Maximum Entropy Decomposition of Quadrupole Mass Spectra

U. v. Toussaint[1], V. Dose[1], and A. Golan[2]

[1]*Centre for Interdisciplinary Plasma Science,*

Max-Planck-Institut für Plasmaphysik, EURATOM Association,

Boltzmannstr. 2, D-85748 Garching, Germany and

[2]*American University, Roper 200, 4400 Massachusetts Ave,*

NW Washington, DC 20016-8029, USA

(Dated: February 23, 2004)

Abstract

We present an information-theoretic method called Generalized Maximum Entropy (GME) for decomposing mass spectra of gas mixtures from noisy measurements. In this GME approach to the noisy underdetermined inverse problem the joint entropies of concentration, cracking and noise probabilities are maximized subject to the measured data. This provides a robust estimation for the unknown cracking patterns and the concentrations of the contributing molecules. The method is applied to mass spectroscopic data of hydrocarbons and the estimates are compared with those received from a Bayesian approach [1, 2]. We show that the GME method is efficient and is computationally fast.

I.   INTRODUCTION

Quadrupole mass spectrometers [3, 4] are convenient and cheap instruments with various applications such as residual gas analysis in vacuum systems, process control [5, 6] and low temperature process plasma diagnostics [7, 8]. In the latter case plasma assisted gas phase chemical reactions are analyzed. The evaluation of such spectra is fairly complicated except for the case of non interfering mass numbers of cracking products of the gas constituents to be analyzed. However, the case of non interfering mass numbers nearly never arises in practice. Interference has therefore to be considered quantitatively. Let us choose, as an example, the chemical erosion of graphite by atomic hydrogen. This process is known to produce a series of hydrocarbons (e.g., CH_4,C_2H_2,C_2H_4,C_2H_6,C_3H_6,C_3H_8, neglecting for the purpose of the present discussion even heavier hydrocarbons) in which a particular component contributes to mass fractions of nearly all lighter members of the series.

This complicates the analysis. The cracking pattern of a quadrupole mass spectrometer is specific to the instrument and may even change drastically with variations of the field axis potential, the ionization energy and the frequently front panel controlled mass resolution. Disentangling a composite spectrum requires in general a calibration of the system that always suffers from corrupting noise. Such a calibration comprises the response measurement of the instrument to each of the constituents of the mixture. These data are then taken as a basis for decomposing the spectrum of a mixture and a frequently adopted simple decomposition proceeds as follows: First identify an entirely (or nearly) non interfering mass signal from the heaviest component of the mixture. For the above case of six hydrocarbons this would be mass 44 of C_3H_8. According to the cracking pattern obtained, the C_3H_8 contribution is subtracted from the mixture spectrum and the whole process is repeated for the second heaviest component, in the model case C_3H_6. It has been pointed out [9] that difficulties may be encountered in such a procedure since the lower mass hydrocarbons will be greatly affected by even small errors in the calibration as well as in the mixture signal measurements.

One step beyond this simple method is the solution of the (here) overdetermined system of linear equations, which results in our first case of twelve different masses contained in a mixture spectrum from the concentrations of only six different gas species. A solution of such an overdetermined system of linear equations has been offered using least-squares

2

methods [10, 11]. There are a number of basic shortcomings in using the least square method. The first is that the least-squares approach does not take advantage from the fact that the cracking matrix can be viewed as a set of proper probability distributions - it has to be imposed as restriction in the least squares type estimator. The second is that the "cracking pattern", which was obtained by just the same type of measurement that yields the mixture "data" and consequently suffers from the same precision limits, is assumed to be known exactly. The third is that due to measurement errors on cracking matrix and data the concentration of a trace component in the mixture may turn out to be negative. Therefore, such a mathematically possible result must be rejected as being logically and physically unacceptable. Another estimation method developed for analyzing overdetermined (over identified) systems of equations is the generalized method of moments [12]. This method can handle the over-identification issues by giving a weight to each piece of information (data), but it is not fully applicable to our estimation problem as it suffers from some of the basic issues discussed earlier. Under certain assumptions, however, the method we proposed here, is related to the above mentioned generalized method of moments [13, 14]. Furthermore, we have considered so far the favourable case in which it is possible to measure the cracking matrix. This may not always be the case. In case of mass spectroscopic detection of radicals the cracking pattern is not accessible in the usual way as a matter of principle. Nevertheless, radicals are of particular interest in the diagnostic of low temperature process plasmas. The method developed and presented here can handle the case of radicals as well.

