

Restoration for Comprehensive Two-Dimensional Gas Chromatography*

Jiazheng Shi Stephen E. Reichenbach
Computer Science and Engineering Department
University of Nebraska – Lincoln
Lincoln NE 68588-0115 USA

Abstract

This paper develops an approach for restoration of data from comprehensive two-dimensional gas chromatography (GCxGC), a powerful new technology for chemical separations. GCxGC restoration is required to separate coeluting (i.e., overlapping) peaks. The GCxGC process is modeled as a two-dimensional linear, shift-variant system, based on the properties of chemical compounds and gas chromatography. The model can account for nonhomogeneous peak shapes and separability of the two instrument columns. The restoration problem is formulated to minimize the difference between observed GCxGC data and an ideal separation, subject to the physically meaningful constraints: nonnegativity, volume preservation, and unimodality. The paper develops a constrained alternating least-squares (CALS) method for solving the restoration problem. Experimental results based on simulation and real GCxGC data indicate that the proposed model and CALS method perform well for GCxGC restoration.

1 Introduction

This paper develops an approach for restoration of data from comprehensive two-dimensional gas chromatography (GCxGC), a powerful new technology for chemical analysis [1]. GCxGC offers an order-of-magnitude improvement in separation capacity compared to traditional, one-dimensional GC, but for complex samples, peaks associated with some compounds still may coelute (i.e., overlap) in the data. Computer-based restoration methods are used to separate overlapping chromatographic peaks and produce better qualitative and quantitative results.

If GC is modeled as a linear system, restoration can be implemented by deconvolution to separate overlapping chromatographic peaks (if the peak centers are somewhat separated). Several papers have presented deconvolution methods for one-dimensional GC [2, 3, 4]. The two

columns of GCxGC are designed to be orthogonal, so one-dimensional deconvolution theoretically can be extended to GCxGC in the separable fashion.

One important issue of deconvolution is estimation of chromatographic peak shapes which are regarded as the point spread function (PSF) of a linear system. In one-dimension, peak shapes can be described effectively by mathematical functions such as empirically transformed Gaussian (ETG), polynomial modified Gaussian (PMG), generalized exponentially modified Gaussian (GEMG), and hybrid of Gaussian and truncated exponential (EGH) [5, 6, 7]. For GCxGC, however, there are at least two challenges. One challenge is that mathematical functions that can approximately model two-dimensional peak shapes have many parameters. The other challenge is that optimal parameter values for two-dimensional peak shapes are non-homogeneous, subject to different analytes (even from the same chemical family) and experimental conditions. Deconvolution methods that require homogeneous peak shapes cannot account for variation of peak shapes in GCxGC data. It is desirable to develop a more general linear model which can account for nonhomogeneous peak shapes.

Based on properties of chemical compounds and gas chromatography, GCxGC can be modeled as a shift-variant linear system:

$$\mathbf{D} = \mathbf{R}\mathbf{P}\mathbf{C}^T + \mathbf{E}, \quad (1)$$

where \mathbf{D} is an $I \times J$ matrix for observed GCxGC data, \mathbf{R} is an $I \times I$ matrix for the shift-variant point spread function (SVPSF) of the first column, \mathbf{C} is a $J \times J$ SVPSF matrix for the second column, \mathbf{P} is an $I \times J$ matrix of the desired impulse image (i.e., a single impulse for each chemical compound), \mathbf{E} is an $I \times J$ matrix accounting for additive noise and/or model error, and exponent ‘ T ’ denotes matrix transpose. The linear model exploits the orthogonality of GCxGC and can account for various chromatographic peak shapes.

Because the desired impulse image \mathbf{P} , the SVPSF \mathbf{R} , and the SVPSF \mathbf{C} are all unknown, the alternating least-squares (ALS) approach is employed to find their joint solutions. ALS iteratively estimates each unknown in the least-squares sense conditionally on the other unknowns [8]. The

*This work was supported by NSF Award IIS-0431119. Contact information: email: [jshi|reich]@cse.unl.edu, phone: 1.402.472.2401, fax: 1.402.472.7767.

estimates are refined iteratively until there are no significant changes in them or of the fit of the model to the observed data. The ALS algorithm may yield many solutions, so in order to find a desired solution, physically meaningful constraints are imposed on estimates during iteration.

