A Novel Determination of Peak Areas for Application on Large Computers

F. Caesar / M. Klier
BASF Aktiengesellschaft, D-6700 Ludwigshafen, Germany

Summary

A computer program with some new algorithms for the determination of peak areas of gas chromatograms has been developed which has been used for several years with a satellite computer system. In contrast to most gc-programs the first and second derivatives of the curve are not used for peak detection. The maximum of a peak is defined by ordinates of the sample points alone; the base line is constructed by drawing curves of higher order through those parts of the chromatogram which are defined to be base line by special criteria. Consequently, the peak areas on the tailing of a solvent are determined more correctly than with skimming and, furthermore, the calculated base line of chromatograms with temperature program and subsequent isothermal run can be approximated to the real base line. The base line divides the chromatogram into several peak groups which are further separated by the “democratic distribution” method. This program is best suited for nonroutine analysis in research laboratories, because only a few input parameters are necessary for peak area determination with unknown chromatograms.

In the Analytical Laboratories of the BASF in Ludwigshafen gas chromatographs are connected among other instruments to a process control computer, which works as a satellite of a big computer in the computing centre [1]. The digital sample points of the gas chromatographs are collected and after the end of the run the data are transmitted to the computing centre together with the input parameter necessary for computation.

A software package has been developed which enables an excellent peak area determination in non-routine chromatograms with often changing mixtures. Because all raw data are at disposal the program dispenses with the real-time peak detection criteria used in dedicated computer system. In place of that, methods for peak determination — specially for base line construction — were developed which take into account all sample points of the chromatogram.

This software package consists of the following parts:

1. Noise analysis with the first hundred sample points of the chromatogram for determination of the mean square deviation
2. Elimination of spikes in the whole chromatogram
3. Construction of base line, dividing the chromatogram into sample point groups
4. Smoothing the raw data
5. Detection of peaks and shoulders
6. Checking the peak width
7. Separation of overlapping peaks and shoulders
8. Calculation of retention time

The parts 1 to 8 of the program running in this sequence will further be explained in this paper.

1. Noise Analysis

A fitted straight line is placed into the first hundred sample points which must be free from peaks and greater disturbances. The degree of noise is described as the mean square deviation of the sample points from the fitted line. After omitting all sample points, whose distance from the fitted line is greater than 2σ, the regression is repeated. σ, determined by this procedure, is the basis for many criteria in the consecutive steps of the software, specially in the parts 2, 3, and 5.

2. Elimination of spikes

To detect spikes with some certainty also at critical parts of the chromatogram a simple test as described in 1. is not sufficient. The following procedure has proved its utility in practice:

A frame (2 × 3 sample points) with an enclosed window (3 points) is moved through the chromatogram (Fig. 1). In each position a line inside the window is computed from the sample points of the frame and then compared with the actual three sample points of the window. For this purpose two fitted straight lines are placed into the three points of each part of the frame with the condition that the two straight lines intersect in the middle of the window. Sample points inside the window with a deviation of more than A from the straight line in question are defined to be a spike. A is calculated from the mean deviation σ, the mean deviation of the points in the frame from the fitted line σ_R and the difference Δm between the slopes of the two lines.

\[ A = \sqrt{\left(\sigma^2 + \sigma_R^2\right) \left(3.2 + \frac{\Delta m}{2\sigma}\right)^2} \]

3. Construction of Base Line

Two different kinds of base line construction can be chosen

a) consisting of a sequence of straight lines for isotherm chromatograms
b) consisting of a sequence of polynomials of second order chosen in temperature programmed chromatograms with a complicated base line.

First of all, in both cases appropriate fix points (base points) are searched for, which are then connected by
In a second run the largest differences between the straight lines and that parts of the chromatogram lying under the lines are located and taken as additional preliminary base points (6 and 7 in Fig. 2). By repeated application of the "rubber band method" the final base points are found.

With a more curved base line the case b) (parabolas) should be chosen. After searching for preliminary base points in an analog mode additional base points are placed into valleys suspected to be base line in order to improve the constructed base line in these sections (A and B in Fig. 2); the number of additional points is proportional to the length of the valleys.

Normally the additional base points are determined within the whole chromatogram in the same manner. The following special cases can — either separate or combined — be chosen:

a) Peaks to be determined are situated on the tailing of a solvent peak.

b) Chromatograms with temperature program and isothermal postrun.

In case a) roughly estimated values of the abscissa of the begin and the end of the peak tailing which is taken for base line must be given. Depending on the slope of the tailing these values are corrected. After dividing the chromatogram at the end of the tailing the base points are searched for in both parts of the chromatogram. By appropriate choice of parameters care has to be taken that the number of base points on the tailing is high enough for fitting the constructed base line to the real base line.

In case b) an approximate end of the temperature program is indicated by the user. This value is automatically revised, for example placed into a valley if the indicated endpoint lies on a peak. From the begin of the chromatogram to the corrected endpoint of the temperature program the base points are found in the usually manner, from this point to the end of the chromatogram, however, after new aspects. The base points determined in this region are finally revised by a "negative rubber band" criterion.