Journal of Mathematical Chemistry, Vol. 43, No. 4, May 2008 (© 2007)
DOI: 10.1007/s10910-007-9260-8

Existence, uniqueness, and calculation of the physically acceptable solution of a particular case of the Sherman equations

Víctor Delgado*

Física Médica,Departamento de Radiología, Universidad Complutense, 28040 Madrid, Spain
E-mail: vdelgado@med.ucm.es

Rodolfo Figueroa

Departamento de Ciencias Fisicas, Universidad de la frontera, Av. Francisco Salazar 01145,
Casilla Postal 54-D Temuco, Chile

Received 2 April 2007; Revised 17 April 2007

When a chemical sample made of N elements is analyzed by using sequential selective excitation by monochromatic X-ray beams and selective measurement of the characteristic X-rays, the production of secondary fluorescence does not interfere with the measurements. This experimental situation leads to a particular simple case of the Sherman equations which can be written in this instance as linear equations. The linear equations thus obtained are shown to be very similar to the equations appearing in the classical models of Beattie and Brissey and of Lachance and Traill. The linear algebra proves the existence of N different sets of solutions, but the Perron Frobenius theorem ensures that there is one and only one physically feasible solution, and also leads to the method for obtaining it. This equation solution method can be extended to the equations appearing when standard samples of pure elements are also measured.The propagation of the errors in the measurements to the errors in the sample concentrations has been calculated and simulated, and the results have shown that the solution is well conditioned.

KEY WORDS: x-ray fluorescence analysis, Sherman equation, fundamental parameters method, Perron theorem, iterative methods, Beattie and Brissey equations, Lachance and Traill equations

1. Introduction

Let us consider the following experimental situation of selective excitation and selective measurement: we have a thick sample made of N chemical elements which is excited at 45° by a tunable monoenergetic photon source of known intensity, and the fluorescence produced is measured also at 45° with a detector

*Corresponding author.

of known relative efficiency. When the sample is excited sequentially at N different energies, each one over the K edge of each element but under the K edge of the following element, and we measure the K alpha fluorescence of such element each time, there is no contribution of secondary fluorescence to the measurements, and the Sherman equations can thus be written as:

$$G_i = \frac{\varepsilon\varepsilon_{ri}}{\sqrt{2}} C_i \frac{I_i \sigma_i\left(h\nu_i\right)}{\sum_{j=1}^{j=n} \mu_j\left(h\nu_i\right) C_j + \sum_{j=1}^{j=n} \mu_j (h\nu_{E_{K\alpha i}}) C_j}, \tag{1}$$

$$\sum_{j=1}^{j=n} C_j = 1, \tag{2}$$

$$C_i > 0 \ \forall i,$$
$$\varepsilon > 0,$$

where G_i is the K-fluorescence of element i; C_i is the concentration of element i; ε is the unknown global efficiency and ε_{ri} is the relative efficiency at the energy of the characteristic radiation of the element i; I_i is the intensity of the exciting beam, $\sigma_i\left(h\nu_i\right)$ is the fluorescence production cross-section of element i at its corresponding excitation energy $h\nu_i$; $n = N$ is the number of elements; $\mu_j\left(h\nu_i\right)$ is the attenuation coefficient of element j at its excitation energy $h\nu_i$, and $\mu_j\left(h\nu_{E_{K\alpha i}}\right)$ is the attenuation coefficient of element j at the energy of the characteristic radiation of element i, i.e. $E_{K\alpha i}$. Equation 2 is the condition of normalization of the concentrations: $\|C\|_1 = 1$. These concentrations and the global efficiency have to be positive.

The above experimental situation can be achieved by monochromatising a synchrotron beam or a intense X-ray beam as proposed by Figueroa [1], but using now selective counting instead of integral counting as Figueroa used.

These equations can be written as a linear system similar to the one proposed by Beattie and Brissey [2].

By defining $a_i = \sigma_i\left(h\nu_i\right)$, and $a_{i,j} = \mu_j\left(h\nu_i\right) + \mu_j\left(h\nu_{E_{K\alpha i}}\right)$, the equation (1) can be written as:

$$F_i = K C_i \frac{a_i}{\sum_{j=1}^{j=n} a_{i,j} C_j}, \tag{3}$$

where K is $\frac{\varepsilon}{\sqrt{2}}$ and F_i is G_i / I_i. By rearranging equation (3) we get:

$$\frac{F_i}{a_i} \sum_{j=1}^{j=n} a_{i,j} C_j - K\, C_i = 0, \tag{4}$$

$$\sum_{j=1}^{j=n} C_j = 1$$

or in matricial form:

$$\begin{pmatrix} (F_1\frac{a_{1,1}}{a_1} - K) & F_1\frac{a_{1,2}}{a_1} & .. & F_1\frac{a_{1,n}}{a_1} \\ F_2\frac{a_{2,1}}{a_2} & (F_2\frac{a_{2,2}}{a_2} - K) & .. & . \\ . & . & .. & . \\ . & . & .. & . \\ F_n\frac{a_{n,1}}{a_n} & . & .. & (F_n\frac{a_{n,n}}{a_n} - K) \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ . \\ . \\ C_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ 0 \end{pmatrix}, \tag{5}$$

$$\sum_{j=1}^{j=n} C_j = 1$$

and by rearranging the above equation, we can write it in a form similar to that used by Beattie and Brissey:

$$\begin{pmatrix} (1 - K\frac{a_1}{F_1\, a_{1,1}}) & \frac{a_{1,2}}{a_{1,1}} & .. & \frac{a_{1,n}}{a_{1,1}} \\ \frac{a_{2,1}}{a_{2,2}} & (1 - K\frac{a_2}{F_2\, a_{2,2}}) & .. & . \\ . & . & .. & . \\ . & . & .. & . \\ \frac{a_{n,1}}{a_{n,n}} & . & .. & (1 - K\frac{a_n}{F_n\, a_{n,n}}) \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ . \\ . \\ C_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ 0 \end{pmatrix}, \tag{6}$$

$$\sum_{j=1}^{j=n} C_j = 1.$$

The equations can be also written in a similar way to that proposed by Lachance and Traill [3]:

