การพัฒนาโปรแกรมวิเคราะห์ตัวประกอบอย่างง่ายสำหรับการประยุกต์ทางสเปกโทรเคมี

นายอรุณชัย ตั้งเจริญบำรุงสุข

DEVELOPMENT OF SIMPLE FACTOR ANALYSIS PROGRAM
FOR SPECTROCHEMICAL APPLICATION

MR ARUNCHAI TANGCHAROENBUMRUNGSUK

Thesis Title	Development of Simple Factor Analysis Program for Spectrochemical Application
By	Mr Arunchai Tangcharoenbumrungsuk
Department	Chemistry
Thesis Advisor	Associate Professor Vudhichai Parasuk, Ph.D.
Co-advisor	Associate Professor Siri Varothai, Ph.D.
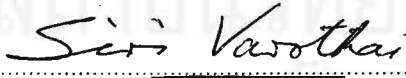
Accepted by the Graduate School, Chulalongkorn University in Partial Fulfilment of the Requirement for the Master's Degree.
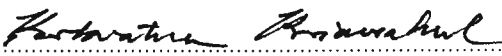
.... Dean of Graduate School
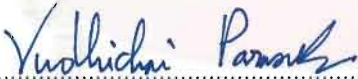
(Associate Professor Suchada Kiranandana, Ph.D.)

Thesis Committee

.... Chairman

(Associate Professor Sirirat Kokpol, Ph.D.)

.... Thesis Advisor

(Associate Professor Vudhichai Parasuk, Ph.D.)

.... Thesis Co-advisor

(Associate Professor Siri Varothai, Ph.D.)

.... Member

(Assistant Professor Korbratna Kriausakul, Ph.D.)

อรุณชัย ตั้งเจริญบำรุงสุข: การพัฒนาโปรแกรมวิเคราะห์ตัวประกอบอย่างง่ายสำหรับ
การประยุกต์ทางสเปกโทรเคมี (DEVELOPMENT OF SIMPLE FACTOR ANALYSIS
PROGRAM FOR SPECTROCHEMICAL APPLICATION) อาจารย์ที่ปรึกษา:
รศ.ดร.วุฒิชัย พาราสุข, อาจารย์ที่ปรึกษาร่วม: รศ.ดร.ศิริ วโรทัย, 95 หน้า.
ISBN 974-333-076-3.

การวิเคราะห์ตัวประกอบเป็นวิธีการศึกษาแบบหลายตัวแปรของข้อมูลชนิดเมตริกซ์ ได้
พัฒนาโปรแกรม MTFA เวอร์ชัน 1.0 ขึ้นเพื่อใช้วิเคราะห์ข้อมูลจากสเปกตรัมดูดกลืนรังสี
อัลตราไวโอเลต/วิสิเบิล โปรแกรม MTFA ประกอบด้วยไฟล์ 3 ไฟล์ คือ MTFA1.FOR, Matrix.FOR
และ Print.FOR ซึ่งเป็นโปรแกรมที่เขียนด้วยภาษาฟอร์แทรนของ Microsoft FORTRAN 5.1 บนเครื่อง
คอมพิวเตอร์แบบส่วนบุคคล ในขั้นแรกจะเป็นการทดสอบประสิทธิภาพและพิสูจน์ความถูกต้องของ
โปรแกรมโดยสร้างสเปกตรัมจำลองจากองค์ประกอบที่สมมติขึ้น 4 องค์ประกอบแล้วบวกหรือลบด้วย
ความผิดพลาดแบบสุ่ม 2 ระดับ คือ 0.0005 และ 0.0015 หน่วยแอบซอร์บแบนซ์ พบว่า ผลของการ
ทดสอบสเปกตรัมจำลองมีความถูกต้องสูง ต่อมาได้ใช้โปรแกรมดังกล่าววิเคราะห์สเปกตรัมจากการ
ทดลองสเปกโทรเมทริกไทเทรชันของการเกิดสารประกอบเชิงซ้อนในระบบระหว่างทองแดง (Cu^{2+})
กับไกลซีน (GlyH) และระหว่างทองแดงกับอะลานีน (AlaH) ในช่วงที่เป็นกรดระหว่างพีเอช 1 ถึง 7
ผลจากการทดสอบยืนยันว่า ทั้ง 2 ระบบมีองค์ประกอบที่ดูดกลืนแสง 4 องค์ประกอบ คือ สำหรับระบบ
ของทองแดงกับไกลซีนจะเป็น Cu^{2+}, $CuGlyH^{2+}$, $CuGly^+$ และ $CuGly_2$ และระบบของทองแดงกับ
อะลานีนจะเป็น Cu^{2+}, $CuAlaH^{2+}$, $CuAla^+$ และ $CuAla_2$ ตามลำดับ จากการวิเคราะห์โดยโปรแกรมจะ
สามารถหาค่าความเข้มข้นและโมลาร์แอบซอร์บติวิตีของแต่ละองค์ประกอบ รวมถึงค่าคงที่สมดุลของ
การเกิดสารประกอบเชิงซ้อนของทั้ง 2 ระบบได้

ภาควิชา   เคมี
สาขาวิชา   เคมี ฟิสิคัล
ปีการศึกษา   2542

ลายมือชื่อนิสิต อรุณชัย ตั้งเจริญบำรุงสุข
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

By means of the Factor Analysis (FA), a multivariate technique for studying matrices of data, the Program MTFA version 1.0 was developed to resolve the UV/visible absorption spectra. The MTFA contains three files: MTFA1.FOR, Matrix.FOR, and Print.FOR implemented in fortran language using Microsoft FORTRAN 5.1 on personal computer. At first, the efficiency and validation of the program were tested by the simulated spectra which generated by four artificial components and plus/minus by two levels of random error: 0.0005, and 0.0015 absorbance units. The simulated spectra were able to be resolved by the program with high accuracy. The program was later used to discriminate the experimental spectra of the complexes formed between Cu(II) and glycine (GlyH), and between Cu(II) and alanine (AlaH) during the pH titration in the acid region ranging from 1 to 7. The program results strongly establish that there are "four" components associated the both systems. In such system of Cu-GlyH, the four components are Cu^{2+}, $CuGlyH^{2-}$, $CuGly^{+}$ and $CuGly_2$; and Cu^{2+}, $CuAlaH^{2+}$, $CuAla^{-}$ and $CuAla_2$ for the system of Cu-AlaH, respectively. In addition, by the use of concentration profiles and molar absorptivities obtained from the program, the complex formation constants of Cu-GlyH and Cu-AlaH were obtained.

ภาควิชา........เคมี........

สาขาวิชา........เคมี ฟิสิคัล........

ปีการศึกษา........2542........

ลายมือชื่อนิสิต........................

ลายมือชื่ออาจารย์ที่ปรึกษา........................

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม........................

ACKNOWLEDGEMENTS

CONTENTS

จุฬาลงกรณ์มหาวิทยาลัย

LIST OF FIGURES

pages

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In recent years, the use of computers to solve chemical problems has increased tremendously. Chemometrics is one of those areas where mathematical and statistical methods are used to handle, interprete, and predict chemical data.[1-2] Powerful methodologies have opened new vistas for chemist and provided useful solutions for many complex chemical problems. Factor analysis has proved to be one of the most potent techniques in the chemometric arsenal.

1.1 Definition of Factor Analysis

Factor analysis (FA) was founded by the behavioral scientists in 1901 but the first real development was accomplished in 1933. Although an idea tool for solving chemical problem, the method went unnoticed by the chemical profession until the birth of chemometrics in 1970s and the arrival of the computer era.

Today, chemists and scientists in general are familiar with computers, mathematics, and statistics, the prerequisites for factor analysis. The field has grown so large that it is impossible to examine all the methodologies. So, the definition of factor analysis has changed over the years, encompassing a much wider selection of techniques than originally intended. The global definition of factor analysis is given as "a multivariate technique for reducing matrices of data to their lowest dimensionality by the use of orthogonal factor space and transformations that yield predictions and/or recognizable factors."[1]

1.2 Generalizations

This section concerns notation and terminology, which can be used in a general manner. Data matrix $\mathbf{D}$, consisting of r rows and c columns, is writen as

$$\mathbf{D} = \underset{\text{row designee}}{} \begin{bmatrix} d_{11} & d_{12} & \dots d_{1c} \\ d_{21} & d_{22} & \dots d_{2c} \\ \vdots & \vdots & \vdots \\ d_{r1} & d_{r2} & \dots d_{rc} \end{bmatrix} \qquad (1.1).$$

column designee

The row and column headings of the matrix are called *designee*. Each measured data point in $\mathbf{D}$ is specified by a subscript denoting its row and column position in the matrix. The symbol d_{ik} represents the data point associated with the ith row and kth column of the matrix.

Along the way of factor analysis, we assume that each point in the data matrix must be a linear sum of product terms. The number of terms in the sum, n, is called the *number of factors*. Specifically, we seek solutions of the form

$$d_{ik} = \sum_{j=1}^{n} r_{ij} c_{jk} \qquad (1.2)$$

where r_{ij} and c_{jk} are called factors. For the jth factor in the sum, the *row factor* r_{ij} is associated with the ith row of data matrix and the corresponding *column factor* c_{jk} is associated with the kth column of the matrix.

For data modeled by (1.2), the data matrix can be decomposed into two matrices

$$\mathbf{D} = \mathbf{R}_{\text{abstract}} \cdot \mathbf{C}_{\text{abstract}} \qquad (1.3)$$

data matrix   row matrix   column matrix

where

$$\mathbf{R}_{abstract} = \begin{array}{c} \textit{factor} \\ \textit{row designee} \end{array} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{r1} & r_{r2} & \cdots & r_{rn} \end{bmatrix}$$

$$\mathbf{C}_{abstract} = \begin{array}{c} \textit{column designee} \\ \textit{factor} \end{array} \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1c} \\ c_{21} & c_{22} & \cdots & c_{2c} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nc} \end{bmatrix}.$$

Since this solution is purely mathematical thus it devoids physical meaning, these matrices are called *abstract matrices*. The column of $\mathbf{R}_{abstract}$ are called *abstract factors*. *Row matrix* $\mathbf{R}_{abstract}$ contains a row for each of the r row designees and a column for each of the n factors, while *column matrix* $\mathbf{C}_{abstract}$ has a column for each of the c column designees and a row for each factor. Explicitly, the factor analytical solution isolates the row-designee factors from the column-designee factors.

The following step of the factor analysis is the development of a complete, physically meaningful model for the data. To do this, the abstract factors are mathematically "transformed" into physically significant, "real" factors. Transforming the abstract solution into a real solution is a difficult but realizable goal of factor analysis.

To carry out the transformations, an appropriate information matrix, $\mathbf{T}$, is required. Postmultiplying $\mathbf{R}_{abstract}$ by $\mathbf{T}$ and premultiplying $\mathbf{C}_{abstract}$ by the inverse of the transformation $\mathbf{T}^{-1}$, the data matrix in (1.3) can be expressed as

$$\begin{aligned} \mathbf{D} &= \{\mathbf{R}_{abstract}\mathbf{T}\}\{\mathbf{T}^{-1}\mathbf{C}_{abstract}\} \\ &= \mathbf{R}_{transformed}\mathbf{C}_{transformed} \end{aligned} \qquad (1.4).$$

If the transformed solution can be shown to have physical significance, a real solution to the problem will have been found so that

$$\mathbf{D} = \mathbf{X}_{real}\mathbf{Y}_{real} \qquad (1.5)$$

where $\mathbf{X}_{real} = \mathbf{R}_{transformed}$ and $\mathbf{Y}_{real} = \mathbf{C}_{transformed}$. This equation summarizes the ultimate objective of factor analysis.

1.3 Chemical Application

Factor analysis has been applied to several chemical problems such as chemical kinetic study from overlapped fluorescence emission spectra [3], overlapped chromatograms of binary and ternary mixtures from gas chromatography/mass spectrometry (GC/MS) [4], overapped chromatographic peaks of eight-components mixtures from liquid chromatography with photodiode-array ultraviolet detecter (LC/UV) [5], quantitative analysis of overlapped Raman spectra [6], impurities monitoring using high performance liquid chromatography (HPLC) [7], acid-base equilibria of monoprotic organic acid-base pairs from UV/VIS spectrophotometric titration [8] and so on.

Let us consider the simple case of UV/VIS spectrophotometry. Suppose a data matrix, $\mathbf{A}$, involving the ultraviolet absorbance of five different mixtures of the same absorbing components measured at six wavelengths:

$$
\mathbf{A} =
\begin{array}{c}
\text{wavelength} \\
278\,\text{nm} \\
274\,\text{nm} \\
270\,\text{nm} \\
266\,\text{nm} \\
262\,\text{nm} \\
258\,\text{nm}
\end{array}
\begin{array}{ccccc}
 & & \text{mixture} & & \\
1 & 2 & 3 & 4 & 5 \\
\left[\begin{array}{ccccc}
0.005 & 0.031 & 0.063 & 0.091 & 0.046 \\
0.040 & 0.172 & 0.356 & 0.444 & 0.218 \\
0.103 & 0.283 & 0.484 & 0.471 & 0.208 \\
0.116 & 0.323 & 0.562 & 0.548 & 0.241 \\
0.125 & 0.318 & 0.516 & 0.450 & 0.185 \\
0.104 & 0.267 & 0.430 & 0.376 & 0.154
\end{array}\right]
\end{array}
\qquad (1.6).
$$

The main problem here is to determine the number of components, to identify the chemical constituents, and to ascertain their concentrations.

According to (1.2), factor analysis will **automatically** furnish an abstract solution for each absorbance datum, A_{ik}, in the form

$$A_{ik} = \sum_{j=1}^{n} w_{ij} m_{jk} \tag{1.7.}$$

Here w_{ij} and m_{jk} are the jth abstract row and column factors associated with the ith wavelength and the kth mixture, respectively. To account for the absorbance within experimental error, n factors are included in the sum. According to (1.7), the absorbance data matrix has an abstract factor analytical solution expressed by

$$\mathbf{A} = \mathbf{W}_{abstract} \mathbf{M}_{abstract} \tag{1.8}$$

where $\mathbf{W}_{abstract}$ and $\mathbf{M}_{abstract}$ are wavelength-factor and mixture-factor matrices, respectively.

The most important feature of the abstract solution is that it reveals the number of factors responsible for the absorbance data. Ultimately, we search for an appropriate transformation matrix that will convert the abstract solution into a physically significant real solution

$$\mathbf{A} = \mathbf{W}_{real} \mathbf{M}_{real} \tag{1.9.}$$

Going from (1.8) to (1.9) is not automatic. On the contrary, this step presents the most difficult challenge to chemists, requiring a great deal of effort, knowledge, and intuition. If theoretical speculations can be invoked, the transformation has a better chance of being successful.

In this case, the absorbance data obeys the Beer's law, therefore, the factor can be interpreted chemically. For a mixture containing n absorbing components, Beer's law models each absorbance datum by the equation

$$A_{ik} = \sum_{j=1}^{n} \varepsilon_{ij} c_{jk} \tag{1.10.}$$

Here ε_{ij} is the molar absorptivity per unit pathlength of component j at wavelength i, and c_{jk} is the molar concentration of component j in the kth mixture. Equation (1.10) involves a

linear sum of products analogous to (1.7); therefore, data that obey Beer's law should have meaningful factor analytical solution. To solve the problem completely, we must find the transformation matrix that will convert the abstract solution into the real solution. When this is done correctly, (1.9) will take the form

$$\mathbf{A} = \mathbf{E}_{real}\mathbf{C}_{real} \tag{1.11}$$

Each column of the molar absorptivity matrix, $\mathbf{E}_{real}$, corresponds to the absorbance of one of the pure components at the five wavelengths, essentially tracing out the spectrum of the pure component. Each row of the molar concentration matrix, $\mathbf{C}_{real}$, corresponds to the concentrations of one of the n components in each of the five mixtures.

In summary, the ultimate payoff from factor analysis in this type of problem is to determine:

1. The number of absorbing components.
2. The concentration of each component in each mixture.
3. The spectrum of each component.

The factor analysis approach is far more useful than the popular determinant method for finding the concentrations of components in multicomponent mixtures, since the spectra of all components must be specified initially in the latter approach. By contrast, factor analysis can furnish the number of components, the concentrations, and the spectral information via a purely mathematical route.

1.4 Scope of this Study

In this study, the Program MTFA version 1.0 was developed to execute the factor analysis. The program was written in fortran language using Microsoft FORTRAN V5.1 compiler for personal computer. The main application here is to resolve the UV/VIS absorption spectra for the acid-base equilibria of the polyprotic organic acid-base pairs and also the formation of metal-ligand complexes. The disposition of the program should be interactive to user, and provided the manual and automatic selections. At first, the efficiency and validation of the program are tested by simulated spectra and the experimental spectra of the complexes formed between Cu(II) and glycine (GlyH), and between Cu(II) and alanine (AlaH) during the pH titration in acid region ranging from 1 to 7.

CHAPTER 2

THEORETICAL CONSIDERATION

This chapter summarizes methodologies of the factor analysis, focusing on spectrochemical problems with details of mathematical formulation.

2.1 Factor Analysis Programing

Factor analysis involves five main steps; preparation, reproduction, transformation, combination, and prediction. Figure 2.1 shows the sequencing of the steps and the most important information resulting from each step.
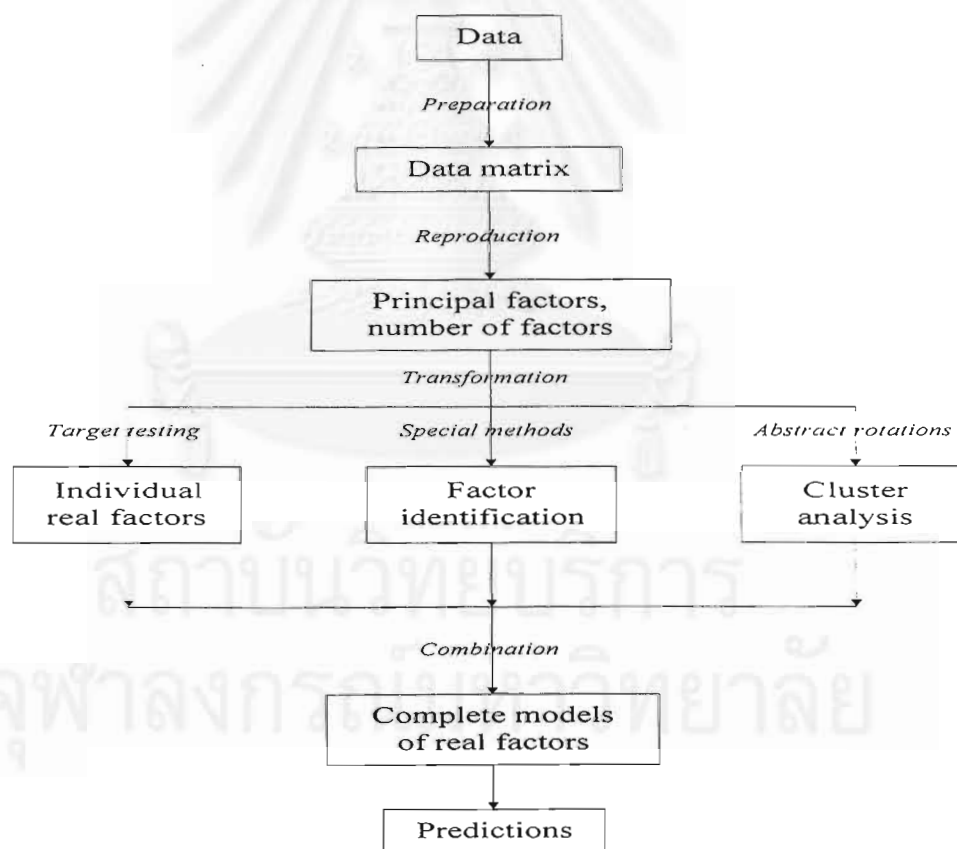
Figure 2.1 Block diagram of the main steps in factor analysis

2.1.1 Preparation

The objective of the data preparation step is to obtain a data matrix best suited for factor analysis. This step involves formulating the problem, selecting the data, and mathematically pretreating the data to conform with appropriate theoretical or statistical criteria.

2.1.2 Reproduction

The abstract reproduction step is the mathematical underpinning of factor analysis. Reproduction involves two procedures: obtaining the "principal" factor solution and determining the correct number of factors.

Principal Factor Solution

The procedure for calculating the abstract solution are called *eigenanalysis*, yielding eigenvalues and associated eigenvectors. Several commonly used methods are listed below:

- Power method [2;9]
- Jacobi method [2;9]
- Singular Value Decomposition (SVD) [2;8-9]
- Nonlinear Iterative Partial Least Squares (NIPALS) [2;9-10]
- Principal Factor (or Component) Analysis (PFA or PCA) [2;11]

Number of Factors

After following the eigenanalysis, we seek to discover how many of the c factors are physically important. The abstract factors can be divided into two sets; a primary set of n factors which account for the real measurable features of the data, and a secondary set of $c-n$ factors called the *null* set, which are associated entirely with experimental error. To eliminate

the secondary factors from the initial solution, various techniques were developed. In general we classify these into three classes:

1. Methods based on experimental error

   - Residual Standard Deviation (RSD)

   - Root-Mean-Square (RMS) Error

   - Average Error

   - Chi-Squared [2;12]

   - Standard Error in the Eigenvalue [2;13]

2. Empirical methods

   - Imbedded Error Function (IE) [2;14]

   - Factor Indicator Function (IND) [2;14]

   - Cumulative Percent Variance [2]

3. Statistical methods

   - Reduced Eigenvalue and Statistical F-Test (or Percentage of Significant Level, %SL) [2;9;15-17]

When the correct number of factors is employed, the reproduced data matrix should be reconstructed. This means that the reproduced data matrix is more accurate than the original data matrix.

2.1.3 Transformation

Transformation of principal factors into recognizable parameters is the most important dividend of factor analysis. As explained in Section 1.2, a transformation matrix is employed to carry out a transformation. Three distinct methods are introduced to transform the eigenanalysis solution:

1. Abstract Factor Analysis (AFA); involving mathematical rotations.

2. Target Factor Analysis (TFA); involving target testing.

3. Special methods; involving known chemical constraints e.g. Key Set Factor
   Analysis (KSFA), Partial Least Squares (PLS), Evolving Factor Analysis (EFA),
   Rank Annihilation Factor Analysis (RAFA), Iterative Factor Analysis (IFA),
   Evolutionary Factor Analysis (EVOLU), Variance Diagram (VARDIA), and so on.

By transforming the principal factors, the *real* factors, which describe the properties of the designees in the data matrix, are obtained.

2.1.4 Combination and Predictions

In the combination step, real factors are combined to complete models and tested by principal component regression (PCR). Finally, the models are used to predict the missing data and/or the unknown samples in multivariate calibration.

2.2 Spectrochemical Problems of Copper-Glycine and Copper-Alanine Complexations

As described in Section 1.3, the data obtained by the UV/VIS spectrophotometry are applicable to the factor analysis. In this study, the absorption of copper(II)-glycine (GlyH), and copper(II)-alanine (AlaH) complexes were used to test our own-written factor analysis program.

Glycine, H_2NCH_2COOH, is the simplest amino acid. It is a zwitterion, containing an amino group and a carboxylic acid group, exhibiting properties of both acid and base. The equilibria of glycine are known as [18;34-35]

$$H_3N^+CH_2CO_2H \xrightarrow{\;pK_{a1}=2.350\;} H_3N^+CH_2CO_2^- \xrightarrow{\;pK_{a2}=9.778\;} H_2NCH_2CO_2^- \qquad (2.1).$$
$$\;(GlyH_2^+) \qquad\qquad\qquad \text{zwitterion (GlyH)} \qquad\qquad (Gly^-)$$
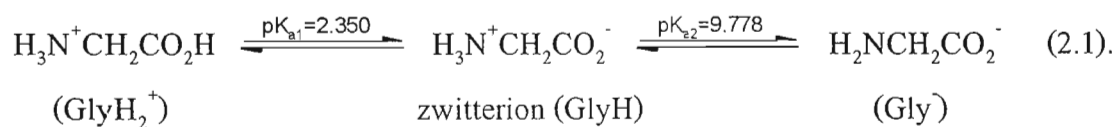
Alanine, $H_2NCH(CH_3)COOH$, is a simple amino acid. Similar to glycine, the equilibria of alanine are known as [34-35]

$$H_3N^+CH(CH_3)CO_2H \xrightarrow{pK_{a1}=2.36} H_3N^+CH(CH_3)CO_2^- \xrightarrow{pK_{a2}=9.78} H_2NCH(CH_3)CO_2^- \quad (2.2).$$

$\quad$ (AlaH$_2^+$) $\qquad\qquad\qquad$ zwitterion (AlaH) $\qquad\qquad\qquad$ (Ala$^-$)
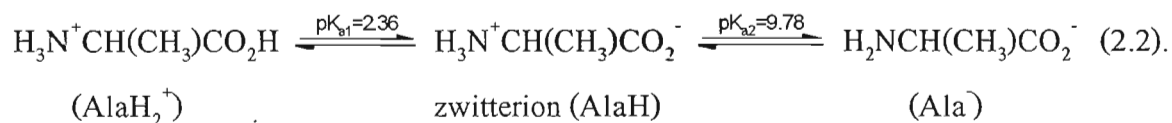
Both glycine and alanine are known to complex with copper(II) forming different species. Deciphering the nature of the complexes formed between cation and complexing agents (such as copper-glycine, and copper-alanine) is not an easy task when the complexes exist in dynamic equilibrium and cannot be isolated chemically. In practice, it depends on the model suggestion and sensitivity of the analytical techniques. Spectroscopic methods are in general highly sensitive and suitable for studying chemical equilibria in solution but the spectra are often complicated and difficult to interprete owing to high spectral overlapping. Thus the spectroscopic methods were overlooked until the advent of chemometrics and the factor analysis.

