Journal of Chromatography, 513 (1990) 349–353
Elsevier Science Publishers B.V., Amsterdam

CHROM. 22 473

Note

Information theory of column chromatography on the basis of the information measure FUMI

YUZURU HAYASHI* and RIEKO MATSUDA
National Institute of Hygienic Sciences, 18–1, Kamiyoga 1-Chome, Setagaya-ku, Tokyo 158 (Japan)
(First received December 8th, 1988; revised manuscript received January 23rd, 1990)

Precision and accuracy are of the greatest importance in all aspects of the theory and practice in analytical chemistry[1]. The information measure FUMI (FUnction of Mutual Information) quantitatively describes the precision of measurements in chromatography[2–4] and the concept involved is known in information theory as mutual information[5]. The optimum condition can be clearly defined as one which can provide the maximum amount of mutual information FUMI.

The aim of this paper is to consider the amount of information obtained from peaks in some chromatographic circumstances from the viewpoint of information theory based on FUMI. The treatment of the optimum peak separation has disparate views in the so-far developed optimization methods using the resolution R_s[6–9]. The optimum separation can be deduced as a theoretical consequence from the maximum amount of FUMI.

UNDERLYING CONCEPT

The optimum peak separation under a particular operating condition can be derived from FUMI[10]. If the leading peak ($j = 1$) is fixed at a retention time τ_1, then the optimum retention time $\hat{\tau}_2$ ($>\tau_1$) for the second peak is expressed as

$$\hat{\tau}_2 = \frac{N^{1/2} + \gamma}{N^{1/2} - \gamma} \cdot \tau_1 \tag{1}$$

where N denotes the plate number of the column and γ is a constant ($= 2.07$). Using the well known approximation[11,12] $(1 + x/2)/(1 - x/2) \approx 1 + x$, we obtain the least capacity factor $\hat{k}_2$ corresponding to $\hat{\tau}_2$:

$$\hat{k}_2 = \left(1 + \frac{2\gamma}{N^{1/2}}\right) k_1 + \frac{2\gamma}{N^{1/2}} \tag{2}$$

In the above, the peak areas A_j are assumed to be the same. The optimum peak

0021-9673/90/$03.50

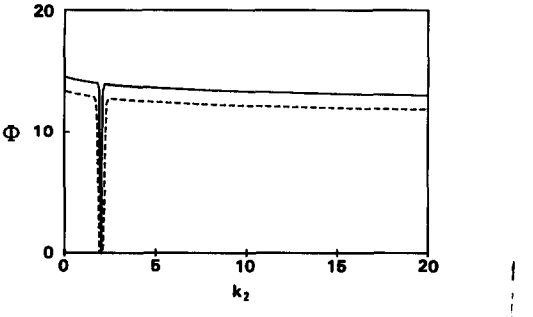

Fig. 1. Effect of the degree of peak overlap on FUMI for a two-peak system. The capacity factor k_1 is fixed (= 2); k_2 is varied. A_j = 5000 (j = 1, 2); N = 10 000 (solid line) or 1000 (dashed line); the hold-up time τ_0 = 200; $\tilde{a}$ = 1; γ = 2.07. Filtering-off points: $\kappa_f(1) = \tau_2 - \beta\sigma_2$; $\kappa_f'(2) = 2\tau_2 - (\tau_1 + \beta\sigma_1)$, where $\beta = 2\gamma - \pi^{1/2}/2^{10}$.

separation was derived under the condition that the optimum suffers no strong peak overlap, *i.e.*, the information loss $\delta\Phi = 0$.

CHROMATOGRAPHIC POWERS

Fig. 1 shows the influence of peak overlap and separation on the mutual information Φ. It is assumed that one peak is fixed at a position τ_1 and the other moves; the other variables, N, A_j, etc., are invariant. There exists a deep trough, which stems from strong peak overlap. In addition to the trough, the FUMI Φ gradually decreases with increasing capacity factor k_2 of the moving peak because of the peak-widening effect[3].