By using equation (2) in the denominator of equation (3) we get:

$$F_i = C_i K \frac{a_i}{\sum_{j \neq i} a_{i,j} C_j + a_{ii}(1 - \sum_{j \neq i} C_j)},$$

$$\sum_{j=1}^{j=n} C_j = 1$$

and rearranging the latter equation we can write it in a form similar to that used by Lachance and Traill:

$$\frac{1}{K} F_i \frac{a_{ii}}{a_i} = \frac{C_i}{1 + \sum_{j \neq i}(\frac{a_{i,j}}{a_{ii}} - 1)C_j}, \tag{7}$$

$$\sum_{j=1}^{j=n} C_j = 1.$$

2. Solution of the equations

2.1. Solution of the equations when written in a form similar to that proposed by Beattie and Brissey: the Perron–Frobenius theorem and the von Mises power method

Let us write equation (5) in a shorter way by defining $A_{i,j} = F_i \frac{a_{i,j}}{a_i}$:

$$\begin{pmatrix} (A_{1,1} - K) & A_{1,2} & .. & A_{1,n} \\ A_{2,1} & (A_{2,2} - K) & .. & . \\ . & . & .. & . \\ . & . & .. & . \\ A_{n,1} & . & .. & (A_{n,n} - K) \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ . \\ . \\ C_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ 0 \end{pmatrix}, \qquad (8)$$

$$\sum_{j=1}^{j=n} C_j = 1.$$

Here we have a system of $N+1$ equations and $N+1$ unknowns: C_1 to C_n and K. This system is linear in C_i and non-linear in K. The first N equations form a linear homogeneous system, which in general has only the trivial solution, that is, $C_i = 0 \; \forall \; i$. The condition for the existence of non-trivial solutions is that the determinant of the matrix of coefficients be 0, then, there are infinite solutions , in fact, a straight line of solutions. The intersection of this straight line of solutions with the equation which expresses the condition of normalization, equation (2), gives us a unique solution. So, when looking for non-trivial solutions, we have to impose that the determinant of equation (8) be zero, and this condition provides N solutions for K, and, from each value of K, a set of concentrations C_n will be obtained.

Let us see it in detail: equation (8) can be rewritten as:

$$\begin{pmatrix} A_{1,1} & A_{1,2} & .. & A_{1,n} \\ A_{2,1} & A_{2,2} & .. & . \\ . & . & .. & . \\ . & . & .. & . \\ A_{n,1} & . & .. & A_{n,n} \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ . \\ . \\ C_n \end{pmatrix} = K \begin{pmatrix} C_1 \\ C_2 \\ . \\ . \\ C_n \end{pmatrix},$$

which is an eigenvalue problem. The values of K that provide non-trivial solutions are the eigenvalues of matrix A whose elements-are $A_{i,j}$ and the non trivial solutions are the eigenvectors [normalized to 1 according to equation(2)] of matrix A. The eigenvalues are the roots of the characteristic polynomial, which is a polynomial of degree N in K coming from imposing that the determinant of the matrix of coefficients be equal to 0. The fundamental theorem of algebra ensures that a polynomial of degree N has N roots, and, if the coefficients of the polynomial are real, the roots are real or complex conjugated numbers.

Up to now we have solved the system of equations formally, but the situation is far from satisfactory, since we have no *a priori* guarantee of the existence and uniqueness of physically acceptable solutions. We need to calculate all the eigenvalues and eigenvectors of a matrix (which is a numerically complicated task), and we have to look for positive eigenvalues and eigenvectors hoping that there be one and only one positive solution. We have performed numerical experiments with ternary, quaternary and even quinary samples, and we have always found one and only one positive eigenvalue. Thus, we hoped that by looking through the eigenvalue problem throughly, the existence and uniqueness of a positive solution would be proven, and so has happened.

The elements $A_{i,j}$ of matrix A are always positive, then matrix A belongs to the kind of matrices called positive matrices. There are strong theorems about the eigenvalues and eigenvectors of positive matrices. In particular, the Perron–Frobenius theorem [4] proves the following statements:

If A is a $n \times n$ positive matrix, there is one and only one positive eigenvalue (called the Perron eigenvalue) such that its norm is the greatest of the norms of all eigenvalues.

The Perron eigenvector – corresponding to the Perron eigenvalue – is the only eigenvector with all its elements positive and their sum equal to 1.

Then, the Perron–Frobenius theorem ensures the existence and uniqueness of a solution which fulfills the physical requirement of positivity and normalization. Moreover, it also provides us with an efficient algorithm for obtaining the values of K and C_i: The Perron theorem asserts that the Perron eigenvalue is the dominant eigenvalue, since it is the eigenvalue with greater norm, and the calculation of the dominant eigenvalue of a matrix and its corresponding eigenvector can be accomplished by the method of repeated multiplication from von Mises [5, 6], which can be implemented as follows:

Step 1: Let us take an unnormalized positive vector $C_{un}0$ (for example $C_{un}0_i = F_i$) and normalize it. Call the result $C0$, and take $K0 = \sum_{j=1}^{j=n} C_{un}0_j$ as in:

$$C0_i = \frac{C_{un}0_i}{K0}.$$

Step 2: Multiply $C0$ once by matrix A to get a new vector, $C_{un}1$. Normalize the result and obtain $K1$ and $C1$:

$$C_{un}1_i = \sum_{j=1}^{j=n} A_{i,j} C0_j, \qquad K1 = \sum_{j=1}^{j=n} C_{un}1_j, \qquad C1_i = \frac{C_{un}1_i}{K1}.$$