Darj and Malinowski [18] used the window factor analysis (WFA), a self-modeling chemometric method, to evaluate the visible spectra of Cu(II) and glycine complexes, and expressed the complex formation as;

$$Cu^{2+} + GlyH \xrightarrow{K_1} CuGlyH^{2+}$$

$$CuGlyH^{2+} \xrightarrow{K_2} CuGly^+ + H^+$$

$$CuGly^+ + GlyH \xrightarrow{K_3} CuGly_2 + H^+ \qquad (2.3)$$
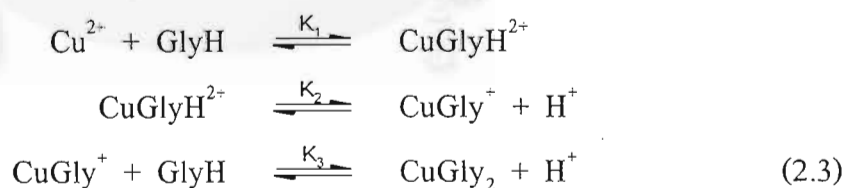
where K_1, K_2, and K_3 represent the stepwise equilibrium constants. These constants can be expressed in terms of activity coefficients, f_i, and concentrations;

$$K_1 = \frac{f_{CuGlyH^{2+}}[CuGlyH^{2+}]}{f_{Cu^{2+}}[Cu^{2+}]f_{GlyH}[GlyH]}$$

$$K_2 = \frac{f_{CuGly^+}[CuGly^+]f_{H^+}[H^+]}{f_{CuGlyH^{2+}}[CuGlyH^{2+}]}$$

$$K_3 = \frac{f_{CuGly_2}[CuGly_2]f_{H^+}[H^+]}{f_{CuGly^+}[CuGly^+]f_{GlyH}[GlyH]} \tag{2.4}.$$

Analogous to the system of glycine, the complex formation of Cu(II) and alanine ought to be

$$Cu^{2+} + AlaH \quad \xrightleftharpoons{K_1} \quad CuAlaH^{2+}$$

$$CuAlaH^{2+} \quad \xrightleftharpoons{K_2} \quad CuAla^+ + H^+$$

$$CuAla^+ + AlaH \quad \xrightleftharpoons{K_3} \quad CuAla_2 + H^+ \tag{2.5}$$
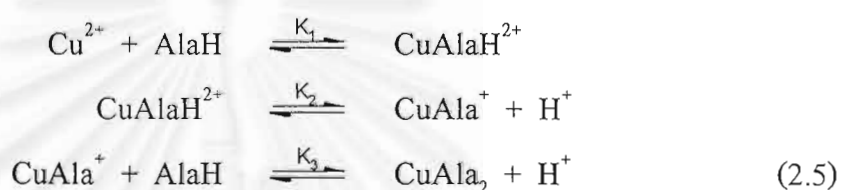
where K_1, K_2, and K_3 represent the stepwise equilibrium constants and express as

$$K_1 = \frac{f_{CuAlaH^{2+}}[CuAlaH^{2+}]}{f_{Cu^{2+}}[Cu^{2+}]f_{AlaH}[AlaH]}$$

$$K_2 = \frac{f_{CuAla^+}[CuAla^+]f_{H^-}[H^+]}{f_{CuAlaH^{2+}}[CuAlaH^{2+}]}$$

$$K_3 = \frac{f_{CuAla_2}[CuAla_2]f_{H^+}[H^+]}{f_{CuAla^-}[CuAla^+]f_{AlaH}[AlaH]} \tag{2.6}.$$

Here the modeling method, a subdivision of evolutionary factor analysis (EVOLU), was selected since they tend to be more force fitting and are called "hard" models, whereas self-modeling methods tend to be more revealing but more ambiguous (in some cases) and are called "soft" models such as window factor analysis (WFA), rank annihilation factor analysis (RAFA), and so on. Along this way, the number of components comes out first and makes a clue to propose the chemical models. If the models are well-established, the equilibrium constants, concentration profiles and spectral absorptivities of the Cu(II)-glycine, and Cu(II)-alanine complexs can be obtained successively.

2.3 Fundamental Concept of Chemical Equilibria

This section concerned about the fundamental concept of chemical equilibria that underlying the problem of Cu(II)-glycine and Cu(II)-alanine complexes, and the system of acid deprotonation.

There are two particular equilibrium constants that are commonly given special names. [19] Thus when the Lewis acid is a proton, the inverse of the equilibrium constant for the reaction

$$H^+ + X^- \xrightleftharpoons{K_{HX}} HX$$

$$K_a^\theta = \frac{1}{K_{HX}} = \frac{f_{H^+}[H^+]f_{X^-}[X^-]}{f_{HX}[HX]} \qquad (2.7)$$
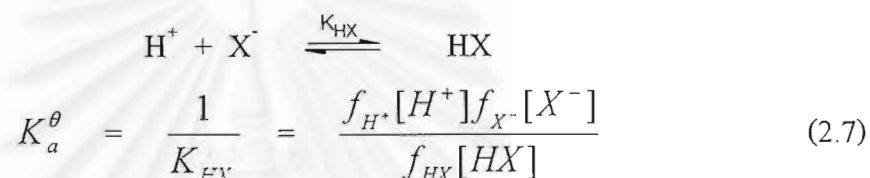
is known as the *acid dissociation constant* (K_a^θ) of the acid HX. The second special case is when the Lewis acid is a metal ion (M) and the Lewis base is a ligand (L), then for the reaction

$$M + L \xrightleftharpoons{K_{ML}} ML$$

$$K_{ML}^\theta = \frac{f_{ML}[ML]}{f_M[M]f_L[L]} \qquad (2.8)$$
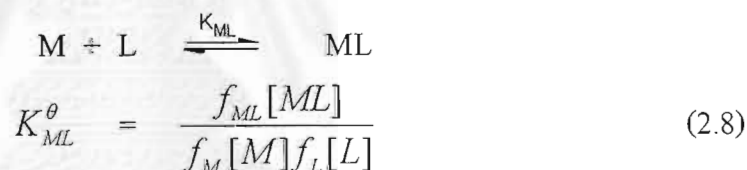
the equilibrium constant K_{ML}^θ is known as a *stability (or formation) constant* of the complex ML. The activity coefficients are in general tedious and difficult to measure. They also depend very significantly on the nature and concentrations of the other species present in solution. To avoid this problem, the background electrolyte (or sometimes called ionic-strength adjustor) is used to maintain the activity coefficients effectively constant. Hence we can incorporate the f_i terms into K_a^θ or K_{ML}^θ and obtain the practical forms as

$$K_a = \frac{[H^+][X^-]}{[HX]} \qquad (2.9)$$

$$K_{ML} = \frac{[ML]}{[M][L]} \qquad (2.10)$$

where K (no superscript) are known as *stoichiometric equilibrium constants* whereas K^θ are known as *thermodynamic equilibrium constants*.

2.3.1 Acid-Base Equilibria

Consider a system of k steps of dissociation of an acid denoted as H_kA. The equilibria present are:

$$H_kA \rightleftharpoons H^+ + H_{k\text{-}1}A^- ; \quad K_{a1} = \frac{[H^+][H_{k-1}A^-]}{[H_kA]}$$

$$H_{k\text{-}1}A^- \rightleftharpoons H^+ + H_{k\text{-}2}A^{2-} ; \quad K_{a2} = \frac{[H^+][H_{k-2}A^{2-}]}{[H_{k-1}A^-]}$$

$$\vdots \qquad\qquad \vdots$$

$$HA^{(k\text{-}1)\text{-}} \rightleftharpoons H^+ + A^{k\text{-}} ; \quad K_{ak} = \frac{[H^+][A^{k-}]}{[HA^{(k-1)-}]} \qquad (2.11).$$
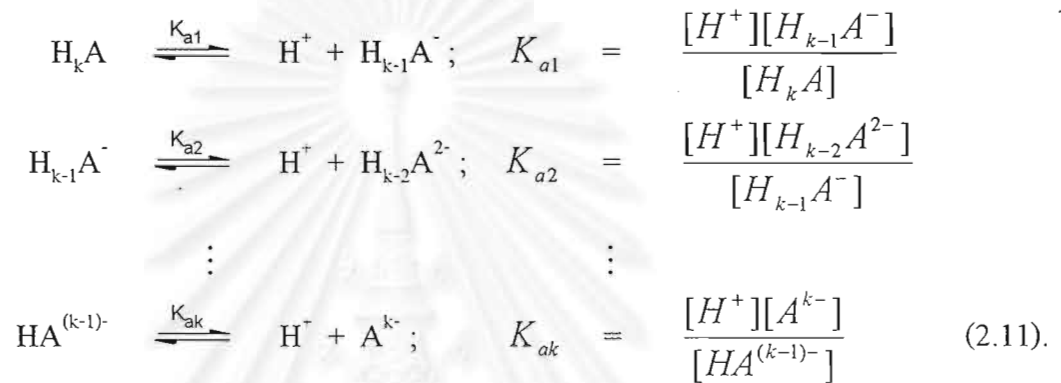
To express the distribution of each component, we introduce the concept of degree of formation, α, the mole ratio of one component with respect to all components. Then

$$\alpha_1 = \frac{[H_kA]}{[H_kA]+[H_{k-1}A^-]+...+[A^{k-}]}$$

$$\alpha_2 = \frac{[H_{k-1}A^-]}{[H_kA]+[H_{k-1}A^-]+...+[A^{k-}]}$$

$$\vdots \qquad\qquad \vdots$$

$$\alpha_{k+1} = \frac{[A^{k-}]}{[H_kA]+[H_{k-1}A^-]+...+[A^{k-}]} \qquad (2.12).$$

Substituting in (2.12) for (2.11) and rearranging, we obtain:

$$\alpha_1 = \cfrac{1}{1 + \cfrac{K_{a1}}{[H^+]} + \cfrac{K_{a2}K_{a1}}{[H^+]^2} + \ldots + \cfrac{K_{ak}K_{a(k-1)}\ldots K_{a1}}{[H^+]^k}}$$

$$\alpha_2 = \cfrac{1}{\cfrac{[H^+]}{K_{a1}} + 1 + \cfrac{K_{a2}}{[H^+]} + \cfrac{K_{a3}K_{a2}}{[H^+]^2} + \ldots + \cfrac{K_{ak}K_{a(k-1)}\ldots K_{a2}}{[H^+]^{k-1}}}$$

$$\alpha_3 = \cfrac{1}{\cfrac{[H^+]^2}{K_{a1}K_{a2}} + \cfrac{[H^+]}{K_{a2}} + 1 + \cfrac{K_{a3}}{[H^+]} + \cfrac{K_{a4}K_{a3}}{[H^+]^2} + \ldots + \cfrac{K_{ak}K_{a(k-1)}\ldots K_{a3}}{[H^+]^{k-2}}}$$

$$\vdots \qquad\qquad \vdots$$

$$\alpha_{k+1} = \cfrac{1}{\cfrac{[H^+]^k}{K_{a1}K_{a2}\ldots K_{ak}} + \cfrac{[H^+]^{k-1}}{K_{a2}K_{a3}\ldots K_{ak}} + \ldots + \cfrac{[H^+]}{K_{ak}} + 1} \qquad (2.13).$$

The general form of the equation (2.13) may be written as

$$\frac{1}{\alpha_m} = \left(\sum_{i=0}^{m-1} \frac{[H^+]^i}{\prod_{j=1}^{i} K_{a(m-j)}} \right) + \left(\sum_{\substack{i=1 \\ m<k+1}}^{k-m+1} \frac{\prod_{j=m}^{i+m-1} K_{a(j)}}{[H^+]^i} \right) \qquad (2.14).$$

2.3.2 Metal-Ligand Complex Equilibria

Regarding to a system of p steps of formation of metal (M) and protonated ligand (HL) that are expressed by the equation

$$M + HL \xrightleftharpoons{K_1} ML + H^+; \qquad K_1 = \frac{[ML][H^+]}{[M][HL]}$$

$$ML + HL \xrightleftharpoons{K_2} ML_2 + H^+; \qquad K_2 = \frac{[ML_2][H^+]}{[ML][HL]}$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$ML_{p-1} + HL \xrightleftharpoons{K_p} ML_p + H^+; \qquad K_p = \frac{[ML_p][H^+]}{[ML_{p-1}][HL]} \qquad (2.15).$$
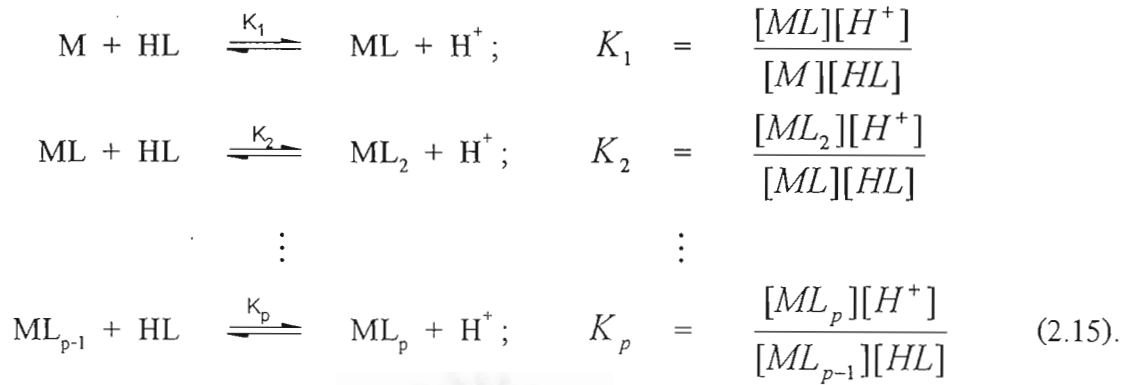
In many literature, the concept of *overall* or *cumulative stability constants*, usually denoted by β_i, were introduced as

$$M + HL \xrightleftharpoons{\beta_1} ML + H^+$$

$$M + 2HL \xrightleftharpoons{\beta_2} ML_2 + 2H^+$$

$$\vdots$$

$$M + pHL \xrightleftharpoons{\beta_p} ML_p + pH^+$$

where
$$\beta_1 = K_1 = \frac{[ML][H^+]}{[M][HL]}$$

$$\beta_2 = K_1 K_2 = \frac{[ML_2][H^+]^2}{[M][HL]^2}$$

$$\vdots \qquad\qquad \vdots$$

$$\beta_p = \prod_{i=1}^{p} K_i = K_1 K_2 ... K_p = \frac{[ML_p][H^+]^p}{[M][HL]^p} \qquad (2.16).$$
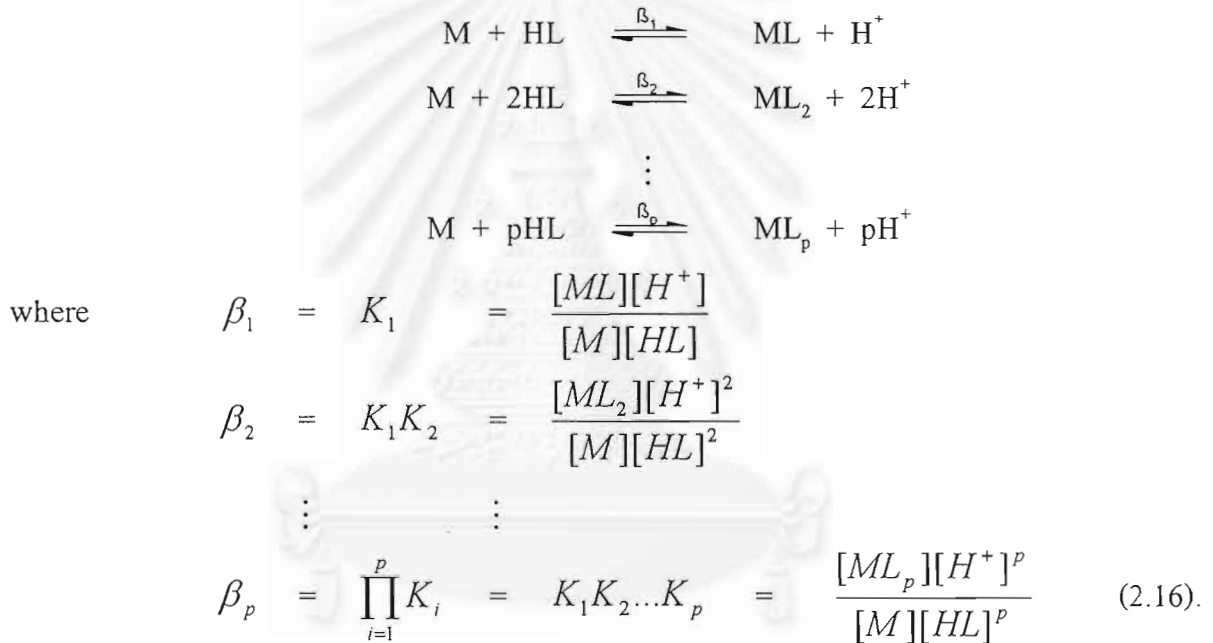
To make it more general, we applied the variable-chemical model as

$$(x_1)M + (y_1)HL \xrightleftharpoons{\beta_1} ML^{(1)} + (z_1)H^+$$

$$(x_2)M + (y_2)HL \xrightleftharpoons{\beta_2} ML^{(2)} + (z_2)H^+$$

$$\vdots$$

$$(x_p)M + (y_p)HL \xrightleftharpoons{\beta_p} ML^{(p)} + (z_p)H^+$$
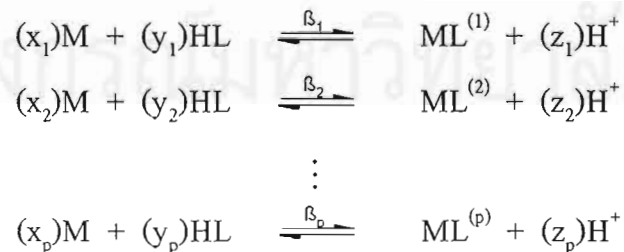
where
$$\beta_1 = \frac{[ML^{(1)}][H^+]^{z_1}}{[M]^{x_1}[HL]^{y_1}}$$

$$\beta_2 = \frac{[ML^{(2)}][H^+]^{z_2}}{[M]^{x_2}[HL]^{y_2}}$$

$$\vdots$$

$$\beta_p = \frac{[ML^{(p)}][H^+]^{z_p}}{[M]^{x_p}[HL]^{y_p}} \tag{2.17}.$$

By the principle of mass balance and the assumption of mononuclear complex, we obtain

$$M_{tot} = [M] + [ML^{(1)}] + [ML^{(2)}] + \ldots + [ML^{(p)}]$$

$$[M] = \frac{M_{tot}}{1 + \beta_1 \dfrac{[HL]^{y_1}}{[H^+]^{z_1}} + \beta_2 \dfrac{[HL]^{y_2}}{[H^+]^{z_2}} + \ldots + \beta_p \dfrac{[HL]^{y_p}}{[H^+]^{z_p}}} \tag{2.18}$$

where M_{tot} is the total concentration of metal (M). Since the protonated ligand may exist in dynamic equilibrium with its conjugated acid and/or base as

$$H_2L \xrightleftharpoons{K_{a1}} H^+ + HL; \qquad K_{a1} = \frac{[H^+][HL]}{[H_2L]}$$

$$HL \xrightleftharpoons{K_{a2}} H^+ + L; \qquad K_{a2} = \frac{[H^+][L]}{[HL]}.$$
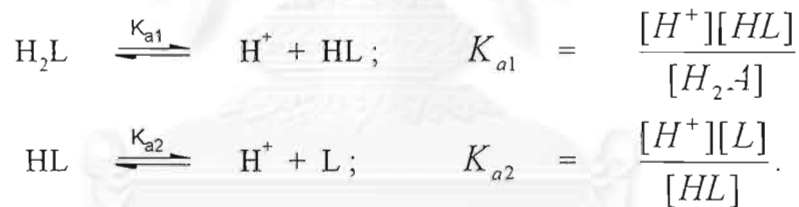
Similar as above, the mass balance of ligand can be expressed as

$$L_{tot} = [H_2L] + [HL] + [L] + y_1[ML^{(1)}] + y_2[ML^{(2)}] + \ldots + y_p[ML^{(p)}]$$

$$= [HL]\left(\frac{[H^+]}{K_{a1}} + 1 + \frac{K_{a2}}{[H^+]}\right) + [M]\left(y_1\beta_1\frac{[HL]^{y_1}}{[H^+]^{z_1}} + y_2\beta_2\frac{[HL]^{y_2}}{[H^+]^{z_2}} + \ldots + y_p\beta_p\frac{[HL]^{y_p}}{[H^+]^{z_p}}\right) \tag{2.19}$$

where L_{tot} is the total concentration of ligand. Substitute (2.18) into (2.19), and applying the binary search [9;20] for determining the concentration of HL. Consequently, evaluate the concentration of M, and substitute successively into (2.17) then obtain the mole ratio of $ML^{(1)}$, $ML^{(2)},\ldots, ML^{(p)}$ as

$$\alpha_1 = \frac{[M]}{[M]+[ML^{(1)}]+[ML^{(2)}]+...+[ML^{(p)}]}$$

$$= \frac{1}{1+\beta_1\dfrac{[HL]^{y_1}}{[H^+]^{z_1}}+\beta_2\dfrac{[HL]^{y_2}}{[H^+]^{z_2}}+...+\beta_p\dfrac{[HL]^{y_p}}{[H^+]^{z_p}}}$$

$$\alpha_2 = \frac{[ML^{(1)}]}{[M]+[ML^{(1)}]+[ML^{(2)}]+...+[ML^{(p)}]}$$

$$= \frac{1}{\dfrac{[H^+]^{z_1}}{\beta_1[HL]^{y_1}}+1+\dfrac{\beta_2}{\beta_1}\dfrac{[HL]^{y_2-y_1}}{[H^+]^{z_2-z_1}}+\dfrac{\beta_3}{\beta_1}\dfrac{[HL]^{y_3-y_1}}{[H^+]^{z_3-z_1}}+...+\dfrac{\beta_p}{\beta_1}\dfrac{[HL]^{y_p-y_1}}{[H^+]^{z_p-z_1}}}$$

$$\alpha_3 = \frac{[ML^{(2)}]}{[M]+[ML^{(1)}]+[ML^{(2)}]+...+[ML^{(p)}]}$$

$$= \frac{1}{\dfrac{[H^+]^{z_2}}{\beta_2[HL]^{y_2}}+\dfrac{\beta_1}{\beta_2}\dfrac{[H^+]^{z_2-z_1}}{[HL]^{y_2-y_1}}+1+\dfrac{\beta_3}{\beta_2}\dfrac{[HL]^{y_3-y_2}}{[H^+]^{z_3-z_2}}+\dfrac{\beta_4}{\beta_2}\dfrac{[HL]^{y_4-y_2}}{[H^+]^{z_4-z_2}}+...+\dfrac{\beta_p}{\beta_2}\dfrac{[HL]^{y_p-y_2}}{[H^+]^{z_p-z_2}}}$$

$$\vdots$$

$$\alpha_{p+1} = \frac{[ML^{(p)}]}{[M]+[ML^{(1)}]+[ML^{(2)}]+...+[ML^{(p)}]}$$

$$= \frac{1}{\dfrac{[H^+]^{z_p}}{\beta_p[HL]^{y_p}}+\dfrac{\beta_1}{\beta_p}\dfrac{[H^+]^{z_p-z_1}}{[HL]^{y_p-y_1}}+\dfrac{\beta_2}{\beta_p}\dfrac{[HL]^{y_p-y_2}}{[H^+]^{z_p-z_2}}+...+1} \qquad (2.20).$$

2.4 Mathematical Synopsis

In this section, the following convention will be used. Scalar quantities (i.e., numbers) are represented by lowecase letters – a, b, c, x, y, and z. Bold, uppercase letters or enclosures in square brackets [ ] signify matrices. Vectors (*i.e.*, one-dimensional arrays of number) are symbolized by bold, lowercase letters – **s**, **t**, **u**, **x**, **y**, and **z**. (In particular, all vectors are considered to be column vectors unless otherwise indicated.) Row vectors are denoted by a prime – **s'**, **t'**, **u'** and **z'**. Matrix transposition, whereby rows and columns are interchanged, are

denoted by a prime. The ^, called "hat," above a quantity signifies an estimated (or calculated) quantity.

In experiments, the absorbance data matrix, $\mathbf{D}$, is constructed whose r rows correspond to the number of wavelengths (channels) and c columns corespond to the number of measured solutions, *i.e.* the dimension of raw data matrix is $r{\times}c$ (assuming that $c \leq r$). The conditions to obtain the data are detailed in Chapter 5.

According to the Beer's law, the total absorbance per unit cell pathlength, d_{ik}, of the kth solution at the ith wavelength is the sum of the absorbance of the n absorbing components as described by

$$d_{ik} = \sum_{j=1}^{n} \varepsilon_{ij} c_{jk} \qquad (1.10), (2.21)$$

where ε_{ij} is the molar absorptivity of the jth component at wavelength i, and c_{jk} is the concentration of the jth component of the solution k. Equation (2.21) can be rewritten in the matrix form, where $\mathbf{E}$ is an $r{\times}n$ matrix with the pure absorption spectra of the n components in its columns, and $\mathbf{C}$ is an $n{\times}c$ matrix with its rows representing the concentration profiles of the n components:

$$\mathbf{D} = \mathbf{EC} \qquad (1.11), (2.22)$$

The problem to be solved by factor analysis here is the estimation of $\mathbf{E}$ and $\mathbf{C}$.

2.4.1 Constructing the Covariance Matrix and Decomposition

The covariance matrix, $\mathbf{Z}$, is obtained by premultiplying the data matrix by its transpose;

$$\mathbf{Z} = \mathbf{D}'\mathbf{D} \qquad (2.23).$$

The resulting covariance matrix is decomposed by use of principal factor analysis (PFA) to determine the eigenvectors that span the factor space. The theory of factor analysis states that the eigenvectors of the covariance matrix are the same vectors that span the space of the data matrix. To obtain the eigenvectors, we employ the method of iteration. Initially, numerical values for the elements of the first principal eigenvector, c_1, were chosen at random, and multiply this vector by the covariance matrix:

$$Zc_1 = \lambda_1 c_1 \qquad (2.24)$$

Here, λ_1 is the corresponding eigenvalue. The product of Zc_1 is then normalized to obtain the new c_1 and again multiply to Z to give a better approximation to c_1 and λ_1. This process is repeated again and again, each time generating newer and better approximation to c_1 and λ_1, until (2.24) is satisfied.

To obtain the second principal eigenvector, we proceed by calculating the first-residual matrix, R_1, as dictated by (2.25)

$$R_1 = Z - \lambda_1 c_1 c_1' \qquad (2.25)$$

and continue an iteration procedure analogous to the method used to obtain the first eigenvector, but evolving (2.26)

$$R_1 c_2 = \lambda_2 c_2 \qquad (2.26).$$

In vice versa, the second-residual matrix, R_2, is calculated by means of (2.27)

$$R_2 = R_1 - \lambda_2 c_2 c_2' \qquad (2.27).$$

Carry out this further computation to extract the remaining eigenvectors until the residual is essentially zero or the number of principal factors is equal to the number of column or row in the data matrix, whichever is smaller.

2.4.2 Identifying the Number of Real Factors

According to the Section 2.1.2, only n eigenvectors are correspondent to the significant real factors. To deduce the exact size of the factor space, various criteria have been developed. However, no single "magic" criterion is applicable to all types of data. They are limited to matrices which contain a relative uniform error throughout.[15] For the best result, the various criteria ought to be mentioned together. Normally, the methods based on experimental error are preferred when the error is known. Often such information is lacking and the empirical and statistical methods must be employed.

1. Methods Based on Experimetal Error

Several criteria have been introduced here for determining the size of the true factor space when accurate estimations of the experimental error are known.