Iterative methods have been studied extensively in digital image restoration [9, 10, 11]. Major advantages of the iterative methods are: i) closed-form solutions are not necessary; ii) *a priori* knowledge or constraints, such as nonnegativity and unimodality of peaks, can be incorporated conveniently; iii) iterative processes can be monitored for desired solutions; and iv) parameters which control solutions can be updated during iteration. With sufficient iterations, the CALS method can yield the desired image \mathbf{P} , together with SVPSFs \mathbf{R} and \mathbf{C} .

The rest of this paper is organized as follows. Section 2 gives the mathematical formulation of restoration of GCxGC data. Section 3 presents the algorithm of CALS and discusses some practical issues for its implementation. Section 4 demonstrates experimental results of both simulation and real GCxGC data. Finally, Section 5 gives conclusions and describes future work.

2 Formulation of the problem

In this paper, elements of vectors and matrices are denoted by lowercase italics, vectors are denoted by lowercase bold characters, and matrices are denoted by uppercase bold characters. The dimensions of matrices \mathbf{R} , \mathbf{C} , \mathbf{P} , and \mathbf{D} are respectively $I \times I$, $J \times J$, $I \times J$, and $I \times J$.

Solving Equation (1) for the desired impulse image \mathbf{P} is an inverse problem and requires a well-defined objective function. A typical objective function is the least-squares error defined as:

$$\mathcal{F}(\mathbf{R}, \mathbf{C}, \mathbf{P}) = \|\mathbf{D} - \mathbf{R}\mathbf{P}\mathbf{C}^T\|_F^2, \quad (2)$$

which is a functional of \mathbf{R} , \mathbf{C} , and \mathbf{P} and where $\|\cdot\|_F$ is the Frobenius norm.

In most GCxGC applications, only \mathbf{D} is known, so an infinite number of solutions of \mathbf{R} , \mathbf{C} , and \mathbf{P} can account for \mathbf{D} according to theories of matrix factorization. In order to guarantee the uniqueness of the solution, the orthogonormal constraint can be imposed on columns of \mathbf{R} and \mathbf{C} like singular value decomposition (SVD). But the orthogonormal constraint does not have chemical or physical meaning in factor analysis. Fortunately, there are several physically meaningful constraints such as nonnegativity and unimodality [8] which can reduce the number of solutions significantly. Subject to physically meaningful constraints and efficient mathematical models for peaks, minimizing the objective function $\mathcal{F}(\mathbf{R}, \mathbf{C}, \mathbf{P})$ can give a reasonable solution of the desired impulse image \mathbf{P} . If sufficient constraints are

Mathematical formulation of the problem

Given observed $I \times J$ GCxGC data \mathbf{D} , minimize $\mathcal{F}(\mathbf{R}, \mathbf{C}, \mathbf{P}) = \|\mathbf{D} - \mathbf{R}\mathbf{P}\mathbf{C}^T\|_F^2$, subject to

1. Nonnegativity:
 - (a) $r_{ij} \geq 0, \forall 1 \leq i, j \leq I$,
 - (b) $c_{ij} \geq 0, \forall 1 \leq i, j \leq J$,
 - (c) $p_{ij} \geq 0, \forall 1 \leq i \leq I, 1 \leq j \leq J$.
2. Constant volume:

$$\forall 1 \leq j \leq I, \sum_{i=1}^I r_{ij} = 1; \forall 1 \leq j \leq J, \sum_{i=1}^I c_{ij} = 1$$

$$\implies \sum_{i=1}^I \sum_{j=1}^J d_{ij} = \sum_{i=1}^I \sum_{j=1}^J p_{ij}$$
3. Unimodality:

If $\exists i$ and j , such that $p_{ij} \neq 0$, then $r_{1i} \leq \dots \leq r_{ii} \geq \dots \geq r_{iI}$ and $c_{1j} \leq \dots \leq c_{jj} \geq \dots \geq c_{jJ}$
 else $r_{1i} = \dots = r_{iI} = 0$ and $c_{1j} = \dots = c_{jJ} = 0$.

Figure 1: Mathematical formulation of the problem of finding the desired impulse image \mathbf{P} from observed GCxGC data \mathbf{D} .

available, the solution can be unique. Figure 1 mathematically formulates the problem of finding the desired impulse image \mathbf{P} from observed GCxGC data \mathbf{D} .