In this article we present an information-theoretic method called Generalized Maximum Entropy (GME) to solve and interpret quadrupole mass spectra. It provides positive definite composition fractions. In addition, this method treats measurements of cracking matrix and mixture data on the same terms and incorporates the specified errors for both in a consistent and coherent way. Further, it allows us to introduce approximate prior knowledge of cracking data such as those in the published tables [15, 16]. We determine from a mixture measurement not only constituent concentrations but also improved values of the cracking matrix. In that way, we use the fact that a mixture spectrum carries always informations on both, the concentrations and the cracking coefficients. Finally, our method is constructed in a generic way for over-determined, well-determined, or underdetermined problems and provides efficient estimates in each case. The latter case arises in radical detection, since the cracking pattern of radicals is usually not even approximately known. The analysis remains

the same in all three cases.

II. SPECTRUM DECOMPOSITION

The ideal mass spectrum of a gas mixture containing J components of number densities $\hat{X}_j$ with $1 \leq j \leq J$ can be written to a good approximation as

$$\hat{m}_i = \hat{s} \sum_{j=1}^{J} \hat{X}_j \epsilon_j h_i \hat{P}_{ij}. \tag{1}$$

$\hat{m}_i$ is the mass signal in channel i, $1 \leq i \leq I$, ϵ_j the sensitivity of the instrument for gas species j, and $\hat{P}_{ij}$ the cracking matrix. The factor $\hat{s}$ is assumed to be constant; $\hat{s}$ carries a dimension which accommodates the relation of mass signal $\hat{m}_i$ to the species number densities $\hat{X}_j$. The factor h_i characterizes the mass dependent sensitivity of a (quadrupole) spectrometer which is composed of the transmission and the mass dependent ion detection efficiency of the channeltron. This is the place where an approximation enters. The ion detection efficiency of a channeltron depends not only on ion mass and energy but also on the ion species. This variation, though important in special cases, will be neglected in our analysis. We also assume that the instrument works in the linear response regime, meaning that at sufficiently low gas densities, the linear relation (1) holds. We shall now absorb ϵ_j and h_i in the concentrations $\breve{X}_j$ and the cracking matrix $\breve{P}_{ij}$ by defining $\breve{X}_j = \epsilon_j \hat{X}_j$ and $\breve{P}_{ij} = h_i \hat{P}_{ij}$. Further we generalize (1) to the case of N measurements providing mass signals $\hat{\vec{m}}_1 \ldots \hat{\vec{m}}_N$ and include measurement errors $\hat{e}_{ni}$:

$$\hat{m}_{ni} = \hat{s}_n \sum_{j=1}^{J} \breve{X}_{nj} \breve{P}_{ij} + \hat{e}_{ni}. \tag{2}$$

We allow for a change in overall sensitivity between the different measurements by indexing the factor $\hat{s}$. Finally we redefine (2) by introducing the quite natural normalizations $\breve{X}_{nj} \rightarrow X_{nj}$ and $\breve{P}_{ij} \rightarrow P_{ij}$. We also adopt the common practice of normalizing the mass signals $\hat{\vec{m}} \rightarrow \vec{m}$ such that

$$\sum_{i=1}^{I} P_{ij} = 1 \qquad \sum_{j=1}^{J} X_{nj} = 1 \qquad \sum_{i=1}^{I} m_{ni} = 1. \tag{3}$$

Note that the normalization of the mass signals has an influence on the error statistics and has to be handled with care in the standard χ^2-approaches that rely on a Gaussian error distribution. The various factors following from these normalizations are all absorbed in

$\hat{s}_n \to s_n$. $\hat{e}_{ni}$ is scaled with the same factor as m_{ni} to obtain e_{ni}. From (3) it follows that s_n will turn out to be close to unity since all the quantities are normalized and the error contributions have a random sign. After being normalized $\vec{P}_j$ and $\vec{X}_n$ can be interpreted as discrete probability distributions.