Residual Standard Deviation

The residual standard deviation (RSD) or real error (RE) is defined as

$$\text{RSD} = \left(\frac{\sum\limits_{j=n+1}^{c} \lambda_j^0}{r(c-n)} \right)^{1/2} \tag{2.28}$$

where λ_j^0 is the eigenvalue associated with the residual error. The criterion of judgement is met when the RSD approximately equals to the estimated error.

Root-Mean-Square Error

The root-mean-square (RMS) error is defined by the equation

$$\text{RMS} = \left(\frac{\sum_{j=n+1}^{c} \lambda_j^0}{rc} \right)^{1/2} \tag{2.29a}$$

$$= \left(\frac{c-n}{c} \right)^{1/2} (\text{RSD}) \tag{2.29b}.$$

Although RMS and RSD are closely related, they measure two entirely different errors. The RMS measures the difference between raw data and factor analysis-regenerated data. The RSD measures the difference between raw data and pure data possesing no experimental error.

Average Error

The average error, $\bar{e}$, is simply the average of the absolute values of the differences between the original and regenerated data, and directly proportional to the root-mean-square error

$$\bar{e} = \left(\frac{2}{\pi} \right)^{1/2} (\text{RMS}) \tag{2.30}.$$

Chi-Squared

Bartlett [12] defined the (calculated) chi-squared (χ^2) as

$$\chi_n^2 (\text{calculated}) = \sum_{i=1}^{r} \sum_{k=1}^{c} \frac{\left(d_{ik} - d_{ik}^{\neq} \right)^2}{\sigma_{ik}^2} \tag{2.31}$$

where σ_{ik} is the standard deviation associated with the measurable d_{ik}, $d_{ik}^{\neq}$ is the value of the corresponding point regenerated from factor analysis using the n largest eigenvalues, and the sum is taken over all experimental points. For each set of eigenvectors, χ_n^2 is compared to its expectation value given by the product

$$\chi_n^2 \text{(expected)} = (r-n)(c-n) \tag{2.32}.$$

The criterion of judgement is met when the calculated χ^2 is less than its corresponding expectation value.

Standard Error in the Eigenvalue

Hugus and El-Awady [13] showed that the standard error in an eigenvalue is related to the standard deviations of the data points, and defined by the equation

$$\sigma_m = \left(\sum_{j=1}^{c} \sum_{k=1}^{c} c_{mj}^2 c_{mk}^2 \sigma(Z)_{jk}^2 \right)^{1/2} \tag{2.33}$$

where σ_m is the standard error in the mth eigenvalue, c_{mj} and c_{mk} are the jth and kth components of the mth eigenvector, and

$$\sigma(Z)_{jk}^2 = \begin{cases} \sum_{i=1}^{r} \left(d_{ij}^2 \sigma_{jk}^2 + d_{ik}^2 \sigma_{ij}^2 \right) & \text{for } j \neq k \\ \sum_{i=1}^{r} 4 d_{ij}^2 \sigma_{ij}^2 & \text{for } j = k \end{cases}$$

where σ_{ij} is the error in d_{ij}.

This criterion are allowed when the σ_m is less than its eigenvalue.

2. Empirical Methods

Since the information of error may be either not available or is highly suspected, the empirical methods have to be developed to solve this challenging problem.

Imbedded Error Function

The imbedded error (IE) function is defined by the equation:

$$IE = \left(\frac{n \sum_{j=n+1}^{c} \lambda_j^0}{rc(c-n)} \right)^{1/2} \tag{2.34a}$$

$$= RSD \sqrt{\frac{n}{c}} \tag{2.34b}$$

Because the information of secondary eigenvalues, the number of rows and columns in the data matrix, and the number of factors is always available to us when perform factor analysis. So, we can calculate IE as a function of n, as n goes from 1 to c. The IE function should decrease as more and more primary eigenvector sets are used, and increase when the secondary eigenvectors in the reproduction are included. If the errors are distributed uniformly, the dimensionality of the true factor space should be evaluated.

Factor Indicator Function

Malinowski [2;14] discovered an empirical function, called factor indicatior function, which appears to be much more sensitive than the IE function in its ability to pick out the proper number of factors and seems to be the best choice in general cases.[14;21] The factor indicator function (IND) is defined as

$$IND = \frac{RSD}{(c-n)^2} \tag{2.35}$$

The IND function, similar to the IE function, reaches a minimum when the correct number of factors are employed.

Cumulative Percent Variance

The cumulative percent variance is a measure of the percentage of the total variance in the data which is accounted for by abstract reproduction. It is defined as follow;

$$\text{Cumulative percent variance} \quad = \quad 100 \left[\sum_{i=1}^{r} \sum_{k=1}^{c} \left(\frac{d_{ik}^{\neq}}{d_{ik}} \right)^2 \right] \qquad (2.36\text{a})$$

$$= \quad 100 \left(\frac{\sum_{j=1}^{n} \lambda_j^{\neq}}{\sum_{j=1}^{c} \lambda_j} \right) \qquad (2.36\text{b}).$$

Here $d_{ik}^{\neq}$ is the value of a data point reproduced by AFA, and d_{ik} is the raw, experimental data point.

The percent variance criterion accepts the set of largest eigenvalues required to account for the variance within a chosen specification.

3. Statistical Methods

Reduced Eigenvalue and Statistical F-Test

Malinowski [2;15] defined the reduced eigenvalue (REV) as

$$\text{REV}_j \quad = \quad \frac{\lambda_j}{(r - j + 1)(c - j + 1)} \qquad (2.37).$$

Because the reduced error eigenvalues are equally proportional to the standard deviation, a statistical F-test was invoked:

$$F(1, s - n) \quad = \quad \frac{\sum_{j=n+1}^{s} (r - j + 1)(c - j + 1)}{(r - n + 1)(c - n + 1)} \frac{\lambda_n}{\sum_{j=n+1}^{s} \lambda_j^0} \qquad (2.38)$$

Here s is equal to r or c, whichever is smaller. This is designed to test the null hypothesis

$$\text{H}_0: \quad \text{REV}_n \quad = \quad \text{REV}_{pool}^0$$

against the alternative hypothesis (a one-tail-test)

$$\text{H}_a: \quad \text{REV}_n \quad > \quad \text{REV}_{pool}^0$$

where REV_{pool}^0 is the weighted average of the pool of reduced error eigenvalues.

The percent significance level (%SL) for F-distribution are defined as the probability that F would be as large as it is if the first sample's underlying distribution actually has smaller variance than the second's denoted by $Q(F/v_1, v_2)$ where v_1 and v_2 are the number of degree of freedom in the first and second samples, respectively.[9]

$$
\begin{aligned}
\%\text{SL} &= 100 \times Q(F/v_1, v_2) \\
&= 100 \frac{\Gamma(v_1 + v_2)}{\Gamma(v_1)\Gamma(v_2)} \int_0^x F^{(v_1-1)}(1-F)^{(v_2-1)} dF
\end{aligned}
\tag{2.39}
$$

where
$$
x = \frac{v_2}{v_2 + v_1 F/2}
$$

and $\Gamma(v) \equiv$ Gamma function of argument $v = \int_0^\infty F^{(v-1)} e^{-F} dF$.

The number of significant factors is determined when the significant level for n eigenvalues is less than some desired value (such as 5% or 10%) *i.e.* the null hypothesis is rejected and the alternative is accepted.[16]

By carrying out the above methods, we have already obtained the correct factor size, and then the *reduced* column matrix, $\overline{C}$, is constructed by the complete set of n primary eigenvectors as

$$
\overline{C} = \begin{bmatrix} \mathbf{c}_1' \\ \mathbf{c}_2' \\ \vdots \\ \mathbf{c}_n' \end{bmatrix} = [\mathbf{c}_1 \, \mathbf{c}_2 \ldots \mathbf{c}_n]'
\tag{2.40}
$$

The *reduced* row matrix, $\overline{R}$, is yielded by premultiplication of the inverse of the $\overline{C}$ by data matrix as

$$
\overline{R} = \mathbf{D}\overline{C}^{-1} = \mathbf{D}\overline{C}'
\tag{2.41}
$$

Then multiply $\overline{\mathbf{C}}$ by $\overline{\mathbf{R}}$ to generate the *reproduced* data matrix, $\overline{\mathbf{D}}$, as

$$\overline{\mathbf{D}} = \overline{\mathbf{R}}\,\overline{\mathbf{C}} \qquad (2.42).$$

At the end of this stage, we have already compressed the factor model by deleting the error factors, and obtained the more accurate reproduced data. However, from the standpoint of a theoretical chemist, the analysis should not terminate here. The main objective is to gain insight into the nature of the factors, and the abstract solution must be transformed to a more meaningful solution as described in the next section.

2.4.3 Transformation by Modeling Methods

Kankare [22] was the first investigator to use models with adjustable parameters in factor analytical studies. Models, based on well-established scientific knowledge and theory, are formulated to express the evolutionary profile, c_j, of component j as a function of instrumental and phenomenological parameters, g_{ijk} and p_{ijk}, respectively;

$$c_j = f\!\left(g_{ijk}, p_{ijk}\right) \qquad (2.43).$$

In the case of spectrometric titration of Cu(II)-glycine, and Cu(II)-alanine complexes, the g_{ijk} and p_{ijk} parameters are correspondent to the pH of each solution and equilibrium constants respectively. To overcome this transformation problem, a unique function of $f\!\left(g_{ijk}, p_{ijk}\right)$ is created by utilizing of the information of chemical equilibrium (as described in the Section 2.2) and the principle of mass balance (as described in the Section 2.3).

As a starting point, values are assigned to each of these parameters and assembled into a profile matrix, $\mathbf{C}_{\text{model}}$. The loading matrix, $\mathbf{L}$, is then computed by the pseudoinverse;

$$\mathbf{L} = \mathbf{D}\mathbf{C}^{+}_{\text{model}} \qquad (2.44).$$

These matrices are multiplied to regenerate the *estimated* data,

$$\hat{\mathbf{D}} = \mathbf{LC}_{model} \tag{2.45}.$$

The fit is evaluated by examining χ^2, the sum of squares of the differences between the raw data and the estimated data

$$\chi^2 = \sum_{i=1}^{r}\sum_{k=1}^{c}\left(d_{ik} - \hat{d}_{ik}\right)^2 \tag{2.46}.$$

The parameters are varied by the optimization methods until χ^2 reaches minimum.

2.4.4 Optimization Methods

As stated in the previous section, the modeling transformation depends on the optimization methods. On the realm of numerical methods, the optimization methods can be classified into *derivative* and *non-derivative*. Simplex method [9;23;26], a derivative-free strategy, is the most commonly used in chemical problems and may be improved by merging with the Gauss-Newton method, Fibonacci Unidirectional search and Stochastic Initialization algorithm.

Simplex Method

The original non-adaptive simplex method, proposed by Spendley *et al.* [24] is rarely used today. The first useful modification, by Nelder and Mead [25], led to a simple and widely applicable algorithm.

The simplex is a (flexible) polyhedron, having (m+1) vertices constructed in m-dimensional parameter space. For m = 2, the simplex is a triangle and for m = 3, a tetrahedron.

The process of minimization by a simplex method involves three steps:

(1) Construction of initial simplex body

(2) Iterative search for minimum

(3) Identification of a search termination

In the minimization procedure at each cycle, step (2) and (3) are repeated. Step (1) affects the speed of convergence.
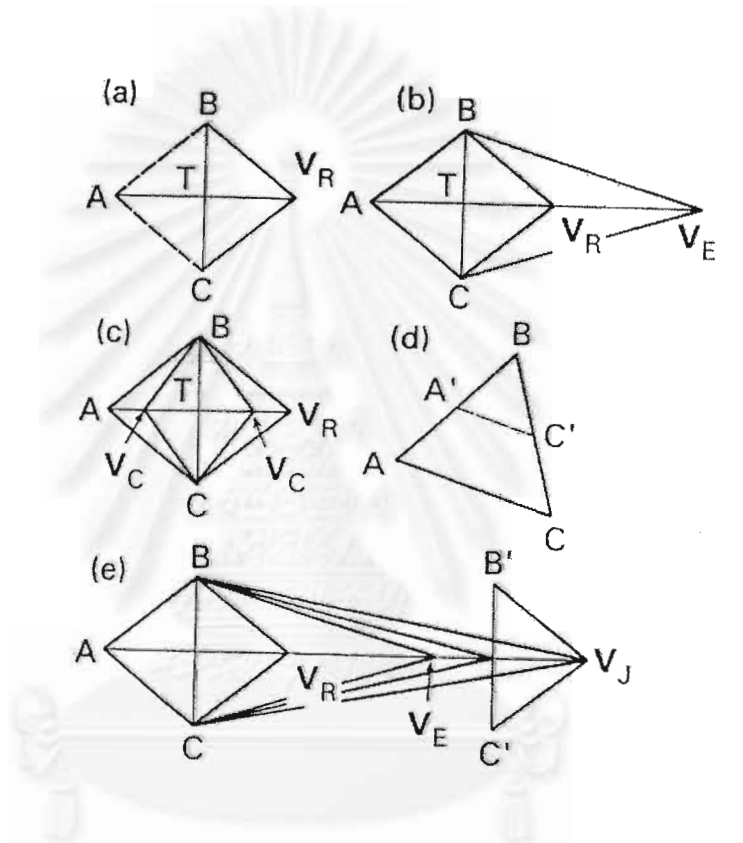
Figure 2.2 The simplex operations: (a) reflection, (b) expansion, (c) contraction,

(d) reduction, and (e) transfer.

Construction of the initial simplex body

The co-ordinates of the vertices of a simplex create rows of the matrix V of dimension $[(m+1) \times m]$. When an initial guess $\beta^{(0)}$ is proposed, the initial simplex is constructed such that the first row of the matrix V contains this initial guess as the co-ordinates of the component. The jth row $(j=2,\dots,m+1)$, is given by

$$V_{ji} = \begin{cases} \beta_i^{(0)} + \dfrac{0.5\beta_j^{(0)}}{m\sqrt{2}}\left(m - 1 + \sqrt{m+1}\right) & \text{for } i \neq j \\[4mm] \beta_i^{(0)} + \dfrac{0.5\beta_j^{(0)}}{m\sqrt{2}}\left(\sqrt{m+1} - 1\right) & \text{for } i = j \end{cases} \qquad (2.47).$$

Iterative search for minimum

The procedure calculates the functional value $U(\beta_j)$ for all vertices of the simplex and then discriminates the worst point (V_U) corresponding to the maximum (U_U) and the best point (V_L) corresponding to the minimum (U_L).

First, a *reflection* of the vertex V_U is made through the centre of gravity (P) of the other vertices as

$$V_R = P + \alpha(P - V_U) \qquad (2.48)$$

where

$$P = \frac{\displaystyle\sum_{\substack{i=1 \\ i \neq U}}^{m+1} V_i}{m} \qquad (2.49a)$$

or

$$P_w \equiv \text{weighted centre of gravity}$$

$$= \frac{\displaystyle\sum_{i=1}^{m+1}(U_i - U_U)V_i}{\displaystyle\sum_{i=1}^{m+1}(U_i - U_U)} \qquad (2.49b)$$

and α is called *reflection parameter*. When, for at least one vertex $V_i(i \neq L, i \neq U)$, $U_i > U_R > U_L$, the vertex V_U is replaced by the vertex V_R and the kth cycle is finished.

If, however, $U_L > U_R$, an *expansion* is made, to obtain point V_E:

$$V_E = P + \gamma(V_R - P) \qquad (2.50)$$

where γ is called *expansion parameter*. If $U_E \leq U_R$, then V_U or V_R may be replaced by V_E. Sometimes additional expansion to the point V_J is performed, that is

$$V_J = (J+1)V_E - JP \qquad (2.51)$$

where $J = 2, 3, 4, \ldots$, until $U_{J+1} > U_J$. Then $\mathbf{V}_U$ is replaced by $\mathbf{V}_J$ and a *transfer* of another vertex $\mathbf{V}_i$ into a new position $\mathbf{V}_i^N$ is performed in order to keep the original form of the simplex

$$\mathbf{V}_i^N = \mathbf{V}_i + J(\mathbf{V}_E - \mathbf{P}), \; i \neq U \tag{2.52}$$

until the kth cycle is once again completed.

When $U_R > U_i$ for all $i \neq U$ after a reflection, a *contraction* is performed by using the point $\mathbf{V}_C$ given by

$$\mathbf{V}_C = \begin{cases} \mathbf{P} + \beta(\mathbf{V}_U - \mathbf{P}) & \text{for } U_U < U_R \\ \mathbf{P} + \beta(\mathbf{V}_R - \mathbf{P}) & \text{for } U_U > U_R \end{cases} \tag{2.53}$$

where β is called *contraction parameter*. If $U_C < U_U$, $\mathbf{V}_U$ is replaced by $\mathbf{V}_C$, completing the kth iteration.

If, despite the contraction, $U_C \geq U_U$, the simplex is reduced around $\mathbf{V}_L$ with the smallest criterion value. *Reduction* involves replacement of vertices $\mathbf{V}_i (i \neq L)$ with the new vertices $\mathbf{V}_i^Z$ such that

$$\mathbf{V}_i^Z = \mathbf{V}_L + \lambda(\mathbf{V}_i - \mathbf{V}_L) \tag{2.54}$$

where λ is called *reduction parameter*. This reduction procedure completes the kth cycle.

Identification of termination criteria

Nelder and Mead [25] recommend that at the end of each cycle, an examination of the magnitude of the decrease of the criterion function and of the relative changes of the simplex vertices be made by use of

$$|U_U - U_L| < \varepsilon_1 \tag{2.55a}$$

and

$$\frac{1}{m+1} \sum_{i=1}^{m+1} d\left[\mathbf{V}_i^{(k)}, \mathbf{V}_i^{(k-1)}\right]^2 < \varepsilon_2 \tag{2.55b}.$$

The constant ε_1 and ε_2 should have the value less than 10^{-4} and 10^{-8}, respectively. The term $d\left[\mathbf{V}_i^{(k)}, \mathbf{V}_i^{(k-1)}\right]$ denoted the distance between vertices $\mathbf{V}_i^{(k)}$ and $\mathbf{V}_i^{(k-1)}$ of the cycle k and $(k\text{-}1)$, respectively.

Gauss-Newton Method

Spendley [24] proposed a simple procedure which combines the simplex with the Gauss-Newton method, a derivative optimization. The procedure starts from a linearized objective function $f(x_i, \beta)$ for $\beta = \mathbf{V}_j$ or $\beta = \mathbf{V}_L$. It can be shown that

$$e_i\left(\mathbf{V}_j\right) - e_i\left(\mathbf{V}_L\right) \quad \approx \quad \mathbf{J}_i^T\left(\mathbf{V}_j - \mathbf{V}_L\right) \tag{2.56}$$

where $\mathbf{J}_i^T$ is the ith row of the Jacobian $\mathbf{J}$ and the symbol $e_i\left(\mathbf{V}_j\right) = y_i - f(x_i, \mathbf{V}_j)$ denotes the ith residual for the estimate $\mathbf{V}_j$. Similarly $e_i\left(\mathbf{V}_L\right)$ denotes the ith residual for the estimate, $\mathbf{V}_L$. In matrix notation, (2.56) can be written as

$$\mathbf{T} \quad \approx \quad \mathbf{JA} \tag{2.57}$$

where $\mathbf{T}$ is the (n×m) matrix with elements

$$\mathbf{T}_{ij} \quad = \quad e_i\left(\mathbf{V}_j\right) - e_i\left(\mathbf{V}_L\right), \ i = 1, \ldots, n, \ j = 1, \ldots, m \tag{2.58}$$

and $\mathbf{A}$ is the (m×m) matrix with elements

$$A_{jk} \quad = \quad V_{jk} - V_{Lk}, \ j = 1, \ldots, m, (j \neq L) \ k = 1, \ldots, m \tag{2.59}$$

If the simplex vertices in the kth iteration are known, the matrix $\mathbf{J}$ may be estimated from (2.57). Let us assume that the criterion for the least-square method is valid. The increment vector, $\mathbf{L}$, of the Gauss-Newton [23] for this criterion may be calculated from the approximate expression

$$\mathbf{L} \quad = \quad \mathbf{A} - \mathbf{D}^{-1}\mathbf{w} \tag{2.60}$$

where elements of the matrix $\mathbf{D}$ are given by

$$D_{jk} \quad = \quad \sum_{i=1}^{n}\left[e_i\left(\mathbf{V}_j\right) - e_i\left(\mathbf{V}_L\right)\right]\left[e_i\left(\mathbf{V}_k\right) - e_i\left(\mathbf{V}_L\right)\right], \ j, k = 1, \ldots, m \ (\neq L) \tag{2.61}$$

and those of vector $\mathbf{w}$ by

$$w_j = \sum_{i=1}^{n} \left[e_i(V_j) - e_i(V_L) \right] e_i(V_L), \quad j \neq L \qquad (2.62).$$

In the kth iteration of a given simplex optimization, the procedure determines L, from (2.60) and calculates the criterion value, $U(V_L + L)$. This value then determines whether the procedure continues according to the original simplex method or replaces vertex V_U of the maximum value $U(V_U)$ by a vertex $(V_L + L)$, and then uses an approximate Gauss-Newton method.

Fibonacci Unidirectional Search

Marsili-Libelli and Castelli [27] improved convergent speed of the simplex by combining with Fibonacci unidirectional search. Along this context, the distinction between the two operations of reflection and expansion are unified into a single outward search procedure, namely a unidirectional optimization, based on the Fibonacci interval elimination method.

According to the modified simplex algorithm, the reflection parameter α is retained, it is now only an initial estimate. In addition an incremental reflection parameter, δ, is defined, and the interval selection for the Fibonacci search is determined iteratively as follows.

First, the primary and secondary reflection are defined as

primary: $X_1(0) = P + \alpha(P - V_U)$ (2.48),(2.63)

secondary: $X_2(0) = P + (\alpha - \delta)(P - V_U)$ (2.64)

The incremental reflection parameter is selected in order to determine the search descent direction; therefore it should be such that $U(X_2(0)) < U(X_1(0))$. If this condition is not satisfied, in other words if the local minimum is already bracketed in the initial search interval,

then the incremental parameter δ is halved and equation (2.63) and (2.64) are recomputed. This procedure is repeated until

$$U(\mathbf{X}_2(0)) \quad < \quad U(\mathbf{X}_1(0)).$$

Then the initial search step is computed as

$$\mathbf{d} \quad = \quad \mathbf{X}_2(0) - \mathbf{X}_1(0) \tag{2.65}$$

and the first point of the forward interval determination is set as

$$\mathbf{X}_3(0) \quad = \quad \mathbf{X}_2(0) + \mathbf{d}Fib(1) \tag{2.66}$$

where $Fib(1)$ is the first Fibonacci number.[28]

The initial search interval is then enlarged by computing the sucessive triplets

$$\mathbf{X}_1(g) \quad = \quad \mathbf{X}_2(g-1) \tag{2.67a}$$

$$\mathbf{X}_2(g) \quad = \quad \mathbf{X}_3(g-1) \tag{2.67b}$$

$$\mathbf{X}_3(g) \quad = \quad \mathbf{X}_2(g) + [\mathbf{X}_2(g) - \mathbf{X}_1(g)]\frac{Fib(g)}{Fib(g-1)} \tag{2.67c}$$

where g is the forward iteration counter. New triplets are determined through (2.67) until

$$U(\mathbf{X}_3(g)) \quad > \quad U(\mathbf{X}_2(g)).$$

Then the forward expression is terminated and the interval $[\mathbf{X}_1(g), \mathbf{X}_3(g)]$ is taken as the starting Fibonacci interval, as it does contain the required minimum in the search direction.

Within the initial search interval $[\mathbf{X}_1(g), \mathbf{X}_3(g)]$ a new point is located at

$$\mathbf{X}_p \quad = \quad \mathbf{X}_1(g) + \frac{Fib(g)}{Fib(g-1)}[\mathbf{X}_2(g) - \mathbf{X}_1(g)] \tag{2.68}.$$

If $g = 2$, a comparison is made between $U(\mathbf{X}_2(g))$ and $U(\mathbf{X}_p)$, and the interval containing the minimum is retained. If $g > 2$, the interval elimination procedure requires as many iterations as the predetermined number g of the Fibonacci sequence.

Defining the ratio

$$R(i) = \frac{Fib(g-2)}{Fib(g)} \tag{2.69},$$

the new boundary $[\mathbf{P},\mathbf{Q}]$, with $U(\mathbf{P}) < U(\mathbf{Q})$, are located through the following backward iteration. Starting with $i = g$ and an initial search interval with boundaries

$$\mathbf{P} = \mathbf{X}_1(g) \text{ and } \mathbf{Q} = \mathbf{X}_3(g)$$

at each step determine

$$\mathbf{X}_q(i) = \mathbf{P} + R(i)(\mathbf{Q} - \mathbf{P}) \tag{2.70}$$

$$\mathbf{X}_p(i) = \mathbf{Q} - R(i)(\mathbf{Q} - \mathbf{P}) \tag{2.71}.$$

If $U(\mathbf{X}_q(i)) \le U(\mathbf{X}_p(i))$ then

$$\mathbf{Q} = \mathbf{X}_p(i) \tag{2.72a}$$

$$\mathbf{X}_p(i-1) = \mathbf{X}_q(i) \tag{2.72b}$$

$$\mathbf{X}_q(i-1) = \mathbf{P} + R(i)(\mathbf{Q} - \mathbf{P}) \tag{2.72c}.$$

If $U(\mathbf{X}_q(i)) > U(\mathbf{X}_p(i))$ then

$$\mathbf{P} = \mathbf{X}_q(i) \tag{2.73a}$$

$$\mathbf{X}_q(i-1) = \mathbf{X}_p(i) \tag{2.73b}$$

$$\mathbf{X}_p(i-1) = \mathbf{Q} - R(i)(\mathbf{Q} - \mathbf{P}) \tag{2.73c}.$$

This procedure is repeat until $i = 2$.

In some cases [27], the combination of Fibonacci unidirectional search and simplex method shows less "wobbling" in the search direction and improves the speed of convergence.

Stoichastic Initialization Algorithm

A weak point common to all search algorithms is their inherent inability to avoid the local minima. In fact it is well-known that search algorithms tend to converge to the optimum in whose domain of attraction the starting point is located, regardless of whether it is a local or

a global optimum. Hence careful selection of the starting point helps to circumvent the problem, but this implies a lengthy trial and error procedure without general guarantees. Marsili-Libelli and Castelli [27] proposed a stochastic initialization algorithm for selecting the starting point in such a way that convergence to the global minimum is assured.