Let $\mathbf{R} = [\mathbf{r}_1 \dots \mathbf{r}_I]$ and $\mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_J]$. From the aspect of factor analysis of matrices, the observed GCxGC data \mathbf{D} can be factorized as the summation of $I * J$ rank-one matrices weighted by p_{ij} as follows:

$$\begin{aligned} \mathbf{D} &= \mathbf{R}\mathbf{P}\mathbf{C}^T + \mathbf{E} \\ &= \sum_{1 \leq i \leq I} \sum_{1 \leq j \leq J} p_{ij} \mathbf{r}_i \mathbf{c}_j^T + \mathbf{E} \\ &\approx \sum_{1 \leq i \leq I} \sum_{1 \leq j \leq J} p_{ij} \mathbf{r}_i \mathbf{c}_j^T \\ &= \sum_{1 \leq i \leq I} \sum_{1 \leq j \leq J} p_{ij} \begin{bmatrix} r_{i1} \\ \vdots \\ r_{iI} \end{bmatrix} [c_{j1} \dots c_{jJ}]. \quad (3) \end{aligned}$$

Because the number of chromatographic peaks in GCxGC data \mathbf{D} is typically much less than the number $I * J$ of elements of \mathbf{D} , the desired impulse image \mathbf{P} is a sparse matrix and most columns of \mathbf{R} and \mathbf{C} are all zeros and so don't contribute to \mathbf{D} .

3 Constrained alternating least-squares

ALS is a basic technique for Tuck3 model and PARAFAC (parallel factor analysis) model [12, 13]. The essential idea of ALS is to estimate each unknown in the least-squares

1. Initialize \mathbf{X} ,
2. $\mathbf{Y} = \mathbf{D}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$,
3. $\mathbf{X} = \mathbf{D} \mathbf{Y} (\mathbf{Y}^T \mathbf{Y})^{-1}$,
4. Go to step 2 until stop conditions are satisfied.

Figure 2: The classic ALS algorithm for a two-way problem.

sense conditionally on the other unknowns. For example, let a two-way problem $\mathbf{D} = \mathbf{X}\mathbf{Y}^T$, where \mathbf{D} is a matrix of observed data and \mathbf{X} and \mathbf{Y} are two unknown matrices. Figure 2 presents the classic ALS algorithm for solving

$$\operatorname{argmin}_{\mathbf{X}, \mathbf{Y}} \{ \|\mathbf{D} - \mathbf{X}\mathbf{Y}\|_F^2 \}. \quad (4)$$

This classic ALS implementation requires the pseudoinverse (*e.g.*, Moore-Penrose) of matrices, which may cause frequent violations of the unimodality constraint.

The ALS restoration of minimizing $\mathcal{F}(\mathbf{R}, \mathbf{C}, \mathbf{P})$ with respect to \mathbf{R} , \mathbf{C} , and \mathbf{P} is equivalent to two successive two-way problems. Let $\mathbf{W} = \mathbf{C}\mathbf{P}^T$ be an intermediate matrix. The two two-way problems can be formulated as:

$$\operatorname{argmin}_{\mathbf{R}, \mathbf{W}} \{ \mathcal{F}_1(\mathbf{R}, \mathbf{W}) = \|\mathbf{D} - \mathbf{R}\mathbf{W}^T\|_F^2 \} \quad (5)$$

and

$$\operatorname{argmin}_{\mathbf{C}, \mathbf{P}} \{ \mathcal{F}_2(\mathbf{C}, \mathbf{P}) = \|\mathbf{W} - \mathbf{C}\mathbf{P}^T\|_F^2 \}. \quad (6)$$

The optimal solution of \mathbf{W} in Equation (5) is used successively for optimizing \mathbf{C} and \mathbf{P} in Equation (6).

Iterative methods without matrix pseudoinverse have lower possibility of violating the constraints. In particular, when an iterative estimate is close to the optimal solution, the update is very small and the unimodal shape can be maintained. One popular iterative method is the steepest descent which updates unknowns along the opposite direction of the gradient of an objective function at the latest estimate of the unknowns. Mathematically,

$$\begin{aligned} \mathcal{F}_1(\mathbf{R}, \mathbf{W}) &= \|\mathbf{D} - \mathbf{R}\mathbf{W}^T\|_F^2 \\ &= \sum_{i=1}^I \|\mathbf{d}_i - \mathbf{R}\mathbf{w}_i\|_F^2 \\ &= \sum_{i=1}^I \|\mathbf{d}_i - \mathbf{r}_i \mathbf{W}^T\|_F^2. \end{aligned} \quad (7)$$