With this background, our objective is to estimate $\mathbf{P}$ and $\mathbf{X}$ with minimal assumptions on the noise component $\vec{e}$. Regardless of the size of the matrix $\mathbf{P}$ and the number of observations N, the problem is always under-determined; there are more unknowns than data. Possible approaches to solve such problems are the Maximum Entropy (ME) method developed by Jaynes [26] and the Generalized ME (GME) method [19]. For detailed discussion see for example [13, 14].

In order to apply the maximum entropy principle [17] or rather the principle of minimum cross entropy [18], we follow [19] and express the measurement errors as the expectation values of proper discrete probability distributions $\vec{W}_{ni}$:

$$e_{ni} = \sum_{k=1}^{K} W_{nik} r_{nik} \qquad W_{nik} \geq 0 \qquad \sum_{k=1}^{K} W_{nik} = 1, \tag{4}$$

which accounts for the fact that the noise can be positive or negative. The vector $\vec{r}_{ni}$ has K components symmetric around zero, $\vec{r}_{ni} = (r_{ni1}, \ldots, 0, \ldots, r_{niK})$. These vectors form the support of W_{nik}. The boundaries (r_{ni1} and r_{niK}) are a multiple of the measurement standard deviation σ_{ni} of mass signal i in the nth measurement. For a detailed discussion of the influence of the size and range of the error support on the estimations see [19]. For the present case we chose K = 3. The estimates have been shown to be very insensitive to variations in these settings. This completes our definitions and specifications.

Next we employ the principle of minimum cross entropy of the probabilities in $\mathbf{P}, \mathbf{X}$ and $\mathbf{W}$ subject to the constraints. The Lagrange functional to be optimized is

$$\begin{aligned}
\Phi = & \sum_{i,j} P_{ij} \log \left(\frac{P_{ij}}{P_{ij}^0} \right) + \sum_{n,j} X_{nj} \log \left(\frac{X_{nj}}{X_{nj}^0} \right) + \sum_{n,i,k} W_{nik} \log \left(\frac{W_{nik}}{W_{nik}^0} \right) + \\
& \sum_{j} \mu_j \left(1 - \sum_{i} P_{ij} \right) + \sum_{n} \nu_n \left(1 - \sum_{j} X_{nj} \right) + \sum_{n,i} \rho_{ni} \left(1 - \sum_{k} W_{nik} \right) + \\
& \sum_{n,i} \lambda_{ni} \left(m_{ni} - s_n \sum_{j} P_{ij} X_{nj} - \sum_{k} W_{nik} r_{nik} \right).
\end{aligned} \tag{5}$$

The first line of (5) is the sum of the cross entropies of the distributions $\mathbf{P}$, $\mathbf{X}$, $\mathbf{W}$ versus their default (prior) values $\mathbf{P}^0$, $\mathbf{X}^0$, $\mathbf{W}^0$, where $\mathbf{W}^0$ are always taken to be uniform, so the

priors on the noise components reflect $e_{ni} = 0 \, \forall n,i$. The second line includes the proper normalizations via Lagrange multipliers $\vec{\mu}$, $\vec{\nu}$, $\vec{\rho}$ and the third line introduces the the observed data represented as constraints on $\mathbf{P}$, $\mathbf{X}$ and $\mathbf{W}$ resulting from measured mass spectra $\vec{m}_1, \ldots, \vec{m}_N$ via Lagrange multipliers $\vec{\lambda}$. The optimal solution can be computed with standard numerical packages (e.g. [20, 21]) or other optimization software for reasonably sized problems (6 measurements, 12 mass channels, 6 source gases, support space size 3, resulting in a total number of variables of 492) in a few minutes.

III. THE UNCONSTRAINED (CONCENTRATED) DUAL MODEL

The special structure of the equation system with its constraints allows one to derive a fast iterative solution scheme already fulfilling the normalization constraints [19, 22]. At the same time the number of unknowns is reduced to the number of measured data points represented by the basic set of unknowns $\vec{\lambda}$. In order to find the optimum of (5) we start differentiating it with respect to $\mathbf{P}$,$\mathbf{X}$ and $\mathbf{W}$ and equating to zero. Thus

$$\frac{\partial \Phi}{\partial P_{ij}} = \log \left(\frac{P_{ij}}{P_{ij}^0} \right) + 1 - \sum_{n=1}^{N} s_n X_{nj} \lambda_{ni} \overset{!}{=} 0. \tag{6}$$