Let $U(\mathbf{X})$, with $\mathbf{X} \in \mathfrak{R}^m$, be the function whose minimum is to be found within the search interval $[\mathbf{A}, \mathbf{B}]$. The algorithm considers a stochastic search based on a random variable w with uniform distribution in $[0,1]$. As the search goes on, more points in $[\mathbf{A}, \mathbf{B}]$ are selected according to the equation

$$\mathbf{X}(i+1) \quad = \quad \mathbf{X}(i) - \mathbf{W}(i)(2w-1)^k \tag{2.74}$$

for each component ($j = 1,2,\ldots,m$), where i is the iteration counter, k is a desired integer, and $\mathbf{W}(i)$ is the search interval at the ith iteration, defined as

$$\mathbf{W}(i) \quad = \quad max[\mathbf{X}(i) - \mathbf{A}, \mathbf{B} - \mathbf{X}(i)] \tag{2.75}.$$

After checking that the new point $\mathbf{X}(i+1)$ is inside the search interval $\mathbf{W}(i)$ the corresponding function is evaluated, and if it represents an improvement, it is retained in place of $\mathbf{X}(i)$.

The iterations are terminated by use of

(1) a maximum number of iterations, i_{max}.

(2) an uncertainty reduction criterion, *i.e.* an algorithm for increasing k.

The merits and liabilities of this method are those typical of stochastic search algorithms. They are little sensitive to function irregularities and can tackle multimodal functions.[27] On the other hand, convergence is assured only in probabilistic terms and depends on the law for increasing k.

CHAPTER 3

PROGRAM IMPLEMENTATION

The factor analysis program, "MTFA" version 1.0, was developed. The source code was written in FORTRAN using Microsoft FORTRAN V5.1 (Microsoft Corp. version of FORTRAN-77) [29-30] for personal computer (PC). Here, main application is to resolve the UV/VIS absorption spectra for the acid-base equilibria of the polyprotic organic acid-base pairs and the formation of metal-ligand complexes. However, the program can also be adapted to resolve other problems.

In particular, the program should be interactive, provided both manual and automatic selections, and generalized enough for further development.

MTFA version 1.0 contains three files:

1) MTFA1.FOR, consisting of main program and all of necessary functions and subroutines that involved the operations of factor analysis

2) Matrix.FOR, consisting of all subroutines that involved the general operations of matrix such as addition, subtraction, multiplication, transpose, inverse, pseudoinverse, etc [37]

3) Print.FOR, consisting of only two subroutines for all output printings

The program manual and examples were given in Appendix.

3.1 Program Strategies

For easy understanding, the program was separated into nine parts:

1. Input Elementary Data.

2. Constructing Covariance Matrix.

3. Decomposition of the Covariance Matrix.

4. Error Treatment.

5. Abstract Column and Row Matrices.

6. Input Additional Data.

7. Stochastic Initialization Algorithm and Optimization Methods.

8. Data Manipulation and Predictions.

9. Print Results.

Nonetheless, the programmer intended to separate part 2, 3 and 5 from each other, to enable the implementation of NIPALS and SVD algorithms in the future version. The flowchart of MTFA program were shown in Figure 3.1.
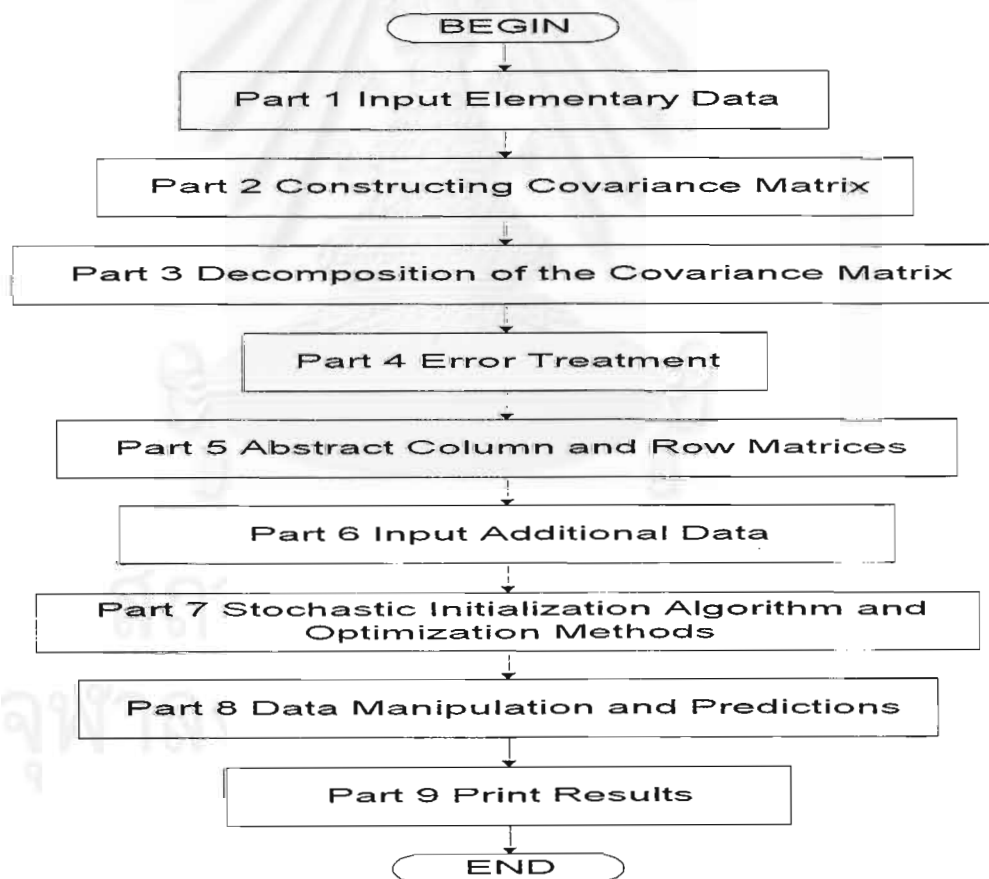
Figure 3.1 The sequence of MTFA program version 1.0

3.1.1 Input Elementary Data

This part was designed for user to input elementary data and select appropriate methodologies, which are:

- Data filename, *i.e.*, spectra filename

- Dimension of data matrix, *i.e.*, number of rows and columns

- Type of chemical problems:

  (1) Acid-Base Equilibrium,

  or (2) Metal-Ligand Complex Equilibrium

- Type of optimization methods:

  (1) Simplex Method,

  (2) Simplex Method with Approximate Gradient,

  (3) Combination of Simplex and Approximate Gauss-Newton Methods,

  or (4) Modified Simplex Method with Fibonacci Unidirectional Search

- pH of each solution

- Total concentration in mol/L of acid or ligand and metal

- Dissociation of ligand

- Gradient for the procedure of principal factor analysis

- Tolerance and Convergence in the procedure of optimizations

- Simplex parameters

The sequence of input was shown in Figure 3.2.
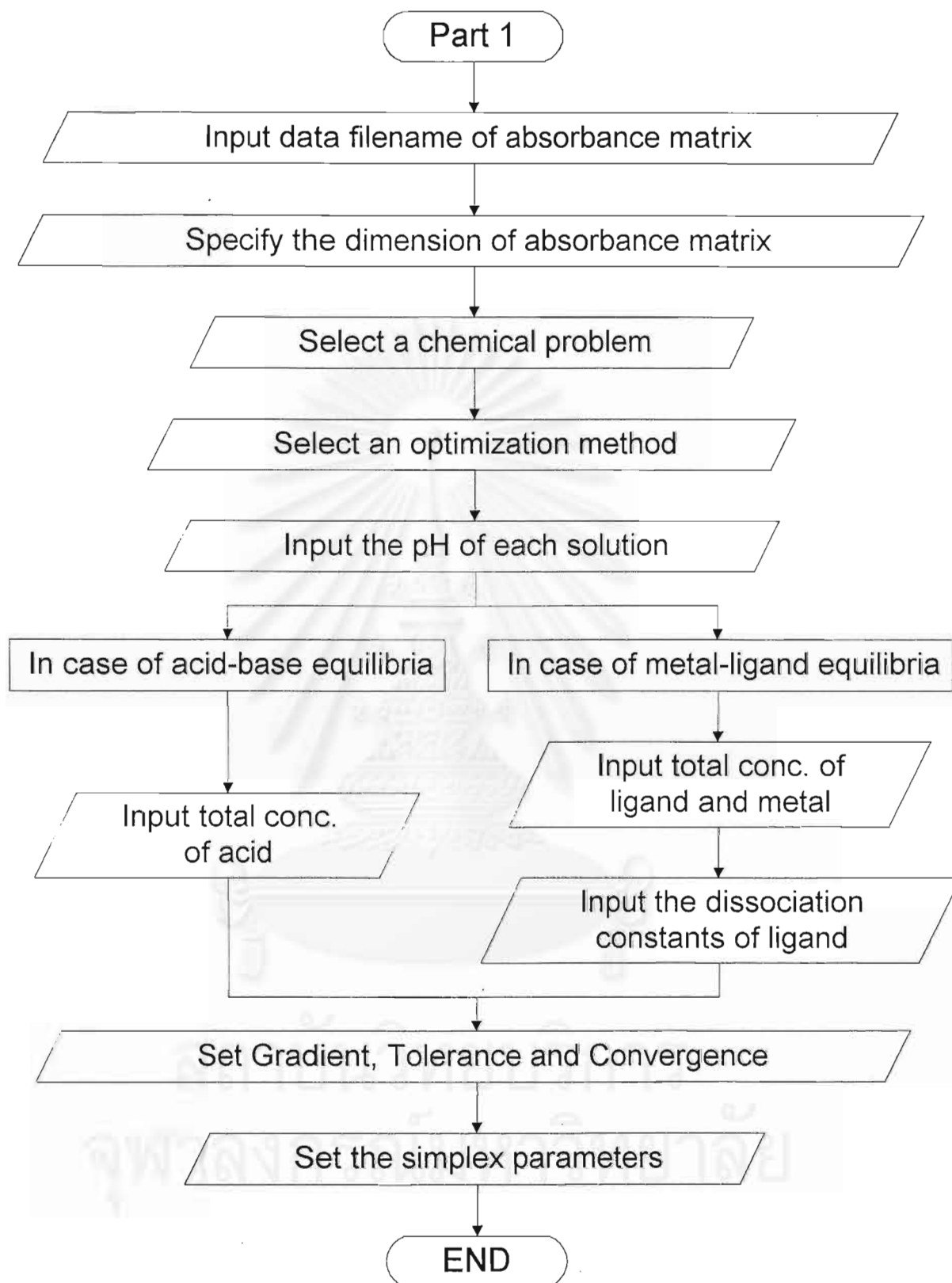
Figure 3.2 The flowchart of Part 1; Input Elementary Data

3.1.2 Constructing Covariance Matrix

This part was designed to compute the covariance matrix $\mathbf{Z}$, and may be extended to calculate the correlation matrix $\mathbf{Z}_N$. The procedure of this part began with reading the absorbance matrix ($\mathbf{D}$) which was specified the filename in Part 1, transposed $\mathbf{D}$ to obtain $\mathbf{D}'$, and premultiplied $\mathbf{D}'$ to $\mathbf{D}$ yield $\mathbf{Z}$ as given in Figure 3.3.
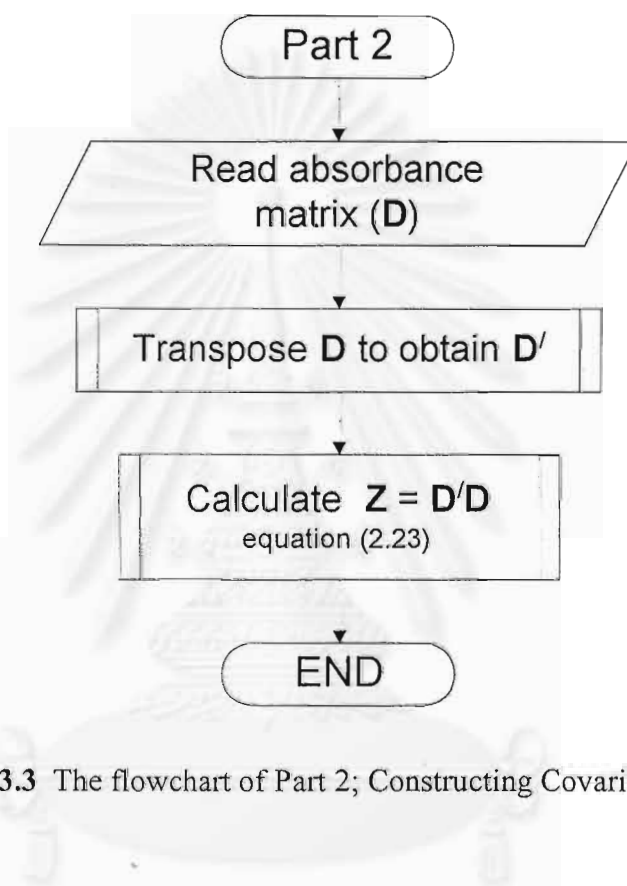
Figure 3.3 The flowchart of Part 2; Constructing Covariance Matrix

3.1.3 Decomposition of the Covariance Matrix

This part was designed for user to select the automatic or manual setting for the initial guess eigenvectors. Then decompose the covariance matrix to obtain the complete set of abstract eigenvectors ($\mathbf{C}_{col}$ matrix) and eigenvalues (λ). The procedure of this part is to perform the principal factor analysis (PFA) which described in Section 2.4.1 and illustrated in Figure 3.4.
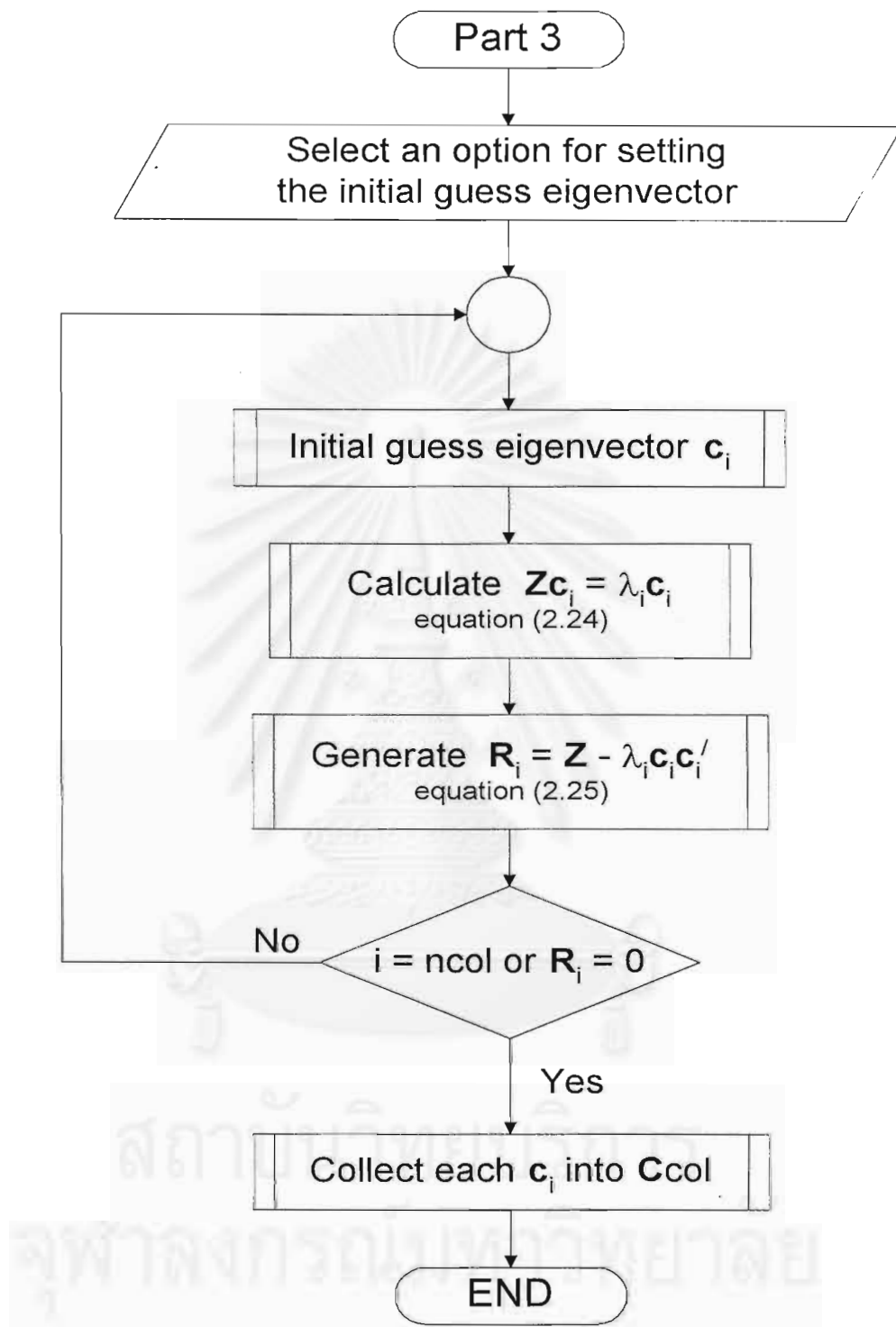
Figure 3.4 The flowchart of Part 3; Decomposition of the Covariance Matrix

3.1.4 Error Treatment

This part was designed to calculate all error indicators. These are RSD, RMS, IE, IND, cumulative percent variance and %SL. At the end of this part the program will suggest automatically the plausible number of primary eigenvectors. If the user does not agree with the suggestion, he/she could input the number of primary eigenvectors manually as wishes. The algorithm for this is shown in Figure 3.5.
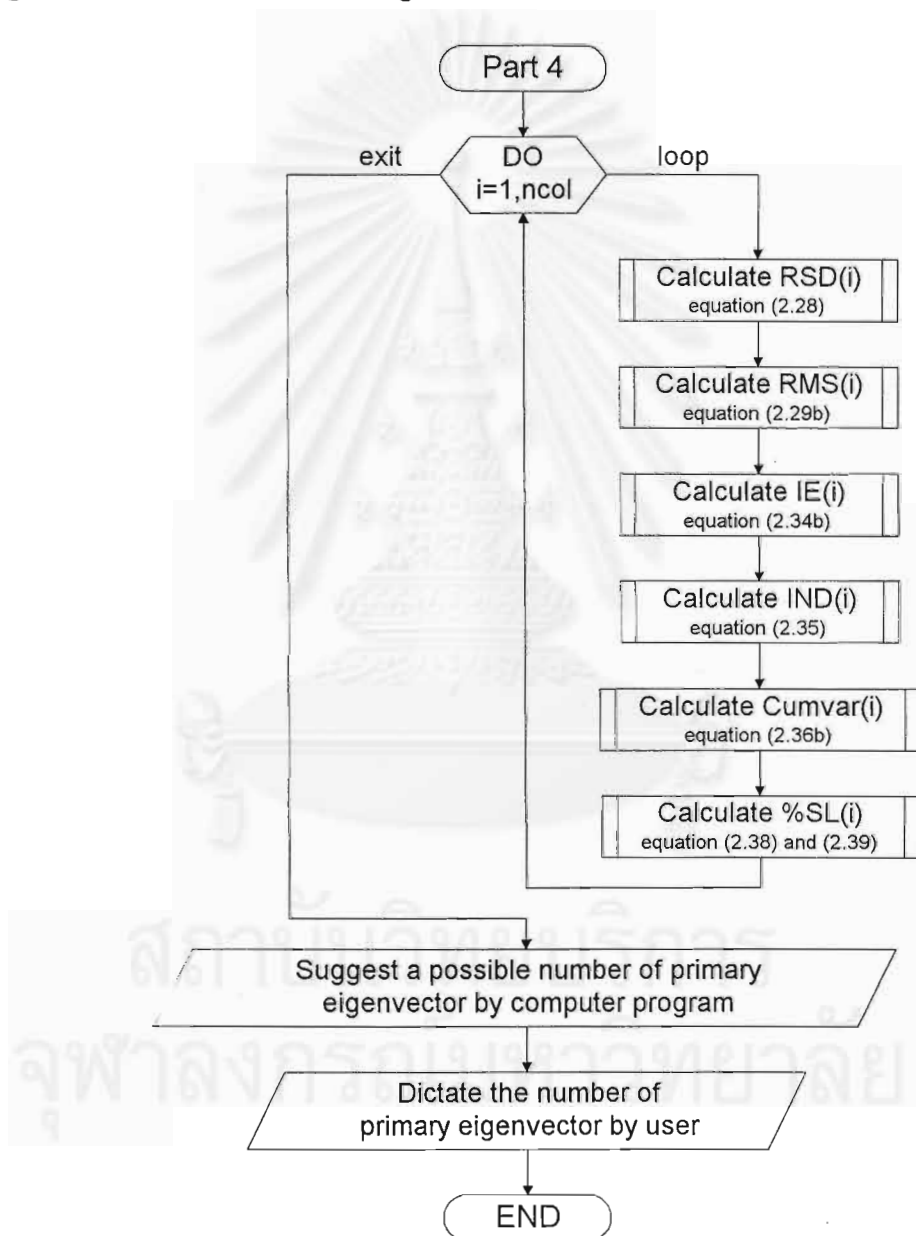
Figure 3.5 The flowchart of Part 4; Error Treatment

3.1.5 Abstract Column and Row Matrices

This part was designed to determine the reduced column (C_{red}), reduced row (R_{red}) and reproduced data (D_{rep}) matrices from the matrix of abstract eigenvectors obtained in Section 3.1.3. The procedure was listed in Figure 3.6.
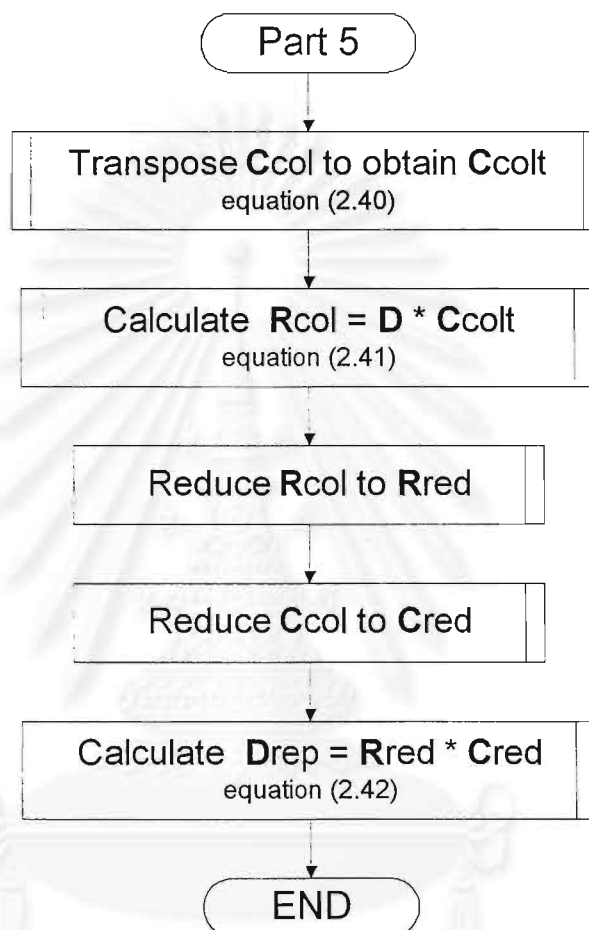
Figure 3.6 The flowchart of Part 5; Abstract Column and Row Matrices

3.1.6 Input Additional Data

This part was designed for user to confirm the number of acid-dissociation or complex formation steps. In case of complex formation, the user should carefully propose the formation model. If the mistake model was introduced, the following procedures would lead to unreliable results. Then the user should desire to do the stochastic initialization algorithm.
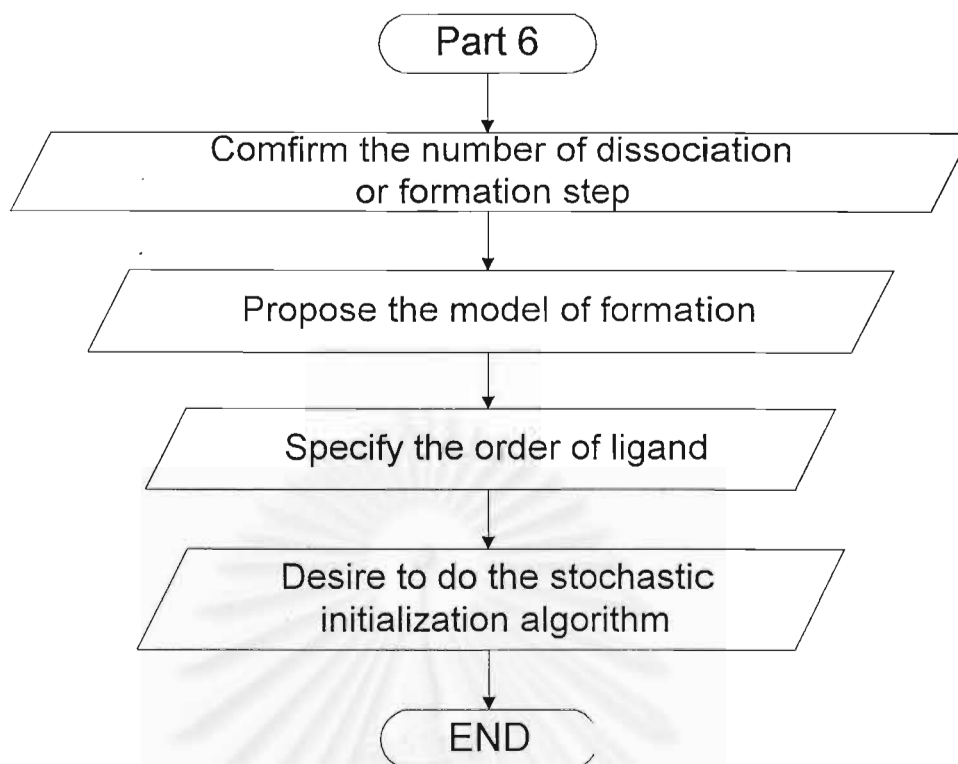
Figure 3.7 The flowchart of Part 6; Input Additional Data