The gradients of the objective function $\mathcal{F}_1(\mathbf{R}, \mathbf{W})$ with respect to \mathbf{R} and \mathbf{W} are:

$$\frac{\partial \mathcal{F}_1}{\partial \mathbf{R}} = \left\{ \frac{\partial \mathcal{F}_1}{\partial \mathbf{r}_1}, \dots, \frac{\partial \mathcal{F}_1}{\partial \mathbf{r}_I} \right\} = (\mathbf{R}\mathbf{W}^T - \mathbf{D})\mathbf{W} \quad (8)$$

$$\frac{\partial \mathcal{F}_1}{\partial \mathbf{W}} = \left\{ \frac{\partial \mathcal{F}_1}{\partial \mathbf{w}_1}, \dots, \frac{\partial \mathcal{F}_1}{\partial \mathbf{w}_I} \right\} = (\mathbf{W}\mathbf{R}^T - \mathbf{D}^T)\mathbf{R}. \quad (9)$$

Similarly, the gradients of the objective function $\mathcal{F}_2(\mathbf{C}, \mathbf{P})$ with respect to \mathbf{C} and \mathbf{P} are:

$$\frac{\partial \mathcal{F}_2}{\partial \mathbf{C}} = \left\{ \frac{\partial \mathcal{F}_2}{\partial \mathbf{c}_1}, \dots, \frac{\partial \mathcal{F}_2}{\partial \mathbf{c}_J} \right\} = (\mathbf{C}\mathbf{P}^T - \mathbf{W})\mathbf{P} \quad (10)$$

$$\frac{\partial \mathcal{F}_2}{\partial \mathbf{P}} = \left\{ \frac{\partial \mathcal{F}_2}{\partial \mathbf{p}_1}, \dots, \frac{\partial \mathcal{F}_2}{\partial \mathbf{p}_J} \right\} = (\mathbf{P}\mathbf{C}^T - \mathbf{W}^T)\mathbf{C}. \quad (11)$$

Figure 3 proposes the CALS algorithm for solving the problem as defined in Figure 1. \otimes stands for Khatri Rao columnwise product [13]. The algorithm has two inner iterations and one outer iteration. Each inner iteration minimizes one two-way problem. The optimal \mathbf{W} of $\mathcal{F}_1(\mathbf{R}, \mathbf{W})$ is used to optimize \mathbf{C} and \mathbf{P} in $\mathcal{F}_2(\mathbf{C}, \mathbf{P})$. The outer iteration re-initializes \mathbf{R}_0 , \mathbf{C}_0 , and \mathbf{P}_0 in step 4 so that optimal estimates for $\mathcal{F}_2(\mathbf{C}, \mathbf{P})$ can be used to update \mathbf{R} and \mathbf{W} in $\mathcal{F}_1(\mathbf{R}, \mathbf{W})$ as further optimization. The algorithm converges eventually because each iteration never degrades its previous estimates in the least-squares sense.

An important concern for iterative algorithms is the risk that minimization of the objective function \mathcal{F} with respect to \mathbf{R} , \mathbf{C} , and \mathbf{P} may fall into local minima. Complicated algorithms such as simulated annealing [14, 15] can help find the global minima, but they are commonly time-consuming. One effective method which can avoid many local minima is to initialize \mathbf{R} , \mathbf{C} , and \mathbf{P} with appropriate values rather than random values or zeros. An initialization close to the global minima can reduce the risk of stopping at local minima and the convergence time. For the initial \mathbf{R}_0 and \mathbf{C}_0 of the CALS algorithm, chromatographic peak shapes can be assumed to be homogeneous. Either peaks that can be identified clearly from GCxGC data or the Gaussian function (if clear peaks are unavailable) can be good candidate PSFs for deconvolution. Digital image restoration methods for linear, shift-invariant systems can be used to initialize the desired image \mathbf{P}_0 . Blind image deconvolution also can be exploited for more accurate initialization [16, 17].