The maximum of FUMI at the right-hand side of the trough shown in Fig. 1 indicates the optimum separation for peaks in a column of plate number N. The maximum information can be obtained from the optimally separated peaks and can be derived from eqn. 2 and FUMI[3,4]:

$$\Phi = \psi_1^* + \psi_2^* - \frac{1}{2}\log\left[(k_1 + 1)^2\left(1 + \frac{2\gamma}{N^{1/2}}\right)\right] \tag{3}$$

where

$$\psi_j^* \equiv \frac{1}{2}\log\left(\frac{A_j^2 N^{1/2}}{2\pi^{1/2}\tau_0\tilde{a}}\right) \tag{4}$$

τ_0 is the hold-up time and $\tilde{a}$ denotes the power spectrum intensity of white noise[4]. The function $N/(1 + 2\gamma/N^{1/2})$ involved in Φ increases monotonously with respect to N and then the plate number effect varies accordingly. If N = 1000 and k_1 = 2.00, then $\hat{k}_2$ = 2.39 and Φ = 12.76. If N is improved, FUMI selects sharper but closer peaks of higher precision as the optimum. If N = 10 000 and k_1 = 2.00, then $\hat{k}_2$ = 2.14 and Φ = 13.95.

The information Φ can vary, even if the relative retention α (= k_2/k_1) is kept

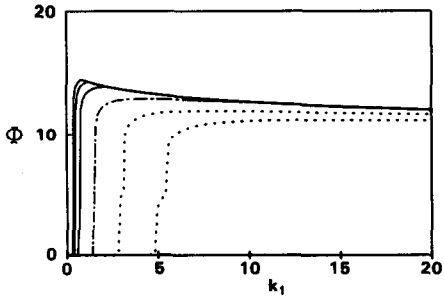

Fig. 2. Effect of capacity factor k_1 on FUMI for various values of the relative retention α. $A_j = 5000$; $\tilde{a} = 1$; $\gamma = 2.07$; $N = 10\,000$; $\tau_0 = 200$. $\hat{\alpha} = 1.041$ (theoretical). From top to bottom, $\alpha = 1.100$ [the maximum of $\Phi = 14.48$ ($\delta\Phi = 0$) at $k_1 = 0.78$ from simulation (0.73 from eqn. 5)]; 1.080 [$\Phi_{max} = 14.28$ (0.005) at $k_1 = 1.16$ (1.12)]; 1.060 [$\Phi_{max} = 13.88$ (0.025) at $k_1 = 2.16$ (2.38)]; 1.040 [$\Phi_{max} = 12.89$ (0.509) at $k_1 = 4.28$]; 1.030 [$\Phi_{max} = 11.92$ (0.956) at $k_1 = 7.92$]; 1.025 [$\Phi_{max} = 11.18$ (1.181) at $k_1 = 13.88$].

constant. As the capacity factors k_1 and k_2 of the peaks increase, the mutual information increases abruptly, reaches a maximum and gradually decreases [see Fig. 2 (solid lines)]. The small amounts of information at both sides of the Φ maximum can be attributed to too strong peak overlap and excessive widening of peaks. The maximum information in Fig. 2 is given approximately by substituting in FUMI the optimum capacity factors of the constant α:

$$\hat{k}_1 \approx \frac{2\gamma}{N^{1/2}(\alpha - 1) - \gamma(\alpha + 1)} \tag{5}$$

(and $\hat{k}_2 = \alpha\hat{k}_1$ when eqn. 1 is used).

A column of plate number N has a definite lower limit $\hat{\alpha}$ and any peaks characterized by α ($<\hat{\alpha}$) cannot be quantified with the required precision in the column:

$$\hat{\alpha} = 1 + \frac{2\gamma}{N^{1/2}} \tag{6}$$

This can be obtained from the restriction that the denominator in eqn. 5 should be positive. The behaviour of the information Φ towards α is shown in Fig. 2 (N = constant). If α is larger than the lower limit, then the mutual information Φ takes a high peak (solid lines). A drastic change in Φ occurs at the lower limit as a boundary (dot-dashed line). Below the limit $\hat{\alpha}$, FUMI can no longer reach the information obtained above $\hat{\alpha}$ and the peaks fail to provide the precision in any k_1 or k_2 examined (dotted lines). This degraded information is attributed to the unavoidable information loss $\delta\Phi$ arising from strong peak overlap.