3.1.7 Stochastic Initialization Algorithm and Optimization

This part was designed to perform the stochastic initialization algorithm and the optimization methods such as simplex or the combination of simplex and Gauss-Newton or Fibonacci unidirectional searches. In this study, the simplex method as described in Section 2.4.4 is the default in which its algorithm showed in Figure 3.8
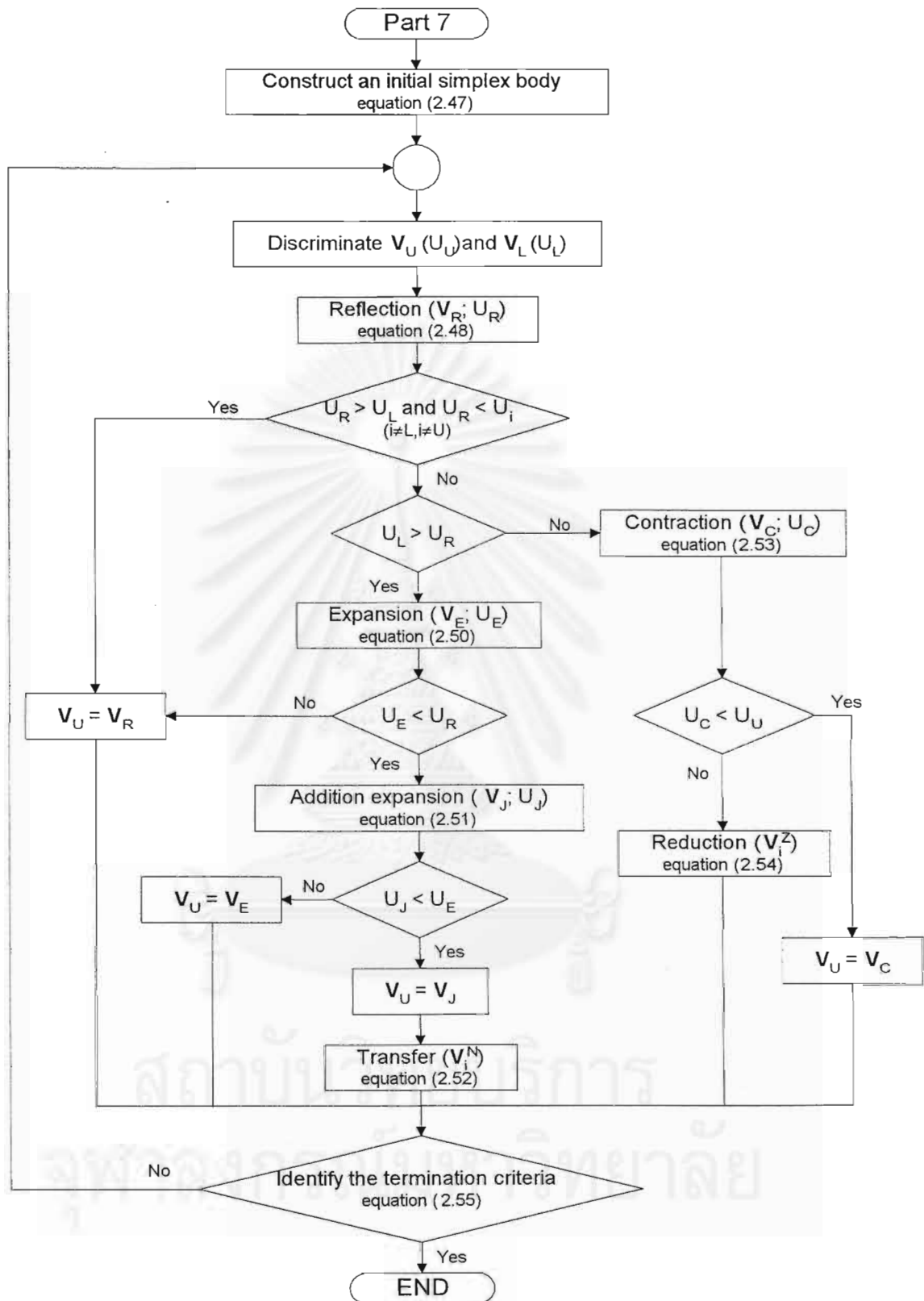
Figure 3.8 The flowchart of Part 7; Simplex Method

3.1.8 Data Manipulation and Predictions

This part was designed to process the output data from previous part and carried out calculations.
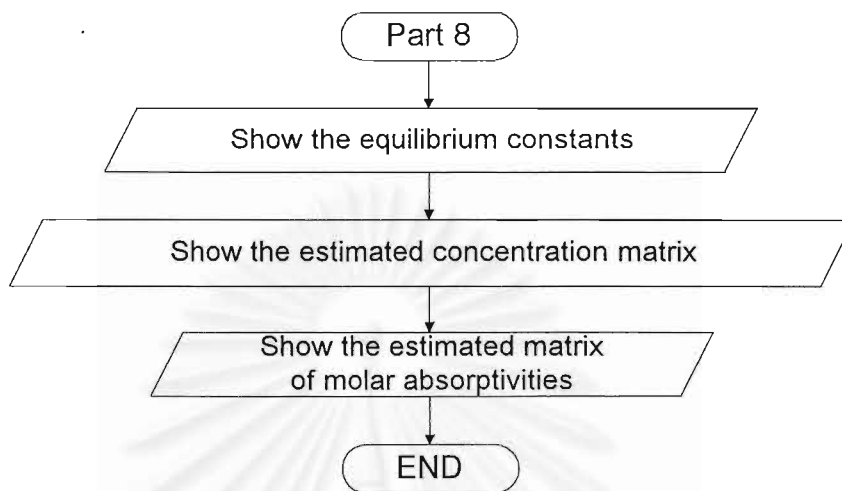
Figure 3.9 The flowchart of Part 8; Data Manipulation and Predictions

3.1.9 Print Results

This part was designed to enable user to select the output unit, *i.e.* printer, computer-saved file (.log), display, or none.
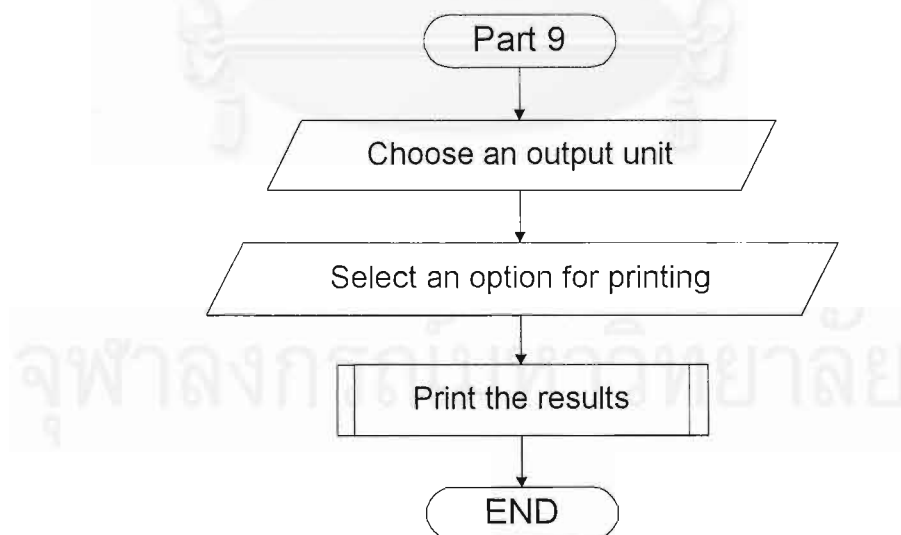
Figure 3.10 The flowchart of Part 9; Print Results

3.2 Summary of Input Requirements

a) Prepare a data file (.txt) of spectra and give 8-letter name.

b) Determine the dimension of data matrix.

c) Select a type of chemical problem.

d) Select an optimization method.

e) Specify the pH of each solution.

f) Input the total concentration (mol/L) of acid or ligand and metal, respectively.

g) Input the number of dissociation step of ligand.

h) In case of complex formation, specify the dissociation constants of ligand.

i) Set the Gradient, Tolerance and Convergence. The default values are recommended.

j) Set the simplex parameters. The default values are recommended.

k) Choose an option for setting the initial guess of eigenvector. The automatic setting is recommended.

l) Dictate the number of primary eigenvectors. The program suggestion is recommended.

m) Give the number of dissociation or formation step of acid or complex, respectively. The program estimation is recommended.

n) In case of complex formation, propose the chemical model.

o) Desire to do the stochastic initialization algorithm.

p) Make an initial guess for equilibrium constants.

q) Select a choice of output unit.

s) Select an option for printing.

CHAPTER 4

VALIDATION OF PROGRAM

The mission of this chapter is to validate the MTFA program using simulated spectral data. The program works very well in solving the simulated spectra with very good accuracy. However, in the instance where the spectra of different components are (nearly) identical, the failure of the program is expected.

4.1 Validation Approach

According to the program, there are two facets to be examined. These are

(I) Validation of program calculations and sensible logic,

and (II) Validation of whole program.

The first facet involves the calculations of matrix operation (such as addition, subtraction, multiplication, transpose, inverse, pseudoinverse and capture of submatrix), the principal factor analysis, the error indicators and the simplex methods. Calculations and logic of all subprograms mentioned above were verified and tested individually and/or in group. Their results were not discussed here.

Only whole-program testing were reported. Two chemical applications *i.e.* acid-base equilibria and metal-ligand complex equilibria were used to validate the performance of the whole program.

The detail of which was given below:

1) **Simulated system of acid-base equilibria**. In this case, an acid, namely H_3A, is dissociated as

$$H_3A \quad \xrightleftharpoons{K_{a1}=1.23 \times 10^{-4}} \quad H^+ + H_2A^-$$

$$H_2A^- \quad \xrightleftharpoons{K_{a2}=2.34 \times 10^{-7}} \quad H^+ + HA^{2-}$$

$$HA^{2-} \quad \xrightleftharpoons{K_{a3}=3.45 \times 10^{-9}} \quad H^+ + A^{3-} \qquad (4.1)$$
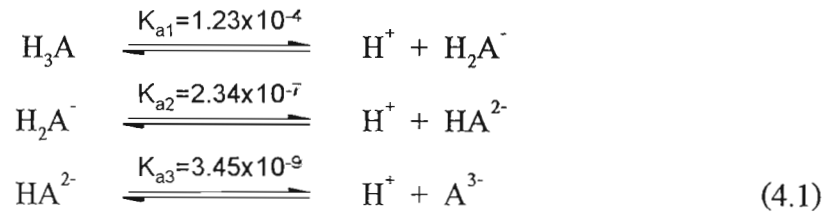
where K_{a1}, K_{a2} and K_{a3} are represented the dissociation (or deprotonation) constants.

Evidently, there are 3 steps of dissociation, and 4 principal components: H_3A, H_2A^-, HA^{2-} and A^{3-}.

 2) **Simulated system of metal-ligand complex equilibria.** In this case, a metal (M^{2+}) and protonated ligand (HL^-) are in the equilibria as

$$H_2L \quad \xrightleftharpoons{K_{a1}=4.467 \times 10^{-4}} \quad HL^- \quad \xrightleftharpoons{K_{a2}=1.667 \times 10^{-10}} \quad L^{2-} \qquad (4.2)$$

$$M^{2+} + HL^- \quad \xrightleftharpoons{\beta_1=7.59} \quad M(HL)^+$$

$$M^{2+} + HL^- \quad \xrightleftharpoons{\beta_2=5.37 \times 10^{-2}} \quad ML + H^+$$

$$M^{2+} + 2HL^- \quad \xrightleftharpoons{\beta_3=3.39 \times 10^{-5}} \quad ML_2^{2-} + 2H^+ \qquad (4.3)$$
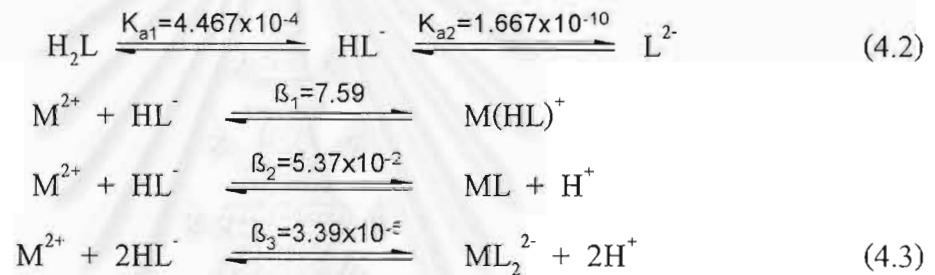
where K_{a1} and K_{a2} are the acid-dissociation constants of ligand, and β_1, β_2 and β_3 are the overall stability constants of complexation, respectively. In the other hand, the chemical equilibria in

(4.3) can be expressed in the matrix form as $\begin{array}{ccc} M & HL^- & H^+ \end{array}$ $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 2 \end{bmatrix}$.

Here, we assign 3 steps of complexation, and 4 principal components: M^{2+}, $M(HL)^-$, ML and ML_2^{2-}.

Along the above approach, if the final output from the program agrees with the former settings (simulated data), the program has been validated.

4.2 Building up the Simulated Spectra

Since we were interested in two chemical problems as stated in Section 4.1, the simulated spectral data were produced separately for each problem. In case of acid-base equilibria, the simulated spectra were generated by imaginating as there are ten solutions of acid (H_3A) with constant-total concentration (A_{tot}) of 0.2000 M and pH varying from 3 to 12. The another case of metal-ligand complex equilibria, the simulated spectra were generated by imaginating as there are twelve solutions of metal (M) and ligand (HL) with constant-total concentration of metal (M_{tot}) 0.0020 M and ligand (L_{tot}) 0.1000 M, and pH varying from 1.5 to 7. Ten solutions of H_3A were recorded from 300 to 700 nm at 2-nm intervals, while the twelve solutions of M-HL were from 450 to 850 nm at 3-nm intervals (totally 201 intervals).

Hence the matrix of simulated spectra for both cases, denoted by $\mathbf{A}$, were created by the multiplication product of the matrix of artificial molar absorptivities, denoted by $\mathbf{E}$, and the matrix of artificial concentration, denoted by $\mathbf{C}$, as

$$\mathbf{A} = \mathbf{EC} \qquad (1.11), (4.4).$$

The dimensions of each matrix are shown in Table 4.1.

Table 4.1 Dimensions of each matrix

Cases of Equilibria	Dimensions		
	Data Matrix $\mathbf{A}$	Matrix of Absorptivities $\mathbf{E}$	Matrix of Concentrations $\mathbf{C}$
Acid protonation	(201×10)	(201×4)	(4×10)
Metal-ligand complex	(201×12)	(201×4)	(4×12)

We remarked that the number "4" were used since we have assigned 4 principal components as stated in Section 4.1.

It is well-known that experimental uncertainty does exist and blends into the pure data. Therefore, the error matrix (S) should be added to the equation (4.4). Thus,

$$A = EC + S \qquad (4.5).$$

The elements S_{ij} are generated using pseudo-random number. Since the standard deviation in absorbance was estimated to vary between 0.0005 and 0.0015 absorbance unit [2], therefore, two levels of error *i.e.* ±0.0005 and ±0.0015 were introduced respectively.

The molar absorptivity matrix, E, were generated by arbitrary drawing and shown in Figure 4.1 and 4.2.

The concentration matrix, C, for the case of acid-base equilibria were formulated by the equation (2.13) or (2.14) where $k = 3$

$$\mathbf{c}_1 = \alpha_1 A_{tot} = \frac{A_{tot}}{1 + \dfrac{K_{a1}}{[H^+]} + \dfrac{K_{a2}K_{a1}}{[H^+]^2} + \dfrac{K_{a3}K_{a2}K_{a1}}{[H^+]^3}}$$

$$\mathbf{c}_2 = \alpha_2 A_{tot} = \frac{A_{tot}}{\dfrac{[H^+]}{K_{a1}} + 1 + \dfrac{K_{a2}}{[H^+]} + \dfrac{K_{a3}K_{a2}}{[H^+]^2}}$$

$$\mathbf{c}_3 = \alpha_3 A_{tot} = \frac{A_{tot}}{\dfrac{[H^+]^2}{K_{a1}K_{a2}} + \dfrac{[H^+]}{K_{a2}} + 1 + \dfrac{K_{a3}}{[H^+]}}$$

$$\mathbf{c}_4 = \alpha_4 A_{tot} = \frac{A_{tot}}{\dfrac{[H^+]^3}{K_{a1}K_{a2}K_{a3}} + \dfrac{[H^+]^2}{K_{a2}K_{a3}} + \dfrac{[H^+]}{K_{a3}} + 1} \qquad (4.6)$$

where $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{c}_3$ and $\mathbf{c}_4$ are row vectors of the matrix C.
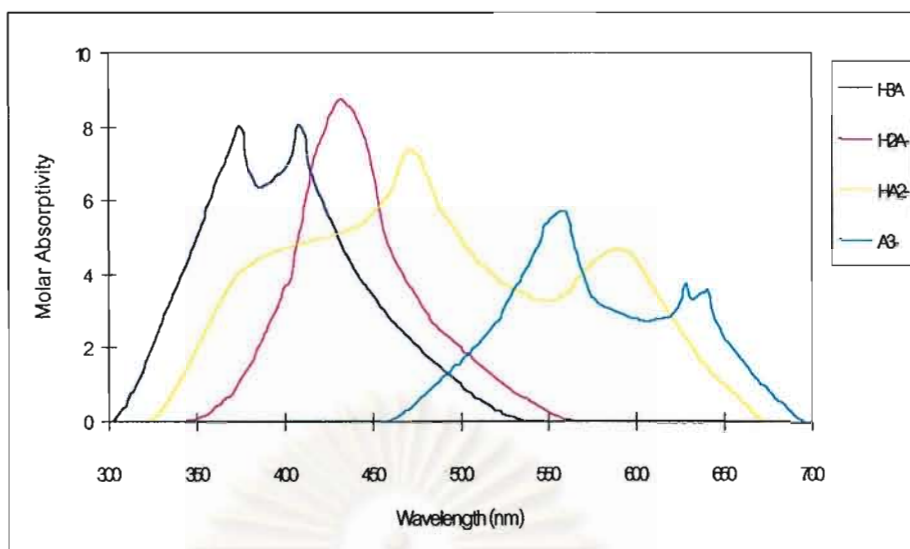
Figure 4.1 Artificial molar absorptivity spectra of 4 components for the acid-base equilibria.
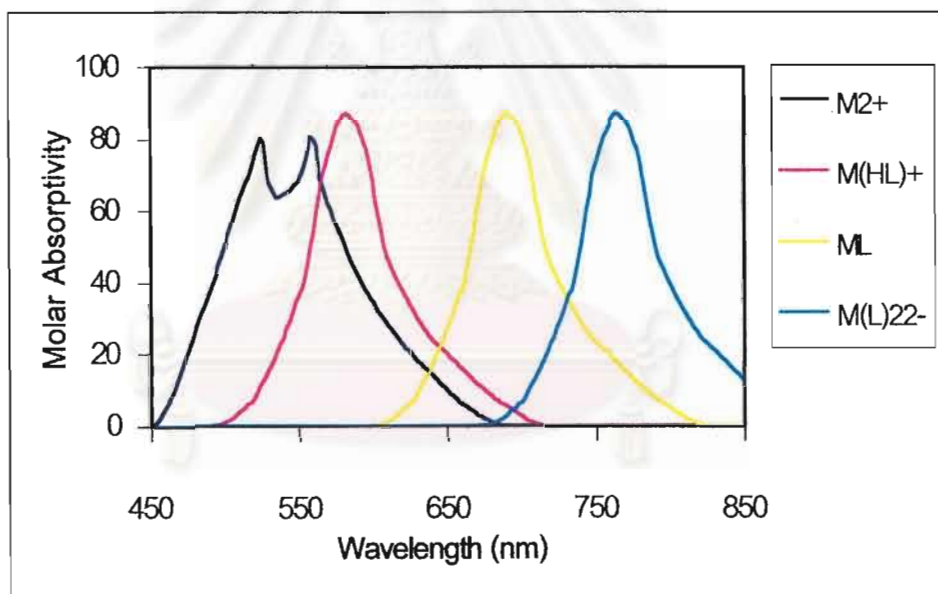
Figure 4.2 Artificial molar absorptivity spectra of 4 components for the complex equilibria.

In case of metal-ligand complex equilibria, the matrix $\mathbf{C}$ were formulated by the equation (2.20) where $p = 3$, and using the model $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 2 \end{bmatrix}$

$$\mathbf{c}_1 = \alpha_1 M_{tot} = \frac{M_{tot}}{1 + \beta_1[HL] + \beta_2\dfrac{[HL]}{[H^+]} + \beta_3\dfrac{[HL]^2}{[H^+]^2}}$$

$$\mathbf{c}_2 = \alpha_2 M_{tot} = \frac{M_{tot}}{\dfrac{1}{\beta_1[HL]} + 1 + \dfrac{\beta_2}{\beta_1}\dfrac{1}{[H^+]} + \dfrac{\beta_3}{\beta_1}\dfrac{[HL]}{[H^+]^2}}$$

$$\mathbf{c}_3 = \alpha_3 M_{tot} = \frac{M_{tot}}{\dfrac{1}{\beta_2}\dfrac{[H^+]}{[HL]} + \dfrac{\beta_1}{\beta_2}[H^+] + 1 + \dfrac{\beta_3}{\beta_2}\dfrac{[HL]}{[H^+]}}$$

$$\mathbf{c}_4 = \alpha_4 M_{tot} = \frac{M_{tot}}{\dfrac{1}{\beta_3}\dfrac{[H^+]^2}{[HL]^2} + \dfrac{\beta_1}{\beta_3}\dfrac{[H^+]^2}{[HL]} + \dfrac{\beta_2}{\beta_3}\dfrac{[HL]}{[H^+]} + 1} \quad (4.7).$$

The concentration profiles (matrix $\mathbf{C}$) in both cases were given in Figure 4.3 and 4.4.
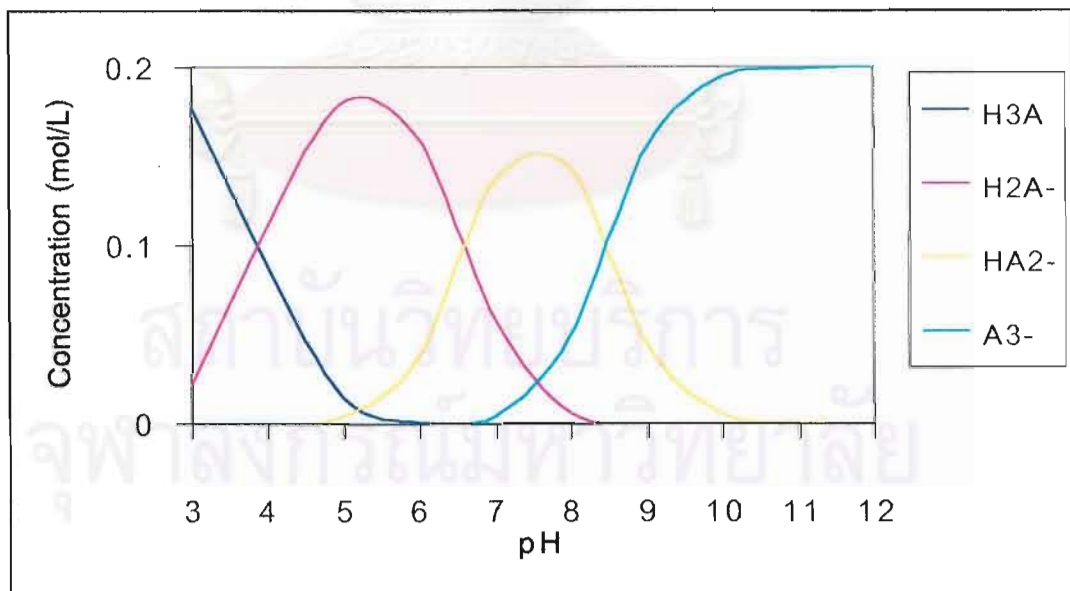
Figure 4.3 Artificial concentration profiles of 4 components computed by equation (4.6).
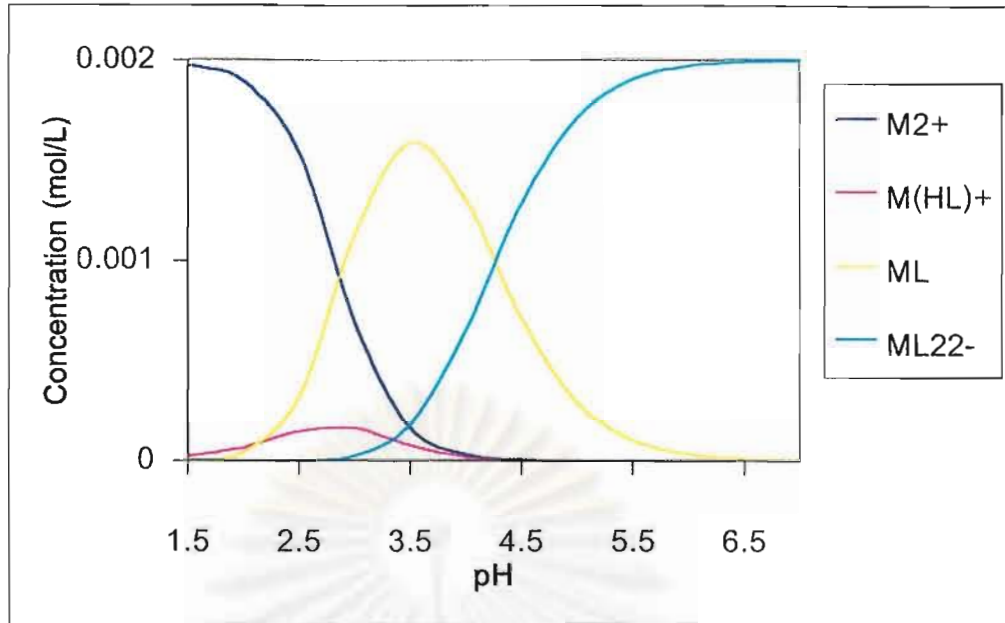
Figure 4.4 Artificial concentration profiles of 4 components computed by equation (4.7).

