

Разделение перекрывающихся пиков при обработке масс-спектров на ЭВМ. Б е с ч а с т н о в О. И., З д а - н о в и ч В. В., С и н и ц ы н а Н. Л. — В кн.: Научное приборостроение. Л., «Наука», 1983, с. 122—124..

В работе рассматривается метод разделения перекрывающихся пиков, которые обнаружены предшествующей обработкой. Разделение основано на применении статистических методов при учете аппаратной функции. Окончательное разделение перекрывающихся пиков должно производиться по методу наименьших квадратов. Проведен численный эксперимент, который показал, что при наличии шумов и малом числе (5—6) дискретов, приходящихся на пик, исследуемый метод дает хорошую точность разделения наложившихся пиков. Лит. — 4 назв., ил. — 2.

АВТОМАТИЗАЦИЯ ПРИБОРОВ ДЛЯ НАУЧНЫХ ИССЛЕДОВАНИЙ

О. Н. Бесчастнов, В. В. Зданович, Н. Л. Смирнова

РАЗДЕЛЕНИЕ ПЕРЕКРЫВАЮЩИХСЯ ПИКОВ ПРИ ОБРАБОТКЕ МАСС-СПЕКТРОВ НА ЭВМ

В большинстве аналитических приборов сигналом, несущим информацию о пробе, является спектр, который показывает зависимость тока или напряжения на выходе прибора от времени. Чаще всего этот сигнал представляет собой последовательность пиков различной формы.

Если два пика достаточно далеко отстоят друг от друга (т. е. седловина между ними меньше половины амплитуды меньшего по высоте пика), то их можно считать неперекрывающимися и при обработке каждый из пиков рассматривать независимо. Минимальное перекрытие пиков, которое следует учитывать, определяется точностью измерения спектра. Наличие перекрытия пиков в спектре затрудняет, а в ряде случаев делает невозможной его интерпретацию. Поэтому разделение перекрывающихся пиков является важной задачей первичной обработки информации. Ее решение часто дает возможность получить дополнительную информацию о составе и структуре вещества, исследуемого с помощью аналитических приборов [1].

При решении проблемы разделения перекрывающихся пиков встают две задачи: во-первых, определить алгоритм, с помощью которого можно осуществить указанное разделение; во-вторых, построить подходящую модель спектра, получаемого с аналитического прибора, чтобы иметь возможность в отсутствие аналитического прибора исследовать выбранный алгоритм. Решению этих двух задач и посвящена данная статья.

Во многих случаях удается разделить масс-спектр на ряд неперекрывающихся участков, содержащих один пик или несколько наложившихся пиков. Назовем такие участки наблюдаемыми пиками. Очевидно, что каждый наблюдаемый пик можно рассматривать отдельно. Пусть $y(t)$, $t \in (a, b)$ — участок спектра, соответствующий одному наблюдаемому пику. Функция $y(t)$ может быть представлена в виде

$$y(t) = \int_{t-c}^{t+c} x(u) \rho(u-t) du, \quad (1)$$

где $\rho(u)$, $u \in (-c, c)$ — аппаратная функция прибора, предполагаемая для простоты одинаковой для всех точек данного участка спектра; $x(u)$, $u \in (a-c, b+c)$ — исходный сигнал до искажения аппаратной функцией.

Будем предполагать, что в процессе измерений производится дискретизация спектра по времени с шагом Δt , тогда результаты измерений представляются в виде последовательности

$$y_i = y(t_i), \quad t_i = a + \Delta t(i-1), \quad i = 1, 2, \dots, k.$$

В данной работе рассматривается задача компенсации влияния аппаратной функции и определения исходного сигнала $x(t)$. Целесообразно стремиться определить функцию $x(t)$ в тех же точках, в которых была измерена функция

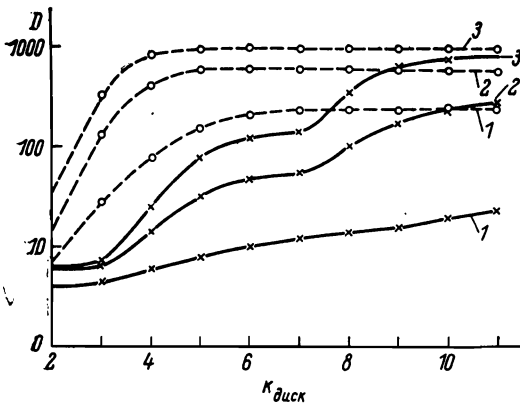


Рис. 1. Относительные ошибки восстановления формы сигнала при различном числе дискретов в исходном сигнале ($k_{\text{диск}}$) и аппаратной функции (1—3).

Число дискретов в аппаратной функции: 1 — 11; 2 — 21; 3 — 31. Сплошные линии — интегрирование по методу Симпсона, пунктир — по методу трапеций.