The solution is given by

$$P_{ij}^* = \frac{P_{ij}^0 \exp \left[\sum_n s_n X_{nj}^* \lambda_{ni}^* \right]}{\Omega_j^P \left(\vec{\lambda}^* \right)}, \tag{7}$$

with

$$\Omega_j^P \left(\vec{\lambda} \right) = \sum_i P_{ij}^0 \exp \left[\sum_n s_n X_{nj} \lambda_{ni} \right]. \tag{8}$$

The denominator $\Omega_j^P \left(\vec{\lambda} \right)$ takes care of the normalization in $\mathbf{P}$. Next we take the derivatives with respect to $\mathbf{X}$

$$\frac{\partial \Phi}{\partial X_{nj}} = \log \left(\frac{X_{nj}}{X_{nj}^0} \right) + 1 - \sum_{i=1}^{I} s_n P_{ij} \lambda_{ni} - \nu_n \overset{!}{=} 0 \tag{9}$$

and obtain

$$X_{nj}^* = \frac{X_{nj}^0 \exp \left[s_n \sum_i P_{ij}^* \lambda_{ni}^* \right]}{\Omega_n^X \left(\vec{\lambda}^* \right)} \tag{10}$$

and the normalizing factor is given by

$$\Omega_n^X \left(\vec{\lambda} \right) = \sum_j X_{nj}^0 \exp \left[s_n \sum_i P_{ij} \lambda_{ni} \right]. \tag{11}$$

6

In the same way the expression for W is derived as

$$W^*_{nik}\left(\vec{\lambda}\right) = \frac{W^0_{nik} \exp\left[\lambda^*_{ni}\mathfrak{r}_{nik}\right]}{\Omega^W_{ni}\left(\vec{\lambda}^*\right)} \qquad (12)$$

with

$$\Omega^W_{ni}\left(\vec{\lambda}\right) = \sum_k W^0_{nik} \exp\left[\lambda_{ni}\mathfrak{r}_{nik}\right]. \qquad (13)$$

We use these relations to determine λ_{ni}. To this end we substitute the results obtained so far (eq. (7), (10) and (12)) into (5) and obtain

$$\Phi = \sum_{ni} \lambda_{ni}m_{ni} - \sum_j \log \Omega^P_j\left(\vec{\lambda}\right) - \sum_n \log \Omega^X_n\left(\vec{\lambda}\right) - \sum_{ni} \log \Omega^W_{ni}\left(\vec{\lambda}\right) + \sum_{ni} \lambda_{ni} \sum_j P_{ij}X_{nj}. \qquad (14)$$

Note, since eqs.(7),(10) and (12) are properly normalized the second line of eq.(5) which contains the normalizing conditions is equal to zero upon back substitution. Regarding eq.(14) now as the unconstrained problem in λ, or in statistical terms, the concentrated model with parametric dependencies in **X**,**P** and **W** we solve for the N×I unknowns λ_{ni} instead of the much larger number of unknowns in the direct approach (in our first example the number of variables is reduced from 492 to 72). Using

$$\frac{\partial \Phi}{\partial s_n} = -\sum_i \lambda_{ni} \sum_j P_{ij}X_{nj} \overset{!}{=} 0 \qquad (15)$$

equation (14) simplifies even more. Instead of using the above equation (15) to determine $\vec{s}$ it is easier to return to the model equation

$$m_{ni} = s_n \sum_j X_{nj}P_{ij} + \sum_k W_{nik}\mathfrak{r}_{nik} \qquad (16)$$

and to sum over i

$$\sum_i m_{ni} = 1 = s_n \underbrace{\sum_j X_{nj}}_{=1} \underbrace{\sum_i P_{ij}}_{=1} + \sum_{i,k} W_{nik}\mathfrak{r}_{nik}, \qquad (17)$$

resulting in the simple expression

$$s_n = 1 - \sum_{i,k} W_{nik}\mathfrak{r}_{nik}. \qquad (18)$$

Nevertheless, it is recommended to fix the s_n to 1 for the iterative solution. It has been shown that the scaling variables deteriorate the otherwise very stable iteration process.