Step 3: Multiply $C1$ once again by A to get a new vector. Normalize the result and obtain $C2$ and $K2$:

$$C_{un}2_i = \sum_{j=1}^{j=n} A_{i,j}\, C1_j, \qquad K2 = \sum_{j=1}^{j=n} C_{un}2_j, \qquad C2_i = \frac{C_{un}2_i}{K2}.$$

Repeat m times:

The von Mises method ensures that $C = \lim_{m\to\infty} Cm$ is the Perron eigenvector of matrix A, and that $K = \lim_{m\to\infty} \sum_{j=1}^{j=n} C_{un}m_j$ is the Perron eigenvalue. The successive iterations converge promptly and Cm, when m is large enough, is a good approximation to the Perron eigenvector, while Km is also a good approximation to the Perron eigenvalue K. At iteration m the relative error in the estimate of the dominant eigenvalue K is of the order of $\left\| \frac{\lambda_2}{K} \right\|^m$, where λ_2 is the eigenvector of greater norm among the non-dominant eigenvectors.

The normalized eigenvector C is the unique physically acceptable solution and K is the global efficiency.

The need of taking a vector with all its terms positive as the first term of the iteration should be pointed out. Let us explain this concept: since matrix A is positive, each intermediate result, Cm_i, is positive too, and the iterative multiplication method converges to the unique physically acceptable solution. If there were negative elements in the seed, the iterative method could diverge, so, mathematically, we have to impose the condition that F_i, – the production of fluorescence for each element – be greater than zero, which is a natural physical constraint to the results of our fluorescence measurements. The iteration can also be performed without normalizing the $C_{un}m_i$at each step [7], but computational problems of underflow or overflow are likely to appear.

2.2. Solution of the equations when written in a form similar to that used by Lachance and Traill: an iterative method

Equations (7) can be written as:

$$C_i = \frac{1}{K}\frac{F_i a_{ii}}{a_i}\left(1 + \sum_{j\neq i}(\frac{a_{i,j}}{a_{ii}} - 1)C_j\right), \tag{9}$$

$$\sum_{j=1}^{j=n} C_j = 1.$$

The first n linear equations can be solved by standard algebraic methods and their solution are the concentrations C_j as a function of the global efficiency

$K, C_j(K)$. Concentrations $C_j(K)$ can be substituted for C_j in the last equation, thus obtaining:

$$\sum_{j=1}^{j=n} C_i - 1 = 0. \tag{10}$$

When the value of n is small, it is quite straightforward to verify, by direct algebraic calculation, that this last equation, which is a polynomial of degree n in K, is the characteristic polynomial of matrix A. The direct verification is unfeasible for large n, but the system of equations (7) is obtained by linear operations from the system of equations (4), so, both systems are equivalent and have the same solutions, and 10 is, in fact, the characteristic polynomial of matrix A.

These equations in C_i and K can be solved in the following fashion: the solution for K is the greater root of the polynomial in K 10. This root can be obtained by using standard numerical methods, where the values of $C_j(K)$ for each concrete value of K needed in the search for the root can be obtained by solving the complete linear equations (9) by standard algebraic methods. When the solution for K is found, it is introduced in equations (9) and the values for C_i are obtained by solving the linear system 9. This is a solid approach, but it is very costly computationally.

However equation (9) can be written as:

$$\frac{F_i a_{i,i}}{a_i} \left(1 + \sum_{j \neq i} \left(\frac{a_{i,j}}{a_{i,i}} - 1 \right) C_j \right) = K C_i, \tag{11}$$

$$\sum_{j=1}^{j=n} C_j = 1$$

and it becomes apparent that we are dealing with a linear eigenvalue problem. This eigenvalue problem has exactly the same solution as the eigenvalue problem which has been discussed above (in section II A), because the equations (4) and (7) are equivalent. In fact, equation (11) can be written as:

$$\frac{F_i}{a_i} \sum_{j=1}^{j=n} a_{i,j} C_j = K\, C_i,$$

$$\sum_{j=1}^{j=n} C_j = 1,$$

which is exactly the same as equation (4). So we have the guarantee that there is one and only one positive dominant eigenvalue with an associated positive normalized eigenvector, and the von Mises iterative method for calculating the dominant eigenvalue of a linear eigenvalue problem can be applied now in the following way:

Step 1: Let us take a positive unnormalized vector $C_{un}0$ (for example $C_{un}0_i = F_i$) and normalize it. Call the result $C0$, and take $K0 = \sum_{j=1}^{j=n} C_{un}0_j$:

$$C0_i = \frac{C_{un}0_i}{K0}.$$

Step 2: Substitute $C0$ for C in the right-hand term of equation (9) to get $C_{un}1$. Normalize the result and obtain $K1$ and $C1$:

$$C_{un}1_i = \frac{F_i\, a_{i,i}}{a_i}\left(1 + \sum_{j\neq i}\left(\frac{a_{i,j}}{a_{i,i}} - 1\right) C0_j\right), \quad K1 = \sum_{j=1}^{j=n} C_{un}1_j, \quad C1_i = \frac{C_{un}1_i}{K1}.$$

Step 3: Substitute $C1$ for C and $K1$ for K in the right-hand term of equation (9), to get $C_{un}2$. Normalize the result and obtain $K2$ and $C2$:

$$C_{un}2_i = \frac{F_i\, a_{i,i}}{a_i}\left(1 + \sum_{j\neq i}\left(\frac{a_{i,j}}{a_{i,i}} - 1\right) C1_j\right), \quad K2 = \sum_{j=1}^{j=n} C_{un}2_j, \quad C2_i = \frac{C_{un}2_i}{K2}.$$

Repeat m times.