At each iteration, the CALS algorithm searches estimates along the steepest descent. Setting the values of α_i ($i = 1, 2, 3, 4$), called step size in the numerical analysis literature, is important. In simple iterative methods, step sizes are fixed within a safe convergence interval [10, 11]. Updating step sizes α_i ($i = 1, 2, 3, 4$) at each iteration so that they minimize the objective functions \mathcal{F}_1 and \mathcal{F}_2 can speed convergence. Because constraints of unimodality and nonnegativity are nonlinear projection, line search methods can work well [15, 18]. The conjugate gradient is inapplicable here because conjugation of two successive searching directions may be destroyed by nonlinear projections.

The stop condition is determined by iteration times and the expected quality of an estimate. The quality measures how close an estimate is to its expected value or global minimum. There are several stop conditions in the numerical analysis literature. For example, i) two successive iterations

CAL S Algorithm for GCxGC Restoration

1. Initialize $\mathbf{R}_0, \mathbf{C}_0, \mathbf{P}_0$ and let $\mathbf{W}_0 = \mathbf{C}_0 \mathbf{P}_0^T$ be an intermediate estimate.
2. Optimize $\mathcal{F}_1(\mathbf{R}, \mathbf{W}) = \|\mathbf{D} - \mathbf{R}\mathbf{W}^T\|_F^2$ with respect to \mathbf{R} and \mathbf{W} . At the k^{th} inner iteration:
 - (a) Use steepest descent (or line search) method to estimate optimal α_1 ,
 - (b) $\mathbf{R}^k = \mathbf{R}^{k-1} + \alpha_1 \otimes (\mathbf{D} - \mathbf{R}^{k-1} \mathbf{W}^{k-1 T}) \mathbf{W}^{k-1}$,
 - (c) Impose constraints on \mathbf{R}^k ,
 - (d) Use line search method to estimate optimal α_2 ,
 - (e) $\mathbf{W}^k = \mathbf{W}^{k-1} + \alpha_2 \otimes (\mathbf{D}^T - \mathbf{W}^{k-1} \mathbf{R}^{k T}) \mathbf{R}^k$,
 - (f) Impose constraints on \mathbf{W}^k ,
 - (g) $\mathbf{R}^{k-1} = \mathbf{R}^k$ and $\mathbf{W}^{k-1} = \mathbf{W}^k$ if $\mathcal{F}_1^{k-1} > \mathcal{F}_1^k$,
 - (h) Go to step 2.a until inner stop conditions are satisfied.
3. Optimize $\mathcal{F}_2(\mathbf{C}, \mathbf{P}) = \|\mathbf{W} - \mathbf{C}\mathbf{P}^T\|_F^2$ with respect to \mathbf{C} and \mathbf{P} . At the k^{th} inner iteration:
 - (a) Use steepest descent (or line search) method to estimate optimal α_3 ,
 - (b) $\mathbf{C}^k = \mathbf{C}^{k-1} + \alpha_3 \otimes (\mathbf{W} - \mathbf{C}^{k-1} \mathbf{P}^{k-1 T}) \mathbf{P}^{k-1}$,
 - (c) Impose constraints on \mathbf{C}^k ,
 - (d) Use line search method to estimate optimal α_4 ,
 - (e) $\mathbf{P}^k = \mathbf{P}^{k-1} + \alpha_4 \otimes (\mathbf{W}^T - \mathbf{P}^{k-1} \mathbf{C}^{k T}) \mathbf{C}^k$,
 - (f) Impose constraints on \mathbf{P}^k ,
 - (g) Update $\mathbf{C}^{k-1} = \mathbf{C}^k$ and $\mathbf{P}^{k-1} = \mathbf{P}^k$ if $\mathcal{F}_2^{k-1} > \mathcal{F}_2^k$,
 - (h) Go to step 3.a until inner stop conditions are satisfied.
4. Reinitialize $\mathbf{R}_0, \mathbf{C}_0$, and \mathbf{P}_0 with solutions from steps 2 and 3 if $\mathcal{F}^{l-1} > \mathcal{F}^l$ at the l^{th} outer iteration.
5. Go to step 2 until the outer stop conditions are satisfied.

Figure 3: The CAL S algorithm for GCxGC restoration.

have little difference about the estimates or the objective function; or ii) the gradient is sufficiently close to 0.

Physically meaningful constraints are crucial to guarantee that the algorithm can converge to an expected solution. The nonnegativity and volume preservation constraints can be implemented with low computational cost. Regarding the unimodality of columns of \mathbf{R} and \mathbf{C} , Bro and Sidirpoulos have developed an efficient algorithm in the least-squares sense for an N-way problem [8].