Furthermore the values of $\vec{s}$ are expected to be very close to 1 (usually to better than 1% as comparisons with the solutions of the direct approach have shown).

The key to the solution of the problem is evidently to find the maximum of $\Phi\left(\lambda | P_{ij}, X_{nj}, W_{nik}\right)$ since we have obtained explicit solutions of the conditioning variables. This sketches the iteration to be followed. Find λ starting from $\mathbf{P}^{(0)}=\mathbf{P}^0$, $\mathbf{X}^{(0)}=\mathbf{X}^0$ and $\mathbf{W}^{(0)}=\mathbf{W}^0$, then calculate $\mathbf{P}^{(1)}$, $\mathbf{X}^{(1)}$, $\mathbf{W}^{(1)}$ until convergence is obtained.

IV.   RESULTS

To illustrate and evaluate the performance of this inversion procedure a comparison with a recently introduced Bayesian solution for the mass spectroscopy decomposition problem [1, 2] has been carried out. To obtain the uncertainty of the estimated values sampling simulations were done. According to the given data uncertainty 100 sets of simulated data points were generated using a normal distribution for each data point centered at the measured value. For each set the optimal $\mathbf{P}^*$ and $\mathbf{X}^*$ have been computed and those values were used to estimate the mean and variance for $\mathbf{P}$ and $\mathbf{X}$. The time required for a point estimate is typically below one second on a contemporary PC. The estimation of the uncertainties proceeded initially according to the traditional likelihood and least squares approaches in which the uncertainty is directly derived from the inverse of the Hessian matrix of eq.(14)[19]. However this takes into account only the curvature of the function at the maximum which turned out to be misleading in the present problem which involves random variables in both sides of the equations. The resulting estimates were at variance with the Bayesian estimates [1] and therefore a bootstrap approach, which is a consistent method for estimating the sampling variances, has been used instead.

1.   First Example

The analysis shown in the previous section was first applied to analyse the data set given in Schwarz-Selinger et al [2]. The mass spectroscopic data from the pyrolysis of $(CH_3)_2N_2$ (used as a source for methyl radicals) have been measured with a differentially pumped mass spectrometer (Hiden Hal 201) in the line of sight of the capillary gas line. Furthermore, since cracking patterns can depend on gas pressure [24] five calibration measurements were

performed for the main fragments of $(CH_3)_2N_2$ namely N_2,CH_4,C_2H_2, C_2H_4 and C_2H_6 at signal levels comparable to those observed in the pyrolysis. A detailed description of the experiment and the data is given in [2].

For the present evaluation the same prior information about the cracking matrix $\mathbf{P}^0$ has been used as in [2] and a flat prior distribution for the concentration $\mathbf{X}^0$ of the individual fragments was applied (see first line of Table I). The results of the analysis for the radical CH_3 are given in Table II and compared with the results of [2] obtained by a thorough Bayesian analysis. The latter analysis requires time consuming Markov Chain Monte Carlo sampling [23]. The agreement is excellent. Not only do the point estimates for the cracking matrix elements coincide well within the uncertainties but also within the estimated error margin. The computed fractions of the mixture constituents are also in very good agreement. In accordance with the CH_3-concentration given in Table I (0.127 ± 0.003) an independent result of (0.13 ± 0.04) using ionization threshold mass spectroscopy has been published [25].

2. Second example

As a second example we chose the data set given in [1] applying the same analysis as before. A mixture of three carbon hydrates $(C_2H_6,C_3H_8, C_4H_{10})$ was analysed, together with calibration measurements for each of the carbon hydrates used. The results for the concentrations are shown in table III and compared with the values obtained by Preuss et al [1]. Again the results match each other within the estimated uncertainty and also the uncertainties are in very good agreement. The same is valid for the estimated cracking matrix elements of this mixture experiment (not shown).

V. CONCLUSION

We have used the GME-approach to perform a joint estimation of cracking matrices and fractions taking into account all relevant measurements and uncertainties. The GME results are practically identical (within the error bars) to the more elaborate Bayesian approach using MCMC-techniques. Therefore this comparison provides a valuable mutual validation of both approaches (Bayes and GME). The GME method is about 100-times faster than the Bayesian method and therefore allows online monitoring of the plasma composition. Finally,

the GME approach being able to take into account uncertainties in the cracking matrix, overcomes a main difficulty of other approaches that rely on poorly justified assumption of exact cracking patterns.