Consequently, the simulated spectra were computed by (4.5) at different levels of error yielding absorbance matrices with dimensions of (201×10) and (201×12), respectively. Then the simulated spectra were shown in Figure 4.5, 4.6, 4.7, 4.8, 4.9 and 4.10.
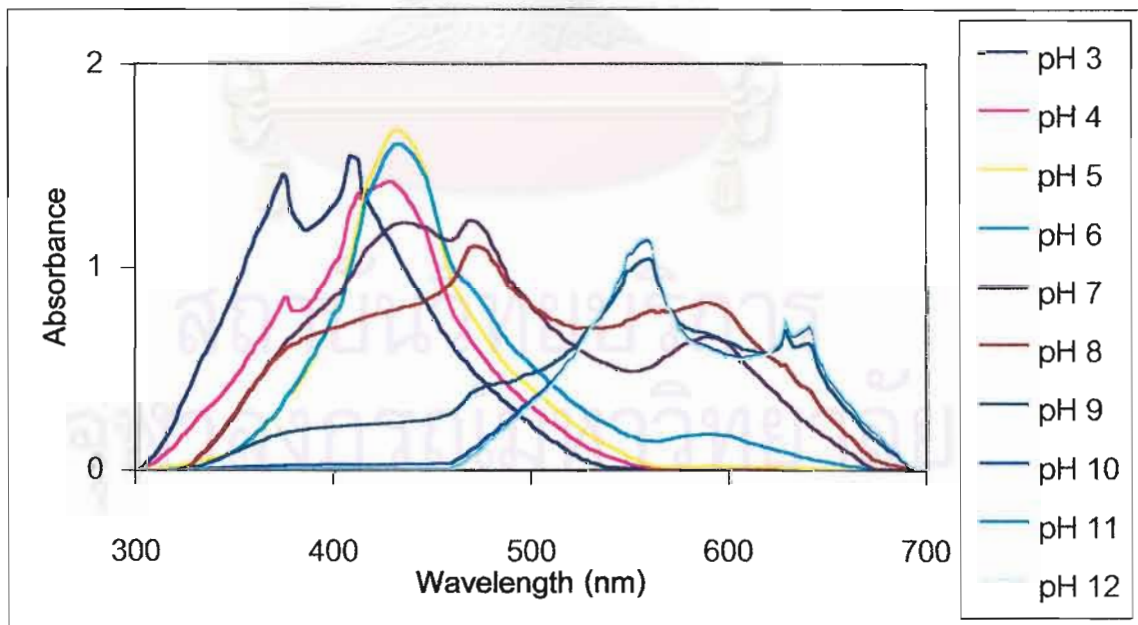
Figure 4.5 Ten simulated spectra of H_3A dissociation without added error.
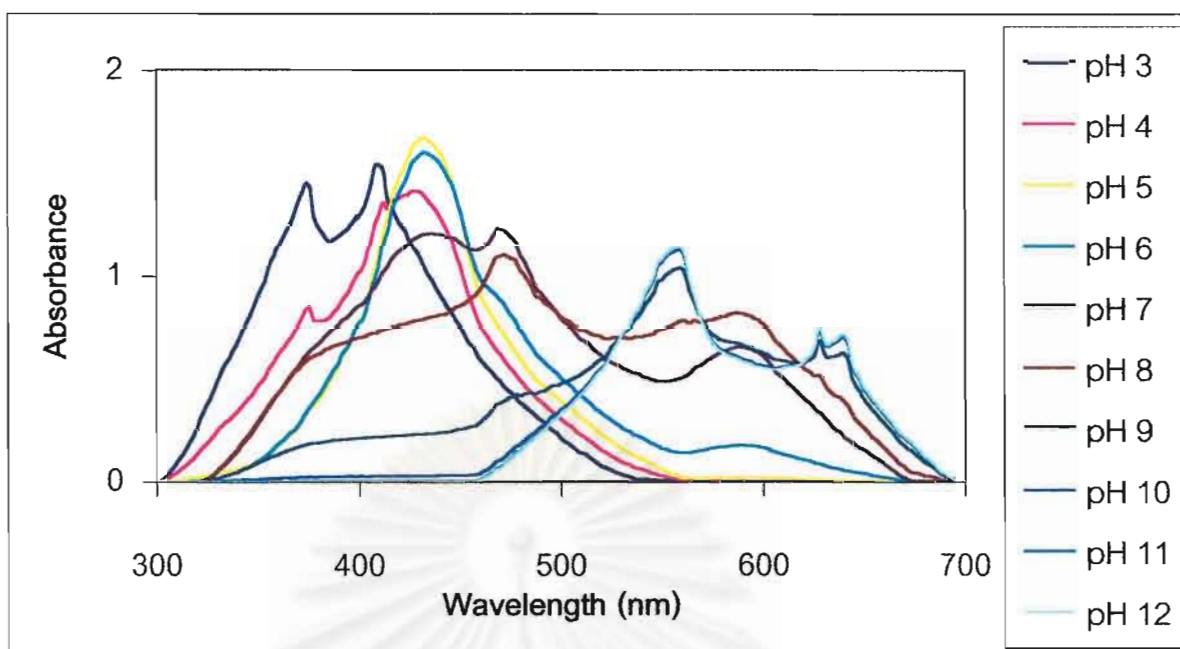
Figure 4.6 Ten simulated spectra of H$_3$A dissociation with ±0.0005-added error.
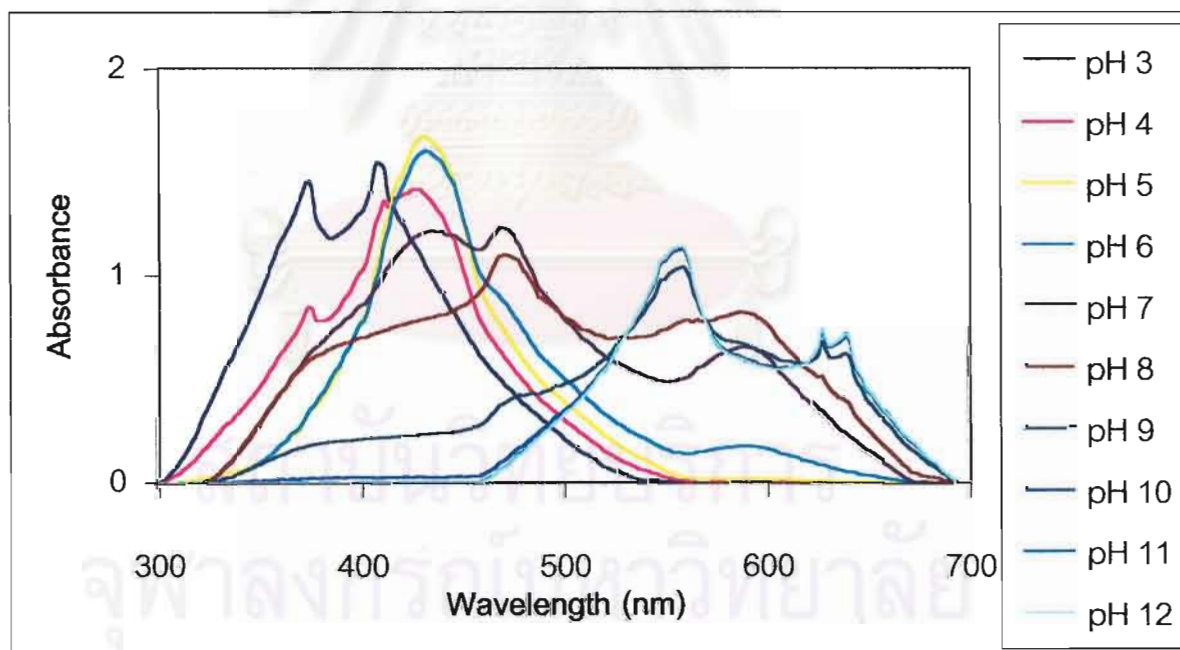
Figure 4.7 Ten simulated spectra of H$_3$A dissociation with ±0.0015-added error.
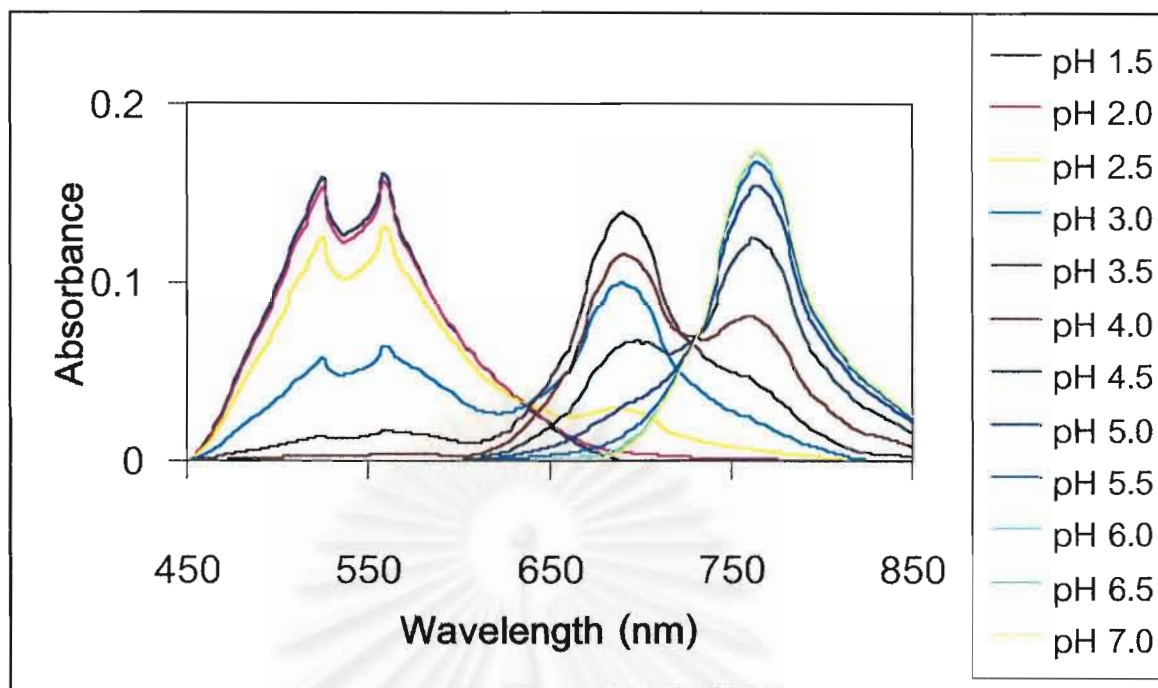
Figure 4.8 Twelve simulated spectra of M-HL complexation without added error.
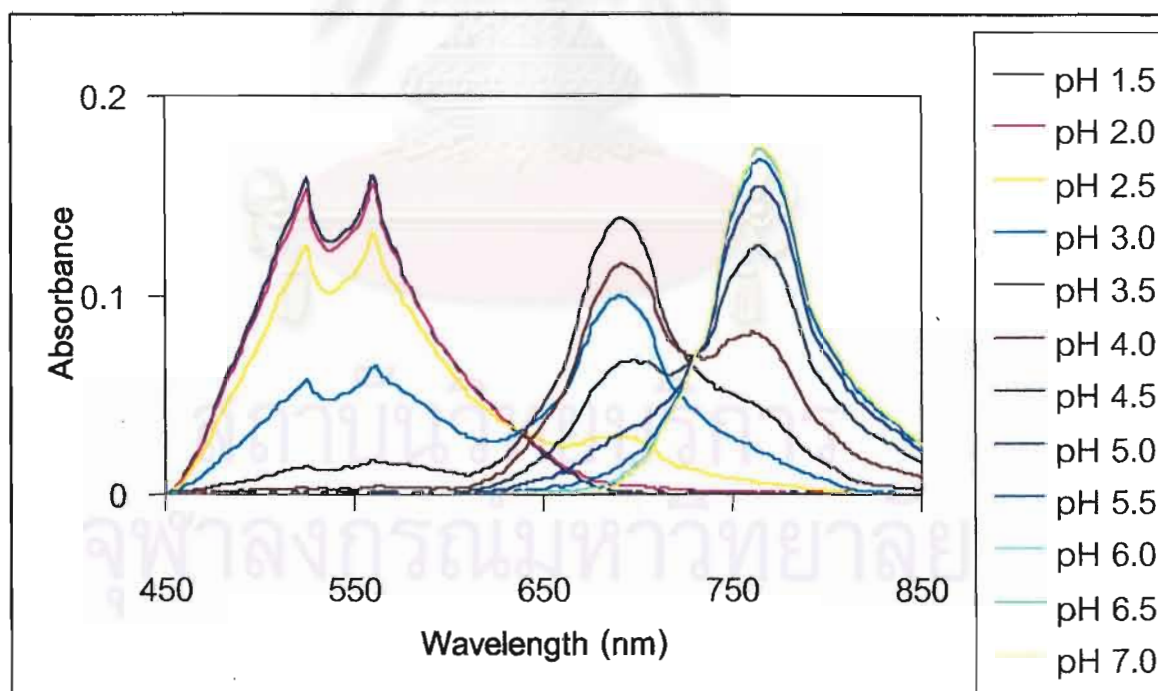
Figure 4.9 Twelve simulated spectra of M-HL complexation with ±0.0005-added error.
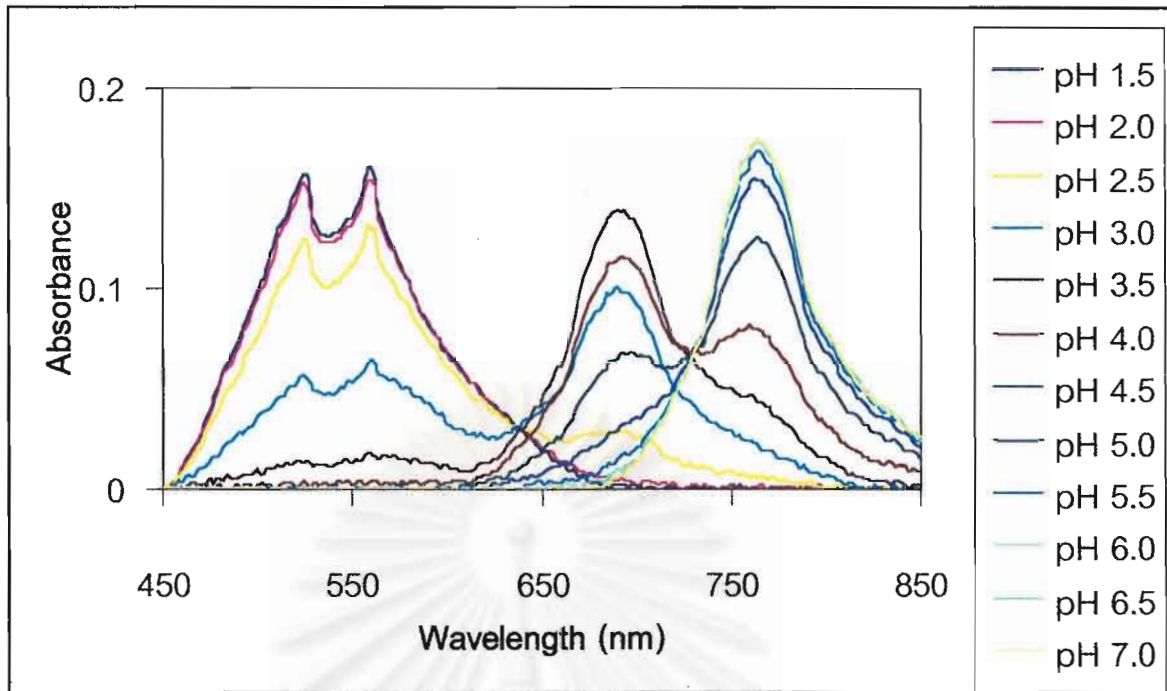
Figure 4.10 Twelve simulated spectra of M-HL complexation with ±0.0015-added error.

4.3 Program Testing by Simulated Spectra

The simulated absorbance matrices as detailed in Section 4.2 were created as text files (.txt) using text editor, and then input to the Program MTFA. After resolving the absorbance data, the testing results of the systems of H_3A dissociaion and M-HL complexation were tabulated in Table 4.2 and 4.3, respectively. Further details of testing were reported in Appendix. In the tables, RMS and Norm were signified the root mean squars and Euclidean norm of a matrix $\mathbf{X}$ ($r \times c$) [36-37], and defined as

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^{r}\sum_{k=1}^{c} x_{ik}^2}{rc}} \tag{4.8}$$

$$\text{Norm} = \sqrt{\sum_{i=1}^{r}\sum_{k=1}^{c} x_{ik}^2} \tag{4.9}$$

And the "Diff." was an abbreviation of difference.

From the results by Program MTFA, the followings are concluded:

- The number of components is distinctly "4", and consistent to the former settings.

- The output constants of dissociation and stability are in very good agreement with the beforehand settings, albeit the errors were added.

- The concentration matrix differences between the program output and the original settings are small, *i.e.* the fifth digit (or more than) after the decimal point of the concentration in mol/L, and quite acceptable in the range of error.

- The differences of molar absorptivities showed similar behavior with the differences of concentration matrix, but more deviation was observed.

- The extractable error is approximately 77% of the added error. This showed the prominent feature of factor analysis in reducing the error from the raw data.

- Explicitly, the system of acid-base equilibria showed highly accurate results with respect to the complex equilibria. Undoubtedly, these were caused by the simplicity and uniqueness of the model function as stated in Section 2.4.3.
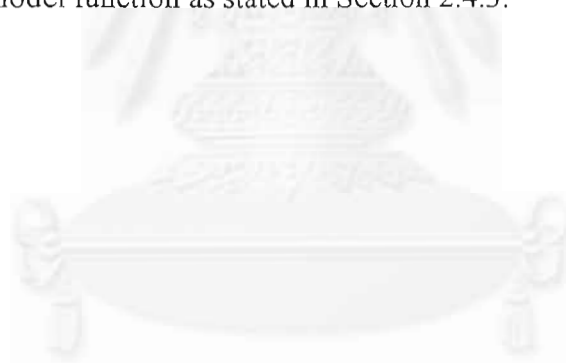
Table 4.2 Testing results of H_3A dissociation by program MTFA

Results		Beforehand settings	Pure data (±0.0000 error)	Pure data (±0.0005 error)	Pure data (±0.0015 error)
No. of components		4	4	4	4
Added Error:	RMS	-	-	2.904×10^{-4}	8.712×10^{-4}
	Norm	-	-	1.302×10^{-2}	3.906×10^{-2}
Extractable Error:	RMS	-	2.144×10^{-7}	2.208×10^{-4}	6.356×10^{-4}
	Norm	-	9.611×10^{-6}	9.897×10^{-3}	2.849×10^{-2}
K_{a1}		1.230×10^{-4}	1.230×10^{-4}	1.230×10^{-4}	1.232×10^{-4}
K_{a2}		2.340×10^{-7}	2.340×10^{-7}	2.340×10^{-7}	2.340×10^{-7}
K_{a3}		3.450×10^{-9}	3.450×10^{-9}	3.451×10^{-9}	3.450×10^{-9}
χ^2-value		-	9.801×10^{-11}	6.765×10^{-6}	7.760×10^{-5}
Diff. of matrix C:	RMS	-	1.589×10^{-8}	5.106×10^{-6}	2.822×10^{-5}
	Norm	-	1.005×10^{-7}	3.229×10^{-5}	1.444×10^{-4}
Diff. of matrix E:	RMS	-	5.802×10^{-7}	1.124×10^{-3}	3.544×10^{-3}
	Norm	-	1.645×10^{-5}	3.188×10^{-2}	0.101

Table 4.3 Testing results of M-HL complexation by program MTFA

Results		Beforehand settings	Pure data (±0.0000 error)	Pure data (±0.0005 error)	Pure data (±0.0015 error)
No. of components		4	4	4	4
Added Error:	RMS	-	-	2.878×10^{-4}	8.568×10^{-4}
	Norm	-	-	1.414×10^{-2}	4.208×10^{-2}
Extractable Error:	RMS	-	2.175×10^{-7}	2.126×10^{-4}	6.274×10^{-4}
	Norm	-	1.068×10^{-5}	1.044×10^{-2}	3.081×10^{-2}
β_1		7.590	7.590	7.678	7.883
β_2		5.370×10^{-2}	5.370×10^{-2}	5.356×10^{-2}	5.282×10^{-2}
β_3		3.390×10^{-5}	3.390×10^{-5}	3.395×10^{-5}	3.366×10^{-5}
χ^2-value		-	4.061×10^{-11}	2.582×10^{-5}	1.894×10^{-4}
Diff. of matrix C:	RMS	-	2.041×10^{-8}	8.484×10^{-7}	2.963×10^{-6}
	Norm	-	1.414×10^{-7}	5.878×10^{-6}	2.053×10^{-5}
Diff. of matrix E:	RMS	-	1.132×10^{-3}	0.832	2.817
	Norm	-	3.209×10^{-2}	23.582	79.877

CHAPTER 5

EXPERIMENTAL

5.1 Chemicals and Instruments

5.1.1 Chemicals

All used chemicals were tabulated in Table 5.1.

Table 5.1 List of used chemicals

Chemicals	Purity and Source
Glycine	99.7% from Riedel-deHaen (Germany).
Alanine	99% from Fluka (Switzerland).
Cupric chloride anhydrous	97% from Fluka (Switzerland).
Sodium chloride	Analytical reagent grade from Mallinckrodt (USA).
Sodium hydroxide	98% (pellet) from Eka Nobel Industries (Sweden).
Hydrochloric acid	Concentrated acid from BDH Laboratory (England).

5.1.2 Instruments

1. Ultraviolet-visible spectrophotometer: Spectronic Genesys 5 from Spectronic Instruments, Inc. (USA).

2. Analog pH meter coupled with combined glass electrode: Scholar 425 from Corning, Inc. (USA).

3. Temperature-controlled shaker: Maxi-Shake from Heto Lab Equipment A/S (Denmark).

5.2 Preparation of Solutions

1. <u>Stock solution of 0.01 M cupric chloride</u>. Prepare by dissolution of accurate quantity of anhydrous cupric chloride in 0.6 M sodium chloride.

2. <u>Stock solution of 0.5 M glycine</u>. Prepare by dissolution of accurate quantity of glycine in 0.6 M sodium chloride.

3. <u>Stock solution of 0.5 M alanine</u>. Prepare by dissolution of accurate quantity of alanine in 0.6 M sodium chloride.

4. <u>2 M hydrochloric acid</u>. Prepare by dilution of approximate volume of conc. hydrochloric acid in 0.6 M sodium chloride.

5. <u>2, 0.2 and 0.002 M sodium hydroxide solution</u>. Prepare by dissolution of approximate quantities of sodium hydroxide in 0.6 M sodium chloride.

6. <u>0.6 M sodium chloride solution</u>. Prepare by dissolution of approximate quantity of sodium chloride in doubly distilled water.

7. <u>18 solutions of copper-glycine</u>. Prepare by pipetting appropriate amount from stock solutions of cupric chloride, glycine, and sodium hydroxide or hydrochloric acid into 50-mL volumetric flasks, and then diluting the mixture with 0.6 M sodium chloride. Each solution contained 0.002 M copper(II) and 0.10 M glycine but different pH ranging from 1 to 7.

8. <u>13 solutions of copper-glycine</u>. Prepare by pipetting appropriate amount from stock solutions of cupric chloride, alanine, and sodium hydroxide or hydrochloric acid into 50-mL volumetric flasks, and then diluting the mixture with 0.6 M sodium chloride. Each solution contained 0.002 M copper(II) and 0.10 M alanine but different pH ranging from 1 to 7.

5.3 Measurements

The visible absorption spectra of all sample solutions were recorded with a UV/vis spectrophotometer. The baseline was recorded using 0.6 M sodium chloride and the

"baseline" option from the instrument's program menu. Absorbances of copper-glycine samples were recorded at 3-nm intervals from 450 to 850 nm yielding 134 points for each solution, while each copper-alanine sample was at 2-nm intervals yielding 201 points. All solutions also were temperature-controlled at 33°C before measuring.

The pH of each solution was measured by an analog pH meter which was calibrated by standard buffer of pH 4.0 and 7.0 supplied by Corning, Inc.

CHAPTER 6

RESULTS AND DISCUSSION

From the experiment, eighteen solutions of copper-glycine, and thirteen solutions of copper-alanine were prepared with pHs varying from 1 to 7. The pH of greater than 7 can not be used since copper would be precipitated. The visible spectra of the copper-glycine solutions were recorded from 450 to 850 nm at 3-nm interval (Figure 6.1), yielding a 134×18 absorbance matrix, while the solutions of copper-alanine were recorded at the same range but 2-nm intervals (Figure 6.2) yielding a 201×13 absorbance matrix. Then the text file (.txt) of absorbance matrices were created using text editor such as Microsoft Editor, Microsoft FORTRAN or a word processor program, and input to the Program MTFA.

Along the program, the filename of text file (absorbance matrix) and dimension of matrix should be specified, and then the program would read the data automatically as described in Section 3.1.1 and 3.1.2. All input and settings were detailed in the reports in Appendix.
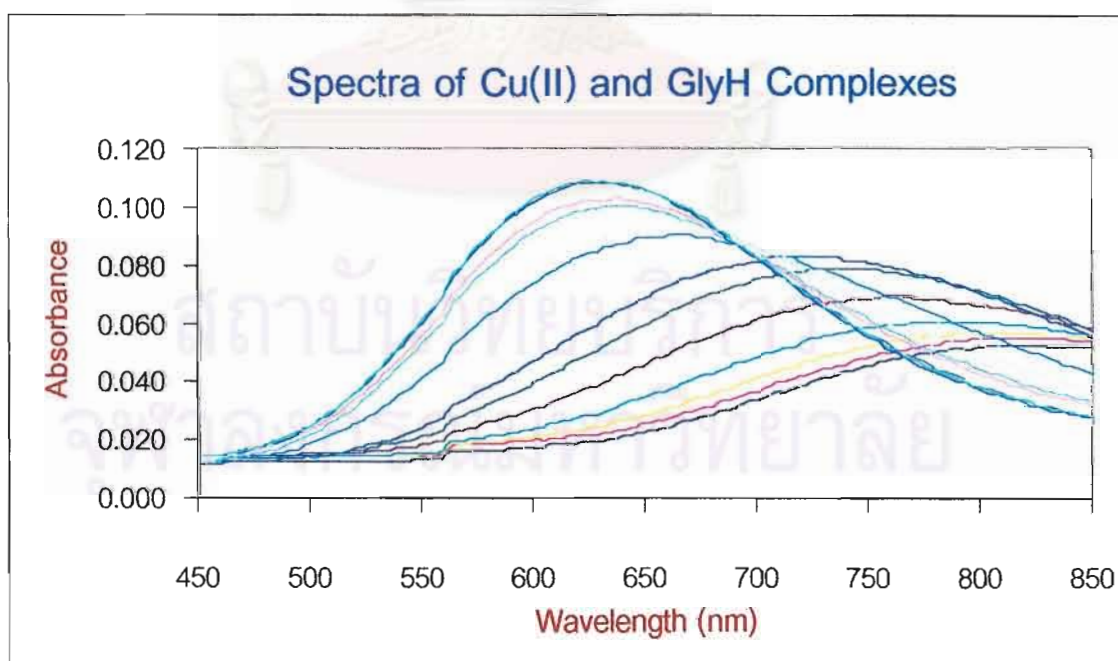
Figure 6.1 Visible spectra of 18 copper-glycine solutions with pH ranging from 1 to 7.
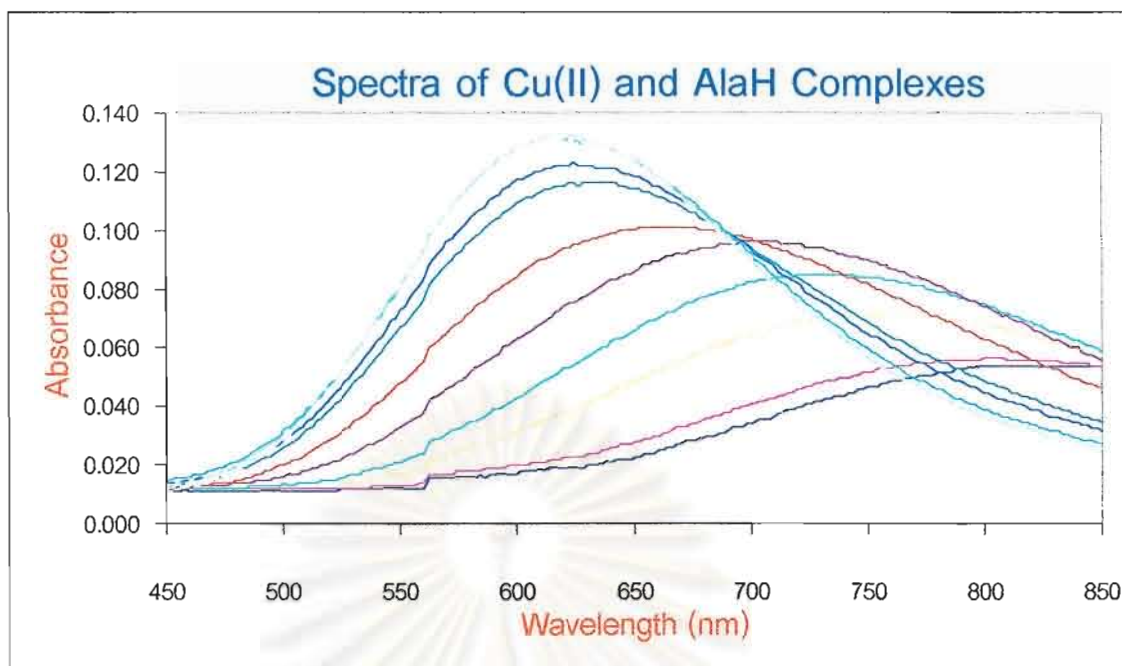
Figure 6.2 Visible spectra of 13 copper-alanine solutions with pH ranging from 1 to 7.