4 Experimental results

4.1 Simulation

The simulation is implemented in Matlab 6.5. Figure 4(A) illustrates a 31×21 simulated-GCxGC image \mathbf{D} with three overlapping chromatographic peaks, synthesized as in Equation 1. Columns of the SVPSF matrices \mathbf{R} and \mathbf{C} are simulated by Gaussian functions:

$$\begin{aligned} r_{ij} &= A_{rj} e^{-i^2/(2\sigma_{rj}^2)}, \quad 1 \leq i, j \leq 31 \\ c_{ij} &= A_{cj} e^{-i^2/(2\sigma_{cj}^2)}, \quad 1 \leq i, j \leq 21, \end{aligned} \quad (12)$$

where σ_{rj} and σ_{cj} are set randomly in the intervals (2, 3) and (0.7, 1) respectively, with uniform distribution, in order

to simulate the shift-variance in \mathbf{R} and \mathbf{C} . The constraint of volume preservation (from Figure 1) implies that for this example:

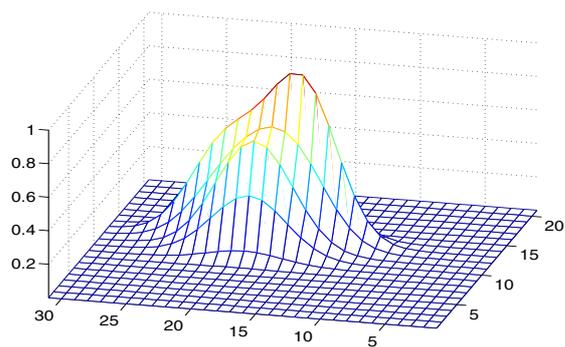
$$\begin{aligned} A_{rj} &= 1 / \sum_{i=1}^{31} e^{-i^2/(2\sigma_{rj}^2)}, \quad 1 \leq j \leq 31 \\ A_{cj} &= 1 / \sum_{i=1}^{21} e^{-i^2/(2\sigma_{cj}^2)}, \quad 1 \leq j \leq 21. \end{aligned} \quad (13)$$

The desired image \mathbf{P} has three impulses, corresponding to the three chromatographic peaks in \mathbf{D} . Tables 4.1 list locations and heights of the impulses. The noise matrix \mathbf{E} is simulated by the normal distribution with 0 mean and 0.001 standard deviation.

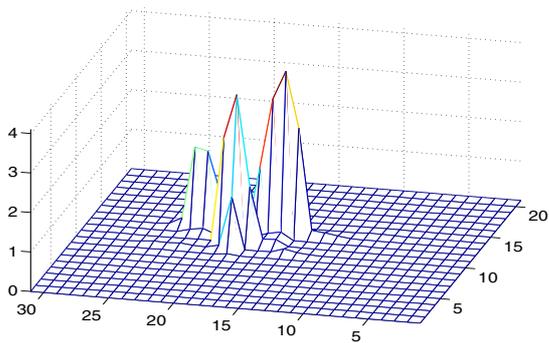
ID	Impulses for synthesis			Peaks from restoration		
	row	col	height	row	col	volume
1	16	13	11.681	16	13	12.928
2	19	11	10.841	19	11	10.596
3	22	13	6.855	23	13	5.555
4	-	-	-	23	10	0.160

Table 1: Parameters values for the synthesis (left part). Locations and volumes of the restored peaks (right part).

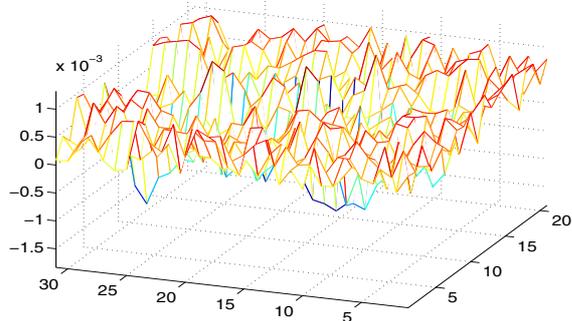
\mathbf{R}_0 and \mathbf{C}_0 are initialized respectively by the Gaussian function with $\sigma_{rj} = 3$ ($1 \leq j \leq 31$) and $\sigma_{cj} = 1$ ($1 \leq j \leq 21$) in the shift-invariant fashion. The initial impulse image P_0 is estimated by the traditional iterative image restoration [10]. Figure 4(B) illustrates the resulting image. The three peaks are separated completely in this image. Table 4.1 lists the locations and volumes of the restored peaks. The major peaks (ID=1, 2, and 3) have good estimates of location and shapes. Peak 4 is a pseudo-peak which has very small peak volume, so it can be removed by appropriate thresholding. Figure 4 illustrates the residual, which is close to the noise matrix \mathbf{E} .