APPENDIX A: INFERENCE, DIAGNOSTICS AND STATISTICS

We now provide the basic statistics and diagnostics necessary to evaluate the GME estimates.

1. The Entropy Ratio Statistic

Following the logic of the maximum likelihood ratio test, it is possible to construct an entropy-ratio test that has a limiting χ^2 distribution, and is used to obtain confidence intervals for the estimated parameters. Let λ be the vector of Lagrange multipliers for the N$\times$ I observations. Let $H_u(\lambda^*)$ be the objective (total entropy) value of the GME model where *none of the parameters are constrained*, or similarly, none of the λ's are constrained. Thus, $H_u(\lambda^*)$ is just the optimal value of the objective functional represented in (5) or in its dual-concentrated model (14). Next let $H_C(\lambda_0)$ be the entropy value of the constrained problem where λ=0, or equivalently all the parameters are constrained to zero (or posteriors are constrained to their correspondent priors). Thus $H_C(\lambda_0)$ is the maximum *possible* value of the joint entropies (objective function) if all priors are uniform. It can be obtained by maximizing the joint entropies subject to no constraints (except for the requirements that all distributions are proper). Doing so yields the maximal total entropy value of the three sets of discrete, uniform distributions $\mathbf{P}$, $\mathbf{X}$ and $\mathbf{W}$. In case of non-uniform priors $H_C(\lambda_0)$ is equal to the total entropies of the priors $\mathbf{P}^0$, $\mathbf{X}^0$ and $\mathbf{W}^0$. The entropy ratio statistic, ER, for testing the null hypothesis H_0:$\lambda = 0$ is then

$$ ER(\lambda = 0) = 2\left[H_C(\lambda_0) - H_u(\lambda^*)\right]. \tag{A1} $$

Under certain (mild) assumptions, $ER(\lambda = 0) \xrightarrow{d} \chi^2_{(I-1)N}$ as N $\to \infty$ when H_0 is true, and $(I-1)N$ is the number of basic parameters (or restrictions), where $\xrightarrow{d}$ stands for convergence in probability. The appropriate α-level confidence interval for λ (and similarly for $\mathbf{P}$ and $\mathbf{X}$) is obtained by setting $ER(\lambda) \leq C_\alpha$, where C_α is chosen so that $\Pr(\chi^2 < C_\alpha) =$

α. Similarly, we can test any other hypothesis of the form $H_0 : \lambda = \lambda_0$, or $H_0 : \mathbf{P} = \mathbf{P^0}$ for all, or any subset, of the parameters. We use this statistic to test whether our estimates are significantly different than the priors.

2. The Normalized Entropy Statistic

Within the information approach used here, one can investigate the amount of information in the estimated coefficients and the informational content in each one of the observations. The *Normalized Entropy* (information) measure for the whole cracking matrix is

$$S\left(\mathbf{P}^*\right) \equiv \frac{-\sum_{i,j} p_{ij}^* \log\left(p_{ij}^*\right)}{J \times \log\left(I\right)} \tag{A2}$$

with $S(\mathbf{P}^*) \in [0, 1]$, where 1 reflects complete ignorance or complete uncertainty (all estimated probabilities are uniform, or conversely $\lambda = 0$), and 0 reflects perfect knowledge (perfect in-sample prediction). If priors are non-uniform, substitute the denominator with the entropy of the priors. Similarly the normalized entropy for each one of the P_j^* can be constructed, as well as for X_n^*. In summary, the lower S, the more information there is in the analyzed sample relative to the prior, meaning the associated estimated Lagrange multipliers λ^* are different from zero, yielding in turn a post-data (posterior) distribution that is "far" away from the prior distribution.

3. A "Goodness of Fit" Measure

Using the same line of reasoning as above (each constraint, or data point, represents additional potential information that may lower the value of the objective function but can never increase it) the "goodness of fit" measure for this GME estimator is

$$R = 1 - \frac{H_u\left(\mathbf{P}^*, \mathbf{X}^*\right)}{H_C\left(\mathbf{P}^0, \mathbf{X}^0\right)} = 1 - S\left(\mathbf{P}^*, \mathbf{X}^*\right) \tag{A3}$$

where R=0 implies the data set has no new information in it, and R=1 implies perfect certainty or perfect "in-sample" prediction. The relationship between the "goodness of fit"-measure and the normalized entropy reveals the natural connection between the traditional information-theoretic (or entropy) and the ML-type methods.