The von Mises method ensures that: $\lim_{m\to\infty} Cm = C$ is the dominant eigenvector of the linear operator, and that $\lim_{m\to\infty} \sum_{j=1}^{j=n} C_{un}m_j = K$ is the corresponding eigenvalue. The successive iterations converge promptly and Cm when m is large enough, is a good approximation to the positive eigenvector, while Km is also a good approximation to the dominant eigenvalue K.

Thus, the normalized eigenvector C is the unique physically acceptable solution, and K is the global efficiency.

The need of taking a vector with all their terms positive as the firs term of the iteration should be also pointed out. The seed and all intermediate steps have to be now normalized in order to avoid that term $a_{i,i}(1+\sum_{j\neq i}(\frac{a_{i,j}}{a_{i,i}} - 1)C_j)$ take negative values. By normalizing $C_{un}m_j$at each step we also avoid computational problems of underflow or overflow.

3. Error calculation

3.1. Analytical bounds

Equations (1) and (2)

$$F_i = K\,C_i \frac{a_i}{\sum_{j=1}^{j=n} a_{i,j}\,C_j},$$

$$\sum_{j=1}^{j=n} C_j = 1,$$

$$C_i > 0\ \forall i,$$
$$K > 0$$

can be regarded as defining the fluorescence values F_i as a function of the concentrations C_i and K. Since the values of $a_{i,j}$ and C_i are positive, the function is continuous and derivable, so, we can calculate the differential as a function of the partial derivatives:

$$\mathrm{d}F_i = \sum_{j=1}^{j=n} \frac{\partial}{\partial C_j}\left(K C_j \frac{a_i}{\sum_{l=1}^{l=n} a_{i,j} C_j} \right)\mathrm{d}C_j + \frac{\partial}{\partial K}\left(K C_i \frac{a_i}{\sum_{j=1}^{j=n} a_{i,j} C_j} \right)\mathrm{d}K,$$

$$\sum_{j=1}^{j=n} \mathrm{d}C_j = 0.$$

We have $N+1$ linear equations with $N+1$ unknowns, which can be solved in order to obtain the values of $\mathrm{d}C_i$ and $\mathrm{d}K$ as a function of $\mathrm{d}F_i$ (theorem of the implicit function). This ensures a continuous dependence on the errors in C_i and K as a function of the errors in F_i, and formally allows us to obtain the uncertainties in C_i and K as a function of the uncertainties in the measured F_i.The practical situation is, again, uncomfortable, because the actual expressions, even in a three component sample, are very cumbersome, but the propagation of relative errors can be dealt with in a rather simple way and simple results can be obtained for majoritary and minoritary elements.

Let us see first the general behavior of the errors of the minoritary and majoritary elements associated only to the positivity and normalization properties of the concentrations: since the error propagation is continuous, when the absolute errors in F_i are small and symmetrical, the absolute errors in C_i are small an symmetrical as well. Then the following results are obtained:

First: since the solutions are always positive, this implies that $C_i \pm \Delta C_i, > 0$ $\forall\ i \Rightarrow |\Delta C_i| < C_i \forall\ i$ and then $\Delta C_i \to 0$ if $C_i \to 0$, and the relative error of any concentration is less than 100%.

Second: we have that $\sum_{j=1}^{j=n} C_j = 1$, which implies that $\sum_{j=1}^{j=n} \Delta C_j = 0$. Suppose that there is a majoritary element, i.e., element number 1, such that $C_1 \to 1$; let us now calculate its relative error:

$$\frac{\Delta C_1}{C_1} = \frac{-\sum_{j=2}^{j=n} \Delta C_j}{C_1} = \frac{-\sum_{j=2}^{j=n} \Delta C_j}{1 - \sum_{j=2}^{j=n} C_j}$$

and, since ΔC_j and $C_j \to 0$ as $C_1 \to 1$, this implies that $\frac{\Delta C_1}{C_1} \to 0$ when $C_1 \to 1$.

These two previous results about the errors in minoritary and majoritary elements are simply associated to the continuous propagation of errors and to the conditions of normalization and positivity of the concentrations, and they are true as well as the solution algorithm produces positive and normalized values of concentrations. They are also independent of the actual values of relative uncertainties in F_i (provided that they are small).

In order to go a step further, we will analyze the case of a majoritary element and $N - 1$ minoritary elements. By taking the logarithmic derivative in Eqs. 3 we get:

$$C_i = \frac{1}{K}\frac{F_i}{a_i}\sum_{j=1}^{j=n} a_{ij} C_j,$$

$$\frac{\Delta C_i}{C_i} = -\frac{\Delta K}{K} + \frac{\Delta F_i}{F_i} + \frac{\sum_{j=1}^{j=n} a_{ij} \Delta C_j}{\sum_{j=1}^{j=n} a_{ij} C_j}. \tag{12}$$

Let us take element 1 as the majoritary element, then, when $C_1 \to 1$ the values of $\Delta C_1 / C_1$, C_j and ΔC_j all go to zero and the previous equation is simplified to:

$$\frac{\Delta F_1}{F_1} = \frac{\Delta K}{K}$$

that is to say, the relative error of the global efficiency K is equal to the relative error of the fluorescence produced by the majoritary element.

The rest of the elements are minoritary, so, $C_i \to 0 \,\forall i \neq 1$ and the equation (12) is simplified to:

$$\frac{\Delta C_i}{C_i} = -\frac{\Delta K}{K} + \frac{\Delta F_i}{F_i}.$$

The relative error of the minoritary elements is bounded by the sum of the relative error of their fluorescence F_i and the relative error of K.