A. Synthesized GCxGC image



B. Restored GCxGC image



C. Residual

Figure 4: Experimental results for simulation.

4.2 Real GCxGC data

Figure 5(A) illustrates a 34×21 subimage cropped from GCxGC data “oven040209run5” (Zoex Corporation, Lincoln NE) with an oven-temperature ramp-rate of $5^\circ\text{C}/\text{min}$. This subimage has two chemical constituents: hexamethylbenzene and 1-dodecanol. Again, the Gaussian function is used to initialize \mathbf{R}_0 and \mathbf{C}_0 homogeneously with $\sigma_{rj} = 2$ ($1 \leq j \leq 34$) and $\sigma_{cj} = 0.5$ ($1 \leq j \leq 21$) respectively and traditional iterative image restoration [10] is used to initialize \mathbf{P}_0 . Figure 5(B) illustrates the resulting image. After restoration the two chemical compounds are separated, with the hexamethylbenzene peak located at pixel (18, 9) and the 1-dodecanol peak at (16, 11). After restoration, the volume ratio of the two peaks is 8.96. By comparison, in GCxGC runs with different temperature ramp rates which fully separated the peaks, the ratios were 9.07 at $2^\circ\text{C}/\text{min}$, 9.12 at $3^\circ\text{C}/\text{min}$, and 9.04 at $4^\circ\text{C}/\text{min}$. The difference of the restored ratio with the mean ratio for separated data is about 1.3%, which is about 2.7 times the inferred standard deviation from the separated data. The residual data \mathbf{E} (illustrated in Figure 5(C)) is close to a noise distribution, but in the real GCxGC data, compounds produce peaks that are not completely separable, so the residual also contains slight model error.

5 Conclusions

This paper develops an approach for restoration of data from comprehensive two-dimensional gas chromatography (GCxGC). The GCxGC process is modeled as a shift-variant linear system, based on the properties of chemical compounds and gas chromatographic instruments. The model can account for nonhomogeneous peak shapes and separability of two columns of instruments. The restoration problem is formulated to minimize the difference between observed GCxGC data and an ideal separation, subject to the physically meaningful constraints: nonnegativity, volume preservation, and unimodality.

The paper also proposes a constrained alternating least-squares (CALs) method for GCxGC restoration. The essential idea is to estimate each unknown in the least-squares sense conditionally on the other unknowns and the constraints. The approach is implemented by iteration with low computational cost because the linear model is separable.

Experimental results based on the simulation and the real GCxGC data indicate that the proposed linear model and CALs perform well for GCxGC restoration. It is an effective approach to improving separation, identification, and quantification of analytes. Ongoing work is evaluating the quantitative performance for more chemical analytes.