4. The CE and χ^2 Tests

Recall the CE definition for $\mathbf{P}$ as

$$\mathrm{I}\left(\mathbf{P}; \mathbf{P}^0\right) = \sum_{ij} p_{ij} \log \frac{p_{ij}}{p_{ij}^0}. \tag{A4}$$

Given the null hypothesis $\mathrm{H}_0 : \mathbf{P} = \mathbf{P}^0$, then

$$\chi^2_{(\mathrm{I}-1)\mathrm{J}} = \sum_{ij} \frac{\left(p_{ij} - p_{ij}^0\right)^2}{p_{ij}^0}. \tag{A5}$$

Now a second order approximation to the CE above is

$$\mathrm{I}\left(\mathbf{P}; \mathbf{P}^0\right) \approx \frac{1}{2} \sum_{ij} \frac{\left(p_{ij} - p_{ij}^0\right)^2}{p_{ij}^0} \tag{A6}$$

which is just one-half of the entropy-ratio statistic (for evaluation $\mathbf{P}^*$ versus $\mathbf{P}^0$) that we discussed above. We use this approach to test hypotheses regarding our estimates $\mathbf{P}^*$ and $\mathbf{X}^*$.

5. The Covariance

Given the dual (concentrated) model, the covariance is basically the inverse of the Hessian evaluated at λ^*. The Hessian of (14) is

$$\mathrm{H}\left(\lambda\right) = \left(\frac{\partial^2 \Phi}{\partial \lambda \partial \lambda^{\mathrm{T}}}\right) = \frac{\partial^2}{\partial \lambda \partial \lambda^{\mathrm{T}}} \left[\log \Omega\left(\lambda\right)\right], \tag{A7}$$

where $\Omega\left(\lambda\right) \equiv \Omega^{\mathrm{P}}\left(\lambda\right) \Omega^{\mathrm{X}}\left(\lambda\right) \Omega^{\mathrm{W}}\left(\lambda\right)$. Given the Hessian $\mathrm{H}(\lambda^*)$, and the Covariance $Cov(\lambda^*)$ we can calculate the Covariance of $\mathbf{P}^*$ (or $\mathbf{X}^*$) via the transformation

$$Cov\left(\mathbf{P}^*\right) = \left(\frac{\partial \mathbf{P}^*}{\partial \lambda^*}\right)^{\mathrm{T}} Cov\left(\lambda^*\right) \left(\frac{\partial \mathbf{P}^*}{\partial \lambda^*}\right). \tag{A8}$$

[1] PREUSS, R., KANG, H., SCHWARZ-SELINGER, T. and DOSE, V., Quantitative Analysis of Multicomponent Mass Spectra, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 21st International Workshop*, edited by R. L. Fry, AIP 2002, p.155-162.

[2] SCHWARZ-SELINGER, T., PREUSS, R. DOSE, V. and von der LINDEN, W., J. Mass Spectrom. **36** (2001) p.866-874.

[3] PAUL, W. and STEINWEDEL, H., Z. Naturforschung A**8** (1953) p.448.

[4] DAWSON, P. *Quadrupole Mass Spectrometry*, Elsevier, Amsterdam, 1976.

[5] TISONE, T.C., BOLKER, B.F.T. and LATOS, T.S., J. Vac. Sci. Technol. **17** (1980) p.415

[6] AITA, C.R., MYERS, T.A. and LA-ROCCA, W.J., J. Vac. Sci. Technol. **18** p. 324

[7] COBURN, J.W., Rev. Sci. Instr. **41** (1970) p. 1219

[8] VASILE, M.J. and SMOLINSKY, G., Int. J. Mass. Spectrom. Ion Phys., **12** (1973), p.133.

[9] DOBROZEMSKY, R. and FÄRBER, W., Vakuumtechnik **20** (1971) p.231.

[10] RAIMONDI, D.L., WINTERS,H.F., P. G. and CLARKE, D., IBM J. Res. Dev. **15** (1971) p.307.