In the general case, considering that $C_i = 1 - \sum_{j \neq i} C_j$, the equation (12) can be written as:

$$\frac{\Delta C_i}{C_i} = -\frac{\Delta K}{K} + \frac{\Delta F_i}{F_i} + \frac{\sum_{j \neq i} \left(\frac{a_{i,j}}{a_{ii}} - 1\right) \Delta C_j}{1 + \sum_{j \neq i} \left(\frac{a_{i,j}}{a_{ii}} - 1\right) C_j}.$$

Then, we can deduce that the relative error of the element i is bounded as follows:

$$\left|\frac{\Delta C_i}{C_i}\right| < \left|\frac{\Delta K}{K}\right| + \left|\frac{\Delta F_i}{F_i}\right| + \frac{\sum_{j \neq i} \left|(\frac{a_{i,j}}{a_{ii}} - 1) \Delta C_j\right|}{1 + \sum_{j \neq i} (\frac{a_{i,j}}{a_{ii}} - 1) C_j} < \left|\frac{\Delta K}{K}\right| + \left|\frac{\Delta F_i}{F_i}\right| + \sum_{j \neq i} \left|\frac{\Delta C_j}{C_j}\right|$$
$$< N \left|\frac{\Delta K}{K}\right| + \sum_{j=1}^{j=n} \left|\frac{\Delta F_i}{F_i}\right|.$$

This is a rather crude bound, and it is never approached in practice, as can be seen in the numerical simulation made.

3.2. *Numerical simulation*

We have simulated the analysis of ternary samples made of Cr, Fe, and Ni where the interelement effects by absorption are intense. The simulation has been done by calculating the fluorescence production in ternary samples with concentrations taken at random. The values of the fluorescence cross sections and attenuation coefficients have been taken from [8, 9], and the random concentration are generated in this way: for each element we select a random number uniformly distributed between zero and one, and the concentrations are calculated normalizing these random numbers by making their sum equal to one.

From these random concentrations the fluorescence production for each element has been calculated, and it has been varied by adding or subtracting a fixed percentage of the production at random: 0.1% for Cr, 0.5% for Fe, and 0.9% for Ni. From these varied fluorescence productions the concentrations of Cr, Fe, and Ni, and K have been calculated by the von Mises method and they have been compared with their true value.

The results are presented in figure 1, where each graph represents the relative error of K and of each concentration C_i as a function of each concentration C_j. The number of simulated points are over 10,000.

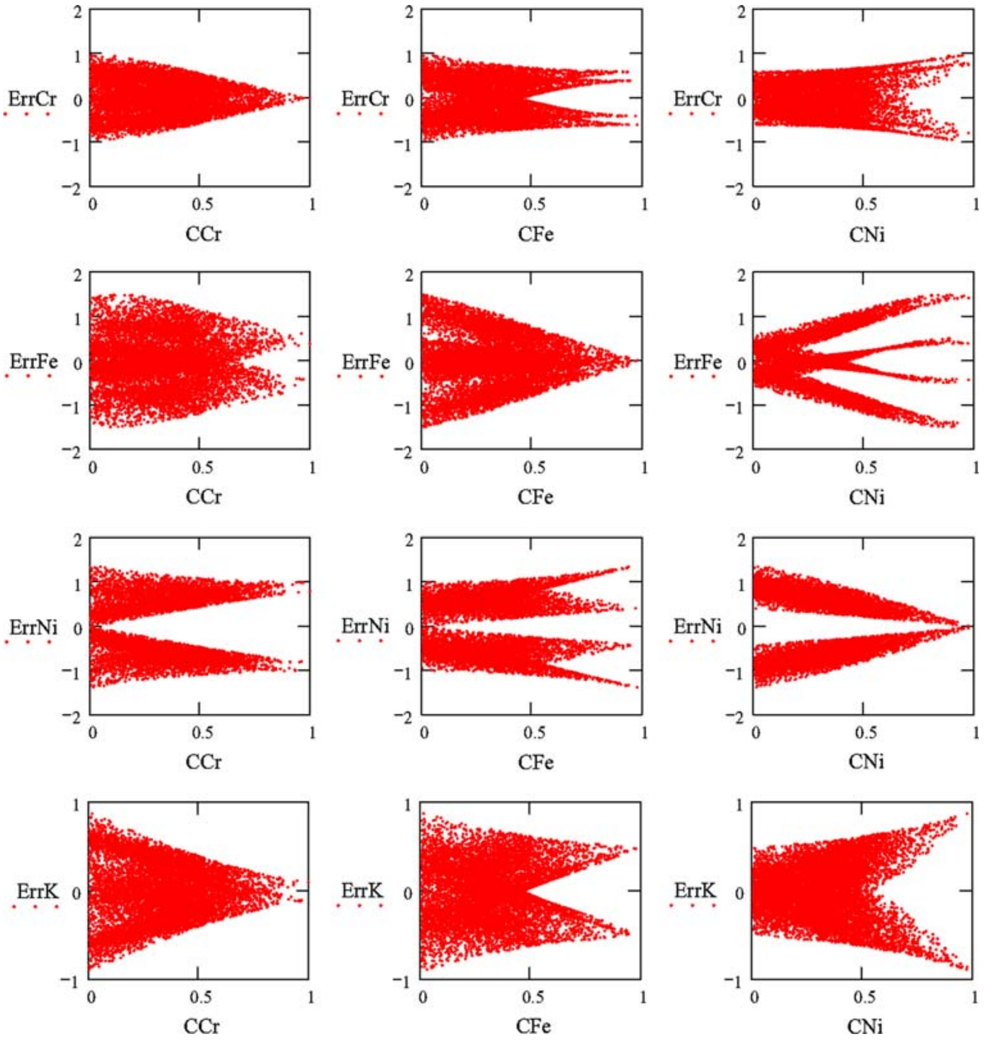

Figure 1. Monte Carlo simulated errors in the concentrations of Cr, Fe and Ni and in the eignen value K in function of the concentrations of Cr, Fe and Ni. The simulated errors have been obtained by assuming errors of 0.1% in the measurement of Cr, of 0.5% in the measurement of Fe and of 0.9% in the measurement of Ni. The errors are measured in %

4. The importance of positivity and normalization in the calculation of the physically acceptable solution when the global efficiency is determined by measurements

We have discussed the importance of the positivity and normalization of the seed used in the iterative methods. Now we will sketch the importance of such conditions when the value of K is determined by measurements, which corresponds to more realistic experimental situations.