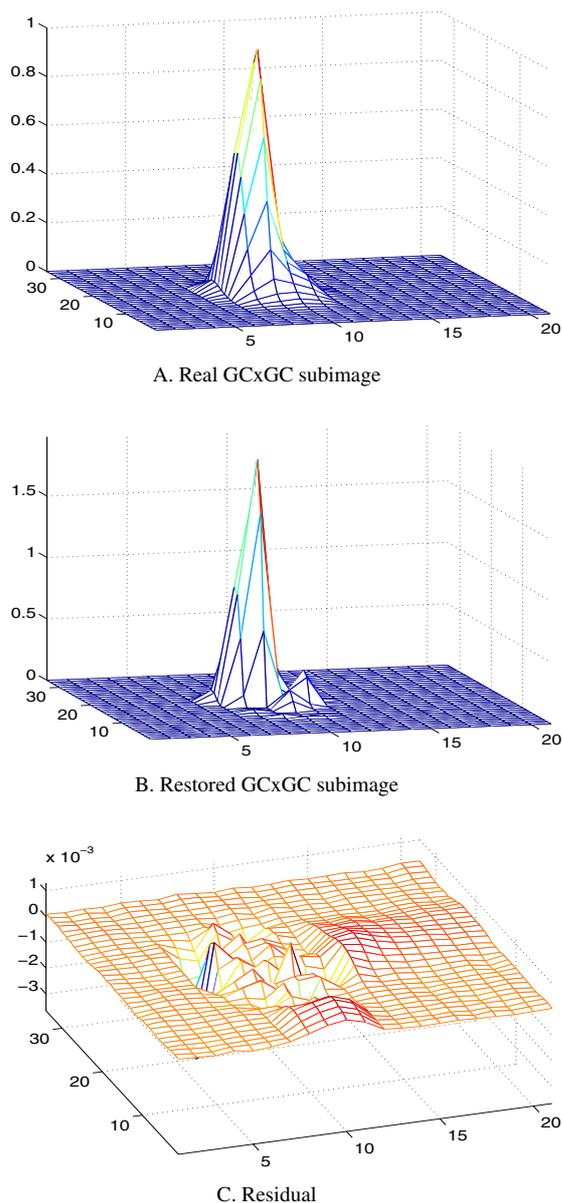


Figure 5: Experimental results for real GCxGC data.

References

- [1] J. Dallüge, J. Beens, and U. A. Brinkman, "Comprehensive two-dimensional gas chromatography: a powerful and versatile analytical tool," *Journal of Chromatography A*, vol. 1000, pp. 69–108, 2003.
- [2] R. D. Caballero, M. C. García-Alvarez Coque, and J. J. Baeza-Baeza, "Parabolic-Lorentzian modified gaussian model for describing and deconvolving chromatographic peaks," *Journal of Chromatography A*, vol. 954, pp. 59–76, 2002.
- [3] G. Vivó-Truyols, J. Torres-Lapasió, R. D. Caballero, and M. C. García-Alvarez Coque, "Peak deconvolution in one-dimensional chromatography using a two-way data approach," *Journal of Chromatography A*, vol. 958, pp. 35–49, 2002.
- [4] P. Nikitas, A. Pappa-Louisi, and A. Papageorgiou, "On the equations describing chromatographic peaks and the problem of the deconvolution of overlapped peaks," *Journal of Chromatography A*, vol. 912, pp. 13–29, 2001.
- [5] J. Li, "Comparison of the capability of peak functions in describing real chromatographic peaks," *Journal of Chromatography A*, vol. 952, pp. 63–70, 2002.
- [6] Z. Pápai and T. L. Pap, "Analysis of peak asymmetry in chromatography," *Journal of Chromatography A*, vol. 953, pp. 31–38, 2002.
- [7] T. L. Pap and Z. Pápai, "Application of a new mathematical function for describing chromatographic peaks," *Journal of Chromatography A*, vol. 930, pp. 53–60, 2001.
- [8] R. Bro and N. D. Sidiropoulos, "Least squares algorithms under unimodality and non-negativity constraints," *Journal of Chemometrics*, vol. 12, pp. 223–247, 1998.
- [9] A. K. Katsaggelos, "Iterative image restoration algorithms," *Optical Engineering*, vol. 28, no. 7, pp. 735–748, 1989.
- [10] J. Biemond, R. L. Lagendijk, and R. M. Mersereau, "Iterative methods for image deblurring," *Proceedings of the IEEE*, vol. 78, no. 5, pp. 856–883, 1990.
- [11] A. K. Katsaggelos and C.-J. Tsai, "Iterative image restoration," in *Handbook of Image and Video Processing*. A. Bovik, Ed. Academic Press, San Diego CA, 2000.
- [12] E. R. Malinowski, *Factor Analysis in Chemistry*. New York, NY: John Wiley and Sons, 1991.
- [13] R. Bro, "Multi-way analysis in the food industry: Models, algorithms, and applications," Ph.D. dissertation, Department of Dairy and Food Science, Royal Veterinary and Agricultural University, Denmark, 1998.
- [14] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.
- [15] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*. New York, NY: Cambridge University Press, 1997.
- [16] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, 1996.
- [17] —, "Blind image deconvolution revisited," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 61–63, 1996.
- [18] J. J. Moñe and D. J. Thuente, "Line search algorithms with guaranteed sufficient decrease," *ACM Transactions on Mathematical Software*, vol. 20, no. 3, pp. 286–307, 1994.