After analyzing the absorbance data by MTFA program, the results were reported and collected in Appendeix.

Interpreting the Results

Explicitly, the output results of the both systems of copper-glycine and copper-alanine were closed together, and possessed a similar trend in the properties.

Determination of the Number of Copper Species. For both systems of copper-glycine and copper-alanine, three important error indicators, i.e., IND, IE and %SL, given the consensus of "4" primary eigenvectors which corresponding to 4 absorbing species in the equilibria. Since glycine has no absorption in the region of measurement, the absorption ought to be contributed by copper and copper complexes, *i.e.* Cu^{2+}, $Cu(GlyH)^{2+}$, $Cu(Gly)^{+}$, and $Cu(Gly)_2$. These species given were taken from suggestions of previous literature.[18] Analogous to copper-glycine system, the 4 absorbing species for the system of copper-alanine

would be Cu^{2+}, $Cu(AlaH)^{2+}$, $Cu(Ala)^{+}$, and $Cu(Ala)_{2}$. It should be noted that the fourth component has very small eigenvalue with respect to the first three components. This means that the fourth component exists in very small amount in the equilibria.

Determination of the Equilibrium Constants, Concentration Profiles and Absorptivities. By the program suggestion, there are "3" steps of copper-glycine complexation. From Darj and Malinowski [18], the corresponding model should be

$$Cu^{2+} + GlyH \xrightleftharpoons{\beta_1} Cu(GlyH)^{2+}$$

$$Cu^{2+} + GlyH \xrightleftharpoons{\beta_2} Cu(Gly)^{+} + H^{+}$$

$$Cu^{2+} + 2(GlyH) \xrightleftharpoons{\beta_3} Cu(Gly)_{2} + 2H^{+} \qquad (6.1).$$
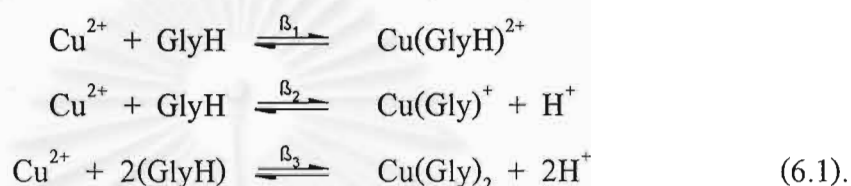
Incorporating this to the input, the factor analysis program yielded 3 overall stability constants (β);

$$\beta_1 = 4.147$$
$$\beta_2 = 3.446 \times 10^{-2}$$
$$\beta_3 = 4.415 \times 10^{-4} \qquad (6.2).$$

In the same way, the complexation model of copper-alanine would be

$$Cu^{2-} + AlaH \xrightleftharpoons{\beta_1} Cu(AlaH)^{2+}$$

$$Cu^{2+} + AlaH \xrightleftharpoons{\beta_2} Cu(Ala)^{+} + H^{+}$$

$$Cu^{2+} + 2(AlaH) \xrightleftharpoons{\beta_3} Cu(Ala)_{2} + 2H^{+} \qquad (6.3).$$
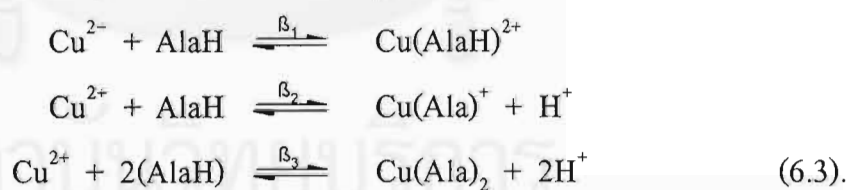
Then the program yielded 3 overall stability constants as

$$\beta_1 = 11.68$$
$$\beta_2 = 4.605 \times 10^{-2}$$
$$\beta_3 = 3.753 \times 10^{-4} \qquad (6.4).$$

However, the other complexation models expressed in term of the matrix form such as

$$\begin{array}{ccc} Cu^{2+} & GlyH_2^+ & H^+ \end{array}$$
$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 2 & 4 \end{bmatrix}$$ and $$\begin{array}{ccc} Cu^{2+} & Gly^- & H^+ \end{array}$$ $$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \end{bmatrix}$$ also were introduced to fit the data, but the output

results were unreasonable or impossible in real conditions such as negative concentrations, negetive stability constants, high noise in molar absorptivities, and so on. Moreover, some

models such as $$\begin{array}{ccc} Cu^{2+} & AlaH_2^+ & H^+ \end{array}$$ $$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 4 \end{bmatrix}$$ and $$\begin{array}{ccc} Cu^{2+} & Ala^- & H^+ \end{array}$$ $$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \end{bmatrix}$$ fail to converge with acceptable

χ^2 as defined by equation (2.46) in Section 2.4.3.

The output matrices of concentration and absorptivities were exported to create the charts as shown in Figure 6.3, 6.4, 6.5 and 6.6.
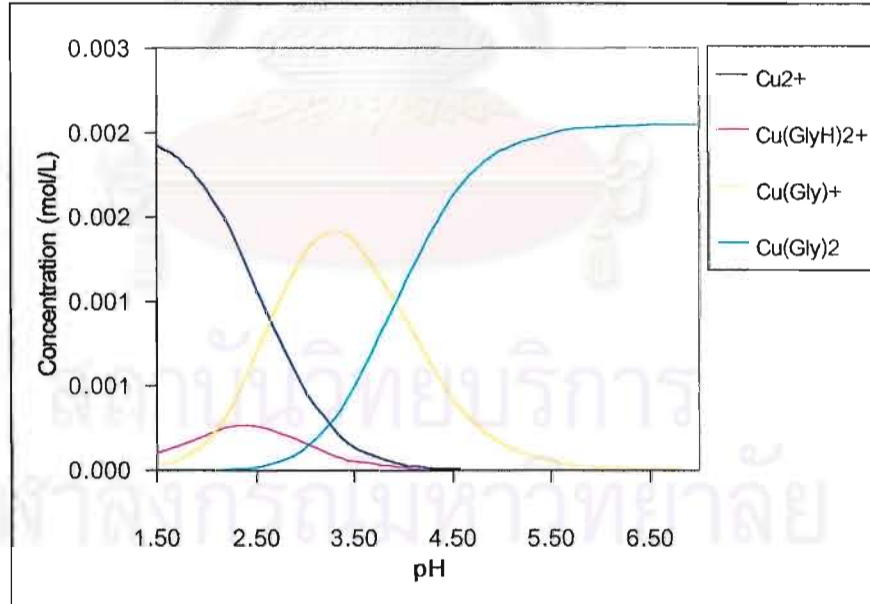
Figure 6.3 Estimated concentration profiles for the system of copper-glycine.
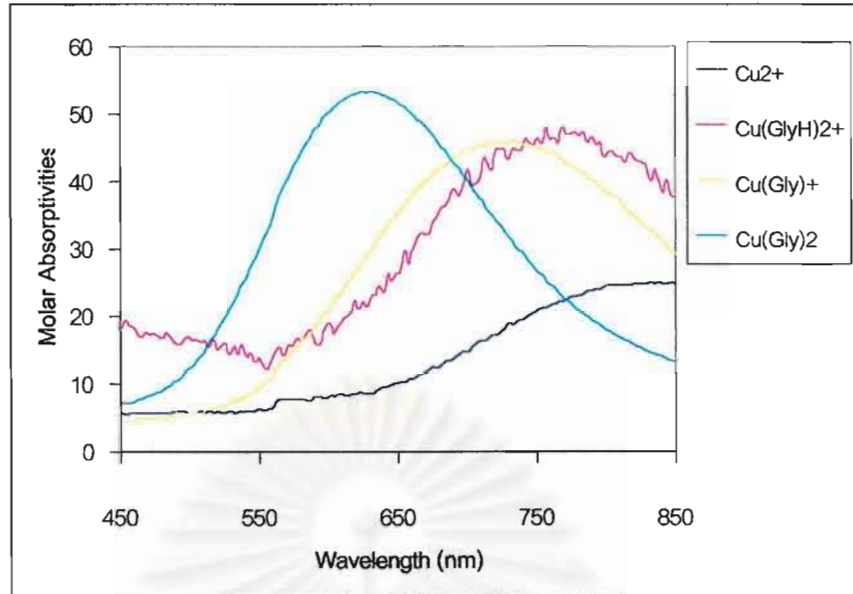
Figure 6.4 Estimated molar absorptivity spectra for the system of copper-glycine.
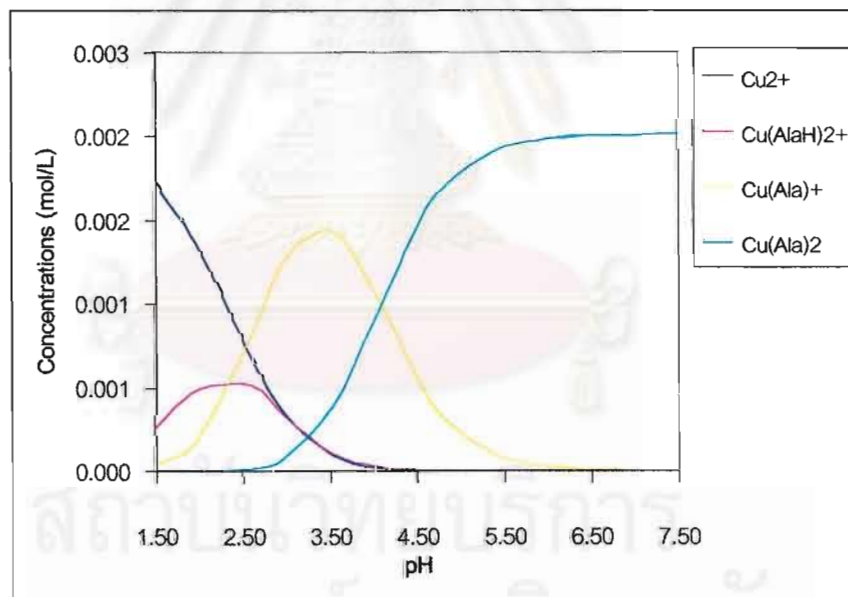
Figure 6.5 Estimated concentration profiles for the system of copper-alanine.
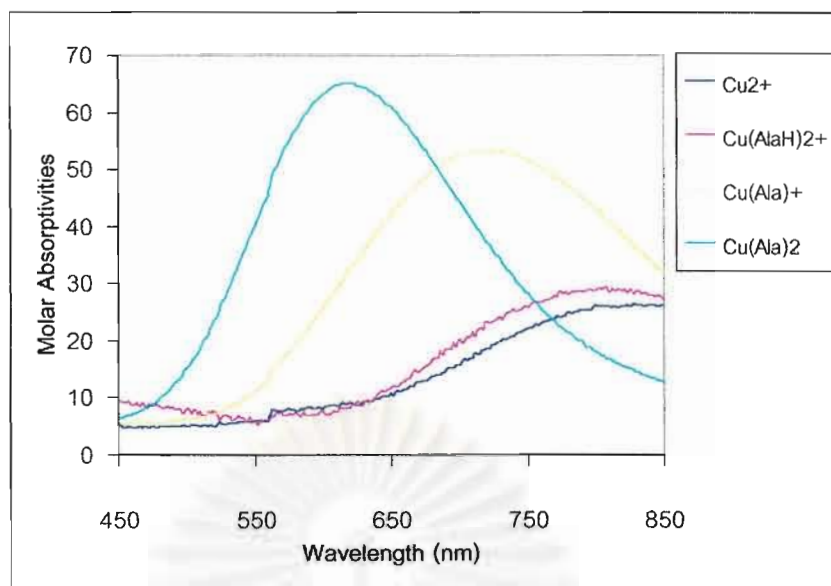
Figure 6.6 Estimated molar absorptivity spectra for the system of copper-alanine.

Evidently, the concentrations of $Cu(GlyH)^{2+}$ and $Cu(AlaH)^{2+}$ are smaller than the other copper forms, and exist in a narrow region of the pH range. These conform the results of eigenvalues as stated above. However, most of errors were accumulated in the molar absorptivities especially the species which contained small amount such as $Cu(GlyH)^{2+}$ and $Cu(AlaH)^{2+}$ as seen in Figure 6.3 to 6.6.

It should be noted that the concentration of $Cu(GlyH)^{2+}$ is less than those of $Cu(Gly)^{+}$ and $Cu(Gly)_2$ as shown in Figure 6.3, even if the value of β_1 is greater than those of β_2 and β_3, respectively. The reason is that the concentrations of each component are comparable as

$$[CuGlyH^{2+}] \quad > \quad [CuGly^{+}][H^{+}] \quad > \quad \frac{[CuGly_2][H^{+}]}{[GlyH]} \qquad (6.5).$$

This implied that the concentration of $Cu(GlyH)^{2+}$ is less than those of $Cu(Gly)^{+}$ and $Cu(Gly)_2$, but larger than the products of $[CuGly^{+}][H^{+}]$ and $\dfrac{[CuGly_2][H^{+}]}{[GlyH]}$, respectively. Similar for copper-glycine, the comparison of concentrations of copper-alanine system should be

$$[CuAlaH^{2+}] \quad > \quad [CuAla^{+}][H^{+}] \quad > \quad \frac{[CuAla_2][H^{+}]}{[AlaH]} \qquad (6.6)$$

Discussion with Other Results   Our results seem to contradict to the literature. [34-35]   Irving and Pettit [35] used the potentiometric method to study the complexation of copper and glycine in 0.10 M KCl at 20°C, and proposed two complexation steps as

$$Cu^{2-} + Gly^- \underset{\beta_1=1.32\times10^8}{\xrightleftharpoons{\hspace{2cm}}} Cu(Gly)^+$$

$$Cu^{2+} + 2(Gly^-) \underset{\beta_2=1.08\times10^{15}}{\xrightleftharpoons{\hspace{2cm}}} Cu(Gly)_2 \qquad (6.7).$$
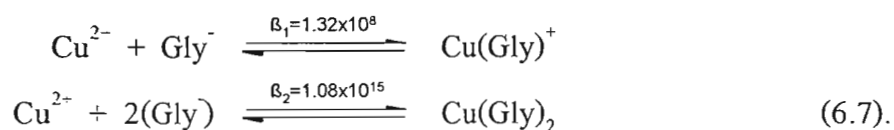
The sensitivity of spectrophotometry may be considered as higher than that of potentiometry because the chemical species contribute the total absorption directly, while the potentiometry they must impact on the hydrogen concentration gradient which consider as indirected to electrical potential.   So, the potentiometric technique may unable to discriminate $Cu(GlyH)^{2+}$ species from $Cu(Gly)^+$, and then concluded these two-step to one-step complexation because $Cu(GlyH)^{2+}$ exist only in small amount as seen in Figure 6.3.   In addition, the above results of copper-glycine complexes, *i.e.* number of components, number of complexation steps, and complexation model, were corresponding to Darj and Malinowski's study [18] where the window factor analysis (WFA), a self-modeling method, was used to evaluate the visible spectra of copper-glycine.   Nonetheless, their report of overall stability constants was (compared to equation (6.2))

$$\beta_1 = 7.59$$
$$\beta_2 = 5.37\times10^{-2}$$
$$\beta_3 = 3.39\times10^{-4} \qquad (6.8).$$

Their results are in the same order of magnitude as ours.   In addition, the stability constants of copper-glycine system are quite closer to that of copper-alanine one.   They are also expected because the molecular structure and acid dissociation constants of glycine are similar to the alanine.   The deviations may be caused by

-   Temperature difference.   Regrettably that the study of Darj and Malinowski [18] has no report about the experimental temperature, and made a suspect to compare the results.

- Variation of base-line. When long-time experiment was performed, the base-line may be changed significantly from the beginning until ceasing, and these may cause by the electric fluctuation.

- Nature of components. Along the way of principal factor analysis, the chemical species which existed in the system with a little amount would be hardly discriminated with the error or noise, and then given a more deviated results.

- Limitation of factor analysis and modeling transformation. According to the modeling transformation in Section 2.4.3, the method tried to evaluate the concentrations at the first, and then used the results to calculate the molar absorptivities. If the concentration values contained error were used to calculate, more error would be accumulated in the molar absorptivities.

CHAPTER 7

CONCLUSIONS

Program War version 1.0 was developed to execute the factor analysis and modeling transformation. The program was validated and utilized to determine the equilibrium constants and the concentration and spectral profiles of the components in chemical equilibrium from UV/VIS spectroscopic titrations. As demonstrated in Chapter 4 and 6, the program works well on the data that are obtained by simulation and experiment. However, some comments should be noted here.

Number of Components. A common problem to the methods of chemometrics and factor analysis is to determine the number of principal components. The possible complications due to the mistake might be base-line variations, contribution from solvents, formation of unexpected species, and so on. Fortunately, this is rarely a problem for the system considered here.

Stability of the Program. The stability of the program or the method of factor analysis depends on several factors such as the number of spectra analyzed, the number of data points in each spectrum, the degree of spectral overlap between the components, the signal to noise ratios of the spectra, and how close to the titration end points one can reach. The effect of these factors causes the errors in calculations of concentrations, absorptivities and equilibrium constants. For example in Section 4.3, when the level of error in simulated spectra was increased, the accuracy of calculation was decreased.

Application of the Program. Essentially any equilibrium systems could be analyzed by the program. Especially, the systems of acid-base equilibria would be analyzed with high accuracy albeit existing the high level of noise. When the equilibrium expression is not

known, the program can be used to test different models. The only requirement is that the spectral response is linear, and the program is therefore applicable to most UV/VIS spectroscopic techniques. The analysis could be completed in a few minutes on a standard personal computer.

REFERENCES

1. Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., de Jong, S., Lewi, P. J., and Smeyers-Verbeke, J. Handbook of chemometrics and qualimetrics: Part A. Netherlands: Elsevier Science B. V., 1997.

2. Malinowski, E. R. Factor analysis in chemistry. 2nd ed. New York: John Wiley & Sons, 1991.

3. Knorr, F. J., and Harris, J. M. Analytical Chemistry 53, No.2, (1981): 272-276.

4. Knorr, F. J., Thorsheim, H. R., and Harris, J. M. Analytical Chemistry 53, No.6, (1981): 821-825.

5. Frans, S. D., McConnell, M. L. and Harris, J. M. Analytical Chemistry 57, No.8, (1985): 1552-1559.

6. Sasic, S. Analyst 123, (1998): 1193-1197.

7. Zissis, K. D., Brereton, R. G., and Escott, R. Analyst 122, (1997): 1007-1013.

8. Frans, S. D., and Harris, J. M. Analytical Chemistry 57, No.8, (1985): 1718-1721.

9. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. Numerical recipes in FORTRAN. 2nd ed. USA: Cambridge University Press, 1992.

10. Kubista, M., Sjoback, R., and Albinsson, B. Analytical Chemistry 65, No.8, (1993): 994-998.

11. Brereton, R. Analyst 120, (1995): 2313-2336.

12. Bartlett, M. S. British Journal of Psychology (Statistical Section) 3, (1950): 77-85.

13. Hugus, Z. Z., Jr., and El-Awady, A. A. Journal of Physical Chemistry 75, No.19, (1971): 2954-2957.

14. Malinowski, E. R. Analytical Chemistry 49, No.4, (1977): 606-617.

15. Malinowski, E. R. Journal of Chemometrics 1, (1987): 33-40.

16. Malinowski, E. R. Journal of Chemometrics 3, (1988): 49-60.

17. Lorber, A. Analytical Chemistry 56, No.6, (1984): 1004-1010.

18. Darj, M. M., and Malinowski, E. R. Analytical Chemistry 68, No.9, (1996):
    1593-1598.

19. Hartley, F. R., Burgess, C., and Alcock, R. M. Solution equilibria. Great Britain:
    Ellis Horwood, 1980.

20. Faires, J. D., and Burden, R. L. Numerical methods. USA: PWS Publishing, 1993.

21. Malinowski, E. R. Journal of Chemometrics 6, (1992): 29-40.

22. Kankare, J. J. Analytical Chemistry 42, No.12, (1970): 1322-1326.

23. Meloun, M., Militky, J., and Forina, M. Chemometrics for analytical chemistry
    volume 2: PC-aided regression and related methods. Great Britain:
    Ellis Horwood, 1994.

24. Spendley, W. in Optimization. London: Academic Press, 1969.

25. Nelder, J. A., and Mead, R. Journal of Computation 7, (1965): 308.

26. Deming, S. N., and Morgan, S. L. Analytical Chemistry 45, No.3, (1973):
    278A-283A.

27. Marsili-Libelli, S., and Castelli, M. Applied Mathematics and Computation 23,
    (1987): 341-357.

28. http://www.mcs.surrey.ac.uk.

29. Nyhoff, L., and Leestma, S. FORTRAN 77 for engineering and scientists. 3rd ed.
    Singapore: Macmillan Publishing, 1988.

30. เดือน สินธุพันธ์ประทุม. เทคนิคฟอร์แทรน 77. กรุงเทพมหานคร: สำนักพิมพ์
    จุฬาลงกรณ์มหาวิทยาลัย, 2535.

31. Thompson, C. J., Danielson, J. D. S., and Callis, J. B. Analytical Chemistry 69,
    No.1, (1997): 25-35.

32. Saurina, J., Hernandez-Cassou, S., and Tauler, R. Analytical Chemistry 67, No.20,
    (1995): 3722-3726.

33. Vandeginste, B. G. M., Derks, W., and Kateman, G. <u>Analytica Chimica Acta</u> 173, (1985): 253-264.

34. The Chemical Society <u>Stability constants supplement No.1</u>. Great Britain: Alden & Mowbray, 1971.

35. Irving, H., and Pettit, L. D. <u>Journal of Chemical Society</u> (1963): 1546-1553.

36. Goldberg, J. L. <u>Matrix theory with applications</u>. Singapore: McGraw-Hill, 1991.

37. Lipschutz, S. <u>Theory and problems of beginning linear algebra</u>. New York: McGraw-Hill, 1997.

APPENDIX

```
***********************************************************************
*    MTFA v.1 - The Factor Analysis Programme for Chemical Analysis   *
*                                                                     *
*    Created by Arunchai Tangcharoenbumrungsuk                        *
*    Conducted by Dr.Vudhichai Parasuk                                *
***********************************************************************


    1. Elementary Data
         - Data file name                      Acid0015
         - Dimension of data matrix            201 x 10
         - Type of problem                     Acid-Base Equilibrium
         - Optimization method                 Simplex Method
         - Set the initial guess eigenvector   Automatic
         - Stochastic Initialization Algor.    No

    2. Optimization Method
         - Number of steps of dissociaton of the acid  =  3
         - The starting guess
           Value of parameter no. 1 = 1.000E-04
           Value of parameter no. 2 = 1.000E-07
           Value of parameter no. 3 = 1.000E-09

         - The convergence are obtained
           Value of parameter no. 1 = 1.232E-04
           Value of parameter no. 2 = 2.340E-07
           Value of parameter no. 3 = 3.450E-09

         - The chi-squared value      = 7.760E-05

         - Total optimization time  =    8 s 84 cs

         - Number of iteration cycle =   74

    3. Value Setting
         - Gradient (determine the pricipal factor)   = 1.000E-12
         - Tolerance (simplex method)                 = 1.000E-20
         - Convergence (simplex method)               = 1.000E-16
         - Simplex parameters
           Reflection parameter    =  1.000   (default)
           Contraction parameter   =   .550   (default)
           Expansion parameter     =  2.900   (default)
           Reduction parameter     =   .500   (default)

    4. Value Input
         - Number of solutions    =  10
         - pH of solutions
           The pH of solution no. 1  =   3.000
           The pH of solution no. 2  =   4.000
           The pH of solution no. 3  =   5.000
           The pH of solution no. 4  =   6.000
           The pH of solution no. 5  =   7.000
           The pH of solution no. 6  =   8.000
           The pH of solution no. 7  =   9.000
           The pH of solution no. 8  =  10.000
           The pH of solution no. 9  =  11.000
           The pH of solution no.10  =  12.000


         - Total concentration of acid    =  .2000      mol/L
```

5. Eigenvectors and Error Functions

Results of Error Treatment :

n	Eigenvalue	RSD	RMS	IE	IND	%SL
1	443.214753	.326	.309	.103	4.019E-03	1.09
2	161.058484	.138	.124	6.177E-02	2.158E-03	.189
3	22.643744	7.554E-02	6.320E-02	4.138E-02	1.542E-03	1.69
4	8.028203	8.206E-04	6.356E-04	5.190E-04	2.279E-05	2.694E-1
5	.000167	8.014E-04	5.667E-04	5.667E-04	3.206E-05	46.1
6	.000149	7.856E-04	4.968E-04	6.085E-04	4.910E-05	48.3
7	.000137	7.717E-04	4.227E-04	6.457E-04	8.574E-05	50.6
8	.000129	7.570E-04	3.385E-04	6.771E-04	1.893E-04	53.4
9	.000122	7.334E-04	2.319E-04	6.957E-04	7.334E-04	59.0
10	.000108	-	-	-	-	-

By following methods, the number of primary eigenvectors are obtained :
 - Imbedded Error Function (IE)   =   4
 - Factor Indicator Function (IND)   =   4
 - Percent Significant Level (%SL)   =   4

By automatic discrimination :
The most possible number of primary eigenvectors   =   4

