Multicomponent analysis of stripping voltammograms by target transformation factor analysis

Ling Gao and Shouxin Ren
Department of Chemistry, Inner Mongolian University, Huhhot 010021, Inner Mongolia, China

Target transformation factor analysis was studied for simultaneous determination of four simulated stripping voltammograms. A Gaussian function was used to generate synthetic stripping voltammograms. Two programs, SPGRELEP and ELECTTFA, were designed to perform voltammogram simulation and target transformation factor analysis. The method made use of full information of voltammogram and matrix data processing. The ELECTTFA program was used to determine the number of components, to identify the components and to yield a quantitative analysis of unknowns. Experimental results showed the method to be successful.

Introduction

Chemometric methods have been applied to improve results in the multicomponent analytical techniques in spectroscopy [1-3], but they have rarely been used in electrochemistry. Target transformation factor analysis (TTFA) is a technique that is especially valuable for achieving meaningful transformations of the abstract factors obtained by abstract factor analysis (AFA) [4]. TTFA can be used to yield both qualitative and quantitative information. It also can be used to decide whether or not a suspected substance is present in the mixture and to deduce the composition of each mixture. This method has been applied to spectroscopic, nuclear magnetic resonance, chromatography and kinetic analysis [5, 6]; attention has been paid, however, to applying TTFA to voltammetry. Many voltammetric techniques are of a local character, using only the peak currents or a limited number of points around the peak, thus losing much of the information contained in the remaining parts of the voltammogram. This paper describes the improvement of multicomponent determinations using the full information of voltammogram and matrix data processing.

Theoretical background and program algorithms

Building the data matrix \( D(P, S) \)

The original data matrix, \( D \), contains the anodic current response at \( P \) voltage values of \( S \) mixtures; each voltammogram being a column and the data taken at the same voltage being a row.

Abstract factor analysis of the experimental data matrix \( D \)

The covariance matrix \( Z(D^TD) \) can be subjected to singular value decomposition for calculation of the eigenvalues and eigenvectors, \( Z = USV^T \), where \( S \) is a diagonal matrix of the eigenvalues and \( U \) is the matrix of eigenvectors. The \( D \) matrix can be represented as the product of two matrices: \( D = RC \), where \( R(P, F) \) are the current profiles of the present species and \( C(F, S) \) their concentration matrix (where \( F \) is the number of components). It is easy to compute: \( C = U^T \) and \( R = DU \) due to the orthonormality of \( U^{-1} \) and \( U^T \). The number of principal components is estimated by several criteria, which are based on the theory of error in factor analysis [7]. Not all of the eigenvectors produced convey useful information; some are due to noise. After deleting the insignificant eigenvectors \( R \) and \( C \) become \( R(P, a) \) and \( C(a, S) \), where \( a \) is the number of principal components.

Target transformation

The abstract matrices must be rotated into significant real matrices. This oblique rotation is accomplished by transforming matrix \( T \). The matrix \( T \) can be determined by using the target vector \( R_2 \) from the equations \( T = (R_2^TR_2)^{-1}R_2^TR_1 \) or \( T = S^{-1}R_1^TR_2 \), and then the concentrations and the anodic current profiles of sample component are given by \( C_0 = T^{-1}C_1 \) and \( R_0 = R_1T \), respectively. In traditional factor analysis, the current profiles of pure components are always used as the \( R_2 \) matrix to calculate \( T \). This may cause errors due to interaction between the components. To overcome these, an \( R_2 \) matrix, calculated from the standard mixtures by classical least squares analysis \( R_2 = D_5C_3^T(C_3C_3^T)^{-1} \), and a non-zero intercept added to each voltage value, are used as the target vector instead of the pure component standards. \( C_3 \) is the concentration matrix of the standard mixtures, \( D_5 \) is the response matrix of the standard mixtures.

