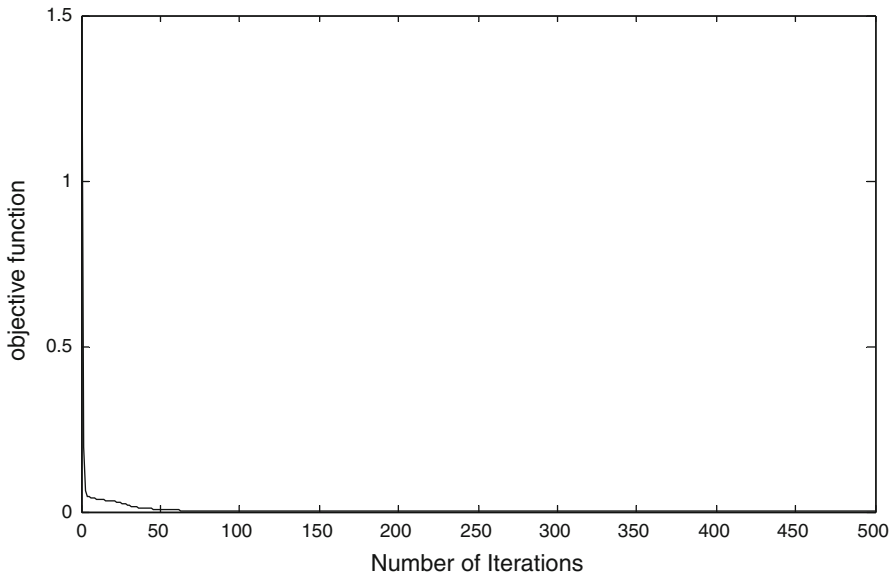
Fig. 16 Objective function versus number of iterations for proposed ML-NMFmse implementation using Robust initialization

help of standard chromatograms. But, the obtained chromatograms are overlapping due to almost similar retention times of acetone and acrolein (i.e., 1.128 and 1.188 min respectively).Hence, during the peak identification, only one peak has been identified instead of two in the overlapped region. If these overlapping peaks were not resolved, the chromatograms of mixtures obtained could not get a good match with that of a pure component in the database. Hence to resolve the individual components, initially, the MCR–ALS algorithm was applied on the experimental overlapped chromatograms of acetone and acrolein mixture and the results are shown in Fig. 12. The resolved components are compared with the standard chromatograms. The correlation coefficients are 0.9043, 0.9948 and the signal recoveries (Pr) are 7.5689 dB, 20.5206 dB for the acetone and acrolein respectively.

The proposed ML-NMFmse algorithm was then applied on the experimental overlapped chromatograms obtained to resolve acetone and acrolein.

The W and H matrices were initialized with Robust initialization and then the proposed ML-NMFmse was used to find the resolved components. The standard chromatograms of acetone and acrolein, overlapped chromatogram of acetone and acrolein mixture and the resolved chromatograms are shown in Fig. 13. The same dataset was again tested using proposed ML-NMFmse by taking EFA estimate as initial matrices and the resolved results are shown in Fig. 14. The results are same for all the runs. Then the same chromatograms were resolved using proposed ML-NMFmse by taking NNDSVD based estimate as initial matrices and the results are shown in Fig. 15 and are same for all the runs. The correlation coefficients and the signal recoveries (Pr) for the resolved acetone and acrolein are calculated and given in Table 4. The results show that a better solution can be obtained with proper initial estimate of W and

H matrices. Figure 16 shows the convergence history of the proposed ML-NMFmse with Robust initialization. It shows that the algorithm converges to a stable limit point.

On comparing the resolution of MCR–ALS and proposed algorithm, latter gives better resolution than that of MCR–ALS. This is evident from the correlation coefficient and signal recovery of resolved components given in Table 4. The comparison of different initialization methods on the experimental chromatogram separation shows that the NNDSVD based initialization gives better performance than the other initialization methods.

4 Conclusion

In this paper, the proposed ML-NMFmse algorithm with three different initialization methods were used for the overlapped chromatogram separation and the results were compared with MCR–ALS method.

The application of the above said algorithms on the simulated overlapped chromatograms of acetone and acrolein mixture shows that the proposed method fails to perform better than MCR–ALS method for partially overlapped chromatograms. But performs better than MCR–ALS method for severely overlapped and embedded chromatograms. In the case of embedded chromatogram, Robust method of initialization performs better than all other methods. For a resolution of severely overlapped chromatogram, NNDSVD method of initialization outperforms all other methods. This is evident from simulation study as well as experimental study.

The results show that the proposed ML-NMFmse algorithm converges to a stable limit point and the resolution results are found to be same for all runs, which are the advantages of the proposed ML-NMFmse method.