After judgement by user :
The number of primary eigenvectors   =   4

RMS of extractable error matrix   =   6.35633E-04

Norm of extractable error matrix   =   2.84973E-02

```
***********************************************************************
*    MTFA v.1 - The Factor Analysis Programme for Chemical Analysis   *
*                                                                     *
*    Created by Arunchai Tangcharoenbumrungsuk                        *
*    Conducted by Dr.Vudhichai Parasuk                                *
***********************************************************************


   1. Elementary Data
         - Data file name                     Comp0015
         - Dimension of data matrix           201 x 12
         - Type of problem                    Metal-Ligand Complex Equilibrium
         - Optimization method                Simplex Method
         - Set the initial guess eigenvector  Automatic
         - Stochastic Initialization Algor.   No

   2. Optimization Method
         - Number of steps of complex formation      =   3
         - The starting guess
           Value of parameter no. 1 =  1.00
           Value of parameter no. 2 = 1.000E-02
           Value of parameter no. 3 = 1.000E-05

         - The convergence are obtained
           Value of parameter no. 1 =  7.883
           Value of parameter no. 2 = 5.282E-02
           Value of parameter no. 3 = 3.366E-05

         - The chi-squared value     = 1.894E-04

         - Total optimization time  =  51 s 19 cs

         - Number of iteration cycle = 149

   3. Value Setting
         - Gradient (determine the pricipal factor)   = 1.000E-12
         - Tolerance (simplex method)                 = 1.000E-20
         - Convergence (simplex method)               = 1.000E-16
         - Simplex parameters
           Reflection parameter   =  1.000   (default)
           Contraction parameter  =   .550   (default)
           Expansion parameter    =  2.900   (default)
           Reduction parameter    =   .500   (default)

   4. Value Input
         - Number of solutions    =   12
         - pH of solutions
           The pH of solution no. 1  =    1.500
           The pH of solution no. 2  =    2.000
           The pH of solution no. 3  =    2.500
           The pH of solution no. 4  =    3.000
           The pH of solution no. 5  =    3.500
           The pH of solution no. 6  =    4.000
           The pH of solution no. 7  =    4.500
           The pH of solution no. 8  =    5.000
           The pH of solution no. 9  =    5.500
           The pH of solution no.10  =    6.000
           The pH of solution no.11  =    6.500
           The pH of solution no.12  =    7.000


         - Total concentration of ligand  =  .1000       mol/L
         - Total concentration of metal   = 2.0000E-03   mol/L
```

- Number of dissociation step of the ligand = 2
  Dissociation constant no. 1 = 4.467E-04
  Dissociation constant no. 2 = 1.667E-10

- Model of complex formation
  step 1 :   1   1   0
  step 2 :   1   1   1
  step 3 :   1   2   2

- Order of ligand that form complex with metal = 2

5. Eigenvectors and Error Functions

Results of Error Treatment :

n	Eigenvalue	RSD	RMS	IE	IND	%SL
1	4.361018	3.869E-02	3.705E-02	1.117E-02	3.198E-04	2.21
2	2.568931	1.921E-02	1.753E-02	7.841E-03	1.921E-04	.208
3	.738950	1.159E-03	1.004E-03	5.797E-04	1.431E-05	4.163E-0
4	.001482	7.684E-04	6.274E-04	4.436E-04	1.201E-05	3.83
5	.000169	7.445E-04	5.686E-04	4.806E-04	1.519E-05	41.6
6	.000155	7.200E-04	5.091E-04	5.091E-04	2.000E-05	42.5
7	.000131	7.013E-04	4.527E-04	5.356E-04	2.805E-05	45.6
8	.000121	6.811E-04	3.932E-04	5.561E-04	4.257E-05	46.7
9	.000115	6.542E-04	3.271E-04	5.665E-04	7.269E-05	47.5
10	.000094	6.379E-04	2.604E-04	5.823E-04	1.595E-04	52.8
11	.000086	6.231E-04	1.799E-04	5.966E-04	6.231E-04	59.5
12	.000078	-	-	-	-	-

By following methods, the number of primary eigenvectors are obtained :
  - Imbedded Error Function (IE)   =   4
  - Factor Indicator Function (IND)   =   4
  - Percent Significant Level (%SL)   =   4

By automatic discrimination :
The most possible number of primary eigenvectors   =   4

After judgement by user :
The number of primary eigenvectors   =   4

RMS of extractable error matrix   =   6.27389E-04

Norm of extractable error matrix =   3.08124E-02

```
*****************************************************************
*   MTFA v.1 - The Factor Analysis Programme for Chemical Analysis   *
*                                                               *
*   Created by Arunchai Tangcharoenbumrungsuk                   *
*   Conducted by Dr.Vudhichai Parasuk                           *
*****************************************************************

   1. Elementary Data
      - Data file name                    CuGlyH
      - Dimension of data matrix          134 x 18
      - Type of problem                   Metal-Ligand Complex Equilibrium
      - Optimization method               Simplex Method
      - Set the initial guess eigenvector Automatic
      - Stochastic Initialization Algor.  No

   2. Optimization Method
      - Number of steps of complex formation     =  3
      - The starting guess
        Value of parameter no. 1 =  1.00
        Value of parameter no. 2 = 1.000E-02
        Value of parameter no. 3 = 1.000E-05

      - The convergence are obtained
        Value of parameter no. 1 =  4.147
        Value of parameter no. 2 = 3.446E-02
        Value of parameter no. 3 = 4.415E-05

      - The chi-squared value      = 8.718E-04

      - Total optimization time  =   1 min. 17 s 44 cs

      - Number of iteration cycle = 180

   3. Value Setting
      - Gradient (determine the pricipal factor)   = 1.000E-12
      - Tolerance (simplex method)                 = 1.000E-20
      - Convergence (simplex method)               = 1.000E-16
      - Simplex parameters
        Reflection parameter   =  1.000   (default)
        Contraction parameter  =   .550   (default)
        Expansion parameter    =  2.900   (default)
        Reduction parameter    =   .500   (default)

   4. Value Input
      - Number of solutions    =   18
      - pH of solutions
        The pH of solution no. 1  =    1.470
        The pH of solution no. 2  =    1.700
        The pH of solution no. 3  =    1.980
        The pH of solution no. 4  =    2.250
        The pH of solution no. 5  =    2.570
        The pH of solution no. 6  =    2.950
        The pH of solution no. 7  =    3.240
        The pH of solution no. 8  =    3.990
        The pH of solution no. 9  =    4.080
        The pH of solution no.10  =    4.360
        The pH of solution no.11  =    4.480
        The pH of solution no.12  =    4.600
        The pH of solution no.13  =    4.770
        The pH of solution no.14  =    5.040
        The pH of solution no.15  =    5.560
        The pH of sulution no.16  =    6.040
        The pH of solution no.17  =    6.600
        The pH of solution no.18  =    6.960
```

- Total concentration of ligand   = .10011   mol/L
- Total concentration of metal   = 2.0440E-03   mol/L

- Number of dissociation step of the ligand =   2
    Dissociation constant no. 1 = 4.467E-03
    Dissociation constant no. 2 = 1.667E-10

- Model of complex formation
    step 1 :   1   1   0
    step 2 :   1   1   1
    step 3 :   1   2   2

- Order of ligand that form complex with metal = 2

5. Eigenvectors and Error Functions

Results of Error Treatment :

n	Eigenvalue	RSD	RMS	IE	IND	%SL
1	8.000366	1.372E-02	1.333E-02	3.233E-03	4.746E-05	6.929E-0
2	.415320	2.484E-03	2.342E-03	8.281E-04	9.705E-06	4.743E-0
3	.012884	4.170E-04	3.807E-04	1.702E-04	1.853E-06	6.154E-0
4	.000190	2.913E-04	2.569E-04	1.373E-04	1.486E-06	1.33
5	.000019	2.840E-04	2.414E-04	1.497E-04	1.681E-06	37.8
6	.000017	2.766E-04	2.259E-04	1.597E-04	1.921E-06	38.3
7	.000016	2.695E-04	2.107E-04	1.681E-04	2.228E-06	39.2
8	.000014	2.629E-04	1.959E-04	1.752E-04	2.629E-06	40.4
9	.000014	2.548E-04	1.802E-04	1.802E-04	3.145E-06	39.5
10	.000012	2.482E-04	1.655E-04	1.850E-04	3.878E-06	42.0
11	.000012	2.399E-04	1.496E-04	1.875E-04	4.895E-06	41.1
12	.000011	2.310E-04	1.333E-04	1.886E-04	6.415E-06	41.7
13	.000010	2.205E-04	1.162E-04	1.874E-04	8.821E-06	41.9
14	.000008	2.138E-04	1.008E-04	1.885E-04	1.336E-05	46.6
15	.000007	2.063E-04	8.424E-05	1.884E-04	2.293E-05	48.3
16	.000007	1.979E-04	6.597E-05	1.866E-04	4.948E-05	51.2
17	.000006	1.899E-04	4.476E-05	1.845E-04	1.899E-04	58.5
18	.000005	-	-	-	-	-

By following methods, the number of primary eigenvectors are obtained
    - Imbedded Error Function (IE)   =   4
    - Factor Indicator Function (IND)   =   4
    - Percent Significant Level (%SL)   =   4

By automatic discrimination :
The most possible number of primary eigenvectors =   4

After judgement by user :
The number of primary eigenvectors   =   4

RMS of extractable error matrix   =   2.56942E-04

Norm of extractable error matrix =   1.26190E-02

```
*********************************************************************
*   MTFA v.1 - The Factor Analysis Programme for Chemical Analysis  *
*                                                                   *
*   Created by Arunchai Tangcharoenbumrungsuk                       *
*   Conducted by Dr.Vudhichai Parasuk                               *
*********************************************************************

  1. Elementary Data
       - Data file name                    CuAlaH
       - Dimension of data matrix          201 x 13
       - Type of problem                   Metal-Ligand Complex Equilibrium
       - Optimization method               Simplex Method
       - Set the initial guess eigenvector Automatic
       - Stochastic Initialization Algor.  No

  2. Optimization Method
       - Number of steps of complex formation       =  3
       - The starting guess
       Value of parameter no. 1 =  1.00
       Value of parameter no. 2 = 1.000E-02
       Value of parameter no. 3 = 1.000E-05

       - The convergence are obtained
       Value of parameter no. 1 =  11.68
       Value of parameter no. 2 = 4.605E-02
       Value of parameter no. 3 = 3.753E-05

       - The chi-squared value      = 9.418E-04

       - Total optimization time  =   1 min.  5 s 74 cs

       - Number of iteration cycle = 184

  3. Value Setting
       - Gradient (determine the pricipal factor)   = 1.000E-12
       - Tolerance (simplex method)                 = 1.000E-20
       - Convergence (simplex method)               = 1.000E-16
       - Simplex parameters
          Reflection parameter   =  1.000   (default)
          Contraction parameter  =   .550   (default)
          Expansion parameter    =  2.900   (default)
          Reduction parameter    =   .500   (default)

  4. Value Input
       - Number of solutions   =   13
       - pH of solutions
          The pH of solution no. 1  =    1.500
          The pH of solution no. 2  =    1.980
          The pH of solution no. 3  =    2.610
          The pH of solution no. 4  =    2.990
          The pH of solution no. 5  =    3.540
          The pH of solution no. 6  =    4.020
          The pH of solution no. 7  =    4.590
          The pH of solution no. 8  =    4.840
          The pH of solution no. 9  =    5.510
          The pH of solution no.10  =    6.470
          The pH of solution no.11  =    6.130
          The pH of solution no.12  =    6.510
          The pH of solution no.13  =    7.530
```

- Total concentration of ligand = .10324   mol/L
- Total concentration of metal   = 2.0202E-03 mol/L

- Number of dissociation step of the ligand = 2
  Dissociation constant no. 1 = 4.365E-03
  Dissociation constant no. 2 = 1.659E-10

- Model of complex formation
    step 1 : 1   1   0
    step 2 : 1   1   1
    step 3 : 1   2   2

- Order of ligand that form complex with metal = 2

5. Eigenvectors and Error Functions

Results of Error Treatment :

n	Eigenvalue	RSD	RMS	IE	IND	%SL
1	11.874195	1.697E-02	1.630E-02	4.706E-03	1.178E-04	3.539E-0
2	.675174	2.948E-03	2.712E-03	1.156E-03	2.436E-05	2.836E-0
3	.018828	4.377E-04	3.839E-04	2.103E-04	4.377E-06	2.587E-0
4	.000229	2.934E-04	2.441E-04	1.628E-04	3.623E-06	3.12
5	.000024	2.861E-04	2.244E-04	1.774E-04	4.470E-06	42.0
6	.000021	2.803E-04	2.057E-04	1.904E-04	5.720E-06	44.4
7	.000020	2.738E-04	1.860E-04	2.009E-04	7.604E-06	44.8
8	.000019	2.670E-04	1.656E-04	2.095E-04	1.068E-05	45.8
9	.000017	2.606E-04	1.445E-04	2.168E-04	1.629E-05	47.6
10	.000015	2.561E-04	1.230E-04	2.246E-04	2.846E-05	50.7
11	.000014	2.499E-04	9.802E-05	2.299E-04	6.248E-05	52.9
12	.000014	2.382E-04	6.607E-05	2.289E-04	2.382E-04	58.1
13	.000011	-	-	-	-	-

By following methods, the number of primary eigenvectors are obtained :
  - Imbedded Error Function (IE)   =   4
  - Factor Indicator Function (IND)   =   4
  - Percent Significant Level (%SL)   =   4

By automatic discrimination :
The most possible number of primary eigenvectors =   4

After judgement by user :
The number of primary eigenvectors   =   4

RMS of extractable error matrix   =   2.44145E-04

Norm of extractable error matrix =   1.24801E-02

MTFA version 1.0

Program Manual

Program MTFA version 1.0 was developed to execute the factor analysis and modeling transformation. The source codes were written in fortran language using Microsoft FORTRAN V5.1 for personal computer. The main application here is to resolve the UV/VIS absorption spectra for the acid-base equilibria of the polyprotic organic acid-base pairs and also the formation of metal-ligand complexes. This compact manual provided step-by-step instructions, and demonstration with an example of input and/or output in an easy understanding.

Running the MTFA Program

1) If you have the source code of MTFA program, you should first compile the MTFA program by Microsoft FORTRAN V5.1. (see the compile procedures in the manual of Microsoft FORTRAN) In the other hand, if you have in form of execute file, *i.e.* MTFA.EXE, you can run the program by typing MTFA in the DOS promt as:

```
C:\MTFA
```

After pressing the ENTER key, the program should clear the screen and start to run the following factor analysis as:

```
----------------------------------------------------------------------------------
*  MTFA v.1 – The Factor Analysis Program for Chemical Analysis    *
*  Created by Arunchai Tangcharoenbumrungsuk                       *
*  Conducted by Dr. Vudhichai Parasuk                              *
----------------------------------------------------------------------------------
```

- Total concentration of ligand = .10324   mol/L
- Total concentration of metal = 2.0202E-03 mol/L

- Number of dissociation step of the ligand = 2
  Dissociation constant no. 1 = 4.365E-03
  Dissociation constant no. 2 = 1.659E-10

- Model of complex formation
     step 1 :   1   1   0
     step 2 :   1   1   1
     step 3 :   1   2   2

- Order of ligand that form complex with metal = 2

5. Eigenvectors and Error Functions

Results of Error Treatment :

n	Eigenvalue	RSD	RMS	IE	IND	%SL
1	11.874195	1.697E-02	1.630E-02	4.706E-03	1.178E-04	3.539E-0
2	.675174	2.948E-03	2.712E-03	1.156E-03	2.436E-05	2.836E-0
3	.018828	4.377E-04	3.839E-04	2.103E-04	4.377E-06	2.587E-0
4	.000229	2.934E-04	2.441E-04	1.628E-04	3.623E-06	3.12
5	.000024	2.861E-04	2.244E-04	1.774E-04	4.470E-06	42.0
6	.000021	2.803E-04	2.057E-04	1.904E-04	5.720E-06	44.4
7	.000020	2.738E-04	1.860E-04	2.009E-04	7.604E-06	44.8
8	.000019	2.670E-04	1.656E-04	2.095E-04	1.068E-05	45.8
9	.000017	2.606E-04	1.445E-04	2.168E-04	1.629E-05	47.6
10	.000015	2.561E-04	1.230E-04	2.246E-04	2.846E-05	50.7
11	.000014	2.499E-04	9.802E-05	2.299E-04	6.248E-05	52.9
12	.000014	2.382E-04	6.607E-05	2.289E-04	2.382E-04	58.1
13	.000011	-	-	-	-	-

By following methods, the number of primary eigenvectors are obtained
   - Imbedded Error Function (IE)   =   4
   - Factor Indicator Function (IND)   =   4
   - Percent Significant Level (%SL)   =   4

By automatic discrimination :
The most possible number of primary eigenvectors =   4

After judgement by user :
The number of primary eigenvectors   =   4

RMS of extractable error matrix =   2.44145E-04

Norm of extractable error matrix =   1.24801E-02

MTFA version 1.0

Program Manual

   Program MTFA version 1.0 was developed to execute the factor analysis and modeling transformation. The source codes were written in fortran language using Microsoft FORTRAN V5.1 for personal computer. The main application here is to resolve the UV/VIS absorption spectra for the acid-base equilibria of the polyprotic organic acid-base pairs and also the formation of metal-ligand complexes. This compact manual provided step-by-step instructions, and demonstration with an example of input and/or output in an easy understanding.

Running the MTFA Program

   1) If you have the source code of MTFA program, you should first compile the MTFA program by Microsoft FORTRAN V5.1. (see the compile procedures in the manual of Microsoft FORTRAN) In the other hand, if you have in form of execute file, *i.e.* MTFA.EXE, you can run the program by typing MTFA in the DOS promt as:

```
C:\MTFA
```

After pressing the ENTER key, the program should clear the screen and start to run the following factor analysis as:

```
--------------------------------------------------------------------------------
* MTFA v.1 – The Factor Analysis Program for Chemical Analysis   *
* Created by Arunchai Tangcharoenbumrungsuk                      *
* Conducted by Dr. Vudhichai Parasuk                             *
--------------------------------------------------------------------------------
```

2) Begin to input the data filename by **typing 8-character name** or less than such as:

```
Input Elementary Data:
      -   Please type the data filename: CuAlaH01
```

3) Input the dimension, *i.e.* the number of row and column, of data matrix.

```
      -   Please specify the dimension of data matrix (nrow,ncol): 201 13
```

If you don't known the dimension, you can count them by opening the data file by a text editor such as Microsoft Editor, Microsoft FORTRAN, or a word processing. It should be noted that the data file was created by a text editor such as Microsoft Editor, Microsoft FORTRAN or a word processing program. The format of data matrix is very simple where each row associated to the ith wavelength, and each column associated to the jth mixture. The example for data file was shown below.

0.012	0.013	0.013	0.012	0.013	0.012	0.013	0.014	0.015	0.013
0.011	0.012	0.013	0.012	0.013	0.012	0.014	0.015	0.016	0.014
0.011	0.012	0.013	0.012	0.014	0.016	0.019	0.021	0.023	0.021
0.011	0.012	0.013	0.013	0.015	0.018	0.023	0.025	0.028	0.027
0.011	0.012	0.013	0.014	0.018	0.024	0.032	0.035	0.039	0.038
0.012	0.012	0.015	0.017	0.025	0.036	0.050	0.054	0.060	0.060

....

....

4) Select a chemical problem from the list by pressing a number such as:

```
      List of problem types:
         (1)  Acid-Base Equilibrium
         (2)  Metal-Ligand Complex Equilibrium
      Please select one of the problem:   2
```

Noted that the program should protect you to input unreasonable choise such as "3", "4", etc.

5) Select an optimization method from the list by pressing a number such as:

List of the optimization methods:
   (1) Simplex Method
   (2) Simplex Method with Approximate Gradient (Weighted
       .Centroid Method)
   (3) Combination of Simplex and Approximate Gauss-Newton
       Methods v.1
   (4) Modified Simplex Method with Unidirectional Fibonacci
       Search
Please select one of the list: **1**

Noted that the program should protect you to input unreasonable choise such as "5", "6", etc.

6) Input the pH of each solution. Here there are 13 solutions.

- Hence the number of solutions = 13 Then specify the pH of solution: The pH of solution no. 1 = **1.50** The pH of solution no. 2 = **1.98** The pH of solution no. 3 = **2.61** The pH of solution no. 4 = **2.99** The pH of solution no. 5 = **3.54** The pH of solution no. 6 = **4.02** The pH of solution no. 7 = **4.59** The pH of solution no. 8 = **4.84** The pH of solution no. 9 = **5.51** The pH of solution no. 10 = **6.13** The pH of solution no. 11 = **6.47** The pH of solution no. 12 = **6.51** The pH of solution no. 13 = **7.53**

The user should know the pH of each solution along the experiment.

7) Since we chose the metal-ligand complex equilibrium, now we must input the total concentrations (in mol/L) of ligand and metal, and the dissociation constants (known from literatures) of ligand such as:

> - Please input the total concentration of ligand = **0.10324**
>
> - Please input the total concentration of metal = **0.0020202**
>
> - Please input the number of step of dissociation of the Ligand = **2**
>   Then specify the dissociation constants for each step.
>   Dissociation constant no. 1 = **4.365E-3**
>   Dissociation constant no. 2 = **1.659E-10**

Noted that the format descriptor "E" in FORTRAN are signified as an exponent e.g. 4.365E-3 means that 4.365×10^{-3}.

8) Arrive here you should specify three constants of "Gradient", "Tolerance" and "Convergence" as stated in Section 2.4.1, 2.4.4 and 3.1.1. For this example, we would like to use the default values such that Gradient $= 1.0 \times 10^{-8}$, Tolerance $= 1.0 \times 10^{-16}$, and Convergence $= 1.0 \times 10^{-14}$.

> In the process of determination of principal factor
> - Do you want to set the Gradient?
>   If "yes", please input the value.
>   If "no", please press <ENTER> to use the default value:
>
> In the process of optimization method
> - Do you want to set the Tolerance?
>   If "yes", please input the value.
>   If "no", please press <ENTER> to use the default value:
> - Do you want to set the Convergence?
>   If "yes", please input the value.
>   If "no", please press <ENTER> to use the default value:

9) Set the simplex parameters. (see detail in Section 2.4.4 Simplex Method)   Here we would like to use the default values such that Reflection, Contraction, Expansion and Reduction parameters are set to default as 1.0, 0.55, 2.9 and 0.5, respectively.

> - Would you like to set the Simplex parameters?
>   If "yes", please input the value.
>   If "no", please press <ENTER> to use the default value:

10) Choose an option of setting the initial guess eigenvector. Here the automatics are reccommended.

```
Log into Factor Analysis Programme:
Eigenvectors and Eigenvalues:
        Please select the option for setting the initial guess eigenvector
        (1)  Automatic setting
        (2)  Manual setting
        Enter the selection   1
```

At this stage, the program should calculate all eigenvectors and eigenvalues by the method of principal component analysis (PCA). The examples of output were shown below.

```
Number of Eigenvector  =  13
Eigenvalues:
    11.874195
    0.675174
    0.018828
    0.000229
    0.000024
    0.000021
    0.000020
    0.000019
    0.000017
    0.000015
    0.000014
    0.000014
    0.000011
```

11) In this stage, the program should calculate all error indicators, and automatically suggest the possible number of primary eigenvectors. On this example, the number of primary eigenvectors are equal to "4".

Results of Error Treatment:

n	Eigenvalue	RSD	RMS	IE	IND	%SL
1	11.874198	1.697E-02	1.630E-02	4.706E-03	1.178E-04	3.539E-05
2	0.675174	2.948E-03	2.712E-03	1.156E-03	2.436E-05	2.836E-06
3	0.018828	4.377E-04	3.839E-04	2.103E-04	4.377E-06	2.587E-06
4	0.000229	2.934E-04	2.441E-04	1.628E-04	3.623E-06	3.12
5	0.000024	2.861E-04	2.244E-04	1.774E-04	4.470E-06	42.0
6	0.000021	2.803E-04	2.057E-04	1.904E-04	5.720E-06	44.4
7	0.000020	2.738E-04	1.860E-04	2.009E-04	7.604E-06	44.8
8	0.000019	2.670E-04	1.656E-04	2.095E-04	1.068E-05	45.8

9	0.000017	2.606E-04	1.445E-04	2.168E-04	1.629E-05	47.6
10	0.000015	2.561E-04	1.230E-04	2.246E-04	2.846E-05	50.7
11	0.000014	2.499E-04	9.802E-05	2.299E-04	6.248E-05	52.9
12	0.000014	2.382E-04	6.607E-05	2.289E-04	2.382E-04	58.1
13	0.000011	-	-	-	-	-

- By computation, the estimated number of primary eigenvector = 4

Note: If the estimation = 0, it means that the data matrix may consist of unsufficient data or contain only pure error. The data, therefore, are not factor analyzable.

- Do you agree with this estimation?
  If "yes", please prees <y>.
  If "no", please input your estimation: **Y**

12) In this stage, you should input additional data, *i.e.* the number of complex formation (recommend to agree with the program estimation), the formation model, and the initial guess of equilibrium constants sucha as:

Given the Further Information:
  - By computation, the estimated number of steps of complex formation = 3

  - Do you agree with the estimation?
    If "yes", please press <y> or <Enter>.
    If "no", please input your estimation:

  - Please propose the model of formation: (M L H)
    step 1: **1 1 0**
    step 2: **1 1 1**
    step 3: **1 2 2**

  Then make the initial guess for all overall stability constants.
  - Do you like to use the Stochastic Initialization Algorithm or ordinary initial guess?
    If "yes", please prees <y>.
    If "no", please press <N> or <Enter>.

  Value of parameter no. 1 = **1.0**
  Value of parameter no. 2 = **1.0E-02**
  Value of parameter no. 3 = **1.0E-05**

13) In this step, you must wait for the program optimization.   When the program finish, the convergence equilibrium constants were obtained such as:

Data Manipulation and Predictions:
- The equilibrium constants are:
  11.68   4.605E-02   3.753E-05

- Estimated column matrix:
  0.001928 0.001846 0.001672 0.001405 0.000987
  ....

- Estimated row matrix:
  5.649   18.103   4.559   7.098
  ....

---

14) Finally, the user must select the options for output unit and printing by selecting a number from the list such as:

Results Printing:

- Would you like to print the summary of results?
  (1)   Printer
  (2)   File (.Log)
  (3)   Monitor
  (4)   None
  Please select one from the list: 3

- Option for printing
  (1)   Full information
  (2)   Only necessary information
  Please select one from the list: 2

CURRENT VITA

ARUNCHAI TANGCHAROENBUMRUNGSUK

Date of Birth: December 15, 1969

Birth Place: Petchaburi, Thailand

Marital Status: Single

Current Address:

68 Pongsuriya road,

Amper Muang,

Petchaburi 76000

Phone: (032)-426010

Education:

Master degree student, Chemistry, October 1999, Chulalongkorn University, Bangkok.

Bachelor of Science, Chemistry, May 1993, Mahidol University, Bangkok.

Professional Experience:

11/95-04/96   *Chemist*, Sithiporn Associates Co., Ltd. 101 Sirinthorn, Bangplud, Bangkok.

05/93-05/95   *Chemistry Lecturer*, Rajabhat Institute Suan Dusit, Ratchasrima road, Dusit, Bangkok.