Target testing

The aim of target testing is to decide whether a proposed target can be accepted as a real factor. The criteria used for accepting/rejecting target vectors is based on the similarity between the target vector and the predicted target vector. Malinowski and Howery suggested using the \( \text{SPOIL} \) function to decide whether a proposed target is acceptable or not [4]. The \( \text{SPOIL} \) function is defined as the ratio of the real error in the target vector \( \text{RET} \) to the real error in the predicted vector \( \text{REP} \):

\[
0142-0453/96 $12.00 © 1996 Taylor & Francis Ltd.
175
SPOIL = RET/REP. The apparent error in the test vector \( \langle AET \rangle \) is calculated using:

\[
AET = \left[ \frac{\sum_{i=1}^{a} (R_0 - R_i)^2}{a} \right]^{1/2}.
\]

\( REP \) is defined as \( REP_i = RE_i \| T_i \| \) with \( i = 1, 2, \ldots, a \). Where \( \| T_i \| \) is the Euclidean norm of the vector; \( RET \) is then evaluated by \( RET = [(AET)^2 - (REP)^2]^{1/2} \).

Malinowski and Howery divided the value of the SPOIL function into three regions: (1) an acceptable region (0–3); (2) a fair region (3–6); and (3) an unacceptable region (>6).

Voltammogram simulation

A Gaussian peak shape is assumed for the voltammograms of the pure component with peak position \( M \), peak height \( C \), and widths \( W \):

\[
cur(i) = C(i) \exp\left\{-\frac{(vol - M)}{W} \right\}^2
\]

The \( \text{vol} \) denotes the voltage values of the full voltammogram and the symbol \( i \) denotes the component number. The second part of the simulation consists of generating the measurement noise \( \langle N_{\text{os}} \rangle \). The measurement noise is assumed to be normally distributed with zero mean:

\[
\mathbb{E}[N_{\text{os}}(\text{vol})] = 0 \quad \text{for all Vol}_{k}
\]

and white with variance \( R(\text{vol}) \):

\[
\mathbb{E}[N_{\text{os}}(\text{vol}) \cdot N_{\text{os}}(\text{vol})] = R(\text{vol}) \delta_{k,j}
\]

\[
\delta_{k,j} = 1 \quad \text{if } k = j
\]

\[
\delta_{k,j} = 0 \quad \text{if } k \neq j
\]

\( \mathbb{E}[\cdot] \) denotes the expectation of the quantity within the brackets, \( \delta \) is Kronecker’s symbol.

Two programs, SPGRELEP and ELECTTFA, which are based on the algorithms were designed to perform voltammogram simulation and TFA.

Experimental

**Instrument**

A GW 286 EX/16 microcomputer with a maths coprocessor was used for the calculations.

Simulated voltammograms

Simulated data were used to evaluate the potential of TFA for stripping voltammograms. The peak potentials of copper, lead, cadmium and zinc had the values of -0.35, -0.53, -0.70 and -1.10 V (VS.SCE). The peak widths of the four components were 0.05 V. Gaussian distributed noise was added to the voltammograms to simulate experimental noise as 1% of the average signal value. The background was assumed to be normally distributed with zero mean; the peak height of the background was 0.05 ppm. Taking background as baselines, the simulated voltammograms of nine unknown samples and baselines were plotted as shown in figure 1. The voltammograms of the standard samples were simulated in the same way as the unknown samples. Figure 2 is a three-dimensional plot of the simulated standard data matrix \( D_3 \).