Consider that an experimental value of K has been obtained by direct measurement of the global efficiency or by using an standard sample made of a pure element.

When the value of K is known, the equations (1) and (2) become a system of $N+1$ equations with N unknowns without solution due to the presence of experimental errors, so, we have to seek for approximate solutions.

We will present now the results of numerical simulations of some solution methods made with ternary samples of Cr, Fe, and Ni.

First method (Beatie and Brissey): when looking at the equations written in the form similar to that of Beattie and Brissey 6 the first N equations are now "almost dependent", so, we can drop any of them and solve the resulting system composed by the remaining $N-1$ equations and the equation of normalization, thus obtaining an approximated set of solutions for the concentrations. If this process is repeated for each one of the N first equations, N sets of solutions are obtained, and the mean and even the standard deviation for each C_i can be calculated.

This approach works well for small errors (a small percentage) in the data and intermediate values of concentrations. The values of the concentrations are always normalized because the condition of normalization is always included in the linear system solved, but when the concentrations are small, the linear system is ill conditioned, the propagation of errors is very bad, and sometimes negative concentration values appear. The solution method guarantees the normalization but does not guarantees positivity of the concentrations, so the method is not acceptable.

Second method (Lachance and Traill): by considering the equations written in the form similar to that of Lachance and Traill 7 we found that the first N equations have an independent term and can be solved. The solution is not normalized because we have dropped the condition of normalization, but, after obtaining the solution, it can be normalized, and thus a normalized set of concentrations is obtained. This method produces positive and normalized results for small errors in the data and for any concentration values with ternary samples made of Cr, Fe, and Ni, but it is sensitive to the errors in K, even producing negative solutions when the error of K reaches 80%. If we look at equation (7) we see that the positivity is not guaranteed because the term $1+\sum_{j\neq i}(\frac{a_{i,j}}{a_{ii}}-1)C_j$ can take negative values since C_j, which are unnormalized, can take values greater than 1. The method may seem practically acceptable, but it does not guarantee the positivity, and with samples of more components, the sensitivity to errors in K increases because the condition of normalization 10 is a polynomial in K of degree N, the latter being the number of components in the sample.

Third method (positive and normalized iteration): we can write the $N+1$ equations(3) as:

$$C_i = \frac{F_i}{K\,a_i}\sum_{j=1}^{j=n} a_{i,j}\,C_j, \tag{13}$$

$$\sum_{j=1}^{j=n} C_j = 1$$

and perform the following iterative procedure which preserves positivity and normalization in each step:

Step 1: Let us take a positive unnormalized vector $C_{un}0$ (for example $C_{un}0_i = F_i$) and normalize it. Call the result $C0$, and calculate the normalization factor of the seed $N0$: $N0 = \sum_{j=1}^{j=n} C_{un}0_j$:

$$C0_i = \frac{C_{un}0_i}{N0}.$$

Step 2 : Substitute $C0$ for C in the right-hand term of equation(13) to get $C_{un}1$. Normalize the result and obtain $N1$ and $C1$:

$$C_{un}1_i = \frac{F_i}{K\,a_i}\sum_{j=1}^{j=n} a_{i,j}\,C0_j =, \qquad N1 = \sum_{j=1}^{j=n} C_{un}1_j, \qquad C1_i = \frac{C_{un}1_i}{N1}.$$

Step 3 : Substitute $C1$ for C and $K1$ for K in the right-hand term of 13 to get $C_{un}2$. Normalize the result and obtain $N2$ and $C2$:

$$C_{un}2_i = \frac{F_i}{K\,a_i}\sum_{j=1}^{j=n} a_{i,j}\,C1_j =, \qquad N2 = \sum_{j=1}^{j=n} C_{un}2_j, \qquad C2_i = \frac{C_{un}2_i}{N2}.$$

Repeat m times.

This iterative procedure is exactly the same as that of von Mises because matrix $A_{i,j}$ has simply been scaled by dividing it by the value of K, so, the new matrix has an eigenvalue which is 1. If the experimental value of K is exact, Nm goes to 1. When K is affected by some error, the value of K converges to a number different from 1, but the values of concentrations do not change. This method of solution is insensitive to the errors in K.

Final remark (measurement of N standards): a final consideration must be done with respect to the use of standards. If we measure N standards made of the same pure elements which compose the sample analyzed and said sample is excited in the same conditions, we have the experimental advantage of avoiding the need to know the exciting intensities I_i and the relative efficiency ε_{ri}.The equations for the sample and the standard will read:

$$G_i = \frac{\varepsilon\varepsilon_{ri}}{\sqrt{2}} C_i \frac{I_i \sigma_i\,(h\nu_i)}{\sum_{j=1}^{j=n} \mu_j\,(h\nu_i)\; C_j + \sum_{j=1}^{j=n} \mu_j (h\nu_{E_{K\alpha i}})\, C_j}, \tag{14}$$

$$\sum_{j=1}^{j=n} C_j = 1$$

for the sample and

$$S_i = \frac{\varepsilon\varepsilon_{ri}}{\sqrt{2}} \frac{I_i \sigma_i\,(h\nu_i)}{\mu_i\,(h\nu_i) + \mu_i (h\nu_{E_{K\alpha i}})} \tag{15}$$

for the standard made of the pure element i.