References

1. D.D. Lee, H. Seung, Nature **401**, 788–791 (1999)
2. D.D. Lee, H. Seung, in *Advances in Neural Information Processing NIPS* 2000, (2000)
3. P.O. Hoyer, J. Mach. Learn. Res. **5**, 1457–1469 (2004)
4. S.Z. Li, X.W. Hou, H.J. Zhang, Q.S. Cheng, in *International Conference on Computer Vision and Pattern Recognition* (2001), pp. 207–212
5. S. Zafeiriou, A. Tefas, I. Buciu, I. Pitas, IEEE Trans. Neural Netw. **17**(3), 683–695 (2006)
6. I. Buciu, I. Pitas, in *IEEE Workshop on Machine Learning for Signal Processing*, pp. 539–548 (2004)
7. M.N. Schmidt, M. Morup, *Nonnegative Matrix Factor 2-D Deconvolution for Blind Single Channel Source Separation.* Independent Component Analysis and Blind Signal Separation. Lecture Notes in Computer Science, vol. 3889 (Springer, Heidelberg, 2006), pp. 700–706
8. H.-T. Gao, T.-H. Li, K. Chen, W.-G. Li, X. Bi, Talanta **66**, 65–73 (2005)
9. L. Mingyu, J. Hongbing, Z. Chunhong, *Non-negative Matrix Factorization and Its Application in EEG Signal Processing.* IEEE Xplore. ISBN: 978-1-4244-1748-3 (2008)
10. H. Li, T. Adali, W. Wang, D. Emge, A. Cichocki, J. VLSI Sign. Process. **48**(1–2), 83–97 (2007)
11. Z. Yuan, E. Oja, in *14th Scandinavian Conference on Image, Analysis* (2005) pp. 333–342

12. M. Morup, K.H. Madsen, L.K. Hansen, IEEE Workshop on Machine Learning for Signal Processing (2007)
13. S.S. Bucak, B. Gunsel, O. Gursoy, in *ICEIS 8th International Workshop on Pattern Recognition in Information Systems* (2007)
14. M. Morup, L.K. Hansen, S.M. Arnfred, *Algorithms for Sparse Non-negative Tucker Decompositions*. Technical Report (2006)
15. H. Gampp, M. Maeder, C.J. Meyer, A.D. Zuberbuhler, Talanta **32**, 1133–1139 (1985)
16. M. Maeder, Anal. Chem. **59**, 527–530 (1987)
17. W. Windig, J. Guilment, Anal. Chem. **63**, 1425–1432 (1991)
18. R. Tauler, A.K. Smilde, B.R. Kowalski, J. Chemometr. **9**, 31–58 (1995)
19. R. Tauler, Chemom. Intell. Lab. Sys. **30**, 133–146 (1995)
20. P.J. Gemperline, J. Chem. Inf. Comput. Sci. **24**, 206–212 (1984)
21. S. Anbumalar, P. Rameshbabu, R. Anandanatarajan, Chromatogram separation using Matrix decomposition. Int. J. Comput. Appl. **27**(3), 24–32 (2011)
22. S. Anbumalar, P. Rameshbabu, R. Anandanatarajan, An algorithm to resolve embedded chromatograms in chemical analysis. Int. J. Comput. **1**(4), 610–619 (2011)
23. S. Anbumalar, P. Rameshbabu, R. Anandanatarajan, Overlapped chromatograms separation using nonnegative Matrix factorization. Int. J. Comput. Electr. Eng. **3**(5), 654–658 (2011)
24. C.-J. Lin, *Information and Support Services Techincal Report*, Department of Computer Science, National Taiwan University (2005a)
25. E.F. Gonzales, Y. Zhang, *Accelerating the Lee-Seung Algorithm for Non-negative Matrix Factorization*. Technical Report (Department of Computational and Applied Mathematics, Rice University, 2005)
26. C.-J. Lin, *Information and Support Services Technical Report ISSTECH-95-013* (Department of Computer Science, National Taiwan University, 2005b)
27. A. Cichocki, R. Zdunek, A.H. Phan, S.-I. Amari, *Non-negative Matrix and Tensor Factorizations—Applications to Exploratory Multi-way Data Analysis and Blind Source Separation* (John Wiley & Sons Ltd., 2009, preprint)
28. C. Boutsidis, E. Gallopoulos, Pattern. Recon. **41**, 1350–1362 (2008)
29. P. O. Hoyer, in *Proceedings of IEEE Workshop on Neural Networks for Signal Processing* (2002) pp. 557–565
30. J. Piper, P. Pauca, R. Plemmons, M. Giffin, in *Proceedings of AMOS Technical Conference* (2004)