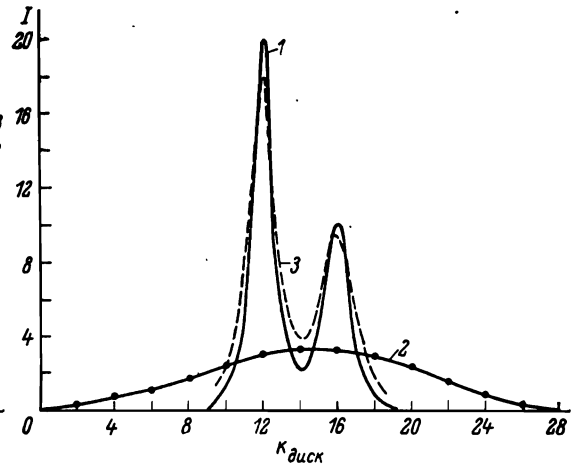


Рис. 2. Пример восстановления формы сигнала.

Сигналы: 1 — исходный; 2 — искаженный аппаратной функцией; 3 — восстановленный.

$y(t)$. Выберем величину c кратной шагу дискретизации Δt ; тогда интеграл (1) заменяется конечной суммой

$$y_i = \sum_{j=1}^s x_{i+j-1} \rho_j, \quad i = 1, 2, \dots, k, \quad (2)$$

где $s = 2c/\Delta t$, $x_i = x(t_i)$, ρ_j , $j = 1, 2, \dots, s$ — коэффициенты, определяемые принятой схемой численного интегрирования.

Так как в масс-спектрометрии спектры представляют собой неотрицательные функции и за пределами отрезка (a, b) функция $y(t) = 0$, то функция $x(t)$ из выражения (1) отличается от нуля только на участке $(a+c, b-c)$. Поэтому последовательность x_i , $i = 1, 2, \dots, k+s-1$ обладает следующим свойством: значения $x_i = 0$ при $s \geq i \geq k$. Таким образом, отличны от нуля только $k-s$ членов последовательности.

Учитывая это и производя соответствующую перенумерацию последовательности x_i , запишем систему уравнений (2) в матричной форме

$$RX = Y, \quad (3)$$

где R — матрица размером $(k-s) \times k$; X — искомый $(k-s)$ -мерный вектор; Y — известный k -мерный вектор.

Решая систему (3) с помощью метода наименьших квадратов, получим

$$X = (R^T R)^{-1} R^T Y. \quad (4)$$

Для оценки влияния шумов на точность вычисления по формуле (4) было использовано выражение [2]

$$P = (R^T R)^{-1} \sigma^2,$$

где P — ковариационная матрица ошибок определения вектора X ; σ^2 — дисперсия присутствующей в векторе Y случайной составляющей, в которой амплитуды имеют нормальное распределение, а значения статистически независимы.

Результаты расчета относительной среднеквадратичной ошибки D , вычисленной для одной координаты вектора X в зависимости от числа дискретов в том же векторе, приведены на рис. 1. Можно видеть, что размерность вектора X должна быть небольшой (5—6) и повышение точности численного интегрирования при применении метода Симпсона является весьма существенным, так как обратная матрица из (4) получается плохо обусловленной.

Для проверки работоспособности рассматриваемого метода разделения пиков был создан алгоритм, имитирующий сигнал с перекрывающимися пиками на выходе масс-спектрометра.

Форма пика выбиралась гауссовской [3]:

$$A = A_m^{(i)} \exp[-(x - m_i)^2/2G^2], \quad (5)$$

где $A_m^{(i)}$ — амплитуда i -го пика; m_i — положение i -го пика в спектре; G — коэффициент формы пика; x — текущее значение по оси абсцисс спектра.

Параметры $A_m^{(i)}$, m_i , G и число пиков могут задаваться оператором произвольно. Суммарный дискретизированный спектр получается суммированием значений пиков на заданном наборе дискретов спектра. Так как гауссовский пик имеет длинные «хвосты», то для ограничения времени вычислений значений дискретизированного спектра вводился порог V и получаемые по формуле значения A , меньшие, чем порог, не участвовали в создании спектра. Сильно перекрывающиеся пики получаются при малой разности $m_{i+1} - m_i$ и небольшом коэффициенте G .

На основании изложенного метода создан алгоритм разделения перекрывающихся пиков, реализованный на языке Фортран-4 для ЭВМ СМ-4. Численный эксперимент был проведен при отсутствии шума и следующей исходной информации: $m_i = 12$, $m_{i+1} = 16$, $A_m^{(i)} = 20$, $A_m^{(i+1)} = 10$, $G = 0.24$.

Его результаты приведены на рис. 2. Из рис. 2 видно, что погрешность восстановления в отсутствие шума пренебрежимо мала. При наличии шума погрешность оказывается небольшой лишь при малой размерности вектора X .

Таким образом, предложенный метод разделения перекрывающихся пиков обеспечивает хорошее разделение при малом числе точек в результирующем сигнале, в то время как существующие методы [4] хорошо разделяют пики только при наличии большого числа точек в сигнале.

ЛИТЕРАТУРА

1. Meerwall E. von. — Comp. Phys. Commun., 1975, v. 9, N 2, p. 117.
2. Лу Р. Оптимальные оценки, определение характеристик и управление. М., 1966.
3. Gladney H. M., Dowden B. F., Swalen J. D. — Analyt. Chem., 1969, v. 41, N 7, p. 883.
4. Силмс Я. Я., Кофман А. М., Розенблит А. Б. Первичная обработка хроматограмм и спектров на ЭВМ. Принципы. Алгоритмы. Программы. Рига, 1980.