Dividing equation(15) by the equation(14) and finding the value of G_i we get:

$$G_i = C_i \frac{S_i\left(\mu_i\,(h\nu_i) + \mu_i (h\nu_{E_{K\alpha i}})\right)}{\sum_{j=1}^{j=n} \mu_j\,(h\nu_i)\, C_j + \sum_{j=1}^{j=n} \mu_j (h\nu_{E_{K\alpha i}}) C_j} = C_i \frac{S_i a_{i,i}}{\sum_{j=1}^{j=n} a_{i,j} C_j},$$

which can be written as:

$$C_i = \frac{G_i}{S_i} \sum_{j=1}^{j=n} \frac{a_{i,j}}{a_{i,i}} C_j = r_i \sum_{j=1}^{j=n} \frac{a_{i,j}}{a_{i,i}} C_j,$$

$$\sum_{j=1}^{j=n} C_j = 1,$$

where the ratio r_i is defined as: $r_i = G_i / S_i$. In order to solve this system of equations we can consider the following eigenvalue problem:

$$\lambda C_i = \sum_{j=1}^{j=n} S_{i,j}\, C_j, \tag{16}$$

$$\sum_{j=1}^{j=n} C_j = 1,$$

where $S_{i,j} = r_i\,(a_{i,j}/a_{i,i})$. One of its eigenvectors is that formed by the actual concentrations,C_i, which is positive and normalized, and whose corresponding eigenvalue is 1. Since matrix $S_{i,j}$ is positive, the Perron theorem ensures that there is one and only one positive normalized eigenvector with one associated positive dominant eigenvalue, so the Perron eigenvector and the Perron eigenvalue of matrix $S_{i,j}$ are, in fact, respectively, C_i and 1.The von Mises method will provide the positive and normalized concentrations and the value of λ. In a real case, with errors in the measurements, λ will differ from the unity. The propagation of errors in the ratios r_i to errors in the concentrations and λ will be as discussed above for the propagation of errors in F_i to errors in K and in concentrations. In a real case, the between λ and 1 will be an indication of the actual errors in the measurements and in the concentrations calculated.

5. The Sherman algorithm revisited

With the methodology here proposed we will revise now the algorithm introduced by Sherman for the correction of matrix effects on X-ray fluorescence analysis [10]. The Sherman algorithm is described by the following system of equations:

$$\begin{pmatrix} A_{1,1}-t_1 & A_{1,2} & .. & A_{1,n} \\ A_{2,1} & A_{2,2}-t_2 & .. & . \\ . & . & .. & . \\ . & . & .. & . \\ A_{n,1} & . & .. & A_{n,n}-t_n \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ . \\ . \\ C_n \end{pmatrix} = 0, \tag{17}$$

where C_i is the concentration of each element and $A_{i,j}$ represents the influence coefficient of the element j on the analyte i, and t_i is the time in seconds required to accumulate a preset number of counts. It is claimed in [11] that, since system 17 is homogeneous, only ratios among the unknown C_i can be obtained, so an extra equation is needed, that is, the condition that the sum of all concentrations has to be the unit:

$$\sum_{j=1}^{j=n} C_j = 1. \tag{18}$$

Then, it is stated that by using equation(18), one of the equations from 17 is eliminated and so one solution is obtained. Since there are N equations in 18, the elimination of each equation gives rise to N different systems of equations and a different solution is obtained in each case. This is pointed out as a serious drawback of the method because there is no way to decide among the N solutions which is the correct one. It is also pointed out that another disadvantage is that the numerical values of the coefficients depend on the excitation conditions and on the number of accounts accumulated.

Nevertheless, the system of equations 17 can be viewed under a perspective: equations(17) are a linear homogeneous system of N equations. Linear Algebra tells us that such a system has only the trivial solution, except if one of the equations is a linear combination of the other $N-1$ equations, and then the determinant of the matrix of coefficients is zero. Only in that case ratios among the concentrations can be obtained in a coherent way. Without any *a priory* condition on the values of the coefficients, the N equations are linearly independent, and system 17 has only the trivial solution $C1 = C2 = \cdots Cn = 0$; this is incompatible with equation (18), so the linear system of equations formed by equations(17) and (18) has no solution. If one of the equations of 17 is arbitrarily replaced by equation (18), an arbitrary system of equations is obtained, and an arbitrary solution for the concentrations is also obtained. This solution has no relation with the one obtained by changing another equation of 17 by 18. Only when the equations of 17 are linearly dependent there is a non-trivial solution, the ratios among Ci can be calculated, and any of the equations in 17 can by replaced by equation(18), leading to the same solution for Ci.

In a real situation, Equations (17) are “almost dependent” and the substitution of equation (18) for any of the equations of system 17 gives rise to a situation similar to the one discussed when we were dealing with the first method – Beatie and Brissey – of obtaining approximate solutions.

Without any *a priori* condition, the determinant of the matrix of the coefficients does not have to be null, but it should be pointed out that there is a parameter on the matrix of the coefficients which can be taken as a supplementary variable. Let us see it in detail: the times t_i in seconds required to accumulate a preset number of counts can be written as $t_i = \mathrm{Nacc}/R_i$, where Nacc is the total number of accounts and R_i is the counting rate for each element. Since the physical problem does not depend either on the actual number of accounts measured, or on the unit of time, we can write t_i as $t_i = K/R_i$, where K is a free parameter. Substituting K/R_i for t_i in the system of equations (17), Eq. 17 can be written as an eigenvalue problem:

$$\begin{pmatrix} R_1\,A_{1,1} & R_1\,A_{1,2} & .\,. & R_1\,A_{1,n} \\ R_2\,A_{2,1} & R_2\,A_{2,2} & .\,. & . \\ . & . & .\,. & . \\ . & . & .\,. & . \\ R_n\,A_{n,1} & . & .\,. & R_n\,A_{n,n} \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ . \\ . \\ C_n \end{pmatrix} = K \begin{pmatrix} C_1 \\ C_2 \\ . \\ . \\ C_n \end{pmatrix}. \tag{19}$$