Eqn. 6 also indicates that sufficiently separated peaks ($\alpha \geqslant \hat{\alpha}$) can be quantified with satisfactory precision even if a column of lower N [but $N^{1/2} \geqslant 2\gamma/(\alpha - 1)$] is used. It involves at least two cases: (i) a column of the same size but lower N; (ii) a short column of the same plate height. In case (i), the information decreases according to the

equation of FUMI. In contrast, shortening of the column leads to an increase in the information of the sufficiently separated peaks ($\alpha \geqslant \hat{\alpha}$).

A change in the column length L affects the retention time τ_j and plate number N. If the length of the colunn L is changed to xL, FUMI is described by

$$\Phi = \psi_1^* + \psi_2^* - \frac{1}{2}\log[(k_1 + 1)(k_2 + 1)] - \log(x^{1/2}) \tag{7}$$

By replacing N in eqn. 2 with $\hat{x}N$, we can obtain the optimum column length $\hat{x}L$ for the peaks:

$$\hat{x}^{1/2} = \frac{2\gamma}{N^{1/2}} \cdot \frac{k_1 + 1}{k2 - k_1} \tag{8}$$

Eqns. 3–8 will be useful for determining the amount of information that can be obtained from peaks with the optimum separation in various situations.

Plots of R_s under the same conditions as shown in Figs. 1 and 2 will display no similarity to the plots of FUMI. R_s plots are monotonously increasing functions of k_2 at a fixed k_1 and of k_1 at a fixed α[6]. The plot of R_s against N will be similar in shape to that of FUMI, but there is a difference: $R_s \propto N^{1/2}$; $\Phi = \log N^{1/2} + \text{constant}$. The above results reflect a clear distinction between information and separation.

A two-peak system has been treated here for simplicity. The theory of FUMI can easily be applied to optimization of multi-peak systems without any other conditions because of its additivity: $\Phi = \sum_{j=1}^{q} \varphi_j$ (φ_j denotes the individual peak information). The calculation of FUMI is also very easy: for Gaussian peaks, it requires only peak area, peak width, retention time and noise level.

FUMI contains an arbitrary constant γ which can be determined according to the conformity with linearity of the chromatograph and peak-resolving power of the data processing sytem. In this paper, γ is set at 2.07 for mathematical peak resolution by Kalman filter. FUMI, however, holds for other data processing such as the perpendicular dropping if γ takes another appropriate value. For the optimum peak resolution, γ is closely related to R_s[10]. If the least acceptable separation (*e.g.*, $R_s = 1.5$ or 3.0) is specified, then FUMI can select the optimum conditions according to the corresponding γ. The optima from FUMI and R_s-based criteria will not always be the same because of the difference in the basic concepts.

Analytical applications of FUMI are now under development. Its merits and limitation should be elucidated by further study of the optimization of many chromatographic variables.

REFERENCES

1 D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte and L. Kaufman, *Chemometrics: a Textbook*, Elsevier, Amsterdam, 1988.
2 Y. Hayashi, S. Yoshioka and Y. Takeda, *Anal. Sci.*, 5 (1989) 329.
3 Y. Hayashi and R. Matsuda, *Anal. Chim. Acta*, 222 (1989) 313.

4 Y. Hayashi, *Anal. Sci.*, 6 (1990) 15.
5 S. Arimoto, *Johou Riron (Information Theory)*, Kyoritsu Shuppan, Tokyo, 1976.
6 P. J. Schoenmakers, *Optimization of Chromatographic Selectivity*, Elsevier, Tokyo, 1986.
7 H. J. G. Debets, B. L. Bajema and D. A. Doornbos, *Anal. Chim. Acta*, 151 (1983) 131.
8 C. E. Goewie, *J. Liq. Chromatogr.*, 9 (1985) 1431.
9 J. C. Berridge, *Anal. Chim. Acta*, 191 (1986) 243.
10 Y. Hayashi, *Anal. Sci.*, 6 (1990) 257.
11 J. M. Gidding, *Anal. Chem.*, 39 (1967) 1027.
12 Japan Society of Analytical Chemistry, Kanto Region Branch, *Kosoku Ekitai Kuromatogurafu Bunseki (High-Performance Liquid Chromatographic Analysis)*, Sangyo Tosho, Tokyo, 1982.