[11] DOBROZEMSKY, R., J. Vac. Sci. Technol. **9** (1972) p.220.

[12] MATYAS,L., *Generalized Method of Moment Estimation*, edited by Laszlo Matyas, Cambridge University Press, Cambridge,1999

[13] GOLAN, A., Information and Entropy Econometrics- Editor's View, J. Econometrics, **107**,1-2 (2002),p.1-15.

[14] MITTELHAMMER,R.C., JUDGE,G., MILLER, D., *Econometric Foundations*, Cambridge University Press, Cambridge, 2000.

[15] CORNU, A. and MASSOT, R., *Compilation of Mass Spectral Data*, Heyden London, 1979.

[16] SPITELER, M. and SPITELER, G., *Massenspektrensammlung von Loesungsmitteln, Verunreinigungen, Saeulenbelegmaterialien und einfachen aliphatischen Verbindungen,* Springer Verlag, Wien/New York, 1973.

[17] SIVIA, D., *Data Analysis: A Bayesian Tutorial*, Oxford University Press, 1996.

[18] KAPUR, J. and KESAVAN, H.K., *Entropy Optimization Principles with Applications* Academic Press, Boston,MA, 1992.

[19] Golan, A., JUDGE, G. and MILLER, D., *Maximum Entropy Econometrics*, John Wiley & Sons, West Sussex, 1996.

[20] Optimization routine nag_nlp_sol, mark18 from NAG LTD, Oxford,OX2 8DR, UK, http://www.nag.co.uk

[21] PRESS, W., TEUKOLSKY, S., VETTERLING, W., and FLANNERY, B., *Numerical Recipes in Fortan*, Cambridge University Press, Cambridge, 1992.

[22] AGMON, N., ALHASSID, Y. and LEVINE, R.D., J. of Computational Physics **30**, p. 250-259, 1979.

[23] GILKS, W., RICHARDSON, S., and SPIEGELHALTER, D., editors, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London, 1996.

[24] BRETH, A. and DOBROZEMSKY, R., Deviation of cracking patterns and their influence on rga accuracy, in *SASP 82, Symposium on atomic and surface physics*, edited by W. LINDINGER, F. HOWORKA, T. M. and EGGER, F., (STUDIA, Maria Alm/Salzburg, Austria, 1982), p.75.

[25] SCHWARZ-SELINGER, T., DOSE, V., JACOB, W., and von Keudell, A., J. Vac. Sci. Technol. A **19** (2001) p.110.

[26] JAYNES,E.T., Information Theory and Statistical Mechanics Phys. Rev. **106** (1957) p.620.

TABLE I: Prior and posterior fraction of the mixture constituents using either the Bayes approach with an exponential prior on the cracking elements or the GME method. To be comparable the results from [2] are scaled so that the mean values add up to one.

	N_2	CH_4	C_2H_2	C_2H_4	C_2H_6	CH_3
Prior	0.167	0.167	0.167	0.167	0.167	0.167
Bayes	0.398±0.002	0.208±0.002	0.001±0.002	0.043±0.006	0.225±0.004	0.125±0.004
GME	0.396±0.004	0.209±0.003	0.001±0.001	0.044±0.004	0.223±0.005	0.127±0.003

TABLE II: Posterior expectation of the cracking coefficients of the radical CH_3 using the Bayesian method with an exponential prior (first entry) or the GME method (second entry) with estimated error margin, respectively. The entries are scaled to a maximum value of 1000, a notation commonly used in organic mass spectrometry.

Mass	12	13	14	15
Prior	65 ± 10	136 ± 13	823 ± 82	1000 ± 100
Bayes	10 ± 1	38 ± 3	358 ± 20	1000 ± 20
GME	9 ± 1	40 ± 3	374 ± 20	1000 ± 19

TABLE III: Prior and posterior fraction of the mixture constituents using either the Bayesian approach with an exponential prior or the GME ansatz. To be comparable the results from [1] are scaled so that the mean values add up to one.

	C_2H_6	C_3H_8	C_4H_{10}
Prior	0.444	0.333	0.222
Bayes	0.443 ± 0.011	0.290 ± 0.017	0.267 ± 0.019
GME	0.443 ± 0.014	0.296 ± 0.021	0.261 ± 0.014