As already discussed, this eigenvalue problem, when all the influence coefficients $A_{n,n}$ are positive, has one and only one positive normalized solution for the concentrations which can be obtained in an iterative way. From a practical point of view, the Sherman model 17 can be written (by dividing each equation by its t_i) as follows:

$$\begin{pmatrix} A_{1,1}/t_1 & A_{1,2}/t_1 & .. & A_{1,n}/t_1 \\ A_{2,1}/t_2 & A_{2,2}/t_2 & .. & . \\ . & . & .. & . \\ . & . & .. & . \\ A_{n,1}/t_n & . & .. & A_{n,n}/t_n \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ . \\ . \\ C_n \end{pmatrix} = 1 \begin{pmatrix} C_1 \\ C_2 \\ . \\ . \\ C_n \end{pmatrix} \tag{20}$$

and the corresponding eigenvalue problem to be solved in order to obtain a positive normalized solution is:

$$\begin{pmatrix} A_{1,1}/t_1 & A_{1,2}/t_1 & .. & A_{1,n}/t_1 \\ A_{2,1}/t_2 & A_{2,2}/t_2 & .. & . \\ . & . & .. & . \\ . & . & .. & . \\ A_{n,1}/t_n & . & .. & A_{n,n}/t_n \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ . \\ . \\ C_n \end{pmatrix} = \lambda \begin{pmatrix} C_1 \\ C_2 \\ . \\ . \\ C_n \end{pmatrix}, \tag{21}$$

where, in the ideal case, the value of λ is 1. The deviation from 1 of the actual value of λ gives an indication of the goodness of the experimental data, $A_{i,j}$ and t_i.

6. Discussion

In a particular experimental situation characterized by selective excitation and selective measurements, the Sherman equations become linear and have a closed form solution with N sets of solutions when dealing with a sample with a sample made of N components. Nevertheless, we can assure the existence of one and only one physically acceptable solution whose concentration values are positive and normalized. The existence and uniqueness of that physically acceptable solution is strongly bounded to the positivity of the fluorescence production cross-sections and of the attenuation coefficients.

We have to point out the crucial importance of imposing the conditions of positivity and normalization of the concentrations in the iterative solution methods; this is equivalent to saying that the vector used as seed cannot be orthogonal to the dominant eigenvector. Otherwise, the iterative process is not suitable, since it will try to converge to one of the $N-1$ sets of non-physical solutions corresponding to the non-dominant complex or negative eigenvalues of the matrix. These conditions are also of crucial importance when solving the equations associated to other experimental situations which use standard samples. In particular, the Sherman algorithm has been reanalyzed and an stable way of obtaining unique positive and normalized solutions has been given.

The propagation of errors is stable, and *a priory* bounds on the relative errors in the concentrations have been obtained. This bounds are also based on the conditions of positivity and normalization of the concentrations.

We hope that this kind of analysis based on the *a priori* physical conditions the concentrations have to fulfill can be extended to polyenergetic beams and to the production of secondary fluorescence

7. Conclusions

It has been shown that in a particular experimental situation characterized by selective excitation and selective measurement, the Sherman equations become linear and can be solved in a closed form thus obtaining N sets of solutions when dealing with a sample of N components.

It has also been shown that, in the physical case of positive cross-sections and positive attenuation coefficients, there is one and only one physically acceptable solution which can be obtained in a stable iterative way.

An iterative convergent method has been found for calculating such a solution.

The method has been extended to experimental situations which use standard samples and to the Sherman algorithm.

The propagation of errors has been studied analytically and numerically thus obtaining $a\ priory$ bounds on the errors and showing that the propagation of errors is stable.

It seems convenient to study the possibility of extension of the methods here analyzed to polyenergetic beams and to the production of secondary fluorescence.

Acknowledgments

The authors are indebted to Dr. Richard Rousseau from the Geological Survey of Canada for his encouraging discussions and for providing copies of the references of Beattie and Brissey and Lachance and Traill.

References

[1] R.M. Figueroa, X-Ray Spectrom. 32 (2003) 299–306.
[2] H.J. Beattie and R.M. Brissey, Anal. Chem. 26 (1954) 980–983.
[3] G.R. Lachance, and R.J. Traill, Can. Spectrosc. 11 (1966) 43–48.
[4] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra* (SIAM, Philadelphia, 2000).
[5] von Mises, R. and P.-G. Hilda Ange. Math. Mech. 9 58–77, 1929 152–164.
[6] G.W. Stewart, *Introduction to Matrix Computation* (Academic Press, Orlando, 1973).
[7] A.R.G. Heesterman *MATRICES and their Eigenvalues* (World Scientific Publishing Co, Pte. Ltd. P O Box 128, Fatter Road, Singapore 9128, 1990).
[8] W.H. McMaster, N. Kerr Del Grande, J.H. Mallet and J.H. Hubbell, Lawrence Radiation Laboratory (Livermore), UCRL - 50174 Sec II Rev. 1 *Compilation of X-Ray Cross Sections* (1969).
[9] M.J. Berger and J. H. Hubbell, *XCOM: Photon Cross Sections on a Personal Computer,* NBSIR 87-3597 (National Bureau of Standards: Gaithersburg, MD, 1987).
[10] J. Sherman, *The Correlation Between Fluorescent X-ray Intensity and Chemical Composition* (ASTM Special Publication 157, Philodelphia, 1953), p. 27.
[11] R.E. Van Grieken and A. Markowicz, *Handbook of X-Ray Spectrometry* Spectroscopy Series, Vol. 29 (Marcel Dekker, New York, 2002).