*Figure 1. Simulated voltammograms of the unknown samples.*

*Figure 2. Three-dimensional plot of the standard data matrix \( D_3 \).*

Three-level orthogonal array design of standard samples

For any orthogonal array design, a matrix, which consists of columns and rows, with various numbers at the intersections of each column and row, must be constructed. Table 1 displays an \( L_9(3^4) \) matrix. This is a three-level orthogonal array matrix which is made up of four columns and nine rows. Each column represents a factor, which is an independent variable, and each row represents an experimental trial. The numbers at the intersections indicate the level settings that apply to

| No. | 1  | 2  | 3  | 4  | Cu | Pb | Cd | Zn |
|-----|----|----|----|----|----|----|----|----|
| 1   | 1  | 1  | 1  | 4  | 2  | 3  | 3  |
| 2   | 1  | 2  | 2  | 4  | 5  | 6  | 6  |
| 3   | 1  | 3  | 3  | 4  | 8  | 9  | 9  |
| 4   | 2  | 1  | 2  | 7  | 2  | 6  | 3  |
| 5   | 2  | 2  | 3  | 7  | 5  | 9  | 6  |
| 6   | 2  | 3  | 1  | 7  | 8  | 3  | 9  |
| 7   | 3  | 1  | 3  | 10 | 2  | 9  | 9  |
| 8   | 3  | 2  | 1  | 10 | 5  | 3  | 3  |
| 9   | 3  | 3  | 2  | 10 | 8  | 6  | 1  |
Table 2. Results of the factor analysis on the unknown samples.

| N  | EV       | RE       | IND | XE       | IE       | ER       | REV       | Frac     |
|----|----------|----------|-----|----------|----------|----------|-----------|----------|
| 1  | 3.1797E+3 | 2.9521E-1 | 4.6127E-3 | 2.7833E-1 | 9.8404E-2 | 1.1519E+2 | 5.3531E+0 | 9.8573E-1 |
| 2  | 2.7604E+1 | 1.9963E-1 | 4.0741E-3 | 1.7606E-1 | 9.4106E-2 | 2.3928E+0 | 5.3084E-2 | 8.5573E-3 |
| 3  | 1.9000E+1 | 1.4003E-1 | 3.8898E-3 | 1.1434E-1 | 8.0848E-2 | 1.3771E+0 | 2.3764E-2 | 3.3004E-3 |
| 4  | 7.7310E+0 | 1.0195E-2 | 4.0781E-4 | 7.5991E-3 | 6.7969E-3 | 7.0926E+2 | 2.0452E-2 | 2.3966E-3 |
| 5  | 1.9000E-2 | 9.4150E-3 | 5.8844E-4 | 6.2767E-3 | 7.0175E-3 | 1.2814E+0 | 3.5161E-5 | 3.3791E-6 |
| 6  | 8.5067E-3 | 8.6734E-3 | 9.6371E-3 | 5.0076E-3 | 7.0818E-3 | 1.4724E+0 | 3.4863E-5 | 2.6371E-6 |
| 7  | 5.7776E-3 | 8.3109E-3 | 2.0777E-3 | 3.9187E-3 | 7.3296E-3 | 1.0768E+0 | 3.2098E-5 | 1.7911E-6 |
| 8  | 5.5506E-3 | 7.5407E-3 | 7.5407E-3 | 7.5407E-3 | 7.1700E-3 | 1.4234E+0 | 4.3935E-5 | 1.6660E-6 |
| 9  | 7.6090E-3 | 7.8043E-3 | 6.0433E-3 | 1.6599E-6 | 1.0784E+0 | 3.9584E-5 | 1.6599E-6 |

The factors for the experimental trials. From this matrix it can be noted that each of four columns is varied over three level setting, each level setting repeats three times, and thus a total of $3 \times 3 = 9$ experimental trials are necessary for each column. Furthermore, in any two columns, the horizontal combinations of any two level numbers appear at the same number of times. That is, each combination of the nine ordered pairs $(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)$ appear exactly once. The above features of the $L_9(3^4)$ matrix provide orthogonality among all the four columns. Three-level orthogonal array design with the $L_9(3^4)$ matrix provides a systematic procedure for selecting the concentration of the training set samples (see table 1).

Results and discussion

Determination of the number of factors

Eight criteria were used to calculate the number of factors and the results are shown in table 2. The IND function reached a minimum at $N = 4$; the magnitude of the first four eigenvalues were larger than those of $5-9$; and a maximum of the eigenvalue ratio (ER) function appeared at 4. The magnitude of the reduced eigenvalues (REV) decreased rapidly until $N = 4$, then it stabilized. The minimum of the Frac function appeared at 5; after 5 the change in Frac was quite small (the appropriate number of eigenvectors is one less than that giving the minimum Frac value). Both the IE and XE functions show a sharp drop between the first and fourth eigenvectors and then they level off. A plot of logarithm of REV versus factor number is shown in figure 3. The REVs of factor 1–4 are clearly much larger than the rest. Some idea of the information contained in the various factors can be obtained by inspecting the corresponding eigenvector. The number of significant eigenvectors is determined by plotting them as a function of the voltage. From these figures, it is fairly obvious that the eigenvectors of the first four factors represent what appears to be linear combinations of the original pure component voltammograms. In contrast, the last five factors look like noise. Figures 4 and 5 show the eigenvector corresponding to factor 4 and 5, respectively. The fourth eigenvectors provide significant information, and the fifth look like noise. From these criteria, it was concluded that four significant components were present.

![Figure 3](image1.png)  
Figure 3. Log REV versus factor number for matrix D.

![Figure 4](image2.png)  
Figure 4. Plot of the fourth eigenvector of matrix D.

![Figure 5](image3.png)  
Figure 5. Plot of the fifth eigenvector of matrix D.
Determination and validation of target vector

According to the algorithms, the target matrix $R_2$ was calculated from the standard samples by the least squares method. The SPOIL function was developed, based on the similarity between the target vector and the predicted target vector, to validate the validity and usefulness of a predicted target vector. In order to judge whether the predicted target vectors were a successful target, the SPOIL values were calculated and the results are given in Table 3. The SPOIL values of Cu, Pb and Cd were below 3.0 and the value of Zn was a little over 3.0, indicating that all the predicted target vectors can be considered to be true targets.

Simultaneous determination of unknown samples

The abstract factor matrices, $R_1$ and $C_1$, were calculated from data matrix $D$. Because the target vectors for the four metals were known from the standard samples, target transforming matrix $T$ was calculated as shown in Table 4. By target transformation the concentration and the current profile matrix, $C_0$ and $R_0$, were found. Using the ELECTTFA program, the concentrations of unknown samples and their recoveries, as well as relative deviation, are listed in tables 5 and 6. The experimental results proved that target transformation factor analysis provides satisfactory results for simultaneous determination of synthetic stripping voltammogram.

**Conclusion**

A method has been described for simultaneous determination of four stripping voltammograms. It can be used to analyse the whole voltammogram rather than just picking out a few characteristic values. The analysis of simulated stripping voltammograms shows how the method can be used to identify the components and to yield a quantitative analysis of each mixture. The target matrix is calculated
from the standard samples by least squares, instead of the pure component standard. The number of components is determined by inspection of both computed results and figures. The method was testified by simulated voltammograms with satisfactory results.

Acknowledgement

The authors would like to thank the National Natural Science Foundation of China for financial support.

References

1. Windig, W., Chemometrics and Intelligent Laboratory system, 23 (1994), 71.
2. Xie, Y., Liang, Y. and Yu, R., Analytica Chimica Acta, 272 (1993), 61.
3. Kubista, M., Sjoberg, R., and Nygren, J., Analytica Chimica Acta, 302 (1995), 121.
4. Malinowski, E. R. and Howery, D. E., Factor Analysis in Chemistry (Wiley, New York, 1985).
5. Wierz, D. F. and Blades, M. W., Talanta, 37 (1990), 39.
6. Cladera, A., Gomez, E., Estela, J., and Cerdá, V., Analytical Chemistry, 65 (1993), 707.
7. Ren, S. and Gao, L., Journal of Automatic Chemistry, 17 (1995), 115.